DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

Together in Practice: Comparing LCC and DDC Assignment Across Library of Congress Bibliographic Records

Kai Li[1], **Inkyung Choi**[2], **Jessica Yi-Yun Cheng**[3], **Zach Jenkins**[1], and **Brian Dobreski**[1]

[1] University of Tennessee, Knoxville, [2] Sungkyunkwan University, [3] Rutgers University

Abstract

Library of Congress Classification (LCC) and Dewey Decimal Classification (DDC) are two of the most widely used knowledge organization systems in libraries, yet empirical understanding of how they align and diverge in cataloging practice at scale remains limited. This paper examines co-assignment patterns between LCC and DDC classes using 4,042,962 dual-classified bibliographic records drawn from the Library of Congress's book catalog. Through descriptive quantitative analysis and bipartite network analysis, we identify areas of strong and weak structural correspondence between the two systems. Results reveal that well-defined humanities disciplines — including law, fine arts, religion, literature, and history — exhibit high one-to-one alignment, while broader and more applied domains such as social sciences, technology, and computer and information science show markedly dispersed cross-system mappings. A structural asymmetry is also evident: LCC classes tend to map more sharply to single DDC counterparts than vice versa. Network analysis identifies seven disciplinary communities and highlights Social Sciences and Technology as key interdisciplinary hubs, while second-level classes reveal contrasting topologies — a hub-centric star structure for LCC:G and a fragmented, multi-polar constellation for DDC:6XX. These findings carry practical implications for library reclassification projects, cataloging workflows, and the reuse of bibliographic metadata in emerging technological environments.

1. Introduction

Bibliographic records, standardized metadata descriptions for resources within a library's collection, are widely shared and reused among institutions. Among the most prominent providers of these records is the Library of Congress (LC). As the *de facto* national library of the United States, LC is recognized as a source of authoritative and high-quality metadata, and their records have been reused in libraries around the world [1]. While, unsurprisingly, LC uses Library of Congress Classification (LCC) for its collections, their records also often contain classification numbers from Dewey Decimal Classification (DDC). Bibliographic classification numbers do more than identify a resource's place on the shelves; they also position the resource within a coherent, conceptual context [2]. As knowledge organization systems (KOS), LCC and DDC feature different approaches, design principles, and conceptual coverage and contexts.

Traditionally, these differences have been explored via comparisons of the schedules themselves. Prior work has largely focused on theoretical correspondences between classification schemes or small-scale mappings between classes [3], [4]. While these approaches reveal structural similarities, they do not capture how classification systems are applied in real-world cataloging practice at scale [5], [6]. As a result, there remains

Full Paper

Available Aug 01, 2026

Correspondence to
Brian Dobreski
bdobresk@utk.edu

limited empirical understanding of how LCC and DDC align or diverge when jointly assigned to the same resources.

The increasing availability of large-scale bibliographic datasets and advances in computational methods offer new opportunities to address this gap. By leveraging millions of co-classified records, it becomes possible to systematically examine patterns of alignment and divergence that are not observable through traditional manual or small-scale analyses. Records from LC, then, provide a unique opportunity to empirically compare the systems in practice within a large, high-quality bibliographic dataset.

In this paper, we present findings from a comparison of assigned LCC and DDC numbers within a set of 11 million bibliographic records from LC. As such, the key research question of this research paper is: how are LCC and DDC classes co-assigned with each other in practice in bibliographic records? To address this research question, we examined the agreement of co-assignment (measured by the share of records assigned to the dominant corresponding class) and the structural topology of the network composed of the classes at both the first and second classificatory levels.

Using descriptive quantitative analysis and network analysis, we provide initial insight into these questions and visualize several trends and patterns of interest. The results of this work offer new contributions through the use of empirical data to show how, in practice rather than theory, LCC and DDC align or diverge in practice, and provide further insight into significant conceptual differences between these systems. More broadly, findings hold implications for library practice, metadata reuse, and cataloger training and workflows.

2. Literature Review

Library of Congress Classification and Dewey Decimal Classification are two of the most prominent systems used to classify library resources. LCC was originally devised in the early nineteenth century to classify the holdings of the Library of Congress. This classification began as an enumerative system requiring little number building, but over time has moved toward faceting and number building relying on an increasing number of tables [7]. Although LCC is detailed and widely-used, its original design was to meet the classification needs of the Library of Congress's own collection, rather than universal application. This explains the uneven distribution of LCC classes and bias toward materials from the United States [8]. DDC is similarly widely-used for classification in libraries and is distinguished from LCC by its brevity and usage in libraries with comparatively smaller collections. This system was first published by Melvil Dewey in 1876. Dewey created ten main classes that reflected the areas of study in his time. In the time since, a disparity has developed between different classes, as classes such as "Language" have remained stable, while others like "Technology" have changed considerably. Additionally, the strict structural hierarchy of the DDC has resulted in frequently criticized placement of certain subjects, including "Psychology" as a division of "Philosophy" and "Sports and Amusements" inside "The Arts" [8]. Both systems maintain their own distinct structures that have led to various differences and distinct benefits and drawbacks.

Existing literature comparing Library of Congress Classification and Dewey Decimal Classification heavily focuses on the utilization of classification numbers within the schemes, as well as the ways that reclassification, which typically occurs in academic libraries and from DDC to LCC, affects student and staff perspectives of the library. Morris used bibliographic data from OCLC and found that classification numbers largely mirrored the Bradford distribution, with twenty percent of classification numbers containing eighty percent of items [9]. Sultanik surveyed library student workers who worked at her institution during reclassification and found that three-fourths of library student workers preferred LCC to DDC, citing ease of use and intuitiveness [10]. O'Neill et al. analyzed the LC MARC Class File and found classes between the classification schemes that map to one another very effectively and, conversely, very ineffectively [11]. Other findings from existing LCC-DDC comparisons include highlighting use cases for either system [3], [12], drawbacks of either system [12], and presence of bias toward White Anglo-Saxon Protestants in both LCC and DDC [3]. While these studies provide important insights into structural differences, user perceptions, and biases within classification systems, they are largely limited to theoretical comparisons, small-scale mappings, or localized case studies. As a result, they do not fully capture how LCC and DDC interact when jointly applied to the same resources at scale.

At the same time, a growing body of research in bibliometrics and information science has demonstrated the value of network-based approaches for analyzing relationships between large-scale data, particularly the topology and dynamics in any type of social network [13], [14]. However, such approaches have rarely been applied to the comparison of large-scale bibliographic records. This study builds on and extends prior works by adopting this network-based approach to examine co-assignment patterns between LCC and DDC.

3. Methods

In this study, we used the open-access version of the dataset provided by the MARC Distribution Services [15] published in 2019. This dataset contains more than 11 million metadata records for books in the library catalog. We selected this dataset because of the key role played by the Library of Congress in the development and maintenance of cataloging standards [16] and the high quality and representativeness of the dataset as proven by previous research [17]. Because this dataset represents a single snapshot from 2019, our findings reflect co-assignment patterns as they existed at that point in time and should be interpreted accordingly; longitudinal analysis tracking changes to these patterns is reserved for future work.

From all the records in the dataset, we extracted all records with any classification number from both classification systems (i.e., the record contains both 050 and 082 fields from the MARC21 format) to construct networks to understand the relationship between the two classification systems. We excluded alternative numbers from both systems, such as LAW in the LCC system and [Fic] in the DDC system. Out of all 11.28 million records, 9.98 million records contain a valid LCC number and 4.80 million contain a valid DDC number. We selected all 4,042,962 records that have both numbers into our sample.

In network science terms, a *graph* is a mathematical structure consisting of *nodes* connected by *edges*. In our study, the two classification systems are jointly represented as a *bipartite graph* — a graph whose nodes are divided into two disjoint sets such that edges only connect nodes across sets, never within the same set. Specifically, LCC classes and DDC classes constitute the two node sets, and an edge is drawn between an LCC node and a DDC node whenever the two classes co-appear in the same bibliographic record. Edges are *weighted* by the number of records supporting each co-occurrence pair, where stronger co-assignment relationships are reflected by higher edge weights. This bipartite representation makes it possible to analyze the structural topology of cross-system co-assignment.

To understand the relationship between both classification systems, we first constructed bipartite networks between the two systems, where we used the second-level categories (such as BC and GA in LCC and 010 and 240 in DDC) as nodes and the co-appearance of two categories in the same bibliographic record as the path between two nodes. By using this level, we were able to reveal more granular features within the network of the two classification systems.

We further applied community detection algorithms on the network to identify communities within the network, i.e., clusters of LCC and DDC categories that are similar with each other. All the network graphs are based on the full population (i.e., 4,042,962 dual-classified records). Community detection methods are discussed under individual graphs in the Results section.

4. Results

4.1. How are first-level classes co-assigned together?

Table 1 presents LCC categories with strong one-to-one DDC alignment, based on 4,042,962 dual-classified bibliographic records. For instance, LCC K (Law) maps overwhelmingly to DDC 300 (Social Sciences) at 96.58%, while LCC N (Fine Arts) maps to DDC 700 (Arts and Recreation) at 93.72%, and LCC P (Language and Literature) maps to DDC 800 (Literature) at 76.97%. Similarly, the history-related LCC classes C (Auxiliary Sciences of History), D (World History), E (History of the Americas), and F (History of the Americas) all converge on DDC 900 (History and Geography), with shares ranging from 62.85% (E) to 86.62% (F). These high-concentration mappings suggest strong structural correspondence between the two systems in well-defined disciplinary areas.

In contrast, certain LCC classes display more dispersed mappings, indicating conceptual breadth that does not reduce neatly to a single DDC counterpart. LCC G (Geography,

Table 1. *LCC categories with strong one-to-one DDC mappings*

LCC Main Class	Records in class	Top DDC Main Class	Top Class Share	Second DDC Main Class	Second Class Share
C (Auxiliary Sciences of History)	54,848	900 (History & Geography)	80.25%	000 (Computer Science, Information & General Works)	5.00%
D (World History)	294,542	900 (History & Geography)	81.95%	300 (Social Sciences)	14.52%
E (History of the Americas)	78,128	900 (History & Geography)	62.85%	300 (Social Sciences)	30.57%
F (History of the Americas)	139,888	900 (History & Geography)	86.62%	300 (Social Sciences)	9.56%
K (Law)	232,680	300 (Social Sciences)	96.58%	600 (Technology)	0.95%
N (Fine Arts)	137,690	700 (Arts & Recreation)	93.72%	900 (History & Geography)	2.15%
P (Language and Literature)	756,593	800 (Literature)	76.97%	400 (Language)	9.09%

Table 2. *LCC categories with dispersed DDC mappings*

LCC Main Class	Records in class	Top DDC Main Class	Top Class Share	Second DDC Main Class	Second Class Share
G (Geography, Anthropology, Recreation)	140,261	700 (Arts & Recreation)	39.28%	300 (Social Sciences)	27.99%
V (Naval Science)	11,812	300 (Social Sciences)	47.34%	600 (Technology)	44.48%

Table 3. *DDC categories with strong one-to-one LCC mappings*

DDC Main Class	Records in class	Top LCC Main Class	Top Class Share	Second LCC Main Class	Second Class Share
200 (Religion)	240,150	B (Philosophy, Psychology, Religion)	95.23%	H (Social Sciences)	0.85%
400 (Language)	73,557	P (Language and Literature)	93.51%	L (Education)	3.45%
500 (Science)	276,905	Q (Science)	88.52%	G (Geography, Anthropology, Recreation)	5.26%
800 (Literature)	595,272	P (Language and Literature)	97.83%	M (Music)	0.49%

Anthropology, Recreation) splits between DDC 700 (39.28%) and DDC 300 (27.99%), reflecting its interdisciplinary scope across recreation, social science, and physical geography. Perhaps most striking is LCC V (Naval Science), where the top DDC category (300, at 47.34%) and the second category (600, at 44.48%) are nearly equal, suggesting that naval science materials are almost evenly distributed between the social science and technology dimensions of DDC. The social sciences hub is also apparent from the LCC side: six LCC classes, including H (Social Sciences), J (Political Science), K (Law), L (Education), U (Military Science), V (Naval Science), have DDC 300 as their top category, confirming the broad absorptive scope of the DDC Social Sciences division.

Tables 3-4 reverses the perspective, showing the top two LCC categories for selected DDC main classes. As with Table 1, several DDC categories exhibit highly concentrated mappings to a single LCC class, confirming strong bidirectional alignment in certain disciplines. DDC 800 (Literature) maps almost exclusively to LCC P (Language and Literature) at 97.83%, DDC 200 (Religion) to LCC B (Philosophy, Psychology, Religion) at 95.23%, and DDC 400 (Language) to LCC P at 93.51%. DDC 500 (Science) also shows a dominant mapping to LCC Q (Science) at 88.52%, with only a modest secondary link to LCC G (Geography, Anthropology, Recreation) (5.26%).

By contrast, DDC categories that cover broader or more applied domains show markedly more dispersed mappings. DDC 300 (Social Sciences), the largest category by record count (1,116,008 records), distributes across LCC H (Social Sciences) (36.30%) and LCC K (Law) (20.14%), with the remaining share spread across multiple other LCC

Table 4. *DDC categories with disperse LCC mappings*

DDC Main Class	Records in class	Top LCC Main Class	Top Class Share	Second LCC Main Class	Second Class Share
000 (Computer Science, Information & General Works)	159,576	Q (Science)	31.17%	Z (Bibliography, Library Science)	22.79%
300 (Social Sciences)	1,116,008	H (Social Sciences)	36.30%	K (Law)	20.14%
600 (Technology)	606,217	T (Technology)	36.53%	R (Medicine)	28.71%
700 (Arts & Recreation)	353,082	N (Fine Arts)	36.55%	G (Geography, Anthropology, Recreation)	15.61%

classes — reflecting the fragmentation of social science topics across the LCC scheme. DDC 600 (Technology) similarly splits between LCC T (Technology) (36.53%) and LCC R (Medicine) (28.71%), consistent with the division of applied sciences into engineering and medical subdomains within LCC. DDC 000 (Computer Science, Information, and General Works) stands out as the most dispersed category, with its top LCC class Q (Science) accounting for only 31.17% of records and the second Z (Bibliography, Library Science) for 22.79%, highlighting the residual, catch-all nature of this DDC division. Comparing the two tables, the asymmetry between the systems becomes evident: while many LCC classes map sharply to a single DDC counterpart, the reverse is not always true, particularly for DDC's broader divisions (000 (Computer Science, Information & General Works), 300 (Social Sciences), 600 (Technology), 700 (Arts & Recreation)), which absorb materials from multiple LCC classes.

4.2. *Topology of co-assignment networks between the two systems*

Figure 1 visualizes the top-level bipartite network linking Library of Congress Classification (LCC) classes and Dewey Decimal Classification (DDC) hundreds. The network comprises all 21 LCC categories and all 10 DDC classes. Overall, this graph shows seven different communities identified by the Louvain community detection method, namely Philosophy / Psychology / Religion (brown), Physical / Information Sciences (red), History (orange), Technology / Medicine (green), Social Sciences (blue), Arts and Recreation (purple), and Languages and Literatures (pink). The structure of the communities suggests that the Social Sciences (DDC 300, LCC H) and Technology (DDC 600) serve as the interdisciplinary topics that connect to many other nodes/topics.

Next, we took a closer look at some of the most dispersed classes from the two systems. Figure 2 presents two network graphs composed of second-level LCC and DDC nodes centered on LCC:G (left) and DDC:6XX (right). To improve visual clarity, both subgraphs retain only the top 50% of nodes ranked by weighted degree strength — the sum of edge weights across all of a node's connections — thereby excluding peripheral nodes with minimal co-assignment activity.

The network of LCC:G domain (Geography, Anthropology, Recreation) comprises 99 nodes (LCC = 9, DDC = 90) and 189 edges, with density of 0.039. Topologically, the structure is strongly hub-centric, with one dominant connector (GV, i.e., Recreation, Leisure) and multiple smaller, more specialized G-subdomains forming a tighter cluster. This "star" topology is reflected in a low average path length of 2.29, indicating a very direct, one-to-many mapping. Notably, there is one stronger GV and 79X association (shown by the much larger 79X orange node), suggesting frequent co-occurrence be-

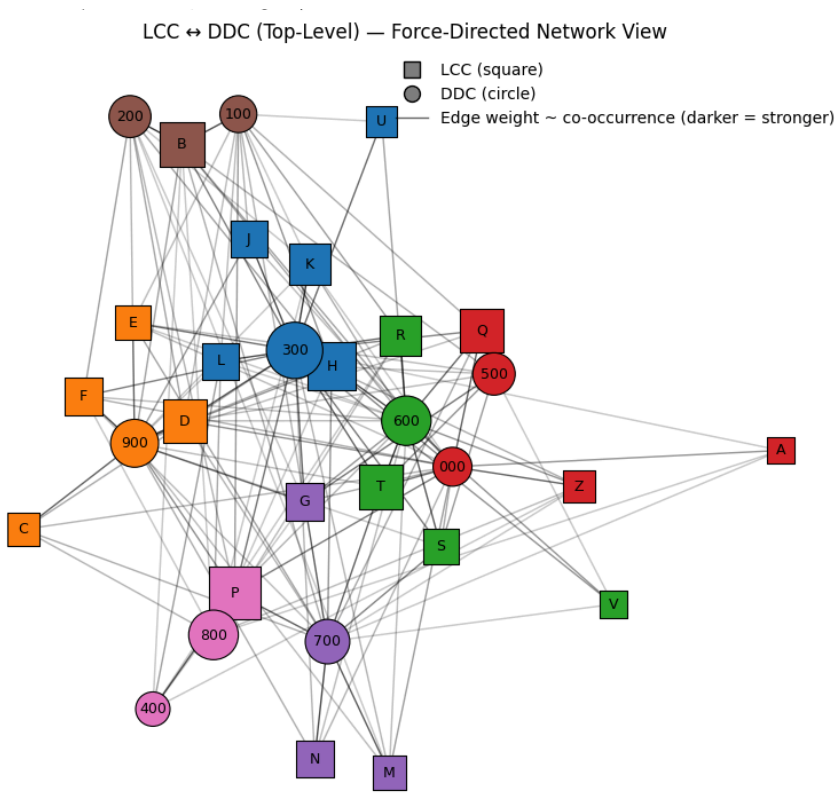

Figure 1. *First-level class bipartite network between LCC and DDC*

tween LCC recreation and DDC recreational/performing arts areas. Other G-subclasses (e.g., GA/GB/GN/GR/GT) connect to only a small set of DDC categories. This means the LCC "G" area is not evenly cross-classified, some parts (like GV) overlap with many topics, but these other G subcategories stay more narrowly tied to specific subjects.

As compared, the network of DDC 6XX class presents a more fragmented, multi-polar landscape. This network comprises 91 nodes (LCC = 81, DDC = 10) and 167 edges, with density of 0.041. Despite having a similar total node count (91), it possesses a significantly higher average path length of 3.17, signaling wider distribution of topics within this network. Visually, the network transitions from a single hub to a "constellation" of specialized clusters—most notably around 61X (Medicine & Health), 62X (Engineering), where 62X links strongly to technology-oriented LCC subclasses (e.g., T in LCC), and 61X links broadly to medical/health-oriented LCC subclasses (e.g., R in LCC). Smaller 6XX subclasses, such as 63X (Agriculture), 64X (Home & family management), and 65X (Management & public relations) link towards a range of specialized LCC topics such as agriculture (e.g., S*), management (e.g., HF), and manufacturing (e.g., TS), indicating that these topics are present in the network but less central than medicine or technology-oriented classes.

5. Discussion and Conclusion

While, expectedly, the majority of records in the dataset contained LCC numbers, around 43% contained DDC, resulting in a large set of records with numbers from both systems. Overall, points of greatest agreement witnessed between the two systems

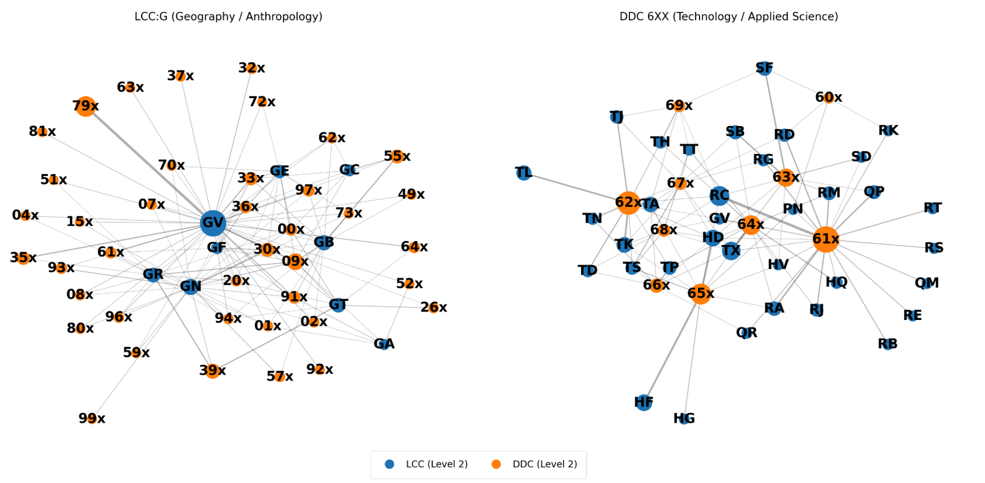

Figure 2. *Second-level class bipartite networks centered on G class of LCC (left) and 6XX class of DDC (right)*

tended toward topics from the humanities, including history, fine arts, religion, and literature, as well as law. Points of greatest disagreement tended to occur with the social and applied sciences; geography, technology, social science, and computer and information science showed the least consistent co-assignments. To some extent, this may reflect disciplinary areas that have undergone the most growth and transformation since the inception of the two systems in the nineteenth century. It may also be indicative of areas of study that are more interdisciplinary in methods and approach. Finally, some points of disagreement may simply show where each respective system is less conceptually coherent. As visualized above, the LCC:G class encompasses a variety of topics and maps inconsistently across a wide range of DDC numbers. Since 1910, the G class has grown to include a variety of loosely related topics, reflecting its role as a "bridge class" between history and social science, as well as a more abstract organizing principle (i.e., the place of humans in the world) [18]. Similarly, DDC's 600 class has grown in size and scope as technology has progressed, leading to inconsistent mappings across various areas of LCC over time.

Findings presented here hold implications for libraries, library staff, and users of library metadata. For libraries involved in reclassification projects, results show parts of the collection where the process may be more challenging, for example, the DDC 600s when moving from DDC to LCC. Similarly, for catalogers, areas of disagreement noted above represent topics where assigning a number from one system based on one from another could require more caution. Finally, results offer additional insight into classification data within bibliographic datasets that may see increasing reuse in emerging technological environments.

Future work on this ongoing project will continue to examine co-assignment between the two systems at a finer level of detail, and move beyond books to examine classification patterns for other formats as well. We will also explore how the topological networks between the two systems shifted over time in response to changes in LCC and DDC. In addition, we will apply the node2vec algorithm to investigate how the vectorized networks change over time, which will enable more nuanced analysis of structural relationships between classification categories and the drift of nodes over time. We hope that these future directions will lead to a deeper understanding of how human knowledge evolves and how knowledge organization systems can better support such developments and their evolving needs.

References

[1] R. Fischer and R. Lugg, "Study of the North American MARC records marketplace," Library of Congress, 2009.

[2] S. Batley, *Classification in theory and practice*. Chandos Publishing, 2014.

[3] P. Hider and M. Coe, "Academic disciplines in the context of library classification: Mapping university faculty structures to the DDC and LCC schemes," *Cataloging & Classification Quarterly*, vol. 60, no. 2, pp. 194–213, 2022, doi: 10.1080/01639374.2022.2040675.

[4] A. Slavic, "Faceted classification: Management and use," *Axiomathas*, vol. 18, no. 2, pp. 257–271, 2008, doi: 10.48550/arxiv.1705.07047.

[5] B. Hjørland, "Classification," *Knowledge Organization*, vol. 44, no. 2, pp. 97–128, 2017, doi: 10.5771/0943-7444-2017-2-97.

[6] J. Mai, "The modernity of classification," *Journal of Documentation*, vol. 67, no. 4, pp. 710–730, 2011, doi: 10.1108/00220411111145061.

[7] L. M. Chan and T. L. Hodges, Eds., "Library of Congress Classification (LCC)," in *Encyclopedia of Library and Information Science*, 4th ed., CRC Press, 2017, pp. 2847–2855. doi: 10.1081/E-ELIS4-120043714.

[8] L. M. Chan and A. Salaba, *Cataloging and Classification: An Introduction*. Rowman & Littlefield Publishers, Incorporated., 2017.

[9] L. R. Morris, "The frequency of use of Library of Congress Classification numbers and Dewey Decimal Classification numbers in the MARC file in the field of library science," *Technical Services Quarterly*, vol. 8, no. 1, pp. 37–49, 1991, doi: 10.1300/J124v08n01_05.

[10] A. G. Sultanik, "A case study on a library classification's effect on library student workers' perception on library work," *Christian Librarian*, vol. 65, no. 1, pp. 3–15, 2022.

[11] E. T. O'Neill, M. Dillon, and D. Vizine-Goetz, "Class dispersion between the Library of Congress Classification and the Dewey Decimal Classification," *Journal of the American Society for Information Science*, vol. 38, no. 3, pp. 197–205, 1987, doi: 10.1002/(SICI)1097-4571(198705)38:3%253C197::AID-ASI8%253E3.0.CO;2-2.

[12] N. Salem and A. M. K. Shehata, "Electronic games classification in the library of congress and Dewey classification schemes: A comparative study," *Global Knowledge, Memory and Communication*, vol. 71, no. 6/7, pp. 468–484, 2022, doi: 10.1108/GKMC-10-2020-0155.

[13] E. Otte and R. Rousseau, "Social network analysis: a powerful strategy, also for the information sciences," *Journal of information Science*, vol. 28, no. 6, pp. 441–453, 2002, doi: 10.1177/016555150202800601.

[14] J. McLevey and R. McIlroy-Young, "Introducing metaknowledge: Software for computational research in information science, network analysis, and science of science," *Journal of Informetrics*, vol. 11, no. 1, pp. 176–197, 2017, doi: 10.1016/j.joi.2016.12.005.

[15] Library of Congress Cataloging Distribution Service, "Books All : MDS Connect Dataset." [Online]. Available: https://www.loc.gov/item/2020445551

[16] J. Y. Cole Jr and A. J. Kinney, Eds., "Library of Congress," in *Encyclopedia of Library and Information Science*, 3rd ed., Routledge, 2003, p. 1593.

[17] K. Li, B. Dobreski, and M. Busch, "How are library cataloging metadata used differently over time: a large-scale quantitative analysis of the Library of Congress catalog," *Journal of Documentation*, vol. 81, no. 2, pp. 385–402, 2025, doi: 10.1108/JD-08-2024-0199.

[18] L. M. Chan and A. Salaba, *Cataloging and Classification: An Introduction*. Rowman & Littlefield Publishers, Incorporated., 2017.