Discrepancy and Sparsity

Patrice Ossona de Mendez — CAMS, CNRS/EHESS, France
Mario Grobler — University of Bremen, Bremen, Germany
Yiting Jiang — Université de Paris, France and Zhejiang Normal University, China
Sebastian Siebertz — University of Bremen, Bremen, Germany
Alexandre Vigny — University of Bremen, Bremen, Germany

Abstract

We study connections between combinatorial discrepancy and graph degeneracy. We prove that the maximum discrepancy over all subgraphs H of a graph G of the neighborhood set system of H is in $\Omega(\log \deg(G))$ and $O(\deg(G))$, where $\deg(G)$ denotes the degeneracy of G. We also relate weak coloring numbers and discrepancy of graph powers.

We prove that a monotone class of graphs has bounded expansion if and only if all the set systems FO-definable in this class have bounded hereditary discrepancy. We also give a characterization of nowhere dense classes in terms of discrepancy.

We derive a corollary on the discrepancy of neighborhood set systems of edge colored graphs, a polynomial-time algorithm to compute ε-approximations of size $O(1/\varepsilon)$ for set systems definable in bounded expansion classes, an application to clique coloring, and the non-existence of a quantifier elimination scheme for nowhere dense classes.

1 Introduction

Discrepancy theory emerged from the study of the irregularities of statistical distributions and number sequences. It developed and became a central tool in computational geometry. Two decades ago, Matoušek initiated the study of *combinatorial discrepancy*, which became a significant subject in its own right. The combinatorial discrepancy measures the inevitable irregularities of set systems and the inherent difficulty to approximate them.

Discrepancy theory offers powerful tools and techniques with many applications in computational geometry, probabilistic algorithms, derandomization, communication complexity, searching, machine learning, pseudorandomness, optimization, computer graphics, and more. Central notions in this theory are also the well known notions of VC-dimension, ε-nets and ε-approximations, the latter corresponding to the expected properties of a pseudorandom set.

A structural theory of classes of sparse graphs emerged recently, which is based on the study of densities of shallow minors, generalized coloring numbers, and constrained orientations. In this setting, two central notions are those of *classes with bounded expansion*, which generalize classes excluding a topological minor, and *nowhere dense classes*, which generalize classes locally excluding a topological minor. These classes have strong algorithmic and structural properties. In particular, in a nowhere dense class it can be checked in almost linear time whether a first-order formula is satisfied in a given graph from the class. This last example is only one among others that witness a strong connection between sparsity theory and first-order logic.

2 Motivating examples

We illustrate our results with a few motivating examples. All these results mentioned here are obtained as special cases of our general theorems on combinatorial discrepancy of set systems definable in sparse graph classes.

Problem 1. *Assume that a graph G has the property that for every red/blue coloring of the edges of G there exists a partition (A, B) of the vertex set of G such that the number of red (resp. blue) neighbors in A and B of any vertex differ by at most 1. Does G contain a vertex with small degree?*

It follows from our results that such a graph G is 1303-degenerate.

Problem 2. *Given a planar graph G, find a small subset F of edges such that, for every pair u, v of distinct vertices of G, the probability that an edge of G belongs to a uv-path of length at most 100 differs from the probability that an edge in F belongs to a uv-path of length at most 100 by at most ε.*

We prove that a set F of edges of size $\mathcal{O}(1/\varepsilon)$ with the prescribed properties can be constructed deterministically in polynomial time.

Problem 3. *Does there exist a constant c, such that the vertices of every map graph*[1] *G can be colored red or blue, in such a way that the difference between the number of red and blue vertices in every maximal clique of G is at most c?*

Although there are quite a few reasons to believe that such a constant would not exist (it is not even possible in general to color the vertices of a perfect graph red and blue in such a way that no maximal clique is monochromatic) we prove that such a constant c exists for map graphs.

Last, we also consider the following (seemingly completely unrelated) problem from sparse finite model theory. It is known that every class of finite graphs with bounded expansion has a quantifier elimination scheme involving unary relations and functions. As it is known that the fixed-parameter tractability of first-order model-checking extends from bounded expansion classes to the more general nowhere dense classes, it is natural to ask whether the quantifier elimination scheme also extends. It has been conjectured that this is not the case, but no proof of this fact was known.

Problem 4. *Give an example of a nowhere dense class $\mathscr{C}$ of graphs such that there exists no expansion σ of the signature of graphs by unary relation and function symbols with the property that every first-order formula is equivalent on a σ-expansion $\mathscr{C}^+$ of $\mathscr{C}$ to a quantifier-free first-order σ-formula.*

We prove that the class $\mathscr{C}$ of 1-subdivisions of bipartite graphs whose girth exceeds the maximum degree has the above property. Precisely, there is no expansion σ of the signature of graphs by unary relation and function symbols in which the formula $\varphi(x, y)$ expressing that x and y are at distance 2 in the graph is equivalent on a σ-expansion of $\mathscr{C}$ to a quantifier-free first-order σ-formula.

[1] A *map graph* is the vertex-face incidence graph of a planar map.

3 Preliminaries: Combinatorial discrepancy and Sparse classes

Let $(U, \mathscr{S})$ be a *set system*, where $\mathscr{S}$ is a collection of subsets of the *ground set* U. When the ground set is clear from the context, we refer to the set system as $\mathscr{S}$. The *discrepancy* of a mapping $\chi\colon U \to \{-1,1\}$ on a set $S \in \mathscr{S}$ is $\operatorname{disc}_\chi(S) = \left|\sum_{v\in S}\chi(v)\right|$; the *discrepancy* of χ on $\mathscr{S}$ is the maximum of $\operatorname{disc}_\chi(S)$ over all $S \in \mathscr{S}$, that is, $\operatorname{disc}_\chi(\mathscr{S}) = \max_{S\in\mathscr{S}} \operatorname{disc}_\chi(S)$. The *(combinatorial) discrepancy* of $\mathscr{S}$ is the minimum discrepancy of a mapping $\chi\colon U \to \{-1,1\}$ on $\mathscr{S}$, that is,

$$\operatorname{disc}(\mathscr{S}) = \min_{\chi:U\to\{-1,1\}} \max_{S\in\mathscr{S}} \left|\sum_{v\in S}\chi(v)\right|.$$

A more robust notion is the *hereditary discrepancy* of a set system $(U, \mathscr{S})$, defined as $\operatorname{herdisc}(\mathscr{S}) = \max_{U'\subseteq U} \operatorname{disc}(\mathscr{S}|_{U'})$, where $\mathscr{S}|_{U'}$ denotes the set system $\{S\cap U' : S \in \mathscr{S}\}$. Moreover, bounding the hereditary discrepancy allows to bound the sizes of ε-nets and ε-approximations (what is not the case for discrepancy).

A graph G is *d-degenerate* if every non-empty induced subgraph of G has minimum degree at most d. The minimum integer d such that a graph G is d-degenerate is the *degeneracy* $\deg(G)$ of G. A class $\mathscr{C}$ is *degenerate* if there is an integer d such that all the graphs in $\mathscr{C}$ are d-degenerate. A class $\mathscr{C}$ of graphs is called *monotone* if it is closed under taking subgraphs and *hereditary* if it is closed under taking induced subgraphs.

For a graph G and an integer r, a *$\leq r$-subdivision* of a graph G is a graph obtained by subdividing every edge of G at most r times. A graph H is a *topological minor* of a graph G *at depth* r if a $\leq 2r$-subdivision of H is a subgraph of G. For a class $\mathscr{C}$ of graphs, we denote by $\mathscr{C} \mathbin{\widetilde{\nabla}} r$ the class of the topological minors at depth r of graphs in $\mathscr{C}$. A class $\mathscr{C}$ has *bounded expansion* if, for every integer r, the class $\mathscr{C} \mathbin{\widetilde{\nabla}} r$ is degenerate. For example, every proper minor-closed classes of graphs and every class of graphs with bounded maximum degree have bounded expansion. The class $\mathscr{C}$ is *nowhere dense* if no class $\mathscr{C} \mathbin{\widetilde{\nabla}} r$ is the class of all graphs. For example, every class with bounded expansion is nowhere dense, as well as the class of all graphs whose maximum degree is bounded by the girth, are nowhere dense.

Generalized coloring numbers have been introduced by Kierstead and Yang as a generalization of the so-called coloring number. Let G be a graph and let L be a linear ordering of $V(G)$. We say that a vertex u is *weakly d-reachable* from a vertex v if there exists in G a path P of length at most d (possibly 0) linking u and v such that u is the minimum vertex of P with respect to L, and we denote by $\operatorname{WReach}_d[G, L, v]$ the set of all vertices weakly d-reachable from v. The *weak coloring number* $\operatorname{wcol}_d(G)$ is defined as the minimum over all possible linear orderings L of $\max_{v\in V(G)} |\operatorname{WReach}_d[G, L, v]|$. Zhu proved that a class $\mathscr{C}$ of graphs has bounded expansion if and only if, for each integer d, the weak coloring numbers with rank d of the graphs in $\mathscr{C}$ are bounded. (Note that nowhere dense classes can also be characterized in terms of bounds on the weak coloring numbers.)

Structural and algorithmic properties of classes with bounded expansion and nowhere dense classes have strong links with first-order logic. In particular, the fixed-parameter linear time first-order model-checking algorithm of Dvořák, Kráľ, and Thomas for bounded expansion classes is based on a quantifier elimination scheme. This also justifies to extend the study of the discrepancy of neighborhood set systems to first-order definable set systems, that is to set systems definable using first-order logic formulas.

4 Our results.

We prove that the notions of discrepancy and degeneracy are deeply linked:

Theorem 1. *For every graph G we have*

$$\frac{\log_2(\pi \deg(G))}{4} - 2 \le \max_{H \subseteq G} \operatorname{disc}(\mathscr{S}^E(H)) < 3\deg(G),$$

where $\mathscr{S}^E(H)$ denotes the neighborhood set system of H.

We extend this result to inequalities relating weak coloring numbers and discrepancy of graph powers.

Theorem 2. *Let G be a graph and let d be a positive integer. Then*

$$\frac{\log_2(\mathrm{wcol}_{\lceil d/2 \rceil}(G))}{6(d+1)} - \frac{\log_2(d+1)}{3} - \frac{3}{2} \le \max_{d' \le d} \max_{H \subseteq G} \operatorname{herdisc}(\mathscr{S}^E(H^{d'})) < (2d\,\mathrm{wcol}_{d-1}(G) + 1)\,\mathrm{wcol}_d(G).$$

We deduce that a monotone class $\mathscr{C}$ has bounded expansion if and only if the hereditary discrepancy of $\mathscr{S}^E(G^k)$ is bounded on $\mathscr{C}$ for each positive integer k.

In order to extend these results further, we switch to a model theoretic point of view. We introduce a theory of structures with only unary relations and unary functions, which we prove has quantifier elimination and mirrors the properties of the classes of graphs with bounded expansion. From this, we deduce a characterization of bounded expansion classes.

Theorem 3. *Let $\mathscr{C}$ be a monotone class of graphs. Then the following are equivalent (where $\mathscr{S}^\varphi(G)$ denotes the set system defined by a partitioned first order formula $\varphi(\bar{x}; \bar{y})$ on a vertex-colored graph G):*

1. *the class $\mathscr{C}$ has bounded expansion;*
2. *the hereditary discrepancy of every set system $\mathscr{S}^\varphi(G)$ definable on a monadic expansion of $\mathscr{C}$ (i.e. using arbitrary vertex colorings of the graphs in $\mathscr{C}$) is bounded;*
3. *for each positive integer k, the hereditary discrepancy of $\mathscr{S}^E(G^k)$ for $G \in \mathscr{C}$ is bounded.*

Then, using the bounds on the VC-density of set systems definable in nowhere dense classes of Pilipczuk, Siebertz, and Toruńczyk, we give a characterization of nowhere dense classes in terms of discrepancy. We believe that our upper bounds on discrepancy can be improved in this case, and we propose a conjecture for the optimal bound.

We provide some corollaries on edge colored graphs, ε-approximations, clique coloring, and quantifier elimination schemes, which allows us to solve the motivating problems presented in the introduction.