Scalable Epidemiological Agent-based Modeling with Dynamic Behaviors

Joy Kitson
University of Maryland jkitson@umd.edu
Supervisor: Abhinav Bhatele

I. Introduction

Over the course of the COVID-19 pandemic, a wide variety of modeling approaches have been used to inform policymakers at various levels of government. This process has helped illuminate both the benefits and the limitations of using different modeling techniques in this role. Agent-based models (ABMs) – where the behaviors of individual members of a population are simulated directly – have proved particularly well-suited in use cases like counterfactual analysis. Counterfactual analysis seeks to understand the impact of different sets of public health interventions in various what-if scenarios, which are often easier to directly represent in ABMs.

However, this resolution comes at a cost; ABMs are generally orders of magnitude more complex and computationally expensive than other modeling techniques. This generally means that such ABMs must be highly scalable parallel applications. Existing ABMs are generally either (1) complex, small simulations (at most around a million agents) with a focus on epidemiological results rather than computational efficiency, or (2) simpler, large models where much of the complexity lies in the underlying datasets and the behavioral model of agents remains relatively simple (e.g. a sequence of top-down interventions determines behavior)

With these limitations of existing simulations in mind, we propose to enable ABMs of infectious disease spread to efficiently scale to large populations and core counts while efficiently modeling a combination of top-down and bottom-up behaviors that are both complex and dynamic.

II. Background and Related Work

As a class of simulations, epidemiological ABMs provide a wide variety of ways to answer the same basic questions: (1) who does each individual agent make contact with in a given period of time? and (2) given a set of contacts, is an agent infected by one of them? Whether the first question is answered based on an input network of pairwise contact probabilities, individual travel schedules, or stochastic movement within a region, answering the second is a matter of summing infection probabilities based on the state of those contacts. Protective behaviors and public health interventions can change the answer to either question. For example, school closures can change contacts, while vaccinations can change infection probabilities even when contacts are fixed.

While many models of this type have been developed, few have been scaled to the hundreds of millions of agents needed to simulate the population of a country like the US. EpiCast [1] was one of the first models to achieve these scales. EpiCast adapted the SPaSM molecular dynamics code by treating agents as particles interacting within fixed communities based on where they lived and worked. Another framework that achieved similar scales is EpiSimdemics [2], which proposed a novel parallel algorithm of combining time-stepping and discrete event simulation (DES) to simulate disease spread given a visit schedule for each agent with discrete locations.

Orthogonal to efforts to scale ABMs to large populations, there have been efforts to better model how those populations behave. Most current models rely on coarse-grained top-down interventions or relatively rigid data-driven behaviors, neither of which capture the feedback loops between disease spread and people's behaviors observed in the real world. Coupled contagion models [3], where information about – or fear of – a disease spreads alongside the disease itself and influences behavior, provide one way of representing the interplay between disease and fear spread. However, these have yet to be applied to ABMs at scales beyond 10 million agents.

III. Research Questions and Approach

Our work involves answering three main questions:

- **Q1** How can we efficiently scale ABMs to large populations of agents?
- **Q2** How can we incorporate granular dynamic behavior into ABMs without sacrificing scalability?
- **Q3** How do the dynamics of large-scale ABMs change as dynamic behavior is introduced?

For **Q1**, we build on top of the general framework used in EpiSimdemics [2], first creating our own implementation of the algorithm and then introducing optimizations to address issues such as load imbalance and communication overhead. For **Q2** and **Q3**, we begin by implementing a variant of Epstein et al.'s coupled contagion model [3] in EpiCast [1] with purely local fear spread, then seek to extend it to model the impact of fear spread through other mechanisms, such as broadcasters and social media.

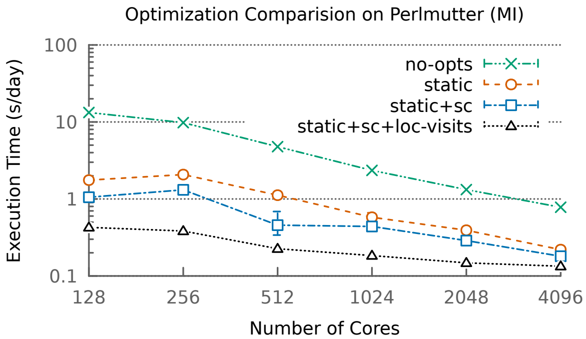

Fig. 1: Performance comparison of combinations of different performance optimizations with each added optimization reducing runtimes. Execution times averaged over three runs, extrema shown in error bars.

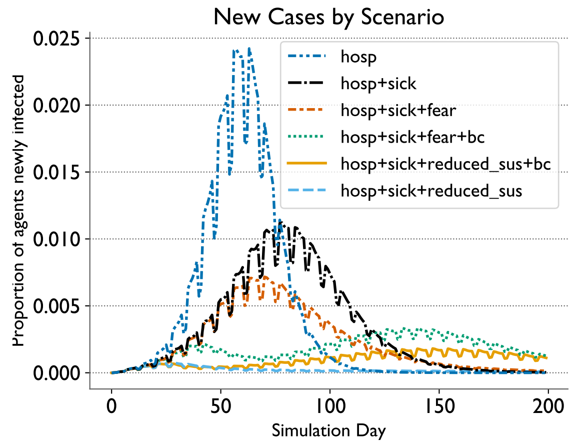

Fig. 2: New cases for six different fear spread scenarios run in EpiCast.

IV. Contributions and Achievements

A. Research results

For the scaling work (**Q1**), we have developed a new simulator, Loimos, to test our scaling strategies. Loimos uses a combination of time-stepping and DES, and is implemented on top of Charm++, an asynchronous task-driven parallel runtime system used for other highly heterogenous applications such as the NAMD molecular dynamics simulator.

We implement three main performance optimizations in Loimos, the impact of which is shown in Figure 1: (1) a static load balancing algorithm that maintains geographic locality by partitioning locations using linear cuts after sorting by ZIP codes (static), (2) a short-circuit optimization where we only evaluate the DES for locations visited by at least one infectious person on a given day (sc), and (3) a modification to the core algorithm for computing contacts which reduces communication by storing visit information on the process where the location data is stored, not the person data (loc-visits). This allowed us to send minimal data on each person's current state rather than their full visit schedule before processing contacts for a given simulation day. Altogether, these optimizations result in a $31.03\times$ speedup on 128 cores, when run on a realistic synthetic dataset of ~ 9.3 million agents representing the population of the US state of Michigan. This work is being published in IPDPS this year.

For **Q2-Q3**, we have implemented a dynamic behavioral model in EpiCast. This model couples the spread of fear of a disease with that of the disease itself. Interactions with afraid or symptomatic agents spread fear, which causes agents to engage in protective behaviors that reduce disease spread. We have also developed a second mechanism for fear spread based on a network of local broadcasters, which can spread fear or mitigate fear spread based on a combination of local fear and disease spread. We have also reworked the transmission computation to bin contacts in a given context rather than compute them pairwise, turning a $7.3\times$ slowdown over the original code into a 11% speedup. This model has already proved capable of producing multiple epidemic waves in a simulation of the contiguous United States (CONUS). Figure 2 shows the how the proportion of new cases (top) and fear prevalence vary for a range of different fear-spread scenarios.

B. Dissemination and Community Engagement

I have presented talks on the development of Loimos at the 2022 and 2024 Charm++ workshops, along with a poster at SC22 and talks at the UMD booth for SC23 and SC24. I also presented a lightning talk on the behavioral work in EpiCast at Los Alamos National Laboratory in 2024.

V. Summary and Outlook

We are currently investigating ways to better understand how the properties of the underlying visit network impact the performance of Loimos runs. This involves building and empirically validating a performance model based on a variety of network theoretic properties and searching for network sparsification techniques that preserve the statistical distribution of simulation outcomes.

We are working on extending our fear spread model to use an arbitrary person-person network as input, which we can then tune to represent real world social networks, such as those enabled by social media. We intend to use a variation on the communication scheme developed for Loimos to minimize communication overhead.

References

[1] T. C. Germann, K. Kadau, I. M. Longini Jr, and C. A. Macken, "Mitigation strategies for pandemic influenza in the united states," *Proceedings of the National Academy of Sciences*, vol. 103, no. 15, pp. 5935–5940, 2006.

[2] C. L. Barrett, K. R. Bisset, S. G. Eubank, X. Feng, and M. V. Marathe, "Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks," in *SC'08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*. IEEE, 2008, pp. 1–12.

[3] J. M. Epstein, J. Parker, D. Cummings, and R. A. Hammond, "Coupled contagion dynamics of fear and disease: mathematical and computational explorations," *PloS one*, vol. 3, no. 12, p. e3955, 2008.