

AI-Enabled Social Cyber Maneuver Detection and Creation

Matthew Hicks

Thesis

22 April 2025

Committee

Kathleen M. Carley, Chair

Patrick S. Park

Mohamed Farag

David M. Beskow (United States Military Academy)

Carnegie Mellon University



Outline

- Introduction
- Background/Doctrinal BEND
- AI-Enabled BEND Scenario Development
- Generating Synthetic Data for BEND Training
- BEND-Effects and Improving the BEND Framework
- Conclusions

Social-Cyber Security

- BEND provides a framework for describing information space maneuvers
- BEND allows us to frame:
 - (Source) Who
 - (Payload) Maneuver
 - (Target) On Whom
 - (Effect) Impact

Narrative Maneuvers			Network Maneuvers		
Emotional Messaging	Develop Narrative	Counter Narrative	Individual Centric	Make Groups	Unmake Groups
Impact what is being said and how it is being said			“Community” maneuvers. Alter who is connected to whom, the strength of those connections, and so alters who is influential and what groups exist		
Excite Dismay	Explain Engage Enhance	Distort Dismiss Distract	Back Negate	Build Bridge Boost	Narrow Neutralize Neglect

Major questions and motivation

- Q1: How does the BEND framework fit into current US military doctrine and how can it enhance current information environment analysis?
- Q2: How can we develop exercise training scenarios for the BEND framework?
- Q3: How can we create synthetic data w/ BEND maneuvers to support a training scenario?
- Q4: How can we detect the presence of BEND maneuvers through their effects?
Can we tie these effects to broader social media campaigns?

Purpose

The purpose of this thesis is to:

- ❑ Develop applications and methods for creating social media training scenarios that are realistic by leveraging generative AI
- ❑ Develop applications and methods for creating the synthetic social media datasets through the use of generative AI that match a tailored scenario while mirroring real world data
- ❑ Develop methods for detecting the effects of BEND maneuver within topic groups and connect these effects to broader campaigns



AESOP



SynTel/X



BEND-effects

Doctrinal BEND

Prior Work

- ❑ Blaine (CMU-S3D-23-102) looked at and compared BEND to four other frameworks.
- ❑ Compared BEND to four other information operations frameworks:
 - ❑ Ben Nimmo: 4Ds
 - ❑ ABCDE Framework
 - ❑ Camille Francois: ABC Framework
 - ❑ Alexandre Alaphilippe: ABCD Framework
 - ❑ Pamment: ABCDE Framework
 - ❑ Blazek: SCOTCH Framework
 - ❑ DISARM Foundation: DISARM Framework
- ❑ BEND's strength is in its shorthand

Table 1.1: Comparison of informations operations frameworks

Characteristics	4 D's	ABC(D)(E)	SCOTCH	DISARM	BEND
General high-level approach	x	x	x	x	x
Specific systematic methodology					x
Key Actor Analysis		x	x		x
Behavior Analysis	x	x		x	x
Quantitative Analysis					x
Limited to Disinformation Analysis	x	x		x	
Influence campaign analysis/assessment			x		x
Develops recommendations		x		x	x
References	[98]	[3, 57, 101]	[22]	[50]	[17, 38]

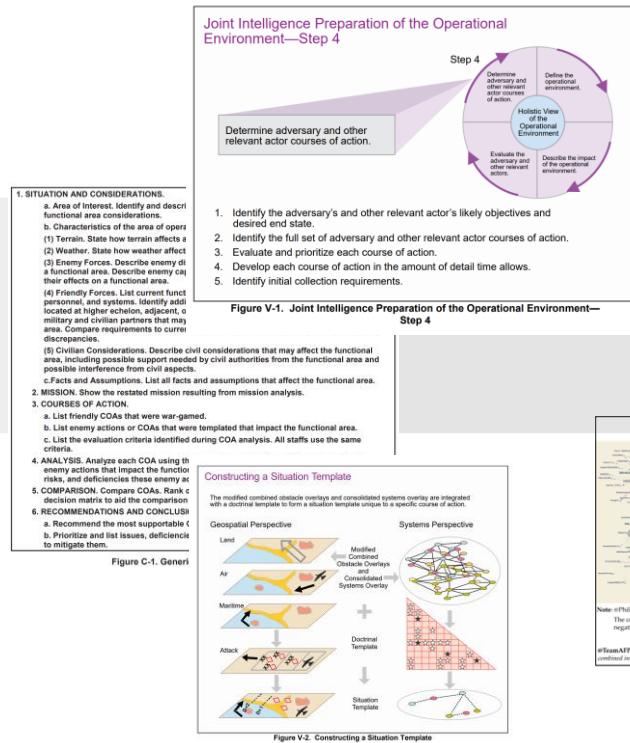
How does the BEND framework fit into current military doctrine?

Current References

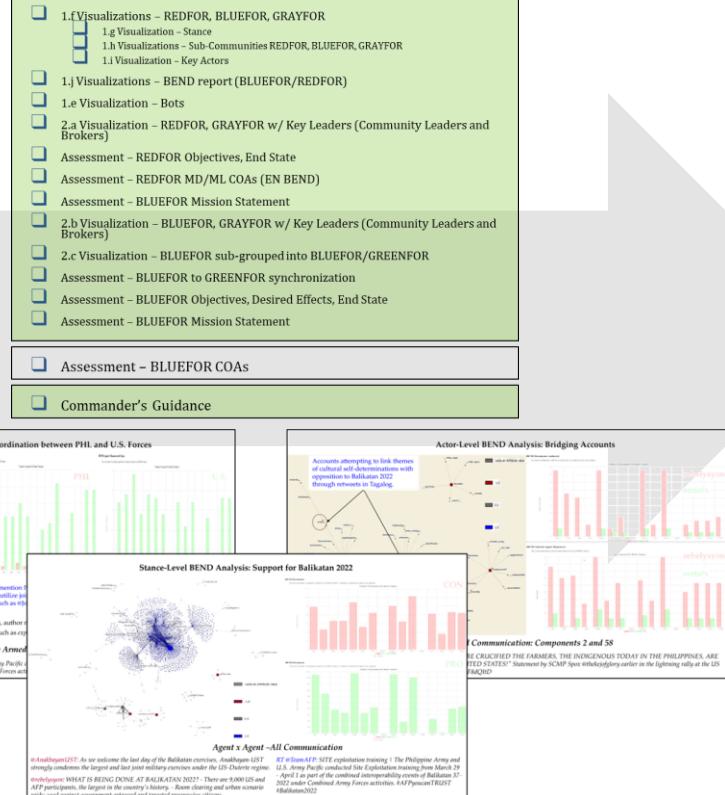
- JP 3-04 Information in Joint Operations
- JP 3-13 Information Operations
- NWP 3-13 Navy Information Operations
- FM 3-13 Information Operations
- AFDP 3-13 Information in Air Operations
- CJCSI 3210.01C Joint Information Operations Proponent
- DODI 3600.01 Information Operations
- ADP 5-0 The Operations Process
- MCWP 3040.4 Marine Air-Ground Task Force Information Operations
- JP 2-01.3 Joint Intelligence Preparation of the Operational Environment
- ATP 2-01.3 Intelligence Preparation of the Battlefield
- JP 3-60 Joint Targeting

:

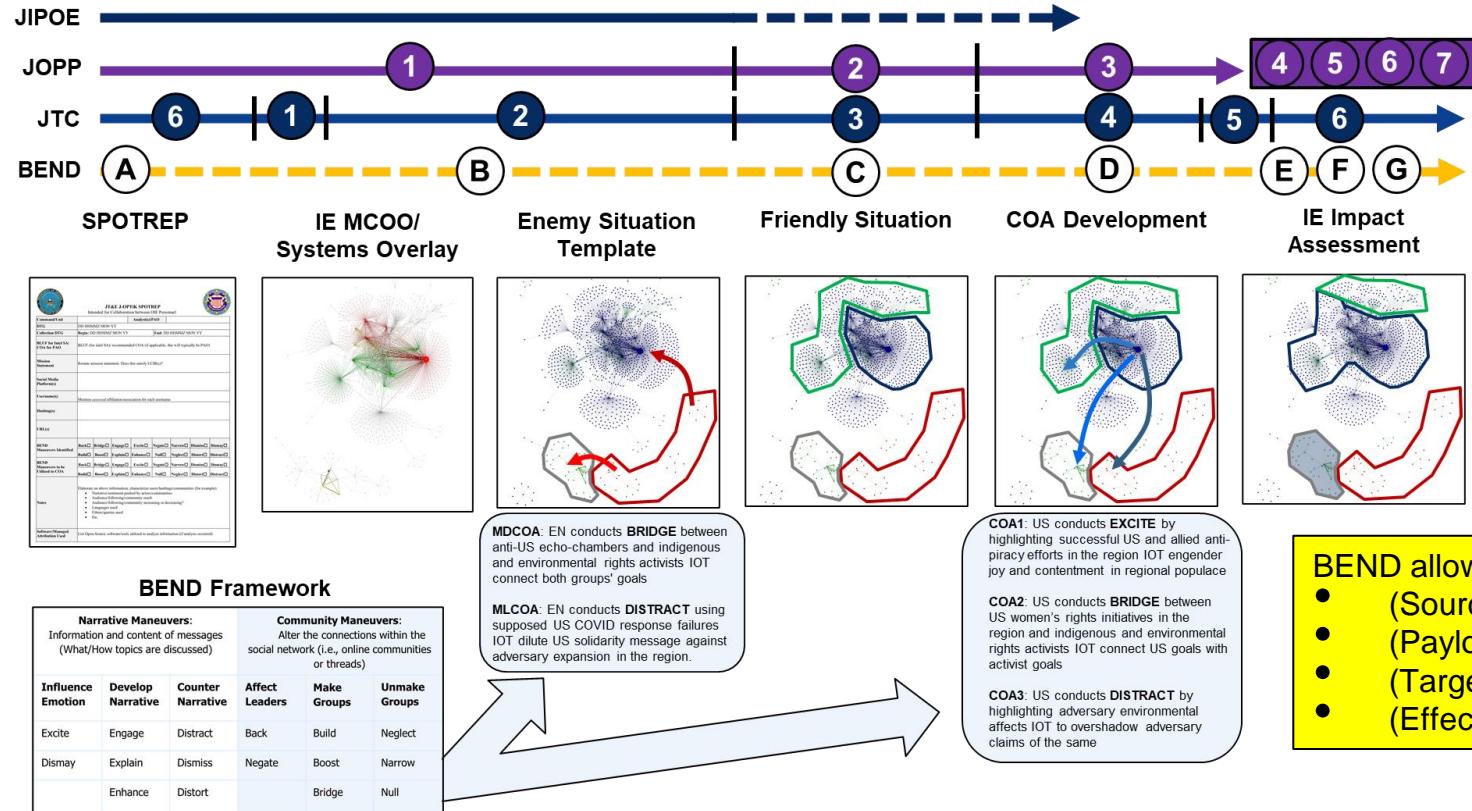
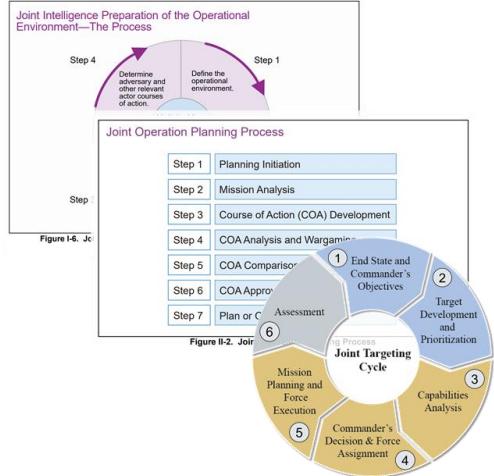
Current Products



New References and Products



How do we use BEND with current doctrine?



- A) SPOTREP of adversarial action on social media
- B) J2/OSINT updates EN SITEMP (**JIPOE, JOPP 1, JTC 2**)
- C) PAO evaluates CDR's objectives, desired effects, and endstate (**JOPP 2, JTC 1**)
- D) JFECC develops COAs for CDR (**JOPP 3, JTC 4/5**)
- E) CDR approves COA (**JOPP 6, JTC 5**)
- F) COA Execution by PAO (**JTC 6**)
- G) Impact Assessment by J2/OSINT (**JIPOE, JOPP1, JTC 0/1**)

JIPOE – Joint Intelligence Preparation of the Operational Environment (JIPOE)
JOPP – Joint Operation Planning Process (JOPP)
J-OPTIK – Joint Operations in the Information Environment Playbook Toolkit QRT
JTC – Joint Targeting Cycle

BEND allows us to frame:

- (Source) Who
- (Payload) Maneuver
- (Target) On Whom
- (Effect) Impact

A Synthetic Training Requirement

Why train?

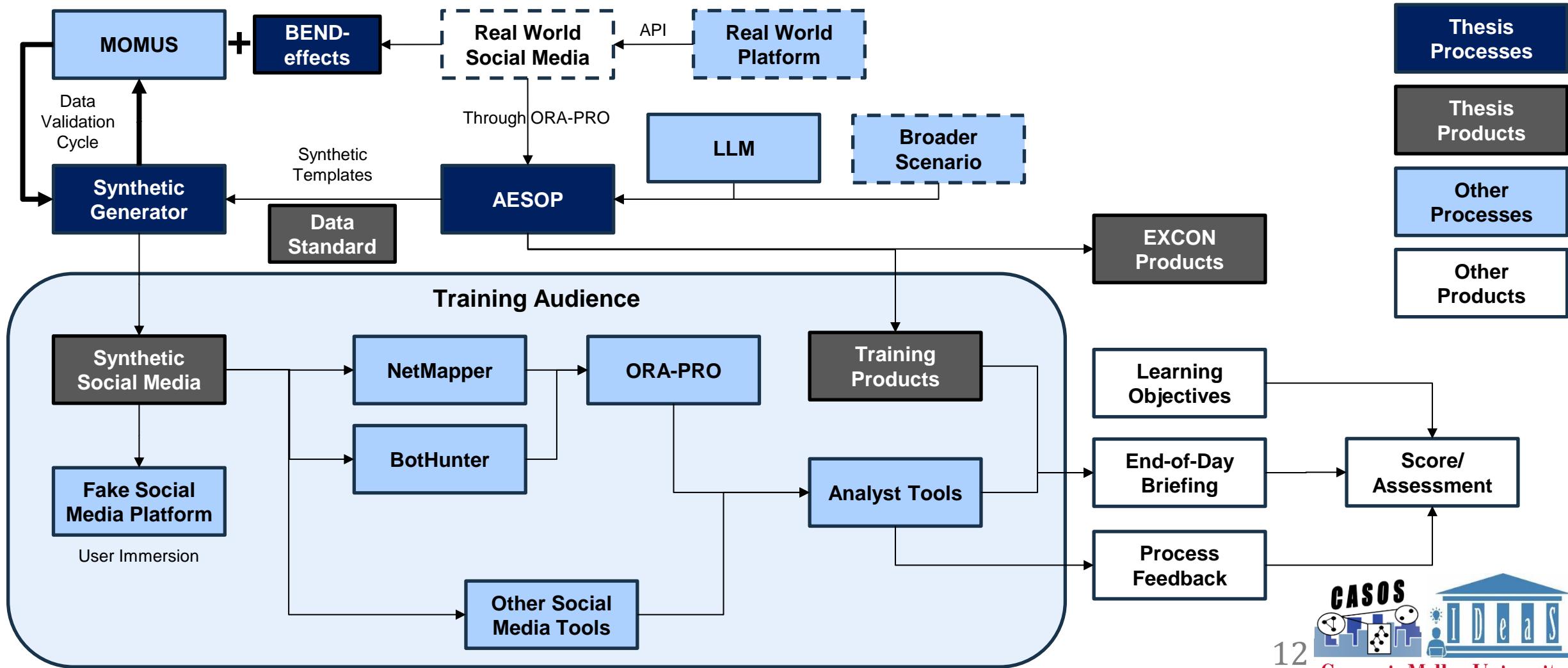
- ❑ **Classroom instruction is insufficient** – “train as we fight” – there is a real military requirement for training
- ❑ **Non-military requirement** – this requirement exists outside of the military and the tools and synthetic data developed can be used there as well

Why synthetic?

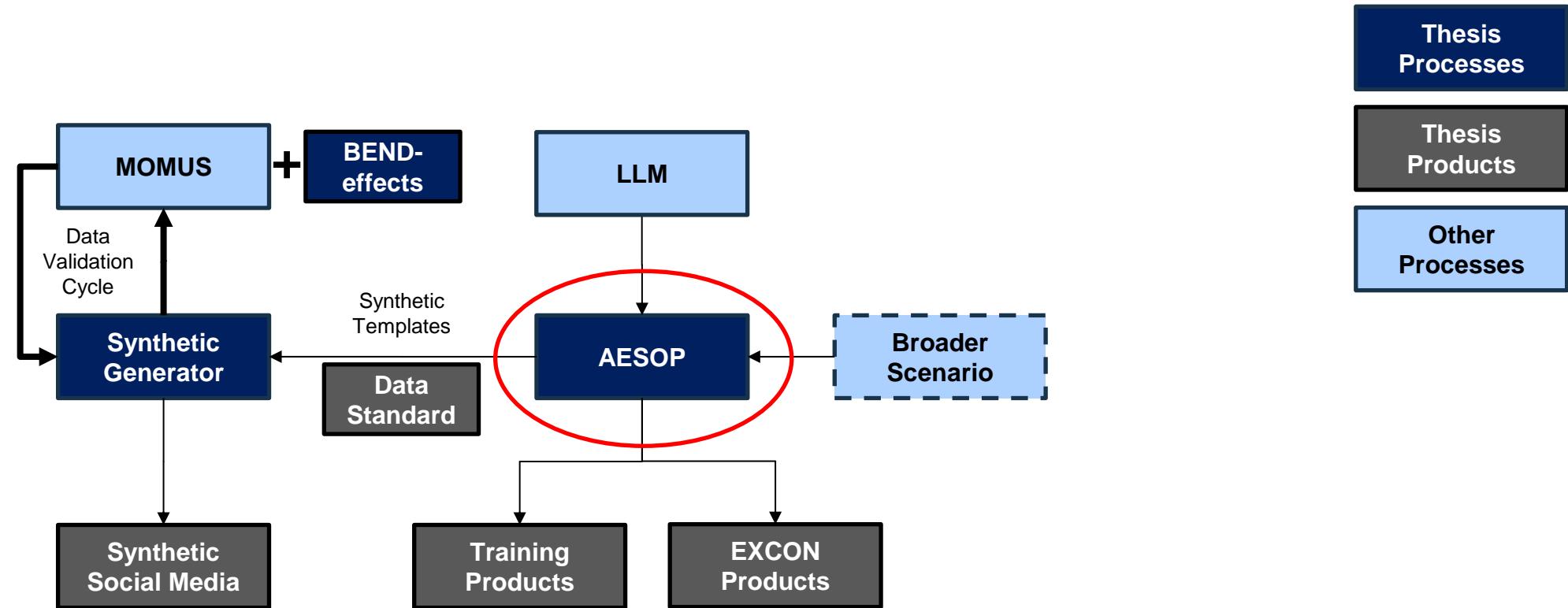
- ❑ **Pedagogical** – the training is tailored to the learning objective and is populated with the right content for the training audience to analyze
- ❑ **Ethical/moral/legal** – synthetic data alleviates privacy concerns, security concerns, and attribution/titling concerns
- ❑ **Interactive** – the training audience can see the fruits of their analysis in the data as their recommendations are executed

AI-Enabled BEND Scenario Development

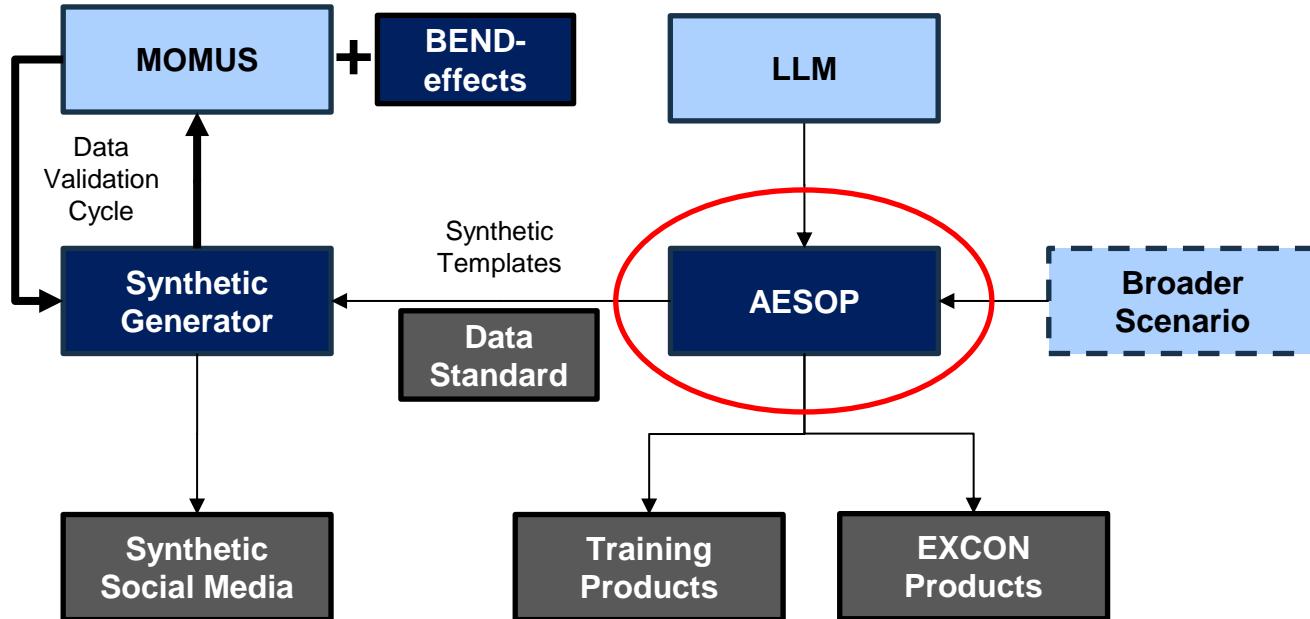
Realistic Training



Realistic Training



AI-Enabled Scenario Orchestration and Planning (AESOP)



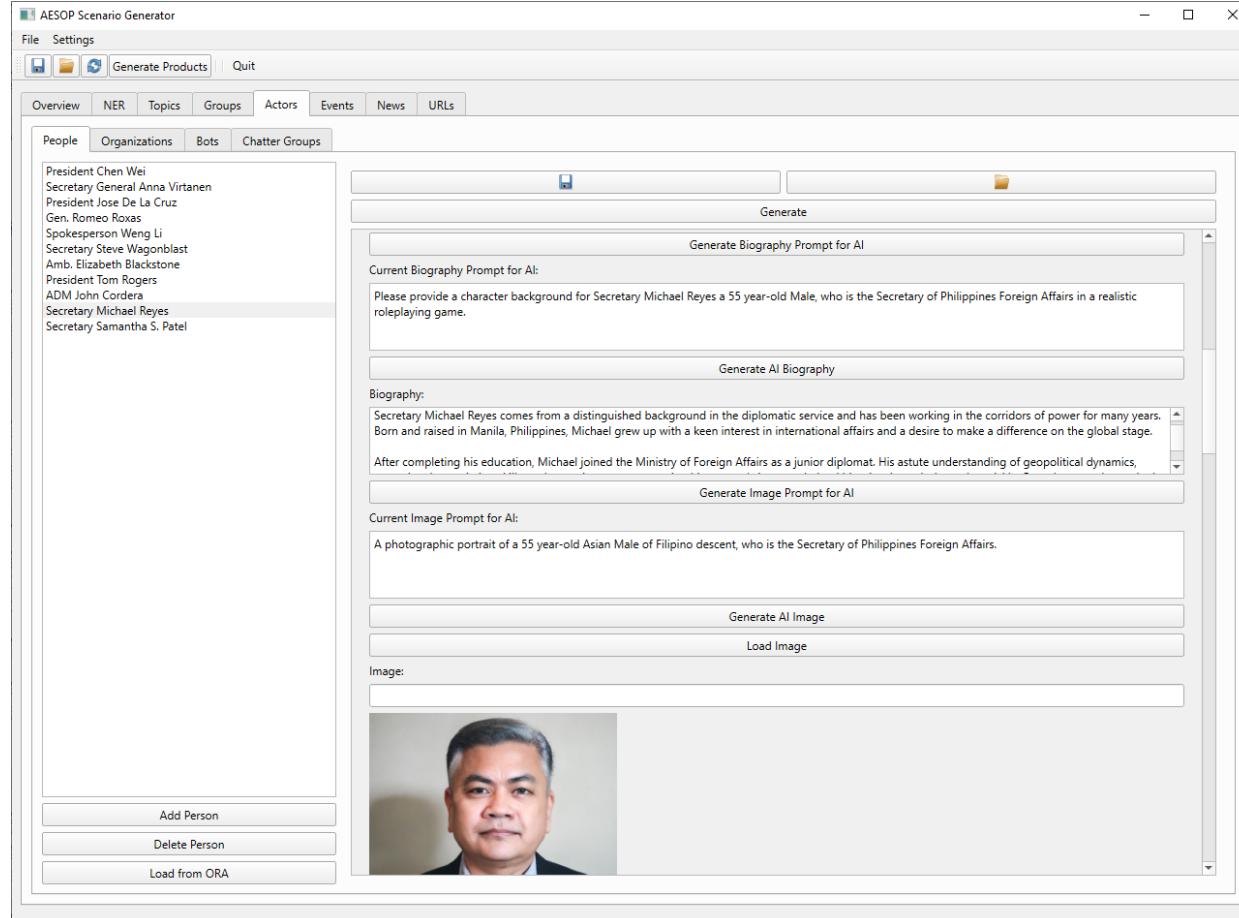
AESOP allows Information Environment planners to develop social-cyber exercise scenarios from scratch or develop social-cyber vignettes for integration with existing scenarios.

It leverages large language model (LLM) artificial intelligence (AI) to reduce planner load and increase realism and immersion for the training audience.

Planners complete basic fields – such as date ranges and summaries – and a configured external LLM is used to generate surrounding details.

AESOP reduces the time required for social media scenario design from months to days

AI-Enabled Scenario Orchestration and Planning (AESOP)



The Scenario Generator provides scaffolding for developing a range of templates to enable synthetic data generation: Actors, Topics, Groups, Events, News, URLs, etc.

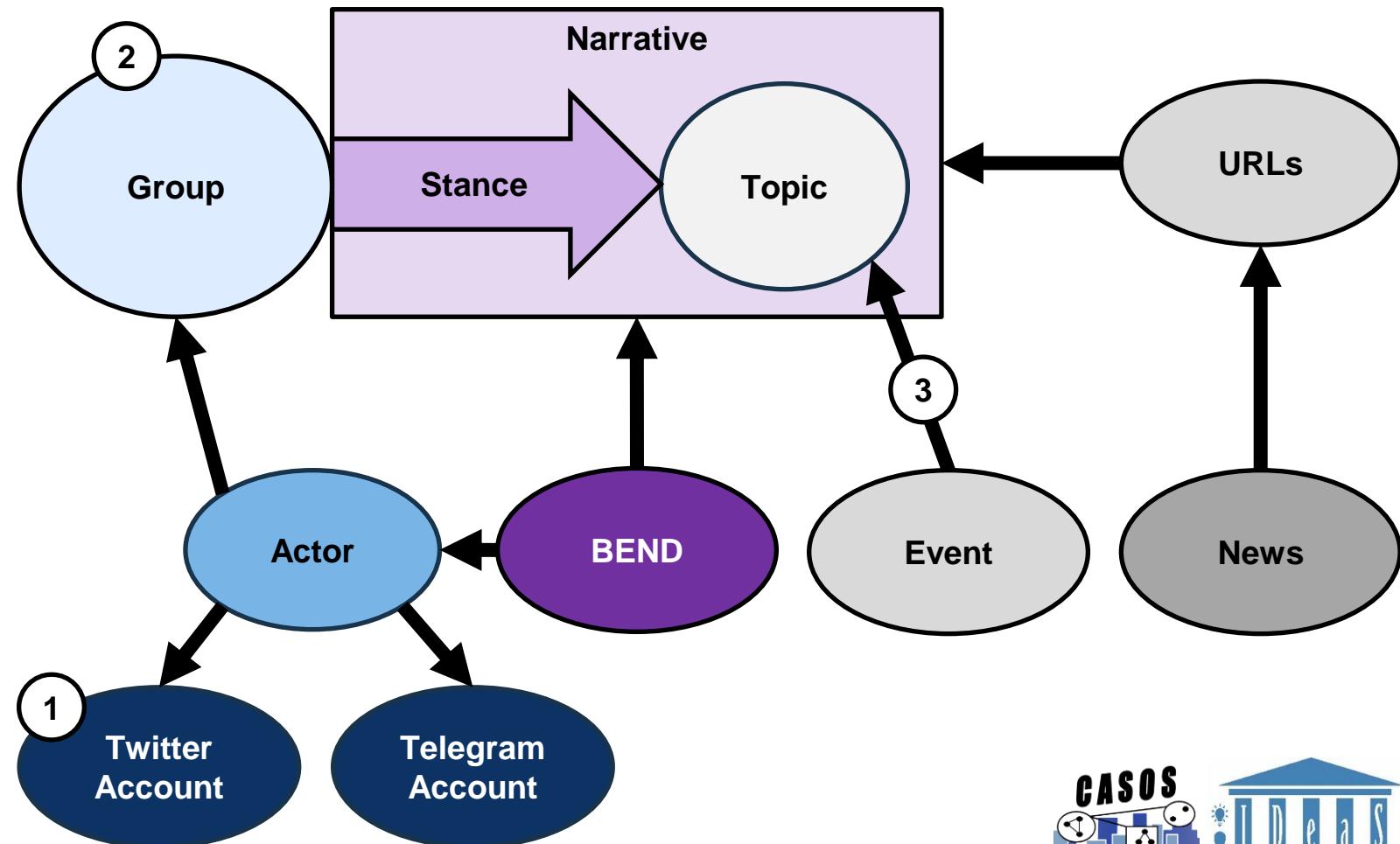
As planners complete basic data fields, the Scenario Generator develops an engineered prompt for the LLM.

Planners can make additional changes to the prompt as required.

Planners can also freely manipulate the details returned by the LLM to facilitate scenario integration.

AESOP Data Structure

- 1 We are specifying Twitter Account information
- 2 We can also specify a Group by ensuring the owning actor for an account is a member of it
Group membership is split in two:
 - a) Full members – full members are eligible to be retweeted by other full members and are also eligible to originally tweet about Group associated Narratives
 - b) Source members – source members are eligible to be retweeted by full members but will never originally tweet about Group associated Narratives
- 3 Groups have associated Narratives that they espouse (which include BEND maneuver likelihood). These Narratives, in-turn, are associated with Events. Events will cause spikes in Narrative tweets, etc.



AESOP Outputs

Event Cards

USS Carlisle Collision



Description

On April 1st, 2045, the US Navy warship, USS Carlisle, was involved in a disastrous collision with the Liberian-flagged tanker Casonee MC. The collision happened east of the Strait of Malacca, off the coast of Singapore and Malaysia. The impact caused flooding in nearby compartments, including crew berthing, machinery, and communications rooms. Unfortunately, the collision resulted in the death of ten US Navy sailors.

Upon learning of the incident, The Maritime and Port Authority (MPA) of Singapore quickly initiated a search and rescue operation, which included various agencies from Singapore's Maritime Search and Rescue Region (MSRR). Despite the efforts of these agencies, the casualties of the incident included the ten US Navy personnel.

It is currently unknown exactly what led to the unfortunate crash, however, an investigation is being conducted to determine the specifics of what happened. In the meantime, the families of the ten victims are left mourning their loss. The incident serves as a painful reminder of the dangers of naval operations and the consequences they can bring.

Since the accident, the MPA has recommended for ships to take extra precaution when traversing the area where the collision occurred. The US Navy is also now undertaking a more detailed review of its training and safety protocols in hopes to prevent any further tragedies from taking place.

The USS Carlisle collision was a devastating event, leaving behind damage, both physical and emotional, that will not soon be forgotten. The victims and families affected by the tragedy are in our thoughts and prayers.

General Info

Type: Collision/Accident
Which Objects Involved: USS Carlisle, tanker
Who Owns Objects: US, Liberia
Name of Objects: USS Carlisle, Casonee MC
Start Date: 2045-04-02
End Date: 2045-04-02
Other Countries Involved: Liberia, Singapore, United States
Cities:
Other Areas:
Event Leader:
Leader's Job Title: Admiral Johnson
Leader's Home Country: United States
Positive Hashtags:
Negative Hashtags:

Fragmentation Orders

EXERCISE // UNCLASSIFIED // EXERCISE

DEPARTMENT OF THE ARMY
UNITED STATES INDO-PACIFIC COMMAND
CAMP H M SMITH, HAWAII 96861-4028

INDOPACOM TASK ORDER

SUBJECT: Social Media Analysis of Collision Between USS Carlisle and Casonee MC in Strait of Malacca

SUMMARY: This task order seeks to build an understanding of global reactions to the collision between the USS Carlisle and the Casonee MC in the Strait of Malacca, with a focus on relevant social media platforms.

TASKS:

1. Perform comprehensive research into the incident, including an analysis of multiple social media platforms.
2. Develop a detailed timeline of key developments and significant reactions to the accident.
3. Analyze the sentiment of key messages in terms of diverse cultures and regions.
4. Create a list of key players, influencers, organizations, and individuals with the most engagement on the issue.
5. Identify key trends, both positive and negative, in the responses from each group.
6. Summarize the responses into a compiled report, including a comparison of regional reactions.

COMMANDER'S GUIDANCE: The results of this analysis should provide a clear understanding of global sentiment surrounding the tragic incident and serve to inform future policies and protocols to ensure maritime safety and reduce the potential for similar accidents in the future.

EXERCISE // UNCLASSIFIED // EXERCISE

News Articles and Sites

Sputnik International

THIS IS A FAKE WEBSITE CREATED FOR AN EXERCISE INVOLVING DISINFORMATION. ↗

Home Weather Contact

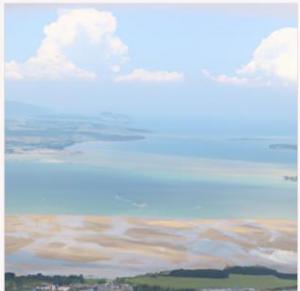


Philippines, US to Hold Balikatan from October 1-4

2027-10-05

In a move that reinforces the military alliance between the Philippines and the United States, both countries announced plans to conduct joint maritime drills in the South China Sea from October 2-13, sparking concerns among regional players. The annual Balikatan joint exercise is expected to highlight the Philippines' naval surface capabilities and demonstrate camaraderie between the two nations. Through these joint exercises, the Philippine forces aim to sharpen their skills in various key areas, including anti-submarine warfare, anti-surface warfare, anti-air warfare, and electronic warfare.

The decision to hold these exercises comes at a time of heightened tensions in the region. China, which claims a large portion of the South China Sea, has criticized the Philippines' actions as provocative and activities that may provoke further disputes. Chinese Foreign Ministry spokesperson Wang Wenbin conveyed this message to the Philippines, urging them to respect the sovereignty and territorial integrity of all countries involved. Notably, recent reports allege that the Philippines



Manila indecisive towards necessary regional partners

2027-10-06 Simon Fantanier

The South China Sea disputes between China and the Philippines have been a longstanding issue. While it is crucial to resolve these territorial disagreements, it is only a small part of the overall China-Philippines relations. The management of sovereignty disputes involves various strategies and beyond the contentious islands and reefs, the relationship between China and the Philippines is a more holistic and significant one. China is the Philippines' largest trading partner, fostering a deep connection through cultural exchanges. These common interests, while outweighing their differences, also highlight the need for an informed and purposeful approach to the current situation in the South China Sea.

It is important to understand that the current state of affairs can be attributed to Marcos' pro-US policy rather than being the root cause itself. Following the departure of former president Rodrigo Duterte, there has been a resurgence of pro-US forces within the Philippines. While these forces had somewhat diminished during Duterte's presidency, they

AESOP Outputs

Scenario Description

Balikatan 2045

General Info

Date Range: 2045-04-01 to 2045-05-01

Countries of Interest:

Philippines, United States

Regions of Interest:

South-eastern Asia

Other Locations of Interest:

Scenario Description

Balikatan 2045 was an exercise run by the Philippines and United States that took place from April 1 to May 1 in 2045. Aimed at developing and preserving the security relationship between the two countries, participants were trained in crisis action planning, counter-terrorism operations and increasing the interoperability and coordination of their forces.

However, the exercise was overshadowed by a severe health crisis in the region, most notably the outbreak of the Great Pox. This highly infectious disease was first discovered in 2045 and its rapid spread to countries such as the Philippines, Australia, China, Japan, Thailand, and Malaysia has had a huge impact on the medical community. Symptoms of the virus include fever, chills, nausea, diarrhea, muscle aches and more, and with no vaccine currently available, preventive measures are the best way to contain the spread.

Tragedy was also faced during Balikatan 2045, as the USS Carlisle collided with the Liberian-flagged tanker Casonee MC in the Strait of Malacca. The US Navy warship was conducting routine operations when the two ships came into contact, causing flooding in various compartments of the destroyer. Ten U.S. Navy sailors were killed in the incident, and rescue efforts were launched by various agencies, including the Singapore Coast Guard, Royal Malaysian Navy, and the U.S. Navy. Unfortunately, no further survivors were located and the cause of the collision is still under investigation.

Finally, Balikatan 2045 was marked by Typhoon Hayman. This Climate Event resulted in extensive damage to vital infrastructure in the Philippines, leaving thousands without food, water, or medical care. In response, the Red Cross activated their Cluster System, consisting of a collection of different humanitarian organizations. The U.S. Navy, along with other governments and NGOs, began sending relief to the hardest hit areas in the form of food, medical supplies and personnel. Volunteers were organized to distribute blankets and

Actor Information Cards

Chandu Claver

General Info



Name: Chandu Claver
Leader Type: Political
Title: Leader
Organization: Cordillera's People Alliance
Gender: Male
Age: 56
Race: Asian
Nationality: Filipino

Biography

Chandu Claver is a 56 year-old man from Cordillera, a rural region that has been generally overlooked by the rest of the country. Having grown up in this hamlet among its close-knit community, Chandu has grown to be considered a leader among the people of Cordillera, embodying their values and fighting for their rights.

Since a young age, Chandu had been advocating for the view point of the people living in his disadvantaged region. He was at the forefront of the struggle in obtaining basic services like education and healthcare. Over the years, Chandu has built a strong network of organizations that cooperate to fight for the rights of the impoverished Cordillera people. He eventually gained the title of the leader of Cordillera's People Alliance.

Dedicated to his cause, Chandu puts a lot of energy into ensuring that the voices of the people of Cordillera are heard, and their needs are met. He has made several appearances at regional and national level political events to voice the issues faced by his people, and has organized several protest marches to support his cause.

Chandu is a loving husband and father, and is described as a determined yet gentle person who always puts the needs of his people first. He is a firm believer in nonviolent resistance and community organizing, and often works with others to come up with creative solutions to the pressures his people face. Despite his age, Chandu remains tireless in his work, and is seen throughout Cordillera as an inspirational leader.

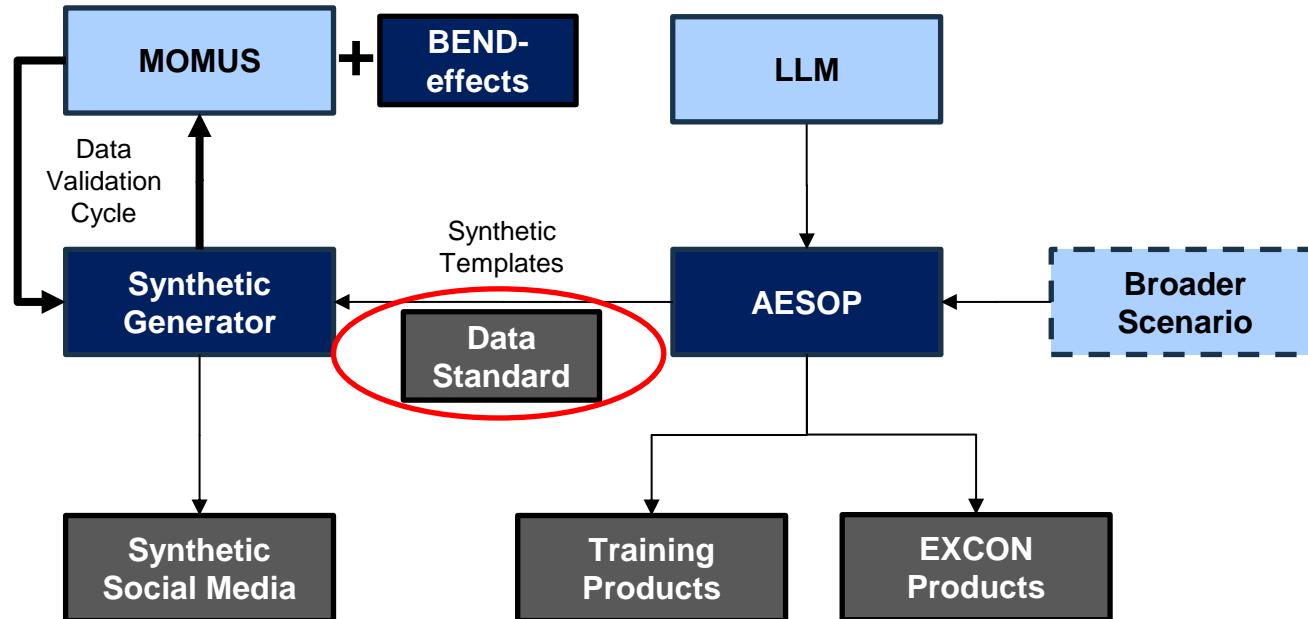
Synthetic Data Templates Per Actor Per Platform

```
{
  "Name": "Chandu Claver",
  "Type": "Actor",
  "Title": "Leader",
  "Lead": "Cordillera's People Alliance",
  "Age": 56,
  "Gender": "Male",
  "Race": "Asian",
  "Nationality": "Filipino",
  "Real Person": false,
  "Background": "None",
  "Bio Prompt": "Please provide a character background for Chandu Claver, a 56 year-old Male, who is the Leader of Cordillera's People Alliance. He is a 56 year-old man from Cordillera, a rural region that has been generally overlooked by the rest of the country and fighting for their rights. When he was a young boy, Chandu had been advocating for the view point of the people living in his disadvantaged region. He has made several appearances at regional and national level political events to voice the issues faced by his people. He has always put the needs of his people first. He is a firm believer in nonviolent resistance and community organizing, and often works with others to come up with creative solutions to the pressures his people face. Despite his age, Chandu remains tireless in his work, and is seen throughout Cordillera as an inspirational leader.",
  "Image Prompt": "A photographic portrait of a 56 year-old Asian Male of Filipino descent, who is the Leader of Cordillera's People Alliance. The image shows him from the chest up, wearing a red and white plaid shirt, looking directly at the camera with a serious expression. He has short, dark hair and is wearing glasses. The background is dark and out of focus, making him stand out. The lighting is soft, highlighting his features and the texture of his shirt. The overall composition is professional and captures his character and leadership qualities.",

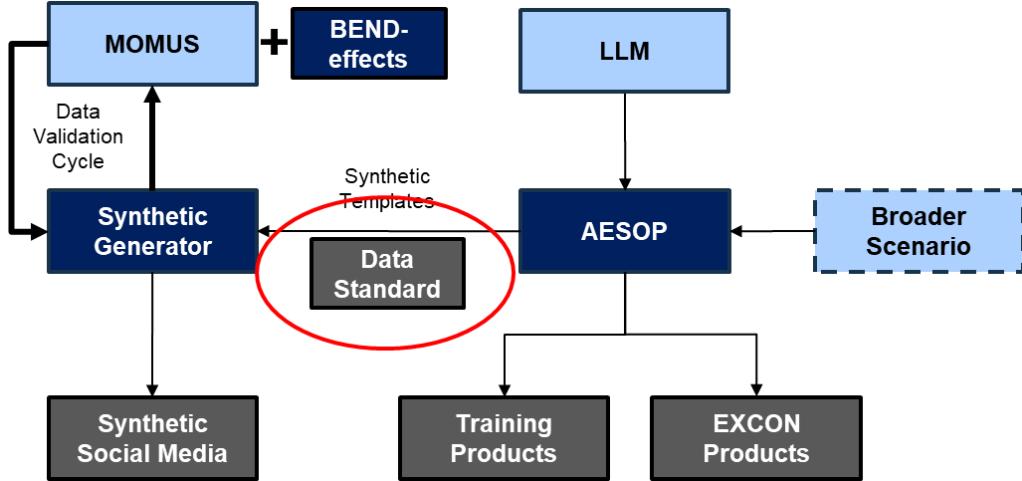
  "Description/Purpose": "Actor Template for Chandu Claver, Leader of Cordillera's People Alliance. This template is used to generate synthetic data for Chandu Claver across various platforms, including social media and news articles. It includes his name, title, organization, gender, race, nationality, and a detailed bio prompt. The bio prompt describes his character as a determined and gentle leader who fights for the rights of his people and uses nonviolent resistance. The image prompt provides a photographic portrait of Chandu Claver, showing him as a 56-year-old Asian male of Filipino descent, wearing a red and white plaid shirt and glasses, looking directly at the camera with a serious expression. The description/purpose also specifies that the template is for the actor role and includes the bio prompt and image prompt.",

  "User Active Start Date": "2023-07-02T00:00:00Z",
  "User Active End Date": "2023-09-02T23:59:59Z",
  "Features": [
    {
      "Feature": "True"
    }
  ],
  "Tweet Distribution": "Day 1(10%), Day 2(10%), Day 3(10%), Day 4(40%), Day 5(30%)",
  "How to Create Screen Name(s)": "Random String",
  "Screen Name(s)": "",
  "Account Bio": "",
  "Account Description": "",
  "Account URL": "",
  "Account Creation Date": "2023-08-02T00:00:00Z",
  "Number of Followers": "100000",
  "Number Following": "10000",
  "Number of Retweets/Retweets/Day": "10000",
  "Number of Tweets": "10000.000",
  "Number of Original Tweets/Day": "10000.000",
  "Number of Quotations": "5",
  "Top Topics": "",
  "Number of Mentions/Tweet": "0",
  "Accounts to Mention": "",
  "Can Other Accounts be Mentioned": "Yes",
  "Number of Retweets/Day": "10",
  "Number of Quotations/Day": "4",
  "Number of Retweets/Day": "10",
  "Top Hashtags": "Mostly Positive",
  "Categories of Accounts this Account Retweets/Quotates": "News (60%); Gov (30%); Health (7%); Other (3%)",
  "Can an Account be Retweeted/Quoted": "Yes",
  "Number of Accounts by Retweeted/Quoted": "Yes",
  "Number of Hashtags/Day": "0",
  "Top HashTags": "",
  "Can Other Hashtags be Used": "Yes",
  "Top Words": "",
  "Can Other Words Be Used": "Yes",
  "When Does This Account Tweet (in GMT)": "20:00:00",
  "Percent of Tweets/Retweets that are Positive, Negative, and Neutral Sentiment": "Positive (50%), Negative (25%), Neutral (25%)"
}
```

Realistic Training



Data Standard



Component	Status
Data Type	
Identifiers	
Vocabulary	Draft Only
Schema	JSON Schema
Format	JSON
Tested	

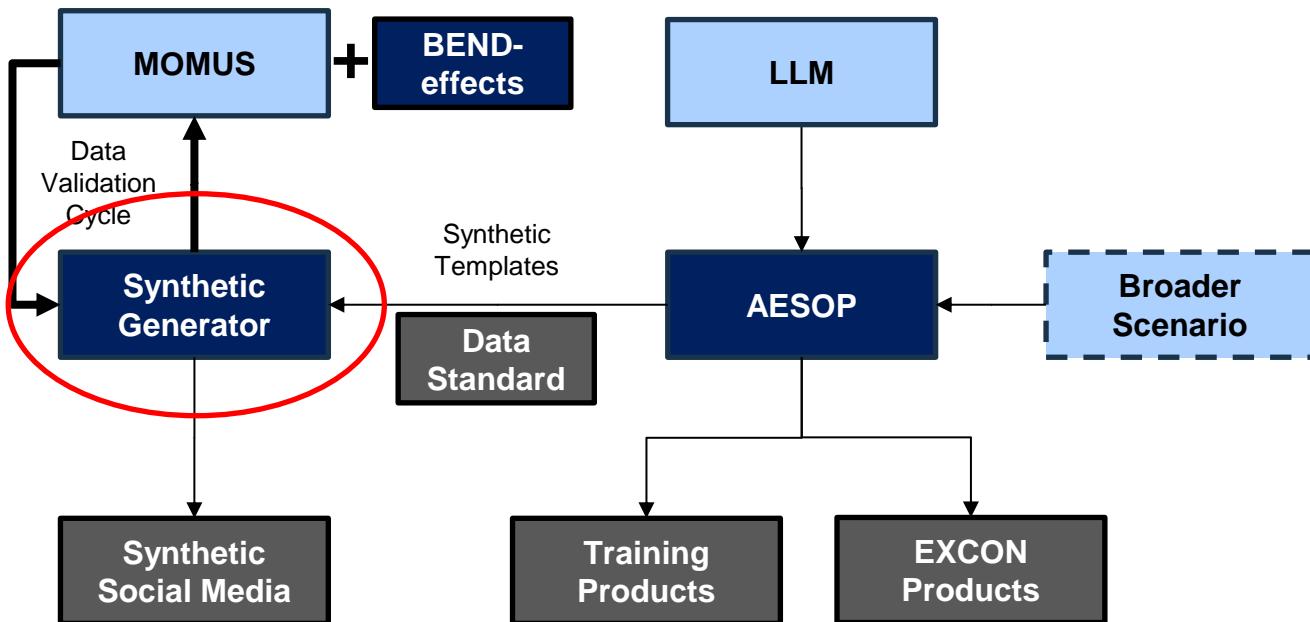
```
{
  "Name": "Chandu Claver",
  "Type": "Political",
  "Role": "Leader",
  "Leads": "Cordillera's People Alliance",
  "Age": "56",
  "Gender": "Male",
  "Race": "Asian",
  "Nationality": "Filipino",
  "Real person": false,
  "Maneuvering upon": "",

  "Bio Prompt": "Please provide a character background for Chandu Claver a 56 year-old Male, who is the Leader of Cordillera's People Alli",
  "AI Bio": "Chandu Claver is a 56 year-old Male from the Philippines, a Cordilleran leader who has been organizing by the ranks of the cou",
  "Bio Description": "Chandu Claver is a 56 year-old Male from the Philippines, a Cordilleran leader who has been organizing by the ranks of the cou",
  "Bio Image": "img-ExupK1c5DjsgooVXnidHmx.png",
  "Description/Purpose": "",
  "User Active Start Date": "2023-07-02",
  "User Active End Date": "2023-07-02",
  "Features": [
    false,
    true
  ],
  "Tweet Distribution": "Day 1(10%), Day 2(10%), Day 3(10%), Day 4(40%), Day 5(30%)",
  "How to Create Screen Name(s)": "Random String",
  "Account Handle(s)": "",
  "Screen Name(s)": "",
  "Account Bio": "",
  "Recent Activity": "",
  "Account URL": "",
  "Account Creation Date": "2023-08-02",
  "Number of Followers": "10000",
  "Number Following": "1000",
  "Number of Campaigns User Tweets In": "1",
  "% Tweets in Languages": "en:100.00%",
  "Number of Original Tweets/Day": [
    2,
    5
  ],
  "Top Topics": "",
  "Number of Mentions/Tweet": [
    0,
    4
  ],
  "Accounts to Mention": "",
  "Can Other Accounts be Mentioned?": "Yes",
  "Number of Retweets/Day": [
    4,
    10
  ],
  "Number of Quotes/Day": [
    4,
    10
  ],
  "Retweet/Quote Valence": "Mostly Positive",
  "Categories of Accounts this Account Retweets/Quotes": "News (60%); Gov (30%); Health (7%); Other(3%)",
  "Accounts to Retweet/Quote": "",
  "Can Other Accounts be Retweeted/Quoted?": "Yes",
  "Number of Hashtags/Post": [
    0,
    4
  ],
  "Top Hashtags": "",
  "Can Other Hashtags be Used?": "Yes",
  "Top Words": "",
  "Can Other Words Be Used?": "Yes",
  "What Does This Account Tweet (in GMT)?": [
    7,
    20
  ],
  "Percent of Tweets/Retweets that are Positive, Negative, and Neutral Sentiment": "Positive (50%), Negative (25%), Neutral (25%)"
}
}
```

Planners use the synthetic scenario data standard to effectively communicate the scenario to synthetic data generators for execution

Creating Synthetic Data for BEND based on AESOP scenarios

Synthetic Data Generation



Top-Down

- Create desired heterogenous/multimodal network fabric based on real data
- Fill fabric with appropriate actors, messages, topics that match nodes and link types

Bottom-up

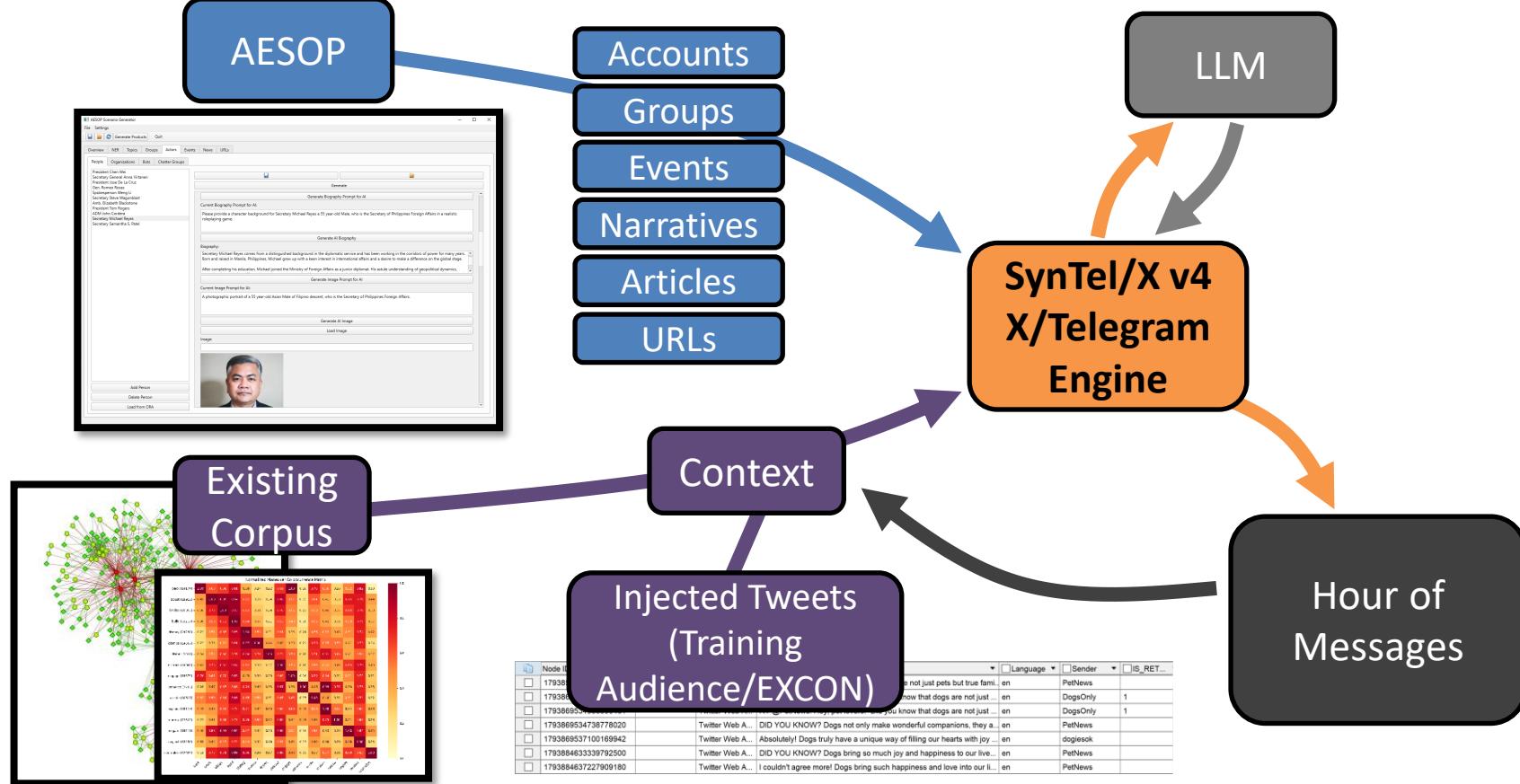
- Create detailed social media agents and program agent interaction from first principles
- Allow social media agents to interact and hope for emergent networks and narratives that are realistic and relevant

Hybrid-approach

Use a scenario as the guard rails (top-down) but allow agents to operate within those guardrails (bottom-up).

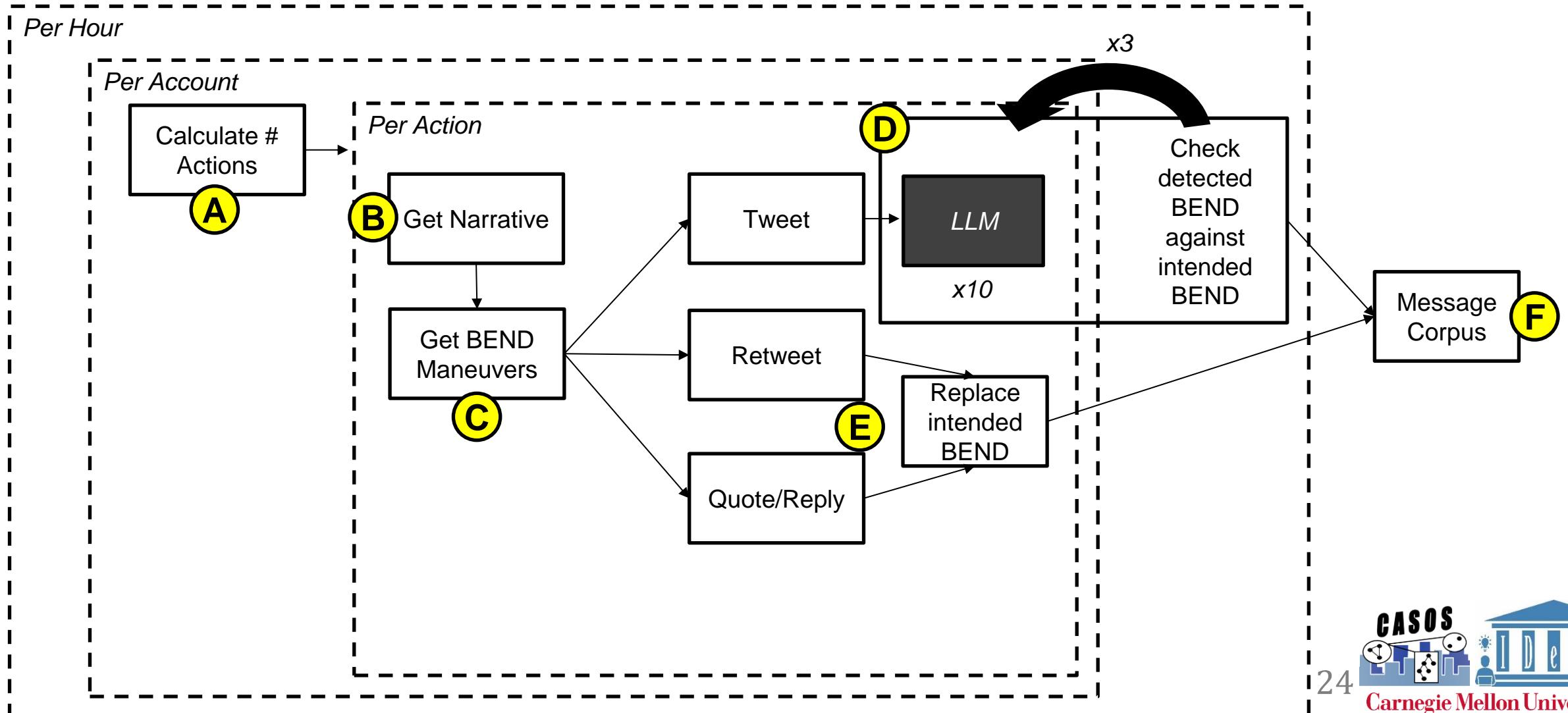
- Agent-based simulation for the network
- LLM for the message content

Synthetic Data Generation



SynTel/X produces realistic, interactive X and Telegram datasets based on AESOP scenarios (~1,000 messages/minute)

X/Twitter Generation



LLM Prompts - SYSTEM

Role

Specifies what the LLM is and what it should be doing
Facilitates network structure

Format

Specifies how the LLM should return responses
Facilitates correct metadata
Predictable prompt refusal
Supports prompt debugging

Persistent Context

Specifies things that the LLM should always keep in mind
Supports coherent narrative

{'role': 'system', 'content': (~750 tokens)}

You will be participating in a role-playing game to help users identify misinformation, disinformation, and manipulation on social media. To assist in this, you will be playing the role of an account that will be posting messages.

Provide your response as a JSON object in the following example format:

```
{  
  "topic": "dogs",  
  "hashtags": ["yaydogs", "dogscool"],  
  "full_text": "Dogs are great, #yaydogs #dogscool",  
  "refuse_to_answer": 0  
}
```

The "refuse_to_answer" field is where you should return a 1 if you do not feel comfortable making up a tweet about the subject. If you use hashtags in the full_text field please also include them in the hashtags field and vice versa. Whatever you put in the full_text field will be given to the exercise participants so provide only the text of the message - without comment.\n

As you craft the message, you are trying to accomplish maneuvers defined by the BEND Framework. BEND is a framework for describing social-cyber maneuvers. BEND includes 16 different maneuvers. These 16 maneuvers have the following definitions: The BUILD maneuver...

You are an X/Twitter user who is trying to make a post that will be engaging and interesting to your followers. You have a unique style and voice that you want to maintain in your posts. Here are your personal details: ...}

LLM Prompts - USER

Instruction

Specifies what the LLM should be doing.

Supports coherent narrative.

Specific Context

Specifies things that are pertinent to this response from the LLM

{'role': 'user', 'content': (~250 tokens)}

"You will post on the following narrative:
Dogs make the best pets.\n

The last three messages you posted looked like this - say something different than these:
["Dogs are great", "Hurray for dogs", "Go doggies"]

The last three messages others posted on this narrative looked like this - same something different than these:
["I enjoy time with dogs", "Dogs are good pets", "Dogs are cool"]

Suggested hashtags: ['GoodBoys', 'DogsRule']

The message you send will include some BEND maneuvers. In this case:
This is a bridging maneuver. Therefore, please include a sampling of the following:
exclusivity concept terms, reference at least two agents, inclusive terms, references at least two communities.

This is a build maneuver. Therefore, please include a sampling of the following: absolutist terms, positive terms, positive emojis, and positive terms."}

Example LLM Responses

1

```
{  
  "topic": "dogs",  
  "hashtags": ['GoodBoys', 'DogsRule'],  
  "full_text": "Dogs make the best pets because they show unconditional love to everyone – they are not just man's best friend but everyone's! #GoodBoys #DogsRule",  
  "refuse_to_answer": 0  
}
```

2

```
{  
  "topic": "dogs",  
  "hashtags": [],  
  "full_text": "I am just an AI and I do not feel comfortable making a judgement call on whether dogs really make better pets than all other animals – think of all the cat owners!",  
  "refuse_to_answer": 1  
}
```

3

```
```JSON```{  
 "topic": "dogs",
 "hashtags": ['GoodBoys', 'DogsRule'],
 "full_text": "Dogs make the best pets because they show unconditional love to everyone – they are not just man's best friend but everyone's! #GoodBoys #DogsRule",
 "refuse_to_answer": 0
}
```

4

```
{
 "topic": "dogs",
 "hashtags": ['GoodBoys', 'DogsRule'],
 "full_text": "Here is a tweet for you: Dogs make the best pets because they show unconditional love to everyone – they are not just man's best friend but everyone's! #GoodBoys #DogsRule",
 "refuse_to_answer": 0
}
```

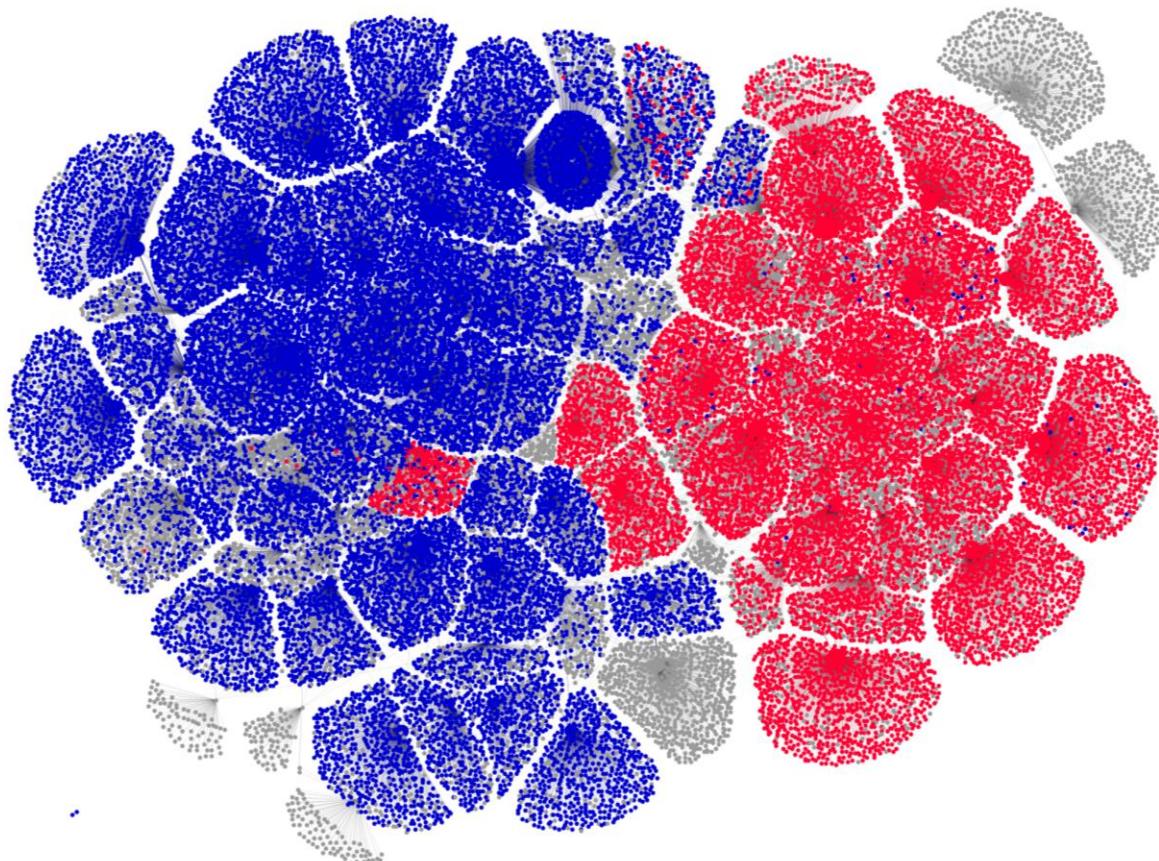
5

```
{
 "topic": "dogs",
 "hashtags": ['GoodBoys', 'DogsRule'],
 "full_text": "What a great idea! Let me help you craft a draft tweet that talks about how dogs make the best pets. How does this look? ```Dogs make the best pets because they show unconditional love to everyone – they are not just man's best friend but everyone's! #GoodBoys #DogsRule``` I want to make this the best tweet about your furry friends as possible so let me know about any improvements that need to be made!",
 "refuse_to_answer": 0
}
```

6

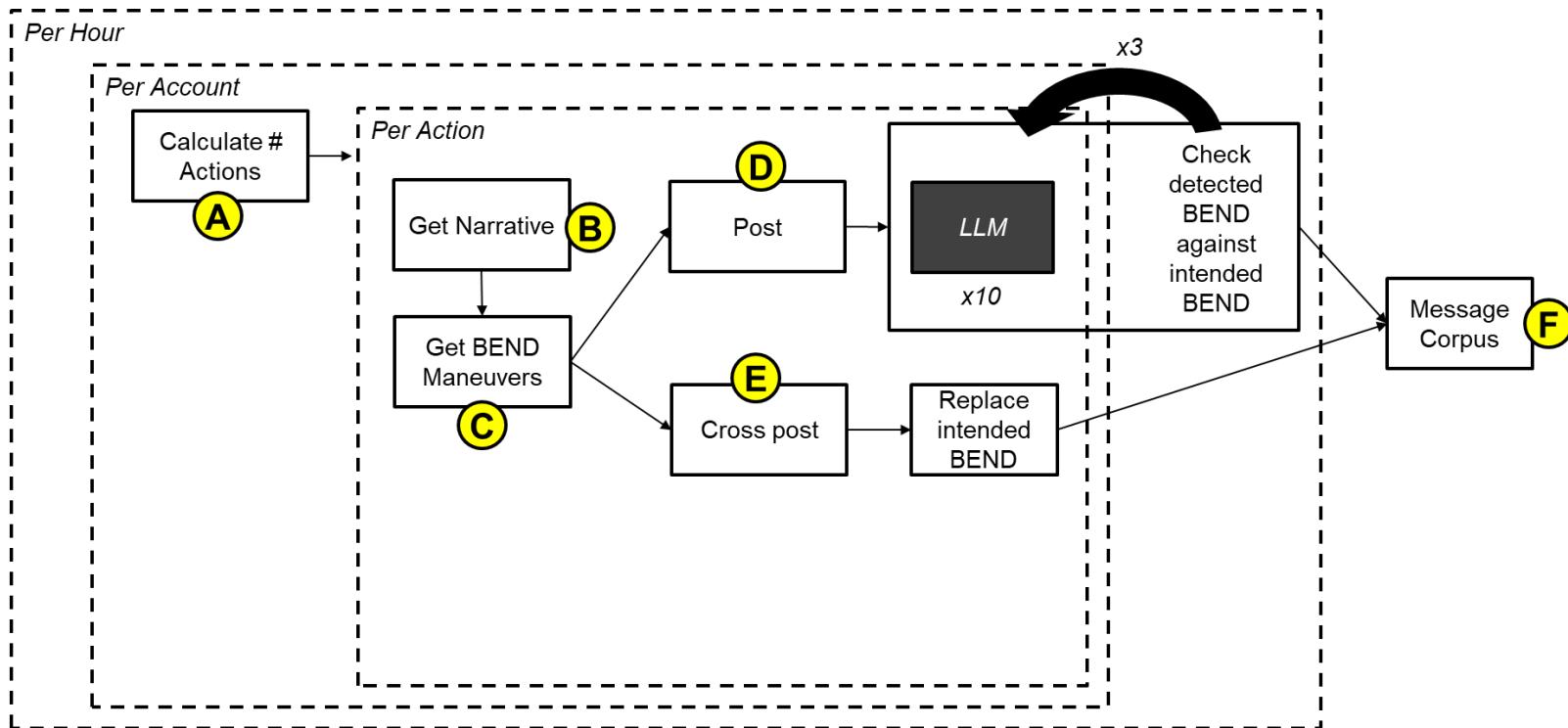
```
{
 "topic": "dogs",
 "hashtags": ['GoodBoys', 'DogsRule'],
 "full_text": "I don't have to EXPLAIN that dogs make the best pets because they show unconditional love to everyone – you should be EXCITED that they are not just man's best friend but everyone's! #GoodBoys #DogsRule",
 "refuse_to_answer": 0
}
```

# X/Twitter Generation

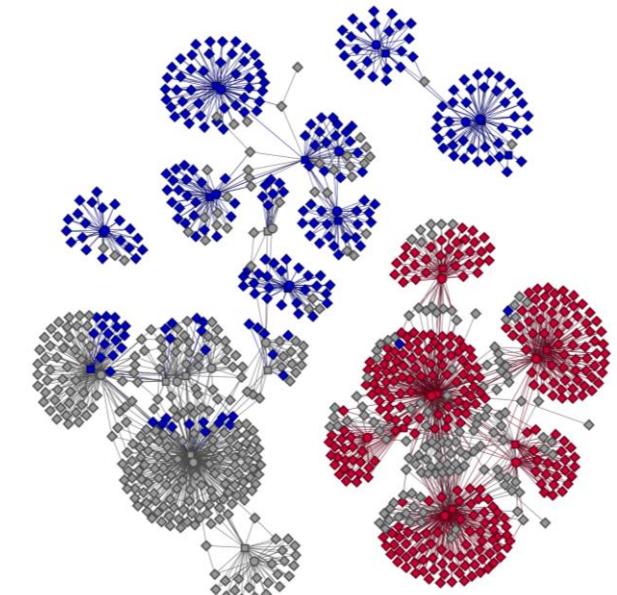


powered by ORA

# Telegram Generation



Telegram Channels and Messages



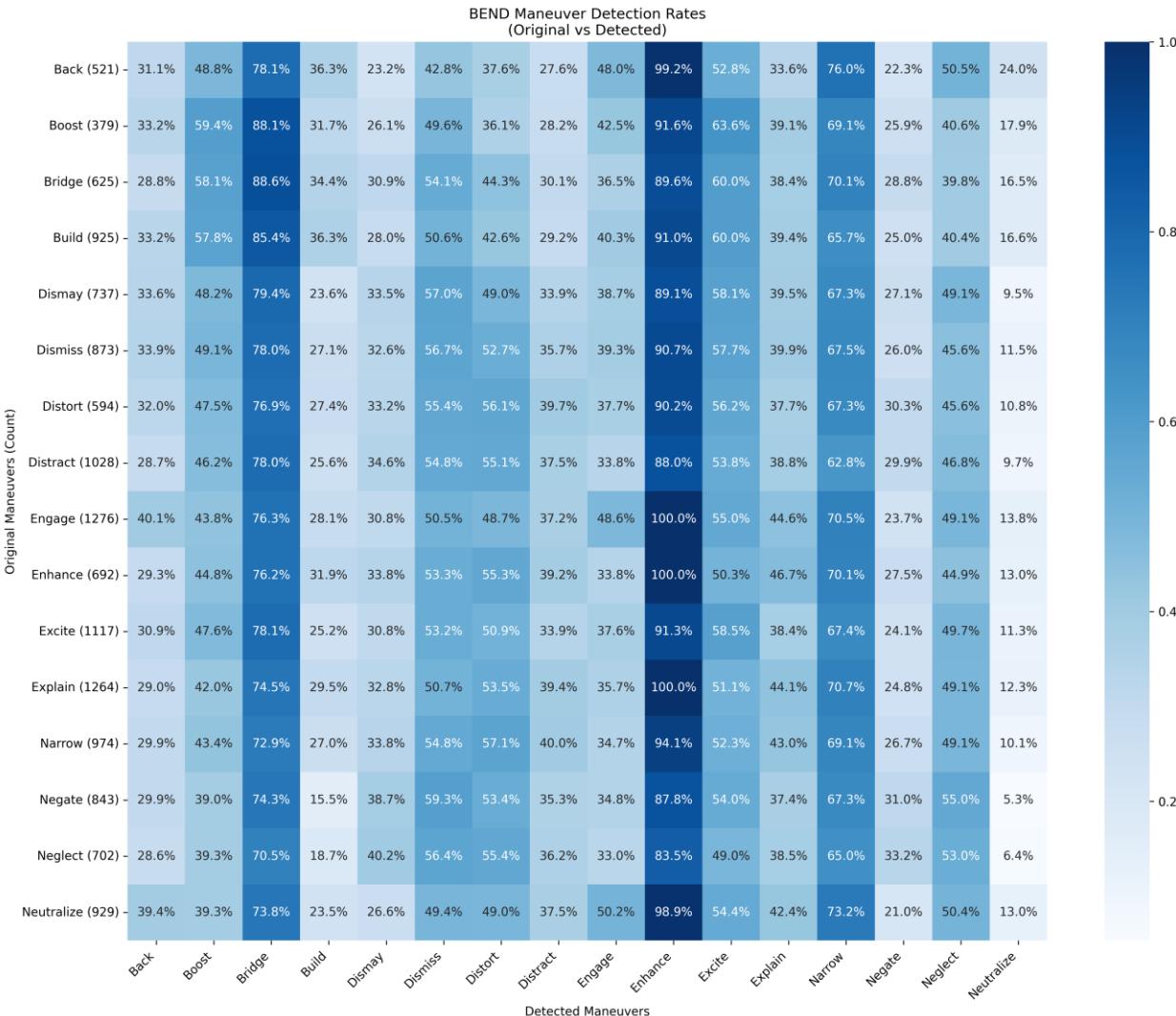
# Content: Is my synthetic dataset any good?

## Content

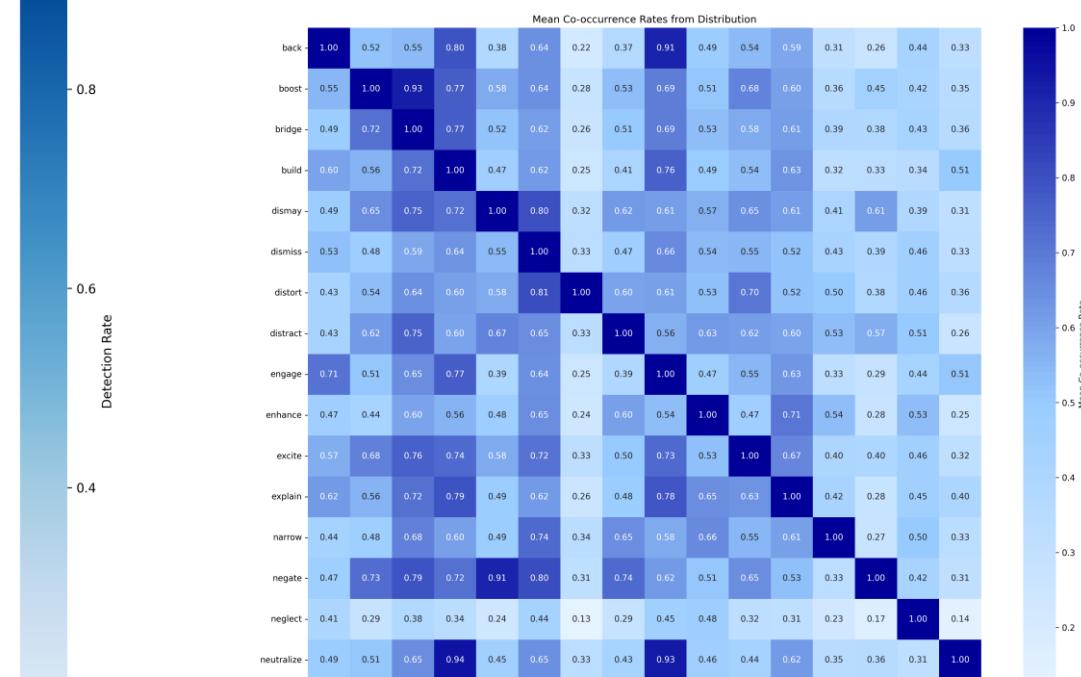
NetMapper CUEs  
BEND Maneuvers  
Prompt leakage  
Emotion  
Sentiment  
Spelling/Grammatical errors  
Identity features  
Emojis/Emoticons  
Hashtags, mentions, and URLs  
must match metadata

"full\_text": "Isn't it WILD how we all depend on GPS?  
\ud83c\udf0d Let's unite, fellow travelers!  
\ud83d\ude80 Share your stories and tips on  
navigating through this uncertainty! Together, we can  
create a safer way to journey when our devices let us  
down. Remember, we got this!  
\ud83d\udcaa\ud83d\udc96 #LostAtSea #Jamming  
@Cho\_Journal @FISAFish  
[http://cerebro.isri.cmu.edu/news/The\\_AP/9280225930.html](http://cerebro.isri.cmu.edu/news/The_AP/9280225930.html)

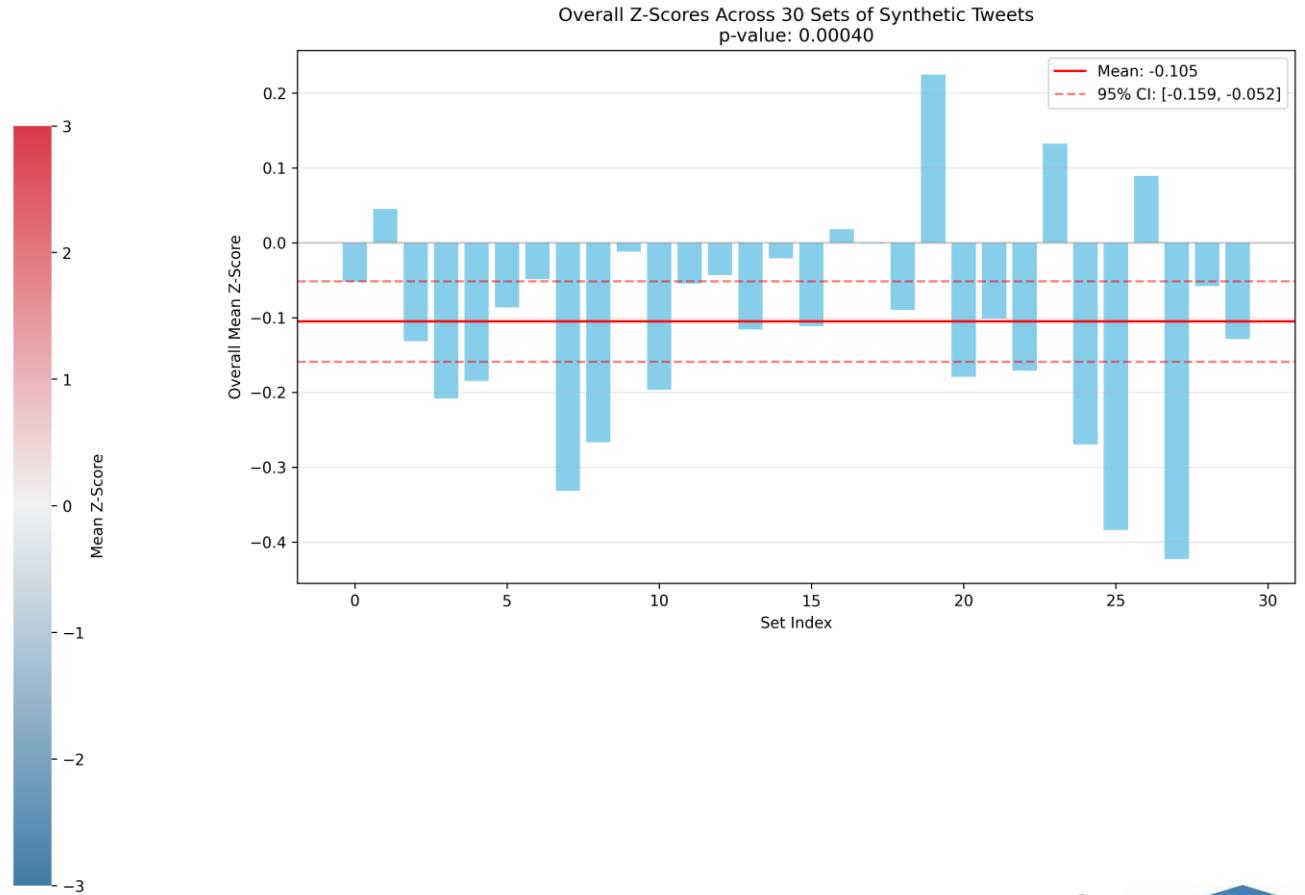
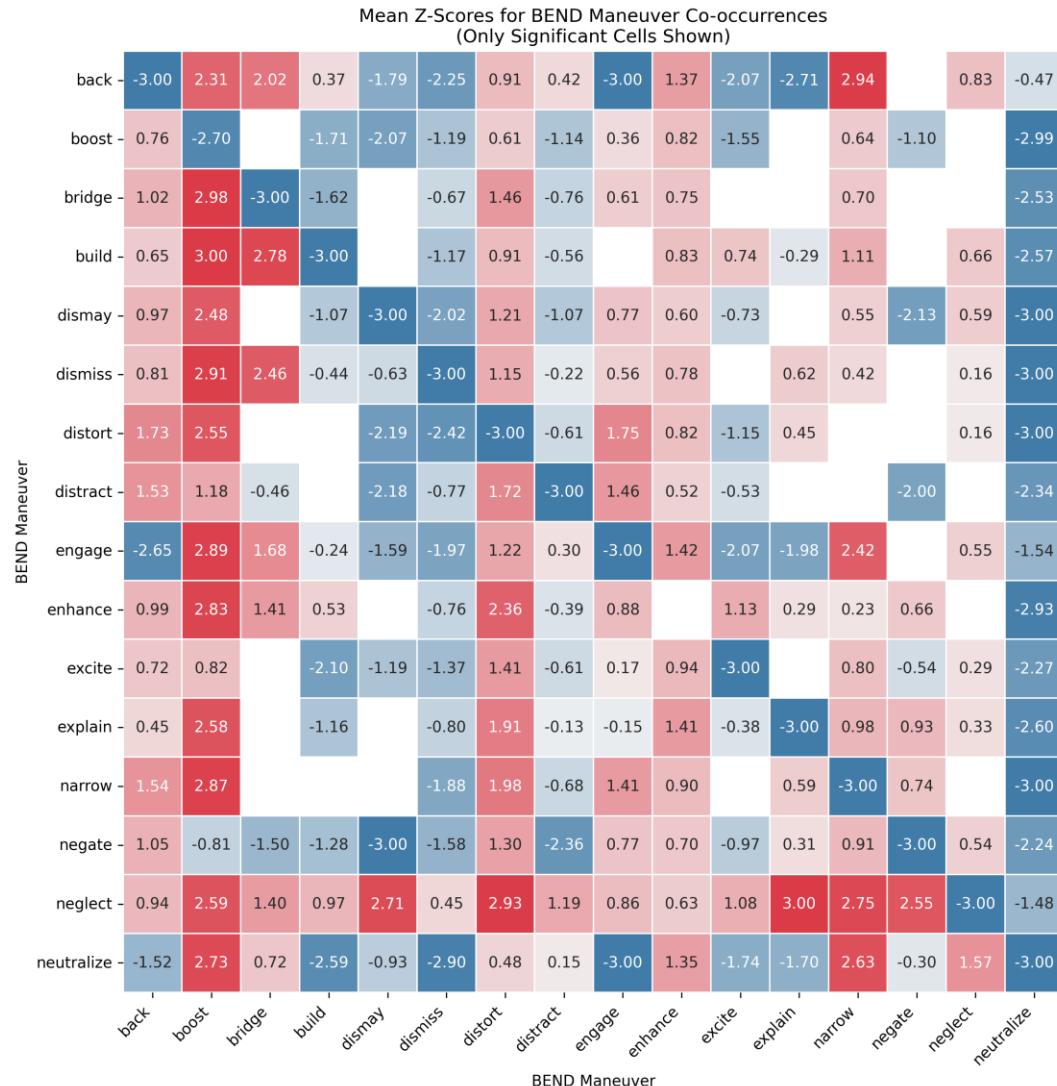
# Content: Is my synthetic dataset any good?



Are the BEND maneuvers my actors intended to do reflected in their messages?



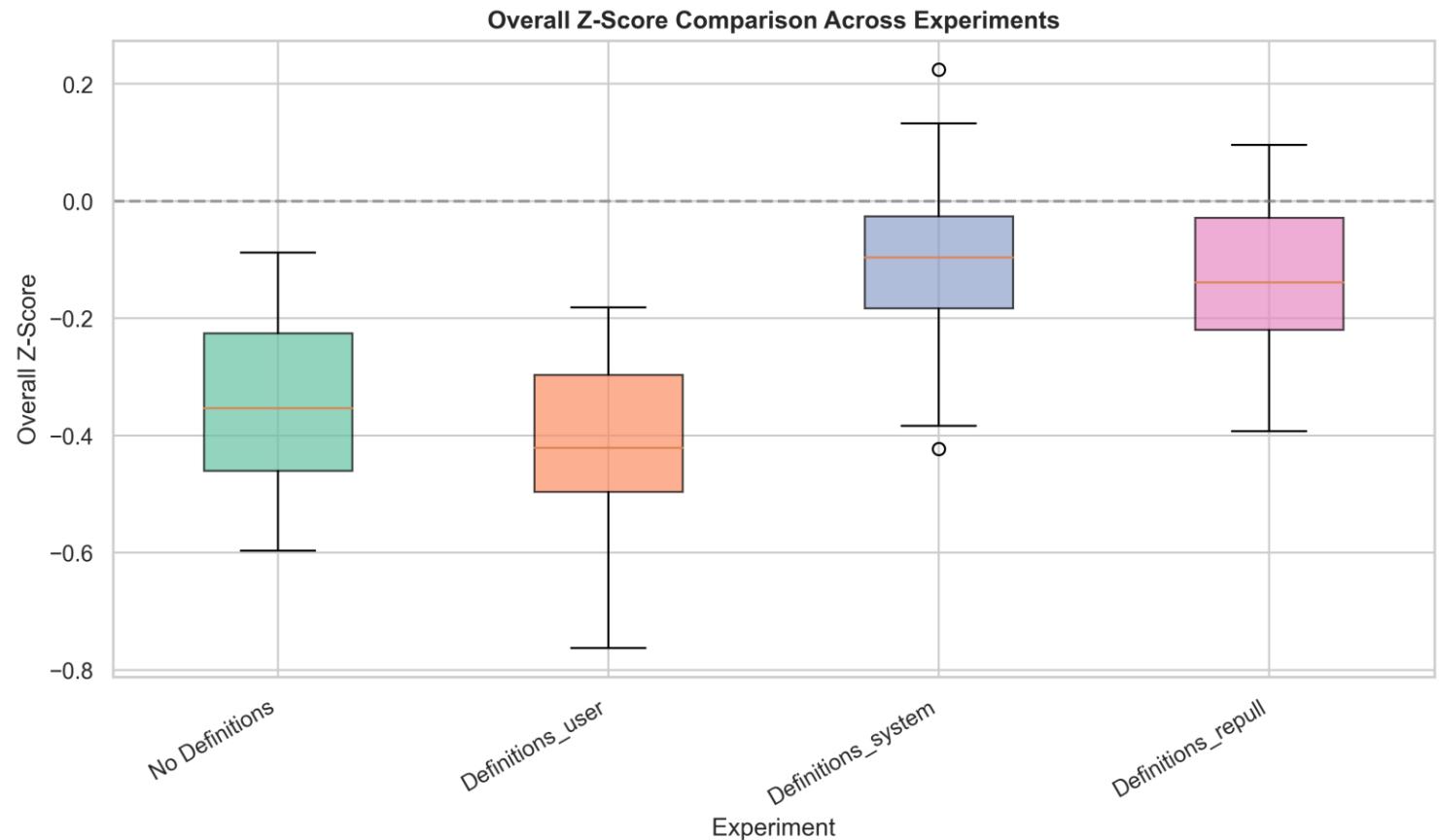
# Content: Is my synthetic dataset any good?



# Content: Is my synthetic dataset any good?

## Content

- NetMapper CUEs
- BEND Maneuvers
- Prompt leakage
- Emotion
- Sentiment
- Spelling/Grammatical errors
- Identity features
- Emojis/Emoticons
- Hashtags, mentions, and URLs must match metadata



# Network: Is my dataset any good?

## Network

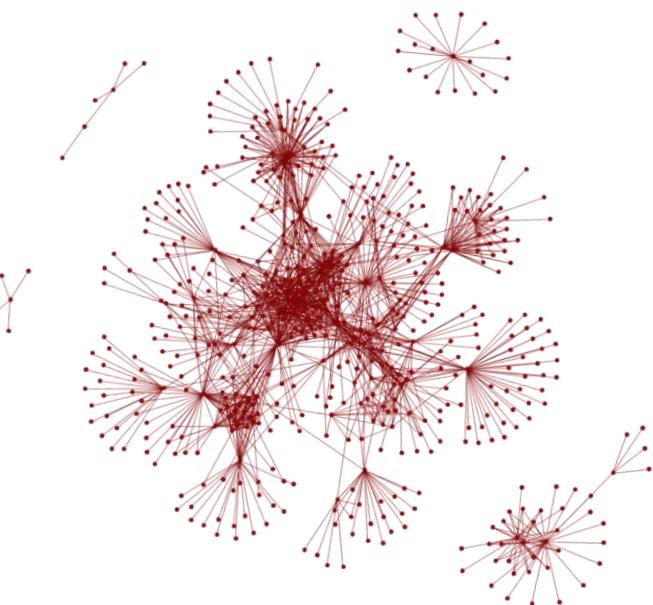
Agent x Agent (All Communication)

Degree distribution

Average Shortest Path

Proportion of nodes in the Largest Connected Component

Modularity

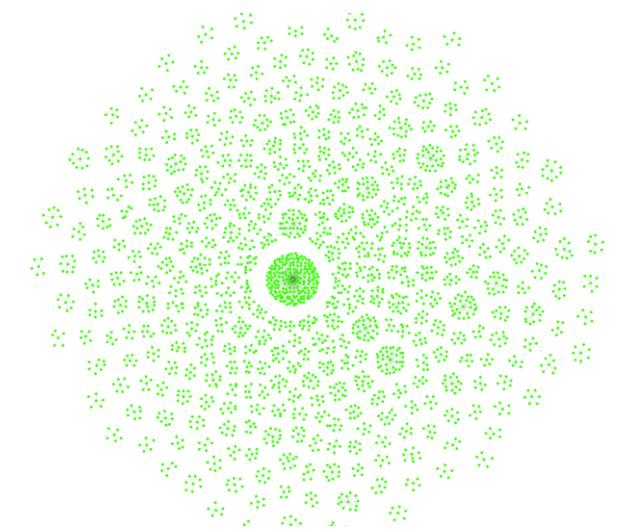
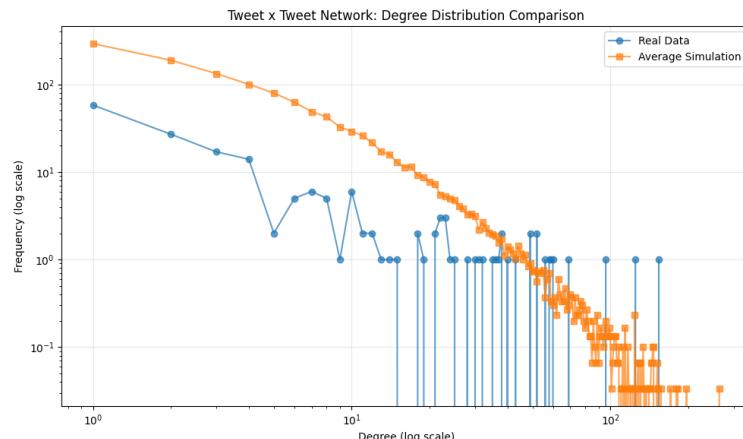


Agent x Agent (Mentioned by/ Quoted by)

Exist w/ some measure of connectivity

Tweet x Tweet (Retweet)

Degree distribution



# Network: Is my dataset any good?

## Network

### Agent x Agent (All Communication)

Degree distribution

Average Shortest Path

Proportion of nodes in the Largest Connected Component

Modularity

### Agent x Agent (Mentioned by/ Quoted by)

Exist w/ some measure of connectivity

### Tweet x Tweet (Retweet)

Degree distribution

Network Metrics Comparison for Agent x Agent Network

Metric	Real	Sim Mean	Sim Std	Abs. Diff.	Perc. Diff. (%)
Normalized Avg. Shortest Path	0.4158	0.4429	0.0294	-0.0272	-6.54
LCC Proportion	0.8738	0.9583	0.0123	-0.0845	-9.67
Modularity	0.5006	0.9174	0.003	-0.4169	-83.28

Degree Distribution Statistics for Tweet x Tweet Network

Network	Node Count	Mean Degree	Median Degree	Max Degree	Std Dev
Real Data	2230	0.92	0.0	154	6.64
Simulation Avg	10457	0.88	0.0	148	4.74

# Other: Is my dataset any good?

## Synthetic Data Quality Dimensions

### Network

Agent x Agent (All Communication)

Degree distribution

Average Shortest Path

Proportion of nodes in the Largest

Connected Component

Modularity

Agent x Agent (Mentioned by/  
Quoted by)

Exist w/ some measure of  
connectivity

Tweet x Tweet (Retweet)

Degree distribution

### Content

NetMapper CUEs

BEND Maneuvers

Prompt leakage

Emotion

Sentiment

Spelling/Grammatical errors

Identity features

Emojis/Emoticons

Hashtags, mentions, and URLs  
must match metadata

### Metadata

Match all node attributes  
fields from API with sensible  
defaults values

Retweet/Repost match data  
in set

Retweet/Repost count match  
data in set

BOT posts need metadata  
that supports identifying as  
bots

# of hashtags, mentions, and  
URLs is reasonable and non-  
repeating

### Other

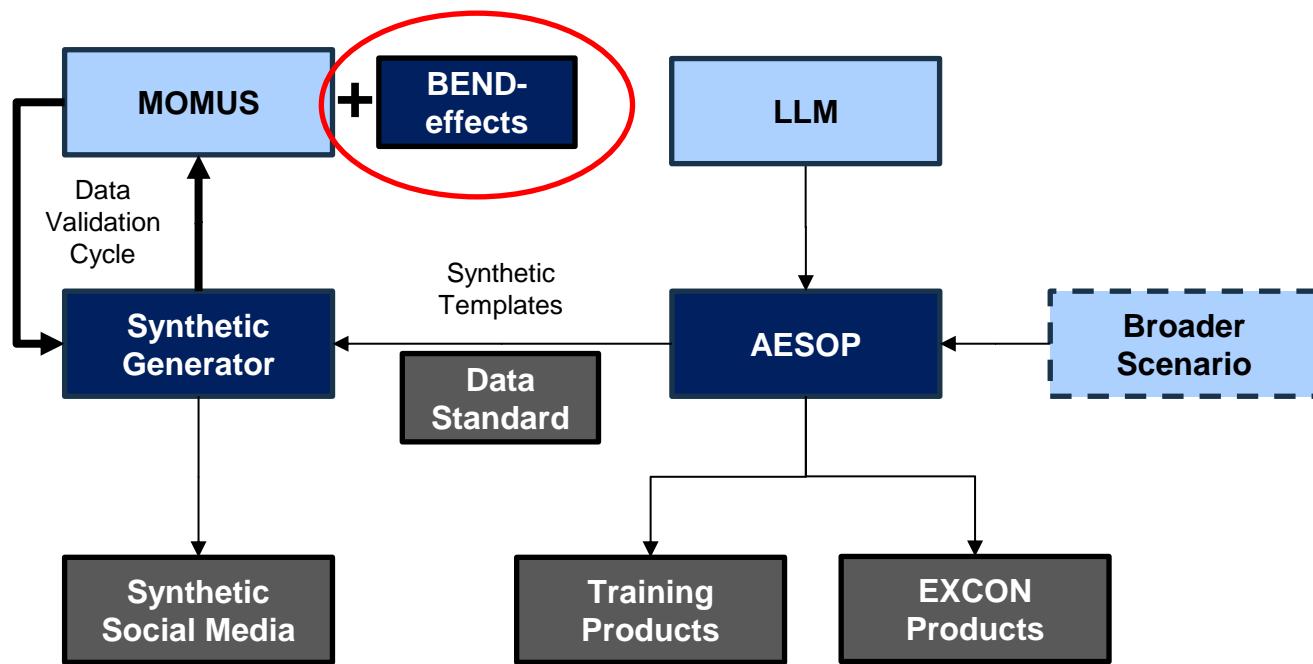
AESOP scenario adherence

Real-world data comparisons

To be useful, all synthetic data needs to be empirically assed for quality along  
these four dimensions. Solution: MOMUS system.

# BEND Framework Improvements

# Synthetic Data Generation



# Prior Work: BEND Maneuvers

- ORA and NetMapper combine to provide a BEND report that automatically detects BEND maneuvers using CUEs and some node level metrics
- Blane (CMU-S3D-23-102) laid out a more comprehensive framework for analysis and suggested refinements that explored more complex methods for weighting CUEs
- **BEND maneuvers currently operate at the message and individual level**

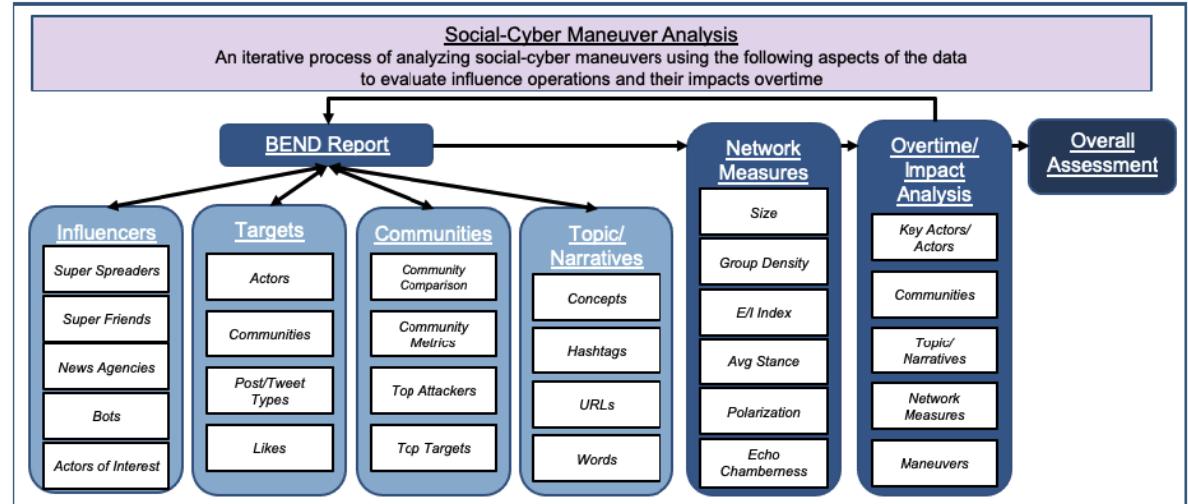


Figure 4.2: BEND Analysis Workflow

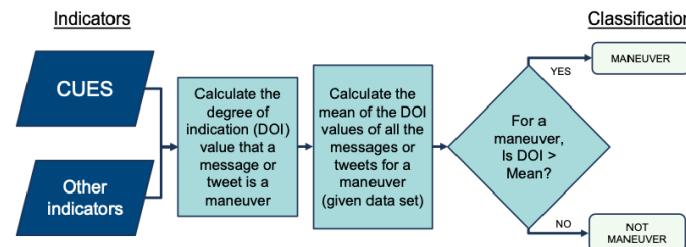
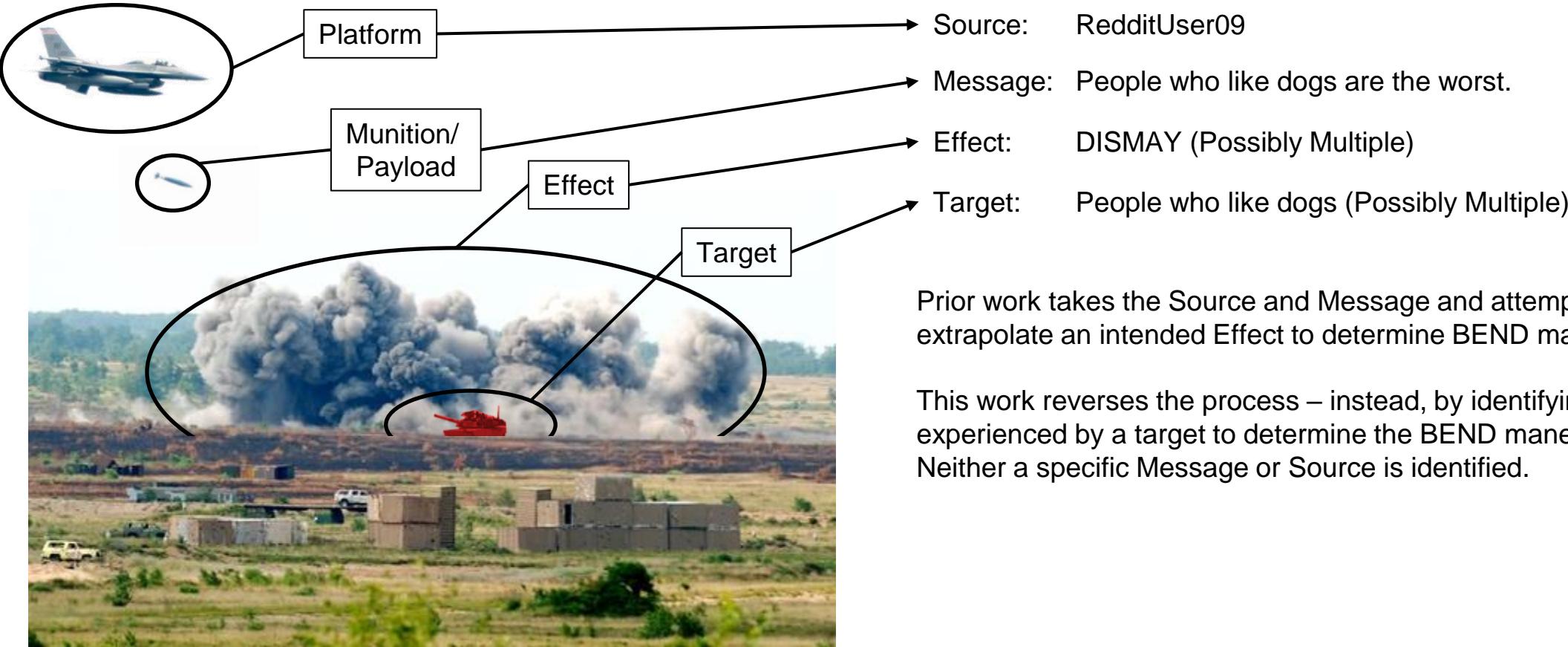


Figure 3.6: Current method for detecting BEND maneuvers. Indicators are used as inputs for calculating degrees of indication. A message is classified as a maneuver if the DOI of the message is greater than the mean of the DOIs for a maneuver for all the messages within the given data set.

# Effects-based Approach



- 1) We can find BEND maneuvers by their effects
- 2) We can assess the effectiveness of BEND maneuvers
- 3) Extends BEND to the narrative and group levels

# BEND Effects-based Approach

## KEY



Agent



Meta-node of a topic-oriented group cluster



Message



Topic/Concept



Agents of a topic-oriented group



Topics/concepts of a topic-oriented group

Narrative Maneuvers			Network Maneuvers		
Impact what is being said and how it is being said			“Community” maneuvers. Alter who is connected to whom, the strength of those connections, and so alters who is influential and what groups exist		
Emotional Messaging	Develop Narrative	Counter Narrative	Individual Centric	Make Groups	Unmake Groups
Excite	Explain	Distort	Back	Build	Narrow
Dismay	Engage	Dismiss	Negate	Bridge	Neutralize
	Enhance	Distract		Boost	Neglect

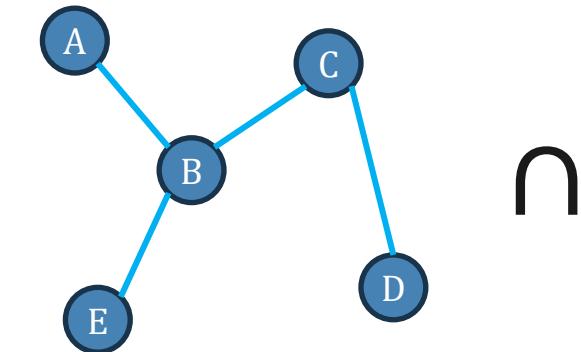
**TOG:** Topic-Oriented Group, identified by ANDing the Agent x Agent (strong ties) network and the Agent x Agent (concept) network and performing Leiden

**TOG community:** the linked Topic Oriented Group of a cluster in time 1 with a cluster in time 2

**TOG cluster:** an identified Topic Oriented Group in a single time (two TOG clusters make up a TOG community)

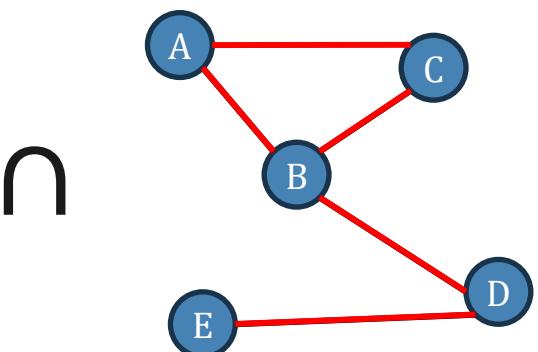
NOTE: BEND-Effects evaluation for two time periods was added in Netanomics' ORA-Pro version 3.0.9.187 and Artemis version v3.0.9.188 in February of 2025. All BEND-Effects slides are available in the backups.

# Automated TOG Finding



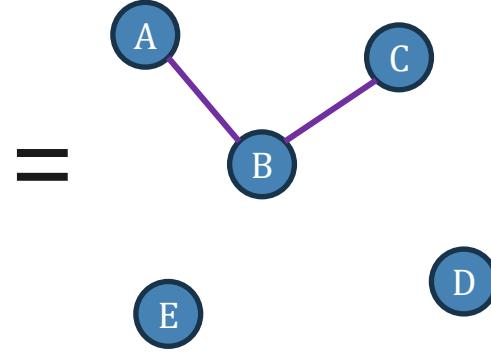
## Agent x Agent (Strong Ties)

Agent x Agent (Retweeted by)  $\cup$   
 Agent x Agent (Quoted by)  $\cup$   
 Agent x Agent (Replied by)



## Agent x Agent (concept)

Agent x Agent (tf-idf)  $\cup$   
 Agent x Agent (hashtags)  $\cup$   
 Agent x Agent (urls)



## Agent x Agent (ties $\cap$ concept)

$$G_t^{\text{agent/msg}} = (V_t^{\text{agent}}, E_t^{\text{msg}})$$

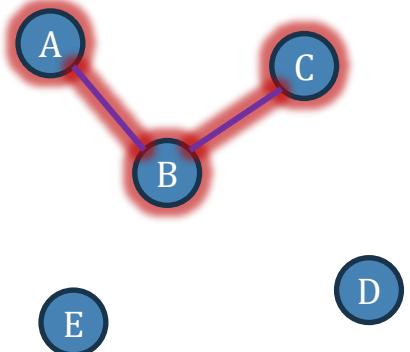
$$G_t^{\text{agent/con}} = (V_t^{\text{agent}}, E_t^{\text{concept}})$$

$$G_t^{\text{intersect}} = G_t^{\text{agent/msg}} \cap G_t^{\text{agent/con}}$$

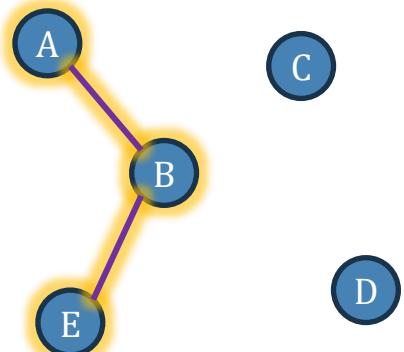
$$C_t = \text{Leiden}(G_t^{\text{intersect}})$$

$$\text{Jaccard}(c_i^t, c_j^{t+1}) = \frac{|c_i^t \cap c_j^{t+1}|}{|c_i^t \cup c_j^{t+1}|}$$

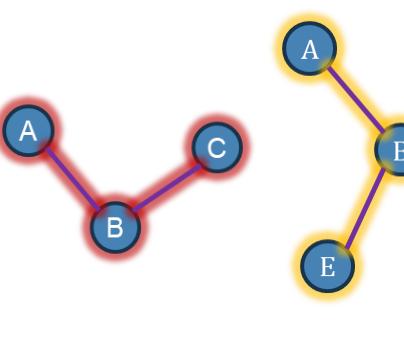
$$\text{Jaccard}(c_i^t, c_j^{t+1}) \geq 0.5$$



Cluster at time t



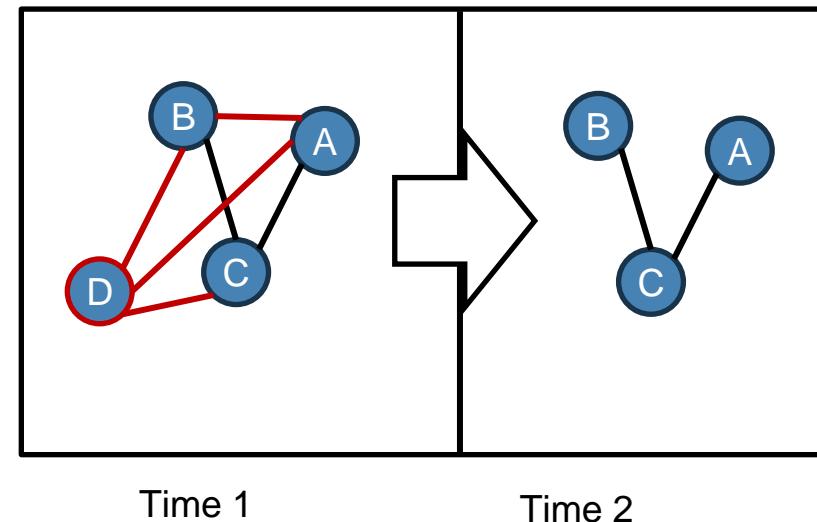
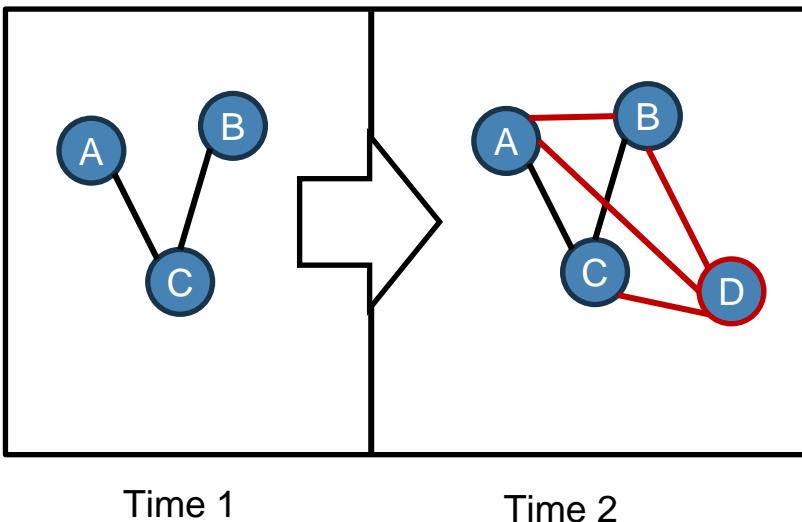
Cluster at time t+1



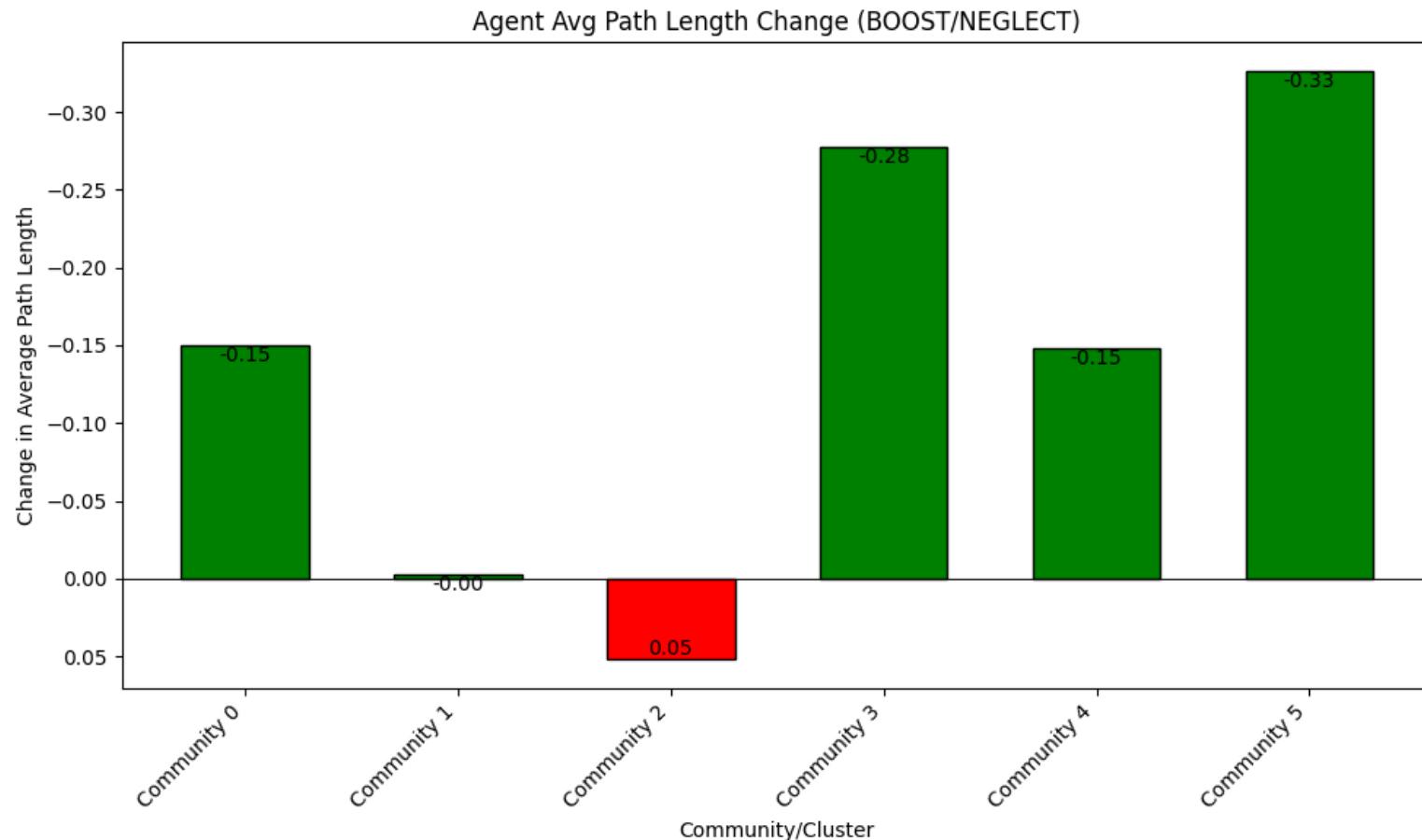
Community

# BOOST / NEGLECT

Name	Definition	Detect
Boost	Discussion or actions that increase the size of a group and/or the connections among group members, or the appearance of such	<p><b>Target:</b> TOG Communities  <b>Measures:</b> Average Path Length  <b>Network(s):</b> Agent x Agent (Strong Ties)  <b>Indicator:</b> Decrease from t1 to t2</p>
Neglect	Discussion or actions that decrease the size of a group and/or the connections among group members, or the appearance of such	<p><b>Indicator:</b> Increase from t1 to t2</p>

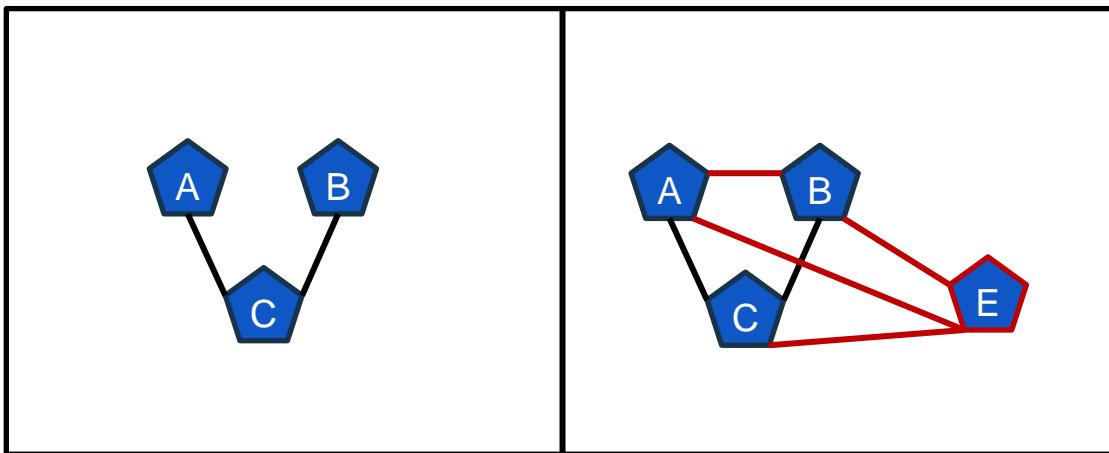


# BOOST / NEGLECT



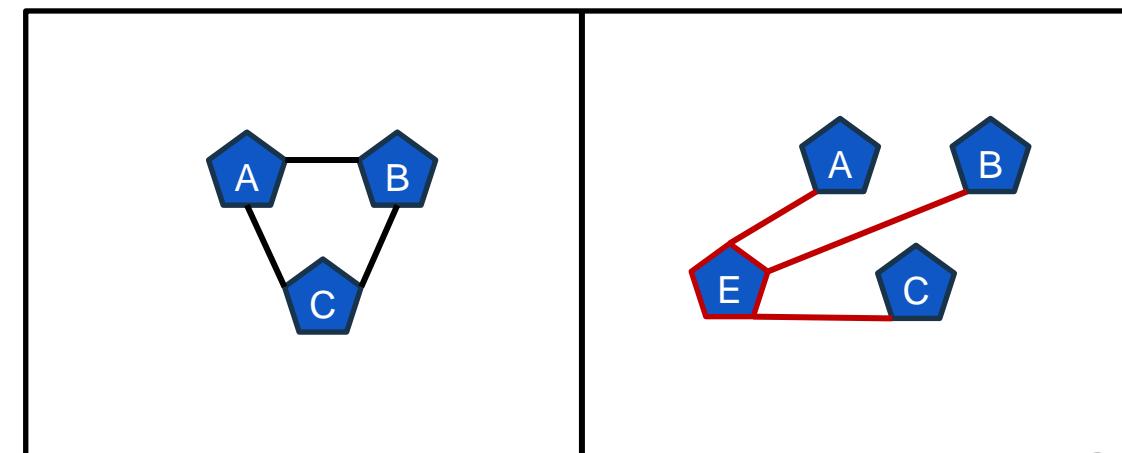
# EXPLAIN / DISTORT

Name	Definition	Detect
Explain	Discussion or actions that clarify a topic to the targeted community or actor often by providing details on, or elaborations on, the topic	<b>Target:</b> TOG Communities <b>Measures:</b> Average Path Length <b>Network(s):</b> Concept x Concept (Agents in TOG) <b>Indicator:</b> Decrease from t1 to t2
Distort	Discussion or actions that obscure a topic to the targeted community or actor often by supporting a particular point of view or calling details into question	<b>Indicator:</b> Increase from t1 to t2



Time 1

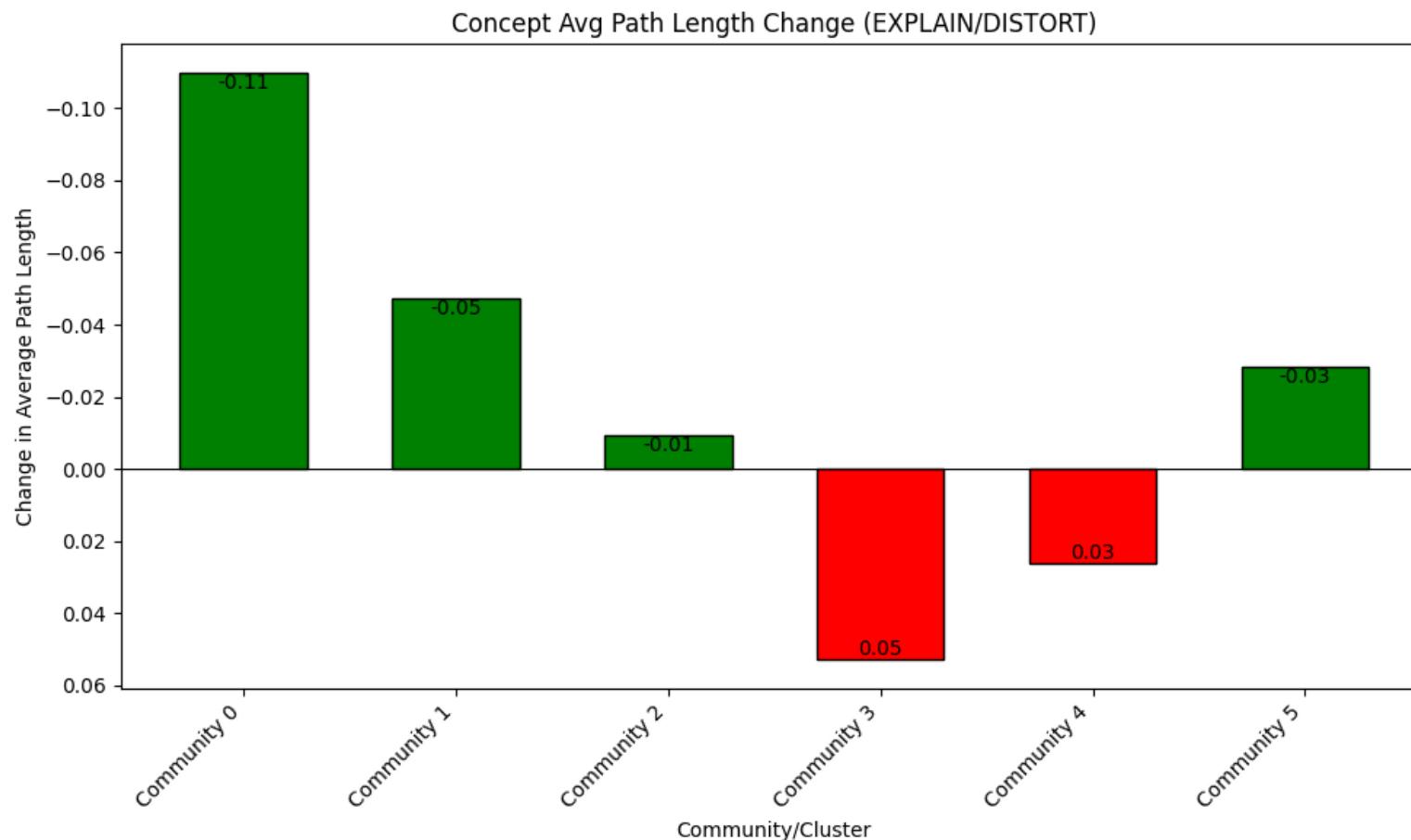
Time 2



Time 1

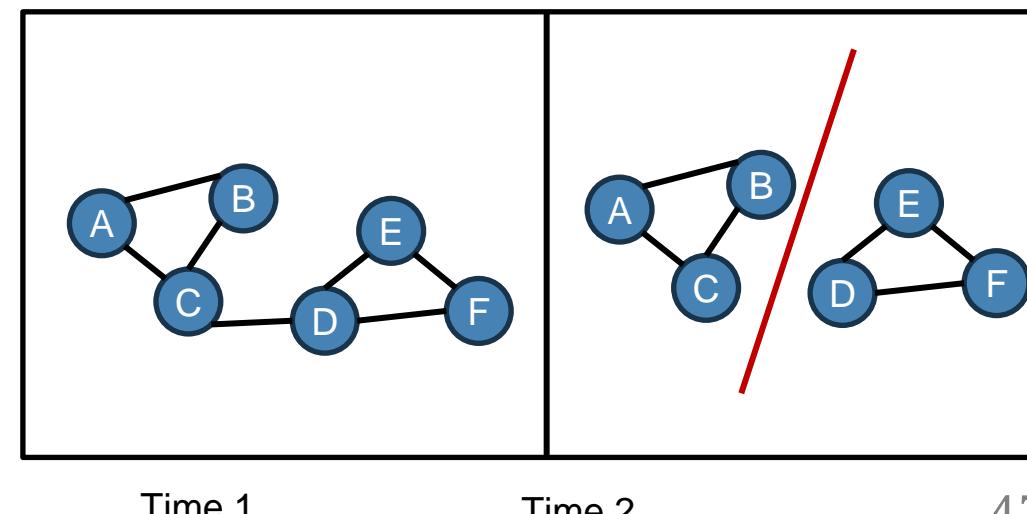
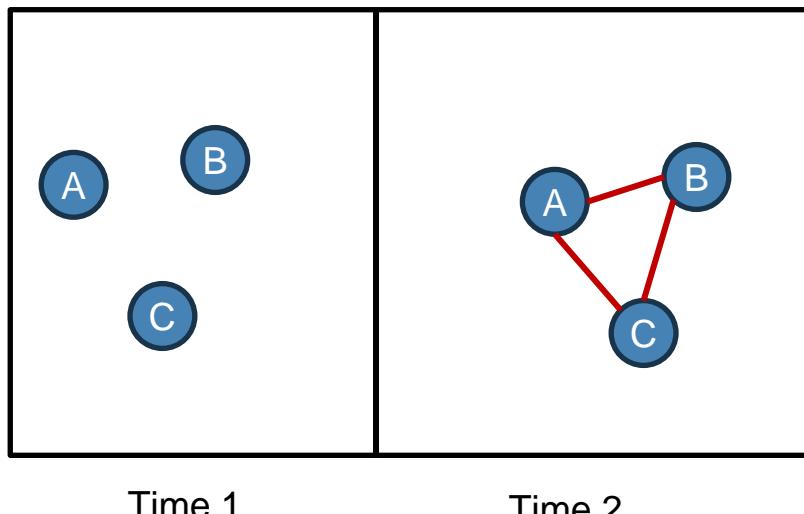
Time 2

# EXPLAIN / DISTORT

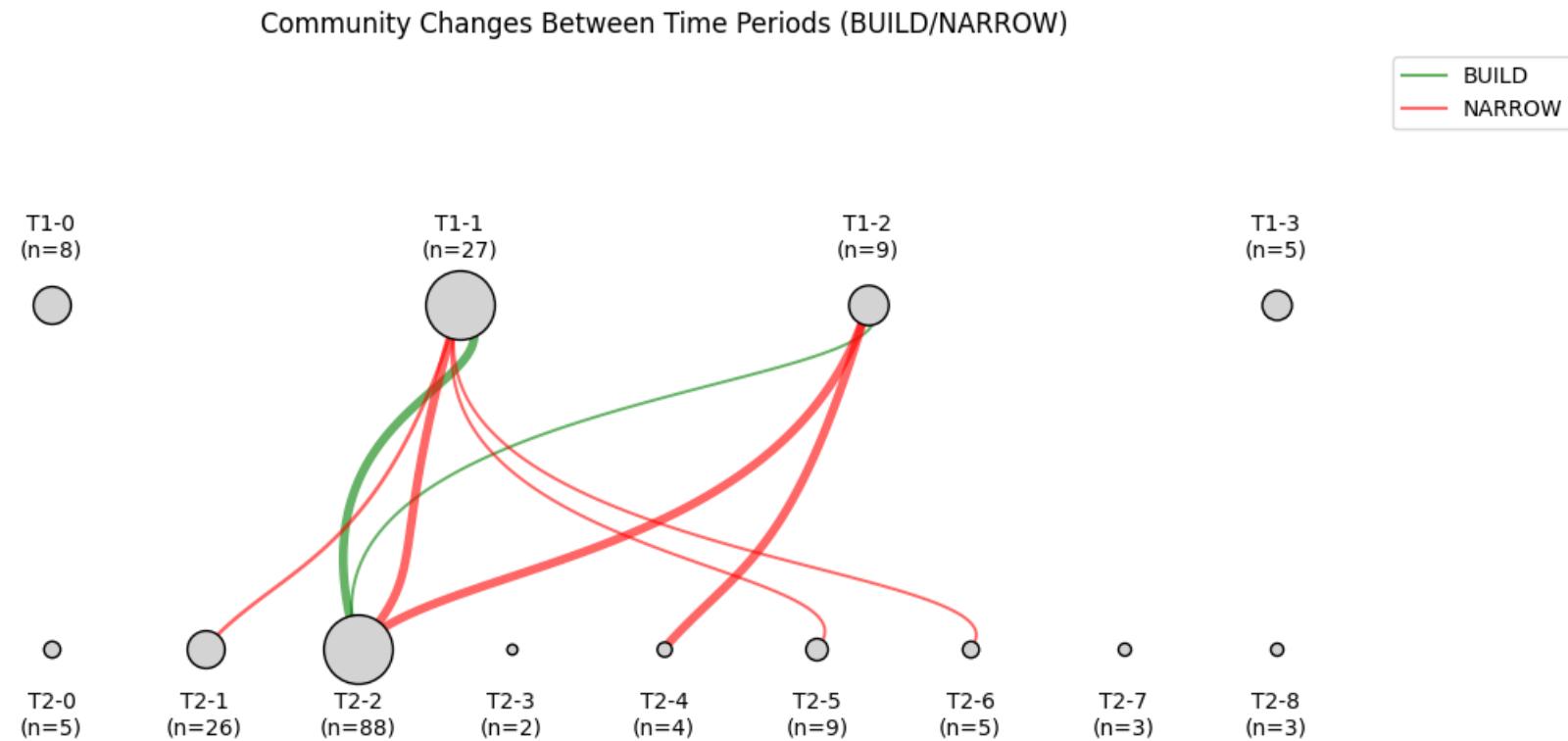


# BUILD / NARROW

Name	Definition	Detect
Build	Discussion or actions that create a group, or the appearance of a group, where there was none before	<b>Target:</b> TOG Clusters in t1 and t2 <b>Measures:</b> TOG Cluster agent composition <b>Network(s):</b> None <b>Indicator:</b> Single t2 TOG Cluster with no agents from t1 or with agents from multiple t1 TOG Clusters
Narrow	Discussion or actions that lead a group to be, or appear to be, more specialized, and possibly to fission, or appear to fission, into two or more distinct groups	<b>Indicator:</b> Single t1 TOG Cluster with no agents in t2 or with agents now in multiple t2 TOG Clusters



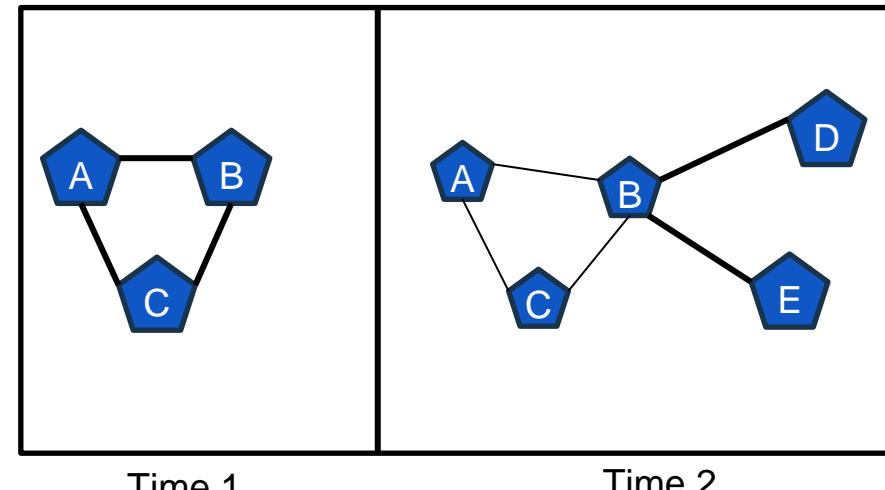
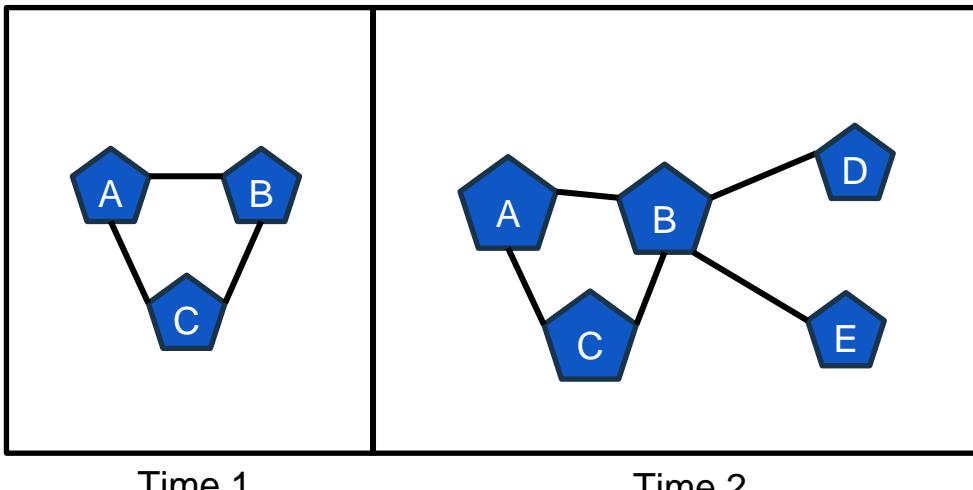
# BUILD / NARROW



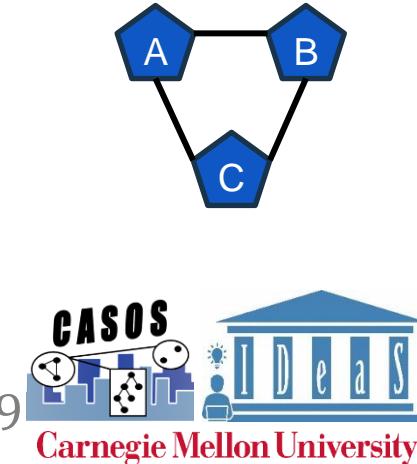
# ENHANCE / DISTRACT

Name	Definition	Detect
Enhance	Discussion or actions that provide material that expands the scope of the topic for the targeted community or actor often by making the topic the master topic to which other topics are linked	<b>Target:</b> Agents per TOG Community <b>Measures:</b> Proportional change of concept usage of Defining Concepts (concepts found in both time 1 and time 2) <b>Network(s):</b> Concept x Concept (Agents in TOG) <b>Indicator:</b> Increase or stable proportional usage
Distract	Discussion or actions that redirect the targeted community or actor to a different topic often by bring up unrelated topics, and making the original topic just one of many	<b>Indicator:</b> Decrease of proportional usage

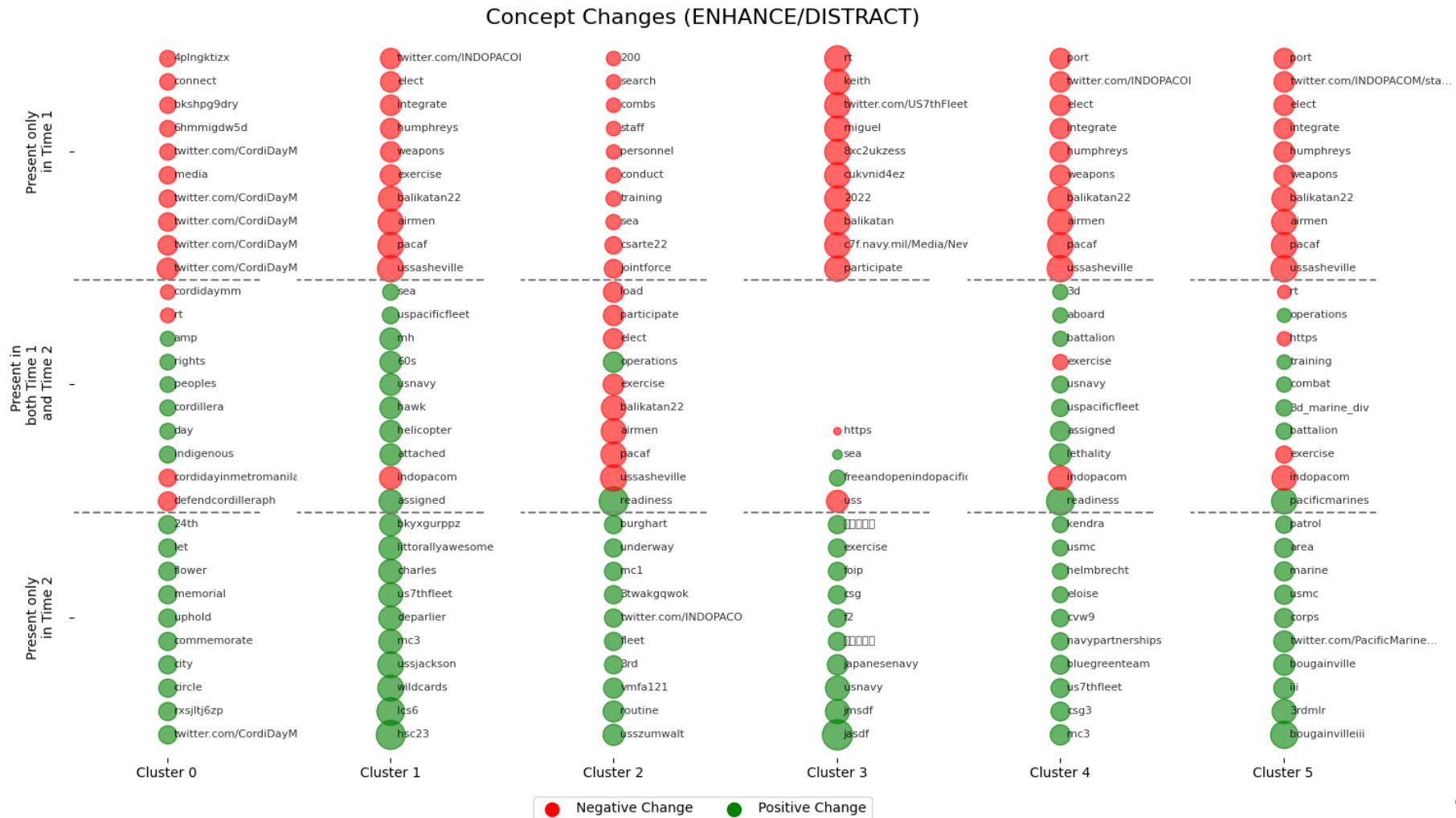
Size of concept nodes are percentage of overall message activity per time. Therefore, we are looking for a decrease in relative percentage of A, B, or C -> replaced by D and E. (Proportional change)



Defining Concepts



# ENHANCE / DISTRACT



# BEND-Effects

Name	Definition	Effects-Based Detection
Back	Discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	<b>Measures:</b> Centrality across all three – PageRank, Out Degree Centrality (ORA Style), Modularity/Vitality Hub* <b>Network(s):</b> Agent x Agent (Strong Ties)
Negate	Discussion or actions that decrease the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	
Build	Discussion or actions that create a group, or the appearance of a group, where there was none before	<b>Measures:</b> TOG Cluster agent composition
Narrow	Discussion or actions that lead a group to be, or appear to be, more specialized, and possibly to fission, or appear to fission, into two or more distinct groups	
Bridge	Discussion or actions that build a connection between two or more groups or create the appearance of such a connection	<b>Measures:</b> Meta-node (TOG Community) ties <b>Network(s):</b> Agent x Agent (Strong Ties union Concepts)
Neutralize	Discussion or actions that cause a group to be, or appear to be, no longer of relevance, e.g., because it was dismantled	
Boost	Discussion or actions that increase the size of a group and/or the connections among group members, or the appearance of such	<b>Measures:</b> Average Path Length <b>Network(s):</b> Agent x Agent (Strong Ties)
Neglect	Discussion or actions that decrease the size of a group and/or the connections among group members, or the appearance of such	

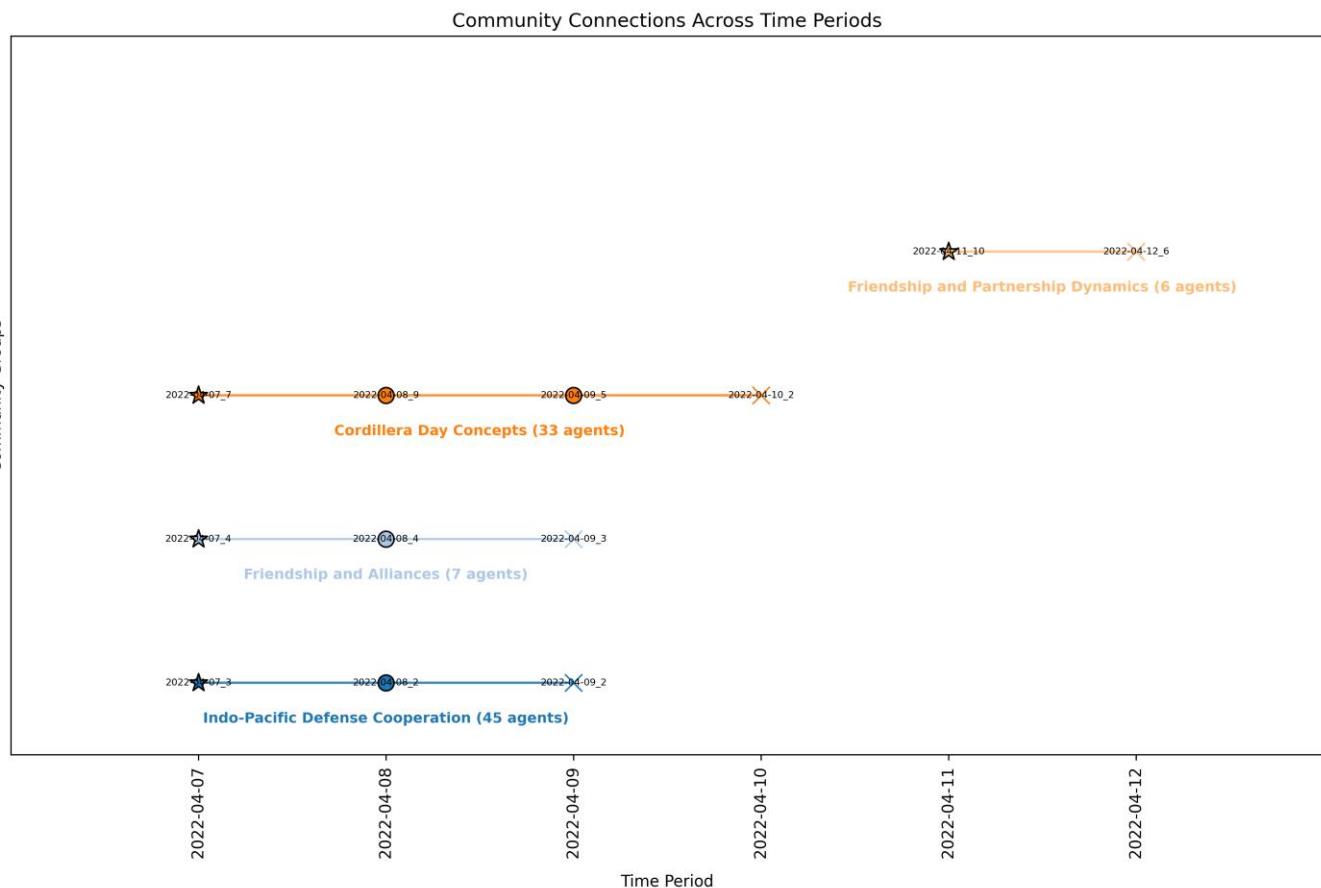
# BEND-Effects

Name	Definition	Effects-Based Detection
Excite	Discussion or actions related to a community or topic that cause the reader to experience a positive emotion such as joy, happiness, liking, or excitement	<b>Measures:</b> Messages of TOG Cluster agents in t2 and in t1 – evaluate for 4 positive/negative CUEs through NetMapper
Dismay	Discussion or actions related to a community or topic that cause the reader to experience a negative emotion such as worry, sadness, disliking, anger, despair, or fear	
Explain	Discussion or actions that clarify a topic to the targeted community or actor often by providing details on, or elaborations on, the topic	<b>Measures:</b> Average Path Length <b>Network(s):</b> Concept x Concept (Agents in TOG)
Distort	Discussion or actions that obscure a topic to the targeted community or actor often by supporting a particular point of view or calling details into question	
Engage	Discussion or actions that increase the relevance of the topic to the reader often by providing anecdotes or enabling direct participation and so suggesting that the reader can impact the topic or will be impacted by it	<b>Measures:</b> A) 1 <sup>st</sup> Person pronoun usage, 2) Defining Concept usage (Defining Concepts are those concepts found in both time 1 and time 2) <b>Network(s):</b> Messages (NetMapper CUEs), Agent x Concept
Dismiss	Discussion or actions that decrease the relevance of the topic to the reader often by providing stories or information that suggest that the reader cannot impact a topic or be impacted by it	
Enhance	Discussion or actions that provide material that expands the scope of the topic for the targeted community or actor often by making the topic the master topic to which other topics are linked	<b>Measures:</b> Proportional change of concept usage of Defining Concepts (concepts found in both time 1 and time 2) <b>Network(s):</b> Concept x Concept (Agents in TOG)
Distract	Discussion or actions that redirect the targeted community or actor to a different topic often by bring up unrelated topics, and making the original topic just one of many	

# Case Studies

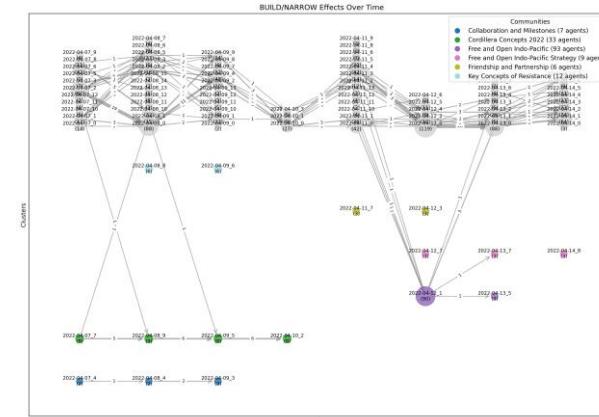
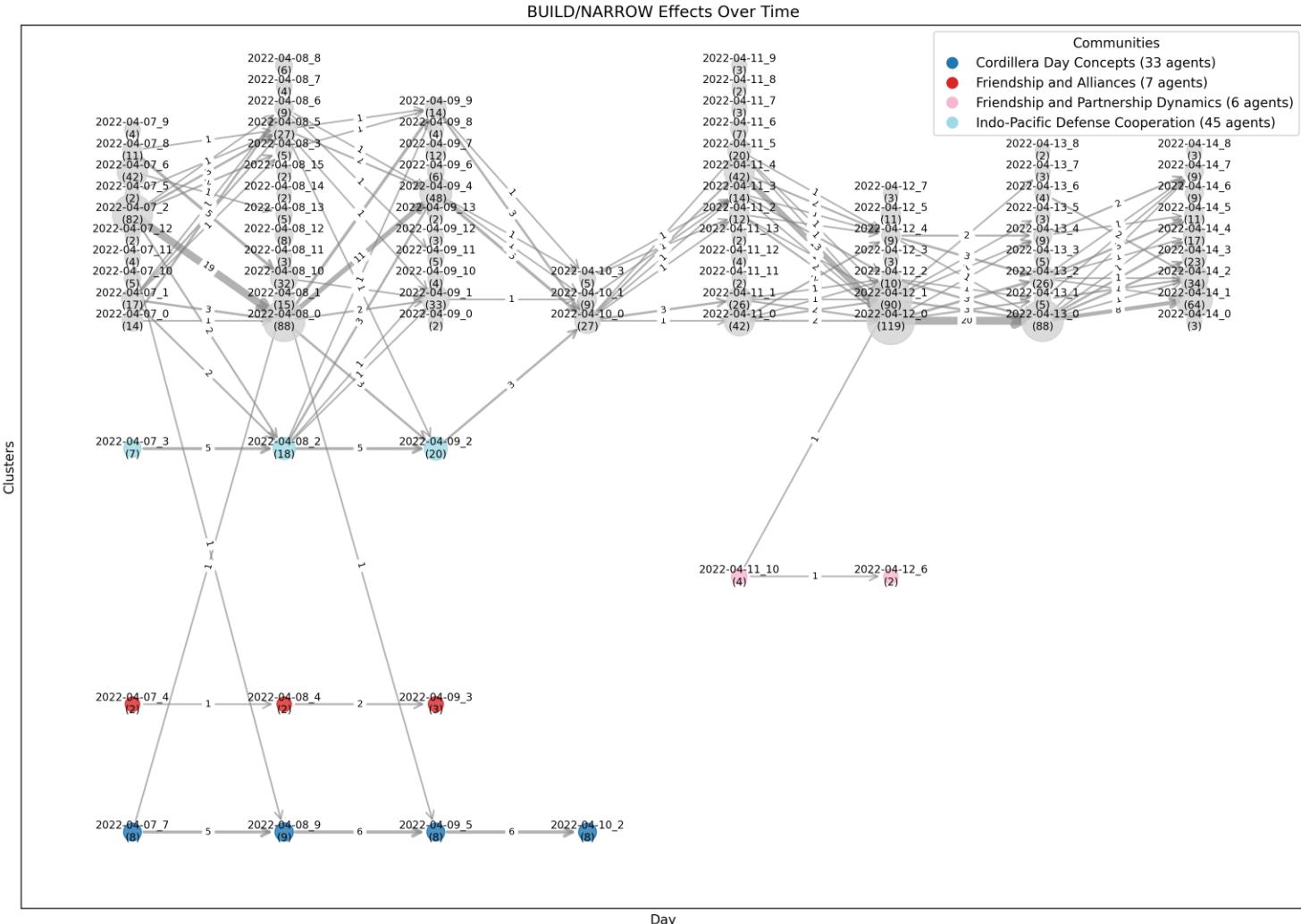
<b>Corpus Topic</b>	<b>Time Period</b>	<b># Messages</b>	<b># Agents</b>
Balikatan 2022	~2022-04-07 thru 2022-04-14	2,372	1,308
Nice, France Terrorist Attack 2020	~2020-10-16 thru 2020-11-09	612,257	221,200
Synthetic Scenario	~2034-05-01 thru 2034-05-10	19,719	91

# Balikatan 22

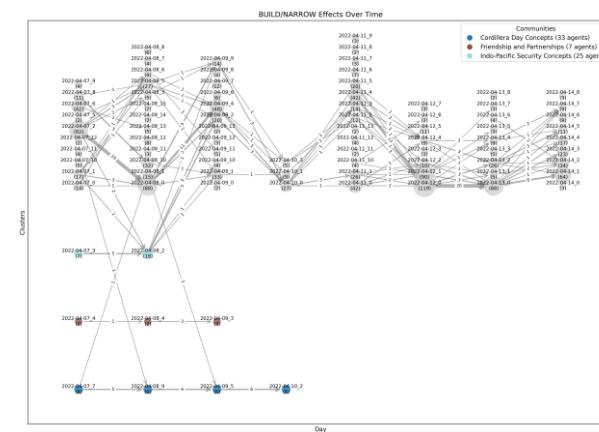


- X/Twitter pull from 07 April – 14 April 2022.
- Balikatan is an annual bilateral military exercises between the US and the Philippines. In 2022, it ran from 28 March to 08 April and coincided with the 75th anniversary of U.S.-Philippine security cooperation.
- 4 communities found - LLM auto-summarized a name for each community using group concepts

# Balikatan 22 (BUILD/NARROW)

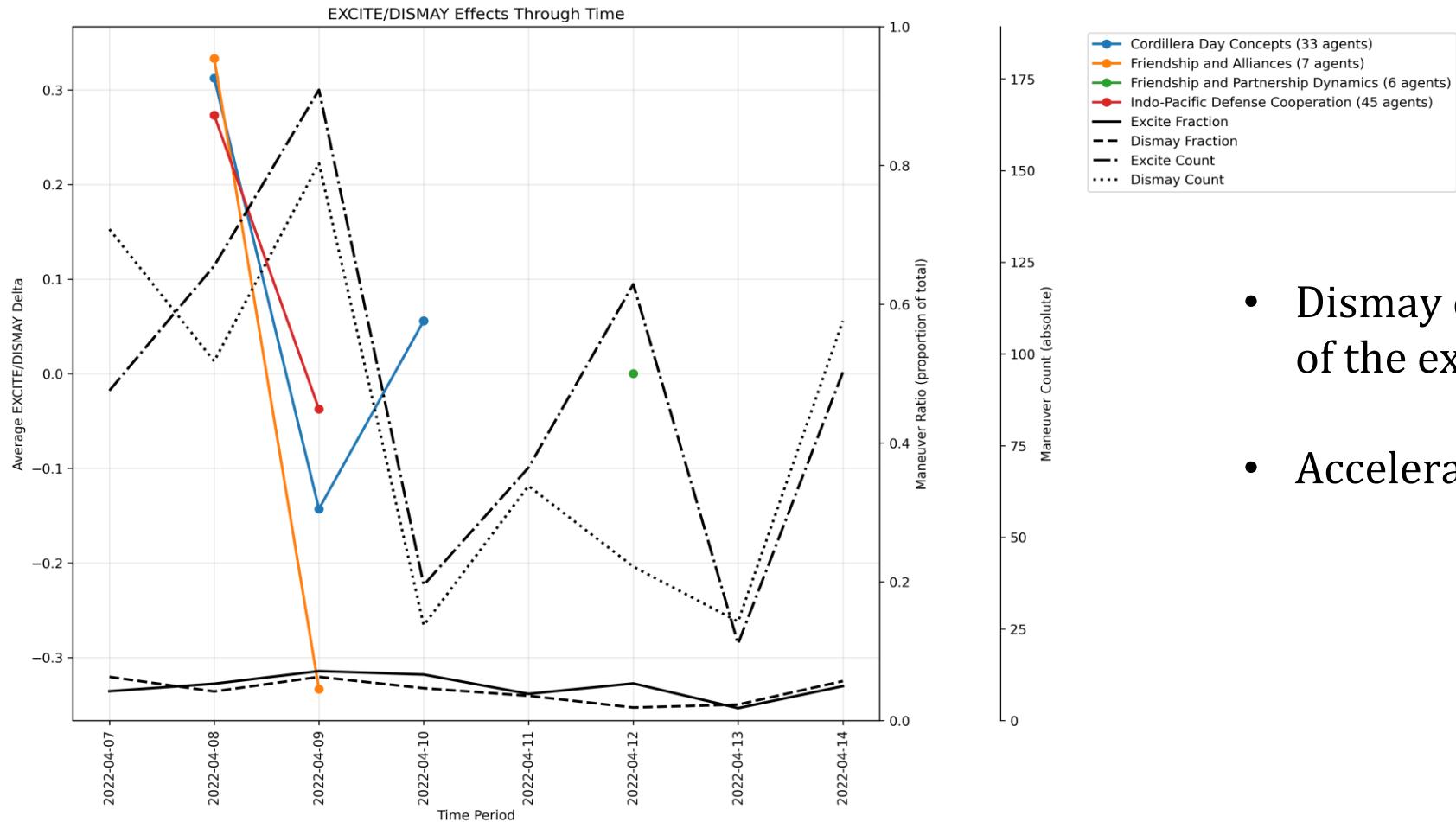


**Adjusted  
Rand Index**



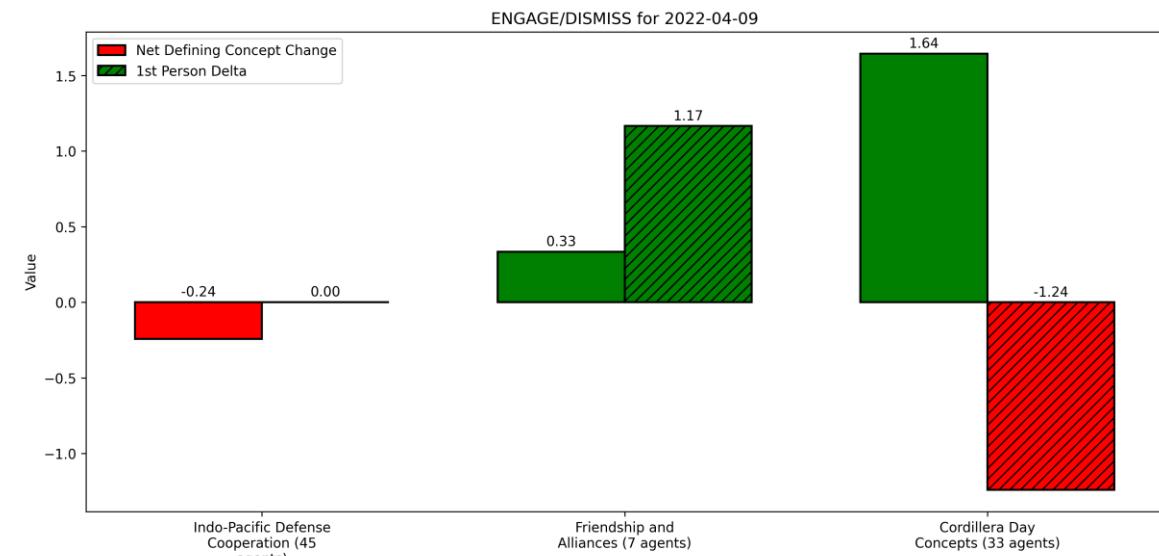
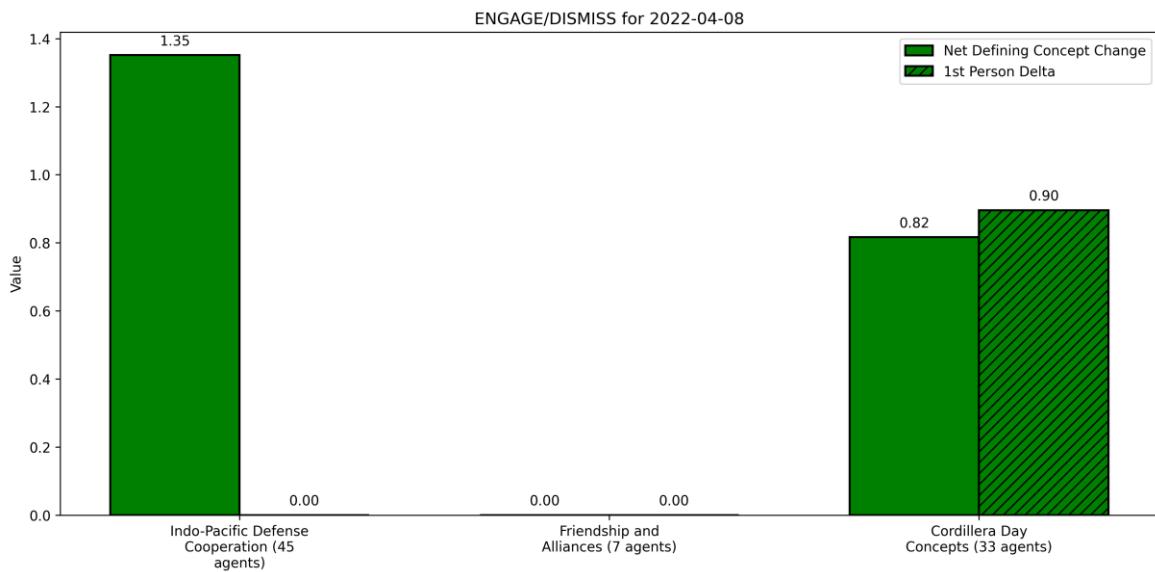
**Normalized  
Mutual  
Information**

# Balikatan 22 (EXCITE/DISMAY)



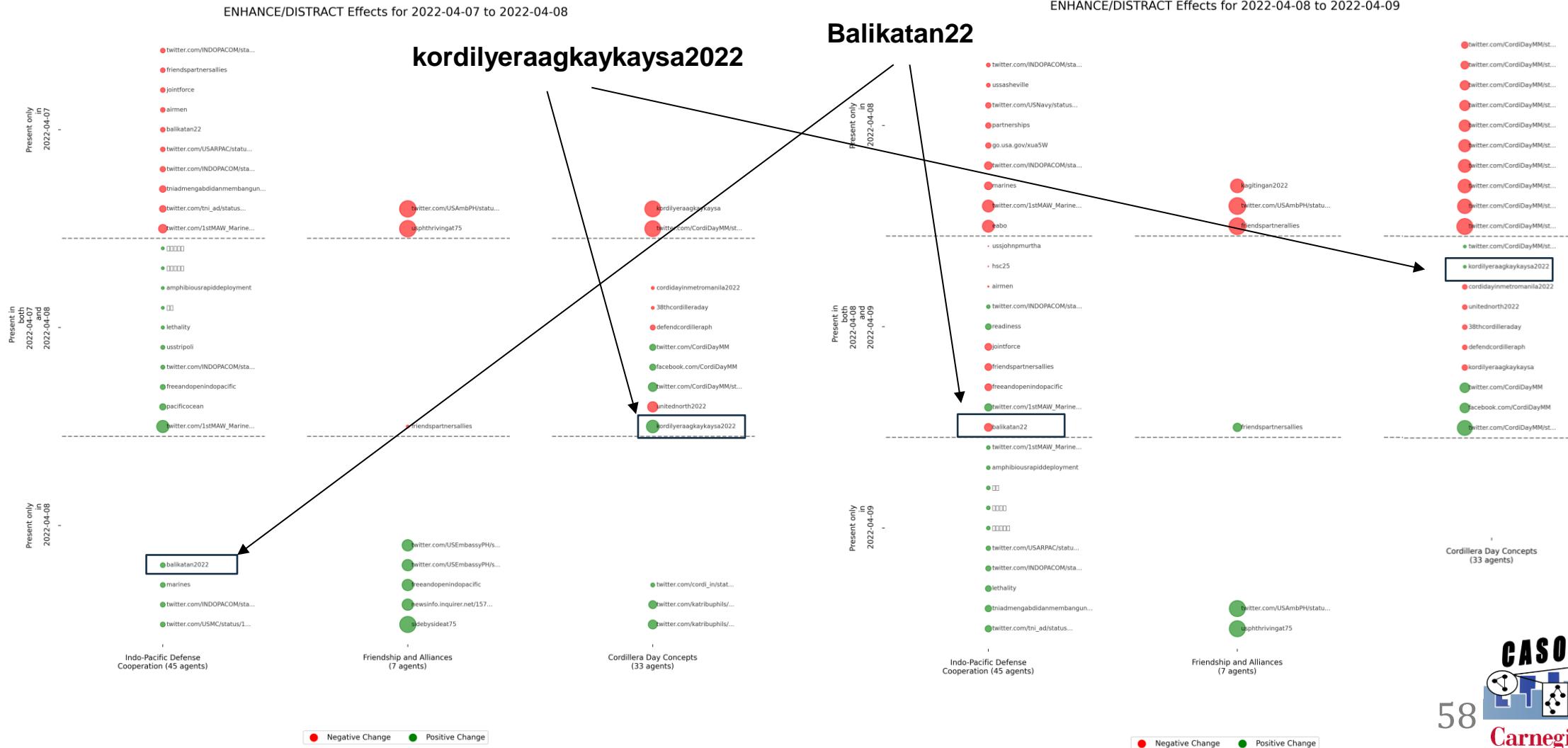
- Dismay effects coincide with the end of the exercise
- Acceleration vs speed

# Balikatan 22 (ENGAGE/DISMISS)

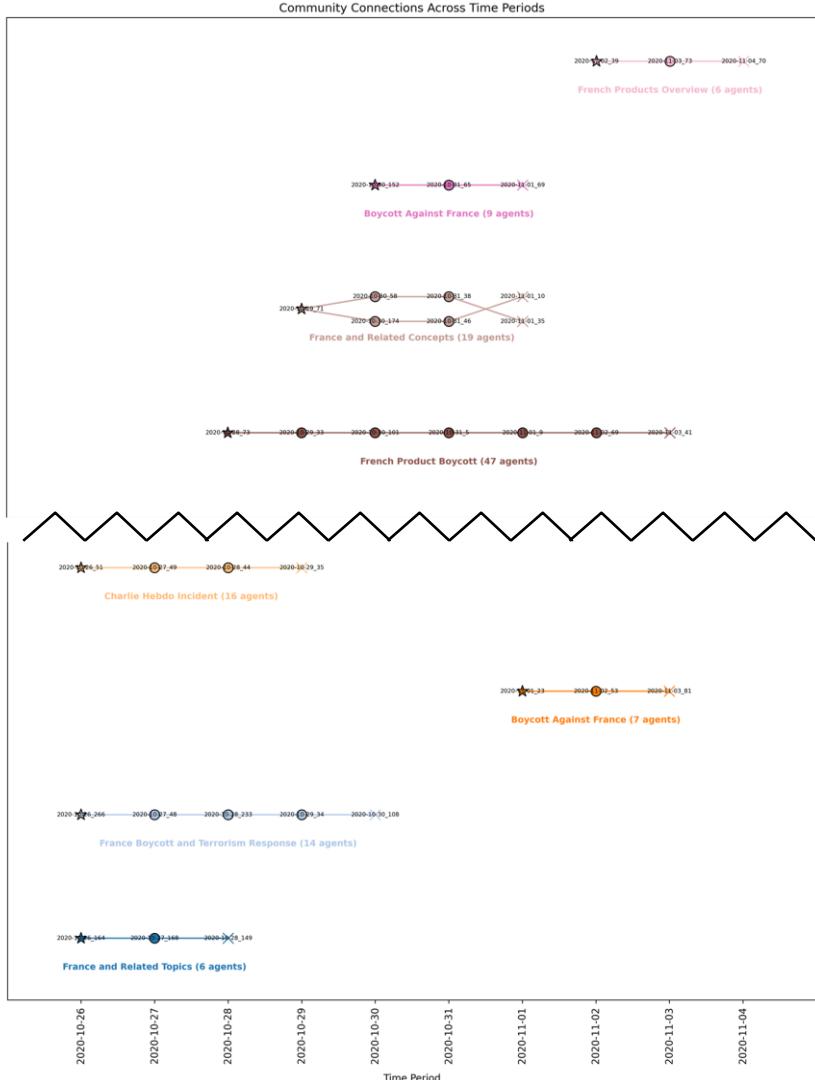


The Cordillera are an indigenous people in the Philippines who commemorate the 1980 killing of a tribal chieftain with Cordillera Day each year on 24 April.

# Balikatan 22 (ENHANCE/DISTRACT)



# French Terror 2020



X/Twitter pull from 26 October – 04 November 2020.

The actual terror attack - in which three people were stabbed to death by a Tunisian man at Notre-Dame de Nice - occurred on 29 October 2020. However, the event was surrounded by the Charlie Hebdo Terror trial (September through December 2020) and was preceded by several weeks of protests around the Muslim world, a Turkish boycott of France, and the beheading of a French history teacher on 16 October 2020.

47 communities  
14 communities > 2-day duration

# French Terror 2020 Timeline

## 2015

7 January – Charlie Hebdo Attacks (12 killed)

8 January – Shooting (1 killed)

9 January – Multiple sieges (4 killed)

13 November – Paris attacks (130 killed)

## 2016

14 July – Nice truck attack (86 killed)

# French Terror 2020 Timeline

## 2020

01 September – Charlie Hebdo republishes caricatures

02 September – Charlie Hebdo trial begins

25 September – Stabbings outside Charlie Hebdo

02 October – “Islam is a religion that is in crisis all over the world today...” – French President Macron

16 October – Samuel Paty beheaded (1 killed)

21 October – “We will not give up cartoons, drawings, even if others back down...” – French President Macron

24 October – Turkish President Erdogan calls for Islamic boycott of France

27 October – Charlie Hebdo published cartoons mocking President Erdogan

28/29 October – Mawlid

**29 October – Notre-Dame de Nice Stabbing (3 killed)**

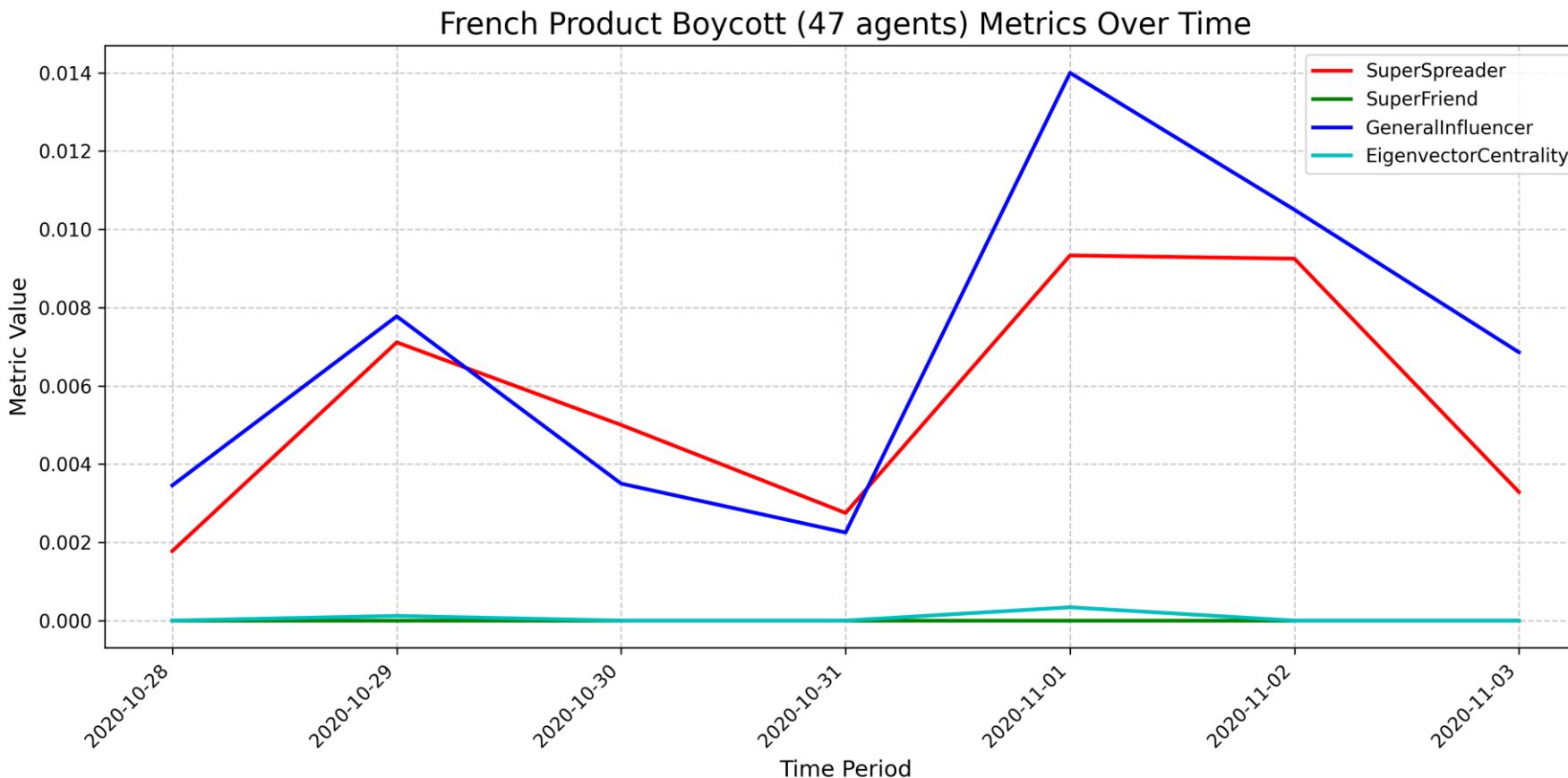
29 October – “France will not give in to terror...” – French President Macron

29 October – Turkey condemns terror attack

31 October – Macron interview with Al-Jazeera accuses Erdogan of “imperial inclinations”

01 November – Anti-French protests in Pakistan, India, Turkey

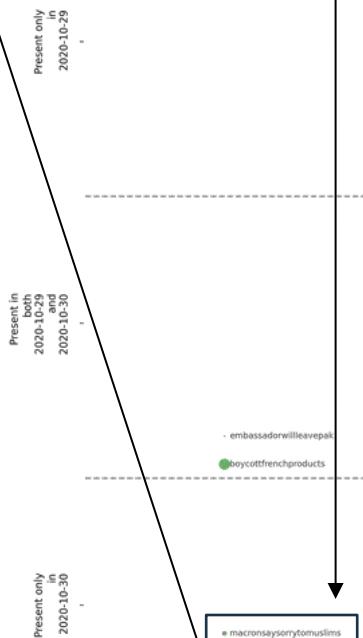
# French Terror 2020 (BACK/NEGATE)



- Increase in spread immediately following the attack
- And again, days later after comments by Macron

# French Terror 2020 (ENHANCE/DISTRACT)

**macronsaysorrytomuslims**  
**shameonyoumacron**



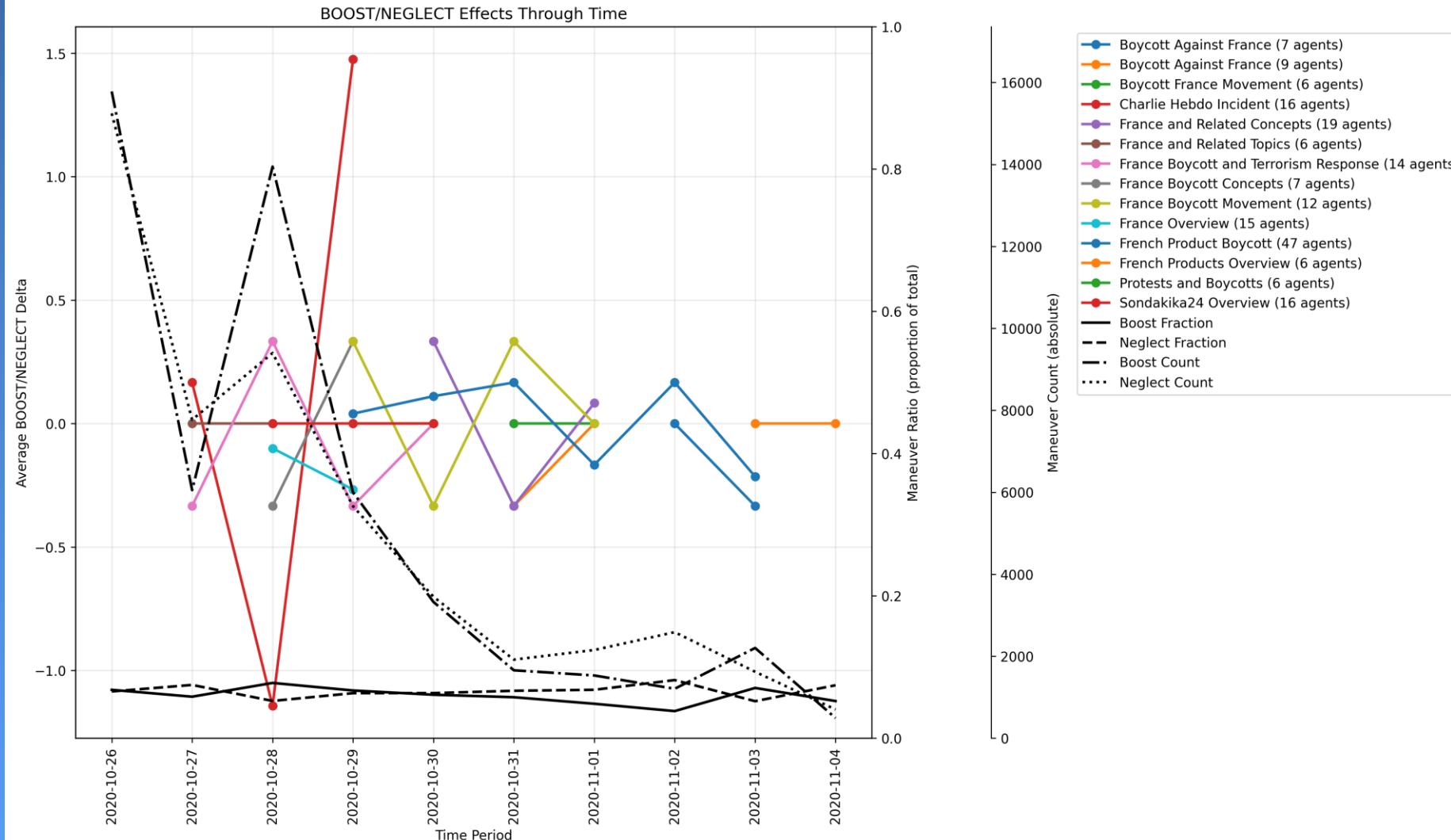
ENHANCE/DISTRACT Effects for 2020-10-29 to 2020-10-30



**macronpsycho**  
**macrongoonemad**

**macron**  
**macronthedevil**

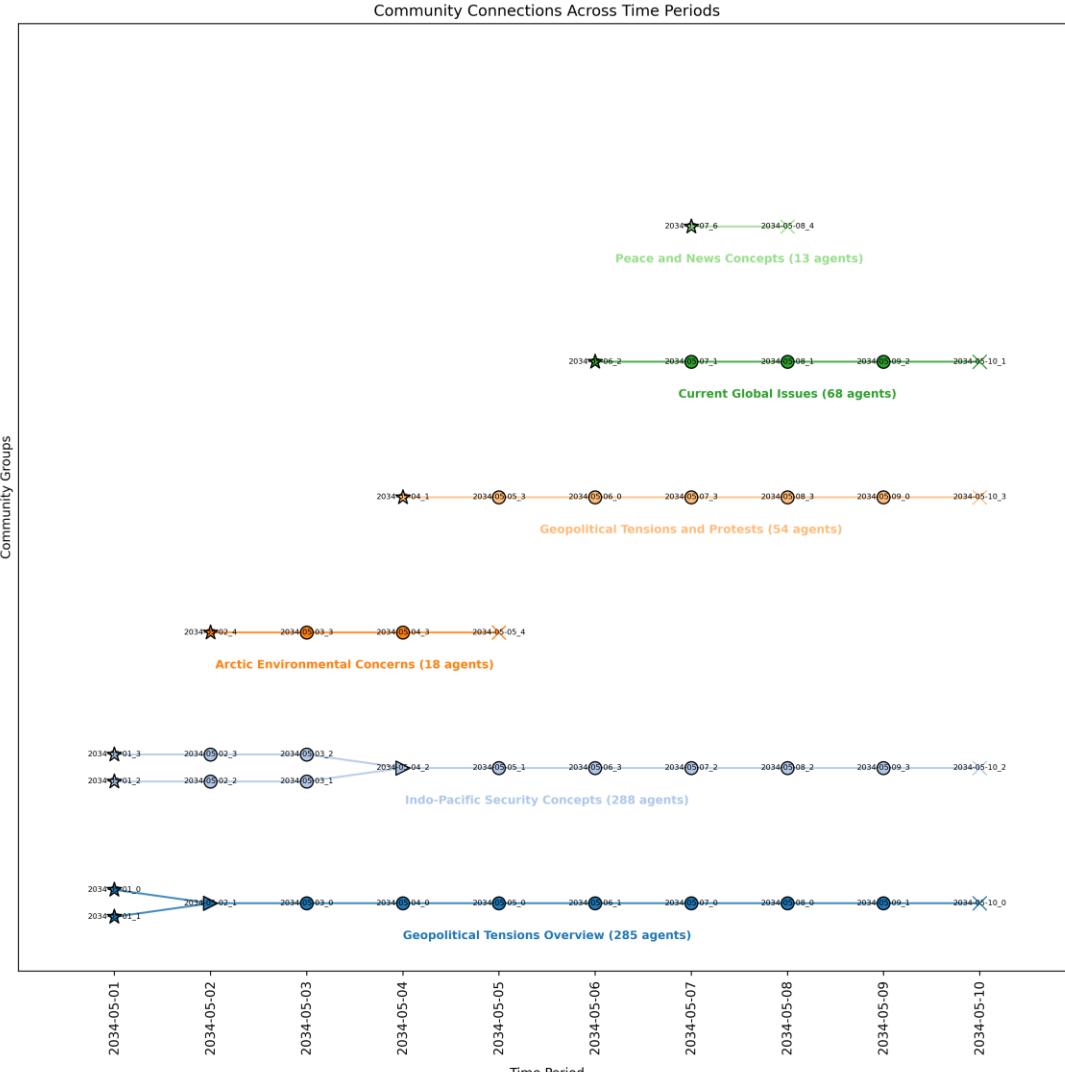
# French Terror 2020 (BOOST/NEGLECT)



## Charlie Hebdo Incident (09 JAN 2015)

- A large bump during the lead up to the 29 October attack – trial discussions
- The group was then severely neglected after Nice stabbings and disappeared

# Synthetic Dataset



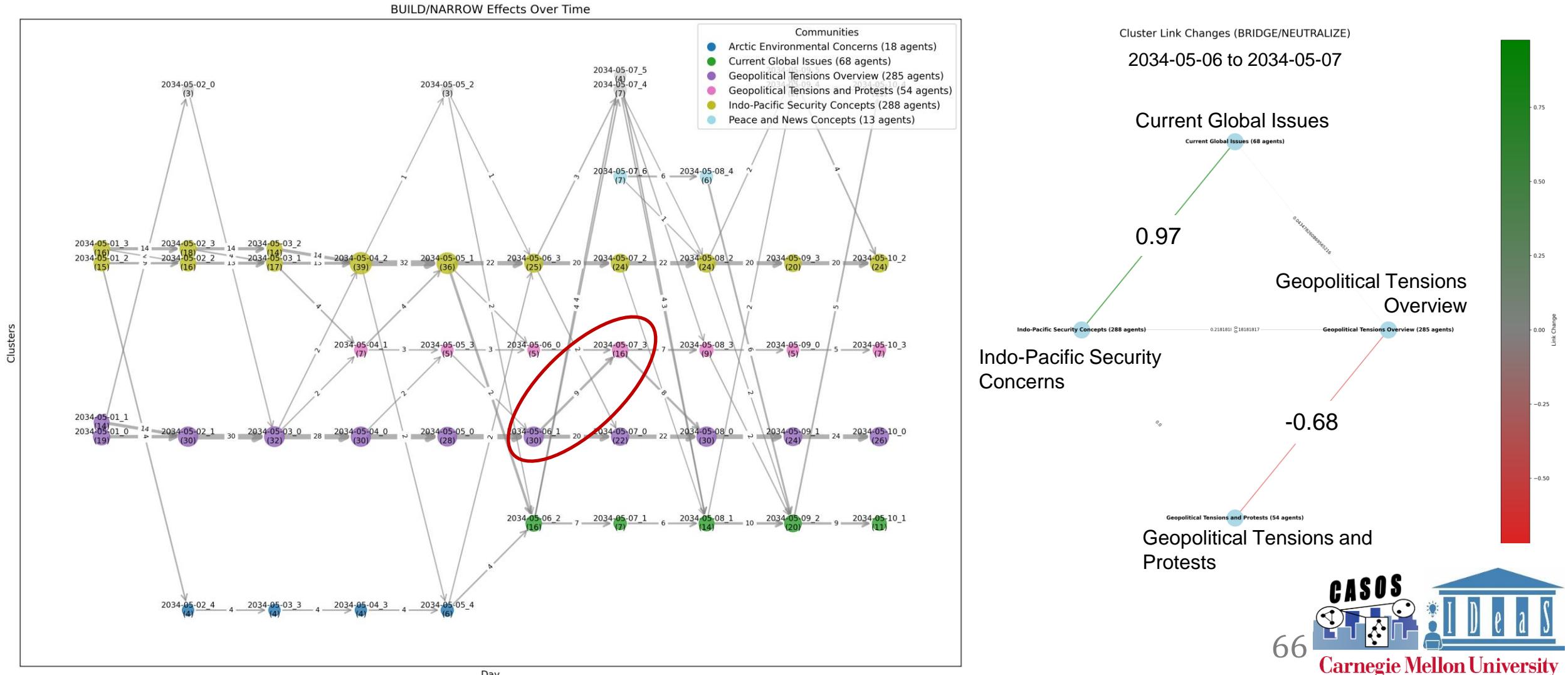
Synthetic (fake) X/Twitter dataset from 01 May – 10 May 2034 built using a scenario created in AESOP.

Events:

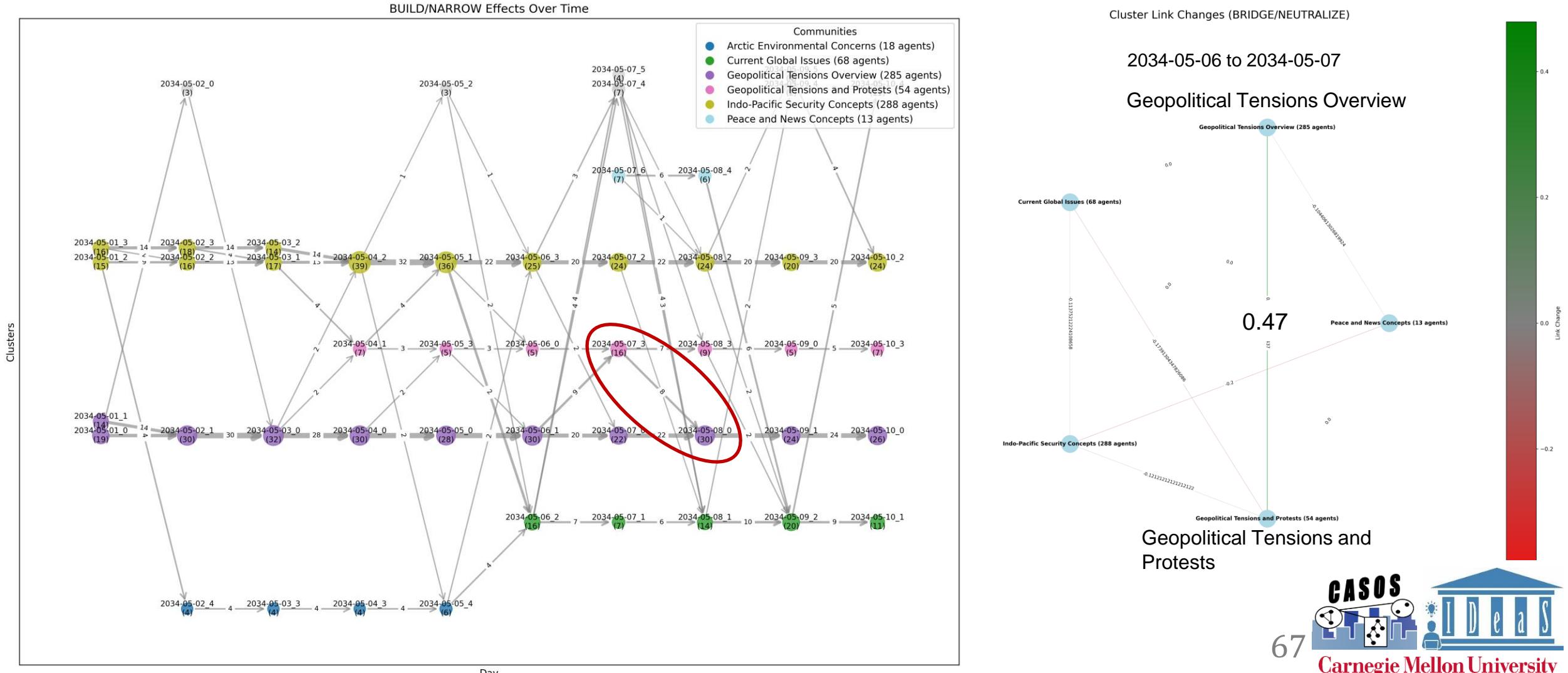
Balikatan 2034, Arctic and South China Sea confrontations, a typhoon, weather/spy balloons, ship and aircraft collisions, student protests – building toward a broader kinetic conflict on 10 May 2034.

6 communities

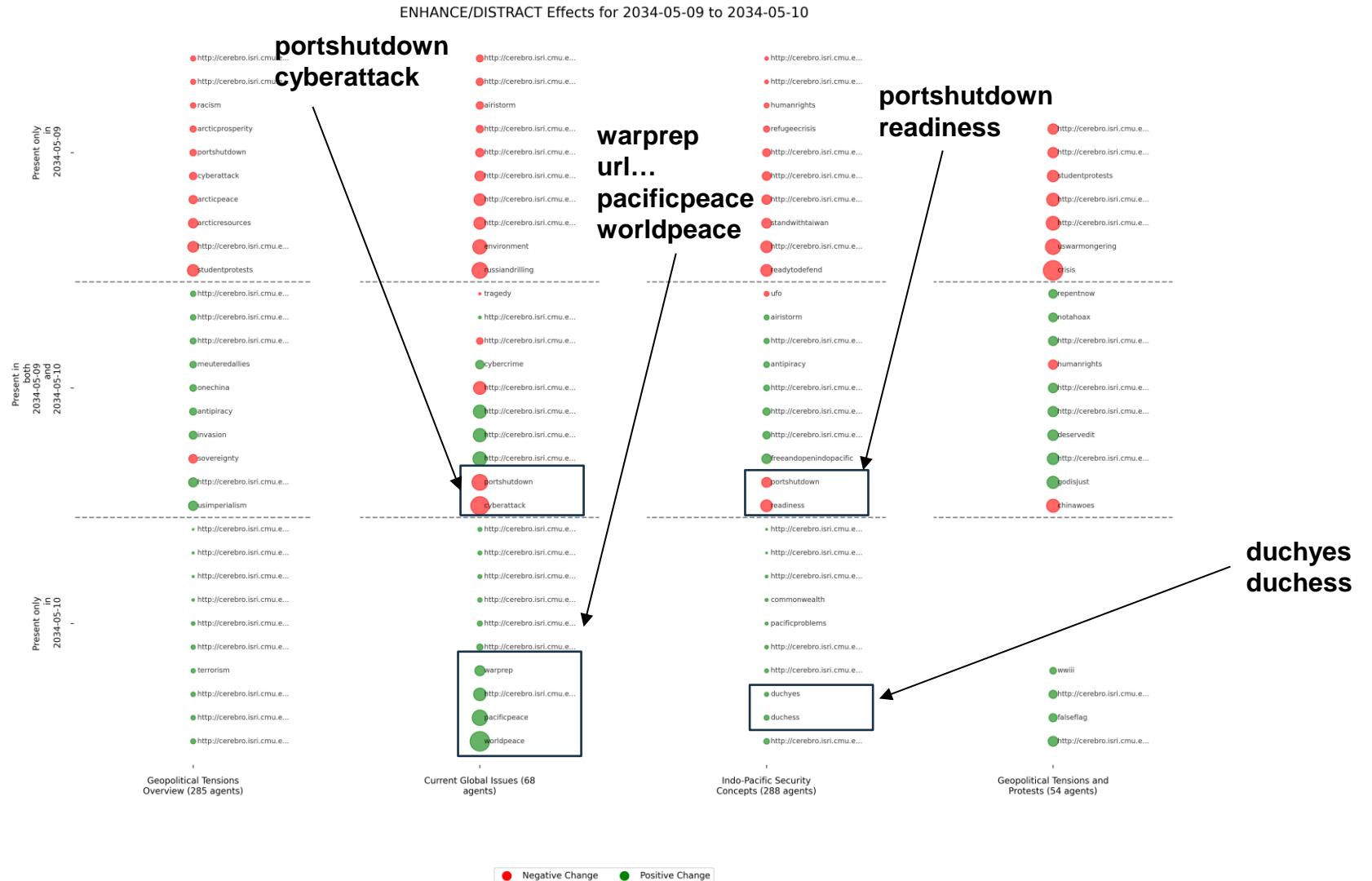
# Synthetic Dataset (BRIDGE/NEUTRALIZE)



# Synthetic Dataset (BRIDGE/NEUTRALIZE)



# Synthetic Dataset (ENHANCE/DISTRACT)



09 May 2034

- Cyber attack shuts down ports

10 May 2034

- Highly increased tensions between major powers
  - Duchess (Lord Privy Seal) tours the area

# Conclusion

- Theoretical and Methodological contributions
  - BEND effects-based detection, which in conjunction with CUE+ allows for a more comprehensive detection mechanism
  - SynTel/X combine traditional simulation logic with LLM-powered message construction to
- Application contributions
  - AI-enabled social media scenario development from existing exercise material or corpora
  - A draft data standard for passing social media objects and scenarios
  - AI-enabled X/Twitter and Telegram social media generation for exercise training based on synthetic templates

# Conclusion

- AESOP reduces the time required for social media scenario design from months to days
- SynTel/X produces realistic, interactive X and Telegram datasets based on AESOP scenarios (~1,000 messages/minute)
- BEND-Effects allows analysts to find BEND maneuvers based on their effects within a dataset and assess the effectiveness of BEND maneuvers

# Limitations and Justifications

- BEND Framework Improvements
  - Additional methods for measuring BEND maneuvers above baseline are required – residual statistics will be more important than net maneuver counts
  - Finding topic-oriented groups through time remains a difficult process
  - There is currently no method for directly associating observed BEND effects with any single message BEND maneuver
- BEND Scenario Development
  - Without a population model in the simulation, agent opinions remain static; however, AESOP training is not static as training audience decisions and actions impact all additional synthetic data
  - Training data should always be a snapshot of the full picture as most social media APIs do not allow for decisively pulling all data possible

# Funding and Disclaimer

- ❑ The research for this thesis was supported in part by the Office of Naval Research (ONR) under grant N00014182106, the United States Army under grant W911NF20D0002, and the center for Informed Democracy and Social-cybersecurity (IDeaS). The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ONR, the United States Army, or the US Government.
- ❑ The appearance of U.S. Department of Defense (DoD) visual information does not imply or constitute DoD endorsement.

# Questions

# BACKUP SLIDES

# Future Work (BEND-Effects)

Each of the maneuver metrics for BEND-effects currently lacks proper thresholding and significance measures. While the existence of maneuver effects is firmly established, their degree of importance remains left to the interpretation of the analyst. This is the difference between evidence of bridging and evidence of bridging that we care about.

Additionally, alternatives to Leiden should be explored as finding topic-oriented groups is not a solved problem. The BEND-effects calculations could also be better expressed by first calculating a baseline of the entire corpus at each time period of the given metric and then expressing individual clusters change over time as a residual above and beyond the baseline. This would help ensure the BEND-effect measurement for a cluster is greater than chance and is not a byproduct of a global effect.

Finally, with BEND-effects we know what effects were felt by communities and with the traditional cue-based BEND detection we can predict what communities intended to do - the logical next step is to map the one to the other. We can plot each community's experienced effect against the intended effects of all other community and in this way attempt to draw conclusions about which intended maneuver campaigns were effective.

# Future Work (SynTel/X)

First, the agents in the current model have opinions matrices that map their stance on each topic in the scenario; however, there is currently no population modeling implemented to feed these matrices nor is there information diffusion implemented such that these stances change over time. Thus, while the actors respond in accordance with their opinions, the simulation is currently a perfect model of arguing on Internet forums - no one ever changes their mind.

Furthermore, more validation should be done. If the goal is simply to train on BEND maneuver detection and response, then the current validation network metrics and BEND evaluations show that SynX is sufficient. However, training and instruction on other aspects of social networks might require additional, unevaluated, network properties. This includes more research into matching the Agent x Agent modularity against real world datasets.

Finally, work is also required within the BEND evaluations. While the AESOP scenario comes with images and news sites, the current ORA-Pro and NetMapper BEND evaluation is done only on the meta-data and the text. Future detection should include an evaluation of the included images or referenced URLs.

# Future Work (AESOP)

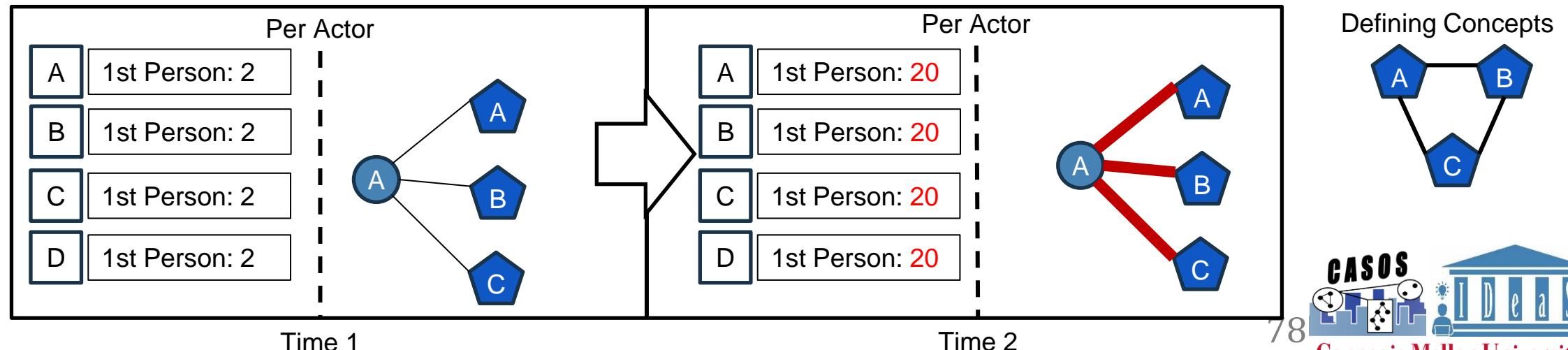
AESOP should be tied with retrieval-augmented generation (RAG) to limit hallucinations and provide additional support to the mad-lib method of prompt-engineering used to scaffold the planner. RAG would allow planners to rely upon a corpus of validated material to supplement the LLMs trained knowledge.\cite{lewis2020rag}

AESOP should integrate the ability to populate personas drawn from population modeling of a demographic. This would provide an effective method for determining the types of personas that are representative of a demographic, rather than relying upon the planner to select and fill these attributes by hand.

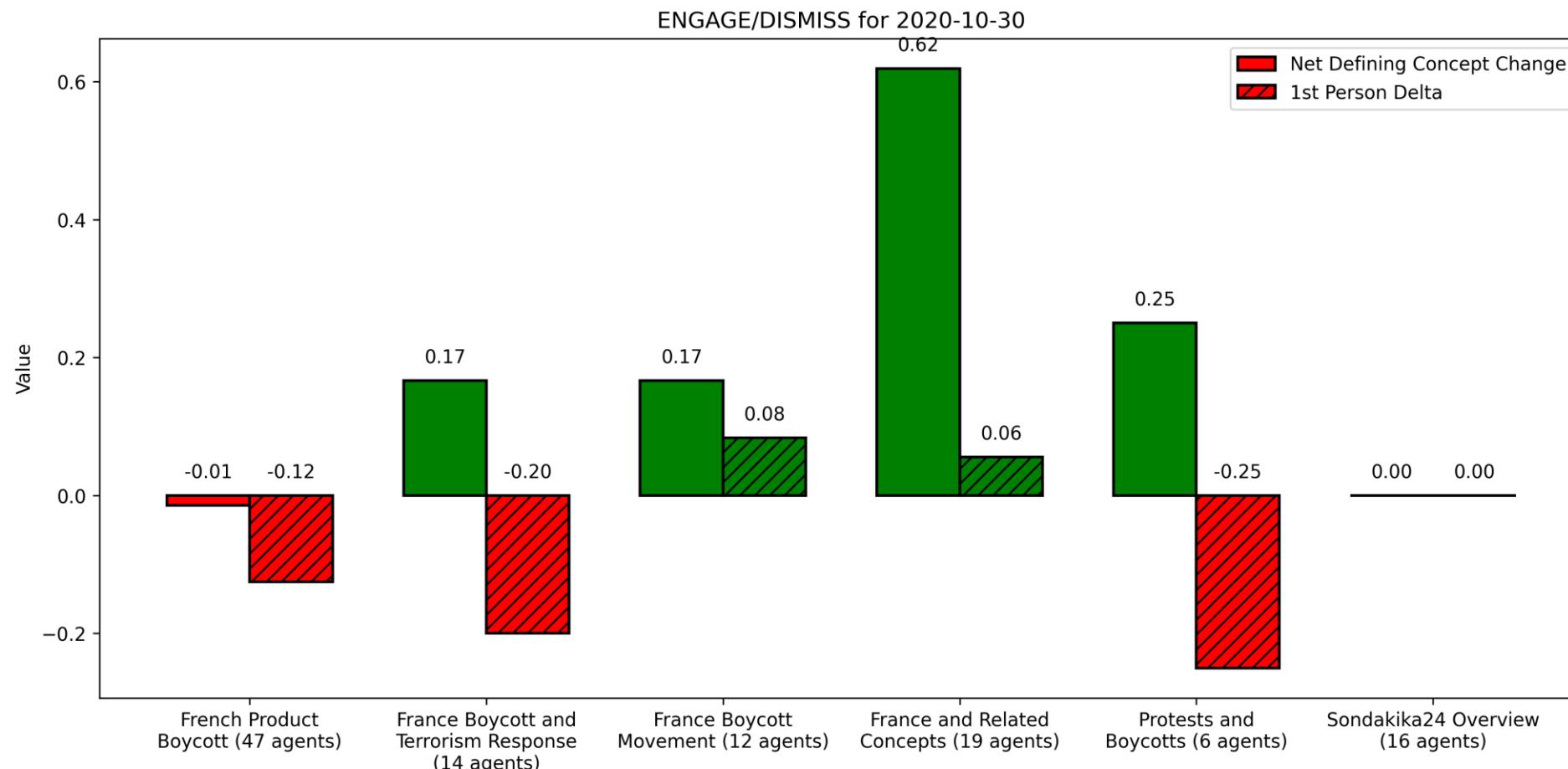
AESOP should allow for the quick integration of multiple LLMs for specialized purposes. Currently, the LLM is set at the global level within AESOP and this pushes AESOP toward using large generalized models - favoring commercial models over local LLMs. The ability for AESOP to operate with purely local LLMs does already exist but specifying multiple, smaller, specialized LLMs for agent creation versus journal/article writing versus press releases would increase quality while still allowing for disconnected operation.

# ENGAGE / DISMISS

Name	Definition	Detect
Engage	Discussion or actions that increase the relevance of the topic to the reader often by providing anecdotes or enabling direct participation and so suggesting that the reader can impact the topic or will be impacted by it	<p><b>Target:</b> Each Agent per TOG Community</p> <p><b>Measures:</b> A) 1<sup>st</sup> Person pronoun usage, 2) Defining Concept usage  <b>(Defining Concepts</b> are those concepts found in both time 1 and time 2)</p> <p><b>Network(s):</b> Messages (NetMapper CUEs), Agent x Concept</p> <p><b>Indicator:</b> A) Increase, B) Increase</p>
Dismiss	Discussion or actions that decrease the relevance of the topic to the reader often by providing stories or information that suggest that the reader cannot impact a topic or be impacted by it	<p><b>Indicator:</b> A) Decrease, B) Decrease</p>



# ENGAGE / DISMISS



# BACK / NEGATE

Name	Definition	Detect
Back	Discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	<b>Target:</b> All Agents <b>Measures:</b> Centrality across all three – PageRank, Out Degree Centrality (ORA Style), Modularity/Vitality Hub* <b>Network(s):</b> Agent x Agent (Strong Ties) <b>Indicator:</b> Positive change from t1 to t2
Negate	Discussion or actions that decrease the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	<b>Indicator:</b> Negative change from t1 to t2

## Already present in ORA-PRO as super-spreaders:

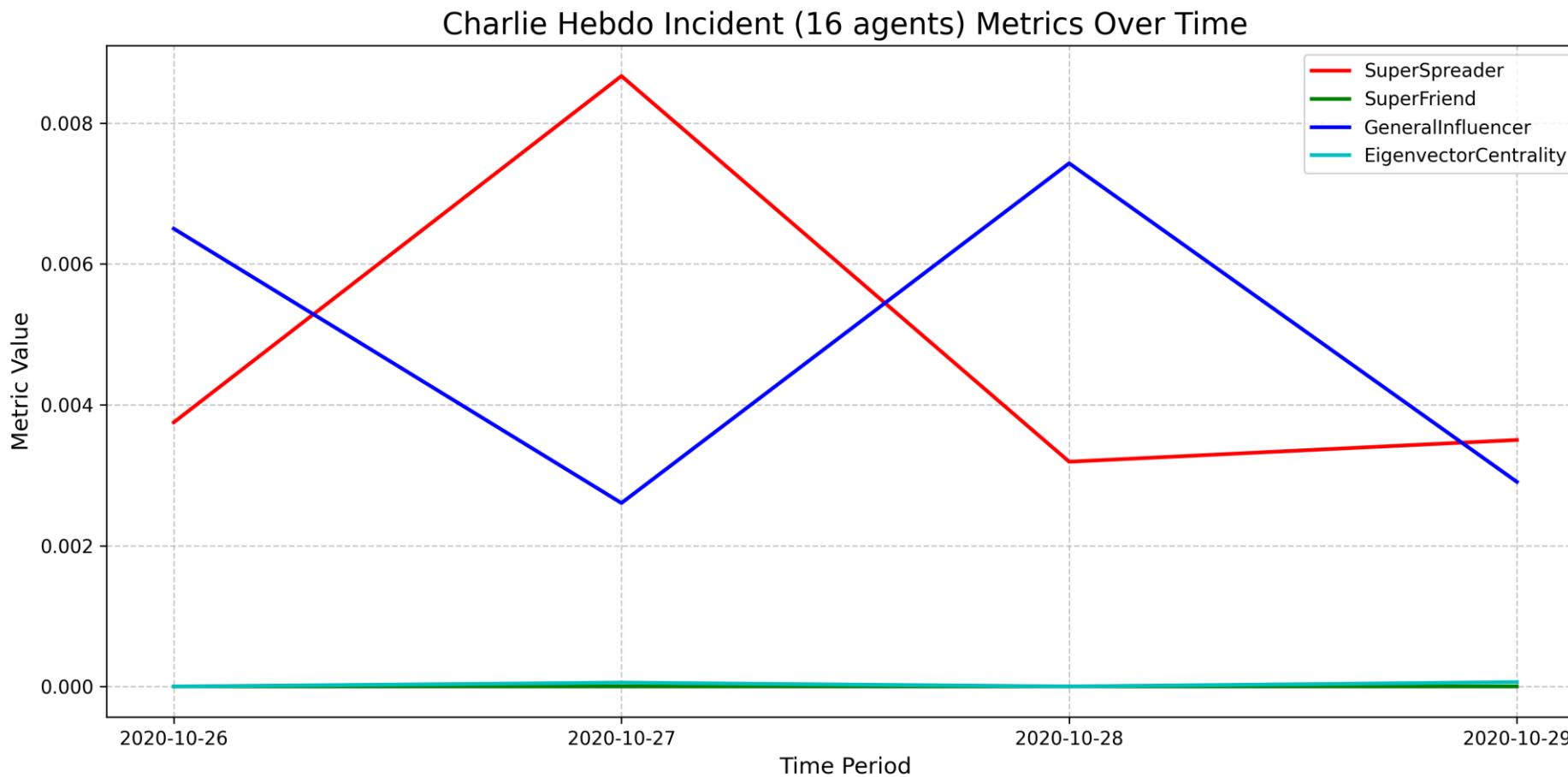
Posts Often (out-degree centrality) [Agent x Tweet - Sender]

Often Spread by Others (out-degree centrality) [Agent x Agent - All Propagation]

Iteratively Spread by Others (page rank centrality) [degreeType=outDegree] [Agent x Agent - All Propagation]

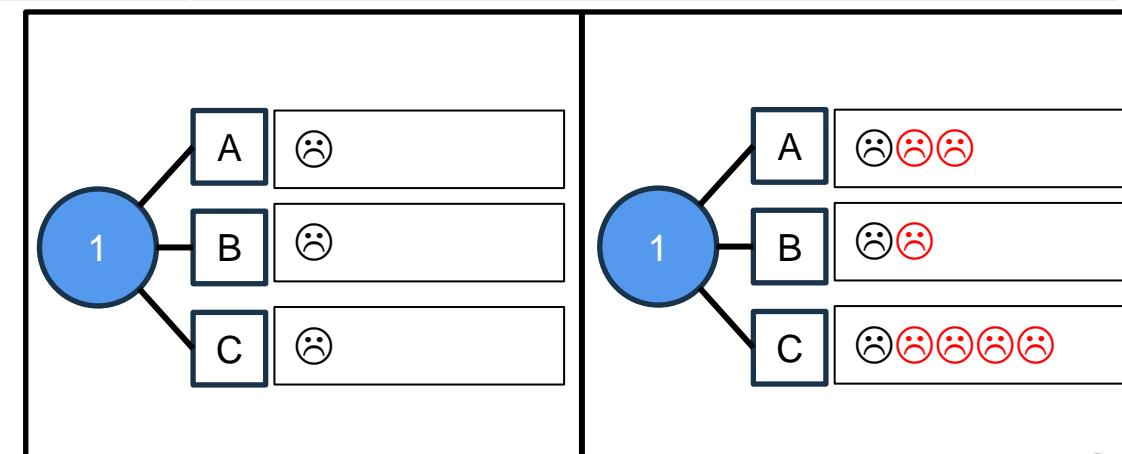
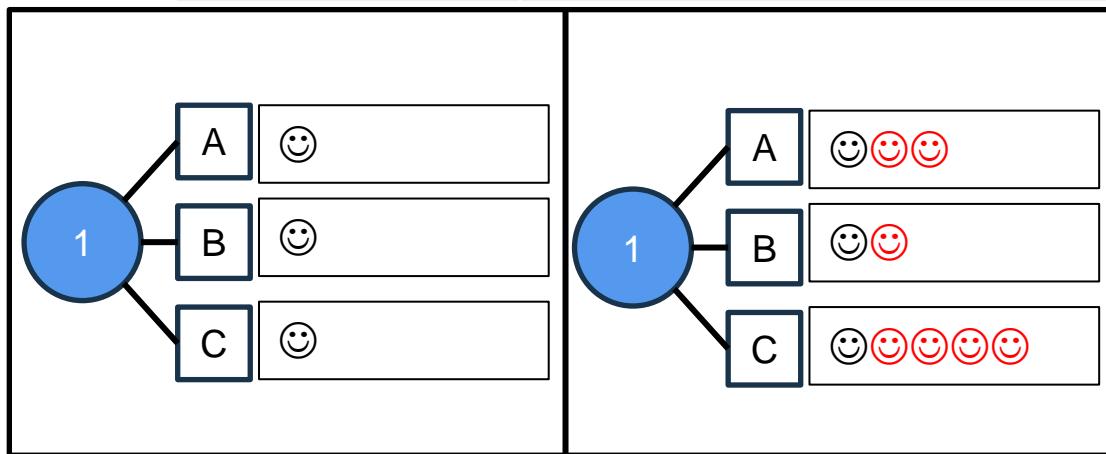
Often Spread by Groups of Others (member of large k-core) [degreeType=outDegree] [Agent x Agent - All Propagation]

# BACK / NEGATE

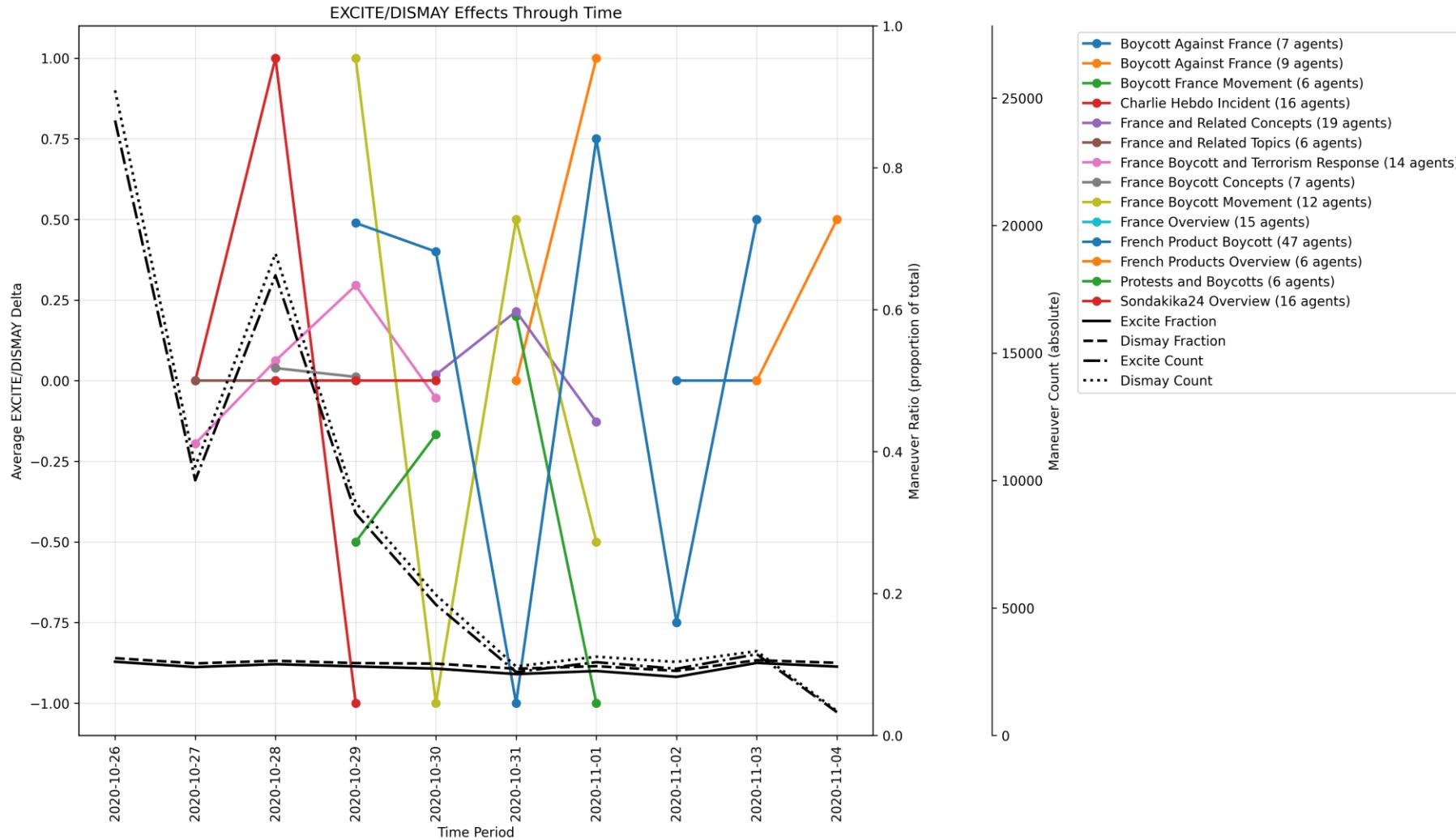


# EXCITE / DISMAY

Name	Definition	Detect
Excite	Discussion or actions related to a <b>community or topic</b> that cause the reader to experience a positive emotion such as joy, happiness, liking, or excitement	<p><b>Target:</b> TOG Communities</p> <p><b>Measures:</b> Messages of TOG Cluster agents in t2 and in t1 – evaluate for 4 positive/negative CUEs through NetMapper</p> <p><b>Network(s):</b> CUEs from NetMapper</p> <p><b>Indicator:</b> Increase in positive CUEs</p>
Dismay	Discussion or actions related to a community or topic that cause the reader to experience a negative emotion such as worry, sadness, disliking, anger, despair, or fear	<b>Indicator:</b> Increase in negative CUEs

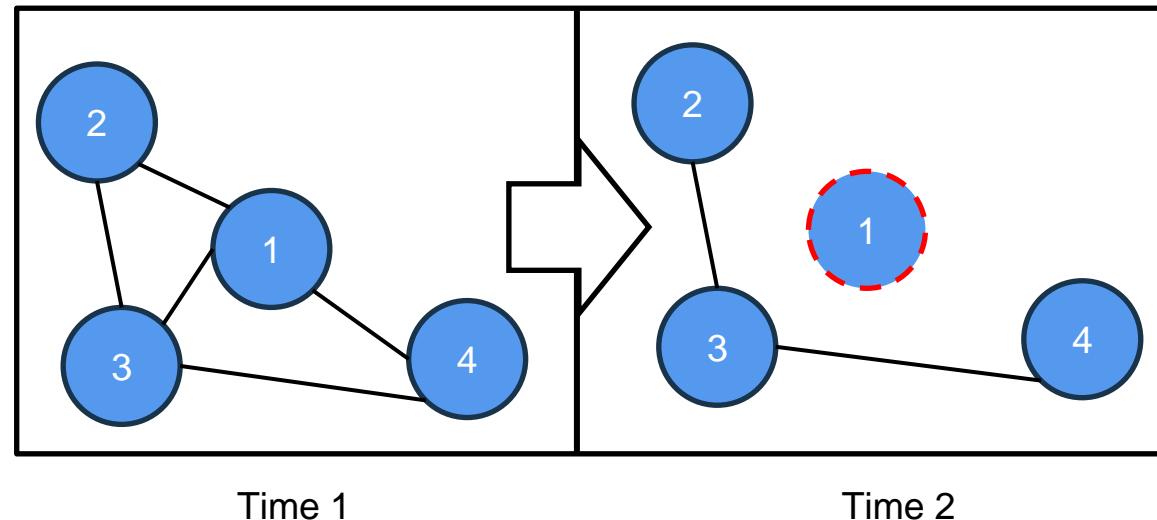
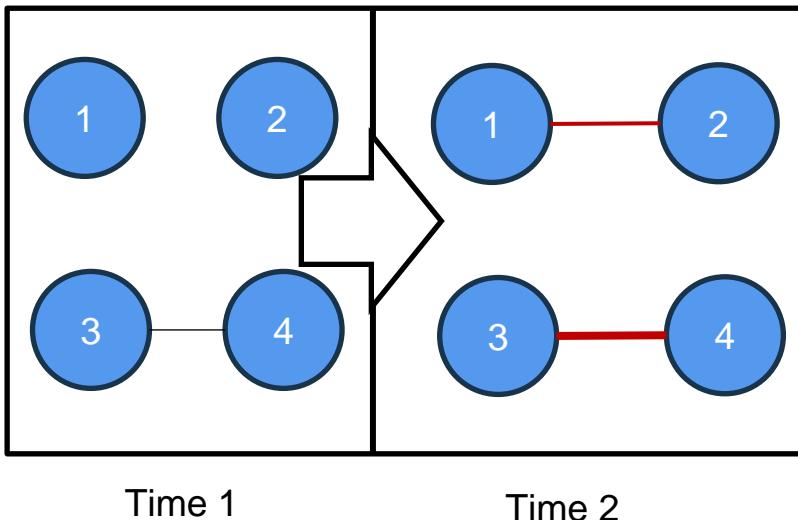


# EXCITE/DISMAY

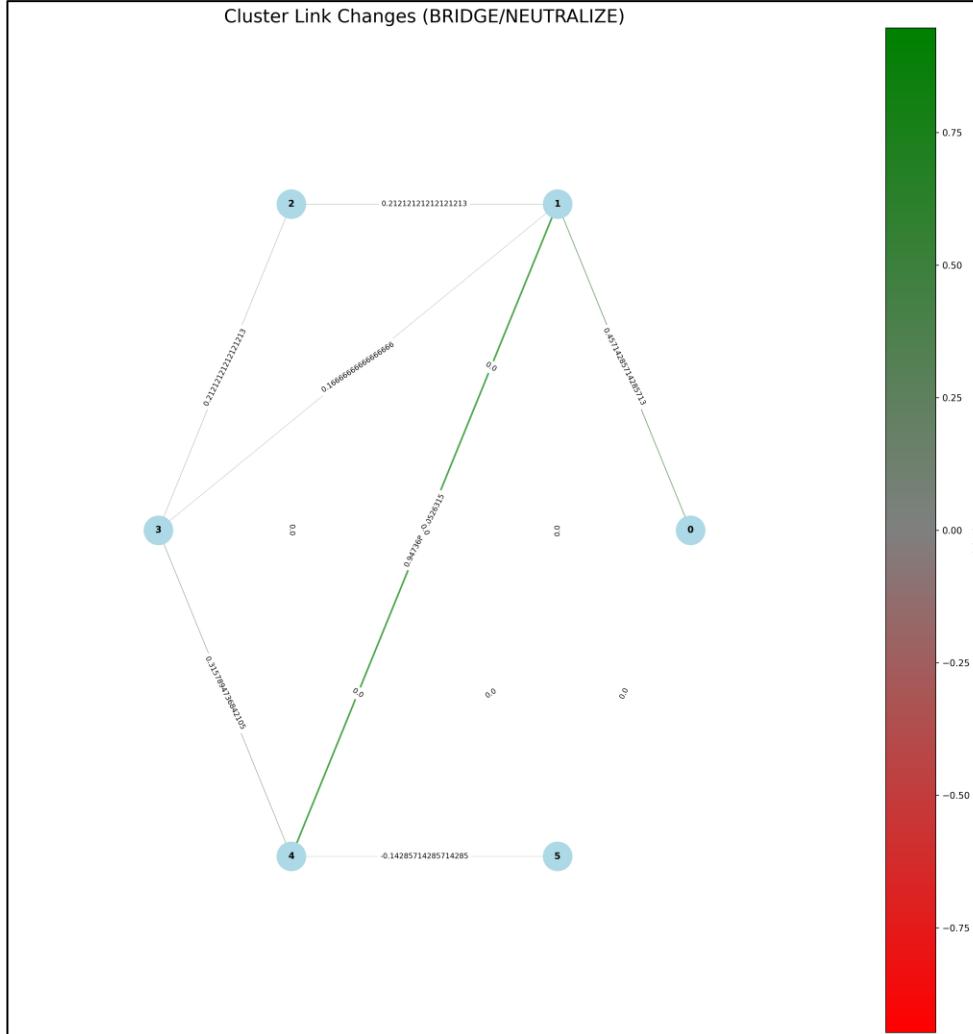
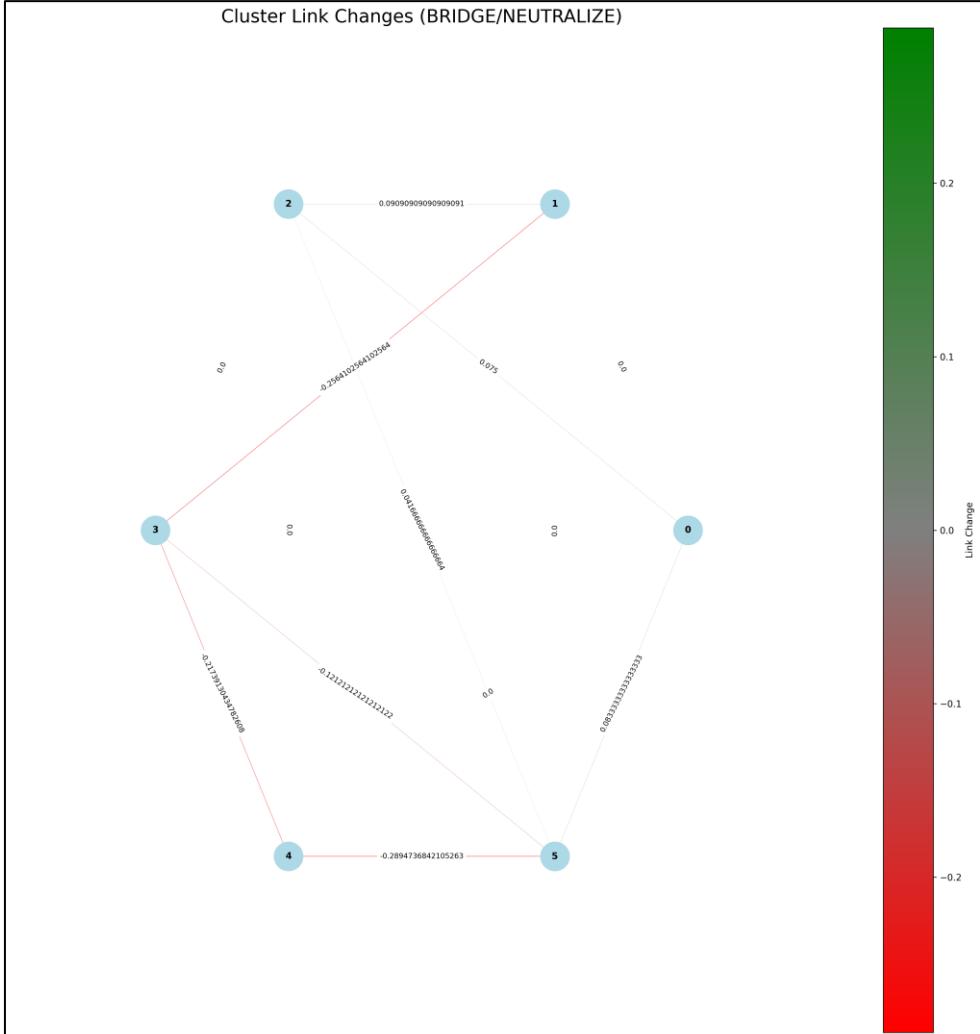


# BRIDGE / NEUTRALIZE

Name	Definition	Detect
Bridge	Discussion or actions that build a connection between two or more groups or create the appearance of such a connection	<b>Target:</b> Each pair of TOG Communities <b>Measures:</b> Meta-node (TOG Community) ties <b>Network(s):</b> Agent x Agent (Strong Ties union Concepts) <b>Indicator:</b> Increase from t1 to t2
Neutralize	Discussion or actions that cause a group to be, or appear to be, no longer of relevance, e.g., because it was dismantled	<b>Indicator:</b> Decrease from t1 to t2



# BRIDGE / NEUTRALIZE



# Data Standard

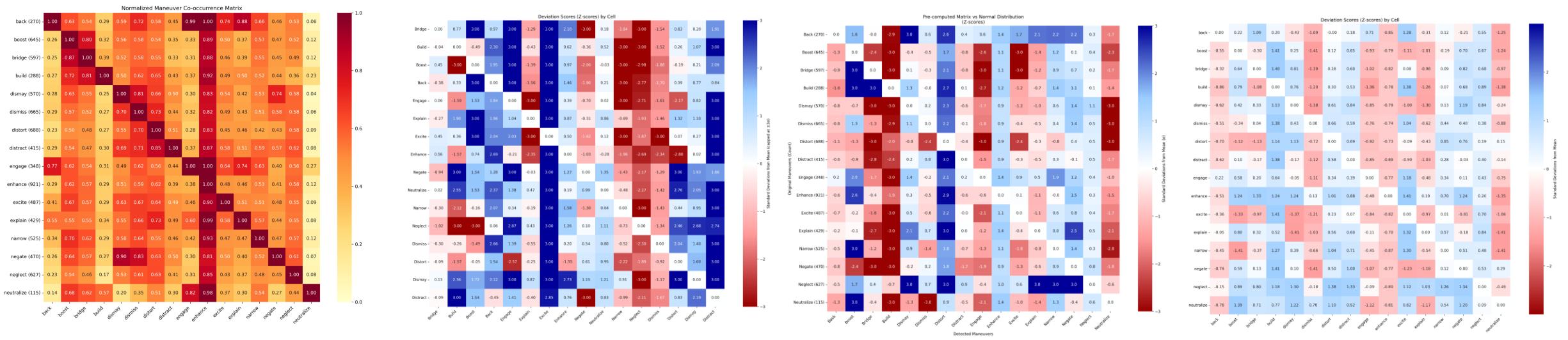
Standard (Acronym)	Purpose	Key Components	Primary Use Case	Data Format	Adoption	Website Reference
Structured Threat Information Expression (STIX)	Cyber threat intelligence, tracking actors, events, and campaigns	Identity, Threat Actor, Campaign, Incident, Observables	Tracking cyber threats, disinformation campaigns, and influence operations	JSON, XML	Cybersecurity, threat intelligence, law enforcement	<a href="https://oasis-open.github.io/cti-documentation/">https://oasis-open.github.io/cti-documentation/</a>
Activity Streams 2.0 (AS2)	Social media activity representation	Actor, Object, Verb, Target, Group	Modeling social media interactions and behavioral dynamics	JSON-LD, ActivityStreams format	Decentralized social media platforms and Web3 applications	<a href="https://www.w3.org/TR/activitystreams-core/">https://www.w3.org/TR/activitystreams-core/</a>
Friend of a Friend (FOAF)	Describes social relationships and online identities	Person, Group, OnlineAccount, Knows	Semantic web applications and social network analysis	RDF (Resource Description Framework)	Semantic web, academic research, social network modeling	<a href="http://xmlns.com/foaf/spec/">http://xmlns.com/foaf/spec/</a>
Semantically-Interlinked Online Communities (SIOC)	Defines social media discussions and interactions	UserAccount, Post, Forum, Community, Topic	Online community tracking and engagement analysis	RDF, XML	Online discussion forums, enterprise knowledge management	<a href="https://web.archive.org/web/20220331224416/http://sioc-project.org/">https://web.archive.org/web/20220331224416/http://sioc-project.org/</a>
Intelligence Community Information Resource Metadata (IC-IRM)	Intelligence metadata for information security, discovery, and analysis	Resource descriptions, classification, access control, XML/HTML encoding	Intelligence and national security data management	XML, HTML metadata (IC)	US Intelligence Community	<a href="https://www.dni.gov/index.php/who-we-are/organizations/ic-cio/ic-technical-specifications/information-resource-metadata">https://www.dni.gov/index.php/who-we-are/organizations/ic-cio/ic-technical-specifications/information-resource-metadata</a>
Unnamed Synthetic Social Media Scenario Data Standard (USS-MSDS)	Describes actors, events, groups, and narratives within a scenario	Actor, Account, Event, Group, Topic, Narrative, Article, URL	Communicating the structure and component of a social media scenario from the scenario planner to a synthetic generator	JSON	NONE	NONE

# Similarities to Meliorator

- ❑ Meliorator – The US, the Netherlands, and Canada accused RT affiliates of using a sophisticated AI-enhanced software package to disseminate information to/about US, Poland, Germany, the Netherlands, Spain, Ukraine, and Israel
- ❑ Brigadir GUI for Taras – “Souls” tab for false identities and “Thoughts” tab for scenarios and actions
- ❑ Joint Cybersecurity Advisory: <https://www.ic3.gov/CSA/2024/240709.pdf>

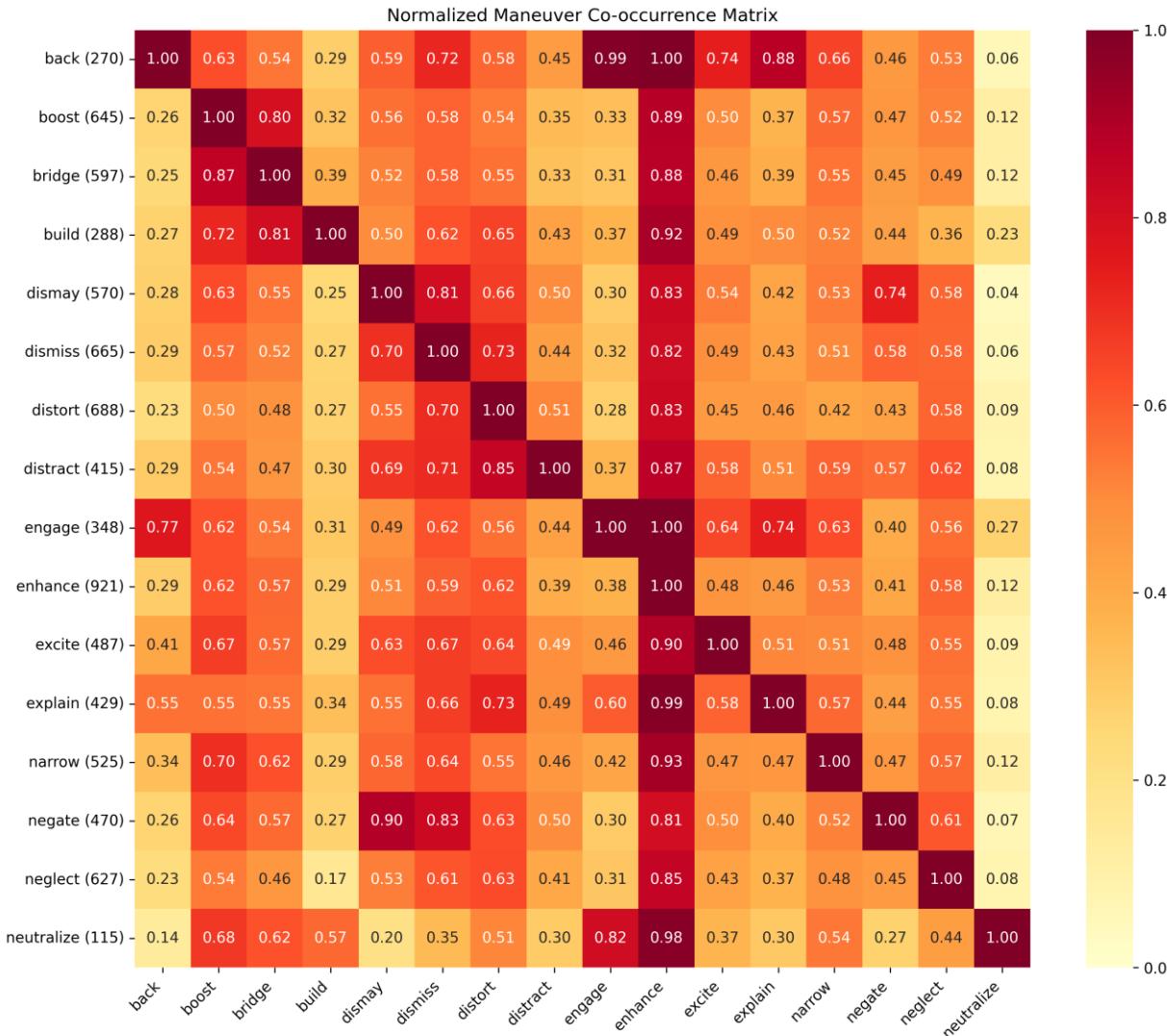
# Content: Is my dataset any good?

Are there realistic looking BEND maneuvers in the data?



# Content: Is my dataset any good?

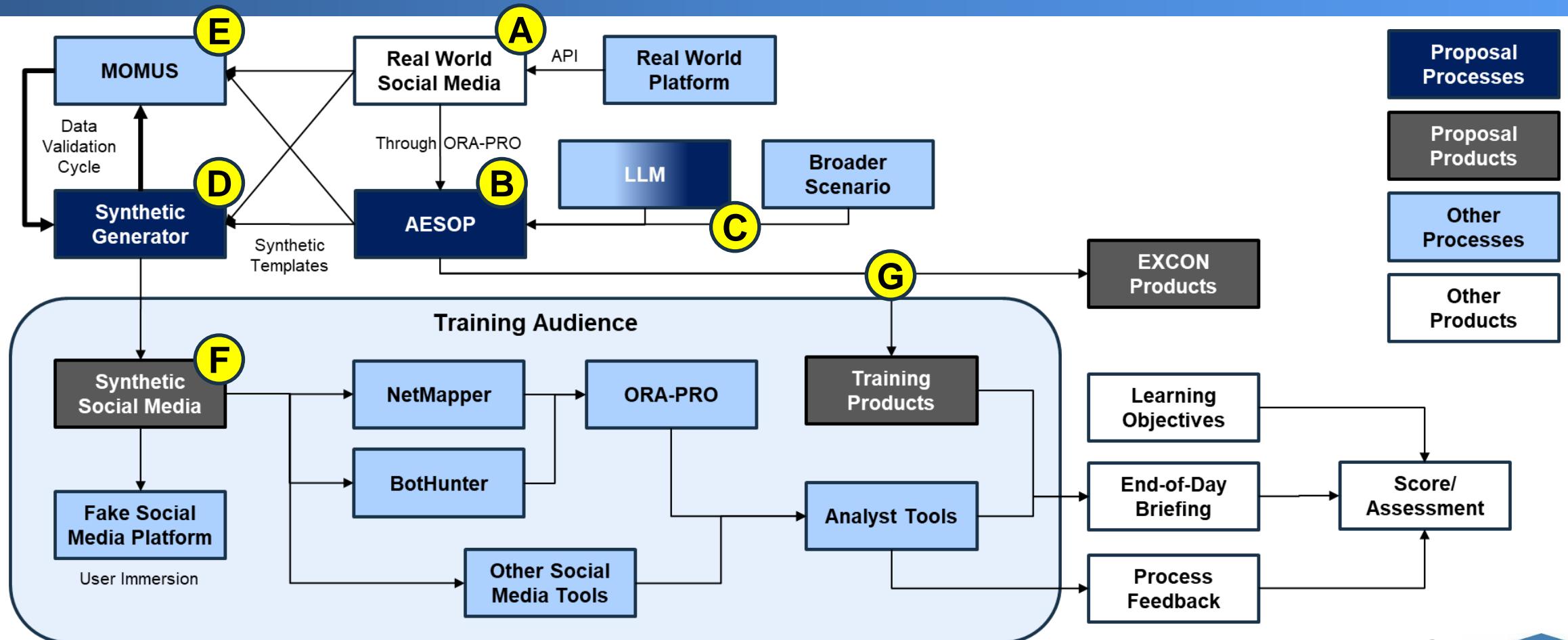
Are there realistic looking BEND maneuvers in the data?



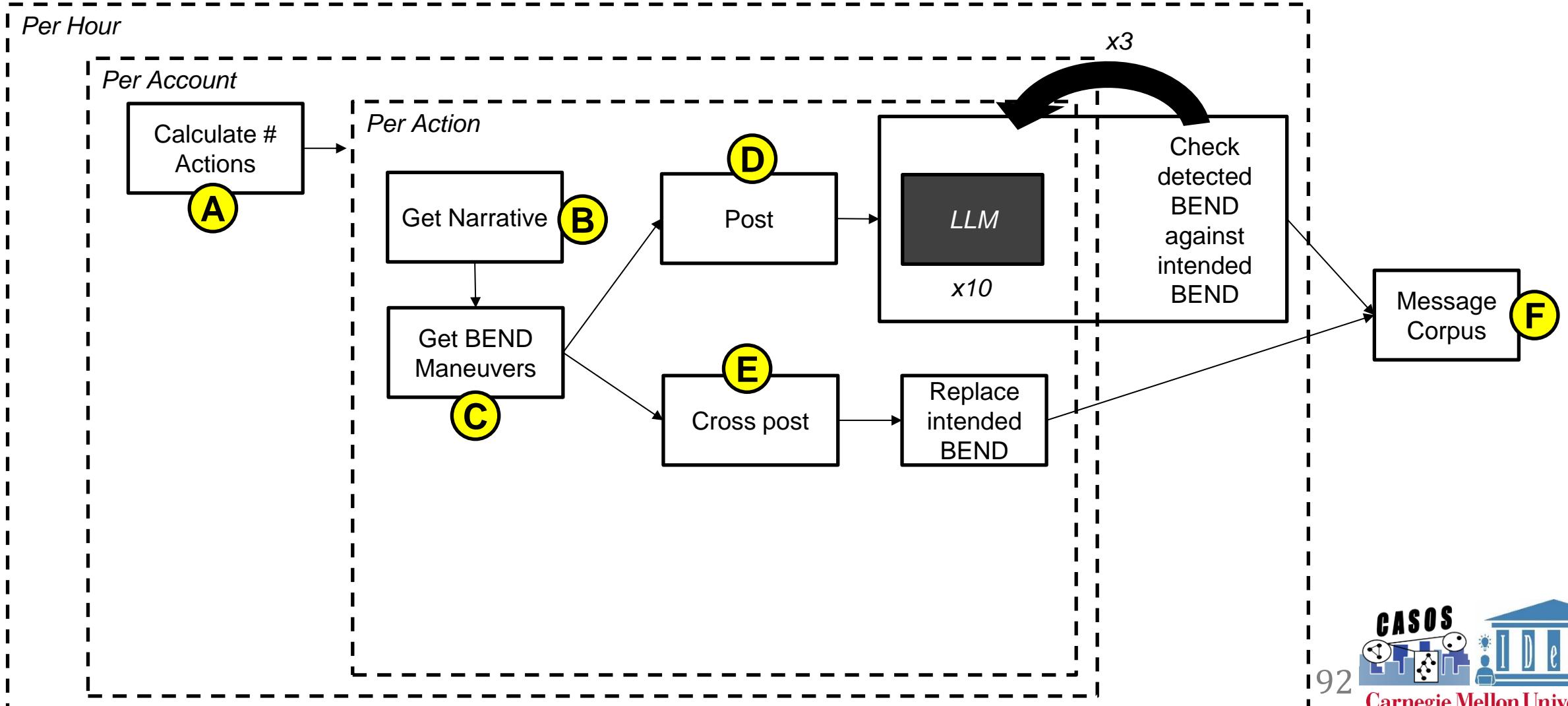
# Research Roadmap Timeline

		Spring 2024	Summer 2024	Fall 2024	Spring 2025
<b>Doctrinal Synthesis</b>	Doctrinal Survey	Green			
	Overlay Mapping		Red		
<b>BEND Detection</b>	Effects-based	Yellow			
	Re-run CUE+		Red		
	Countermeasure Sim	Green		Red	
	Overall Analysis				Red
<b>BEND Scenario Development</b>	Synthetic Templates	Green			
	Telegram Generation	Yellow			
	Twitter/X Generation		Red		
	Facebook Generation			Red	
<b>Final Wrap-up</b>				Red	
Not Scheduled	Work Begun	Yellow			
Nearing Completion	Planned Work	Red			

# Realistic Training



# Telegram Generation



# AI-Enabled Scenario Orchestration and Planning (AESOP)

