Intrinsically Motivated Graph Exploration Using Network Theories of Human Curiosity

Shubhankar P. Patankar* **Mathieu Ouellet*** **Juan Cerviño** **Alejandro Ribeiro**
Kieran A. Murphy **Dani S. Bassett**
University of Pennsylvania
{spatank,ouellet,jcervino,aribeiro,kieranm,dsb}@seas.upenn.edu

Abstract

Intrinsically motivated exploration has proven useful for reinforcement learning, even without additional extrinsic rewards. When the environment is naturally represented as a graph, how to guide exploration best remains an open question. In this work, we propose a novel approach for exploring graph-structured data motivated by two theories of human curiosity: the information gap theory and the compression progress theory. The theories view curiosity as an intrinsic motivation to optimize for topological features of subgraphs induced by nodes visited in the environment. We use these proposed features as rewards for graph neural-network-based reinforcement learning. On multiple classes of synthetically generated graphs, we find that trained agents generalize to longer exploratory walks and larger environments than are seen during training. Our method computes more efficiently than the greedy evaluation of the relevant topological properties. The proposed intrinsic motivations bear particular relevance for recommender systems. We demonstrate that next-node recommendations considering curiosity are more predictive of human choices than PageRank centrality in several real-world graph environments.

1 Introduction

Providing a task-agnostic incentive for exploration as an intrinsic reward has proven useful for reinforcement learning, even in the absence of any task-specific (extrinsic) rewards [1, 2]. Termed *curiosity* in reference to the analogous drive in humans, prior formulations are based on different means of quantifying the novelty or surprisal of states encountered by an agent [3]. If states are represented as graphs, the task-agnostic motivation to explore can additionally be content-agnostic, depending only on the topological properties of the visited state subgraph. Leading theories of human curiosity are similarly content-agnostic, based only on the structural properties of a relational graph connecting atoms of knowledge without regard to their actual content [4].

Theories of curiosity seek to describe the intrinsic motivations that underlie human decision-making when acquiring information through exploration. The *information gap theory* (IGT) argues that curiosity collects knowledge to regulate gaps in our understanding of the world [5]. Exposure to a small amount of novel information pushes an individual's uncertainty about the environment past an acceptable threshold, creating an information gap. Curious agents are driven to resolve the discrepancy by acquiring information to close the gap [6, 7]. An alternative account, the *compression progress theory* (CPT), posits that information-seeking behavior is motivated to build increasingly compressible state representations [8, 9]. Compression enables abstraction and improved generalization by emphasizing the essential latent structures of knowledge [10–12]. Both theories provide optimization objectives for the human exploration of graph-structured environments.

*Equal contribution. Code available at: https://github.com/spatank/curiosity-graphs.

Patankar et al., Intrinsically Motivated Graph Exploration Using Network Theories of Human Curiosity. *Proceedings of the Second Learning on Graphs Conference (LoG 2023)*, PMLR 231, Virtual Event, November 27–30, 2023.

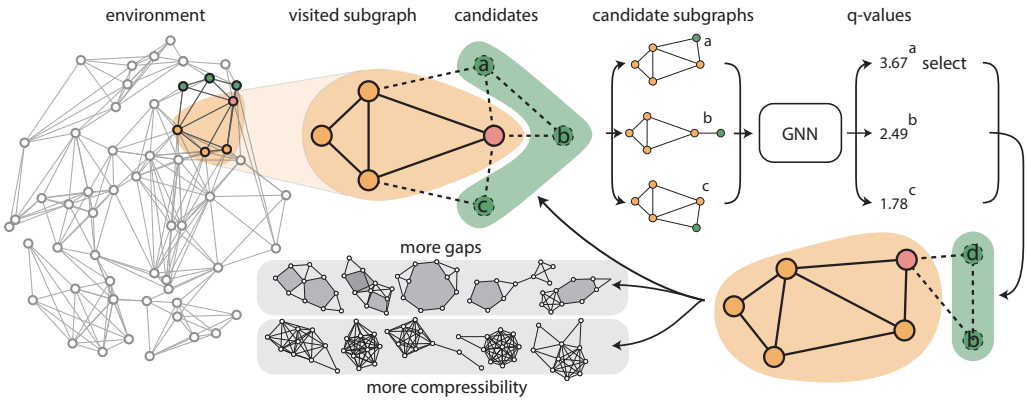

Figure 1: Neural network for graph exploration. The subgraph induced by the set of currently visited nodes is denoted in orange. Candidate nodes to visit at the next time step are denoted in green. We build candidate subgraphs by adding each neighbor to the already visited subgraph. The candidates are processed with a GNN to obtain Q-values, denoting their long-term potential to create or close gaps or to improve compressibility. Two example trajectories are shown: one with a high number of gaps and one with greater compressibility.

In this work, we demonstrate that network theoretic measurements of information gaps and compression progress can be meaningful exploration incentives for graph neural network (GNN)-based reinforcement learning (RL). Here, similar to human curiosity, which is typically conceived of as being non-instrumental, and unlike traditional RL formulations, where artificial curiosity is a means to an end, exploration itself is the broader goal. To that end, we train GNN agents to explore graph-structured environments while optimizing for gap creation and improved compression (Figure 1). Additionally, we use GNNs trained with human curiosity rewards to modify PageRank centrality. We measure an alternative form of PageRank by biasing underlying random walks towards nodes that create information gaps or improve network compressibility. We use human exploration trajectories acquired from spaces that can be naturally represented as graphs—movies, books, and Wikipedia—to evaluate predictions of user choices made by PageRank against those made by our new metric.

Our primary contributions are the following:

- We adapt intrinsic motivations for human curiosity as reward functions for reinforcement learning.
- We replace expensive reward computations with graph neural networks. Subject to training costs, our method is computationally efficient and generalizes to longer exploratory walks and larger environments than are seen during training.
- We demonstrate that incorporating curiosity into PageRank centrality leads to better predictions of human preferences compared to standard PageRank.

2 Related work

Human curiosity as graph exploration. Curiosity in humans is conceptualized as the intrinsic motivation to gather information from the environment [5, 13, 14]. Humans acquire information even when it is expensive [15, 16] and may have no tangible utility [17, 18], suggesting that exploration is inherently valuable. Recent work has expanded the traditional knowledge acquisition perspective on curiosity by also considering how units of knowledge relate to each other. This perspective defines curiosity as an exploratory walk on a graph. Here, curiosity entails building a growing knowledge network by acquiring informational units as nodes and their relationships as edges [4, 19]. The state of an individual's knowledge is viewed as the subgraph of the environment induced by the visited nodes [20, 21]. Under this formulation, humans explore Wikipedia via trajectories with fewer information gaps and greater network compressibility than relevant null models [21].

Intrinsic motivations in reinforcement learning. The need for improved exploration has led reinforcement learning to incorporate curiosity-like intrinsic motivations into its algorithmic framework [22, 23]. Exploration rewards in RL take several forms. At the core of all approaches is an inducement for the learning agent to seek novelty. Count-based approaches encourage visits to unfamiliar or infrequently visited states [24–28]. When the state space is large, enumerating the frequencies

of visits to all possible states is expensive. To overcome this challenge, density models derive uncertainty-based pseudo-counts [24, 25]. A complementary perspective emphasizes model building and formulates curiosity in terms of learning progress and surprisal [1, 9, 29–32]. For instance, in the prediction error approach—alongside an extrinsic task—the agent attempts to learn a model of the environment's dynamics. Curiosity rewards are proportional to the error when predicting transitions between states. Memory-based methods assign rewards considering how different a newly visited state is from those stored in memory [33, 34]. Instead of a prescriptive approach, parametric methods attempt to explicitly learn an intrinsic reward function [35–38]. In general, improved exploration is a means to an end, with intrinsic rewards supplementing extrinsic task-specific rewards.

Graph combinatorial optimization and reinforcement learning. Combinatorial optimization entails selecting elements from a finite set of options such that the chosen subset satisfies an objective function [39]. Graph analyses often involve combinatorial optimization, with graph structure imposing constraints on the solution space. Recent work combines graph neural networks and reinforcement learning to construct solutions by incrementally adding nodes to a partial set [40–42]. First, a GNN constructs an embedding for the candidate solution; second, an agent, for instance, a deep Q-network (DQN), trained via RL, selects an action to expand the solution [43]. The two networks can be trained end-to-end with an optimization objective driving gradients for learning. This approach solves various graph combinatorial tasks, such as the traveling salesperson problem [43–45], finding the maximum independent set [46], or the minimum vertex cover [43, 47], and identifying isomorphic subgraphs [48]. Instead of uncovering nodes, GNNs can also sequentially collapse nodes into each other with implications for matrix multiplication [49]. Recent work has also sought to formulate the graph exploration task explicitly as a Markov decision process, using domain-specific node features and novelty rewards [50, 51]. GNNs, in combination with RL, have also been used to build and rewire graphs such that they possess high values of specific features of interest [52, 53].

PageRank and human navigation on graphs. PageRank seeks to model and predict online human browsing preferences. PageRank assigns centrality scores to web pages, considering their importance and relevance, determined by the number and quality of links between pages [54]. It employs a random walks-based approach that includes occasional jumps, known as teleportation, to simulate the likelihood that users will transition between different pages [55]. Beyond its initial application to serving search results, PageRank has found broad utility in modeling human navigation in other graph-structured environments [54]. PageRank-based recommendations show alignment with empirical observations of human behavior [56–59]. Further, topological characteristics of the underlying graph impact user navigation, underscoring the importance of considering connectivity patterns when forecasting noisy human preferences [60]. The PageRank algorithm explicitly factors topology into its computations by adjusting the underlying random walk process, aligning either the transition or teleportation weights with the environment's topological features [57, 61].

3 Methods

Our goal is to train an agent to explore while optimizing for a structural property of the visited subgraph. Consider a graph-structured environment $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node set $\mathcal{V}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Let $\mathcal{V}_T = \{v_1, v_2, \cdots, v_T\} \subseteq \mathcal{V}$ be an ordered set of explored nodes at time T. The corresponding subgraph trajectory is the sequence $\mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_T$, wherein the t-th subgraph $\mathcal{S}_t$ is induced by the first t visited nodes. Specifically, given the graph $\mathcal{G}$, the number of nodes to visit T, a graph feature function $\mathcal{F} : 2^{\mathcal{G}} \to \mathbb{R}$, and a discount factor $\gamma \in [0, 1]$, we seek an ordered set $\mathcal{V}_T^*$ such that $\sum_{t=1}^{T} \gamma^{t-1} \mathcal{F}(\mathcal{S}_t)$ is maximal. The function $\mathcal{F}$ acts as an intrinsic reward to encourage exploration. The discounting parameter determines the extent to which future values of $\mathcal{F}$ factor into the decision-making at each step. Drawing inspiration from human curiosity, we adopt information gap theory and compression progress theory to design two reward functions, $\mathcal{F}_{IGT}$ and $\mathcal{F}_{CPT}$.

3.1 Network theories of curiosity

Information gap theory views human curiosity as an intrinsic motivation to regulate gaps in knowledge. Exposure to new information pushes the level of uncertainty about the environment past an acceptable threshold, creating an uncertainty gap. Curiosity seeks to find information units to close this gap. By modeling the state of knowledge as a graph, we can characterize information gaps as topological cavities. In a graph, cavities can take several forms: dimension 0 cavities represent disconnected network components, whereas those of dimension 1, known as 1-cycles, represent

non-triangular loops of edges (Figure 2A). In order to identify and count topological cavities, a graph is first converted into a higher-order relational object known as a *simplicial complex* [62]. A simplicial complex is comprised of simplices. Geometrically, a d-simplex is a shape with flat sides formed by connecting $d+1$ points. For $0 \leq d \leq 2$, by definition a node is a 0-simplex, an edge is a 1-simplex, and a filled triangle is a 2-simplex. We can construct a simplicial complex by assigning a d-simplex to each $(d+1)$-clique in a binary graph. In a simplicial complex, a d-dimensional topological cavity is identified as an enclosure formed by d-simplices that cannot be filled by a higher-dimensional simplex. We refer the reader to Refs. [63–67] for more details on algebraic topology.

Given a simplicial complex, the d-th *Betti number* β_d counts the number of topological gaps of dimension d. Prior work examining human curiosity finds compelling evidence in support of information gap theory. In particular, humans create induced subgraphs with an increasing number of 1-dimensional cavities [21]. Therefore, in this work, at each time step t with a visited subgraph $\mathcal{S}_t$, we assign rewards equal to β_1, that is, we set $\mathcal{F}_{IGT} = \beta_1(\mathcal{S}_t)$.

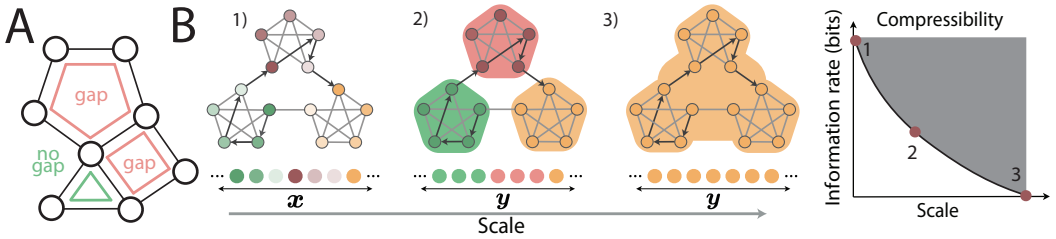

Figure 2: Quantifying network theories of human curiosity. *(A)* Gaps or cavities in a graph can be formalized using algebraic topology. A 1-dimensional cavity, also known as a 1-cycle, is a non-triangular loop of edges. *(B)* The information rate of a random walk $\boldsymbol{x}$ on a graph is given by its entropy. If we cluster the nodes, the walk sequence $\boldsymbol{x}$ is compressed into a new sequence $\boldsymbol{y}$, where y is the cluster that contains node x. The new sequence has a lower information rate than the original sequence. The number of clusters defines the scale at which the network is described. We can find an optimal clustering at every scale of description that maximally lowers the information rate. These values can be recorded in a rate-distortion curve. Network compressibility is the maximal reduction in the information rate, averaged across all scales. Graphically, this value represents the area above the rate-distortion curve bounded by the entropy of the unclustered random walk.

Compression progress theory posits that curiosity is a drive to compress the state of knowledge [8]. During graph exploration, at each step t in a trajectory, a reward for compression can be assigned as network compressibility [68]. Consider a subgraph $\mathcal{S}_t$ with t nodes and q edges, represented by a symmetric adjacency matrix $M \in \mathbb{R}^{t \times t}$. Information about the subgraph's structure can be encoded in the form of a random walk $\boldsymbol{x} = (x_1,\ x_2,\ \dots)$. The walk sequence is generated by randomly transitioning from a node to one of its neighbors. Thus, for a random walk on $\mathcal{S}_t$, the probability of transitioning from node i to node j is $P_{ij} = M_{ij} / \sum_j M_{ij}$. Since the walk is Markovian, its information content (or *entropy*) is given by $H = -\sum_i \pi_i \sum_j P_{ij} \log P_{ij}$. Here, π_i is the stationary distribution representing the long-term probability that a walk arrives at node i, given by $\pi_i = \sum_j M_{ij}/2q$.

Assigning nodes to clusters leads to a coarse-grained sequence $\boldsymbol{y} = (y_1,\ y_2,\ \dots)$. The number of clusters n can be used to define a scale of the network's description $s = 1 - \frac{n-1}{t}$. When $n = t$, the network is described at a fine-grained scale $s = 1/t$; at the other extreme, when $n = 1$ the network is described at the coarsest scale $s = 1$. At every description scale in between, it is possible to identify a clustering of nodes that minimizes the information rate (Figure 2B). After computing these optimal clusterings across all scales, we arrive at a rate-distortion curve $R(s)$, representing a bound on the information rate as a function of the scale s. The compressibility C of the network is then given as the average reduction in the information rate across all scales [68], $C = H - \frac{1}{t}\sum_s R(s)$. Human curiosity in graph-structured environments leads to induced subgraphs with increasing network compressibility [21]. Therefore, we assign compression rewards as $\mathcal{F}_{CPT} = C(\mathcal{S}_t)$, where $C(\mathcal{S}_t)$ denotes the compressibility of subgraph $\mathcal{S}_t$.

3.2 Reinforcement learning for graph exploration

We formulate the graph exploration problem as a Markov decision process (MDP) [69]:

- **States**: The state is defined as the subgraph induced by the visited nodes at time t, $\mathcal{S}_t = \mathcal{G}[\mathcal{V}_t]$. We specify the initial state $\mathcal{S}_1$ by randomly selecting a starting node $v_1 \in \mathcal{V}$. Each state represents a partial solution to the broader sequential exploration task.
- **Actions**: The agent can transition to a neighbor of the most recently visited node. We denote the neighborhood of a node v as $\mathcal{N}(v) = \{u \in \mathcal{V} \mid (v, u) \in \mathcal{E}\}$. Therefore, given the state at time t, the set of available next nodes is $\mathcal{A}(\mathcal{S}_t) = \mathcal{N}(v_t) \backslash \mathcal{V}_t$. If no nodes are available in the immediate neighborhood, we expand the action set to include all neighbors of the explored subgraph.
- **Transitions**: Given the pair $\mathcal{S}_t$ and $v \in \mathcal{A}(\mathcal{S}_t)$, the transition to state $\mathcal{S}_{t+1}$ is deterministic with $P\left(S_{t+1} \mid S_t, v\right) = 1$.
- **Rewards**: The reward at time t is defined as $R_t = \mathcal{F}(\mathcal{S}_t)$. Considering information gap theory, we reward agents for visiting nodes that create 1-cycles ($\mathcal{F}_{IGT}$). Considering compression progress theory, we reward agents for visiting nodes that improve network compressibility ($\mathcal{F}_{CPT}$).

The policy $\pi(v \mid \mathcal{S}_t)$ maps states to actions, fully describing the agent's behavior in the environment. At each step, the agent makes decisions using a value function $Q(\mathcal{S}_t, v)$, which evaluates candidate nodes $v \in \mathcal{A}(\mathcal{S}_t)$ in the context of the currently explored subgraph $\mathcal{S}_t$. The function measures the total (discounted) reward that is expected to accumulate if the agent selects action v in state $\mathcal{S}_t$ and thereafter follows policy π. In turn, the policy can be viewed as behaving greedily with respect to the value function, $\pi = \arg\max_{v \in \mathcal{A}(\mathcal{S}_t)} Q\left(\mathcal{S}_t, v\right)$. Solving an MDP entails finding an optimal policy that maximizes the expected discounted sum of rewards.

We parameterize the value function Q using a GNN $\Phi(\cdot) : \mathcal{G} \rightarrow \mathbb{R}$. GNNs build vector embeddings for nodes by iteratively aggregating their features with those from their local neighborhoods [70]. Each aggregation step is typically followed by a fully connected layer and a non-linear activation function. Depending on the number of rounds of aggregation, features from more distant locations in the graph can inform the embedding for each node. Specifically, we use the *GraphSAGE* architecture [71], where at the l-th round of feature aggregation, the embedding for node u is given as,

$$h_u^{(l)} = f^{(l)}\left(h_u^{(l-1)}, h_{\mathcal{N}(u)}^{(l-1)}\right) = g\left[\theta_C^{(l)} h_u^{(l-1)} + \theta_A^{(l)} \tilde{A}\left(h_{\mathcal{N}(u)}^{(l-1)}\right)\right], \tag{1}$$

where $\tilde{A}$ represents the aggregation operator, $g\left[.\right]$ is the activation function, and θ_C and θ_A are parameters for combination and aggregation, respectively [42, 71]. Our choice of GraphSAGE is motivated not by its sampling-based approach but rather by its capacity for inductive learning. Therefore, we do not perform neighbor sampling during feature aggregation. Further, considering that leading network theories of human curiosity are agnostic to the precise content of individual nodes, we only use the local degree profile (LDP) of each node as the initial feature set [72]. LDP comprises features of a node's neighborhood, including its degree, the minimum and maximum degrees of its neighbors, and the average and standard deviation of the degrees of its neighbors.

We train GNNs for exploration using the DQN algorithm, with a replay buffer for experience sampling, a target network, and a decaying ϵ-greedy exploration rate [73]. Details of the full neural network architecture and the training process are included in the Supplement.

3.3 Curiosity-biased node centrality

Several graph theoretical quantities can be defined in terms of random walk processes. We can use agents trained to explore graphs to bias random walkers and, by extension, the corresponding quantities. PageRank is a widely recognized algorithm that assigns node centrality scores to graph data [54, 55, 74, 75]. The per-node score η can be interpreted as the stationary distribution of a random walk process on a network. With probability α, a random walker moves along an edge from node v_i to one of its neighbors. The probability of reaching a connected node v_j is P_{ij}. Alternatively, with probability $1 - \alpha$, the walker jumps, or teleports, to a random node in the network. The probability of jumping to node v_k is q_k. Under conditions of irreducibility and aperiodicity [76], the stationary distribution is given as

$$\sum_i (I - \alpha P_{ij}^t)\eta_i = (1 - \alpha) q_j. \tag{2}$$

The PageRank algorithm follows a random walk that is entirely Markovian. Typically, the probability P_{ij} depends solely on the out-degree of node v_i and, in the case of node-weighting, on the vector q. Personalized PageRank biases the random walk process using q_k by taking into account nodes that are already visited in the network [77].

Adaptations of PageRank often incorporate biases in the weighting scheme or the teleportation mechanism to better predict human navigation in graph-structured environments [56–61]. We can integrate agents trained to optimize for the exploration objectives described earlier into the PageRank algorithm in the form of similar biases. Specifically, given an already visited subgraph, we propose to modify transition probabilities using the Q-values assigned to candidate nodes. Consider a non-Markovian random walker sitting at node v_l with a path history $V_l = \{v_1, \cdots, v_{l-1}, v_l\}$. The visited nodes in the path induce a corresponding subgraph $\mathcal{S}_l$. Paths are built starting from the most recent initialization or teleportation event. We use a Q-value function trained to optimize for an objective $\mathcal{F}$ to bias the walker. The transition probability from node v_l to node v_m can be re-defined as,

$$P^{\mathcal{F}}_{lm}(\mathcal{S}_l) \equiv \begin{cases} \frac{(1-p_g)p_g^{rank(Q(\mathcal{S}_l, v_m))-1}}{1-p^{|\mathcal{A}(\mathcal{S}_l)|}}, & v_m \in \mathcal{A}(\mathcal{S}_l), \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

where $rank(Q(\mathcal{S}_l, v_m))$ is the rank for v_m considering the Q-values for the candidate nodes and $p_g \in [0, 1]$ is a parameter that controls how likely the walker is to select nodes greedily. To compute biased per-node PageRank values, we simulate a walker using $P^{\mathcal{F}}_{ij}(\mathcal{S}_i)$ until probabilities converge.

4 Experiments

4.1 Exploration in synthetically generated networks

We train a curiosity-based GNN agent to explore synthetically generated graph environments that exhibit a broad range of degree profiles and topologies [78, 79]. We examine synthetic networks generated using the random geometric (RG), Watts-Strogatz (WS), Barabási-Albert (BA), and Erdös-Rényi (ER) graph models. Details surrounding the generation process are available in the Supplement. For each of the four graph models, we build 100 training, 10 validation, and 10 testing environments. Each environment is constructed to have $N = 50$ nodes. Each episode lasts for 10 steps and, therefore, consists of visits to 10 distinct nodes. After training, we evaluate the GNN agent in the testing environments against four baseline approaches:

- Random: Select a candidate node at random.
- Greedy: For each candidate node, build a candidate state subgraph. Evaluate the reward function for each subgraph and select the node that results in the biggest one-step improvement.
- Max Degree: Select the candidate node with the largest degree.
- Min Degree: Select the candidate node with the smallest degree.

$\mathcal{F}$	$\mathcal{G}$	Random	Max Degree	Min Degree	Greedy	GNN
IGT	RG	$0.312_{\pm 0.034}$	$0.010_{\pm 0.007}$	$0.144_{\pm 0.027}$	$1.495_{\pm 0.079}$	$\mathbf{2.308_{\pm 0.092}}$
	WS	$1.141_{\pm 0.068}$	$1.048_{\pm 0.068}$	$1.586_{\pm 0.082}$	$2.707_{\pm 0.103}$	$\mathbf{3.303_{\pm 0.106}}$
	BA	$7.593_{\pm 0.145}$	$2.565_{\pm 0.083}$	$3.932_{\pm 0.115}$	$19.332_{\pm 0.206}$	$\mathbf{21.970_{\pm 0.169}}$
	ER	$9.197_{\pm 0.144}$	$9.638_{\pm 0.162}$	$4.953_{\pm 0.127}$	$\mathbf{25.20_{\pm 0.164}}$	$24.058_{\pm 0.183}$
CPT	RG	$8.607_{\pm 0.027}$	$8.928_{\pm 0.027}$	$7.864_{\pm 0.033}$	$\mathbf{9.615_{\pm 0.014}}$	$9.271_{\pm 0.017}$
	WS	$7.117_{\pm 0.021}$	$6.788_{\pm 0.025}$	$6.937_{\pm 0.021}$	$\mathbf{7.668_{\pm 0.012}}$	$7.174_{\pm 0.014}$
	BA	$6.926_{\pm 0.023}$	$8.526_{\pm 0.015}$	$5.899_{\pm 0.016}$	$\mathbf{8.669_{\pm 0.016}}$	$8.556_{\pm 0.010}$
	ER	$6.767_{\pm 0.020}$	$6.931_{\pm 0.019}$	$6.022_{\pm 0.017}$	$\mathbf{8.262_{\pm 0.016}}$	$7.880_{\pm 0.015}$

Table 1: Performance of GNN-based agents using information gap theory (IGT) and compression progress theory (CPT) compared to four baseline methods (random, max degree, min degree, greedy). We compare results using the total average return gathered by agents in four types of synthetic graph environments (random geometric - RG, Watts-Strogatz - WS, Barabási-Albert - BA, Erdős-Rényi - ER).

The total average reward gathered by the different agents is presented in Table 1. For the IGT reward, in all graph models except for ER, the GNN outperforms the greedy agent. By contrast, the one-step-ahead greedy agent consistently performs best for CPT, with the GNN a close second. Baseline approaches broadly perform well compared to the GNN for CPT than they do for IGT. When exploring a graph with the IGT objective, adding a single node can close several topological gaps simultaneously, requiring careful consideration of options. By contrast, compressibility is less

sensitive to the choice of node at each step due to its strong correlation with the clustering coefficient [68]. If exploring inside a cluster, neighbors of a node are likely to be neighbors of each other, lowering the likelihood that a single choice will significantly alter long-term network compressibility. For instance, the max degree baseline performs well for the CPT objective in random geometric graphs because high-degree nodes are centrally placed and surrounded by dense, highly clustered neighborhoods [79]. Barabási-Albert graphs, similarly, have highly clustered cores due to preferential attachment in their generative process [80]. Watts-Strogatz networks have high clustering when the edge rewiring probability is low. As a result, even random exploration in such topologies tends to occur inside clusters leading to greater compressibility. In support of this view, the minimum degree baseline, which is likely to select a node outside of a cluster, is typically further apart from the performance of the GNN compared to the other baselines.

4.1.1 Trajectory length and environment size generalization

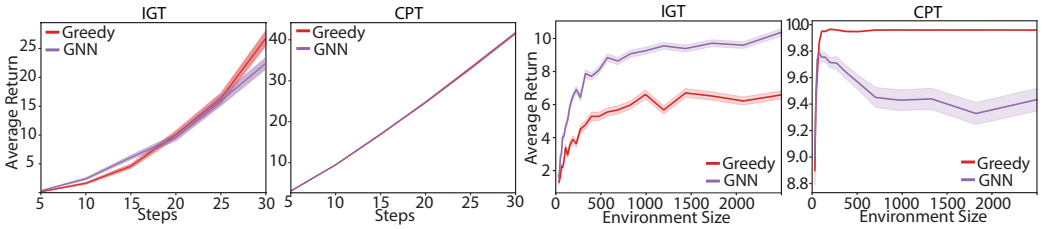

Figure 3: Trajectory length and environment size generalization. GNNs trained for graph exploration generalize to shorter and longer trajectories and to smaller and larger environments than are seen during training. We train GNNs to explore 10 steps for IGT and CPT in random geometric environments with 50 nodes. Performance does not degrade for exploratory walks of a different length in 50-node environments. Similarly, when taking 10 steps, GNN-based agents outperform or match the greedy agent in smaller and larger environments than those of size 50 that are seen during training. Bands denote standard error.

After training the GNN agent to explore 10 nodes in RG graph environments with 50 nodes, we evaluate generalization performance for longer trajectories and larger environments. We test trajectory length generalization while holding environment size fixed at 50 nodes. For walks shorter and longer than 10 steps, the GNN performs comparably to the greedy agent for both IGT and CPT (Figure 3). Next, we test environment size generalization by evaluating 10-step walks in graph environments that are larger than 50 nodes. In environments that are up to two orders of magnitude larger than those seen during training, the GNN is consistently superior to the greedy agent for IGT and exhibits comparable performance for CPT (with an average return of 9.4 compared to 10). In summary, the performance of trained GNNs does not degrade for settings outside the training regime. These results indicate that we can train GNNs for graph exploration in regimes where reward computations are relatively inexpensive due to the smaller size of subgraphs and expect them to scale to longer walks and larger networks.

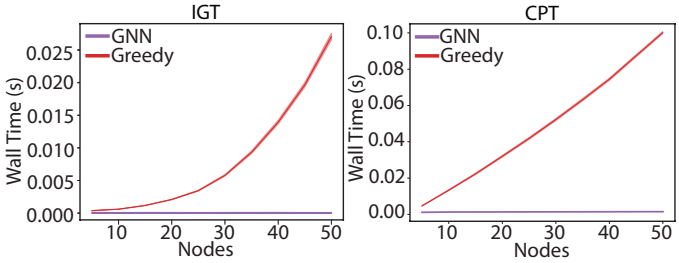

Figure 4: Wall time. Wall time for a forward pass through the GNN compared to the greedy evaluation of rewards. Bands denote standard error over computations for 50 networks.

4.1.2 Time complexity

Using graphs of different sizes, we evaluate the computational efficiency of our approach by comparing the wall time for a forward pass through the GNN with that for a greedy evaluation of the two

reward functions. Figure 4 displays results for RG graphs. Wall time for greedy evaluation of the IGT and CPT objectives grows quickly with subgraph size, while the GNN offers a faster alternative. Calls to $\mathcal{F}_{IGT}$ scale according to $O(|\mathcal{S}_t|^2)$ [81], whereas those to $\mathcal{F}_{CPT}$ scale according to $O(|\mathcal{S}_t|^3)$. By contrast, the computational complexity for the GNN grows linearly as $O(|\mathcal{S}_t|)$. Comparing the rewards for the two theories of curiosity, the information gap reward is significantly cheaper to evaluate compared to network compressibility. Therefore, in addition to approximating human intrinsic motivations for exploration, we find that the GNN offers a route to efficient computation of meaningful topological features of graphs.

4.2 Predicting human choices during graph navigation

Next, we evaluate the utility of curiosity-trained agents for predicting human choices in graph-structured environments. We gather human trajectories of graph exploration from three real-world datasets: MovieLens [82], Amazon Books [83, 84], and Wikispeedia [85, 86]. Each dataset can be naturally represented as a graph-structured environment. Details on how we process the data are available in the Supplement. We train GNNs for graph exploration in each environment for both information gap theory and compression progress theory. We use GNNs trained with curiosity rewards to bias PageRank centrality to predict next nodes visited by humans.

To incorporate person-specific data when computing PageRank, we modify the hop vector q to be zero for all nodes except a user's $n_{\text{burn-in}}$ most recently visited nodes [77]. We assign a uniform jump probability to the $n_{\text{burn-in}}$ nodes, with $q_k = 1/n_{\text{burn-in}}$. Each graph feature function $\mathcal{F}$ yields a PageRank vector $\eta_i^{\mathcal{F}}$. We combine these vectors linearly to obtain a final PageRank vector, denoted as η' such that $\eta' \equiv \tilde{\beta}\eta_{\text{PR}}(\alpha) + \tilde{\gamma}\eta_{\text{IGT}}(\alpha) + \tilde{\delta}\eta_{\text{CPT}}(\alpha)$ where $\tilde{\beta}^2 + \tilde{\gamma}^2 + \tilde{\delta}^2 = 1$ and η_{PR} is the score vector obtained using standard PageRank. To evaluate this approach, we optimize the set of variables $\alpha, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta}$ using a training set of transitions. We then compare performance against unbiased PageRank, where only α is optimized. We split the set of user trajectories acquired from each dataset into $\mathbf{S}_{\text{test}}$ and $\mathbf{S}_{\text{train}}$. These sets consist of portions of human trajectories with a length of $n_{\text{burn-in}} + 1$. Next, we perform Bayesian optimization to compute parameters $\hat{a}$ and $\hat{a}_{\text{bias}}$ for the two sets,

$$\hat{a} \equiv \arg\max_{\alpha} \sum_{S \in \mathbf{S}_{\text{train}}} \text{rank}_{v_{\text{burn-in}}}(\eta_{\text{PR}}(\alpha)) \tag{4}$$

$$\hat{a}_{\text{bias}} \equiv \arg\max_{\alpha, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta}} \sum_{S \in \mathbf{S}_{\text{train}}} \text{rank}_{v_{\text{burn-in}}}(\eta'(\alpha, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta})). \tag{5}$$

To evaluate our method, we calculate the ratio of improvement on the test set, given as

$$r_{\mathbf{S}_{\text{test}}} \equiv \sum_{S \in \mathbf{S}\text{test}} \text{rank}_{v_{\text{burn-in}}}(\eta'(\hat{a}_{\text{bias}})) / \sum_{S \in \mathbf{S}\text{test}} \text{rank}_{v_{\text{burn-in}}}(\eta_{\text{PR}}(\hat{a})). \tag{6}$$

Table 2 displays $r_{\mathbf{S}_{\text{test}}}$ in percentage terms for the three datasets when considering curiosity theories alone or in combination. Across all combinations, improvement ranges from 2.9% to 32.2%, indicating that incorporating curiosity for the biasing of walks is useful. The IGT or CPT-trained agents perform better with roughly similar values depending on the dataset. In the Wikispeedia data, however, CPT leads to improvement nearly four times higher than IGT. The books and movie datasets exhibit similarities since the selection mechanism in both environments is not directed towards a goal. By contrast, the Wikispeedia dataset involves goal-directed navigation. We provide illustrative examples from the MovieLens dataset of user paths alongside predictions made by both unbiased and curiosity-biased PageRank in the Supplement.

Graph dataset $\mathcal{G}$	IGT	CPT	IGT + CPT
MovieLens	$+4.2\%_{\pm 2.1\%}$	$+5.1\%_{\pm 1.5\%}$	$\mathbf{+7.9\%_{\pm 1.7\%}}$
Amazon Books	$+5.4\%_{\pm 1.9\%}$	$+4.6\%_{\pm 1.6\%}$	$\mathbf{+9.6\%_{\pm 1.9\%}}$
Wikispeedia	$+2.9\%_{\pm 2.9\%}$	$+12.2\%_{\pm 3.2\%}$	$\mathbf{+32.2\%_{\pm 7.7\%}}$

Table 2: Percentage improvement ($r_{\mathbf{S}_{\text{test}}}$) with curiosity-biased centrality for the MovieLens, Amazon Book, Wikispeedia datasets.

Figure 5B shows the improvement in predicting the transitions made by humans in the Wikispeedia dataset. We compare percentile ranks for each transition made by the human when making predictions

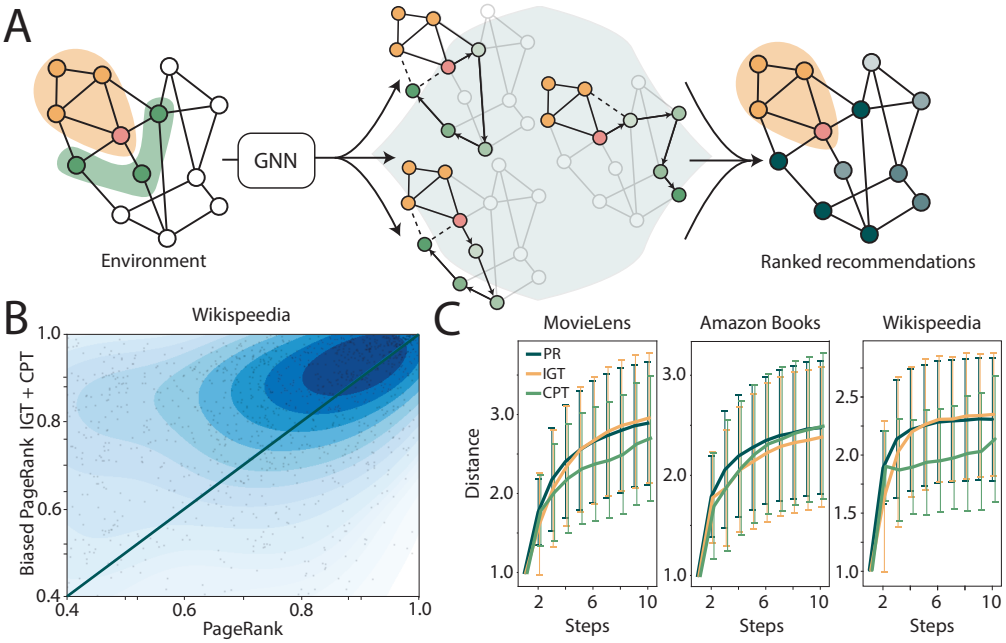

Figure 5: Re-defining centrality using agents trained for curiosity. *(A)* We measure curiosity-biased PageRank centrality using a set of biased walkers that explore the graph starting from a subset of already visited nodes. Biases are incorporated using GNNs trained for IGT and CPT rewards. *(B)* Example demonstrating the improvement in predicting human transitions when using curiosity-biased versus standard PageRank. Biased curiosity assigns higher percentile ranks to actual transitions than standard PageRank. *(C)* Random walker diffusion, measured as the distance from the initial node for each graph. A comparison is made between the unbiased (blue), IGT-biased (orange), and CPT-biased walkers (green).

with and without biasing the random walk process. We find that biased curiosity assigns higher percentile ranks to actual transitions than standard PageRank. We also analyze the distance from the initial node with respect to time for individual random walk trajectories (Figure 5C). In general, observed differences between the biased walkers are small and fall within the standard deviation of the walk process. However, the CPT-biased walker stands out as it tends to remain closer to the initial node in both the MovieLens and the Wikispeedia datasets (see Supplement). These observations suggest that the differences observed in the biased PageRank algorithm are not solely attributable to changes in the diffusion properties of the random walks.

5 Discussion

How to measure curiosity best remains an open question both in the context of humans and for reinforcement learning. While our specific choices—1-cycles and network compressibility—are motivated by recent work studying human behavior, other topological features may be more suited to drive graph exploration. Nonetheless, through our work, we demonstrate the utility of content-agnostic topology-aware intrinsic motivations. Similar to leading theories of human curiosity that are intentionally content independent, our method 1) uses no node information other than topological statistics and 2) uses no information in the reward other than what is available in the structure of visited subgraphs. Even at this level of abstraction, where topology takes precedence over content, we show that agents learn generalizable and transferable graph exploration strategies. Further, we show that agents trained with human-like motivations can help devise centrality measures that predict human behavior better than PageRank. This result has two critical implications. First, we can use our method to test hypotheses about human motivations when navigating graph-structured environments. Trained agents can act as hypothesis testers by examining whether their choices for the subsequent nodes to visit align with human choices. Second, our method can be used to design recommender systems for environments where human navigation of graphs is largely goalless.

References

[1] Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 2778–2787. JMLR.org, 2017. 1, 3

[2] Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, and Alexei A. Efros. Large-scale study of curiosity-driven learning. In *ICLR*, 2019. 1

[3] Zhaohan Daniel Guo, Shantanu Thakoor, Miruna Pislar, Bernardo Avila Pires, Florent Altché, Corentin Tallec, Alaa Saade, Daniele Calandriello, Jean-Bastien Grill, Yunhao Tang, Michal Valko, Remi Munos, Mohammad Gheshlaghi Azar, and Bilal Piot. BYOL-explore: Exploration by bootstrapped prediction. In *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=qHGCH75usg. 1

[4] Dale Zhou, David M. Lydon-Staley, Perry Zurn, and Danielle S. Bassett. The growth and form of knowledge networks by kinesthetic curiosity. *Current Opinion in Behavioral Sciences*, 35:125–134, 2020. ISSN 2352-1546. doi: https://doi.org/10.1016/j.cobeha.2020.09.007. URL https://www.sciencedirect.com/science/article/pii/S235215462030142X. 1, 2

[5] George Loewenstein. The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116(1):75–98, 1994. 1, 2

[6] Min Jeong Kang, Ming Hsu, Ian M. Krajbich, George Loewenstein, Samuel M. McClure, Joseph Tao yi Wang, and Colin F. Camerer. The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, 20(8):963–973, 2009. doi: 10.1111/j.1467-9280.2009.02402.x. URL https://doi.org/10.1111/j.1467-9280.2009.02402.x. PMID: 19619181. 1

[7] Nabil Daddaoua, Manuel Lopes, and Jacqueline Gottlieb. Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates. *Scientific Reports*, 6(1):20202, 2016. doi: 10.1038/srep20202. URL https://doi.org/10.1038/srep20202. 1

[8] Juergen Schmidhuber. Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *arXiv pre-print*, 2008. 1, 4

[9] Jürgen Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, 2010. doi: 10.1109/TAMD.2010.2056368. 1, 3

[10] Joshua B. Tenenbaum, Charles Kemp, Thomas L. Griffiths, and Noah D. Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022):1279–1285, 2011. doi: 10.1126/science.1192788. URL https://www.science.org/doi/abs/10.1126/science.1192788. 1

[11] Anne G. E. Collins. The cost of structure learning. *Journal of Cognitive Neuroscience*, 29 (10):1646–1655, 10 2017. ISSN 0898-929X. doi: 10.1162/jocn_a_01128. URL https://doi.org/10.1162/jocn_a_01128.

[12] Ida Momennejad. Learning structures: predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences*, 32:155–166, 2020. 1

[13] Jacqueline Gottlieb, Pierre-Yves Oudeyer, Manuel Lopes, and Adrien Baranes. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, 17(11):585–593, 2013. ISSN 1364-6613. doi: https://doi.org/10.1016/j.tics.2013.09.001. URL https://www.sciencedirect.com/science/article/pii/S1364661313002052. 2

[14] Celeste Kidd and Benjamin Y. Hayden. The psychology and neuroscience of curiosity. *Neuron*, 88(3):449–460, 2015. ISSN 0896-6273. doi: https://doi.org/10.1016/j.neuron.2015.09.010. URL https://www.sciencedirect.com/science/article/pii/S0896627315007679. 2

[15] Christopher K. Hsee and Bowen Ruan. The Pandora effect: The power and peril of curiosity. *Psychological Science*, 27(5):659—-666, 2016. 2

[16] Jaydin Clark, Asia Vincent, Xinyi Wang, Amanda L. McGowan, and David M. Lydon-Staley. Smokers' curiosity for tobacco-related trivia aids memory of tobacco-related information. *PsyArXiv*, 2021. 2

[17] Daniel Bennett, Stefan Bode, Maja Brydevall, Hayley Warren, and Carsten Murawski. Intrinsic valuation of information in decision making under uncertainty. *PLOS Computational Biology*, 12(7):1–21, 07 2016. doi: 10.1371/journal.pcbi.1005020. URL https://doi.org/10.1371/journal.pcbi.1005020. 2

[18] Maja Brydevall, Daniel Bennett, Carsten Murawski, and Stefan Bode. The neural encoding of information prediction errors during non-instrumental information seeking. *Scientific Reports*, 8 (1):6134, 2018. doi: 10.1038/s41598-018-24566-x. URL https://doi.org/10.1038/s41598-018-24566-x. 2

[19] Perry Zurn and Danielle S. Bassett. On curiosity: A fundamental aspect of personality, a practice of network growth. *Personality Neuroscience*, 1:e13, 2018. doi: 10.1017/pen.2018.3. 2

[20] David M. Lydon-Staley, Dale Zhou, Ann Sizemore Blevins, Perry Zurn, and Danielle S. Bassett. Hunters, busybodies and the knowledge network building associated with deprivation curiosity. *Nature Human Behaviour*, 5(3):327–336, 2021. doi: 10.1038/s41562-020-00985-7. URL https://doi.org/10.1038/s41562-020-00985-7. 2

[21] Shubhankar P. Patankar, Dale Zhou, Christopher W. Lynn, Jason Z. Kim, Mathieu Ouellet, Harang Ju, Perry Zurn, David M. Lydon-Staley, and Dani S. Bassett. Curiosity as filling, compressing, and reconfiguring knowledge networks. *Collective Intelligence*, 2(4): 26339137231207633, 2023. doi: 10.1177/26339137231207633. URL https://doi.org/10.1177/26339137231207633. 2, 4

[22] Arthur Aubret, Laetitia Matignon, and Salima Hassas. A survey on intrinsic motivation in reinforcement learning. *arXiv pre-print*, 2019. 2

[23] Pawel Ladosz, Lilian Weng, Minwoo Kim, and Hyondong Oh. Exploration in deep reinforcement learning: A survey. *Information Fusion*, 85:1–22, 2022. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2022.03.003. URL https://www.sciencedirect.com/science/article/pii/S1566253522000288. 2

[24] Marc G. Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Rémi Munos. Unifying count-based exploration and intrinsic motivation. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 1479–1487, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819. 2, 3

[25] Georg Ostrovski, Marc G. Bellemare, Aäron van den Oord, and Rémi Munos. Count-based exploration with neural density models. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 2721–2730. JMLR.org, 2017. 3

[26] Haoran Tang, Rein Houthooft, Davis Foote, Adam Stooke, Xi Chen, Yan Duan, John Schulman, Filip De Turck, and Pieter Abbeel. #exploration: A study of count-based exploration for deep reinforcement learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 2750–2759, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

[27] Marlos C. Machado, Marc G. Bellemare, and Michael Bowling. Count-based exploration with the successor representation, 2018.

[28] Daejin Jo, Sungwoong Kim, Daniel Wontae Nam, Taehwan Kwon, Seungeun Rho, Jongmin Kim, and Donghoon Lee. Leco: Learnable episodic count for task-specific intrinsic reward, 2022. 2

[29] Pierre-Yves Oudeyer and Frederic Kaplan. What is intrinsic motivation? a typology of computational approaches. *Frontiers in Neurorobotics*, 1, 2007. ISSN 1662-5218. doi: 10.3389/neuro.12.006.2007. URL https://www.frontiersin.org/articles/10.3389/neuro.12.006.2007. 3

[30] Pierre-Yves Oudeyer, Frdric Kaplan, and Verena V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2): 265–286, 2007. doi: 10.1109/TEVC.2006.890271.

[31] Bradly C. Stadie, Sergey Levine, and Pieter Abbeel. Incentivizing exploration in reinforcement learning with deep predictive models. *arXiv pre-print*, 2015.

[32] Rein Houthooft, Xi Chen, Yan Duan, John Schulman, Filip De Turck, and Pieter Abbeel. Vime: Variational information maximizing exploration. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 1117–1125, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819. 3

[33] Justin Fu, John D. Co-Reyes, and Sergey Levine. Ex2: Exploration with exemplar models for deep reinforcement learning, 2017. 3

[34] Nikolay Savinov, Anton Raichuk, Raphaël Marinier, Damien Vincent, Marc Pollefeys, Timothy Lillicrap, and Sylvain Gelly. Episodic curiosity through reachability. *arXiv pre-print*, 2018. 3

[35] Jonathan Sorg, Richard L Lewis, and Satinder Singh. Reward design via online gradient ascent. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010. URL https://proceedings.neurips.cc/paper_files/paper/2010/file/168908dd3227b8358eababa07fcaf091-Paper.pdf. 3

[36] Zeyu Zheng, Junhyuk Oh, and Satinder Singh. On learning intrinsic rewards for policy gradient methods. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 4649–4659, Red Hook, NY, USA, 2018. Curran Associates Inc.

[37] Farzan Memarian, Wonjoon Goo, Rudolf Lioutikov, Scott Niekum, and Ufuk Topcu. Self-supervised online reward shaping in sparse-reward environments. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, page 2369–2375. IEEE Press, 2021. doi: 10.1109/IROS51168.2021.9636020. URL https://doi.org/10.1109/IROS51168.2021.9636020.

[38] Rati Devidze, Parameswaran Kamalaruban, and Adish Singla. Exploration-guided reward shaping for reinforcement learning under sparse rewards. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=W7HvKO1erY. 3

[39] Christos Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Dover Publications, 1982. 3

[40] Yun Peng, Byron Choi, and Jianliang Xu. Graph learning for combinatorial optimization: A survey of state-of-the-art. *Data Science and Engineering*, 6(2):119–141, Jun 2021. ISSN 2364-1541. doi: 10.1007/s41019-021-00155-3. URL https://doi.org/10.1007/s41019-021-00155-3. 3

[41] Nina Mazyavkina, Sergey Sviridov, Sergei Ivanov, and Evgeny Burnaev. Reinforcement learning for combinatorial optimization: A survey. *Computers & Operations Research*, 134: 105400, 2021. ISSN 0305-0548. doi: https://doi.org/10.1016/j.cor.2021.105400. URL https://www.sciencedirect.com/science/article/pii/S0305054821001660.

[42] Sai Munikoti, Deepesh Agarwal, Laya Das, Mahantesh Halappanavar, and Balasubramaniam Natarajan. Challenges and opportunities in deep reinforcement learning with graph neural networks: A comprehensive review of algorithms and applications. *arXiv pre-print*, 2022. 3, 5

[43] Hanjun Dai, Elias B. Khalil, Yuyu Zhang, Bistra Dilkina, and Le Song. Learning combinatorial optimization algorithms over graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6351–6361, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964. 3

[44] Chaitanya K. Joshi, Quentin Cappart, Louis-Martin Rousseau, and Thomas Laurent. Learning the travelling salesperson problem requires rethinking generalization. *arXiv pre-print*, 2020. doi: 10.4230/LIPIcs.CP.2021.33.

[45] Yujiao Hu, Yuan Yao, and Wee Sun Lee. A reinforcement learning approach for optimizing multiple traveling salesman problems over graphs. *Knowledge-Based Systems*, 204:106244, 2020. ISSN 0950-7051. doi: https://doi.org/10.1016/j.knosys.2020.106244. URL https://www.sciencedirect.com/science/article/pii/S0950705120304445. 3

[46] Sungsoo Ahn, Younggyo Seo, and Jinwoo Shin. Learning what to defer for maximum independent sets. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020. 3

[47] Faisal N. Abu-Khzam, Mohamed M. Abd El-Wahab, Moussa Haidous, and Noureldin Yosri. Learning from obstructions: An effective deep learning approach for minimum vertex cover. *Annals of Mathematics and Artificial Intelligence*, Aug 2022. ISSN 1573-7470. doi: 10.1007/s10472-022-09813-2. URL https://doi.org/10.1007/s10472-022-09813-2. 3

[48] Hanchen Wang, Ying Zhang, Lu Qin, Wei Wang, Wenjie Zhang, and Xuemin Lin. Reinforcement learning based query vertex ordering model for subgraph matching. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 245–258, 2022. doi: 10.1109/ICDE53745.2022.00023. 3

[49] Eli A. Meirom, Haggai Maron, Shie Mannor, and Gal Chechik. Optimizing tensor network contraction using reinforcement learning. *arXiv pre-print*, 2022. 3

[50] Hanjun Dai, Yujia Li, Chenglong Wang, Rishabh Singh, Po-Sen Huang, and Pushmeet Kohli. Learning transferable graph exploration. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/afe434653a898da20044041262b3ac74-Paper.pdf. 3

[51] Fanfei Chen, Paul Szenher, Yewei Huang, Jinkun Wang, Tixiao Shan, Shi Bai, and Brendan Englot. Zero-shot reinforcement learning on graphs for autonomous exploration under uncertainty. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5193–5199, 2021. doi: 10.1109/ICRA48506.2021.9561917. 3

[52] Victor-Alexandru Darvariu, Stephen Hailes, and Mirco Musolesi. Goal-directed graph construction using reinforcement learning. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 477(2254):20210168, 2021. doi: 10.1098/rspa.2021.0168. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2021.0168. 3

[53] Christoffel Doorman, Victor-Alexandru Darvariu, Stephen Hailes, and Mirco Musolesi. Dynamic network reconfiguration for entropy maximization using deep reinforcement learning. Proceedings of the First Learning on Graphs Conference (LoG 2022), PMLR 198:49:1-49:15, 2022. 3

[54] David F Gleich. PageRank beyond the web. *siam REVIEW*, 57(3):321–363, 2015. 3, 5

[55] Darui Zhu, Haifeng Wang, Rui Wang, Jiandong Duan, and Jing Bai. Identification of key nodes in a power grid based on modified pagerank algorithm. *Energies*, 15(3):797, 2022. 3, 5

[56] Dániel Ficzere, Gergely Hollósi, Attila Frankó, and András Gulyás. Random walk for generalization in goal-directed human navigation on Wikipedia. In *International Conference on Complex Networks and Their Applications*, pages 202–213. Springer, 2022. 3, 6

[57] Tiziano Piccardi, Martin Gerlach, Akhil Arora, and Robert West. A large-scale characterization of how readers browse Wikipedia. *ACM Transactions on the Web*, 17(2):1–22, 2023. 3

[58] Salvador Aguinaga, Aditya Nambiar, Zuozhu Liu, and Tim Weninger. Concept hierarchies and human navigation. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 38–45. IEEE, 2015.

[59] Hisham Benotman and David Maier. Comparing personalized PageRank and activation spreading in Wikipedia diagram-based search. In *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 41–50. IEEE, 2021. 3

[60] Dimitar Dimitrov, Philipp Singer, Florian Lemmerich, and Markus Strohmaier. What makes a link successful on Wikipedia? In *Proceedings of the 26th International Conference on World Wide Web*, pages 917–926, 2017. 3

[61] Lisette Espín-Noboa, Florian Lemmerich, Simon Walk, Markus Strohmaier, and Mark Musen. Hoprank: How semantic structure influences teleportation in PageRank (a case study on bioportal). In *The World Wide Web Conference*, pages 2708–2714, 2019. 3, 6

[62] Jonathan L. Gross and Thomas W. Tucker. *Topological Graph Theory*. Wiley-Interscience, USA, 1987. ISBN 0471049263. 4

[63] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46: 255–308, January 2009. 4

[64] Robert Ghrist. Barcode: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45:61–75, October 2007.

[65] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[66] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005. doi: 10.1007/s00454-004-1146-y. URL https://doi.org/10.1007/s00454-004-1146-y.

[67] Ginestra Bianconi. *Higher-Order Networks*. Elements in Structure and Dynamics of Complex Networks. Cambridge University Press, 2021. doi: 10.1017/9781108770996. 4

[68] Christopher W. Lynn and Danielle S. Bassett. Quantifying the compressibility of complex networks. *Proceedings of the National Academy of Sciences*, 118(32):e2023473118, 2021. doi: 10.1073/pnas.2023473118. URL https://www.pnas.org/doi/abs/10.1073/pnas.2023473118. 4, 7

[69] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, 2018. 4

[70] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2021. doi: 10.1109/TNNLS.2020.2978386. 5

[71] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 1025–1035, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964. 5

[72] Chen Cai and Yusu Wang. A simple yet effective baseline for non-attributed graph classification. *arXiv pre-print*, 2018. 5

[73] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, Feb 2015. ISSN 1476-4687. doi: 10.1038/nature14236. URL https://doi.org/10.1038/nature14236. 5

[74] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998. 5

[75] Baoning Wu and Kumar Chellapilla. Extracting link spam using biased random walks from spam seed sets. In *Proceedings of the 3rd international workshop on Adversarial information retrieval on the web*, pages 37–44, 2007. 5

[76] Olle Häggström et al. *Finite Markov chains and algorithmic applications*, volume 52. Cambridge University Press, 2002. 5

[77] Taher Haveliwala, Sepandar Kamvar, Glen Jeh, et al. An analytical comparison of approaches to personalizing PageRank. Technical report, Technical report, Stanford University, 2003. 5, 8

[78] Mark Newman. *Networks: An Introduction*. Oxford University Press, Inc., USA, 2010. ISBN 0199206651. 6

[79] Jesper Dall and Michael Christensen. Random geometric graphs. *Physical review E*, 66(1): 016121, 2002. 6, 7

[80] Luciano da Fontoura Costa and Gonzalo Travieso. Exploring complex networks through random walks. *Physical Review E*, 75(1):016102, 2007. 7

[81] Barbara Giunti, Guillaume Houry, and Michael Kerber. Average complexity of matrix reduction for clique filtrations. In *Proceedings of the 2022 International Symposium on Symbolic and Algebraic Computation*, pages 187–196, 2022. 8

[82] F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19, 2015. 8

[83] Ruining He and Julian McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*, pages 507–517, 2016. 8

[84] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 43–52, 2015. 8

[85] Robert West and Jure Leskovec. Human wayfinding in information networks. In *Proceedings of the 21st international conference on World Wide Web*, pages 619–628, 2012. 8

[86] Robert West, Joelle Pineau, and Doina Precup. Wikispeedia: An online game for inferring semantic distances between concepts. In *Twenty-First International Joint Conference on Artificial Intelligence*, 2009. 8