Practical Applications of Graph Data for Agent-based Models

Clarence Dillon[1] 0000-0002-4739-0558

George Mason University, Fairfax, VA 22030, USA (cdillon2@gmu.edu) https://cos.gmu.edu/cds/

Abstract

This is a use case for a Neo4j graph database[5] supporting an agent-based simulation of the social processes that generate the world's political order. The schema-free and object-oriented nature of graph databases make them useful companions to agent-based modelling and simulation. Integrating multiple data sets into a simulation can be problematic. To initialize and validate the model, I imported historical data about world order into a graph database. The resulting database integrates spatial, temporal, and event-based facts. The simulation output is also stored to the database so that simulated and empirical data can be directly compared.

Introduction

Agent-based models (ABM) simulate change in a system over time. Validation of ABM typically focuses on patterns of change while a simulation runs, comparing empirical evidence to the resulting state of a simulation, or both. Graph database schema are inherently object-oriented, which parallels the programming model in ABM. Practical experiences creating and using this graph database in conjunction with ABM may be useful, not only to the academic community studying international relations and peace science but, also to those searching for ways to manage empirical and simulation data in their research. Scripts are available on GitHub for integrating published datasets[2] and creating hex grids for historical political boundaries[3]. The resulting *World Order* database includes:

1. Most of the Correlates of War datasets[6] and others, like Systemic Peace's Polity IV dataset[1]
2. Spatial tiles derived from Uber's H3 hierarchical, discrete spatial grid[4]
3. A temporal mapping of dates and periods to a calendar tree: years, months, weeks, days

Problem Statement

ABM require data to be denormalized—disassembled and attributed to programmatic objects—and then organized chronologically into discrete time intervals (calibrated to simulation steps). The bulk of peace science datasets are organized first around nation states (individually or by dyad) and time (by year of record or event date). Examples include: the volume of trade between two states during a given year or the beginning and end of a war between two states. The data must be reshaped to be useful in the agent-based paradigm, particularly when agents are not equivalent to the normalized data records. Collecting and integrating multiple datasets can be tedious because the schema typically do not align and sometimes overlap. The provenance of each (possibly competing) fact needs to be recorded so that simulation results can be controlled and experiments can be validated.

Metadata and Provenance

It is necessary to maintain source metadata for each data element while attributing it to the relevant entities. This is accomplished by creating nodes representing sources of the data and the datasets they provide: the provenance. As depicted in figures 1 and 2, some relationship are divided and the *Fact* node is inserted with links to its metadata.

Attribution and Facts

$$(System) \xleftarrow{MEMBER\_OF} (State)$$

Figure 1: Relationship without attribution.

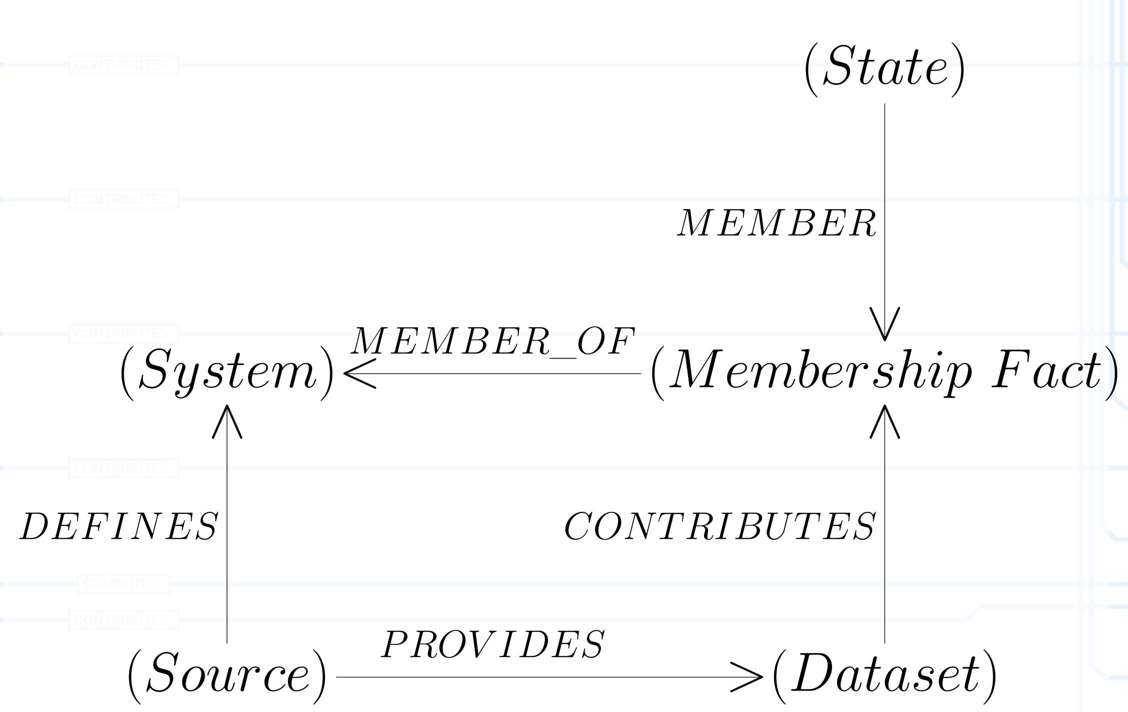

Figure 2: State system membership with Fact and metadata

Practical Example

A practical query for the world order simulation: one State will consider which territories it can attack within two border crossings, without attacking any allies. The attack will be focused on the most populous grid within the target territory. The query includes data from four different sources. Results are returned within 50ms.

Practical Result

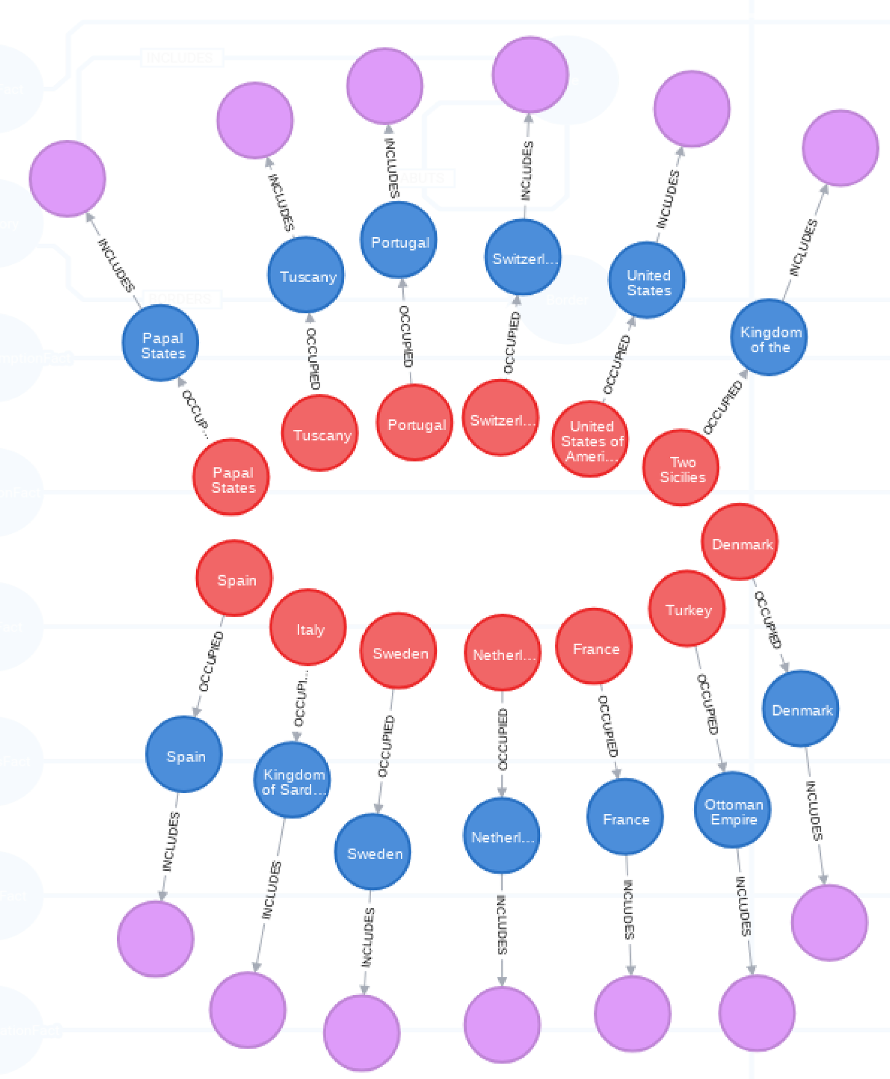

Figure 3: Most populous grid in each target territory.

Conclusion

Denormalizing structured datasets into a graph schema adapts data to the object-orientation and relatedness that characterize ABM, without sacrificing analytic requirements of scientific analysis. The schema-free nature of graph data is a natural environment to integrate separately-published datasets into a single repository that can be adapted to a variety of uses and querying. A graph database has made it possible to integrate data from algorithmic, computed and static sources. Maintaining accurate metadata and the provenance of each fact within the database has proven to be a critical component of the database and is required to support validation.

References

[1] Center for Systemic Peace. *Polity IV*. URL: http://www.systemicpeace.org/polityproject.html.

[2] Clarence W. Dillon. *Historical Basemaps*. URL: https://github.com/usuallycwdillon/historical-basemaps.

[3] Clarence W. Dillon. *World Order Data Integration*. URL: https://github.com/usuallycwdillon/vvna.

[4] David Ellis, Nick Rabinowitz, and Isaac Brodsky. *H3: Hexagonal Hierarchical Geospatial Indexing System*. URL: https://uber.github.io/h3/.

[5] Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases*. First. Sebastopol: O'reilly Media, Inc, 2013, p. 223. ISBN: 9781449356262. DOI: http://dx.doi.org/10.1016/B978-0-12-407192-6.00003-0.

[6] Correlates of War. *Correlates of War Data Sets*. URL: http://correlatesofwar.org/data-sets.

Acknowledgments