

Co-Loan Analysis of Finnish Public Library Loan Data

Olli Nurmi¹, Kati Launis² and Erkki Sevänen³

¹ VTT Technical Research Centre of Finland LTD, Espoo Finland

² University of Eastern Finland, Dept. of Literature, Joensuu, Finland

³ University of Eastern Finland, Joensuu, Finland

olli.nurmi@vtt.fi, klaunis@utu.fi, erkki.sevanen@uef.fi

Abstract. This paper analyses public library loan data of the two most popular book genres, novels and crime fiction, to illustrate the cultural and social literacy prevailing in Finland. Using a social network analysis method we are able to identify and visualize distinct book clusters – rather than simply groups of books – providing a considerably more nuanced insight. Firstly, we generated book networks based on library customers’ loan transactions where books were associated with each other when they were co-loaned. We then applied a modularity maximization method to identify book clusters. The most influential books and authors were identified by examining their involvement in the network. The results show that the reading culture is no longer uniform but is fragmented into multiple smaller clusters. Additionally, the position of national classics, popular among Finnish readership some decades ago, has radically weakened. The results also show that the library users typically borrow multiple books in the same series. Through this study, we found that a social network analysis leads to a better interpretation of the library collection usage and overview of the reading culture. The presented approach benefits library users, librarians, and literary scholars.

Keywords: Social Network Analysis, Library Loan Data, Reading Culture.

1 Introduction and Applied Method

Social network analysis studies have rarely been conducted using public library user interaction data. A major constraint on data access relates to public libraries’ long-standing commitment to data privacy. As a general practice, public libraries neither collect nor store their member-generated transactional logs in any form.

In our study, we had access to a sample (2016-17, 1.5 million loans) of anonymized library loan data, which allowed us to generate a network of books based on their loan transactions. Each book was linked to others when they were co-loaned, creating a co-occurrence book network.

We made an assumption that with a large volume of library loan data we could effectively detect books that were related to each other, forming book clusters. Ideally, these clusters would consist of a manageable number of books that could then be interpreted and characterized. Here we refer to the distributional hypothesis in which the underlying idea is that “a book is characterized by the company it keeps”.

In literary research, social network analyses have occasionally been used as a means to visualize certain structural features of a text or a corpus (Moretti, 2013 and Michel et al. 2010). A common usage is the visualization of relationships between the texts based on the similarities of the textual contents, and relationships between textual entities such as words. A further application is the visualization of the relations between characters appearing in the texts (Jänicke, Franzini, Cheema, & Scheuermann, 2015).

Over time, book clusters might grow or shrink in size, may split into smaller clusters or merge forming larger ones. New clusters may also emerge, and old ones can disappear. In this study, we used modularity maximization (Newman 2006) which is a state-of-the-art method for community detection in static networks (Javed, Younis, Latif, Qadir, & Baig, 2018). As a tool, we used the Gephi open-source network analysis and visualization software package.

2 Data Sample and Data Preparation

Vantaa City Library is a part of the Helmet network (Helsinki Metropolitan Area Libraries), consisting of the city libraries of Helsinki, Espoo, Kauniainen, and Vantaa. The Helmet collection, consisting of 3.4 million items, is available for Vantaa City Library users through this network.

Our data sample include all records collected from the Vantaa City Library during the period 20 July 2016–22 October 2017 involving about 1.5 million loan interactions. In this data women borrowed the majority of fictional titles: they borrowed 76 percent of the fiction books, which, of course, has a great effect on the results (Launis, Cherny, Neovius, Nurmi & Vainio, 2018).

Typically, the number of loans peak during holiday periods, when so called light reading titles are more frequently borrowed. The data collection period of 15 months was long enough to smoothen these sorts of temporal fluctuations from the overall loan pattern.

In this work, we selected adult loan data for novels and crime fiction for closer analysis. To study the interconnections between the literary works (i.e. book titles) we merged the loan data regarding different editions, manifestations, and copies for each book title. We then identified the co-loans based on the paired presence of the books within a specific loan cart. The lists of co-loaned books were not sampled from unique library visits but was each an aggregation of data across multiple library visits. An intuitive selection criterion to establish a robust link between the books was an aggregated number of co-loans of at least five. The size of the resulting two co-occurrence networks is presented in the Table 1.

Table 1. The size of the co-occurrence networks. The nodes represent books and the edges indicate robust associations between the books.

Nodes (books)	Edges (associations)	Genre
533	395	novels
618	572	crime fiction

Surprisingly, the number of robust associations between the books was rather low. The reason for this was that the number of book loans per title had a heavy-tailed distribution where only a fraction of the books were popular with many co-occurrences far from the “head” or central part of the distribution.

The frequently loaned books were followed by a majority of less popular books which gradually “tailed off” asymptotically. The books at the far end of the tail have a very low probability of co-occurrence and robust associations.

3 Results

3.1 Book Clusters

The method identified six major clusters of novels, which mostly contain entertaining fiction written by women for women. Many of the books in these clusters are published in series.

Figure 1 depicts that four of the six book clusters were formed around contemporary female writers, writing entertaining fiction in series and under a pseudonym. Three of these ‘entertaining’ clusters – popular fiction or ‘light reading’ written in series and targeted for female readership – were formed around Finnish, contemporary female authors, with the pseudonyms Enni Mustonen, Marja Orkoma and Anneli Kivelä. In their novels, the story is typically located in idyllic, rural areas of Finland. The novels discuss not only love but also everyday life from the women’s point of view.

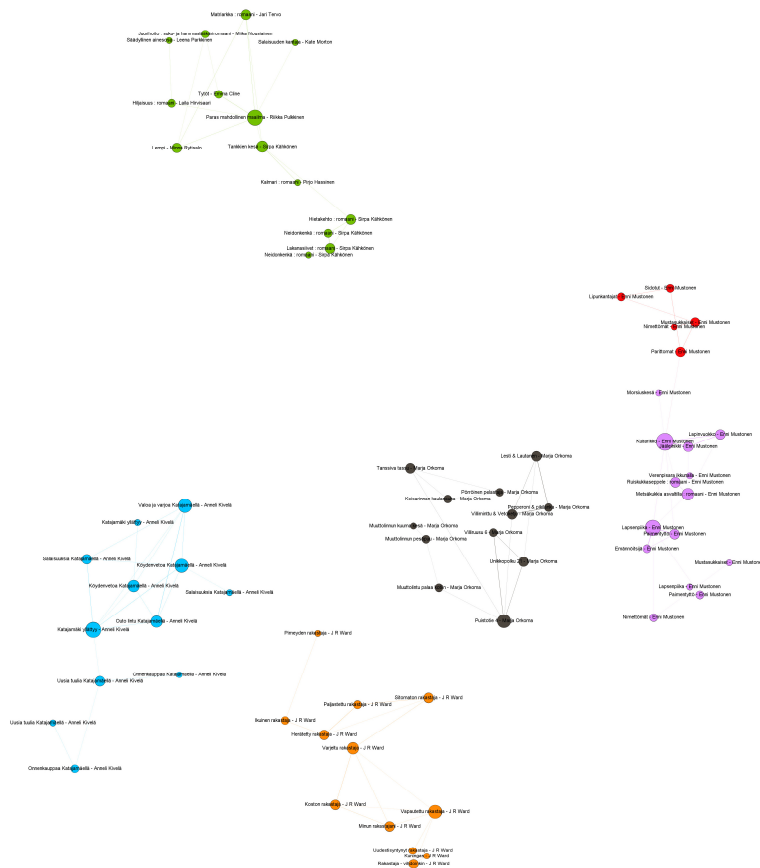


Fig. 1. Co-occurrence network of novels showing six of the largest novel clusters.

At the top and middle right are clusters containing entertaining novels with lively historical depictions, from the popular Finnish female author Kirsti Manninen (alias Enni Mustonen). In Mustonen’s series “Järjen ja tunteen tarinoita” (Stories of Reason and Emotion), borrowed in this cluster, women’s life stories and the history of Finland are intertwined. The series takes place in the 18th–20th century and contains the novels *Nimettömät* (The Nameless, 2004), *Mustasukkaiset* (The Jealous Ones, 2005), *Lipunkantajat* (The Flag Bearers, 2006), *Sidotut* (The Bound Ones, 2007) and *Parittomat* (The Unpaired Ones, 2008). Mustonen is accurate and skilful in presenting how historical vortexes have had influence and have been felt in the lives of ordinary contemporaries. Her most recent series “Syrjästäkatsojan tarinoita” (Stories of an Onlooker, 2013) has been a success both in libraries and in bookstores; one of the novels of this series (*Ruokarouva*, Housekeeper, 2016) was the most popular fictional work among women in the Vantaa City Library data used for this article. One reason for this could

be that the genre this novel represents of historical novels and historical romance have been popular in Finland since the mid nineteenth century (See Launis et al., 2018).

In the middle is the cluster with Pirkko Syyrimaa's (alias Marja Orkoma) novels. These novels take place in the fictional Blackbird Valley. Syyrimaa is best known for her crime novels, published under the pseudonym Pirkko Arhippa. At the bottom right is the blue cluster including Anne Seppälä's (alias Anneli Kivelä's) novels that are published in the series titled "Katajamäki" (2007–). The stories, again, take place in an idyllic countryside village. The series is conveniently written so that each part can also be read as a separate story. The main characters vary from one story to another. They are women at different stages of their lives, leaving behind the city dust of Tampere and Helsinki and their problems by choosing rural peace. They want to evaluate their past, weigh their future options, and seek a new direction.

The green cluster at the top left contains novels from contemporary Finnish authors Sirpa Kähkönen, Minna Rytisalo, Miika Nousiainen, Riikka Pulkkinen, Jari Tervo, Leena Parkkinen and Pirjo Hassinen. Their books are considered high-quality novels, valued by critics and often referred to in the media. The cluster also includes novels from the international bestselling authors Kate Morton and Emma Cline.

3.2 Central Authors

We can identify the central authors by looking the total number of associations. There are at least two relevant ways of doing this which give different results.

Firstly, we can consider the total number of the associations per author regardless of the quality of the associations. In many cases, books are written in series and part of these associations are between books by the same author. Second way is to omit these self-loops and consider the total number of associations between different authors books. The result of this analysis is shown in Table 2 for crime fiction.

Table 2. Authors whose books had the most associations in total and authors with most associations between each other.

Authors with highest number of associations between books		Authors with highest number of associations between each other	
Mari Jungstedt	SE	Anna Jansson	SE
Seppo Jokinen	FIN	Mari Jungstedt	SE
Jarkko Sipilä	FIN	Jo Nesbø	NO
Outi Pakkanen	FIN	Jarkko Sipilä	FIN
Christian Rönneck	FIN	Lars Kepler	SE
Ann Cleeves	UK	Stefan Tegenfalk	SE
Kati Hiekkapelto	FIN	Leena Lehtolainen	FIN
Katarina Wennstam	SE	Katarina Wennstam	SE
Anna Jansson	SE	Kristina Ohlsson	SE
Jussi Adler-Olsen	DK	Samuel Bjørk	NO

Interestingly there is a difference in the share of Finnish or Swedish authors in the two lists being 5/3 in the first and 2/6 in the second list. This indicates that people borrowing crime fiction written by a Finnish author keep up reading the same book series and do not often borrow crime fiction by some other author. This pattern was not visible among the borrowers of Swedish authors' books.

Well-networked authors included Mari Jungstedt, Jarkko Sipilä, Katarina Wennstam and Anna Jansson. Mari Jungstedt and Anna Jansson have sold millions of copies which have been translated into various languages.

4 Reading Culture in Finland

The results indicate that the Finnish reading culture has changed radically since the classical studies by Katarina Eskola focusing on the readers and their literary manners and taste in the 1970s and 1980s. The position of national classics (such as Väinö Linna and Mika Waltari), which were popular among the Finnish readership some decades ago, has radically declined. The reading culture is not uniform anymore (cf. Eskola, 1979, 1990) and it is fragmented into multiple smaller clusters. In another connection (Launis & Mäkikalli 2020) we have shown that the young readers in our data favour the translations of the newest Anglo-American young-adult fiction, novels such as John Green's *The Fault in Our Stars* and Estelle Maskame's *DIMILY*-trilogy. Interestingly, the authors favoured by young readers are highly visible on the Internet and on social media; the impact of digitalization is clearly visible in the data.

On the other hand, some aspects of the literary taste of Finnish readers seem to be quite permanent. Even though brand new titles, domestic fiction and winners of the annual literary prize (Finland-prize) are very much favoured by the readers, also depictions of Finnish history narrated in a realistic manner and depicting hard work and the countryside still seem to tempt readers, as can be seen in the clusters addressed in this article. A good example of this is the book *Ruokarouva* (Housekeeper) by Kirsi Manninen (pen name Enni Mustonen) (Launis, Cherny, Neovius, Nurmi & Vainio, 2018).

Our results are in line with Kimmo Jokinen's (1997) study. He stated that stories picturing everyday life in a realistic and detailed fashion are typical for the Finnish reading culture – both the previous article focusing on women readers in Vantaa City Library (Launis, Cherny, Neovius, Nurmi & Vainio, 2018) and the co-occurrence analysis used here reveal that this kind of entertaining fiction is still very much favoured by Finnish readers, forming the book clusters shown above. In the Finnish reading culture, there is not much room for literature that deviates from this form of literature – nor for literature that plays with experimental forms.

The results also show that library users typically borrow multiple books in the same series. This can be explained by the increased use of branding, where a set of marketing and communication methods are applied to distinguish the author from competitors, aiming to create a lasting impression in the minds of the readers. An author brand is, in essence, a promise to its readers, including emotional benefits. When readers are familiar with an author's brand, they tend to favour it over competing others.

5 Discussion

A co-loan data analysis enables grouping books according to their readerships. This enriches and opens up new directions in literary and library/information studies. Visualizing book relationships enables researchers to interpret the usage of a library collection in new ways rather than simply looking at the relative popularity of the books.

Overviews can vary from visualizations that display the individual books in a collection and their relationships (i.e. the document space) to displays that show themes or topics associated with the contents of the books (i.e. the semantic space). We can look at relationships between local and temporal clusters, and clusters across time, as well as clusters grouped by class or gender, as well as physical or temporal proximity.

One of the obvious advantages in carrying out a network analysis is that very complex connections can be made clear and structured. This way conclusions can be drawn about a large quantity of connections that would otherwise possibly appear to be irrelevant or overly complicated.

Several algorithms can be used to calculate the importance of any given node in a network. In the libraries' case, we can use these algorithms to identify authors or books with influence over the whole network. By promoting these influential authors or books, librarians could increase their effect on promoting reading.

One of the disadvantages of using a network analysis is that people might have the tendency to solely focus on "the big picture", thus neglecting all the small and sometimes personal/human factors that also play a role in a certain analysis. Looking at library loans in such a calculating way might invoke a certain utilitarian viewpoint that does not take into account other library functions.

Library users may benefit from visualizations of book clusters, which may offer them new ways in finding books to read. Placing the books in clusters facilitates the library users' ability to shift smoothly from one cluster to another when searching and selecting new books. The visualization is also useful in educating library users to understand how modern collaborative recommendation engines work.

The method offers a new tool for librarians in their collection management and helps to identify book clusters and their relations. Books may be divided into natural groups and category management can be aligned accordingly. The books may even be placed in new ways that could help users to find interesting books. This type of analysis can also facilitate new ways to create book recommendations. In addition, the results show that series of books should be marked to enable the readers to locate them easily.

Literary scholars can use the described methods to investigate social structures if they have access to the large data sources available in libraries. For example, they could identify local and global patterns and gain a better picture of the prevailing literary culture. Changes and trends can be detected by repeating this kind of analysis periodically.

In practical applications, we need datasets that contain thousands or millions of transactions to create meaningful book clusters. The results must be considered with care because the method cannot say anything about how the reader has understood, used, or reflected on the written texts.

The quality of the results obtained depends on a variety of factors, such as the quality of the descriptors used for the relationship, the scope of the database, and the adequacy of statistical methods for simplifying and representing the findings.

Topics for further research include the tie formation or tie strength connecting the books together. In this study, we used co-loans as an indication of a link between the books, but there might be other forms of linking such as co-searching in a web catalogue. Even the absence of links between book clusters may reveal some interesting literary phenomena. A lack of interconnectedness between groups may result in “filter bubbles” in terms of information exchange.

Additional visualisations are accessible at Libdat-project webpage as interactive graphs (<http://virtual.vtt.fi/virtual/libdat/index.htm>).

References

1. Eskola, K.: *Suomalaiset kirjanlukijoina*. Tammi, Helsinki (1979).
2. Eskola, K.: *Lukijoiden kirjallisuus Sinuhesta Sonja O:hon*. Tammi, Helsinki (1990).
3. Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G.: On Close and Distant Reading in Digital Humanities : A Survey and Future Challenges. *Eurographics Conference on Visualization, EuroVis*, 1–21 (2015).
4. Javed, M. A., Younis, M. S., Latif, S., Qadir, J., & Baig, A.: Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications*, 108 (January), 87–111 (2018).
5. Jokinen, K.: *Suomalaisen lukemisen maisemaihanteet*. Jyväskylä, Jyväskylän yliopisto (1997).
6. Launis, K., Cherny E., Neovius M., Nurmi O. & Vainio, M.: Mitä naiset lukevat? *AVAIN - The Finnish Review of Literature Studies*, (4), 4–21 (2018).
7. Launis, Kati & Mälikalli, Aino: Mitä tehdä, kun Shakespeare ei vloggaa eikä Waltari twiittaa? Kirjasto, koulu ja nuorten uudistuvat lukemiskulttuurit. Kirjallinen elämä markkinaperaustaisessa mediayhteiskunnassa. Toim. Elina Arminen, Anna Logrén ja Erkki Sevänen. *Vastapaino, Tampere*, 333–360 (2020).
8. Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Aiden, E. L.: Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331 (6014), 176–182 (2010).
9. Moretti, F.: *Distant Reading*. Verso, London & New York (2013).
10. Neovius, M., Launis, K., & Nurmi, O.: Exploring Library Loan Data for Modelling the Reading Culture: Project LibDat. In E. Mäkelä, M. Tolonen, & J. Tuominen (Eds.), *Humanities in the Nordic Countries 3rd. Conference*, pp. 386–393 (2018).
11. Newman, ME: Modularity and community structure in networks. *Proc Nat Acad Sci* 103(23):8577–8582 (2006).
12. Libdat-project homepage, <http://virtual.vtt.fi/virtual/libdat/index.htm>, last accessed 2021/3/2.