Infinite Jest: An Elegant Hairball

K. Hunter Wapman
Brian Lubars
Carl Mueller

1 INTRODUCTION

Does a story mirror the world, or show a different one? How does a story's structure impact its narrative?

We pursue an analysis of David Foster Wallace's *Infinite Jest* to answer variants of these two questions, viewing the book as a growing network:

(1) Does the story's social network mirror the world or show a different one?

(2) Why did the author choose to structure the narrative out of chronological order?

1.1 A mirror of the world?

Infinite Jest is of sufficient length to present a complex social network of characters and interactions between them. We make qualitative comparisons of its network of characters and their interactions with those one could expect to find in real-world social networks and random graphs.

Mediums for storytelling have evolved from an oral tradition to a range of genres with the written word and invention of the printing press.

1.2 How does structure impact narrative?

It's a non-statement to say that a novel's structure impacts its narrative, but to answer "how?" is very difficult. *Infinite Jest*, in particular, reveals the narrative world in pieces and leaves it as an exercise for the reader to stitch together the complete picture and maintain an updated view as the story progresses.

With this in mind, we may additionally examine the relationship between events that expose the story's world (expositions) building and driving the plot forward.

Does this choice have a purpose reflected in the network structure which might make the book more or less comprehensible to the reader, or is it purely stylistic?

TODO: cite papers on narrative structures?

In this work we seek to expand network analysis techniques for real-world social networks to examine artificial social networks in narratives.

Specifically, we extract and analyze a network of character interactions from *Infinite Jest*, using character co-occurances to produce a growing weighted undirected graph where higher edge weights indicating more interactions.

We compare this narrative social network to properties exhibited by real-world networks, including a heavy-tailed degree distribution, the small-world property, inverse degree assortativity, and community structure.

We find that centrality measures such as degree centrality and betweenness accurately identify the main characters, and comment on gender bias in the network.

Finally, we examine the narrative structure of *Infinite Jest* by comparing the dynamics of the growing network in authored book-order versus chronological order.

2 BACKGROUND

2.1 Infinite Jest Book

CHRONOLOGY

Sierpinski Gasket

âĂIJCertain kind of parallel lines are supposed to start converging in such a way that an "end" can be projected by the reader somewhere beyond the right frame. If no such convergence or projection occured to you, then the book's failed for you.âĂİ âĂŤ DFW

Endnotes

In *Infinite Jest*, the story is split into sections which are ordered non-chronologically.

The first few sections in the book are actually the last to occur chronologically!

2.2 Why a Network?

There are many possible definitions of an edge in a social network: dialogue, physical interactions, sentiment. We chose to present character co-mentions in the text as edges. This gives us the broadest view of a book as a network of character interactions, and additionally eases the burden of natural-language parsing because we can disregard all semantic information in the interactions.

2.3 Questions..questions that need answering

Would the book be more understandable if told in chronological order?

How is the narrative world structured? (static-network analysis) Main characters: centrality measures Small-world, community detection, assortativity

How does the structure evolve throughout the narrative? (dynamic analysis) Chronological order vs. book order Attachment: sparsification/densification? Main character centralities over time

Why is the book structured this way? Chronological vs. non-chronological How does the structure impact the narrative?

3 RELATED WORK

A quick search shows that we are not the first to mine networks from literary works and apply network analysis techniques.

In fact, there are a few interesting works which use data-driven analysis to evaluate hypotheses in literary analysis. Though it

doesn't use networks, Reagan et al.'s uses NLP and sentiment analysis to confirm Kurt Vonnegut's rejected Master's thesus on emotional story arcs [6]. Analyzing 1,327 stories from Project Gutenberg, they find six core emotional trajectory arcs and compare modern popularity of stories for each arc evaluated by number of downloads.

Several social networks have been built from narratives. Alberich et al. (2002) build bipartite social network out of Marvel comic book character appearances in comic books, finding many similar characteristics to real-life scientific collaboration networks such as short distances and a power-law degree distribution tail with a cutoff [1]. Notably, they find a lower clustering coefficient than expected in real networks.

Bonato et al. [3] compared networks for three books: *Twilight* by Stephanie Meyer, *The Stand* by Steven King, and J.K. Rowling's *Harry Potter and the Goblet of Fire.* They found the Chung-Lu model best fits the co-occurance networks. Game of Thrones and Lord of the Rings have also been similarly analyzed [2, 7].

Infinite Jest differs from the other narratives touched on in prior works because of its complexity, supposedly wide and in-depth wordview, and its unusual (non-chronological) structure. We therefore hope to find how this artificial social network compares to real social networks and to those previously analyzed. We suspect we may see statistics more similar to real-world social networks.

4 METHODS

4.1 Data Preprocessing

4.1.1 Data Source & Preprocessing. To generate analyzable text this project utilizes an existing '.mobi' file of *Infinite Jest*. This file is put through a .mobi to HTML conversion to generate an HTML formatted file as well as a .mobi to raw text conversion. This enables the use of regular expressions to find and parse various text features automatically. Thus the HTML text is used to identify endnote locations such that simplified endnote tags can be inserted into the raw text file. Similarly, regular expressions are used to indentify where section breaks exist according to the section annotations given by the book Elegant Complexity [4]. The raw text is split on these sections and saved into separate files. Likewise each endnote is saved into a separate file. In each of these files, we remove all inessential special characters (e.g. special quotes). As we employ the Python 3.7 standard, all text files are imported as unicode.

4.1.2 Named Entitiy Recognition. A major challenge with *Infinite Jest* is the abundant use of pseudonyms and aliases of the 200+ characters in the novel. As our network uses characters as nodes, identifying named entities and their synonym coreference resolutions (not for pronouns) requires an extensive hand-engineered approach. We utilize the Named Entity Recognition (NER) parser of the Python library SpaCy [5] augmented with its own Matcher parser. By running the existing NER model on the text, we compare candidate entities generated from the NER parser with our own indetified character matches. Manually parsing over these results enables us to build an entity-synonym list that feeds into the Matcher in subsequent NER parses. This enables Spacy to indentify a large number of the pseudonyms and aliases and their locations within each section of the text.

CREATING GRAPHS examples

4.1.3 Challenges and Considerations. One major affront to the Matcher approach is the extensive use of pronoun coreferences. Similarly, the extensive use of dialog presents a challenge in identifying named entities since the the identification of the speaking character is often hidden. For example, in the following quote, two characters never explicitly mention each other but are refering to the same character, Bob.

Problems: Point of View: who is âĂĲIâĂİ? Reference vs Interaction False positives Disambiguation: Which John is John? Pseudonyms âĂĲMadame PsychosisâĂİ is a pseudonym of Joelle van Dyne Hypergraph: âĂĲone of the Vaught sistersâĂİ

> "*I* am really concerned about *Bob*"
> "*I* am concerned about *him* too."
> "What should *we* do about *his* problem?"

This exchange would be overlooked by our approach. While there are some approaches for coreference resolution at this granularity, most are trained on models not well suited for the unstructured and informal style of David Foster Wallace's novel. As such, the use of named entities as nodes, and our methodology for identifying where they exist in text, enables the generation of a co-mention network. However, such a network may not perfectly capture true latent interaction structure of the book.

4.2 Network Design

4.2.1 Nodes and Edges. Network nodes constitute each character identified in the set of found named entities. These were referenced against online resources to ensure proper coverage of the characters in the book. Edges in the book represent a co-mention between two entities in the text. A threshold number of tokens (words) under which the number of tokens between the mention of one entity and another determines if an edge is established. If the edge already exists, the weight is updated. The current entity i is only matched with proceeding entities j within this threshold. Once no match is found, the next available entity is checked for proceeding matches.

The threshold number of tokens is determined by a semi-objective measure of the effect of the threshold length on the average clustering coefficient and the giant component size (see Figure 1). The intuition behind the use of these metrics is that we choose the minimum threhold length required to produce a large giant component and ample enough clustering best capturing the highly connected nature of characters in the novel. Our chosen threshold is 50 tokens.

5 STATIC-NETWORK ANALYSIS

5.1 Statistics

5.1.1 Degree Distribution. Average Degree
Power law?

$$P(k) \sim k^{-\tau}$$
$$P(k) \sim k^{-\tau} 10^{-k/c}$$

TODO: 'logarithmically bin data and perform a linear regression of Log(P(5)) on log(r) to get the power law fit

5.1.2 Assortativity.

5.1.3 Small World. Comparison of diameter to real social networks?

(A) Clustering Coefficient vs. Threshold Size

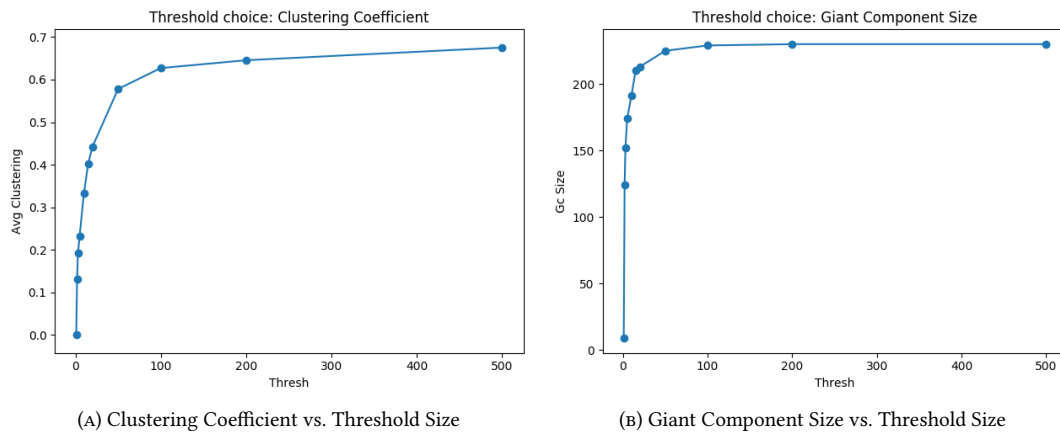(B) Giant Component Size vs. Threshold Size

FIGURE 1: Comparing threshold size effect on clustering coefficient and giant component size in order to determine the ideal threshold size.

5.2 Modularity

- talk about NMI - talk about modularity

5.2.1 Clustering. - compare against config model expectations?
look at triangles, how many are closed/open, compare to expectations from a normal social network?

5.3 Gender

6 DYNAMIC ANALYSIS

6.1 Narrative

6.2 Sparsification vs Densification

7 FUTURE WORK

Given more time and resources, there are several interesting questions to explore. We only ran our analysis with a single book, but it would be interesting to compare character interaction networks within and between different genres. For example, given a few dozen books, we could start to determine how *Infinite Jest*'s character network structure compare to other postmodern works, or how postmodern works differ from others like mysteries, historical fiction, fantasy, or short stories.

It would also be interesting to note if we can structurally identify common narrative techniques or plot devices – for example, the climax, backstories, cliffhangers, plot twists, or stories-within-stories.

8 CONCLUSION

REFERENCES

[1] R. Alberich, J. Miro-Julia, and F. Rossello. 2002. Marvel Universe looks almost like a real social network. *arXiv e-prints*, Article cond-mat/0202174 (Feb. 2002), cond-mat/0202174 pages. arXiv:cond-mat.dis-nn/cond-mat/0202174

[2] Andrew Beveridge and Jie Shan. 2016. Network of Thrones. *Match Horizons* (2016). https://www.maa.org/sites/default/files/pdf/Mathhorizons/NetworkofThrones%20%281%29.pdf

[3] Anthony Bonato, David Ryan D'Angelo, David F. Elenberg, Ethan R.and Gleich, and Yangyang Hou. 2016. Mining and modeling character networks. *CoRR* abs/1608.00646 (2016). arXiv:1608.00646 http://arxiv.org/abs/1608.00646

[4] Greg Carlisle. 2007. *Elegant complexity: a study of David Foster Wallaces Infinite jest.* Sideshow Media Group.

[5] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To appear* (2017).

[6] 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. *CoRR* abs/1606.07772 (2016). arXiv:1606.07772 http://arxiv.org/abs/1606.07772

[7] Mauricio Aparecido Ribeiro, Roberto Antonio Vosgerau, Maria Larissa Pereira Andruchiw, and Sandro Ely de Souza Pinto. 2016. The complex social network of The Lord of Rings. *Revista Brasileira de Ensino de Física* 38, 1 (2016).