A City of Millions: Mapping Literary Social Networks At Scale

Sil Hamilton[1*], **Rebecca M. M. Hicke**[2*], **David Mimno**[1], **Matthew Wilkens**[1]
[1]Department of Information Science
[2]Department of Computer Science
Cornell University
{srh255,rmh327,mimno,wilkens}@cornell.edu
*Equal contribution

Abstract

We release 70,509 high-quality social networks extracted from multilingual fiction and nonfiction narratives. We additionally provide metadata for ∼30,000 of these texts (73% nonfiction and 27% fiction) written between 1800 and 1999 in 58 languages. This dataset provides information on historical social worlds at an unprecedented scale, including data for 2,510,021 individuals in 2,805,482 pair-wise relationships annotated for affinity and relationship type. We achieve this scale by automating previously manual methods of extracting social networks; specifically, we adapt an existing annotation task as a language model prompt, ensuring consistency at scale with the use of structured output. This dataset serves as a unique resource for humanities and social science research by providing data on cognitive models of social realities.
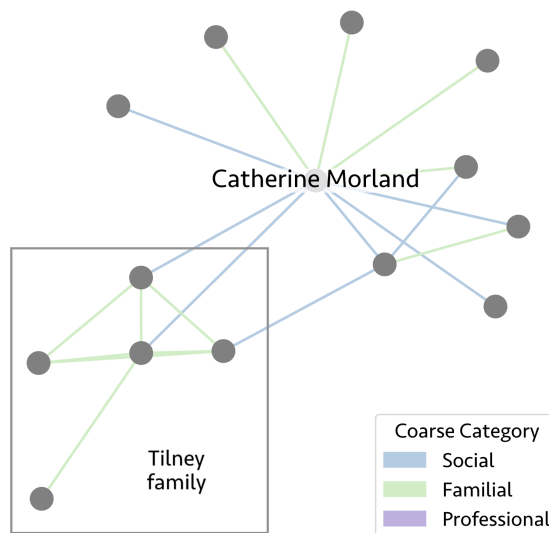
Figure 1: The graph of relationships in *Northanger Abbey* by Jane Austen created by our model. Note the presence of the Tilney family at the bottom left of the figure.

1 Introduction

Literary scholars have long been interested in the social worlds of novels. Novels depict social configurations across time and space at varying levels of abstraction, from the grand descriptions of geopolitical intrigue in *War and Peace* to the personal relationships underpinning *In Search of Lost Time*. While social networks cannot represent the full detail and nuance of literary works, they provide a uniform format to identify large-scale patterns. However, prior attempts at extracting social networks from literary texts have been hindered by a dependency on supervised machine learning models limited in accuracy and scalability.

In this work, we present a dataset of high quality social networks extracted from 70,509 literary texts, such as that shown in Figure 1. We extract the networks, including affinities and relationship types, using a novel method that passes a modified prompt from Massey et al. (2015) to Gemini 1.5 Flash configured to output JSON. We validate this

approach by demonstrating that it produces annotations similar to the manual annotations provided by Massey et al. (2015). In addition to networks, we also provide extended metadata for a subset of ∼30,000 works, of which 22,015 are nonfiction and 7,331 fiction.

This dataset will provide researchers with the opportunity to evaluate literary and social hypotheses at scale. As an initial example, we show that nonfiction networks consist of more communities and are less clustered than fiction networks. This may help explain why characters in non-fiction texts travel more (Wilkens et al., 2024) as, intuitively, a text consisting of fewer social communities may feature fewer locations.[1]

2 Related Work

Literary social network extraction. Significant previous research has addressed extracting

---

[1]We provide the full dataset here.

social networks from literary texts. One traditional approach involves creating networks by hand (Moretti, 2011; Smeets et al., 2021; Sugishita and Masuda, 2023), but manual annotations are time intensive and do not scale to large datasets. Alternative approaches look for character co-occurrences in windowed units like sentences or chapters (Way, 2018; Evalyn et al., 2018; Fischer and Skorinkin, 2021). Identifying co-occurrences is computationally lightweight, but their dependency on surface-level features limits their accuracy and applicability. Neural networks have also been used more widely in recent years for this task (Nijila and Kala, 2018; Kim and Klinger, 2019; Chen et al., 2020; Mellace et al., 2020). Specifically, Piper et al. (2024) and Zhao et al. (2024) both use generative models to extract literary social networks, but their approaches are semi-supervised and thus not easily scaled, limiting their studies to datasets in the low hundreds of volumes.

Literary social networks in use. Literary social networks are often used to study particular character or character-relationship traits such as prominence (Masías et al., 2017; Sudhahar and Cristianini, 2013), cooperativeness (Chaturvedi et al., 2016), relationship trajectory (Chaturvedi et al., 2017; Mellace et al., 2020), and relationship valence (Nijila and Kala, 2018; Kim and Klinger, 2019; Piper et al., 2024). Some studies also use social networks to ground characters in particular locations (Lee and Lee, 2017; Lee and Yeung, 2012). Social networks are likewise useful for studying aspects of plot, including conflict (Smeets et al., 2021), narrative trajectory (Min and Park, 2016; Moretti, 2011), textual genre (Agarwal et al., 2021; Evalyn et al., 2018), and text veracity (Sugishita and Masuda, 2023; Volker and Smeets, 2020). They also provide data for studies comparing differences within a corpus (Fischer and Skorinkin, 2021), over time (Algee-Hewitt, 2017), and between different social theories (Elson et al., 2010; Falk, 2016; Bonato et al., 2016; Stiller and Hudson, 2005; Stiller et al., 2003). However, these studies make use of relatively small corpora, limiting the statistical significance of their results.

3 Methods

Data. We draw volumes from the Project Gutenberg (PG) corpus (Hart, 1971). PG is an online collection of public domain literary volumes developed by volunteers. It currently contains over 75,000 works, and continues to grow. The size and historical breadth of the corpus makes it popular with researchers working in literary analysis (Brooke et al., 2015; Reagan et al., 2016; Piper, 2022) and corpus linguistics (Gerlach and Font-Clos, 2020).

To create our dataset, we first download the full corpus from PG, resulting in 72,875 volumes totaling 25GB of raw text.[2] We then supplement the limited metadata provided by PG (author and title) by aligning texts with MultiHATHI, an extended multilingual edition of the HathiTrust Digital Library catalog (Hamilton and Piper, 2023), containing metadata such as publication date, language, and fiction/nonfiction status. We use title and author edit distance (Levenshtein, 1965) to find the closest match for each PG text in MultiHATHI, only considering matches where the MultiHATHI title and author matches both exceed 80% similarity. This process yields 33,919 well-documented texts.

Model selection. We consider two qualities when selecting a suitable large language model (LLM) for generating social networks from arbitrary-length texts. The first is maximum context length. The longest work in our full corpus is 13,551,565 tokens (4,233,776 words) when tokenized with the SentencePiece-based Gemma 2 tokenizer (Kudo and Richardson, 2018; Team Gemma et al., 2024). For comparison, the mean word count of all volumes in our full corpus is 63,656 words ($P_{95} = 170{,}601$). Prompts of this magnitude can quickly exhaust the capacity of recent "open weight" LLMs, which most commonly offer context windows equal to or less than 128,000 tokens, despite the growing popularity of positional embedding modifications like RoPE and YaRN (Peng et al., 2023; Su et al., 2023; Jiang et al., 2023; Dubey et al., 2024).

Our second consideration is support for structured output. When we generate output for ~70,000 documents from a stochastic language model, there is no default guarantee that the output will be consistent. Recent methods for guaranteeing consistent output include grammar-constrained decoding, where tokens are selectively masked at sampling time according to some context-free grammar (Gerganov, 2024; Microsoft, 2024; Rickard, 2024; Beurer-Kellner et al., 2024). A competing method is structured output, where the model emits JSON according to a JSON Schema

---

[2]Our copy was obtained on September 29, 2024.

passed at inference time (Shorten et al., 2024). Along with larger context windows, proprietary models have made structured output a common feature. From the pool of presently-available proprietary LLMs satisfying both these conditions, we select Google's `Gemini 1.5 Flash`, which features a context window of 1×10^6 tokens and supports structured output via JSON Schema (Team Gemini et al., 2024).

Pipeline. To create an appropriate prompt for extracting social networks from texts, we turn to a public dataset released by Massey et al. (2015). This dataset contains 2,170 annotated character relationships produced from 109 fictional narratives. Each character pair is labeled with three attributes: the valence of the relationship (positive, negative, or neutral) and two descriptors ("coarse-grained" and "fine-grained," with 3 and 30 possible labels respectively) further clarifying the relationship in terms of social function and connection (e.g., whether the characters are lovers). Massey et al. (2015) release the annotation prompt they used for Mechanical Turk workers alongside the dataset. We adapt their prompt for `Gemini 1.5 Flash` in JSON Schema, effectively requiring the model to return a JSON array of characters and their relationships for each text.[3]

4 Results

We use this pipeline to process the entire Project Gutenberg Corpus. It returns 71,836 networks from a total 72,875 volumes after omitting 1,039 volumes that fail to pass the Gemini API safe content filter.[4] Removing duplicate networks and malformed relationships (correcting attribute labels where possible) reduces this to 70,509 total networks, of which 29,346 (22,015 nonfiction and 7,331 fiction) have HathiTrust metadata available.

Validation. We assess the validity of our approach by comparing our Gemini-based pipeline against the human-annotated results reported in Massey et al. (2015). For each network in Massey et al. (2015), we retrieve the original text and pass it through Gemini together with our prompt. We additionally instruct the model to only return an-

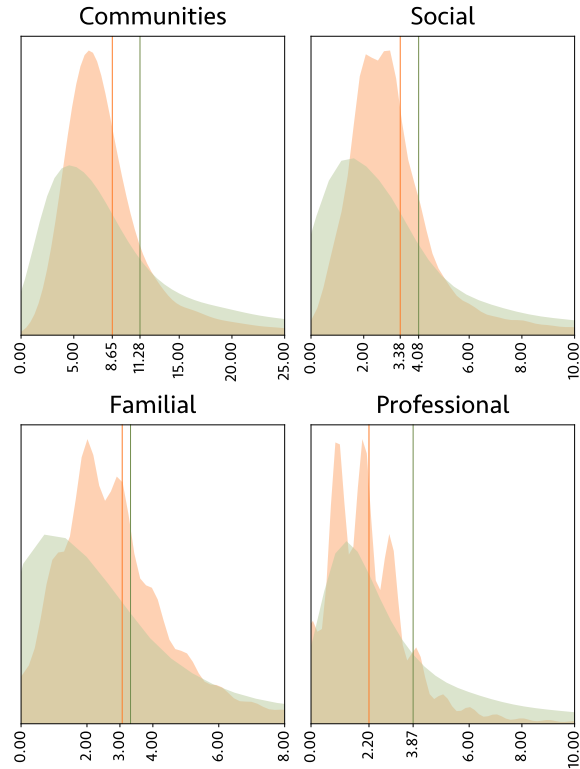
Figure 2: Density of community distributions for nonfiction (in green) and fiction texts (in orange). Vertical lines represent distribution means and graphs are truncated to show 90% of the data. In all cases, fiction texts feature smaller mean community counts than nonfiction texts.

notations for the character pairs pre-identified by Massey et al. (2015). We then calculate the ratio of true positive annotations over all annotations on a per-attribute basis to assess accuracy.

Identifying character networks and attributes is hard: Massey et al. (2015) report inter-annotator agreement rates of $\kappa = 0.812, 0.744$ and 0.364 for these tasks. Our pipeline achieves a promising 81% accuracy for valence and 74% for "fine category." However, Gemini does noticeably worse for "coarse category" annotations (55%) despite the fact that each fine label is unique to a single coarse label (e.g., the fine label "husband/wife" implies the coarse label "familial"). We therefore make use of the coarse label annotations corresponding to the Gemini provided fine category labels in place of the originally produced values.

5 Network Properties

Previous research has suggested that fictional worlds are smaller than nonfictional worlds. For example, Wilkens et al. (2024) showed that fictional

---

[3]We provide an example prompt together with a list of rejected volumes here.

[4]Two works that fail to pass the safe content filter are Abraham Lincoln's *Gettysburg Address* and an illustrated copy of Edgar Allen Poe's poetry. We do not investigate further the reasons for rejection in our corpus.

protagonists travel smaller distances, follow more routine paths, and more frequently spend time in domestic or private spaces than their nonfictional counterparts. With our new access to large-scale social network information, we can test whether similar distinctions hold between community structures. In the same way that fictional characters travel less than do people in non-fiction, they may also participate in more tightly-knit social networks.

Network characteristics. We test the validity of this hypothesis by assessing the network characteristics of the nonfiction and fiction volumes with metadata available in our dataset. We find that non-fiction networks are on average significantly larger than fiction networks by both number of nodes (22.14 v. 42.69) and number of edges (27.42 v. 42.91).[5] Two other metrics of network complexity include the number of disconnected components (groups of characters that do not interact) and transitivity (the probability that two nodes that share a mutual connection are themselves connected). Fictional networks have significantly fewer disconnected components (2.20 vs. 5.12) and their mean transitivity is significantly larger (0.22 vs. 0.12) than nonfiction networks. Thus, we see that fiction networks are smaller and more clustered than nonfiction networks.

Community detection. While completely disconnected sub-graphs are easy to identify, there are also often denser *communities* embedded in larger graphs. Since our networks can contain multiple edges between two nodes representing different relationships, we first divide the full network into three networks (familial, social, professional). We then use the Louvain method (Blondel et al., 2008) to partition each graph into communities, with no pre-set number of communities per graph.

The mean number of communities in non-fiction networks is indeed larger than in fiction networks (11.28 vs. 8.65).[6] This is true even if we look only at social (3.38 vs. 4.08), familial (3.07 vs. 3.32), or professional (2.20 vs. 3.87) communities. Figure 2 shows the distribution of community counts for fiction and non-fiction. Fictional networks tend to have a more consistent number of communities,

while non-fiction networks have a wider range; if a work has very few or very many communities, it is more likely to be non-fiction.

Relationship types. We observe that fiction networks consist of a significantly larger proportion of social (48.08% v. 39.44%) and familial (30.07% v. 21.15%) relationships on average, whereas non-fiction networks have a larger average proportion of professional relationships (39.41% v. 21.86%). This aligns with Wilkens et al. (2024)'s finding that fictional characters spend more time in domestic and private spaces.

6 Conclusion

This work presents a novel dataset containing 70,509 high-quality social networks extracted from fiction and nonfiction narratives. It additionally includes metadata for 29,346 texts written between 1800 and 1999 in 58 languages. We release this resource to support researchers in the humanities and social sciences studying the development of social worlds over time, and the work of behavioral scientists who seek to understand how cognitive models of social communities compare with real-world social communities.

Our dataset-construction process also contributes to a growing literature on adapting annotation task instructions for language models. Our results demonstrate that we can use LLMs to generate large-scale datasets for complicated and nuanced annotations on volume-length data. We find that constrained output such as JSON Schema is critical to maintaining consistency and compatibility at scale. We also observe that more concrete, descriptive annotations are more successful than more abstract annotations, even when these appear logically identical to humans.

Next steps. While our dataset is a step forward for researchers studying social networks, there remains room for progress. Generative language models improve over older social network extraction methods based on surface-level features, but we need better open and locally runnable alternatives to inefficient and costly proprietary models. These models have low interpretability, do not permit unlimited token lengths, and block content considered inappropriate for opaque reasons that may be inappropriate for historical data. We similarly lack good evaluation methods. Modeling social networks as graphs is an inherently *interpretive*

---

[5]All reported distinctions are significant under Welch's t-test at $p < 10^{-9}$.

[6]This method of calculating community count assumes that a majority of communities will only contain relationships of a single edge type. Results with communities drawn from the entire graph are similar.

act, sometimes literally so in the context of literary data. To that end, social networks may change depending on the perspective of the narrator. Our current method does not allow these perspectives be reflected in graph structures. We believe future research should consider further methods for assessing the validity of extracted social networks.

Limitations

We note three primary limitations impacting this work. First, our textual data is sourced from predominantly European authors. Because it is an American project, the vast majority of volumes in Project Gutenberg are written in English. While the dataset does contain volumes written in at least 58 languages, the three most dominant are English, French, and German. The second limitation is that `Gemini 1.5 Flash` has a maximum context window of one million tokens. This means our pipeline could not process ∼374 volumes whose token counts exceed this maximum (although we note that the average text in our dataset contains 63,656 words, two orders of magnitude below this maximum). Finally, `Gemini 1.5 Flash` can only emit a maximum of 8,000 tokens in one API call. Our results indicate that some volumes contain social networks exceeding this maximum, suggesting some networks included in our dataset are incomplete.

Acknowledgments

References

Divya Agarwal, Devika Vijay, et al. 2021. Genre Classification Using Character Networks. In *Proceedings of the 5th International Conference on Intelligent Computing and Control Systems (ICICCS 2021)*, pages 216–222. IEEE.

Mark Algee-Hewitt. 2017. Distributed Character: Quantitative Models of the English Stage, 1550–1900. *New Literary History*, 48(4):751–782.

Luca Beurer-Kellner, Marc Fischer, and Martin Vechev. 2024. Guiding LLMs The Right Way: Fast, Non-Invasive Constrained Generation. In *Proceedings of the 41st International Conference on Machine Learning (ICML 2024)*, pages 3658–3673.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

Anthony Bonato, David Ryan D'Angelo, Ethan R Elenberg, David F Gleich, and Yangyang Hou. 2016. Mining and Modeling Character Networks. In *Algorithms and Models for the Web Graph: 13th International Workshop*, pages 100–114. Springer.

Julian Brooke, Adam Hammond, and Graeme Hirst. 2015. GutenTag: An NLP-driven Tool for Digital Humanities Research in the Project Gutenberg Corpus. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 42–47. Association for Computational Linguistics.

Snigdha Chaturvedi, Mohit Iyyer, and Hal Daume III. 2017. Unsupervised Learning of Evolving Relationships Between Literary Characters. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 3159–3165.

Snigdha Chaturvedi, Shashank Srivastava, Hal Daume III, and Chris Dyer. 2016. Modeling Evolving Relationships Between Characters in Literary Novels. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1):2704–2710.

Yi-Ting Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. MPDD: A multi-party dialogue dataset for analysis of emotions and interpersonal relationships. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 610–614, Marseille, France. European Language Resources Association.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, and collaborators. 2024. The Llama 3 Herd of Models.

David K. Elson, Nicholas Dames, and Kathleen R. McKeown. 2010. Extracting Social Networks from Literary Fiction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, pages 138–147.

Lawrence Evalyn, Susan Gauch, and Manisha Shukla. 2018. Analyzing Social Networks of XML Plays: Exploring Shakespeare's Genres. In *Proceedings of DH2018*, pages 368–370.

Michael Falk. 2016. Making Connections: Network Analysis, the Bildungsroman and the World of The Absentee. *Journal of Language, Literature and Culture*, 63(2-3):107–122.

Frank Fischer and Daniil Skorinkin. 2021. *Social Network Analysis in Russian Literary Studies*, pages 517–536. Springer International Publishing.

Georgi Gerganov. 2024. llama.cpp.

Martin Gerlach and Francesc Font-Clos. 2020. A Standardized Project Gutenberg Corpus for Statistical Analysis of Natural Language and Quantitative Linguistics. *Entropy*, 22(1):1–14.

Sil Hamilton and Andrew Piper. 2023. MultiHATHI: A Complete Collection of Multilingual Prose Fiction in the HathiTrust Digital Library. *Journal of Open Humanities Data*, 9(1):1–7.

Michael S. Hart. 1971. Project Gutenberg.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7B.

Evgeny Kim and Roman Klinger. 2019. Frowning Frodo, Wincing Leia, and a Seriously Great Friendship: Learning to Classify Emotional Relationships of Fictional Characters. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 647–653, Minneapolis, Minnesota. Association for Computational Linguistics.

Taku Kudo and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

James Lee and Jason Lee. 2017. Shakespeare's Tragic Social Network; Or Why All the World's A Stage. *Digital Humanities Quarterly*, 11(2).

John Lee and Chak Yan Yeung. 2012. Extracting Networks of People and Places From Literary Texts. In *Proceedings of the 26th Pacific Asia Conference on Language, Information and Computation*, pages 209–218, Bali, Indonesia. Faculty of Computer Science, Universitas Indonesia.

V. I. Levenshtein. 1965. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics Doklady*, 10:707–710.

Víctor Hugo Masías, Paula Baldwin, Sigifredo Laengle, Augusto Vargas, and Fernando A Crespo. 2017. Exploring the Prominence of Romeo and Juliet's Characters Using Weighted Centrality Measures. *Digital Scholarship in the Humanities*, 32(4):837–858.

Philip Massey, Patrick Xia, David Bamman, and Noah A. Smith. 2015. Annotating character relationships in literary texts.

Simone Mellace, Alessandro Antonucci, et al. 2020. Temporal Embeddings and Transformer Models for Narrative Text Understanding. In *Proceedings of the Text2Story'20 Workshop*, pages 71–77, Lisbon, Portugal.

Microsoft. 2024. Guidance. Microsoft, Inc.

Semi Min and Juyong Park. 2016. Network Science and Narratives: Basic Model and Application to Victor Hugo's Les Misérables. In *Complex Networks VII: Proceedings of the 7th Workshop on Complex Networks CompleNet 2016*, pages 257–265. Springer International Publishing.

Franco Moretti. 2011. Network Theory, Plot Analysis. *New Left Review*, 68.

M Nijila and MT Kala. 2018. Extraction of Relationship Between Characters in Narrative Summaries. In *2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR)*, pages 1–5. IEEE.

Bowen Peng, Jeffrey Quesnelle, Honglu Fan, and Enrico Shippole. 2023. YaRN: Efficient Context Window Extension of Large Language Models. In *Proceedings of the 12th International Conference on Learning Representations (ICLR 2023)*.

Andrew Piper. 2022. Biodiversity Is Not Declining in Fiction. *Journal of Cultural Analytics*, 7(3):1–15.

Andrew Piper, Michael Xu, and Derek Ruths. 2024. The Social Lives of Literary Characters: Combining Citizen Science and Language Models to Understand Narrative Social Networks. In *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, pages 472–482, Miami, USA. Association for Computational Linguistics.

Andrew J. Reagan, Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Peter Sheridan Dodds. 2016. The Emotional Arcs of Stories Are Dominated by Six Basic Shapes. *EPJ Data Science*, 5(1):1–12.

Matt Rickard. 2024. ReLLM.

Connor Shorten, Charles Pierse, Thomas Benjamin Smith, Erika Cardenas, Akanksha Sharma, John Trengrove, and Bob van Luijt. 2024. StructuredRAG: JSON Response Formatting with Large Language Models.

Roel Smeets, Maarten De Pourcq, and Antal van den Bosch. 2021. Modeling Conflict: Representations of Social Groups in Present-Day Dutch Literature. *Journal of Cultural Analytics*, 6(3).

James Stiller and Mathew Hudson. 2005. Weak Links and Scene Cliques Within the Small World of Shakespeare. *Journal of Cultural and Evolutionary Psychology*, 3(1):57–73.

James Stiller, Daniel Nettle, and Robin IM Dunbar. 2003. The Small World of Shakespeare's Plays. *Human Nature*, 14:397–408.

Jianlin Su, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. 2023. RoFormer: Enhanced Transformer with Rotary Position Embedding. *Neurocomputing*, 568(C).

Saatviga Sudhahar and Nello Cristianini. 2013. Automated Analysis of Narrative Content for Digital Humanities. *International Journal of Advanced Computer Science*, 3(9):440–447.

Kashin Sugishita and Naoki Masuda. 2023. Social Network Analysis of Manga: Similarities to Real-World Social Networks and Trends Over Decades. *Applied Network Science*, 8(1):79.

Team Gemini, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy Lillicrap, Angeliki Lazaridou, Orhan Firat, James Molloy, Michael Isard, and et al. 2024. Gemini: A Family of Highly Capable Multimodal Models.

Team Gemma, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, Johan Ferret, Peter Liu, Pouya Tafti, Abe Friesen, Michelle Casbon, Sabela Ramos, Ravin Kumar, Charline Le Lan, Sammy Jerome, Anton Tsitsulin, Nino Vieillard, Piotr Stanczyk, Sertan Girgin, Nikola Momchev, Matt Hoffman, Shantanu Thakoor, Jean-Bastien Grill, Behnam Neyshabur, Olivier Bachem, Alanna Walton, Aliaksei Severyn, Alicia Parrish, and et al. 2024. Gemma 2: Improving Open Language Models at a Practical Size.

Beate Volker and Roel Smeets. 2020. Imagined Social Structures: Mirrors or Alternatives? A Comparison Between Networks of Characters in Contemporary Dutch Literature and Networks of the Population in the Netherlands. *Poetics*, 79:1–15.

Thomas Way. 2018. A Framework for Unsupervised Extraction of Social Networks from Textual Materials. In *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, pages 244–249, New York, NY, USA. Association for Computing Machinery.

Matthew Wilkens, Elizabeth F Evans, Sandeep Soni, David Bamman, and Andrew Piper. 2024. Small Worlds: Measuring the Mobility of Characters in English-Language Fiction. *Journal of Computational Literary Studies*, 3(1):1–16.

Runcong Zhao, Qinglin Zhu, Hainiu Xu, Jiazheng Li, Yuxiang Zhou, Yulan He, and Lin Gui. 2024. Large Language Models Fall Short: Understanding Complex Relationships in Detective Narratives. In *Findings of the Association for Computational Linguistics: ACL 2024*, Bangkok, Thailand. Association for Computational Linguistics.