Measuring Information Propagation in Literary Social Networks

Matthew Sims
School of Information
UC Berkeley
msims@berkeley.edu

David Bamman
School of Information
UC Berkeley
dbamman@berkeley.edu

Abstract

We present the task of modeling information propagation in literature, in which we seek to identify pieces of information passing from character A to character B to character C, only given a description of their activity in text. We describe a new pipeline for measuring information propagation in this domain and publish a new dataset for speaker attribution, enabling the evaluation of an important component of this pipeline on a wider range of literary texts than previously studied. Using this pipeline, we analyze the dynamics of information propagation in over 5,000 works of English fiction, finding that information flows through characters that fill structural holes connecting different communities, and that characters who are women are depicted as filling this role much more frequently than characters who are men.

1 Introduction

With the rise of sociological approaches to narrative, work in literary criticism has increasingly turned to the ways in which authors depict social networks in their texts. This includes critical attention to both *network topologies*, such as understanding characters and their structural relationships with others (Levine, 2009), and *information flow*, such as theorizing the representation of disease and gossip (Levine, 2009; Margolis, 2012; Spacks, 1985). Much computational work in NLP has arisen to support the former line of research, including extracting social networks from text (Elson et al., 2010), predicting familial relationships (Makazhanov et al., 2014), and modeling the interactions between characters (Iyyer et al., 2016; Chaturvedi et al., 2017). This in turn has driven work in the digital humanities examining the structure of literary networks (Moretti, 2011; Algee-Hewitt, 2017; Piper et al., 2017; Alexander, 2019).

At the same time, however, there remains a substantial gap in computational work to support the
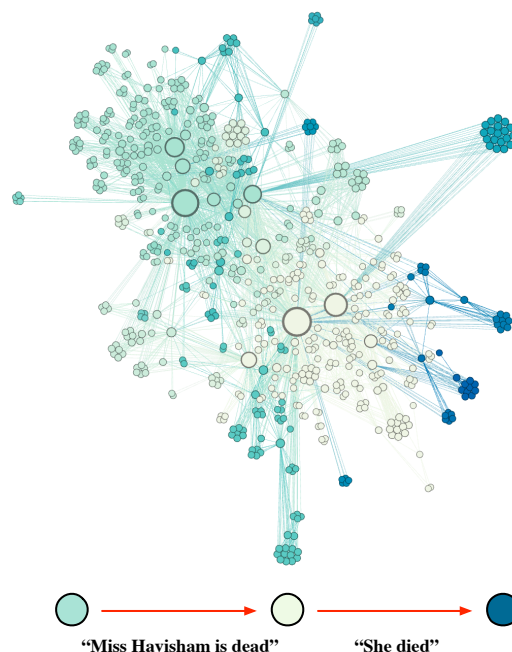

Figure 1: The character co-occurrence network for *Great Expectations*. Nodes represent characters and edges represent conversational interactions. Below the network, we illustrate an example of information transmission across a character triad.

latter research goal of modeling the flow of information within depicted networks. Yet understanding how the transmission of information is represented in these imagined worlds has the potential to be of great value to scholars in the humanities, since the resulting models can serve as a basis for broader insights about the social structures embedded in narratives, the role of characters based on attributes such as race and gender, and the informational dynamics of gossip (Spacks, 1982, 1985; Martin, 2014).

In this work, we specifically aim to fill this gap by developing methods to track the flow of information in novels by extracting instances of a message

passing from character A to character B to character C, only given a depiction of their conversational interactions. We develop a methodology for modeling this mode of propagation in both *explicit* networks (where one character provides information that is explicitly attributed to another character, such as "Bob told me that Jack escaped"); and in *implicit* networks, where information is repeated by multiple characters without such attribution. While the results of the methods enable a range of potential analyses—for instance, comparative analysis between authors, characters, and dyads—we focus on two illustrative case studies. First, we examine the linchpins of information flow—the characters who are most responsible for the propagation of information—and how they are positioned relative to the overall network topology; and second, we examine the gender dynamics of information propagation and what it tells us about how novelists represent men and women as the means and agents for transmitting facts, gossip, and other details about the social workings of these imagined worlds.

We make the following contributions with this work:

1. We present a new NLP pipeline for determining information propagation in literary texts, incorporating a range of different sub-tasks, including coreference resolution, speaker attribution, character network identification, and information extraction.

2. We present a new dataset for speaker attribution, comprised of 1,765 quotations linked to their speakers in 100 different literary texts, allowing us to evaluate a critical component of this pipeline on a wider range of literary texts than previously studied.

3. We leverage our pipeline to analyze the dynamics of information propagation in a collection of 5,345 works of English fiction from Project Gutenberg. We find that information flows through characters that fill structural holes connecting different communities, and that characters who are women are depicted as filling this role much more frequently than characters who are men.

2 Related work

Much of the computational research into information propagation and diffusion has focused on the domain of social media (Bakshy et al., 2012). Research in this area includes analyses of information diffusion in blogs (Gruhl et al., 2004; Leskovec et al., 2007), the spread of news across online networks (Leskovec et al., 2009), and in particular, the spread of rumor and misinformation (Kwon et al., 2013; Friggeri et al., 2014; Del Vicario et al., 2016; Vosoughi et al., 2018).

A core aspect of this work that strongly differs from networks in fiction is that the individual components of social media networks (the nodes, edges, and instances of propagation) are often directly observed. In modeling retweet dynamics in Twitter, for instance, nodes are defined as unique users, edges are directly observed friend and follow links defined by the platform, and propagation occurs when one user retweets a message posted by another they are connected to. More closely related to the challenges posed by detecting propagation in fiction is work that may directly observe the node and edge structure of a network, but must infer an act of *propagation*, including work in tracking the diffusion of memes (Leskovec et al., 2009), text reuse across legislative bills (Wilkerson et al., 2015) and quotations in news (Niculae et al., 2015).

While information propagation has yet to inform work in narrative (hence the purpose of this study), network structure has increasingly informed literary scholarship. Following the work of Bourdieu (1996), literary scholars have in recent years begun to explore the role that social networks play both in authorial composition (So and Long, 2013; Mazanec, 2018) and in the narrative representation of "networked social experience" (Levine, 2009).

Treating literary works *themselves* as networks, however, poses distinct computational challenges. While research into information propagation in social media tends to presume access to explicit networks, the character networks represented in novels are implicit. To determine these networks, we draw on previous work by Elson et al. (2010), who build edges between character nodes through conversational interactions. Various computational work to extract social networks from literature has built on this research over the past ten years,[1] including fundamental methods designed to extract networks for other languages like German (Jannidis et al., 2016), incorporate other categories of nodes such as locations (Lee and Yeung, 2012) and objects (Sudhahar and Cristianini, 2013), and analyze the

---

[1] See Labatut and Bost (2019) for a review.

643

structure of networks to test specific hypotheses (Elson et al., 2010; Agarwal et al., 2012; Coll Ardanuy and Sporleder, 2014; Piper et al., 2017). Our work builds on this tradition by introducing methods to reason about the phenomenon of propagation in fiction based on these constructed networks.

3 Methods

Our goal in this work is to investigate the behavior of information propagation in literary texts. In order to identify acts of propagation in this context, we need to determine the underlying network structure of a novel, including the nodes (by inferring characters) and the edges (by inferring some interaction between them). We describe first our pipeline for doing so, which involves identifying a set of unique characters from their mention in a text using coreference resolution (§3.1), attributing dialogue to those characters (§3.2), building a social network of speakers and listeners from that data (§3.3), and operationalizing a measure of "information" that we can treat as an atomic unit involved in propagation using slot-based information extraction (§3.4). With these constructed networks, we can measure acts of implicit propagation (§3.5) and explicit propagation (§3.6) within it.

3.1 Coreference resolution

Most contemporary systems for coreference resolution are trained on the benchmark OntoNotes dataset (Hovy et al., 2006), which primarily consists of news and conversation; literature is represented there only in the narrow genre of the Bible.

In order to use coreference resolution specifically trained on literature, we use the coreference annotations and trained model described in Bamman et al. (2020). This model is a neural coreference system inspired by Lee et al. (2017), augmented with BERT contextual representations (Devlin et al., 2019), and trained on 210,532 tokens in LitBank, comprising 100 different works of English-language fiction. Bamman et al. (2020) report its cross-validated average F-score on LitBank to be 68.1, notably higher than the performance for a model trained on OntoNotes (which has an average F1 score of 62.9).

3.2 Speaker attribution

Data. Previous work in literary speaker attribution has focused on a relatively small set of novels. Both He et al. (2013) and Muzny et al. (2017)

annotate Austen's *Pride and Prejudice* and *Emma* as well as Chekhov's *The Steppe*. Similarly, the Columbia Quoted Speech Corpus includes six texts by Austen, Dickens, Flaubert, Doyle and Chekhov. While these datasets have been able to drive much work in the development of models for speaker attribution, they represent a comparatively narrow slice of how dialogue is depicted in literature.

In order to evaluate the robustness of models across a diverse range of novels and authors, we annotate all 100 texts in LitBank (Bamman et al., 2019) with the boundaries for all true quotations and link each to the entity who spoke it. Here we are able to draw on the coreference annotations present in LitBank, which already link each mention to a unique entity. All annotations were carried out using the BRAT annotation interface (Stenetorp et al., 2012) by four annotators after a period of initial training, prompted to identify all quotations and attribute each one to the speaker who uttered it. Given the high agreement rate observed by Muzny et al. (2017) (κ of 0.97 for quote-speaker labels), each quotation is attributed by a single annotator. To check consistency, we double-annotate a sample of 10 texts (10% of the entire collection) at the end of the annotation process and find a similarly high inter-annotator agreement rate (Cohen's κ of 0.962). In total, 1765 quotations were annotated across all 100 works of fiction. This data is freely available under a Creative Commons ShareAlike 4.0 license at https://github.com/dbamman/litbank.

Quotation identification. For the task of quotation identification, we use the method implemented in BookNLP (Bamman et al., 2014), which uses simple regular expressions (text contained between an opening quote and a closing quote). On our gold annotations, this method results in an F1 score of 90.8 for quotation identification (87.1 precision and 95.0 recall). False positive failure cases of strings wrapped in quotation marks that do not constitute dialogue include various typographical uses of quotation for signifying other phenomena, including scare quotes for emphatic use (to introduce jargon, neologisms, or irony), titles of works of art, the mention of a term (as distinct from its use), and written use (see Brendel et al. (2011) for a survey). False negatives primarily arise due to regex matching errors (such as a stray quotation mark that results in an inversion of the subsequent speech and narration), or texts that do not delimit speech with

	B^3	MUC	CEAF$_{\phi_4}$	Average	Δ
Predicted coreference	68.0	84.9	61.0	71.3	–
–Trigram matching	67.7	84.5	60.1	70.8	-0.5
–Dependency parses	66.5	83.4	56.8	68.9	-2.4
–Singleton mention detection	67.0	85.9	59.6	70.8	-0.5
–Paragraph final mention linking	68.0	84.9	61.0	71.3	0.0
–Vocatives	68.9	85.9	62.1	72.3	+1.0
–Conversational pattern	66.8	85.0	58.9	70.3	-1.0
Oracle coreference	80.2	89.7	74.7	81.5	+10.2

Table 1: Metrics for cluster overlap between the gold set of clusters $\mathcal{G}$ and predicted set of clusters $\mathcal{C}$. Each cluster is defined as the set of quotations spoken by the same speaker. We also present the upper bound of carrying out speaker attribution using gold coreference labels (oracle coreference), which suggests that there is much to be gained in improving quotation attribution not only by improving coreference, but independently of it as well.

quotation marks at all (such as Joyce's *Ulysses*, which introduces direct speech with dashes).

Attribution. For speaker attribution, we reimplement the deterministic method of Muzny et al. (2017) using the full coreference information predicted above. Muzny et al. (2017) describes a series of deterministic sieves for the two tasks of a.) mapping quotes to the nearest speaker mention and b.) linking identified speaker mentions to character entities. The Quote→Mention phase includes sieves such as high-precision regular expressions for predefined Quote/Mention/Verb patterns (e.g., ["...,"]$_{QUOTE}$ [said]$_{VERB}$ [Jane]$_{MENTION}$), originally defined in Elson and McKeown (2010); dependency structure information (identifying mentions that hold an NSUBJ relation to a verb of communicating); and vocatives in the previous quotation. Quotations unattributed after running all sieves are assigned the majority speaker in the context.

To separate out the task of quotation identification from quotation attribution, we evaluate quotation attribution with gold quotation boundaries. While previous work on quotation attribution in literary texts, including Muzny et al. (2017) and He et al. (2013), evaluate system performance using classification accuracy and precision/recall (where each quotation in a test book is judged to be assigned to the correct true speaker from a predefined gold character list), we do not presume access to a gold character list during prediction time. Like Almeida et al. (2014), we evaluate performance using a measure of cluster overlap (here, the suite of metrics used in evaluating coreference resolution), where each cluster is defined by the set of quotations spoken by the same speaker.

As Table 1 illustrates, our reimplementation of Muzny et al. (2017) for the task of speaker attribution yields an average F1 score of 71.3 across the three cluster metrics when evaluated on all 100 books in our newly annotated data. As we ablate different aspects of the Muzny et al. (2017) model, performance generally degrades, attesting to the value of individual sieves.
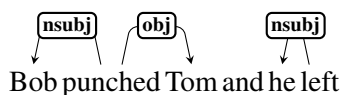
3.3 Identifying character networks

Similar to previous approaches for determining character networks in literary fiction (Elson and McKeown, 2010; Moretti, 2011), we use conversation as the basis for determining the edges in our graph. However, rather than trying to identify specific speaker-listener interactions, we instead extract dialogue blocks, drawing an edge between all characters mentioned outside of a quotation in a given block. Edges are weighted by the total number of dialogue blocks in which a given pair of characters are found to be co-present. We use a simple heuristic to identify these conversation blocks: if three or more contiguous sentences do not contain any quoted dialogue, we treat this as the termination of the block. The resulting graph serves as the basis for identifying information propagation in a given novel, as detailed in the following subsections.

3.4 Defining information

Whereas large-scale corpora such as social media data sets provide networks in which fuzzy matching of information may be appropriate and in which information repetition is often substantial (Leskovec et al., 2009), in the context of novels such methods are unlikely to have sufficient precision. As

a result, we select an information approach which allows us to maximize precision at the cost of potentially missing some instances of propagation. Our approach entails identifying quoted speech that references at least one character. One way to define this type of speech would be to simply describe it as *gossip*, though we feel that this is an overly narrow term given the general nature of our approach. Specifically, we select propositional tuples of the form (subject, verb, object), such that the subject holds an `nsubj` dependency relation to the verb and the object holds an `obj` relation (using the terminology of the Universal Dependencies (Nivre et al., 2016)); the subject and object may be filled by a character entity, a non-character nominal phrase, or a null token if neither is present. We ignore any tuples which contain *I*, *you*, or *we* (along with their variants) in either the subject or object slots, since they have comparatively higher errors in coreference. For the verb slot, we always select the lemma form of the proposition's head verb. Character entities in a proposition are identified by their unique character IDs established through coreference resolution (and not by the surface form of their mention).

Consider the following hypothetical example:

Bob punched Tom and he left

Given the operation of coreference resolution mapping "Bob" and "he" to the entity `Bob-id1` and "Tom" to `Tom-id2`, the extracted tuples for this sentence would be: [`Bob-id1`, punch, `Tom-id2`] and [`Bob-id1`, leave, ∅]. We extract all propositional tuples using a set of rules applied to the dependency parse of a given sentence. Although reductive to some degree, defining and extracting information in this way allows us to avoid informational noise and only select consistent propositional units.

To further reduce potential informational noise, we also only select tuples containing words that are likely to have some intrinsic interest to the plot and which have a relatively fixed meaning. After analyzing the 100 words that occur most frequently across all the tuples extracted from our corpus, we select tuples containing terms associated with the following four categories: *amorous*, *hostile*, *juridical*, and *vital*. For each category, we include the following words along with any synonyms that are also present in the top 100 tuple words: *amorous* (love, marriage), *hostile* (hurt, hit, shoot, kill), *juridical* (arrest, escape, innocent, guilty), and *vital* (alive, sick, dead). Since the Gutenberg corpus primarily contains nineteenth-century novels, these topics reflect many of the key events that these works of fiction tend to focalize.

3.5 Defining implicit propagation

We identify instances of implicit information propagation simply by determining whether a propositional tuple passes between a minimum of three characters. In other words, we look for an informational triad of the form character $A \rightarrow$ character $B \rightarrow$ character C, such that character A and character B are co-present when character A voices the initial instance of the proposition (but character C is not), and character B and character C are co-present when character B repeats the proposition during a different conversation block.

3.6 Defining explicit propagation

Along with implicit instances of information propagation, we note that novels often contain explicit propagation as well. We define explicit propagation as occurring when a character reports what another character said to a third character. In other words, we simply search for variations of the pattern "`[character-id]` said" in the context of quoted speech. Specifically, the variations considered include synonyms for "say" along with any arguments or modifiers that are relevant to introducing reported speech (e.g., "declared," "told me," "mentioned that," "claimed to," etc.).

The benefit to capturing instances of explicit propagation is that such instances can be extracted with very high precision regardless of the informational topic being discussed. Consequently, unlike for implicit propagation, we make no constraints on the nature of the information itself (in contrast to the four topics defined above). After identifying instances of explicit propagation, we incorporate coreference resolution and speaker attribution to determine the specific characters of a given propagating triad. Section 5.2 discusses how the resulting data from this approach can be used to analyze the role that gender plays in the depiction of information propagation within novels.

4 Experiments

Given instances of information propagation extracted from novels, we seek to understand the structural roles of the literary network and its topography that contribute to information passing between dyads. In particular, we seek to disentangle two possible alternatives:

H1. Information propagates through bridges who pass information between otherwise disconnected communities.

H2. Information propagates among densely connected strong ties (such as between members of the same family who interact frequently).

These alternatives correspond to different functions of gossip in literature, as theorized by Spacks (1985): while gossip primarily involves the "deliberate circulation of information," it also functions to reinforce existing relationships among strong ties (a point taken up in real-world social networks in Foster (2004)). We operationalize this distinction for understanding the dynamics of implicit propagation by describing information-propagating characters and non-information-propagating characters through six network measures that capture their topological properties within the network:

1. **Closeness centrality**: the average inverse distance between a given node and all other nodes in a graph.

2. **Betweenness centrality**: the fraction of shortest paths that pass through a node, summed over all node pairs.

3. **Average neighbor degree**: the average degree of the nodes in a given node's neighborhood.

4. **Effective size**: the measure of the non-redundancy between a node and its contacts—specifically how connected a node's contacts would be in its absence (i.e., the resulting structural hole).

5. **Efficiency**: the effective size of a node divided by its degree.

6. **Triangle count**: the number of triangles for which a node serves as a vertex, where a triangle is defined as a set of three nodes that are directly connected to each other.

We use the above measures to describe all nodes that either function as the B node in an $A \rightarrow B \rightarrow C$ information triad, or could function as such a node. More specifically, whenever we observe an instance of propagation $A \rightarrow B \rightarrow C$, in which at least one separate character B' was co-present with B when hearing A's information (but did not propagate it further), we select a pair comprised of B as a propagating node and B' as a non-propagating node (sampling B' from the set of co-present characters if more than one was present). In cases in which no non-propagating character was co-present, we instead sample a B' from the set of propagating instances that had multiple co-present characters. When sampling the non-propagating B' nodes, we only select characters that have been observed to speak at least once in the text based on our speaker attribution model (we hypothesize that selecting these characters is a better way to judge the efficacy of a propagation model, since they at least vocalize some information in the narrative, and hence are more likely to resemble propagating nodes in terms of their narrative functions).

4.1 Results

In order to test implicit information propagation in literature, we run tuple extraction on 5,345 works of fiction from the Project Gutenberg corpus. We find that roughly 3,600 of these books contain at least one instance of a repeated tuple containing a word from our four topics of interest (indicating the possibility of propagation based on our criteria). We proceed to run the rest of our pipeline on this subset of books.[2] In total, we find that 35% of these works contain at least one instance of implicit information propagation.

To distinguish between the two hypotheses outlined above, we scale all the features of the data between 0 and 1 and train a non-regularized logistic regression model to distinguish between information propagating B nodes and non-propagating B' nodes. We run the model on 1,730 B nodes and 1,730 B' nodes. The results are shown in Table 2 and discussed in more detail in the next section.

To ensure that our results are not simply caused by aspects of each network's general topology (irrespective of the unique qualities of propagating B nodes) we also run a degree-preserving randomiza-

---

[2]We process all books on a high-performance computing cluster using 24-core Intel Xeon Haswell processors and 64 GB of RAM; the average runtime of this pipeline on one book on this platform is five minutes and two seconds.

Graph Measure	Coefficient
Efficiency	3.0*
Effective size	2.7
Betweenness centrality	0.5
Closeness centrality	0.1
Triangles	−0.4
Average neighbor degree	−4.9*

Table 2: Logistic regression model coefficient values. Stronger positive values are indicative of information-propagating nodes; stronger negative values are indicative of non-propagating nodes. * denotes $p < 0.01$.

tion experiment (Miller and Hagberg, 2011) as a more stringent means for testing significance. For each network containing a propagating node, we generate 10 expected degree graphs and use them to calculate network measures for the corresponding propagating B and non-propagating B' nodes in the original network, producing a set of 10 randomized measures for each of the 1,730 original nodes in each class. We then randomly sample a single measure from each of these sets, yielding 1,730 randomized node measures for both classes, and re-run our logistic regression model on that resample. We repeat this process 10,000 times to generate an expected null distribution for each coefficient and assess the frequency with which a null coefficient value was observed to be as extreme as the value we observe under the true network—analogous to a p-value in a bootstrap hypothesis test (Efron, 1982; Berg-Kirkpatrick et al., 2012; Dror et al., 2018).

For the two node measures found to be significant under our original model, efficiency has a p-value of 0.08 (8% of 10,000 random trials observe a statistic as extreme as 3.0), no longer rising to the level of significance at $\alpha = 0.01$, while average neighbor degree has a p-value of 0 (no random trial sees as a statistic as extreme as −4.9), providing further evidence of its significance as a feature for discriminating information-propagating nodes.

5 Analysis

5.1 Implicit propagation and weak ties

As Table 2 shows, average neighbor degree and efficiency are both found to be significant at a threshold of $\alpha = 0.01$, while average neighbor degree is confirmed to be significant under a degree-preserving randomization experiment. These results support the first of our two postulated hypotheses (introduced in §4): information in novels

propagates through characters that serve as bridges between otherwise disconnected communities.

Average neighbor degree has the largest coefficient (by absolute value) and is negatively correlated with propagation. High values of average neighbor degree denote communities that are already well-connected (both to each other and to the rest of the network). In such a information-rich neighborhood, instances of propagation would be of less value or necessity, and hence would be less likely to be observed.

Support for the first hypothesis is further confirmed by the strong positive coefficient for efficiency. Like effective size, efficiency is a means of determining the extent to which a structural hole would occur if a specific node were removed from the network. Whereas effective size indicates the possibility of such a structural hole in general, efficiency measures how much each one of a node's connections on average contribute to linking otherwise disconnected neighborhoods. Thus high efficiency suggests that a node is not only serving as a useful bridge between other nodes, but that it is doing so productively relative to its total number of connections.

In a sense, these results suggest that we are observing a version of the weak tie theory first proposed by Granovetter (1973). By virtue of the fact that a character's connections are not themselves closely connected, that character can in turn serve an essential informational role for the community.

5.2 Explicit propagation and gender

While our methods for extracting implicit propagation for amorous, hostile, juridical and vital events identified 1,730 instances in 5,345 novels, our method for identifying explicit propagation yields far more—93,948 instances of propagation involving 258,619 triads (since there may be multiple listeners for a single instance). Although the analysis carried out on implicit propagation is not possible for the explicit case (since there is no way to identify co-present B' nodes when the initial instance of a proposition remains unobserved in the text), the size of the explicit results are conducive to other analyses. Specifically, we consider here the role that gender plays in the depiction of propagation. As Spacks (1985) points out, women are often stereotyped (both within the real world and in representations in literature) as more likely to engage in gossip; from a networked perspective, they are also

often cast as intermediaries between men, "serving as points through which to triangulate male-to-male desire or power" (Selisker, 2015). Analyzing gender (and other demographic attributes) in the context of information propagation enables scholars to consider how authors construct the informational ecology of their novels given the functional roles played by different characters.

To measure the role that gender plays in how authors represent information propagation in novels, we calculate the relative proportion of different gender configurations for propagating triads compared to all triads present across our entire data set (we determine the gender of a character by counting up all the male and female nouns and pronouns in that character's coreference chain). This allows us to answer the question: given the overall structural opportunity to transmit information, how often does transmission actually occur based on gender?
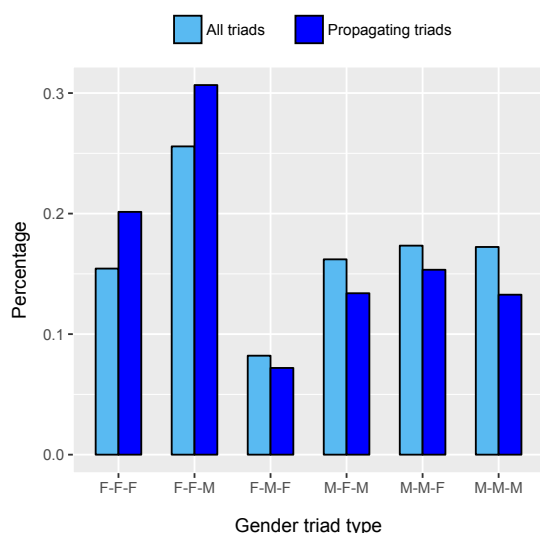

Figure 2: Comparison of the relative proportions of triad variations based on gender. All triads (light blue, $n = 158,250,238$) represent every observed triad across 5,345 books. Propagating triads (dark blue, $n = 258,619$) indicate only those triads observed to explicitly propagate information. The widest 95% confidence interval across all proportions is ± 0.0018, so that all differences within a gender triad type are significant.

Figure 2 illustrates the proportion of each gender configuration compared to the total; for instance, while 15.4% of all character triads are comprised of three women (F-F-F), 20.1% of triads involved in information propagation involve three women. Overall, we find that not only are female characters more likely to serve as propagators than male characters in this dataset, but that female characters fill

this role more frequently than one would expect given the proportion of female connector nodes across all triads. The proportion of propagation when the middle node is male, conversely, is lower than the expected value for every configuration. In other words, authors represent women as information propagators comparatively more frequently than men relative to their overall expectation.

Although literary criticism tends to envision the role of women in novels as being intermediaries between men (Woolf, 1929; Sedgwick, 1985; Schantz, 2008; Selisker, 2015), our analysis of information propagation tells a slightly different story. While women do indeed appear to serve as intermediaries/connectors more frequently than men do, women propagate information *between* men much more rarely than they do in other configurations (i.e., F-F-F, F-F-M). Though women may of course still connect men in these narratives, they do not appear to do so by notably passing on information. We leave determining the broader significance of this result to future work.

6   Conclusion

We introduce the task of identifying information propagation in literary social networks, designing an NLP pipeline for extracting both *implicit* and *explicit* propagation. This work offers a new perspective on the analysis of social networks in literary texts by considering the dynamics of how information flows through them—both as a result of the structural topology of the network (characters who successfully propagate are information bridges between communities), and as a result of the specific characteristics of each node (women are depicted more frequently as successful propagators than men).

This study, of course, contains limitations: readers of fictional works are only afforded a partial perspective of the world that is represented—namely the interactions an author chooses to describe (and not, for example, the dialogue we might presume takes place "off-screen"). Considered from a narratological perspective, however, this is a benefit rather than a drawback, since our goal is not to understand the underlying reality of these imagined worlds but rather how authors opt to represent the informational dynamics from which their stories are constructed. In developing this pipeline to examine how authors depict the transmission of information within nar-

rative texts, we hope to drive a variety of future research in this space, including not only such narratological questions as how "gossip impels plots" (Spacks, 1985), but also questions pertaining to issues of bias in representation, the flow of information, and factuality. Code to support this work can be found at https://github.com/mbwsims/literary-information-propagation.

Acknowledgments

References

Apoorv Agarwal, Augusto Corvalan, Jacob Jensen, and Owen Rambow. 2012. Social network analysis of alice in wonderland. In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, pages 88–96, Montréal, Canada. Association for Computational Linguistics.

Sam Alexander. 2019. Social network analysis and the scale of modernist fiction. *Modernism/modernity*, 3.

Mark Algee-Hewitt. 2017. Distributed character: Quantitative models of the English stage, 1550–1900. *New Literary History*, 48(4).

Mariana S. C. Almeida, Miguel B. Almeida, and André F. T. Martins. 2014. A joint model for quotation attribution and coreference resolution. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 39–48, Gothenburg, Sweden. Association for Computational Linguistics.

Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. 2012. The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, page 519–528, New York, NY, USA. Association for Computing Machinery.

David Bamman, Olivia Lewke, and Anya Mansoor. 2020. An annotated dataset of coreference in English literature. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 44–54, Marseille, France. European Language Resources Association.

David Bamman, Sejal Popat, and Sheng Shen. 2019. An annotated dataset of literary entities. 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 2138–2144, Minneapolis, Minnesota. Association for Computational Linguistics.

David Bamman, Ted Underwood, and Noah A. Smith. 2014. A Bayesian mixed effects model of literary character. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 370–379, Baltimore, Maryland. Association for Computational Linguistics.

Taylor Berg-Kirkpatrick, David Burkett, and Dan Klein. 2012. An empirical investigation of statistical significance in NLP. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 995–1005. Association for Computational Linguistics.

P. Bourdieu. 1996. *The Rules of Art: Genesis and Structure of the Literary Field*. Meridian (Stanford, Calif.). Stanford University Press.

Elke Brendel, Jörg Meibauer, and Markus Steinbach, editors. 2011. *Understanding Quotation*. De Gruyter.

Snigdha Chaturvedi, Mohit Iyyer, and Hal Daumé III. 2017. Unsupervised learning of evolving relationships between literary characters. In *Association for the Advancement of Artificial Intelligence*.

Mariona Coll Ardanuy and Caroline Sporleder. 2014. Structure-based clustering of novels. In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, pages 31–39, Gothenburg, Sweden. Association for Computational Linguistics.

Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. 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 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Rotem Dror, Gili Baumer, Segev Shlomov, and Roi Reichart. 2018. The hitchhiker's guide to testing statistical significance in natural language processing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1:*

Long Papers), pages 1383–1392, Melbourne, Australia. Association for Computational Linguistics.

Bradley Efron. 1982. *The jackknife, the bootstrap, and other resampling plans*, volume 38. Siam.

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 '10, pages 138–147, Stroudsburg, PA, USA. Association for Computational Linguistics.

David K. Elson and Kathleen R. McKeown. 2010. Automatic attribution of quoted speech in literary narrative. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, AAAI'10, pages 1013–1019. AAAI Press.

Eric K. Foster. 2004. Research on gossip: Taxonomy, methods, and future directions. *Review of General Psychology*, 8(2).

Adrien Friggeri, Lada Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor cascades. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Mark S Granovetter. 1973. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380.

Daniel Gruhl, Ramanathan Guha, David Liben-Nowell, and Andrew Tomkins. 2004. Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*, pages 491–501. ACM.

Hua He, Denilson Barbosa, and Grzegorz Kondrak. 2013. Identification of speakers in novels. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1312–1320, Sofia, Bulgaria. Association for Computational Linguistics.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: the 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, NAACL-Short '06, pages 57–60, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *North American Association for Computational Linguistics*.

F. Jannidis, I. Reger, M. Krug, L. Weimer, L. Macharowsky, and F. Puppe. 2016. Comparison of methods for the identification of main characters in German novels. In *Digital Humanities*.

Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *2013 IEEE 13th International Conference on Data Mining*, pages 1103–1108. IEEE.

Vincent Labatut and Xavier Bost. 2019. Extraction and analysis of fictional character networks: A survey. *CoRR*, abs/1907.02704.

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.

Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 188–197, Copenhagen, Denmark. Association for Computational Linguistics.

Jure Leskovec, Lars Backstrom, and Jon Kleinberg. 2009. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 497–506. ACM.

Jure Leskovec, Mary McGlohon, Christos Faloutsos, Natalie Glance, and Matthew Hurst. 2007. Patterns of cascading behavior in large blog graphs. In *Proceedings of the 2007 SIAM international conference on data mining*, pages 551–556. SIAM.

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

Aibek Makazhanov, Denilson Barbosa, and Grzegorz Kondrak. 2014. Extracting family relationship networks from novels. *CoRR*, abs/1405.0603.

Stacey Margolis. 2012. Network theory circa 1800: Charles Brockden Brown's "Arthur Mervyn". *Novel: A Forum on Fiction*, 45(3).

Nicholas Martin. 2014. Literature and gossip: An introduction. *Forum for Modern Language Studies*, 50(2).

Thomas J Mazanec. 2018. Networks of exchange poetry in late medieval China: Notes toward a dynamic history of Tang literature. *Journal of Chinese Literature and Culture*, 5(2):322–359.

Joel C. Miller and Aric Hagberg. 2011. Efficient generation of networks with given expected degrees. In *Algorithms and Models for the Web Graph*, pages 115–126, Berlin, Heidelberg. Springer Berlin Heidelberg.

Franco Moretti. 2011. *Network theory, plot analysis*. Stanford Literary Lab.

Grace Muzny, Michael Fang, Angel Chang, and Dan Jurafsky. 2017. A two-stage sieve approach for quote attribution. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, volume 1, pages 460–470.

Vlad Niculae, Caroline Suen, Justine Zhang, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Quotus: The structure of political media coverage as revealed by quoting patterns. In *WWW*.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1659–1666.

Andrew Piper, Mark Algee-Hewitt, Koustuv Sinha, Derek Ruths, and Hardik Vala. 2017. Studying literary characters and character networks. In *DH 2017*.

Ned Schantz. 2008. *Gossip, Letters, Phones: The Scandal of Female Networks in Film and Literature*. Oxford University Press.

Eve Sedgwick. 1985. *Between Men: English Literature and Male Homosocial Desire*. Columbia University Press.

Scott Selisker. 2015. The bechdel test and the social form of character networks. *New Literary History*, 46(3).

Richard Jean So and Hoyt Long. 2013. Network analysis and the sociology of modernism. *boundary 2*, 40(2):147–182.

Patricia Spacks. 1982. In praise of gossip. *The Hudson Review*, 35(1).

Patricia Spacks. 1985. *Gossip*. Knopf.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. BRAT: A web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 102–107, Stroudsburg, PA, USA. Association for Computational Linguistics.

Saatviga Sudhahar and Nello Cristianini. 2013. Automated analysis of narrative content for digital humanities. In *International Journal of Advanced Computer Science*.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.

John Wilkerson, David Smith, and Nicholas Stramp. 2015. Tracing the flow of policy ideas in legislatures: A text reuse approach. *American Journal of Political Science*, 59(4):943–956.

Virginia Woolf. 1929. *A Room of One's Own*. Harcourt Brace Jovanovich.