Transcript of Y.C. Tay's Q&A session at International Youth STEM Conference (https://www.iysc2022.org/)

Question:
I saw on your bio that you are both a computer scientist and mathematician, in what specific areas do those two fields intersect?

Answer:

Mathematics and Computer Science interact in many ways. Let me give some examples.

Machine Learning: http://www.d2l.ai/chapter_appendix-mathematics-for-deep-learning/index.html

Hardware Design: https://www.cs.ox.ac.uk/teaching/courses/2009-2010/computeraidedverification/

Animation: https://www.maa.org/meetings/calendar-events/math-in-the-movies

Cryptography: https://nrich.maths.org/2200

Computational Complexity: https://www.claymath.org/millennium-problems/p-vs-np-problem

Question:
What are you currently investigating?

Answer:

3 topics:

(1) *Synthetic generation of student data*:
Universities collect a lot of data on their students (their majors, courses enrolled, grades, WiFi connection, etc.), and some researchers do research with this data (e.g. using WiFi connection times to deduce sleep behavior for different majors). However, there are privacy issues with giving researchers access to real data. My students and I are using Generative Adversarial Networks to generate synthetic data that is similar to real data, so the researchers can use the synthetic data for analysis. See https://nus.edu.sg/alset/2020/06/08/alset-seminar-w-feng-tay-student-data/)

(2) *Eigenvalue computation for large graphs*:
Graphs are a common structure in computer science (e.g. as models for social networks). These graphs can be represented as matrices. Matrices have something called eigenvalues and eigenvectors that capture important properties (e.g. https://www.math.arizona.edu/~glickenstein/math443f08/bryanleise.pdf). The state-of-the-art techniques for calculating eigenvalues/eigenvectors have issues (e.g. convergence) for some matrices. We have worked out a different way of doing the calculation.

(3) *An equation for citation curves*:
If you go to Google Scholar (https://scholar.google.com/), you can find the citations for an author's publications, together with metrics like h-index that summarize the citation distribution. I have devised a simple equation that can be used to analyze citation distributions and design citation indices. See https://arxiv.org/abs/2201.04353 .

Question:
When you work with performance modelling, do you code a lot? What language primarily?

Answer:

Yes and no.

In performance modelling, you essentially work out equations that describe a computer system (e.g. https://www.comp.nus.edu.sg/~tayyc/mattyc_html/CME.html); this does not involve coding.

However, computer systems are very complicated, so the equations are usually based on some assumptions and approximations. This raises 2 questions: (1) Is the model accurate? (2) Are the conclusions derived from analyzing the equations actually properties of the real system, or merely artifacts of the model?

Checking (1) and (2) may require coding up a simulation of the system, or software to process measurements from a real system. My students use whatever language (C, C++, Java, …) that they are comfortable with, and which are technically feasible (e.g. Java may be too slow for a simulation).

Question:
How does the use local time apply to distributed computing in your research?

Answer:

In distributed computing, you have multiple computers collaborating on a task. E.g. the computers in a Mars rover, Mars reconnaissance orbiter and mission control on earth send information to each other for the software to decide what the rover should do. How can you prove that the algorithms in the software are correct?

The standard way for such correctness proofs is to define a "global state" of the system, e.g. a vector <s1(t), s2(t), s3(t)> where s1(t) is the state of the computer for the rover, s2(t) for the orbiter, and s3(t) for mission control for a "global time" t. However, Einstein already said that there is no such thing as "global time", since time is relative.

To me, it is dodgy to base a theory (correctness in distributed computing) on a concept that does not exist (global time). I therefore advocate redoing such theories by starting with "local time", i.e. using <s1(t1), s2(t2), s3(t3)> where t1 is time on the rover, t2 is time on the orbiter, and t3 is time at the mission control.

In case relativity theory seems too far from computing, note that the GPS in your cell phone (and in the aeroplane cockpit) would not work if GPS does not take relativity theory into account. See https://physicscentral.com/explore/writers/will.cfm .

Question:
I saw that you don't read other people's papers before writing your own. I find that really cool! As a high school student, do you recommend me do that as well?

Answer:

Umm, that's not quite right. It should be "I don't read other people's papers before finding my own solution to a problem." The reason: Looking at others' solutions first may get in the way of your own train of thought (and undermine your own creativity).

However, once you have a solution, you should look at what others have done, so you can relate your solution to theirs (Does your solution make more/less assumptions? Would your solution require more/less computation? …).

Or, if you get stuck, then looking at others' solutions will show you how they work around the difficulty, or point out some issue with their solution, etc.

For a high school student, the same principle applies with academic exercises: you should take on the problem yourself before you look at others' solutions, so you can compare your solutions, or else understand where your thinking went in the wrong direction, etc.

Question:
Usually when you write papers, how many other authors do you work with?

Answser:

The co-authors on most of my papers have directly contributed to the papers. E.g. one of my recent papers has just 2 co-authors X and Y, where X suggested the problem and we had multiple technical discussions, whereas Y did the data collection. This means that most of my papers only have a small number of co-authors.

In cases where there were many co-authors, that usually happened when I was not the main contributor to the paper, and the main author had some other consideration (e.g. who got the funding for the research, who their supervisor was) for including other names in the author list.

In some disciplines, the number of co-authors can be very large. E.g.
https://www.nature.com/articles/nature.2015.17567

Question:
In math & cs, do you often make new discoveries?

Answer:

The mathematics that I do are in service of computer science, so there is usually no new mathematical discovery (just application of known mathematics).

In the example that I gave above (https://arxiv.org/abs/2201.04353), the equation has a simple form that high school students would be comfortable with. The discovery there lies in finding that such a simple-looking function can tractably model citation distributions.

In another example on WiFi (https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.456.8771&rep=rep1&type=pdf), the mathematics is again simple, and the discovery there concerns how WiFi behavior can be characterized by various combinations of WiFi parameters.

Question:
Is pure mathematics or applied mathematics more worth pursuing?

Answer:

This depends on your interest and aptitude, and the particular topic.

Some people find beauty in pure mathematics, some people find satisfaction in applied mathematics; either way, they can be immensely interesting.

Aptitude plays a part: someone who is very good at applying mathematics to physics (say) may find some areas of pure mathematics very dry (e.g. proof theory). Conversely, a person who is very good at some area of pure mathematics (e.g. algebraic geometry) may find some areas of applied mathematics (numerical analysis, say) uninteresting.

Note, however, there is often no clear distinction between pure and applied mathematics. E.g. elliptic curves belong to abstract algebra, but have found important applications in cryptography (https://crypto.stanford.edu/pbc/notes/elliptic/).

Question:
What is the most important area of mathematics?

Answer:
One way to judge may be to look at the 7 Millennium Problems (https://www.claymath.org/millennium-problems). These are problems that mathematicians think will have a huge impact on mathematics if progress is made on solving them.

Another way to answer the question is that there is no "most important area", since much of mathematics are inter-related. E.g. the Langlands Program (https://mathworld.wolfram.com/LanglandsProgram.html) is a very important open problem that relates two seemingly different areas of mathematics.