



Article

Analysis of Concentrations of Loans by Using Book Circulation Data in Korea University Library

Ji-Ann Yang

The Institute for Korean Christianity Culture, Soongsil University, 369 Sangdo-Ro, Dongjak-Gu, Seoul 06978, Korea; jiannyang@gmail.com; Tel.: +82-2-820-0838

Received: 14 October 2020; Accepted: 8 December 2020; Published: 10 December 2020



Abstract: In this paper, data of almost 8 million loans of books recorded for 15 years by the Korea University Library are analyzed by using big data analytic techniques. During this period, book circulation decreased with an average annual rate of decline of 4.4%. The use factor of books in each Dewey decimal classification (DDC) class was evaluated to measure how efficiently books were used by library users. Loan frequencies of books were analyzed and meaningful results regarding loan concentrations and the half-lives of books were obtained. It was observed that 50% of the total loans in each year were for 20% of all borrowed books in that year. This phenomenon will be called the 20/50 loan rule, and the set of the top 20% most borrowed books, whose cumulative loan frequencies reach 50% of total loans, will be called a core collection. The 20/50 loan rule shows the loan concentration of library books. The extent of loan concentration gets stronger if loans for two or more consecutive years are concerned. It was found that with high probability, books in a core collection at a specific year are also categorized as a core collection in next years. Moreover, books categorized as a core collection in consecutive years have longer half-lives compared with all other circulating books.

Keywords: library; loan; 20/50 loan rule; loan concentration; core collection; half-life

1. Introduction

The library collection is the most important part the various elements that make up a library. Most libraries have difficulties in providing users with optimal collections due to their limited budgets and space problems caused by increases in the volume of books. So, collection development has become one of the most difficult tasks for librarians. To overcome these issues, library are trying to organize user-centered collections by analyzing book use patterns and grasping users' preferences for collections. Libraries have also invested a lot of effort and time in the development of collections so that users can effectively use the library materials needed for research and learning activities, since the collections of academic libraries are an important basis for research and education.

In addition, recent changes in the library information environment have diversified resource types and information delivery methods, and have changed the information-seeking behaviors of users affected by information technology [1–3]. Due to these effects, the use of e-books or electronic resources has increased every year, while the use of printed books has decreased every year. A number of studies have been conducted on comparing the utilization rates of printed books and e-books as well as on the decreasing purchasing of physical materials such as books [4]. It was found that developing a collection with printed books and e-books has different effects on users' book usage [5]. Rose-Wiles showed that the circulation of printed books in university libraries has been reduced by 23% in five years [6]. Thus, a more detailed analysis of circulation data is needed to develop print collections effectively as the use of electronic materials increases. For this, circulation data are often used to analyze the characteristics of a library's collection or to identify the current status of books' use. A library has a vast collection of materials, and at the same time, users produce vast amounts of circulation data every day using

library materials. These circulation data in libraries are both big data and structured data that can be analyzed immediately. Because of these points, it has become increasingly important to identify the books borrowed intensively by library users and utilize them efficiently for the purpose of collection development and management in libraries.

Collection evaluation through the analysis of circulation statistics is also very important for accurately grasping the current status of collections in the library and predicting users' collection usage patterns to keep the collections up to date and improve the quality. It is the basis for solving space problems, which are seriously considered by academic libraries, and increasing the usability of collections by weeding out low-use collections [6,7]. Similarly, circulation statistics are an invaluable source of information for figuring out users' preferences for a specific subject in a collection in that books are still an important communication channel between library users in all subject areas. Both librarians and users consider circulation statistics to be the primary criteria for assessing library values. That is, the analysis of circulation data is an important factor that gives actual users an actual insight into which books they use, which can be used as basic data for organizing collections and improving library services. Therefore, it is necessary to develop a method of analyzing data accumulated in a library by utilizing circulation data in various ways.

This study aims to find circulation patterns of library users based on the circulation data of print books with a big data analysis method. This paper is organized as follows: Section 2 reviews relevant studies on circulation patterns and analyses of library books and users. Section 3 introduces the data collection and methods. Section 4 describes the results of our research with analytical techniques. Then, Section 5 discusses our findings and directions for the future research.

2. Related Work

Many studies have been conducted to analyze the circulation patterns of users based on a library's book collection statistics and circulation statistics. Walters et al. studied the differences among librarians, faculty, and students in the selection of library books [8]. Dinkins evaluated the collection development policy by using circulation statistics gathered in five departments for five years in order to check whether the books purchased through the library acquisition met the needs of users. The utilization rates of books selected by librarians were generally higher than those of books selected by professors in the departments. Therefore, it could be more desirable to select library materials through collaboration of librarians and professors in each department [9].

Using the one-year circulation data of university library users, Kim analyzed the use pattern of library collections of 16 subject areas and five user groups [10]. The preference for new materials was investigated for each subject and user group. Yoo et al. classified one year of circulation records of university libraries into 16 subjects and five user groups, and analyzed them in terms of use factors and publication years by using statistical analysis methods, such as correlation coefficients [11].

Galvin et al. analyzed the usage of books and academic journals to contribute to establishing the library development policy of the Pittsburgh University Library [12]. In research-oriented university libraries, interrelationships were analyzed in the aspect of budgets, usage of books and academic journals, and the aging rate of literature. Of the total collections, 40% of books were never checked out and 73% of the loaned books were borrowed within two years of purchase. This is clear evidence of the preference for the recently published books. Moreover, such a phenomenon had shown almost the same pattern for six years. Based on this analysis, books checked out for the first time within 6–7 years of purchase and the aging patterns of book usage could be predicted.

Adams et al. used one-year circulation statistics obtained from the Indiana University Science Library to analyze circulation patterns for collections depending on publishers, publication years, and subjects [13]. Books published earlier tended to be checked out more than recently published books. It was also found that books of a specific publisher were rarely checked out, and thus, the publisher is a factor influencing the circulation pattern. Using five-year book loan statistics from the Lied Library at Nevada University, Tucker conducted a comparative analysis of collection usage patterns for books

purchased by liaison librarians and books purchased through the library acquisition [2]. The lending of the entire collection had been decreasing every year for five years, and the usage rate of books purchased through the liaison librarian was also low. So, there was a need to redistribute the book budget because of the decrease in book usage.

Cheung et al. analyzed the loan patterns of books obtained from the university library's 15-year checkout statistics in various categories defined by year, language, and Library of Congress classification to determine the utilization rate of books [14]. It was found that the utilization rate for the first 5–7 years after the book was purchased tended to increase, and then decreased slightly. In addition, books that had not been loaned within the first 1–2 years after they were purchased had never been loaned for 15 years, and the usage rate of books purchased by the library was higher than that of the donated books. In addition, one-third of all books purchased at the library were never checked out. So, books that were not circulated within the first 5–7 years could be stored in a separate storage library to solve the problem of insufficient space.

Some studies focused on developing a core collection using citation analysis for the purpose of collection development. Knievel et al. used 9131 citations included in 2002 journals to analyze citation patterns [15]. The type of literature mainly cited in the science and technology field was predominantly in academic journals, but most of the citations in the humanities field were in books. It was found that such citation analysis could be immediately applied to the library collection development policies. Black derived core journals based on the form of citations in specific subject areas and the distribution of cited academic journals by publication year [16]. Scholarly journals and books were typical forms of citations, with 13% of scholarly journals accounting for 80% of citations. In addition, the period for which articles in academic journals were mainly cited was observed to be at least eight years. Keeping track of important journals and books through citation analysis helped researchers save time when they could find the literature.

Trueswell conducted a number of studies to determine a core collection of books [17,18]. He analyzed circulation transaction data to form the optimal core collection out of total library holdings. He introduced a ratio of the last circulation date and the loan period, and obtained a cumulative distribution of the resultant ratio. It became possible to predict the core collection satisfying 99% of users' requirements and to reduce the volume of library holdings by 60% to 70% at the same time. He also refined the pattern of determining the core collection by using circulation transaction data obtained from university libraries [19]. He asserted the necessity of selecting a core collection as an objective criterion for distinguishing between books that were highly circulated and books to be stored in stacks in the library. In addition, by examining the patterns of library users' behaviors, Trueswell demonstrated the 80/20 rule, meaning that 80% of usage can be attributed to 20% of the collection [20].

Oh et al. analyzed the usability of collections for each subject in terms of Bonn's use factor, turnover rate of the collection, and Trueswell's 80/20 rule by using circulation data for 10 years of the Gwangjin District Public Library [21]. They found core collections from the borrowed books according to use patterns and presented the basis for effective collection development policy. This analysis of circulation data could be used as a collection evaluation index to contribute to collection development and collection evaluation.

Through the previous studies introduced above, it was found that the circulation data can be used for establishing collection evaluation and collection management policies by analyzing the usage patterns of books based on various criteria, such as subject, language, publication year, etc. This study, however, is not limited to describing the technical analysis of the loan data, but rather investigates core collections in which loans intensively occur and users' loan patterns from specific perspectives. Furthermore, the variations in loan concentration depending on all subject areas are explored and the half-lives of books are also analyzed.

3. Data Collection and Methods

In this paper, circulation data of books recorded for 15 years by the Korea University Library, which is one of top-ranked research-oriented universities in Korea, are analyzed. The Korea University Library held 2,616,202 books as of 2019, which is at a very high level among universities in Korea. Collection statistics and circulation data were gathered for each fiscal year from 2005 to 2019. The circulation data were extracted through Tulip, a library integration system, and written as a text format. Each entry of data contains various fields, including title, Dewey decimal classification (DDC), author, publisher, publication year, loan and return date for each transaction, number of renewals, number of multiple copies, etc. The original circulation data of all fields acquired were pre-processed and refined to exclude entries containing erroneous records or missing entities. In-house use of books could not be quantified because there is no system for monitoring in-house uses of books in the library. Thus, this study analyzed only books that were checked out and did not include in-house use.

Library users were grouped according to their affiliations with humanities/social sciences, science, engineering, and others. Circulation data were analyzed by subject categories based on the DDC, which is a system for organizing all subjects of books from general to specific, indicated by numbers. The DDC has ten main classes, and the first digit in each three-digit number represents the main class: 000 stands for computer science, information, and general works, 100 for philosophy and psychology, 200 for religion, 300 for social sciences, 400 for language, 500 for science, 600 for technology, 700 for art and recreation, 800 for literature, and 900 for history and geography. To analyze the 15-year circulation data collected for this study, the open-source language R, which is widely used for big data analysis and data mining, was utilized.

4. Results

4.1. General Statistics of Circulation Data

The number of books registered in the Korea University Library increased gradually every year. It increased from 1,527,355 in 2005 to 2,616,202 in 2019, and the average number of books per year was approximately 2.2 million during this period. On the other hand, the number of loans extracted for each year decreased gradually. It decreased from 631,346 in 2005 to 330,584 in 2019, and the total number of loans for 15 years reached 7,983,869. As for the collection status by subject, the largest collection was in the fields of 300 social sciences and 800 literature, while the lowest was in the fields of 400 language and 200 religion. Considering the large variations in the number of books in each subject area, it is not appropriate to simply analyze the number of loans to measure the efficiency of book use. To resolve this problem, this paper uses the use factor proposed by Bonn to measure how books in a specific category are utilized by library users [22]. The use factor of books in a specific category m is defined as

$$\text{Use factor of books in subject category } m = \frac{\frac{\text{number of loans in a specific subject category } m}{\text{total number of loans in all subject categories}}}{\frac{\text{number of books in a specific subject category } m}{\text{total number of books in all subject categories}}} \quad (1)$$

The high ratio of the number of loans over the number of books in a specific subject category indicates that books in that category are actively used in terms of loans. The higher the use factor of a specific category is, the more actively books in this category are used by library users. If the use factor is greater than 1, it is evaluated as an actively used collection. The number of books in the library and the number of loans recorded from 2005 to 2019 are summarized in Table 1.

It was observed that books in 300 social sciences and 800 literature were the most frequently borrowed, while books in 200 religion were the least borrowed among all subjects. Considering the number of books in each subject, however, books in 500 science seemed to be borrowed relatively less than others. The books in 500 science showed the lowest use factor among all DDC classes. On the other hand, 800 literature, 700 art and recreation, and 100 philosophy and psychology showed

relatively high use factors. The use factor can be effectively used by libraries to establish their policies for purchase and acquisition of new books. For example, a library needs to purchase books in DDC classes with high use factors each year because they are relatively frequently borrowed compared with the number of books.

Table 1. The numbers of books and loans and the use factor in each Dewey decimal classification (DDC) class for each year.

Year	Item	000	100	200	300	400	500	600	700	800	900	Total
2005	Books	94,956	74,121	32,422	318,087	50,603	294,583	192,946	55,399	282,237	132,001	1,527,355
	Loans	31,005	34,191	9301	120,538	23,868	51,942	77,909	34,414	196,144	52,034	631,346
	Use Factor	0.79	1.12	0.69	0.92	1.14	0.43	0.98	1.50	1.68	0.95	-
2006	Books	102,986	85,170	36,792	354,707	55,544	312,101	230,254	64,520	316,264	147,541	1,705,879
	Loans	29,032	34,923	8775	127,046	24,036	40,067	80,763	34,688	181,411	51,352	612,093
	Use Factor	0.79	1.14	0.66	1.00	1.21	0.36	0.98	1.50	1.60	0.97	-
2007	Books	109,414	93,807	41,535	386,080	60,647	327,311	253,655	71,960	344,013	161,797	1,850,219
	Loans	29,937	39,791	9455	135,392	26,997	45,920	86,043	35,397	181,298	56,838	647,068
	Use Factor	0.78	1.21	0.65	1.00	1.27	0.40	0.97	1.41	1.51	1.00	-
2008	Books	114,956	100,930	45,352	419,947	64,749	339,334	271,140	80,695	364,031	175,043	1,976,177
	Loans	29,477	41,191	9371	139,702	26,998	45,222	87,334	36,123	184,593	56,784	656,795
	Use Factor	0.77	1.23	0.62	1.00	1.25	0.40	0.97	1.35	1.53	0.98	-
2009	Books	119,774	106,568	48,410	451,066	68,626	352,390	283,016	86,548	381,858	185,548	2,083,804
	Loans	27,412	42,438	10,394	145,068	26,625	43,890	86,554	36,464	192,983	60,571	672,399
	Use Factor	0.71	1.23	0.67	1.00	1.20	0.39	0.95	1.31	1.57	1.01	-
2010	Books	123,158	111,546	50,165	472,932	70,639	363,777	294,060	90,999	393,764	191,961	2,163,001
	Loans	25,174	44,797	10,144	142,230	24,629	49,278	80,959	34,767	173,423	58,949	644,350
	Use Factor	0.69	1.35	0.68	1.01	1.17	0.45	0.92	1.28	1.48	1.03	-
2011	Books	124,519	113,675	51,402	482,151	71,978	369,926	298,768	93,542	400,954	195,437	2,202,352
	Loans	24,236	43,895	9942	136,996	23,308	43,383	76,448	32,023	150,725	54,411	595,367
	Use Factor	0.72	1.43	0.72	1.05	1.20	0.43	0.95	1.27	1.39	1.03	-
2012	Books	127,045	117,559	53,446	499,043	73,755	380,479	306,535	97,257	410,721	202,325	2,268,165
	Loans	22,831	42,448	8721	129,997	21,028	38,950	70,135	29,639	130,309	50,292	544,350
	Use Factor	0.75	1.50	0.68	1.09	1.19	0.43	0.95	1.27	1.32	1.04	-
2013	Books	129,151	124,326	55,783	513,917	77,253	391,202	313,454	101,246	424,448	209,748	2,340,528
	Loans	21,268	40,452	8294	123,197	20,202	35,005	64,253	28,934	117,629	47,596	506,830
	Use Factor	0.76	1.50	0.69	1.11	1.21	0.41	0.95	1.32	1.28	1.05	-
2014	Books	130,740	126,839	57,518	524,326	79,790	405,345	318,475	103,665	432,654	213,601	2,392,953
	Loans	20,199	37,804	7671	113,292	19,692	32,133	57,297	25,238	106,675	44,529	464,530
	Use Factor	0.80	1.54	0.69	1.11	1.27	0.41	0.93	1.25	1.27	1.07	-
2015	Books	132,402	129,167	58,717	534,472	81,866	411,969	322,879	105,941	440,736	217,097	2,435,246
	Loans	19,685	37,079	7403	114,916	19,676	33,234	53,349	28,576	107,162	45,537	466,617
	Use Factor	0.78	1.50	0.66	1.12	1.25	0.42	0.86	1.41	1.27	1.09	-
2016	Books	133,691	131,181	59,776	543,441	82,922	420,985	326,559	107,633	445,845	220,174	2,472,207
	Loans	19,488	34,952	6869	108,457	18,139	31,185	48,433	27,978	102,048	42,077	439,626
	Use Factor	0.82	1.50	0.65	1.12	1.23	0.42	0.83	1.46	1.29	1.07	-
2017	Books	135,342	133,995	60,895	553,089	84,604	433,165	331,082	109,543	453,830	223,763	2,519,308
	Loans	19,380	31,768	6327	99,893	17,220	27,574	42,727	22,663	93,687	39,140	400,379
	Use Factor	0.90	1.49	0.65	1.14	1.28	0.40	0.81	1.30	1.30	1.10	-
2018	Books	137,488	136,472	62,148	564,923	86,091	440,652	338,243	111,839	459,755	226,868	2,564,479
	Loans	18,242	29,172	6589	92,896	15,916	24,978	40,135	20,717	87,071	35,819	371,535
	Use Factor	0.92	1.48	0.73	1.14	1.28	0.39	0.82	1.28	1.31	1.09	-
2019	Books	140,865	139,303	63,409	576,819	87,018	451,054	346,361	114,084	466,097	231,192	2,616,202
	Loans	16,228	26,923	5924	85,309	13,412	22,438	35,946	18,135	75,677	30,592	330,584
	Use Factor	0.91	1.53	0.74	1.17	1.22	0.39	0.82	1.26	1.28	1.05	-

4.2. Concentration of Loans and the 20/50 Loan Rule

The number of borrowed books and the loan frequency of each borrowed book were investigated for 15 years from 2005 to 2019. Books with multiple copies were considered a single book, and the loan frequency was obtained by adding loan frequencies of all multiple copies. After each book's loan frequency was computed in this manner, books were ranked in descending order of loan frequency. Then, the loan frequencies of books were arranged in descending order and plotted as in Figure 1,

where the x-axis denotes the percentage ranking of books in terms of loan frequency. Note that Figure 1 was obtained by using book loan data of the year 2010. It was observed that the arranged loan frequency in descending order formed a power law distribution. This implies that only a small proportion of books were intensively borrowed, and they occupied high proportion of total loans. In short, loans were highly concentrated on a small proportion of books.

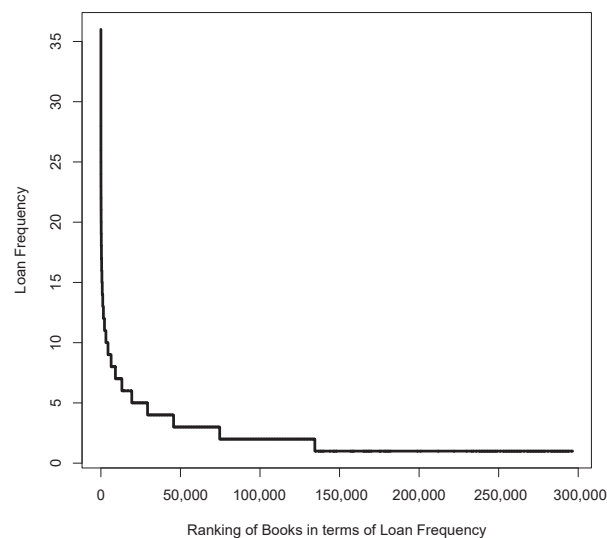


Figure 1. One-year loan frequency of books arranged in descending order, where the book loan data of year 2010 were used.

To see the extent of loan concentration quantitatively, the cumulative distribution of loan frequency with respect to the percentage ranking of books was investigated. In Figure 2, cumulative distributions of one-year loan frequencies arranged in descending order are plotted, where for visual clarity, cumulative distributions for only four sample years, 2007, 2010, 2013, and 2018, are plotted. It was observed that every year, the cumulative distribution of loan frequency had an identical shape, and 50% of total loans were concentrated on the top 20% most borrowed books. This observation will be called the 20/50 loan rule [23]. The top-ranked most borrowed books occupying half of the total loans for each year will be called the core collection [23]. The proportions of the core collection out of all borrowed books for each year from 2005 to 2019 are listed in Table 2. The proportion of the core collection among all borrowed books varies year by year, but it is around 20%. Thus, the 20/50 loan rule is observed to be satisfied every year.

Table 2. Proportion of the core collection considering all DDC classes as a whole for each year.

Year	Proportion (%)	Year	Proportion (%)	Year	Proportion (%)
2005	17.95	2010	20.62	2015	21.22
2006	18.39	2011	21.21	2016	21.09
2007	18.73	2012	21.72	2017	20.95
2008	19.49	2013	21.77	2018	21.14
2009	20.11	2014	21.89	2019	21.99

For more detailed analysis of the loan concentration, loans in each DDC class were investigated. Loan frequencies of individual borrowed books belonging to each DDC class were counted. Then, the cumulative distribution of loan frequency arranged in descending order was obtained, and the percentage ranking of books corresponding to the cumulative distribution of 50% was determined. The set of books whose rankings were higher than or equal to the determined ranking were defined as a core collection in the DDC class under investigation.

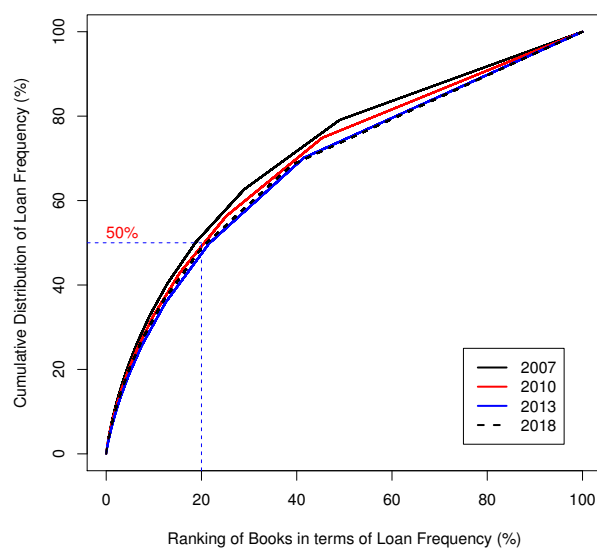


Figure 2. Cumulative distribution of one-year loan frequency arranged in descending order, where book loan data of years 2007, 2010, 2013, and 2018 are considered. The 20/50 loan rule is observed.

Table 3 lists percentage rankings of books corresponding to the 50% point of cumulative distribution in each DDC class. It was observed that 50% of loans of one year in each DDC class were concentrated on approximately 20% of borrowed books, and thus, the 20/50 loan rule was generally satisfied in each DDC class. It is notable that the equivalent level of loan concentration was observed for all DDC classes. The circulation pattern of books in each DDC class was expected to be identical, although each DDC class shows different numbers of books, different numbers of loans, and different use factors. It was also observed that proportions of core collections in 500 science and 800 literature were lower than those in other DDC classes. Thus, 500 science and 800 literature experienced higher loan concentration than others. Note that low proportion of core collections indicates high loan concentration. On the other hand, 200 religion and 700 art and recreation showed lower loan concentrations because the core collections were defined by higher proportions of borrowed books than others.

Table 3. Proportion (%) of core collections in each DDC class for each year.

Year	000	100	200	300	400	500	600	700	800	900
2005	20.24	20.71	21.78	20.25	20.35	16.91	19.87	21.71	15.01	19.22
2006	21.20	20.15	22.82	20.27	20.90	18.80	20.26	22.12	15.39	19.19
2007	21.25	20.93	24.02	20.82	21.94	18.27	20.55	22.31	16.10	19.73
2008	21.44	21.08	24.96	21.33	22.61	19.24	20.93	23.09	16.46	21.12
2009	22.08	21.46	25.21	21.50	23.21	20.26	21.59	23.80	16.95	21.47
2010	22.22	21.59	25.78	21.45	23.63	20.33	21.65	24.77	17.73	21.46
2011	22.19	21.51	25.77	21.64	23.83	20.55	22.15	25.80	18.64	21.93
2012	22.26	21.94	26.08	22.20	23.96	20.98	22.92	26.06	20.02	22.65
2013	22.11	21.76	26.69	22.53	24.05	20.89	23.08	26.22	20.39	22.63
2014	21.91	21.55	25.71	22.70	23.90	20.92	23.10	26.18	20.36	22.37
2015	21.36	21.68	27.19	21.47	22.74	20.06	22.46	24.14	20.37	21.38
2016	20.81	21.36	26.31	21.44	22.67	19.87	22.57	23.54	20.51	21.29
2017	20.00	21.61	26.78	20.94	22.89	19.76	22.69	25.39	20.35	21.38
2018	19.65	21.50	26.57	21.04	23.25	19.48	22.41	26.03	20.36	21.94
2019	20.56	21.79	29.07	21.76	24.06	19.66	22.45	26.76	20.94	23.61

4.3. Variation of Loan Concentrations by Adding or Merging Categories

When books are grouped by a certain classification, it is possible that a single book can belong to multiple groups. If this happens in the core collection, the extent of loan concentration is expected to get weaker. As introduced in Section 4.2, classifying books as DDC classes does not significantly change

the extent of loan concentration of all books. Other kinds of classification, however, can influence the extent of loan concentration.

To investigate the variation of loan concentration as a result of adding categories, grouping books according to users' affiliations is considered. Table 4 lists the proportions of core collections according to the affiliations of users and for each year, where it was observed that the proportion of core collection exceeded 20%. This shows that weaker concentrations of loans resulted from dividing books into groups. Nevertheless, the proportions of core collections did not deviate seriously from 20%, which shows that the 20/50 loan rule was still satisfied in general.

Table 4. Proportions (%) of core collections according to the affiliations of users.

Year	Humanities/Social Sciences	Science	Engineering
2005	19.78	22.15	19.25
2006	19.75	24.78	20.65
2007	20.50	23.68	21.07
2008	20.94	24.62	23.00
2009	21.27	25.69	24.04
2010	21.88	26.27	24.69
2011	22.78	27.95	25.24
2012	23.81	29.33	25.49
2013	24.15	30.56	25.50
2014	24.23	31.33	25.68
2015	23.45	30.32	25.71
2016	23.54	29.49	25.71
2017	23.68	29.99	25.70
2018	23.93	29.63	25.69
2019	24.52	30.95	25.87

Next, we considered merging categories, with the resultant loan concentration getting stronger. Merging categories was achieved by merging loan data of consecutive years when counting loan frequencies of borrowed books. The number of consecutive years merged together will be called a window size. The relationship of loan concentration and the window size was investigated by varying window size from one year to 15 years. By applying a sliding window technique, the average value of loan frequency for a given window size was computed. Figure 3 shows the cumulative distributions of average loan frequencies arranged in descending order with respect to the percentage ranking of books. As the window size got longer, 50% of loans were accomplished by lower proportions of borrowed books. In other words, the core collection had higher dominance as the window size got longer. This tendency is also shown in Table 5, which lists the average proportion of books occupying half of the total loans for a given number of consecutive years, i.e., window size.

Table 5. Average proportion (%) of books occupying 50% of total loans for each window size.

Window Size (Years)	1	2	3	4	5	10	15
Proportion (%)	20.55	16.72	14.97	13.96	13.39	12.02	11.45

The above observation is explained by the fact that books in a core collection at a certain year can belong to a core collection in the next year with a high probability. Table 6 lists the average proportion of books in a core collection at a certain year that are also categorized as a core collection in the next years. For obtaining average proportions, a sliding window technique was applied. At a certain year, a core collection was determined and stored. Then, proportions of books in the core collection that satisfied the following three properties were obtained:

- They were also categorized as a core collection in the next year,
- They were also categorized as a core collection in at least one year in the next two years,
- They were also categorized as a core collection in next two consecutive years.

It was observed that, with high probability, books in a core collection in a certain year were also in a core collection in the next year. Thus, doubling the window size does not guarantee doubling the volume of core collection. The growing speed of the volume of core collection was slower than the growing speed of the window size of consecutive years. It follows that the loan concentration gets stronger as the window size of consecutive years grows. The 500 science and 800 literature classes showed a higher extent of loan concentration with respect to the window size than other DDC classes.

Table 6. Probability (%) that a core collection in a certain year becomes a core collection in the next year(s).

	000	100	200	300	400	500	600	700	800	900	Total
Next year	45.40	44.95	31.69	44.42	43.11	52.86	42.15	34.79	51.21	46.86	46.34
At least 1 year in next 2 years	54.37	55.66	39.95	53.65	54.24	62.35	51.08	45.37	60.54	56.48	55.79
Next 2 consecutive years	25.13	26.31	16.05	25.62	24.58	35.50	23.41	17.28	34.33	28.04	28.20

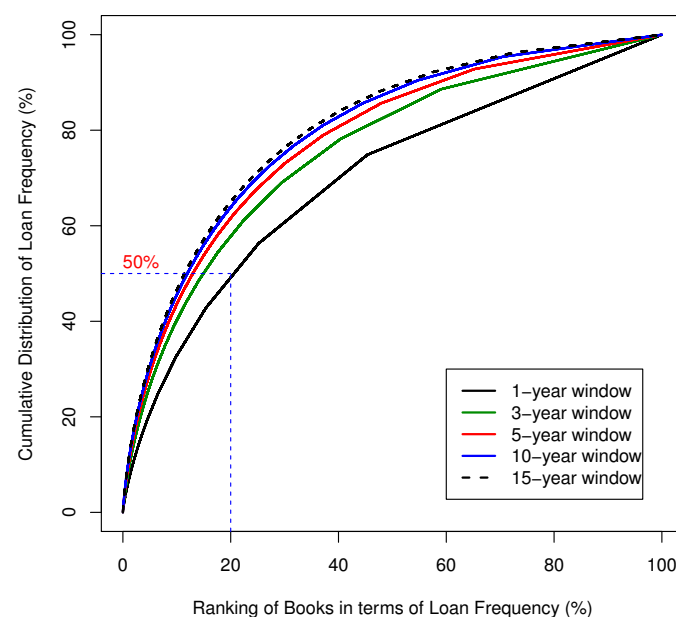


Figure 3. Cumulative distribution of loan frequency arranged in descending order with varying window sizes.

4.4. Half-Lives of Books

In general, books are intensively borrowed right after they are published and shelved in the library [10–12,14]. Thus, the loan frequency of books is highly correlated with their book ages, where a book age is defined by the number of years elapsed after the book is published. The typical loan frequency of books exponentially decays with respect to the book age, as shown in Figure 4. A slowly decaying pattern of loan frequency implies that library users utilize the book for a long time.

In physics, the half-life or half-life duration was introduced as the period of time required for one half of the atomic nuclei of a radioactive material to decay [24]. It can also be regarded as the time interval required for a certain entity to have the cumulative distribution of 50%. In addition, many previous studies have defined the half-life as a measure to indicate the life span of academic journals [25,26]. People want to evaluate books in terms of the time duration in which they are intensively utilized by library users [23]. As a result, the half-life was adopted in the field of library and information sciences as a useful parameter to measure characteristics of library books.

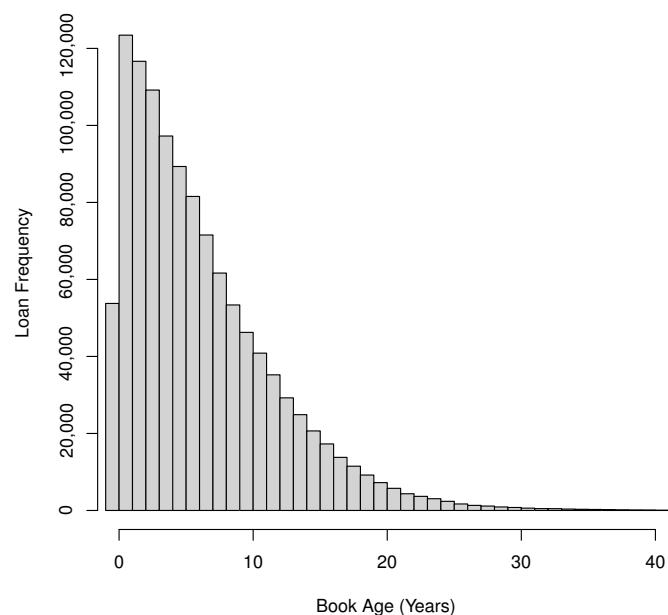


Figure 4. Loan frequency with respect to book age.

Figure 5 shows a typical cumulative distribution of the loan frequency of books, which was obtained by using loan data in 800 literature during the years from 2005 to 2019. The half-life is defined as the book age corresponding to the 50% point of cumulative distribution. After plotting the cumulative distribution of loan frequency as in Figure 5, a curve-fitting technique was applied to obtain an equation describing the cumulative distribution with respect to the book age [23]. Then, by solving the equation, the half-life was determined as the book age corresponding to 50% of the cumulative distribution.

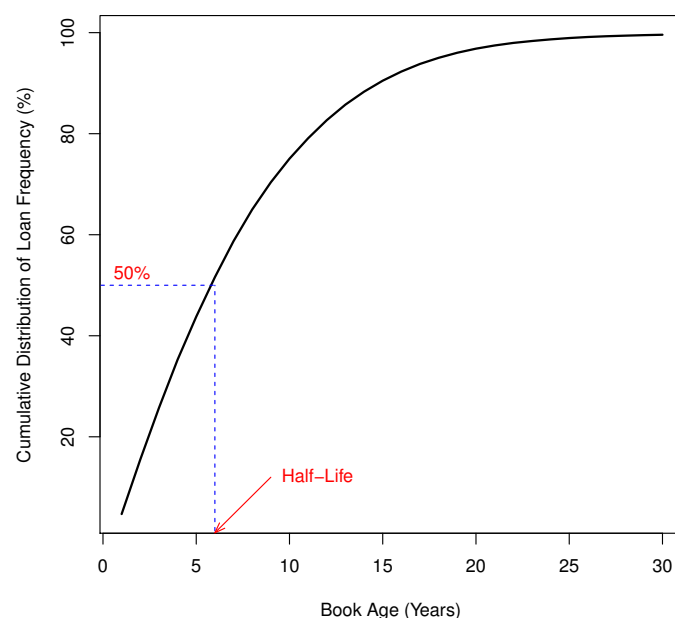


Figure 5. Cumulative distribution of loan frequency with respect to book age, where the half-life is marked.

Table 7 lists the values of half-lives of books belonging to each DDC class for the years from 2005 to 2019. It was observed that books in 200 religion and 400 language had longer half-lives than books in other DDC classes. It is notable that books belonging to 300 social sciences have short half-lives, though they have high loan frequency, as listed in Table 1.

Table 7. Half-lives of books in each DDC class.

DDC	000	100	200	300	400	500	600	700	800	900	Books	Loans
Half-life	4.04	5.91	6.61	5.40	6.77	6.38	4.69	6.23	6.27	6.17	1,088,742	7,983,869

The values for the half-life of a core collection in consecutive years are listed in Table 8. As shown in Table 8, core collections in consecutive years had longer half-lives than other books. This observation follows the observation provided in Table 6 that books in a core collection at a certain year were classified as a core collection in the next years with high probability. Thus, the analysis of half-lives can also explain why loan concentration gets stronger as the time window size grows.

Table 8. Half-lives of books that were considered core collections for consecutive years.

Number of Years	000	100	200	300	400	500	600	700	800	900	Books	Loans
5 years	4.79	6.27	6.37	5.83	6.75	6.14	5.16	6.48	6.36	6.38	45,987	2,183,993
10 years	6.25	7.51	7.92	7.65	8.37	7.61	7.15	7.75	7.93	8.93	5855	488,644

5. Discussion

In this paper, the loan concentrations of print books in a library were intensively analyzed. It was found that a majority of loans were attributed to a small portion of the collection, called a core collection. Library users steadily utilized this core collection, causing the core collection to have a longer half-life than other books.

Circulation data regarding loan activities recorded for 15 years in the Korea University Library were thoroughly investigated and analyzed by using big data analytic techniques. The Korea University Library is one of top-ranked research libraries in Korea, and it held 2.6 million books as of 2019 and 8 million loan histories for 15 years from 2005 to 2019. Book circulation gradually decreased over 15 years from 631,346 in 2005 to 330,584 in 2019. The number of loans in 2019 decreased by 47.6% compared to the number of loans in 2005. In this study, almost 8 million loans and corresponding books' bibliographic information were intensively analyzed to obtain interesting and meaningful results.

The use factor of books in each DDC class for each year was evaluated to quantify how efficiently books were used by library users in terms of loans. A high use factor of books in a specific DDC class means a high number of loans per book in that class. Thus, a library needs to purchase books in this class to improve the efficiency of book usage.

The loan frequencies of books were the main ingredients of this study, and they were analyzed from various perspectives to obtain useful results. The loan concentrations and half-lives of books are examples of practical observations gained from the analysis of loan frequency. Every year, 50% of total loans were accomplished by 20% of all borrowed books. This observation is named the 20/50 loan rule. It was observed that the 20/50 loan rule was valid for the 15 years from 2005 to 2019. This means that usage patterns of library users were highly consistent and very predictable for the core collection. This rule also held for each DDC class every year. A core collection is also defined as books whose cumulative loan frequencies correspond to 50% of total loans in each year. The existence of a core collection may provide libraries with a guideline for plans of service, acquisition, and collection management. As such, libraries may operate a core collection section to offer library users quick and easy access to popular books.

It was observed that the extent of loan concentration gets stronger if multiple categories of books are merged as one. Lengthening the window size for looking into loan frequencies is an example of merging categories. If loan frequencies of books are counted for two or more consecutive years, the proportion of borrowed books accomplishing 50% of loans is lower than 20%. Thus, the loan concentration gets stronger in this case. On the other hand, if books are further classified by extra categories, the loan concentration gets weakened. If books are further categorized according to

affiliations of users, the proportion of books regarded as a core collection in each category is higher than 20%. This indicates weaker concentration of loans.

The reason for the stronger loan concentration resulting from considering longer window size in loan frequency analysis is provided. With high probability, books in a core collection at a specific year are categorized as a core collection in the next years. Books included in core collections of consecutive years have longer half-lives than ordinary books. These can explain why loan concentration can be stronger by increasing the window size used for loan frequency analysis. In this paper, loan frequency is thoroughly analyzed and many meaningful results are obtained. Loan duration is regarded as another meaningful factor to judge how actively books are utilized in terms of loans. Thus, an analysis of the loan durations of borrowed books will be performed as future research. Since library users have distinct maximum loan periods according to their status, e.g., undergraduate student, graduate student, faculty, and staff, more sophisticated analysis techniques may be required in future studies.

Funding: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A6A3A01042723).

Conflicts of Interest: The author declares no conflict of interest.

References

1. Rose-Wiles, L.M.; Shea, G.; Kehnemuyi, K. Read in or check out: A four-year analysis of the distribution and internal use of printed books. *J. Acad. Librariansh.* **2020**, *46*. [\[CrossRef\]](#)
2. Tucker, J.C. Collection evaluation of monograph purchased by the University of Nevada, Las Vegas libraries. *Collect. Manag.* **2009**, *34*, 157–181. [\[CrossRef\]](#)
3. Ching, S.H.; Hsiao, M. Impact of the e-book on total circulation transactions—a case study for Hong Kong academic libraries. *Libr. Collect. Acquis. Tech. Serv.* **2015**, *39*, 1–8. [\[CrossRef\]](#)
4. Waters, J.; Roach, J.; Emde, J.; Mceathron, S.; Russell, K. A comparison of e-book and print book discovery, preferences, and usage by science and engineering faculty and graduate students at the university of Kansas. *Issues Sci. Technol. Librariansh.* **2014**, *75*. [\[CrossRef\]](#)
5. Fry, A. Factors affecting the use of print and e-books: Use research and discussion. *Univ. Res. Libr.* **2018**, *79*, 68–85.
6. Rose-Wiles, L.M. Are printed books dead? an Investigation of book circulation at a mid-sized academic library. *Tech. Serv. Q.* **2013**, *30*, 129–152. [\[CrossRef\]](#)
7. White, B. Citations and circulation counts: Data sources for monograph deselection in research library collections. *Coll. Res. Libr.* **2017**, *78*, 53–65. [\[CrossRef\]](#)
8. Walters, W.H.; Gormley, J.; Handfield, A.E.; López-Fitzsimmons, B.M.; Markgren, S.; Paradise, L.; Sheehan, S.E. Library book selection decisions and selectors? effectiveness: Differences among librarians, faculty, and students. *Univ. Res. Libr.* **2020**, *81*, 617–645. [\[CrossRef\]](#)
9. Dinkins, D. Circulation as assessment: Collection development policies evaluated in terms of circulation at a small academic library. *Coll. Res. Libr.* **2003**, *64*, 46–53. [\[CrossRef\]](#)
10. Kim, S.-A. A study on the collection use of an academic library: Focused on the analysis of circulation statistics of the a-academic library users. *J. Korean Biblia Soc. Libr. Inf. Sci.* **2013**, *24*, 115–131.
11. You, K.-J.; Park, I.-J. A study about the use survey by the circulation record analysis of an academic library. *J. Inf. Manag.* **2008**, *39*, 109–129.
12. Galvin, T.J.; Kent, A. Use of a university library collection. *Libr. J.* **1977**, *102*, 2317–2320.
13. Adams, B.; Noel, B. Circulation statistics in the evaluation of collection development. *Collect. Build.* **2008**, *27*, 71–73. [\[CrossRef\]](#)
14. Cheung, S.; Chung, T.; Nesta, F. Monograph circulation over a 15-year period in a liberal arts university. *Libr. Manag.* **2011**, *32*, 419–434. [\[CrossRef\]](#)
15. Knievel, J.; Kellsey, C. Citation analysis for collection development: A comparative study of eight humanities fields. *Univ. Libr. Fac. Staff Contrib.* **2005**, *49*, 142–168. [\[CrossRef\]](#)
16. Black, S. Using citation analysis to pursue a core collection of journals for communication disorders. *Libr. Resour. Tech. Serv.* **2001**, *45*, 3–9. [\[CrossRef\]](#)

17. Trueswell, R.L. A quantitative measure of user circulation requirements and its possible effect on stack thinning and multiple copy determination. *Am. Doc.* **1965**, *16*, 20–25. [[CrossRef](#)]
18. Trueswell, R.W. Determining the optimal number of volumes for a library's core collection. *Libri* **1966**, *16*, 49–61. [[CrossRef](#)]
19. Trueswell, R.W. User circulation satisfaction vs. size of holdings at three academic libraries. *Coll. Res. Libr.* **1969**, *30*, 204–213. [[CrossRef](#)]
20. Trueswell, R.W. Some behavioral patterns of library users: The 80/20 rule. *Wilson Libr. Bull.* **1969**, *43*, 458–461.
21. Oh, J.-E.; Jeong, D.Y. A study on the analysis of rate of use and core collection for collection evaluation in public libraries: In the case of Gwangjin district public library. *J. Korean Soc. Libr. Inf. Sci.* **2015**, *49*, 201–221.
22. Bonn, G.S. Evaluation of the collection. *Libr. Trends* **1974**, *22*, 265–304.
23. Yang, J.-A. A study on use behaviors of books through circulation data analysis in an academic library. *J. Korean Library Inf. Sci. Soc.* **2017**, *48*, 263–293.
24. Burton, R.E.; Kebler, R.W. The half-life of some scientific and technical literature. *Am. Doc.* **1960**, *11*, 18–22. [[CrossRef](#)]
25. Line, M.B. The influence of the type of sources used on the results of citation analyses. *J. Doc.* **1979**, *35*, 265–284. [[CrossRef](#)]
26. So, M.-H.; Ko, S.-S. Research on the holding value of academic journals based on the half-life index-number in science and technology. *J. Korean Libr. Inf. Sci. Soc.* **2008**, *39*, 377–395.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).