Snapshots of modern mathematics
from Oberwolfach

№ 5/2026

The 4-Sample Theorem on planar graphs

Carlos Améndola • Thomas Kahle

The famous *4-Color Theorem* from graph theory states that the vertices of any planar graph can be colored with four colors, so that no neighboring vertices have the same color. The *4-Sample Theorem* from algebraic statistics says that the maximum likelihood estimator for a *Gaussian graphical model* of a planar graph exists with probability 1 if one has at least four samples. This number of necessary samples, the *maximum likelihood threshold*, is a new graph invariant from algebraic statistics and connected not only to parameter estimation, but also to matrix completion, the theory of filling partial matrices, and rigidity theory, which deals with stability of objects.

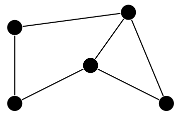
1 The 4-Sample Theorem

Statistics and algebra look back on a long joint history, maybe starting with Pearson's algebraic analysis of the distribution of crabs in the Bay of Naples.[1] In the 21st century, algebraic statistics is both an exciting branch of statistics [10] and a vibrant community that continuously uncovers new connections between

[1] Pearson analyzed the ratio of forehead breadth and body length of the crabs and employed algebraic methods in order to find some statistical parameters. [8]

data science and algebraic, geometric, and discrete structures in mathematics. Here we want to tell one such story, namely how the theory of estimation for "multivariate normal distributions", through topics like matrix completion and rigidity theory, is connected to basic invariants of graphs.

Graphs are mathematical objects that formalize the idea of elements of a set being pairwise related. Consider, for example, the public transportation map in your town: you have a collection of stops (the *vertices*), and two stops are connected (there is an *edge* between them) if a route links them. Graphs are visualized by diagrams like this:



Formally, a graph $G = (V, E)$ consists of a set of vertices V and a set of edges E as in the picture. We only consider finite graphs. Thus, our set of vertices is always a finite set $V = \{1, \ldots, m\}$ for some $m \in \mathbb{N}$. If there is an edge between two vertices $i, j \in V$, we write $(i, j) \in E$. In this snapshot we are mostly interested in planar graphs, that is, graphs that can be drawn in the plane so that no two edges cross.

In what follows, we explore "Gaussian graphical models" – statistical models that use graphs to capture the dependencies between Gaussian random variables. Our main goal is to establish the *4-Sample Theorem*, which contains the surprising insight, that a finite number of samples typically suffices for the parameters of the model to be "estimable". Our path towards proving this theorem takes us from multivariate statistics all the way to the theory of rigid polyhedra. The (very convincing) name of this theorem was suggested by Steffen Lauritzen during the 2022 Oberwolfach Workshop *Algebraic Structures in Statistical Methodology.*

In Gaussian graphical models, the number of vertices of the graph G corresponds to the number of Gaussian random variables (which we call *covariates*). One exciting consequence of the theorem is that the number of required samples does not depend on the number of covariates. This is important in biological applications and machine learning, which both feature a very large number of variables and very few samples.

2 Gaussian graphical models

Consider a random vector $X = (X_1, X_2, \ldots, X_m)$, which has m random variables as components, with a *multivariate normal distribution*. Multivariate normal distributions generalize normal distributions to higher dimensions. A normally distributed random variable takes values in $\mathbb{R}$ and is determined by its mean and variance: its values concentrate around the mean, while the variance quantifies how widely the values are spread. In the multivariate case, this means that the probability that X takes values in a subset of $\mathbb{R}^m$ is determined by a *mean vector* $\mu \in \mathbb{R}^m$ and a *covariance matrix* $\Sigma \in \mathbb{R}^{m\times m}$, which is *symmetric*[2] and *positive definite*[3] (PD). Analogously to the one-dimensional normal distribution, the mean vector specifies the center of the multivariate normal distribution. The covariance matrix describes the variation of a multivariate normal distribution in the different directions. To simplify matters, we can assume here that each of our variables is centered, so that $\mu = 0$ in what follows and the distribution is completely determined by $\Sigma \in \mathrm{PD}_m$.[4]

We receive n samples of X and wish to estimate the distribution of X. Here we should think of m as being large compared to the sample size n. Such an estimation is most often based on model assumptions. This means that we assume certain rules to hold, which could for example be empirically observed. An extensively used class of models is the class of *Gaussian graphical models*. There, we use graphs $G = (V, E)$ on the vertex set $V = \{1, 2, \ldots, m\}$ and we assume that each vertex corresponds to one component of the random vector X. Intuitively, an edge in G should represent dependence of the corresponding components of X. Here we mean dependence in the sense of probability theory: one random variable depends on the other if the value it takes correlates to the value the other one takes.

The Gaussian graphical model associated to G is the set of all multivariate normal distributions whose covariance matrices lie in the set

$$\mathcal{M}(G) = \left\{\Sigma \in \mathrm{PD}_m : \Sigma^{-1}_{ij} = 0 \text{ for } (i,j) \notin E\right\}.$$

In other words, we say that a covariance matrix belongs to the model if its inverse has a 0 in the entry (i, j) whenever there is no edge between i and j in G.[5] For example, consider $G = C_4$ to be the 4-cycle below:

$$\begin{array}{ccc} 1 & — & 2 \\ | & & | \\ 4 & — & 3 \end{array}$$

[2] A matrix is called symmetric if its entries are symmetric with respect to the diagonal.

[3] Geometrically, positive definiteness means that a matrix does not flip a vector across the origin when applied. For our covariance matrix this means that the components of the random vector correlate "smoothly" without any flips.

[4] Positive definite matrices of size $m \times m$.

[5] This condition is natural because it is equivalent to the statement that X_i is "conditionally independent" of X_j given the rest of the components of the Gaussian vector X.

Then,

$$\mathcal{M}(C_4) = \left\{ \Sigma \in \mathrm{PD}_4 : \Sigma^{-1} = \begin{pmatrix} * & * & 0 & * \\ * & * & * & 0 \\ 0 & * & * & * \\ * & 0 & * & * \end{pmatrix} \right\}.$$

3 Estimating the graph from data

In practical applications, the graph G is often unknown. The estimation would then proceed in two steps. First, a graph G is estimated.[6] Then, maximum likelihood estimation[7] for the graphical model $\mathcal{M}(G)$ is performed based on that graph. The goal of maximum likelihood estimation is to estimate the unknown parameters of a model in a way that makes the observed data most likely. In our case, we assume the data follows a multivariate normal distribution, with the unknown parameters being the mean vector and the covariance matrix. Therefore, the objective is to estimate these parameters such that the observed data vectors are most likely to occur under the assumed distribution. This optimization problem comes with its own difficulties, which we will not delve into here. Notice, however, that a *maximum likelihood estimate* (MLE) might not exist. Its existence is one of the central problems addressed in this snapshot.

A concrete example of this approach is in [12]. The authors consider a climate network grid with 2562 nodes (variables) and they have 756 samples, which are real vectors of length 2562. They want to estimate both the graph structure and the parameters of the corresponding Gaussian graphical model. They employ the graphical LASSO, and they find that the MLE does exist.

The results in this snapshot explain why and how the existence of the MLE is connected to the structure of the graph. As always in mathematics, it is useful to first give a name to the thing one wants to study. Algebraic statisticians have coined the term *maximum likelihood threshold* (MLT) for the number of samples that is necessary to have an MLE with probability 1. Interestingly, this number is independent of the concrete data (up to events of probability 0). An even more interesting fact is that this number only depends on certain graph-theoretic invariants and thus becomes independent of the concrete number of covariates, as long as the graph structure does not change too much. For example, in any square grid every vertex has exactly four neighbors, independently of its size. The fact that the MLT does not grow as the number of covariates grows, gives an advantage over classical results in statistics that do not consider the graph structure. One such classical result is that, if the number of samples exceeds

---

[6] A common method for this step is the "graphical LASSO" [6].

[7] We will not go into the details of maximum likelihood estimation here. It is arguably the most common estimation method in parametric statistics and also featured in Snapshot 1/2018 by Anna Seigal [9].

the number of covariates, then the MLE exists with probability 1.[8]

The input for algorithms which we use to approximate the covariance matrix is the *empirical covariance matrix* S. This is an approximation to the covariance matrix computed from the vector-valued observations. This matrix need not be positive definite (like an actual covariance matrix). In fact, if the number of samples is less than the number of covariates, it can never be positive definite. In such cases we have the following *completion theorem* [5, Theorem 2.1]: the MLE exists with probability 1 in $\mathcal{M}(G)$ if and only if the empirical covariance matrix S restricted to the diagonal entries and those corresponding to edges of G has a PD-completion. Here, the *PD-completion* is a positive definite matrix that has the same entries like S on the diagonal and in those positions whose index pair (i, j) corresponds to an edge in the graph G. If the number of samples exceeds the number of variables, the empirical covariance matrix itself is PD.

To illustrate the PD-completion, consider the small 2×3 grid below. The six vertices mean that we deal with a 6×6-completion problem.

$$\begin{array}{ccccc} 1 & — & 2 & — & 3 \\ | & & | & & | \\ 4 & — & 5 & — & 6 \end{array}$$

As it turns out, for any graphical model of a rectangular grid, four samples suffice for the MLE to exist with probability 1, no matter the size of the grid. Consider the following partial matrices, where each could be the empirical covariance matrix of some given data:

$$\begin{pmatrix} 6 & 1 & & -2 & & \\ 1 & 4 & 2 & & 3 & \\ & 2 & 2 & & & 3 \\ -2 & & & 5 & 1 & \\ & 3 & & 1 & 5 & -5 \\ & & 3 & & -5 & 10 \end{pmatrix} \qquad \begin{pmatrix} 6 & 1 & & -2 & & \\ 1 & 4 & 2 & & 3 & \\ & 2 & 2 & & & 3 \\ -2 & & & 5 & 1 & \\ & 3 & & 1 & 5 & -5 \\ & & 3 & & -5 & 5 \end{pmatrix}$$

The entries that are left out correspond exactly to the non-edges of the grid. To find a PD-completion means filling in these entries so that the resulting matrix is positive definite. Whether this is possible can depend subtly on the matrix. Indeed, even though they differ only on their last entry, the matrix on the left has a PD-completion, while the matrix on the right does not. The advanced reader might try to verify this claim.

The key insight is that *generically*, we are always in the situation that the completion exists, once the empirical covariance matrix has been computed from enough samples. The sporadic cases for which this is not true happen with probability 0 and arise, for example, if by chance all samples would be exactly equal. However, for many modern applications, for example in biology,

[8] This goes back to the theory of “exponential families” of Barndorff-Nielsen and Dempster.

the number of variables m is *very large*, and gathering samples is *very expensive*. It is therefore natural to ask for MLE existence results that are independent of the number of samples. That such bounds exist and that they are connected to graph theory became clear already in the 1990s in [5], but only recently, through the connection to and renewed interest in matrix completion, several breakthroughs have been possible. In particular, the matrix completion problem has been connected to "rigidity theory" of graphs in the work of Bernstein, Gross, Sullivant, and others [4, 3, 7].

4 The maximum likelihood threshold

As explained above, the maximum likelihood threshold of a graph G is the least number of samples required so that the maximum likelihood estimate in the Gaussian graphical model associated to the graph G exists with probability 1. We denote this number by $\mathrm{mlt}(G)$. The definition of mlt is attractive because it is a *graph invariant* – $\mathrm{mlt}(G)$ is a specific number that only depends on the structure of the graph G. There is no general method known to compute $\mathrm{mlt}(G)$ from G. So far the community has focused on providing bounds.

From the completion theorem we can see that $\mathrm{mlt}(G)$ is bounded by the number of vertices of G: if the number of samples exceeds the number of variables, the empirical covariance matrix is PD with probability 1 and therefore the MLE exists with probability 1. This bound is achieved for the complete graph K_m, which is the graph containing every possible edge on m vertices: $\mathrm{mlt}(K_m) = m$. For applications such as the ones outlined above, it is desirable to bound $\mathrm{mlt}(G)$ *independently of the number of vertices*, or in terms of graph invariants.

The algebraic statistics community got interested in this problem around 2008, when Steffen Lauritzen posted the benchmark problem of understanding the maximum likelihood threshold for the $3{\times}3$-grid, see [11]. Improved combinatorial bounds have been achieved by transforming the problem again, this time to a geometric problem! The question of determining $\mathrm{mlt}(G)$ can be formulated as a question about geometric structures made of rigid bars connected at freely rotating joints. This is *rigidity theory*. We come back to it in Section 6 after talking about planar graphs, which are crucial for the 4-Sample Theorem.

5 Planar graphs

The first reduction in the proof of the 4-Sample Theorem arises from considering only *maximal planar graphs*. These are planar graphs, such that adding any further edge makes them non-planar. To every planar graph, one can add edges until it is a maximal planar graph. It can be checked that the number of edges of a maximal planar graph with m vertices is $3m - 6$. If G is maximal planar, then all regions arising when drawing the graph have exactly three sides. This also applies to the unbounded outside region. Additionally, any maximal planar graph is 3-*connected*, meaning that it stays *connected*[9] after removing two or fewer vertices. An interesting maximal planar graph is the Goldner–Harary graph in Figure 1.

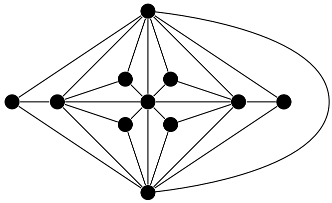


Figure 1: The Goldner–Harary graph is a maximal planar graph on 11 vertices with $27 = 3 \cdot 11 - 6$ edges.

Let us now finally formulate the 4-Sample Theorem precisely:

Theorem (4-Sample Theorem, [7, Corollary 3.11]). *Let G be a planar graph. If there are at least four samples from a distribution in the Gaussian graphical model of G, the maximum likelihood estimator exists with probability 1.*

To prove the 4-Sample Theorem, we can assume that G is maximal planar: Adding edges to a planar graph in order to obtain a maximal planar graph corresponds to prescribing more non-0s for the matrices belonging to the Gaussian graphical model associated to the graph. Therefore, a PD-completion of the empirical covariance matrix that belongs to the model associated to the maximal planar graph also belongs to the model associated to the actual graph. Hence, if the MLE exists with probability 1 for the maximal planar graph, then it also exists with probability 1 for the actual graph.

6 Graphs and frameworks

Our goal below is to reduce the 4-Sample Theorem to another breakthrough theorem of two algebraic statisticians: Elizabeth Gross and Seth Sullivant. Their result below is about *frameworks*, which are embeddings of the vertices (and thus the entire graph) into a space where one can then argue about angles and lengths. To be more precise, a *framework* in $\mathbb{R}^d$ is a pair (G, p) where G is

[9] A graph is called *connected* if you can start at any vertex and reach any other vertex by following edges.

a graph with m vertices and p is a collection of m points $p_1, \ldots, p_m \in \mathbb{R}^d$. Two frameworks (G, p) and (G, q) for the same graph are called *equivalent* if for each edge $(i, j) \in E$ the distance between the points i and j agrees. Equivalence is a somewhat weaker notion than that of congruence. Two frameworks are called *congruent* if one can be transformed into the other without changing distances and angles.

To formulate rigidity of frameworks, one considers continuous changes to the points $p_1, \ldots, p_m$ so that all frameworks along the continuous deformation are equivalent. That is, one moves the points around in space such that the distances stay the same. A framework is called *rigid* if any two equivalent frameworks arising from such a continuous deformation are congruent. For example, a triangle is rigid but a square is not, as seen in Figure 2.

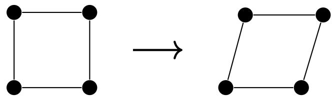


Figure 2: Two equivalent frameworks of the 4-cycle in $\mathbb{R}^2$ that are not congruent.

Let us also consider the following complementary property. A framework is called *independent* if for any choice of an edge and any sufficiently small $\varepsilon > 0$, there exists a framework for the same graph but such that the chosen edge has its length modified by ε while all other lengths stay the same. A graph G is called *d-independent* if d is the minimal dimension for which an independent framework for the graph G exists. For example, the 4-cycle is 2-independent because it can be embedded in the plane as an independent framework, albeit not rigidly. Any edge can be made a little shorter or longer, while keeping the combinatorics and the other edge lengths (just the angles would change).[10] The diamond graph, see Figure 3, with one added edge is still 2-independent, but the complete graph is not.[11] The intuition is that a framework is more rigid if it has more edges and more independent if it has fewer edges.

While it is conceivable that the rigidity of a framework for a given graph could depend on the concrete framework chosen, this is not the case. It has been known since the 1970s [2] that for a fixed graph and a fixed dimension d, either almost all d-dimensional frameworks for that graph are rigid, or almost all are not rigid. This means that the rigidity of a framework is determined entirely by its graph, so we can treat rigidity of a framework as a property of its graph.

[10] Of course, one also needs to check that the square cannot be embedded as an independent framework in $\mathbb{R}$.

[11] Stop for a moment and figure out *why* the complete graph is not 2-independent.

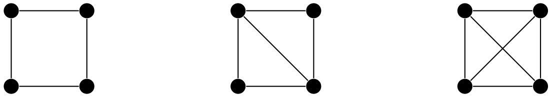


Figure 3: The 4-cycle, the diamond graph and the complete graph.

Independence of frameworks, on the contrary, is not directly a graph property. For every graph G with m vertices there indeed does exist an independent framework with that graph. Choose, for example, the framework in $\mathbb{R}^m$ such that each edge corresponds to one of the m directions of $\mathbb{R}^m$. Nevertheless, d-independence is indeed a property of the graph and not of the framework. Thus, the diamond graph is rigid and 2-independent.

We can now formulate [7, Theorem 1.2] from Gross and Sullivant in the following way. If a graph G is $(n-1)$-independent, then $\operatorname{mlt}(G) \leq n$. Therefore, if we show that any (maximal) planar graph can be realized by an independent framework in the 3-dimensional space, we are able to conclude the 4-Sample Theorem. This is what we aim for in the next section.

7 Proof of the 4-Sample Theorem

Let G be a maximal planar graph. The celebrated Steinitz's Theorem[12] from 1916 says that the graphs formed by the vertices and edges of polyhedra are precisely the finite 3-connected planar graphs. Consequently, since G is 3-connected, there exists a polyhedron which corresponds to our graph G. Of course, we now view the vertices of this polyhedron as a framework in $\mathbb{R}^3$, but is it rigid? At this point, we employ a theorem of Cauchy[13] which originated in Euclid's "Elements" and has a proof from "THE BOOK" [1, Chapter 14].[14] Cauchy's Theorem says that if two convex polyhedra have the same graph according to Steinitz's Theorem and if further under this correspondence of graphs, all facets are pairwise congruent (as 2-dimensional convex polygons), then the 3-dimensional convex polyhedra are themselves congruent. In particular, the framework is rigid.[15]

[12] Ernst Steinitz (1871–1928) was a German mathematician who made fundamental contributions to modern algebra.

[13] Augustin Louis Cauchy (1789–1859) was a French mathematician who made significant contributions to the burgeoning field of analysis.

[14] At least after Steinitz, Schoenberg, and Aleksandrov ironed out all the issues in Cauchy's proof of the main lemma.

[15] In the world of non-convex polyhedra there also exist "flexible" polyhedra which have an interesting history too.

This can be illustrated by a fun experiment: construct a graph of a 3-dimensional polytope from pieces of maccheroni (that is, rigid tubes) and connect the edges flexibly, for example by running a thread through all edges. A cube built like this falls flat (it is not rigid), but a tetrahedron does not (it is rigid).

Back in the proof, since maximal planar graphs subdivide the plane into triangles, our polyhedron associated to the graph G is a simplicial polytope – all of its faces are triangles. By the SSS-Theorem[16], each face is rigid. Hence, since the congruence of faces is guaranteed, Cauchy's Theorem can be applied. Thus, we have produced a rigid framework from the polyhedron that we obtained from the maximal planar graph G. In $\mathbb{R}^3$, the minimal number of edges of a rigid framework turns out to be $3m - 6$. By this minimality, our framework is rigid and independent, and hence G is 3-independent. Therefore, we have achieved $\text{mlt}(G) \leq 4$. This means that the MLE of the Gaussian graphical model associated to any planar graph exists with probability 1 as soon as we have at least four samples, which is precisely the 4-Sample Theorem. □

8 Conclusion, questions, outlook

Graphical models associated to planar graphs are very important in spatial statistics. The 4-Sample Theorem tells us that only a small, constant number of samples is needed for consistent estimation in these models. Nevertheless, in some applications, it might still be too much to ask for the MLE to exist with probability 1. In such a situation, a positive probability for the MLE to exist would suffice. This is captured by the *weak maximum likelihood threshold*. For example, the weak MLT of any square-grid graph, such as the one in the example on page 5, is 2, although the MLT is 3.

The bound in the 4-Sample Theorem is sharp, and there exist examples of planar graphs for all values of the MLT. A graph G with no edges satisfies $\text{mlt}(G) = 1$. Any *tree*[17] T has $\text{mlt}(T) = 2$, the 4-cycle C_4 has $\text{mlt}(C_4) = 3$ and finally, the complete graph on four vertices K_4 satisfies $\text{mlt}(K_4) = 4$.

As always, the mathematical story continues. *Matroid theory* is a crucial tool to express and generalize the notions of independence in the proof. It is widely recognized as the right language to talk about rigidity and the MLT as a graph invariant. See [4] for recent bounds on $\text{mlt}(G)$ in terms of rigidity-theoretic properties of G. This connection is actively being developed in algebraic statistics.

---

[16] The Side-Side-Side Theorem states that two triangles with equal corresponding side lengths must be congruent.

[17] A tree is a graph where any two vertices are connected by exactly one path of edges.

References

[1] M. Aigner and G. M. Ziegler, *Proofs from THE BOOK*, 6th ed., Springer, 2018.

[2] L. Asimow and B. Roth, *The rigidity of graphs*, Transactions of the American Mathematical Society **245** (1978), 279–289.

[3] D. I. Bernstein, S. Dewar, S. J. Gortler, A. Nixon, M. Sitharam, and L. Theran, *Computing maximum likelihood thresholds using graph rigidity*, Algebraic Statistics **14** (2023), no. 2, 287–305.

[4] ______, *Maximum likelihood thresholds via graph rigidity*, The Annals of Applied Probability **34** (2024), no. 3, 3288–3319.

[5] S. L. Buhl, *On the existence of maximum likelihood estimators for graphical Gaussian models*, Scandinavian Journal of Statistics **20** (1993), no. 3, 263–270.

[6] J. Friedman, T. Hastie, and R. Tibshirani, *Sparse inverse covariance estimation with the graphical lasso*, Biostatistics **9** (2008), no. 3, 432–441.

[7] E. Gross and S. Sullivant, *The maximum likelihood threshold of a graph*, Bernoulli **24** (2018), no. 1, 386–407.

[8] K. Pearson, *Contributions to the mathematical theory of evolution*, Philosophical Transactions of the Royal Society of London (1894), no. 185.

[9] A. Seigal, *The algebraic statistics of an Oberwolfach workshop*, Snapshots of Modern Mathematics from Oberwolfach (2018), no. 01, https://doi.org/10.14760/SNAP-2018-001-EN.

[10] S. Sullivant, *Algebraic statistics*, Graduate Studies in Mathematics, vol. 194, American Mathematical Society, 2018.

[11] C. Uhler, *Geometry of maximum likelihood estimation in Gaussian graphical models*, The Annals of Statistics **40** (2012), no. 1, 238–261.

[12] T. Zerenner, P. Friederichs, K. Lehnertz, and A. Hense, *A Gaussian graphical model approach to climate networks*, Chaos: An Interdisciplinary Journal of Nonlinear Science **24** (2014), no. 2.

Carlos Améndola *is a professor of algebraic and geometric methods in data analysis at the institute of mathematics of the Technical University of Berlin.*

Thomas Kahle *works at Otto-von-Guericke University in Magdeburg, Germany on topics of algebra and combinatorics with applications to statistics. He is active in the popularization of mathematics for example on Social Media or through his Podcast Eigenraum.*

Mathematical subjects
Discrete Mathematics and Foundations, Probability Theory and Statistics

DOI
10.14760/SNAP-2026-005-EN

Snapshots of modern mathematics from Oberwolfach provide exciting insights into current mathematical research. They are written by participants in the scientific program of the Mathematisches Forschungsinstitut Oberwolfach (MFO). The snapshot project is designed to promote the understanding and appreciation of modern mathematics and mathematical research in the interested public worldwide. All snapshots are published in cooperation with the IMAGINARY platform and can be found on www.imaginary.org/snapshots and on www.mfo.de/snapshots.

ISSN 2626-1995

Mathematisches Forschungsinstitut
Oberwolfach gGmbH
Schwarzwaldstr. 9 – 11
77709 Oberwolfach
Germany

Director
Gerhard Huisken

Mathematisches Forschungsinstitut Oberwolfach

Member of Leibniz Association

IMAGINARY open mathematics