Unsupervised learning of spatiotemporal motifs in dynamically self-organized cells

Lim Ying Chen [1] **Rakesh Das** [2] **Tetsuya Hiraiwa** [3] **N. Duane Loh** [1 4]

[1]*Department of Physics, National University of Singapore, 117551 Singapore, Singapore* [2]*Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany* [3]*Institute of Physics, Academia Sinica, Taiwan 115201, Taiwan* [4]*Department of Biology, National University of Singapore, 117557 Singapore, Singapore*. Correspondence to: LYC e0014951@u.nus.edu, NDL duaneloh@nus.edu.sg.

1. Introduction

Understanding cell motility is key to predicting and manipulating such systems. Quantitatively studying such motility requires good featurization of its structural dynamics.

Global properties of large complex cell systems are often described using features like nematic and hexatic order parameters [1, 2, 3], angular regularity between neighbouring cells [4], and dynamic order parameters for groups of cells [4]. However, these features may be insufficient for covering the wide range of cellular behaviour that can emerge from such complex systems.

We surmise that persistent and prevalent patterns in cell groups, which we call spatiotemporal motifs, make good machine learning features for detailed studies on cell structures and dynamics. Such motif patterns can efficiently quantify coarse-grained cell group properties by omitting each group's internal degrees of freedom. This simplification of features makes correlations between the motifs and system more readily apparent.

Here, we find motifs in a system of simulated cells [5] that can dynamically self-organise (DSO) into various large multicellular patterns depending on its intercellular interaction strength parameters α_{CIL} and α_{CF}. Contact inhibition of locomotion (CIL), which describes how cells tend to separate after contact, was experimentally observed in neural crest cells [6] and cancer cells [7]. Relatedly, contact following (CF), which describes how contacting cells follow each other, was proposed to explain the behaviour of the slime mold Dictyostelium Discoideum [8]. Each cell's neighbourhood is quantified via six structural and dynamic motif features (see table in figure 3) relating the cell to its Voronoi neighbours.

2. Identifying prevalent and persistent motifs

In this model [5], one prevalent and persistent motif is cells that neighbour six other cells (6NNs). Under most α_{CIL} and α_{CF} conditions, 6NNs are prevalent since they often make up between 30% to 80% of cell groups present in a simulated system when averaged over time (see A1). Similarly, 6NNs are persistent when $\alpha_{\mathrm{CIL}} < 0$ as they tend to stay with their neighbours longer than many other cell groups.

While 5NNs and 7NNs do not appear prevalent and persistent, we later see them form meaningful defect patterns in large cell collectives. Hence, they can also be seen to make up motifs under certain conditions.

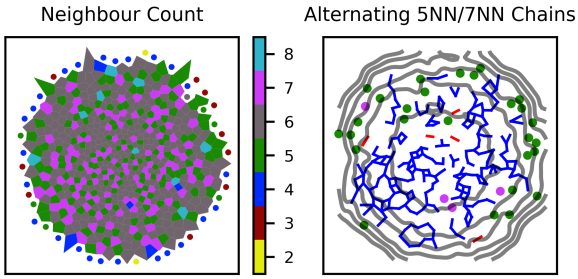

Fig. 1: (LEFT) Cells coloured by their number of neighbours, showing the prevalence of 6 nearest neighbours (6NN) groups in static aggregates. (RIGHT) Packing defects are evident from the isolated, 5NN and 7NN cell groups next to one another (red segments) and the extended chains of alternating adjacent 5NN and 7NN groups (blue lines).

3. Motifs sense strain and defects in packing

When $\alpha_{\mathrm{CIL}} < 0$, cells tend to attract one another to form large quasi-stable and compact cell collections (e.g., static, motile and rotating aggregates). There, the prevalent 6NN motifs are commonly interspersed with 5NN and 7NN motif defects (see figure 1) and resemble those found in hexagonal nets in materials, albeit with some structural differences.

Unlike hexagonal crystals, the motifs in this system are highly compressible. This compressibility appears as strain patterns (e.g., increasingly dense 6NN packing deep inside aggregates) which are unusual for solid-state materials.

The 5NN and 7NN motif defects in aggregates and rings can move and fluctuate dynamically. This is unlike many solid-state systems, where defects are generally stationary and prevent the underlying objects from moving. These defects also tend to form extended and alternating chains of 5NNs and 7NNs.

4. Detectable quasi-periodic motif ensembles

When cells organise into different DSO patterns, this is often reflected through changes in their local structural properties. Hence, we expect that the structural features of nearest neighbour features can change while a system evolves.

By tracking changes to the Voronoi neighbour distribution of cell groups, we found that spiral DSO cells show a recurrent and persistent motif of periodically cycling between two modes, as seen in figure 2. Cells showing this pattern weakly attract and have a

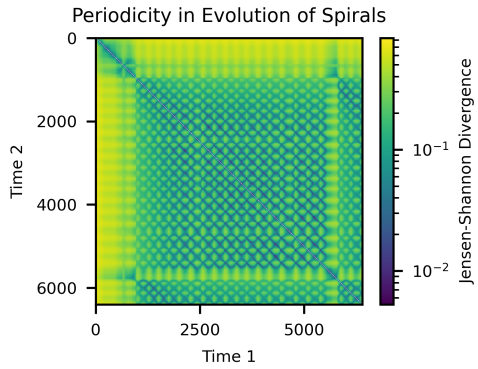

Fig. 2: Periodicity in spiral DSO pattern cell evolution (see parameter space in figure 3) revealed by the periodic Jensen-Shannon divergence in mean neighbour distance of cell groups between different time points.

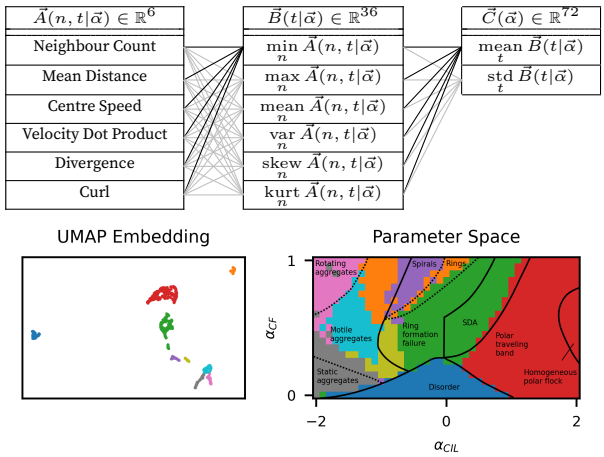

Fig. 3: 'Phase' spaces uncovered from coarse-grained motif features embedded as two-dimensional points using UMAP. These points, each representing a unique parameter pair, were clustered using HDBSCAN and colourized by its cluster index alongside its associated parameter space grid. The DSO phases identified in [5] are drawn in the parameter space to highlight similarities between the phases and clustering.

strong tendency to follow one another ($\alpha_{\mathrm{CIL}} \approx -0.4$, $\alpha_{\mathrm{CF}} \approx 0.7$).

While a system evolves, we characterise the local dynamics of cells at several time points using the mean distance of each cell from its Voronoi neighbours. This quantity encodes both the local cell packing density and the distance-dependent interaction forces acting on each cell. The distance distribution at each time point is then quantised into a histogram. The difference in features between pairs of time points is calculated from the Jensen-Shannon divergence between their corresponding histograms, and a clear periodic pattern is visible from this divergence in figure 2.

5. Unsupervised classification of collective spatiotemporal behaviours using coarse-grained descriptors

In [5], the author determined a 'phase diagram' of the DSO cell dynamics simulated by a dynamics model primarily using visual inspection guided by particle densities. Here, we explore if such phases can also be obtained directly from the motif features using unsupervised learning.

In simulated systems where cells evolve under fixed interaction strength parameters α_{CIL} and α_{CF}, we extracted six structural and dynamic cell features from each cell across multiple time points. These six features of the nth motif at time t, $\vec{A}(n,t|\alpha_{\mathrm{CIL}},\alpha_{\mathrm{CF}})$, were coarse-grained over the complex dynamics of each system into a 72-dimensional feature vector $\vec{C}(\alpha_{\mathrm{CIL}},\alpha_{\mathrm{CF}})$: first by summarising $\vec{A}(n,t|\alpha_{\mathrm{CIL}},\alpha_{\mathrm{CF}})$ over all n cells at each time point t to $\vec{B}(t|\alpha_{\mathrm{CIL}},\alpha_{\mathrm{CF}})$, then again over all t to $\vec{C}(\alpha_{\mathrm{CIL}},\alpha_{\mathrm{CF}})$ (see figure 3).

The 72-dimensional feature vectors $\vec{C}(\alpha_{\mathrm{CIL}},\alpha_{\mathrm{CF}})$ were then embedded as points in a two-dimensional space using Uniform Manifold Approximation and Projection (UMAP), which attempts to place vectors with similar features close together (see figure 3). Points in the UMAP embedding were clustered through Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) and colourized by their cluster indices along with their corresponding grid position in the parameter space.

The HDBSCAN clustering in figure 3 shows several similarities to the phase boundaries drawn via manual [5]. This similarity is clearest with the disordered phase, where the visually distinctive cell scattering resulted in clear differences in the motif features compared to other phases. However, our clustering differed in several other phases, and was unable to distinguish quantifiable and persistent differences between others (i.e., polar travelling bands vs. homogeneous polar flocks).

6. Conclusion

By extracting structural and dynamic features from cells grouped with their Voronoi neighbours, we were able to identify motifs present in a system of simulated cells [5]. Using these features, we studied several phenomena in the system (i.e., strain and defects in cell packings, quasi-periodic evolution of spiral DSO patterns, and unsupervised learning of interaction parameter phase spaces). These results showcase the usefulness of this motif for quantifying the structural dynamics of such complex systems. Furthermore, this work also suggests the possibility of using machine learning to induce non-linear order parameters for such complex systems.

Acknowledgments

We thank the support of the Nation University of Singapore Physics department for the PhD scholar-

ship.

References

[1] Robert Großmann, Igor S. Aranson, and Fernando Peruani. A particle-field approach bridges phase separation and collective motion in active matter. *Nature Communications*, 11(1):5365, October 2020.

[2] Xia-qing Shi and Hugues Chaté. Self-Propelled Rods: Linking Alignment-Dominated and Repulsion-Dominated Active Matter, July 2018.

[3] Sebastian Weitz, Andreas Deutsch, and Fernando Peruani. Self-propelled rods exhibit a phase-separated state characterized by the presence of active stresses and the ejection of polar clusters. *Physical Review E*, 92(1):012322, July 2015.

[4] Luciano Da Fontoura Costa, Fernando Rocha, and Silene Maria Araújo De Lima. Characterizing polygonality in biological structures. *Physical Review E*, 73(1):011913, January 2006.

[5] Tetsuya Hiraiwa. Dynamic Self-Organization of Idealized Migrating Cells by Contact Communication. *Physical Review Letters*, 125(26):268104, December 2020.

[6] Elena Scarpa, András Szabó, Anne Bibonne, Eric Theveneau, Maddy Parsons, and Roberto Mayor. Cadherin Switch during EMT in Neural Crest Cells Leads to Contact Inhibition of Locomotion via Repolarization of Forces. *Developmental Cell*, 34(4):421–434, August 2015.

[7] Benjamin Lin, Taofei Yin, Yi I. Wu, Takanari Inoue, and Andre Levchenko. Interplay between chemotaxis and contact inhibition of locomotion determines exploratory cell migration. *Nature Communications*, 6(1):6619, April 2015.

[8] Tamiki Umeda and Kei Inouye. Possible Role of Contact Following in the Generation of Coherent Motion of Dictyostelium Cells. *Journal of Theoretical Biology*, 219(3):301–308, December 2002.

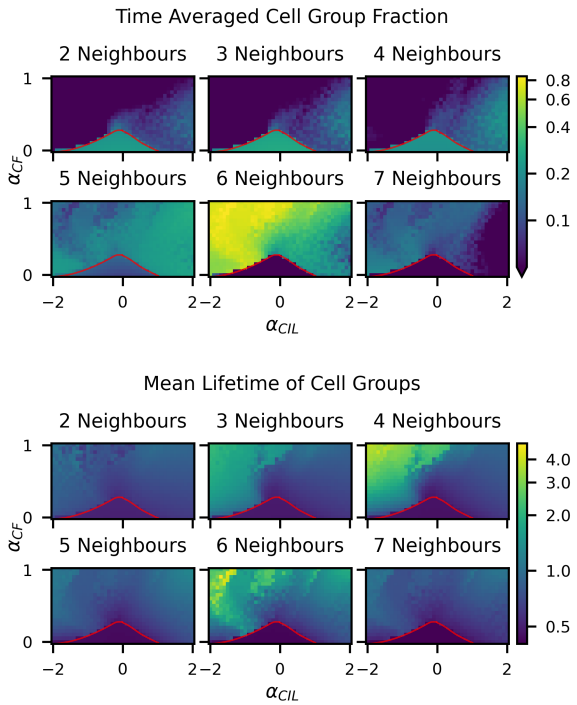

Fig. A1: Prevalence (top) and persistence (bottom) of different nearest neighbour (NN) cell groups at various contact following (α_{CF}) and contact inhibition of locomotion (α_{CIL}) strengths. 6NN groups are prevalent under most condition, accounting for more than 30% of the type of cell groups present. 4NNs and 6NNs are persistent when $\alpha_{\mathrm{CIL}} < 0$, staying with their neighbours longer than the other type of cell groups. A red line is drawn to highlight the region with disordered dynamic self-organising cell pattern.