DISTANCIA – Metric Graph Theory [Théorie Métrique des Graphes]

ANR Proposal, 2017

DEFI 7: Société de l'Information et de la communication

1. Proposal's context, positioning and objectives

1.1. **Proposal's context.** This proposal is concerned with graphs and metrics, both on theoretical foundations and applications. Each of the themes developed here has many applications due to the growing importance of this subject. Such applications can be found in real world networks. For example, the hub labelling problem in road networks can be directly applied to car navigation applications. Understanding key structural properties of large-scale data networks is crucial for analyzing and optimizing their performance, as well as improving their reliability and security. In prior empirical and theoretical studies researchers have mainly focused on features such as small world phenomenon, power law degree distribution, navigability, and high clustering coefficients. Although those features are interesting and important, the impact of intrinsic geometric and topological features of large-scale data networks on performance, reliability and security is of much greater importance. Recently, there has been a surge of empirical works measuring and analyzing geometric characteristics of real-world networks, namely the Gromov hyperbolicity (called also the negative curvature) of the network. It has been shown that a number of data networks, including Internet application networks, web networks, collaboration networks, social networks, and others, have small hyperbolicity. Based on the experimental observation by Narayan and Saniee, Jonckheere et al. conjectured that the property, observed in real-world networks, in which traffic between vertices tends to go through a relatively small core of the network, as if the shortest path between them is curved inwards, may be due to global curvature of the network. We proved this conjecture recently by using metric methods, namely Helly type theorems and properties of injective hulls of hyperbolic graphs.

Metric graph theory was also indispensable in solving some open questions in concurrency and learning theory in computer science and geometric group theory in mathematics. Median graphs are exactly the 1–skeletons of CAT(0) cube complexes (which have been characterized by Gromov in a local-to-global combinatorial way). They play a vital role in geometric group theory (for example, in the recent solution of the famous Virtual Haken Conjecture). Median graphs are also the domains of event structures of Winskel, one of the basic abstract models of concurrency. This correspondence is very useful in dealing with questions on event structures. For instance, we used it to disprove Rozoy-Thiagarajan and Thiagarajan's conjectures about nice labelings and regular labelings of event structures. Lopsided sets–a generalization of median graphs and particular partial cubes–seem to acquire an importance in designing sample compression schemes for concept classes of bounded VC-dimension in computational learning theory. We showed that they do not admit corner peelings, and this provided counterexamples to some influential published work. Analogously to median graphs, bridged graphs have been characterized as the 1–skeletons of systolic complexes. Systolic complexes satisfy many global properties of CAT(0) spaces (contractibility, fixed point property) and were suggested as a variant of simplicial complexes of combinatorial nonpositive curvature. A remarkable appearance of modular graphs occurred in classifying the complexity of the so-called 0–extension problem, a combinatorial optimization problem generalizing the minimum cut problem and having applications in computer vision. A dichotomy characterization of tractability of the 0–extension problem was provided by Hirai: if G is an orientable modular graph, then the 0–extension problem on G is polynomial, otherwise it is NP-hard.

Many classical algorithmic problems concern distances: shortest path, center and diameter, Voronoi diagrams, TSP, clustering, etc. Algorithmic and combinatorial problems related to distances also occur in data analysis. Low-distortion embeddings into ℓ_1-spaces (theorem of Bourgain and its algorithmical use by Linial et al.) were the founding tools in metric methods. Recently, several approximation algorithms for NP-hard problems were designed using metric methods. Other important algorithmic graph problems related to distances concern the construction of sparse subgraphs approximating inter-node distances and the converse, augmentation problems with distance constraints. Finally, in the distributed setting, an important problem is that of designing compact data structures allowing very fast computation of inter-node distances or routing along shortest or almost shortest paths.

Besides computer science and mathematics, applications of structures involving distances can be found in archeology, computational biology, statistics, data analysis, etc. The problem of characterizing isometric subgraphs of hypercubes has its origin in communication theory and linguistics. In the search for a method for chronologically ordering archaeological deposits, the archeologist Robinson introduced in 1951 a distance measure which now bears his name (Robinson dissimilarity) and is the standard distance model for seriation. To take into account the recombination effect in genetic data, the mathematicians Bandelt and Dress developed in 1991 the theory of canonical decompositions of finite metric spaces. Together with geneticists, Bandelt successfully used it over the years to reconstruct phylogenies, in the evolutional analysis of mtDNA data in human genetics. One important step in their method is to build a reduced median network that spans the data but still contains all most parsimonious trees. As mentioned above, the median graphs occurring there constitute a central notion in metric graph theory.

With this proposal, we aim to participate at the elaboration of this new domain of Metric Graph Theory, which requires experts and knowledge in combinatorics (graphs, matroids), geometry, and algorithms. This expertise is distributed over the members of the consortium and a part of the success of our project it will be to share these knowledges among all the members of the consortium. This way we will create a strong group in France on graphs and metrics.

1.2. **Scientific objectives.** The central subject of *metric graph theory* (MGT) is the investigation and characterization of graph classes whose metric satisfies the main metric and convexity properties of classical metric geometries like $\mathbb{R}^n$ with l_2, l_1, or l_∞–metric, hyperbolic spaces, hypercubes, trees. Such central properties are convexity of balls, Helly property for balls, geometry of geodesic or metric triangles, isometric and low-distortion embeddings, retractions, treelikeness, uniqueness or existence of medians, etc. The main classes of graphs central to MGT are median graphs, Helly graphs, partial cubes and ℓ_1–graphs, lopsided sets, bridged graphs, Gromov hyperbolic graphs, modular and weakly modular graphs. Other classes arise from combinatorics and geometry: basis graphs of matroids, tope graphs of oriented matroids, dual polar spaces. Later, it turned out that many classes of graphs from MGT give rise to important cubical and simplicial complexes, like CAT(0) cube complexes or systolic complexes.

In the project, we will try to make substantial contributions to combinatorial, algorithmic, geometric, and topological structure of graph classes occurring in metric graph theory and of their associated cell complexes by developing and exploiting theory for common structures and structural similarities that occur in problems from diverse areas. We hope to use this knowledge for solving questions in computer science (concurrency and machine learning), discrete mathematics, and combinatorial optimization. We would also like to develop algorithms for distance problems related to data analysis, network analysis, distributed computing, and seriation.

1.3. **Main themes.** The project focuses on two main subjects "Structure in metric graph theory" and "Algorithms in metric graph theory", consisting of strongly interconnected research themes:

- S1. Local-to-global characterizations
- S2. Median graphs and event structures
- S3. Lopsided sets and sample compression
- S4. Matroidal structures
- S5. Isometric and low distortion embeddings
- S6. Packing and covering with balls, identifying codes, and χ-boundedness
- A1. Algorithmic aspects of hyperbolic graphs
- A2. Algorithms for graph classes from MGT
- A3. Finite metric spaces: approximation and realization
- A4. Seriation and classification.

For each theme we will formulate the main problems on which we would like to work and show their relevance, the state-of-art and our previous contributions. We summarize the description of each theme with the formulation of main objectives. The missing definitions to the first subject can be found in the survey [56].

1.4. **Structure in Metric Graph Theory.**

Theme S1: *Local-to-global characterizations.* The local-to-global approach originated in Riemannian geometry (Cartan-Hadamard theorem) and was extended by M. Gromov to nonpositively curved metric spaces. In the particular setting of CAT(0) spaces, Gromov's theorem asserts that a complete simply connected geodesic metric space is CAT(0) iff it is locally CAT(0). Equivalently, it asserts that the universal cover of a locally CAT(0) space is CAT(0). In the case of CAT(0) cube complexes, Gromov specified this result in a purely combinatorial way: these are simply connected cube complexes in which the links of vertices are flag (clique) complexes.

A *local-to-global approach* in MGT consists in associating to a graph G a cell complex $X(G)$ (triangle, square, triangle-square complex) so that G belongs to a class of graphs $\mathcal{G}$ iff $X(G)$ is simply connected and satisfies a local combinatorial property. This way, we showed in [73] that median graphs are exactly the graphs whose square complex is simply connected and satisfies the cube condition (any three squares pairwise intersecting in edges live in a cube) and (using Gromov's characterization) are exactly the 1–skeletons of CAT(0) cube complexes. Analogously to median graphs, bridged graphs have been characterized in [73] as the 1–skeletons of simply connected simplicial flag complexes in which the links of vertices do not contain induced 4– and 5–cycles. Those complexes were rediscovered in [98, 106, 137] and dubbed *systolic complexes*. Systolic complexes satisfy many global properties of CAT(0) spaces (contractibility, fixed point property) and were suggested in [106] as a variant of simplicial complexes of combinatorial nonpositive curvature. This research culminated recently in [22] and [19] with a local-to-global characterization of basis graphs of matroids (solving a conjecture by Maurer [116]) and weakly modular graphs, Helly, and dual polar graphs. For example, we showed that basis graphs of matroids are exactly the graphs whose triangle-square complexes are simply connected and which are locally basis graphs (balls of radius 3 are like balls in basis graphs). Weakly modular graphs have been characterized in the same vein: these are exactly the graphs whose triangle-square complexes are simply connected and which are locally weakly modular (balls of radius 3 are like balls in weakly modular graphs). We also showed that Helly graphs are exactly the clique-Helly graphs with simply connected triangle complexes. Analogously, our characterizations of dual polar graphs from one hand simplifies that of Cameron [70] and shows that dual polar graphs represent a natural subclass of weakly modular graphs, and, on the other hand, provides an alternative proof of a difficult result of Brouwer and Cohen [88].

The results of [19] provide a deep and quite complete theory of weakly modular graphs and their associated cell complexes. Nevertheless, there are still many open question about them, some of them raised in [19]. One of them will be to *investigate the properties of groups acting on Helly graphs. Are they biautomatic? Do they admit a geodesic bicombing?* It will be also important to extend the results of [19] to larger classes of graphs, for example to meshed graphs (generalizing weakly modular graphs and basis graphs of matroids) or to graphs with convex balls (generalizing bridged graphs).

The *graphs with convex balls* have been characterized in [130, 92] via forbidden isometric cycles. For example, in the case of graphs G with convex balls, all isometric cycles have length 3 or 5. Thus, if we define a 2-dimensional cell complex $X(G)$ with a cell for each triangle and each induced pentagon of G, then $X(G)$ is simply connected. Thus one can ask *under which local conditions the 1-skeleton of a simply connected triangle-pentagonal complex has convex balls?*

A graph $G = (V, E)$ is called *meshed* if for any three vertices u, v, w with $d(v, w) = 2$, there exists a common neighbor x of v and w such that $2d(u, x) \leq d(u, v) + d(u, w)$. Meshed graphs are thus characterized by some (weak) convexity property of the radius functions $d(\cdot, u)$ for $u \in V$. This condition ensures that all balls centered at cliques induce isometric subgraphs of G. Analogously to the theory of weakly modular graphs, we would like *to develop a metric and local-to-global theory for meshed graphs. What meshed graphs lead to CAT(0) cell complexes or to cell complexes with combinatorial nonpositive curvature?* This is a much harder task because meshed graphs constitute a much larger class with a different behavior. We showed in [19] that triangle-square complexes of meshed graphs cannot be characterized in a local-to-global way. Nevertheless, we hope that for some classes of meshed graphs and under some stronger conditions such results are possible. It will be interesting to see what classes of graphs occurring in point-line geometries (as dual polar graphs), in complexes arising from topology of surfaces (curve, arc, and Kakimizu complexes), or in combinatorics (like basis graphs of matroids) are meshed or satisfy similar properties. In our opinion, one candidate is the class of constant-parity jump systems of Bouchet and Cunningham [65], which found numerous applications in discrete convexity [119]. A *jump system* is a set of integer points with an exchange property generalizing that of bases of matroids. Let E be a finite set. For $X = (X_e), Y = (Y_e) \in \mathbb{Z}^E$, define $[X, Y] = \{Z \in \mathbb{Z}^E : \min\{X_e, Y_e\} \leq Z_e \leq \max\{X_e, Y_e\}, \forall e \in E\}$. In other words, $[X, Y]$ is the interval between the integer vectors X and Y in the grid $\mathbb{Z}^E$. For $X, Y \in \mathbb{Z}^E$, a unit vector $s \in \mathbb{Z}^E$ is called an *(X, Y)-increment* if $s = (0, \ldots, 0, s_e, 0 \ldots, 0)$ where $s_e \in \{-1, +1\}$ and $X + s \in [X, Y]$. Then $\mathcal{J} \subseteq \mathbb{Z}^E$ is a *constant-sum jump system* [65, 119] if for any $X, Y \in \mathcal{J}$ and for any (X, Y)-increment s, there exists an $(X + s, Y)$-increment t such that $X + s + t \in \mathcal{J}$. Analogously to characterizations of basis graphs of matroids and even Δ-matroids [116, 74, 22], we would like to investigate the metric and local-to-global structure of jump systems. *Are they meshed? How to characterize them metrically or in a local-to-global way?*

Objectives: *Local-to-global characterization of graphs with convex balls and meshed graphs; metric and local-to-global characterizations of graphs of jump systems; Helly groups.*

Related themes: S2,S3,S4,S5,A2

Theme S2: *Median graphs and event structures.* Event structures introduced by Nielsen, Plotkin, and Winskel [122, 136, 135] are a widely recognized abstract model of concurrent computation. Thiagarajan [132] and Rozoy and Thiagarajan [125] formulated two interesting and important combinatorial conjectures about event structures and their equivalence with other models of concurrency. Using the equivalence between domains of event structures, median graphs, and CAT(0) cube complexes we disproved both these conjectures. In the current project, we would like to prove that both conjectures hold for some important classes of event structures.

An *event structure* is a triple $\mathcal{E} = (E, \leq, \smile)$, where E is a set of *events*, $\leq$ is a partial order on E, called *causal dependency*, and $\smile$ is a binary relation on E called *conflict*. For all e, e', e'', if $e \smile e'$ and $e' \leq e''$, then $e \smile e''$ (conflict $e \smile e''$ is inherited from conflict $e \smile e'$). A conflict $e \smile e'$ is *minimal* if it is not inherited from another conflict. The events which are not is causal dependency or in conflict are called *concurrent*. The events e and e' are *independent* if they are either concurrent or in minimal conflict. An *independent set* is a set of pairwise independent events. The *degree* of $\mathcal{E}$ is the maximum size of an independent set. A *labeling* is a map $\lambda : E \to \Lambda$, where Λ is some alphabet, and λ is a *nice labeling* if $\lambda(e) \neq \lambda(e')$ whenever e and e' are independent. A *configuration* is a downward closed conflict-free subset of events. The *domain* $\mathcal{D}(\mathcal{E})$ is the set of configurations ordered by inclusion. The *filter* of a configuration C is the set of all configurations containing C. An event structure $\mathcal{E}$ is *regular* if $\mathcal{D}(\mathcal{E})$ contains only a finite number of isomorphic classes of filters. A *regular labeled event structure* is an event structure $\mathcal{E}$ which admits a finite nice labeling λ such that the domain $\mathcal{D}(\mathcal{E})$ has only a finite number of classes of isomorphisms of colored filters. Rozoy and Thiagarajan [125] conjectured that *any event structure with finite degree has a finite nice labeling.* Thiagarajan's [132] conjecture asserts that *any regular event structure admits a regular nice labeling*, or, equivalently, that *regular event structures correspond exactly to finite 1-safe Petri nets* (a related conjecture of Badouel et al. (1999) [54] asserts that *the domains of regular event structures are recognizable*).

The domain $\mathcal{D}(\mathcal{E})$ of an event structure $\mathcal{E}$ naturally gives rise to a median graph G and an accompanying CAT(0) cube complex X. Indeed, let G be the graph whose vertices are the configurations with C and C' joined by an edge iff $C = C' \cup \{e\}$ for some event e. It was shown in [59] that G is a median graph, and thus its cube complex X is CAT(0) [73]. The hyperplanes of X correspond to the events in E. Conversely, each CAT(0) cube complex X and any vertex $v \in X$, gives rise to an event structure whose events are the hyperplanes of X. Hyperplanes H and H' define concurrent events iff they cross, and $H \prec H'$ iff H separates H' from v. The events defined by H and H' are in conflict iff H and H' do not cross and neither separates the other from v. Therefore, rephrasing questions about event structures in the language of median graphs and CAT(0) cube complexes can lead to their solution.

We used this correspondence in [24] and [18] to disprove both Rozoy and Thiagarajan's and Thiagarajan's and Badouel et al. conjectures. Haglund, Niblo, Sageev, and Chepoi conjectured that any CAT(0) cube complex of bounded degree can be isometrically embedded into a finite number of trees. In [32], we adapted the counterexample from [24] to also disprove the embedding conjecture by Haglund et al. and we solved in the positive nice labeling and embedding conjectures for two-dimensional CAT(0) cube complexes.

Even if all previous conjectures turned out to be false, it would be important to exhibit classes of event structures for which these fundamental conjectures are true. Badouel et al. [54] showed (with a difficult proof) that Thiagarajan's conjecture hold for context-free domains. Context-free graphs are particular Gromov-hyperbolic graphs. An interesting challenge would be to *establish Thiagarajan's conjecture for Gromov-hyperbolic domains.* A positive answer to this question would show that *deciding if a regular event domain admits a regular nice labeling is undecidable* [18]. Haglund [99] proved that Haglund et al. embedding conjecture is true for hyperbolic CAT(0) cube complexes. Adapting his proof, one can also show that Rozoy and Thiagarajan's conjecture is also true in this case. Thiagarajan's conjecture was positively solved by Nielsen and Thiagarajan [123] for conflict-free event structures. A possible way to extend their result is to *consider this conjecture for confusion-free domains* introduced by Nielsen et al. [122]. From geometric and combinatorial points of view, context-free and conflict-free domains have quite different structural properties and give rise to different kinds of CAT(0) cube complexes. For instance, in context-free domains (and more generally, hyperbolic domains), isometric square-grids are bounded while conflict-free domains can contain infinite square-grids.

Objectives: *Thiagarajan's conjecture for hyperbolic and confusion-free domains.*

Related themes: S1,S3,S4,A1

Theme S3: *Lopsided sets and sample compression.* A long-standing open problem in computational learning theory is a problem by Littlestone and Warmuth (1986) asking whether for any concept class $\mathcal{C} \subseteq \{0,1\}^U$ of VC-dimension d there always exists a compression scheme whose size is of order of d. In these schemes the input sample is compressed to a small subsample that encodes a hypothesis consistent with the input sample. The way to design sample compression schemes is to construct for each concept class $\mathcal{C}$ a *representation map*, i.e., an injective map $r : \mathcal{C} \to \{0,1\}^U$ such that $|r(C)| = O(d)$ and for any $C, C' \in \mathcal{C}$ the following *non-clashing condition* holds: $C \cap (r(C) \cup r(C')) \neq C' \cap (r(C) \cup r(C'))$. At first glance, this has nothing to do with the current project. However, as we will show below, a crucial particular case for which this problem is open is that of lopsided sets (a class of partial cubes generalizing median graphs). In this case, the problem can be restated in a truly combinatorial way and its solution can open perspectives for solving the general case.

Recall that a subset $Y \subseteq U$ is *shattered* by $\mathcal{C} \subseteq \{0,1\}^X$ if $\{Y \cap C : C \in \mathcal{C}\} = \{0,1\}^Y$. The VC-*dimension* d of $\mathcal{C}$ is the largest size of a shattered set. The famous *Sauer lemma* asserts that if the VC-dimension of $\mathcal{C}$ is d and $|U| = n$, then $|\mathcal{C}| \leq \sum_{i=0}^{d} \binom{n}{i}$. The concept classes for which the upper bound is sharp are called *maximum classes*, and constitute important and well-studied objects. Kuzmin and Warmuth [112] constructed representation maps for maximum classes. They also noticed that the so-called *corner peelings* define natural representation maps and conjectured that maximum classes always admit corner peelings (a *corner peeling* of $\mathcal{C}$ is a dismantling order of the concepts $C_1, \ldots, C_n$ of $\mathcal{C}$ such that each C_i belongs to a unique cube in the concept class $\{C_i, C_{i+1} \ldots, C_n\}$). This conjecture was resolved in the affirmative by Rubinstein and Rubinstein [126] using geometric and topological techniques.

Another important general result for concept classes $\mathcal{C}$ is the following *sandwich lemma.* Denote by $\overline{X}(\mathcal{C})$ the collection of all shattered by $\mathcal{C}$ subsets of U. Denote also by $\underline{X}(\mathcal{C})$ the collection of all strongly shattered by $\mathcal{C}$ subsets of U, i.e., subsets Y such that there exists an Y-cube included in $\mathcal{C}$. Both $\overline{X}(\mathcal{C})$ and $\underline{X}(\mathcal{C})$ are simplicial complexes and $\underline{X}(\mathcal{C}) \subseteq \overline{X}(\mathcal{C})$. The sandwich lemma [63, 86, 57] asserts that $|\underline{X}(\mathcal{C})| \leq |\mathcal{C}| \leq |\overline{X}(\mathcal{C})|$ holds for any $\mathcal{C}$. The concept classes $\mathcal{C}$ for which equality $|\mathcal{C}| = |\overline{X}(\mathcal{C})|$ were called *ample* [86, 57] or *extremal* [63]. It turned out that this is equivalent to the equality $\underline{X}(\mathcal{C}) = \overline{X}(\mathcal{C})$ (due to this, we will denote this complex by $X(\mathcal{C})$) and they are equivalent to *lopsided sets* studied before in [114]. Lopsided sets generalize median structures and maximum classes and can be characterized in a multitude of combinatorial, recursive, and metric ways. In particular, they are exactly the isometric subgraphs of the hypercube $\{0,1\}^U$ such that for any strongly shattered set Y any two Y-cubes Q', Q'' included in $\mathcal{C}$ can be connected by a shortest gallery consisting of Y-cubes of $\mathcal{C}$.

Due to the equality $|\mathcal{C}| = |X(\mathcal{C})|$ for lopsided sets, one can ask if *there exists a bijective map* $r : \mathcal{C} \to X(\mathcal{C})$ *which is a representation map.* This question was asked recently by Moran and Warmuth [118]. They also conjectured that lopsided (extremal) classes admit a corner peeling. The same question was asked before by Weideman [134] and by Chepoi (unpublished, 1997). Since 2004-2005 we know (unpublished) that it has a negative answer: we showed that corner peeling of a lopsided set $\mathcal{C}$ is equivalent to an isometric dismantling of $\mathcal{C}$ and to the extendable shellability of the octahedron dual to the cube $\{0,1\}^X$. However, in her PhD thesis from 2004, Tracy Hall [101] showed that octahedra are not extendably shellable. Her counterexample provides us with a lopsided set $\mathcal{C}$ on 299 vertices (concepts) of a 12-dimensional cube not having any corner. In fact, this counterexample is a maximum class of VC-dimension 3, showing that the result of [126] is false. After communicating this example to Moran and Warmuth, they noticed that it also contradicts the representation map constructed in [112], thus the *question of existence of representation maps for maximum classes is also open.*

In the attempt of constructing representation maps for lopsided classes, we proved a *local-to-global result* which makes a link between representation maps and unique sink orientations (USO) (such orientations were previously intensively studied for cubes [131]). We hope that this *structural characterization can be helpful in constructing representation maps.* If representation maps for lopsided classes will be constructed, then the general Littlestone and Warmuth problem can be attacked from the following combinatorial angle: *show that any maximal (by inclusion) concept class* $\mathfrak{C}$ *of VC-dimension* d *can be covered by* $O(d)$ *lopsided sets of VC-dimension* $O(d)$.

Objectives: *Representation maps for lopsided and maximum classes; covering of concept classes of VC-dimension* d *with* $O(d)$ *lopsided classes of* $O(d)$ *dimension.*

Related themes: S1,S2,S4,S6,A2

Theme S4: *Matroidal structures.* A *matroid* on a set X is a collection $\mathcal{B}$ of subsets of X, called *bases*, satisfying the exchange property: for all $A, B \in \mathcal{B}$ and $i \in A \setminus B$ there exists $j \in B \setminus A$ such that $A \setminus \{i\} \cup \{j\} \in \mathcal{B}$. All bases have the same cardinality. The *basis graph* of $\mathcal{B}$ is the graph whose vertices are the bases of $\mathcal{B}$ and whose edges are the pairs A, B such that $|A \triangle B| = 2$. The *basis matroid polytope* is the convex hull of the indicator vectors of bases. It is well-known that the basis graph is the 1-skeleton of the basis matroid polytope. Basis graphs have been nicely characterized in a metric way by Maurer [116] (a local-to-global characterization was provided in [22]); for similar characterizations of even Δ-matroids see [74, 22].

Some of the most important open questions about matroids are about their bases (for example, Rota's conjecture for which there is a polymath project) or their *basis graphs* (cyclic ordering of bases, matroids are expanders). Some of these questions can be considered also for even Δ-matroids, dual polar graphs, or other subclasses of meshed graphs. We would like to use our knowledge of basis graphs to approach this kind of difficult questions. The first question is a conjecture by Wiedemann [134] about cyclic ordering of bases of a matroid. Our interest to this conjecture comes from the fact that it can be formulated in completely metric terms: *any two bases* B', B'' *of a matroid lie on a common isometric cycle of the basis graph.* Notice that this property holds for dual polar graphs [19]. Another question is the difficult conjecture by Mihail and Sudan (unpublished) that matroids are expanders (more exactly, that *basis graphs of matroids are expanders*). A graph $G = (V, E)$ is an *expander* if for any partition of V into V', V'' the number of edges running between V' and V'' is at least $\min\{|V'|, |V''|\}$. One method of attempting to prove that a graph G is an expander is the *canonical path method.* It consists in constructing for each pair of vertices x, y an (x, y)- and an (y, x)-path such that for each edge uv of G there exists at most $|V|$ paths passing via uv. We believe that this conjecture is related with the previous one and that as the canonical paths between two bases B', B'' one can select the paths in the isometric cycle passing via B' and B''.

The next problem was asked by Mayr and Plaxton [117] for graphic matroids and is motivated by the problem of computing k-minimal bases of matroids. We will formulate it the general context of matroids. Let $f : V \to \mathbb{R}$ be a function on a graph $G = (V, E)$. Let $\alpha_1 < \alpha_2 < \ldots < \alpha_m$ be the values taken by f. A vertex v has *rank* i if $f(v) = \alpha_i$. Mayr and Plaxton [117] proved that if T is a minimum spanning tree of a some graph and $i \leq m$, then there exists a spanning tree T' or rank i such that the distance $d(T, T')$ between T and T' in the basis graph is $\leq i - 1$. They also formulated the following conjecture: *if* $1 \leq j < i \leq m$ *and* T *is a spanning tree of rank* j*, then there exists a spanning tree of rank* i *at distance* $\leq i - 1$ *from* T. We call a function f on a graph G *square-additive* if for any square $uvwx$ of G, $f(u) + f(w) = f(v) + f(x)$. Any function on bases which is the sum of the weights of the elements is square-additive; compare with discrete convex functions [119]. The result and the conjecture of Mayr and Plaxton [117] can be extended to square-additive functions on bases graphs of matroids: *f is a square-additive function on a basis graph taking m different values,* $1 \leq j < i \leq m$*, and B is a basis of rank j, then there is a basis B' of rank i at distance* $\leq i - 1$ *from B*. This kind of results are relevant to the structure of landscapes of square-additive functions and provide bounds for the number of steps in descent-type algorithms of finding k-minima of such functions.

The *Tutte polynomial* of a matroid is a famous combinatorial invariant, involving a huge literature and appearing in a number of contexts, ranging from algebraic graph theory and enumerative combinatorics to knot theory and statistical physics. It can be seen as witnessing various structural properties of bases in matroids and of orientations (or signatures) in oriented matroids (or hyperplane arrangements), by means of counting or decomposing objects (bases with respect to activities, no-broken-circuit subsets, orientations with respect to activities, etc.). The active bijection ([43, 44] and other papers in this series) relates together those objects and properties in matroids and oriented matroids. In relation with MGT, *these objects and properties can often be expressed in terms of distances and partitions in associated graphs, polytopes and hypercubes:* basis activities, whose generating series is the Tutte polynomial, can be interpreted as distances of bases to the minimal/maximal lexicographic bases; region activities can be viewed as distances in the face lattice, and hence in the cocircuit/tope graphs, with respect to the minimal/maximal flag of faces of a topological representation; subset activities can be understood in terms of a partition of the (subset) hypercube into boolean lattices associated with bases (themselves forming a graded lattice whose lattice-rank is given by basis activities); orientation activities can be understood in terms of a partitions of the (reorientation) hypercube into boolean lattices given by the above region activities along with duality. The way these objects and properties can be understood in these terms, they are likely to be more understandable or generalizable in metric graph theory or in the covector signature studied in this project (partial cubes, lopsided sets, COMs, etc.), and be related to other combinatorial structures.

We would also like to continue the development of the theory of complexes of oriented matroids (COMs), recently introduced and investigated in [58]. In [58] we proposed a common generalization of oriented matroids (OMs), median graphs, and lopsided sets, called COMs, and intimately related to MGT. In this generalization, global symmetry and the existence of the zero sign vector, required for OMs, are replaced by local relative conditions. COMs can be viewed as complexes whose cells are OMs and which are glued together in a lopsided fashion These novel structures can be characterized in terms of two axioms, generalizing the familiar characterization for oriented matroids. Like hyperplanes of CAT(0) cube complexes, the hyperplanes of COMs are COMs and we characterize COMs in terms of their hyperplanes. We also described a gluing scheme by which every COM can successively be erected as a certain complex of oriented matroids, in essentially the same way as a lopsided set can be glued together from its maximal cubes. A realizable COM is represented by a hyperplane arrangement restricted to an open convex set. Among these are the examples formed by linear extensions of ordered sets, generalizing the OMs corresponding to the permutohedra. Relaxing realizability to local realizability, we capture a wider class of combinatorial objects: we show that CAT(0) Coxeter zonotopal complexes give rise to locally realizable COMs. The characterizations of COMs via hyperplanes or carriers is metric; moreover, the tope graphs of COMs are partial cubes. More recently, the paper [109] (generalizing and interpreting the results of [114] and [83] in a metric way) presented a characterization of tope graphs of COMs as partial cubes in which antipodal subgraphs are gated. In the project, we would like to generalize the theory of COMs along the lines the theory of OMs was developed. Namely, we would like to define for COMs the duality theory and the basis orientations (in case of OMs they are characterized via the chirotope). One of the principal conjectures of [58] is that COMs are *convex subgraphs of OMs* (i.e., they can be represented as the intersection of halfspaces of an OM). This extends the conjecture of Lawrence [114] that the lopsided sets are exactly the convex subgraphs of uniform oriented matroids. Another (quite opposite) conjecture is that of Las Vergnas claiming that any simple pseudohyperplane arrangement admits a simplicial cell. In metric terms, this is equivalent to say: any antipodal partial cube G whose intersection with any proper subcube is lopsided has a vertex of degree at most the dimension of G (the dimension of the largest cube of G). It will be interesting to see how these two conjectures work for Tracy Hall's example [101].

Objectives: *Wiedemann cyclic ordering of bases conjecture; Mayer and Plaxton landscape conjecture; generalizations of properties involving the Tutte polynomial for matroids and oriented matroids in terms of cubes and distances; duality theory for COMs; COMs as convex subgraphs of OMs; Las Vergnas's conjecture about simplicial cells.*

Related themes: S1,S2,S3,A2

Theme S5: *Isometric and low distortion embeddings.* A metric space (X, d) is *isometrically embeddable* into a host metric space (Y, d') if there exists a map $\varphi : X \mapsto Y$ such that $d'(\varphi(x), \varphi(y)) = d(x, y)$ for all $x, y \in X$. More generally, $\varphi : X \mapsto Y$ is an *embedding with multiplicative distortion* $\lambda \geq 1$ if $d(x, y) \leq d'(\varphi(x), \varphi(y)) \leq \lambda \cdot d(x, y)$ for all $x, y \in X$. In this project, we would like to address some questions about isometric embeddings of graphs into hypercubes and half-cubes, and about existence of bounded distortion ℓ_1-embeddings of planar graphs and 1-skeletons of some 2-dimensional cell complexes.

Djoković [85] characterized graphs isometrically embeddable into hypercubes in the following simple but pretty way: $G = (V, E)$ can be isometrically embedded into a hypercube iff G is bipartite and for any edge uv, the (disjoint) sets $W(u, v)$ and $W(v, u)$ are convex, where $W(u, v) = \{x \in V : d(x, u) < d(x, v)\}$. The pairs of disjoint halfspaces $\{W(u, v), W(v, u)\}$ define a space with walls. Recently we found a Djoković-type characterization of graphs isometrically embeddable into Johnson graphs [25]. Shpectorov [127] provided an efficient characterization of all ℓ_1-graphs (ℓ_1-*graphs* are the graphs which can be isometrically embedded into an ℓ_1-space): a finite graph G is an ℓ_1-graph iff G isometrically embeds in a Cartesian product of octahedra and half-cubes. For the moment a structural characterization of ℓ_1-graphs or isometric subgraphs of half-cubes is missing; see [84, Problem 21.4.1]: *provide a structural characterization of ℓ_1-graphs and of graphs isometrically embeddable into half-cubes. Give a forbidden-subgraph or a local-to-global characterization of ℓ_1-weakly modular and ℓ_1-meshed graphs.* Due to the major role played by the dual polar graphs in the structure of (s)weakly modular graphs [19], it would be interesting to *characterize isometric subgraphs of Cartesian products of dual polar graphs.*

The following question concerns a Menger-type characterization (in [51] called local-to-global characterization) of isometric embedding into ℓ_1-spaces of dimension n. Let $c_p(n)$ denote the least positive integer such that a metric space (X, d) isometrically embeds into $\mathbb{R}^n$ with ℓ_p-metric iff each subspace of X with at most $c_p(n)$ points embeds. The Menger theorem asserts that $c_2(n) = n + 3$. In [55], we showed that $c_1(2) = c_\infty(2) = 6$. J. Edmonds [89] proved that $c_\infty(n) = \infty$ for $n \geq 3$, i.e., the ℓ_∞-metrics with $n \geq 3$ do not admit a Menger-type theorem. An important remaining open question is *Is $c_1(n)$ for $n \geq 3$ finite? Compute $c_1(3)$.* We conjecture that $c_1(3)$ is finite (it was noticed in [55] that $c_1(3) \geq 10$). The simpler variants can be asked for ℓ_1-embedding of *graph-metrics* and for *isometric embedding into any normed plane.*

Finally, we would like to investigate the question of embeddability with constant distortion of several graph classes into ℓ_1-spaces. The motivation stems from the famous *planar embedding conjecture* of Linial, London, and Rabinovich [72, 115, 97] asserting that *metrics of planar graphs can be embedded into ℓ_1 with constant distortion.* We would like to find an approximate

Djoković-like result for this problem, which can be useful for constructing the embeddings. For this probably one has to appropriately relax the notion of convexity. In [20] we proved that all planar graphs which are 1-skeletons of planar CAT(0) complexes with regular Euclidean polygons as cells are ℓ_1-embeddable with distortion at most 2 (this significantly improved and simplified the result of [128]). Finding other large classes of planar graphs ℓ_1-embeddable with bounded distortion is an interesting and important question. It can be shown that to solve the planar embedding conjecture it suffices to solve it for planar quadrangulations, i.e., for planar graphs in which all inner faces are 4-cycles. One interesting class of such quadrangulations is constituted by the *planar special cube complexes* of Haglund and Wise [100]. For example, the square grid from which we removed a set of disjoint rectangular subgrids is a special cube complex. Already for this particular class we do not know how to perform the ℓ_1-embedding.

Beside planar metrics, only very few classes of finite metrics or graph-metrics are known to have bounded-distortion ℓ_1-embedding. Finding such an embedding would also imply that respective metrics admit a system of walls with strong properties. We intend to investigate 1-skeletons of some simplicial and cubical complexes, occurring in geometry and geometric group theory: 2-dimensional systolic and special cube complexes, 2-dimensional folder complexes.

Objectives: *Isometric subgraphs of half-cubes; Menger-type results for ℓ_1-embedding into $\mathbb{R}^3$; bounded distortion ℓ_1-embedding for classes of 1-skeletons of 2-dimensional complexes, in particular, for planar special cube complexes.*

Related themes: S1,S5,A2,A3

Theme S6: *Packing and covering with balls, identifying codes, and χ-boundedness.* The packing and covering problems are classical themes in computer science and combinatorics. In the *set covering problem*, given a collection $\mathcal{F}$ of subsets of a domain X, the task is to find a subcollection of $\mathcal{F}$ of minimum size $\rho(\mathcal{F})$ whose union is X. The *set packing problem* asks to find a maximum number $\nu(\mathcal{F})$ of pairwise disjoint subsets of $\mathcal{F}$. A problem closely related to set covering is the hitting set problem. A subset T is a *hitting set* of $\mathcal{F}$ if $T \cap S \neq \varnothing$ for any $S \in \mathcal{F}$. The *minimum hitting set problem* asks to find a hitting set of smallest cardinality $\tau(\mathcal{F})$. All these three problems are NP-hard and difficult to approximate. If X is a metric space and $\mathcal{F}$ is the set of its balls of equal radii, then the minimum covering and the minimum hitting set problems are equivalent, i.e., $\rho(\mathcal{F}) = \tau(\mathcal{F})$. The inequality $\tau(\mathcal{F}) \geq \nu(\mathcal{F})$ holds for any family $\mathcal{F}$ and any domain X. Of particular importance are the families $\mathcal{F}$ for which there exists a universal constant c such that $\tau(\mathcal{F}') \leq c\nu(\mathcal{F}')$ holds for any subfamily $\mathcal{F}'$ of $\mathcal{F}$ (we will say that such families have the *bounded covering-packing property*). Establishing the bounded covering-packing property is a notoriously difficult problem and it is open for many simple particular cases, for instance for axis-parallel rectangles in $\mathbb{R}^2$.

We are interested in classes of graphs for which the set of balls of equal radii R has the bounded covering-packing property. In [76], we proved that if S is a compact subset of a geodesic δ-hyperbolic space or graph, then $\rho_{R+2\delta}(S) \leq \nu_R(S)$, where $\rho_{R+2\delta}(S)$ is the covering number of S with balls of radius $R + 2\delta$ and $\nu_R(S)$ is the packing number of S with balls of radius R. This result is useful if the hyperbolicity is much smaller than the radius R of balls used in the covering. (In [29] we obtained similar relationships between the sizes of packings and hitting sets of quasiconvex sets in hyperbolic graphs). More recently, in [30] we proved that Busemann surfaces satisfy the bounded covering-packing property: namely, if S is a compact subset of a Busemann surface, then $\rho(S) \leq 19\nu(S)$. In [31] it was shown that planar graphs of diameter $2R$ can be covered with a constant number of balls of radius R. This result was generalized to all graphs on surfaces of a given genus [64]; see also [67] for other generalizations.

It was conjectured in [31] that the class of planar graphs has the bounded covering-packing property, and this is the main problem in this domain we want to solve: *is it true that for any R, the balls of radius R of any planar graph satisfy the bounded covering-packing property (with a universal constant not depending of R)?* A positive answer to this question can be established by showing that planar graphs satisfy the *weak doubling property*: *there exists a universal constant c such that any planar graph G contains a ball of radius $2R$ which can be covered by c balls of radius R.* Notice also that questions similar to those for planar graphs can be also considered for *any compact (finite) subset of points of an arbitrary polygon (with holes) endowed with the geodesic metric.* The existing proofs of [31, 64, 67] are strongly related with the notions of VC-*dimension* and the *Hadwiger-Debrunner* (p,q)*-property* for balls. We would like to understand and exploit further the relation between VC-dimension, (p,q)-property, and packing/covering in order to generalize and extend these results to new classes of graphs. Namely, *what other classes of graphs or metric spaces have the bounded covering-packing property, weak doubling property, bounded VC-dimension or the Hadwiger-Debrunner* (p,q)*-property?*

Similarly to the nice labeling problem from S2, the covering/packing problem with balls can be viewed as a coloring problem of the complement of the intersection graph of balls and the question can be restated as the χ-boundedness of the resulting class of graphs (a class $\mathcal{G}$ of graphs is χ*-bounded* if there exists a constant c such that for any $G \in \mathcal{G}$ the chromatic number $\chi(G)$ and the clique number $\omega(G)$ are related by the inequality $\chi(G) \leq c \cdot \omega(G)$). We hope that a better understanding of ball coverings of planar graphs can be useful in the coloring of *exact* d*-powers* of graphs and metric spaces (for example, the hyperbolic plane), in which vertices at distance exactly d must receive distinct colors. For instance, in the case of trees, this question is deeply related with the coloring of hyperbolic spaces but also with some questions on the chromatic number of powers of bounded expansion classes of graphs. Very recently, the paper [66] (including among the coauthors N. Bousquet, a member of this project) presented a counterexample to the conjecture by van den Heuvel and Naserasr (another member of this project) that exact odd powers of planar graphs can be colored in a constant number of colors. It will be interesting to investigate *how* χ*-boundedness of exact* d*-powers of planar graphs and of general metric spaces is related to bounded doubling dimension.*

The classical *three longest path conjecture* by T. Gallai asserts that any three longest paths of a connected graph G have a common point. This is easily true for two longest paths and false for seven and more longest paths (the question remains open for 3,4,5, and 6 paths). Given three longest paths P_1, P_2 and P_3, we considered $f(P_1,P_2,P_3) = \min\{d(v,P_1)+d(v,P_2)+d(v,P_3) : v \in V\}$ to be the shortest total distance of a vertex to the paths and defined $\hat{f}(n)$ to be maximum over all $f(P_1,P_2,P_3)$ and over all graphs on n vertices. The conjecture claims that $f(P_1,P_2,P_3) = 0$ for any triplet P_1,P_2,P_3 of longest paths and thus $\hat{f}(n) = 0$ for all n. In an ongoing work, we proved that if the conjecture is false and there exits a graph G and longest paths P_1,P_2,P_3 such $f(P_1,P_2,P_3) \neq 0$, then we can extend this example with a linear growth of $\hat{f}(n)$ (hence to prove the conjecture it suffices to show that the function $\hat{f}(n)$ is sublinear).

A set M of vertices of a graph G is a *metric basis* if for each pair u,v of distinct vertices, there exists a vertex x of M with $d(x,u) \neq d(x,v)$. The smallest size of a metric basis of G is the *metric dimension* of G. We would like to investigate the metric dimension for planar graphs, namely how the order of the graph could be bounded in terms of its diameter and metric dimension. For planar graphs of metric dimension 2, it is known that $n = O(D^2)$, where D is the diameter of the graph. This bound is reached by any square grid. There exist planar graphs with metric dimension $k = 3$ and order $\Theta(D^3)$ and for general k it is known that $n = O(D^4k^4)$ [60]. Hence the following question is interesting: *Do there exist planar graphs with small metric dimension and order* $O(D^4)$*? Can the bound* $n = O(D^4k^4)$ *be improved?*

For an integer $\rho \geq 1$, a set S of vertices of a graph $G = (V,E)$ is an ρ*-identifying code* if (1) S is a hitting set of the set $\mathcal{B} := \{B(v,\rho) : v \in V\}$ of ρ-balls of G and (2) S *separates* the vertices of G: $B(u,\rho) \cap S \neq B(v,\rho) \cap S$ for all $u,v \in V$. Notice the following surprising link between the separation condition for identifying codes and the non-clashing condition for representation maps: viewing $\mathcal{B}$ as a concept class of $\{0,1\}^V$, if S is an ρ-identifying code for G, then setting $r(B(v,\rho)) := B(v,\rho) \cap S$ for any $v \in V$, the map r satisfies the non-clashing condition and thus is a representation map for $\mathcal{B}$. Vice-versa, if $r : \mathcal{B} \to \{0,1\}^V$ is a representation map such that $r(B(v,\rho)) \subset B(v,\rho), v \in V$, then the set $S := \bigcup_{v\in V} r(B(v,\rho))$ is an ρ-identifying code. *This link merits future investigations and can enrich both problems.* Another question about ρ-identifying codes is the surprising monotonicity conjecture by Moncel [105] for hypercubes and studied extensively since then: *the minimum size* $\gamma_{id}(Q_d)$ *of an* ρ*-identifying code of a* d*-hypercube* Q_d *is monotonically increasing with its dimension:* $\gamma_{id}(Q_d) \leq \gamma_{id}(Q_{d+1})$. On the other hand, for

$\rho = 1$ it is known that $\gamma_{id}(Q_{n+2}) \leq 4\gamma id(Q_n)$ and that $\gamma_{id}(Q_{n+1}) \leq (2 + \frac{1}{n+1})\gamma_{id}(Q_n)$ [91]. Surprisingly, the following is open: *show for $\rho = 1$ that $\gamma_{id}(Q_{n+1}) \leq 2\gamma_{id}(Q_n)$* (this holds [62, 124] if Q_n has an identifying code with no isolated vertices).

Objectives: *Bounded covering-packing and weak doubling properties for balls in planar graphs; coloring of exact powers of graphs and metric spaces; metric dimension of planar graphs and ρ-identifying codes in hypercubes, the three longest path conjecture.*

Related themes: S3,S5,A2,A3

1.5. **Algorithms in Metric Graph Theory.**

Theme A1: Algorithmic aspects of hyperbolic graphs. *δ-Hyperbolic metric spaces* have been defined by M. Gromov via a simple 4-point condition: for any four points u, v, w, x, the two larger of the distance sums $d(u, v)+d(w, x), d(u, w)+d(v, x), d(u, x)+d(v, w)$ differ by at most 2δ (hyperbolicity of graph-metrics and geodesic metric spaces can be characterized in a multitude of ways: thin and slim geodesic triangles, linear isoperimetric inequality, exponential divergence of geodesics, etc). Geodesic hyperbolic spaces and infinite hyperbolic graphs play a very important role in modern metric geometry and geometric group theory: the word problem in hyperbolic groups in decidable and many graphs and simplicial complexes (curve complex, arc complex, etc.) arising from the topology of surfaces are hyperbolic. Like the treewidth measures how close a graph is to a tree from a connectivity point of view, the hyperbolicity measures how close a graph is to a tree from a metric point of view. In particular, the metric spaces of trees are 0-hyperbolic. Designing fast and accurate algorithms for *hyperbolic graphs* is motivated by numerous empirical results establishing that many real world networks and graphs have small hyperbolicity. An active line of research consists in using the specific metric properties of δ-hyperbolic graphs to derive structural and algorithmic results relevant for such real-life graphs and networks. Many algorithmic results on graphs are typically obtained under structural conditions like bounded treewidth, bounded genus, or minor-closedness. Thus, one can ask *to what extent for metric problems such conditions can be replaced by hyperbolicity δ?*

The participants of this project from different centers already substantially contributed to this line of research. The distance query problem in δ-hyperbolic graphs has been investigated in [95]. Simple linear time algorithms for diameter and center problems have been proposed in [26] and polynomial time algorithms for covering and packing with balls (with an additive error depending of δ) have been proposed in [76]. Answering on open question from [21], a game theoretical characterization of δ-hyperbolicity was given in [23] where it is shown that all graphs in which a cop moving at speed s can catch a robber moving at speed larger than s are δ-hyperbolic with $\delta = O(s^2)$. This allows to provide a constant factor approximation of the hyperbolicity of a graph G (the smallest δ such that G is δ-hyperbolic) in $O(n^2)$ time (n is the number of vertices). Very recently, answering a question by Jonckheere et al. [107], we proved in [29] that any δ-hyperbolic network admits a *core*, i.e., a ball of radius $O(\delta)$ which intercepts all shortest paths between at least $\frac{n^2}{4}$ pairs of n arbitrary vertices of G. Such a result explains the experimental observation of Narayan and Saniee [120] that real-world networks with small hyperbolicity have a core congestion. In [29], we also extended the results of [76] about covering and packing with balls to arbitrary quasiconvex sets of hyperbolic graphs.

These results (in particular, those about covering and packing with quasiconvex sets) suggest the existence of a kind of *meta-theorem* (à la Courcelle) asserting that efficient algorithms and good characterizations that works for trees have a $O(\delta)$-approximated counterpart for δ-hyperbolic graphs. Our experience shows that finding the correct formulation of these results and their proofs requires a deep understanding of the metric properties of δ-hyperbolic spaces. In this project, we intend to continue this research and obtain this kind of results for multiflow and multicut problems, k-server problem, hub labeling, distance and routing labeling schemes.

In the *k-server problem*, an online algorithm ALG controls k mobile agents (servers) located at the points of a metric spaces (X, d). A sequence of requests $\sigma = r_1, r_2, \ldots, r_n \in X$ has to be *served* by the agents, i.e. some agent has to move to the point r_i. For a sequence of requests σ, the cost of ALG is the total distance traveled by his agents to serve σ. The *k-server conjecture* asserts that for any metric space (X, d) there exists a k-competitive algorithm (an algorithm is said to be *c-competitive* for a problem Π if there is a constant α such that for each instance $\sigma \in \Pi$, $\mathrm{ALG}(\sigma) \leq c \cdot \mathrm{OPT}(\sigma) + \alpha$, where $\mathrm{OPT}(\sigma)$ is the cost of an optimal offline algorithm). There is a $(2k-1)$-competitive algorithm for any metric space [111] and the conjecture holds for $k = 2$. On the other hand, there exists a simple k-competitive algorithm for trees [81] (with a nice competitiveness analysis). We would like to adapt the algorithm for trees to *all δ-hyperbolic graphs and geodesic metric spaces to obtain a $(k + O(\delta))$-competitive algorithm.*

Given a graph $G = (V, E)$, an edge-capacity function $c : E \to \mathbb{R}^+$ and a set of k pairs of terminals $s_1t_1, s_2t_2, \ldots, s_kt_k$, the *multicut problem* is to find a subset of edges $F \subseteq E$ of minimum capacity that intercepts all paths between s_i and t_i for $i = 1, \ldots, k$. This problem is already NP-hard for trees but, in this case, there is a polynomial 2-approximation algorithm [94]. An *R-multicut* is an analog of a multicut in which the set of edges that intercept all paths between terminal pairs is replaced by a set of balls of radius R that intercept all shortest paths between terminal pairs. We would like to answer the following non-trivial question: *Is it possible to design a constant factor approximation algorithm for the R-multicut problem when G is δ-hyperbolic graph and $R = O(\delta)$?* We will also consider other multicut-multiflow problems (for example, the sparsest cut problem) and other optimization problems on δ-hyperbolic graphs related with distances (for example, the multifacility location problem).

One of the recent advances in practical computation of shortest paths is based on hub labelings of graphs. A *hub labeling* of a graph $G = (V, E)$ associates to each vertex $v \in V$ a subset $H(v) \subset V$ called the *hub of* v and the set of distances $\{d(v, x) : x \in H(v)\}$, such that given any two vertices $u, v \in V$, there exists $x \in H(u) \cap H(v) \cap I(u, v)$ (recall that $I(u, v)$ is the set of all shortest paths between u and v). Then $d(u, v)$ can be easily retrieved as $d(u, x) + d(x, v)$. The goal is to find a hub labeling of small size, say a hub labeling of smallest ℓ_1-*size* $\sum_{v \in V} |H(v)|$ or of smallest ℓ_∞-*size* $\max\{|H(v)| : v \in V\}$. Recently, Angelidakis et al. [50] provided a linear programming analysis of a simple algorithm for hub labeling of trees (which assign labels of size $\log n$ to all vertices) and proved that this algorithm is a factor 2 approximation algorithm for the L_1-size of hubs. We would like to *extend this algorithm and its analysis to δ-hyperbolic graphs.* This is more challenging than it seems because of lower bounds of the size of distance labeling schemes in hyperbolic graphs established by Gavoille and Ly [95].

Objectives: *Fast and accurate approximation algorithms for optimization problems with metric data for δ-hyperbolic graphs; a $(k+O(\delta))$-competitive algorithm for k-server problem; constant factor algorithms for the R-multicut problem and for the hub labeling of minimal ℓ_1-size.*

Related themes: S2,S4,S6,A2

Theme A2: *Algorithms for graph classes from MGT.* Metric Graph Theory supplies classes of graphs with interesting metric properties. It is natural to ask various algorithmic questions about these graphs. The first one is *efficient testing* if a graph G belongs to a given class of graphs. This "property testing" was already done for some classes. We would like to *investigate it for various classes of isometric subgraphs of hypercubes, half-cubes, and Johnson graphs.* In relation with efficient testing, we would like to investigate properties and algorithmic applications of Breadth First Search (BFS), Lexicographic Breadth First Search (LexBFS), and Lexicographic Depth First Search (LexDFS). For example, for numerous classes of graphs, LexBFS provides total orders with useful properties (perfect elimination order for chordal and interval graphs, domination order for bridged and weakly bridged graphs) and has numerous applications in efficient testing of graph properties (for example, in the recent papers [82, 113] they were used, respectively, for designing linear time algorithms for cocomparability graphs and to recognize Robinsonian dissimilarities). In the project, we intend to *use LexBFS to recognize convex geometries.* We noticed that on some classes of graphs, LexBFS and LexDFS orders have properties close to greedoids. We plan to investigate in details the relationships between convex geometries and these classes of graphs. Another important question is to *design more efficient* *algorithms for classical metric problems like shortest path, rectilinear TSP with constraints, diameter, center, median problems, and other facility location problems* for particular (but quite general) graph classes (for example, diameter, center, and median problems on Helly graphs).

In the distributed setting, it will be very natural to *investigate which classes from MGT admit compact encoding schemes with fast answers to (a) adjacency queries, (b) distance queries and geodesic routing. Which classes admit encoding schemes with polylogarithmic labels?* The size of the adjacency scheme for G is often related to the arboricity or density of G. We will investigate these parameters for subgraphs G of Cartesian products of graphs, for subgraphs of half-cubes and of Johnson graphs. In case of subgraphs of Cartesian products, we expect that in many cases they are bounded by the maximum of respective parameters of factors and the VC-dimension of G (upper bounded by $\log |V(G)|$). This will generalize the classical result of Haussler et al. [103] that for any subgraph $G = (V, E)$ of a hypercube, $\frac{|E|}{|V|}$ does not exceed the VC-dimension of the set family defined by V. It would be interesting and important if this inequality can be extended to subgraphs of half-cubes and of Johnson graphs (for an appropriate definition of the VC-dimension). We also plan to investigate distance queries and geodesic routing problems based on hub labeling for such classes of graphs as median graphs and bridged graphs. Another algorithmic question is *which classes of graphs from MGT admit sparse spanners (with $O(n)$ or $O(n \log n)$ edges) and constant distortion or are quasi-isometric to sparse graphs.* Also we plan to investigate the notions (alternative to VC-dimension and doubling dimension) of *highway dimension* [48] and *skeleton dimension* [110], both related to hub labeling schemes. We also propose to investigate generalizations of hub labeling. One interesting approach concerns *h-hopsets*, that is a set of additional (transitive) edges so that any pair of nodes is connected by a shortest path of at most h hops (i.e., at most h edges). Such structure are particularly interesting for parallel computing of shortest path trees and for approximating shortest paths [90]. A hub labeling can be seen as a 2-hopset where the L_1-size of the hub labeling corresponds to the number of edges in the 2-hopset. We will investigate how graph classes from MGT enable better trade-offs between number of hops h and number of edges in the h-hopset.

An important topic in this theme is the investigation of *algorithmic and complexity aspects of geodesic convexity in graphs* and the *construction of different types of envelopes relaxing the classical convex hull.* The first question concerns the computation of such convexity parameters as the hull number, Helly, Radon, and Caratheodory numbers, and their fractional analogs in classes of graphs (it is known that computing the hull number is NP-hard already for partial cubes [3]). The geodesic convex hull $\mathrm{conv}(S)$ of a subset S of vertices of a graph G can be constructed in polynomial time, however in many cases the size of $\mathrm{conv}(S)$ is much bigger than the size of S (can be exponential in $|S|$ and close to the size of G). Thus one can ask to construct subsets of $\mathrm{conv}(S)$ containing S, of size polynomial in parameters of S, and satisfying metric properties of the host space G. In case of subsets S of median graphs (ℓ_1-spaces) and in case of Helly graphs (injective spaces), there are canonical ways to construct smallest median subgraph $\mathrm{med}(S)$ and Helly subgraph $\mathrm{helly}(S)$ containing S: $\mathrm{med}(S)$ is the median closure of S (called also the cubulation of S) and $\mathrm{helly}(S)$ is the Hellification of S (the discrete counterpart of the injective hull of S [87, 104]). $\mathrm{med}(S)$ and $\mathrm{helly}(S)$ are subsets of $\mathrm{conv}(S)$ but still may be exponential in $|S|$. If S is a subset of $\{0, 1\}^X$, one can define analogous envelopes with respect to properties being *lopsided* or being a *partial cube*: $\mathrm{lop}(S)$ is a smallest lopsided set containing S and $\mathrm{pc}(S)$ is a smallest partial cube containing S (however $\mathrm{lop}(S)$ and $\mathrm{pc}(S)$ are no longer canonically defined). One can ask if *for any $S \subset \{0, 1\}^X$ of VC-dimension d, $lop(S)$ and/or $pc(S)$ have size polynomial in S and VC-dimension $O(d)$?* (For lopsided sets, this question is related to the compression conjecture of Littlestone and Warmuth). Instead of VC-dimension one can ask the same question about the diameter or the TSP-perimeter of S (i.e., the length of a shortest TSP tour for S). Similar questions can be asked for other host graphs: Johnson graphs, half-cubes, bridged graphs. In [77] we investigated Pareto envelopes in $(\mathbb{R}^n, \ell_\infty)$ and $(\mathbb{R}^3, \ell_1)$; the *Pareto envelope* of a set S consists of all points of the space whose distance-vector to the points of S is not Pareto-dominated. Even if they coincide with injective hulls in the first case and with median hulls in the second case, Pareto envelopes of subsets S of $\{0, 1\}^X$, $|X| > 3$, are no longer median. In relation with previous questions, it will be interesting to *investigate the structure and the complexity of Pareto envelopes of subsets of hypercubes.*

Our last topic concerns complexity and algorithmic issues about identifying codes and metric dimension. The identifying codes ($\rho = 1$) and metric dimension problems were shown to be NP-complete for interval graphs and permutation graphs [93]. What about more restricted classes of graphs: *unit interval graphs and bipartite permutation graphs?* More wide and general question is *What is the impact of hyperbolicity on various identification parameters (like ρ-identifying codes, locating ρ-dominating sets, or metric dimension)?* For graphs with bounded tree-length (which are hyperbolic) the metric dimension problem is FPT when parametrized by the solution size [61]. Maybe similar bounds can be derived in terms of hyperbolicity.

Objectives: *Fast algorithms for efficient testing (in particular, using BFS, LexBFS or LDFS); fast algorithms for classical metric problems (shortest path, diameter, center, median, rectilinear TSP); adjacency schemes and density results based on VC-dimension for subgraphs of Cartesian products; distance queries and geodesic routing based on hub labeling; lopsided hulls, pc-hulls and Pareto envelopes for subsets of hypercubes and other host graphs; metric dimension and identifying codes for some classes of graphs.*

Related themes: S1,S3,S4,S5,S6,A1,A4

Theme A3: Finite metric spaces: approximation and realization. Approximations and low-distortion embeddings of metric spaces into simpler metric spaces having a nice geometric structure is a fundamental (and very rich) mathematical question, having numerous algorithmic and combinatorial applications. Optimal realizations of finite metrics by network-metrics is another rich source of challenging algorithmic and structural questions with applications.

Let (X, d) be a *finite metric space.* An edge-weighted graph $G = (V, E, l)$ is called a *realization* of (X, d) if $X \subseteq V$ and for any two points $x, y \in X$ the equality $d(x, y) = d_l(x, y)$ holds, where $d_l(x, y)$ is the shortest-path distance in the graph G weighted by l. We will say that (X, d) is respectively a *tree-metric, an outerplanar-metric, a series-parallel metric, a planar-metric* if (X, d) admit a realization whose support is respectively a tree, an outerplanar graph, a series-parallel graph, a planar graph, etc. For example, K_n admits a realization as a star with $V(K_n)$ corresponding to leaves of the star and all edges of length $\frac{1}{2}$ (thus K_n belongs to all these classes). The *recognition problem* can be formulated as follows: given a class $\mathcal{M}$ of metric spaces and an input space (X, d), does (X, d) belong to $\mathcal{M}$? There are many positive algorithmic results on this problem: the recognition of ultrametrics, tree-metrics, outerplanar metrics, (Euclidean) ℓ_2-metrics, and ℓ_∞-metrics is polynomial. On the other hand, the recognition of ℓ_1-metrics is NP-complete, however the recognition of graphic ℓ_1-metrics is polynomial. The first basic (and intriguing) question is about the algorithmic recognition of series-parallel and planar metrics: *given a finite metric space* (X, d), *is* (X, d) *series-parallel or is* (X, d) *planar?*

The second type of questions concerns low-distortion embeddings of graph-metrics. Given a class $\mathcal{M}$ of host metric spaces and an input finite metric space (X, d), one can formulate several general approximation-optimization problems of the following nature: find a best approximation of (X, d) with a metric from $\mathcal{M}$. By a *best approximation* one can understand the additive or the multiplicative distortion. Formally, given two metric spaces (X, d) and (Y, d'), a map $\varphi : X \mapsto Y$ is an *embedding with multiplicative distortion* $\lambda \geq 1$ if $d(x, y) \leq d'(\varphi(x), \varphi(y)) \leq \lambda \cdot d(x, y)$ for all $x, y \in X$. Given a metric space (X, d) and a class $\mathcal{M}$ of host metric spaces, we denote by $\lambda^* := \lambda^*(X, \mathcal{M})$ the minimum distortion of an embedding of (X, d) into a member of $\mathcal{M}$. Analogously, $\varphi : X \mapsto Y$ is an *embedding with additive distortion* $\lambda \geq 0$ if $d(x, y) - \lambda \leq d'(\varphi(x), \varphi(y)) \leq d(x, y) + \lambda$ for all $x, y \in X$. In a similar way, we can define the minimum additive distortion for embedding of a metric space (X, d) into a class $\mathcal{M}$ of host metric spaces.

The paper [53] presents a large (around 100) constant factor approximation (which was improved in [52] to a factor 27) for optimal multiplicative distortion of embedding a graph metric into a tree metric. In [75], a simple factor 6 algorithm for this problem was provided. The paper [75] also presents a constant factor algorithm for approximating the optimal distortion of embedding a graph metric into an outerplanar metric. The open questions we would like to consider is a logical continuation of [75] and can be formulated in the following way: *design constant-factor* *or/and polylog-factor approximation algorithms for embedding graph-metrics (finite metrics) into series-parallel metrics or/and planar metrics with least additive or multiplicative distortion.*

A realization $G = (V, E, l)$ of a metric space (X, d) is called an *optimal realization* if the total edge length of G is minimum over all realizations of (X, d). Dress [86] showed that optimal realizations always exist. However, Althöfer [49] proved that finding such an optimal realization is NP-hard. One of the main algorithmic questions about optimal realizations is: *design an approximation algorithm for constructing optimal realizations of finite metrics and graph-metrics.*

Let $G = (V, E)$ be a (possibly weighted) graph and a set of pairs $P \in V \times V$, a subgraph $H \subset G$ is called a *pairwise distance-preserver* for (G, P) if $d_H(u, v) = d_G(u, v)$ for all $(u, v) \in P$, i.e. all distances in G between pairs in P are preserved in H. The goal is to find a pairwise distance-preserver of total minimum length. This problem is known to be NP-complete already when G is the grid graph. In this case, if P is the set of all pairs, then the problem is equivalent to the minimum *Manhattan network problem*, which was proven NP-complete in [80]. We want to *characterize polynomial cases of* (G, P) *and to find constant factor approximation algorithms for other NP-complete cases* (in [71, 78] we obtained factor 2 and 2.5 algorithms for the classical minimum Manhattan network problem in the ℓ_1-plane and normed plane with polygonal balls). For example, with P all possible pairs of a subset S of V, for *which classes of graphs the problem can be solved in polynomial time (i.e., for any graph G from the class and any subset S of V(G)?*

Objectives: *Recognition of series-parallel and planar metrics; approximation algorithms for embedding graph-metrics into series-parallel and planar metrics; approximation algorithm for constructing optimal realizations and pairwise distance-preservers.*

Related themes: S5,S6,A4,

Theme A4: Seriation and classification. As we already mentioned in the introduction, Robinson dissimilarities are the standard distance model for seriation. A *dissimilarity* on a set of objects X is just a map $d : X \times X \to \mathbb{R}^+ \cup \{0\}$ such that $d(x, y) = d(y, x)$ and $d(x, y) = 0$ iff $x = y$ (distances additionally satisfy the triangle condition). A dissimilarity d on X is called *Robinsonian* if there exists a total order $\prec$ on X such that if $x_i \prec x_j \prec x_k$ then $d(x_i, x_k) \geq \max\{d(x_i, x_j), d(x_i, x_k)\}$. This corresponds to a total ordering of the lines and columns of the distance matrix such that the distances are increasing while moving along the lines and the columns. The ultrametrics (the standard model in phylogenetics) are particular instances of Robinsonian dissimilarities. Deciding if a dissimilarity d on X is Robinsonian is equivalent to deciding if the hypergraph of balls $\mathcal{B} = \{B(x, r) : x \in X, r \in \mathbb{R}^+\}$ is an interval hypergraph and thus can be tested in polynomial time. Recently, in [47] we designed an optimal $O(n^2)$-time algorithm to recognize if d is Robinsonian ($|X| = n$) based on PQ-trees (subsequently, in [16] we also showed that one can embed the associated PQ-trees into a distributive lattice). An $O(n^2 \log n)$ algorithm based on LexBFS was designed in [113] (other algorithms of the same complexity were known before). Robinsonian structures have both strong and nice structural properties (they establish links between distances, graphs, clustering systems and lattices) and interesting applications (seriation, phylogenetic problems, knowledge representation and management, etc). On structural problems, *we will work on metrical characterizations for some classical structures. This metric approach should lead to new efficient algorithms able to work on larger data sets than usually.* We also intend to apply the tools developed in [47] and [16] to *design approximation algorithms for Robinson fitting, efficient for large data sets.* A factor 16 algorithm for the NP-hard problem of the best *additive* distortion approximation of a dissimilarity d by a Robinsonian dissimilarity $\hat{d}$ was proposed in [79]. A similar (and more difficult) problem of the *best multiplicative approximation of a dissimilarity by a Robinsonian dissimilarity is widely open* (there are no known constant factor algorithms for this problem in the much simpler case of ultrametrics). It would be interesting to investigate if repetitive applications of LexBFS can lead to faster than $O(n^2 \log n)$ (i.e., those in [113]) algorithms. Due to the importance of the problem and similarity to classical sorting, it will be interesting to see if Quicksort can be extended to a randomized $O(n^2)$ algorithm for this recognition problem. It would be also very interesting to obtain a simple optimal deterministic algorithm for this problem, perhaps based on some extension of graph search. All these algorithms will be implemented and tested. We intend to publish a library where all these methods will be available. Finally, it will be important to extend the existing efficient recognition algorithms and concepts from classical Robinsonian (binary) dissimilarities to ternary Robinson dissimilarities in the sense of [133] and to arboreal dissimilarities [69].

Objectives: *Approximation algorithms for Robinson fitting, efficient for large data sets; optimal recognition algorithms based on graph search; average behavior of near-optimal algorithms; extend recognition algorithms and structural properties to ternary Robinson and arboreal dissimilarities.*

Related themes: A2, A3.

1.6. **Risk management and methodology.** We intend to work on some difficult problems from MGT and its applications. Succeeding to solve some of them will be a great success for the project and a reward for us. However, working on this kind of questions represents a certain level of risk. We are ready to invest the necessary time to get a deep understanding of structures and properties necessary to have a chance to answer these problems. On the other hand, our own experience shows that to obtain very good results it is necessary to regularly think and work hard on difficult questions. Our results from 2016 on the existence of cores in hyperbolic graphs and a counterexample to Thiagarajan's conjecture bring some confidence in our forces and in the feasibility of the project. Several formulated problems have a pronounced combinatorial, graph-theoretical, algorithmic, combinatorial optimization, data analysis flavor and require an expertise from these areas of research. The consortium has internationally recognized experts in all those areas (see the next section and the tables). On some problems we will collaborate with internationally recognized fellows, some of them are our coauthors.

Even if the project comprises 10 different themes of research and a long list of questions, there are strong relationships between most of themes. For each of the themes, we already enumerated the related themes. We outline now some of these links as well as some differences between the themes. In some themes the object of study is the same but the perspectives and the motivation are different. For example, S2 is about Thiagarajan's conjecture for hyperbolic domains while A1 is about algorithmic problems on hyperbolic graphs. Even if these two objects are close to each other and the hyperbolicity of median graphs is well understood, the techniques used in both themes will be very different. S3 is about the existence of representation maps for lopsided concept classes and the second part of S4 is about COMs, generalizing lopsided sets. Again, the difference between S3 and S4 is that S3 requires a solution of a concrete combinatorial question (using all previous known results about lopsided classes), while S4 is about the development of a more general theory of COMs. Notice also the link between the representation maps and identifying codes in S3 and S6. Analogously, the local-to-global characterizations developed in S1 can be useful in dealing with questions about basis graphs in S4 and inspired the local-to-global characterization of representation maps S3. The themes S5 and A3 are also very close to each other: the first one is dealing with the existence of low-distortion embeddings (in the case of planar metrics and ℓ_1), while A3 is about the algorithmic computation of such embeddings (into given host spaces). Notice also that the themes (S2)-(S5) are dealing with similar objects but interpreted in completely different ways: the events of an event structure are the same as the hyperplanes of a CAT(0) cube complexes, lopsided sets, or COMs, cuts in the ℓ_1-embeddings, and wall systems in geometric group theory. The vertices of a lopsided class are called concepts in learning theory, bases in the theory of matroids, topes in the theory of oriented matroids, and configurations in the theory of event structures. The fact that all such objects lead to partial cubes show their structural richness. Each of the specific application domain provides various combinatorial and structural problems which can be treated from a common point of view.

The main questions of themes S5, S6, and A3 are about planar metrics. Understanding the structure and the intrinsic difficulties of planar metrics is one goal of our project (one can say that *planar metrics are still waiting for their Kuratowski*). In some proofs of results of theme S4 the pc-minors (partial cube minors) are used while for low-distortion embeddings into trees and outerplanar graphs (theme A3) metric minors are used. This shows that an important work must be done in developing the concepts of metric minors in different contexts, and the consortium seems to have the required expertise. Notice also that the themes and the main questions are concentrated around a few core objects and subjects: *local-to-global, median and CAT(0) cube complexes, lopsided sets, VC-dimension, and COMs, basis graphs, low distortion ℓ_1-embeddings, hitting set and packing with balls, planar metrics, hyperbolic graphs, Robinson dissimilarities.*

Concerning the methodology, we would like to mention that several difficult questions of the project are about objects which have numerous characterizations and a rich structure: CAT(0) cube complexes, lopsided sets, basis graphs of matroids, tope graphs of COMs, hyperbolic graphs. Our positive experience with hyperbolic graphs and CAT(0) cube complexes as domains of event structures shows that depending on the problem to solve, the choice of an appropriate definition of the respective object is crucial (in addition to all other additional work related to the problem). We hope that a similar scheme will work for dealing with the problems of the project about the same objects or about lopsided classes and basis graphs. On the other hand, there is no such rich structure for planar metrics, and finding it will be our challenge.

2. Project organisation and means implemented

2.1. **Coordinator.** V. Chepoi (CV in the appendix). VC is professor in CS at Aix-Marseille U. since 1998. He was awarded an AvHumboldt Fellowship in 1994. VC is the head of ACRO team of LIF. Metric Graph Theory is his main (and probably single) research topic, on which he regularly and intensively works since 1981. He obtained in MGT several foundational results (some mentioned above) and advised 5 PhD theses. He published about 112 journal publications and 22 proceedings of international conferences. Main results have been published in premium and leading journals like JTCB, Algorithmica, J. Algorithms, SIAM J. Discr. Math., SIAM J. Comput., Adv. Math., Adv. Appl. Math., Trans. AMS, Discr. Comput. Geom., Europ. J. Combin., Cybernetics, SODA, SoCG, APPROX-RANDOM. He formulated the most part of the questions in the project (some based on his current work) and he plans to devote to the project the most time of his research. He will be involved in all themes of the project.

2.2. **Consortium.** The consortium consists of 24 fellows, comprising experts in metric graph theory, algorithmics of distances, approximation algorithms, combinatorial optimization, graph theory and graph algorithms, distributed algorithms, matroids and oriented matroids, and seriation. 12 participants are from LIF, Aix-Marseille U., 4 participants are from IRIF, U. Paris-Diderot, and 8 other participants are from six French Universities: two from Bordeaux (LABRI), two from Grenoble (G-SCOP), and four from Paris (LAMSADE, LIPN), Clermond-Ferrand (LIMOS), and Montpellier (LIRMM). Due to this geographical dispersion, still having two centers and a few outliers, we decided to have *two partners, one at LIF (Marseille) and one at IRIF (Paris)*, and to assign all the participants to those two centers via the nearest neighbor principle. Victor Chepoi, the scientific coordinator of the project, will coordinate the pole in Marseille and Pierre Charbit (CV in the appendix) will coordinate the pole in Paris. Each of the two partners will ensure that the participants from outside affiliated with them (further named "isolated participants") will be fully engaged in the themes on which they are assigned in the project and that the partner will cover the travel and related expenses of isolated participants in the limits of their involvement in the project. Here is the full list of participants:

LIF (Marseille): F. Brucker (FB, 40%, Pr), J. Chalopin (JC, 30%, CR1), V. Chepoi (VC, 75%, Pr), B. Couëtoux (BC, 50%, MdC), B. Estellon (BE, 40%, MdC), K. Knauer (KK, 40%, MdC), A. Labourel (AL, 25%, MdC), K. Nouioua (KN, 50%, MdC), G. Naves (GN, 50%, MdC), P. Prea (PP, 60%, MdC), P. Valicov (PV, 40%, MdC), Y. Vaxès (YV, 50%, Pr).

IRIF (Paris): P. Charbit (PC, 40%, MdC), M. Habib (MH, 30%, Pr), R. Naserasr (RN, 30%, CR1), L. Viennot (LV, 30%, DR).

Other centers: M. Bonamy (MB, 30%, CR2) and C. Gavoille (CG, 20%, Pr), both LABRI; L. Beaudou (LB, 30%, MdC, LIMOS), N. Bousquet (NB, 30%, CR CNRS, G-SCOP) and N. Catusse (NC, 30%, MdC, G-SCOP), D. Cornaz (MC, 30%, MdC, LAMSADE), E. Gioan (EG, 40%, CR1, LIRMM), R. Grappe (RG, 40%, MdC, LIPN).

Most of participants (this concerns almost all participants from Marseille) previously did research in one or both main subjects of the project (some of them, NC, BE, KN, LB, VC completed their PHD on these topics), as well as on graph algorithms, combinatorics and graph theory. FB, PP are experts in data analysis and classification, VC, BC, NC, BE, KN, YV are also experts in approximation algorithms, RG, DC, GN in combinatorial optimization, MB, RN, PV in graph colorings, CG, AL in distributed representations of graphs, while JC, VC, EG, KK in matroids and oriented matroids. We plan to have a PhD fellowship at LIF on the theme A1 "Algorithmic aspects of hyperbolic graphs" and two postdocs, one at LIF on "Structure in metric graph theory" and the second at IRIF on "Algorithms in metric graph theory". We asked L. Viennot (DR2 Inria) to join the consortium because of his unique knowledge on algorithms and protocols for networks and his theoretical and practical work on hub labeling in road networks.

VC will coordinate themes S1 and S3, JC will coordinate S2. The theme S4 will be coordinated by EG and KK, the first on basis graphs aspects and Tutte polynomial and the second on oriented matroids and COMS. Analogously, the theme S5 will be co-directed by NB and BE. LB will coordinate the theme S6. The theme A1 will be coordinated by YV, while the theme A2 will be coordinated by CG and MH. The themes A3 and A4 will be coordinated by GN and PP, respectively. The short CVs of coodinators of all themes are presented in the appendix.

In the next table we present the involvement of members quantified in months, the list of themes on which they will be involved, and the distribution of isolated participants among the two partners (AMU stands for Aix-Marseille University and PDU for Paris Diderot University).

Partner	Affiliation	Name	First Name	Position	Involvement	Themes
IRIF	Cl.-Ferrand U., LIMOS	Beaudou	Laurent	MdC	15	S5,A2
IRIF	CNRS, LABRI	Bonamy	Marthe	MdC	15	S4,S6,A2
LIF	CNRS, G-SCOP	Bousquet	Nicolas	CR CNRS	15	S6,A1,A2,A3
LIF	ECM, LIF	Brucker	Francois	Professor	20	A4
LIF	Grenoble U., G-SCOP	Catusse	Nicolas	MdC	15	S6,A2,A3
LIF	CNRS, LIF	Chalopin	Jéremie	CR CNRS	15	S1-S5, A1,A2
IRIF	PDU, IRIF	Charbit	Pierre	MdC	20	S6,A2
LIF	AMU, LIF	Chepoi	Victor	Professor	36	Coordinator
IRIF	Dauphine U., LAMSADE	Cornaz	Deniz	MdC	15	A2,A3
LIF	AMU, LIF	Couetoux	Basile	MdC	24	S4, S6, A1,A2,A3
LIF	AMU, LIF	Estellon	Bertrard	MdC	20	S6,A1,A2
IRIF	Bordeaux U., LABRI	Gavoille	Cyril	Profesor	10	S6,A1,A2
LIF	CNRS, LIRMM	Gioan	Emeric	CR CNRS	20	S2,S4
IRIF	Paris-Nord U., LIPN	Grappe	Roland	MdC	20	S5,A2,A3
IRIF	PDU, IRIF	Habib	Michel	Professor	15	A1,A2,A4
LIF	AMU, LIF	Labourel	Arnaud	MdC	12	A1,A2
LIF	AMU, LIF	Knauer	Kolja	MdC	20	S4,A2
LIF	AMU, LIF	Nouioua	Karim	MdC	24	S6, A2,A3
IRIF	CNRS, IRIF	Naserasr	Reza	CR CNRS	12	S6,A2
LIF	AMU, LIF	Naves	Guyslain	MdC	24	S5,S6,A1,A2,A3
LIF	AMU, LIF	Prea	Pascal	MdC	30	A4
LIF	AMU, LIF	Valicov	Petru	MdC	20	S6,A2
LIF	AMU, LIF	Vaxès	Yann	Professor	24	S5,S6,A1,A2,A3
IRIF	INRIA, IRIF	Viennot	Laurent	DR INRIA	15	S5,A1,A2
LIF	AMU, LIF	XXX	XXX	PhD Student	36	A1,A2
LIF	AMU, LIF	XXX	XXX	Postdoc	12	S1-S6
IRIF	PDU, IRIF	XXX	XXX	Postdoc	12	A2-A4

2.3. **Justification of requested resources.** The project duration is 48 months. Including the PhD and the two Postdocs funded by the project, we obtain about 500 person.months for the 4 years of the project, which corresponds to more that 10 full-time researchers per year. We will use this number to evaluate further costs.

Personnel. For the realization of the project, we would like to require one PhD student, two post-doctoral students, and several master students.

PhD student: The aim of the project is a perfect fit for a PhD thesis in theoretical computer science. As specified above, the PhD student will be involved in the themes A1 and A2, on a subject "Algorithmic aspects of hyperbolic graphs". The PhD thesis will be performed at LIF. We plan to hire a PhD student quickly after the beginning of the project.

Estimated cost: 91 K€.

Post-docs: We would like to have two post-doctoral researchers, one at LIF on "Structure in metric graph theory" (on one of the themes S4, S5, or S6) and the second one at IRIF on "Algorithms in metric graph theory" (at the theme A2). Each of the post-docs will be hired for the second or third year of the project.
Estimated cost: $54,5 + 56,1$ K€=110,6K€.
Master students: Several questions in the project are easy to understand and may attract undergraduate or graduate students as the first research subject. We would like to require 8 interships for master students, for 4 months each, 4 interships for each partner.
Estimated cost: $8 \times 2,25K$€=18 K€.

Travel expenses. As for any fundamental research project, it is very important to have national and international collaborations. Therefore we need fundings for our visits and for inviting our coauthors. For the dissemination of our research results, it is important to participate in international conferences and workshops (which are also standards of Computer Science). For a better collaboration and a successful start, we plan to have three meetings of the participants of this project (which may be attended by other researchers).

Considering that the project covers one international mission or two national missions per researcher and taking into account PhD and pos-doctoral students, we require 97 K€ for travel expenses, which will be distributed evenly between the two partners. For the three meetings of the participants we will need 27 K€; this amount will be affected to each partner "au prorata". We would like to require 18 K€ for 6 months for visiting fellows (3 K€per month).
Estimated cost: 97 K€+27 K€+18 K€=142 K€.

Equipment. We will need personal computers for a part of the members of the project. We evaluate the cost of a laptop to 2000K€ and we require a total of 14 K€.
Estimated cost: 14 K€.

Conference organization. Continuing the long tradition of conferences on discrete metric spaces, we would like to organize one at CIRM, Luminy in 2020 or 2021.
Estimated cost: 5 K€.

The distribution of this budget among the two partners (taking into account the number of participants at each partner) is as follows:

	PhD	PostDoc	Travel	Equip.	LocalMeet.	Visitors	Interships	Conf.	EnvTax
LIF	91 K€	54,5 K€	64 K€	8 K€	18 K€	9 K€	9 K€	6 K€	21 K€
IRIF	0 K€	56,1 K€	33 K€	6 K€	9 K€	9 K€	9 K€	0 K€	9,6 K€

3. Impact and benefits of the project

3.1. **Scientific, economic, social or cultural impact.** Metric spaces were introduced a century ago and remain a subject of active research. The concepts of geodesic space, CAT(0) space, delta-hyperbolicity, quasi-isometry, injective hull play central roles in geometry and geometry of groups. Since the 50s, graphs endowed with metrics found an increasing number of applications. Graphs as metric objects occur in the investigation of groups, matroids, incidence geometry, etc. Many classical algorithmic problems concern distances: shortest path, center and diameter, Voronoi diagrams, TSP, clustering, etc. Algorithmic and combinatorial problems related to distances also occur in data analysis. Low-distortion embeddings into ℓ_1 (Bourgain's theorem and applications by Linial et al.) and the dimension reduction lemma were the founding tools of so-called metric methods, used in designing approximation algorithms. Other important algorithmic graph problems related to distances concern the construction of sparse subgraphs approximating inter-node distances and the converse, augmentation problems with distance constraints. In the distributed setting, an important problem is that of designing compact data structures allowing very fast computation of inter-node distances or routing along shortest or almost shortest paths. Finally, hyperbolic graphs introduced by Gromov in geometric group theory becomes a source of applications in network analysis, motivated by empirical works showing that a number of data networks, including Internet application networks, web networks, collaboration networks, social networks, and others, have small hyperbolicity.

Besides mathematics and CS, applications of structures related to distances can be found in archeology, computational biology, statistics, data analysis, etc. The problem of characterizing isometric subgraphs of hypercubes has its origin in communication theory and linguistics. In the search for a method for chronologically ordering archaeological deposits, the archeologist Robinson introduced in 1951 a distance measure which now bears his name and is the standard distance model for seriation. To take into account the recombination effect in genetic data, the mathematicians Bandelt and Dress developed in 1991 the theory of canonical decompositions of finite metric spaces. One important step in this method is building a median network that contains all Steiner (most parsimonious) trees. Median graphs occurring there constitute a central notion in MGT. Together with their algebraic and geometric counterparts, they have many nice properties and numerous characterizations. These structures have been investigated in several contexts independently for more than half a century. What is surprising is that they coincide with objects from completely different domains: CAT(0) cube complexes from geometric group theory (pure mathematics) and domains of event structures from concurrency (theoretical computer science), and that these bijections are essential in solving open problems on event structures and concurrent automata. Lopsided sets investigated in MGT and generalizing median structures have recently found applications in computational learning theory, in designing compression schemes for concept classes. Lopsided sets are also particular COMs (complexes of oriented matroids), related to structures investigated in discrete geometry and combinatorics.

This shows that MGT is a research domain common to mathematics and CS, with strong and deep connections to many applied areas and domains of fundamental research. Our research project concerns all previously mentioned themes and applications of metric spaces and graphs as metric objects. In view of the transversal nature of metric spaces and their applications, we believe that our research and results on this project will have the same level of impact.

3.2. **Describe how the results address an ANR 2017 Work Programme challenge.** We believe that our project fits best into DEFI 7, devoted to CS, because: (i) all participants are affiliated to CS laboratories, (ii) our research subjects are directly connected to topics traditionally developed in CS: algorithms, combinatorics, graph theory, combinatorial optimization, concurrency, computational learning theory, data analysis.

3.3. **Dissemination strategy.** Continuing the long tradition of conferences on discrete metric spaces, we would like to organize one at CIRM, Luminy. We will also organize several meetings of the participants, which may be attended by other researchers. The main methods of dissemination of our results remain: (i) publishing them in major CS and mathematical journals and CS conferences, (ii) presenting them as invited or contributed speakers at various high-quality mathematical conferences and workshops, (iii) giving lectures and seminars at universities, in France and worldwide. The fellowships for Master students as well as the PhD and postdocs is yet another way of spreading the results.

3.4. **Involvement in other projects.** We indicate here the recent projects in which the members of our consortium are or have been involved:

Years	Coordinator	ANR Project	Name	Person.Month
2006–2010	G. Cornuejols	OPTICOMB ANR-06-BLAN-0375	V. Chepoi, Y. Vaxès, P. Préa, K. Nouioua	24 each
2011–2014	J. Ramírez Alfonsín	TEOMATRO ANR-10-BLAN 0207	V. Chepoi, E. Gioan	24 each
2011–2015	C. Druţu	GGAA ANR-10-BLAN-0116	V. Chepoi Y. Vaxès	24 12
2013–2017	J. Chalopin	MACARON ANR JCJC	J. Chalopin A. Labourel	36 28
2015–2018	S. Das/M. Mihalak	ANCOR ANR/SNF	J. Chalopin A. Labourel	9 14