Encoding Graph Topology with Randomized Ising Models

Domenico Tortorella[1], Antonio Brau[2], Alessio Micheli[1]

1- University of Pisa - Department of Computer Science
Largo B. Pontecorvo 3, 56127 Pisa - Italy

2- University of Pisa - Department of Physics 'E. Fermi'
Largo B. Pontecorvo 3, 56127 Pisa - Italy

Abstract. The increasing popularity of deep learning on graphs has motivated the need for the co-design of hardware and graph representation models. We propose Randomized Ising Model (RIM), a reservoir computing model for encoding topological information of graph nodes, that is amenable to physical implementation via neuromorphic hardware. Our experiments demonstrate that RIM's node embeddings are able to provide sufficient topological information to be suitable to address node classification tasks, exhibiting an accuracy in line with Graph Echo State Networks.

1 Introduction

Deep learning on graph has recently enjoyed an ever-increasing popularity on a wide range of applications, since diverse data such as molecular compounds, protein interactions, social networks, etc. can be all represented as graphs. As a consequence, a plethora of graph models have been proposed to address those application as learning problems on graphs [1]. However, since both the dimension of graph data to be processed as well as the dimension of deep learning models is approaching scales that stretch the capabilities of currently available computational resources, the need to rethink both the design of graph model and computing hardware arises. A promising line to address these challenges has sprung up at the intersection of two research fields: reservoir computing and neuromorphic computing [2]. In the reservoir computing paradigm, input data is encoded via a randomly-initialized reservoir, while only a readout classifier for the downstream task requires training. Graph Echo State Networks (GESN) [3] are an efficient model for graph representation that follows this approach. The feasibility of a GESN implementation on memristive neuromorphic hardware has been recently explored, showing promising results on node and graph classification tasks [2]. In this paper, we propose *Randomized Ising Model* (RIM) as a neuromorphic reservoir-based model for encoding graph topological information as the spin configuration of a Lenz–Ising mathematical model of ferromagnetic interaction. As a proof-of-concept, we perform simulations instead of physically implementing our model, and we focus the experimental analysis to assess the ability of RIM's node embeddings to encode the topological information necessary to address node classification tasks as effectively as other methods.

2 Representing topological information in nodes

Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be a graph with set of nodes $\mathcal{V}$ and set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. We denote by $\mathcal{N}(v)$ the neighborhood of node v, and by $\mathbf{A}$ the graph adjacency matrix. A node embedding vector $\mathbf{h}_v \in \mathbb{R}^H$ should encode enough information on the graph topology to represent the node's relationship with the rest of the network it belongs to. Several approaches have been proposed so far to obtain such embeddings, briefly surveyed as follows.

Positional encoding As the transformer architecture has been generalized from sequences in natural language processing to graph learning [4], positional encodings (PE) have been generalized to represent the position of a node within the graph topology. Laplacian PE (LPE) adopt the H leading eigenvectors of the symmetric graph Laplacian matrix $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, where $\mathbf{D}$ is the degree matrix, as the node positional encoding. Laplacian eigenvectors have previously been applied in tasks such as spectral clustering of graph nodes [5]. Random-walk PE (RWPE) are instead computed as the diagonal entries of the first H powers of the random walk matrix $\mathbf{A}\mathbf{D}^{-1}$ [6]. For both PEs, the computational cost becomes cubic in the number of nodes as the embedding dimension H approaches $|\mathcal{V}|$.

Representation learning Neural models for graphs consisting of a deep hierarchy of L convolutional layers that perform local aggregation of node features are most popular approach for learning node representations. Among the plethora of models proposed in literature [1] that differ just by the type of convolution layer, we consider two examples in this paper. GCN [7] defines its convolution operation as $\mathbf{h}^{(\ell)} = \mathrm{ReLU}\left(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\,\mathbf{h}^{(\ell-1)}\,\mathbf{W}^{(\ell)}\right)$, whereas GIN [8] defines it as $\mathbf{h}^{(\ell)} = \mathrm{MLP}^{(\ell)}\left((\mathbf{I} + \mathbf{A})\,\mathbf{h}^{(\ell-1)}\right)$, where $\mathrm{MLP}^{(\ell)}$ is a multi-layer neural networks. Both the weighs $\mathbf{W}^{(\ell)}$ and the parameters of $\mathrm{MLP}^{(\ell)}$ in each layer are trained end-to-end with the downstream task, which uses the final layer representations $\mathbf{h}^{(L)}$ as node embeddings.

Reservoir computing Graph Echo State Networks (GESN) [3] are an efficient model within the Reservoir Computing (RC) paradigm. To encode graph topology, node embeddings are recursively computed by the dynamical system

$$\mathbf{h}_v^{(k)} = \tanh\left(\mathbf{w}_{\text{in}} + \textstyle\sum_{v' \in \mathcal{N}(v)} \hat{\mathbf{W}}\,\mathbf{h}_{v'}^{(k-1)}\right), \quad \mathbf{h}_v^{(0)} = \mathbf{0}, \tag{1}$$

where $\mathbf{w}_{\text{in}} \in \mathbb{R}^H$ and $\hat{\mathbf{W}} \in \mathbb{R}^{H \times H}$ are the input-to-reservoir and the recurrent weights, respectively. Reservoir weights are randomly initialized from a uniform distribution in $[-1, 1]$, and then rescaled to the desired input scaling and reservoir spectral radius, without requiring any training. Previous results [9] have demonstrated that GESN is particularly effective in encoding topological information when $\rho(\hat{\mathbf{W}}) \gg 1/\rho(\mathbf{A})$, where $\rho(\mathbf{A})$ denotes the graph spectral radius, i.e. the largest absolute eigenvalue of its adjacency matrix $\mathbf{A}$. In this case, as the dynamical system does not converge [10], equation (1) is iterated over for a number of steps K at least as large as the graph diameter [9]; the final state $\mathbf{h}_v^{(K)}$ is used as the node embedding.

3 Randomized Ising models for node embedding

Lenz–Ising models [11] are mathematical models that represent magnetic interactions in statistical mechanics. The model consists in a set of discrete variables h_i that represent magnetic *spins* that can assume the two values ± 1. A particular assignment of spin values $\mathbf{h}$ is called a *spin configuration*. The probability of the system assuming a particular configuration is given by the Boltzmann distribution with inverse thermodynamic temperature $\beta \geq 0$, $P_\beta(\mathbf{h}) = e^{-\beta E(\mathbf{h})}/Z_\beta$, where $E(\mathbf{h})$ is the *energy* of the configuration and Z_β is the partition function.

We propose to employ the spin configuration as a way to encode node representations in a graph. To this end, we define the energy function of our Ising model as the Hamiltonian

$$E(\mathbf{h}) = -\mathbf{h}^\top (\mathbf{A} \otimes \hat{\mathbf{W}})\, \mathbf{h}, \tag{2}$$

where the interaction matrix $\mathbf{A} \otimes \hat{\mathbf{W}}$ is defined as the Kronecker product of the graph adjacency $\mathbf{A}$ and a random matrix $\hat{\mathbf{W}} \in \mathbb{R}^{H \times H}$. The spin configuration $\mathbf{h} \in \mathbb{R}^{|\mathcal{V}|H}$ is partitioned in groups of H spins, each of whom corresponds to a node embedding $\mathbf{h}_v$. We call our model RIM for *Randomized Ising Model*.

The Ising model defined in equation (2) presents no external magnetic field. In this paper, we further assume that the random matrix $\hat{\mathbf{W}}$ is sparse, symmetric, with ± 1 entries sampled from an unskewed Bernoulli distribution, and that the system is at near-zero thermodynamic temperature, i.e. $\beta \to \infty$. In this setting, for the particular case $H = 1$, our Ising model is analogous to a Curie–Weiss model [12], inducing local energy minima at near-zero besides the two ground state spin configurations $\mathbf{h} = \pm \mathbf{1}$. Those local minima have been demonstrated to correspond to non-trivial solutions to the MINCUT problem, thus hinting at the topological informative content that RIM's node embedding can provide.

As a general mathematical model, RIM can be physically implemented by unconventional computing techniques such as memristive circuits [13] or spintronic devices [14]. Our experiments simulate the Ising model via the Metropolis–Hastings algorithm, a well-established Markov chain Monte Carlo method to approximate high-dimensional distributions by iterative sampling.

4 Experiments and discussion

As a practical approach to measure the quality of node embeddings, we rely on their effectiveness as representations for semi-supervised node classification tasks. Node embeddings $\mathbf{h}_v$ are used as features for a linear readout classifier $\mathbf{y}_v = \mathbf{W}_{\text{out}}\, \mathbf{h}_v + \mathbf{b}_{\text{out}}$. The weights $\mathbf{W}_{\text{out}} \in \mathbb{R}^{C \times H}, \mathbf{b}_{\text{out}} \in \mathbb{R}^C$ are trained by ridge regression on one-hot encoding of target class $y_v \in 1, ..., C$ for all models

Encoding model		Cora	CiteSeer	PubMed
	Baseline	28.9 ± 2.2	20.6 ± 2.2	39.7 ± 0.5
	LINK	75.4 ± 2.5	58.3 ± 2.8	77.1 ± 0.9
Positional encoding	LPE	82.9 ± 1.3	$\mathbf{67.4} \pm 0.3$	82.2 ± 0.5
	RWPE	34.1 ± 2.0	29.3 ± 2.8	OOT
Representation learning	GCN	56.8 ± 3.3	37.2 ± 3.6	62.6 ± 1.3
	GIN	71.2 ± 2.3	50.4 ± 4.9	70.6 ± 2.0
Reservoir computing	GESN	84.0 ± 1.2	61.8 ± 5.7	$\mathbf{82.5} \pm 0.6$
	RIM	$\mathbf{84.2} \pm 1.1$	62.4 ± 4.9	81.7 ± 0.6

Table 1: Node classification accuracy, average over 10 folds with standard deviation. Our proposed model is highlighted. The majority class baseline is reported for reference. 'OOT' denotes that computation exceeded 3 hours.

except GCN and GIN, where the embedding models and the readouts are trained end-to-end.

We perform our experiments on the popular node classification tasks Cora, CiteSeer, PubMed without input features, following the public 10-fold scaffold training/validation/test splits of [15]. For RIM, we fix the sparsity of $\hat{\mathbf{W}}$ at 100 non-zero entries per row; we average the results over 5 Metropolis–Hastings simulations with different random initial spin configurations and number of iterations $50|\mathcal{V}|H$. For GESN, we explore the same hyper-parameter ranges of [9]. GCN and GIN are trained via Adam optimizer with early stopping for 2500 epochs with cross-entropy loss, selecting the number of layers $L \in \{1, ..., 5\}$, the learning rate in $\{10^{-2}, 10^{-3}\}$, and the weight decay in $\{0, 10^{-5}, 10^{-3}\}$. In all models we consider embedding dimension $H \in \{2^4, ..., 2^{12}\}$.

In addition to the models of Sec. 2, we report also the accuracy of the majority-class baseline and of LINK, which uses directly the rows of $\mathbf{A}$ as node embeddings. The results of Tab. 1 show that all embedding model are able to perform better than the baseline, a sign that topological information is required to address the tasks. Despite its high computational cost, RWPE performs barely above the baseline. Both representation learning models GCN and GIN perform significantly worse than LINK on all three tasks. On Cora, RIM achieves top accuracy, closely followed by GESN and LPE. On CiteSeer, LPE is instead the best performing node embedding, followed at a significant distance by RIM and GESN; however, the accuracy of LPE must be considered together with its demanding computational cost of $O(|\mathcal{V}|^3)$ due to the eigendecomposition of $\mathbf{L}$. Finally, on PubMed, GESN is the achieves top accuracy, closely followed by LPE and RIM. Overall, the experiments demonstrate that the reservoir computing models, namely GESN and our proposed RIM, generally rank as the best accuracy models, achieving similar performance on the considered tasks.

In Fig. 1 we report the node classification accuracy as a function of the embedding dimension for the three best-performing models of Tab. 1. The analysis

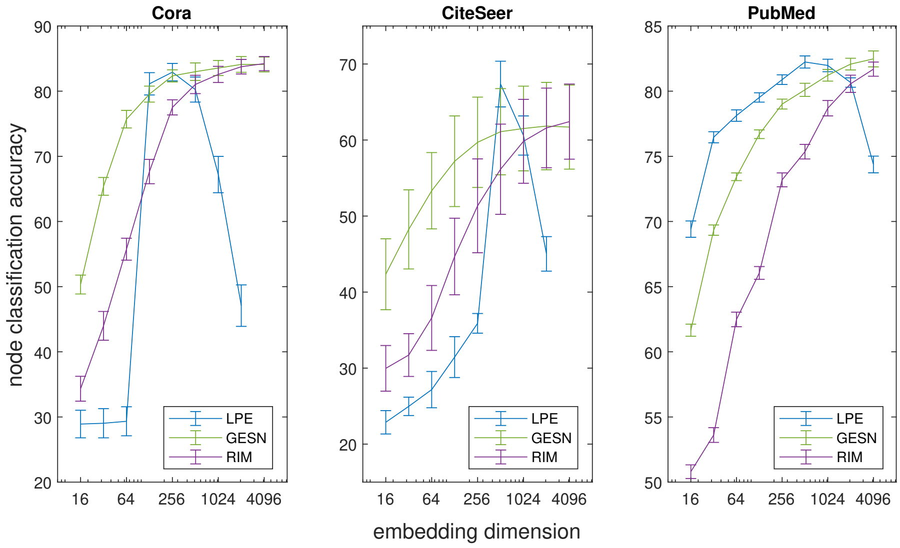

Fig. 1: Node classification accuracy as a function of embedding dimension H for LPE, GESN and our proposed model RIM.

confirms that LPE requires an embedding dimension large enough to bring its computational cost close to cubic in order to match or improve the accuracy of reservoir computing models. While GESN achieves better accuracy compared to RIM for smaller embedding dimensions, the gap closes as both reservoir models arrive at similar performances for $H = 2^{12}$. This suggests the need for physical implementations of RIM to be able to scale to large number of spin units.

5 Conclusion and future directions

In this paper, we have proposed Randomized Ising Model (RIM) as a reservoir-based model for node embedding amenable to physical implementation on neuromorphic hardware [13, 16, 14]. As a proof-of-concept, we have simulated RIM on three citation networks to evaluate its effectiveness in encoding graph topology sufficiently well to address node classification tasks. The experiments have demonstrated that RIM performs in line with Graph Echo State Networks, a different reservoir computing model that has recently been considered for neuromorphic implementation [2].

Building on these preliminary results, in future works we will extend RIM to take into account also input features in its node encoding. We will also expand our analysis to consider embedding quality metrics such as Shannon entropy or Dirichlet energy that are task-independent [17], as well as analyze more in depth the factors that can affect the properties of Ising models such as thermodynamic temperature and interaction sparsity [18].

Acknowledgments Research partly supported by PNRR, PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1, funded by European Commission under the NextGeneration EU programme.

References

[1] D. Bacciu, F. Errica, A. Micheli, and M. Podda. A gentle introduction to deep learning for graphs. *Neural Networks*, 129:203–221, 2020.

[2] S. Wang, Y. Li, D. Wang, W. Zhang, X. Chen, D. Dong, S. Wang, X. Zhang, P. Lin, C. Gallicchio, X. Xu, Q. Liu, K. Cheng, Z. Wang, D. Shang, and M. Liu. Echo state graph neural networks with analogue random resistive memory arrays. *Nature Machine Intelligence*, 5(2):104–113, 2023.

[3] C. Gallicchio and A. Micheli. Graph echo state networks. In *Proceedings of the 2010 International Joint Conference on Neural Networks*, pages 3967–3974, 2010.

[4] V. P. Dwivedi and X. Bresson. A generalization of transformer networks to graphs. In *AAAI Workshop on Deep Learning on Graphs: Methods and Applications*, 2021.

[5] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[6] V. P. Dwivedi, A. T. Luu, T. Laurent, Y. Bengio, and X. Bresson. Graph neural networks with learnable structural and positional representations. In *10th International Conferenceon Learning Representations*, 2022.

[7] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations*, 2017.

[8] K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks? In *7th International Conference on Learning Representations*, 2019.

[9] A. Micheli and D. Tortorella. Addressing heterophily in node classification with graph echo state networks. *Neurocomputing*, 550:126506, 2023.

[10] D. Tortorella, C. Gallicchio, and A. Micheli. Spectral bounds for graph echo state network stability. In *The 2022 International Joint Conference on Neural Networks*, 2022.

[11] S. G. Brush. History of the Lenz-Ising model. *Reviews of Modern Physics*, 39(4):883–893, 1967.

[12] R. Gheissari, C. M. Newman, and D. L. Stein. Zero-temperature dynamics in the dilute Curie–Weiss model. *Journal of Statistical Physics*, 172(4):1009–1028, 2018.

[13] V. J. Dowling and Y. V. Pershin. Memristive Ising circuits. *Physical Review E*, 106(5):054156, 2022.

[14] D. Bhowmik, R. S. Yadav, N. Garg, A. Holla, and P. K. Muduli. Spintronics-based neuromorphic and Ising computing. In *8th IEEE ETDM Conference*, 2024.

[15] H. Pei, B. Wei, K. C. Chang, Y. Lei, and B. Yang. Geom-GCN: Geometric graph convolutional networks. In *8th International Conference on Learning Representations*, 2020.

[16] G. Milano, G. Pedretti, K. Montano, S. Ricci, S. Hashemkhani, L. Boarino, D. Ielmini, and C. Ricciardi. In materia reservoir computing with a fully memristive architecture based on self-organizing nanowire networks. *Nature Materials*, 21(2):195–202, 2022.

[17] A. Micheli and D. Tortorella. Designs of graph echo state networks for node classification. *Neurocomputing*, 597:127965, 2024.

[18] A. Pournaki, E. Olbrich, S. Banisch, and K. Klemm. Order-disorder transition in the zero-temperature Ising model on random graphs. *Physical Review E*, 107(5):054112, 2023.