ORA Datasets 2017

Neal Altman, Kathleen M. Carley & Jeffrey Reminga
August 23, 2017
CMU-ISR-17-112

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Center for the Computational Analysis of Social and Organizational Systems
CASOS technical report.

Keywords: ORA, Social Networks, Dynamic Networks, Meta-network, Network Data

Abstract

This report briefly describes network datasets produced at the Center for Computational Analysis of Social and Organizational Systems (CASOS). These datasets are related to the network visualization and analysis tool ORA and articles for publication in PLOS ONE, a peer-reviewed open access scientific journal published by the Public Library of Science (PLOS). The files and datasets are available for download at the CASOS web site (http://www.casos.cs.cmu.edu/tools/data2.php).

Contents

1 Datasets and Descriptions

1.1 Network Datasets

This report describes the networks available for download at the CASOS web site as of the report date. These networks cover a broad variety of network and content types and are grouped by source.

1.1.1 CASOS Produced Datasets

The CASOS produced datasets were created by the Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University and are oriented towards the use of the ORA network tool and the DyNetML network data format.

1.1.2 PLOS ONE Datasets

Per requirement, data relating to articles submitted to PLOS ONE, a peer-reviewed open access scientific journal published by the Public Library of Science (PLOS) are provided on the CASOS web site. These datasets employ a variety of networking tools and network format, many of which are compatible with ORA.

1.1.3 Public Datasets

In addition to the CASOS produced datasets described in this report, a number of public datasets are available on the CASOS web site.

The public datasets were produced by other scholars for a variety of tools and purposes. Released for general use they have been collected by CASOS for research and instructional use and copies are hosted on the CASOS server. Many of the public datasets have been converted into formats which are compatible with ORA and other CASOS tools.

These datasets **are not** further described in this report but can be accessed at the CASOS web site: http://www.casos.cs.cmu.edu/tools/data2.php under **Other Public Datasets**.

1.2 Dataset Descriptions

The remainder of this report briefly describes each of the CASOS and PLOS ONE datasets using the following format and conventions. The general format is:

Dataset Name

Textual description of the dataset.

Dataset Characteristics					
Created by:	{ individual , organization name}		**Creation Date:**		
# Files:		**File Format:**	[1]	**In ZIP Archive:**	{yes/no} [2]
Download URL:	{Link to CASOS web site for dataset download.}				
References:	{Relevant references about the dataset, creation and analysis.}				
How to Cite:	{Recommended citation for dataset itself.}				
Notes:	{Additional information on dataset such as file layout details.}				

*The **Dataset Characteristics** table describes the dataset file(s) and how to access and cite them.*

Notes:

[1] **File Format:** – The file format for a dataset, may include multiple types. This field describes the *contents* of the ZIP archive (if present) not the archive itself.

[2] **In ZIP Archive:** – Indicates whether the dataset file(s) are encoded as a ZIP archive. The ZIP format encapsulates multiple files in a lossless data compressed format. Using the format also precludes automatic attempts by web browsers to display certain file types (e.g. XML files may be automatically displayed inline rather than downloaded by some browsers). It follows that certain dataset ZIP archives may contain a single file or achieve very low compression ratios.

Network Characteristics					
Network Type:	*{Dynamic Meta-Network/ Meta-Network/ Meta-Network / Nodeset}* [1]		**# of Networks:**	DM-Net: #/ MetaNet: #/Net: # [2]	
Link Type:	*{Binary / Weighted / N/A}* [3]	**Graph Type:**	*{Directed / Symmetric/ N/A }* [4]	**Density:**	*{Sparse / Dense/ N/A }* [5]
Has Geo Data:	*{yes / no / N/A}* [6]	**Has Time Data:**	*{yes / no / N/A}* [7]		
Number of Nodes by Class [8]					
Agent:		Organization:		Knowledge:	
Resource:		Belief:		Event:	
Task:		Location:		Role:	
Action:		Unknown:		**Total Nodes:**	
Notes:					

*The **Network Characteristics** table describes the network(s) contained in the dataset or representative networks built from the dataset contents. This table may be omitted if the dataset does not contain prebuilt networks or uses non-CASOS network formats.*

Notes:

[1] **Network type:** – is defined using the general CASOS hierarchy. From lowest complexity to highest, where each level contains the lower level entities:

- **Nodeset** – List of entities of a similar type (e.g. a list of persons) along with optional additional data about the entities (**attributes**: such as age, alternative names or computed measures).[1]

---

[1] A nodeset will be encapsulated in a meta-network when delivered in DyNetML format. A meta-network containing only nodesets will be classified as a nodeset in this document.

- **Network** – A matrix indicating connections between one or two nodesets by marking the presence or absence of links connecting them (optionally including the connection strength and direction).[2]

- **Meta-Network** – A data structure contain zero to many nodesets and networks describing a complex structure made of multiple different nodesets and networks. For example, a business might be considered to consist of persons, sub-organizations, locations, tasks and processes defining a single large network.

- **Dynamic Meta-Network** – A data structure containing zero to many meta-networks recording how an entity changes over time. For example the evolution of an online community of interest about a news event can be recorded as a dynamic meta-network.

- **N/A** – indicates dataset does not contain a network, whether predefined or as raw data, or the network data cannot be characterized using the CASOS hierarchy.

[2] **# of Networks:** – The number of constituent components as defined under network type. Does not necessarily correspond to the number of files (e.g. more than one file may be required to encode a network or more than one network may be contained in a file):

- **DM-Net:** – The number of dynamic meta-networks contained in the dataset.

- **MetaNet:** – The total number of meta-networks contained in the dataset (sometimes referred to as "frames" when meta-networks are contained in a dynamic meta-network).

- **Net:** – The total number of networks contained in the dataset.

[3] **Link Type**: – In ORA, connections in a network (links) are encoded as number. A 0 value indicates no connection; any other positive or negative decimal value indicates a link.

- **Binary** – All link weights are either 0 (no link) or 1 in all networks.

- **Weighted** – One or more link weights are not 0 or 1 in any network.

- **N/A** – The link type is not relevant or unknown.

[4] **Graph Type:** – Links within an ORA network may directed (a link occurs from X to Y) or undirected (a link connects both X to Y and Y to X).

- **Directed** – One or more link in a constituent network is directed.

- **Symmetric** – All links are undirected.

- **N/A** – Not relevant or unknown.

---

[2] In the ORA network tool, a network requires one or two nodesets exist in order to create a network both of which are associated together as part of a meta-network. However, the term network, as generally used, does not require this association and CASOS datasets may encoded using file formats that ORA will translate into meta-networks prior to analysis. Hence this report may describe a dataset as a "network" per general usage.

[5] **Density:** – Density is a per-network measure of the proportion of actual links to possible links.

- **Sparse** – All constituent networks have a density value of 0.50 or less.
- **Dense** – One or more constituent networks have a density value greater than 0.50.
- **N/A** – Not relevant or unknown.

[6] **Has Geo Data:** – Locational data (latitude and longitude) is present as an attribute in one or more nodesets contained in the dataset or is present in a data file that can be used to create a network.

- **Yes** – Locational data is present in at least one nodeset.
- **No** – No locational data is included in any nodeset.
- **N/A** – Not relevant or unknown.

[7] **Has Time Data:** – Time values are present in meta-networks or time data is available in a data file that can be used to create time stamped meta-networks; values may be dates (14 January 2003) or relative (e.g. A, B, C). Note that this is a meta-network specific data field in ORA and the presence of time information in attributes is not relevant.

- **Yes** – Time information is present for all meta-networks.
- **No** – Time information is not included for one or more meta-networks.
- **N/A** – Not relevant or unknown.

[8] **Number of Nodes by Class** – The number of nodes in each of ORA's nodeset classes. A value of 0 indicates a nodeset of a given class is present but empty. A dash (-) indicates that no nodeset of that class is present. When multiple nodesets of a nodeset class are present, the counts are given for each nodeset separated by a slash (e.g. 5/2/0/1). Multiple nodesets may be present in a single network or when multiple networks are contained in a dynamic meta-network. Notes distinguish between these cases.

2   CASOS Produced Datasets

Network datasets described in this section were produced at the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The individual networks were produced for research and training purposes. These datasets were developed in conjunction with ORA, a dynamic meta-network assessment and analysis tool developed by CASOS at Carnegie Mellon University.

2.1   a2c2_team_d_scuds

The a2c2_team_d_scuds dataset.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**	May 2007	
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://casos.cs.cmu.edu/tools/datasets/internal/a2c2_team_d_scuds/a2c2_team_d				

4

	_scuds.xml.zip
References:	
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php
Notes:	

Network Characteristics					
Network Type:	Meta-Network	**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 4		
Link Type:	Binary	**Graph Type:**	Directed	**Density:**	Dense
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	6	Organization:	-	Knowledge:	-
Resource:	94	Belief:	-	Event:	-
Task:	152	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	252
Notes:					

2.2 a2c2_team_d_no_scuds

The a2c2_team_d_no_scuds dataset.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**	May 2007	
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://casos.cs.cmu.edu/tools/datasets/internal/a2c2_team_d_no_scuds/a2c2_team_d_no_scuds.xml.zip				
References:					
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics					
Network Type:	Meta-Network	**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 4		
Link Type:	Binary	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	6	Organization:	-	Knowledge:	-
Resource:	83	Belief:	-	Event:	-
Task:	152	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	241
Notes:					

2.3 Company

The Company dataset.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**	May 2007	
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/company/company.xml.zip				
References:					
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 3	
Link Type:	Binary	**Graph Type:**	Symmetric	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	16	Organization:	-	Knowledge:	19
Resource:	-	Belief:	-	Event:	-
Task:	24	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	59
Notes:					

2.4 Hatfield-McCoy

This dataset contains a list of individuals (agents) involved in the Hatfield-McCoy feud between two rural families who lived in the border area between the US states of West Virginia and Kentucky along the Tug Fork of the Big Sandy River in the years 1863–1891. The list of agents is augmented by a multiple attributes describing agent affiliation, participation and associations.

The dataset was constructed based on the family webpages and Wikipedia and is presented in two forms: as an Excel spreadsheet and as an ORA DyNetML nodeset.

Networks can be constructed using ORA based on these attributes but the dataset itself contains only the agent list.

Dataset Characteristics					
Created by:	Dr. Kathleen M. Carley, CASOS		**Creation Date:**	2017	
# Files:	1	**File Format:**	DyNetML, Excel Spreadsheet	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/HatfieldMcCoy/HatfieldMcCoy.zip http://www.casos.cs.cmu.edu/tools/datasets/internal/HatfieldMcCoy/HatField%20and%20McCoy.zip				

References:	Carley, Kathleen M., Morgan, Geoffrey M., & Levine, Joel. (2017). Socio-cultural Cognitive Mapping. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-17-115.
How to Cite:	Carley, Kathleen M., Morgan, Geoffrey M., & Levine, Joel. (2017). Socio-cultural Cognitive Mapping. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-17-115.
Notes:	

Network Characteristics					
Network Type:	Nodeset[*]		# of Networks:	DM-Net: 0/ MetaNet: 0/Net: 0	
Link Type:	N/A	Graph Type:	N/A	Density:	N/A
Has Geo Data:	No	Has Time Data:	No		
Number of Nodes by Class					
Agent:	66	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	Total Nodes:	66
Notes:	* The nodeset is encapsulated as part of a meta-network, per DyNetML convention. The following attributes are associated with the Agent nodeset: *Node Name, Node Title, Devil Anse Kids, Harmed Hatfield, Harmed McCoy, Hatfield, Intermarried, Killed due to Feud, McCoy, Randolph Kids, female, male, man, woman*.				

2.5 Tanzania Embassy CT

This dataset contains five meta-networks with similar structure describing the plot to bomb the US embassy in Tanzania at different periods in time.

Dataset Characteristics					
Created by:	CASOS		Creation Date:		
# Files:	5	File Format:	DyNetML	In ZIP Archive:	Yes[*]
Download URL:	embassy_tanzania_ct_period1.xml : http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/embassy_tanzania_ct_period1.zip embassy_tanzania_ct_period2.xml : http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/embassy_tanzania_ct_period2.zip embassy_tanzania_ct_period3.xml : http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/embassy_tanzania_ct_period3.zip embassy_tanzania_ct_period4.xml : http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/embassy_tanzania_ct_period4.zip embassy_tanzania_ct_period5.xml : http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/embassy_tanzania_ct_period5.zip				
References:					
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and				

Notes:	* Files are in separate ZIP archives.

Network Characteristics – embassy_tanzania_ct_period1.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 14	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	10	Organization:	-	Knowledge:	4
Resource:	4	Belief:	-	Event:	3
Task:	5	Location:	54	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	80
Notes:					

Network Characteristics – embassy_tanzania_ct_period2.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 14	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	14	Organization:	-	Knowledge:	4
Resource:	4	Belief:	-	Event:	3
Task:	5	Location:	54	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	84
Notes:					

Network Characteristics – embassy_tanzania_ct_period3.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 14	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	16	Organization:	-	Knowledge:	4
Resource:	4	Belief:	-	Event:	3
Task:	5	Location:	54	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	86
Notes:					

Network Characteristics – embassy_tanzania_ct_period4.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 14	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	16	Organization:	-	Knowledge:	4
Resource:	4	Belief:	-	Event:	3
Task:	5	Location:	54	Role:	-

Action:	-	Unknown:	-	**Total Nodes:**	86
Notes:					

Network Characteristics – embassy_tanzania_ct_period5.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 14	
Link Type:	Weighted	**Graph Type:**	Weighted	**Density:**	Sparse
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	15	Organization:	-	Knowledge:	4
Resource:	4	Belief:	-	Event:	3
Task:	5	Location:	54	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	85
Notes:					

2.6 Tanzania Embassy Bombing - 2009

These are two different versions of Tanzania over time. One is specifically for trails: things move across time, only have one location at a time. In this one, only the location edges are changing. All other information is known from the start.

The other is based on a learning/surveillance scenario, where more and more information is slowly learned about the target. In this one, you start out with only a little information about the targets and slowly the meta-network is built up over time. This one has trails, too, but they tend to be shorter. For both of these datasets, the temporal component is completely made up, but if they're aggregated across all time periods, you get the "tanzaniab_2009.xml" meta-network.

Dataset Characteristics			
Created by:	CASOS	**Creation Date:**	
# Files:	2	**File Format:** DyNetML	**In ZIP Archive:** Yes
Download URL:	tanzania_3_2009b_trails.xml.zip: http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/tanzania_3_2009b_trails.xml.zip tanzania_3_2009b_learning.xml.zip: http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/tanzania_3_2009b_learning.xml.zip		
References:			
How to Cite:	CASOS Center, Institute for Software Research, Carnegie Mellon University. (2009). *Tanzania Embassy Bombing 2009 data set* [Data set]. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php		
Notes:			

Network Characteristics – tanzania_3_2009b_trails.xml					
Network Type:	Dynamic Meta-Network		**# of Networks:**	DM-Net: 1/ MetaNet: 10/Net: 170	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		

Number of Nodes by Class[*]					
Agent:	30[*]	Organization:	10	Knowledge:	24
Resource:	18	Belief:	3	Event:	5
Task:	37	Location:	34	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	161
Notes:	* All meta-networks (frames) contain the same number of nodesets and nodes.				

Network Characteristics – tanzania_3_2009b_learning.xml					
Network Type:	Dynamic Meta-Network	**# of Networks:**	DM-Net: 1/ MetaNet: 10/Net: 170		
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Dense
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class[*]					
Agent:	3/6/9/12/15/ 18/21/24/27/ 30[*]	Organization:	1/2/3/4/5/6/7/ 8/9/10	Knowledge:	2/4/7/9/12/14/ 16/19/21/24
Resource:	1/3/5/7/9/10/ 12/14/16/18	Belief:	0/0/0/1/1/1/2/ 2/2/3	Event:	0/1/1/2/2/3/3/ 4/4/5
Task:	3/7/11/14/18/ 22/25/29/33/ 37	Location:	5/6/11/15/22/ 23/24/30/32/ 34	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	15/29/47/64/ 84/97/110/ 130/144/161
Notes:	* Node counts are listed by first to last frame in the dynamic meta-network.				

2.7 Tanzania Embassy Bombing - 2006

The Tanzania Embassy Bombing - 2006 dataset.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**		
# Files:	3	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	tanzania_3_2006.xml.zip: http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/Tanzania_Embassy_ Bombing_2006/tanzania_3_2006.xml.zip tanzania_4_2006.xml.zip: http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/Tanzania_Embassy_ Bombing_2006/tanzania_4_2006.xml.zip tanzania_5_2006.xml.zip: http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/Tanzania_Embassy_ Bombing_2006/tanzania_5_2006.xml.zip				
References:					
How to Cite:	CASOS Center, Institute for Software Research, Carnegie Mellon University. (2008). *Tanzania Embassy Bombing 2006 data set* [Data set]. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics – tanzania_3_2006.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 17	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Dense
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	27	Organization:	2	Knowledge:	26
Resource:	19	Belief:	2	Event:	5
Task:	35	Location:	29	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	145
Notes:					

Network Characteristics – tanzania_4_2006.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 19	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Dense
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	26	Organization:	2	Knowledge:	23
Resource:	22	Belief:	2	Event:	6
Task:	43	Location:	32	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	156
Notes:					

Network Characteristics – tanzania_5_2006.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 20	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Dense
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	21	Organization:	2	Knowledge:	24
Resource:	21	Belief:	2	Event:	7
Task:	44	Location:	38	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	159
Notes:					

2.8 Tanzania Embassy Bombing - 2004

The Tanzania Embassy Bombing - 2004 dataset.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**		
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/embassy/embassy.xml.zip				
References:					
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining, Springer*. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				

11

Notes:	

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Binary	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	16	Organization:	-	Knowledge:	4
Resource:	4	Belief:	-	Event:	-
Task:	5	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	29
Notes:					

2.9 Tanzania Kenya

SData based off the Tanzania - Kenya bombing.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**		
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/tanzania_kenya/tanzania-kenya-imoon.xml.zip				
References:					
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 12	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	18	Organization:	-	Knowledge:	14
Resource:	13	Belief:	-	Event:	-
Task:	25	Location:	5	Role:	-
Action:	-	Unknown:		**Total Nodes:**	75
Notes:					

2.10 Tanzania Geospatial

SData based off the Tanzania bombing. Data includes geospatial information to be used in ORA.

Dataset Characteristics

Created by:	CASOS		Creation Date:			
# Files:	1		File Format:	DyNetML	In ZIP Archive:	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/tanzaniaGis/tanzania_3_2009 b.xml.zip					
References:						
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php					
Notes:						

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: / MetaNet: /Net:	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	30	Organization:	10	Knowledge:	24
Resource:	18	Belief:	3	Event:	5
Task:	37	Location:	34	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	161
Notes:					

2.11 Glitch

This dataset is a collection of data describing many of the economic and social interactions that occurred in Glitch, a recently closed massively multiplayer online game aimed at casual players. The bulk of the data are records of sales of virtual goods by different players during the game's life, stored as CSV files. The records depict a history of the Currant, the game's principal currency. The data also includes both a snowball-sample based scrape of all of the explicit friendships between different players (where "friendships" are minimally defined relationships that players may create for any reason) and a scrape of the seven different Glitch forums. This scrape is presented as both raw HTML and in a minimally parsed format sufficient for generating a basic network of posters and responders, while poster and responder networks are provided in DyNetML format. This data was used and collected in Tech Report, CMU-ISR-13-100.

The data is divided into two sections based on format: Network DyNetML files are contained in the "DyNetML files" directory, while CSV files and raw HTML are contailed in the "CSV and HTML files" directory. The directory tree is as follows:

- **CSV and HTML files**
  - **economicData**
    - allGlitchAuctions.csv - All of the auctions carried out during Glitch's life that were successfully scraped via the API.
    - allGlitchSDBsales.csv - All SDB-based sales carried out during Glitch's life that were successfully scraped via the API.

- glitchStreetPricesDecember.csv - The official "street price" of every item listed in the official Glitch Encyclopedia in December, 2012. (The game includes many items not present in the encyclopedia.)
  - o **forumDigests**
    - announcementsDigest.csv - Digest of all of the posts and responses in the Announcements Forum.
    - bugsDigest.csv - Digest of all of the posts and responses in the Bugs Forum.
    - developerDigest.csv - Digest of all of the posts and responses in the Developer Forum.
    - generalDigest.csv - Digest of all of the posts and responses in the General Forum.
    - ideasDigest.csv - Digest of all of the posts and responses in the Ideas Forum.
    - marketplaceDigest.csv - Digest of all of the posts and responses in the Marketplace Forum.
    - offtopicDigest.csv - Digest of all of the posts and responses in the Offtopic Forum.
  - o **forumsRaw**
    - **announcementsThreads** (directory) - HTML scrape of all threads posted to the Announcements Forum.
    - **bugsThreads** (directory) - HTML scrape of all threads posted to the Bugs Forum.
    - **developerThreads** (directory) - HTML scrape of all threads posted to the Developer Forum.
    - **generalThreads** (directory) - HTML scrape of all threads posted to the General Forum.
    - **ideasThreads** (directory) - HTML scrape of all threads posted to the Ideas Forum.
    - **marketplaceThreads** (directory) - HTML scrape of all threads posted to the Marketplace Forum.
    - **offtopicThreads** - HTML scrape of all threads posted to the Off Topic Forum.
  - o playerFriendships.csv - A CSV file of all explicit player friendships in Glitch. These connections are not necessarily mutual.
- **DyNetML files**
  - o Player Friendships.xml - A DyNetML network of all explicit player friendships in Glitch. These connections are not necessarily mutual.
  - o **forumNetworks**
    - Announcements Network.xml - DyNetML network linking posters, repliers, posts, and replies in the Announcements Forum.

14

- Bugs Network.xml - DyNetML network linking posters, repliers, posts, and replies in the Bugs Forum.
- Developer Network.xml - DyNetML network linking posters, repliers, posts, and replies in the Developer Forum.
- General Network.xml - DyNetML network linking posters, repliers, posts, and replies in the General Forum.
- Ideas Network.xml - DyNetML network linking posters, repliers, posts, and replies in the Ideas Forum.
- Marketplace Network.xml - DyNetML network linking posters, repliers, posts, and replies in the Marketplace Forum.
- Offtopic Network.xml - DyNetML network linking posters, repliers, posts, and replies in the Off Topic Forum.

Dataset Characteristics					
Created by:	CASOS	**Creation Date:**			
# Files:	31901	**File Format:**	Various	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/glitch/glitchDataset.xml.zip				
References:	Landwehr, Peter. (2013). A Collection of Economic and Social Data from Glitch, a Massively Multiplayer Online Game. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-13-100.				
How to Cite:	Landwehr, Peter. (2013). A Collection of Economic and Social Data from Glitch, a Massively Multiplayer Online Game. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-13-100. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics – Player Friendships.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Binary	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	62060	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	62060
Notes:					

Network Characteristics – Announcements Network.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 4	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	2046	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	416/8150[*]
Task:	-	Location:	-	Role:	-

Action:	-	Unknown:	-	**Total Nodes:**	10612
Notes:	* Contained in two nodesets, Posts and Replies.				

Network Characteristics – Bugs Network.xml

Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 4	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	2980	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	4389/13957*
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	21326
Notes:	* Contained in two nodesets, Posts and Replies.				

Network Characteristics – Developer Network.xml

Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 4	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	188	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	188/739*
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	1115
Notes:	* Contained in two nodesets, Posts and Replies.				

Network Characteristics – General Network.xml

Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 4	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	8555	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	12323/ 209874*
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	230752
Notes:	* Contained in two nodesets, Posts and Replies.				

Network Characteristics – Ideas Network.xml

Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 4	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	4264	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	8009/40669*
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	52942

Notes:	* Contained in two nodesets, Posts and Replies.

Network Characteristics – Marketplace Network.xml

Network Type:	Meta-Network			# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 4	
Link Type:	Weighted	Graph Type:	Directed	Density:	Sparse	
Has Geo Data:	No	Has Time Data:	No			
Number of Nodes by Class						
Agent:	1764	Organization:	-	Knowledge:	-	
Resource:	-	Belief:	-	Event:	3691/16485[*]	
Task:	-	Location:	-	Role:	-	
Action:	-	Unknown:	-	**Total Nodes:**	21940	
Notes:	* Contained in two nodesets, Posts and Replies.					

Network Characteristics – Offtopic Network.xml

Network Type:	Meta-Network			# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 4	
Link Type:	Weighted	Graph Type:	Directed	Density:	Sparse	
Has Geo Data:	No	Has Time Data:	No			
Number of Nodes by Class						
Agent:	2269	Organization:	-	Knowledge:	-	
Resource:	-	Belief:	-	Event:	958/18713[*]	
Task:	-	Location:	-	Role:	-	
Action:	-	Unknown:	-	**Total Nodes:**	21940	
Notes:	* Contained in two nodesets, Posts and Replies.					

2.12 kc_node_aa

The kc_node_aa dataset.

Dataset Characteristics

Created by:	CASOS		Creation Date:		
# Files:	1	File Format:	DyNetML	In ZIP Archive:	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/kc_node_aa/kc_node_aa.xml.zip				
References:					
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics

Network Type:	Meta-Network			# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Binary	Graph Type:	Directed	Density:	Sparse	
Has Geo Data:	No	Has Time Data:	No			
Number of Nodes by Class						

Agent:	5	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	5
Notes:					

2.13 kc_node_af

The kc_node_af dataset.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**		
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/kc_node_af/kc_node_af.xml.zip				
References:					
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Binary	**Graph Type:**	Symmetric	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	5	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	5
Notes:					

2.14 Sci-Fi Books

The attributes (which can also be viewed as knowledge nodes) are:

- **frequency** - the number of lists of top N science fiction books that these appeared in circa 2016
- **date** - the date first published; if it is a series, date listed is the date first book was published
- **author** - author of the book or series
- **century** - which century it was written in; 1 = prior to 1900, 2 = prior to 2000, 3 = prior to 2100

18

- **quarter century** - 1 = 1800 to 1824; 2 = 1825 to 1849; 3 = 1850 to 1874; 4 = 1875 to 1899; 5 = 1900 to 1924; 6 = 1925 to 1949; 7 = 1950 to 1974; 8 = 1975 to 1999; 9 = 2000 to 2024
- **author gender** - 1 = male; 2 = female
- *content of story* - for each of the following, rates were given about the content of the story. A 4 point scale was used and each book was read by 2 people with disagreements argued until agreement was reached; 0 = that was not present; 1 = that was present but peripheral; 2 = that was present at a stronger level but not strongly integral to the story; 3 = that was present, strong and integral to the story.
  - **robots, androids or AI computers**
  - **battles**
  - **romance**
  - **magic**
  - **time travel**
  - **interplanetary**
  - **multi-species**
  - **sentient species**
  - **beasts**
  - **psychic powers**
  - **novel technology** - (not AI-ish, ex. steam based technology is considered novel)
  - **after catastrophe** - often post-apocalyptic

Dataset Characteristics					
Created by:	Dr. Kathleen M. Carley, CASOS		Creation Date:		
# Files:	1	File Format:	Excel Spreadsheet	In ZIP Archive:	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/sci-fi/Sci-Fi%20Books.zip				
References:					
How to Cite:	Kathleen M. Carley (2017). *Sci-Fi Books data set* [Data set]. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics					
Network Type:	N/A*		# of Networks:	DM-Net: 0/ MetaNet: 0/Net: 0	
Link Type:	N/A	Graph Type:	N/A	Density:	N/A
Has Geo Data:	No	Has Time Data:	Yes		
Number of Nodes by Class					
Agent:	-	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	Total Nodes:	157
Notes:	* Dataset does not contain a pre-built network.				

2.15 Syria in the News

This is data collected from the world-wide web indicating features of the various groups in Syria. These are the same groups as in another Syrian dataset so it can be used for comparison. The data is binary, a 1 indicates that feature is present; a 0 indicates it is not.

Dataset Characteristics				
Created by:	CASOS	**Creation Date:**		
# Files:	1	**File Format:**	Excel Spreadsheet	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/syria/Syria%20from%20news.zip			
References:				
How to Cite:	Kathleen M. Carley (2017). *Syria in the news data set* [Data set]. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php			
Notes:				

Network Characteristics					
Network Type:	N/A*		**# of Networks:**	DM-Net: 0/ MetaNet: 0/Net: 0	
Link Type:	N/A	**Graph Type:**	N/A	**Density:**	N/A
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	-	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	11
Notes:	* Dataset does not contain a pre-built network.				

2.16 Startup Company

The Startup Company dataset hand coded by Dr. Kathleen M. Carley and company Netanomics. This is based on ethnographic observations of a CMU based startup, May 2012.

Dataset Characteristics				
Created by:	Dr. Kathleen M. Carley, CASOS	**Creation Date:**		
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/startup_company/startup_company_2.0.zip			
References:				
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php			
Notes:				

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Binary	**Graph Type:**	Directed	**Density:**	Dense
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	5	Organization:	-	Knowledge:	19
Resource:	-	Belief:	-	Event:	-
Task:	5	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	29
Notes:					

2.17 Welsh Canals

A single input file in ORA's advanced table format encodes geographical, location and time information relating to the network of canals in Wales (as they were described in Wikipedia in 2016). The data covers the canal network's evolution from 1790 to 2016. In several cases, individual canals were closed as commercial enterprises but have been restored and reopened for recreational use causing canals to appear, disappear and then reappear when viewed over time.

As described in the ORA help entry "Import Advanced Table" the file can be used to create a dynamic meta-network of the Welsh canal network over time. In this encoding, a series of geographic places form the nodes which are connected by links representing the canals. Since the number of places is small, the resulting network is only approximate in a geographic sense but, like a subway map, provides an abstract representation of the network. The date information defines the duration of the link from completion to abandonment on a decade scale. (Interestingly, the choice of places mentioned in Wikipedia varies between articles, depending on the author's orientation and might link original canal features, towns and villages or modern transportation features.)

The dynamic meta-network resulting from the ORA help procedure is also provided.

Dataset Characteristics			
Created by:	Neal Altman, CASOS	**Creation Date:**	2016
# Files:	3	**File Format:** Various	**In ZIP Archive:** Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/welshcanals/welshcanalsdataset.zip		
References:			
How to Cite:	Neal Altman (2016). *Welsh Canals data set* [Data set]. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php		
Notes:			

Network Characteristics[*]					
Network Type:	Dynamic Meta-Network		**# of Networks:**	DM-Net: 1/ MetaNet: /Net:	
Link Type:	Binary	**Graph Type:**	N/A	**Density:**	N/A
Has Geo Data:	Yes	**Has Time Data:**	Yes		
Number of Nodes by Class					

Agent:	-	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	77	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	77
Notes:	* Describes the network produced by following the procedure "Import Advanced Table" in the ORA help.				

2.18 West Bank

The West Bank dataset. This data was collected at CASOS. It is about six major terrorist groups that operate in the West Bank. These groups are the Al Aksa Martyrs Brigades, Al Fatah, Al Qaeda, Hamas, Hezbollah, and the Islamic Jihad. We gathered the 18 texts from that the networks were extracted from LexisNexis Academia via exact matching Boolean keyword search for each of the groups. The media searched with LexisNexis were The Economist, The Washington Post, and The New York Times. The time frame of our data set ranges from articles published in 2000 to 2003.

Dataset Characteristics	
Created by:	CASOS
# Files:	1
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/west_bank_18/west_bank_18.xml.zip
References:	
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php
Notes:	

Note: The Dataset Characteristics table also contains "Creation Date:", "File Format: DyNetML", and "In ZIP Archive: Yes" columns.

Network Characteristics					
Network Type:	Meta-Network	**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 36		
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	47	Organization:	38	Knowledge:	32
Resource:	59	Belief:	-	Event:	-
Task:	22	Location:	66	Role:	21
Action:	-	Unknown:	-	**Total Nodes:**	285
Notes:					

2.19 Stargate

Stargate dataset based on the TV series Stargate.

Note: If using the Stargate dataset please refer to the ORA Users Guide.

Dataset Characteristics

Created by:	CASOS		Creation Date:		
# Files:	2	File Format:	DyNetML	In ZIP Archive:	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/stargate/stargate_dynetml.xml.zip				
References:					
How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:	If using the Stargate dataset please refer to the ORA Users Guide.				

Network Characteristics – stargate.xml					
Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 12	
Link Type:	Weighted	Graph Type:	Directed	Density:	Sparse
Has Geo Data:	No	Has Time Data:	No		
Number of Nodes by Class					
Agent:	16	Organization:	-	Knowledge:	9
Resource:	7	Belief:	-	Event:	14
Task:	18	Location:	7	Role:	-
Action:	-	Unknown:	-	Total Nodes:	71
Notes:	File stargate.xml				

Network Characteristics – sg1.xml					
Network Type:	Meta-Network		# of Networks:	DM-Net: / MetaNet: 1/Net: 12	
Link Type:	Binary	Graph Type:	Directed	Density:	Sparse
Has Geo Data:	No	Has Time Data:	No		
Number of Nodes by Class					
Agent:	5	Organization:	-	Knowledge:	9
Resource:	7	Belief:	-	Event:	14
Task:	18	Location:	7	Role:	-
Action:	-	Unknown:	-	Total Nodes:	60
Notes:	File sg1.xml				

2.20 Star Wars

This dataset was created by Dave Columbus (CASOS) and Jon Storrick (CASOS) for use in the ORA Help documentation. The 24 separate meta-network files contain time information and can visualized using the Loom analysis or combined into a dynamic meta-network.

Dataset Characteristics					
Created by:	Dave Columbus and Jon Storrick , CASOS		Creation Date:		
# Files:	24	File Format:	DyNetML	In ZIP Archive:	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/star_wars/star_wars.xml.zip				
References:					

How to Cite:	Carley, Kathleen M. (2014). ORA: A Toolkit for Dynamic Network Analysis and Visualization, In Reda Alhajj and Jon Rokne (Eds.) *Encyclopedia of Social Network Analysis and Mining*, Springer. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php
Notes:	

Network Characteristics – StarWarsP01.xml .. StarWarsP24.xml[*]					
Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 24/Net: 48	
Link Type:	Weighted	Graph Type:	Symmetric	Density:	Sparse
Has Geo Data:	No	Has Time Data:	Yes		
Number of Nodes by Class					
Agent:	11	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	13	Role:	-
Action:	-	Unknown:	-	Total Nodes:	24
Notes:	* All 24 files have the nodeset characteristics as shown here.				

2.21 OSN Threat Groups

The following files are contained in this dataset:

- CJTC_readme.txt – contains information about the data files.
- attributes.tsv – contains node attribute information for users associated with the 2 hop snowball sample described in Benigni and Carley (2016) (see References in Dataset Characteristics, below). The file contains the following fields:
    - o **userID**
    - o **ScreenName**
    - o **followingCount**
    - o **followerCount**
    - o **tweetCount**
    - o **tweetsCollected**
    - o **lastTweet**
    - o **creation_date**
    - o **urlCount**
    - o **mentionCount**

    The fields correspond to fields provided by the Twitter API.
- friend_edgefile.xml – a directed network edge list of the following or friend ties associated with all nodes listed in attributes.tsv.
- mention_edgefile.xml – a directed network edge list of the mention ties associated with all nodes listed in attributes.tsv. Additionally epoch time for each edge is provided in the 'date' field.
- user_ht_edgefile.tsv – a bipartite network edge list of the user to hash tag ties associated with all nodes listed in attributes.tsv. Additionally epoch time for each

edge is provided in the 'date' field. This data can be analyzed using R source code provided at: https://github.com/mbenigni/OSNThreatGroups.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**		
# Files:	5	**File Format:**	Various	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/osnThreatGroups/cjtc.zip				
References:	Matthew Benigni and Kathleen M. Carley, 2016, "From Tweets to Intelligence: Understanding the Islamic Jihad Supporting Communities in Twitter," In *Proceedings of the International Conference SBP-BRiMS 2016*, Kevin S. Xu, David Reitter, Dongwon Lee and Nathaniel Osgood (Eds.) June 28-July 1, 2016 Washington DC, Springer, DOI: 10.1007/978-3-319-39931-7.				
How to Cite:	Matthew Benigni and Kathleen M. Carley (2016) "From Tweets to Intelligence: Understanding the Islamic Jihad Supporting Communities in Twitter," In *Proceedings of the International Conference SBP-BRiMS 2016*, Kevin S. Xu, David Reitter, Dongwon Lee and Nathaniel Osgood (Eds.) June 28-July 1, 2016 Washington DC, Springer, DOI: 10.1007/978-3-319-39931-7. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics – friend_edgefile.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 2	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	2000/2000[*]	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	4000
Notes:	* Divided into 2 separate nodesets.				

Network Characteristics – mention_edgefile.xml					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 4	
Link Type:	Directed	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	2000/2000/2000[*]	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	6000
Notes:	* Divided into 3 separate nodesets.				

Network Characteristics – user_ht_edgefile.xml					
Network Type:	Meta-Network[*]		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse

Has Geo Data:	No	Has Time Data:	No		
Number of Nodes by Class				
Agent:	249014	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	249014
Notes:	* If imported to ORA as a meta-network.				

2.22 OEC Detection

This dataset was created by CASOS graduate student Matthew Benigni for the 2016 SBP-BRiMS Tutorial on Covert Groups.

Dataset Characteristics			
Created by:	Matthew Benigni, CASOS	**Creation Date:**	
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	brims.zip: http://www.casos.cs.cmu.edu/tools/datasets/internal/brims/brims.zip Euromaidan_Twitter_Community.zip: http://www.casos.cs.cmu.edu/tools/datasets/internal/brims/Euromaidan_Twitter_Community.zip		
References:				
How to Cite:	Benigni, Matthew (2016).Detection, Analysis, and Distruption of Online Extremist Communities [Tutorial Presentation]. 2016 SBP-BRiMS International Conference, Washington, DC. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php		
Notes:				

Network Characteristics – Euromaidan_Twitter_Community.xml					
Network Type:	Meta-Network	**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 3		
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class				
Agent:	1094	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	1094
Notes:					

2.23 Good Flightpaths

The flightpaths datasets are near identical, developed serve as a tutorial on how bringing data into the ORA GIS works.

Good Flightpaths showcases the proper form. Airport terminals are of node type Location, each has their own latitude and longitude, and the agents have <Agent x Location> connections, which ORA takes to be an "Agent is located at..." connection.

When you import data into the ORA GIS, it asks two questions in the backend. "Does this data have coordinate data (LatLon, MGRS, etc)" and "What nodesets have coordinate data?" If no coordinate data is found, you have to configure it. If it finds coordinate data in ONLY nodesets of type Location (as in Flightpaths good), then all locations are placed, and any network that TARGETS a location nodeset is said to have a "Is located at" relationship. If it finds coordinate data in nodesets of ANY type besides location, all nodes are placed according to their coordinates, and we make no assumptions about networks. Links are just links.

Dataset Characteristics					
Created by:	Jon Storrick		**Creation Date:**	February 2017	
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/good_flightpaths/Flightpaths.good.zip				
References:					
How to Cite:	Storrick, Jon. *Good flightpaths* [dataset]. Available from: http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 6	
Link Type:	Weighted	**Graph Type:**	Symmetric	**Density:**	Sparse
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	11	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	285[*]	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	296
Notes:	* Contained in 3 separate nodesets.				

2.24 Not Good Flightpaths

The flightpaths datasets are near identical, developed serve as a tutorial on how bringing data into the ORA GIS works.

Not Good Flightpaths is mostly the same as the Flightpaths Good dataset, but the nodesets with geographic data aren't of type location. Not Good Flightpaths, at the time it was developed, would not load in the ORA Visualizer at all. Using it was an exercise in showing users how to configure non-formatted data to load it into GIS. In current versions of ORA, the dataset will load and display without error.

Dataset Characteristics					
Created by:	Jon Storrick		**Creation Date:**	February 2017	
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/bad_flightpaths/Flightpaths.notgood.zip				

References:	
How to Cite:	Storrick, Jon. *Not Good flightpaths* [dataset]. Available from: http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php
Notes:	

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 6	
Link Type:	Weighted	**Graph Type:**	Symmetric	**Density:**	Sparse
Has Geo Data:	Yes	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	11	Organization:	285[*]	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	296
Notes:	* Contained in 3 separate nodesets.				

2.25 Marvel Phase 1

This dataset was created by Netanomics staff and is based on early Marvel super hero comics.

Dataset Characteristics			
Created by:	Netanomics	**Creation Date:**	June 2016
# Files:	1	**File Format:** DyNetML	**In ZIP Archive:** Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/marvel/Marvel%20Phase%201.zip		
References:			
How to Cite:	Netanomics. *Marvel Phase 1* [dataset]. Available from: http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php		
Notes:			

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 10	
Link Type:	Weighted	**Graph Type:**	Binary	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	56	Organization:	17	Knowledge:	3
Resource:	28	Belief:	-	Event:	-
Task:	-	Location:	49	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	153
Notes:					

2.26 PR1

PR1 is a random network for larger data demonstrations.

Dataset Characteristics					
Created by:	Jon Storrick		**Creation Date:**	February 2017	
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/pr1/PR1.zip				
References:					
How to Cite:	Storrick, Jon. *PR1* [dataset]. Available from: http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 4	
Link Type:	Binary	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	657	Organization:	-	Knowledge:	-
Resource:	671	Belief:	-	Event:	-
Task:	-	Location:	668	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	1996
Notes:					

2.27 Raiders of the Lost Ark

Based on the movie "Raiders of the Lost Ark" (1981), directed by Steven Spielberg. Encodes the location of the characters for each time interval (where the location of characters is known).

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**	June 2015	
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://casos.cs.cmu.edu/tools/datasets/internal/raiders/Raiders%20of%20the%20Lost%20Ark%20-%20Dynamic.zip				
References:					
How to Cite:	CASOS Center. *Raiders of the Lost Ark* [dataset]. Available from: http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php.				
Notes:					

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 1/ MetaNet: 27/Net: 27	
Link Type:	Binary	**Graph Type:**	N/A	**Density:**	Sparse
Has Geo Data:	Yes	**Has Time Data:**	Yes		
Number of Nodes by Class					
Agent:	9[*]	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	19[*]	Role:	-

Action:	-	Unknown:	-	**Total Nodes:**	28[*]
Notes:	* All meta-networks contain the same number of nodesets and nodes; 756 nodes in total are contained in the dynamic meta-network as a whole.				

2.28 Valkyria Chronicles

Valkyria Chronicles is a mapping of character affinities and powers from the video game Valkyria Chronicles.

Dataset Characteristics					
Created by:	Jon Storrick		**Creation Date:**	February 2017	
# Files:	1	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/valkyria/Valkyria_Chronicles.zip				
References:					
How to Cite:	Storrick, Jon. *Valkyria Chronicles* [dataset]. Available from: http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php				
Notes:					

Network Characteristics					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 2	
Link Type:	Binary	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	55	Organization:	-	Knowledge:	-
Resource:	-	Belief:	162	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	217
Notes:					

2.29 NATO 2016 Exercises

NATO conducted one of its largest exercises over the past decade, Trident Juncture, from October 21st to November 6th 2015. As part of this exercise, the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon was asked to assess, in partnership with the Data Mining and Machine Learning Lab at Arizona State University, the social media response to Trident Juncture.

The dataset provides social media observations for the period of one week, October 5-11, 2015.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**		
# Files:	7	**File Format:**	DyNetML	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/tools/datasets/internal/nato/Twitter_JSON_merge_10_11_nato.zip				

References:	Frankenstein, Will, Huang, Binxuan, and Carley, Kathleen M. (2016). NATO Trident Juncture on Twitter: Public Discussion. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-16-100.
How to Cite:	Frankenstein, Will, Huang, Binxuan, and Carley, Kathleen M. (2016). NATO Trident Juncture on Twitter: Public Discussion. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-16-100. Retrieved from http://www.casos.cs.cmu.edu/tools/datasets/internal/index.php
Notes:	

Network Characteristics – Twitter JSON merge_10_5_nato.xml

Network Type:	Meta-Network			# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 7
Link Type:	Weighted	Graph Type:	Directed	Density:	
Has Geo Data:	Yes	Has Time Data:	No		
Number of Nodes by Class					
Agent:	611	Organization:	-	Knowledge:	132
Resource:	-	Belief:	-	Event:	571
Task:	-	Location:	49	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	1363
Notes:					

Network Characteristics – Twitter JSON merge_10_6_nato.xml

Network Type:	Meta-Network			# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 12
Link Type:	Weighted	Graph Type:	Directed	Density:	
Has Geo Data:	Yes	Has Time Data:	No		
Number of Nodes by Class					
Agent:	653	Organization:	-	Knowledge:	186
Resource:	-	Belief:	-	Event:	658
Task:	-	Location:	44	Role:	-
Action:	-	Unknown:	42	**Total Nodes:**	1583
Notes:					

Network Characteristics – Twitter JSON merge_10_7_nato.xml

Network Type:	Meta-Network			# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 7
Link Type:	Weighted	Graph Type:	Directed	Density:	Sparse
Has Geo Data:	Yes	Has Time Data:	No		
Number of Nodes by Class					
Agent:	418	Organization:	-	Knowledge:	159
Resource:	-	Belief:	-	Event:	434
Task:	-	Location:	24	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	1035
Notes:					

Network Characteristics – Twitter JSON merge_10_8_nato.xml

Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 7

Link Type:	Weighted	Graph Type:	Directed	Density:	Sparse
Has Geo Data:	Yes	Has Time Data:	No		

Number of Nodes by Class					
Agent:	690	Organization:	-	Knowledge:	221
Resource:	-	Belief:	-	Event:	679
Task:	-	Location:	44	Role:	-
Action:	-	Unknown:	-	Total Nodes:	1634
Notes:					

Network Characteristics – Twitter JSON merge_10_9_nato.xml					
Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 7	
Link Type:	Weighted	Graph Type:	Directed	Density:	Sparse
Has Geo Data:	Yes	Has Time Data:	No		

Number of Nodes by Class					
Agent:	429	Organization:	-	Knowledge:	112
Resource:	-	Belief:	-	Event:	335
Task:	-	Location:	26	Role:	-
Action:	-	Unknown:	-	Total Nodes:	902
Notes:					

Network Characteristics – Twitter JSON merge_10_10_nato.xml					
Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 7	
Link Type:	Weighted	Graph Type:	Directed	Density:	Sparse
Has Geo Data:	Yes	Has Time Data:	No		

Number of Nodes by Class					
Agent:	400	Organization:	-	Knowledge:	71
Resource:	-	Belief:	-	Event:	335
Task:	-	Location:	29	Role:	-
Action:	-	Unknown:	-	Total Nodes:	835
Notes:					

Network Characteristics – Twitter JSON merge_10_11_nato.xml					
Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 7	
Link Type:	Weighted	Graph Type:	Directed	Density:	Sparse
Has Geo Data:	Yes	Has Time Data:	No		

Number of Nodes by Class					
Agent:	452	Organization:	-	Knowledge:	113
Resource:	-	Belief:	-	Event:	399
Task:	-	Location:	31	Role:	-
Action:	-	Unknown:	-	Total Nodes:	995
Notes:					

3 Datasets for the PLOS ONE Articles

Datasets described in this section are related to articles published or under review in the journal PLOS ONE an open access scientific journal published by the Public Library of Science (PLOS).

3.1 The Role of Datasets on Scientific Influence within Conflict Research

3.1.1 OriginalText

3.1.1.1 Conflictall_noreviews.txt

The original, uncleaned text file, as extracted from the Web of Science.

Dataset Characteristics					
Created by:	Van Holt and Moates	**Creation Date:**			
# Files:	1	**File Format:**	Text	**In ZIP Archive:**	No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/1OriginalText/conflictall_noreviews.txt				
References:					
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

3.1.1.2 Clean.WoS

This is a file that was cleaned by WoS file to standardize files. This file can be imported into Pajek.

Dataset Characteristics					
Created by:	Van Holt and Moates	**Creation Date:**			
# Files:	1	**File Format:**	Text	**In ZIP Archive:**	No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/1OriginalText/clean.WoS				
References:					
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

3.1.2 Files for the Critical Path (Fig 1 in the manuscript)

3.1.2.1 Cite.net

This is a Pajek format file that was processed using Pajek WoS, which is the input file for the critical path analysis (see codes in Appendix 2).

Dataset Characteristics					
Created by:	Van Holt and Moates		**Creation Date:**		
# Files:	1	**File Format:**	Pajek .net	**In ZIP Archive:**	No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/2CriticalPathFiles/Cite.net				
References:					
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

Network Characteristics (As Imported by ORA)					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Binary	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	-	Organization:	-	Knowledge:	1543961[*]
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	1543961
Notes:	* Nodeset class was selected during import to ORA.				

3.1.2.2   Figure1_criticalpath.vna

The Netdraw critical path file.

Dataset Characteristics					
Created by:	Van Holt and Moates		**Creation Date:**		
# Files:	1	**File Format:**	Netdraw VNA	**In ZIP Archive:**	No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/2CriticalPathFiles/figure1_criticalpath.vna				
References:					
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

3.1.3 Keywords from only critical path articles (Fig S2)

3.1.3.1   WKAff.##d and WKAff.##h

From only the 49 critical path articles, files which are works by keywords files for UCINET.

Dataset Characteristics – WKAff.##d[*]				
Created by:	Van Holt and Moates		**Creation Date:**	
# Files:	1	**File Format:**	UCINET	**In ZIP Archive:** Yes
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/3keywordcriticalpath/WK-Affd.zip			
References:				
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.			
Notes:	* Paired with WKAff.##h to represent a single network.			

Dataset Characteristics – WKAff.##h[*]				
Created by:	Van Holt and Moates		**Creation Date:**	
# Files:	1	**File Format:**	UCINET	**In ZIP Archive:** Yes
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/3keywordcriticalpath/WK-Affh.zip			
References:				
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.			
Notes:	* Paired with WKAff.##d to represent a single network.			

Network Characteristics – WKAff.##h and WKAff.##d (As Imported by ORA)					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Weighted	**Graph Type:**	Symmetric	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	-	Organization:	-	Knowledge:	1047[*]
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	1047
Notes:	* Nodeset class was selected during import to ORA.				

3.1.3.2   excel sheet

Contains all keywords from the WK affiliation matrix.

Dataset Characteristics				
Created by:	Van Holt and Moates		**Creation Date:**	
# Files:	1	**File Format:**	Excel Spreadsheet	**In ZIP Archive:** No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/3keywordcriticalpath/testaff.xlsx			
References:				
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on*			

	scientific influence within conflict research. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

3.1.3.3   kwfigurefinal2.vna
The final figure file, which contains only the keywords that were retained.

Dataset Characteristics					
Created by:	Van Holt and Moates		**Creation Date:**		
# Files:	1	**File Format:**	Netdraw VNA	**In ZIP Archive:**	No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/3keywordcriticalpath/kwfigurefinal2.vna				
References:					
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

3.1.4 Keywords from 2010 only (Figure 1 in the manuscript)

3.1.4.1   2010deg59+.vna
A UNINET file (vna format) that is the affiliation matrix of a works by keyword file just for 2010. For a node to link, >59 articles had to discuss that keyword pair in common.

Dataset Characteristics					
Created by:	Van Holt and Moates		**Creation Date:**		
# Files:	1	**File Format:**	Netdraw VNA	**In ZIP Archive:**	No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/4Keywords2010/2010deg59+.vna				
References:					
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

3.1.5 Other Files (not used in manuscript)

3.1.5.1   WA.net
Contains works by author file as generated by WoS Pajek from the file clean.WoS..

Dataset Characteristics					
Created by:	Van Holt and Moates		**Creation Date:**		

# Files:	1	File Format:	Pajek .net	In ZIP Archive:	No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/5Others/WA.net				
References:					
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

Network Characteristics (As Imported by ORA)					
Network Type:	Meta-Network		# of Networks:	DM-Net: / MetaNet: /Net:	
Link Type:	Binary	Graph Type:	N/A	Density:	Sparse
Has Geo Data:	No	Has Time Data:	No		
Number of Nodes by Class					
Agent:	1542961[*]	Organization:	-	Knowledge:	575644
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	Total Nodes:	2118605
Notes:	* Nodeset class was selected during import to ORA.				

3.1.5.2   WJ.net

File contains works by journal as generated by WoS Pajek from the file clean.WoS.

Dataset Characteristics					
Created by:	Van Holt and Moates		Creation Date:		
# Files:	1	File Format:	Pajek .net	In ZIP Archive:	No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/5Others/WJ.net				
References:					
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

Network Characteristics (As Imported by ORA)					
Network Type:	Meta-Network		# of Networks:	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Binary	Graph Type:	N/A	Density:	Sparse
Has Geo Data:	No	Has Time Data:	No		
Number of Nodes by Class					
Agent:	1542961	Organization:	-	Knowledge:	75914
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	Total Nodes:	1618875
Notes:					

3.1.5.3   WK.net

File contains works by author as generated by WoS Pajek from the file clean.WoS.

Dataset Characteristics					
Created by:	Van Holt and Moates		**Creation Date:**		
# Files:	1	**File Format:**	Pajek .net	**In ZIP Archive:**	No
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/conflictdataset/5Others/WK.net				
References:					
How to Cite:	Van Holt, T., Johnson, J. C., Moates, S., Carley, K.M. (2016) *The role of datasets on scientific influence within conflict research*. PLOS ONE. DOI: 10.1371/journal.pone.0154148.				
Notes:					

Network Characteristics (As Imported by ORA)					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Binary	**Graph Type:**	N/A	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					
Agent:	1542961	Organization:	-	Knowledge:	575644
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	2118605
Notes:					

3.2 The Broad Reach of Online Extremism: Understanding the ISIS Supporting Community on Twitter

This dataset is part of an article under review by PLOS ONE and can be analyzed using R source code provided at: https://github.com/mbenigni/OSNThreatGroups.

3.2.1 Dataset Files

The following files are derived from the 2 hop snowball sample described in the article

3.2.1.1   deIdentified_attributes.csv

This file contains node attribute information for users associated with the 2 hop snowball sample described in the aforementioned work. The file contains the following fields:

- **anonID** - refers to a unique identifier assigned to each user and corresponds to nodes in the provided edge lists.
- **followingCount** - corresponds to field provided by the Twitter API.
- **followerCount** - corresponds to field provided by the Twitter API.
- **tweetCount** - corresponds to field provided by the Twitter API.
- **lastTweet** - corresponds to field provided by the Twitter API.
- **creation_date** - corresponds to field provided by the Twitter API.
- **lang** - corresponds to field provided by the Twitter API.
- **suspended** - refers to accounts that were suspended by Twitter between NOV14 and MAR15. Some of these suspended accounts were used as positive case labels. A full explanation is provided in the article.

- **official** - refers to a list of human verified media, government, and celebrity accounts used to train the 'official classifier' in our presented work.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**		
# Files:	1	**File Format:**	CSV	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/deIdentified/deIdentified_attributes.zip				
References:					
How to Cite:	Matthew Benigni, Kenneth Joseph and Kathleen M. Carley, 2017-forthcoming, "Online Extremism and the Communities that Sustain It: Detecting the ISIS Supporting Community on Twitter," PLOS ONE.				
Notes:					

Network Characteristics					
Network Type:	N/A		**# of Networks:**	DM-Net: 0/ MetaNet: 0/Net: 0	
Link Type:	N/A	**Graph Type:**	N/A	**Density:**	N/A
Has Geo Data:	No	**Has Time Data:**	Yes		
Number of Nodes by Class					
Agent:	125605[*]	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	125605
Notes:	* Nodeset class was selected arbitrarily.				

3.2.1.2 deIdentified_friend_edges.csv

This file contains a directed network edge list of the following or friend ties associated with all nodes listed in deIdentified_attributes.csv.

Dataset Characteristics					
Created by:	CASOS		**Creation Date:**		
# Files:	1	**File Format:**	CSV	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/deIdentified/deIdentified_friend_edges.zip				
References:					
How to Cite:	Matthew Benigni, Kenneth Joseph and Kathleen M. Carley, 2017-forthcoming, "Online Extremism and the Communities that Sustain It: Detecting the ISIS Supporting Community on Twitter," PLOS ONE.				
Notes:					

Network Characteristics (As Imported by ORA)					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	No		
Number of Nodes by Class					

Agent:	125584[*]	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	125584
Notes:	* Nodeset class was selected during import to ORA.				

3.2.1.3   deIdentified_mention_edges.csv

This file contains a directed network edge list of the mention ties associated with all nodes listed in deIdentified_attributes.csv. Additionally the epoch time for each edge is provided in the 'date' field.

Dataset Characteristics				
Created by:	CASOS	**Creation Date:**		
# Files:	1	**File Format:**	CSV	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/deIdentified/deIdentified_mention_edges.zip			
References:				
How to Cite:	Matthew Benigni, Kenneth Joseph and Kathleen M. Carley, 2017-forthcoming, "Online Extremism and the Communities that Sustain It: Detecting the ISIS Supporting Community on Twitter," PLOS ONE.			
Notes:				

Network Characteristics (As Imported by ORA)					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: 0/ MetaNet: 1/Net: 1	
Link Type:	Weighted	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	Yes[*]		
Number of Nodes by Class					
Agent:	108837[#]	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	108837
Notes:	* Time data is available in the input file; use in network creation is optional. # Nodeset class was selected during import to ORA.				

3.2.1.4   deIdentified_user_ht_edges.csv

This file contains a bipartite network edge list of the user to hash tag ties associated with all nodes listed in deIdentified_attributes.csv. Additionally epoch time for each edge is provided in the 'date' field.

Dataset Characteristics				
Created by:	CASOS	**Creation Date:**		
# Files:	1	**File Format:**	CSV	**In ZIP Archive:**	Yes
Download URL:	http://www.casos.cs.cmu.edu/projects/plosone/deIdentified/deIdentified_user_ht_edges.zip			

References:	
How to Cite:	Matthew Benigni, Kenneth Joseph and Kathleen M. Carley, 2017-forthcoming, "Online Extremism and the Communities that Sustain It: Detecting the ISIS Supporting Community on Twitter," PLOS ONE.
Notes:	

Network Characteristics (As Imported by ORA)					
Network Type:	Meta-Network		**# of Networks:**	DM-Net: / MetaNet: /Net:	
Link Type:	Binary	**Graph Type:**	Directed	**Density:**	Sparse
Has Geo Data:	No	**Has Time Data:**	Yes[*]		
Number of Nodes by Class					
Agent:	4242843[#]	Organization:	-	Knowledge:	-
Resource:	-	Belief:	-	Event:	-
Task:	-	Location:	-	Role:	-
Action:	-	Unknown:	-	**Total Nodes:**	4242843
Notes:	* Time data is available in the input file; use in network creation is optional. # Nodeset class was selected during import to ORA.				