Mining Large-scale Corporate Networks

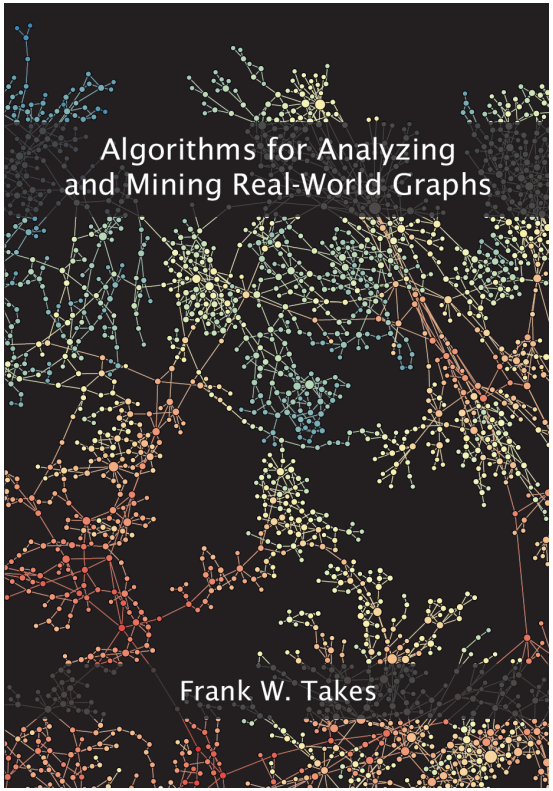
Frank Takes

LIACS, Leiden University
AISSR, University of Amsterdam

Leiden Complex Networks Network — LCN2
January 29, 2016

Introduction

Introduction



Introduction

- F.W. Takes, *Algorithms for Analyzing and Mining Real-World Graphs*, PhD thesis, Leiden University, 2014.

Introduction

- F.W. Takes, *Algorithms for Analyzing and Mining Real-World Graphs*, PhD thesis, Leiden University, 2014.
- E.M. Heemskerk and F.W. Takes, The Corporate Elite Community Structure of Global Capitalism, in *New Political Economy* 21(1): 90–118, 2016. `dx.doi.org/10.1080/13563467.2015.1041483`

Introduction

- F.W. Takes, *Algorithms for Analyzing and Mining Real-World Graphs*, PhD thesis, Leiden University, 2014.
- E.M. Heemskerk and F.W. Takes, The Corporate Elite Community Structure of Global Capitalism, in *New Political Economy* 21(1): 90–118, 2016. `dx.doi.org/10.1080/13563467.2015.1041483`
- Leiden Institute of Advanced Computer Science (LIACS)
- Amsterdam Institute for Social Science Research (AISSR)

Introduction

- F.W. Takes, *Algorithms for Analyzing and Mining Real-World Graphs*, PhD thesis, Leiden University, 2014.
- E.M. Heemskerk and F.W. Takes, The Corporate Elite Community Structure of Global Capitalism, in *New Political Economy* 21(1): 90–118, 2016. `dx.doi.org/10.1080/13563467.2015.1041483`
- Leiden Institute of Advanced Computer Science (LIACS)
- Amsterdam Institute for Social Science Research (AISSR)
- Interdisciplinary research

Corporate networks

- Networks
- Nodes are firms
- Edges/links indicate for example:
  - trading
  - borrowing/lending
  - ownership
  - (board) interlocks
- Aim is to understand:
  - Corporate control
  - Economy at a macro level
  - Corporate elites

Corporate network

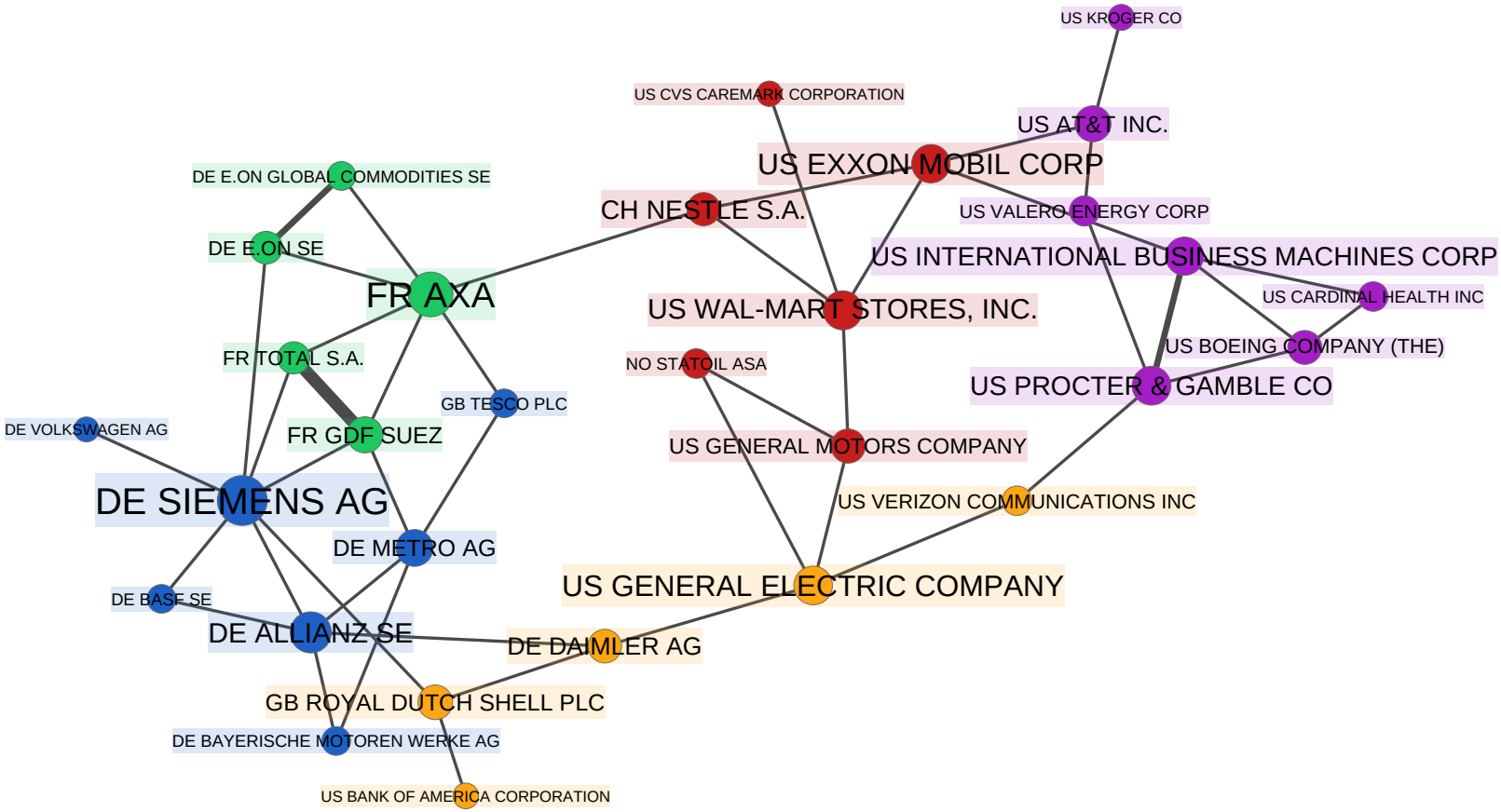


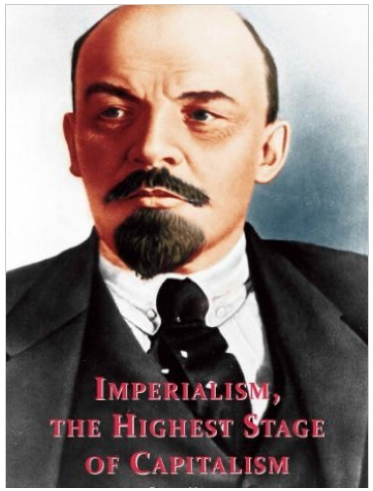
Figure: Board interlock network of 30 firms.

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."



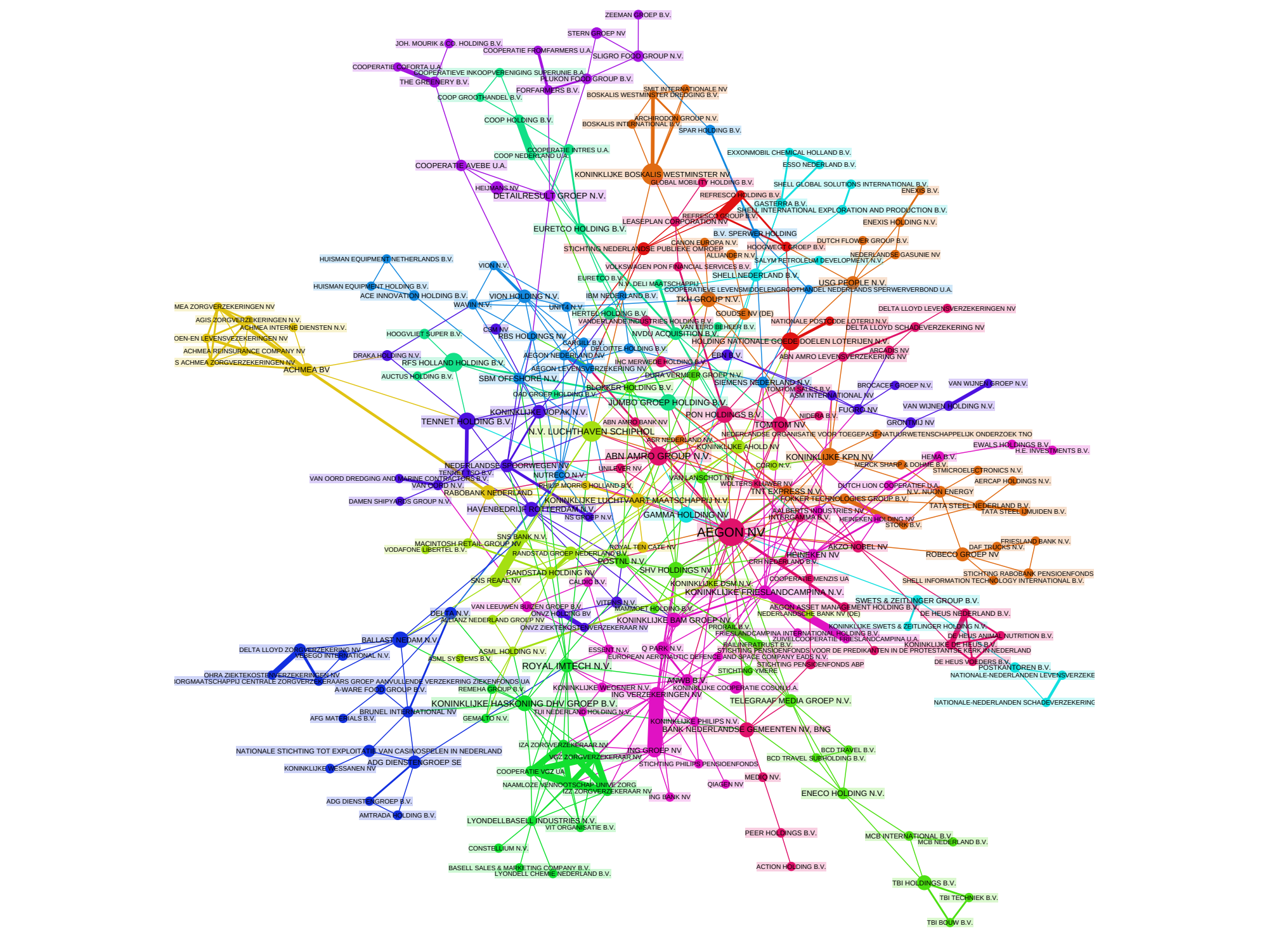
Board interlocks

- **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, in *Annual review of Sociology* 22: 271–298, 1996.


ZEEMAN GROEP B.V.
STERN GROEP NV
JOH. MOURIK & CO. HOLDING B.V.
COOPERATIE FROMFARMERS U.A.
SLIGRO FOOD GROUP N.V.
COOPERATIE COFORTA U.A.
COOPERATIEVE INKOOPVERENIGING SUPERUNIE B.A.
PLUKON FOOD GROUP B.V.
THE GREENERY B.V.
FORFARMERS B.V.
SMIT INTERNATIONALE NV
BOSKALIS WESTMINSTER DREDGING B.V.
COOP GROOTHANDEL B.V.
ARCHIRODON GROUP N.V.
BOSKALIS INTERNATIONAL B.V.
COOP HOLDING B.V.
SPAR HOLDING B.V.
COOPERATIE INTRES U.A.
EXXONMOBIL CHEMICAL HOLLAND B.V.
COOP NEDERLAND U.A.
ESSO NEDERLAND B.V.
COOPERATIE AVEBE U.A.
KONINKLIJKE BOSKALIS WESTMINSTER NV
SHELL GLOBAL SOLUTIONS INTERNATIONAL B.V.
GLOBAL MOBILITY HOLDING B.V.
HEIJMANS NV
ENEXIS B.V.
DETAILRESULT GROEP N.V.
REFRESCO HOLDING B.V.
GASTERRA B.V.
SHELL INTERNATIONAL EXPLORATION AND PRODUCTION B.V.
REFRESCO GROUP B.V.
ENEXIS HOLDING N.V.
LEASEPLAN CORPORATION NV
EURETCO HOLDING B.V.
B.V. SPERWER HOLDING
CANON EUROPA N.V.
DUTCH FLOWER GROUP B.V.
STICHTING NEDERLANDSE PUBLIEKE OMROEP
HOOGWEGT GROEP B.V.
ALLIANDER N.V.
NEDERLANDSE GASUNIE NV
HUISMAN EQUIPMENT NETHERLANDS B.V.
SALYM PETROLEUM DEVELOPMENT N.V.
VOLKSWAGEN PON FINANCIAL SERVICES B.V.
VION N.V.
SHELL NEDERLAND B.V.
EURETCO B.V.
N.V. DELI MAATSCHAPPIJ
USG PEOPLE N.V.
HUISMAN EQUIPMENT HOLDING B.V.
COOPERATIEVE LEVENSMIDDELENGROOTHANDEL NEDERLANDS SPERWERVERBOND U.A.
ACE INNOVATION HOLDING B.V.
VION HOLDING N.V.
IBM NEDERLAND B.V.
TKH GROUP N.V.
MEA ZORGVERZEKERINGEN NV
WAVIN N.V.
UNIT 4 N.V.
GOUDSE NV (DE)
DELTA LLOYD LEVENSVERZEKERINGEN NV
AGIS ZORGVERZEKERINGEN N.V.
HERTEL HOLDING B.V.
VANDERLANDE INDUSTRIES HOLDING B.V.
NATIONALE POSTCODE LOTERIJ N.V.
ACHMEA INTERNE DIENSTEN N.V.
VAN EERD BEHEER B.V.
DELTA LLOYD SCHADEVERZEKERING NV
HOOGVLIET SUPER B.V.
CSM NV
RBS HOLDINGS NV
NVDU ACQUISITION B.V.
OEN-EN LEVENSVERZEKERINGEN NV
HOLDING NATIONALE GOEDE DOELEN LOTERIJEN N.V.
CARGILL B.V.
DELOITTE HOLDING B.V.
ARCADIS NV
ACHMEA REINSURANCE COMPANY NV
DRAKA HOLDING N.V.
AEGON NEDERLAND NV
EBN B.V.
ABN AMRO LEVENSVERZEKERING NV
RFS HOLLAND HOLDING B.V.
IHC MERWEDE HOLDING B.V.
S ACHMEA ZORGVERZEKERINGEN NV
AEGON LEVENSVERZEKERING NV
ACHMEA BV
DURA VERMEER GROEP N.V.
AUCTUS HOLDING B.V.
SBM OFFSHORE N.V.
SIEMENS NEDERLAND N.V.
BROCACEF GROEP N.V.
VAN WIJNEN GROEP N.V.
BLOKKER HOLDING B.V.
TOMTOM SALES B.V.
ASM INTERNATIONAL NV
GAD GROEP HOLDING B.V.
JUMBO GROEP HOLDING B.V.
FUGRO NV
VAN WIJNEN HOLDING N.V.
KONINKLIJKE VOPAK N.V.
PON HOLDINGS B.V.
NIDERA B.V.
GRONTMIJ NV
TENNET HOLDING B.V.
ABN AMRO BANK NV
TOMTOM NV
N.V. LUCHTHAVEN SCHIPHOL
NEDERLANDSE ORGANISATIE VOOR TOEGEPAST-NATUURWETENSCHAPPELIJK ONDERZOEK TNO
ASR NEDERLAND NV
EWALS HOLDINGS B.V.
KONINKLIJKE AHOLD NV
H.E. INVESTMENTS B.V.
ABN AMRO GROUP N.V.
KONINKLIJKE KPN NV
HEMA B.V.
NEDERLANDSE SPOORWEGEN NV
UNILEVER NV
CORIO N.V.
MERCK SHARP & DOHME B.V.
STMICROELECTRONICS N.V.
TENNET TSO B.V.
NUTRECO N.V.
VAN OORD DREDGING AND MARINE CONTRACTORS B.V.
VAN LANSCHOT NV
AERCAP HOLDINGS N.V.
VAN OORD N.V.
PHILIP MORRIS HOLLAND B.V.
WOLTERS KLUWER NV
TNT EXPRESS N.V.
DUTCH LION COOPERATIEF U.A.
N.V. NUON ENERGY
RABOBANK NEDERLAND
FOKKER TECHNOLOGIES GROUP B.V.
DAMEN SHIPYARDS GROUP N.V.
KONINKLIJKE LUCHTVAART MAATSCHAPPIJ N.V.
TATA STEEL NEDERLAND B.V.
HAVENBEDRIJF ROTTERDAM N.V.
AALBERTS INDUSTRIES N.V.
TATA STEEL IJMUIDEN B.V.
GAMMA HOLDING NV
INTERGAMMA B.V.
HEINEKEN HOLDING NV
NS GROEP N.V.
STORK B.V.
AEGON NV
SNS BANK N.V.
FRIESLAND BANK N.V.
MACINTOSH RETAIL GROUP NV
ROYAL TEN CATE NV
AKZO NOBEL NV
DAF TRUCKS N.V.
VODAFONE LIBERTEL B.V.
RANDSTAD GROEP NEDERLAND B.V.
HEINEKEN NV
ROBECO GROEP NV
POSTNL N.V.
CRH NEDERLAND B.V.
STICHTING RABOBANK PENSIOENFONDS
RANDSTAD HOLDING NV
SHV HOLDINGS NV
SHELL INFORMATION TECHNOLOGY INTERNATIONAL B.V.
SNS REAAL NV
CALDIC B.V.
KONINKLIJKE DSM N.V.
COOPERATIE MENZIS UA
KONINKLIJKE FRIESLANDCAMPINA N.V.
SWETS & ZEITLINGER GROUP B.V.
VITENS N.V.
VAN LEEUWEN BUIZEN GROEP B.V.
MAMMOET HOLDING B.V.
AEGON ASSET MANAGEMENT HOLDING B.V.
DELTA N.V.
ONVZ HOLDING BV
NEDERLANDSCHE BANK NV (DE)
DE HEUS NEDERLAND B.V.
ALLIANZ NEDERLAND GROEP NV
KONINKLIJKE BAM GROEP NV
ONVZ ZIEKTEKOSTENVERZEKERAAR NV
PRORAIL B.V.
KONINKLIJKE SWETS & ZEITLINGER HOLDING N.V.
FRIESLANDCAMPINA INTERNATIONAL HOLDING B.V.
DE HEUS ANIMAL NUTRITION B.V.
ZUIVELCOOPERATIE FRIESLANDCAMPINA U.A.
BALLAST NEDAM N.V.
KONINKLIJKE DE HEUS B.V.
FAILLIENETRUST B.V.
DELTA LLOYD ZORGVERZEKERING NV
ASML HOLDING N.V.
ESSENT N.V.
Q PARK N.V.
STICHTING PENSIOENFONDS VOOR DE PREDIKANTEN IN DE PROTESTANTSE KERK IN NEDERLAND
VEGEO INTERNATIONAL N.V.
EUROPEAN AERONAUTIC DEFENCE AND SPACE COMPANY EADS N.V.
ASML SYSTEMS B.V.
DE HEUS VOEDERS B.V.
ROYAL IMTECH N.V.
STICHTING PENSIOENFONDS ABP
POSTKANTOREN B.V.
OHRA ZIEKTEKOSTENVERZEKERINGEN NV
STICHTING YMERE
NATIONALE-NEDERLANDEN LEVENSVERZEKE
ORGMAATSCHAPPIJ CENTRALE ZORGVERZEKERAARS GROEP AANVULLENDE VERZEKERING ZIEKENFONDS UA
REMEHA GROEP B.V.
KONINKLIJKE WEGENER N.V.
ANWB B.V.
A-WARE FOOD GROUP B.V.
KONINKLIJKE COOPERATIE COSUN U.A.
ING VERZEKERINGEN NV
TELEGRAAF MEDIA GROEP N.V.
NATIONALE-NEDERLANDEN SCHADEVERZEKERING
KONINKLIJKE HASKONING DHV GROEP B.V.
TUI NEDERLAND HOLDING N.V.
BRUNEL INTERNATIONAL NV
GEMALTO N.V.
AFG MATERIALS B.V.
KONINKLIJKE PHILIPS N.V.
BANK NEDERLANDSE GEMEENTEN NV, BNG
NATIONALE STICHTING TOT EXPLOITATIE VAN CASINOSPELEN IN NEDERLAND
IZA ZORGVERZEKERAAR NV
BCD TRAVEL B.V.
ING GROEP NV
BCD TRAVEL SUBHOLDING B.V.
ADG DIENSTENGROEP SE
VGZ ZORGVERZEKERAAR NV
STICHTING PHILIPS PENSIOENFONDS
KONINKLIJKE TESSANEN NV
COOPERATIE VGZ UA
MEDIQ NV
NAAMLOZE VENNOOTSCHAP UNIVE ZORG
QIAGEN NV
ENECO HOLDING N.V.
IZZ ZORGVERZEKERAAR NV
ADG DIENSTENGROEP B.V.
ING BANK NV
AMTRADA HOLDING B.V.
LYONDELLBASELL INDUSTRIES N.V.
VIT ORGANISATIE B.V.
PEER HOLDINGS B.V.
MCB INTERNATIONAL B.V.
MCB NEDERLAND B.V.
CONSTELLIUM N.V.
ACTION HOLDING B.V.
BASELL SALES & MARKETING COMPANY B.V.
LYONDELL CHEMIE NEDERLAND B.V.
TBI HOLDINGS B.V.
TBI TECHNIEK B.V.
TBI BOUW B.V.

Corporate networks

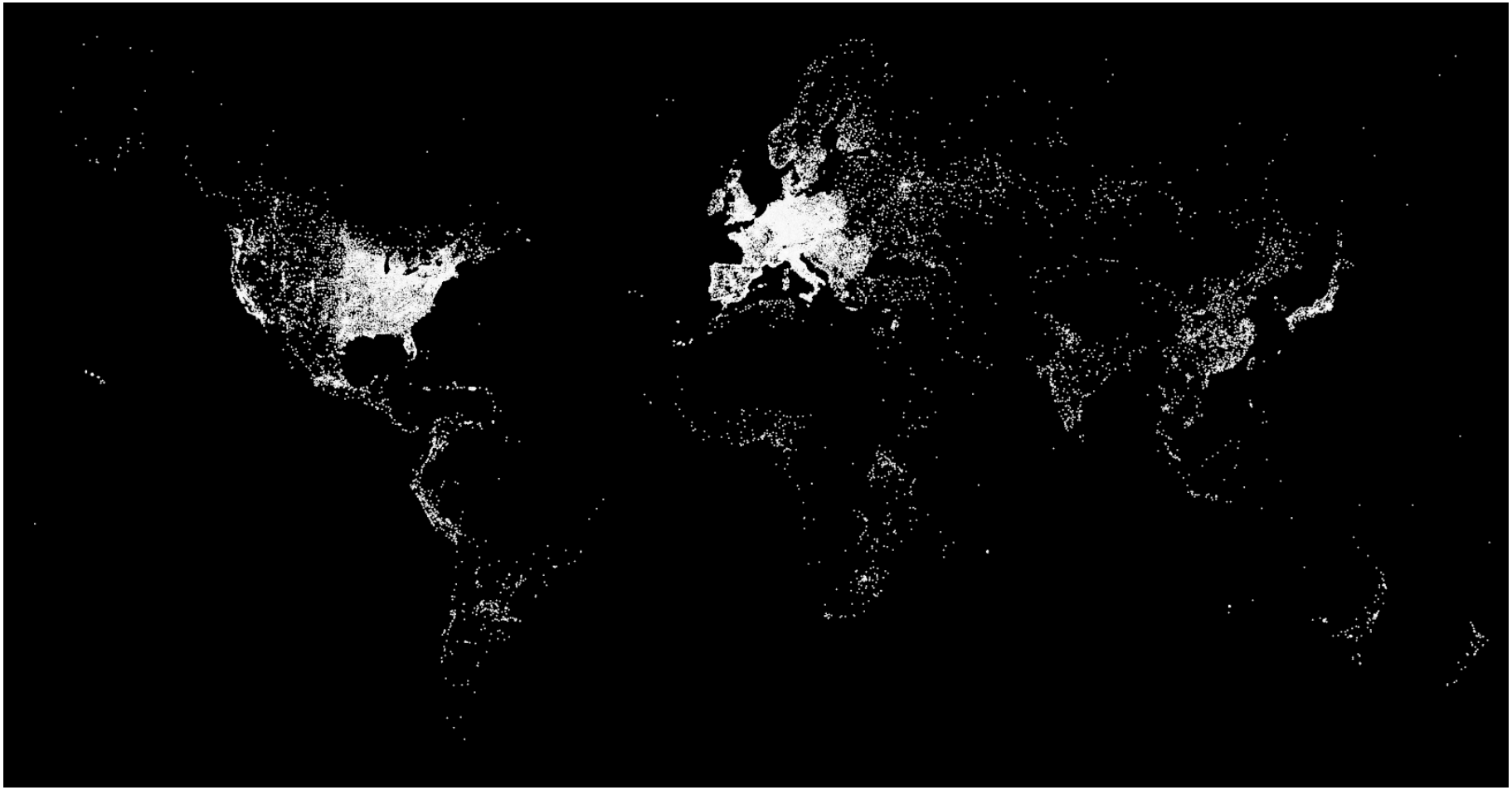


Figure: 400,000 largest firms globally, plotted based on latitude/longitude.

Corporate networks

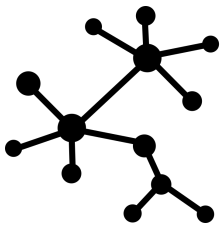
Figure: Global corporate network: over $1,000,000$ board interlocks.

CORPNET

- **CORPNET** — Corporate Network Governance: Power, Ownership and Control in Contemporary Global Capitalism
- *What are the* **features**, **origins** *and* **power** *political consequences of corporate governance networks in modern economic life?*
  - Nature: map and analyze the network
  - Origins: uncover generating mechanisms
  - Power: understand how it operates

CORPNET

- **CORPNET** — Corporate Network Governance: Power, Ownership and Control in Contemporary Global Capitalism
- *What are the* **features**, **origins** *and* **power** *political consequences of corporate governance networks in modern economic life?*
  - Nature: map and analyze the network
  - Origins: uncover generating mechanisms
  - Power: understand how it operates
- Work with Eelke Heemskerk and Javier Garcia-Bernardo



CORPNET
University of Amsterdam

Corporate network analysis

Universiteit Leiden

- Apply techniques from (social) **network analysis** to corporate data
- **Nodes** represent around firms across the globe
- **Edges** denote different relationships:
  - (Undirected) **board interlocks**: shared senior level directors
  - (Directed) ownership ties based on shareholder information
- Node attributes: country, sector, performance indicators, number of employees, ...
- Edge attributes: number of interlocks, type of shares, number of shares, ultimate share percentage, ...
- Data source: ORBIS database

Three topics

Universiteit Leiden

1. Network topology & centrality — large-scale
2. Community detection — large-scale
3. Data quality — very-large-scale

Universiteit Leiden

Network topology & centrality

Based on: F.W. Takes and E.M. Heemskerk, Centrality in the Global Network of Corporate Control, *forthcoming*, 2016.

Dataset

- ORBIS database (Bureau van Dijk)
- Firms listed as “large” or “very large”, and “active”
- Personal interlocks at senior management and board level
- Snapshot from December 2013
- Two-mode network of $971,891$ firms and $3,272,523$ top executives
- $579,924$ firms did not have any interlocks

Dataset

- ORBIS database (Bureau van Dijk)
- Firms listed as “large” or “very large”, and “active”
- Personal interlocks at senior management and board level
- Snapshot from December 2013
- Two-mode network of $971,891$ firms and $3,272,523$ top executives
- $579,924$ firms did not have any interlocks
- The remaining $391,967$ nodes form the nodes in the one-mode **global firm-by-firm network**

Topological properties

Global network	
Nodes	$391,967$
Edges	$1,711,968$
Density	$2.229 \cdot 10^{-5}$
Average degree	8.746
Clustering coefficient	0.755
Degree assortativity	0.260
Components	$55,616$

Topological properties

Global network	
Nodes	$391,967$
Edges	$1,711,968$
Density	$2.229 \cdot 10^{-5}$
Average degree	8.746
Clustering coefficient	0.755
Degree assortativity	0.260
Components	$55,616$

Giant component	
Nodes	$238,859$ nodes (60.9%)
Edges	$1,533,030$ (89.5%)
Density	$5.374 \cdot 10^{-5}$
Average degree	12.83
Clustering coefficient	0.751
Degree assortativity	0.202
Average distance	7.775
Radius	18
Diameter	34

Topological distributions

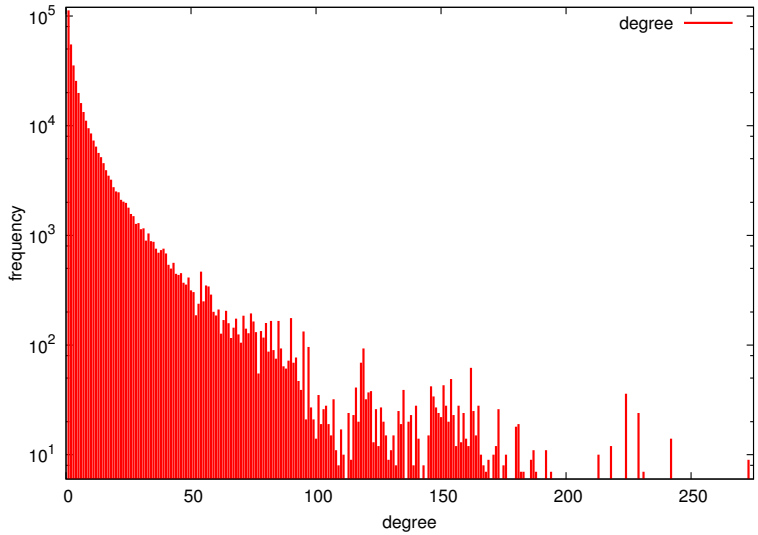


Figure: Degree distribution of the giant component

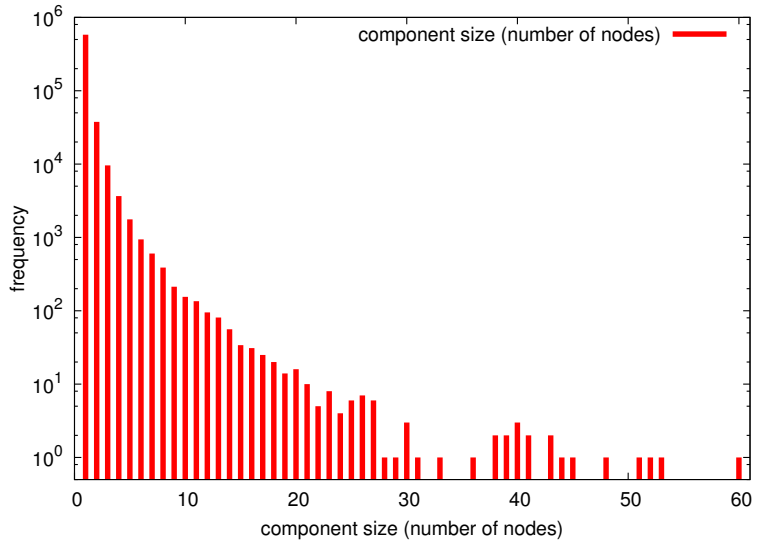


Figure: Component size distribution (excluding giant component)

Topological distributions

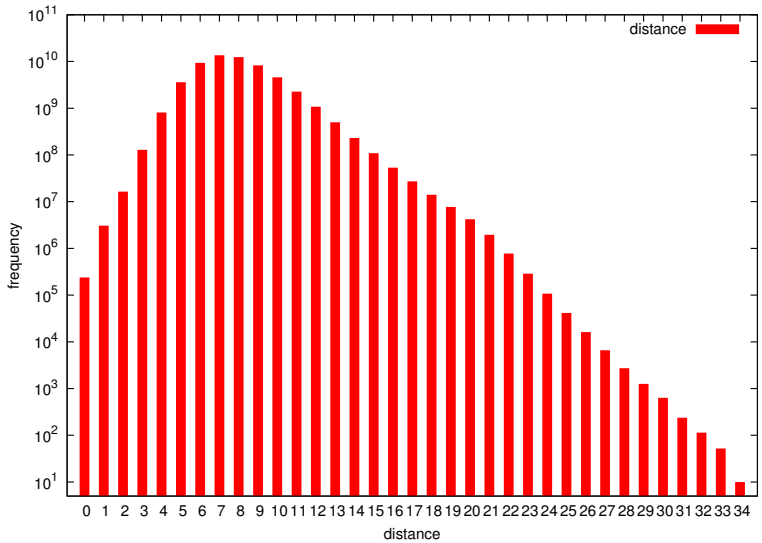


Figure: Distance distribution

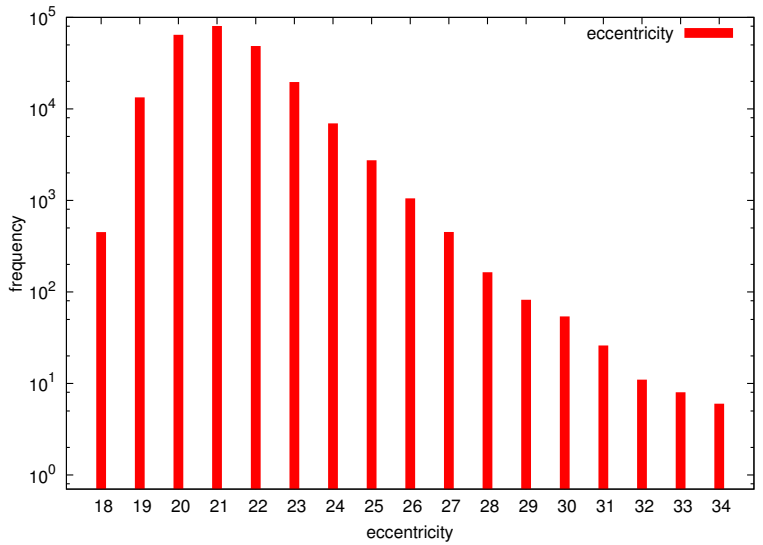
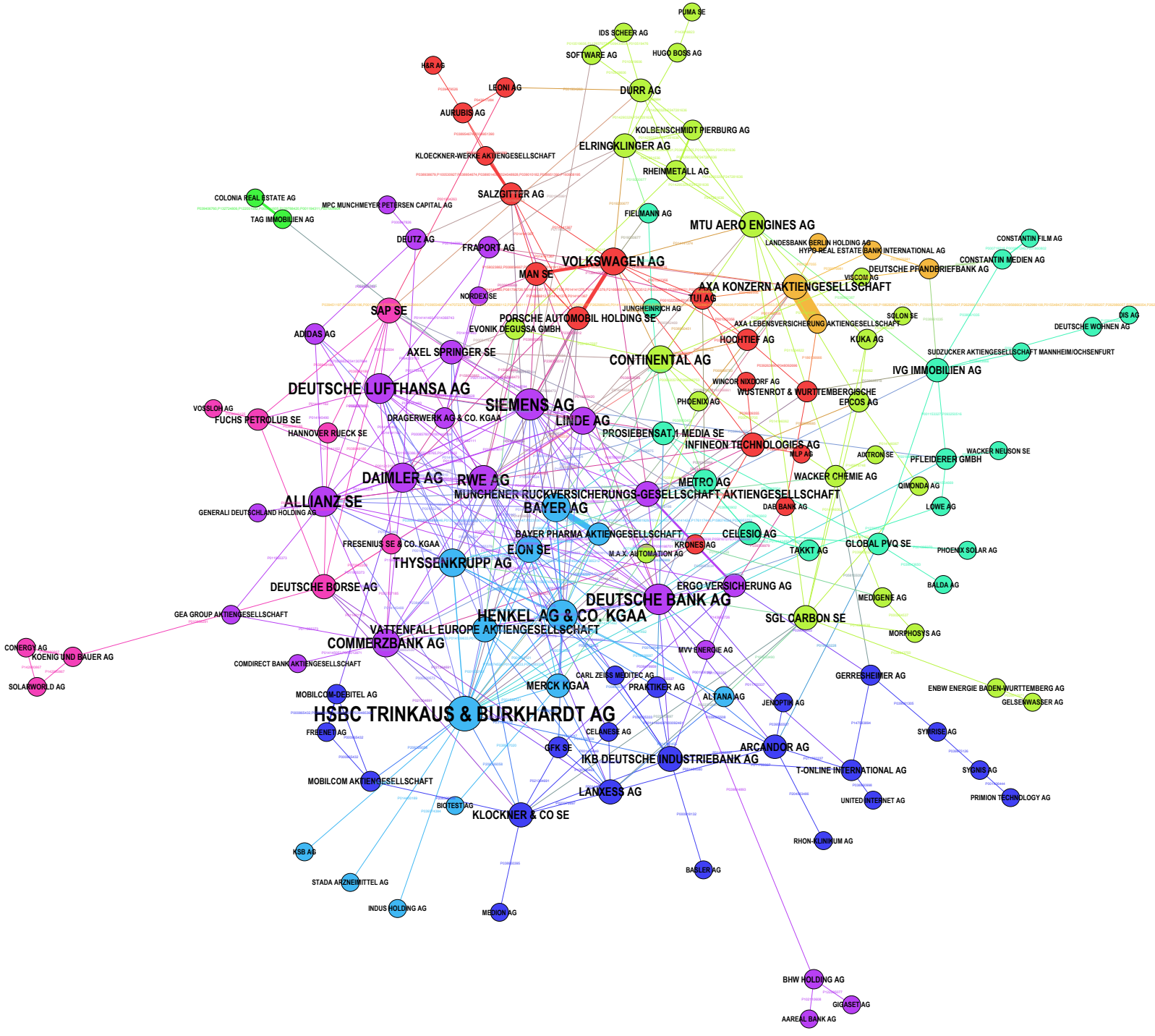


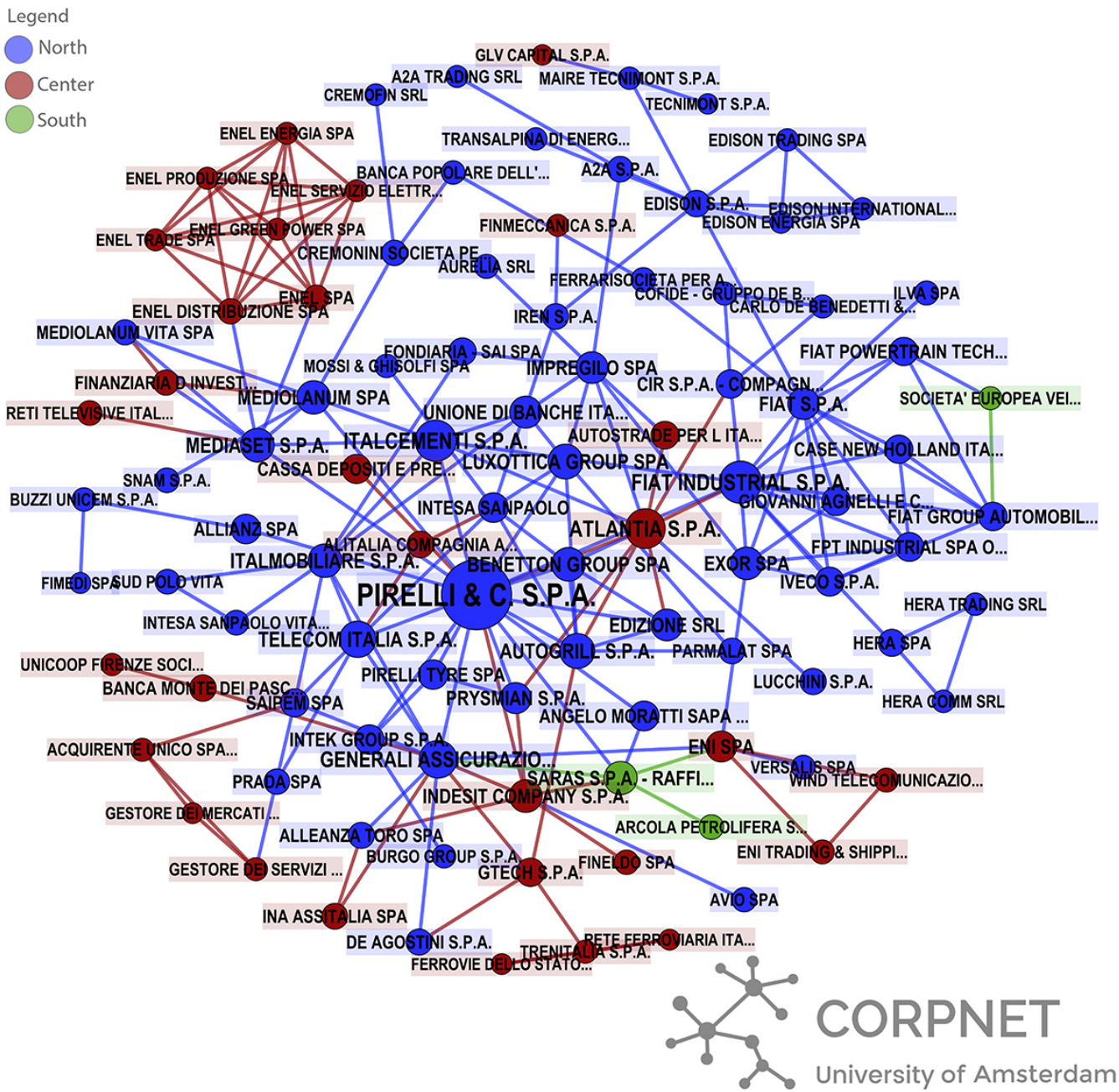
Figure: Eccentricity distribution

National networks

Universiteit Leiden

Country	Nodes	Density	Clust. coeff.	Degree assort.	Avg. dist.	Transnat. factor
GB	32,962	0.00067	0.356	0.845	6.63	0.26
US	24,802	0.00024	0.228	0.741	6.71	0.48
ES	11,102	0.00143	0.156	0.849	6.30	0.25
NO	8,963	0.00130	0.173	0.613	5.69	0.40
FR	8,896	0.00083	0.170	0.445	6.13	0.77
MY	7,878	0.00398	0.115	0.785	4.50	0.07
DE	7,224	0.00142	0.320	0.799	8.15	0.63
SE	6,656	0.00166	0.430	0.829	6.40	0.79
NL	6,083	0.00271	0.225	0.785	7.61	0.84
IN	5,911	0.00173	0.047	0.332	4.72	0.20
CA	5,439	0.00146	0.072	0.352	5.20	0.52
DK	4,517	0.00229	0.163	0.549	5.61	0.78
IT	4,483	0.00125	0.198	0.524	7.57	0.88
BE	3,264	0.00254	0.123	0.416	5.17	1.57
RU	2,939	0.00263	0.102	0.556	6.57	0.08
KR	2,802	0.00174	0.124	0.356	5.83	0.05
FI	2,626	0.00294	0.174	0.539	5.52	1.11
JP	2,605	0.00119	0.113	0.208	7.20	0.21
IE	2,497	0.01479	0.178	0.747	5.78	0.39
AT	2,142	0.00440	0.273	0.670	5.58	0.79
PT	2,120	0.00488	0.138	0.620	5.45	0.56
AU	1,897	0.00382	0.085	0.414	4.94	0.58
LU	1,484	0.00705	0.196	0.720	6.72	1.55
SG	1,472	0.00709	0.080	0.421	4.14	0.90
VN	1,393	0.00558	0.090	0.501	4.44	0.01
CH	999	0.00620	0.077	0.316	4.78	1.63
CN	891	0.00475	0.132	0.465	5.80	1.18
KY	642	0.00693	0.098	0.387	5.40	3.90


PUMA SE
IDS SCHEER AG
SOFTWARE AG
HUGO BOSS AG
H&R AG
LEONI AG
DURR AG
AURUBIS AG
KOLBENSCHMIDT PIERBURG AG
ELRINGKLINGER AG
KLOECKNER-WERKE AKTIENGESELLSCHAFT
RHEINMETALL AG
SALZGITTER AG
COLONIA REAL ESTATE AG
MPC MUNCHMEYER PETERSEN CAPITAL AG
TAG IMMOBILIEN AG
FIELMANN AG
MTU AERO ENGINES AG
DEUTZ AG
FRAPORT AG
LANDESBANK BERLIN HOLDING AG
HYPO REAL ESTATE BANK INTERNATIONAL AG
CONSTANTIN FILM AG
CONSTANTIN MEDIEN AG
VOLKSWAGEN AG
MAN SE
DEUTSCHE PFANDBRIEFBANK AG
VISCOM AG
AXA KONZERN AKTIENGESELLSCHAFT
TUI AG
NORDEX SE
SAP SE
JUNGHEINRICH AG
PORSCHE AUTOMOBIL HOLDING SE
SOLON SE
DIS AG
DEUTSCHE WOHNEN AG
EVONIK DEGUSSA GMBH
AXA LEBENSVERSICHERUNG AKTIENGESELLSCHAFT
ADIDAS AG
HOCHTIEF AG
KUKA AG
AXEL SPRINGER SE
CONTINENTAL AG
SUDZUCKER AKTIENGESELLSCHAFT MANNHEIM/OCHSENFURT
IVG IMMOBILIEN AG
WINCOR NIXDORF AG
DEUTSCHE LUFTHANSA AG
WUESTENROT & WURTTEMBERGISCHE
EPCOS AG
SIEMENS AG
VOSSLOH AG
PHOENIX AG
FUCHS PETROLUB SE
DRAGERWERK AG & CO. KGAA
LINDE AG
HANNOVER RUECK SE
PROSIEBENSAT.1 MEDIA SE
INFINEON TECHNOLOGIES AG
MLP AG
AIXTRON SE
WACKER NEUSON SE
PFLEIDERER GMBH
DAIMLER AG
RWE AG
METRO AG
WACKER CHEMIE AG
QIMONDA AG
MUNCHENER RUCKVERSICHERUNGS-GESELLSCHAFT AKTIENGESELLSCHAFT
ALLIANZ SE
BAYER AG
DAB BANK AG
LOWE AG
GENERALI DEUTSCHLAND HOLDING AG
BAYER PHARMA AKTIENGESELLSCHAFT
CELESIO AG
FRESENIUS SE & CO. KGAA
KRONES AG
GLOBAL PVQ SE
TAKKT AG
PHOENIX SOLAR AG
E.ON SE
M.A.X. AUTOMATION AG
THYSSENKRUPP AG
BALDA AG
DEUTSCHE BORSE AG
ERGO VERSICHERUNG AG
DEUTSCHE BANK AG
MEDIGENE AG
GEA GROUP AKTIENGESELLSCHAFT
HENKEL AG & CO. KGAA
SGL CARBON SE
VATTENFALL EUROPE AKTIENGESELLSCHAFT
MORPHOSYS AG
CONERGY AG
KOENIG UND BAUER AG
COMMERZBANK AG
MVV ENERGIE AG
COMDIRECT BANK AKTIENGESELLSCHAFT
SOLARWORLD AG
CARL ZEISS MEDITEC AG
GERRESHEIMER AG
MERCK KGAA
PRAKTIKER AG
ENBW ENERGIE BADEN-WURTTEMBERG AG
MOBILCOM-DEBITEL AG
ALTANA AG
GELSENWASSER AG
JENOPTIK AG
HSBC TRINKAUS & BURKHARDT AG
FREENET AG
CELANESE AG
SYMRISE AG
GFK SE
ARCANDOR AG
IKB DEUTSCHE INDUSTRIEBANK AG
T-ONLINE INTERNATIONAL AG
SYGNIS AG
MOBILCOM AKTIENGESELLSCHAFT
LANXESS AG
UNITED INTERNET AG
PRIMION TECHNOLOGY AG
BIOTEST AG
KLOCKNER & CO SE
RHON-KLINIKUM AG
KSB AG
BASLER AG
STADA ARZNEIMITTEL AG
INDUS HOLDING AG
MEDION AG
BHW HOLDING AG
GIGASET AG
AAREAL BANK AG



Work by master student Roberto Lucchese

Findings

- Small world phenomenon
- Average node-to-node distance
  - Global network: 7.775
  - National networks: 5.692 (average) or 6.188 (weighted average)
- National footprints still visible?

Findings

- Small world phenomenon
- Average node-to-node distance
  - Global network: 7.775
  - National networks: 5.692 (average) or 6.188 (weighted average)
- National footprints still visible?
  - Competing elites
  - Globalization

Findings

- Small world phenomenon
- Average node-to-node distance
  - Global network: 7.775
  - National networks: 5.692 (average) or 6.188 (weighted average)
- National footprints still visible?
  - Competing elites
  - Globalization
- Let's investigate more complex embeddedness measures!

Centrality

- **Node centrality**: the importance of a node with respect to the other nodes based on the structure of the network
- **Centrality measure**: computes the centrality value of all nodes in the graph
  - **Degree centrality**: number of connections
  - **Closeness centrality**: average distance to all other nodes
  - **Betweenness centrality**: relative number of times a node is on a shortest path
- But what is the ground truth to verify these measures?

Centrality

- **Node centrality**: the importance of a node with respect to the other nodes based on the structure of the network
- **Centrality measure**: computes the centrality value of all nodes in the graph
  - **Degree centrality**: number of connections
  - **Closeness centrality**: average distance to all other nodes
  - **Betweenness centrality**: relative number of times a node is on a shortest path
- But what is the ground truth to verify these measures?
  - Hard to say!

Centrality

- **Node centrality**: the importance of a node with respect to the other nodes based on the structure of the network
- **Centrality measure**: computes the centrality value of all nodes in the graph
  - **Degree centrality**: number of connections
  - **Closeness centrality**: average distance to all other nodes
  - **Betweenness centrality**: relative number of times a node is on a shortest path
- But what is the ground truth to verify these measures?
  - Hard to say!
  - Correlate with **firm prominence** (revenue)?

Centrality measures compared

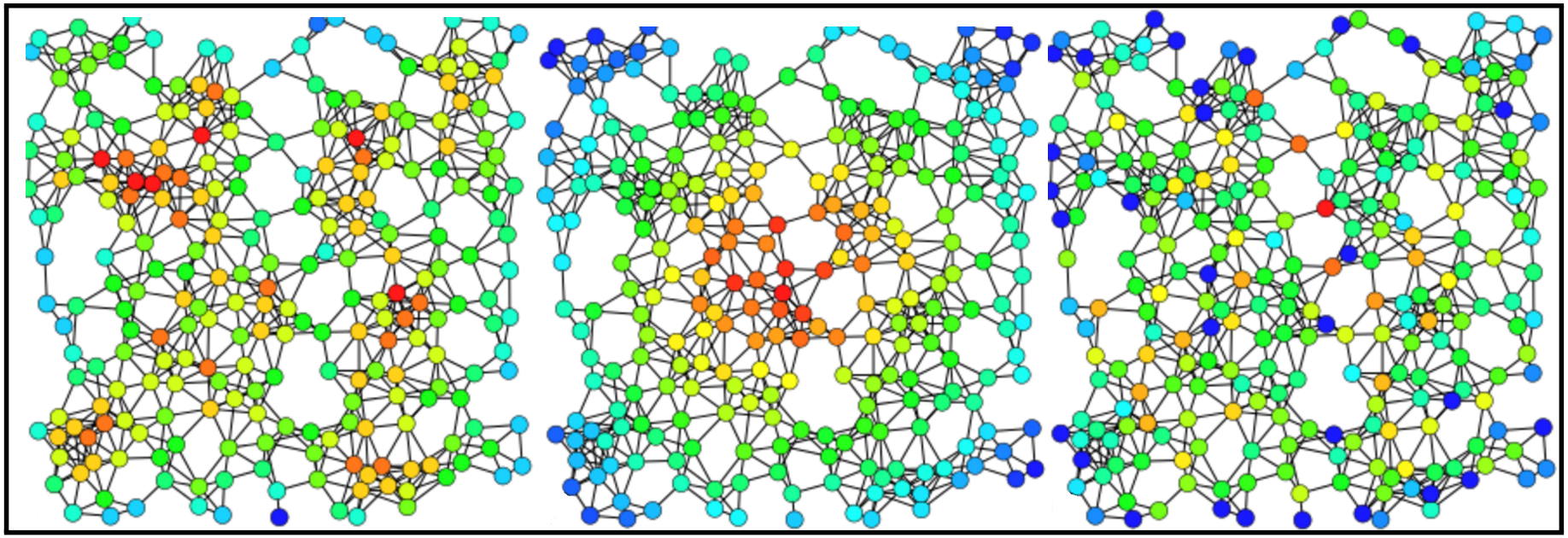


Figure: Degree, closeness and betweenness centrality

Source: "Centrality" by Claudio Rocchini, Wikipedia `File:Centrality.svg`

Global vs. National centrality

Global

1. US AT&T INC.
2. US 7-ELEVEN INC.
3. GB ROYAL DUTCH SHELL
4. GB ERNST & YOUNG EUROPE
5. KR SAMSUNG ELECTRONICS
6. GB PRICEWATERHOUSECOOPERS
7. CH RAIFFEISEN SCHWEIZ
8. GB KPMG EUROPE

Global vs. National centrality

Global

1. US AT&T INC.
2. US 7-ELEVEN INC.
3. GB ROYAL DUTCH SHELL
4. GB ERNST & YOUNG EUROPE
5. KR SAMSUNG ELECTRONICS
6. GB PRICEWATERHOUSECOOPERS
7. CH RAIFFEISEN SCHWEIZ
8. GB KPMG EUROPE

Great Britain

1. GB ERNST & YOUNG EUROPE
2. GB PRICEWATERHOUSECOOPERS
3. GB KPMG EUROPE
4. GB ROYAL DUTCH SHELL
5. GB DELOITTE
6. GB JP MORGAN
7. GB EASYJET
8. GB DLA PIPER INTERNATIONAL

Global centrality

Table: Correlation between centrality measures and with firm prominence (revenue), $n = 238,859$.

	Betweenness	**Closeness**	**Degree**	**Eigenvector**
Betweenness	1.000	0.430	0.521	0.356
Closeness	0.430	1.000	0.495	0.902
Degree	0.521	0.495	1.000	0.498
Eigenvector	0.356	0.902	0.498	1.000
Firm prominence	0.192	0.109	-0.046	0.064

National centrality

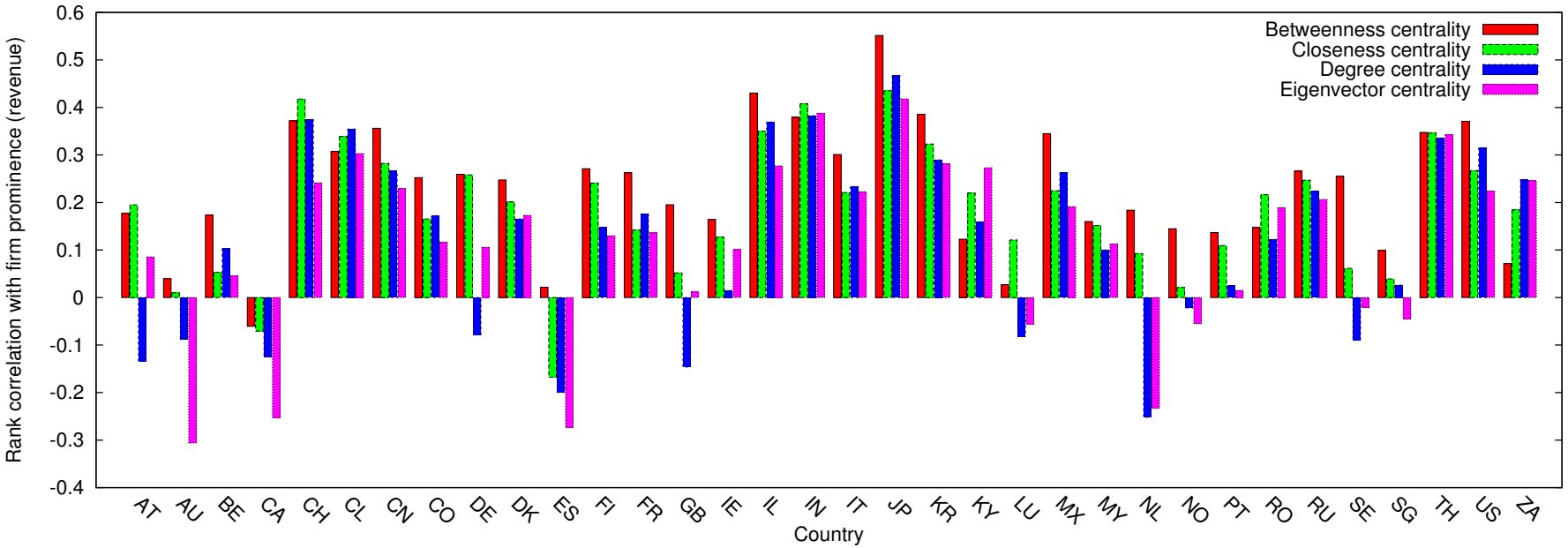


Figure: Correlation between firm prominence (revenue) and national centrality

National vs. global centrality

- **Centrality persistence**: correlation between global centrality (in the full network) and national centrality (in a partition)

Centrality persistence

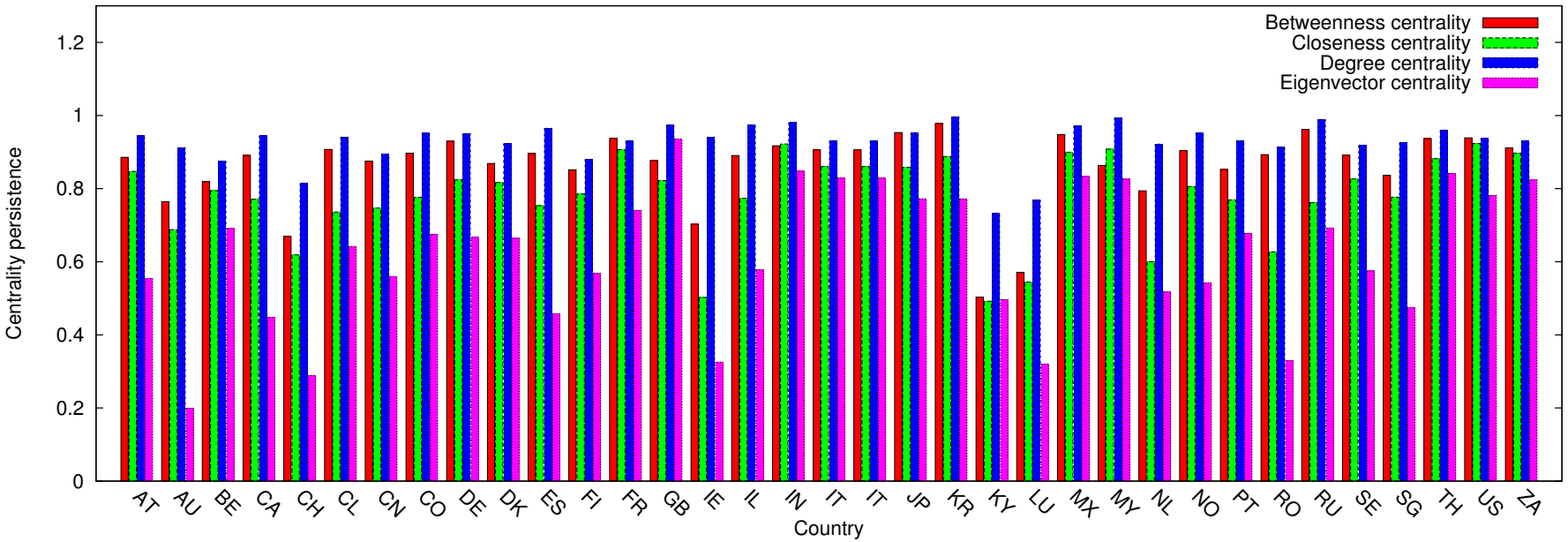


Figure: Centrality persistence for the 35 largest countries.

Centrality persistence

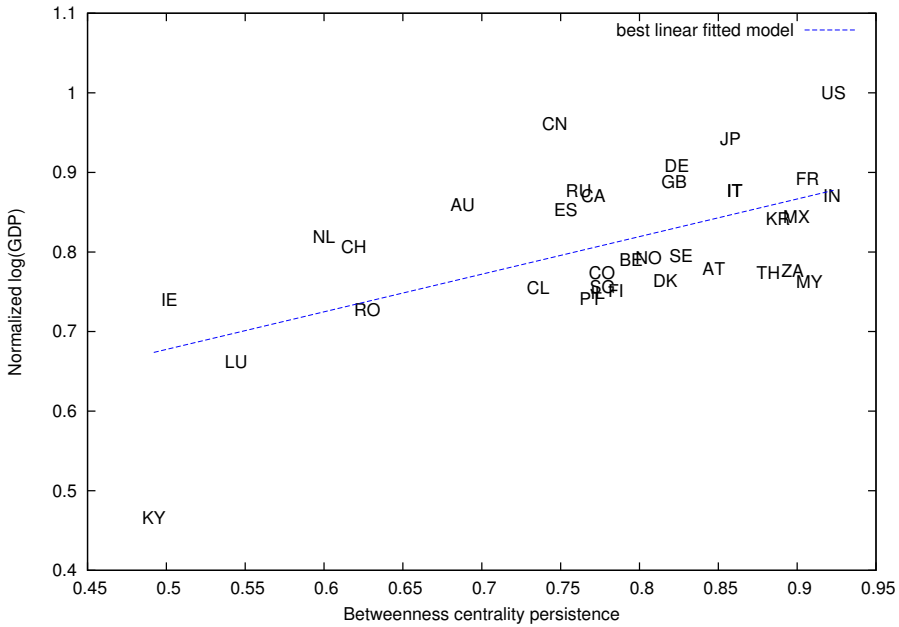


Figure: Betweenness centrality persistence vs. normalized log(GDP).

National vs. global centrality

- **Partition ranking dominance**: the ranking of a partition within the full network.
  Assume $S \subseteq V$ in a graph $G = (V, E)$.
  Assume that a node $v \in S$ according to some centrality ranking has rank $r(v) \in [0, |V|]$ in the full ranking of all nodes in V.
  Partition ranking dominance $pcr(S, V)$ is then defined as:

$$pcr(S, V) = 0.5 - \frac{\sum_{v \in S} r(v)}{|S| \cdot |V|}$$

  - value > 0 means the partition is less central than expected
  - value < 0 means it is more central than expected

Partition ranking dominance

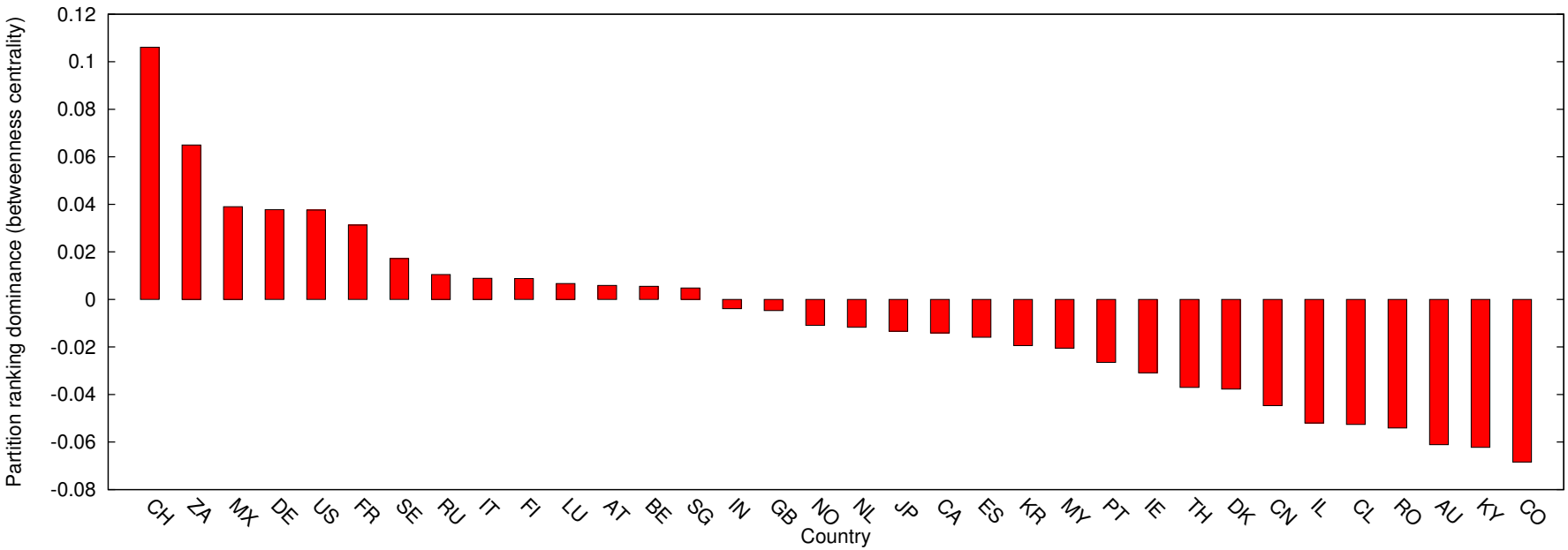


Figure: Partition ranking dominance (based on betweenness centrality).

Universiteit Leiden

Community detection

Based on: E.M. Heemskerk and F.W. Takes, The Corporate Elite Community Structure of Global Capitalism, in *New Political Economy* 21(1): 90–118, 2016. dx.doi.org/10.1080/13563467.2015.1041483

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
- Country network: aggregate firms from the same country

Community detection

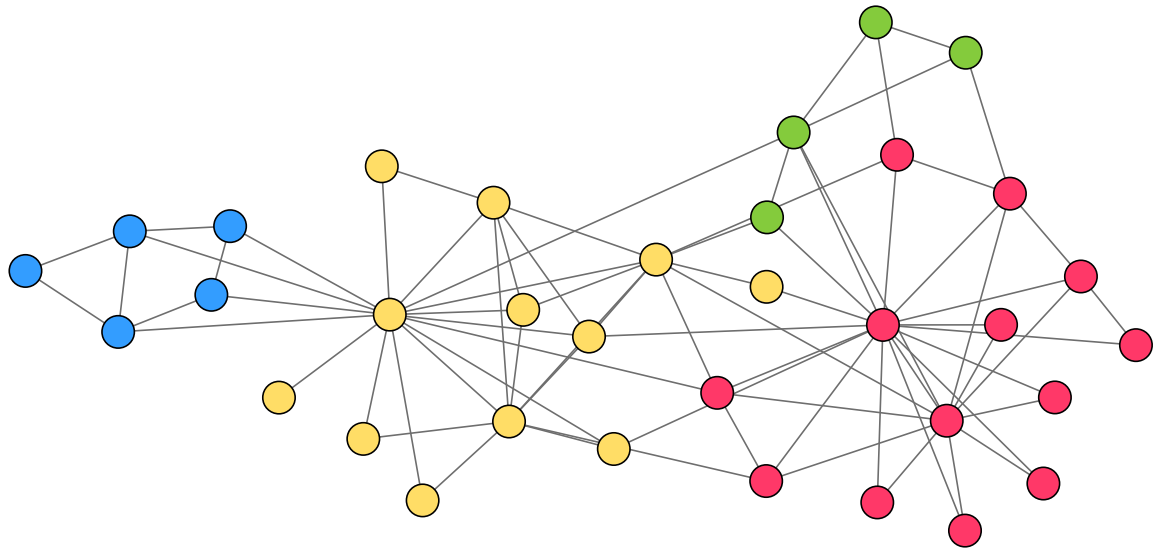


Figure: Communities: node subsets connected more strongly with each other

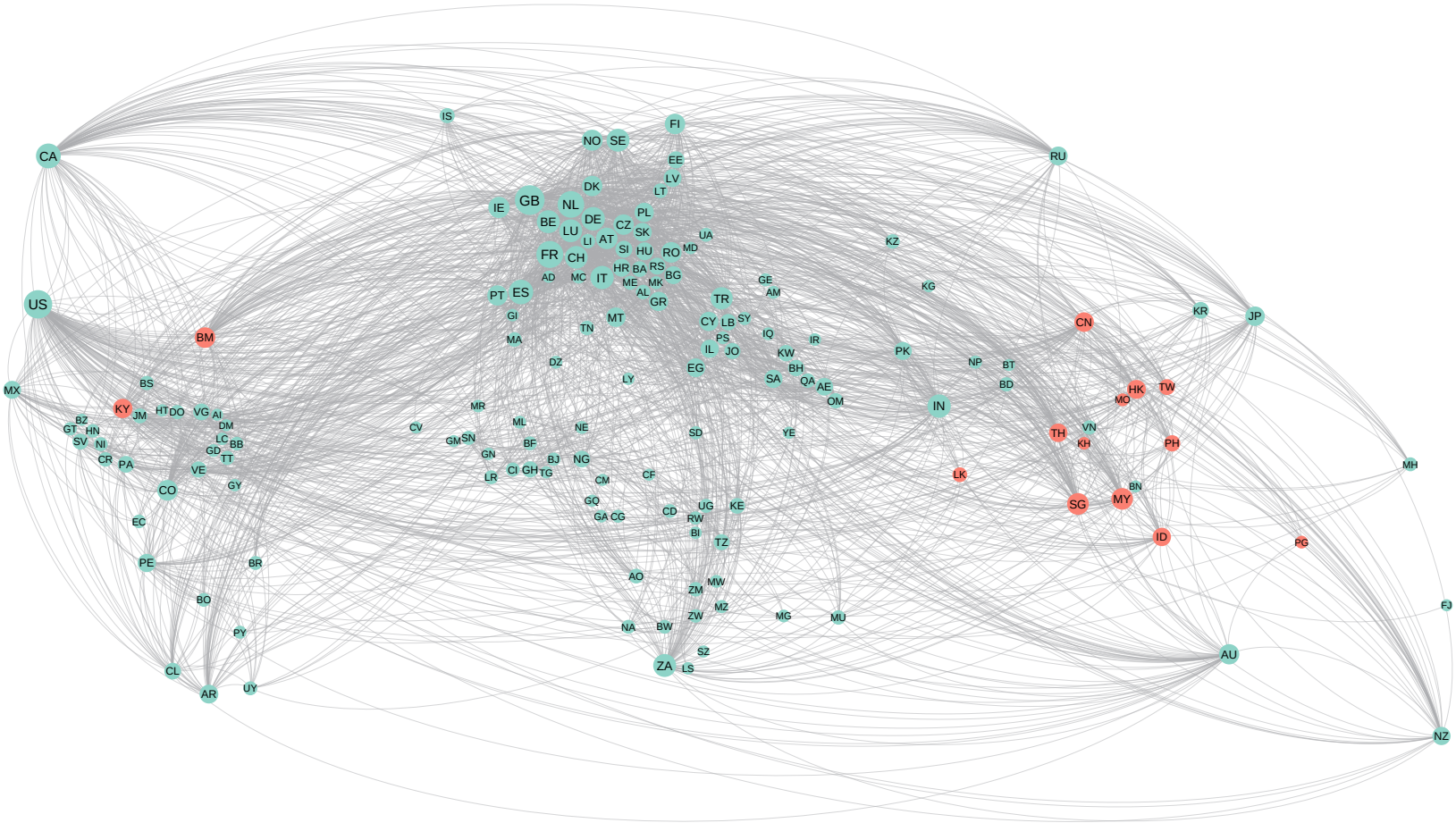
Modularity

- **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)

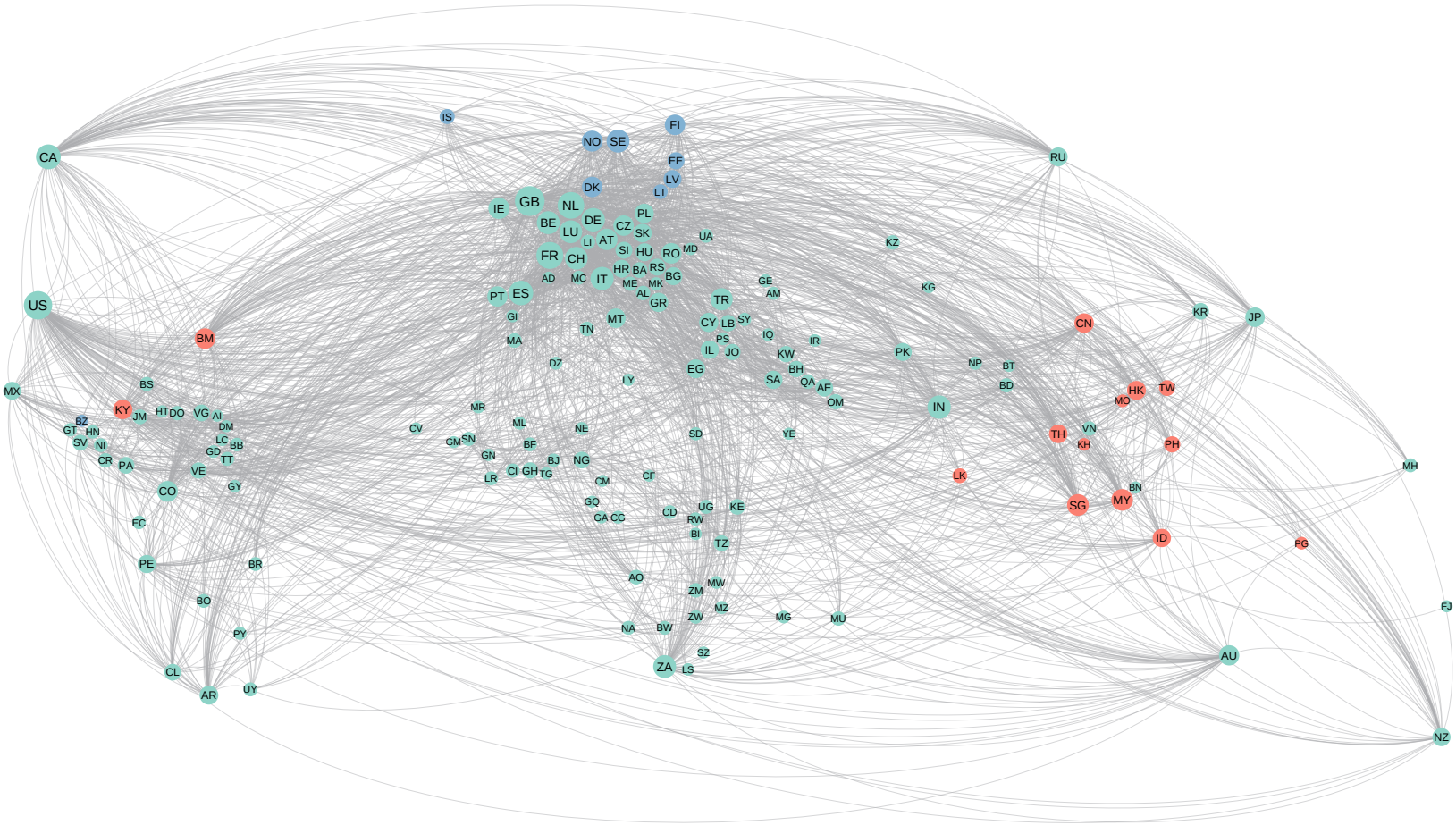
V.D. Blondel, J-L. Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communities in large networks in *Journal of Statistical Mechanics: Theory and Experiment* 10: P10008, 2008.

Community detection

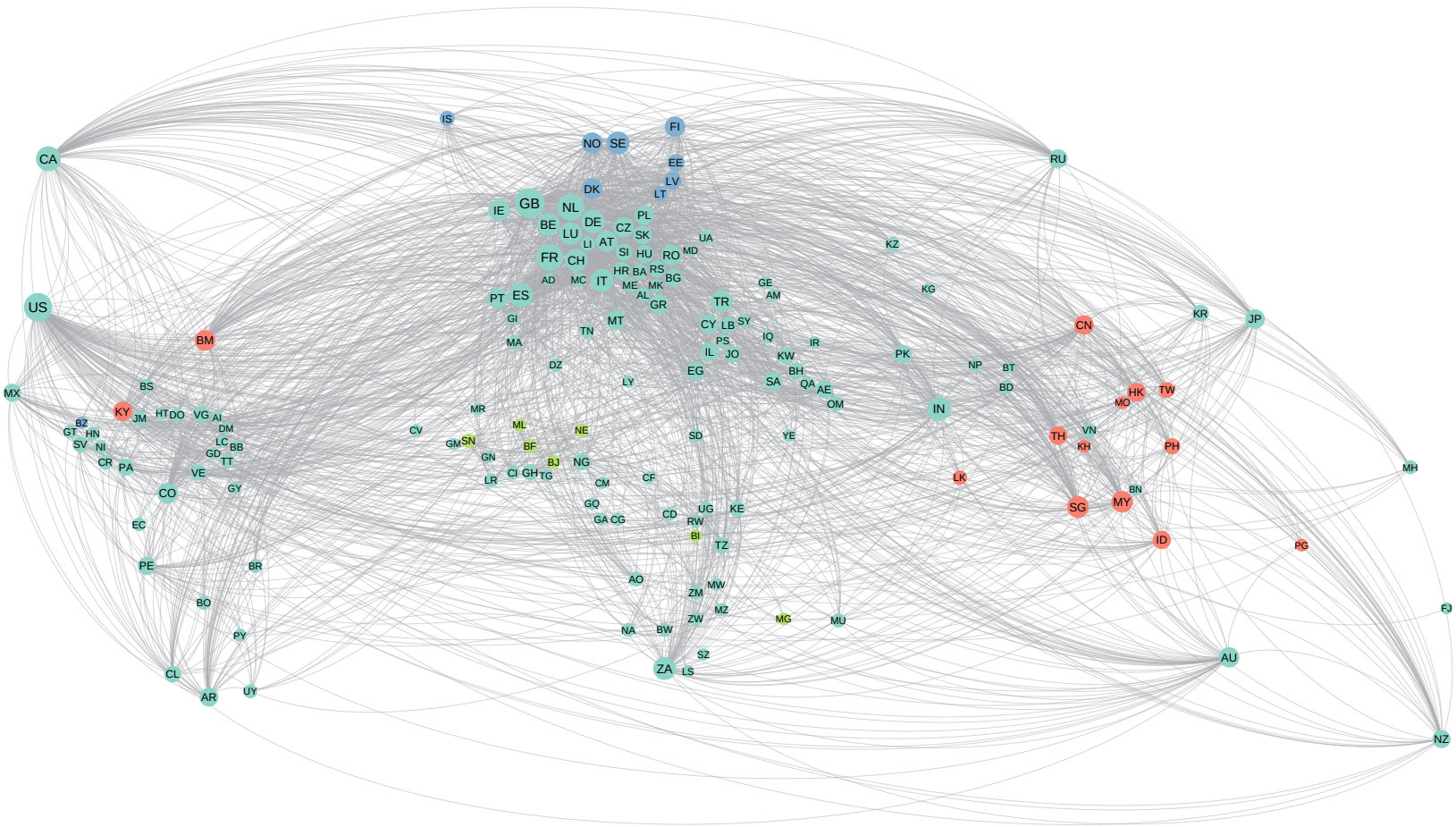
Community detection



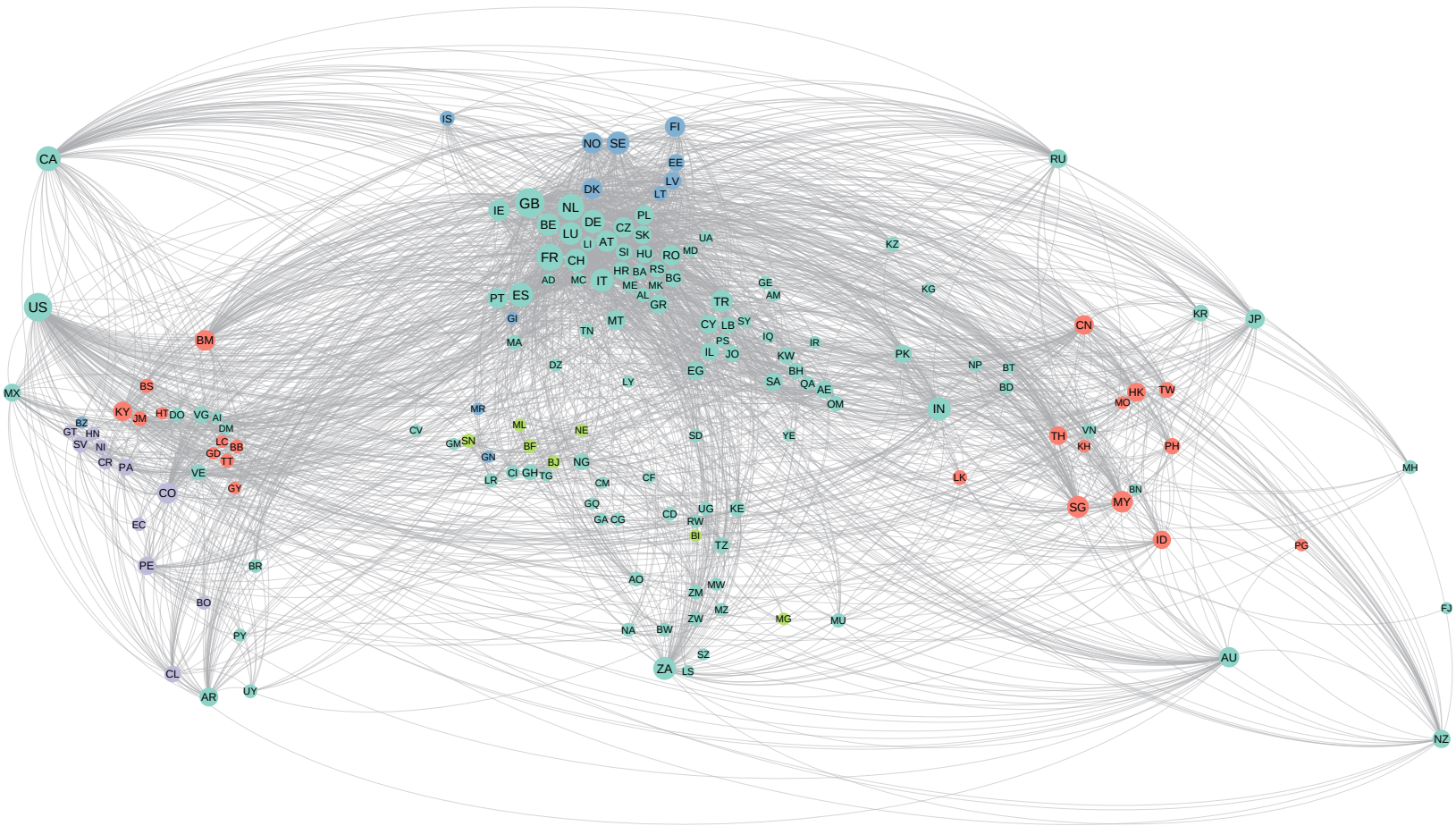
Community detection



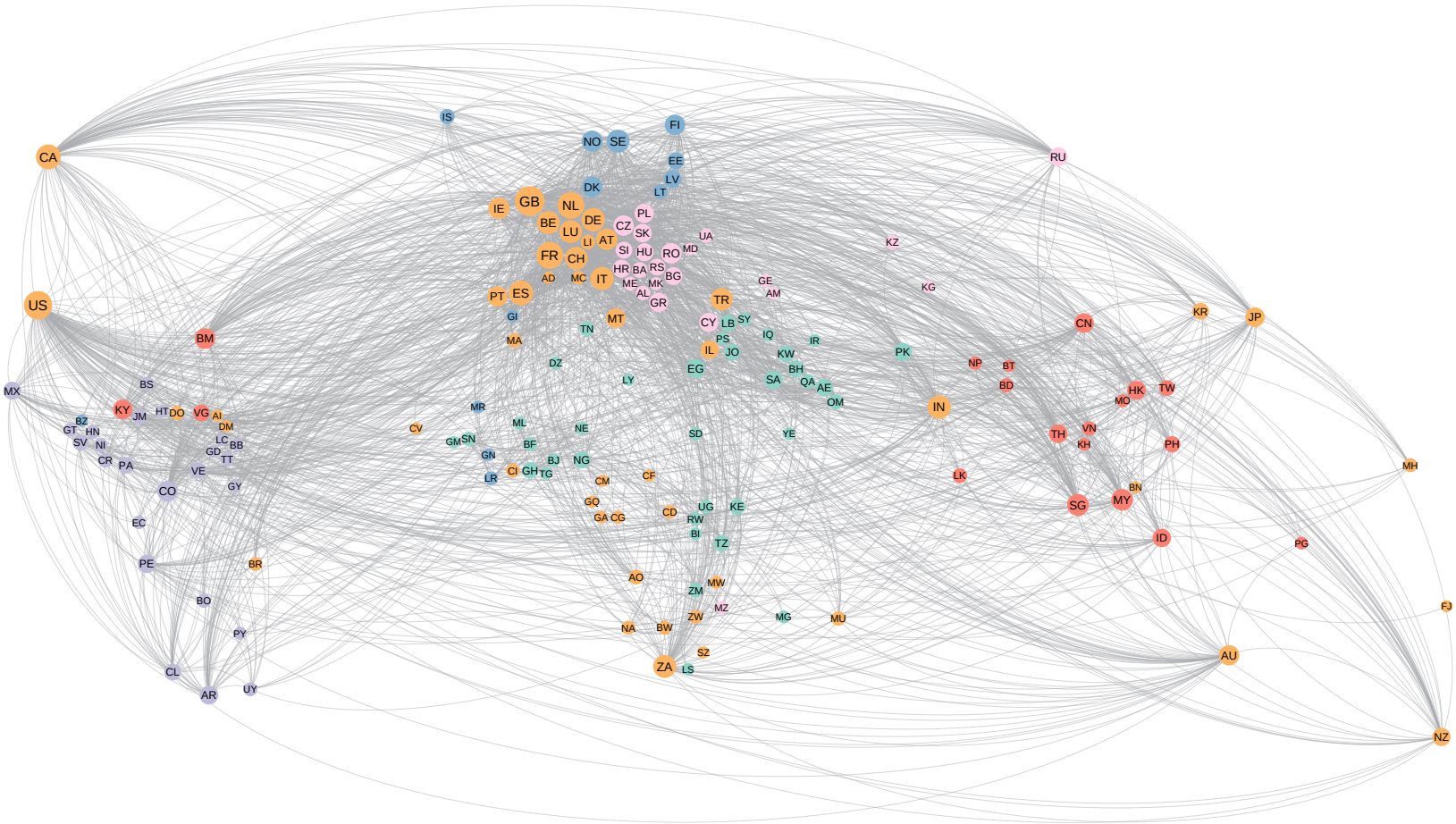
Community detection



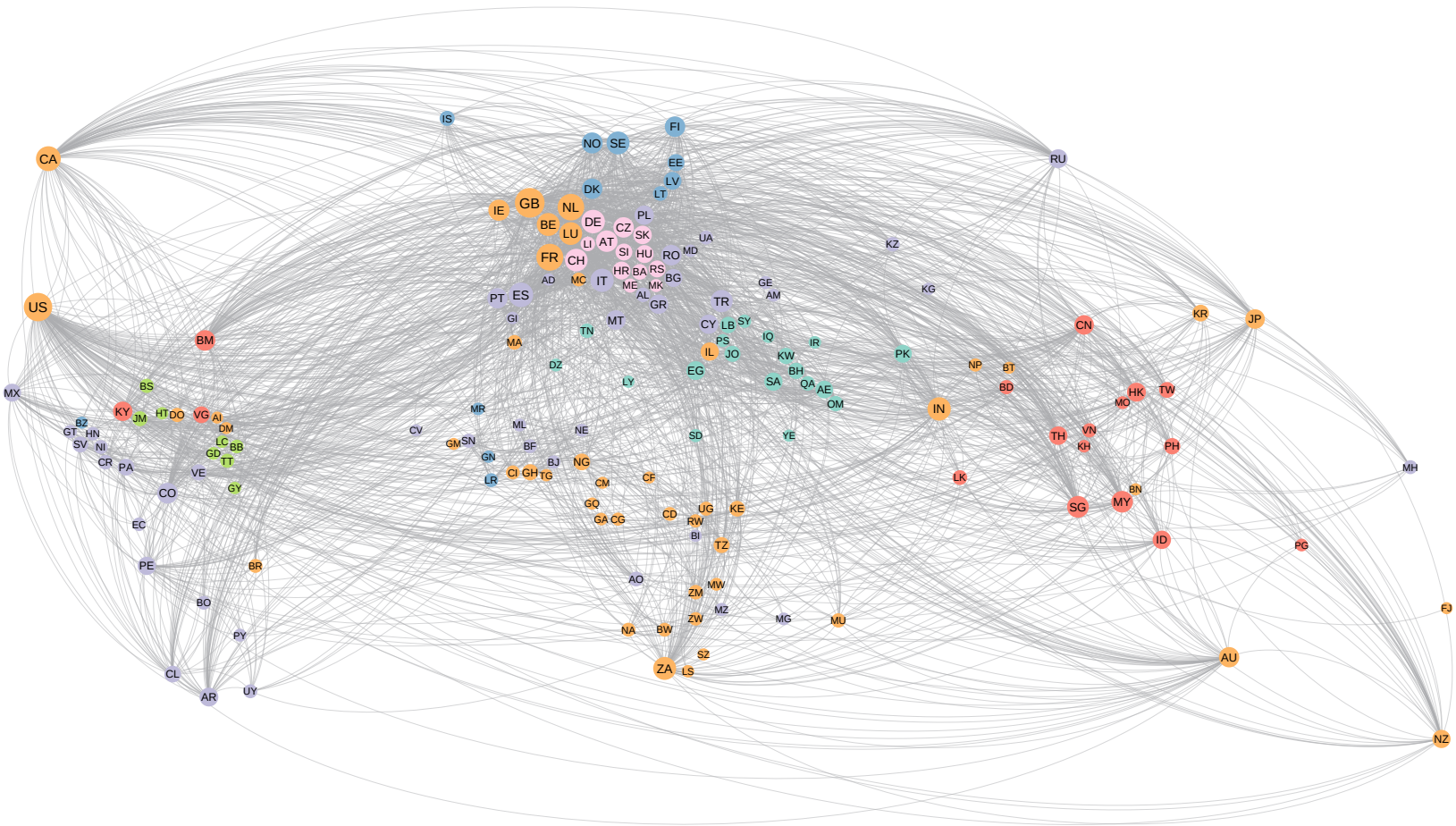
Community detection



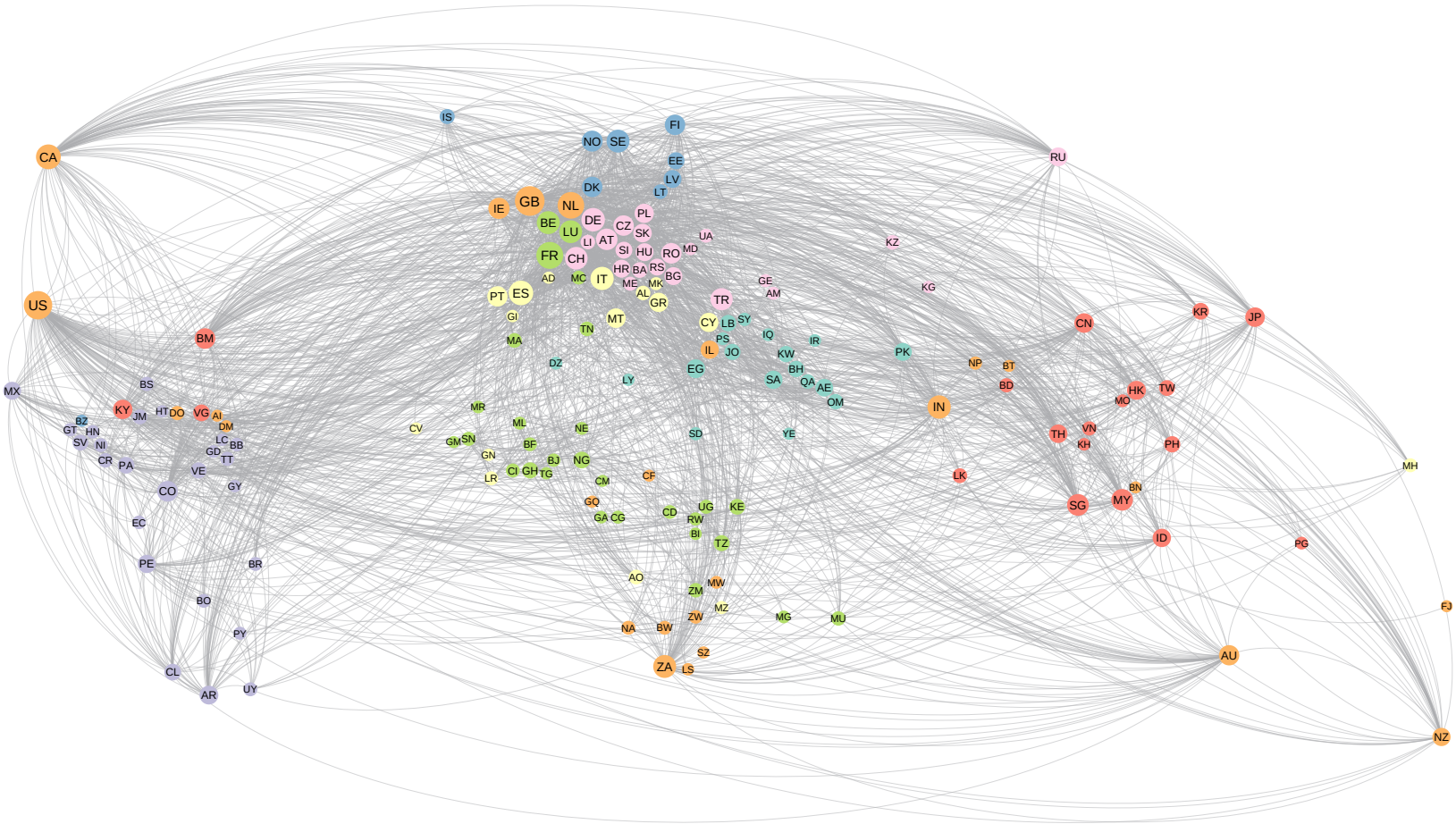
Community detection



Community detection



Community detection



Community detection

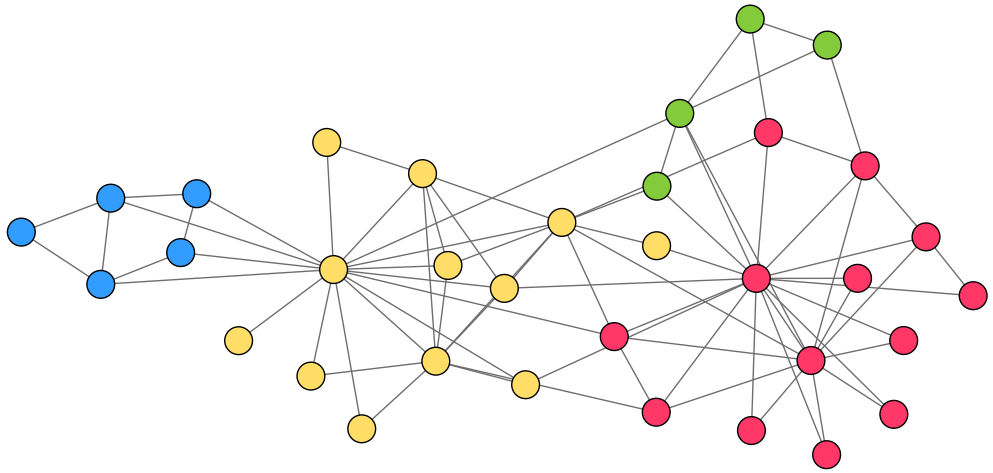


Figure: Communities: node subsets connected more strongly with each other

- Communities in corporate networks have a **regional** character and **financial ties** are clearly visible
- Historical events and cultural similarities between countries correlate with interlocks

Community detection

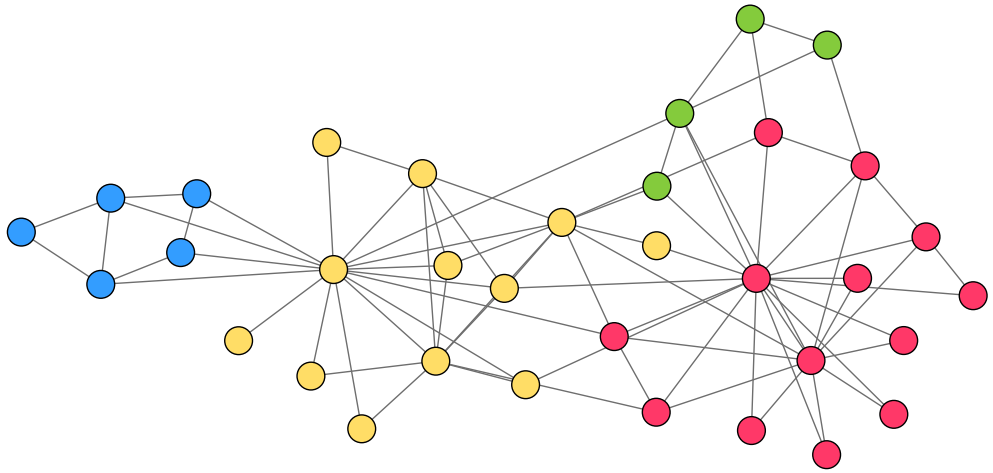


Figure: Communities: node subsets connected more strongly with each other

- Communities in corporate networks have a **regional** character and **financial ties** are clearly visible
- Historical events and cultural similarities between countries correlate with interlocks
- Outliers and effects of randomization

Computing infrastructure

- Server grade hardware
- Dual high-frequency CPU architecture with
  2 $\times$ Intel Xeon (Haswell) E5-2643, 6 cores, 12 threads, 3.4GHz
- Memory: DDR4-2133 RAM, 24 x 64GB = 1536GB = **1.5TB**
- Storage: 7TB solid state disk (SDD) storage in RAID6
- 1GBit uplink to the world

Made possible by the High Performance Computing and Networking (HPCN) fund (summer 2015 call) of the University of Amsterdam.

Universiteit Leiden

Data quality

Data quality

- Previous dataset was from September 2013
- CORPNET: study **all** firms
- More than 200 million firms
- Are all firms equally important?
- Do we have all the firms?
- What is the quality of the data?

Data quality

- Data quality
  - Accuracy: the data is true
  - Consistency: data remains clear and verifiable over time
  - Integrity: data has not suffered from corruption
  - **Completeness**: do we have all the data?
- We "found" that the Spanish market size was ten times larger than the US market: one outlier in the data.

Average operating revenue

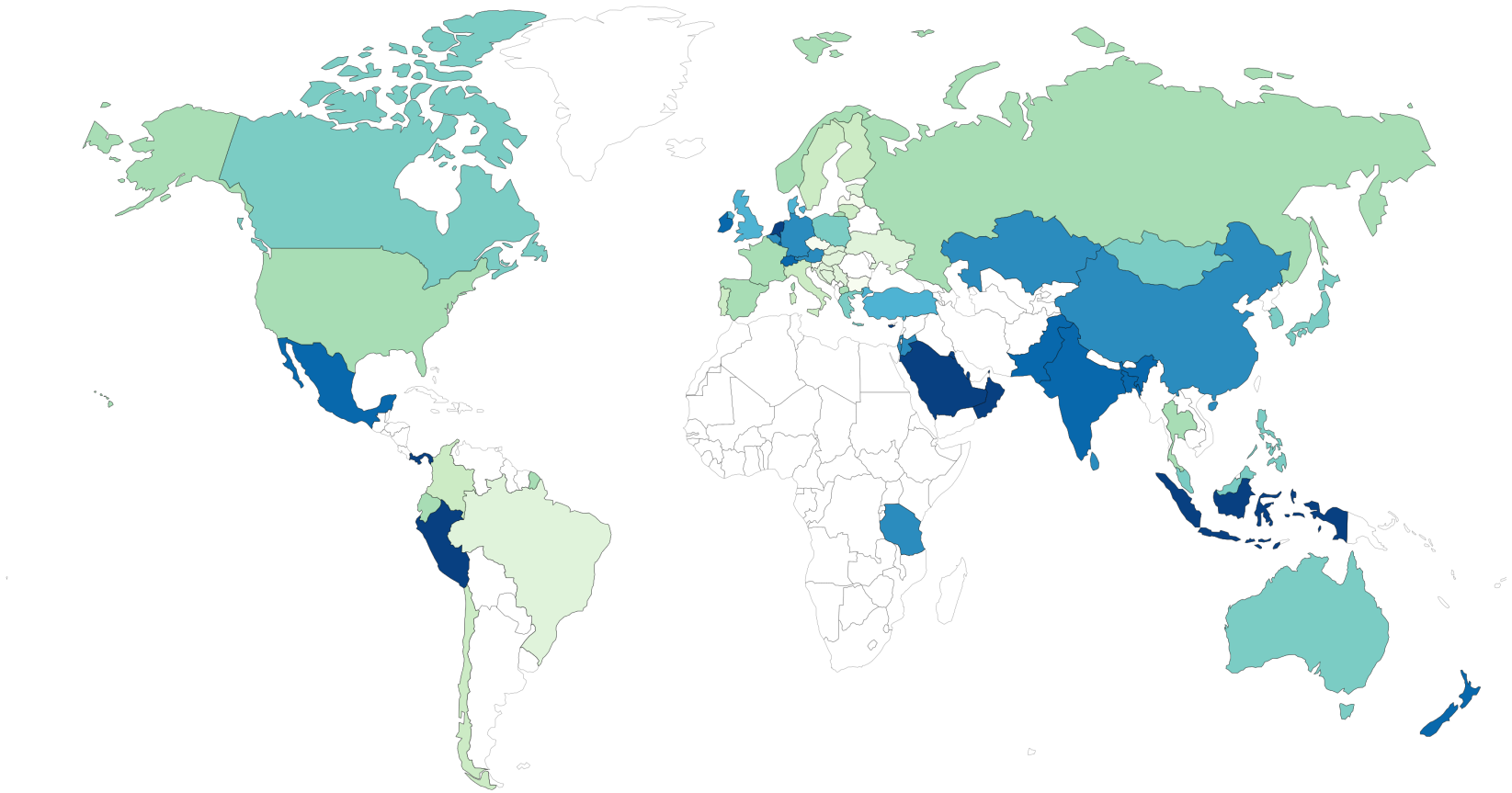


Figure: Observed average revenue per country for 200 million firms

Data quality

- Assess firm data quality based on comparing intrinsic factors of countries using:
  - Worldbank data on GDP per capita for each country
  - Eurostat data on the number of firms in each county
  - Distribution of sum of revenues per country in our data

Data quality

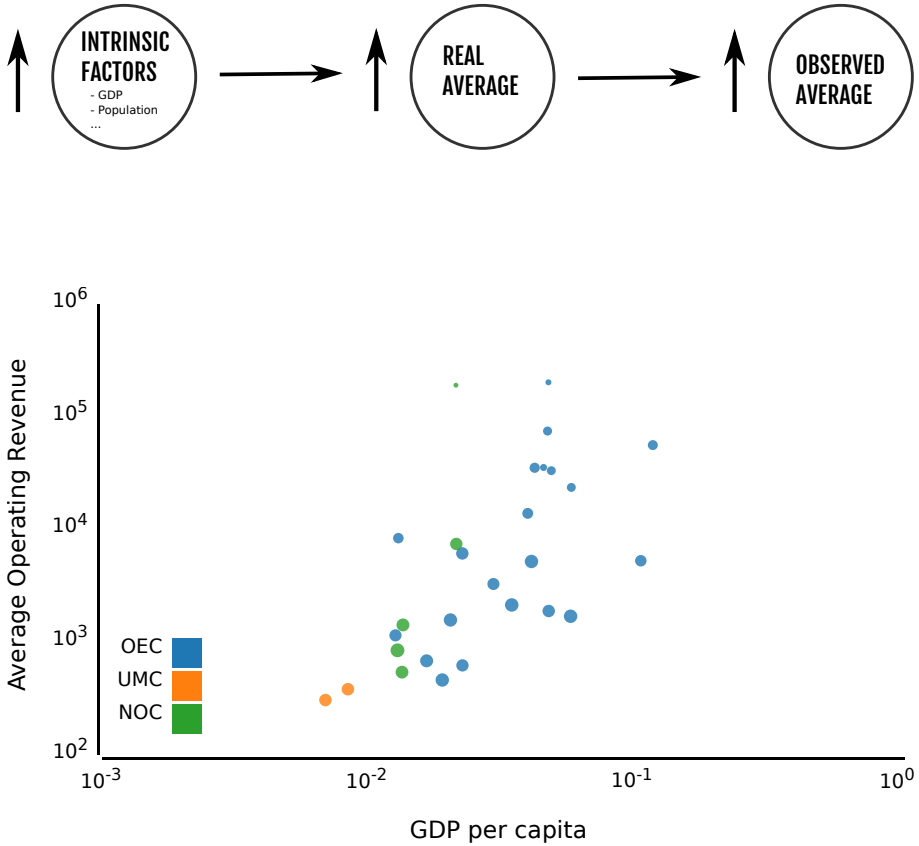


Figure: Richer countries have larger firms

Data quality

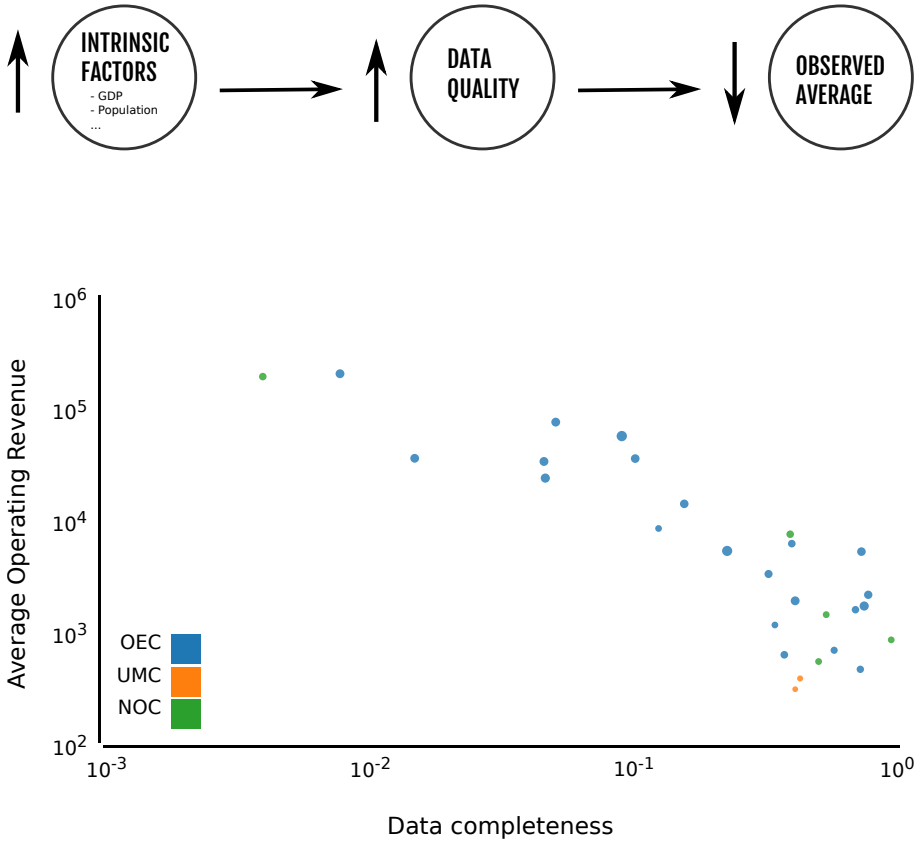


Figure: Richer countries have better quality

Data quality

- Rich countries have higher average revenue, but better quality, which decreases the observed average (hard to decouple).
- We are interested in the real average (given complete data):
  1. Real average $\propto \frac{\text{GDP}}{number\ \ of\ \ firms}$
  2. Calculate the effect of intrinsic factors and extrapolate to other countries
  3. Calculate the quality of our global firm data

Data quality

- The distribution of firm operating revenues follows a lognormal distribution for 95% of firms, with consistent variance
- Larger firms are well-represented. Richer countries have higher data quality. Higher quality decreases the observed average.

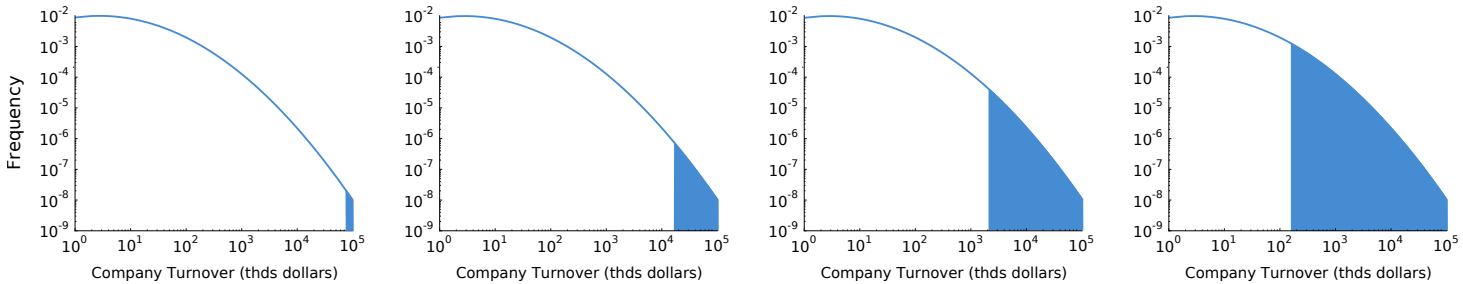


Figure: Lognormal distribution and addition of firms

Data quality

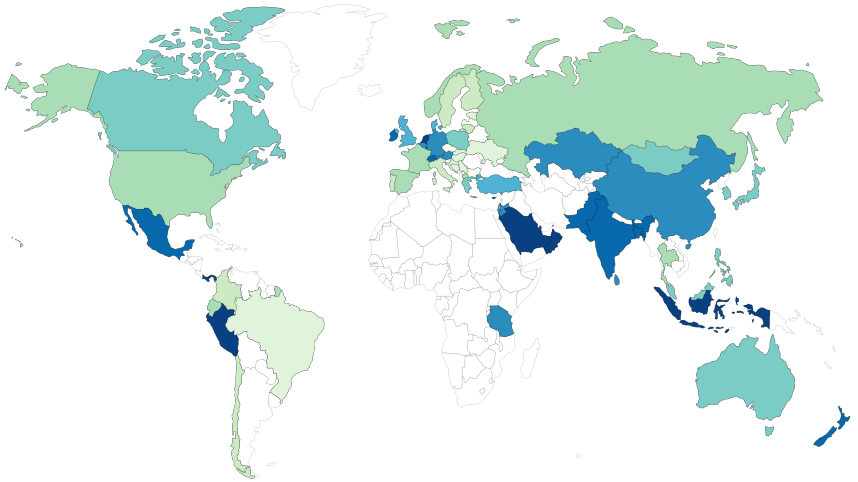


Figure: Observed average revenue

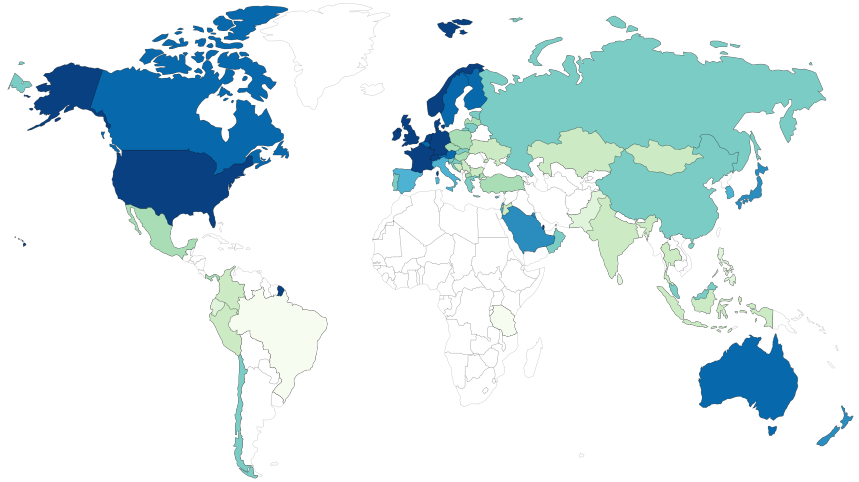


Figure: Estimated average revenue

Data quality

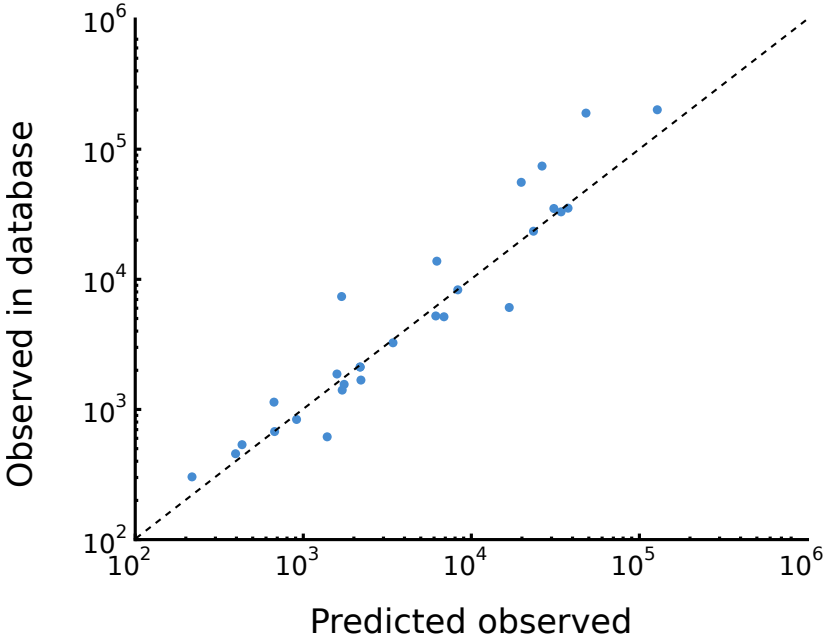


Figure: $\log(\textit{predicted observed}) = 3.15 \log(\textit{estimated real}) + \log(\textit{completeness}) - 1.05$

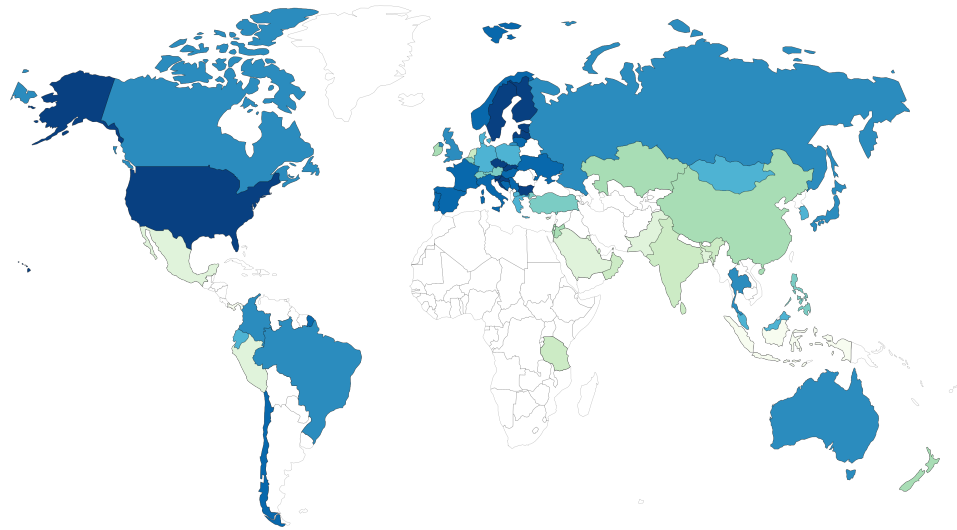
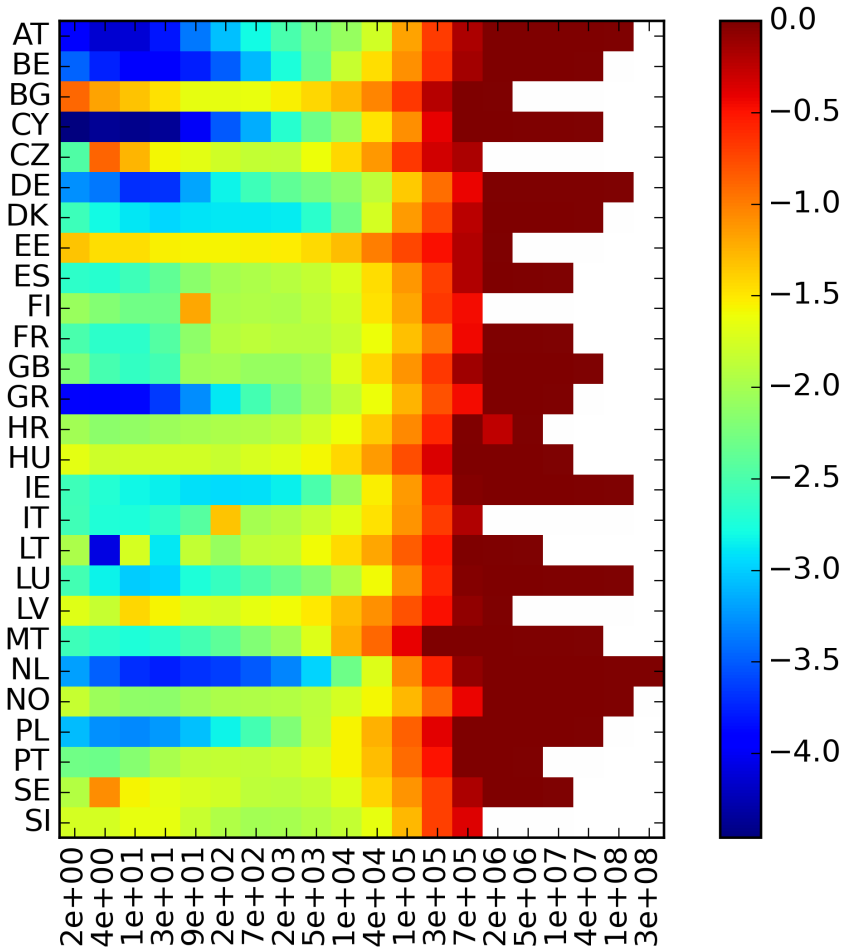


Figure: Actual completeness of our data

Completeness per country



Other directions

- Revolving doors
- Top income compensation
- Firms in occupied territories
- Public tender, procurement
- Relation with patent networks
- Exchange Traded Funds (ETFs)
- Analysis of particular national networks

Conclusion

- Big corporate network data provides interesting insight in firm power and control across the globe
- Topological properties, centrality analysis and community detection reveal regional patterns in the global network
- Interpretation of measures is crucial and depends on data quality
- We understand the completeness of our 200 million firm dataset, now we can assess the effect on the network

Conclusion

- Big corporate network data provides interesting insight in firm power and control across the globe
- Topological properties, centrality analysis and community detection reveal regional patterns in the global network
- Interpretation of measures is crucial and depends on data quality
- We understand the completeness of our 200 million firm dataset, now we can assess the effect on the network
- **CORPNET** has a challenging yet exciting time ahead!
  Website: `http://corpnet.uva.nl`
- We are open to sharing data, best practices and ideas!

Thank you!

Universiteit Leiden

- Questions?

`http://franktakes.nl`
`http://corpnet.uva.nl`