Universiteit Leiden

Business Intelligence & Process Modelling

Frank Takes

Universiteit Leiden

Lecture 11 — Applications and Case Studies

Recap

- Business Intelligence: anything that aims at providing actionable information that can be used to support business decision making (February and March)
- Process Modelling (April and May)

Universiteit Leiden

Network science (recap)

Context

- Data
- Data Analysis
- Data Mining
- Data Analytics
- Data Science
- Big Data

Context

- Data
- Data Analysis
- Data Mining
- Data Analytics
- Data Science
- Big Data

- Network Data
- Social Network Analysis
- Network Science
- Graph Mining

Network science

- **Network science**: understanding data by investigating interactions and relationships between individual data objects as a network
- **Networks** are the central model of computation

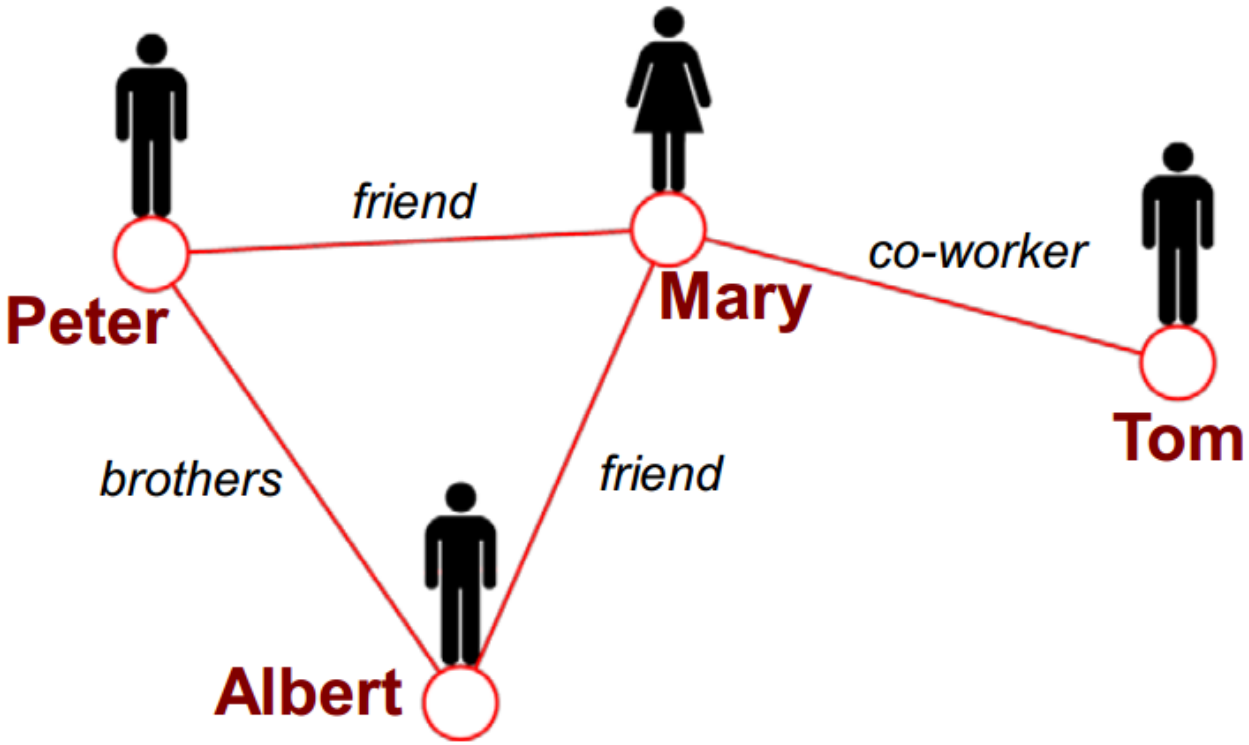
Network science

- **Network science**: understanding data by investigating interactions and relationships between individual data objects as a network
- **Networks** are the central model of computation
- Branch of data science focusing on network data
- Method in complexity research
- Complex systems approach: the behavior emerging from the network reveals patterns not visible when studying the individuals

Networks

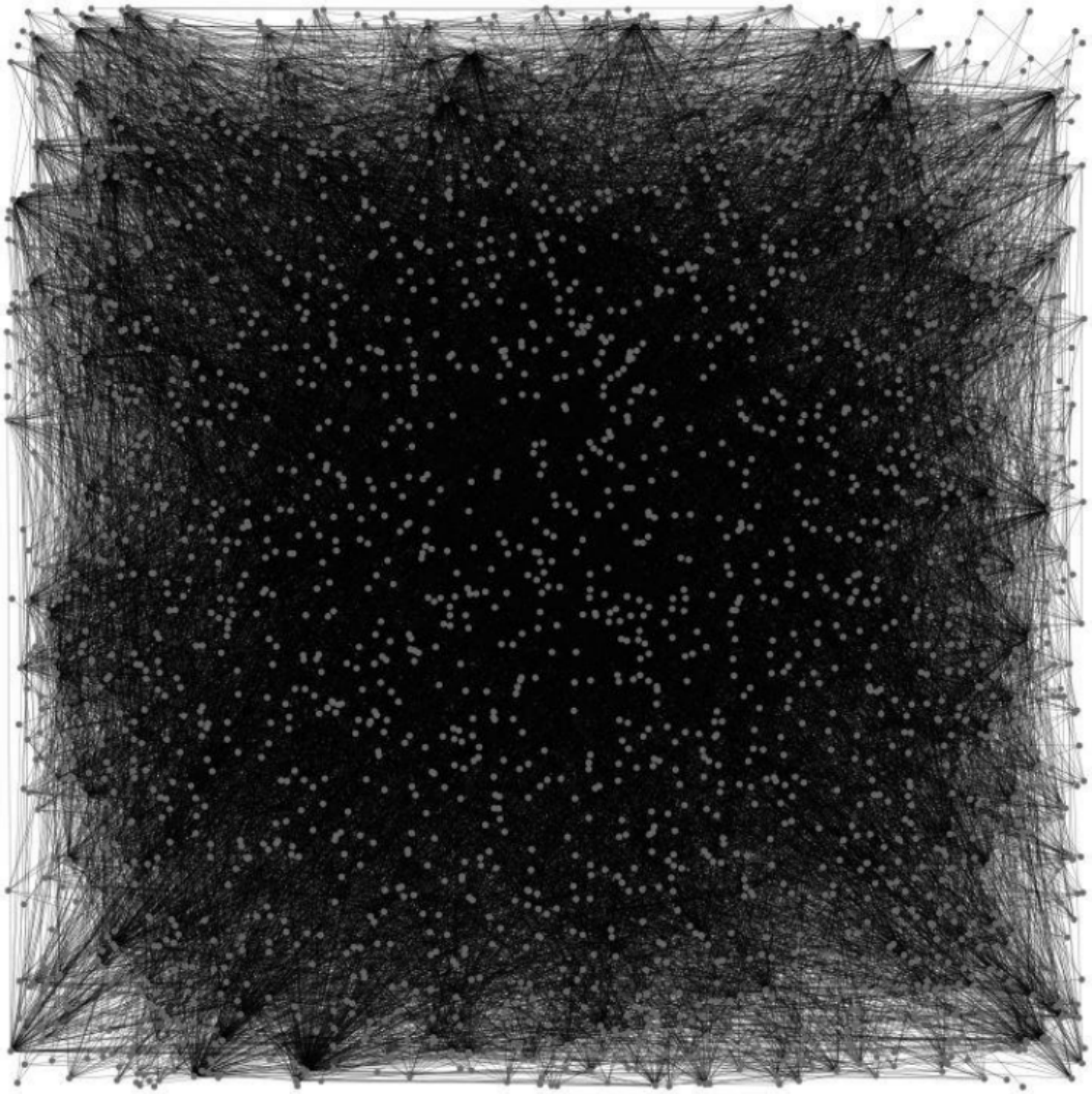
- Network (graph) $G = (V, E)$
- Objects (nodes/vertices) V
- Relations (links/edges) E
- Number of nodes — $|V|$ n
- Number of edges — $|E|$ m
- Examples:
  - Online social networks
  - Scientific citation and collaboration networks
  - Webgraphs
  - Biological networks
  - Communication networks
  - Financial networks
  - Corporate networks

Social network

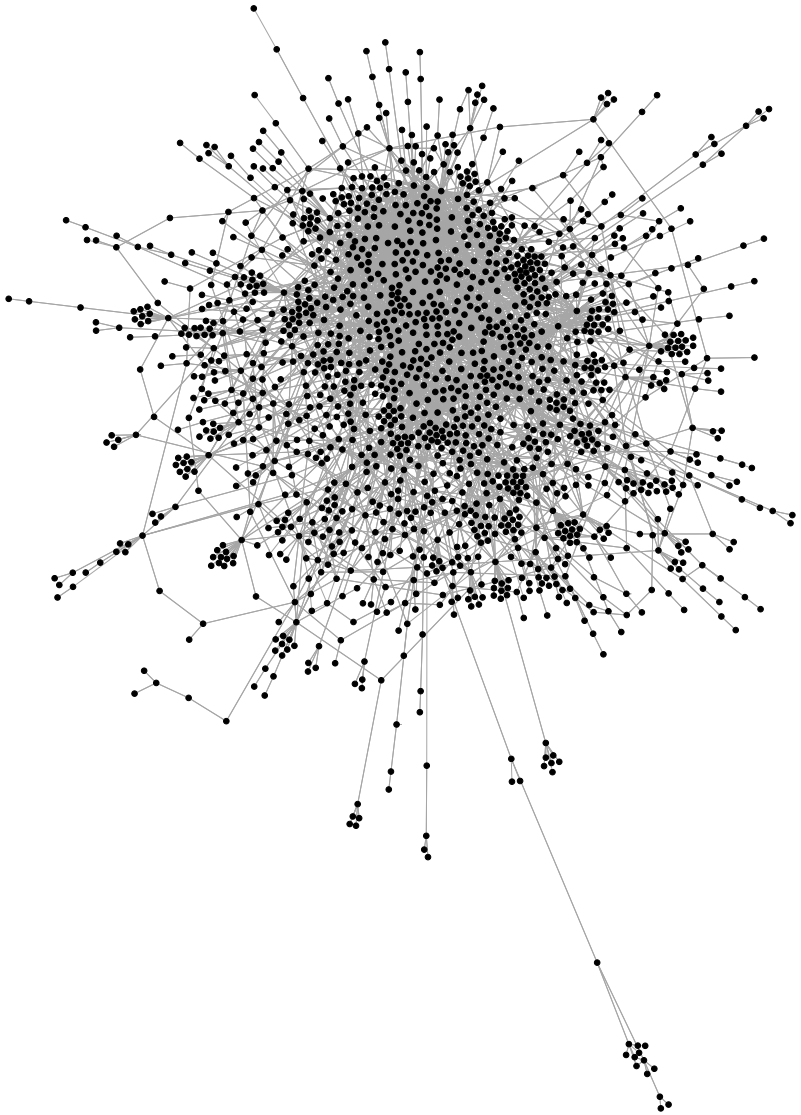


Source: `http://web.stanford.edu/class/cs224w`

Large network data



Visualization algorithms



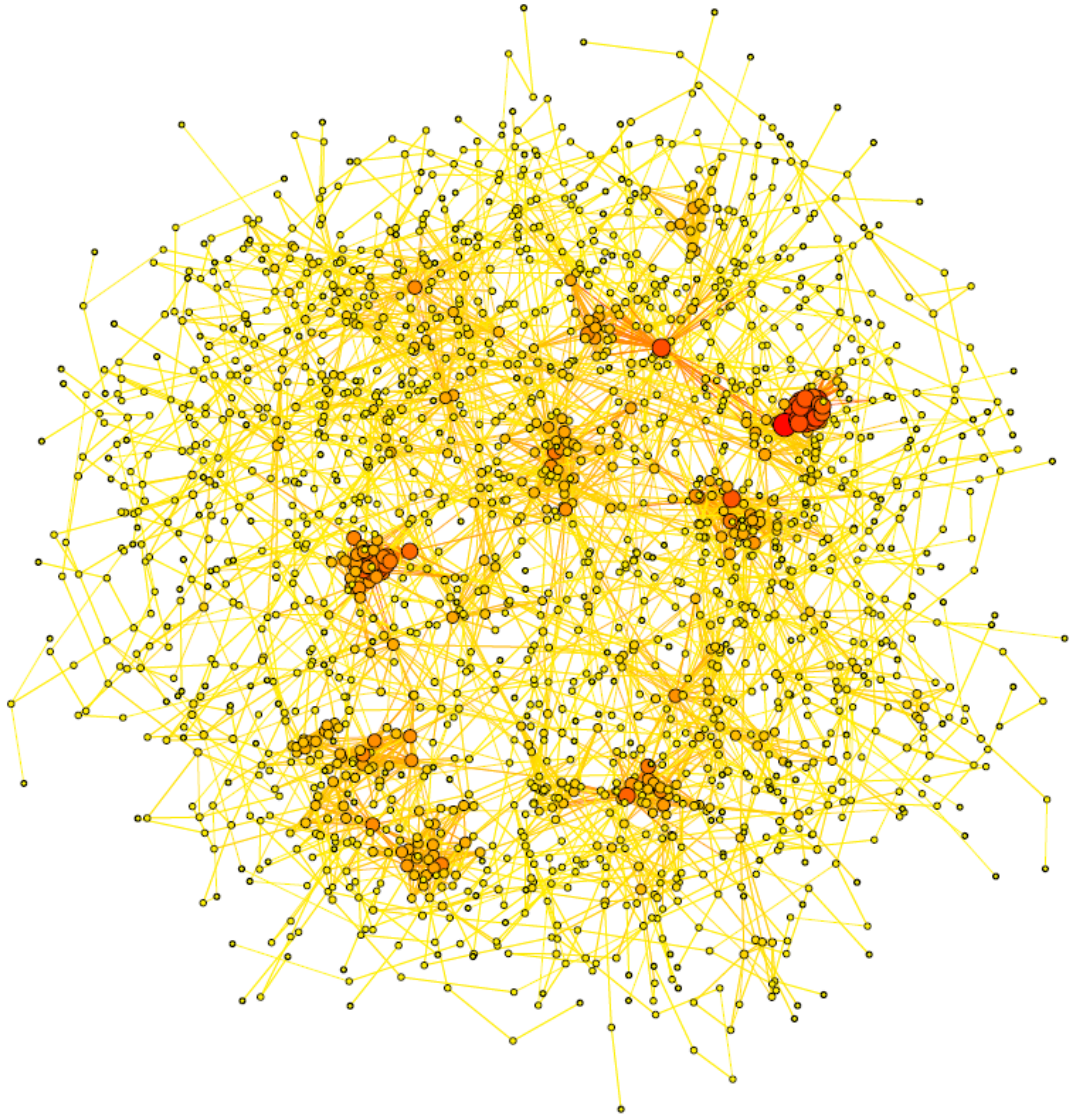


Figure : Sample of online social network with 1 876 nodes and 8 070 edges.

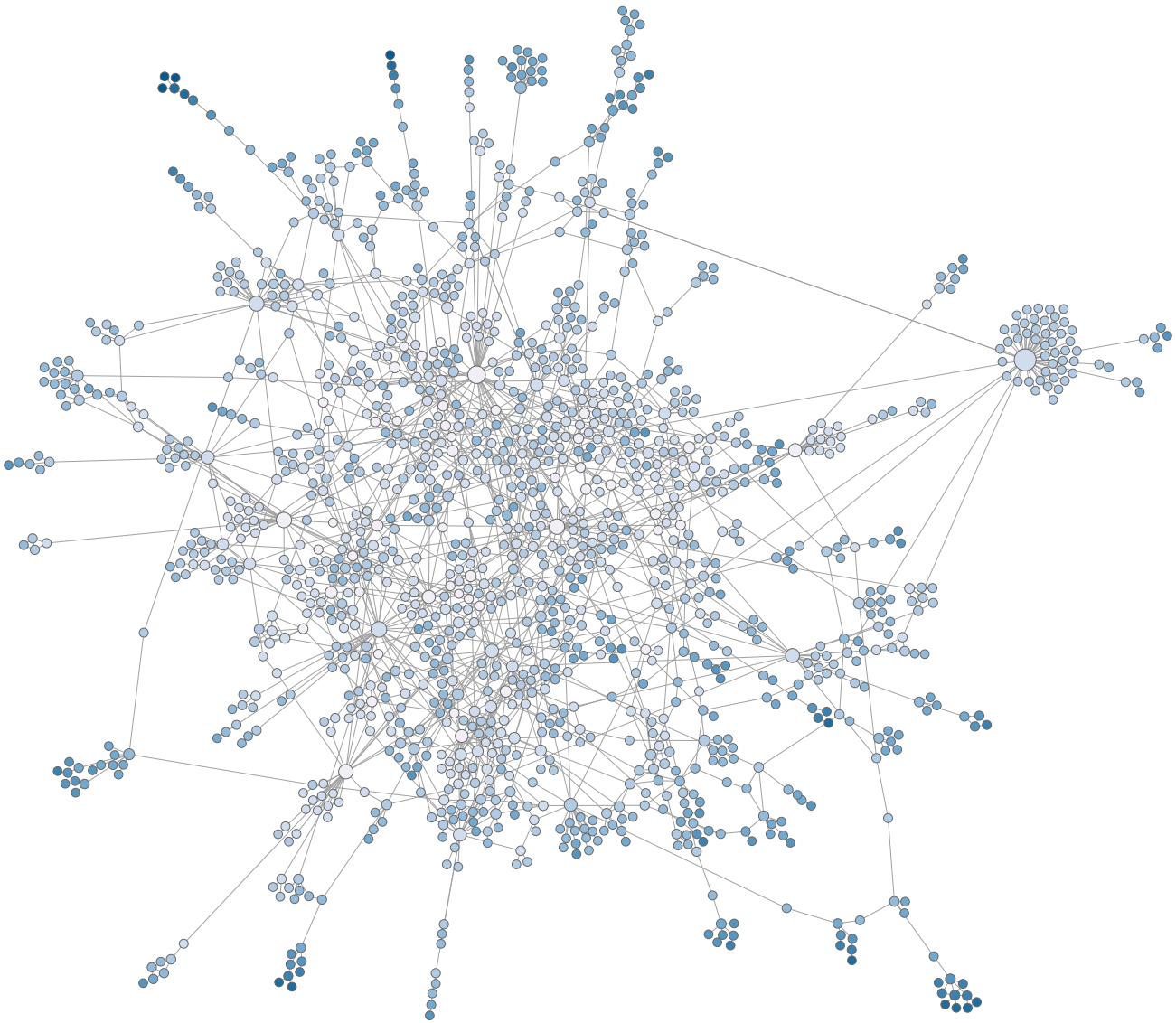


Figure : Yeast protein interaction network with 1 458 nodes and 1 948 edges

Types of graphs

- Directed vs undirected graphs
  - Reciprocity/Symmetry: extend to which directed links are mutual
- Weighted vs. unweighted graphs
  - Unweighted: weight of 1 for computational reasons
  - Signed networks: positive and negative weights
- Labeled (annotated) vs. unlabeled graphs
  - Labels on nodes and/or edges
- One-mode (homogenic) vs. two-mode networks
  Or: multi-mode (heterogenic) networks
- Static vs. dynamic (temporal) networks
  - Timestamps on nodes and/or edges

Community detection

- **Community**: set of nodes connected more strongly with eachother than with the rest of the network
- Community detection algorithms:
  - Clique-based methods
  - Hierarchical clustering
  - Divisive algorithms (centrality-based)
  - **Modularity maximization** algorithms

Community detection

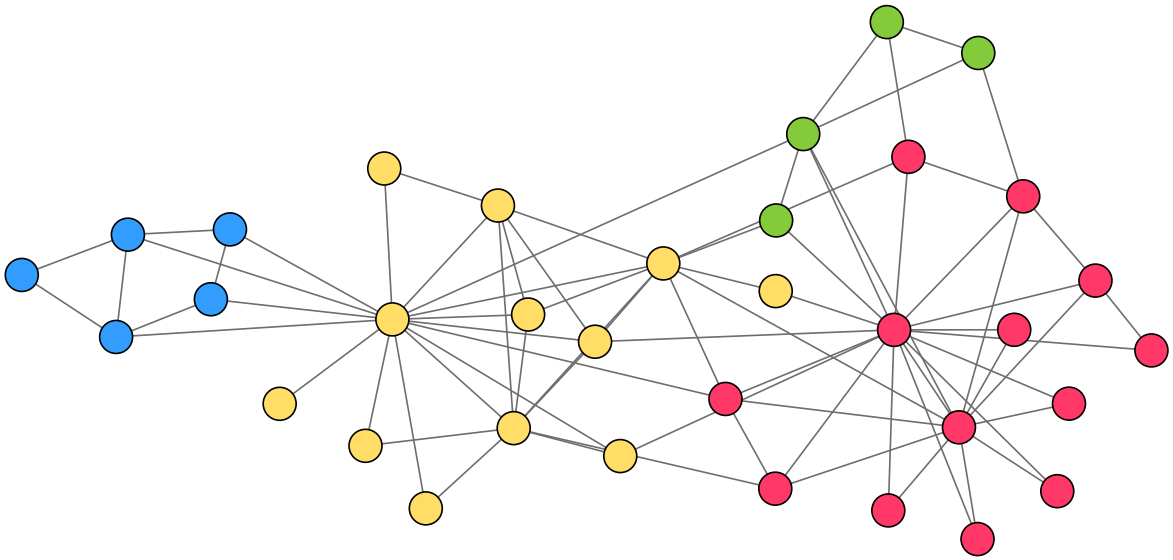


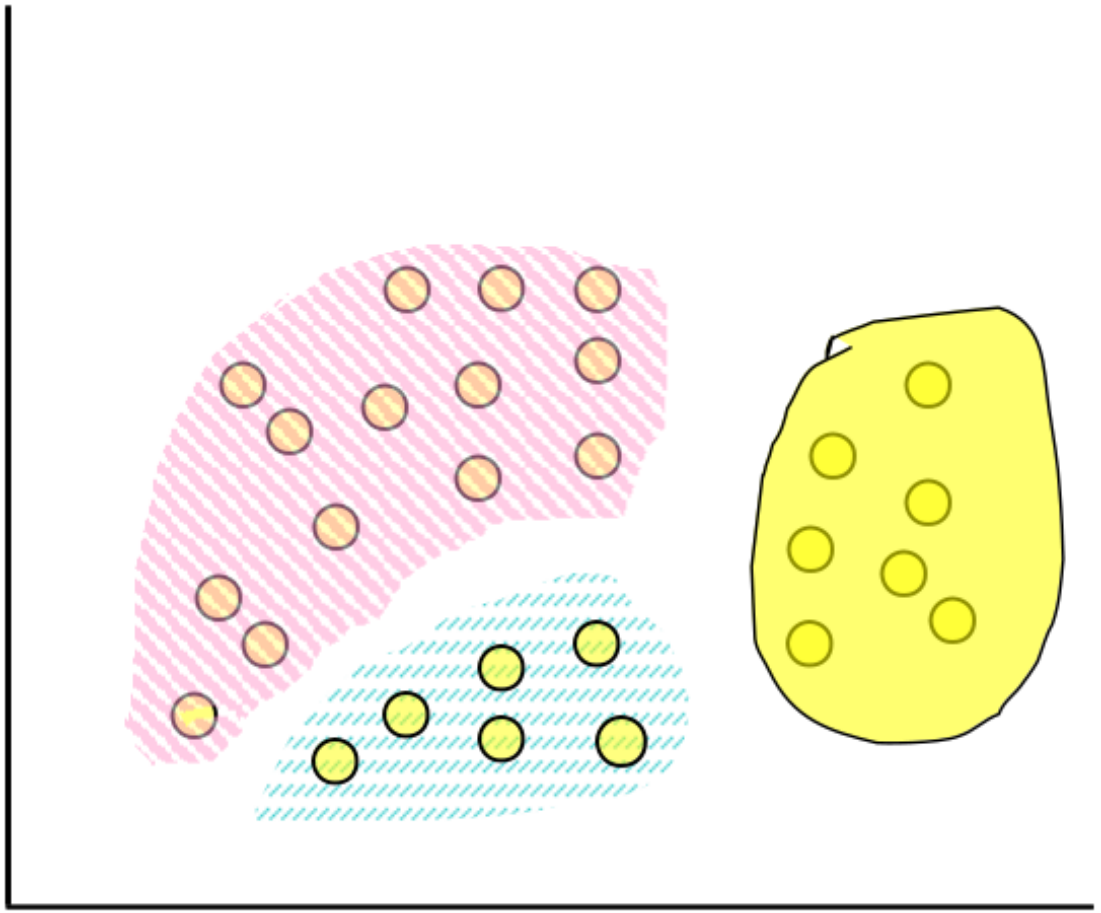
Figure : Communities: node subsets connected more strongly with each other

Modularity

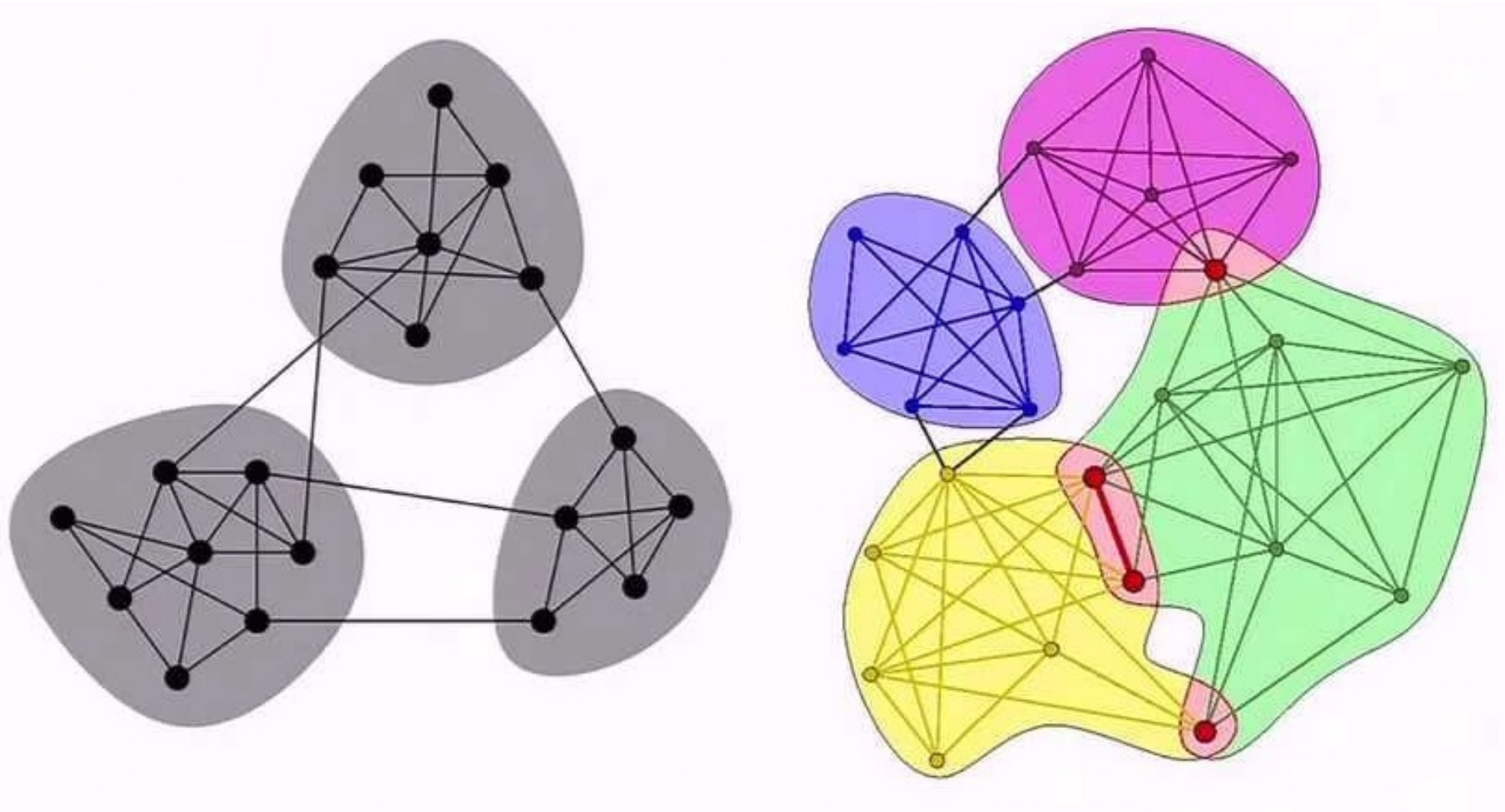
- **Modularity**: numerical value indicating the quality of a division of a network into communities
- **Community**: subset of nodes for which the fraction of links inside the community is higher than expected in a random network
- Modularity $Q \in [0, 1]$
- Resolution parameter r indicating how "tough" the algorithm should look for communities
- Algorithms optimize the modularity score Q given some r (using hill climbing, heuristics, genetic algorithms and many more optimization techniques)

V.D. Blondel, J-L. Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communities in large networks in *Journal of Statistical Mechanics: Theory and Experiment* 10: P10008, 2008.

Related: clustering



Partitions vs. communities



J. Leskovec, Affiliation Network Models for Densely Overlapping Communities, MMDS 2012.

Evaluating communities and partitions

- **Communities**: groups of nodes that are more connected amongst each other than with the other nodes of the network
- **Partitions**: non-overlapping communities
- Compare with groups of nodes based on common attributes
- Human interpretation by hand can suffer from subjective bias

Corporate networks

- **Nodes** are organizations/firms/companies/corporations

Corporate networks

- **Nodes** are organizations/firms/companies/corporations
- **Links** represent:
  - Trade
  - Loans
  - Ownership
  - Social ties

Corporate networks

- **Nodes** are organizations/firms/companies/corporations
- **Links** represent:
  - Trade
  - Loans
  - Ownership
  - Social ties

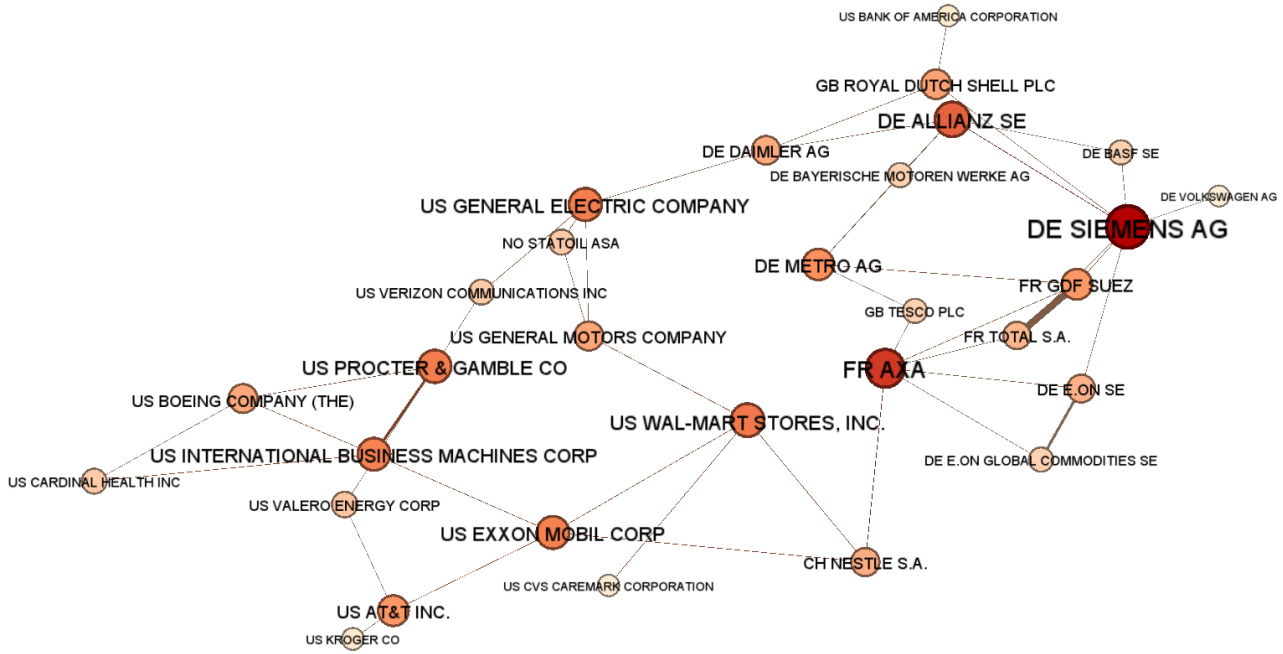


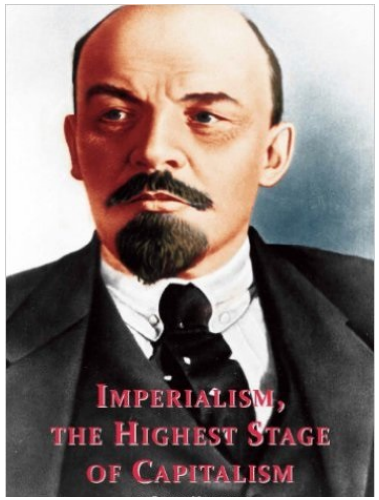
Figure : Board interlock network of 30 firms

Board interlocks

- **Board interlock**: there is a relationship between firms because they share a board member or director

Board interlocks

- **Board interlock**: there is a relationship between firms because they share a board member or director
- Vladimir I. Lenin, *Imperialism, The Highest Stage of Capitalism*, 1916.
- "... a personal union, so to speak, is established between the banks and the biggest industrial and commercial enterprises, the merging of one with another through the acquisition of shares, through the appointment of bank directors to the Supervisory Boards (or Boards of Directors) of industrial and commercial enterprises, and vice versa."



Board interlocks

- **Causes** of interlocks:
  - Collusion
  - Cooptation and monitoring
  - Legitimacy
  - Career advancement
  - Social cohesion
- **Consequences** of interlocks:
  - Corporate control
  - Economic performance
  - Access to resources

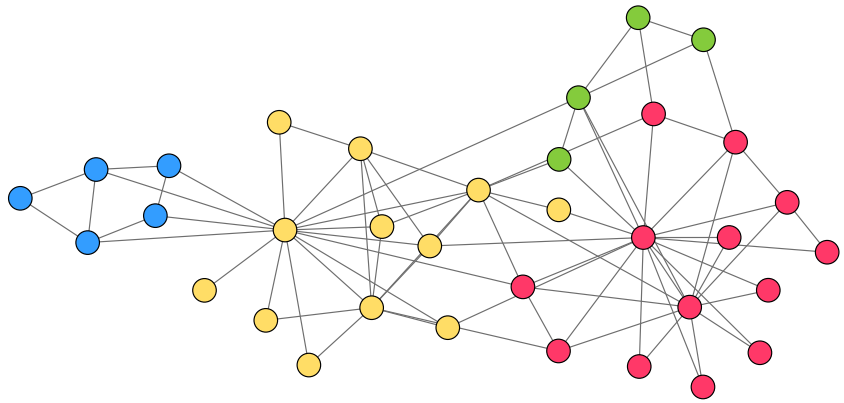
Board interlocks

- **Causes** of interlocks:
  - Collusion
  - Cooptation and monitoring
  - Legitimacy
  - Career advancement
  - Social cohesion
- **Consequences** of interlocks:
  - Corporate control
  - Economic performance
  - Access to resources

M. Mizruchi, What do interlocks do? An analysis, critique, and assessment of research on interlocking directorates, *Annual review of Sociology* 22: 271–298, 1996.

Corporate board interlock networks

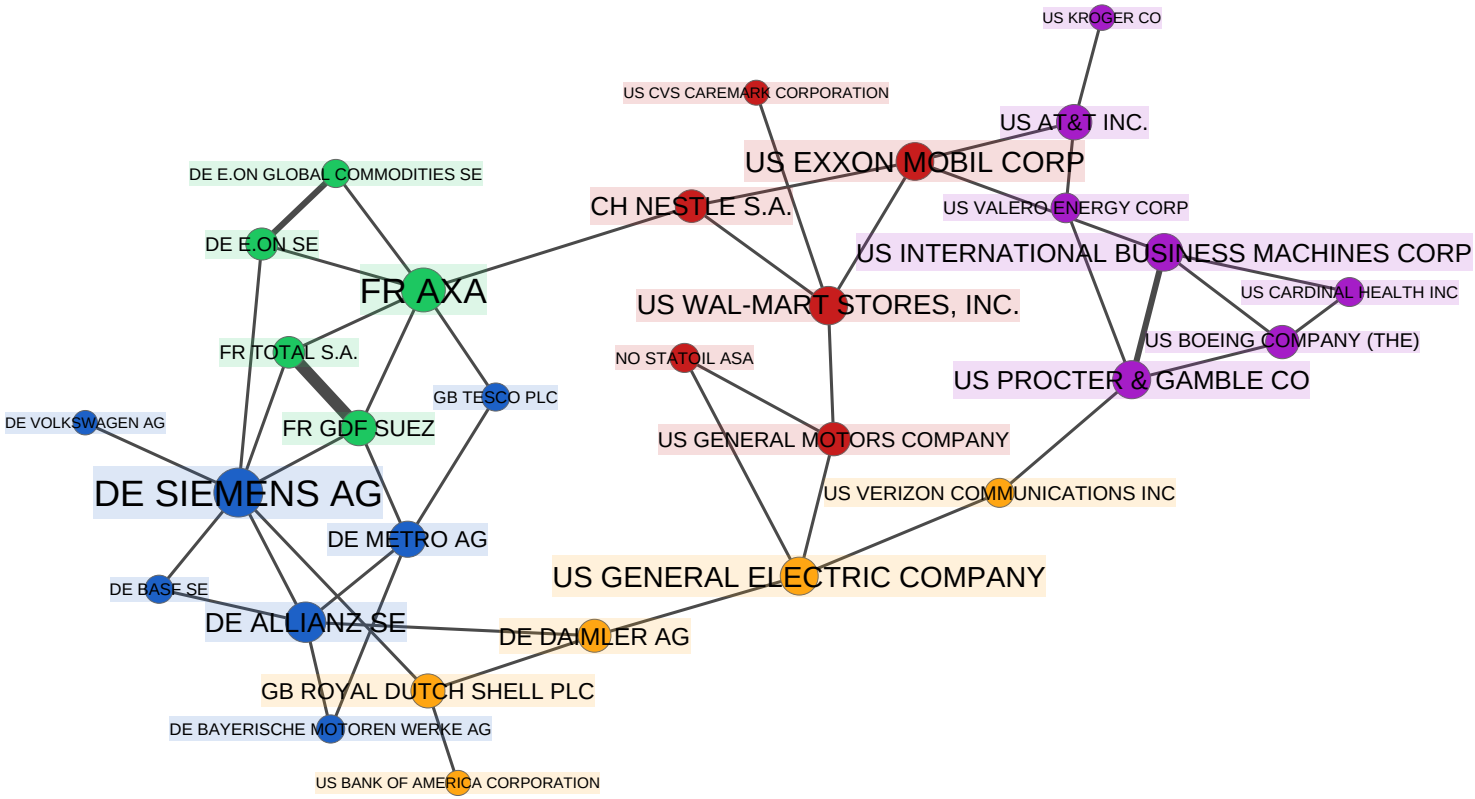
- Nodes are firms, edges represent shared board members
- Power and control
- Globalization
- Influence of countries
- Local or global business groups?
- **Community detection** in networks

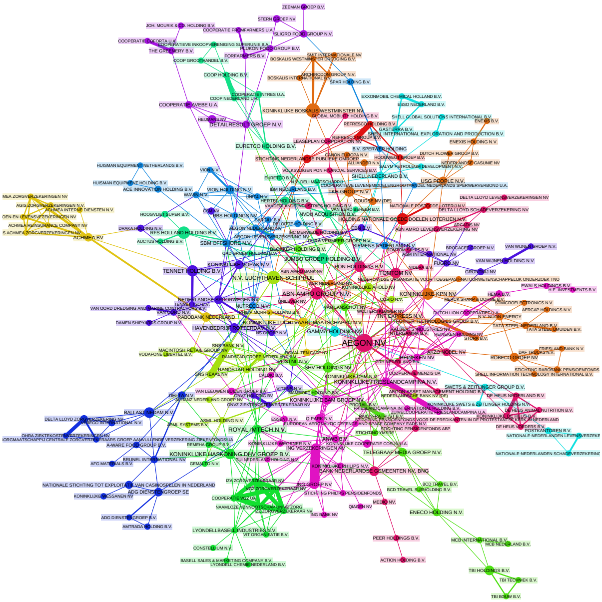


Corporate board interlock networks

- Nodes are **firms** (world-wide)
- Links are social ties between firms (board interlocks)
- **Board interlock**: two firms share a senior level director
- 1,068,409 firms
- 3,262,413 interlocks
- 80 countries
- National subgraphs based on country attribute

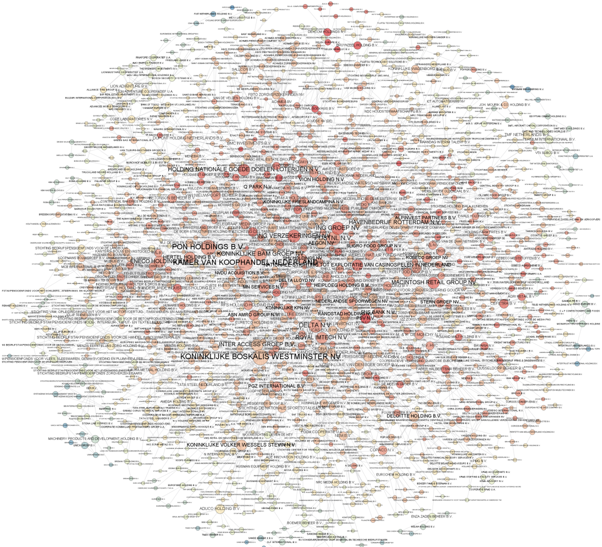
Corporate board interlock networks





Universiteit Leiden

Corporate network NL (1948 nodes)

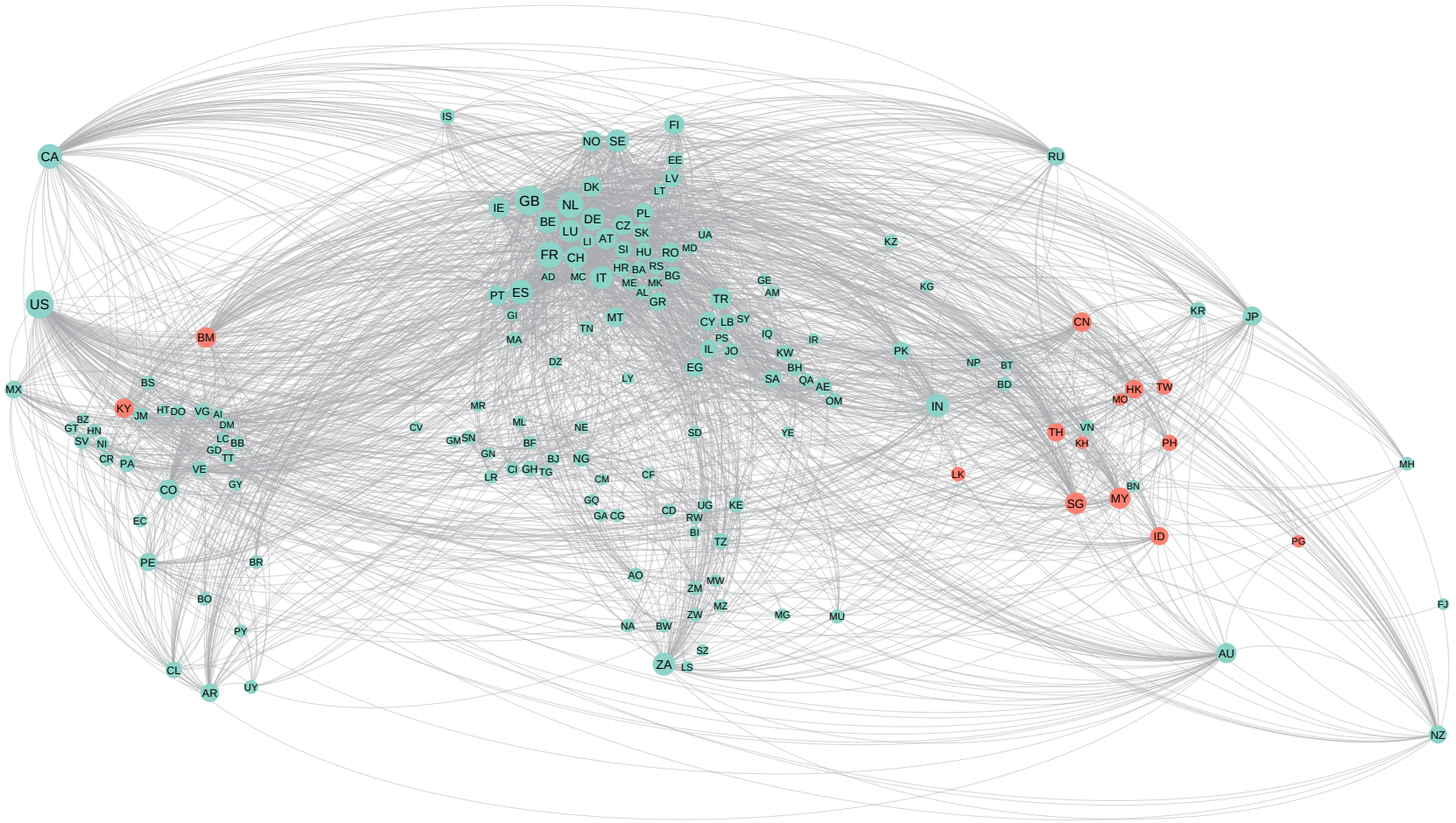


Corporate country network

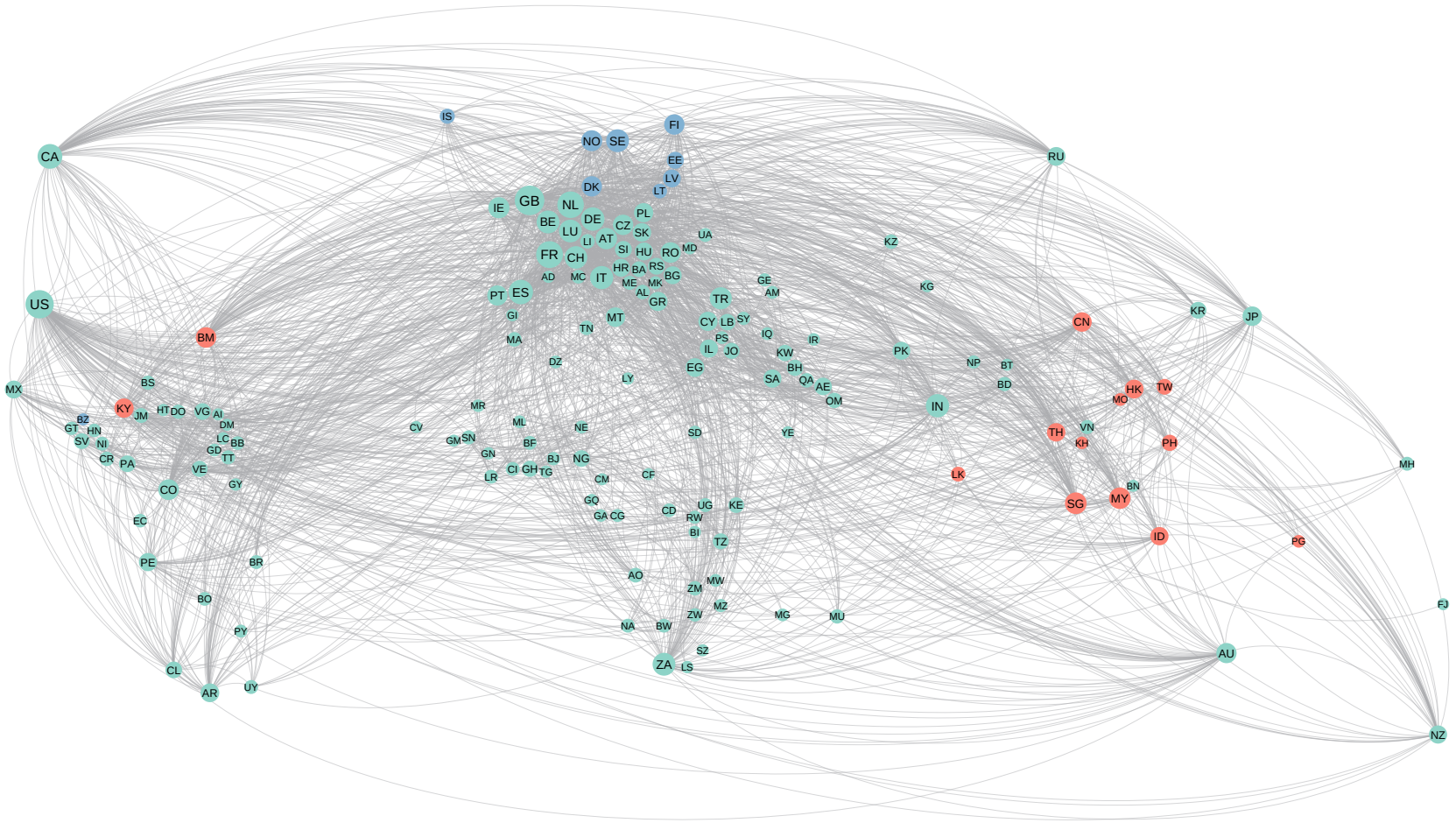
- Aggregation of corporate network
- Nodes are **countries**
- Edges represent social interaction between firms
- Layout on geographic Mercator map
- Communities represent social ties between countries

Corporate network communities

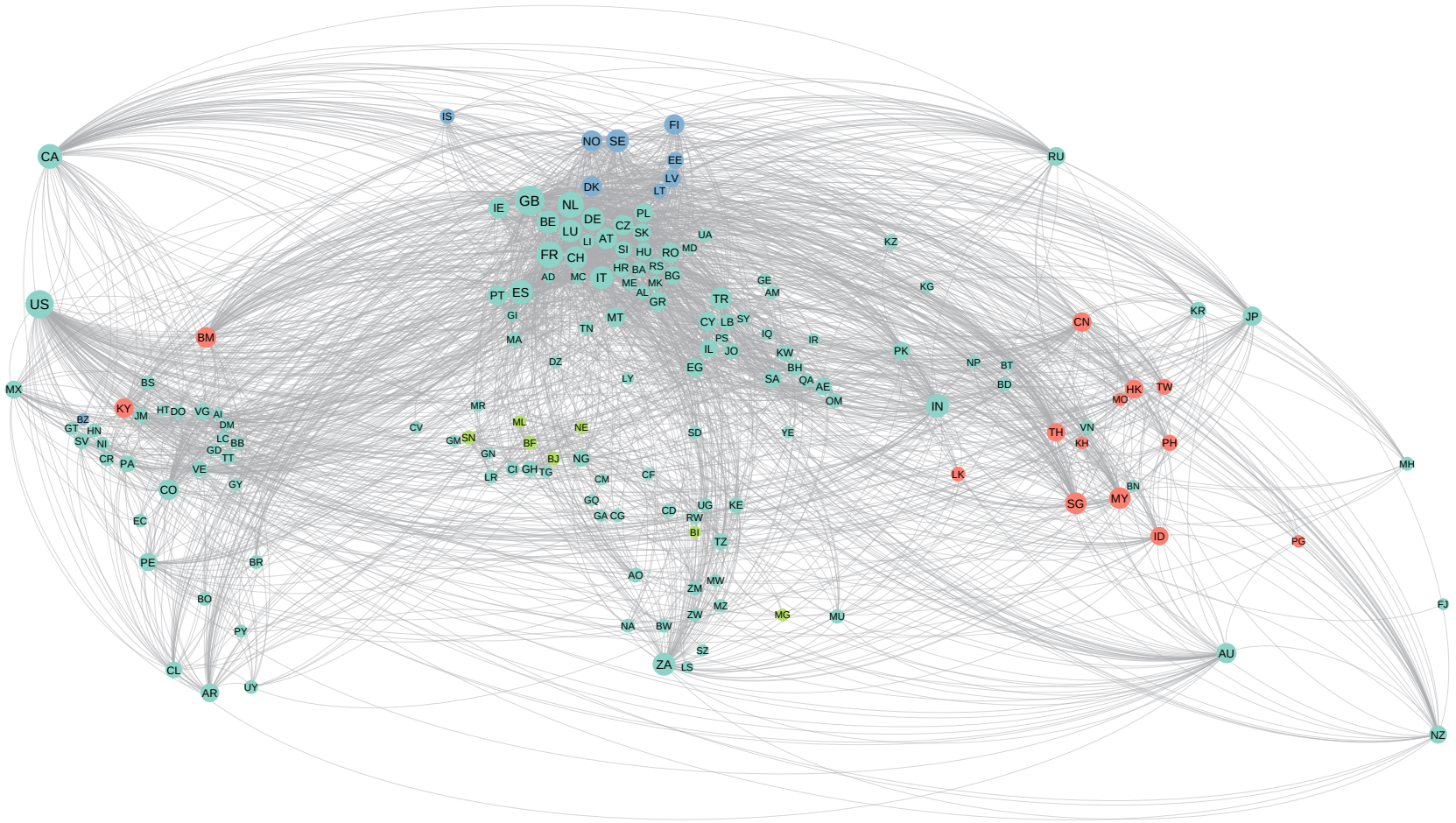
Communities, resolution = 2.0



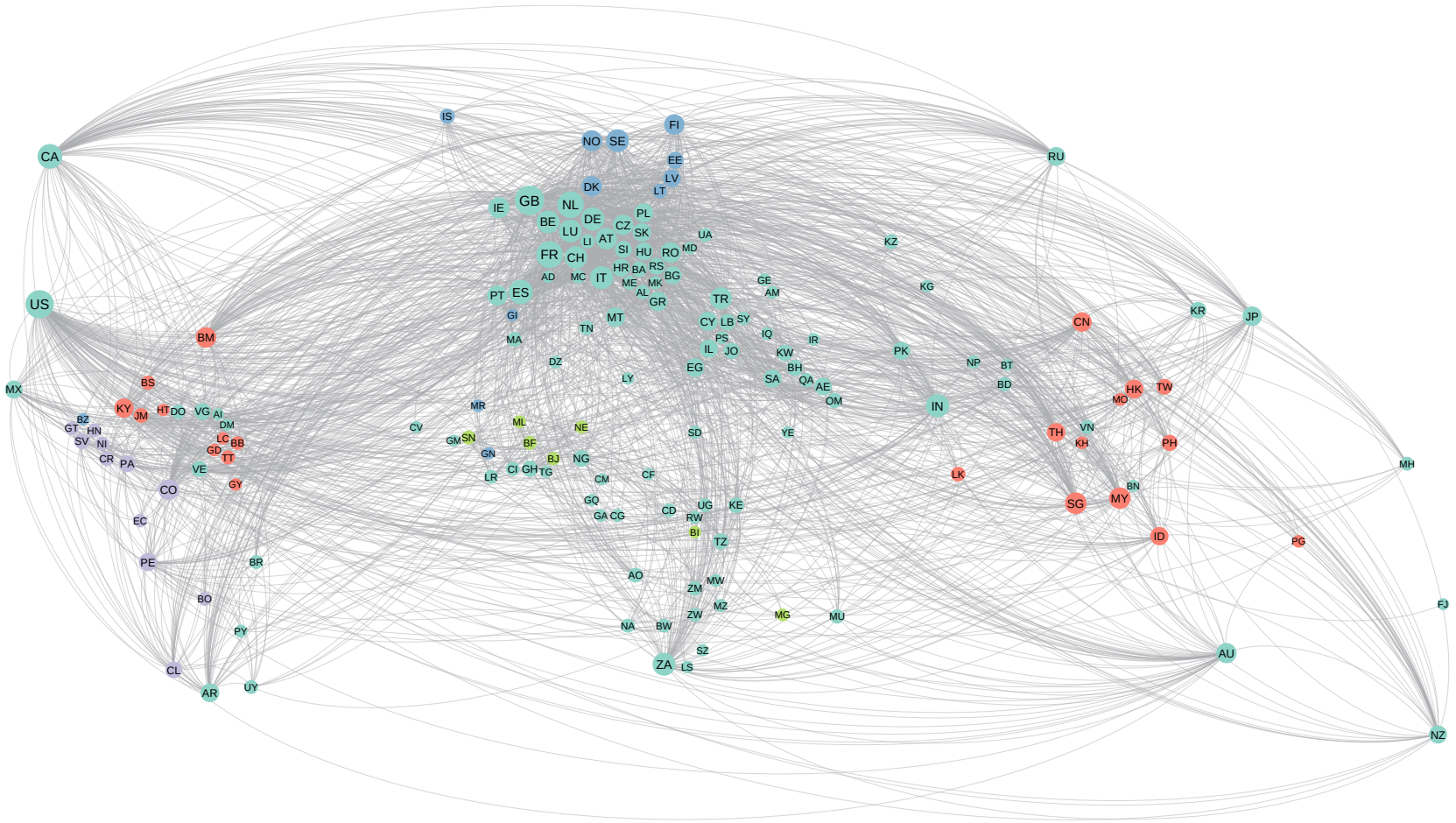
Communities, resolution = 1.75



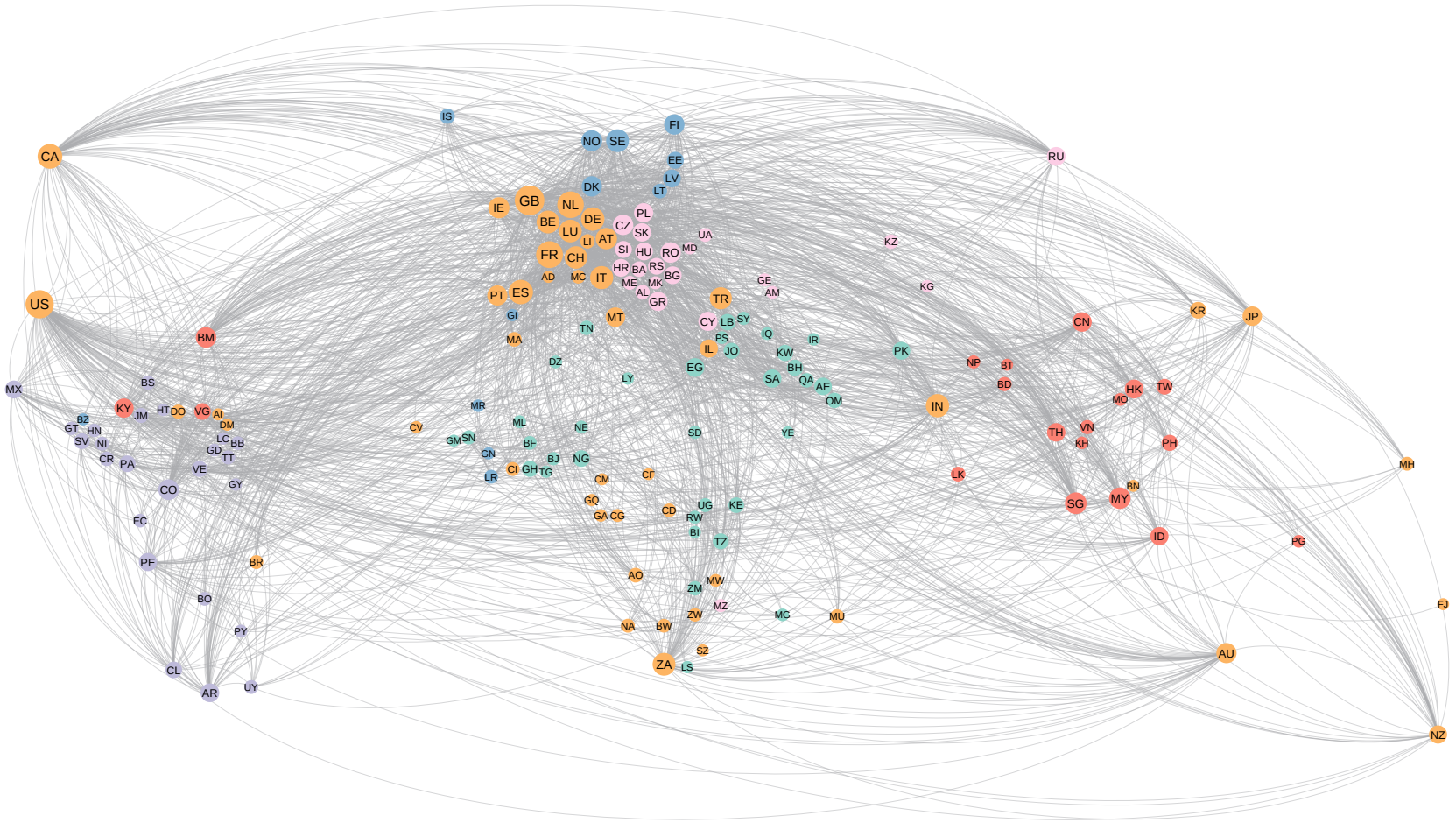
Communities, resolution = 1.55



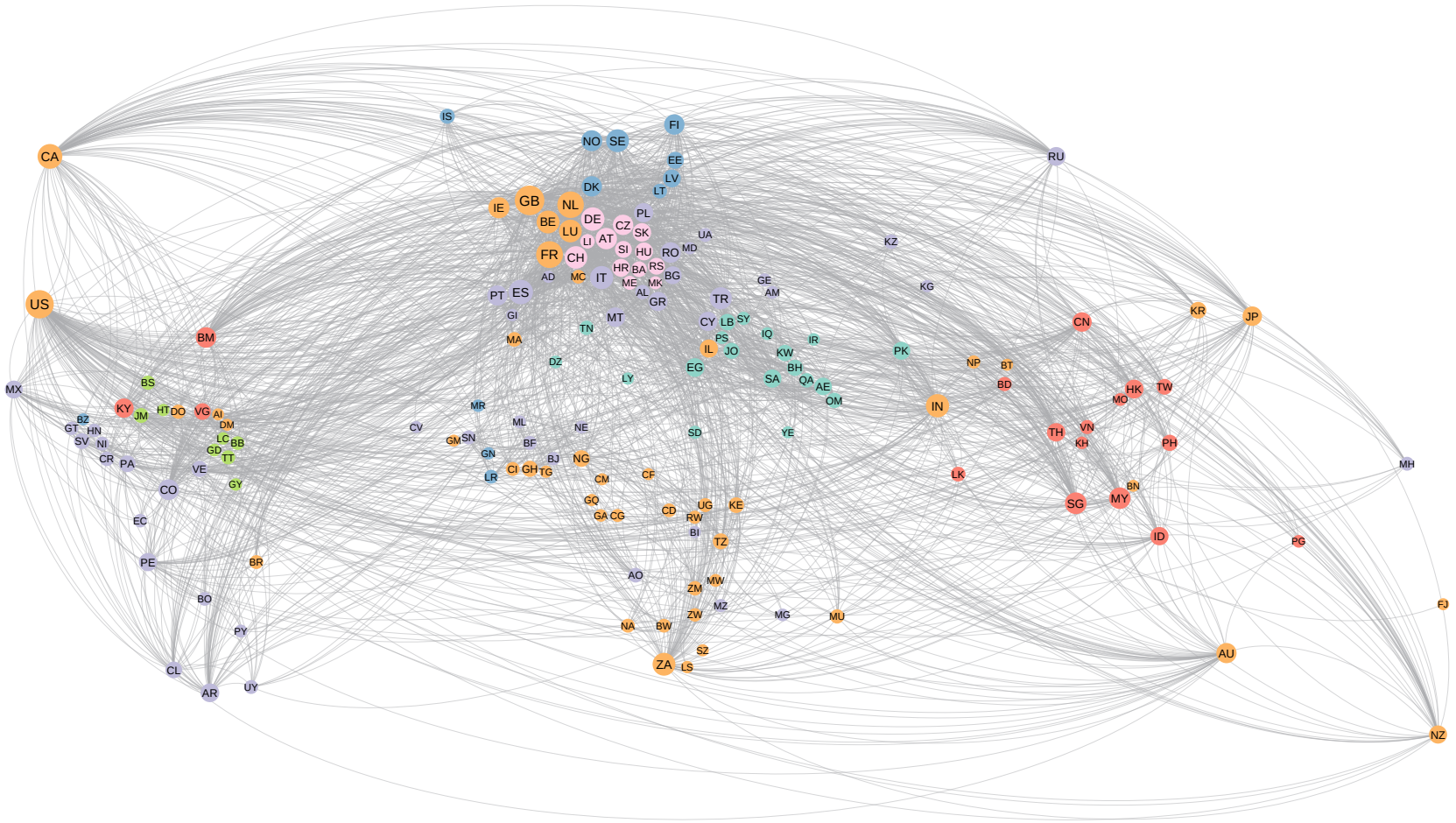
Communities, resolution = 1.5



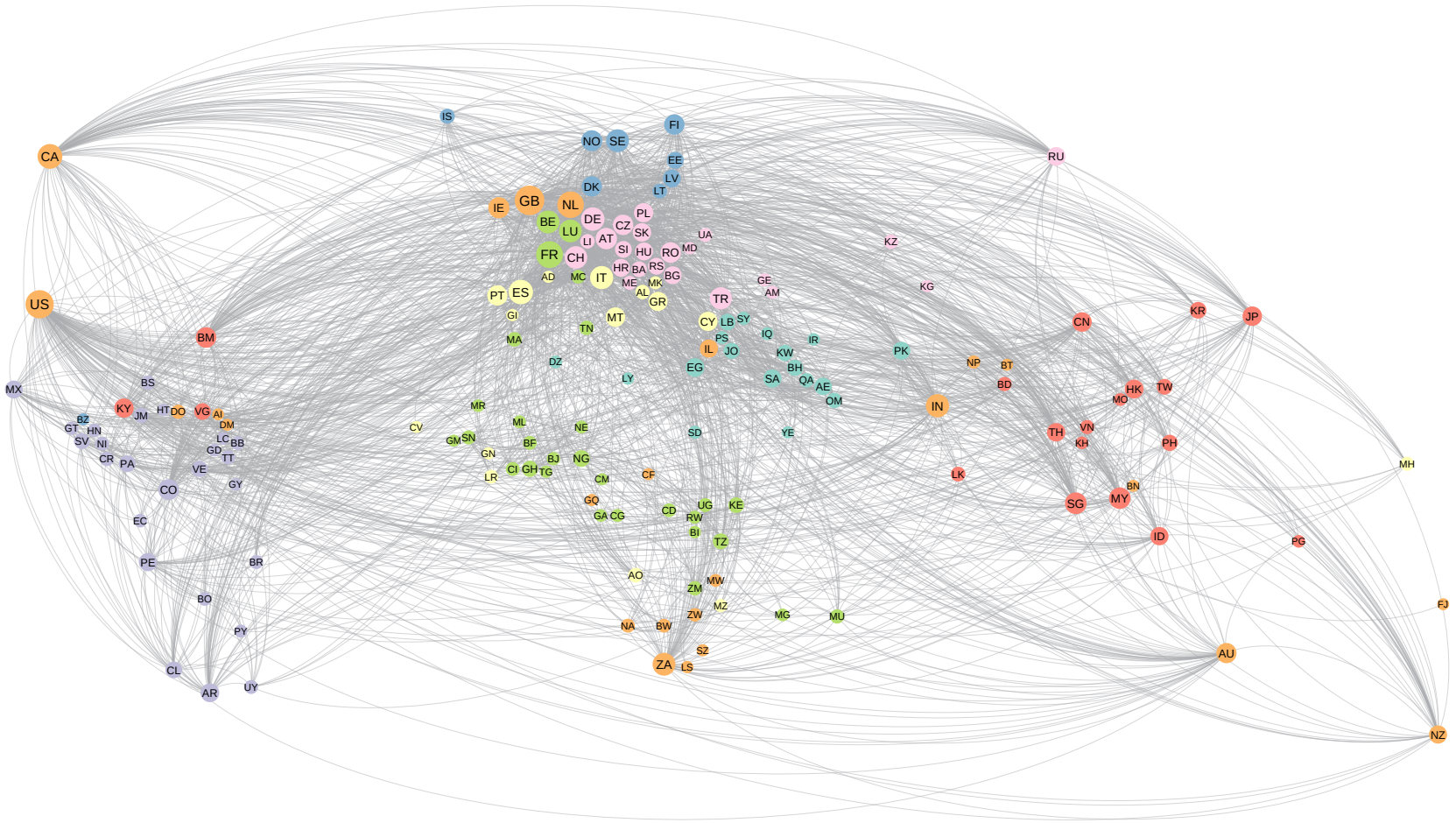
Communities, resolution = 1.0



Communities, resolution = 0.75



Communities, resolution = 0.5



Appearing communities

1. Eastern Asia cluster
2. Baltic cluster
3. Former French colonies
4. Spanish/Portugese LatAm ties
5. Western/Eastern Europe
6. UK/US ties with NL, financial world

Universiteit Leiden

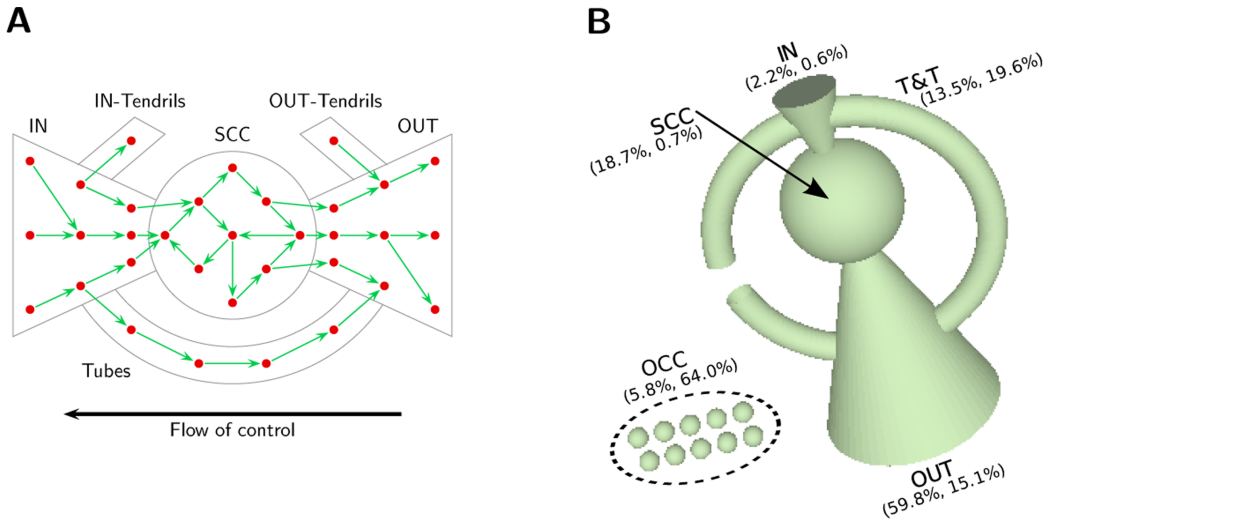
Network flow

Uncovering Offshore Financial Centers in the Global Corporate Ownership Network

Corporate networks

- **Nodes** are organizations/firms/companies/corporations
- **Links** represent:
  - Trade
  - Loans
  - **Ownership**
  - Board interlocks

Ownership network



S. Vitali, J.B. Glattfelder and S. Battiston, S, The network of global corporate control, *PloS one* 6(10), e25995, 2011.

Offshore financial centers

- **Offshore Financial Center** (OFC): jurisdiction (country) that attracts financial activities from abroad through low taxation and lenient regulation.

Offshore financial centers

- **Offshore Financial Center** (OFC): jurisdiction (country) that attracts financial activities from abroad through low taxation and lenient regulation.
- OFCs "process" around $\$6\,000\,000\,000\,000$ yearly

Offshore financial centers

- **Offshore Financial Center** (OFC): jurisdiction (country) that attracts financial activities from abroad through low taxation and lenient regulation.
- OFCs "process" around $6 000 000 000 000 yearly
- Problematic because:
  - Unnecessary system complexity
  - Accountability
  - Tax avoidance

Characterizing OFCs

- So, which countries are OFCs?

Characterizing OFCs

- So, which countries are OFCs?
  - Heavily debated
  - Definitions differ
  - Highly contested and politicized

Characterizing OFCs

- So, which countries are OFCs?
  - Heavily debated
  - Definitions differ
  - Highly contested and politicized
- FDI/GDP ratio approach:
  substantially more Foreign Direct Investment than expected based on GDP? $\rightarrow$ OFC

Characterizing OFCs

- So, which countries are OFCs?
  - Heavily debated
  - Definitions differ
  - Highly contested and politicized
- FDI/GDP ratio approach:
  substantially more Foreign Direct Investment than expected based on GDP? $\rightarrow$ OFC...
  Problematic because:
  - FDI systematically underreported and subject to politics
  - No exact money flows, just A to B
  - OFC homogeinity assumption

Characterizing OFCs

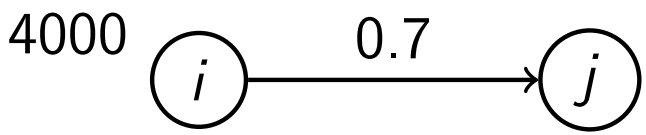
- So, which countries are OFCs?
  - Heavily debated
  - Definitions differ
  - Highly contested and politicized
- FDI/GDP ratio approach:
  substantially more Foreign Direct Investment than expected based on GDP? $\rightarrow$ OFC...
  Problematic because:
  - FDI systematically underreported and subject to politics
  - No exact money flows, just A to B
  - OFC homogeinity assumption
- Solution: **Networks!**

Data

- Orbis database
- Based on data at Chambers of Commerce
- 71,201,304 ownership links between firms
- Better coverage for high-income countries
- Poor data quality $\rightarrow$ fiscal secrecy
- Findings likely represent lower bound

Ownership network

- Ownership network $N = (V, E)$
- Nodes are firms $i \in V$
  - Have value associated to them R_i
  - Are based in a particular country $\phi(i)$
- Directed links indicate ownership relation / value flow $(i, j) \in E$
  - Have an associated weight in $[0; 1]$ $w(i, j)$
- $w(i, j) = 0.7$, $R_i = 4000$ gives:



Ownership paths and value

- Ownership path

$$p = (v_1, v_2, \ldots, v_\ell) \text{ with } (v_i, v_{i+1}) \in E \text{ for } 1 \leq i < \ell$$

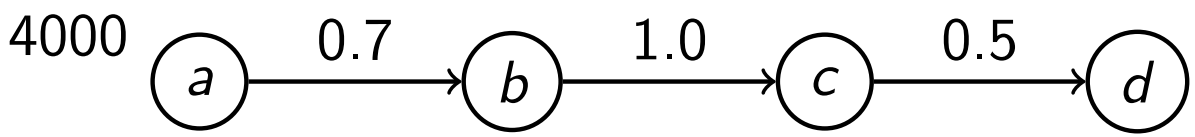
- Multiplicative ownership

$$w_p = \prod_{i=1}^{\ell-1} w(v_i, v_{i+1})$$

- Path value

$$V_p = R_{v_1} \cdot w_p$$

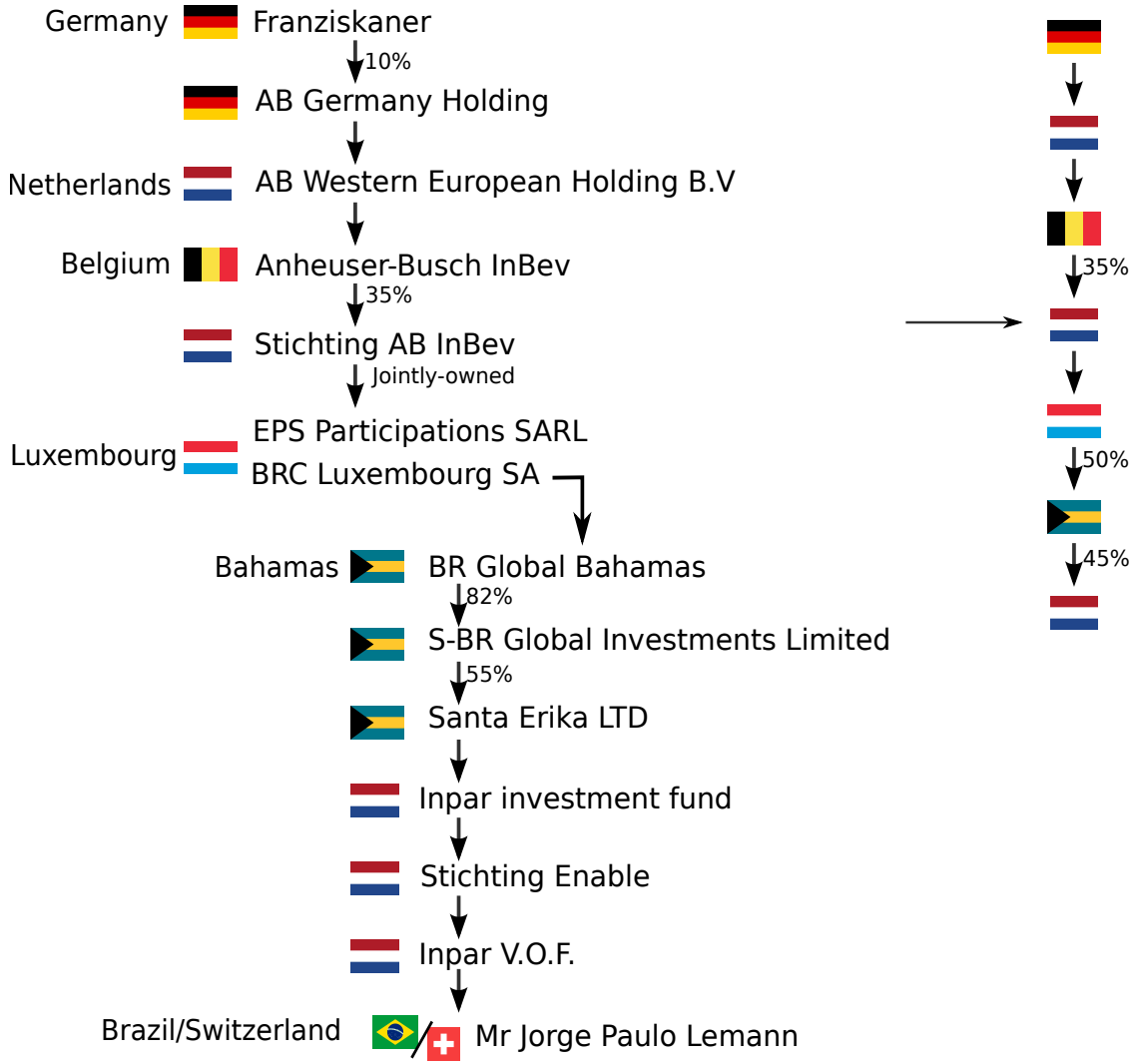
- $p = (a, b, c, d)$ and $w_p = 0.35$ and $V_p = 1400$



Ownership paths $\rightarrow$ OFCs

Ownership paths → OFCs

Ownership paths → OFCs



Ownership chains

- An **ownership chain** of a firm v is an ownership path p which:
  1. starts at node v,
  2. is a simple path (has no repeated nodes),
  3. has multiplicative ownership value greater than θ, i.e., $w_p \geq \theta$ and
  4. is maximal in length, i.e., cannot be extended by adding another node.

Ownership chains

- An **ownership chain** of a firm v is an ownership path p which:
  1. starts at node v,
  2. is a simple path (has no repeated nodes),
  3. has multiplicative ownership value greater than θ, i.e., $w_p \geq \theta$ and
  4. is maximal in length, i.e., cannot be extended by adding another node.
- Set of all ownership chains starting at v is denoted C_v
- Enumeration is possible using a DFS search for each node

Ownership chains

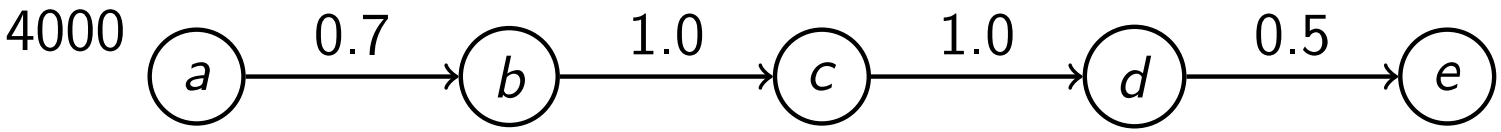
- An **ownership chain** of a firm v is an ownership path p which:
  1. starts at node v,
  2. is a simple path (has no repeated nodes),
  3. has multiplicative ownership value greater than θ, i.e., $w_p \geq \theta$ and
  4. is maximal in length, i.e., cannot be extended by adding another node.
- Set of all ownership chains starting at v is denoted C_v
- Enumeration is possible using a DFS search for each node
- Ultimately, $C = \bigcup_{v \in V} C_v$ is the set of all ownership chains

Ownership chunks

- **Ownership chunk**: subpath of length $2, 3, \ldots, \ell$ (generated from a chain of length ℓ)
- From all chains C, generate all ownership chunks, forming set H
- Ownership chunks of length x: H^x.
- Chunk $q \in H$ has an associated value $V^p(q)$ or in short V^p_q.

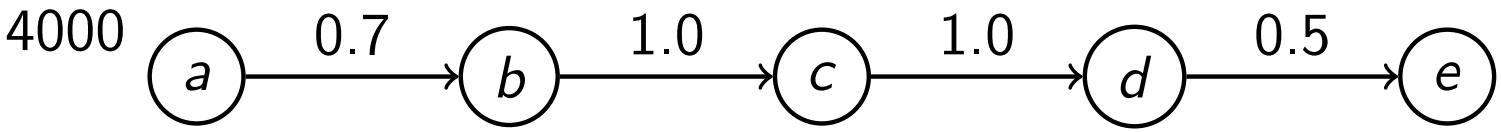
Ownership chunks

- **Ownership chunk**: subpath of length $2, 3, \ldots, \ell$ (generated from a chain of length ℓ)
- From all chains C, generate all ownership chunks, forming set H
- Ownership chunks of length x: H^x.
- Chunk $q \in H$ has an associated value $V^p(q)$ or in short V^p_q.



Ownership chunks

- **Ownership chunk**: subpath of length $2, 3, \ldots, \ell$ (generated from a chain of length ℓ)
- From all chains C, generate all ownership chunks, forming set H
- Ownership chunks of length x: H^x.
- Chunk $q \in H$ has an associated value $V^p(q)$ or in short V^p_q.



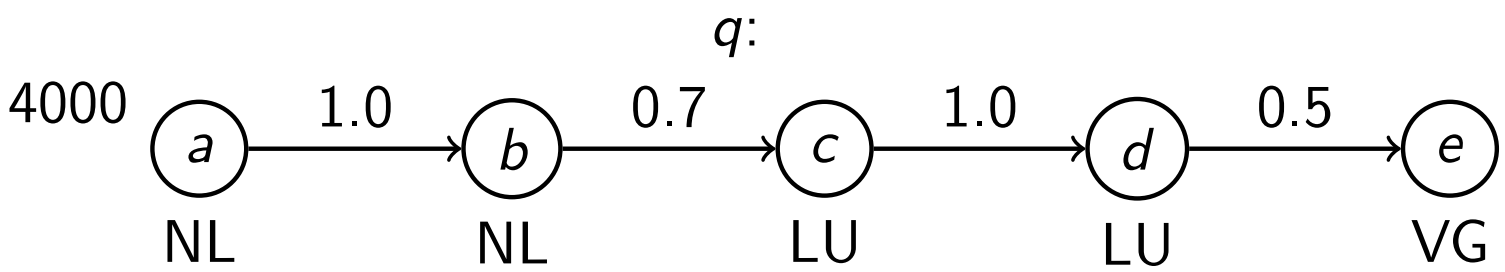
- Chunks of length 3 of $p = (a, b, c, d, e)$:
  - $q_1 = (a, b, c)$ with $V^p_{q_1} = 2800$
  - $q_2 = (b, c, d)$ with $V^p_{q_2} = 2800$
  - $q_2 = (c, d, e)$ with $V^p_{q_3} = 1400$

Country chains

- Map each node v in a chunk q to its country $\phi(v)$
- Merge subsequent same country nodes into **country chain** g
- New valuation function $V^{\phi}(g)$ for country chains, aggregating value through chunks with different value origins to the country level

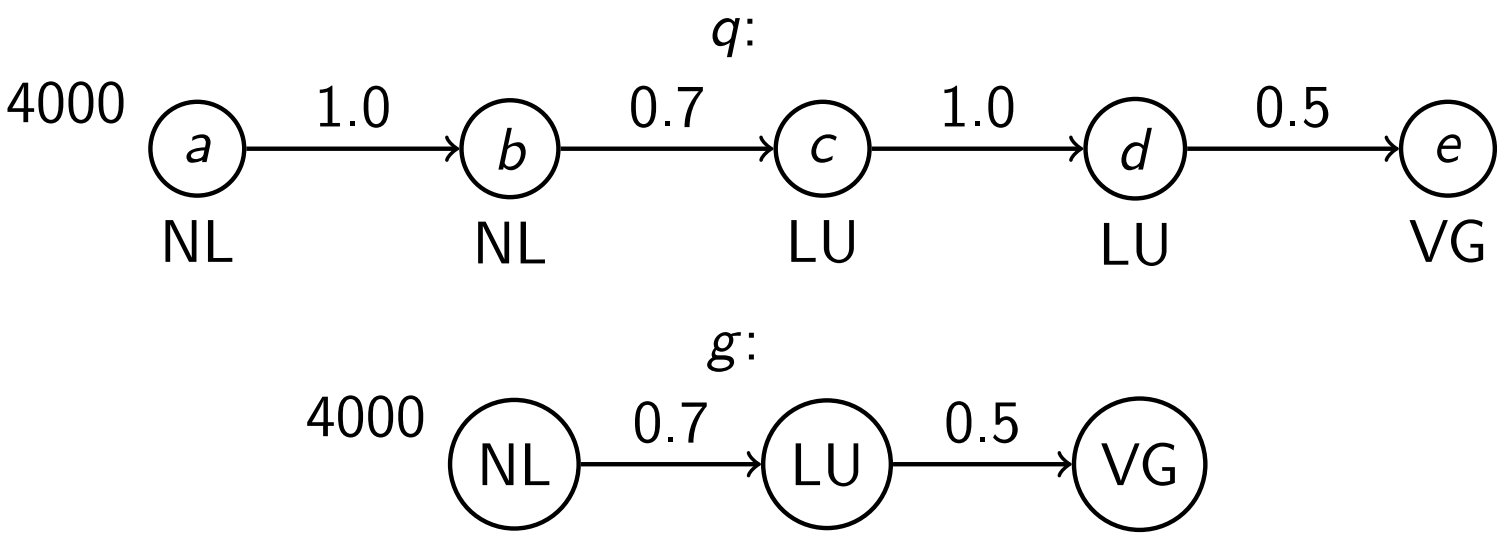
Country chains

- Map each node v in a chunk q to its country $\phi(v)$
- Merge subsequent same country nodes into **country chain** g
- New valuation function $V^{\phi}(g)$ for country chains, aggregating value through chunks with different value origins to the country level

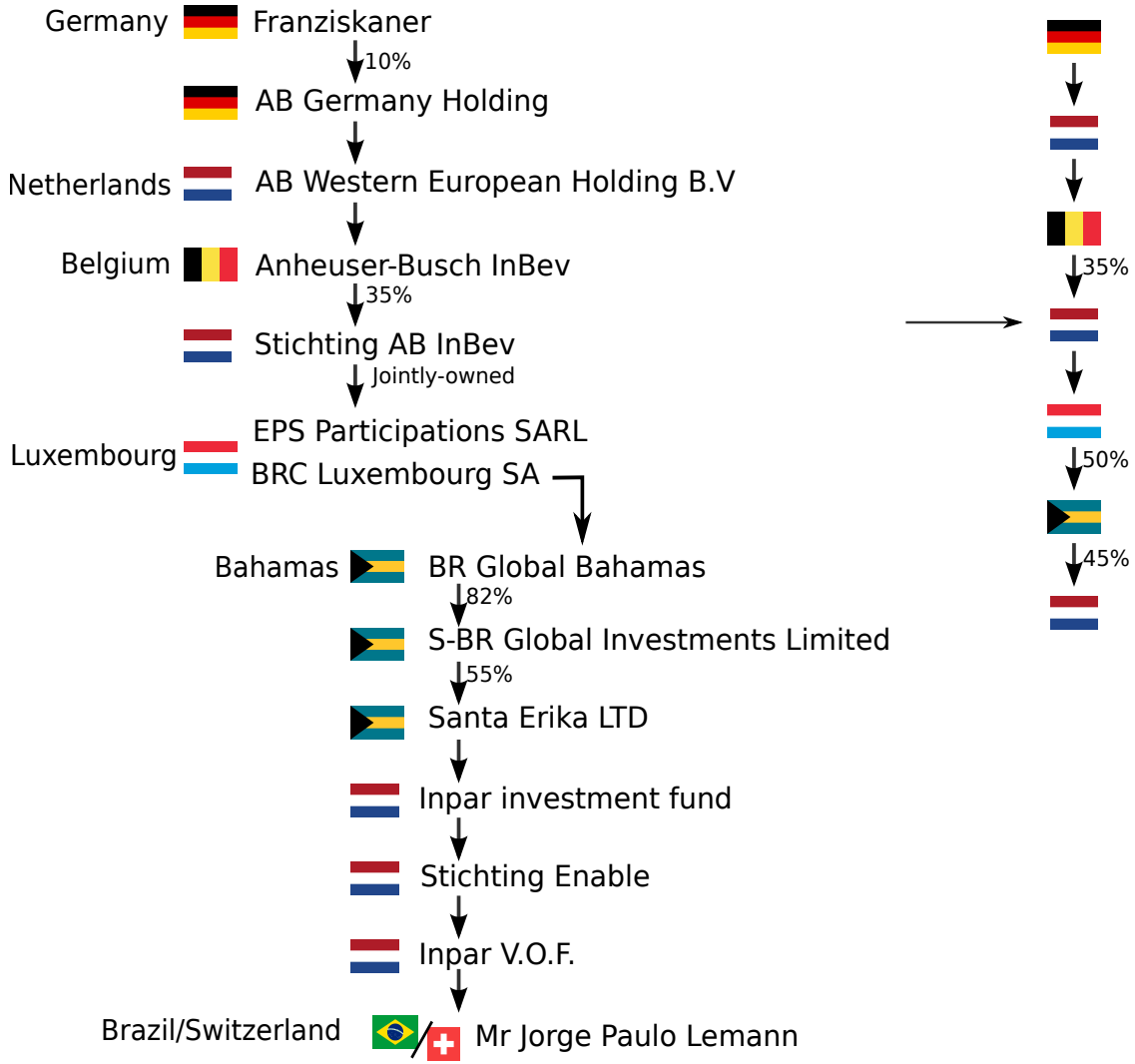


Country chains

- Map each node v in a chunk q to its country $\phi(v)$
- Merge subsequent same country nodes into **country chain** g
- New valuation function $V^{\phi}(g)$ for country chains, aggregating value through chunks with different value origins to the country level



Country chains



Results on our data

- 71,201,304 ownership links between firms
- 11,404,819 transnational ownership chains
  - threshold $\theta = 10^{-3}$
  - chain value > 1
- 108,159,506 chunks
- 377,098 country chains

Results on our data

- 71,201,304 ownership links between firms
- 11,404,819 transnational ownership chains
  - threshold $\theta = 10^{-3}$
  - chain value > 1
- 108,159,506 chunks
- 377,098 country chains
  - 7,172 chains of size two
  - 52,655 chains of size three

Sink countries

- A country is a **sink** if less value leaves the country than enters the country. Measured using G^2: country chains of length 2.
- **Sink-OFC centrality** of a country c is defined as

$$C_{sink}(c) = \frac{\sum_{g \in G^2 \wedge g[1]=c} V^{\phi}(g) - \sum_{g \in G^2 \wedge g[0]=c} V^{\phi}(g)}{\sum_{g \in G^2} V^{\phi}(g)}$$

Sink countries

- A country is a **sink** if less value leaves the country than enters the country. Measured using G^2: country chains of length 2.
- **Sink-OFC centrality** of a country c is defined as

$$C_{sink}(c) = \frac{\sum_{g \in G^2 \wedge g[1]=c} V^{\phi}(g) - \sum_{g \in G^2 \wedge g[0]=c} V^{\phi}(g)}{\sum_{g \in G^2} V^{\phi}(g)}$$

- Problem: metric depends on size of country.
- Solution: normalize based on value GDP_c of the country, obtaining:

$$C_{sink-normalized}(c) = C_{sink}(c) \cdot \frac{\sum_{i \in I} GDP_i}{GDP_c}$$

Sink countries

ISO	Country name	C_s	ISO	Country name	C_s	ISO	Country name	C_s
VG	British Virgin Islands	5235	MH	Marshall Islands	100	BZ	Belize	38
TW	Taiwan	2278	MT	Malta	100	GI	Gibraltar	34
JE	Jersey	397	MU	Mauritius	75	AI	Anguilla	27
BM	Bermuda	374	LU	Luxembourg	71	LR	Liberia	17
KY	Cayman Islands	331	NR	Nauru	67	VC	St. Vincent & Gran.	14
WS	Samoa	277	CY	Cyprus	62	GY	Guyana	14
LI	Lichtenstein	225	SC	Seychelles	60	HK	Hong Kong	14
CW	Curacao	115	BS	Bahamas	40	MC	Monaco	11

Table : List of sink-OFCs, ordered by sink centrality value C_s.

Conduits

- The extent to which a country is **conduit** is measured using G^3: country chains of length 3.
- **Inward conduit-OFC centrality** measures the value flowing from a sink-OFC, into the conduit country c, out to any country:

$$C_{conduit_{in}}(c) = \frac{\sum_{g \in G^3 \wedge g[1]=sink \wedge g[2]=c} V^{\phi}(g)}{\sum_{g \in G^3} V^{\phi}(g)}$$

  Here, $g[i] = sink$ denotes that the i-th node in the chain is a sink.
- **Outward conduit-OFC centrality** measures the value flowing from any country, into conduit c, out to a sink-OFC:

$$C_{conduit_{out}}(c) = \frac{\sum_{g \in G^3 \wedge g[3]=sink \wedge g[2]=c} V^{\phi}(g)}{\sum_{g \in G^3} V^{\phi}(g)}$$

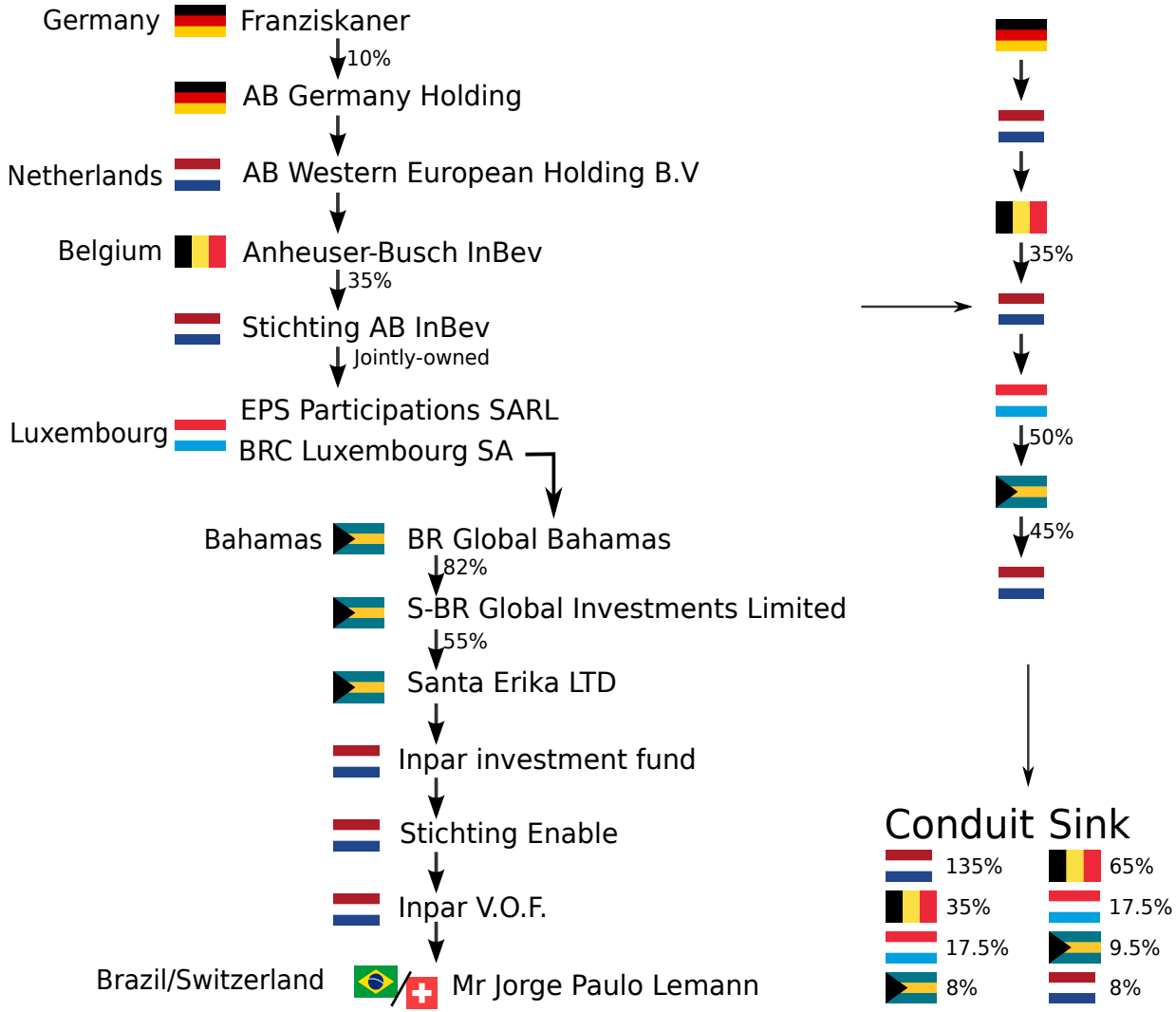
- Normalize by GDP

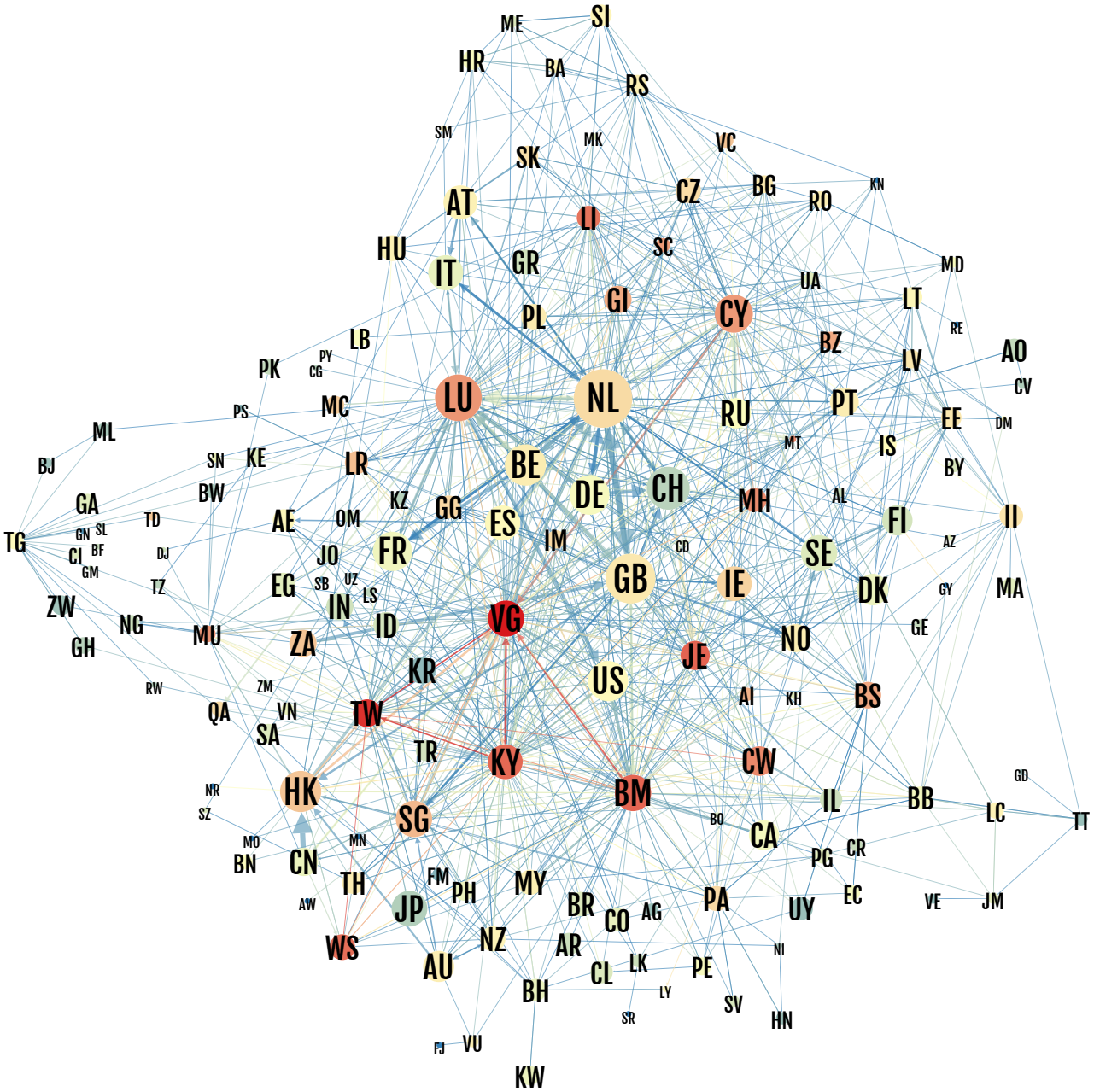
Conduits

ISO	Country name	Non-normalized $C_{c_{out}}$	Non-normalized $C_{c_{in}}$	$C_{c_{out}}$	$C_{c_{in}}$
NL	The Netherlands	$7.4 \cdot 10^{11}$	$3.8 \cdot 10^{11}$	18.6	22.5
GB	United Kingdom	$3.8 \cdot 10^{11}$	$1.3 \cdot 10^{11}$	3.1	2.4
CH	Switzerland	$2.2 \cdot 10^{11}$	$2.7 \cdot 10^{10}$	6.9	2.0
SG	Singapore	$7.2 \cdot 10^{10}$	$2.2 \cdot 10^{10}$	5.1	3.8
IE	Ireland	$6.4 \cdot 10^{10}$	$3.3 \cdot 10^{10}$	5.9	7.2

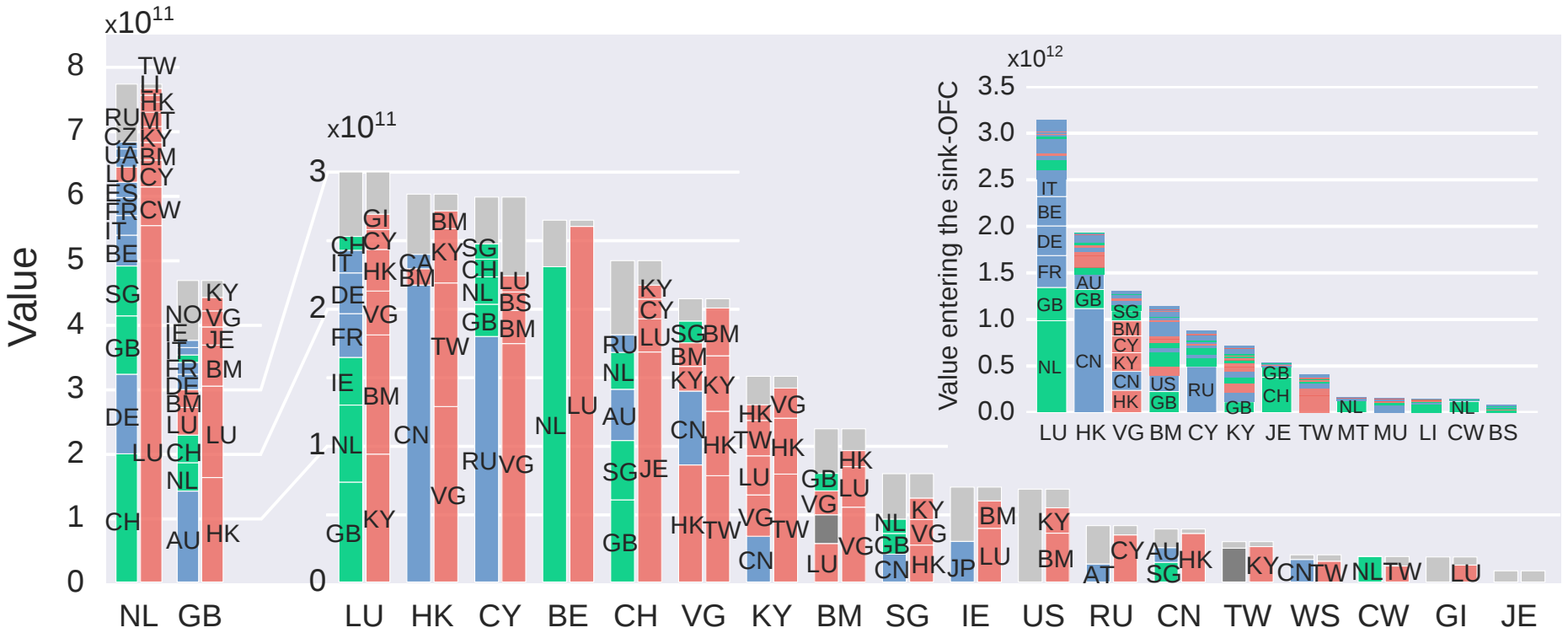
Table : List of conduit-OFCs, ordered by non-normalized $C_{c_{out}}$.

Country chains

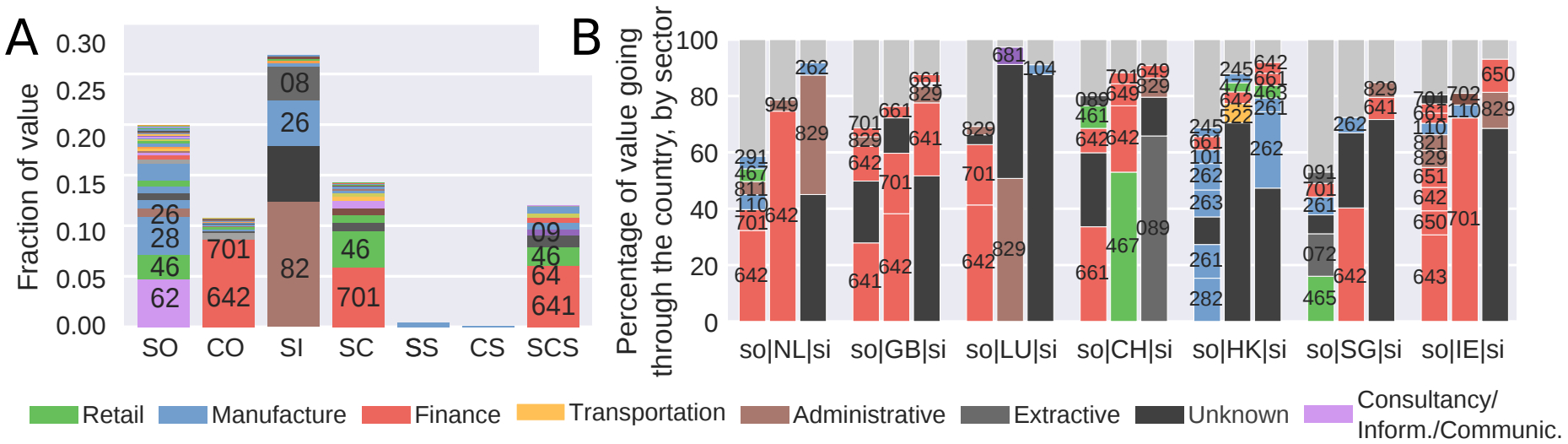



SI
ME
HR
BA
RS
SM
MK
VC
SK
AT
CZ
BG
RO
KN
LI
HU
SC
MD
IT
GR
UA
LT
GI
CY
PL
RE
LB
BZ
AO
PK
PY
CG
LV
CV
LU
NL
MC
PS
RU
PT
EE
DM
ML
MT
IS
BJ
SN
KE
LR
BE
BY
DE
CH
BW
MH
AL
GA
KZ
GG
FI
II
TD
AE
OM
ES
IM
AZ
TG
GN
SL
FR
CD
SE
CI
BF
JO
DJ
GB
IE
GM
EG
SB
UZ
DK
GY
MA
TZ
IN
LS
ZW
VG
NG
ID
MU
ZA
NO
GE
GH
JE
KR
US
RW
ZM
AI
KH
BS
QA
VN
TW
SA
TR
KY
CW
NR
GD
HK
BM
IL
BB
SZ
SG
BO
LC
TT
MO
MN
CA
BN
CN
PG
CR
TH
FM
MY
PH
BR
PA
EC
AW
JP
CO
AG
UY
VE
JM
NZ
WS
AR
NI
AU
CL
LK
PE
BH
LY
SV
SR
HN
PJ
VU
KW

Conduit country specialization



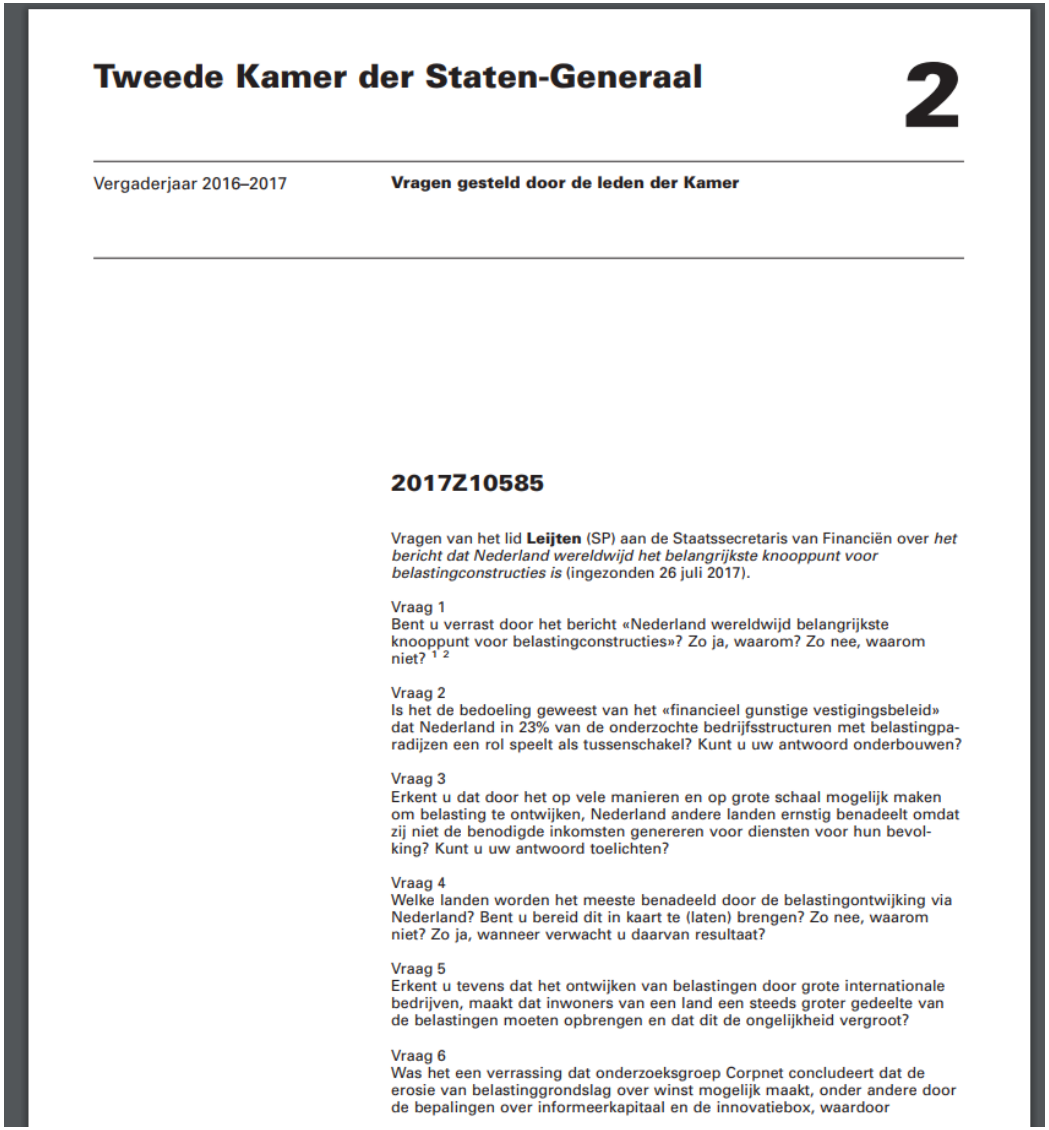
Conduit sector specialization



Findings

- The Netherlands is the conduit between European companies and Luxembourg
- Hong Kong (for China) and Luxembourg (for EU countries) serve as the main countries in the route to typical tax havens
- The United Kingdom is the conduit between European countries and former members of the British Empire, such as Hong Kong, Jersey, Guernsey and Bermuda
- Cyprus is the conduit for Russian firms
- Ireland is prominent in financial leasing and head offices
- Luxembourg specializes in support activities
- Hong Kong and Switzerland are dominant in financial intermediation and derivatives dealing

Questions in parliament . . .



Tweede Kamer der Staten-Generaal **2**

Vergaderjaar 2016–2017 **Vragen gesteld door de leden der Kamer**

2017Z10585

Vragen van het lid **Leijten** (SP) aan de Staatssecretaris van Financiën over *het bericht dat Nederland wereldwijd het belangrijkste knooppunt voor belastingconstructies is* (ingezonden 26 juli 2017).

Vraag 1
Bent u verrast door het bericht «Nederland wereldwijd belangrijkste knooppunt voor belastingconstructies»? Zo ja, waarom? Zo nee, waarom niet? [1] [2]

Vraag 2
Is het de bedoeling geweest van het «financieel gunstige vestigingsbeleid» dat Nederland in 23% van de onderzochte bedrijfsstructuren met belastingparadijzen een rol speelt als tussenschakel? Kunt u uw antwoord onderbouwen?

Vraag 3
Erkent u dat door het op vele manieren en op grote schaal mogelijk maken om belasting te ontwijken, Nederland andere landen ernstig benadeelt omdat zij niet de benodigde inkomsten genereren voor diensten voor hun bevolking? Kunt u uw antwoord toelichten?

Vraag 4
Welke landen worden het meeste benadeeld door de belastingontwijking via Nederland? Bent u bereid dit in kaart te (laten) brengen? Zo nee, waarom niet? Zo ja, wanneer verwacht u daarvan resultaat?

Vraag 5
Erkent u tevens dat het ontwijken van belastingen door grote internationale bedrijven, maakt dat inwoners van een land een steeds groter gedeelte van de belastingen moeten opbrengen en dat dit de ongelijkheid vergroot?

Vraag 6
Was het een verrassing dat onderzoeksgroep Corpnet concludeert dat de erosie van belastinggrondslag over winst mogelijk maakt, onder andere door de bepalingen over informeerkapitaal en de innovatiebox, waardoor

Answer by Wiebes

Tweede Kamer der Staten-Generaal **2**

Vergaderjaar 2016–2017 **Aanhangsel van de Handelingen**

Vragen gesteld door de leden der Kamer, met de daarop door de regering gegeven antwoorden

2540

Vragen van het lid **Leijten** (SP) aan de Staatssecretaris van Financiën over *het bericht dat Nederland wereldwijd het belangrijkste knooppunt voor belastingconstructies is* (ingezonden 26 juli 2017).

Antwoord van Staatssecretaris **Wiebes** (Financiën) (ontvangen 25 augustus 2017).

Vraag 1, 2
Bent u verrast door het bericht «Nederland wereldwijd belangrijkste knooppunt voor belastingconstructies»? Zo ja, waarom? Zo nee, waarom niet? [1] [2]
Is het de bedoeling geweest van het «financieel gunstige vestigingsbeleid»

become a supporter subscribe find a job

theguardian

news / opinion / sport / arts / life

world / europe / US / americas / asia / australia / middle east / africa / more

Tax havens

Netherlands and UK are biggest channels for corporate tax avoidance

The two countries are conduits for 37% of money heading to tax havens, most of which have strong links to Britain

Conclusions

- **Corporate networks** explicitly and implicitly contain rich information on corporations and jurisdictions
- Offshore financial centers (OFCs) come in two flavors: **conduits** and **sinks**
- OFCs can be detected using **corporate ownership chains**
- Conduits are highly **specialized** in serving particular countries and industrial sectors

Conclusions

- **Corporate networks** explicitly and implicitly contain rich information on corporations and jurisdictions
- Offshore financial centers (OFCs) come in two flavors: **conduits** and **sinks**
- OFCs can be detected using **corporate ownership chains**
- Conduits are highly **specialized** in serving particular countries and industrial sectors
- The Netherlands is the largest value conduit in the world