Draft Part IV Project Pitch

Sam Skinner

22 Sep 2023

Elevator Pitch

In a nutshell, this project looks to investigate the possibility of real-time, 'on-the-ground' contagious disease modelling. The goal of such a system is to alert individuals that they are latently infected before they become symptomatic and infectious to others.

Contents

1 The Concept

In 2020, the world was not prepared for a pandemic. This unpreparedness was multifaceted, and the virus spread fast. Eventually a vaccine helped us to gain control, but all vaccines take a long time to develop and test and this one was no different.

In the initial stages of the pandemic, contact tracing was used extensively to try and slow the spread of the virus. Contact tracing is a labour-intensive task and was always by nature days or weeks behind.

Digital tools were deployed to augment this contact tracing, but the function of these tools were limited in scope. They were designed to help people remember things more reliably in case a contact tracer called. DP-3T[1] and similar protocols attempted to use Bluetooth®-enabled smartphones to track contact between individuals, and the NZ COVID Tracer app tracked individuals' exposure to locations. These tools didn't significantly process the data they collected. It was up to contact tracers to connect the dots.

I believe that in such situations, early in a pandemic before a vaccine is available, a tool which processes its own data in a clever way could be useful. *What if the tool could predict that you were infected by a virus in question before you could spread it, let alone return a positive test for it?*

Done right, such a tool could be both technically possible and almost as effective as a vaccine, but much quicker to deploy. The research carried out in this project would create the required engineering evidence to allow informed decisions to be made about the viability of such a tool.

1.1 Likely approach

Each person included in the system carries a device. This device acts as an electronic proxy for its bearer.

These are electronic devices. They behave in a predefined manner and can communicate between one another. If one person comes into contact with another person, one device comes into contact with another device. Contact between people translates to an exchange between devices. This allows each device to use its collected data to tally the chances its bearer is latently infected, and to broadcast this probability to neighbouring devices.

1.2 The engineering perspective

Essentially, this can be viewed as a real-time, computationally distributed prediction system. Each device continually recalculates the probability its bearer has contracted the disease in question.

The exact nature of the devices (such as whether they are digital or analog) and other such implementation details will be decided during the course of the project.

A likely challenge with such a system is that an imperfect model for transmission will cause the predictions to become less and less accurate with time — and the model will never be perfect. The simulated transmission between the devices will cease to resemble the real transmission between the humans, so its predictions will no longer be useful. Consequently, we need a way to retroactively update predictions for a large number of people whenever new data (such as a positive or negative test result) becomes available.[2] Making this work this within the bounds of causality is a key part of this project.

---

[1] Decentralised Privacy-Preserving Proximity Tracing. See https://en.wikipedia.org/wiki/Decentralized_Privacy-Preserving_Proximity_Tracing.

[2] A span of several days exists between someone becoming sick, and them receiving a positive test result. If the system is broadcasting probabilities of infection amongst devices, *historic probabilities* need to be updated and cascaded between devices each time new information becomes available.

If the system has been working on the basis that someone is not sick when really they were sick, all of the people who have been around them recently will have been exposed. The system needs some mechanism that cascades new information between devices which are no longer nearby to one another.

1.3 Outcomes & scope

It would never be possible in a Part IV project to fully implement such a system. A full-scale version would need to satisfy many other criteria (for example, privacy and statistical requirements) other than just the engineering requirements. Additionally, such a system is designed to work amongst a large population pursued by a disease — millions of people means millions of devices.

However, a lab-scale test is achievable and useful. Charting the waters as far as the engineering design is concerned would provide a lot of insight into the practical aspects of the concept; the biggest unknowns with this project are the technical aspects.[3]

Therefore, to inform future thought, research, and implementations, the project will do the following.

- Devise a method of observing and integrating exposure between people, and a way to link this information to probabilities of infection.
- Build the required devices and attempt to demonstrate infection prediction in a simulated, lab-scale environment.
- Report on the theoretical and practical aspects and identify the possibilities and drawbacks of the concept system.
- Comment on the feasibility of such a system.

The objective here is not to develop the final solution. It's to do the required research into this concept and test if it's technically feasible.

A Possible Working Principle

Half of the battle with a project like this is working out how to approach the problem. The goal is to design a system the predicts that you're latently infected before you even come down sick. This is much easier said than done.

What's to say that the problem isn't so open ended that it'll take half a year to decide how to even approach the project? To address this question, I will provide a high-level and abstract analogy. This should provide a more definite frame of reference through which the theoretical side of this project can be viewed and approached.[4]

A.1 The gist of the idea (an analogy)

Let's assume that everyone in a particular population in a particular geographic area deals in 'shares.' Imagine there are a wide variety, and they are akin to stock market shares: they are distinct and independent, and there is no limit on the number that can be held at one time.

---

[3] I would suppose that writing an acceptable virus model (and even the mass distribution of the hardware and/or software running such a model) is achievable. Whether the underpinning hardware could ever work in practice is less certain, but the best way to find out is to try to build it.

[4] Note however that the analogy is not a blueprint for a final implementation. It's just a draft and the concept will necessarily evolve over the course of the project.

Members of this population give away copies of their shares to others who are physically nearby. (They are all guilty of counterfeiting.) Suppose that people give away copies of their shares randomly but at a certain and specific average rate. The rate at which they give away shares of a particular variety depends on how many shares they hold of that same variety.

As an example, let's consider John. He holds lots of 'banana' shares and only a few 'grapefruit' ones. He holds no 'mango' shares. Consequently he will give away banana shares to people nearby very generously (at a high frequency) but will give away grapefruit ones only occasionally. He cannot give away any mango shares because he has none.

Let's say John has 50 banana shares on Tuesday. Suppose that this means he will give away a banana share to anyone nearby 50 times on Tuesday. Consider Bob, who spends half of a day with John. Bob can expect to receive about 25 banana shares from John.

Now, because these are 'shares,' they have an associated value. This value might change from day to day. Perhaps on Tuesday, banana shares were worth $1 each. Consequently, banana shares contributed $50 to John's 'net worth,' and $25 to Bob's net worth, at the end of Tuesday.

A.2 Relating the analogy to virus modelling

We have created a situation where Bob's net worth has become linked to John's net worth. Both depend on the same 'market conditions.' This is crucial. Suppose that on Thursday (two days later), banana shares double in value. This would affect Bob's net worth just the same as it affects John's. This wouldn't happen if John had simply given Bob the original value in cash.

Let's step out of the analogy now. A 'share' is an abstraction of the engineering result of this Part IV project. The 'market conditions' by which shares are valued is in reality a stream of positive and negative virus test results. An individual's 'net worth' is their probability of being infected by the virus.

On Tuesday, John thought he had a 50% chance of being infected with the virus, on account of holding $50 in banana shares. We made an implicit assumption above that spending a full day with John would cause Bob to have a 100% chance of catching a virus from John.[5] As Bob spent 50% of a day with someone with a 50% chance of being sick, he now has a 25% chance of being sick — hence Bob's $25 net worth at the end of Tuesday.

On Thursday, John tests positive. On Tuesday, he thought he had a 50% chance of being positive. We now know, with the benefit of hindsight, that he actually had a 100% chance of being positive. He gave Bob $25 worth of banana shares, so Bob thought he now had a 25% chance of being sick. Knowing what we know now, he actually has a 50% chance of being sick. He's been 'short-changed' with the shares he's been given.

The number of shares given to Bob is a done deal. We can't go back in time to change that. Instead we will use the sharemarket to pass this new information on to Bob retroactively. Let's 'revalue' banana shares at $2

---

[5] This analogy contains many such implicit assumptions. In the interests of clarity, I'm choosing simple numbers where possible even if they are inaccurate.

each. This means that John now has $100 in banana shares, which corresponds to our measurement of the probability he's sick. Crucially, Bob's shares *also* grow to $50 in value, meaning that his probability of being sick has been retroactively corrected by the new information.

We noted earlier that cash would not afford us this useful property. Had John given Bob $25 in cash, there would be no way to retroactively inform Bob of John's positive result. This would mean that when Bob spends time with other people, he reports to them an incorrect probability of being sick. But because shares can made to change in value, we have gained centralised control over a distributed phenomenon.

(Picture this: Bob received 25 banana shares from John on Tuesday. Let's say on Wednesday he spends 4.8 hours with Sue. That's one fifth of a day, and he has 25 banana shares from John. He therefore gives Sue 5 banana shares, to a value of $5. When John tests positive on Thursday, Sue's banana shares double in value. She now has $10 worth of banana shares. It's as if Sue had been given the correct information in the first place!)[6]

Giving away 'cash' is very similar to the principle employed by DP-3T, PEPP-PT,[7] and the other tracing protocols. Yes, you can connect the dots from one person to the next, but you have to do this manually. You're always playing catch up. Giving away 'shares' removes this burden: you simply publish a new list of valuations and the probabilities take care of themselves.[8] You've effectively gone back in time and corrected all of the estimates from days prior.

A.3 The bigger picture

To summarise, we have found a (metaphorical) way to pass 'probabilities of sickness' from person to person. We have also found a way to intervene and correct any errors and uncertainties naturally building up along the way.

During the covid-19 pandemic, governments hired people with mathematical backgrounds to make population-level predictions. We now have a conceptual starting point to allow us to run these disease models on the ground, in real time, and at an individual level.

This concept system isn't a direct replacement for conventional contact tracing (nor vaccines) but if done right it could be more rapid and efficient. It would work even with a large number of infected people.[9]

By cascading accurate probabilities of infection faster than the virus can spread, any outbreaks can be nipped in the bud.

---

[6] In practice, it may be necessary to limit the speed at which information propagates through the model. A virus spreads fast, but not *that* fast.

[7] Pan-European Privacy-Preserving Proximity Tracing. See https://www.pepp-pt.org.

[8] This requires a sufficiently wide variety of shares. Revaluing shares only works if it's possible to target the infected group without the noticeably affecting any unconnected people holding the same shares by coincidence only.

[9] Arguably, this system would be *more effective* when there are more cases, at least from a layman's perspective. There's more data available, and the accuracy of any individual prediction is less critical.