Using the Panama Papers to Explore the Financial Networks of the Middle East

Abdullateef Rabab'ah,* Mahmoud Al-Ayyoub,* Mohammed A. Shehab,*
Yaser Jararweh* and Bernard J. Jansen ‡
* Jordan University of Science and Technology, Irbid, Jordan
‡ Qatar Computing Research Institute, HBKU, Doha, Qatar
{maalshbool, yijararweh}@just.edu.jo, jjansen@ist.psu.edu

***Abstract*—In what has been described as the WikiLeaks of the financial world, the release of millions of documents (known as the "Panama Papers") have placed at the center of global media attention the elaborate ways used by some of the elite to hide their financial assets leading to serious allegation of financial corruption. In this work, we explore the information contained in these documents using social network analytics. Due to the large size of the network constructed from the Panama Papers, we limit our attention to a specific region, which is the Middle East. The analysis reveals that while the constructed network enjoys some typical characteristics, there are many interesting observations and properties worth discussing. Specifically, using the extracted network consisting of 37,442 nodes and 79,544 edges, our social network analysis finding show that, perhaps surprisingly, the nodes or the social network are not necessarily directly correlation with perceived financial influence.**

***Index Terms*—Social Network Analysis; Panama Papers; Middle East.**

I. Introduction

During April 2016, 2.6 terabytes of data containing 11.5 million confidential documents about more than 214,000 shell companies from more than 200 countries was leaked from the Panama-based law firm, Mossack Fonseca. These documents contained transactions spanning more than four decades and revealing how wealthy individuals and powerful corporations from around the world have been using elaborate and complex corporate ownership and control structures and offshore lightly-regulated tax havens to hide their beneficial ownership of companies and, thus, obscure their personal financial assets. While such things might have legitimate and legal grounds, they are often looked at with scrutiny, since there are many cases where they were used for tax evasion or for hiding criminal activities. The documents also revealed the need for more and stricter disclosure and reporting rules in addition to having a platform for global tax administrations where the different types of economies (developed and developing) work closely together and exchange information in order to better mitigate the effect of aggressive tax evasion techniques [1]–[7].

The release of these document sent waves of shock throughout the world with serious allegations against top government official, as well as companies, ranging from tax evasion to financial corruption and money laundering. In fact, in a simple study of a subset of the documents, McGee [5] reported more than 80 studies on tax evasion and 11 studies on bribery. Even the less serious allegations, such as the lack of transparency, have lead to dramatic consequences on may politicians and government officials such as the resignation of Iceland's Prime Minister [1], [2]. Understanding the social relationships identified within the Panama Papers may lead to a clearer understanding of how these methods of offshore structures are employed in different regions of the world.

In this work, we use social network analytics to explore the information within the Panama Papers. A network has been constructed based on the released documents. However, the constructed network is very large. So, we filter it out to keep only the parts pertaining to Middle Eastern entities in order to investigate entities within the Middle East North Africa (MENA) region, which has received little to no prior investigation. We then compute several centrality measure and report our observations in order to identify the key MENA players, if any, from this data set. The contributions of this work are novel as the data is still new and the existing efforts on using a social network analysis (SNA) approach to study it are limited. See Section II.

The rest of this paper is organized as follows. The following section discusses the efforts invested on relevant tasks. Section III discusses the steps we follow in this work. Finally, the paper is concluded in Section IV with a brief discussion of the future directions of this work.

II. Related Works

In this section, we discuss the efforts relevant to this study. We start with general works about network analysis before discussing the works pertaining to the Panama Papers.

With the rise of online social media platforms, we have witnessed surges of interest in complex networks and their analysis leading to several breakthroughs [8]–[11]. One type of networks that is of interest to us is constructed by considering the relationships between different corporate actors and entities such as investment flows, and co-ownership ties creating what some authors have called corporate or financial networks [11]. While earlier work on such networks focused on manual inspection of Top 100 or Top 500 global firms (as provided through lists such as the Fortune 500) [12]–[16]. or on dozens or hundreds of firms in a certain region, such as Europe [17]–[19], other works considered much larger networks consisting of hundreds of thousands of firms world wide [20], [21] leading to what is called *Big Corporate Network Data (BCND)* [11]. Nonetheless, these networks are often built based on public information. Of course, data leaks such as the Panama Papers can lead to more insightful construction of such networks.

The focus of this work is on the Panama Papers. This leak received a lot of attention across the globe. Some papers prior research focused on revealing unethical/illegal actions such as [2], [5], while other research focused on studying the potential bias in the media coverage of this leak [22]. Other more technical direction explored the way journalists and investigators explored the leaked data.

The International Investigative Journalist Consortium (ICIJ) claims that more than 370 journalists working in more than 25 languages and over 76 countries searched through the approximately 2.6 terabytes (the equivalent of approximately 600 DVDs) of leaked data [23]. According to [24], the ICIJ used a search engine and a visualization tool to navigate through the massive amount of data extracted from the Panama Papers. The provider of this tool, a company called Linkurious, designed the interface to cater for non tech-savvy investigators interactively searching through data [25].

The SNA community took immediate interest in the Panama Papers. Within a month after their release, Dmitry Zinoviev (of Suffolk University, NY, USA) wrote a Python code to perform simple analysis on the network constructed from the Panama Papers. The analysis focused on the size and the components of the network.[1] Another noteworthy effort was made by Moses Boudourides and Sergio Lenis (of the University of Patras, GR) who performed detailed analysis of the Panama Papers network focusing on the entities pertaining to Greece, Cyprus and Russia.[2]

III. Methodology

This section discusses the methodology we follow in this work. The aim is to use social network analytics to explore the information within the Panama Papers. Specifically, a dataset has been released by the Organized Crime and Corruption Reporting Project (OCCRP),[3] and it has been processed to create a social network where the nodes are companies, individuals, etc., and the edges represent relations connecting them such as "Beneficiary of" and "Shareholder of". Then, the network data is transformed into a format readable by network analysis packages such as Pajek [26].[4] The resulting data is too large to be processed efficiently, so, we limit our attention to one specific region. The region of choice is the MENA due to its strategic importance , especially as a world financial center. Given the popularity and widespread employment of these shadowy tactics around the globe, one would expect MENA actors to also be involved. We filter out the network to keep only the nodes pertaining to Middle Eastern entities and the edges between them. We then perform social network analysis steps to identify the most central nodes in the network. The analysis aims to reveal whether any node plays special roles any special role such as being a "Gray Cardinal" or "Boundary Spanner" [27].

In the following subsections, we present a quantitative description of the constructed network. We then discuss visualizing and analyzing the network with a focus on the ME region.

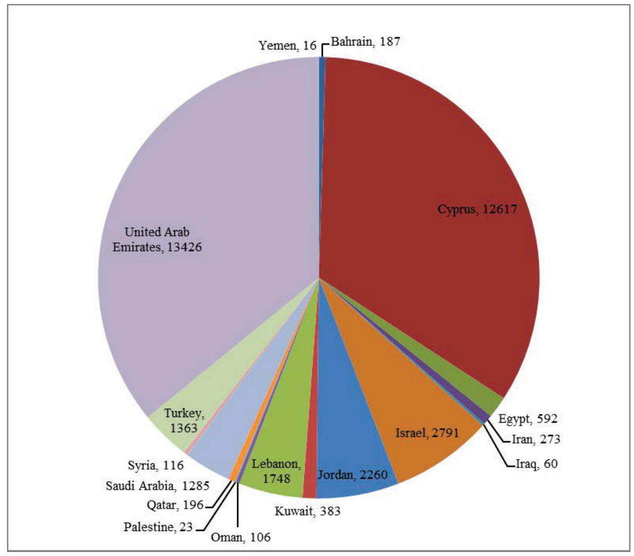

Fig. 1. Nodes distribution over ME countries.

TABLE I
Edges Distribution Over Relation Types

Relation type	Number of instances
Auditor of	1
Authorised Person / Signatory of	4
Bank Signatory of	1
Beneficial Owner of	23
Beneficiary of	1,894
Connected of	3
Director of	7,603
Intermediary of	20,148
Investment Advisor of	1
Nominated Person of	1
Officer of	1
Owner of	6
Power of Attorney of	9
President of	2
Protector of	3
Register of Shareholder of	1
Registered address	15,890
Same address as	51
Same name and registration date as	51
Secretary of	854
Shareholder of	30,664
Similar name and address as	2,550
Trust Settlor of	10
Trustee of Trust of	72
Vice President of	1

[1] https://www.linkedin.com/pulse/panama-papers-exercise-network-analysis-dmitry-zinoviev

[2] http://mboudour.github.io/2016/09/24/Interlocking-International-Networks-from-Panama-Papers-GR-CY-RU.html

[3] https://www.occrp.org/en/panamapapers/database

[4] http://vladowiki.fmf.uni-lj.si/doku.php?id=notes:net:pa

A. Network Information

The Panama dataset we study in this research consists of 838,105 nodes and 126,9796 edges. Given this huge number of nodes and edges, it is difficult to navigate through the network and perform analysis over the relations between the nodes. In this research, we focus mainly on the Middle East (ME) region network.

To extract the ME region network from the Panama network we keep edges connecting nodes from the ME at least one end point. The extracted network consists of 37,442 nodes and 79,544 edges. Figure 1 shows the nodes distribution over the ME countries. The figure shows that the largest number of nodes in the network are from the United Arab Emirates (UAE) cementing its role as one of the main financial hubs in the region. On the other hand, countries with weak investment record and long history of conflict and instability such as Yemen have very low percentage of the nodes (0.04%).

The edges represent different relation types. Table I shows the list of relation types along with the number of instances of each type. The table shows that the Shareholder and Intermediary edges compose the majority of the edges appearing in this network.

Based on the nodes they are connecting, each edges falls into one of the following three categories:

1) Edges connecting two nodes from ME.
2) Edges connecting a node from ME to a node outside ME.
3) Edges connecting a node outside ME to a node from ME.

The distribution of the edges based on this categorization is shown in Table II. For the first category, address-related relations represent the majority. As for the second and third categories, the Shareholder, Director and Intermediary relations represent the majority.

The number of nodes from outside ME participating in the edges list are 23,114 nodes from 132 countries. Figure 2 shows the distribution of participating nodes from outside ME. We keep in the figure countries from which there are more than 200 nodes. The remaining countries are grouped under the name "others." The figure shows that 27% of the nodes participating in a connection with a ME node are omitted from the Panama network. The country with the largest number of known nodes is the British Virgin Islands with 22% of the whole nodes.

B. Network Visualization and Analysis

We use the Gephi tool [28] to perform the network visualization and analysis. Gephi is open-source software that is widely used for graph visualization and network analysis [29]. It provides a set of tools to handle the complexity of graph mathematics, allowing its users to focus on the meaning of network connections using a set of alternative visualization layouts that display the network connections. The visualization is done through the use of the ForceAtlase2 algorithm [30] implemented in Gephi. The ForceAtlas2 algorithm is used to

TABLE II
EDGES DISTRIBUTION BASED ON CATEGORY

Category	Type	Number of instances
From ME to ME	Director of	341
	Intermediary of	17
	Registered address	14,347
	Same address as	2
	Same name and registration date as	5
	Secretary of	146
	Shareholder of	175
	Similar name and address as	2,433
From ME	Authorised Person / Signatory of	4
	Bank Signatory of	1
	Beneficial Owner of	17
	Beneficiary of	714
	Director of	3,403
	Intermediary of	2,919
	Investment Advisor of	1
	Nominated Person of	1
	Officer of	1
	Owner of	6
	Power of Attorney of	9
	President of	2
	Protector of	3
	Register of Shareholder of	1
	Registered address	452
	Same name and registration date as	19
	Secretary of	151
	Shareholder of	9,558
	Similar name and address as	19
	Trust Settlor of	10
	Trustee of Trust of	72
	Vice President of	1
To ME	Auditor of	1
	Beneficial Owner of	6
	Beneficiary of	1,180
	Connected of	3
	Director of	3,859
	Intermediary of	17,212
	Registered address	1091
	Same address as	49
	Same name and registration date as	27
	Secretary of	569
	Shareholder of	20,619
	Similar name and address as	98

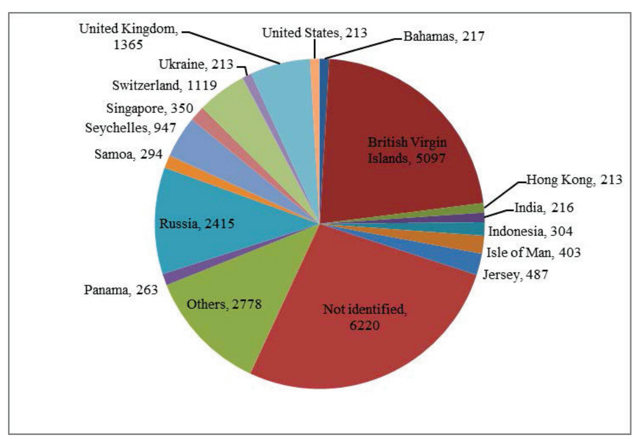

Fig. 2. Distribution of the nodes outside ME that have relations with ME nodes of our network.

layout nodes by placing nodes according to their dependency on the other nodes based on the connections between nodes.

For the Network Analysis, we apply following network analysis measures.

- *Average degree*, which computes the average number of edges connecting the nodes on the network.
- *Average path length*, which computes the number of steps required to navigate from one node to another considering the shortest path between nodes.
- *Network diameter*, which computes the shortest distance between the most distance nodes of the network.
- *Network modularity*, which measures the strength of dividing the network into a set of groups. It measures the ability to detect communities in the network.
- *Connected component*, which computes the number of component in the network whose nodes are connected. Connected component means that node inside the component are connected to each other and you can navigate from a node to any other nodes.
- *Average clustering coefficient*, which measures how neighboring nodes are connected to each other.

These network analysis measures can be used to study the power and influence of a node on the network using node centrality measures. Centrality measurement in social media is a way of measuring the power and influence of individuals in the network based on the connections in which they are participating. Measuring the power or influence aims to identify the importance of a node over other in the network [27]. Centrality can be measured using different metrics; the most common measurements are degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality.

Degree centrality signifies the celebrities in the network by considering the number of connections with other nodes in the network. The more connections the node has the more important it is. Node degrees can be either in-degree, which measures the number of connection pointing to a node, or out-degree, which measures the number of nodes that the current node pointing to.

Closeness centrality aims to find the gossipmongers in the network. Closeness measures the important of a node by its ability to send information to other nodes and receive information from other nodes according to the distance from others. The more the nodes are close to each other the more the node can share information with others. The value of the closeness measure is within the range [0, 1]; a higher value indicates greater closeness [29].

Betweenness centrality aims to find the bridges or bottlenecks in the network. Betweenness measures how a node acting as a bottleneck node in the network [29]. When the node is in a bottleneck state, then lots of the nodes in the network have to pass through the node in order to communicate with others. From this point, the Betweenness could be used to measure the power of a node in the network.

In some networks, none of the above measures can detect the importance of a node. One such case is known as the Gray Cardinal where the decision maker operates secretly or unofficially through persons who surround him/her and pass his/her decisions to others. Such cases can be detected using another measure called eigenvector centrality. In *eigenvector centrality*, the node's centrality is affected by how the node is connected to other central nodes.

In this work, we compute the set of centrality measures (degree, closeness, betweenness, and eigenvector) for each of the networks under study.

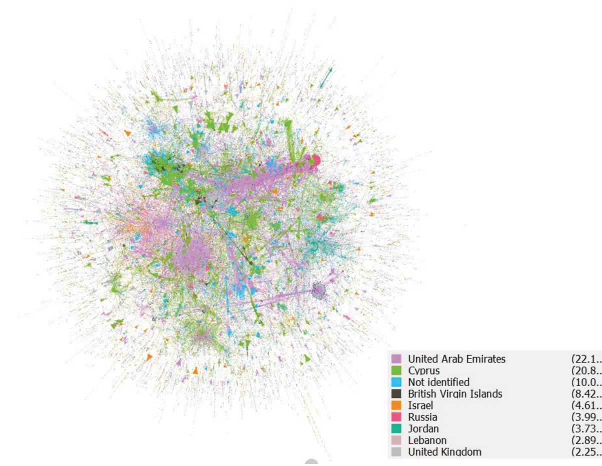

Fig. 3. Visualization of the ME network.

C. Analyzing the ME Network

Figure 3 shows the ME network visualization. The ME network consists of a set of edges connecting nodes from ME with nodes from outside ME. The visualization shows that the network has 60,556 nodes and 77,659 edges. The nodes in the network colored based on the country. The figure shows that the largest ratio of a node are for UAE and Cyprus from ME countries and "Not Identified" and British Virgin Islands from non-ME countries.

Applying the analysis measures to the network at hand shows that it has a modularity value of 0.959 along with 3,165 as a number of communities with average clustering coefficient of 0.90%. The network consists of 3,077 connected components with an average degree of 2.57 and network diameter of 5.

Table III presents the top 20 nodes in each centrality measure: degree, betweenness, and eigenvector. The closeness centrality values are not presented in the table since we have a large number of nodes with a closeness value of 1, which represents the maximum value for this measure.

The top twenty 20 centrality nodes are from 8 countries, three of which are from ME: UAE, Cyprus, and Turkey. The remaining are either unidentified or from outside ME: UK, Hungary, Switzerland, and Saint Kitts and Nevis. The top-twenty betweenness centrality nodes are from three countries (Bahrain, UAE, and Cyprus) all of which from ME. The same thing can be said about the three countries from which the

TABLE III
TOP 20 NODES IN EACH CENTRALITY MEASURE

Degree centrality			Betweenness centrality			Eigenvector centrality		
ID	Country	Value	ID	Country	Value	ID	Country	Value
11009351	UAE	3,168	75906	Bahrain	193	263908	Cyprus	1
75595	UAE	2,682	20151	UAE	185.9	270987	Cyprus	0.255
298170	Cyprus	2,097	298252	Cyprus	177	272170	UAE	0.202
11010502	UK	1,479	12215344	Cyprus	151	259943	Cyprus	0.179
11012146	Cyprus	1,315	22049	UAE	127.4	285434	Cyprus	0.158
11014056	Cyprus	1,140	12152234	Cyprus	101	269424	Cyprus	0.147
297669	Cyprus	955	12216202	Cyprus	95.3	14026869	Cyprus	0.144
263908	Cyprus	954	12216562	Cyprus	94.2	280367	Cyprus	0.139
11000382	UAE	892	12216243	Cyprus	82.1	260017	Cyprus	0.13
11011192	Hungary	603	12216654	Cyprus	79.6	264114	Cyprus	0.128
11004124	Cyprus	406	12220770	Cyprus	75.5	268983	Cyprus	0.116
11012191	Not Identified	370	12216681	Cyprus	75.5	260970	UAE	0.113
11009152	Switzerland	299	12220783	Cyprus	68.8	12220783	Cyprus	0.090
12162249	Saint Kitts and Nevis	261	12220769	Cyprus	74.3	10060693	Jordan	0.100
54613	Turkey	249	12111126	Cyprus	72.4	271917	Cyprus	0.096
22184	UAE	245	12220768	Cyprus	71.3	263912	Cyprus	0.091
298161	Cyprus	235	12214820	Cyprus	68.8	12220769	Cyprus	0.090
270987	Cyprus	234	12214823	Cyprus	66.8	14046928	Cyprus	0.086
11005942	Cyprus	234	12216655	Cyprus	66	12220770	Cyprus	0.085
11006109	UAE	228	12112605	Cyprus	65.8	14060675	Cyprus	0.084

top-twenty eigenvector centrality nodes are (Cyprus, UAE, and Jordan).

Each of the considered centrality measures presents a different picture about the importance of nodes on the network. In order to get a better idea on the role and importance of these nodes, it is useful to consider combining more than one measure. For example, according to [27], nodes with low betweenness and high degree or high closeness are not special in the sense that they do not represent crucial bridges connecting different parts of the network. This is the case for all of the top 20 nodes in Table III. However, if you explore a little further down the list of top nodes in terms of degree and closeness centrality, we would find nodes 75906, 20151 and 298252 (from UAE, Bahrain and Cyprus, respectively), which have high degree, closeness and betweenness. Usually, nodes with high degree and low closeness and betweenness values are worth investigating. However, in our network, we notice that such nodes are in fact sinks (they have 0 outgoing edges).

The analysis reveals many interesting cases such as Node 12152234 from Cyprus. This node has high betweenness (101) and low degree (4 in-degree and 4 out-degree) and closeness (0.42), which means that its few ties are crucial for the network flow and it monopolizes the ties from a small number of nodes to many others [27]. Finally, while there are many nodes with high eigenvector centrality values, perhaps, the most interesting one is node 12220783 which has high closeness and betweenness and low degree, which makes is “boundary spanner,” which is essentially standing between two dense and popular clusters, but not being a full-time member of any of them [27].

IV. CONCLUSION

In this work, we explored the information contained in the newly released documents known as the Panama Papers using social network analytics. We limited our attention to the Middle East (ME) and constructed a network consisting of nodes from the ME and the edges between them. The analysis revealed that while the constructed network enjoys some typical characteristics, there are many interesting observations and properties worth discussing.

In the future, we plan on performing a more careful cleaning up of the data. We also plan on performing deeper analysis on the communities involved in the network. We also plan on expanding our study to cover more regions.

ACKNOWLEDGMENT

The authors would like to thank the Deanship of Research at the Jordan University of Science and Technology for supporting this work (Grant #20160081).

REFERENCES

[1] M. Fenwick and E. P. Vermeulen, “Disclosure of beneficial ownership after the panama papers,” *Lex Research Topics in Corporate Law & Economics Working Paper*, no. 2016-3, 2016.

[2] J. O’Donovan, H. F. Wagner, and S. Zeume, “The value of offshore secrets–evidence from the panama papers,” *Available at SSRN 2771095*, 2016.

[3] L. J. Trautman, “Following the money: Lessons from the panama papers part 1: Tip of the iceberg,” *Available at SSRN*, 2016.

[4] R. K. Weisbord, “A catharsis for us trust law: American reflections on the panama papers,” *COLUMBIA LAW REVIEW ONLINE*, vol. 116, pp. 93–107, 2016.

[5] R. W. McGee, “The panama papers: A discussion of some ethical issues,” *Available at SSRN 2785954*, 2016.

[6] A. Jallow, “Panama papers: A celebral global madness and what options to mitigate this scourge in the realm for prudent financial accountability through taxation,” *Available at SSRN*, 2016.

[7] M. Vermeiren and W. Lips, “The panama papers and the international battle against tax havens: Lessons for the eu,” 2016.

[8] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[9] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, “Economic networks: The new challenges,” *science*, vol. 325, no. 5939, pp. 422–425, 2009.

[10] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *science*, vol. 323, no. 5916, pp. 892–895, 2009.

[11] E. M. Heemskerk, K. Young, F. Takes, B. Cronin, J. Garcia-Bernardo, V. Popov, W. Winecoff, L. F. Henriksen, and A. Laurin-Lamothe, "Big corporate network data: Problems, diagnostics, and fixes," 2016.

[12] W. K. Carroll and M. Fennema, "Is there a transnational business community?" *International sociology*, vol. 17, no. 3, pp. 393–419, 2002.

[13] G. F. Davis, M. Yoo, and W. E. Baker, "The small world of the american corporate elite, 1982-2001," *Strategic organization*, vol. 1, no. 3, pp. 301–326, 2003.

[14] W. K. Carroll and J. P. Sapinski, "The global corporate elite and the transnational policy-planning network, 1996-2006 a structural analysis," *international Sociology*, vol. 25, no. 4, pp. 501–538, 2010.

[15] B. Cronin, "Networks of corporate power revisited," *Procedia-Social and Behavioral Sciences*, vol. 10, pp. 43–51, 2011.

[16] J. Murray, "Evidence of a transnational capitalist class-for-itself: the determinants of pac activity among foreign firms in the global fortune 500, 2000–2006," *Global Networks*, vol. 14, no. 2, pp. 230–250, 2014.

[17] K. Van der Pijl, O. Holman, and O. Raviv, "The resurgence of german capital in europe: Eu integration and the restructuring of atlantic networks of interlocking directorates after 1991," *Review of International Political Economy*, vol. 18, no. 3, pp. 384–408, 2011.

[18] W. K. Carroll, M. Fennema, and E. M. Heemskerk, "Constituting corporate europe: a study of elite social organization," *Antipode*, vol. 42, no. 4, pp. 811–843, 2010.

[19] E. M. Heemskerk, F. Daolio, and M. Tomassini, "The community structure of the european network of interlocking directorates 2005–2010," *PloS one*, vol. 8, no. 7, p. e68581, 2013.

[20] S. Vitali, J. B. Glattfelder, and S. Battiston, "The network of global corporate control," *PloS one*, vol. 6, no. 10, p. e25995, 2011.

[21] E. M. Heemskerk and F. W. Takes, "The corporate elite community structure of global capitalism," *New Political Economy*, vol. 21, no. 1, pp. 90–118, 2016.

[22] K. Kanagaretnam, J. Lee, C. Y. Lim, and G. J. Lobo, "Cross-country evidence on the role of independent media in constraining corporate tax aggressiveness," *Journal of Business Ethics*, pp. 1–24, 2016.

[23] C. P. Rettig, "The panama papers and lessons learned from years of offshore leaks," 2016.

[24] K. E. Bichler, "Graphic design for the real world? how can the potential of visual communication for solving problems be used in an activist way?" *Artifact*, vol. 3, no. 4, pp. 11–1, 2016.

[25] S. Heymann, "Panama papers: How linkurious enables icij to investigate the massive mossack fonseca leaks. retrieved 25 may, 2016," 2016.

[26] W. De Nooy, A. Mrvar, and V. Batagelj, *Exploratory social network analysis with Pajek*. Cambridge University Press, 2011, vol. 27.

[27] M. Tsvetovat and A. Kouznetsov, *Social Network Analysis for Startups: Finding connections on the social web*. " O'Reilly Media, Inc.", 2011.

[28] M. Bastian, S. Heymann, M. Jacomy *et al.*, "Gephi: an open source software for exploring and manipulating networks." *ICWSM*, vol. 8, pp. 361–362, 2009.

[29] K. Cherven, *Network graph analysis and visualization with Gephi*. Packt Publishing Ltd, 2013.

[30] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, "Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software," *PloS one*, vol. 9, no. 6, p. e98679, 2014.