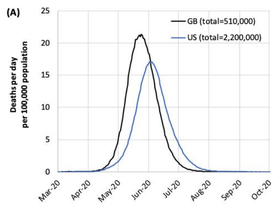
Generalizing Bulk-Synchronous Parallel Model for Data Science: From Data to Threads and Agent-Based Simulations

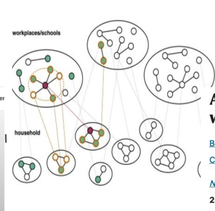
Zilu Tian[1], Peter Lindner[1], Christoph Koch[1], Markus Nissl[1], Val Tannen[2]
[1]DATA, EPFL [2]University of Pennsylvania

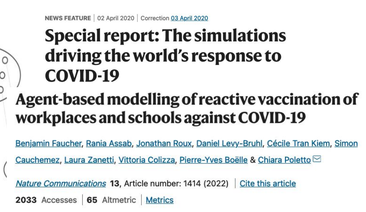
Why agent-based simulations matter?

- Epidemics

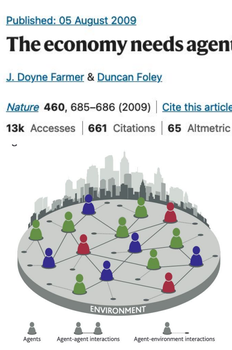



London's Imperial College predicts millions to die from coronavirus pandemic in UK and
The New York Times
LIVE Covid-19 Updates | Map and Cases | Nearing 1 Million U.S. Deaths
Behind the Virus Report That Jarred the U.S. and the U.K. to Action
It wasn't so much the numbers themselves, frightening though they were, as who reported them: Imperial College London.




NEWS FEATURE | 02 April 2020 | Correction 03 April 2020
Special report: The simulations driving the world's response to COVID-19
Agent-based modelling of reactive vaccination of workplaces and schools against COVID-19
Benjamin Faucher, Rania Assab, Jonathan Roux, Daniel Levy-Bruhl, Cécile Tran Kiem, Simon Cauchemez, Laura Zanetti, Vittoria Colizza, Pierre-Yves Boëlle & Chiara Poletto
Nature Communications 13, Article number: 1414 (2022) | Cite this article
2033 Accesses | 65 Altmetric | Metrics

- Economics


Published: 05 August 2009
The economy needs agent-based modelling
J. Doyne Farmer & Duncan Foley
Nature 460, 685–686 (2009) | Cite this article
13k Accesses | 661 Citations | 65 Altmetric | Metric

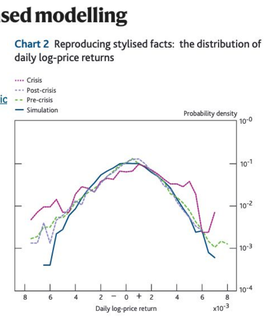


Chart 2 Reproducing stylised facts: the distribution of daily log-price returns

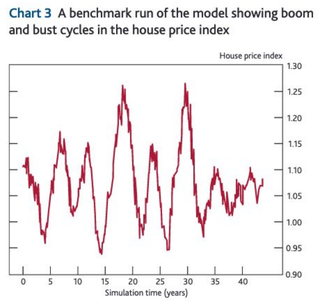


Chart 3 A benchmark run of the model showing boom and bust cycles in the house price index

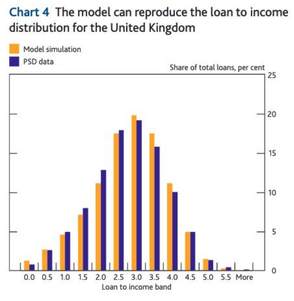


Chart 4 The model can reproduce the loan to income distribution for the United Kingdom

Challenges of agent-based simulations

- Agent-based simulations are flexible, but inefficient to execute
  - High concurrency
    - A realistic simulation has billions of agents
  - Code heterogeneity
    - "Think like a vertex" is homogeneous
  - Communication-intensive
    - Existing frameworks assume little or no communication

Pregel | APACHE GIRAPH | GraphX | APACHE Spark

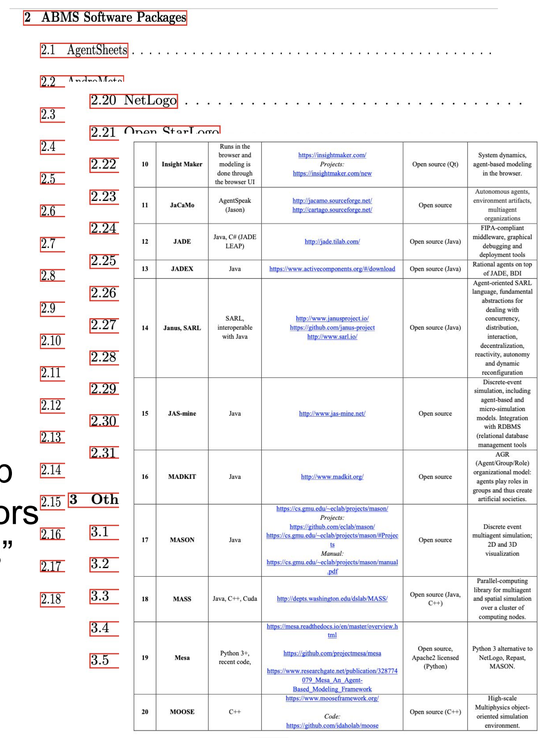
- For data management
  - Simulations generate a large amount of data
    - Image long-running simulations with billions of agents
  - Simulations form part of complex analytics pipelines
    - "How does the average wealth of the top 30% change?"
  - Simulations are can be viewed as model samples
    - "What is the average wealth of the population if we increase the initial wealth by 10%, 20% and 50%, respectively?"

What are agent-based simulations, really?

- Depend on who you ask!
  - A recent survey in 2020 listed
    - 36 general-purpose frameworks
    - 100+ specialized frameworks
  - Different assumptions about agents
    - NetLogo considers turtles as agents, along with patches and links
    - DMASON assumes each agent belongs to a temporal region
    - Repast Symphony assumes that agents actions are scheduled
  - Different assumptions about interaction
    - NetLogo assumes spatial-based interaction
    - DMASON is based on publish-subscribe paradigm
    - Repast Symphony allows instant changes to other agents' states

- The lack of formal models causes high heterogeneity
  - Increase users' learning curves
  - Decrease cross-platform result verification
  - Hard to select the right tool
  - Limit performance optimizations to framework-dependent

- Generally speaking, frameworks are round-based or asynchronous
  - whether agents proceed in lockstep
  - But frameworks have different flavors of "round-based" or "asynchronous"


2 ABMS Software Packages
2.1 AgentSheets
2.20 NetLogo
2.21 Open StarLogo
3 Oth

10	**Insight Maker**	Runs in the browser and modeling is done through the browser UI	https://insightmaker.com/ *Projects:* https://insightmaker.com/new	Open source (Qt)	System dynamics, agent-based modeling in the browser.
11	**JaCaMo**	AgentSpeak (Jason)	http://jacamo.sourceforge.net/ http://cartago.sourceforge.net/	Open source	Autonomous agents, environment artifacts, multiagent organizations
12	**JADE**	Java, C# (JADE LEAP)	http://jade.tilab.com/	Open source (Java)	FIPA-compliant middleware, graphical debugging and deployment tools
13	**JADEX**	Java	https://www.activecomponents.org/#/download	Open source (Java)	Rational agents on top of JADE, BDI
14	**Janus, SARL**	SARL, interoperable with Java	http://www.janusproject.io/ https://github.com/janus-project http://www.sarl.io/	Open source (Java)	Agent-oriented SARL language, fundamental abstractions for dealing with concurrency, distribution, interaction, decentralization, reactivity, autonomy and dynamic reconfiguration
15	**JAS-mine**	Java	http://www.jas-mine.net/	Open source	Discrete-event simulation, including agent-based and micro-simulation models. Integration with RDBMS (relational database management tools
16	**MADKIT**	Java	http://www.madkit.org/	Open source	AGR (Agent/Group/Role) organizational model: agents play roles in groups and thus create artificial societies.
17	**MASON**	Java	https://cs.gmu.edu/~eclab/projects/mason/ *Projects:* https://github.com/eclab/mason/ https://cs.gmu.edu/~eclab/projects/mason/#Projects *Manual:* https://cs.gmu.edu/~eclab/projects/mason/manual.pdf	Open source	Discrete event multiagent simulation; 2D and 3D visualization
18	**MASS**	Java, C++, Cuda	http://depts.washington.edu/dslab/MASS/	Open source (Java, C++)	Parallel-computing library for multiagent and spatial simulation over a cluster of computing nodes.
19	**Mesa**	Python 3+, recent code,	https://mesa.readthedocs.io/en/master/overview.html https://github.com/projectmesa/mesa https://www.researchgate.net/publication/328774079_Mesa_An_Agent-Based_Modeling_Framework	Open source, ApacheZ licensed (Python)	Python 3 alternative to NetLogo, Repast, MASON.
20	**MOOSE**	C++	https://www.mooseframework.org/ *Code:* https://github.com/idaholab/moose	Open source (C++)	High-scale Multiphysics object-oriented simulation environment.

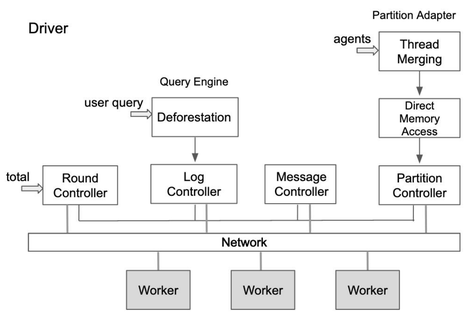
Contributions

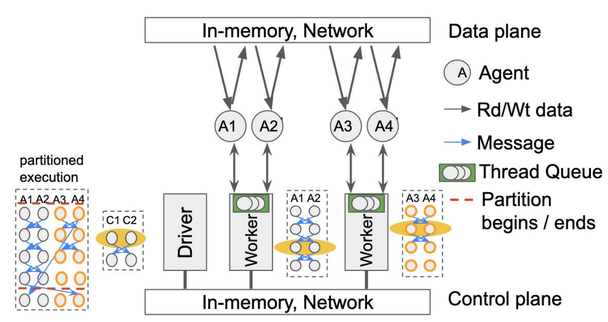
- Formal models that define agents and their interactions
  - Programming model
    - Agents are sequential processes that communicate through messaging
    - A simulation is, conceptually, concurrent execution of interacting agents
    - "Simulate" as an operator for integrating with data science pipeline
  - Computational model
    - Weighted hierarchical BSP model

- Optimizations
  - Thread merging
    - Tame high-concurrency
  - Direct memory accesses
    - Bypass messaging overhead
  - Deforestation
    - Reduce the volume of generated data

Frontend	Query	Simulate Op.		DSL
Optimizations	Deforestation	Thread Merging		Direct Memory Access
Core	Round Controller	Query Engine	Message Controller	Partition Adapter
Backend	Akka			

- Implementation
  - An eDSL in Scala for parallel agent programming
  - A system architecture based on the BSP-like model



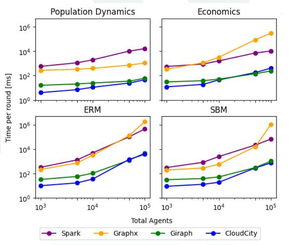


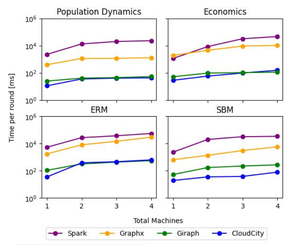
Benchmark Description

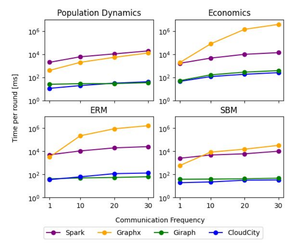
- Population Dynamics
  - Simulate the game of life example in a 2D grid
  - Model each cell in the grid as an agent
- Economics
  - Simulate the bidding process in the stock market
  - Model traders and the stock market as agents
- Epidemics
  - Simulate individuals of states Susceptible, Infectious, Recovered, Hospitalized, or Deceased
  - Model the population and locations as agents
  - Use random graph models to simulate population connectivity
    - Erdos-Renyi Model (ERM)
    - Stochastic Block Model (SBM)

Performance

- Our system has on par or better performance than current BSP-like systems







- Effective optimizations