

On Data Quality, Communities and Centrality in Big Corporate Network Analysis

Javier Garcia-Bernardo, Eelke Heemskerk & **Frank Takes**

University of Amsterdam, The Netherlands

GESIS Computational Social Science Winter Symposium
Köln, December 2–3, 2015

Corporate networks



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

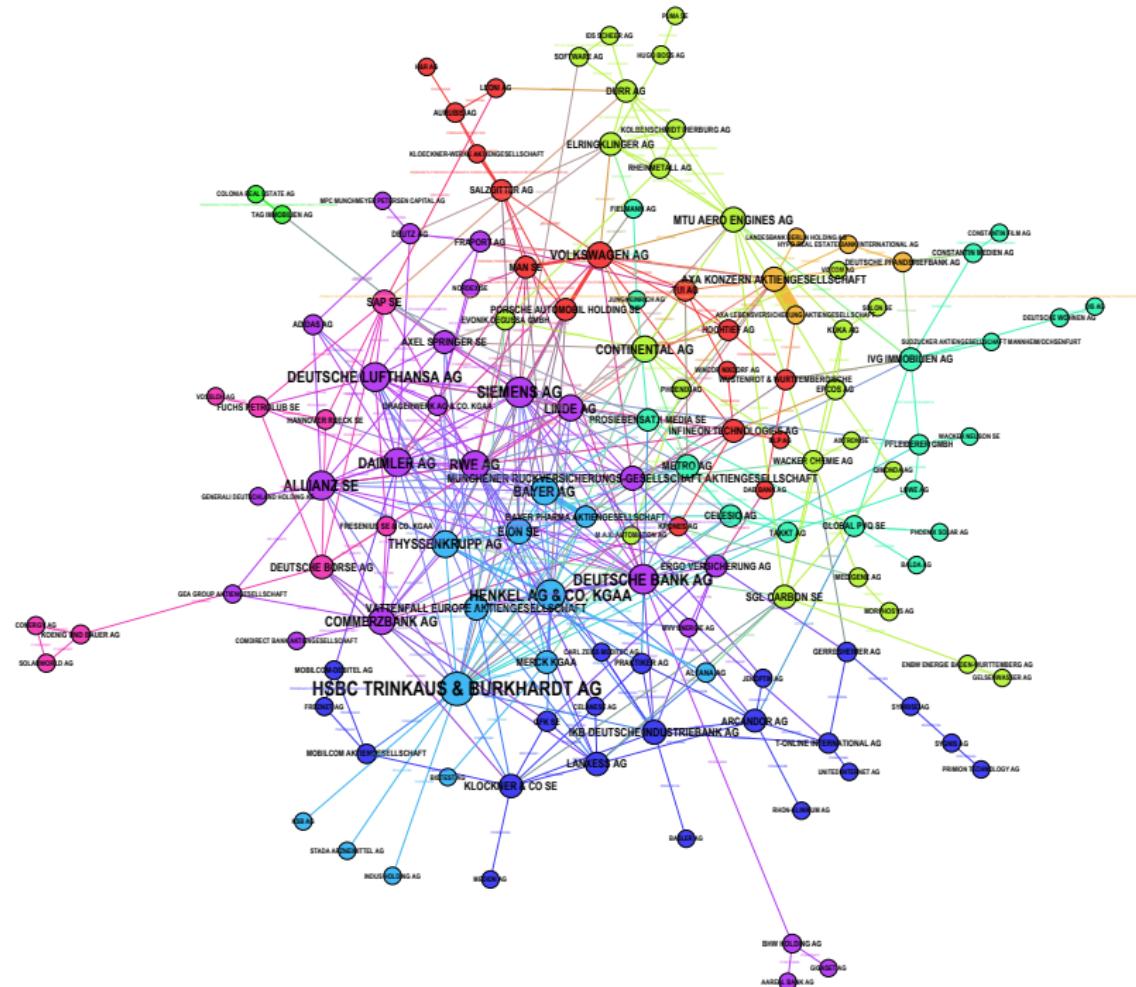
Corporate networks



Figure: Global board interlock network.

Corporate network analysis

- Apply techniques from (social) **network analysis** to corporate data
- **Nodes** represent around 200 million firms across the globe
- **Edges** could 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



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?*



European Research Council
Established by the European Commission



CORPNET

University of Amsterdam

- **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?*



European Research Council
Established by the European Commission



CORPNET

University of Amsterdam

- Previous work: **community detection** and **centrality** analysis on the largest 400,000 firms (nodes) and 1,700,000 interlocks (edges)

- **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?*



European Research Council
Established by the European Commission



CORPNET

University of Amsterdam

- Previous work: **community detection** and **centrality** analysis on the largest 400,000 firms (nodes) and 1,700,000 interlocks (edges)
- Focus in the first few months of the CORPNET project: **data quality** of the 200 million firm dataset

Community detection

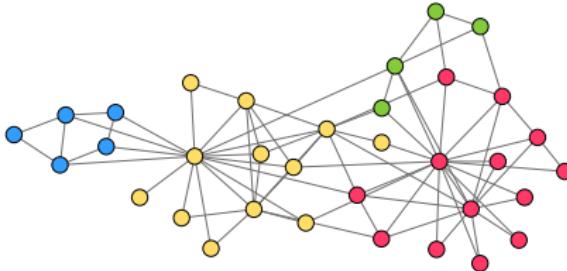


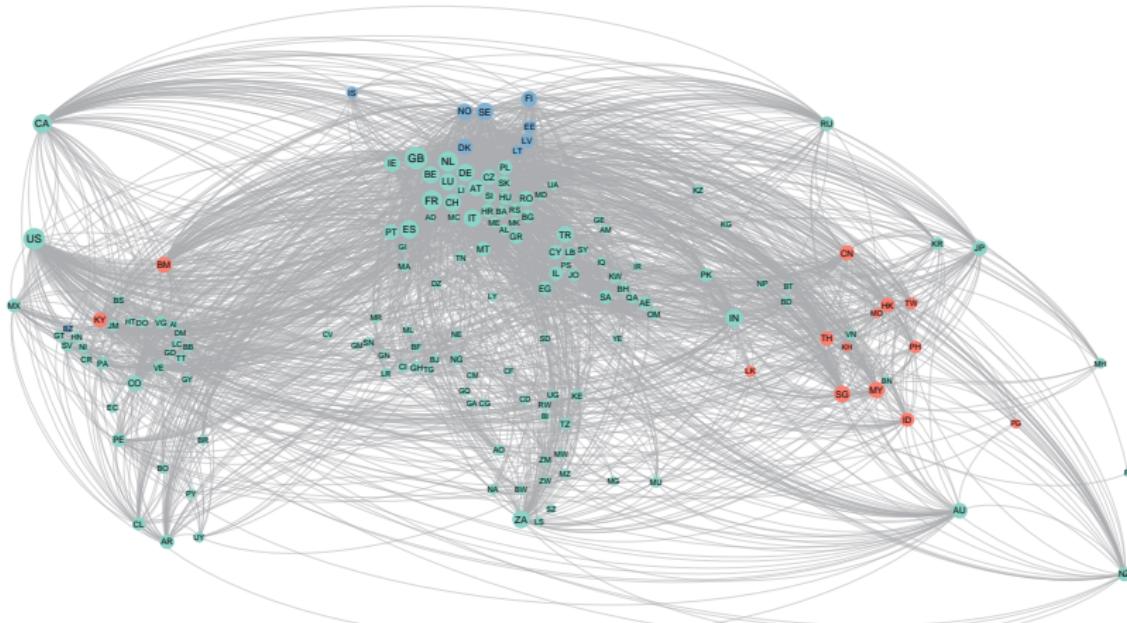
Figure: Communities: node subsets connected more strongly with each other

- Country network: aggregate firms from the same country
- 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](https://doi.org/10.1080/13563467.2015.1041483)

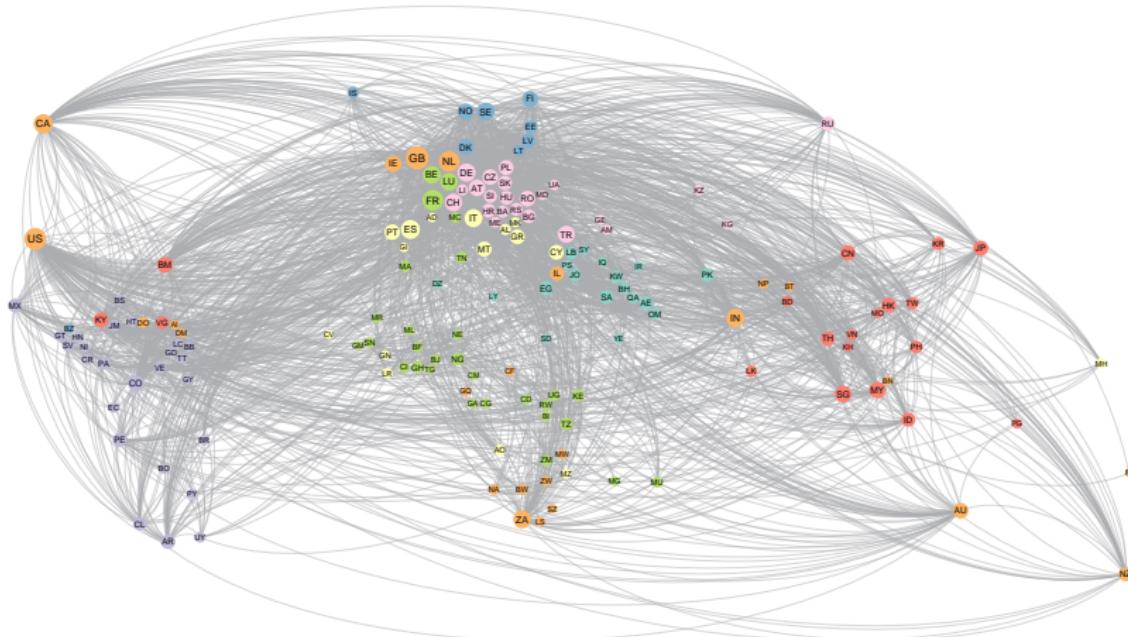
Community detection



Community detection



Community detection



Community detection

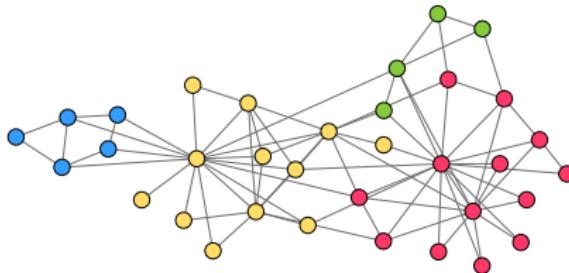


Figure: Communities: node subsets connected more strongly with each other

- 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](https://doi.org/10.1080/13563467.2015.1041483)
- Main result: communities in corporate networks have a **regional** character and **financial ties** are clearly visible

Community detection

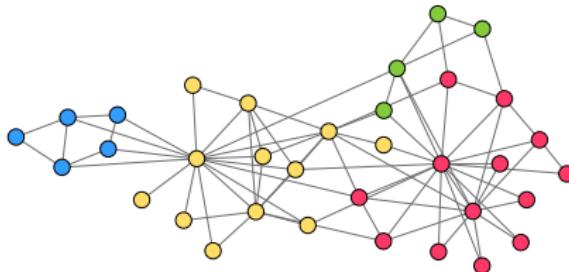


Figure: Communities: node subsets connected more strongly with each other

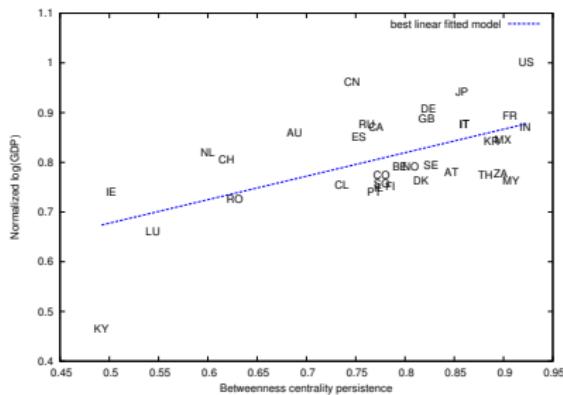
- E.M. Heemskerk and F.W. Takes, The Corporate Elite Community Structure of Global Capitalism, in *New Political Economy* 21(1): 90–118, 2016. [dx.doi.org/10.1080/13563467.2015.1041483](https://doi.org/10.1080/13563467.2015.1041483)
- Main result: communities in corporate networks have a **regional** character and **financial ties** are clearly visible
- Outliers and effects of randomization

Centrality analysis

- **Centrality** measures allow the most important nodes (based on the structure of the network) to be identified
- Degree centrality, betweenness centrality, etc.
- F.W. Takes and E.M. Heemskerk, Centrality in the Global Network of Corporate Control, *forthcoming*, 2016.

Centrality analysis

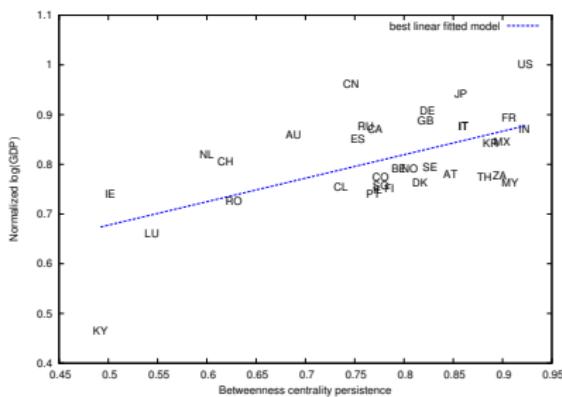
- **Centrality** measures allow the most important nodes (based on the structure of the network) to be identified
- Degree centrality, betweenness centrality, etc.
- F.W. Takes and E.M. Heemskerk, Centrality in the Global Network of Corporate Control, *forthcoming*, 2016.



- Main result: large differences between countries in terms of **centrality persistence**, measuring national vs. global power of a country's firms

Centrality analysis

- **Centrality** measures allow the most important nodes (based on the structure of the network) to be identified
- Degree centrality, betweenness centrality, etc.
- F.W. Takes and E.M. Heemskerk, Centrality in the Global Network of Corporate Control, *forthcoming*, 2016.



- Main result: large differences between countries in terms of **centrality persistence**, measuring national vs. global power of a country's firms
- Outliers due to cliques with high degree nodes of administrative interlock ties

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

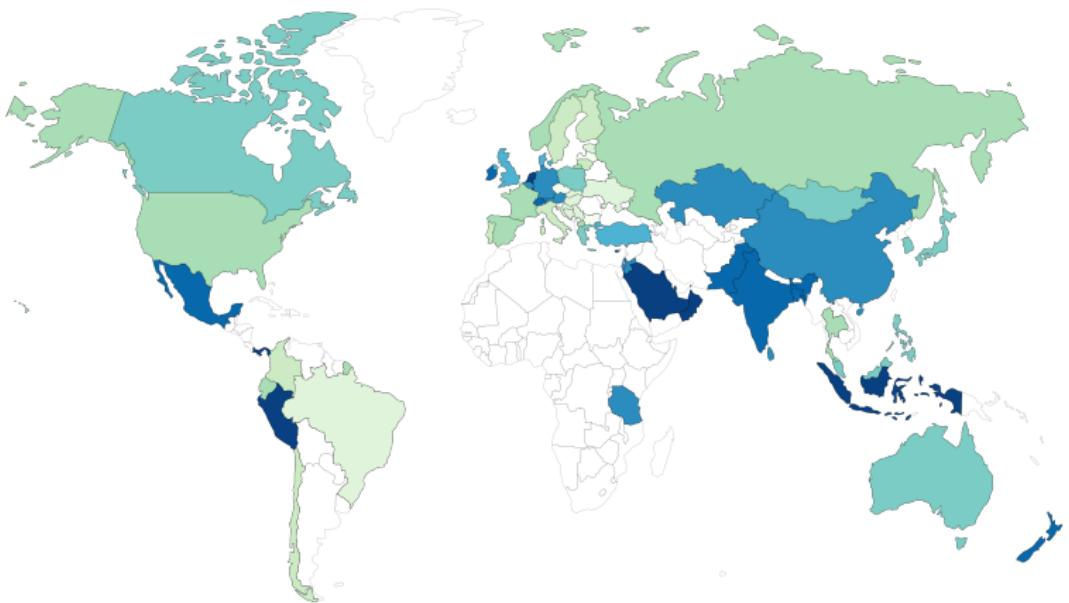


Figure: Observed average revenue per country for 200 million firms

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

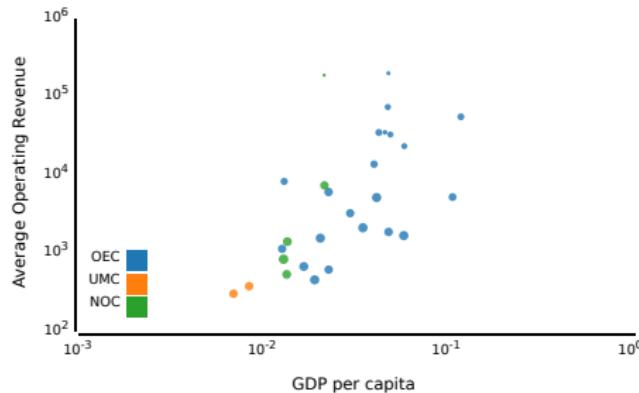


Figure: Richer countries have larger firms

Data quality

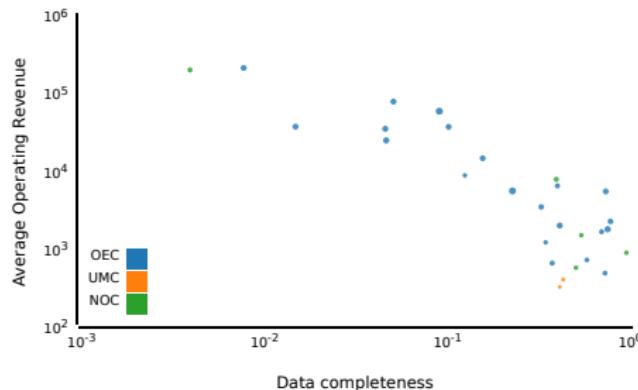


Figure: Richer countries have better quality

- 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 $\propto \frac{\text{GDP}}{\text{number of firms}}$
- 2 Calculate the effect of intrinsic factors and extrapolate to other countries
- 3 Calculate the quality of our global firm data

Data quality

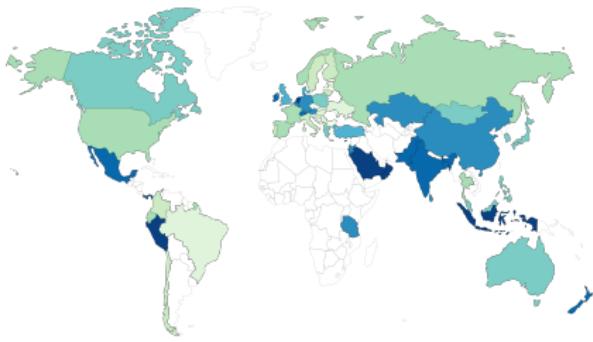


Figure: Observed average revenue

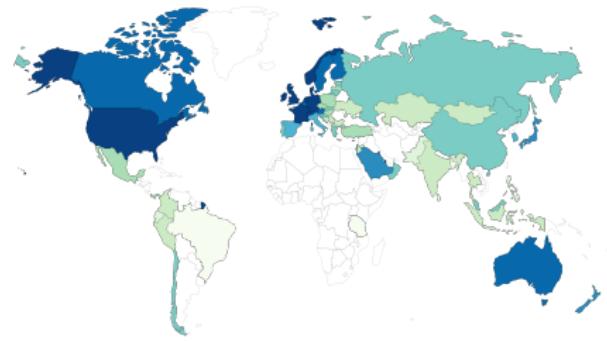


Figure: Estimated average revenue

Data quality

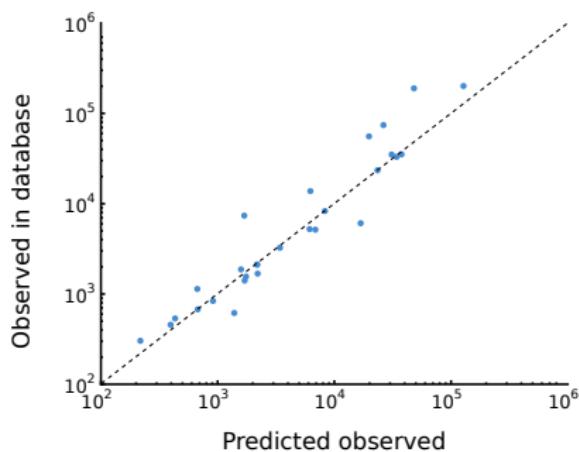


Figure: $\log(\text{predicted observed}) = 3.15 \log(\text{estimated real}) + \log(\text{completeness}) - 1.05$

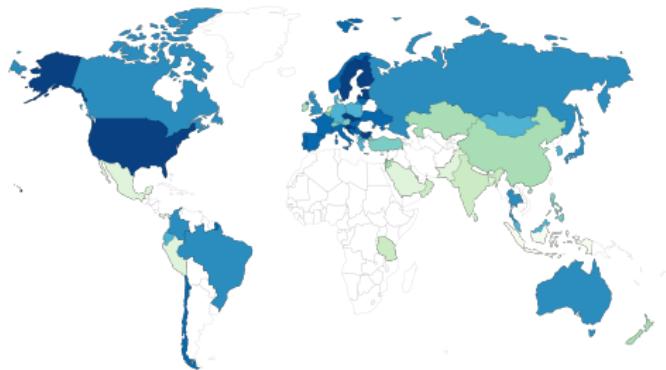
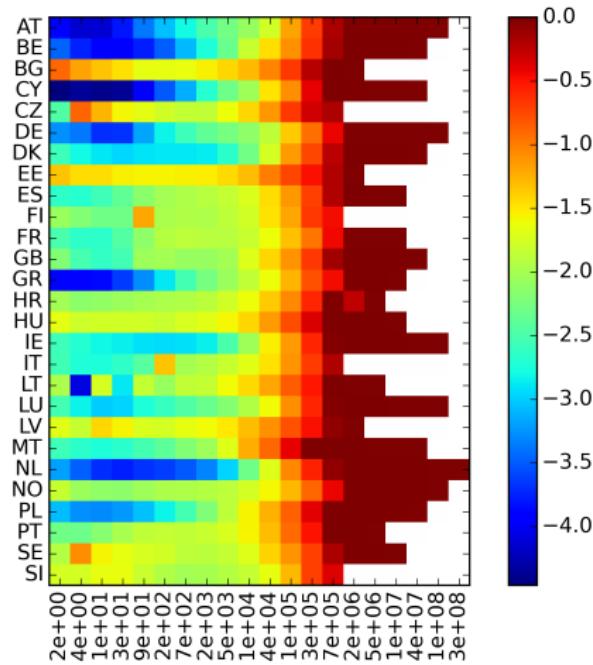


Figure: Actual completeness of our data

Completeness per country



Conclusion

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

Conclusion

- Big corporate network data provides great insight in firm power and control across the globe
- Topological properties, centrality analysis and community detection all show regional patterns in the global network
- Interpretation of measures is crucial and depends on data quality
- We understand the completeness of our 200 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!

- Questions?



<http://corpnet.uva.nl>