東京大学　関本研究室 / Sekimoto Lab. IIS, the University of Tokyo.

GNN-based Core Company Identification via Real-World Data of Commercial Flows

Hideaki KANO, Yoshiki OGAWA, Soohyun JOO, Yoshihide SEKIMOTO

Backgrouds

Important to identify the core companies in the distribution channel in order to build a resilient supply chain

What is Core Company Identification?

A new concept of Core Company was defined to encompass both bottleneck firms and connector hub firms, Detecting companies that play an important role in business-to-business transactions using the Graph Neural Network framework.

Specific things we aim to achieve

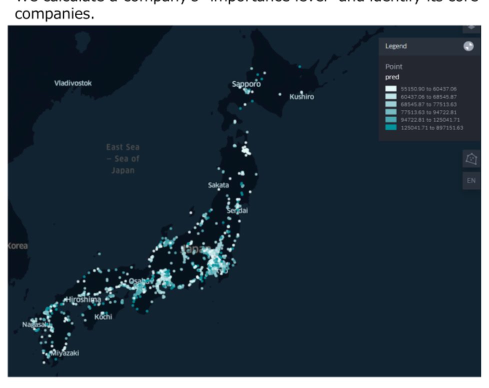
Community Detection

Based on the structure of trading networks throughout Japan
Divide the network to maximize cohesion within the community

Comparison of Hokkaido Community and Aichi Community

Detection of core companies

Geographic/financial/transactional as well as community considerations, We calculate a company's "importance level" and identify its core companies.



Geographic Distribution of Top Scoring Firms

Prior Research | Detecting Bottleneck Companies

Based on business-to-business transaction data and company profile data
Identifies bottleneck companies by supervised learning on table data

A model for extracting bottleneck firms in the supply chain [Ogawa et al. 2022].

Business-to-Business Transaction Data

By about 840,000 companies obtained from GIMT (2019)
business-to-business transaction data

Bottleneck Company Data

28 companies were selected based on a survey conducted by Nikkei Shimbun (2012).

Features to be utilized for learning

The following is prepared from transaction data and summary information
- Basic company attributes (sales, capital, number of employees)
- Centrality of supply chain network
- Indicators for local network structure

Classification model

Training and classification with multiple machine learning models
- Logistic regression
- Support Vector Machines
- Random Forest

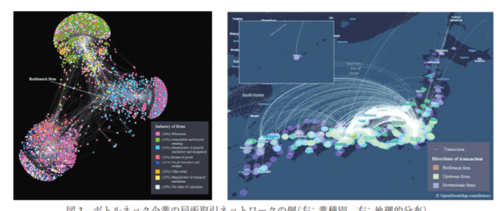


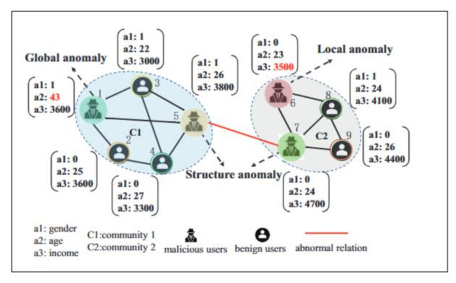
図3 ボトルネック企業の局所取引ネットワークの例(左: 業種別, 右: 地理的分布)

Prior Research | Graph Anomaly Detection

Recent progress in detecting singular nodes using Graph Neural Networks
Singularity is detected by considering not only node characteristics but also graph connectivity and community.

Community-focused singular node detection

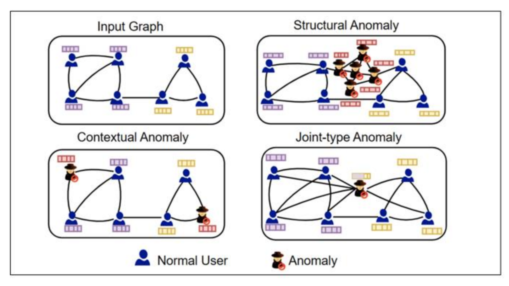
Considering the community information of the graph
Detect nodes with specific features



Luo et al. (WSDM'22)

Singular node detection focusing on graph structure

Not only nodes with singular features,Detecting singular nodes by considering the connectedness of the graph



Roy et al. (WSDM'24)

Contributions

Considering the distribution channel as a graph, the core companies are identified using GNN based on data on business-to-business transactions across Japan

Data | Utilized Data and Data Processing

Features are created for each company based on NIHACHI and ZAIMUH data, and used to train GNN models.

Data to be utilized

NIHACHI Data

Utilizes 4,528,642 annual transaction data for all of Japan in 2021
620,068 companies

Ordering Company CD	Ordering Company Corporate CD	Estimated trading volume
ABC001	XYZ001	291.24
ABC002	XYZ001	34.79
...	...	...

ZAIMUH Data

Utilize the following data from 802,017 companies listed in ZAIMUH
Number of Employees/Total Capital/Equity/Sales/Recurring Revenue/Operating Expenses/
SG&A Expenses / Advance Payables
Increase/Decrease/Inventory Increase/Decrease/Capital

Model Input/Output

Model Input

Company Code	joint feature	Financial Features	Community feature value
ABC001	(1, 0, 49, ...)	(10000, 132, ...)	0
ABC002	(3, 1, 22, ...)	(3000, 212, ...)	24
...	...	...	...

Model output

Company Code	score
ABC001	4.5
ABC002	5.9
...	...

Graph Neural Network | Core Company Identification Framework

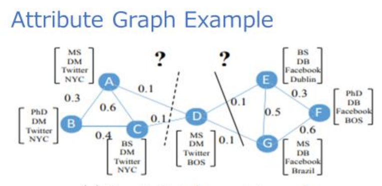
Intercompany transactions are viewed as a graphical problem, detecting community and core companies.

Formulation as a GNN problem

Business-to-business transactions as graphs

Think of business-to-business transactions as an attribute graph
- Consider node as a company and consider the following
- Consider the edge as a business-to-business transaction as follows
- The attribute associated with the node is considered to be a characteristic of the company and Consider the following

Attribute Graph Example

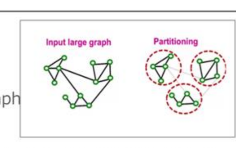


(a) An attributed uncertain graph

Main Modules

Community Detection

Partitioning a graph using METIS



This makes the size of the partitioned graph size is approximately equal, and

the number of edges between subgraphs is optimized to be minimal,
The number of edges between subgraphs is optimized to be minimal.
In other words, maximum transactions in the community,
and divide them in such a way as to minimize inter-community transactions

Detection of core companies

The following values are learned as objective variables for each company, and treat the predicted value as a score representing the level of importance.

- $\mu_{i取引社数} = \Sigma_{j\neq i}(1[受注取引_{i,j}]) + \Sigma_{j\neq i}(1[発注取引_{i,j}])$
- $\mu_{i取引額} = \Sigma_{j\neq i}(受注取引額_{i,j}) + \Sigma_{j\neq i}(発注取引額_{i,j})$
- $\mu_{iコア} = \{\Sigma_{j\neq i}(1[受注取引_{i,j}]) + \Sigma_{j\neq i}(1[発注取引_{i,j}])\} \times \{\log(\Sigma_{j\neq i}(受注取引額_{i,j})) + \log(\Sigma_{j\neq i}(発注取引額_{i,j}))\} \times \{\Sigma_{j\neq i}(1[異なるコミュニティ_{i,j}])\}$

Algorithms

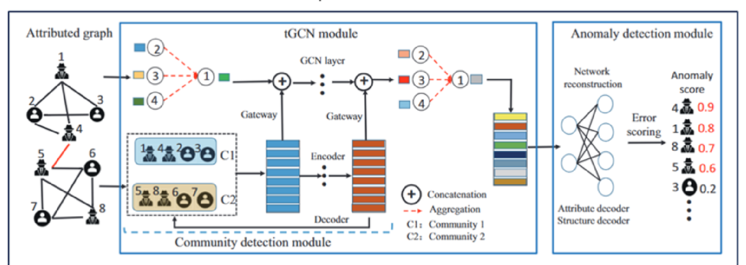
After detecting communities from business flows, the higher the score for the number of transactions / the number of companies / the number of communities involved in transactions, the higher the score.

Graph Neural Network | Community Detection

Computational difficulties with the methods of previous studies when the entire Japan is the target of calculations. Therefore, focusing on the fact that bottlenecks and connector hubs are at the center of the network, communities are detected

Challenges in dealing with data on a national scale in Japan - Computational complexity

In the method in the previous study,
Calculate the community by a calculation based on the adjacency matrix However, this method is computationally expensive, difficult to calculate using NIHACHI, which is a nationwide transaction in Japan.

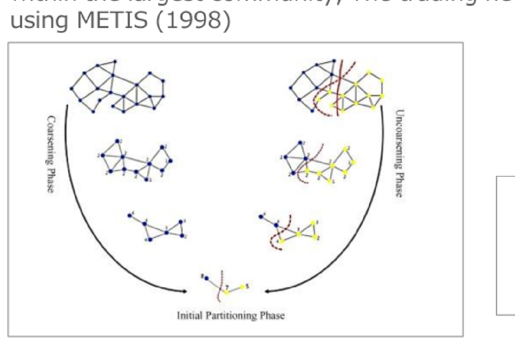


	ground-truth anomaly		injected anomaly					
Datasets	Amazon	Enron	Cora	Citeseer	Pubmed	BlogCatalog	Flickr	ACM
nodes	1,418	13,533	2,708	3,327	19,717	5,196	7,575	16,484
edges	3,695	176,897	5,429	4,732	44,338	171,743	239,738	71,980
attributes	28	20	1,433	3,703	500	8,189	12,074	8,337
anomalies	28	5	150	150	600	300	450	600

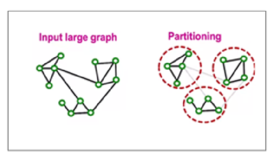
Luo et al. (WSDM'22)

Graph partitioning using METIS

Therefore, bottleneck companies and connector hub companies Focusing on the fact that the company exists at the center of the trading network between the smallest community - to be within the largest community, The trading network is partitioned using METIS (1998)



Sketch of partitioning a graph by METIS

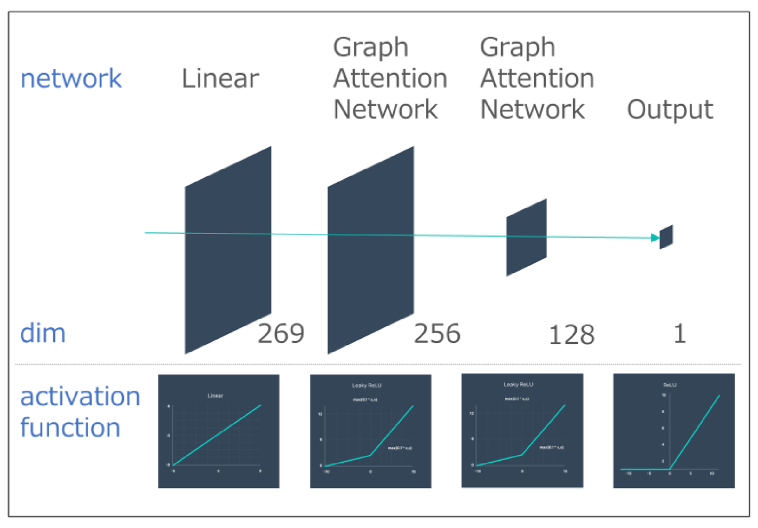


MEITS recursively traverses the graph,partitioning the graph to minimize the number of edge cuts.

Graph Neural Network | Detection of Core Companies

Learns scores and extracts core companies using GNN architecture that can learn linearity/nonlinearity

GNN architecture to detect core companies



To prevent gradient loss in the activation function, Leaky ReLU is used in the intermediate layer

Types of Scores for Core Company Detection

The following values are learned as objective variables for each company, and treat the predicted value as a score representing the level of importance.

Score that considers companies with a large number of order-receiving companies as important.

- $\mu_{i取引社数} = \Sigma_{j\neq i}(1[受注取引_{i,j}]) + \Sigma_{j\neq i}(1[発注取引_{i,j}])$

Score that considers companies with a large amount of transactions for receiving and placing orders as important.

- $\mu_{i取引額} = \Sigma_{j\neq i}(受注取引額_{i,j}) + \Sigma_{j\neq i}(発注取引額_{i,j})$

Score that considers companies with high transaction value/number of trading companies/trading communities for orders received and placed to be important

- $\mu_{iコア} = \{\Sigma_{j\neq i}(1[受注取引_{i,j}]) + \Sigma_{j\neq i}(1[発注取引_{i,j}])\} \times \{\log(\Sigma_{j\neq i}(受注取引額_{i,j})) + \log(\Sigma_{j\neq i}(発注取引額_{i,j}))\} \times \{\Sigma_{j\neq i}(1[異なるコミュニティ_{i,j}])\}$

Results

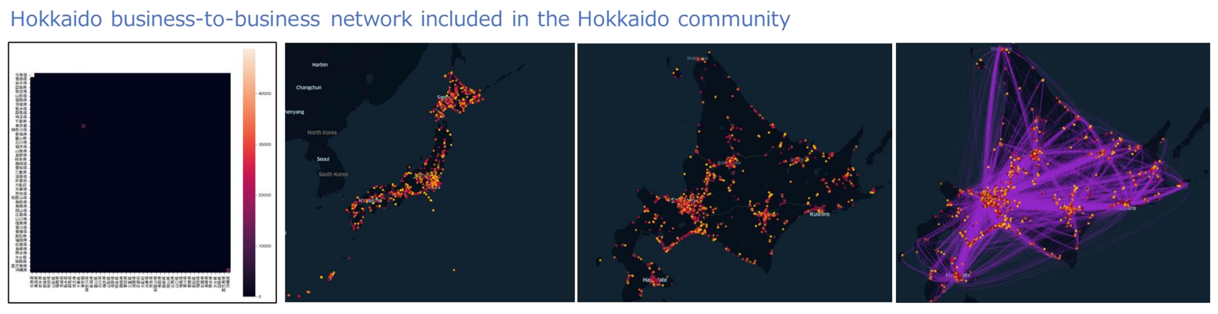
Confirmed that the detected community's regional characteristics and companies with high scores are at the core of the graph.

Geographic trends in b2b tx in each community

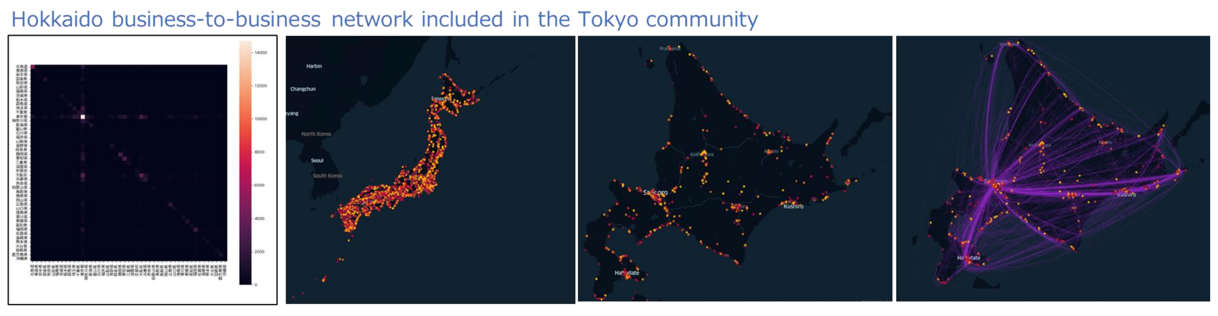
Local interpretation of each community

北海道	東京 大阪 愛知 九州	東京	東京 福井 愛知 大阪	東京 大阪
東京-大阪 大阪-東京	東京-大阪 大阪-東京	山梨	東京	東京 神奈川 大阪 福岡
大阪 東京 愛知 広島 福岡	神奈川	東京	富山 福井 石川 千葉 京都	福岡 熊本 長崎 大分
長野 静岡 愛知 岐阜	兵庫 大阪 岡山 四国	福島 新潟 宮城 （東北）	東京	東京
東京 大阪 東京-大阪 大阪-東京	大阪 東京-大阪 大阪-東京	東京 大阪	茨城	愛知 東京

Hokkaido business-to-business network included in the Hokkaido community

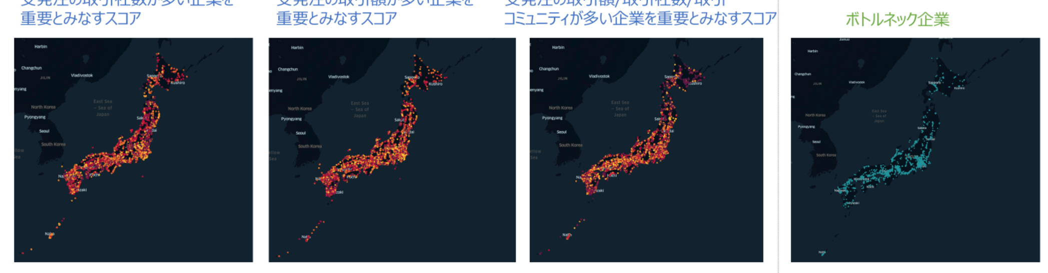


Hokkaido business-to-business network included in the Tokyo community



Experiments and Discussions | Detection of Core Companies

Companies with each score in the top 95% were selected. Companies with higher scores considering the community confirmed that they do business with many companies



	受発注の取引社数が多い企業を重要とみなすスコア			受発注の取引額が多い企業を重要とみなすスコア			受発注の取引額/取引社数/取引コミュニティが多い企業を重要とみなすスコア			ボトルネック企業		
	取引社数	取引額	コミュニティ数	取引社数	取引額	コミュニティ数	取引社数	取引額	コミュニティ数	取引社数	取引額	コミュニティ数
"01%"	2	25.6	2	2	20.1	2	4	27.1	3	0	0.0	0
"05%"	5	87.0	3	5	75.0	3	7	81.9	4	0	0.0	0
"25%"	14	432.6	7	13	405.6	6	17	408.1	8	5	186.5	3
中央値	24	1481.7	10	24	1544.8	10	28	1460.5	11	22	1584.2	9
"75%"	45	6926.4	15	48	7548.6	16	51	6464.1	16	51	9514.4	16
"95%"	225	89456.3	32	239	94741.9	33	228	76221.1	33	288	117295.2	36
"99%"	828	468066.4	47	912	555284.3	48	853	488163.8	47	1035	708265.4	49