

# TOPOS: Pinsky was wrong, Euler was right



Robert Adler presents the third in his series of four articles on TOPOS, Topology, Probability and Statistics:

This month, I want to direct most of my words to my fellow probabilists, and, ultimately, to issue them a challenge. Accepting the challenge will, I am certain, benefit all of the three components of TOPOS.

Ross Pinsky, a close friend of mine and probabilist with a strong analytic side, once reacted to my late-in-life love affair with topology by claiming that he could think of no two topics that were further apart than Probability and (especially algebraic) Topology.

On the face of it, it would seem that Ross was right. After all, Topology deals with large scale, global concepts, like ‘round’, ‘holey’, and ‘wholly’ (fortunately avoiding ‘holy’) while Probability is best at handling local questions, like, ‘Is a stochastic process continuous, or differentiable?’ Even if you think about the theorems of Probability and Statistics that say global things about large systems, almost all work from the infinitesimal up. Our most basic results, the laws of large numbers and the central limit theorem, work by saying that in a sum of many things, no single term can dominate, and it is this insignificance of the individual that eventually leads to the global phenomena at the core of Probability and Statistics.

Why should this matter? Well, in the first TOPOS column I argued that people who care about analyzing high dimensional data should care about Topology. Data invariably involves randomness, and so, despite what some of my Computer Science friends might claim, its analysis requires statistical thinking. In turn, the theory (and practice!) of Statistics is based on Probability. So, if you buy into this theme, we are going to have to marry Topology and Probability, despite Ross’s heartfelt angst.

In the second column I tried to introduce the uninitiated to Homology—the heart of Topology—via the notion of persistence homology and its pictorial representation via barcodes. I also noted that Homology is all about  $k$ -dimensional “holes” in  $n$ -dimensional sets. The number of such holes is called the  $k$ -th Betti number, after the 19th century Italian mathematician, Enrico Betti, and is denoted by  $\beta_k$ . For the rest of this column that is all you need to know about Homology; i.e. That there is a simple numerical variable called a Betti number, and a richer, more informative, and statistically

important construct known as a barcode.

Returning to the setting of the first column, there I was, in 2010, a statistical-probabilistic fish out of water, gasping in the dry air of my first Applied Topology conference, when I was accosted by a tumultuous troop of titillated topologists delighted by the belief that “now that we have captured a probabilist, he can tell us about the distributions of Betti numbers for random simplicial complexes”.

What complex? What is random? What do you expect to be able to discover? Why would you care?

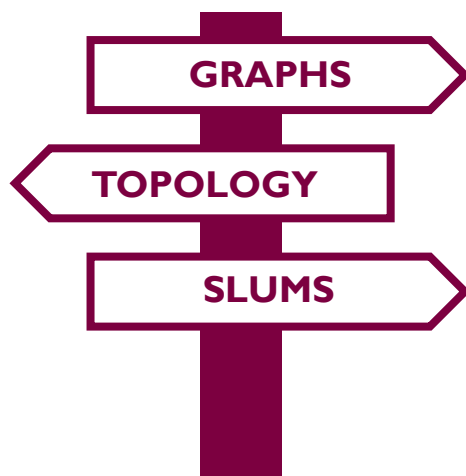
The “why would you care” was the question of the first column of this series, so let’s assume that issue is settled. As for “what complex”, here is their example:

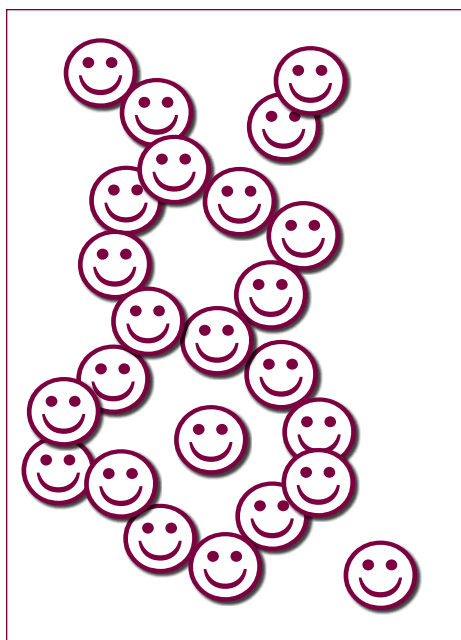
Take a set of points. (The points of a Poisson process, homogeneous or not, or iid observations from a distribution, perhaps a mixture distribution.) Join points that are close, thus obtaining a (very simple) random simplicial complex.

For a first question, they wanted to know what could be said about  $\beta_0$  of this object. “Ahh,” I said wisely, stroking my white beard (it has to be there for something useful beyond saving time shaving) “That’s easy. I know all about that.” Why so easy? Well,  $\beta_0$  just counts connected components, and this “random simplicial complex” of theirs was familiar. We cover it when we talk about things like connectivity in random graphs or networks, or percolation, or even graphical models. The truth is that I don’t know too much about these things, but my bookshelf is full of books by probabilists and statisticians covering these topics. Since I was representing the rest of you, saying “I” instead of “we” or “they” seemed like something I could get away with.

It turned out that mentioning graphs was a mistake, since their retort was the saying (attributed, apparently incorrectly, to Whitehead) that “*graphs are the slums of Topology*”. So, smirking, they upped the ante, placing a smiley on each of the random points, as in the picture on the next page. “Now tell us about the distribution of the homology of the union of smileys. Of course, the smileys are high dimensional, and so is the homology”.

It turns out that this particular model also has a familiar history. Stochastic geometers call it (a special case of) the “Boolean model”, and concentrate on finding formulae for expectations of geometric quantities like volume, surface area, etc. Others call it “continuum percolation” and ask for what critical smiley radius will such a structure,





when the number of smileys becomes large, become connected. Like most percolation problems, this one is easy to ask, but very hard to answer.

However, none of this helped me with the topologists. They did not care about geometric measurements. For a topologist, there is no such thing as “large” or “small”, only “hole-iness”.

Simple connectivity, of course, is beneath their contempt.

So, what can we say about the expectations of the Betti numbers of the union of  $n$ -dimensional smileys (leaving even the definition of “distribution of homology”, let alone results, to a later generation)?

Ross would claim that this problem is hard, and he is right. That, however, is not to say that it is impossible, and a brave cohort of young researchers is attacking these problems and proving some fascinating results. Taking the lead is Matthew Kahle, who stands out from the crowd in a number of ways, not the least of which is that he is the only mathematician I know who is so dedicated to his profession that he has a tattoo of a fractal on his biceps. Kahle and others have analysed the large scale behavior of mean Betti numbers for this and a number of other recipes for generating random simplicial complexes. There are now results about expectations, laws of large numbers, and even central limit theorems. (See the recent review by Kahle and Omer Bobrowski at [arxiv.1409.4734](https://arxiv.org/abs/1409.4734).)

This is not the place to give details about theorems, other than one. Actually, it is not yet a theorem, but more a collection of observations based on simulation, experience, and real theorems, all pointing in the direction of a universality phenomenon that surprised even seasoned topologists. The phenomenon is the following: In all models of random structure studied so far in which Betti numbers become large, there is always one that is an order of magnitude larger than the others. In other words, if we return to our version of homology that says it is all about gluing together spheres of different dimension, then there is one dimension that dominates all the others. Either we have, in three dimensions, lots of connected components and very few rings

and holes, or lots of rings, all in a few connected components with few voids, and so on. It seems that when Nature plays dice to build large random structures, she has a great deal of trouble building complicated ones, but concentrates on building sets with rather uniform homologies.

All of these results are about large systems, which is when Probability can apply its tools to ignore local structure. Saying something about small systems turns out to be harder, unless we take the lead from Euler.

Among the multitude of mathematical constructs that bear his name, lies the “Euler characteristic” of sets. For a 3D set, this is the number of components minus the numbers of rings plus the number of holes, and in general it is an alternating sum of Betti numbers.

The Euler characteristic (EC) is ubiquitous—it appears just about everywhere the words “topology” or “geometry” do, and it has a multitude of seemingly different definitions. The result is that much about the EC for random systems—both those of the smiley kind and those arising from continuous systems—is known. Indeed, a good part of my own career in random fields has involved the EC, and Jonathan Taylor has achieved mathematical miracles in working this up to a beautiful theory. My sorely missed friend Keith Worsley spent much of his career in Biostatistics developing what he and the British neuroscientist Karl Friston dubbed Topological Inference, applying results about the EC to analyze fMRI brain images.

So it turns out that if one wants ready-made results linking Probability and Topology, then one way to go is with the EC. Euler was right. He found something topologically deep, but simple enough for probabilists.

But the big challenge to Probability is still out there. Although results about Betti numbers are coming in, they are still mainly about means and limit theorems, More importantly, almost nothing is known about the distributional properties of barcodes, and these are the main tool of Applied Topology. Probabilistic results about barcodes are going to be crucial to developing a serious statistical theory behind their application.

In summary, we urgently need to develop new tools to attack one of the most challenging, interesting, and ultimately applicable problems around: describing the distributional properties of the algebraic topological structure of random systems. We need, simultaneously, to prove that Ross was wrong, and to move Probability out of the slums of thinking only in the trivial topology of graphs, moving it into a far richer and more promising domain of real topology.

**THEOREM:**  
When Nature plays  
dice, she generates  
topological simplicity