Using ORA for Social Media

By Kathleen M. Carley
Professor of Computation, Organizations and Society CMU
CEO Netanomics Inc. aka Carley technologies Inc.

Overview

ORA-NetScenes aka ORA is a network analytic tool, designed to help the analyst rapidly analyze, visualize, and engage in what-if reasoning about both social networks and high-dimensional and even geo-temporal networks. In this exercise you will launch ORA, import social media data, and then run the twitter report. You can then, if you choose, visualize the data or conduct secondary analyses. If you have multiple time periods of data you can load each separately and do trend analysis.

Launch ORA

To launch ORA you simply click on the ORA icon. Select use trial version. Then ORA will open in a window.

Import Data

In the toolbar, on the upper left, click on the wand – data import wizard. This launches a window. Select the option to import other data formats and then select twitter. On the lower right of the wizard

window select next. Then ORA will ask you to select the twitter data format. Scroll down and select SR. Under select one or more data files – select the file you want. You probably want to aggregate by 4 hours, to store the message as a tweet node, and to create the re-tweeted by network. Then select next then finish. This automatically converts the twitter JSON file that you have into a network. You will now see data in the meta-network manager and you will see information about the network in the editor panel. Each time you import data you will be asked if you want to replace or append the data to previous data imported.

By default, ORA parses hashtags used in Twitter messages as Knowledge nodes. For larger datasets, this may be sufficient; however, for smaller datasets, you may wish to parse the Twitter message using the Universal Thesaurus to identify trends in your data.

Run Twitter Report

In the toolbar, under analysis, under Key Entities select Twitter report. You should now see a window open containing results from a large number of analyses on the twitter data. You can peruse this to identify the critical tweeters and topics.

Visualize Data

In the editor window (and in the toolbar) you will see an option to visualize your network. This will visualize the network highlighted in the meta-network manager. Use the legend to display or hide different classes of nodes and edges. Note – for really large networks ORA will automatically group the nodes; for moderately large networks it will ask you want you want to not show – e.g., node names, isolates, links; and, for small networks it will just show everything.

Remove isolates

To simplify the visualization you may want to remove isolates. To do this select actions from the visualization toolbar and then select delete isolates.

Partition data by key actors

If you have been collecting ongoing data and are interested in only seeing tweets from a subset of Twitter users, e.g. only looking at verified users, you can partition your data to only look at those actor's tweets. Under "Data Management >> Attribute Manager", select Partition, and then navigate to the attribute of the agent that you want to highlight. After partitioning, it is recommended that you re-generate inferred networks (e.g. Hashtag x Hashtag network by multiplying Agent x Hashtag networks) as the partitioning will only affect networks that directly involve the nodeset (e.g. Agent x Agent retweet network).

Highlight key nodes

Some nodes are critical, such as the individual most retweeted, or the most connected to hashtag. These can be highlighted by selecting node appearance from the visualization toolbar and then select Node Color then select Color Nodes by Attribute or Measure. This will pop up a window. Under select a measure, pick a measure – e.g., centrality total degree.

Find groups

To identify groups in the visualized network you can run a grouping algorithm such as Newman groups and ORA will automatically color those nodes that are in the same group. To do this select Node Appearance from the visualization toolbar, then select Node Color, then select Color Nodes by Newman Grouping. Each group will be a different color.

Run Other Reports

Other reports can be run on this data. Under analysis, try running, e.g., the key entity report. In the toolbar go to analysis then select Generate Reports, then Locate Key Entities, then Key Entity. This will open a window. Simply respond to all questions and press Next until done.

A common clustering approach in text analysis is latent dirchlet allocation (LDA) – this can also be run on Twitter data. To run this in ORA, you will first need to take the transform of the automatically generated Tweet x Topic network to use as input for the "Topic Analysis" report in ORA, which can be done under the "Data Management >> Matrix Algebra" menu . You can set a new metanetwork as an output in this report to view correlations and overlapping topic memberships between different clusters.

General Hints

You can increase processing speed by deleting isolates.

You can increase visualization speed by showing only one network, or not showing isolates, labels, or lines.

In the visualizer, if you have an image you like save it in any format you like.

If you delete things from your data, save the new meta-network under a new name. Version control is a good thing.

Read the ORA QuickStart guide for additional information.

If you have sentiment ratings for your data, whether by category or by numerical assessment, you can import this data by selecting "Import Data >> Import Excel or text delimited files >> Table of node attributes", and matching Twitter messages to nodes using Tweet ID. It is highly recommended to NOT use Excel for this – Twitter IDs do not render in Excel, and their length can also trip up R's default importer. If using R to combine Tweet IDs and sentiment ratings, we recommend importing Tweet IDs as character strings or utilizing the R package "bit64", which will correctly render Tweet IDs.

Twitter Data

It is usually good to identify critical tweeters and hashtags. It is good to visualize the retweet network – often without labels and to visualize the hashtag network with labels.

ORA-PRO

Note: If you are using social media you will be better off with ORA-PRO as it is not node limited.