Complex Adaptive Systems: Energy, Information and Intelligence: Papers from the 2011 AAAI Fall Symposium (FS-11-03)

# NEH Panel
# Simulating Plot: Towards a Generative Model of Narrative Structure

**Graham Alexander Sack**
English & Comparative Literature Dept., Columbia University
gas2117@columbia.edu

**Abstract**

This paper explores the application of computer simulation techniques to the fields of literary studies and narratology by developing a model for plot structure and characterization. Using a corpus of 19th Century British novels as a case study, the author begins with a descriptive quantitative analysis of character names, developing a set of stylized facts about the way narratives allocate attention to their characters. The author shows that narrative attention in many novels appears to follow a "long tail" distribution. The author then constructs an explanatory model in NetLogo, demonstrating that basic assumptions about plot structure are sufficient to generate output consistent with the real novels in the corpus.

## Introduction

Although computer-based analysis remains a minority pursuit in literary criticism, it has gained particular traction over the past 25 years within the subfields of stylistics and authorship attribution. Studies in this area generally utilize statistical analysis of word frequencies, punctuation, and other countable features to identify similarities and differences in authorial style (see "Burrow's Delta"[1]). The study that follows draws inspiration from this body of research by counting the frequency and co-occurrence of a generally ignored sub-class of common words: character names.[2] However, my approach and intentions differ in two crucial respects from previous studies.

First, rather than style, this paper is concerned with plot and characterization, two areas about which computational analysis has had comparatively little to say. As critic Franco Moretti has argued, plot is the crucial element that must be quantified if computational methods are to gain traction in mainstream literary criticism.[3] This paper is an effort to do so.

Second, the overwhelming majority of prior computational studies in literary criticism have been *descriptive*—counting and classifying the surface features of a text. This study, however, is focused on *generative models*. Although I make use of descriptive analysis, the intent is to motivate a computer simulation that I will show is sufficient to reproduce several key stylized facts about actual narratives.

This paper is divided into two parts:

Part 1 uses descriptive quantitative analysis to develop a set of stylized facts about plot and characterization based on a corpus of sixty 19th Century British novels.

Part 2 develops and reports the results from a computer simulation of narrative structure.

## Part 1: Descriptive Analysis

### The "Long Tail" in Narrative Attention

In *The One vs. The Many* (2003), literary critic Alex Woloch repositions the questions of plot and characterization with which narratologists and formalists have traditionally been concerned in terms of the concept



[1] Burrows, John. "Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship." Literary and Linguistic Computing. Vol. 17, No. 3, 2002. P. 267 – 287.

[2] Character names are often regarded as noise and excluded from authorship and stylistics analysis because they are not consistent across texts.

[3] Moretti, Franco. "Network Theory, Plot Analysis." NEH Network Analysis and the Humanities Conference. UCLA: August, 2010. Unpublished Conference Proceedings.


Figure 1: Distribution of Narrative Attention (Frequency of Name Mentions by Character) – *The Pickwick Papers*

Table 1: Goodness of Fit by Novel

| Title of Novel | Power Law | Exponential | Logarithmic | Linear |
|---|---|---|---|---|
| A Study in Scarlet | 0.796 | 0.896 | 0.815 | 0.576 |
| Adam Bede | 0.978 | 0.831 | 0.715 | 0.315 |
| Alton Locke | 0.923 | 0.723 | 0.385 | 0.144 |
| Annals of the Parish | 0.957 | 0.773 | 0.401 | 0.152 |
| Aurora Floyd | 0.964 | 0.912 | 0.809 | 0.449 |
| Barchester Towers | 0.963 | 0.878 | 0.786 | 0.395 |
| Belinda | 0.953 | 0.923 | 0.732 | 0.368 |
| Bleak House | 0.911 | 0.962 | 0.441 | 0.173 |
| Castle Rackrent | 0.930 | 0.982 | 0.928 | 0.704 |
| Daniel Deronda | 0.957 | 0.929 | 0.776 | 0.383 |
| David Copperfield | 0.966 | 0.889 | 0.281 | 0.079 |
| Deerbrook | 0.896 | 0.967 | 0.827 | 0.460 |
| Doctor Thorne | 0.946 | 0.927 | 0.724 | 0.357 |
| Dracula | 0.869 | 0.870 | 0.884 | 0.693 |
| East Lynne | 0.954 | 0.941 | 0.759 | 0.366 |
| Emma | 0.939 | 0.959 | 0.804 | 0.436 |
| Hard Cash | 0.990 | 0.835 | 0.524 | 0.181 |
| Henry Esmond | 0.958 | 0.900 | 0.473 | 0.188 |
| History of Pendennis | 0.991 | 0.830 | 0.523 | 0.170 |
| In the Year of Jubilee | 0.919 | 0.937 | 0.790 | 0.470 |
| Jack Sheppard | 0.954 | 0.924 | 0.772 | 0.383 |
| Jane Eyre | 0.963 | 0.811 | 0.316 | 0.098 |
| Jude the Obscure | 0.970 | 0.822 | 0.754 | 0.406 |
| Lady Audley's Secret | 0.947 | 0.937 | 0.711 | 0.365 |
| Little Dorrit | 0.866 | 0.987 | 0.816 | 0.460 |
| Mansfield Park | 0.933 | 0.954 | 0.820 | 0.460 |
| Mary Barton | 0.940 | 0.937 | 0.778 | 0.450 |
| Middlemarch | 0.963 | 0.865 | 0.787 | 0.372 |
| New Grub Street | 0.913 | 0.951 | 0.886 | 0.613 |
| North and South | 0.911 | 0.935 | 0.696 | 0.367 |
| Oliver Twist | 0.869 | 0.968 | 0.843 | 0.503 |
| Our Village | 0.888 | 0.664 | 0.575 | 0.291 |
| Paul Clifford | 0.949 | 0.900 | 0.850 | 0.495 |
| Persuasion | 0.865 | 0.987 | 0.939 | 0.654 |
| Phineas Finn | 0.894 | 0.965 | 0.765 | 0.400 |
| Pride and Prejudice | 0.898 | 0.985 | 0.893 | 0.575 |
| Sybil | 0.951 | 0.923 | 0.794 | 0.399 |
| Tess of the d'Ubervilles | 0.978 | 0.831 | 0.638 | 0.308 |
| Ambassadors | 0.876 | 0.960 | 0.859 | 0.615 |
| Bride of Lammermoor | 0.973 | 0.904 | 0.762 | 0.399 |
| Egoist | 0.846 | 0.986 | 0.947 | 0.684 |
| Heart of Mid-Lothian | 0.964 | 0.920 | 0.708 | 0.328 |
| Mill on the Floss | 0.952 | 0.938 | 0.784 | 0.404 |
| Moonstone | 0.880 | 0.980 | 0.940 | 0.649 |
| Richard Feverel | 0.941 | 0.954 | 0.800 | 0.434 |
| Pickwick Papers | 0.972 | 0.895 | 0.426 | 0.140 |
| Picture of Dorian Gray | 0.901 | 0.931 | 0.909 | 0.746 |
| Portrait of a Lady | 0.912 | 0.955 | 0.839 | 0.502 |
| Return of the Native | 0.927 | 0.897 | 0.826 | 0.526 |
| Sign of the Four | 0.988 | 0.861 | 0.772 | 0.440 |
| Jekyll and Hyde | 0.802 | 0.958 | 0.928 | 0.936 |
| Tenant of Wildfell Hall | 0.907 | 0.879 | 0.366 | 0.134 |
| Way We Live Now | 0.939 | 0.955 | 0.765 | 0.373 |
| Wings of the Dove | 0.876 | 0.990 | 0.936 | 0.680 |
| Woman in White | 0.961 | 0.939 | 0.802 | 0.441 |
| Tom Brown's School Days | 0.980 | 0.790 | 0.466 | 0.181 |
| Vanity Fair | 0.980 | 0.795 | 0.622 | 0.235 |
| Villette | 0.964 | 0.881 | 0.415 | 0.158 |
| Waverley | 0.979 | 0.884 | 0.662 | 0.296 |
| Wuthering Heights | 0.930 | 0.943 | 0.736 | 0.412 |
| **Average** | **0.931** | **0.907** | **0.721** | **0.406** |

of "narrative attention." Woloch announces his intention to…

> redefine literary characterization in terms of [a] distributional matrix: how the apportioning of attention to any specific individuals is intertwined with the narrative's continual apportioning of attention to different characters who jostle for limited space within the same fictive universe (Woloch, 13).

Woloch argues that "narrative attention" in novels (and, by extension, in narratives generally) is a scarce resource that authors must choose how to allocate amongst the characters populating their stories.

Taking a cue from *The One vs. The Many*, this paper begins by applying quantitative rigor to the concepts of "distribution" and "apportioning of narrative attention," terms that Woloch uses qualitatively.

By way of example, Figure 1 depicts the statistical distribution of character name mentions in Charles Dickens' *The Pickwick Papers*. The distribution of name mentions (an observable metric) can be used as an instrumental variable for the distribution of narrative attention (a latent, unobservable variable). The result is striking—109 characters organized into what one might term "the long tail": a small set of central characters represented by the spike on the left followed by a steep drop off to a long but shallow tail consisting of dozens of characters who are mentioned fewer than 10 times. Mr. Mallard, Mr. Price, Mr. Grundy, Bill—even a reader exceptionally well-versed in this novel is unlikely to recognize these names or remember the existence of these characters; and indeed, that seems to be the point. The characters at the far end of "the long tail"—which roughly correspond to what Woloch calls "minor minor characters" (Woloch, 116)—exist to be forgotten. The large volume of such characters is inseparable from the paucity of name mentions: readers experience them as a depersonalized mass rather than as individuals, as narrative scaffolding, on the border between character and landscape. Beyond the right edge of the distribution lie even deeper levels of obscurity and invisibility into which the characters in the "long tail," barely individuated as it is, are in constant danger of falling: anonymous groups of "choral" characters, unnamed strangers, unreferenced servants.

Table 1 shows the goodness for fit for power law, exponential, logarithmic, and linear curves against the character name distributions for sixty 19th Century British novels. The data shows that the distribution of narrative attention in most novels from the period approximates either a power law or exponential distribution, implying that the "long tail" is a common pattern in novelistic form.

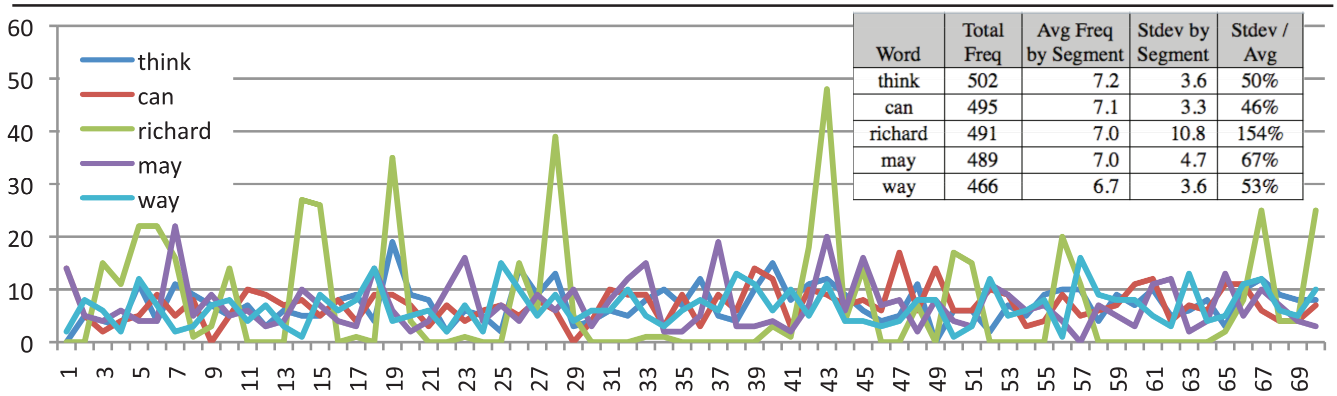
A wide range of phenomena are also known to follow a long tail: wealth distribution, website hits, and online books sales, for example, all obey a power law. The data for the novels sampled suggests that character name mentions and, by extension, narrative attention, are similarly distributed. That the distribution of attention within a novel should closely resemble the distribution of wealth within a nation is a provocative fact that calls for explanation.

One answer may be that the long tail in narrative attention is merely a special case of Zipf's law, which states that word frequencies in a large corpus follow a power law. Since character names are a subset of the words in a novel (accounting for ~2-4% of all word occurrences on average), it may seem intuitive that they too should follow a power law. But there are a few problems with this explanation.

First, although character name mentions in nearly all of novels in the sample follow a long tail, they do not all follow a power law: names in many novels lack the sharp peak typical of power laws and are better approximated by an exponential distribution (see Table 1).

Second, character names are not distributed across a text in the same way as other classes of words. The frequency of common vocabulary words is relatively consistent across all segments of a text: high frequency words like "of," "and," and "the" are high frequency everywhere. The prevalence of character names, on the other hand, varies

Figure 2: Word Frequency by 5000 Word Segment (*Bleak House*)




| Word | Total Freq | Avg Freq by Segment | Stdev by Segment | Stdev / Avg |
|---|---|---|---|---|
| think | 502 | 7.2 | 3.6 | 50% |
| can | 495 | 7.1 | 3.3 | 46% |
| richard | 491 | 7.0 | 10.8 | 154% |
| may | 489 | 7.0 | 4.7 | 67% |
| way | 466 | 6.7 | 3.6 | 53% |

substantially. For example, of the 250 most frequent words in Dickens' *Bleak House*, 19 are character names and 231 are common vocabulary words. If the text is divided into 5000 word segments, the frequency of the typical common vocabulary word varies from segment to segment with a normalized standard deviation of 60%. For character names, the standard deviation across segments is 214%. Figure 2 provides a clear picture of the difference: the most frequent name in *Bleak House* is "Richard" (a reference to the character Richard Carstone). "Richard" appears roughly the same number of times as the words "think," "can," "may," and "way," but it has 2-3 times the standard deviation. This difference reflects the fact that high frequency vocabulary words are determined by an author's style, which, at least for 19th Century novels, tends to be fairly consistent across a text, while character name prevalence is determined by the plot, which varies substantially. The distribution of attention in novels, then, is best approached by looking at how characters are instantiated on a scene-by-scene basis in the plot.

## How Narrative Attention Accumulates

To better understand the long tail distribution, it is helpful to do an inspection of the way narrative attention accumulates over the course of a novel. I begin by using a word frequency analysis program (the Intelligent Archive[4]) to divide each novel into 5000 word segments (ignoring chapter breaks) and I then count the number of times that each character is mentioned in each segment. The result is a set of time-varying "character prevalence vectors" that can be graphed to provide a visualization of plot and character development. I graph the name mentions for the top 25 characters in each novel on both (1) a segment-by-segment basis and (2) a cumulative basis.

Consider two representative cases: Jane Austen's *Pride and Prejudice* and Charles Dickens' *Bleak House*.

*Pride and Prejudice* provides a base-case for the way narrative attention accumulates over time in novels. As Figure 3(a) shows, Elizabeth Bennett dominates narrative attention in Pride and Prejudice: she is named ~800 times, twice that of the next most mentioned character. The remainder of the *dramatis personae* fall off in development gradually, with no sharp breaks or discontinuities. Figure 3(b) shows the attention paid to each character in each 5000 word segment of the novel. Elizabeth (represented by the dark blue line) is the dominant presence in almost every segment of the novel. The secondary cast is represented episodically by a succession of peaks: Jane (segment 2), Darcy and Charles Bingley (segment 3), etc. Narrative attention cycles through these secondary characters, returning to each every 2-4 segments to allocate a "peak." Two tiers of characters emerge: Elizabeth, the consistent, primary object of narrative attention and a secondary cast of 6-10 characters, who occupy background positions in the narrative with occasional moments of foregrounding. While Elizabeth is the source of narrative consistency, it is via this process of rotation through secondary figures that the novel generates a sense of plot development and variety. Figure 3(c) offers a cumulative perspective on this process. One notes the near-perfect linearity of Elizabeth's development and the relative straightness of all the other paths. The linearity of Elizabeth's path reflects the extreme consistency of the narrative attention devoted to her: Elizabeth's name is mentioned roughly 25-40 times in nearly every one of the novels 24 5,000 word segments so that her cumulative appearance by the nth segment is roughly n times her appearance in the first. Moreover, we note that the lines do not cross in the cumulative diagram. The relative position of each line indicates the corresponding character's ranking in terms of overall narrative importance. Elizabeth is 1st, Darcy is 2nd, Jane is the 3rd, etc. The fact that the lines do not (or rarely ever) cross means that these rankings never change. The structure of character development is static: the characters that are marked as narratively important in the first several chapters of the novel remain so throughout the remainder of the novel. Likewise, characters initially assigned to minority positions will never change their place in the narrative order of things. Narrative attention in entirely predictable: once a secondary character, always a secondary character.

Dickens' mid-Victorian multi-plot behemoth, *Bleak House*, provides a striking contrast. *Bleak House* consists of 69 5,000 word segments, features an enormous cast of characters (81 by my count), and mixes first and third person point of view. Looking at figure 4(b), we note the obvious differences from *Pride and Prejudice*: narrative attention is distributed as a dizzying series of disconnected, sharp peaks with no overarching source of consistency: characters appear for a segment or two and then step out of frame. The development of attention devoted to the primary characters in *Bleak House* proceeds in a manner analogous to that of the secondary characters in a single-plot novel such as *Pride and Prejudice*, that is, through an organizing logic of rotation. The novel cycles through its enormous cast characters episode by episode, developing them in fits and starts. The wavy, plateauing paths in figure 4(c) are symptomatic of this episodic pattern of development: a character receives a burst attention and then is ignored for a half dozen segments until there is another burst of attention. Moreover, there is a thorough confusion of narrative ranking, evidenced by the innumerable crossings and re-crossings of the narrative paths. The status of characters in *Bleak House* is constantly shifting as they are upgraded and downgraded in terms of

[4] Intelligent Archive (2010). [Computer software]. Hugh Craig, University of Newcastle, Australia.

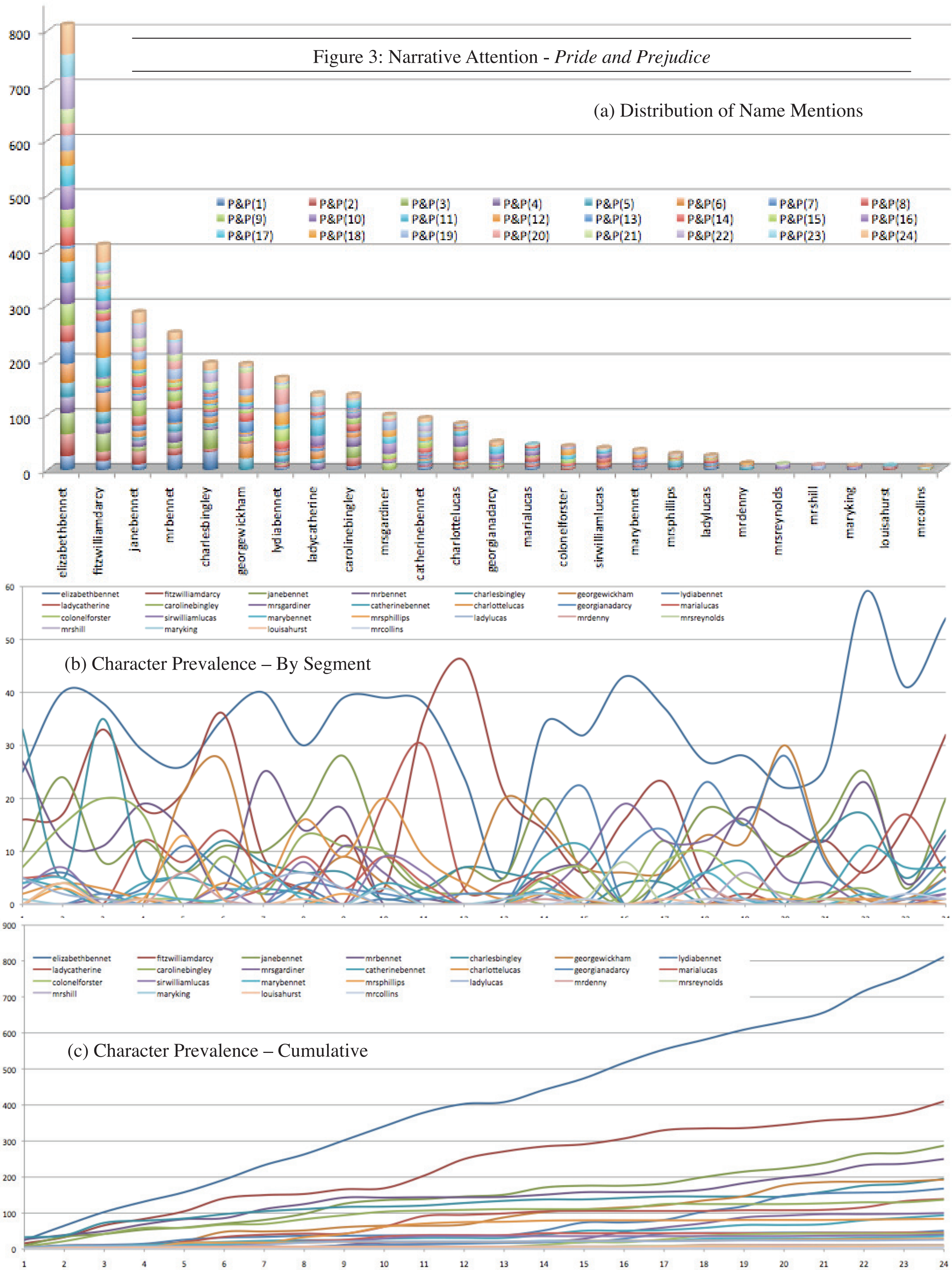
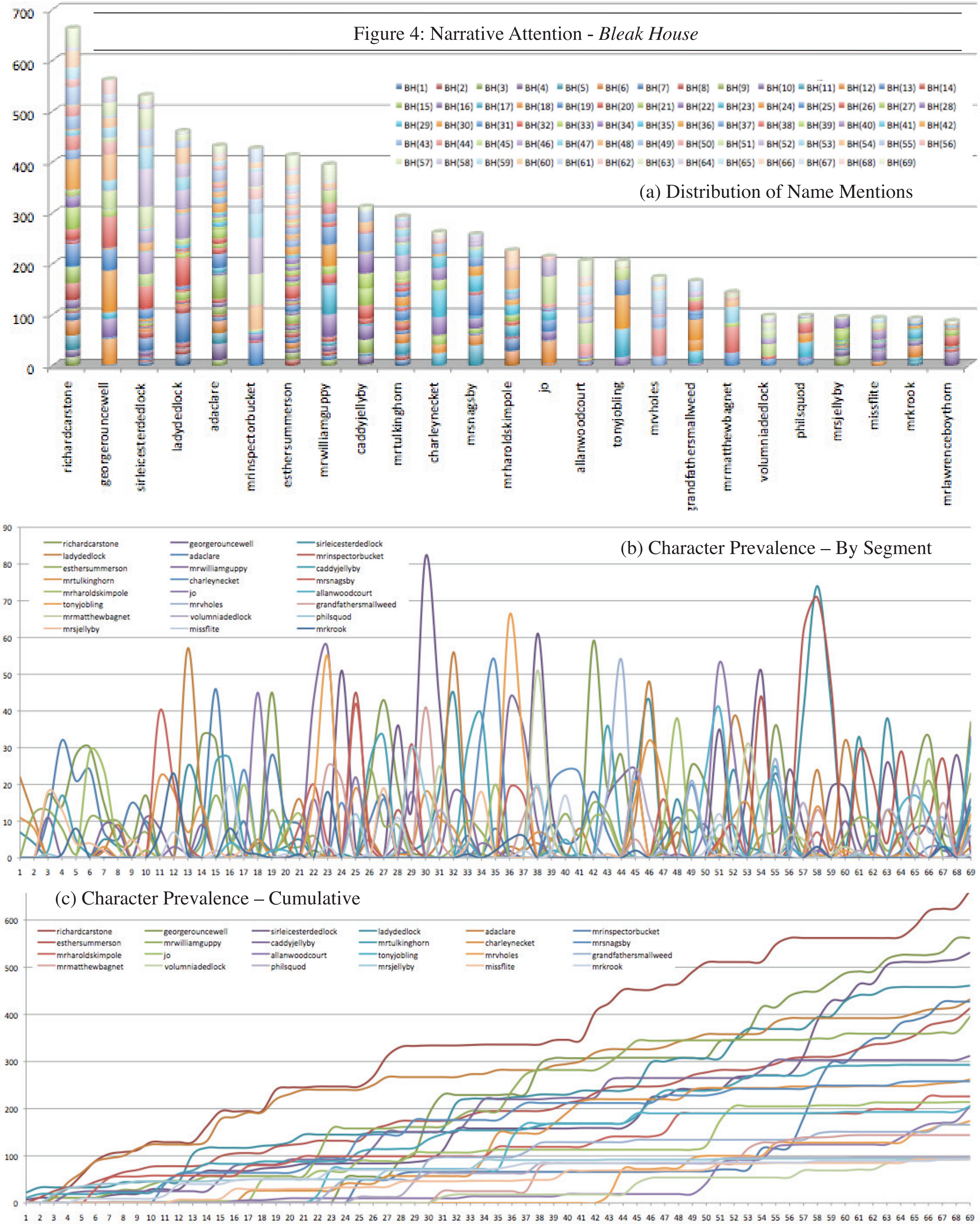



Figure 3: Narrative Attention - *Pride and Prejudice*

Figure 4: Narrative Attention - *Bleak House*



narrative importance: characters that appear in the background of narrative attention in one segment may step into the foreground in another. It is impossible to predict who the primary characters will be by the novel's end based on the allocation of attention at the novel's beginning. *Bleak House* likewise lacks a high peak, with name mentions dropping off very gradually. As Table 1 shows, it is best fit by an exponential distribution rather than a power law.

*Pride and Prejudice* and *Bleak House* represent two poles in the temporal dynamics of narrative attention—one adhering strictly to a logic of consistency and predictability and the other to a logic of variety and unpredictability. Most other novels fall between these poles and their graphs appear as linear combinations of the contrasting temporal processes represented.

## Part 2: Generative Models

### Simulating Narratives

Computer simulation techniques can play a valuable role in elucidating the dynamics behind narrative attention and plot described above. There are a number of potential approaches.

Characters in a narrative could be treated as independent agents in an agent-based model (ABM). Features of the narrative's structure, such as the distribution of narrative attention, would then be understood as an emergent property of rule-based character interactions. Character-agents might pursue pre-specified motives (e.g., to get married, to solve a murder); alternatively, Woloch's proposition that characters compete for scarce narrative attention could be represented by an objective function that characters seek to optimize. The dynamics of the system would be impacted by starting conditions related to a narrative's form and genre, such as the size of a novel's cast, character development conventions (e.g., whether minor characters are fixed in subordinate roles or may become the center of dramatic action in a subplot or parallel plot), and plot development protocols (e.g., linear vs. episodic plot structure, single vs. multiple plots, number and relation between subplots). Different starting conditions and rules of interaction would produce different distributions in narrative attention, which could be calibrated against actual novels to provide a better understanding of what parameters (character number, plot structure, etc.) drive structure. This would help literary critics and narratologists to situate extant authors, genres, and national and historical traditions within the range of narrative possibilities.

Such an approach treats narratives as self-organizing complex adaptive systems (CAS). One drawback of this approach is that it downplays the role of the author by making characters entirely self-directed. The "author," under this rubric, is present only in the starting conditions pre-specified by the choice of parameters: he is entirely non-interventionist. Although evidence certainly exists to support this version of authorship—Henry James, for example, speaks of the autonomy of his characters in the prefaces to *Roderick Hudson* and *The Portrait of a Lady*—this approach is at odds with the intuition most of us have that novels are meticulously crafted objects that undergo extensive revision; nor does this model seem adequate to describe narrative forms in which the consistency and believability of character behavior is sacrificed to other concerns, as in agit-prop political fiction.

A more realistic simulation that accounts for authorial intervention might model a narrative as the interaction between two levels of agency: an author-agent and a set of character-agents. Character-agents would pursue motives, while the author-agent would intervene to optimize an objective function related either to aesthetic criteria ("Is there sufficient conflict?"), narrative interest ("Is the plot too simple or too complex?"), or thematic content ("Does the narrative illustrate the desired themes?").

Yet another modeling approach is to use a system dynamics sensibility, eschewing character agency in favor of a structuralist approach that envisions narrative as composed of sub-structures with combinatorial rules. By way of illustration, it is this approach that I will focus on in the remainder of this paper. My central concern will be to construct an explanatory model of narrative structure using a few basic assumptions.

### Assumptions

I begin by assuming that a plot structure is composed of a set of interwoven "plot strands." For a concrete example one might think of the plot *structure* of a serialized novel such as *Bleak House* or a television series like *The Wire*. Such narratives generally have multiple plot strands (in TV parlance, referred to as an "A plot," "B plot," "C plot," etc.). Each plot strand is instantiated in scenes. A plot structure, then, consists of a particular realized sequence of scenes. For example, if there are three plot strands (A, B, C), one possible plot structure might be A, B, A', C, B', A" while another might be B, A, C, B', A'. With no combinatorial restrictions, the number of possible plot sequences is n^m, where n = # of plot strands and m = # of scenes. I further assume that plot strands interweave, alternating with one another such that the same strand cannot be instantiated in two consecutive scenes, which reduces the number of possible sequences to n*(n-1)^(m-1). I further require that each strand must be instantiated as a scene at least once in a plot structure. This requires m >= n, and reduces the number of possible plot sequences to ~[m!/(m-n)!]*(n-1)^(m-n).

I next assume that each strand has an internal hierarchy consisting of main characters, supporting characters, and incidental characters who appear in its scenes. These characters occupy different levels of importance to the plot and therefore receive varying levels of narrative attention. For modeling purposes, consistent with the data I have gathered for novels, I assume that "narrative attention" can be measured instrumentally by the number of times that a character's name is mentioned. For dramatic rather than narrative plot structures this instrument could be modified—for example, for a film or TV series one might measure screen time, while for a play one might measure the number of lines that a character speaks. Main characters are assumed to be the primary focus of a plot strand and therefore must appear in all of a strand's scenes and receive the greatest level of narrative attention. Supporting characters may or may not appear in any given scene and receive less attention than main characters. Incidental character may or may not appear in any given scene and receive less attention than either main or supporting characters.

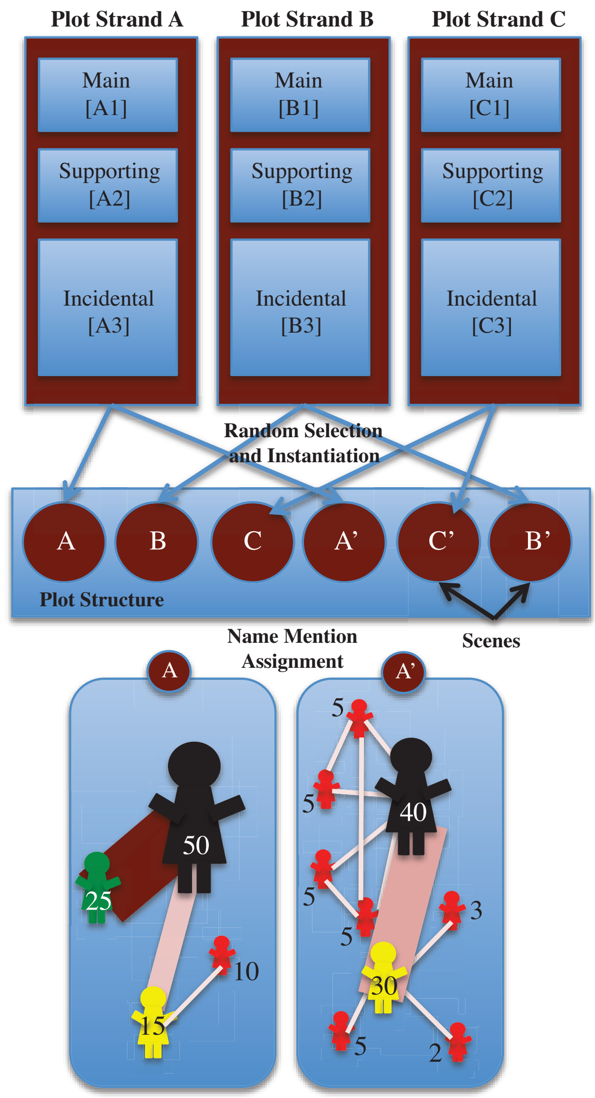
## Methodology

NetLogo was used to implement this model.[5] The user specifies the number of characters, plot strands, and scenes. At set-up, the model generates (1) a character hierarchy for each strand consisting of main, supporting, and incidental characters, and (2) a random plot sequence consistent with the combinatorial rules specified above. The model then progresses sequentially through the plot, instantiating each strand as a scene in the predetermined order. Each time a strand is instantiated as a scene, three things happen:

1. A list of characters is selected to appear in the scene from the strand's hierarchy.
2. A quantity of narrative attention (measured by name mentions) is allocated to each character. The total amount of attention available is fixed by chapter length and name prevalence, which are user specified. As a result, attention is a scarce resource and allocation is a zero-sum game, consistent with Woloch's thesis.
3. To represent character interactions, a weighted undirected link is formed between each pair of characters appearing in a scene. The link is weighted as a random overlap between the number of name mentions of each character it links.

The model is stochastic, with 5 sources of randomness: (1) the order in which characters are assigned to strands, (2) the order in which strands are called as scenes, (3) the set of characters called each time a strand is instantiated as a scene, (4) the number of name mentions assigned to each

[5] A JAVA applet of the model is available at http://www.columbia.edu/~gas2117. The source code is available by request (email: gas2117@columbia.edu)




character, and (5) the value of the edge weightings assigned to each character interaction.

The model generates output in several formats: (1) time-plots of the scene-by-scene and cumulative number of name mentions assigned to each character, (2) an overall distribution of narrative attention along with a measure of the fit of this distribution against power law and exponential functions, and (3) a social network diagram and network metrics describing the character interactions.

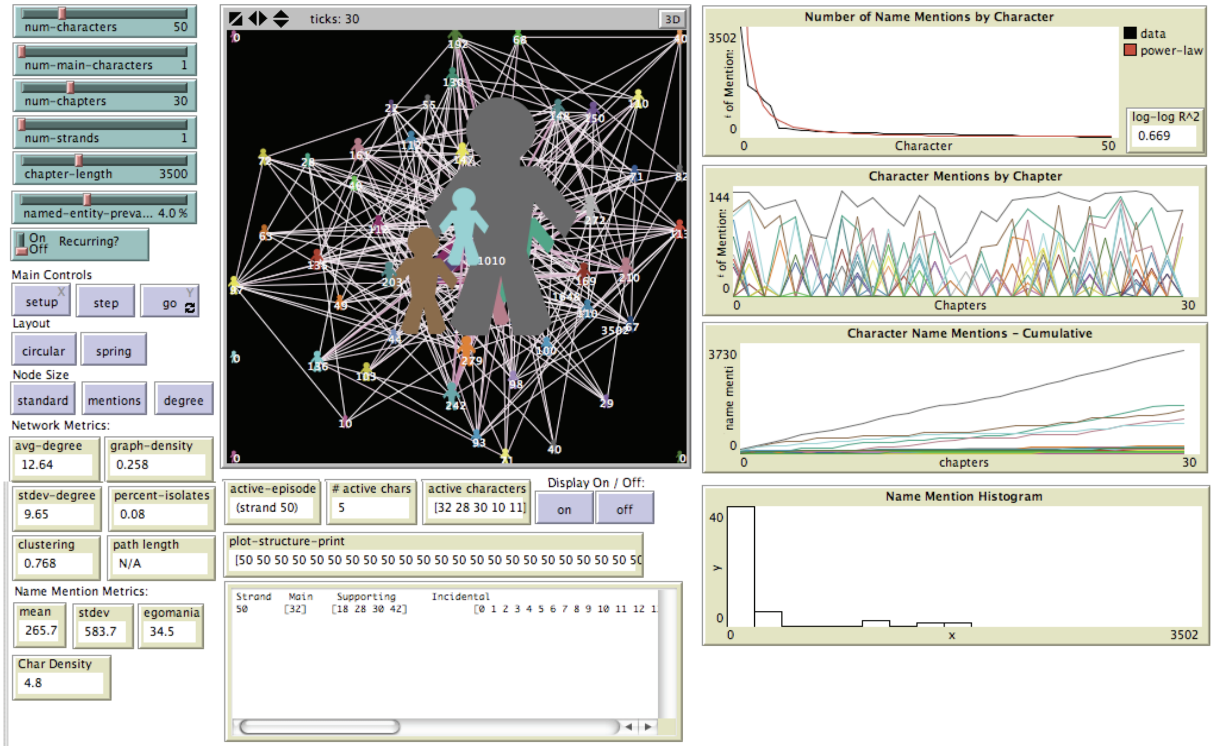
## Results

Although simplistic in its assumptions, this simulation is sufficient to reproduce a number of the salient features of narrative attention in the novels sampled.

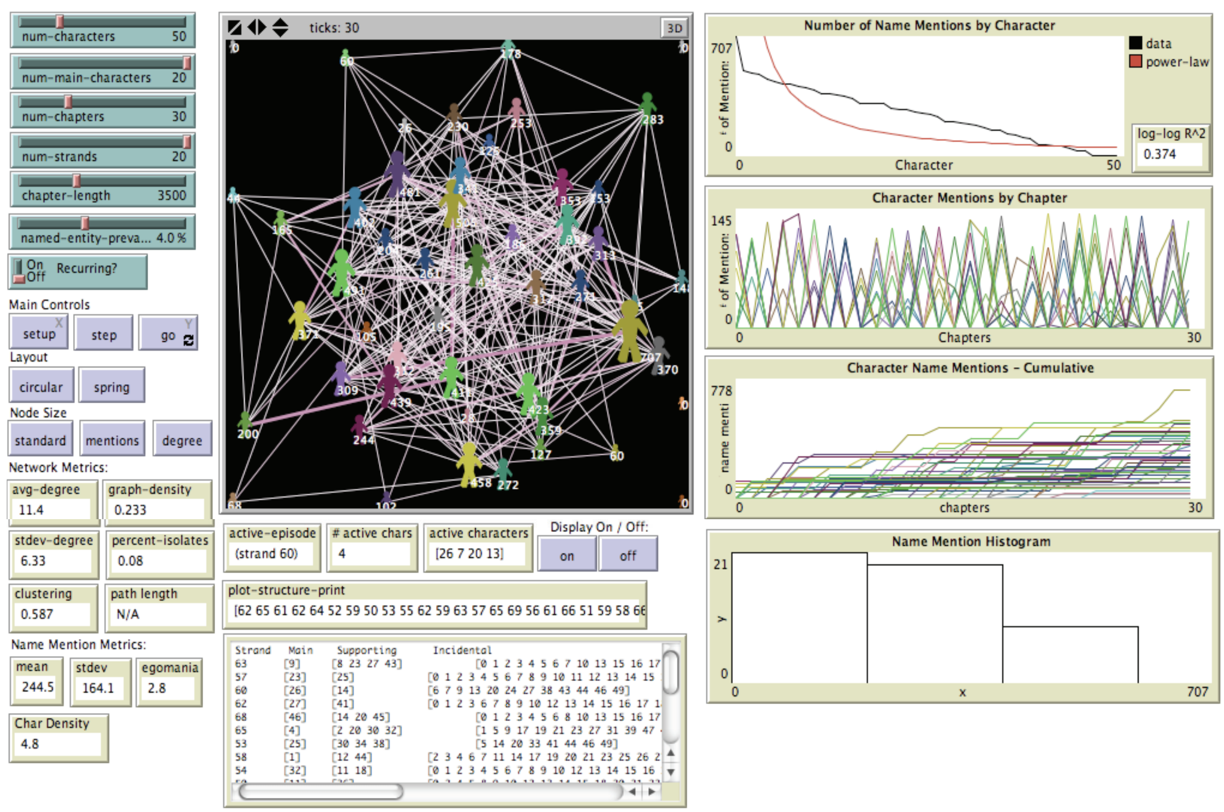
If the number of plot strands and main characters are set low—corresponding to a narrative that is tightly focused around one or a few characters in a single story line—the results closely resemble those observed for a *Bildungsroman* such as *Pride and Prejudice*. See figure

Figure 5: Examples of Model Output

(a) # of main characters = 1; # of plot strands = 1



(b) # of main characters = 20; # of plot strands = 20



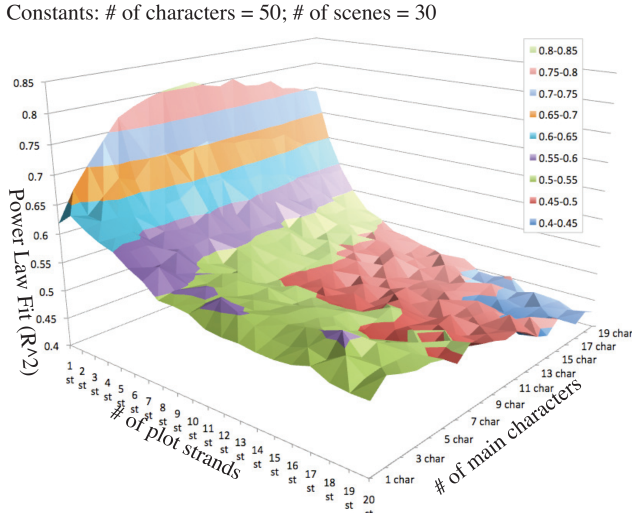
5(a). The cumulative diagram output by the model depicts discrete, non-intersecting trajectories similar to those we saw for Austen's now, reflecting consistency in character development and rigidity in the rankings of narrative importance. The distribution of attention across the characters fits a power law with a high R-squared.

If the number of plot strands and main characters are set high—corresponding to a narrative focused around a large ensemble of characters across many subplots or parallel plots—the results closely resemble those observed for a sweeping social problem novel such as *Bleak House*. See figure 5(b). The model reproduces the many-peakedness of the scene-by-scene diagram and the plateau or wave shape of the lines in the cumulative diagram (indicative of limited bursts of narrative attention rotating through the large cast) as well as the many crossings of the character development trajectories (indicative of the unpredictable / shifting status of the characters in terms of importance to the plot).

Figure 6 shows a sweep of the model's output in parameter space. The z-axis is the average goodness of fit of a power-law distribution. The x-axis represents the number of main characters (from x = 1 to x = 20) and the y-axis the number of plot strands (from y = 1 to y = 20). The number of characters is held constant at 50 and the number of scenes is held constant at 30. The model is run 40 times for each (x,y) pair, for a total of 16,000 runs. As the graph shows, the distribution of narrative attention fits a power law well for a low number of plot strands. As the number of plot strands increases, the fit erodes, particularly if the number of characters is increased along with the strands.

Figure 6: Parameter Sweep of Model Output

Constants: # of characters = 50; # of scenes = 30




## Conclusion

The simulation that I have developed is intentionally simplistic: I have modeled plot structure and characterization only in terms of combinatorial rules for plot strands. I have not attempted to give any internal sophistication to characters (such as motives), nor have I attempted to represent anything in terms of thematic or generic content. The "agents" in this model have little in the way of agency. Nevertheless, this simple model of plot structure is sufficient to generate results directionally consistent with the way narrative attention is allocated in actual novels. I have not shown that this assumption is necessary, merely that it is sufficient, and there are a number of other models that may be capable of generating similar results, such as the agent-based models of character interaction I outlined above. Other models may also have the advantage of explaining why particular individuals are assigned to main character positions within plot strands, an issue that is bracketed in this model with the assumption that characters have no internal attributes.

This has been intended as both a methods paper and a case study. The author hopes that it has offered an example of the way that simulations can empower computational literary criticism to move beyond the description of surface features to the testing of hypotheses about plot, character, and narrative structure.

## References


1. The texts of all novels used in this study were downloaded from Project Gutenberg. http://www.gutenberg.org/catalog/

2. Anderson, Chris. *The Long Tail.* Hyperion: New York, 2008.

3. N. Dames, D. Elson, K. McKeown, 2010. "Extracting Social Networks from Literary Fiction." Proceedings of 48th Annual Meeting of Assoc. for Computational Linguistics, Sweden.

4. D. Easley, J. Kleinberg. "Power Laws and Rich Get Richer Phenomena." *Networks, Crowds, and Markets.* Cambridge UP, 2010.

5. Garrett, Peter. *The Victorian Multiplot Novel.* New Haven: Yale UP, 1980.

6. Gass, William. "The Concept of Character in Fiction." *Fiction and the Figures of Life*. New York: Vintage Books, 1971.

7. James, Henry. Preface to *Roderick Hudson* and *The Portrait of a Lady*. *Henry James: Literary Criticism. European Writers and The Prefaces.* New York: The Library of America, 1984.

8. Leeming, Glenda. *Who's Who in Jane Austen and the Brontes.* New York: Taplinger, 1974.

9. Newlin, George. "A Taxonomy." *Everyone in Dickens.* Vol. III. Westport: Greenwood Press, 1995.

10. Wilensky, U. 1999. NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL. http://ccl.northwestern.edu/netlogo/.

11. Woloch, Alex. *The One vs. the Many*. Princeton: Princeton UP, 2003.