Using Social Network Analysis to Study Diversity in Business Partnerships

Rola Abdulkarim
British University in Dubai
Sherief Abdallah (sherief.abdallah@buid.ac.ae)
British University in Dubai
Abdulla AlHossani
Department of Economy and Tourism, Dubai

Research Article

Keywords: Graph mining, Social network analysis, Community detection, Business entrepreneurship analysis

Posted Date: May 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1582025/v1

Abstract

Understanding the key elements of successful business partnerships has significant economic value. In this paper we analyze, for the first time, real world trade license data as a social network. The data we analyze cover the license registrations from 2015 to 2017 in a major cosmopolitan city in the middle east. The dataset consists of more than twenty thousand trade licenses and thirty-nine thousand individuals (business partners). We view the data as two complementary networks that evolve over time: one network where a node represents a partner and an edge represents the existence of partnership (common trade licenses), and another network where a node represents a license and an edge connects two licenses that have common partners. We apply community detection algorithms to both networks and carefully investigate how the top clusters evolve over time. Our analysis shows that business partnerships exhibit diversity with respect to age and nationality, but not when it comes to gender. We also identify common motifs that repeat in these networks and remain consistent over time.

I. Introduction

Understanding the key elements of successful business partnerships is of interest to both government issuing licenses and investors who form such partnerships. Between 70% and 80% of business partnerships eventually fail [1,2]. Surprisingly, very little research was done on analyzing real-world business partnership data, particularly as a social network. This in part due to the lack of publicly available data. In this paper we take advantage of having access to data set of trade licenses in a cosmopolitan city. To our knowledge, we provide the first analysis of such data as a social network of partners and licenses.

The business partners dataset was obtained from the Department of Economy and Tourism in Dubai city. Having been recognized as the most cosmopolitan city in the world in 2015 with 83% foreign residents [3] and hosting residents from more than 200 nationalities [4], Dubai city provides a fertile ground for studying how partnerships form and evolve in a diverse society.

We propose and analyze two complementary encodings of the data as a network: in one encoding nodes represent licenses and an edge connects nodes that share at least one partner, while the other encoding represents each partner as a node and an edge connects two partners if they share at least one license.

In our analysis, we aim to answer the following research questions:

1) If we apply community detection algorithms to networks representing business partnerships and licenses, can we identify common motifs?

2) How diverse are the identified communities/clusters with respect to partners demographics (age, nationality, and gender) and business activities (tourism, real estate, etc.)?

3) Do the networks follow the well-known power-law distribution?

4) Do the answers to the above questions remain consistent over the studied three years?

To our knowledge, we provide the first analysis of trade licenses data using social network analysis methodology. We identify interesting patterns both in terms of common motifs and the edge weight distribution. Our research lays the foundation for studying the effect of the diversity on success of partnerships.

Ii. Literature Review

Aiming to examine the spatial network structure of the tourism economy within urban agglomeration in the middle reaches of the Yangtze River (UAMRYR), researchers have applied social network analysis(SNA). The tourism economic gravity model have been also applied to evaluate the tourism economic quality of each city of the 28 cities in UAMRYR. SNA was applied using ArcGIS10.2 software to gain more insight into the social characteristics using multiple network metrics and measures. For instant, network density gave an indication of the closeness between different cities within the network. Also, betweenness centrality used to identify cities of significant role in controlling the tourism economic connection[17]. The researchers focused also on the out-degree and in-degree network metrics to measure the level of tourism economic development of each city within the constructed network. Moreover, the researchers split the cities into four plates based on the convergent correlation for further analysis. The utilization of SNA assisted in identifying the cities of higher tourism economic quality and better transportation accessibility, and that the spatial network structure of the tourism economy within UAMRYR is loose and unstable[17]. In another research, SNA have been exploited to understand the Italian tourism system. The unite of analysis was the Italian travel agencies, in which these agencies represented the nodes, and the links are the connections established between these agencies. Using modularity, seven communities were discovered, two of them are larger than others. Modularity measure was around 0.6 and the network was relatively sparse. The researchers claimed the low network connectivity is relative to the general efficiency of the system. Whereas low level of collaboration and cooperation between actors was observed in terms of hospitality. Using SNA It have been concluded that the collaboration between travel agencies was of poor efficiency and highlighted the importance of considering the qualitative analyses of the network’s parameters for enhanced recommendations[18]. We see that applying SNA in the tourism sector extend our perception on the effectiveness of SNA in the business domain. On the other hand, considering the role of entrepreneurship in economic growth and development has gained increasing interest in recent research as economists and policymakers argue that the level of entrepreneurship contributes to public success [6]. Similarly, in consideration to a significant impact of entrepreneurship, two entrepreneurship analysis studies have been conducted aiming to support decision making for successful start-ups [7][8]. The focus of the first research is the analysis of the business models of data-driven start-up firms to identify commonalities between the models pursued by the firms. The second research on the other hand, focused on applying data mining classification techniques to anticipate a successful business project and discover key factors to entrepreneurship success or failure [8]. Distinct methodologies were applied by each study where one exploited unsupervised data mining techniques [7]. And the other implemented a supervised technique [8].

The unsupervised data mining application involved the utilization of the k-medoids algorithm in implementing the data mining technique of clustering. Two steps were applied to identify attributes of interest. First, they identified six dimensions common among business models, including key resources possessed by the firm, data-related main activities, such as data preprocessing and transformation, value proposition, defined as the product or service the customer's value, targeted customers, revenue model, and cost structure. The second step was to identify the features related to each dimension. For example, the main features related to a data-driven firm's resources were the data and data sources, and the main activities dimension's features included data generation, acquisition, processing, aggregation, analytics, visualisation, and dissemination. The second step resulted in 35 features covering the six business model dimensions. The output of these two process steps is a data-driven business model (DDBM) framework [7]. The researchers prepared a sample of 100 randomly selected data-driven firms for the analysis where data on these firms' business models were collected from different resources, which were then coded against the developed DDBM framework resulting in binary feature vectors [7]. With the prepared dataset, the analysis was applied through four stages of selection of clustering variables, selection of a clustering algorithm and similarity metrics, specifying the number of clusters, and results in validation and analysis. The authors selected nine related variables out of the 35 features captured by the DDBM framework for the clustering task. Seven clusters were next specified, and the k-medoids and the Euclidean distance measure algorithms were used to execute the clustering task. For verification, the clustering analysis was repeated using a different algorithm, the silhouette coefficient [9], to verify the quality of the cluster, and case studies were applied by the researchers to review clusters significance. The result of the clustering task was seven clusters of data-driven start-up firms, where one of these clusters has been neglected due to insufficient similarity among its firms. Each of the six clusters was analysed in terms of the seven cluster variables where four distinct business models were discovered for the data source variable and three patterns were identified in term of key activities. As described, the study was centred around a business domain of data-driven companies, and did not consider aspects related to business owners [7]. For the approach adopted by the second research, classification was the data mining technique chosen for analysing business start-ups data. The three classification algorithms were applied are decision tree algorithm, Apriori algorithm, and Logistic regression algorithm [8]. The analysis process began with the acquisition of entrepreneurship datasets where records of 63 entrepreneurs who participated in the CCEEmprende program for entrepreneurial development support between 2007 and 2010 were collected [8]. The dataset includes a set of features related to the entrepreneurs including age, gender, education level, employment status, reasons for establishing entrepreneurships, type of support (family, economic, moral), reason for participating in the CCEEmprende program, project funding, and success indicators specified as the creation of the enterprise and generated income [8]. For the implementation of the classification task, researchers first applied the decision tree algorithm using SPSS to identify the path associated with enterprise success or failure. Next, Tanagra data mining software was used to generate the association rules. Two rules were related to enterprise success, called 'having funds' and 'independent labour situation,' and two for enterprise failure, called 'having no fund' and 'dependent labour situation' [8]. The authors used support and confidence measures to observe association rules quality verifications. Finally, logistic regression including the Wald test was performed highlighting the two significant variables of ‘having fund’ and ‘entrepreneur pre-existing employment situation’ [8]. The researchers concluded the two key factors related to entrepreneurship success are the entrepreneur’s financing and previous employment status. In summary, both reviewed studies analysed entrepreneurship from different perspectives using alternate tools and methodologies. One focused on clustering data-driven start-ups relying on firms' attributes [7]. While clustering is also the approach in this paper, we utilise the data related to entrepreneurs and the partnerships regardless of the firms' economic activities. On the other hand, the second study focused on entrepreneurs' attributes with an objective to predict entrepreneurship success or failure [8]. Due to present limitations and unavailability of related variables, a prediction of enterprise status is out of scope for this research.

Iii. Dataset

The dataset was provided by the Department of Economy and Tourism in Dubai. The data covered active trade licenses that were established over three years period: 2015 to 2017. TABLE I presents main dataset attributes and the details of each. The dataset includes 89,547 records for 2015, 89,260 for 2016, and 82,710 for 2017. Each record represents an economic activity for each license where one trade license can be attached to multiple economic activities.

Iv. Business Licenses Network Analysis

The aim of this section is to understand the diversity in terms of business activities over the years of analysis and whether there are common motifs in the networks. We focus on two attributes: Activity master group and Activity category (TABLE I). Figure 1, Figure 2, and Figure 3 show the distribution of the four main business activities (outer donut chart) as well as the sub-activities (inner donut chart) for each of the 3 years. We can observe the consistence of the activity and sub-activity distributions over the three years.

A. *Data preprocessing*

We applied the following preprocessing tasks in order to make the dataset ready for network visualization and analysis:

- Filtration: we selected licenses with at least one shared business partner exists with another license. Remaining licenses have been excluded.

TABLE I. The MAIN attributes of the business licneses dataset analyzed in this study.

Attribute Name	Data Type	Description
Issue Year	numeric	The year in which the license was issued.
License Category	text	The name of the authority that issued the license.
License No	numeric	A unique number for each license.
Legal Type	text	The legal type of the license (e.g., Limited Liability Company).
Actv Master Grp	text	The primary economic activity attached to each license (e.g., Commercial/ Professional).
Activity Category	text	A detailed description of the economic activity attached to the license.
Person Serial No	numeric	A unique identifier for each business partner
Partner Birth Date	date	The date of birth for each business partner.
Gender	text	The business partners gender (Female/Male).
Partner Nationality	text	The name of the country.
Person Category	text	The category of the business partner (Person/Body Corporate).
Partner Share	numeric	The ownership percentage as a decimal for the amount each partner has in the corresponding business.

- Generating new feature: for each license we computed “Number of partners” which specified the total number of business partners registered for the corresponding license (note that each record in the original data set connects a license to single partner).
- Encoding data as a network: A node represent a license and an edge connects two licenses if they share at least one partner. The weight of an edge corresponds to the number of shared partners between connected two licenses. TABLE V shows some summary statistics of the resulting network over the years. We can observe that most of the statistics are stable over the 3-year period, including the average degree, number of nodes, and number of edges.

Figure 4 shows the edge weight distribution for the licenses network over the different years. The lines in the chart are the corresponding LOESS regression. Interestingly, we can observe a pattern that is consistent over the 3-year period: a power-law distribution with a hump toward the distribution tail. While such pattern was reported in real world biological networks [20] we are not aware of any previous work that highlights similar distribution in social networks. The pattern highlights the nonzero probability of new edge being created within the existing nodes using preferential attachment (traditional power-law distribution assumed edges are created only when a new node is created).

B. *Visual Encoding*

Figure 5. illustrates the suggested network visual encoding. The network is undirected. The edge thickness encodes the weight (which is the number of business partners shared between the two connected licenses). The shape of a node encodes the master business activity color encodes the (sub)activity category(TABLE III). There are 16 different (sub)activity categories defined in the dataset. Each business license may attach with more than one. Using RGB color encoding, top three activity categories have been selected and mapped to one of the three colors: red, green, and blue. Black color indicates that the business license's activity category falls under categories other than the top three. The node size encodes to the number of business partner each license has. Finally, an edge color is a mix of the colors of the two connected nodes.

C. *Analysis*

ForceAtlas 2 layout have been used to explore the network visually.[1]

TABLE II. Node Shape encoding

Shape	Master Business Activity
Circle	Commercial
Triangle	Professional
Square	Industrial
Pentagon	Tourism

TABLE III. DATA VALUES AND INTIUITION OF COLOR COLUMN used to color the nodes of THE BUISNESS LICENSES network

Value (color hexadecimal code)	Color name	Activity category
#FF0000	**Red**	Trading & Services
#00FF00	**Green**	Real Estate,Renting,Bus Servic
#0000FF	**Blue**	Construction
#FFFF00	**Yellow**	Trading & Services + Real Estate
#FF00FF	**Purple**	Trading & Services + Construction
#00FFFF	**Cyan**	Real Estate + Construction
#000000	**Black**	Others
#FFFFFF	**White**	Trading & Services + Real Estate + Construction

After building the network using the transformed and preprocessed dataset in Gephi, we applied communities' detection algorithm to identify the patterns of discovered clusters.

a) *Graph exploration* After importing the edges and nodes lists, ForceAtlas 2 layout have been used to explore the initial state of the graph. The graph is consisting of multiple unconnected undirected sub graphs of different sizes. Since no filtration is needed, number of graph metrics have been calculated for each year dataset for better understanding as per shown in TABLE V.

TABLE IV. Summary statistics of the graphed datasets prepared for 2015, 2016, and 2017. The nodes represent business partners, and edges represent a connection of at least one trade license shared between two business partners.

Graph Statistics	Year		
	2015	**2016**	**2017**
Number of Nodes	20,237	20,178	20,212
Number of Edges	98,766	103,850	102,280
Network Diameter	14	11	11.46
Average Path length	1.87	1.65	1.97
Avg. Degree	9.76	10.29	10.12
Avg. Weighted Degree	12.097	15.8	11.46
Avg. Edge Weight	1.24	1.54	1.13
Avg. Clustering Coefficient	0.968	0.969	0.968

b) *Community detection* is the process of discovering the cohesive groups or clusters in the network [12]. The method we utilized to discover the hidden clusters is based on the modularity. Modurality involves maximizing the number of edges within a community while minimizing the number of edges across communities [19].

The community detection algorithm has been executed for each year separately. Leiden algorithm detected 10,486 clusters for the year 2015 (with quality measure of 0.994), 10,504 clusters for the year 2016 (quality measure of 0.995), and 10,983 clusters for the year 2017 (quality measure of 0.993). For meaningful visualization and further analysis, we focus on relatively big clusters with 0.20% or more of network's nodes.

For the discovered communities to properly stand out, Force Atlas 2 layout[2] have been applied first followed by Fruchterman Reingold[3] layout. TABLE VI. provides details on the communities detected within the business partnerships network for 2015.

TABLE V. The top seven largest communities detected within the business licenses network for year 2015. The percentage of the number of nodes that form each discovered community is listed in descending order such that cluster #1 is the largest in terms of the number of nodes forming it.

Cluster #	Number of Nodes		Avg. Betweenness Centrality	Avg. Clustering Coefficient
	Percentage	Count		
1	0.62%	126	28.71	0.99
2	0.45%	91	63.16	0.88
3	0.33%	67	46	0.97
4	0.33%	67	21.73	0.94
5	0.29%	58	0	1
6	0.27%	55	31.54	0.88
7	0.25%	50	191.16	0.92

Figure 6. The largest seven communities detected in year 2015. Nodes represent the business licenses and edges reflect shared business parter(s) between the connected nodes/licenses. Node shape encodes master business activity type while the node color encodes the sub activity. Node size reflects the number of business partners associated with the corresponding license. Edge thickness encodes the number of shared business partner(s) betweentwo nodes. The clusters take one of three motifs, a clique (such as Cluster # 3 and Cluster # 5),, a clique connected with sub-graph (such as Cluster # 1, Cluster # 4, Cluster #

6, and Cluster # 7), and multiple less dense subgraphs linked together through one or more nodes (such as Cluster # 2).

Figure 6. Shows the top seven clusters, with each cluster containing at least 0.2% of the total number of nodes. We identified three motifs in the licenses network: 1) a clique (Cluster 3 and Cluster 5), 2) a clique connected with subgraph(s) (Clusters 1,4,6, and 7), and 3) a core-periphery structure in which multiple less dense graphs are connected together through one or more nodes (Cluster 2). Of the 3 motifs, we observed that the clique was the most common over the study period (2015–2017). The diversity of the clusters in terms of both master activities and (sub)activities is worth noting (as highlighted by the different node colors and shapes in Clusters 1, 3, 4, 5, 6, and 7). In fact, only Cluster 2 was mostly homogeneous in terms of commercial activities. This means that many investors prefer to diversify their businesses instead of focusing on one type of commercial activity. TABLE VII further quantifies the commercial activity diversity of the top clusters for the year of 2015.

TABLE VI. The main characteristics of the top seven clusters identified within the business licenses network data of year 2015

Cluster #	Activity Master Category				Avg # of Partners
	Commercial	Professional	Industrial	Tourism	
1	65.87%	33.33%	0.79%	0	3.65
2	97.8%	2.2%	0	0	7.82
3	55.22%	44.78%	0	0	3.64
4	94.03%	2.99%	0	2.99	3.72
5	72.41%	25.86%	1.72%	0	3.78
6	54.55%	40%	1.82%	3.64%	3.87
7	86%	14%	0	0	4.16

In same manner, business licenses network of years 2016 and 2017 have been analyzed. For the year 2016, 8 clusters had more than 0.20% of the nodes (licenses). The three motifs we identified in 2015 can also be observed in the clusters of the year 2016 (Figure 7.) and the year 2017 (Figure 8.). Also the diversity we observed in 2015, remain consistent in 2016 (TABLE VIII and TABLE IX.) and 2017 (TABLE X. and TABLE XI.)

TABLE VII. The top eight largest communities detected within the business licenses network for year 2016. The percentage of the number of nodes that form each discovered community is listed in descending order such that cluster #1 is the largest in terms of the number of nodes forming it.

Cluster #	Number of Nodes		Avg. Betweenness Centrality	Avg. Clustering Coefficient
	Percentage	Count		
1	0.53%	106	0	1
2	0.34%	69	0	1
3	0.31%	63	78.07	0.82
4	0.28%	57	8.39	0.93
5	0.28%	56	35.69	0.89
6	0.24%	48	0	1
7	0.23%	47	0	1
8	0.23%	46	23.11	0.97

TABLE VIII. The main characteristics of the top eight clusters identified within the business licenses network data of year 2016

Cluster #	Activity Master Category				Avg # of Partners
	Commercial	Professional	Industrial	Tourism	
1	74.53%	23.58%	1.89%	0	3.92
2	100%	0	0	0	6.29
3	49.21%	50.79%	0	0	3.1
4	94.74%	3.51%	0	1.75%	3.56
5	98.21%	1.79%	0	0	7.52
6	62.5%	37.5%	0	0	3.48
7	97.87%	2.13%	0	0	4.96
8	82.61%	13.04%	0	4.35%	3.67

TABLE IX. The top nine largest communities detected within the business licenses network for year 2017. The percentage of the number of nodes that form each discovered community is listed in descending order such that cluster #1 is the largest in terms of the number of nodes forming it.

Cluster #	Number of Nodes		Avg. Betweenness Centrality	Avg. Clustering Coefficient
	Percentage	Count		
1	0.62%	125	4.96	0.99
2	0.31%	63	51.87	0.99
3	0.31%	62	3.84	0.95
4	0.27%	54	90.56	0.91
5	0.27%	54	110.52	0.81
6	0.26%	52	10.98	0.97
7	0.25%	51	8.20	0.97
8	0.24%	49	13.96	0.99
9	0.23%	46	2.83	0.96

TABLE X. The main characteristics of the top nine clusters identified within the business licenses network data of year 2017

Cluster #	Activity Master Category				Avg # of Partners
	Commercial	Professional	Industrial	Tourism	
1	68.8%	30.4%	0.8%	0	3.81
2	68.25%	31.75%	0	0	3.65
3	56.45%	43.55%	0	0	3.55
4	70.37%	16.67%	1.85%	11.11%	3.26
5	79.63%	14.81%	5.56%	0	5
6	53.85%	46.15%	0	0	3.08
7	76.47%	21.57%	1.96%	0	3.63
8	63.27%	34.69%	0	2.04%	3.43
9	80.43%	19.57%	0	0	3.59

1 The tool used in the explanatory data analysis is Gephi [20]

2 https://github.com/gephi/gephi/wiki/Force-Atlas-2.

3 https://github.com/gephi/gephi/wiki/Fruchterman-Reingold

V. Business Partners Network Analysis

The data preprocessing under this part included the following tasks:

- Filtration: only individual partners were considered (to allow us to study demographic diversity), while partners who are corporations were excluded. Also, licenses with single business owners were excluded (since they will represent single disconnected nodes). Furthermore, only well connected nodes (have more than the overall average degree), and heavy-weight edges (have weight more than the overall average edge weight) were retained and analyzed further.
- Missing Values: Missing values were identified in the business partner date of birth and gender fields. Since the count of records with missing values was small, these records were simply eliminated.
- Feature Extraction: Additional fields were added to the dataset to enhance the analysis results, including age of the business partner in relation to the license year of issuance.
- Adding new features: two new features were added to the dataset describing the longitude and latitude of each country corresponding to the business partners' nationalities. The data for longitude and latitude fields were obtained from the Harvard WorldMap dataset[4]. A mapping between the partner's nationality field and corresponding country name from the Harvard WorldMap dataset provided accurate longitude and latitude values for each country listed in the business licenses datasets. The longitude and latitude fields have been used to visualize diversity of business partners' nationalities on a world map, as shown in Figure 14.
- Encoding data as a network: In this part of our analysis we encode every partner as a node, and an edge exists between two partners if they are partners in at least one license. The weight of each edge corresponds to the number of licenses that the connected partners/nodes share as per Figure 12. TABLE XII. shows some summary statistics of the network over the years. Figure 11. shows the edge weight distribution for the partners' network for the different years. The lines in the chart are the corresponding LOESS regression. Similar to Figure 4. , we also observe a pattern that is consistent over the 3-year period: a power-law distribution with a hump toward the distribution tail.

TABLE XI. Summary statistics of the graphed datasets prepared for 2015, 2016, and 2017. the nodes represent business partners, and edges represent a connection of at least one trade license shared between two business partners

Year	Count		Average			
	Nodes	*Edges*	*Degree (AD)*	*Weighted Degree (AWD)*	*Edge Weight (AEW)*	*Clustering Coefficient*
2015	33,650	43,333	2.576	2.921	1.134	0.905
2016	31,434	37,988	2.417	2.988	1.236	0.899
2017	28,886	33,937	2.35	2.681	1.141	0.894

As discussed earlier, nodes represent partners and edges encode sharing one or more licenses among partners. Edge thickness reflects the edge weight, which is the number of licenses shared between the two linked partners The node size is proportion to the betweenness centrality value of each node, which highlights the key partners within each cluster.

c. Analysis

We applied the community detection algorithm and focused on clusters that contained 0.12% or more of the nodes/partners. Nodes with highest betweenness centrality are assumed to be the ones that contribute most to the connectivity between the other nodes within the network. Figure 13. Shows the top clusters discovered in for the year of 2015. We observe here two of the 3 motifs we identified in the first part of our analysis: a clique (which represents a close-knit of business partners), and core-periphery structure in which multiple sub-graphs connected together through a dominant node (which highlights the existence of critical partners who link multiple groups of business that would be otherwise disconnected).

We then analyzed the diversity of the different clusters with respect to gender, age, and nationality. TABLE XIII. shows the main statistics of the three demographics for the year of 2015. There is diversity with respect to age and nationality, and not as much diversity when it comes to gender. Four out of the seven biggest clusters have no female members. Even for the clusters that do have female members, the percentage is no more than 12%. Figure 14. projects business partnerships on a world map based on nationality for enhanced visual perception. The figure reveals the diversity of business partners nationality especially for cluster 1 and cluster 3. Also, in the prominent-member type of clusters, the nationality of the prominent member is different from most common nationality of the cluster except for cluster number 7.

To investigate the evolution of partnerships over the years, we traced the same nodes(partners) that formed the top seven clusters of the year 2015, across the following two years

TABLE XII. The main characteristics of the top seven clusters identified within the business partnerships network data of year 2015

Cluster #	Count		Age		Nationality characteristic
	Female	Male	Avg	STDev	
1	3	26	43.2	7.1	Multinational
2	0	20	43.7	8.9	India/Oman/UK
3	0	17	42.6	6.7	Multinational
4	1	15	52.6	10.7	India/UAE/UK
5	0	15	50.5	9.2	India
6	1	14	62	14.7	UAE/UK/Turkey
7	0	14	53.2	11.1	Multinational

Figure 15. displays the seven clusters detected across the three years. The cluster structures are similar over the three years. However, some clusters show increased or decreased business activities. For example, the business partners of cluster 6 participated in an increased number of new business licenses in 2016 compared to 2015 and 2017. The partners of cluster 5 experiences fewer business activities in 2017 compared to 2015 and 2016, and such behavior is observed through the changing in edge thickness connecting these partners within the cluster. Finally, a world map is utilized again to visualize the nationalities of business partners linked together within each cluster between 2016 and 2017 as per in Figure 16. and Figure 17.

[4] https://worldmap.harvard.edu/data/geonode:country_centroids_az8.

Vi. Conclusion

This study proposed a novel approach in analyzing business relationships. The application of social network analysis techniques to the business licenses and partners dataset, showcases a successful utilization of SNA in the economic domain. The work has been divided into two main parts. The first part analyzed the licenses network, where nodes represent business licenses and edges connect licenses that share partners. The analysis identified 3 motifs that are common in the detected clusters of licenses: 1) clique, 2) core-periphery, and 3) clique-subgraph. The majority of the clusters showed significant diversity in terms of business activities.

Such information can help subject matter experts and decision makers in setting the policies and regulations concerned with issuing business licenses and managing the economic business types. It's worth mentioning that our dataset includes limited attributes which in turn limits our findings and recommendations.

The second part of the paper analyzed the partners network, where nodes represent partners and edges connect partners that share one or more licenses.

The top seven largest clusters in 2015 were identified and traced over 2016 and 2017. Two distinct motifs were identified within the discovered clusters: a clique and a core-periphery structure. The clustering coefficient was relatively high over the three years, and the network diameter was small, which are common measures for naturally occurring graphs [5]. Also, high modularity and clustering coefficient values determined over the three years indicated and verified the presence of community structures within the business partnerships network [11]. In general, the seven clusters identified in 2015 maintained the same structure into 2016 and 2017. It worth mentioning that clique and core-periphery motifs are claimed to be the most common network structures discriminating social networks which is another interesting finding of this work [5]. We also observed partners diversity when it comes to age and nationality but not when it comes to age. This specific observation highlights the need for policies that address this lack of gender diversity.

Declarations

Acknowledgment

We thank Department of Economy and Tourism in Dubai for providing business partnerships data and Dubai Statistics Center for their support.

References

1. Ward, S., (2018). Why Business Partnerships Fail. The Balance Small Business [online]. [Accessed 09 November 2018]. Available at: https://www.thebalancesmb.com/why-business-partnerships-fail-4107045
2. Neville, A., (2013). Why Partnership Is Harder Than Marriage. Forbes [online]. [Accessed 09 November 2018]. Available at: https://www.forbes.com/sites/amandaneville/2013/03/01/why-partnership-is-harder-than-marriage/#55abb6e87ec9.
3. Kapadia, Y. (2016). Dubai most cosmopolitan city in world; 83% foreign residents. Khaleej Times. [online] 16 January. [Accessed 10 November 2018]. Available at: https://www.khaleejtimes.com/nation/dubai/dubai-most-cosmopolitan-city-in-the-world.
4. Gulf News. (2018). UAE promotes tolerance and respect for others. 23 January. [Accessed 10 November 218]. Available at: https://gulfnews.com/opinion/editorials/uae-promotes-tolerance-and-respect-for-others-1.2161752.
5. Chakrabarti, D. and Faloutsos, C., (2006). Graph mining: Laws, generators, and algorithms. ACM computing surveys (CSUR), 38(1), p.2.
6. Glaeser, E.L., Kerr, S.P. and Kerr, W.R., (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. Review of Economics and Statistics, 97(2), pp.498–520.

7. Hartmann, P.M., Zaki, M., Feldmann, N. and Neely, A., (2016). Capturing value from big data–a taxonomy of data-driven business models used by start-up firms. International Journal of Operations & Production Management, 36(10), pp.1382–1406.
8. Hochsztain, E., Tasistro, A. and Messina, M., (2015), September. Data Mining applications in entrepreneurship analysis. In Data Mining with Industrial Applications (DMIA), 2015 International Workshop on (pp. 25–29). IEEE.
9. Han, J., Pei, J. and Kamber, M., (2011). Data mining: concepts and techniques. Elsevier.
10. Pushpa, S., 2012. An efficient method of building the telecom social network for churn prediction. International Journal of Data Mining & Knowled Management Process, 2(3), pp.31–39.
11. Newman, M.E., (2006). Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23), pp.8577–8582.
12. Bedi, P., & Sharma, C. (2016). Community detection in social networks. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 6(3), 115–135.
13. Abdulkarim, R., & Abdallah, S. (2019, October). Using social network analysis to study diversity in business partnerships. In 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 126–133). IEEE.
14. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10), P10008.
15. Goel, R., & Sharma, R. (2021). Studying leaders & their concerns using online social media during the times of crisis-A COVID case study. Social Network Analysis and Mining, 11(1), 1–12.
16. Singh, M., Jakhar, A. K., & Pandey, S. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. Social Network Analysis and Mining, 11(1), 1–11.
17. Gan, C., Voda, M., Wang, K., Chen, L., & Ye, J. (2021). Spatial network structure of the tourism economy in urban agglomeration: A social network analysis. Journal of Hospitality and Tourism Management, 47, 124–133.
18. Valeri, M., & Baggio, R. (2021). Italian tourism intermediaries: A social network analysis exploration. Current Issues in Tourism, 24(9), 1270–1283.
19. Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Scientific reports, 9(1), 1–12.
20. Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Figures

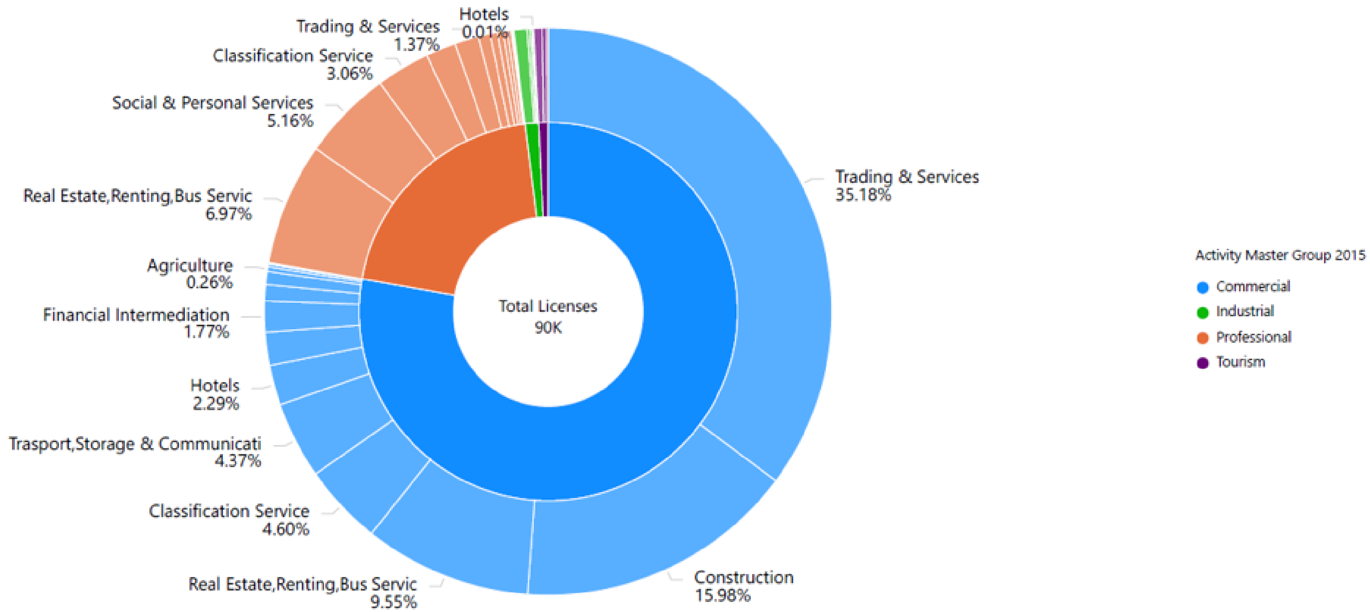

Figure 1

Distribution of master business activities and activity category within business licenses dataset of year 2015

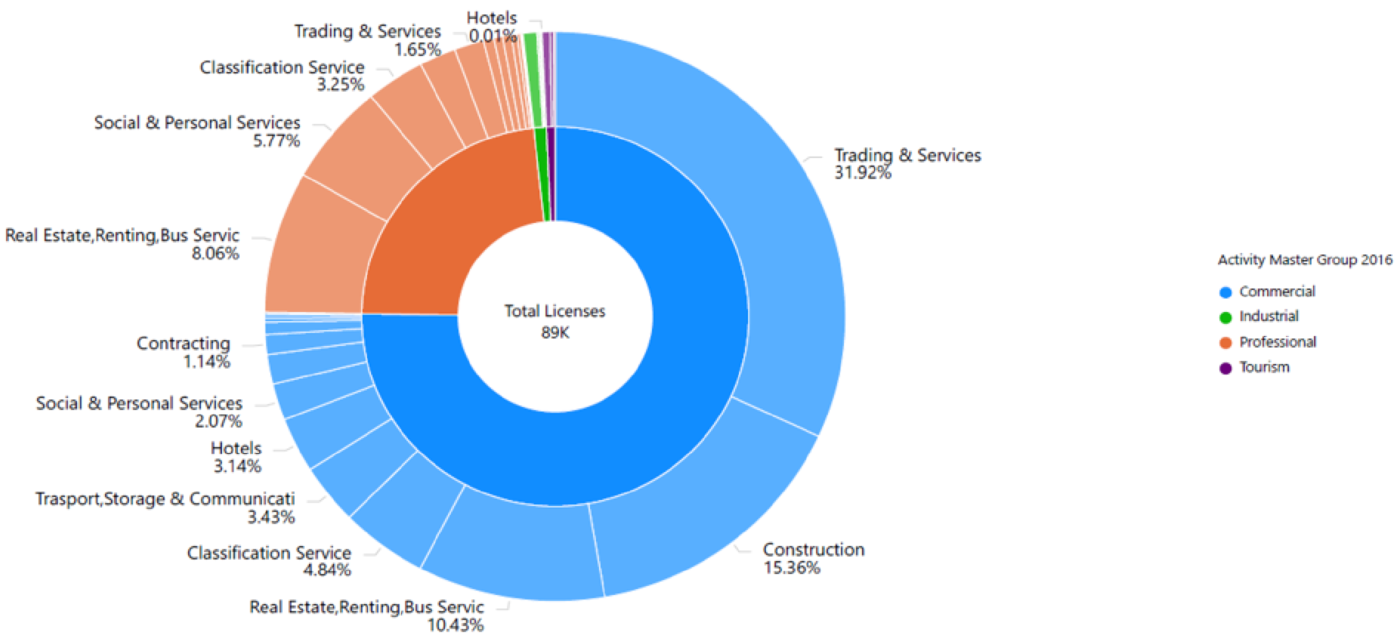

Figure 2

Distribution of master business activities and activity category within business licenses dataset of year 2016

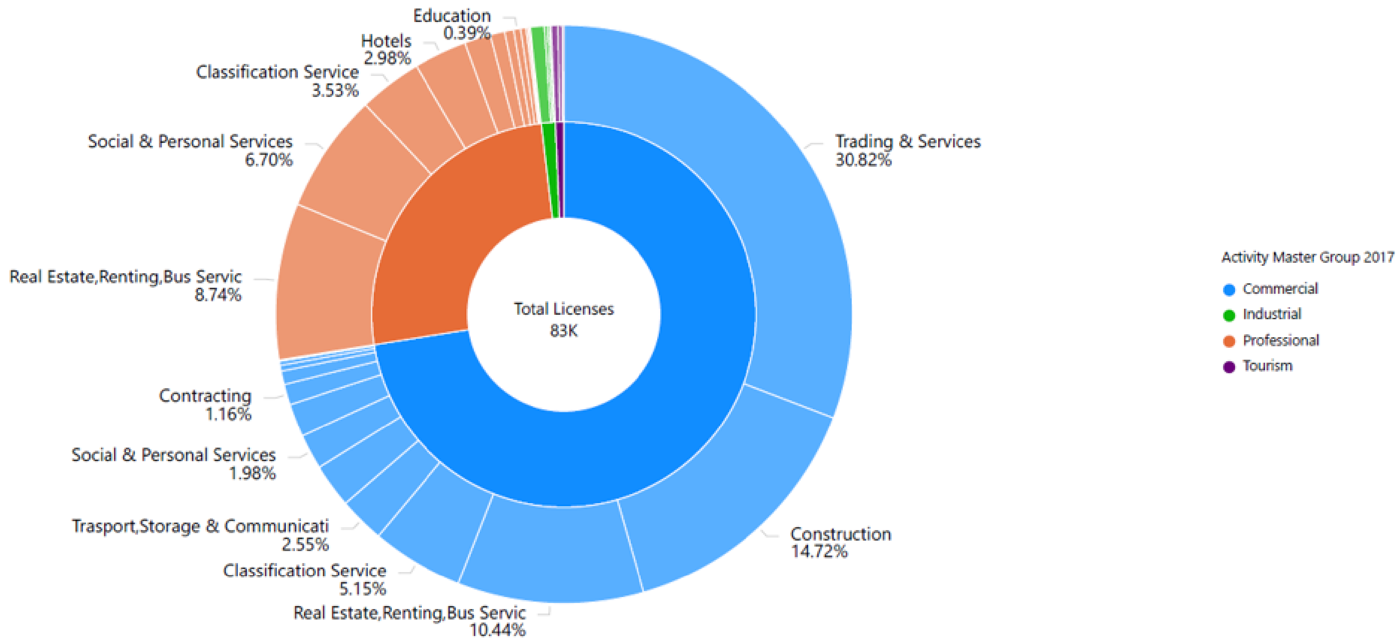

Figure 3

Distribution of master business activities and activity category within business licenses dataset of year 2017

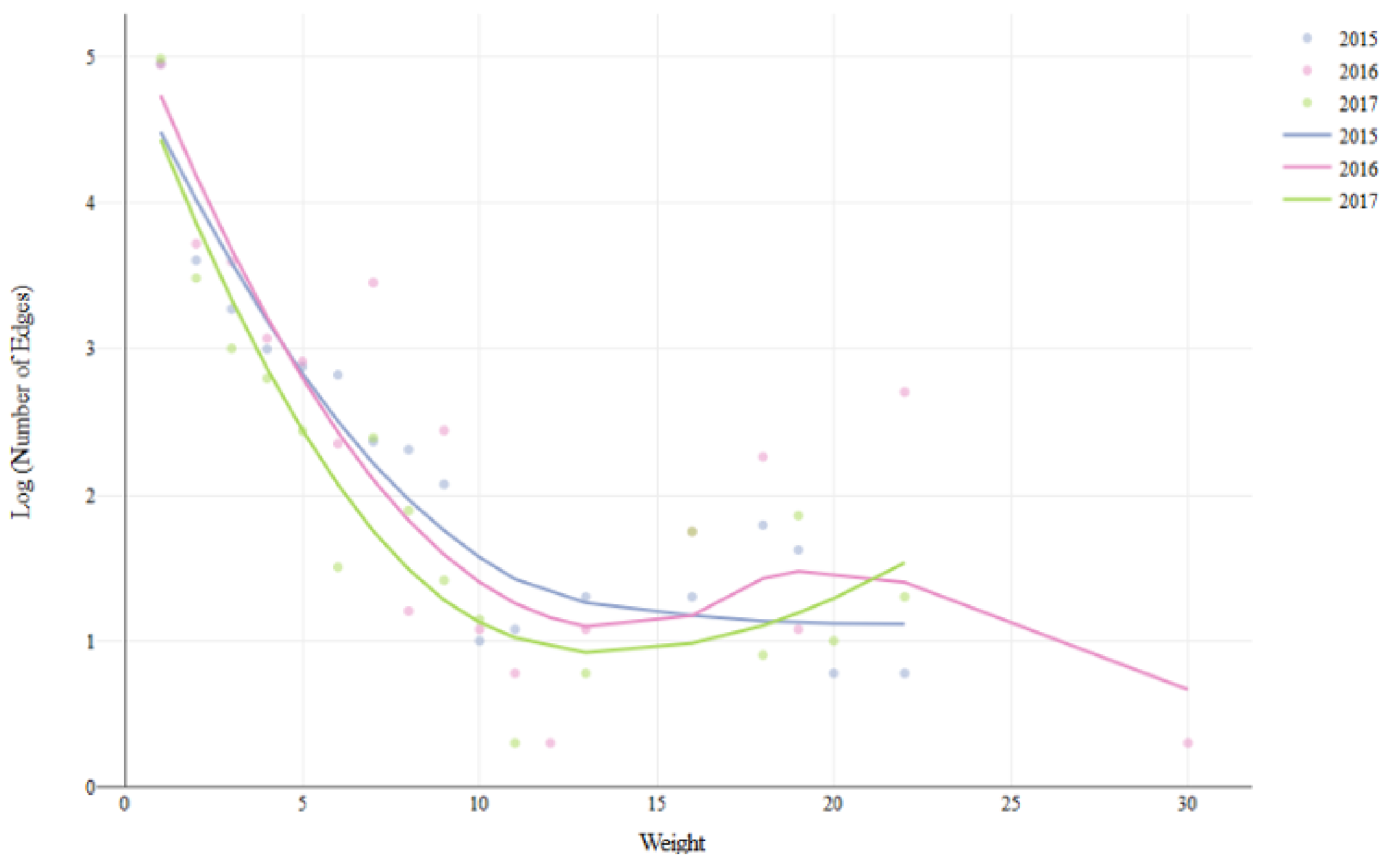

Figure 4

Edges Weight distribution within licenses networks over the years 2015, 2016, and 2017.

Figure 5

License-License network structure proposed for constructing business partnerships as a graph

Figure 6

The largest seven communities detected in year 2015. Nodes represent the business licenses and edges reflect shared business parter(s) between the connected nodes/licenses. Node shape encodes master business activity type while the node color encodes the sub activity. Node size reflects the number of business partners associated with the corresponding license. Edge thickness encodes the number of shared business partner(s) betweentwo nodes. The clusters take one of three motifs, a clique (such as Cluster # 3 and Cluster # 5), , a clique connected with sub-graph (such as Cluster # 1, Cluster # 4, Cluster # 6, and Cluster # 7), and multiple less dense subgraphs linked together through one or more nodes (such as Cluster # 2).

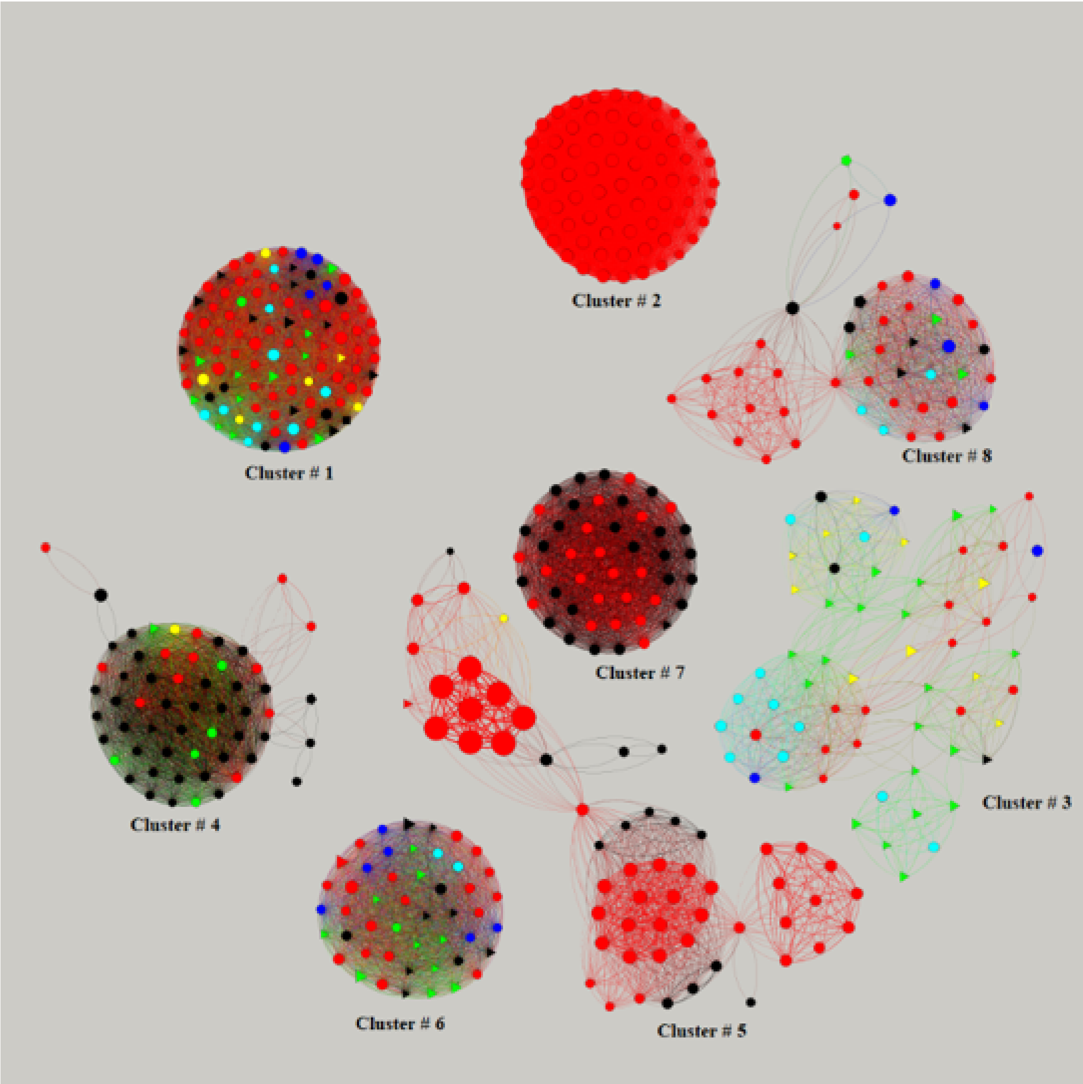

Figure 7

The largest communities detected in year 2016. Similar to the clusters identified in year 2015, the clusters take one of three motifs, a clique like (Clusters 1,2,4,6, and 7), a clique connected with sub-graph (Cluster 8), and multiple less dense subgraphs linked together through one or more nodes(Clusters 3 and 5)

Figure 8

Top clusters detected in year 2017. Similar to the clusters identified in years 2015 and 2016, the clusters take one of three three motifs, a clique (Clusters 1,2,3, and 9), a clique connected with sub-graph (Clusters 6,7, and 8), and multiple less dense subgraphs linked together through one or more nodes (Clusters 4 and 5).

Figure 9

Distribution of business parnters in terms of age and gender over the three years of anaylsis 2015,2016, and 2017. Young business partners denotes the effect of inheritance and/or family businesses.

Figure 10

Distribution of business partners in terms of top 30 nationalities over the three years of analysis 2015,2016, and 2017.

Figure 11

Edges Weight distribution within partners networks over the years 2015, 2016, and 2017.

Figure 12

Parnter-Partner network structure proposed for constructing business partnerships as a graph

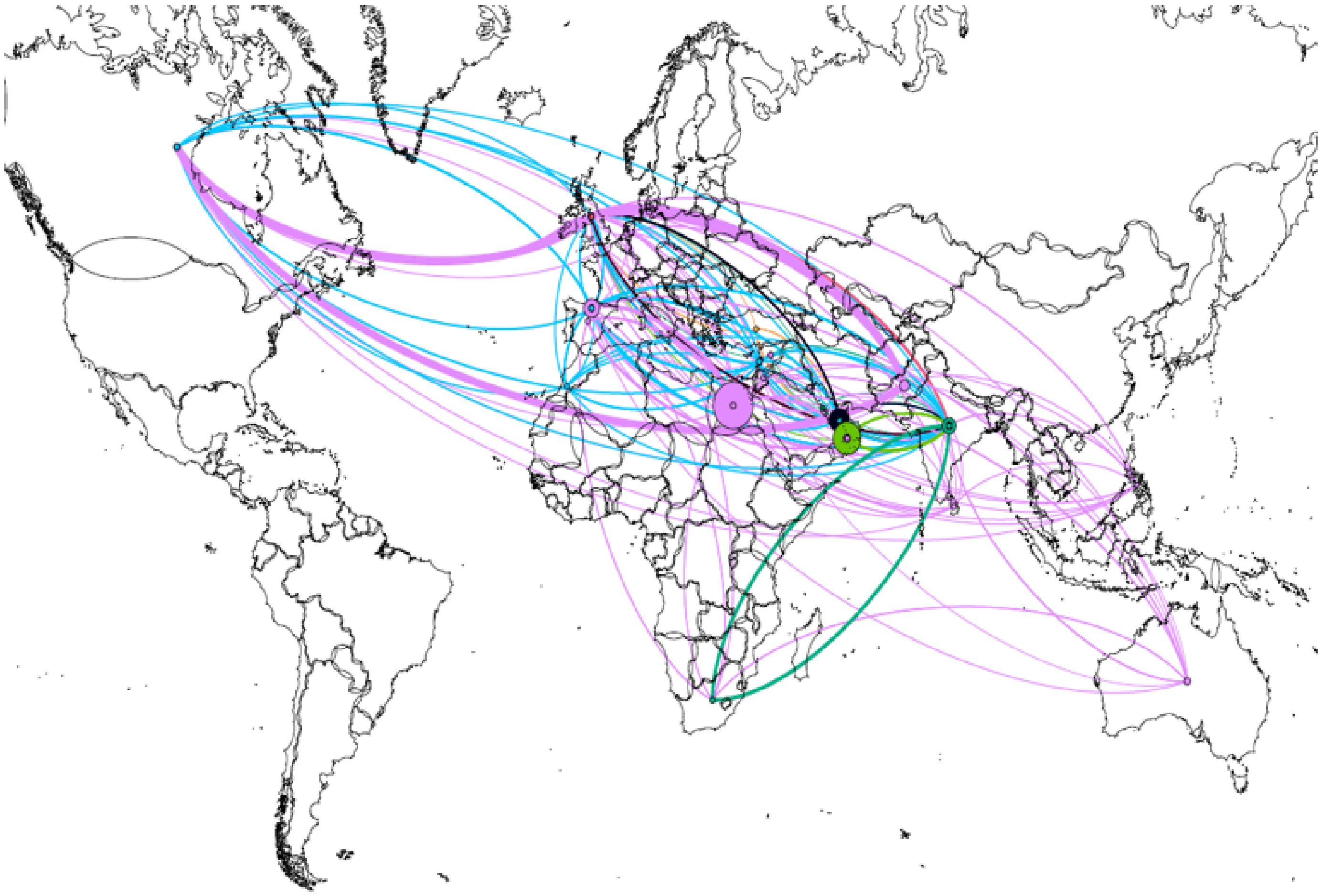

Figure 14

A world map showing the top seven largest business partner clusters discovered within the business partnership dataset for 2015. Each is labelled with a unique colour where the business partner's nationality is projected to the corresponding country.

Figure 15

The top seven communities detected in the business partnerships network over (a) 2015, (b) 2016, and (c) 2017. The clusters maintained their structure over the three years. However, a variation of the node sizes and edge thickness identify change affecting each cluster during the three years of analysis. Cluster 3, for example, shows increased partnerships among its members in 2017 comparing to 2015 and 2016

Figure 16

A world map showing the business partnerships network for each cluster discovered in 2016. The node size is proportional to the betweenness centrality, and edge thickness is proportional to the number of business partnerships established between two partners

Figure 17

A world map showing the business partnerships network for each cluster discovered in 2017. The node size is proportional to the betweenness centrality, and edge thickness is proportional to the number of business partnerships established between two partners.