Check for updates

On Fully Diverse Sets of Geometric Objects and Graphs

Fabian Klute(✉) and Marc van Kreveld

Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
{f.m.klute,m.j.vankreveld}@uu.nl

Abstract. Diversity is a property of sets that shows how varied or different its elements are. We define full diversity in a metric space and study the maximum size of fully diverse sets. A set is *fully diverse* if each pair of elements is as distant as the maximum possible distance between any pair, up to a constant factor. We study metric spaces based on geometry, embeddings of graphs, and graphs themselves. In the geometric cases, we study measures like Hausdorff distance, Frechét distance, and area of symmetric difference between objects in a bounded region. In the embedding cases, we study planar embeddings of trees and planar graphs, and use the number of swaps in the rotation system as the metric. In the graph cases, we use the number of insertions and deletions of leaves or edges as the metric. In most cases, we show (almost) tight lower and upper bounds on the maximum size of fully diverse sets. Our results lead to a very simple randomized algorithm to generate large fully diverse sets in several cases.

Keywords: Diversity · Distance Measures · Diverse Geometric Objects · Diverse Graphs · Diverse Embeddings

1 Introduction

When generating data, for example for benchmarks, it may be important that the generated set is sufficiently diverse. The same is true for systems that assist in choosing a desired layout or configuration by showing various options. For example, in graph drawing this observation has led to systems that present several drawings of a graph. A user can now choose a drawing, or indicate preference, after which more drawings like the preferred one can be generated [2].

But what does diversity mean in this context? We address this question in a formal way. We introduce a framework that allows us to study diversity of "objects", and analyze the maximum number of objects that are pairwise far apart. This framework is applicable in many contexts.

Supported by the Netherlands Organisation for Scientific Research (NWO) under project no. 612.001.651 and the Austrian Science Foundation (FWF) grant J4510.

M. A. Bekos and M. Kaufmann (Eds.): WG 2022, LNCS 13453, pp. 328–341, 2022.
https://doi.org/10.1007/978-3-031-15914-5_24

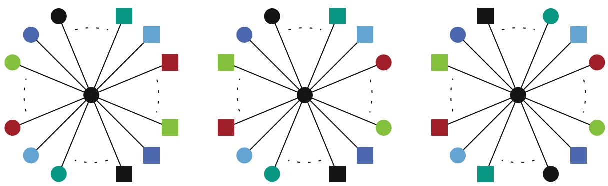


Fig. 1. Fully diverse set of labeled (color+symbol) n-vertex stars. Opposite leaves have the same color but a different symbol. The second embedding has one opposite-pair exchange per triple when compared to the first embedding, and the third embedding has two opposite-pair exchanges per triple. Any two of the three embeddings have distance $\Omega(n^2)$ if distance is measured by the number of swaps of adjacent edges.

Diversity as a Counting Problem. Let $(\mathcal{S}, \mu)$ be a metric space where $\mathcal{S}$ is a base set or class of objects and μ is a distance measure that takes a pair from $\mathcal{S}$ and assigns a distance. We consider the cases where $\mu(a, b)$ is bounded for all $a, b \in \mathcal{S}$; let $M = \sup_{a,b\in\mathcal{S}} \mu(a, b)$ be the highest value that is attained (possibly in the limit) by μ on $\mathcal{S}$.

Definition 1. *For a given $c \geq 1$, a subset $\hat{\mathcal{S}} \subseteq \mathcal{S}$ is called* $\frac{1}{c}$-diverse *if for all $x, y \in \hat{\mathcal{S}}$, we have $\mu(x, y) \geq \frac{1}{c} \cdot M$. If c can be chosen constant, independent of $|\hat{S}|$, then $\hat{\mathcal{S}}$ is called* fully diverse.

Intuitively, we relate the distance of all pairs of the subset to the maximum distance within the base set. We are interested in the question how large fully diverse subsets can be. As a simple example, consider all points in a unit square region in the plane and Euclidean distance as the metric. Then the maximum distance is $\sqrt{2}$, so a $\frac{1}{c}$-diverse (sub)set of points must have pairwise distances of at least $\sqrt{2}/c$. It is easy to see that any $\frac{1}{c}$-diverse set has size $O(c^2)$ by a packing argument, and any maximal $\frac{1}{c}$-diverse set has size $\Omega(c^2)$. The maximum size of a fully diverse set of points is $\Theta(c^2) = \Theta(1)$ if c is a constant.

When considering more complex objects, like polygonal lines, triangulations, drawings of graphs, and graphs themselves, we need a distance between any two objects. We consider geometric distance measures for geometric objects and discrete measures for graphs. In several geometric cases, we need to assume that the objects reside in a bounded space to make the metric space bounded. Let U be a unit diameter disk in the plane. Some geometric distance measures are:

- For any two simple polygons inside U, their Hausdorff distance.
- For any two polygonal lines inside U, the Fréchet distance between them.
- For two simple polygons inside U, their symmetric difference.
- For two drawings of a given labeled graph inside U, the total vertex displacement (summed distance between vertices with the same label).
- For any two drawings of a given labeled graph with the same embedding, the L_1-distance of the vector of angles of adjacent edges (sum of differences of corresponding angles).

The maximum distance between any two objects in these cases is bounded by 1, 1, $\pi/4$, $|V|$, and $4\pi|V|$, respectively.

Some discrete measures for two embeddings of a given labeled graph are:

- For two embeddings of a labeled graph, the L_1-distance of the vectors on the edges, where an edge gives 1 if it intersects any other edge and 0 otherwise.
- For two embeddings of a labeled tree, the number of swaps of adjacent edges at a vertex to convert the embedding of one into the other.
- For two planar embeddings of a labeled graph, the number of swaps around cut-vertices and split pairs to convert one into the other.

See Fig. 1 for a fully diverse set of embeddings of a labeled star graph.

For general graphs (independent of embedding), we study measures based on the number of additions and removals of edges or leaves; in other words, the edit distance for a set of possible edits [10].

Relation to Diversity and Similar Notions in Science. Diversity has been studied in a variety of scientific contexts. One well-known example is in ecosystems, specifically, the diversity of species that are represented in a sample of animals or plants, see for instance [12,21]. The Shannon index is commonly used, also known as Shannon entropy in information theory.

In computer science, diversity has been studied in a variety of areas. For example, the diversity of the output in selection tasks in big data [9] or recommender systems [17], the diversity of input data sets for machine learning [16], or the diversity of colored point sets in computational geometry [15].

Diversity without a priori assigned categories is of interest in the study of the diversity of a population in genetic algorithms, e.g. [23]. Following similar ideas, researchers later studied the diversity of sets of solutions in satisfiability problems [14], multicriteria optimization problems [22], and, recently, parameterized algorithms, e.g. [3]. Similar to our work, the diversity measures found in this line of work are commonly based on the Hamming distance. Hebrard et al. [11] introduced the maximization problem to find the set of solutions that maximizes this sum or minimum distance over all sets of solutions.

With respect to drawings of graphs, Biedl et al. [5] studied how to heuristically generate a set of different not necessarily planar straight-line drawings. Their measure of distance between any two drawings is obtained by greedily matching vertices using a composite measure of Euclidean distance and position in the drawing. Bridgeman and Tamassia [7] investigated geometric measures for the distance between two orthogonal drawings.

Counting the overall number of structures such as planar triangulations or crossing-free geometric graphs on a point set (without requiring that they pairwise be far) has been widely studied, e.g. in [13]. See also the blog entry by Sheffer with a list of references on this topic.[1] Finally, for a given planar graph the number of embeddings it admits has been studied in the context of algorithms to count them [8,20].

[1] https://adamsheffer.wordpress.com/numbers-of-plane-graphs/.

Table 1. Lower and upper bounds on the maximum size of fully diverse sets in various metric spaces where the measure is geometric. U denotes a unit diameter disk.

Object	Metric	Space	Diameter	Lower bound	Upper bound
Polygons	Hausdorff distance	U	$\Theta(1)$	$\Omega(1)$	$O(1)$
Polylines	Fréchet distance	U	$\Theta(1)$	$2^{\Omega(n)}$	$2^{O(n)}$
Polygons	Area symm. diff.	U	$\Theta(1)$	$2^{\Omega(n)}$	$2^{O(n \log n)}$

The Approach and the Results. We initiate the study of the size of fully diverse sets in bounded metric spaces, as described in this introduction. We believe that such a study is important in all algorithmic problems where different objects are generated, for example, in graph drawing and benchmark construction.

We use a unified approach to obtain our results, which we can explain best on bit strings of length n with the metric the number of bit flips (or Hamming distance). We choose a sufficiently large constant c. Then we show that for any bit string, the number of bit strings that can be obtained by at most n/c bit flips is bounded from above, while the total number of bit strings is 2^n. Dividing the latter quantity (2^n) by the former gives a lower bound on the maximum size of a fully diverse set. When we apply this simple scheme to the various metric spaces listed before, we encounter different kinds of challenges.

In Sect. 2 we investigate the maximum size of a fully diverse set of fair bit strings (fair = equally many 0s as 1s). We show that for fair bit strings of length n, the maximal size of a fully diverse set is exponential. The same is true for fair *cyclic* bit strings. These results are used as a core ingredient in later proofs.

In Sect. 3 we present results for three metric spaces where distance is geometric. Bounds on the maximum size of fully diverse sets are given in Table 1, where diameter specifies the maximum distance in the metric space. The main challenge is the suitable discretization of the space of all possible polylines or polygons, so that the desired distance between pairs can be analyzed.

Next we consider embeddings of trees and planar graphs, that is, the cyclic order of neighboring nodes, in Sect. 4. The metric is the minimum number of swaps of adjacent neighbors to get from the one embedding to the other. Our results are given in Table 2.

In Sect. 5 we consider graphs as combinatorial objects and base the metric on edit distance. We distinguish labeled and unlabeled graphs, and consider trees, planar graphs, and general graphs. Table 3 gives the results.

There exists a very simple randomized algorithm to generate fully diverse sets of size k (provided k is small enough). It works as follows, starting with an empty set S and a constant $c \geq 1$, and a known maximum diameter M of the base set: (i) Generate a random element e from the base set. (ii) Test if e has distance at least M/c to all elements in S. If so, add e to S, and if not, discard it and continue at (i). Stop when set S contains k elements. This algorithm leads to fully diverse sets of large size with high probability for several examples in this paper, if distances can be computed easily.

Table 2. Lower and upper bounds on the maximum size of fully diverse sets in various metric spaces concerning embedded labeled graphs with n nodes. The lower bound for trees holds for any tree, whereas the upper bound holds for some trees (Theorem 4).

Object	Metric	Diameter	Lower bound	Upper bound
Ternary trees	# adjacent swap	$\Theta(n)$	$2^{\Omega(n)}$	$2^{O(n)}$
Star graphs	# adjacent swap	$\Theta(n^2)$	$2^{\Omega(n)}$	$2^{O(n \log n)}$
Trees	# adjacent swap	$\Theta(\sum_{v \in V} \deg^2(v))$	$2^{\Omega(\sqrt{n})}$	$[\, 2^{O^*(\sqrt{n})} \,]$
Star graphs	# any swap	$\Theta(n)$	$2^{\Omega(n \log n)}$	$2^{O(n \log n)}$
Planar graphs	# adjacent swap	Theorem 5	Theorem 5	Theorem 5

Table 3. Lower and upper bounds on the maximum size of fully diverse sets in various metric spaces where the measure is edit distance and the objects are graphs with n nodes. Intermediate graphs must be in the same class.

Object	Metric	Diameter	Lower bound	Upper bound
Trees	# reattach leaf	$\Theta(n)$	$2^{\Omega(n \log n)}$	$2^{O(n \log n)}$
Planar graphs	# insert/delete edge	$\Theta(n)$	$2^{\Omega(n \log n)}$	$2^{O(n \log n)}$
Graphs	# insert/delete edge	$\Theta(n^2)$	$2^{\Omega(n^2)}$	$2^{O(n^2)}$
Trees (unlabeled)	# reattach leaf	$\Theta(n)$	$2^{\Omega(n)}$	$2^{O(n)}$
Planar graphs (unlab.)	# insert/delete edge	$\Theta(n)$	$n^{\Omega(1)}$	$2^{O(n)}$
Graphs (unlabeled)	# insert/delete edge	$\Theta(n^2)$	$2^{\Omega(n^2)}$	$2^{O(n^2)}$

2 Fair Bit Strings

Let B be a bit string of length $n \geq 8$. We say that B is a *fair* bit string if it contains at least $\lfloor \frac{n}{2} \rfloor$ ones and at least $\lfloor \frac{n}{2} \rfloor$ zeros. Moreover, we say two fair bit strings B_1 and B_2 of length n are *far* if they differ in at least $\lfloor \frac{n}{8} \rfloor$ positions. Conversely, if B_1 and B_2 are not far we say they are *close*. Since rounding does not influence our results, we omit rounding to integers from now on. We obtain the following lemma using a bound by Robbins [19] and Stirling's approximation.

Lemma 1 ($\star$). *Let B be a fair bit string with n bits, the number of fair bit strings close to B is at most*

$$\frac{2}{3\pi} \cdot \left(\frac{256}{27}\right)^{n/4} = O(1.754...^n).$$

Lemma 1 allows us to show (in Lemma 2) that there are exponentially many fair bit strings of length n that are all pairwise far from each other when we consider the number of bit flips as the distance measure. Since we need at most n bit flips to transform any bit string of length n into any other, upper-bounding M in Definition 1, a set of pairwise far fair bit strings is fully diverse.

Lemma 2. *For fair bit strings of length n, any maximal fully diverse set of fair bit strings, using Hamming distance, has size at least*

$$\Omega\left(n^{-1} \cdot \left(\frac{27}{16}\right)^{n/4}\right) = \Omega\left(1.139...^n\right).$$

Proof. Since $\sum_{i=0}^{n} \binom{n}{i} = 2^n$ and $\binom{n}{n/2}$ is the largest term of $n+1$ terms, it is at least $2^n/n$ (taking the first and last term as one term), which is a lower bound on the number of fair bit strings of length n.

A maximal set of fully diverse fair bit strings can be obtained by starting with the set of all fair bit strings, selecting any member, removing all that are close, and repeating. By Lemma 1, we know how many we maximally delete in one step, so the number of iterations (and size of a maximal set of fair fully diverse bit strings) is at least

$$\frac{2^n/n}{\frac{2}{3\pi} \cdot \left(\frac{256}{27}\right)^{n/4}} = \frac{3\pi \cdot 2^n \cdot 27^{n/4}}{2n \cdot 256^{n/4}} \geq \Omega\left(n^{-1} \cdot \left(\frac{27}{16}\right)^{n/4}\right) = \Omega(1.139...^n). \qquad \square$$

When considering *cyclic* bit strings (a bit string is equivalent to any of its $n-1$ cyclically shifted versions), the above analysis does not apply directly. For two fair cyclic bit strings B_1 and B_2 of length n, we say that B_1 and B_2 are *far* if they differ in at least $\frac{n}{8}$ positions for all of their cyclically shifted versions. Conversely, if B_1 and B_2 are not far we say they are *close*.

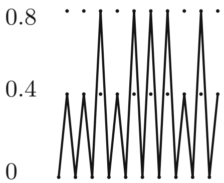
Lemma 3 ($\star$). *For fair cyclic bit strings of length n, any maximal fully diverse set of fair cyclic bit strings, using Hamming distance, has size*

$$\Omega\left(n^{-2} \cdot \left(\frac{27}{16}\right)^{n/4}\right) = \Omega\left(1.139...^n\right).$$

3 Geometric Diversity

Given two closed subsets A and B of a metric space, the *Hausdorff distance* between A and B is defined as the maximum distance of any point in A to its closest point in B or vice versa. For the *Fréchet distance* let A and B be two curves in the plane. Informally, the Fréchet distance between A and B is the minimum length of a leash that allows a person to walk along A and a dog along B with neither of them ever walking backwards. See Alt and Godau [1] for the formal definitions. The area of symmetric difference between two polygons is the total area inside exactly one of the polygons.

Let $\mathcal{S}$ be any set of simple polygons inside a unit diameter disk U. Any two polygons inside U have Hausdorff distance ≤ 1. Assume $\mathcal{S}$ is fully diverse, so a constant $c \geq 1$ exists such that for any two $P_i, P_j \in \mathcal{S}$ ($i \neq j$), their Hausdorff distance is at least $1/c$. We partition U by horizontal and vertical lines spaced $1/(2c)$, resulting in $O(c^2)$ cells. If P_i and P_j occupy exactly the same cells of this grid, then their Hausdorff distance is at most $1/(\sqrt{2}c) < 1/c$, a contradiction, so there must be a cell occupied by exactly one of P_i and P_j. This property holds for every pair of polygons in $\mathcal{S}$, so $\mathcal{S}$ cannot contain more than $2^{O(c^2)} = O(1)$ polygons and be fully diverse. The size of $\mathcal{S}$ does not depend on the descriptive complexity of the polygons. The upper bound also applies to polygons with holes or that are disconnected, and to drawings of graphs.



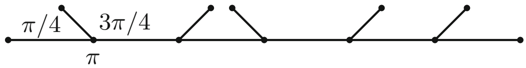


Fig. 2. Left, lower bound construction for Fréchet distance. Right, lower bound construction for sum of angle differences by encoding the bit string 01011.

Theorem 1. *A fully diverse set of polygons inside a bounded region has size $O(1)$ when we measure the distance by the Hausdorff distance.*

Next we show by construction that a fully diverse set of polygonal lines with n vertices inside a unit diameter disk U can have exponential size when using the Fréchet distance. We choose points on three horizontal lines $y = 0$, $y = 0.4$, and $y = 0.8$; on the first line we take points with x-coordinates i/n for $1 \leq i \leq n/2$, and on the second and third line we take points with x-coordinates $i/n + 1/(2n)$ for $1 \leq i \leq n/2$. We make x-monotone polygonal lines by using all points on the line $y = 0$, and between two such points, we choose either the point on $y = 0.4$ or on $y = 0.8$. See Fig. 2(left). Any two of the $2^{n/2}$ different options has Fréchet distance at least 0.4, hence these options together give a set of size $2^{\Omega(n)}$ that is fully diverse. The construction is easily adapted to simple polygon boundaries.

For area of symmetric difference, we can use the construction in Fig. 2(left) if we add one vertex at the bottom right to close the polyline with one straight-line segment to a polygon. Having $\Omega(n)$ spikes different implies an area of symmetric difference of $\Omega(1)$. Hence, the spikes encode the bits of a bit string, and Lemma 2 gives the lower bound.

We obtained $2^{\Omega(n)}$ lower bounds on the size of fully diverse sets in two cases. Is it the right lower bound, or can we also achieve a bound like $2^{\Omega(n \log n)}$?

Concerning the Fréchet distance, assume a unit diameter disk U and let a constant $c \geq 1$ be given. We partition U by a square grid of line spacing $1/(2c)$, so that any two points in the same grid cell have distance $< 1/c$. There are $O(c^2)$ cells, which is constant. We can encode any polyline of n vertices by the sequence of cells in which the vertices lie. It is straightforward to see that two polylines that have the same sequence of cells, have Fréchet distance $< 1/c$, so they cannot be in the same fully diverse set. Consequently, the size of a fully diverse set is bounded by the number of sequences of cells: $(O(c^2))^n = 2^{O(n)}$.

Theorem 2. *A fully diverse set of polygonal lines or simple polygon boundaries with n vertices in a bounded region, may have size $2^{\Omega(n)}$ and has size at most $2^{O(n)}$, if distance is measured by the Fréchet distance.*

For area of symmetric difference we need a much finer grid in order to ensure that visiting the same cells implies a distance of at most $\pi/(4c)$. Consider a grid with cells of diameter $< 1/(2cn)$. Then two simple polygons that have the same vertices in the same cells in the same order have an area of symmetric difference of at most $1/(\sqrt{2}c) < \pi/(4c)$ because each pair of corresponding edges causes a symmetric difference of at most $\sqrt{2}/(2cn)$. This leads to an upper bound of $((2cn)^2)^n = 2^{O(n \log n)}$.

Theorem 3. *A fully diverse set of simple polygons with n vertices, may have size $2^{\Omega(n)}$ and has size at most $2^{O(n \log n)}$, if distance is measured by the area of symmetric difference.*

Remark 1. The techniques presented in Sects. 2 and 3 are quite versatile. Without any new ideas, we can also show that for drawings of labeled star graphs in a bounded region, the maximum size of a fully diverse set is $2^{\Theta(n)}$ when distance is measured as sum of vertex displacements. The lower bound uses an encoding of a fair bit string to generate drawings that are far apart. The construction is in fact the one shown in Fig. 1, used for a different metric space. The upper bound uses the partition of the bounded region into a grid of size $O(c^2)$. Similarly, we can show that for drawings of ternary trees with the same embedding whose distance is measured by the sum of absolute differences of corresponding angles, we also get $2^{\Theta(n)}$ as the maximum size of a fully diverse set. Figure 2(right) shows how a bit string can be converted to a drawing so that far bit strings give far drawings.

4 Embedding Diversity

In this section we investigate the existence of large sets of embedded graphs that are diverse according to a topological measure. We show that there are superpolynomially many fully diverse sets of embedded trees and planar graphs when we use the number of changes in the rotation system as the distance measure. An *adjacent-edge swap* exchanges the position of two edges that are incident to the same vertex and adjacent in its rotation. Notice that degree-2 vertices can be omitted or ignored, since their rotation system is not changed by a swap. In this section, all graphs are assumed to be labeled.

Trees. To start, we consider *ternary trees*, i.e., trees that contain only degree 3 vertices as non-leaf vertices. Let $T = (V, E)$ be such a ternary tree with n leaves and $n - 2$ non-leaf vertices. Observe that at every non-leaf vertex there are exactly two possible cyclic orders of the incident edges. We derive a bit encoding of the possible embeddings of T as a bit string B that contains a bit for every non-leaf vertex of T. For each such vertex we associate its bit set to 0 with one of the cyclic orders, and its bit set to 1 with the other cyclic order.

Lemma 4 (⋆). *Let T be a labeled ternary tree with n leaves and B_1 and B_2 two bit encodings of embeddings of T, such that B_1 and B_2 are fair bit strings and far from each other, then they correspond to embeddings of T that are $\Omega(n)$ adjacent-edge swaps apart.*

Applying the analysis from Sect. 2 we obtain the following.

Lemma 5 **($\star$).** *For a labeled ternary tree with n leaves, a fully diverse set of embeddings may have size $2^{\Omega(n)}$ if distance is the number of adjacent-edge swaps.*

Next, we consider labeled star graphs. Let $S = (V, E)$ be a labeled star with central vertex $u \in V$ and leaves $v_1, \ldots, v_n \in V$ incident to edges $e_1, \ldots, e_n \in E$ for some even $n \in \mathbb{N}$. We define a cyclic bit string B describing orders of the edges incident to u as follows. Consider the edges $e_1, \ldots, e_n$ around u, ordered by their indices. For each antipodal pair of edges e_i, e_j in S (where $j = i + \frac{n}{2}$), we add one bit b_i to B and let it be 1 if e_i, e_j have exchanged their positions in the cyclic order and 0 if not; see Fig. 1. Clearly, B has length $\frac{n}{2}$; recall that two cyclic bit strings of length $\frac{n}{2}$ are fair if they contain at least $\frac{n}{4}$ zeros and at least $\frac{n}{4}$ ones, and they are far if they differ in at least $\frac{n}{16}$ positions.

Lemma 6 **($\star$).** *Let S be a labeled star graph with n leaves and B_1 and B_2 two bit encodings of embeddings of S, such that B_1 and B_2 are fair cyclic bit strings and far from each other, then they correspond to embeddings of S that are $\Omega(n^2)$ adjacent-edge swaps apart.*

Lemma 7 **($\star$).** *For a labeled star graph with n leaves, a fully diverse set of embeddings may have size $2^{\Omega(n)}$ if distance is the number of adjacent-edge swaps.*

It remains to combine the two previous cases to handle any tree $T = (V, E)$ that does not contain degree 2 vertices. This is non-trivial, and in fact, for some trees, we no longer have a fully diverse set of embeddings of exponential size. First, observe that the maximum distance between two embeddings of a tree whose internal nodes that have degrees $d_1, \ldots, d_k$ is proportional to $\sum_{i=1}^{k} d_i^2$.

Lemma 8 **($\star$).** *For any labeled tree with n leaves, there exists a fully diverse set of embeddings of size $2^{\Omega(\sqrt{n})}$.*

Proof Sketch. Assume that a labeled tree T is given whose internal vertices $v_1, \ldots, v_k$ are sorted by degrees $d_1 \geq d_2 \geq \cdots \geq d_k$. Let j be the smallest value such that $\sum_{i=1}^{j} d_i^2 \geq \frac{1}{2} \sum_{i=1}^{k} d_i^2$. We distinguish two cases, $d_j \geq \sqrt{n}$ and $d_j < \sqrt{n}$. In the former case, we only use the vertices $v_1, \ldots, v_j$ to make fully diverse sets. Each such vertex already admits a fully diverse set of size $2^{\Omega(d_i)}$ by Lemma 7. This allows us to just combine the embeddings and choose embeddings for the remaining vertices at random.

If $d_j < \sqrt{n}$ we use the vertices $v_j, \ldots, v_k$. We group them into sets $V_1, \ldots, V_z$ with $z = \Theta(\sqrt{n})$ such that for each V_h, $h = 1, \ldots z$, the sum of its squared degrees is in $\Theta\left(\sum_{i=1}^{k} d_i^2 / \sqrt{n}\right)$. We then fix two far embeddings for each group V_h. Using these two embeddings to encode a bit string we then get a fully diverse set of sufficient size using Lemma 3 essentially in the same manner as for Lemma 5. $\square$

To prove that no better bound exists that applies to all trees, consider a tree with n leaves, one vertex v with degree $\sqrt{n} \log n$, and all other internal vertices with degree 3. The maximum distance between two embeddings is determined by vertex v only: it is $\Theta(n \log^2 n)$. The linearly many vertices of degree 3 require only

$O(n)$ adjacent-edge swaps, so they play no role in obtaining a fully diverse set. Considering v and its neighbors as a star graph then implies that the maximum size of a fully diverse set is $2^{O(\sqrt{n}\log n)}$. Intuitively, we have just $O(\sqrt{n}\log n)$ bits in an encoding that are effective to realize a fully diverse set. We can give v degree $\sqrt{n}\log\log n$ or even smaller for a slightly better bound.

Theorem 4. *For any labeled tree with n leaves, there is a fully diverse set of embeddings of size $2^{\Omega(\sqrt{n})}$, and there exists a tree whose size of a fully diverse set of embeddings is $2^{O^*(\sqrt{n})}$, where $O^*(\sqrt{n})$ denotes $O(f(n))$ for any function $f(n)$ that is asymptotically larger than $\sqrt{n}$.*

Suppose we consider a different metric, namely the number of edge *relocations* for embedded trees. A relocation on the cyclic order around a vertex places one of its edges anywhere else in the order in a single step. For ternary trees this is equivalent to an adjacent-edge swap, but for a star graph, the maximum distance between any two embeddings of stars is $\Theta(n)$ instead of $\Theta(n^2)$.

Lemma 9 ($\star$). *For a labeled star graph with n leaves, a fully diverse set of embeddings may have size $2^{\Omega(n\log n)}$ if distance is measured by edge relocations.*

Planar Graphs. Here we give a sketch of how the results just given can be extended to planar embeddings of planar graphs. Let $G = (V, E)$ be an embedded labeled planar connected simple graph. Since swapping two adjacent edges in G does not necessarily preserve planarity we instead consider swaps of components separated by *cut-vertices* and *split pairs* [4]. A cut-vertex $u \in V$ is a vertex such that G is not connected after u is removed. Similarly, a split pair $\{u, v\} \subset V$ of G is a pair of vertices such that G is not connected after u and v are both removed from G. The *incident components* of a cut-vertex or a split pair are the connected components obtained after removing this cut-vertex or split pair.

We consider the rotation of the incident components around a cut-vertex or split pair, and so-called *adjacent-component swaps* between them. To ensure that every possible embedding can be reached, we allow the operation of mirroring a triconnected component at no cost. Each cut-vertex or split pair can be treated as the central vertex of a star and its incident components as the leaves. Swapping the order of two leaves corresponds one-to-one to swapping the order of two of its incident components. To ensure this we first resolve nesting components around cut-vertices. Then, it suffices to only consider rotations around cut-vertices and split pairs in which the respective incident components appear one after another. This allows us to derive analogous versions of Lemmas 4 and 6 which in turn enables us to argue in the same fashion as for Theorem 4 to obtain the lower bound in the following theorem.

Moreover, the upper bound of Theorem 4 translates immediately since trees are planar graphs and adjacent-edge swaps in trees are equivalent to swapping the incident components around a cut-vertex.

Theorem 5. *For any labeled planar graph $G = (V, E)$ with n_c cut-vertices and n_p split pairs each with at least 3 incident connected components, a fully diverse* *set of planar embeddings may have size* $2^{\Omega(\sqrt{n_c+n_p})}$ *and there exists a planar graph whose size of a fully diverse set of embeddings is* $2^{O^*(\sqrt{n_c+n_p})}$ *if distance is measured by adjacent-component swaps.*

5 Abstract Graphs

In this section we consider the diversity of abstract graphs of some given graph class and a distance based on edits. Throughout, we require that if a graph is in graph class $\mathcal{G}$, then after applying an operation to it the resulting graph is still in $\mathcal{G}$. We consider trees, planar graphs, and general graphs, and discuss diversity for the labeled and unlabeled cases. As most of the ideas are the same as the ones used earlier, we keep the description short. The results are given in Table 3.

Trees. For trees, we use the following edit operation: Take a leaf, unattach it from its neighbor, and attach it to a different vertex. We consider the edit distance measure: the distance between two trees of n vertices is the number of leaf reattachments needed to convert one tree into the other. Note that any two trees with n vertices have a finite distance, since every tree can easily be turned into the star graph (in the labeled case, a specific node must become the central vertex to use a star as a canonical tree). The maximum distance is $\Theta(n)$, since we need at most $n-2$ edits to convert any tree into a star.

We start with the labeled case. To construct a large size fully diverse set of labeled trees, we can restrict ourselves to paths. A path essentially encodes a permutation of its labels and reverse permutations are identified. We can use essentially the same proof ideas as for the case of labeled stars and their embedding under swaps (where cyclic shifts were identified). The upper bound is trivial, and we obtain $2^{\Theta(n \log n)}$ for the maximum size of a fully diverse set.

Next we switch to unlabeled trees. The situation is quite different, because there is only one path now, and in fact, it is known that there are only $2^{O(n)}$ different unlabeled trees [18]. To show an exponential lower bound for unlabeled trees, we start out with a path of $2+n/4$ vertices. We attach either one or two leaves to the middle $n/4$ vertices, encoding a 0 or 1 in a bit string. We attach all remaining vertices equally as paths to the ends of the initial path. These two tails have length at least $n/16$, ensuring that we need $n/16$ operations to operate on the bit string in unwanted ways. Using our knowledge on the full diversity of bit strings, we obtain $2^{\Theta(n)}$ as the bound.

General Graphs. We consider general graphs of n vertices with edge insertion or deletion as the elementary operation. The edit distance is the distance between two graphs. We again distinguish in the labeled and unlabeled cases. The two graphs furthest apart are the empty graph and the complete graph in both cases.

We start with the labeled case. Every labeled edge can be seen as a bit in a bit string, where absence encodes 0 and presence 1. We immediately get a bound of $2^{\Theta(n^2)}$ by Sect. 2. For the unlabeled case, we observe that there are at least $2^{n(n-1)/2}/n!$ graphs, since we can assign labels in at most $n!$ ways. This is still $2^{\Omega(n^2)}$. The upper bound follows from the labeled case.

Planar Graphs. For planar graphs we use the same edit operation as for general graphs. Since we can always first remove edges and then insert them, any edit sequence can be turned into an edit sequence that stays within the class of planar graphs. An upper bound on the number of operations needed is clearly $6n - 12$.

For labeled planar graphs, we obtain a lower bound by analyzing how many labeled planar graphs are within $(6n - 12)/c$ edits from a given graph. This number is certainly bounded by $(n^2)^{(6n-12)/c} = n^{O(n/c)}$. At the same time, the number of labeled paths is already $n!/2$. By choosing c sufficiently large, we obtain $2^{\Theta(n \log n)}$ as the maximum size of a fully diverse set.

The most intriguing case turns out to be unlabeled planar graphs. It is known that the number of unlabeled planar graphs is bounded by $2^{O(n)}$ [6], which is obviously also an upper bound on the size of a fully diverse set.

For a lower bound, the idea is to consider graphs that are unions of stars. We can connect them into one connected graph if needed, but the argument is cleanest for these unconnected graphs. We consider only stars with 2^i vertices, $0 \le i \le \log n - \log\log n$. Suppose we have $n/(2^i \log n)$ stars of size 2^i, then it takes $n/(2 \log n)$ edge insertions (and a number of edge deletions) to convert this into $n/(2^{i+1} \log n)$ stars of size 2^{i+1}. Converting in the other direction also takes at least $n/(2 \log n)$ operations.

In the fully diverse set we construct, we choose stars with either 2^i or 2^{i+1} vertices, for $i = 0, 2, 4, \ldots, (\log n) - (\log\log n) - 1$, the latter value rounded down to the nearest even number, henceforth denoted by m. We then have roughly $m/2$ different sizes in any single set, out of the twice as many sizes used in the whole construction. Notice that a set indeed has size n. We can see the choice between stars of size 2^i and 2^{i+1} as an encoding of a bit, and hence we have a bit string of length roughly $m/2$. We choose a fully diverse set of bit strings, which implies the choice of stars in a graph in the set. By Lemma 2, a fully diverse set of fair bit strings of this length has maximum size $2^{\Omega(m/2)}$, which is $n^{\Omega(1)}$.

6 Conclusions and Open Problems

We introduced the concept of a fully diverse set of objects, like polygons and graphs, in a metric space, by relating the inter-distance between any two objects in that set to the maximum distance possible. We then studied a number of distance measures, both geometric and combinatorial, and proved bounds on the maximum size of fully diverse sets. There are two cases where the lower and upper bounds do not match, giving rise to the two main open problems of this paper. We also sketched a simple randomized algorithm to generate fully diverse sets of a certain type of objects.

As our full diversity definition can be applied to any class of objects in a metric space provided the maximum distance is bounded, there are many other cases to be explored. For example, 2-dimensional distributions with the Wasserstein distance, or graphs with different edit distances than the ones used in this paper. Furthermore, a definition of full diversity that does not require the metric space to be bounded is worth examination.

References

1. Alt, H., Godau, M.: Computing the Fréchet distance between two polygonal curves. Int. J. Comput. Geom. Appl. **5**, 75–91 (1995). https://doi.org/10.1142/S0218195995000064
2. Bach, B., Spritzer, A., Lutton, E., Fekete, J.-D.: Interactive random graph generation with evolutionary algorithms. In: Didimo, W., Patrignani, M. (eds.) GD 2012. LNCS, vol. 7704, pp. 541–552. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36763-2_48
3. Baste, J., et al.: Diversity of solutions: an exploration through the lens of fixed-parameter tractability theory. In: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI 2020), pp. 1119–1125 (2020). https://doi.org/10.24963/ijcai.2020/156
4. Battista, G.D., Tamassia, R.: On-line planarity testing. SIAM J. Comput. **25**(5), 956–997 (1996). https://doi.org/10.1137/S0097539794280736
5. Biedl, T., Marks, J., Ryall, K., Whitesides, S.: Graph multidrawing: finding nice drawings without defining nice. In: Whitesides, S.H. (ed.) GD 1998. LNCS, vol. 1547, pp. 347–355. Springer, Heidelberg (1998). https://doi.org/10.1007/3-540-37623-2_26
6. Bonichon, N., Gavoille, C., Hanusse, N., Poulalhon, D., Schaeffer, G.: Planar graphs, via well-orderly maps and trees. Graphs Combin. **22**(2), 185–202 (2006). https://doi.org/10.1007/s00373-006-0647-2
7. Bridgeman, S., Tamassia, R.: Difference metrics for interactive orthogonal graph drawing algorithms. In: Whitesides, S.H. (ed.) GD 1998. LNCS, vol. 1547, pp. 57–71. Springer, Heidelberg (1998). https://doi.org/10.1007/3-540-37623-2_5
8. Cai, J.: Counting embeddings of planar graphs using DFS trees. SIAM J. Discret. Math. **6**(3), 335–352 (1993). https://doi.org/10.1137/0406027
9. Drosou, M., Jagadish, H.V., Pitoura, E., Stoyanovich, J.: Diversity in big data: a review. Big Data **5**(2), 73–84 (2017). https://doi.org/10.1089/big.2016.0054
10. Gao, X., Xiao, B., Tao, D., Li, X.: A survey of graph edit distance. Pattern Anal. Appl. **13**(1), 113–129 (2010)
11. Hebrard, E., Hnich, B., O'Sullivan, B., Walsh, T.: Finding diverse and similar solutions in constraint programming. In: Proceedings of 20th National Conference on Artificial Intelligence (AAAI 2005), pp. 372–377 (2005)
12. Hill, M.O.: Diversity and evenness: a unifying notation and its consequences. Ecology **54**(2), 427–432 (1973). https://doi.org/10.2307/1934352
13. Huemer, C., Pilz, A., Silveira, R.I.: A new lower bound on the maximum number of plane graphs using production matrices. Comput. Geom. **84**, 36–49 (2019). https://doi.org/10.1016/j.comgeo.2019.07.005
14. Ingmar, L., de la Banda, M.G., Stuckey, P.J., Tack, G.: Modelling diversity of solutions. In: Proceedings of 34th AAAI Conference on Artificial Intelligence (AAAI 2020), pp. 1528–1535 (2020)
15. van Kreveld, M., Speckmann, B., Urhausen, J.: Diverse partitions of colored points. In: Lubiw, A., Salavatipour, M. (eds.) WADS 2021. LNCS, vol. 12808, pp. 641–654. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-83508-8_46
16. Kulesza, A., Taskar, B.: Determinantal point processes for machine learning. Found. Trends Mach. Learn. **5**(2–3), 123–286 (2012). https://doi.org/10.1561/2200000044
17. Kunaver, M., Pozrl, T.: Diversity in recommender systems - a survey. Knowl. Based Syst. **123**, 154–162 (2017). https://doi.org/10.1016/j.knosys.2017.02.009

18. Otter, R.: The number of trees. Ann. Math. **49**(3), 583–599 (1948)
19. Robbins, H.: A remark on Stirling's formula. Am. Math. Mon. **62**(1), 26–29 (1955)
20. Stallmann, M.F.M.: On counting planar embeddings. Discret. Math. **122**(1–3), 385–392 (1993). https://doi.org/10.1016/0012-365X(93)90316-L
21. Tuomisto, H.: A consistent terminology for quantifying species diversity? Yes, it does exist. Oecologia **164**(4), 853–860 (2010). https://doi.org/10.1007/s00442-010-1812-0
22. Ulrich, T., Bader, J., Thiele, L.: Defining and optimizing indicator-based diversity measures in multiobjective search. In: Schaefer, R., Cotta, C., Kołodziej, J., Rudolph, G. (eds.) PPSN 2010. LNCS, vol. 6238, pp. 707–717. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15844-5_71
23. Wineberg, M., Oppacher, F.: The underlying similarity of diversity measures used in evolutionary computation. In: Cantú-Paz, E., et al. (eds.) GECCO 2003. LNCS, vol. 2724, pp. 1493–1504. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45110-2_21