The Matrix: An Agent-Based Modeling Framework for Data Intensive Simulations

P. Bhattacharya[1], S. Ekanayake[3], C. J. Kuhlman[1], C. Lebiere[2], D. Morrison[2], S. Swarup[1], M. L. Wilson[1], and M. G. Orr[1]

[1]Network Systems Science and Advanced Computing Division
Biocomplexity Institute and Initiative, University of Virginia

[2]Human-Computer Interaction Institute
School of Computer Science, Carnegie Mellon University

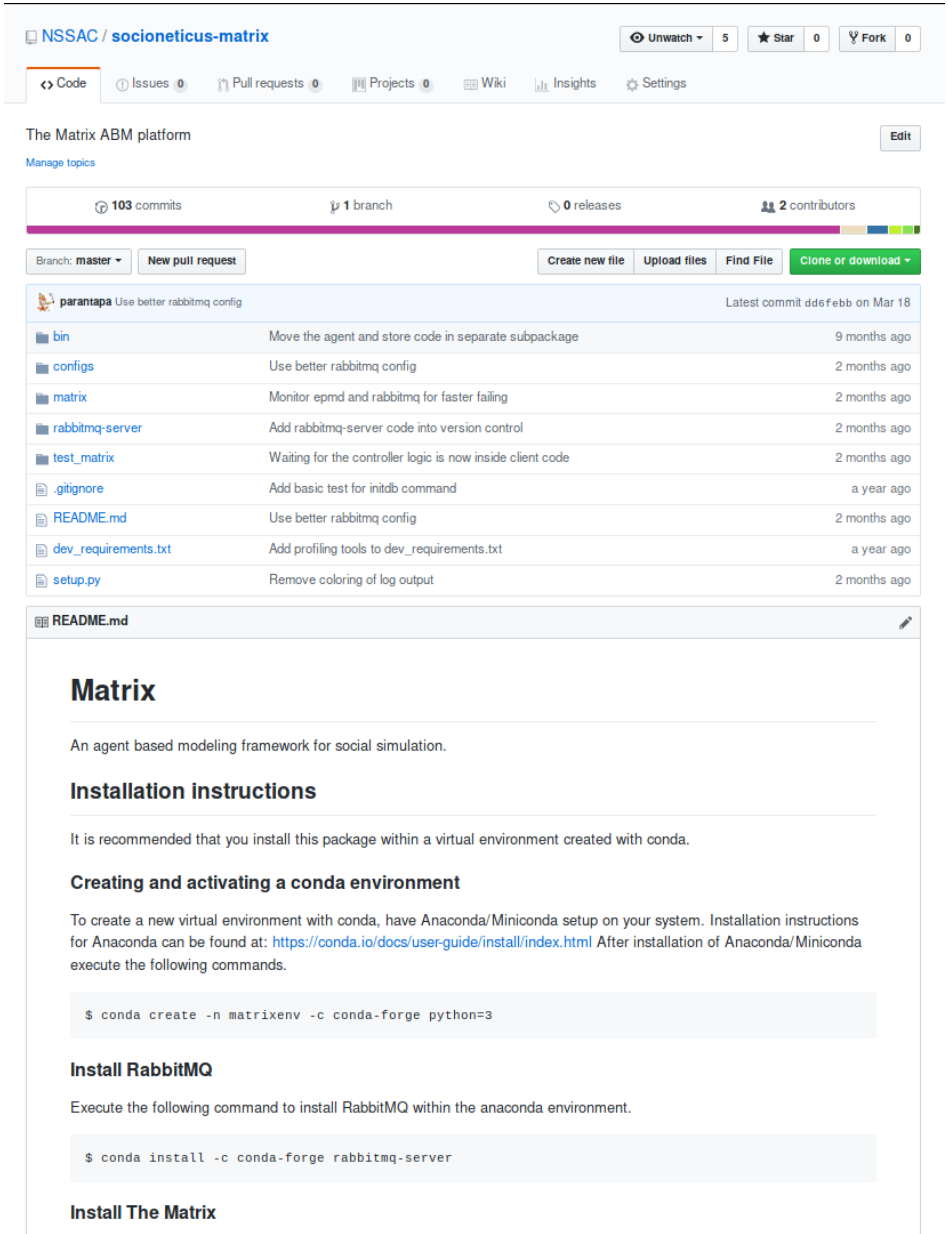
[3]Lawrence Berkeley National Laboratory

AAMAS 2019

What is The Matrix?

What is The Matrix?

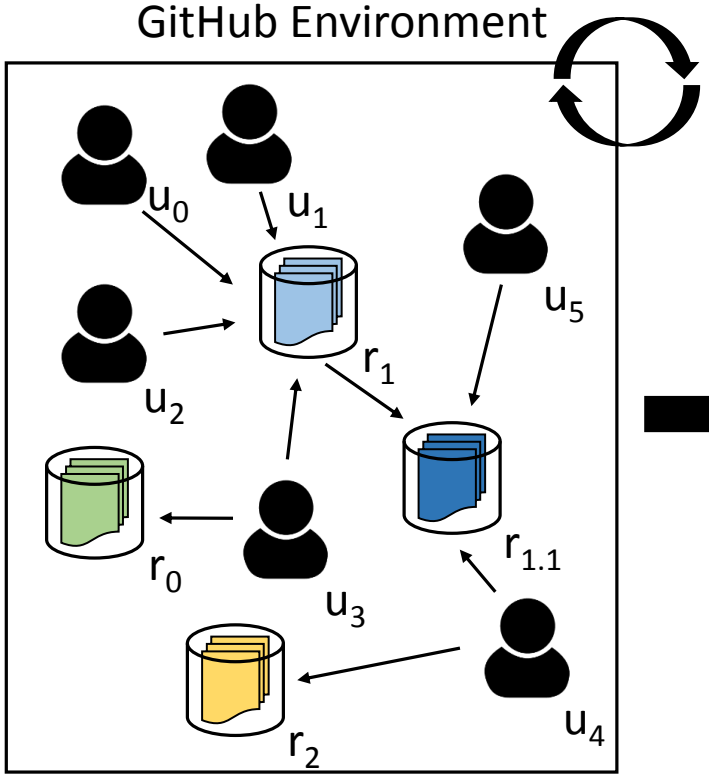
- The Matrix is an agent based modeling (ABM) framework
- The Matrix is free and open source software
- Specialized for 'compute and data intensive' simulations
  - Such as large number individual cognitive agents



`github.com/NSSAC/socioneticus-matrix`

Simulating GitHub

GitHub Environment



Event Trace

User	Repo	Type	Time
$user_3$	$repo_0$	PushEvent	2018-02-01T00:00:00Z
$user_1$	$repo_1$	CreateEvent	2018-02-01T00:01:22Z
$user_2$	$repo_1$	IssueEvent	2018-02-01T00:03:08Z
$user_1$	$repo_1$	DeleteEvent	2018-02-01T00:10:45Z
⋮	⋮	⋮	⋮
$user_i$	$repo_j$	IssueEvent	2018-02-28T11:57:39Z
$user_k$	$repo_l$	PushEvent	2018-02-03T11:59:50Z

Simulating GitHub

Continued …

- Event trace: Chronologically ordered sequence of events
- Event: A four tuple (user, repo, event type, time)
- GitHub simulation problem:
  - Input: training event trace (ground truth)
  - Output: simulated event trace
- Evaluation: compare simulated trace with held out trace (ground truth)

Simulating GitHub

Continued …

- Input event trace
  - 9.56 million users
  - 44.61 million repositories
  - 32 months of data
  - 797 million events (user-repository interactions)
  - Collected by Leidos[1]
- Output event trace
  - 1 month of interactions

[1] https://www.leidos.com/

Accommodating diverse modeling requirements

Name	Model Type	Prog. Lang.
Freq-Stat	Frequentist statistical model	Python
Soc-Th	Social structure theory model	Python
CM-ANN	Artificial neural network model	C++
CM-Bayes	Bayesian cognitive theory model	R
CM-ACTR	ACT-R cognitive theory model	Common Lisp

ABM Framework Wishlist

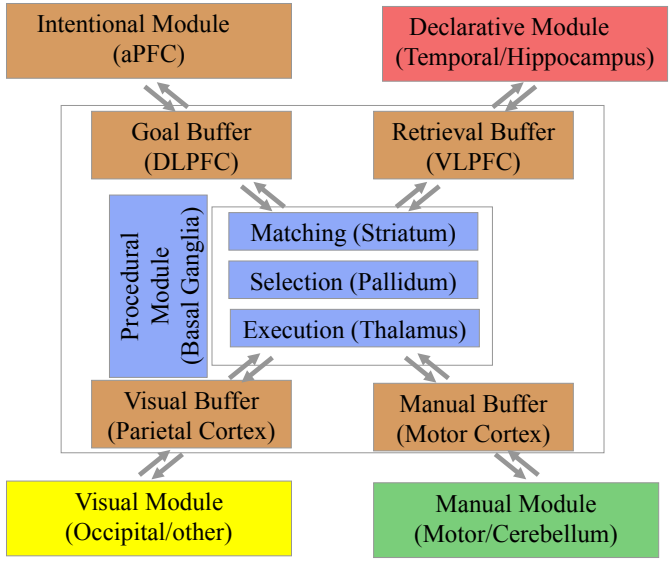
- Ability to rapidly prototype and test heavyweight agent models
- Ability to write agents in popular programming languages
  - Python, R, C++, Java, Lisp, ...
- Ability to use GPU units, and popular neural network libraries
  - TensorFlow, PyTorch, Keras, Lens, ...
- Ability to use cognitive system libraries like ACT-R
- Ability to run simulations on commodity clusters
  - Clusters without RDMA backed networks
  - Popular cloud computing platforms:
    Amazon EC2, Google Compute Cloud, and Microsoft Azure
- Ability to efficiently store, update, and query large amounts of system state
  - large amounts $\approx$ hundreds of gigabytes
- Ability to use run simulations with millions of active agents

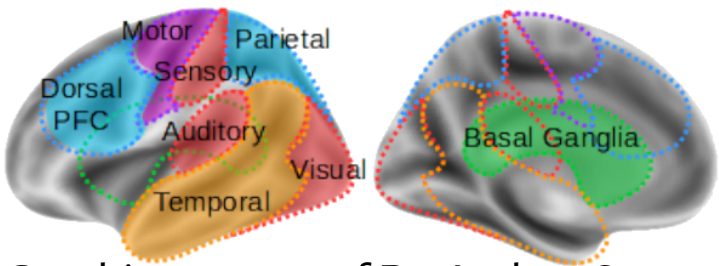
Existing ABM Frameworks

- Repast and Repast HPC (Collier and North [2013])
- FLAME and FLAME GPU (Coakley et al. [2012], Kiran et al. [2010])
- MIRAGE (Park et al. [2017])
- Swarm (Minar et al. [1996])
- Mason (Luke et al. [2005])
- AnyLogic (Huang et al. [2016])
- NetLogo (Collier and North [2013], Kiran et al. [2010])

CM-ACTR for Simulating GitHub

- ACT-R is a high-fidelity cognitive architecture (Anderson et al. [2004])
- Successfully used to develop hundreds of human behavior models from simple cognitive psychology experiments.
- Various modules of ACT-R model different aspects of cognition
- ACT-R library (written Common Lisp) provides APIs to write cognitive models





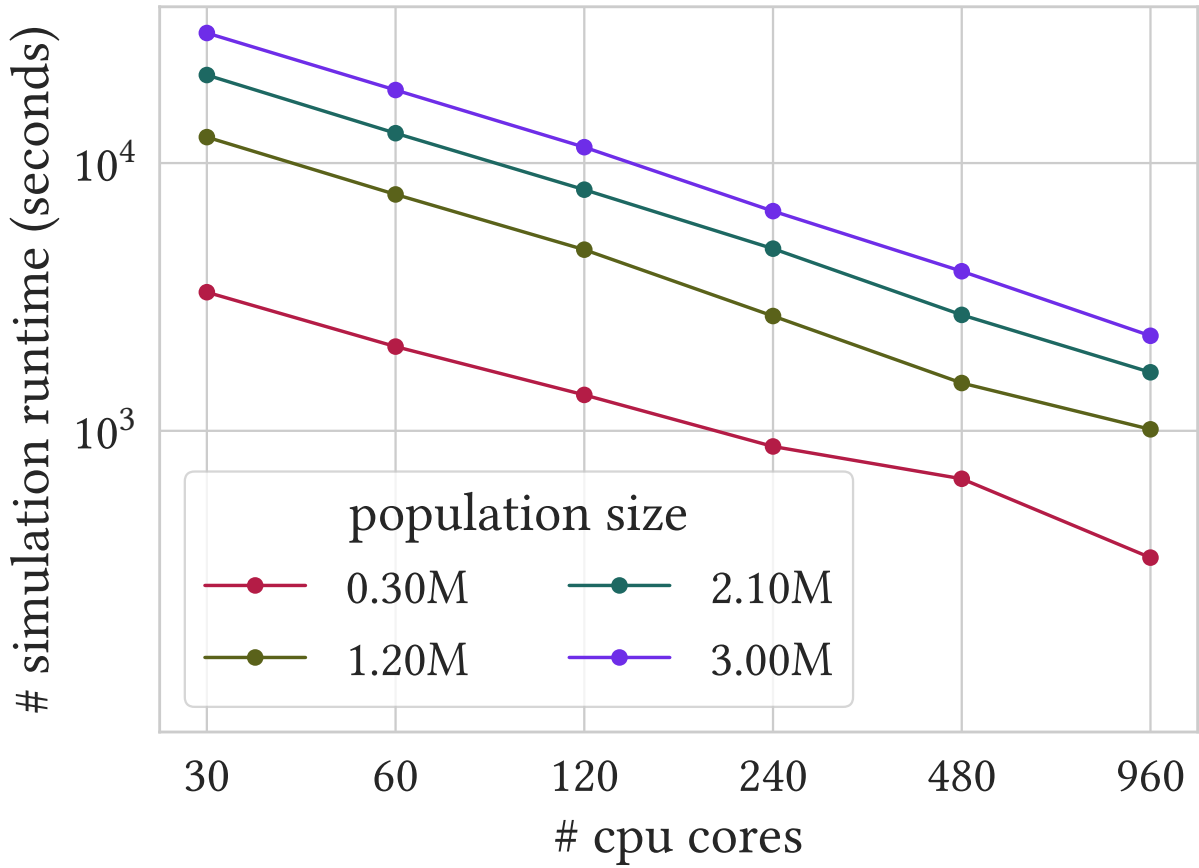
Graphic courtesy of Dr. Andrea Stocco

CM-ACTR for Simulating GitHub

Continued …

- CM-ACTR used the ACT-R library and written in Common Lisp
- CM-ACTR used only declarative memory and procedural modules
- Agents store previously seen events in memory
- New event computed one element of the tuple at a time
- Chosen components used as retrieval context for next components

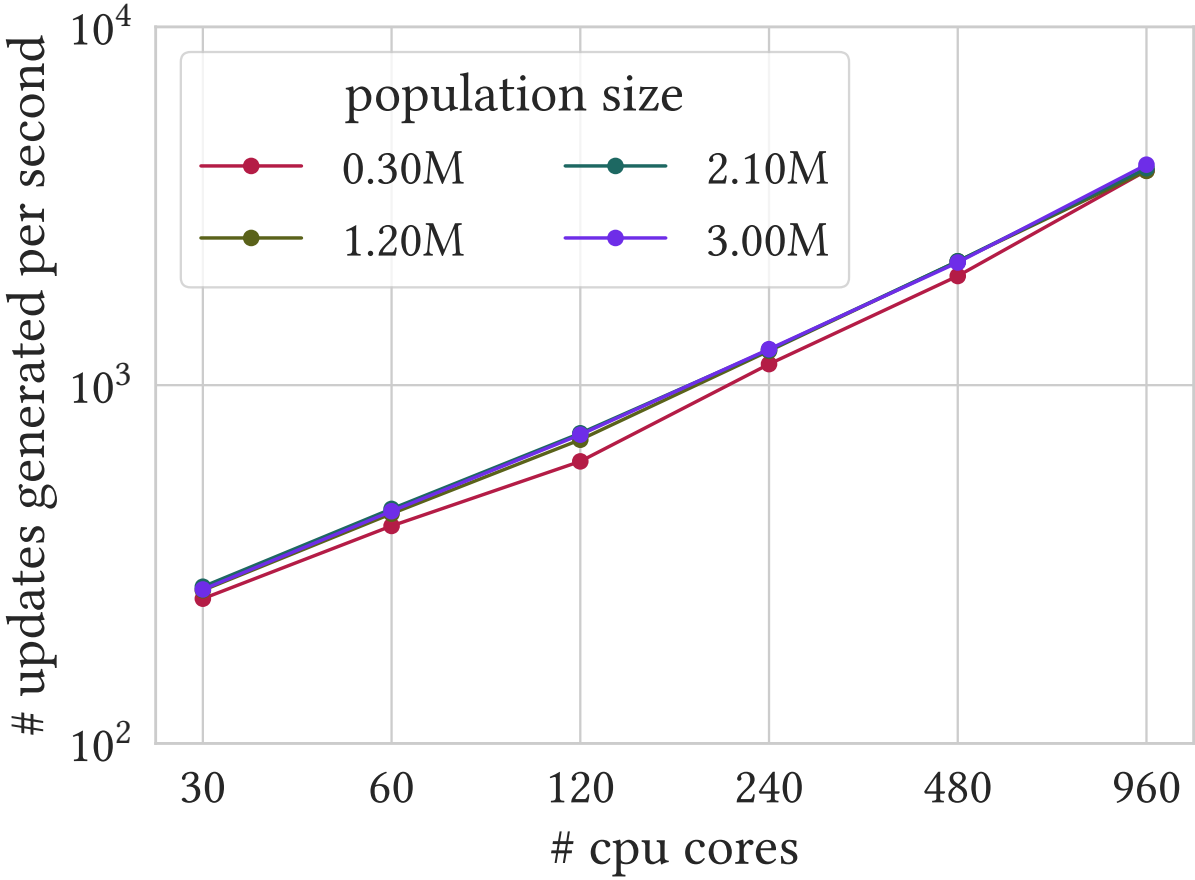
Scaling up CM-ACTR



Reduction in simulation runtime of CM-ACTR simulation
with increasing number of cpu cores,
for different population sizes of GitHub agents.

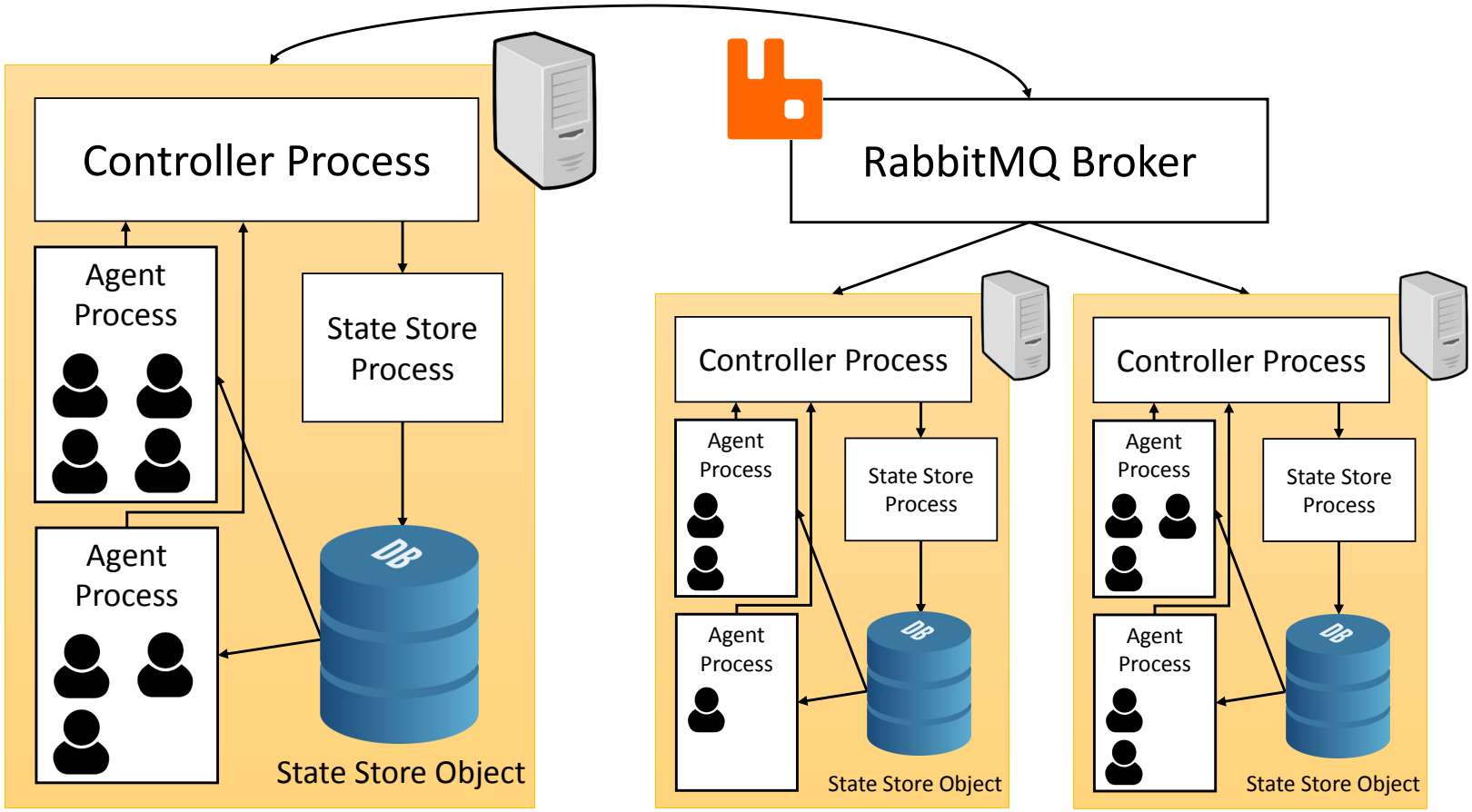
Scaling up CM-ACTR

Continued …



Number of updates generated per second by CM-ACTR simulation with increasing number of cpu cores.

Overview of Matrix ABM Runtime



- Matrix exposes API interface using JSON RPC over TCP/IP
- Subsumes complexities of distributed computation and synchronization

Conclusion

- The Matrix ABM platform
  - Facilitates rapid prototyping of ‘compute and data intensive’ agent models
  - Allows flexibility in use of programming languages and libraries
  - Targets commodity clusters and popular cloud computing platfroms

References

John R Anderson, Daniel Bothell, Michael D Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. An integrated theory of the mind. *Psychological review*, 111(4):1036, 2004.

Simon Coakley, Marian Gheorghe, Mike Holcombe, Shawn Chin, David Worth, and Chris Greenough. Exploitation of High Performance Computing in the FLAME Agent-Based Simulation Framework. In *IEEE 14th International Conference on High Performance Computing and Communications*, pages 538–545, 2012.

N. Collier and M. North. Parallel agent-based simulation with Repast for High Performance Computing. *SIMULATION*, 89(10):1215–1235, 2013. doi: 10.1177/0037549712462620.

Jingsi Huang, Lingyan Liu, and Leyuan Shi. Auction Policy Analysis: An Agent-Based Simulation Optimization Model of Grain Market. In *Winter Simulation Conference (WSC)*, pages 3417–3428, 2016.

M. Kiran, P. Richmond, M. Holcombe, L. S. Chin, D. Worth, and C. Greenough. FLAME: Simulating Large Populations of Agents on Parallel Hardware Architectures. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1633–1636, 2010.

Sean Luke, Claudio Cioffi-Revilla, Liviu Panait, Keith Sullivan, and Gabriel Balan. MASON: A Multi-Agent Simulation Environment. *Simulation: Transactions of the society for Modeling and Simulation International*, 82: 517–527, 2005.

N. Minar, R. Burkhart, C. Langton, and M. Askenazi. The Swarm simulation system: A toolkit for building multi-agent simulations. Technical Report 96-06-042, Santa Fe Institute, 1996.

B. H. Park, M. R. Allen, D. White, E. Weber, J. T. Murphy, M. J. North, and P. Sydelko. MIRAGE: A Framework for Data-Driven Collaborative High-Resolution Simulation. In D. A. Griffith, Y. Chun, and D. J. Dean, editors, *Advances in Geocomputation*, pages 395–403. Springer International Publishing, 2017.

Thank you.

`github.com/NSSAC/socioneticus-matrix`

The research is based upon work supported by the Defense Advanced Research Projects Agency (DARPA), via the Air Force Research Laboratory (AFRL). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA, the AFRL or the U.S. Government.