Tags, Borders, and Catalogs: Social Re-Working of Genre on LibraryThing

MARIA ANTONIAK, Cornell University
MELANIE WALSH, Cornell University
DAVID MIMNO, Cornell University

Through a computational reading of the online book reviewing community *LibraryThing*, we examine the dynamics of a collaborative tagging system and learn how its users refine and redefine literary genres. LibraryThing tags are overlapping and multi-dimensional, created in a shared space by thousands of users, including readers, bookstore owners, and librarians. A common understanding of genre is that it relates to the content of books, but this resource allows us to view genre as an intersection of user communities and reader values and interests. We explore different methods of computational genre measurement within the open space of user-created tags. We measure overlap between books, tags, and users, and we also measure the homogeneity of communities associated with genre tags and correlate this homogeneity with reviewing behavior. Finally, by analyzing the text of reviews, we identify the thematic signatures of genres on LibraryThing, revealing similarities and differences between them. These measurements are intended to elucidate the genre conceptions of the users, not, as in prior work, to normalize the tags or enforce a hierarchy. We find that LibraryThing users make sense of genre through a variety of values and expectations, many of which fall outside common definitions and understandings of genre.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **Information systems** → **Digital libraries and archives**.

Additional Key Words and Phrases: tagging, book reviews, natural language processing

ACM Reference Format:
Maria Antoniak, Melanie Walsh, and David Mimno. 2021. Tags, Borders, and Catalogs: Social Re-Working of Genre on LibraryThing. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 29 (April 2021), 29 pages. https://doi.org/10.1145/3449103

1 INTRODUCTION

LibraryThing is an online book reviewing community that allows users to catalog their reading, curate their own book collections, and organize these collections according to unique systems and categories. This website is used not only by readers but also by maintainers of small lending libraries, who rely on LibraryThing's cataloging service to organize and distribute their collections. Tagging plays a particularly important role in this community, as it is used for LibraryThing's paid cataloging service and frequently (though not always) used to define a book's genre—that is, to define what *kind* of book it is, in the eyes of the user. The definition of literary genres has traditionally been mediated by academic scholarship and commercial publishers, but the rise of online communities of readers has brought a new perspective. LibraryThing readers work together

Authors' addresses: Maria Antoniak, Cornell University, maa343@cornell.edu; Melanie Walsh, Cornell University, melanie.walsh@cornell.edu; David Mimno, Cornell University, mimno@cornell.edu.

2573-0142/2021/4-ART29 $15.00
https://doi.org/10.1145/3449103

to organize the global library of books in an altruistic, encylopedia-like project. But tags and genres are also used to help map the community itself, to implicitly identify groups of readers who engage with certain kinds of books in certain kinds of ways.

Prior work on tagging systems in online book reviewing communities has explored how tags enhance reading experiences, help define communities, and serve as bridges between online and offline worlds. For example, Gruning [16] explores how the display of virtual objects can support identity-building and help users find ongoing value in their collections, arguing that maintenance work—tagging, shelving, reviewing—supports the user's awareness and remembrance of objects in their collection and strengthens their view of those objects as valuable. Prior work on tagging systems at CSCW has shown how users' tagging behaviors are influenced both by personal tendencies and by the patterns they observe in their community [32, 37], which we discuss in §2.1.

One of the most interesting trends that we observe in the LibraryThing community is the creative proliferation of genre, which we believe is the result of these collaborative tagging and reviewing practices. LibraryThing users can pick from an unconstrained vocabulary when adding tags, and they can also add free-text reviews. Tags can include familiar genres such as *science fiction* or *fantasy* but they can also identify new subgenres, microgenres, and personal memos. Text reviews provide an even richer opportunity for users to express subjective experiences. As a result, genre blossoms into an expansive grassroots literary taxonomy that incorporates familiar genres but also splinters into new forms that incorporate reviewer preferences and expectations.

As many literary scholars have maintained about genre elsewhere [14, 29, 46, 55], LibraryThing genres are often blurry and overlap. We show in this paper that different reading communities have different values, perspectives, and tagging practices, and these differences lead to contradictions in assigning and defining genres via tags. Worrall [56] proposes that LibraryThing and similar websites function as *boundary objects*, which connect different communities and allow them to collaborate without consensus. We suggest that tags and genres might be viewed as boundary objects, as well. In Star [40]'s original conception, boundary objects maintain coherence across different communities, despite mismatches between the meanings ascribed by these communities. In the same way, tags connect LibraryThing users who hail from different reading communities, putting them into a shared space that allows them to collaborate.

We explore what tags, reviews, and genres help capture about the LibraryThing community in the ways they relate, overlap, and conflict. We combine computational social science methods (natural language processing and information theory) with digital humanities scholarship to study the tagging systems generated by the LibraryThing community. Our experiments seek to answer two overarching research questions.

> ***Research Question 1:*** *Given the unconstrained lexical space of collaborative tags, which community dimensions can we use to map and measure a set of target genres?*

We would like to orient ourselves within the space created by the LibraryThing users, not by forcing their tags to fit a preconceived idea of what that structure should look like, but by following the signposts that have already been erected for us. We experiment with different features of the tags to curate complementary and conflicting maps of the genres, including a tag's overlap with other tags, the internal lexical consistency of a tag's reviews, and the homogeneity of a tag's reviewer community. These maps can be used to both "zoom out" from individual tags, discovering inter-genre patterns, and "zoom in" by identifying outlier reviews for close reading. In particular, we are alert to unexpected genre pairings and distances, and we compare our findings to traditional understandings of genre drawn from sources such as Wikipedia taxonomies and literary scholarship. We look for boundaries not to enforce separation but to illuminate which qualities of books are valuable to this community, as they collectively make sense of their reading experiences.

> ***Research Question 2:*** *How can we use review texts and tags to detect genre-dependent values and expectations among online book reviewers?*

LibraryThing allows us to evaluate not only the "content" of a genre but also the utility and significance of the genre to readers. Taggers and reviewers bring individual preferences to their tags and review texts. Some reviewers more strongly value an exciting plot with a surprise ending, while others value realistic descriptions of historical settings. We hypothesize that some of these values are genre-dependent and that a reviewer's expectations of a book frame the discourse of their review. We use topic modeling to create genre signatures that represent review themes, allowing us to distinguish dissimilar genres and to group genres with similar signatures. We use codifications of genre on Wikipedia as a proxy for common understandings of genre and as a source of comparison. Definitions of genre on Wikipedia tend toward clean hierarchies, while LibraryThing takes a more active, creative, and open approach, with visible contradictions between different users' tagging.

Our contributions include the following:

- **Mapping:** We map genres using lexical surprisal, community homogeneity, themes and values as expressed in review texts, and book and user overlap. We find that each of these dimensions highlights new clusters of genres.
- **Overlap:** Even when genres are distinct in terms of works, they can be overlapping in terms of user communities. Readers of *memoir* tend to also be readers of *crime*, though books in these genres are rarely tagged as the other genre.
- **Borders:** By building predictive models of genre given review text, we learn computational representations of typical reviews for a genre. These models can then identify unexpected or surprising examples. Close readings of these "misclassifications" help capture the fluidity of genres and reader experiences that cannot be easily represented by tags.
- **Values:** We find that tight-knit communities of readers who often review books in the same genre have shared values and expectations. Genres with more homogenous communities (e.g., *romance*, *horror*) tend to have books with lower average ratings, suggesting that they have higher shared standards.
- **Context:** We compare our results to genres collected on another crowdsourced website, Wikipedia, and put our results in conversation with work on boundary objects and other collaborative tagging systems. We find that tags and genres are digital sites of both cooperation and disagreement.

2 RELATED WORK

2.1 Tagging Systems

A *collaborative tagging system* allows multiple users in a community to tag the same object, and aggregations of these tags are then shown as features of the object [38]. These tagging systems are also referred to as *folksonomies*, a neologism for "folk taxonomy" [47, 48, 54]. Crucially, collaborative tagging systems and folksonomies rely on *uncontrolled* vocabularies rather than pre-defined hierarchies and taxonomies and include interacting levels of personal and community tagging.

Why do users choose to participate in collaborative tagging systems? Motivations can include organization of one's personal data as well as social recognition from other users [53] and identification of functions of the object (e.g., what the object is, who owns it) [15]. Through a set of surveys, Bartley [1] finds that LibraryThing users usually add tags for collection management, to add factual information, and to help others find books. Tagging systems can also be seen as collaborative sensemaking, "orienteering", or information foraging [26, 30, 44]. Individual tagging decisions are sometimes influenced by other users [15, 37], indicating that users learn from other users and make sense of the tagging space together. These motivations and habits will likely vary depending on the design and functionality of the website that houses a given tagging system.

Several attempts have been made to categorize the tags used in folksonomies. For example, Golder and Huberman [15] proposes three tag classes—*factual*, *subjective*, *personal*—which are used in later work by Sen et al. [37] to categorize movie tags. Heymann et al. [18] divide the tags into six types—*objective and content-based*, *opinion*, *personal*, *physical*, *acronym*, and *junk*—and find that the majority of LibraryThing and Goodreads tags are objective and content-based, while Goodreads has more personal tags than LibraryThing.

Other work seeks to categorize the taggers themselves. For example, Körner et al. [22] divide users into *categorizers*, who use a small set of hierarchical tags, and *describers*, who use many creative and non-hierarchical tags. They explore the emergent semantics in collaborative tagging systems and find that *describers* contribute more than their more rigid counterparts. Some work has found that users tag independently of other users [32], while Zubiaga et al. [60] finds that certain groups of users assign higher quality tags that are more useful for tag prediction systems.

Much prior work has focused on tagging systems as problems to be solved. If the tags are to be used as input for the creation of canonical systems and hierarchies, then the tags should be normalized. Hypothesized synonyms should be conflated and ambiguities should be resolved to enhance information retrieval, recommendation, automatic tagging, and ontology construction [19, 23, 60]. For example, Heymann et al. [18] emphasize three qualities of collaborative tagging systems—consistency, quality, and completeness—and compare to systems designed by experts.

We take a different view. Rather than learning a hierarchy, we want to use the associations of users to learn nuances about their understanding of genre and usage of tags. The point of collaborative tagging is to escape the hierarchical view of data and instead favor an inclusive, flexible structure [15]. The non-hierarchical tagging system allows each object to be about several things simultaneously [15]; this quality is exactly what allows genres on LibraryThing to overlap and intertwine. This overlap allows us to learn about cooccurrences, correlations, and relationships between genres according to a community in ways previously not possible.

2.2 Online Book Reviews

The internet and social media have greatly increased the amount of available evidence about readers and reading communities. Earlier research about readers relied on sources such as archival materials (e.g., personal diaries), ethnographies, and surveys [28, 33]. These sources typically offer rich data about a small number of readers or more cursory data about a large number of readers, with little in between. Online book review websites such as LibraryThing and Goodreads, where readers publish records of their thoughts in their own words and organically form social bonds with other readers, offer invaluable resources for the study of readers and reading communities.

Researchers in the fields of digital humanities and cultural analytics have started to take advantage of online book ratings and reviews to study readers, though they have mostly focused on Goodreads data. For example, the Stanford Literary Lab uses Goodreads ratings as metrics for general book popularity among readers [31]. Bourrier and Thelwall similarly use Goodreads ratings and reviews to understand the contemporary reception of 19th-century literature [4], while English et al. [10] explore the overlap between Goodreads users who read "popular" books and users who read "prestigious" books.

A variety of predictive tasks have also been studied in the context of book reviews. These include popularity prediction [25] and automatic recommendation systems that incorporate user specialties [52]. Resources such as the UCSD Book Graph, a dataset of scraped and labeled Goodreads reviews and user data, are intended for the tasks of item recommendation [50] and spoiler detection [51]. Unlike these works, we use prediction only as a lens and not a tool to enforce an ontology.

2.3 Mapping Online Communities

In our study of LibraryThing, we are interested in both the collective LibraryThing community and in the sub-communities that gather around genres. In particular, we are interested in mapping these sub-communities in order to compare them and learn how their members understand genre. The growth, dynamics, and interplay of online communities has received extensive attention from the research community. Prior work has analyzed membership life cycles, loyalty, expertise levels [7, 17, 27, 35], as well as automatically identified user roles and the typical progression between these roles [58]. Other work has shown that genealogies of communities are predictive of community life cycles [43].

Measuring and mapping communities has taken several forms. Cunha et al. [6] compare a variety of success measures for different communities, including membership growth, membership retention, community survival, and activity volume; we will use a similar set of popularity and activity metrics as an initial mapping of the genres. Zhang et al. [59] map groups to a two-dimensional space, where the axes are distinctiveness (specialized language) and dynamicity (temporal variation in topics). We do not account for dynamicity in our models, but we take a similar approach in modeling language specificity and mapping communities for comparison.

2.4 Genre

Many readers understand "genre" as a way of classifying literary works based on shared textual characteristics, such as similar plot structures, character types, or settings. By this logic, if a novel takes place in outer space, then it might be classified as *science fiction*. Or if a novel features a detective as its main character, then it might be classified as *mystery* or *detective fiction*. This conception of genre is reflected in Wikipedia descriptions of genres, unsurprising given its descriptive goals as a popular encyclopedia.

Yet many literary scholars resist understanding genre as a neat classification system. They emphasize that genres are blurry, change over time, and depend on context [29]. There is no "master list" of genres [14]. Even scholars such as Underwood and Wilkens, who have demonstrated that computational models can detect surprisingly clean boundaries between genres, insist that genre cannot be easily defined [46, 55]. They rely on computational classification precisely because other forms of classification fail. Genre, according to other scholars, is not something that books *have*, or something that can be found in the texts themselves. Rosen and Pavel argue that genre is a tool that authors use to write books, akin to a "set of recipes" [29, 34]. From another angle, Radway and others have explored genre as a product of the publishing industry, as categories that are used to market and sell books [33].

Within the fields of natural language processing and computational social science, research has focused on learning fixed genre categories from texts. A variety of approaches have been proposed for automatic genre identification [2, 20, 39, 57], most focusing on book-length texts as training data. These works raise the question of what genre is: is it a set of surface level facets [20] or is abstraction required [57]? Genre has also been successfully incorporated into book recommendation systems [24] and used for analysis of emotional and narrative arcs [21]. While we are similarly focused on genre definitions, similarities, and boundaries, we focus not on the book texts but on user reviews and tags; our goal is not to predict the "correct" genre label but to learn from users about genres are understood and used in the LibraryThing community.

3 DATA DESCRIPTION

3.1 LibraryThing: A Web of Books

LibraryThing[1] is a social website where users can track their reading, share book reviews and ratings, create personal and shared collections of books, and interact in forums. As of May 2020, LibraryThing boasts 2,519,906 members and 3,853,370 reviews for 1,389,542 works. We were unable to locate demographic data for LibraryThing, but the *Local Statistics*[2] page on LibraryThing indicates that the majority of associated bookstores and libraries are in the United States, with smaller but significant numbers in the United Kingdom, France, Canada, Japan, Germany, and Argentina.

LibraryThing is an explicitly encyclopedic website, providing both an accessible wiki called Common Knowledge[3] and professional cataloging services, TinyCat,[4] designed for small working libraries. A major section of the website, labeled *Zeitgeist*,[5] contains ranked lists and statistics describing the web of books, users, and authors. Users can document fine-grained metadata for the books in their collections, including the physical description of the book (e.g., dimensions, weight), official identification numbers, and reading dates. Users can also volunteer as *Helpers* to curate works and authors, fight spam, and contribute information, which can earn them profile badges for their efforts. These qualities make LibraryThing an extraordinarily rich source of data and suggest that its tagging system has broader consequences for many small library catalogs.

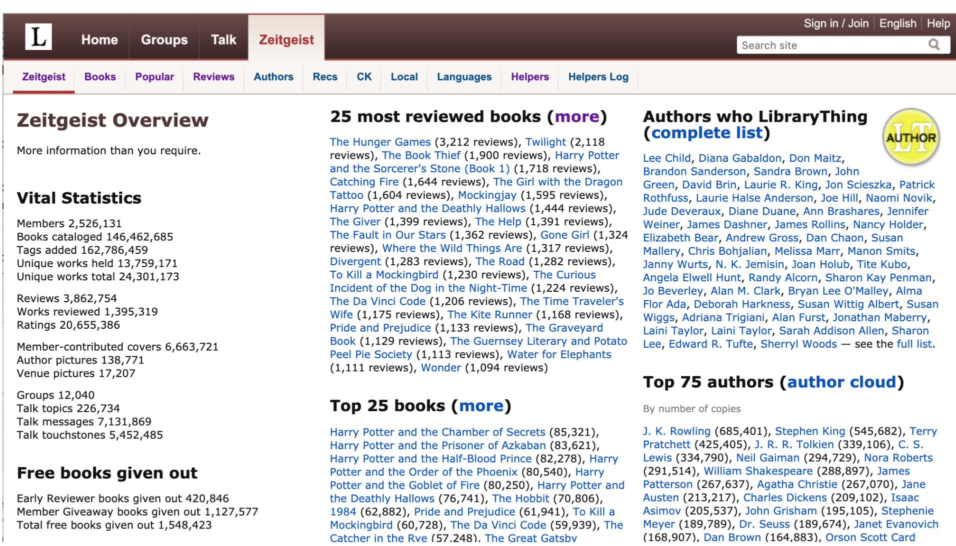

Fig. 1. A screenshot of the LibraryThing Zeitgeist (accessed May 28, 2020).

Tagging plays a central role in organizing both the global library of works and each user's reading profile. Genres and tags overlap; not all tags correspond to genres, but all genres correspond to at least one tag. Each user profile comes pre-equipped with six *collections—Your Library*, *Wishlist*, *Currently reading*, *To read*, *Read but unowned*, *Favorites*. But other tags are organic, free-text fields that users can use creatively. While many users stick to traditional genre tags, others use humorous

[1] https://www.librarything.com/
[2] https://www.librarything.com/local/stats
[3] https://www.librarything.com/commonknowledge/
[4] https://www.librarycat.org/
[5] https://www.librarything.com/zeitgeist

or niche tags, tag their books with year numbers (often to catalogue their reading over time), or use specialized tags unique to the user. Prior work has found that users have various motivations underlying their tagging practices, as we discuss in §2.1.

LibraryThing shares many similarities with Goodreads[6], a larger book reviewing website with over 120 million members. The ability to personally and collectively tag books plays an important role on both websites, though LibraryThing divides these tasks between *collections*, *tags*, and *lists*, while Goodreads focuses on *tags* (also referred to as *shelves* and *genres*) and *lists*. Goodreads was acquired by Amazon in 2013, and while a portion of LibraryThing was purchased by Abe Books (since also acquired by Amazon), LibraryThing has remained independent and ad-free. Unlike Goodreads, until May 18, 2020, LibraryThing charged its users a $10-$25 fee to catalog more than 200 books. It is now free for all users, and LibraryThing relies on its cataloging services for income. This focus on paid services rather than selling user data has likely influenced both users' perceptions of the website and the design of the website, allowing LibraryThing to maintain data-rich features that would perhaps be overwhelming and unappealing to the average Goodreads user.

Most importantly for data science research, Goodreads throttles the number of visible reviews for each book. Only 300 reviews are visible for a limited set of sort options (e.g., newest reviews, oldest reviews) [49]. Since many books have thousands of reviews, this means that for both the users of the website and researchers seeking to study review texts, the majority of the data is inaccessible. Our decision to study LibraryThing partially stems from a desire to include the complete set of reviews for each book in our analysis, rather than a biased sample.

3.2 Data Collection

As described above in §3.1, LibraryThing contains an enormous number of books and reviews. After years of user input, it also contains an enormous number of tags: over 167 million. We find that these tags form a "long tail" in which the majority of the tags are applied to a very small number of books. We cannot analyze all of these tags, both because of lack of space in this paper and because many of the tags are not associated with enough reviews to make reliable comparisons with other tags. As we discuss in §3.3, our methods require that we control for several review characteristics—including rating polarity, review length, and book title—and most tags do not have enough data to properly control for all of these features. Restricting and holding constant our target tags also allows us to more easily make comparisons across different metrics. While the unconstrained, creative use of tags is part of what makes LibraryThing genres so interesting, we must find ways to scope down the tags for analysis.

Therefore, we manually identify a set of 20 target genres (shown in Table 1) by examining the most frequent 75 tags on LibraryThing. We discard tags that are too broad (e.g., *fiction*, *to-read*) or that are near duplicates of other tags (e.g., *classic* and *classics*).[7] Prior to our collection, LibraryThing already combined some synonymous tags, e.g., the *fantasy* tag includes *Fantasy*, *fantasia*, *fantasía*, and *FANTASY*. We choose these target genres rather than more creative or user-specific tags, because we are interested in how LibraryThing users re-imagine more conventional literary genres. While this decision leaves unexplored many areas of LibraryThing, and perhaps could be read as re-imposing traditional genres on the collaborative tagging system, we see these target genres as both touchstones and starting points. But there are also surprising and unconventional genres even in the most frequent 75 tags, like *vampires*, *family*, and *animals*, which do not fit traditional or scholarly conceptions of genre.

[6]https://www.goodreads.com/

[7]There are cases where tags that are close in name operate very differently; e.g., books tagged *french* are usually books written in French while books tagged *france* are usually books set in France.

We scrape metadata for the 1,000 top books for each of these 20 target genres, where "top" books are those that have most often received the target tag. Scraped book metadata includes the title, author, rating distribution, publication date, and tag cloud (counts for all the tags that all users have applied to the book). We scrape the full set of public reviews (review text, user ID, date, star rating) for each book, and for each reviewer, we scrape their public tag cloud (the tags they have personally applied). This results in a total of 17,440 books, 319,850 reviews, and 33,849 users.

Genre	**Mean Review Length**	**Lexical Density**	**Top 5 Most Tagged Books**
politics	312 words	8.2	*The Prince, The Communist Manifesto, Animal Farm, The Audacity of Hope: Thoughts on Reclaiming the American Dream, 1984*
classics	290 words	7.9	*Pride and Prejudice, The Odyssey, Jane Eyre, The Iliad, The Great Gatsby*
science fiction	289 words	8.2	*Ender's Game, Dune, The Hitchhiker's Guide to the Galaxy, Fahrenheit 451, Neuromancer*
psychology	278 words	8.2	*Blink: The Power of Thinking Without Thinking, Man's Search for Meaning, The Man Who Mistook His Wife for a Hat and Other Clinical Tales, Man and his Symbols, Quiet: The Power of Introverts in a World That Can't Stop Talking*
romance	275 words	8.9	*Pride and Prejudice, Twilight, New Moon, Jane Eyre, The Time Traveler's Wife*
young adult	274 words	9.2	*The Hunger Games, Harry Potter and the Sorcerer's Stone (Book 1), Harry Potter and the Chamber of Secrets, Harry Potter and the Half-Blood Prince, Harry Potter and the Goblet of Fire*
vampires	273 words	9.2	*Twilight, New Moon, Eclipse, Dracula, Breaking Dawn*
horror	271 words	8.6	*Dracula, Frankenstein, The Shining, It, Salem's Lot*
historical fiction	267 words	8.4	*The Pillars of the Earth, The Book Thief, The Other Boleyn Girl, Outlander, Memoirs of a Geisha*
biography	260 words	8.3	*John Adams, Alexander Hamilton, Steve Jobs, Truman, The Diary of a Young Girl*
fantasy	258 words	8.7	*The Hobbit, Harry Potter and the Sorcerer's Stone (Book 1), Harry Potter and the Chamber of Secrets, Harry Potter and the Prisoner of Azkaban, Harry Potter and the Half-Blood Prince*
memoir	252 words	8.3	*The Glass Castle: A Memoir, Angela's Ashes, Running with Scissors, Reading Lolita in Tehran: A Memoir in Books, Eat, Pray, Love*
family	241 words	8.5	*Love You Forever, The Relatives Came, Guess How Much I Love You, A Chair for My Mother, The Lovely Bones*
graphic novel	239 words	8.3	*Watchmen, The Sandman Vol. 1: Preludes and Nocturnes, Persepolis: The Story of a Childhood, Maus I: A Survivor's Tale: My Father Bleeds History, V for Vendetta*
mystery	238 words	8.3	*The Da Vinci Code, And Then There Were None, The Girl with the Dragon Tattoo, Murder on the Orient Express, Angels & Demons*
crime	238 words	8.2	*The Girl with the Dragon Tattoo, In Cold Blood, The Girl Who Played with Fire, The Girl Who Kicked the Hornet's Nest, The Devil in the White City*
humor	225 words	7.6	*The Hitchhiker's Guide to the Galaxy, Good Omens: The Nice and Accurate Prophecies of Agnes Nutter, Witch, Me Talk Pretty One Day, The Restaurant at the End of the Universe, America (The Book): A Citizen's Guide to Democracy Inaction*
children	216 words	8.7	*Harry Potter and the Sorcerer's Stone (Book 1), Harry Potter and the Chamber of Secrets, Harry Potter and the Prisoner of Azkaban, Harry Potter and the Goblet of Fire, Harry Potter and the Half-Blood Prince*
animals	195 words	9.1	*Brown Bear, Brown Bear, What Do You See?, Charlotte's Web, Watership Down, All Creatures Great and Small, The Mitten*
picture book	175 words	9.8	*Where the Wild Things Are, The Very Hungry Caterpillar, Goodnight Moon, If You Give a Mouse a Cookie, Alexander and the Terrible, Horrible, No Good, Very Bad Day*

Table 1. The mean review length and the lexical density for each of the 20 target genres. Measurements shown are the result of data sampling (§3.3). Lexical density indicates the number of unique words divided by the total number of words; genres with lower scores indicate genres with greater lexical diversity.

Total Unique Tags	12,115 unique tags
Number of Target Tags	20 target tags
Mean Unique Tags per Book	30 unique tags / book
Mean Total Tags per Book	2,483 tags / book
Mean Unique Tags per User	452 tags / user
Mean Total Tags per User	5,186 tags / user

Table 2. Tag statistics for the sampled dataset.

Genre	**Sub-Genres**	**Related Tags**
romance	paranormal romance, historical romance, Regency romance, contemporary romance, necromancer	historical romance, Regency, paranormal romance, erotica, werewolf
fantasy	urban fantasy, Science Fiction/Fantasy, epic fantasy, dark fantasy, high fantasy	Regency romance, NF, calibre, historical mystery, Pulitzer Prize
historical fiction	alternate historical fiction, Tudor historical fiction	historical mystery, historical, historical novel, Middle Ages, historical romance
science fiction	Science Fiction/Fantasy, military science fiction, classic science fiction, feminist science fiction, science fiction romance	Terry Pratchett, futuristic, sf, dystopian, dystopia
children	children's, children's literature, children's fiction, children's books, children's book	children's classics, board book, Newberry, juvenile fiction, historical romance
young adult	young adult fiction, young adult literature, young adult fantasy, adults& young adults ficiton, genre: young adult	Newberry, children's classics, YA, Newbery Medal, Newbery
humor	dark humor, humorous, black humor, political humor, humorous fiction	Terry Pratchett, dark fantasy, erotica, borrowed, calibre
vampires	Morganville Vampires, morganville vampires series, Chicagoland Vampires, Argeneau Vampires, New Tales of the Vampires	vampire, werewolves, werewolf, paranormal romance, paranormal
picture book	wordless picture book, _Picture Books, children's picture book, _Hardback Picture Books, picture book: easy	Caldecott Medal, Caldecott Honor, collection:Fiction, shelf:Fiction, colors
family	family saga, dysfunctional family, family history, family secrets, family relationships	siblings, Regency romance, erotica, anthropology, pb
graphic novel	graphic novels, Comics & Graphic Novels, Graphic Novels & Comics, comic/graphic novel	graphic novels, Vertigo, comics, superheroes, comic
psychology	social psychology, evolutionary psychology, neuropsychology, positive psychology, Jungian psychology	personal development, self-help, neuroscience, psychiatry, brain
biography	autobiography, biography/memoir, Biography & Autobiography, autobiography/memoir, literary biography	autobiography, memoir, presidents, American Presidents, US history
memoir	biography/memoir, graphic memoir, autobiography/memoir, travel memoir, lady trent's memoirs	autobiography, biography, essays, travel, Islam
horror	horror fiction, gothic horror, survival horror, classic horror, Christian horror	Stephen King, zombies, king, dark fantasy, Neil Gaiman
classics	children's classics, Penguin Classics, Harvard Classics, NYRB Classics, Christian Classics	children's classics, classic literature, 19th century literature, classic fiction, allegory
crime	crime fiction, true crime, crime and mystery, Nursery Crime, Hard Case Crime	Agatha Christie, Christie, crime fiction, police procedural, historical mystery
mystery	historical mystery, murder mystery, British mystery, mystery-thriller, cozy mystery	Agatha Christie, Christie, crime fiction, detective, police procedural
politics	American politics, US politics, geopolitics, world politics, Christianity and politics	presidents, American Presidents, communism, political science, US history
animals	farm animals, talking animals, stuffed animals, zoo animals, forest animals	bears, mice, pets, board book, cats

Table 3. Users provide extensive sub-categorization in target genres. Sub-genres are the most frequent tags containing the genre n-gram. Related tags are the tags with the highest pointwise mutual information (PMI) with the target genre (omitting tags with low frequency).

The top books for each tag are not mutually exclusive. For example, a top book for the tag *fantasy* might also be a top book for the tag *science-fiction*. Even if the book is tagged *science-fiction* more often than *fantasy*, we will still add the book to the *fantasy* genre if its fantasy ranking is in the top 1,000. In other words, the top books are the most popular books for the tag, not the books most specific to the tag. However, Table 1 shows that the top books are still distinctive.

3.3 Genre Sampling

We find significant differences between the target genres, including mean review length, vocabulary size, mean star rating, and mean number of ratings. For example, *picture books* have a very high mean star rating and a very low mean number of ratings, while *horror* has a higher number of ratings but a much lower mean star rating. Users infrequently review *picture books*, but when they do, they rate them very positively, whereas users tend to be more critical of *horror* books, even though they review them more overall. For most genres, the vocabulary size is correlated with the mean length of the reviews, but outliers include *vampires* and *young adult*, which have small vocabulary sizes given their mean review lengths. These outliers suggest that reviews of *vampires* and *young adult* books tend to discuss more similar subjects in similar ways. In order to compare the genre features of research interest, we use the following sampling sequence to control for features like review length which are not of interest. This method also controls for the influence of extremely popular books such as the *Twilight* series. Results of this sampling are shown in Table 1.

We remove reviews without ratings, reviews not written in English (using a simple filter requiring at least five English stopwords and fewer than five Spanish stopwords), duplicate reviews (where duplicates require identical review IDs, user IDs, and book IDs), and reviews with fewer than 100 words. To control for polarity, we randomly sample two positive and two negative reviews for each book. We define negative reviews as those with ratings between 0.5-3.5 stars and positive reviews as those with ratings between 4-5 stars. We choose a higher cut-off for negative reviews, rather than choosing the midpoint 2.5, because there is a strong skew across the book reviews towards positive ratings, and qualitatively, we find that a rating of 3.5 stars usually indicates serious criticisms of the book.

To control for the review length, which can vary significantly by genre, we retain only the last 100 words of each review text. This is a common preprocessing step in NLP analyses of texts with variables lengths; for example, see the discussion and sampling decisions in Danescu-Niculescu-Mizil et al. [7]. Controlling for review length is particularly important in our analyses of diversity of themes present in reviews (e.g., our use of topic entropy in §6), as longer reviews could by nature of their length contain more diverse language. In the case of online book reviews, we observe that reviewers are more likely to begin reviews with meta-content (e.g., where they read the book, personal stories unrelated to the book) while they are more likely to end the reviews with summaries of their thoughts, re-stating the different themes mentioned earlier. We use the last 100 words because our analysis is focused on the reviewer's judgements of the books.

Books that do not meet these all of these filtering requirements are discarded. Of the remaining books, we randomly sample 300 books per genre. We allow books to appear in multiple categories, as this reflects the reality of genre-crossing books, and we allow multiple books from the same author, as this reflects the outsized influence of prolific authors. Our sampling results in a total of 4,934 unique books (100 words per review, 2 reviews per polarity per book, 300 books per genre).

3.4 Data Description

We observe some general patterns in the tags gathered via our sampling method. Table 2 shows the number of tags across the sampled dataset, as well as the number of tags per user. On average, tags are used heavily by the users in our dataset, with a mean of 452 unique tags per user and a mean of

5,186 total tags per user. This high average is driven by superusers who have tagged thousands of books; many of these users are collectors, bookstore owners, or librarians.

One of the most exciting aspects of online tagging, for both researchers and users, is users' freedom to tag books as they wish and to create their own tags. This allows traditional genres to become associated with different sets of books, and it also allows micro-genres to proliferate. Table 3 shows a sample of tag associations for our target genres, including sub-genres (genres containing the target genre's n-gram) and related tags (tags whose PMI with the target genre is high). The sub-genres sometimes represent a particular series or author (e.g., sub-genres for *vampires*), while sub-genres for the *classics* genre indicate a frequent preoccupation with official lists. Some indicate the overlap of two genres (e.g., the frequency of the word "memoir" in the *biography* sub-genres), but many do appear as true sub-genres (e.g. *romance*, *fantasy*, *science fiction*, *psychology*).

LibraryThing reviews range in length and vocabulary size, with some genres featuring longer reviews and bigger vocabularies. Reviews often touch on many different themes, including plot, characters, writing style, the author's biography, and why the reviewer read the book. For example, in a review of the novel *Flowers for Algernon* (1966), the reviewer discusses their emotional reaction to the book, mentions awards the book has received, offers a plot summary, and reflects about the social stigma surrounding dementia:

> *It's knocked my socks off, thrown me in the corner and left me a crumbling, emotional wreck...the character and story development, the writing, the way in which it stirs the emotions and its sheer humanity have all hit the right spot. It is also an emotional study into how a person may react to the possibility of the onset of dementia...Yes, Flowers for Algernon has won awards for science fiction and yes it is in the SF Masterworks list but ultimately it is a story of humanity and a person struggling to gain acceptance for who he really is not for who others want him to be.* —lilywren

We use the texts of the reviews to study the relationship between these themes and genres, and we also use these review texts to measure how surprising the review is given the genre. By focusing on the review text rather than the book text, we interpret genre as it is received, understood, and used by readers—rather than how it is used or understood by authors, critics, and publishers.

4 SHIFTING BOUNDARIES: MEASURING GENRE OVERLAP

Because of the open lexicon of the collaborative tagging system, the shape and size of the genre space is always changing and growing. In order to measure genres, we cannot rely on any cardinal directions; instead, we need to set the tags in relation with one another, which allows us to infer connections and distinctions that may or may not be obvious. How can we map genres into a single space where we can draw comparisons between them?

We first display some descriptive statistics of genre popularity: the number of ratings and mean star rating. Figure 2 shows the genres plotted along these two measures of popularity. These scores represent the full set of all ratings for the books included in our sample, but not the full set of books on LibraryThing. There is no clear relationship between these variables, but we can observe that genres have different properties. Genres targeted towards children (*picture book*, *children*, *animals*) are the most positively rated, but vary in rating frequency. The most popular genres, by both metrics, are *children*, *fantasy*, and *family*. The lowest rated genres include *mystery*, *crime*, and *horror*, while non-textual formats (*picture book*, *graphic novel*) are the least popular in terms of frequency. Even with these basic statistics we can find differences between seemingly similar genres: *memoir* is less positively rated than *biography*, while *fantasy* is much more positively rated than *science fiction*. These results reinforce that there are measurable differences between genres, even though we leave open the possibility that a single book could appear in multiple genres.

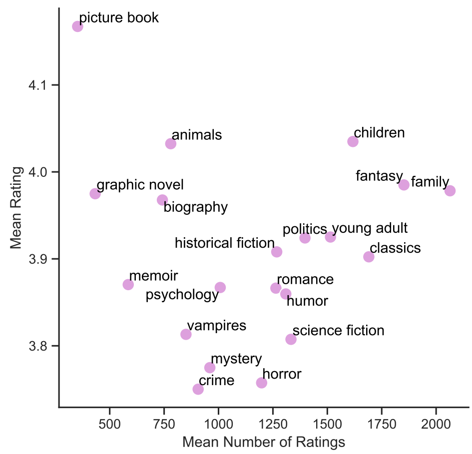

Fig. 2. The average popularity of the target genres, including the mean number of ratings per book and the mean star rating (out of 5). Results are shown for our sampled data (see §3.3).

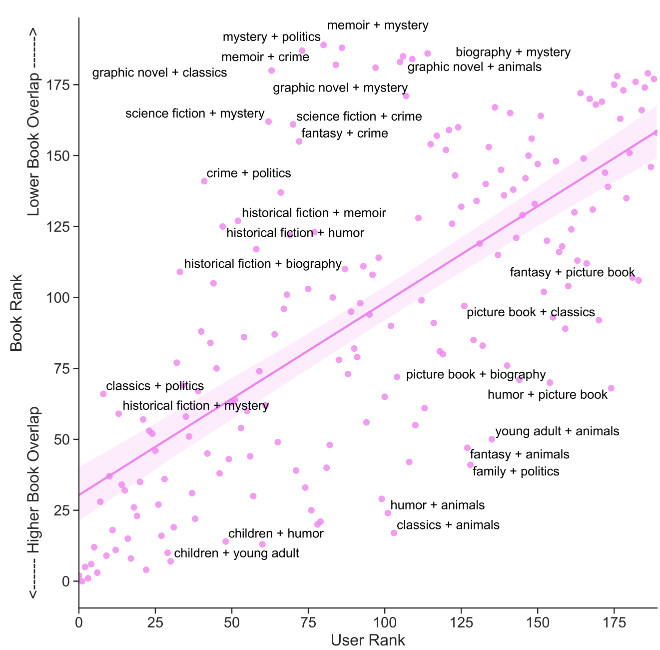

Fig. 3. User overlap between genre pairs correlates with book overlap, but there are outliers. Each point represents two genres, and the axes represent the *rank* of the genre pair, where lower numbers indicate higher ranks and therefore higher overlap. For example, the genre pair *classics + animals* has a mid-range user overlap rank and a high book overlap rank, indicating that these genres share surprisingly few users given how many books are shared. Pearson correlation between book and user overlap is significant ($r = 0.68$, $p < 0.05$).

Next, we measure genre similarity using two metrics. First, using our sampled book sets, we measure the *book overlap* between each pair of genres—that is, how many books have been tagged as both one genre and another genre. Some genre pairs share no book overlap (e.g., *politics* and *mystery*; *classics* and *graphic novel*) while others share many books in common (e.g., *children* and *animals*; *memoir* and *biography*). 24% of the genre pairs have no book overlap. Second, we measure the *reviewer overlap* between each pair of genres—that is, how many reviewers have tagged a book in one genre and a book in another genre. We convert both measurements into ranks, where the genre pair with the greatest book overlap has Rank 0.

We expect the user overlap rankings to largely mirror the book overlap rankings. If two genres share many books in common, it follows that they would also share many reviewers in common. First, the shared books will necessarily include shared reviewers and, second, the high overlap in books implies that the genres are thematically related. If a user finds one of the genres appealing, they are likely to also find the other genre appealing. However, some outliers emerge. For example, in Figure 3, we notice that *classics* and *animals* have higher book overlap than we would expect given their user overlap. In contrast, *classics* and *politics* have higher user overlap than we would expect given their book overlap.

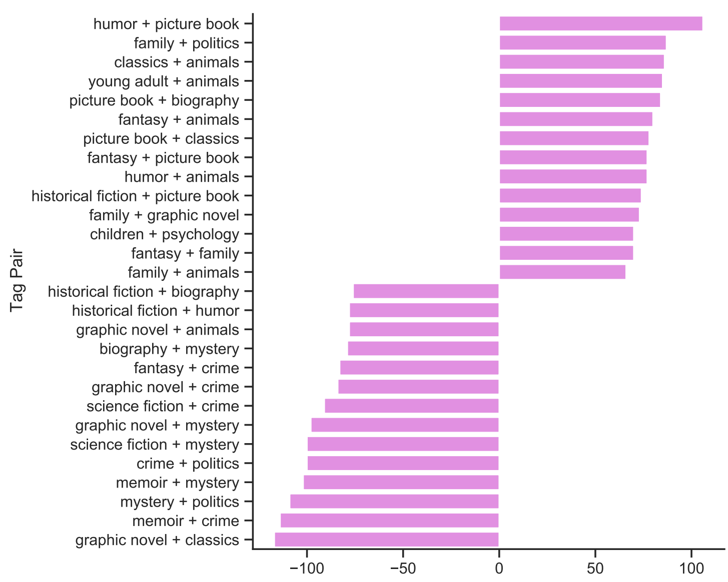

Fig. 4. The pairs of genres with the greatest difference between their book overlap and their user overlap. Overlap is represented as a ranking where rank 0 is the genre pair with the highest book overlap.

We quantify these patterns by taking the difference between the user and book overlap rankings.

$$\text{overlap difference} = \text{user overlap rank} - \text{book overlap rank} \tag{1}$$

Genre pairs with very high or very low scores are outlier pairs, which deviate from the expectation that book overlap rank should match user overlap rank. We show these outliers in Figure 4. For example, given the low number of books that have been tagged as both *graphic novel* and *classics*, it is surprising to see how many users read within both of these genres. This high user overlap could be explained by a tendency of users who review within the *graphic novel* tag to also review within the *classics* tag—or it could be explained by a tendency of users who review classics to read widely across many genres, including graphic novels. On the other hand, *humor* and *picture book* have relatively high book overlap but relatively low user overlap. Perhaps *picture book* reviewers read more frequently in other genres and only occasionally review a picture book, e.g., when they give a book as a gift or when reading a book to a child.

Other gaps help to reveal more meaningful connections between genres and identify subcommunities that straddle multiple genres. For example, *memoir* and *crime* have relatively few books in common (low book overlap), but they share many users in common (high user overlap). The low book overlap seems to indicate that users draw a clear distinction between these two genres, which we speculate might involve a distinction between fiction and non-fiction. *Crime* novels are often fictional while *memoirs* are often non-fictional. This divide is reflected in the Wikipedia genre hierarchy, which separates *crime* and *memoir* into separate supergenres. Nevertheless, these genres share many reviewers, suggesting that there are significant similarities between them. Both genres are also commonly written in a sensationalized style, and both genres focus on psychology and personality. The high user overlap clues us into reader interests that span different genres.

5 QUANTIFYING LEXICAL FIT: "MISTAKES" AND SURPRISES

We have established that we can measure similarity and difference between genres in terms of their users and book overlap, but what about the reviews themselves? We find that we can classify the genre of a book being reviewed based on the text of the review alone—without the book's own text, title, or author—because users focus on different aspects (e.g., characters, plot, suspense) for certain genres in their reviews. We follow similar work that has sought to predict genres from texts [19, 46], but our training set is reviews, rather than book texts, so that we can focus on the *reception* of a book rather than its content. Note that although we are training a classifier to quantify the association between words and labels, we are not running a predictive experiment with held-out testing data, but rather an evaluation on the full data set, more like a standard linear regression. Our goal is not to maximize predictive performance, but rather simply to computationally represent ambiguity and similarity between genres. As a result, our results should be interpreted as an upper bound for predictive accuracy, and not as a measure of generalization. This approach allows us to analyze the collection in two ways: first, if reviews for two genres cannot be easily distinguished *even when the labels are available at training time*, that is evidence that they serve the same values and expectations, and second, if a review is "surprising," it may describe a setting in which a reader has a unique or idiosyncratic experience of a book.

Genre	Precision	Recall	F1	Genre	Precision	Recall	F1
graphic novel	0.83	0.81	0.82	*mystery*	0.86	0.65	0.74
science fiction	0.78	0.84	0.81	*picture book*	0.63	0.84	0.72
psychology	0.71	0.90	0.79	*romance*	0.65	0.82	0.72
biography	0.72	0.80	0.75	*classics*	0.74	0.67	0.70
animals	0.90	0.61	0.73	*historical fiction*	0.70	0.70	0.70
crime	0.76	0.78	0.77	*fantasy*	0.62	0.78	0.69
politics	0.81	0.74	0.77	*young adult*	0.72	0.66	0.69
vampires	0.72	0.77	0.75	*humor*	0.67	0.66	0.67
memoir	0.78	0.73	0.75	*family*	0.68	0.59	0.63
horror	0.78	0.69	0.74	*children*	0.52	0.60	0.56

Table 4. Results for the genre classifier are above random, but nowhere close to perfect. We report results on training data to provide an upper-bound on learnability: if a classifier cannot make a correct prediction even with full access at training time, the distinction is difficult. F1 represents the harmonic mean of precision and recall; higher scores indicate better classification performance and reviews that are more systematically lexically distinct. Macro F1 across all of the genres was 0.73.

We train a supervised classifier on the review texts, using as labels the genre of each review in our sampled data. We use a logistic regression (one-vs-all) model with TF-IDF weighted unigram features, using the last 100 words of each review (as described in our sampling procedure in §3.3). Genre labels are the most frequent tags assigned to the book, after filtering out a small set of high level tags that do not resemble genres.[8] We measure the *surprisal* of the review text given the genre using the probability of the true label: surprisal = $1 - P(\text{true label})$. High surprisal scores indicate that the predicted probability of the true label was low and that the review was difficult to classify as its target genre. Low surprisal scores indicate that the predicted probability of the true label was high and that the classifier was able to predict the review's target genre. This method generally works well at identifying reviews for books that blend different genres. When averaged, these scores can tell us which genres blend more with other genres.

Using this simple classification method, we find a macro F1 score of 0.73 across all of the genres on our training data, indicating that genres are learnable from text above random guessing (see Table 4 for full results). However, we emphasize that we are using the classifier as a tool to explore the relationship between genres, not as a reliable genre predictor (in fact, we are interested in its inconsistencies). We use the classifier to obtain surprisal scores for each review in our dataset. In contrast to many classification setups where we are concerned about *overfitting*, in this case we are interested in *underfitting*, such that some genre classifications are not easily learnable by classifiers even with full information. The classifier acts as one of many possible metrics of lexical similarity and allows us to probe where two genres' reviews might have some level of semantic overlap.

Figure 5 shows the relationship between misclassification counts and book overlap for each pair of genres. The high number of misclassifications between *memoir* and *biography* seems to conform to our expectations, as both genres are stories of a person's life. Similarly, the frequent misclassifications of reviews for the *animals*, *picture book*, and *children* genres points to their commonality, as these sets of genres have high book and reviewer overlap. By comparing to book overlap, we can identify pairs with unusually high or low numbers of misclassifications given their similarity. For example, *romance* and *horror* have an unusually low number of misclassifications given their high book overlap, while *animals* and *psychology* have an unusually high number of misclassifications given that they share no books in common.

Often, the classifier's mistakes indicate similarities and overlaps between genres. But on other occasions, the classifier's mistakes indicate a mismatch between the reviewer's priorities and the typical priorities for that genre. For example, the following review of Ann Bronte's novel *The Tenant of Wildfell Hall* (1848) was misclassified as *psychology* when the book was actually tagged as *romance*:

> *I was in awe of Anne Bronte's ability to tell such a relevant story in 1848. There are so many women who find themselves in the same situation today. She was young and naïve when she married Arthur Huntingdon and by the time she learned his true character it was too late. The writing is wonderful and for me that story pulled me in completely. The author tells the story from Gilbert's point-of-view at times and from Helen's at other times. The changing narrative flowed well and never rang false.Bronte covers some intense subjects in the book. In addition to infidelity and alcoholism, she makes some disturbing observations about women's rights during this time period. Sometimes it's easy to forget how far we've come in the last few years.*
> —bookworm12

The reviewer focuses on elements of *The Tenant of Wildfell Hall* that pertain to the characters' psychological states and mental and physical health, as well as how these conditions relate to broader society of the 19th century and of the present. A review that more easily conformed to

[8] [*fiction, non-fiction, to-read, ebook, kindle, literature, unread, own, hardcover, wishlist*]

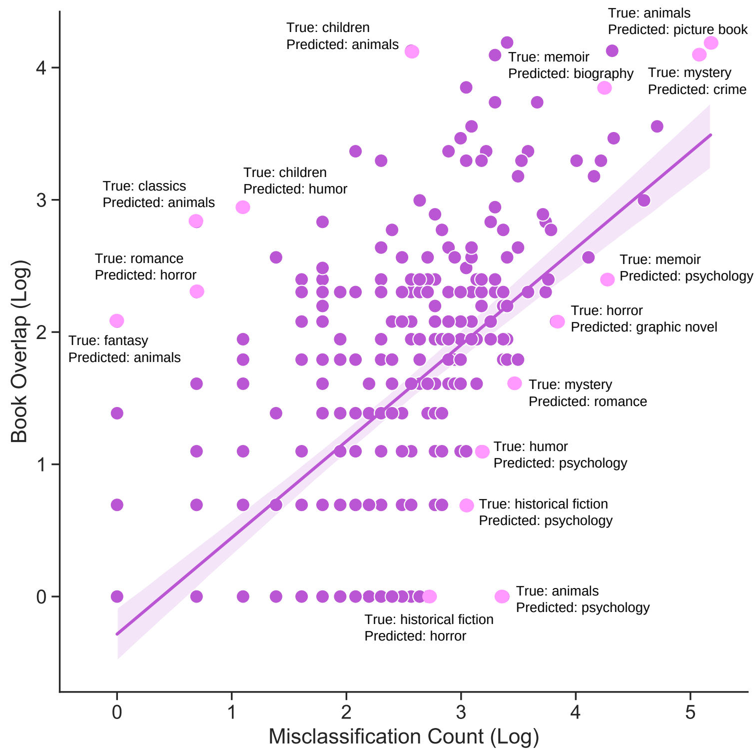

Fig. 5. The number of overlapping books and the number of genre misclassifications of user reviews for each pair of genres. Each point represents a pair of genres in which one is the *true* tag applied to the review text and one is the *predicted* tag from our model. As expected, we find a significant relationship using Pearson correlation ($r = 0.65$, $p < 0.05$) between the book overlap and misclassification count, but we highlight outlier genre pairs, e.g., *animals* and *psychology* have an unusually high misclassification count given their very low book overlap.

the *romance* genre might have discussed the ending, the romantic plot, or the attractiveness of the characters. But these are not the elements that this particular reviewer discussed. The surprisal scores thus helps us better understand the elements that readers seem to really care about or gravitate toward in a particular genre, as well identify and interpret outlier reviews for the genre.

We show examples of the classifier output and surprisal scores in Table 5. For example, we show a review excerpt of Anna Sewell's *Black Beauty*. The most popular tag for this book was *animals* but our model misclassified this review as *classics* with high confidence, resulting in a high surprisal score. The reviewer writes about the book's popularity and compares its sales rates to well-known classics. The misclassification, in other words, flags the distinctiveness of both the book and the review, suggesting that what shapes genre perception is likely more than the text itself, which is discussed in the context of other books tagged as *classics*.

We can also arrive at single surprisal score for each genre by taking the mean of the surprisal scores for the reviews assigned to that genre. The most surprising genres include *young adult*, *family*, *classics*, *children*, and *fantasy*. Genres that have higher mean surprisal scores are harder to classify; these genres are "fuzzier" and the language used in the reviews for these genres is more wide-ranging. These genres are often mistaken for similar genres, but it could also be the case that these genres are simply broader. For example, the *classics* genre contains a wide range of themes and discourses, in both its books and reviews. The high surprisal scores emphasize the view of genres as fuzzy, overlapping tags, rather than the rigid hierarchy sought by Wikipedia editors. We use these summary surprisal scores later to compare with community homogeneity (§7).

True Genre	Predicted Genre	Surprisal	Example Misclassified Reviews
romance	*romance*	0.00	*I am not normally a fan of romance novels as I find them too mushy and cutesy, but this one had a sense of humor about it that I really enjoyed...The heroine was very independent and snarky and the main romance was full of comedic situations with a smattering of seriousness that made it seem fairly realistic for the genre. It was a book that was absorbing and fast to read.* —Arualanne (*The Perfect Rake*)
romance	*historical fiction*	0.03	*There were some moments though where I had to wonder about the historical accuracy of some of the attitudes and that broke the reading spell for me.Pretty predictable but I enjoyed the ride. Almost a 4 read for me but not quite.* —wyvernfriend (*Simply Unforgettable*)
animals	*classics*	0.20	*...it's no wonder it's been so popular since it was first published. I was surprised to learn that Black Beauty is one of the top thirty best-selling books of all time in the English language, selling over 50 million copies–more than The Odyssey, To Kill a Mockingbird, Pride and Prejudice, and Gone with the Wind...* —nsenger (*Black Beauty*)

Table 5. Examples classifications and surprisal scores. Excerpts are selected from the last 100 words of the reviews. Higher surprisal indicates greater confidence in the incorrect label.

6 VALUES AND EXPECTATIONS: MEASURING THEMATIC SIGNATURES

So far, we have summarized genres as one or two dimensional scores. This has allowed us to map genres onto an interpretable space where we can compare genres, measure their similarity, and identify outliers. However, while user and book overlap, predictive surprise, and community density are strong signals of genre similarity, they do not tell us *why* these genres are or are not similar. Learning review *aspects* might help answer this question. Aspects are themes of a review, usually focused on features of the product being reviewed; in the case of books, these might include plot, characters, and writing style. Reviews are generally written to explain a rating, not a genre tag, but by measuring the amount users choose to write about particular aspects and averaging over reviews for a specific genre, we hope to approximate which aspects are most significant for that genre. Measuring which aspects users focus on for each genre will teach us the expectations and values that the LibraryThing community attaches to each genre.

To answer these questions, we measure the thematic similarity of the review texts for our target genres. For our purposes, we take a relatively simple unsupervised approach, as we would like to discover themes rather than ordain them. We train a latent Dirichlet model (LDA) [3] on the full set of scraped reviews, removing duplicate texts. Before training, we probabilistically downsample words associated with specific genres by using the Authorless Topic Models package [45].[9] This downsampling reduces the incidence of genre-specific topics, as we are more interested in cross-cutting themes that could be important for more than one genre (e.g., a *Harry Potter* topic would not be useful outside a narrow band of genres).

We experiment with different numbers of topics and find that 30 topics produce interpretable and not overly broad or narrow topics. For readability, we remove a set of common stopwords from the topic keywords, and we assign labels to each topic through manual examination of each topic's most probable words and highest ranked documents. Table 6 shows the resulting topics. Each topic is associated with a full probability distribution over the vocabulary, and we display the ten most probable words for each topic and its hand-annotated label.

[9] https://github.com/laurejt/authorless-tms

	Label	Top Words	Example Review
0	reading log	*read, reading, time, one, first, years, [NUMBER], ago, back, library*	*"Edit: 4-4-18 Just finished, full review coming soon. I read this for book club and it was an interesting..."*
1	expectations + mixed opinions	*really, like, didn, story, read, think, felt, much, liked, would*	*"At the beginning, I loved this book! But as I read further, I went back and forth. Maybe I just didn't get it..."*
2	visual descriptions	*one, like, little, could, every, back, night, man, time, see*	*"Beautifully written and detailed descriptions of the scenery and animals. 'In the pale moonlight...'"*
3	numbers (ratings, dates, pages)	*[NUMBER], read, stars, first, pages, story, would, really, like, good*	*"I would give the first 100 pages a 2 star rating, but the next 200 pages are a solid 4 stars..."*
4	simple plots	*would, could, never, one, made, didn, thought, wanted, like, even*	*"Joey is a chipmunk and today is his friend's birthday. He organizes a party but can't find a present..."*
5	city/town/village/country setting	*new, town, life, small, city, story, people, time, house, home*	*"Maisie lives in a small village, in a rural part of the country. Her journey as an immigrant, traveling..."*
6	movie adaptations + short stories	*read, story, stories, written, one, movie, review, would, reading, novel*	*"This book was ok, but nothing compared to the TV show. If you've seen the Netflix series, then you know..."*
7	writing style (themes, coherence, prose)	*one, even, much, rather, would, seems, though, quite, might, perhaps*	*"Praise for this novel is well-deserved, despite the dense opening. Perhaps even too polished, but beautiful..."*
8	sarcastic/angry review	*like, really, get, know, even, one, think, good, much, thing*	*"I can't even with this book. The main character is a complete twit! Why the author wrote it is beyond..."*
9	strong emotional reaction by reviewer	*love, read, loved, one, story, characters, even, know, like, way*	*"I don't know where to start writing; this book broke my heart and swept me away to places I never expected..."*
10	life lessons for empathy + relationships	*people, like, things, think, one, life, way, know, make, even*	*"This is a wonderful book that helps you understand that everyone is unique. Looking from the outside..."*
11	writing style (characters, plot, pacing)	*novel, story, characters, reader, character, first, narrative, plot, time, one*	*"Adeptly written character examinations. By the end of the book, the point of view could have been switched..."*
12	coming-of-age story	*school, young, girl, high, friends, year, life, adult, girls, new*	*"Tommy is a teenager working through some complicated family dynamics. The novel follows one year..."*
13	audiobook	*read, audio, funny, fun, humor, one, like, great, voice, bit*	*"Reread by listening to the audiobook. (Streamed from my library.) Narrator made me laugh a lot, better..."*
14	life lessons from terrible experiences	*life, story, love, death, one, human, world, reader, yet, man*	*"A powerful memoir about an unimaginable nightmare. Moving and through-provoking experiences..."*
15	action plot summary	*world, one, war, find, must, evil, human, power, fight, save*	*"Action-packed fantasy! Three friends set out on a quest to save the world. The evil that awaits them is..."*
16	character development	*characters, character, plot, story, novel, main, interesting, well, much, good*	*"This book was well-written, but I couldn't empathize with the characters. The personalities aren't complex..."*
17	book background (author, history, curation)	*novel, first, [NUMBER], read, one, years, time, work, century, history*	*"This book was relatively unknown until recently. Parts had been printed in magazines, but this collection..."*
18	fantasy adventures	*story, world, tale, fantasy, stories, young, characters, adventure, magic, find*	*"Short stories about a fantasy land and one boy's struggle to find his magical talent. Interesting creatures..."*
19	romantic relationships	*love, woman, women, man, young, life, men, novel, husband, marriage*	*"A classic romance, gripping to the end. The central figure, Philip, falls irredeemably in love with... "*
20	family relationships	*family, mother, father, story, life, children, child, parents, sister, brother*	*"I felt sorrow and joy for the twin brothers, surviving without their parents and adopted by their..."*
21	short plot summaries about children	*get, gets, one, day, boy, back, girl, goes, little, home*	*"One morning a little boy and his father wake up and make waffles. The boy gets dressed and goes to school..."*
22	series + plot summary	*series, first, characters, one, still, new, story, next, much, time*	*"This book picks up where the last one ended. The first three books moved quickly, and once again, the... "*
23	emotional reaction to characters	*really, love, series, like, read, characters, character, one, also, good*	*"A fun, silly read. I really liked the romance, and the characters aren't too dramatic. Claire was my fav..."*
24	military + racial + class history	*war, american, history, people, white, time, america, black, world, country*	*"This autobiography of WWII tells a story of discrimination. At that time, companies wouldn't hire..."*
25	politics / sociology / religion / psychology	*world, human, society, science, people, one, future, new, god, also*	*"A classic treatise on historians struggling to fit ancient culture into modern times. Social and political..."*
26	mystery	*one, case, murder, mystery, death, man, dead, police, find, series*	*"A series of murders plagues the English countryside. After her encounter with the killer, Investigator..."*
27	pop psychology + self-help	*author, also, life, interesting, many, read, work, well, much, information*	*"An effective tool and a good guide with a wide-ranging perspective to relationships. Offers insights..."*
28	illustrations + child-appropriate writing style	*story, illustrations, reader, also, readers, author, liked, like, different, main*	*"Clearly for children, and in my opinion, a great introduction with easy words. The illustrations are..."*
29	suitability for school reading/activities	*would, children, story, students, could, read, great, use, also, kids*	*"Fun to read, and elementary students would enjoy this book. Summary and classroom extensions:..."*

Table 6. Descriptions of 30 topics learned via LDA and authorless topic models. Example texts are paraphrased amalgamations of the three most probable reviews for each topic.

Examination of these topics yields both anticipated results and surprises. As expected, characters and plot are frequent themes in LibraryThing book reviews. Perhaps less anticipated are the *life lesson* topics, Topic 10 and Topic 14. Topic 14 is associated with reviews that discuss powerful lessons learned from terrible and tragic circumstances—such as wars and genocides—while Topic 10 emphasizes lessons of empathy, human relationships, and the uniqueness of each person. Also surprising is the *reading log* topic, Topic 0, which chronicles the reviewer's acquisition and reading of a book. Topic 3 is associated with numbers of any form, probably due to our preprocessing of numbers into normalized placeholder tokens; it includes dates, page numbers, and star ratings. As such, it could also be considered a meta-review topic.

We also measure the entropy of the topic distribution associated with each review. High entropy scores point to reviews that range widely over many topics. For example, in a high entropy review of the novel *Aristotle and Dante Discover the Secrets of the Universe* (2012), which was tagged as a *young adult* book, one LibraryThing user began by discussing films and feminism:

> *Goddamn, this book is good. The first thing I did when I walked out of Mad Max: Fury Road was call my male best friend and asked him if all movies felt like this to him. I'd never before watched an action movie that felt like it was written for me as a woman, produced for me as a woman, meant for me as a woman, and all of that completely unapologetically.* —anonymous reviewer

But then the user quickly moved on to discuss other topics, praising the novel's romance, characters, and writing style, and finally recommending the book explicitly:

> *And added on top of that was a careful, honest examination of what it might be like to be gay as a young Mexican-American man, all the pieces of identity and adolescence and dawning understanding there are to stumble over, without that quick cloying of traditional M/M romance novels. The words, the flow, the characters – it was an amazing reading experience. This is easily one of the best books I've read in the last few years and I'll be recommending it to everyone who I think can truly appreciate it.* — anonymous reviewer

This review demonstrates that *young adult* books often generate dynamic responses about a rich bouquet of themes and subjects, part of what makes them hard to classify and leads to their high surprisal score.

We find that some of the topics are more strongly correlated with negative reviews: these include Topic 1 (expectations and mixed opinions), Topic 3 (numbers), Topic 7 (writing style), Topic 8 (sarcastic/angry), and Topic 16 (character development). Most of these are high-entropy topics, with probability distributions spread across many genres. This indicates that these negative patterns cut across genres, rather than being specific to a single genre. Indeed, we do not find genre-level differences between topics associated with positive and negative reviews. The only topic strongly correlated with positive reviews is Topic 23 (emotional reaction to characters), which has the highest probability for the genres *young adult*, *romance*, and *vampires*. All of these genres correspond to tight-knit communities of reviewers (see Figure 7); perhaps what brings them together is a shared love of certain character types.

Figure 6 shows the topic probability distributions for each genre, averaged across the sampled reviews for that genre. Some genres have predictably similar topic signatures (e.g., *animals*, *picture book*, and *children*; *crime* and *mystery*) while others are surprisingly distinct. For example, the *graphic novel* genre has a lower probability for the *illustrations + child appropriate writing style* topic than the *picture book* genre; this supports our finding in §5 that the *graphic novel* genre is more lexically distinct. The *graphic novel* genre also has greater probability on *series + plot summary* and appears to elicit more passionately angry reviews. The *classics* have a higher probability for the topic *writing style (themes, coherence, prose)*, whose top-ranked documents aim toward higher-brow literary criticism. In contrast, reviewers for *historical fiction* also focus on writing style, but their reviews are more casual and focused on characters, plot, and pacing.

Fig. 6. Topic probabilities for the target genres. Columns are normalized so that probability distinctions for topics with universally lower probabilities are still interpretable. Rows (topics) are sorted by entropy over the averaged genre distribution.

The rows (topics) in Figure 6 are sorted by the entropy of their genre distribution. Some of the topics have high entropy and cut across all of the genres. For example, Topic 0 (reading log) and Topic 3 (numbers) have near uniform distributions across the genres, which is unsurprising. Other topics have "spikier" distributions, rising sharply for specific topics. For example, Topic 13 (audiobook) spikes for *memoir* and *humor*, while Topic 20 (family relationships) spikes for *family* and *memoir*. The topics with the lowest entropy are child-specific topics, where the understood purpose and audience of the books are constrained.

We expected that the topic entropy scores would be strongly correlated with the surprisal scores described in §5. Intuitively, reviews that discuss fewer themes should be easier to classify. However, the relationship between surprisal and topic entropy is not straightforward. For example, the *graphic novel* genre has relatively high topic entropy, indicating a discussion of many different themes. But it also has a relatively low surprisal score, indicating that the texts of the reviews are easy to classify as belonging to a book in the *graphic novel* genre. Also surprising is the *children* genre, which usually aligns closely with the *picture book* and *animals* genres. It has relatively higher topic entropy but much higher surprisal than the other two genres, indicating that while its reviews are topically similar (as seen in Figure 6), its review texts are much harder to classify.

7 MAPPING GENRES BY COMMUNITY HOMOGENEITY

We turn now to a consideration of the people tagging and writing reviews. Do users "specialize" in specific genres—that is, often tag books in the same genre or write reviews that are lexically similar to other reviews in the genre? If so, how can we best measure this specialization, and what can we learn from this specialization about tagging and genre on LibraryThing? We hypothesize that there are different kinds of genre specialization. (1) A reviewer could be well-read in a particular genre and write reviews that are lexically similar to other reviews for books in that genre. (2) A reviewer could fit a genre lexically but only read one or two books in that genre. (3) A reviewer could be well-read in a particular genre but their reviews might be lexical outliers, indicating that they apply a different framework to these books from other reviewers.

We explore both of these possibilities through measures of *lexical homogeneity* and *community homogeneity* for each genre. For lexical homogeneity, we rely on the surprisal scores learned in §5. For community homogeneity, we use the personal tag cloud associated with each user that represents all the tags they have assigned to books. We filter tags that occur in fewer than 20 tag clouds, and we find the cosine similarity between each pair of normalized vectors, where each vector represents the tags used by a user who has reviewed in that genre. A high cosine similarity indicates a high degree of similarity between the reviewers. This tagging similarity could be interpreted as a similarity in reading habits.

Relying on a user's tagging history comes with some limitations. Users often tag books that they have not read, either for personal reasons (e.g., to mark the book for future reading) or as volunteer labor for the community (e.g., to add missing metadata for unpopular books). Users also employ tags for different functions, including personal cataloging (using idiosyncratic tags) and community contribution (see §2.1), and it's possible that these preferences align with different communities. However, by limiting our comparison sets to those who have written at least one review for the target genre, we enforce a lower bound on user-genre relatedness.

We might expect that genres with high lexical surprisal (reviews that are difficult to classify) are written by reviewers with dissimilar tagging habits (indicating dissimilar values and interests)—and indeed, in Figure 7a, we find that some genres (e.g., *classics*, *children*) fit this pattern. However, we do not find a significant correlation between community homogeneity and review surprisal. Unexpectedly, other genres (e.g., *picture book*) are easy to classify using the review text but their reviewers have differing tagging habits; the language used in these reviews is distinctive, but the set of reviewers is not. Other genres (e.g., *young adult*, *fantasy*) have high community homogeneity scores but also high surprisal scores. Their reviewers are similar but these reviews are more difficult to classify—they are lexically more dissimilar.

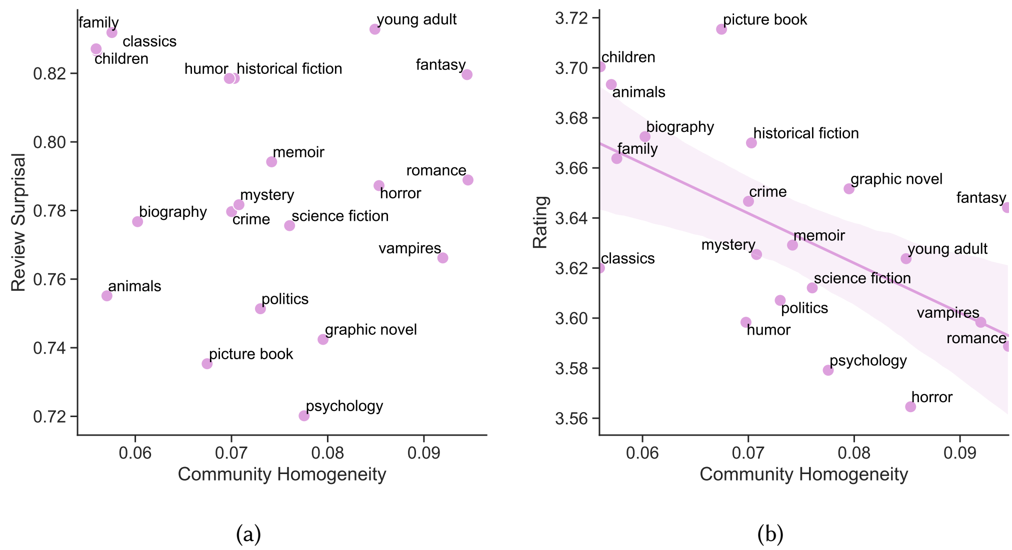

Fig. 7. Are tighter communities easier to predict? Are tighter communities more critical? Figure 7a shows the target genres plotted along surprisal (the ability of a classifier to predict the genre of a review) and community homogeneity (averaged cosine similarities between reviewers' tagsets). Figure 7b shows the target genres plotted along rating and community homogeneity. Genres whose reviewers have more similar reading habits tend to also have higher ratings according to a Pearson correlation test (r = -0.60, p < 0.05).

Figure 7b shows the genres plotted by their community homogeneity and their mean rating. We find a significant correlation between these measurements. Genres with higher rated books tend to have less similar reviewers; these include *children*, *animals*, *biography*, and *family*. Perhaps these reviewers only read that genre for specific occasions, e.g., when a book is particularly well-known and well-liked, confirming our observations when examining user overlap (see Figure 4). Genres with lower rated books, like *romance*, *vampires*, and *horror*, tend to have more similar reviewers. We hypothesize that these tight-knit communities have higher shared standards for their ratings.

8 DISCUSSION

There is not a single right way to map tags and genres in the LibraryThing community. Different maps reveal different outliers, pairings, and patterns. Reducing the rich tags to two dimensions will not answer all of our questions, but creating multiple mappings and comparing them has allowed us to tease apart some of the ways in which LibraryThing users see genre. Unlike much prior work, we do not seek to normalize the tags. While the unconstrained vocabulary of tags on LibraryThing means that "errors" like synonyms, typos, and overly personalized tags do exist, we take advantage of this information and use it to discover what is new, rather than force it to fit a traditional structure.

By exploring thematic signatures of LibraryThing genres, we learn which aspects of the reading experience are valued by LibraryThing reviewers and how these values vary depending on the genre of the book being reviewed. We discover strange similarities—e.g., the resemblance between *young adult* and more "adult" genres like *horror*—and we also find peculiarities in the topic signatures of strongly related genres, as in the case of *memoir* and *biography*. These patterns connect to a broader set of themes which we discuss below.

Audience and Reception. There are many parallels between our work and recent digital humanities studies of genre. For example, our use of classification to measure an aspect of literary style is similar to Underwood [46] and our attempts to map genres are similar to the book clusters in Wilkens [55]. Our approach is founded in this tradition, which often uses computational tools in a non-normative way to explore ambiguity and to find outliers and "misclassifications" rather than to make "accurate" predictions. However, much prior computational work on genre in the digital humanities has focused not on reception but on book texts [46], whereas we focus on reception via online book reviews.

Reception scholars such as Fish [13] have argued that readers' experiences of texts are strongly shaped by their "interpretive communities"—groups that share common strategies for interpreting texts (e.g., a group of professional literary critics). We find that LibraryThing users' reviews and tagging behaviors similarly correspond to their audiences on the site, with reviewers for certain genres writing more about certain aspects than others (§6). The shared norms in this tagging community might be driven not only by personal tagging motivations (tagging and curating one's own library) but by communal and performative ones, too (publishing reviews, ratings, and tags).

Our analysis of community homogeneity (§7) shows that some genres attract tighter communities of reviewers, whose reading habits are more similar to each other—*fantasy*, *vampires*, *romance*—while other genres attract reviewers with more diverse reading habits, such as *classics* and *children*. However, review writing style is not always consistent with community homogeneity, suggesting that there are other influences and other interpretive communities that shape LibraryThing users as well (e.g., literary criteria learned from school communities).

Virtual Collections Grounded in Real Places and Objects. LibraryThing is not just a virtual meeting place for book lovers; it also provides a cataloging service, TinyCat, to physical lending libraries around the world. TinyCat allows librarians to input their own metadata, but it also provides genre labels for books, which saves these librarians additional work. The process for genre assignment is not publicly explained but presumably relies to some extent on the tags provided by users on LibraryThing. Our exploration of how genre is defined on LibraryThing thus has implications for small libraries in addition to online communities. The non-conventional genres of LibraryThing may be shaping how today's library patrons discover books. It could be the case that "non-prestigious" genres are shaping our libraries and that patrons will be able to find books categorized by *vampires* enthusiasts on LibraryThing.

Open vs Closed Systems. Feinberg [11] notes that the value of a digital object, such as a book catalogued in online libraries, can take form at different levels, such as the document, the text, and the work. We could interpret the goal of LibraryThing's crowd-sourced metadata and volunteer efforts to clean and organize the digital library as seeking to normalize various objects and texts into a single *intellectual work*. In many ways, LibraryThing's cataloging goals—with incentives for users to combine tags and clean metadata—resemble the goals of a traditional library index. The TinyCat cataloging service and Common Knowledge wiki rely both on the richness and clarity of the data input by LibraryThing users. But the open, free-form tagging system on LibraryThing also gives the community creative license to diverge from traditional catalogs. Our findings highlight that changing shapes and multiple perspectives, rather than rigid taxonomies, are necessary when understanding genre on LibraryThing.

Comparison to Wikipedia Genres. LibraryThing and Wikipedia are united in their reliance on the crowd and their goals of creating internet catalogs, but their opposing philosophies on what those catalogs should look like, and how much power is given to the individual user, differentiates the websites and their resulting genre definitions. Our study of genre surprisal on LibraryThing provides strong evidence for an understanding of genres as shifting, overlapping entities in a living system, rather than fixed points in a descriptive, historical hierarchy. Frequent misclassifications based on the review texts show that boundaries are difficult to draw between many of the genres. Our study of book and user overlap bolsters this view, as even seemingly unlikely pairs of genres sometimes share significant numbers of books and/or users. Wikipedia genre definitions, in comparison, are more delineated and structured, which reflects their genesis as forced agreements between editors. For example, *fantasy* is defined on Wikipedia in terms of magical plot elements and settings, but we find that *fantasy* books on LibraryThing are associated with discussions of action plots, character development (Topic 16), and book series. These aren't simply qualities of the books tagged as fantasy but important themes that the reviewers value and spend time writing about.

The maps we built to compare lexical and community homogeneity on LibraryThing reveal strange groupings in comparison to the Wikipedia genre definitions. The pragmatic similarities of *psychology*, *politics*, *picture book*, and *graphic novel*, as shown in Figure 7a, are unexpected, given that their Wikipedia definitions center around topical themes rather than functional features of the audience and review text. On the other hand, the grouping of *fantasy*, *young adult*, *vampires*, and *romance* in Figure 7b is unsurprising given some of the thematic similarities between these genres, but it is surprising that they also function in very similar ways (dense community with low ratings; perhaps a community of picky readers). These patterns point to the importance of looking beyond themes and examining functional aspects of genres, which the tagging system allows us to do.

Genres as Boundary Objects. Our results support a view of genres through the analytic frame of *boundary objects*. The concept of boundary objects has been widely applied, including in diverse areas such as architecture and engineering design [36, 42] as well as ecological and environmental tools [8]. We rely on the original definition of boundary objects from Star [40] and the clarifications in Star [41], which describe objects that are coherent between different groups and allow for collaboration, while at the same time maintaining interpretive flexibility between those groups. These definitions were used in Worrall [56] to discuss the Goodreads and LibraryThing websites as boundary objects and sites of collaboration between different communities.

We find that LibraryThing genres are also a site of collaboration between communities with different perspectives; a *romance* fan is able to communicate with *history* enthusiasts even if their standards when evaluating *romance* and *history* texts are distinct (as we find when evaluating reviewer values in §6). Collaborative tagging allows these reading communities to work together, and the communication work of the reviews is often mediated through the reviewer's perception of the text's genre. These literary categories bring readers into conversation, though not always consensus, about the reviewing standards that should be applied to a genre and which tags to apply to works. This collaboration allows the community to build effective resources despite disagreements over specific tagging assignments. Efforts to clean the tags will always be in conflict with the long tail of tags applied to any popular work—and yet the cooperative project of LibraryThing still succeeds, supporting a wiki and cataloging service, allowing readers to "cooperate without consensus" despite the flexible boundaries of genres. This constructive friction allows LibraryThing to thrive through its diversity of reading experiences.

9 ETHICAL CONSIDERATIONS

Online book reviews pose challenges for ethical data science, especially with regard to citation and quotation. On the one hand, LibraryThing reviews are public and usually intended to be read by a wide audience of other book lovers. Many reviewers clearly take pride in their reviews and tags, as evidenced by their profiles full of badges, descriptions of their reading habits, and interactions with other reviewers. Reviewers often use their real names or include information in their public profile (e.g., location, age, profession, photos) that make them easily identifiable. Some reviewers are compensated for their reviews by authors or publishers, or they receive free books in exchange for reviews. All of this suggests that LibraryThing users and their labor deserve credit.

On the other hand, book reviews represent personal opinions on a wide variety of sensitive topics, and this information could be harmful if revealed in a new context or to an unexpected audience. This tension between honoring users' artistry and protecting their privacy is highlighted in Bruckman [5]. We can view the reviewers as "amateur artists" who deserve credit for their work, or we can view them as people who might not want or expect their work to appear beyond LibraryThing. Studies of Twitter users [12] and online fandom participants [9] have found that users have varying levels of comfort with researchers using their data. Most likely, different reviewers will have different perspectives on these questions, and so we err on the side of privacy.

Our study was considered exempt from our institution's IRB. Given the tensions discussed above, we do not release review texts or any data that is not easily viewable on the LibraryThing Zeitgeist web page.[10] Instead, we release the names of the 20 target genres as well as the 300 book IDs for each genre.[11] This maintains the review authors' abilities to edit and delete their reviews, while still giving credit to the creative work that enabled this study [5].

For the reviews that we directly quote in this article, we contacted the authors, disclosed our identities and publication intentions, and asked permission for use of their creative work and whether they would like their username credited. If the authors did not want to be included or did not respond, we replaced these quotations with reviews written by authors who have given consent. Our motivation in contacting the reviewers before publication is to honor users' wishes with regard to privacy and to grant them agency in how their creative work is presented in our paper; some reviewers prefer to be named as authors, while others prefer anonymity.

10 LIMITATIONS AND OPEN QUESTIONS

We intentionally focus on a single community, LibraryThing, because we are interested in how the users of one community have collaborated to form their own understanding of genre. This means that our analysis is constrained to this community, and we cannot guarantee that our results will transfer to other online book reviewing communities, or other online tagging communities. However, we expect that our methods of genre measurement would extend not only to other book reviewing websites like Goodreads but also to other art objects (e.g., movies, games) and other collaborative tagging systems in general.

We do not have comprehensive demographic information for the users on LibraryThing (it might be possible to scrape this information, but we did not feel that this was a worthwhile sacrifice of privacy), but it is very likely that these demographics are skewed and that this skew is reflected in the reading habits and tagging choices of the users. The communities that we observe around specific genres are thus only indicative of patterns on LibraryThing.

We might have made different decisions in our selection of 20 hand-picked genres. This choice leaves much unexplored. For example, it would have been useful to compare the thematic signatures of *historical fiction* and *history*, but we were unable to do so because we did not include *history* in our sample. We could instead have clustered all the tags using an unsupervised method and used these clusters as focal points, rather than single tags that resemble traditional genres.

Our sampling choices, while well-motivated, could also have introduced some unexpected effects. Although we had access to large sets of LibraryThing reviews, our sampling procedure (e.g., controlling for review length, controlling for polarity, controlling for genre) greatly limited the

[10] https://www.librarything.com/zeitgeist

[11] https://github.com/maria-antoniak/librarything-genres

final amount of data available for each genre. We were left with just two positive and two negative reviews per book and 300 books per genre—from a data science perspective, a very small sample. Perhaps most importantly, we have limited our analysis to only reviews written in English.

Finally, we see many open questions. A stronger foundation in sensemaking would enrich our discussions of how users individually and collaboratively make sense of difficult topics and works. We do not attempt to model direct relationships between users—on their profiles or on forum pages—though these interactions could help illuminate the community dynamics that we touch on in our measure of community homogeneity. These and other themes, we leave to future work.

11 CONCLUSION

We have taken a computational view of the LibraryThing book reviewing community. This approach has allowed us to map the open space of tags and compare the community's understandings of different genres. We find where genres overlap and where communities intersect, and by modeling the language used in reviews, we uncover the values and expectations that reviewers bring to particular genres. These comparisons and mappings emphasize that features outside of the book contribute to the community's understanding of genre. Similarity to other readers and perceptions of audience-appropriateness can also affect genre perceptions. Rather than take a prescriptive view that seeks to normalize tags, we describe one community's understanding of genre and genres.

12 ACKNOWLEDGEMENTS

The authors thank Matthew Wilkens, Rosamond Thalken, Gregory Yauney, Lillian Lee, Esin Durmus, Lauren Kilgour, Sharifa Sultana, Katherine Antoniak, and our anonymous reviewers for their feedback on this paper. We also thank the LibraryThing users who generously gave us permission to quote their reviews. This work was supported by NSF #1652536 and the Sloan Foundation.

REFERENCES

[1] Peishan Bartley. 2009. Book tagging on LibraryThing: how, why, and what are in the tags? *Proceedings of the American Society for Information Science and Technology* 46, 1 (2009), 1–22.

[2] Douglas Biber. 1986. Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language* (1986), 384–414.

[3] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *JMLR* 3, Jan (2003), 993–1022.

[4] Karen Bourrier and Mike Thelwall. 2020. The Social Lives of Books: Reading Victorian Literature on Goodreads. *Journal of Cultural Analytics* (Feb. 2020), 12049. https://doi.org/10.22148/001c.12049

[5] Amy Bruckman. 2002. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology* 4, 3 (2002), 217–231.

[6] Tiago Cunha, David Jurgens, Chenhao Tan, and Daniel Romero. 2019. Are All Successful Communities Alike? Characterizing and Predicting the Success of Online Communities. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 318–328. https://doi.org/10.1145/3308558.3313689

[7] Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13)*. Association for Computing Machinery, New York, NY, USA, 307–318. https://doi.org/10.1145/2488388.2488416

[8] Tshering Dema, Margot Brereton, Jessica L. Cappadonna, Paul Roe, Anthony Truskinger, and Jinglan Zhang. 2017. Collaborative Exploration and Sensemaking of Big Environmental Sound Data. *Comput. Supported Coop. Work* 26, 4–6 (Dec. 2017), 693–731. https://doi.org/10.1007/s10606-017-9286-9

[9] Brianna Dym and Casey Fiesler. 2020. Ethical and privacy considerations for research using online fandom data. *Transformative Works and Cultures* 33 (2020).

[10] James F. English, Scott Enderle, and Rahul Dhakecha. 2022. Bad Habits on Goodreads?. In *In preparation for James F. English and Heather Love, eds., Literary Studies and Human Flourishing* (New York). Oxford UP.

[11] Melanie Feinberg. 2013. Beyond Digital and Physical Objects: The Intellectual Work as a Concept of Interest for HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association

for Computing Machinery, New York, NY, USA, 3317–3326. https://doi.org/10.1145/2470654.2466453

[12] Casey Fiesler and Nicholas Proferes. 2018. "Participant" Perceptions of Twitter Research Ethics. *Social Media + Society* 4 (2018).

[13] Stanley Fish. 1982. *Is There a Text in This Class? The Authority of Interpretive Communities*. Harvard University Press, Cambridge Mass.

[14] John Frow. 2014. *Genre*. Routledge.

[15] Scott A Golder and Bernardo A Huberman. 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science* 32, 2 (2006), 198–208.

[16] Jane Gruning. 2018. Displaying Invisible Objects: Why People Rarely Re-Read E-Books. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173713

[17] William L Hamilton, Justine Zhang, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. 2017. Loyalty in online communities. In *Eleventh International AAAI Conference on Web and Social Media*.

[18] Paul Heymann, Andreas Paepcke, and Hector Garcia-Molina. 2010. Tagging Human Knowledge. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining* (New York, New York, USA) *(WSDM '10)*. Association for Computing Machinery, New York, NY, USA, 51–60. https://doi.org/10.1145/1718487.1718495

[19] Sudipta Kar, Suraj Maharjan, A. Pastor López-Monroy, and Thamar Solorio. 2018. MPST: A Corpus of Movie Plot Synopses with Tags. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), Miyazaki, Japan. https://www.aclweb.org/anthology/L18-1274

[20] Brett Kessler, Geoffrey Nunberg, and Hinrich Schutze. 1997. Automatic Detection of Text Genre. In *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Madrid, Spain, 32–38. https://doi.org/10.3115/976909.979622

[21] Evgeny Kim, Sebastian Padó, and Roman Klinger. 2017. Investigating the Relationship between Literary Genres and Emotional Plot Development. In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Association for Computational Linguistics, Vancouver, Canada, 17–26. https://doi.org/10.18653/v1/W17-2203

[22] Christian Körner, Dominik Benz, Andreas Hotho, Markus Strohmaier, and Gerd Stumme. 2010. Stop Thinking, Start Tagging: Tag Semantics Emerge from Collaborative Verbosity. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA) *(WWW '10)*. Association for Computing Machinery, New York, NY, USA, 521–530. https://doi.org/10.1145/1772690.1772744

[23] TK Landauer. 1984. Statistical semantics-analysis of the potential performance of keyword information-systems, and a cure for an ancient problem. In *Journal of psycholinguistic research*, Vol. 13. PLENUM PUBL CORP 233 SPRING ST, NEW YORK, NY 10013, 495–496.

[24] Suraj Maharjan, Manuel Montes, Fabio A. González, and Thamar Solorio. 2018. A Genre-Aware Attention Model to Improve the Likability Prediction of Books. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 3381–3391. https://doi.org/10.18653/v1/D18-1375

[25] Suman Kalyan Maity, Ayush Kumar, Ankan Mullick, Vishnu Choudhary, and Animesh Mukherjee. 2018. Understanding Book Popularity on Goodreads. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork* (Sanibel Island, Florida, USA) *(GROUP '18)*. Association for Computing Machinery, New York, NY, USA, 117–121. https://doi.org/10.1145/3148330.3154512

[26] Lynne M. Markus. 2001. Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems* 18, 1 (2001), 57–93.

[27] Julian John McAuley and Jure Leskovec. 2013. From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise through Online Reviews. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13)*. Association for Computing Machinery, New York, NY, USA, 897–908. https://doi.org/10.1145/2488388.2488466

[28] National Endowment for the Arts. 2004. Reading at Risk: A Survey of Literary Reading in America.

[29] Thomas Pavel. 2003. Literary Genres as Norms and Good Habits. *New Literary History* 34, 2 (2003), 201–210. https://www.jstor.org/stable/20057776

[30] Peter Pirolli. 2005. Rational analyses of information foraging on the web. *Cognitive Science* 29, 3 (2005), 343–373.

[31] J. D. Porter. 2018. Popularity/Prestige. *Stanford Literary Lab* Pamphlet 17 (2018).

[32] Emilee Rader and Rick Wash. 2008. Influences on Tag Choices in Del.Icio.Us. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, CA, USA) *(CSCW '08)*. Association for Computing Machinery, New York, NY, USA, 239–248. https://doi.org/10.1145/1460563.1460601

[33] Janice A. Radway. 1991. *Reading the Romance: Women, Patriarchy, and Popular Literature*. The University of North Carolina Press, Chapel Hill.

[34] Jeremy Rosen. 2018. Literary Fiction and the Genres of Genre Fiction. *Post45: Peer-Reviewed* (Aug. 2018). http://post45.research.yale.edu/2018/08/literary-fiction-and-the-genres-of-genre-fiction/

[35] Tiago Santos, Florian Lemmerich, Markus Strohmaier, and Denis Helic. 2019. What's in a Review: Discrepancies Between Expert and Amateur Reviews of Video Games on Metacritic. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 140 (Nov. 2019), 22 pages. https://doi.org/10.1145/3359242

[36] K. Schmidt and I. Wagner. 2004. Ordering Systems: Coordinative Practices and Artifacts in Architectural Design and Planning. *Computer Supported Cooperative Work (CSCW)* 13 (2004), 349–408.

[37] Shilad Sen, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper, and John Riedl. 2006. Tagging, Communities, Vocabulary, Evolution. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (Banff, Alberta, Canada) *(CSCW '06)*. Association for Computing Machinery, New York, NY, USA, 181–190. https://doi.org/10.1145/1180875.1180904

[38] Gene Smith. 2007. *Tagging: people-powered metadata for the social web.* New Riders.

[39] Efstathios Stamatatos, Nikos Fakotakis, and George Kokkinakis. 2000. Automatic Text Categorization In Terms Of Genre and Author. *Computational Linguistics* 26, 4 (2000), 471–495. https://www.aclweb.org/anthology/J00-4001

[40] Susan Leigh Star. 1989. The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving. In *Distributed Artificial Intelligence*.

[41] Susan Leigh Star. 2010. This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values* 35 (2010), 601–617.

[42] Eswaran Subrahmanian, Ira Monarch, Suresh Konda, Helen Granger, Russ Milliken, Arthur Westerberg, and The N-Dim Group. 2003. Boundary Objects and Prototypes at the Interfaces of Engineering Design. *Comput. Supported Coop. Work* 12, 2 (May 2003), 185–203. https://doi.org/10.1023/A:1023976111188

[43] Chenhao Tan. 2018. Tracing community genealogy: how new communities emerge from the old. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[44] Jaime Teevan, Christine Alvarado, Mark S. Ackerman, and David R. Karger. 2004. The Perfect Search Engine is Not Enough: A Study of Orienteering Behavior in Directed Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 415–422. https://doi.org/10.1145/985692.985745

[45] Laure Thompson and David Mimno. 2018. Authorless Topic Models: Biasing Models Away from Known Structure. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, Santa Fe, New Mexico, USA, 3903–3914. https://www.aclweb.org/anthology/C18-1329

[46] William E. Underwood. 2016. The Life Cycles of Genres. *Journal of Cultural Analytics* (May 2016), 11061. https://culturalanalytics.org/article/11061-the-life-cycles-of-genres

[47] Thomas Vander Wal. 2005. Folksonomy Definition and Wikipedia. http://www.vanderwal.net/random/entrysel.php?blog=1750

[48] Thomas Vander Wal. 2007. Folksonomy. https://www.vanderwal.net/essays/051130/folksonomy.pdf

[49] Melanie Walsh and Maria Antoniak. 2021. The Goodreads "Classics": A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism. *Post45 and Journal of Cultural Analytics* (2021).

[50] Mengting Wan and Julian J. McAuley. 2018. Item recommendation on monotonic behavior chains. In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*, Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan (Eds.). ACM, 86–94. https://doi.org/10.1145/3240323.3240369

[51] Mengting Wan, Rishabh Misra, Ndapa Nakashole, and Julian McAuley. 2019. Fine-Grained Spoiler Detection from Large-Scale Review Corpora. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 2605–2610. https://doi.org/10.18653/v1/P19-1248

[52] Jianling Wang, Ziwei Zhu, and James Caverlee. 2020. User Recommendation in Content Curation Platforms. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 627–635.

[53] Rick Wash and Emilee Rader. 2007. Public bookmarks and private benefits: An analysis of incentives in social computing. *Proceedings of the American Society for Information Science and Technology* 44, 1 (2007), 1–13.

[54] Jonathan Weber. 2006. Folksonomy and controlled vocabulary in LibraryThing. *Unpublished Final Project, University of Pittsburgh* (2006), 5–6.

[55] Matthew Wilkens. 2016. Genre, Computation, and the Varieties of Twentieth-Century U.S. Fiction. *Journal of Cultural Analytics* (1 11 2016).

[56] Adam Worrall. 2015. "Like a Real Friendship": Translation, Coherence, and Convergence of Information Values in LibraryThing and Goodreads. *iConference 2015 Proceedings* (2015).

[57] Joseph Worsham and Jugal Kalita. 2018. Genre Identification and the Compositional Effect of Genre in Literature. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, Santa Fe, New Mexico, USA, 1963–1973. https://www.aclweb.org/anthology/C18-1167

[58] Diyi Yang, Robert E. Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Jurafsky. 2019. Seekers, Providers, Welcomers, and Storytellers: Modeling Social Roles in Online Health Communities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 344, 14 pages. https://doi.org/10.1145/3290605.3300574

[59] Justine Zhang, William L Hamilton, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. 2017. Community identity and user engagement in a multi-community landscape. In *Eleventh International AAAI Conference on Web and Social Media*.

[60] Arkaitz Zubiaga, Christian Körner, and Markus Strohmaier. 2011. Tags vs Shelves: From Social Tagging to Social Classification. In *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia* (Eindhoven, The Netherlands) *(HT '11)*. Association for Computing Machinery, New York, NY, USA, 93–102. https://doi.org/10.1145/1995966.1995981