The Discrepancy of Shortest Paths

Greg Bodwin
Computer Science and Engineering, University of Michigan, Ann Arbor, MI, USA

Chengyuan Deng
Department of Computer Science, Rutgers University, Piscataway, NJ, USA

Jie Gao
Department of Computer Science, Rutgers University, Piscataway, NJ, USA

Gary Hoppenworth
Computer Science and Engineering, University of Michigan, Ann Arbor, MI, USA

Jalaj Upadhyay
Management Science and Information Systems, Rutgers University, Piscataway, NJ, USA

Chen Wang
Department of Computer Science, Rice University, Houston, TX, USA
Computer Science and Engineering, Texas A&M University, College Station, TX, USA

Abstract

The *hereditary discrepancy* of a set system is a quantitative measure of the pseudorandom properties of the system. Roughly speaking, hereditary discrepancy measures how well one can 2-color the elements of the system so that each set contains approximately the same number of elements of each color. Hereditary discrepancy has numerous applications in computational geometry, communication complexity and derandomization. More recently, the hereditary discrepancy of the set system of *shortest paths* has found applications in differential privacy [Chen et al. SODA 23].

The contribution of this paper is to improve the upper and lower bounds on the hereditary discrepancy of set systems of unique shortest paths in graphs. In particular, we show that any system of unique shortest paths in an undirected weighted graph has hereditary discrepancy $O(n^{1/4})$, and we construct lower bound examples demonstrating that this bound is tight up to polylog n factors. Our lower bounds hold even for planar graphs and bipartite graphs, and improve a previous lower bound of $\Omega(n^{1/6})$ obtained by applying the trace bound of Chazelle and Lvov [SoCG'00] to a classical point-line system of Erdős.

As applications, we improve the lower bound on the additive error for differentially-private all pairs shortest distances from $\Omega(n^{1/6})$ [Chen et al. SODA 23] to $\widetilde{\Omega}(n^{1/4})$, and we improve the lower bound on additive error for the differentially-private all sets range queries problem to $\widetilde{\Omega}(n^{1/4})$, which is tight up to polylog n factors [Deng et al. WADS 23].

2012 ACM Subject Classification Theory of computation → Shortest paths; Theory of computation → Computational geometry

Keywords and phrases Discrepancy, hereditary discrepancy, shortest paths, differential privacy

Digital Object Identifier 10.4230/LIPIcs.ICALP.2024.27

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: `https://arxiv.org/abs/2401.15781` [9]

Funding *Greg Bodwin and Gary Hoppenworth*: Supported by NSF:AF 2153680.
Jie Gao and Chengyuan Deng: Supported partially by NSF DMS-2311064, DMS-2220271, IIS-2229876, CCF-2208663, CCF-2118953.
Jalaj Upadhyay: Funded by Decanal Research Grant number 302918.

1 Introduction

In graph algorithms, a fundamental problem is to efficiently compute distance or shortest path information of a given input graph. Over the last decade or so, the community has increasingly sought a principled understanding of the *combinatorial structure* of shortest paths, with the goal to exploit this structure in algorithm design. That is, in various graph settings, we can ask:

> *What notable structural properties hold for **shortest** path systems, that do not necessarily hold for **arbitrary** path systems?*

The following are a few of the major successes of this line of work:

- An extremely popular strategy in the literature is to use *hitting sets*, in which we (often randomly) generate a set of nodes S and argue that it will hit a shortest path for every pair of nodes that are sufficiently far apart. Hitting sets rarely exploit any structure of shortest paths, as evidenced by the fact that most hitting set algorithms generalize immediately to arbitrary set systems. However, they have inspired a successful line of work into graphs of bounded *highway dimension* [1, 6, 7]; very roughly, these are graphs whose shortest paths admit unusually efficient hitting sets of a certain kind.
- Shortest paths exhibit the notable structural property of *consistency*, i.e., any subpath of a shortest path is itself a shortest path. This fact is used throughout the literature on graph algorithms [21, 22, 8], including e.g. in the classic Floyd-Warshall algorithm for All-Pairs Shortest Paths. A recent line of work has sought to characterize the additional structure exhibited by shortest path systems, beyond consistency [8, 21, 19, 20, 17, 4, 2].
- Planar graphs have received special attention within this research program, and planar shortest path systems carry some notable additional structure. For example, it is known that planar shortest paths have unusually efficient tree coverings [5, 11], and that their shortest paths can be compressed into surprisingly small space [12, 13]. Shortest path algorithms also often benefit from more general structural facts about planar graphs, such as separator theorems [29, 28].

The main result of this paper is a new structural separation between shortest path systems and arbitrary path systems, expressed through the lens of *discrepancy theory*. We will come to formal definitions of discrepancy in just a moment, but at a high level, discrepancy has been described as a quantitative measure of the combinatorial pseudorandomness of a discrete system [18], and it has widespread applications in discrete and computational geometry, random sampling and derandomization, communication complexity, and much more[1]. We will show the following:

▶ **Theorem 1** (Main Result, Informal). *The discrepancy of unique shortest path systems in weighted graphs is inherently smaller than the discrepancy of arbitrary path systems in graphs.*

This separation between unique shortest paths and arbitrary paths is due to the structural property of *consistency* of unique shortest path systems, which is well-studied in the literature [21, 22, 8].

[1] We refer to the excellent textbooks of Alexander, Beck, and Chen [3], Chazelle [14], Matoušek [33] for discussion and further applications.

Our results can be placed within a larger context of prior work in computational geometry. A classical topic in this area is to determine the discrepancy of incidence structures between points and geometric range spaces such as axis-parallel rectangles, half-spaces, lines, and curves (cf. [14, Section 1.5]). These results have been used to show lower bounds for geometric range searching [37, 34].

Indeed, systems of unique shortest paths in graphs capture some of the geometric range spaces studied in prior work. For instance, arrangements of straight lines in Euclidean space can be interpreted as systems of unique shortest paths in an associated graph, implying a relation between the discrepancies of these two set systems. This connection has recently found applications in the study of differential privacy on shortest path distance and range query algorithms [16, 23].

More generally, discrepancy on graphs have also found applications in proving tight lower bounds on answering cut queries on graphs [26, 32]. We provide a detailed literature review for discrepancy on graphs in the full version of our paper [9]. The full version of our paper further discusses the connection between our results and the discrepancy of arrangements of curves.

1.1 Formal Definitions of Discrepancy

We first collect the basic definitions needed to understand this paper.

▶ **Definition 2** (Edge and Vertex Incidence Matrices). *Given a graph $G = (V, E)$ and a set of paths Π in G, the associated* vertex incidence matrix *is given by $A \in \mathbb{R}^{|\Pi| \times |V|}$, where for each $v \in V$ and $\pi \in \Pi$ the corresponding entry is*

$$A_{\pi,v} = \begin{cases} 1 & \text{if } v \in \pi \\ 0 & \text{if } v \notin \pi. \end{cases}$$

The associated edge incidence matrix *is given by $A \in \mathbb{R}^{|\Pi| \times |E|}$, where for each $e \in E$ and $\pi \in \Pi$ the corresponding entry is*

$$A_{\pi,e} = \begin{cases} 1 & \text{if } e \in \pi \\ 0 & \text{if } e \notin \pi. \end{cases}$$

▶ **Definition 3** (Discrepancy and Hereditary Discrepancy). *Given a matrix $A \in \mathbb{R}^{m \times n}$, its discrepancy is the quantity*

$$\mathrm{disc}(A) = \min_{x \in \{1,-1\}^n} \|Ax\|_\infty.$$

Its hereditary discrepancy is the maximum discrepancy of any submatrix A_Y obtained by keeping all rows but only a subset $Y \subseteq [n]$ of the columns; that is,

$$\mathrm{herdisc}(A) = \max_{Y \subseteq [n]} \mathrm{disc}_v(A_Y).$$

For a system of paths Π in a graph G, we will write $\mathrm{disc}_v(\Pi), \mathrm{herdisc}_v(\Pi)$ to denote the (hereditary) discrepancy of its vertex incidence matrix, and $\mathrm{disc}_e(\Pi), \mathrm{herdisc}_e(\Pi)$ to denote the (hereditary) discrepancy of its edge incidence matrix.

For intuition, the vertex discrepancy of a system of paths Π can be equivalently understood as follows. Suppose that we color each node in G either red or blue, with the goal to balance the red and blue nodes on each path as evenly as possible. The discrepancy associated to that particular coloring is the quantity

$$\max_{\pi \in \Pi} \left| \ |\{v \in \pi \ \mid \ v \text{ colored red}\}| - |\{v \in \pi \ \mid \ v \text{ colored blue}\}| \ \right|.$$

The discrepancy of the system Π is the minimum possible discrepancy over all colorings. The hereditary discrepancy is the maximum discrepancy taken over all *induced path subsystems* Π' of Π; that is, Π' is obtained from Π by selecting zero or more vertices from G, deleting these vertices, and deleting all instances of these vertices from all paths.[2] We may delete nodes from the middle of some paths $\pi \in \Pi$, in which case Π' may no longer be a system of paths in G, but rather a system of paths in some other graph G' with fewer nodes and some additional edges. Nonetheless, its vertex incidence matrix and therefore $\text{herdisc}_v(\Pi')$ remain well-defined with respect to this new graph G'. Edge discrepancy can be understood in a similar way, coloring edges rather than vertices.

1.2 Our Results

Our main result is an upper and lower bound on the hereditary discrepancy of unique shortest path systems in weighted graphs, which match up to hidden polylog n factors.

▶ **Theorem 4** (Main Result).

- **(Upper Bound).** *For any n-node undirected weighted graph G with a unique shortest path between each pair of nodes, there exists a polynomial-time algorithm that finds a coloring for the system of shortest paths Π such that:*

$$\text{herdisc}_v(\Pi) \leq \widetilde{O}(n^{1/4}) \qquad \textit{and} \qquad \text{herdisc}_e(\Pi) \leq \widetilde{O}(n^{1/4}).$$

- **(Lower Bound).** *There are examples of n-node undirected weighted graphs G with a unique shortest path between each pair of nodes in which this system of shortest paths Π has $\text{herdisc}_v(\Pi) \geq \widetilde{\Omega}(n^{1/4})$ and $\text{herdisc}_e(\Pi) \geq \widetilde{\Omega}(n^{1/4})$. In fact, in these lower bound examples we can take G to be planar or bipartite.*

This theorem has immediate applications in differential privacy; we refer to Theorem 6 discussed below. We can strengthen the hereditary discrepancy lower bound into a vertex (non-hereditary) discrepancy lower bound in the undirected and directed settings. We leave open whether our lower bound extends to (non-hereditary) edge discrepancy as well, and to vertex or edge discrepancy of planar graphs. We refer to Table 1 for a list of our results in these settings.

Table 1 Overview of vertex/edge (hereditary) discrepancy on general graphs and special families of graph: tree, bipartite and planar graphs. Here n is the number of vertices of the graph and m is the number of edges. D is the graph diameter or the longest number of hops of paths considered.

		Tree	Bipartite	Planar	Undirected Graph	Directed Graph
V	disc	$\Theta(1)$	$\Theta(1)$	$O(n^{1/4})$	$\widetilde{\Theta}(n^{1/4})$	$\widetilde{\Theta}(n^{1/4})$
	herdisc	$\Theta(1)$	$\widetilde{\Theta}(n^{1/4})$	$\widetilde{\Theta}(n^{1/4})$	$\Omega(n^{1/6})$[15] $\rightarrow \widetilde{\Theta}(n^{1/4})$	$\widetilde{\Theta}(n^{1/4})$
E	disc	$\Theta(1)$	$\Theta(1)$	$O(n^{1/4})$	$O(n^{1/4})$	$\min\left\{O(m^{1/4}), \widetilde{O}(D^{1/2})\right\}$
	herdisc	$\Theta(1)$	$\widetilde{\Theta}(n^{1/4})$	$\widetilde{\Theta}(n^{1/4})$	$\Omega(n^{1/6})$[15] $\rightarrow \widetilde{\Theta}(n^{1/4})$	$\widetilde{\Omega}(n^{1/4})$

[2] In the coloring interpretation, hereditary discrepancy allows a different choice of coloring for each subsystem Π', rather than fixing a coloring for Π and considering the induced coloring on each Π'.

The upper bound in Theorem 4 is constructive and algorithmic; that is, we provide an algorithm that colors vertices (resp. edges) of the input graph to achieve vertex (resp. edge) discrepancy $\widetilde{O}(n^{1/4})$ on its shortest paths (or on a given subsystem of its shortest paths). Notably, Theorem 4 should be contrasted with the fact that the maximum possible discrepancy of *any* simple path system of polynomial size in a general graph is known to be $\widetilde{\Theta}(n^{1/2})$.[3] In fact, the lower bound on discrepancy (as well as hereditary discrepancy) for a grid graph for a polynomial number of *simple* paths can be $\Omega(\sqrt{n})$ (see the full version of our paper [9] for a proof and more discussion on grid graphs). Thus, Theorem 4 represents a concrete separation between unique shortest path systems and general path systems.

The main open question that we leave in this work is on the hereditary edge discrepancy of shortest paths in *directed* weighted graphs. We show the following:

▶ **Theorem 5.** *For any n-node, m-edge directed weighted graph G with a unique shortest path between each pair of nodes, the system of shortest paths Π satisfies*

$$\mathrm{herdisc}_v(\Pi) \leq O(n^{1/4}) \qquad \textit{and} \qquad \mathrm{herdisc}_e(\Pi) \leq O(m^{1/4}).$$

Lower bounds in the undirected setting immediately apply to the directed setting as well, and so this essentially closes the problem for directed hereditary *vertex* discrepancy. It is an interesting open problem whether the bound for directed hereditary *edge* discrepancy can be improved to $\widetilde{O}(n^{1/4})$ as well.

Applications to Differential Privacy. One application of our discrepancy lower bound on unique shortest paths is in differential privacy (DP) [24, 25]. An algorithm is differentially private if its output distributions are relatively close regardless of whether an individual's data is present in the data set. More formally, for two databases Y and Y' that are identical except for one data entry, a randomized algorithm $\mathcal{M}$ is (ε, δ) differentially private if for any measurable set A in the range of $\mathcal{M}$, $\Pr[\mathcal{M}(Y) \in A] \leq e^{\varepsilon} \Pr[\mathcal{M}(Y') \in A] + \delta$.

The topic of discrepancy of paths on a graph is related to two problems already studied in differential privacy: *All Pairs Shortest Distances (APSD)* ([16, 27, 36]) and *All Sets Range Queries (ASRQ)* ([23]), both assuming the graph topology is public. In APSD problem, the edge weights are not publicly known. A query in APSD is a pair of vertices $(u, v) \in V \times V$ and the answer is the shortest distance between u and v. In contrast, in ASRQ problem, the edge weights are assumed to be known, and every edge also has a private *attribute*. Here, the range is defined by the shortest path between two vertices (based on publicly known edge weights). The answer to the query $(u, v) \in V \times V$ then is the sum of private attributes along the shortest path. In what follows, we give a high-level argument for the lower bound on DP-APSD problem; the lower bound of $\widetilde{\Omega}(n^{1/4})$ for the DP-ASRQ problem also follows nearly the same arguement (see the full version of our paper [9] for details).

Chen et al. [16] showed that DP-APSD can be formulated as a linear query problem. In this setting, we are given a vertex incidence matrix A of the $\binom{n}{2}$ shortest paths of a graph and a vector x of length n and asked to output Ax. They show that the hereditary discrepancy of the matrix A provides a lower bound on the ℓ_∞ error for any (ε, δ)-DP mechanism for this problem. With this argument, our new discrepancy lower bound immediately implies:

[3] A path system is *simple* if no individual path repeats nodes. The upper bound of $\widetilde{O}(n^{1/2})$ follows by coloring the nodes randomly and applying standard Chernoff bounds. The lower bound is nontrivial and follows from an analysis of the Hadamard matrix; see [14], Section 1.5.

▶ **Theorem 6** (Informal version of Corollaries 7.1 and F.1 in [9]). *The (ε, δ)-DP APSD problem and (ε, δ)-DP ASRQ problem require additive error at least $\widetilde{\Omega}(n^{1/4})$.*

The best known additive error bound for the DP-ASRQ problem is $\widetilde{O}(n^{1/4})$ [23], which, by Theorem 6, is tight up to a polylog(n) factor. Prior to this work, the only known lower bounds for DP-ASRQ and DP-APSD were from a point-line system with hereditary discrepancy of $\Omega(n^{1/6})$ [16]. The best known additive error upper bound for DP-APSD is $\widetilde{O}(n^{1/2})$ [16, 27]. Closing this gap remains an interesting open problem.

In addition to differential privacy, our hereditary discrepancy results also have implications for matrix analysis. In short, we can show that the factorization norm of the shortest path incidence matrix is $\widetilde{\Theta}(n^{1/4})$. We delay a detailed discussion to the full version of our paper [9].

1.3 Our Techniques

We will overview our upper and lower bounds on discrepancy separately.

Upper Bound Techniques. A folklore structural property of unique shortest paths is *consistency*. Formally, a system of undirected paths Π is consistent if for any two paths π_1, π_2, their intersection $\pi_1 \cap \pi_2$ is a (possibly empty) contiguous subpath of each. It is well known that, for any undirected graph $G = (V, E, w)$ with unique shortest paths, its system of shortest paths Π is consistent. An analogous fact holds for directed graphs. Our discrepancy upper bounds will actually apply to *any* consistent system of paths – not just those that arise as unique shortest paths in a graph.

We give our upper bounds on the discrepancy of consistent systems in two steps. First, we prove the existence of a low-discrepancy coloring using a standard application of *primal shatter functions* (see the the full version of our paper [9] for a definition). For consistent paths, the primal shatter function has degree two in both directed and undirected graphs. This immediately gives us an upper bound of $O(n^{1/4})$ for vertex discrepancy and $O(m^{1/4})$ for edge discrepancy (since edge discrepancy is defined on a ground set of m edges in the graph G).

When the graph is dense, this upper bound on edge discrepancy deteriorates, becoming trivial when $m = \Theta(n^2)$. We thus present a second proof of $\tilde{O}(n^{1/4})$ for *both* vertex and edge discrepancy, which explicitly constructs a low-discrepancy coloring. This improves the bound for vertex discrepancy by polylogarithmic factors and edge discrepancy by polynomial factors. The main idea in this construction is to adapt the *path cover* technique, used in the recent breakthrough on shortcut sets [30]. That is, we start by finding a small base set of roughly $n^{1/2}$ node-disjoint shortest paths in the distance closure of the graph. These paths have the property that any other shortest path π in the graph contains at most $O(n^{1/2})$ nodes that are not in any paths in the base set. We then color *randomly*, as follows:

- For every node that is not contained in any path in the base set, we assign its color randomly. Thus, applying concentration bounds, the contribution of these nodes to the discrepancy of π will be bounded by $\pm\widetilde{O}(n^{1/4})$.
- For every path in the base set, we choose the color of the first node in the path at random, and then alternate colors along the path after that. Then we can argue that by consistency, the nodes in each base path randomly contribute $+1$ or -1 (or 0) to the discrepancy of π (see Figure 1 for a visualization). Since there are only $n^{1/2}$ paths in the base set, we may again apply concentration bounds to argue that the contribution to discrepancy from these base paths will only be $\pm\widetilde{O}(n^{1/4})$.

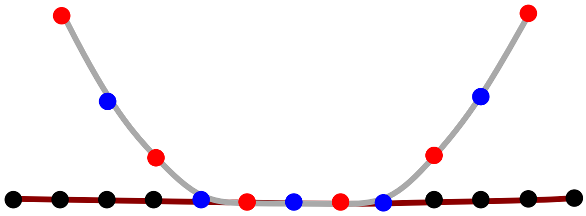

Figure 1 If we color the nodes of a unique shortest path with alternating colors, then its nodes will contribute discrepancy 0, $+1$, or -1 to all unique shortest paths that intersect it.

Summing together these two parts, we obtain a bound of $\widetilde{O}(n^{1/4})$ on discrepancy, which holds with high probability. We can translate this to a bound on *hereditary* discrepancy using the fact that consistency is a hereditary property of path systems.

Lower Bound Techniques. For lower bounds, we apply the trace bound of [15] on hereditary discrepancy together with an explicit graph construction [10] that was recently proposed as a lower bound against *hopsets* in graphs. An (exact) hopset of a graph G with hopbound β is a small set of additional edges H in the distance closure of G, such that every pair of nodes has a shortest path in $G \cup H$ containing at most β edges.

Until recently, the state-of-the-art hopset lower bounds were achieved using a point-line construction of Erdős [35], which had n points and n lines in $\mathbb{R}^2$ with each point staying on $\Theta(n^{1/3})$ lines and each line going through $\Theta(n^{1/3})$ points. This point-line system also implies tight lower bounds for the Szemerédi-Trotter theorem and the discrepancy of arrangements of lines in the plane [15], as well as the previous state-of-the-art lower bound on the discrepancy of unique shortest paths.

This point-line construction can be associated with a graph that possesses useful properties derived from geometry. If edges in this graph are weighted by Euclidean distance, then the paths in the graph corresponding to straight lines are unique shortest paths by design. On the other hand, two such shortest paths (along straight lines) only intersect at most once.

Recently, a construction in Bodwin and Hoppenworth [10] obtained stronger hopset lower bounds with a different geometric graph construction, which still took place in $\mathbb{R}^2$ but allowed shortest paths to have many vertices/edges in common. We show that this construction can be repurposed to derive a stronger lower bound of $\widetilde{\Omega}(n^{1/4})$ on vertex hereditary discrepancy, by applying the trace bound of [15]. Combined with our upper bounds, this substantially improves our understanding of the discrepancy of unique shortest paths.

The above upper and lower bounds are for general graphs. Naturally, one can ask if we have better bounds for special families of graphs. We further show that the lower bounds remain the same for two interesting families: planar graphs and bipartite graphs. The lower bound construction mentioned above is not planar, and so this requires some additional work. A natural attempt is to restore planarity by adding vertices to the construction wherever two edges cross. However, this comes at a cost of an increase in the number of vertices, and also with a potential danger of altering the shortest paths. In the full version of our paper [9], we first show that the number of crossings is not too much higher than n. Then, by carefully changing the weights of the edges and by exploiting the geometric properties of the construction, we show that the topology and incidence of shortest paths are not altered. For bipartite graphs, although the vertex discrepancy can be made very low – by coloring the vertices on one side $+1$ and vertices on the other side -1 – the hereditary discrepancy can be as high as the general graph setting. Specifically, we show a 2-lift of any graph G to a bipartite graph which essentially keeps the same hereditary discrepancy. Details can be found in the full version of our paper [9].

2 Preliminaries

A *path system* is a pair $S = (V, \Pi)$ where V is a ground set of nodes and Π is a set of vertex sequences called *paths*. Each path may contain at most one instance of each node. We now formally define consistency, a structural property of unique shortest paths that will be useful.

▶ **Definition 7.** *A path system $S = (V, \Pi)$ is consistent if no two paths in S intersect, split apart, and then intersect again later. Formally:*

- *In the undirected setting, consistency means that for all $u, v \in V$ and all $\pi_1, \pi_2 \in \Pi$ such that $u, v \in \pi_1 \cap \pi_2$, we have that $\pi_1[u, v] = \pi_2[u, v]$, i.e., the intersection of π_1 and π_2 is a contiguous subpath (subsequence) of π_1 and π_2.*
- *In the directed setting, consistency means that for all $u, v \in V$ and all $\pi_1, \pi_2 \in \Pi$ such that u **precedes** v in both π_1 and π_2, we have that $\pi_1[u, v] = \pi_2[u, v]$.*

In every weighted graph for which all pairs shortest paths exist (i.e. no negative cycles), we can represent all-pairs shortest paths using a consistent path system. In particular, if all shortest paths are *unique*, then consistency is implied immediately.

We will investigate the combinatorial discrepancy of path systems (V, Π). Usually, we will assume that $|V| = n$ and $|\Pi|$ is polynomial in n. We define a vertex coloring $\chi : V \mapsto \{-1, 1\}$ and define the *discrepancy* of Π as

$$\mathrm{disc}(\Pi) = \min_{\chi} \chi(\Pi), \quad \text{where } \chi(\Pi) = \max_{\pi \in \Pi} |\chi(\pi)|, \, \chi(\pi) = \sum_{v \in \pi} \chi(v).$$

Using a random coloring χ, we can guarantee that for all paths $\pi \in \Pi$ [14]:

$$|\chi(\pi)| \leq \sqrt{2|\pi| \ln(4|\Pi|)}.$$

This immediately provides a few observations.

▶ **Observation 8.** *When Π is a set of paths with size polynomial in n, then $\mathrm{disc}(\Pi) = O(\sqrt{n \log n})$. This bound is true even for paths that are possibly non-consistent.*

▶ **Observation 9.** *When the longest path in Π has D vertices we have $\mathrm{disc}(\Pi) = O(\sqrt{D \log n})$. Thus, for graphs that have a small diameter (e.g., small world graphs), the discrepancy of shortest paths is automatically small.*

Hereditary discrepancy is a more robust measure of the complexity of a path system (V, Π), defined as $\mathrm{herdisc}(\Pi) = \max_{Y \subseteq V} \mathrm{disc}(\Pi|_Y)$, where $\Pi|_Y$ is the collection of sets of the form $\pi \cap Y$ with $\pi \in \Pi$. Clearly, $\mathrm{herdisc}(\Pi) \geq \mathrm{disc}(\Pi)$. Sometimes the discrepancy of a set system may be small while the hereditary discrepancy is large [14]. Thus in the literature, we often talk about lower bounds on the hereditary discrepancy.

Now that we have defined vertex and edge (hereditary) discrepancy, one may wonder if there is an underlying relationship between vertex and edge (hereditary) discrepancy since they share the same bounds in most settings presented in Table 1. The following observation shows that vertex discrepancy bounds directly imply bounds on edge discrepancy.

▶ **Observation 10.** *Denote by $\mathrm{disc}(n)$ (and $\mathrm{herdisc}(n)$) the maximum discrepancy (minimum hereditary discrepancy, respectively) of a consistent path system of a (undirected or directed) graph of n vertices. We have that*

1. *Let $g(x)$ be a non-decreasing function. If $\mathrm{herdisc}_v(n) \geq g(n)$, then $\mathrm{herdisc}_e(n) \geq g(n/2)$.*
2. *Let $f(x)$ be a non-decreasing function. If $\mathrm{disc}_v(n) \leq f(n)$, then $\mathrm{disc}_e(m) \leq f(m)$.*

The proof of Observation 10 is deferred to the full version of our paper [9]. We also use some technical tools from discrepancy theory and statistics. For details please refer to the full version of our paper [9].

3 Undirected Graphs: Lower Bound and Explicit Colorings

We now discuss the main result (Theorem 4). We first show in Section 3.1 a hereditary discrepancy lower bound of $\Omega(n^{1/4}/\sqrt{\log n})$ for both edge and vertex discrepancy in general undirected graphs. Then in Section 3.2 we present a vertex coloring achieving hereditary discrepancy of $\widetilde{O}(n^{1/4})$. Finally, we present an explicit edge coloring with the same hereditary discrepancy bound in Section 3.3.

3.1 Lower Bound

As suggested by Observation 10, we focus on the vertex hereditary discrepancy, and our goal is to prove the following statement (Theorem 11). In Theorem 10 of the full version of our paper [9], we show that this theorem implies the same lower bound on (non-hereditary) vertex discrepancy as well.

▶ **Theorem 11.** *There are examples of n-vertex undirected weighted graphs G with a unique shortest path between each pair of vertices in which this system of shortest paths Π has*

$$\mathrm{herdisc}_v(\Pi) \geq \Omega(n^{1/4}/\sqrt{\log n}).$$

To obtain the lower bound, we employ the new graph construction by [10], which shows that any exact hopset with $O(n)$ edges must have at least $\widetilde{\Omega}(n^{1/2})$ hop diameter. Despite seeming unrelated, this construction also sheds light on our problem. Another technique we use to show the hereditary discrepancy lower bound is the trace bound [31] (and restated in the full version of our paper [9]). In the following proof section, we first summarize the construction related to our objective, then show the calculation using the trace bound that leads to our lower bound.

Proof. The key properties of the graph construction in [10] (see also Section 5 of [9]) that we need can be summarized in the following lemma.

▶ **Lemma 12** (Lemma 1 of [10]). *For any $p \in [1, n^2]$, there is an infinite family of n-node undirected weighted graphs $G = (V, E, w)$ and sets Π of p paths in G such that*

- *G has $\ell = \Theta\left(\frac{n}{\sqrt{p \log n}}\right)$ layers. Each path in Π starts in the first layer, ends in the last layer, and contains exactly one node in each layer.*
- *Each path in Π is the unique shortest path between its endpoints in G.*
- *For any two nodes $u, v \in V$, there are at most $\frac{\ell}{h(u,v)}$ paths in Π that contain both u and v, where $h(u, v)$ is the hopdistance (number of edges on the shortest path) between u and v in G and $1 \leq h(u, v) \leq \ell$.*
- *Each node $v \in V$ lies on at most $O\left(\frac{\ell p}{n}\right)$ distinct paths in Π.*

We will make use of the shortest path vertex incidence matrix of this graph. Recall that hereditary discrepancy considers the sub-incidence matrix induced by columns corresponding to a set of vertices. We select the set of vertices occurring in the paths in Π, and show it leads to hereditary discrepancy at least $\Omega(n^{1/4}/\sqrt{\log n})$. Specifically, take A as the incidence matrix such that each row corresponds to one path in Π. A has dimension $p \times n$ where n is the number of vertices in G and the (i, j)-th entry of A is 1 is the vertex j is in the path i.

Now define $M = A^T A$. Recall that $\mathrm{tr}(M)$ is the number of 1s in the matrix A. Since by construction, every path has length ℓ, we have $\mathrm{tr}(M) = p\ell$. Furthermore, let m_{ij} be the (i, j)-th element of matrix M, and observe that it is exactly the number of paths that contain vertices i and j. (Note that $m_{ij} = m_{ji}$.) Additionally, $\mathrm{tr}(M^2)$ is the number of length 4 closed walks in the bipartite graph representing the incidence matrix A. This implies that

$$\begin{aligned}\mathrm{tr}(M^2) &= \sum_{j=1}^{p}\sum_{\substack{u,v\in P_j,\\ u\neq v}} m_{u,v} + \sum_{i=1}^{n} m_{ii}^2 = \sum_{j=1}^{p}\sum_{i=1}^{\ell}\sum_{\substack{u,v\in P_j,\\ h(u,v)=i}} m_{u,v} + n\cdot O\left(\frac{p\ell}{n}\right)^2 \\ &\leq \sum_{j=1}^{p}\sum_{i=1}^{\ell} \ell\cdot\frac{\ell}{i} + O\left(\frac{p^2\ell^2}{n}\right) \leq p\ell^2\log\ell + O\left(\frac{p^2\ell^2}{n}\right) = \frac{np\ell^2\log(\ell)+O(p^2\ell^2)}{n}. \quad (1)\end{aligned}$$

By setting $p = n\log n$, it follows that $\ell = \Theta(\sqrt{n}/\log n)$ and $\mathrm{tr}(M) = p\ell$. Further,

$$np\ell^2\log\ell = O(n\cdot n\log n\cdot\frac{n}{\log^2 n}\cdot\log n) = O(n^3) = O(p^2\ell^2).$$

By Equation (1), we have $\mathrm{tr}(M^2) = O(p^2\ell^2/n)$. Using this and $\mathrm{tr}(M) = p\ell$ in the trace bound of [31] gives us

$$\begin{aligned}\mathrm{herdisc}(A) &\geq \frac{(\mathrm{tr}(M))^2}{8e\min\{p,n\}\cdot\mathrm{tr}(M^2)}\sqrt{\frac{\mathrm{tr}(M)}{\max\{p,n\}}} = \frac{(\mathrm{tr}(M))^2}{8en\cdot\mathrm{tr}(M^2)}\sqrt{\frac{\mathrm{tr}(M)}{p}} \\ &\geq \Omega\left(\frac{p^2\ell^2}{p^2\ell^2}\sqrt{\ell}\right) = \Omega(\sqrt{\ell}) = \Omega\left(\frac{n^{1/4}}{\sqrt{\log n}}\right).\end{aligned}$$

The trace bound is formally stated in the full version of our paper [9]. ◀

3.2 Vertex Discrepancy Upper Bound – Explicit Coloring

In this subsection, we will upper bound the discrepancy $\chi(\Pi)$ of a consistent path system (V,Π) with $|V| = n$ and $|\Pi| = \mathrm{poly}(n)$. This will immediately imply an upper bound for the hereditary vertex discrepancy of unique shortest paths in undirected graphs.

▶ **Theorem 13.** *For a consistent path system $S = (V,\Pi)$ where $|V| = n$ and $|\Pi| = \mathrm{poly}(n)$, there exists a labeling χ such that $\chi(\Pi) = O(n^{1/4}\log^{1/2} n)$. Consequently, every n-vertex undirected graph has hereditary vertex discrepancy $O(n^{1/4}\log^{1/2} n)$.*

Let $S = (V,\Pi)$ be a consistent path system with $|V| = n$ and $|\Pi| = \mathrm{poly}(n)$. As the first step towards constructing our labeling $\chi : V \mapsto \{-1,1\}$, we will construct a collection of paths Π' on V that will have a useful covering property over the paths in Π.

Constructing path cover Π'. Initially, we let $\Pi' = \emptyset$. We define V' to be the set of all nodes in V belonging to a path in Π', i.e., $V' := \bigcup_{\pi'\in\Pi'}\pi'$. While $|\pi\setminus V'| \geq n^{1/2}$ for some $\pi\in\Pi$, find a (possibly non-contiguous) subpath of π of length $n^{1/2}$ that is vertex-disjoint from all paths in Π'. Formally, find a subpath $\pi'\subseteq\pi$ such that $|\pi'| = n^{1/2}$ and $\pi'\cap V' = \emptyset$. Add path π' to path cover Π' and update V'. Repeatedly add paths to path cover Π' in this manner until $|\pi\setminus V'| < n^{1/2}$ for all $\pi\in\Pi$.

▶ **Proposition 14.** *Path cover Π' satisfies the following properties:*
1. *for all $\pi\in\Pi'$, $|\pi| = n^{1/2}$,*
2. *the number of paths in Π' is $|\Pi'| \leq n^{1/2}$,*
3. (Disjointness Property). *The paths in Π' are pairwise vertex-disjoint,*
4. (Covering Property). *For all $\pi\in\Pi$, the number of nodes in π that do not lie in any path in path cover Π' is at most $n^{1/2}$. Formally, let $V' = \cup_{\pi'\in\Pi'}\pi'$. Then $\forall\pi\in\Pi, |\pi\setminus V'| \leq n^{1/2}$,*
5. (Consistency Property). *For all $\pi\in\Pi$ and $\pi'\in\Pi'$, the intersection $\pi\cap\pi'$ is a (possibly empty) contiguous subpath of π'.*[4]

[4] Note that it may not be true that $\pi\cap\pi'$ is a contiguous subpath of π.

Proof. Properties 1, 3, and 4 follows from the construction of Π'. Property 2 follows from Properties 1 and 3 and the fact that $|V| = n$. The Consistency Property of Π' is inherited from the consistency of path system S. Specifically, by the construction of Π', path $\pi' \in \Pi'$ is a subpath of a path $\pi'' \in \Pi$. Recall that by the consistency of path system S, the intersection $\pi \cap \pi''$ is a (possibly empty) contiguous subpath of π''. Then $\pi \cap \pi'$ is a contiguous subpath of π' since $\pi' \subseteq \pi''$. This concludes the proof. ◀

Constructing labeling χ. Let $\pi' = (v_1, \dots, v_k) \in \Pi'$ be a path in our path cover. We will label the nodes of π' using the following random process. With probability $1/2$ we define $\chi : \pi' \mapsto \{-1, 1\}$ to be

$$\chi(v_i) = \begin{cases} 1 & i \equiv 0 \mod 2 \text{ and } i \in [1,k] \\ -1 & i \equiv 1 \mod 2 \text{ and } i \in [1,k] \end{cases},$$

and with probability $1/2$ we define $\chi : \pi' \mapsto \{-1, 1\}$ to be

$$\chi(v_i) = \begin{cases} -1 & i \equiv 0 \mod 2 \text{ and } i \in [1,k] \\ 1 & i \equiv 1 \mod 2 \text{ and } i \in [1,k] \end{cases}.$$

The labels of consecutive nodes in π' alternate between 1 and -1, with vertex v_1 taking labels 1 and -1 with equal probability. Since the paths in path cover Π' are pairwise vertex-disjoint, the labeling χ is well-defined over $V' := \cup_{\pi' \in \Pi'} \pi'$. We choose a random labeling for all nodes in $V \setminus V'$, i.e., we independently label each node $v \in V \setminus V'$ with $\chi(v) = -1$ with probability $1/2$ and $\chi(v) = 1$ with probability $1/2$. An illustration can be found in Figure 2.

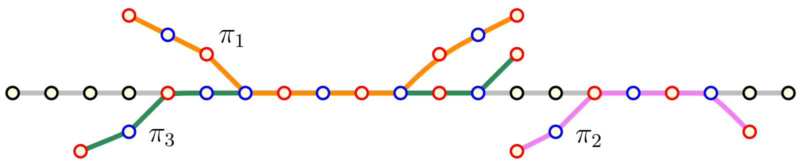

Figure 2 In this figure, paths $\pi_1, \pi_2, \pi_3 \in \Pi'$ from the path cover are intersecting a path $\pi \in \Pi$. Paths in the path cover are pairwise vertex-disjoint, and each path in the cover contributes discrepancy 0, -1, or $+1$ to π.

Bounding the discrepancy $\chi(\Pi)$. Fix a path $\pi \in \Pi$. We will show that $\left|\sum_{v \in \pi} \chi(v)\right| = O(n^{1/4} \log^{1/2} n)$ with high probability. Theorem 13 will follow as $|\Pi| = \text{poly}(n)$.

▶ **Proposition 15.** *For each path π' in path cover Π',*

$$\sum_{v \in \pi \cap \pi'} \chi(v) \in \{-1, 0, 1\}.$$

If $|\pi \cap \pi'| \equiv 0 \mod 2$, then $\sum_{v \in \pi \cap \pi'} \chi(v) = 0$. Moreover,

$$\Pr\left[\sum_{v \in \pi \cap \pi'} \chi(v) = -1\right] = \Pr\left[\sum_{v \in \pi \cap \pi'} \chi(v) = 1\right].$$

Proof. By the Consistency Property of Π' (as proven in Proposition 14), path $\pi \cap \pi'$ is a (possibly empty) contiguous subpath of π'. Then since consecutive nodes in π' alternate between -1 and 1, it follows that $\sum_{v \in \pi \cap \pi'} \chi(v) \in \{-1, 0, 1\}$.

Now note that $\sum_{v \in \pi \cap \pi'} \chi(v) \neq 0$ iff $|\pi \cap \pi'|$ is odd. Moreover, the first vertex of $\pi \cap \pi'$ takes labels 1 and -1 with equal probability. This concludes the proof of Proposition 15. ◀

We are now ready to bound the discrepancy of π.

▶ **Proposition 16.** *With high probability,* $\chi(\pi) = O(n^{1/4} \log^{1/2} n)$.

Proof. We partition the nodes of π into two sources of discrepancy that we will bound separately. Let $V' := \cup_{\pi' \in \Pi'} \pi'$.

Discrepancy of $\pi \cap V'$. For each path $\pi' \in \Pi'$, let $X_{\pi'}$ be the random variable defined as

$$X_{\pi'} := \sum_{v \in \pi \cap \pi'} \chi(v).$$

We can restate the discrepancy of $\pi \cap V'$ as

$$\left| \sum_{v \in \pi \cap V'} \chi(v) \right| = \left| \sum_{\pi' \in \Pi'} X_{\pi'} \right|.$$

By Proposition 15, if $|\pi \cap \pi'| \equiv 0 \mod 2$, then $X_{\pi'} = 0$, so we may assume without any loss of generality that $|\pi \cap \pi'|$ is odd for all $\pi' \in \Pi'$. In this case, $\Pr[X_{\pi'} = -1] = \Pr[X_{\pi'} = 1] = 1/2$, implying that $\mathbb{E}[\sum_{\pi' \in \Pi'} X_{\pi'}] = 0$. Then by Proposition 14 and Chernoff, it follows that for any constant $c \geq 1$,

$$\Pr\left[\left| \sum_{\pi' \in \Pi'} X_{\pi'} \right| \geq c \cdot n^{1/4} \log^{1/2} n \right] \leq e^{-c^2 \frac{n^{1/2} \log n}{2|\Pi'|}} \leq e^{-c^2/(2 \cdot \log(n))} = n^{-c^2/2}.$$

Discrepancy of $\pi \setminus V'$. Note that by the Covering Property of the path cover (as proven in Proposition 14), $|\pi \setminus V'| \leq n^{1/2}$. Moreover, the nodes in $V \setminus V'$ are labeled independently at random, implying that $\mathbb{E}[\sum_{v \in \pi \setminus V'} \chi(v)] = 0$. Then we may apply a Chernoff bound to argue that for any constant $c \geq 1$,

$$\Pr\left[\left| \sum_{v \in \pi \setminus V'} \chi(v) \right| \geq c \cdot n^{1/4} \log^{1/2} n \right] \leq \exp\{-c^2 \frac{n^{1/2} \log n}{2|\pi \setminus V'|}\} \leq e^{-c^2/(2 \cdot \log(n))} = n^{-c^2/2}.$$

We have shown that with high probability, the discrepancy of our labeling is $O(n^{1/4} \log^{1/2} n)$ for $\pi \cap V'$ and $O(n^{1/4} \log^{1/2} n)$ for $\pi \setminus V'$, so we conclude that the total discrepancy of π is $O(n^{1/4} \log^{1/2} n)$, completing the proof of Proposition 16. ◀

Extending to hereditary discrepancy. Let A be the vertex incidence matrix of a path system $S = (V, \Pi)$ on n nodes, and let A_Y be the submatrix of A obtained by taking all of its rows but only a subset Y of its columns. Then there exists a subset $V_Y \subseteq V$ of the nodes in V such that A_Y is the vertex incidence matrix of the path system $S[V_Y]$ (path system S induced on V_Y). Moreover, if path system S is consistent, then $S[V_Y]$ is also consistent. Then we may apply our explicit vertex discrepancy upper bound to $S[V_Y]$. We conclude that the hereditary vertex discrepancy of S is $O(n^{1/4} \log^{1/2} n)$.

3.3 Edge Discrepancy Upper Bound – Explicit Coloring

By Theorem 5, the edge discrepancy of the unique shortest paths of a (possibly directed) graph on m edges is $O(m^{1/4})$. However, in the case of undirected graphs and DAGs, we can improve the edge discrepancy to $O(n^{1/4} \log^{1/2} n)$, where n is the number of vertices in the graph, by modifying the explicit construction for vertex discrepancy in Section 3.2. Our proof strategy will follow the same framework as the explicit construction for vertex discrepancy, but with some added complications in the construction and analysis.

We first introduce some new notation that will be useful in this section. Given a path π and nodes $u, v \in \pi$, we say that $u <_\pi v$ if u occurs before v on path π. Additionally, given a path system $S = (V, \Pi)$, we define the edge set $E \subseteq V \times V$ of the path system as the set of all pairs of nodes $u, v \in V$ that appear consecutively in some path in Π. Likewise, for any path π over the vertex set V, we define the edge set of π, $E(\pi) \subseteq \pi \times \pi$, as the set of all pairs of nodes $u, v \in \pi$ such that u, v appear consecutively in π and $(u, v) \in E$. Note that if path system S corresponds to paths in a graph G, then E will be precisely the edge set of G.

Recall that we wish to construct an edge labeling $\chi : E \mapsto \{-1, 1\}$ so that

$$\chi(\Pi) = \max_{\pi \in \Pi} \left| \sum_{e \in E(\pi)} \chi(e) \right|$$

is minimized. We will upper bound the discrepancy $\chi(\Pi)$ of consistent path systems such that $|V| = n$ and $|\Pi| = \text{poly}(n)$. This will immediately imply an upper bound on the edge discrepancy of unique shortest paths in undirected graphs.

▶ **Theorem 17.** *For all consistent path systems $S = (V, \Pi)$ where $|V| = n$ and $|\Pi| = \text{poly}(n)$ with edge set E, there exists a labeling $\chi : E \to \{-1, 1\}$ such that*

$$\chi(\Pi) = O(n^{1/4} \log^{1/2} n).$$

Consequently, every n-vertex undirected graph has hereditary edge discrepancy $O(n^{1/4} \log^{1/2} n)$.

Let $S = (V, \Pi)$ be a consistent path system with $|V| = n$ and $|\Pi| = \text{poly}(n)$. As the first step towards constructing our labeling $\chi : E \mapsto \{-1, 1\}$, we will construct a collection of paths Π' on V that will have a useful covering property over the paths in Π.

3.3.1 Constructing path cover Π'

Initially, we let $\Pi' = \emptyset$. We define V' to be the set of all nodes in V belonging to a path in Π', i.e.,

$$V' := \bigcup_{\pi' \in \Pi'} \pi'.$$

While there exists a path $\pi \in \Pi$ such that $|\pi \setminus V'| \geq n^{1/2}$, find a (possibly non-contiguous) subpath of π of length $n^{1/2}$ that is vertex-disjoint from all paths in Π'. Specifically, let $\pi' \subseteq \pi$ be a (possibly non-contiguous) subpath of π containing exactly the first $n^{1/2}$ nodes in $\pi \setminus V'$. Add path π' to path cover Π' and update V'. Repeatedly add paths to path cover Π' in this manner until $|\pi \setminus V'| < n^{1/2}$ for all $\pi \in \Pi$.

Note that our path cover Π' is very similar to the path cover used in the explicit vertex discrepancy upper bound. Indeed, path cover Π' inherits all properties of the path cover defined in Subsection 3.2. The key difference here is that we require subpaths $\pi' \subseteq \pi$ in Π' to contain the *first* $n^{1/2}$ nodes in $\pi \setminus V'$. This will imply an additional property of our path cover, which we call the No Repeats Property.

▶ **Proposition 18.** *Path cover Π' satisfies all properties of Proposition 14, as well as the following additional properties:*

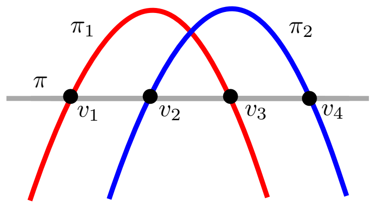

Figure 3 In this figure, paths $\pi_1, \pi_2 \in \Pi'$ are intersecting a path $\pi \in \Pi$. This arrangement of paths is forbidden by the No Repeats Property of Proposition 18.

- (Edge Covering Property). *For all $\pi \in \Pi$, the number of edges in π that are not incident to any node lying in a path in path cover Π' is at most $n^{1/2}$. Formally, let $V' = \cup_{\pi' \in \Pi'} \pi'$. For all $\pi \in \Pi$,*

$$|\{(u, v) \in E(\pi) \mid u \notin V' \text{ and } v \notin V'\}| \leq n^{1/2},$$

- (No Repeats Property). *For all paths $\pi \in \Pi$, $\pi_1, \pi_2 \in \Pi'$, and nodes $v_1, v_2, v_3, v_4 \in \pi$ such that $v_1, v_3 \in \pi_1$ and $v_2, v_4 \in \pi_2$, the following ordering of the vertices in Π is impossible:*

$$v_1 <_\pi v_2 <_\pi v_3 <_\pi v_4,$$

where $x <_\pi y$ indicates that node x occurs in π before node y.

Proof. All properties from Proposition 14 follow from an identical argument as in the original proof. The Edge Covering Property follows immediately from the Covering Property of Proposition 14. What remains is to prove the No Repeats Property.

Suppose for the sake of contradiction that there exist paths $\pi \in \Pi$, $\pi_1, \pi_2 \in \Pi'$, and nodes $v_1, v_2, v_3, v_4 \in \pi$ such that $v_1, v_3 \in \pi_1$ and $v_2, v_4 \in \pi_2$, where $v_1 <_\pi v_2 <_\pi v_3 <_\pi v_4$. We will assume that path π_1 was added to Π' before path π_2 (the case where π_2 was added to Π' first is symmetric). By the construction of Π', path $\pi_1 \in \Pi'$ is a (possibly non-contiguous) subpath of a path $\pi_1'' \in \Pi$ that it was constructed from. Additionally, by the consistency of the path system S, the intersection $\pi \cap \pi_1''$ is a contiguous subpath of π. Then $v_2 \in \pi \cap \pi_1''$, and specifically, $v_2 \in \pi_1''$.

We assumed that $v_2 \in \pi_2$, which implies that $v_2 \notin \pi_1$, since paths in Π' are pairwise vertex-disjoint. Since path π_1 was added to Π' before path π_2, this means that when π_1 was added to Π', node v_2 did not belong to any path in Π' (i.e., v_2 was not in V'). Recall that in our construction of Π', we constructed subpath $\pi_1 \subseteq \pi_1''$ so that it contained exactly the *first* $n^{1/2}$ nodes in $\pi_1'' \setminus V'$. However, $v_2 \notin \pi_1$, but $v_3 \in \pi_1$, and v_2 comes before v_3 in π_1''. This contradicts our construction of path π_1 in path cover Π'. ◀

3.3.2 Constructing labeling χ

Let $\pi' \in \Pi'$ be a path of length k in our path cover. Let $e_1, \dots, e_k \in E(\pi')$ be the edges in π' listed in the order they appear in π'. Note that since π' is a possibly non-contiguous subpath of a path in Π, pairs of nodes $u, v \in V$ that appear consecutively in π do not necessarily correspond to edges in edge set E.

We will label the edges in $E(\pi')$ using the following random process. With probability $1/2$ we define $\chi : E(\pi') \mapsto \{-1, 1\}$ to be

$$\chi(e_i) = \begin{cases} 1 & i \equiv 0 \mod 2 \text{ and } i \in [1, k] \\ -1 & i \equiv 1 \mod 2 \text{ and } i \in [1, k] \end{cases},$$

and with probability 1/2 we define $\chi : E(\pi') \mapsto \{-1, 1\}$ to be

$$\chi(e_i) = \begin{cases} -1 & i \equiv 0 \mod 2 \text{ and } i \in [1, k] \\ 1 & i \equiv 1 \mod 2 \text{ and } i \in [1, k] \end{cases}.$$

Note that the labels of consecutive edges e_i, e_{i+1} in π' alternate between 1 and -1, with edge e_1 taking labels 1 and -1 with equal probability.

Since the paths in path cover Π' are pairwise vertex-disjoint, the labeling χ is well-defined over $E' := \cup_{\pi' \in \Pi'} E(\pi')$. We take a random labeling for all edges in $E \setminus E'$, i.e., we independently label each edge $e \in E \setminus E'$ with $\chi(e) = -1$ with probability 1/2 and $\chi(e) = 1$ with probability 1/2.

3.3.3 Bounding the discrepancy ϕ

Fix a path $\pi := \pi[s, t] \in \Pi$. We will show that

$$\left| \sum_{e \in E(\pi)} \chi(e) \right| = O(n^{1/4} \log^{1/2} n)$$

with high probability. This will complete the proof of Lemma 17 since $|\Pi| = \text{poly}(n)$. The proof of the following proposition follows from an argument identical to Proposition 15 and hence omitted.

▶ **Proposition 19.** *For each path π' in path cover Π',*

$$\sum_{e \in E(\pi) \cap E(\pi')} \chi(e) \in \{-1, 0, 1\}.$$

If $|E(\pi) \cap E(\pi')| \equiv 0 \mod 2$, then $\sum_{e \in E(\pi) \cap E(\pi')} \chi(e) = 0$. Moreover,

$$\Pr\left[\sum_{e \in E(\pi) \cap E(\pi')} \chi(e) = -1 \right] = \Pr\left[\sum_{e \in E(\pi) \cap E(\pi')} \chi(e) = 1 \right].$$

We are now ready to bound the edge discrepancy of π. Define

$$V' := \bigcup_{\pi' \in \Pi'} \pi' \quad \text{and} \quad E' := \bigcup_{\pi' \in \Pi'} E(\pi').$$

We partition the edges of the path π into three sources of discrepancy that we will bound separately. Specifically, we split $E(\pi) \subseteq \pi \times \pi$ into the following sets E_1, E_2, E_3:

- $E_1 := E(\pi) \cap E'$,
- $E_2 := E(\pi) \cap ((V \setminus V') \times (V \setminus V'))$, and
- $E_3 := E(\pi) \setminus (E_1 \cup E_2)$.

Sets E_1 and E_2 roughly correspond to the two sources of discrepancy considered in the vertex discrepancy upper bound, while set E_3 corresponds to a new source of discrepancy that will require new arguments to bound. We begin with set E_1.

▶ **Proposition 20** (Discrepancy of E_1). *With high probability, $\left|\sum_{e \in E_1} \chi(e)\right| = O(n^{1/4} \log^{1/2} n)$.*

Proof. The proposition follow from an argument similar to Proposition 16. For each path $\pi' \in \Pi'$, let $X_{\pi'}$ be the random variable defined as

$$X_{\pi'} := \sum_{e \in E(\pi) \cap E(\pi')} \chi(e).$$

We can restate the discrepancy of E_1 as

$$\left| \sum_{e \in E_1} \chi(e) \right| = \left| \sum_{\pi' \in \Pi'} X_{\pi'} \right|.$$

By Proposition 19, if $|E(\pi) \cap E(\pi')| \equiv 0 \mod 2$, then $X_{\pi'} = 0$, so without any loss of generality, we may assume that $|E(\pi) \cap E(\pi')|$ is odd for all $\pi' \in \Pi'$. In this case,

$$\Pr[X_{\pi'} = -1] = \Pr[X_{\pi'} = 1] = 1/2,$$

implying that $\mathbb{E}[\sum_{\pi' \in \Pi'} X_{\pi'}] = 0$. Then by Proposition 18 and the Chernoff bound, it follows that for any constant $c \geq 1$,

$$\Pr\left[\left| \sum_{\pi' \in \Pi'} X_{\pi'} \right| \geq c \cdot n^{1/4} \log^{1/2} n \right] \leq e^{-c^2 \frac{n^{1/2} \log n}{2|\Pi'|}} \leq e^{-c^2/2 \cdot \log n} \leq n^{-c^2/2}. \qquad \blacktriangleleft$$

We now bound the discrepancy of E_2.

▶ **Proposition 21** (Discrepancy of E_2). *With high probability,* $\left|\sum_{e \in E_2} \chi(e)\right| = O(n^{1/4} \log^{1/2} n)$.

Proof. The proposition follow from an argument similar to Proposition 16. Note that by the Edge Covering Property of the path cover (Proposition 18),

$$|E_2| = |\{(u, v) \in E(\pi) \mid u, v \notin V'\}| \leq n^{1/2}.$$

Moreover, the edges in $E \setminus E'$ are labeled independently at random, so we may apply a Chernoff bound to argue that for any constant $c \geq 1$,

$$\Pr\left[\left| \sum_{e \in E_2} \chi(e) \right| \geq c \cdot n^{1/4} \log^{1/2} n \right] \leq e^{-c^2 \frac{n^{1/2} \log n}{2|E_2|}} \leq e^{-c^2/2 \cdot \log n} \leq n^{-c^2/2}.$$

completing the proof. ◀

Finally, we upper bound the discrepancy of E_3.

▶ **Proposition 22** (Discrepancy of E_3). *With high probability,* $\left|\sum_{e \in E_3} \chi(e)\right| = O(n^{1/4} \log^{1/2} n)$.

Proof. Let

$$k := |\{\pi' \in \Pi' \mid \pi \cap \pi' \neq \emptyset\}|$$

denote the number of paths in our path cover that intersect π. We define a function $f : \mathbb{Z}_{\geq 0} \mapsto \mathbb{Z}_{\geq 0}$ such that $f(\phi)$ equals the largest possible value of $|E_3|$ when $\phi = k$. Note that f is well-defined since $0 \leq |E_3| \leq |E|$. We will prove that $f(\phi) \leq 4\phi$, by recursively decomposing path π.

When $\phi = 1$, there is only one path $\pi' \in \Pi'$ that intersects π. Then the only edges in E_3 are of the form

$$E(\pi) \cap ((V' \times (V \setminus V')) \cup ((V \setminus V') \times V')) = E(\pi) \cap ((\pi' \times (V \setminus \pi')) \cup ((V \setminus \pi') \times \pi')).$$

By the Consistency Property of Proposition 18, path π' can intersect π and then split apart at most once. Then

$$f(1) = |E_3| = |E(\pi) \cap ((\pi' \times (V \setminus \pi')) \cup ((V \setminus \pi') \times \pi'))| \leq 2.$$

When $\phi > 1$, we will split our analysis into the two cases:

- **Case 1.** There exists paths $\pi_1', \pi_2' \in \Pi'$ and nodes $v_1, v_2, v_3 \in \pi$ such that $v_1, v_3 \in \pi_1'$ and $v_2 \in \pi_2'$ and $v_1 <_\pi v_2 <_\pi v_3$. In this case, we can assume without any loss of generality that $\pi[v_1, v_3] \cap \pi_1' = \{v_1, v_3\}$ (e.g., by choosing v_1, v_3 so that this equality holds). Let x be the node immediately following v_1 in π, and let y be the node immediately preceding v_3 in π. Recall that s is the first node of π and t is the last node of π. It will be useful for the analysis to split π into three subpaths:

  $$\pi = \pi[s, v_1] \circ \pi[x, y] \circ \pi[v_3, t],$$

  where $\circ$ denotes the concatenation operation. Let

  $$\phi_1 := |\{\pi' \in \Pi' \mid \pi[x, y] \cap \pi' \neq \emptyset\}| \text{ and } \phi_2 := |\{\pi' \in \Pi' \mid (\pi[s, v_1] \circ \pi[v_3, t]) \cap \pi' \neq \emptyset\}|.$$

  We claim that $\phi_1 < \phi$, $\phi_2 < \phi$, and $\phi_1 + \phi_2 = \phi$. We will use these facts to establish a recurrence relation for f. By our assumption that $\pi[v_1, v_3] \cap \pi_1' = \{v_1, v_3\}$, it follows that $\pi[x, y] \cap \pi_1' = \emptyset$, and so $\phi_1 < \phi$. Likewise, by the No Repeats Property of Proposition 18,

  $$(\pi[s, v_1] \circ \pi[v_3, t]) \cap \pi_2' = \emptyset,$$

  so $\phi_2 < \phi$. Finally, observe that more generally, if there exists a path $\pi' \in \Pi'$ such that $\pi' \cap \pi[x, y] \neq \emptyset$ and $\pi' \cap (\pi[s, v_1] \circ \pi[v_3, t]) \neq \emptyset$, then the No Repeats Property of Proposition 18 is violated. We conclude that $\phi_1 + \phi_2 = \phi$.
  Now $|E_3|$ can be upper bounded by the following inequality:

  $$|E_3| \leq |E_3 \cap E(\pi[x, y])| + |E_3 \cap E(\pi[s, v_1] \circ \pi[v_3, t])| + 2.$$

  Then using the observations about ϕ_1, ϕ_2, and ϕ in the previous paragraph, we obtain the following recurrence for f:

  $$f(\phi) \leq f(\phi_1) + f(\phi_2) + 2 = f(i) + f(\phi - i) + 2,$$

  where $0 < i < \phi$.
- **Case 2.** There exists a path $\pi' \in \Pi'$ and $v_1, v_2 \in \pi$ such that $\pi \cap \pi' = \pi[v_1, v_2] \cap V'$. Let x be the node immediately preceding v_1 in π, and let y be the node immediately following v_2 in π. Again, we split π into three subpaths:

  $$\pi[s, t] = \pi[s, x] \circ \pi[v_1, v_2] \circ \pi[y, t].$$

  Let

  $$\phi_1 := |\{\pi' \in \Pi' \mid \pi[v_1, v_2] \cap \pi' \neq \emptyset\}| \text{ and } \phi_2 := |\{\pi' \in \Pi' \mid (\pi[s, x] \circ \pi[y, t]) \cap \pi' \neq \emptyset\}|.$$

By our assumption in Case 2, it follows that $\phi_1 = 1$ and $\phi_2 = \phi - 1$. Since $|E_3|$ can be upper bounded by the inequality

$$|E_3| \le |E_3 \cap E(\pi[v_1, v_2])| + |E_3 \cap E(\pi[s,x] \circ \pi[y,t])| + 2,$$

we immediately obtain the recurrence

$$f(\phi) \le f(\phi_1) + f(\phi_2) + 2 \le f(1) + f(\phi - 1) + 2.$$

Taking our results from Case 1 and Case 2 together, we obtain the recurrence relation

$$f(\phi) \le \begin{cases} \max\{f(i) + f(\phi - i) + 2,\ f(1) + f(\phi-1) + 2\} & \phi > 1 \text{ and } 1 < i < \phi \\ 2 & \phi = 1 \end{cases}$$

Applying this recurrence $\le \phi$ times, we find that

$$f(\phi) \le \phi \cdot f(1) + 2\phi \le 4\phi.$$

Finally, since $k \le |\Pi'| \le n^{1/2}$ and we defined f so that $f(k)$ equals the largest possible value of $|E_3|$, we conclude that

$$|E_3| \le f(k) \le f(n^{1/2}) = O(n^{1/2}).$$

Since the edges in $E_3 \subseteq E \setminus E'$ are labeled independently at random, we may apply a Chernoff bound as in Proposition 21 to argue that $\chi(E_3) = O(n^{1/4} \log^{1/2} n)$ with high probability. ◀

We have shown that with high probability, the discrepancy of our edge labeling is $O(n^{1/4} \log^{1/2} n)$ for E_1, E_2, and E_3, so we conclude that the total discrepancy of π is $O(n^{1/4} \log^{1/2} n)$. A straightforward extension of this argument implies identical bounds for hereditary discrepancy. We defer this proof to the full version of our paper [9].

4 Conclusion and Open Problems

This paper reported new bounds on the hereditary discrepancy of set systems of unique shortest paths in graphs. An open problem is to improve our edge discrepancy upper bound in directed graphs. Standard techniques in discrepancy theory imply an upper bound of $\min\{O(m^{1/4}), \widetilde{O}(D^{1/2})\}$ for this problem, leaving a gap with our $\Omega(n^{1/4}/\sqrt{\log n})$ lower bound when $m = \omega(n)$. Unfortunately, we were not able to extend our low-discrepancy edge and vertex coloring arguments for undirected graphs to the directed setting, due to the pathological example in Figure 4.

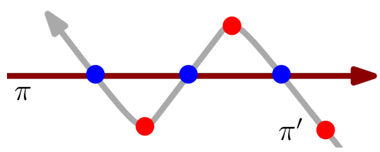

Figure 4 An example in directed graphs that demonstrates how coloring unique shortest paths with alternating colors can fail to imply low discrepancy.

References

1 Ittai Abraham, Amos Fiat, Andrew V Goldberg, and Renato F Werneck. Highway dimension, shortest paths, and provably efficient algorithms. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 782–793. SIAM, 2010.

2 Shyan Akmal and Nicole Wein. A local-to-global theorem for congested shortest paths. In *31st Annual European Symposium on Algorithms (ESA 2023)*, pages 8:1–8:17, 2023.

3 J Ralph Alexander, József Beck, and William WL Chen. Geometric discrepancy theory and uniform distribution. *Handbook of discrete and computational geometry*, page 279, 2004.

4 Saeed Akhoondian Amiri and Julian Wargalla. Disjoint shortest paths with congestion on DAGs. *arXiv preprint*, 2020. arXiv:2008.08368.

5 Lorenzo Balzotti. Non-crossing shortest paths are covered with exactly four forests. *arXiv preprint*, 2022. arXiv:2210.13036.

6 Hannah Bast, Daniel Delling, Andrew Goldberg, Matthias Müller-Hannemann, Thomas Pajor, Peter Sanders, Dorothea Wagner, and Renato F Werneck. Route planning in transportation networks. *Algorithm engineering: Selected results and surveys*, pages 19–80, 2016.

7 Johannes Blum, Yann Disser, Andreas Emil Feldmann, Siddharth Gupta, and Anna Zych-Pawlewicz. On Sparse Hitting Sets: From Fair Vertex Cover to Highway Dimension. In Holger Dell and Jesper Nederlof, editors, *17th International Symposium on Parameterized and Exact Computation (IPEC 2022)*, volume 249 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 5:1–5:23, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.IPEC.2022.5.

8 Greg Bodwin. On the structure of unique shortest paths in graphs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2071–2089. SIAM, 2019.

9 Greg Bodwin, Chengyuan Deng, Jie Gao, Gary Hoppenworth, Jalaj Upadhyay, and Chen Wang. The discrepancy of shortest paths. *arXiv preprint*, 2024. arXiv:2401.15781.

10 Greg Bodwin and Gary Hoppenworth. Folklore sampling is optimal for exact hopsets: Confirming the $\sqrt{n}$ barrier. In *Proceedings of the 64th IEEE Symposium on Foundations of Computer Science (FOCS) 2023*, pages 701–720, 2023.

11 Hsien-Chih Chang, Jonathan Conroy, Hung Le, Lazar Milenkovic, and Shay Solomon. Covering planar metrics (and beyond): O(1) trees suffice. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2231–2261, 2023. doi:10.1109/FOCS57990.2023.00139.

12 Hsien-Chih Chang, Pawel Gawrychowski, Shay Mozes, and Oren Weimann. Near-optimal distance emulator for planar graphs. In *26th Annual European Symposium on Algorithms (ESA 2018)*, pages 16:1–16:17, 2018.

13 Hsien-Chih Chang, Robert Krauthgamer, and Zihan Tan. Near-linear ε-emulators for planar graphs. *arXiv preprint*, 2022. arXiv:2206.10681.

14 Bernard Chazelle. *The Discrepancy Method: Randomness and Complexity*. Cambridge University Press, 2000.

15 Bernard Chazelle and Alexey Lvov. A trace bound for the hereditary discrepancy. In *Proceedings of the sixteenth annual symposium on Computational geometry*, SCG '00, pages 64–69, New York, NY, USA, May 2000. Association for Computing Machinery.

16 Justin Y. Chen, Badih Ghazi, Ravi Kumar, Pasin Manurangsi, Shyam Narayanan, Jelani Nelson, and Yinzhan Xu. Differentially private all-pairs shortest path distances: Improved algorithms and lower bounds. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 5040–5067, 2023.

17 Maria Chudnovsky, Daniel Cizma, and Nati Linial. The structure of metrizable graphs. *arXiv preprint*, 2023. arXiv:2311.09364.

18 Fan R. K. Chung, Ronald L. Graham, and Richard M. Wilson. Quasi-random graphs. *Combinatorica*, 9:345–362, 1989.

19 Daniel Cizma and Nati Linial. Geodesic geometry on graphs. *Discrete & Computational Geometry*, 68(1):298–347, 2022.

20 Daniel Cizma and Nati Linial. Irreducible nonmetrizable path systems in graphs. *Journal of Graph Theory*, 102(1):5–14, 2023.

21 Don Coppersmith and Michael Elkin. Sparse sourcewise and pairwise distance preservers. *SIAM Journal on Discrete Mathematics*, 20(2):463–501, 2006.

22 Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2022.

23 Chengyuan Deng, Jie Gao, Jalaj Upadhyay, and Chen Wang. Differentially private range query on shortest paths. In *Algorithms and Data Structures Symposium*, pages 340–370. Springer, 2023.

24 Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, pages 265–284. Springer, 2006.

25 Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

26 Marek Eliáš, Michael Kapralov, Janardhan Kulkarni, and Yin Tat Lee. Differentially private release of synthetic graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 560–578. SIAM, 2020.

27 Chenglin Fan, Ping Li, and Xiaoyun Li. Private graph all-pairwise-shortest-path distance release with improved error rate. In *Proceedings of Advances in Neural Information Processing Systems*, volume 35, pages 17844–17856, 2022.

28 Sariel Har-Peled. A simple proof of the existence of a planar separator. *arXiv preprint*, 2011. `arXiv:1105.0103`.

29 Monika R Henzinger, Philip Klein, Satish Rao, and Sairam Subramanian. Faster shortest-path algorithms for planar graphs. *Journal of Computer and System Sciences*, 55(1):3–23, 1997.

30 Shimon Kogan and Merav Parter. New diameter-reducing shortcuts and directed hopsets: Breaking the barrier. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1326–1341. SIAM, 2022.

31 Kasper Green Larsen. Constructive discrepancy minimization with hereditary L2 guarantees. In Rolf Niedermeier and Christophe Paul, editors, *36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019)*, volume 126 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 48:1–48:13, Dagstuhl, Germany, 2019. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

32 Jingcheng Liu, Jalaj Upadhyay, and Zongrui Zou. Optimal bounds on private graph approximation. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1019–1049. SIAM, 2024.

33 Jiři Matoušek. *Geometric Discrepancy: An Illustrated Guide*. Springer Science & Business Media, May 1999.

34 Shanmugavelayutham Muthukrishnan and Aleksandar Nikolov. Optimal private halfspace counting via discrepancy. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1285–1292, 2012.

35 János Pach and Pankaj K Agarwal. *Combinatorial Geometry*. John Wiley & Sons, October 2011.

36 Adam Sealfon. Shortest paths and distances with differential privacy. In *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, pages 29–41, 2016.

37 Csaba D Toth, Joseph O'Rourke, and Jacob E Goodman. *Handbook of discrete and computational geometry*. CRC Press, 2017.