Have Graph — Will Lift? The Case for Higher-Order Benchmarks

Bastian Rieck BASTIAN.GROSSENBACHER@UNIFR.CH
AIDOS Lab, University of Fribourg, Switzerland;
Institute of AI for Health, Helmholtz Munich, Germany

Editors: Michael Bleher, Freya Jensen, Levin Maier, Diaaeldin Taha, and Anna Wienhard

Abstract

After a somewhat rocky start, geometry and topology have established a foothold in machine learning. Message passing, either on graphs or higher-order complexes, is one of the main drivers of *geometric deep learning*, and paradigms that were once considered to be firmly in the realm of the abstract—like sheaves—have been "tamed" to serve as novel inductive biases for model architectures in *topological deep learning*. The veritable diversity of models, however, is in stark contrast to the scarcity of suitable benchmark datasets. As a result, researchers often resort to *lifting* existing graph datasets to include higher-order information. In this opinion paper, I want to encourage the community to also source new datasets, which may be used to prop up the foundations of our research field.

1. Introduction

Graphs are an ubiquitous data modality, permitting researchers to model complex phenomena in numerous domains (Veličković, 2023). For a long time, the analysis of graphs, commonly referred to as *graph learning*, was dominated by *graph kernels*, i.e., similarity measures based on structural information such as the isomorphism class of a graph (Morris et al., 2023; Borgwardt et al., 2020). In their seminal article, Bronstein et al. (2017) outlined the field of *geometric deep learning*, which provides "blueprints" for developing models that are capable of tackling data modalities like graphs and manifolds. The focus on continuous and metric properties of data by geometric deep learning was recently complemented by an algebraic and combinatorial perspective. Methods of this sort are now commonly assigned to the field of *topological deep learning* (Bodnar, 2022), which promises to improve relational learning (Papamarkou et al., 2024).

With both *geometric deep learning* (GDL) and *topological deep learning* (TDL) making use of computational paradigms like *message passing*—albeit often on different modalities, i.e., graphs (GDL) versus higher-order data like simplicial complexes (TDL)—the lines between the fields become increasingly blurry. As such, attempts to provide a precise separation are doomed to fail and this article will refrain from making them. Instead, I want to focus on the way GDL/TDL researchers *analyze* their proposed methods. A typical evaluation pipeline involves working with graph benchmark datasets and, depending on whether one is interested in assessing the performance of a higher-order method or not, *lifting* such graphs to a higher-order representation like a simplicial complex. Hence, instead of working on the original graph $\mathcal{G}$, one rather works on an associated complex $\mathcal{C}$. Any improvements in

(a) Clique complex (b) Neighborhood complex

Figure 1: An illustration of two different lifting strategies for a graph $G = (V, E)$. The *clique complex* $\mathcal{C}_{\text{cl}}$ contains all subsets whose induced graphs are *complete*, i.e., $\mathcal{C}_{\text{cl}} \coloneqq \{A \subseteq V \mid G[\sigma] \text{ is complete}\}$. The *neighborhood complex* $\mathcal{C}_{\text{nb}}$, by contrast, contains all simplices whose vertices have a common neighbor, i.e., $\mathcal{C}_{\text{nb}} \coloneqq \{A \subseteq V \mid \text{CN}(A) \neq \emptyset\}$, with $\text{CN}(A) \coloneqq \{v \in V \mid (v, a) \in E \text{ for all } a \in A\}$. The striking difference between the two constructions is that the clique complex contains only a single 2-simplex, whereas the neighborhood complex contains two of them. In both cases, the original graph appears as the shadow of the respective complex; the "right" choice of lifting strategy is unclear a priori and depends on the dataset as well as the respective task. Lifting is thus an inherent part of the model selection process and should be treated as such.

predictive performance are then considered to reflect the positive influence of the proposed method. While this approach is not wrong *per se*, I want to outline some of its shortcomings and, later on, propose alternatives, which I believe would be generally beneficial for the GDL/TDL community at large.

2. Quo Vadis, Graph?

Lifting graphs to simplicial complexes or cellular complexes has been a staple of TDL research since the beginning. In Bodnar et al. (2021b), graphs are lifted to simplicial complexes by building their *clique complex*, i.e., a simplicial complex in which each clique of k vertices is represented by a $(k-1)$-dimensional simplex. This is later on extended to cellular complexes (Bodnar et al., 2021a). Other complexes have been described in the literature, though, such as the *neighborhood complex* (Matoušek, 2003, p. 130), in which simplices are subsets of vertices that share a common neighbor, or the *Dowker complex* (Dowker, 1952), which requires the existence of a binary relation defined on the vertices. Even more (simplicial) complexes are described by Jonsson (2008), typically with the goal of providing insights into the topology of the underlying graphs and its properties. In the context of *topological data analysis*, lifting is often conflated with the process of creating a *filtration* of a graph (or complex), i.e., a nested way of indexing individual substructures, leading to an independent rediscovery of some lifting methods (Aktas et al., 2019).

Figure 1 depicts two lifting strategies. The resulting complexes reflect different aspects of the underlying graph. Hence, the lifting strategy needs to be treated as a crucial part of a model and depends on the respective task. I cannot help but draw a parallel to "The Bitter Lesson" (Sutton, 2019) at this point: In many cases, static lifting strategies seem to be hand-crafted features, which, ultimately, may be replaced by a more general, highly-scalable computational paradigm. This is not to say that lifting strategies are not useful,

though. In fact, the existence of modern TDL libraries (Telyatnikov et al., 2025) can be a boon for improved benchmarking, provided researchers are *aware* of how to interpret the choice of lifting strategy. It is unlikely that there is a "one-size-fits-all" lifting procedure, hence any specific choice of strategy should always be reported together with the model and may not generalize. The best way to reason about lifting strategies is to consider them as a part of a *multiverse analysis*, i.e., an analysis that explicitly encodes different choices (such as the selection of a lifting strategy or the selection of a model as well as the tuning of hyperparameters), aiming to make research more transparent and more reproducible (Steegen et al., 2016). Our previous work found that latent representations of generative models, for instance, are highly sensitive with respect to their initial parameter choices (Wayland et al., 2024). The necessary ablation studies are, unfortunately, often absent from papers, making it hard to assess the origin of gains in task performance. In particular if the use of a lifting strategy results in a model with more parameters, care must be taken so as not to confuse the source of any observed improvements.

In a nutshell

For GDL/TDL applications using graph datasets, this implies that *all* choices of lifting, including the *absence* of a lifting strategy altogether, should be reported as ablations.

This piece of advice is *not* obviated by recent work, which presents the first instance of *learnable* lifting strategies (Franco et al., 2026). While learnable strategies are certainly a step in the right direction, they need to be assessed critically like any other crucial choice of model architecture. All of this pontificating is not meant to downplay any of the numerous advances in GDL/TDL. However, there have been some recent discussions concerning the overall quality of our benchmark datasets and our research culture (Bechler-Speicher et al., 2025). More precisely, there is some disconcerting evidence that the graph structure of many datasets is largely *irrelevant*, implying that task performance can be improved by training on an empty graph or on a randomly-rewired one (Coupette et al., 2025). Put somewhat differently, we, as a community, are just now starting to understand the intricate interplay between graph *structure* and graph *features*. Moving to higher-order domains will only exacerbate the complexity of this problem—we have already seen evidence for this when extending *message passing* to simplicial complexes, for example (Taha et al., 2025).

In a nutshell

Due to its influence on model design and training, lifting procedures may hide the *relevance* of using higher-order information, especially for datasets in which the interplay between features and structure is not (yet) well understood.

3. Ad Astra!

Given that lifting may, in the worst case, hide the very thing we are interested in studying, what are our options? I believe a viable path to be the collection and construction of new datasets. This should make gainful use of all the hard lessons gleaned from graph learning. Inspired by the issues raised in Bechler-Speicher et al. (2025), here are some suggestions:

1. Datasets need to be constructed in a *meaningful* way. The higher-order structure needs to be *crucial* and *advantageous* for the task at hand.

2. Task *complexity* should be known, for instance by including the performance of simple (graph) baselines. This point is closely related to the previous one.

3. Standard *splits* should be available to ensure reproducibility and understand predictive performance. This should be accompanied by a certain minimum *size* of the dataset to enable proper statistical reasoning.[1]

4. Datasets should be under *version control* to enable their subsequent modification and error correction. While not always necessary, a proper *maintenance plan* is required to account for community-led improvements.

All of these items require a better understanding of the *provenance* of the data. That is, among other things, we need to understand *who* collected it for *what* purpose and *how* it is supposed to be maintained. For many datasets in graph learning, these questions cannot be reliably answered, unfortunately. A quick analysis[2] of patterns in the source code of `pytorch-geometric` (Fey et al., 2025), arguably one of the standard software packages for graph learning, shows that datasets are hosted on a variety of domains, including Amazon AWS (5), `dropbox.com` (6), and the currently defunct `graphmining.ai` (6). Several datasets thus do not seem to be directly under the control of the research community, setting us up for a replication crisis in the future—even with the best intentions, datasets might vanish because people move on to different jobs or delete data by accident. There is no need for pessimism: Since this is an open-source project, *we*, as a research community, can make things better. In the end, we will also benefit from it because we presumably want to test our new models under the best conditions to make sure that any signal we are receiving is actually part of the data and not just an artifact.[3] If the recent upheavals in graph learning tells us only one thing, it should be that we need to adopt a more critical approach towards scholarship. Lipton and Steinhardt (2018) already mentioned general patterns that we would do well to avoid, but in the age of large datasets, adopting improved practices for "data stewardship" will only help us advance. Better reporting like paper checklists or

1. Older graph datasets are often limited in size and the lack of predefined splits necessitates using techniques like cross-validation. However, cross-validation needs to be *repeated* to properly capture standard deviations (Errica et al., 2020; O'Bray et al., 2021).
2. Due to the way code in `pytorch-geometric` is structured, the following pattern will find most uses of dataset URLs:

```
rg -t py "^\s*url ="                                    \
  | grep -Eo "(http|https)://[a-zA-Z0-9./?=_%:-]*" \
  | awk -F/ '{print $3}'                                \
  | sort | uniq -c | sort -rn
```

3. The words of Feynman (1985) should always be at the forefront of our mind:

   > The first principle is that you must not fool yourself—and you are the easiest person to fool.

datasheets (Gebru et al., 2021) can play a part here but only if we treat the remainder of the data collection and curation process as something worthwhile *per se*.

In a nutshell

We need to treat *data collection*, *data curation*, and *data maintenance* as integral steps of the model development process.

4. Acta Non Verba

After pointing out issues, it is time for something more concrete and savory. As a first step towards eventually establishing a large-scale database of well-curated higher-order datasets, we started creating MANTRA,[4] the "Manifold Triangulations Assemblage" (Ballester et al., 2025). MANTRA is based on the work of the late Frank H. Lutz (Lutz, 2017, 2008), who enumerated (minimal) triangulations of 2-manifolds and 3-manifolds and, where possible, provided a classification in terms of topological invariants. This resulted in a vast dataset of more than 43,000 2-manifolds and more than 250,000 3-manifolds, all of which are represented as *abstract simplicial complexes*. Unlike other datasets, these triangulations are intrinsically higher-order—there is no lifting or prior modeling involved. Moreover, the original data contains only combinatorial information, i.e., there are no features for vertices, edges, and simplices. Even in the case of 2-manifolds, where a *random realization algorithm* (Lutz, 2008) can provide coordinates, these triangulations have no "canonical" way of being imbued with coordinates. As Figure 2 illustrates, the random realizations do not help in recognizing the underlying manifold, thus increasing dataset complexity. On the bright side, however, an abstract, coordinate-free representation lends itself to a simple textual description. Here is an example entry of the database we curated, showing one particular triangulation of the 2-sphere as well as some associated properties:


```
{
  "id": "manifold_2_4_5_1",
  "triangulation": [[1,2,3], [1,2,4], [1,3,4], [2,3,4]],
  "dimension": 2,
  "n_vertices": 4,
  "betti_numbers": [1, 0, 1],
  "torsion_coefficients": ["", "", ""],
  "name": "S^2",
  "genus": 0,
  "orientable": true,
  "vertex_transitive": true
}
```


4. https://github.com/aidos-lab/MANTRA

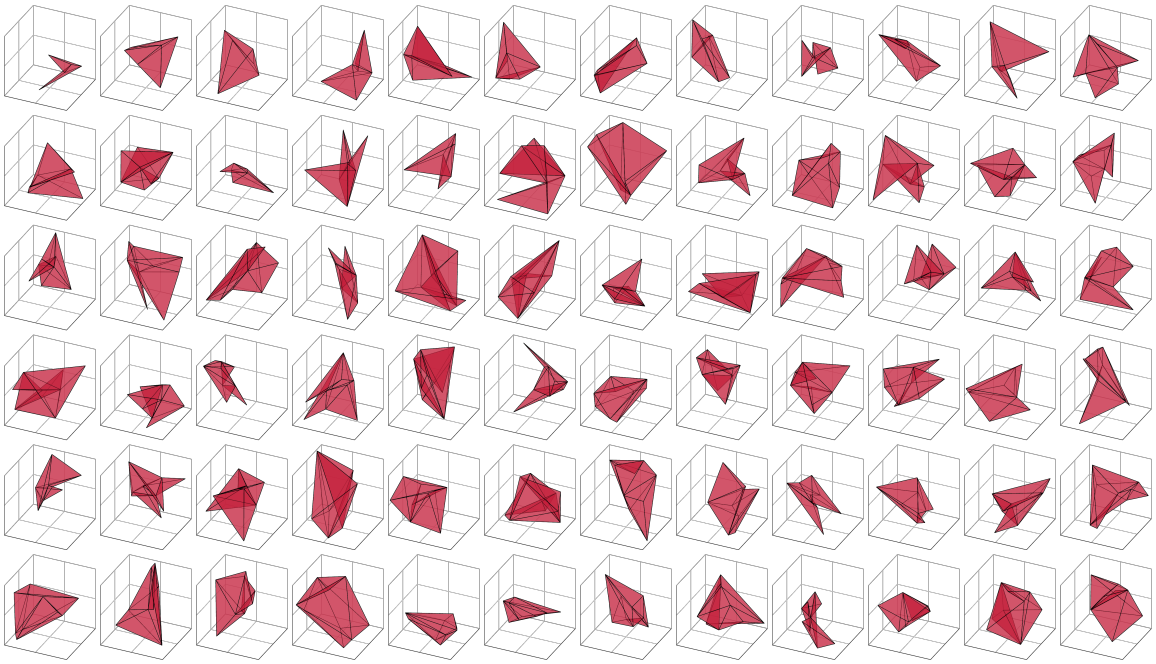

Figure 2: A visualization of some triangulations of S^2, the 2-dimensional sphere, contained in the MANTRA dataset. The task of recognizing the topological type is well-defined but, due to the absence of *canonical* coordinates, difficult for contemporary GDL/TDL models.

Some of this information is *redundant*—for instance, *orientability* can also be determined purely based on the last Betti number—or *specific* to 2-manifolds (like the *genus*). However, even in this abstract form, the dataset gives rise to *regression* or *classification* tasks:

1. Predict Betti numbers or torsion coefficients (regression).
2. Predict orientability, vertex transitivity, or canonical name (classification).

These tasks can be addressed in a myriad of ways, making, for instance, use of higher-order message passing methods (Telyatnikov et al., 2025), exploiting graph representations like the 1-skeleton, or employing transformer-like structural encodings (Ballester et al., 2024; Carrasco et al., 2025).

Are these tasks hard? However, in light of the issues raised above, we should first think about what we are trying to accomplish here—these tasks appear somewhat "natural," but what about their utility and complexity? Here, we need to go on a brief tangent: Topologists have fully classified (closed) 2-manifolds already. Every closed 2-manifold is known to be homeomorphic to either S^2, a *connected sum* of k tori, i.e., $\#^k T^2$, or a *connected sum* of k projective planes, i.e., $\#^k \mathbb{RP}^2$. Consequently, the *Euler characteristic* and *orientability* are sufficient to classify a 2-manifold. Thus, for 2-manifolds, the tasks described above serve as an important "smoke test" of overall model quality: Any *topological deep learning* model should aim to exhibit perfect predictive performance here. Success would indicate

that the model is sufficiently *expressive* to learn basic topological invariants. The situation for 3-manifolds is more complex—while the classification based on canonical name is simple for the current version of the dataset, there is still value in models trying to predict other properties (Betti numbers, torsion coefficients), *despite* the fact that homology algorithms solve them. Moreover, there is no need to stop at dimension 3. Even though higher-dimensional manifolds are known to be impossible to classify (Kreck, 2014), a large-scale dataset could help address research-level questions for topologists. While we are not there yet, I am convinced that if our TDL models remain incapable of going well beyond simplicial homology calculations, we are likely experiencing *fundamental challenges* in model design that we have yet to overcome.

The performance of contemporary models. Our preliminary experiments (Ballester et al., 2025) indicate that both GDL and TDL models struggle to solve the aforementioned tasks. While TDL models (specifically, those making use of higher-order message passing) exhibit overall slightly better performance than their GDL counterparts (which are by necessity restricted to a graph representation of the data, i.e., the 1-skeleton), even the case of 2-manifolds is far from being solved. And this is not even scratching the surface of the transformations that we have at our disposal when dealing with triangulations. For instance, classical constructions like *barycentric subdivision* do not change the homeomorphism type of a manifold. Nevertheless, when subdividing a triangulation, we observe performance decreases. The (sobering) conclusion of these experiments is that the current generation of TDL models are *not* making use of the inherent topology of the data but rather perform local aggregations on combinatorial structures. As such, they are focusing on a *combinatorial* prior that is not aligned with the *topological* equivalence relation required to solve the tasks.

We are still at the beginning. I strongly believe that this should only be the *beginning* of the story, however. To better understand the failure modes of our models, we need other datasets of the same ilk, i.e., with known provenance, well-defined tasks, and intrinsically higher-order, both *with* and *without* node coordinates. It is my hope that, through community contributions, MANTRA may become a repository of many diverse triangulations, a higher-order companion to collections like GraphBench (Stoll et al., 2026). MANTRA already provides versioning information, release tracking, and many additional conveniences that ensure proper maintenance and reproducibility—everyone is cordially invited to participate!

5. Conclusion

Having hopefully demonstrated the necessity and utility of new and extended benchmarks, I want to raise some open questions, which may serve as inspiration for advancing the field:

1. How can we move *beyond* expressivity measures borrowed from graph learning? We recently looked into how structural measures that explicitly incorporate feature information could be defined (Carrasco et al., 2026); an equivalent for higher-order data would seem appropriate and useful.

2. What other (manifold-based) datasets would be appropriate as benchmarks? Can we find some that are *intrinsically* equipped with coordinates/features?

3. Which *other* invariants or properties do we want to predict? Even in the case of 3-manifolds, computational topology has developed new algorithms to measure numerous properties that are characteristic of a triangulation or its underlying manifold (Burton, 2004; Burton et al., 1999–2025).

Whatever we decide to do, we should ensure that TDL methods are not merely *inspired* by topological concepts but are also equipped with strong topological guarantees.

Acknowledgments

I am grateful for the opportunity to discuss these issues with the community and extend heartfelt thanks to the organizers of the Workshop on Geometry, Topology, and Machine Learning (GTML 2025) for accommodating me. The opinions in this article have been honed in many discussions with my collaborators, in particular Mathieu Alain, Rubén Ballester, Ernst Röell, Martin Carrasco, Corinna Coupette, Kelly Maggs, Emily Simons, and Jeremy Wayland. Any errors in judgment are solely my responsibility.

References

Mehmet E. Aktas, Esra Akbas, and Ahmed El Fatmaoui. Persistence homology of networks: Methods and applications. *Applied Network Science*, 4(1), 2019.

Rubén Ballester, Pablo Hernández-García, Mathilde Papillon, Claudio Battiloro, Nina Miolane, Tolga Birdal, Carles Casacuberta, Sergio Escalera, and Mustafa Hajij. Attending to topological spaces: The cellular transformer, 2024. URL https://arxiv.org/abs/2405.14094.

Rubén Ballester, Ernst Röell, Daniel Bīn Schmid, Mathieu Alain, Sergio Escalera, Carles Casacuberta, and Bastian Rieck. MANTRA: The manifold triangulations assemblage. In *International Conference on Learning Representations*, 2025.

Maya Bechler-Speicher, Ben Finkelshtein, Fabrizio Frasca, Luis Müller, Jan Tönshoff, Antoine Siraudin, Viktor Zaverkin, Michael M. Bronstein, Mathias Niepert, Bryan Perozzi, Mikhail Galkin, and Christopher Morris. Position: Graph learning will lose relevance due to poor benchmarks. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 81067–81089. PMLR, 2025.

Cristian Bodnar. *Topological Deep Learning: Graphs, Complexes, Sheaves*. PhD thesis, University of Cambridge, 2022.

Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Liò, Guido F. Montufar, and Michael Bronstein. Weisfeiler and Lehman go cellular: Cw networks. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors,

Advances in Neural Information Processing Systems, volume 34, pages 2625–2640. Curran Associates, Inc., 2021a.

Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido F. Montufar, Pietro Lió, and Michael Bronstein. Weisfeiler and Lehman go topological: Message passing simplicial networks. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 1026–1037. PMLR, 2021b.

Karsten Borgwardt, Elisabetta Ghisu, Felipe Llinares-López, Leslie O'Bray, and Bastian Rieck. Graph kernels: State-of-the-art and future challenges. *Foundations and Trends® in Machine Learning*, 13(5–6):531–712, 2020.

Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Benjamin A. Burton. Introducing Regina, the 3-manifold topology software. *Experimental Mathematics*, 13(3):267–272, 2004.

Benjamin A. Burton, Ryan Budney, William Pettersson, et al. Regina: Software for low-dimensional topology, 1999–2025. URL https://regina-normal.github.io.

Martin Carrasco, Guillermo Bernardez, Marco Montagna, Nina Miolane, and Lev Telyatnikov. HOPSE: Scalable higher-order positional and structural encoder for combinatorial representations, 2025. URL https://arxiv.org/abs/2505.15405.

Martin Carrasco, Olga Zaghen, Kavir Sumaraj, Erik Bekkers, and Bastian Rieck. Graph homomorphism distortion: A metric to distinguish them all and in the latent space bind them, 2026. URL https://arxiv.org/abs/2511.03068.

Corinna Coupette, Jeremy Wayland, Emily Simons, and Bastian Rieck. No metric to rule them all: Toward principled evaluations of graph-learning datasets. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 11405–11434. PMLR, 2025.

Clifford H. Dowker. Homology groups of relations. *Annals of Mathematics*, 56(1):84–95, 1952.

Federico Errica, Marco Podda, Davide Bacciu, and Alessio Micheli. A fair comparison of graph neural networks for graph classification. In *International Conference on Learning Representations*, 2020.

Matthias Fey, Jinu Sunil, Akihiro Nitta, Rishi Puri, Manan Shah, Blaž Stojanovič, Ramona Bendias, Alexandria Barghi, Vid Kocijan, Zecheng Zhang, Xinwei He, Jan Eric Lenssen, and Jure Leskovec. PyG 2.0: Scalable learning on real world graphs. In *Temporal Graph Learning Workshop @ KDD 2025*, 2025.

Richard P. Feynman. Cargo cult science. In Edward Hutchings, editor, *Surely You're Joking, Mr. Feynman!: Adventures of a Curious Character*, pages 338–346. W. W. Norton, New York, NY, USA, 1985.

Jorge Luiz Franco, Gabriel Duarte, Alexander V. Nikitin, Moacir A. Ponti, Diego Mesquita, and Amauri H. Souza. Differentiable lifting for topological neural networks. In *International Conference on Learning Representations*, 2026.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021.

Jakob Jonsson. *Simplicial Complexes of Graphs*, volume 1928 of *Lecture Notes in Mathematics*. Springer, Heidelberg, Germany, 2008.

Matthias Kreck. A guide to the classification of manifolds. In Jonathan M. Rosenberg, Sylvain Cappell, and Andrew Ranicki, editors, *Surveys on Surgery Theory, Volume 1*, pages 121–134. Princeton University Press, Princeton, NJ, USA, 2014.

Zachary C. Lipton and Jacob Steinhardt. Troubling trends in machine learning scholarship, 2018. URL https://arxiv.org/abs/1807.03341.

Frank H. Lutz. Enumeration and random realization of triangulated surfaces. In Alexander I. Bobenko, John M. Sullivan, Peter Schröder, and Günter M. Ziegler, editors, *Discrete Differential Geometry*, pages 235–253. Birkhäuser, Basel, Switzerland, 2008.

Frank H. Lutz. The manifold page, 2017. URL https://www3.math.tu-berlin.de/IfM/Nachrufe/Frank_Lutz/stellar/.

Jiří Matoušek. *Using the Borsuk–Ulam Theorem: Lectures on Topological Methods in Combinatorics and Geometry*. Universitext. Springer, Heidelberg, Germany, 2003.

Christopher Morris, Yaron Lipman, Haggai Maron, Bastian Rieck, Nils M. Kriege, Martin Grohe, Matthias Fey, and Karsten Borgwardt. Weisfeiler and Leman go machine learning: The story so far. *Journal of Machine Learning Research*, 24(333):1–59, 2023.

Leslie O'Bray, Bastian Rieck, and Karsten Borgwardt. Filtration curves for graph representation. In *Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1267–1275, New York, NY, USA, 2021. Association for Computing Machinery. doi: 10.1145/3447548.3467442.

Theodore Papamarkou, Tolga Birdal, Michael Bronstein, Gunnar Carlsson, Justin Curry, Yue Gao, Mustafa Hajij, Roland Kwitt, Pietro Liò, Paolo Di Lorenzo, Vasileios Maroulas, Nina Miolane, Farzana Nasrin, Karthikeyan Natesan Ramamurthy, Bastian Rieck, Simone Scardapane, Michael T. Schaub, Petar Veličković, Bei Wang, Yusu Wang, Guo-Wei Wei, and Ghada Zamzmi. Position: Topological deep learning is the new frontier for relational learning. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 39529–39555. PMLR, 2024.

Sara Steegen, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11 (5):702–712, 2016.

Timo Stoll, Chendi Qian, Ben Finkelshtein, Ali Parviz, Darius Weber, Fabrizio Frasca, Hadar Shavit, Antoine Siraudin, Arman Mielke, Marie Anastacio, Erik Müller, Maya Bechler-Speicher, Michael Bronstein, Mikhail Galkin, Holger Hoos, Mathias Niepert, Bryan Perozzi, Jan Tönshoff, and Christopher Morris. GraphBench: Next-generation graph learning benchmarking, 2026. URL https://arxiv.org/abs/2512.04475.

Richard S. Sutton. The bitter lesson, 2019. URL http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

Diaaeldin Taha, James Chapman, Marzieh Eidi, Karel Devriendt, and Guido Montufar. Demystifying topological message-passing with relational structures: A case study on oversquashing in simplicial message-passing. In *International Conference on Learning Representations*, 2025.

Lev Telyatnikov, Guillermo Bernardez, Marco Montagna, Mustafa Hajij, Martin Carrasco, Pavlo Vasylenko, Mathilde Papillon, Ghada Zamzmi, Michael T Schaub, Jonas Verhellen, Pavel Snopov, Bertran Miquel-Oliver, Manel Gil-Sorribes, Alexis Molina, VICTOR GUALLAR, Theodore Long, Julian Suk, Patryk Rygiel, Alexander V Nikitin, Giordan Escalona, Michael Banf, Dominik Filipiak, Liliya Imasheva, Max Schattauer, Alvaro L. Martinez, Halley Fritze, Marissa Masden, Valentina Sánchez, Manuel Lecha, Andrea Cavallo, Claudio Battiloro, Matthew Piekenbrock, Mauricio Tec, George Dasoulas, Nina Miolane, Simone Scardapane, and Theodore Papamarkou. TopoBench: A framework for benchmarking topological deep learning. *Journal of Data-Centric Machine Learning Research*, 2025.

Petar Veličković. Everything is connected: Graph neural networks. *Current Opinion in Structural Biology*, 79:102538, 2023.

Jeremy Wayland, Corinna Coupette, and Bastian Rieck. Mapping the multiverse of latent representations. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 52372–52402, 2024.