Challenges of High-Dimensional Combinatorics

Nati Linial

The Hebrew University

Jerusalem, Israel

July 9, 2018

What is High-Dimensional Combinatorics?

This area starts from the observation that essentially all the elementary combinatorial concepts that we teach to our undergraduates at their basic Discrete Math course are one dimensional.

These objects have interesting high-dimensional counterparts that we want to investigate. There are several particularly significant examples, some of which we'll encounter today.

Some favorite high-dimensional creatures

- A graph is a one-dimensional simplicial complex. We will discuss some extremal problems on the combinatorics of higher-dimensional simplicial complexes.
- We view Latin squares as 2-dimensional permutations.
- There is a fascinating notion of hypertrees due to Gil Kalai. We still cannot answer many basic questions about them.
- Steiner systems can be viewed as high-dimensional partitions.
- There are several significant recurring themes. E.g.,:
  - How to efficiently generate them at random.
  - Exploring their typical properties.
  - How to minimize their discrepancy.
  - Maximizing their girth.

An Extremal Problem

A major theme in extremal graph theory is the search for statements of the form "a graph with many edges must have small dense subgraphs".

Here our motivation is more geometric in nature and we seek to prove statements such as An n-vertex 3-uniform hypergraph with many faces must contain a triangulation of the 2-dimensional sphere S^2.

An early extremal high-dimensional result

Theorem (Brown Erdős, Sós, 1973)

Every n-vertex 3-uniform hypergraph $H = (V, X)$ with at least $c \cdot n^{5/2}$ hyperedges contains a triangulation of the 2-sphere S^2. The bound is tight up to the value of the constant $c > 0$.

Proof of upper bound

Consider the $n \times \binom{n}{2}$ zero-one matrix whose rows are indexed by the n vertices and its columns by unordered pairs.

$$A = \begin{pmatrix} a_{1,e_1} & a_{1,e_2} & \cdots \\ \vdots & \ddots & \\ a_{n,e_1} & & a_{n,e_{\binom{n}{2}}} \end{pmatrix}$$

where the v, e entry is one or zero according to whether or not $x \cup e$ is a hyperedge in H.

Let u_i be the i-th row in this matrix. It is an indicator vector of the edge set of a graph called the link of vertex i.
Namely, those edges, which together with vertex i form a hyperedge in H.

Proof (upper bd.)

In particular, the inner product $< u_i, u_j >$ is the number of edges which are both in the link of vertex i and in the link of vertex j. As we soon show, there is a pair $1 \leq i < j \leq n$ for which

$$< u_i, u_j > \quad \geq \quad n.$$

But this implies that the intersection of these two links must contain a cycle. Such a cycle together with the vertices i, j forms a triangulation of the 2-sphere.

In the topological terminology such a 2-sphere is called the *suspension* or *double cone* over a cycle.

Proof (upper bd. - last step)

Let d_e be the number of triples that contain the edge e. A standard double-counting argument yields:

$$\sum_{i<j} < u_i, u_j >= \sum \binom{d_e}{2}.$$

By convexity,

$$\sum \binom{d_e}{2} \geq \binom{n}{2} \cdot \binom{\frac{1}{\binom{n}{2}} \sum d_e}{2}$$

By assumption, the r.h.s. is at least $\Omega(n^3)$ so that

$$\exists i < j \quad \text{with} \quad < u_i, u_j > \quad \geq \quad n$$

as planned. $\square$

Proof (Lower bound)

The original proof of the lower bound in the BES paper was constructive. Namely, they gave an explicit construction of a 3-uniform hypergraph with $c' \cdot n^{5/2}$ hyperedges that contains no triangulation of S^2.

We now present a probabilistic argument that yields this conclusion.
Its advantage: It is more widely applicable (see below).

More concretely, the argument is based on the alteration method.

Proof (Lower bound, contd.)

Namely, we select each triple independently and with probability $p = \frac{c}{\sqrt{n}}$ for some appropriately chosen $c > 0$ and then remove a triple from every triangulated 2-sphere that we have created.

The crux of the matter is to show, e.g., that we have eliminated only a small fraction of our originally chosen hyperedges.

Proof (Lower bound, contd.)

As Tutte proved, there are at most

$$v! \cdot \lambda^v$$

v-vertex triangulations of the plane, for some constant $\lambda > 1$.
By Euler's formula

$$v - e + f = 2$$

In a triangulation

$$3f = 2e$$

So that

$$f = 2v - 4$$

Lower bound contd.

So, what is the expected number of triangulated spheres that we create?

$$\mathbb{E}X = \sum_{v \geq 4} \binom{n}{v} p^f \cdot \text{ number of } v\text{-vertex triangulated spheres}$$

But $p = \frac{c}{\sqrt{n}}$, $f = 2v - 4$, and there are at most $v! \cdot \lambda^v$ such spheres.

Therefore,

$$\mathbb{E}X < \sum_{v \geq 4} \binom{n}{v} (\frac{c^2}{n})^{v-2} v! \cdot \lambda^v = O(n^2)$$

The claim follows, since $n^2 \ll n^{5/2}$. $\square$

Conjecture and related problems

Conjecture

Every n-vertex 3-uniform hypergraph with $c \cdot n^{5/2}$ hyperedges contains a triangulation of a torus.

As we saw, if true this is tight up to the value of c.

Many similar problems suggest themselves.

What about other 2-dimensional manifolds?
In higher dimensions?

What's in a tree?

Every basic book in discrete mathematics has this characterization of trees:

Theorem

TFAE for a graph $G = (V, E)$ with n vertices and $n - 1$ edges

- *G is connected.*
- *G is acyclic.*
- *G is collapsible.*

An elementary collapse is a step at which we discard a vertex of degree 1 and the only edge that contains it. If all edges of G can be discarded by a sequence of elementary collapses, we say that G is collapsible.

A trick? A one dimensional miracle?

Enter G's incidence matrix $=$ the one-dimensional boundary operator ∂_1.

$$\begin{array}{cc} & E \\ V & \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ \dots & & \dots & \dots & & \\ 0 & 0 & -1 & 0 & \dots & 0 \\ \vdots & \ddots & & & & \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \dots & & \dots & \dots & & \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix} \end{array}$$

The equivalence between connectivity and acyclicity is just the equality row $\mathrm{rank}(\partial_1) = $ column $\mathrm{rank}(\partial_1) = n - 1$.

A trick?

Collapsibility means that row $\mathrm{rank}(\partial_1) = \text{column rank}(\partial_1) = n - 1$ can be proved by the following simple algorithm.

Algorithm: If some row has a single nonzero entry, eliminate it and the corresponding column. Repeat.

Rationale: This column cannot participate in any column dependence, i.e., this edge does not participate in any cycle.

Connectivity of graphs is the vanishing of the zeroth homology.
Acyclicity of graphs is the vanishing of the first homology.

In higher dimension collapsibility implies acyclicity, but is not implied by it. The equivalence between the two is a one-dimensional miracle.

Bye bye Wonderland

Recall the phase transition at $p = \frac{1}{n}$ in $G(n,p)$ theory.

When $p = \frac{1-\epsilon}{n}$ a $G(n,p)$ graph is a forest with positive probability.

But when $p = \frac{1+\epsilon}{n}$ the graph is a forest with vanishingly small probability.

So, what happens in higher dimensions?

High-dimensional $G(n, p)$?

Roy Meshulam and I have introduced $X_d(n, p)$, a high-dimensional counterpart to $G(n, p)$.

This is a d-dimensional n-vertex simplicial complex with a full $(d-1)$-dimensional skeleton in which every d-face is included independently with probability p.

And if your topology is a bit rusty: This is an n-vertex $(d+1)$-uniform hypergraph where every hyperedge is included independently with probability p, (plus every hyperedge of size $\leq d$).

At any event, note that $X_1(n, p)$ is identical with $G(n, p)$.

Some tools of the trade

The analogue of the incidence matrix of a graph is the *d*-dimensional boundary operator ∂_d. In matrix form

$$\begin{array}{cc}
 & \binom{n}{d+1} \\
\binom{n}{d} & \begin{pmatrix}
0 & 0 & 0 & 0 & \dots & 0 \\
\dots & & \dots & \dots & & \\
0 & 0 & 1 & 0 & \dots & 0 \\
\vdots & \ddots & & & & \\
0 & 0 & -1 & 0 & \dots & 0 \\
\dots & & \dots & \dots & & \\
0 & 0 & 1 & 0 & \dots & 0 \\
\vdots & \ddots & & & & \\
\dots & & \dots & \dots & & \\
0 & 0 & 0 & 0 & \dots & 0
\end{pmatrix}
\end{array}$$

Some tools of the trade - ∂_d in the random case

$$\begin{array}{cc} & p\binom{n}{d+1} \\ \binom{n}{d} & \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ \dots & & \dots & \dots & & \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & & & & \\ 0 & 0 & -1 & 0 & \dots & 0 \\ \dots & & \dots & \dots & & \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & & & & \\ \dots & & \dots & \dots & & \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix} \end{array}$$

Here the left kernel encodes "connectivity" and the right kernel "cycles".

What $X_d(n, p)$ tells us

Theorem (A bit inaccurate - Combining several papers by Aronshtam, Linial, Luczak, Meshulam, Peled)

- *The threshold for d-collapsibility in $X_d(n, p)$ is $p \approx \frac{\log d}{n}$.*
- *The threshold for acyclicity in $X_d(n, p)$ is $p \approx \frac{d}{n}$.*
- *The actual numerators can be computed to desired accuracy.*

Terra incognita: There is a whole range of the parameter p in $X_d(n, p)$ for which the resulting simplicial complex is acyclic, yet, not d-collapsible.

Sporadic examples of this sort were known for a long time, but this may be the first systematic supply of such complexes. This is a fascinating yet uncharted territory.

Climbing the tree, again

The incidence matrix ∂_1 of the complete graph K_n

$$
\begin{array}{cc}
 & \begin{array}{cccccc} & & \binom{n}{2} & & & \end{array} \\
[n] & \begin{pmatrix}
0 & 0 & 0 & 0 & \dots & 0 \\
\dots & & \dots & \dots & & \\
0 & 0 & -1 & 0 & \dots & 0 \\
\vdots & \ddots & & & & \\
0 & 0 & 1 & 0 & \dots & 0 \\
\dots & & \dots & \dots & & \\
0 & 0 & 0 & 0 & \dots & 0
\end{pmatrix}
\end{array}
$$

has rank $n-1$. What does a column basis look like?

Answer: a tree.

Gil Kalai's $\mathbb{Q}$-acyclic complexes, or d-Hypertrees

The matrix of the d-dimensional boundary operator ∂_d has rank $\binom{n-1}{d}$. According to Kalai - A d-hyertree is a column basis of this matrix, thought of as a simplicial complex or a $(d+1)$-uniform hypergraph.

$$
\begin{array}{cc}
 & \binom{n}{d+1} \\
\binom{n}{d} & \begin{pmatrix}
0 & 0 & 0 & 0 & \dots & 0 \\
\dots & & \dots & \dots & & \\
0 & 0 & 1 & 0 & \dots & 0 \\
\vdots & \ddots & & & & \\
0 & 0 & -1 & 0 & \dots & 0 \\
\dots & & \dots & \dots & & \\
0 & 0 & 1 & 0 & \dots & 0 \\
\vdots & \ddots & & & & \\
\dots & & \dots & \dots & & \\
0 & 0 & 0 & 0 & \dots & 0
\end{pmatrix}
\end{array}
$$

The Cayley-Kalai enumeration formula

Theorem (Cayley-Kalai Formula)

$\sum_{T \text{ is a } d-\text{hypertree}} |H_{d-1}(T)|^2 = n^{\binom{n-2}{d}}$

Problem

How many n-vertex d-hypertrees are there?

Already Kalai's work yields some estimates, which Yuval Peled and I have recently improved, but the answer still eludes us.

Conjecture

There are constants $c_d > 0$ such the number of n-vertex d-hypertrees is

$$((1 + o_n(1)) \cdot c_d \cdot n)^{\binom{n-1}{d}}$$

Something on collapsible hypertrees

Clearly, a collapsible n-vertex, d-dimensional complex whose number of d-faces is $\binom{n-1}{d}$, is a hypertree. But these hypertrees seem extremely rare.

Conjecture of Kalai

Almost none of the d-dimensional hypertrees are collapsible.

Conjecture and problem - Linial and Peled

There are constants $\gamma_d > 0$ such that the number of collapsible d-hypertrees is

$$((1 + o_n(1)) \cdot \gamma_d \cdot n)^{\binom{n}{d}}$$

Is it possible that, in fact, $\gamma_d = c_d$?
.... and how could this be consistent with the previous conjecture?

... and indeed - How do you sample efficiently and uniformly d-hypertrees, collapsible d-hypertrees etc.?

In order to numerically generate d-hypertrees, we consider them as bases of a $\mathbb{Q}$-matroid.

Algorithm: Starting from a given d-hypertree, add a d-face uniformly at random. Exactly one cycle C is created. Remove a random d-face of C.

problem

The limit distribution of the above Markov Chain is clearly uniform on d-hypertrees, but is this chain rapidly mixing?

Collapsible d-hypertrees are even more mysterious

We have no idea about the following:

problem

How to generate collapsible d-trees efficiently uniformly at random?

Collapsible d-hypertrees are not the bases of a matroid, but we ask:

problem

Do collapsible d-hypertrees offer an interesting extension to the notion of bases in Matroid theory?

Generation - Somehow, quickly, uniformly...

Many one-dimensional combinatorial objects can be sampled uniformly and efficiently rather easily: Graphs, permutations, partitions etc.

As usual, the high-dimensional world is more tricky.

We start with One Factorization $= \mathrm{OF}_n$.
It is well-known that for even n, the edges of K_n can be $(n-1)$-colored. This is sometime described as schedule for a league of n teams. If edge ij is colored k, then team i meets team j on round k.

Generating one-factorizations

It is easy (and fun) to find Markov Chains that quickly come close to a OF.

Here is one (of many...) that we examined.

- Start at any edge-coloring of K_n with colors $\{1, \ldots, n-1\}$.
- At each step attempt to randomly recolor a random edge.
- Accept the change iff it reduces the potential function Ψ.

$\Psi :=$ Number of pairs of intersecting edges that have the same color.

Note that $\Psi = 0$ iff the coloring is a OF.

Even though this process usually gets stuck...

This process gets almost always stuck in a local optimum. However, we have a good description of these local optima:

Lemma 1

In a local optimum of Ψ, every color class is the union of disjoint edges and two-edge paths.

Lemma 2

At every proper local optimum of Ψ, one can efficiently find two vertices whose edges can be recolored so as to decrease Ψ.

Theorem (Maya Dotan, Linial)

There is an $(S_n \times S_{n-1})$-invariant Markov Chain that reaches a OF_n in $O(n^3)$ steps from every starting point.

Going a little more lightly on Ψ

In a strict walk (aka Hill climbing) the potential improves at every step.

What about a more relaxed attitude to Ψ ?

In a mild random walk we adopt a proposed recoloring whenever Ψ does not increase. The following conjecture is supported by simulations:

Conjecture

The mild edge-color-switch walk converges w.h.p. in poly(n) time.

Enter Metropolis Filter

We also consider walks with occasional steps in which Ψ increases.

Every proposed color switch in which Ψ does not increase is accepted. If the proposed switch increases Ψ by Δ, we accept it with probability ϵ^{Δ}, for some choice of $\epsilon > 0$. The limit distribution of such a walk is uniform on OFs. The challenge is to guarantee that:

- The MC mixes rapidly (in poly(n) time), and
- It spends at least $\frac{1}{\text{poly}(n)}$ of the time on OFs.

Metropolis Filter to the rescue?

Numerical simulations suggest:

Conjecture (Made jointly with Yuval Peled)

With a proper choice of ϵ the walk spends $\frac{1}{\mathrm{poly}(n)}$ of its time on OFs.

Concerning rapid mixing we do not know much.

A (somewhat defective) hierarchy

Both a OF and a Steiner Triple System can be viewed as Latin squares.

A OF is a Latin square that is

- Symmetric w.r.t. the main diagonal: $a_{ij} = a_{ji}$ for all i and j.
- The terms of the main diagonal: $a_{ii} = n$ for all i.

An STS is a Latin square A that is

- Symmetric w.r.t. the main diagonal: $a_{ij} = a_{ji}$ for all i and j.
- The terms of the main diagonal: $a_{ii} = i$ for all i.
- Additional symmetry: If $a_{ij} = k$ then $a_{ik} = j$.

Latin squares $\supset$ One-factorizations $\supset$ STS's ?

The second inclusion is invalid as stated, e.g., OF_n exist iff n is even whereas STS_n exists iff $n \equiv 1, 3 \mod 6$. However, we think that it is a fruitful concept to consider, especially in asymptotic questions.

How many are they?

- Latin squares

$$((1 + o(1))\frac{n}{e^2})^{n^2}$$

- One-factorizations

$$((1 + o(1))\frac{n}{e^2})^{n^2/2}$$

- Steiner Triple Systems

$$((1 + o(1))\frac{n}{e^2})^{n^2/6}$$

The big picture

- Investigate combinatorial designs under the basic paradigms of modern graph theory: Ask asymptotic, extremal, local questions.
- Use their structural similarities and seek fundamental differences.

How come this was not done before?

1. It was, but on a limited scale. E.g., the Erdős-Hanani conjecture. Rödl, who proved it, went on (with several coauthors) to study questions such at independent sets in STS etc.
2. There was recent dramatic progress in the field - Keevash's proof that Steiner Systems exist, then work by Glock, Kühn, Lo and Osthus.
3. Now that we know these things exist, we can start to investigate them within our usual paradigms.

Back to generation

Latin squares: Jacobson and Matthews '96 gave a MC based on a simple local step. This chain is connected, but we do not know if it mixes rapidly.

Here is another efficient method to randomly generate Latin squares (joint with Zur Luria). First put a random permutation in every row. Find a column where some entry is missing and another one where the same entry is abundant. Balance between the two columns by a zigzag move.

Higher dimensional permutations? A 3-D permutation is an $n \times n \times n$ array with entries from $[n]$ such that every line (row, column or shaft) contains every $n \geq i \geq 1$ exactly once. Is there a MC that generates 3-D and even higher-dimensional permutations?

Girth

The girth of graphs has been a personal obsession for me. I cannot resist telling the most embarrassing question on this which I cannot answer.

Start with n isolated vertices. At each step put an edge between some pair of vertices at distance $> D$, until no such pairs remain. The resulting graph has $\mathrm{diam}(G) \leq D$. Also, no step creates a cycle of length $< D + 2$. So, this yields a graph with $\text{girth} - \text{diameter} \geq 2$.

Problem

Are there graphs G with minimum degree ≥ 3 for which

$$\text{girth}(G) - \text{diameter}(G)$$

is arbitrarily large?

Girth in higher dimension

What is the girth of a connected graph G? Any connected set of v vertices in G contains at least $v-1$ edges. The girth of G is the smallest v such that G has v vertices that span at least v edges. Similar considerations on STS lead to the following notion.

Definition

The *girth* of an STS is the smallest v such that there is a set of v vertices that span at least $v-2$ triples of the system.

Erdős STS girth conjecture

There exist Steiner Triple Systems of arbitrarily high girth.

Girth in higher dimension

Despite considerable efforts, this conjecture is still wide open. This prompted the following question by Rödl, and by me.

Problem

Do there exist a $c > 0$ and infinitely many n-vertex 3-uniform hypergraphs with at least cn^2 hyperedges and arbitrarily large girth?

A very recent work gives a strikingly positive answer.

Theorem (Glock, Kühn, Lo, and Osthus)

For every $\epsilon > 0$ and for every integer K, if $n \equiv 1, 3 \bmod 6$, is large enough, then there is an n-vertex 3-uniform hypergraph with at least $(\frac{1}{6} - \epsilon)n^2$ hyperedges and girth $\geq K$.

This is remarkably close to $\frac{n(n-1)}{6}$, the number of triples in an STS.

More high-dimensional girth

In the same spirit we can define a cycle in a Latin square as a set of rows A, a set of columns B and a set of colors C such that the $A \times B$ sub matrix has at least $|A| + |B| + |C| - 2$ symbols from C.

We define the girth of a Latin square as the minimum of $|A| + |B| + |C|$ over all cycles $(|A|, |B|, |C|)$, and we ask

Problem

Are there Latin squares of arbitrarily high girth?

Even more high-dimensional girth

A cycle in a OF is a set of vertices A and a set of colors C such that there are at least $|A| + |C| - 2$ edges induced by A with colors from C.
Again the girth of a OF is the minimum of $|A| + |C|$ over all cycles $(|A|, |C|)$, and we ask

Problem

Are there one factorizations of arbitrarily high girth?

MAZALTOV

LACI