

# MC1

07 April 2020 22:08

- 123,892,863 records
- Times are relative to Jan. 1, 2025 (Unix Timestamp)
- 6 different channels of data: 1 (Phone), 2 (Email), 3 (Procurement), 4 (Co-authorship), 5 (Demographics), 6 (Travel)
- 5 node types
- 7 edge types.
- Edges always go from node type 1 to some other node type
- Type 1 nodes are the only nodes with a spatial location assigned.
- Node types are as follows:
  1. Person (used in all channels)
  2. Product category (for the procurement channel, eType = 3)
  3. Document (from the co-authorship channel, eType = 4)
  4. Financial category (from financial demographics channel, eType = 5)
  5. Country (from the travel channel, eType = 6)
- Edge types are as follows:
  0. Email
  1. Phone
  2. Sell (procurement)
  3. Buy (procurement)
  4. Author-of
  5. Financial (income or expenditure, depending on direction)
  6. Travels-to
- Communications channels (eType 0 and 1):
  - represent phone and email transactions
  - Some records have location information
  - The weight for communications is always 1, representing 1 call or email.
- Procurement channels (eType 2 and 3):
  - represent sale (eType 2) and purchase (eType 3)
  - Item always listed in the Target column.
  - Whenever there is a sale of an item in channel 2 there will be a corresponding entry in channel 3 at the same time.
  - The two people can be linked via the item they are both connected to.
  - There may also be communications between the two people.
  - The weight for procurements represents the value of the item.
  - Procurements do not have location information.
- Co-authorship channel (eType 4):
  - The co-authorship channel represents publication of scientific or technical articles.
  - The source column contains the person who is the author, and the target column contains a unique identifier for the publication.
  - Multiple authors may be connected to the same publication.
  - The time for publications occurs before all other records in the data so the values are negative.
  - The weight column indicates the fraction of the authors for the given publication.
  - Authorship does not have location information.
- Demographics channel (eType 5):
  - The demographics channel is a graph representation of the spending characteristics of each person in up to 30 categories.
  - This channel may also be thought of as attributes of each person.
  - The person is listed in the source column when money is spent in a category and listed in the target column when money is received in that category (such as income or gifts).
  - The categories are listed in the file DemographicCategories.csv.

- A person may not be connected to each category, such as the case where a homeowner does not have rent expenses.
- The time for all records in this channel is 31536000, which is the 365th day, and comes after all other record types.
- The weight channel shows how much is spent (or received) in a given category.
- This channel creates many edges that do not represent person-to-person connections in the same way as the other channels.
- Demographic records do not have location information.
- Travel channel (eType 6):
  - The travel channel connects people (always listed in the source column) with locations (listed in the target column).
  - Time represents the start of a trip and weight represents the length of the trip in days.
  - All location columns should have data for each record.
  - The SourceLocation and TargetLocation columns have identifiers for countries of the origin and destination of each trip.
  - More specific latitude and longitude values are also provided.