Extract Document Co-Citation Network (Core Only)

Menu path
Data Preparation > Database > ISI > Extract Document Co-Citation Network (Core Only)

Description
Extracts the document co-citation network from an ISI database.

Each document in the input database is represented by a node. An edge is drawn between the nodes for two documents if and only if they were cited by a common document in your dataset.

Core Document vs. Non-Core Document
There is a distinction drawn between documents contained in your dataset and documents in general. A document in your dataset is called a "core document". Your documents may (and probably do) reference non-core documents.

The output network of this algorithm will contain nodes representing only core documents. For an algorithm that will represent non-core documents too, see Extract Document Co-Citation Network (Core and References).

Analyses
The output network will include the following summaries of your dataset:

- **Node (Document)**
  - A prettified label for identifying this document.
  - A prettified string giving the journal (or other source) of this document (called 'source').

- **Edge (Co-Citation)**
  - The number of documents in your dataset which cited both of these documents.
  - This figure is also given with three common normalizations: Jaccard, cosine, and Dice.
  - Publication year of the earliest co-citing document.
  - Publication year of the most recent co-citing document.

Usage Hints
Load an ISI file into the tool, then create a database from it using the ISI database loader.

It is strongly recommended that the database be cleaned before extracting any co-citation networks from it.

For a quick analysis of a small dataset you may wish to merge together author entities with identical names. For a scientifically sound analysis of a larger dataset, you can find author entity merging suggestions (or manually set your own merging orders from scratch) and perform the merge.

Then, you will probably want to merge together journal entities according to recognized variants.

Finally, you must match references up to documents in your dataset (there are no citations to analyze, otherwise).

Implementation Details
The specific queries run by the tool can be found in the source code.