

[http://www.bostonglobe.com/ideas/2015/08/10/computer-scientists-have-looked-for-solution-that-doesn-exist/tXO0qNRnbKrClfUPmavifK/story.html?p1=Article\\_Trending\\_Most\\_Viewed](http://www.bostonglobe.com/ideas/2015/08/10/computer-scientists-have-looked-for-solution-that-doesn-exist/tXO0qNRnbKrClfUPmavifK/story.html?p1=Article_Trending_Most_Viewed)

# The Boston Globe

For 40 years, computer scientists looked for a solution that doesn't exist

By Kevin Hartnett GLOBE CORRESPONDENT AUGUST 10, 2015



For 40 years, computer scientists have tried in vain to find a faster way to do an important calculation known as “edit distance.” Thanks to groundbreaking work from two researchers at MIT, they now know the reason they’ve continually failed is because a faster method is actually impossible to create. “Edit distance” is a straightforward and satisfyingly tangible idea. Imagine you have two sequences of numbers and want to know how many steps it takes to transform one into the other. In this transformation you can add a digit, remove a digit, or substitute one digit for another. For example, you have the data strings “1234” and “2345.” To turn the first into the second takes two steps — remove the 1, add a 5.

The number of steps required to make this transformation is the edit distance.

It's a cool concept and also a useful one. Biologists use edit distance all the time when comparing genomes of different organisms. A genome is essentially a long string of data — the sequence of A, G, T, and C that make up our DNA. By calculating the edit distance between the genomes of two organisms, biologists can estimate how far back in evolutionary time the organisms diverged from each other.

Edit distance is useful, but also very time-consuming to calculate. The current method for measuring it is known as the Wagner-Fischer algorithm. It was developed 40 years ago and works essentially by placing data on a grid. One string of data goes along the top row, the other goes down the left-hand column, and the algorithm fills in the grid diagonally, counting changes as it goes.

The Wagner-Fischer algorithm is a computationally-intensive method that operates in what computer scientists call “quadratic time.” In plain terms, this means that when the length of the data strings goes up a little, the number of steps needed to compare them goes up a lot. For example, if you have two sequences, each with 10 digits, it takes 100 operations (10-squared) to compute the edit distance. But if you have two sequences, each with 100 digits, it takes you 10,000 operations to compare them (100-squared). The number of digits went up a little (just 90). The number of operations went up a lot (9,900).

- **Like Airbnb, but for algorithms?**

The fact that edit distance is only computable in quadratic time has big implications for genomics. The human genome contains about 3 billion base pairs. To compute the edit distance between two human genomes takes 3 billion-squared operations (to appreciate how big that is, write a 9 followed by 18 zeroes). That's a number of operations, explains Piotr Indyk of MIT, that is “computationally infeasible.” Which is to say that our best computers chugging away for a really long time still won't generate an answer.

Because it's computationally infeasible to compute the edit distance between two human genomes, biologists have to rely on approximations. They'd love a faster method, but Indyk and his coauthor, MIT graduate student Arturs Backurs, recently demonstrated a faster method is impossible to create. And

by impossible they don't mean "very hard" or "our technology has to improve first." They mean something like, "by the laws of the mathematics, it can't be done."

Indyk and Backurs presented their work at the annual Symposium on the Theory of Computing conference in Portland, Ore., in June. The details of how they demonstrated this impossibility are hard to describe, but their approach is easy enough to apprehend. In computer science, the most famous open question is the P equal to NP problem. It's a mega-sized issue. Whoever proves it first would receive millions of dollars in prize money and be all over the international news. Most if not nearly all computer scientists believe the P equal to NP problem is false. Indyk and Backurs developed a clever strategy by which they showed that in order for there to be a faster way of calculating edit distance, a stronger variant of the P equal to NP problem must be true. And since most everyone is convinced that this variant of the P equal to NP problem is false, it follows there's almost certainly no way to really improve on the Wagner-Fischer algorithm.

Indyk and Backurs' result has been greeted with something like relief among computer scientists, who can now stop beating their heads against a wall in search of a faster method that doesn't exist.

"I remember wondering as a student, 15 years ago, whether you could substantially beat the quadratic-time algorithm for [edit distance]. Thanks to Backurs and Indyk, we now know for the first time that you can't," says Scott Aaronson, a computer scientist at MIT.

For Indyk, too, his recent work acts as a permission slip to move onto other questions. Like hundreds of other computer scientists, he has spent years searching fruitlessly for a faster way to compute edit distance. Now, he says, "I will stop trying to solve the problem."