THE CUBISM PROJECT:
BELIEF AND SENTIMENT CLASSIFICATION

TAC 2016 Workshop
November, 2016
Gaithersburg, Maryland USA

Adam Dalton, Morgan Wixted, and Yorick Wilks
Institute for Human and Machine Cognition, Ocala, FL

Meenakshi Alagesan, Gregorios Katsios, Ananya Subburathinam, and Tomek Strzalkowski
State University of New York - University at Albany

Belief and Sentiment Evaluation

- The basis of the evaluation are private state tuples (PSTs), which are 4-tuples of the following form:

  (source-entity, target-object, value, provenance-list)
- The target can be any relation, or any event (the target can also be any entity for sentiment)
- English, Chinese, and Spanish
- The value is:
  - A sentiment value (positive, negative), or
  - A belief value (CB, NCB, ROB)
- **Participants had access to files specifying EREs of interest; this includes in-document co-reference of entity mentions and event mentions**

Main Takeaways

- Belief
  - Make use of the existing structure of Rich ERE annotations
  - Evaluate impact of communities of belief created based on that structure
  - Evaluate the impact of dialogue act features
  - Language agnostic
- Sentiment
  - Adapted an affect calculus algorithm originally designed to compute affect in metaphors
  - Combine syntactic and semantic structure with base polarity values of words and phrases
  - Base polarity values for English words are obtained from automatically derived ANEW+ polarity lexicon

Our Approach to Beliefs

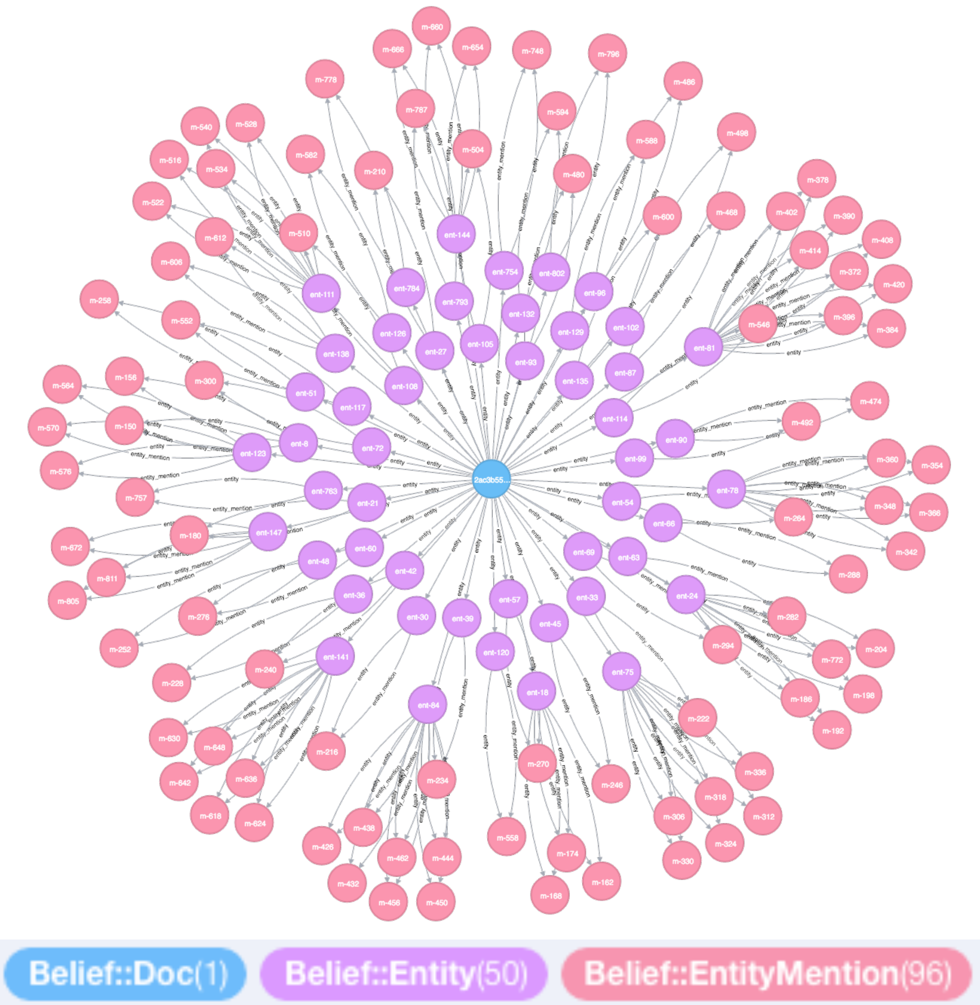
- Base
  - Construct graph from Rich ERE annotations
  - Augment graph with source information using parsing expression grammar
  - Nodes based on Rich ERE elements
    - Heterogeneous node and relation types
- Communities of Belief
  - Initialize all nodes with a unique label
  - Propagate label based on neighboring labels
  - No pre-defined objective function or prior information about communities
- Dialogue Acts
  - Predict discourse structure in the form of labeled dependency relationships between posts

Network Construction

- Start with a document

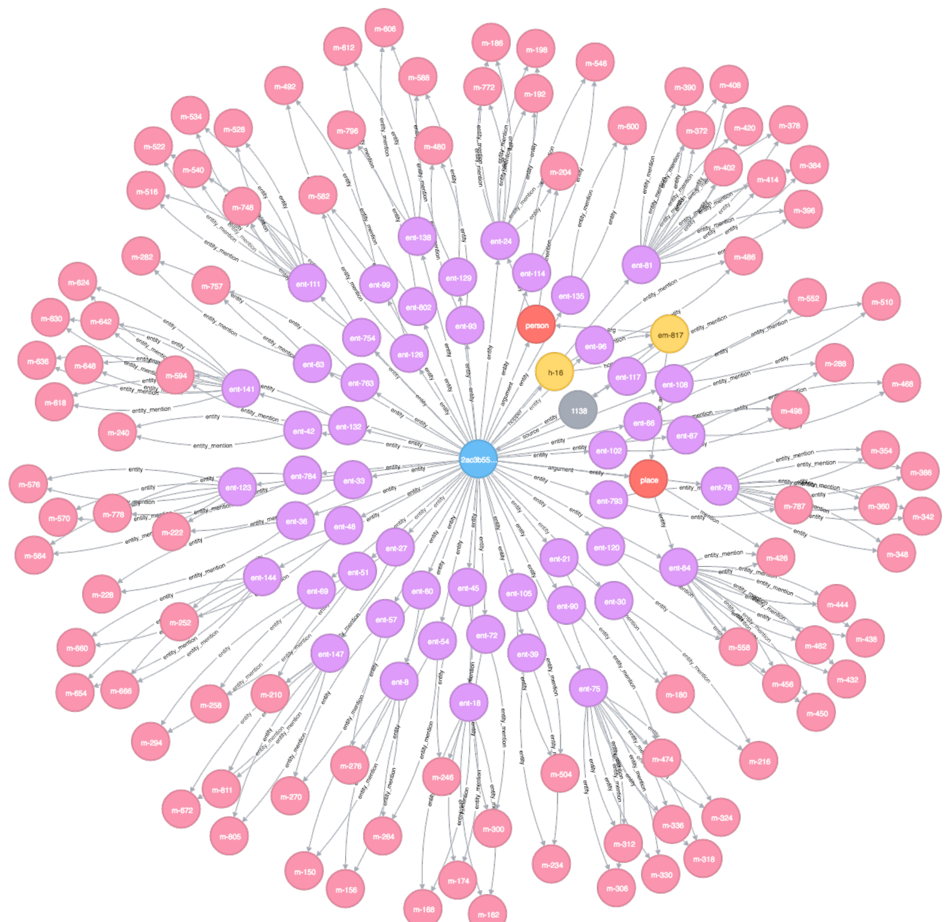
Network Construction

- Include Entities and Entity Mentions



Network Construction

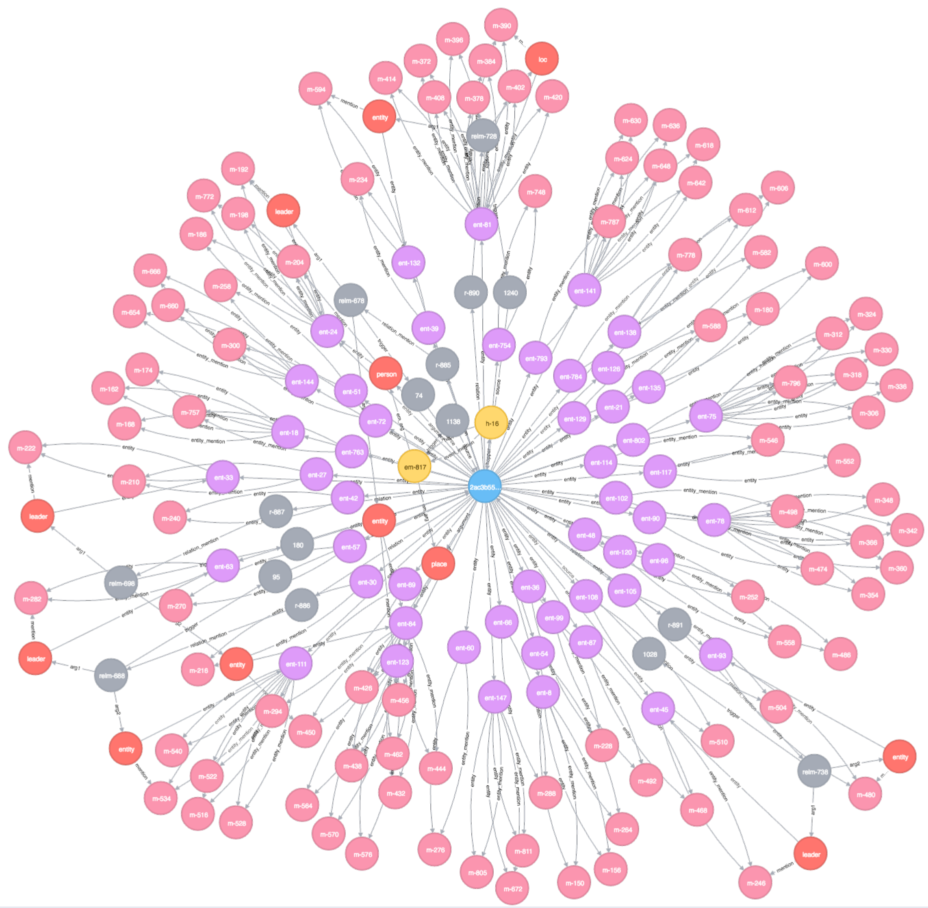
- Add event mentions, triggers, and arguments



Belief::Argument(2) | Belief::Doc(1) | Belief::Entity(50) | Belief::EntityMention(96) | Belief::EventMention(1) | Belief::Hopper(1) | Belief::Trigger(1)

Network Construction

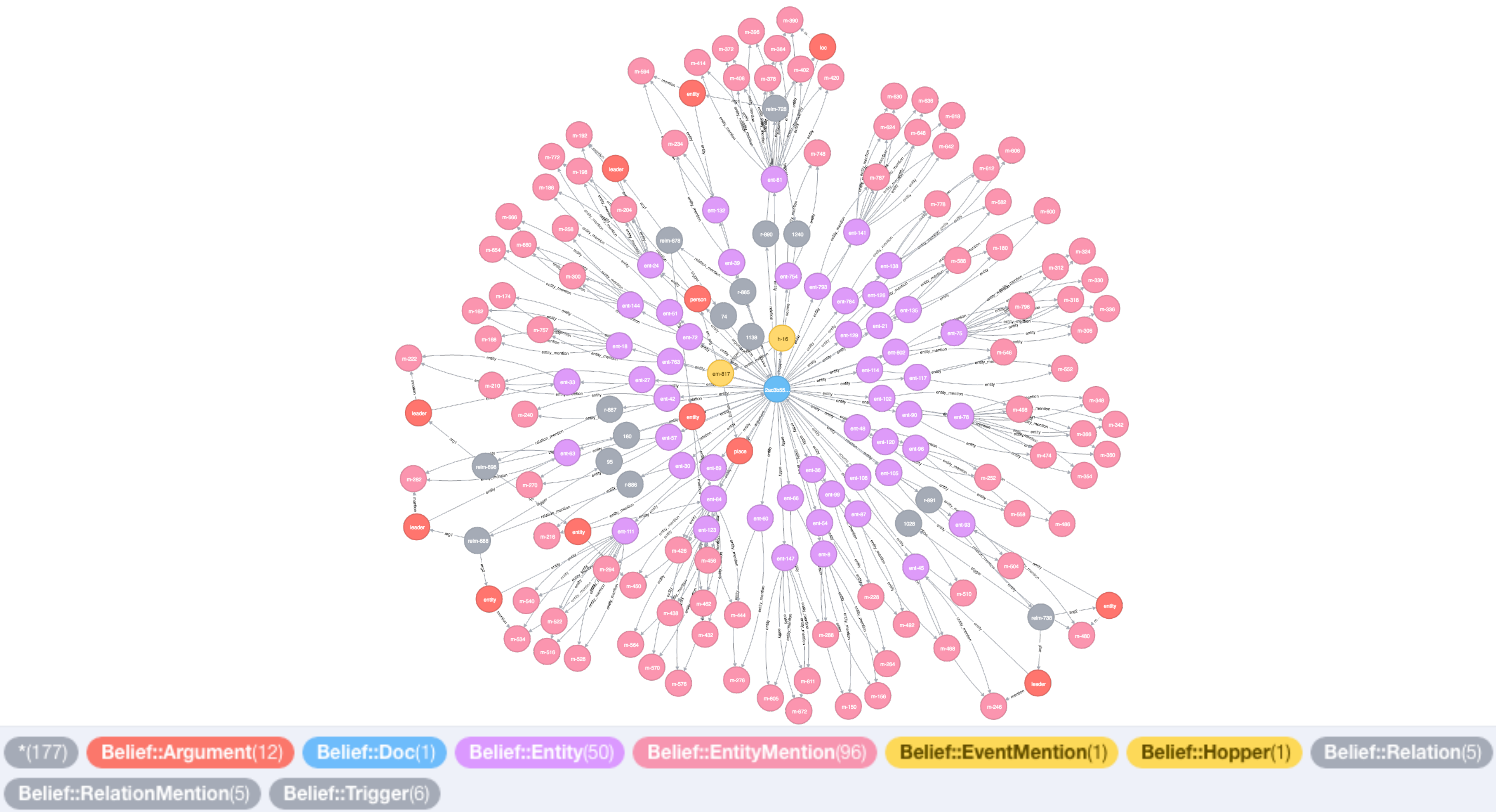
- Now relations, relation triggers, and relation arguments



*(177) Belief::Argument(12) Belief::Doc(1) Belief::Entity(50) Belief::EntityMention(96) Belief::EventMention(1) Belief::Hopper(1) Belief::Relation(5) Belief::RelationMention(5) Belief::Trigger(6)

Network Construction

- This is essentially a graph of possible targets



Parsing Expression Grammar for Source

```
<post author="randman" datetime="2011-12-04T23:21:00" id="p205">
<quote>
There are terrorist plots in the world, there just aren't terrorist plots like on "24."

</quote>
Interesting. 24 didn't involve the Illuminati or aliens so according to some here, no conspiracies.
</post>
<post author="randman" datetime="2011-12-04T23:26:00" id="p206">
<quote orig_author="Gazpacho">
The existence of the Trilateral Commission, and of its project to halt radical political movements around the world and restore a kind of liberal-authoritarian stability, are documented facts of history.
</quote>
Good point. How is it a conspiracy theory when the globalists openly call for world government.
</post>
```

Source document

Parsing Expression Grammar for Source

Best Annotation

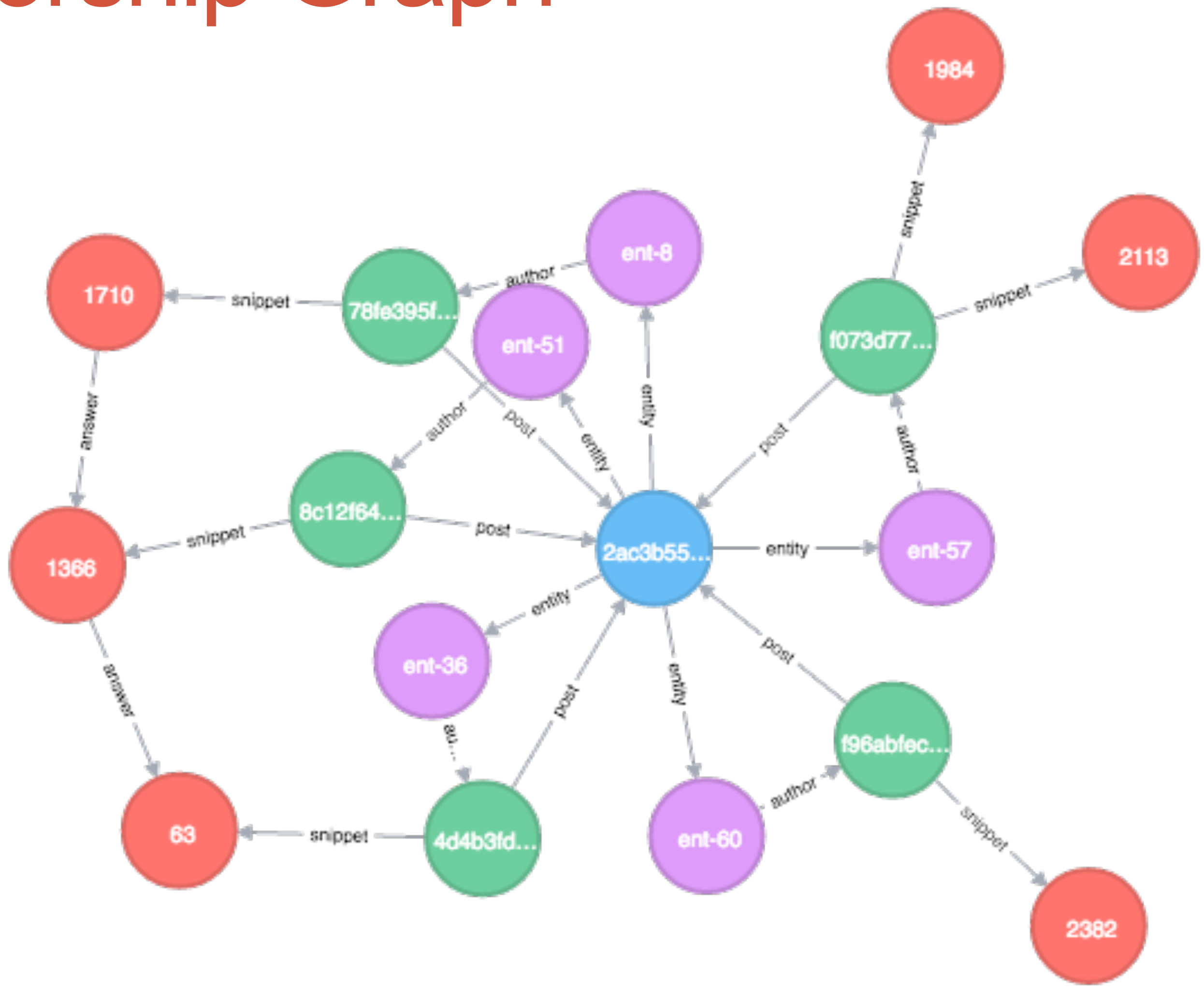
```
<event ere_id="em-976">
  <trigger offset="2113" length="7">killing</trigger>
  <sentiment polarity="neg" sarcasm="no">
    <source ere_id="m-126" offset="943" length="7">randman</source>
  </sentiment>
</event>
```

Linked at mention level

```
<entity_mention id="m-126" noun_type="NAM"
source="010aaf594ae6ef20eb28e3ee26038375" offset="943" length="7">
  <mention_text>randman</mention_text>
</entity_mention>
<entity_mention id="m-132" noun_type="NAM"
source="010aaf594ae6ef20eb28e3ee26038375" offset="5256" length="7">
  <mention_text>randman</mention_text>
</entity_mention>
```

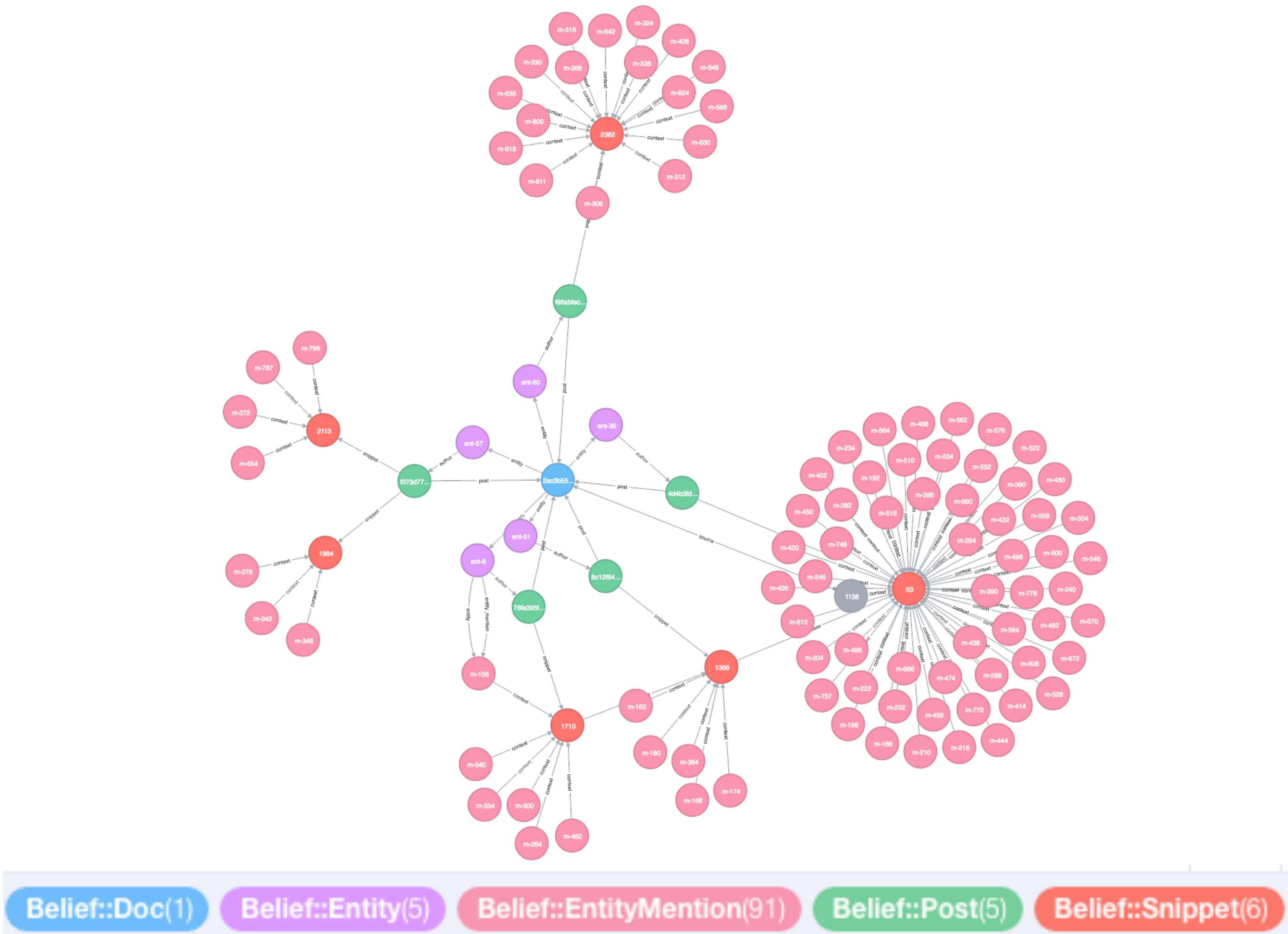
Rich ERE Annotation

Authorship Graph



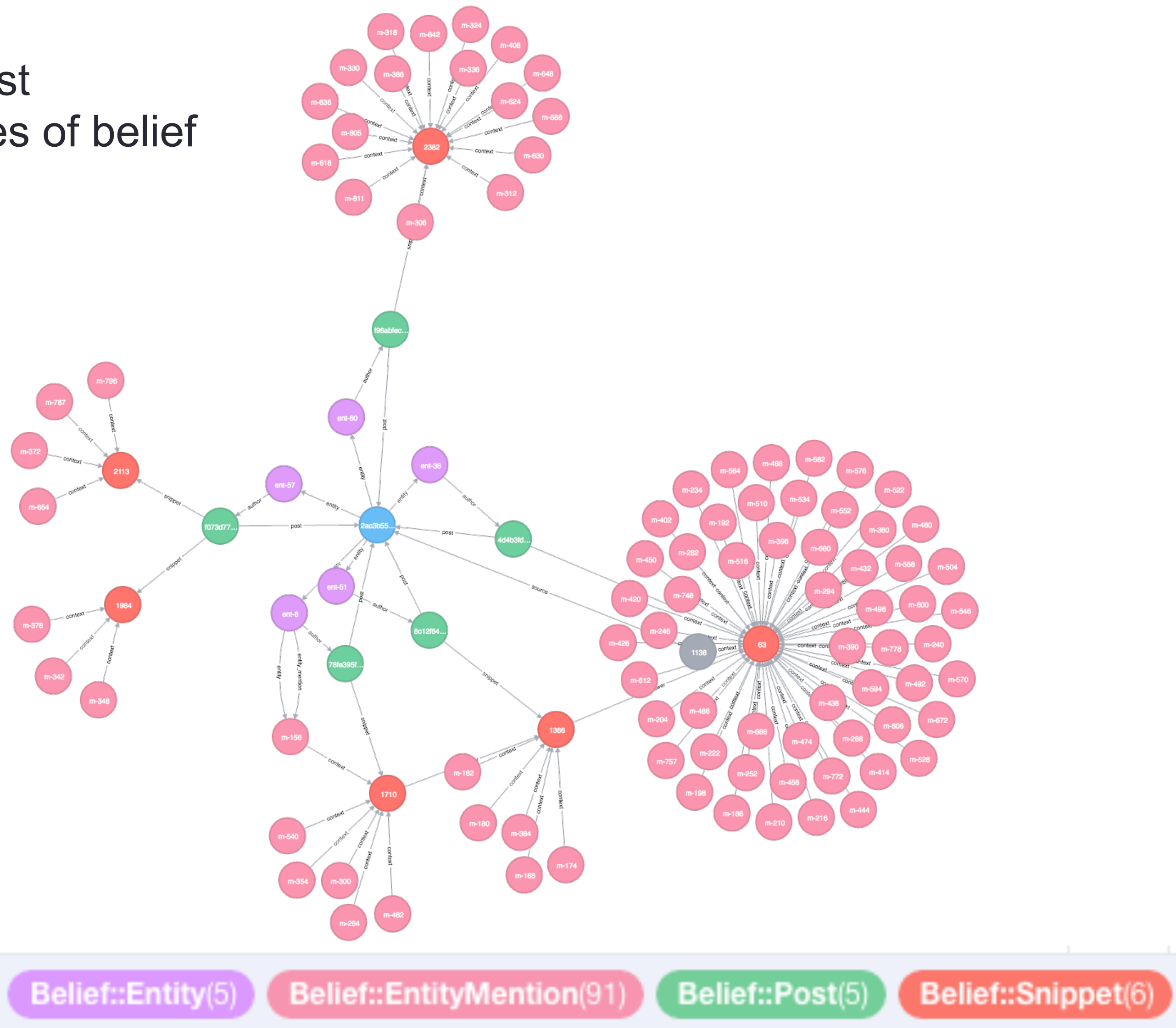
Belief::Doc(1) Belief::Entity(5) Belief::Post(5) Belief::Snippet(6)

Authorship Graph with ERE Data



Authorship Graph with ERE Data

Authors are most common sources of belief



Run 1: Naïve Bayes Labeling

- Process for Run 1 belief submissions
- Label belief nodes attached to event triggers, event arguments, and relation mentions with training data
- Features include
  - Nominals (event type, subtype, and realis; argument role and realis; relation type, subtype, and realis)
  - Strings (argument context; surrounding context)
- Graph structure not used

Results


English Belief
1
0.8
0.6
0.4
0.2
0
Baseline_sys1
Columbia_GW…
Columbia_GW…
Columbia_GW…
cornpittmich_sy…
cornpittmich_sy…
cornpittmich_sy…
CUBISM_sys1
Spanish Belief
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Submission:…
Submission:…
Submission:…
Submission:…
Submission:…
Submission:…
Submission:…
Submission:…
Chinese Belief
1.2
1
0.8
0.6
0.4
0.2
0
Baseline_s…
Columbia_…
Columbia_…
Columbia_…
cornpittmic…
cornpittmic…
cornpittmic…
CUBISM_…
DF Prec
DF Recall
DF F-Measure
NW Prec
NW Recall
NW F-Measure

Next Steps: Motivated by ViewGen

- Represents beliefs of agents as explicit, partitioned proposition-sets known as environments
- Includes the notion of “stereotypes”
  - Pre-existent models that fit stereotypical groups of people
  - By determining which stereotypes fit an individual we can ascribe the beliefs of those stereotypes to the agent
  - This might work for belief type as well
- Issues for the evaluation
  - No predefined models of particular groups of agents, so
  - Need unsupervised stereotype assignment

Graph Aware Mining

- Community Detection
  - Group nodes that are similar to each other and dissimilar from the rest of the network
  - Communities can provide insight into the beliefs of its members

- Relaxation Labeling (Future work)
  - Boost automated classification by considering neighbors
  - “Context-free” approaches don’t take advantage of networked information
    - Authors and genres
    - Football teams and conference opponents
    - Source, target, and type of belief?

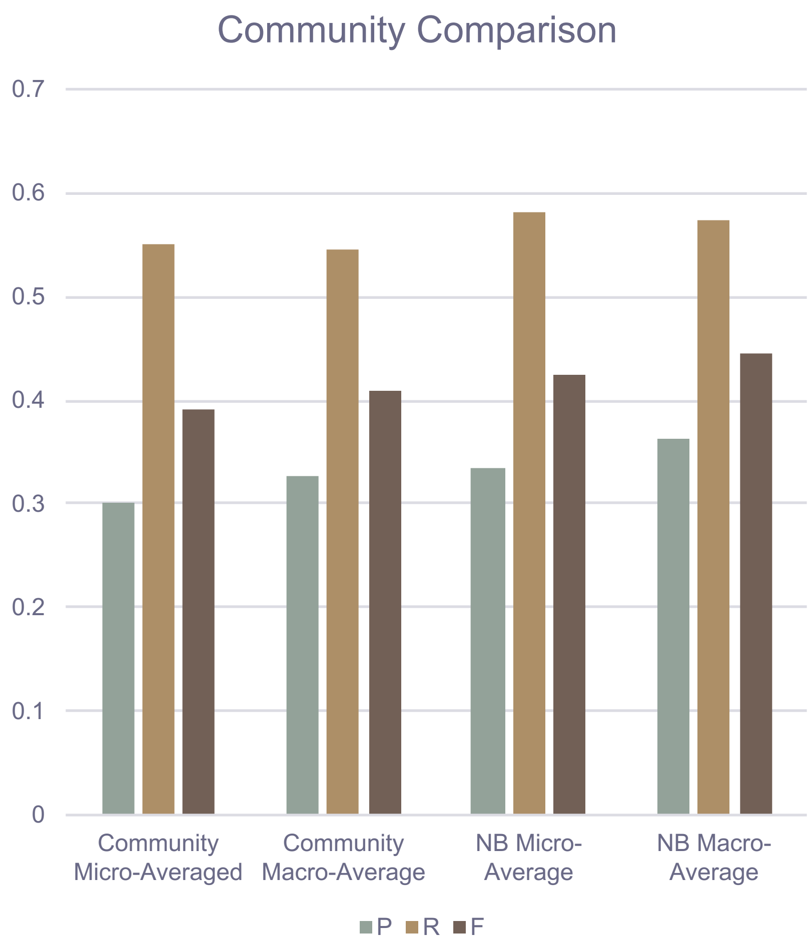
Community Detection Approach

- Unsupervised, near-linear time
- Number and size of communities are not predefined
- Label Propagation
- Has been effectively applied to detect communities in
  - Football conferences
  - Citation networks

1. Initially assign each node a unique label
2. Randomly order the nodes
3. For each node in that order, set the community label to the label that occurs most frequently in its neighbors
4. Stop when each node has a label that the maximum number of its neighbors have

Raghavan, Usha Nandini, Réka Albert, and Soundar Kumara. "Near linear time algorithm to detect community structures in large-scale networks."*Physical review E* 76.3 (2007): 036106.

Community Features



- Removed string features
- Added community profile features
  - Distribution in the community of each event/relation type-subtype combo

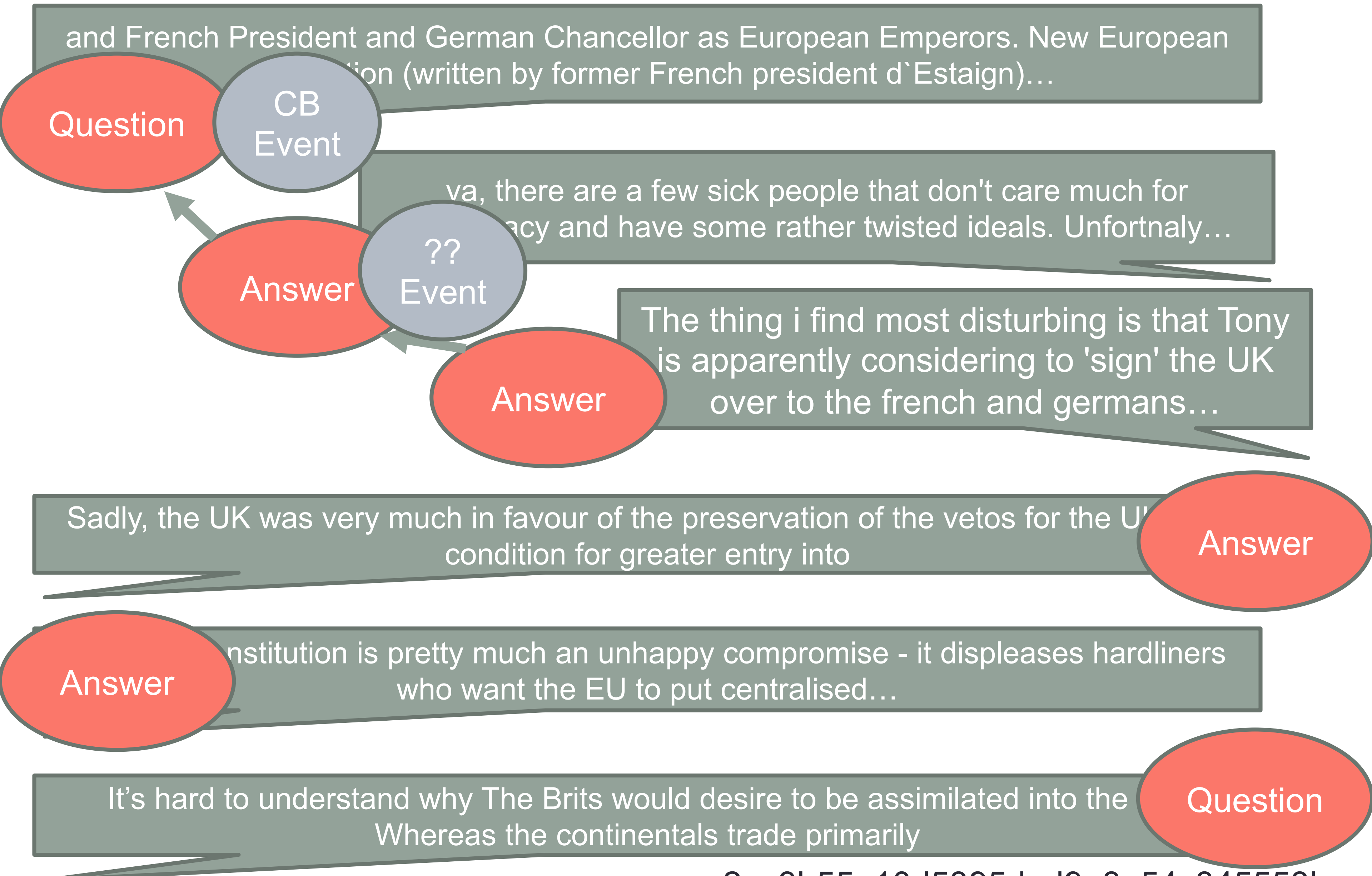
Issues with Graph-based classification

- Within document coref only, so most communities are dominated by source document
  - Link on event and relation subtypes
  - Simplistic cross-document coref
  - Still only 3 communities with > 1 document
- Wide range of document origins means authors don't repeat
  - Graph-based features might still aid classification, but misses thesis

Dialogue Acts

- Are beliefs classifications influenced by beliefs expressed in linked posts?
- Does the dialogue act of the post impact the belief class?
- Used MaltParser and the approach to predicting thread discourse structure described in (Wang, 2011)
  - One feature of MaltParser that makes it well suited to this task is it is possible to define feature models of arbitrary complexity for each token
- Used paragraphs as tokens rather than full posts as described in (Wang, 2011)
  - Attempt to scope tokens closer to the events and relations they contain

Wang, Li, et al. "Predicting thread discourse structure over technical web forums." *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011.



2ac3b55a10d5395ded9e8e54c345553b

Dialogue Act Features

- **Initiator –** is the paragraph in the initial post
- **Position –** position of paragraph in thread, between 0 and 1
- **Post Similarity –** distance from current paragraph to most similar other paragraph
- **Punctuation –** counts of ‘?’, ‘!’, and URLS
- **Author Profile –** percentage of paragraphs written by the author

- Previous work found that the author profile feature was the most useful when it was an author’s first post

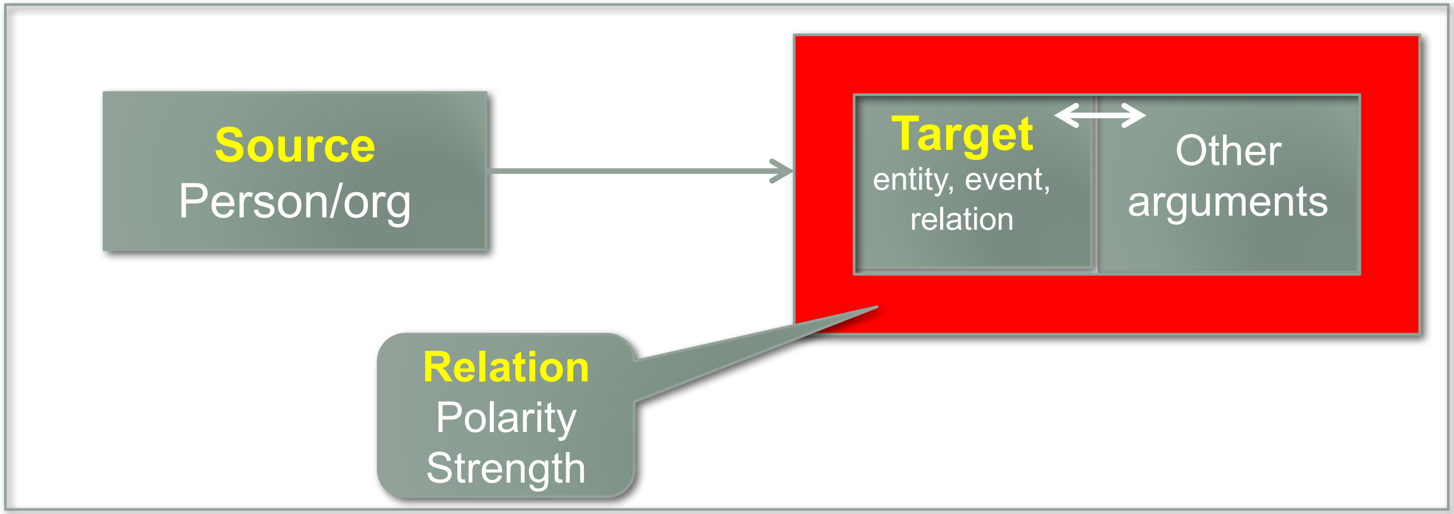
Next Steps for Dialogue Act

- Make use of the joint classification
  - We are currently using the dialogue act as a feature for a Naïve Bayes Classifier, but could apply it directly to joint classification of source, target, and type
- Domain specific training data
- Works in situ, so just getting the beginning of a thread should work fine
  - Starting in the middle needs to be evaluated

SENTIMENT

Target Focused Sentiment Extraction

- Detects **<Source, Relation, Target>** triplets
  - Source: a writer or an entity in agent role
  - Target An entity, relation and event within Source's scope
  - Relation: a verb or other item with Target as an argument.



- Sentiment computed using the Affect Calculus.
  - Baseline sentiment from an expanded ANEW+ affect lexicon

Types of sentiment-carrying relations

- ***GMOs*** *pollute the environment*.
  - Relation type: Agentive
  - Relation is highly negative (1.85)

- *Also, certain* ***GMO's*** *are nutrient enriched, so that's an advantage.*
  - Relation type: Propertive
  - Relation is highly positive (7.7)

- *It is easier for farmers to grow* ***GMOs*** *with less loss*.
  - Relation type: Patientive
  - Relation is slightly positive (5.6)

Base Polarity: Affect Lexicon

- Affective Norms of Words (ANEW)
  - c.f. Bradley and Lang, 1999
  - Also includes arousal and dominance values
- Scores words/phrases on a 9 point scale
  - Lower scores → negative valence ([1, 4) negative)
  - Higher scores → position valence ((5, 9] positive)
  - Neutral scores ([4, 5])
- Expanded ANEW+ lexicon using WordNet
  - Modeled after Liu et al. 2014
  - Original contains scores for ~2500 words
  - Expanded contains 22755 words

Affect Calculus: computing sentiment towards target

Relation type	**Type 1 (propertive)** *Rel(Target)*	**Type 2 (agentive)** *Rel (Target, X)*		**Type 3 (patientive)** *Rel(X, Target)*	
Relation/X		$X \geq$ *neutral*	$X <$ *neutral*	$X \geq$ *neutral*	$X <$ *neutral*
Positive	POSITIVE	POSITIVE	$\leq$ UNSYMP	POSITIVE	$\leq$ SYMPAT
Negative	NEGATIVE	$\leq$ UNSYMP	$\geq$ SYMPAT	$\leq$ SYMPAT	$\geq$ SYMPAT
Neutral	NEUTRAL	NEUTRAL	$\leq$ NEUTRAL	NEUTRAL	$\leq$ NEUTRAL

Expressed sentiment: negative

GMOs pollute the environment.

Target is agent
Relation is agentive

Relation is negative
(ANEW score 1.85)

Arg. X is neutral
(ANEW score 5.0)

Chinese and Spanish versions

- Chinese and Spanish systems are the same as English except:
  - ANEW+ is replaced by an equivalent Chinese and Spanish affective lexicons
    - Spanish is derived from English and human validated
    - Chinese is derived from Chinese sentiment vocabulary (VCA) and from English ANEW+ translation
  - Dependency Parsers:
    - Chinese – Stanford; Spanish – Freeling

Evaluation Results

System	Precision	Recall	F-measure
TAC SentEval 2014	30%	22%	26%
TAC BeST 2016 DF NW	 15% 5%	 15% 2%	 15% 3%

Discussion

- IHMC/Albany sentiment system has been designed for meaningful population of ADEPT KB
- Aim for high confidence extraction of sentiment triples geared towards high precision in subsequent aggregation
- This design was evident in 2014 KBPSent evaluation, where precision and high confidence were critical for producing best results.
- 2016 BeSt evaluation emphasizes recall and does not reward high confidence decisions
- Our system was designed to refrain from outputting low-confidence items (as those would simply add noise into the aggregate), hence low recall, even as our precision was relatively high compared to other systems.

Suggestions for Next Year

- Consider sources where authors appear in multiple documents
- Cross document event and entity resolution
- Extend evaluation to quotes and arguments
- Include polarity of belief
- Reward confidence scores