Herbert Edelsbrunner, David Letscher, and Afra J. Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4):511–533, 2002.

William H Guss and Ruslan Salakhutdinov. On characterizing the capacity of neural networks using algebraic topology. *arXiv preprint arXiv:1802.04443*, 2018.

Kazuyuki Hara, Daisuke Saitoh, and Hayaru Shouno. Analysis of dropout learning regarded as ensemble learning. In Alessandro E.P. Villa, Paolo Masulli, and Antonio Javier Pons Rivero (eds.), *Artificial Neural Networks and Machine Learning (ICANN)*, number 9887 in Lecture Notes in Computer Science, pp. 72–79, Cham, Switzerland, 2016. Springer.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 1633–1643, 2017.

Danijela Horak, Slobodan Maletić, and Milan Rajković. Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03034, 2009.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141, 2018.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 448–456. PMLR, 2015.

Valentin Khrulkov and Ivan Oseledets. Geometry score: A method for comparing generative adversarial networks. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2621–2629. PMLR, 2018.

Pek Y. Lum, Gurjeet Singh, Alan Lehman, Tigran Ishkanov, Mikael Vejdemo-Johansson, Muthu Alagappan, John Carlsson, and Gunnar Carlsson. Extracting insights from the shape of complex data using topology. *Scientific Reports*, 3:1–8, 2013.

Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2017.

James R. Munkres. *Elements of algebraic topology*. CRC Press, Boca Raton, FL, USA, 1996.

Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. On the expressive power of deep neural networks. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 2847–2854. PMLR, 2017.

Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Michaela Hardt, Peter J Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, et al. Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, 1(1):18, 2018.

Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew Lungren, and Andrew Y. Ng. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

Bastian Rieck and Heike Leitte. Exploring and comparing clusterings of multivariate data sets using persistent homology. *Computer Graphics Forum*, 35(3):81–90, 2016.

Bastian Rieck, Ulderico Fugacci, Jonas Lukasczyk, and Heike Leitte. Clique community persistence: A topological visual analysis approach for complex networks. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):822–831, 2018.

Andrew Michael Saxe, Yamini Bansal, Joel Dapello, Madhu Advani, Artemy Kolchinsky, Brendan Daniel Tracey, and David Daniel Cox. On the information bottleneck theory of deep learning. In *International Conference on Learning Representations (ICLR)*, 2018.

Ravid Shwartz-Ziv and Naftali Tishby. Opening the black box of deep neural networks via information. *arXiv preprint arXiv:1703.00810*, 2017.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015.

Ann Sizemore, Chad Giusti, and Danielle S. Bassett. Classification of weighted networks through mesoscale homological features. *Journal of Complex Networks*, 5(2):245–273, 2017.

Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. In *Workshop Track of the International Conference on Learning Representations (ICLR)*, 2015.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 3104–3112, 2014.

Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2015.

Michael Tsang, Dehua Cheng, and Yan Liu. Detecting statistical interactions from neural network weights. In *International Conference on Learning Representations (ICLR)*, 2018.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (eds.), *European Conference on Computer Vision (ECCV)*, volume 8689 of *Lecture Notes in Computer Science*, pp. 818–833, Cham, Switzerland, 2014. Springer.

Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. In *International Conference on Learning Representations (ICLR)*, 2017.

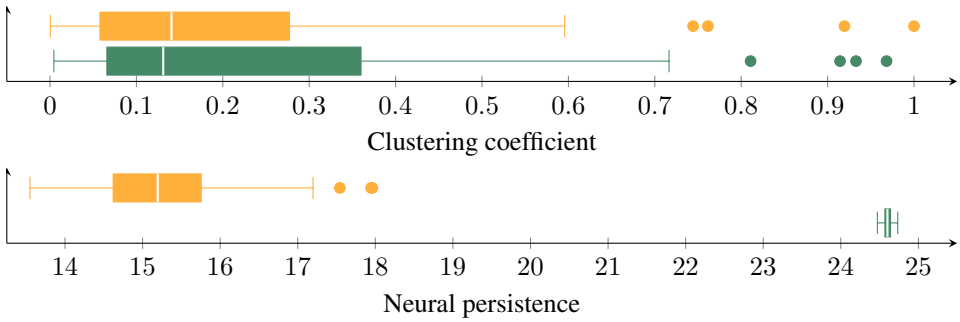

Figure A.1: Traditional graph measures (top), such as the clustering coefficient, fail to detect differences in the complexity of neural networks. Our novel *neural persistence* measure (bottom), by contrast, shows that trained networks with $\eta = 0.5$ (green), which have an accuracy of ≈ 0.91, obey a different distribution than networks trained with $\eta = 1 \times 10^{-0.5}$ (yellow), which have accuracies ranging from 0.38–0.65.

A Appendix

A.1 Comparison with graph-theoretical measures

Traditional complexity/structural measures from graph theory, such as the clustering coefficient, the average shortest path length, and global/local efficiency are already known to be insufficiently accurate to characterize different models of complex random networks Sizemore et al. (2017). Our experiments indicate that this holds true for (deep) neural networks, too. As a brief example, we trained a perceptron on the MNIST data set with batch stochastic gradient descent ($\eta = 0.5$), achieving a test accuracy of ≈ 0.91. Moreover, we intentionally 'sabotaged' the training by setting $\eta = 1 \times 10^{-5}$ such that SGD is unable to converge properly. This leads to networks with accuracies ranging from 0.38–0.65. A complexity measure should be capable of distinguishing both classes of networks. However, as Figure A.1 (top) shows, this is *not* the case for the clustering coefficient. Neural persistence (bottom), on the other hand, results in two regimes that can clearly be distinguished, with the trained networks having a significantly smaller variance.

A.2 Proof of Theorem 2

Proof. We may consider the filtration from Section 3.1 to be a subset selection problem with constraints, where we select n out of m weights. The neural persistence $\mathrm{NP}(G_k)$ of a layer thus only depends on the *selected* weights that appear as tuples of the form $(1, w_i)$ in $\mathcal{D}_k$. Letting $\widetilde{\mathbf{w}}$ denote the vector of selected weights arising from the persistence diagram calculation, we can rewrite neural persistence as $\mathrm{NP}(G_k) = \|\mathbb{1} - \widetilde{\mathbf{w}}\|_p$. Furthermore, $\widetilde{\mathbf{w}}$ satisfies $\|\mathbf{w}_{\min}\|_p \leq \|\widetilde{\mathbf{w}}\|_p \leq \|\mathbf{w}_{\max}\|_p$. Since all transformed weights are non-negative in our filtration, it follows that (note the reversal of the two terms)

$$\|\mathbb{1} - \mathbf{w}_{\max}\|_p \leq \mathrm{NP}(G_k) \leq \|\mathbb{1} - \mathbf{w}_{\min}\|_p \,, \tag{6}$$

and the claim follows. □

A.3 Additional visualizations and analyses for early stopping

Due to space constraints and the large number of configurations that we investigated for our early stopping experiments, this section contains additional plots that follow the same schematic: the top row shows the differences in accuracy and epoch for our measure when compared to the commonly-used validation loss. Each cell in the heatmap corresponds to a single configuration of b and g. In the heatmap of accuracy differences, blue represents parameter combinations for which we obtain a *higher* accuracy than validation loss for the same parameters; white indicates combinations for which we obtain the same accuracy, while red highlights combinations in which our accuracy decreases. Similarly, in the heatmap of epoch differences, green represents parameter combinations for which we stop *earlier* than validation loss for the same parameter. The scatterplots in Section 4.2 show an