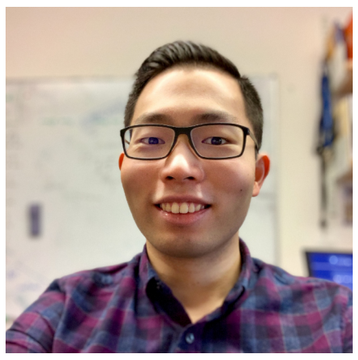


Function Follows Form: Decoding the Secrets of Shapes

Ka Man (Ambrose) Yim and Jacob Leygonie

Mathematical Institute, University of Oxford yim@maths.ox.ac.uk

The Problem

The geometric structure of molecules, brains, and social networks determine the physical properties of these systems. We can model these structures as networks: for example, a molecule is a network of atoms linked by chemical bonds.

We want to develop machine learning algorithms that predict real life attributes of objects based on their network structure.

New algorithms are necessary: traditional algorithms cannot analyse networks due to fundamental mathematical limitations.

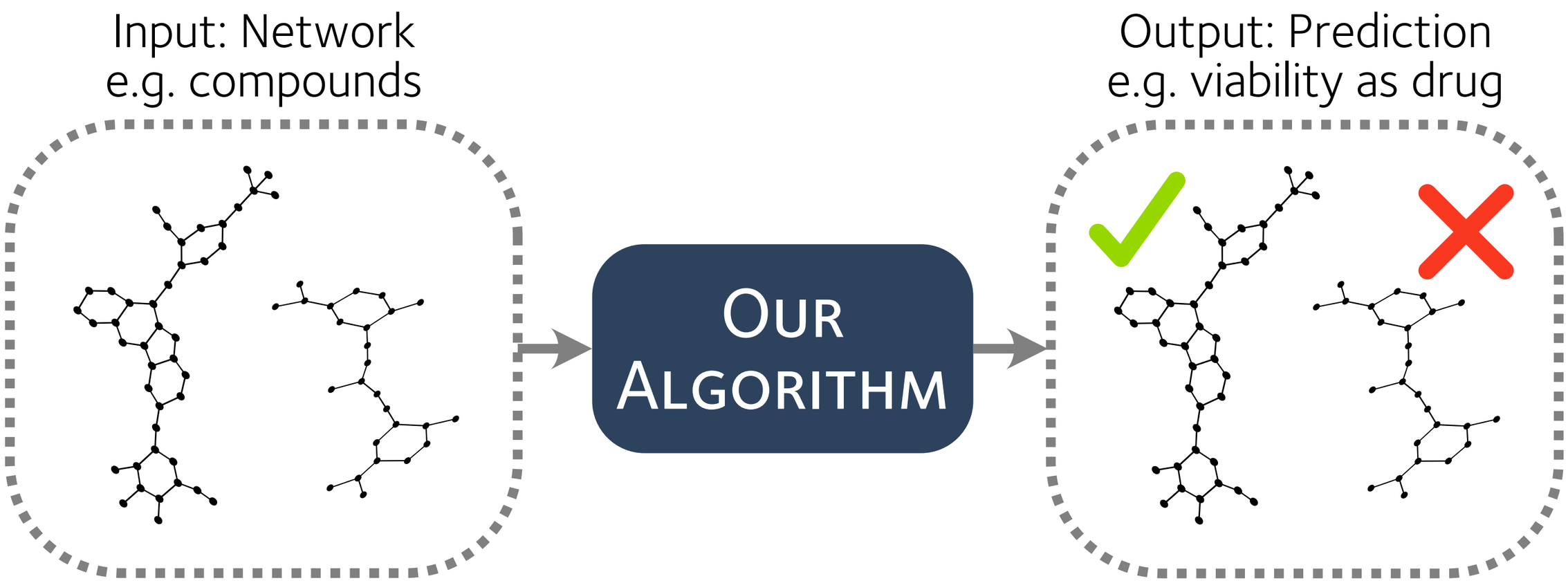
Why should you care?

Machine learning algorithms for networks can help industry save money, and generate industrial innovations and scientific insight.

Accelerate drug discovery. Algorithms can recommend compounds that are likely drug candidates. This can save costly lab time, potentially saving millions of pounds per drug.

Improve medical diagnosis. Algorithms can screen for ailments by analysing networks in our circulatory and respiratory systems.

Better recommendations to consumers. Companies such as Netflix can use algorithms to analyse their network of consumers and products to inform product and content recommendations.



We develop a new machine learning algorithm that analyses the structure of networks to predict their physical properties.

Our Solution

Our algorithm uses Persistent Homology (PH) to capture structural features of networks and make predictions.

PH is derived from algebraic topology, the mathematical study of shapes. PH tells us how structures such as loops are created and destroyed when we assemble a network from constituent parts.

By training on large amounts of data, our algorithm learns an optimal way of assembling a network, so that the structural information PH reveals lead to more accuracte predictions.

Research Outcomes

We tested the performance of our algorithm in predicting the biochemical properties of molecules, based only on their network structure.

We are able to achieve state-of-the-art prediction accuracies on many benchmark tests. In one benchmark test, we are able to predict whether a class of molecules causes mutations in a type of bacteria with over 90% accuracy. In another, we were able to predict whether another class of molecules is a viable cancer treatment drug with almost 75% accuracy.

EPSRC Engineering and Physical Sciences Research Council

InFoMM Industrially Focused Mathematical Modelling

The pre-print of our paper is now on arxiv!