Computer science is to biology what calculus is to physics. If Elon Musk wants to make sure that a billionaire can go to space and make sure he gets home safely, he can brush up on his integrals and derivatives and then he can calculate how much force to give his rocket.

But in a field like molecular biology, the questions we ask have a different flavor. Many are about making choices between distinct but extremely numerous options. Given this bacterial DNA I found in the wild and that huge database of known genes, which one does it best match? No matter how much I like doing calculations with a pencil and paper --- which, to be clear, I do love --- I am never going to come up with the answer myself, because it just involves too many comparisons of one piece of DNA to another. But computers are great at straightforward tasks that need to be done over and over again.

However, even computers can be stumped if the problem is hard enough or the input size is big enough. The human genome is 3 billion characters long. If we use a naive algorithm, it would take about an hour for a computer to check whether a small bit of DNA is contained in that genome. But today we can do this basically instantaneously. And it's not because we have faster computers --- it's because we have better algorithms.

In my dissertation, I am also designing better algorithms for problems like the ones I just mentioned. RNA transcripts are molecules similar to DNA that are a product of gene expression. In order to get an accurate picture of what is going on inside a cell, we want to read the code of the RNA molecules inside it. But unfortunately, we don't have the technology to get that code directly. Instead, we get chopped up bits of RNA, and we'd like to fit them back together like a jigsaw puzzle. Except this puzzle has millions of pieces, and instead of four sides to fit together, there might be hundreds of ways to align each piece.

In computer science, big advances are often made when we take a tricky problem and represent it using something more abstract. In my work, that abstraction is a graph, where nodes and edges encode the pieces of RNA from the sample, and by looking for the right set of paths in the graph, we can recover the code of the transcripts. But even finding those paths is one of those problems that is too hard for computers if we just use simple approaches. My work has given new ways to find the paths more accurately and quickly. Thank you.