ORA & NetMapper

L. Richard Carley, Jeff Reminga and Kathleen M. Carley
Netanomics and Carnegie Mellon University
lrc@netanomics.com

Abstract. Network analytics is one of the key pillars of data science. Today, however, network analysis is more than just analyzing small networks with one or two types of nodes. Instead, big data, high dimensional networks, temporal networks and geo-spatial networks are key. New methods are needed for assessing such networks. Many of these methods are embedded in the ORA system. In addition, new methods are needed to extract networks from texts. NetMapper is designed for this purpose. These tools are interoperable and together support the rapid and accurate network analytics for raw data often in the form of comments, news articles, or blogs.

Keywords: Social Network Analysis, High Dimensional Network Analysis, Network Visualization, Network Extraction from Texts

1 Introduction

Network analytics are widely used in many fields. Increasingly though, the networks of interest are high dimensional. That is, they are networks of networks with multiple types of nodes and links, the links are weighted, directional, and temporal, and the nodes have attributes. For example, studies of organizations might look at the social network within the organization as well as the task assignment and knowledge network. Studies using Twitter frequently look at the mentions network at the same time as the hashtag network. In both cases these networks may exist at one time or many, and the nodes may be spatially set. For example, in disasters, a tweeter may list their latitude and longitude.

Network analysis techniques allow the analyst to answer questions about who is doing what to whom, who is influential, what messages are critical, whether there are hidden groups, and how these change by time and location. This is done by examining the networks connecting people, places, ideas and things.

Sometimes network data is easy to extract from meta-data or pre-exists in a data base. For example, header information in email (from-to), sender and hashtags in twitter, or authors in citation data. Other types the network is implicit and appears embedded in texts such as when police use comment fields to discuss co-conspirators, or news articles provide discussions of which leader is meeting with which, or court cases provide records of interaction. In these latter cases networks need to be extracted from these raw texts.

Today network analytics requires interoperability. That is network analytic and visualization engines need to interoperate with network extraction technologies. ORA, for example interoperates with many extraction technologies. These include, but are not limited to: NetMapper, Blog Trackers, YouTube Tracker, and TweetTracker.

2 ORA

ORA is a powerful network analysis and visualization tool (see Carley, 2017 for additional details). ORA supports the assessment of standard social network data, organizational network data, high-dimensional network data, meta-network data, geo-spatial network data, and dynamic network data. ORA supports dynamic network analysis (DNA) and geo-spatial network analysis. This extends social network analysis and link analysis to the assessment of multiple networks of different entities and links that may (or may not) be temporally or geo-spatially linked. In addition ORA offers many special network analytic techniques designed for assessing social media data.

Using well validated social science and cognitive science findings a series of techniques were created and implemented in ORA for inferring missing links, evolving networks over time, forecasting information diffusion and belief change, and inferring the command structure for specific tasks. A key new method in ORA is Socio-Cultural Cognitive Mapping – an approach fro inferring a network from node attributes and creating a network mapping in which the layout is interpretable for more than just identification of clusters. Graph theoretic, machine learning, statistical and agent-based modeling techniques are used to identify patterns and assess change. ORA can assess any networks involving who, what, where, why, how or the connections between them and how they changes over time. ORA can also assess trails connecting who was where when. ORA can convert networks to trails and trails to networks. This capability supports geo-spatial analytics.

Common networks that analysts use ORA to assess include: social networks, knowledge networks, capabilities networks, activity networks, task networks, resource networks, communication networks, alliance networks, retweet/reply/mention or like networks, and co-hashtag networks. ORA provides the capability to assess network performance and vulnerability in complex what-if scenarios, visualize networks and comparative network metrics such as agent betweenness, centrality, and closeness, and extract networks from text sources such as email. ORA can also link to data in other forms such as PenLink, AnalystNotebook, CSV files, and SQL databases.

ORA uses a combined "meta-network" representation combining multiple interlocking networks and a variable tie (link) representation and a trail representation of who was where providing temporal and geospatial context among nodes. Graph metric, standard social network metrics, pattern identification algorithms for both trail and network data, clustering algorithms, and visualization algorithms can then be run to assess social networks, semantic networks, knowledge networks, activity networks, etc. and to asses and identify patterns in trails. In addition, there are several integrated models for belief formation and an ABM simulation that can be run using this representation to evolve the underlying networks change in diffusion and effort are directly measured and a canonical classification task is also used to measure accuracy and agreement, and diffusion in the organization. ORA also includes basic capabilities for community detection, motif identification, clustering, and topic identification.

ORA supports many forms of visualization including 2D and 3D visual network representation, block models, heat maps, geo-network visualization 2D and 3D, wordl for nodes – nodel, and basic graphing capabilities. Some of these forms of visualization are shown in Figure 1. ORA also has many built in features for data management and cleaning. ORA has been parallelized and includes multiple features to support big-data analysis. ORA also includes extensive on-line and off-line help – see Neal et al, 2017.

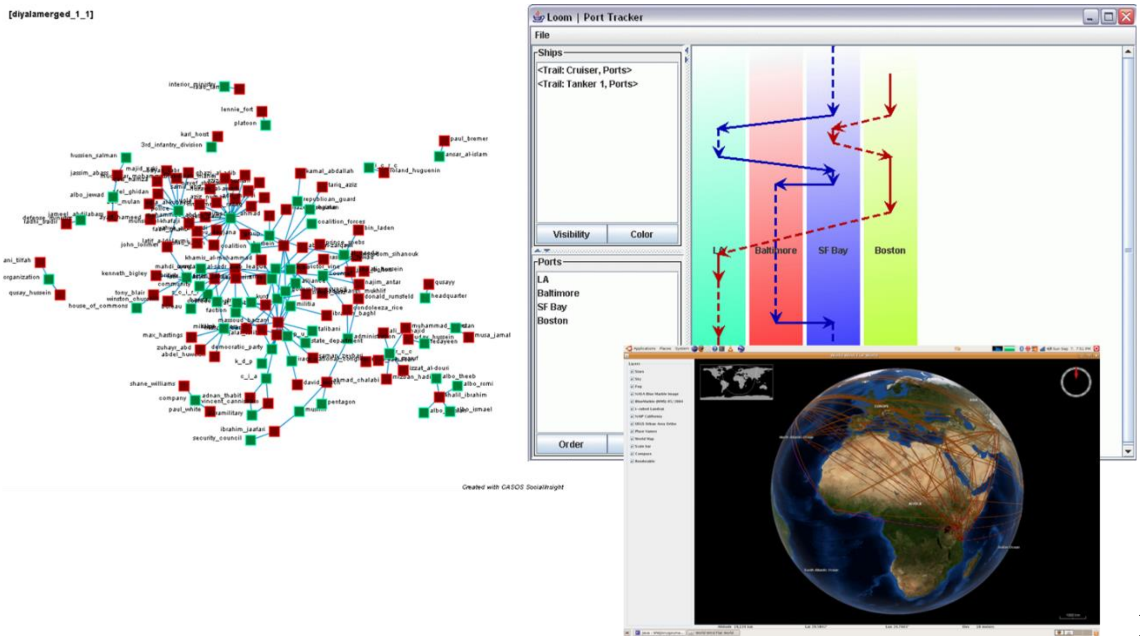

Figure 1. Illustrative Visualizations in ORA

3 NetMapper

NetMapper is a tool that supports extracting networks from texts and assigning sentiment at the context level. Given a text, NetMapper extracts both concepts and links among them and assesses the sentiment attached to the concepts. NetMapper can be rapidly customized to support extraction of just terms of interest

Each text is processed separately. Three things can be exported. First, a network representation of the text in which all concepts are of the same type – knowledge. This is known as the semantic network. Second, a network representation of the texts in which all concepts are classified by type. This is known as a meta-network. Finally, each concept in the text is evaluated for its context level sentiment. With data in a json or csv format, e.g. twitter or citation data, NetMapper cam be used to extract one network across all entries (so all tweets or all citations.)

The output from NetMapper is in CSV and in the xml format read by ORA. NetMapper is partially lexicon based. It contains an extensive set of thesauri, translation files and delete lists in over 40 languages. In addition, it also supports the use of user-generated domain thesauri and delete lists. Hence, users who are working in a specialized area or on a specialized topic can fine tune what concepts are extracted using the domain files.

4 Conclusion

Interoperability, support for big data, and special support for social media are a few of the hallmarks of these technologies. They are continually upgraded. Application domains include, but are not limited to, public health, medical informatics, counter-terrorism, gang analysis, criminal investigations, social media studies, and team analysis.

References

Kathleen M. Carley, 2017, "ORA: A Toolkit for Dynamic Network Analysis and Visualization." In Reda Alhajj and Jon Rokne (Eds.) Encyclopedia of Social Network Analysis and Mining, Springer. DOI:10.1007/978-1-4614-7163-9_309-1

Neal Altman, Kathleen M. Carley and Jeffrey Reminga, 2017, ORA User's Guide 2017, Carnegie Mellon University, School of Computer Science, Institute for Software Research, Pittsburgh, Pennsylvania, Technical Report CMU-ISR-17-100.