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 analyze David Foster Wallace's *Infinite Jest* as a network in an attempt to answer variants of these two questions:

(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?

To answer these questions, we extract a network of character interactions that grows as the book progresses. We pursue our analyses in two contexts:

(1) **Statically**: analyzing the state of the book's network at the time of its end.
(2) **Dynamically**: analyzing the state of the book's network as it grows throughout the telling.

1.1 A mirror of the world?

There is a complex social network of characters and their interactions in *Infinite Jest*. 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.

We compare this narrative social network to properties exhibited by real-world networks, examining its degree distribution, inverse degree assortativity, and community structure.

We find that centrality measures – degree centrality and betweenness specifically – accurately identify important characters.

We also explore gender bias in the network, an aspect of the book that has received significant criticism.

1.2 How does structure impact narrative?

In *Infinite Jest* the world is revealed only in pieces and indirectly, and it is up to the reader to stitch those pieces together to create complete picture and to maintain a coherent view of the story as it progresses.

It's a non-statement to say that a novel's structure impacts its narrative, but it is far from so easy to answer the question "how?" In an effort to answer this question quantitatively, we ask: "does the decision to tell the story out of chronological order make it more or less comprehensible for the reader?"

Our examination of the book's chronology leads us to a new understanding of the kinds of functions events in the book take on and encourages a distinction between events that expose the world (functioning as exposition), and those that drive the narrative forward.

Network Analysis and Design, December, 2018, Boulder, CO, USA
2018.

2 BACKGROUND

"Certain 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." – David Foster Wallace [1]

2.1 The structure of *Infinite Jest*

Infinite Jest's structure differs from that of a typical book in a number of ways.

2.1.1 Chronology. The story is made up of 192 sections which are ordered non-chronologically; as seen in figure 1, the first sections to occur in the book appear last chronologically.

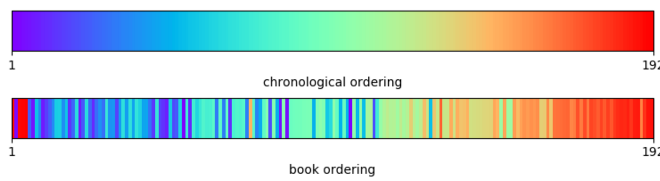


Figure 1: Contrasting the ordering of events per the book with the chronology presented in [5]

2.1.2 Sierpinski Gasket. The book's structure is ostensibly based on a Sierpinski Gasket, a triangular fractal.

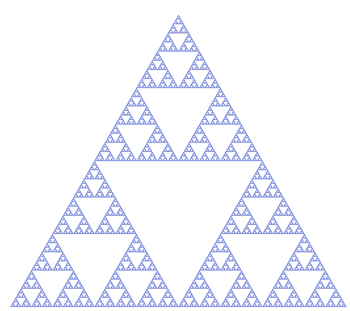


Figure 2: Sierpinski Gasket

2.1.3 Endnotes. The book uses endnotes extensively to convey additional information of varying degrees of relevance. The gamut of information can run from multi-page mini-stories containing important narrative information (*Infinite Jest*, endnote 304) to "No clue." (*Infinite Jest*, endnote 216). Endnotes can reference other endnotes (*Infinite Jest*, endnote 45), and can themselves have endnotes (*Infinite Jest*, endnotes 388). These present a challenge both for the reader and for our approach to generating the network.

2.2 Related Work

We are not the first to mine networks from literature or to apply data-driven techniques to evaluate a hypothesis.

Reagan et al. use NLP and sentiment to analyze 1,327 stories from Project Gutenberg in an effort to confirm Kurt Vonnegut's rejected Master's thesis categorizing story arcs into six types by emotion; they indeed find six core emotional trajectories, and they use the number of downloads of stories with each archetypal arc-type to evaluate each type's popularity [8].

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 [2]. Notably, they find a lower clustering coefficient than expected in real networks.

Bonato et al. [4] 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 [3, 9].

Infinite Jest differs from the narratives explored in prior works due its unorthodox narrative structure, the sheer number of characters, and the stylistic diversity of its prose.

3 *INFINITE JEST* AS 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.

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 [5]. 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 [6] 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?"

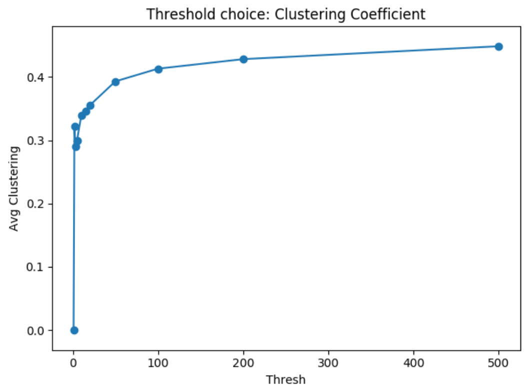
This exchange would be overlooked by our approach. While there are some approaches for coreference resolution at this granularity, most are trained on models not well suited for the unstructured and informal style of David Foster Wallace's novel. As such, the use of named entities as nodes, and our methodology 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

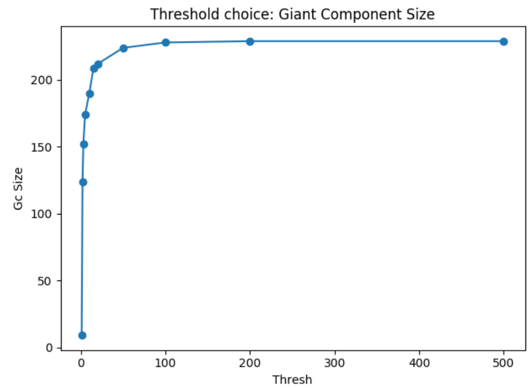
4.2.1 Nodes and Edges. Network nodes constitute each character identified in the set of found named entities. These were referenced against online resources to ensure proper coverage of the characters in the book. Edges in the book represent a co-mention between two entities in the text. A threshold 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 by adding '1' to the current weight value. 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 3). 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.

4.2.2 Section Snapshots and Modeling Recency. One of the major considerations of our analyses concerns the ordering of sections. The book does not follow a chornological ordering with contiguous secitons potentially taking place anywhere within the time range of the book. However, the book Elegant Complexity [5] provides



(A) Clustering Coefficient vs. Threshold Size



(B) Giant Component Size vs. Threshold Size

FIGURE 3: Comparing threshold size effect on clustering coefficient and giant component size in order to determine the ideal threshold size.

a chronological ordering of sections. To perform a comparison of these two orders ('chronological' and ''booktime'), we constructed a timelapse of graph snapshots that represents the aggregate graph structure up to the current section in the given ordering.

Regardless of the ordering, the graph at the end of each sequence will be identical. To model the differences in exposure to narratives and characters between the two orderings, we perform a decay mechanism of the weights of the graph. At each section, the corresponding graph snapshot's weights are mapped through a decay function developed orginially by psychologist Hermann Ebinghaus dubbed a 'forgetting curve'. This curve is as follows://

$$R = e^{\frac{-t}{(10000*s)}}$$

Here R represents the memory retention. t represents the time, which we measure using the number of tokens per section. s represents memory stability. The exponent is scaled by $\frac{1}{10000}$ to ensure reasonable memory retention values between 0 and 1. This approach enables us build a rough model the decaying memory of a reader to the narrative and character exposure differences between 'chronological' and 'booktime' section orderings.

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. We see an unweighted degree assortative mixing coefficient of -0.0945. Degree assortative networks typically reflect core-periphery structures, where a dense core of highly-connected nodes is surrounded by successively less-dense periphery nodes. Degree disassortative networks, on the other hand, are more stars with high-degree nodes connected to low-degree. According to Newman in *Networks*, social networks are unusual in that they typically have a positive degree assortativity [7]. Therefore it is strange for us to see disassortative mixing by degree in *Infinite Jest*, possibly indicating more of a star-like structure or fewer community structures than real-world social networks.

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

5.2 Modularity

- talk about NMI - talk about modularity

5.2.1 Clustering. Using the transitivity definition of clustering coefficient, we can examine the fraction of closed triads.

$$C = \frac{(\text{number of triangles}) \cdot 3}{\text{number of connected triples}}$$

This metric disregards the edge weights, looking only at connections between characters. We find $C = 0.3930$, reflecting the relatively dense connections – perhaps within communities such as the Tennis Academy or Halfway House – as opposed to a tree-like structure rooted at the highest-degree (main) characters.

We can compare the clustering coefficient against the configuration model to determine if this effect is due to the degree sequence alone or perhaps reflects a conscious author choice. We find configuration model we get an average local clustering coefficient of 0.1728, vs 0.3930 in the book.

TODO: how does this compare to real social networks?

5.3 Gender

6 DYNAMIC ANALYSIS

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

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

6.1 Narrative

6.2 Sparsification vs Densification

7 CONCLUSION

8 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.

REFERENCES

[1] 1996. Live Online with David Foster Wallace. http://www.badgerinternet.com/~bobkat/jest11a.html

[2] 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

[3] 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

[4] 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

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

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

[7] M. E. J. Newman. 2010. *Networks: An Introduction.*

[8] 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

[9] 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).