
Books that have stayed with us

By **Lada Adamic**, www.facebook.com

([http://getpocket.com/redirect?](http://getpocket.com/redirect?url=https%3A%2F%2Fwww.facebook.com%2Fnotes%2Ffacebook-data-science%2Fbooks-that-have-stayed-with-us%2F10152511240328859)

[url=https%3A%2F%2Fwww.facebook.com%2Fnotes%2Ffacebook-data-science%2Fbooks-that-have-stayed-with-us%2F10152511240328859](http://getpocket.com/redirect?url=https%3A%2F%2Fwww.facebook.com%2Fnotes%2Ffacebook-data-science%2Fbooks-that-have-stayed-with-us%2F10152511240328859))

View Original ([http://getpocket.com/redirect?](http://getpocket.com/redirect?url=https%3A%2F%2Fwww.facebook.com%2Fnotes%2Ffacebook-data-science%2Fbooks-that-have-stayed-with-us%2F10152511240328859)

[url=https%3A%2F%2Fwww.facebook.com%2Fnotes%2Ffacebook-data-science%2Fbooks-that-have-stayed-with-us%2F10152511240328859](http://getpocket.com/redirect?url=https%3A%2F%2Fwww.facebook.com%2Fnotes%2Ffacebook-data-science%2Fbooks-that-have-stayed-with-us%2F10152511240328859))

by Lada Adamic and Pinkesh Patel

Favorite books are something friends like to share and discuss. A Facebook meme facilitates this very interaction. You may have seen one of your friends post something like “List 10 books that have stayed with you in some way. Don't take more than a few minutes, and don't think too hard. They do not have to be the 'right' books or great works of literature, just ones that have affected you in some way.” If not great works of literature, what are the books that have stayed with us?

The following analysis was conducted on anonymized, aggregate data.

To answer this question we gathered a de-identified sample of over 130,000 status updates matching “10 books” or “ten books” appearing in the last two weeks of August 2014 (although the meme has been active over at least a year). The demographics of those posting were as follows: 63.7% were in the US, followed by 9.3% in India, and 6.3% in the UK. Women outnumbered men 3.1:1. The average age was 37. We therefore expect the books chosen to be reflective of this subset of the population.

We programmatically segmented the posts into lists, and found the most frequently occurring substrings, which corresponded to different books, e.g. “Anna Karenina by Leo Tolstoy”. However, the same book could appear as different substrings: e.g. just “Anna Karenina” or “Anna Karenina - Leo Tolstoy”. We clustered similar variants programmatically, hand tuning where the algorithm had failed to merge two popular variants. We then used the clusters to automatically match the book lists against the common variants of the top 500 most popular books.

Here are the top 20 books, along with a percentage of all lists (having at least one of the top 500 books) that contained them.



1. 21.08 Harry Potter series - J.K. Rowling
2. 14.48 To Kill a Mockingbird - Harper Lee
3. 13.86 The Lord of the Rings - JRR Tolkien
4. 7.48 The Hobbit - JRR Tolkien
5. 7.28 Pride and Prejudice - Jane Austen
6. 7.21 The Holy Bible
7. 5.97 The Hitchhiker's Guide to the Galaxy - Douglas Adams
8. 5.82 The Hunger Games Trilogy - Suzanne Collins
9. 5.70 The Catcher in the Rye - J.D. Salinger
10. 5.63 The Chronicles of Narnia - C.S. Lewis
11. 5.61 The Great Gatsby - F. Scott Fitzgerald
12. 5.37 1984 - George Orwell
13. 5.26 Little Women - Louisa May Alcott

14. 5.23 Jane Eyre - Charlotte Bronte
15. 5.11 The Stand - Stephen King
16. 4.95 Gone with the Wind - Margaret Mitchell
17. 4.38 A Wrinkle in Time - Madeleine L'Engle
18. 4.27 The Handmaid's Tale - Margaret Atwood
19. 4.05 The Lion, the Witch, and the Wardrobe - C.S. Lewis
20. 4.01 The Alchemist - Paulo Coelho

While there are many great 'serious' books on the list, the Hitchhiker's Guide to the Galaxy makes an appearance at #7, and Harry Potter reigns supreme (although enjoying the advantage that it was most often referred to as a series and our clustering algorithm lumped all Harry Potter books into the same cluster). Stephen King's dark novels have stayed with their readers as well (The Stand at #14 and the Dark Tower series at #64). In the complete list of the top 100, included at the end of this post, we see a number of children's books appear as well. Although these may not normally be considered great works of literature, they tend to stay with us through the decades. In particular, two of Shel Silverstein's books (the Giving Tree and Where the Sidewalk Ends) make it into the top 100, as does the Little Prince.

One can also look at connections between the books, e.g. 'people who listed X also listed Y', using pointwise mutual information. In the network visualization, each node represents a book, sized by the frequency with which it was mentioned, as an edge represents an unusual number of co-occurrences of the two books in the lists.

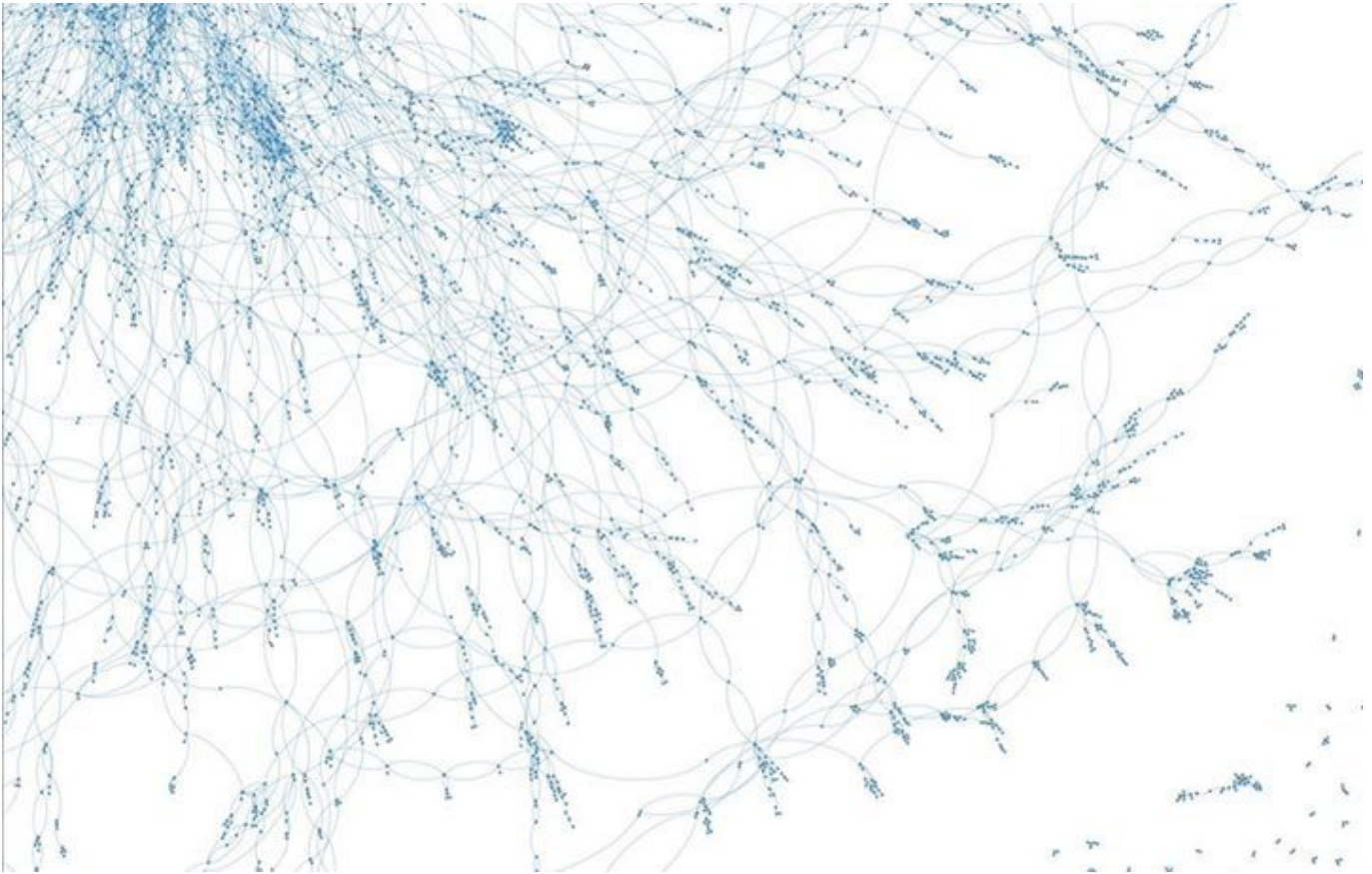


Each book is linked to another it occurs with more often than expected. The color represents whether the book was more often mentioned by women

Photo by: red

There is actually another kind of network that forms. While some people shared the meme without tagging, calling on all their friends to make their own posts, others tagged specific friends whose favorite books

they'd like to know about. Even a small fragment of the cascade shows long (tangled) tagging chains through which it diffused.



Tagging links posts about favorite books.

Do friends tend to like the same books? We computed the number of books shared between lists linked via tags, which was a mere 0.4 books on average! This number was 4 times greater than the overlap of 0.1 books between any two random lists. It is also an underestimate, since our automated matching identifies only 5.3 books/list on average (rather than the full 10), due to matching on just the 500 most commonly mentioned titles. Nevertheless, the low overlap underlines that even in a world of relatively few highly successful bestsellers, lists of favorites tend to be rather different, even between friends.

Finally, the remaining top 100 books were:

- 55 2.31 The Count of Monte Cristo - Alexandre Dumas
- 56 2.31 Pillars of the Earth - Ken Follett
- 57 2.29 The Da Vinci Code - Dan Brown
- 58 2.24 Brave New World - Aldous Huxley
- 59 2.21 A Tale of Two Cities - Charles Dickens
- 60 2.21 Les Miserables - Victor Hugo
- 61 2.16 Great Expectations - Charles Dickens
- 62 2.12 Night - Elie Wiesel
- 63 2.12 The Dark Tower Series - Stephen King
- 64 2.07 Outlander - Diana Gabaldon
- 65 1.92 The Color Purple - Alice Walker
- 66 1.89 A Thousand Splendid Suns - Khaled Hosseini
- 67 1.88 The Art of War - Sun Tzu
- 68 1.85 Catch 22 - Joseph Heller
- 69 1.85 The Bell Jar - Sylvia Plath
- 70 1.83 The Perks of Being a Wallflower - Stephen Chbosky
- 71 1.78 The Old Man and the Sea - Ernest Hemingway
- 72 1.76 Memoirs of a Geisha - Arthur Golden
- 73 1.75 Tuesdays with Morrie - Mitch Albom
- 74 1.73 The Road - Cormac McCarthy
- 75 1.72 Watership Down - Richard Adams
- 76 1.72 A Tree Grows in Brooklyn - Betty Smith
- 77 1.68 Where the Sidewalk Ends - Shel Silverstein
- 78 1.65 The Girl with the Dragon Tattoo - Stieg Larsson
- 79 1.65 A Song of Ice and Fire - George R. R. Martin
- 80 1.65 Are You There God? It's Me, Margaret - Judy Blume
- 81 1.64 Charlotte's Web - E.B. White
- 82 1.63 The Time Traveler's Wife - Audrey Niffenegger
- 83 1.62 Anna Karenina - Leo Tolstoy
- 84 1.62 Crime and Punishment - Fyodor Dostoyevsky

- 85 1.61 The Adventures of Huckleberry Finn - Mark Twain
- 86 1.58 The Shack - William P. Young
- 87 1.56 Watchmen - Alan Moore
- 88 1.55 Interview with the Vampire - Anne Rice
- 89 1.54 The Odyssey - Homer
- 90 1.54 The House of the Spirits - Isabel Allende
- 91 1.53 The Stranger - Albert Camus
- 92 1.52 Call of the Wild - Jack London
- 93 1.51 The Five People You Meet in Heaven - Mitch Albom
- 94 1.51 Siddhartha - Herman Hesse
- 95 1.50 East of Eden - John Steinbeck
- 96 1.50 Matilda - Roald Dahl
- 97 1.49 The Picture of Dorian Gray - Oscar Wilde
- 98 1.47 Zen and the Art of Motorcycle Maintenance - Robert Pirsig
- 99 1.45 Love in the Time of Cholera - Gabriel Garcia Marquez
- 100 1.45 Where the Wild Things Are - Maurice Sendak

[An earlier version of this post had 2 clusters representing the Chronicles of Narnia series. When these were merged, the series rose up to #10]