

EPPS 6354 Information Management Project Proposal

Alden Felix

February 23, 2023

Database Purpose

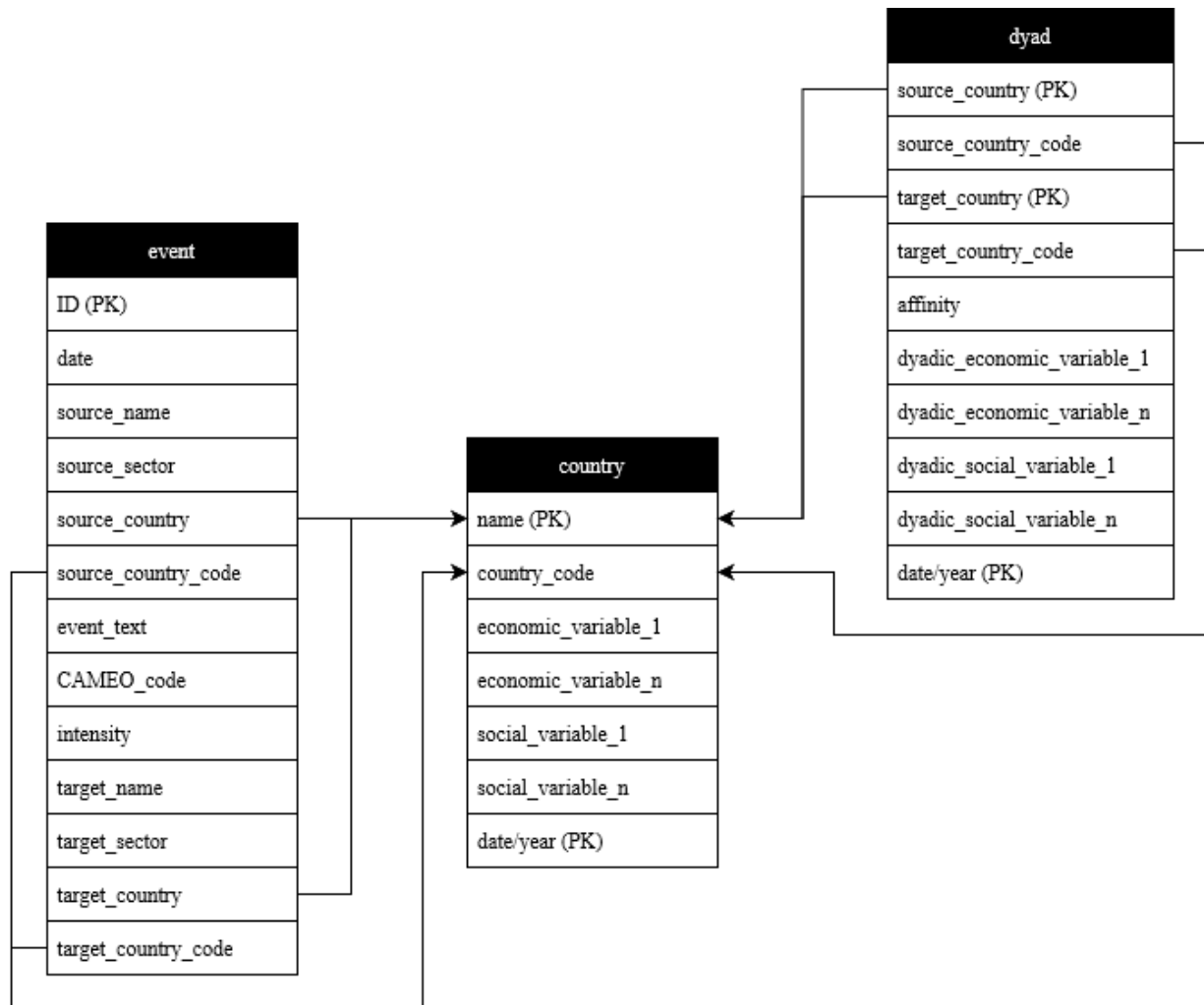
The purpose of this database project is to help fill in a gap in the existing literature surrounding affinity communities. While conflict has been studied extensively, and predicting conflict through machine learning and big data is attracting attention, the same is not true for the study of network effects on global relationships and investigating the origin of these networks. This database will directly aid a machine learning project I am involved in that seeks to predict interstate affinity using multiple social and economic variables. The data gathering process for the project revealed that the data we are seeking is currently scattered throughout multiple unrelated databases and contains many variables that are not of interest. Collecting, merging, and tidying the data into a form that can be used in a machine learning model to predict affinity will be made much easier with a relevant database.

Schema

The most significant element of the database for affinity studies will be the dyad relation. There will be two tuples for every dyad to be able to build a model with any state in the database serving as the home state, which all other states will compare to. For example, one tuple with state A as the source and state B as the target, and another tuple with the roles reversed. This is also necessary to account for differences in statistical measurement by state, which causes state A's imports from state B to not equal state B's exports to state A. Below is a relational schema

for the database with placeholders for the social and economic variables as they are yet to be formally defined.

Database Relation Schema



This is an initial stage of the relation schema with the most important elements laid out to be developed further as the machine learning study progresses. Particularly, the measure of affinity between states, the dependent variable of the study, has multiple ways of being computed which would modify the schema.