

A Data Processing Pipeline for Socio-Technical Network Analysis

May 2023

Gabriel Arthur Weaver, Daniel





DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

A Data Processing Pipeline for Socio-Technical Network Analysis

Gabriel Arthur Weaver, Daniel

May 2023

Idaho National Laboratory Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the U.S. Department of Energy Under DOE Idaho Operations Office Contract DE-AC07-05ID14517

A Data Processing Pipeline for Socio-Technical Network Analysis INCOSE WG

Motivation

The Problem: How do adversarial business practices impact critical infrastructure systems within a given geographic region of interest?

Specialization of 'cyber' may lead to blindspots for dependencies that achieve influence but are exogenous to traditional system boundaries.

Socio-Technical Network Analysis (STNA) provides an approach to consider interactions between social and cyber domains.



However, the company denied the allegation in a June 27 email to Fifth Domain saving, "there was no breach of information" from Newport News Shipvard, nor

were their systems connected to a foreign server controlled by a Chinese group,

Tactics achieved through the cyber domain may also be achieved through social/business domains.

Adversarial tactics seen in the cybersecurity domain may also be achieved through other domains.

Our research focuses on adversarial techniques in the business domain that affect infrastructure.

Want to think about how such techniques across domains can be composed by adversaries.

Cross-domain adversarial killchains (inspired by Cyber Kill Chain [Assante et al. 2015])

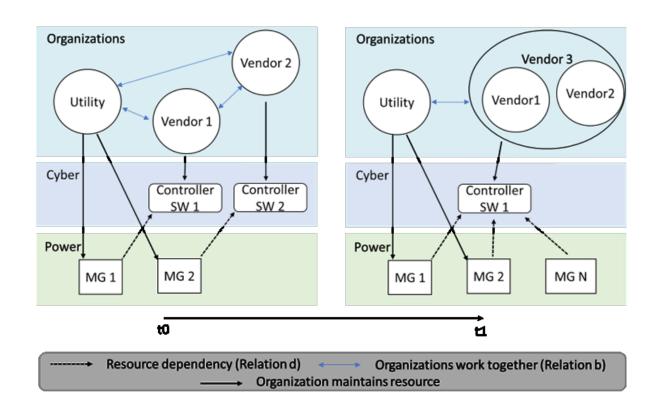
	Acquisition	Ownership	Company Ownership Proxy	Due Diligence for Mergers & Acquisitions	Ownership of Nearby Asset	Parent Company	Foreign- Owned Facilities
	Use	Board Membership	Board Membershi p Proxy	Legal Subpoena	Pilot Studies		Disinformation
Social	Leasing	Asset Maintenanc e		Digital Twin Creation	Business Data Availability Platform Capture		Degrade Service Quality
	Employment	Subcontract/ Supporting Vendor					
		Investing or Lending					
Cyber	Drive-by Compromise	Hardcoded Credentials	Change Operating Mode	Network Connection Enumeratio n	Data from Information Repositories	Commonly Used Port	Unauthorized Command Message

[Alexander et al. 2020]

Our Approach: Adversarial STNA

Contributions

- 1. Develop a data curation and processing pipeline to generate networks that enable Adversarial Socio-Technical Network Analysis.
- 2. Networks produced by our pipeline may be processed as knowledge, geospatial, or dynamic graphs.
 - Enable a variety of analyses to characterize and detect adversarial behaviors



Outline

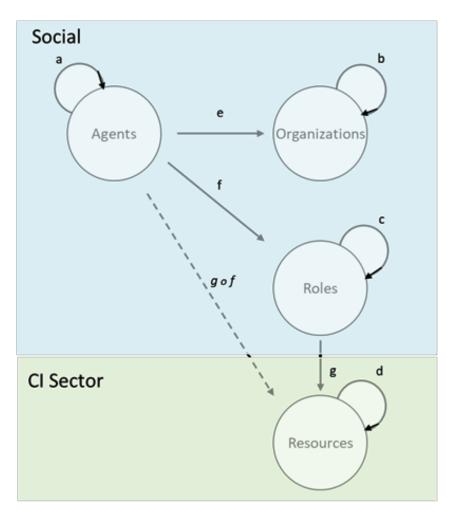
- Introduction
- Theory
- Architecture
- Use Cases
- Conclusion

Theory

Multilayered Networks for Adversarial STNA

Want a systematic approach to study structured relations between domains that can influence critical infrastructure systems (e.g. cyber, social).

- 1. Define types of entities, relations, and attributes via *ontologies*
- 2. Represent structured relations among entities using *multilayered networks* [Kivela 2014]



Network Construction: Entity Occurrences as Vertices

Ontology

- Classes define entity types
- <u>Data properties</u> define attributes of entities.

Multilayered Network

- Entity types provide values for STNA network's semantic layer.
- Attributes affect vertex weighting functions

Social Network: Entity Types and Data Properties							
Entity Type	Data Property	Description	Network Component				
	hasCIK	Central Index Key (CIK)	V				
atno para Organization	hasSIC	Standard Industrial Classification (SIC)	w_semantic				
stna_core.Organization	hasBusinessAddress	Business address	L_Geospatial				
	hasStateOfIncorporation	State of Incorporation	L_Geospatial				
	hasConformedName	Conformed Name	V				
stna_core.Person	hasOfficerTitle	Officer title	w_semantic				
	hasBusinessAddress	Business Address	L_Geospatial				

Network Construction: Relations as Edges

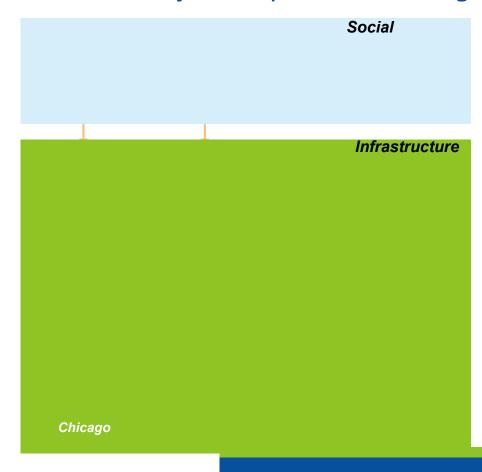
Ontology

Object properties define *relations* between entities.

Object Properties and Features to Construct EM						
Object Property	Subject Class	Object Class				
Social Network						
hasBeneficialOwnership	Organization	Organization				
naosenerrorare merenrp	Person	Organization				
isDirectorOf isOfficerOf isTenPercentOwnerOf	Person	Organization				
Infrastructure Network						
hasEVNetwork	EVChargingStation	EVChargingNetwork				
hasPaymentNetwork	EVChargingStation	PaymentNetwork				
Inter-Layer (Socio-Technical)						
owns maintains	Organization Person	CIAsset				

Multilayered Network

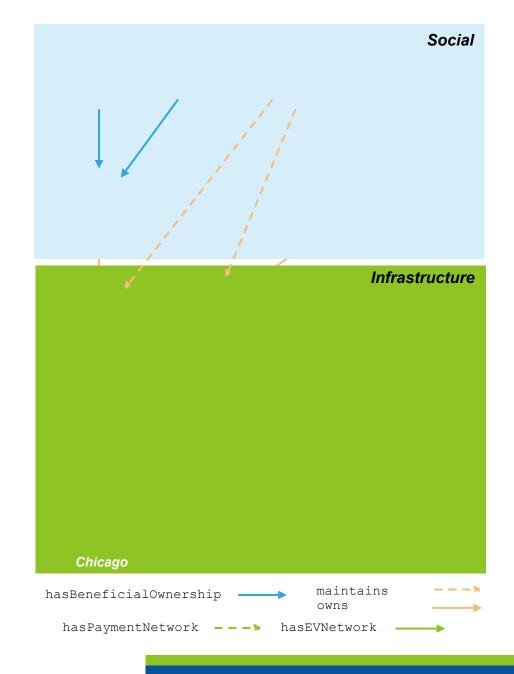
Relations may be expressed as edges.



Analyses Enabled

Our intent is to construct multilayered networks for adversarial STNA that can be adapted to a wide variety of analyses.

- Semantic Layer (L_{semantic})
 - Inference Algorithms
- Geospatial Layer (Lgeospatial)
 - Spatially-Localized Attacks (SLA)[Ouyang 2016]
- Temporal Layer (Ltemporal)
 - Temporal Motifs [Paranjape et al. 2017]



STNA Network Generation

Architecture and Design

Architecture

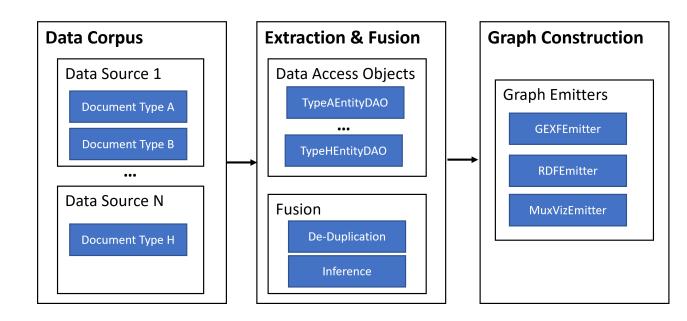
The intent is to construct multilayered networks for adversarial STNA that can be adapted to a wide variety of analyses.

Constructing such networks is difficult

- Diverse, heterogeneous data sources
- Difficult to update, adapt, repurpose

Therefore, we created a data processing pipeline to construct STNA

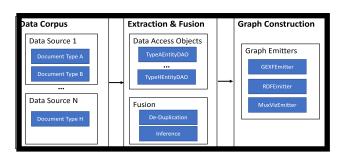
- Domain model for data access
- Networks composed of explicit data types for vertices, nodes, and other components.



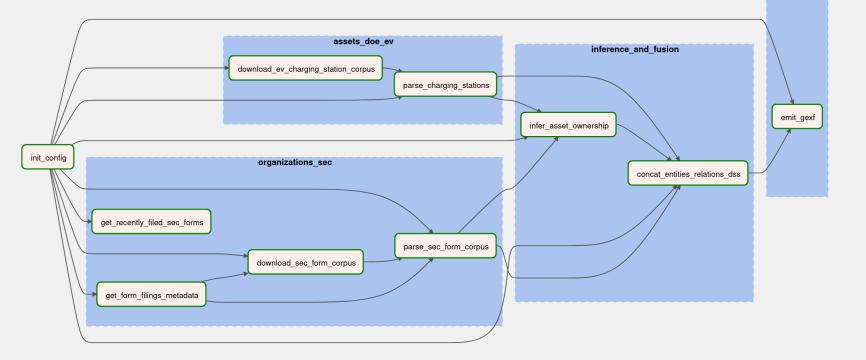
Architecture - Implementation

Init Config

```
{'edgar.data_dir_path': 'build/chicago/1_23_2023/SEC',
  'evasset.data_dir_path': 'build/chicago/1_23_2023/EVChargingStation',
  'fusion.data_dir_path': 'build/chicago/1_23_2023/fusion',
  'editorial.data_dir_path': './build/chicago/1_23_2023/editorial',
  'edgar.cik_list': ['0001819584'],
  'edgar.form_types': ['SC 13D', 'S-4'],
  'evasset.latitude': '41.8781',
  'evasset.longitude': '-87.6298',
  'evasset.radius': '30.0'}
```



emit knowledge graph



Data Corpus

The intent is to construct multilayered networks for adversarial STNA that can be adapted to a wide variety of analyses.

Many types of data sources are available to construct social networks

- Twitter [Debreceny 2017]
- Pitchbook
- Facebook
- SEC Data [Bichler et al. 2015]

3	Initial statement of beneficial ownership of securities
4	Statement of changes in beneficial ownership
S-4	Registration statement after a merger or acquisition between two companies
SC 13D	Filed when an investor or entity purchases more than 5% shares of a public company.

Filed whenever an entity makes a tender offer as

Social Network

SEC Form

TO-T

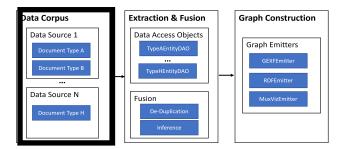
SEC Forms for Organizational Influence

part of a takeover bid.

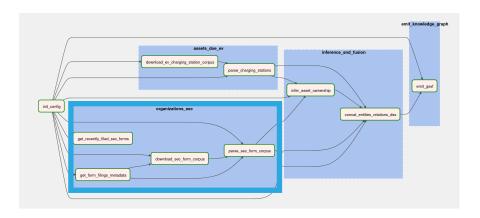
Description

Intent is to be able to

- 1. Construct the social network layers
- 2. Compose with layers for a Critical Infrastructure Sector



Data Corpus - Implementation



view filings metadata (stenv) pasithea@pasithea:~/Documents/LDRD/stna\$ python src/bin/view filings metadata.py --forms metadata path=data/chicago data/1 23 2023/editorial/form filings met∥ adata.json Form SC 13D CIK 0001819584 accessionNumber filingDate I fileNumber primaryDocument I form d9441195_13-d.htm 0000919574-22-002704 | 1970-01-01 00:27:29.721600 005-91687 SC 13D ea149046-13dtough volta.htm | SC 13D 0001213900-21-053569 1970-01-01 00:27:14.688000 005-91687 volta scottmercer-sch13d.htm | SC 13D 0001068238-21-000191 | 1970-01-01 00:27:11.750400 005-91687 volta chriswendel-sch13d.htm | SC 13D 0001068238-21-000190 | 1970-01-01 00:27:11.750400 | 005-91687 Form S-4 CIK 0001819584 accessionNumber I fileNumber | primaryDocument 0001213900-21-026615 | 1970-01-01 00:27:01.209600 | 333-256173 | fs42021 tortoiseacq2.htm | S-4

Social Network Entity Extraction

Social Network - Volta Inc

SC 13D Filings

Accession Number	File Number	File Name
0000919574-22-002704	005-91687	d9441195_13-d.htm
0001213900-21-053569	005-91687	ea149046-13dtough_volta.htm
0001068238-21-000191	005-91687	volta_scottmercer-sch13d.htm

S-4 Filings

0001213900-21-026615	333-256173	fs42021_tortoiseacq2.htm
----------------------	------------	--------------------------

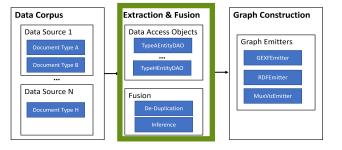
Social Network

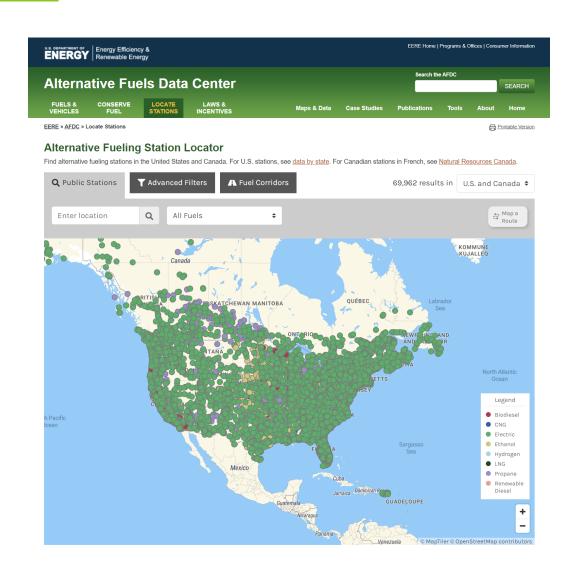
stna_core.Organization

hasName	hasLocation	ownsPercentage	Start date	Data source
Virgo Hermes, LLC	Delaware	10.1%	2022-04-12	d9441195_13-d.htm
Virgo Investment Group LLC	Delaware	10.1%	2022-04-12	d9441195_13-d.htm
Energize Ventures Fund LP	Delaware	6.0%	2021-10-20	Ea149046-13dtough_volta.htm

SC 13D 1 d94411	95_13-d.htm		
	UNITED STATES SECURITIES AND EXCHANGE COMMISSION WASHINGTON, DC 20549		
	SCHEDULE 13D THE SECURITIES EXCHANGE ACT OF 1934 (Amendment No.)*		
	Volta Inc.		
	(Name of Issuer)		
	Class A Common Stock, par value \$0.0001 per share		
	(Title of Class of Securities)		
	92873V102		
	(CUSIP Number)		
	Jesse C. Watson 1201 Howard Avs, Suite 300 Burlingame, California, 94010 United States of America (650) 437-3368		
	(Name, Address and Telephone Number of Person Authorized to Receive Notices and Communications)		
	March 31, 2022		
	(Date of Event Which Requires Filing of this Statement)		
If the filing person	on has previously filed a statement on Schedule 13G to report the acquisition that is the subject of this Schedule 13D, and is filing this schedule because of ss.240.13d-1(e), 240.13d-1(f) or 240.13d-1(g), check the following box	[X].	
* The remainder of	of this cover page shall be filled out for a reporting person's initial filing on this form with respect to the subject class of securities, and for any subsequent amendment containing information which would alter disclosures provi	led in a prior	cover page.
CUSIP No.	92873V102		
1.	NAME OF REPORTING PERSONS LR.S. IDENTIFICATION NOS. OF ABOVE PERSONS (ENTITIES ONLY)		
	Virgo Hermes, LLC		
2.	CHECK THE APPROPRIATE BOX IF A MEMBER OF A GROUP	(a) (b)	
i		(-)	1-1
3.	SEC USE ONLY		
4.	SOURCE OF FUNDS		
	AF		
5.	CHECK BOX IF DISCLOSURE OF LEGAL PROCEEDINGS IS REQUIRED PURSUANT TO ITEMS 2(d) OR 2(e)		[_]
6.	CITIZENSHIP OR PLACE OF ORGANIZATION		
	Delaware		
NUMBER OF S	HARES BENEFICIALLY OWNED BY EACH REPORTING PERSON		
7.	SOLE VOTING POWER		
	0		
8.	SHARED VOTING POWER		
	16,222,891		
9.	SOLE DISPOSITIVE POWER		

Critical Infrastructure Sector Entity Extraction

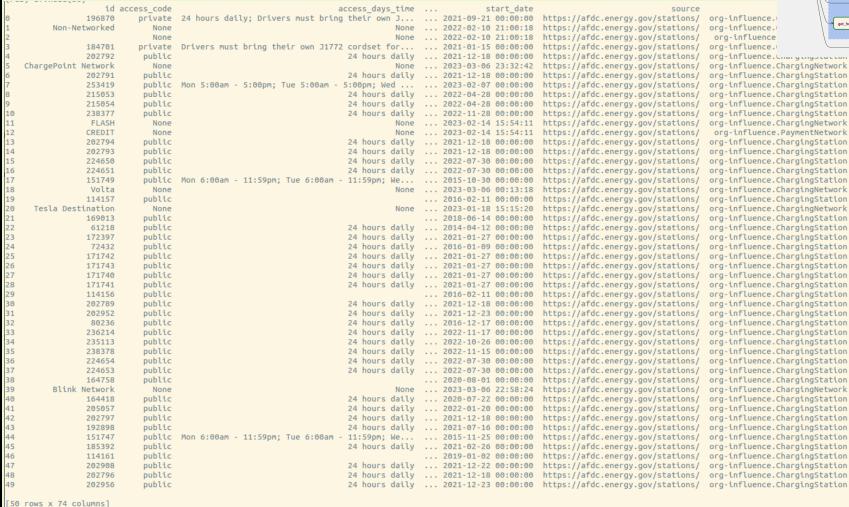


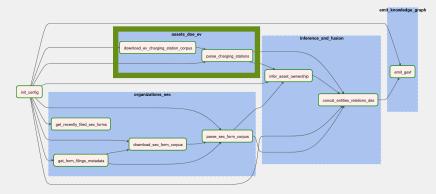


EV Network							
alt-fuel-stations.EVCh	argingStation						
hasName	hasLocation	hasEVNetwork	Start date				
Paul Simon Chicago JCC	Chicago	Non-Networked	2021-09-21				
INTERPARK ADAMWABASH 2	Chicago	ChargePoint Network	2021-12-18				
Grant Park South	Chicago	Tesla Destination	2016-02-11				

Entity Extraction - Implementation

view_entities_relations



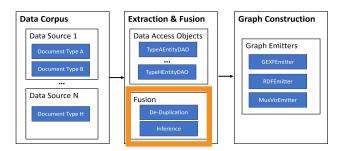


Fusion

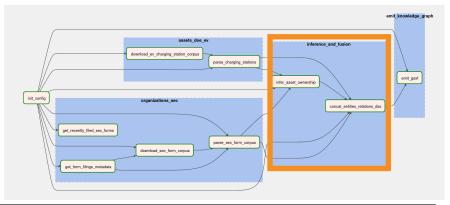
EV Network

alt-fuel-stations.EVChargingStation

hasName	hasLocation	hasEVNetwork	Start date
Paul Simon Chicago JCC	Chicago	Non-Networked	2021-09-21
INTERPARK ADAMWABASH 2	Chicago	ChargePoint Network	2021-12-18
Grant Park South	Chicago	Chicago Tesla Destination	
alt-fuel-stations.EVChar			
ChargePoint Network	None	n/a	n/a
Tesla Destination	None	n/a	n/a
Blink Network	None	n/a	n/a



Fusion - Implementation



V	iew entities	relations					
	, – , , –	•					
	subject_id	relation	object_id	subject_name	start_date		
520	A D M V	isNetworkOf	166665	0	1970-01-01 00:00:00.000000		
519	EV Connect	isNetworkOf	166665	0	1970-01-01 00:00:00.000000		
314	Non-Networked	isNetworkOf	60987	0	1970-01-01 00:00:00.000000		
666	Non-Networked	isNetworkOf	184418		1970-01-01 00:00:00.000000	1 11 32 3 1	
5		acquiresBeneficialInterest			1970-01-01 00:27:11.750400		
6		acquiresBeneficialInterest				build/chicago/1_23_2023/SEC/0001819584/0001068 1970-01-01 00:00:00.000000	
9						build/chicago/1_23_2023/SEC/0001819584/0001213 1970-01-01 00:00:00.000000	
7		acquiresBeneficialInterest				build/chicago/1_23_2023/SEC/0001819584/0001213 1970-01-01 00:00:00.000000	
8		acquiresBeneficialInterest			1970-01-01 00:27:14.688000		
10				3	1970-01-01 00:27:14.688000		
11		acquiresBeneficialInterest		3	1970-01-01 00:27:14.688000		
12		acquiresBeneficialInterest			1970-01-01 00:27:14.688000		
13		acquiresBeneficialInterest		3	1970-01-01 00:27:14.688000		
297	A CREDIT D Debit M V	isNetworkOf	187544		1970-01-01 00:00:00.000000		
443	Non-Networked	isNetworkOf	43358		1970-01-01 00:00:00.000000		
296	Non-Networked	isNetworkOf	187544		1970-01-01 00:00:00.000000		
511	Non-Networked	isNetworkOf	53859		1970-01-01 00:00:00.000000		
232	AMPUP	isNetworkOf	169699		1970-01-01 00:00:00.000000		
15	Non-Networked	isNetworkOf	184701		1970-01-01 00:00:00.000000		
14	Non-Networked	isNetworkOf	196870	Θ	1970-01-01 00:00:00.000000	https://afdc.energy.gov/stations/ 1970-01-01 00:27:24.526818	2022-02-10 21:00:18

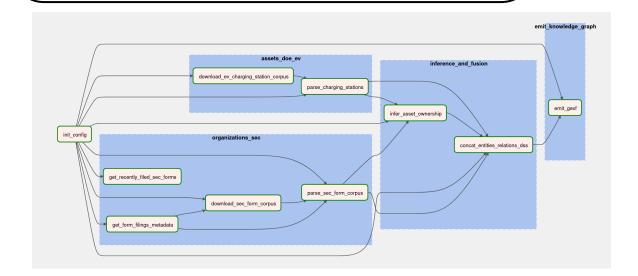
Iterative Discovery

During this analysis of Volta, we discovered the following:

- 1. Identification of all EV assets within a geographic region of interest.
- 2. Several other Organizations that own EV Stations in the region.
 - 000177393
 - 0001318605
 - 001429764
 - 0001530649
- 3. Identification of People/Orgs with a beneficial ownership over Volta
 - CIK: 0001890372
 - CIK: 0001879633

Init Config (n=0)

```
{'edgar.data_dir_path': 'build/chicago/1_23_2023/SEC',
  'evasset.data_dir_path': 'build/chicago/1_23_2023/EVChargingStation',
  'fusion.data_dir_path': 'build/chicago/1_23_2023/fusion',
  'editorial.data_dir_path': './build/chicago/1_23_2023/editorial',
  'edgar.cik_list': ['0001819584'],
  'edgar.form_types': ['SC 13D', 'S-4'],
  'evasset.latitude': '41.8781',
  'evasset.longitude': '-87.6298',
  'evasset.radius': '30.0'}
```



Init Config (n=1)

```
{'edgar.data_dir_path': 'build/chicago/1_23_2023/SEC',
  'edgar.max_cik_form_types_downloaded': 10,
  'evasset.data_dir_path': 'build/chicago/1_23_2023/EVChargingStation',
  'fusion.data_dir_path': 'build/chicago/1_23_2023/fusion',
  'editorial.data_dir_path': './build/chicago/1_23_2023/editorial',
  'edgar.cik_list': ['0001819584',
  '0001777393',
  '0001318605',
  '0001329764',
  '0001530649'],
  'edgar.form_types': ['SC 13D', 'S-4'],
  'evasset.latitude': '41.8781',
  'evasset.longitude': '-87.6298',
  'evasset.radius': '30.0'}
```

Regional Infrastructure Asset Ownership

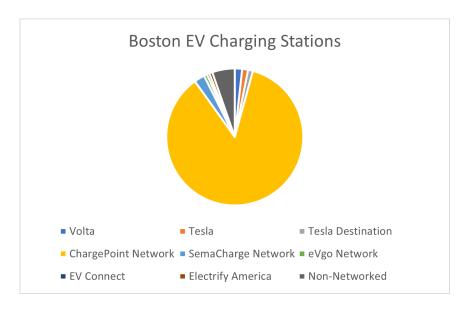
Semantic and Geospatial Analysis

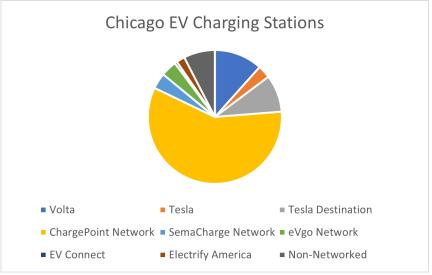
Resource Ownership within a Region

Our intent is for adversarial STNA networks to enable a wide variety of analyses of *organizational influence over physical infrastructure systems*.

Example: Consider the ownership of the communications network upon which EV Charging Stations in a region depend.

- Semantic attributes (Lsemantic) enable us to infer an ownership (owns) relation given the name of the EVChargingNetwork and an Organization.
- The data analysis pipeline enables comparison across different regions.

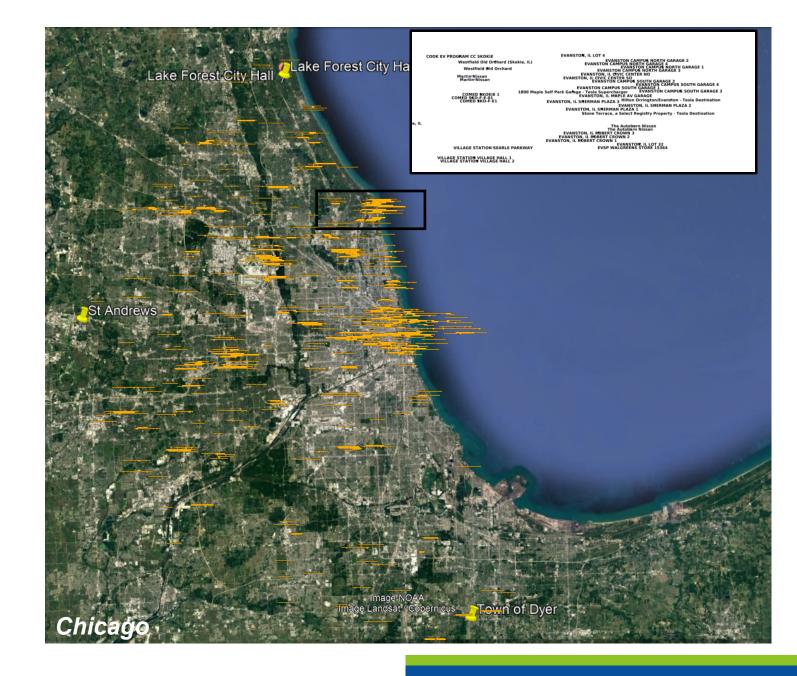




Resource Ownership within a Region

Geospatial Attributes (Lgeospatial) within the Critical Infrastructure layers enable us to consider organizational influence over specific region.

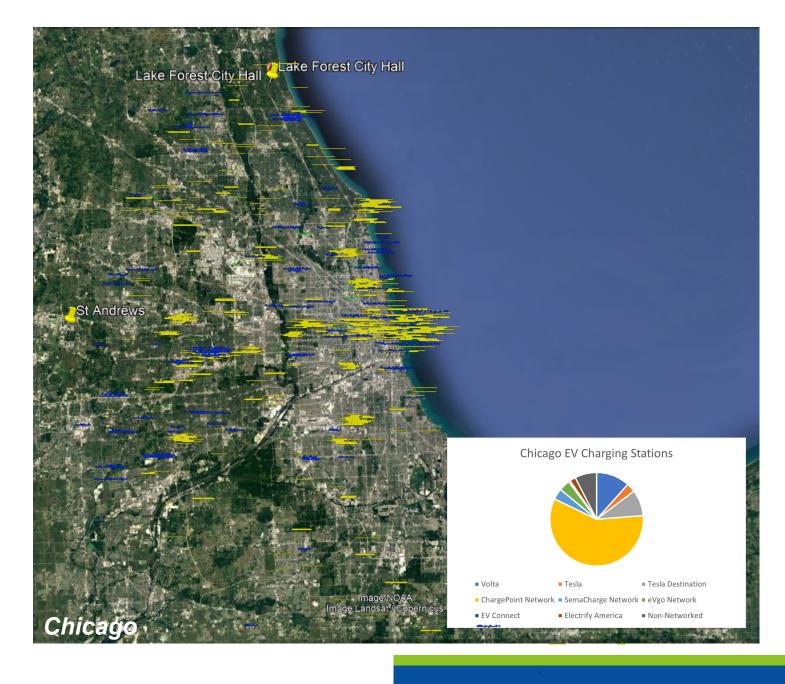
- This provides more detail than a simple pie chart of proportions
- Enables the ability to zoom in on a given region of interest



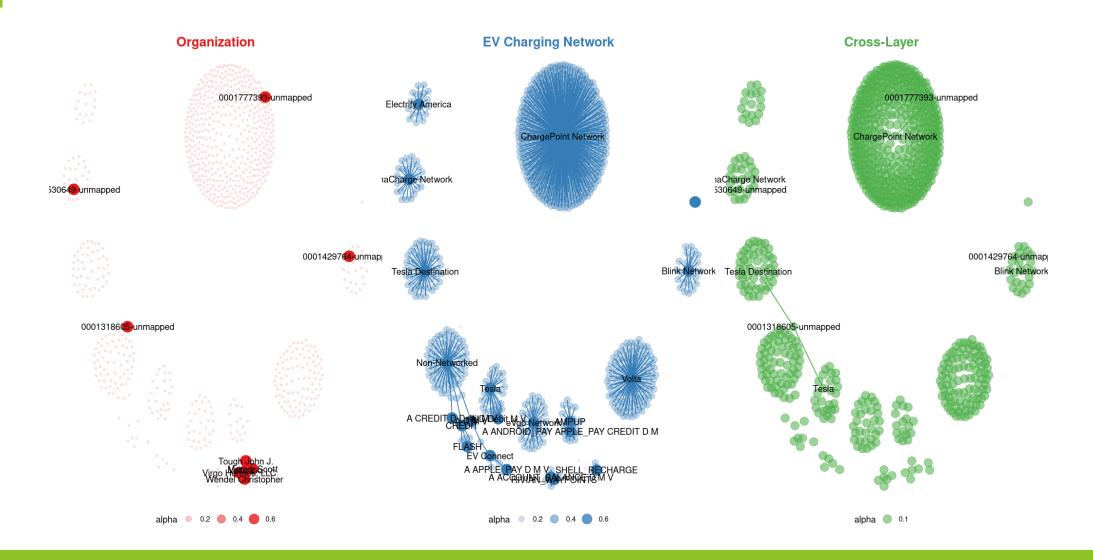
Resource Ownership within a Region

Geospatial Attributes (Lgeospatial) within the Critical Infrastructure layers enable us to consider organizational influence over specific region.

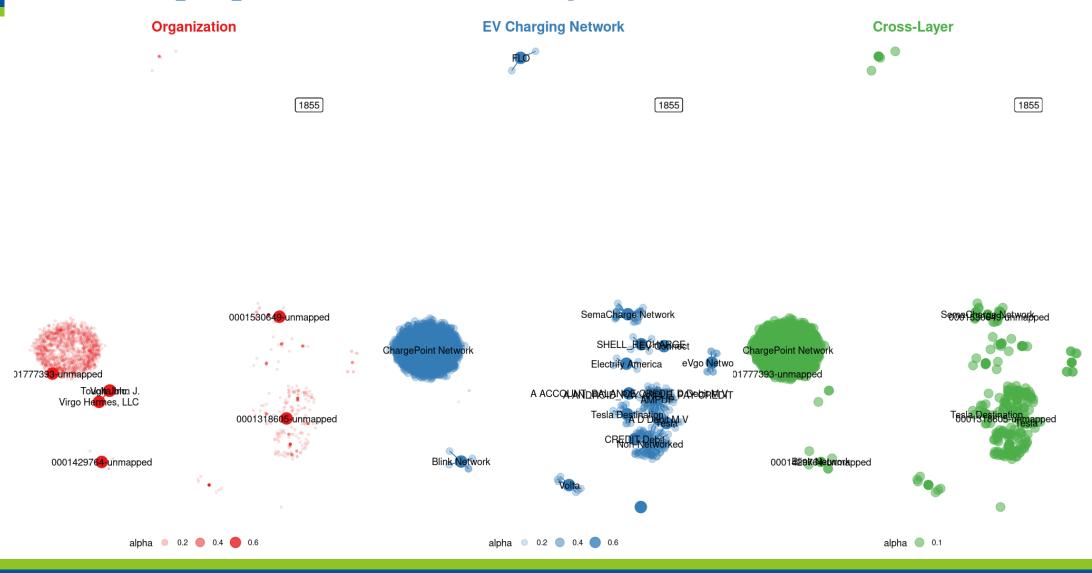
- This provides more detail than a simple pie chart of proportions
- Enables the ability to zoom in on a given region of interest and see who has influence and where
 - ChargePoint
 - Volta
 - eVgo



EV Charging Station Ownership Around Chicago



EV Charging Station Ownership Around Boston



Newport News Shipping

Historical Study, Indirect Data Access via Subsidiary



IT and Networks

Hackers are repeatedly targeting Navy contractors

By Mark Pomerleau

iii Jun 27, 2019









Huntington Ingalls Newport News Shipyard on Aug. 22, 2015. Huntington Ingalls was subject of a sophisticated hacking spree by organs of the Chinese government, according to Reuters. (Mark D.

Huntington Ingalls, the Navy's largest shipbuilder, was compromised by a largescale hacking campaign waged by several organs of the Chinese government, according to a Reuters report.

However, the company denied the allegation in a June 27 email to Fifth Domain, saying, "there was no breach of information" from Newport News Shipyard, nor were their systems connected to a foreign server controlled by a Chinese group, known as APT10

Featured Video



Electric tanks and upgraded targeting drones | Defense News Weekly Full Episode 10.22.22



This smart shirt might save your life on the battlefield

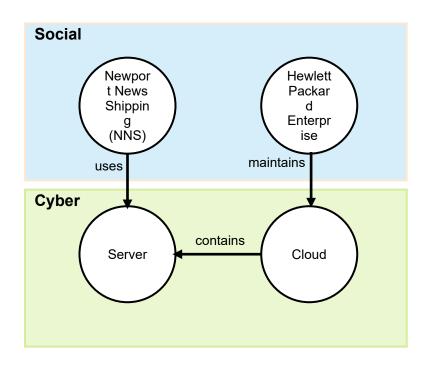


rigorous testing



offerings at AUSA

Newport News Shipping Case Study



Construction of G 1:

- 1. NNS uses a Server
- Cloud contains Server
- 3. HPE/CSC maintains Cloud

Time:

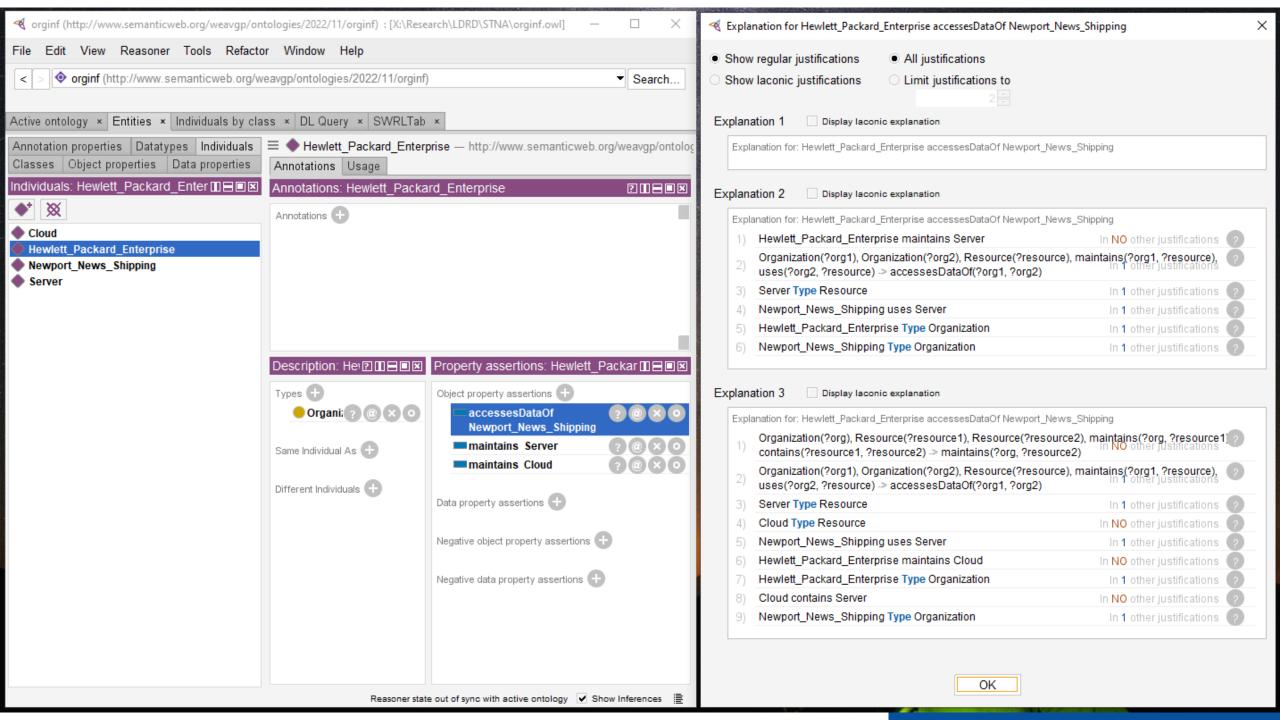
From time t1_1 to time t2_1.

Interence:

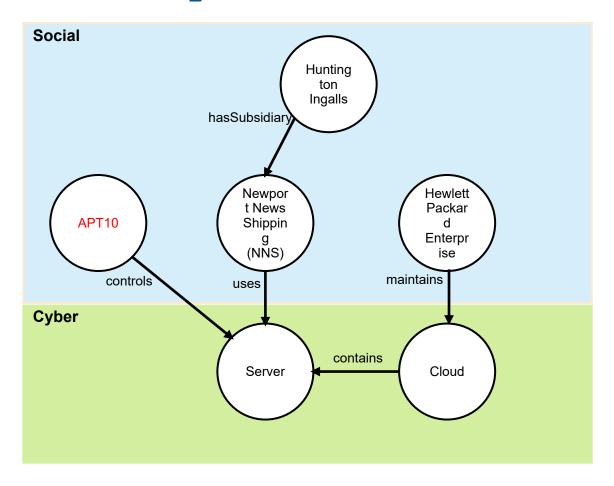
- 1. <org1> maintains <resource1> & <resource1> contains <resource2> → <org1> maintains <resource2>
- 2. <org1> maintains <resource2> & <org2> uses <resource2> □ <org1> accessesDataOf <org2>

Relations Inferred:

- HPE maintains Server [I1]
- HPE accessesDataOf NNS [I2]



Upstream Impacts via Social Network Analysis



Inference Rules:

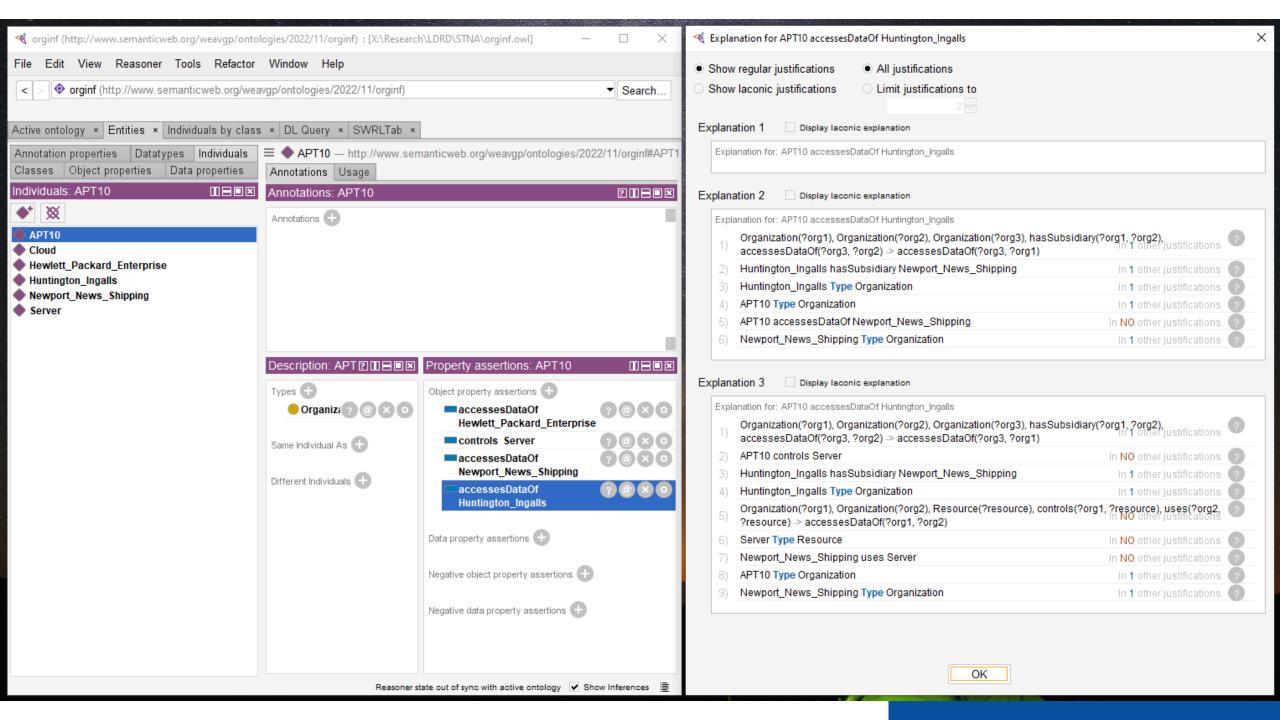
- <org1> maintains <resource1> & <resource1> contains
 <org1> maintains <resource2>
- 2. <org1> maintains <resource2> & <org2> uses <resource2> □ <org1> accessesDataOf <org2>
- 3. <org1> controls <resource2> & <org2> uses <resource2> □ <org1> accessesDataOf <org2>
- 4. <org1> controls <resource2> & <org2> maintains <resource2> □ <org1> accessesDataOf <org2>
- 5. <org3> hasSubsidiary <org2> & <org1> accessesDataOf <org2> □ <org1> accessesDataOf <org3>

Data Sets:

hasSubsidiary via OpenCorporates (?)

Inferred Relations

- HPE maintains Server [I1]
- HPE accessesDataOf NNS [I2]
- APT10 accessesData of NNS [I3]
- APT10 accessesData of HPE [I4]
- APT10 accessesData of HuntingtonIngalls [I5]



Conclusions

The Problem: How do adversarial business practices impact critical infrastructure systems within a given geographic region of interest?

 Specialization of 'cyber' may lead to blindspots for dependencies that achieve influence but are exogenous to traditional system boundaries.

Our Approach: Our research focuses on adversarial techniques in the business domain that affect infrastructure.

- Want to think about how such techniques across domains can be composed by adversaries.
- Support a variety of analyses using semantic, geospatial, and temporal attributes.

	Acquisition	Ownership	Company Ownership Proxy	Due Diligence for Mergers & Acquisitions	Foreign- Owned Facilities
	Use	Board Membership	Board Membershi p Proxy	Legal Subpoena	Disinformation
Social	Leasing	Asset Maintenanc e		Digital Twin Creation	Degrade Service Quality
Cyber	Drive-by Compromise	Hardcoded Credentials	Change Operating Mode	Network Connection Enumeratio n	Unauthorized Command Message



Battelle Energy Alliance manages INL for the U.S. Department of Energy's Office of Nuclear Energy. INL is the nation's center for nuclear energy research and development, and also performs research in each of DOE's strategic goal areas: energy, national security, science and the environment.