Application of Hodge decomposition to money flow among firms' bank accounts

Yoshi Fujiwara

Graduate School of Information Science
University of Hyogo, Kobe, JAPAN

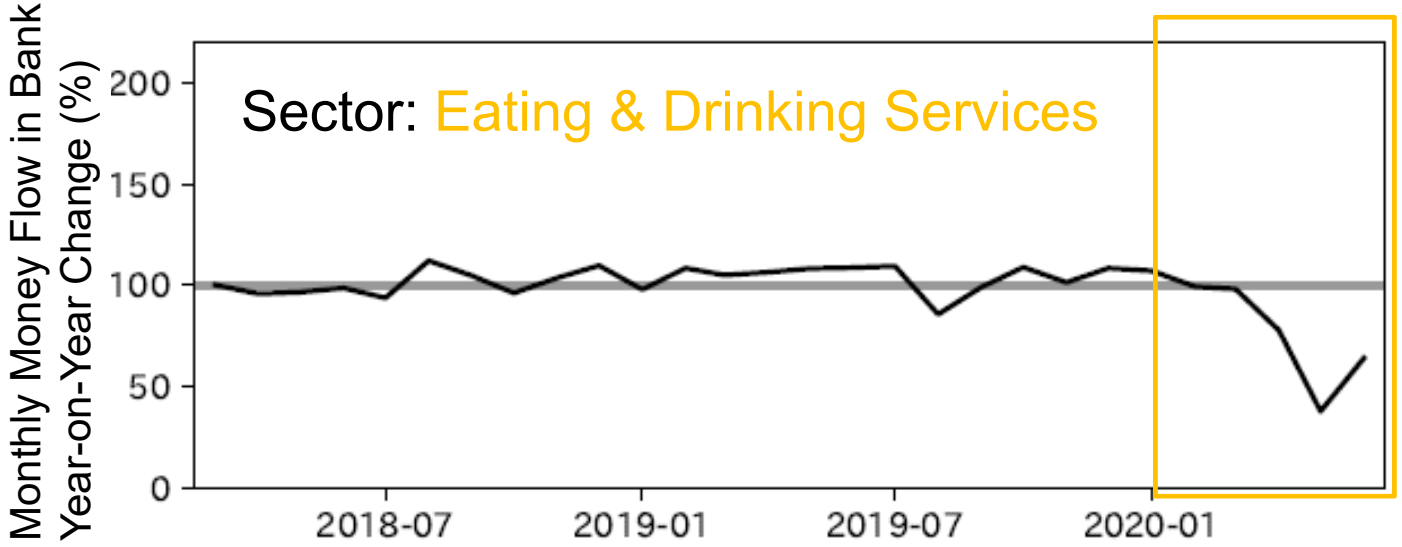
yoshi.fujiwara@gmail.com

Contents

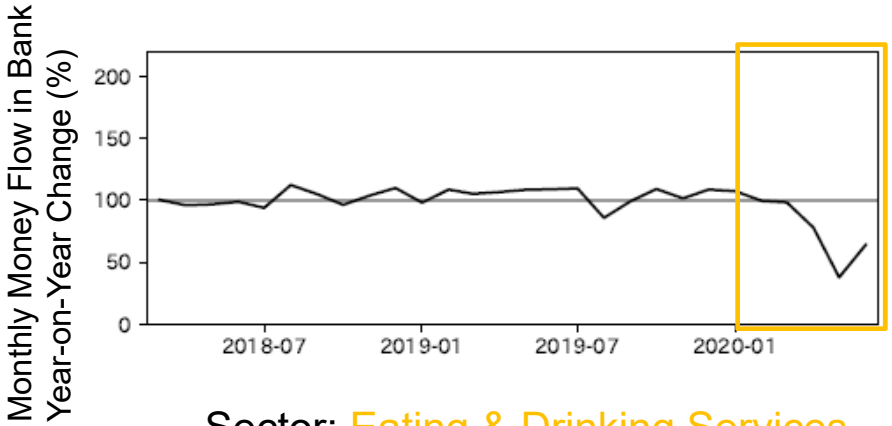
1. Background
2. Big Data of Bank Remittance among Firms
3. Network Structure
   - Topology
   - Hodge decomposition
   - NMF (non-negative matrix factorization)
4. Implications

Background

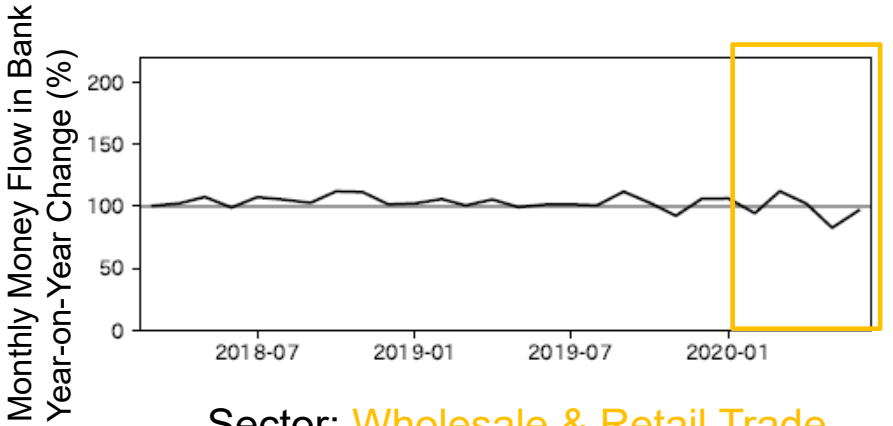
- Big data of production (Supplier-Customer) Network
  - ✓ Observation : year, quarter at most
  - ✓ Not real-time
- E.g. COVID-19 effect to real economy



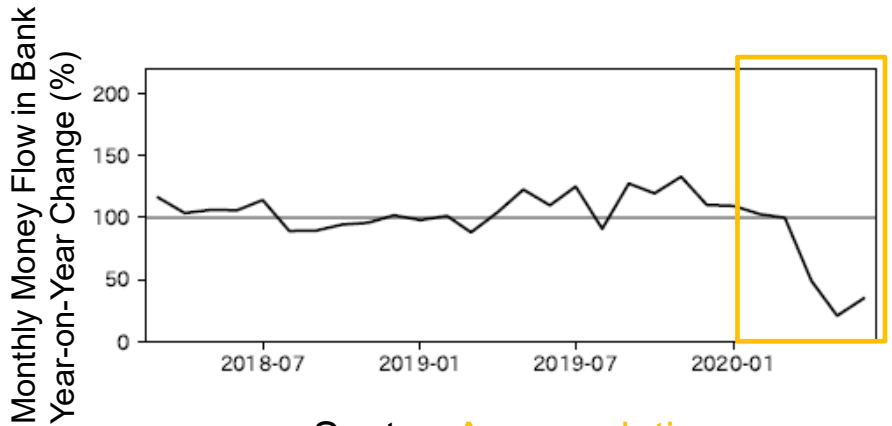
T. Yamaguchi et al.,
"Covid-19 Effect Big Data of Bank Remittance"
Shiga University, DP, October 2020



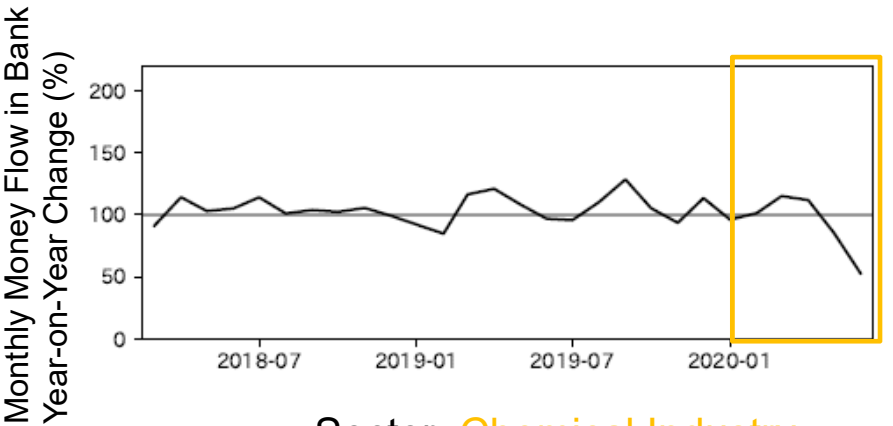
Sector: Eating & Drinking Services



Sector: Wholesale & Retail Trade



Sector: Accomodations



Sector: Chemical Industry

T. Yamaguchi et al.,
"Covid-19 Effect Big Data of Bank Remittance"
Shiga University, DP, October 2020

Contents

1. Background
2. Big Data of Bank Remittance among Firms
3. Network Structure
   - Topology
   - Hodge decomposition
   - NMF (non-negative matrix factorization)
4. Implications

Data : Bank Remittance

- Regional Bank: medium-sized, largest in a region (Shiga pref. 1M people)
- All accounts and remittances
  Period : 2017/March to 2019/July ＝ 29m = 883d
- Firms' accounts and intra-bank remittances selected

滋賀銀行

Table 1: **Bank accounts and transfers: summary**

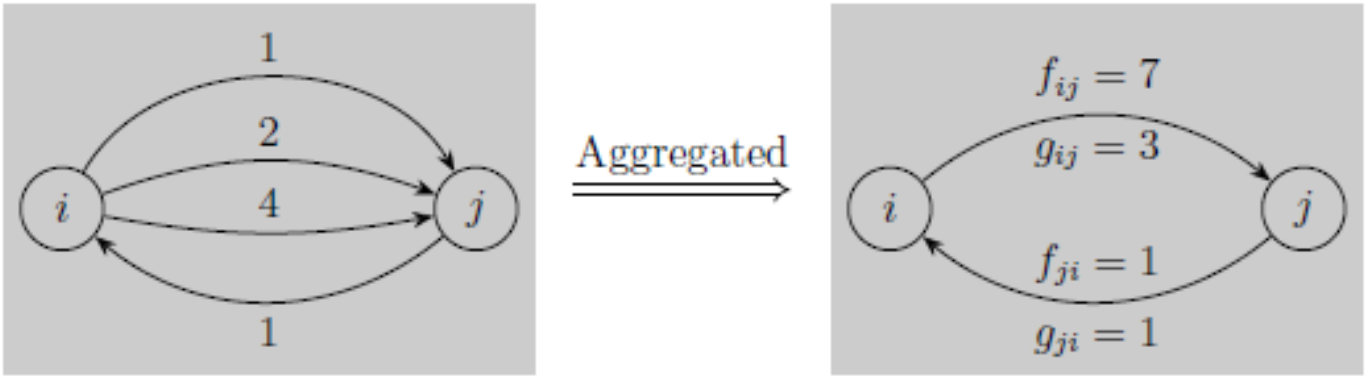
Number/Amount	Entire data	Within Bank A	
		all	firms
#Accounts	1.71 M	642,411	30,613
#Transfers	23.06 M	12,847,963	2,409,619
#Links	3.13 M	1,470,107	280,864
Transfer (Yen)	17.43 T	5.26 T	2.15 T

For a transfer $i \to j$, the column "Entire data" includes the cases in which either i or j is not an account of Bank A. The column "Within Bank A" corresponds to the case in which both i and j are accounts of Bank A. "firms" implies that both the source and the target of a link are firm accounts. M and T denote million and trillion, respectively.

- ✓ Accounts = 30K
- ✓ Remittances = 2.4M
- ✓ #edges = 0.28M
- ✓ Amount = 2.15 T ¥ (21.5 B $)

Construction of “Money Flow” NW

Remittances between accounts *i*, *j*



#Nodes= 30,613 = *N*
#Edges= 280,864 = *M*
Directed
Weighted : freq. and amount of remittance

Table 2: **Summary statistics for links' flows and frequencies**

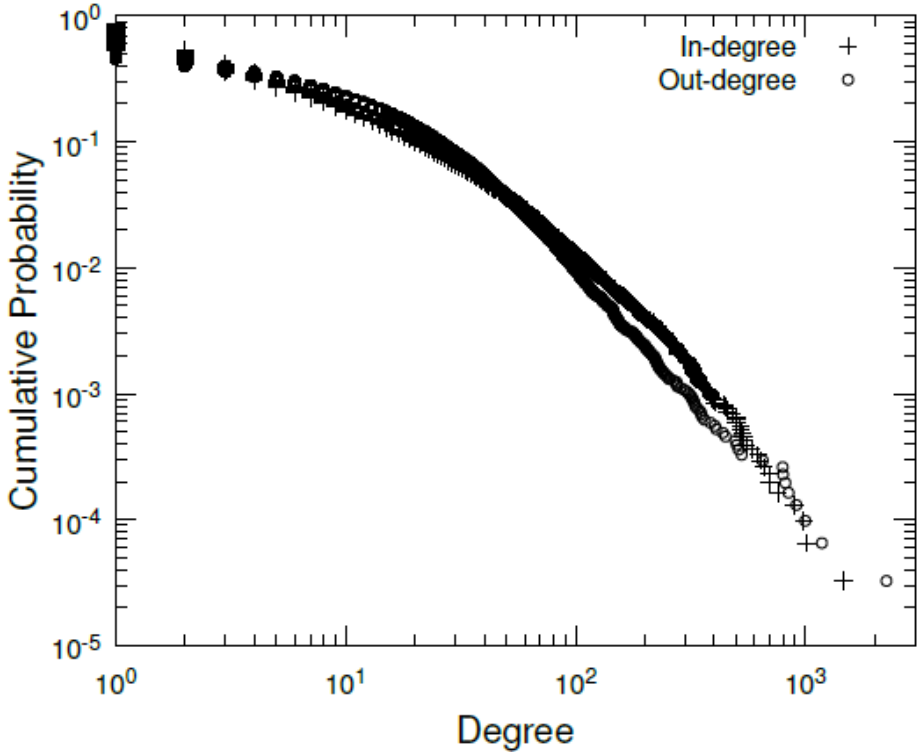
Stats.	Flow (Yen)	Frequency
Min.	1	1
Max.	3.00×10^{10}	2,616
Median	0.20×10^{6}	3
Avg.	7.65×10^{6}	8.58
Std.	1.53×10^{8}	19.92

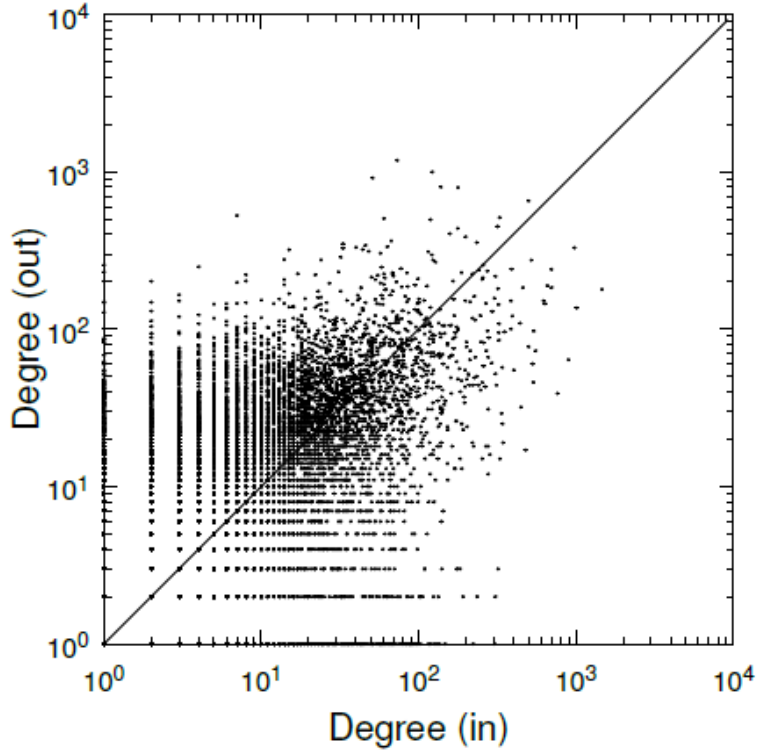
Contents

1. Background
2. Big Data of Bank Remittance among Firms
3. Network Structure
   - Topology
   - Hodge decomposition
   - NMF (non-negative matrix factorization)
4. Implications

Network Structure

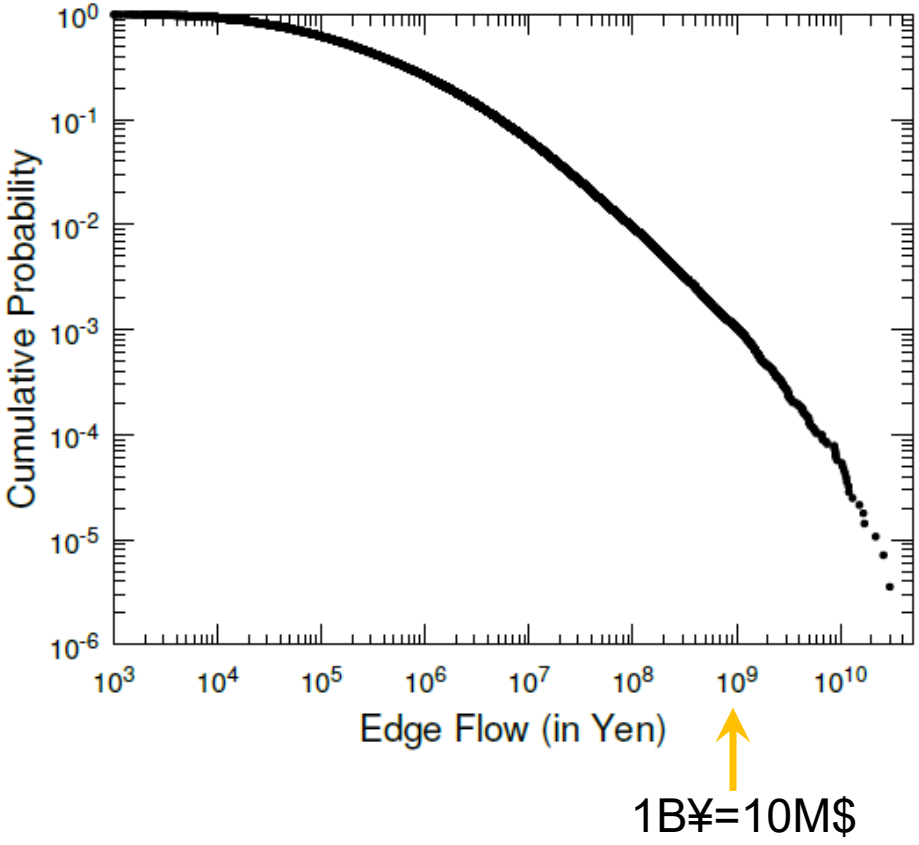
- NW is a sparse graph
  Density:= $M/(N*(N-1)) = 0.02\%$
- Heavy-tailed distributions for #souces (in-degree), #destinations (out-degree)
- Positive weak correlation between in-degree and out-degree

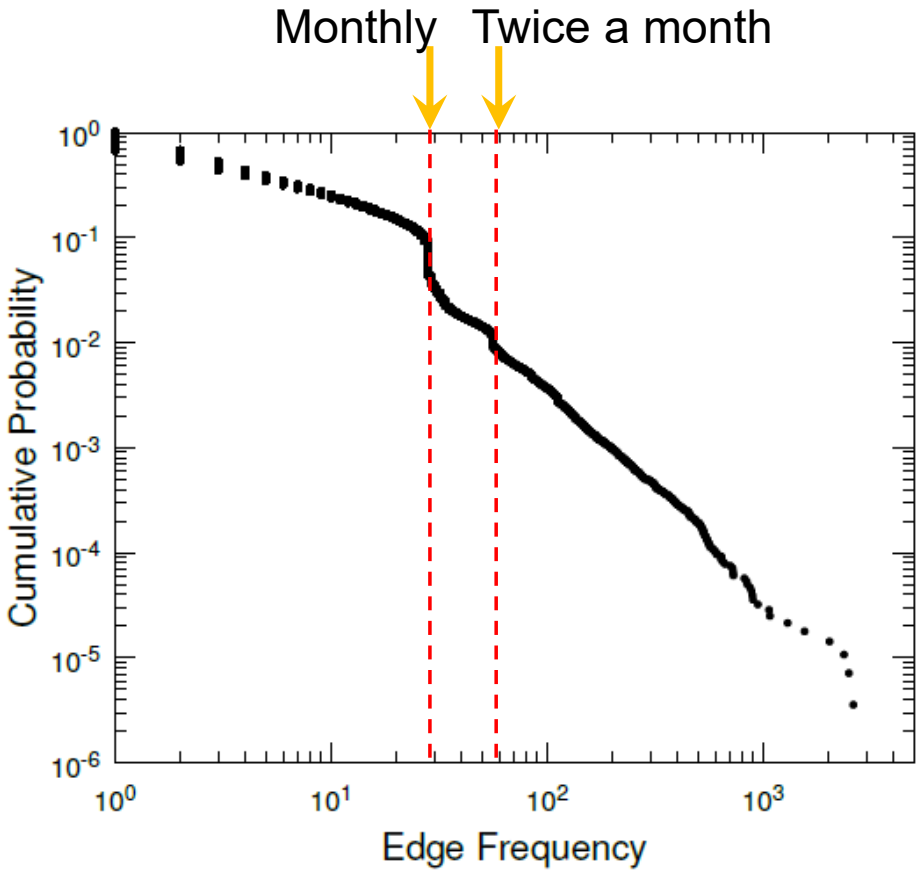




Pearson's $r = 0.303$ $(p < 10^{-6})$; Kendall's $\tau = 0.164$ $(p < 10^{-6})$

- Heavy-tail also for edge weight = freq.
  Presence of stable remittance (monthly, twice a month)





Period：2017/March to 2019/July＝29 months

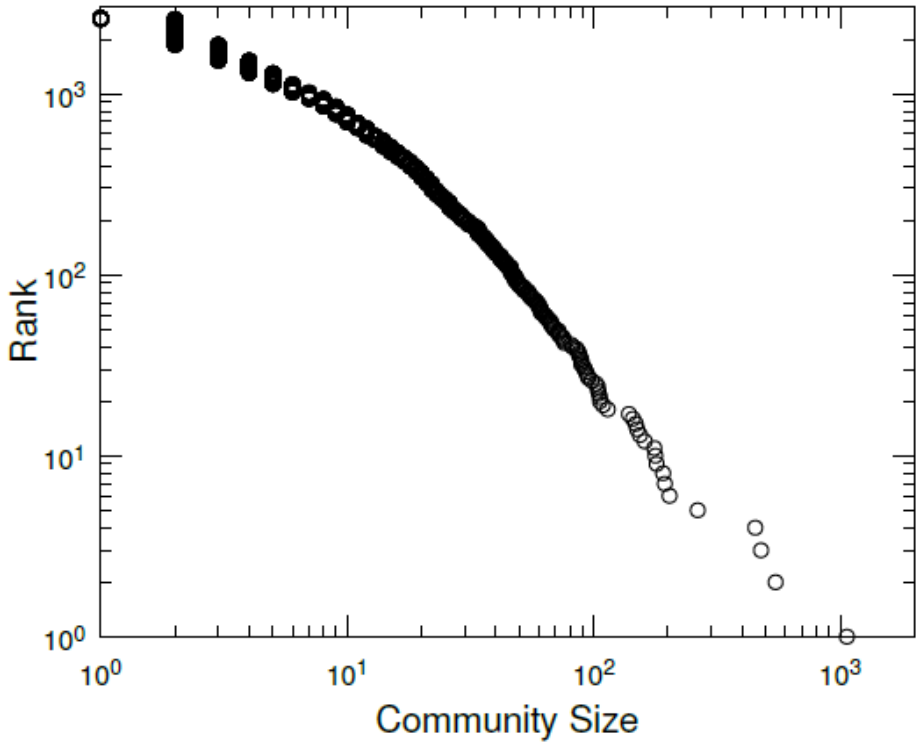
Community structure

- Hierarchical community analysis by Infomap
  - ✓ Similar to production network in Japan

Table 3: Numbers of communities, irreducible communities, and accounts at each level of community analysis using Infomap

Level	#comm.	#irr. comm.	#accounts	Ration(%)
1	164	143	355	0.012
2	2,327	2,264	28,948	94.5
3	215	215	1,310	0.043
Total	—	2,621	30,613	100.0

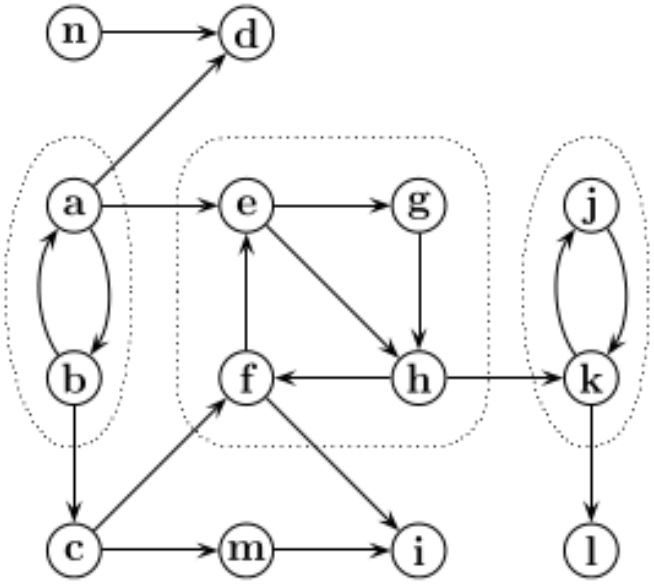
Each level corresponds to the hierarchical level in the Infomap community analysis [9]. A community at a level can be decomposed at the next lower level (from top to bottom). If a community cannot be decomposed further, it is called an irreducible community. The numbers of irreducible communities are listed in the third column. The fourth column lists the numbers of accounts belonging to these irreducible communities at each level.



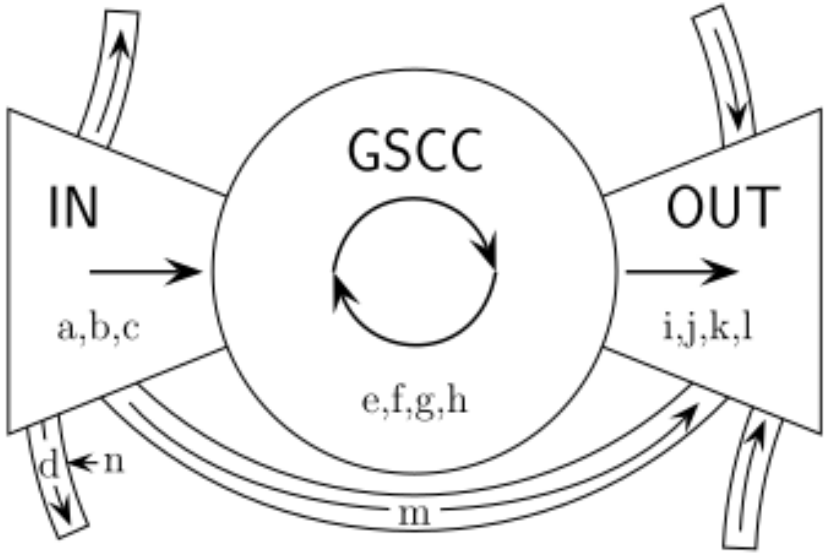
“Bowtie” structure of a directed graph

Strongly connected components (SCC) of a directed graph

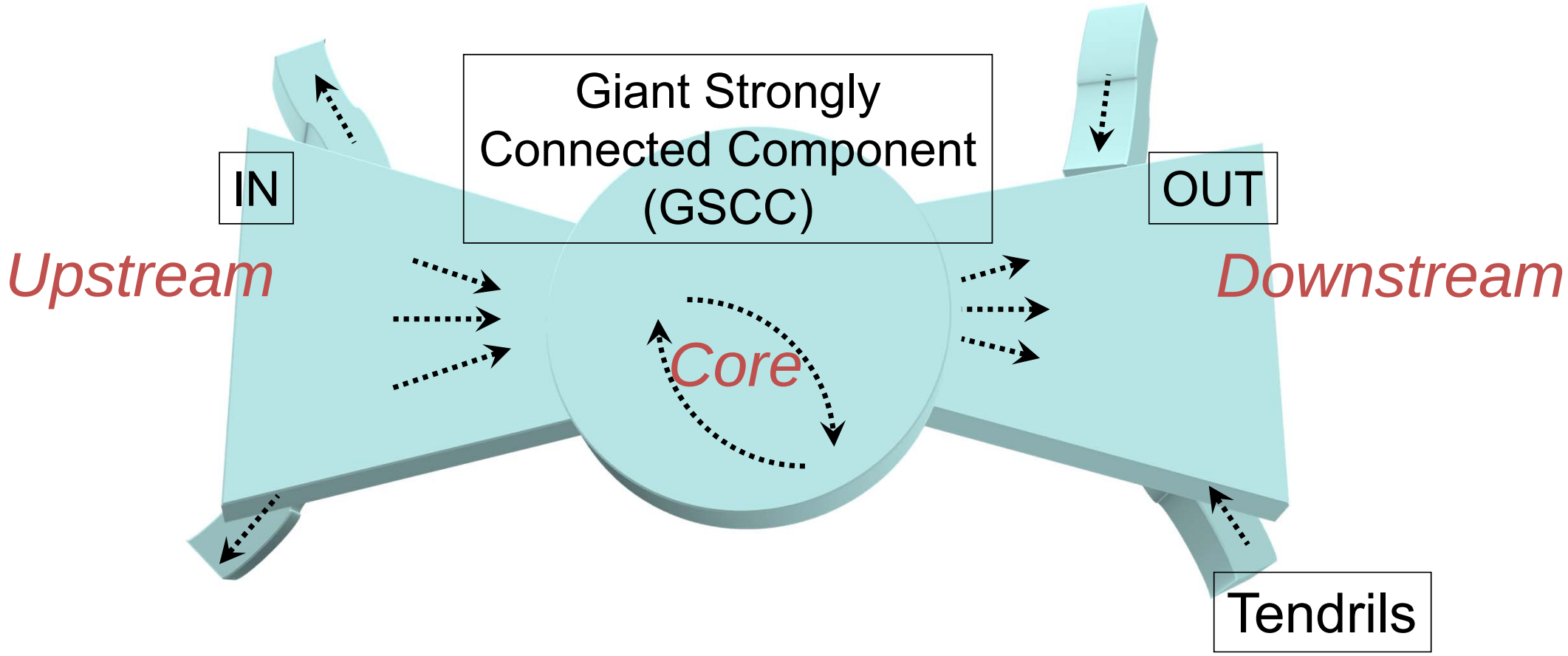
Example



Giant SCC (GSCC), *IN/OUT components* and others



Money flow: “bowtie” structure



Bow-tie or *walnut* structure

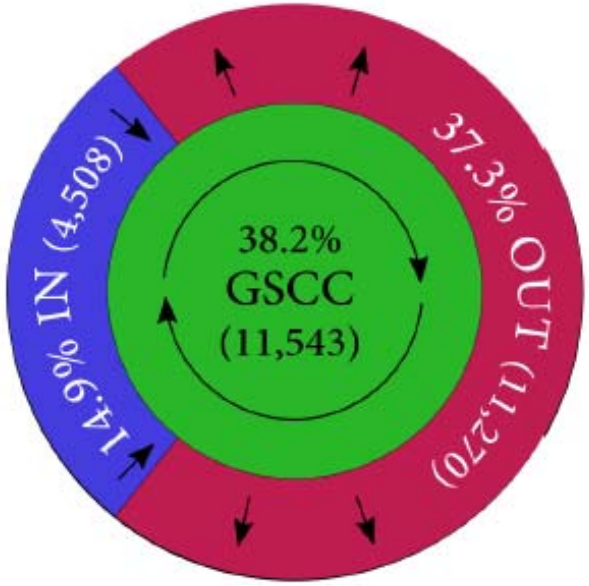
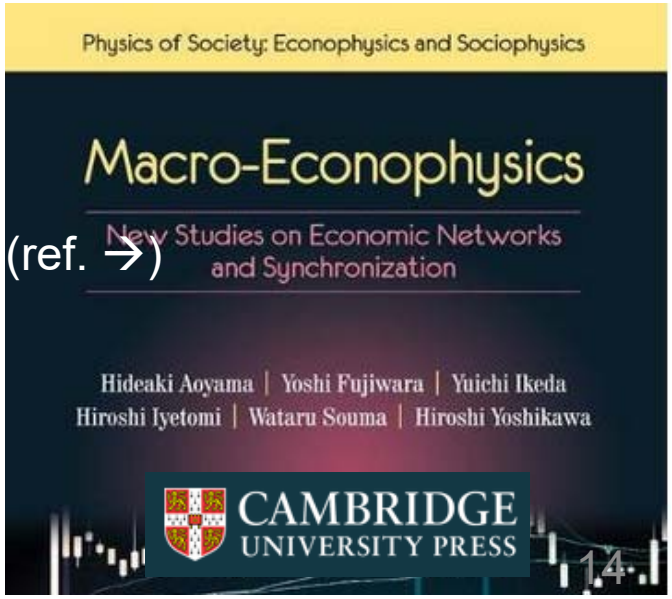


Table 5: "Walnut" structure: shortest distance from GSCC to IN/OUT.

IN to GSCC			OUT from GSCC		
Distance	#accounts	Ratio(%)	Distance	#accounts	Ratio(%)
1	4,346	96.41%	1	11,051	98.06%
2	144	3.19%	2	208	1.85%
3	8	0.18%	3	11	0.10%
4	10	0.22%	4	0	0.00%
Total	4,508	100%	Total	11,270	100%

The left half lists the number of accounts in the IN component connected to the GSCC accounts with the shortest distances within 4 at most. The right half represents the OUT component similarly.

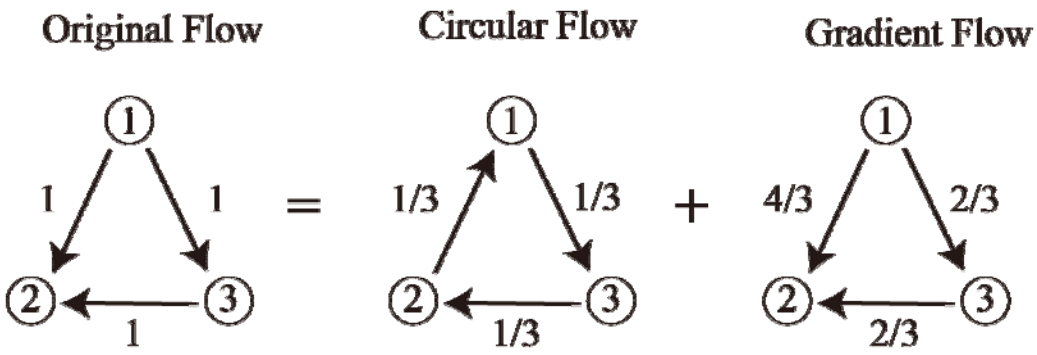
- Upstream/Downstream/Core of mony flow
  - ✓ IN/OUT connected to GSCC with short distances
  - ✓ Similar to production network at a nationwide scale in Japan (ref. →)



Q. But how to identify upstream/downstream and circulation of flow?

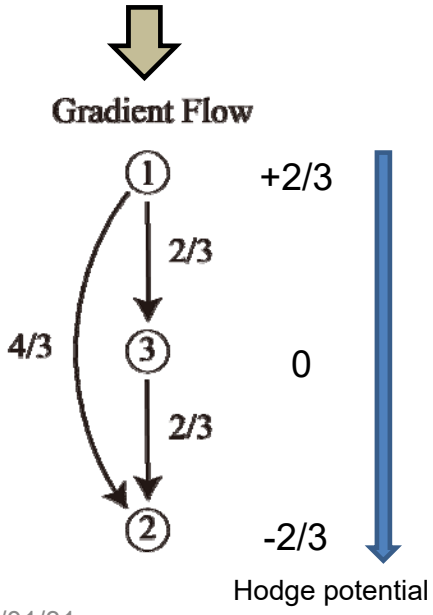
Helmholtz-Hodge decomposition applied to network flow

A simple example



H. von Helmholtz (physics)

W. Hodge (math)



Hodge decomposition for physicists

Adjacency

$$A_{ij} = \begin{cases} 1 & \text{if there is a directed edge from node } i \text{ to node } j \\ 0 & \text{otherwise} \end{cases}$$

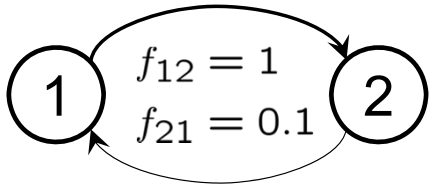
Flow along each edge

$f_{ij} > 0$ if there is a flow from node i to node j

Definition: Flow (net) $F_{ij} = f_{ij} - f_{ji}$

Definition: Weight or strength $w_{ij} = f_{ij} + f_{ji}$

Case: a pair of reciprocal edges



$F_{12} = 0.9 = -F_{21}$

$w_{12} = 1.1 = w_{21}$

Hodge decomposition

$$F_{ij} = w_{ij}(\phi_i - \phi_j) + F_{ij}^{(\text{circ})}$$

Circular flow is divergence-free by definition: $\sum_j F_{ij}^{(\text{circ})} = 0$

→ Simultaneous linear equations (eigenvalue problem) to determine the potentials:

$$\sum_j L_{ij}\phi_j = \sum_j F_{ij} \qquad (i = 1, \ldots, \#\text{nodes})$$

where Laplacian is given by

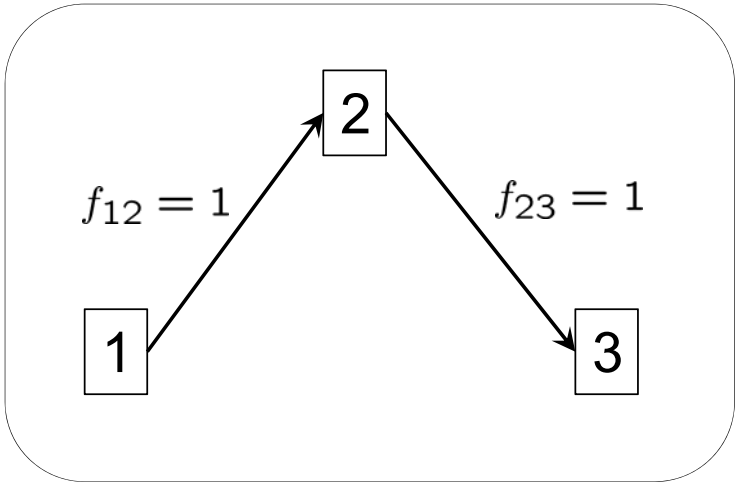
$$L_{ij} = \delta_{ij} \sum_k w_{ik} - w_{ij}$$

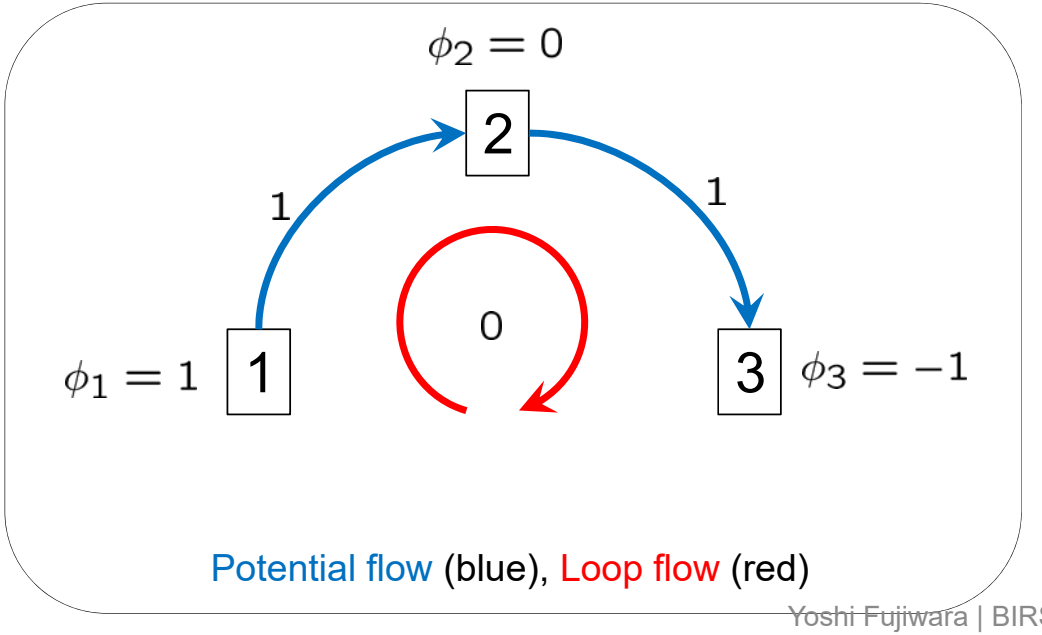
N.B. The equations are not independent, corresponding to an arbitrary origin of potential.
i.e. the eigenvalue problem has one trivial solution.

Let us call each part as follows:

$$F_{ij} = \underbrace{w_{ij}(\phi_i - \phi_j)}_{\text{Potential flow}} + \underbrace{F_{ij}^{(\text{circ})}}_{\text{Circular flow}}$$

Ex.1





Potential flow (blue), Loop flow (red)

Adjacency

$$||A_{ij}|| = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Flow

$$||f_{ij}|| = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

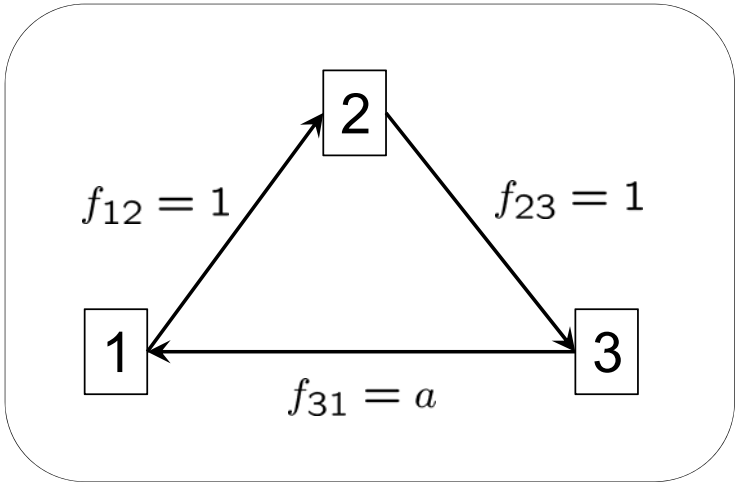
$$||w_{ij}|| = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

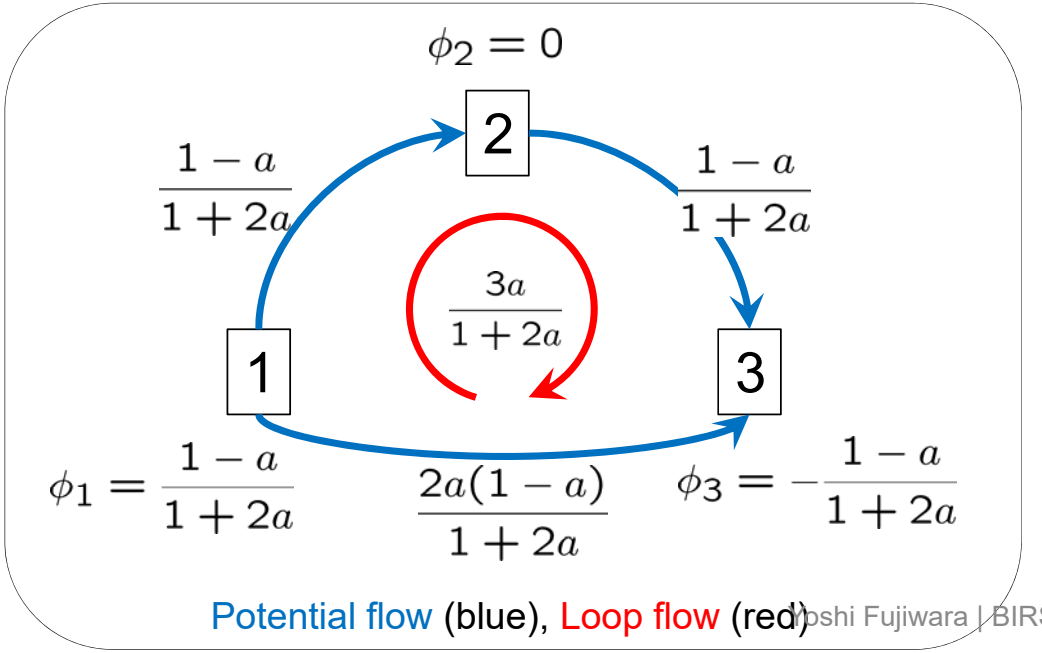
$$||F_{ij}|| = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

$$||L_{ij}|| = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

Solve $\sum_j L_{ij}\phi_j = \sum_j F_{ij}$

Ex.2





Potential flow (blue), Loop flow (red)

Adjacency

$$||A_{ij}|| = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Flow

$$||B_{ij}|| = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a & 0 & 0 \end{bmatrix}$$

$$||w_{ij}|| = \begin{bmatrix} 0 & 1 & a \\ 1 & 0 & 1 \\ a & 1 & 0 \end{bmatrix}$$

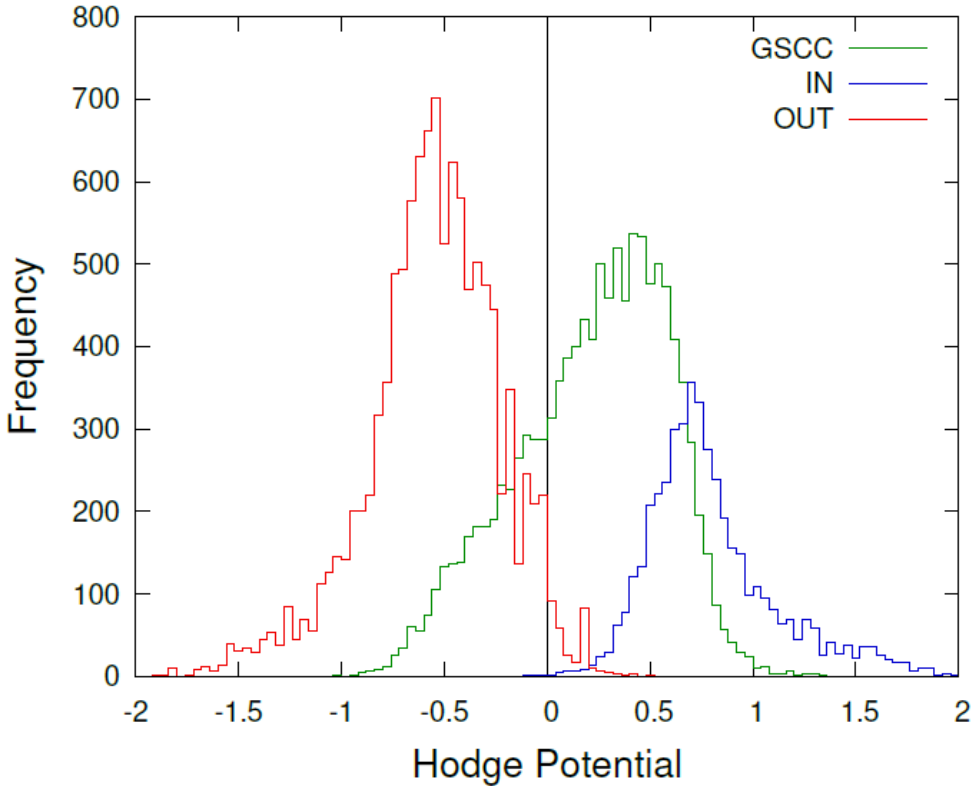
$$||F_{ij}|| = \begin{bmatrix} 0 & 1 & -a \\ -1 & 0 & 1 \\ a & -1 & 0 \end{bmatrix}$$

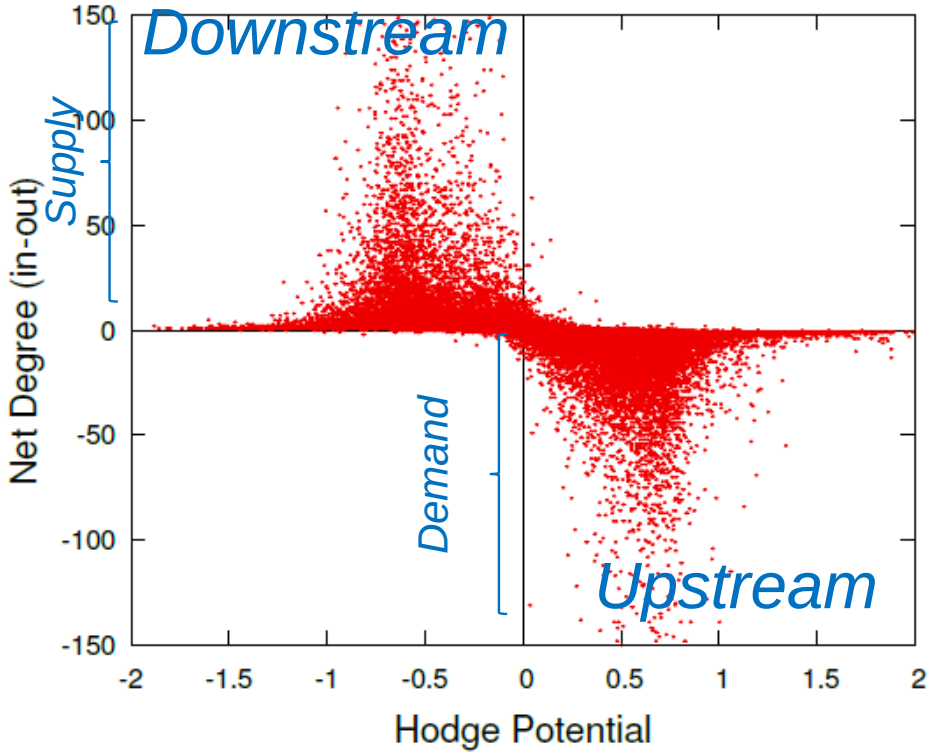
$$||L_{ij}|| = \begin{bmatrix} 1+a & -1 & -a \\ -1 & 2 & -1 \\ -a & -1 & 1+a \end{bmatrix}$$

Solve $\sum_j L_{ij}\phi_j = \sum_j F_{ij}$

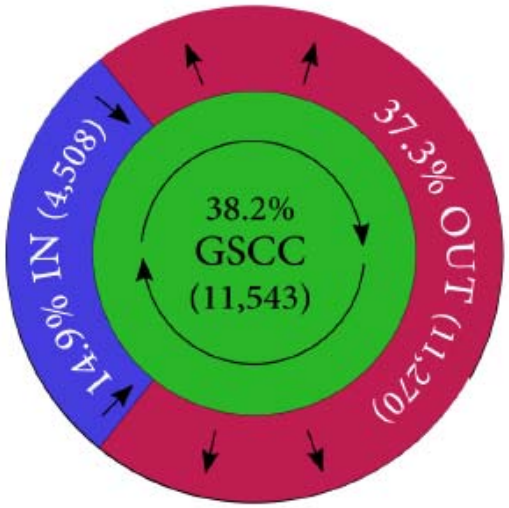
Hodge decomposition

- Hodge potential of each account
  - ✓ Cf. bow-tie structure（left）
  - ✓ Cf. supply and demand（right）



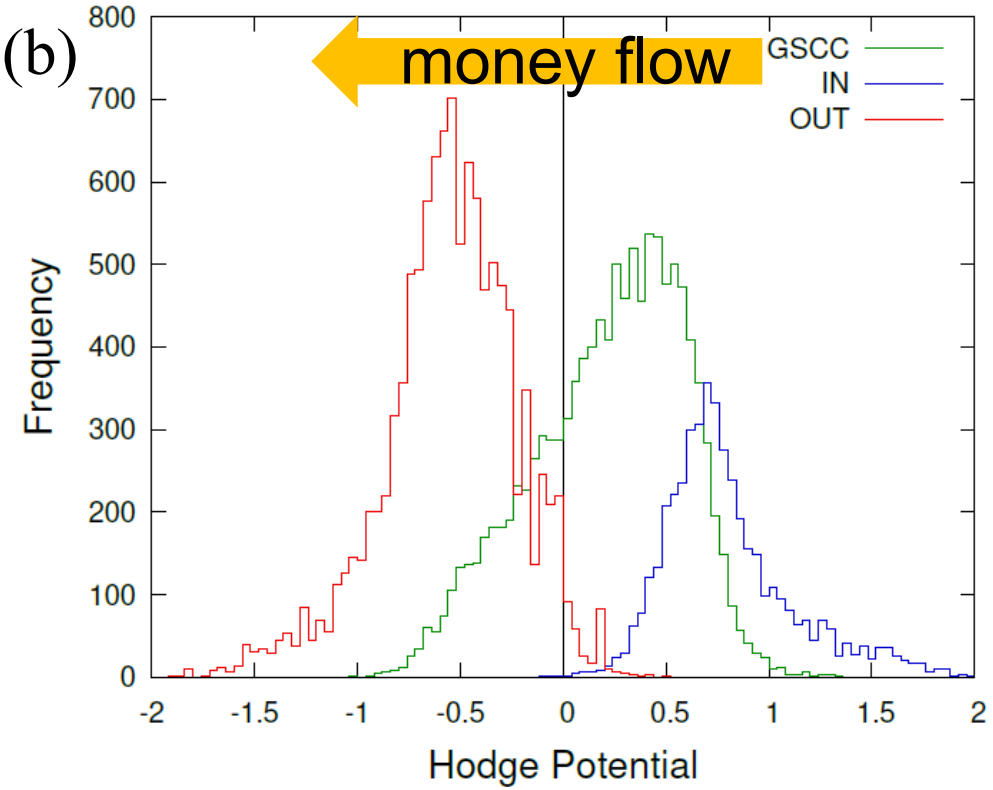


(a)



money flow

(b)

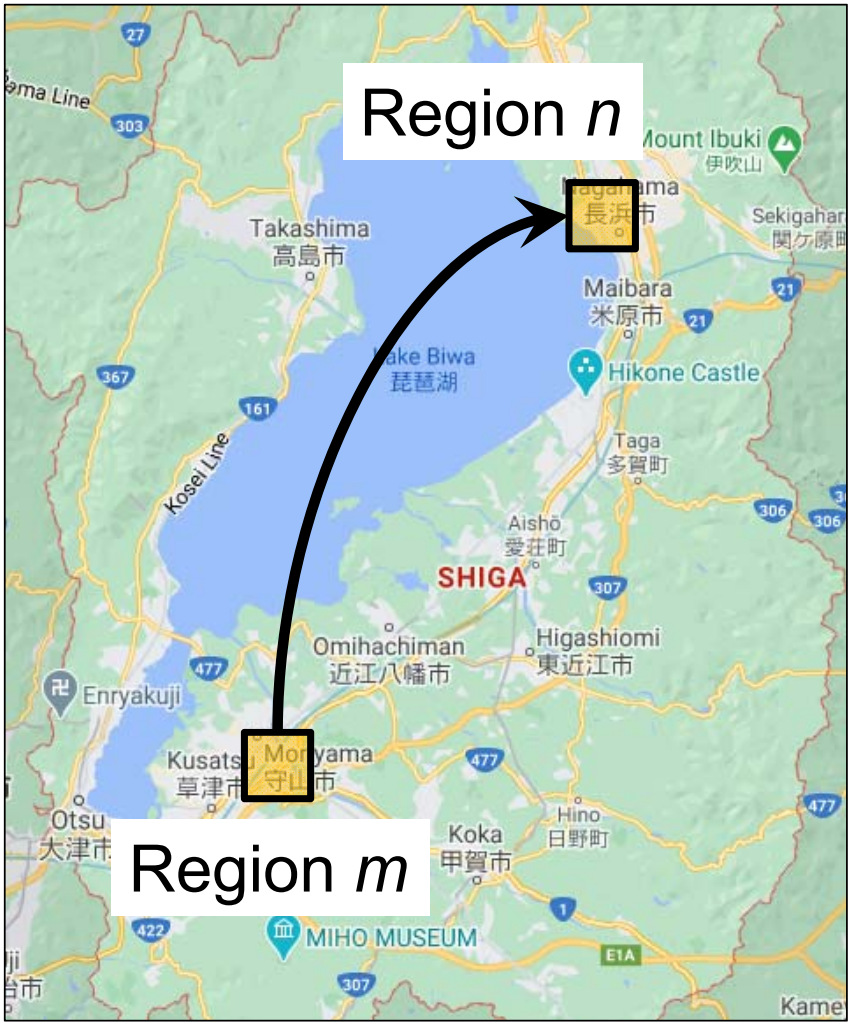


NMF(non-negative matrix factorization)

- Money flow
  - ✓ “principal components”
  - ✓ how money flow from one region to another
- Convert remittances based on geographical locations of source and destination

$$V_{m,n}$$

Flow from region *m* to region *n*

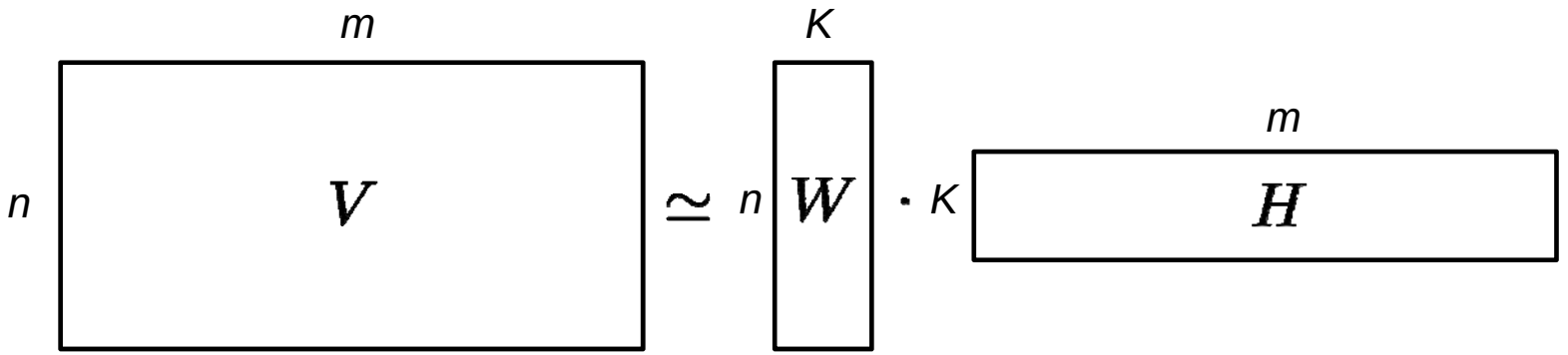


NMF = non-negative matrix factorization

$V_{m,n} \geq 0$

$$V \simeq W\,H$$

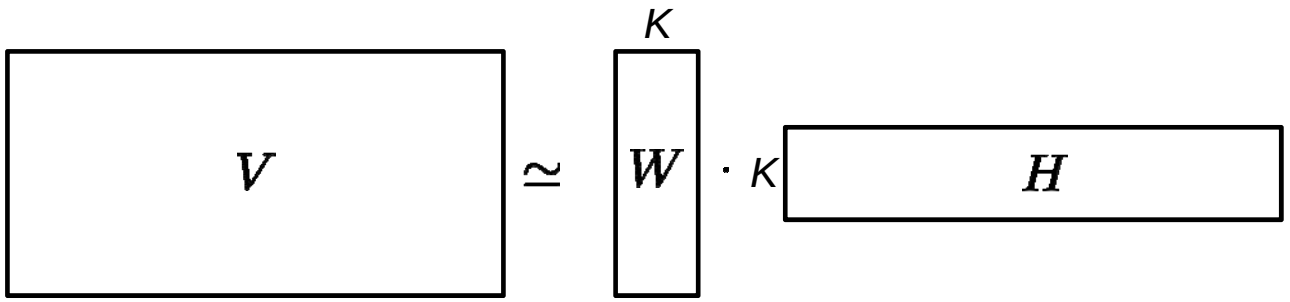
where $W \geq 0, H \geq 0$



If V is sparse, $K \ll n, m$ K components

for our study $n = m$

How to determine the number of components, *K* ?
NMF is a "probabilistic latent semantic analysis" in ML text analysis



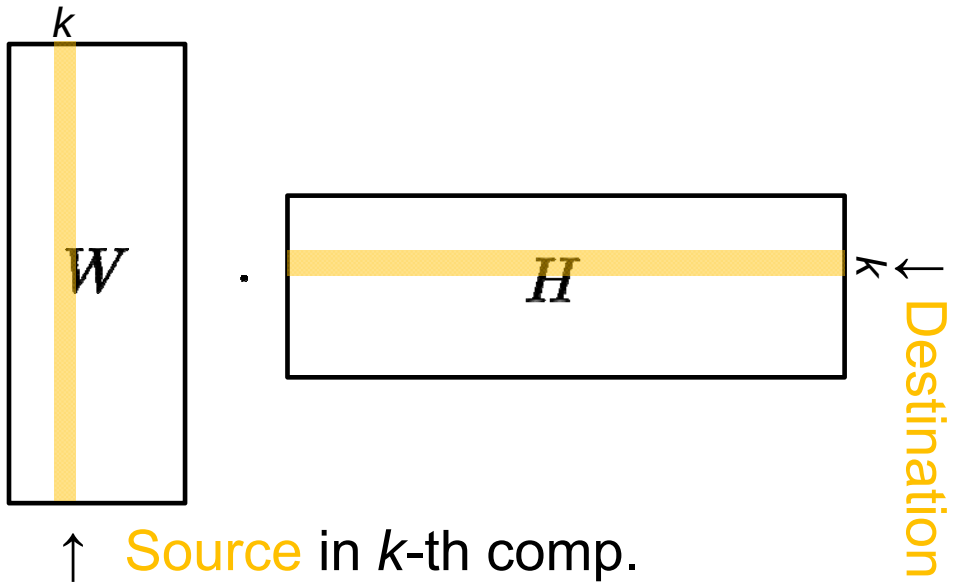
document-word frequency *doc. = a mix of topics / topic = a mix of words*

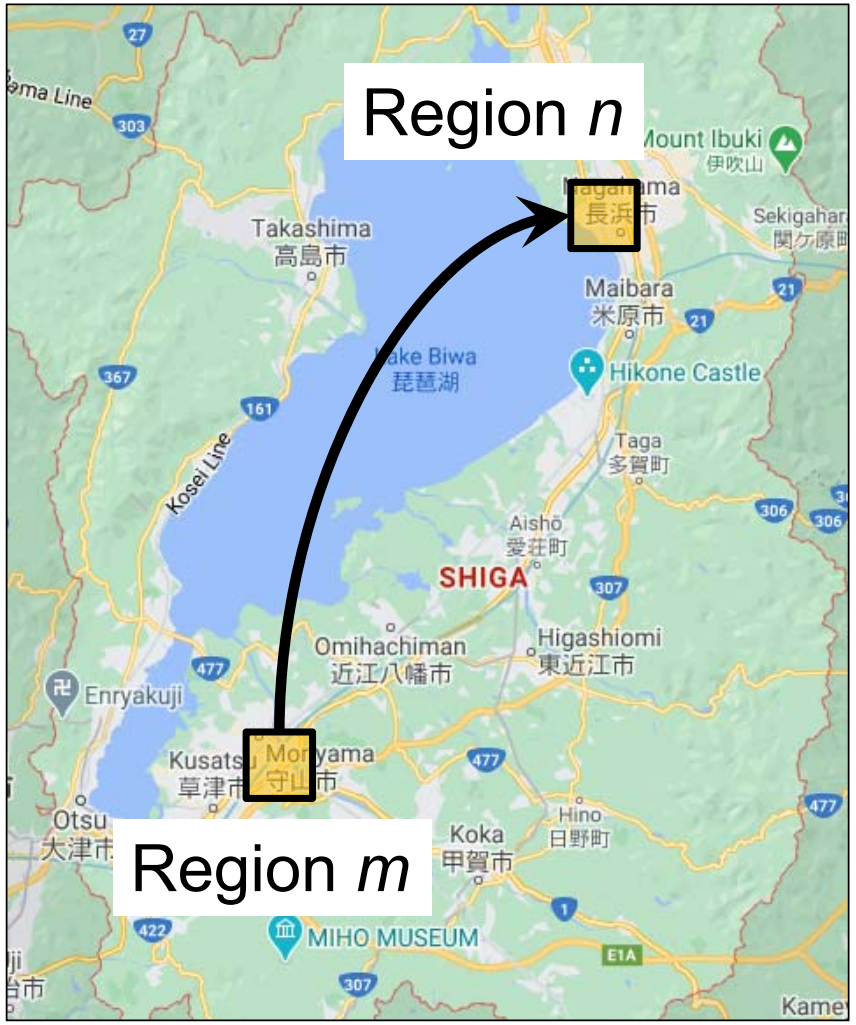
✓ Cross validation and other methods are used to tells us about "number of topics", *K*

See *Y. Fujiwara, R. Islam*, Bitcoin's Crypto Flow Network, *https://arxiv.org/abs/2106.11446*

Geographical information

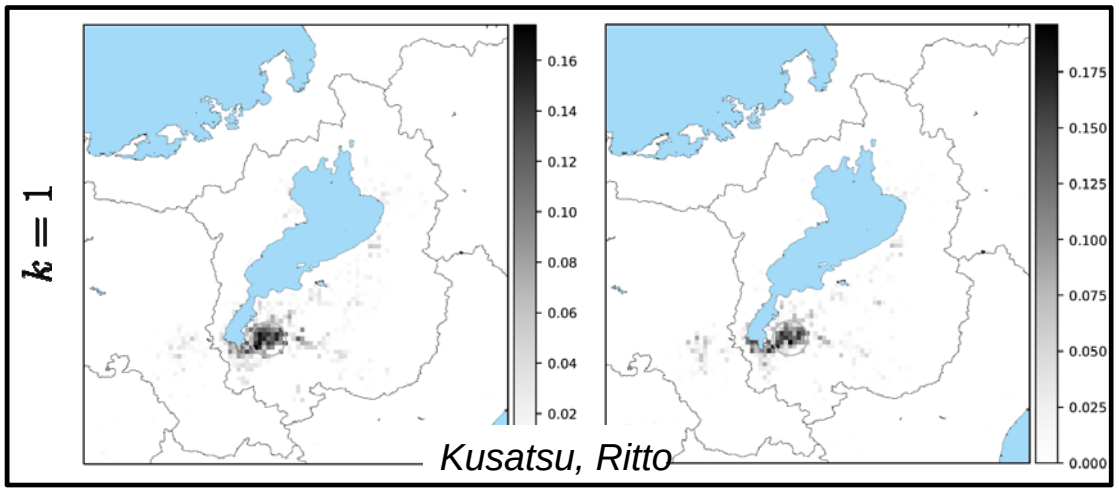
$$V_{m,n} \simeq \sum_{k}^{d} W_{mk} H_{kn}$$



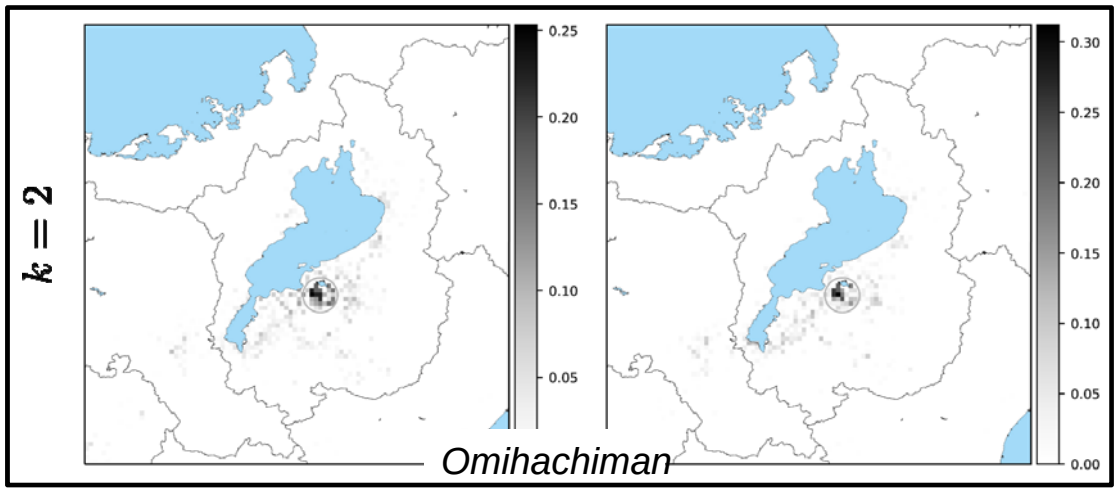


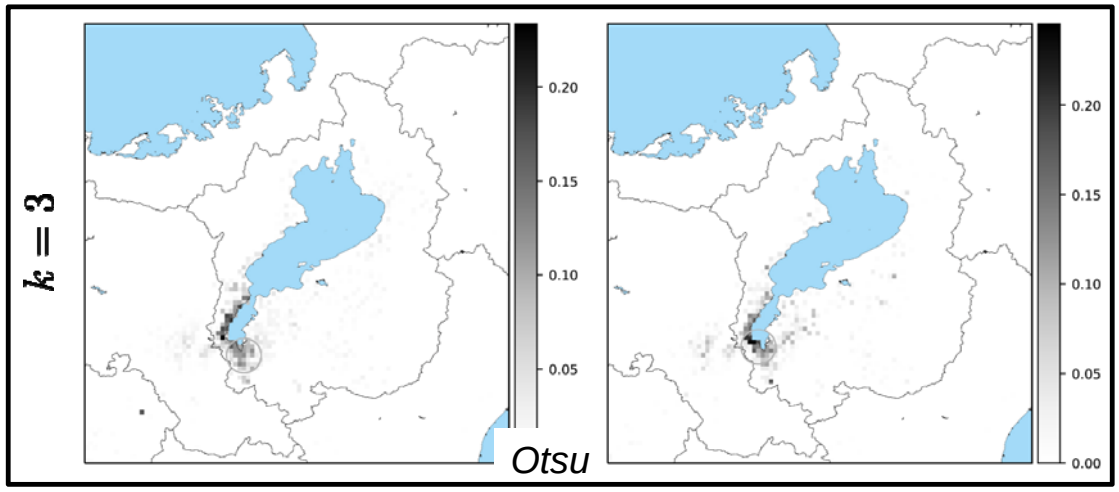
滋賀銀行ディスクロージャー誌2020「しがぎん」店舗配置図

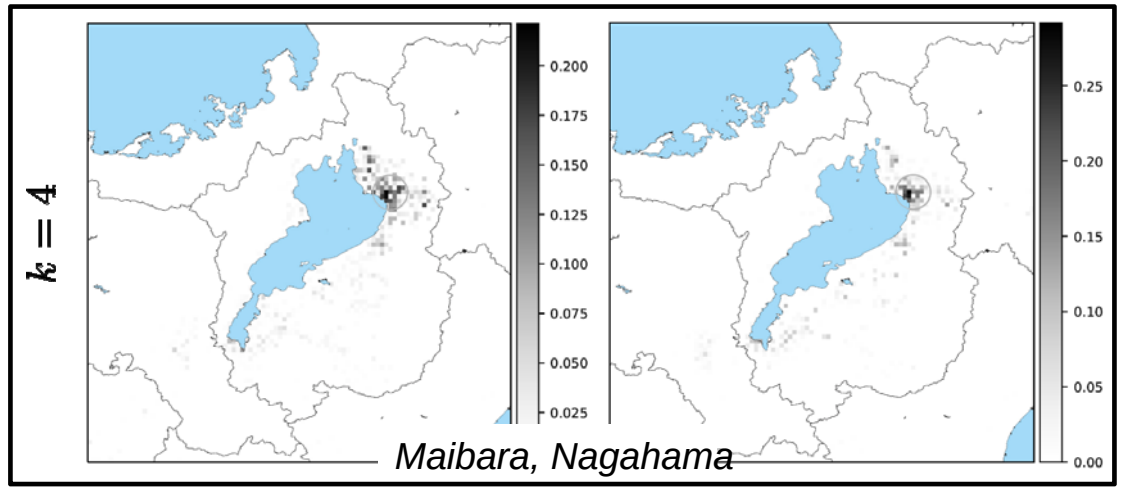
Source → Destination



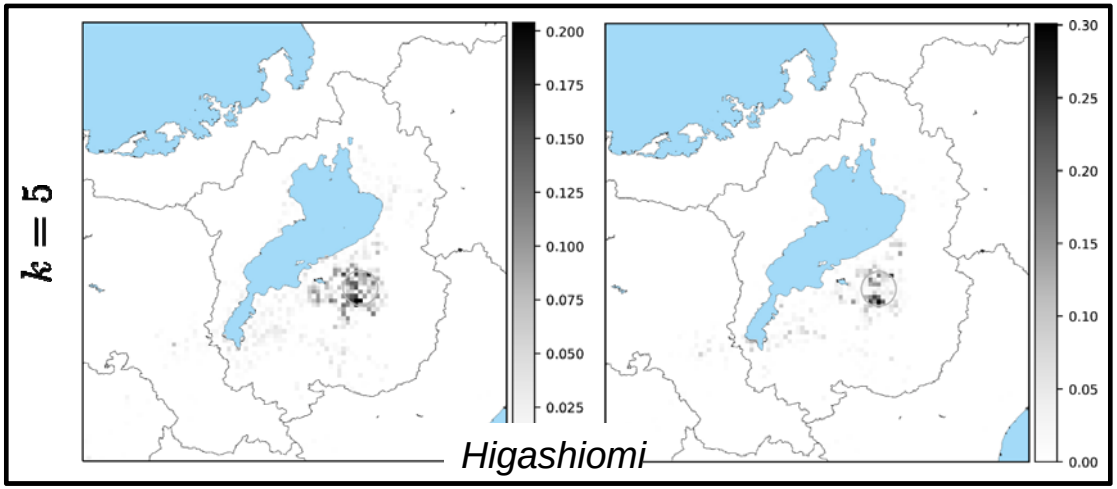
Source → Destination



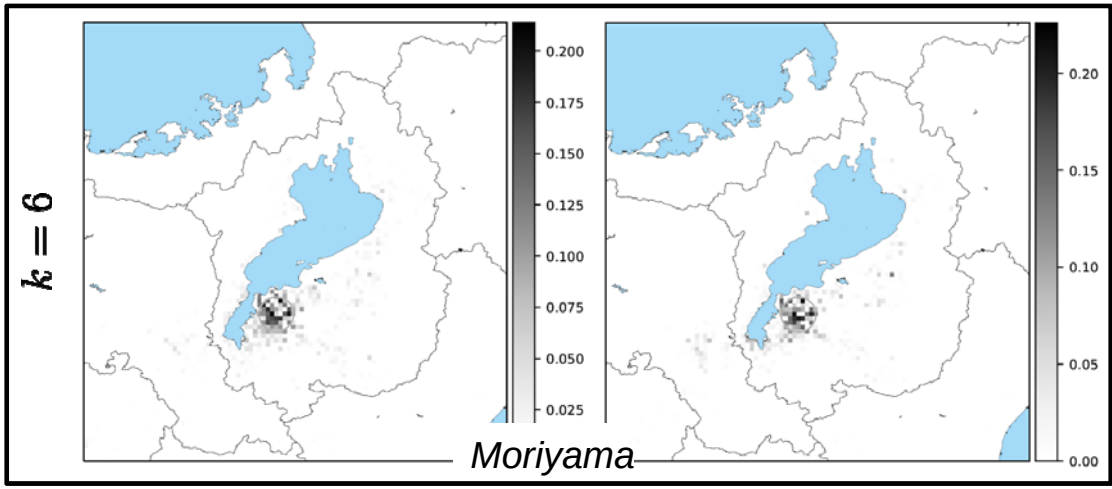


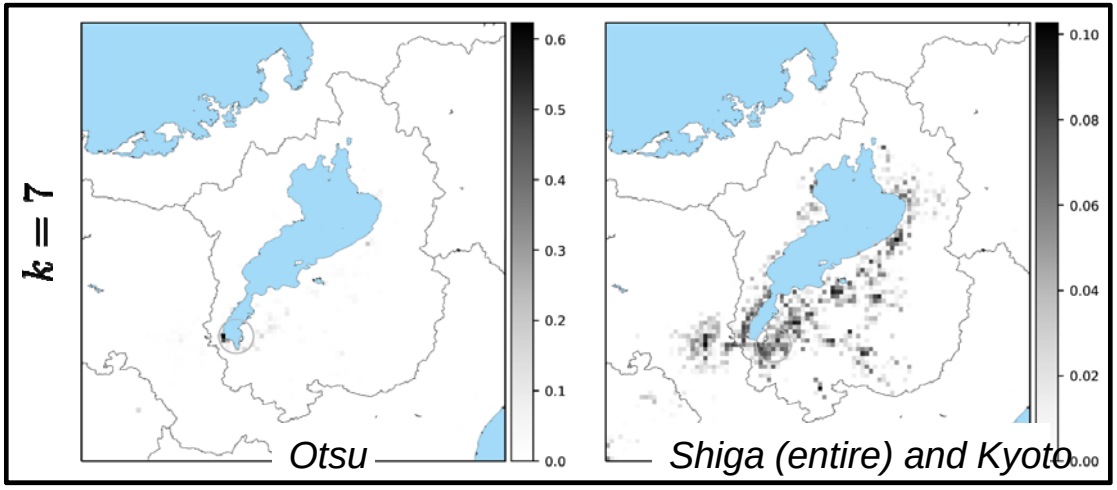


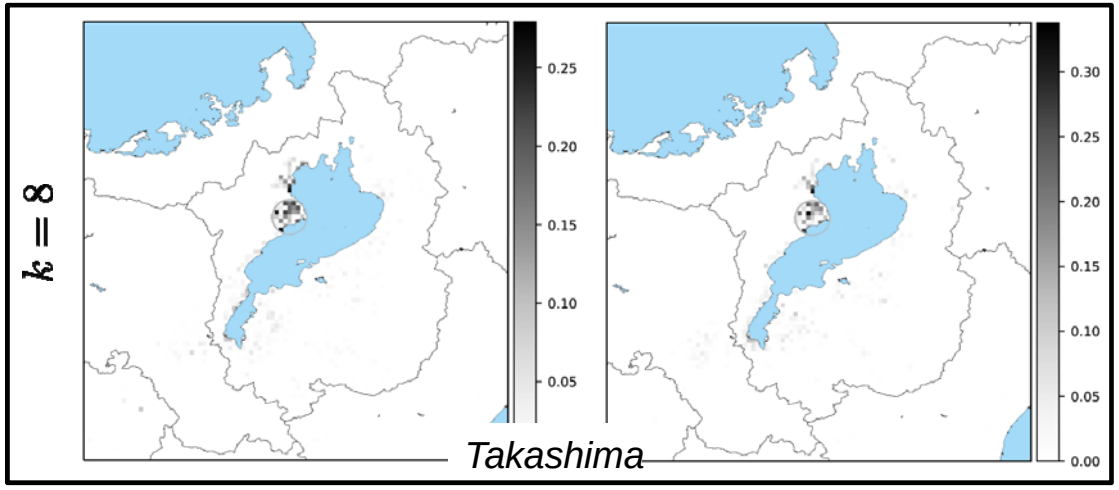
Source → Destination



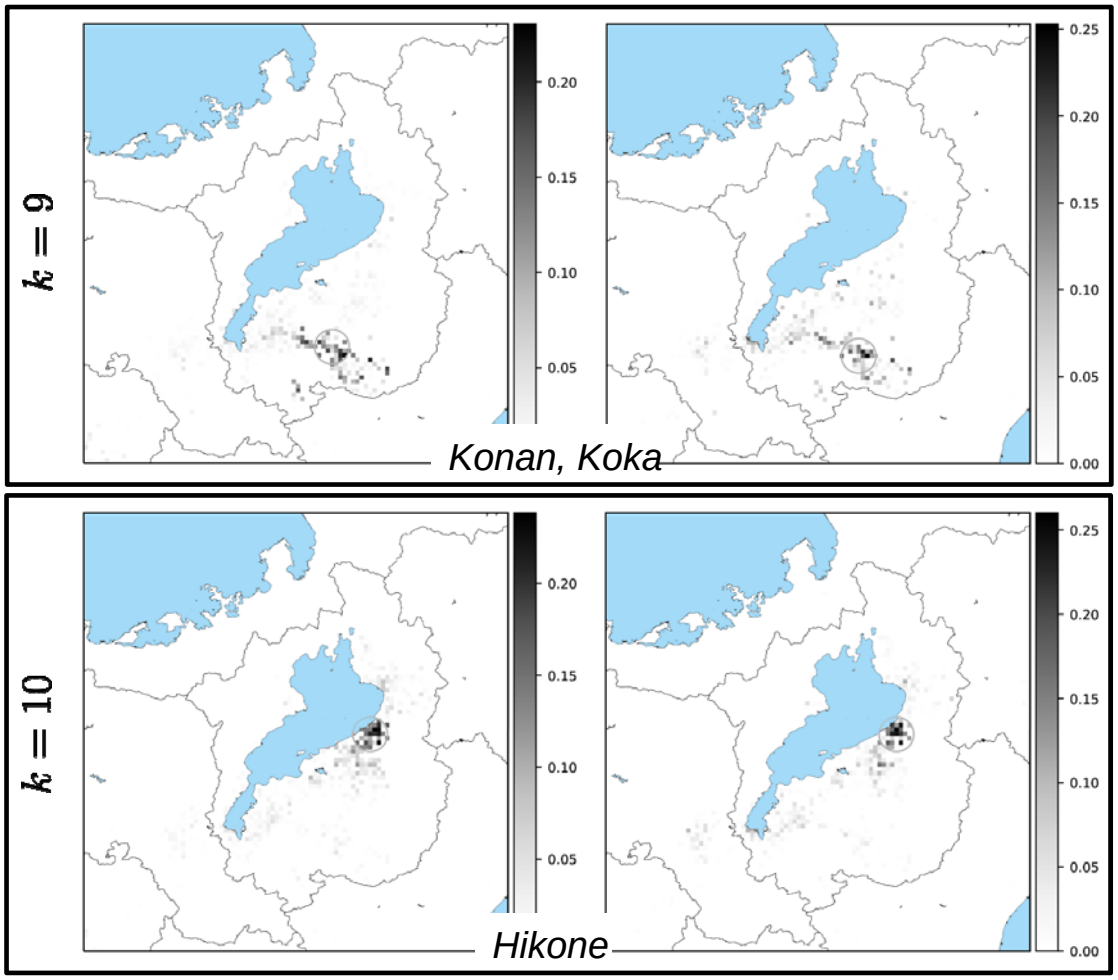
Source → Destination





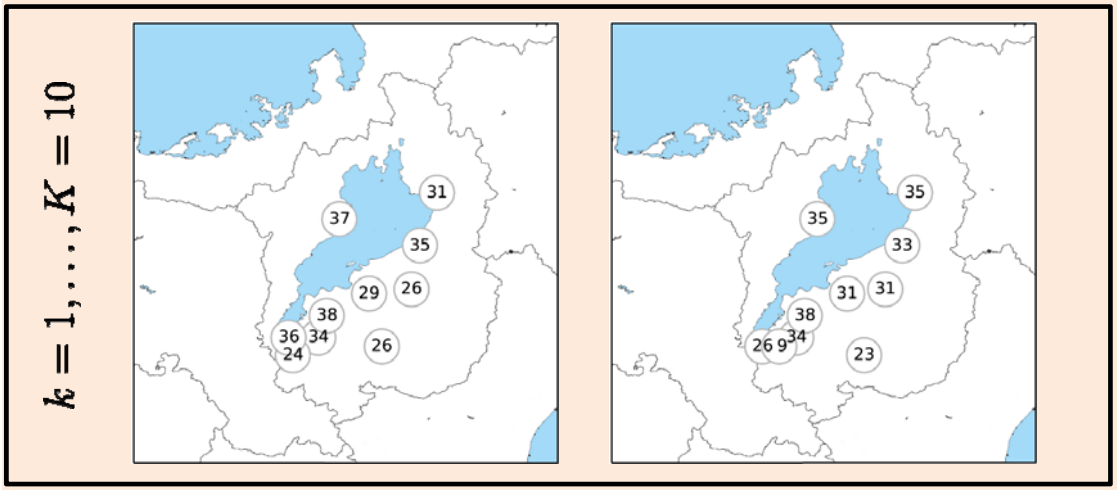


Source → Destination



10 components

Circle's numbers: concentration (%)



Source → Destination

Contents

1. Background
2. Big Data of Bank Remittance among Firms
3. Network Structure
   - Topology
   - Hodge decomposition
   - NMF (non-negative matrix factorization)
4. Implications

Implications

Money flow between firms' bank accounts in a region

1. Ideal for studies on the regional real economy
2. Real-time monitoring posssible
3. Applications
   - ✓ COVID-19, economic stress and effect
   - ✓ Firms default/bankruptcy and influence
   - ✓ Illegal or outlier activities of money flow

Conclusion

- Big data of money flow in a bank's accounts of firms
  - ✓ Bank remittance data
- Background
  - ✓ Real-time observation of real economy
    → Many applications, e.g. Covid-19, financial stress
  - ✓ Purpose: First, we studied the long-term and main structure
- Results
  - ✓ Bow-tie structure to identify upstream/downstream
  - ✓ Hodge potential: Each firm's location in the money flow
    → correlated with supply/demand of money
  - ✓ NMF (non-negative matrix factorization=NMF)
    "principal components" of flow
    → geographical features extracted






Yoshi Fujiwara *Hideaki Aoyama* *Hiroyasu Inoue* *Takayuki Yamaguchi* *Takuma Tanaka* *Kentaro Kikuchi*

Fujiwara et al. *EPJ Data Science* (2021) 10:19
https://doi.org/10.1140/epjds/s13688-021-00274-x

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE Open Access

Check for updates

Money flow network among firms' accounts in a regional bank of Japan

Yoshi Fujiwara[1,2*], Hiroyasu Inoue[1], Takayuki Yamaguchi[2], Hideaki Aoyama[3,4], Takuma Tanaka[2,5] and Kentaro Kikuchi[6]

https://doi.org/10.1140/epjds/s13688-021-00274-x

Thank you!

yoshi.fujiwara@gmail.com