Community Detection Methods in Library's Books and Borrowers Social Network Segmentation

Tedy Setiadi [1,2,*], Mohd Ridzwan Yaakub [2], and Azuraliza Abu Bakar [2]

[1] Department of Informatics, Universitas Ahmad Dahlan, Indonesia
[2] Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Malaysia; Email: ridzwanyaakub@ukm.edu.my (M.R.Y), azuraliza@ukm.edu.my (A.A.B.)
*Correspondence: tedy.setiadi@tif.uad.ac.id (T.S.)

***Abstract*—In this paper, we discuss the application of community detection methods to book-borrowers networks in libraries. The aim is to obtain a segment of books and borrowers that are closely linked to the lending network in the library. This study applies six community detection methods, namely Louvain, Spinglas, Walktrap, Infomap, Label Propagation Algorithm (LPA), and Greedy to identify groups of books and borrowers. Meanwhile, evaluating the effectiveness of this method uses the modularity, performance, coverage, density, community size, and community fit metrics. The results showed that the community detection method was effective in identifying book segments and related borrowers in the library lending network. The Louvain method was found to be most effective in identifying communities with higher quality and better interpretation. The results of segmentation of books and borrowers can support improving library collection management and increasing demand for books, provide insight into patterns of borrowing books to improve library services and user satisfaction.**

***Keywords*—community detection, book and borrower networks, collection management, library service**

I. Introduction

Currently, academic libraries have become centers of learning resources and sources of knowledge that have challenges in serving the needs of their users [1]. The library also acts as a data generator unit. Data in academic libraries are increasing in collections, visitors, and transactions such as borrowing books, accessing online collections, and borrowing library places. Book management is the leading business in the library. The library's book management challenges are the limited book procurement budget, limited book storage capacity, and non-optimal use of books [2, 3]. Although there has been a rapid digital transformation in e-books and other digital publications, borrowing physical books is still the library's primary service to its readers. The pattern of borrowing books can be identified by utilizing the books borrowing transaction data in the library database. The number of borrowed books directly reflects readers' demand, and used to measure the effectiveness of books' usage, and an essential factor in supplying books. In addition, it is crucial to know the user's behavior in borrowing books.

A common problem with book-lending is that libraries treat book borrowers equally, not considering the users' book loan terms. The users must follow specific processes to get the book they want. Another problem is that many books interest a few people, some of which have never been borrowed. On the other hand, some books are often borrowed, but the number is limited and often unavailable.

The book-lending process generates a particular type of data where each book-lending transaction connects a particular borrower to a particular book. The book-lending transaction can be represented as a book-lending network, called a bipartite network. In this network, a borrower is connected to one or more books that he borrows, but he is not connected to other borrowers. Similarly, a book is related to one or more borrowers who have borrowed it but is not related to any other book.

The Social Network Analysis (SNA) can provide valuable insights about borrowers and books on a macro scale, thus providing book lending services that are more proactive than reactive [4]. SNA involves analysing data for application areas, such as market segmentation, crime detection, and recommendation systems. In library management, SNA utilize author data such as articles co-authors, co-citation, and co-keyword networks for community detection and co-author recommendation [5]. For example, community detection and co-author recommendations use the co-author network model to nurture collaboration for significant academic research [6]. Researchers analyzed the book-borrowing data for the book recommendations system [2, 4, 7], online reading behavior [8], and community detection for book-borrower segmentation [9].

Community detection is one of the SNA approaches to identify groups of nodes in a network that are more dense than others [6, 10]. Several applications of community detection in library management are online learning communities detection in learning repositories [11], co-

Manuscript received January 6, 2023; revised March 1, 2023; accepted July 11, 2023; published November 10, 2023.

author networks analysis [6, 12], and book recommendations [11, 13]. The network in SNA can be represented in a mathematical model graph consisting of a set of points, or vertices, connected in pairs by lines or edges. Many networks are not homogenous, consisting of diverse clusters rather than a uniform mass of nodes [14–16]. There are many edges between the vertices within the groups but fewer edges between the groups. The structures are depicted in Fig. 1, where three communities are denoted by dotted circles, which have denser internal relationships than relationships between groups. The problem of community detection is finding communities in large networks automatically.

The contribution of this research lies in the application of the community detection method to social networks of books and borrowers in libraries. By identifying closely related communities or groups of books and borrowers, this study provides insight into the borrowing patterns of library users and can help improve library management and services. This study also evaluates the effectiveness of various community detection methods in identifying clusters and provides recommendations for the most suitable methods for segmenting social networks of books and borrowers.

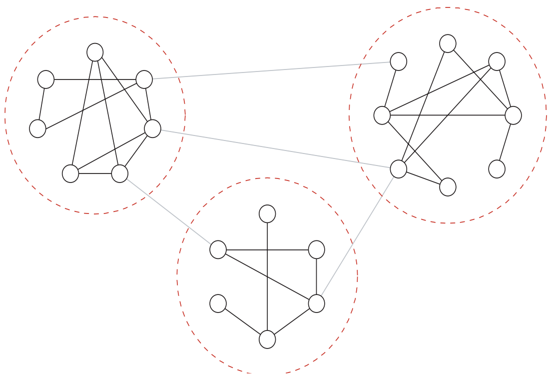


Figure 1. Community in a social network.

II. Literature Review

A. *Library's Books and Borrowers Social Network Segmentation*

Social Network Analysis (SNA) over the last decade has become a challenging problem to analyze data in various application fields such as market segmentation, crime detection, recommendation systems, co-authorship networks [5], market basket analysis [6], trending topic analysis, and many more. SNA can be composed into two areas; community detection and sentiment analysis. State-of-the-art application of SNA in libraries, proposed by the following researchers; Wu and Lee *et al.* [3] employed association rules mining and statistical circulation in discovering knowledge from library collection borrowing data. The limitation of this study is that model validation related to performance evaluation has not been carried out. Xin and Haihong *et al.* [4] proposed the community detection method using book borrowing data. They transformed the data into a reader-reader similarity graph and used them for system recommendation. The limitation of this research is that it only uses loyal readers and greedy algorithms for community detection and uses modularity as an evaluation metric. Several recent SNAs in library management, Yassine and Kadry *et al.* [7] conducted a study on the Khan repository to determine the development of the community and the performance of the online learning network by applying the Parallel Louvain and Parallel Label Propagation Algorithm (LPA) community detection algorithms. Erfanmanesh and Hoseeini [8] used SNA and scientometrics to visualize and analyze research development data on libraries and information science in various countries. Studies on the analysis of scientific publication writing in the form of the relationship between authors and co-authors with the SNA approach were carried out by several researchers. Liu and Nelson *et al.* [9] conducted a network analysis of co-authors at the ACM, IEEE, or combined ACM and IEEE digital library conferences. A weighted directed network model is introduced to represent the co-author network and author rank as indicators of each author's contribution to the network. Said and Wegman *et al.* [11] analyzed the co-authoring network of scientific publications and proposed categorizing journal writing into a solo model (without co-authors), a mentor model, an entrepreneurial model, and a team model. Zheng and Gong *et al.* [12] studied the author networks to investigate academic communities and community evolution. Choi and Yi *et al.* [13] analyzed the relationships between articles through keyword networks to predict future widespread knowledge in the management information system area. Lozano and Sabastian *et al.* [17] studied work on the keyword co-occurrences in library network analysis. Citation network analysis was conducted by studying university citation patterns for five Web Science Subject Categories with a PageRank algorithm. It combines citation and network analysis for a system recommendation for academic papers [18]. Li and Zhang [19] analyzed the bipartite network topology from the book lending database at the Shangai University library. Then, they proposed a book recommendation system using network data. Yan and Zhang *et al.* [20] analyzed the factors that influence user behavior using network analysis and data mining techniques using historical data on borrowing books at Peking University. Han and Zhang *et al.* [21] studied online reading behavior in university digital libraries through network analysis. They show that the degree distribution of the book-lending network follows an exponential distribution and follows a "small-world" phenomenon. Tunali and Tümer *et al.* [22] analyzed book lending with social network analysis and community detection, using data on book lending from the Turkish Ardahan University Library. They compiled a projected graph of books and readers, then used Louvain's algorithm to find book communities. The results were the size of the book community and the genre of books from each community. Lee [23] analyzed data on borrowing books at the university library, based on the behavior of returning borrowed books, analyzing the distribution of book returns.

This section described the books, borrowers' social network analysis, and the community detection approach.

Previous works showed the limitation of the algorithms used. Besides, many CD algorithms have yet to be applied to library books and borrowers' networks. This paper examines the performances of community detection algorithms on the book and borrowers' projected bipartite graph for community segmentation. The performance is evaluated using the metrics of modularity, performance, coverage, density, and community size.

B. Community Detection Methods

The fundamental problem in SNA is Community Detection (CD). The issue of CD is how to automatically identify the relevant group of nodes in a network. Along with developing CD applications in the real world, research to develop CD algorithms is also increasing. Several algorithms classified as unipartite graphs are the Louvain algorithm [24], Infomap [25], LPA [26], and Walktrap [27], and Spinglass [28]. The greedy algorithm proposed by Clauset and Newman *et al.* [29] is based on modularity optimization to find the best partition in the network. Each node is considered as its community using hierarchical clustering. Then, the neighboring communities are merged iteratively, where each node is moved to the community that maximizes the modularity function. These aggregated communities are combined until the obtained modularity function can no longer be increased. The computational complexity on the sparse graph is $O(N\log_2 N)$.

The Louvain algorithm is proposed by Blonde and Guillaume *et al.* [24] with a heuristic method based on modularity optimization, such as the greedy algorithm. Louvain includes a community aggregation step to improve processing on large networks. The Louvain algorithm consists of two steps. The first step is maximizing the objective function by moving nodes to the community. The second step is the community aggregation process. In this process, the communities generated from the first step are merged into supernodes. Both processes are repeated until the objective function can no longer be increased. The computational complexity of the Louvain algorithm is $O(N\log N)$. The Infomap algorithm proposed by Rosvall and Bergstrom [25] is based on the information theory principles, which find the optimal community to find the minimum information description of a random walk in the graph. The algorithm maximizes the objective function, the minimum description length. Previous studies by Zeng and Yu [30] have found Infomap's performance remains stable for networks with up to 100,000 nodes, this algorithm runs in $O(m)$.

The Walktrap algorithm proposed by Pons and Latapy [27] finds the community in the graph based on a random walk. The basic intuition of this algorithm is that random walks on a graph tend to get trapped into densely connected parts corresponding to the community. Random walk is used to calculate the distance between nodes. The nodes are then assigned into groups with smaller intra-community and larger inter-community distances through bottom-up hierarchical clustering. In the worst case, the computational complexity is $O(mn^2)$ and space $O(n^2)$. For sparse networks the computational complexity is $O(N^2 \log(N))$.

Raghavan and Albert *et al.* [26] proposed the LPA algorithm, an iterative process to find a stable community on a graph without using an objective function. LPA is based on the concept that nodes should belong to the community of most of their neighbors. Therefore, they gradually renew their membership according to their incident nodes. This method starts by assigning a unique label to each graph node. Then, iteratively simulates each node in the graph and adopts the most common label among its neighbors. The process is repeated until the label reaches the maximum occurrence among its neighbors. The computational complexity of the label propagation algorithm is $O(m)$.

Spinglass algorithm proposed by Reichardt and Bornholdt [28], adopting ideas from statistical mechanics and identifying communities using the spinglass model, with requirements: (a) for nodes in the same spin state (on the same group) reward internal edges between nodes; (b) penalize missing edges between nodes, for nodes in the different spin state (on the different group); (c) penalize existing edges between different groups; and (d) reward non-links between different groups. This method is similar to optimization based, which aims to minimize the spinglass energy with the spin state being the community index. In a sparse graph, the computational complexity of this algorithm is approximately $O(N^{3.2})$.

C. Community Detection Metrics

This section describes the metrics used to evaluate the quality of community detection. The metrics are modularity, performance score, coverage, and density are among the metrics [31, 32]. Table I depicts the notations and symbols used in the metrics.

TABLE I. NOTATIONS AND SYMBOLS OF COMMUNITY DETECTION METRICS

Terms	Symbol	Condition
the number of nodes in graph	n	
the number of edges in graph	m	
the sum of degrees of the nodes in community c	k_c	
the number of intra-community edges	E_c	$E_c = \|\{i,j\} \in E, C_i = C_j\|$
the number of inter-community edges	$E_{c'}$	$E_{c\prime} = \|\{i,j\} \in E, C_i \neq C_j\|$
the number of inter-community non-edges	E'_c	$E'_c = \|\{i,j\} \notin E, C_i \neq C_j\|$
the number of potential edges in graph	E	$E = n(n-1)/2$

1) Modularity

Modularity is a popular validation metric for measuring the strength of community structures. Good communities have a more significant number of internal edges and a smaller number of edges between communities than expected when compared to a random graph [33].

$$mod = \sum_1^n \left[\frac{E_c}{m} - \left(\frac{k_c}{2m}\right)^2\right] \tag{1}$$

The modularity value falls between 0 and 1, with 1 indicating that the partitioning process resulted in a strong community. Networks with modularity in the range of 0.3 and 0.7 indicate that they have a strong community structure. However, in some cases, the limit is not met, which can give a negative value [33]. The limitation of the modularity metric is that it may fail to detect a community smaller than a specific scale called the resolution limit. It depends on the total network measure and the community connectivity degree.

2) Coverage

The coverage of a community is the ratio of the number of intra-community edges to the total number of edges [26]. The coverage value is shown in Eq. (2),

$$cov = \frac{E_c}{m} \quad (2)$$

Coverage values range from 0 to 1. A higher coverage value means more edges within the cluster than edges connecting different clusters, which means better clustering. The coverage with value 1 means the ideal cluster structure, where all clusters are disconnected because all edges are in the cluster. Coverage is not a good quality metric for finding communities. Assigning all network nodes as one large community will result in the maximum coverage value. An additional information, such as the number of communities is needed to detect the community. Since such information does not exist about the actual community, this measure cannot be helpful for community detection.

3) Performance score

Performance score is a quality function that counts the number of pairs of vertices that are "interpreted" correctly, i.e., two vertices belonging to the same community and connected by an edge or two vertices belonging to different communities and not connected by an edge. The performance value is the ratio of intra-comm edges plus intercom non-edges divided by the total potential edges [26]. The definition of performance for the partition is in Eq. (3), where *n* is the number of nodes, and *Ci* is the community with *i* nodes in it. Performance scores range from 0 to 1; higher values indicate a community with high internal and sparse external density, so the community is better.

$$per = \frac{E_c + E'_c}{E} \quad (3)$$

The limitation of this metric in large-scale sparse networks, there is a high probability that the number of non-adjacent nodes belonging to different communities becomes very high, so the performance is biased towards high scores in the network.

4) Density

The strength of the relationship between vertices can be analyzed using the density function, which is the ratio between the number of edges represented in the community and the total number of edges in the entire graph [31]. Ratios are computed for all communities to evaluate the overall impact. The density function is defined in Eq. (3). The density value lies in the interval [0–1]; the higher the density indicates, the better the separation of the community.

$$den = \frac{1}{m}\sum_{i=1}^{k} E_c \quad (4)$$

The main drawback of this metric is that it only considers internal relationships between community nodes without regard to external ones and partitions whose communities consist of edge-connected pairs of vertices. Thus, combining it with quality metrics that consider external community relationships is necessary. In addition to a high internal density, the community must have a low external density.

5) Size of the community

The size of the community metric is used to indicate the goodness of the resulting community distribution. Wagenseller *et al.* suggested "the desirable community" term, namely the size of a community with strong ties in the range of 4–150 members [34]. The basic principle, a community that is too large, has weak connections and is therefore unstable, whereas a community that is too small has no practical value. Based on the community size distribution, we measure the percentage of users assigned to the desirable community is called the fit of community. The fit of community f_c is defined in Eq. (5), where C_d is the number of the desirable community and C is the total number of the community.

$$f_c = \frac{c_d}{c} \times 100\% \quad (5)$$

III. Materials and Methods

The research methodology diagram is shown in Fig. 2.

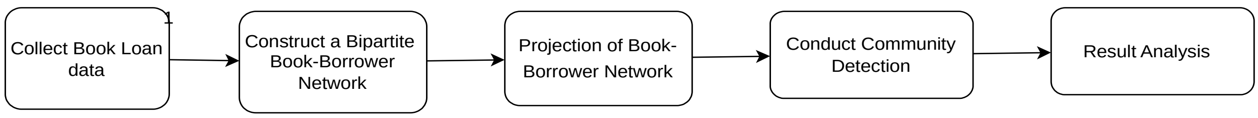


Figure 2. The research methodology.

A. Dataset and Data Representation

The books borrowing data was collected from the library of Ahmad Dahlan University, Yogyakarta, Indonesia, involving 6576 records of a historical dataset in 2018–2021. This book borrowing data has barcode attributes, book title, borrower number, borrower name, borrower major, borrowing date, and return date. From all these attributes, we choose the attributes relevant for community construction in the following: the book title's attributes, borrower number, and borrower department.

B. Projected Bipartite Graph

We construct a community of bipartite graph of the library's books borrowing transaction, where there is a book relationship with borrowers. Then we make projections the bipartite graph into book graphs and borrower graphs [35]. The bipartite graph modelling consists of two phases; first, we represent the book borrowing data into bipartite graph, and projected bipartite graph. Then we develop the community detection algorithms on the standard bipartite graph and the projected bipartite graph representation which projects the bipartite graph of the book's borrowing data into the book's graph and the borrower's graph to obtain the books and borrowers segmentation.

Many real-world issues can be modeled as a graph. In general, a graph is denoted by two tuples G(V, E), with V(G) is known as the vertex set and E(G) as the edges set. A bipartite graph is a graph where the vertices consist of two disjoint subsets where each edge does not come from the same subset. The mathematical definitions are as follows:

Definition 1 [36]: A graph G(B, S, E) is called a bipartite graph if $V(G) = B(G) \cup S(G)$ and $B(G) \cap S(G)=\phi$ and for each edge $(uv) \in E(G)$ either $u \in B(G)$, $v \in S(G)$ or $v \in B(G)$, $u \in S(G)$. G is a complete bipartite graph if $\forall u \in B(G)$ and $\forall v \in S(G)$, $(uv) \in E(G)$. A bipartite graph is generally represented by a b-adjacency matrix. For any vertex $b_i \in B$, $s_j \in S$ with $|B(G)| \times |S(G)|$ where the matrix elements are as follows:

$$B(x) = \begin{cases} 1, if\left(b_i s_j\right) \in E(G) \\ 0 \quad\quad ,\text{otherwise} \end{cases} \tag{6}$$

Because most algorithms and network analysis are designed for general graphs, one commonly used technique for bipartite graphs is to project one edge of the vertex based on connectivity with the other side. Formally, the projection from a bipartite graph to a unipartite graph can be defined as follows:

Definition 2 [37]: Let G(B,S,E) is bipartite graph. Projection of the bipartite graph G for the vertex set B to the vertex set S is to construct a unipartite or one mode network G'(B, E') where V(G)=B and $(b_i b_j) \in E(G')$ if $N(b_i) \cap N(b_j) \neq \emptyset$. The projection of a bipartite graph will always produce a pair of unipartite graphs. The set of vertices U associated with S and the set of vertices S associated with U.

An illustration of the projection of a bipartite graph into two unipartite graphs is shown in Fig. 3. The bipartite Books-borrowers graphs are projected into two unipartite graphs: the book and the borrower graphs. Two vertices in the projection graph have a relationship because they have a relationship with the same vertex in the bipartite graph. For example, vertices 1 and 2 in a book graph are connected because in a bipartite graph, vertices 1 and 2 are connected to the same vertices, namely vertices A and B.

Algorithm 1 present the construction of the projected bipartite graph from the bipartite graph. The algorithm employs exhaustive search is to search all possible pairs of vertices, whether they have at least one neighbouring node. The lines 4–6 of algorithm 1 are the exhaustive search process. The element "1" will be added to the adjacency matrix of graph G' if they have at least one neighbour in common. Line 7 checks whether it has the same neighbours; if yes, then the adjacency matrix is added with the element "1" (line 8). Table II describes the properties of both graphs.

Algorithm 1. Projected Bipartite Graph Construction

Input : Bi-Adjacency Matrix (B) of bipartite graph G
Output: Adjacency Matrix (A) of unipartite graph G'

```
1    n1 ← number of rows (B)
2    n2 ← number of columns (B)
3    CreateMatrix(A, n1, n2,0)
4    for i ← 1 to n1 do
5       for j ← i+1 to n2 do
6          for k ← i+1 to n2 do
7            if B[i][k]==1 && B[j][k]==1 then
8                   A[i][j] ← 1
9                   break
10             else
11                   A[i][j] ← 0
```

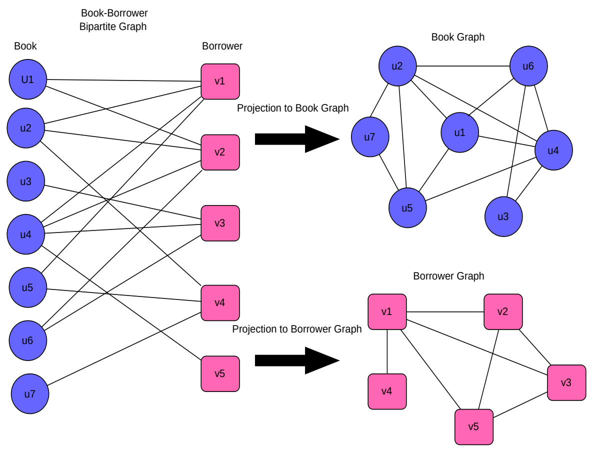


Figure 3. Projection of book-borrower graph.

C. Community Detection Methods

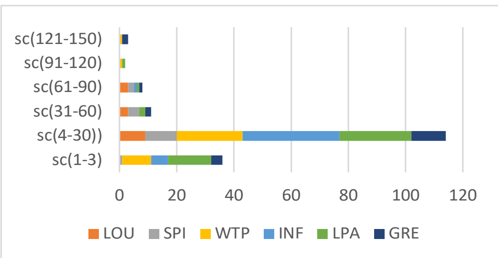
The community detection algorithms are employed on the book and borrower projected bipartite graph. Six community detection algorithms as described in Section III, such as Louvain, Infomap, LPA, Walktrap, Spinglass, and Greedy algorithms were employed. The community detection performance was evaluated on modularity, coverage, performance, density, and community size.

Mathematically, community detection can be formulated as an optimization problem, where the objective is to find a partition of the nodes into non-overlapping communities that maximizes some quality function. One of the most widely used quality functions is modularity, which measures the degree of deviation of the observed network from a null model in which the nodes are connected randomly. Mathematical concepts and notations, evaluation matrices and others have been discussed previously.

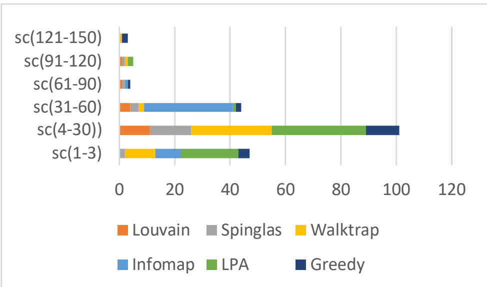
The resulting book and borrower segmentation were evaluated on the distribution of community size and its important node (have the highest degree). The book and borrower graphs test the community detection process by running a community detection algorithm that only considers the information structure. After finding the best community detection algorithm, the algorithm is used to calculate the community size distribution. The community size distribution represents the books and borrowers' segmentation in the book borrowing process.

IV. Result and Discussion

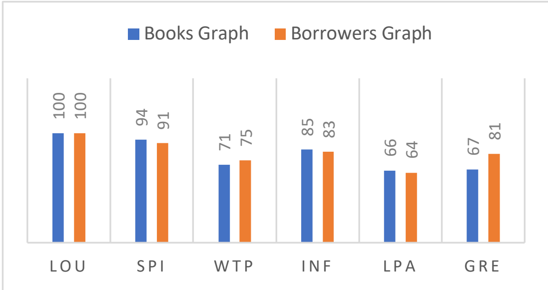
In this section, we present the performance result of community detection on the books' and borrowers' projected bipartite graphs. The Louvain, Spinglas, Walktrap, Infomap, LPA, and Greedy algorithms were employed for the books' and borrowers' projected bipartite graph. We evaluate the performance of the community detection algorithms in both books and borrowers segmentation. The results of the modularity (mod), performance (per), coverage (cov), density (den), number of community (nc), interval size (int size) are shown in Table II. Interval size (int size) is the range of community sizes obtained is from smallest to largest, a good range if it is in the range 4–150 ("the desirable community"). The number of communities (nc) in Table II and the community size (sc) in Figs. 4(a) and 4(b) can show the goodness of community distribution. The more the number of communities that have a community size in the 'desired community' range, the better the community distribution. Meanwhile, Community fit expressed the percentage of suitability of the community obtained to the expected community, the greater the value, the better the community obtained (shown in Fig. 4(c)). The number of communities (nc) in Table II and the community size (sc) in Figs. 4(a) and 4(b) can show the goodness of community distribution. The more the number of communities that have a community size in the 'desired community' range, the better the community distribution. Meanwhile, Community fit expresses the percentage of suitability of the community obtained to the expected community, the greater the value, the better the community obtained (shown in Fig. 4(c)).



(a)



(b)



(c)

Figure 4. Distribution on bipartite graph: (a) Community size (*sc*) Distribution on books graph; (b) Community Size (sc) distribution on borrower graph; (c) Community fit.

TABLE II. Projected Bipartite Graph

Type	Method	*mod*	*cov*	*per*	*den*	*nc*	*int_size*
Books Graph	Lou	0.71	0.87	0.91	0.37	15	5–84
	Spi	0.70	0.86	0.91	0.28	18	3–86
	Wal	0.68	0.89	0.87	0.65	35	2–146
	Inf	0.67	0.80	0.96	0.59	41	2–84
	LPA	0.67	0.84	0.92	0.70	44	2–94
	Gre	0.68	0.90	0.86	0.32	12	4–109
Borrowers Graph	Lou	0.70	0.89	0.91	0.36	17	5–97
	Spi	0.70	0.86	0.92	0.32	22	2–94
	Wal	0.68	0.89	0.91	0.65	44	2–122
	Inf	0.67	0.80	0.96	0.62	52	2–88
	LPA	0.65	0.86	0.93	0.73	58	2–106
	Gre	0.65	0.90	0.86	0.48	21	2–141

A. *Performances of Community Detection Algorithms on Books and Borrowers Projected Bipartite Graph*

1) *Analysis based on metrics*

Table II respectively describe the performance of the community detection algorithm on the projected bipartite on books graph and borrowers graph. All algorithms gave comparative results in modularity (mod), performance (per), and Coverage (cov)). Louvain and Spinglas obtained the highest modularity, Infomap gave the highest performance and Greedy with the highest coverage (cov) values. Due to the small size of the books' and borrowers' graphs on the modularity metric, it may result that all community algorithms have high performance and are almost as good. According to the previous modularity and coverage relationship analysis, the coverage value is greater than 0.5 and higher than the modularity in all algorithms. The table shows that all algorithms' coverage value is greater than 0.5. Also, the coverage value is higher than the modularity value in all algorithms. Because the modularity value is already high, the coverage value is very high. In terms of performance metrics, the result is a very high value of average 91% because the book and borrowers' graphs are of the sparse graph type.

In the density (den) metric, according to the analysis, density is affected by the number of communities (nc) produced. However, there is a significant difference in the results for the LPA, Walktrap, and Infomap algorithms which have high density, while Louvain, Spinglas, and Greedy are classified as low density. This difference is not only affected by the number of communities produced but also related to the community size distribution. The LPA, Walktrap, and Infomap algorithms produce communities with small sizes (1–3) high enough to contribute high-density values (shown in Figs. 4(a) and 4(b)).

The community size distribution metrics in Fig. 4(a) and Fig. 4(b) show that all algorithms produce the largest community size with nodes (4–30). The most centralized range of community distribution is obtained by Louvain, while Walktrap and Greedy obtain the widest. The desired (best) community measurement result from the community fit metric obtained by Louvain is the best with all the resulting communities according to the expected community (100%), both book and borrower graphs. The order slightly below this is Spinglas. Meanwhile, the worst is Walktrap, followed by Greedy algorithm (shown in Fig. 4(c)).

2) Analysis of the algorithms

The high value of modularity produced is because the metric is meant to maximize its value. The Louvain algorithm is better than Greedy because there is a mechanism for improving the performance of modularity in the second stage, namely the formation of an aggregation community. The Spinglas algorithm uses the based spins applying the optimization to minimize Spinglas in forming communities, thus obtaining good results. Meanwhile, the Walktrap and the Infomaps are based on random walks, and the LPA is based on similarity nodes. The detection results show that the number of communities (nc) and the density is high, whereas intuitively because the book and borrower graph are sparse, the density should also be low. This scenario is justified because the resulting communities are too small and do not meet the desired community. For example, LPA only gives 64% of the resulting communities that match the desirable communities. The LPA is the worst algorithm in modularity and community suitability. It could be due to the simple algorithm mechanism that relies on the similarity of neighbouring nodes regardless of modularity. Even though it is fast, the results are unstable.

The experimental results show that modularity (mod) and community size (sc) are dominant evaluation metrics. Our experiments prove that modularity is indeed prevalent for measuring quality. Our experimental results reinforce the findings studied Wickramasingh [38], and Yang [39] with research that the Louvain and Spinglass algorithms have the best performance for small graph sizes. In addition to the modularity metric, the community size metric also plays a role in being considered because it can indicate the good distribution of community sizes obtained. Interval size (int_size) is useful for knowing a good interval of community size (community is of good size if 4< int_size <150). In this study, referring to the value of modularity, interval size and community size distribution, the Louvain algorithm has the best performance.

B. Analysis of Books and Borrowers Segmentation

In the book graph, community detection is done using the Louvain algorithm because it has the best modularity value. It can be seen in Fig. 5 that there are 15 communities (segments) of books that the borrower borrows together. The largest segment comprises 83 members (17.9%) of the community. This segment's most popular book (most important node) is Qualitative Research Methodology, with 67 degrees. The top five most popular books are shown in Table III.

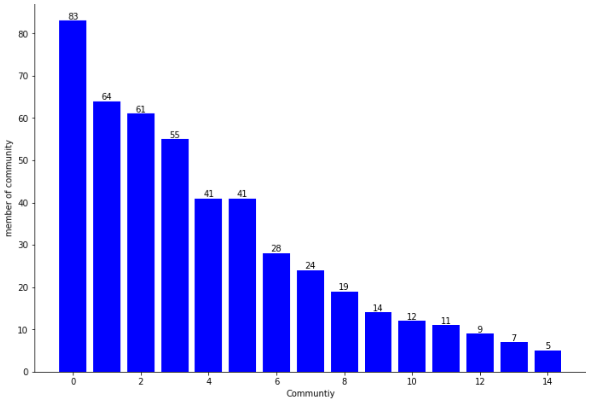


Figure 5. Segmentation of books community.

TABLE III. TOP FIVE COMMUNITY SIZE AND POPULAR BOOKS

Community Size	Books Title	Degree
85	Qualitative Research Methodology	67
62	Public Health Science	64
54	Indonesian Herbal Pharmacopoeia	61
48	Educational Research Methods Quantitative, Qualitative Approach	54
41	Health Promotion Theory and Application	44

In the borrower's graph, the community detection is done using the Louvain algorithm. The result is that the segmentation of borrowers graph and the influential nodes of each community are shown in Fig. 6 and Table IV. From these results, it was found that there were 16 community (segments) of borrowers who borrowed the same book, with one prominent community that was large, with 97 members (17.5%) with the most active borrowers borrowing books with degree 58 from the Pharmacy department. Interestingly, it was also found that borrowing communities from the Pharmacy department for different book borrowers were in the top three. The top five most popular books are shown in Table IV.

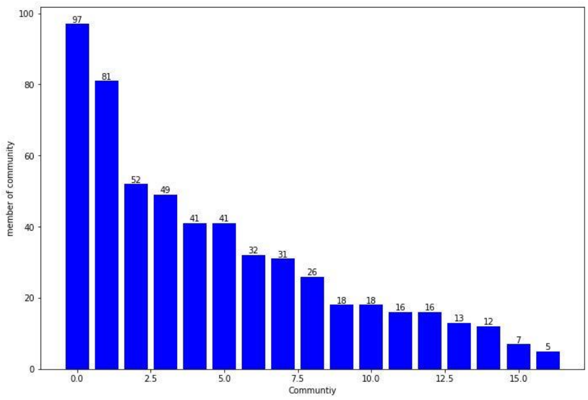


Figure 6. Segmentation of borrower's community.

TABLE IV. TOP FIVE COMMUNITY SIZE AND BORROWER ACTIVE

Community Size	Borrowers No	Department	Degree
88	225	Pharmacy	58
52	739	Public Health Science	53
26	239	Pharmacy	49
40	18	Informatic	49
33	424	Biology Education	45

V. CONCLUSION

The experimental results on the book graph and borrowers graph show that in four metrics, there is a trade-off in determining the best community detection algorithm. Louvain algorithm and Spinglass gives the best in modularity, the greedy algorithm in coverage, Infomap in performance, and LPA in density. By referring to the algorithm with the best modularity and size community distribution, in the book segmentation, 15 communities were generated, and the Qualitative Research Methodology book is the most popular. In comparison, there were 16 communities for the borrowers, and the borrowers from the Pharmacy department were the most diligent in borrowing books. Analyzing the process of borrowing library books can provide significant insights for library management in improving service quality and user satisfaction. For example, the policy of procuring printed books per title can be optimized depending on the borrower's expectations. In addition, library managers can recommend books to borrowers based on identified borrower segments from the same book borrowers. Likewise, borrower segmentation information can be used to recommend appropriate promotion strategies. For example, for a community (department) that rarely borrows books, the library manager is more active in book promotion activities.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Azuraliza Abu Bakar: conceptualization, methodology, supervision, project administration, funding acquisition, writing—reviewing and editing. Tedy Setiadi: conceptualization, methodology, investigation, data curation, writing—original draft. Mohd Ridzwan Yaakub: supervision, resources, validation, writing—reviewing and editing. All authors had approved the final version.

FUNDINGS

The research is supported by Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme FRGS/1/2020/ICT02/UKM/01/2.

REFERENCES

[1] K. Barner, "The library is a growing organism: Ranganathan's fifth law of library science and the academic library in the digital era," *Libr. Philos. Pract.*, 2011.

[2] L. Lembinen, "Academic library directors' strategic decision-making process," *Lib. Q.*, vol. 28, no. 1, pp. 1–21, 2018.

[3] C. H. Wu, T. Z. Lee, and S. C. Kao, "Knowledge discovery applied to material acquisitions for libraries," *Inf. Process. Manag.*, vol. 40, no. 4, pp. 709–725, 2004.

[4] L. Xin, E. Haihong, and S. Junde, "Community detection based on readers' borrowing records," in *Proc. 2013 IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput. GreenCom-iThings*, 2013, pp. 1001–1005.

[5] T. T. Aung and T. T. S. Nyunt, "Community detection in scientific co-authorship networks using Neo4j," in *Proc. 2020 IEEE Conf. Comput. Appl.*, 2020.

[6] T. Setiadi, M. R. Yaakub, and A. A. Bakar, "Evaluation of community detection algorithms for artificial and real networks," in *E-Proc. Seminar on Information Retrieval and Knowledge Management*, 2022.

[7] S. Yassine, S. Kadry, and M. A. Sicilia, "Application of community detection algorithms on learning networks. The case of Khan Academy repository," *Comput. Appl. Eng. Educ.*, 2020.

[8] M. Erfanmanesh and E. Hosseini, "Using social network analysis method to visualize library & information science research," *J. Adv. Inf. Technol.*, vol. 7, no. 3, pp. 177–181, 2016.

[9] X. Liu, J. Bollen, M. L. Nelson, and H. Sompel, "Co-authorship networks in the digital library research community," *Inf. Process. Manag.*, vol. 41, no. 6, pp. 1462–1480, 2005.

[10] A. Mahmoudi, A. A. Bakar, M. Sookhak, and M. R. Yaakub, "A temporal user attribute-based algorithm to detect communities in online social networks," *IEEE Access*, vol. 8, pp. 154363–154381, 2020. doi: 10.1109/ACCESS.2020.3018941

[11] Y. H. Said, E. J. Wegman, W. K. Sharabati, and J. T. Rigsby, "Social networks of author-coauthor relationships," *Comput. Stat. Data Anal.*, vol. 52, no. 4, pp. 2177–2184, 2008.

[12] J. Zheng, J. Gong, R. Li, K. Hu, H. Wu, and S. Yang, "Community evolution analysis based on co-author network: A case study of academic communities of the journal of 'Annals of the Association of American Geographers'," *Scientometrics*, vol. 113, no. 2, pp. 845–865, 2017

[13] J. Choi, S. Yi, and K. C. Lee, "Analysis of keyword networks in MIS research and implications for predicting knowledge evolution," *Inf. Manag.*, vol. 48, no. 8, pp. 371–381, 2011.

[14] A. Mahmoudi, M. R. Yaakub, and A. A. Bakar, "New time-based model to identify the influential users in online social networks," *Data Technologies and Applications*, vol. 52, no. 2, pp. 278–290, 2018. https://doi.org/10.1108/DTA-08-2017-0056

[15] A. Mahmoudi, M. R. Yaakub, and A. A. Bakar, "A new method to discretize time to identify the milestones of online social networks," *Soc. Netw. Anal. Min.*, vol. 8, no. 1, pp. 1–20, 2018.

[16] A. Mahmoudi, M. R. Yaakub, and A. A. B. U. Bakar, "The Relationship between online social network ties," *ACM Transactions on Knowledge Discovery from Data*, vol. 13, no. 3, 2019.

[17] L. Sebastián *et al.*, "Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature," *Scientometrics*, vol. 120, pp. 609–629, 2019.

[18] F. A. Massucci and D. Docampo, "Measuring the academic reputation through citation networks via PageRank," *J. Informetr.*, vol. 13, no. 1, pp. 185–201, 2019.

[19] N. N. Li and N. Zhang, "The bipartite network study of the library book lending system," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng.*, vol. 4, part 1, pp. 773–782, 2009.

[20] F. Yan, M. Zhang, J. Tang, T. Sun, Z. Deng, and L. Xiao, "Users' book-loan behaviors analysis and knowledge dependency mining," *Lect. Notes Comput. Sci.*, vol. 6184, pp. 206–217, 2010.

[21] L. Han, G. Zhang, B. Yong, Q. He, F. Feng, and Q. Zhou, "Statistical study of characteristics of online reading behavior networks in university digital library," *World Wide Web*, vol. 22, no. 3, pp. 1175–1187, 2019.

[22] V. Tunali and C. Tümer, "Analysis of book-borrowing network using complex network analysis," *BEU Journal of Science*, vol. 10, no. 1, pp. 133–142, 2021.

[23] T. H. Lee and J. W. Lee, "Self-organized human behavioral patterns in book loans from a library," *Phys. A Stat. Mech. its Appl.*, vol. 563, 125473, 2021.

[24] V. D. Blondel, J. Guillaume, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, pp. 1–12, 2008.

[25] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," in *Proc. Natl. Acad. Sci. U. S. A*., 2008, vol. 105, no. 4, pp. 1118–1123.
[26] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E-Stat. Nonlinear, Soft Matter Phys*., vol. 76, no. 3, pp. 1–11, 2007.
[27] P. Pons and M. Latapy, "Computing communities in large networks using random walks," *J. Graph Algorithms Appl.*, vol. 10, no. 2, pp. 191–218, 2006
[28] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 74, no. 1, pp. 1–14, 2006.
[29] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E-Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 70, no. 6, p. 6, 2004.
[30] J. Zeng and H. Yu, "Effectively unified optimization for large-scale graph community detection," in *Proc. 2019 IEEE International Conference on Big Data*, *Big Data*, 2019, pp. 475–482. https://doi.org/10.1109/BigData47090.2019.9005481
[31] H. S. Pattanayak, H. K. Verma, and A. L. Sangal, "Community detection metrics and algorithms in social networks," in *Proc. ICSCCC 2018—1st Int. Conf. Secur. Cyber Comput. Commun.*, 2018, pp. 483–489.
[32] T. Chakraborty, A. Dalmia, A. Mukherjee, and N. Ganguly, "Metrics for community analysis: A survey," *ACM Comput. Surv*., vol. 50, no. 4, 2017.
[33] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E-Stat. Nonlinear, Soft Matter Phys.*, vol. 69, no. 22, pp. 1–15, 2004
[34] P. Wagenseller, F. Wang, and W. Wu, "Size matters: A comparative analysis of community detection algorithms," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 4, pp. 951–960, 2018. https://doi.org/10.1109/TCSS.2018.2875626
[35] R. Arthur, "Modularity and projection of bipartite networks," *Phys. A Stat. Mech. its Appl*., vol. 549, 124341, 2020.
[36] A. Abiad, C. Dalfo, and M. A. Fiol, "Algebraic characterizations of regularity properties in bipartite graphs," *European Journal of Combinatorics*, vol. 34, no. 8, pp. 1223–1231, 2013
[37] H. Gmati, A. Mouakher, and I. Hilali-Jaghdam, "BI-COMDET: Community detection in bipartite networks," *Procedia Comput. Sci.*, vol. 159, pp. 313–322, 2019.
[38] A. Wickramasinghe and S. Muthukumarana, "Assessing the impact of the density and sparsity of the network on community detection using a Gaussian mixture random partition graph generator," *Int. J. Inf. Technol.*, vol. 14, no. 2, pp. 607–618, 2022.
[39] Z. Yang, R. Algesheimer, and C. J. Tessone, "A comparative analysis of community detection algorithms on artificial networks," *Sci. Rep.*, vol. 6, no. 128, pp. 1–18, 2016.

Copyright © 2023 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.