Natural Image Classification via Quasi-Cyclic Graph Ensembles and Random-Bond Ising Models with Enhanced Nishimori Temperature Estimation

Классификация естественных изображений с помощью ансамблей квазициклических графов и моделей Изинга со случайными связями с улучшенной оценкой температуры Нисимори

Usatyuk V., Researcher, South-West State University
Sapoznikov D., Senior Research Engineer, T8 LTD
Egorov S., Professor, South-West State University

The 9th International Conference on Deep Learning in Computational Physics

SINP MSU, Moscow, Russia
2 July, 2025

Goal of Research:

Improve of DLCP 2024 results *: from GAN generated Images to Natural Images

Propose:

1. constructing a ensemble **priori graphs (Stochastic Block Model of Image features)** of graph spectral embedding (for MLP at CNN/Transformer, Diffusion graph) with the property of unsupervised clustering on Random-Bond Ising Models which under
2. Bethe-temperature (in the finite) tends to the minimum of Bethe Free Energy (maximum local Entropy) under high (fraction/fractal) dimension of Natural Images (compare to GAN/Diffusions)

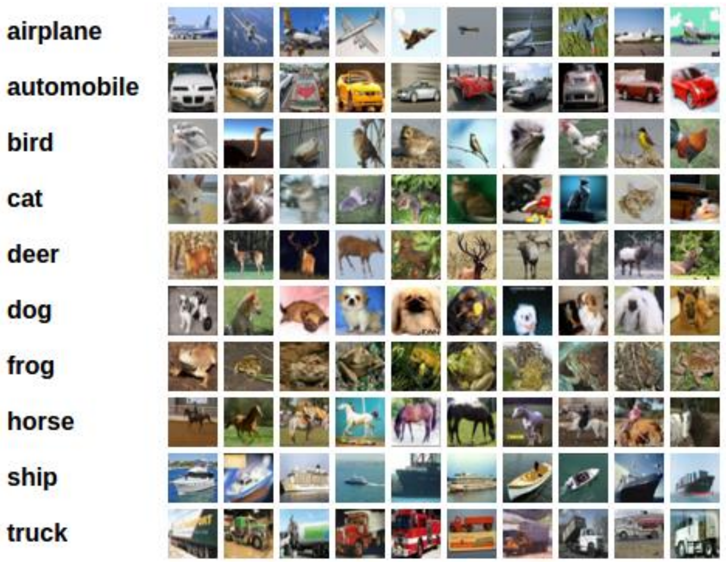
*Usatyuk, V.S., Sapozhnikov, D.A. & Egorov, S.I. Enhanced Image Clustering with Random-Bond Ising Models Using LDPC Graph Representations and Nishimori Temperature. *Moscow Univ. Phys.* **79** (Suppl 2), S647–S665 (2024).

Illustration of Problem Statement

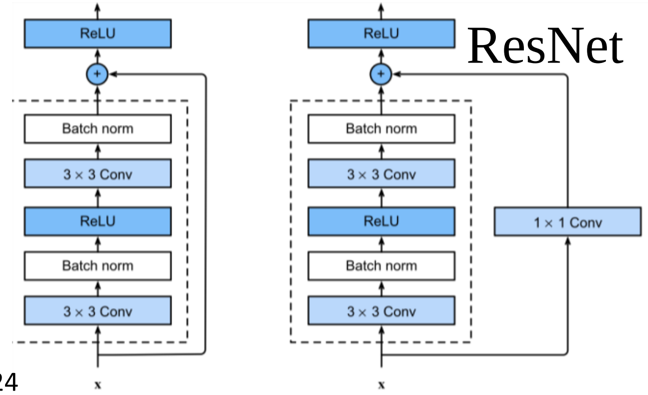
Layer Name	Activation Type	Embedding Dimension m_k	Estimated Dimension, D_k
Input	None	3072	32.4
relu1	relu	16384	51.2
stack1block4relu3	relu	65536	93.5
stack2block3relu3	relu	32768	67.0
stack3block2relu3	relu	16384	44.9
global average pooling	pooling	256	16.2
fully connected	linear	10	8.1
softmax	softmax	10	2.8

[1,2]

CIFAR10



Eight layers ResNet, the embedding dimension [1,2] and the estimated dimension for the embedded manifolds, sampled by the N = 10^4 points cloud.



ResNet

Current DNN are far from efficient, $m_k > \frac{D_k(D_k+1)}{2}$

[3]

[1] Gäfvert O. et al. On the hidden layer-to-layer topology of the representations of reality realised within neural networks. Engineering Computations, June 11, 2024

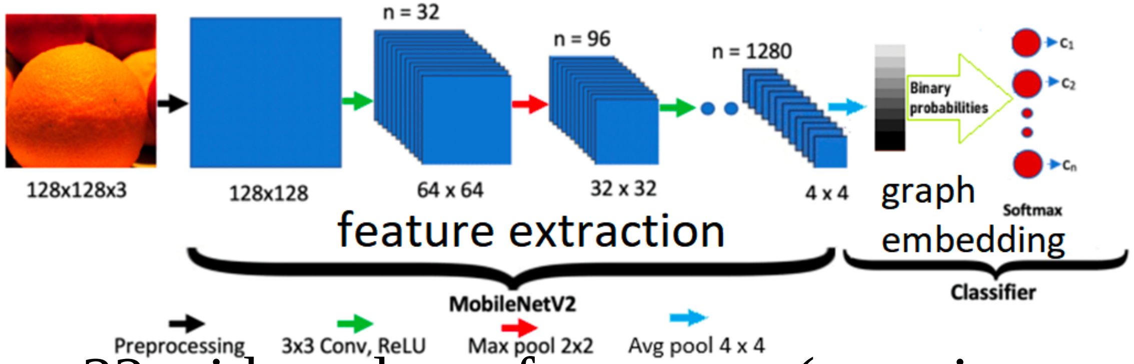
[2] Rocco, S.D., Edwards, P.B., Eklund, D., Gäfvert, O., & Hauenstein, J.D. (2022). Computing geometric feature sizes for algebraic manifolds. *ArXiv, abs/2209.01654*.

[3] Günther, Matthias **Isometric embeddings of Riemannian manifolds.** *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*, 1137–1143. *Mathematical Society of Japan, Tokyo*, 1991

Empirical Application of Research:

Democratize AI by reducing compute costs—akin to how message passing (BP) revolutionized error-correcting codes.

Replace MLP (Feed-Forward) Layers in CNN/Transformer and (Diffusion via parallel walks on a QC graph) graph embedding.



40 times reduction of param., improve false detection, robust to weight prune and quantizat.

1280 features became 32 without lost of accuracy (even improved and robust to puncture)

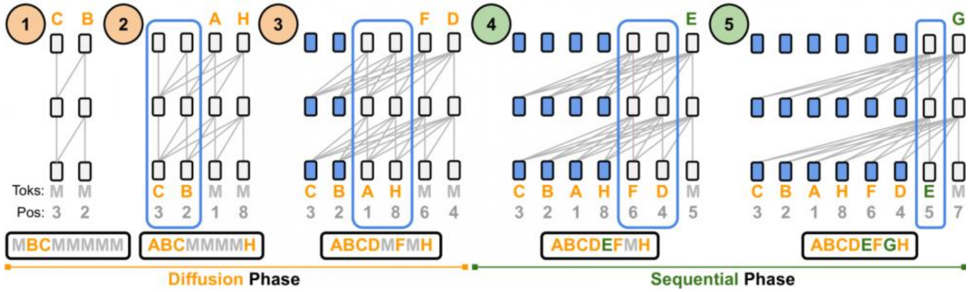
Visual Transformers similar

Language Model Diffusion Perplexity superior

With block structured parallelism and

Easy scale (recalculate) among model family

8B-32B-128B-256B …, and using QC graph lifting any size



Eso-LM Diffusion LLMs

Auto Adapting , prediction of activation pattern and SDE closed description, BP/min-sum train

Theoretical Application of Research:

Symmetry turns the Bethe-Hessian into a computational assembly map– a discrete bridge to Novikov conjection, [1].
*Can topological invariants of manifolds be recovered from analytic data (e.g., spectra of differential operators **H**)? [2]*

$$\frac{dE_\lambda}{d\lambda} = \left\langle \psi_\lambda \middle| \frac{d\hat{H}_\lambda}{d\lambda} \middle| \psi_\lambda \right\rangle$$

Spectral Degeneracy Meets Group Theory

In **sparse k-regular quasi-cyclic bipartite graphs with**:
- Girth ≥6 (no small cycles),
- Large Hamming distance or max permanent, (structural rigidity by perfect match)

The **Bethe-Hessian** H **(deformated Laplacian) matrix** exhibits:

Eigenvalues: Explicitly linked to **group characters** (via Fourier transform on $\mathbb{Z}/n\mathbb{Z}$)

Zero modes: Multiplicity = number of invariant harmonic cycles under group action

Under such Graphs - Spectral degeneracy encodes symmetry-protected topology!

Novikov Conjecture Analogy

Novikov Context	Quasi-Cycle Graphs
Group $\pi_1(M)$ actions	$\mathbb{Z}/n\mathbb{Z}$ symmetry
L^2-Betti numbers [5]	Bethe-Hessian H zero modes
Assembly maps (K-theory)	Spectral projections of $H(r)$

The **group-averaged Bethe-Hessian** $\langle \boldsymbol{H}(\boldsymbol{r}) \rangle_{\boldsymbol{G}}$ under sym. group G mimics the assembly map in Novikov conjectures.
Persistent zero modes under r-filtration ≈ **homotopy-invariant signatures**

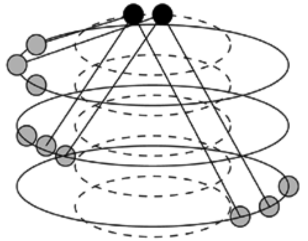
Ollivier-Ricci Curvature (ORC, [3]):

Quasi-cycles have **uniform curvature** (all edges equivalent under symmetry).
ORC detects symmetry-induced topological rigidity.

Graphon linear quadratic regulation problem, **[4]**:

Subspace decomposition aligns with group-invariant subspaces.
Subspace decomposition ≈ Spectral projection of Bethe-Hessian.
Nishimori temperature $\boldsymbol{r}$ acts like *scale-separation parameter* in control theory.

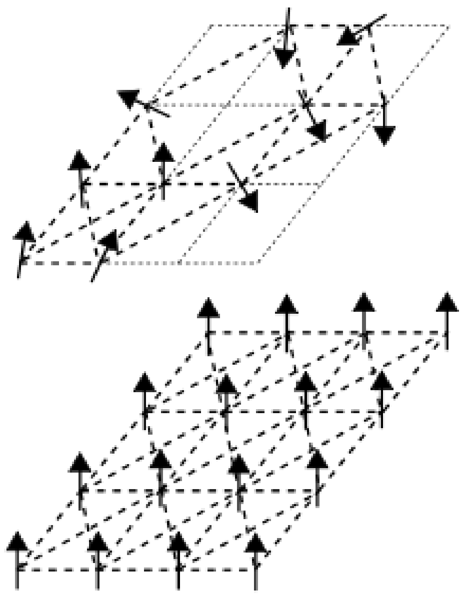
QC Toric symmetry-protected topology characterized by long-range entanglement and robustness against local perturbations



[1] S P. Novikov, Algebraic construction and properties of Hermitian analogs of k-theory over rings with involution from the point of view of Hamiltonian formalism. Some applications to differential topology and to the theory of characteristic classes. Izv.Akad.Nauk SSSR, v. 34, 1970 I N2, pp. 253–288; II: N3, pp. 475–500.
[2] Hellmann–Feynman theorem. Hellmann H (1937). *Einführung in die Quantenchemie*. Leipzig: Franz Deuticke. p. 285.Feynma R. P. (1939). "Forces in Molecules". *Physic. Rev.*. **56** (4): 340–343. Güttinger P. (1932). *Pauli W. (1933).*
[3] Sia, J., Jonckheere, E. & Bogdan, P. Ollivier-Ricci Curvature-Based Method to Community Detection in Complex Networks. *Sci Rep* **9**, 9800 (2019)
[4] S. Gao and P. E. Caines, "Subspace Decomposition for Graphon LQR: Applications to VLSNs of Harmonic Oscillators," in *IEEE Transactions on Control of Network Systems*, vol. 8, no. 2, pp. 576-586, 2021
[5] Danijela Horak, Jürgen Jost, Spectra of combinatorial Laplace operators on simplicial complexes, Advances in Mathematics, Volume 244, 2013, pp. 303-336

Random Bond Ising Model (RBIM) - 2D Ising model on Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

$$\mathcal{H}(\sigma) = -\sum_{(i,j)\in\mathcal{E}} J_{ij}\sigma_i\sigma_j,$$



Each node $i \in \mathcal{V}$ is associated with a spin variable, $\sigma_i \in \{-1, +1\}$, $\mathcal{E}$ - is the set of edges (bonds)
J_{ij}- bond strengths, which are random assigned according to a specified probability distribution.
In the random bond Ising model, J_{ij} can take different values, typically following a distribution such as

$$P(J_{ij}) = p\delta(J_{ij} - J) + (1-p)\delta(J_{ij} + J),$$

where δ is the Dirac delta function, J is a positive constant representing the bond strength, and p is the probability of a bond being ferromagnetic ($J_{ij} = J$) versus antiferromagnetic ($J_{ij} = -J$).

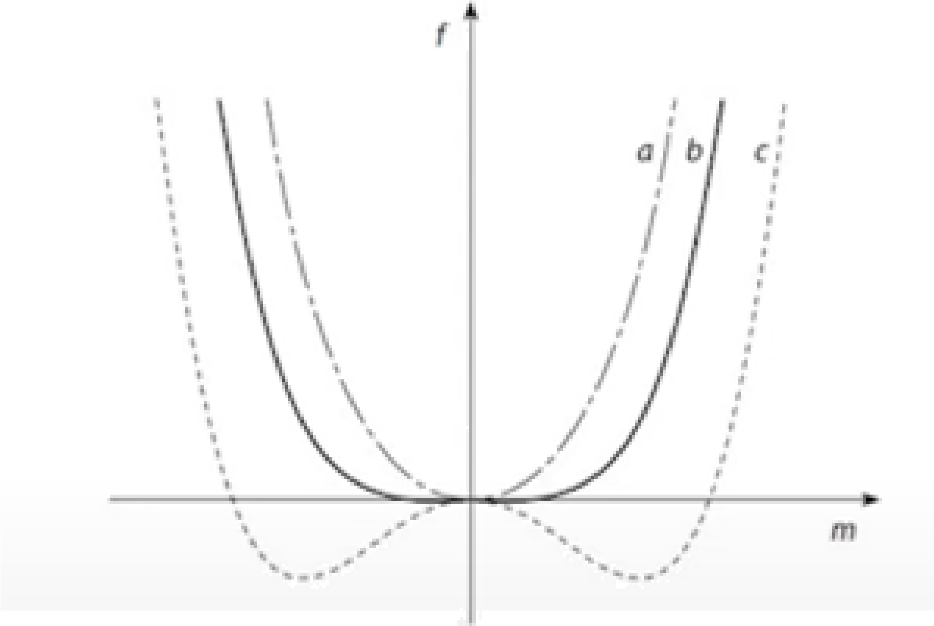
The partition function Z of the system, which is used to study the statistical properties, is given by

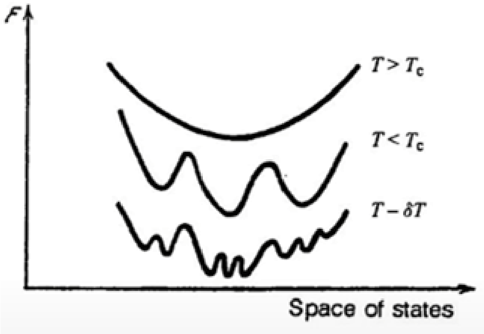
$$Z = \sum_{\sigma} e^{-\beta\mathcal{H}(\sigma)},$$

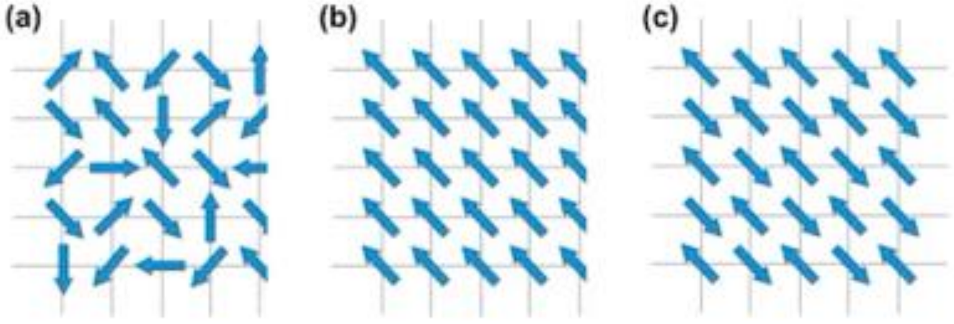
where $\beta = \frac{1}{k_B T}$ is the inverse temperature, with k_B being the Boltzmann constant and T the temperature.

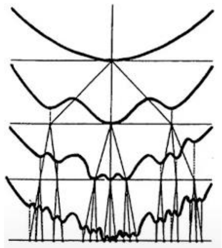
Rapaport D. C.. , *J. of Phys.* , 1972 , vol. C5 1830[IoP STACKS]; D. C. Rapaport, J. of Phys. **C5** (1972),2813[IoP STACKS], K. I. Grozdev and A. M. Kosevich, Phys. Stat. Sol. **b66** (1974), 77; K. I. Grozdev and A. M. Kosevich, Soviet Phys.-JETP **66** (1974), 2269

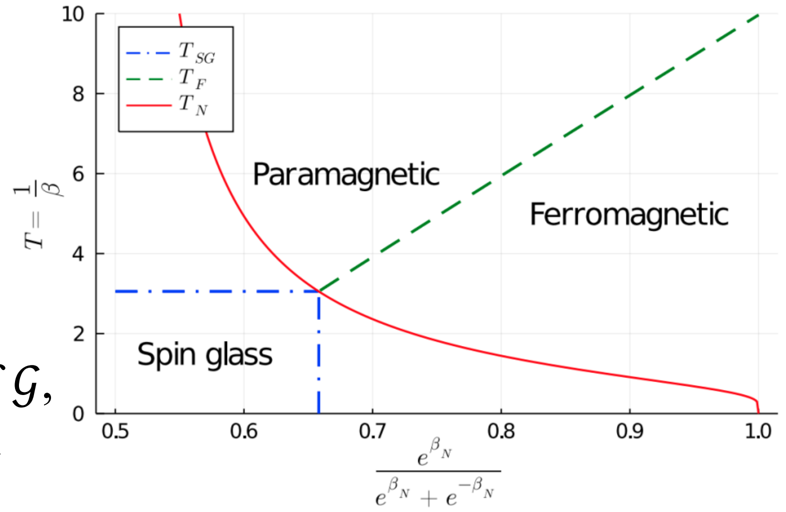
RBIM phase transaction under different temperature











At different temperature T we can see a phase transition

Under Nishimori Temp. Let $A \in \{0,1\}^{n\times n}$ be the symmetric adjacency matrix of $\mathcal{G}$, defined by $A_{ij} = 1$ if $(ij) \in \mathcal{E}$, and $A_{ij} = 0$ otherwise. Let $D \in \mathbb{N}^{n\times n}$ be the diagonal degree matrix $D = \mathrm{d}iag(A\mathbf{1}_n)$.

$$\mathbb{E}\left[H_{\beta_{\mathrm{N}},J}\right] = I_n + \mathbb{E}\left[\frac{\mathrm{th}(\beta J_{ij})}{1-\mathrm{th}^2(\beta J_{ij})}\right](D - A).$$

A $- \, non\ diagonal\ element$ $completly\ defined$ by graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

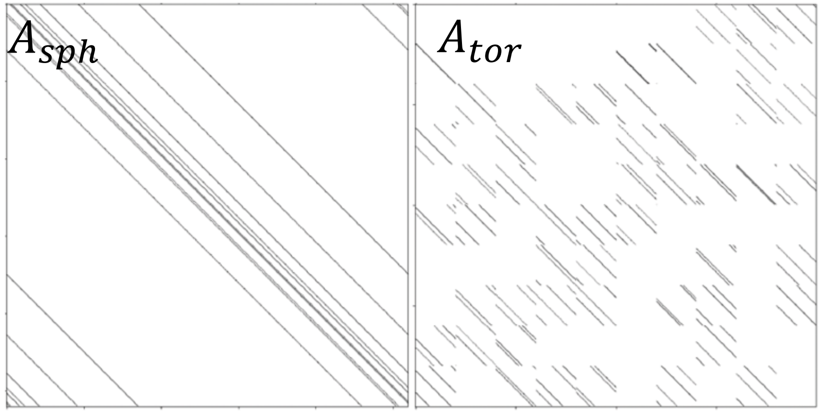
Viktor Dotsenko, Introduction to the Replica Theory of Disordered Statistical Systems *Landau Institute for Theoretical Physics, Moscow, 2001, 219 p*
Mezard M . Parisi G., Virasoro M. Spin Glass Theory and Beyond An Introduction to the Replica Method and Its Applications.1986, 476 p.
H. Nishimori, J. Phys. C: Solid State Phys. 13(21), 4071 (1980) , H. Nishimori, Prog. Theor. Phys. 66(4), 1169 (1981), H. Nishimori, Prog. Theor. Phys. 76, 305 1986
Lorenzo Dall'Amico et al Nishimori meets Bethe: a spectral method for node classification in sparse weighted graphs J. Stat. Mech. (2021) 093405

$$\mathbb{E}[H_{\beta_N,J}] = I_n + \mathbb{E}\left[\frac{th(\beta J_{ij})}{1 - th^2(\beta J_{ij})}\right](D - A).$$

Quasi-Cyclic Sparse Graphs of spin interaction in RBIM

$A\ in\ Bethe - Hessian\ \ defined$ by graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

A **quasi-cyclic Low Density parity-check (QC-LDPC) code** has a sparse parity-check matrix H composed of m×n blocks, each of size *p×p, [1,2]*. Each block is either:
A **zero matrix**, or a **cyclic permutation matrix (CPM)** P^a ($a \in \mathbb{Z}$), defined by cyclically shifting the rows of the identity matrix I_p by a positions.
Exponents a of CPM are element of the cyclic group $\mathbb{Z}/p\mathbb{Z}$.
Such matrix represented by bipartite graph (Tanner-graph).

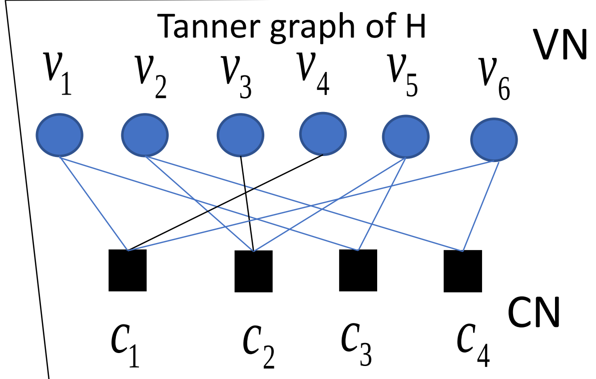


Quasi-Cyclic Spherical Graph Quasi-Cyclic Toric Graph

$p = 2$

$$H = \begin{array}{c} \begin{array}{cccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{array} \\ \left[\begin{array}{cc|cc|cc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \end{array} \begin{array}{c} \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{array}$$

$$H_{QC} = \begin{bmatrix} I^0 & I^1 & I^1 \\ I^0 & I^{-1} & I^0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$



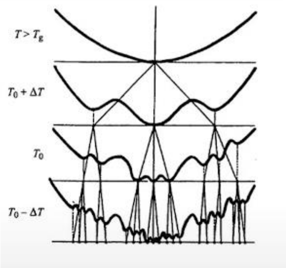
Closed path from variable node form cycle.
Odd degree cycle (imbalance) form basin in energy lands.

Cycle 8:
$v_1 \to c_3 \to v_5 \to c_2 \to v_2 \to c_4 \to v_6 \to c_1 \to v_1$

When we construct [3-6] such QC graphs we optimize cycle (Bethe-permanent [7], pseudocodewords, EMD/ACE) and code (Hamming) distance properties (permanent, [8,9], weight spectrum enumerator, graph perfect match properties).

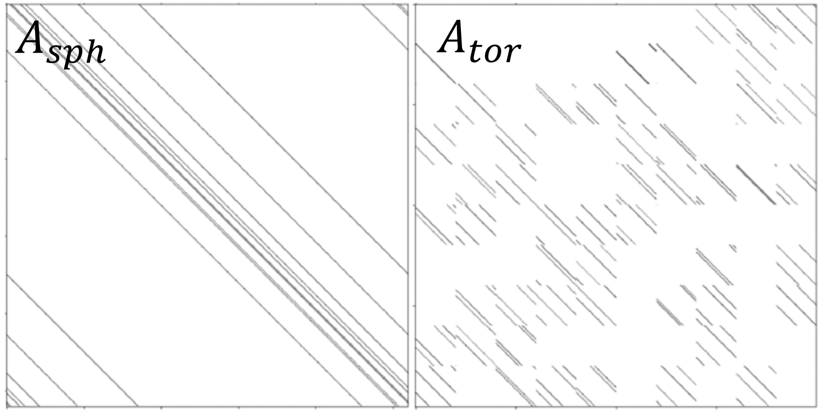
[1] R. M. Tanner, D Sridhara, and T. Fuja, “A class of group-structured LDPC codes,” Proc. ISCTA 2001, 2001. [2] Fossorier M.P.C., “Quasi-cyclic low-density parity-check codes from circulant permutation matrices ”, *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1788–1793, 2004.
[3] Usatyuk, V.S., Sapozhnikov, D.A. & Egorov, S.I. Enhanced Image Clustering with Random-Bond Ising Models Using LDPC Graph Representations and Nishimori Temperature. *Moscow Univ. Phys.* **79** (Suppl 2), S647–S665 (2024).
[4] V. Usatyuk and I. Vorobyev, "Simulated Annealing Method for Construction of High-Girth QC-LDPC Codes," *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, Athens, Greece, 2018, pp. 1-5
[5] V. Usatyuk and I. Vorobyev, "Construction of High Performance Block and Convolutional Multi-Edge Type QC-LDPC codes," *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, Budapest, Hungary, 2019
[6] V. S. Usatyuk and S. I. Egorov, "Mixed Integer Linear Programming Method for Energy-Based Model Trapping Sets Enumerating," *2024 26th International Conference on Digital Signal Processing and its Applications (DSPA)*, Moscow, Russian Federation, 2024, pp. 1-6
[7] R. Smarandache, "Pseudocodewords from Bethe permanents," *2013 IEEE International Symposium on Information Theory*, Istanbul, Turkey, 2013, pp. 2059-2063
[8] Smarandache R., and Vontobel P.O., “Quasi-cyclic LDPC codes: influence of proto- and Tanner-graph structure on minimum hamming distance upper bounds ”, in *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 585–607, Feb. 2012.
[9] P. O. Vontobel, "Counting in Graph Covers: A Combinatorial Characterization of the Bethe Entropy Function," in *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 6018-6048

Nishimori Temperature approximation for Natural images



The hyerarchical tree of spin-glass states

To perform accurate node clustering, knowledge of the Nishimori temperature β_{N} is crucial for obtaining a precise estimate of the true node classes σ. A powerful estimator of σ is derived from the signs of the entries of the eigenvector $\mathbf{x}$ of the Bethe-Hessian matrix $H_{\beta_{\mathrm{N}},\tilde{J}}$, associated with its smallest amplitude eigenvalue, which is close to zero.



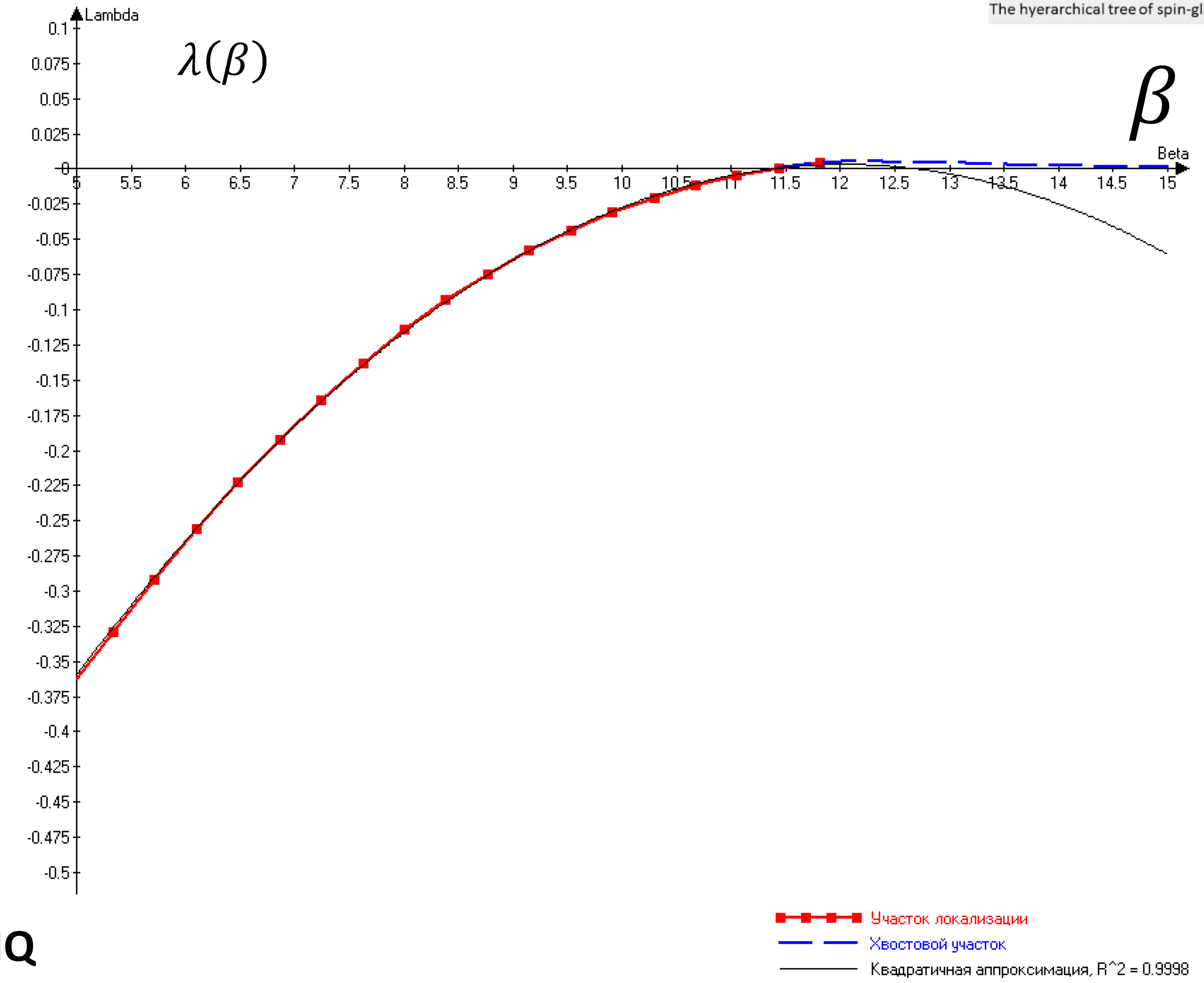
Quasi-Cyclic Spherical Graph   Quasi-Cyclic Toric Graph

Input: Initial affinity matrix A_{initial}, mean c of the adjacency matrix, and initial approximation $\beta_{\text{initial}} = \frac{1}{c}$.
Define: β_{SG} from the equation $c(\overline{A}_{\beta_{\text{SG}}})^2 - 1 = 0$ and β_L : $\lambda_{\perp}(L(J_{\beta_L})) \to 0_+$.
Procedure:

1: Initialize exact points: $\beta_1 = \beta_L - \beta_{\text{SG}} < \beta_2 < \beta_3 < \beta_4 < \beta_5 = \beta_L$.
2: Compute eigenvalues $\lambda_{\perp}(L(J_{\beta_i}))$ for $i = 1, \ldots, 5$.
3: Build the Lagrange polynomial: $\sum_{j=0}^{4} a_j \beta^j$.
4: Solve for root β^* that minimizes $|\beta_L - \beta^*|$.
5: Update $\beta_L = \beta^*$.

Output: $\beta_N = \beta^*$.

λ



ImageNET
Flickr-Faces-HQ

Stochastic Block Model [1]

SBM(n, p, Θ)

Graph $G = (V, E)$
n vertices: $V = \{1, \ldots, n\}$

Community generation
$p = (p_1, \ldots, p_k)$ prob. vector
$X \sim p$ community label
$C_i = \{u \in V \mid X_u = i\}$

Edge set sampling
Connectivity: $\Theta \in [0,1]^{k \times k}$ symmetric
$\mathbb{P}(u \sim v) = \Theta_{i,j}$ if $u \in C_i$ and $v \in C_j$

Special cases
1. If $P_{i,j} = \tilde{p}, \forall i,j \Rightarrow$ Erdős–Rényi $G(n, \tilde{p})$
2. $P = \begin{pmatrix} p_1 & & p_2 \\ & \ddots & \\ p_2 & & p_1 \end{pmatrix} \Rightarrow$ Planted partition

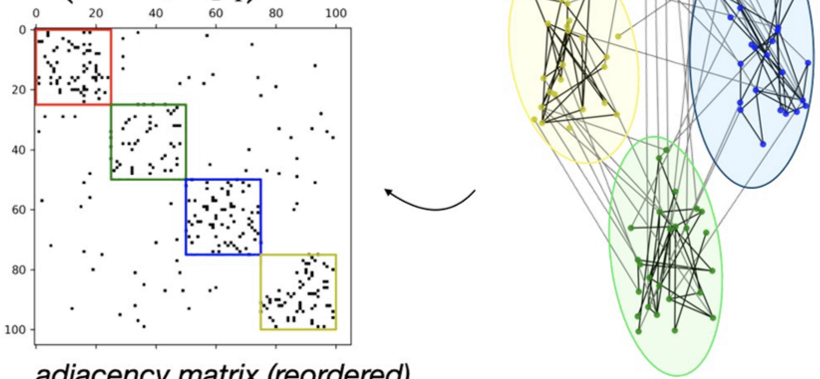
Social Networks 5 (1983) 109–137
North-Holland

STOCHASTIC BLOCKMODELS: FIRST STEPS *

Paul W. HOLLAND
Educational Testing Service **

Kathryn Blackmond LASKEY and Samuel LEINHARDT
Carnegie-Mellon University †

(Sometimes the community label X is taken deterministic)



adjacency matrix (reordered)

S. Fortunato, D. Hric / Physics Reports 659 (2016) 1–44 37

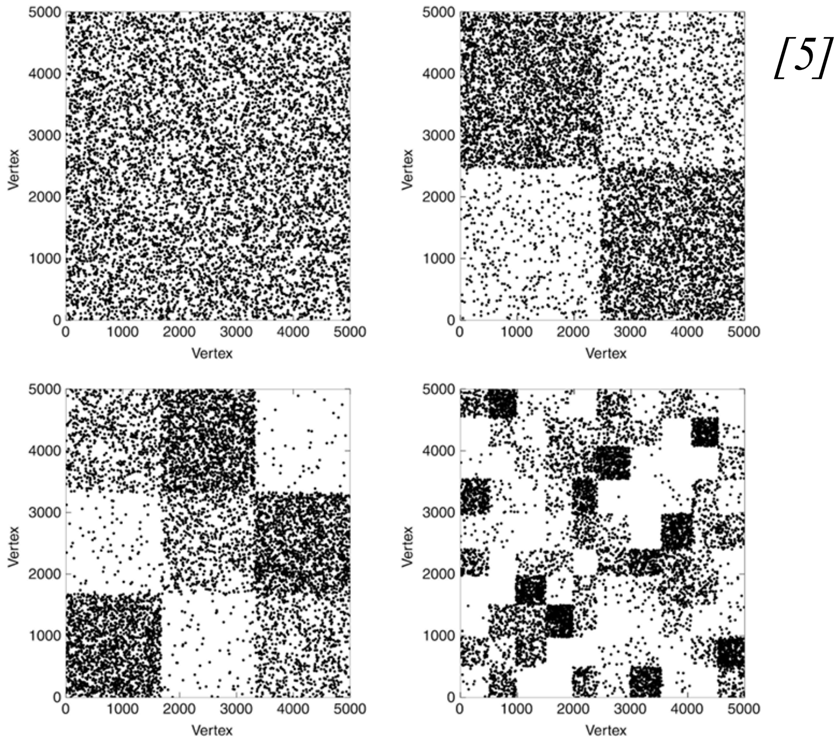


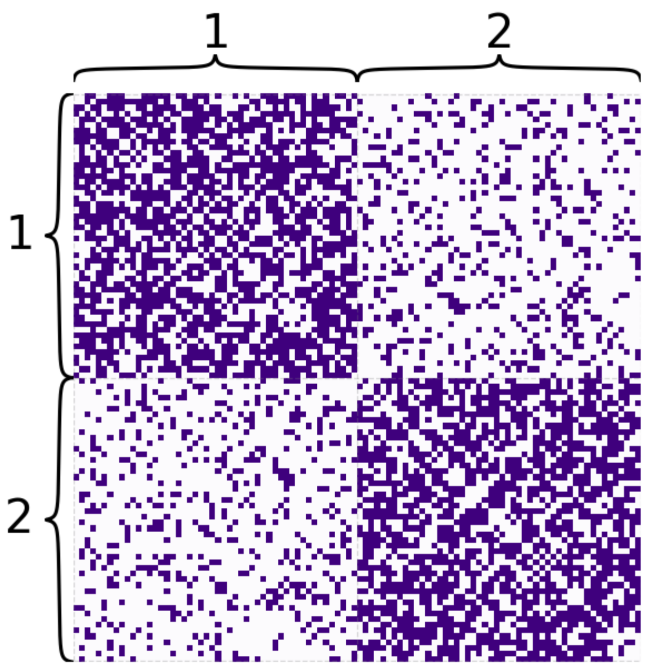
Fig. 29. Artificial groups in random networks. The elements of the adjacency matrix of an Erdős-Rényi random graph (top left) can be rearranged, by suitably reshuffling the list of vertices. This procedure may produce a block structure, with the blocks becoming increasingly more visible the smaller their size. Such groups are not real, though, but they are generated by random fluctuations in the edge patterns among the vertices. The construction principle of the network does not give any special role to groups of vertices, since all pairs of vertices have identical probability to be joined.
Source: Courtesy by Tiago P. Peixoto.

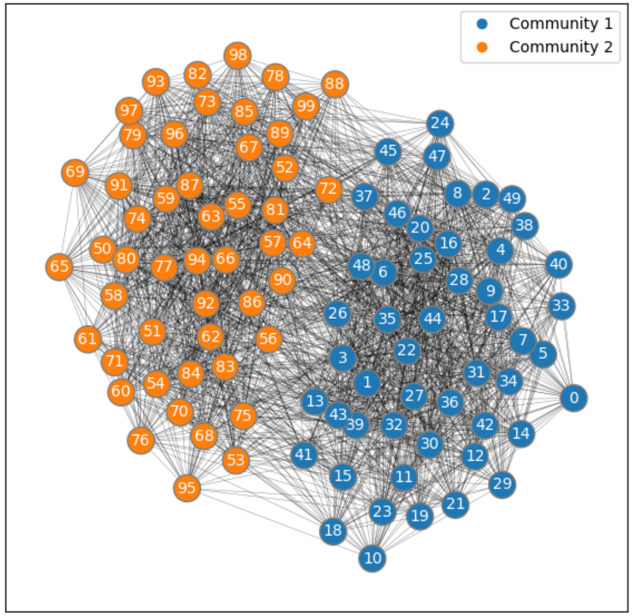
The RBIM serves as the fundamental statistical physics "graph" model for analyzing SBM's phase transitions and inference limits, [3,4]. **SBM ↔ RBIM Mapping**: Community labels (SBM) ≡ Ising spins (RBIM). **Phase Transition Universality**: Detectability threshold (SBM) = Spin-glass transition (RBIM). **Statistical Physics Foundation**: Both models share replica symmetry breaking (cavity and etc) and belief propagation (message passing) analysis [2].

[1] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, June 1983.
[2] Mézard M., A. Montanari **Information, Physics, and Computation,** 2009, 569 p. [3] Decelle, A., Krzakala, F., Moore, C., & Zdeborová, L. (2011). *Inference and Phase Transitions in the Detection of Modules in Sparse Networks. Physical Review Letters, 107(6).* [4] Zdeborová, L., & Krzakala, F. (2016). *Statistical physics of inference: thresholds and algorithms. Advances in Physics, 65(5), 453–552.* [5] Fortunato, S., & Hric, D. (2016). *Community detection in networks: A user guide. Physics Reports, 659, 1–44.*

Nishimori Temperature-Based Node Classification Algorithm

$SBM_n(z, B)$ Simulation





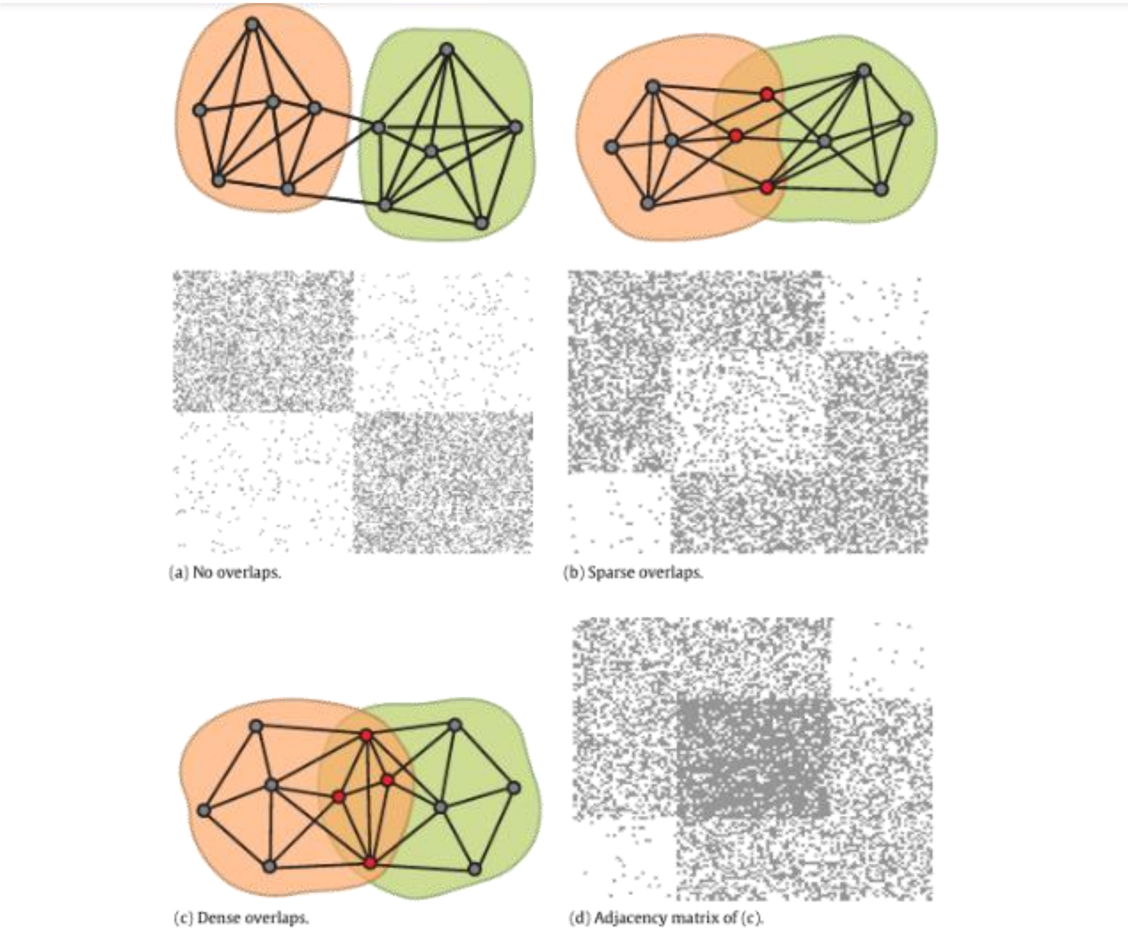


Fig. 20. Stylised views of community structure. In (a) and (b) we show the conventional pictures of non-overlapping and overlapping clusters, respectively. Under the network diagrams we see the corresponding adjacency matrices. The overlaps have a lower edge density than the rest of the communities. The analysis by Yang and Leskovec suggests that a more realistic model could be the one shown in (c, d), where the overlaps are denser than the non-overlapping parts.
Source: Reprinted figure with permission from [20].

http://docs.neurodata.io/graph-stats-book/representations/ch5/single-network-models_SBM.html
Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, June 1983.
Emmanuel Abbe. Community Detection and Stochastic Block Models. *arXiv*, March 2017. arXiv:1703.10146

Nishimori Temperature-Based Node Classification Algorithm

Input: Test sample $x_j \in \overline{X}^{\text{test}}$, embedding weights w_i, discriminatory features Σ_i.

for $i = 1$ to k **do**

  Compute embedding value: $e_i^j = w_i \cdot x_j$.

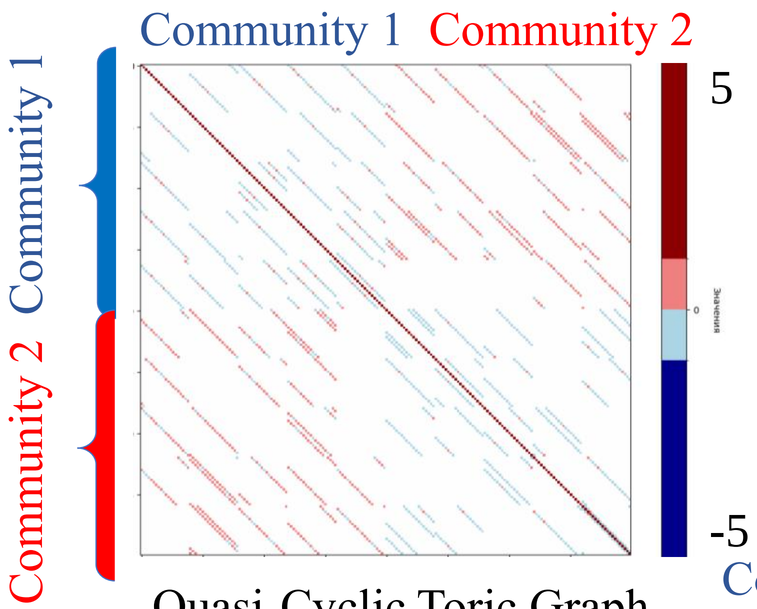
  Estimate probabilities using Bayes' theorem: $P_i(H_1^i \mid x_j) = \frac{P_i(x_j|H_1^i)P_i(H_1^i)}{P_i(x_j)}$.

end for

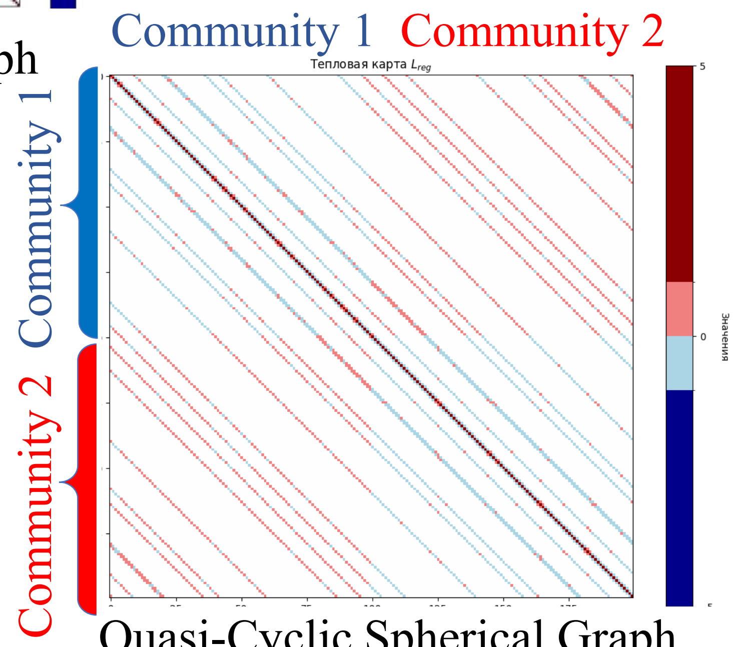
Normalize probabilities: $\overline{P}_i = \text{Sigmoid}(P_i) = \frac{\exp(P_i)}{\sum_{j=1}^{k}\exp(P_j)}$.

Assign predicted class: $\widehat{y}_j = \arg\max_i \overline{P}_i$.

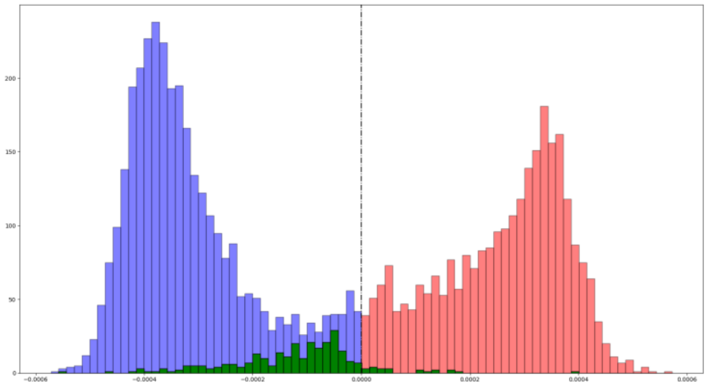
Output: Predicted class $\widehat{y}_j$.

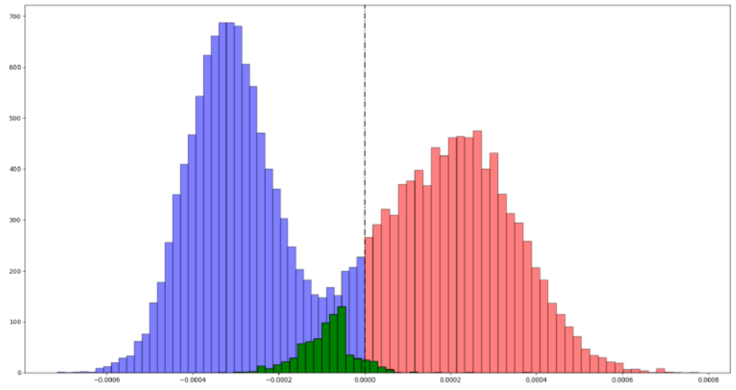


Quasi-Cyclic Toric Graph

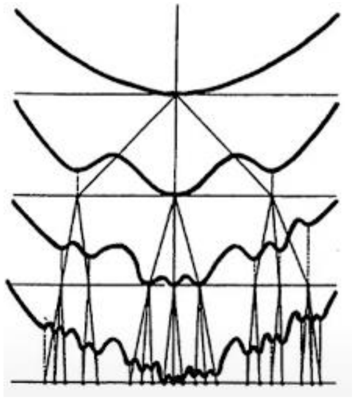


Quasi-Cyclic Spherical Graph





One graph not enough to archive* fraction dimension of Natural image manifold



Finite size(limit our graph structures, graph negative curvature – bottleneck bridges and community), different topology invariants extracted by different graphs (ex. high and low freq. comp), metric learning - cosine metric not enough (large number of classes, coarse feature).

Hard ensembles of graph, with expert:

Discrete labels $\hat{y}_1, \hat{y}_2, \hat{y}_3$ from base classifiers graphs.

$$\hat{y}_{res} = \begin{cases} \hat{y}_{i,} \text{ unanimous agreement} \\ Arbitor(\hat{y}_i), majority\ vote \end{cases}$$

Hard ensembles overcome limitations of single-graph methods near phase transitions (e.g., detectability thresholds in SBMs), where spectral classifiers exhibit high variance.

Stacked Soft Graph Ensemble:

Class probability vectors p_i from n −graph classifiers.

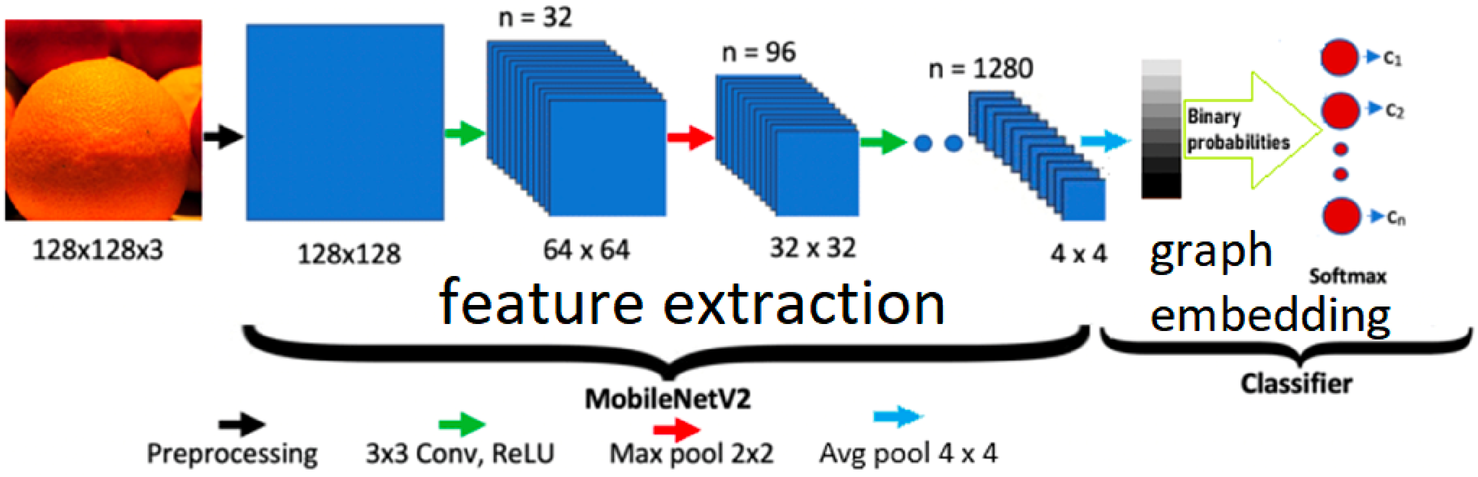
Step 1. Stacked classes probability p_i, $\boldsymbol{x_{meta}} = [p_i]$

Step 2. Train ($\boldsymbol{x_{meta}}$, $\hat{y}_{true}$) a **meta-classifier** using MLP, KRR, RF, GBM (LightGMB/XGBoost).

*Topology Surgery of Data Manifold using Bethe-Hessian graph

Opitz, D., & Maclin, R. (1999). Popular Ensemble Methods. Journal of Artificial Intelligence Research 11 (1999) 169-198

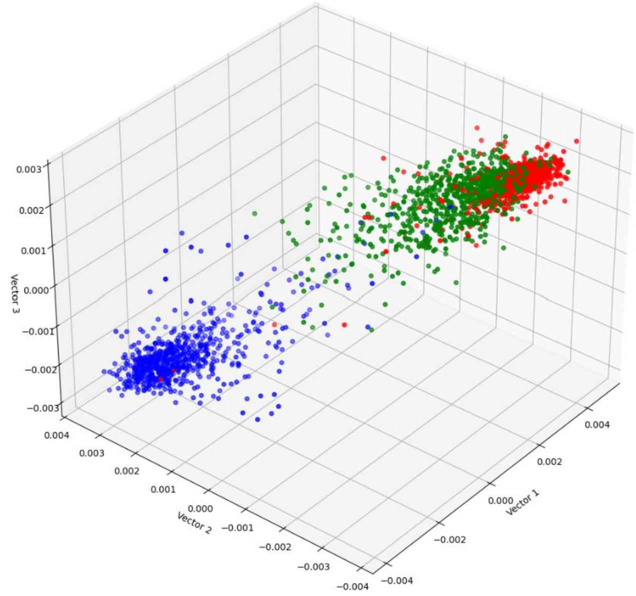
R. M. O. Cruz, R. Sabourin and G. D. C. Cavalcanti, "On Meta-learning for Dynamic Ensemble Selection," *2014 22nd International Conference on Pattern Recognition*, Stockholm, Sweden, 2014, pp. 1230-123

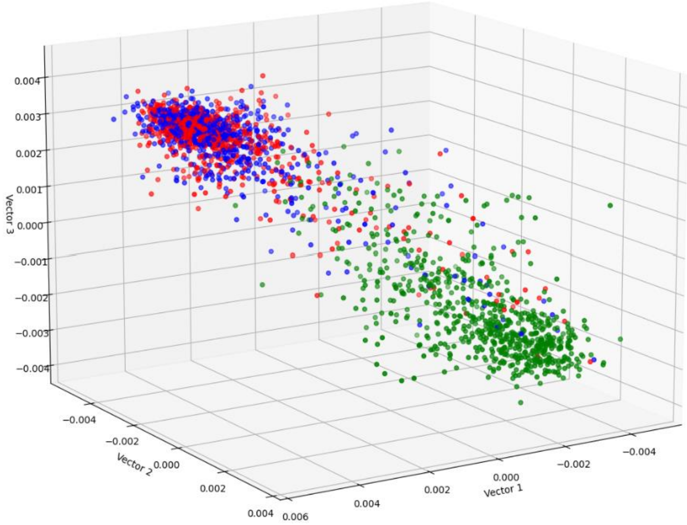


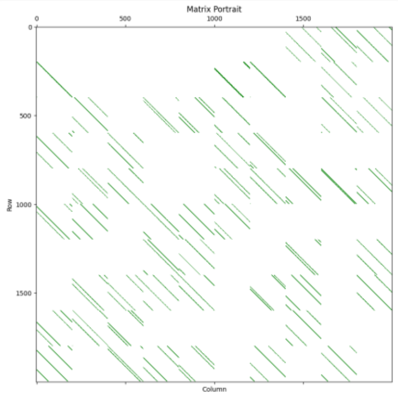
k Classes	10	100
M^{train} dataset	10000	130000
M^{test} dataset	3000	5000
m^{train} Embedding per class	1000	1300
m^{test} Embedding per class	300	50
S Features	1280	1792
s Embedded features	32	64

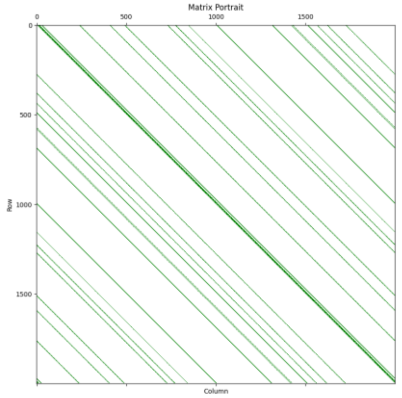
Class	Precision	Recall	F1-Score	Support
0	1	0.98	0.99	300
1	0.99	0.99	0.99	300
2	1	0.99	0.99	300
3	0.99	0.94	0.96	300
4	0.97	0.97	0.97	300
5	0.99	0.99	0.99	300
6	0.94	0.95	0.95	300
7	0.98	0.98	0.98	300
8	0.94	0.98	0.96	300
9	0.96	1	0.98	300
Total				
Accuracy			0.98	3000
Macro Avg	0.98	0.98	0.98	3000
Weighted Avg	0.98	0.98	0.98	3000

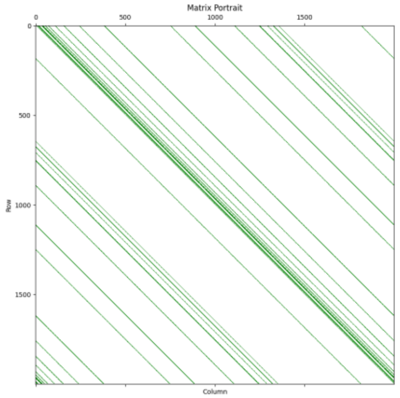
TABLE I: Classification results for 32-dimensional graph embedding on the ImageNet10 dataset.

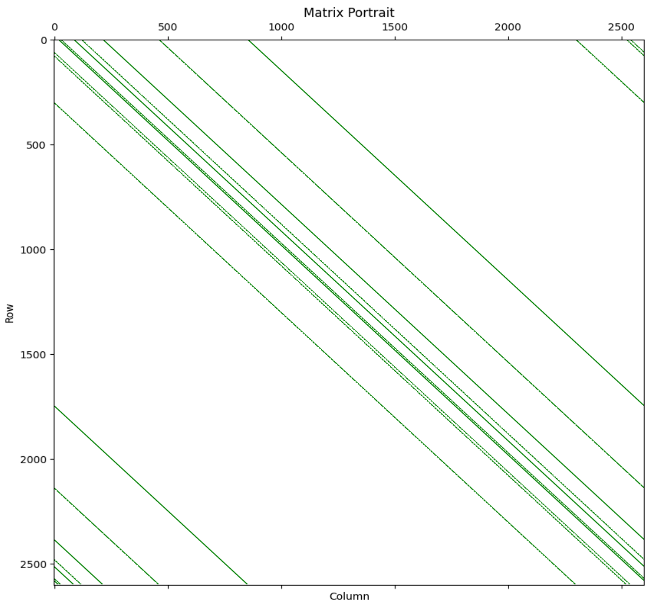


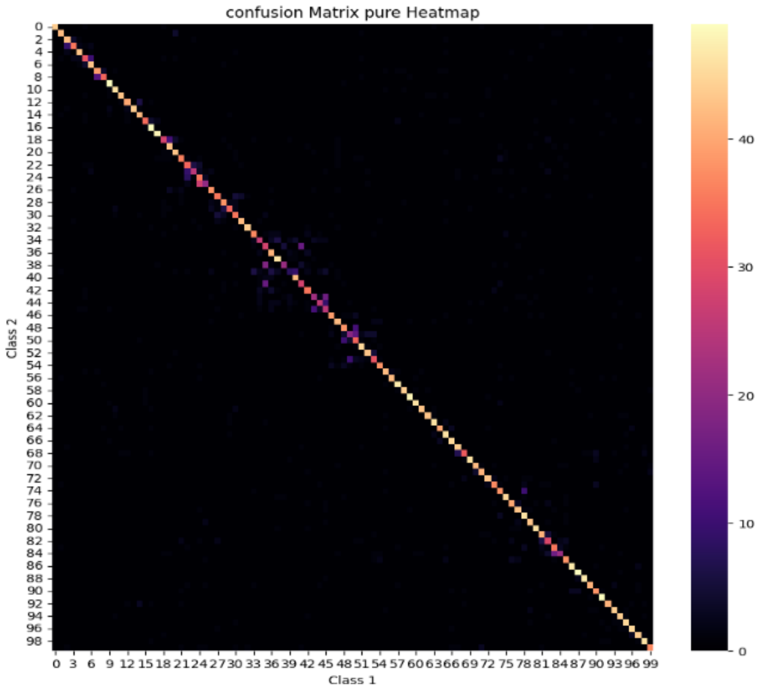




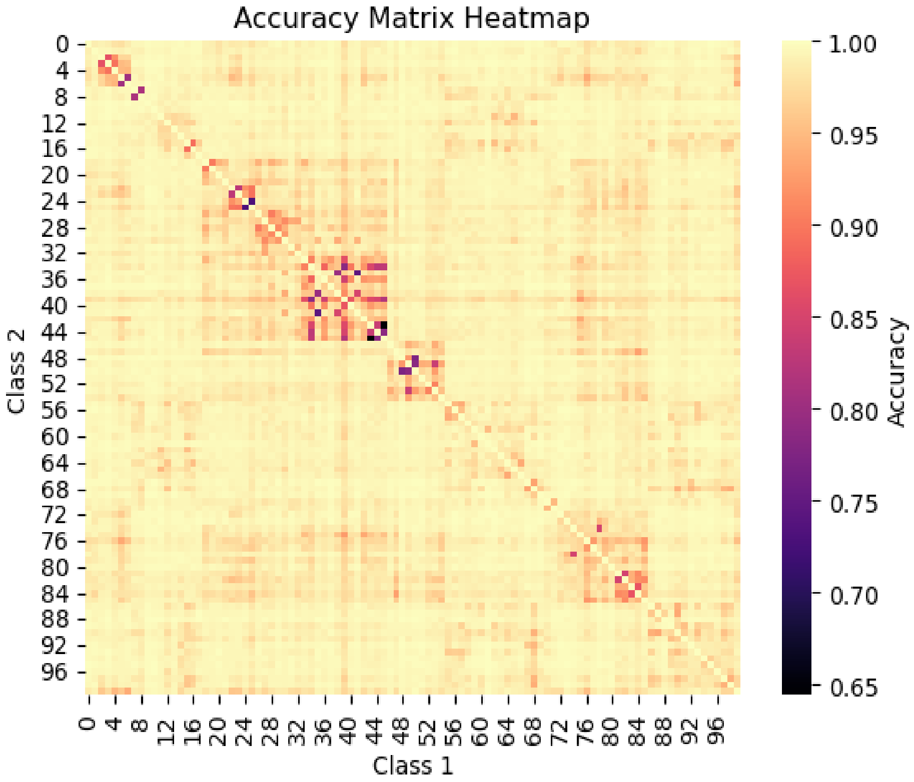








Imagenet100



Hard ensembles 3 Accuracy 82.73%.

$$y^{\text{Arbitrator}} = \begin{cases} y^{1st}, P_{1st} - P_{2nd} > T \\ \text{Arbitrator}\left(y^{1st}, y^{2nd}, y^{3rd}\right), \text{otherwise} \end{cases}$$

Let us set a probability threshold T; if the difference between the most probable and the second most probable solutions is less than this threshold, we will submit the three most probable classes to the arbitrator for decision.

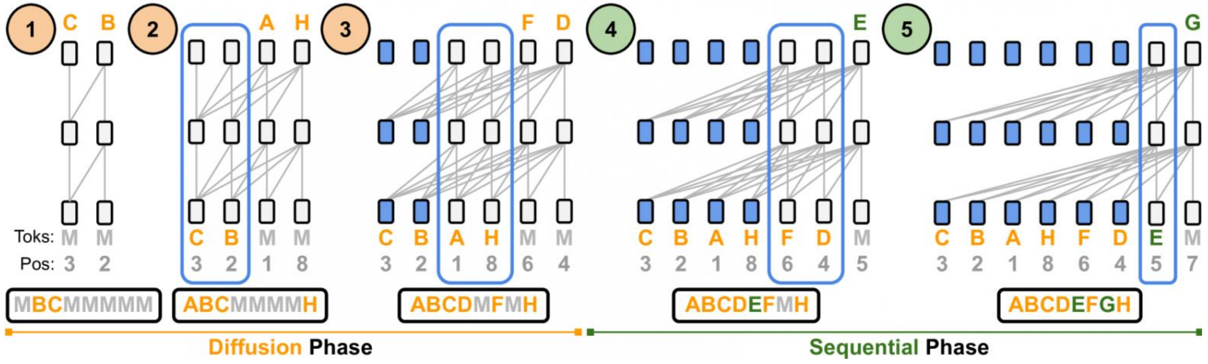
Soft MLP ensemble 87.9% accuracy.

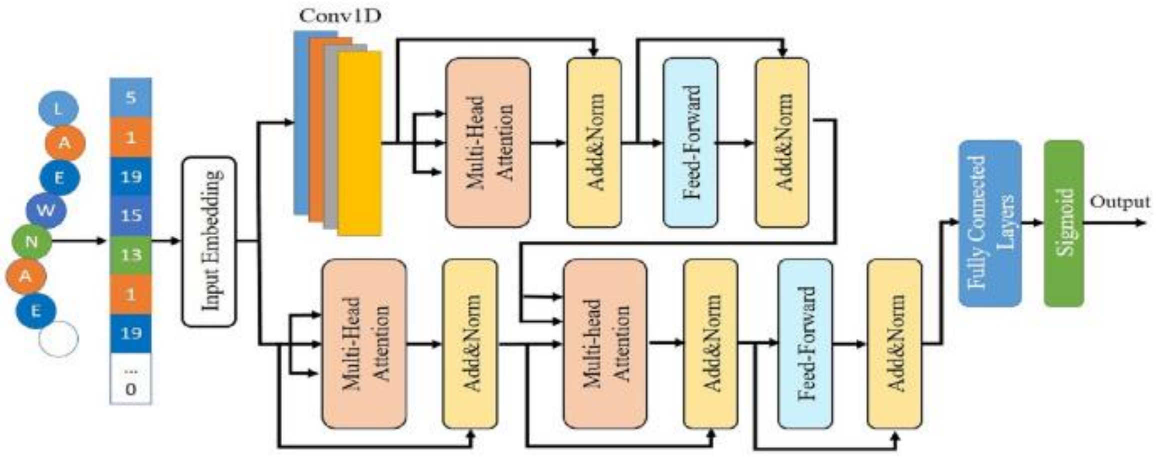
Summary

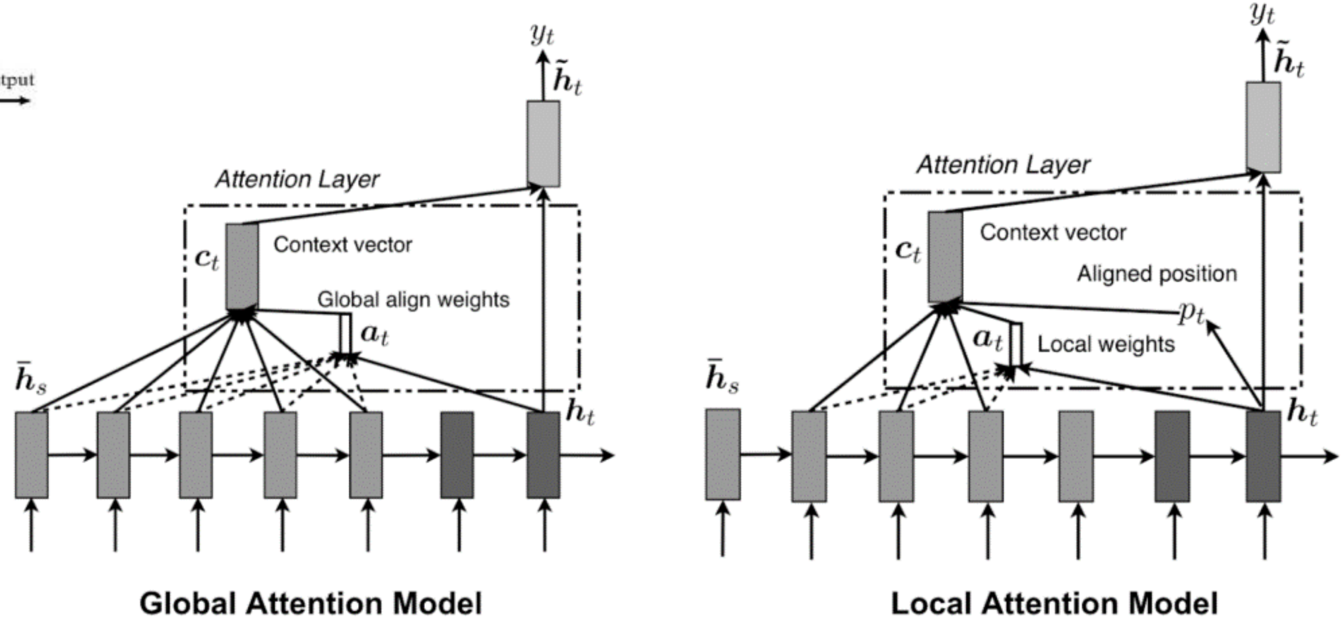
We proposed

1. RBIM QC sparse graph Stochastic Block models (sphere and toric families) for spectral embedding
2. approximation of Bethe-Nishimori temperature and check it under ImageNET (10K in progress)
3. decrease number of requirement feature from 1280(1720) to 32/64 respectively under Mobilenetv2 CNN for Natural image classification;
4. hard and soft Ensemble of QC sparse graphs approach.

Thank You For Your Attention!



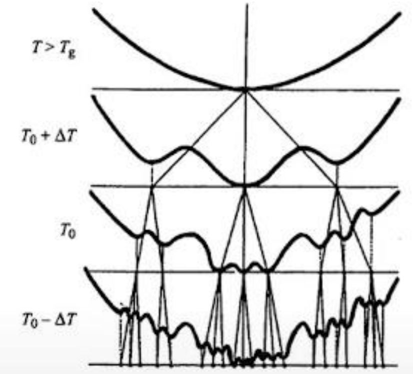




Hierarchical tree (Negative Curvature topology of Graph) of spin glass state

Def. 1 (Weighted non-backtracking matrix) *Given a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$ and a function $f:\mathcal{E}\to\mathbb{R}$, so that $\forall e\in\mathcal{E}$, $f(e)=\omega_e$ is the weight corresponding to the edge e, the weighted non backtracking matrix $B\in\mathbb{R}^{2|\mathcal{E}|\times 2|\mathcal{E}|}$ is defined on the set of directed edges of $\mathcal{G}$ as*

$$B_{(ij),(k\ell)}=\delta_{jk}(1-\delta_{i\ell})\omega_{k\ell}.$$



The hyerarchical tree of spin-glass states

Local (hierarchical) Hessian

Def. 2 (Bethe-Hessian matrix) *Given a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, a function $f:\mathcal{E}\to\mathbb{R}$ so that $\forall e\in\mathcal{E}$, $f(e)=\omega_e$ and a parameter $x\in\mathbb{C}\backslash\{\pm\omega_{ij}\}_{(ij)\in\mathcal{E}}$, the Bethe-Hessian matrix $H(x)\in\mathbb{C}^{n\times n}$ is defined as*

$$H_{ij}(x)=\left(1+\sum_{k\in\partial i}\frac{\omega_{ik}^2}{x^2-\omega_{ik}^2}\right)\delta_{ij}-\frac{x\omega_{ij}}{x^2-\omega_{ij}^2}.$$

Since $\mathcal{G}$ is an undirected graph, $H(x)$ is symmetric but not Hermitian, unless $x\in\mathbb{R}$. The relation between the spectra of the matrices B and $H(x)$ is given by the Watanabe-Fukumizu formula.

Prop. 1 (Watanabe-Fukumizu) *Let $H(x)$ and B on the same graph $\mathcal{G}$ and for the same weighting function f. $x\in\mathbb{C}\backslash\{\pm\omega_{ij}\}_{(ij)\in\mathcal{E}}$:*

$$\det[xI_{2|\mathcal{E}|}-B]=\det[H(x)]\prod_{(ij)\in\mathcal{E}}(x^2-\omega_{ij}^2),$$

x in the spectrum of B, $\det[H(x)]=0$.

Yusuke Watanabe and Kenji Fukumizu. Loopy belief propagation, bethe free energy and graph zeta function. arXiv preprint arXiv:1103.0605, 2011
Iwao Sato, Hideo Mitsuhashi, and Hideaki Morita. A matrix-weighted zeta function of a graph. Linear and Multilinear Algebra, 62(1):114–125, 2014.
A. Saade, F. Krzakala, and L. Zdeborova, ``Spectral clustering of graphs with the Bethe Hessian,'' NIPS, 2014.
F. Krzakala, et al., Spectral redemption in clustering sparse networks, Proc. Natl. Acad. Sci. U.S.A. 110 (52) 20935-20940

Nishimori Temperature-Based Node Classification Algorithm

To perform accurate node clustering, knowledge of the Nishimori temperature β_{N} is crucial for obtaining a precise estimate of the true node classes σ. A powerful estimator of σ is derived from the signs of the entries of the eigenvector $\mathbf{x}$ of the Bethe-Hessian matrix $H_{\beta_{\mathrm{N}},\tilde{J}}$, associated with its smallest amplitude eigenvalue, which is close to zero.First, consider $\mathbf{y}$, the eigenvector associated with the smallest eigenvalue of $H_{\beta_{\mathrm{N}},J}$. Let $A \in \{0,1\}^{n\times n}$ be the symmetric adjacency matrix of $\mathcal{G}$, defined by $A_{ij} = 1$ if $(ij) \in \mathcal{E}$, and $A_{ij} = 0$ otherwise. Let $D \in \mathbb{N}^{n\times n}$ be the diagonal degree matrix $D = \mathrm{d}iag(A\mathbf{1}_n)$.

$$\mathbb{E}\left[H_{\beta_{\mathrm{N}},J}\right] = I_n + \mathbb{E}\left[\frac{\mathrm{th}(\beta J_{ij})}{1-\mathrm{th}^2(\beta J_{ij})}\right](D - A).$$

Alg. 2 The Nishimori-Bethe relation for node classification

Input: Weighted adjacency matrix of a graph $\tilde{J} \in \mathbb{R}^{n\times n}$, precision error $\varepsilon \in \mathbb{R}$
Return:Value of $\hat{\beta}_{\mathrm{N}} \in \mathbb{R}^{+}$, estimated label vector $\hat{\sigma} \in \{-1,1\}^n$

- Shift the non-zero $\tilde{J}_{ij}$ as: $\tilde{J}_{ij} \leftarrow \tilde{J}_{ij} - \frac{1}{2|\mathcal{E}|}\mathbf{1}_n^T\tilde{J}\mathbf{1}_n$
- Compute $\hat{\beta}_{\mathrm{N}} \leftarrow$ Compute_$\hat{\beta}_{\mathrm{N}}$
- Compute $H_{\beta_t,J} = \delta_{ij}\left(1 + \sum_{k\in\partial i}\frac{\mathrm{th}^2(\beta J_{ik})}{1-\mathrm{th}^2(\beta J_{ik})}\right) - \frac{\mathrm{th}(\beta J_{ij})}{1-\mathrm{th}^2(\beta J_{ij})}$.
- Compute $\mathbf{x} \leftarrow$ the eigenvector associated to $\gamma_{\min}(H_{\hat{\beta}_{\mathrm{N}},\tilde{J}})$
- Estimate $\hat{\sigma}$ as the output of 2-class *k-means* on the entries of $\mathbf{x}$

return: β_t, $\hat{\sigma}$.

Lorenzo Dall'Amico et al Nishimori meets Bethe: a spectral method for node classification in sparse weighted graphs J. Stat. Mech. (2021) 093405

Simulated Annealing method for Quasi-Cyclic LDPC bipartite graph construction

Algorithm 1 Simulated Annealing method for construction of QC-LDPC codes

Require: $M(\mathbf{H})$–mother matrix, L– circulant size, g-girth of lifted matrix, EMD-minimal EMD value, $Iter$– maximal number of iterations, $seed$- a seed to be used in a pseudo-random number generator, $Temp$– initial value of temperature

$Nstep = 0$
$i, j = rnd(seed)$
for $it = 0; it <= Iter; it = it + 1$ **do**
    while $M_{ij}(\mathbf{H}) = 0$ **do**
        $i, j = rnd(seed)$
    end while
    for $k = 0; k <= L - 1; k = k + 1$ **do**
        $\Theta_k = enumcirccycles(i, j, k, g, EMD)$

$$P(k) = w(k) / \sum_{m=0}^{L-1} w(m),$$

$$w(k) = e^{\frac{-\Theta_k}{Temp}},$$

where Θ_k-number of cycles through $E_{ij}(\mathbf{H})$-CPM with shift value k, $P(k)$– probability of k-shift CPM value choice, $w(k)$–probability weight function;
    end for

$$\Phi = enumcycles(E(\mathbf{H}), g, EMD),$$

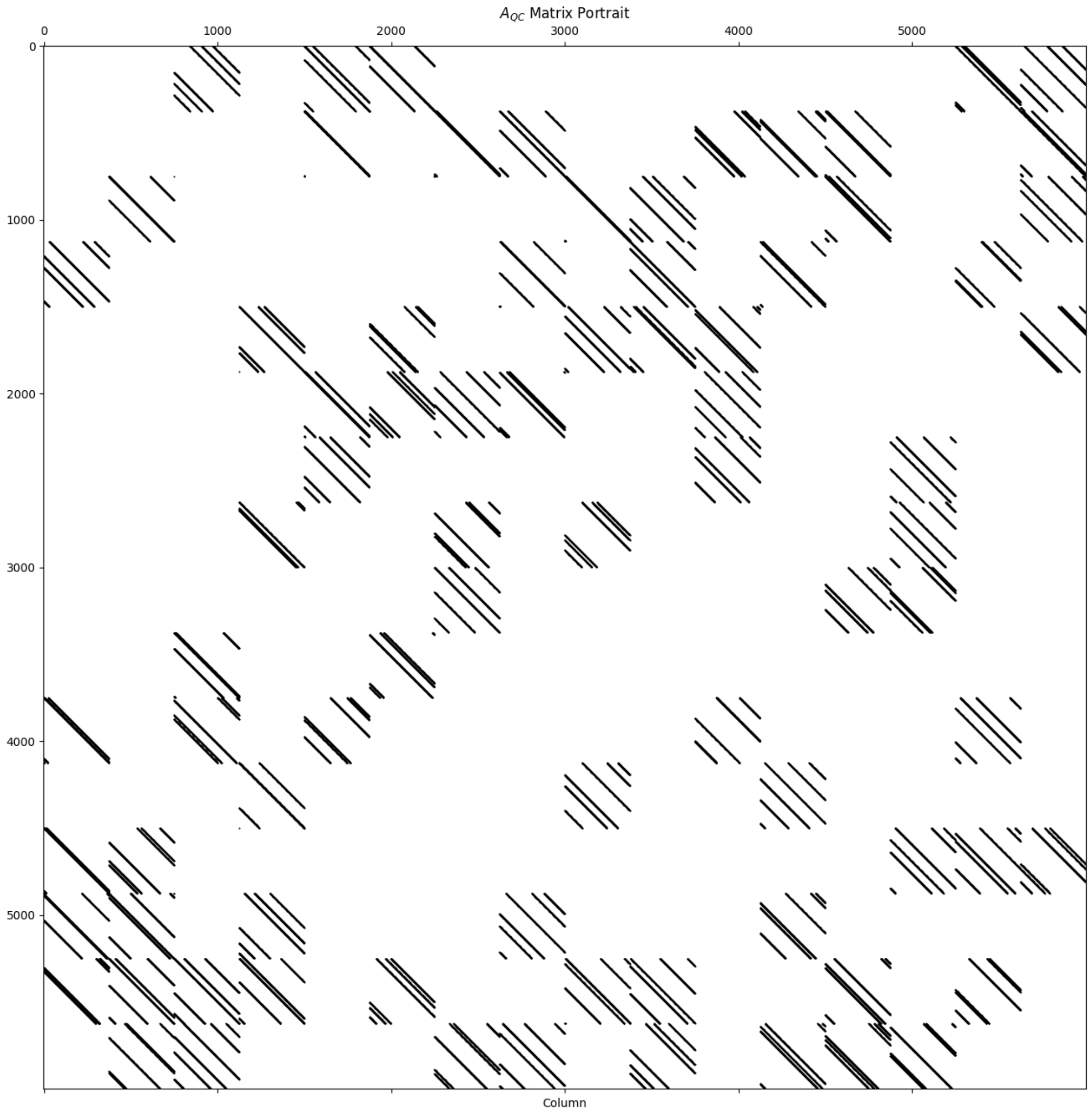
where Φ-total number of cycles in exponent matrix $E(\mathbf{H})$;

$$E_{ij}(\mathbf{H}) = rndshift(P, Temp)$$

    Nstep=Nstep+1

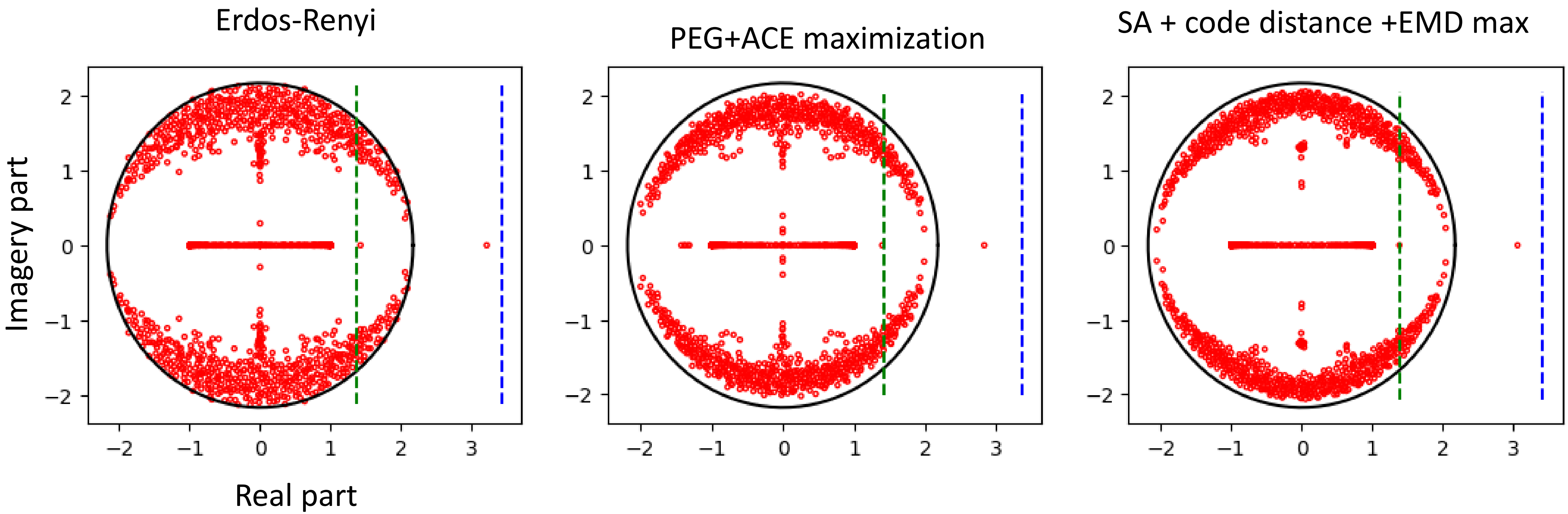
$$Temp = \eta \frac{\Phi}{Nstep^2},$$

where η-some constant value;
end for
    return $E(\mathbf{H})$



V. Usatyuk and I. Vorobyev, "Simulated Annealing Method for Construction of High-Girth QC-LDPC Codes," *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, Athens, Greece, 2018, pp. 1-5

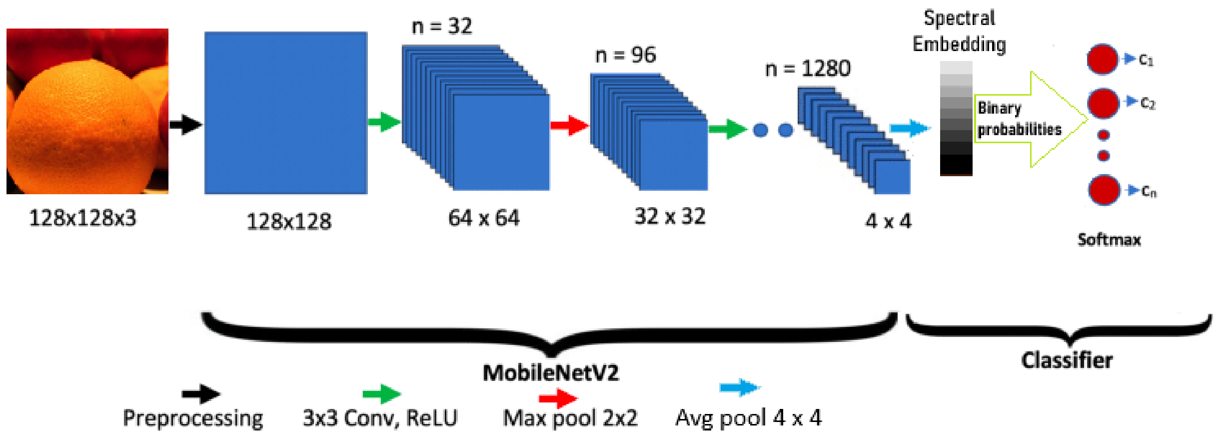
Spectrum of the weighed matrix B

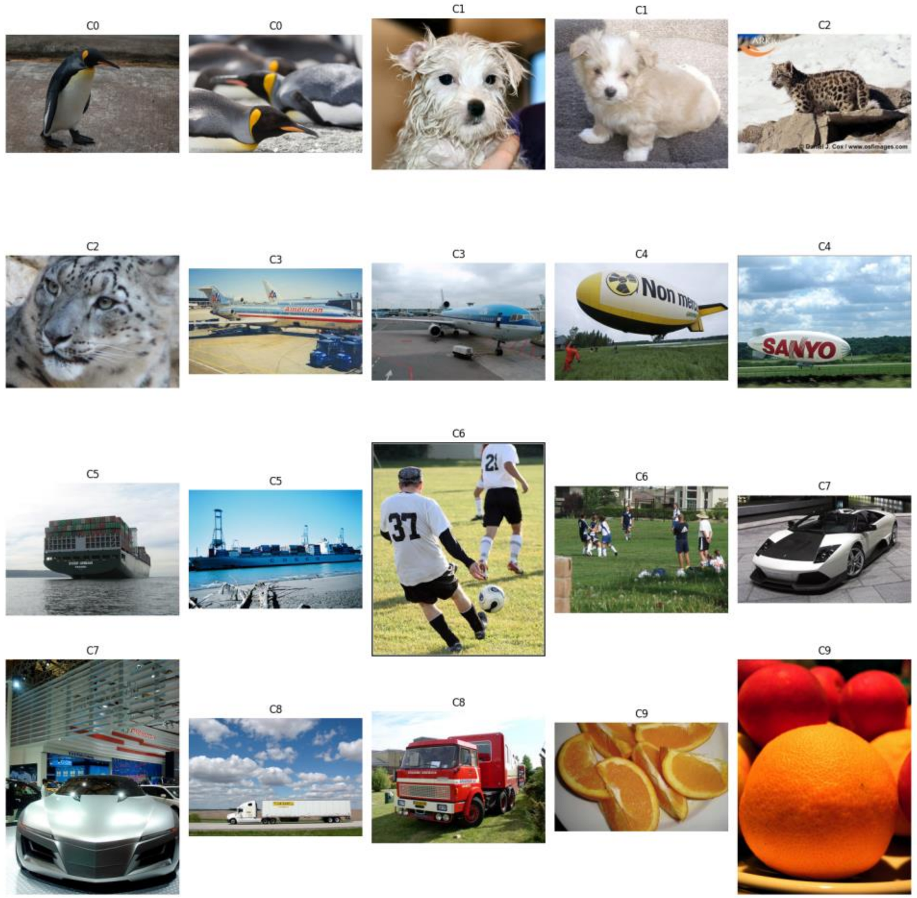


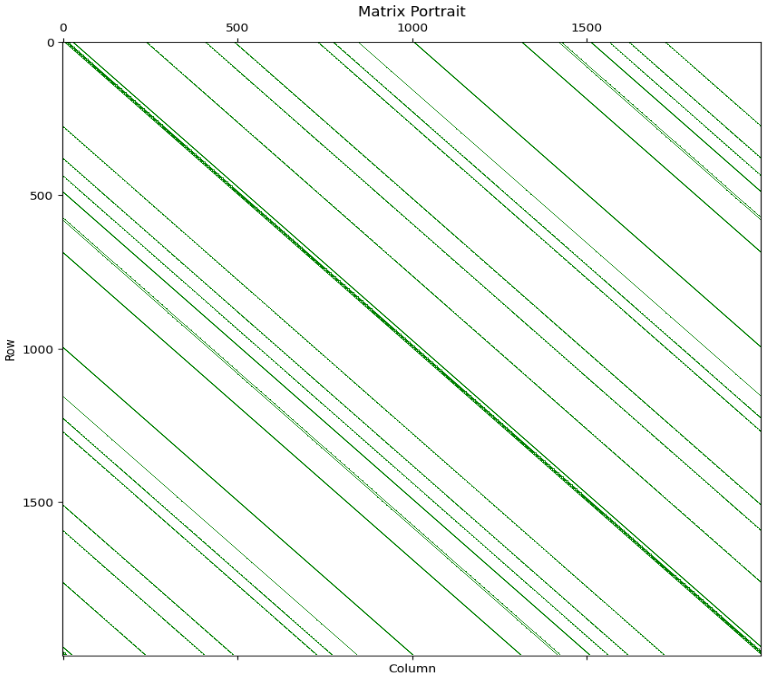
K classes	10	100	1000
Train dataset	10000	130000	1281167
Test dataset	3000	5000	100000
Emd train	1000	1300	1281
Emd test	300	50	100
features	1280	1792	2304
Emb features	32	64	128

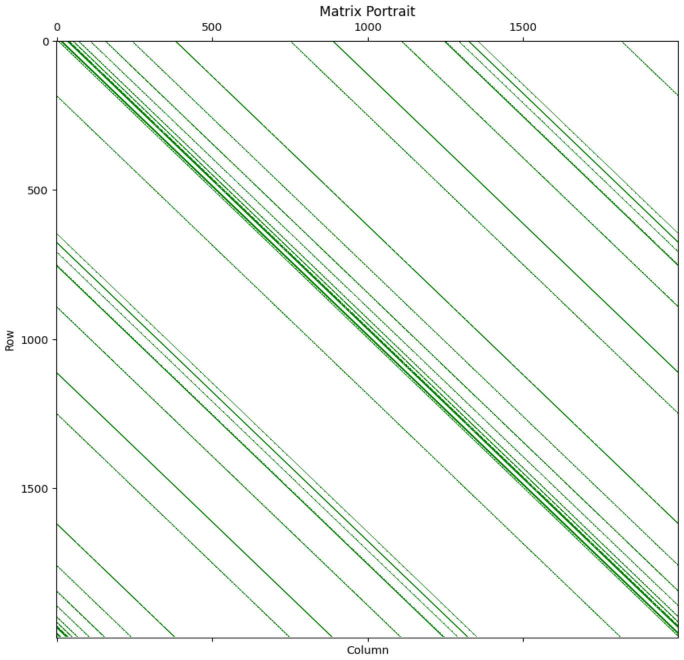
13000 images, 10 classes: "king penguin, Aptenodytes patagonica", "Maltese dog, Maltese terrier, Maltese", "snow leopard, ounce, Panthera uncia", "airliner", "airship, dirigible", "container ship, containership, container vessel", "soccer ball", "sports car, sport car", "trailer truck, tractor trailer, trucking rig, rig, articulated lorry, semi", "orange"

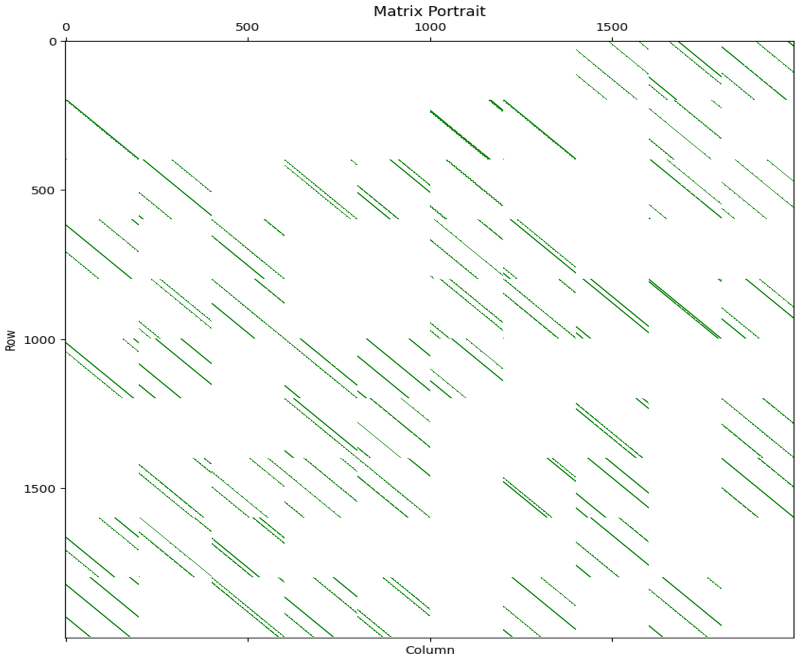
MobileNetV2






Matrix Portrait
0
500
1000
1500
Row
Column

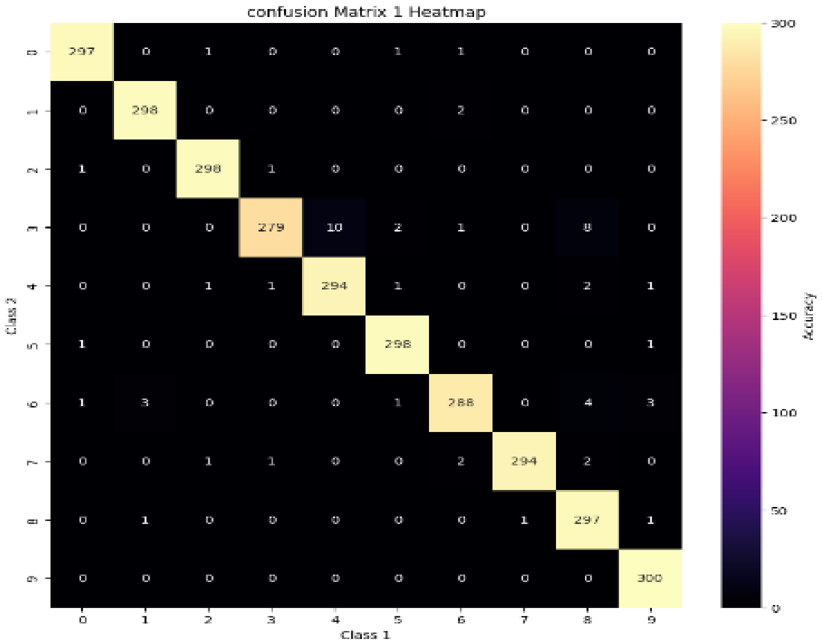

Matrix Portrait
0
500
1000
1500
Row
Column

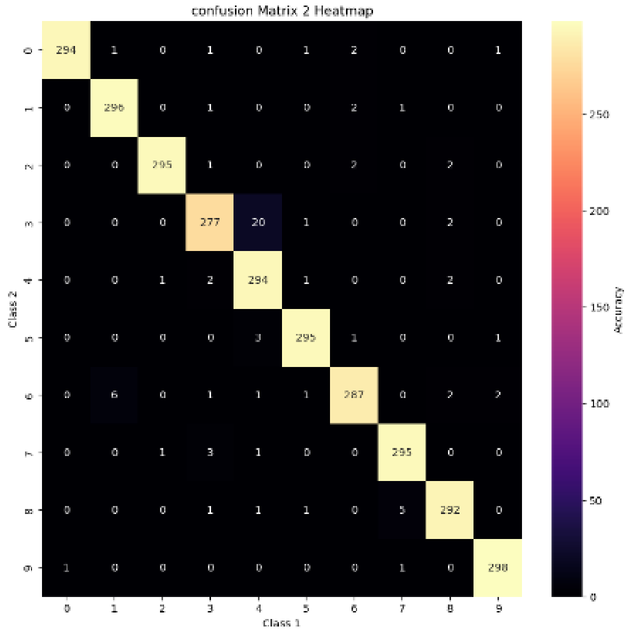

Matrix Portrait
0
500
1000
1500
Row
Column

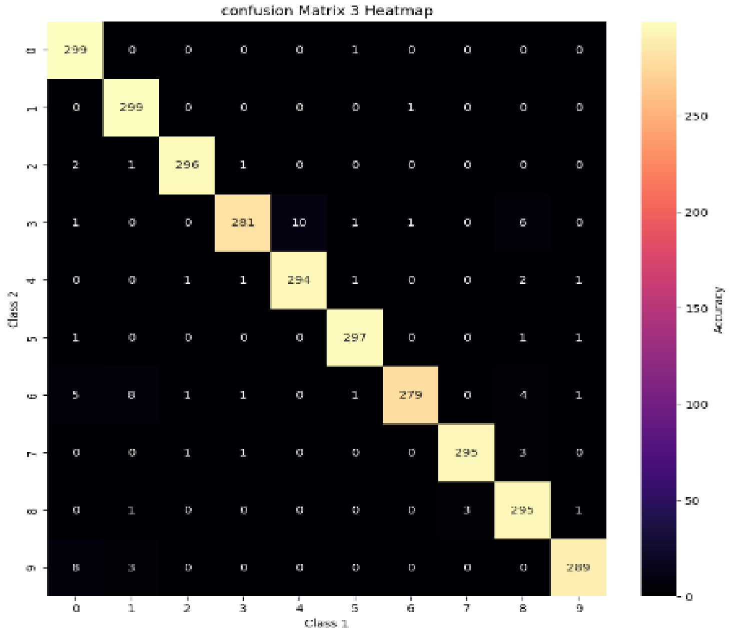
	precision	recall	f1-score	support
0	0.99	0.99	0.99	300
1	0.986755	0.993333	0.990033	300
2	0.990033	0.993333	0.991681	300
3	0.989362	0.93	0.958763	300
4	0.967105	0.98	0.97351	300
5	0.983498	0.993333	0.988391	300
6	0.979592	0.96	0.969697	300
7	0.99661	0.98	0.988235	300
8	0.948882	0.99	0.969005	300
9	0.980392	1	0.990099	300
accuracy	0.981	0.981	0.981	0.981
macro avg	0.981223	0.981	0.980941	3000
weighted avg	0.981223	0.981	0.980941	3000

	precision	recall	f1-score	support
0	0.99661	0.98	0.988235	300
1	0.976898	0.986667	0.981758	300
2	0.993266	0.983333	0.988275	300
3	0.965157	0.923333	0.943782	300
4	0.91875	0.98	0.948387	300
5	0.983333	0.983333	0.983333	300
6	0.97619	0.956667	0.96633	300
7	0.976821	0.983333	0.980066	300
8	0.973333	0.973333	0.973333	300
9	0.986755	0.993333	0.990033	300
accuracy	0.974333	0.974333	0.974333	0.974333
macro avg	0.974711	0.974333	0.974353	3000
weighted avg	0.974711	0.974333	0.974353	3000

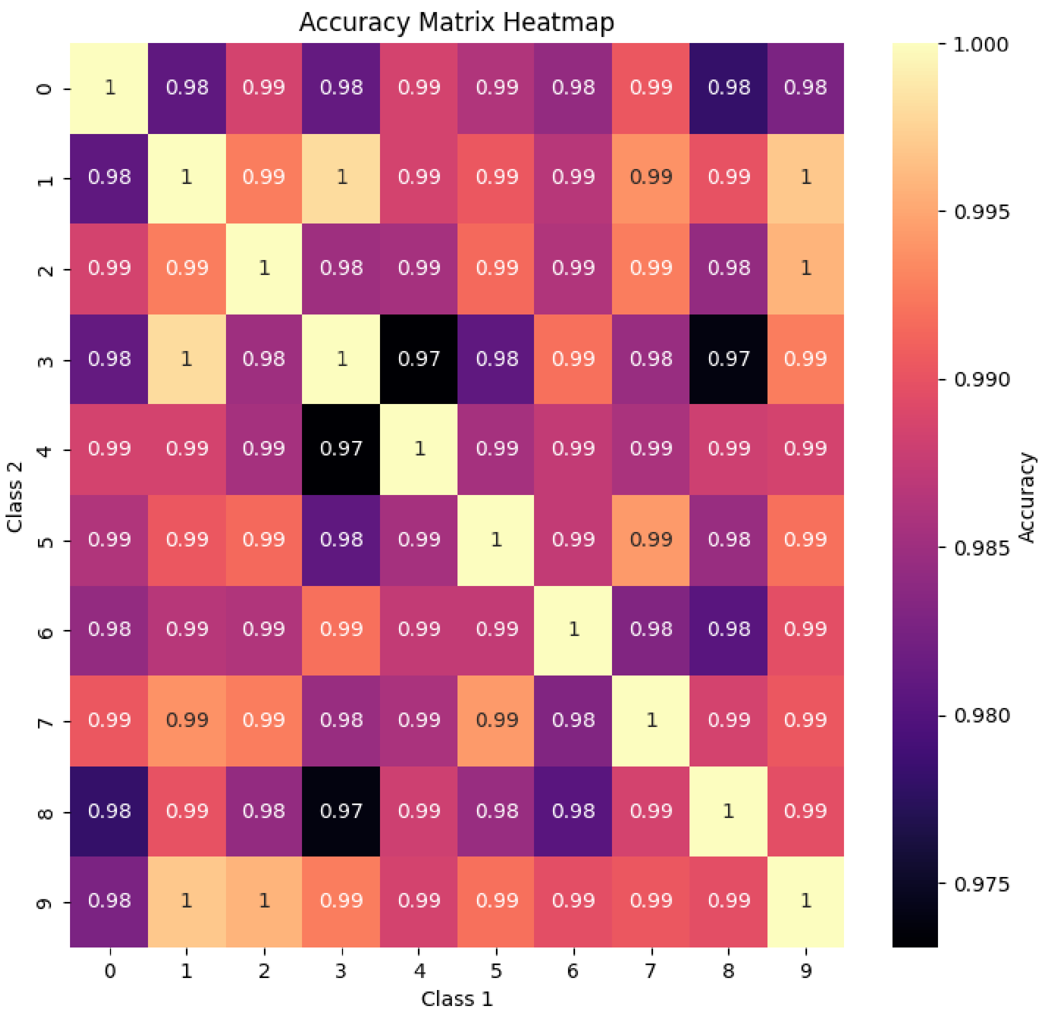
	precision	recall	f1-score	support
0	0.946203	0.996667	0.970779	300
1	0.958333	0.996667	0.977124	300
2	0.989967	0.986667	0.988314	300
3	0.985965	0.936667	0.960684	300
4	0.967105	0.98	0.97351	300
5	0.986711	0.99	0.988353	300
6	0.992883	0.93	0.960413	300
7	0.989933	0.983333	0.986622	300
8	0.948553	0.983333	0.96563	300
9	0.986348	0.963333	0.974705	300
accuracy	0.974667	0.974667	0.974667	0.974667
macro avg	0.9752	0.974667	0.974613	3000
weighted avg	0.9752	0.974667	0.974613	3000

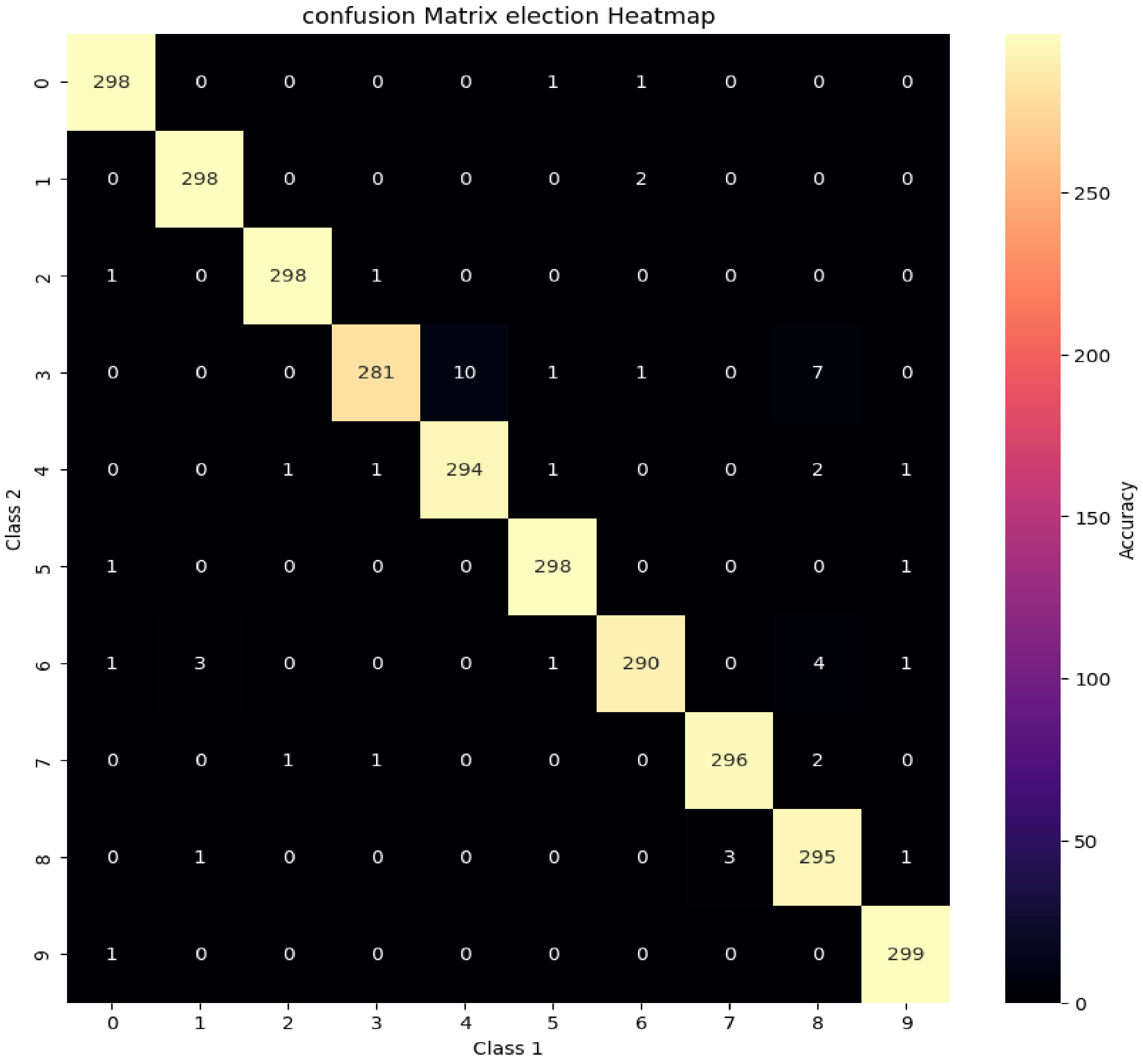

confusion Matrix 1 Heatmap
Class 1
Class 2
Accuracy

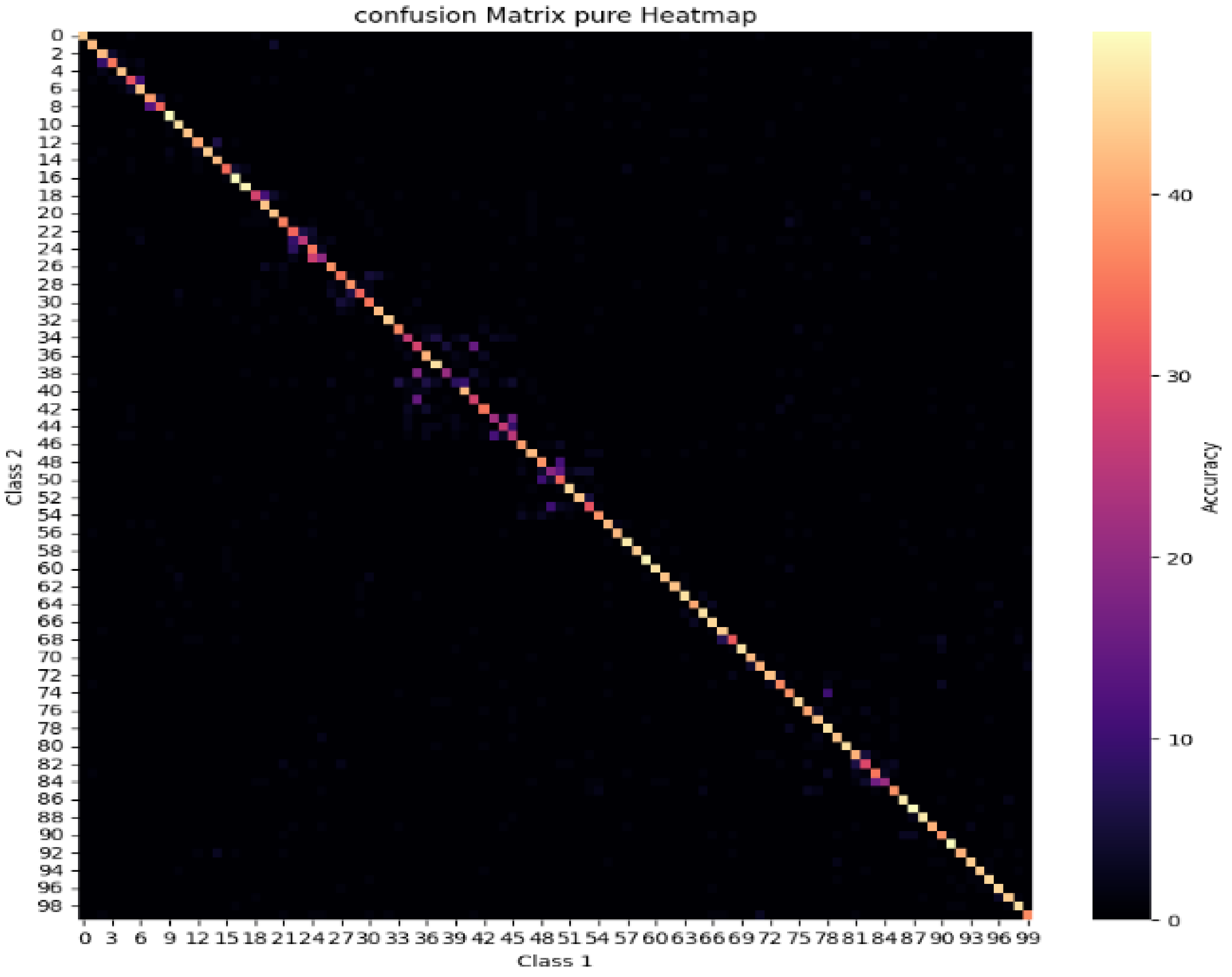

confusion Matrix 2 Heatmap
Class 1
Class 2
Accuracy

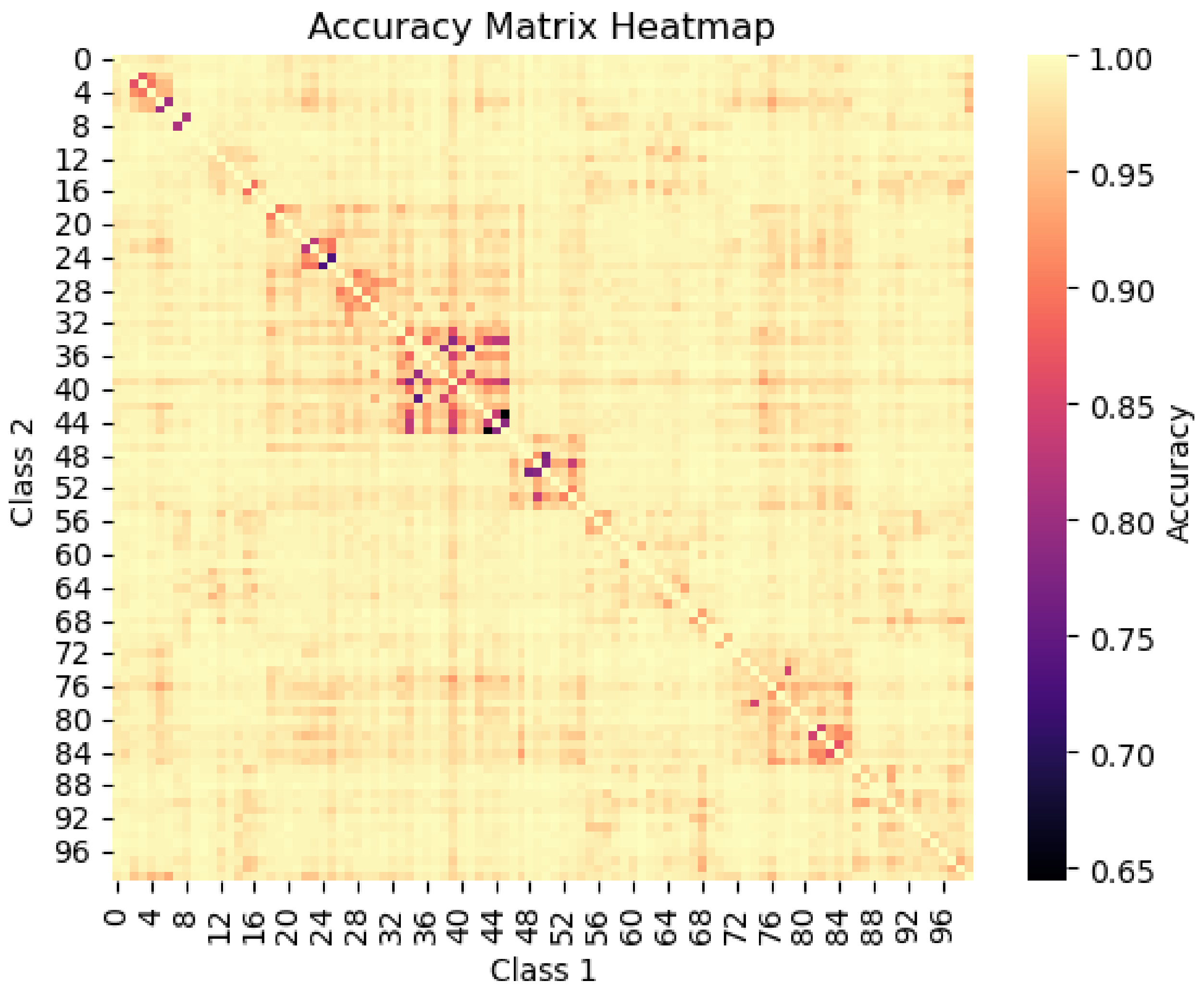

confusion Matrix 3 Heatmap
Class 1
Class 2
Accuracy

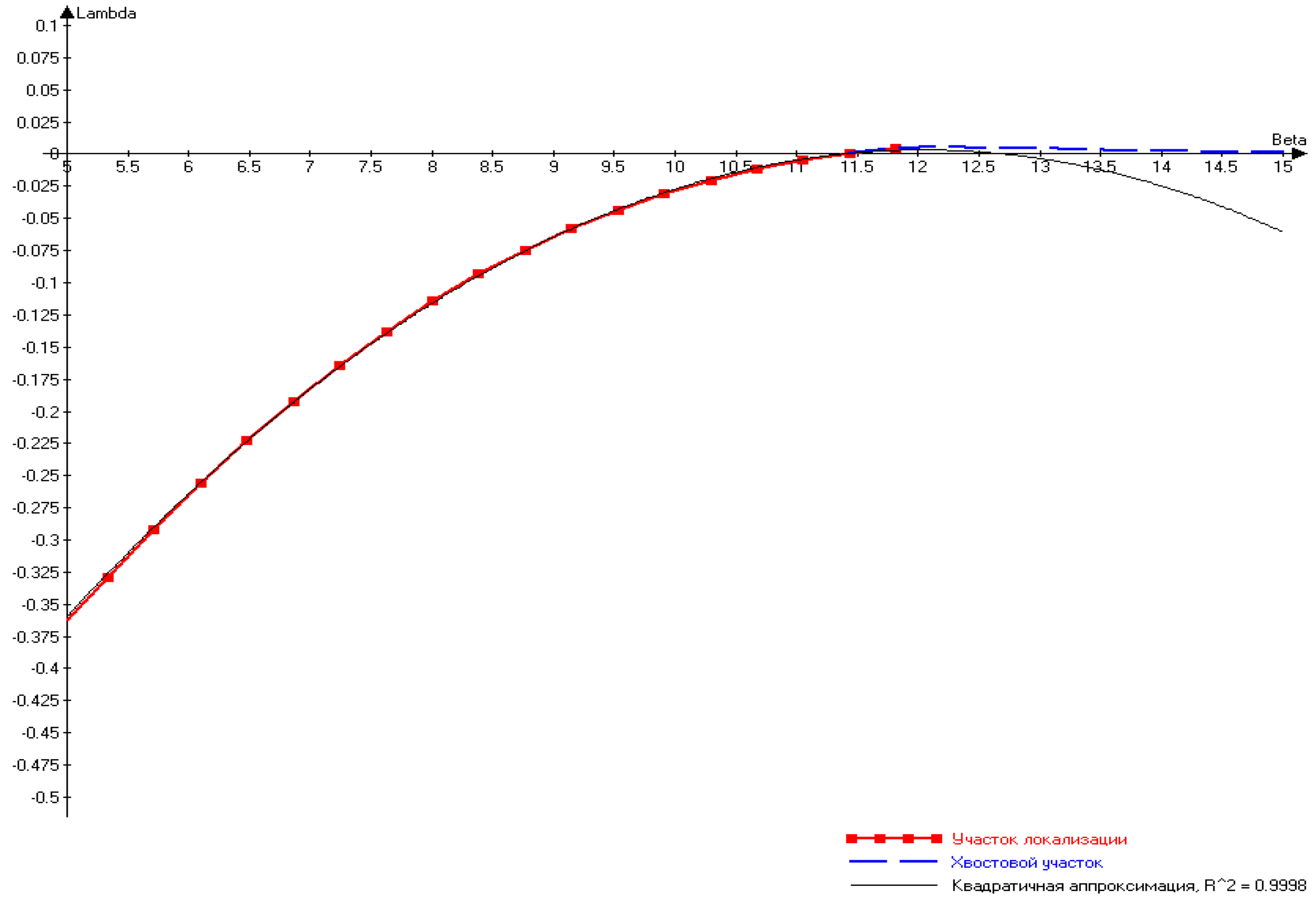
Arbitor

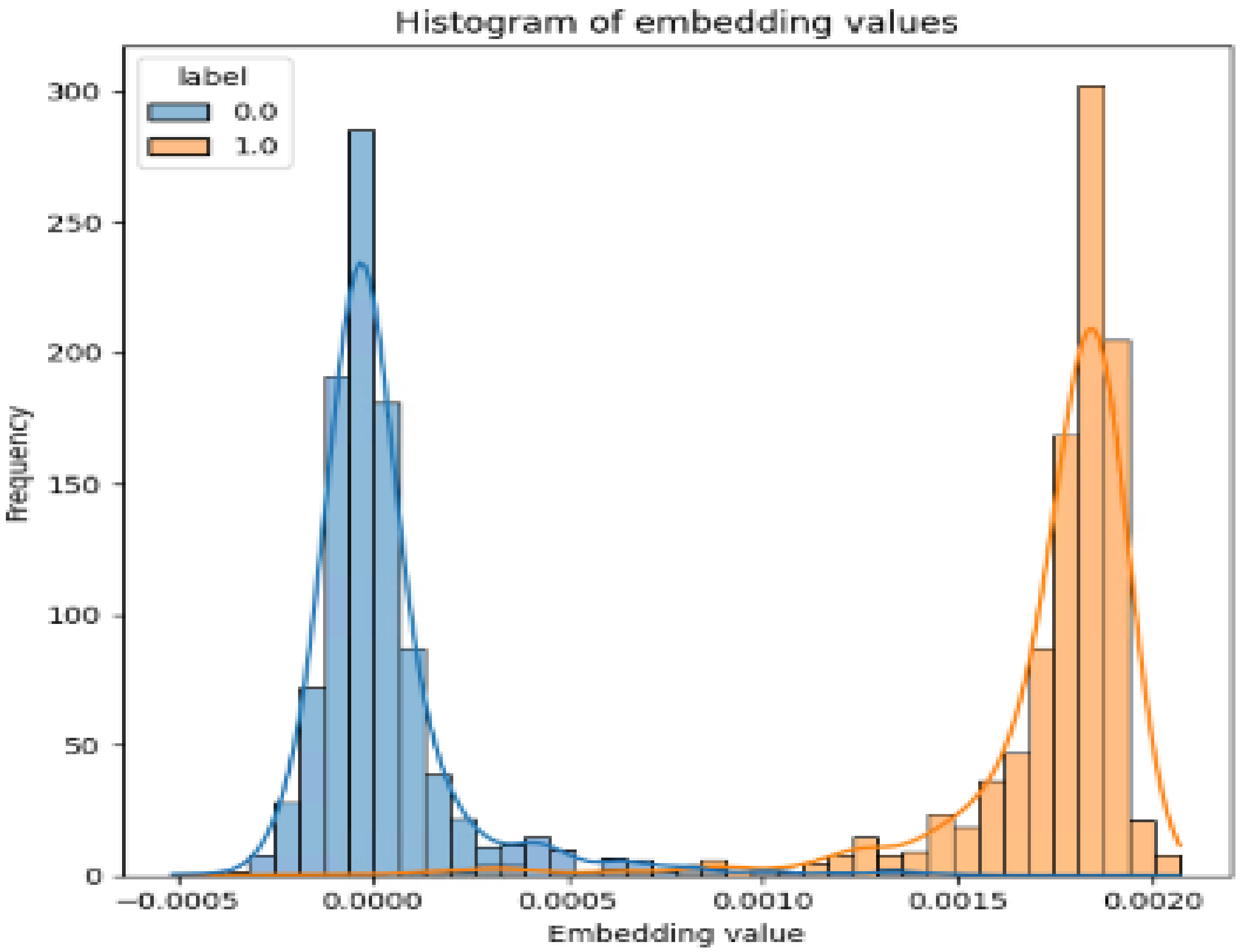

Accuracy Matrix Heatmap
1 0.98 0.99 0.98 0.99 0.99 0.98 0.99 0.98 0.98
0.98 1 0.99 1 0.99 0.99 0.99 0.99 0.99 1
0.99 0.99 1 0.98 0.99 0.99 0.99 0.99 0.98 1
0.98 1 0.98 1 0.97 0.98 0.99 0.98 0.97 0.99
0.99 0.99 0.99 0.97 1 0.99 0.99 0.99 0.99 0.99
0.99 0.99 0.99 0.98 0.99 1 0.99 0.99 0.98 0.99
0.98 0.99 0.99 0.99 0.99 0.99 1 0.98 0.98 0.99
0.99 0.99 0.99 0.98 0.99 0.99 0.98 1 0.99 0.99
0.98 0.99 0.98 0.97 0.99 0.98 0.98 0.99 1 0.99
0.98 1 1 0.99 0.99 0.99 0.99 0.99 0.99 1
Class 2
0 1 2 3 4 5 6 7 8 9
Class 1
Accuracy
1.000
0.995
0.990
0.985
0.980
0.975


confusion Matrix election Heatmap
Class 2
Class 1
Accuracy
0
1
2
3
4
5
6
7
8
9
298
281
294
290
296
295
299
10
7
250
200
150
100
50


confusion Matrix pure Heatmap
Class 2
Class 1
Accuracy
0
10
20
30
40


Accuracy Matrix Heatmap
Class 2
Class 1
Accuracy
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76
80
84
88
92
96


Lambda
0.1
0.075
0.05
0.025
0
-0.025
-0.05
-0.075
-0.1
-0.125
-0.15
-0.175
-0.2
-0.225
-0.25
-0.275
-0.3
-0.325
-0.35
-0.375
-0.4
-0.425
-0.45
-0.475
-0.5
Beta
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5
10
10.5
11
11.5
12
12.5
13
13.5
14
14.5
15
Участок локализации
Хвостовой участок
Квадратичная аппроксимация, R^2 = 0.9998


Histogram of embedding values
label
0.0
1.0
300
250
200
150
100
50
0
Frequency
−0.0005
0.0000
0.0005
0.0010
0.0015
0.0020
Embedding value

In an energy-based model, the probability of a configuration σ is defined in terms of an energy function, similar to the Hamiltonian of the Ising model. For the random bond Ising model, the probability $P(\sigma)$ of a spin configuration σ is given by the Boltzmann distribution:

$$P(\sigma) = \frac{e^{-\beta\mathcal{H}(\sigma)}}{Z},$$

where $\mathcal{H}(\sigma)$ is the Hamiltonian and Z is the partition function.

Similarly, deep neural networks can be used to model complex probability distributions. Consider a neural network with parameters θ that models a probability distribution $Q_\theta(\sigma)$. The goal is to approximate the true distribution $P(\sigma)$.

Variational inference involves approximating the true posterior distribution $P(\sigma)$ with a parameterized distribution $Q_\theta(\sigma)$ by minimizing the KL divergence between them. The KL divergence is defined as:

$$\mathrm{KL}(Q_\theta \parallel P) = \sum_\sigma Q_\theta(\sigma)\log\frac{Q_\theta(\sigma)}{P(\sigma)}.$$

Substituting $P(\sigma)$ with its Boltzmann form, we get:

$$\mathrm{KL}(Q_\theta \parallel P) = \sum_\sigma Q_\theta(\sigma)(\log Q_\theta(\sigma) + \beta\mathcal{H}(\sigma) + \log Z).$$

Since log Z is constant with respect to θ, it can be omitted from the optimization process. The optimization objective becomes:

$$\mathrm{KL}(Q_\theta \parallel P) \propto \sum_\sigma Q_\theta(\sigma)(\log Q_\theta(\sigma) + \beta\mathcal{H}(\sigma)).$$

Training Neural Networks

The training of a neural network to approximate the distribution $P(\sigma)$ can thus be viewed as minimizing the KL divergence. This is equivalent to minimizing the variational free energy:

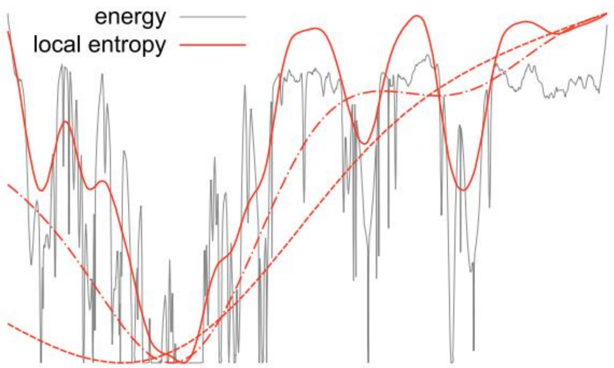
$$\mathcal{F}(\theta) = \sum_{\sigma} Q_\theta(\sigma)\left(\mathcal{H}(\sigma) + \frac{1}{\beta}\log Q_\theta(\sigma)\right).$$

Minimizing this free energy involves adjusting the neural network parameters θ to ensure that $Q_\theta(\sigma)$ closely approximates the true distribution $P(\sigma)$. This process is analogous to learning the weights in the neural network, where the energy function $\mathcal{H}(\sigma)$ can be interpreted as a parameterized function (such as a neural network) that captures the dependencies between the variables.

Local entropy*

$$\Phi(\sigma, \beta, \theta(\gamma)) = log \sum_{\{\sigma\prime\}} e^{-\beta E(\sigma') - \gamma d(\sigma, \sigma'),}$$

where $d(\cdot,\cdot)$ – distance between configuration, γ –weight, σ – reference spin configuration



Energy landscape compared with local entropy landscape

Global extremum represented by symmetrical cycle in graph.

Local extremum around represented by odd degree connected cycle in graph.

*Carlo Baldassia, Christian Borgsc , Jennifer T. Chayesc , Alessandro Ingrosso, Carlo Lucibello, Luca Sagliettia , and Riccardo Zecchina Unreasonable effectiveness of learning neural networks: From accessible states and robust ensembles to basic algorithmic schemes *Proceedings of the National Academy of Sciences* (PNAS), 113 (48), **November 15, 2016,** E7655-E7662 https://www.pnas.org/doi/full/10.1073/pnas.1608103113

Carlo Baldassi, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina Subdominant Dense Clusters Allow for Simple Learning and High Computational Performance in Neural Networks with Discrete Synapses. Physical Review Letters 115, 128101, p. 1-5., 2015

Usatyuk V., Sapozhnikov D., Egorov S. Spherical and Hyperbolic Toric Topology-Based Codes On Graph Embedding for Ising MRF Models: Classical and Quantum Topology Machine Learning https://ui.adsabs.harvard.edu/abs/2023arXiv230715778U

Quasi-Instatones– local minimum TS(a,b) under local entropy Φ

solving an incorrectly posed operator problem*

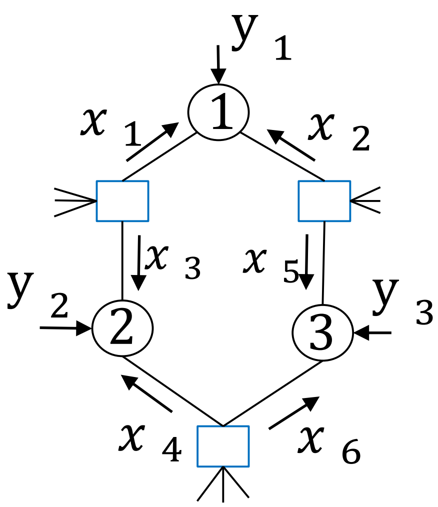
$A \subset B, A - submatrices$ (local solution) contained in B connected through variable y_1, y_2, y_3.

A subgraph with cycle 6

TS(a,b)

$x_1, x_2, x_3, x_4, x_5, x_6$

$$Ax = y,$$



$$\begin{cases} x_1 = (y_2 + x_4)/\alpha \\ x_2 = (y_3 + x_6)/\alpha \\ x_3 = (y_1 + x_2)/\alpha \\ x_4 = (y_3 + x_5)/\alpha \\ x_5 = (y_1 + x_1)/\alpha \\ x_6 = (y_2 + x_3)/\alpha \end{cases}$$

$$A = \begin{pmatrix} \alpha & 0 & 0 & -1 & 0 & 0 \\ 0 & \alpha & 0 & 0 & 0 & -1 \\ 0 & -1 & \alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha & -1 & 0 \\ -1 & 0 & 0 & 0 & \alpha & 0 \\ 0 & 0 & -1 & 0 & 0 & \alpha \end{pmatrix}$$

Choise of normalize α influences the convergence to global minima

VN ◯

CN □

For $\alpha > 1, A$ – has full rank and, accordingly, a unique solution

The selection of α allows you to find a solution to an ill-posed problem (the inverse of Tikhonov's regularization**, "balances" the variables in the cycle with the subgraph external to them***)

*Тихонов А. Н. О некорректных задачах линейной алгебры и устойчивом методе их решения // Доклады Академии наук СССР. — 1965. — Т. 163, № 3. — С. 591—594.
**М. И. Сумин, "О роли множителей Лагранжа и двойственности в некорректных задачах на условный экстремум. К 60-летию метода регуляризации Тихонова", Вестник российских университетов. Математика, 28:144 (2023), 414–435
*****M. Chertkov et al. Interaction screening: efficient and sample-optimal learning of ising models.** NIPS'16 pp. 2603-2611

The Weighted Laplacian Matrix (Laplacian Method)

A classical spectral clustering method for weighted graphs uses the *weighted Laplacian* matrix $L = \overline{D} - \tilde{J}$, where $\overline{D} = \mathrm{d}iag(|\tilde{J}|\mathbf{1}_n)$. The eigenvector associated with the smallest eigenvalue of L provides a relaxed solution to the NP-hard optimization problem of signed ratio-cut graph clustering.

For *signed graphs*, where $\tilde{J}$ entries are ± 1, the Bethe-Hessian matrix $H_{\beta,\tilde{J}}^{signed}$ is defined as:

$$H_{\beta,\tilde{J}}^{signed} = (1 - \mathrm{t}h^2(\beta))I_n + \mathrm{t}h^2(\beta)D - \mathrm{t}h(\beta)\tilde{J}.$$

As $\beta_{\mathrm{N}} \to \infty$, $H_{\beta_{\mathrm{N}},\tilde{J}}^{signed} \to L$. The signed Laplacian can be seen as the zero temperature limit of the Bethe-Hessian matrix.

The Mean Field Approximation (Naïve Bayes)

Using a naive mean field approximation, the probability distribution is:

$$p_{\mathbf{m}}(\mathbf{s}) = \prod_{i \in \mathcal{V}} \frac{1+m_i s_i}{2},$$

where m_i is the average of s_i over the distribution.

The associated variational free energy is:

$$\tilde{F}^{\mathrm{M}F}_{\tilde{J},\beta}(\mathbf{m}) = -\sum_{(ij)\in\mathcal{E}} \beta \tilde{J}_{ij} m_i m_j + \sum_{i\in\mathcal{V}} \sum_{s_i} \frac{1+m_i s_i}{2} \log\left(\frac{1+m_i s_i}{2}\right).$$

The Hessian of the free energy at the paramagnetic point $\mathbf{m} = \mathbf{0}$ is:

$$H^{\mathrm{M}F}_{\beta,\tilde{J}} = I_n - \beta \tilde{J}.$$

Thus, the eigenvectors of $H^{\mathrm{M}F}_{\beta,\tilde{J}}$ are simply the eigenvectors of $\tilde{J}$, making β irrelevant.

The Spin Glass Bethe-Hessian

The spin glass Bethe-Hessian matrix, with $\beta = \beta_{SG}$, can also achieve non-trivial clustering as soon as theoretically possible. The temperature β_{SG} is estimated by solving:

$$c\mathbb{E}[\mathrm{t}h^2(\beta_{\mathrm{SG}}\tilde{J}_{ij}\sigma_i\sigma_j)] = c\mathbb{E}[\mathrm{t}h^2(\beta_{\mathrm{SG}}\tilde{J}_{ij})] = 1.$$

However, for realistic heterogeneous graphs, this choice may be suboptimal.

Alaa Saade, Marc Lelarge, Florent Krzakala, and Lenka Zdeborov´a. Clustering from sparse pairwise measurements. In 2016 IEEE International Symposium on Information Theory (ISIT), pages 780–784. IEEE, 2016.

Квазициклические Низкоплотностные коды и их ключевые свойства

$$
\begin{array}{cc}
 & \begin{array}{cccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{array} \\
H = & \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{array}{c} c_1 \\ c_2 \\ c_3 \end{array}
\end{array}
$$

Низкоплотностные-коды (Low Density Parity Check,LDPC-коды), – это блочный линейный код размерностью k и длиной кодового слова n, задаваемый проверочной матрицей H размерностью $(n\text{-}k)\cdot n$, имеющей небольшую плотность отличных от нуля символов. Матрица размера $(n\text{-}k)\cdot n$ содержит порядка n проверок.

Каждая строка проверочной матрицы H задает уравнение проверки на четность:

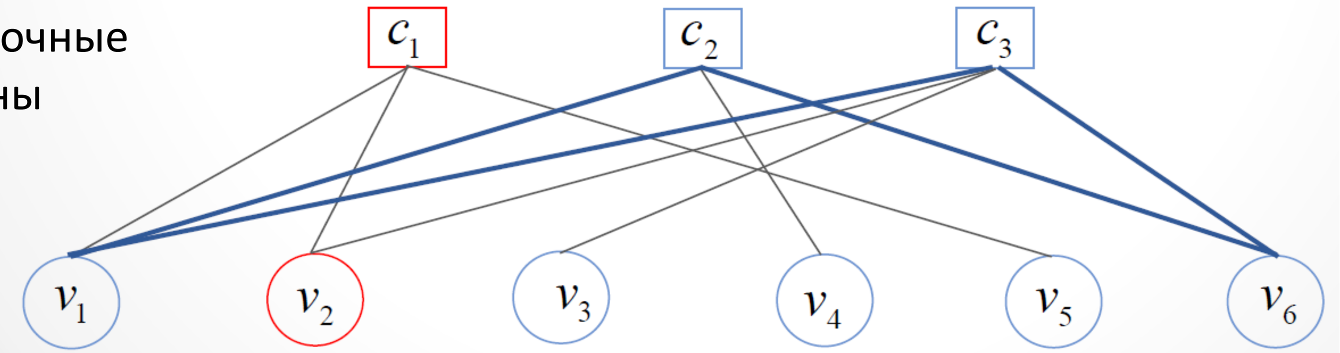
$$\sum_{l=0}^{n-1} v_l \cdot h_{\mathrm{j},l} = 0$$

где j – номер строки проверочной матрицы (номер проверочного уравнения); l – номер символа кодового слова; $h_{j,l}$ – элемент проверочной матрицы.

Граф Таннера –это двудольный граф, соответствующей проверочной матрице LDPC- кода

проверочные вершины

кодовые вершины



Замкнутый простой путь в Таннер графе с узлами $v_1 \rightarrow c_2 \rightarrow v_6 \rightarrow c_3 \rightarrow v_1$ образуют цикл длины 4

Квазициклические низкоплотностные коды (Quasi Cyclic-LDPC)

$$H = \begin{array}{cc} & \begin{array}{cccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{array} \\ & \left[\begin{array}{cc|cc|cc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \begin{array}{c} c_1 \\ c_2 \\ c_3 \\ c_4 \end{array} \end{array}$$

$$H_{QC} = \begin{bmatrix} I^0 & I^1 & I^1 \\ I^0 & I^{-1} & I^0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

Квазициклический LDPC (коды QC-LDPC) - LDPC-коды с матрицей проверки на четность, определяемой структурированной блочной подматрицей - матрицей циклических перестановок.

Матрица циклических перестановок размера 2x2, веса 1 и ее сжатая форма представления

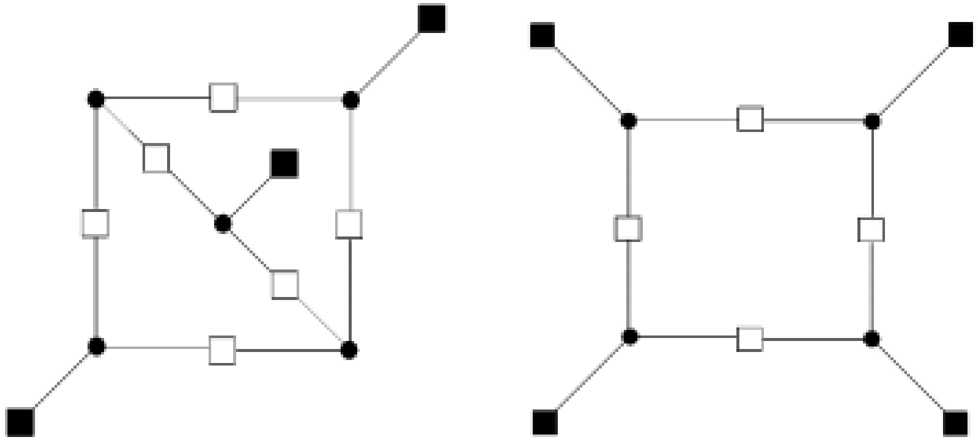
$$I^0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, I^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, I^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Традиционно используются матрицы веса 1.

[redacted]LDPC codes”, Proc. ICSTA 2001, Ambleside, England, 2001

Свойства низкоплотностных кодов. Трэппин-сеты (a,b), a,b∈ $\mathbb{Z}$

На свойства низкоплотностых кодов отрицательно влияют циклы в графе Таннера, образующих «Трэппин-сеты». (Trapping set, TS,) или (*a*,*b*)-подграфы (подграфы в графе Таннера, состоящие из *a* символьных узлов, *b* из которых инцидентны проверочным узлам с нечетными степенями). На рисунке ниже изображены два (a,b)-подграфа. Левый подграф образован пересечением трех циклов длины 8, правый одним циклом длины 8.



(a) (b)

Trapping sets: (a) $(5, 3)$ and (b) $(4, 4)$

-символьные узлы,

- проверочные узлы с четной степенью инцидентности

- проверочные узлы с нечетной степенью инцидентности

NP-сложная задача

A. McGregor and O. Milenkovic, "On the Hardness of Approximating Stopping and Trapping Sets in LDPC Codes," *2007 IEEE Information Theory Workshop*, Tahoe City, CA, USA, 2007, pp. 248-253

Поиск Трэппин-сетов (a,b), b≠ 0

*

Полный перебор требует анализ

$$\binom{4896}{48} \approx 8.2812e{+}115$$

Gurobi 7.5.1 solver линейного программирования 32 threads

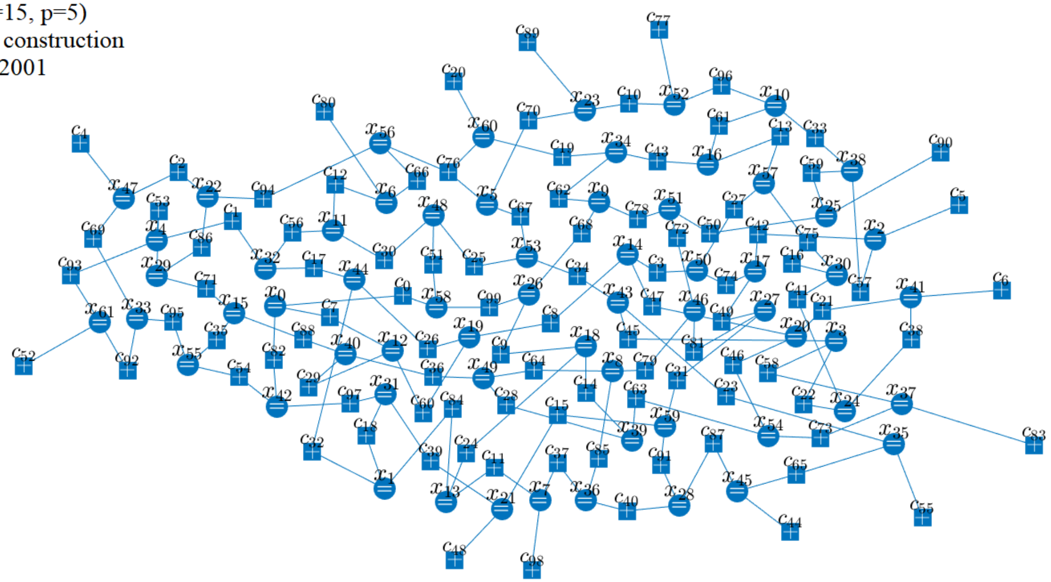
SS 48 в коде Маргулиса (4896,2474) Margulis за 700 451 с.

*Velasquez A., Subramani K. , Wojciechowski P. , On the complexity of and solutions to the minimum stopping and trapping set problems,Theoretical Computer Science, Volume 915, 2022, Pages 26-44,

Полный перебор требует анализ

$$\binom{4896}{62} \approx 1.2701E{+}143$$

TS(62, 16) at Margulis(4896, 2448)
Ramanujan graph(q=15, p=5)
Vontobel-Rosenthal construction
Allerton 2000, ISIT 2001
true_K=2474



Предложенный метод**

TS(62,16) with
TS variable nodes, (x_0)...(x_end) on Fig: 6 41 65 159 364 546 719 769 880 970 981 1097 1132 1140 1164 1274 1279 1385 1508 1561 1625 1681 1716 1819 1823 1979 1988 2194 2216 2369 2460 2506 2738 2765 2795 2855 2950 2976 3024 3154 3182 3271 3444 3566 3575 3836 3940 4012 4109 4141 4162 4355 4396 4401 4463 4547 4581 4671 4710 4800 4859 4864

CPLEX 12.8 LP solver 32 threads
Total (root+branch&cut) = 3.67 секунды. (292.92 ticks)

V. S. Usatyuk, "Low Error Floor QC-LDPC Codes Construction Using Modified Cole's Trapping Sets Enumerating," *2023 25th International Conference on Digital Signal Processing and its Applications (DSPA)*, Moscow, Russian Federation, 2023, pp. 1-6

Кодовые слова - Трэппин-сеты TS(a,0), поиск слов малого(минимального веса)

Задача оценки кодового расстояния

Пусть матрица H, размера $(n-k)\times n$, заданная над F_q определяет линейное отображение $F_q^n \to F_q^{n-k}$ такое, что $x \to Hx$, $\ker(H) = \left\{ c \in F_q^n \middle| Hc = 0 \right\}$.

Оценка кодового расстояния линейного блочного кода: Найти вектор с минимальным число отличных от нуля координатных компонент, $c \in ker(H) \setminus \{0\}$?

d_{min}-кодовое расстояние

Пример поиска кодового расстояния - Трэппин-сета (d_{min},0)

$$Hc = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$GF(2)$

$$\overline{c}_0^{\,T} = [0 \quad 0 \quad 0 \quad 0 \quad 0] \qquad w(\overline{c}_0^{\,T}) = \left\|\overline{c}_0^{\,T}\right\| = 0$$

$$\overline{c}_1^{\,T} = [0 \quad 0 \quad 0 \quad 1 \quad 1] \qquad w(\overline{c}_1^{\,T}) = \left\|\overline{c}_1^{\,T}\right\| = 2$$

$$\overline{c}_2^{\,T} = [0 \quad 1 \quad 1 \quad 0 \quad 1] \qquad w(\overline{c}_2^{\,T}) = \left\|\overline{c}_2^{\,T}\right\| = 3$$

$$\overline{c}_3^{\,T} = [0 \quad 1 \quad 1 \quad 1 \quad 0] \qquad w(\overline{c}_3^{\,T}) = \left\|\overline{c}_3^{\,T}\right\| = 3$$

Полином весового спектра кода

$$f(z) = 1 + z^2 + 2z^3$$

Слово веса 0

Весовой спектра кода $\quad W(c) = [1, 0,1,2,0,0]$

Слово веса 2

Слова веса 3

$$\mathrm{d}_{\min} = 2;$$

Спектр связности можно улучшать и в частности TS(d_{min},0): $d_{min} \to max$

Международный конкурс по оценки кодового расстояния от объединения: Французского национального центра научных исследований (CNRS), Национального института исследований в области цифровых наук и технологий (Inria Paris), национального исследовательский институт математики и информатики в Нидерландах (CWI)

Best solutions

Weight	Authors	Algorithm	Details
214	Vasiliy Usatyuk	Lattice: Kannan emb, SBP (SBP), SVP	See details
215	Samuel Neves	-	See details
220	Valentin Vasseur	Dumer	See details

https://decodingchallenge.org/low-weight/

Low-weight Codeword Problem

This page is dedicated to the problem of finding low-weight codewords for random binary linear codes.

Low Weight Codeword problem. Given integers n, k, w such that $k \leq n$ and $w \leq n$, an instance of the problem $\mathrm{LWC}(n,k)$ consists of a full rank parity-check matrix $\mathbf{H} \in \mathbb{F}_2^{(n-k)\times n}$. A solution to the problem is a non-zero vector $\mathbf{c}$ with Hamming weight $\leq w$ such that $\mathbf{Hc}^\top = \mathbf{0}$.

The challenge. Here, we focus on instances with code rate $R = 0.5$, that is $n = 2k$. We fix $n = 1280$ (see below). At this length, there exists on average a unique codeword of weight 144 (the Gilbert-Varshamov bound) and finding it should requires at least 2^{128} operations (see below). Finding words of higher weight is easier. The goal is to find codewords with a weight as low as possible. The current record is Array.

Choice of n. The complexity of finding a codeword of weight equal to the Gilbert-Varshamov bound in a code of rate $R = 0.5$ using the BJMM algorithm asymptotically requires $2^{0.0999852\,n}$ operations, therefore with $n = 1280$, finding the smallest codeword should require at least 2^{128} operations.

Compared to the Syndrome Decoding challenge, here the size of the instance is **cryptographically large**. The goal is to assess that finding codewords close to the GV-bound is hard. This problem is close to the Shortest Vector Problem for lattices.

Instance generation. The instances are generated using a Python script. This script takes as input the length of the code and a seed (but for this challenge we only use the length $n = 1280$).

How to participate?

1. Choose an instance on the right. All instances have the same size, the only difference is the value of the random seed used to generate them. You can also use the generator to generate an instance with another random seed.
2. Parse the instance to get the values of $\mathbf{H}$. Find a non-zero codeword **c of weight** < **Array** such that $\mathbf{Hc}^\top = \mathbf{0}$.
3. Submit your solution using the submission form. If your solution is correct, your name will appear in the hall of fame.

Submit your solution

Hall of fame

Download instances

Instance generator

Format of instances

Download instances

All instances have the same size.
They are indexed by seed.

0	1	2	3
4	5	6	7
8	9	10	11

NP-сложная задача

Best solutions

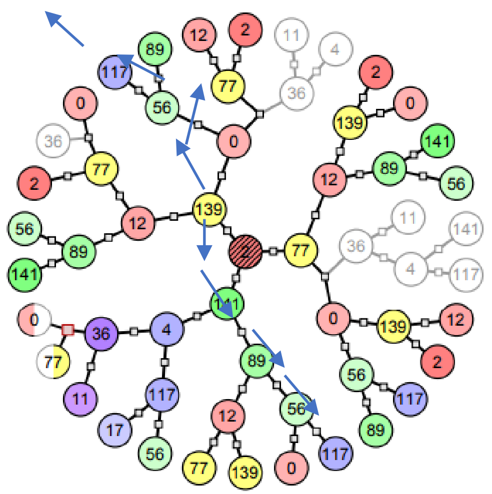
Weight	Authors	Algorithm	Details
211	Leo Ducas, Marc Stevens	Hybrid Wagner-Babai using Code Reduction [eprint:2020/869]	See details
212	Vasiliy Usatyuk	Lattice:SBP(BKZ), SVP	See details
214	Vasiliy Usatyuk	Lattice: Kannan emb, SBP (SBP), SVP	See details

I. Dumer, D. Micciancio and M. Sudan, "Hardness of approximating the minimum distance of a linear code," in *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 22-37, Jan. 2003

Модель Шеррингтона-Киркпатрика (Sherrington-Kirkpatrick)

$$H_E = -\sum_{i=1,\ldots,N}\ \sum_{a=1,\ldots,M} C_{ij} J_{ij} \sigma_i \sigma_j$$

C_{ij} - матрица связности, элемент которого равен 1, если два спина взаимодействуют, и 0 в противном случае,
J_{ij}- вес (потенциал) взаимодействия между спинами, σ_i- спины модели Изинга



Число слоев p –это число

Пар спинов соединённых

Проверочной матрицей кода.

Число итераций декодирования

H_E можно преобразовать:

$$H_{SK} = -\sum_p \sum_{i_1,\ldots,i_p=1,\ldots,n} C^{(p)}_{i_1,\ldots,i_p} J_{i_1,\ldots,i_p} \sigma_{i_1} \ldots \sigma_{i_p},$$

где $J_0^{(p)}, \Delta\ J^2_{(p)}$ М.О. и дисперсия J_{ij}.

Проверочная матрица квазициклического кода H определяет $C^{(p)}_{i_1,\ldots,i_p}$.

x_i n-битное кодовое слов кодирующее спин $\sigma_i = 2x_i - 1$.

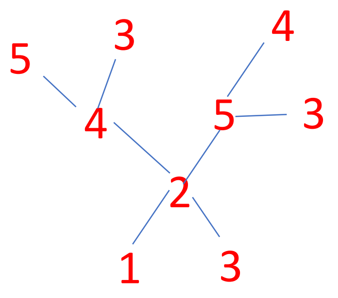
Взаимодействие пар спинов (без кратных ребер) происходит на глубину p.

$J_{i_1,\ldots,i_p}$ - двух спинового взаимодействия и принимаются как независимые случайные величины с известным распределением вероятностей.

$$\bar{x}_0^{\,T} = [0 \quad 0 \quad 0 \quad 0 \quad 0], \bar{x}_1^{\,T} = [0 \quad 0 \quad 0 \quad 1 \quad 1]$$

$$\bar{x}_2^{\,T} = [0 \quad 1 \quad 1 \quad 0 \quad 1], \bar{x}_3^{\,T} = [0 \quad 1 \quad 1 \quad 1 \quad 0]$$

Множество спинов соответствующих кодовым словам
является состояниями минимальной энергии Гамильтониана



$$Hx = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Sourlas, N.. Spin Glasses, Error-Correcting Codes and Finite-Temperature Decoding. EPL (Europhysics Letters). 25. 159. (2007)

Граф Таннер (мультиграф, в случае наличия не тривиального X_1) Квазициклического низкоплотностного кода соответствует калибровочному поля

Кодовые вершины | Ребра Графа | Проверочные узлы

$$X = \left(X_0, X_1, X_2\right)$$

$$X_1 \subset X_0 \times X_2$$

$$(j,\alpha) \in X_1 \Leftrightarrow j \in \alpha \qquad \alpha = 1,\cdots,M$$

$$\left|X_0\right| = N \quad \left|X_2\right| = M$$

$$\left|X_1\right| = \sum_{j \in X_0} m_j = \sum_{\alpha \in X_2} n_\alpha$$

Кодовые слов (закодированные в спин)

$$\sigma = \left\{\sigma_j\right\}_{j \in X_0}; \sigma_j = \pm 1$$

Эквивалентные коды (калибровочная инвариантность)

$$X \sim X' \Leftrightarrow C_X = C_{X'}$$

$$C_X = \left\{\sigma \,\middle|\, \prod_{j \in \alpha} \sigma_j = 1, \forall \alpha \in X_2\right\}$$

Калибровочная группа

$$G = GL\left(M, F_2\right)$$

Эквивалентные коды (калибровочная инвариантность)

$$Hx = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$H'x = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Графы не эквивалентны, мы добавили цикл

Статистическая модель независимых ошибок спина

Намагниченности спинового стекла

Вероятность результата измерения

$$P_j(\zeta) = \left(2\pi s^2\right)^{-N/2} \int dh\, \delta\left(m_j(h) - \zeta\right) \exp\left(-F(h)\right)$$

Вероятность индивидуальных ошибок спина

$$B_j = \int_{-\infty}^{0} d\zeta P_j(\zeta)$$

Для симметрично распределенного Гауссова шума

$$F(h) = \left(2s^2\right)^{-1} \sum_{j \in X_0} \left(h_j - s^2\right)^2$$

Отношения сигнал-шум (ОСШ), Температура

В функции плотности вероятности преобладает наиболее вероятная конфигурация шума инстантон (седловая точка)

$$\frac{\partial}{\partial h}\left(F(h) - \lambda\left(m_j(h) - \zeta\right)\right)\Big|_{h=\bar{h}} = 0$$

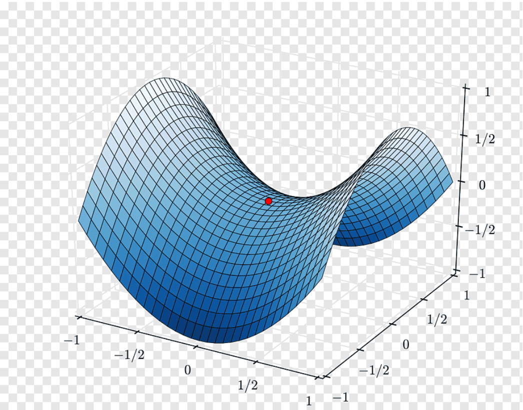
$$m_j\left(\bar{h}\right) - \zeta = 0$$

Множитель Лагранжа

$$P_j(\zeta) \approx \kappa^{-1}(\zeta) \exp\left(-F\left(\bar{h}\right)\right)$$

$$B_j \approx \kappa_0^{-1} \exp\left(-F\left(\bar{h}\right)\right)$$

TS(a,0)
Кодовые слова



V. Chernyak, M. Chertkov, M. Stepanov, and B. Vasić, "Local theory of BER for LDPC codes: instantons on a tree," *Los Alamos Workshop on Applications of Statistical Physics to Coding Theory,* Jan. 2005.

Топологический инварианты и кривизна Изинг модели полученной из графов Таннера (мультиграфа в случае веса циркулянта больше 1)

Локальная система:
Каждый узел имеет древовидную “окрестность”

$$j \in U_j^{(l)} \subset X$$

Универсальное покрывающее дерево
(аналогично римановым поверхностям)

$$p : \overline{X} \to X$$

Фундаментальная группа

свободная группа с генератором

$$\overline{X}(a;l) \subset \overline{X}; a \in p^{-1}(j) \cong \pi_1(X) \cong F(g)$$

Род поверхности 3 (genus)

Теорема Черна-Гаусса-Бонне

Род поверхности (genus)

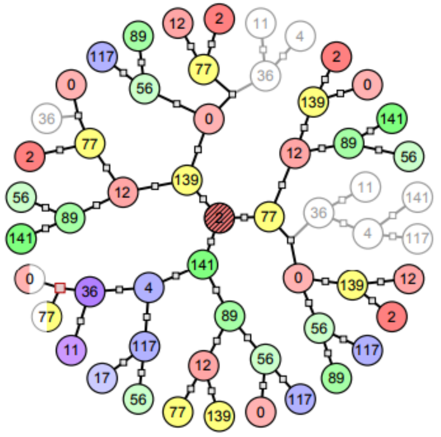
$$\sum_{j\in X_0} \frac{2-m_j}{2} + \sum_{\alpha \in X_2} \frac{2-n_\alpha}{2} = 1-g$$

Локальная кривизна

Графы с постоянной кривизной

$$m_j = m \qquad mN = nM$$

$$n_\alpha = n \qquad \overline{X} \cong Y(m,n)$$



Покрывающее дерево универсально и обладает высокой симметрией
Блочной структурой

Wiberg Codes and iterative decoding on general graphs 1995
Weiss Correctness of Local Probability Propagation in Graphical Models with Loops 2000
Knill A GRAPH THEORETICAL GAUSS-BONNET-CHERN THEOREM arXiv:1111.5395 2011
Forney, "Codes on Graphs: Models for Elementary Algebraic Topology and Statistical Physics 2018
Meshulam, Roy. (2018). Graph codes and local systems.

Квази-инстантоны – псевдокодовые слова TS(a,b)

Дискретные значения спина

$$A(\overline{\eta}) \subset \overline{X}(0;l)$$

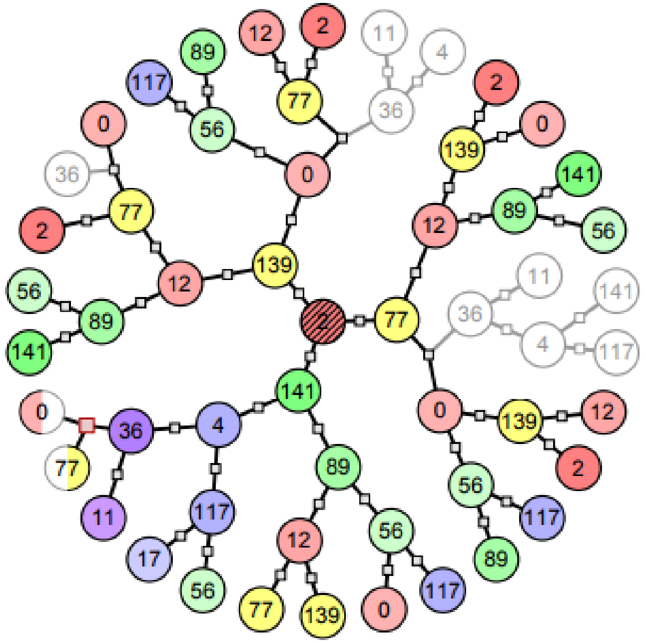
$$\sigma_j = \prod_{i\in\gamma_j}^{i\neq j} \operatorname{sgn}(\overline{\eta}_i);\ j\in A_0(\overline{\eta})$$

$$(A,\sigma)$$

Намагниченность спинового стекла

$$\overline{\eta}_0 = \overline{h}_0 + \sum_{k\in A_0}^{k\neq 0} s_k \overline{h}_k = \sum_{i\in X_0} n_i \overline{h}_i \qquad s_k = \prod_{a\in C_k^{(0)}}^{a\neq 0} \sigma_a$$

Вместо глобального минимума множество локальных



Дерево квазициклического Низкоплотносного кода Таннера [155,64,20] для p=4 слоев (после 4 итераций)

TS(a,b), b≠ 0

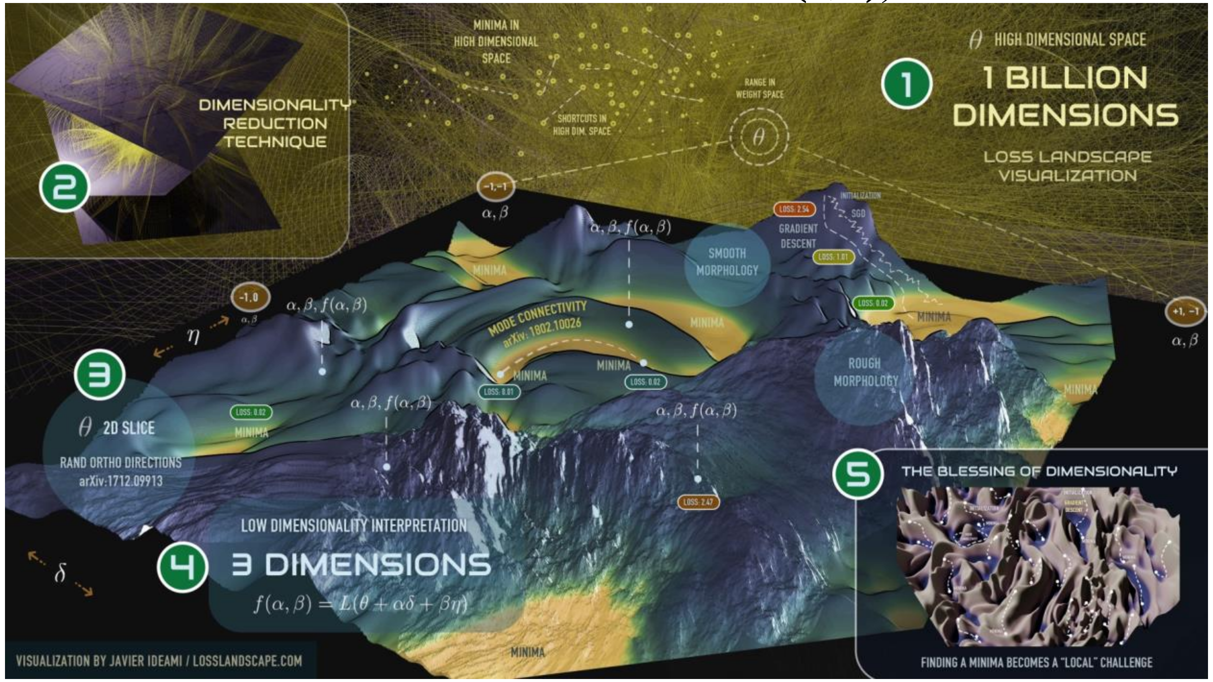
Квази-инстантоны (субоптимальные решения в результате неполноты ранга ***решаемой системы, циклов в графе)***

$$\frac{\partial}{\partial \overline{h}_j} \sum_{i\in X_0} \left(\frac{s^2(\overline{h}_i - 1)^2}{2} - \lambda n_i \overline{h}_i \right) = 0; \sum_{i\in X_0} n_i \overline{h}_i = 0 \qquad m_j = s^2 \overline{\eta}_0$$

$$\overline{h}_j = 1 - n_j \left(\sum_{k\in X_0} n_k \right) \left(\sum_{k\in X_0} n_k^2 \right)^{-1} \quad \overline{S}(n) = \frac{1}{2} \left(\sum_{k\in X_0} n_k \right)^2 \left(\sum_{k\in X_0} n_k^2 \right)^{-1}$$

Beom Jun Kim* Performance of networks of artificial neurons: The role of clustering. PHYSICAL REVIEW E 69, 045101(R) (2004)

Ландшафт потерь квантовых моделей на калибровочном поле заданном Низкоплотностным кодом. “Гладкость” (хорошая обусловленность), возникает при улучшение спектров связности (* устранения треппин-сетов TS(a,b)) и максимизации кодового расстояния TS(a,0)



*Ландшафты поверхностей образуются в результате формирования ячеек вокруг кодовых слов TS(a,0) и псевдокодовых слов (TS(a,b)) образованных циклами

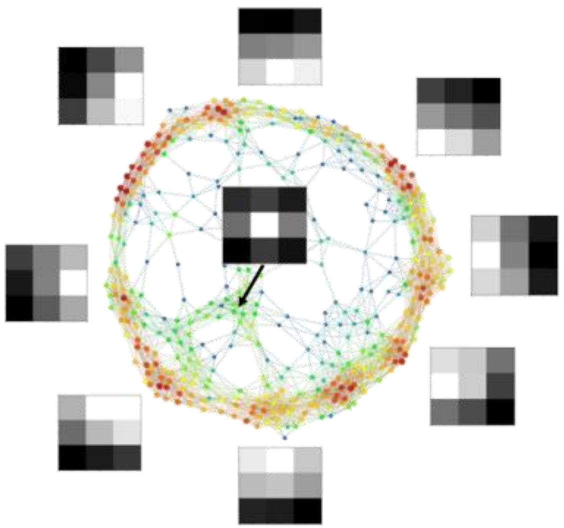


Figure 6: CIFAR-10 layer 2, gray scale

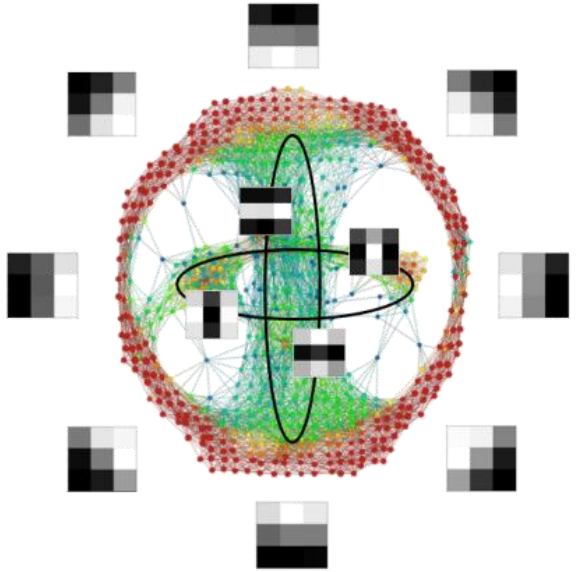


Figure 7: First layer, CIFAR-10

** Топология структуры (комплекс) активации тензоров глубокой нейронной сети (CNN, 3x3), обученной на CIFAR-10

*Beom Jun Kim Performance of networks of artificial neurons: The role of clustering. PHYSICAL REVIEW E 69, 045101(R) (2004)

** [1712.09913] Visualizing the Loss Landscape of Neural Nets (arxiv.org)

***Carlsson, G., Gabrielsson, R.B. (2020). Topological Approaches to Deep Learning. In: Baas, N., Carlsson, G., Quick, G., Szymik, M., Thaule, M. (eds) Topological Data Analysis. Abel Symposia, vol 15. Springer