Higher-Order Methods for Data: Phase transitions from stochastic topology to point clouds and complex systems

Lecture Notes Companion - BeyondTheEdge

Fernando A. N. Santos

Korteweg–de Vries Institute for Mathematics (KdVI), Institute for Advanced Studies (IAS),
Dutch Institute for Emergent Phenomena (DIEP), University of Amsterdam (UvA), The Netherlands

f.a.nobregasantos@uva.nl

December 1, 2025

Higher-Order Methods for Data – Lecture Notes

These lecture notes are a companion to the *BeyondTheEdge* course *"Higher-Order Methods for Data"*, given online by Fernando A. N. Santos. The course draws on Dr. Santos's experience on topological phase transitions and connects to work by other scholars in the BeyondTheEdge network. It consists of two parts: **Part 1** introduces classical graph phase transitions and their higher-order generalizations, and **Part 2** explores topological phase transitions in brain networks and complex systems, giving one example of how and where algebraic topology and data analysis converge. The notes are informal, i.e, in a narrative (or storytelling) style, following the same structure as the online lecture. The papers used for preparing this course are quoted in the end of these notes. The course videos can be found at:

`https://www.beyondtheedge.network/event/course-higher-order-methods-for-data`

Part 1: From Erdős–Rényi Graphs to High-Dimensional Phase Transitions

Random Graphs and the Giant Component. A natural starting point is the classic Erdős–Rényi (ER) model of random graphs. In an $ER(n, p)$ graph with n vertices, each possible edge is present independently with probability p. As p increases, the graph undergoes a *phase transition*: around the critical value $p_c \sim \frac{1}{n}$, a single connected component containing a finite fraction of all nodes suddenly emerges. This *giant component* is absent for $p \ll 1/n$ but almost surely exists for $p \gg 1/n$. At exactly $p \approx 1/n$, the largest component size scales as $n^{2/3}$, a hallmark of criticality. Below p_c, all components remain tiny (size $O(\log n)$), and above p_c one finds one giant component and many small ones. Notably, this transition is *continuous* or second-order: as p passes the threshold, the giant component grows gradually (no instantaneous jump in size). In physics terms, the order parameter (fraction of nodes in the giant cluster) increases smoothly with p, analogous to a continuous phase transition with no latent heat. This behavior reflects the fact that adding edges just below p_c typically connects small tree-like clusters, whereas above p_c those clusters coalesce into one large network, growing steadily as p increases further.

Higher-Order Analogues – Random Simplicial Complexes. The question arises: do similar phase transitions occur in *higher-order networks* (hypergraphs or simplicial complexes) when we move beyond pairwise links? In combinatorial topology, a d-dimensional simplicial complex generalizes a graph by including not only vertices (0-simplices) and edges (1-simplices), but also filled triangles (2-simplices), tetrahedra (3-simplices), etc. Just as an ER graph is built by including edges at random, one can define *random simplicial complexes* by including higher-dimensional simplices with some probability.

A foundational result by Linial & Meshulam (2006) and by Kahle (2009) showed that these complexes indeed exhibit *thresholds* for the emergence of higher-dimensional topological features. In particular, there are critical probabilities at which homology in a given dimension appears with high probability. For example, Matthew Kahle identified that in the clique complex - a simplicial complex defined by all cliques of a graph G - $X(G(n,p))$ of an ER graph $G(n,p)$, the k-th homology H_k (loops, voids, etc.) almost surely "turns on" (becomes nontrivial) when p passes through a certain window (in terms of $p = n^{\alpha}$). In a random clique complex $X(G(n,p))$, the threshold at which the k-dimensional homology H_k becomes nontrivial is $p = n^{-1/(k+1)}$. Thus, in the parametrization $p = n^{\alpha}$, each Betti number β_k "turns on" at $\alpha = -\frac{1}{k+1}$. Below that range H_k vanishes, and above it H_k persists. This was a higher-dimensional analogue of the Erdős–Rényi connectivity threshold for H_0 (connected components).

Abrupt Topological Transitions. A surprising discovery was that, unlike the graph case ($d = 1$), higher-dimensional phase transitions can be *discontinuous* (first-order). Linial and Peled (2016) proved that for random complexes in dimension $d \geq 2$, the onset of a "giant" cycle or void is a sharp, abrupt transition. At a critical probability p_c, a giant $(d-1)$-dimensional hole emerges essentially fully formed rather than growing gradually from small cycles. They introduced the notion of a *giant shadow*, which generalizes the giant component: for $d = 1$ a giant shadow corresponds to the giant connected component, but for $d \geq 2$ it signifies a spanning homological cycle appearing at the threshold. Crucially, the birth of the giant shadow is a *first-order phase transition* for $d \geq 2$. In other words, just below p_c there are no extensive cycles, and just above p_c a system-spanning cycle exists – an abrupt jump. This finding marked one of the first examples where moving to higher-order interactions fundamentally changed the nature of a phase transition. Intuitively, near the threshold almost every new higher-dimensional simplex (e.g. a triangle, tetrahedron, etc.) you add closes a loop, so the complex suddenly gains many cycles at once. This contrasts with graphs, where adding a single edge near criticality typically creates at most one new cycle.

Connections with complex systems, inspiration, and applications. These mathematical insights by Linial, Peled, Meshulam, and Kahle formed the theoretical backbone for much recent work. In fact, the Euler characteristic χ of a complex serves as a succinct summary of all Betti numbers (alternating sum), and its expected value crosses zero at the moment a new homology group appears. The zeros of $\mathbb{E}[\chi]$ act as estimators of the critical points for giant-cycle formation – an idea we will see applied to real data in Part 2.

Beyond stochastic topology, analogous discoveries emerged in other domains of pure and applied maths. For example, discontinuous transitions induced by higher-order interactions have been observed in dynamics on networks. One prominent case is *explosive synchronization* in higher-order oscillator systems. Millán et al. (2020) showed that extending the Kuramoto model to include phase-coupling not just along edges but also on triangles, etc., can yield an *explosive synchronization transition* (a sudden jump to global phase-locking). In their higher-order Kuramoto model, the presence of simplex-based coupling (combined with an adaptive feedback) causes the normally second-order sync transition to become abrupt. This dynamical result mirrors the combinatorial findings: higher-order interactions can induce first-order phase transitions where only continuous ones existed in pairwise systems. Such phenomena appear widespread – from epidemic dynamics on hypergraphs to social contagion models, many "beyond pairwise" systems show qualitatively new, often discontinuous behaviors.

The general lesson from Part 1 is that when we move beyond graphs to consider triangles, tetrahedra, and higher simplices, we enter a new regime of phase transition behavior. Next, we explore how these concepts carry into data analysis and neuroscience, drawing connections between topology, physics, and real-world complex systems.

Part 2: Topological Phase Transitions in High-Order Networks

Topology Meets Complex Systems. Real-world complex systems (like the brain) often defy easy modeling – their "microscopic" dynamics and interactions are not fully known. However, we can still measure activity or connectivity data and ask: *does the system exhibit phase transition-like phenomena?* Here is where algebraic topology, especially *Topological Data Analysis* (TDA), becomes invaluable. Topology studies properties of shapes that remain invariant under continuous deformations. By applying it to data (e.g. treating a dataset or network as a topological space), we can detect robust structural features (holes, loops, connectivity) without requiring a detailed mechanistic model. In other words, topology provides an intrinsic, model-free framework to probe complex systems. This approach has gained traction in many fields – from neuroscience to materials science – precisely because topological features are coordinate-free and noise-tolerant, picking up patterns that traditional analyses might miss. In neuroscience, for instance, network science traditionally uses graph theory to study brain connectivity, but recently TDA techniques have been introduced to go beyond pairwise connections and capture higher-order structure, offering improved robustness against noise.

Simplicial Complexes from Data. To apply topology to data, we first need to construct a space from the data points or network. One common method is to use a Vietoris–Rips complex or a clique complex. In a graph (e.g. a functional brain network where nodes are brain regions and edges link significantly correlated regions), every clique (complete subgraph) can be "filled in" as a simplex. For example, three nodes that all connect pairwise form a triangle (2-simplex) in the clique complex, four fully interconnected nodes form a tetrahedron, and so on. The result is a simplicial complex capturing higher-order interactions present in the data.

Importantly, we can build a *filtration*: imagine a threshold τ on the edges (for brain data, τ

could be a correlation strength). Start with τ high so that no edges are present, then gradually lower τ to add more and more connections. This generates a nested sequence of complexes from empty to full graph. At each threshold, one can compute topological invariants like Betti numbers β_k (the number of k-dimensional holes) or the Euler characteristic.

Euler Characteristic as a Global Invariant. The Euler characteristic χ is a classic topological invariant defined as an alternating sum of the number of simplices of each dimension:

$$\chi = \#\text{vertices} - \#\text{edges} + \#\text{faces} - \#\text{tetrahedra} + \cdots .$$

For a familiar 3D object like a convex polyhedron, this formula reduces to $V - E + F = 2$ (Euler's formula for polyhedra), which is actually the Euler characteristic of a sphere. If the shape has a hole (like a donut has a genus-1 hole), the Euler characteristic is different (for a torus, $\chi = 0$). In higher dimensions, we continue the alternating sum for higher-order simplices.

The Euler characteristic has several key advantages for data analysis:

- It is easy to compute even for large complexes (just count simplices), making it a "cheap" summary of topological structure.
- It is intrinsic – invariant under continuous deformations of the space (so robust to noise or smooth changes in data).
- Crucially, it encodes information about all Betti numbers at once, albeit in aggregate form (in fact $\chi = \sum_k (-1)^k \beta_k$). Changes in homology (appearance or disappearance of holes) will be reflected as changes in χ.

In particular, when χ crosses zero or has a singular behavior as a function of the filtration parameter, it often signifies a topological phase transition. This idea has support on stochastic topology, especially in the works of Kahle, Linial and Peled, that the Betti numbers are localized in specific probabilities intervals.

Topological Phase Transitions and Euler Entropy. There is a deep analogy between topological features in data and thermodynamic phase transitions in physics. In physics, a phase transition is often signaled by a non-analytic change in a thermodynamic potential or order parameter (e.g. the jump in magnetization in a ferromagnet or a divergence of specific heat at criticality). In fact, from a mathematical viewpoint, phase transitions often happen at the vicinity of zeros of partition functions. Similarly, if one sees χ as a topological partition function, one can treat $-\ln|\chi|$ as a kind of topological entropy measuring the logarithm of the count of "states" (simplices). When this Euler entropy shows singularities as a function of some control parameter, it indicates a profound structural reconfiguration – essentially a *topological phase transition.*

Indeed, past research had shown in certain spin models that phase transitions coincide with changes in the topology of configuration space sublevel-sets. For example, in the mean-field XY model and the ϕ^4 lattice model, the known thermodynamic transition was retrieved by persistent homology analysis of the energy landscape (Donato et al., 2016). This finding provided an early validation that TDA can detect phase transitions from data alone, without prior knowledge of the order parameter. The Euler characteristic has been proposed as an entropy-like quantity in various contexts, with the idea that zeros or sign changes of χ signal topological or thermodynamic phase transitions in the thermodynamic limit – an idea now borne out in several models in point cloud data and complex systems. Bobrowski & Skraba (2020) later sharpened this perspective by showing in various random models that zeros of the expected Euler characteristic $\mathbb{E}[\chi]$ closely approximate critical p_c values for formation of giant loops.

Brain Networks and complex systems. Functional brain networks offer a prime example of a complex system where these ideas have been applied. The brain can be viewed as a network of regions (nodes) with weighted links representing functional connectivity (e.g. correlation in activity). Because brain activity data are noisy and extremely complex, traditional graph metrics give only a partial view. By constructing a clique complex of the brain network at varying correlation thresholds, one can analyze the multi-scale topology of connectivity. Santos and colleagues did exactly this, seeking topological phase transitions in brain data. They computed the Euler characteristic of the clique complex as a function of the threshold (from no edges to all edges) for resting-state fMRI networks. Strikingly, they observed multiple points where χ crosses zero – corresponding to the emergence of giant loops, voids, etc., in the network structure. In fact, the first zero of χ occurs around the threshold where a giant connected component forms (the classic percolation transition), and the next zero occurs where the first giant 1-dimensional cycle (loop) appears, followed by higher-dimensional cycles. Each zero of the Euler characteristic thus marks a *topological phase transition* in the brain network, wherein a new "giant" topological feature spans the system.

Plotting the Betti numbers β_k as well, one sees each β_k peaks around its respective transition, mirroring the behavior in random simplicial complexes. The Euler entropy $-\ln|\chi|$, when plotted, shows sharp spikes (singularities) at those thresholds – just like specific heat spikes in physical phase transitions.

This finding – reported as the discovery of topological phase transitions in functional brain networks (Santos et al., 2019) – suggests that the brain's functional connectivity is organized in a way that as you tune connection density, the network "reconfigures" through abrupt topological changes. Moreover, the geometrical nature of these transitions can be interpreted as a high-dimensional generalization of percolation: instead of just a giant cluster of nodes forming, we see giant loops and voids forming, implying the brain network has a latent higher-order organization. This was directly inspired by the theoretical work of Linial, Kahle, and others from Part 1, now applied to empirical data. It also aligns with Bobrowski & Skraba's homological percolation concept – indeed the brain results appeared shortly before that terms was coinned, and provided a real-world confirmation that Euler characteristic zeros pinpoint critical points for topological transitions, or homological "percolation".

Applications and Further Insights. Once we have these topological phase transition points, we can treat them as system fingerprints. Just as the boiling point of water (100°C at 1 atm) is a material constant, the threshold for a topological transition might characterize a given network or state of a system. In the brain, researchers compared healthy vs. diseased states using such topological markers. For example, using a cohort of healthy subjects versus patients with brain tumors (gliomas), they found statistically significant differences in the Euler characteristic curves and transition points between the groups. In other words, the critical thresholds at which loops or voids emerge in the functional network differed between healthy and pathology, suggesting a potential biomarker of network organization. This parallels the idea that a material's critical temperature is an intrinsic property; here the brain's network has intrinsic critical points that may shift with disease or other conditions.

Beyond global invariants like χ, one can also examine local topological measures. An example is network curvature, based on a discrete analogue of the Gauss–Bonnet theorem (which relates the integral of curvature over a surface to its Euler characteristic). For networks, certain definitions of curvature (e.g. Forman–Ricci curvature) allow assigning a scalar to each node or link. It turns out that as a network passes through a topological phase transition, the distribution of node curvatures changes dramatically. In the brain data, near the critical threshold almost all nodes attained curvature ≈ 0 (flatness), whereas below the threshold many nodes had negative curvature and above it many had positive curvature – indicating a curvature sign-flip as the system reorganizes into a new phase. This is consistent with the Euler characteristic (global curvature sum) going through zero at the transition. Such local signatures could prove useful for identifying critical regions or "hub" nodes driving the transition.

Universality and Outlook. The convergence of algebraic topology with phase transition theory opens many questions. One exciting development is the discovery of universal patterns in the topological characteristics of random data. Bobrowski & Skraba (2023) recently provided evidence that when persistence diagrams (the multiscale summary of topological features) from random point clouds are properly normalized, they obey a universal probability law, independent of the specific data distribution. In other words, much like many physical systems fall into a few universality classes of critical behavior, random topological features might follow a single universal distribution. This unexpected finding – analogous to a topological Central Limit Theorem – suggests a deep underlying order in the "noise" part of TDA. It allows researchers to define a universal null model for persistence, enabling rigorous significance testing for observed topological features. Thus, one can quantify how unlikely a given large hole or loop is, relative to a null hypothesis of random structure.

In summary, Part 2 demonstrated how higher-order topology provides a powerful lens on complex data:

- we can detect phase transitions without knowing the order parameter,
- identify multi-scale homological structures (from clusters to loops to voids) in networks,
- and possibly uncover universal laws governing these structures.

The lecture shows that concepts from algebraic topology and statistical physics are not only mathematically analogous but can be unified to yield insights in neuroscience and beyond. As the Beyond the Edge network continues to connect these ideas across disciplines, we anticipate further developments – for instance, mapping other physics phenomena (Bose–Einstein condensation, synchronization, etc.) to network analogues, and refining the notion of universality classes in high-dimensional data.

References

[1] Erdős, P., & Rényi, A. (1960). *On the evolution of random graphs.* Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17–61.

[2] Linial, N., & Meshulam, R. (2006). *Homological connectivity of random 2-complexes.* Combinatorica, **26**(4), 475–487. doi:10.1007/s00493-006-0027-9

[3] Kahle, M. (2009). *Topology of random clique complexes.* Discrete Mathematics, **309**(6), 1658–1671. doi:10.1016/j.disc.2008.02.037

[4] Linial, N., & Peled, Y. (2016). *On the phase transition in random simplicial complexes.* Annals of Mathematics, **184**(3), 745–773. doi:10.4007/annals.2016.184.3.3

[5] Millán, A. P., Torres, J. J., & Bianconi, G. (2020). *Explosive higher-order Kuramoto dynamics on simplicial complexes.* Physical Review Letters, **124**(21), 218301. doi:10.1103/PhysRevLett.124.218301

[6] Donato, I., Gori, M., Pettini, M., Petri, G., De Nigris, S., Franzosi, R., & Vaccarino, F. (2016). *Persistent homology analysis of phase transitions.* Physical Review E, **93**(5), 052138. doi:10.1103/PhysRevE.93.052138

[7] Santos, F. A. N., Raposo, E. P., Coutinho-Filho, M. D., Copelli, M., Stam, C. J., & Douw, L. (2019). *Topological phase transitions in functional brain networks.* Physical Review E, **100**(3), 032414. doi:10.1103/PhysRevE.100.032414

[8] Bobrowski, O., & Skraba, P. (2020). *Homological percolation and the Euler characteristic.* Physical Review E, **101**(3), 032304. doi:10.1103/PhysRevE.101.032304

[9] Centeno, E. G. Z., Moreni, G., Vriend, C., Douw, L., & Santos, F. A. N. (2022). *A hands-on tutorial on network and topological neuroscience.* Brain Structure and Function, **227**(3), 741–762. doi:10.1007/s00429-021-02435-0

[10] deÁmorim Filho, E. C., Moreira, R. A., & Santos, F. A. N. (2022). *The Euler characteristic and topological phase transitions in complex systems.* Journal of Physics: Complexity, **3**(2), 025003. doi:10.1088/2632-072X/ac664c

[11] Bobrowski, O., & Skraba, P. (2023). *A universal null-distribution for topological data analysis.* Scientific Reports, **13**(1), 12274. doi:10.1038/s41598-023-37842-2

[12] Castro, D. M., Raposo, E. P., Copelli, M., & Santos, F. A. N. (2025). *Interdependent scaling exponents in the human brain.* Physical Review Letters, **135**(19), 198401. doi:10.1103/lvwj-hjr3