Multi-Graph Encoder-Decoder Model for Location-Based Character Networks in Literary Narrative

Avi Bleiweiss
BShalem Research, Sunnyvale, U.S.A.

Keywords: Narrative Intelligence, Geo-Location, Graph Theory, Community Discovery, Neural Networks.

Abstract: In the past decade, an extended line of research developed a broad range of methods for reasoning about narrative from a social perspective. This often revolved around transforming literary text into a character network representation. However, there remain inconsistent traits of narrative structure produced computationally by either neural language technology or network theory tools. In this paper, we propose an encoder-decoder model with a main objective to mitigate the apparent computational divergence. Our encoder novelty lies in generating hundreds of location-based network graphs to render a fine-grained narrative. We further formalize a decoder task for detecting character communities and analyze modularity and membership affiliation. Through empirical experiments, we present visualization of stages in our computational process for four literary fiction novels.

1 INTRODUCTION

Over a decade ago, the study of literature underwent a major shift from close reading of individual texts to the construction of abstract models. The quantitative approach to literature has Moretti (2005) graphically map out text according to history, geography, and social connections. Moreover, by turning time into space, a narrative plot can be further represented as a social network of characters and interactions (Moretti, 2011). Network analysis also offers a powerful mode of intrinsic criticism, by providing empirical measures of the novel social scale and density (Alexander, 2019).

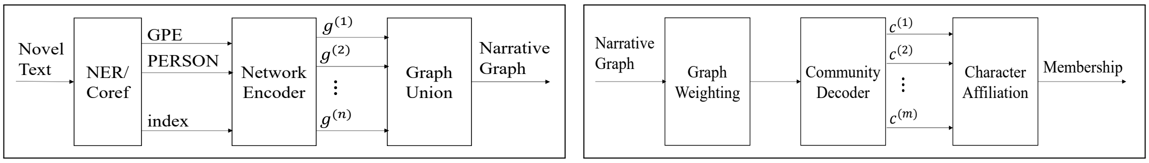
A character network is a graph describing a narrative by representing the characters as its vertices, and their structural relationships with others through its edges. Edges are often attributed a weight and direction properties to express multidimensional relatedness. Because much of what the characters do or say is narrated, a direct discourse only covers a small part of the plot and thus the transformation of plots into networks is a lot less consistent. Our work centers around automating network extraction from fictional novel text by applying advanced natural language processing (NLP) technology.

The emergence of sociological approaches to a narrative, cast characters as points of social intersection rather than centered subjects. In her seminal work, Levine (2009) suggests the networked novel extends alternatives to conventional constructions and offers a perceptive account of how disease reveals social networks. Characters in text are drawn into one great distributed network, but often they act as nodes on two or more different networks. Our study is motivated by having the unfolding plot revolve around multiple principles of interconnections to capture social experience.

Understanding how spoken language is represented in a novel over time is a key question in the digital humanities. In particular, representing social relationships between characters is an important component of literature, as Elson et al. (2010) took the first steps toward automating the task of mention-level quote attribution for literary text. In our work, instead of conversational networks we render the relatively understudied narrative dimension of named places into a literary network structure that captures character-specific physical settings and geo-locations. We present the network as a collection of many small location-based graphs, thus encoding an explicit spatial logic of the narrative. Our study leverages recent advances in neural NLP for the tasks of named entity recognition (NER) and coreference resolution.

This paper offers the following key contributions: (1) we propose a graph encoder-decoder model that consistently retains a defined narrative structure at each of its computational stages, (2) we introduce a

Bleiweiss, A.
Multi-Graph Encoder-Decoder Model for Location-Based Character Networks in Literary Narrative.
DOI: 10.5220/0011776400003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 3*, pages 790-797
ISBN: 978-989-758-623-1; ISSN: 2184-433X



(a) Encoder. (b) Decoder.

Figure 1: A graph encoder-decoder model for extracting and analyzing a location network that represents a narrative in literature. The encoder input is a literary text sequence that follows location and character NER along with coreference resolution of aliases. In the next stage, the encoder computes n small geo-location graphs $g_{1:n}$, each linking a handful of characters. The graphs are then merged into a single narrative network G. The decoder operates on graph G and performs pairwise graph weighting and community discovery to produce m communities $c_{1:m}$. This process compresses the narrative representation of hundreds of geo-location graphs down to under a dozen communities. Lastly, the decoder extracts character memberships from each community.

formal decoder that is tasked with detecting character communities that reduce the literary network complexity from hundreds of small graphs to less than a dozen entities, and (3) through extensive qualitative and quantitative analysis we provide visualization of the narrative network drawn from different perspectives.

2 RELATED WORK

Detecting character relations plays an important role in narrative understanding built upon social network theory. In this context, one of the more vital contributions to literary theory in the past two decades is the concept of character-space relation, coined by Woloch (2009). In his account of fictional characterization, a figure, central or minor, emerges in a predetermined position within the narrative. The configuration and intersections between many character-spaces within a single narrative are essential to the dynamics of literary representation. In our work, we follow this core abstraction and generate many location-based network graphs to render a novel. We then unionize the graphs to form a single social network and analyze the distribution of character communities.

The critical review by Labatut and Bost (2019) presents extensive scientific research related to extracting and analyzing fictional character networks. We follow in more detail on the methods most relevant to our study.

Considered by many the more influential work, Elson et al. (2010) characterize a text of literary fiction by extracting a network of social conversations. They obtain character mentions from conversational segments in nineteenth-century British novels by analyzing their dialogue interaction. Using the NER pipeline in CoreNLP to discover character names, they expand on aliases by using manually drawn coreference chains. The use of explicit quotations in text for inferring links between characters remains however ambiguous at times.

To identify interacting agents, Agarwal and Rambow (2010) build character networks using tree kernel-based relation extraction from structured parse tree information. A text snippet may thus describe social relations between two individuals explicitly by the type of a relation, or implicitly by a social event. Whereas Lee and Yeung (2012) assert that quoted speech need not assumed to be the main course of encoding interpersonal relations (Elson et al., 2010). Besides people and events, they also integrate locations in their networks. Our work expands on their model and a geographical node is more than just a global staging position, instead we generate many network graphs to represent the novel text, each of a star topology with a hub that captures a location chronologically.

A machine learning model devised by Celikyilmaz et al. (2010) describes a probabilistic approach for detecting conversations between actors in a novel, and analyzing networks built based on topical similarity in actor speech. Their method follows the linguistic intuition that rich contextual information can be useful in understanding dialogues. On the other hand, He et al. (2013) proposed an alternate venue for identifying speaker references in novels, using a probabilistic model that exploits lexical and syntactic clues in the text itself. However, they manually construct a list of characters and their aliases, unlike Vala et al. (2015), who proposed an eight stage pipeline for detecting characters automatically, which builds a graph where nodes are names and edges connect names belonging to the same character.

Compelling is Edwards et al. (2020) study that model and compare manual to automatic co-occurrence and unsupervised machine learning methods for extracting social networks from narratives. In their findings, automatic extraction methods produced commensurable results for density and centrality measures with the more accurate but by far more time consuming manual approach. While edge weights were only moderately correlated with a 0.8 Spearman coefficient for NLP networks. Although their experiments conducted on a television show, their conclusions are likely to extend to literary narratives.

More recently, Schmidt et al. (2021) applied both a rule-based pipeline and an end-to-end deep learning model to NER and coreference resolution (Lee et al., 2018), and showed that neural networks outperform the rule-based approach on most evaluation settings. Lastly, the effort by Piper et al. (2021) seeks to provide a coherent theoretical foundation for implementing NLP computational solutions and identify narrative as an important basis for understanding human behavior.

3 MODEL

In Figure 1, we provide an overview of our proposed graph encoder-decoder model.

The encoder process for transforming novel text to a list of location-based networks of characters is straightforward. We scan the entire novel text and extract words defined as either Geo-Political (GPE) or PERSON entities, using the spaCy named entity recognition (NER) and the neural coreference resolution (Coref) tools. [1] A GPE occurrence triggers both terminating the creation of the current network and also starting to construct a new network. After that we collect all the PERSON names that are delimited between a pair of GPEs or a GPE and end of text. The graphs we construct are each of an undirected star topology— they have the GPE as the root node, and all the PERSON nodes are the leaves. In addition, we draw timestamp attributes from the running word index of either the GPE or PERSON entities. The indices aid in retaining the chronology of the story telling.

More formally, using a colon notation, we denote a collection of k PERSON entities $p_{1:k} = (p_1, \dots, p_k)$. Given k characters in a location, the graph g that we construct has $k+1$ vertices and k edges emanating from $1:k$ all to vertex $k+1$. The graph cardinality is thus $|g| = k$. PERSON names p_i and location identity l are attached to their corresponding vertex labels. We unionize all n graphs $g^{(j)}$ into one network $G = \left(g^{(1)}, \dots, g^{(n)}\right)$ that represents the narrative, expecting location graphs $g^{(j)}$ to connect each a small number of characters— about a handful on average. The algorithm for integrating subgraphs $g^{(j)}$ has linear node complexity with the total number of characters in a narrative. Thus, the same character residing in multiple locations is assigned an identical node ID to effectively address large-scale and dynamic narratives.

Unlike the encoder components that mostly operate in NLP space, the decoder computationally functions entirely in the network theory domain. The narrative network G links the encoder with the decoder, and the latter produces a set of m communities, where $m << n$. Each community c is a disjoint union of separable geo-location graphs. We denote a community set $C = \left(c^{(1)}, \dots, c^{(m)}\right)$. While there are several methods for community detection in networks, in our work, we chose the Louvain (Blondel et al., 2008) hierarchical clustering algorithm, owing to both its ability to detect high-modularity community partitions and its exceptional computational efficiency with a runtime of $O(n \log n)$. The last stage of the decoder generates character membership in each community, we further use for visual analysis.

4 EVALUATION

In this section, we provide for our experiments visualization of network modeling and report computational metrics using igraph (Csardi and Nepusz, 2006). [2]

Table 1: Our test set of fiction literary novels.

Title	Chapters	Tokens
Sign of the Four	12	43,736
Portrait of a Lady	27	112,661
Emma	55	159,950
David Copperfield	64	361,938

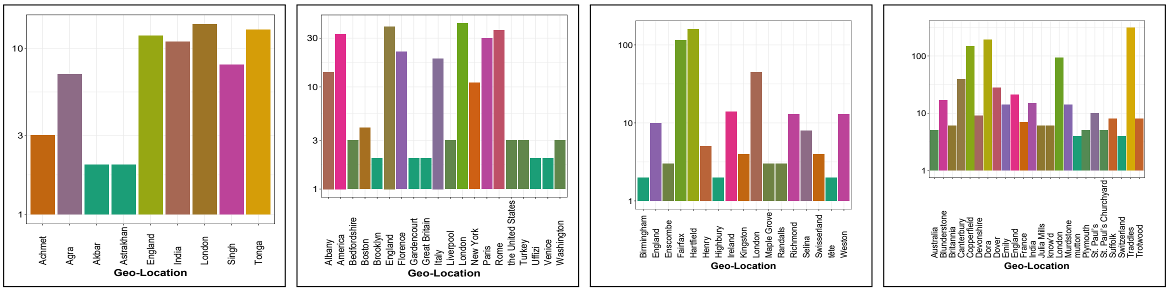
Table 2: Location character density across test novels.

Title	Characters	Locations	Density
Sign of the Four	558	106	5.26
Portrait of a Lady	1,436	319	4.50
Emma	3,980	443	8.98
David Copperfield	5,845	1,106	5.28

Novel Test Set. We obtained unicode encoding of the fictional literature text from Project Gutenberg, and carried our work on four 19-century British novels including `The Sign of the Four` by Co-

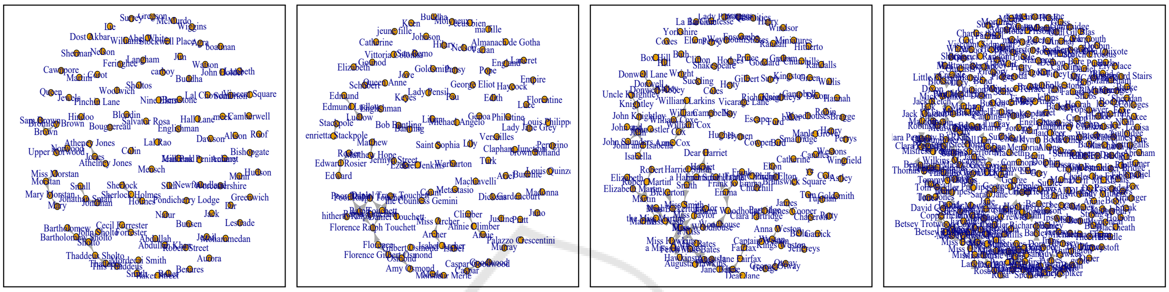
[1] https://spacy.io/

[2] https://igraph.org



(a) Sign of the Four. (b) Portrait of a Lady. (c) Emma. (d) David Copperfield.

Figure 2: Named geo-location distribution in logarithmic scale over our test novels.



(a) Sign of the Four. (b) Portrait of a Lady. (c) Emma. (d) David Copperfield.

Figure 3: Network representation of our test novels. Vertices are labeled with unique character names, as edges initially connect multi-word coreferent variations of the name.

nan Doyle (1890), [3] `The Portrait of a Lady` by Henry James (1881), [4] `Emma` by Jane Austen (1815), [5] and `David Copperfield` by Charles Dickens (1849). [6] These texts totaled more than half a million words (Table 1). Our choice of the test corpus concurs with the test set in Elson et al. (2010) and facilitates compare with a baseline. Although, unlike randomly selecting a handful of chapters from each book, we used each novel text of its entirety.

Past the NER stage, the encoder has sufficient data to assess both geo-location density and distribution in the novel narratives. In Table 2, we show location density ranging from a handful up to nine characters across our test novels. This is suggestive of the number of location networks, as well as the cardinality of narrative graphs our encoder constructs. While in Figure 2, we show in logarithmic scale the distribution of identified named geo-locations in each of our test novels, excluding places that are mentioned only once. We expect graphs with hubs of same-name locations to be merged in the community discovery stage of the decoder. To ensure high quality NER, we conducted several iterations of increasing the train set offered by spaCy and circumvent false-negative named entities for failing to flag an entity. However, a few false positives for wrongly tagging a non-entity as a place entity were unavoidable. For example, the word 'Know'd' in the novel `David Copperfield` was singled out as GPE (Figure 2).

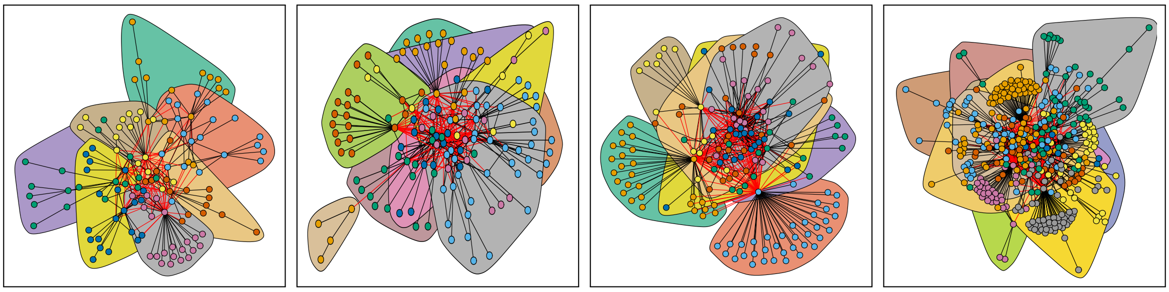
Our encoder retains a skeletal character network for each literary narrative to aid analyzing character aliases visually. Aliases transpire for mentioned multi-word names and are learned by the neural coreference resolver. We outline aliases in each of the narrative network representations in Figure 3. Graph vertices are labeled with unique character names, and the edges, shown at the bottom left of the network, connect vertices of character aliases.

Community Discovery. Detecting communities in a literary narrative network based on geo-locations is useful to broaden character social relations and complement the downstream task of quote attribution. To this extent, community discovery can answer the question of association by singling out subgraphs with identical named locations and effectively merging them together into a concise representation. Similarly, a community may link subgraphs of different named locations while sharing the same character node attributes. To help understand the narrative space concept, a community structure facilitates events that are likely to disseminate across a multi-

[3]https://www.gutenberg.org/files/2097/2097-0.txt
[4]https://www.gutenberg.org/files/2833/2833-0.txt
[5]https://www.gutenberg.org/files/158/158-0.txt
[6]https://www.gutenberg.org/files/766/766-0.txt

Table 3: Statistical summarization of graph node distribution that characterizes the input to our community discovery stage.

Title	Graphs	Nodes	Min	Max	Mean	STD
Sign of the Four	66	320	1	26	4.84	5.06
Portrait of a Lady	216	635	1	10	2.93	2.06
Emma	349	1,707	1	20	4.89	3.59
David Copperfield	672	1,890	1	20	2.81	2.50



(a) Sign of the Four. (b) Portrait of a Lady. (c) Emma. (d) David Copperfield.

Figure 4: Character community discovery across our test novels.

tude of places. Furthermore, uncovering organizational principles in networks sets the sphere of activity bounds where the narrative plot unfolds.

The input to the community discovery task consists of independent geo-location subgraphs. In Table 3, we provide subgraph count and statistical summarization of the subgraph character nodes for each of our test novels. As expected, literary networks are represented for the most part with hundreds of small networks, each of a handful of characters, on average.

In Figure 4, we provide visualization of community partitions. The Louvain algorithm we used uncovers community structures at different resolutions (Lambiotte et al., 2014), and uses modularity as an objective function to optimize for finding the best subdivision of a network. Identified communities are both small and large and suggest a plausible representation compression of at least an order of magnitude, for a narrative network with hundreds of geo-subgraphs reduced to around ten communities. In Table 4, we show the number of discovered communities, along with the modularity quality scale, a numerical scalar that measures the relative density of edges inside communities with respect to edges outside communities. Given the $[-1,+1]$ range, our modularity scores proved sufficiently compelling.

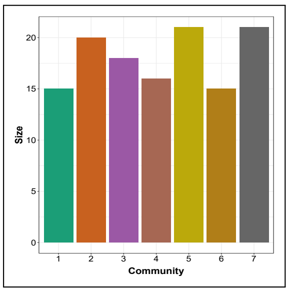
The geo-location distributions shown in Figure 2 present the first order for predicting community partitions in a literary narrative. Qualitatively, we anticipated subgraphs with the same named location at their root to be merged into a single community. Thus, the number of communities m for each of the narratives in our novel test set ought not to exceed $m \in \{9, 20, 17, 24\}$, respectively. The Louvain algorithm effectively reduced the number of subdivisions to $m \in \{7, 10, 7, 10\}$, respectively, with an impressive 2X compression ratio for the larger three narratives.

Character Affiliation. Community graph partitions are a vital resource to learn about character distributions. Allocations of member affiliation with a community gives the division of the vertices across communities and is outlined in Table 4. Character apportions for `The Sign of Four` novel are evidently fairly balanced, however, the remainder of the literary narratives rather offer diverse cluster sizes. In Figure 5, we show a dendrogram plot of the membership extracted from a sampled narrative community. We applied a hierarchy plane cutoff and rendered a handful of clearly distinct character clusters for improved visualization. We note that named entities at the leaf nodes of each subtree may consist of both persons and geo-locations, of which the latter can be further filtered if so desired. The community dendrogram split introduces a representation with a new and concise set of character relationships and thus much more simpler to understand.

Temporal Relations. Our encoder captures the temporal aspect of narrative evolution by the association of named entities with a token running index over a narrative text sequence. In Figure 6, we show the decoder interpretation of timestamp intervals for twenty named entities including both character and geo-location samples obtained from `The Portrait of a Lady` narrative. A few intervals shown with a

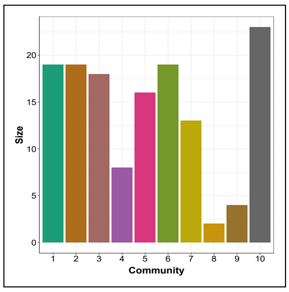
Table 4: Output community distribution across our test novels.

Communities	7
Modularity	0.43



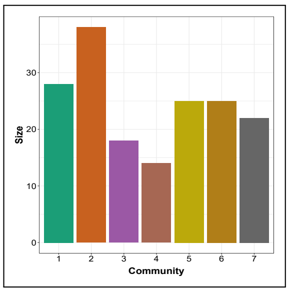
(a) Sign of the Four.

Communities	10
Modularity	0.37



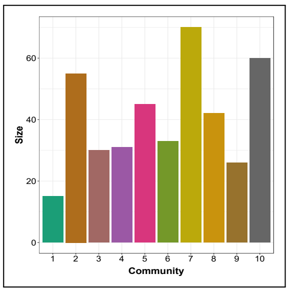
(b) Portrait of a Lady.

Communities	7
Modularity	0.32

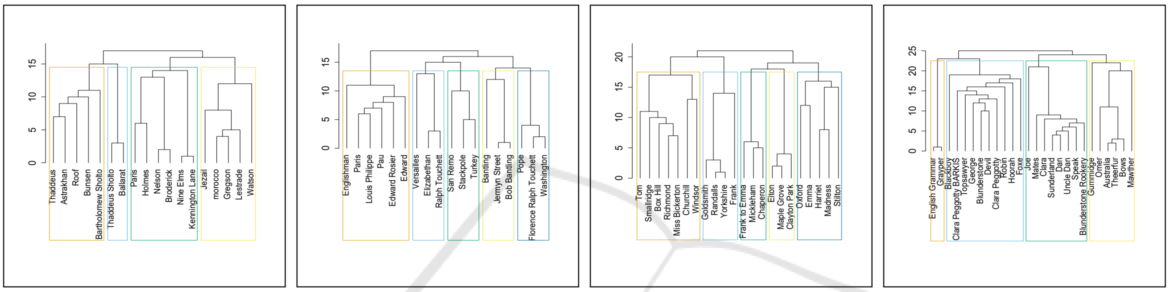


(c) Emma.

Communities	10
Modularity	0.44



(d) David Copperfield.



(a) Sign of the Four. (b) Portrait of a Lady. (c) Emma. (d) David Copperfield.

Figure 5: Dendrogram visualization of a sample of character affiliation with a community from each of our test novels.

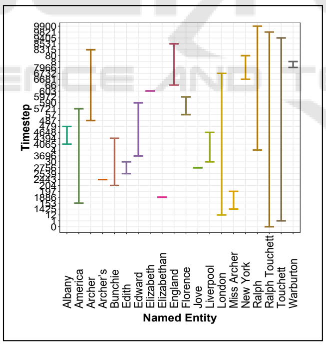


Figure 6: Timestamp intervals of named entity samples.

horizontal bar are of an empty order, thus indicating a single appearance instance in the entire plot. Instead of using network theory features (Besnier, 2020), the unfolding of role importance over time for either a character or a location for this matter is rather implicit in our framework— the longer the interval the higher the significance attached to the entity plot contribution. In an absolute sense, an entity is given importance priority that we qualify by how close is a long-standing interval endpoint to the conclusion of the narrative.

Table 5: Physical and virtual location-based narrative accuracy across our test set.

Title	Physical	Virtual	Accuracy
Sign of the Four	54	12	0.82
Portrait of a Lady	150	66	0.70
Emma	295	54	0.85
David Copperfield	412	260	0.61

Error Analysis. In our model, character entities share a location based on the proximity of the person mention to the location mention in the narrative text. We distinguish between physical and virtual plot bound geo-locations, with the latter represented as single-node graph occurrences that render no social relationship (Table 3). While virtual locations are perfectly valid entities in an automated graph construction, we consider them an exception. This lets us attach a quality measure for transforming narrative text to a multitude of geo-location subgraphs, using an accuracy form: $l_p/(l_p+l_v)$, where l_p and l_v are physical and virtual locations, respectively. In Table 5, we show narrative accuracy measures across our test novels with a plausible corpus mean of 0.75.

5 DISCUSSION

In the context of a narrative space, we sought after practical graph-based computational methods that match our proposed model. We attend to the role space plays in a narration as a feature capacity of the story plot, with places that make up the physical environment in which the characters of a narrative live and move (Brasher, 2017).

Narrative Graph to Graphormer. In this section, we contrast our graph encoder-decoder approach with the well established Transformer (Vaswani et al., 2017) architecture in modeling natural language data. We reviewed whether the Transformer is suitable to model graphs and make graph representation learning work for the task of narrative network understanding end-to-end, while considering feeding the Transformer with our encoder output G.

Table 6: Train scores for GNN node classification.

Title	Precision	Recall	F1	Support
Sign of the Four	0.33	0.48	0.37	79
Portrait of a Lady	0.29	0.44	0.33	88
Emma	0.26	0.34	0.27	107
David Copperfield	0.11	0.22	0.13	255

Table 7: Comparing performance to an external baseline.

System	Precision	Recall	F1
Elson et al. (2010)	0.54	0.55	0.48
Ours	0.25	0.37	0.28

To this end, mainstream variants of graph neural networks (GNNs; Scarselli et al., 2009) have shown to outperform the Transformer on many graph-level classification tasks. Recently, Ying et al. (2021) introduced the Graphormer, [7] built upon a standard Transformer neural network that encodes directly the structural information of graphs. Their proposed centrality and spatial encodings proved many GNN variants may be cast as special Graphormer cases, and shown to lead the state-of-the-art performance on a wide range of graph-level prediction tasks. However, in its current state the Graphormer mostly attends to node classification and less on loose subgraph clustering— vital for learning character-centric narratives. Moreover, the training datasets for benchmarking the Graphormer were primarily scraped from physical and biological science graphs and would have to be augmented with literary-specific network data to benefit the performance of narrative understanding.

Narrative Graph to GNN. To reason about our system performance, we linked the narrative graph structure G to a GNN. In Table 6, we outline weighted-average train scores of GNN node classification across our literary narratives, noting that recall and accuracy measures are nearly identical. We apportioned our graph data for each novel into train, validation, and test splits of 70/10/20 percent, respectively, and used degree as the node embedding representation. Our neural model was constructed from a two-layer graph convolutional network (GCN), and we ran 500 training epochs on each of our narratives using the Adam optimizer with a fixed dropout of 0.2. Although not intended to demonstrate performance, the observation that predicting F1 scores decline with the length of the narratives is compelling on its own.

We sought after comparing our classification performance to an external baseline. Using a test corpus of four novels, with identical titles to the ones used in Elson et al. (2010) for system evaluation, served our purpose well, although the goals, tools, and resources vary greatly between the implementations. For example, they used a single large character network, where we applied many small geo-location graphs. While they explored 60 novels for their train set with over ten million words, their novel test set is a considerable small subset of our texts. In table 7, we compare our average classification scores across test novels to the mean scores across the methods for detecting conversations in Elson et al. (2010). We anticipated lower scores for connecting our system to GNN, shown at about 0.6X of the baseline.

6 CONCLUSIONS

In this paper, we propose an encoder-decoder computational model to reason about a narrative spatially. Surprisingly, many questions can be answered by measuring the relationships between places mentioned in text. Geo-location networks are especially useful to narrow down the search space of a narrative, before deploying a realistic discourse structure.

Our analyses carve several avenues of future research, such as alias resolution in mentioned location names, introduce character proximity relationships by using graph weights to represent physical distance between places, and address a more authentic temporal evolution of a community over an unveiling narrative.

[7] https://github.com/Microsoft/Graphormer

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their insightful suggestions and feedback.

REFERENCES

Agarwal, A. and Rambow, O. (2010). Automatic detection and classification of social events. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1024–1034, Cambridge, MA.

Alexander, S. (2019). Social network analysis and the scale of modernist fiction. *Modernism/modernity*, 3(4).

Besnier, C. (2020). History to myths: Social network analysis for comparison of stories over time. In *Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (SIGHUM)*, pages 1–9, Online.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

Brasher, J. P. (2017). Narrating space/spatializing narrative: Where narrative theory and geography meet. *The American Association of Geographers Review of Books*, 5(3):180–182.

Celikyilmaz, A., Hakkani-Tu, D., He, H., Kondrak, G., and Barbosa, D. (2010). The actor-topic model for extracting social networks in literary narrative. In *Neural Information Processing Systems (NIPS) in Machine Learning for Social Computing Workshop*, page 7, Whistler, Canada.

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.

Edwards, M., Tuke, J., Roughan, M., and Mitchell, L. (2020). The one comparing narrative social network extraction techniques. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 905–913.

Elson, D., Dames, N., and McKeown, K. (2010). Extracting social networks from literary fiction. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 138–147, Uppsala, Sweden.

He, H., Barbosa, D., and Kondrak, G. (2013). Identification of speakers in novels. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1312–1320, Sofia, Bulgaria.

Labatut, V. and Bost, X. (2019). Extraction and analysis of fictional character networks: A survey. *ACM Computing Surveys*, 52(5):1–40.

Lambiotte, R., Delvenne, J.-C., and Barahona, M. (2014). Random walks, markov processes and the multiscale modular organization of complex networks. *IEEE Transactions on Network Science and Engineering*, 1(2):76–90.

Lee, J. and Yeung, C. Y. (2012). Extracting networks of people and places from literary texts. In *Pacific Asia Conference on Language, Information, and Computation*, pages 209–218, Bali, Indonesia.

Lee, K., He, L., and Zettlemoyer, L. (2018). Higher-order coreference resolution with coarse-to-fine inference. In *The North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, pages 687–692, New Orleans, Louisiana.

Levine, C. (2009). Narrative networks: Bleak house and the affordances of form. *Novel: A Forum on Fiction*, 42(3):517–523.

Moretti, F. (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso Books, Brooklyn, New York.

Moretti, F. (2011). *Network Theory, Plot Analysis*. Stanford Literary Lab, Palo Alto, California.

Piper, A., So, R. J., and Bamman, D. (2021). Narrative theory for computational narrative understanding. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 298–311, Online.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80.

Schmidt, D., Zehe, A., Lorenzen, J., Sergel, L., Düker, S., Krug, M., and Puppe, F. (2021). The FairyNet corpus - character networks for German fairy tales. In *Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (SIGHUM)*, pages 49–56, Online).

Vala, H., Jurgens, D., Piper, A., and Ruths, D. (2015). Mr. bennet, his coachman, and the archbishop walk into a bar but only one of them gets recognized: On the difficulty of detecting characters in literary texts. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 769–774, Lisbon, Portugal.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems (neurIPS)*, volume 30. Curran Associates, Inc.

Woloch, A. (2009). *The One vs. the Many: Minor Characters and the Space of the Protagonist in the Novel.* Princeton University Press, Princeton, New Jersey.

Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., and Liu, T.-Y. (2021). Do transformers really perform badly for graph representation? In *Advances in Neural Information Processing Systems (neurIPS)*, volume 34, pages 28877–28888. Curran Associates, Inc.