

The Stuff of Science Fiction: An Experiment in Literary History

Stefania Forlini, University of Calgary, Canada
 Uta Hinrichs, University of St Andrews, Scotland
 Bridget Moynihan, University of Calgary, Canada

Abstract: This article argues for a speculative, exploratory approach to literary history that incorporates information visualization early on into, and throughout, the research process. The proposed methodology combines different kinds of expertise—including that of fans and scholars in both literary studies and computer science—in processing and sharing unique cultural materials. Working with a vast fan-curated archive, we suggest tempering scholarly approaches to the history of science fiction (SF) with fan perspectives and demonstrate how information visualization can be incorporated into humanistic research processes, supporting exploration and interpretation of little-known cultural collections.

A Shared-Expertise Approach to Cultural Analysis

In her recent work, *How We Think*, N. Katherine Hayles suggests that Digital Humanities (DH) research offers new opportunities for creative collaborations across disciplines and with “expert amateurs” beyond “academic walls” [Hayles 2012, 36]. Such collaborations require that we acknowledge the value of different kinds of expertise, welcome cross-pollination of approaches and promote open access to rare materials and to analytical activities. In this article, we argue for a collaborative, shared-expertise approach to literary and/or cultural collections and showcase our own take on such an approach. The work we present is the result of the collaboration between an expert amateur who compiled one of “the most important research archives” of SF [Latham 2010, 161] and scholars in both literary studies and computer science.

In an attempt to reconsider numerous neglected specimens of proto/early SF and to devise a DH approach to literary history suited to popular genres, our project specifically aims:

- to investigate how the Bob Gibson anthologies of speculative fiction—unique, hand-crafted and fan-curated anthologies of SF’s “great unread” [Moretti 2000, 54]—can contribute to scholarly assessments of the evolution of SF, and
- to develop information visualizations that enable researchers, students, fans, and the general public to explore the collection from different perspectives, promoting fluid movement between close and distant reading.

Visualizations are key to our collaborative and exploratory approach. Rather than using visualizations solely to display the final *results* of our research, we are experimenting with evolving interactive visualizations in tandem with our research questions and ongoing investigations of Gibson’s untapped collection of little-known materials. The visualizations are thus integral to our research *process*, shaping and being shaped by our research questions and ongoing findings, and are not simply tools used as a means to an end. This exploratory approach is necessary because very little is known about the primary materials we are investigating. Moreover, we want to make visible our research process as it is impacted by digital tools and to remain open to unforeseen questions and research avenues that arise through our first-hand interactions with the collection and with the developing visualizations.

In describing this process, our article makes two main contributions: we outline our emergent method for the study of early SF in the context of the Gibson Collection, adapting Franco Moretti's [Moretti 2005] evolutionary approach to genre, and we demonstrate how integrating information visualization early on in the research process can support exploration and guide the interpretation of vast, largely unknown literary collections. We begin by introducing the Gibson anthologies and their unique characteristics before describing our theoretical approach to investigating and classifying their material contents, and the design considerations underlying our visualizations. Because the focus of this paper is primarily on the value of visualizations employed throughout the research process, we describe our initial visual explorations and findings as well as new iterations of our visualizations as they evolved. Finally, we identify the benefits and limitations of our approach to early SF as it has developed so far and outline future steps.

The Gibson Anthologies and the History of SF

The Bob Gibson anthologies of speculative fiction are unique compilations of SF stories hand-crafted by Canadian collector and devoted SF fan, Bob Gibson (1908-2001). Gibson harvested a wide range of science-fictional materials for his more than 890 anthologies from primarily English-language magazines published from the 1840s onwards.¹ He then bound these materials himself into unique compilations, illustrated many of his covers (see Fig. 1), and provided for each anthology a hand-written table of contents, which—significantly—includes symbols through which he rated the “SF content” of the items he collected (see Fig. 2). Although Gibson left no key to these symbols, we argue that his system of classification will help elucidate the history of periodical-based SF.

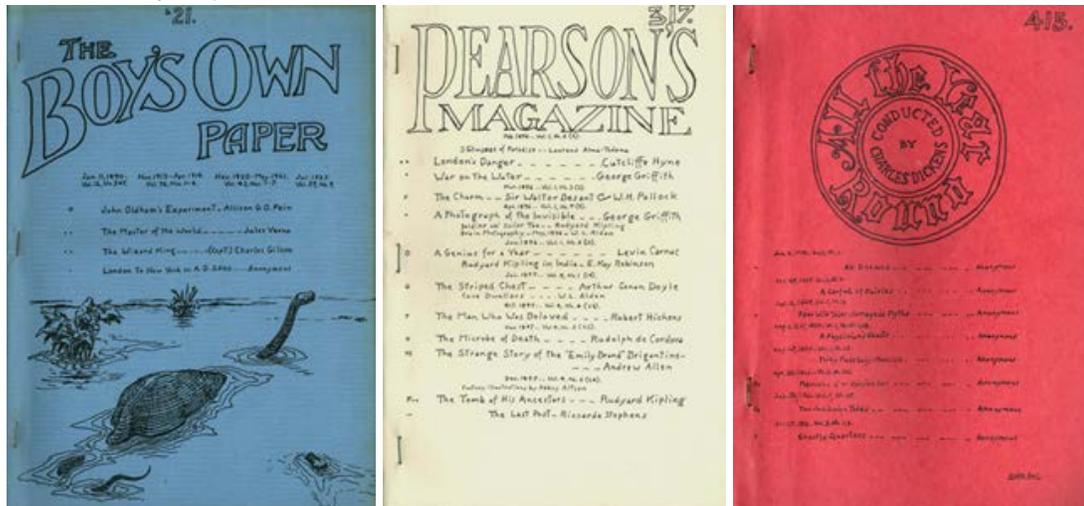


Figure 1: Three anthologies that are part of the Gibson Collection².

¹ Although not numerous, there are some items from French-language magazines.

² Images are courtesy of Special Collections, Taylor Family Digital Library of the University of Calgary.

2010]. In Gibson's more than 40,000-item collection held at the University of Calgary, the approximately 890 anthologies are a relatively small part—but no less important for all that, given the uniqueness of their content.⁴ As far as we know, no other resource undertakes the cataloguing, editing, and classification of early SF published before the advent of pulp magazines devoted to the genre the way Gibson's unique anthologies do. In fact, even an initial perusal of the anthologies suggests that they have the potential to alter drastically our understanding of early SF in particular. In the 167 anthologies containing works published before 1930, that is, before SF is named as such and before the emergence of pulp magazines devoted to this “new” genre, we have already identified 83 women writers not mentioned in Everett Franklin Bleiler's authoritative bibliography of early SF⁵; their very presence in the Gibson anthologies offers a substantial revision to accounts of early SF that see this history as dominated by male writers. Of course, the contributions of these (and many other) forgotten writers included in the anthologies cannot be fully assessed without also assessing Gibson's challenge to the definition of SF and its related genres that led him to include these writers in the first place.

If the Gibson anthologies offer us the tantalizing possibility of a more rigorous account of the early years of SF, they also present us with what Matthew Wilkens calls “the problem of abundance” [Wilkens 2012, 250] and challenge us to develop new methodologies adequate to the task. In developing our own method, we adopt Moretti's conception of genre as a “diversity spectrum” of forms that evolve over time [Moretti 2005, 77] while drawing on Gibson's anthologies to guide the detection of formal experimentations in the early years of SF when the genre is at its most amorphous. We use custom-built information visualizations to tap into Gibson's unique expertise, allowing us to explore his own classification and to promote hypothesis-making and testing, playful exploration, and serendipitous discoveries.

The potential of information visualization, that is, “the use of computer-supported, interactive, visual representations of abstract data to amplify cognition” [Card, Mackinlay, and Shneiderman 1999, 7], is increasingly recognized for supporting humanistic inquiry ([Jänicke et al., 2015, Jockers 2013, Moretti 2005], among others). Information visualization techniques make visible patterns within data collections that are otherwise difficult or impossible to see. In this way, information visualization can help to communicate an argument by providing concise visual evidence (as shown in a literary context by Moretti [Moretti 2005]). However, we believe that it can also facilitate more micro-level analyses. As Hayles suggests, information visualization allows us “both to see large-scale patterns and to zoom in” to examine specific details more carefully [Hayles 2012, 77], and, thus, can help balance the quantitative analyses of distant reading with the thoughtful, deliberative interpretations of close reading. In addition, we are testing the extent to which information visualization helps raise new research questions and hypotheses and, through interactive capabilities, supports the in-depth investigation of these throughout the research process. One could say that our experiment is a double one: it explores the early stages of SF's evolution through Gibson's archive even as it explores the use of information visualization in generating new research questions, hypotheses, and

⁴ For more information on the Gibson Collection housed at the University of Calgary's Special Collections, see Boyd 2010, Hemmings 2005 and www.ucalgary.ca/gibson-speculative-fiction/.

⁵ This figure does not include women who wrote anonymously or used initials or a male pseudonym.

classifications. Of these two, necessarily entangled, experiments, this paper will focus primarily on the latter.

Evolutionary Approaches to Science Fiction

Following Moretti, we are interested in studying the evolution of genre by tracking the successes and failures of “stylistic mutations” [Moretti 2005, 91]. However—and this is where we divert from Moretti—we do not decide in advance what counts as a meaningful formal feature⁶. To do so would be problematic in the case of this little-known body of early SF. While we could track imagined technological inventions or another widely recognized “novum” of more recent SF, we instead examine specimens of early SF in a more open-ended fashion to ensure that we remain sensitive to the unique features of this under-studied moment in the genre’s evolution. As Moretti himself notes, “when a new genre first arises [...] no ‘central’ convention has yet crystallized” and, as such, it is “open to the most varied experiments” [Moretti 2005, 77]; its diversity spectrum is “[q]uite wide” [Moretti 2005, 77]. Ultimately, we aim to provide a more robust account of the diversity spectrum of early SF, but our experiment begins with, and is fundamentally shaped by, the materials meticulously curated by Gibson to acknowledge, as others have, that fans are uniquely poised to contribute to the study of popular culture [Hills 2002, Jenkins 1992].

To our knowledge, there is to date only one other DH project on SF that also adopts an evolutionary approach: Rabkin and Simon’s Genre Evolution Project (GEP). The GEP tracks the evolution of American SF stories published from 1927 onwards through a relational database, which links individual works to major historical events and keeps track of reprints and translations of stories as evidence of their “survival”—unlike Moretti’s evolutionary approach, which tracks the emergence and evolution of *formal features* rather than stories. GEP is remarkable for its focus on periodical-based short stories, its attention to the wider cultural context in which these stories were published, and for its unique approach to developing keywords for each story. However, because it begins with stories published in 1927, it effectively skips over the messiest part of the genre’s evolution.⁷ Moreover, the project’s definition of survival commits it to a major works/major authors’ approach, which continues to underestimate the contributions of so-called minor works and writers in the genre’s development. In contrast, we adopt Gibson’s more inclusive approach and Moretti’s definition of survival based on formal features to tackle the wide diversity spectrum of an emerging genre. This approach is already showing the limitations of current scholarly histories and classifications of early SF, most of which are based on the features of 20th-century SF rather than the unique features of earlier stages of evolution.

For the purposes of this exploratory pilot study, we have digitized a subset of 50 Gibson anthologies (those with works that are in the public domain) and collected data on these 50 and on 22 more anthologies for a total of 72 anthologies, containing the earliest works (more than 1,500 items). We are collecting metadata on every item, including authors, publication years,

⁶ See, for example, Moretti’s examination of clues in detective fiction in *Graphs, Maps, Trees* and “The Slaughterhouse of Literature”.

⁷ Many critics agree that the SF genre begins to achieve its modern form in the 1880s and 90s, but even then it obviously “did not simply spring into being” [Stableford 2004, 3]. GEP skips this formative period altogether. It also excludes serialized fiction, which was popular in the early years of SF.

plot summaries, and keyword motifs. This information is stored in a relational database, which in turn becomes the basis of the visualizations that we are developing to help us track patterns across items in the anthologies. To highlight our scholar/fan collaboration, we began by exploring what counts as SF in Gibson's system of classification.

Toward a Living Classification

Classification systems are foundational to any human practice of knowledge production and maintenance, particularly in digital media ecologies where information must be coded and stored in databases to be digitally accessible. However, teasing out and exploring the aspects that these systems make invisible, exclude, and otherwise silently define are equally necessary if we are to create what G. C. Bowker and S. L. Star call a "living classification" [Bowker and Star 1999, 326]—a classification that remains flexible and finely tuned to its parameters and limitations. We believe that the Gibson anthologies—with their 78 "SF content" symbols used by Gibson to classify each item⁸—offer an opportunity to challenge existing classifications of SF and to explore new and more flexible systems. The openness of the Gibson anthologies themselves therefore guides our classification choices and our presentation of these anthologies, which in turn forms the basis of our database structures and our visualizations. However, we also recognize the need to compare Gibson's classification with well-established scholarly systems to understand more fully Gibson's potential contribution to ongoing scholarly work.

We began our efforts to conceptualize the Gibson anthologies by surveying previous SF bibliographies, dictionaries, and encyclopedias, aiming for a system that would allow us to classify the formal features, plot devices, and content of the anthology items. Many compilations were rejected, as they provided only a broad definition of SF, typically based on 20th-century texts⁹. Others provided a detailed classification structure, but still focused on full-blown SF and/or canonical novels, while using terms, such as cyberpunk and nanotechnology, that are anachronistic to early/proto SF¹⁰. Similar anachronistic terms, as well as an understanding of "Hard SF" based on problematic 20th-century understandings of scientific disciplines, inform the classification structure used by GEP [Rabkin and Simon 2014]. Relying substantially on any of these systems would superimpose modern definitions of SF onto the anthologies and inherently exclude much of their content from consideration.

Of the classification systems we reviewed, only Bleiler's *Science-Fiction: The Early Years* [Bleiler 1990] focuses on the rich, porous period of proto-SF in order to take account of early periodical publications, including stories and authors that may only have been published

⁸ Based on the careful examination of Gibson's marginalia in the anthologies, we have come to understand these symbols as Gibson's self-conscious efforts to classify the works he compiled in his anthologies. His notes, although cryptic at times, provide some clues about the significance of individual symbols and his reasons for choosing one symbol over another.

⁹ These include D. H. Tuck's *The Encyclopedia of Science Fiction and Fantasy* (1978), E. Naha's *The Science Fictionary: An A-Z Guide to the World of SF Authors, Films, & TV Shows* (1980), J. Gunn's *The New Encyclopedia of Science Fiction* (1988), M. Ashley's *Time Machines: The Story of the Science Fiction Pulp Magazines From the Beginning to 1950* (2001), and B. Stableford's *Historical Dictionary of Science Fiction Literature* (2004).

¹⁰ These include T. A. Shippey and A. J. Sobczak's *Magill's Guide to Science Fiction and Fantasy Literature* (1996), N. Barron's *The Anatomy of Wonder: A Critical Guide to Science Fiction* (2004), and *The Encyclopedia of Science Fiction* <<http://www.sf-encyclopedia.com/>>.

once. His extensive bibliography covers over 3000 stories written between the third century BC and 1930, with most works falling between 1870 and 1930. Bleiler also includes a list of nine major SF motifs that are then divided into sub-motifs, all of which were developed using terminology drawn from the stories that comprise his bibliography, thereby minimizing the imposition of anachronistic terms (see Fig. 3). Bleiler's system allows us to categorize the diverse science-fictional items in each Gibson anthology according to the nine major motifs, while also allowing us to account for these items' more granular differences through the sub-motifs.

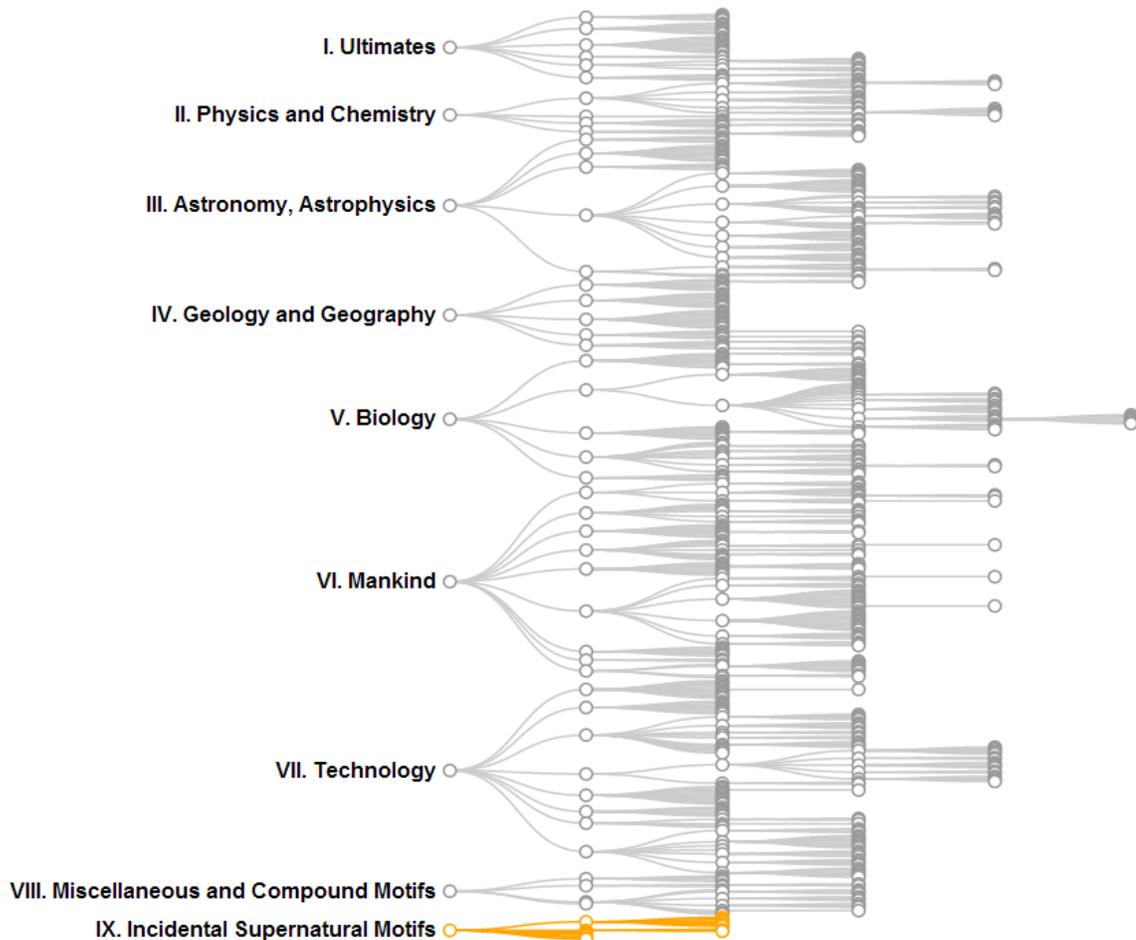


Figure 3: Bleiler's SF classification system, branching from nine major motif categories.

While Bleiler's system in many ways compliments Gibson's, their points of tensions prove at least as productive as their points of similarity. For example, like many other critics, Bleiler underestimates the importance of the supernatural in early SF. This is reflected in the positioning of the supernatural last among Bleiler's major motifs and in the motif's title: "Incidental supernatural motifs occasionally found, particularly in the earlier works" [Bleiler 1990, xviii]. Furthermore, this motif is the least differentiated in Bleiler's hierarchy; in contrast to other motifs, it contains few sub-motifs (see Figure 3, highlighted in orange), thus limiting the specificity with which supernatural elements can be coded through Bleiler. In contrast, our preliminary analysis of the earliest subset of anthology items shows that Gibson does not

exclude the supernatural, but rather seems to suggest its affinity with early SF. In fact, many of the items in our subset include supernatural elements together with more recognizable SF elements, such as technology, medical developments, and/or experiences with other worlds or planets. As we will show in our discussion section, such hybridity indicates that the supernatural has haunted SF since its beginnings, and that it is modern understandings, rather than the historical moment, that have sought to exclude it. Consequently (and perhaps unsurprisingly) such stories, rigorously tracked by Gibson, rarely occur in Bleiler's bibliography.

Adopting Bleiler's motifs does not mean that we neglect relationships such as the one between SF and the supernatural, nor does it mean that we seek to override Gibson's classifications with Bleiler's. Rather, the productive tension between Bleiler's and Gibson's approaches to categorization highlights the need for precisely the type of close analysis enabled by Gibson's anthologies and our visualizations. Moreover, we are not relying exclusively on Bleiler as a point of comparison, as we are allowing trends in our data set itself to indicate when new keywords, perhaps unique to this collection, are needed. These keywords are being developed through our own reading of the stories, and so continue Bleiler's efforts to create keywords inductively from the SF content itself. Maintaining an open space of interaction between existing systems (e.g., Bleiler and Gibson) and our own emerging classification system allows us to apply multiple lenses through which to zoom in on the messy areas of early SF that have been bypassed to this point and to move toward a living classification of SF.

Visualization Perspective

Much like our attempts to classify items in the Gibson anthologies, our design approach to visualize the items and their connections is exploratory. Before discussing our early visual experiments and explorations, we will outline a number of considerations that we began to establish early on in this experiment to guide our choice of visualization techniques and the design of the visualizations:

- *Supporting Open-Ended Explorations:* Recent discussions in the area of Information Sciences (see, e.g, [Marchionini 2006]) and information visualization [Dörk et al. 2011, Thudt et al. 2012]) suggest supporting more open-ended, exploratory strategies toward digital collections, moving beyond query-based search interfaces. In our project, the ability to examine Gibson's anthologies in an exploratory way is essential because of the open-ended nature of the research questions and because of the need for both researchers and general-interest users to first familiarize themselves with this unknown collection. We aim at addressing an open-ended interaction paradigm in the design of our visualizations.
- *Addressing a Broad Range of Audiences:* An information visualization ideally enables a dialogue between people and data. The intended audience (e.g, their interest and background) therefore necessarily influences the design. Creating a single visual exploration tool that satisfies the different audiences interested in the Gibson anthologies is a challenge. The analytical focus of researchers may be difficult (perhaps even impossible) to reconcile with a more curiosity-driven approach of SF fans and the broader public. We consider these different motivations and explore how different types of visualizations can meet these, individually or in combination.

- *Linking Different Aspects of the Collection:* The metadata we extracted manually for anthology items provides a rich source for the analysis of corresponding temporal and contextual patterns. Drawing from and extending existing visualization techniques, we investigate how this manifold data can be visually represented individually and in relation to each other. Similar to previous approaches that have been applied to cultural collections [e.g., Dörk et al. 2008, Hinrichs et al. 2008, Thudt et al. 2012, Whitelaw 2015], these techniques will then be linked to form a visual exploration tool, wherein the interaction with one visualization influences the others and shows how the different facets of the Gibson anthologies relate to each other.
- *Synergy between Distant and Close Reading:* Following up on Jockers’s argument on macroanalysis [Jockers 2013], we aim at creating a synergy between distant and close reading by providing a range of overviews that show the Gibson anthologies from different high-level perspectives, which act as an entry point into further analysis and exploration by revealing curious patterns [Shneiderman 1996]. At the same time, we will provide a strong connection to the individual anthology items so that direct access to, and close reading of, the original sources is possible at all times during the exploration process.
- *Aesthetic Considerations:* As literary collections are being made accessible through digital interfaces, we need to find ways to preserve their unique material and aesthetic character, which fundamentally shapes our experience and understanding of them. Throughout this project we will explore how to maintain and convey at least some of the unique visual and material characteristics of the Gibson anthologies. One approach is to give the unique hand-illustrated anthology covers a strong presence in the interface.

Visual Explorations

Our first visual explorations focused on cracking the Gibson code of symbols. From our work in Special Collections, we suspected that the symbols served at least two functions. Notes and marginalia left by Gibson indicate that he was coding for “SF content” but also rating the stories for the “quality” of such content. However, Gibson’s notes do not explicitly outline the system of symbols, nor do they provide any key. Given the sheer number of symbols (26) and the number of items (1513), even in our relatively small subset of 72 anthologies (out of the total 890), we brainstormed ways that visualizations might help us explore the meaning of specific symbols and their relation to other similar-looking symbols. We therefore relied on both literary studies and computer science perspectives on the data and its conceptualization, eventually deciding to juxtapose anthology items and their corresponding Gibson symbols with Bleiler’s motifs through a radial visualization that correlates these motifs with Gibson’s symbols. The resulting visual explorations are shown in Figures 4 – 7. As we will explain, these visualizations helped us begin to characterize the Gibson symbols, helped make the collection more accessible by beginning to clarify the contents of the anthologies for users unfamiliar with the collection, and, last but not least, invited further exploration while helping to generate new

hypotheses, research questions, and new iterations of visualizations designed to help us address these new hypotheses and questions.

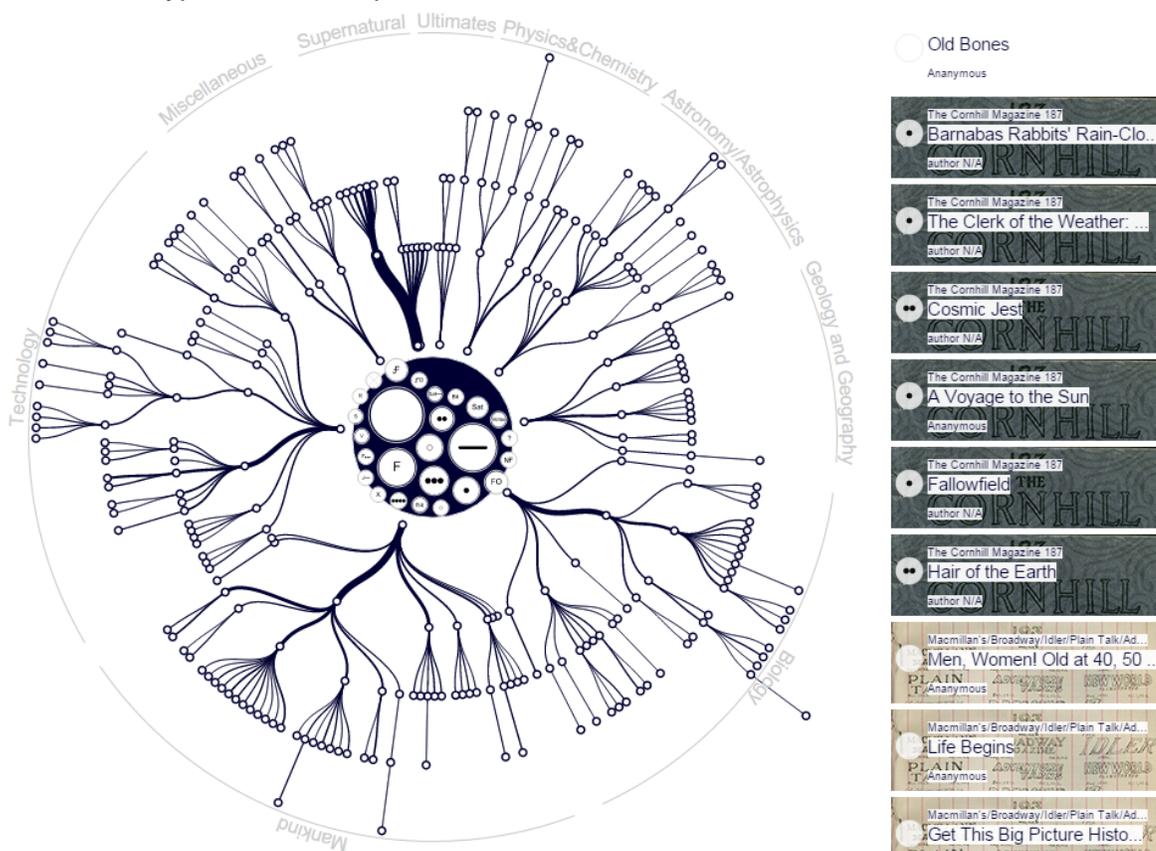


Figure 4: Distribution of anthology items within Bleiler's SF motif hierarchy.

Figure 4 shows an initial visualization that includes an interactive radial tree diagram (left), which includes Gibson's symbols at the centre and Bleiler's hierarchical motifs in the radiating branches, and is interlinked with a list of anthology items (right), providing multiple points of entry to explore the contents of the anthologies. The cluster of circles at the middle of the radial tree diagram shows all of Gibson's symbols present in our subset, with the size of each symbol representing its frequency¹¹; for example, we can see that the “dash” and the letter “F” are the most frequent symbols used in our subset, as they appear largest. The branches that radiate from centre of the diagram show Bleiler's hierarchical motifs, with the stroke weight of each radial branch representing the frequency of the motif (measured by the number of corresponding anthology items associated with the motif). The list to the right of the radial tree diagram shows each anthology item's symbol, title, author and publication year with a cover image of the corresponding anthology. This list is an entry point to more details about each item (e.g., a content summary and a link to the original item, where available), while also revealing information about the symbols when a particular symbol is selected. Any selection in one part of

¹¹ The empty circle that occurs with the symbols indicates those anthology items to which Gibson has not assigned a symbol.

the visualization will result in corresponding adjustments to all other parts that invite possible hypotheses and/or interpretations.

Figure 5, for instance, shows all the motifs, symbols, and items that correspond to a selected symbol, a particularly ambiguous Gibson symbol, which might either be described as a conjoined “JF” or a stylized F with a tail. By reading the individual stories marked with this symbol, one might suspect that this symbol refers to “Juvenile Fiction” or “Fairytale”, as it includes titles such as “The Little Mermaid” and “Wyemarle and the Forest Fairies”. However, the visualization brings all these stories together onto one screen, which allows for easy comparison and confirmation of the hypothesis. Although using database queries might also lead to similar results, the visualization makes the discovery much more accessible and immediately meaningful, because it allows the user to see simultaneously how, for example, these fairytales intersect with Bleiler’s motifs, how many other stories share this symbol, and what anthologies and source magazines contain these stories.

The simultaneous communication provided by the visualization is particularly important for symbols that have hundreds of items associated with them, as in the case of the “F” or “-“ symbols. In order to decipher these symbols without the help of the visualization, one would either have to retrieve dozens of anthologies from Special Collections and peruse hundreds of stories, keeping careful notes of any observable patterns, or scrolling through database search results, all while (again) manually keeping track of observable potential patterns (which becomes difficult the more items one attempts to track). All such work would have to be done before being able to develop—let alone confirm—hypotheses about a symbol’s meaning. With the help of the visualization, however, certain hypotheses are suggested by the simultaneous visual intersection of variables (symbols and motifs, and story titles) and can be quickly ruled out and/or validated. For example, in the case of the “F” symbol, a perusal of the abstracts and motifs provided in the visualization reveals that this symbol is not likely related to the stylized F (or JF) that signifies fairytales or juvenile fiction, as the stories associated with the “F” symbol are much more eclectic, tend to deal with ghosts, disappearances and mysteries, and are directed at a more adult audience. This prompted us to look into the “F” symbol items further, and with the guidance of Gibson’s marginalia, we now believe that this symbol refers to “Fortean” stories, stories that focus on unexplainable observed phenomena. Returning to the visualization and perusing the motifs and abstracts of stories labeled with this symbol confirms this hypothesis. In the cases of both the stylized F or JF and the plain “F” symbols, it is clear that visualization is effectively incorporated into the research *process* (which includes more traditional humanistic modes of inquiry such as archival work), even at its earliest, most exploratory stages, and not, as is often the case, used solely as a way to showcase *results* visually.

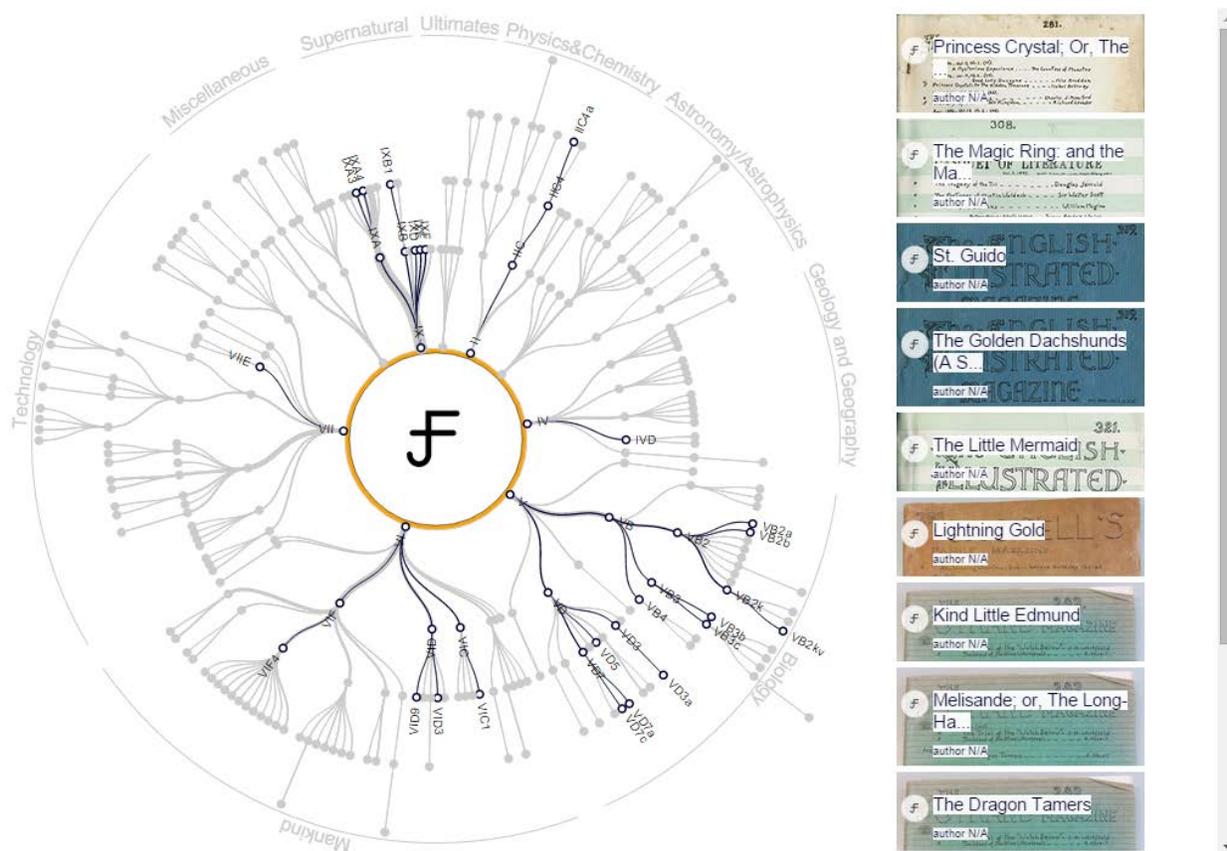


Figure 5: Selecting a symbol shows its corresponding motif branches (in blue) and anthology items.

In addition to Gibson symbols, the visualization offers Bleiler motifs as another point of entry to Gibson's anthologies. The nine branches of Bleiler's hierarchy are named in the text that surrounds the radial diagram (for comparison, refer to Fig. 3 and Fig. 4) and, as mentioned, the stroke weight of each radial branch represents the frequency of the motif. While Bleiler's numbered hierarchy suggests an order of importance, moving from the first ranked category, Ultimates, to the last, Incidental Supernatural Motifs, we intentionally laid out the hierarchy in a radial geometry that does not privilege certain motifs over others so that the anthologies themselves could suggest the significance of certain motifs over others. Selecting a motif highlights its corresponding branches (in orange) and filters the item list to show all stories associated with the selected motif and its sub-motifs, while the Gibson symbols in the middle circle are adjusted to show their distribution within this motif selection. For example, Figure 6 shows that, so far, 521 of 1513 items in our subset are associated with the motif branch "Incidental Supernatural Motifs" and that that this motif is related to 19 of Gibson's symbols. Here, the motif to which Bleiler granted the least importance appears in Gibson's anthologies to be the most prominent.

Moreover, the visualization shows how the motifs inter-relate. When a motif is selected, blue branches represent sub-motifs that co-occur with the selected motif across at least one anthology item (see Fig. 6). Using the blue branches as a guide to correlations, several motif branches can be selected at once in order to trace connections between motifs, their associated items, and the Gibson symbols. For example, while one might expect that the supernatural

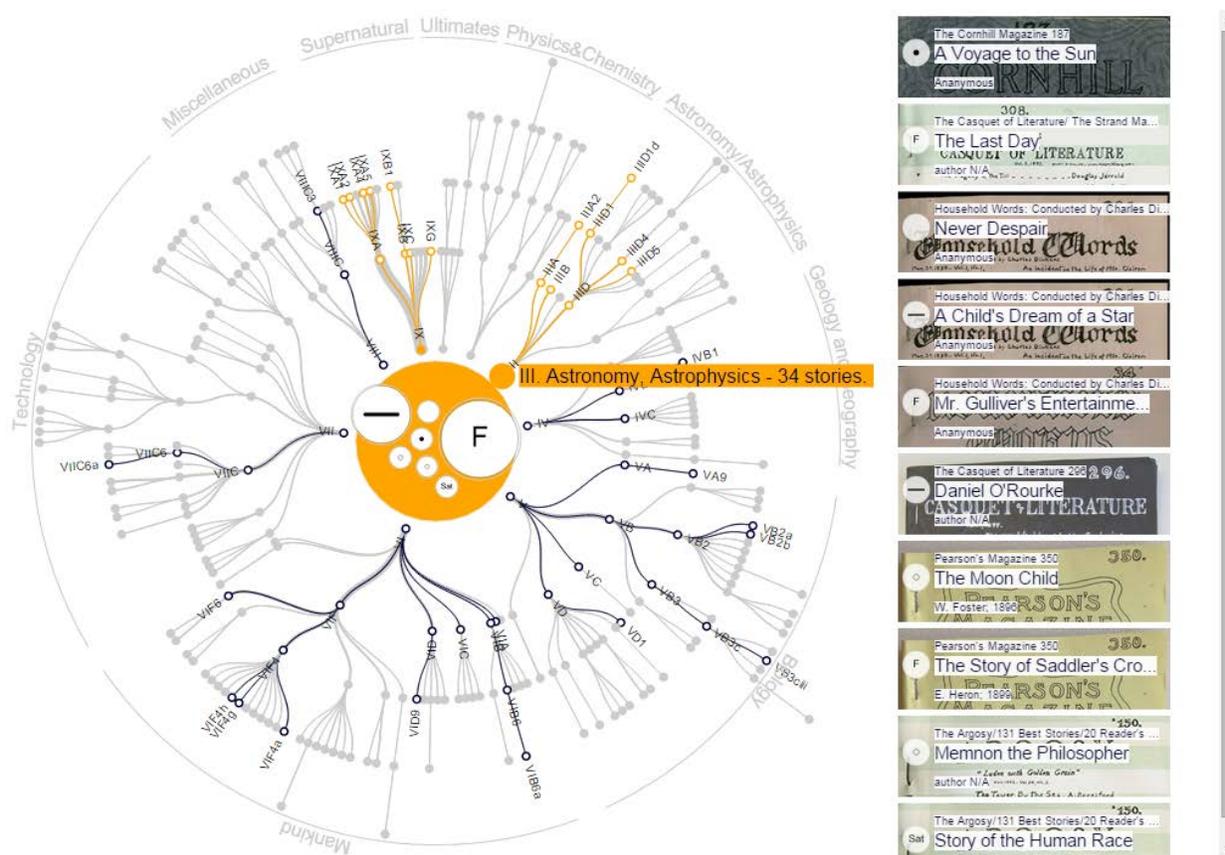


Figure 7: Selecting several motif branches provides a list of items that share motifs from both branches. The middle circle shows the frequency of corresponding symbols.

Beyond hypotheses about the symbols, the collection's content, and the evolution of SF traits, this early visualization also sparked new research questions (discussed in greater detail in the next section) and ideas about further kinds of visual explorations. We have since updated the original visualization through several phases in order to respond to specific research questions that arose during our interactions with the early visualizations and to feedback that we solicited from non-specialists. We have discussed the details of these focus groups elsewhere¹² and will not repeat them here. Instead, we highlight briefly the resulting changes to the early visualization interface that resulted in a more elaborate visualization – the Speculative W@nderverse – to show how collaboration with mixed-expertise audiences and with a developing visualization tool has intrinsically shaped our research process, which in turn continues to shape the development of the visualizations (see Fig. 8).

In presenting our early visualizations to mixed expertise audiences, we found that even for people with little or no familiarity with the contents of the Gibson anthologies, interacting with the visualizations led to discoveries and generated hypotheses about the contents of the collection. Users reported learning things they had not known in mere minutes of interacting with

¹² Full details of these feedback sessions can be found in our article, "Speculative Practices: Utilizing InfoVis to Explore Untapped Literary Collections", forthcoming in *Transactions on Visualization and Computer Graphics*.

the tool and many left with questions they hoped to explore further. Based on user feedback we have implemented a number of new features.

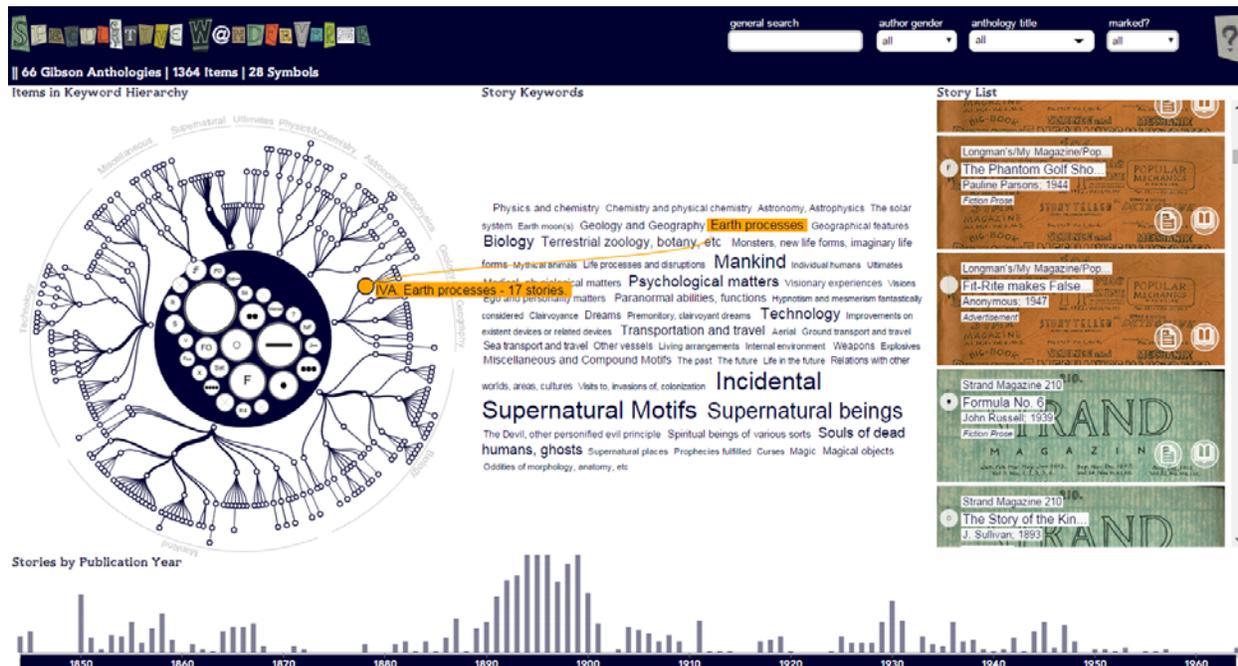


Figure 8: The Speculative W@nderverse – a result of our experimental visualization approach.

In addition to the aforementioned interactive timeline that allows users to narrow searches to items published during specific segments of time, and/or to plot the distributions of items associated with user-selected symbols and motifs, we have included more targeted search functions that allow users to search the collection by author, keyword, and publication venue to address new research questions that arose as we interacted with the early visualization. We have also re-designed the entry point to the visualization which reflects on the interpretative character of our visualization approach to the Gibson collection; the letters composing the new title of our visualization interface are from Gibson's anthology covers themselves and thus show how our approach borrows from Gibson's own scrapbooking approach (see Fig. 9).



Figure 9: Loading Screen of the Speculative W@nderverse visualization.

The Speculative W@nderverse visualization includes a more prominent view of the anthology covers with Gibson's original tables of contents and artwork; clicking on an anthology cover in the item view brings up a high resolution image of this very cover. We have also since incorporated another view of Bleiler's motifs into the visualization in the form of a tag cloud. Once again challenging Bleiler's own hierarchy, the frequency of the keywords across Gibson's items determines the size of the keyword tag, rather than the numerical order created by Bleiler, and each of these tags reflects a node in the radial diagram. While the radial diagram is easily used by scholars familiar with Bleiler's motifs, the cloud allows general-interest users to scan quickly through the most relevant keywords for the Gibson anthologies. Clicking on a word highlights the node, just as clicking on a node highlights the word in the tag cloud. Following the feedback of our users, we also added a connecting line between the node and keyword that appears when the user hovers over the node, thus emphasizing further the connection between these two views. Just as with the radial diagram, the tag cloud adjusts based on any filters utilized, and when a motif is selected, only co-occurring motifs remain in the cloud. Our visualizations are therefore evolving and developing along with our research process, and are engaged in a continual effort to meet the needs of the tool's diverse audiences. The Speculative W@nderverse in its current form can be considered a milestone in our process, but not a final result – further visualization experiments that focus, for example, on incorporating more strongly the material and visual aesthetics of the Gibson anthologies, will follow.

Discussion and Conclusion

With the help of our evolving experimental visualizations, we have deciphered numerous Gibson symbols to date, have begun to rethink definitions and categorizations of early SF, and have broken down our necessarily broad initial research questions into more targeted questions, which have led us to supplement exploratory with more targeted search functions in our visualization, and which are now fueling new, more analytical visualizations.

As with any project that hopes to grapple with what Moretti calls the “great unread”, progress is slow, but this is particularly true when grappling with an unusually long-lived genre. As Moretti notes in passing while discussing the average longevity of different genres, SF is unique in that it has lasted for more than 100 years with no sign of abating [Moretti 2005, 31]—unlike most other genres, which he found lasted about 25-30 years on average [Moretti 2005, 20-21]. This longevity and the great diversity of works considered part of the SF genre suggests the need to consider historically-specific (and thus changing) sub-genres, hybrid genres, and micro-genres of what might be considered a complex mega-genre uniquely characteristic of modernity.

While expansive—but certainly not exhaustive—the Gibson collection offers possible ways to begin characterizing the vast array of periodical-based SF as it changed over time. We will continue to compare his choices with those of scholars and to consider the limitations of his collection. This pilot study, focusing on 72 of the total 890 Gibson anthologies, begins with the earliest works compiled by Gibson (starting in the 1840s through to the early 1900s). Even this more modest time period entails going through more than a thousand items. The current exploratory visualizations (which continue to evolve and result in new visualization experiments) and the process of developing them have certainly assisted us in beginning to characterize the anthologies and to begin to understand their potential contribution to SF studies. However, they

also highlight the messiness inherent in the evolution of genres that might be obscured by neatly organized, result-oriented static visuals that seem to promise self-evident conclusions. For example, we have several items that occur within the Gibson anthologies that cannot be coded with any of Bleiler's motifs, and so become difficult to represent within our current visualizations¹³. We are also aware that different research questions or interests would necessarily create different classifications for the Gibson items than the ones that we are applying. Our tool is thus best understood as a particular lens that allows us to zoom in on the messy period of proto-SF, offering a view that must be tempered with the views of existing and ongoing SF scholarship. This means recognizing that our tool, no less than any other interpretation of the period, comes with its own exclusions and limitations. We therefore maintain that, more than simply providing certain kinds of results (such as helping to decipher the Gibson symbols), developing and incorporating interactive visualizations throughout a research process inevitably showcases the limitations of one's choices at every step, the multiplicity of approaches possible and the multitude of questions that splinter into further questions, without ever suggesting the incontestability or stability of results.

The deciphering of Gibson symbols continues to rely both on our careful perusal of Gibson's marginalia and any clues we may find therein and on the visualizations that help suggest and test a variety of hypotheses about their meaning. In this case, the visualizations effectively work in tandem with more traditional modes of archival work. The rethinking of definitions and categorizations of SF is a more fraught process, in part because it everywhere leads to the limitations both of early scholarly attempts and of the current attempt, but such rethinking also offers a sobering check that balances the promise of definitive knowledge, which implicitly comes with attempts to account for the "great unread".

The visualizations have perhaps been most successful in helping to refine and rethink research questions and possible ways to work towards answers. We began this project with fairly broad questions about how a fan-curated collection might contribute to scholarly understandings of the history of SF. However, as we gathered and began to visualize data from the collection, we started to think through the differing perspectives of the diverse audiences the tool might reach and we began to ask more clearly articulated questions. Some questions directly responded to the early visualization, such as: 1) What are the relationships or points of contrast between two particular symbols? or 2) How could we better represent the interaction between Gibson's classifications and those of Bleiler through the visualization?. Such questions led to changes to the visualization itself through both our radial tree diagram and our tag cloud. Other questions were suggested by our interactions with the visualization as we attempted to decipher the Gibson symbols. These include, for example: 1) Are certain Gibson symbols more commonly associated with certain source magazines and/or with certain authors and/or time periods? 2) Do Gibson's symbols and/or Bleiler's motifs suggest specific hybrid or sub-genres of SF that are specific to particular decades and/or periodicals? 3) Is there a difference between the kinds of topics treated by male or female writers? Such questions emerged from the ease with which the visualization provides simultaneous information on multiple variables and led us to incorporate more targeted filters mentioned above. As our questions narrow, we are also considering whether we should maintain a tool that tries to address both public and academic

¹³ These items are grouped together for now under a single node in the hierarchy, labeled "No applicable keyword" and in the tag cloud as "undefined".

audiences simultaneously, or whether the increasing specificity of our research questions will eventually necessitate multiple separate tools. As we think of new iterations and/or new separate visualizations, we are also considering questions for future research, including: 1) Could we develop a way to compare Gibson's selections from specific periodicals with the materials he rejected from these same periodicals in order to better understand his selection process? 2) How can we include images of Gibson's notes and marginalia in the visualizations so as to promote the analysis of his editorial function? 3) How can we further represent the material qualities of Gibson's unique hand-crafted anthologies in the visualization?

In this paper, we have focused primarily on the early stages of our research into the untapped Gibson anthologies to showcase specifically the value of incorporating visualization early on into the research process, when research is arguably in its most exploratory stages. This approach is in direct alliance with our developing methodology, which takes inspiration from Moretti's attention to the evolution of genres through stylistic mutations, even while adapting his approach in order to include expert amateur knowledge and to integrate visualizations *throughout the research process* (not just to support an argument about results). Furthermore, we utilize open-ended data explorations in developing more flexible classification systems of SF. Based on the premise that, as many have noted, "we think through, with, and alongside media" [Hayles 2012, 1], this pilot project specifically showcases that the interactive co-development of tool and research processes can be fruitful, especially in fostering discoveries in complex, little-known collections. While the benefits of displaying empirical research results through visualizations is well-known and visualizations have historically been called upon to serve this purpose more than any other [Drucker 2014], the use of interactive visualizations to help researchers and non-researchers experiment and play with collected data is increasingly recognized.¹⁴ Developing visualizations collaboratively (with scholars in both literary studies and computer science) promotes creative play and problem-solving across disciplines, and allows for visualizations that are uniquely suited to the collection under study. If the trend in DH seems to be toward developing generic tools that are applicable to almost any data set, we argue that there is also room for tools tailored to the specific questions raised by utterly unique collections that may be of interest both within and beyond scholarly circles. While we recognize that such an approach is resource-intensive, it also promises a more substantial pay-off for a wider, mixed-expertise audience.¹⁵

As we continue our work, our team will be providing the latest iteration of this visualization online and soliciting more feedback on the tool from a wider audience to continue to tailor the tool to the needs of different audiences and to allow for possible collaborations with expert amateurs in future research projects on this unique collection. For now, however, we hope we have begun to show the value of an exploratory, fan-driven approach to the "great unread" of early SF, and of open-ended, interactive information visualization tailored to the content and aesthetic characteristics of a unique collection.

¹⁴ See for example S. Brown et al. 2010 and T. Clement 2012.

¹⁵ The lack of usability studies in the development of new DH tools and the low adoption rate of such tools suggest that DH tool development is already resource-intensive in light of its uncertain pay-off. As such, new approaches to tool development should be explored to increase both usability and tool adoption. See Gibbs and Owens 2012 and Schreibman 2010.

Acknowledgements

We thank the Social Sciences and Humanities Research Council of Canada, the Calgary Institute for the Humanities, the University of Calgary's Special Collections and Libraries and Cultural Resources, and Jason Reid (University of Calgary, Faculty of Arts IT) for their support of this project.

References

- Ashley 2000** Ashley, M. (2000). *The Time Machines: The Story of the Science-Fiction Pulp Magazines from the Beginning to 1950*. Liverpool: Liverpool University Press.
- Bleiler 1990** Bleiler, E. F. (1990). *Science Fiction: The Early Years*. Kent: Kent State University Press.
- Bowker and Star 1999** Bowker, G. C. and Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, Massachusetts: MIT Press.
- Boyd 2010** Boyd, M. (2005). "The Bob Gibson Collection of Speculative Fiction at the University of Calgary." *Science-Fiction Studies*. 37(2), 180-181.
- Brown 2010** Brown, S. Ruecker, J. Antoniuk, S. Farnel, M. Gooding, S. Sinclair, M. Patey, and S. Gabriele. "Reading Orlando with the Mandala Browser: A Case Study in Algorithmic Criticism via Experimental Visualization." *Digital Studies*. 2(1), 2010
- Card, Mackinlay, and Shneiderman 1999** Card, S., Mackinlay, J., and Shneiderman, B. (1999). "Chapter 1: Information Visualization", Card, S., Mackinlay, J., and Shneiderman, B. (eds), *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers, 1-34.
- Clement 2012** Clement, T. (2012). "Distant Listening or Playing Visualisations Pleasantly with the Eyes and Ears." *Digital Studies*. 3(2), 2012.
- Dörk et al. 2008** Dörk, M., Carpendale, S., Collins, C., and Will, C. (2008) "VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery." *IEEE TVCG*, 14(6), 1205-1212.
- Dörk et al. 2011** Dörk, M., Carpendale, S., and Williamson, C. (2011). "The Information Flaneur: A Fresh Look at Information Seeking," in *Proc. of CHI'11*. New York: ACM, 1215-1224.
- Drucker 2014** Drucker, J. (2014). *Graphesis*. Boston: Harvard University Press.
- Gibbs 2012** Gibbs, F., Owens, T. (2012). "Building Better Digital Humanities Tools: Toward Broader Audiences and User-Centered Designs." *Digital Humanities Quarterly*, 6(2), 2012.

Hayles 2012 Hayles, N. K. (2012). *How We Think: Digital Media and Contemporary Technogenesis*. Chicago: University of Chicago Press.

Hemmings 2005 Hemmings, M. (2005). "Collection Overview: Outstanding Significance and National Importance of the Bob Gibson Collection."
http://people.ucalgary.ca/~mhemming/Gibson/collection_overview.htm

Hills 2002 Hills, M. (2002). *Fan Cultures*. London: Routledge.

Hinrichs et al. 2008 Hinrichs, U., Schmidt, H., and Carpendale, S. (2008). "EMDialog: Bringing Information Visualization into the Museum," *IEEE TVCG*, 14(6),1181-1188.

Jänicke et al. 2015 Jänicke, S., Franzini, G., Cheema, M., and Scheuermann G. (2015). *On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges*. In Proceedings of EuroVis—STARs, pages 83–103, 2015.

Jenkins 1992 Jenkins, H. (1992). *Textual Poachers: Television Fans and Participatory Culture*. London: Routledge.

Jockers 2013 Jockers, M. (2013). *Macroanalysis: Digital Methods and Literary History*. Chicago: University of Illinois Press.

Latham 2010 Latham R. (2010). "Science Fiction Studies Showcase: Library Collections and Archives of SF and Related Materials." *Science-Fiction Studies*, 37(2): 161-162.

Luckhurst 2005 Luckhurst, R. (2005). *Science Fiction*. Cambridge: Polity Press.

Luckhurst 2010 Luckhurst, R. (2010). "Science Fiction and Cultural History." *Science-Fiction Studies*, 37(1): 3-15.

Marchionini 2006 Marchionini, G. (2006). "Exploratory Search: From Finding to Understanding." *Communications of the ACM*, 49(4):41-46.

Moretti 2000 Moretti, F. (2000). "Conjectures on World Literature." *New Left Review*, 1(January/February): 54-68.

Moretti 2000 Moretti, F. (2000). "The Slaughterhouse of Literature." *Modern Language Quarterly*. 61(1), 207-227.

Moretti 2005 Moretti, F. (2005). *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.

Rabkin 2004 Rabkin, E. (2004). "Science Fiction and the Future of Criticism." *PMLA* 119(3), 457-473.

Rabkin and Simon 2001 Rabkin E. and Simon, C. (2001). "Age, Sex, and Evolution in the Science Fiction Marketplace." *Interdisciplinary Literary Studies*. 2.2 (2001): 45-58. Print.

Rabkin and Simon 2014 Rabkin, E. and Simon, C. (2014). *The Genre Evolution Project*. Available from <<http://www.umich.edu/~genreevo/>>. (Accessed 15 August 2014).

Stableford 2004 Stableford, B. (2004). "The Emergence of Science Fiction, 1516-1914," *Anatomy of Wonder: A Critical Guide to Science Fiction*. Ed. Neil Barron. Westport, Conn.: Libraries Unlimited, 3.

Schreibman 2010 Schreibman, J., Hanlon A. (2010). "Determining Value for Digital Humanities Tools: Report on a Survey of Tool Developers." *Digital Humanities Quarterly*, 4(2), 2010.

Shneiderman 1996 Shneiderman, B. (1996). "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," in *Proc. of Visual Languages*, IEEE, 336-343.

Thudt et al. 2012 Thudt, A., Hinrichs, U., and Carpendale, S. (2012). "The Bohemian Bookshelf: Supporting Serendipitous Discoveries through Information Visualization," in *Proc. of CHI*. New York:ACM, 1461-1470.

Wilkens 2012 Wilkens, M. (2012). "Canons, Close Reading, and the Evolution of Method," *Debates in the Digital Humanities*. Ed. Matthew K. Gold. Minneapolis: University of Minnesota Press, 250.

Whitelaw 2015 Whitelaw, M. (2015). *Generous Interfaces for Digital Cultural Collections*. DHQ: Digital Humanities Quarterly. 9(1).