Erschienen in: EuroVis 2017 Eurographics / IEEE VGTC Conference on Visualization 2017 / Heer, Jeffrey et al. (Hrsg.). - (Computer Graphics Forum ; 36,3). - S. 213-225
https://dx.doi.org/10.1111/cgf.13181

NEREx: Named-Entity Relationship Exploration in Multi-Party Conversations

Mennatallah El-Assady[1,2], Rita Sevastjanova[1], Bela Gipp[1], Daniel Keim[1], and Christopher Collins[2]

[1]University of Konstanz, Germany
[2]University of Ontario Institute of Technology, Canada

Abstract

We present NEREx, an interactive visual analytics approach for the exploratory analysis of verbatim conversational transcripts. By revealing different perspectives on multi-party conversations, NEREx gives an entry point for the analysis through high-level overviews and provides mechanisms to form and verify hypotheses through linked detail-views. Using a tailored named-entity extraction, we abstract important entities into ten categories and extract their relations with a distance-restricted entity-relationship model. This model complies with the often ungrammatical structure of verbatim transcripts, relating two entities if they are present in the same sentence within a small distance window. Our tool enables the exploratory analysis of multi-party conversations using several linked views that reveal thematic and temporal structures in the text. In addition to distant-reading, we integrated close-reading views for a text-level investigation process. Beyond the exploratory and temporal analysis of conversations, NEREx helps users generate and validate hypotheses and perform comparative analyses of multiple conversations. We demonstrate the applicability of our approach on real-world data from the 2016 U.S. Presidential Debates through a qualitative study with three domain experts from political science.

1. Introduction

Multi-party conversations, such as political debates or oral court arguments, are characterized by a rapid exchange of opinions, arguments, and information, producing lengthy verbatim text transcripts rich with interruptions, disfluencies, repetitions, and other characteristics not often found in highly edited text. These events are often of long-lasting political, economic, and social importance, and the subject of much analysis. Aside from exploring the underlying social dynamics of conversations, investigation into the participation of speakers over the course of a discussion, the thematic evolution of a debate, and the different argumentation strategies are of interest in the social sciences (e.g., [Hab84, PEG13, NHBR13]).

Extracting structured information from this semi-structured data can be time-consuming, requiring close reading, annotating important passages and keywords, and structuring these meaningfully. A common approach for structuring this information is to create lists and mind-maps, using keywords extracted from the text [JFCS15]. Keywords are elements in a text that reflect its content and point to specific concepts, expressions, or abbreviations. In data-mining, the automatic classification of such elements in text corpora is known as named-entity extraction. Named-entities are typically grouped into categories, such as persons, locations, and organizations.

We developed **NEREx**, an interactive visual analytics framework for **N**amed-**E**ntity **R**elationship **Ex**ploration. Our approach was developed in an iterative design process with continuous refinement guided by periodic participation of linguists and political science scholars. Collaboratively, we identified six requirements for the effective support of the analysis process of our domain experts, these are: (1) getting an overview of important named-entities and their relations; (2) enabling close reading; (3) exploring the influence of different speakers; (4) supporting focused analysis of specific topics; (5) allowing for a temporal review of the complete conversation; (6) identifying the emotional context of entities and highlighting politeness. To arrive at the final design presented in this paper, we conducted three informal observational studies with a total of 12 participants to improve the usability and effectiveness of our approach, as well as, a qualitative pair-analytics study with three political scientists, discussed in Section 7.

Our approach was developed to support the exploration and analysis of multi-party conversations, in particular to provide an overview and entry point for unknown data. Addressing the described requirements, NEREx offers several linked perspectives on text data, as well as, powerful interaction capabilities. It uses a two-level abstraction of the text to construct high-level views of the semantic structure of relevant keywords and their relations. First, the automatic abstraction from the text-level, using named-entity extraction, grouping, and categorization. Second, an interactive aggregation of the extracted entities into concept clusters supports the specific analysis task of the user. Our framework is text-type

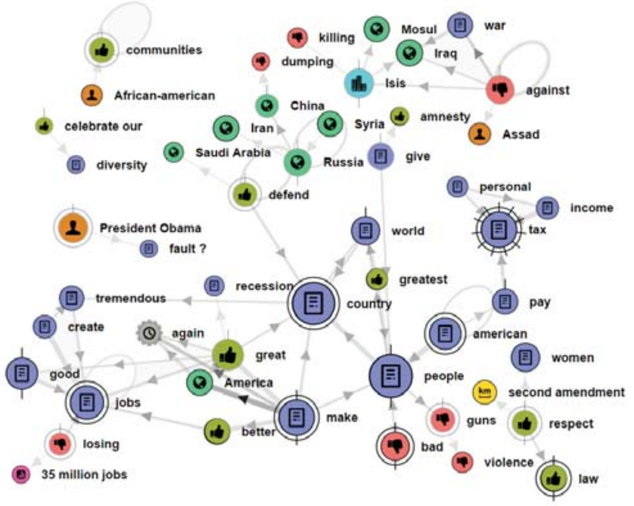


Figure 1: *Entity Graph of the combined transcripts of the three presidential debates between Trump and Clinton. The minimum entity-pair frequency for this graph is set to 3, resulting in a high-level overview of all important entity pairs and the influence of certain topics in the debate, such as taxes, jobs, and ISIS.*

and language independent, however, its design and the examples throughout this paper target text data with conversational character. We combine supervised and unsupervised learning methods to extract and categorize named-entities and other relevant keywords, such as dates, locations, or units of measurement. To analyze the semantic structure of the categorized entities, we apply a distance-restricted entity-relationship model to build pairs of named-entities. NEREx incorporates six linked views to support the following analysis tasks: Data Exploration, Hypothesis Generation, Temporal Analysis, Hypothesis Verification, and Comparative Analysis.

The ***Text-Level View*** **(TLV)** provides for reading of the text with entities in context, while the ***Entity-Level View*** **(ELV)** reveals entity sequences. We construct ***Entity Graphs*** **(EG)** (see Figure 1) by combining frequent entity pairs into an interactive graph structure and ***Speaker Graphs*** **(SG)** by connecting speakers who use common entity pairs. ***Concept Graphs*** **(CG)** are created from user-specified concepts, aggregating named-entities in a second abstraction level. To facilitate the task-driven analysis of the data, we designed a variety of interaction techniques, such as search and filter options for the data exploration and hypothesis generation, as well as visual querying for hypothesis verification. Using an animated reconstruction of concept graphs, ***Temporal Graphs*** **(TG)** allow the temporal analysis of the evolution of a conversation over time.

The main contribution of this paper is a framework for the exploratory analysis of multi-party conversations using six linked views to offer different perspectives on the data. We introduce a classification scheme for named-entities tailored to conversational text and a distance-restricted model to extract their relations. Moreover, we propose a graph clutter reduction technique through node groupings, to enhance the scalability of the overview. Lastly, we discuss our findings from a qualitative study with domain experts.

2. Related Work

Our framework for visual analysis of multi-party conversations is informed by related research from the fields of name-entity extraction and visual content analysis.

Named-Entity Extraction — Also known as Named-Entity Recognition (NER), named-entity extraction is a widely studied classification problem that refers to extracting elements in a text that belong to specified categories. Early works also describe this problem as a task of recognizing proper names [CS92]. Although according to the definition, the categories for the NER are predefined, there is a wide variety of categories that are considered in the literature. The most studied categories are names, times, and numbers. These types are commonly tagged as *enamex*, *timex*, and *numex*, respectively [NS07]. Common subcategories for names are persons, organizations, or locations. Expressions of time or date are examples of times. Numbers could be monetary values or percentages. The categorization of named-entities depends on the application scenario and on the underlying data. For example, for research papers in the natural sciences, a further differentiation of numbers is appropriate, into distance, speed, etc. If geo-location is of importance to the analysis, it can additionally be sub-categorize into cities, countries, and other landmarks. Sekine and Nobata [SN04] have defined a named-entity hierarchy that includes about 200 categories, covering frequent entities in news articles.

For extracting named-entities, techniques can be categorized into supervised, semi-supervised and unsupervised extraction. Supervised learning approaches are the established method for NER [NS07]. They rely on large annotated corpora and derive disambiguation rules using discriminative features of the entity classes for the extraction. Algorithms like Hidden Markov Models [BMSW97], Maximum Entropy Models [SGS98], and Support Vector Machines [AM03] have been successfully applied to NER. However, Conditional Random Fields [ML03] have proven to be the most reliable technique [NS07]. Semi-Supervised approaches use seed words to learn a categorization that is used for classifying unseen data, e.g., via bootstrapping [DA08]. Unsupervised techniques are based on lexical knowledge and statistical patterns in large unannotated corpora [TS12, RCE*11].

Examining the relations between named-entities has been widely studied [BB07], with two main types of relations in the literature: Relations based on syntactic structure and relations based on entity co-presence in documents. The first relation type is based on the assumption that deep linguistic knowledge is required for the comprehensive modeling of entity-relations (e.g., [ZAR03, Kam04, GT09, SGS11]). These models are mainly used for applications that require the modeling of linguistic complexity, such as question answering approaches. However, parsing-based techniques often fail on verbatim conversation transcripts due to the ungrammatical nature of the text. The second relation type is used mainly for text summarization purposes, since it is based on statistical correlations between entities. Tools such as Jigsaw [SGL08], Contexter [GM04], and others use such models to visualize thematic relations in text corpora. Our approach is a hybrid, using a restricted distance window to related entities in the text.

Visual Content Analysis — Jänicke et al. give an overview of the state-of-the-art visualization approaches in their recent survey [JFCS15]. They describe the tension between text-level close-reading and abstract distant-reading and conclude that both perspectives on the data are important for a holistic analysis. They classify close-reading techniques as augmenting the text using different colors, font sizes, glyphs, or connections. From the visual design of the close-reading view, there are several comparable approaches to NEREx [ARLC*13, AKV*14, GWFI14]. Distant-reading techniques abstract the text using different features and exhibit, therefore, a larger design space for the visualization. Depending on the features selected for the analysis tasks, distant-reading visualizations can be categorized into structure overviews, heat-maps, tag clouds, geospatial-maps, timelines, and graphs. NEREx consists of different views that contain visual elements from many of these categories. Yet, the most central element are the node-link diagrams used to visualize named-entity relations. Related approaches to creating text networks include Phrase Nets [VHWV09] and others [Cob05, AGL*07, VHWV09].

Visualization approaches based on conversational text data are also related to our work. Conversation Clusters [BK09], Chat Circles [DV02], and GroupMeter [LPH*09] group the content of conversations dynamically to show thematic structure. However, in contrast to NEREx, these approaches do not allow a deeper analysis of concept connections within a discussion. Trains of Thought [SGH12] goes a step further in connecting different themes together, however, it does not distinguish between different classes of concepts. ConVis [HC14] focuses on the analysis of opinions, ConToVi [EAGA*16] enables the exploration of speaker dynamics, uVSAT [KSBK*16] facilitates the analysis of stances in online social media, and Conceptual Recurrence Plots [ASW12] are used to provide insight on the coherence of a discourse. These approaches target specific analysis tasks or conversation types. In contrast, NEREx is designed as a general entry point for the analysis of different aspects of conversations.

3. Named-Entity Abstraction Model

After standard data cleaning, lemmatization, and n-gram extraction, we use a combination of supervised and unsupervised learning techniques and heuristic approaches to extract relevant elements from the text. Since these techniques are language-dependent, we designed NEREx to work on both English and German data, with the option of extending the supported languages in the future. In this paper, we focus on the pipeline for the English language.

To extract the first basic entity categories, we use the Stanford Named Entity Recognition system [FGM05]. This stable approach uses supervised learning through a linear-chain Conditional Random Field model [FM09] to predict the most likely sequence of named-entity labels in a corpus. The system uses multiple features (lemmas, POS tags, capitalization, etc.). This established recognizer has stable performance with high accuracy [NS07]. We apply a 7-class model on our data to extract: Location, Person, Organization, Money, Percent, Date, and Time.

Additionally, we rely on unsupervised learning through topic modeling [EA15] and lexical-chaining algorithms [GREA15] to extract content-related keywords for the particular text of the analysis. These keywords are based on statistical correlations and give an insight into the thematic composition of a text corpus. Both unsupervised methods do not require prior knowledge about the content of the conversation, in particular, no compulsory parameter for the number of topics. However, depending on the focus of the analysis, the user can optionally specify the desired number of topics.

Finally, to extract other relevant elements from the text, we use a set of heuristic approaches. These include rule-based classifications using word-lists, lemmas, POS-tags, and regular expressions. These rules are created manually to tag the text using lists of tokens or phrases that can be updated interactively. The lists include units of measurement, date and time keywords, politeness indicators, and positive or negative emotion indicators. One way in which these rule-based classifications complement the supervised learning techniques is by improving the recognition of titles or honorifics by using word lists that include, e.g., *Dr.*, *Prof.*, *Sen.*, *Judge*, etc.

3.1. Entity Classification

Through extracting the elementary entities in the first step, we lay the foundation for the classification of these entities into categories relevant for the analysis. These categories are text-type dependent. For conversational text data, we derived 10 general categories. These are generated through a rule-based combination of their relevant elementary entities.

One of the most important categories for multi-party conversations are persons, as they allow the tracing of the mentions of different people in the conversation and indicate the active roles in a debate. **Person** names are identified by the NER and additional titles and abbreviations are adapted using the rule-based classification scheme. In most conversations we analyzed, speakers refer to **Geo-Locations** in the context of their utterances. This category allows, therefore, the exploration of geo-special information in the text. Another important category is **Organization**, which classifies names of organizations mentioned in the text. For creating this category, we rely on the NER in addition to given user-generated word-lists for specific conversations. The category **Date-Time** indicates not only the presence of time- or date-identifiers in the text, but also word phrases that point to a time span or have a reference to a specific year. **Measuring Units**, such as *kilometer*, *mile*, or *gram* are extracted using word-lists. We define measuring units as a separate category, since they indicate the orders of magnitude that speakers mention in their utterances. If a measuring unit is identified in close proximity to a number or numerical-expression, we use regular expressions to match these into measures. Hence, the category **Measure** does not contain single tokens, but combined ones, such as *400 km*. In addition to measures that contain a measuring unit, we include statistical data into this category, identified by a number or percentage followed by an optional preposition and a Context-Keyword, e.g., *thousands of people*. Measures are important cues in the conversation, since they indicate facts that speakers bring into the discussion. Given that measures are often disputed, showing measures in their semantic context gives insight into the controversies of a debate. Examples of controversial discussions include factual disputes over the cost of a particular project or the number of people involved.

To place the extracted entities in context, we rely on statistical algorithms for extracting **Context-Keywords**. These are not named-entities, but are nevertheless important for understanding thematic relations. Moreover, to reveal the attitude of speakers towards certain concepts and the other participants in the debate, we use sentiment analysis algorithms to extract **Positive-** and **Negative-Emotion Indicators**, as well as word-lists to tag **Politeness** keywords, such as *thank you* and *please*. These three categories were defined based on a request from our domain experts. In the remainder of the paper, we refer to all categories of extracted terms collectively as *entities*.

To remove duplicates of the same concept, we group entities that are based on the same token but identified in different categories. If instances of an entity are classified in multiple categories, the category used in graph visualizations is the most frequent one. Furthermore, if two entities form a stable collocate over the whole corpus, they are automatically grouped together as a single entity.

3.2. Distance-Restricted Entity-Relationship Model

In order to explore the relations between entities, we use a distance-restricted model for creating entity-pairs. As described in Section 2, prior work considers relations between entities either as semantic relations based on linguistic knowledge or regards two entities as related due to their presence in the same document. The often ungrammatical structure of utterances in verbatim text transcripts (including non-standard lexical items, syntactic patterns, interruptions, repetitions, and crosstalk) requires a simple extralinguistic model. To overcome these limitations, we introduce a model that bounds the scope in which we consider two entities to be related, using a distance threshold (*maxDist*). Our method creates a pair of entities if the entities appear in the same sentences within *maxDist* words of one another.

Using this definition, we can ensure an efficient computation of all entity pairs. As depicted in Figure 2, to compute the entity pairs a sliding window, *maxDist* words wide, moves along each sentence and finds for each named-entity all following entities falling within the window and sentence boundaries. Since we are interested in the direction of the entity-pair-relations, we only look for all following entities within the window for each entity. Exploring the entity-pairs extracted by our model gives insight into the direction of the relationship between entities and allows a better understanding of the semantic structure of a text corpus. The most frequent entity-pairs may represent the topic of a discussion and frequent pairs of a speaker may give a hint to their stance.

However, not only are frequent entity-pairs important, but also pairs that are semantically similar (e.g., *cut \$5 trillion*, *cut \$4 trillion*). These often present opposing speaker opinions about a given topic. To maximize the chance that entity-pairs represent true semantic relations, we set the default value of *maxDist* to a low value (5 words). To vary the granularity of the analysis, the parameter can be changed interactively depending on the analysis task and data. Nevertheless, some infrequent longer-distance entity-pairs are discovered which do not represent salient relations. To reduce their impact, we calculate the average observed distance (in words) for each entity pair. We use both frequency and average distance in the visualization to reflect the strength of the entity-pair.

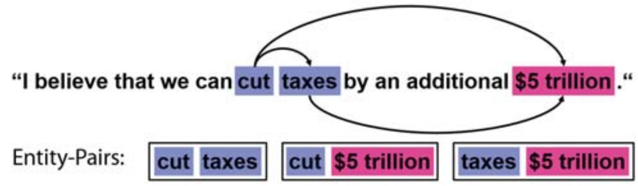


Figure 2: *Example of generating entity-pairs with the distance-restricted entity-relationship model.*

4. Visualization Components

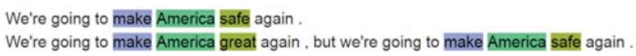
After processing the data using our model, the six linked full-screen views of NEREx, arranged in a tabular interface, are populated with word, entity, and entity-pair statistics and relations, speakers and their metadata, and the complete utterances of the conversation. All views of NEREx are connected through brushing and linking, keeping selections and filters consistent across the views.

The extracted name entity categories are the most important elements for all the views. To make them pre-attentively recognizable, we choose to map them to a discriminative visual variable for nominal data, such as color or shape. Since we additionally encode the frequency of each entity using size, and comparing the sizes of different shapes yields less accurate results than comparing objects of the same shape, we chose colored circles to represent the different entity categories. We chose the specific hues to be easily distinguished and mnemonic where possible (e.g., red as a Negative-Emotion-Indicator and green as a Positive-Emotion-Indicator).

The following sections describe the six interactive, linked views. All views, except for the Speaker Graph (SG) could be used to analyze non-conversational text in addition to conversations.

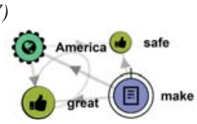
4.1. Text-Level View (TLV)

The first view represents the complete text of the corpus with all entities highlighted in their respective colors. This visualization is a very important reference for the analysis because it allows the users to go back to the original text for a close-reading. This view supports search and filter interactions, as well as selection and highlighting. Through brushing and linking, users can make selections in other views, then return to TLV to inspect a related segment of the text in detail. Figure 3a shows a snippet of the TLV.



(a) *Text-Level View (TLV)*

(b) *Entity-Level View (ELV)*



(c) *Entity Graph (EG)*

Figure 3: *Abstraction of named-entities from the text-level to the abstracted distant-reading views.*

4.2. Entity-Level View (ELV)

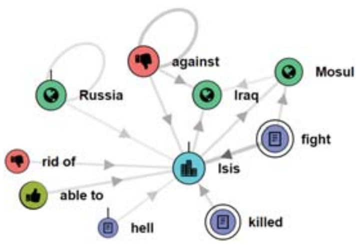
The second view is an abstraction of the single entities in the text. This visualization abstracts the sentences as lines and shows the abstract entity-circles in their respective position along these lines. An example of an utterance from this view is seen in Figure 3b. All sentences that make up one speaker-turn are enclosed by a bounding box (not shown). An adjacent detail panel shows the full-text sentences upon hovering on an entity in the ELV. A navigation pane showing a compressed representation of the complete corpus, with utterances containing selected entities highlighted. This allows users to jump directly to any point in the conversation.

The main purpose of the ELV is to support detection of patterns and anomalies in the occurrences of entities over the course of a discussion. To support this interactive analysis, we implemented a number of methods that go beyond simple search, filter, and highlighting operations. These are explained in more detail in Section 5.

4.3. Entity Graph (EG)

The EG is a directed node-link graph of the extracted entities (nodes) and their entity-pair relations (edges). The position of each node in the graph is determined using a force-directed layout. Edge lengths are proportional to the average entity-pair distance, so entities closer in the text are closer in the graph. Edge thickness and brightness are related to the frequency, so frequent pairs are more visible. Node size is proportional to the frequency of the corresponding entity in the text. Figure 3c shows the representation of the slogan “Make America Great [Safe] Again!” in the EG.

Figure 4 shows the entity graph of the first 2016 presidential debate. This graph gives an overview of the discussed topics in the debate, e.g., *taxes*, *jobs*, *gun law*, *the war on terror*, *cyber-warfare*, etc. In addition to the debate content, a moderation topic cluster can be found on the top left corner of the figure. To adjust the view’s level of detail, the user can interactively adjust the minimum occurrence frequency of an entity-pair to be included in the EG. By lowering this parameter, the resulting graph becomes more dense and connected. By gradually increasing the minimum frequency, the graph divides into several components representing different subtopics. The sub-graph on the side shows one such connected component, related to the entity *ISIS*. Some of the keywords in this sub-graph can also be found in the complete graph in Figure 4.

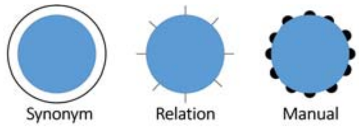


In addition to the interactions supported by the whole framework, the EG incorporates interactions that support navigation and readability of the graph. In addition to panning and zooming, the user can adjust the spacing of the nodes using a slider that varies the global repulsive force of the layout. Details about nodes and edges are provided on demand with tooltips, and hovering over an element highlights its direct neighbors.

To focus on a single entity and its relations throughout the conversation, the user can select a node in the EG. Related nodes and edges are highlighted, while other elements are de-emphasized. The selection is propagated to the other views. To explore the relations between nodes, the user can enable node-anchoring and fix the position of nodes of interest. The layout of related nodes will update to the new anchor positions.

Graph Clutter Reduction

To reduce visual clutter in the EG and to improve the scalability of the view, we introduce three methods for grouping nodes, namely *Synonym*, *Relation*, and *Manual* groupings. As shown on the side-figure, we use different node contours to indicate the type of node grouping. The contours are designed so that they can be overlaid to indicate a combined grouping in a node. For example, the node *Mr. Trump* in Figure 4 contains all three types of groups. In the following, we will explain the different grouping types in more detail.



Synonym Group — This type of grouping clusters together nodes that have a high similarity (using a user-defined parameter for the minimum similarity threshold). These similarities are based on the Levenshtein [Lev66] edit distances on the word and n-gram levels. Entities in this group are sorted according to their frequency in the text and the most frequent entity defines the group node and category. To avoid duplications in the graph, the automatic synonym grouping sorts ambiguous entities to the group they are most similar to. To match entities beyond simple token similarities, we use a heuristic for the different categories. For example, for person entities, the first and last names are compared. To avoid matching two persons who share either their first or last name, we only apply this grouping if it does not introduce ambiguities into the data.

Relation Group — Some entities are only connected to one other entity (i.e., leaf nodes). To reduce clutter, these nodes are grouped into their connected neighbor (i.e., parent node), which is subsequently labeled as a relationship group. If a connected component in the graph only contains two nodes, the first entity of the pair is the parent. Visually, relation nodes are distinguished using edge-stubs that indicate the number of relations they hide. Relationship group nodes can be toggled open to show all contained relations.

Manual Group — In addition to automatic grouping, entities can be modified manually. The manual grouping and modification is marked with a flower-pattern-contour. This editing consists of grouping, merging and splitting nodes and groups, renaming entity-groups, changing a group or node category, and deleting nodes.

4.4. Speaker Graph (SG)

The purpose of the fourth visual component is highlighting the concurrence of named-entity pairs between the utterances of different speakers. We designed the speaker graph using common entity-pairs between speakers as a measure of their proximity.

In this graph, speakers are depicted as nodes and are connected by an edge, if they have at least one entity-pair in common. The size of each speaker node represents the number of their

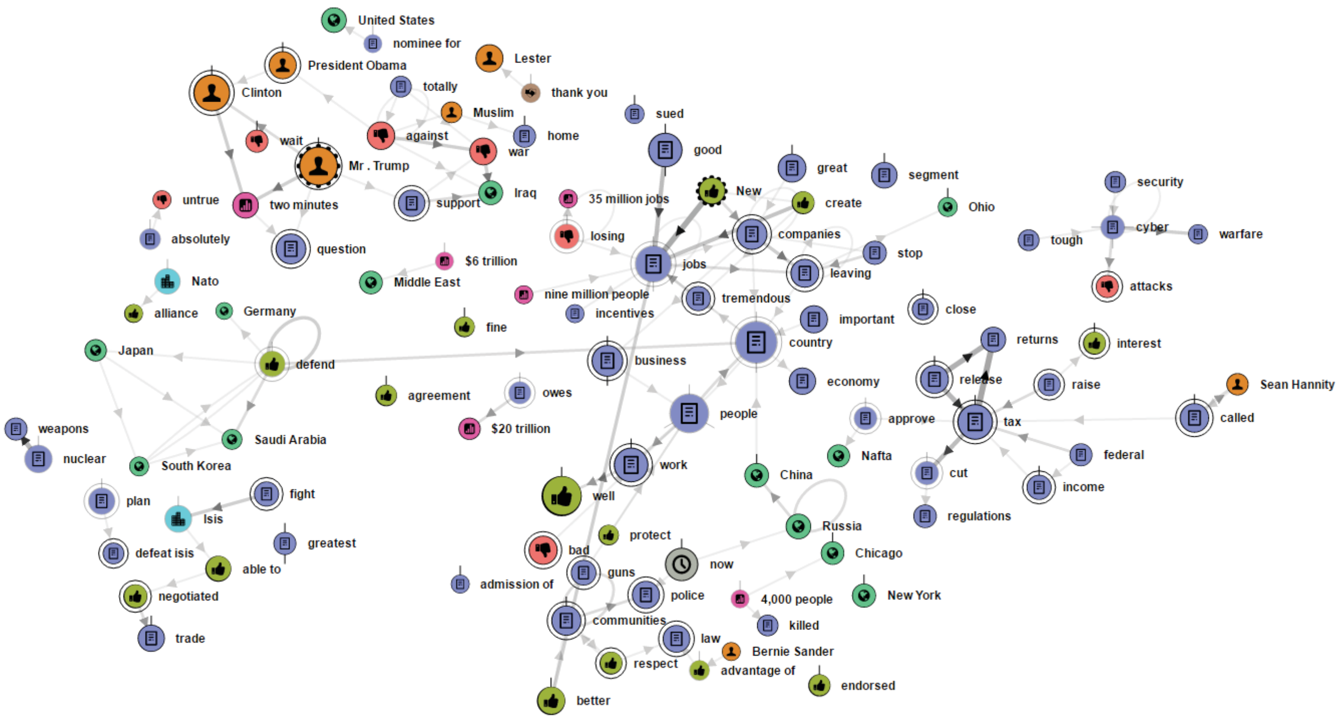
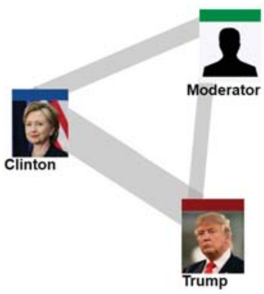


Figure 4: *Entity graph of the first presidential debate between Trump and Clinton, with a minimum entity-pair frequency of 2.*

utterances while the thickness of an edge connecting two speakers is proportional to the number of their common entity-pairs. The average frequency of these common pairs is used as a weight for the edges. The side figure shows the speaker graph of all three presidential debates between Trump and Clinton. This visualization is more insightful for multi-party conversations with a larger number of participants. Figures 9a and 9b show the speaker graphs of the republican and democratic candidacy debates for the US presidential elections of 2016.



This view integrates a detail panel. By hovering over the nodes of the graph, the panel shows an ordered list of the most frequent entity-pairs used by a single speaker. Hovering over edges shows entity-pairs used in common by two speakers. The user can select an entity-pair to explore which speakers mention it. Selecting a speaker reveals all their connections in the graph. This might indicate the activity of a speaker in the debate or the centrality of the speaker's utterances to the overall discussion. For additional details about the speakers, users can explore their profiles and statistics on their participation in a sidebar.

4.5. Concept Graph (CG)

Concept graphs are designed for a focused analysis of user-defined concepts. These visualizations create a second level of abstraction on top of the named-entity abstraction. This second abstraction level is achieved through a manual aggregation of relevant named-entities into concept containers. This aggregation is particularly useful for the analysis of relations between different topics across a discussion. Before selecting the concepts to visualize, the user creates concept containers. This is done using a responsive interface that allows the searching and filtering of entities, as well as their reordering according to different features. To create a container, the user just has to drag-and-drop the selected entities from the general list into the specific container panel.

To enhance the selection of relevant entities for the concept containers, we implemented a recommendation system that relies on the Levenshtein [Lev66] edit distance between entities, as well as the co-occurrence of entities in the corpus. After selecting an entity, the system automatically suggests the similar entities to add to the newly created container. In addition, the topic of each entity is used to recommend possible candidates who could be semantically related to the already selected entities.

After creating a set of concept containers (at least two), the user can select two or more containers to create a concept graph. This graph structure is based on the entity relationships across containers and ignores entity-pairs within the same container. This allows the analysis of relations between different topics or concepts, while reducing clutter.

The layout and interaction of the node-link diagram of a CG is the same as an EG to facilitate the usability of the tool. In addition to the entity nodes, speaker-nodes can also be included in this graph, to connect each speaker to all the entity-pairs they mentioned. Since CGs are more focused graphs that are intended to be used for a detailed analysis, in addition to the force-directed layout of the graph, three anchored layouts are supported. These are based on the analysis task and can be combined and adapted by the user.

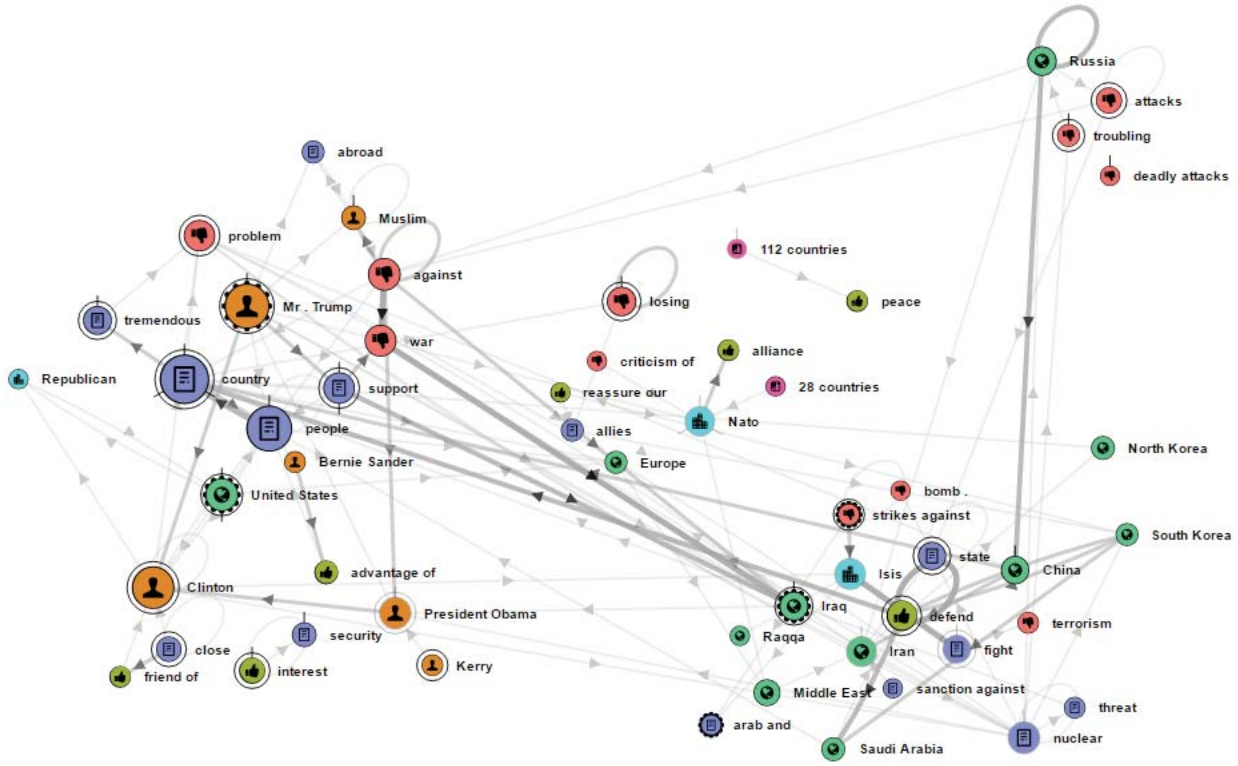


Figure 5: *A location-anchored concept graph depicting entities related to the topic "war on terror" in the first 2016 presidential debate.*

Speaker-anchored CGs enable the positioning of speaker nodes as fixed anchors in the force-directed layout to explore the tension-field created between them. When the speaker-position is fixed, all other nodes are positioned automatically, getting pulled by each speaker with a specific force (corresponding to the frequency with which a speaker mentions that entity). This layout is particularly useful for analyzing the contributions of the different speakers to the debate and for finding speakers that share similar views.

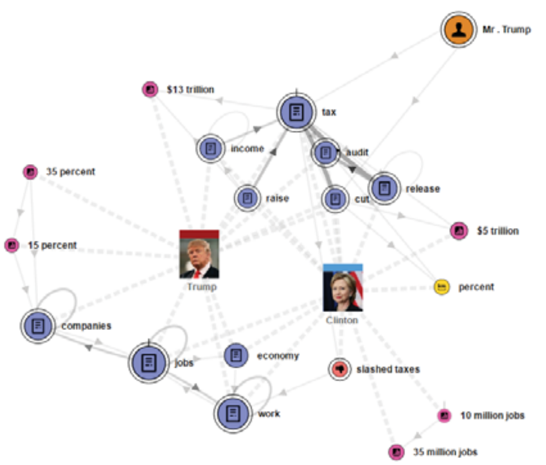


Figure 6: *Speaker-anchored concept graph from the first presidential debate, focusing on the concept taxes. The graph is anchored by the two candidate speaker-nodes.*

While creating a speaker-anchored CG, the speaker nodes tend to be fixed according to their similarity, which reduces the forces pulling the nodes in opposite directions. Figure 6 shows a speaker-anchored CG of the first 2016 presidential debate with the focus on the concept *taxes*. When using this visualization, all nodes in the graph can be selected interactively, showing only their related nodes and edges to reduce clutter.

Location-anchored CGs use the approximate coordinates of geo-location entities to anchor them on the canvas. This is especially useful for use-cases that analyze the relations between different geo-locations. Figure 5 shows a location-anchored concept map created using the containers for *geo-location* and *war-on-terror* in the first presidential debate of 2016. The location entities in this figure are fixed in their approximate position on a world map while all other entities are place around them by the force-directed layout.

Entity-anchored CGs are designed to reveal the relation between selected entities. As in anchored EGs, by fixing the position of selected entities, the graph is reorganized to show the most related nodes to each entity and bridge-nodes that connect entities.

4.6. Temporal Graph (TG)

Temporal graphs allow the exploration and analysis of the evolution of a conversation over time. This capability is relevant for the analysis of speaker participation in a conversation and for showing the interactions among speakers in a debate. TGs are based on CGs with additional timeline and animation controls. For the time-varying exploration of the graphs, we use animation in addition to

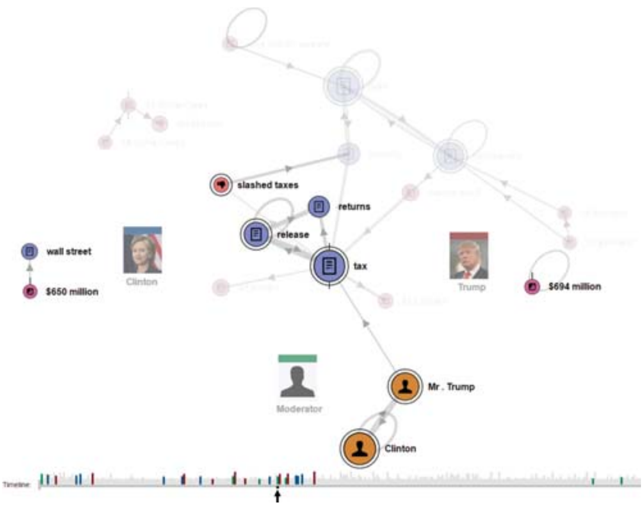


Figure 7: *A temporal graph from all presidential debates concatenated, showing different perspectives on the concept taxes and that the concept occurs almost entirely during the first debate.*

highlighting to overcome change-blindness. The timeline, as shown in Figure 7, consists of vertical bars that refer to the utterances of the discussion, with their height indicating the utterances' length. All utterances that contain entity-pairs that appear in the graph are highlighted using the color of their respective speaker.

The animation starts with an empty canvas and with each passing utterance, entity-pairs are added to the graph. To reduce clutter and focus the viewer on the current topics of conversation during the animation, we gradually fade out inactive nodes as the animation progresses. If entity pairs are re-used, they are highlighted to draw attention. To allow the conversation to be followed from the perspectives of different speakers, we blend-in the node of the currently active speaker in each frame of the animation. A frame of the presidential debate animation for the TG on taxes is shown in Figure 7. The speed of the animation can be changed interactively.

5. Interactive Analysis and Exploration

To facilitate the analysis and exploration using NEREx, we implemented a wide range of interactions. In addition to the specific interactions discussed for each view (e.g., parameter adjustments, linking and brushing, etc.), we provide further interactions, which are usable across all components of the framework. In this section, we describe the most important of these techniques in more detail.

Adaptive Entity Extraction — To improve the accuracy of our entity extraction and classification, we implemented an interactive learning system that adjusts the rule-based classification and word-lists. Entity classification corrections provided by users are retained and incorporated into future sessions, converging over time to a more accurate classification of entities.

Search and Filter — Using a rich search and filter interface, users can select any element in the visualizations for a detailed inspection. We apply search using auto-complete and matching on sub-strings. Lists of entities can be sorted alphabetically, according to the entity category, or by frequency. Users can search and filter single named-entities, entity-pairs (ordered or unordered), or speakers. By selecting any of these elements, each visualization is updated, revealing a different view into the data. Figure 8 depicts the search and filter interface of the NEREx framework.

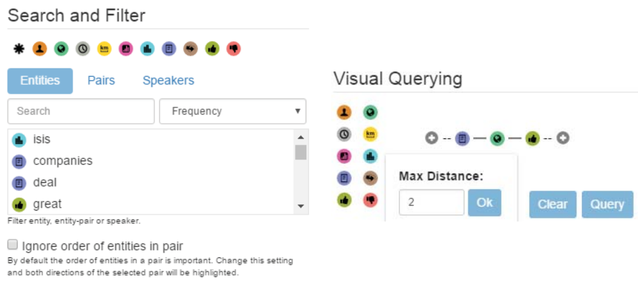


Figure 8: *Search, filter, and visual query interfaces.*

Visual Querying — Visual queries consist of a chain of entities and the maximum distance between each entity. An entity can be defined as a complete category or as a set of single entities. The visual querying interface is mainly used for hypothesis verification. By dragging-and-dropping the entity-icons on the empty query placeholders, the query chain expands in both directions, creating new placeholders. Individual entities can also be selected and included in or excluded from a category in the query. The maximum distance between two entities in a query chain can be adjusted by interacting with the connection between them. Figure 8 shows an example of a query that looks for a location, followed by an organization and then a person.

6. Implementation and Scalability

NEREx is implemented as a client-server application. The back-end implements a set of text processing algorithms as described in Section 3, while the front-end web-application is built on an AngularJS framework with visualizations in D3 [BOH11].

Entity extraction and classification is completed once in a pre-processing step on the server, and thus we do not encounter challenges scaling pre-processing to very large datasets. However, since several of our views are based on graphs, we are limited by the common challenges of scalablity of force-directed graph layouts. While other approaches, such as a matrix diagram, may scale to a higher number of entity pairs, we chose to work with graphs due to their intuitive readability. In addition to interactive features, such as hover and selection to focus on a local neighborhood of the graph, we have included several ways for users to limit graph growth.

The number of unique entities and entity pairs extracted from the text will affect the level of clutter in each of the views of NEREx. The growth of entities with text size depends on settings of the extraction algorithm, as well as the particular content of the text. Views based on texts which are repetitive and focused on a few topics will not grow cluttered as quickly as views based on widely varying text with many relations.

To reduce the number of visible nodes in our views, we implemented the three aforementioned types of node groupings. We also allow the user to adjust the entity extraction process by modifying the *maxDist* parameter. The minimum number of appearances of entity pairs required for inclusion in the views can similarly be adjusted. To facilitate selection, defaults which have yielded good results across several corpora are provided. A typical use case would be to limit the EG size by removing distant and infrequent pairs to create manageable overviews, then adjust the parameters to provide more data for focused views of specific contexts in CG and TG.

7. Expert Case Studies

To evaluate the applicability of our approach we conducted a qualitative user study with three *subject matter experts* (SMEs) from political science (in the following referred to as E1, E2, and E3). All three experts analyze multi-party conversations in their research; in particular they focus on the effects of different modes of communication for reaching consensus. We used the pair analytics method of Kaastra et al. [KF14], in which one member of our research team acted in the *visual analytics expert* (VAE) role working with the SME. In addition, another researcher was present to observe the interaction between the VAE and SME and aspects of insight generation. The VAE and observer roles were consistent for all participants. Due to the large number of features and views in NEREx and the limited time of highly qualified experts to learn a new interface, pair analytics is appropriate. It removes any confounds due to SMEs learning the interface, while focusing the team on domain-specific questions and insights. Each two hour session began with an overview of the views and features of NEREx, and a semi-structured interview to gain feedback on these aspects. This was followed by an open-ended analysis of two different datasets, which participants selected from a set of three corpora:

D1 Three moderated debates of the 2016 US presidential election between Donald Trump and Hillary Clinton.
D2 One day of oral arguments of the US Supreme Court for the case Bush vs. Gore (Dec 11th, 2000).
D3 2016 US presidential primary debates between the leadership candidates, by party (8 Democratic and 11 Republican debates).

The VAE controlled the interface, with input from the SME, who was given a pointing device to indicate regions of interest. For simplicity of explanation, in the description below, interaction events attributed to the SMEs were directly requested to the VAE who carried them out. Sessions were audio-recorded, screen-captured, and observed by another member of the research team, who took notes.

In the following, we report results from datasets [D1] and [D3], since these conversations are well known to a large audience and do not contain legal jargon, as in [D2]. During the study, we observed that analysis tasks were generally performed in iterative cycles. Analyses started with data exploration to find a topic of interest. The SME raised hypotheses during this initial exploration, which were then verified using the different views created by the SME and VAE in collaboration, before moving on to a new question. Thus we structure the following discussion of the study outcomes according to the high-level analysis tasks supported by NEREx.

7.1. Data Exploration and Hypotheses Generation

The pre-election presidential debates in the United States have a long tradition and are customary for the candidates of the two major political parties before the general election. These moderated debates are broadcast on television and radio and watched by millions in the US and abroad. Due to the relevance of these debates to the research of direct democracy, our political science experts were interested in exploring patterns in the most recent debates [D1].

To get an overview of the complete corpus, each pair started by exploring the entity graphs of all three debates combined. E2 was interested in a high-level overview of the data, therefore, the VAE increased the minimum entity-pair frequency and the similarity threshold. The resulting graph depicted some general content clusters, as shown in Figure 1. E2 discovered some predominant topics on this high-level graph, such as *Taxes*, *War on Terror*, *Women*, *Jobs*, and *Gun Regulations*. These subjects were also consistent with the findings of the two other SMEs and confirmed their expectations. They were also all quick to find pointers to *Trump's populist rhetoric* by spotting, for example, his slogan "*Make America Great Again!*", as well as the entity pair *Obama* $\rightarrow$ *fault?* on the high-level entity graph. This made all SMEs wonder about the role of slogans and populist language in political debates. In particular, E2 derived the hypothesis **[H1]** that "Trump will have a more populist rhetoric and will not be as inclusive as Clinton".

E1, on the other hand, was interested in exploring the mentions of the keyword *women* throughout the debate. Therefore, the VAE searched for this keyword to find it in the different views — it was mentioned 68 times throughout the three debates. E1 started by exploring the relations between entities in the graph view that are linked to *women*. The related entities he found were mostly negative emotion indicators, e.g., *belittling*, *embarrass*, *insult*, *grabbing*, *attacks*, *sexual assault*, and *pigs*. He also found entities with a positive connotation like *kiss* and *respect*, as well as the number *nine*. After exploring this subgraph surrounding the entity *women*, E1 suggested that **[H2]** "Clinton will be raising more issues about women throughout the debate".

E1 was also interested in the further analysis of entities related to the topics *war on terror* and *foreign policies of the US*, as he expected these topics to be more dominant in the overall entity graph. To explore this hypothesis, the VAE created a concept graph using containers about these topics. The VAE and E1 agreed to try a location-anchored layout, see Figure 5. This concept graph revealed a number of interesting subtopics, beyond the *war on terror*, such as the developments in Iraq after the war; the fight against *ISIS* and the alliance with *NATO* and Europe; the Russian airstrikes in Syria; the nuclear thread and related sanction in Iran; some mentions of prominent political figures in the US. Using linking and brushing the expert could quickly identify the statements of both politicians towards each of these subtopics.

7.2. Hypothesis Verification

Throughout their analysis, the three experts used the linked views to verify their individual hypotheses. While E1 and E2 were interested in analyzing the structure and content of the debates and conversations, E3 was mainly interested in analyzing the development of emotions throughout the debates. In particular, he was interested in exploring how the amount of positive and negative emotion indicators change over time and the usage of diverse usage of emotion indicators by the candidates. In this section, we will discuss how the two example hypotheses from Section 7.1 were verified by E1 and E2, respectively.

[H1] In order to inspect populist rhetoric, E2 selected relations indicating such discourse from the entity graph. He then requested the VAE to switch to the entity-level view in order to inspect the relations between these entities and read their respective text-segments. He noticed that most of the utterances in the selection were attributed to Trump. Hence, he was especially interested in analyzing utterances of other speakers using the same rhetoric. He concluded that these instances were occasions where Clinton attacked Trump on the ground of his populist language. E2 could verify his hypothesis and find specific text passages as references to support his claim.

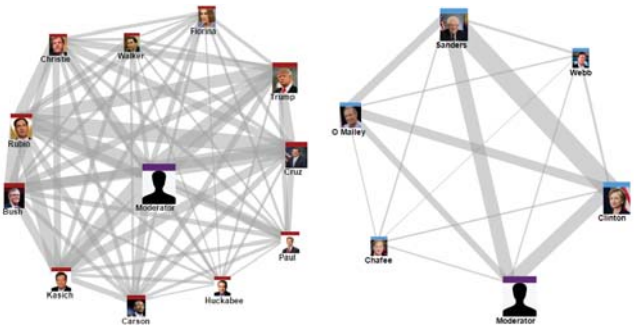
[H2] To inspect the claim that Clinton would be raising more issues about women throughout the debate, E1 filtered for the entity *women* and saw in the overview entity level view that Clinton indeed has more mentions of this entity. He also was interested in relations associated with *women* and selected these for further inspection. E1 was especially focused on exploring mentions of these relations by the different speakers. The VAE, therefore, switched to the speaker graph and used it to analyze entity pairs related to *women*, which E1 selected for a detailed analysis in the text level view. Overall, E1 observed that both Clinton and the moderator attacked Trump on his behavior towards women. However, Trump consistently repeated that he has great respect towards women.

7.3. Temporal Analysis

During the analysis of the first presidential debate, E2 observed a strong relation between the entities *release*, *tax*, and *returns* (see Figure 4). He therefore became interested in looking at the topic *taxes* in more detail. To do so, the VAE created a concept graph with relevant entities (Figure 6), using a speaker-anchored layout to analyze what the contributions of the two candidates were for this topic. E2 noticed that Clinton had a strong correlation to the entity *release*, as she was pushing her opponent to *release his tax returns* of the last years and accusing him of *slashing taxes*. Trump, on the other hand, defended himself by mentioning that he is under *routine audit* and would release his tax returns when it is finished. E2 also commented that this graph shows the proposed cuts by the two candidates, for Trump *\$13 trillion*, while Clinton was talking about *\$5 trillion*. Clinton also talked about increasing jobs from 10 to 35 million. While Trump talked about lowering the tax rates from 35% to 15%. E2 could verify his hypothesis and get more background knowledge though interactively selecting interesting entity-relations and switch to the entity level view to get more context and read the corresponding utterances.

After the detailed analysis of the topic *taxes* in the first debate, E2 became interested in exploring the development of this topic throughout the three debates. Based on the news he had heard, he suspected that Clinton would continue bringing up the release of her opponents tax returns at multiple points in time throughout the debates. For this analysis, the VAE created a temporal graph (Figure 7) for the concept *taxes* for all debates combined. Using the temporal animation, E2 could confirm that the topic *taxes* was strongest in the first debate, but was relevant in the other two as well. He also saw that Clinton and the Moderator brought up the release of Trump's tax return at multiple occasions in the debate, noting from the text level view that they compared his behavior to all previous presidential candidates who all have released their tax returns. In his defense, Trump mentioned that he had released some financial statements showing an income of *\$694 million* in the past year. E2 was also astonished to observe Trump switching the topic to the failures in the financial system, the need for new jobs, and how he would improve the situation of American businesses.

7.4. Comparative Analysis



(a) *Republican Graph.* **(b)** *Democratic Graph.*

Figure 9: *Comparison of the speaker graphs of the presidential candidates of both major parties.*

For the comparative analysis, we collected all debates from the 2016 U.S. presidential primary elections of the two major political parties. We grouped together all debates for each party to create two large corpora. All experts in our study were unfamiliar with the details of these debates.

Figure 9 shows the speaker graphs for both parties. By comparing both speaker-networks in Figures 9a and 9b, it is immediately apparent that the graph of the Republican candidates is denser and includes more speakers than that of the Democrats. The Republican party had more leadership candidates in the 2016 election. However, during the first few months, many candidates from both parties withdrew their candidacy. A glance at the graphs quickly makes obvious the most influential candidates in both parties. For the Democrats, they are Clinton, Sanders, and O'Malley. For the Republicans, they are Trump, Cruz, and Kasich. However, there are also other candidates in the Republican graph, like Rubio, Carson, and Bush, who have a significant presence. This is due to the very late withdrawal of their candidacy.

When analyzing the complete semantic maps of both datasets, the experts identified a set of common topics between both parties, such as *education*, *health-care*, *immigration-reform*, *gun-control*, *economy*, *foreign policies*, and the *war against terror*. Yet, some topics had a greater focus in one corpus and were not discussed as vigorously in the other. One example is the *tax-cut* topic that was very dominant in the Republican entity graph but not present in the Democratic graph. However, in both, the most salient named-entities related to the *war on terror*. E2 was interested in the subtle

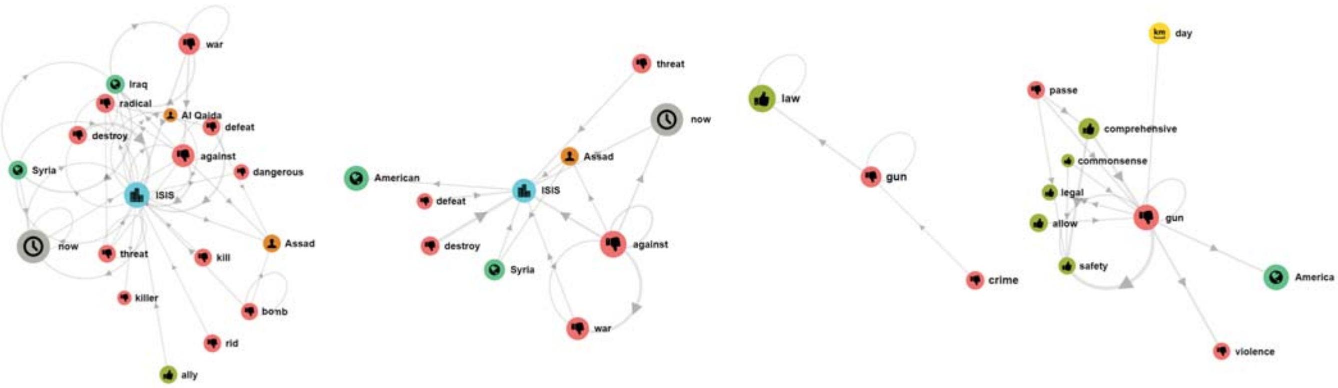


Figure 10: *Comparison between Democratic (D) and Republican (R) debates. From l-r: (R) on ISIS, (D) on ISIS, (R) on guns, (D) on guns.*

differences between parties on this topic. To support this analysis, the VAE created several containers for concepts, such as war, terror, and geo-locations for both datasets, and generated concept graphs using these containers. E2 then derived a hypothesis that controversial topics, such as war on *terrorism* or *gun-control*, are not treated with the same priority by both parties. Therefore, he selected the central entity for each topic in the entity graphs of both parties. By analyzing the sub-graphs around the selected entities, he drew conclusions about the importance of a topic for each party. Figure 10 shows the sub-graphs around the entity nodes *ISIS* and *gun* for both parties. These graphs show that while a topic, such as *gun-control* was important in the Democratic debates, it was not a central topic for the Republicans. Republicans, however, discussed the subject of *terrorism* more intensely.

7.5. Pair Dynamics and Usability

Through the study sessions, we recorded and observed the dynamics between the SMEs and VAE, and provide some comments here, following Kaastra et al. [KF14]. We observed instances of each of the three communicative mechanisms of joint action: grounding, advancing, and repairing. For example, when analyzing the presidential debates, common ground was obtained conversationally at the start of the session through reflection on shared knowledge of the recent debates (which were heavily discussed in the news). Deictic gestures such as pointing with the hand (usually the SME), or the mouse pointer (usually the VAE) were also used alongside verbal cues both specific "the node labelled 'election"' and general "the red node *there*" to establish common frames of reference.

To advance the analysis, the VAE employed a series of suggestions worded as questions to the SME, e.g., "Are you interested in the high occurrence of negative emotion?", "What about the temporal evolution of the debate?". The SME tended to reflect on the on-screen views aloud, and pose data-oriented questions "I wonder if there is more negative emotion associated with Trump than Clinton?" in order to guide the VAE to views of interest. As in past pair analytics studies [AHKGF11] we also observed the use of different continuation words ("mmm-hmm, yeah") to indicate continued interest in a view, and interjections to indicate a vertical transition to a new question ("okay, all right, no"). Some usability issues arose when the SME asked questions which were not easily answered by the tool, for example, to see all emotion words related to a single concept (concept graphs require two concepts). In these instances, a repair event occurred. Repair coordinations took place mainly through verbal corrections: the VAE explaining the capabilities of the tool or correcting a misinterpretation of an on screen view; the SME clarifying information needs ("no, I want to just see one concept and the associated emotions"). Gesture was also used here — in a few instances, the SME acquired the mouse to carry out a pointing or selection action directly.

Some parts of NEREx were more requested than others by the SMEs. The entity graph was the most commonly requested view as a starting point in analysis. This may be due to the relatively straightforward nature of this view, which gives an overview of the data. In addition, there were powerful features of the system which SMEs did not initially remember to use, perhaps because they were not immediately visible on screen. When reminded of the visual querying feature in particular, SMEs made heavy use of it.

7.6. Mediation Analysis

To demonstrate the general nature of NEREx , in addition to political debates and court arguments, we used NEREx to analyze a more complex multi-party conversation, the mediation process of the Stuttgart 21 (S21) project. This nine-day arbitration on a controversial railway and urban development project in Germany comprises a corpus of around 6,000 utterances, involving 60 speakers. In contrast to the datasets used for the pair-analytics study, this conversation contains more crosstalk, off-topic discussions, and has

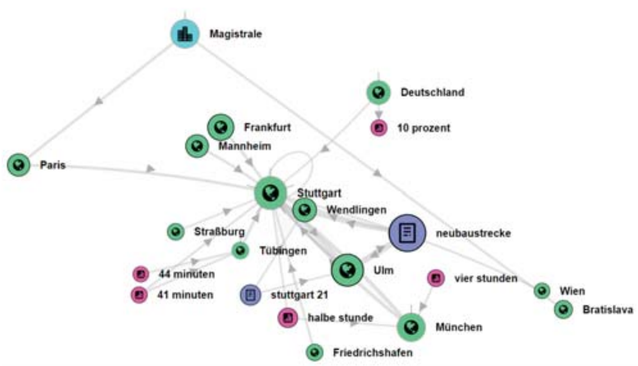


Figure 11: *A location-anchored concept graph based on one day (Nov 4, 2010) of the Stuttgart 21 mediation.*

many interruptions and ungrammatical sentences. In this use case, we illustrate the applicability of our approach on a German multi-party conversation.

Due to the complex nature of the mediation, the discussion was broken into multiple topics which were heavily discussed by the proponents and opponents of the S21 project. In addition, invited experts illustrated plans from both camps. One central discussion point was the construction of a new high-speed rail track (DE: Neubaustrecke) between the cities of Wendlingen and Ulm. This new construction will contribute to significantly shorten the commute time of the main connection (Magistrale) for trains between Paris and Bratislava. As shown in Figure 11, the discussion on Nov 4, 2010 focused on the construction of this new track and its positive effect of reducing commute times. In the figure, the affected connection locations related to the trains passing through the city of Stuttgart reconstruct an abstract railroad map of southern Germany. This location-anchored graph was constructed from all locations and frequent concepts connected to the central node of Stuttgart.

Given the complexity of the mediation and the large scale of the dataset, using NEREx for exploration and analysis has an immense added value, especially in getting an overview of all subtopics and their relation to different speakers.

8. Discussion and Lessons Learned

NEREx has been generally well received by the domain experts. They said the tool added value to analysis of deliberative conversations, especially for previously unknown data. All experts confirmed that our tool supports all our targeted analysis tasks. However, E1 commented that although he could perform all tasks with the tool, he "trusted the numbers more" when it comes to hypothesis verification, referring to the traditional empirical analysis methodologies in political science.

Generally, all experts appreciated the interface design and interactivity of the tool, commenting that "[...] the tool has good aesthetics and well-chosen colors." (E2). Overall, all experts could generate findings and insights with the tool, e.g., E3 commented: "It's fascinating to see how Trump manages to get from any question to ISIS." He found the entity graph to be a particularly useful view of NEREx, as it gives a high-level overview of a complete discussion while allowing providing details on demand. Additionally they expressed their desire for extensions and improvements. For example, they would like to split the screen to compare two views side by side. Two experts also suggested giving the tool a stronger focus on sentiment analysis of conversations. One expert suggested we continue to investigate alternative methods for relating entities through grammar.

When asked about the contributions of NEREx, the experts commented: "The tool is very good for exploration. It helps in generating many ideas that can lead to hypotheses. It also helps in finding out if we have 'enough data' to analyze a particular subject in a debate." (E1), i.e., whether certain keywords can be expected to correlate in his statistical analysis. "The main strength of this tool is to find questions and get a good idea of what the answer might be." (E2). "This tool does not only support the generation of hypotheses about the data but also about how things can be measured."(E3), referring to the refinement of traditional statistical models in political science for measuring certain aspects of a debate. In addition, E2 suggested that NEREx "[...] might be good to use for educational purposes or as an exploratory presentation tool to get a better picture of the key elements of a debate."

We learned several key lessons about the design of NEREx which could influence future systems created for domain experts. In particular, throughout the iterative design process we attempted to balance ease of use with powerful functionality. We provided multiple views on the data, and chose to reduce complexity by keeping each on a separate screen, eschewing a coordinated multiple views (CMV) approach. While the end result is less cluttered than a CMV, the placement of views off screen meant that SMEs often forgot they existed. We also learned that providing such a rich tool can be effective to explore complex relations in the data, but there were opportunities to provide greater utility to the domain experts. For example one expert requested to have statistics integrated in the views for further analysis of causal relations. Another wanted to extract the structured findings in a way he could use other software. Two experts suggested that the system should *learn* from manual grouping actions and propagate the groupings to other nodes automatically, to speed up the graph curation process.

9. Conclusion and Future Work

We presented NEREx, a visual analytics framework for the exploratory analysis of verbatim conversation transcripts. Our approach explores the relations of named-entity pairs based on a distance-restricted entity-relationship model. We presented six linked, interactive views tailored to the analysis of multi-party conversations. We evaluated the applicability of our approach for five analysis tasks with a pair analytics study with three political scientists. Overall, NEREx has been well received by the domain experts, who gained new insight into familiar and unfamiliar datasets.

In future work, we would like to extend the capabilities of our approach to incorporate new features and improve data extraction. In particular, we would like to categorize entities by their specificity to the given corpus and use this information to highlight potential transitive chains in the entity graph. Furthermore, we plan to design more tailored views to support other text types and analysis tasks. To achieve this, we plan to create additional elementary categories by using other text features and named-entities. Finally, we would like to extend NEREx by logging provenance and interaction data to enable storytelling. The software is available as a web-service to the public for non-commercial purposes, as part of the VisArgue framework - `http://visargue.inf.uni.kn/`.

References

[AGL*07] AUVIL L., GROIS E., LLORÀ X., PAPE G., GOREN V., SANDERS B., ACS B., MCGRATH R.: A flexible system for text analysis with semantic network. *Digital Humanities* (2007), 17–20. 3

[AHKGF11] ARIAS-HERNANDEZ R., KAASTRA L., GREEN T., FISHER B.: Capturing reasoning processes in collaborative visual analytics. In *Proc. of Hawaii Int. Conf. on System Sciences 44* (2011). 11

[AKV*14] ALEXANDER E., KOHLMANN J., VALENZA R., WITMORE M., GLEICHER M.: Serendip: Topic model-driven visual exploration of text corpora. In *Proc. of Visual Analytics Science and Technology* (2014), pp. 173–182. 3

[AM03] ASAHARA M., MATSUMOTO Y.: Japanese named entity extraction with redundant morphological analysis. In *Proc.of NAACL HLT* (2003), Association for Computational Linguistics, pp. 8–15. 2

[ARLC*13] ABDUL-RAHMAN A., LEIN J., COLES K., MAGUIRE E., MEYER M., WYNNE M., JOHNSON C. R., TREFETHEN A., CHEN M.: Rule-based visual mappings–with a case study on poetry visualization. In *Computer Graphics Forum* (2013), vol. 32, pp. 381–390. 3

[ASW12] ANGUS D., SMITH A., WILES J.: Conceptual recurrence plots: Revealing patterns in human discourse. *Visualization and Computer Graphics, IEEE Transactions on 18*, 6 (2012), 988–997. 3

[BB07] BACH N., BADASKAR S.: A review of relation extraction. *Literature review for Language and Statistics II* (2007). 2

[BK09] BERGSTROM T., KARAHALIOS K.: Conversation clusters: Grouping conversation topics through human-computer dialog. In *Proc. of SIGCHI* (2009), ACM, pp. 2349–2352. 3

[BMSW97] BIKEL D. M., MILLER S., SCHWARTZ R., WEISCHEDEL R.: Nymble: A high-performance learning name-finder. In *Proc. of the Conf. on Applied Natural Language Processing* (1997), Association for Computational Linguistics, pp. 194–201. 2

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D3: Data-driven documents. *Trans. Visualization & Comp. Graphics (InfoVis)* (2011). 8

[Cob05] COBURN A.: Text modeling and visualization with network graphs. In *Proc. of the Digital Humanities* (2005). 3

[CS92] COATES-STEPHENS S.: The analysis and acquisition of proper names for the understanding of free text. *Computers and the Humanities 26*, 5-6 (1992), 441–456. 2

[DA08] DANG V. B., AIZAWA A.: Multi-class named entity recognition via bootstrapping with dependency tree-based patterns. In *Advances in Knowledge Discovery and Data Mining*. Springer, 2008, pp. 76–87. 2

[DV02] DONATH J., VIÉGAS F. B.: The Chat Circles series: Explorations in designing abstract graphical communication interfaces. In *Proc. of the Conf. on Designing Interactive Systems* (2002), ACM, pp. 359–369. 3

[EA15] EL-ASSADY M.: *Incremental Hierarchical Topic Modeling for Multi-Party Conversation Analysis*. Master's thesis, University of Konstanz, 2015. 3

[EAGA*16] EL-ASSADY M., GOLD V., ACEVEDO C., COLLINS C., KEIM D.: ConToVi: Multi-Party Conversation Exploration using Topic-Space Views. *Computer Graphics Forum 35*, 3 (2016), 431–440. 3

[FGM05] FINKEL J. R., GRENAGER T., MANNING C.: Incorporating non-local information into information extraction systems by gibbs sampling. In *Proc. of the Annual Meeting of the Association for Computational Linguistics* (2005), pp. 363–370. 3

[FM09] FINKEL J. R., MANNING C. D.: Nested named entity recognition. In *Proc. of Empirical Methods in Natural Language Processing* (2009), Association for Computational Linguistics, pp. 141–150. 3

[GM04] GROBELNIK M., MLADENIC D.: Visualization of news articles. *Informatica (Slovenia) 28*, 4 (2004), 375–380. 2

[GREA15] GOLD V., ROHRDANTZ C., EL-ASSADY M.: Exploratory text analysis using Lexical Episode Plots. In *Eurographics Conference on Visualization (EuroVis) - Short Papers* (2015), Bertini E., Kennedy J., Puppo E., (Eds.), The Eurographics Association. 3

[GT09] GONZALEZ E., TURMO J.: Unsupervised relation extraction by massive clustering. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on* (2009), IEEE, pp. 782–787. 2

[GWFI14] GOFFIN P., WILLETT W., FEKETE J.-D., ISENBERG P.: Exploring the placement and design of word-scale visualizations. *Trans. on Visualization and Computer Graphics 20*, 12 (2014), 2291–2300. 3

[Hab84] HABERMAS J.: *The Theory of Communicative Action*, vol. 1. Beacon Press, 1984. 1

[HC14] HOQUE E., CARENINI G.: Convis: A visual text analytic system for exploring blog conversations. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 221–230. 3

[JFCS15] JÄNICKE S., FRANZINI G., CHEEMA M. F., SCHEUERMANN G.: On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. *The Eurographics Association* (2015), 83–103. 1, 3

[Kam04] KAMBHATLA N.: Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In *Proc. of ACL Interactive poster and demonstration sessions* (2004), Association for Computational Linguistics, p. 22. 2

[KF14] KAASTRA L. T., FISHER B.: Field experiment methodology for pair analytics. In *Proc. of the Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization (BELIV)* (2014), ACM Press, pp. 152–159. 9, 11

[KSBK*16] KUCHER K., SCHAMP-BJEREDE T., KERREN A., PARADIS C., SAHLGREN M.: Visual analysis of online social media to open up the investigation of stance phenomena. *Information Visualization 15*, 2 (2016), 93–116. 3

[Lev66] LEVENSHTEIN V. I.: Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (1966), vol. 10, pp. 707–710. 5, 6

[LPH*09] LESHED G., PEREZ D., HANCOCK J. T., COSLEY D., BIRNHOLTZ J., LEE S., MCLEOD P. L., GAY G.: Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proc. of SIGCHI* (2009), ACM, pp. 537–546. 3

[ML03] MCCALLUM A., LI W.: Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *Proc. of NAACL HLT* (2003), Association for Computational Linguistics, pp. 188–191. 2

[NHBR13] NGUYEN V., HU Y., BOYD-GRABER J. L., RESNIK P.: Argviz: Interactive Visualization of Topic Dynamics in Multi-party Conversations. In *Human Language Technologies* (2013), pp. 36–39. 1

[NS07] NADEAU D., SEKINE S.: A survey of named entity recognition and classification. *Lingvisticae Investigationes 30*, 1 (2007), 3–26. 2, 3

[PEG13] PERSSON M., ESAIASSON P., GILLJAM M.: The effects of direct voting and deliberation on legitimacy beliefs: An experimental study of small group decision-making. *European Political Science Review 5*, 3 (Nov 2013), 381–399. 1

[RCE*11] RITTER A., CLARK S., ETZIONI O., ET AL.: Named entity recognition in tweets: an experimental study. In *Proc. of the Conf. on Empirical Methods in Natural Language Processing* (2011), Association for Computational Linguistics, pp. 1524–1534. 2

[SGH12] SHAHAF D., GUESTRIN C., HORVITZ E.: Trains of thought: Generating information maps. In *Proc. of the Int. Conf. on World Wide Web* (2012), ACM, pp. 899–908. 3

[SGL08] STASKO J., GÖRG C., LIU Z.: Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization 7*, 2 (2008), 118–132. 2

[SGS98] SEKINE S., GRISHMAN R., SHINNOU H.: A decision tree method for finding and classifying names in Japanese texts. In *Proceedings of the Sixth Workshop on Very Large Corpora* (1998). 2

[SGS11] SUN A., GRISHMAN R., SEKINE S.: Semi-supervised relation extraction with large-scale word clustering. In *Proc. of NAACL HLT* (2011), Association for Computational Linguistics, pp. 521–529. 2

[SN04] SEKINE S., NOBATA C.: Definition, Dictionaries and Tagger for Extended Named Entity Hierarchy. In *LREC* (2004), pp. 1977–1980. 2

[TS12] TKACHENKO M., SIMANOVSKY A.: Named entity recognition: Exploring features. In *KONVENS* (2012), pp. 118–127. 2

[VHWV09] VAN HAM F., WATTENBERG M., VIÉGAS F. B.: Mapping text with Phrase Nets. *IEEE Trans. on Visualization & Computer Graphics*, 6 (2009), 1169–1176. 3

[ZAR03] ZELENKO D., AONE C., RICHARDELLA A.: Kernel methods for relation extraction. *The Journal of Machine Learning Research 3* (2003), 1083–1106. 2