IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 2 March 2023, accepted 21 March 2023, date of publication 24 March 2023, date of current version 29 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3261331

RESEARCH ARTICLE

Identifying Firm Significance and Positions in the Patent Innovation Based on Centrality Measures' Clustering Approach

PRIYANKA C. BHATT AND TZU-CHUEN LU
Department of Information Management, Chaoyang University of Technology, Taichung 413, Taiwan
Corresponding author: Tzu-Chuen Lu (tclu@cyut.edu.tw)

This work was supported in part by the Ministry of Science and Technology (MOST), Taiwan, Republic of China, under Grant MOST 109-2221-E-324-025-MY3.

ABSTRACT Organizations strive to achieve technological competence in the current era of inevitable technological progress. One way to measure the adaptability of firms to huge technological shifts is through various parameters, including patenting activities. This study presents a method for identifying the significance of firms in an innovation network using patent citation analysis and centrality measures. Specifically, the study employs k-means clustering to classify firms into similar clusters based on network-based centrality measures such as betweenness, closeness, and eigenvector centrality. The study then develops a cluster relational network by establishing a cluster adjacency network and identifying firm positions within and between clusters. By examining the relationship between clusters, the cluster network identifies the significance of firms. The study identifies four positions, namely, leader, follower, knowledge inertia, and significantly emerging, that align with the status of firms in patenting innovation capability. The method is implemented using blockchain technology as a case study. The novelty of the study lies in the structured approach to identifying firm significance by adding another layer of adjacency network to existing patent citation analysis techniques.

INDEX TERMS Patent analysis, innovation assessment, k-means clustering, patent centrality analysis, social network analysis.

I. INTRODUCTION

Innovation and technological competence are indispensable for a firm to obtain market dominance and operational excellence. Firms invest more in technological innovation, research, and development to develop novel consumer solutions and services [1]. A firm's innovation capability can be measured through patenting activity. A direct correlation exists [2] between patenting activity and a firm's performance. Firms can secure and assess their research and development outcomes through their patent activity. Patent analysis not only aids organizations in assessing their technological strategies but also monitors the competitor's technological endeavors. Patents represent technological novelty, firm innovation capability, and countries' innovation development index.

Patent data can be analyzed to identify multiple dimensions of an innovation landscape. By utilizing patent analysis, previous studies have identified trends and trajectories of evolving and established technologies [3], [4], technology convergence and divergence, and firm innovation capabilities. Patent citation analysis is employed to assess the knowledge flow [5], [6] and knowledge providers and absorbers [7] across the technology trajectory. Patent citation analysis allows for a multi-dimensional examination of the innovation landscape. The network relationships can be obtained based on the citation analysis. Citation relationships can be obtained for the patent documents and associated organizations, technology domains, and patent families. It is possible to identify the central nodes, i.e., the actors with a significant place in the network, through citation relationships.

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

Distinctive characteristics of a node's importance can be identified by employing different centrality measures defining distance or inflow-outflow-based dimensions. Centrality analysis can also identify the correlations across subnetworks in a network community. Various centrality measures include degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and PageRank centrality. Previous studies have employed centrality measures to identify significant patents in the technology domain, identifying convergence patterns [8], knowledge flow [9], the core-periphery structure of the network [10], knowledge positions in the market domain, and so on. The network analysis for patent positions is often restrictive in identifying firms' core competence or position in the market. On the other hand, the technological network developed via the examination of patents is not adequate for firms to acquire knowledge regarding their competitive strategy. Previous researchers have studied the roles and positions perspective using patent analysis, specifically, main path analysis. However, limited research has focused on identifying the prominence of firms in the technology innovation landscape. Moreover, it is feasible to identify firms with similar characteristics concerning patent innovations and their citation relationships. The study consequently tries to address the following research questions:

RQ1: Do firms that belong to a particular innovation cluster have unique features that influence the whole patent innovation landscape of a technology domain?

RQ2: What relationship do firms have among and across the innovation clusters in the different positions?

This study addresses the research questions by proposing a method for identifying the significance of firms' innovations and their respective positions in the citation network. The method extends from previous research by utilizing cluster analysis based on centrality measures of firm citations. In addition, the study adds another dimension to the citation network analysis by proposing a cluster adjacency network for firms' innovation and developing significance indicators for firms' innovation based on cluster citations.

The paper is further classified into five sections. The following section discusses the background studies for the method adopted, followed by the research methodology in section three. Results are described in section four, and the study is concluded in section five.

II. RELATED WORK

Patents, in general, are the primary source of invention and technology evidence. However, when patents cite each other, the citations form a hierarchy across that invention path. Patents and their citations become a significant factor in the technology innovation network and their relevance in the ecosystem as a whole [11]. When a patent cites a previous patent(s), it is called a backward citation, and when that patent is cited by future patent(s), those citations are called forward citations. More backward citations indicate the dependence of an invention on previous technologies/knowledge, and more forward citations indicate the impact of the invention on its successors [12]. Inventors typically file patent applications, often associated with assignees such as firms, organizations, and institutions. Previous studies have examined various aspects of knowledge flow in research and development, including technology knowledge flow across different organizations, innovation parameters across organizations, highly central organizations in the innovation landscape, and more [13], [14], [15], [16], [17], [18] by analyzing the patent data and citations. Organizations are investing more in their research and development [19], and the output is seen through their patenting activities and the commercial product or service thus produced. Organizations aim to identify promising technologies and their future to create technological strategies [20]. Building successful technological capabilities to have a niche in the market has been gripped by incumbents and startups [21]. Previous research [22], [23] has shown that most citations flow between a few firms in the network, and these significant firms are likely to drive activity in the market. The previous research formulated the multivariate analysis model to develop the organization innovation relationship, briefed in Table 1.

TABLE 1. Research gap and novelty.

Previous Research	Method Used	Gaps	This study
[23], [24]	Multivariate analysis of variance	The choice of centrality measures, both degree and path-based measures are chosen for the clustering	Path based centrality measures are chosen to identify the network significance instead of node significance
[25], [26]	Technology knowledge status and Technology knowledge redundancy	The patent affiliation relationship and firm citation relationship does not identify the exact correlation between the firm's technology flow	The technology innovation cluster is obtained based on the firm relation network created from patent affiliation network.

Previous studies have used all centrality measures to identify their effect on each other. Furthermore, the centrality measures are either degree-based or path based. Degree-based centrality measures focus on node significance in the whole network; however, path-based centrality measures signify the influence of associated nodes and the sub-network. This study focused on path-based centrality measures to identify the firm network significance. Additionally, previous studies utilized the patent affiliation and firm citation relationships to obtain the technology knowledge status and redundancy across the network. The affiliation matrix is created with additional values of firm relationships. Therefore, this study leverages the patent affiliation to create the firm adjacency matrix, which is further utilized to obtain the firm clusters and innovation

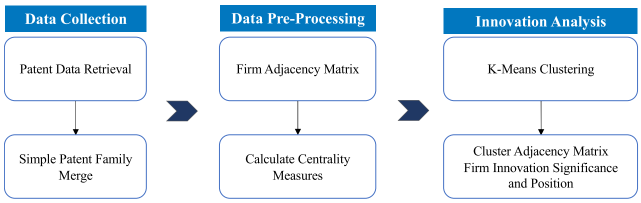


FIGURE 1. Research process.

significance analysis. There has been a lack of research utilizing the method employed in this study. Previous research has focused on patent or firm citations and has progressed to perform trajectory analysis for technology evolution. Centrality measures have been employed to obtain important patents or significant firms in the trajectory. However, this study performs clustering based on centrality measures and identifies groups of firms with similar characteristics and their influence across the innovation network.

III. RESEARCH PROCESS

The research process is apportioned into three phases; the first phase includes data collection from the Webpat database using MTrends software. The data is extracted for US patents, followed by a simple family merge to avoid duplicity of the data. The second phase involves data pre-processing, where this study develops a firm citation dataset, also known as firm adjacency matrix. Furthermore, centrality measures are obtained for the extracted citation dataset. Finally, innovation analysis is performed in phase 3, where k-means clustering is applied to classify firms into similar clusters. A cluster adjacency matrix is created to find firm innovation significance. The research process is shown in Fig. 1, and the process is detailed further in this section.

A. DATA COLLECTION

1) PATENT DATA RETRIEVAL

The data is retrieved from the Webpat database using MTrends patent analysis software based on keywords and the selected technology's International Patent Classification (IPC) codes. The keywords and IPC codes are extracted based on previous studies [26], [27], [28], [29] and patent reports. Previous studies have used single keywords, such as Blockchain [26], for identifying the trends; however, if we must look past and find the technology's path dependency, we must also understand the underlying technology terms. The description of retrieved data is shown in Table 2. The data retrieved includes details such as Patent number, application date, publication date, assignee, citations, assignee country.

2) SIMPLE PATENT FAMILY MERGE

This study, after receiving the entire patent dataset, uses simple patent family merge to eliminate duplicate entries under different filings for the same innovation. Simple patent families are groups of patent documents that typically refer to the same invention [30], [31]. The technical content of such patents is generally identical because same inventors

TABLE 2. Data description.

Patent Data	Description
Patent Number (P.No.)	Number given to a patent after it is filed, such as US10853341, where the first two digits are the country code, and the next group of numbers is the patent's application number
Application Date	The date when patent was applied
Assignee Country	The country of the organization/institution associated with the patent
Assignee	Organization/Institution name of the author(s) of the patent document
Citations	Citations obtained by the patent from other patents

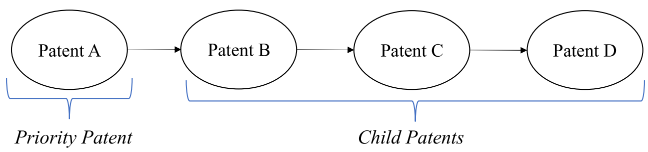


FIGURE 2. A simple patent family structure.

file multiple region/office applications for patents worldwide to secure their invention. The patent documents in a patent family usually have the same priority patent. In these applications, the original application is typically referred to as the priority application in the patent family. The simple patent family structure, depicted in Fig. 2, indicates that a priority patent A is followed by subsequent descendant patents B, C, and D.

Patents obtained after the simple family merge are used for the further analysis.

B. DATA PRE-PROCESSING

1) FIRM ADJACENCY NETWORK

A particular patent entry obtained has an assignee entity associated with it. Assignee is the term given to firms or institutions filing patents. The premise is that each patent is assigned one assignee, whereas one assignee may have more than one patent affiliated with it. Patent citation relationship can be further leveraged to obtain firm citation relationship since each firm has at least one patent affiliated to it in the dataset. It is also possible to affiliate the firm citations from patent citation data, i.e., when one patent cites another patent, it also indicates the citation flow across the firms. If C denotes the firm adjacency matrix and that each cell in the matrix at the position, i-th row and j-th column, is denoted by $f_{i,j}$, for all $1 \leq i, j \leq m$, then the notation of firm adjacency matrix is denoted in Eq. 1.

$$
\begin{aligned}
\mathrm{C} := & \left(f_{i,j}\right)_{m\times m} \\
& \{where \\
& \qquad f_{i,j} = \sum P_{i,j} \\
& \qquad P_{i,j} = 1,\ if\ patent(s) affiliated with firm i cite \\
& \qquad\qquad patent(s) affiliated with firm j, \\
& \qquad else\ P_{i,j} = 0;
\end{aligned}
$$

$$i, j \leq m, mis the total number of distinct firms$$
$$with citation relationship$$
$$\} \tag{1}$$

For example,

$$C = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,m} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,m} \\ \cdots & \cdots & \cdots & \cdots \\ f_{m,1} & f_{m,2} & \cdots & f_{m,m} \end{bmatrix}.$$

The rows in the matrix indicate the firms citing the patents from firms in the columns, i.e., $f_{1,2}$ indicates the total number of citations from patents affiliated with Firm 1 citing patent(s) affiliated with Firm 2. for instance, if Firm 1 has one patent, which cites two patents from Firm 2, then $f_{1,2} = 2$. the citation relationship across the firms can establish the focal firms based on the adjacency network.

2) CALCULATE CENTRALITY MEASURES

The effective evolution of any technology results from radical innovations cumulatively coming together one after another [32], which becomes significant enough in establishing a particular technology. Technological evolution can be measured in the context of technological trajectories, which may or may not be linear. Centrality analysis is considered an essential method in social network analysis, representing the significance of a node in a trajectory. Depending on the application and viewpoint, what counts as central or significant may differ in context. Consequently, there are various ways to identify centralities in different contexts. This study considers three types of centralities: betweenness centrality, closeness centrality, and eigenvector centrality. One or more of these measures can be analyzed in network analysis to gain a better perspective on the network.

3) BETWEENNESS CENTRALITY

The number of times a node is on the shortest path between other nodes is measured by betweenness centrality. This metric identifies which nodes in a network act as bridges between other nodes. Betweenness can be used to examine communication dynamics. A high betweenness count could indicate that nodes in a network have authority over different clusters [33] or are on both clusters' peripheries. The likelihood that a node must mediate a link between two other nodes that are not directly involved is referred as the degree of betweenness. The formula to calculate betweenness centrality for a node is shown in Eq. 2.

$$C_{bw}(x) = (\sum_{i \neq x \neq j}^{n} \frac{\sigma_x(i,j)}{\sigma(i,j)}) \times \frac{(n-1)(n-2)}{2} \tag{2}$$

where, $\sigma_x(i,j)$ is the shortest path between nodes i and j that pass-through node x, and $\sigma(i,j)$ indicates all the shortest paths between nodes i and j. Thus, betweenness centrality of node x can be described as the ratio of the shortest paths that go through node x by all the shortest paths between every two nodes i and j. Further the value is normalized with number of node pairs (excluding the current node). Betweenness centrality indicates the probability of the passing of information from node i to j will go through x. Entities or nodes having high betweenness centrality are considered significant controllers of power of information in the network. For instance, if there were five shortest paths between a pair of nodes, and three of them went through node k, then the fraction of the node k would be $3/5 = 0.6$. The process is repeated for every pair of nodes in the network. The fractions are thus added up to obtain the betweenness centrality for the node.

4) CLOSENESS CENTRALITY

Closeness centrality is the measure that scores each node based on its closeness to all other nodes in the network. It stands for the convenience and ease of connections between the focused node and the rest of the nodes in the network [22]. Closeness centrality is determined by calculating the average inverse distance (distance) between a node and every other node in a network. Consequently, it has been commonly operationalized as a measure of reachability and proximity to other collaborators (nodes in a network) and attributed to facilitating ease of access and interactions with others, immediacy in acquiring information [34], degree of dependence of the learner on others, the potential for control over sharing of resources and communication in a network, and awareness of opportunities. The formula to calculate closeness centrality for a node is shown in Eq. 3.

$$C_{cl}(x) = \frac{n}{\sum_{j=1}^{n} \rho(x,j)} \tag{3}$$

where, $\rho(x,j)$ is the distance between nodes x and j, and n is the total number of nodes in network. The closeness centrality measure of x is the ratio of network size and the summation of the distance of x with every other node.

5) EIGENVECTOR CENTRALITY

Eigenvector centrality measures a node's importance based on the total centralities of its neighbors [6]. Eigenvector centrality has been utilized to measure social capital, ego network strength, and robust network connections. The fundamental premise is that links from important nodes are valued more than unimportant ones [35] and will have higher eigenvector scores. Initially, all nodes are treated equally, but nodes with more edges acquire prominence as the calculation advances. Their significance spreads to the nodes with which they are connected. The equation for eigenvector centrality is shown in Eq. 4, where λ is an eigenvalue of the adjacency matrix.

$$C_{egn}(x) = \left(\frac{1}{\lambda}\right) \times (\sum_{i \neq j}^{n} P(i,j)\, x) \tag{4}$$

A simple example is illustrated for the three centrality measures in Table 3 for a network shown in Fig. 3. to calculate betweenness centrality of each node (vertex) of the network, based on Eq. 2, there is only one path between any pair of nodes. for calculation we need to

TABLE 3. Centrality measure example.

Node	Betweenness	Closeness	Eigenvector
1	1	1.25	0.707
2	0	0.714	0.408
3	0	0.714	0.408
4	0	0.714	0.408
5	0	0.714	0.408

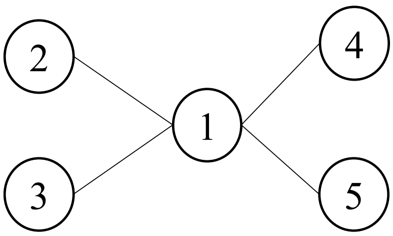


FIGURE 3. An example network to explain centrality measures.

consider all pairs of nodes from {2, 3, 4, 5}, which is 6 {(2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5)}. furthermore, only one path exists between each pair that passes through node 1. Thus, the betweenness centrality for node 1 is $(\sum_{i\neq x\neq j}^{n} \frac{\sum_x (i,j)}{\sum (i,j)}) \times \frac{(n-1)(n-2)}{2} = \left(\frac{1}{6}\right) \times \frac{(5-1)(5-2)}{2} = 1$, and for other nodes no paths pass through the nodes, therefore the betweenness for other nodes is 0. to calculate the closeness centrality, the distance of node 1 from every other node is 1, therefore the closeness centrality for node 1 is $\frac{5}{4} = 1.25$. Further, for every other node, the distance of nodes from every other node is 7, therefore the closeness centrality is $\frac{5}{7} = 0.714$. Eigenvector centrality considers more steps for the calculation. First, each node's centrality is proportional to the centrality values of its neighbors. the centrality value sum is the most elementary function. A scaling factor λ is employed to permit more generalized solutions. Therefore, if x is the vector for each node, and a is the adjacency network for the graph, then $x = \left(\frac{1}{\lambda}\right)(Ax)$. for the adjacency matrix of the graph, the equation becomes: $\lambda x = Ax$. therefore, for the example, the Eq. 4 becomes, $\lambda 4 - 3\lambda^2 = 0$, and eigenvalues become, $-\sqrt{3}, 0, 0, and \sqrt{3}$. therefore, computing the eigenvector centrality based on Eq. 4 is given in Table 3. centrality for the source (beginning) and sink (end) nodes will be similar since they won't have incoming or outgoing links from the nodes.

C. INNOVATION ANALYSIS

To identify the innovation significance and positions of the firms in the patent dataset, this study first utilizes *k-means* clustering method to classify the dataset into similar clusters based on the centrality measures identified in the previous step. Further cluster adjacency network is developed to calculate the technology knowledge flow between the firms in identified innovation clusters to finally obtain the innovation significance and position of the firms.

1) K-MEANS CLUSTERING

K-means clustering is a data mining method employed to identify a group of objects in a dataset with similar properties [36]. The clusters are identified after calculating the similarities (the distance between objects in each cluster to the centroid), where each centroid has an average cluster value. Higher similarity indicates closer distance and vice versa. The euclidean distance calculation is shown in Eq. 5.

$$\mu\left(x_j, c_j\right) = \sqrt{\sum_{j=1}^{n} (x_j - c_j)^2} \tag{5}$$

K-means clustering process first defines the cluster numbers, k, from the dataset. Secondly, k points are chosen from the dataset as centroid and set the identified datapoints to the nearest cluster centroid and then calculate the centroid again for the new formed clusters. The steps are repeated unless the centroid remains unchanged. The cluster validation process to identify the number of appropriate clusters is applied on the dataset [37]. The average distance to the centroid is then plotted to determine the k, the method is called elbow method, where the value of k is chosen after which the sum of the squared distance, within cluster sum of squares (WCSS) between each data point of the cluster and its centroid seems to be decreasing.

Previous research [1] has used k-means clustering to group similar data points, particularly in unsupervised datasets. K-means is an appropriate choice when the study aims to explore patterns in a dataset or identify groups of similar observations. Patent-based research has utilized k-means clustering to identify similar groups of patents and patent portfolios [2]. This study builds on previous research by employing k-means clustering to identify similar firm innovation characteristics in the patent landscape based on centrality measures. Since centrality measures are path-based characteristics of the network, grouping nodes or entities in the network with similar characteristics is possible. A citation-based analysis is limited to identifying the main technology path, while clustering can indicate the significance and relationships among entities within the citation networks.

2) CLUSTER ADJACENCY MATRIX

The firm cluster adjacency matrix is developed based on the clusters obtained after applying k-means algorithm in the previous step. The cluster adjacency matrix indicates the citation relationship between the corresponding firms. If total clusters obtained after k-means is N, and C_{adj} indicates $N \times N$ square cluster adjacency matrix, then each element in the matrix at the position, *i-th* row and *j-th* column, is denoted by $c_{i,j}$, which indicates the total number of citations from firms in cluster i to firms in cluster j. The citations of firms are obtained from firm adjacency matrix formulated in Eq. 1. The cluster adjacency matrix equation is shown mathematically in Eq. 6, and Table 4, where the firm citation relationship between each cluster is signified.

$$C_{adj} := (c_{i,j})_{N\times N}$$

TABLE 4. Cluster adjacency network.

Cluster	C_1	C_2	…	C_N
C_1	$c_{1,1}$	$c_{1,2}$	…	$c_{1,N}$
C_2	$c_{2,1}$	$c_{2,2}$	…	$c_{2,N}$
…	…	…	…	…
C_N	$c_{N,1}$	$c_{N,2}$	…	$c_{N,N}$

$$where,\ c_{i,j} = \sum\nolimits_{i,j\leq N} (f_{i,j})$$
$$N = total\ number\ of\ distinct\ clusters \quad (6)$$

The cell value $c_{1,2}$ in table 4 indicates the total number of citations from firms belonging to cluster 1 to the firms belonging to cluster 2.

3) FIRM INNOVATION SIGNIFICANCE NETWORK

Previous studies have utilized technology role and position analysis using frequencies between and within innovation clusters and analyzing technology knowledge status and technology knowledge redundancy based on the patent and firm affiliation network [25], [26]. This study leverages the firm citation relationship to identify the significant firm clusters based on centrality measures. The resulting firm innovation significance network is developed from the cluster adjacency matrix. First, a knowledge flow network utilizing the cluster adjacency matrix is developed. If, S indicates the knowledge flow matrix, where each cell in *i-th* row and *j-th* column is represented by $t_{i,j}$, where diagonal values are represented by knowledge flow within the same clusters, i.e., where $i = j$, and knowledge flow between the different clusters is represented by the rest of the values, the matrix equation is shown in Eq. 7, where N is the total number of clusters, and k represents the total number of firms in each cluster.

$$S := (t_{i,j})_{N\times N} = \left\{ \begin{array}{ll} \sum_{i,j=1}^{N} \frac{C_{\text{adj}}(i,j)}{k_i\times(k_i-1)}, & if\ i=j \\ \sum_{i,j=1}^{N} \frac{C_{\text{adj}}(i,j)}{k_i\times k_j}, & if\ i\neq j \end{array} \right\} \quad (7)$$

The knowledge flow within firm clusters, in Eq.7, for example when $i = j$, is calculated by the ratio of the number of citations between firms in the same cluster by the normalized total number of firms in the corresponding cluster. While the knowledge status between different firm clusters, is calculated by the ratio of the total number of citations from firms in cluster i to cluster j by the product of total number of firms in cluster i and cluster j.

$$I := \begin{cases} 1, & if\ S(i,j) \geq \tau, \\ 0, & if\ S(i,j) < \tau. \end{cases}$$
$$where,\ \tau = \overline{C}\ and\ i,j \leq N \quad (8)$$

Once the knowledge flow network in Eq. 7 is obtained from the cluster adjacency network, the knowledge flow matrix, S, is converted into the innovation significance matrix, I, by setting a threshold value, τ, equal to the average of the total citation counts of the firms from the firm adjacent matrix. The values in the knowledge flow network less than the threshold are replaced with 0 and those equal or greater than threshold are replaced with 1, revealing the final firm innovation significance relationship network.

IV. EMPIRICAL ANALYSIS: CASE OF BLOCKCHAIN TECHNOLOGY

A. DATA COLLECTION

This research selected blockchain technology as the case for empirical analysis of the proposed scheme. The main reason to select blockchain technology for the case is that this technology has been termed the most disruptive technology of the decade. Incumbents and new-entrant firms have recently started investing in the research and development of products, services, and novel processes using this technology. Therefore, the proposed method fits the assessment of this technology to understand the significance and positions of the firms investing and innovating in blockchain technology. Few studies have focused on the patent-based analysis of blockchain technology. The US patents for blockchain technology are retrieved from the MTrends database. The patents are searched based on the search strategy. Literature and patent reports were consulted to formulate the appropriate terminology. Previous studies have used keywords like Blockchain [26] to identify trends. The bibliometric analysis literature [27], [28], [29] provides keyword networks for the technology or domain that can be used best for the proper search strategy. Furthermore, this study used the IPC codes (IPCs) to limit the patents to a specific technology domain. IPCs are an ordered representation of interconnected technology categories. Previous research [39] identified key IPC codes for blockchain technology, including H04L (transmission of digital information), G06F (electric digital data processing), G09C (ciphering or deciphering apparatus for cryptographic or other purposes requiring secrecy), H04N (pictorial communication), and G06Q. (data processing systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory, or forecasting purposes). Table 7 shows the technical nomenclature for the BLT, where the study divides the base technology, i.e., blockchain, into sub-technological fields, followed by IPC codes.

The base technology was categorized further into sub-technological fields (Table 5), and the fields were used to search subject terms across the database's title, abstract, and claims. The initial patents obtained were 10,919. However, the results obtained were sorted using the simple patent family merge function to avoid data duplication. The study obtained the resulting 6,206 patents after merging the results based on the simple patent family. Sample data is shown in

TABLE 5. Search strategy & data collection.

	Sub-Technological Fields	Patent Count	Final Data
Base Technology: Blockchain	Cryptocurrency	**10919** *(3106 - Published, 7813 - Applied)*	**Simple Family Merge**: 6206 patents
	Bitcoin		
	Decentralized database		
	Smart Contracts		
	Decentralized Ledger		
	Distributed Ledger		
	Proof-of-work		
	Ethereum		
	Namecoin		
	Litecoin		
Base IPC Codes	H04L		
	G06F		
	G09C		
	H04N		
	G06Q		

TABLE 6. Sample data.

Publication No.	Application Date	Country	Assignees
US11334439	29/08/18	US	IBM Co.
US10853341	01/12/20	KY	ANT
US11055442	06/07/2021	US	Capital One
US11044254	22/06/2021	US	BOA

table 6. The firm name coding and details are provided in APPENDIX A.

B. DATA PRE-PROCESSING

1) FIRM ADJACENCY NETWORK

The firm data is obtained for creating the firm citation matrix, where the minimum total citation for a firm is limited to 2 to have at least a link of two nodes within a network. The network is obtained after the patent affiliation and patent adjacency network. The firm adjacency matrix for 47 firms associated with the 6,206 patents is developed. The firm citations obtained are 2,046 for 47 firms after eliminating the citations of less than two from the preliminary network. The resulting firm adjacency network is used to obtain the centrality measures of the firms in the network. A sample network of the top five firms is shown in Table 7.

TABLE 7. Firm citation matrix sample for five firms.

FIRM	ANT	IBM	CAPITAL ONE	BOA
ANT	83	157	4	35
IBM	12	74	3	23
CAPITAL ONE	3	12	5	15
BOA	1	22	1	23

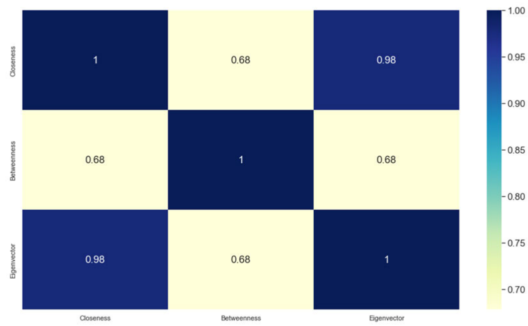


FIGURE 4. Correlation matrix for centrality measures.

2) CENTRALITY MEASURES

the centrality measures chosen in this study are betweenness, closeness, and eigenvector centrality. the reason for choosing these three centralities for the cluster analysis is that these centrality measures are network-based (path based) and highlight the node's importance across the network and sub-networks. The centrality measures of the 46 firms used as input for the clustering algorithm are provided in APPENDIX B. The centrality measures were obtained using the python networkx package. The correlation between the different centrality measures is consistent compared to other centralities. the correlation between the centralities is shown in Fig. 4. The correlation matrix shows the higher correlation between eigenvector and closeness centrality, i.e., 0.98. In contrast, betweenness has an equal correlation with both eigenvector and closeness centrality, i.e., 0.68. This indicates that the firms with higher closeness centrality have higher eigenvector centrality as well.

C. INNOVATION ANALYSIS

1) K-MEANS CLUSTERING

The elbow method is applied to identify the optimal number of clusters (k) for the classification. The study employed a python *scikit-learn* package to perform the k-means clustering algorithm. First, the study identifies the within cluster sum of squares (WCSS) for the different numbers of clusters. From the relation between WCSS and the

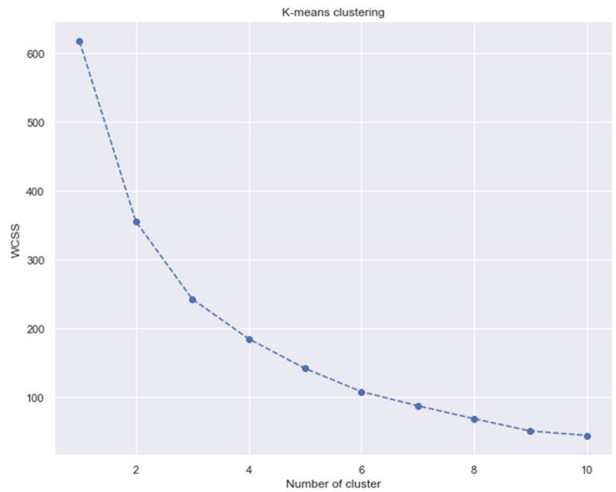


FIGURE 5. Choosing the number of k (elbow method).

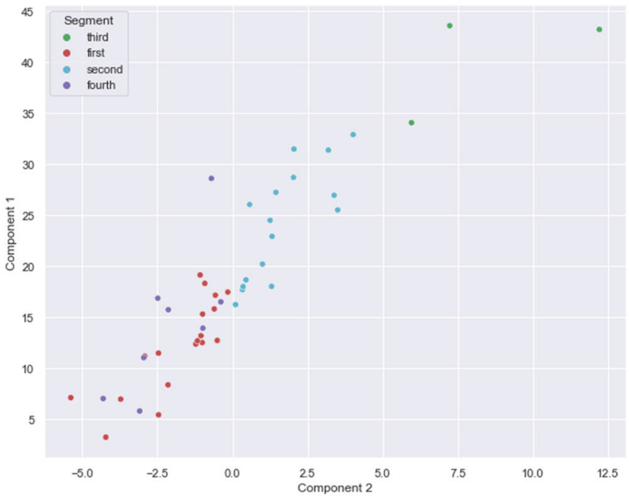


FIGURE 6. K-means clustering result.

number of clusters, we can identify the optimal number of clusters where the WCSS starts to drop, shown in Fig. 5. Finally, the elbow method indicates the optimal number of clusters to be four for this study. Finally, k-means clustering is applied based on the two highly correlated components, closeness, and eigenvector centrality. The visualization of item points in the cluster is shown in Fig. 6.

The clusters are summarized in table 8. where cluster one consists of 39% of the firms, cluster two consists of 35% of the firms, cluster three consists of 7% and cluster four consists of 20% of the total 46 firms. Once the clusters are obtained, the study develops the cluster citation network. The cluster adjacency network is shown in Table 9, where we can identify the highly cited cluster and the least cited cluster, as well as the highly citing and least citing clusters. Cluster two is the highly cited as well as highly citing cluster, which means the firms in cluster two have the highest citation correlation across them. moreover, least cited cluster is cluster one and least citing cluster is cluster four.

To obtain the final significance and position network relationship, we first identify the technology knowledge status within and between the clusters using the calculations explained in Eq. 7.

TABLE 8. Firm clusters.

Cluster ID	Firms in the cluster	Count	%
C_1	1. Mongo DB, Inc. (US), 2. United Services Automobile Association (US), 3. Wells Fargo Bank, National Association (US), 4. Gemini Ip, Llc (US), 5. VMware, Inc. (US), 6. AT&T Intellectual Property, L.P. (US), 7. American Express Travel Related Services Company, Inc. (US), 8. Citibank, National Association (US), 9. Nokia Technologies Oy. (FI), 10. The Toronto-Dominion Bank (CA), 11. NEC Corporation (JP), 12. Coinplug, Inc. (KR), 13. Fujitsu Limited (JP), 14. JPMorgan Chase Bank, National Association (US), 15. Red Hat, Inc. (US), 16. Walmart Apollo, Llc (US), 17. PayPal, Inc. (US), 18. Pivotal Software, Inc. (US)	18	39%
C_2	1. Capital One Services, LLC (US), 2. Salesforce.Com, Inc. (US), 3. State Farm Mutual Automobile Insurance Company (US), 4. Accenture Global Solutions Limited (IE), 5. Amazon Technologies, Inc. (US), 6. Oracle International Corporation (US), 7. British Telecommunications Public Limited Company (GB), 8. The Bank Of New York Mellon Trust Company, N.A. (US), 9. Sap Se (DE), 10. Hitachi, Ltd. (JP), 11. Microsoft Technology Licensing, LLC. (US), 12. Cisco Technology, Inc. (US), 13. Mastercard International Incorporated (US), 14. Intel Corporation (US), 15. Visa International Service Association (US), 16. Keir Finlow-Bates (FI)	16	35%
C_3	1. Advanced New Technologies Co., Ltd. (US), 2. International Business Machines Corporation (US), 3. Bank of America Corporation (US)	3	7%
C_4	1. Alibaba Group Holding Limited (KY), 2. Hewlett Packard Enterprise Development LP (US), 3. Deutsche Bank Trust Company Americas (US), 4. Verizon Patent and Licensing Inc. (US), 5. Kyndryl, Inc. (US), 6. Dell Products L.P. (US), 7. EMC IP Holding Company LLC (US), 8. Google LLC. (US), 9. Huawei Technologies Co., Ltd. (CN)	9	20%

TABLE 9. Cluster adjacency network.

Clusters	C_1	C_2	C_3	C_4
C_1	15	37	17	2
C_2	48	93	30	9
C_3	33	43	9	6
C_4	12	44	18	7

TABLE 10. Knowledge flow network.

Clusters	C_1	C_2	C_3	C_4
C_1	0.05	0.13	0.31	0.01
C_2	0.17	0.39	0.63	0.06
C_3	0.61	0.90	1.50	0.22
C_4	0.07	0.31	0.67	0.10

*τ=0.20

TABLE 11. Firm innovation significance and positions network.

Clusters	C_1	C_2	C_3	C_4
C_1	0	0	1	0
C_2	0	1	1	0
C_3	1	1	1	1
C_4	0	1	1	0

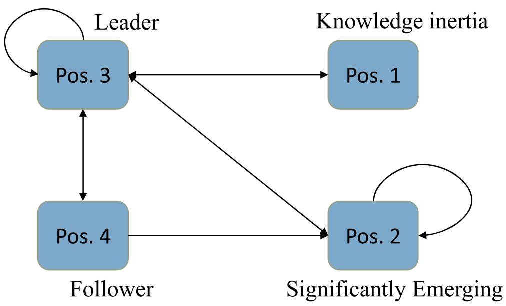


FIGURE 7. Firm innovation significance and positions.

2) IDENTIFYING INNOVATION SIGNIFICANCE

To determine the relevance of the firm citation link strength between innovation clusters, we calculate the average of firm citation network, i.e., $\tau = 0.20$. as shown in Table 10, firm citation strength less than 0.20 implies that the citation link between clusters is not significant.

Table 11 displays the resulting technological inter- and intra-cluster connections that are used to construct a network of innovation significance and positions. The four clusters are therefore mapped based on the network, and their respective significance and positions are determined. After evaluating the significance and positions of firms in different groups, the distinct categories are depicted in Fig. 7.

The firm description for each cluster related to the country, patent count and citation count is provided in APPENDIX C. the firm innovation significance and positions identified are described further.

a: POSITION 1: KNOWLEDGE INERTIA

The firms in the position of knowledge inertia, can be considered as the firms that are innovating within their own bubble, but also creating important enough knowledge to be flowing to and from the firms in the role of leader itself. The innovations from firms in this role are mostly incumbents scaling up to innovate in the breakthrough technology and in the process gaining the technological knowledge from the leader firms.

The number of patents and their citations have an increasing relationship in this cluster, as shown in Fig. 8. although the frequency of patent innovation is not higher, but higher citation count indicates the growing impact of the innovations from the firms in this cluster.

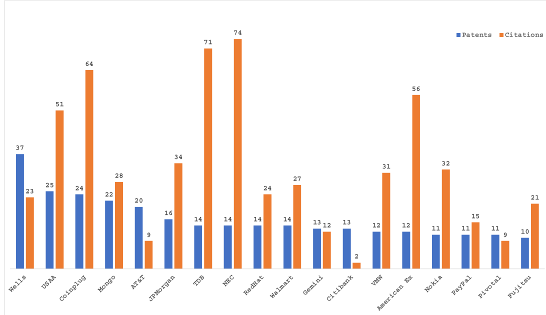


FIGURE 8. Patent and citation count for cluster 1.

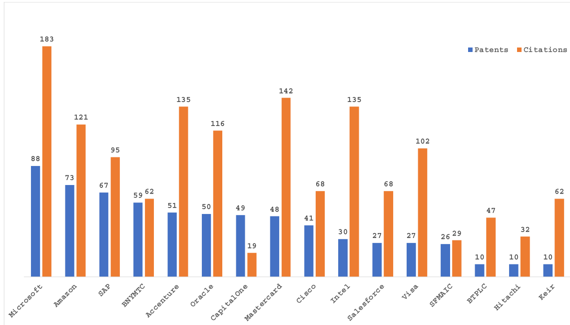


FIGURE 9. Patent and citation count for cluster 2.

b: POSITION 2: SIGNIFICANTLY EMERGING

The firms in the position of significantly emerging can be considered as prolific innovators in the technology domain. The firms not only follow the knowledge path of their innovation output but are also emerging on the innovation output of firms in other positions. The firms in this cluster are technological incumbents innovating significantly in the select technology domain, and therefore significantly emerging on the knowledge flow path.

The patent and citation count for the firms in cluster 2 is shown in Fig. 9.

c: POSITION 3: LEADER

The firms in leader position have the highest technology knowledge flow within and between the technological innovation domains. The firms in this role, are the most exhaustive innovators in the technology domain, where all other positions follow the innovation output from the firms in this role. The firms in this role are the pioneer incumbents or the first movers in the disruptive marketplace.

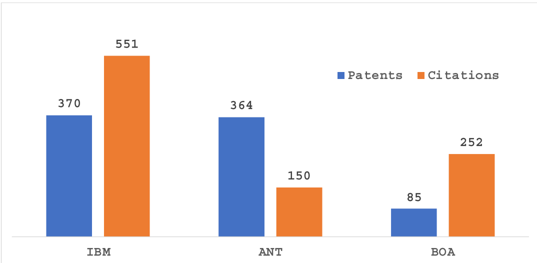


FIGURE 10. Patent and citation count for cluster 3.

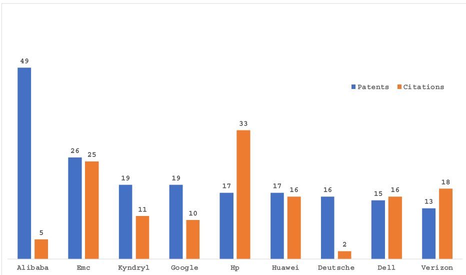


FIGURE 11. Patent and citation count for cluster 4.

The firms in this cluster have higher number of patents and higher citation count as well, shown in Fig. 10

d: POSITION 4: FOLLOWER

The firms in the follower position are considered to follow the technological flow of firms in the other positions, however, they act as the absorbers of knowledge rather than producers. This cluster also has the lowest cluster citation and cited score in the cluster adjacency matrix, indicating their lower contribution to the knowledge flow path of the network.

Firms in this position have comparatively a lesser number of patents and citations than the rest of the positions, as shown in Fig. 11.

V. CONCLUSION, LIMITATIONS, AND FUTURE SCOPE

The study developed a structural method to identify firms' technology knowledge significance and positions in a patent innovation network. The study used the case of blockchain technology to assess the effectiveness of the proposed method. The proposed method employed patent and firm affiliation relationships to obtain firm citation network and calculate centrality measures for the firms in the network. Furthermore, k-means clustering is used to categorize firms based on centrality measures. The study utilized a cluster adjacency matrix to develop the technology knowledge flow network, which was then utilized to create the innovation significance network. This approach builds upon previous research, such as the method proposed by [23] for identifying positions and roles in patent litigation networks using multivariate analysis of variances (manova), as well as the technology knowledge status and redundancy methods developed by [25] for identifying firm roles and positions. However, this

TABLE 12. Firm Coding.

Code	Abbreviation	Firm Name
F1	ANT	Advanced New Technologies Co., Ltd.
F2	IBM	International Business Machines Corporation
F3	MONGO	Mongo DB, Inc.
F4	CAPITALONE	Capital One Services, LLC
F5	BOA	Bank of America Corporation
F6	SALESFORCE	Salesforce.Com, Inc.
F7	SFMAIC	State Farm Mutual Automobile Insurance Company
F8	ALIBABA	Alibaba Group Holding Limited
F9	ACCENTURE	Accenture Global Solutions Limited
F10	AMAZON	Amazon Technologies, Inc.
F11	ORACLE	Oracle International Corporation
F12	BTPLC	British Telecommunications Public Limited Company
F13	HP	Hewlett Packard Enterprise Development LP
F14	BNYMTC	The Bank Of New York Mellon Trust Company, N.A.
F15	SAP	Sap Se
F16	USAA	United Services Automobile Association
F17	HITACHI	Hitachi, Ltd.
F18	MICROSOFT	Microsoft Technology Licensing, LLC.
F19	DEUTSCHE	Deutsche Bank Trust Company Americas
F20	CISCO	Cisco Technology, Inc.
F21	MASTERCARD	Mastercard International Incorporated
F22	WELLS	Wells Fargo Bank, National Association
F23	INTEL	Intel Corporation
F24	GEMINI	Gemini Ip, Llc
F25	VERIZON	Verizon Patent and Licensing Inc.
F26	VISA	Visa International Service Association
F27	VMW	VMware, Inc.

TABLE 12. ***(Continued.)*** **Firm Coding.**

F28	KYNDRYL	Kyndryl, Inc.
F29	KEIR	Keir Finlow-Bates
F30	DELL	Dell Products L.P.
F31	EMC	EMC Ip Holding Company LLC
F32	AT&T	At&T Intellectual Property, L.P.
F33	AMERICAN EX	American Express Travel Related Services Company, Inc.
F34	CITIBANK	Citibank, National Association
F35	NOKIA	Nokia Technologies Oy
F36	TDB	The Toronto-Dominion Bank
F37	NEC	NEC Corporation
F38	COINPLUG	Coinplug, Inc.
F39	FUJITSU	Fujitsu Limited
F40	GOOGLE	Google LLC.
F41	HUAWEI	Huawei Technologies Co., Ltd.
F42	JPMORGAN	JPMorgan Chase Bank, National Association
F43	REDHAT	Red Hat, Inc.
F44	WALMART	Walmart Apollo, Llc
F45	PAYPAL	PayPal, Inc.
F46	PIVOTAL	Pivotal Software, Inc.

study extends the methodology by incorporating a k-means unsupervised clustering algorithm and integrating the technology knowledge flow with firm cluster networks to identify firm significance and positions. The resulting analysis categorized firms into four clusters based on their position and roles. By formulating clear research questions and leveraging these methods and results, this study provides valuable insights into the technological innovation landscape.

RQ1: Do firms that belong in a particular innovation cluster have unique features that influence the whole patent innovation landscape?

Each innovation cluster obtained after the analysis has unique features, and the significance is based on those features. Firms belonging to the leader innovation cluster have the higher influence in the innovation network. Based on

TABLE 13. **Firm centrality measures.**

Code	Closeness	Betweenness	Eigenvector
F1	0.544585	0.142863	0.145361
F2	0.796708	0.314328	0.390394
F3	0.387588	0.000603	0.036692
F4	0.488889	0.027951	0.096187
F5	0.672222	0.120347	0.309449
F6	0.524661	0.024437	0.146194
F7	0.500258	0.004444	0.124754
F8	0.349774	0.006047	0.015854
F9	0.605947	0.038986	0.232099
F10	0.605947	0.07257	0.213414
F11	0.55873	0.005886	0.210556
F12	0.488889	0.00726	0.107575
F13	0.443528	0.021453	0.064627
F14	0.55873	0.06006	0.17042
F15	0.51834	0.013936	0.152255
F16	0.51834	0.002553	0.111977
F17	0.512169	0.003673	0.100984
F18	0.651852	0.054364	0.282261
F19	0.355556	0	0.013797
F20	0.55873	0.030353	0.197428
F21	0.589346	0.021397	0.213316
F22	0.434568	0.002928	0.049907
F23	0.589346	0.015117	0.205499
F24	0.425963	0.001185	0.056913
F25	0.452865	0.000574	0.047689
F26	0.605947	0.002056	0.232328
F27	0.478025	0.007142	0.095023
F28	0.361531	0.001002	0.014773
F29	0.531139	0.036414	0.167031
F30	0.361531	0	0.020939
F31	0.361531	0	0.020939
F32	0.472772	0.001003	0.08163
F33	0.524661	0.002595	0.140871
F34	0.377388	0	0.020256
F35	0.494508	0.002887	0.099144
F36	0.506144	0.000519	0.126816
F37	0.500258	0.000816	0.1044
F38	0.512169	0.003993	0.121757
F39	0.478025	0.002342	0.086568

TABLE 13. ***(Continued.)*** **Firm centrality measures.**

F40	0.384127	0.000571	0.030313
F41	0.358519	0	0.013797
F42	0.512169	0.001347	0.144975
F43	0.51834	0.001593	0.14403
F44	0.452865	0	0.052254
F45	0.439002	0.003013	0.059133
F46	0.48913	0	0.082058

the technology knowledge flow within (1.65) and between (0.61 for C_1, 0.90 for C_2, and 0.22 for C_4) the firm clusters, the higher technology status within clusters indicates higher patent innovations across leader position. Firms in this cluster were identified as ANT, IBM, and BOA. IBM has been at the top in terms of patenting activities in advanced technologies. Furthermore, in the analyzed data IBM has the highest number of patents, i.e., 370, ANT with 193 patents and BOA with 85 patents. The leader firms also dominate the citation count in the dataset. Furthermore, the firms in leader position have citation links with all other firms in different positions in the network, indicating the higher influence. Moreover, the innovation clusters in significantly emerging and knowledge inertia position highlight the adaptability of the incumbents and traditional firms to respond to the disruptions in the marketplace.

RQ2: What relationship do firms have among and across the innovation clusters in the different positions?

Although different innovation clusters have different positions and influence across the network. However, the influence is developed based on the inter-connectedness across the network. Looking at the firm innovation significance and positions network in Table 10, innovation position two has the higher interconnectedness among its own firms, i.e., the significantly emerging firms have higher activity compared to firms in other positions. Furthermore, the firms in position three and position four are also more connected with firms in cluster two in their knowledge flow path. The interconnectedness of the innovation clusters allows researchers to understand the knowledge flow of the technological innovation further in the patent landscape. Firms in the position of leader (C_3), follower (C_4) and significantly emerging (C_2) are more closely connected in the technology innovation path. Firms in cluster two follow firms in leader position more than the other positions since the knowledge flow between these two positions is 0.63 (C_2 *to* C_3,) and 0.90 (C_3 *to* C_2). This indicates that even though the network is influenced by position three, leader, however, the firms in this position have their knowledge dependency tied with position two firms. The firms in position one, i.e., knowledge inertia have high corre-

TABLE 14. **Firm description.**

Code	Country	Cluster	Patents	Citations
F3	US	1	22	28
F16	US	1	25	51
F22	US	1	37	23
F24	US	1	13	12
F27	US	1	12	31
F32	US	1	20	9
F33	US	1	12	56
F34	US	1	13	2
F35	FI	1	11	32
F36	CA	1	14	71
F37	JP	1	14	74
F38	KR	1	24	64
F39	JP	1	10	21
F42	US	1	16	34
F43	US	1	14	24
F44	US	1	14	27
F45	US	1	11	15
F46	US	1	11	9
F4	US	2	49	19
F6	US	2	27	68
F7	US	2	26	29
F9	IE	2	51	135
F10	US	2	73	121
F11	US	2	50	116
F12	GB	2	10	47
F14	US	2	59	62
F15	DE	2	67	95
F17	JP	2	10	32
F18	US	2	88	183
F20	US	2	41	68
F21	US	2	48	142
F23	US	2	30	135
F26	US	2	27	102
F29	FI	2	10	62
F1	US	3	364	150
F2	US	3	370	551
F5	US	3	85	252

TABLE 14. ***(Continued.)*** **Firm description.**

F8	KY	4	49	5
F13	US	4	17	33
F19	US	4	16	2
F25	US	4	13	18
F28	US	4	19	11
F30	US	4	15	16
F31	US	4	26	25
F40	US	4	19	10
F41	CN	4	17	16

lation with firms in leader position, i.e., the knowledge flow from position one to position three is 0.61 and 0.31 for vice-versa, which further indicates the dependency of innovation from each cluster. This research established the method to identify the significance across the positions in the knowledge flow network.

Patent analysis provides valuable insights into technology's past, present, and future. By analyzing patent data, decision-makers and researchers can better understand the technological landscape. Firms tend to secure their inventions through patents before publishing their outcomes in the public domain. Timely analysis of patent outputs can provide competitors with information to devise their technological and market strategies based on the results. Therefore, by utilizing citation analysis to identify clusters of firms, the study can identify the technology roles and firm positions, distinguishing first movers from followers. For example, the study's results identify ANT, IBM, and BOA as leaders in Blockchain technology, with high patent activity in general. In contrast, firms such as alibaba, hp, and dell are identified as followers, despite being incumbents. Understanding a firm's position in the innovation domain and their patent strategies is crucial for forecasting competitor patent activities, as a firm may have various patents related to different components when developing a product.

The study makes a significant contribution to the field of patent analysis by proposing a novel method for identifying the significance and positions of firms in the patent innovation landscape. The uniqueness of this study lies in the structural model used to analyze firm significance. While previous studies have explored various dimensions of the technology innovation landscape, such as patent citations, patent and firm affiliations, and firm citations, this study adds a new dimension of cluster adjacency networks to identify similar characteristics in the innovation landscape. This study identifies firm significance and positions by combining citation analysis, centrality measure analysis, and k-means clustering. In social network analysis, it is crucial to understand the status of different entities and their relationships with other entities in the network. The results of this method can be helpful to decision-makers in identifying their technological strategies by identifying the positions of firms innovating in the selected domain. This can help evaluate the strategic position where an organization wants to prioritize technological innovation. The positions are identified based on inter- and intra-relationship across the firm clusters, providing additional insights into firm innovation beyond solely utilizing a citation analysis network. The study, however, has its limitations too, for example, it utilizes centrality-based clustering followed by adjacency relationships between the firms to identify positions. Furthermore, the choice of the dataset is limited to US patents for the sake of data completeness. Future research can employ different methods apart from centrality measures to identify novel firm characteristics and relationships using different country office patent datasets or technology.

APPENDIX A

See Table 12.

APPENDIX B

See Table 13.

APPENDIX C

See Table 14.

REFERENCES

[1] P. Soto-Acosta, S. Popa, and I. Martinez-Conesa, "Information technology, knowledge management and environmental dynamism as drivers of innovation ambidexterity: A study in SMEs," *J. Knowl. Manag.*, vol. 22, no. 4, pp. 824–849, 2018.

[2] T.-M. Chen, C.-C. Wei, and H.-C. Che, "Contribution of patent indicators to China stock performance," in *Proc. 7th Int. Congr. Adv. Appl. Informat. (IIAI-AAI)*, Jul. 2018, pp. 793–798.

[3] F. Altuntas and M. S. Gök, "Analysis of patent documents with utility mining: A case study of wind energy technology," *Kybernetes*, vol. 50, no. 9, pp. 2548–2582, Aug. 2021.

[4] K.-K. Lai, H.-C. Chen, Y.-H. Chang, V. Kumar, and P. C. Bhatt, "A structured MPA approach to explore technological core competence, knowledge flow, and technology development through social network patentometrics," *J. Knowl. Manag.*, vol. 25, no. 2, pp. 402–432, Mar. 2021.

[5] S. H. Baruffaldi and M. Simeth, "Patents and knowledge diffusion: The effect of early disclosure," *Res. Policy*, vol. 49, no. 4, May 2020, Art. no. 103927.

[6] L. J. Aaldering, J. Leker, and C. H. Song, "Analyzing the impact of industry sectors on the composition of business ecosystem: A combined approach using ARM and DEMATEL," *Exp. Syst. Appl.*, vol. 100, pp. 17–29, Jun. 2018.

[7] Z. Wang, A. L. Porter, X. Wang, and S. Carley, "An approach to identify emergent topics of technological convergence: A case study for 3D printing," *Technol. Forecasting Social Change*, vol. 146, pp. 723–732, Sep. 2019.

[8] Y. Cho, Y. J. Han, J. Hwang, J. Yu, S. Kim, C. Lee, S. Lee, and K. P. Yi, "Identifying technology opportunities for electric motors of railway vehicles with patent analysis," *Sustainability*, vol. 13, no. 5, pp. 1–14, 2021.

[9] I. Park and B. Yoon, "Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network," *J. Informetrics*, vol. 12, no. 4, pp. 1199–1222, Nov. 2018.

[10] B. Yan and J. Luo, "Multicores-periphery structure in networks," *Netw. Sci.*, vol. 7, no. 1, pp. 70–87, Mar. 2019.

[11] V. Kumar, K.-K. Lai, Y.-H. Chang, P. C. Bhatt, and F.-P. Su, "A structural analysis approach to identify technology innovation and evolution path: A case of M-payment technology ecosystem," *J. Knowl. Manag.*, vol. 25, no. 2, pp. 477–499, Mar. 2021.

[12] S. J. Jee, M. Kwon, J. M. Ha, and S. Y. Sohn, "Exploring the forward citation patterns of patents based on the evolution of technology fields," *J. Informetrics*, vol. 13, no. 4, Nov. 2019, Art. no. 100985.

[13] Y. Hu, J. Wen, and Y. Yan, "Measuring the performance of knowledge resources using a value perspective: Integrating BSC and ANP," *J. Knowl. Manag.*, vol. 19, no. 6, pp. 1250–1272, Oct. 2015.

[14] L. Feng, Y. Li, Z. Liu, and J. Wang, "Idea generation and new direction for exploitation technologies of coal-seam gas through recombinative innovation and patent analysis," *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, p. 2928, Apr. 2020.

[15] S. Choi, M. Afifuddin, and W. Seo, "A supervised learning-based approach to anticipating potential technology convergence," *IEEE Access*, vol. 10, pp. 19284–19300, 2022.

[16] A. Sun and R. Li, "Can patent quality alleviates information asymmetry in Chinese intellectual property market: Evidence from auditing field," *IEEE Access*, vol. 10, pp. 79335–79347, 2022.

[17] W. Lin, R. Lai, and R. Xiao, "Patent data-driven product innovative design based on complex network," *Zhongguo Jixie Gongcheng/China Mech. Eng.*, vol. 31, no. 7, pp. 842–852, 2020.

[18] S. Choi, M. S. Noh, J. Yoon, H. Park, and W. Seo, "Analyzing technological spillover effects between technology classes: The case of Korea technology finance corporation," *IEEE Access*, vol. 6, pp. 3573–3584, 2018.

[19] J. Park, E. Heo, and D. Lee, "Effective R&D investment planning based on technology spillovers: The case of Korea," *Scientometrics*, vol. 111, no. 1, pp. 67–82, Apr. 2017, doi: 10.1007/s11192-017-2276-3.

[20] X. Xu and M. Gui, "Applying data mining techniques for technology prediction in new energy vehicle: A case study in China," *Environ. Sci. Pollut. Res.*, vol. 28, no. 48, pp. 68300–68317, Dec. 2021.

[21] A. Sukumar, V. Jafari-Sadeghi, A. Garcia-Perez, and D. K. Dutta, "The potential link between corporate innovations and corporate competitiveness: Evidence from IT firms in the UK," *J. Knowl. Manag.*, vol. 24, no. 5, pp. 965–983, Jun. 2020.

[22] Y. Zhao, Z. Li, L. Li, and N. Qi, "Research on technological divisive faultlines in cliques: Evidence from China's biomedical industry innovation network," *IEEE Access*, vol. 10, pp. 89745–89757, 2022.

[23] Y.-H. Chang, K.-K. Lai, C.-Y. Lin, F.-P. Su, and M.-C. Yang, "A hybrid clustering approach to identify network positions and roles through social network and multivariate analysis," *Scientometrics*, vol. 113, no. 3, pp. 1733–1755, Dec. 2017.

[24] C.-L. Chang, K.-K. Lai, H.-C. Chen, and H.-J. Chang, "A structural approach to identify the position and role of the litigation relation network of smartphone companies," in *Proc. Int. Conf. Eng., Technol. Innov. (ICE/ITMC)*, Jun. 2017, pp. 111–118.

[25] H. J. Chang, H. C. Chen, Y. H. Chang, V. Kumar, C.-Y. Lin, and Y. R. Lee, "A structured approach to locate the technological position by technology knowledge redundancy-patent citation network perspective," *Int. J. Inf. Manag. Sci.*, vol. 31, no. 1, pp. 55–78, 2020.

[26] T. Daim, K. K. Lai, H. Yalcin, F. Alsoubie, and V. Kumar, "Forecasting technological positioning through technology knowledge redundancy: Patent citation analysis of IoT, cybersecurity, and blockchain," *Technol. Forecasting Social Change*, vol. 161, Dec. 2020, Art. no. 120329.

[27] P. C. Bhatt, V. Kumar, T.-C. Lu, and T. Daim, "Technology convergence assessment: Case of blockchain within the IR 4.0 platform," *Technol. Soc.*, vol. 67, Nov. 2021, Art. no. 101709.

[28] S. Miau and J.-M. Yang, "Bibliometrics-based evaluation of the blockchain research trend: 2008—March 2017," *Technol. Anal. Strategic Manag.*, vol. 30, no. 9, pp. 1029–1045, Sep. 2018.

[29] H. Yalcin and T. Daim, "Mining research and invention activity for innovation trends: Case of blockchain technology," *Scientometrics*, vol. 126, no. 5, pp. 3775–3806, May 2021.

[30] K.-K. Lai, P. C. Bhatt, V. Kumar, H.-C. Chen, Y.-H. Chang, and F.-P. Su, "Identifying the impact of patent family on the patent trajectory: A case of thin film solar cells technological trajectories," *J. Informetrics*, vol. 15, no. 2, May 2021, Art. no. 101143.

[31] T. Wada, "The choice of examiner patent citations for refusals: Evidence from the trilateral offices," *Scientometrics*, vol. 117, no. 2, pp. 825–843, Nov. 2018, doi: 10.1007/s11192-018-2885-5.

[32] P. Das, R. Verburg, A. Verbraeck, and L. Bonebakker, "Barriers to innovation within large financial services firms," *Eur. J. Innov. Manag.*, vol. 21, no. 1, pp. 96–112, Jan. 2018.

[33] C. Tang and G. Zhang, "The moderating effects of firm's and industrial co-inventive networks on the relationship between R and D employees mobility and firm creativity," *IEEE Trans. Eng. Manag.*, vol. 69, no. 5, pp. 2102–2116, Oct. 2022.

[34] A. Ficara, G. Fiumara, P. De Meo, and A. Liotta, "Correlation analysis of node and edge centrality measures in artificial complex networks," in *Proc. 6th Int. Congr. Inf. Commun. Technol.*, 2022, pp. 901–908.

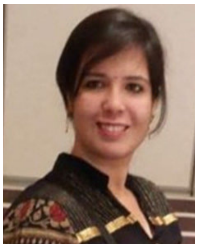
[35] M. Saqr, R. Elmoazen, M. Tedre, S. López-Pernas, and L. Hirsto, "How well centrality measures capture student achievement in computer-supported collaborative learning?—A systematic review and meta-analysis," *Educ. Res. Rev.*, vol. 35, Feb. 2022, Art. no. 100437.

[36] P. Ampornphan and S. Tongngam, "Exploring technology influencers from patent data using association rule mining and social network analysis," *Information*, vol. 11, no. 6, p. 333, Jun. 2020.

[37] J. Jeon and Y. Suh, "Multiple patent network analysis for identifying safety technology convergence," *Data Technol. Appl.*, vol. 53, no. 3, pp. 269–285, Sep. 2019.

[38] K.-K. Lai, C.-Y. Lin, Y.-H. Chang, M.-C. Yang, and W.-G. Yang, "A structured approach to explore technological competencies through R&D portfolio of photovoltaic companies by patent statistics," *Scientometrics*, vol. 111, no. 3, pp. 1327–1351, Jun. 2017.

[39] P. C. Bhatt, V. Kumar, T. C. Lu, R. L. T. Cho, and K. K. Lai, "Rise and rise of blockchain: A patent statistics approach to identify the underlying technologies," in *Proc. Intell. Inf. Database Syst., 12th Asian Conf.*, 2020, pp. 456–466.



PRIYANKA C. BHATT has worked with institutes of national importance, such as SIDBI, IIT Kanpur; HBCSE—Tata Institute of Fundamental Research, Mumbai; and National Medical Library, New Delhi. She is currently a Ph.D. Research Scholar with the Department of Information Management, College of Informatics, Chaoyang University of Technology, Taiwan. She has an extensive experience of more than five years in the areas of computer engineering and information science. Her research interests include scientometric, patent statistics, social network analysis, data mining, cloud computing, and ICT applications.

TZU-CHUEN LU received the B.M. and M.S.I.M. degrees in information management from the Chaoyang University of Technology, Taichung, Taiwan, 1999 and 2001, respectively, and the Ph.D. degree in computer engineering from National Chung Cheng University, in 2006. She is currently a Professor with the Department of Information Management, Chaoyang University of Technology. Her research interests include computational complexity, steganography, data encapsulation, and decision support systems.