Extract Co-Occurrence Network

Description

Extract Co-occurrence Network from Table allows the user to create a network based on columns in a table. The user selects a column that contains multiple values separated by a delimiter. The algorithm splits those values based on the delimiter and creates an edge between between each unique value. It can also perform calculations on other columns present in the table.

Extract Co-occurrence Network from Table takes 3 parameters.

- **Column Name -** The name of the column that contains the values one wishes to extract.
- **Text Delimiter -** The character value that separates the values in the extraction column.
- **Aggregation Function File -** An optional file that contains functions to analyze the table.This algorithm takes a table with with several delimited values in a single column. See Usage hints for a sample table and aggregation function file.

Applications

You can use this algorithm to extract a network based on values that co-occur in a column of a table and calculate some statistics based on other values in that table. This can be used to extract coauthorship networks, co-citation networks, co-PI networks, etc.

Usage Hints

A simple coauthorship network described in a .csv file might look like the following:

```
"Authors","Times Cited","Title"
"Fred Flintstone|Wilma Flintstone|Barney Rubble|Betty Rubble","20","Somebody Walk my Dinosaur: Kids and Pets in
the Stone Age"
```

The algorithm can then be run with the following parameters:

Column Name

Authors

Text Delimiter

|

If you wish to perform calculations on the different columns you can specify a set of aggregation functions. The following aggregation functions are supported:

- Arithmetic Mean: arithmeticmean
- Sum: sum
- Count: count
- Geometric Mean: geometricmean
- Max: max
- Min: min

Aggregate functions may not work on all types of data.

Define an aggregate function file in the following way:

```
<node|edge>.<newcolumnname>=<columnnameinoriginaltable>.<function>
```

Below is a sample aggregation function file for co-authorship networks.

```
node.numberOfWorks = Authors.count
edge.numberOfCoAuthoredWorks = Authors.count
node.timesCited = Times Cited.sum
```

If the column specified in the aggregation function file does not exist, then the calculation is skipped, but the rest of the network is still generated.

Missing Data

The current implementations of the aggregation functions do not deal well with missing input data (see bug SCISQUARED-528). If the algorithm does not run correctly while using an aggregation function file, try examining the input data, and removing any rows that contain missing values, or placeholders like the empty string "" or "N/A".

Links

- Source Code

See Also

⚠ The license could not be verified: License Certificate has expired! Generate a Free license now.