Copula-SVI: Vine-Copula Variational Inference with Stein Refining for Instance-Level Correlation Capturing

Junxi Xiao[1], Qinliang Su[1,2*]

[1] School of Computer Science and Engineering, Sun Yat-sen University
[2] Guangdong Key Laboratory of Big Data Analysis and Processing

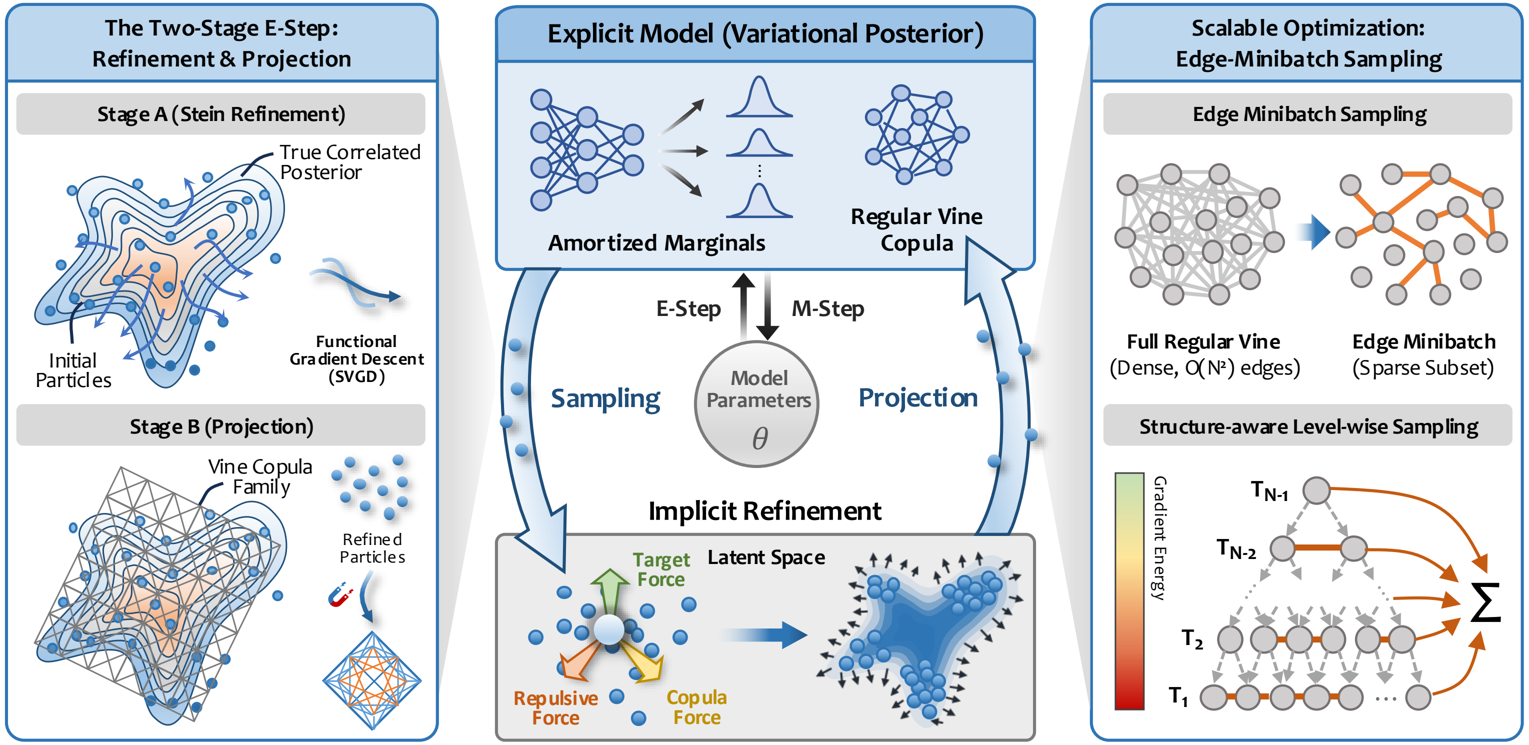
Why mean-field / SVGD-only VI is not enough

- **Real data:** time series, graphs, constraints ⇒ correlated instances
- **Mean-field VI:** $q(\mathbf{Z}) = \prod_i q(\mathbf{z}_i \mid \mathbf{x}_i)$ ⇒ misses cross-instance posterior dependence
- **Prior instance-level VI:** tree / local high-order structures are useful, but can be restrictive or costly
- Latent-space SVGD with factorized projection creates a “refine-and-forget” loop
- **Goal: scalable, non-Gaussian, higher-order instance-level posterior**

Key Idea: copula posterior + Stein refinement

❖ **Separate marginal modeling from dependence modeling**

$$q_{\Phi}(\mathbf{Z}\,|\,\mathbf{X}) = c(\mathbf{U};\boldsymbol{\psi}) \cdot \prod_i q_{\phi}(\mathbf{z}_i\,|\,\mathbf{x}_i), \quad \text{where } u_i^d = F_{\phi}(z_i^d\,|\,\mathbf{x}_i)$$

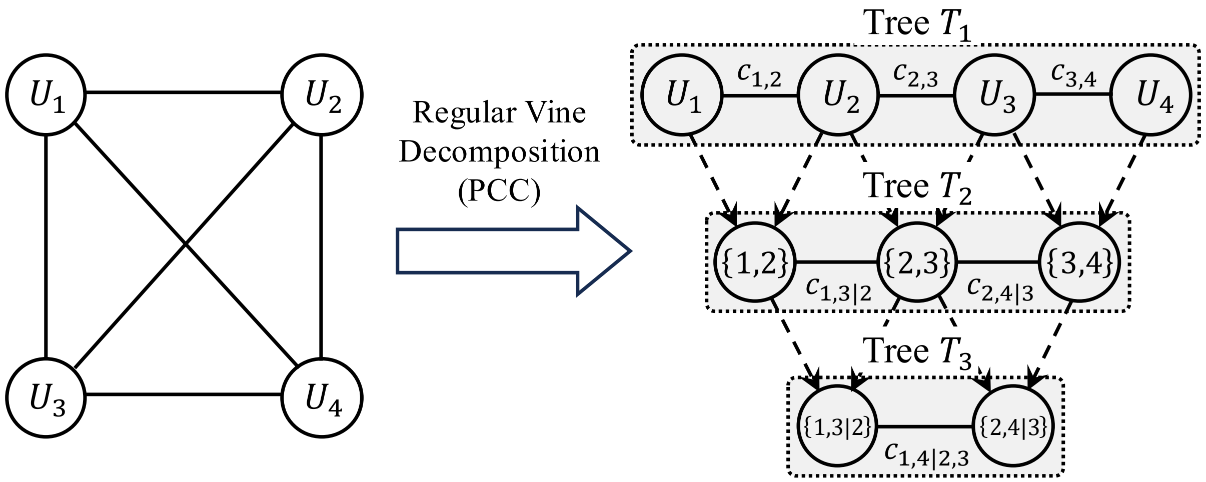


Core benefit: refined cross-instance dependence is preserved instead of erased.

Vine-copula decomposition: high-order dependence

- A full N-variate copula is expressive but impractical for large N
- **Regular Vine:** decompose high-dimensional dependence into bivariate pair-copulas
- Layered edges capture first-order and higher-order conditional dependencies

$$\log c(\mathbf{U};\boldsymbol{\psi}) = \sum_{d}\sum_{\ell}\sum_{e\in E_\ell} \log c(u^d_{e,1}, u^d_{e,2}; \psi_{e,d})$$



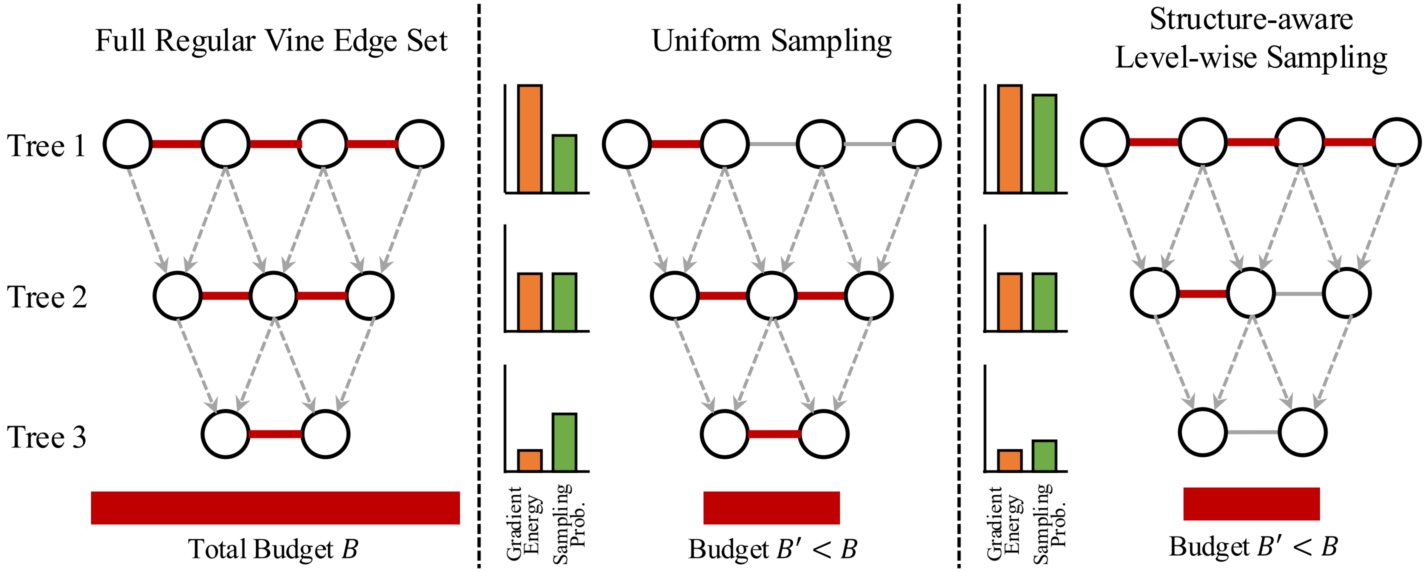
Vine turns dense dependence into an edge-additive objective.

Scalable vine learning via level-wise sampling

- Full vine has $\mathcal{O}(N^2)$ edges ⇒ full updates are expensive
- Edge-minibatching with Horvitz–Thompson weights gives an unbiased gradient

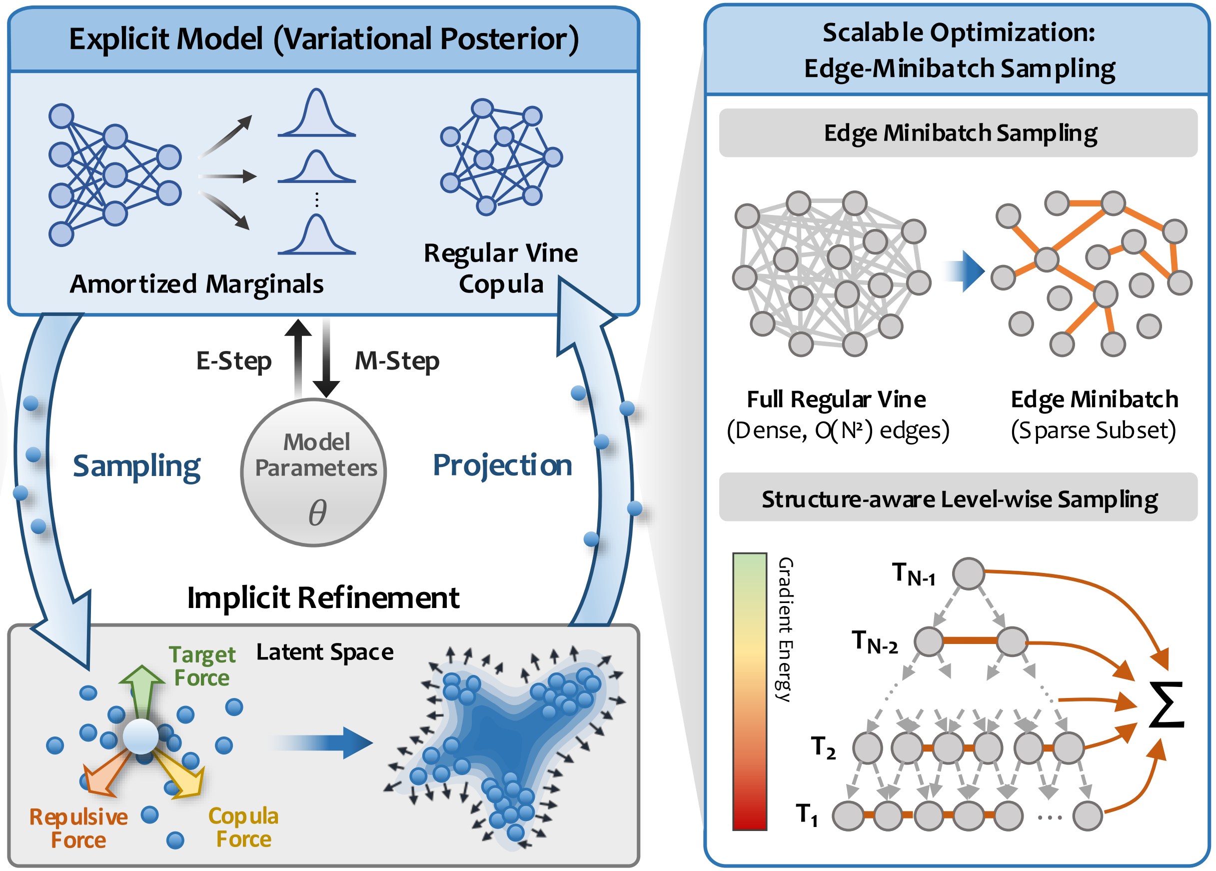
$$\hat{g}_{\boldsymbol{\psi}} \log c(\mathbf{U};\boldsymbol{\psi}) = \sum_{d} \sum_{e \in S} \frac{1}{\pi_e} \cdot \nabla_{\boldsymbol{\psi}} \log c_{e,d}(u^d_{e,1}, u^d_{e,2}; \psi_{e,d})$$

- Uniform sampling is unbiased but high variance
- Level-wise allocation samples more edges from high-gradient-energy vine levels



Efficient sampling: sparse-vine initializer + Stein correction

- Sampling from the full vine is also expensive
- Reuse the sampled edge set S to build a sparse vine initializer $q_{\Phi,S}(\mathbf{Z} \mid \mathbf{X})$
- SVGD then corrects sparse-initialized particles using the exact posterior score



Theory: sparse-init bias is transient; finite gap is governed by omitted-edge energies.

Results: time-series anomaly detection

- **Datasets:** SMAP, MSL, SMD; **metric:** F1 and ELBO
- Higher vine order ⇒ better F1 and ELBO; Copula-SVI > HoT-VI at comparable orders

Table 2. Anomaly detection results (F1-score and ELBO).

Dataset		SMAP		MSL		SMD	
Metric		F1	ELBO	F1	ELBO	F1	ELBO
DAGMM		0.7105	-115.2820	0.7007	-277.7380	0.7094	-155.9460
LSTM-VAE		0.7298	-116.9500	0.6780	-281.3220	0.7842	-146.0540
OmniAnomaly		0.8434	-98.9217	0.8849	-161.0002	0.8857	-72.0419
SISVAE		0.8299	-101.1878	0.8766	-182.6060	0.8775	-72.5832
HoT-VI	3-order	0.8552	-95.2314	0.8940	-157.2134	0.9153	-67.4001
	10-order	0.8636	-92.2948	0.9145	-134.0815	0.9284	-65.0345
Ours	3-order	0.8628	-93.9194	0.9032	-154.5485	0.9224	-65.1764
	10-order	0.8753	-90.9828	0.9245	-132.6568	0.9371	-63.7555
	50-order	0.8846	-87.4992	0.9387	-103.7108	0.9474	-62.0640
	100-order	**0.8871**	**-85.0005**	**0.9453**	**-91.1058**	**0.9504**	**-61.1717**

Interpretation: cross-window dependence helps detect deviations from temporally coherent context.

Results: time-series forecasting

❖ Average over horizons H ∈ {24, 48, 168, 336, 720}; lower MSE / MAE is better

Method		Informer†		GRU-NVP†		DeepAR†		VRAE†		HoT-VI				Copula-SVI *(Ours)*							
										3-order		10-order		3-order		10-order		50-order		100-order	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	24	0.577	0.549	3.540	0.733	1.166	0.836	0.743	0.762	0.543	0.505	0.363	0.376	0.521	0.472	0.344	0.349	0.330	0.337	**0.321**	**0.326**
	48	0.685	0.625	2.549	0.622	1.154	0.827	0.826	0.801	0.578	0.528	0.392	0.392	0.545	0.487	0.370	0.373	0.348	0.354	**0.341**	**0.350**
	168	0.931	0.752	3.831	0.774	1.083	0.778	1.070	0.938	0.721	0.615	0.510	0.464	0.684	0.568	0.489	0.442	0.472	0.417	**0.462**	**0.413**
	336	1.128	0.873	6.877	1.008	1.043	0.766	1.199	1.016	0.883	0.702	0.616	0.525	0.817	0.671	0.575	0.494	0.543	0.465	**0.537**	**0.459**
	720	1.215	0.896	5.377	1.060	1.075	0.795	1.426	1.164	1.021	0.781	0.763	0.630	0.959	0.726	0.704	0.582	0.663	0.551	**0.643**	**0.537**
ETTm1	24	0.453	0.444	0.605	0.437	1.360	0.871	0.687	0.646	0.409	0.417	0.253	0.298	0.385	0.387	0.242	0.278	0.233	0.270	**0.226**	**0.263**
	48	0.494	0.503	2.787	0.701	1.334	0.866	0.817	0.724	0.535	0.488	0.330	0.345	0.497	0.453	0.310	0.325	0.294	0.315	**0.287**	**0.306**
	168	0.678	0.614	4.212	0.824	1.170	0.838	0.853	0.794	0.578	0.521	0.368	0.373	0.547	0.492	0.349	0.346	0.333	0.333	**0.323**	**0.324**
	336	1.056	0.786	5.062	1.019	1.249	0.846	1.091	0.975	0.641	0.567	0.434	0.415	0.605	0.526	0.410	0.396	0.386	0.382	**0.377**	**0.369**
	720	1.192	0.926	5.799	1.075	1.075	0.770	1.165	0.996	0.737	0.626	0.528	0.474	0.686	0.583	0.501	0.441	0.475	0.427	**0.462**	**0.419**
Electricity	24	0.312	0.387	3.514	1.844	0.211	0.330	0.279	0.396	0.256	0.346	0.134	0.238	0.240	0.330	0.124	0.222	0.120	0.209	**0.117**	**0.206**
	48	0.392	0.431	3.318	1.786	0.332	0.398	0.317	0.410	0.277	0.363	0.152	0.255	0.266	0.347	0.140	0.244	0.135	0.233	**0.133**	**0.229**
	168	0.515	0.509	3.482	1.833	1.065	0.811	0.366	0.475	0.303	0.382	0.174	0.273	0.283	0.363	0.162	0.255	0.154	0.240	**0.150**	**0.236**
	336	0.759	0.625	3.921	1.941	1.040	0.795	0.402	0.515	0.319	0.395	0.194	0.293	0.301	0.363	0.183	0.281	0.175	0.271	**0.172**	**0.266**
	720	0.969	0.788	4.232	2.020	1.048	0.804	0.450	0.556	0.348	0.416	0.230	0.323	0.331	0.383	0.217	0.299	0.207	0.281	**0.204**	**0.273**
Exchange	24	0.611	0.626	1.557	0.877	1.328	0.692	0.140	0.310	0.093	0.227	0.033	0.126	0.053	0.145	0.028	0.108	0.026	0.101	**0.025**	**0.097**
	48	0.680	0.644	1.589	0.883	1.345	0.701	0.238	0.435	0.171	0.306	0.058	0.164	0.082	0.163	0.049	0.141	0.046	0.134	**0.043**	**0.128**
	168	1.097	0.825	1.663	0.903	1.434	0.745	0.642	0.703	0.368	0.458	0.196	0.326	0.226	0.311	0.165	0.275	0.155	0.254	**0.150**	**0.242**
	336	1.672	1.036	1.682	0.905	1.489	0.778	1.050	0.953	1.165	0.821	0.496	0.515	0.576	0.556	0.410	0.432	0.385	0.408	**0.367**	**0.390**
	720	2.478	1.310	1.748	0.928	1.526	0.793	3.003	1.593	2.029	1.090	1.508	0.857	1.368	0.823	1.247	0.744	1.180	0.703	**1.139**	**0.681**
Weather	24	0.162	0.235	1.222	0.909	0.205	0.250	0.227	0.315	0.186	0.281	0.129	0.179	0.174	0.262	0.123	0.167	0.117	0.158	**0.116**	**0.156**
	48	0.348	0.400	2.319	1.287	0.229	0.267	0.449	0.495	0.291	0.361	0.186	0.230	0.271	0.345	0.174	0.215	0.166	0.203	**0.164**	**0.197**
	168	0.444	0.463	2.174	1.165	0.344	0.343	0.563	0.648	0.429	0.486	0.294	0.313	0.406	0.453	0.271	0.289	0.256	0.277	**0.253**	**0.268**
	336	0.578	0.523	2.119	1.221	0.568	0.527	0.781	0.841	0.625	0.575	0.550	0.430	0.599	0.545	0.514	0.404	0.487	0.386	**0.482**	**0.380**
	720	1.059	0.741	2.621	1.303	**0.571**	0.533	1.125	1.058	0.808	0.653	0.772	0.510	0.761	0.623	0.710	0.488	0.678	0.470	0.658	**0.461**
Average		0.819	0.660	3.112	1.122	0.978	0.678	0.796	0.741	0.573	0.516	0.387	0.373	0.487	0.455	0.352	0.344	0.335	0.327	**0.326**	**0.319**

Key message: higher-order instance dependence improves long-horizon predictive representations.

Results: constrained clustering and ablations

- **Constrained clustering:** constraints induce posterior dependence across instances
- Copula-SVI propagates constraint information through learned dependence, not only direct edges

Dataset	Metric	VaDE	C-IDEC	DGG†	DC-GMM	TreeVI	HoT-VI		Copula-SVI *(Ours)*			
							5-order	10-order	5-order	10-order	50-order	100-order
MNIST	ACC	89.0±5.0	96.3±0.2	95.8±0.1	96.5±0.2	97.4±0.3	98.3±0.4	98.5±0.3	98.3±0.3	98.6±0.2	98.8±0.3	**99.0±0.3**
	NMI	82.8±3.0	91.8±1.0	91.2±0.2	91.4±0.3	93.1±0.6	94.2±0.3	94.6±0.3	95.6±0.4	96.2±0.3	96.4±0.3	**96.5±0.4**
	ARI	80.9±3.0	92.1±0.4	91.4±0.3	92.5±0.5	93.7±0.7	95.2±0.5	95.6±0.5	96.4±0.2	96.7±0.1	96.9±0.4	**97.0±0.2**
fMNIST	ACC	55.1±2.2	68.1±3.0	79.9±0.4	80.5±0.8	81.4±0.6	83.2±0.5	83.4±0.4	83.6±0.3	83.8±0.3	84.2±0.4	**84.4±0.3**
	NMI	57.9±2.7	66.7±2.0	70.1±0.3	72.0±0.4	73.9±0.6	74.8±0.5	75.1±0.4	75.2±0.3	75.6±0.2	75.9±0.2	**76.1±0.3**
	ARI	41.6±3.1	52.3±3.0	64.9±0.3	66.4±0.5	67.9±0.9	69.1±0.5	69.2±0.4	69.4±0.1	69.5±0.2	69.9±0.3	**70.0±0.2**
Reuters	ACC	76.0±0.7	94.7±0.6	93.5±0.6	95.4±0.2	95.9±0.6	97.2±0.6	97.6±0.5	97.5±0.2	98.0±0.4	98.2±0.3	**98.4±0.2**
	NMI	50.1±1.3	81.4±0.7	81.2±0.8	82.7±0.7	83.4±0.5	85.1±0.6	85.4±0.5	85.6±0.2	85.8±0.2	86.3±0.2	**86.4±0.5**
	ARI	58.0±1.4	87.7±0.9	87.8±0.5	89.0±0.6	90.2±0.4	91.6±0.5	92.0±0.4	91.8±0.3	92.5±0.3	92.6±0.4	**92.8±0.3**
STL-10	ACC	77.3±0.5	81.6±3.8	89.9±0.3	89.5±0.5	90.4±0.9	92.2±0.6	92.4±0.4	92.7±0.2	92.8±0.2	93.0±0.3	**93.3±0.3**
	NMI	70.6±0.4	77.3±1.7	80.9±0.5	80.2±0.7	81.3±0.8	82.8±0.6	83.1±0.5	83.0±0.3	83.4±0.3	83.8±0.3	**84.0±0.2**
	ARI	62.7±0.4	71.8±3.4	79.0±0.4	78.4±0.9	79.5±0.7	81.3±0.5	81.5±0.5	81.5±0.3	81.9±0.2	82.2±0.3	**82.3±0.4**

Method	ACC	NMI	ARI
Factorized	96.5±0.2	91.4±0.3	92.5±0.5
Factorized+SVGD	96.8±0.2	91.8±0.3	93.1±0.4
Copula-SVI	**98.6±0.2**	**96.2±0.3**	**96.7±0.1**

SVGD alone helps only modestly; the explicit copula posterior is needed to preserve refined dependence.

Takeaways and outlook

- **Summary:**
  - Copula posterior separates local marginals from instance-level dependence
  - Regular vine gives expressive high-order dependence through pair-copulas
  - Edge-minibatching + level-wise sampling make full-vine learning practical
  - Sparse-vine initialization + SVGD refinement reduce cost while preserving dependence
- **Limitations / Future:**
  - Automatic pair-copula family selection and adaptive vine order
  - Combine instance-level copulas with dimension-level structured posteriors
  - More efficient GPU implementation for very large correlated batches