Topology and Community Structure of the Global System of Corporate Control

Frank W. Takes
LIACS, Leiden University &
AISSR, University of Amsterdam
ftakes@liacs.nl takes@uva.nl

May 23, 2016

Introduction

With the nowadays abundantly available amount of data, there is an ever increasing need to better understand and find patterns in this data. Data objects are usually not mere individual entities within some dataset, but in fact they interact, communicate or relate to other data objects. Think of an online social network, in which users (data objects) interact or form explicit friendships with other users, or how webpages on the internet are linked by means of the hyperlinks on these pages. Data that has some sort of relational aspect is typically modeled using the notion of a *network*, in which *nodes* (also called vertices, objects or actors) are linked by means of *edges* (also called links, relationships, ties or lines). There is wide interest in analyzing data using the notion of a network: mathematicians and computer scientists traditionally have a fierce interest in the underlying mathematical structure of a *graph*, and "graph mining" is an active research field working on finding patterns in this type of data. Physicists study network data as "complex networks", and people in the social sciences have been studying networks of interacting people under the umbrella term "social network analysis". Today, these fields come together within the field of *network science*, an interdisciplinary research field with network data as the common object of study [2].

In this article we focus on the *global corporate network*, modeling the relationships that exist between firms, corporations and organizations in our global economy. We live in a highly connected world, where firms do not operate as individualistic market actors, but are instead part of a connected network of business. For example because firms trade with each other, because they own a percentage of one another, or because they lend money to each other. Here, we focus on social ties between firms. We say that two firms are connected by an undirected edge if the two firms are governed by at least one common director or executive board member, resulting in the so-called *board interlock network* (see Figure 1 for an example). Directors or CEO's often have more than one appointment, and it is well known that their involvement in multiple firms, referred to as a board *interlock*, opens up these firms to other firms's information, resources and expertise, and in general fosters social cohesion [8]. There is an enormous body of research in the social sciences about the effects and consequences of board interlocks, dating back more

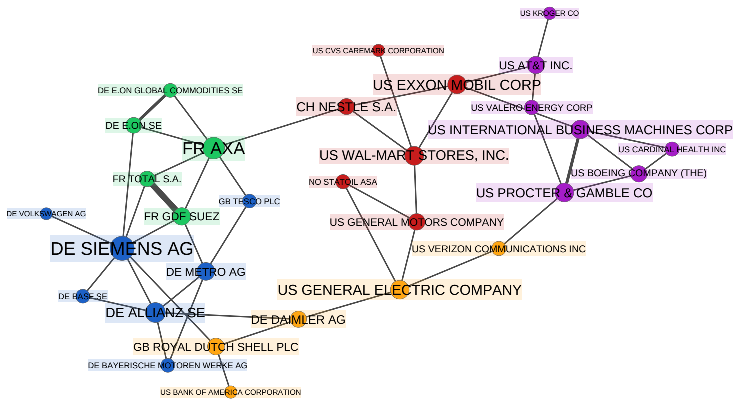

Figure 1: A sample of the board interlock network.

than 100 years to (amongst others) Vladimir Lenin's 1916 book "Imperialism, the highest stage of capitalism" [7], in which Lenin notes using a case study of Germany that firms and banks often share directors for a number of reasons and with a number of possible consequences.

The interdisciplinary CORPNET research group at the University of Amsterdam (see `http://corpnet.uva.nl`) takes advantage of recent developments in the field of network science. It aims to better understand the power-political causes and consequences of the network of corporate control by studying a large-scale dataset consisting of millions of firms connected via hundreds of millions of ties based on board interlocks and ownership.

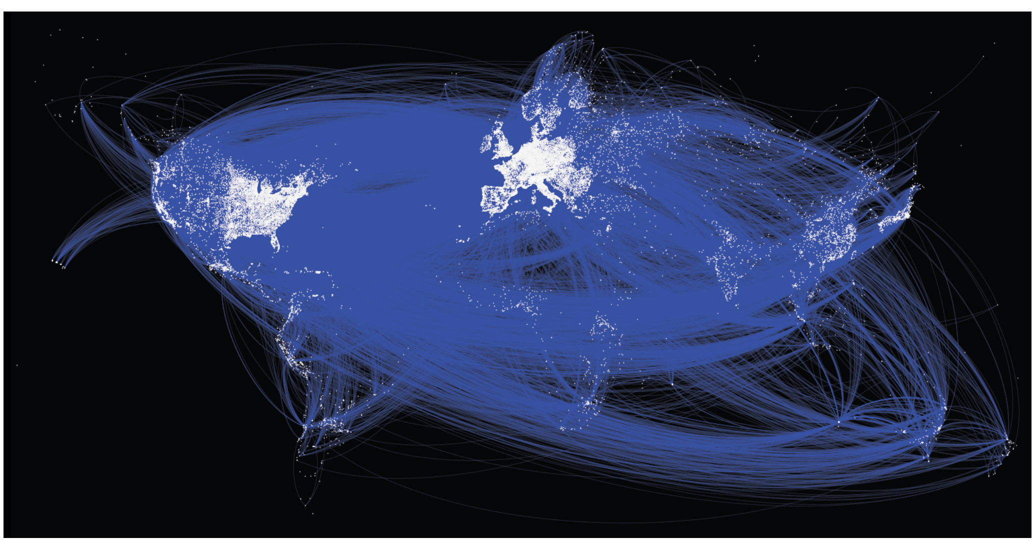

Figure 2: Geographical visualization of the global network of corporate control: around $400,000$ firms and over $1,700,000$ board interlocks.

Since 2013, a sample of this network was studied that consisted of the largest one million firms across the globe. A geographical visualization of this network is given in Figure 2, highlighting how in our global economy, firms are densely connected through board interlocks. We will further investigate the structure and patterns in this network in the remainder of this paper.

Network topology

The considered board interlock network, linking firms if they share at least one senior level board member or director, consists of $391,967$ firms that form at least one interlock, in total having more than $1,711,968$ interlock ties between them. Some properties of the network are given in Table 1. It turns out that the structure of the global network resembles that of many other real-world networks. Specifically, it is sparse (measured by the low density), meaning that there are very few links compared to the theoretically maximal number of links. Furthermore, there is a power law degree distribution (see Figure 3), meaning that the number of nodes with very few connections is large, whereas there are a smaller number of hub-like nodes with a very high number of connections. Here, the tail of this distribution is the result of a few (note the logarithmic axes) firms having a large number of economic entities between which all directors are shared. Moreover, nodes densely cluster together, forming a larger than random number of closed triangles of connections, as measured by the clustering coefficient of 0.76. A randomly generated network with the same degree distribution would have a clustering coefficient of less than 1.0×10^{-6}. Not all firms are directly or indirectly connected, but there is one giant component of size $238,859$ connecting the majority of the nodes. The vast majority of the smaller components (all with 60 nodes or less, distribution in Figure 4) represent simple "parent/subsidiary"-structures from the same country that do not share directors with

Table 1: Global network properties.

Global corporate network	
Nodes (firms)	$391,967$
Edges (interlocks)	$1,711,968$
Density	$2.229 \cdot 10^{-5}$
Average degree	8.746
Connected components	$55,616$
Giant component	
Nodes (firms)	$238,859$ nodes (60.9%)
Edges (interlocks)	$1,533,030$ (89.5%)
Density	$5.374 \cdot 10^{-5}$
Average degree	12.83
Clustering coefficient	0.751
Average distance	7.775
Radius	18
Diameter	34

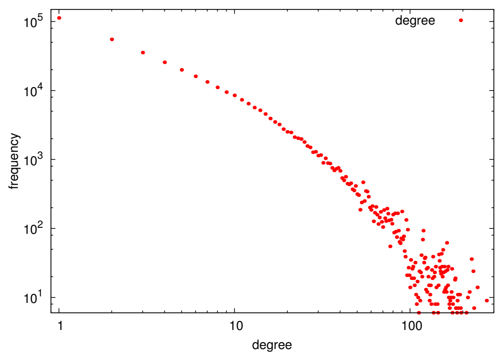

Figure 3: Degree distribution.

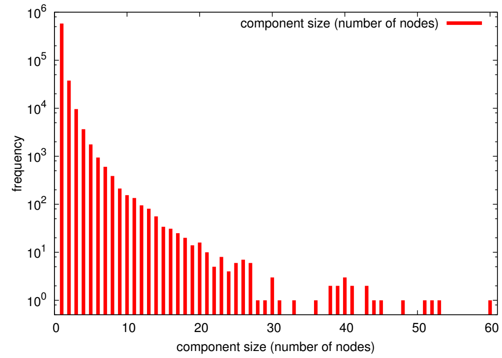

Figure 4: Component size distribution.

the giant component. Going from the full network to the giant component, the number of edges drops only by 10%, indicating that the majority of interlocking activity is captured in the 90% of edges that reside within the giant component, which is why the main focus of our study is on this giant component.

Despite the fact that the network is very sparse, the average distance (i.e., the minimal number of edges) between two nodes is relatively low, as can be seen in Figure 5. This is known as the *small world* property [6], a phenomenon that occurs in a number of real-world networks, such as (online) social networks, information networks such as Wikipedia and webgraphs. Indeed, the corporate boards of firms across the globe are on average connected in 7.775 steps. To put this into context: given that boards of the larger companies typically meet once every month, it is often anecdotally noted that a deadly disease among the corporate elite could wipe out the majority of corporate leaders in little over half a year [3]. However, although the average is low, the extremes are much higher. The *eccentricity* of a node indicates the length of a longest shortest path (maximal distance) from that particular node, and the radius and diameter are the minimal and maximal eccentricity values (see Table 1 and in particular Figure 6). Given a (too simple) model in which the disease starts at a random firm in the network and each month

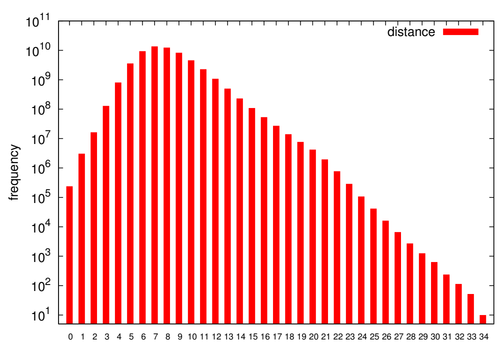

Figure 5: Distance distribution.

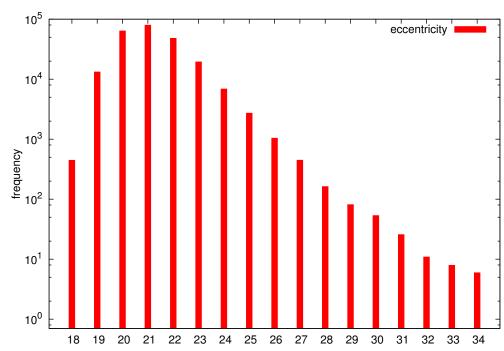

Figure 6: Eccentricity distribution.

spreads to all its neighbors, a very tiny (note the logarithmic vertical axis of Figure 6) number of firms will survive at least 18 months and depending on where the disease started, at most up to 34 months. There are a number of challenges related computational complexity involved in computing the toplogical properties, distances and eccentricities in large networks, and the interested reader may see for example [9, 10].

Community detection

In traditional data mining on tabular data, clustering can be used to find data objects with similar attributes. In network data, it may very well be that certain groups of nodes are more connected with each other, than with the rest of the network, resembling attribute similarity in clustering. Such groups of nodes are called *communities*, and *community detection* algorithms take as input the structure of the network and output a division of this network into (usually non-verlapping) communities (also called partitions). We typically then indicate these communities using colors such as is done in Figure 1.

One way of finding communities is by *modularity maximization*. Modularity is a number that indicates the quality of a particular division of a network into communities, where a higher number represents a better division into communities. Well-known heuristic search algorithms such as hill-climbing methods, genetic algorithms or local search, e.g., the popular Louvain algorithm [1], can then be used to optimize the modularity value. These algorithms typically start with each node assigned to its own community, and then iteratively merge two nodes (or a node and a previously merged community) into the same community (node), as long as the value of the modularity increases. The iteration at which maximum modularity is attained then gives the optimal number and division of the network into communities. The algorithm can furthermore take a so-called *resolution parameter* that indicates how tough the algorithm should look for communities (at the sacrifice of the quality of the solution), resulting in more or fewer communities depending on whether we are considering a high or low resolution.

If we use community detection on the corporate board interlock network, we obtain, for a particular “low”, “medium” and “high” resolution, the divisions into communities shown in Figure 7, 8 and 9. Here, we have aggregated the firms into nodes representing countries, connected through weighted links denoting the number of firms that share directors between the linked countries. The position of nodes is determined based on the latitude and longitude of the center of the countries they represent.

In Figure 7, we see how the first communities that “appear”, i.e., are visible at the lowest resolution, show a clear regional character. There is a Scandinavian/Baltic community, indicating that apparently, firms in these countries are more connected with each other than with the rest of the world. There is furthermore a community around China and other Asian countries, corresponding with the frequently made observation that although now participating in the world-wide economy, Asia is not that well integrated with the rest of the world. The outlier cases of Bermuda and the Cayman Islands underline the importance of a sensible interpretation of large-scale network analysis results. These countries are not per se part of an Asian business community, but upon inspection of the underlying data appear to be linked via a number of ties to real estate firms in Malaysia. Indeed, the Cayman Islands have frequently been identified as a tax haven. There is furthermore an African community containing a number of tightly connected former French

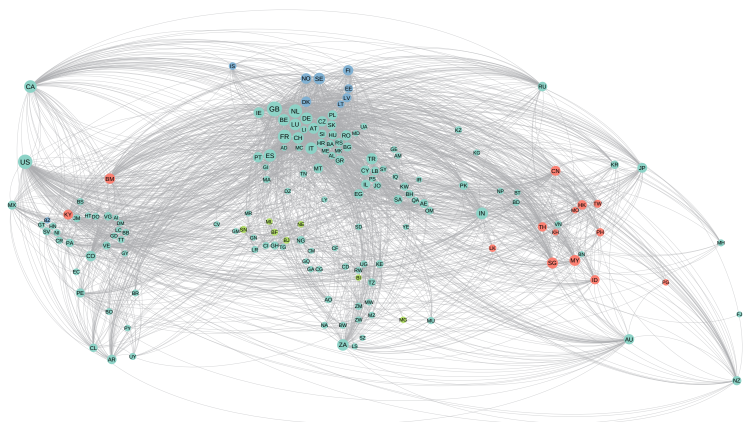

Figure 7: Communities in the aggregated global board interlock network, “low” resolution.

colonies. Yet apparently, France itself has more edges to the western world than to these countries, so is in a different community.

At a bit higher resolution, Figure 8 shows a more fine-grained division of the network into communities. Here we see how a central European community forms, and how the middle East distinguishes itself from the rest of the world in a separate community. A number of African countries now connects to a large community of mostly western countries, including the United States, Canada, Ireland and the Benelux countries, but also Australia, India and New Zealand. Indeed, this community includes most of the former British commonwealth. Latin America has separated from the rest of South America due to its strong ties with southern European countries.

In an even more fine-grained division of the network into communities as shown in Figure 9, southern Europe falls apart into a Russia-oriented and Mediterranean-oriented community. We furthermore observe strong ties between the US, Canada, Ireland and the Netherlands, hinting towards a common well-known advantageous fiscal construction. Although this pattern is interesting, it does not resemble the real social ties between corporations that interlocking directorates research focuses on (as discussed in the beginning of this paper). Rather, this pattern indicates the influence of administrative ties on the observed results. This again highlights the importance of sensible interpretation step of network analysis results.

In summary, we note that although globalization has led to a world-wide connected economy, the resulting communities have a clear regional character, showing cultural ties as well as former colonial patterns. Interestingly enough, none of this type of data was put into the analysis; the only input of the algorithm was the board interlock network. For a more thorough discussion and interpretation of these results, the reader is referred to [4, 5].

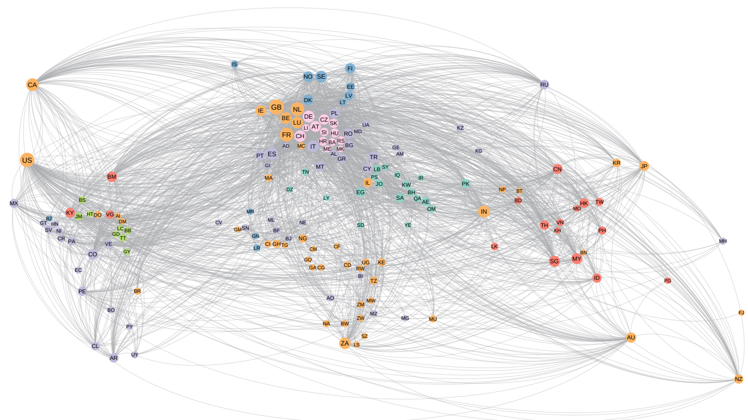

Figure 8: Communities in the aggregated global board interlock network, "medium" resolution.

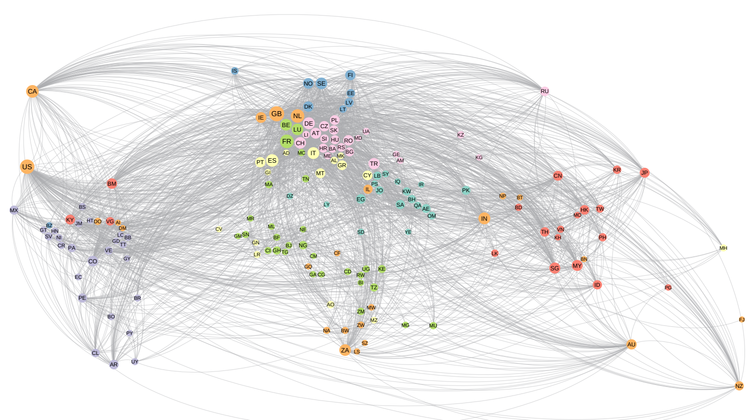

Figure 9: Communities in the aggregated global board interlock network, "high" resolution.

Concluding remarks

Throughout this article, we have highlighted the results of using network analysis techniques to better understand corporate network data based on board interlocks. Network analysis reveals a number of patterns that are not evident in the underlying data, but become visible in the network perspective. By modeling the data as a system of interaction rather than a set of objects, we are able to better understand the dense connectedness of the "small world" global economy. Futhermore, network community detection discovers cultural, historical, geographical and financial patterns that are far from visible in the underlying raw data, demonstrating the added value of *network science* for extracting knowledge from large-scale interaction data.

References

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

[2] U. Brandes, G. Robins, A. McCranie, and S. Wasserman. What is network science? *Network Science*, 1(01):1–15, 2013.

[3] G. F. Davis. Agents without principles? The spread of the poison pill through the intercorporate network. *Administrative Science Quarterly*, 36(4):583–613, 1991.

[4] E. M. Heemskerk and F. W. Takes. The community structure of the corporate elite of global capitalism. *New Political Economy*, 21(1):90–118, 2016.

[5] E. M. Heemskerk, F. W. Takes, J. Garcia-Bernardo, and M. J. Huijzer. Where is the global corporate elite? A large-scale network study of local and nonlocal interlocking directorates. *arXiv 1604.04722, 29 pages*, 2016.

[6] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 163–170, 2000.

[7] V. I. Lenin. *Imperialism, the highest stage of capitalism*. 1916.

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

[9] F. W. Takes. Afstanden in online sociale netwerken. *Eureka! Magazine*, 9(33):8–11, 2011.

[10] F. W. Takes and W. A. Kosters. Computing the eccentricity distribution of large graphs. *Algorithms*, 6(1):100–118, 2013.

This text appeared in Eureka! Magazine volume 14, number 54, pp. 20–23, 2016.