Quantifying the Effects of Edge Uncertainty on Node Embeddings

Narayan G. Sabhahit and **Moritz Laber** *Northeastern University*
Franziska Heeg and **Jan von Pichowski** and **Ingo Scholtes** *University of Würzburg*
Dmitri Krioukov and **Tina Eliassi-Rad** *Northeastern University*
TINA@ELIASSI.ORG

Abstract

Uncertainty quantification is an essential step for developing reliable and trustworthy machine learning models. While a plethora of methods have been developed for quantifying predictive uncertainty of graph neural networks, theoretical insight into the impact of uncertain graph structure on node embeddings remains limited. Here, we leverage tools from the study of exponential random graphs to rigorously derive node-wise preactivation distributions for graph convolutional networks, highlighting sparsity as the key parameter governing the form of this distribution. On random graphs with diverging average degrees, message passing is asymptotically deterministic; however, on random graphs with constant average degrees, node preactivation distributions retain finite variance.

Keywords: graph neural networks, uncertainty quantification, random graph models

1. Introduction

Graph machine learning is used in various high-stakes domains such as drug design (Stokes et al., 2020), weather forecasting (Lam et al., 2023), fraud detection (Dou et al., 2020), and autonomous driving (Gao et al., 2020). In these critical applications, proper uncertainty quantification is as important as the prediction itself (Sullivan, 2015; Soize, 2017). Following a larger trend within deep learning (Abdar et al., 2021; Hüllermeier and Waegeman, 2021), several methods for uncertainty quantification have been developed for graph neural networks (GNNs) (Wang et al., 2024; Ju et al., 2025). This includes conformal prediction (Zargarbashi et al., 2023; Clarkson, 2023), deep ensembles (Mallick et al., 2024; Vieira et al., 2026), and Bayesian approaches (Hasanzadeh et al., 2020; Zhang et al., 2019). When dealing with graph-structured data, sources of uncertainty include node and edge features (e.g., Chernikova et al. (2026)), as well as the graph structure (e.g., Shafi et al. (2025)). A large body of work has highlighted edge uncertainty as a key challenge in the analysis of relational data (Peel et al., 2022; Cimini et al., 2019). Thus, we focus on edge uncertainty. Prior work in graph machine learning has developed algorithms, e.g., GNNs with learnable latent graphs (Manenti et al., 2025; Franceschi et al., 2019) and edge dropout (Rong et al., 2019; Hasanzadeh et al., 2020), to address practical challenges associated with uncertain graphs. However, there are only a few theoretical perspectives on the influence of edge uncertainty on GNN predictions (Duranthon and Zdeborová, 2025).

Here, we approach the problem of quantifying uncertainty in node embeddings due to uncertain edges using tools from network science and statistical physics. We focus on graph convolutional networks (GCNs) (Kipf and Welling, 2017) and graphs sampled from

© 2026 N.G. Sabhahit, M. Laber, F. Heeg, J. von Pichowski, I. Scholtes, D. Krioukov & T. Eliassi-Rad.

exponential random graph models (Park and Newman, 2004; Squartini and Garlaschelli, 2017), specifically, the Erdős-Rényi (ER) $G(n, p_n)$ model (Erdős and Rényi, 1959, 2011; Gilbert, 1959) and soft random geometric graphs $\mathrm{SRGG}(n, \rho_n, \beta)$ (Penrose, 2003).

Our **contributions** are as follows: **(1)** We establish that sparsity, quantified by the average degree $\bar{k}$, is a key property governing the behavior of node embedding uncertainty in the limit of large graphs. **(2)** We show that whenever the average degree diverges, message passing updates on large graphs become asymptotically deterministic and all embeddings collapse to the same value. We emphasize the difference to oversmoothing which occurs as the number of layers, in contrast to the average degree, diverges (Rusch et al., 2023; Li et al., 2018). **(3)** In contrast, when the average degree is constant, the embedding distribution no longer concentrates. We rigorously derive the exact asymptotic form of the preactivation distribution for a single-layer GCN, and highlight the need for more sophisticated analytical techniques for a multilayer GCN in the constant-degree case.

2. Setup

We aim to understand how uncertainty in the edges of a graph G propagates through a GCN. In this work, we consider the GCN architecture (Kipf and Welling, 2017):

$$x_i^{(\ell+1)} = \sigma\left(h_i^{(\ell+1)}\right); \quad h_i^{(\ell+1)} = w^{(\ell)} \sum_{j=1}^{n} \frac{G_{ij}}{\sqrt{k_i k_j}} x_j^{(\ell)}, \tag{1}$$

where $x_i^{(0)} \in \mathbb{R}$ is the input feature of node i, $x_i^{(\ell)} \in \mathbb{R}$ its latent representation at layer $1 \leq \ell \leq \ell_{\max}$, and $h_i^{(\ell+1)}$ its corresponding preactivation. We denote the adjacency matrix elements of the graph G as G_{ij} and the degree of node i as $k_i = \sum_{j=1}^{n} G_{ij}$. The trainable weight is $w^{(\ell)} \in \mathbb{R}$ and $\sigma(\cdot)$ is a non-linear activation function.

To model edge uncertainty, we rely on (conditionally) edge-independent exponential random graphs (ERGMs) (Park and Newman, 2004; Squartini and Garlaschelli, 2017). That is, we assume the probability $\mathbb{P}(G)$ of a graph G is $\mathbb{P}(G) = \prod_{i<j} p_{ij}^{G_{ij}} (1 - p_{ij})^{1-G_{ij}}$ where p_{ij} is the connection probability between nodes i and j. We have $G_{ij} = 1$ with probability p_{ij} and $G_{ij} = 0$ with probability $1 - p_{ij}$. In addition, we consider self-loops, $G_{ii} = 1$ with probability one. We focus on $G(n, p)$ (Section 3) and $\mathrm{SRGG}(n, \rho_n, \beta)$ (Appendix E).

We assume: **(A1)** Weights $\boldsymbol{w} = \{w^{(\ell)}\}_{1 \leq \ell \leq \ell_{\max}}$ are fixed and $\mathcal{O}(1)$. This corresponds to training the GCN on a fixed graph and deploying it on graphs sampled from $\mathbb{P}(G)$ (assumed to encode the true data-generating process). **(A2)** Input node features $\boldsymbol{x}^{(0)} = \{x_i^{(0)}\}_{1 \leq i \leq n}$ are assumed to be deterministic and $\mathcal{O}(1)$. This means that the graph structure is the only source of uncertainty. **(A3)** Node representations and weights are scalar, $x_i^{(\ell)}, w^{(\ell)} \in \mathbb{R}$ (Equation (1)). In this setting, we seek to characterize the asymptotic marginal distribution of node embeddings $\mathbb{P}(x_i^{(\ell)} | \boldsymbol{x}^{(0)})$ at layer ℓ, thereby quantifying how uncertainty in the graph structure translates into uncertainty in the node embeddings. Appendix A summarizes our notation and Appendix B provides a schematic depiction of our setup.

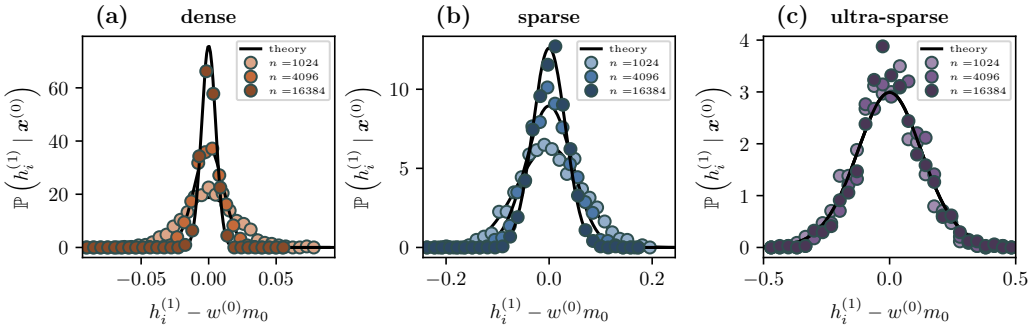

Figure 1: Distribution $\mathbb{P}(h_i^{(1)} \mid \boldsymbol{x}^{(0)})$ of the node preactivation $h_i^{(1)}$ of an arbitrary node i at the first layer for **(a)** dense ($p_n = 0.3$, orange), **(b)** sparse ($p_n = 1.5/\sqrt{n}$,blue), and **(c)** ultra-sparse ($p_n = 10/n$, purple) graphs match the theoretical predictions (black,(a), (b), Equation (2), (c), Equation (4)). The distribution concentrates in $G(n, p_n)$ as $n \to \infty$ for dense and sparse but not for ultra-sparse graphs.

3. Results

We present analytical results for the preactivation distribution $\mathbb{P}(h_i^{(1)}|\boldsymbol{x}^{(0)})$ at layer $\ell = 1$ induced by a GCN on $G(n, p_n)$, where every pair of distinct nodes is connected independently with probability p_n. We call a random graph model *dense* if the average degree $\bar{k} = 1 + (n-1)p = \mathcal{O}(n)$, *sparse* if $\bar{k} = o(n)$ and *ultra-sparse* if $\bar{k} = \mathcal{O}(1)$. First, we present results on the dense and sparse case, where the average degree is diverging. We then turn to the ultra-sparse case. We detail the experimental setup in Appendix C.

Dense & Sparse Graphs: Let $m_\ell = (1/n)\sum_{i=1}^n x_i^{(\ell)}$ and $q_\ell = (1/n)\sum_{i=1}^n (x_i^{(\ell)})^2$. In Appendix F, we show that the distribution of the preactivations $h_i^{(1)}$ in layer one is:

$$\mathbb{P}\left(h_i^{(1)} \mid \boldsymbol{x}^{(0)}\right) \approx \mathcal{N}\left(w^{(0)}m_0, \frac{(w^{(0)})^2(1-p)q_0}{\bar{k}}\right). \tag{2}$$

In the limit $n \to \infty$, the distribution collapses to a delta function; as the variance scales as $\mathcal{O}(1/\bar{k})$ and $\bar{k} \to \infty$. Hence, the message passing updates become asymptotically deterministic, $h_i^{(1)} \xrightarrow{(P)} w^{(0)}m_0$, and, $x_i^{(1)} = \sigma(w^{(0)}m_0)$. All the node representations collapse to the same point, despite the randomness in the edges. Since we know m_0 given initial features $\boldsymbol{x}^{(0)}$, we can compute embeddings for any $\ell \geq 1$ as

$$m_\ell = \sigma(w^{(\ell-1)}m_{\ell-1}) \quad \text{and} \quad x_i^{(\ell)} = m_\ell. \tag{3}$$

Figure 1(a,b) compares Equation (2) for an arbitrary node in the case of dense and sparse $G(n, p_n)$ with numerical simulations showing agreement and decreasing variance as the graph size n increases. In Appendix D, we show that this holds across nodes and verify that the power law exponent matches the theoretical prediction (Equation (2)). Figure 2 shows that the node embeddings deterministically converge to the theoretical predictions in

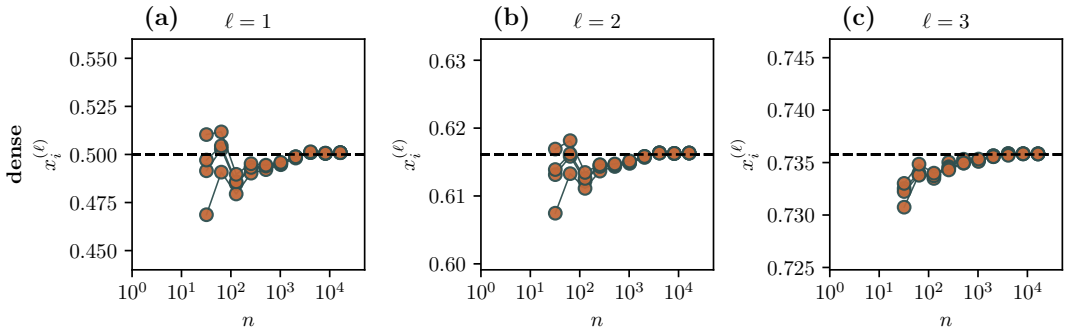

Figure 2: As $n \to \infty$ in dense $G(n, p_n)$ ($p_n = 0.3$), the node embeddings $x_i^{(\ell)}$ (shown for four arbitrary nodes) at different layers (1, left, 2, center, 3 right) converge to their value under deterministic updates (black dashed line, Equation (3)).

Equation (3) in dense $G(n, p_n)$ as the number of nodes n increases. We show the analogous result for sparse graphs in Appendix D.

Ultra-sparse Graphs: In Appendix G, we derive the moment-generating function (MGF) $\Phi(t)$ of $\mathbb{P}(h_i^{(1)}|\boldsymbol{x}^{(0)})$ in a single-layer GCN on ultra-sparse $G(n, p_n)$,

$$\Phi(t) = \exp\left[\frac{tw^{(0)}x_i^{(0)}}{\bar{k}} - \bar{k} + \bar{k} \cdot M_0\left(\frac{tw^{(0)}}{\bar{k}}\right)\right], \tag{4}$$

where $M_0(t)$ is the MGF of the distribution of the initial features. For an arbitrary node, Figure 1(c) compares the distribution of preactivations at layer 1 (per Equation (4)) with numerical simulation results. In the ultra-sparse regime finite variance in preactivations persists as the graph size increases.

4. Discussion

We develop a rigorous theory for node embedding uncertainty in GCNs due to edge uncertainty in ERGMs with diverging average degree. We also take first steps towards the case of constant average degree. Our findings have implications for advancing graph machine learning by (1) highlighting the usefulness of random graphs in this domain (e.g., Aliakbarisani et al. (2026); Palowitch et al. (2022)); and (2) showcasing how statistical physics (Zdeborová and Krzakala, 2016) and network science (Blöcker et al., 2026) provide insight into GNNs.

While we focus on the ER and SRGG models, our analysis is based on (conditionally) edge-independent ERGMs, and thus easily extendable to other random graphs, e.g., stochastic block models (Holland et al., 1983; Peixoto, 2019), configuration models (Chatterjee et al., 2011; Chung and Lu, 2002), and certain latent space models (Krioukov et al., 2010; Boguñá et al., 2021). The focus on low-dimensional features and embeddings is a limitation of the current work, but the extension to the more realistic high-dimensional case poses only a technical but no conceptual challenge.

The most promising and challenging direction for future work are ultra-sparse random graphs, due to their importance for modeling of complex systems (Cimini et al., 2019) and theoretical challenges posed by quenched randomness (Metz, 2025; Metz and Castillo, 2026).

Acknowledgments

M.L. and T.E.R. were supported by the Inaugural Joseph E. Aoun Endowment.

References

Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297, 12 2021. ISSN 1566-2535. doi: 10.1016/j.inffus.2021.05.008.

Roya Aliakbarisani, Robert Jankowski, M. Ángeles Serrano, and Marián Boguñá. HypBench: Hyperbolic Benchmark for Graph Neural Network Performance. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–14, 2026. ISSN 2162-2388. doi: 10.1109/TNNLS.2026.3697597.

Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 3 edition, May 1995. ISBN 978-0-471-00710-4.

Christopher Blöcker, Martin Rosvall, Ingo Scholtes, and Jevin D. West. Deep Graph Learning will stall without Network Science, February 2026. URL https://doi.org/10.48550/arXiv.2502.01177.

Marián Boguñá, Ivan Bonamassa, Manlio De Domenico, Shlomo Havlin, Dmitri Krioukov, and M. Ángeles Serrano. Network Geometry. *Nature Reviews Physics*, 3(2):114–135, February 2021. ISSN 2522-5820. doi: 10.1038/s42254-020-00264-4.

Sourav Chatterjee, Persi Diaconis, and Allan Sly. Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4):1400–1435, August 2011. ISSN 1050-5164, 2168-8737. doi: 10.1214/10-AAP728.

Alesia Chernikova, Moritz Laber, Narayan G. Sabhahit, and Tina Eliassi-Rad. Robustness and Generalization in Uncertainty-Aware Message Passing Neural Networks. In *The 29th International Conference on Artificial Intelligence and Statistics*, February 2026. URL https://openreview.net/forum?id=BcqtGTw9OZ.

Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, December 2002. doi: 10.1073/pnas.252631999.

Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and Guido Caldarelli. The statistical physics of real-world networks. *Nature Reviews Physics*, 1(1):58–71, January 2019. ISSN 2522-5820. doi: 10.1038/s42254-018-0002-6.

Jase Clarkson. Distribution Free Prediction Sets for Node Classification. In *Proceedings of the 40th International Conference on Machine Learning*, pages 6268–6278. PMLR, July 2023. URL https://proceedings.mlr.press/v202/clarkson23a.html.

Yingtong Dou, Zhiwei Liu, Li Sun, Yutong Deng, Hao Peng, and Philip S. Yu. Enhancing Graph Neural Network-based Fraud Detectors against Camouflaged Fraudsters. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 315–324, New York, NY, USA, October 2020. Association for Computing Machinery. ISBN 978-1-4503-6859-9. doi: 10.1145/3340531.3411903.

O. Duranthon and L. Zdeborová. Statistical Physics Analysis of Graph Neural Networks: Approaching Optimality in the Contextual Stochastic Block Model. *Physical Review X*, 15(4):041026, 11 2025. doi: 10.1103/lfxj-hbsk.

P. Erdős and A. Rényi. On the evolution of random graphs. In *The Structure and Dynamics of Networks*, pages 38–82. Princeton University Press, October 2011. ISBN 978-1-4008-4135-6. URL https://www.degruyterbrill.com/document/doi/10.1515/9781400841356.38/html.

Paul Erdős and Alfred Rényi. On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

Luca Franceschi, Mathias Niepert, Massimiliano Pontil, and Xiao He. Learning Discrete Structures for Graph Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning*, pages 1972–1982. PMLR, May 2019. URL https://proceedings.mlr.press/v97/franceschi19a.html.

Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir Anguelov, Congcong Li, and Cordelia Schmid. VectorNet: Encoding HD Maps and Agent Dynamics From Vectorized Representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11525–11533, 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Gao_VectorNet_Encoding_HD_Maps_and_Agent_Dynamics_From_Vectorized_Representation_CVPR_2020_paper.html.

E. N. Gilbert. Random Graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959. ISSN 0003-4851. URL https://www.jstor.org/stable/2237458.

Arman Hasanzadeh, Ehsan Hajiramezanali, Shahin Boluki, Mingyuan Zhou, Nick Duffield, Krishna Narayanan, and Xiaoning Qian. Bayesian Graph Neural Networks with Adaptive Connection Sampling. In *Proceedings of the 37th International Conference on Machine Learning*, pages 4094–4104. PMLR, November 2020. URL https://proceedings.mlr.press/v119/hasanzadeh20a.html.

Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, June 1983. ISSN 0378-8733. doi: 10.1016/0378-8733(83)90021-7.

Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3):457–506, 3 2021. ISSN 1573-0565. doi: 10.1007/s10994-021-05946-3.

Wei Ju, Siyu Yi, Yifan Wang, Zhiping Xiao, Zhengyang Mao, Hourun Li, Yiyang Gu, Yifang Qin, Nan Yin, Senzhang Wang, Xinwang Liu, Philip S. Yu, and Ming Zhang. A Survey of Graph Neural Networks in Real World: Imbalance, Noise, Privacy and OOD Challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–20, 2025. ISSN 1939-3539. doi: 10.1109/TPAMI.2025.3630673.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*, February 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, September 2010. doi: 10.1103/PhysRevE.82.036106.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, December 2023. doi: 10.1126/science.adi2336.

Qimai Li, Zhichao Han, and Xiao-ming Wu. Deeper Insights Into Graph Convolutional Networks for Semi-Supervised Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), April 2018. ISSN 2374-3468. doi: 10.1609/aaai.v32i1.11604.

Tanwi Mallick, Jane Macfarlane, and Prasanna Balaprakash. Uncertainty Quantification for Traffic Forecasting Using Deep-Ensemble-Based Spatiotemporal Graph Neural Networks. *IEEE Transactions on Intelligent Transportation Systems*, 25(8):9141–9152, August 2024. ISSN 1558-0016. doi: 10.1109/TITS.2024.3381099.

Alessandro Manenti, Daniele Zambon, and Cesare Alippi. Learning Latent Graph Structures and their Uncertainty. In *Forty-Second International Conference on Machine Learning*, June 2025. URL https://openreview.net/forum?id=TMRh3ScSCb.

Fernando L. Metz. Dynamical Mean-Field Theory of Complex Systems on Sparse Directed Networks. *Physical Review Letters*, 134(3):037401, January 2025. doi: 10.1103/PhysRevLett.134.037401.

Fernando L. Metz and Isaac Pérez Castillo. Dynamical cavity method for continuous-time complex systems on sparse random graphs, June 2026.

John Palowitch, Anton Tsitsulin, Brandon Mayer, and Bryan Perozzi. GraphWorld: Fake Graphs Bring Real Insights for GNNs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, pages 3691–3701, New York, NY, USA, August 2022. Association for Computing Machinery. doi: 10.1145/3534678.3539203.

Juyong Park and M. E. J. Newman. Statistical mechanics of networks. *Physical Review E*, 70(6):066117, December 2004. doi: 10.1103/PhysRevE.70.066117.

Leto Peel, Tiago P. Peixoto, and Manlio De Domenico. Statistical inference links data and theory in network science. *Nature Communications*, 13(1):6794, November 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-34267-9.

Tiago P. Peixoto. Bayesian Stochastic Blockmodeling. In *Advances in Network Clustering and Blockmodeling*, chapter 11, pages 289–332. John Wiley & Sons, Ltd, 2019. ISBN 978-1-119-48329-8. doi: 10.1002/9781119483298.ch11.

Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, May 2003. ISBN 978-0-19-170785-8. doi: 10.1093/acprof:oso/9780198506263.001.0001.

Yu Rong, Wenbing Huang, Tingyang Xu, and Junzhou Huang. DropEdge: Towards Deep Graph Convolutional Networks on Node Classification. In *International Conference on Learning Representations*, September 2019. URL https://openreview.net/forum?id=Hkx1qkrKPr.

T. Konstantin Rusch, Michael M. Bronstein, and Siddhartha Mishra. A Survey on Oversmoothing in Graph Neural Networks, March 2023. URL https://doi.org/10.48550/arXiv.2303.10993.

Zohair Shafi, Germans Savcisens, and Tina Eliassi-Rad. REGE: A method for incorporating uncertainty in graph embeddings. In *Proceedings of the 2025 SIAM International Conference on Data Mining (SDM)*, pages 376–385. Society for Industrial and Applied Mathematics, 2025. doi: 10.1137/1.9781611978520.40.

Christian Soize. *Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering*, volume 47 of *Interdisciplinary Applied Mathematics*. Springer International Publishing, Cham, 2017. ISBN 978-3-319-54338-3 978-3-319-54339-0. doi: 10.1007/978-3-319-54339-0.

Tiziano Squartini and Diego Garlaschelli. *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics*. SpringerBriefs in Complexity. Springer International Publishing, Cham, 2017. ISBN 978-3-319-69436-8 978-3-319-69438-2. doi: 10.1007/978-3-319-69438-2.

Jonathan M. Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M. Donghia, Craig R. MacNair, Shawn French, Lindsey A. Carfrae, Zohar Bloom-Ackermann, Victoria M. Tran, Anush Chiappino-Pepe, Ahmed H. Badran, Ian W. Andrews, Emma J. Chory, George M. Church, Eric D. Brown, Tommi S. Jaakkola, Regina Barzilay, and James J. Collins. A Deep Learning Approach to Antibiotic Discovery. *Cell*, 180(4):688–702.e13, February 2020. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2020.01.021.

T.J. Sullivan. *Introduction to Uncertainty Quantification*, volume 63 of *Texts in Applied Mathematics*. Springer International Publishing, Cham, 2015. ISBN 978-3-319-23394-9 978-3-319-23395-6. doi: 10.1007/978-3-319-23395-6.

Pedro C. Vieira, Pedro Ribeiro, and Viacheslav Borovitskiy. Do Deep Ensembles Actually Capture Uncertainty in Graph Neural Networks?, May 2026. URL https://doi.org/10.48550/arXiv.2605.22593.

Fangxin Wang, Yuqing Liu, Kay Liu, Yibo Wang, Sourav Medya, and Philip S. Yu. Uncertainty in Graph Neural Networks: A Survey. *Transactions on Machine Learning Research*, 3 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=0e1Kn76HM1.

Soroush H. Zargarbashi, Simone Antonelli, and Aleksandar Bojchevski. Conformal Prediction Sets for Graph Neural Networks. In *Proceedings of the 40th International Conference on Machine Learning*, pages 12292–12318. PMLR, July 2023. URL https://proceedings.mlr.press/v202/h-zargarbashi23a.html.

Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, September 2016. ISSN 0001-8732. doi: 10.1080/00018732.2016.1211393.

Yingxue Zhang, Soumyasundar Pal, Mark Coates, and Deniz Ustebay. Bayesian Graph Convolutional Neural Networks for Semi-Supervised Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):5829–5836, July 2019. ISSN 2374-3468. doi: 10.1609/aaai.v33i01.33015829.

Table 1: Summary of the notations used in this paper.

Symbol	Explanation
G, G_{ij}	undirected graph and entries of its adjacency matrix.
$\mathbb{P}(G)$	distribution over graphs, i.e., random graph model.
$p_{ij} \in [0,1]$	probability that an undirected edge exists between nodes i and j.
n	number of nodes.
p, p_n	connection probability in the model $G(n,p)$ (with n nodes).
β	inverse temperature in the model $\mathrm{SRGG}(n,\rho,\beta)$.
ρ, ρ_n	density parameter in the model $\mathrm{SRGG}(n,\rho,\beta)$.
$y_i \in [0,n)$	the latent coordinate of node i in the model $\mathrm{SRGG}(n,\rho,\beta)$.
d_{ij}	the distance between nodes i and j in the model $\mathrm{SRGG}(n,\rho,\beta)$.
k_i	degree, i.e., number of neighbors, of node i.
$\bar{k}$, $\bar{k}_n$	average degree of a graph (with n nodes)
$\ell \in \{1,\dots,\ell_{\max}\}$	index into GCN layers.
$w^{(\ell)} \in \mathbb{R}$	weight at GCN layer ℓ
$\boldsymbol{w}$	collection of all GCN weights.
$\sigma : \mathbb{R} \to \mathbb{R}$	a non-linear activation function.
$x_i^{(0)} \in \mathbb{R}$	input feature of node i.
$\boldsymbol{x}^{(0)} \in \mathbb{R}^n$	input features of all nodes
$x_i^{(\ell)}$	embedding of node i in layer $\ell \geq 1$.
$\boldsymbol{x}^{(\ell)} \in \mathbb{R}^n$	embeddings of all nodes.
$h_i^{(\ell)}$	preactivation of node i in layer $\ell \geq 1$.
$\mathbb{P}(x_i^{(\ell)})$	node i's marginal embedding distribution at layer ℓ.
$\mathbb{P}(x_i^{(\ell)} \mid \boldsymbol{x}^{(0)})$	node i's embedding distribution at layer ℓ given all input features.
$\mathbb{P}(h_i^{(\ell)} \mid \boldsymbol{x}^{(\ell-1)})$	node i's preactivation distribution at layer ℓ given all embeddings at layer $\ell-1$.
$\mathcal{N}(\mu,\sigma^2)$	normal distribution with mean μ and variance σ^2.
m_ℓ	empirical first moment of node features at layer ℓ in a graph.
q_ℓ	empirical raw second moment of node features at layer ℓ in a graph.
$\Phi(t)$	moment generating function of the conditional preactivation distribution $\mathbb{P}(h_i^{(1)} \mid \boldsymbol{x}^{(0)})$.
$M_0(t)$	moment generating function of the initial feature distribution.
f_{ij}	a scaled Bernoulli random variables indexed by node pairs i,j.
$\phi_j(t)$	moment generating function of f_{ij} (independent of i).
μ_0	mean of the initial node feature.
σ_0^2	variance of the initial node feature.

Appendix A. Notation

Table 1 summarizes the notation used throughout the paper.

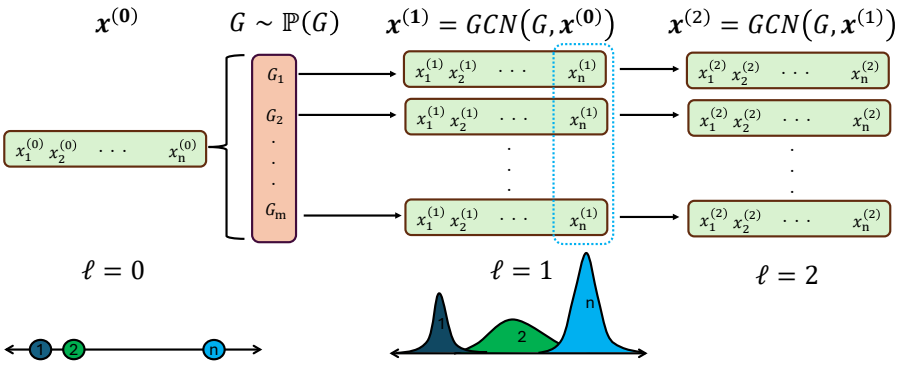

Figure 3: The problem setup. We consider GCN on a random graph model $\mathbb{P}(G)$ with n nodes. Each node has a deterministic initial scalar feature $\boldsymbol{x}^{(0)} = \{x_i^{(0)}\}_{1 \le i \le n}$ and fixed GCN weights $\boldsymbol{w} = \{w^{(\ell)}\}_{0 \le \ell \le \ell_{\max}}$. Since the graph $G \sim \mathbb{P}(G)$ is random, the node embeddings $\boldsymbol{x}^{(\ell)}$ at different layers ℓ are also random variables. Our goal is to quantify the asymptotic nature of these distributions.

Appendix B. Schematic

Our setup is schematically depicted in Figure 3.

Appendix C. Experimental Setup

All our experiments use the GCN architecture described in Equation (1) with $\ell_{\max} = 3$ layers, and one-dimensional weights. The weights are sampled from the standard normal distribution and fixed, throughout. We use the sigmoid function σ as the non-linear function. This means we do not train the GCN model.

Node features are drawn once from the standard normal distribution and fixed, throughout. We ensure that the same node has the same feature for different sampled graphs.

Random graphs are sampled either from $G(n, p_n)$ or $\mathrm{SRGG}(n, \rho_n, \beta)$. We vary the number of nodes, $n \in \{32, 64, \ldots, 16,384\}$, and use 10^3 sampled graphs per setting of the number of nodes for a given average degree, $\bar{k}$.

We use three different functional forms for the average degree. In the dense case, we use $\bar{k}_n = 0.3(n-1) \approx 0.3n$, in the sparse case we use $\bar{k}_n = 1.5\frac{(n-1)}{\sqrt{n}} \approx 1.5\sqrt{n}$, and in the ultra-sparse case we set $\bar{k}_n = 10$. For $G(n, p_n)$, we determine the connection probability as $p_n = \frac{\bar{k}}{n-1}$. For $\mathrm{SRGG}(n, \rho_n, \beta)$, Appendix E describes how we obtain ρ_n from $\bar{k}_n$; we set $\beta = 1.9$.

Appendix D. Additional Results for Erdős-Rényi Graphs

Figure 4 shows the decrease in the average variance of the preactivation distribution for nodes in graphs of different size in the dense, sparse, and ultra-sparse regime. The agreement between theory and simulation confirms that Equation (2) accurately describes the decrease

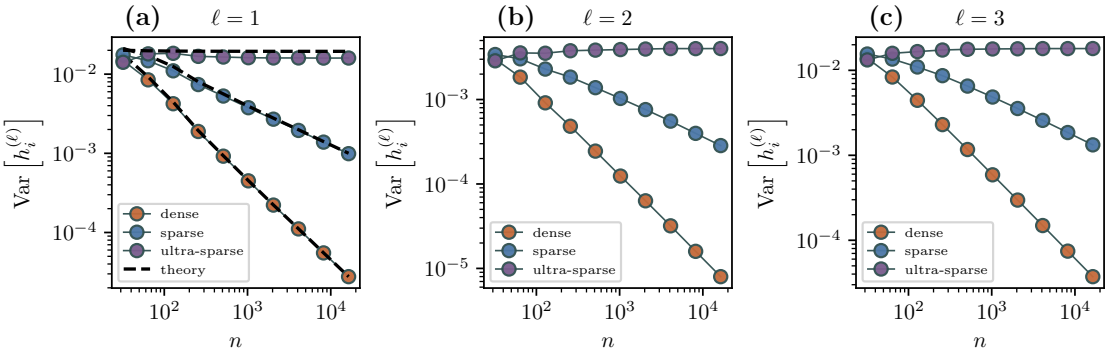

Figure 4: The variance of the node-wise preactivations $h_i^{(\ell)}$ at each layer **(a)** $\ell = 1$, **(b)** $\ell = 2$, and **(c)** $\ell = 3$ decreases as the number of nodes n increases for dense (orange, $p_n = 0.3$) and sparse (blue, $p_n = 1.5/\sqrt{n}$) graphs, but not for ultra-sparse graphs (purple, $p_n = 10/n$). Results are averaged over 10^3 graphs sampled from the model $G(n, p_n)$, and initial node features are drawn independently from the standard normal distribution, $x_i^{(0)} \sim \mathcal{N}(0, 1)$.

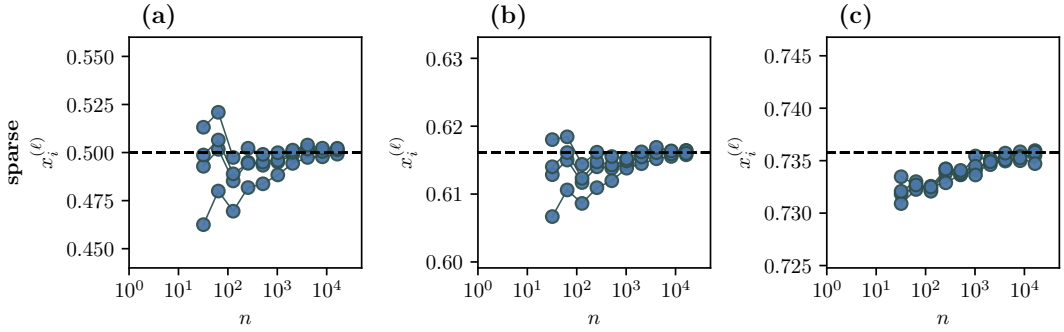

Figure 5: As the graph size increases, the node embeddings $x_i^{(\ell)}$ (shown for four arbitrary nodes) at different layers (1, left, 2, center, 3 right) converge to their value under deterministic updates (black dashed line, Equation (3)) for sparse graphs ($p_n = 1.5/\sqrt{n}$). Results are averaged over 10^3 graphs drawn from $G(n, p_n)$.

in first-layer preactivation variance in both the dense and the sparse cases. For the ultra-sparse case, the average variance does not decrease as the number of nodes in the graph increases.

Figure 5 illustrates the convergence of node embeddings to the theoretically predicted value (Equation (3)) in sparse $G(n, p_n)$ as the number of nodes n increases.

Appendix E. Results for Soft Random Geometric Graphs

Our approach is not restricted to $G(n,p)$ but extends to other random graph models. Here, we showcase this through experiments on soft random geometric graphs (SRGGs).

Setup: The SRGG model is a latent space model. That is, each node $i \in \{1,\dots,n\}$ is assigned a latent coordinate $y_i \in [0,n)$ and the probability p_{ij} of an edge between nodes i and j is a function of the distance d_{ij} between them:

$$p_{ij} = \min\left\{\left(\frac{\rho_n}{d_{ij}}\right)^{\beta}, 1\right\}, \tag{5}$$

where $\beta \in (1,2)$ is called the inverse temperature and controls the amount of clustering and ρ_n determines the average degree. Conditioned on the node coordinates, $\{y_i\}_{i=1}^n$, the probability of a graph is given by,

$$\mathbb{P}(G) = \prod_{i<j} p_{ij}^{G_{ij}}(1-p_{ij})^{1-G_{ij}} \tag{6}$$

In our experiments, we sample the hyper-soft version of this model. For each realization, coordinates are sampled uniformly at random $y_i \sim \mathrm{U}[0,n)$ and distances are computed using periodic boundary conditions:

$$d_{ij} = \frac{n}{2} - \left|\frac{n}{2} - |y_i - y_j|\right|. \tag{7}$$

We fix the inverse temperature at $\beta = 1.9$ to obtain strong clustering, and use numerical root finding to determine the value of ρ_n that results in a given average degree:

$$\bar{k}(n,\rho,\beta) = \frac{2\rho}{\beta-1}\frac{n-1}{n}\left(\beta - \left(\frac{2\rho}{n}\right)^{\beta-1}\right). \tag{8}$$

We choose $\bar{k}_n = 0.3(n-1) \approx 0.3n$ in the dense case, $\bar{k}_n \approx \frac{1.5(n-1)}{\sqrt{n}} \approx 1.5\sqrt{n}$ in the sparse case, and $\bar{k}_n = \frac{10(n-1)}{n} \approx 10$ in the ultra-sparse case.

We use the same GCN architecture as in the main text, and draw node features once from $x_i^{(0)} \sim \mathcal{N}(0,1)$ and hold them constant throughout.

Results: Our findings for SRGGs mirror those for $G(n,p)$. The distribution $\mathbb{P}\left(h_i^{(1)} \mid x_i^{(0)}\right)$ of the preactivations $h_i^{(1)}$ of an arbitrarily chosen node i in the first layer $\ell = 1$, concentrates in the dense case (Figure 6**(a)**) and the sparse case (Figure 6**(b)**), but not in the ultra-sparse case (Figure 6**(c)**). The observed distributions are closely aligned with the theoretical predictions; see Equation (2) for the dense and sparse cases and Equation (4) for the ultra-sparse case.

In both the dense and sparse cases, as the preactivation distributions concentrate at each node, their average node-wise variance decays as a power law in the number of nodes n (Figure 7). In the ultra-sparse case, for which preactivation distributions do not concentrate, the variance remains constant as the number of nodes increases n. For the first layer (Figure 7**(a)**), we correctly predict the exponent of the variance decay (Equation (2)).

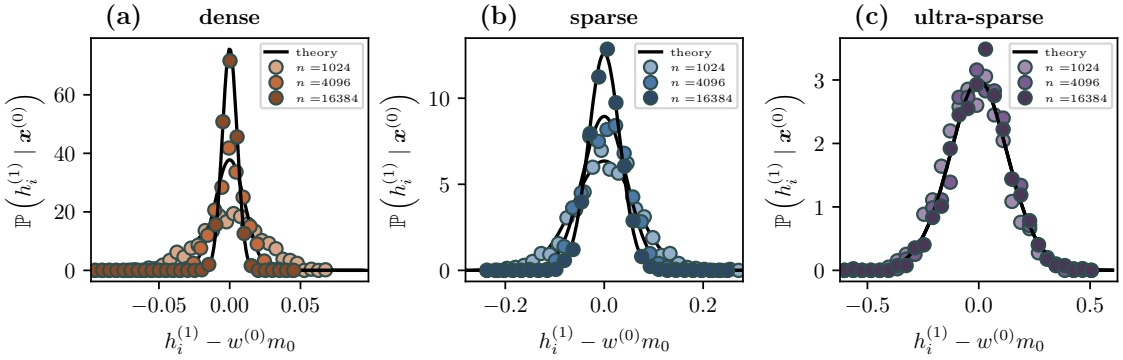

Figure 6: Distribution $\mathbb{P}(h_i^{(1)} \mid x_i^{(0)})$ of the node preactivation $h_i^{(1)}$ of an arbitrarily chosen node i at the first layer for **(a)** dense ($\bar{k}_n \approx 0.3n$, orange), **(b)** sparse ($\bar{k}_n \approx 1.5\sqrt{n}$, blue), and **(c)** ultra-sparse ($\bar{k} \approx 10$, purple) graphs match the theoretical predictions (black, Equation (2) and Equation (4), respectively). The distribution concentrates with increasing graph size $n \to \infty$ for dense and sparse graphs but not for ultra-sparse graphs. Distributions are based on 10^3 graphs sampled from $\mathrm{SRGG}(n, \rho_n, \beta)$ for $\beta = 1.9$ and ρ_n determined via $\bar{k}_n$.

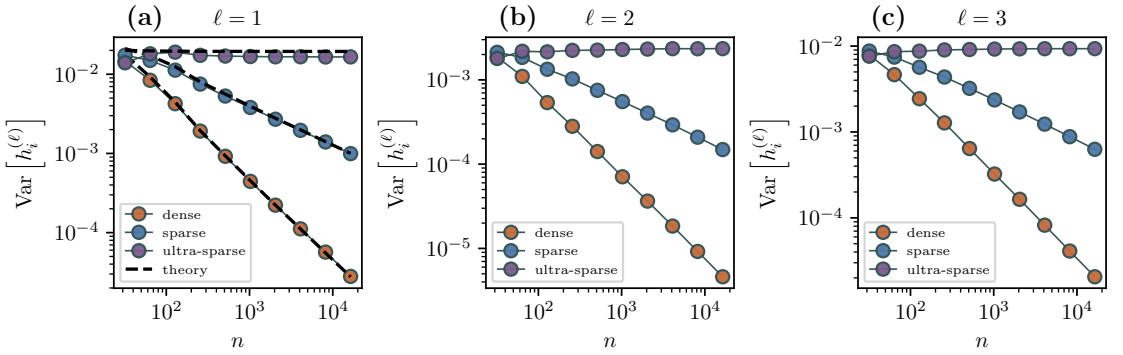

Figure 7: The variance of the node-wise preactivations $h_i^{(\ell)}$ at each layer **(a)** $\ell = 1$, **(b)** $\ell = 2$, and **(c)** $\ell = 3$ decreases as the number of nodes n increases for dense (orange, $\bar{k}_n \approx 0.3n$) and sparse (blue, $\bar{k}_n \approx 1.5\sqrt{n}$) graphs but not for ultra-sparse graphs (purple, $\bar{k}_n \approx 10$). Results are averaged over 10^3 graphs sampled from the model $\mathrm{SRGG}(n, \rho_n, \beta)$ with $\beta = 1.9$, ρ_n determined by the average degree, and initial node features are drawn independently from the standard normal distribution, $x_i^{(0)} \sim \mathcal{N}(0, 1)$.

If the distribution of node embeddings concentrates (i.e., for dense and sparse but not ultra-sparse graphs), updates become deterministic on large graphs (Equation (3)). We

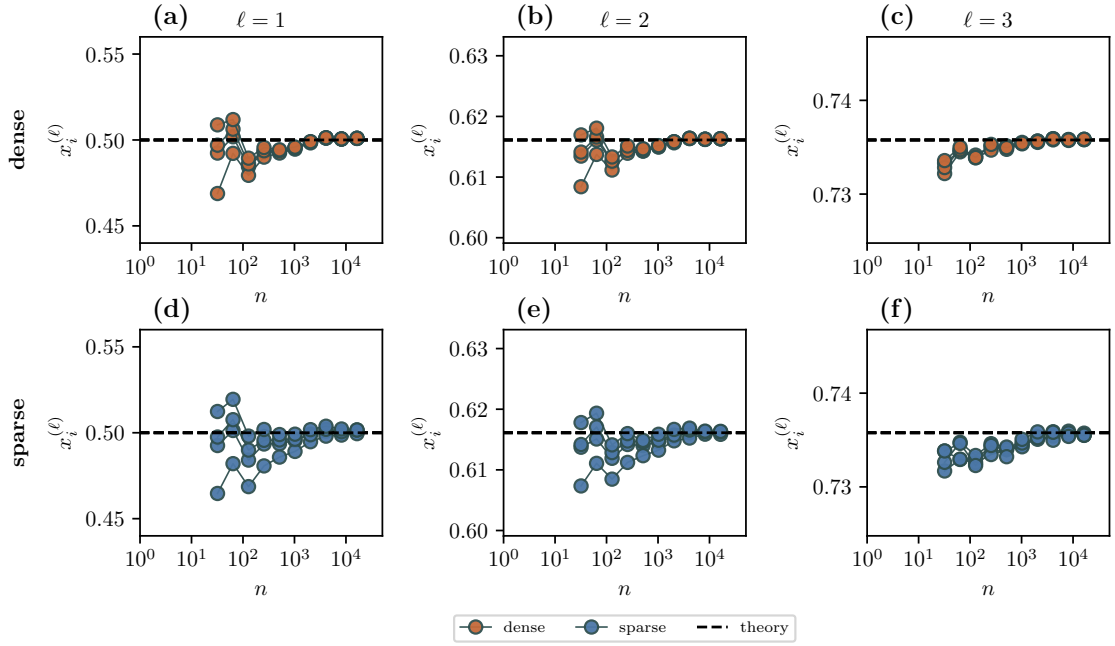

Figure 8: As the graph size n increases, the node embeddings $x_i^{(\ell)}$ (shown for four arbitrary nodes) at layers $\ell = 1$ (left column), $\ell = 2$ (center column), and $\ell = 3$ (right column) converge to their value under deterministic updates (black dashed line, Equation (3)) for dense ($\bar{k}_n \approx 0.3n$, upper row) and sparse ($\bar{k}_n \approx 1.5\sqrt{n}$, lower row) graphs. Results are averaged over 10^3 graphs drawn from SRGG with ρ_n chosen to create a given $\bar{k}_n$ and node features are initialized as $x_i^{(0)} \sim \mathcal{N}(0, 1)$.

illustrate this in Figure 8, where we show, for both dense and sparse graphs, how node embeddings $x_i^{(\ell)}$ at layer $\ell \in \{1, 2, 3\}$ converge to the embeddings expected under deterministic updates as the number of nodes n in the graph increase.

Appendix F. Embedding Uncertainty for Dense or Sparse Erdős-Rényi Graphs

We obtain the asymptotic form of $\mathbb{P}(h_i^{(1)}|\boldsymbol{x}^{(0)})$ on dense/sparse $G(n, p_n)$ graphs, and show that the distribution has vanishing variance. Consequently, all node embeddings collapse to the same deterministic point in the limit $n \to \infty$.

The degree distribution in $G(n, p_n)$ is the binomial distribution, which concentrates around the mean when $\bar{k} \to \infty$. Hence, $k_i = \bar{k} + \mathcal{O}(\sqrt{\bar{k}}) = \bar{k} \cdot (1 + o(1))$. In this setting, Equation (1) reduces to

$$x_i^{(\ell+1)} = \sigma\left(h_i^{(\ell+1)}\right); \quad h_i^{(\ell+1)} = \frac{w^{(\ell)}}{\bar{k}} \sum_{j=1}^{n} G_{ij} x_j^{(\ell)}. \tag{9}$$

Further let,

$$m_\ell = \frac{1}{n}\sum_{j=1}^{n} x_j^{(\ell)}, \quad q_\ell = \frac{1}{n}\sum_{j=1}^{n}\left(x_j^{(\ell)}\right)^2. \tag{10}$$

We begin by considering the case of single-layer GCN on dense and sparse $G(n, p_n)$ graphs. Given $\boldsymbol{x}^{(0)}$, we have:

$$\begin{aligned}
\mathbb{E}[h_i^{(1)}|\boldsymbol{x}^{(0)}] &= \mathbb{E}\left[\frac{w^{(0)}}{\bar{k}}\sum_{j=1}^{n} G_{ij}x_j^{(0)}\right] \\
&= \frac{w^{(0)}}{\bar{k}}\left(x_i^{(0)} + p\sum_{j\neq i} x_j^{(0)}\right) \\
&= \frac{w^{(0)}}{\bar{k}}x_i^{(0)}(1-p) + w^{(0)}m_0 \\
&= w^{(0)}m_0 + \mathcal{O}(\bar{k}^{-1}),
\end{aligned} \tag{11}$$

The conditional expectation of the preactivation of node i at the first layer is given by:

$$\mathbb{E}[h_i^{(1)}|\boldsymbol{x}^{(0)}] = w^{(0)}m_0 + o(1). \tag{12}$$

The variance is:

$$\begin{aligned}
\mathrm{Var}\left[h_i^{(1)}|\boldsymbol{x}^{(0)}\right] &= \mathrm{Var}\left[\frac{w^{(0)}}{\bar{k}}\sum_{j=1}^{n} G_{ij}x_j^{(0)}\right] \\
&= \left(\frac{w^{(0)}}{\bar{k}}\right)^2 \sum_{j=1}^{n}\mathrm{Var}[G_{ij}](x_j^{(0)})^2 \\
&= \left(\frac{w^{(0)}}{\bar{k}}\right)^2 \left[p(1-p)\sum_{j=1}^{n}(x_j^{(0)})^2 - p(1-p)(x_i^{(0)})^2\right] \\
&= \frac{(w^{(0)})^2(1-p)q_0}{\bar{k}} + o(\bar{k}^{-1}),
\end{aligned} \tag{13}$$

Under mild regularity conditions ensuring the Lindeberg condition—such as uniformly bounded features, non-vanishing layer zero empirical second moment—the Lindeberg-Feller central limit theorem yields (Billingsley, 1995):

$$\mathbb{P}\left(h_i^{(1)}|\boldsymbol{x}^{(0)}\right) \approx \mathcal{N}\left(w^{(0)}m_0, \frac{(w^{(0)})^2(1-p)q_0}{\bar{k}}\right). \tag{14}$$

In the thermodynamic limit $n \to \infty$, the distribution collapses to a delta function since the variance scales as $\mathcal{O}(1/\bar{k})$.

Deriving the law for the embeddings $\mathbb{P}(x_i^{(1)}|\boldsymbol{x}^{(0)})$ for finite but large n requires computing the push forward through the non-linear activation function σ since $x_i^{(1)} = \sigma\left(h_i^{(1)}\right)$. However, in the thermodynamic limit, since $\bar{k} \to \infty$, we have $h_i^{(1)} \xrightarrow{(P)} w^{(0)}m_0$, hence:

$$x_i^{(\ell+1)} = \sigma(w^{(\ell)}m_\ell), \tag{15}$$

where $\xrightarrow{(P)}$ denotes convergence in probability.

The message passing updates of the forward pass are thus asymptotically deterministic for all nodes i. Moreover, the node representations collapse to the *same* deterministic point. This same argument also holds for a GCN with arbitrary finite layers $\ell_{\max}$. Therefore, the deterministic update equations are:

$$\begin{aligned} m_0 &= \frac{1}{n}\sum_{j=1}^{n} x_j^{(0)} \quad \text{for } \ell = 0, \\ m_\ell &= \frac{1}{n}\sum_{j=1}^{n} x_j^{(\ell)} = \sigma(w^{(\ell-1)} m_{\ell-1}) \quad \text{and} \\ x_i^{(\ell)} &= \sigma\left(w^{(\ell-1)} m_{\ell-1}\right) \quad \text{for } \ell \geq 1. \end{aligned} \tag{16}$$

Two consequences are worth explicitly stating. First, the node-wise $\mathbb{P}(x_i^{(\ell)})$ are degenerate. The entire ensemble is summarized by the scalar recursion Equation (16), so that a single deterministic forward pass propagating only the mean feature carries the same information as the full ensemble. Second, the arguments above hold when $\bar{k} \to \infty$. In the ultra-sparse case, $\bar{k} = \mathcal{O}(1)$, such a collapse does not happen, and the edge randomness truly induces uncertainty in node embeddings. Deriving the form of $\mathbb{P}(x_i^{(\ell)})$ in the ultra-sparse regime requires advanced methods from statistical physics to handle quenched disorder.

Appendix G. Embedding Uncertainty for Ultra-sparse Erdős-Rényi Graphs

We derive the moment-generating function of the node embedding distribution in the case of a single-layer GCN on ultra-sparse $G(n, p_n)$. Consider the ultra-sparse $G(n, p_n)$ model with $\bar{k} = \mathcal{O}(1)$. That is, p_n equals $\mathcal{O}(1/n)$ on a single-layer GCN (Equation (1)). For analytical tractability, we reduce Equation (1) to:

$$x_i^{(1)} = \sigma\left(\frac{w^{(0)}}{\bar{k}} \sum_{j=1}^{n} G_{ij} x_j^{(0)}\right). \tag{17}$$

In ultra-sparse $G(n, p_n)$, the degree distribution is given by the Poisson distribution, which is no longer self-averaging. Hence, $\sqrt{k_i k_j} \approx \bar{k} + \mathcal{O}(\sqrt{k})$ is an approximation here and is not exact as in the case of dense and sparse graphs. But the behavior of the architecture (Equation (17)) is qualitatively the same as that of Equation (1). As before, let:

$$h_i^{(1)} = \sum_{j=1}^{n} f_{ij}; \;\; f_{ij} = \frac{w^{(0)}}{\bar{k}} G_{ij} x_j^{(0)} \tag{18}$$

Given the initial node feature vector $\boldsymbol{x}^{(0)}$, we wish to find the distribution of $\mathbb{P}(h_i^{(1)})$. Notice that $h_i^{(1)}$ is the sum of independent nonidentical scaled Bernoulli random variables:

$$f_{ij} = \begin{cases} \dfrac{w^{(0)}x_j^{(0)}}{\bar{k}}, & \text{with probability } p, \\ 0, & \text{with probability } 1-p. \end{cases} \tag{19}$$

with the moment-generating function of the form:

$$\phi_j(t) = \mathbb{E}[e^{tf_{ij}}] = (1-p) + pe^{tw^{(0)}x_j^{(0)}/\bar{k}}. \tag{20}$$

By independence, the MGF of $\mathbb{P}(h_i^{(1)}|\boldsymbol{x}^{(0)})$ is given by:

$$\begin{aligned} \Phi(t) &= e^{tw^{(0)}x_i^{(0)}/\bar{k}} \prod_{j\neq i, j=1}^{n} \left(1 - p(1 - e^{tw^{(0)}x_j^{(0)}/\bar{k}})\right), \\ &= e^{tw^{(0)}x_i^{(0)}/\bar{k}} \prod_{j\neq i, j=1}^{n} \left(1 - \frac{\bar{k}}{n}(1 - e^{tw^{(0)}x_j^{(0)}/\bar{k}})\right), \\ \log(\Phi(t)) &= \frac{tw^{(0)}x_i^{(0)}}{\bar{k}} + \sum_{j\neq i, j=1}^{n} \log\left(1 - \frac{\bar{k}}{n}(1 - e^{tw^{(0)}x_j^{(0)}/\bar{k}})\right), \\ &= \sum_{j=1}^{n} -\frac{\bar{k}}{n}(1 - e^{tw^{(0)}x_j^{(0)}/\bar{k}}) + \frac{tw^{(0)}x_i^{(0)}}{\bar{k}} + \mathcal{O}(1/n), \\ \log(\Phi(t)) &= \frac{tw^{(0)}x_i^{(0)}}{\bar{k}} - \bar{k} + \frac{\bar{k}}{n} \sum_{j\neq i, j=1}^{n} e^{tw^{(0)}x_j^{(0)}/\bar{k}} + \mathcal{O}(1/n) \end{aligned} \tag{21}$$

Further, under the assumption that $x_i^{(0)} \sim \mathbb{P}(x_i^{(0)})$, we have:

$$\begin{aligned} \frac{1}{n} \sum_{j\neq i, j=1}^{n} e^{tw^{(0)}x_j^{(0)}/\bar{k}} &\approx \int_{-\infty}^{\infty} e^{tw^{(0)}x_j^{(0)}/\bar{k}} \mathbb{P}(x_j^{(0)}) dx_j^{(0)}, \\ &= M_0\left(\frac{tw^{(0)}}{\bar{k}}\right), \end{aligned} \tag{22}$$

where $M_0(t) = \mathbb{E}[e^{tx^{(0)}}]$ is the MGF of the initial node features.

Hence, in the limit $n \to \infty$:

$$\log\left(\Phi(t)\right) = \frac{tw^{(0)}x_i^{(0)}}{\bar{k}} - \bar{k} + \bar{k} \cdot M_0\left(\frac{tw^{(0)}}{\bar{k}}\right). \tag{23}$$

Differentiating Equation (23) and evaluating at $t = 0$ together with $M_0(0) = 1$, $M_0'(0) = \mu_0$ and $M_0''(0) = \mu_0^2 + \sigma_0^2$ with μ_0 and σ_0^2 being the mean and the variance of the initial node features, allows us to calculate the mean and variance of $h_i^{(1)}$ as follows:

$$\begin{aligned}\mathbb{E}[h_i^{(1)}|\boldsymbol{x}^{(0)}] &= w^{(0)}\left(\mu_0 + \frac{x_i^{(0)}}{\bar{k}}\right) \\ \mathrm{Var}[h_i^{(1)}|\boldsymbol{x}^{(0)}] &= \frac{\left(w^{(0)}\right)^2(\mu_0^2+\sigma_0^2)}{\bar{k}}.\end{aligned} \tag{24}$$