A glimpse of high-dimensional combinatorics

Nati Linial

ITW, Jerusalem, May 1, 2015

What are we talking about?

The geometric viewpoint of combinatorics suggests that many basic combinatorial constructs are one-dimensional. Our purpose here is to explore their fascinating high-dimensional counterparts.

- Latin squares are the two-dimensional analogs of permutations.
- Hypertrees extend the notion of a tree.
- There is an emerging theory of high-dimensional tournaments
- Simplicial complexes offer a high-dimensional perspective of graph theory.

Yes, graphs are everywhere, but why?

One major reason for the phenomenal success of graphs in real life applications is this:
In numerous real-life situations we need to understand a large complex system whose elementary constituents are pairwise interactions.

- Interacting elementary particles in physics.
- Proteins in some biological system.
- Partners in an economic transaction.
- Humans in some social context.

But what can we do about multi-way interactions?

- Proteins come, more often than not, in complexes that involve several proteins at once.
- Human social networks tend to include several individuals.
- Economics transactions often involve several parties at once.
- Distributed systems are many-sided by their very nature.

Hypergraphs, anyone?

There is a combinatorial theory of hypergraphs. A hypergraph (V, F) consists of a set of *vertices* V and a collection F of subsets of V. The sets that belong to F are called *hyperedges*.
If every hyperedge contains exactly two vertices we are back to graphs.
These are the good news. The bad news are that the theory of hypergraphs is not nearly as well developed as graph theory.

Never despair - Simplicial complexes to the rescue

We only need to make a small modification to the notion of hypergraph to arrive at simplicial complexes. This way we make contact with a rich body of powerful mathematics in topology and geometry that can help us.
What's more - many fascinating new connections and perspectives suggest themselves.

Definition

Let V be a finite set of *vertices*. A collection of subsets $X \subseteq 2^V$ is called a *simplicial complex* if it satisfies the following condition:

$$A \in X \text{ and } B \subseteq A \Rightarrow B \in X.$$

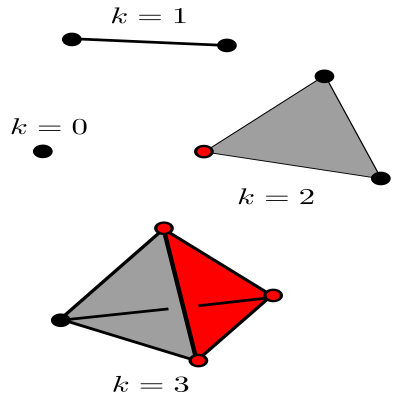
A member $A \in X$ is called a simplex or a face of dimension $|A| - 1$.
The dimension of X is the largest dimension of a face in X.

Up up and away

- A one-dimensional simplicial complex = A graph.
  - A zero-dimensional face = A vertex.
  - A one-dimensional face = an edge.
- Higher dimensional complexes offer a wonderful mix of combinatorics with geometric (mostly topological) ideas.
- The challenge - to develop a combinatorial perspective of higher dimensional complexes.

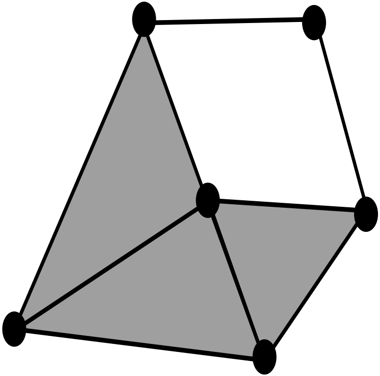
Simplicial complexes as geometric objects

Assign to $A \in X$ with $|A| = k + 1$ a k-dim. simplex



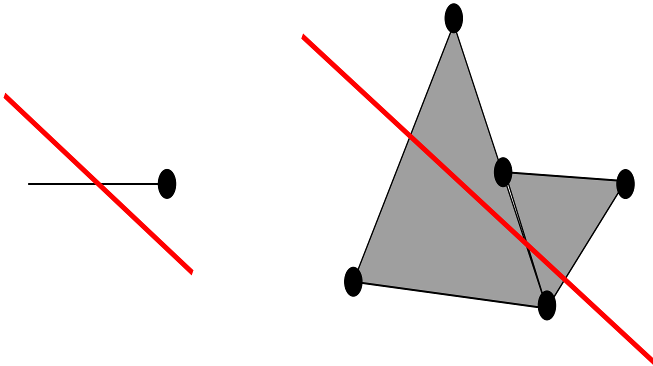
Putting simplices together properly

The intersection of every two simplices in X is a common face.



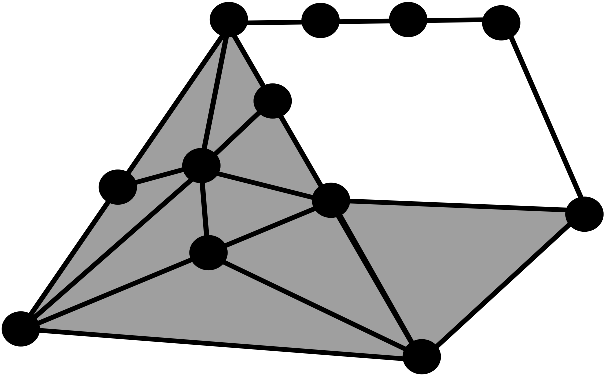
How NOT to do it

Not every collection of simplices in $\mathbb{R}^d$ is a simplicial complex



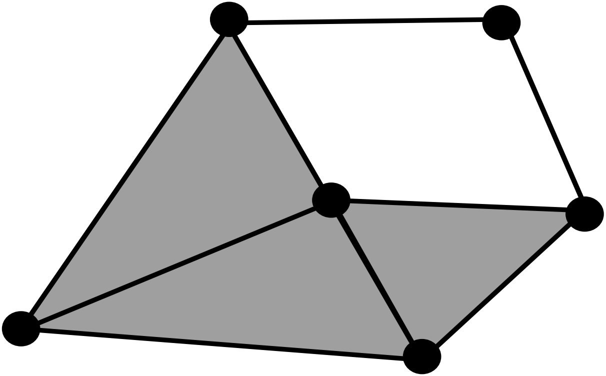
Geometric equivalence

Combinatorially different complexes may correspond to the same geometric object (e.g. via subdivision)



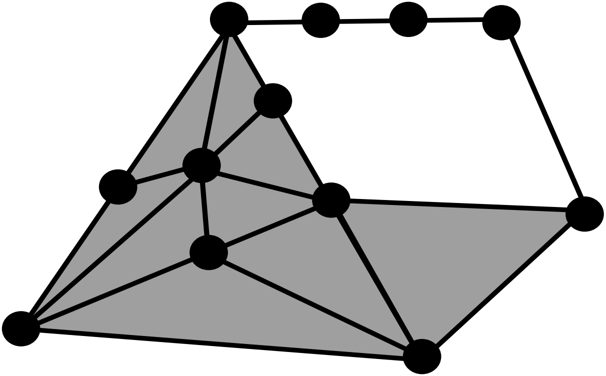
Geometric equivalence

So



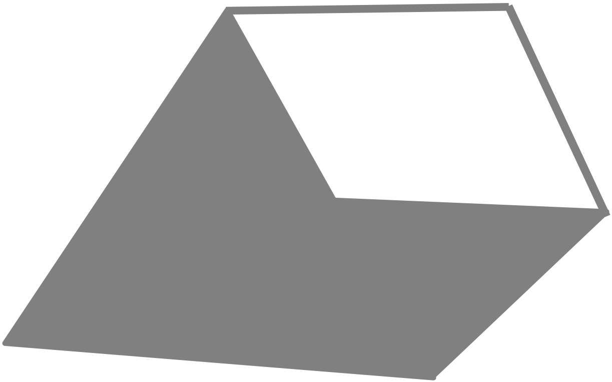
Geometric equivalence

and



Geometric equivalence

are two different combinatorial descriptions of the same geometric object



Recall: The incidence matrix of a graph

$V \times E$ Vertices vs. edges.

$$A_G = \begin{array}{c} \\ \vdots \\ i \\ \vdots \\ j \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} \dots & ij & \dots & \dots & \dots \end{array} \\ \begin{pmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & +1 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & -1 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \end{array}$$

The incidence matrix tells many things

- G is connected iff A_G has a trivial left kernel.
  - Because A_G's left kernel is the linear span of the indicator vectors of G's connected components.
- The cycle space of G is the right kernel of A_G.
  - Because A_G's right kernel is the linear span of the indicator vectors of G's cycle.

Recall: Equivalent descriptions of trees

Theorem

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

1. *G is connected.*
2. *G is acyclic.*
3. *The columns corresponding to $E(G)$ are linearly independent.*
4. *They form a column basis for A_{K_n}, the incidence matrix of the complete graph.*
5. *G is collapsible.*

The equivalence of conditions 1, 2, 3, 4

The rank of A_{K_n} is $n-1$: There is exactly one linear dependence among the n rows

$$1A_{K_n} = 0.$$

1. G is connected $\Leftrightarrow$ the left kernel of A_G is trivial.
2. G is acyclic $\Leftrightarrow$ the right kernel of A_G is zero.
3. The columns corresponding to $E(G)$ are linearly independent.
4. They form a column basis for A_{K_n}, the incidence matrix of the complete graph.

Collapsibility

An elementary collapse is a step where you remove a vertex of degree one and the single edge that contains it.

A graph G is collapsible if by repeated application of elementary collapses you can eliminate all of the edges in G.

Collapsing - a linear algebra perspective

Let A_G be the incidence matrix of graph G. In an elementary collapse we erase row i and column e of A_G where the (i, e) entry is the only nonzero entry in the i-th row. Recall: e is the one and only edge incident with vertex i.

G is collapsible if it is possible to eliminate all its columns by a series of elementary collapses.

This implies that G is acyclic - Collapsing yields a proof that the right kernel is empty.

But note

Whereas conditions 1-4 are linear algebraic, collapsibility is a purely combinatorial condition. Indeed we will soon see that in higher dimensions

collapsibility implies conditions 1-4, but the reverse implication does not hold.

Setting up the ground

Here is the high-dimensional analog of the incidence matrix.

Boundary operators of simplicial cplexes

$(d-1)$-dimensional faces vs. d-dimensional faces.

$$\partial = \begin{array}{c} \\ \vdots \\ ij \\ \vdots \\ ik \\ \vdots \\ \vdots \\ jk \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccccc} \cdots & \cdots & ijk & \cdots & \cdots & \cdots & \cdots & \cdots \end{array} \\ \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & +1 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & -1 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & +1 & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \end{array}$$

Does this tell us what a hypertree is?

We only know where to start:

Q: What is the rank of ∂_d ?

A: $\binom{n-1}{d}$

because $\partial_{d-1}\partial_d = 0$.

What is a d-dimensional hypertree?

It is a d-dimensional simplicial complex with

- A full $(d-1)$-dimensional skeleton.
- It has $\binom{n-1}{d}$ d-dimensional faces.

So that

- ∂_d has a trivial left kernel.
- ∂_d has a zero right kernel.
- The columns of ∂_d for a column basis to boundary operator of the full matrix of all $(d-1)$-faces vs. all d-faces.

What about collapsibility?

Let X be a d-dimensional complex.
If some $(d-1)$-dimensional face τ is contained in a unique d-dimensional face σ, then the corresponding elementary collapse is to eliminate both τ and σ from X.
X is d-collapsible if it is possible to eliminate all its d-faces by a series of elementary collapses.
Collapsibility implies acyclicity. But....

A little surprise

$\binom{6-1}{2} = 10$

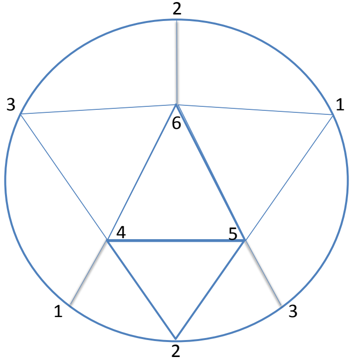


Figure: A triangulation of the projective plane

A little surprise

This example is showing us (at least) two things: Unlike the 1-dimensional case of graphs, the definition of a *d*-dimensional hypertree depends on the underlying field.

Indeed: The 6-point triangulation of the projective plane is a $\mathbb{Q}$-hypertree, but not a $\mathbb{F}_2$-hypertree.

In dimension ≥ 2 collapsibility is stronger than being a hypertree.

In fact we state

Conjecture

For every $d \geq 2$ and for every field $\mathbb{F}$ and $n \to \infty$ almost none of the n-vertex d-dimensional $\mathbb{F}$-hypertrees are collapsible.

If so...

Q: Can you, at least, come up with more examples of non-collapsible hypertrees?

A construction: Let n be prime and $d \geq 2$. Fix a subset $A \subset \mathbb{Z}_n$ of cardinality $|A| = d + 1$. The sum complex X_A corresponding to A has a full $(d-1)$-dimensional skeleton and contains a d-face σ iff $\sum_{x \in \sigma} x \in A$.

Theorem (L., Meshulam, Rosenthal)

The complex X_A is always a $\mathbb{Q}$-hypertree. It is collapsible iff A forms an arithmetic progression.

An old mystery

$\mathbb{Q}$-hypertrees were introduced by Kalai (1983). He proved a beautiful enumeration formula, analogous to Cayley's formula that there are n^{n-2} labeled trees on n vertices. However, we still do not know:

Open Problem

For $d \geq 2$ and large n, find (at least approximately) the number of d-dimensional n-vertex $\mathbb{Q}$-hypertrees.

A recent surprise

Let $G = (V, E)$ be a disconnected graph, and let $ij \notin E$. We say that ij is in G's shadow if i and j belong to the same connected component of G.
In other words ij is in G's shadow iff the column corresponding to the edge ij is in the linear span of the columns of A_G.

Easy Observation

Let G be an "almost tree", i.e., an n vertex forest with $n-2$ edges (and hence with two connected components). Then at least $(1-o(1))\frac{n^2}{4}$, i.e., at least half of the remaining edges, are in G's shadow.

Shadows in higher dimension

Construction: Let X be a 2-dimensional n-vertex complex with a full 1-dimensional skeleton. The 2-faces of X are the arithmetic triples of difference $\neq 1$. Easy fact: The number of 2-faces in X is $\binom{n-1}{2} - 1$ (one less than a 2-dimensional hypertree).

Theorem (L., Yuval Peled)

The complex X is $\mathbb{Q}$-acyclic. Assuming the Riemann hypothesis[1]*, there are infinitely many primes n for which X has an empty shadow.*

[1] It actually suffices to assume the weaker Artin's conjecture

What next?

We want to develop a theory of random simplicial complexes, in light of to random graph theory. Specifically we seek a higher-dimensional analogue to $G(n, p)$.

Recollections of $G(n, p)$

This is the grandfather of all models of random graphs. Investigated systematically by Erdős and Rényi in the 60's, a mainstay of modern combinatorics and still an important source of ideas and inspiration.

Start with n vertices.

For each of the $\binom{n}{2}$ possible edges $e = xy$, choose independently and with probability p to include e in the random graph that you generate.

Closely related model: the evolution of random graphs starts with n vertices and no edges. At each step add a random edge to the evolving graph.

Back to the classics

Theorem (Erdős and Rényi '60)

The threshold for graph connectivity in $G(n, p)$ is

$$p = \frac{\ln n}{n}$$

Specifically, if $p \leq (1 - \epsilon)\frac{\ln n}{n}$, then a graph in $G(n, p)$ is, whp, disconnected.
On the other hand, if $p \geq (1 + \epsilon)\frac{\ln n}{n}$, then a graph in $G(n, p)$ is, whp, connected.

One part of this theorem is really easy

If $p < (1-\epsilon)\frac{\ln n}{n}$, then a random graph in $G(n,p)$ is not only almost surely disconnected.

In fact, in this range of p, the graph almost surely has some isolated vertices.
This is an easy consequence of the coupon-collector principle from probability theory.

That G is almost surely connected for $p > (1+\epsilon)\frac{\ln n}{n}$ requires proof.

A d-dimensional analog of $G(n, p)$

About 10 years ago, with R. Meshulam we introduced the following model of a random d-dimensiona n-vertex complex $X_d(n, p)$. It is set up so that in the one-dimensional case $d = 1$ the $X_1(n, p)$ model is identical with $G(n, p)$.
Start with a full $(d - 1)$-dimensional skeleon. (In the case of graphs - start with n vertices.)
For each d-dimensional face σ, independently and with probability p, decide whether $\sigma \in X$. (For graphs - same with every edge).

"Connectivity" in higher dimensions

Unlike the situation in graphs, there is more than one way to capture the idea of "connectivity" in higher-dimensional simplicial complexes. Here we concentrate on what is arguably the simplest one:

The boundary operator ∂_d has a trivial left kernel. But

$$\partial_{d-1}\partial_d = 0$$

So, for every d-complex X

$$\text{row space}(\partial_{d-1}(X)) \subseteq \text{left kernel}(\partial_d(X)).$$

The row space of $\partial_{d-1}(X)$ is the trivial part of $\partial_d(X)$'s left kernel. We consider X "connected" when $\partial_d(X)$ has a trivial left kernel, i.e., when

$$\text{left kernel}(\partial_d(X)) = \text{row space}(\partial_{d-1}(X)).$$

In mathematical parlance the name of this condition is the vanishing of the $(d-1)$-st homology of X.

Remark

When $d = 1$ *(i.e., for graphs)*

$$\textit{row space}(\partial_0(G)) = \{\alpha\mathbf{1} | \alpha \in \mathbb{F}\}$$

is one-dimensional, and we recover the usual definition of graph connectivity.

...and the answer is...

Theorem (L. - Meshulam, and Meshulam-Wallach)

The threshold for connectivity of $X_d(n,p)$ is

$$p = \frac{d \ln n}{n}.$$

Specifically, whp, left kernel$(\partial_d(X))$ is

- *nontrivial for $p < (1-\epsilon)\frac{d \ln n}{n}$, and*
- *trivial for $p > (1+\epsilon)\frac{d \ln n}{n}$.*

Again, one part of the theorem is easy

When $p < (1-\epsilon)\frac{d \ln n}{n}$
the matrix $\partial_d(X)$ almost surely contains an all-zeros row
and consequently it has a nontrivial left kernel.

Such a row corresponds to an $(d-1)$-dimensional face that is not contained in any of the randomly chosen d-dimensional faces.

The proof that such an "isolated" $(d-1)$-face exists, is a straightforward coupon-collector argument.

Back to $G(n,p)$ theory - the evolution of random graphs

The most dramatic chapter in Erdős-Rényi papers on $G(n,p)$ is the phase transition in the evolution of random graphs.

Start with n isolated vertices and sequentially add a new random edge, one at a time.
Observe the connected components of the evolving graph.

Prelude - The early stages

At the very beginning we see only isolated edges (a matching).

As we proceed, more complex connected components start to appear, but still they are all small and simple.

- small = cardinality $O(\log n)$.
- simple = a tree.
- Possibly a constant number of exceptions which are a small tree plus one edge = unicylic graphs with $O(\log n)$ vertices.

Crescendo - The phase transition

Around step $\frac{n}{2}$ and over a very short period of time A GIANT COMPONENT EMERGES.

GIANT= cardinality $\Omega(n)$, i.e., a constant fraction of the whole vertex set.

Note: Time $\frac{n}{2}$ corresponds to $p = \frac{1}{n}$.

In the wake of the revolution

Around step $\frac{n}{2}$ many other parameters are undergoing an abrupt change.

In particular, for $p < \frac{1-\epsilon}{n}$, the probability that the evolving graph contains a cycle is bounded away from both zero and one.

However, for $p > \frac{1+\epsilon}{n}$, the graph almost surely contains a cycle.

In other words, it almost surely ceases to be a forest.

It's not obvious what the analogous high dimensional phenomenon is

There is no obvious notion of a connected component in dimensions $d \geq 2$.

So what can the analogous statement be?

In search of the high-dimensional analog

There are at least two high-dimensional analogs of the forest/non-forest transition in graphs.

- Collapsible/non-collapsible complex.
- Acyclic/acyclic (The right kernel of ∂_d is zero/non-zero) complex.

Recall: collapsible complexes are acyclic, so clearly

$$p_{collapse} \leq p_{acyclic}$$

But is the inequality strict?

Let the experiment speak

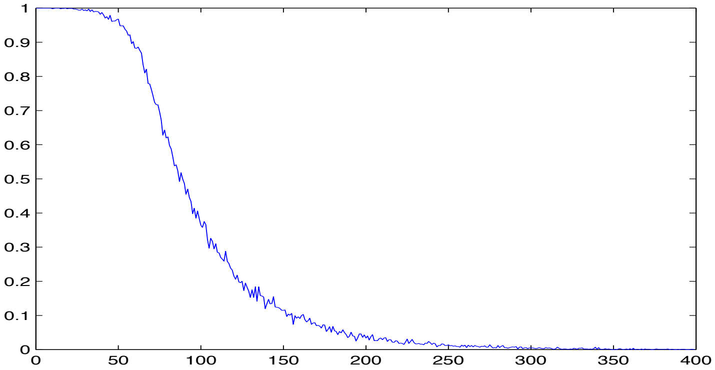
Experimenting with $G(n, p)$: Start with n vertices. Sequentially add a random edge and record whether or not this edge connects two distinct connected components.

Equivalently: Is this edge in the sun/in the shade?

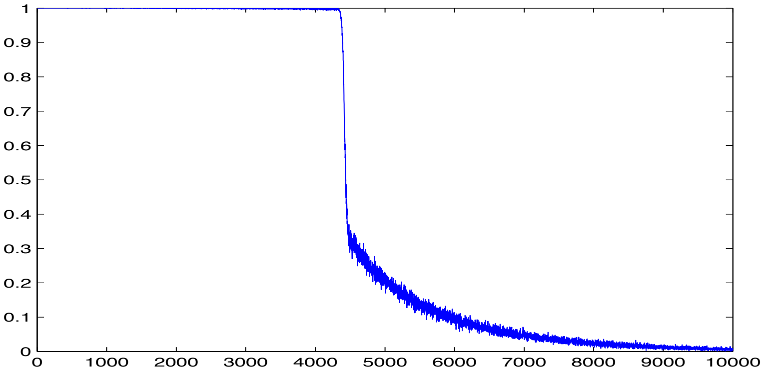
In other words: The addition of an edges can only increase the right kernel of A_G. Does it stay the same or does it go up?

Experimenting with $X_d(n,p)$: Start with a full $(d-1)$-dimensional skeleton. Sequentially add a random d-face and record whether or not this new face is in the sun/in the shade.
In other words: Does the right kernel of $\partial_d(X)$ stay the same or does it get larger as the new face is added?

A view of phase transition in $G(n, p)$



Phase transition in $X_2(n, p)$ complexes



Mysteries resolved

- The collapsibility threshold is substantially smaller than the acyclicity threshold.
- (Easy) In $G(n, p)$ the emergence of the giant component is concurrent with the emergence of the giant shadow. "Giant" means $\Omega(n^2)$ edges.
- In all dimensions $d \geq 1$ the acyclicity threshold coincides with the emergence of a giant shadow ($\Omega(n^{d+1})$ faces of dimension d).
- Whereas this is a second order phase transition in graphs, for $d \geq 2$ this is a first order phase transition.

Mysteries resolved

Theorem[Lior Aronshtam, L., Tomasz Łuczak, Roy Meshulam, Yuval Peled]

- The collapsibility threshold in $X_d(n,p)$ is

$$(1+o_d(1))\frac{\log d}{n}.$$

- The threshold for almostly having a cycle in $X_d(n,p)$ is

$$\frac{d+1-o_d(1)}{n}.$$

Mysteries resolved

- This is also where a giant shadow of $\Omega(n^{d+1})$ faces of dimension d shows up.
- In the evolution of random simplicial complexes with $d \geq 2$ the first occurring cycle is, almost surely, either
  - The boundary of a $(d+1)$-dimensional simplex, or
  - A cycle that includes $\Omega(n^d)$ faces of dimension d.

That's all folks