Morning everybody. I am read here to present our work titled ‘An integrated epidemic simulation workflow for submodular intervention strategies’.

As a result of the ongoing Covid-19 pandemic there has been a dramatic increase in the need for epidemic simulators. There has also been algorithmic advancement in the field of network science based vaccination strategies.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The objective of this work is simple: integrate these two together into a workflow so that one can devise various individual level vaccination strategies and simulate its effect on an epidemic simulator. The simulator we use here is called covasim which is developed by the institute for disease modeling based in Seattle. More on that later. Though we use covasim the workflow is generic enough to accommodate other agent-based simulators.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A bit of background before we get started. The main motivation behind network science based intervention strategies, come from the problem formulation of Influence maximization.

It is when you're given a graph g with vertex set v and a weighted edge set E. There is a diffusion process which dictates how influence spreads on the graph.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The problem is to identify k nodes to influence initially such that the final expected influence spread from these initial seeds is maximized.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

This brings us to Preempt which is a framework published in 2020, which reformulates Inf Max problem as the epiControl problem, Given a contact network g and initial set of infected nodes b, the problem is to find k seeds to vaccinate so that the expected number of infections at the end of the diffusion process is minimized or the number of lives saved, which is the number of infections from doing nothing minus the number of infections from vaccinating seed set S, is maximized.

In the paper they prove that the objective function is sub modular when a particular realization of the contact network is a rooted tree, or the disease propagation happens in the form of a rooted tree, for example in sexually transmitted diseases, but it serves a very good heuristic for other contact networks as well where those guarantees are not met.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

This brings us to the simulator component of the workflow. For that we use covasim which is a stochastic agent-based simulator used to simulate the spread of covid-19

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

over synthetic populations based on real world demographic data.  
The population is in the form of a contact network where every agent is a node. Two agents can be connected by an edge if they interact with each other, let's say in a household or school, work or out in the community. For an edge from u to v the weight represents the probability that u will infect v provided u is infectious already.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Lastly, it supports highly customizable intervention policies at the individual level

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Putting these together, this is the rough workflow where we start the simulation and let it run for a month so that there is a pandemic or the pandemic develops into pandemic,

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

then on 31st day, we extract the contact network.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Use preempt or any other seed selection strategy to select seeds to vaccinate

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

and vaccinate them.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Then we let it run for a week.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

We repeat the same process every week.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

This continues for n such rounds.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Then we extract the metrics to compare the efficacy of the strategy. For baselining purposes, we also use random and degree of a node as cheap heuristics to select seeds to see how well preempt does in minimizing the infections.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

This brings us to the set of initial experiments we conducted using this workflow. As input, we use a synthetic population based on the demographic data of India in terms of the age and household size distribution. The contact network is of size 100k. The degree distribution of the network is shown on the right. As you can see this is a very typical real world contact network which are known to have a multimodal degree distribution. The simulation duration is over five months.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The metrics we use to evaluate the efficacy of one vaccination scheme against another are cumulative infections, daily number of infections, and cumulative deaths.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

in the first set of experiments we've vaccinated certain percentages of the population in just one round, which is shown in the x-axis by using different seed selection strategies.

The value in using preempt as a seed selection strategy is increasingly more as you increase the percentage of population vaccinated. random seed selection strategy is always the worst which is why it is excluded from the future charts to increase the readability. Please note that social distancing, masks etc are not enabled here in the simulator to make the numbers reflect the effects of vaccinations only. That is why the percentage of people infected is so inflated. So it makes sense to look at them in a relative manner.

Now vaccinating around 20% of the population in one go like shown here is unrealistic at the beginning of the pandemic so it makes sense to batch it out.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In this set of experiments we vaccinate a total of 20% of the population but in batches of thousand or 1% in each of the 20 rounds represented by the vertical lines. And compare the curves with that of single round where twenty percent is vaccinated

The higher the curve, worse is the vaccination strategy.

We can see that single round is better as expected but unrealistic as I mentioned before. In the batched version pre-empt outperforms degree, and in fact, it is comparable to the single round degree for daily infections and cumulative deaths.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Next we see the effect of non-uniform batch size. For that we use a top heavy strategy to reflect the real world aggressive vaccinations at the beginning of the pandemic. 2000 per round for the first five rounds, thousand per round for the next five and five hundred per round for the last ten. We see that irrespective of the batch size, preempt is better than degree.

Top heavy batched preempt is better than uniform batch preempt.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In conclusion we can say that this work provides a framework that integrates graph theoretic intervention strategies into epidemic simulation.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Inf Max based heuristics are better at identifying key nodes to vaccinate based on what they represent in the context of the contact network.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Identifying which nodes to vaccinate isn't enough timing also plays a key role in not letting the pandemic get out of control.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

This an acknowledgement of the funds and a shoutout to the developers at Covasim for their help while I was familiarising myself with the simulator.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Thank you.