DHQ: Digital Humanities Quarterly

2014 Volume 8 Number 4

Visualizing and Analyzing the Hollywood Screenplay with ScripThreads

Eric Hoyt <ehoyt_at_wisc_dot_edu>, University of Wisconsin-Madison Kevin Ponto <kbponto_at_wisc_dot_edu>, University of Wisconsin-Madison Carrie Roy <croy_at_wisc_dot_edu>, University of Wisconsin-Madison

Abstract

Of all narrative textual forms, the motion picture screenplay may be the most perfectly predisposed for computational analysis. Screenplays contain capitalized character names, indented dialogue, and other formatting conventions that enable an algorithmic approach to analyzing and visualizing film narratives. In this article, the authors introduce their new tool, ScripThreads, which parses screenplays, outputs statistical values which can be analyzed, and offers four different types of visualization, each with its own utility. The visualizations represent character interactions across time as a single 3D or 2D graph. The authors model the utility of the tool for the close analysis of a single film (Lawrence Kasdan's *Grand Canyon* [1991]). They also model how the tool can be used for "distant reading" by identifying patterns of character presence across a dataset of 674 screenplays.

1. Introduction

In November 2009, the website xkcd.com published a series of info-graphics that visualized character interactions in movies such as the *Lord of the Rings* trilogy (2001-2003), the *Star Wars* trilogy (1977-1983), and *Jurassic Park* (1993). Exhibiting both attention to detail and a nice sense of humor, the xkcd charts allowed time to play out across the x-axis and showed how the characters exist in different spaces within the narrative world (the difference between storylines occurring on the Death Star and Tatooine, in the *Star Wars* example). These info-graphics became the motivation for Tanahasi and Ma's "Design Considerations for Optimizing Storyline Visualizations," a computer science paper published in *IEEE Transactions on Visualization and Computer Graphics* (2012). Tanahasi and Ma propose an algorithmic approach to generating the design of these types of visualizations. While these types of visualizations were of interest to scholars and researchers, they came with several short-comings, most notably that all of the information that was to be visualized needed to be gathered externally. This meant that a human would need to watch a film and manually fill in information as to when and where certain characters appeared and interacted. It greatly diminished the utility of these visualizations as a tool of data exploration; the analysis more or less needed to be completed before the tool could be used.

THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS. THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.



Figure 1. xkcd's 2009 charts of character interactions in movies proved to be popular on the Internet and influential to visualization designers.

The challenge of conveying complex information (character appearances, relations with other characters, and absences) as it changes over time — along with our shared interests in the Digital Humanities and our more singular scholarly focuses on cinema and television (Hoyt), visualization and storytelling (Roy), and computational approaches (Ponto) — spurred us to develop a tool that algorithmically analyzes and visualizes screenplays. In building a tool to aid Humanists, we sought to heed Abello, Broadwell, and Tanghrelini's call for computational approaches that "combine distant reading and close reading" [Abello, Broadwell, and Tanghrelini 2012]. One appeal of a fully algorithmic approach is that it scales nicely — enabling the "distant reading" that Franco Moretti first proposed and, in our case, allowing a researcher to study 1,000 screenplays rather than 10 or 20 [Moretti 2000]. Keeping in mind the value of close reading, though, we also sought to create a narrative analysis tool that offered direct access to every line of a screenplay — similar to the "Reading View" window in the Watching the Script software prototype developed for the visualization of theatrical text [Roberts-Smith et al. 2013]. In our case, we wanted to connect this "Reading View" window that displays lines from the screenplay to a series of visualizations that account for scenes, pacing, and character interactions. Rather than reducing a screenplay simply to statistical aggregates, we wanted to map the way a screenplay unfolds as it moves from page to page. From generating several hundred of these narrative profiles, we can compare and contrast large numbers of screenplays across decades, by author, by genre, by narrative structure and more.

2

3

4

In this article, we introduce our tool, ScripThreads, and discuss some of our initial research findings from using ScripThreads to analyze and visualize hundreds of screenplays from the American Film Scripts Online collection. We model how the tool can be productively used in film analysis as a tool for close reading by analyzing and comparing two screenplays co-written and directed by Lawrence Kasdan, *The Big Chill* (1983) and *Grand Canyon* (1991). We also model how the tool can be used for distant reading by searching across hundreds of screenplays for the pattern of the "hyper-present protagonist" — movies that place a main character in every scene or nearly every scene.

In the process of building the tool's prototype and writing this article, we have come to appreciate the many ways that a computer reads a screenplay differently from you or me. Humans gather insight from watching and experiencing the emotion, tension, and dynamics of a movie or screenplay. Rather than attempting to train a computer to understand a film in the exact same way we do, we would prefer to ask a computer to do tasks that it is designed to do well and that we humans struggle with. Humans have memory limitations when it comes to matters of sequential timing and the entrances and exits of dozens of characters. In contrast, computers are excellent at gathering and recording these sorts

of details from structured texts. Lev Manovich has suggested that one of the most valuable things that comes from combining computational analysis and the visualization of vast amounts of information in a single image is that it defamiliarizes our understanding of the works that we study in the Humanities [Manovich 2012]. As we hope to demonstrate, ScripThreads is a powerful framework for defamiliarization, provoking new questions, and producing new answers through the combined strengths of the human analyst and the computer.

2. Researching the Screenplay: Opportunities and Challenges

Of all narrative textual forms, the motion picture screenplay may be the most perfectly pre-disposed for computational analysis. As Murtagh, Ganz, and Reddington explain, "a filmscript is a semi-structured textual document in that it is subdivided into scenes and sometimes other structural units. Both character dialog and also descriptive and background metadata is provided in a filmscript. The metadata is partially formalized" [Murtagh, Ganz, and Reddington 2011]. As semi-structured documents with formatting conventions analogous to a metadata schema, screenplays are ideally suited for automated computer parsing. There is no need for laborious TEI encoding to detect character dialogue exchanges and interactions. When a character speaks, his or her name is capitalized and centered on the page. The character's dialogue generally appears one line below. As we discuss below, there are variations within this general format that create parsing challenges. Nevertheless, the extraction of character interactions in a screenplay is far easier to automate than in a novel or epic poem.

5

6

7

8

Our research and software development contributes to a lively and growing area of research about screenwriting history, form, theory, and practice. As a discipline, Film Studies has long lived in the shadow of the "auteur theory," which views the director as the author of a film. Film scholars and critics seem unable to let go of an "auteur desire," in the words of Dana Polan [Polan 2001]. Even scholars who acknowledge that film is a highly collaborative medium will refer to "Scorsese's *Taxi Driver*" — elevating the director to the status of author rather than the studio (Columbia Pictures), screenwriter (Paul Schrader), or cinematographer (Michael Chapman). Over the last decade, however, screenwriting studies has emerged as its own sub-field within the discipline of film and media studies. The Screenwriting Research Network held its first conference in 2008; the *Journal of Screenwriting* published its first issue in 2010.

A small contingent of screenwriting researchers are, like us, pursuing the computational analysis of screenplays [Marinov 2011] [Marinov and Stitts 2013a] [McKie 2014] [Murtagh, Ganz, and Reddington 2011]. The 2013 launch of Samuel Marinov and Brock Stitts' Screenplay Owl analysis tool marked an especially exciting development. With its emphasis on dialogic exchange frequencies, Screenplay Owl differs significantly from ScripThreads. However, both ScripThreads and Screenplay Owl show that the initial speculations about computational screenplay analysis are becoming realities. As we worked on revising this article for DHQ in early-2014, we witnessed the launch of another promising screenwriting analytics tool: ScriptFAQ, developed by Stewart McKie. McKie's tool is designed for practicing screenwriters who want to ask questions about a screenplay they are working on (for example, how many scenes is one character in compared to another?). For ScriptFAQ to answer these questions, the screenplay must be imported in the XML-based Final Draft file format and the user must enter a substantial number of additional metadata fields, which require detailed knowledge of the screenplay. For a writer working on a screenplay, these metadata fields are quite easy to complete and the results are well worth the effort. For a researcher wanting to quickly compare hundreds of screenplays across history, however, the need for many additional metadata, prior knowledge about each screenplay, and the Final Draft file format (which most older screenplays aren't readily available in) make ScripThreads better suited than ScriptFAQ for distant reading. Ultimately, for a researcher interested in computer-enhanced close reading, we believe the combination of analyzing a screenplay in ScripThreads, ScreenplayOwl, and ScriptFAQ may generate the richest results of all.

Regardless of the software platform, there are some inherent limitations to an automated approach to screenplay analysis. First, a screenplay nearly always differs to some degree from the completed film version, which may include improvised dialogue from the film's production or lack scenes or characters that were cut in post-production. Screenplay analysis requires us to qualify arguments we may want to make about the entire film. Another major challenge to this line of research is access — specifically, digital access to the authoritative versions of screenplays. There are several websites offering free, downloadable screenplays of contemporary Hollywood movies. The provenance,

authoritativeness, and legal status of these screenplays, though, are not clear. Moreover, these websites generally have very few screenplays for movies produced prior to the 1980s.

American Film Scripts Online (AFSO) is the digital resource that we believe provides access to the largest number of authoritative, digitized screenplays. The resource's creator, Alexander Street Press, licensed 1,009 screenplays from Warner Bros., Universal, and a number of other rights holders. Roughly half of the screenplays are available as PDF facsimiles of the original documents, and all of the screenplays are available as HTML documents, which have been rekeyed, eliminating most of the problems that come from uncorrected OCR text. The HTML mark-up is not consistently or semantically structured, but this productively forced us to code ScripThreads's parsing algorithm so that it could handle a wider variety of screenplays and not depend on rigid mark-up standards. As for AFSO's selection of screenplays, the collection is stronger in some areas than others. Over half of the 1,009 screenplays derive from 1930s and 1940s Hollywood movies — primarily, productions from Warner Bros., RKO, and MGM (it's no coincidence that Time Warner holds the rights to all three of these studio film libraries). AFSO is well suited, then, for research questions focusing on Hollywood's "Golden Age." Other strength areas of AFSO include 1990s American films (both studio and independent) and large script collections from certain contemporary screenwriters, including Paul Schrader, Lawrence Kasdan, and John Sayles. As a result, AFSO is also well suited for researching questions involving screenplay authorship, a topic we explore later in this essay.

3. ScripThreads: Parsing Method and Forms of Visualization

The ScripThreads software prototype is a cross-platform tool for the analysis and visualization of screenplays. The tool is written in C++ and utilizes the QT toolkit for its graphical user interface, making it easily ported to multiple systems. Figure 2 shows a screenshot of the graphical user interface, which offers separate windows for the Reading View, Visualization View, and Character and Settings View. By September 2014, users will be able to download the ScripThreads prototype at http://scripthreads.org. ScripThreads takes in text and HTML file screenplays as an input, parses these files and generates data for visualization and analysis. The features of the tool are described below and showcased with visualizations from *The Big Chill* (1983) and other Hollywood screenplays.



Figure 2. The ScripThreads graphical user interface separate windows for the Reading View, Visualization View, and Character and Settings View.

10

3.1. Parsing Method

While screenplays contain a structure that is far more defined than other mediums, we still found substantial variation between different works. For instance while most works contained indentation and then the character name in full capitalization before each paragraph of dialogue, the number of spaces this indentation took varied greatly between different works. Furthermore, some authors used indentation and capitalization to indicate other screenplay attributes such as sound effects, locations, and times.

For these reasons we used a two-step process to automatically find characters in a screenplay. In the first pass, each line was analyzed to determine if it was potentially the indication of a character. Lines which were considered candidates were pushed to a list. After the first pass, the list was analyzed to determine the most likely amount of indentation before a character name. For instance, a given screenplay may have 12 spaces of indentation before a character name with the spoken lines included in the paragraph below.

A second pass was then undertaken to generate information of which characters were in which scene. This pass consisted of parsing for three different items, character names, scene breaks, and meta-information. Character names were gathered from the first pass. Additional names could also be entered by the user. Scenes were determined by looking for a user defined set of keywords, such as "int." or "ext.". In practice, it was determined that a list of 10-12 keywords well captured scene changes. Finally, meta-information data, such as page numbers, were found using simple matching techniques.

This second pass generated data for each scene as to how long the scene lasted, which characters were involved, and for how many pages in the original screenplay the scene encompassed. From this data, four different types of visualization are available to users, each with its own utility.

3.2. Force Directed Graph

After the characters are detected, ScripThreads applies the framework of network theory by drawing a relationship (edge) between characters (nodes) who share the same scene. Force directed graphs are a common method visualizing this type of information. Digital Humanities scholars have used force directed graphs to represent networks of characters, literary authors, and topics [Simeone 2012]. In a force directed graph, edges are represented as virtual springs and nodes are represented as virtual masses. During runtime, a virtual physics simulation is run which attempts to find an equilibrium position for all nodes. Figure 3 is an example of a typical force directed graph. In this case, we used Gephi to graph the network of characters from the 1983 film, *The Big Chill*.



Figure 3. A typical force directed graph of a character network. This graph shows the network of characters from The Big Chill (1983).

Unfortunately, this approach is only designed for representing connectivity information, and is not designed for a time-16 based approach. One approach to encode temporal information along with connectivity information is to treat each timestep as an individual force directed graph. These graphs can then be converted into a 3D data structure which can be analyzed from arbitrary positions. The limitation of this approach on its own is that each time-step is handled on an individual basis; there is no guarantee that the nodes will appear as temporally connected threads.

To overcome this limitation, ScripThreads accounts for the temporal dimension through use of a single 3D data structure. Each node is placed on a series of time-step planes on the z-axis and are not only connected to the other relevant nodes in their time-step, but are also connected to the previous and future states. This enables each character to be viewed as a virtual thread that can become entangled in other threads when relationships occur. An added value to the vertical alignment of character threads, as opposed to the horizontal arrangement of the xkcd examples, is that the entire script is visible and able to be scrolled alongside the visualization — enabling the integrated close and distant reading noted in the introduction. The rendering system also simulates the idea of a virtual thread, by rendering each character as a continuous character thread — thick when the character is active, thin when not active. This enables unbiased color blending to occur on a model level, often referred to as "color weaving" [Hagh-Shenas et al. 2007]. Lines of dialogue between the characters can be seen as interconnects between the thread with the color representing the speaker.

The force directed visualizations reveal insights into a screenplay's narrative structure, especially for films featuring 18 episodic segments or parallel narratives, parallel protagonists, or parallel lines of action. In screenwriting manuals,

"narrative structure" is often synonymous with "three act structure" and the goal-oriented protagonist whose pursuit of some goal pushes the story from one act to the next. The goal-oriented protagonist is fundamental to Hollywood storytelling, and we acknowledge that ScripThreads does not capture this important dimension of screenplay structure.

^[1] However, we would also argue that there is more to narrative structure than simply the pursuit of goals across three acts. Supporting characters will be introduced along the way, but will they return and, if so, how and when? Are there parallel lines of action as other characters pursue their own goals or serve as thematic foils for the protagonist? The force directed graph of Paul Schrader's *Mishima: A Life in Four Chapters* (1985) clearly captures the specific character threads contained in four sections of the script (see Figure 4). Similar visual patterns in other force directed graphs would suggest the screenplay may be episodic in its overall structure and the way in which it uses supporting characters.



Figure 4. Force directed graph of *Mishima: A Life in Four Chapters* (1985) by Paul Schrader. Circle annotations mark the episodic segment breaks.

Another reoccurring visual pattern we have detected is alternating convergences of different color character threads. Such alternating clusters strongly suggest the screenplay features parallel protagonists, parallel narratives, or parallel lines of action that play out across different spaces. The force directed graph of *The Lord of the Rings: The Return of the King* (2003), for instance, reveals the parallel lines of action as the hobbits and their allies pursue their goals across different spaces (Figure 5). The different lines of action and character threads converge for the climax and lengthy epilogue.



Figure 5. Parallel lines of action (their intersections circled for emphasis) revealed in the force directed graph of Lord of the Rings: The Return of the King (2003).^[2]

In contrast to the *Mishima's* identifiable episodes and the alternating lines of action in *The Return of the King*, the force directed graph of *The Big Chill* (1983) is an example of a film that clusters its characters together in the same spaces and scenes throughout the movie (Figure 6). The narrative of *The Big Chill*, directed by Lawrence Kasdan and written by Kasdan and Barbara Benedek, centers on a group of seven college friends who reunite after a shared friend, Alex, dies. Most of the film's action occurs after the funeral at a spacious vacation house — a setting that facilitates many different types of character interactions, ranging from two characters speaking privately to scenes that bring the entire group together. The frequency of these interactions between the same group of characters is visibly evident in the tight clustering of the threads colored red (Sam), pink (Harold), orange (Michael), light orange (Nick), green (Meg), and light green (Sarah) toward the center of the force directed graph. The character whose thread veers in the largest arc from the central group is Chloe, Alex's girlfriend who did not attend college with the rest of the group and speaks far less than the other characters.



Figure 6. ScripThreads force directed graph of *The Big Chill* (1983), written by Lawrence Kasdan and Barbara Benedek.

3.3. Absence Graph

In ScripThreads' "absence graph," the x-axis measures presence and absence. A thread's distance from the center of xaxis conveys length of absence as a character — measured forward and backward in time. The resulting visualization can be read like a bus map: characters run parallel routes when they both appear in a scene. When a character is not in a scene, his or her bus route splits off. *The Big Chill's* absence graph (Figure 7) calls our attention to the purple thread of Richard, who is Karen's husband and an outsider from the core group of college friends. Karen's sub-plot in the film centers on her decision about whether to stay with Richard or leave him for her college boyfriend, Sam. As the purple thread's wide arcs reveal, Richard is absent for most of the film. However, the graph also shows that he is more important to the narrative than a character who only matters to a single scene.



Figure 7. ScripThreads absence graph of The Big Chill.

3.4. Presence Graph

The "presence graph" provides a quick glance as to when a character is active in a scene. The size of the thread is wider when the character is active and is smaller when the character is not active. Time is shown in the y-axis from top to bottom. Horizontal lines indicate dialog between the characters.

If we return to *The Big Chill*, this screenplay's presence graph (Figure 8) helps us see that the storytellers do not treat the core group of friends with equal emphasis. The male characters of Sam, Harold, Michael, and Nick speak in more scenes and appear with greater frequency than any of the female characters (the character statistics CSV supports this claim). The character of Sam (red thread) is integral to advancing the plotlines of multiple characters and helps motivate transitions between scenes. However, his love interest, Karen (blue), is largely absent unless the entire group comes together for a scene or the focus turns to her plotline with Sam.

22



Figure 8. ScripThreads presence graph of The Big Chill.

3.5. Increasing Graph

The increasing graph is useful for communicating, in a single image, character activity and storytelling techniques across the course of a narrative. Unlike the force directed and convergence graphs, ScripThreads' increasing graph is not rooted in the network theory. Perhaps for this reason, though, we've found that Humanities researchers unfamiliar with networks tend to find the increasing graph the fastest to grasp and interpret.

The increasing graph rotates the axes from the convergence graph: the x-axis becomes time and the y-axis becomes character presence. If a character is present in a scene, then his or her colored thread vertically increases. If a character is not present in a scene, then her thread remains flat.

The Big Chill's increasing graph (Figure 9) shows that the Sam and Harold characters (red and pink) have roughly an equal level of presence throughout the screenplay. Slightly less present are Michael and Nick (orange and light orange), who also appear roughly equally, and they, in turn, are followed very closely by Meg and Sarah (green and light green). The increasing graph reinforces our earlier observation that the male characters play active roles in more scenes than the female characters. *The Big Chill's* gender imbalance is small, though, compared to the vast majority of Hollywood screenplays. Additionally, the proximity of the seven threads representing the seven college friends is, relatively speaking, extremely tight. We have yet to find another "ensemble" or "multi-character" screenplay with such a tight range of presence levels for this many characters.

24



The screenplays that Kasdan wrote and directed immediately before and after *The Big Chill* are more typical of American screenwriting (Figure 9). In most Hollywood movies, the story centers on the goals of one or two protagonists. These protagonists — antiheroes, in the case of the neo-noir *Body Heat* — are the red and pink threads. The orange threads in *Body Heat* and *Silverado* reveal another common storytelling technique that our graphs help to see: introducing a character in the first ten pages who will re-emerge in the second act to increase conflict and complicate the protagonist's pursuit of his goals. Interestingly, and fittingly for the western and crime noir genres, Kasdan made this character an authority figure in both screenplays: the sheriff in *Silverado* and the district attorney friend-turned-threat memorably played by Ted Danson in *Body Heat*.

27

29

30

When we looked at hundreds of screenplay increasing graphs, we noticed a sub-group in which the red thread shoots up diagonally in a straight line, far exceeding any other thread line (Figure 13). These graphs are indicative of screenplays that focus on a single protagonist and in which the protagonist appears in every scene or nearly every scene. We analyze this pattern in greater depth in Section 5 of this article.

3.6. Scene Stats and Character Stats

ScripThreads also gives users the option to export two different types of data: scene statistics and character statistics. In the "Scene Stats" CSV, each row represents one scene, arranged in the order they occur within the script. The fields (columns) indicate the scene's number of lines, number of characters, starting page, ending page, and location (interior or exterior). In the "Character Stats" CSV, the rows represent the screenplay's characters, arranged in descending order of scene activity. The fields here indicate the character's number of active scenes, number of dialogue lines, and percentage of involvement across the film.

4. Close Reading Case Study: Lawrence Kasdan's Grand Canyon

Thus far, we have demonstrated that ScripThreads generates visualizations that reveal storytelling techniques, character interactions, and character activity within a screenplay. In sharing our work, though, we have been asked: how does this tool yield knowledge that couldn't be gained simply through reading the screenplay, watching the film closely, or turning to the existing body of scholarship on narratology, cognitivism, and Hollywood storytelling? While we are enthusiastic about the potential of ScripThreads for distant reading, we also recognize that the close analysis of individual films will always be an important activity of film criticism and scholarship. In this section, we model how ScripThreads can be used as an interpretative tool that enhances — rather than replaces — the use of narrative theory and the method of close reading. To model how a scholar's engagement with ScripThreads can enrich an understanding of a film's narrative structure, we will continue our focus on Lawrence Kasdan's work and analyze the screenplay for *Grand Canyon* (1991).



Figure 10. The marketing of Grand Canyon (1991), written by Lawrence Kasdan and Meg Kasdan, emphasizes the film's ensemble of actors and invites us to think of the film in relationship to Big Chill (1983).

"In the 80's he brought us 'The Big Chill.' Welcome to the 90's." So reads the tag line on the movie poster for Grand Canyon (1991), directed by Lawrence Kasdan and written by Lawrence Kasdan and Meg Kasdan (Figure 10). Grand Canyon's marketing suggests a close relationship between it and The Big Chill, a comparison emphasized further by both films' casting of Kevin Kline and emphases on an ensemble of actors. The question, then, arises - how similar or different are the two films?



Figure 11. ScripThreads force directed graph of Grand Canyon (1991), written by Lawrence Kasdan and Meg Kasdan.

The ScripThreads visualizations for *Grand Canyon* (1991) show that its scene structure and character interactions are significantly different from *The Big Chill* (1983). The wider thread arcs and non-intersecting threads of *Grand Canyon's* force directed graph (Figure 11) show that there are characters who appear numerous times in the film but never share the same scene. Davis (green thread) and Deborah (purple thread), for example, are never present in the same scene. Other characters, such as Dee (light green) and Claire (pink), are only present once in the same scene. Whereas *The Big Chill* is about a network of old friends who physically reunite in the same space, *Grand Canyon* is an example of what some film scholars have referred to as a "network narrative" — a multi-protagonist film that follows numerous characters whose lives intersect at different moments [Bordwell 2006, 94–103]. Of course, one does not need a computer visualization to recognize this difference between *Grand Canyon* and *The Big Chill*. From viewing both films, the difference between the characters gathered spatially in *The Big Chill's* house and the characters dispersed across the city of Los Angeles in *Grand Canyon* is quite apparent. The racially diverse cast of *Grand Canyon* and the film's overriding interests in relations across race and social class are also starkly different from *The Big Chill*, a film in which no non-white character holds narrative significance. So, the question remains — what does ScripThreads offer that simply viewing the films does not?

If we put aside the comparative question and instead focus on the details of *Grand Canyon's* multi-character structure, then more interesting insights begin to emerge. In *The Way Hollywood Tells It*, film scholar David Bordwell writes:

In Lawrence Kasdan's *Grand Canyon* (1991), the married couple Mack and Claire and the brothersister pair of Simon and [Deborah] are given roughly equal emphasis... other plotlines show Mack's

son falling in love with a girl he meets at camp, [Deborah]'s son being alienated, and Mack's friend Davis vowing to stop making ultraviolent movies. The subsidiary characters don't encounter all the customary obstacles and setbacks, yet their wants are developed beyond the limits of a traditional subplot, providing thematic echoes or counterpoints. [Bordwell 2006, 96]

When we look at the increasing graph for *Grand Canyon*, it is striking to note how uneven the distribution of character presence is across the film (see Figure 12). The Kevin Kline character of Mack (red thread) appears in 50% more scenes than any other character. The next two most active threads are those of Mack's wife, Claire (pink), and Simon (orange), the tow truck driver who helps Mack after a car breakdown in one of L.A.'s worst neighborhoods. Simon's sister, Deborah (purple), appears as one of many of the subsidiary characters clustered toward the bottom. She is present in fewer scenes than either Davis (green) or Mack's son Roberto (light orange).



Figure 12. Increasing graph of *Grand Canyon* (1991), written by Lawrence Kasdan and Meg Kasdan.

Does this mean that David Bordwell's analysis of the film is incorrect? No. Bordwell never claims that Mack, Claire, Simon, and Deborah are given equal screen time. Instead, he's suggesting that the storytelling techniques invite the audience to think of the characters as equally important. In fact, Bordwell's book offers insights that explain the discrepancy between Mack's on-screen involvement and the audience's understanding of Mack as one of multiple roughly equal characters. Bordwell describes *Grand Canyon* in the context of contemporary "ensemble films" in which "several protagonists are given equal emphasis, based on screen time, star wattage, control over events, or other spotlighting maneuvers." The star wattage and spotlighting maneuvers are especially significant to our interpretation of *Grand Canyon* as an ensemble drama. As already discussed, the film's marketing emphasized the ensemble of actors. In terms of star power, Danny Glover and Steve Martin had both headlined more commercially successful movies than

Kevin Kline prior to the movie's release.

The screenplay's two major spotlighting maneuvers, which occur at the beginning and end, further encourage us to perceive Mack (Kevin Kline) and Simon (Danny Glover) as equal in narrative importance. The script's most important spotlighting maneuver occurs from pages 6 to 16: the white lawyer Mack's car breaks down at night in South Central Los Angeles; armed young black men approach him and tell him to get out of the car; Simon, a black tow truck driver, pulls up, tells the armed men that Mack is his responsibility, and takes Mack back to his house in an affluent neighborhood. The screenplay interweaves scenes of Mack waiting for the tow truck and scenes of his family back in Brentwood. When Simon and Mack are leaving South Central, Simon's line, "My sister and her kids live near here," motivates a transition to a scene that introduces Simon's sister and her troubled son. The Mack and Simon characters are contrasted by their race and social class, yet treated as equals through the attention to each one's family and their shared sense that, in the words of Simon, "the world ain't supposed to work like this... Everything's supposed to be different than it is." This loss of faith in the social and moral order — counterbalanced by the possibility for human kindness, growth, and, even, miracles — provides the thematic glue for the entire film. The final scene of Mack, Simon, and their loved ones gazing in wonder at the Grand Canyon reestablishes our understanding of the narrative parity between the lives of Mack and Simon. Simon's question, "What do you think?" and Mack's response, "I think... it's not all bad. Not at all," provides an affirmative, glass-half-full answer to the existential guestions that have run throughout the 136-page screenplay.

The many scenes that occur between the car breakdown and visit to the Grand Canyon further encourage the audience to think of "the married couple Mack and Claire and the brother-sister pair of Simon and [Deborah]" as "roughly equal [in] emphasis" [Bordwell 2006]. Bordwell points out that "their lines of action follow Thompson's four-part template," referring to the four act structure that Kristin Thompson identifies as running through most Hollywood movies, including films with multiple protagonists [Thompson 1999]. In *Grand Canyon*, Mack and Simon both move through the four act cycle of setup, complicating action, development, and climax and epilogue. As Thompson argues, it is characters in pursuit of goals that defines classical Hollywood storytelling more than any other feature. The goals often change, and characters generally have both long-term and short-term goals. Mack and Simon are both seeking to restore their faith in the universe and humanity. Yet there are a series of more concrete short-term goals (Mack wanting to return a favor to Simon), appointments (Simon's date with Jane), and deadlines (Mack and Claire's need to make a decision about the baby) that move these characters through the four act structure and make *Grand Canyon* a Hollywood film rather than one of the existential dramas of Ingmar Bergman.

As our analysis has shown, the ScripThreads graphs can help scholars, critics, and practitioners better appreciate how storytelling techniques shape the audience's perception of a narrative. In the case of *Grand Canyon*, the force directed graph (Figure 11) offers an additional insight: the character of Mack functions structurally as the film's key bridge node. Mack (red thread) is active in scenes with nearly all the major characters. He introduces Simon (orange thread) to Jane (blue thread), with whom Simon becomes romantically linked. More importantly, Mack motivates our introductions to his co-worker and one-time lover Dee (light green thread) and movie producer friend Davis (green thread). Searching for happiness and answers in their own lives, Dee and Davis advance the central themes of the film. Yet unlike some modern films, these thematically linked storylines do not occur in isolation to one another. Mack motivates our introductions to these characters and appears in subsequent scenes with them. When Mack visits the wounded Davis, the focus is squarely on Davis and his storyline — encouraging us to think of it as Davis's scene, despite Mack's presence, and furthering the overall notion of *Grand Canyon* as an ensemble film. The ScripThreads force directed graph is useful, then, for reminding us that Mack, as a bridge node, is still vital to the structuring of multiple storylines that are not his own.

5. Distant Reading Case Study: Locating the Hyper-Present Protagonist

Some research in the Digital Humanities begins with a fixed research question and a clear process for gathering evidence. But as Sinclair, Ruecker, and Radzikowska suggest, another important task for the Humanities is "to locate (or discover) new material, with no prior knowledge of the kinds of details used for retrieval" [Sinclair, Ruecker, and Radzikowska 2013]. In our case, we tested out the distant reading possibilities of ScripThreads by using the tool's

35

36

37

"Automate" function to export four types of graphs (force directed, absence, presence and increasing) and the statistical CSV files for the screenplays in the corpus that were produced between 1930 to 2006. We then looked at the graph images for patterns that stood out visually across numerous screenplays. The first pattern that came to our attention was a sub-group of increasing graphs in which the red thread moves diagonally at a straight line, advancing higher and straighter than any other thread line (Figure 13). These graphs pointed us toward the storytelling pattern of what we call the "hyper-present protagonist" — screenplays featuring a main character who appears in every scene or nearly every scene.



Figure 13. Single character increasing graphs for four films with a protagonist who appears in every scene or almost every scene: from left to right, *I Am a Fugitive from a Chain Gang* (1932), *Across the Pacific* (1942), *On Dangerous Ground* (1952), and *Pi* (1998).

After identifying the pattern, we began searching for all instances of the pattern both visually and mathematically. We used R to write and execute a simple algorithm that targeted the exported Character Stats CSV files and extracted information on the character from each screenplay with the highest percentage of involvement. ScripThreads' Character Stats function calculates the percentage of character involvement by: A) identifying whether a character appears in a scene — yes or no; B) calculating how much of the screenplay any given scene takes up as a percentage; C) adding all of the percentage points for the instances when a character is present. In conducting this analysis, roughly one quarter of the 935 screenplays did not parse properly and we chose to disregard their results. We could have opted to use ScripThreads' Advanced Settings and gone one-by-one through the screenplays that returned inaccurate results, adjusting the settings to more clearly identify the way a particular screenplay notes characters and scene breaks. And if we were using ScripThreads for close reading, this is exactly what we would have done. However, because we wanted to test how the ScripThreads prototype performed at scale, we moved forward in analyzing the reduced corpus of 674 screenplays. As we continue to improve the tool, we anticipate the rate of screenplays that accurately parse at the computer's first pass will increase.

Out of the 674 screenplays, the median percentage of maximum character involvement was 80% and the mean percentage was 78%. Only 70 of the screenplays (roughly one-tenth) had a main character present in 94% or more of the screenplay. This group of 70 screenplays formed the sub-set of screenplays featuring a hyper-present protagonist that we analyzed in more detail. Specifically, we were interested in three variables: historical era, genre, and author. Which, if any, of these variables held the most significance for stories featuring the lead character in every scene?

5.1 Historical Era

When we the examined the production dates of our sub-set of data, we found examples of the hyper-present protagonist in screenplays ranging from 1932 to the 2000s.

However, the data indicates that this storytelling tradition did not become prominent in American cinema until the 1940s. Out of the 189 screenplays from the 1930s that we analyzed, we found only two that clearly featured a hyper-present protagonist — *I Am a Fugitive from a Chain Gang* (1932) and *20,000 Years in Sing Sing* (1932).^[3] These two outliers are both social problem films produced by Warner Bros. during the pre-Production Code era. Both screenplays are also

39

set in prison contexts and co-written by Brown Holmes, a point we will return to in our discussion of authorship. Our historical analysis suggests that the hyper-present protagonist grew far more common in Hollywood movies produced during and after World War II. Our findings confirm David Bordwell's argument that Hollywood writers, directors, and producers innovated new storytelling techniques during the 1940s that filmmakers have used ever since [Bordwell 2006] [Bordwell 2013]. Historically, this rise can be understood as part of the effort of filmmakers and writers during the 1940s to produce films of greater psychological realism, moral ambiguity, and experimentation in storytelling.

Ultimately, our historical analysis was a case in which distant reading confirmed what the leading historians of film style and narrative have argued. But as Matthew Jockers points out, we should not expect distant reading and computational literary analysis to always overturn previous understandings. There is value to "bring[ing] a new type of evidence and a new perspective to the matter and in so doing fortify...the existing hypothesis" [Jockers 2014]. Additionally, in this case, our analysis also suggests something interesting about *I Am a Fugitive from a Chain Gang*, a remarkable film about a man whose life is ruined when he is falsely convicted of crime and sentenced to hard labor on a chain gang in the American South. Few would dispute that this film was unusual for its day, with its strong social critique and bleak ending. Our analysis suggests yet another unusual element that contributes to the film's power — the protagonist (played by Paul Muni) is in nearly every scene, increasing the film's psychological intensity and sense of claustrophobia.

5.2. Genre

In our analysis of genre, we confirmed certain existing assumptions about film storytelling and challenged others. Screenwriting guidebooks generally discuss the detective movie as the genre most likely to be presented as a "closed story," a storytelling strategy that keeps the audience aligned with the protagonist's point of view and reveals information to the audience only at the moments when information is revealed to the protagonist. In contrast, an "open story" is one in which the audience learns information that the character does not know [Field 2009] [Hunter 2004]. As we expected, several examples of the hyper-present protagonist were detective films, such as *Murder, My Sweet* (1942), *On Dangerous Ground* (1952), and *8MM* (1999). Our analysis turned up other examples of films that were essentially detective stories, even though the protagonist was not officially a detective, such as *Across the Pacific* (1942). And, in the case of *The Siege* (1998), our analysis allowed us to more clearly see the film's lead investigator character and detective structure, even though the film was marketed on the basis of its action sequences and premise (What if the U.S. responded to a terrorist attack by placing New York City under martial law?).

Interestingly, though, most of the hyper-present protagonist films we found did not belong to the detective genre. Numerous dramas used the "closed story" and hyper-present protagonist technique to place the audience firmly in the perspective of a main character and amplify the film's psychological intensity (examples include *Now Voyager* [1942], *Light Sleeper* [1992], *Pi* [1998], and the previously mentioned *I Am a Fugitive from a Chain Gang* [1932]). We also found dramas that were "open stories" yet still featured the protagonist in nearly every scene. To provide an example familiar to many readers, the character of George Bailey in *It's a Wonderful Life* (1946) appears in nearly every scene of the film. When he goes absent, though, the audience learns vital information that creates a sense of dramatic irony (for example, only the audience and Mr. Potter know that Uncle Billy accidentally put the bank deposit in Potter's lap). Moreover, the first half of the film contains a great deal of voice-over narration that provides information that George doesn't know — not the least of which being that angels in heaven are discussing his plight and reviewing his life. Yet the film also pivots to a "closed" storytelling mode as we learn dramatic news at the same as George, such as the stroke suffered by George's father or the community's run on the bank. Our distant reading analysis of genre helped us distinguish between the presence of a hyper-present protagonist and the question of how a story conveys information. Additionally, the analysis reveals how films can move much more fluidly between "open" and "closed" modes than most screenwriting manuals would suggest.

If the detective genre is especially fertile ground for the hyper-present protagonist, then are there other genres in which such a character is rarely found? We have yet to find any occurrences in the romantic comedy or musical (a genre that frequently sets a romantic comedy story to song). The romantic comedy, by its nature, depends on multiple characters and obstacles to delay their happy union until their end. These conflicts and obstacles are often rooted in misunderstandings, which require the audience to know information that one of the characters does not know. Beyond

45

46

43

the romantic comedy, we found that comedy screenplays, in general, almost never feature a hyper-present protagonist. One explanation might be that writers depend on the protagonist's absence to create situations that serve as set-ups for the jokes later delivered verbally or physically by the protagonist. The very title of *A Night at the Opera* (1935), one of the comedy screenplays in our dataset, is premised on the incongruity between the high-class form of the opera and the low-class antics of its stars, the Marx Brothers. Numerous sequences are structured to play up this incongruity — moving between scenes in which the brothers are absent to set up our expectations, followed by one or more of the brothers entering to disrupt the status quo and deliver the joke.

Only one comedy screenplay in our dataset featured a hyper-present protagonist. This outlier was the Jim Carrey comedy *Liar Liar* (1997), an example of what is known in contemporary Hollywood as "high concept" (a movie with a story that can be distilled and marketed to audiences in 25 words or less). In the case of *Liar Liar*, the high concept is, "What if a hotshot lawyer could not tell a lie for 24 hours due to his son's birthday wish?" The premise enables screenwriters Paul Guay and Stephen Mazur to generate nearly all the jokes with the protagonist present. This comedy depends on incongruity, but it's different than the incongruity derived from placing the Marx Brothers at fancy restaurants and the opera. Instead, the incongruity comes from the difference between Fletcher's (Jim Carrey's) compulsive lying before his son's wish and how he must adapt after he can no longer tell a lie. This example demonstrates that a film's concept (and perhaps the desire of producers to fully capitalize on their highly paid star) can override the pattern of character presence typical for a particular genre.

5.3 Authorship

Finally, we explored the question of authorship as it relates to patterns of the protagonist's presence. Do some screenwriters have a tendency to write films that focus on one protagonist and place that character in nearly every scene? The answer, we found, was yes. As noted earlier, screenwriter Brown Holmes co-wrote the two outliers we identified from the 1930s, *I Am a Fugitive from a Chain Gang* and *20,000 Years in Sing Sing*. This finding suggests that Holmes may have had an especially important role in dramatically structuring and writing the two films, despite working as a contract writer for Warner Bros. and sharing the writing credit with co-authors on both screenplays. As we continue our research, we plan to integrate ScripThreads with other modes of computational analysis and examine more co-authored Warner Bros. screenplays from the 1930s and 40s. We believe the result might offer a better sense of the individual contributions of creative artists on the highly collaborative medium of film. We may find that some writers contributed especially strongly to a film's dialogue, while others, like Brown Holmes, left major contributions to a film's dramatic structure.

As we examined the AFSO corpus for hyper-present protagonists, one author leapt out at us for his tendency to structure films with a main character present in nearly every scene. The AFSO corpus contains the screenplays for fifteen films that were either written or co-written by Paul Schrader, who is best known for writing and directing dark, character-oriented dramas, such as *American Gigolo* (1980) and *Affliction* (1997), and writing some of the best films directed by Martin Scorsese, including *Taxi Driver* (1976), *Raging Bull* (1980), and *The Last Temptation of Christ* (1988). Out of this group of fifteen films, eight films feature a main character present in 94% or more of the screenplay, and only two films place the main character in less than 80% of the screenplay (the two outliers are *Obsession* [1976], a Hitchcockian thriller directed by Brian DePalma, and *Blue Collar* [1978], which Schrader directed about three auto workers who rob their corrupt union).

48

49



Figure 14. Increasing graphs of fifteen produced screenplays either written or co-written by Paul Schrader. The graphs were superimposed onto one another using ImageMagik.

Schrader's tendency to frame stories around single characters who are almost always present can be seen in Figure 14, which superimposes the fifteen increasing graphs of Schrader's screenplays onto one another. The consistency of Schrader's approach is clear from the cluster of red lines (each one representing the most present character from a different Schrader film) thrusting diagonally in a nearly straight line. For a point of comparison, we can turn back to Lawrence Kasdan. Figure 15 superimposes the seven Kasdan screenplays that are available in the AFSO corpus. The Kasdan image shows a screenwriter who works across numerous genres and utilizes a wide variety of storytelling approaches — one film featuring a hyper-present protagonist (*Mumford* [1999]), one multi-character drama with no singular protagonist (*The Big Chill* [1983]), and several films that fall in-between.



Figure 15. Increasing graphs of seven screenplays written or co-written by Lawrence Kasdan. The graphs were superimposed onto one another using ImageMagik.

In some ways, Figure 14 provides an illustration for what film critics and scholars already assume about Schrader: he is a filmmaker who writes character studies about men who are psychologically and/or existentially anguished. American Gigolo (1980), Light Sleeper (1992), and Affliction (1997) are all films that Schrader wrote and directed that fit this paradigm. Schrader calls his protagonists "existential heroes" and, in a widely quoted interview with Garry Wills, remarked, "all my life has been dedicated to the existential hero, and the existential hero seems to have come to the end of his path, replaced by the ironic hero" [Schrader and Wills 2006]. However, there is no inherent reason why an existential hero needs to appear in nearly every single scene of a movie. As noted earlier, Lawrence Kasdan's Grand Canyon keeps its existential protagonist, Mack, absent for numerous scenes, even though he appears in far more scenes than any other character. Similarly, the existential heroes in films directed by Ingmar Bergman and Woody Allen generally exist within an ensemble of characters. In Bergman and Allen films, the protagonist will go absent to facilitate scenes focusing on supporting characters, who may serve as foils to the existential hero, address the story's major themes, and/or advance the plot. There is a long history of this practice in literature, as well. Tolstoy's Anna Karenina would have far less to say about love, society, politics, and life's meaning without the narrative of Levin and Kitty playing out in parallel to Anna's story. The ScripThreads' increasing graph visualizations of Scharder's work reveal that a defining characteristic of Schrader's existential hero is his hyper-presence. The audience stays with the existential hero throughout the entire film. Ingmar Bergman makes films that address the existential themes of Sartre and Camus, but Schrader is the filmmaker who structures his narratives more similarly to their novels — following the "closed story" model that keeps the audience aligned with the protagonist's point-of-view.

Schrader's best known screenplays follow the hyper-present, existential protagonist model we have described. 52

However, a more heterogeneous portrait of Schrader as an author emerges when we examine his twelve unproduced screenplays in the AFSO corpus. To be clear, these are screenplays written by Schrader that were never made into films. Figure 16 shows the superimposed increasing graphs of these 12 screenplays. This group of screenplays includes four hyper-present protagonist scripts: three music-oriented biopics (*Dream Lover: The Bobby Darin Story, Eight Scenes from the Life of Hank Williams*, and *Gershwin*); and one detective film (*The Investigator*), a genre that we know is comparatively likely to have a hyper-present protagonist. What is more surprising is that seven out of the twelve unproduced screenplays place the protagonist in less than 80% of the script. And three screenplays make the protagonist present in less than 70% (Schrader's two produced outliers, *Obsession* and *Blue Collar*, are both higher at 72% and 75% respectively).



Figure 16. Increasing graphs of twelve unproduced screenplays written or co-written by Paul Schrader. The graphs were superimposed onto one another using ImageMagik.

Schrader's unproduced project in which the protagonist is least present (59%) is his script for a retelling of *Snow White* (Figure 17). Here, we see an interesting dynamic between an author's general tendencies and the demands of a particular project. The *Snow White* tale depends on dramatic irony — the audience knows the apple is poisonous, but the protagonist does not. Moreover, the antagonist needs scenes without the protagonist present to interact with the magic mirror, one of the most iconic elements of the myth.





The increasing graphs of Schrader's unproduced work open up a series of questions that we plan to investigate further. What is the relationship between genre and authorship in how narratives are structured? How do certain genres or stories override a screenwriter's established narrative techniques? Finally, how do industrial and cultural assumptions about a particular screenwriter shape the types of projects the writer is offered and that make it into production? By examining Schrader's unproduced works at a distance, we can speculate that industry assumptions about what constitutes a "Paul Schrader film" — namely, a dark, existential character study — may have made studios and financiers more reluctant to produce screenplays by Schrader that fell outside of this paradigm.

6. Conclusion

In this article, we have introduced the ScripThreads tool and demonstrated how it can be used for closely analyzing one screenplay (*Grand Canyon*) and analyzing a much larger group of screenplays to look for a pattern. ScripThreads parses screenplays and extracts data about character presence and interactions, then offers users statistical CSV files and a series of 3D and 2D visualizations that present the data in a time-based manner. The tool is not without its limitations. ScripThreads does not visualize act breaks, shifts in spatial location, or a protagonist's journey toward a particular goal. Yet the data the tool offers can be quite telling. It can reveal details of character presence and character co-occurrences that humans are prone to forget or never remember in the first place.

In using ScripThreads to closely analyze a single film, the continuities and differences between the viewer's perception and the computer's visualizations is a powerful starting point for uncovering storytelling techniques and better understanding cognitive reception. In using ScripThreads to analyze a large group of screenplays, the visualizations and

CSV output files allow researchers to recognize patterns without having prior knowledge of the films. To draw accurate and meaningful conclusions, though, some domain expertise in film history is essential (just as a researcher would want some knowledge of 19th-century literature before making arguments about the century based on topic modeling 100 Victorian novels). Whether applied toward close ready or distant reading, ScripThreads is meant to help researchers gain a richer understanding of the text or texts they are studying. This is a tool to aid Humanities scholars in analysis and interpretation, not a substitute for screenwriting and criticism.

57

ScripThreads offers one additional affordance — the ability to quickly visualize the narrative structure of an unproduced screenplay. Films and television programs play out their stories visually in sequences of photographed and edited action; a produced screenplay has already been visualized for the screen. However, due to the difficulty and expense of making a film, the vast majority of screenplays are never produced, never visualized in this fashion. The increasing graphs of Schrader's unproduced screenplays (Figure 16) provide what may be the first transformative visualizations of these twelve works. The graphs allow us to quickly recognize one way (the protagonist's level of presence) that some of these stories differ from Schrader's better known screenplays. What if we could apply a similar analysis to the screenplay libraries of Hollywood's studios, producers, and talent agencies? The results would yield not simply graphs of character presences, co-occurrences, and absences, but transformative renderings of thousands of stories that have remained absent from audiences. We would no longer be visualizing American film history; we would be visualizing a history that might have been.

Notes

[1] For a brief overview of the three act structure and Kristin Thompson's richer four act structure, see Eric Hoyt's online Prezi, "Hollywood Storytelling: 3 Act or 4 Act Structure," 3 October 2013, http://prezi.com/x7fhnbeofobw/hollywood-storytelling-3-act-or-4-act-structure/. Readers are also strongly encouraged to read Kristin Thompson's *Storytelling in the New Hollywood: Understanding Classical Narrative Technique* (1999).

[2] *The Return of the King* screenplay, written by Peter Jackson, Fran Walsh, and Phillippa Boyens, came from the Internet Movie Script Database (IMSDb). http://www.imsdb.com/Movie%20Scripts/Lord%20of%20the%20Rings:%20Return%20of%20the%20King%20Script.html (accessed 5 October 2013). It is the only screenplay we discuss in this article that does not derive from the American Film Scripts Online Collection.

[3] We should acknowledge that more screenplays from the 1930s failed to parse properly than screenplays from any other decade. Nevertheless, we stand by our analysis and claim that the ever-present protagonist became more common in American cinema beginning in the 1930s.

Works Cited

- AFSO 2009 American Film Scripts Online (2009). Alexander Street Press. http://alexanderstreet.com/products/americanfilm-scripts (accessed 1 November 2013).
- Abello, Broadwell, and Tangherlini 2012 Abello, J., Broadwell, P., & Tangherlini, T. R. (2012). "Computational Folkloristics." *Communications of the ACM*, 55(7), 60-70.
- Bordwell 2006 Bordwell, D. (2006). The Way Hollywood Tells It. Berkeley: University of California Press.
- Bordwell 2013 Bordwell, D. (2013). "The 1940s, mon amour." Observations on Film Art [Blog]. http://www.davidbordwell.net/blog/2013/03/28/the-1940s-mon-amour/ (accessed 14 May 2014).
- Field 2009 Field, S. (2009). *Going to the Movies: A Personal Journey Through Four Decades of Modern Film*. New York: Random House.
- Hagh-Shenas et al. 2007 Hagh-Shenas, H.; Sunghee Kim; Interrante, V.; Healey, C., (2007) "Weaving Versus Blending: a quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color." Visualization and Computer Graphics, IEEE Transactions on, vol. 13, no. 6, pp. 1270, 1277, Nov.-Dec.
- Hunter 2004 Hunter, L. (2004). Lew Hunter's Screenwriting 434: The Industry's Premier Teacher Reveals the Secrets of the Successful Screenplay. New York: Penguin.

Jockers 2014 Jockers, M. (2014). Text Analysis with R for Students of Literature. New York: Springer.

Kasdan and Kasdan 2001 Kasdan, L. and Kasdan, M. (2001). Grand Canyon (screenplay). American Film Scripts Online.

- Manovich 2012 Manovich, L. (2012). How to Compare One Million Images? *Understanding the Digital Humanities*. Ed. David M. Berry. Basingstoke, Hampshire, England and New York: Palgrave Macmillan. 249-278.
- Marinov 2011 Marinov, S. (2011). "Making Plays: Comparative Structural Analysis of Writing for Stage and Screen." 4th Screenwriting Research Network International Conference in Brussels, Belguim. September 2011. (Unpublished paper).
- **Marinov and Stitts 2013a** Marinov, S. and Stitts, B. (2013). "Development of Screenwriting Analysis Software Based on Pfister's Theory of Verbal Communication." Screenwriting Research Network International Conference in Madison, Wisconsin. August 2011. (Unpublished paper.)
- Marinov and Stitts 2013b Marinov, S. and Stitts, B. (2013). Screenplay Owl: Screenplay Analysis Software. http://dramaticanalysis.com/owl/ (accessed 1 November 2013).
- McKie 2014 McKie, S. (2014). ScriptFAQ: Screenplay Analytics. http://phd.tripos.biz/ (accessed 14 May 2014).
- Moretti 2000 Moretti, F. (2000) "Conjectures on World Literature." New Left Review. Jan-Feb 2000. 54-66.
- Murtagh, Ganz, and Reddington 2011 Murtagh, F., Ganz, A., and Reddington, J. (2011). New methods of analysis and semantics in support of interactivity. *Entertainment Computing*, 2: 115-121.
- **Polan 2001** Polan, D. (2001). "Auteur Desire." *Screening the past*, 12, http://www.latrobe.edu.au/www/screeningthepast/firstrelease/fr0301/dpfr12a.htm (accessed 18 October 2013).
- **Roberts-Smith et al. 2013** Roberts-Smith, J., et al. (2013). "Visualizing Theatrical Text: From Watching the Script to the Simulated Environment for Theatre (SET)." *Digital Humanities Quarterly*, 7 (3). http://www.digitalhumanities.org/dhq/vol/7/3/000166/000166.html (accessed 5 May 2014).
- Schrader and Wills 2006 Schrader, P. and Wills, G. (2006). "Paul Schrader: An Interview with Garry Wills." In Shouts and Whispers: Twenty-one Writers Speak about Their Writing and Their Faith. Edited by Jennifer L. Holber. 113-119.
- Simeone 2012 Simeone, Michael. (2012) "Visualizing Topic Models with Force-Directed Graphs." *The Stone and the Shell* [Blog]. 2 December 2014. http://tedunderwood.com/2012/12/02/visualizing-topic-models-with-force-directed-graphs/ (accessed 15 May 2014).
- Sinclair, Ruecker, and Radzikowska 2013 Sinclair, S., Ruecker, R., and Radzikowska, M. (2013). "Information Visualization for Humanities Scholars." In *Literary Studies in the Digital Age: An Evolving Anthology*. Ed. Kenneth M. Price and Ray Siemens. MLA Commons. http://dlsanthology.commons.mla.org/information-visualization-for-humanities-scholars/ (accessed 14 April 2014).
- Tanahasi and Ma 2012 Tanahasi, Y., and Ma, Kwan-Liu. (2012). "Design Considerations for Optimizing Storyline Visualizations." *IEEE Transactions On Visualization And Computer Graphics*, Vol. 18, No. 12: 2679-2688.
- **Thompson 1999** Thompson, K. (1999). *Storytelling in the New Hollywood: Understanding Classical Narrative Technique*. Cambridge, Massachusetts: Harvard University Press.

Xkcd 2009 Xkcd. (Nov 2009). "Movie Narrative Charts," http://xkcd.com/657/ (accessed 2013 October 22).



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.