



Single-cell 3D genome reconstruction in the haploid setting using rigidity theory

Sean Dewar¹ · Georg Grasegger² · Kaie Kubjas³ ·
Fatemeh Mohammadi⁴ · Anthony Nixon⁵

Received: 24 July 2024 / Revised: 20 January 2025 / Accepted: 21 February 2025 /

Published online: 29 March 2025

© The Author(s) 2025

Abstract

This article considers the problem of 3-dimensional genome reconstruction for single-cell data, and the uniqueness of such reconstructions in the setting of haploid organisms. We consider multiple graph models as representations of this problem, and use techniques from graph rigidity theory to determine identifiability. Biologically, our models come from Hi-C data, microscopy data, and combinations thereof. Mathematically, we use unit ball and sphere packing models, as well as models consisting of distance and inequality constraints. In each setting, we describe and/or derive new results on realisability and uniqueness. We then propose a 3D reconstruction method based on semidefinite programming and apply it to synthetic and real data sets using our models.

Keywords 3D genome reconstruction · Hi-C · Rigidity · Semidefinite programming

Mathematics Subject Classification 92E10 · 92E10 · 52C25

1 Introduction

The 3-dimensional (3D) structure of the genome plays an important role in gene regulation (Dekker 2008; Uhler and Shivashankar 2017) and genome misfolding is linked to disease (Norton and Phillips-Cremins 2017). Two different approaches for inferring the

✉ Kaie Kubjas
kaie.kubjas@aalto.fi

¹ School of Mathematics, University of Bristol, Bristol, UK

² Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Linz, Austria

³ Department of Mathematics and Systems Analysis, Aalto University, Espoo, Finland

⁴ Departments of Mathematics and Computer Science, KU Leuven, Leuven, Belgium

⁵ Mathematics and Statistics, Lancaster University, Lancaster, UK

3D genome structure are based on chromosome conformation capture techniques and microscopy-based techniques. The primary microscopy-based technique has been the fluorescent in situ hybridisation (FISH) (Amann et al. 1990). However, this is limited in resolution and restricted to a region of the genome. Recently, in situ genome sequencing (IGS) (Payne et al. 2021), which combines sequencing with imaging techniques, has enabled genome-wide high-resolution 3D genome reconstructions. Currently the availability of IGS data is limited compared to data from chromosome conformation capture experiments (such as 3C, 4C, 5C, Hi-C, ChiA-PET, etc), which record interactions between different fragments of the genome. One of the most popular chromosome conformation capture techniques is Hi-C (Lieberman-Aiden et al. 2009) that records interactions between fragments of a genome on a genome-wide scale. The output of a Hi-C experiment is a Hi-C or a contact matrix, where rows and columns correspond to fragments of the genome and the entries of the matrix record the number of interactions between the fragments. The resolution of Hi-C data is the number of base pairs in a genome fragment.

A recent review of Oluwadare et al. (2019) summarises over 30 different approaches for constructing the 3D genome from contact matrices. The approaches can be divided on the one hand into distance-based, count-based, and probability-based depending on how the interaction frequencies are modelled, and on the other hand into consensus, ensemble and population methods based on the structure of the output (Oluwadare et al. 2019, Figure 2). Distance-based methods first turn contact counts into distances and then infer the positions of loci from the pairwise distances (Zhang et al. 2013; Belyaeva et al. 2022). Different distance-based methods vary in how the interaction frequencies are converted into distances and how 3D positions of loci are inferred from the distances. Contact-based methods directly infer the 3D structure without first turning contacts into distances (Paulsen et al. 2017; Abbas et al. 2019). Probability-based methods model interaction frequencies as random variables, and apply maximum likelihood or Bayesian inference based methods for 3D genome reconstruction (Hu et al. 2013; Varoquaux et al. 2014).

So far, the main focus of 3D genome reconstruction has been on population data, where the contact matrix counts the contacts in a collection of cells. Since the 3D structure can differ in different cells, one may use population data to either infer a single structure that represents the “average” of structures or an ensemble of structures that are consistent with the data. The introduction of single-cell Hi-C by Nagano et al. (2013) has enabled the study of 3D genome structure at the single-cell level. Further single cell Hi-C methods were introduced for haploid organisms by Ramani et al. (2017); Stevens et al. (2017) and for diploid organisms by Tan et al. (2018). Minimization of polymer models for 3D genome reconstruction is used by Nagano et al. (2013); Stevens et al. (2017); Wettermann et al. (2020); Shi and Thirumalai (2021); Kos et al. (2021), Bayesian inference by Rosenthal et al. (2019), and manifold optimization by Paulsen et al. (2015). ChromSDE (Zhang et al. 2013) is a semidefinite optimization based method that was developed for population data, but is also applicable to single cell data. ShRec3D (Lesne et al. 2014) combines shortest path computations with multidimensional scaling for 3D genome reconstruction and this works both for single cell and population Hi-C data.

Our contribution In this paper, we study 3D genome reconstruction from single-cell Hi-C data for haploid organisms. Our main contributions to the abundant literature on this topic are establishing a connection between the 3D genome reconstruction and the mathematics of rigidity theory and the study of uniqueness of 3D constructions.

Rigidity theory is a field of mathematics that investigates whether a point configuration is uniquely determined, up to rigid transformations, from some pairwise distances between the points. One can study the rigidity of a point configuration under various different assumptions. We propose mathematical models associated to penny graphs, unit ball graphs, or some modifications of them (these graphs are defined formally in the sections that follow) that correspond to different biological methods of measuring single-cell data. In each of these mathematical models, we survey known results and derive new results on realisability and uniqueness. Finally, we apply semidefinite programming to obtain 3D genome reconstruction algorithms for three of the proposed models to compare them and analyse whether the reconstruction results are consistent with the uniqueness results.

It is, perhaps, more desirable to investigate 3D genome reconstruction for single-cell data in the diploid setting. However, the reconstruction of the 3D genome from contact matrices in the diploid setting for population data poses numerous challenges (Segal 2022). To the best of our knowledge, the uniqueness of reconstructions for single-cell data has not been studied even in the haploid setting. In this sense, our work serves as a bridge towards establishing a similar theory in the diploid setting.

Structure of the paper. In Sect. 2, we establish our notation and provide the necessary background from graph rigidity theory, along with a concise overview of the problem of 3D genome reconstruction for single-cell data in the haploid setting. The primary objectives of this section are to establish our notation and offer a brief motivation for investigating the rigidity of the proposed families of graphs from a biological perspective.

Sections 3–6 present our models. Each subsequent section introduces additional constraints to the previous models and compares the new model with those presented in the preceding sections. In Sect. 3 we use unit ball graphs to represent the threshold model, discuss realisability in this model and show that uniqueness cannot be achieved without additional constraints being imposed. Section 4 then adds microscopy constraints. The main contributions of this section include a proof that the realisability question can be solved directly by semidefinite programming if we allow the realisation to live in an arbitrary dimension and a proof that inequalities alone do not determine whether there is a finite number of reconstructions. We also consider the case of equality constraints only, that is, where the data comes only from microscopy experiments. This situation turns out to be equivalent to the well-studied problem of global rigidity for bar-joint frameworks. In order to be able to mathematically deduce unique reconstructions, we then consider more specialised models in Sects. 5 and 6. In particular, in Sect. 6, we prove theoretical results, especially about uniqueness and the structure of the possible graphs that could arise from the biological data corresponding to this model.

In Sect. 7, we adapt 3D reconstruction algorithms based on semidefinite programming previously used by Zhang et al. (2013); Belyaeva et al. (2022) to three of the

models considered in this paper. We apply this algorithm to synthetic and real data sets, and compare our results with ShRec3D (Lesne et al. 2014).

2 Preliminaries

We first review the basic theory of rigidity for structures comprising of universal joints linked by stiff bars. Then we define a number of biological models and point to the subsequent sections, where we analyse these models using rigidity theory.

2.1 An introduction to graph rigidity

One of the primary contributions of this paper lies in its application of concepts and tools from rigidity theory to gain insights into the 3D genome reconstruction of single cells. In this section, we provide a concise and accessible description of the topic of rigidity theory.

A (*bar-joint*) *framework* (G, ρ) is the combination of a finite, simple graph $G = (V, E)$ and a realisation $\rho : V \rightarrow \mathbb{R}^d$. The realisation assigns positions to the vertices (represented by universal joints with full rotational freedom) and hence lengths to the edges (unbendable straight line segments).

We are interested in understanding the set of vectors $\rho' \in \mathbb{R}^{d|V|}$ such that the two frameworks (G, ρ) and (G, ρ') have the same edge lengths. Define the rigidity map $f_G : \mathbb{R}^{d|V|} \rightarrow \mathbb{R}^{|E|}$ by putting $f_G(\rho) = (\dots, \|\rho(v_i) - \rho(v_j)\|^2, \dots)_{v_i v_j \in E}$. Then two frameworks (G, ρ) and (G, ρ') are called *equivalent* if $f_G(\rho) = f_G(\rho')$. Instantly, one sees that this set of equivalent frameworks is always infinite since ρ' may be obtainable from ρ by a composition of Euclidean isometries (translations, rotations and reflections). Any such ρ' is now said to be *congruent* to ρ . This is equivalent to specifying that all distances between pairs $\rho(v_i)$ and $\rho(v_j)$ are the same as the distances between the corresponding pairs $\rho'(v_i)$ and $\rho'(v_j)$.

After discarding isometries, we ask whether there are still infinitely many ρ' with the aforementioned property, or equivalently, whether (G, ρ) is *flexible*. If not, then there are finitely many frameworks (G, ρ') equivalent to (G, ρ) (modulo isometries), and (G, ρ) is said to be *rigid*. Furthermore, (G, ρ) is *minimally rigid* if $(G - e, \rho)$ is flexible for all $e \in E$. See Fig. 1 for small examples.

If the framework (G, ρ) is flexible, then one of several equivalent definitions of flexibility (Asimow and Roth 1978) is in terms of motions. A (*continuous*) *motion* of (G, ρ) is a family of continuous functions $x : (0, 1) \times V \rightarrow \mathbb{R}^d$ such that $x(0) = \rho$ and $(G, x(t))$ is equivalent to (G, ρ) for all $t \in (0, 1)$. Then (G, ρ) is flexible if and only if there exists a motion that does not arise from isometries of \mathbb{R}^d , i.e. if at least one pair of vertices of G has a different distance between them in (G, ρ) and in $(G, x(t))$ for some $t \neq 0$. As is standard in the literature, we consider an idealised model where we permit edges to pass through each other in such a continuous motion.

Among rigid graphs, one may ask how many different ways the structure can be realised. This is the global rigidity question, and we say that (G, ρ) is *globally rigid* if no other framework (G, ρ') (up to Euclidean isometries) with the same edge lengths

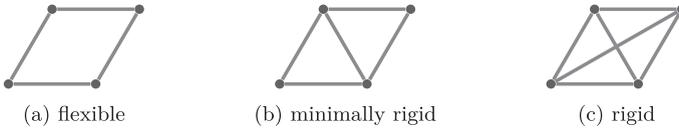


Fig. 1 Graphs with different rigidity properties

exists. In other words, (G, ρ) is globally rigid if every framework in the same dimension as (G, ρ) that is equivalent to it is also congruent to it.

In each of the above definitions, it is clear that the dimension is of crucial importance. For example, every cycle C_n is globally rigid (and hence rigid) on the line, but for all $n \geq 4$, the cycle C_n is flexible in \mathbb{R}^d for all $d \geq 2$. This motivates us to call a framework d -rigid (resp. globally d -rigid) if it is rigid (resp. globally rigid) in \mathbb{R}^d .

Later, we need the following elementary lemma, which has been presented for example by Jackson and Jordán (2005a, Lemma 2.6), that relates vertex connectivity to d -rigidity.

Lemma 2.1 *Let (G, ρ) be d -rigid on at least $d + 1$ vertices. Then G is d -connected.*

A common technique in rigidity theory is to linearise by considering the Jacobian derivative $df_G|_\rho$ of the rigidity map. For consistency with the literature, we put $R(G, \rho) = \frac{1}{2}df_G|_\rho$ and refer to $R(G, \rho)$ as the rigidity matrix of (G, ρ) . Then, the framework (G, ρ) in \mathbb{R}^d is d -independent (in the literature, this is sometimes referred to as d -stress-free) if the rows of $R(G, \rho)$ are linearly independent.

Lemma 2.2 (Maxwell 1864) *Let (G, ρ) be a d -independent graph on at least d vertices. Then $|E| \leq d|V| - \binom{d+1}{2}$ and if $|E| = d|V| - \binom{d+1}{2}$ then G is d -rigid.*

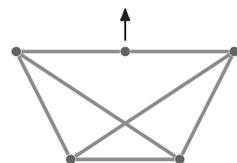
We say that a framework (G, ρ) in \mathbb{R}^d is infinitesimally flexible if there exists a function $\dot{\rho} : V \rightarrow \mathbb{R}^d$ such that

$$\langle \rho(v_i) - \rho(v_j), \dot{\rho}(v_i) - \dot{\rho}(v_j) \rangle = 0 \text{ for all } v_i v_j \in E.$$

An infinitesimal motion $\dot{\rho}$ is trivial if there exists a skew-symmetric matrix S and a vector t such that $\dot{\rho}(v_i) = Sp(v_i) + t$ for all $v_i \in V$. The framework (G, ρ) is infinitesimally rigid if every infinitesimal motion of (G, ρ) is trivial. Equivalently, (G, ρ) is infinitesimally rigid if G is complete on at most $d + 1$ vertices or $|V| \geq d + 2$ and $\text{rank}(R(G, \rho)) = d|V| - \binom{d+1}{2}$.

The rigidity (and global rigidity) of a framework depends on both the underlying graph, and the choice of realisation. See Fig. 2 for an example. However, if we

Fig. 2 A graph which would be infinitesimally rigid if realised as a generic framework, but the chosen framework has a non-trivial infinitesimal motion (as indicated)



restrict attention to generic frameworks, then this complication disappears. A framework (G, ρ) is *generic* if the set of coordinates of ρ forms an algebraically independent set over \mathbb{Q} .

Lemma 2.3 (Asimow and Roth 1978) *Let G be a graph on n vertices. A generic framework (G, ρ) is d -rigid if and only if $n \leq d$ and G is complete or $n \geq d + 1$ and $\text{rank}(R(G, \rho)) = d|V| - \binom{d+1}{2}$.*

It follows from Lemma 2.3 (resp. Gortler et al. 2010) that if (G, ρ) is a generic and d -rigid framework (resp. globally d -rigid framework) then every generic framework (G, ρ') is rigid (resp. globally d -rigid). In other words, in the generic case, rigidity depends only on the graph. Thus, we can define d -rigid and globally d -rigid for graphs (meaning there exists a generic framework that is d -rigid or globally d -rigid).

While genericity is a strong mathematical assumption, it is worth noting that almost all realisations ρ of a given graph yield generic frameworks. Therefore, we can always approximate a given realisation ρ' by a generic realisation ρ within a sufficiently small open neighbourhood of ρ' . Moreover in practical applications noisy data is likely to be generic.

In the generic case, there exist graph-theoretic characterisations of rigidity and global rigidity in the Euclidean plane, as demonstrated by Pollaczek-Geiringer (1927) and Jackson and Jordán (2005b). These characterisations have paved the way for efficient deterministic algorithms, such as network flow algorithms, to test rigidity (Jacobs and Hendrickson 1997). However, extending these results to three dimensions remains a challenging open problem.

2.2 From contacts to distances

In this paper, we study 3D genome reconstruction from Hi-C or contact matrices. These matrices are obtained from Hi-C experiments. The first step (cross-linking) of a Hi-C experiment is to freeze interactions between close DNA segments by chemically linking nearby chromatin segments within the same nucleus spatially. After cross-linking, the DNA segments are cut into fragments (fragmentation) using a restriction enzyme and then ligated together (ligation). This step forms the hybrid DNA molecules representing interactions between different genomic regions. Then, the DNA fragments are purified from contaminants in reverse cross-linking to facilitate DNA sequencing. Finally, using short sequences from both ends of a ligated segment, the segment is matched to a pair of interacting chromosome regions. All such interactions are recorded in a contact matrix. The rows and columns of a contact matrix correspond to regions of the genome, and the entries count the number of interactions between the corresponding regions in a Hi-C experiment.

A Hi-C matrix can record interactions either in a population of cells or in a single cell. In this paper, we focus on the single-cell setting for haploid organisms. The single-cell contact count matrix for a haploid organism is a 0/1-matrix recording whether there is an interaction between two genomic regions. The main problem studied in this paper is the following:

Table 1 Biological (Σ) and mathematical (\sqrt{x}) questions studied in this paper

Existence	Σ	Given a Hi-C matrix and possibly some pairwise distances, is there a 3D genome structure with the given Hi-C matrix and distances?
	\sqrt{x}	Given a graph and possibly some pairwise distances between its vertices, is there a realisation of the graph in a model?
Uniqueness	Σ	Given the Hi-C matrix of a 3D genome structure and possibly some pairwise distances, is there a unique 3D genome structure with the given Hi-C matrix and distances?
	\sqrt{x}	Given a graph in a model and possibly some pairwise distances between its vertices, is there a unique realisation of the graph in the model?
Uniqueness for fixed edge lengths	Σ	Given the Hi-C matrix of a 3D genome structure and all the pairwise distances corresponding to the ones in the Hi-C matrix, is there a unique 3D genome structure?
	\sqrt{x}	Given a graph in a model together with its edge lengths, does it have a unique realisation in the model?
Reconstruction algorithm	Σ	Reconstruct a 3D genome structure from a Hi-C matrix and possibly some pairwise distances.
	\sqrt{x}	Find a realisation of a graph in a model given possibly some pairwise distances between its vertices.

Problem 2.4 *Given a single-cell Hi-C matrix and possibly some pairwise distances between genomic loci for a haploid organism, study the 3D genome reconstruction from this data. The different aspects of the 3D genome reconstruction that we study are given in Table 1.*

We use distance-based methods to study Problem 2.4, which means that we associate distance constraints to a contact matrix. To this end, we model the genome as a string of n beads. The beads correspond to genomic loci in a Hi-C experiment. The positions of the beads are recorded by a matrix $X = [x_1, \dots, x_n]^T \in \mathbb{R}^{n \times d}$. We are primarily interested in the real-life setting when $d = 3$. However, we also consider other values of d when obtaining complete results for $d = 3$ is not feasible. Specifically, we explore less-complex models where the genome is considered to lie in a 1- or 2-dimensional space.

In the subsequent sections, we study Problem 2.4 across various models that incorporate distance constraints alongside the information given by a contact matrix. These models are summarised in Table 2. In the initial model, an interaction between two loci indicates that the beads corresponding to those loci are separated by a distance no greater than a threshold value d_c . Conversely, the absence of an interaction suggests that the beads are more than d_c units apart. This particular model, which we call the threshold model, was introduced by Trieu and Cheng (2014). While it is the most general model, we see that 3D reconstructions are not uniquely identifiable within this framework, which motivates the study of alternative models. The second assumption, that the absence of an interaction suggests that the beads are more than d_c units

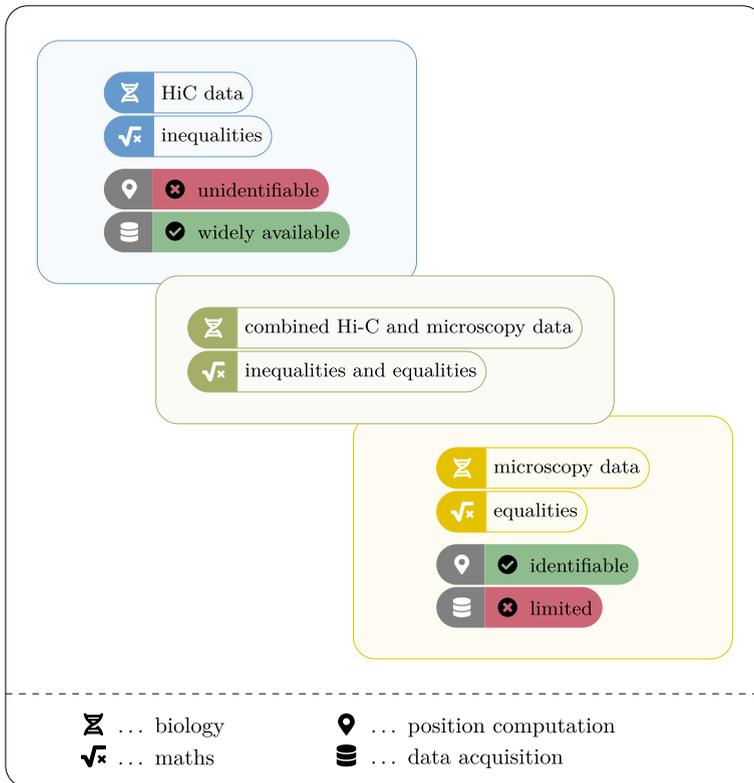


Fig. 3 A schematic overview depicting the two main approaches to obtaining biological data, their mathematical counterparts and the advantages ✓ and disadvantages ✗

apart, in this and other models corresponds to the idealized setting when all contacts are sampled in an Hi-C experiment. In reality, only a small set of actual contacts is sampled, so even if a contact is not observed, then it can be present. We note that if a model is not identifiable with the second assumption and has at least one solution, then it is also not identifiable without any constraints for non-contacts.

In the second model, we assume that in addition to the thresholds, we have knowledge of pairwise distances between some of the beads. This can occur, for example, in the presence of microscopy data or when estimating distances between neighbouring beads. When a sufficient number of pairwise distances are available, the 3D reconstructions become identifiable under this model. However, it may not always be possible to obtain a sufficient number of pairwise distances, which motivates the exploration of two additional models.

In the third model, an interaction between two loci implies that the distance between the corresponding beads precisely equals the threshold value d_c , as opposed to being at most d_c in the threshold model. Similar to the previous models, the absence of an interaction indicates that the beads are separated by a distance greater than d_c .

Table 2 The different models considered in this paper

Biology	Maths	Edges	Non-edges
Threshold	Unit ball	\leq	$>$
Threshold & some distances	Unit ball & some distances	\leq & some $=$	$>$ & some $=$
Same distance	Penny/marble	$= 1$	$>$
Approx. same distance	Generic radii penny	$= 1 \pm \varepsilon$	$>$

Table 3 Theoretical results for the different models and questions

Model	Existence	Uniqueness	Uniqueness for fixed edge lengths
Unit ball	Lemma 3.2 (dim 1)	Lemma 3.4 (no)	Lemma 3.5
	Figure 5 (dim 2)		Theorem 3.6
	Lemma 3.3 (dim 3)		Lemma 4.6
Unit ball & Some distances	–	Proposition 4.2	Lemma 4.4 Theorem 4.5
Penny/marble	NP-hard Breu and Kirkpatrick (1996); Hliněný (1997)	Lemma 5.5	–
Interval radii penny	Lemma 6.2	?	?

While this model is more restrictive than the threshold model, an advantage is that the restriction allows one to get unique 3D reconstructions in some cases.

The fourth model introduces genericity to the third one. More specifically, it assumes that an interaction between two loci implies that the distance between the corresponding beads is within an interval around the threshold d_c . The absence of an interaction implies a certain lower bound on the distance between the corresponding beads, which is made explicit in Sect. 6. Additionally, both the third and fourth models can be enhanced by incorporating the extra assumption that distances between certain beads are known. Utilising all available information is consistently beneficial in all these models.

Most of these models correspond to known frameworks in the rigidity theory. The correspondence between the biological and mathematical models is shown in Table 2. It is important to note, however, that while rigidity theory primarily focuses on edge lengths, our setting introduces additional complexity with constraints on missing edges, resulting in specific inequalities. These additional constraints make our problems more challenging and distinct from the classical rigidity theory.

Table 3 summarises the results on the existence and uniqueness of solutions in the different models referring to the respective theorems or literature.

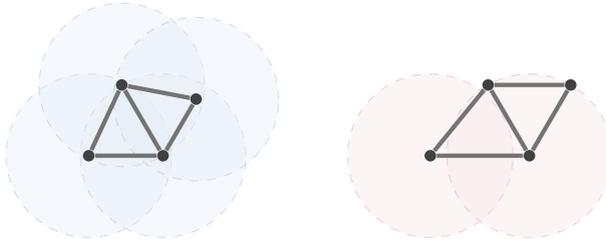


Fig. 4 A unit disk graph with a unit disk realisation (left) and another realisation which does not satisfy the unit disk condition (right)

3 Threshold model—unit ball graphs

We begin by examining a model introduced by Trieu and Cheng (2014), which defines a contact between two loci if their distance is below a threshold value of d_c . Conversely, a pair of loci is classified as non-contact if their distance exceeds the threshold. As the concepts we explore remain invariant under scalings, we can assume without loss of generality that $d_c = 1$. In this particular model, a value of one in the Hi-C matrix indicates that the corresponding distance is smaller than one, while a zero represents a distance greater than one.

Definition 3.1 (Garamvolgyi and Jordán 2020) A graph $G = (V, E)$ is a (d -dimensional) *unit ball graph* if there exists a realisation $\rho : V \rightarrow \mathbb{R}^d$ such that $\|\rho(v) - \rho(w)\| \leq 1$ if and only if $vw \in E$. In such cases, the realisation ρ is called a (d -dimensional) *unit ball realisation* of G and the pair (G, ρ) is called a (d -dimensional) *unit ball framework*. A 1-dimensional unit ball graph is also called a *unit interval graph*, and a 2-dimensional unit ball graph is also called a *unit disk graph*. See Fig. 4 for an example.

3.1 Realisability

Biologically the key interest is in 3D realisations however mathematically we also discuss realisations in lower dimensional spaces where there is already rich theory. For example in the literature, the realisation problem has been studied and solved for the case when $d = 1$. Specifically, a graph is a unit interval graph if and only if it is an *indifference graph* (for a formal definition see Roberts 1969).

A *clique* X in a graph G is a complete subgraph of G . The *clique graph* of G is another graph H in which the vertices correspond to the maximal cliques of G , and two vertices are adjacent if the corresponding maximal cliques intersect in at least one vertex.

Lemma 3.2 (Lekkerkerker and Boland 1962) *A graph G is a unit interval graph if and only if it is chordal, and every connected component of the clique graph of G is a path.*

While such graphs are easy to understand, when we move up to dimension 2, the problem becomes NP-hard, and only partial results are available. We refer to

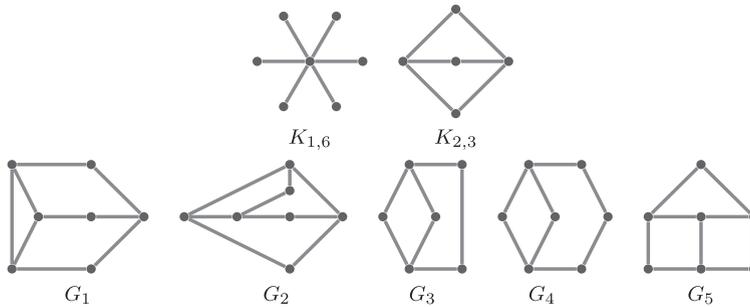


Fig. 5 Some known minimal forbidden subgraphs in dimension two. The existence of such a subgraph prevents a graph from having a unit-disc realisation

Breu and Kirkpatrick (1998) for a detailed analysis of the complexity and to Atminas and Zamaraev (2018) for some forbidden subgraphs. In particular, the class of unit disk graphs is *hereditary*, indicating closure under vertex deletions. Consequently, it raises the question of characterising unit disk graphs based on their minimal forbidden induced subgraphs. The known minimal forbidden subgraphs in dimension 2 are the infinite families described by Atminas and Zamaraev (2018) along with the 7 graphs depicted in Fig. 5.

For $d \geq 3$, the problem of determining whether a graph is a unit ball graph remains NP-hard (Breu and Kirkpatrick 1998). However, to the best of our knowledge, there is essentially no published literature regarding forbidden subgraphs in this context.

Lemma 3.3 *The complete bipartite graph $K_{1,13}$ (with parts of size 1 and 13) is a minimal forbidden subgraph of the family of unit ball graphs in dimension 3.*

Proof The kissing number in dimension three is 12 (see, e.g. Schütte and van der Waerden 1952). In fact, the spheres can be suitably arranged (e.g. in form of an icosahedron) such that they only touch the central one (Pfender and Ziegler 2004), as seen in Fig. 6.¹ Regardless, this shows that $K_{1,12}$ is a unit ball graph but $K_{1,13}$ is not. As the only proper connected subgraphs of $K_{1,13}$ with any edges are isomorphic to $K_{1,12}$, it follows that $K_{1,13}$ is a minimal forbidden subgraph. \square

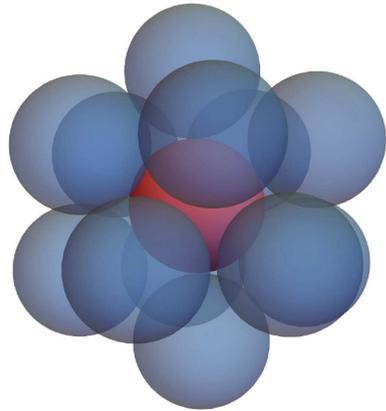
We expect that there are a large number of further minimal forbidden subgraphs, some that arise from extending the examples in Fig. 5, as well as various infinite families (for example extending those given by Atminas and Zamaraev (2018)).

3.2 Uniqueness

In this subsection, we explore the uniqueness of reconstructions derived from the data within the Hi-C matrix. The results here apply in any dimension. Within this model, the Hi-C matrix corresponds precisely to the well-known adjacency matrix, which in turn

¹ Due to this, it was not clear for a long time that the kissing number is indeed 12. This was a famous source of disagreement between Isaac Newton and David Gregory. While Newton believed that the kissing number in 3D is 12, Gregory (incorrectly) believed that the number was in fact 13.

Fig. 6 A central sphere with 12 spheres touching it but none of the outer spheres touch each other



uniquely determines the underlying graph. Consequently, there exists prior theoretical knowledge, particularly when $d = 1$. However, before delving into the details, we first establish a negative result.

Lemma 3.4 *Given a d -dimensional unit ball graph G , there are infinitely many d -dimensional unit ball realisations of G modulo isometries and scalings.*

Proof Let G be a d -dimensional unit ball graph and let ρ be a realisation of G . If none of the edges of ρ have length equal to one, then any small enough perturbation of the realisation ρ is also a unit d -ball realisation of G . If some of the edge lengths of ρ are equal to one, then we first scale the realisation by $(1 - \varepsilon)$. For a small enough $\varepsilon > 0$, this scaling does not affect the number of d -dimensional unit ball realisations. Following this scaling, we may perturb as above. \square

The above lemma indicates that we cannot obtain uniqueness of 3D reconstructions unless we pose further constraints. In the following sections, we explore what kind of additional constraints can be added to obtain uniqueness of 3D reconstructions. Before doing that, we briefly describe results from Garamvolgyi and Jordán (2020), where they fix the edge lengths and from that are able to determine rigidity and global rigidity for some families of unit ball graphs.

A d -dimensional unit ball framework (G, ρ) is *unit ball globally rigid* in \mathbb{R}^d if whenever (G, ρ') is an equivalent unit ball realisation of G in \mathbb{R}^d , then (G, ρ') is congruent to (G, ρ) .

Lemma 3.5 (Garamvolgyi and Jordán 2020, Lemma 3.1) *A generic unit ball globally rigid framework (G, ρ) in \mathbb{R}^d is rigid.*

A minimally d -rigid graph is *special* if every proper subgraph that is d -rigid is a complete graph. Using this definition the following result gives infinite families of unit ball globally rigid graphs when $d = 2$ and when $d = 3$.

Theorem 3.6 (Lemma 5.1, Theorem 4.6 and Subsection 7.3 Garamvolgyi and Jordán 2020) *Let G be a graph.*

- (i) If G is a special minimally 2-rigid graph, then G has a generic realisation in \mathbb{R}^2 that is unit ball global.
- (ii) If G is a 4-connected maximal planar graph that is also a unit ball graph in \mathbb{R}^3 , then G has a generic realisation in \mathbb{R}^3 as a globally rigid unit ball graph.

4 Threshold and microscopy—inequalities and some equalities

Our next model considers both inequality and equality constraints. We first discuss realisability and then discuss the usefulness of inequality constraints for reducing the number of realisations. The section concludes with a brief analysis of uniqueness in the special case when all the constraints are equalities.

4.1 Realisability

Let us, momentarily, ignore the issue of dimensionality for the solutions to our constraint systems. This allows us to use semidefinite programming to solve the realisability question.

Proposition 4.1 *Let $G = (V, E)$ be a graph, let $\lambda : E \rightarrow \mathbb{R}_{\geq 0}$, and let $s : E \rightarrow \{-1, 0, +1\}$. Then determining if there exists a realisation ρ of G in $\mathbb{R}^{|V|-1}$ such that*

$$\|\rho(v) - \rho(w)\|^2 \begin{cases} \leq \lambda(vw) & \text{if } s(vw) = -1, \\ = \lambda(vw) & \text{if } s(vw) = 0, \\ \geq \lambda(vw) & \text{if } s(vw) = +1 \end{cases}$$

can be achieved using a semidefinite program.

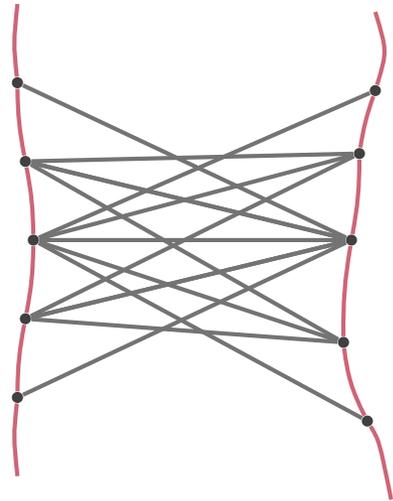
Proof Label the vertices of G as $\{1, \dots, n\}$. Let \mathcal{M} be the closed convex set of all $n \times n$ symmetric matrices M where $M_{ii} = 0$ for all $1 \leq i \leq n$, and

$$M_{ij} \begin{cases} \leq \lambda(ij) & \text{if } s(ij) = -1, \\ = \lambda(ij) & \text{if } s(ij) = 0, \\ \geq \lambda(ij) & \text{if } s(ij) = +1 \end{cases}$$

for each edge $ij \in E$. Let π be the linear map from the space of $n \times n$ symmetric matrices to the space of $(n - 1) \times (n - 1)$ symmetric matrices where $\pi(M)_{ij} = M_{in} + M_{jn} - M_{ij}$ for each $1 \leq i, j \leq n - 1$. The set $\pi(\mathcal{M})$ is now a closed convex set. By Schoenberg (1935, Theorem 1), there exists a realisation ρ of G in $\mathbb{R}^{|V|-1}$ satisfying our desired inequalities if and only if there exists a positive semidefinite matrix $A \in \pi(\mathcal{M})$. Hence, determining the existence of a suitable ρ is equivalent to finding a feasible solution to a semidefinite program. □

The two main issues with Proposition 4.1 are: (i) it does not allow for strict inequalities, and (ii) it does not guarantee the dimension of the realisation which is found.

Fig. 7 Two backbones with additional edges corresponding to contacts between pairs of loci. Lengths of the red edges are assumed to be known while grey edges are only known to be smaller than a threshold



The first issue can be (partially) resolved by replacing all strict inequalities with non-strict inequalities with additional epsilon terms. The second issue is significantly more difficult to deal with. The problem lies in that any equality/inequality for the dimension corresponds to a non-convex polynomial objective function for the optimisation problem. One method for forcing the dimension of realisations to be contained within the desired dimension (e.g., 3-dimensional space) is to solve the semidefinite program given in Proposition 4.1, and then ‘project’ the realisation into the correct dimension. Although this solves the issue of dimension, it can create issues with the required edge-length bounds for the realisation. A follow-up local optimisation can then be performed using gradient descent to improve the output and (if the bounding constraints are not met) obtain a better approximation of a true solution. We analyse one variant of this method in more detail in Sect. 7.1.

4.2 Uniqueness

Due to the non-uniqueness of 3D reconstructions under the unit ball model, we explore the additional constraints that lead to a finite number or unique reconstructions. The unit ball graph model is characterised by a specific set of distance inequalities for edges. In this section, we extend our models, by allowing more general distance inequalities and incorporating equalities defined by distance constraints.

One case in which distance constraints arise is when we assume that the edge lengths between neighbouring beads are known, as depicted in Fig. 7. We call the graph consisting of all beads on a chromosome and edges between neighbouring beads a *backbone*. These edge lengths can be estimated based on the length of a chromosome and the resolution of bins that correspond to individual beads. This assumption was made by Belyaeva et al. (2022).

Another situation that provides edge lengths arises when in addition to Hi-C data, we also have microscopy data, e.g. from Fluorescence In Situ Hybridisation

(FISH) (Amann et al. 1990) experiments. While FISH is limited to studying a specific region of genome, these two types of data (Hi-C and microscopy data) complement each other and, in certain cases, can yield unique reconstructions. This approach was proposed by Abbas et al. (2019).

Proposition 4.2 *Let $X = \{x_1, \dots, x_n\}$ be a generic set of points in \mathbb{R}^d that satisfies a set of distance constraints and distance inequalities between pairs of points. Further, suppose that no distance inequality is satisfied as an equality, e.g. if our distance inequality states that $\|x_i - x_j\| \leq r$ then $\|x_i - x_j\| < r$. Then there exist finitely many sets of points satisfying the same set of distance constraints and strict distance inequalities (modulo isometries) if and only if there exist finitely many sets of points satisfying the same set of distance constraints (modulo isometries).*

Remark 4.3 Recall from Sect. 2.1 that, if there exist only finitely many sets of points satisfying the same set of distance constraints (modulo isometries), then we say that the corresponding bar-joint framework is rigid. Further, a unique solution means that the bar-joint framework is globally rigid.

Proof of Proposition 4.2 If there exist finitely many sets of points satisfying the same set of distance constraints (modulo isometries), then there are also finitely many sets of points satisfying the same set of distance constraints and strict distance inequalities (modulo isometries). For the converse direction, assume that there are infinitely many sets of points satisfying the same set of distance constraints (modulo isometries). Select one set that is also satisfying the inequality constraints and consider the corresponding generic bar-joint framework (G, ρ) .

Let $f_G^{-1}(f_G(\rho))$ denote the configuration space, that is the set of all $q \in \mathbb{R}^{d|V|}$ such that (G, ρ) and (G, q) have the same edge lengths. The fact that $f_G^{-1}(f_G(\rho))$ is a smooth manifold is non-trivial but it is well-known in rigidity theory. This follows from a basic lemma in differential topology (e.g. Milnor 1965, p.11, Lemma1) after proving that $f_G(\rho)$ is a regular value of f_G for any generic ρ , where f_G is the rigidity map. Establishing that $f_G(\rho)$ is a regular value is an application of both Sard's theorem and Tarski-Seidenberg elimination theory (see, e.g., Jackson et al. 2014, Lemma11).

Let \sim be the equivalence relation given by congruences of realisations, and let $U \subset \mathbb{R}^{d|V|}$ be an open neighbourhood of ρ . We claim that, for sufficiently small U , the set $f_G^{-1}(f_G(\rho)) \cap U / \sim$ is a smooth manifold. Since the Euclidean isometries form a Lie group that acts freely on the set of realisations in general position², $f_G^{-1}(f_G(\rho)) \cap U / \sim$ is also a smooth manifold for sufficiently small U . We now note that $f_G^{-1}(f_G(\rho)) \cap U / \sim$ has positive dimension: either G is connected and so $f_G^{-1}(f_G(\rho)) \cap U / \sim$ is compact, and so must have positive dimension since it contains infinitely many points, or G is disconnected, in which it is a simple exercise to show that $f_G^{-1}(f_G(\rho)) \cap U / \sim$ has positive dimension stemming from translating exactly one connected component of the framework (G, ρ) .

Since the framework satisfies the inequalities strictly, all perturbations of the framework in a sufficiently small open neighbourhood of ρ also satisfy the inequalities.

² In that for every isometry g and every general position realisation ρ , the realisation $g \circ \rho$ is equal to ρ if and only if g is the identity map. Every generic realisation is in general position, and the set of general position realisations forms an open dense set.

Since $f_G^{-1}(f_G(\rho)) \cap U/\sim$ has positive dimension, there are infinitely many edge length preserving perturbations modulo isometries, which completes the proof. \square

While inequalities alone do not determine whether there is a finite number of reconstructions, they can reduce the solution space from a finite number to exactly one. An elementary example is given by Fig. 1b which as depicted contains only equalities which leads to exactly two solutions (modulo isometries). However, if we introduce an inequality constraining the distance between the two non-adjacent vertices, then we can make the solution unique.

Whether a set of distance constraints characterises a finite set of $|V|$ points can be checked with high probability by examining the rank of the corresponding rigidity matrix $R(G, \rho)$. If the rank is not $d|V| - \binom{d+1}{2}$ then the constraints do not locally characterise the points up to rigid transformations (Lemma 2.3). On the other hand, if the rank is $d|V| - \binom{d+1}{2}$ then the rank is unchanged if we move to a generic framework (G, ρ') , and so Lemma 2.3 shows that the constraints do characterise the points.

4.3 Equality constraints

A special case of this setting is when we have only equality constraints, i.e. we have data from microscopy experiments but no Hi-C data. In this case, the uniqueness of a solution is equivalent to global rigidity (and finiteness is equivalent to rigidity) of bar-joint frameworks. Here, beyond the material we already presented in Sect. 2.1, there is a large literature that one can exploit (see Asimow and Roth 1978; Connelly 2005; Gortler et al. 2010; Hendrickson 1992; Jackson and Jordán 2005b, inter alia). In particular let us note the following extension of Lemma 2.2 to global rigidity which is a consequence of a theorem of Hendrickson.

Lemma 4.4 (Hendrickson 1992) *Let G be a graph on at least $d + 2$ vertices. If G is globally d -rigid, then $|E| \geq d|V| - \binom{d+1}{2} + 1$.*

There exists a random polynomial time algorithm for testing whether a graph is d -rigid. This algorithm (Jacobs and Hendrickson 1997) follows from Lemma 2.3. A similar situation exists for global d -rigidity, but we need another concept first.

An *equilibrium stress* ω of a framework (G, ρ) is a vector in the cokernel of $R(G, \rho)$. The *stress matrix* Ω is the $|V| \times |V|$ symmetric matrix whose off-diagonal ij -entry is 0 if ij is not an edge and $-\omega_{ij}$ if ij is an edge, and the diagonal entry in row i is the sum $\sum_j \omega_{ij}$. In other words, Ω is the Laplacian matrix of G weighted by ω .

Theorem 4.5 (Connelly 2005, Gortler et al. (2010)) *A graph G on at least $d + 2$ vertices is globally d -rigid if and only if $\text{rank}(\Omega) = |V| - d - 1$.*

Efficient deterministic algorithms are only known for cases where $d \leq 2$, based on a combinatorial characterisation by Jackson and Jordán (2005b) that extends the pebble game algorithm introduced by Jacobs and Hendrickson (1997). However, an alternative approach is presented by Gortler et al. (2010), which provides a polynomial-time randomised algorithm in arbitrary dimensions leveraging Theorem 4.5.

By taking on board information about the backbone of the chromosome, one can infer more detailed rigidity theoretic information. To illustrate this, we consider graph

powers. The k -th *power* of a graph G , denoted G^k , is obtained from G by adding a new edge uv for all non-adjacent vertex pairs u, v of G with distance at most k in G .

The purpose of the next result about graph powers is to illustrate that basic tools from rigidity theory can provide precise information in cases when the locality of the backbone vertices forces enough edges. Since it requires no extra work, we present the result in general dimension.

Lemma 4.6 *Let $G = (V, E)$ be a path. Then G^{d+1} is globally d -rigid.*

Proof If $|V| \leq d + 1$ then G^{d+1} is complete, and hence globally d -rigid. Suppose $|V| \geq d + 2$ and note that the complete graph K_{d+2} is globally d -rigid. Observe that G^{d+1} is obtained from K_{d+2} by a sequence of degree $d + 1$ vertex additions. Adding a vertex of degree $d + 1$ preserves global rigidity as long as the $d + 1$ neighbours affinely span \mathbb{R}^d (which trivially holds in the generic case). So, the proof follows by an elementary induction argument. \square

Returning to the question of realisability, now for systems with only equality constraints, we remark that reducing from a high dimensional realisation, as in Proposition 4.1, i.e. through rank relaxation makes the semidefinite optimisation problem non-convex and, in the worst case, NP-hard. However, this *distance geometry problem* is well known and well studied with many special cases of applied interest being manageable in practice. We direct the interested reader to Liberti et al. (2014); Cassioli et al. (2015) and the references therein for details.

5 Sphere packing models

In this section, we consider a simplification of the unit ball graph model, where contacts are defined by pairwise distances of one, and non-contacts are characterised by distances greater than one. Although this model imposes greater restrictions, it offers the advantage of producing unique reconstructions in certain cases. The model is defined by distance equality and inequality constraints, so it is a special case of the model in Sect. 4.

Definition 5.1 A *penny graph*³ is the contact graph of a collection of unit discs with non-overlapping interiors, i.e. there exists a realisation $\rho : V \rightarrow \mathbb{R}^2$ such that

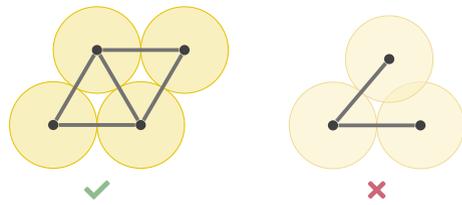
$$\begin{aligned}vw \in E &\implies \|\rho(v) - \rho(w)\| = 1, \\vw \notin E &\implies \|\rho(v) - \rho(w)\| > 1;\end{aligned}$$

such a realisation ρ is said to be a *penny graph realisation* of G . *Marble graphs* and *marble graph realisations* are the 3D analogues of penny graphs and penny graph realisations, respectively.

Figure 8 illustrates an example and a non-example of penny realisations.

³ Penny graphs are also known as unit coin graphs, minimum-distance graphs, smallest-distance graphs and closest-pairs graphs in the literature.

Fig. 8 A penny realisation of a graph (left) and a realisation of another graph that does not fulfil the penny condition since two vertices are too close (right)



Let us first focus on penny graphs. As penny graphs are contact graphs of disc packings, every penny graph is planar. The converse of this statement is not true, as can be seen by the Moser spindle (Fig. 9) which is not realisable as a penny graph. The next lemma lists some basic properties of penny graphs.

Lemma 5.2 *Let $G = (V, E)$ be a penny graph on n vertices. Then:*

- (i) $|E| \leq \lfloor 3n - \sqrt{12n - 3} \rfloor$;
- (ii) the minimum degree is at most 3 and the maximum degree is at most 6;
- (iii) a maximal clique has at most 3 vertices;
- (iv) the chromatic number is at most 4.

Proof The upper bound in (i) was proved by Harborth (1974). This upper bound is also tight; it can be achieved by taking subsets of a triangular lattice. For the lower bound in (ii), at least one vertex of a straight-line planar graph embedding must be convex (i.e. have all neighbours contained in a cone with apex at the vertex), and every convex vertex of a penny realisation has degree at most 3. The upper bound is simply the number of pennies that can be placed around a single penny.

Since at most three points can be equidistant in the plane, conclusion (iii) is immediate. Finally (iv) follows from the 4-colour theorem (or a simple induction argument on the minimum degree). □

Significantly less is known surrounding marble graphs. We summarise the situation in the following lemma with bounds that, unlike the penny graph case, are not tight.

Lemma 5.3 *Let $G = (V, E)$ be a marble graph on n vertices. Then:*

- (i) $|E| \leq 6n - 0.926n^{2/3}$;
- (ii) the minimum degree is at most 8 and the maximum degree is at most 12;
- (iii) a maximal clique has at most 4 vertices;
- (iv) the chromatic number is at most 9.

Proof It was proven by Bezdek and Reid (2013) that an n -vertex marble graph can have at most $6n - 0.926n^{2/3}$ edges. Since the kissing number of a sphere is 12, the maximum degree of a marble graph is 12. The minimum degree of a marble graph is at most 8; this follows from a similar observation that a strictly convex vertex must exist, and a result of Kertész regarding the number of points on the open upper hemisphere that are at least 60° apart (Kertész 1994). The upper bound on the chromatic number was proved by Kertész (1994). □

The current maximum number of edges found in a marble graph is roughly $6n - \sqrt[3]{486n^{2/3}}$, which is achievable with induced subgraphs of the face-centred cubic lattice

Fig. 9 The Moser spindle is an example of a graph that is not realisable as a penny graph

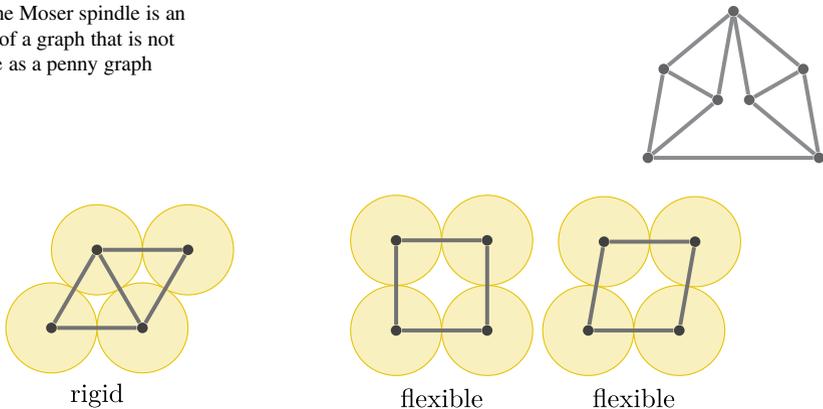


Fig. 10 A rigid penny realisation of a penny graph (left) and two flexible penny realisations of a penny graph (right). One can continuously deform the locations of the pennies, maintaining the given contacts to deform the first realisation into the second

(Bezdek 2012). Similar to penny graphs, we can easily identify that the clique number of a marble graph is at most 4 (since at most four points can be equidistant in \mathbb{R}^3). The maximum possible chromatic number of a marble graph is between 5 (Maehara 2007) and 9 (Kertész 1994), but the exact upper bound is currently open.

Unfortunately, determining whether a graph is a penny graph is NP-hard (Breu and Kirkpatrick 1996). Similarly, determining whether a graph is a marble graph is NP-hard (Hliněný 1997). Interestingly, the problem of identifying contact graphs of unit d -sphere packings (the d -dimensional variant of identifying penny and marble graphs) is known to be NP-hard when $d \in \{2, 3, 4, 8, 24\}$ (Hliněný 1997; Hliněný and Kratochvíl 2001), but remains completely open in all other dimensions greater than 1.

We are particularly interested in the rigidity and global rigidity properties of penny and marble graphs in relation to their penny graph and marble graph realisations. We illustrate this in Fig. 10 in examples. This motivates the following definition.

Definition 5.4 Let G be a penny graph (resp., marble graph). We say that G is *penny-rigid* (resp., *marble-rigid*) if there exist finitely many penny graph realisations (resp., marble graph realisations) modulo isometries. We say that G is *globally penny-rigid* (resp., *globally marble-rigid*) if there exists exactly one penny graph realisation (resp., marble graph realisation) modulo isometries.

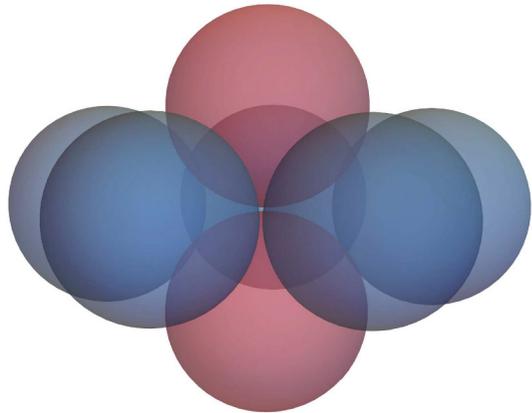
Similarly to the existence problem, determining whether a penny graph is penny-rigid or globally penny-rigid is also computationally difficult.

Lemma 5.5 (Lemma 2.4 Dewar et al. (2025)) Let $G = (V, E)$ be a sphere graph. Then the following properties hold.

- (i) If G is globally sphere-rigid, then G is sphere-rigid.
- (ii) If G is sphere-rigid, $d \leq 3$ and $|V| \geq d + 1$, then G is d -connected.

(Dewar et al. 2025, Theorem 2.6) give a precise, efficient characterisation of when a chordal penny graph is penny-rigid, as well as the same statement for marble-rigidity.

Fig. 11 A marble realisation of the complete bipartite graph $K_{2,5}$



(For chordal graphs, it is shown by Dewar et al. (2025) that penny-global rigidity and penny-rigidity coincide, so the characterisation applies equally to penny-global rigidity.) That paper goes on to illustrate many examples of how penny-rigidity and penny-global rigidity can differ from the standard rigidity notions for bar-joint frameworks (see Dewar et al. 2025, Table 1).

Lemma 5.6 *The graphs $K_{1,13}$ and $K_{2,6}$ are minimal forbidden subgraphs of the family of marble graphs.*

Proof A minor adaptation to the proof given in Lemma 3.3 proves that $K_{1,13}$ is a minimal forbidden subgraph.

For the remainder of the proof we introduce the following notation. For any two graphs G_1, G_2 with distinct vertices, we denote by $G + H$ the graph formed by connecting every vertex in G to every vertex in H . We also denote any graph with n vertices by G_n . We say that a graph is a d -dimensional ball graph if it is the contact graph of a set of d -dimensional balls with non-overlapping interiors. Chen (2016, Corollary 3.7) proved that every graph $G_4 + G_6$, with the exception of G_4 being a cycle and G_6 being the 1-skeleton of the octahedron, is not the contact graph of a set of 4-dimensional balls (here allowing for the balls to have different radii) with non-overlapping interiors. In particular, any graph of the form $K_2 + G_2 + G_6$ is not one of the aforementioned contact graphs. Hliněný and Kratochvíl (2001, Proposition 4.5) showed that a graph $G + K_2$ is the contact graph of a set of $(d + 1)$ -dimensional balls with non-overlapping interiors if and only if G is the contact graph of a set of d -dimensional unit radius balls with non-overlapping interiors. Hence, any graph of the form $G_2 + G_6$ is not a marble graph; in particular, $K_{2,6}$ is not a marble graph. Minimality now follows from noting that every vertex-induced subgraph of $K_{2,6}$ is a marble graph: $K_{1,6}$ is a marble graph as it is a proper induced subgraph of the minimal forbidden subgraph $K_{1,13}$, and $K_{2,5}$ is a marble graph as it is the contact graph of the marble realisation pictured in Fig. 11. \square

Remark 5.7 Holmes-Cerfon (2016) studied the problem of determining the number of all possible sphere-rigid marble graphs with a fixed number of vertices. Specifically,

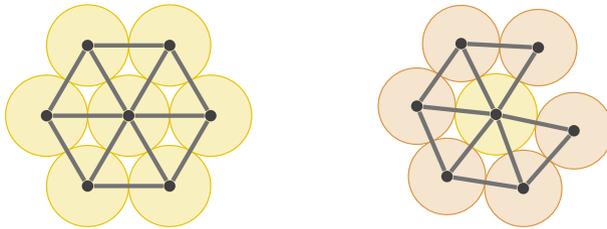


Fig. 12 A graph with a penny-realisation (left) and one that is only realisable with generic radii (right). Note that on the right the outer circles are slightly smaller than the inner one (though for the purpose of the figure still non-generic)

they systematically enumerated the sphere-rigid marble graphs with at most 14 vertices, and the sphere-rigid marble graphs with the maximum number of edges with at most 19 vertices. They found these graphs computationally, with no completeness proof. However, they observed that the list may only miss a negligible number of exceptionally singular graphs.

Recall that Lemma 2.3 implies that a minimally rigid generic framework in 3-dimensions, on a graph G with n vertices, has $3n - 6$ edges. In statistical physics, the term *hypostatic* is used to refer to equal radii sphere packings (or in our language, sphere-rigid marble graphs) with less than $3n - 6$ edges and the term *hyperstatic* is used for sphere-rigid marble graphs with more than $3n - 6$ edges. The computations of Holmes-Cerfon (2016) establish that there is a transition at $n = 10$ where both hypostatic and hyperstatic rigid marble graphs emerge.

6 Generic radii penny and marble graphs

In this section, we consider a generalisation of the penny graphs from Definition 5.1, where the radii of the circles could be slightly larger or smaller than one. See Fig. 12 for an example of a graph that is realisable only with generic radii.

Definition 6.1 A *generic penny/marble graph* is the contact graph of generic radii circles/spheres. The vertices correspond to the centres of the circles/spheres and there is an edge between two vertices if and only if the corresponding circles/spheres touch.

In our target application the lengths of the edges are all expected to be similar. Hence, we consider ε -interval radii penny graphs, that is each penny has radius within an interval from $1 - \varepsilon$ to $1 + \varepsilon$ for some sufficiently small $\varepsilon > 0$, and the set of all radii is algebraically independent.

Connelly et al. (2019) studied the realisability and rigidity properties of generic radii penny graphs. In particular, they provided a tight linear bound on the number of edges in any generic radii penny graph. Further when the bound holds, they classified when the contact graph is rigid and infinitesimally rigid as a penny graph (Connelly et al. 2019, Theorem 1.5). In our setting, this translates to the following lemma.

Lemma 6.2 *The ε -interval radii penny graphs have at most $2n - 3$ edges for any $\varepsilon > 0$ and they are rigid precisely when the bound is attained.*

Observe that this upper bound is less than the minimum number of edges required for a bar-joint framework to be globally rigid in the plane (Lemma 4.4).

It is also worth pointing out that realisations of generic radii penny graphs do not correspond to generic frameworks. Indeed for a generic framework (G, ρ) in the plane on n vertices, the set of coordinates of p is an algebraically independent set of size $2n$, whereas a generic radii penny graph only has an algebraically independent set of size n . From this viewpoint, it should be no surprise that these graphs typically exhibit non-generic behaviour. Hence, none of the classical rigidity theory methods can be directly applied in this setting. Note that our ϵ -interval radii penny graphs are even more restrictive, given that the radii of our circles are in the intervals $(1 - \epsilon, 1 + \epsilon)$. Nevertheless, as Lemma 6.7 shows, global rigidity can be established in some interesting cases.

While generic radii penny graphs can have any maximum degree and adjacent vertices may have multiple common neighbours, we note the following result that illustrates the additional restrictions that interval radii can place.

Lemma 6.3 *Let G be an ϵ -interval radii penny graph for some*

$$\epsilon \leq 1 - \sqrt{2(1 - \cos(2\pi/7))} \approx 0.13223.$$

Then the maximum degree in G is at most 6.

Proof Fix $\epsilon > 0$ to be the smallest value so that it is possible to have a configuration of seven pennies with disjoint interiors and radii contained in the closed interval $[1 - \epsilon, 1 + \epsilon]$ such that one penny is in contact with all other pennies. We call the penny in contact with all others the *centre penny* and the others the *outer pennies*.

We observe that it is always possible to increase the radius of the centre penny and/or decrease the radius of the outer pennies and still find a configuration of seven pennies with the same properties. Hence, we may suppose that the centre penny has radius $1 + \epsilon$ and that each outer penny has radius $1 - \epsilon$. As we chose ϵ as small as possible, it follows that the outer pennies must form a cycle. This set-up produces a regular heptagon with side lengths $2 - 2\epsilon$. Label the centre of the centre penny as o . Then o is distance 2 from each vertex of the heptagon.

Pick two adjacent vertices x, y of the heptagon. The triangle xoy is isosceles with angle $2\pi/7$ radians at o and side lengths $\|x - o\| = \|y - o\| = 2$ and $\|x - y\| = 2 - 2\epsilon$. Using the cosine rule and rearranging, we see that $\epsilon = 1 - \sqrt{2(1 - \cos(2\pi/7))}$. The result now follows since an ϵ -interval radii penny graph does not allow pennies with radius $1 \pm \epsilon$. □

Lemma 6.4 *Let P_0, P_1, P_2, P_3 be four discs with disjoint interiors, with each P_i having radius r_i and centre at p_i . Suppose that P_1, P_2, P_3 are in pairwise contact and p_0 lies in the interior of the triangle formed by the points p_1, p_2, p_3 . Then*

$$3 + 2\sqrt{3} \leq \max\{r_i/r_0 : i \in \{1, 2, 3\}\}.$$

Proof Fix $r_0 = 1$. Suppose that P_1, P_2, P_3 have been chosen so that the value $r_{\max} := \max\{r_i : i \in \{1, 2, 3\}\}$ is minimal. Note that it is possible to shrink each of P_1, P_2, P_3

whilst maintaining all contacts between discs up to the point where P_0 is in contact with all other discs, and this operation decreases the value of r_{\max} . Hence, we can assume that P_0 is in contact with each of P_1, P_2, P_3 . Suppose that $r_{\max} < 3 + 2\sqrt{3}$. For any two vertices $i, j \in \{1, 2, 3\}$, let θ_{ij} be the angle at p_0 in the triangle $p_0 p_i p_j$. Since $\|p_0 - p_i\| = 1 + r_i, \|p_0 - p_j\| = 1 + r_j$ and $\|p_i - p_j\| = r_i + r_j$, it follows from elementary trigonometry that

$$\begin{aligned} \cos \theta_{ij} &= \frac{(1 + r_i)^2 + (1 + r_j)^2 - (r_i + r_j)^2}{2(1 + r_i)(1 + r_j)} = 1 - \frac{2r_i r_j}{(1 + r_i)(1 + r_j)} \\ &< 1 - \frac{2(3 + 2\sqrt{3})^2}{(4 + 2\sqrt{3})^2} = \frac{1}{2}. \end{aligned}$$

Hence, $\theta_{ij} < 2\pi/3$. However, we now see that $\theta_{12} + \theta_{23} + \theta_{13} < 2\pi$, contradicting that the discs P_0, P_1, P_2, P_3 are pairwise touching with P_0 at their centre. Hence, $r_{\max} \geq 3 + 2\sqrt{3}$, which implies the desired inequality. \square

Lemma 6.5 *Let G be an ε -interval radii penny graph for some*

$$\varepsilon \leq \sqrt{3} - 1 \approx 0.73205.$$

Then any pair of adjacent vertices have at most two common neighbours.

Proof Let G be a penny graph with vertices x, y, a, b, c so that x, y are adjacent to each other and each of a, b, c . Let ρ be an ε -interval radii penny realisation of G , and fix r_v to be the radii of the disc associated with vertex v . We observe now that each of the triples xya, xyb, xyc is a triangle in G . By the planarity of ρ , there exists distinct $i, j \in \{a, b, c\}$ such that $\rho(i)$ is contained in the triangle formed by the points $\rho(x), \rho(y), \rho(j)$. By Lemma 6.4 we see that

$$3 + 2\sqrt{3} \leq \max\{r_v/r_i : v \in \{x, y, j\}\} < \frac{1 + \varepsilon}{1 - \varepsilon} \leq 3 + 2\sqrt{3}, \tag{1}$$

which is a contradiction. \square

Remark 6.6 Generic radii penny graphs can be naturally extended to higher dimensions, by replacing the disks with spheres. It is conjectured by Ozkan et al. (2018) that the number of edges in the graph of such a generic radii packing is bounded linearly by $3n - 6$. This conjecture has been further studied by Dewar (2023). Notably, it has been demonstrated that the conjecture holds true for any sphere packing with a contact graph in the form (using the notation described in Lemma 5.6) of $G + K_2$, where G is a penny graph containing no cycles.

Generic radii global rigidity is not a generic property. That is, some graphs have some generic radii realisations that are globally rigid and other generic radii realisations that are not globally rigid. For example the realisation of $K_4 - e$, depicted on the left of Fig. 8 as a penny graph, can be perturbed to a nearby generic radii realisation. This is globally rigid as a generic radii (or interval radii) penny graph. The underlying bar-joint

framework, on the other hand, is not globally rigid as one may ‘flip’ either vertex of degree 2 through the line defined by its neighbours to obtain a second realisation with the same edge lengths. This flip would, of course, destroy the generic radii realisation as the pennies corresponding to the degree 2 vertices would overlap.

Assuming that a graph is an interval radii penny graph, the global rigidity problem now becomes: does it have a unique interval radii penny graph realisation up to rigid transformations of the plane? We give a sample theoretical result illustrating a family of graphs for which uniqueness in this model can be determined from the graph.

A *2-lateration* graph is a graph obtained from K_2 by sequentially adding vertices of degree 2 so that, at each step, the neighbours of the new degree 2 vertex are adjacent. Such graphs are relevant to our study for example if one assumes the backbone is connected there is a reasonably high probability that there is a spanning 2-lateration subgraph.

Lemma 6.7 *Let (G, ρ) be an ε -interval radii penny graph realisation where $\varepsilon \leq \sqrt{3} - 1 \approx 0.73205$. If G is a planar 2-lateration graph then (G, ρ) is globally rigid.*

Proof The lemma trivially holds when (G, ρ) is a triangle. Inductively, suppose that the lemma is true for any ε -interval radii penny graph realisation of a planar 2-lateration graph on up to n vertices. Let (G, ρ) be an ε -interval radii penny graph realisation where G is a planar 2-lateration graph with $n + 1$ vertices. Then G contains a degree 2 vertex v (with neighbours x, y) so that $G - v$ is a planar 2-lateration graph on n vertices and $(G - v, \rho|_{G-v})$ is a ε -interval radii penny graph realisation of $G - v$. By our inductive assumption, $(G - v, \rho|_{G-v})$ is globally rigid. As G is a planar 2-lateration graph with 4 or more vertices, x, y are both adjacent to another vertex z . Given the positions of the vertices of $G - v$ given by ρ , there are exactly two possible locations for the vertex v that satisfy the distance constraints for the edges vx and vy . As $\{x, y, z\}$ defines a triangle, it follows from Lemma 6.4 and Eq. (1) that there is a unique location for the penny corresponding to v so that there is no overlap. \square

7 Reconstruction algorithm

In this section, we study the 3D reconstruction problem. This problem is closely related to the Euclidean distance problem which aims to construct points from pairwise distances between them. We adapt semidefinite programming based algorithms for the Euclidean distance problem and 3D genome reconstruction to the models in Sects. 3, 4, and 5. It seems that our approach can not easily be applied to the model in Sect. 6.

Numerous algorithms for solving the Euclidean distance problem are summarised in the overview article by Liberti et al. (2014). In our setting, by using inequalities we allow interval data (Liberti et al. 2014, Section 3.4), i.e. edges can be weighted with bounded or unbounded real intervals instead of just real numbers. More details on the interval setting can be found (see Cassioli et al. 2015).

We focus on algorithms based on semidefinite programming, as they provide convex formulations that easily allow one to account for noisy measurements. Semidefinite programs for distance geometry problems have been particularly popular in wireless sensor network localisation (Biswas et al. 2006; So and Ye 2007), but also in molecular

conformation problems (Leung and Toh 2010) and computer vision (Weinberger and Saul 2006). In the context of 3D genome reconstruction, they have been used in the population haploid case by Zhang et al. (2013) and the diploid case by Belyaeva et al. (2022); Cifuentes et al. (2024).

7.1 Optimisation formulation

To construct a semidefinite relaxation for the reconstruction problem, we recall the connection between Euclidean distance and positive semidefinite matrices. Let $x_1, \dots, x_n \in \mathbb{R}^d$ be the coordinates of beads. Let G be the corresponding Gram matrix, i.e. $G_{ij} = \langle x_i, x_j \rangle$. It is a positive semidefinite matrix of rank at most d . The squared distance matrix $D^{(2)}$ can be obtained from the Gram matrix G using the relation

$$D_{ij}^{(2)} = G_{ii} + G_{jj} - 2G_{ij}.$$

We define

$$g_{ij}(G) := G_{ii} + G_{jj} - 2G_{ij}.$$

Let our model be defined by

$$\begin{aligned} g_{ij}(G) &= a_{ij}^2 \text{ for } (i, j) \in S_{=}, \\ g_{ij}(G) &\leq \bar{a}_{ij}^2 \text{ for } (i, j) \in S_{\leq}, \\ g_{ij}(G) &\geq \underline{a}_{ij}^2 \text{ for } (i, j) \in S_{\geq}, \\ g_{ij}(G) &< \bar{a}_{ij}^2 \text{ for } (i, j) \in S_{<}, \\ g_{ij}(G) &> \underline{a}_{ij}^2 \text{ for } (i, j) \in S_{>}, \end{aligned} \tag{2}$$

where $S_{=}, S_{\leq}, S_{\geq}, S_{<}, S_{>} \subseteq [n] \times [n]$.

For example, if we have the unit ball graph model, then $S_{=} = \emptyset$, $S_{\leq} = E$ and $S_{>} = ([n] \times [n]) \setminus (\{(1, 1), \dots, (n, n)\} \cup E)$. We did not find an easy way to adapt this formulation to interval radii marble graph model, which is the reason why the model is not considered in this section.

We consider the following convex optimisation formulation:

$$\begin{aligned}
 & \underset{G}{\text{minimise}} && \sum_{(i,j) \in S_{=}} (g_{ij}(G) - a_{ij}^2)^2 \\
 & \text{subject to} && g_{ij}(G) \leq \bar{a}_{ij}^2 \text{ for } (i,j) \in S_{\leq}, \\
 & && g_{ij}(G) \geq \underline{a}_{ij}^2 \text{ for } (i,j) \in S_{\geq}, \\
 & && g_{ij}(G) \leq \bar{a}_{ij}^2 - \varepsilon \text{ for } (i,j) \in S_{<}, \\
 & && g_{ij}(G) \geq \underline{a}_{ij}^2 + \varepsilon \text{ for } (i,j) \in S_{>}, \\
 & && \sum_{1 \leq i, j \leq n} G_{ij} = 0, \\
 & && G \succeq 0.
 \end{aligned} \tag{3}$$

In the formulation (3), we have replaced the strict inequalities $g_{ij}(G) < \bar{a}_{ij}^2$ (resp. $g_{ij}(G) > \underline{a}_{ij}^2$) by $g_{ij}(G) \leq \bar{a}_{ij}^2 - \varepsilon$ (resp. $g_{ij}(G) \geq \underline{a}_{ij}^2 + \varepsilon$) for a sufficiently small epsilon to obtain non-strict inequalities. The constraint $\sum_{1 \leq i, j \leq n} G_{ij} = 0$ removes the freedom that comes from translations of configurations.

Solving the above semidefinite program, gives us a Gram matrix \hat{G} . To find a point configuration in \mathbb{R}^d , we find the d largest eigenvalues $\lambda_1, \dots, \lambda_d \in \mathbb{R}$ of \hat{G} and the corresponding eigenvectors $v_1, \dots, v_d \in \mathbb{R}^n$. For $1 \leq i \leq n$, we define

$$x_i = \left(\sqrt{\lambda_1} \cdot v_{1,i}, \dots, \sqrt{\lambda_d} \cdot v_{d,i} \right).$$

This method is called classical multidimensional scaling (cMDS) (Cox and Cox 2008). If the optimal Gram matrix \hat{G} found by the semidefinite program has rank at most d , the point configuration constructed using cMDS satisfies exactly the lower and upper bounds for distances. If the rank of the Gram matrix is greater than d , then the realisation in \mathbb{R}^d might not satisfy all the bounds and we obtain an approximate solution.

Remark 7.1 Adding $\lambda \cdot \text{trace}(G)$ or $-\lambda \cdot \text{trace}(G)$ to the objective function has been suggested as a heuristic to obtain a low-rank Gram matrix. The idea behind adding $-\lambda \cdot \text{trace}(G)$ to the objective is that, in the case of noisy distances, the points tend to crowd towards the centre and maximising trace spreads out the locations of the beads (Biswas et al. 2006). An alternative approach is to minimise $\text{trace}(G)$, since trace is the convex envelope of the rank function in the unit ball. Minimising trace instead of rank was suggested by Fazel et al. (2001) and it has been applied to the 3D genome reconstruction setting from Belyaeva et al. (2022). In our experiments, modifying the objective function did not improve results, so we used the original formulation (3). The term $\lambda \cdot \text{trace}(G)$ can be added in our implementation by giving $\lambda \in \mathbb{R}$ as an option.

As the final step, we solve the optimisation problem

$$\min \left(\sum_{(i,j) \in S_{\leq} \cup S_{<}} (\|x_i - x_j\| - \bar{a}_{ij})_+^2 + \sum_{(i,j) \in S_{\geq} \cup S_{>}} (\|x_i - x_j\| - \underline{a}_{ij})_-^2 + \sum_{(i,j) \in S_{=}} (\|x_i - x_j\| - a_{ij})^2 \right) \quad (4)$$

using gradient descent initialised at the solution found by the multidimensional scaling. In the above formulation, $\alpha_+ = \max(0, \alpha)$ and $\alpha_- = \max(0, -\alpha)$. This step is the same as by Leung and Toh (2010). Even if the n -dimensional solution corresponding to the optimal Gram matrix satisfies the constraints (2), the d -dimensional solution obtained by cMDS might not. The goal of this step is to improve the d -dimensional solution. To summarise, our reconstruction algorithm performs the following three steps:

- (i) Running the SDP program (3).
- (ii) Using classical multidimensional scaling to obtain a 3D solution from the optimal Gram matrix obtained in step (i).
- (iii) Minimising the objective (4) using gradient descent with initialisation equal to the solution obtained in step (ii).

7.2 Experiments

We run step (i) using `cvx` with the default solver `SDPT3`. For step (iii), we use the `fminunc` command in `MATLAB`. The algorithm used for local optimisation is quasi-Newton. We use two measures of dissimilarity to study the quality of the reconstruction:

- (i) The modification of the objective function (4) where we only keep the inequalities corresponding to the HiC matrix:

$$\min \sum_{(i,j):F(i,j)=1} (\|x_i - x_j\| - 1)_+^2 + \sum_{(i,j):F(i,j)=0} (\|x_i - x_j\| - 1)_-^2. \quad (5)$$

We call the value corresponding to the function (5) the violation.

- (ii) The Procrustes distance between the original structure and the reconstructed structure. The Procrustes transformation finds the affine transformation that minimises the sum of squared distances between two ordered point sets. The Procrustes distance is the sum of squared distances between the aligned point sets divided by the scale of the first point set. We do allow scaling in the Procrustes algorithm.

The Procrustes distance can be computed only if the original structure is known, e.g. for synthetic data sets. Moreover, it is comparing the reconstruction with the original

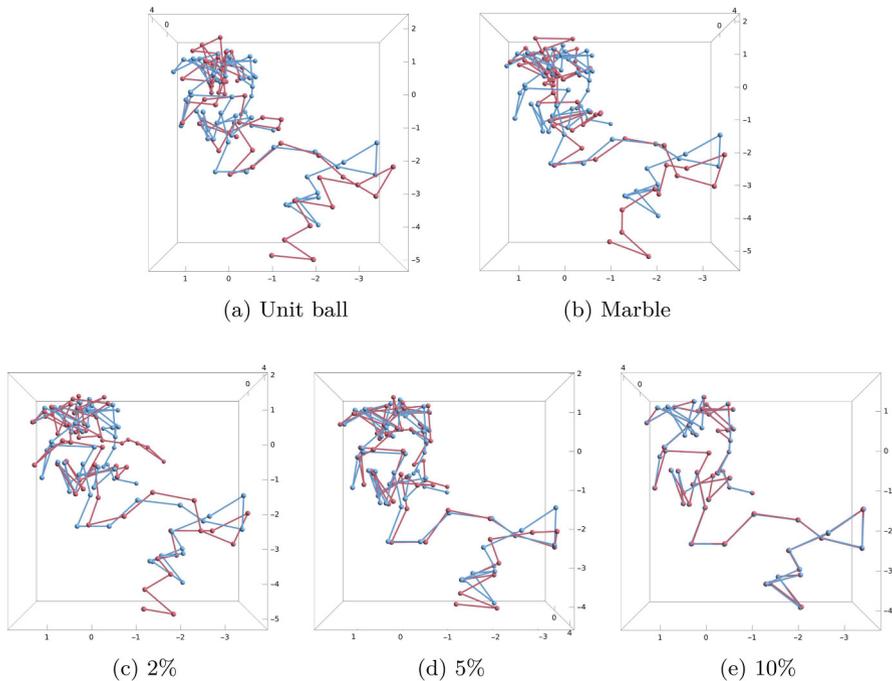


Fig. 13 Examples of reconstructions obtained using our reconstruction method using the backbone information that the distances between neighbouring beads are equal to 1: **a** Unit ball model. **b** Marble graph model. **c–e** Equations and inequalities model with 2%, 5% and 10% of edge lengths known. The chromosome has 60 beads generated from the normalised Brownian motion model and the threshold for a contact is $r = 3$. The backbone of the original data is shown in blue, while red shows the reconstruction

structure, which is more than comparing Hi-C matrices. Therefore, we mostly focus on the first measure of dissimilarity.

Synthetic data. Our synthetic data set consists of random walks with 60 points, where each step of the random walk is generated from the 3D sphere (all steps are of length one). A threshold r is used to construct the Hi-C matrix. If the distance between two points is at most r , then the corresponding value in the Hi-C matrix is set to be 1, and otherwise it is set to be 0. For all models, we assume that we know the distance between neighbouring beads on a chromosome, which in the experiments is equal to one.

For each value $r \in \{1, 2, 3, 4, 5, 6\}$, we generated 50 random walks with 60 vertices. Figure 13 shows example reconstructions using our reconstruction algorithm for unit ball, marble graph and models with equations and inequalities for the threshold $r = 3$. In Fig. 14, we show the median violation (5) for the different models on a log-scale for thresholds r . Some results are numerically 0 which yields the extra line in the graphic. In Fig. 15, we compare the Procrustes distance for these models.

In all three figures, the equations and inequalities model with 10% of edge lengths known gives very good reconstructions. The same model with 5% of edge lengths known also gives a relatively good reconstruction in Fig. 13d and small Procrustes

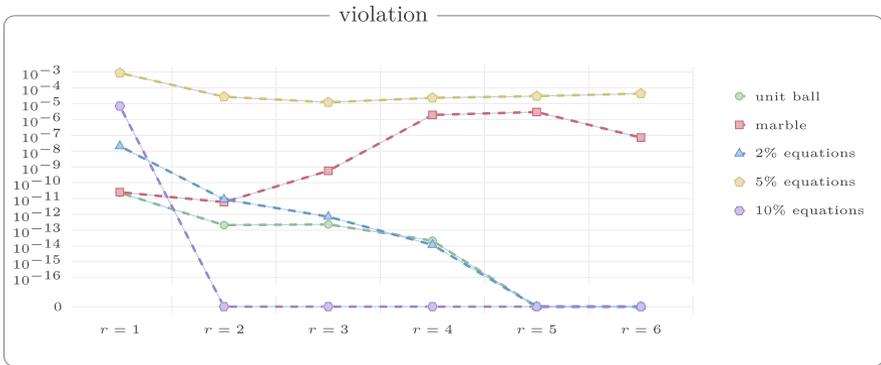


Fig. 14 Medians of the violation to the HiC-matrix for unit ball model, marble graph model and equations and inequalities model with 2%, 5% and 10% of edge lengths known. The medians are plotted in log-scale. The chromosomes have 60 beads generated from the normalised Brownian motion model with backbone. The violation is shown for different thresholds r

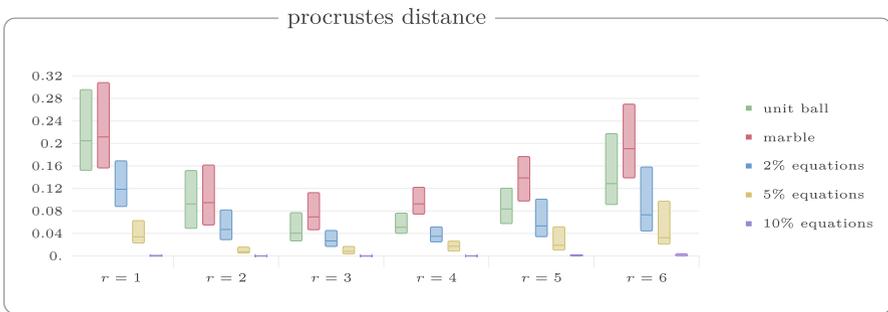


Fig. 15 Medians and 25/75%-quantiles of the Procrustes distance (including scaling) for unit ball model, marble graph model and equations and inequalities model with 2%, 5% and 10% of edge lengths known. The chromosomes have 60 beads generated from the normalised Brownian motion model with backbone. The violation is shown for different thresholds r

distance, but it gives the worst violation out of all the models. We explored the performance of the violation for the equations and inequalities model with different percentage of edge lengths and observed that the violation increases when going from 0% of edges known to about 5% edges known, and decreases after that. One explanation of this could be that with no edge lengths known, i.e. in the unit ball model, it is possible to obtain a 3D structure with low violation although it might be different from the original structure. When adding more edge lengths, initially the violation can get worse, although the Procrustes distance decreases, and if enough edges are added then both violation and Procrustes distance decrease. The other three models get the high-level details of the reconstructions correct, but details are not constructed correctly. These observations are consistent with the identifiability results in the earlier sections, where we observed that a certain number of equality constraints are needed to obtain finitely many or a unique reconstruction.

To have some idea of how good the results are in relation to the number of beads we computed the Procrustes distance allowing scaling for 50 repetitions of the experiment

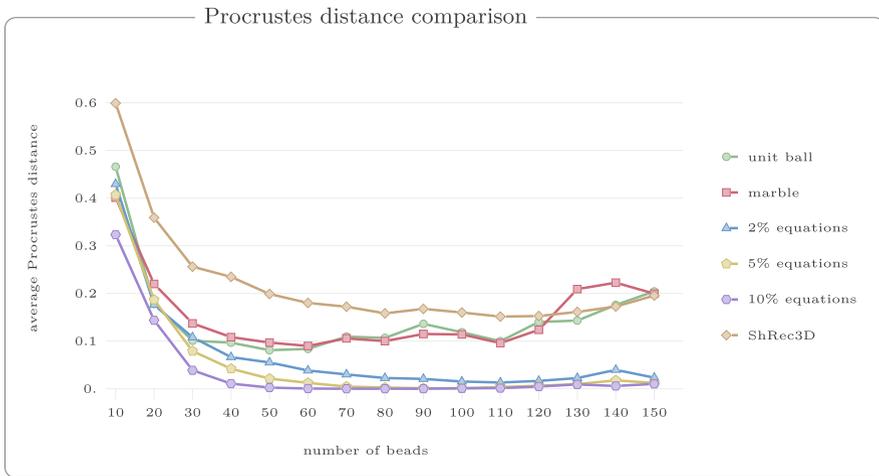


Fig. 16 Comparison of the Procrustes distance. The value shows the average of all the Procrustes distances of 50 repetitions for different methods

with simulated data for 10 to 150 beads. The data was generated using $r = 3$. We also compare the results with ShRec3D algorithm from Lesne et al. (2014) that combines shortest path computations with multidimensional scaling. The result can be seen in Fig. 16. For these small examples our methods turn out to have lower Procrustes distance than the one obtained from ShRec3D.

In Fig. 17 we compare the time needed for different number of beads. The figure shows the average time in seconds over 50 repetitions of the experiment with simulated data. A limitation of the SDP based methods is that they are relatively slow, so we consider up to 150 beads. In this case, ShRec3D algorithm is much faster. In summary, the SDP based methods using models considered in this paper give better reconstructions but they are considerably slower compared to ShRec3D.

Real data. Stevens et al. (2017) combined imaging with a HiC protocol to obtain 3D structures for single G1 phase haploid mouse embryonic stem cells (data accessible at NCBI GEO database Edgar et al. 2002, accessionGSE80280). We applied our reconstruction algorithm with unit ball model on Chromosome 1 in Sample GSM2219497 of this dataset. The chromosome consists of $\sim 193 \cdot 10^6$ base pairs. Each bead in our model corresponds to one million base pairs. There is a contact between two beads if and only if there is at least one contact between base pairs in the corresponding beads. We assume that contacts correspond to distances $r \leq 2$. Our data consists of 193 beads and 802 contacts. Note that this aggregation of data destroys possible backbone information. Transferring the information suitably to the aggregated data is the subject of further investigation. Here, we restrict attention to the case where the backbone information is lost. The 3D reconstruction obtained using the unit ball model without backbone information is depicted in Fig. 18. Figure 18a shows the backbone of the reconstruction and Fig. 18b the contacts in the reconstruction. Extended Data Figure 1d from Stevens et al. (2017) shows an analogous reconstruction at 100kb scale with contacts that are more than distance four apart coloured red. Figure 19 shows

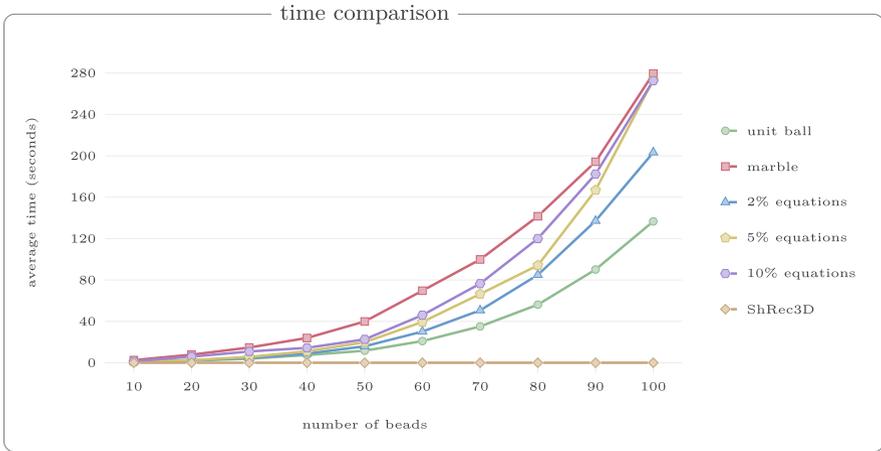


Fig. 17 Time comparison of the methods using backbone data on synthetic examples from before. The time is given on average and is shown in seconds for 50 repetitions.

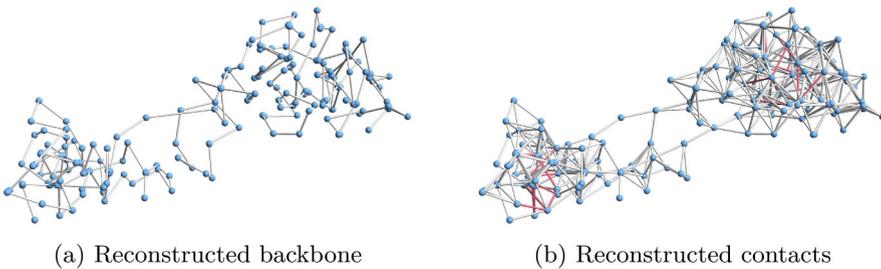


Fig. 18 Reconstruction from a real data set using our reconstruction algorithm for the unit ball model without backbone information. **a** Backbone in the reconstruction. **b** Contacts in the reconstruction, i.e. edges are shown between beads that are in contact according to the aggregated HiC data. The edges which are too large by more than 10% in the reconstruction are coloured in red

the distribution of distances in the reconstructed structure for pairs of beads that are in contact and not in contact as a histogram. The maximal distance of a contact in the reconstruction is ~ 2.37643 and the minimal distance of a non-contact in the reconstruction is ~ 1.48434 .

8 Discussion

In this article, we analysed the identifiability and reconstruction algorithms for single-cell 3D genome reconstruction in haploid organisms. The main contribution of the paper is introducing a correspondence between the biology of 3D genome reconstruction and rigidity theory. We study four different rigidity theoretic models corresponding to different types, or combinations, of biological measurements: unit ball graphs, clas-

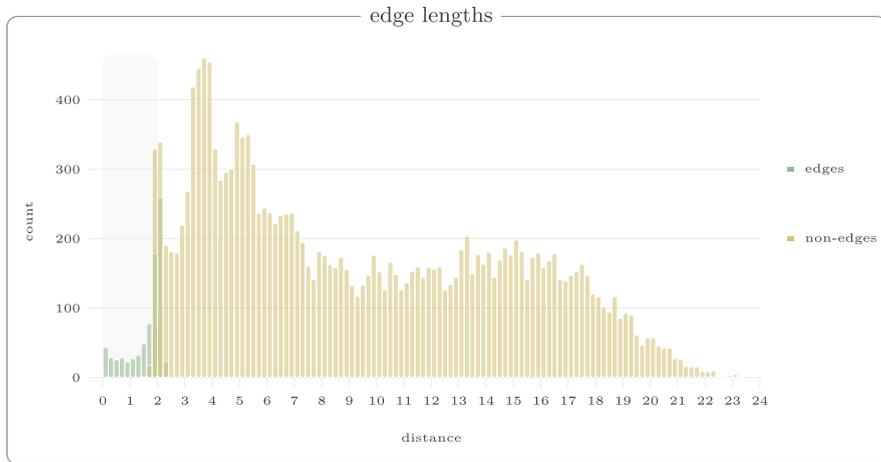


Fig. 19 Histogram showing the edge lengths and non-edge-lengths of the real data reconstruction. The shaded area corresponds to the threshold below which contacts should occur. We see the actual edge lengths (green) and distances of non-edges (yellow) and their respective amounts

sical rigidity together with inequality constraints, penny/marble graphs and interval radii penny and marble graphs.

The unit ball graph model only requires that the pairwise distances corresponding to contacts are less than or equal to some constant, and the rest of the pairwise distances are larger than the constant. It is the simplest and least restrictive of the models, and perhaps biologically the one that makes the most sense. However, we proved that the 3D reconstruction is never finitely identifiable under this model. Our reconstruction experiments illustrate this as the 3D reconstructions do not match with the original structures in detail. Nevertheless, the high-level details are correct and the contact matrices are similar for the original and reconstructed structures, so acceptable reconstructions might be still possible under this model. One direction to explore is the geometry and topology of the space of all the valid 3D reconstructions.

The second model that we studied assumes that we know some pairwise distances and some bounds for the pairwise distances between the beads. The analysis of this model is closely related to the classical rigidity setting, where some pairwise distances between a set of points are known. We show that finite and unique identifiability of 3D reconstructions are possible under this model. The finite identifiability of the reconstruction is determined only by the pairwise distances; the unique identifiability depends both on the pairwise distances and bounds on the distances. We obtain good reconstructions also in our experiments, especially when we know enough pairwise distances. While mathematically this model has desirable properties, in practice it might be difficult to obtain the pairwise distance measurements.

Finally, we study penny/marble graph models that are special cases of the previous model where all known pairwise distances are equal to one and all other pairwise distances are bounded below by one. This model is more restrictive than the unit ball model, as all contacts correspond to equal pairwise distances, and therefore this model

is biologically less realistic. The motivation for studying this model is the observation that identifiability is possible under this model, although studying it is difficult and not much is known. The synthetic reconstruction experiments gave the worst results under this model. We do not know if the reason is that in the particular cases the reconstructions were not identifiable or whether the assumptions of this model are too restrictive. Based on our analysis, we do not recommend using the penny/marble graph model.

Our more specialised interval radii penny/marble graph model allows for interesting mathematics. However a limitation of our reconstruction algorithm is that it does not apply to this model.

As we mentioned in Sect. 2, the assumption that the absence of an interaction suggests that the beads are more than d_c units apart in this and other models corresponds to the idealized setting when all contacts are sampled in an Hi-C experiment. Most of our identifiability results remain the same without this assumption. The unit ball model is never finitely identifiable by Lemma 3.4 and the same remains true when the assumption about non-contacts is dropped. The main uniqueness result for the model with some distance equalities and inequalities Proposition 4.2 does not specify the exact constraints. A consequence of this result is that if we have unique identifiability with the non-contact constraints, then we might get finite identifiability when dropping these constraints. The effect of an imperfect Hi-C matrix for the penny/marble graph model and the ε -interval radii model is to replace every strict inequality for a non-edge with a non-strict inequality. Again, this can only increase the number of solutions, and hence identifiability could be replaced by finitely or infinitely many solutions. In the penny/marble graph model it is also possible for a graph with no solutions to gain solutions, or for an identifiable system to gain infinitely-many solutions. Interestingly, neither of these latter possibilities can occur for the ε -interval radii penny graph model. This is due to the genericity of the radii in conjunction with results of Connelly et al. (2019).

This paper focused on single-cell 3D genome reconstruction in the haploid setting. A future research direction is to conduct a similar study in the diploid setting. We expect the diploid setting, especially the problem of obtaining identifiability results under reasonable constraints, to be considerably more challenging.

Acknowledgements We thank Caroline Uhler for suggesting to study single cell 3D genome reconstruction and Annachiara Korchmaros for helping to improve the paragraph on Hi-C experiments.

Funding Open Access funding provided by Aalto University. S.D. was supported by the Heilbronn Institute for Mathematical Research. G.G. was partially supported by the Austrian Science Fund (FWF): 10.55776/P31888. K.K. was partially supported by the Academy of Finland grant number 323416. F.M. was partially supported by the FWO grants G0F5921N and G023721N, the KU Leuven iBOF/23/064 grant, and the UiT Aurora MASCOT project. A.N. was partially supported by EPSRC grant EP/X036723/1.

Data availability The code used to run the experiments in Sect. 7.2 is available on GitHub: <https://github.com/kaiekubjas/haploid-single-cell-3D-genome-reconstruction>. The real data analysed in Sect. 7.2 is associated to the article by Stevens et al. (2017) and it is accessible at NCBI GEO database of Edgar et al. (2002), accession GSE80280.

Declarations

Conflicts of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbas A, He X, Niu J, Zhou B, Zhu G, Ma T, Song J, Gao J, Zhang MQ, Zeng J (2019) Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. *Nat Commun* 10(1):1–14. <https://doi.org/10.1038/s41467-019-10005-6>
- Amann RI, Krumholz L, Stahl DA (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J Bacteriol* 172(2):762–770. <https://doi.org/10.1128/jb.172.2.762-770.1990>
- Asimow L, Roth B (1978) The rigidity of graphs. *Trans Am Math Soc* 245:279–289. <https://doi.org/10.2307/1998867>
- Atminas A, Zamaraev V (2018) On forbidden induced subgraphs for unit disk graphs. *Discr Comput Geom* 60(1):57–97. <https://doi.org/10.1007/s00454-018-9968-1>
- Bezdek K (2012) Contact numbers for congruent sphere packings in Euclidean 3-space. *Discret Comput Geom* 48(2):298–309. <https://doi.org/10.1007/s00454-012-9405-9>
- Breu H, Kirkpatrick DG (1996) On the complexity of recognizing intersection and touching graphs of disks. In: Brandenburg FJ (ed) *Graph drawing*. Springer, Berlin, pp 88–98. <https://doi.org/10.1007/BFb0021793>
- Breu H, Kirkpatrick DG (1998) Unit disk graph recognition is NP-hard. *Comput Geom* 9(1):3–24. [https://doi.org/10.1016/S0925-7721\(97\)00014-X](https://doi.org/10.1016/S0925-7721(97)00014-X)
- Belyaeva A, Kubjas K, Sun LJ, Uhler C (2022) Identifying 3D genome organization in diploid organisms via Euclidean distance geometry. *SIAM J Math Data Sci* 4(1):204–228. <https://doi.org/10.1137/21M1390372>
- Biswas P, Liang T-C, Toh K-C, Ye Y, Wang T-C (2006) Semidefinite programming approaches for sensor network localization with noisy distance measurements. *IEEE Trans Autom Sci Eng* 3(4):360–371. <https://doi.org/10.1109/TASE.2006.877401>
- Bezdek K, Reid S (2013) Contact graphs of unit sphere packings revisited. *J Geom* 104:57–83. <https://doi.org/10.1007/s00022-013-0156-4>
- Cassoli A, Bardiaux B, Bouvier G, Mucherino A, Alves R, Liberti L, Nilges M, Lavor C, Malliavin TE (2015) An algorithm to enumerate all possible protein conformations verifying a set of distance constraints. *BMC Bioinf* 16:1–15. <https://doi.org/10.1186/s12859-015-0451-1>
- Cox MA, Cox TF (2008) Multidimensional scaling. In: *Handbook of data visualization*, pp. 315–347. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-33037-0_14
- Cifuentes D, Draisma J, Henriksson O, Korchmaros A, Kubjas K (2024) 3D genome reconstruction from partially phased Hi-C data. *Bull Math Biol* 86(33):1–30. <https://doi.org/10.1007/s11538-024-01263-7>
- Connelly R, Gortler SJ, Theran L (2019) Rigidity for sticky discs. *Proc R Soc A*. <https://doi.org/10.1098/rspa.2018.0773>
- Chen H (2016) Apollonian ball packings and stacked polytopes. *Discr Comput Geom* 55:801–826. <https://doi.org/10.1007/s00454-016-9777-3>
- Connelly R (2005) Generic global rigidity. *Discr Comput Geom* 33(4):549–563. <https://doi.org/10.1007/s00454-004-1124-4>

- Dekker J (2008) Gene regulation in the third dimension. *Science* 319(5871):1793–1794. <https://doi.org/10.1126/science.1152850>
- Dewar S (2023) Identifying contact graphs of sphere packings with generic radii. <https://doi.org/10.48550/arXiv.2302.12588>
- Dewar S, Grasegger G, Kubjas K, Mohammadi F, Nixon A (2025) On the uniqueness of collections of pennies and marbles. *Ex Counterexamples* 7:100181. <https://doi.org/10.1016/j.exco.2025.100181>
- Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1):207–210. <https://doi.org/10.1093/nar/30.1.207>
- Fazel M, Hindi H, Boyd SP (2001) A rank minimization heuristic with application to minimum order system approximation. In: *Proceedings of the 2001 American control conference*, vol. 6, pp. 4734–4739. <https://doi.org/10.1109/ACC.2001.945730>. IEEE
- Gortler SJ, Healy AD, Thurston DP (2010) Characterizing generic global rigidity. *Am J Math* 132(4):897–939. <https://doi.org/10.1353/ajm.0.0132>
- Garamvolgyi D, Jordán T (2020) Global rigidity of unit ball graphs. *SIAM J Discret Math* 34(1):212–229. <https://doi.org/10.1137/18M1220662>
- Harborth H (1974) Lösung zu problem 664a. *Elem Math* 29(1):14–15
- Holmes-Cerfon MC (2016) Enumerating rigid sphere packings. *SIAM Rev* 58(2):229–244. <https://doi.org/10.1137/140982337>
- Hu M, Deng K, Qin Z, Dixon J, Selvaraj S, Fang J, Ren B, Liu JS (2013) Bayesian inference of spatial organizations of chromosomes. *PLoS Comput Biol* 9(1):1002893. <https://doi.org/10.1371/journal.pcbi.1002893>
- Hendrickson B (1992) Conditions for unique graph realizations. *SIAM J Comput* 21(1):65–84. <https://doi.org/10.1137/0221008>
- Hliněný P, Kratochvíl J (2001) Representing graphs by disks and balls (a survey of recognition-complexity results). *Discret Math* 229(1):101–124. [https://doi.org/10.1016/S0012-365X\(00\)00204-1](https://doi.org/10.1016/S0012-365X(00)00204-1)
- Hliněný P (1997) Touching graphs of unit balls. In: DiBattista D (ed) *Graph drawing*. Springer, Berlin, pp 350–358. https://doi.org/10.1007/3-540-63938-1_80
- Jacobs DJ, Hendrickson B (1997) An algorithm for two-dimensional rigidity percolation: the pebble game. *J Comput Phys* 137(2):346–365. <https://doi.org/10.1006/jcph.1997.5809>
- Jackson B, Jordán T (2005) The d-dimensional rigidity matroid of sparse graphs. *J Combin Theory Ser B* 95(1):118–133. <https://doi.org/10.1016/j.jctb.2005.03.004>
- Jackson B, Jordán T (2005) Connected rigidity matroids and unique realizations of graphs. *J Combin Theory Ser B* 94(1):1–29. <https://doi.org/10.1016/j.jctb.2004.11.002>
- Jackson B, McCourt TA, Nixon A (2014) Necessary conditions for the generic global rigidity of frameworks on surfaces. *Discr Comput Geom* 52(2):344–360. <https://doi.org/10.1007/s00454-014-9616-3>
- Kertész G (1994) Nine points on the hemisphere. In: Böröczky K, Fejes Tóth G (eds) *Intuitive geometry (Szeged, 1991)*, vol 63. *Colloquia Mathematica Societatis János Bolyai*. North Holland Publishing co., Amsterdam, pp 189–196
- Kos PI, Galitsyna AA, Ulianov SV, Gelfand MS, Razin SV, Chertovich AV (2021) Perspectives for the reconstruction of 3D chromatin conformation using single cell Hi-C data. *PLoS Comput Biol* 17(11):1009546. <https://doi.org/10.1371/journal.pcbi.1009546>
- Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950):289–293. <https://doi.org/10.1126/science.1181369>
- Lekkerkerker CG, Boland JC (1962) Representation of a finite graph by a set of intervals on the real line. *Fundam Math* 51(1):45–64. <https://doi.org/10.4064/fm-51-1-45-64>
- Liberti L, Lavor C, Maculan N, Mucherino A (2014) Euclidean distance geometry and applications. *SIAM Rev* 56(1):3–69. <https://doi.org/10.1137/120875909>
- Lesne A, Riposo J, Roger P, Courmac A, Mozziconacci J (2014) 3D genome reconstruction from chromosomal contacts. *Nat Methods* 11(11):1141–1143. <https://doi.org/10.1038/nmeth.3104>
- Leung N-HZ, Toh K-C (2010) An SDP-based divide-and-conquer algorithm for large-scale noisy anchor-free graph realization. *SIAM J Sci Comput* 31(6):4351–4372. <https://doi.org/10.1137/080733103>
- Maehara H (2007) On configurations of solid balls in 3-space: chromatic numbers and knotted cycles. *Gr Combin* 23:307–320. <https://doi.org/10.1007/s00373-007-0702-7>
- Maxwell JC (1864) On the calculation of the equilibrium and stiffness of frames. *London Edinburgh Dublin Philos Mag J Sci* 27(182):294–299. <https://doi.org/10.1080/14786446408643668>

- Milnor J (1965) Topology from the differentiable viewpoint. University Press of Virginia, Charlottesville
- Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502(7469):59–64. <https://doi.org/10.1038/nature12593>
- Norton HK, Phillips-Cremins JE (2017) Crossed wires: 3D genome misfolding in human disease. *J Cell Biol* 216(11):3441–3452. <https://doi.org/10.1083/jcb.201611001>
- Oluwadare O, Highsmith M, Cheng J (2019) An overview of methods for reconstructing 3-D chromosome and genome structures from Hi-C data. *Biol Proc Online* 21(1):1–20. <https://doi.org/10.1186/s12575-019-0094-0>
- Ozkan A, Prabhu R, Baker T, Pence J, Peters J, Sitharam M (2018) Algorithm 990: efficient atlasing and search of configuration spaces of point-sets constrained by distance intervals. *ACM Trans Math Softw* 44(4):1–30. <https://doi.org/10.1145/3204472>
- Payne AC, Chiang ZD, Reginato PL, Mangiameli SM, Murray EM, Yao C-C, Markoulaki S, Earl AS, Labade AS, Jaenisch R et al (2021) In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* 371(6532):3446. <https://doi.org/10.1126/science.aay3446>
- Pollaczek-Geiringer H (1927) Über die Gliederung ebener Fachwerke. *Z Angew Math Mech* 7(1):58–72. <https://doi.org/10.1002/zamm.19270070107>
- Paulsen J, Gramstad O, Collas P (2015) Manifold based optimization for single-cell 3D genome reconstruction. *PLoS Comput Biol* 11(8):1004396. <https://doi.org/10.1371/journal.pcbi.1004396>
- Paulsen J, Sekelja M, Oldenburg AR, Barateau A, Briand N, Delbarre E, Shah A, Sørensen AL, Vigouroux C, Buendia B et al (2017) Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. *Genome Biol* 18(1):1–15. <https://doi.org/10.1186/s13059-016-1146-2>
- Pfender F, Ziegler GM (2004) Kissing numbers, sphere packings, and some unexpected proofs. *Notices AMS* 51(8):873–883
- Rosenthal M, Bryner D, Huffer F, Evans S, Srivastava A, Neretti N (2019) Bayesian estimation of three-dimensional chromosomal structure from single-cell Hi-C data. *J Comput Biol* 26(11):1191–1202. <https://doi.org/10.1089/cmb.2019.0100>
- Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, Noble WS, Duan Z, Shendure J (2017) Massively multiplex single-cell Hi-C. *Nat Methods* 14(3):263–266. <https://doi.org/10.1038/nmeth.4155>
- Roberts FS (1969) Indifference graphs. In: *Proof Techniques in Graph Theory (Proceedings of the Second Ann Arbor Graph Theory Conference, 1968)*, pp. 139–146. Academic Press, New York
- Schoenberg IJ (1935) Remarks to Maurice Fréchet’s article “sur la definition axiomatique d’une classe d’espace distances vectoriellement applicable sur l’espace de Hilbert”. *Annals of Mathematics*, 724–732 <https://doi.org/10.2307/1968654>
- Segal MR (2022) Can 3D diploid genome reconstruction from unphased Hi-C data be salvaged? *NAR Genom Bioinf* 4(2):038. <https://doi.org/10.1093/nargab/lqac038>
- Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O’Shaughnessy-Kirwan A et al (2017) 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* 544(7648):59–64. <https://doi.org/10.1038/nature21429>
- Shi G, Thirumalai D (2021) From Hi-C contact map to three-dimensional organization of interphase human chromosomes. *Phys Rev X* 11(1):011051. <https://doi.org/10.1103/PhysRevX.11.011051>
- Schütte K, Waerden BL (1952) Das Problem der dreizehn Kugeln. *Mathematische Annalen* 125:325–334. <https://doi.org/10.1007/BF01343127>
- So AM-C, Ye Y (2007) Theory of semidefinite programming for sensor network localization. *Math Program* 109(2):367–384. <https://doi.org/10.1007/s10107-006-0040-1>
- Trieu T, Cheng J (2014) Large-scale reconstruction of 3D structures of human chromosomes from chromosomal contact data. *Nucleic Acids Res* 42(7):52–52. <https://doi.org/10.1093/nar/gkt1411>
- Tan L, Xing D, Chang C-H, Li H, Xie XS (2018) Three-dimensional genome structures of single diploid human cells. *Science* 361(6405):924–928. <https://doi.org/10.1126/science.aat5641>
- Uhler C, Shivashankar GV (2017) Regulation of genome organization and gene expression by nuclear mechanotransduction. *Nat Rev Mol Cell Biol* 18(12):717–727. <https://doi.org/10.1038/nrm.2017.101>
- Varoquaux N, Ay F, Noble WS, Vert J-P (2014) A statistical approach for inferring the 3D structure of the genome. *Bioinformatics* 30(12):26–33. <https://doi.org/10.1093/bioinformatics/btu268>
- Wettermann S, Brems M, Siebert JT, Vu GT, Stevens TJ, Virnau P (2020) A minimal G = o-model for rebuilding whole genome structures from haploid single-cell Hi-C data. *Comput Mater Sci* 173:109178. <https://doi.org/10.1016/j.commatsci.2019.109178>

- Weinberger KQ, Saul LK (2006) Unsupervised learning of image manifolds by semidefinite programming. *Int J Comput Vision* 70(1):77–90. <https://doi.org/10.1007/s11263-005-4939-z>
- Zhang Z, Li G, Toh K-C, Sung W-K (2013) Inference of spatial organizations of chromosomes using semi-definite embedding approach and Hi-C data. In: Deng M, Jiang R, Sun F, Zhang X (eds) *Research in computational molecular biology. RECOMB*. Springer, Berlin, pp 317–332. https://doi.org/10.1007/978-3-642-37195-0_31

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.