

Extract Weighted Document-Document Network

Description

The algorithm creates a weighted, undirected document-to-document network based on the documents similarities from an ISI database.

Menu Path

Extract -> From ISI Database -> Extract Weighted Document-Document Network

Input Parameters

Threshold

Any document-document edge whose calculated similarity is below this value will not be given an edge in the network output by this algorithm.

Comparison Algorithm

The comparison algorithm is used to calculate the similarity between documents. You may select between the following algorithms:

Name	Link to More Information	Theoretical Range of Values	Actual Range of Values for this Algorithm
JaccardCoefficient	Wikipedia	0 to 1	0 to 1
CosineSimilarity	Wikipedia	-1 to 1	0 to 1
SørensenSimilarityIndex	Wikipedia	0 to 1	0 to 1

Output

A network file with all documents connected by a weighted edge as calculated by the *comparison algorithm* if the weight of the was above the *threshold* value.

Usage Hints

The [Cosine Similarity](#) *comparison algorithm* does not preform any real comparison between the terms. It only uses the existence or nonexistence of a term in the set for the calculation.

Implementation Details

The [Cosine Similarity](#) *comparison algorithm* does not preform any real comparison between the terms. It only uses the existence or nonexistence of a term in the set for the calculation.

The network file has the following node properties

Label	Type	Source
label	<i>String</i>	A concatenation of <i>PUBLICATION_YEAR</i> , <i>VOLUME</i> , <i>DIGITAL_OBJECT_IDENTIFIER</i> , <i>BEGINNING_PAGE</i> from the Documents Table , <i>TWENTY_NINE_CHARACTER_SOURCE_TITLE_ABBREVIATION</i> from the Sources Table , and <i>RAW_NAME</i> from the People Table .

The network file has the following edge properties:

Label	Type	Source
weight	<i>float</i>	The <i>comparison algorithm</i> chosen.
DEFAULT_SOURCE_KEY	<i>int</i>	The node given to the edge.
DEFAULT_TARGET_KEY	<i>int</i>	The node given to the edge.

The specific query run by the tool can be found in the [source code](#).

See Also



The license could not be verified: License Certificate has expired! [Generate a Free license now.](#)