

“And That's the Way it is, for Now...”

**Using Network Analysis, Event Extraction, and Named Entity Recognition to
Track Dynamic News Narratives in Written Media Coverage**

Polly McKim, M.S.

Nima Zahadat, Ph.D.

The George Washington University

Abstract

As newsworthy events unfold, their narrative coverage evolves as well. It is difficult to keep up with rapidly changing stories. Combining Natural Language Processing with Network Analysis algorithms and techniques and a highly graphical presentation can help show how narratives change. To achieve this goal, first, network analysis algorithms were employed to aggregate together news narrative chains. Then, event extraction techniques based on journalistic structures were employed to determine the most important elements being reported in each article of the different narratives chains. Other network analysis techniques and visualization techniques were then used to develop a highly visual method for clearly differentiating the changing narrative threads over time. Additionally, Named Entity Recognition was used to detect people, countries, locations and organizations in the finalized narrative clusters. Finally, analysis was done into how mentions and associations of these entities changed over the course of each narrative chain to quantify how a narrative evolved. The resulting analysis of these elements graphically shows how individual news narrative chains change over time.

Keywords: News Narrative Chain Aggregation, Dynamic News Narrative Analysis, Event Extraction, Network Analysis, Dynamic News Tracking, Dynamic Named Entity Recognition Modeling

Introduction

In the modern information age, written news content comes to readers at rates never before seen in history. Today, the public can go online immediately after, or even during, a newsworthy event and read coverage immediately. As events unfold, more information becomes available and the reporting details and focus of the coverage used to convey news to the public changes rapidly. What is reported as one thing often quickly evolves into something else. Keeping up to date, assessing how the important elements evolve, and getting a complete sense of the whole story can be difficult.

News coverage tends to follow generally identifiable patterns. For example, coverage of an extreme weather event would customarily go through several phases. The first articles would report on a predicted severe storm. The coverage would then mutate into stories detailing the location and intensity of the storm. Then, subsequent articles would likely cover relief efforts and effects of the storm. All of these stories together would make up one narrative chain about a single central event, the storm, although other events would fit into this overarching puzzle of coverage surrounding the main event.

While smaller newsworthy events may garner only a few articles, a particularly noteworthy and high-profile situation can generate a large number of stories over many days, weeks or even months. These stories can ultimately have many different threads corresponding to changing subjects, locations, and smaller related events that arise as coverage of the central newsworthy event continues to unfold. These disparate threads come together to create the larger news “narrative” of the noteworthy event. This type of coverage, or “media storm,” differs from the customary news cycle in the volume and time span of coverage (Boydston et al., 2014).

The types of events triggering multiple articles are likely to be more important than those covered only once and are therefore, more worthy of analysis. Analyzing the discrete pieces of the story chain as part of an overall narrative, rather than as individual articles, can potentially offer many benefits. When stories are considered as part of a chain, rather than as individual disparate pieces of information, analysts can develop a better understanding of how and why events unfold as they do. This type of assessment capability would be highly useful in determining ways to improve response strategies to mass casualty events, such as terror attacks or natural disasters. Additionally, evaluating story chains in their entirety would enable analysts to more easily detect where instances of disinformation were introduced into a narrative and how that information subsequently spread.

Despite the potential benefits of identifying narrative chains using written news text, no clear method for solving this problem is commonly accepted or in widespread use. Currently, search engines successfully exploit information retrieval methods to return articles that would likely be of interest based on keywords. However, none of the ranking algorithms currently used by major search engines organize the search results in a fashion that reflects the evolution of the story (Zhu & Oates, 2012).

Many government intelligence organizations analyze a vast amount of open-source intelligence (OSINT), including written news media, in their ongoing efforts to protect national

interests. Currently, media monitoring efforts often aggregate and code stories that track the same overarching event by hand. Analysts then must-read stories in their entirety to track how the stories change and develop overtime (Herrera-Cubides et al., 2020). This is, however, a time-consuming process that does not easily scale.

Additionally, even if stories are organized into groups of articles that cover the same event, comprehension can still be difficult. Many articles summarize previously reported news before reporting on new events, leading to superfluous information (Hamborg et al., 2019). Or, conversely, coverage may not give enough background information to facilitate a full understanding. “Information overload,” a term coined by Shahaf to refer to the difficulty of keeping up with the amount of content constantly being published, gets to the heart of the issue. Too much information and rapidly changing overlapping chains can obscure key points, making it more difficult for readers to follow and to glean helpful knowledge (Shahaf & Guestrin, 2010).

Successfully identifying the central subjects, events and locations is critical to understanding the factors at play and in turn, understanding the overarching narrative and avoiding getting lost in the noise. Further, discerning how the people, locations, and organizations figure into a narrative chain represents another key metric for determining how news narratives change overtime, as understanding the evolution of the entities involved mirrors the progression of the narrative.

Finding a method to easily identify and communicate the key points of a news event and then map how they change over time would also be highly useful for the previously mentioned use cases. This project uses unsupervised machine learning algorithms, network analysis, and visualization techniques to analyze changing news narratives by aggregating articles together into story chains, extracting key information from each article, and graphing the results.

Literature Review

The use of machine learning to analyze various aspects of news coverage is a commonly studied theme in natural language processing (NLP) and many previous projects were critical to the development of this project. This project focused on three distinct tasks, aggregating together the narrative chains, extracting the key information from the articles, and visualizing how the information changes over time. Therefore, research into previously used methods focused on these three specific areas. Although automated story chain aggregation systems are not widely used currently, several recent research projects have focused on developing a suitable algorithm or method. A pioneering project developed in this space sought to combat the issue of “information overload” in news stories, by creating a chain of stories to link together two separate stories. The methodology developed in the project resulted in a system where, if given two different news articles, the model would return a series of articles that “linked” the two input articles together to form a smooth story chain. The project accomplished its goal by constructing a definition for coherence based on influence of the important words, where the value of the coherence is formalized and solved as a Linear Programming problem. From there, the random walk algorithm is used to create bipartite graphs that measure influence between various articles

(Shahaf & Guestrin, 2010). The output from this model is a specially selected set of articles chosen from the original input corpus which best link the provided starting and end point articles. A subsequent project built on the ideas pioneered by Shahaf & Guestrin. This project focused on the issue of limiting redundancy and improving efficiency, which were factors not addressed in Shahaf & Guestrin's algorithm. The project still used a random walk algorithm but introduced additional pruning to remove redundant articles or poorly related articles (Zhu & Oates, 2012).

Additional research into methods to link story chains yielded another, more recent algorithm that was particularly helpful for this project. Similar to the two previous projects described, this proposed method also used key word similarity in determining related articles. However, this method did not call for a fixed start and end point to be pre-identified and instead used network analysis algorithms to cluster all the articles in a corpus into communities where each article was considered to be a member of only one community, or in this case a single narrative chain. This project explored combining two different information retrieval techniques, a novel scoring technique to select the most unusual words in a document and the BM25F algorithm, to assess the pairwise similarity between different news articles published within three days of each other, and then treating these similarity measures as connections in a network of articles. From there, the project used the Infomap algorithm, which was originally developed as a method of community detection in the network analysis field, to identify and cluster related stories (Rosvall & Bergstrom, 2008) & (Bohlin et al., 2014). The paper demonstrated that the proposed technique was successful on a corpus of 40,000 news stories from multiple sources (Nicholls & Bright, 2019). Nichols and Bright, the authors of the paper, also released the code set to accompany their project (Nichols/Bright algorithm). This code was especially useful for this project and ultimately was adapted as the unsupervised learning method used for clustering of the corpus. Information about how the code was modified is detailed in a subsequent portion of this paper.

For analyzing coverage of evolving narratives, applicable projects were more difficult to find. A substantial amount of previous research into shifting narratives has focused on the area of sentiment analysis, including examples of its use to track changing public attitudes, particularly in the political arena (Soelistio & Surendra, 2013). To date, there has not been as much investigation into mechanization for determining how the substantive facts of a story chain change. Nevertheless, information extraction, more specifically event extraction, which focuses on automatically obtaining specific details of instances from raw text, is a particularly popular area of research in NLP and offers possible insight into how to solve this aspect of the project.

Current approaches to event extraction center mainly on three different methods. First, there are data-driven approaches, which aim to convert data to knowledge through the usage of statistics, machine learning, and linear algebra. Second, there are pattern and knowledge-driven methods which extract through representation and exploitation by means of pattern-based approaches. Finally, there are hybrid event extraction approaches which combine the two previous strategies (Hogenboom et al., n.d.). As news articles are reporting on events, it follows

that performing event extraction after successfully clustering story chains together, and mapping the changes overtime, would give a sense of how the narrative chain evolved.

One event extraction method, the “Giveme5W1H” algorithm falls under the third method, using syntactic and domain-specific rules to perform extraction and was uniquely well suited for use in this project since it was developed specifically for event extraction from written news content (Hamborg et al., 2019). The system’s output directly mirrors the “5 Ws 1 H” structure, recommended by leading journalism practices. “5 Ws 1H” which stands for “who”, “what”, “where”, “when”, “why,” and “how”, is a widely taught journalistic framework that encourages writers to home in on the central aspects of an event in order to report the most important information (Ojo & Heravi, 2018). The code for the “Giveme5W1H” algorithm was designated as open source by its creators and served as the foundation for the event extraction portion of this project.

The final focus of research for this project focused on methods for clearly visualizing news information. Two projects were particularly helpful. The first project, also devised by Shahaf, developed a methodology for creating structured summaries of information, which she dubbed “metro maps,” where an algorithm generated a concise structured set of documents which maximized coverage of salient pieces of information. The metro map algorithm also showed the relationships among retrieved pieces in a way that captured the evolution of a story. (Shahaf et al., 2013). Another project, “Story Analyzer,” combined NLP and visualization libraries to provide readers with a “one stop shop” to get all aspects including background, commentary, and other information regarding news stories. The end resulting dashboards gave a highly visual representation of information captured in the news (Mitri, 2020).

Research Methodology

For the purposes of this paper, a news event refers to a noteworthy occurrence covered by the media. A story refers to a single article about an event, while the terms story chain and narrative chain are used interchangeably to describe multiple articles providing coverage about the same event. Article cluster is used to describe all articles detected to be members of the same narrative chain.

The code sets developed and publicly released for the Nichols/Bright algorithm and “Giveme5W1H” algorithm respectively served as the backbone of the data processing for this project. The clustering algorithm was modified and run across the entire corpus and successfully clustered the articles in the corpus into different narrative chains. Changes made to the algorithm included increasing the pairwise calculations window from three days to seven and the threshold for word similarity was also lowered. Although, this resulted in increased computational complexity, as it necessitated an increase in the number of pairwise comparisons that needed to be made, overestimating clusters rather than under estimating was a reasonable trade off. The overall structure of the method which used word scoring and BM25F and Infomap algorithms to determine article clusters remained unchanged. The Nichols/Bright algorithm was run across the entire corpus. After the Nichols/Bright algorithm, some minimal data cleanup was performed.

Then the dataset was fed through the Giveme5W1H event extraction algorithm, previously described in the literature review, which was also adapted for this project.

The Nichols/Bright and the “Giveme5W1H” algorithm components of this project were run on using Amazon Web Services (AWS) cloud computing platform’s EC2s to expedite the run times. Specifically, Terminal multiplex (TMUX) was used to facilitate persistent sessions which enabled asynchronous data processing. Some parts of the algorithms used parallel computing to speed up the data ingest. All configuring done inside AWS was done with BASH command line scripting.

Once, the clusters and event extraction methods were successfully run across the dataset, text standardization was performed. Next, Spacy’s Named Entity Recognition (NER) pipeline was run over the entire text corpus to return the people, places, organizations, and geo-political entities involved in the different narratives. The original variables combined with the engineered features returned by the algorithms and the NER pipeline gave the final set of data points used for analysis. Some of the features were converted into nodes and edges using the Networkx python library to enable additional analysis and visualization of the information in a network format. Converting the finalized dataset into a network format and graphing the result enabled exploration of the interconnectivity of the different elements present in the articles in each narrative chain. The same method was also used to quantify how named entities changed overtime to get a sense of how the story was evolving. Additionally, Gantt charts were used to graph how key elements of the cluster evolved over the course of the narrative chains.

Python 3.7 was used to code most aspects of this project, including procuring the dataset, performing necessary data cleaning and processing, conducting exploratory analysis, adapting non-native code sets, and creating all visualizations. The dataset was obtained by querying an API and was initially downloaded in a multiple CSV format. The network analysis was performed using the Networkx library and visualizations were created using plotly chart studio and matplotlib. Finally, the presentation and corresponding website was developed by customizing a Hypertext Markup Language (HTML) Bootstrap 5 template made available by “Nicepage.” The accompanying Cascading Style Sheet (CSS) and JavaScript (JS) used for the web page were also adapted from “Nicepage” templates. For additional details regarding the technical set up of this project and tools, please consult this project’s accompanying github page (<https://github.com/pmckim1/Capstone>).

Data Source, Processing, and Analysis

It was important for the dataset for this project to mirror the data available for situations where it would be used-- raw, unstructured news text. Therefore, the data used was obtained by querying the API for the online component of “The Guardian,” a British news outlet for articles in "Politics," "Global Development," "US News," "UK News," and "World" sections published between 1/1/2016- 03/03/2021. It is important to note that in this context, “Politics” refers to UK political coverage. In addition to the full text of each article, the API also returned data, title,

section, and the URL for each article. The Guardian was chosen as the source for this project due to the ease of querying its free API and its reputation as a high quality, generally unbiased news source. The global nature of its reporting was also a factor in selecting this source (*The Guardian - Open Platform*, 2021) & (*Theguardian / Open Platform - Documentation / Overview*, 2021). The process used in this project could likely successfully be applied to any corpus of written news media, provided the features included in the dataset were similar. The dataset returned by the API scrapper initially returned 101,213 articles. This initial dataset was then cleaned for extensive formatting issues. Additionally, obituaries, letters to the editors, and streaming blog style articles (e.g. articles that only said “Election Updates Live”) were removed.

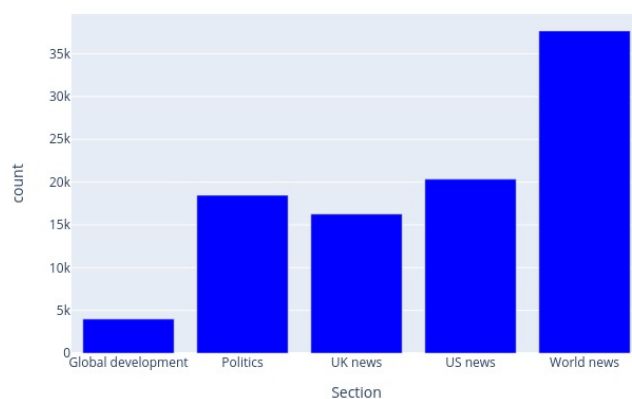
After the initial cleaning steps detailed above, the original dataset was fed into the modified Nichols/Bright and “Giveme5W1H” algorithms. Ultimately, not all of the articles were successfully processed due to textual encoding errors in the text returned by the API. Attempts to fix these encoding errors were not successful and ultimately the corrupted articles were removed from the dataset.

After the data was successfully run through the two algorithms and cleaned, the resulting dataset was 96,799 articles. These articles were then processed through Spacy’s NER pipeline to return the people, locations, geopolitical entities, and organizations. The graph to the right shows the distribution of the section classification of the articles.

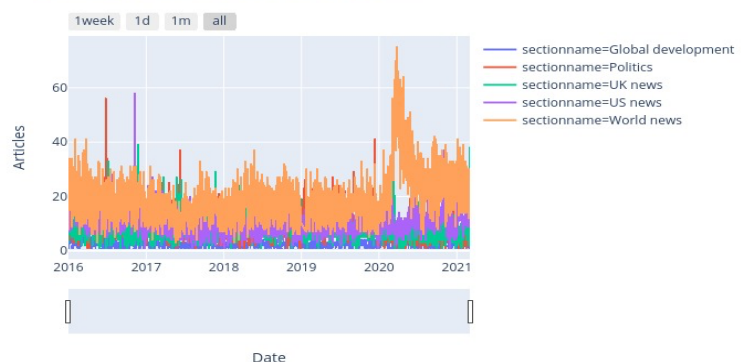
This line graph to the right shows how many articles were published each day in the different sections. World News has a large spike in March 2020, which likely coincides with Coronavirus-19 related coverage. Other spikes were seen in Politics in June 2016, likely coinciding with Brexit and in US News in November 2016, likely coinciding with the US presidential election. To view an interactive version of this graph please visit (pmckim1.github.io/EDA).

Next, looking at the results of the clustering algorithm in the pie chart below-right shows that a higher percentage of articles were not a part of a narrative chain. However, a large percentage of articles were members of a narrative chain community. This tracks with the findings from the paper discussing the

Amount of Articles in Each Section



Total Articles Published Over Time By Section



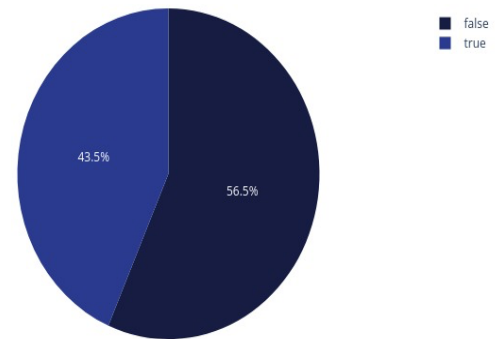
Nichols/Bright algorithm, which found that just under half of all media articles are part of a longer story chain (Nicholls & Bright, 2019).

The charts below-right which show the distribution of the number of articles in each cluster shows that most narrative chains only had two members. As the number of articles in the clusters gets larger the number of communities with that number of articles reduces. The largest narrative chain community had 79 articles and detailed coverage of the 2018 air missile strikes in Syria. Other high article clusters included the death of Qasem Soleimani, Coronavirus-19 vaccine coverage, and Hurricane Ima related articles.

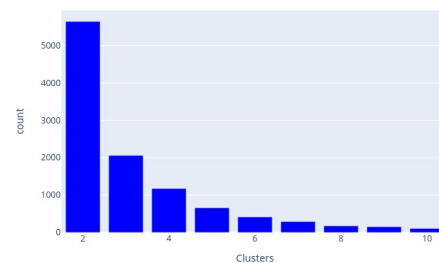
Since this project is focused on the evolution of news narratives, clusters with less than 5 articles in the community were not considered in analysis after the exploratory analysis phase. 2,067 article chains had 5 or more articles with a total of 17,685 articles among them. A visual inspection of the clusters’ headlines showed that the communities detected by the Nichols/Bright algorithm were logical. The information returned by the “Giveme5W1H” algorithm also appeared logical, but in many instances additional descriptive text around the clause of interest was also returned, which necessitated additional data clean up and impeded the analysis of these features.

The Gantt chart below was used to plot the duration of the narratives as well the number of articles in each narrative chain. The example below shows the length and amount for narrative chains with 20 articles or more taking place in the second of 2017. For additional graphs, please see the accompanying github notebooks.

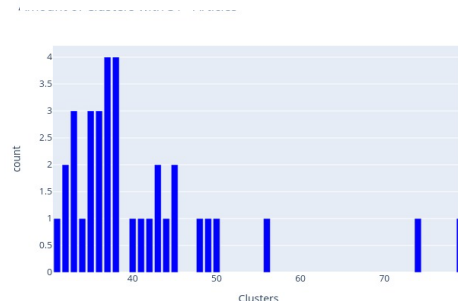
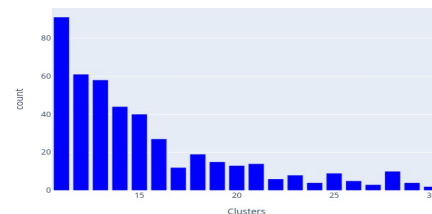
Percent of Articles Assigned to Narrative Chain Clusters



Amount of Clusters with 2-10 Articles



Amount of Clusters with 11-30 Articles



Length and Size of Narrative Clusters Second Half of 2017

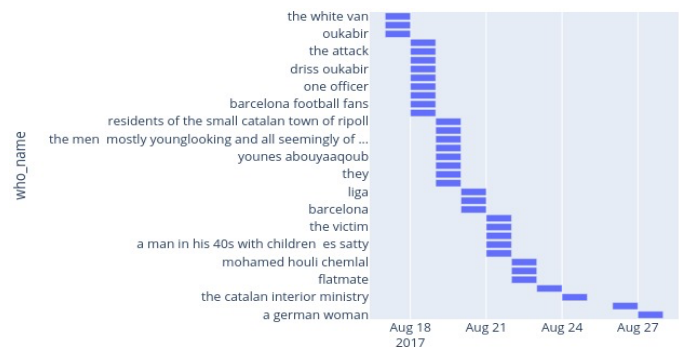


Key Findings

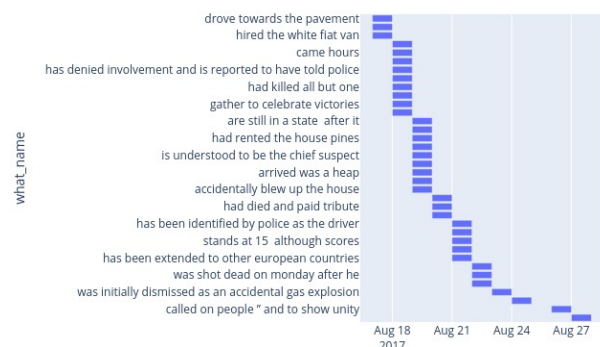
Determining and graphing changes in narrative chains was the driving force behind the design of this project, and unsurprisingly, the most interesting results came when features were analyzed at the narrative chain level. To test the methods developed by this project, the clusters in the finalized dataset were numbered and a random number generator was run to determine which cluster to test. The generator returned cluster “22” which corresponded with a narrative chain detailing the 2017 Barcelona Terror Attacks, where a number of people were killed when a van was driven onto a busy tourist street named “Las Rambla.” The initial attack then evolved into multiple related attacks across Spain. The narrative chain returned by the Nichols/Bright algorithm consisted of 36 articles and lasted around 10 days.

Gantt charts were used to visualize how key elements of the articles, the who, what, where, when, why, and how, changed over time. The graph to the right shows the subjects of the articles, the “whos,” as returned by the event extractor and how their focus in the narrative chain varied overtime. The first story initially was focused on a “the white van.” Subsequent stories then focused on people’s names (mostly of people implicated in the terror attacks) and

Time Spans of 'Whos' Identified by GiveMe5W1H



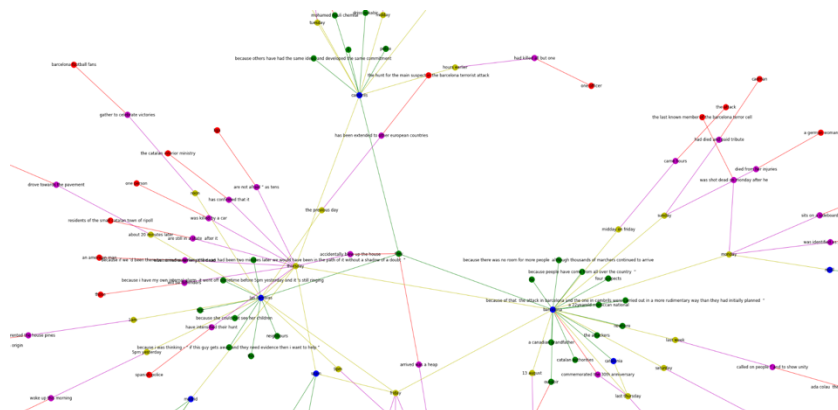
Time Spans of 'Whats' Identified by GiveMe5W1H



other groups such as “Barcelona football fans,” which correspond with a story about memorial events held during a football match soon after the attack. The chart shows that no one subject as identified by the event extractor is the focus for longer than one day. However, the extractor appears to have failed to aggregate all instances of an individual's name into one bin, “oukabir” & “driss oukabir.” The results for the time span of “whats” (above right graph) displays similar results to the “whos.” No one subject is the focus of any article for more than a day, which indicates a rapidly evolving narrative chain.

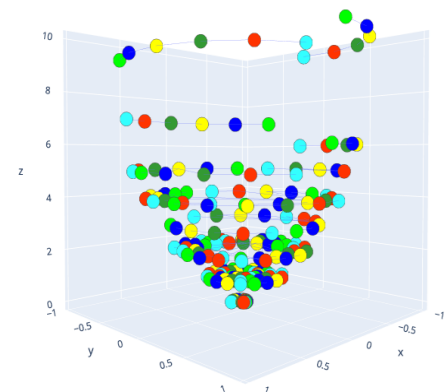
The “wheres” (graphed at right) also work to show the evolution of the narrative. Here, the locations returned by the event extractor, are not as fluid. “Las Ramblas,” the street where the initial attack took place, remains as a leading location for subsequent articles from Aug 17, 2017 through Aug 22, 2017. Additionally, “Barcelona” remained a key location for the entire narrative. “London” is briefly mentioned as a location early in the narrative, which connects to an article published early in the narrative chain that described a similar incident in London a few years prior. The locations “Cambrils,” “Alcazar,” “Ripoll”, and “Madrid” are all mentioned in subsequent articles and correlate with additional events where attacks of violence and subsequent man hunts and arrests rippled across Spain.

The GiveMe5WIH features also provided insight into the narrative change when converted into a network and graphed. When the full narrative is depicted as a network at the same time, the story lines branch off from the common locations. This corresponds with the findings for the same location, where the narrative elements were shown to be less static than the “who” or “what” elements. By investigating these different branches of the network, the different threads of the story become more apparent. This portion of the full network (below and truncated for size) shows how the different threads of the narrative chain interconnect based on location and other shared main elements. For the complete, interactive 3-D version of this plot please see pmckim1.github.io/Case-Study-I.



Converting these key elements to networks at the day-to-day level also gave insight into how narratives evolved. This visualization of the network (below right) uses the “z” axis in order to represent time and all articles released on the same day are on the same level. The initial stories are at the bottom of the plot and the final stories in the narrative chain are at the top. The funnel shape and tighter levels at the bottom of the graph shows how the story initially started with only a few articles, then quickly ramped up with a flurry of related articles. As the narrative chain ends, the number of stories and the clustering dissipates. This visualization shows how the chain first appeared, ramped up quickly, before slowing down and concluding. For the complete, interactive explorable 3-D version of this plot please see pmckim1.github.io/Case-Study-I.

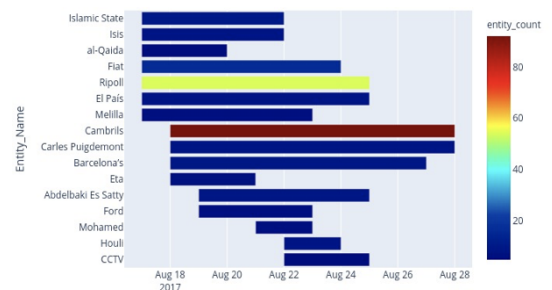
Network of Articles Day by Day



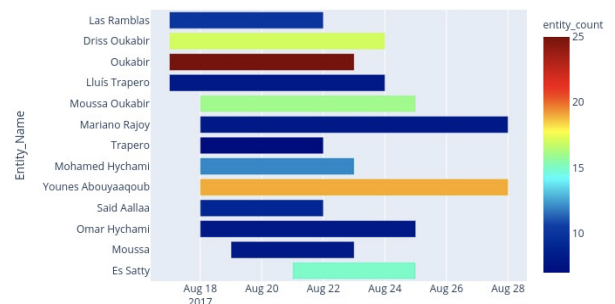
Day by Day Network 3D

The named entities mined from the dataset were also highly useful in determining the changing narrative. Similar to the methods used for the event extraction features, Gantt charts were used to show how the people, locations, and organizations changed over the course of the story. For size reasons, only the most frequently occurring entities are included in the graphs to the right. Here we see some slight mis-tagging in terms of category with “Las Ramblas” listed as a person, but the results are still helpful as they give insight into when additional people and organizations entered and exited the narrative chain and the amount of their involvement.

Named Entity Recognition Evolution of Narrative Chain for Organizations

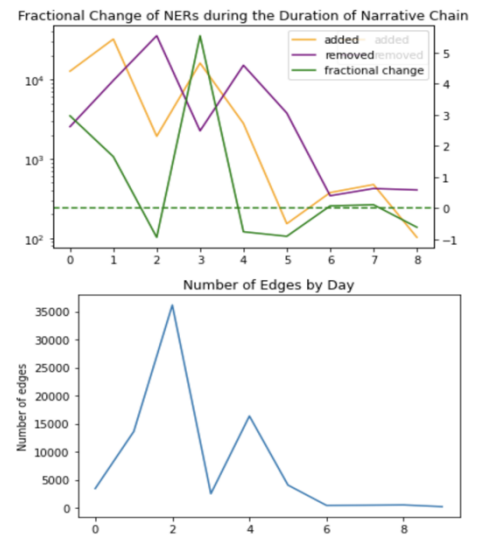


Named Entity Recognition Evolution of Narrative Chain for People

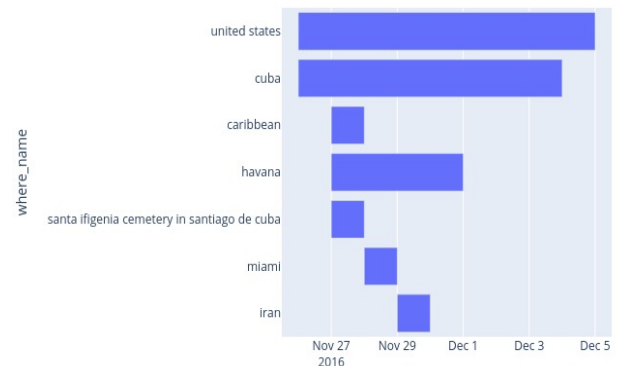


Converting the named entities information into a network at the day level also provided key insights into when and how the narratives were changing. In this portion of the analysis, all entities mentioned in articles published the same day appear in the same network together, entities that appeared in the same article were scored higher for edge connections. Then, the fractional change of edges per day was calculated for the entire narrative chain. The resulting graphs (below right) approximate the ebb and flow of the story itself. The graph spikes when the narrative rapidly evolves as new events occur and more named entities and associations are added. Then, the graph recedes when stories retread the same information and do not include as much overall narrative variation.

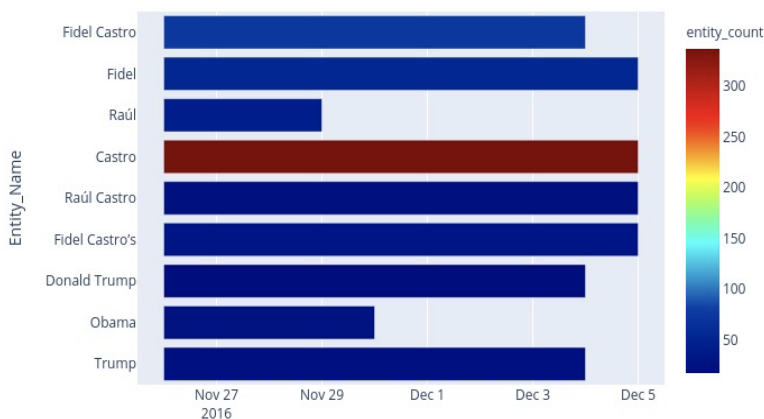
Comparing the results of the Barcelona Truck Attack narrative chain with a more static narrative chain of similar size, highlights how the methods used in this project identify differences and evolutions in the narratives. Considering the “where” chart (below right), there substantially less variation in the locations identified as main elements of the story chain. Additionally, the 3D day to day network visualization of this chain (below right) does not display the tightly clustered funnel shape seen in the Barcelona example. Instead, this narrative chain has a looser more cylindrical structure, suggesting less of a sudden ramp up of articles. The Gantt chart for the named entities (below left) also appears much more stable than the Barcelona example. Additionally, the fractional change graph for this narrative chain has one initial spike before slowly dissipating (below right subsequent page). Testing the methods across a wide range of narrative chains yielded similar results. Fast moving chains with a lot of twists and turns displayed multiple spikes in their fractional change, graphs less consistency in their Gantt plots, and less tightly connected networks. Conversely, slower moving chains displayed the opposite behaviors graphically. For additional case studies of the methods developed for this project in action, please refer to the <https://github.com/pmckim1/Capstone> and pmckim1.github.io/Case-Study-2.



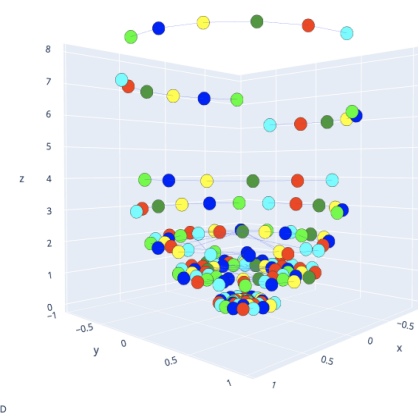
Time Spans of 'Wheres' Identified by GiveMe5W1H



Named Entity Recognition Evolution of Narrative Chain for People



Network of Articles Day by Day



Limitations and Recommendations for Future Research

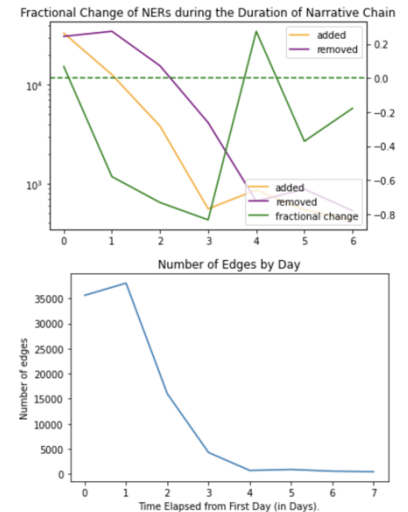
Although this project was largely successful as a proof of concept, it could certainly be improved. The unsupervised nature of the data made assessing success difficult. This project would greatly benefit from devoting time and resources to hand coding the dataset to establish a clear ground truth. Expanding the dataset to include additional news sources would give a more robust picture of the coverage surrounding different newsworthy events. In addition, adding a disparate news source could be used to assess how narratives of the same event differ based on the country or news entity providing the coverage. Further, introducing secondary sources such as social media content and other less formalized text into the project could yield interesting results and help determine how information regarding events spreads.

Additionally, the run times for fully processing the data were lengthy. Prior to putting this model into production to automate the data process, it would be beneficial to find methods to speed up the processing time. Further, building out the text standardization for the elements returned by the “Giveme5W1H” algorithm would also help to make the network analysis of narrative chains more robust.

Finally, although it was not the focus of this project and was not included in the previous sections, some exploratory analysis where the event extraction elements were converted to a network at the corpus level offered insight into how different narrative chains in different clusters fit together. Using the methods described to analyze the evolution of a person's or entity's narrative chain could be an interesting use case for the future.

Conclusion

The constantly evolving nature of news narratives can make it difficult to keep up. This project tested machine learning techniques on original data to find ways to identify narrative chains, extract key information, demonstrate when and how the focus of news narratives chains changed. The processing times for the different algorithms were lengthy and would need to be speed up considerably for this method to be used in production. Ultimately, however, the different steps used to obtain, preprocess, aggregate and analyze the data were successful and resulted in a highly visual way to look at narrative chains. Additionally, this project found that different types of narrative chains displayed different behaviors, depending on whether a narrative chain involves rapidly evolving events or is covering a more static story chain. The use of NER was crucial to the success of this project. In the future, these methods may also be applied to analyze how story chains and the name entities within those story lines interact.



Biography

Polly McKim Halloran is a graduate student in the Data Science Program at The George Washington University. Her interests include natural language processing, automated text analysis, and exploring the role of media coverage in politics and public perception. She has previously worked in legal operations and cyber security. She currently works in national security analysis. She enjoys traveling, researching history, hiking, and spending time with her terrier mix “Winston.”

Dr. Nima Zahadat is a professor of data science, information systems security, and digital forensics. His research focus is on studying the Internet of Things, data mining, information visualization, mobile security, security policy management, and memory forensics. He has been teaching since 2001 and has developed and taught over 100 topics. Dr. Zahadat has also been a consultant with the federal government agencies, the US Air Force, Navy, Marines, and the Coast Guard. He enjoys teaching, biking, reading, and writing.

References

- Bohlin, L., Edler, D., Lancichinetti, A., & Rosvall, M. (2014). Community Detection and Visualization of Networks with the Map Equation Framework. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring Scholarly Impact* (pp. 3–34). Springer International Publishing. https://doi.org/10.1007/978-3-319-10377-8_1
- Boydston, A., Hardy, A., & Walgrave, S. (2014). Two Faces of Media Attention: Media Storm Versus Non-Storm Coverage. *Political Communication*, 31. <https://doi.org/10.1080/10584609.2013.875967>
- Hamborg, F., Breiting, C., & Gipp, B. (2019). Giveme5W1H: A Universal System for Extracting Main Events from News Articles. *ArXiv:1909.02766 [Cs]*. <http://arxiv.org/abs/1909.02766>
- Hamborg, F., Meschenmoser, P., Schubotz, M., & Gipp, B. (2019). NewsDeps: Visualizing the Origin of Information in News Articles. *ArXiv:1909.10266 [Cs]*. <http://arxiv.org/abs/1909.10266>
- Herrera-Cubides, J. F., Gaona-García, P. A., & Sánchez-Alonso, S. (2020). Open-Source Intelligence Educational Resources: A Visual Perspective Analysis. *Applied Sciences*, 10(21), 7617. <https://doi.org/10.3390/app10217617>
- Hogenboom, F., Frasincar, F., Kaymak, U., & de Jong, F. (n.d.). *An Overview of Event Extraction from Text*. 10.
- Mitri, M. (2020). Story Analysis Using Natural Language Processing and Interactive Dashboards. *Journal of Computer Information Systems*, 0(0), 1–11. <https://doi.org/10.1080/08874417.2020.1774442>

- Nicholls, T., & Bright, J. (2019). Understanding News Story Chains using Information Retrieval and Network Clustering Techniques. *Communication Methods and Measures*, 13(1), 43–59. <https://doi.org/10.1080/19312458.2018.1536972>
- Ojo, A., & Heravi, B. (2018). Patterns in Award Winning Data Storytelling. *Digital Journalism*, 6(6), 693–718. <https://doi.org/10.1080/21670811.2017.1403291>
- Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4), 1118–1123. <https://doi.org/10.1073/pnas.0706851105>
- Shahaf, D., & Guestrin, C. (2010). Connecting the dots between news articles. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '10*, 623. <https://doi.org/10.1145/1835804.1835884>
- Shahaf, D., Guestrin, C., & Horvitz, E. (2013). “Metro maps of information” by Dafna Shahaf, Carlos Guestrin and Eric Horvitz, with Ching-man Au Yeung as coordinator. *ACM SIGWEB Newsletter*, Spring, 4:1-4:9. <https://doi.org/10.1145/2451836.2451840>
- Shahaf, D., Guestrin, C., & Horvitz, E. (2012). Trains of thought: Generating information maps. *Proceedings of the 21st International Conference on World Wide Web - WWW '12*, 899–908. <https://doi.org/10.1145/2187836.2187957>
- Soelistio, Y., & Surendra, M. (2013, May 15). *Simple Text Mining for Sentiment Analysis of Political Figure Using Naive Bayes Classifier Method*. <https://doi.org/10.12962/p9772338185001.a18>
- The Guardian—Open Platform*. (n.d.). Retrieved May 3, 2021, from <https://open-platform.theguardian.com/>

theguardian / open platform—Documentation / overview. (n.d.). Retrieved May 3, 2021,
from <https://open-platform.theguardian.com/documentation/>

Zhu, X., & Oates, T. (2012). Finding story chains in newswire articles. *2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)*, 93–100.
<https://doi.org/10.1109/IRI.2012.6302996>