<Your Name>

CASOS

ORA Change List Thesaurus

Jeff Reminga

The CASOS Center
COS Program, School of Computer Science, Carnegie Mellon
Summer Institute 2020

Carnegie Mellon

ISI institute for SOFTWARE RESEARCH

Center for Computational Analysis of Social and Organizational Systems
http://www.casos.cs.cmu.edu/

Carnegie Mellon
ISI institute for SOFTWARE RESEARCH

Overview

- Purpose: clean nodesets to remove incorrect or duplicate nodes, save the changes and apply to other meta-networks
- The Nodeset Editor in ORA lets users:
  – Delete nodes
  – Merge (combine) one or more nodes
  – Move nodes
  – Rename nodes
- Operations are recorded in ORA as a Change List
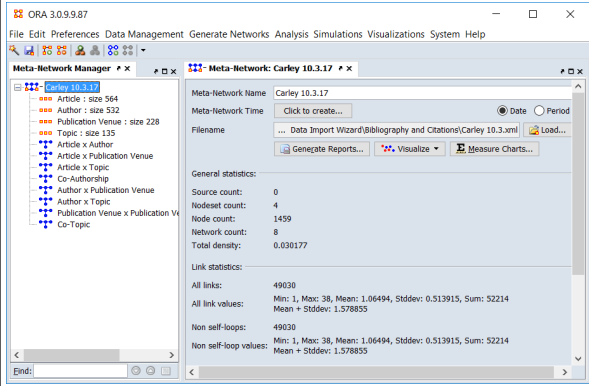- Save, Load, and Apply Change List to other meta-networks

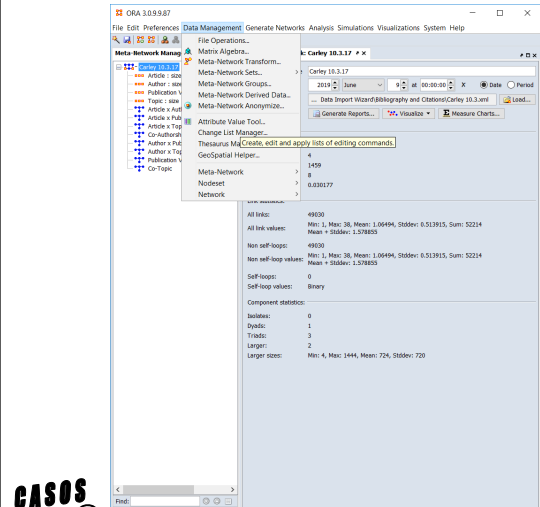### CASOS

<Your Name>

Load meta-networks

- Load the citation meta-network called: Carley 10.3.17.xml
- This is in the Data Import Wizard folder under "Bibliography and Citations"
- This contains 4 nodesets:
  - Article
  - Author
  - Publication Venue
  - Topic

June 2020   © 2020 CASOS, Director Kathleen M. Carley   3
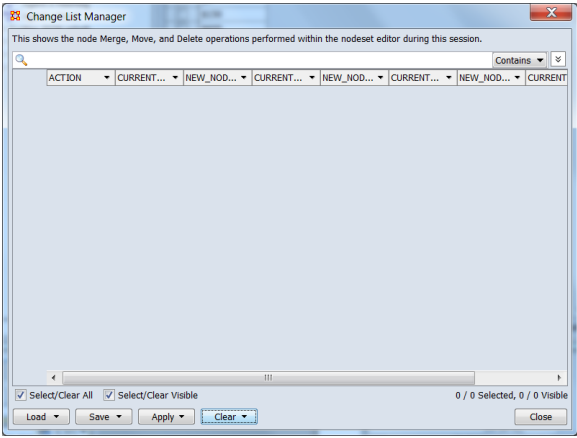

Change List Manager

- Menu item Data Management \ Change List Manager
- Tool that records the node deletes, merges, moves for each nodeset
- First clear before the session of edits

June 2020   © 2020 CASOS, Director Kathleen M. Carley   4
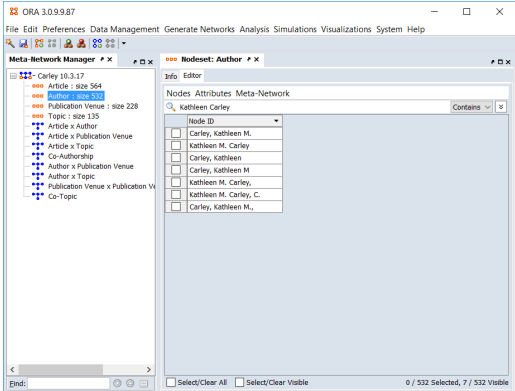
<Your Name>

Change List Manager

- This is an empty list of changes
- Before you begin your editing session of deleting, merging, and moving nodes, click the Clear button

June 2020   © 2020 CASOS, Director Kathleen M. Carley   5
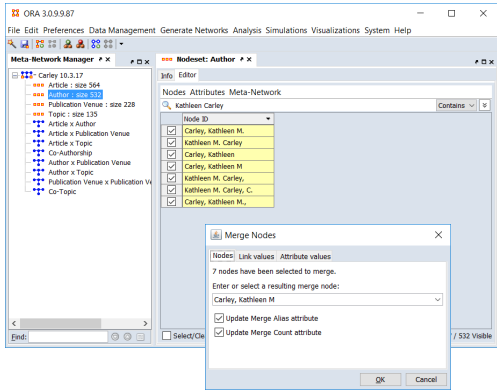

Examine Authors

- Select nodeset Author
- Type "Kathleen Carley" in the search bar
- Notice that there are many variations of her name
- We want to remove these duplicates by combining the nodes and all of their associated links together
- This combining is called merging

June 2020   © 2020 CASOS, Director Kathleen M. Carley   6
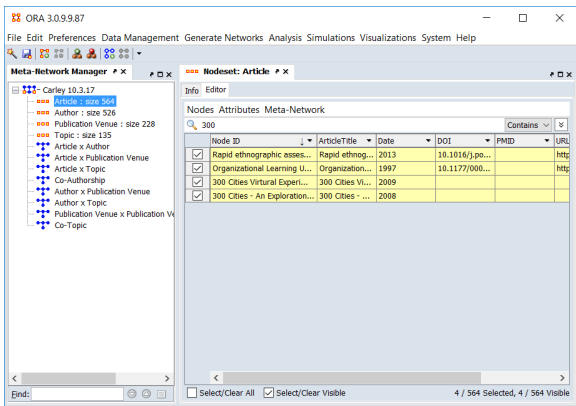
<Your Name>

Select Authors

- Select the Author we want to merge
- Click from the Nodeset Editor menu Nodes \ Merge Selected Nodes
- The Merge Nodes dialog appears (displayed)
- Select the name to combine all Kathleen nodes into
- There are options to control how link values and attributes values are merged
- Click OK to perform the merge

June 2020   © 2020 CASOS, Director Kathleen M. Carley   7

Select Articles to Delete
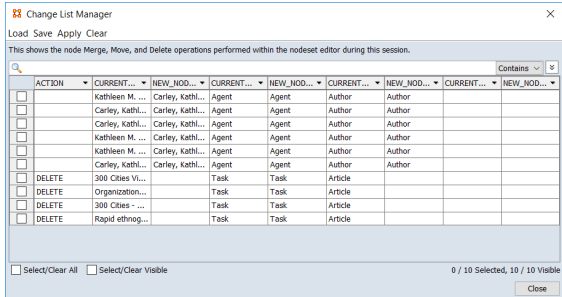
- Suppose we want to delete the Articles that have the word "300" in them
- Select the Article Nodeset
- Type 300 in the search bar
- Select the resulting nodes
- Use the Nodes \ Delete selected nodes menu to delete them

June 2020   © 2020 CASOS, Director Kathleen M. Carley   8

Examine Change List
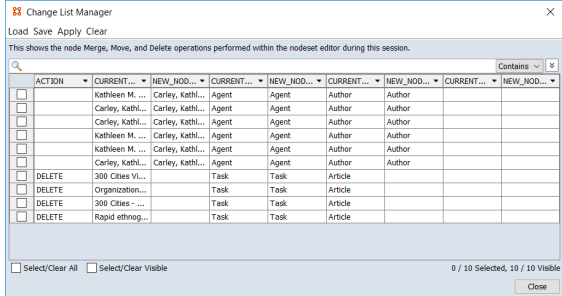
- Open the Change List Manager
- Note the Merge entries for Kathleen
- Note the Delete entries for 300

June 2020 © 2020 CASOS, Director Kathleen M. Carley 9

Apply Change List

- Now suppose we want to perform the same changes to the other meta-networks
- Use the Apply menu item at the top and apply as Change List to do merges and deletes, or Delete List to just perform deletes

June 2020 © 2020 CASOS, Director Kathleen M. Carley 10

Apply Change List

Apply Change List

Choose the change list applicator options, and then choose the meta-networks to change.

Select the apply options:

Apply all changes

☑ Auto detect reduced form

☐ Apply to meta-network in reduced form

☑ Match nodes exactly

Apply the change list to these meta-networks:

☐ **Select All**
  ☐ Carley 10.3.17
  ☐ Carley 10.3.17
  ☐ Carley-1
  ☐ Carley-2
  ☐ Carley-3
  ☐ Carley-4
  ☐ Carley-5
  ☐ Carley-6
  ☐ Carley-7

OK   Cancel

Success

Total changes across all meta-networks:

0   Node IDs changed
25 nodes merged
4   nodes deleted
0   nodes added
0   node types changed

OK

- Select the other meta-networks to apply the change list to
- Click OK
- A dialog will report the changes made to the other meta-networks

June 2020   © 2020 CASOS, Director Kathleen M. Carley   11

Save/Load Change List

Change List Manager

Load Save Apply Clear

This shows the node Merge, Move, and Delete operations performed within the nodeset editor during this session.

	ACTION	CURRENT...	NEW_NOD...	CURRENT...	NEW_NOD...	CURRENT...	NEW_NOD...	CURRENT...	NEW_NOD...
☐		Kathleen M. ...	Carley, Kathl...	Agent	Agent	Author	Author		
☐		Carley, Kathl...	Carley, Kathl...	Agent	Agent	Author	Author		
☐		Carley, Kathl...	Carley, Kathl...	Agent	Agent	Author	Author		
☐		Kathleen M. ...	Carley, Kathl...	Agent	Agent	Author	Author		
☐		Kathleen M. ...	Carley, Kathl...	Agent	Agent	Author	Author		
☐		Carley, Kathl...	Carley, Kathl...	Agent	Agent	Author	Author		
☐	DELETE	300 Cities Vi...		Task	Task	Article			
☐	DELETE	Organization...		Task	Task	Article			
☐	DELETE	300 Cities - ...		Task	Task	Article			
☐	DELETE	Rapid ethnog...		Task	Task	Article			

☐ Select/Clear All   ☐ Select/Clear Visible   0 / 10 Selected, 10 / 10 Visible

Close

- Use the Load and Save menus to save the changes to a file
- The file format looks like what is shown on the screen.
- Note: a thesaurus used in Automap can be loaded and applied

June 2020   © 2020 CASOS, Director Kathleen M. Carley   12

<Your Name>

Automating Nodeset Cleaning

- Navigate to the Article nodeset
- Clean by merging duplicate nodes
- Click Nodes \ Clean Nodeset \ Find Duplicates

June 2020   © 2020 CASOS, Director Kathleen M. Carley   13

Automating Nodeset Cleaning

- Use the dialog to search for nodes that differ by one or more character changes
- Useful to find spelling mistakes and words that differ only in tense
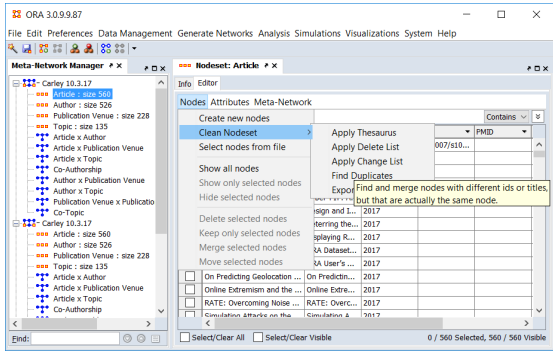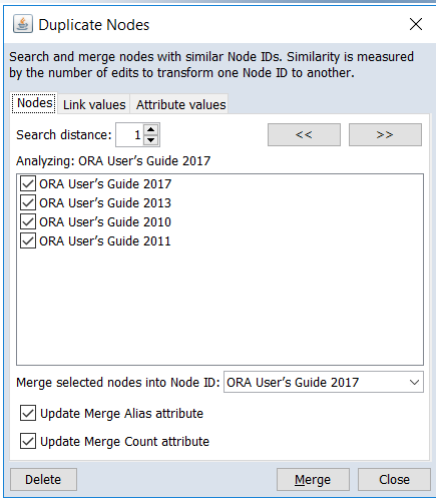- Merges made in this dialog are recorded in the Change List Manager

June 2020   © 2020 CASOS, Director Kathleen M. Carley   14

<Your Name>

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

After Cleaning

- Cleaning data involves merging and deleting nodes
- The Data Import Wizard for the following data formats creates derived data during import:
  - Twitter
  - Bibliography & Citation
  - YouTube
- Derived data is computed on the data directly imported
- The derived data needs to be recomputed after you are done cleaning

CASOS

Carnegie Mellon
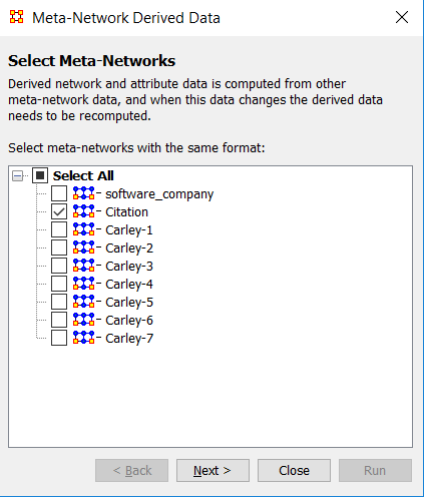isr institute for SOFTWARE RESEARCH

Derive new Data

- ORA contains a tool to recompute data that was imported from the Data Import Wizard
- Open menu item:
  Data Management \ Meta-Network Derived Data

CASOS

CASOS

<Your Name>

Derive new Data…

Meta-Network Derived Data   ✕

Select Meta-Networks

Derived network and attribute data is computed from other meta-network data, and when this data changes the derived data needs to be recomputed.

Select meta-networks with the same format:

- ☐ **Select All**
  - ☐ software_company
  - ☑ Citation
  - ☐ Carley-1
  - ☐ Carley-2
  - ☐ Carley-3
  - ☐ Carley-4
  - ☐ Carley-5
  - ☐ Carley-6
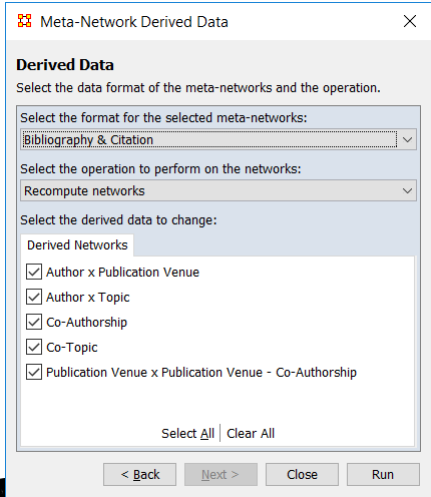 - ☐ Carley-7

< Back   Next >   Close   Run

- You should see this dialog
- The Citation meta-network is selected here
- This data was imported with the Data Import Wizard using the Bibliography & Citation format

Derive new Data…

Meta-Network Derived Data   ✕

Derived Data

Select the data format of the meta-networks and the operation.

Select the format for the selected meta-networks:

Bibliography & Citation ▾

Select the operation to perform on the networks:

Recompute networks ▾

Select the derived data to change:

Derived Networks

- ☑ Author x Publication Venue
- ☑ Author x Topic
- ☑ Co-Authorship
- ☑ Co-Topic
- ☑ Publication Venue x Publication Venue - Co-Authorship

Select All | Clear All

< Back   Next >   Close   Run

- Clicking Next shows the following dialog
- Select the Bibliography & Citation data format
- Select Recompute networks
- Click Run
- The meta-network will have its derived networks recomputed.