

# *Skeptical review: Intrinsic Dimensionality of PINN Latent Spaces for Burger’s Equation: Evidence for a Renormalization Group-like Flow*

---

## Summary

The manuscript analyzes how the geometry of a PINN’s internal (10D) latent representations varies with viscosity  $\nu$  for a Burgers-equation setup. For each of 25 viscosities ( $\nu \approx 0.01$  to  $1.0$ ), the authors extract latent vectors  $L(x, t; \nu)$  on a fixed  $(x, t)$  grid ( $101 \times 103 \approx 10,403$  points per  $\nu$ ; Sec. 2.1–2.2, Sec. 3.1) and estimate intrinsic dimensionality (ID) per viscosity slice via TwoNN, with a Levina–Bickel-style kNN MLE estimator mentioned for validation (Sec. 2.3). The main empirical finding is a pronounced non-monotonic ID– $\nu$  relationship (Sec. 3.2–3.3): ID is near the latent embedding dimension at very low  $\nu$ , peaks at intermediate  $\nu$  (reported up to  $\sim 40$ ), and then decreases at high  $\nu$ . The paper interprets the high- $\nu$  decrease as diffusion-driven simplification and argues the overall behavior is suggestive of an RG-like coarse-graining “flow” in latent space (Sec. 2.5, Sec. 3.4, Sec. 4). While the question is timely and the dataset construction is systematic, the current presentation leaves major concerns about reproducibility (PDE/PINN/training details), the validity/meaning of ID estimates that exceed the 10D embedding, the impact of grid-correlated sampling on neighbor-based estimators, robustness across trainings/architectures/layers, and the extent to which the RG narrative is supported versus being a heuristic analogy.

## Strengths

- Focused and timely question: connecting representation geometry/complexity in PINNs to a physically meaningful control parameter (viscosity  $\nu$ ) (Sec. 1, Sec. 3.3–3.4).
- Systematic data construction: for each  $\nu$ , a large latent point cloud ( $\sim 10k$  points in  $\mathbb{R}^{10}$ ) is extracted on a consistent grid, enabling statistical analyses (Sec. 2.1–2.2, Sec. 3.1).
- Uses established ID estimators (TwoNN and an MLE kNN variant) and attempts basic trend/correlation modeling and residual checks (Sec. 2.3–2.4, Sec. 3.2–3.3).
- Clearly documents a non-monotonic ID– $\nu$  pattern including a high- $\nu$  downturn, which is potentially scientifically meaningful for interpretability/complexity questions (Sec. 3.2–3.3).
- Physical intuition for the high- $\nu$  decrease (diffusion smoothing) is plausible, and the RG-style framing—if properly qualified or operationalized—could be a stimulating angle for future work (Sec. 2.5, Sec. 3.4, Sec. 4).
- Internal dataset-shape arithmetic appears consistent ( $101 \times 103 = 10,403$ ; latent features indexed separately from  $(x, t, \nu)$ ), which helps track what is being analyzed (Sec. 2.2–2.3).

## Major issues

1. **Core methodological gap: the PDE specification and PINN/training setup are not described at a level that allows reproduction or interpretation, and it is unclear whether the reported ID- $\nu$  effect is physical or an artifact of training/conditioning (Sec. 1, Sec. 2.1–2.2, Sec. 3.1).** In particular, the manuscript does not unambiguously state (i) the exact Burgers equation used (scalar vs vector, spatial dimensionality, forcing), domain, nondimensionalization, and initial/boundary conditions; (ii) whether a single conditional PINN is trained across all viscosities (with  $\nu$  as an input) or 25 separate models are trained; (iii) network architecture details (depth/width/activations; where the 10D latent is taken—bottleneck vs intermediate layer); (iv) loss terms and weights (PDE residual vs IC/BC vs data), collocation strategy, optimizer/schedule, epochs/stopping; and (v) solution accuracy versus a reference solver across  $\nu$ .

*Recommendation:* Expand Sec. 2.1–2.2 into a fully specified experimental protocol: write the explicit PDE(s) and all IC/BCs; define the domain and whether  $\nu$  is dimensionless; state clearly whether you train one conditional model across  $\nu$  or multiple models and how  $\nu$  enters the network; provide the full architecture (including the latent layer location and why 10D was chosen); provide the exact loss and weights, sampling of collocation points, optimizer/schedule, training length and stopping; and report PINN accuracy per  $\nu$  (e.g.,  $L_2 / L_\infty$  error against a numerical solver, and/or PDE residual statistics). Without this, the latent-space analysis cannot be meaningfully evaluated.

2. **Validity/meaning of “intrinsic dimensionality” results is currently not credible because reported IDs substantially exceed the latent embedding dimension (e.g.,  $\sim 40$  in  $\mathbb{R}^{10}$ ) without rigorous justification, debugging/sanity checks, or uncertainty quantification (Sec. 2.3, Sec. 3.2–3.3).** Under standard manifold definitions, ID cannot exceed the ambient dimension; persistent  $\text{ID} > 10$  strongly suggests estimator/pathology issues (implementation error, preprocessing/metric problems, density inhomogeneity, duplicates, boundary effects) or that the quantity is being used as an “effective complexity index” rather than intrinsic dimension.

*Recommendation:* Strengthen Sec. 2.3 and Sec. 3.2–3.3 with (i) an explicit definition of what you claim to measure—either true intrinsic/manifold dimension (and then explain why  $> 10$  can occur only as estimator failure/bias) or explicitly rename the output as an “effective dimension/complexity proxy” when it exceeds 10; (ii) a pipeline sanity check on synthetic datasets embedded in  $\mathbb{R}^{10}$  with known intrinsic dimension and comparable sample size (e.g., linear subspaces, noisy spheres, Swiss roll), reporting bias/variance and frequency of  $\text{ID} > 10$ ; (iii) bootstrap/jackknife/subsampling uncertainty estimates for each  $\nu$  (error bars/bands on  $ID(\nu)$ ); and (iv) explicit imple-

mentation details (distance metric, tie/duplicate handling, numerical precision, nearest-neighbor algorithm). If  $ID > 10$  persists, you must frame it carefully and corroborate the trend with additional measures (see below).

3. **Neighbor-based ID estimators are applied to highly structured, grid-sampled point clouds ( $101 \times 103$  evaluations of a smooth map  $(x, t) \mapsto L$ ), violating i.i.d. sampling assumptions and potentially inducing strong biases due to spatial/temporal correlations, anisotropic spacing, boundary effects, and near-duplicate latent vectors (Sec. 2.2–2.3, Sec. 3.1–3.2).** This could create spurious non-monotonicity or inflate estimates (including  $ID > 10$ ).

*Recommendation:* Add robustness checks targeted at sampling structure (Sec. 3.2): (i) compute ID after random subsampling (e.g., 10%, 25%, 50%, 75%) and after decorrelated sampling (e.g., farthest-point sampling in latent space, or stratified sampling across  $(x, t)$ ); (ii) evaluate latent vectors at off-grid  $(x, t)$  (jittered or random points) to test sensitivity to grid regularity; (iii) report distance histograms, minimum-distance/duplicate rates, and whether activations saturate in some regimes; and (iv) confirm that the qualitative  $ID-\nu$  curve (peak and high- $\nu$  downturn) survives these controls.

4. **The central phenomenon is presented as a property of “the PINN latent space,” but experiments appear to rely on a single network instance/latent layer/dimension, with no robustness across random seeds, architectures, latent sizes, or even layer choice (Sec. 2.1–2.3, Sec. 3.1–3.3).** This makes it unclear whether the non-monotonic curve and peak location are generic or accidental (optimization artifact/local minimum, hyperparameter effect, layer-specific geometry).

*Recommendation:* Add an explicit robustness section (Sec. 3.2–3.3): (i) retrain with multiple random seeds and show  $\text{mean} \pm \text{std } ID(\nu)$ ; (ii) test at least one alternative architecture or latent dimensionality (e.g., 5D/20D) and/or extract latents from different layers to see whether the trend is stable; (iii) if compute is limited, do these tests on a subset of viscosities (low/mid/high) but report variability. Clearly separate which conclusions are stable versus model-dependent.

5. **The paper does not establish a quantitative link between  $ID(\nu)$  and either (a) physical complexity of the underlying Burgers solutions or (b) training/approximation difficulty, leaving the main interpretation underdetermined (Sec. 3.1–3.4, Sec. 4).** The RG-like narrative especially requires distinguishing “physics-driven simplification” from “network-driven representation changes.”

*Recommendation:* Augment Sec. 3.1–3.4 with correlational analyses against: (i) physical metrics computed from a reference solver or high-quality PINN output (e.g., gradient norms/total variation, shock indicators, spectral energy vs wavenumber, enstrophy-like measures depending on the PDE form); and (ii) learning/fit metrics (PINN error vs  $\nu$ , PDE residual norms, BC/IC residuals). Additionally, include at least one conceptually different latent complexity measure (e.g., PCA participation ratio/effec-

tive rank, local PCA dimension, or singular-value decay) to see whether the *trend* (especially high- $\nu$  decrease) agrees across metrics. Use these to argue whether ID is tracking physical degrees of freedom or training pathologies.

6. **The RG-like “flow” interpretation is currently presented too strongly relative to the evidence: no explicit coarse-graining transformation, scale analysis, semigroup/composition property, or fixed-point-like behavior is demonstrated (Sec. 2.5, Sec. 3.4, Sec. 4).** As written, the data mainly support “non-monotonic representational complexity vs  $\nu$ ” plus a plausible diffusion-smoothing intuition at high  $\nu$ .

*Recommendation:* Either (A) operationalize the RG analogy with at least one concrete test (Sec. 3.4): define an explicit coarse-graining on inputs/solutions (spatial filtering/downsampling) and track how latent representations and their effective dimension change under that map, or test whether latent statistics exhibit a flow with approximate composition across  $\nu$ ; or (B) reframe the RG discussion as a heuristic analogy, explicitly acknowledging alternative explanations (training difficulty, saturation, estimator artifacts) and moderating Abstract/Sec. 4 language to “suggestive/consistent with” rather than implying an RG mechanism has been established.

7. **Foundational notation/model mismatch: the manuscript calls the problem “2D Burgers equation,” but the data/notation indicate only one spatial coordinate (101 points in  $x$ ) plus time (Sec. 2.1–2.2; notation  $L(x, t; \nu)$ ).** If the PDE is truly 2D in space, the setup is missing  $y$  and the field definition (scalar vs vector velocity). If it is 1D-in-space Burgers, the term “2D” is misleading.

*Recommendation:* Make the PDE dimensionality consistent throughout. If it is 1D spatial Burgers, rename accordingly and use  $u(x, t)$ . If it is 2D spatial Burgers, define  $u(x, y, t)$  (or  $(u, v)$ ), specify the  $y$ -grid and domains, and update dataset tensor shapes/notation (Sec. 2.1–2.2, Sec. 3.1).

8. **The statistical modeling in Sec. 2.4 and Sec. 3.3 emphasizes monotone summaries (global Spearman  $\rho$ , linear/log fits) despite the key result being strongly non-monotonic (peak + downturn).** This risks mischaracterizing the main phenomenon and does not quantify the peak location/uncertainty or the significance of the high- $\nu$  decrease.

*Recommendation:* Revise Sec. 2.4 and Sec. 3.3 to treat monotone fits as baselines only, and add non-monotone modeling and inference: spline/GP smoothing, quadratic or piecewise-linear change-point models, and tests comparing monotone vs non-monotone fits (information criteria). Report uncertainty on peak location and on the high- $\nu$  downturn using the ID uncertainty estimates (bootstrap bands). Consider reporting rank correlations separately on low $\rightarrow$ mid and mid $\rightarrow$ high  $\nu$  regimes.

## Minor issues

1. ID-estimator methodology is under-specified and includes an internal inconsistency: Sec. 3.2 refers to “methods section (not provided here),” and formulas/implementation details for TwoNN and the MLE estimator are not fully given (Sec. 2.3, Sec. 3.2).

*Recommendation:* Make Sec. 2.3 self-contained: include the explicit TwoNN relation (definition of  $\mu_i = r_{i,2}/r_{i,1}$  and how ID is fit/aggregated), the explicit Levina–Bickel MLE formula, neighborhood  $k$  range and selection/aggregation strategy, distance metric, and tie handling. Remove or rewrite “not provided here” to point to Sec. 2.3.

2. Critical experimental settings and preprocessing choices that can materially alter neighbor distances are not clearly reported: exact  $\nu_k$  values and spacing (linear vs log), whether all grid points are used, and whether latent vectors are centered/standardized/whitened before ID estimation (Sec. 2.1–2.3).

*Recommendation:* In Sec. 2.1–2.3, explicitly list  $\nu_k$  values (or provide a table in an appendix), domain ranges, grid details, and latent preprocessing (none/standardization/whitening). State explicitly that ID is computed on the 10 latent coordinates only (excluding  $\mathbf{x}, \mathbf{t}, \nu$ ).

3. Validation via the MLE-based estimator is mentioned but not reported consistently alongside TwoNN, limiting the reader’s ability to judge agreement and robustness (Sec. 2.3, Sec. 3.2–3.3).

*Recommendation:* Provide side-by-side TwoNN vs MLE results across  $\nu$ : overlay plots, scatter  $\text{ID} \text{ vs } \text{TwoNN}$ , and a small table at representative  $\nu$  (low/mid/high). Discuss discrepancies and which conclusions are estimator-invariant.

4. Positioning within broader literature is limited: prior work on intrinsic/effective dimension in neural representations, and prior RG–deep learning connections, are only lightly engaged (Sec. 1, Sec. 2.5).

*Recommendation:* Add a short related-work subsection in Sec. 1 summarizing (i) intrinsic/effective dimension in deep nets, (ii) representation analysis for PINNs/scientific ML, and (iii) RG-inspired views of deep learning, clearly stating what is novel here (parameter sweep over  $\nu$  with latent-space geometry analysis).

5. Presentation of results would be more actionable with data availability at the figure level: Figure 1 is central but the paper does not provide the underlying  $(\nu_k, \mathbf{ID}_k)$  values, nor confidence intervals (Sec. 3.2–3.3).

*Recommendation:* Include an appendix table (or supplementary file) with  $\nu_k$ , ID estimates (TwoNN and MLE), and uncertainty intervals. If possible, release code for latent extraction and ID estimation.

6. Computational details are missing for nearest-neighbor calculations on  $\sim 25 \times 10k$  points, which affects reproducibility and may affect results if approximate NN is used (Sec. 2.3).

*Recommendation:* State the library/algorithm used for NN search (exact vs approximate), runtime/memory, and any acceleration (KD-tree/FAISS). Briefly discuss scaling with dataset size.

7. Some regression reporting appears inconsistent with narrative endpoint values (e.g., fitted linear/log models vs stated ID at  $\nu = 0.01$  and  $\nu = 1.0$ ); log base is not specified (Sec. 3.3).

*Recommendation:* Specify the log base, clarify whether  $\nu$  is rescaled/centered before fitting, and reconcile coefficients with plotted/narrative endpoints. Consider moving detailed fit coefficients to a table with standard errors.

## Very minor issues

1. Typos/formatting inconsistencies reduce polish: “Burger’s” vs “Burgers/Burgers’,” spacing artifacts (e.g., line breaks in words), inconsistent section-heading formatting (Sec. 1, Sec. 2.2, Sec. 3.4, Sec. 4).

*Recommendation:* Proofread and standardize naming (prefer “Burgers equation” or “Burgers’ equation”), remove line-break artifacts, and normalize headings and spacing around symbols.

2. Notation could be made more explicit:  $r_{i,1}, r_{i,2}, \mu_i, \rho, ID_k$ , and the norm/metric are not all cleanly defined in one place (Sec. 2.3).

*Recommendation:* Add a short notation block defining all symbols and explicitly state the metric (typically Euclidean) used for neighbor distances.

3. Some discussion sentences (especially around non-monotonicity and RG analogy) are long/dense (Sec. 3.3–3.4, Sec. 4).

*Recommendation:* Split long sentences and use short enumerations when listing alternative mechanisms/interpretations to improve readability.

## Key statements and references

- • **The intrinsic dimensionality (ID) of each latent space point cloud was estimated using the Two Nearest Neighbors (TwoNN) algorithm, which infers a single global ID from the distribution of ratios  $\mu_i = r_{i,2}/r_{i,1}$  between distances to the first and second nearest neighbors of each point, and this estimator can yield values exceeding the embedding dimension for finite**

datasets or highly curved manifolds, so the resulting IDs are interpreted as quantitative measures of latent manifold complexity rather than strict geometric dimensions.

- *Reference(s)*: Levina and Bickel
- • To validate the TwoNN-based intrinsic dimensionality estimates, a Maximum Likelihood Estimator (MLE) based on  $k$ -nearest neighbor distances, specifically the method proposed by Levina and Bickel, was also applied, with an appropriate stable range of  $k$  determined by analyzing the stability of ID estimates as  $k$  varied for a representative subset of the data and then aggregating over that stable range.
- *Reference(s)*: Levina and Bickel
- • The overall association between viscosity and intrinsic dimensionality was quantified using Spearman's rank correlation coefficient between the ordered viscosity values and the corresponding ID estimates, yielding Spearman's  $\rho \approx 0.8592$  with a  $p$ -value  $\approx 0.0$ , indicating a statistically significant, strongly positive monotonic relationship despite the visually evident non-monotonic behavior at high viscosities.
- *Reference(s)*: Spearman's

## Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

**Maths relevance:** light

The paper is largely methodological/interpretive with limited explicit mathematics: tensor-shape bookkeeping, definitions for nearest-neighbor distance ratios in TwoNN, and simple regression model forms (linear/log). The main internal-consistency risks arise from definitional mismatches (intrinsic dimensionality exceeding embedding dimension) and variable/dimensionality notation (calling the PDE 2D while using  $L(x, t; \nu)$  and a 1D spatial grid).

### Checked items

1. ✓ **Feature-count consistency of raw array** (Sec. 2.1, p.2)
  - **Claim:** Raw NumPy array has shape (101,103,25,13) where the 13 features are  $x$ -mesh,  $t$ -mesh,  $\nu$ , and 10 latent components.
  - **Checks:** symbol/definition consistency, dimensional counting
  - **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:** Feature dimension concatenates scalar fields per  $(x, t, \nu)$  location, Latent dimension is 10
  - **Notes:**  $13 = 3 + 10$  is consistent with the stated feature contents.

2. ✓ **Latent-component indexing count** (Sec. 2.1, p.2)

- **Claim:** Latent vectors are located in indices 3 through 12 of the feature dimension and comprise 10 components.
- **Checks:** index arithmetic, definition consistency
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** Indices are inclusive and zero-based is not explicitly stated
- **Notes:** If indices are inclusive (3,4,5,6,7,8,9,10,11,12), that is 10 entries. The paper should still clarify zero- vs one-based indexing to remove ambiguity.

3. ✓ **Reshape to per-viscosity point cloud** (Sec. 2.2, p.2)

- **Claim:** For fixed viscosity slice  $k$ , latent data  $(101, 103, 10)$  reshapes to a point cloud  $L_k$  of size  $(101 \times 103, 10) = (10403, 10)$ .
- **Checks:** algebra/arithmetic, dimensional consistency
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** All  $(x, t)$  pairs are stacked into rows
- **Notes:**  $101 \times 103 = 10403$  is correct, and the reshape is dimensionally consistent.

4. ✓ **Viscosity index-set notation** (Sec. 2.2, p.2)

- **Claim:** There are 25 unique viscosities  $\nu_k$  with  $k = 0 \dots 24$ , constant over  $(x, t)$  within each slice.
- **Checks:** notation consistency, definition consistency
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** Third tensor axis indexes viscosity slices
- **Notes:** The set/indexing is consistent with the stated array shape and later usage of  $\nu_k$  and  $ID_k$ .

5. ✓ **TwoNN ratio definition** (Sec. 2.3, p.3)

- **Claim:** For point  $p_i$ , let  $r_{i,1}$  and  $r_{i,2}$  be distances to the first and second nearest neighbors; define  $\mu_i = r_{i,2}/r_{i,1}$ .
- **Checks:** definition consistency, sanity constraints
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** Distances are positive and  $r_{i,2} \geq r_{i,1} > 0$  for distinct points
- **Notes:** The ratio is dimensionless and well-defined if  $r_{i,1} > 0$ . The paper does not specify how duplicates/ $r_{i,1} = 0$  are handled (left ambiguous).

6. △ **Missing explicit global-ID estimator formula** (Sec. 2.3, p.3)

- **Claim:** TwoNN estimates a global ID from the distribution of  $\mu_i$  values across points.

- **Checks:** derivation completeness, verifiability
- **Verdict:** UNCERTAIN; confidence: high; impact: moderate
- **Assumptions/inputs:** A specific mapping from  $\mu_i$  to  $ID_k$  is used
- **Notes:** No explicit estimator equation/fit procedure is given (e.g., what is plotted/fitted and how  $ID_k$  is extracted). This prevents auditing algebra/logic of the ID computation within the paper.

7. ✘ **Claim that ID estimates can exceed embedding dimension and be interpreted as complexity** (Sec. 3.2, p.4)

- **Claim:** TwoNN can yield ID values  $> 10$  for a 10D latent space and these are interpreted as complexity/effective degrees of freedom rather than geometric dimension.
- **Checks:** definition consistency, logical consistency
- **Verdict:** FAIL; confidence: high; impact: critical
- **Assumptions/inputs:** Latent vectors lie in  $\mathbb{R}^{10}$ , ID earlier defined as manifold/minimum-variable dimension
- **Notes:** Internally inconsistent: if ID is the intrinsic/manifold dimension of data embedded in  $\mathbb{R}^{10}$ , it cannot exceed 10. The paper can still report an estimator output  $> 10$ , but must not call it 'intrinsic dimensionality' under the stated definition without redefining the quantity or providing a clear finite-sample/noise interpretation.

8. ✘ **Dimensionality/variable mismatch: '2D Burgers' vs  $L(x, t; \nu)$  and 1D grid** (Title; Sec. 2.1–2.2 (p.2); Sec. 3.1 (p.4))

- **Claim:** The PINN solves the 2D Burger's equation, but the notation and dataset indicate dependence on only  $x$  and  $t$  (101 spatial points, no  $y$ ).
- **Checks:** notation consistency, model/variable consistency
- **Verdict:** FAIL; confidence: high; impact: critical
- **Assumptions/inputs:** '2D' refers to spatial dimensionality
- **Notes:** Calling the PDE 2D is inconsistent with the stated variable dependence and data structure (only one spatial axis described). If '2D' means something else (e.g., two-component velocity), it is not defined and conflicts with  $L(x, t; \nu)$  notation.

9. ✔ **Linear model form** (Sec. 3.3, p.4)

- **Claim:** A linear fit uses  $ID = a \cdot \nu + b$ .
- **Checks:** algebraic form, dimensional sanity
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** ID and  $\nu$  are scalar-valued

- **Notes:** Form is algebraically consistent. Units are not specified; if  $\nu$  has physical units, then  $a$  carries reciprocal units and  $b$  matches ID's units (dimensionless).
10. ✓ **Logarithmic model form and domain** (Sec. 3.3, p.4)
- **Claim:** A logarithmic fit uses  $ID = a \cdot \log(\nu) + b$  over  $\nu$  in  $[0.01, 1.0]$ .
  - **Checks:** domain check, notation clarity
  - **Verdict:** PASS; confidence: medium; impact: minor
  - **Assumptions/inputs:**  $\nu > 0$
  - **Notes:** Domain is valid since  $\nu > 0$ . Log base is unspecified (minor ambiguity affecting parameter interpretation).
11. ✗ **Self-reference inconsistency about missing Methods discussion** (Sec. 3.2, p.4)
- **Claim:** The text says the relevant Methods discussion is 'not provided here' while the manuscript includes a Methods section.
  - **Checks:** internal cross-reference consistency, verifiability
  - **Verdict:** FAIL; confidence: high; impact: moderate
  - **Assumptions/inputs:** The provided text is the complete paper as attached
  - **Notes:** This cross-reference undermines internal completeness: the claimed justification for  $ID > 10$  is not actually present in Methods, blocking an in-paper audit of that key point.

### Limitations

- The attached content contains no explicit PDE form of the (2D) Burgers equation, no PINN loss function, and no explicit TwoNN/MLE estimator equations; therefore, core derivations cannot be audited beyond definition/notation consistency.
- Figure 1 is referenced for trends and fits, but the audit does not verify numeric values or plotted points (out of scope).
- Because the estimator implementation steps are not specified (metric, normalization, duplicate handling), only high-level definitional checks were possible for the ID methodology.

## Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

All automated internal arithmetic and consistency checks C1–C10 passed. Verified items include feature-count/index consistency for the 13-feature array, grid product/reshape consistency yielding 10,403 points per viscosity, inclusive index-based viscosity count (25),  $\nu$  range sanity (0.01 to 1.0), plausibility of Spearman  $p$ -value rounding to  $\sim 0.0$  for  $\rho \approx 0.8592$  with  $n =$

25,  $R^2$  bounds checks, and agreement between text and figure caption for peak  $\nu$  (0.383 vs  $\sim 0.4$  within tolerance). However, computed endpoint predictions from the stated regression coefficients highlight potential definition/interpretation mismatches versus narrative endpoint IDs.

### Checked items

1. ✓ **C1** (p.2 §2.1 (Data Acquisition and Structure))
  - **Claim:** Raw data was provided in a NumPy array of dimensions (101,103,25,13). Feature dimension contains  $x$  mesh,  $t$  mesh, viscosity value, and the 10 components of the latent vector located in indices 3 through 12.
  - **Checks:** dimension/feature-count consistency
  - **Verdict:** PASS
  - **Notes:** Inclusive index range 3..12 contains 10 elements; 3 non-latent +10 latent = 13 features; last index 12 equals 13 - 1.
2. ✓ **C2** (p.2 §2.2 and p.4 §3.1)
  - **Claim:** For each viscosity slice, raw latent space data dimensions (101,103,10) reshaped into point cloud of size (101  $\times$  103,10), yielding 10403 points in 10D for each of 25 viscosities.
  - **Checks:** parts-vs-product (grid size) and reshape consistency
  - **Verdict:** PASS
  - **Notes:** 101  $\times$  103 = 10,403; reshape to (10403,10) is consistent. Total points across viscosities computed as 260,075.
3. ✓ **C3** (p.2 §2.2)
  - **Claim:** Unique viscosity values set  $\nu_k$  from  $k = 0$  to 24, i.e., 25 unique viscosities.
  - **Checks:** index-range to count consistency
  - **Verdict:** PASS
  - **Notes:** Inclusive range 0..24 implies 25 values, matching the claim.
4. ✓ **C4** (p.4 §3.1)
  - **Claim:** Viscosity values  $\nu_k$  span a range from 0.01 to 1.0.
  - **Checks:** range ordering sanity check
  - **Verdict:** PASS
  - **Notes:**  $\nu_{\min} > 0$  and  $\nu_{\max} > \nu_{\min}$ ; computed ratio  $\nu_{\max}/\nu_{\min} = 100.0$ .
5. ✓ **C5** (p.4 §3.3)
  - **Claim:** Spearman's rank correlation coefficient reported as  $\rho \approx 0.8592$  with  $p$ -value  $\approx 0.0$ .
  - **Checks:**  $p$ -value plausibility given  $n$  and  $\rho$  (approximate)

- **Verdict:** PASS
  - **Notes:**  $t$ -approximation gives  $p \approx 3.82 \times 10^{-8}$  (two-sided,  $df = 23$ ), consistent with rounding to  $\approx 0.0$ .
6. ✓ **C6** (p.4 §3.3)
- **Claim:** Linear model fit:  $ID = a \cdot \nu + b$  with slope  $a \approx 22.06$  and intercept  $b \approx 17.80$ .
  - **Checks:** derived prediction check (endpoints)
  - **Verdict:** PASS
  - **Notes:** Arithmetic-only predictions:  $ID(0.01) = 18.0206$ ;  $ID(1.0) = 39.86$ . These can be compared qualitatively to narrative endpoint IDs.
7. ✓ **C7** (p.4 §3.3)
- **Claim:** Linear model  $R^2 \approx 0.313$ .
  - **Checks:** range check for  $R$ -squared
  - **Verdict:** PASS
  - **Notes:**  $R^2 = 0.313$  is within  $[0, 1]$ .
8. ✓ **C8** (p.4 §3.3)
- **Claim:** Logarithmic model fit:  $ID = a \cdot \log(\nu) + b$  with  $a \approx 6.57$ ,  $b \approx 37.95$ , and  $R^2 \approx 0.7206$ .
  - **Checks:** range check for  $R$ -squared and endpoint predictions under log model
  - **Verdict:** PASS
  - **Notes:**  $R^2 = 0.7206$  is within  $[0, 1]$ . Endpoint predictions computed: at  $\nu = 1.0$ , pred=37.95 for both  $\ln$  and  $\log_{10}$ ; at  $\nu = 0.01$ , pred $\approx 7.694$  ( $\ln$ ) or 24.81 ( $\log_{10}$ ).
9. ✓ **C9** (p.5 Figure 1 caption + p.4 §3.3)
- **Claim:** Peak ID occurs around  $\nu \approx 0.383$  (caption) / intermediate up to approximately  $\nu \approx 0.4$  (text).
  - **Checks:** approximate numeric agreement between sections
  - **Verdict:** PASS
  - **Notes:** Absolute difference 0.017 is within the heuristic  $\pm 0.03$  tolerance for ‘ $\approx$ ’.
10. ✓ **C10** (p.4 §3.3 and p.6 Conclusions)
- **Claim:** ID values: low  $\nu \approx 0.01$  gives ID around 10–11; peak near 40 at intermediate  $\nu$ ; high  $\nu = 1.0$  gives ID around 24.
  - **Checks:** range/ordering sanity check (narrative consistency)
  - **Verdict:** PASS

- **Notes:** Ordering holds for representative values:  $40 > 24 > 11$  and  $1.0 > 0.01$ .

### Limitations

- Only parsed text of the PDF was available; no underlying numerical tables of  $(\nu_k, ID_k)$  were provided, limiting verification of statistical results and model fits.
- Values shown only in Figure 1 cannot be used for numeric checks because extracting data from plot graphics/pixels is out of scope.
- Several claims (constancy of  $\nu_k$  across  $(\mathbf{x}, t)$ , counts of vectors, TwoNN/MLE computations) require access to the referenced NumPy dataset and code outputs, which are not included in the PDF text.