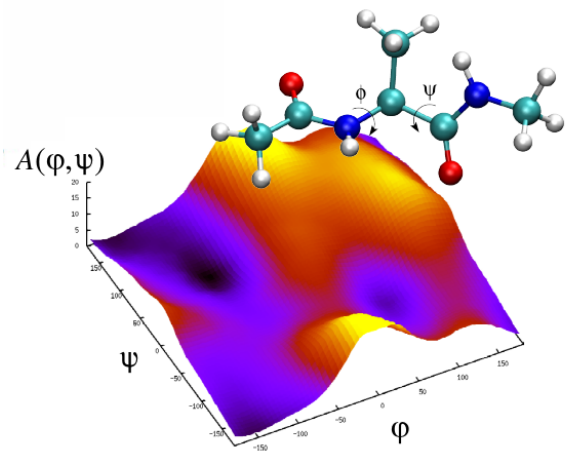
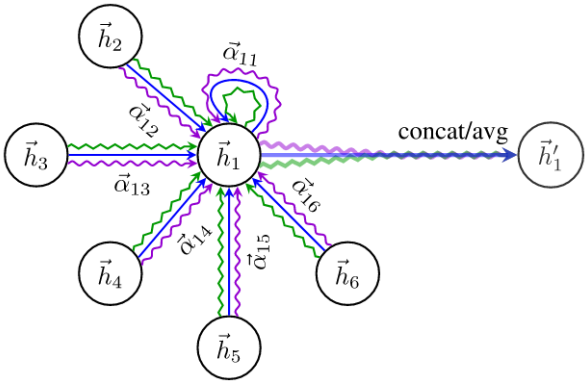
Capturing Molecular Conformations with Graph Neural Networks

Keywords: Deep Learning, Molecular Dynamics, Interpretability
Research team: TAU, Inria Saclay
(In collaboration with Laboratoire de Biochimie Théorique, CNRS, IBPC, Paris)
Location:
Laboratoire de Recherche en Informatique, Université Paris Sud
Bât 660 Claude Shannon, Rue Noetzlin, 91190 Gif-sur-Yvette, France
Supervisors:
Guillaume Charpiat (*guillaume.charpiat@inria.fr*)
Jérôme Hénin (*jerome.henin@ibpc.fr*)
Loris Felardos (*loris.felardos@inria.fr*)



(a) Partial configuration space of Alanine dipeptide



(b) Diagram of a Graph Attention Network

Scientific Context: Numerical simulations are used routinely to study the dynamics of biomolecules at the atomic scale. They produce large amounts of high-dimensional data representing the time trajectories of molecular configurations, but extracting knowledge from such massive data is still a challenging task today. Deep Learning models are already being transformative in the field of Molecular Dynamics by partially addressing some critical issues, as they can identify some inherent structure that is not captured by other methods.

Research Context: The purpose of the underlying research is to capture the high-dimensional configuration space of molecules (usually sampled by means of molecular simulations) by mapping its intrinsic manifold to a more interpretable space of lower dimension. The mapping itself is a learned parametric function (i.e. a neural network) and allows the generation of useful geometric extrapolations [1]. The recently developed Graph Neural Networks [2] already showed impressive results when applied to a variety of molecular studies, where the graphs considered are the ones of molecules [3, 4] and are already used in the context of this research.

Objectives: The goal of this internship is to develop visualization tools, to better analyze such neural networks, during or after training, in order to improve their architecture or training procedures.

- *Analyze the behavior of the network both during and after training.* To better understand the distribution of molecular conformations that the neural network outputs, statistical tools (such

as clustering, or histograms of relevant degrees of freedom) are needed, as well as associated visualizations. Concerning the inner layers of the network, we have simple, new theoretical concepts to detect the parts of the network that lack expressive power; they need to be tested.

- *Participate in the development of the core repository* by (1) extending all the code related to the visualization, interpretation and debug of the network, (2) incorporating new data sources in the training loop and (3) improving the network itself and/or training procedure.

- Propose ideas to refine (or even challenge) the current approach in order to better capture regions of the configuration space that are difficult to naturally sample.

Requirements:
- Mathematics (functional analysis, linear algebra, calculus, probabilities).
- Good Python coding skills (and knowledge of its ecosystem of scientific libraries).
- Basic experience with Pytorch and/or Tensorflow and/or Chainer.
- Interest in or knowledge of Deep Learning.
- Fluency in english.
- Basic knowledge of Organic Chemistry and/or Molecular Biology would also be helpful.

References

[1] Noé, et al. "Boltzmann Generators – Sampling Equilibrium States of Many-Body Systems with Deep Learning." `http://arxiv.org/abs/1812.01729`

[2] Veličković, et al. "Graph Attention Networks." `http://arxiv.org/abs/1710.10903`

[3] Chen, et al. "Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals." `http://arxiv.org/abs/1812.05055`

[4] Schütt, et al. "SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions." `http://arxiv.org/abs/1706.08566`