

Is It Safe to Tell Your Story? Towards Achieving Privacy for Sensitive Narratives

Mohammad Shokri
The Graduate Center,
City University of New York
mshokri@gradcenter.cuny.edu

Allison Bishop
City College,
City University of New York
Proof Trading

Sarah Ita Levitan
Hunter College,
City University of New York

Abstract

Evolving tools for narrative analysis present an opportunity to identify common structure in stories that are socially important to tell, such as stories of survival from domestic abuse. A greater structural understanding of such stories could lead to stronger protections against de-anonymization, as well as future tools to help survivors navigate the complex trade-offs inherent in trying to tell their stories safely. In this work we explore narrative patterns within a small set of domestic violence stories, identifying many similarities. We then propose a method to assess the safety of sharing a story based on a distance feature vector.

1 Introduction

The last two decades have seen an explosion in the development of privacy-preserving techniques for data analysis. Cryptographic techniques like fully homomorphic encryption and secure multi-party computation (e.g. (Gentry, 2009; Ben-Or et al., 1988; Chaum et al., 1988)) have created a rich landscape of choices for how private quantitative data can be delegated, processed, and combined - without revealing underlying details unnecessarily. Definitional and foundational work like the development of differential privacy (Dwork et al., 2006) and its practical deployments (e.g. (Bureau et al., 2023)) has set new high standards for privacy protection, allowing us to fuel the collective benefits of data science without sacrificing individuals.

But - people are ultimately not numbers. Narratives are needed for us to feel heard and to be heard, and to restore emotional depth to experiences that have been flattened into numeric and categorical representations.

The Me Too movement has shown the power of personal narratives in bringing widespread experiences to light, in a way that pure numbers cannot accomplish. This kind of power is often exercised with tremendous risk - including risk to a survivor's

physical safety when telling their narrative publicly makes them identifiable and vulnerable to a perpetrator's wrath. Numbers here can hint at the scale of the hidden stories. The National Intimate Partner and Sexual Violence Survey report estimates that roughly 30% of US women and 10% of US men experience rape, physical violence, or stalking by an intimate partner that impacts their health, safety, or functioning at some point in their lifetimes (Black, 2011). Many of the stories behind these numbers cannot be safely told, and this silence comes at a great cost, robbing survivors of opportunities to heal and connect, and preventing non-survivors from fully absorbing the nature of the phenomenon.

For domestic violence survivors, the complex tradeoffs between the benefits and dangers of telling their stories publicly can be difficult to navigate. How much detail is "safe" to post in an online chat forum? Or what exactly needs to be removed from a personal essay to make it effectively anonymous? A crucially missing tool in this context is a scientifically rigorous notion of safety or anonymity for narrative forms, something that would arm survivors with a firm criterion for making such decisions according to their own goals and values.

Developing such a tool is a challenging task that requires first identifying the kinds of evolving threats that face survivors who want to speak safely. So far, progress in AI seems poised to exacerbate such threats. Current LLMs can infer a large range of personal attributes from users' text material posted online, which could easily violate their privacy (Staab et al., 2023). Neural network based methods have also been shown to be effective for author attribution and author verification based on mere writing style (Rivera-Soto et al., 2021; Fabien et al., 2020).

Such threats are multi-faceted: several different layers of textual analyses could be used individ-

ually or in combination by an adversary to effectively de-anonymize a narrative that is intended to be anonymous. From a cognitive psychology perspective, Bal (2009) proposes a three layered definition for a narrative: the *fabula*, which she defines as a "series of logically and chronologically related events that are caused or experienced by actors"; the *story*, where the narrator selects specific fabula elements to convey; and the *text*, where chosen words express the story in discourse. Using this framework, author attribution attacks based on writing style (Abbasi and Chen, 2008), could be mapped to the text layer of the narrative. . In this work, we focus on the story layer and look for unusual elements of a story that could be dispositive. We might also wonder if a sequence of story elements, even when common individually, might be unique and identifying in combination.

This sets up a familiar tension between privacy and utility: can we smooth any identifiable edges to a survivor narrative without blunting its emotional force? Can we preserve the potential cathartic and connective effects of telling such stories while providing a satisfying level of privacy for survivors?

We do not try to arrive at the answer to this complex and ambitious question in the following few pages. What we intend instead is to begin exploring narrative structure of survivor stories. We view this as a humble first step towards building representations of such narratives that could be useful components in developing new privacy-preserving processing techniques.

We note there are good reasons to believe that such techniques are possible. Afterall, narrative fiction can be viewed as a strong existing technique to preserve privacy while retaining emotional impact, albeit one that still doesn't come with rigorous privacy guarantees and requires high skill from authors. Fiction writers well understand the complex relationship between unique detail and universal themes. James Joyce articulated it best in saying: "In the particular is contained the universal."

There is a tragic but productive irony in this when considered in the context of survivor narratives. The horrifying universality - the sickening commonality - of domestic abuse is exactly why is it crucial to tell firsthand stories about it at scale. And also exactly why doing so safely should be possible. Afterall, the commonality of such stories is the core of what we might want to express, and if such a core is widely shared, it is not inherently identifying. But we need to find an effective path to

that universality that eschews the particular dangers of the particular.

Building on this understanding of complex interplay between the particular and the universal, our research aims to focus on the following two research questions:

- **RQ1:** What common events or patterns can be identified in domestic violence stories?
- **RQ2:** How might the existing patterns contribute to developing strategies to safeguard privacy in narrative analysis while preserving the human impact?

To begin to answer these research questions, we explore the patterns of events in domestic violence narratives to identify commonalities that could inform the development of effective privacy measures. We use a distance vector feature to distinguish domestic violence from non-domestic violence stories and demonstrate that underlying event patterns can be systematically analyzed and leveraged to develop privacy-preserving mechanisms. Our findings suggest that such patterns offer a promising avenue for advancing privacy guarantees in narrative-sharing contexts. We hope this work will inspire further research into the intersection automated narrative understanding and privacy.

2 Related Work

Research on narrative understanding has frequently modeled narratives as sequences of events to capture their structure and progression. This approach has been employed to analyze, generate, and comprehend narratives across various domains (Chambers and Jurafsky, 2008; Goyal et al., 2013; Pichotta and Mooney, 2014; Nguyen et al., 2015; Peng and Roth, 2016; Chaturvedi et al., 2017). Finding similarities between narratives is a challenging tasks that humans also tend to differ in aspects they pay attention for judging the similarity (Nguyen et al., 2014; Fisseni and Löwe, 2012). Computational studies have utilized different features including plot structure (Saldias and Roy, 2020; Chaturvedi et al., 2018), character resemblance (Lee and Jung, 2018; Lee et al., 2018), sentiment progression (Antoniak et al., 2019; Somasundaran et al., 2020), and lexical similarities (Lin et al., 2013; Chaturvedi et al., 2018) to capture narrative similarity.

Domestic violence stories have not received significant attention from NLP researchers. However,

Schrading et al. (2015) examined the language used in domestic violence narratives on Reddit. By comparing these with other emotionally charged stories, such as those expressing *anger* and *anxiety*, they identified distinct linguistic patterns unique to domestic violence stories. In this work, we focus on the events that occur between two main characters in a domestic violence story, namely: the victim and the perpetrator. Inspired by studies that capture plot similarities and lexical similarities, we develop a method to identify narratives that share similar event patterns across many stories.

3 Data

Reddit is a valuable source for finding domestic violence stories due to its large and diverse user base, which provides a wide range of personal experiences and perspectives. The platform allows users to share their stories anonymously, encouraging openness and honesty, which can be crucial for gathering authentic and detailed data. We collected top rated stories posted on the subreddit *r/domesticviolence* between January 1, 2015, and March 31, 2024, focusing on posts exceeding 500 words, as longer narratives are more likely to provide comprehensive accounts and deeper insights into individuals’ experiences. Our initial dataset comprised 220 posts. Upon review, we identified that many of these posts did not necessarily describe the authors’ personal experiences but rather offered general opinions or rants (as used frequently on the platform) about domestic violence. To ensure the quality of the data, we manually assessed the posts, retaining only those that provided personal experiences. This filtering process resulted in a final dataset of 145 stories that include descriptions of domestic violence.

Additionally, we scrape another subreddit to select stories for negative samples in our experiments. We sample 145 stories from the subreddit *r/realstories* where users share their personal stories of hardship, joy, tragedy, etc. To ensure comparability, we only select the stories that have more than 500 and less than 1000 words. Finally, we only keep stories that contain more than one characters. (Mostafazadeh et al., 2016) . Table 1 summarizes the statistics of the datasets we use in our study.

Stats	r/domesticviolence	r/realstories
Number of stories	145	145
Avg sentence length (tokens)	18.5	21.1
Avg sentence count	38.7	36.0
Avg descriptive sentence count	8.7	11.8
Avg event verb count	6.6	9.9

Table 1: Summary Statistics of our domestic violence dataset and the ROCStories dataset

4 Event Extraction

To address RQ1, we focus on identifying recurring events in domestic violence narratives. An event is defined as a specific occurrence involving participants, often characterized as a change of state (Doddington et al., 2004). We hypothesize that domestic violence stories share a common set of events typically occurring between the victim and the perpetrator, with a particular focus on events where the perpetrator acts upon the victim. Using the existing definition of an event from previous studies—as a triplet of subject, predicate, and object (Mousavi et al., 2023)—we aim to extract events in which the perpetrator is the subject and the victim is the object. We employ a method similar to that of (Chaturvedi et al., 2017) to extract events from the narratives. We used the Stanza pipeline (Qi et al., 2020) to process the stories and obtain part-of-speech tags, dependency parses, and co-referent mentions. After obtaining the dependency parse, we identify verbs and their agents and their patients and only extract verbs that have the perpetrator as their agent and the victim as their patient. Based on our preliminary data analysis, we heuristically assume that the narrator is the victim. Finally, in this way, we reduced every story to a set of triples like the following:

He threw me out of the car
 → <He-throw-me>

My ex pushed me to the wall
 → <My ex-push-me>

As a preliminary analysis, we examined the most common verbs used in these events to identify prevalent actions. Figure 1 depicts the results. The prevalence of violent verbs indicates the potential for finding similar patterns within these stories.

To extract events for the stories sampled from *r/realstories*, we first obtain part-of-speech tags and dependency parses. Then we perform named entity recognition. After resolving co-references,

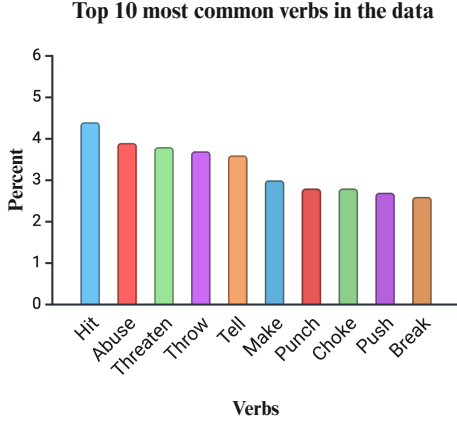


Figure 1: Top 10 most common action verbs lemmatized in the dataset.

we only keep verbs which have a named entity of type *Person* as both their agent and patient.

5 Narrative Representation as Distance Vectors

We operate under the assumption that domestic violence stories contain similar events perpetrated by the abuser against the victim. Previous analytical study on domestic abuse stories have demonstrated that these stories contain distinct language and semantic role labels (Schradling et al., 2015). Rather than representing narratives through prototypical sequences of events, participants, and their causal or temporal relationships, we adopt a more straightforward approach by representing each story as an unordered list of action verbs. We hypothesize that many domestic violence victims undergo similar experiences, leading to narratives with comparable actions. To capture this similarity, we compute the Word Mover’s Distance (WMD) (Kusner et al., 2015) between each story and multiple sets of common events identified in the training set. This process yields a multi-dimensional distance vector that quantifies the degree of similarity across different dimensions of event types, providing a more nuanced representation of the narrative structure. Shortly, WMD calculates the minimum traveling distance between documents in the embedding space by using a *flow matrix* \mathbf{T} to show how much mass of word i in document d should travel to word j in document d' . Let $c(i, j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2$ be the distance between two words x_i and x_j in the embedding space. The solution to the minimum transportation problem is provided by the follow-

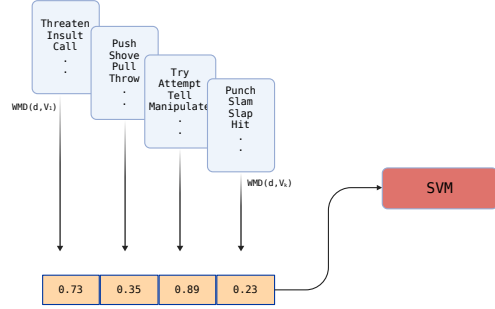


Figure 2: Word Mover’s Distance calculated between the document and each verb cluster to form the feature vector. The Feature vector is then passed to a Logistic Regression model for binary classification.

ing linear program with the constraints that all the mass from d is extracted and pushed to d' :

$$\begin{aligned} \min_{\mathbf{T} > 0} & \sum_{i=1}^{|d|} \sum_{j=1}^{|d'|} \mathbf{T}_{ij} c(i, j) \\ s.t. & \mathbf{T} \mathbf{1} = |d| \\ & \mathbf{T}^\top \mathbf{1} = |d'| \end{aligned}$$

We extract all events and their associated verbs from the training dataset to form the set $V = \{verb_1, verb_2, \dots, verb_n\}$. To capture thematic similarities, we cluster V into K distinct groups of verbs, resulting in: V_1, V_2, \dots, V_k . For each story S , we calculate its WMD to each of these clusters, generating a K -dimensional vector where each dimension represents the distance between the story and its corresponding verb cluster. This vector serves as the feature representation for the narrative, which we then input into a support vector machine (SVM) model to perform binary classification (see Figure 2). The binary classification method directly addresses research question 2 by using the output to assess the safety of publishing a new story. By quantifying the similarities between a new story and existing narratives through word mover’s distance, the model can determine whether the story’s events align with known patterns. If a story is classified as being close to the training data - indicating it shares similarities with many other stories - it suggests that the narrative structure is less likely to reveal sensitive, unique details. This is not a complete evaluation of course, as other features of style or narrative could still be revealing. However, it is one necessary component.

6 Results

In this section, we first assess the performance of our proposed classification method. The task is a binary classification, distinguishing between domestic violence and non-domestic violence stories using distance feature vectors. Our hypothesis is that the distance between events in a story and various event clusters captures the underlying similarities among these stories, making it a valuable metric for privacy. The intuition is that a story achieves anonymity when its events become indistinguishable from those in a large collection of similar stories.

It is noteworthy that the proposed method can serve as an initial step toward ensuring privacy in narratives. If a story is classified as non-domestic violence, it suggests that its events deviate significantly from the typical patterns found in domestic violence stories. This indicates a high degree of uniqueness in the story, which could potentially be exploited for de-anonymization

For our experiments, we use a balanced test set comprising 29 domestic violence stories and 29 common-sense short stories from the ROCStories dataset. The training set includes 116 domestic violence stories and 116 common-sense short stories. To obtain embeddings for each action verb, we pass them through a BERT model (Devlin, 2018) and capture the hidden representation from the [CLS] token of the final layer. These BERT embeddings are used for both verb clustering and Word Mover’s Distance calculations. We use the K-means algorithm to cluster the verbs within the embedding space.

Next, as the the robustness of an ML model is crucial when applied to data with significant linguistic variability, we evaluate the robustness of our model against variations in word choice for actions between the victim and the perpetrator. To do this, we replace the verbs with their synonyms and test our model on the perturbed set. The purpose of this replacement is to ensure that the model can recognize actions that are semantically similar to the original ones, demonstrating that it is not overly sensitive to minor changes in word choice.

6.1 Evaluation of Our Method

Our proposed method calculates the Word Mover’s Distance between each document and various verb clusters in the training set. To determine the optimal number of clusters, we experiment with cluster

counts ranging from 1 to 20. Additionally, we evaluate the performance of our SVM model with a linear kernel using four different values for the regularization parameter C from the set $\{0.1, 1, 10, 100\}$. Lower C values promote a larger margin, which may improve generalization but could also result in a higher rate of misclassification on the training data. Our results, as illustrated in Figure 3, show that performance improves with an increasing number of clusters. This is likely because a higher number of clusters leads to more fine-grained and meaningful verb groupings, which makes the distance calculations more informative.

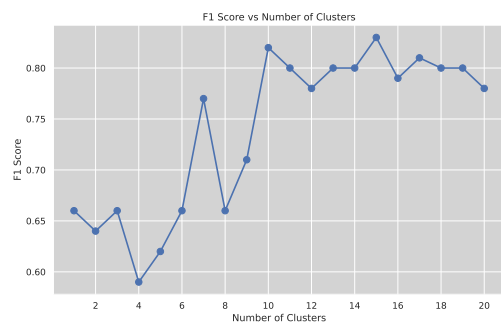


Figure 3: classification f1 score for different numbers of verb clusters. We report the highest achieved for each cluster count with different regularization values in our SVM model.

6.2 Robustness

To assess the robustness of our approach, we replace each action verb with its most prevalent synonym from WordNet (Miller, 1995). This method represents a rigorous adversarial attack due to the limitations in quality of synonym replacements; for instance, the most prevalent synonym for ‘hit’ is ‘reach’, and for ‘punch’ is ‘plug’ which are not always contextually appropriate. The rationale is that synonyms, being closely related in the embedding space, should not significantly alter their distances from verb clusters. We apply this synonym transformation to all data points in the test dataset and evaluate the trained model’s performance on the perturbed test set. Using an SVM model with a linear kernel and a regularization parameter C of 0.1, our results, illustrated in Figure 4, show that as the number of clusters increases, the performance gap between the original and perturbed test sets diminishes. This indicates that our approach becomes more robust with a higher number of verb clusters.

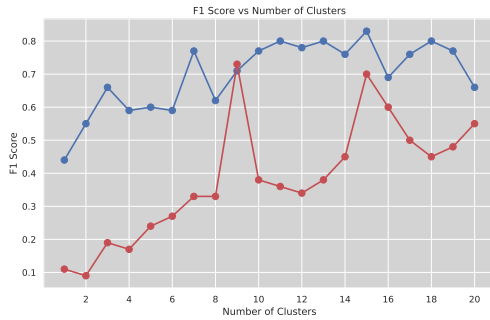


Figure 4: Robustness results for our model tested on the perturbed test set. Performance on perturbed set increases with higher number of clusters. The instability of the red line is due to clustering quality as the method is dependent on obtaining meaningful clusters.

7 Conclusion

In this work, we proposed to use the distance between a new story and a set of available stories as a feature to decide whether it is safe to tell the story. We have seen there is some common structure within stories of domestic violence, enough to separate them from other kinds of stories. This is a humble first step toward a larger understanding of commonality that could help us to define a systematic balance between preserving narrative meaning and protecting individual privacy. Future work could focus on collecting a larger dataset of this type, exploring privacy-preserving methods for the sequential representation of stories, and establishing a formal definition for privacy in the narrative setting.

Acknowledgment

We would like to express our sincere gratitude to the anonymous reviewers for their valuable feedback and insightful suggestions, which have improved the quality of this paper and provided several ideas for future work.

Limitation

One limitation of this study is the relatively small dataset size, with only 145 domestic violence stories collected from Reddit. The limited number of stories restricts the ability to generalize findings across a broader range of narratives and potentially affects the robustness of the classification model. Additionally, the nature of the problem makes it challenging to acquire more data, as stories about domestic violence are often sensitive and not fre-

quently shared in public forums. Another limitation of this study is the selective focus on capturing those verbs as events where the victim and perpetrator are identified as agents/patients. This approach might overlook other significant events and context that do not explicitly involve both the victim and perpetrator in these grammatical roles. Representing stories as a bag of words is another significant limitation of this study. This method ignores the sequential order of events and the narrative structure, which can be crucial for understanding the context and progression of domestic violence stories.

Ethical Considerations

The motivation for this work was developed in collaboration with an anonymous survivor of domestic abuse. Only publicly available data was used in our analyses. Furthermore, as our analyses here are intended to uncover *common* structure among survivor stories and not potentially identifying details, we report only aggregate results. This eschews the kind of granular details that could be a threat to privacy, as damaging privacy of anonymous contributors to a public data set would be antithetical to our goals.

References

- Ahmed Abbasi and Hsinchun Chen. 2008. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems (TOIS)*, 26(2):1–29.
- Maria Antoniak, David Mimno, and Karen Levy. 2019. Narrative paths and negotiation of power in birth stories. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–27.
- Mieke Bal. 2009. *Narratology: Introduction to the theory of narrative*. University of Toronto Press.
- Michael Ben-Or, Shafi Goldwasser, and Avi Wigderson. 1988. [Completeness theorems for non-cryptographic fault-tolerant distributed computation \(extended abstract\)](#). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 1–10. ACM.
- Basile K.C. Breiding M.J. Smith S.G. Walters M.L. Merrick M.T. Chen J. Stevens M.R. Black, M.C. 2011. The national intimate partner and sexual violence survey (nisvs): 2010 summary report. atlanta, ga: National center for injury prevention and control, centers for disease control and prevention.
- The Population Reference Bureau, the U.S. Census Bureau’s 2020 Census Data Products, and Dissemination Team. 2023. [Why the census bureau chose differential privacy](#).

- Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of ACL-08: HLT*, pages 789–797.
- Snigdha Chaturvedi, Mohit Iyyer, and Hal Daume III. 2017. Unsupervised learning of evolving relationships between literary characters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.
- Snigdha Chaturvedi, Shashank Srivastava, and Dan Roth. 2018. Where have i heard this story before? identifying narrative similarity in movie remakes. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 673–678.
- David Chaum, Claude Crépeau, and Ivan Damgård. 1988. [Multiparty unconditionally secure protocols \(extended abstract\)](#). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing, May 2-4, 1988, Chicago, Illinois, USA*, pages 11–19. ACM.
- Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ace) program-tasks, data, and evaluation. In *Lrec*, volume 2, pages 837–840. Lisbon.
- Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. 2006. [Calibrating noise to sensitivity in private data analysis](#). In *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer.
- Maël Fabien, Esaú Villatoro-Tello, Petr Motliceck, and Shantipriya Parida. 2020. Bertaa: Bert fine-tuning for authorship attribution. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON)*, pages 127–137.
- Bernhard Fisseni and Benedikt Löwe. 2012. Which dimensions of narrative are relevant for human judgments of story equivalence?
- Craig Gentry. 2009. [Fully homomorphic encryption using ideal lattices](#). In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 169–178. ACM.
- Amit Goyal, Ellen Riloff, and Hal Daumé Iii. 2013. A computational model for plot units. *Computational Intelligence*, 29(3):466–488.
- Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. 2015. From word embeddings to document distances. In *International conference on machine learning*, pages 957–966. PMLR.
- O-Joun Lee, Nayoung Jo, and Jason J Jung. 2018. Measuring character-based story similarity by analyzing movie scripts. In *Text2Story@ ECIR*, pages 41–45.
- O-Joun Lee and Jason J Jung. 2018. Explainable movie recommendation systems by using story-based similarity. In *Iui workshops*.
- Yung-Shen Lin, Jung-Yi Jiang, and Shie-Jue Lee. 2013. A similarity measure for text classification and clustering. *IEEE transactions on knowledge and data engineering*, 26(7):1575–1590.
- George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James Allen. 2016. A corpus and cloze evaluation for deeper understanding of commonsense stories. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 839–849.
- Seyed Mahed Mousavi, Shohei Tanaka, Gabriel Rocabruna, Koichiro Yoshino, Satoshi Nakamura, and Giuseppe Riccardi. 2023. Whats new? identifying the unfolding of new events in narratives. *arXiv preprint arXiv:2302.07748*.
- Dong Nguyen, Dolf Trieschnigg, and Mariët Theune. 2014. Using crowdsourcing to investigate perception of narrative similarity. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 321–330.
- Kiem-Hieu Nguyen, Xavier Tannier, Olivier Ferret, and Romaric Besançon. 2015. Generative event schema induction with entity disambiguation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 188–197.
- Haoruo Peng and Dan Roth. 2016. Two discourse driven language models for semantics. *arXiv preprint arXiv:1606.05679*.
- Karl Pichotta and Raymond Mooney. 2014. Statistical script learning with multi-argument events. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 220–229.
- Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.
- Rafael A Rivera-Soto, Olivia Elizabeth Miano, Juanita Ordonez, Barry Y Chen, Aleem Khan, Marcus Bishop, and Nicholas Andrews. 2021. Learning universal authorship representations. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 913–919.

- Belen Saldias and Deb Roy. 2020. Exploring aspects of similarity between spoken personal narratives by disentangling them into narrative clause types. *arXiv preprint arXiv:2005.12762*.
- Nicolas Schrading, Cecilia Ovesdotter Alm, Raymond Ptucha, and Christopher Homan. 2015. An analysis of domestic abuse discourse on reddit. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2577–2583.
- Swapna Somasundaran, Xianyang Chen, and Michael Flor. 2020. Emotion arcs of student narratives. In *Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events*, pages 97–107.
- Robin Staab, Mark Vero, Mislav Balunović, and Martin Vechev. 2023. Beyond memorization: Violating privacy via inference with large language models. *arXiv preprint arXiv:2310.07298*.