When Learning on Graphs Breaks

Propagation, Compression, and Collapse in Graph Neural Networks

Marco Lavorini, Department of Computer Science, University of Pisa

March 13, 2026

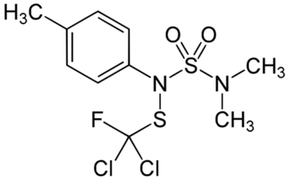
Talk roadmap

- **Learning on graphs:** a minimal introduction
- **Why depth changes behavior**
- **Common pathologies in GNNs**: smoothing, squashing, etc.
- **Open questions**

Goal

Understand where and why learning on graphs can break, and identify a concrete research question worth discussing.

Graphs are **everywhere!**



Molecules

Social Networks

Transportation Networks

And many others...

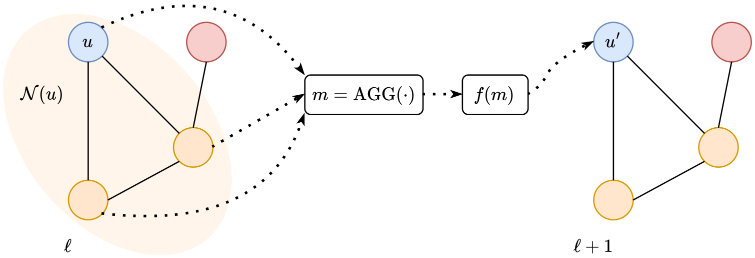
Slide from: *Fabrizio De Castelli, Mauriana Pesaresi Seminar Series* 06/03/2026

Learning on Graphs

Graphs model systems of interacting entities:

- nodes represent entities,
- edges represent relationships.

To learn from such data, we use **Graph Neural Networks**, which rely on nodes iteratively aggregating information from their neighbors.



Message Passing: One Layer

$$\mathbf{h}_i^{(\ell+1)} = \phi^{(\ell)}\left(\mathbf{h}_i^{(\ell)},\ \mathrm{AGG}\{\mathbf{W}^{(\ell)}\mathbf{h}_j^{(\ell)} : j \in \mathcal{N}(i)\}\right)$$

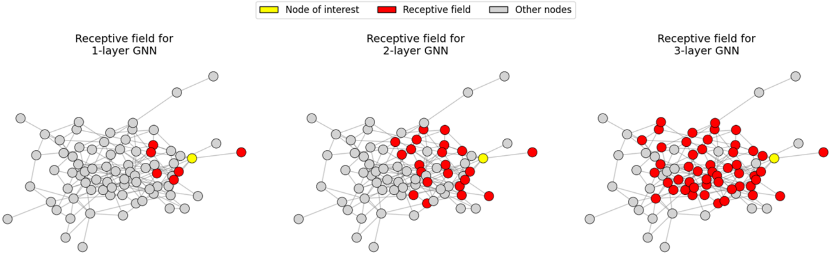
- $\mathbf{W}^{(\ell)}$: **shared weight matrix** (same for all nodes)
- $\mathrm{AGG}(\cdot)$: permutation-invariant aggregation (sum, mean, max...)
- $\phi^{(\ell)}(\cdot)$: update function

Structural objective

Learn node (or graph) representations that are **permutation-equivariant** at the node level and **invariant** at the graph level, so that isomorphic graphs receive identical representations.

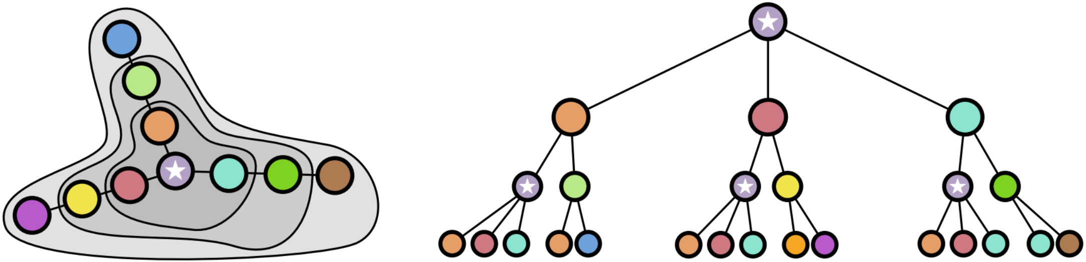
What Happens When We Go Deeper?

- Each layer expands the node's **receptive field**.
- After k layers, a node aggregates information from nodes up to k hops away.



Depth enables long-range interaction. But repeating local aggregation has side effects.

What Happens After Several Layers?



- After k layers, information flows through many paths.
- The same node may appear multiple times in the computation tree.
- All messages are aggregated into a single fixed-size vector.

Image from: Arnaiz-Rodriguez & Errica (ICLR 2026), “Oversmoothing, 'Oversquashing' ...”.

A Taxonomy of Failure Modes in Graph Learning

As depth increases, different distortions can appear:

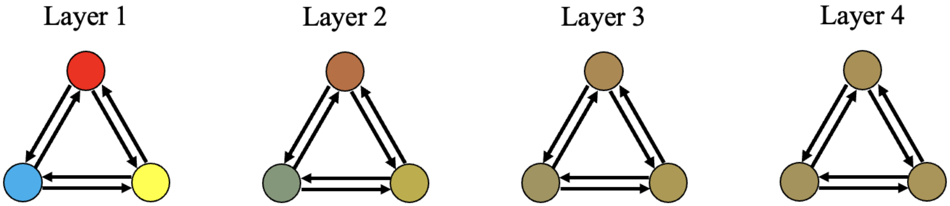
- **Oversmoothing**
  Node representations become too similar, reducing discriminative power [8, 3, 12].
- **Oversquashing**
  Too much information is compressed into fixed-size representations [1, 9, 2].
- **Underreaching**
  Distant but relevant information is not effectively exploited.
- **Heterophily**
  Neighboring nodes differ significantly, challenging simple aggregation schemes [13].

Do these failures share a common underlying mechanism?

Oversmoothing: intuition

Repeated neighborhood mixing tends to:

- reduce local differences,
- make node representations increasingly similar,
- attenuate high-frequency signals on the graph.



As depth increases, representations can collapse toward similarity.

Oversmoothing: one clean (but imperfect) formalization

Let $H^{(\ell)} \in \mathbb{R}^{n \times d}$ be node embeddings at layer ℓ. Oversmoothing is often phrased as:

$$\text{“node similarity increases with } \ell\text{”} \quad \Longleftrightarrow \quad \varepsilon(H^{(\ell)}) \to c.$$

Different families of similarity measures $\varepsilon(\cdot)$ include:

- **Smoothness-based metrics:** Dirichlet Energy (DE) and its normalized version, the Rayleigh Quotient (RQ).
- **Distance-based metrics:** Mean Absolute Deviation (MAD) [3].
- **Subspace criteria:** Rank collapse / convergence to low-dimensional subspaces.

For the rest of this talk, we focus on **Dirichlet Energy** as a representative notion:

$$\mathrm{DE}\big(H^{(\ell)}\big) \to 0.$$

Dirichlet Energy as a Smoothness Measure

A common proxy for smoothness over a graph is the Dirichlet Energy:

$$\mathrm{DE}(H) = \frac{1}{2} \sum_{i \in V} \sum_{j \in \mathcal{N}(i)} \|h_i - h_j\|_2^2.$$

Interpretation:

- High DE: neighboring nodes remain distinguishable.
- Low DE: neighboring nodes become similar.

Oversmoothing narrative:

$$\mathrm{DE}(H^{(\ell)}) \downarrow \quad \text{as } \ell \uparrow .$$

But is low DE always a problem?

Mitigating Oversmoothing

There have been many proposed solutions:

- **Architectural:** residual/skip connections, JK-nets [11].
- **Regularization:** PairNorm [12], DE constraints.
- **Dropout on graph:** DropEdge [7], DropNode [6], DropAGG [4], etc.
- **Alternative propagation:** personalized PageRank [5].

Even if we fix oversmoothing, the tasks can still fail, what is happening?

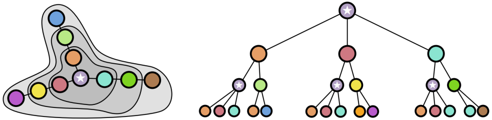
Oversquashing: the core intuition

A node is represented by a vector of fixed dimension d.

However, as depth increases:

- the number of possible information paths grows,
- the number of aggregated signals increases,
- the representation size remains constant.

Compression effect:
many signals must be encoded into one fixed-size vector.



(Recall the computation tree!!!)

Oversquashing: A Terminology Problem

The term *oversquashing* has been used to describe two different phenomena.

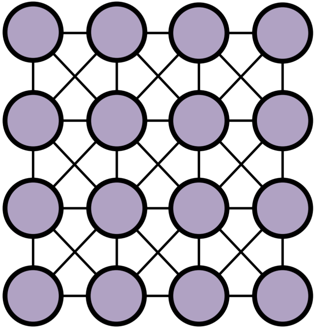
- **Computational bottleneck** [1]
  Exponentially growing information is compressed into a fixed-size vector.
- **Topological bottleneck** [9]
  Structural constraints (e.g., narrow cuts, negative curvature) restrict information flow in the graph.

These problems do not always coexist!

Computational vs Topological Bottlenecks

Grid Graph (No Structural Bottleneck)

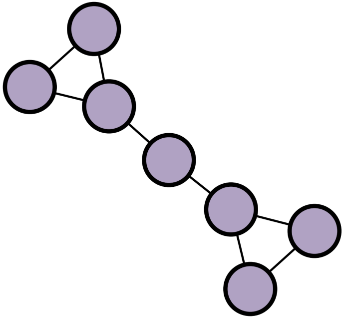
- Information spreads gradually.
- Long paths require many layers.



Computational bottleneck without a topological one.

Graph with a Narrow Cut

- Few edges connect large regions.
- Structural constraints limit flow.



Topological bottleneck.

Oversquashing: what do people actually measure?

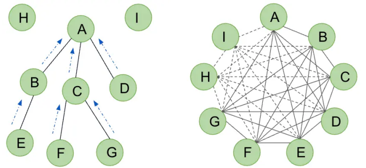
Common approaches to formalize oversquashing include:

- **Influence / sensitivity measures**
  How much does the final representation of a node depend on the initial features of a distant node? [10, 2]
- **Structural bottleneck measures**
  Does the graph contain narrow cuts or negatively curved regions that restrict information flow? [9]
- **Topology-aware distances**
  Distances such as effective resistance estimate how difficult it is for information to travel between two nodes. [2]

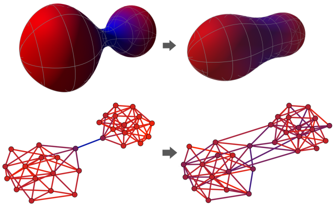
Key issue

Fixing one notion of oversquashing does not necessarily fix the others.

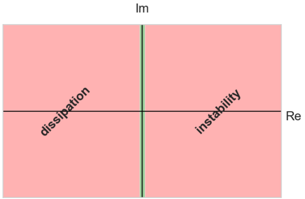
Mitigating Oversquashing



Graph Transformers
Global attention replaces strictly local aggregation.



Rewiring
Modify graph structure to reduce bottlenecks.

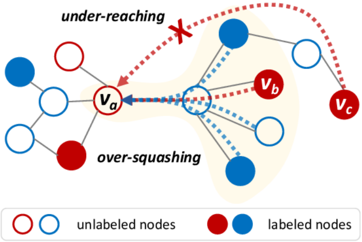


Dynamical Models
Control propagation through stability principles.

Open question

Are these fundamentally new mechanisms, or different ways to control information flow?

Underreaching and Oversquashing

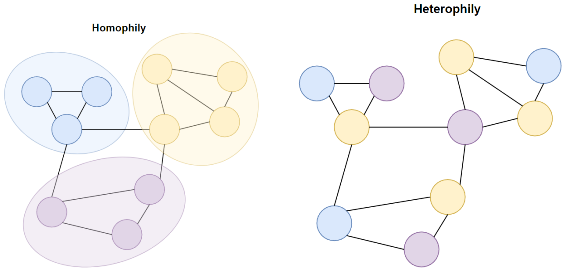


- **Oversquashing:** Information from many nodes is compressed through structural or computational bottlenecks.
- **Underreaching:** Relevant distant information exists, but the model fails to effectively propagate or use it.

Conceptual distinction

Oversquashing limits information capacity.
Underreaching reflects optimization or propagation inefficiency.

Homophily and Heterophily: Definition & Measurement



"Are neighbors informative or misleading?"

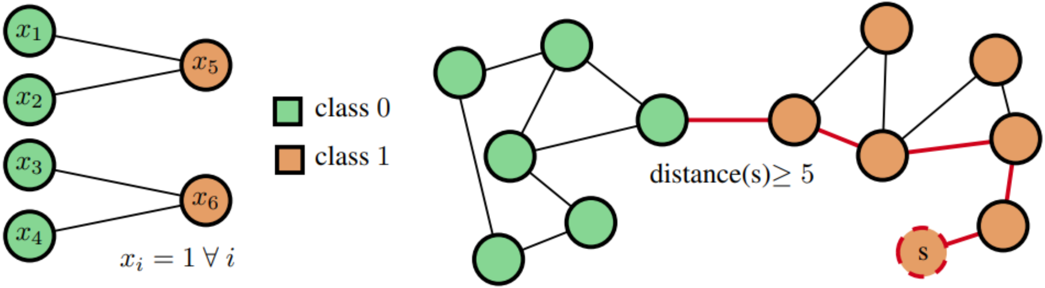
Generic definition (w.r.t. a variable s_i):

$$\mathcal{H}(s) = \frac{1}{|E|} \sum_{(i,j)\in E} \mathbf{1}[s_i = s_j]$$

- if $s_i = y_i \Rightarrow$ **class homophily**
- if s_i bins features $\Rightarrow$ **feature homophily**

Key warning: the same graph can be class-heterophilic but feature-homophilic (and vice versa).

Homophily is not the full story



Takeaways:

- **Homophily/heterophily describe local neighborhoods.** They do not tell us whether the task signal is local or long-range.
- **Depth is a trade-off:** it increases reach, but repeated propagation can wash out distinctions and mix conflicting signals.

Image from: Arnaiz-Rodriguez & Errica (ICLR 2026), "Oversmoothing, 'Oversquashing' ...".

Discussion: open directions

Open problem

Deep message passing can fail for different reasons (collapse, compression, limited reach). We still lack **standard definitions** and **reliable tests** to tell these mechanisms apart.

1) Definitions and measurements

- Can we agree on operational definitions of the main pathologies (oversmoothing, oversquashing, underreaching)?
- What should a metric measure so that it is comparable across models and datasets?

2) What do "solutions" actually solve?

- When do interventions (rewiring, attention, normalization, residuals) address the underlying limitation?

3) Borrowing tools from other fields

- Are there diagnostic principles or design mechanisms from related domains that could better model information flow in GNNs?

Thank you

Thanks!

References I

[1] Uri Alon and Eran Yahav. "On the Bottleneck of Graph Neural Networks and its Practical Implications". In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021.* OpenReview.net, 2021.

[2] Mitchell Black, Zhengchao Wan, Amir Nayyeri, and Yusu Wang. "Understanding Oversquashing in GNNs through the Lens of Effective Resistance". In: *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*. Ed. by Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett. Proceedings of Machine Learning Research. PMLR, 2023.

References II

[3] Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. "Measuring and Relieving the Over-Smoothing Problem for Graph Neural Networks from the Topological View". In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020.

[4] Bo Jiang, Yong Chen, Beibei Wang, Haiyun Xu, and Bin Luo. "DropAGG: Robust Graph Neural Networks via Drop Aggregation". In: *Neural Netw.* C (2023). ISSN: 0893-6080. DOI: `10.1016/j.neunet.2023.03.022`.

[5] Johannes Klicpera, Aleksandar Bojchevski, and Stephan Günnemann. "Predict then Propagate: Graph Neural Networks meet Personalized PageRank". In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[6] Pál András Papp, Karolis Martinkus, Lukas Faber, and Roger Wattenhofer. "DropGNN: Random Dropouts Increase the Expressiveness of Graph Neural Networks". In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. Ed. by Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan. 2021.

[7] Yu Rong, Wenbing Huang, Tingyang Xu, and Junzhou Huang. "DropEdge: Towards Deep Graph Convolutional Networks on Node Classification". In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[8] T. Konstantin Rusch, Michael M. Bronstein, and Siddhartha Mishra. *A Survey on Oversmoothing in Graph Neural Networks*. 2023.

References IV

[9] Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. "Understanding over-squashing and bottlenecks on graphs via curvature". In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

[10] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. "How Powerful are Graph Neural Networks?" In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

References V

[11] Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. "Representation Learning on Graphs with Jumping Knowledge Networks". In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Ed. by Jennifer G. Dy and Andreas Krause. Proceedings of Machine Learning Research. PMLR, 2018.

[12] Lingxiao Zhao and Leman Akoglu. "PairNorm: Tackling Oversmoothing in GNNs". In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[13] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. "Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs". In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. Ed. by Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin. 2020.