Covid-19 and Knowledge Based Computation

Rohit Parikh

ICLA 2021

Thanks to Shirshendu Chatterjee, Bud Mishra, Laxmi Parida, Paul Pedersen, Jayant Shah and two *ICLA* referees

March 7, 2021

Introduction: The problem of dealing with Covid-19, until a vaccine is universally administered, is to decrease the rate of transmission while getting some social and economic activity going.

Infection passes from one person A to another person B when A is infected and B is susceptible. That is to say that B is not infected and not yet immune.

Social activity also takes place when one person interacts with another. Perhaps A is a taxpayer and B is a tax consultant. Then filing the tax return may take the form of the two of them meeting. Much can be done electronically without such a meeting, but if A is a person who needs a haircut and B is a barber then it seems that it is necessary for the two to meet in order that A gets his haircut.

Authorities can block infectious encounters through some kind of lockdown. For instance visits to temples can be banned. Movie theatres can be closed.

At the same time essential services have to be allowed. You must be allowed to go to a grocery store because after all, you do have to eat.

But non-essential encounters may be blocked so that for instance all barber shops may be required to be closed. But this has the consequence that someone who needs to get a haircut is not even allowed to visit a barber who has been infected and is now immune. The ham handed rule is *block one, block all!*.

Random graphs: the study of random graphs begins with a paper by Erdös and Rényi, 1959 in which they looked into the question of *what does it mean to say that a graph is random* and what properties do random graphs (in their sense) have.

Others have followed up their work and in addition to the Erdös Rényi graphs thre are also other kinds of random graphs like those due to Watts-Strogatz and Barabási-Albert. Not being expert in graph theory, I will only say that since human beings are connected in a graph structure, it is important to ask what kind of graph we are talking about.

The issue here is the spread of infection in a graph and what we can do to stop or decrease this spread. Knowing the graph of human encounters is clearly relevant.

Knowledge Based Computation: A knowledge based program could have instructions like *do x if you know that A* where it is not necessarily specified *how* it is known whether A is true.

Such programs were investigated in (4,6). But such instructions can also arise in distributed computing.

The question for us is this. If we do not want blanket instructions like *do not go to the beach* or *do not go to temple*, but replace them by *go to temple only if you* know *that it is safe*, then we are replacing a blanket instruction by a knowledge based one. And clearly the person following this instruction needs the knowledge *whether* it is safe to go to the temple.

Testing is one way to gather this knowledge but attention needs also to be paid to spreading that knowledge to those individuals who have to carry out such instructions.

Consider for example the instruction, pay x amount of tax if your bank interest was y. You need to know what y is and your bank is required to furnish you with that knowledge. So filling out your tax return is actually a knowledge based algorithm and others have the obligation to provide you with the requisite knowledge.

Similarly with Covid-19, the moral burden is on the state to provide you with information which allows you to decide whether it is safe to go to your barber.

The program: If we allowed custom made encounters which are safe, then much more economic and social activity could proceed. And here knowledge based programs can come to the rescue. Suppose everyone follows and is required to follow one the two following programs:

Version A

1. If I know that I am infected then I should not visit someone who, I know, is susceptible.
2. If I know that I am susceptible then I should not visit someone who, I know, is infected.

These two conditions are permissive in case of ignorance. Safer versions of these two conditions are as follows:

Version B

1. If I do not know that I am not infected then I should not visit someone who, for all I know, is susceptible.
2. If I do not know that I am not susceptible then I should not visit someone who, as far as I know, is infected.

Note that in case of ignorance of my status, B1 would forbid me from visiting whereas A1 would allow it. So A1 is more permissive and less safe.

But A is compatible with more economic activity. It has more risk and more value.

If there is a great deal of testing and the results of the tests are generally available, then the two versions A and B will converge. As more information becomes available, what is allowed under B would grow and what is allowed under A would shrink. With perfect knowledge they will coincide.

Also the results need not be made generally available. I need to know that my barber is safe but I do not need to know about the status of a barber in California. But it is not enough if only governor Cuomo knows if my barber is safe. I must have that information myself, but only about my own barber. So there could be a way to balance privacy concerns with knowledge distribution.

Knowledge based algorithms give us a clear path to balancing safety and liveness. What is needed is to get the required information about who is infectious, who is susceptible and who is (currently) immune and make sure that this information is available to those individuals who need to make decisions about their own activities.

Knowledge based algorithms are implicit in many current activities like testing, or contact tracing. But since the formal, knowledge based aspect is not made explicit, some insights can be missed.

The model

Think of the model as consisting of three sets S, I, R (for susceptible, infected and recovered) which are disjoint and whose union is W (the set of all nodes).

At each moment of time t these sets have certain values S(t), I(t), R(t). From time t to t+1 some changes can take place in these three sets. Thus a susceptible person may become infected, an infected person may recover, but a recovered person stays recovered. (As we assume for now). A susceptible person only becomes infected if there is an edge at time t between it and an infected node.

Then an edge is unsafe if one of its nodes is susceptible and the other one is infected.

Now suppose we have a knowledge base and some of the nodes know whether they or their neighbors are infected or susceptible or recovered. Then they can carry out algorithm A or B as the case may be.

Under protocol A you can utilize The edge (i, j) provided you *do not know* that the edge is unsafe. Under protocol B you only utilize that edge if you *know* that the edge is safe. Condition B is more strict. Now utilization of an edge can result in some nodes becoming infected under protocol A but not under B. At the same time an edge being utilized can result in a certain positive value from the social activity.

Should we use protocol A or protocol B? If we do not know very much about who is infected and who is susceptible and so on but we do know that the number of infected nodes is very very small, then with high probability an edge is safe and therefore protocol A is a good protocol to use. It will add a lot of value and will not result in much infection.

If a large proportion of nodes are infected then method A is not safe and then we should go to protocol B. Also, even if the number of infected nodes is currently small, it could increase and we need to continue to test to make sure that it is small. Continuing to make the knowledge base larger and using protocol A with discretion might be the most beneficial method to use.

But protocol B is useful only if you have a lot of knowledge because knowledge is required for the edge to be used and if there is not enough knowledge then edges will not be used very much and there will not be enough social activity. Use of protocol B therefore requires a lot of testing and a *dissemination* of the results obtained, Many states have oscillated between lockdown and a free for all. But an oscillation between protocols A and B might be more pragmatic.

There are three kinds of edges between individuals, hard edges like that between husband and wife living in the same house. These edges are largely involuntary. And there are soft edges where it is up to the two parties concerned whether they will have an encounter. An activity edge is either a hard edge, or a soft edge where two nodes x and y interact to carry out some activity.

The goal then is to allow activity edges only when it is safe and necessary but to allow as many of them as possible.

Moreover, these permissions can be custom made. If I am recovered or my barber is not infected then I can get a haircut. If you are susceptible and your barber might be infected then you cannot get a haircut from him. It is then not necessary to block all visits to barbers.

Such considerations are implicit in seeking testing and in making use of it. But a formal framework helps to make it explicit.

Algorithm: These considerations bring us to this question. Suppose I am a policy maker, I know something about the hard edges and the soft edges, I know the size of W more or less and have a limited capacity to do testing. Say I can test 10% of W.

Then how should I best use my resources? What information shall I gather and to whom shall I convey it?

It could well happen that we come up with an NP-hard problem but it may also be that there are algorithms which are computationally cheap and which would give me a ballpark idea of how to proceed. The problem seems sufficiently simple (logically speaking) so that approximate solutions or scalable solutions would be wihin our grasp.

Evaluation: Let us make the model richer by representing it as $\mathcal{M} = (W, R, I, S, E, H)$ where H represents the hard edges and E represents the soft edges.

The policy maker P knows all the members of H and if an element e in H has the form $(i, j) : \ i, j \in W$ then both i, j know that there is such an edge e. The edges are symmetric so there is no difference between (i, j) and (j, i).

The soft edges in E are situational, they may exist at one moment and not at the next. Both i, j have to agree to the edge $e = (i, j)$ for the edge to be alive. If you want to see your barber, you may need to call him and make an appointment. He will not agree if he knows that you are infected, or that you are susceptible and he is infected.

It is not necessary for the policy maker to know all the soft edges. Suppose certain formulas of the form $S(i)$ or $I(i)$ or $R(i)$ are known to the policy maker. If I am i and I want to make contact with j then perhaps I can consult a database and find out the status of j.

So the total number of formulas that the policy maker needs to know are only at most a constant times the size of W rather than that of $W \times W$.

Let us suppose that at time t, the disjoiint sets of infected, susceptible and recovered are $I(t), S(t), R(t)$ respectively. We asume that $I(t) \cup S(t) \cup R(t) = W$. $R(t) \subseteq R(t+1)$, Thus those who are recovered at time t remain recovered. (This may not be realistic but we assume it for now).

$S(t+1) = S(t) - I(t+1)$. Everyone who is susceptible at t remains so, or is infected at $t+1$.

$R(t+1) = R(t) \cup (I(t) - I(t+1))$ If you are recovered at time $t+1$ then either you were already recovered or you just recovered.

This is an incomplete set. More conditions need to be added.

Note that if j is in I at time $t + 1$ but not in it at time t then there must have been either a hard edge (i, j) at time t with i infected, or such a live soft edge. You can be infected from your spouse whether you chose or not, but you can only be infected from your barber if the two of you chose to meet.

Suppose that P at time t has a knowledge base KB of *some* formulas of the form $R(i, t), S(i, t), I(i, t)$ which tells her who is infected, who is susceptible and who is recovered. She shares these formulas with some members of W or makes them available in case of need. Then they carry out algorithm B. No new people will be infected but also some people who were infected recover (or die).

Proposition

Assume that protocol B is used.

- *If H is empty and all members of W observe protocol B, then* $I(t+1) \subseteq I(t)$. *No one is infected who was not already.*
- *If H is not empty let* C_i *be the hard edge component of node* i. *That is,* C_i *is the smallest set such that* $i \in C_i$ *and If* $(j, k) \in H$ *and* $j \in C_i$ *then* $k \in C_i$. *Then if* i *is infected, then assuming no recovery, all of* C_i *will be eventually infected.*

Proof:

- *Note that protocol B allows no unsafe soft edges. And since H is empty, there are indeed no unsafe edges. So no new person gets infected.*
- *If there are indeed hard edges then infection can spread along a hard edge. But starting from* i, *it can only spread inside the hard component of* i. □

Note that if some infected people recover sponanteously then it will not be the case that everyone in C_i will be infected in the limit. Even if there is a chain from i to j of hard edges, some people in that chain might recover and j might never get infected.

Some questions

a) Given some reasonable conditions on how KB is used, show that if $KB \subset KB'$ then KB' will give better results. To measure this, give a negative value v to each person who is freshly infected from time t to $t+1$ and a positive value v' to each edge in E which is activated between t and $t+1$. The total value of the transition is the sum of all the v' minus the sum of all the v's. Then the value of KB' will be larger. Why? Because some edges will be activated under KB' which were not activated under KB.

Of course we need some uniformity in how KB and KB' are used but the result will hold if every individual who is informed of something under KB is informed of the same fact under KB'. For an edge which is live under KB can also be live under KB'. Under prorocol B, $KB's$ will certainly be better by allowing more activity and no new infections.

Under protocol A, there could be new infections, and the gain from the extra activity will need to be balanced against the loss from the extra infection.

Note that if the negative value v to each person who is freshly infected is infinite then protocol A simply cannot be used because we are not allowing tradeoffs. No number of v' will make up for a single v. It is clear that Sweden used a modest value fo v resulting in more infections and more activity. Was this a wise choice? Depends on the comparison between v and v'.

Note that even if we are given the current state and the KB including its distribution, we cannot *rigorously* predict the next state. Someone who is *allowed* to use an edge, might not use it.

If there is an unsafe edge it still might happen that no infection passes - perhaps because they were not close enough or for long enough.

But still, it is intuitively obvious that the larger knowledge base gives better performance. Anything which can be safely done under KB can also be safely done under KB' and more safe things can be possible under KB' *if* we are using protocol B.

Things are more complex if we are using protocol A.

However if $KB \subseteq KB'$ then every edge allowed under KB' is also allowed under KB. Thus any activity which can be performed under KB' can also be performed under KB. Since we are using protocol A, there are fewer infections under soft edges and the hard edges are the same. Whether KB' gives a gain relative to KB will be more complex.

There is a saying for young men wishing to date, faint heart never won fair lady which seems to be saying, use protocol A. However, New York governor Cuomo used protocol A and he is now in deep trouble!

Conclusion: Some of these considerations will become moot when everyone is vaccinated. But that may take some time and in meanwhile knowledge based algorithms may be of use. Also, when work is theoretical, it can always find application in an area which was not anticipated at the start.

References

1. Brummitt, Charles D., et al. "Jigsaw percolation: What social networks can collaboratively solve a puzzle?." *The Annals of Applied Probability* (2015): 2013-2038.
2. Dean, Nur, *Role of Influence in Complex Network* , Doctoral disssertation, CUNY 2020.
3. Durrett, Rick. "Some features of the spread of epidemics and information on a random graph." Proceedings of the National Academy of Sciences 107.10 (2010): 4491-4498.
4. Fagin, Ronald, et al. "Knowledge-based programs." *Distributed Computing* 10.4 (1997): 199-225.
5. Marathe, Madhav, and Anil Kumar S. Vullikanti. "Computational epidemiology." *Communications of the ACM* 56.7 (2013): 88-96.
6. Parikh, Rohit. "Knowledge based computation." *International Conference on Algebraic Methodology and Software Technology.* Springer, Berlin, Heidelberg, 1995 127-132.
7. Covid-18, symptoms and spread, N. D. Hari Dass, Shantala Hari Dass, Kamal Lodaya & R. V. Vandana, *IWonder*, October 2020