Pseudotime graph diffusion for post hoc visualization of inferred single-cell trajectories

Brandon Lukas[1], Jingbo Pang[2], Timothy J. Koh[2], Yang Dai[1]*

[1]Department of Biomedical Engineering, University of Illinois Chicago, Chicago, USA

[2]Department of Kinesiology and Nutrition, University of Illinois Chicago, Chicago, USA

Correspondence: yangdai@uic.edu

Abstract

Visual representations are widely used to interpret trajectories in single-cell data; however, they do not always faithfully capture inferred trajectory structure. As a result, interpretation of cellular dynamics and downstream analyses may be compromised. Here, we present Pseudotime Graph Diffusion (PGD), a lightweight and interpretable post hoc framework for smoothing cell-level features along pseudotime. PGD operates by performing random-walk diffusion on a pseudotime graph, propagating information along inferred trajectory paths to enhance continuity and structure. We demonstrate that PGD-smoothed embeddings improve visualization of increasingly complex inferred trajectories of monocytes and macrophages during wound healing. We further show that PGD extends naturally to trajectory-aware gene expression smoothing. By improving agreement between visual representations and inferred trajectories, PGD enables more faithful interpretation of dynamic cellular processes.

Keywords: RNA, biological processes, Markov processes, graph theory, information diffusion, smoothing methods

1 Introduction

Single-cell RNA sequencing (scRNA-seq) has enabled the study of dynamic cellular processes at single-cell resolution, providing insights into development, differentiation, and disease-associated state transitions. A central goal of single-cell analysis is to reconstruct continuous cellular trajectories and to order cells along developmental or differentiation axes, motivating the development of trajectory inference and pseudotime estimation methods [1–4].

*Funding: This study was supported by the National Institute of General Medical Sciences (NIGMS) through grant R35GM136228 to T.J.K., and by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Diabetic Complications Consortium through grants DK076169 and DK115255 to Y.D. and T.J.K.

In turn, these methods have prompted widespread use of low-dimensional visual representations to project inferred state transitions and to color cells by pseudotime. These representations play a critical role in data interpretation and hypothesis generation and, in some frameworks, serve as substrates from which quantities are derived [5].

Standard visualization methods such as UMAP often represent trajectories for well-ordered and coherent biological processes, such as embryonic development or pancreatic differentiation [6–8]. However, for less coherent systems, trajectory structures may be obscured. Moreover, when trajectory inference is performed in a higher-dimensional feature space (e.g., principal component space), low-dimensional representations may fail to reflect the inferred structure due to information loss. In such cases, this mismatch can complicate interpretation and downstream analysis.

Several approaches have been specifically proposed to improve trajectory representation in single-cell data. Methods such as PAGA [9] and PHATE [10] aim to encode global or diffusion-based structure in low-dimensions; however, they are not designed to operate in a post hoc setting where inferred trajectories have already been established. Alternatively, one could theoretically visualize a cell-cell transition matrix using any graph layout algorithm; however, this would abstract away cell-level features which could be of interest.

Here, we introduce Pseudotime Graph Diffusion (PGD), a post hoc framework for enhancing agreement between low-dimensional embeddings and existing inferred trajectories or pseudotime estimates. PGD treats a pseudotime graph as a diffusion substrate and smooths cell-level features, such as embeddings or gene expression, along this graph. By diffusing features in a trajectory-aware manner, PGD emphasizes trajectory structure while preserving aspects of the original features.

Using a murine wound-healing dataset, we show that PGD improves visualization of increasingly complex inferred trajectories and naturally extends to trajectory-aware smoothing of gene expression. Our contributions are summarized as follows:

- We introduce PGD, a lightweight post hoc framework that smooths cellular features along a pseudotime graph to enhance trajectory structure.
- We demonstrate that PGD improves the agreement between low-dimensional visual representations and inferred trajectories, particularly for complex branching structures.
- We show that PGD naturally extends to gene expression smoothing, reducing noise while preserving branch-specific trends.

2 Methods

2.1 Data and Preprocessing

To evaluate PGD, we analyzed a publicly available scRNA-seq dataset from the Gene Expression Omnibus (GEO; accession GSE203244), consisting of pooled monocytes and macrophages isolated from murine skin wounds at 3, 6, and 10 days post-injury [11].

Data were preprocessed using scanpy [12]. Count matrices from all timepoints were concatenated, and low-quality cells were filtered using median absolute deviation (MAD) thresholds applied to the number of detected genes, total counts, and mitochondrial read fraction. Counts were then normalized and log-transformed. Principal component analysis (PCA) was performed on highly variable genes, and the resulting PC embeddings were corrected for batch effects using Harmony [13].

2.2 Trajectory Inference with Lamian

Trajectory structure and pseudotime were inferred using Lamian, a statistical framework for pseudotemporal trajectory analysis [14]. Lamian constructs a cluster-based trajectory graph in a low-dimensional feature space and assigns pseudotime values along inferred paths.

As input, we provided Harmony-corrected PC embeddings along with cell-level metadata indicating sampling timepoints. The Lamian cluster with the largest proportion of cells from day 3 post-injury was specified as the pseudotime origin. To model varying branching complexities, Lamian was run across a range of cluster resolutions (5-10 clusters), and the resulting trajectory graphs and pseudotime assignments were used as downstream inputs for PGD.

2.3 Pseudotime Graph Diffusion

Fig. 1 is a schematic of PGD. Let $C = (c_1, \ldots, c_n)$ denote an ordering of n cells by inferred pseudotime, and let $X \in \mathbb{R}^{n \times d}$ denote the corresponding cell-level feature matrix with d features per cell. PGD constructs an undirected cell graph by connecting cells that are nearby along the pseudotime ordering. Specifically, the edge set is defined using a sliding window as

$$E = \left\{ \{c_i, c_j\} \;\middle|\; 0 < |i - j| \leq k,\ i < j \right\}, \tag{1}$$

where $k \geq 1$ is the window radius. For branched trajectories, this construction is applied separately within each branch-specific ordering, and the resulting edge sets are combined by union.

PGD updates cell features using a lazy random-walk diffusion on this graph. At diffusion step $t = 0, 1, 2, \ldots$, features are updated according to

$$X^{(t+1)} = (1 - \alpha) X^{(t)} + \alpha P X^{(t)}, \tag{2}$$

where $X^{(t)}$ denotes the feature matrix at step t, $A \in \mathbb{R}^{n \times n}$ is the symmetric adjacency matrix of the graph, D is the diagonal degree matrix with entries $D_{ii} = \sum_j A_{ij}$, and

$$P = D^{-1} A \tag{3}$$

is the random-walk propagation operator. The parameter $\alpha \in [0, 1]$ controls the diffusion strength by weighting contributions from neighboring cells relative to the original features. When A has no

self-loops, $PX^{(t)}$ corresponds to the local neighborhood mean of each cell's features.

Defining the combined propagation operator

$$\tilde{P} = (1 - \alpha)I + \alpha P, \tag{4}$$

the diffusion process can be written in closed form as

$$X^{(t)} = \tilde{P}^t X^{(0)}, \tag{5}$$

where the exponent t denotes the number of diffusion steps (matrix power), not transposition. Unless otherwise stated, we use $k = 50$, $\alpha = 0.5$, and a single diffusion step ($t = 1$).

2.4 Comparison with MAGIC

We compared PGD against MAGIC [15], a widely used diffusion-based smoothing and imputation method for scRNA-seq data. MAGIC performs diffusion over an expression-derived affinity graph, whereas PGD diffuses over a pseudotime graph. We applied MAGIC using its default parameters. For a controlled comparison, we configured PGD with the same number of diffusion steps ($t = 3$) and set the pseudotime window radius to $k = 8$. Because PGD connects each cell to neighbors in both directions along pseudotime (up to $2k$ neighbors), this setting yields a maximum neighborhood size that is comparable to MAGIC's default maximum of 15 neighbors.

3 Results

3.1 Monocyte and Macrophage Wound-Healing Dataset

After preprocessing, the dataset consisted of 2,848 monocytes and macrophages distributed across days 3 (34.4%), 6 (19.6%), and 10 (45.9%) post-injury (Fig. 2). Visualized using UMAP, the Harmony-corrected PC embeddings display a continuous spectrum of cell states without distinct clustering. While cells from different timepoints occupy generally distinct regions, suggestive of temporal progression, the standard UMAP layout lacks a clearly defined trajectory structure. This lack of visual coherence contrasts with more structured developmental systems, making this dataset an excellent test case for evaluating whether PGD can enhance the representation of trajectories inferred in higher-dimensional spaces.

3.2 PGD-Smoothed Embeddings Improve Visualization of Inferred Trajectories

We first evaluated whether PGD improves visual representations of inferred trajectories. Since UMAP is a standard approach for visualizing single-cell data, we applied PGD to Harmony-corrected

PC embeddings using Lamian-inferred trajectory graphs at varying cluster resolutions, then compared UMAP layouts generated before and after smoothing (Fig. 3).

Lamian inferred progressively more complex trajectory structures using five, six, seven, and nine clusters; however, the UMAP layout constructed from embeddings before PGD remained fixed across these settings. When cells were colored by Lamian cluster or pseudotime, this static layout did not clearly reflect the inferred trajectories, particularly at higher cluster resolutions (Fig. 3a). As a result, cells distant along pseudotime sometimes appeared proximal in the embedding, forcing a mismatch between the visual representation and the inferred dynamics.

In contrast, UMAP layouts constructed from embeddings after diffusion were specific to each inferred trajectory, exhibiting increased continuity and organization along inferred paths. When cells were colored by pseudotime, the layout displayed smoother and more coherent gradients, and separation between trajectory branches became visually apparent (Fig. 3b). Notably, in the nine-cluster trajectory where clusters 5 and 6 originally overlapped despite lying at opposite ends of the pseudotime continuum, PGD clearly separated them while keeping them within the same global neighborhood.

These results suggest that PGD preserves coarse-scale relationships from the original embeddings while enhancing trajectory-consistent structure. By diffusing low-dimensional embeddings along a pseudotime graph, PGD improves the agreement between visual representations and inferred trajectory structures, enabling more faithful interpretation of dynamic processes.

3.3 Effect of Diffusion Strength Parameter

We next examined how the diffusion strength parameter α influences the visual structure of smoothed embeddings. Using the Lamian-inferred trajectory with five clusters, we applied PGD to Harmony-corrected PC embeddings while incrementally increasing α from 0.05 to 0.80 and visualized the results using UMAP (Fig. 4).

The visualization response varied distinctly across three regimes. At low diffusion strengths ($\alpha < 0.30$), smoothing was subtle, and embeddings retained a geometry largely similar to the original input. At moderate strengths ($\alpha \approx 0.30$–0.50), trajectory structure became visually dominant; pseudotime gradients appeared smoother, and branches became more coherent. Conversely, at high strengths ($\alpha > 0.50$), the embeddings began to collapse into one-dimensional artifacts, abstracting away local cell-level variation in favor of the global trajectory backbone.

These results confirm that α offers an interpretable mechanism to tune the balance between data fidelity and trajectory emphasis. For most visualization tasks, intermediate values provide an effective compromise that enhances structural clarity without forcing over-simplification.

3.4 PGD Extends to Trajectory-Aware Smoothing of Gene Expression

Finally, we examined whether PGD can extend to smoothing gene expression along inferred trajectories. This is a useful task in single-cell analysis because scRNA-seq measurements are inherently sparse and noisy due to dropout and technical variability, which can obscure fine-scale expression dynamics.

To evaluate this, we focused on genes known to drive monocyte and macrophage dynamics during wound healing: *Il1b*, *Nfkbia*, *Arg1*, and *H2-Ab1*. In early wounds (0-3 days post-injury), monocytes and macrophages exhibit a pro-inflammatory phenotype characterized by high expression of *Il1b* and *Nfkbia* [11, 16, 17]. As wounds transition to the proliferative phase (3-7 days), cells shift toward a repair phenotype, upregulating genes such as *Arg1* [11, 16, 18, 19]. In parallel, expression of the MHC class II gene *H2-Ab1* increases and reflects antigen presentation functions [11].

Using the Lamian-inferred trajectory with nine clusters, we applied PGD to smooth the expression of these markers along pseudotime. We then visualized the raw (normalized), PGD-smoothed, and MAGIC-smoothed expression patterns on UMAP and as a function of pseudotime (Fig. 5).

Both PGD and MAGIC effectively reduced high-frequency noise from the raw data, yielding highly correlated results overall (Pearson correlation $\rho = 0.882$ across matched cell-gene pairs for the four genes). However, subtle spatial differences emerged on UMAP. For instance, *H2-Ab1* expression in PGD-smoothed data appeared more tightly confined to the lower-left trajectory branch, whereas MAGIC-smoothed expression diffused more broadly across adjacent branches (Fig. 5a). This localization is consistent with previous RNA velocity analyses identifying *H2-Ab1* as a driver for antigen-presenting clusters [11].

Differences became more pronounced when visualizing expression as a function of pseudotime (Fig. 5b). PGD-smoothed expression displayed sharper branch-specific separation and cleaner pseudotemporal gradients. In contrast, MAGIC-smoothed expression appeared noisier along the pseudotime axis and more blended between branches. Notably, for *Nfkbia*, PGD clearly resolved two distinct expression levels across separate branches in early pseudotime (0-1000), a separation that was less distinct in the MAGIC-smoothed result. This separation aligns with the identification of *Nfkbia* as a driver gene along early- to late-stage clusters [11].

Together, these results illustrate that PGD provides a trajectory-aware approach to gene expression smoothing. By enforcing continuity along the inferred pseudotime graph, PGD emphasizes branch-specific dynamics over global similarity, facilitating clearer interpretation of gene expression patterns along cellular trajectories.

4 Discussion

Here, we demonstrated that PGD provides a simple post hoc strategy for improving the alignment between inferred trajectories and low-dimensional visual representations. By diffusing cell-level features along a pseudotime graph, PGD enhances trajectory-consistent structure and continuity

to facilitate interpretation of cellular dynamics. Across increasing levels of trajectory complexity, PGD produced smoother pseudotime gradients and clearer branch separation, reflecting the explicit constraint of diffusion along inferred trajectory paths. We further demonstrated that PGD naturally extends to smoothing gene expression along trajectories, emphasizing pseudotime-dependent and branch-specific expression gradients.

PGD operates entirely downstream of trajectory inference and pseudotime estimation methods, refining cell-level features while leaving underlying trajectories and pseudotime unchanged. This modular design ensures PGD can be readily integrated with diverse trajectory inference frameworks without disrupting established upstream workflows.

Several limitations should be noted. First, the evaluations presented here are primarily qualitative. Second, PGD currently diffuses each feature independently, whereas many biological features exhibit structured dependencies that could be leveraged for improved smoothing. Third, we restricted our analysis to unweighted, undirected pseudotime graphs, although PGD can be generalized to weighted or directed graphs and to more general cell-cell transition matrices. Finally, PGD does not infer trajectories or generate embeddings de novo, and its effectiveness likely depends on the quality and consistency of the supplied trajectory and embeddings. More broadly, PGD offers a simple and extensible framework for enhancing the interpretability of single-cell trajectories.

Acknowledgment

The authors acknowledge the use of AI-assisted tools, including ChatGPT 5.2, Gemini 3 Pro, and GitHub Copilot, solely for language editing and grammar enhancement. The authors take full responsibility for the content of the manuscript.

References

[1] W. Saelens, R. Cannoodt, H. Todorov, and Y. Saeys, “A comparison of single-cell trajectory inference methods,” *Nature biotechnology*, vol. 37, pp. 547–554, 2019.

[2] M. Setty, V. Kiseliovas, J. Levine, A. Gayoso, L. Mazutis, and D. Pe’Er, “Characterization of cell fate probabilities in single-cell data with palantir,” *Nature biotechnology*, vol. 37, pp. 451–460, 2019.

[3] S. V. Stassen, G. G. K. Yip, K. K. Y. Wong, J. W. K. Ho, and K. K. Tsia, “Generalized and scalable trajectory inference in single-cell omics data with via,” *Nature communications*, vol. 12, p. 5528, 2021.

[4] P. Weiler, M. Lange, M. Klein, D. Pe’er, and F. Theis, “Cellrank 2: unified fate mapping in multiview single-cell data,” *Nature Methods*, vol. 21, pp. 1196–1205, 2024.

[5] K. Kamimoto, B. Stringa, C. M. Hoffmann, K. Jindal, L. Solnica-Krezel, and S. A. Morris, "Dissecting cell identity via network inference and in silico gene perturbation," *Nature*, vol. 614, pp. 742–751, 2023.

[6] J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, and F. J. Steemers, "The single-cell transcriptional landscape of mammalian organogenesis," *Nature*, vol. 566, pp. 496–502, 2019.

[7] M. Lange, A. Granados, S. VijayKumar, J. Bragantini, S. Ancheta, S. Santhosh, M. Borja, H. Kobayashi, E. McGeever, and A. C. Solak, "Zebrahub–multimodal zebrafish developmental atlas reveals the state-transition dynamics of late-vertebrate pluripotent axial progenitors," *BioRxiv*, p. 2023, 2023.

[8] A. Bastidas-Ponce, S. Tritschler, L. Dony, K. Scheibner, M. Tarquis-Medina, C. Salinno, S. Schirge, I. Burtscher, A. Böttcher, and F. J. Theis, "Comprehensive single cell mrna profiling reveals a detailed roadmap for pancreatic endocrinogenesis," *Development*, vol. 146, p. dev173849, 2019.

[9] F. A. Wolf, F. K. Hamey, M. Plass, J. Solana, J. S. Dahlin, B. Göttgens, N. Rajewsky, L. Simon, and F. J. Theis, "Paga: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells," *Genome biology*, vol. 20, p. 59, 2019.

[10] K. R. Moon, D. V. Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. van den Elzen, M. J. Hirn, and R. R. Coifman, "Visualizing structure and transitions in high-dimensional biological data," *Nature biotechnology*, vol. 37, pp. 1482–1492, 2019.

[11] J. Pang, M. Maienschein-Cline, and T. J. Koh, "Monocyte/macrophage heterogeneity during skin wound healing in mice," *The Journal of Immunology*, vol. 209, pp. 1999–2011, 2022.

[12] F. A. Wolf, P. Angerer, and F. J. Theis, "Scanpy: large-scale single-cell gene expression data analysis," *Genome biology*, vol. 19, p. 15, 2018.

[13] I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P. ru Loh, and S. Raychaudhuri, "Fast, sensitive and accurate integration of single-cell data with harmony," *Nature methods*, vol. 16, pp. 1289–1296, 2019.

[14] W. Hou, Z. Ji, Z. Chen, E. J. Wherry, S. C. Hicks, and H. Ji, "A statistical framework for differential pseudotime analysis with multiple single-cell rna-seq samples," *Nature communications*, vol. 14, p. 7286, 2023.

[15] D. V. Dijk, R. Sharma, J. Nainys, K. Yim, P. Kathail, A. J. Carr, C. Burdziak, K. R. Moon, C. L. Chaffer, and D. Pattabiraman, "Recovering gene interactions from single-cell data using data diffusion," *Cell*, vol. 174, pp. 716–729, 2018.

[16] Y. Kuninaka, Y. Ishida, A. Ishigami, M. Nosaka, J. Matsuki, H. Yasuda, A. Kofuna, A. Kimura, F. Furukawa, and T. Kondo, "Macrophage polarity and wound age determination," *Scientific Reports*, vol. 12, p. 20327, 2022.

[17] S. Willenborg, L. Injarabian, and S. A. Eming, "Role of macrophages in wound healing," *Cold Spring Harbor perspectives in biology*, vol. 14, p. a041216, 2022.

[18] L. Campbell, C. R. Saville, P. J. Murray, S. M. Cruickshank, and M. J. Hardman, "Local arginase 1 activity is required for cutaneous wound healing," *Journal of Investigative Dermatology*, vol. 133, pp. 2461–2470, 2013.

[19] H. S. Kim, S. A. Jee, A. Einisadr, Y. Seo, H. G. Seo, B. S. Jang, H. H. Park, W.-S. Chung, and B. G. Kim, "Detrimental influence of arginase-1 in infiltrating macrophages on poststroke functional recovery and inflammatory milieu," *Proceedings of the National Academy of Sciences*, vol. 122, p. e2413484122, 2025.

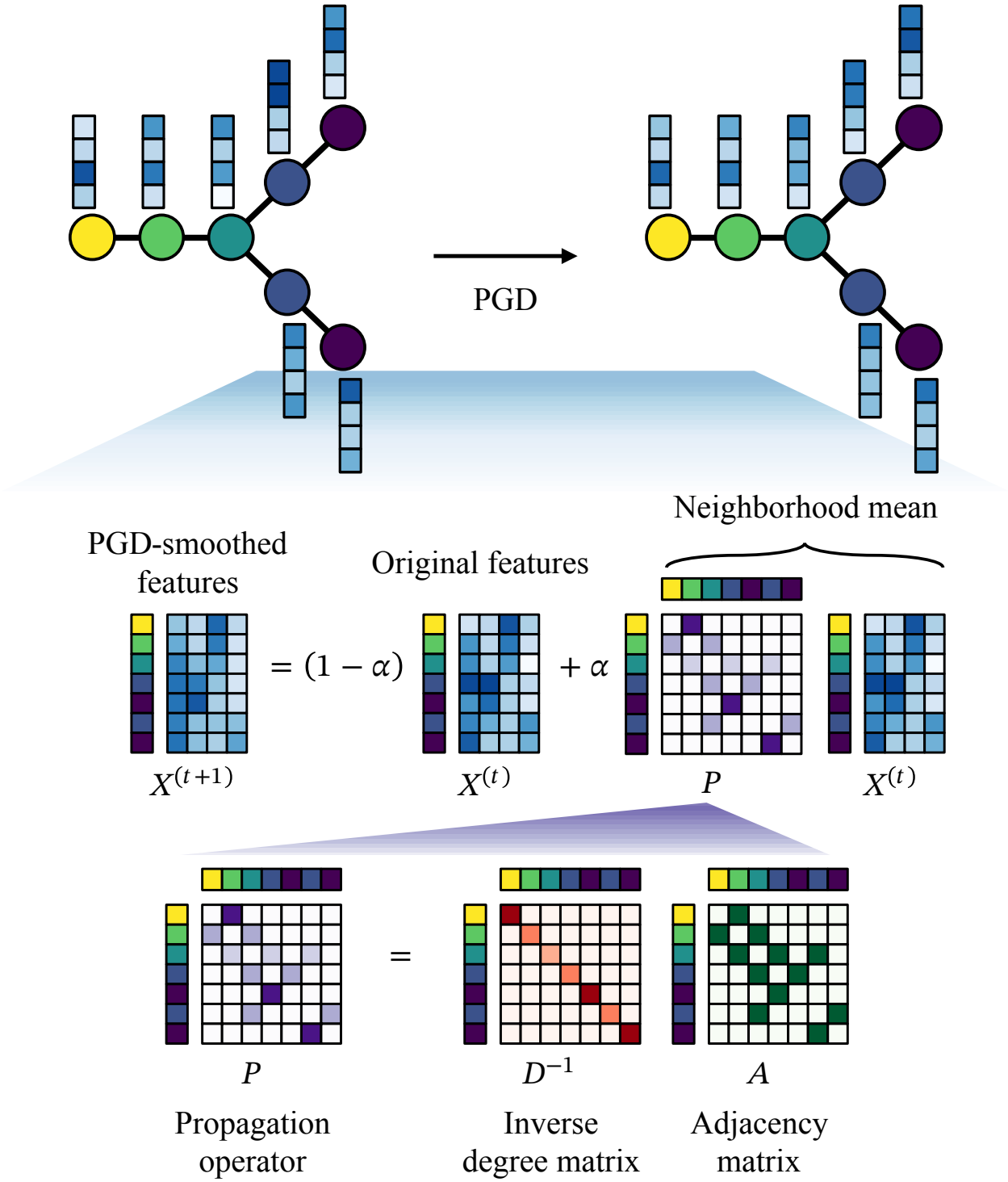

Figure 1: Schematic of PGD.

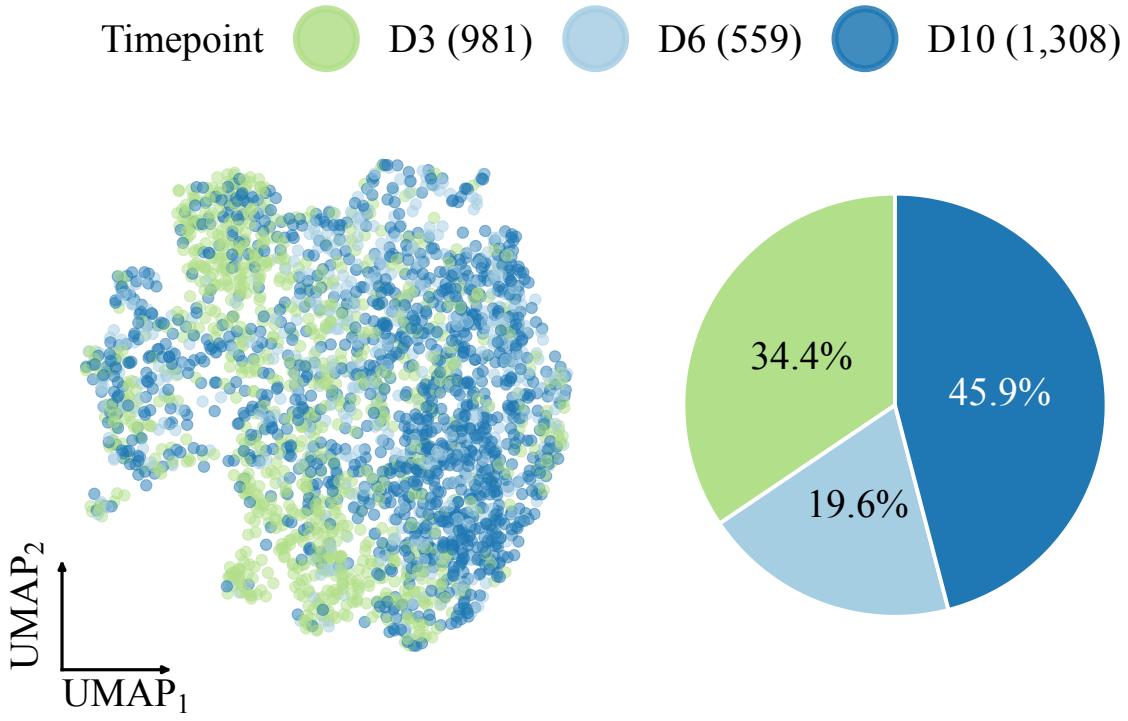

Figure 2: UMAP of the monocyte/macrophage wound-healing dataset colored by timepoint (left) and timepoint composition shown as a pie chart (right).

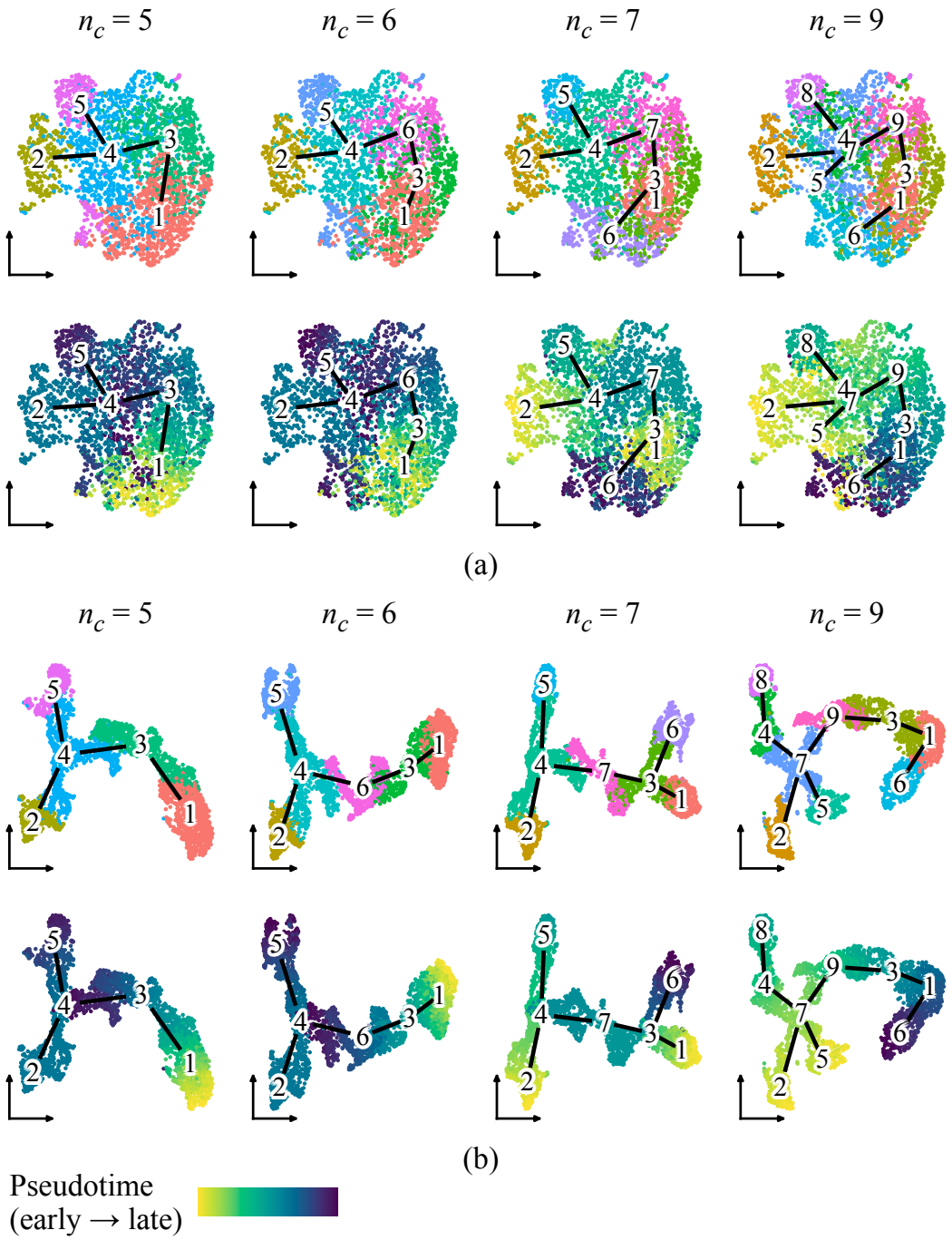

Figure 3: UMAP layouts generated from (a) original embeddings and (b) PGD-smoothed embeddings. Cells are colored by Lamian cluster or pseudotime. Edges indicate the Lamian-inferred trajectory graph.

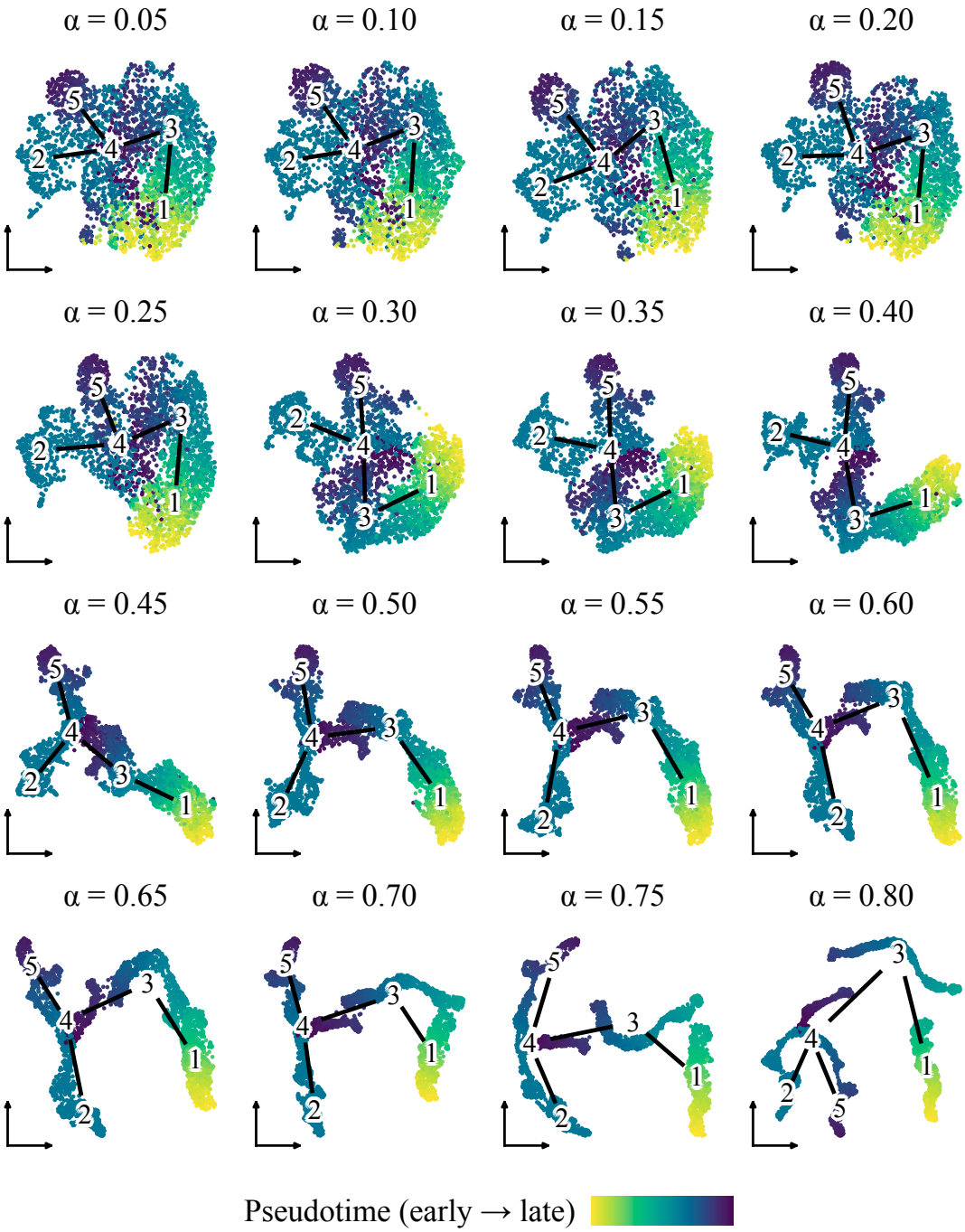

Figure 4: UMAP layouts generated from PGD-smoothed embeddings for incrementally increasing α. Cells are colored by pseudotime. Edges indicate the Lamian-inferred trajectory graph.

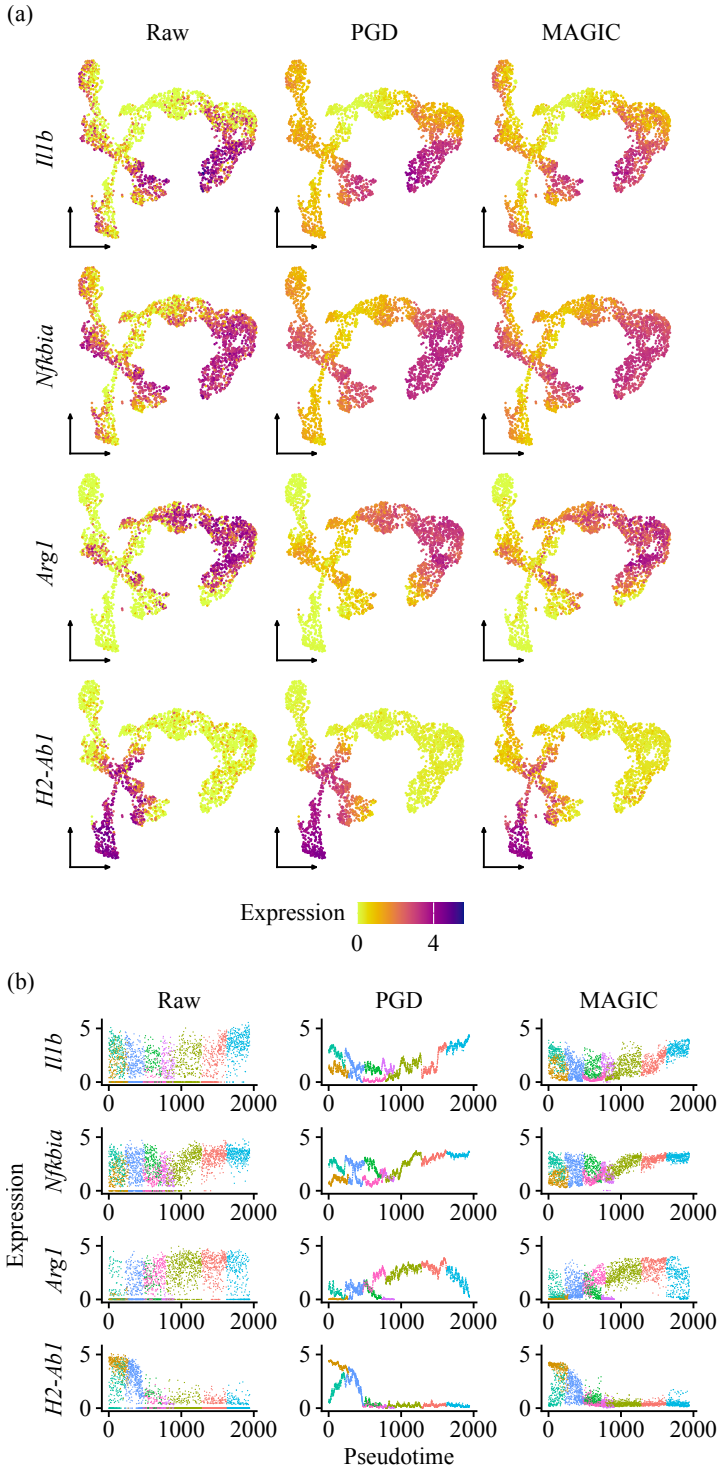

Figure 5: Expression of representative genes (a) on UMAP and (b) as a function of pseudotime. In (b), cells are colored by Lamian cluster.