Frederic.Cazals@inria.fr
Web: `http://team.inria.fr/abs`
Centre Inria at Université Côte d'Azur, France
3IA Côte d'Azur

Master internship proposal

Mining protein dynamics with transformers

Keywords: deep learning, attention mechanisms, transformers, dynamical systems, clustering, (molecular) dynamics, protein structure analysis.

Context. Emerging from the field of natural language processing, (self-)attention mechanisms have proven essential to understand the coupling between tokens in a sentence [1, 2]. Recently, attention mechanisms have also proven key to encode the coupling between spatial patterns observed between amino acids in a protein structure [3]. The corresponding tool, `AlphaFold2` by Deepmind, is considered a major achievement to predict a plausible structure of a protein from its amino-acid sequence. `AlphaFold2` uses internally two steps of attention mechanisms, respectively geared towards multiple sequence alignments and protein structure.

Goals. `AlphaFold2` is a key achievement but outputs a single structure *i.e.* geometry. On the other hand, proteins are highly dynamic molecules, and statistical physics teaches us that observable properties of molecules depend on ensemble of conformations (weighted by Boltzmann's factor). (See also AI, molecular design and the Covid19.) The goal of this internship will be to expand `AlphaFold2` towards protein dynamics. The developments will exploit recent insights related to clusters in self-attention dynamics [4], model based clustering [5], as well as protein structure kinematics [6].

The work envisioned encompasses the design and analysis of algorithms, their coding (C++ and python), as well their experimental evaluation.

Conditions. Internship with *gratification*. Possibility to follow-up with a PhD thesis.

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[2] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. Ccnet: Criss-cross attention for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 603–612, 2019.

[3] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[4] Borjan Geshkovski, Cyril Letrouit, Yury Polyanskiy, and Philippe Rigollet. The emergence of clusters in self-attention dynamics. In *NeurIPS*, 2023.

[5] F. Cazals, L. Goldenberg, and S. Suren. Subspace-embedded spherical clusters: a novel cluster model for compact clusters of arbitrary dimension. *Submitted*, 2024.

[6] T. O'Donnell and F. Cazals. Enhanced conformational exploration of protein loops using a global parameterization of the backbone geometry. *J. Comp. Chem.*, 44(11):1094–1104, 2023.