An Integrated Epidemic Simulation Workflow for Submodular Intervention Strategies

epiDAMIK 4.0: The 4th International workshop on Epidemiology meets Data Mining and Knowledge discovery

Reet Barik [1], Marco Minutoli [2], Mahantesh Halappanavar [2,1], Ananth Kalyanaraman [1,2]

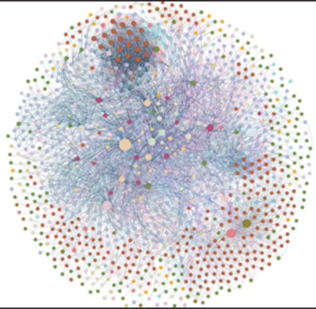
[1]School of Electrical Engineering and Computer Science, Washington State University

[2]Pacific Northwest National Laboratory

August 15, 2021

Objective

Integrate network-based intervention policies at the most granular level into an infectious disease simulation workflow.



Individual-level intervention (vaccination) policies

COVASIM by IDM

Agent-based Covid-19 Simulator

Influence Maximization based Intervention

Influence Maximization: Given a graph $G(V, E)$ of n nodes and m edges, and a diffusion process that dictates how the information is spread from node to node, the classical problem [1] is one of identifying a set of k nodes that is expected to maximize the influence spread based the diffusion process.

The approx. solution based on the greedy algorithm gurantees a $(1 - 1/e - \epsilon)$ bound to the optimal solution for diffusion models with submodular objective functions (like LT and IC).

Targeted Immunization: The problem to identify nodes to vaccinate in a given network, such that the expected disease spread is minimized.

These two problem have been shown to be equivalent under the Linear threshold model in [2].

[1] Kempe, D., Kleinberg, J. and Tardos, É., 2003, August. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 137-146).

[2] Chun-Hung Cheng, Yong-Hong Kuo, and Ziye Zhou. Outbreak minimization vs influence maximization: an optimization framework.BMC Medical Informaticsand Decision Making, 20(1):1–13, 2020

Influence Maximization based Intervention

Influence Maximization: Given a graph $G(V, E)$ of n nodes and m edges, and a diffusion process that dictates how the information is spread from node to node, the classical problem [1] is one of identifying a set of k nodes that is expected to maximize the influence spread based the diffusion process.

The approx. solution based on the greedy algorithm gurantees a $(1 - 1/e - \epsilon)$ bound to the optimal solution for diffusion models with submodular objective functions (like LT and IC).

Targeted Immunization: The problem to identify nodes to vaccinate in a given network, such that the expected disease spread is minimized.

These two problem have been shown to be equivalent under the Linear threshold model in [2].

[1] Kempe, D., Kleinberg, J. and Tardos, É., 2003, August. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 137-146).

[2] Chun-Hung Cheng, Yong-Hong Kuo, and Ziye Zhou. Outbreak minimization vs influence maximization: an optimization framework.BMC Medical Informaticsand Decision Making, 20(1):1–13, 2020

EpiControl

Given a contact network $G = (V, E, w)$, a diffusion model M over G, a set of initially infected vertices $B \subseteq V$, and a fixed budget k, the EpiControl problem is to find an intervention set $S \subseteq V$ of size k such that the expected number of infections at the end of the diffusion process $(\sigma(B, S))$ is minimized.

PREEMPT [3]

EpiControl reformulated by defining *Number of lives saved*:

$$\lambda_{G_i}(B, S) = \sigma_{G_i}(B, \emptyset) - \sigma_{G_i}(B, S).$$

Given a contact network $G = (V, E, w)$, a diffusion model M over G, a set of initially infected vertices $B \subseteq V$, and a fixed budget k, the EpiControl problem is to find an intervention set $S \subseteq V$ of size k such that the expected number of lives saved at the end of the diffusion process $(\lambda(B, S))$ is maximized.

[3] Marco Minutoli, Prathyush Sambaturu, Mahantesh Halappanavar, AntoninoTumeo, Ananth Kalyanaraman, and Anil Vullikanti. Preempt: scalable epidemicinterventions using submodular optimization on multi-gpu systems. In Proceedings of the International Conference for High Performance Computing, Networking,Storage and Analysis, SC 2020, page 55.

EpiControl

Given a contact network $G = (V, E, w)$, a diffusion model M over G, a set of initially infected vertices $B \subseteq V$, and a fixed budget k, the EpiControl problem is to find an intervention set $S \subseteq V$ of size k such that the expected number of infections at the end of the diffusion process ($\sigma(B, S)$) is minimized.

PREEMPT [3]

EpiControl reformulated by defining *Number of lives saved*:

$$\lambda_{G_i}(B, S) = \sigma_{G_i}(B, \emptyset) - \sigma_{G_i}(B, S).$$

Given a contact network $G = (V, E, w)$, a diffusion model M over G, a set of initially infected vertices $B \subseteq V$, and a fixed budget k, the EpiControl problem is to find an intervention set $S \subseteq V$ of size k such that the expected number of lives saved at the end of the diffusion process ($\lambda(B, S)$) is maximized.

[3] Marco Minutoli, Prathyush Sambaturu, Mahantesh Halappanavar, AntoninoTumeo, Ananth Kalyanaraman, and Anil Vullikanti. Preempt: scalable epidemicinterventions using submodular optimization on multi-gpu systems. In Proceedings of the International Conference for High Performance Computing, Networking,Storage and Analysis, SC 2020, page 55.

Covasim [4]

- Stochastic agent-based simulator that is used to simulate the spread of the Covid-19 disease.
- Builds synthetic populations as a network based on real-world demographic data:
  - Every agent is a node in the network
  - Nodes are connected based on if they interact with each other (inside a household, school, work, and out in the community)
  - The edge weights are the probabilities of the source nodes infecting the destination nodes given that the source nodes themselves are infected.
- Supports highly customizable intervention (vaccination) policies.

[4] Cliff C Kerr, Robyn M Stuart, Dina Mistry, Romesh G Abeysuriya, Gregory Hart,Katherine Rosenfeld, Prashanth Selvaraj, Rafael C Nunez, Brittany Hagedorn,Lauren George, et al. Covasim: an agent-based model of covid-19 dynamics andinterventions.medRxiv, 2020.

Covasim [4]

- Stochastic agent-based simulator that is used to simulate the spread of the Covid-19 disease.
- Builds synthetic populations as a network based on real-world demographic data:
  - Every agent is a node in the network
  - Nodes are connected based on if they interact with each other (inside a household, school, work, and out in the community)
  - The edge weights are the probabilities of the source nodes infecting the destination nodes given that the source nodes themselves are infected.
- Supports highly customizable intervention (vaccination) policies.

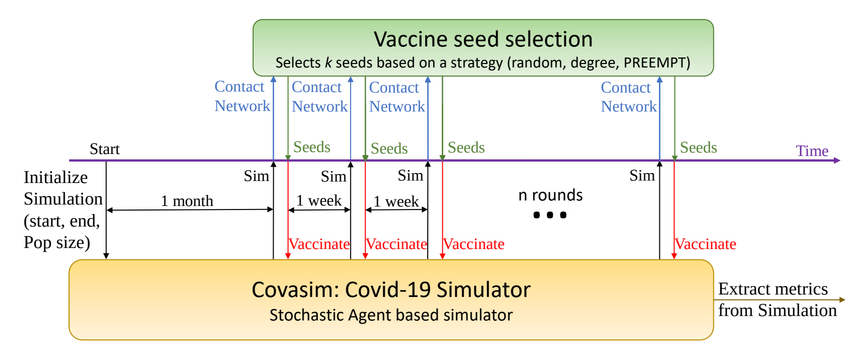
[4] Cliff C Kerr, Robyn M Stuart, Dina Mistry, Romesh G Abeysuriya, Gregory Hart,Katherine Rosenfeld, Prashanth Selvaraj, Rafael C Nunez, Brittany Hagedorn,Lauren George, et al. Covasim: an agent-based model of covid-19 dynamics andinterventions.medRxiv, 2020.

Covasim [4]

- Stochastic agent-based simulator that is used to simulate the spread of the Covid-19 disease.
- Builds synthetic populations as a network based on real-world demographic data:
  - Every agent is a node in the network
  - Nodes are connected based on if they interact with each other (inside a household, school, work, and out in the community)
  - The edge weights are the probabilities of the source nodes infecting the destination nodes given that the source nodes themselves are infected.
- Supports highly customizable intervention (vaccination) policies.

[4] Cliff C Kerr, Robyn M Stuart, Dina Mistry, Romesh G Abeysuriya, Gregory Hart,Katherine Rosenfeld, Prashanth Selvaraj, Rafael C Nunez, Brittany Hagedorn,Lauren George, et al. Covasim: an agent-based model of covid-19 dynamics andinterventions.medRxiv, 2020.

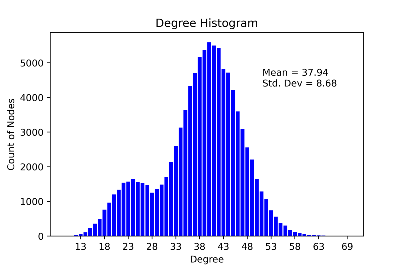
Integrated Workflow



The simulation is allowed to simulate the start of the pandemic unhindered for a month followed by regular vaccination rounds of certain batch sizes every week. Nodes to be vaccinated are specified by a seed selection strategy, which could internally implement various strategies to identify those seeds.

Input

- Simulated location: India
- $|V| = 100,000$
- $|E| = 3,793,826$
- Duration of each run of the simulation: 170 days (over 5 months) – starting on January 1st and ending on June 19th.

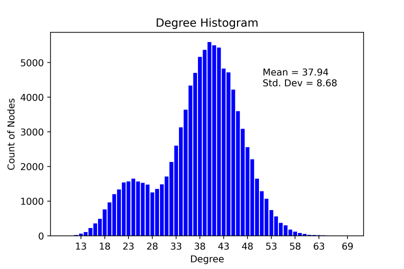


Evaluation Metrics

- the number of *cumulative infections*
- The number of *new infections per day*
- The number of *cumulative deaths*

Input

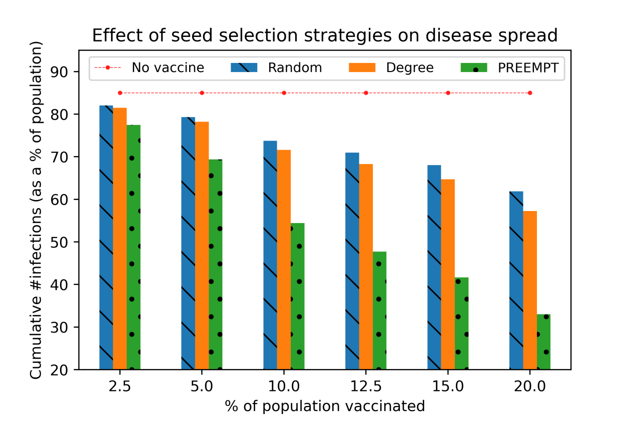
- Simulated location: India
- $|V| = 100,000$
- $|E| = 3,793,826$
- Duration of each run of the simulation: 170 days (over 5 months) – starting on January 1st and ending on June 19th.



Evaluation Metrics

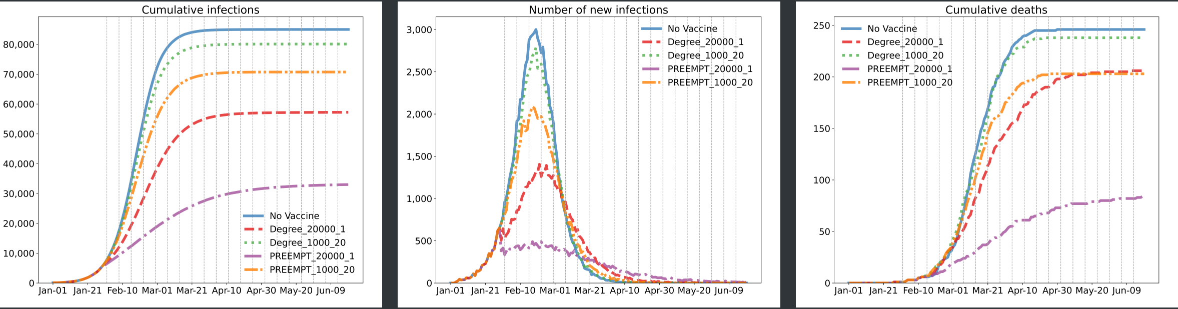
- the number of *cumulative infections*
- The number of *new infections per day*
- The number of *cumulative deaths*

Effect of seed selection strategies



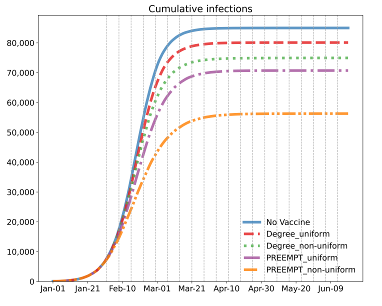
The x-axis represents the % of population vaccinated at a single round, on the 31st day of the simulation. The y-axis represents the cumulative #infections after 5+ months of simulation as a % of the population infected.

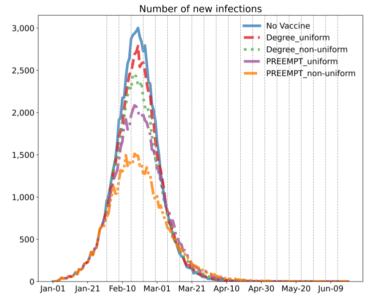
Effect of vaccinating in batches of uniform size

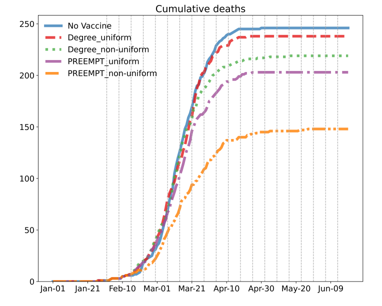


For every curve labeled as 'X_Y_Z', 'X' stands for the seed selection strategy; 'Y' stands for the (uniform) batch size used at each round; and 'Z' stands for the number of rounds. Every vertical dashed line represents a vaccination round. Also shown for comparative reference, is the 'No vaccine' curve that corresponds to zero vaccines given out at each round. We use 20,000 as the total number of vaccines.

Effect of vaccinating in batches of non-uniform sizes







The plots are labeled as ‘X_Y’ where ‘X’ stands for the seed selection strategy and ‘Y’ represents the two batching strategies—*uniform* or *non-uniform*. The uniform strategy applies 1,000 vaccines per round. The non-uniform strategy uses 2,000 vaccines in each of the first 5 rounds; 1,000 vaccines in each of the next 5 rounds; and 500 vaccines per round in the final 10 rounds. Also shown is the ‘No vaccine’ curve for reference.

Conclusion

- Provide a framework that integrates epidemic simulation with graph-theoretic/network science-based interventions.
- There is value in using an influence maximization based approach toward seed selection in the context of epidemic control. In particular, PREEMPT as a seed selection strategy is able to outperform other heuristics like degree or random schemes.
- In addition to a *carefully selected* subset of seeds, our experiments also demonstrate that the *timing* of these vaccination matter—i.e., giving more vaccines early on could save more lives in the long run.

Conclusion

- Provide a framework that integrates epidemic simulation with graph-theoretic/network science-based interventions.
- There is value in using an influence maximization based approach toward seed selection in the context of epidemic control. In particular, PREEMPT as a seed selection strategy is able to outperform other heuristics like degree or random schemes.
- In addition to a *carefully selected* subset of seeds, our experiments also demonstrate that the *timing* of these vaccination matter—i.e., giving more vaccines early on could save more lives in the long run.

Conclusion

- Provide a framework that integrates epidemic simulation with graph-theoretic/network science-based interventions.
- There is value in using an influence maximization based approach toward seed selection in the context of epidemic control. In particular, PREEMPT as a seed selection strategy is able to outperform other heuristics like degree or random schemes.
- In addition to a *carefully selected* subset of seeds, our experiments also demonstrate that the *timing* of these vaccination matter—i.e., giving more vaccines early on could save more lives in the long run.

Acknowledgments

The research is in parts supported by the U.S. DOE ExaGraph project at the Pacific Northwest National Laboratory (PNNL), and by the U.S. National Science Foundation (NSF) grants CCF 1815467, OAC 1910213, and CCF 1919122 to Washington State University. PNNL is operated by Battelle Memorial Institute under Contract DE-AC06- 76RL01830. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

Thank You