The Bipartite Network Study of the Library Book Lending System

Nan-nan Li and Ning Zhang

Business School, University of Shanghai for Science and Technology
Shanghai, China 200093
lnn19850625@126.com, zhangning@usst.edu.cn

Abstract. Through collecting the library lending information of the University of Shanghai for Science and Technology during one year, we build the database between the books and readers, and then construct a bipartite network to describe the relationships. We respectively establish the corresponding un-weighted and weighted bipartite network through the borrowing relationship and the reading days, thereout obtain the statistical properties via the theory and methods of complex network. We find all the properties follow exponential distribution and there is a positive correlation between the relevant properties in un-weighted and weighted networks. The un-weighted properties can describe the cooperation situation and configuration, but the properties with node weight may describe the competition results. Besides, we discuss the practical significance for the double relationship and the statistical properties. Further more, we propose a library personal recommendation system for developing the library humanity design resumptively.

Keywords: complex network, bipartite network, node weight, personal recommendation.

1 Introduction

The analytic methods of complex network have been widely used in various fields [1-7] to describe the relationship between the individual and the collective behavior of the system. Thereinto, each individual corresponds to the different node, and individual interactions correspond to the edge of the two nodes. In the un-weighted network, the edge only gives the qualitative description that whether there is an edge between the nodes. However, in most cases, the distinctness of the interactional strength between nodes plays a vital role. In this way, the edge-weight is introduced to describe the interaction difference, thus form the weighted network [8, 9]. As a major expressive form of complex network, the bipartite network can commendably represent the original information, which has been received more and more attention by researchers. A series of cooperation network in nature and society can be described as the bipartite network constituted with act and actor [10-12].

Library is a treasury trove of spiritual wealth of mankind, an important part of human spiritual civilization and an inexhaustible knowledge of human resources. Books lending is one of the ways to provide services of the library. The quantities of lending

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 773 – 782, 2009.

books directly reflect the readers' demand, also is used to measure the book using effectiveness and be regarded as an important factor toward the purchase of books. From the view of system, library lending network is a typical complex network, as well as a typical cooperation and competition network. Through lending process, we can establish certain links between books and readers, thus constitute the bipartite network. However, the prior empirical researches of the library network [13, 14] are un-weighted and lack definite rationality due to the unclassified books. Compared with the prior research, this paper improves as follows:

(i) In the previous study, book is expressed by book barcode. However, in the library every book has different barcodes, and the number of the book with a same edition is usually more than one. Therefore, if a reader borrows some books with the same edition, it will be disposed as different books. Here, based on the disposal about the bibliography storeroom and library collection, we use call number instead of barcode, and then obtain the statistic of the books with same edition.

(ii) Suppose a reader read a book with one or twenty days, the significance of the book is clearly different for this reader. Hence, if only using the times that the reader borrow a same edition book as the weight, the significance is not great. Dealing with the borrow-return record time based on the library computer system, we get the duration that the reader borrow and return the same edition book with days, and regard them as the weight between this reader and book. If the reader borrows the same edition book several times, the weight will be the sum of each time.

Therefore, through the reasonable classification towards all books, we study the un-weighted and weighted bipartite network of the library lending relations, then obtain several accurate results and analyze the collaboration-competition relationship, hope to provide some new empirical foundation for the library research and help the procurement staffs to better understand the needs situation of readers. Finally, using a proper weighting method [15], we propose a library recommendation system.

2 Constructing the Library Lending Bipartite Network

In this paper, the data are from the library lending situation of the University of Shanghai for Science and Technology (USST) during one year, table 1 shows the specific data format. During this period, the total number of the borrowed books is 83,959 (different book barcodes), which with different editions is about 51,084 kinds (different call numbers), and the total number of readers is 12,610. First, we introduce the statistical situation of original data in the library, as table 1, each reader corresponds to a reader barcode, each book corresponds to a book barcode, and all the barcodes are different with each other. If one reader borrows a book, there will be a connection between the reader barcode and the book barcode. In the format data of the library bibliographic collection, a book call number corresponds to several different book barcodes in table 2. This is because the library will provide several books of the same edition (the title, author, publishing company, publishing time are the same) to lend for readers, these books of a same edition is indicated with the same call number (convenience for readers to lookup), each book has a different bar code for distinction (convenience for the library staffs to take notes). In the actual situation, if a

Table 1. The data format of the library book

Book barcode	Reader barcode	Reader grade	Department	Operation	Dealing time
842851	691	graduate student	Business School	borrow	2007-10-26
801569	983	graduate student	Business School	return	2007-10-26
689855	46	student	English School	return	2007-10-26
848798	7938	student	Business School	return	2007-10-26
850723	246	teacher	Physics School	return	2007-10-26
734438	1658	student	Business School	borrow	2007-10-26

Table 2. The data format of the library bibliography collection

Call number	Book barcode	Superscription
O551/Z53	E031208	Heat and trermodynamics
O551/Z53	E038805	Heat and trermodynamics
I565.44/A933	E010449	Pride and prejudice
I565.44/A933	E028330	Pride and prejudice
I565.44/A933	E031156	Pride and prejudice
I565.44/A933	E028331	Pride and prejudice

reader borrows two books with the same edition, we should regard them as the same book. Thus, in order to exhibit the lending situation more reasonably, we use call number instead of book bar code to establish the library network by SQL.

In the library bipartite network, the nodes can be divided into two types. One type expresses the books, named “acts”; the other expresses the readers who participate in the acts, called “actors”. If there is a borrowing relationship between them, the two will be connected by a line forming an edge. In the entire bipartite network, there is a kind of collaboration-competition relationship among the acts, that is, the lending quantity of all the books present the level of library service , and the quality and popularity of the books form a kind of borrowing competition for readers. On the other hand, the books borrowed by the same reader form a type of collaboration-competition relationship, namely, these books together constitute the reader's knowledge systems, and the books compete against each other to provide service owing to the reader’s limited energy. In each act, the relations between the actors also represent both collaboration and competition, the number of readers that borrow the same book constitutes the reading value of this book, and because the number of readers is relatively larger than the books, the processes form the competition to borrow the same book for the readers.

3 The Properties of the Bipartite Network for the Library Lending System

For un-weighted bipartite network, we study two properties named act size and act degree which merely consider the borrowed times. In the study of the weighted

bipartite network, we regard the reading days as the edge-weight between the reader and the book, and obtain the node strength of books and readers respectively.

3.1 Act Size and Act Degree

Act size is the number of actors connected with an act, that is, the number of readers who borrow this book during the year, roughly indicate the competitiveness of the book. As shown in Fig. 1, except for the small impact on the tail, the distribution can be well approximated by an exponential form

$$y = 0.9519e^{-0.258x}$$

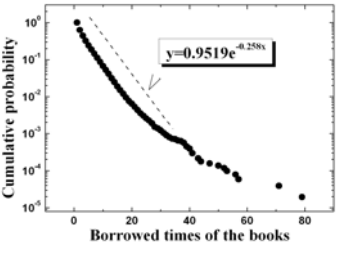

Fig. 1. The cumulative distribution of act size

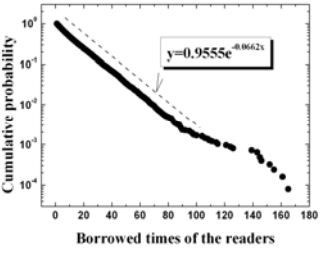

Fig. 2. The cumulative distribution of act degree

Table 3 shows the distribution of the number of borrowed times, two books are borrowed more than 70 times at best, 54.62 percent of the books are borrowed 2-9 times, and 36.70 percent are one time. The average size of the act is 9.44, that is, the average

Table 3. The distribution of the book borrowed times

Borrowed times	79，71	50-59	40-49	30-39	20-29	10-19	2-9	1
Books' number	2	5	13	42	268	3115	27903	18750
Percent	0.00	0.01	0.03	0.08	0.53	6.10	54.62	36.70

Table 4. The category distribution of the former hundred books

Book category	Chinese literature	English language	Computer	Anglo-American Literature	Mathematics	physics
Books' number	29	20	16	11	4	4

Table 5. The borrowed times distribution of the readers

Borrowed times	1	2-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	79-165
Amount	1040	5452	3010	1491	750	434	202	111	57	62
Percent	8.25	43.24	23.9	11.8	5.95	3.44	1.60	0.88	0.45	0.49

borrowed times of all books in the year are 9.44. Surveying the category of the first 100 books with the highest borrowed times as shown in table 4, the most frequent types are Chinese literature, English language and computer science. In the computer category, the number of books on Matlab and C++ is seven and four respectively accounting for the most. Similarly, among the four mathematic books, there are three about probability theory and mathematical statistics. Therefore, the library should be increase relevant procurement, as far as possible to meet the needs of readers.

Act degree of an actor node is defined by the total number of borrowed books for a reader during one year, which expresses the competitive size of the reader. Act degree distribution describes the situation of the readers in the library, Fig. 2 shows the cumulative distribution is approximated by an exponential function $y = 0.9555e^{-0.0662x}$ on the whole. Known from table 5, 43.24 percent of readers borrow 2-9 books in the year. The average of actor degree for all readers is 14.16, that is, the average number of the borrowed books for every reader is 14.16 during the year. Through the statistic about the first 100 readers with the highest times, there are 4 teachers, 71 graduate students and 24 other students. With practice, teachers have abundant knowledge and more engage in single areas, so they borrow books more specifically, also can purchase books continually owing to better economic status, while graduate students have fewer knowledge relatively, thirst for knowledge strongly, and need to gain a large number of relevant literature in study and research, also have better advantage of library privileges (borrowing 10 books one time) than other students (5 books), so their times are higher than other students.

3.2 Node Strength Distribution

Owing to the detailed records that readers borrow-return each book during one year, we can obtain the reading time using SQL and ACCESS. If a reader has been borrowed a book, then we establish an edge between them, and regard the reading time as the weight of this edge. In this way, the weighted bipartite network is established. The node strength is a natural extend corresponding to node degree, and the role of the node strength distribution $p(s)$ is similar to degree distribution $p(k)$, which denote the probability of one node having node strength S. It is defined as

$$s_i = \sum_{j \in N_i} w_{ij} \tag{1}$$

Where N_i represents the collection of neighbor nodes of node i. Node strength considers not only the close neighbor of a node, but also the linked weights between this node and the neighbors, which represent the integrated information of this node.

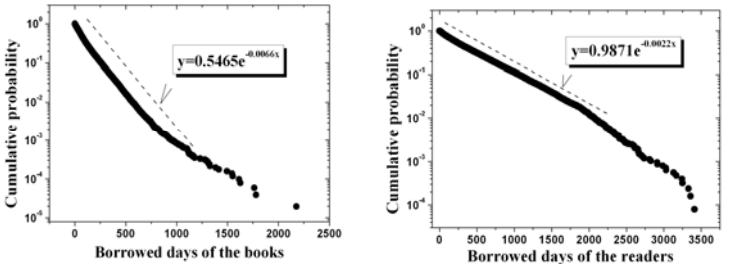

Fig. 3. The cumulative distribution of the node strength, the left is about book and right for reader

Table 6. The distribution of the book node strength

Node strength	0-30	31-60	61-100	101-200	201-300	301-400	401-500	501-600	601-2176
Amount	14431	9740	8635	10204	4139	1955	960	425	415
Percent	28.25	19.07	16.90	19.97	8.10	3.82	1.88	0.83	0.81

Table 7. The former hundred books of the highest node strength

Specialty	English language	Computer	Math	Kinetics	Mechanics	Physics	Anglo-American literature	Telecom
Number	29	25	9	8	6	5	5	4

The node strength of a book denotes the sum of the reading days by every reader for this book, which indicate the competitive size more accurately. Fig. 3 shows the cumulative distribution can well fit with an exponential function $y = 0.5465e^{-0.0066x}$. The average of the node strength is 106.74, namely, every book is kept for 106.74 days by readers during the year. Table 6 lists the specific situation distribution of the days, 92.29 percent is within 300 days, less than a year's time. This shows that the book in the library can meet the reader's demand basically. As the books with larger node strength represent the longer time of this book between borrow and return, namely more competitiveness, which indicate the reader's interest and trend, so we analyze them as focus. Table 7 shows that English language and computer books are still on the top, but the literature books decrease significantly. This indicates the concept of readers is to place study first, entertainment second. At the same time, the readers pay great attention to English and computer books which have become an indispensable tool. In addition, the number of books in table 7 is uniform on dynamic engineering, mechanics, physics, radio and telecommunications technologies, which indicate that the readers of different

specialties all will borrow the relatively professional books during study and research, which further explain the accuracy of the empirical work.

Fig. 3 shows the cumulative distribution of the reader node strength, the main part of the distribution follows an exponential function $y = 0.9871e^{-0.0022x}$. This node strength denotes the total number of the days the reader reads every book in one year, a more accurate competitive expression. The average of the node strength for all readers is 423.44 days, and the node with the greatest strength is a teacher in kinetic college. Through the analysis of prior 100 readers with the greatest strength, these readers are 29 teachers, 70 graduate students and one other student. Because the stated reading time of teacher (90 days) is longer than graduate student (Doctor for 90 days, 30 days for master) and other students (30 days), the teacher total day increase accordingly.

3.3 The Contrast of Un-weighted and Weighted Network

In the bipartite network, a book act size in un-weighted network is the total borrowed times of the book by all readers, and book node strength in weighted network is the sum of the reading days by every reader, both express the book competitiveness. Similarly, actor degree is the total borrowed times of a reader in un-weighted network, reader node strength in weighted network is the total reading days of the reader, and also both show the reader competitiveness. Then, is there a certain relation between the two pairs of distribution in the form of un-weight and weight network? First of all, they follow the same distribution on the whole, the same conclusion in previous study [16], the reasons of such mechanism are explained deep in a network evolution model [17, 18]. Secondly, as shown in Figure 4, there is a positive correlation between them. More exactly, the node strength follows a power law with respective to act size, also to actor degree. This means that, the more times a book is borrowed, the longer time it can be kept. In the same way, the more times a reader borrow, the longer time he can keep. Through studying the statistical relationship between the act degree and node strength, we know that there is larger competitiveness when the actor participates more acts. In practice, as the library restricts the number of the borrowed books and the preserving-book time for the readers, the readers usually faster return the less useful books in order to borrow the needed books for themselves. Then borrowing a book does not represent that the reading value of this book is existent, but the length of the reading

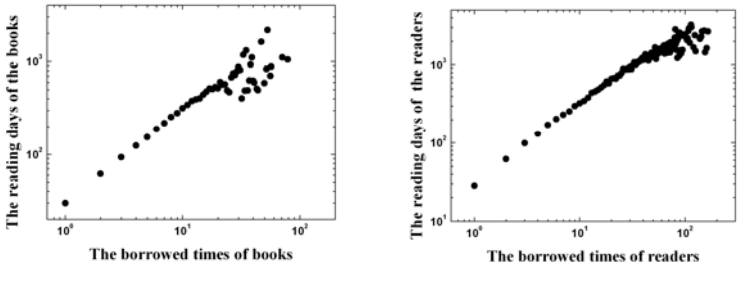

Fig. 4. The empirical results of relationship between book node strength and act size (*the left*), also reader node strength and act degree (*the right*)

days reflect the different value of the book for each readers. Therefore, weighted bipartite network can reflect the actual situation better. All these show that both of them can describe this collaboration situation and configuration, but only the latter can describe the competition results and other competition properties accurately.

4 The Library Personal Recommendation System

There are large numbers of books in library, which maybe have hundreds of kinds in one direction, and most of the same direction books also have different emphasis. Therefore, it is troublesome to find the most correlative book in them for readers. And using the historical books which a reader has been borrowed, we can discover his habits and consider them for him in the future. Here, we propose the library recommendation system which make use of the reader's historical information, and hope to provide several suggestions for the library humanity design.

A reasonable assumption is that the books which a reader have borrowed are what he like, and a recommendation algorithm aims at predicting his personal opinions on those books he have not yet collected. Based on our database, the recommendation system consists of readers and books, and denotes the book-set as $B=\{b_1,b_2,\cdots b_n,\}$ and reader-set as $R=\{r_1,r_2,\cdots,r_m\}$. If readers have borrowed some books, the recommendation system can be described by an n×m adjacent matrix $\{a_{ij}\}$, where a_{ij}=1 if r_j has already borrowed b_i and a_{ij}=0 otherwise. Through the application of the weighting method for bipartite networks presented in [15], we propose a recommendation algorithm. First, we construct the bipartite reader-book network by book-projection which is named G. Then, for a given reader r_i, we set the initial resource located on each node of G as $f(b_j)=a_{ji}$. That is to say, if the book b_j has been borrowed by r_j, then its initial resource is unit, otherwise it is zero. Apparently, the initial configuration which captures personal preferences is different for every reader. The initial resource can be understood as giving a unit recommending capacity to each collected book. According to the weighted resource allocation process discussed in [15], the final resource, denoted by the vector $\overleftrightarrow{f'}$, is $\overleftrightarrow{f'}=W\overleftrightarrow{f}$. Thus components of f' are

$$f'(b_j)=\sum_{l=1}^{n}w_{jl}f(b_l)=\sum_{l=1}^{n}w_{jl}a_{li} \tag{2}$$

For any reader r_i, all his borrowed books $b_j(1\le j\le n,a_{ji}=0)$ are sorted in the descending order of $f'(b_j)$, and those books with highest value of final resource are recommended. Note that, the calculation of $f'(b_j)$ should be repeated m times, since the initial configurations are different for every reader.

5 Conclusions

Through collecting the detailed information of the library of USST during one year, we obtain the characteristics of the library lending network, including act size distribution, act degree distribution, node strength distribution, and find they all follow exponential which imply the connecting numbers of most nodes are similar by nature, namely the nodes that is much higher or lower than average rarely exist. In this paper, it shows that the borrowed books in the library are balanced, the lending system is perfect basically, and the library can meet the readers' needs on the whole. In addition, there is a positive correlation between the distribution of act size and book node strength, the same as act degree and reader node strength. Thus, compared with the un-weighted network, the weighted network can not only describe the cooperation situation and configuration, but also the competition results.

From the analysis of the relevant statistics, we can see that the readers adhere to the concept of learning-oriented, while still read the literary book to enrich life. English and computer books are the favor, therefore the library should increase the corresponding books. In the groups of readers, the node strength of teachers are far greater than students, which fully shows that they always stand at the forefront of the discipline, and enhance academic standards actively. At the same time, graduate student have a strong thirst for knowledge and is the backbone of reader's groups.

Furthermore, we proposed a library personal recommendation system based on a proper weighting method. Through the historical borrowed books for a reader, we will give some suggestions to him when he needs some books. But now, the presented recommendation algorithm is just a rough framework whose details have not been exhaustively explored yet. Next, we will study further and model this mechanism in library.

Acknowledgements. We would like to thank the sublibrarian Ling Han and teacher Lan Chen for supplying the data of the USST library. Thanks also to teacher Da-ren He of Yang Zhou University, Ji-ming Li, Dan-rong Zhang, Geng-sheng Zhao of business School in USST for the helpful discussion. This work is supported by Shanghai Leading Academic Discipline Project (S30501).

References

1. Albert, R., Albert, I., Nakarado, G.L.: Structural vulnerability of the North American power grid. J. Physical Review E 69, 25103 (2004)
2. Albert, R., Jeong, H., Barabási, A.-L.: Diameter of the World Wide Web. J. Nature 401, 130–131 (1999)
3. Motter, A.E., Moura, A.P.S., Lai, Y.C., et al.: Topology of the conceptual network of language. J. Physical Review E 65, 065102(R)(2002)
4. Redner, S.: How popular is your paper? An empirical study of the citation distribution. J. The European Physical Journal B 4, 131–134 (1998)
5. Jeong, H., Mason, S., Barabási, A.-L., et al.: Lethality and centrality in protein networks. J. Nature 411, 41–42 (2001)

6. Jeong, H., Tombor, B., Albert, R., et al.: The large-scale organization of metabolic networks. J. Nature 407, 651–654 (2000)
7. Battiston, S., Catanzaro, M.: Statistical properties of corporate board and director networks. J. The European Physical Journal B 438, 345–352 (2004)
8. Boccaletti, S., Latora, V., Moreno, Y., et al.: Complex Networks: Structure and Dynamics. J. Physics Reports 424, 175–308 (2006)
9. Wu, J.S., Di, Z.R.: C0mplex networks in statistical physics. J. Progress in Physics 24(1), 18–46 (2004) (in Chinese)
10. Lambiotte, R., Ausloos, M.: Uncovering collective listening habits and music genres in bipartite Networks. J. Physical Review E 72, 066107 (2005)
11. Garey, M.R., Johnson, D.S.: A Guide to the Theory of NP Completeness. In: Computers and Intractability. Freeman Publishers, San Francisco (1979)
12. Scott, J.: Social Network Analysis. In: A Handbook, 2nd edn. Sage Publishers, Thousand Oaks (2002)
13. Guo, M.R.: Statistics and analysis of books lent to West Normal University's Library(2001~2004). J. The journal of the library science in Jiangxi 36(4), 27–29 (2006) (In Chinese)
14. Zhang, D.R., Zhang, N.: The study of the library lending network in complex network. Shanghai: Business School in University of Shanghai for Science and Technology, 17–20 (2007) (in Chinese)
15. Zhou, T., Ren, J., Matúš, M., Zhang, Y.C.: Bipartite network projection and personal recommendation. J. Physical Review E 76, 046115 (2007)
16. Fu, C.H., Zhang, Z.P., Chang, H., et al.: A kind of collaboration–competition networks. J. Physica A 387, 1411–1420 (2008)
17. Zhang, P., Chen, K., He, Y., et al.: Model and empirical study on some collaboration networks. J. Physica A 360, 599–616 (2006)
18. Chang, H., Su, B.B., Zhou, Y.P., et al.: Statistical Mechanics and its Applications. J. Physica A 383, 687–702 (2007)