Les membres du jury qui ont évalué ce mémoire ont demandé des corrections mineures.

HEC MONTRÉAL

ESG Affinities of Influencers Within the Business Network

by
Andreea Firanescu

Thesis Supervisor

Gilles Caporossi

HEC Montréal

A Thesis
Presented to the Faculty of the Graduate School of HEC Montreal
In Partial Fulfillment of the Requirements for the Degree of
Master of Science (M. Sc.)
Department of Data Science and Business Analytics

August 2023

Résumé

Les tendances idéologiques actuelles dans le secteur des affaires présentent les influenceurs d'affaires comme susceptibles de modifier les attitudes ESG des entreprises and de leurs dirigeants. Comprendre les affinités réelles des influenceurs d'affaires avec l'ESG nous permet de déterminer s'il y a une base à ces spéculations. Ce projet de recherche cherche à déterminer si les entités qui sont des vecteurs d'influence au sein d'un réseau sont effectivement liées à des attitudes ESG positives. Il le fait en cartographiant le réseau d'affaires sur un graphe bipartite, puis en identifiant des groupes d'influenceurs and en déterminant leurs scores ESG. La recherche s'appuie sur la science des réseaux and l'analyse de réseaux sociaux and réalise des analyses basées sur des graphes à un and deux modes, des mesures de centralité, des graphes aléatoires, la détection de communautés and des corrélations avec les scores ESG. Elle découvre que les entreprises qui sont des influenceurs ont effectivement des scores ESG plus élevés que le réseau dans son ensemble. Elle constate également que l'analyse offre moins de biais lorsqu'elle est effectuée sur le graphe bipartite que sur les projections.

Mots-clés : *Science des réseaux; Graphe bipartite; Réseau à deux modes; ESG; Analyse de réseaux sociaux; Centralité; Réseaux d'affiliation; Conseils d'administration interconnectés.*

Abstract

Current ideological trends in the business sector discuss business influencers as susceptible to alter the ESG attitudes of companies and business leaders. Gaining insights into the real affinities of business influencers with ESG enables us to determine whether there is a basis to those speculations. This paper seeks to find whether entities that are vectors of influence within a network are indeed tied to ESG attitudes. It does so by mapping the business network onto a bipartite graph, then identifying groups of influencers and determining their ESG scores. The research draws on network science and social network analysis and conducts analysis based on one- and two- mode graphs, centrality measures, random graphs, community detection, and correlations with ESG scores. It finds that companies which are influencers indeed have higher ESG scores than the overall network. It also finds that the analysis offers less bias when performed on the bipartite graph than the projections.

Keywords: *Network Science; Bipartite Graph; Two-Mode Network; ESG; Social Network Analysis; Centrality; Affiliation Networks; Interlocking Directorates.*

Acknowledgments

As I was interested in leveraging data science tools for business and social studies, I decided to embark on the journey of completing a master's degree in a field that was not mine. The journey has been enriching, and I am proud of its culmination in a thesis which explores topics important to me – sustainability, social corporate responsibility, and social network analysis.

My first and most important "thank you" is directed to my supervisor, Professor Gilles Caporossi. Besides imparting his passion and incredible knowledge of complex systems, Professor Caporossi has helped me all the way throughout my journey in the Data Science and Business Analytics program and has demonstrated a strong desire for my success in my studies. Most importantly, he distinguished himself with his relentless kindness, which I am very grateful for. He has been a fundamental pillar in the completion of my thesis. I also wish to thank HEC Montréal as a whole for providing me graduate studies established in excellence and also for the quality of its services, as well as providing a well-rounded and supportive learning environment. Lastly, thank my friends and family who offered me endless encouragement when most needed.

As I begin my career in digital strategy consulting, I know that while I stray further away from data science, my studies at HEC Montréal will have provided me with a solid background for me to undertake any project related to leveraging data for digital transformation, and for that I am eternally thankful. I am certain I will see the effects of my degree and HEC Montréal during the entire length of my career.

Figures and Tables

Figure 1. Size of connected components. 38
Figure 2.1. Bipartite structure of the second largest connected component. 38
Figure 2.2. Second largest connected component under the Fruchterman-Reingold layout...................... 38
Figure 3.1. Distribution of ESG scores computed on the companies of graph G 43
Figure 3.2. Distribution of ESG scores computed on individuals and companies...................................... 43
Figure 4.1. Degree distribution on bipartite graph 43
Figure 4.2. Degree distribution on bipartite graph (y-log scale) 43
Figure 4.3. Degree distribution on bipartite graph (log-log scale)... 44
Figure 4.4. Degree distribution by top and bottom nodes on bipartite graph (log-log scale) 44
Figure 5.1. Degree distribution on the Companies projection (log-log scale) 45
Figure 5.2. Degree distribution on the Companies projection (log-log scale) 45
Figure 6. Degree centralities on the bipartite graph, the Companies projection and the Individuals projection. 45
Figure 6.1. Eigenvector centrality of bipartite graph 46
Figure 6.2. Eigenvector centrality of bipartite graph, on a y-log scale 46
Figure 6.3. Eigenvector centrality of bipartite graph, separated by nodes... 46
Figure 6.4. Eigenvector centrality of bipartite graph, separated by nodes, on a y-log scale...................... 46
Figure 7. Eigenvector centrality distributions on projections 47
Figure 8. Dispersion of ESG scores against eigenvector centrality values on the bipartite graph………..47
Figure 9. Distribution of ESG Scores of the influencers with the highest eigenvector centralities……….51
Figure 10. Cumulative ESG assortativity of influencers above the eigenvector centrality's distribution...53
knee, sorted by their eigenvector centrality 53
Figure 11. Cutoff at the knee of the eigenvector centrality distribution 55
Figure 12. Eigenvector centrality distribution of hubs 2 standard deviations above the mean of the degree distribution 56
Figure 13. Random bipartite graphs generated as bipartite graphs 57
Figure 14. Random bipartite graphs generated as unipartite graphs 57
Figure 15. Eigenvector centralities of influencers generated with the lowest and highest cutoff of the random bipartite graphs 58
Figure 16. Random Companies projection 59
Figure 17. Random Individuals projection 59
Figure 18. Community sizes of random bipartite graph…………………………………………...…. .60
Figure 19. Community sizes of Companies projection…………………………………………...…….61
Figure 20. Community sizes of Individuals projection……………………………….…………...……62
Figure 21. ESG distribution of hubs with the 12 highest eigenvector centralities…………………….…69
Figure 22. ESG distribution of hubs with the highest eigenvector centralities, with cutoff at the knee….69
Figure 23. ESG distribution of hubs with the highest degree, with cutoff 2 standard deviations above the mean……………………………….………...…………………………………….………...............69
Figure 24. ESG distribution of hubs with the highest degree, with cutoff at xmin………………………69
Figure 25. ESG distribution of hubs identified by random bipartite graph projections…………………..70
Figure 26. ESG distribution of hubs identified by random graph generated with equivalent density……70

Figure 27. ESG distribution of hubs identified by random graph generated with equivalent edge probability……………………………..…………...…………………………………….……......71
Figure 28: ESG distribution of dominant hubs within communities on the bipartite graph………………71
Figure 29: ESG distribution within the bipartite graph…………………………………………………71
Figure 30: ESG distribution of dominant hubs within communities on the Companies projection………72
Figure 31: ESG distribution within the Companies projection……………………………………………72
Table 1. Initial statistics on the largest connected component (graph G) and its projections 39
Table 2. All possible groups of hubs identified by their cutoffs. 48
Table 3. Assortativities of all possible groups of hubs. 49
Table 4. Eigenvector centralities of the nodes with the 10 highest values computed on the bipartite graph. 50
Table 5. Eigenvector centralities of the nodes with the 10 highest values computed on the Companies projection 52
Table 6. Eigenvector centralities of the nodes with the 10 highest values computed on the Individuals projection 52
Table 7. Eigenvector centrality of 12 potential influencers based on ESG assortativity 54
Table 8. Eigenvector centralities of influencers (companies) identified by hub dominance within the entire graph…61
Table 9. Eigenvector centralities of influencers identified by hub dominance on the Companies projection……….63
Table 10. Eigenvector centralities of influencers identified by hub dominance on the Individuals projection………65
Table 11. Results from Mann-Whitney U test for the companies within the bipartite graph………………………...66
Table 12. Results from Mann-Whitney U test for the individuals within the bipartite graph………………………..67
Table A. Sources of data for public data for S&P's "Professional Data" package (S&P Global)…………78

Table of contents

Résumé 2

Abstract..3

Acknowledgments 4

Figures and Tables..5

Chapter 1: Introduction and Objectives 9

1.1 Problem Description..9

1.2 Identifying influencers 9

1.3 Evaluating ESG affinities..11

1.4 Relevance of the study 11

Chapter 2: Dataset, Preparation, and NetworkX..13

2.1 The Data 13

2.1.1 Professional Data 13

2.1.2 ESG Data 14

2.2 Data Cleaning and Preparation..14

2.3 NetworkX Package...15

Chapter 3: Literature Review and Methodology 16

3.1 ESG and Management Related Network Literature 16

3.2 Methodological Approach..17

3.3 Network Related Concepts...19

3.3.1 The Data as a Graph...19

3.3.2 Centrality 24

3.3.3 Assortativity measures 28

3.3.4 Degree Distribution..30

3.3.5 Random Graphs 32

3.3.6 Additional discarded method 34

3.3.7 Community detection...34

3.3.8 Mann-Whitney U test...35

3.3.9 ESG Measures..37

Chapter 4: Results..38

4.1 Preliminary Analysis of the Network 38

4.1.1 Network Structure 39

4.1.2 Assortativity...41

4.1.3 ESG Distribution..42

4.1.4 Degree distribution......43
4.1.5 Eigenvector Centrality......46
4.2 Groups of Influencers......48
4.2.1 All potential groups of influencers......48
4.2.2 Top 10 Highest Eigenvector Centralities......50
4.2.3 Groups of potential influencers based on ESG assortativity......53
4.2.4 Groups of influencers based on eigenvector centrality......54
4.2.5 Groups of influencers based on the degree distribution......55
4.2.6 Groups of influencers based on random graphs' maximum degrees......56
4.3 Potential influencers based on hub dominance within communities......60
4.5 Mann-Whitney U Test......66
4.6 ESG Results......68
4.6.1 Spearman correlation......68
4.6.2 Visual Exploration......68
Chapter 5: Conclusion......73
Bibliography......75
Appendices......78

Chapter 1: Introduction and Objectives

1.1 Problem Description

One of the current Republican runners for the 2024 presidential election of the United States, Vivek Ramaswamy, has become a vocal critique of the impact of an “ESG current” and what he calls “ESG advocates” within the business world, fervently admonishing their influence on other companies to make decisions oriented towards ESG (Ramaswamy, 2023). ESG stands for a set of standards used to measure a firm’s impact on “Environmental, Social, and Governmental” topics. By 2020, 90% of all S&P 500 companies (arguably industry leaders) had published a sustainability report (Governance & Accountability Institute, 2020). While most of the business world has accepted the need for ESG standards, a debate has spurred pitting together “pro-ESG” and “anti-ESG” factions, both assuming that business influencers are indeed ESG proponents and use that influence to incentivize others into taking ESG-related actions (Smith Judd, 2023). This begs the question - do influential entities within the business world indeed have ESG affinities?

In order to determine whether ESG affinities *can* travel through the business network, this paper sets out with the main objective of *validating whether influencers who are vectors of attitudes within the business network tend to be associated to higher ESG scores.* To accomplish this objective, this paper will draw on different disciplines to carry out the analysis, namely: network science, social network analysis, organizational management, and sustainability.

The research question is separated in two sub-objectives which allow to answer the main objective, as explained in the following sub-sections.

1.2 Identifying influencers

First and foremost, the paper tasks itself with the sub-objective of identifying “meaningful groups of influencers”, that is, delimiting various potential groups of influencers, and establishing whether the influence of those groups is sufficiently different from that of the entire business network in order to consider those groups meaningful. This subobjective is the most challenging aspect of the study.

In order to identify influencers, this paper will make use of network science techniques. Network science allows to turn data into a network, effectively mapping the business network from a set of data in a way that makes it so structural patterns can be identified on said network, including the identification of important entities within the network, in a way that cannot be achieved through analysis of tabular data.

This paper will use the definition of "influencers" as individuals who would have the capacity to spread an ESG attitude within the business network, that is, "vectors of influence". This definition comes from social network analysis, aligned with Borgatti's definition of what an important entity is in a network. In his famous classification of "network flows" and description of centrality measures Borgatti (2005) categorizes several measures of importance of entities that assess "attitudes". Borgatti defines the category of attitudes as movements of "influence"; "Here, the notion is of an influence process in which, through interaction, individuals effect changes in each other's beliefs or attitudes. […] The attitudes spread through replication rather than transfer (I do not lose my attitude the moment I infect you with it)" (Borgatti, 2005 : 58).

At the cross between management studies and social or complex network analysis, influencers within a network are often taken to be board members (Conyon and Muldoon, 2004; Robins and Alexander, 2004; Vasques Filho and O'Neale, 2020). This research is aligned with the convention by using board members as influential entities within the network, however, this study will amplify the scope to include also key business executives. Indeed, there is increasing talk about executives deciding to be ESG leaders within their companies (Luthra *and al.*, 2022; Smith Judd, 2023). In this sense, both board members and top executives can choose to adopt an attitude aligned with ESG expectations. Both board members and top key executives will therefore be considered as "business leaders". The companies that those business leaders affiliate themselves with will also be considered potential business influencers, as oftentimes, it is the company as an entity which exercises said influence, and which of business leaders or companies are the main entities in a network is not set in the literature (Opsahl, 2011 : 2). With regards to ESG, it is specifically companies that are considered to relay ESG attitudes (Ramaswamy, 2023).

Therefore, influencers will include two categories of entities in this research: companies and individuals (also referred to as “business leaders”), the latter category being comprised of key executives and board members, which from henceforth will not be distinguished.

1.3 Evaluating ESG affinities

The second sub-objective is to establish whether influencers indeed have higher ESG scores than overall companies and individuals. The objective is not to identify the presence of influencers that actively try to “push” others towards ESG efforts (causation effect), rather to clarify whether or not companies and business leaders that are arguably network influencers have an affinity with ESG (correlation effect).

As such, the proxy for “ESG affinity” will be the ESG score of the company, or in case of the business leaders, that of the companies they are affiliated with. By using ESG scores as proxies for ESG attitudes, this paper makes the assumption that having a high ESG score is related to having a positive attitude towards ESG. Potential impacts of this assumption are the following:

- That some companies (or business leaders of said companies) with a high ESG score might actually have a negative attitude towards ESG and might not in reality contribute to the spread of ESG attitudes (which is reasonably improbable).
- That some companies (or business leaders of said companies) with a low ESG score might actually have a positive attitude towards ESG and might in reality contribute to the spread of ESG attitudes (which is reasonably probable).

1.4 Relevance of the study

The relevance of the study can be qualified as practical pertinence, for multiple reasons;

1) The combination of *i*. ESG attitudes *ii*. for business leaders *iii*. within a bipartite network has seldom been studied. This paper will leverage and explore data in a novel way by:
   - Combining two datasets: ESG data and company – business leader relationship data.
   - Using said data for a network science approach, specifically for a bipartite network.
2) This study will explore and validate less conventional methods for identifying influencers in bipartite graphs, that is, methods that require less computing power than conventional algorithms:
   - Identification through degree distribution of random graphs.
   - Hub dominance.
   - Assortativity.

Chapter 2: Dataset, Preparation, and NetworkX

2.1 The Data

The data used for the study is comprised of two datasets, one representing company – business leaders' relationship, and the other ESG data for companies. Both are provided by S&P Global Market Intelligence and accessed through Wharton Research Data Services. S&P Global was chosen as a data provider for being recognized as a leader in providing alternative data and for the substantial sizes of its datasets (Wright, 2019).

2.1.1 Professional Data

The first dataset was built from S&P's "Professional Data" package, part of S&P's "People Intelligence" product. The package aggregates publicly available data about 4.5 million professional across the world (S&P Global 2019 : 3). The dataset was constructed through a query with the following variables:

- "companyname": Company name
- "companyid": Company identifier (used across many S&P's packages)
- "personid": Individual identifier
- "boardflag": Binary variable that indicates if an individual is a board member on the company (1) or not (0)
- "keyexecflag": Binary variable that indicates if an individual is a board member on the company (1) or not (0)

The data provided in this package is publicly available data, as indicated in the packages data collection methodology (Appendix A). While the data is of public domain, the names of the individuals were not used, nor are their identifiers conveyed in this paper.

2.1.2 ESG Data

The second dataset was built from S&P's "S&P Global ESG Scores" package, in which S&P provides sustainability ratings to companies based on their SAM Corporate Sustainability Assessment (CSA), one of the oldest and most sophisticated ESG scoring methodologies on the market (S&P Global 2021).

The dataset was constructed through a query with the following variables:

- "companyid": Company identifier (used across many S&P's packages)
- "aspectname": Indicates whether the score is an "S&P Global ESG Score" or another type of disaggregated score
- "scorevalue": Value of the ESG score
- "csascoretypename": Indicates whether a score is "Raw" or "Modeled"
- "assessmentyear": Year on which the ESG score was assigned

2.2 Data Cleaning and Preparation

The professional data was filtered in the following ways:

1) Only individuals which were either board members or key executives were kept in the dataset.
2) Only one row was kept per company – individual relationship. Whilst an individual may occupy various positions in a company, the position help in a company is not studied by this paper, rather the existence of a connection, which is conveyed by a single row. If multiple rows are kept, this would enter a bias in the degree of an individual's node.

The professional data was filtered in the following ways:

1) Only the most recent year was kept (2022 or 2021).
2) Only the scores where the "aspectname" was "S&P Global ESG Score" were kept, as the paper is interested in the overall company score, not disaggregated scores.
3) Modeled scores were kept, and where unavailable, raw scores were used. Modeled scores represent scores where some data has been imputed by S&P, where raw data is unavailable,

so as not to introduce bias in their scores. Raw scores represent scores built exclusively on data reported by the company. Modeled scores were used first as their existence signifies that raw data was incomplete to compute a raw score.

The result is one row per company – business leader and company – ESG score. The two datasets were then joined into one con “companyid”.

2.3 NetworkX Package

The NetoworkX Python package was used to transform the dataset into a graph and perform multiple operations on it. NetowrkX is a package built in 2008 for the exploration and analysis of networks, which contains many traditional algorithms used in network analysis implemented within its functions (Hagberg, Swart and Chult, 2008). The package has become reputed and widely used one, such as for social network analysis (Akhtar, 2014), complex systems (Hadaj *and al.*, 2022), graph visualization, and many other uses (Hagberg *and al.*, 2008). This paper will make use of it for the ease of construction of the network and use of certain readily available algorithms. The functions used will be indicated in the methodology section.

Chapter 3: Literature Review and Methodology

3.1 ESG and Management Related Network Literature

As previously mentioned, few studies cross ESG attitudes and business leaders from a network perspective. The related literature is dominated by:

- Studies related to data science focus on the relationship between ESG and business leaders, yet not from a network perspective.
- Studies of the spread of attitudes amongst business leaders from a network perspective, but unrelated to ESG.

This subsection offers some insights into the latter as a basis for this research. The focus is due to the present study having networks as its main point of research, rather than ESG.

Multiple studies have analyzed the structure of a business network, specifically at the level of boards. This concept is referred to as "interlocking directorates" in social network analysis. Davis and Greve have studied the spread of attitudes in the network of board members and have found that good governance practices spread through interlocking directorates. They concluded that board members serve as vectors of strategic information between companies, creating a diffusion effect (Davis and Greve, 1997). Robins and Alexander studied the structure of Australian and American interlocking directorates as bipartite networks and found that "structures tend to be influenced by the clustering of directors on boards" (2004). This is presents supporting information to the current research as it will help validate findings about the structure of the graph with respect to individuals. The closest study to the current research examined the relationship between board network centrality and British firms' ESG performance and found a positive impact (Harjoto and Wang, 2020). This paper will use some of the same methods, specifically centrality methods, and compare results.

3.2 Methodological Approach

This subsection will outline the methodological approach to the analysis, step by step, and serves to indicate the structure of the presentation and discussion around the results. It is followed by a subsection explaining the methodology with more detail.

1) Creation of the main bipartite graph and its projections.
    a. Creation of the original graph and identification of main connected component.
    b. Initial analyses on the main connected component and Companies and Individuals projections which will serve as a basis for comparison with potential groups of influencers:
        i. Basic measures of size.
        ii. Insights into the structure of the graph and comparison of behaviour with the projections: degree distribution (also performed on top and bottom nodes), degree centrality, redundancy, transitivity, eigenvector centrality.
        iii. Insights into the ESG behaviour of the graph: ESG distribution and ESG assortativity.

2) Analysis of the "most certain influencers".
    a. Identification of the top 10 highest eigenvector centralities.
    b. Comparison of the behaviour of different measures with the results on the original graph and between the graph and its projections. This is done to identify expected behaviour for the groups of influencers.
        i. Eigenvector centralities.
        ii. Degree and ESG assortativity.

3) Identification of potential groups of influencers:

    This sub-section will first make use different measures of importance of nodes. Then, for each measure, it will apply one or different cutoffs, to enclose the group of important influencers. For each potential group of influencers, the same analyses will be performed.
    a. Groups based on the eigenvector centrality on the graph and the projections.

i. Computing the cutoff at the knee of the distribution.
ii. Computing the degree assortativity and the ESG assortativity.

b. Groups based on the degree distribution on the graph.
   i. Computing the cutoffs of the distribution
      1. Two std deviations above the mean
      2. x_{min}
   ii. Computing the degree assortativity, the ESG assortativity, and the eigenvector distribution
c. Groups based on the degrees higher than those of random graphs for graph and projections.
   i. Generation of random graphs based on the equivalent density and edge probability of the original graph
   ii. Cutoff at the maximum degree
d. Identification of most important nodes within communities.
   i. Generation of communities with the Louvain algorithm.
   ii. Identification of dominant nodes through "hub dominance"
   iii. Computing the eigenvector centralities and the assortativities.

4) Evaluation of the "meaningfulness" of the groups of influencers.
   a. Evaluation of group properties to confirm "meaningfulness" of groups.
      i. % of influencers
      ii. Comparison with the properties of the top 10 influencers and of the original graph.
   b. Mann-Whitney U tests between the eigenvector centrality distributions of the identified groups of influencers and the ones of the original graph.

5) Determining correlation between the groups of influencers and ESG scores.
   a. Computing the Spearman correlation between the eigenvector centrality distribution and the ESG scores distribution of the original graph.
   b. Visualization of ESG scores of the groups of influencers against those of all companies.

3.3 Network Related Concepts

This section of the paper will refer to literature related to network science and social network analysis. Specifically, this section will:

- Explain important concepts to the research,
- Outline their implications to this research and expected behaviour,
- Relate to previous research.

3.3.1 The Data as a Graph

"Graphs" are mathematical representations of relations between entities (represented as a set of vertices *V*), whereby said relations are portrayed as a set of edges *E* connecting vertices, such as graph *G = (V, E)*. Graphs are also called "networks", where vertices are "nodes" and edges are "links" (Caporossi and Camby, 2021 : 1). "Graphs" are referred to in graph theory and "networks" in network science and specifically social network analysis yet refer to the same objects. Since this study is interdisciplinary, it will make use of these terms interchangeably, as is commonly accepted (Barabási and Márton, 2016 : 45). When operations are computed on a graph, they are truly computed on its "adjacency matrix" A_{ij}, a square matrix where entities (vertices) are displayed in the rows *i* and columns *j*, and the existence of a relationship (or edge) is represented by a 1 and its inexistence with a 0 (Barabási *and al.*, 2016 : 52). The "degree" k_i of a node represents its number o links to other nodes, such as $k_i = \sum_{j=1}^{N} A_{ij}$. Relationships between nodes are often "undirected", or symmetric, implying that there is no starting point or ending point to the connection (for example, two people knowing each other), resulting in a symmetric matrix such as $A_{ij} = A_{ji}$. The current graph *G* will assign individuals (business leaders) and companies to nodes *V* and the affiliation of an individual to a company and vice-versa as the relationship represented by the edges *E*, resulting in an undirected graph. The presence of two distinct entities (individuals and companies) as nodes in the graph will be discussed in the next sub-section.

The network undergoes the removal of all nodes with connections to a single other node. This is done at the level of the dataset behind the construction of the graph, for which all rows with a nonduplicated company or individual are removed. This process is iterated through until removal of all nodes with a degree of 1. This is done for two reasons:

1) The sheer size of the dataset would make the network too large to work with if it included all individuals holding only one position on a company. Since it is not recommended to study only a portion of a network, this approach was the best in reducing the size.
2) As the research is interested in identifying influencers within a network of attitudes, nodes with a degree of 1 can neither represent an influencer, nor participate actively in the spread of said attitudes.

Many networks are disconnected, that is, are composed of various components (clusters of nodes) that have no edges between them. This characteristic may introduce bias in the results of some operations. Social networks are generally composed of many components of evenly distributed sizes or of a "giant component" (largest connected component) flanked by many small components (Newman, Watts and Strogatz, 2002 : 2568). The presence of a giant component in a network greatly influences the spread of information in social networks, allowing broad communication between large groups. Conversely, without a giant component, communication is limited to smaller groups, restricting information flow to these clusters (Newman *and al.*, 2002 : 2569).

Early results (Chapter 4) showed that the network under consideration displays indeed a giant component. This giant component will be considered as the main graph on which all operations will be performed. This is a common practice in network science when the smallest components are of no research interest, as was demonstrated by Albert and al.'s famous study of the internet (1999). The results will delve deeper into the smallest connected components to demonstrate their little relevance to the rest of the business network.

<u>Converting the bipartite graph into projections</u>

The graph that is studied in the current paper is a "bipartite" graph. As bipartite graphs are a subset of graphs with properties specific to them, this section will offer more detail on the nature and behaviour of these networks.

Bipartite graphs (or "bigraphs"), as opposed to the most common "unipartite" graphs, are networks with two disjoint sets of nodes *U* and *V*, whereby the connections (or edges *E*) lie between those two sets. That is to say, no edges exist between nodes of the same set, but a relationship will be implied between two nodes u_1 and u_2 from one set that are connected to the same node v_1 from the other set, thereby creating the network of connections and graph *G=(U,V,E)* (Barabási *and al.*, 2016).

Some social networks represent "affiliations" under such relationships. Borgatti and Halgin indicate that "[in] social network analysis, the term "affiliations" usually refers to membership or participation data […]" between an individual and the subject of its participation (2011 : 417), such as exists in the network at hand where business leaders are "affiliated" to a company through their participation on its board of directors or as a top executive. This specific business leader – company affiliation has been previously studied in the form of bipartite graphs (Borgatti *and al.*, 2011; Davis *and al.*, 1997; Robins *and al.*, 2004). In social network analysis, these bipartite graphs are referred to as "2-mode graphs" (or "two-mode graphs") to emphasize the existence of two distinct types of entities *U, V* in the rectangular adjacency matrix *A*, represented as the columns and rows.

While there are no direct links between the vertices of one set as stated, affiliation makes it so that social ties can be inferred between nodes of one set. In the current graph, if two business leaders work at the same company, they will have an inferred connection, as will the companies with business leaders in common. These inferred ties are called "co-affiliations" and can be represented on a "projection" of one set of nodes into a graph of its own (*U*-projection), where the edges represent a shared neighbour node in the original graph. Co-affiliation can highlight "[…] an observable manifestation of a social relation that is perhaps unobservable directly" (Borgatti *and al.*, 2011 : 422).

There is no consensus among scholars whether methods for analysis should be extended to the entire bipartite graph (the "direct" approach), or rather use standard techniques on the projections (the "conversion" approach) (Borgatti *and al.*, 2011; Everett and Borgatti, 2013; Latapy, Magnien and Vecchio, 2008). This study will do both, converting the graph into the "Individuals" projection and the “Companies” projection using the *nx.bipartite.projected_graph()* function. The triple graph analysis is conducted for the following reasons:

Conceptual importance of both sets

The current research is interested identifying influencers as potential spreaders of attitudes within networks flows. For this network flow, the relationship company – business leaders is important, and should be evaluated in the bipartite graph. Additionally, this research is interested in in identifying the most important groups of influencers, be they composed of companies or individuals. It is therefore important to conduct influencer detection analyses on the bipartite graph to identify if companies are more important than individuals, or vice-versa. Some techniques might perform better however in identifying influencers within their own set (see below), so they will be computed on projections for comparison (Nacher and Akutsu, 2011 : 4637).

Technique performance

Most graph analysis algorithms are made to be computed on unipartite networks with a square adjacency matrix. NetworkX does store the bipartite adjacency matrix as square one where rows and columns represent all nodes and are populated with 0 and 1, but this in and of itself can lead to issues such as inefficiency of matrix operations, time convergence, misrepresentation of connectivity between sets, spectral analysis, and others. It is expected that in these cases, analysis on projections perform better (Newman, 2010; Opsahl, 2011). On the other hand, projections can be overly densely connected due to induced edges in a projection, which can also lead to inefficiency of matrix operations (Borgatti *and al.*, 2011; Latapy *and al.*, 2008). Indeed, when converting links to the projection, each node of degree d in the bipartite graph induces ($d(d$-1)/2) links in the projection, inflating the number of links (Latapy *and al.*, 2008 : 4). Another phenomenon, a high level of transitivity (global clustering coefficient), will also be present in a projection. This phenomenon is based on the presence of “six-cycles” which results in triangles formed in the projection than would exist in a real unipartite graph, increasing the density of the projection (Vasques Filho *and al.*, 2020 : 3).

Loss and distortion of information

When projecting one set of entities on a graph, the distinction is lost between the two sets, as are those relationships, leading to the loss of important structural data (Borgatti *and al.*, 2011 : 204). One effect is that of redundancy, similar to transitivity, which measures the extent to which a node' neighbours are connected. In a bipartite graph, neighbours are only connected to others from the other set. In a projection, nodes become directly connected, leading to an overlap in their neighbourhoods, and therefore a higher redundancy. Specifically, "four-cycles, that is, pairs of bottom nodes that share more than one (top node) common neighbor […] generate redundant links when creating a simple graph projection" (Vasques Filho *and al.*, 2020 : 3). Redundancy, transitivity, and induced links can therefore heavily skew results in terms of centrality, downplaying the importance of certain connections, and inflating of certain connections, directly affecting centrality, assortativity, and community detection (Latapy *and al.*, 2008 : 13).

Use of the graph and its projections

There will therefore be advantages and disadvantages to the application of most techniques to either the bipartite graph or its projections. However, this paper will follow the majority approach and consider the bipartite graph as the main object of analysis (Borgatti *and al.*, 2011; Latapy *and al.*, 2008). This approach has been adopted in a similar context by Robins and al. (2004). In their study, they used the entire two-mode graph, justified by their interest in the global structure of the interlocking directorship of boards within Australia and the USA, similarly to this paper. This technique will also be justified by the evaluation of redundancy and transitivity in the projections, to ascertain if edges have indeed been induced. In the present research, when results on the bipartite graph for a measure will indicate any bias due to the nature of the two-mode network, they will be discarded or evaluated on the projections.

Therefore, all techniques (where pertinent and feasible in terms of complexity) will be applied to the bipartite graph, to the "Individuals projection", and to the "Companies projection". This will allow comparison and interpretation of results. Results within the section of the identification of potential groups of hubs will determine whether the groups of hubs constituted from projections and the bipartite graph will be used. Statistical analysis techniques to identify meaningfulness of the groups and correlation with ESG scores will be applied to "top nodes" and "bottom nodes", that is, nodes that represent individuals and companies within the structure of the bipartite graph and the groups of hubs. Where relevant, each method will be detailed in terms of expected behaviour on the bipartite graph and unipartite projections.

3.3.2 Centrality

One of the most fundamental concepts in the study of social networks are centrality measures (Borgatti, 2005; Freeman, 1978). They are quantitative node attributes which try to answer the infamous question: "Given a social network, which of its nodes are more central?" (Boldi and Vigna, 2013 : 1). Various centrality measures have been developed over the years to capitalize on different network properties in order to answer that question. Consequently, while all are indicative of nodes' importance, they stress different "network flows" (movements or exchange that happens across the links of a network), and therefore point to different definitions "important nodes" (Borgatti, 2005). It is therefore important to choose the appropriate centrality measures in a way that reflects the desired interpretation of "important nodes". The centrality measures that are applied in this study must highlight important nodes that can be defined as influencers that participate in the spread of attitudes.

Borgatti classifies four measures as the most eminent in capturing those attitude flows: eigenvector centrality, betweenness centrality, closeness centrality, and degree centrality. Following are presented the retained centrality measures and the discarded ones.

Eigenvector centrality

As eigenvector centrality will be the most important measure of centrality in the graph, this section will explain its uses with more detail.

Eigenvector centrality has been determined as a particularly strong and pertinent centrality measure in the present case as it is highly effective in identifying influencers. The eigenvector centrality C^{λ} of a node is computed by finding an eigenvector of the adjacency matrix of the graph that corresponds to its largest eigenvalue (Bonacich, 1987, 2007; Caporossi *and al.*, 2021 : 42):

$$C_i^{\lambda} = \frac{1}{\lambda_1} \sum_{(i,j)\in E} C_j^{\lambda} = \frac{1}{\lambda_1} \sum_{j=1}^{n} a_{ij} C_j^{\lambda}$$

It returns a score of influence because it measures the traffic flows via walks of a specific length from node *i* to node *j* (Borgatti, 2005 : 62), and this, for all neighbours. It is a recursive algorithm that reflects the node's immediate connections and their significance of within the broader network, effectively making it both a local and a global descriptor of nodes. Put simply, a node's eigenvector centrality will be determined by its neighbours' centrality, and their neighbours' centrality, and the following, with decreasing weight in the calculation.

It must be noted that it is the distribution of the eigenvector centrality that will be analyzed due to the size of the network (as will be explained below). Its distribution will be computed on the bipartite graph, on the Companies projection, and on the Individuals projection and compared. Network science do not have a definitive response on whether eigenvector centrality is better computed on the graph or its projections (Everett *and al.*, 2013; Yang *and al.*, 2022). The expected differences are detailed as follows:

Eigenvector centrality distribution on the bipartite graph

On the bipartite graph, the eigenvector centrality of a company is based first on centrality of the individuals it pertains to, and those to that of the companies they belong to, etc., creating a feedback loop of importance. Computing the eigenvector centrality on the bipartite graph is particularly pertinent for this study because it allows for direct comparison. While it is true that companies and business leaders are different entities, for the purpose of this study, they are considered to play the same role in terms of spread of attitudes within the network, making comparison of their centrality a just one. A danger of computing the eigenvector centrality on the bipartite graph is that the eigenvector centrality computed on the adjacency matrix considers every pair as a possible pair, while that is not so for a bipartite graph where pairs are only possible between the two sets of nodes, leading to potential biases (Yang *and al.*, 2022 : 6). To identify potential misleading results, the highest eigenvector centralities on the bipartite graph will be compared to those of the projections.

Eigenvector centrality distribution on the projections

Computing the eigenvector centrality on the projections allows to compare the centralities of the business leaders and individuals between themselves and to confirm the centralities of the bipartite graph. Redundancy however may overplay the influence of certain nodes and downplay that of others. Redundancy can make for dense overlapping parts of the graph, making it so influence does not circulate as well in the overall network, and therefore eigenvector centralities being less sensitive to the network structure (Latapy *and al.*, 2008 : 13). Results will be compared to those of the bipartite graph to evaluate these differences in rank.

Eigenvector centralities as influencers

Two potential groups of influencers will be returned based on eigenvector centrality:

1) The 10 nodes with the highest centralities will be returned for the bipartite graph. These will be considered the "most certain influencers" in the graph, due to this measure being considered the most robust for evaluating centrality. They will be compared to the top 10 values of the projections to ensure there are no misleading results.
2) The "knee" of the distribution will be computed as a cutoff. To detect the "knee" of the degree distribution, the Kneedle algorithm from the Python package "kneed" was computed on the distribution. The algorithm identifies the point of maximum curvature as the "knee" in a distribution. (Satopaa and al., 2011)

Use as a baseline

The eigenvector centrality will be used as a baseline for comparison with other measures. The comparison will be done with the following centralities and in the following fashion:

1) The original graph eigenvector centrality distribution. As other measures will be computed and influencer groups found with their distinct cutoffs, the eigenvector centrality distributions will be computed and compared to that of the original graph and projections, indicating whether the measure or the cutoff show enough meaningful difference with the eigenvector distribution of the initial graph. This will therefore serve as a way to evaluate the robustness of the various methods of identifying influencers. This is due to this measure being considered the most robust for evaluating centrality (Borgatti, 2005).
2) The 10 nodes with the highest centralities will be returned for the bipartite graph. These will be considered the “most certain influencers” compared to the initial graph and the comparison will serve as a to garner insights into expected behaviour of other groups of influencers, which should situate between those “most certain influencers” and the behaviour of the original graph and projections. As other measures will be computed, they will be applied to these 10 centralities.
3) Statistical tests will be performed using eigenvector centralities to identify meaningfulness of the groups (presented at the end of the chapter). Since the tests performed will be based on the median and on ranking the values of two groups and comparing them, the tests will be done separately on companies and individuals. This will be done due to different variances in the eigenvector centralities for the companies and individuals. Additionally, since some groups of hubs will contain more or fewer companies and individuals, results will be more easily interpreted if separated.

Discarded methods

Some centrality measures which are commonly used to identify influencers have been discarded due to their complexity. Indeed, while the complexity of the eigenvector centrality of has been demonstrated to be the centrality of choice for large networks due to its low computational complexity of $O(m+n)$, other measures identified by Borgatti (2005) are too computationally expensive to be computed on the current network (Lohmann *and al.*, 2010; Wandelt, Shi and Sun, 2020 : 68116). Relevant yet discarded methods are listed below with justification for discarding:

- *Betweenness centrality* identifies important nodes through the number of shortest paths passing through them. Those nodes act as bridges within a network, facilitating transfer of information or influence (Borgatti, 2005; Freeman, 1978). Betweenness centrality could therefore capture another insightful aspect of a node's influence. However, as it iterates through shortest paths, it is a computationally expensive method, of complexity $O(nm)$ (Lohmann *and al.*, 2010 : 2)
- *Closeness centrality* also has a computational complexity of $O(nm)$ as it is based on the average length of the shortest paths between nodes (Wandelt *and al.*, 2020 : 68116). Here, most central nodes have shorter distances to others and can therefore receive and impart information sooner, effectively making them designated influencers for imparting attitudes (Borgatti, 2005 : 59).

3.3.3 Assortativity measures

The degree assortativity coefficient is used to measure the degree correlation, that is, the Pearson correlation between the degrees of paired nodes in a graph, from -1 to 1. Positive degree assortativity means that nodes with a high degree tend to connect to other nodes with a high degree, while negative degree assortativity means that nodes with a high degree tend to connect to nodes with a low degree (Barabási *and al.*, 2016 : 236). In the context of this bipartite graph, a positive assortativity would indicate that companies tend to be connected to individuals who themselves are to be affiliated to many companies, and that highly connected individuals would lead companies which would have many board members or top executives. A negative assortativity would indicate that companies tend to be connected to individuals who themselves are to be affiliated to few companies, and that highly connected individuals would lead companies with few board members or top executives.

Most social networks tend to be assortative – that influencers tend to link to nodes who are also influencers (Barabási *and al.*, 2016 : 234; Newman, 2003). However, the current graph might display different behaviour since the graph at hand is 1) a bipartite network, 2) focusing on attitudes as network flows. For exmaple, previous research on the Norwegian network of board of directors has returned lower degree assortativity than that of other social networks, due to transitivity (Vasques Filho *and al.*, 2020 : 5).

Numeric assortativity is similar to degree assortativity but instead of measuring the degree correlation, measures the correlation of a specific node attribute. In the case of the current graph, the node attribute is the ESG score of a company. High assortativity (close to 1) would indicate companies tend to connect with other companies of similar scores, and high disassortativity (close to -1) would indicate they tend to connect with companies of different scores. It is important to note that the ESG assortativity is not indicative of any preference of association by any company as there can be another *z* variable at play causing the correlation in ESG scores.

Both degree and numeric assortativity are computed using existing functions in the NetworkX package, nx.degree_assortativity_coefficient and nx.numeric_assortativity_coefficient respectively, both based on Mark Newman's work (2003). The ESG scores of companies and individuals are then tied to the nodes as nodes attributes. They are computed on the three graphs: the bipartite graph, the Companies projection, and the Individuals projection. Since the bipartite graph and the Individuals projection contain edges respectively companies-individuals and individuals-individuals, the ESG score assigned to individuals is the average ESG score of all the companies associated to said individual. However, the assortativities in the bipartite graph might be skewed towards higher results for individuals working at few companies (lower degree), whereby the ESG score of the company for which the assortativity is computed holds more weight. Therefore, ESG assortativity for the bipartite graph will be used rather as a basis for comparison rather than in and of itself, as opposed to the ones from the projections.

Overall, the value for assortativity is double:

- In and of itself, it offers insights into the graph and projections behaviour and can confirm or infirm expected behaviour.

- It serves also as a basis for comparison between assortativities of the graph and projections to the assortativities of the identified groups of influencers.

3.3.4 Degree Distribution

The degree distribution provides the probability that a node in graph G has degree k, normalized by G's number of nodes N, resulting in $p_k = \frac{N_k}{N}$ (Barabási *and al.*, 2016 : 49). Its shape provides important insights into the nature of the network. Specifically, in their seminal article on "real networks", Barabasi and al. (1999) found that large real networks are "scale free", that is, that their degree sequence follows a power law distribution. Accordingly, many nodes have a small number of connections, while few nodes have a very high number of connections, which Barabasi and al. consider are "hubs" (1999), or influencers (terms to be used interchangeably throughout the paper). This is particularly true of social networks, where some individuals are more connected than others (Caporossi *and al.*, 2021 : 15).

Bipartite networks, by nature, display different degree distributions to those of unipartite graphs, whereby their distribution are highly dependent on two different distributions – that of the top nodes and that of the bottom nodes (Vasques Filho and O'Neale, 2018). Top degrees typically follow power laws, while bottom degrees can either follow a power law or a Poisson distribution (Guillaume and Latapy, 2004 : 218).

The degree distribution will be computed for three main uses:

1) The evaluation of the distribution on the network at hand permits to see whether the network's connections are behaving as expected.
2) It will also serve to identify potential groups of hubs, above the cutoffs of the knee of the distribution and the x_{min} (see below).
3) It will be compared to that of randomly generated graphs, which will point to specific degrees as potential cutoffs for identifying potential groups of hubs (to be explained in the following subsection).

In line with common practices, the degree distribution will be computed analyzed with the following considerations:

1) The degree distribution will be computed on the bipartite graph and then separated between top and bottom nodes to visualize the two degree sequences, on which the analysis will be conducted. While centrality measures such as the eigenvector centrality cannot be computed separately on top or bottom nodes unless projections are created, the degree distribution allows for this separation of degree sequences (Guillaume *and al.*, 2004).
2) Python's "powerlaw" package is used to identify whether the two distributions follow a power law one (Alstott, Bullmore and Plenz, 2014). This is particularly useful as all vertices in the network with a degree of 1 have been eliminated. The algorithm returns α, the power law coefficient, which evaluates the fit of a distribution to a power law distribution by quantifying how quickly the tail of a distribution gradually decreases. In the case of a degree distribution, it indicates how rapidly the frequency of a nodes' degrees decline as their value increases. Considering that the distributions hold a bias due to the removal of degrees of 1, a goodness of fit test will be conducted against an exponential distribution , which can occasionally approximate degree sequence distributions (Vasques Filho *and al.*, 2018). For this reason, the Poisson distribution will also be tested for the nodes pertaining to individuals (which upon visual inspection and testing against a power law distribution returns less certain degrees than the companies' distribution), through a Poisson probability mass function, using the average degree as the lambda parameter.
3) The x_{min}, that is, the "minimal value […] at which the scaling relationship of the power law begins" (Alstott and al., 2014 : 5), similar to a knee, will be computed as a cutoff, and all nodes above will be considered a potential group of hubs, since it represents in the distribution the point at which the "heavy tail" starts.
4) A cutoff will be made arbitrarily at two standard deviations above the mean. This is done to mimic and test the arbitrary approaches that scholars sometimes use to determine which nodes could be hubs. An example of such an approach is to set a threshold at 1%, 5%, or 10% , as the nodes with the highest degrees.
5) The degree distributions will also be computed independently on the projections, which will offer superficial insight into their connections' behaviour, but will not be analyzed any further. While they propose new methods of analysis of projections' degree distributions,

Vasques Filho and O'Neale also indicate; "To date, the interaction between the degree distributions of bipartite networks and their one-mode projections is well understood for only a few cases, or for networks that satisfy a restrictive set of assumptions" (Vasques Filho and al., 2018 : 022307).

6) The degree centrality is also computed. It is similar to the degree distribution but normalized by the maximum possible degree of a graph. It reflects the importance of nodes in terms of their connections, and how said importance is distributed through a network (Freeman, 1978; Lizardo and Jilbert, 2023). It is a crude measure of centrality, which will be computed only in order to identify any major differences with other measures of centrality and the degree distribution.

3.3.5 Random Graphs

Real networks are complex networks modeled on connections between entities happening in real life, such as protein interactions, social networks, the internet, and such. Random graphs are models of graphs that are built to analyze some of their properties, based on the Erdos-Renyi model (Barabási *and al.*, 2016 : 75; Erdős and Rényi, 1960). With respect to bipartite graphs, researchers adopt the technique of creating random graphs and comparing their structure to those of real graphs to see if certain properties appear by chance (randomly) or due to intrinsic network structures (Guillaume and Latapy, 2006; Raphaël, Guillaume and Tarissan, 2015). This has been extensively practiced on bipartite networks, including on networks studying boards of directors. Conyon and Muldoon specifically compare real networks of boards of directors to random graphs they create based on desired degrees distributions to determine whether their real networks display "small-world" properties (Conyon *and al.*, 2004). In the context of the present research, the specific property that will be studies is that of the presence of "hubs".

The Erdos-Renyi models posits that a graph can be built by connecting randomly assigning edges to vertices with the same probability, from which its qualification as a "random graph". Accordingly, some nodes will gain numerous links, while others gain no links or very few links. For unipartite graphs, this results in a degree distribution of Poisson shape, much different from that of scale-free real networks. Barabasi and al. indicate that "in a large network the degree of most nodes is in the narrow vicinity of <k>", where <k> is the average degree (2016 : 81). They compare the degree distribution with that of real networks and indicate that influencers, "nodes with a very large degree", are absent from random networks (Barabási *and al.*, 2016 : 81). They point to the rewritten Poisson distribution $P_k = \frac{e^{-<k>}}{\sqrt{2\pi k}}\left(\frac{e(k)}{k}\right)^k$ where for increasingly large k, k dependent terms decrease rapidly, predicting that "in a random network the chance of observing a hub decreases faster than exponentially" (Barabási *and al.*, 2016 : 81).

Since the degree distribution of the bipartite graph is composed of the distributions of both individuals and companies, the cutoffs will be evaluated for the projections as well as the bipartite graph. For a random bipartite graph, if one of the degree distributions follows a power law, so will that of the random bipartite graph (Guillaume *and al.*, 2006 : 13). This paper will then evaluate the differences randomly generated bipartite graphs and observe any difference in maximum degree.

This paper will use this consideration of hubs as potential influencers within the network, that is, it will consider that any nodes in the real graph with degrees higher than the maximum degree of a randomly generated graph are potential "hubs" and therefore potential influencers in a social network. The generation of random graphs will be done in a fashion as to replicate the size the graph, as opposed to Conyon and Muldoon who use the degree distribution, as this paper is specifically interested in studying the difference between the degree distributions of the random and real networks. Therefore, two methods of replicating the size of the network are used:

- Number of nodes and density of the original graph;
- Number of nodes and edge probability of the graph. Said edge probability is determined on the original graph by $p = \frac{2L}{N(N-1)}$, that is, by the ratio of the actual number of edges to the total possible edges ($\frac{N(N-1)}{2}$).

The random graphs are generated for both the bipartite graph and the projections. On the bipartite graph, the generation will be done through both unipartite and bipartite functions nx.fast_gnp_random_graph nd nx.bipartite.random_graph, both based on the aforementioned Erdos-Renyi random graph.

3.3.6 Additional discarded method

Researchers rely on a variety of techniques to determine what are meaningful cutoffs influencers identification. The present paper has presented multiple methods which are tested to identify which meaningful influencers. As was explained with centrality measures, some of these performant methods cannot be applied to the present network due to their high complexity that would necessitate high computing power considering the size of the network. An important and relevant method that has been discarded is that of testing network robustness without influencers. In a seminal Nature article cited over 10,000 times published in the journal, Jeong and al. find that removing influencers from a network increases its diameter (longest shortest path in a network) (Jeong *and al.*, 2001). This method is applied in an iterative way to identify which nodes are influential. This method is indeed widely used in influencers analysis, however, it is too cumbersome to use on on large graphs due to its complexity of $O(V * (V + E))$.

3.3.7 Community detection

To identify important nodes in the network, central nodes within communities will also be identified. In a graph, nodes can be said to belong to a specific community of nodes, where a community is defined as a "locally dense connected subgraph in a network" (Barabási *and al.*, 2016 : 325). This approach has been tested by Mester and al. (2021), with positive results. They find that both centrality measures and community detection identify similar key nodes. This implies that there is a credible correlation between hub-dominant nodes, as identified by community detection, and nodes that rank highly on centrality measures. Their findings were conclusive especially for less complex, or "lightweight" network (Mester *and al.*, 2021). Since the network at hand is found to be sparse, it is expected that using Mester and al's approach would yield similar results in terms of identifying key nodes. The use of identifying key nodes based on communities and to evaluate their ESG scores is to understand whether nodes that are influential in their own communities, rather than on the overall graph, have higher ESG scores.

The algorithm that will be used is the "Louvain" algorithm, which is based on modularity optimization, where modularity measures the quality of a partition, meaning that the Louvain algorithm aims to find the optimal structure of communities inside a network (Barabási *and al.*, 2016 : 355). The algorithm will be run multiple times, as outputs can differ, and the output with highest modularity will be chosen. A high modularity will be aimed at, to validate the meaningfulness of the communities.

The nodes with the highest "hub dominance" will be identified for each community. The hub dominance is defined as "the ratio between the degree of the largest hub of the component c and the size of the component c" (Diop *and al.*, 2021 : 15). The hub dominance will be implemented for each community C as: $hub\ dominance(C) = \frac{\max(k(n))}{N-1}$, where N is the number of nodes within the community. The implementation follows these steps, iterating over each community:

1. Computing the inner degree of a node within community C
2. Identifying the nodes with the highest degree
3. Calculating the ratio of the hub's inner degree to the size of the community – 1 (effectively the maximum possible degree)
4. Returning the results on a scale of 0 to 1.

The dominant hubs returned will thereafter be evaluated against their eigenvector centrality. The dominant hubs might be different than the nodes with the highest eigenvector centralities as they are ones whose centrality is 1) based on their degree and 2) reflective of the potential to spread an attitude within a community, as opposed to the entire graph.

3.3.8 Mann-Whitney U test

The Mann-Whitney U test will be used to quantify the "meaningfulness" of the groups of influencers. This will be done by comparing the eigenvector centralities of each group of influencers to the overall graph's eigenvector centrality, and evaluating if there is a meaningful difference, thereby confirming the pertinence of the groups of hubs. The Mann-Whitney U test is a non-parametric statistical test used to compare two independent samples, identifying whether the distributions of the two groups are statistically different (Mann and Whitney, 1947). The test ranks the eigenvector centralities from both distributions together and then evaluates if one distribution tends to have higher ranks than the other, essentially testing the medians of the two distributions against each other. The Mann-Whitney U test will be used instead of other statistical tests as it does not require any distribution shape and does not require homogeneity of variance. Two hypotheses are therefore tested:

The null hypothesis

$$H_0 = \textit{There is no difference between the eigenvector distribution of the bipartite graph and the eigenvector distribution of the group of influencers.}$$

The alternative hypothesis

$$H_1 = \textit{There is a difference between the eigenvector distribution of the bipartite graph and the eigenvector distribution of the group of influencers.}$$

The test will follow these steps:

1) The eigenvector centralities of the entire the groups of influencers will be isolated.
2) As an iterative process, they will be removed from the eigenvector centrality distribution of the graph for every group before performing the test.
3) The rank will be turned into a *z*-score for ease of interpretation of the ranks.
4) The *p*-value will be evaluated for statistical significance, where any value above 0.05 will lead to the test being discarded for that possible group of influencers.
5) Significance level will be set at α=0.05 (confidence level of 95%), for a critical z value of |1.96|.
6) The null hypothesis will be rejected for any possible group of influencers if the z value is below -1.96 or above 1.96. Conversely, the alternative hypothesis will be accepted.

It must be noted that a two-tailed Mann-Whitney test will be performed as the interest lies in the difference between the medians, not whether one is higher or lower, that is, whether the groups of hubs have higher or lower eigenvector centralities. A higher centrality would be represented by a lower z (indicating the median of the distribution of hubs is higher than that of the original distribution), as is expected. A negative z score however would point to lower eigenvector centralities in the groups, which is not expected but may happen for groups not created based on eigenvector centralities.

The test will be computed on the original graph and not on the projections, as the eigenvector centrality distribution has been determined too biased on the projections. Instead, the eigenvector centrality distribution of the original graph will be divided between the companies and the individuals, as will the nodes within the groups, and the test will be performed. This will be done due to the high differences between the values of the individuals and companies on the eigenvector centrality distribution. The test will be performed using the "mannwhitneyu" package in python.
Interpretation of meaningfulness will also have been offered previously based on comparison with eigenvector centralities, assortativities, and other factors. Therefore, potential groups of influencers that would have already been deemed not meaningful will have been discarded.

3.3.9 ESG Measures

As a final step, the ESG scores of the groups of hubs identified as meaningful by the Mann-Whitney U test will be compared to those of all companies. This is done in two ways:

1) *Visual comparison between ESG distributions.* Assessment of the ESG distribution is made to understand whether influencers tend to have the same ESG scores as the rest of companies, or whether they tend to find themselves in an extreme or another.
2) *Spearman correlation test.* The Spearman correlation test is computed to help determine if certain groups of hubs tend to have higher or lower ESG scores than the overall distribution of ESG scores of the companies in the graph. This will be done by comparing the eigenvector centrality distribution to the ESG one, for the top and bottom nodes of the overall graph. The Spearman rank-order correlation test therefore evaluates the strength and direction of the relationship between the two ranked sets of data. The test returns the coefficient r_s between -1 and 1, where 1 signifies a perfect negative correlation and 1 a perfect positive correlation, whilst 0 indicates no correlation, as well as the *p*-value, which will be evaluated in the same way as for the Mann-Whitney U test.

Chapter 4: Results

4.1 Preliminary Analysis of the Network

The original network output is unconnected, with 81671 individuals, 12362 companies, 211079 links. As expected, there is a giant connected component of 93850 nodes, and 12 other connected components of sizes below 20.

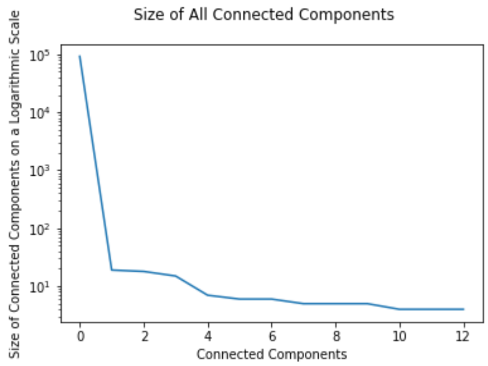


Figure 1. Size of connected components.

A visual inspection of the second largest component offers insight into why these unconnected components can be removed safely from the graph without bias.

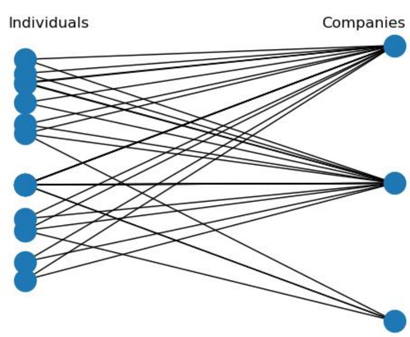


Figure 2.1. Bipartite structure of the second largest connected component.

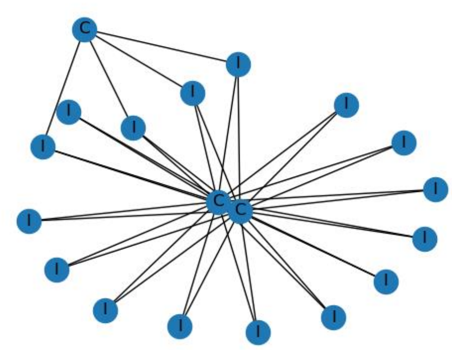


Figure 2.2. Second largest connected component under the Fruchterman-Reingold layout.

Indeed, as can be seen in Figure 2.2, this component is disconnected from the main business network as the individuals present in this component are exclusively affiliated to three companies, and all of them to the same two companies. Therefore, this structure has no individual-company relationship with the original network and can be safely removed.

The Giant Component as Graph *G* and its projections

The largest connected component is used for all subsequent analyses and will be referred to as graph *G*. Some initial statistics on size, structure, and ESG can be viewed in Table 1.

	Bipartite Graph	Companies projection	Individuals projection
# nodes	92143	11297	80846
# edges	210929	165240	4702884
Density	4.97E-05	2.59E-03	1.44E-03
Median ESG score	30	30	41
Average ESG score	32.96	32.96	44.04
Degree assortativity	-0.18	0.16	0.6
ESG assortativity	0.55	0.21	0.42
Redundancy	0.41	0.65	-
Transitivity	0	0.12	0.52

Table 1. Initial statistics on the largest connected component (graph G) and its projections

The following subsections will discuss the three graph's behaviour:

4.1.1 Network Structure

Bipartite graph

- The network has an average redundancy of 0.41, which means that on average, 41% of the neighbours of a node are connected to each other – a moderate overlap in connections. In the context of the bipartite graph, those neighbours are from the opposite sets. The implication for the spread of attitudes is that companies and individuals can receive overlapping information. In terms of ESG scores, since assortativity is moderately high, this indicates that nodes receive similar ESG attitudes from multiple neighbours.

- As expected, its transitivity is of 0, as no triangles could and should be present in a bipartite graph.
- Its density of 4.97E-05 qualifies it as a sparse graph (where the number of links L is qualified as $L \ll L_{max}$), in line with most real networks (Barabási *and al.*, 2016 : 53). As density is measured from a scale of 0 to 1, where a density of 0 represents a graph without edges while a density of 1 indicates a graph where every pair of vertices is connected by a unique edge, it can be concluded that the network is sparse.

Companies projection

- As expected, the redundancy coefficient for the Companies projection is 0.24 higher than that of the bipartite graph, indicating that a high number of overlapping links have indeed been created during the creation of the projection, which will have implications for centrality measures and influencers based on degree distribution.
- A global clustering coefficient of 0.12 indicates moderate level of clustering, where 12% of a randomly picked node's neighbours will be connected. It was expected for the projections to have a certain level of transitivity, as they may share business leaders.
- The Companies projection has the highest density of all three graphs. It is indeed to be expected behaviour that companies would be more connected overall to each other than individuals would be between each other, due to the nature of the entities and since individuals are many in the graph.
- The implication of these results is that companies might act as hubs connecting diverse leaders.

Individuals projection

- The Individuals projection has a high global clustering, where 52% of a randomly picked node's neighbours will be connected. In this case, transitivity indicates which leaders are linked by virtue of being associated with the same company. Considering that many business leaders serve on the board or as executives for multiple companies, they create links with other business leaders with whom they have an overlapping network.

Individuals’ transitivity is also higher than that of the Companies projection as there are many more business leaders than companies in the network. Redundancy was not computed due to time complexity, reflective of a high number of edges.

- The implication is that the Individuals projection is sparse but clustered, meaning that the network of individuals might not be widely connected, rather there might be specific groups of business leaders who collaborate across multiple companies.

These findings validate the approach of this use to use the bipartite graph as the main one for subsequent analyses.

<u>4.1.2 Assortativity</u>

Bipartite graph

- The light degree disassortativity of the bipartite graph (-0.18) indicates that nodes with higher degrees have a slight tendency to tend to connect with nodes with lower degrees and vice-versa. That is, some companies have a slight tendency to connect to many individuals, who themselves are affiliated to few companies, and some companies have a slight tendency to few individuals who themselves are affiliated to many companies. This can be an indication that certain nodes act as influencers and connect to many lesser-connected nodes. These results are in line with those expected for a bipartite social network where the network flow is that of influence (Vasques Filho *and al.*, 2020).
- The fairly strong numeric assortativity (0.55) reveals ESG assortativity may be skewed towards higher results for individuals working at few companies (and vice-versa). This is particularly true given individuals have lower degrees, making the weight of the score of the company represented in the relationship a relatively higher one.

Companies projection

- The light degree assortativity (0.16) and ESG assortativity (0.21) reveal a light tendency for companies to affiliate with companies of similar degrees and ESG scores. This is at odds with the bipartite graph. In the case of degree assortativity, this can be explained by induced edges, redundancy, and transitivity. In this case, degree assortativity for the

Companies projection will be used only as a basis for comparison with groups of influencers.

- With respect to the ESG assortativity being much lower than on the original graph, it is unclear in which proportion it is due to the ESG assortativity on the bipartite graph being indeed skewed or to induced links in the projection creating multiple links to certain companies. This is indicative of ESG assortativity not being a robust measure in this case. ESG assortativity should therefore be used rather as a baseline for comparison for groups of hubs, not for insights in and of itself.

Individuals projection

- The high degree assortativity (0.6) is a clear indication of the strong effects of induced links, transitivity, and redundancy, as for the Companies projection.
- The same factor influence ESG assortativity. Additionally, in this case, the moderate effect of companies associating with others of similar ESG scores can also be due to the way in which ESG scores are computed for individuals (averages of the companies' scores that individuals are affiliated with). Indeed, these are higher than those of companies. This is due to companies' scores being duplicated in those averages – that is, the same company will appear in the scores of all the individuals it is affiliated to. The higher average ESG scores of individuals reflect that companies with higher ESG scores have more weight in those values, that is, that they have more connections to business leaders. This is also reflected in the more symmetric distribution of ESG scores of individuals (Figure 3.2), which resembles more a normal distribution, while the distribution of non-duplicate companies is positively skewed.

4.1.3 ESG Distribution

The ESG distribution is observed in Figure 3.1 skewed to the right shows that a portion of companies are outliers on the higher end of the distribution. This paper will seek to identify whether those companies are also influencers or not.

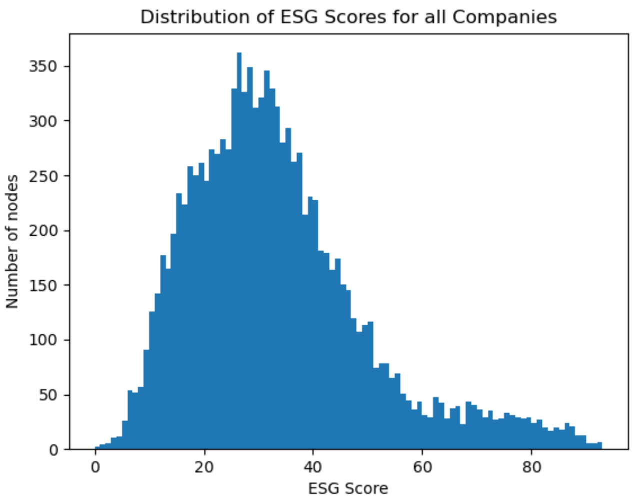


Figure 3.1. Distribution of ESG scores computed on the companies of graph G

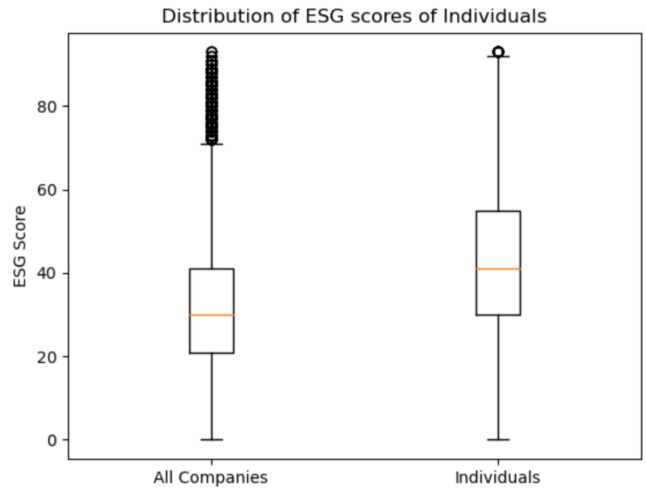


Figure 3.2. Distribution of ESG scores computed on individuals and companies

4.1.4 Degree distribution

The degree distribution clearly shows it is built on a double distribution, whereby individuals have fewer connections than companies. The median degree is 2, which on a non-logarithmic graph can be seen to be caused by the individuals' degree.

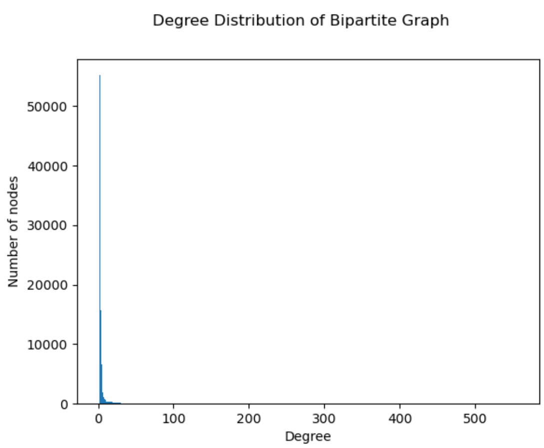


Figure 4.1. Degree distribution on bipartite graph

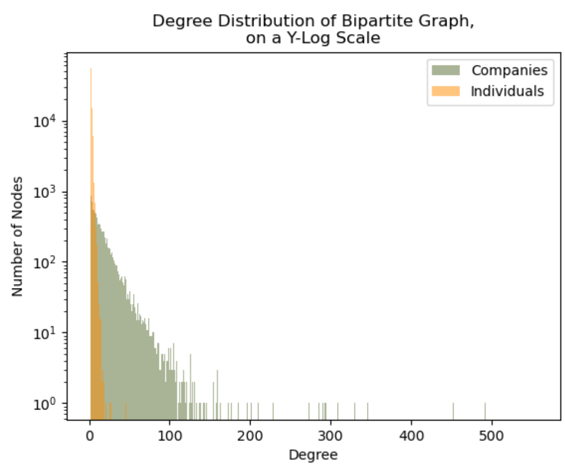


Figure 4.2. Degree distribution on bipartite graph (y-log scale)

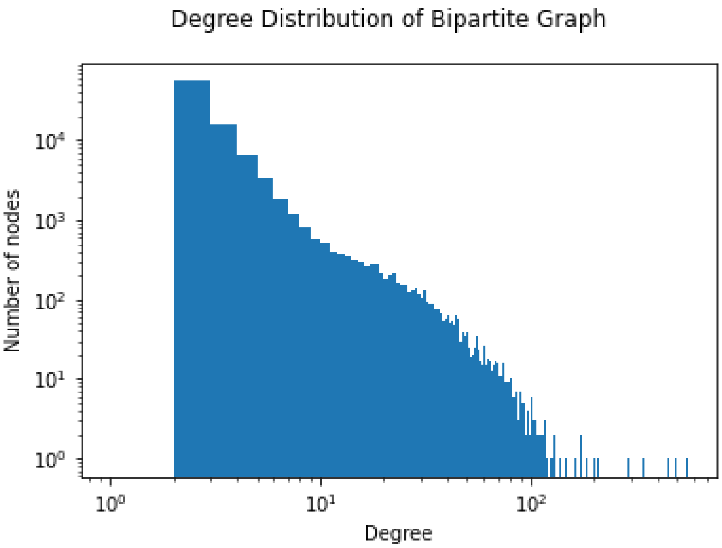


Figure 4.3. Degree distribution on bipartite graph (log-log scale)

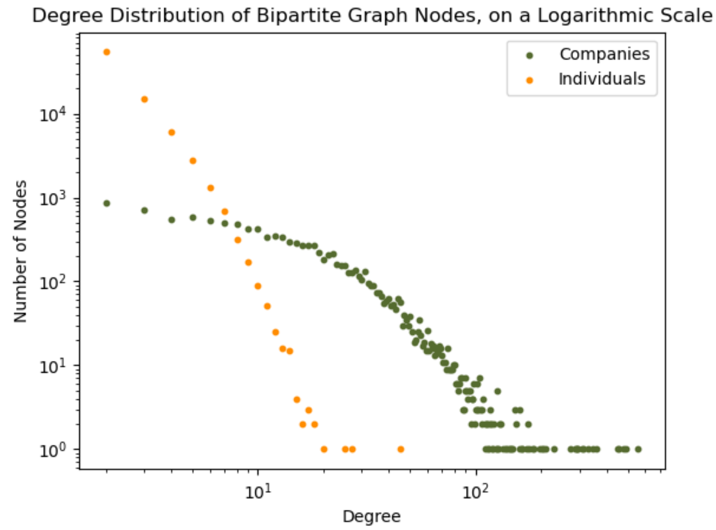


Figure 4.4. Degree distribution by top and bottom nodes on bipartite graph (log-log scale)

Figures 4.3 and 4.4 show on a log-log scale how the two distributions compose the first. The lack of degrees of 1 can be noted as expected. The two distributions were compared fitted against a power law distribution and an exponential one through a log likelihood ratio test. Companies' α resulted in 3.8, whilst individuals' is 5.74, which is indicative that companies' distribution has a heavy tail typical of power law distributions and an overall higher fit with a power law distribution than individuals. The results of the log likelihood ratio test indicated the distributions followed more closely a power law distribution, whereby the result was of 51.58 for companies (with statistical significance of $p = 0.000179$) and of 16.51 for individuals (with low statistical significance of $p = 0.26$). A chi-square test was also performed against the individuals' degree sequence to identify a potential Poisson distribution, refuting it (χ^2 of 6.61E+27 with $p = 0$). The individuals' degree sequence is therefore considered rather an uncertain power law distribution.

While the degree distributions of the projections are biased, their visualization is interesting as it permits to ese the effect of projecting edges, as a contrary effect of the distribution on the bipartite graph. Indeed, the distribution of the Companies' projection reflects lower degrees as nodes' connections reflect business leaders in common, whilst the Individuals' one reflects higher degrees due to being connected to many individuals through common companies.

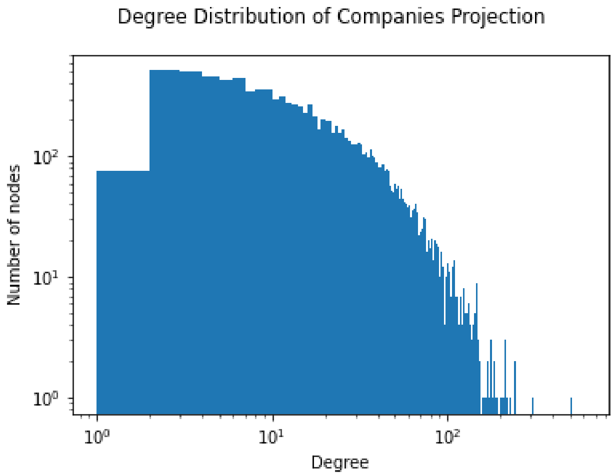


Figure 5.1. Degree distribution on the Companies projection (log-log scale)

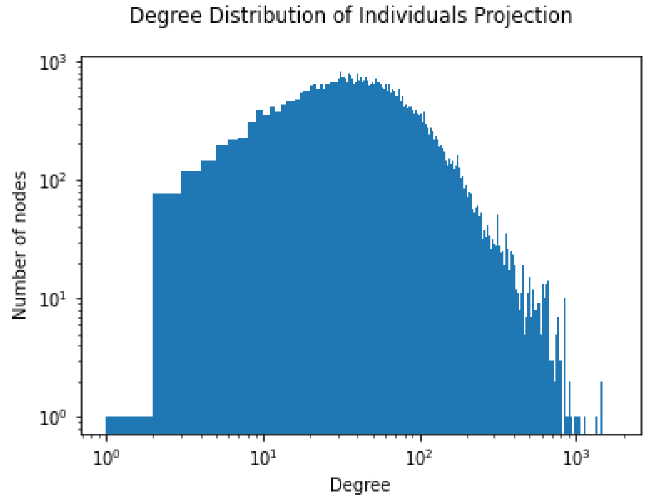


Figure 5.2. Degree distribution on the Companies projection (log-log scale)

The bias introduced by induced links, redundancy, and transitivity can be best visualized in the presence of degrees of 1 in the distributions, that had been removed in the original graph. Therefore, as has been discussed, the degree distribution of the bipartite graph will be utilized for analyses.

The degree centrality distributions reflect the distributions without the removal of the degrees of 1, as they have been normalized over the maximum degree of the graphs.

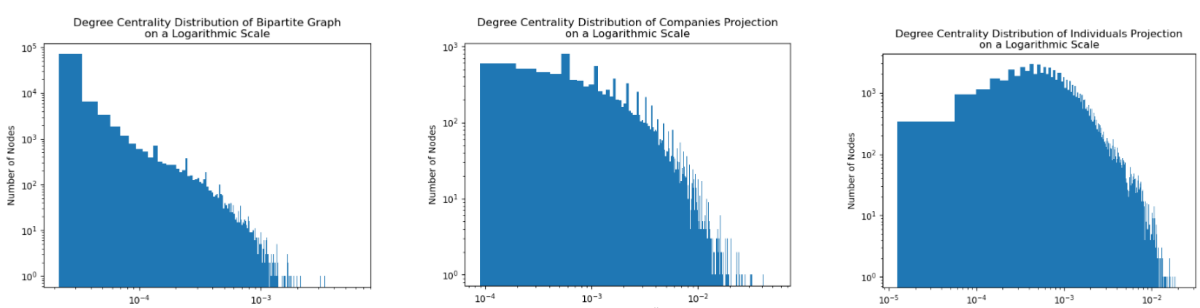
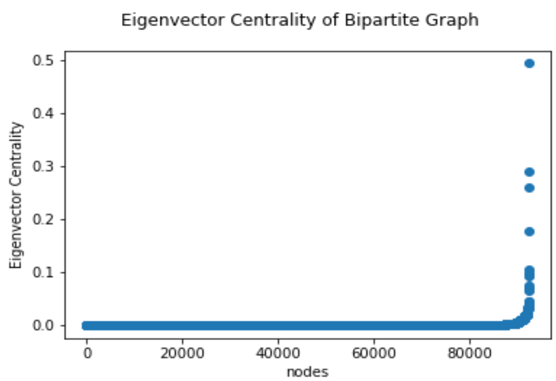


Figure 6. Degree centralities on the bipartite graph, the Companies projection and the Individuals projection.

4.1.5 Eigenvector Centrality



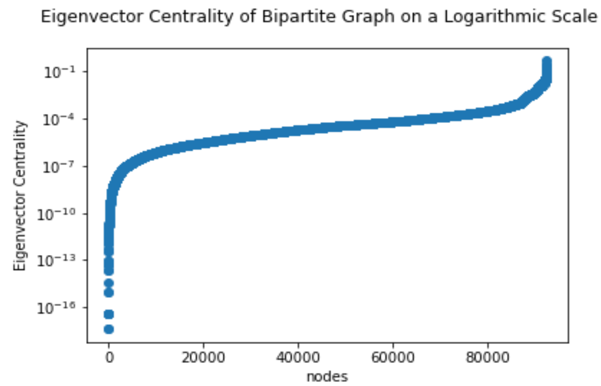
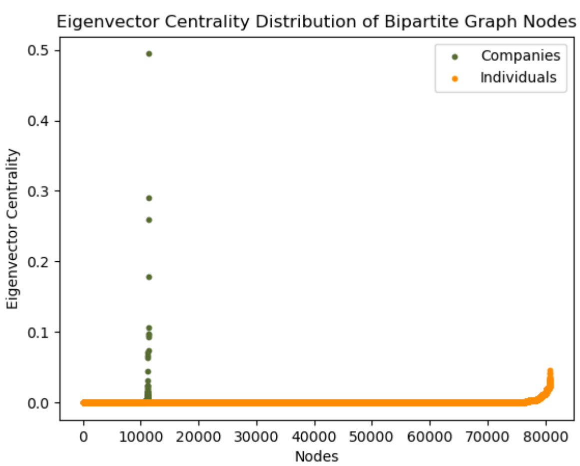


Figure 6.1. Eigenvector centrality of bipartite graph

Figure 6.2. Eigenvector centrality of bipartite graph, on a y-log scale



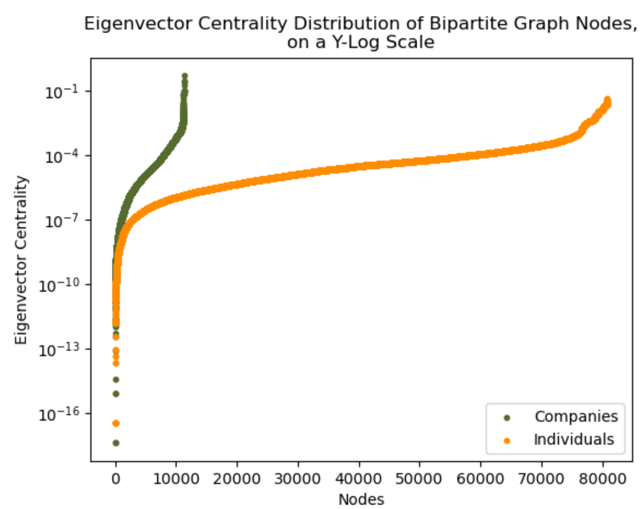


Figure 6.3. Eigenvector centrality of bipartite graph, separated by nodes

Figure 6.4. Eigenvector centrality of bipartite graph, separated by nodes, on a y-log scale

The eigenvector centrality distribution represented in figures 6.1 to 6.4 offer a few main insights:

- Companies have proportionally higher centralities, pointing to companies as the influencers most indicated in spreading attitudes in the business network (figure 6.4).

- A small number of nodes have centralities an order of magnitude higher than the rest, shown by the sharp increase in figure 6.1.
- Related to the abovementioned point, most nodes have a centrality between 10^{-7} and 10^{-3}, distinguishing nodes above the latter centrality, pointing to potential influencers. This will be validated in the next sub-sections.
- The eigenvector centrality shows that influencers can be both individuals and companies (nodes with high centralities), as opposed to the degree distribution, where the highest degrees belong to companies (figure 6.4).

The sharp increase in eigenvector centrality is also observed on the projections, validating the findings on the bipartite graph.

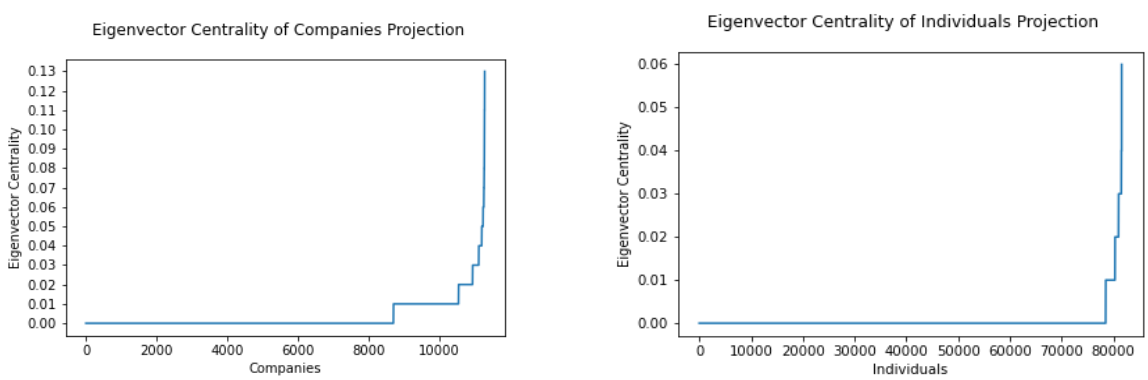


Figure 7. Eigenvector centrality distributions on projections

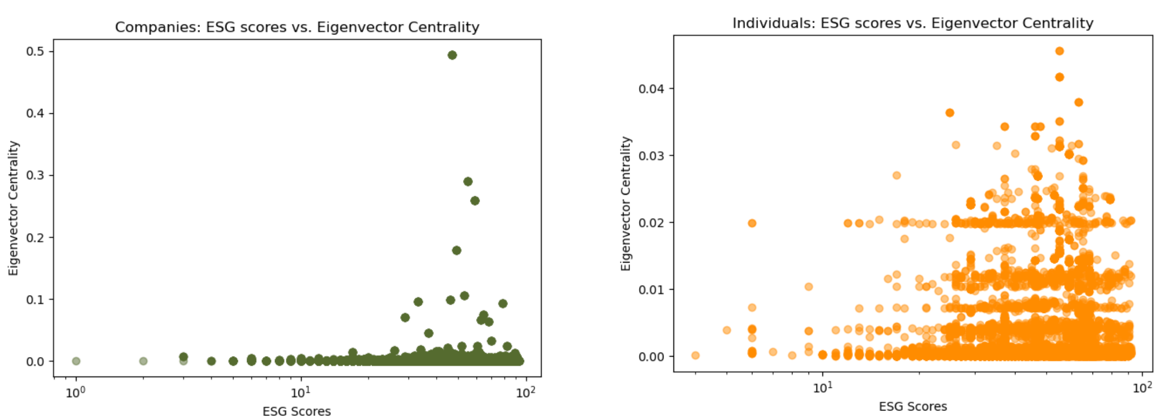


Figure 8. Dispersion of ESG scores against eigenvector centrality values on the bipartite graph

A first view of the eigenvector centralities against ESG scores point to nodes with higher values indeed displaying higher ESG scores.

4.2 Groups of Influencers

4.2.1 All potential groups of influencers

This chapter will indicate the results of the different outputs analyze the different possible groups of influencers, as presented in table 2 and table 3. Influencers identified through hub dominance within communities are not present in the table, as they are not part of continuous groups, and are discussed in the last subsection.

	Cutoff			% hubs (nodes above the cutoff)		
	Bipartite graph	Companies projection	Individuals projection	Bipartite Graph	Companies projection	Individuals projection
Hubs with the 10 highest eigenvector centralities	-	-	-	-	-	-
Hubs with the 12 highest eigenvector centralities	-	-	-	-	-	-
Hubs with the highest eigenvector centralities, with cutoff at the knee	0.0049	0.0078	0.0012	2.36	14.67	6.47
Hubs with the highest degree, with cutoff at 2 Std Dev above the mean	-	-	-	2.92	-	-
Hubs identified by random projections degree	-	54	168	-	14.36	17.56
Hubs identified by random bipartite graph degree generated with equivalent density	15	-	-	5.04	-	-
Hubs identified by random unipartite graph degree generated with equivalent density	18	-	-	4.16	-	-
Hubs identified by random bipartite graph degree generated with equivalent probability	13	-	-	5.7	-	-
Hubs identified by random unipartite graph degree generated with equivalent probability	16	-	-	4.75	-	-
	Bipartite Graph	Companies within graph	Individuals within graph	Bipartite Graph	Companies within graph	Individuals within graph
Hubs with the highest degree, with cutoff at xmin	-	60	7	-	4.06	0.87

Table 2. All possible groups of hubs identified by their cutoffs.

	Degree Assortativity			ESG Assortativity		
	Bipartite Graph	Companies projection	Individuals projection	Bipartite Graph	Companies projection	Individuals projection
Original graph	-0.18	0.16	0.6	0.55	0.21	0.42
Hubs with top 10 eigenvector centralities	-0.92	-0.04	0.31	0.59	0.15	0.27
Hubs with top 12 eigenvector centralities	-0.91	-	-	0.6	-	-
Hubs with the highest eigenvector centrality, with cutoff at the knee	-0.28	-	-	0.46	-	-
Hubs with the highest degree, with cutoff at 2 Std Dev above the mean	-0.28	-	-	0.58	-	-
Hubs identified by random projections degree	-	0.13	0.59	-	0.19	0.42
Hubs identified by random bipartite graph degree generated with equivalent density	-0.22	-	-	0.57	-	-
Hubs identified by random bipartite graph degree generated with equivalent probability	-0.23	-	-	0.57	-	-
Hubs identified by random unipartite graph degree generated with equivalent density	-0.25	-	-	0.57	-	-
Hubs identified by random unipartite graph degree generated with equivalent probability	-0.24	-	-	0.57	-	-
	Bipartite Graph	Companies within graph	Individuals within graph	Bipartite Graph	Companies within graph	Individuals within graph
Hubs with the highest degree, with cutoff at xmin	-	-0.41	-0.43	-	0.28	0.28

Table 3. Assortativities of all possible groups of hubs.

4.2.2 Top 10 Highest Eigenvector Centralities

In this subsection, the 10 nodes highest eigenvector centralities are analyzed. This analysis is conducted on the bipartite graph and the projections. Since these influencers are taken as a baseline for analysis for other influencers, this analysis is conducted on the projections as well, which offers more insight into the difference in results.

Bipartite Graph

Eigenvector centrality rank	Company	Eigenvector Centrality	ESG score
1	Morgan Stanley	0.494	47
2	Citigroup Inc.	0.291	55
3	Deutsche Bank Aktiengesellschaft	0.259	59
4	JPMorgan Chase & Co.	0.179	49
5	Credit Suisse Group AG	0.106	53
6	The Goldman Sachs Group, Inc.	0.098	46
7	Blackstone Inc.	0.095	33
8	UBS Group AG	0.093	79
9	Bank of America Corporation	0.075	65
10	The Carlyle Group Inc.	0.071	29

Table 4. Eigenvector centralities of the nodes with the 10 highest values computed on the bipartite graph.

The 10 highest eigenvector centralities display a significant dispersion in values, going from close to 0.5 to below 0.1 after the first five values. This suggests that first five values are significantly more influential than the others. It is interesting to note that the 10 highest eigenvalues all pertain to companies all from the financial sector. The fact that individuals are absent from the top could indicate that it is companies that are the most influential entities in the network. However, these results account only direct connections and no soft influence or intra-company influence.

Thee ESG scores of these companies are also significantly higher than those of the rest of the companies, although they are not part of the outliers. While this might be due to ESG of the financial industry differing to that of other industries, this is aligned with the current management literature that indicates that financial giants are the ones leading the ESG influence network (Ramaswamy, 2023).

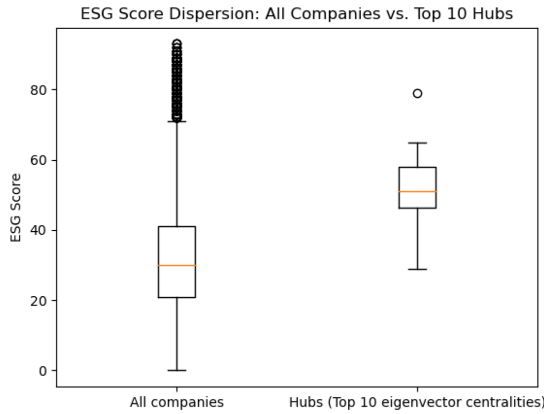


Figure 9. Distribution of ESG Scores of the influencers with the highest eigenvector centralities

These companies have an assortativity of -0.92, exhibiting an almost entirely disassortative behaviour. Whilst very high, this result is in line with the of Vasques and O'Neale who had found an assortativity of -0.65 for the network of Norwegian board members (Vasques Filho *and al.*, 2020 : 5). Behaviour is expected to be exacerbated for the top influencers. Combined with the high eigenvalues, it can be inferred that these are companies that are led by individuals with few connections to other companies, indicating a potential concentration of power. The ESG assortativity is higher than that of the original graph, indicating that they tend to associate with individuals that themselves are affiliated to companies with similar scores.

Companies projection

Eigenvector centrality rank	Company	Eigenvector Centrality	ESG Scores
1	Citigroup Inc.	0.130	55
2	Deutsche Bank Aktiengesellschaft	0.121	59
3	Morgan Stanley	0.118	47
4	Blackstone Inc.	0.111	33
5	General Electric Company	0.110	46
6	KKR & Co. Inc.	0.106	37
7	JPMorgan Chase & Co.	0.105	49

8	The Goldman Sachs Group, Inc.	0.103	46
9	The Carlyle Group Inc.	0.100	29
10	International Business Machines Corporation	0.093	52

Table 5. Eigenvector centralities of the nodes with the 10 highest values computed on the Companies projection

7 out of the 10 companies with the highest eigenvector centralities in the Companies projection also have the highest centralities in the bipartite graph, confirming the results that those companies are indeed the top influencers in the network. It is possible that the three companies that were not part of the projection's top values had their centrality in the overall network diluted due to redundancy, making it so that the 3 other companies have had their neighbourhood centrality inflated. The overall absolute lower eigenvalues in the projection are indicative of the presence of densely connected neighbourhoods, where the centrality of the most influential nodes is distributed evenly across the neighbourhoods but not across the graph, as per the high redundancy of the graph. The degree assortativity of -0.04 and the ESG assortativity of 0.15 being slightly lower than that of the original graph signifies that companies that are influencers have a lesser propensity to associate with similar companies and a more diverse assortativity. This is at odds with the findings for the influencers on bipartite graph, confirming a more random pattern of connections.

Individuals projection

Rank of individuals with highest eigenvector centrality	Eigenvector centrality	ESG Score
1	0.059	55
2	0.055	55
3	0.055	55
4	0.050	64
5	0.048	25
6	0.047	55
7	0.046	48
8	0.046	46
9	0.046	37
10	0.044	46

Table 6. Eigenvector centralities of the nodes with the 10 highest values computed on the Individuals projection

The highest eigenvector centralities of the Individuals projection are significantly lower than in the bipartite graph and Companies projection. This might be due to high the high redundancy and transitivity in the projection, making influence more localized. If the structure of the graph is less distributed than that of the original graph, influencers in the projection might not have an overall influence as high as on the bipartite graph. Since eigenvector centrality extends the neighbourhood over the graph, the behaviour of lower scores is expected. Interestingly, the individual with the highest eigenvector centrality on the projection is affiliated to influencers from the financial industry from the other two groups of 10 highest eigenvector centralities (Citigroup Inc, Deutsche Bank Aktiengesellschaft, Morgan Stanley).

4.2.3 Groups of potential influencers based on ESG assortativity

After an exploration of the ESG assortativity, a sharp increase in assortativity was found up to the 12th node of the bipartite graph, indicative that a distinctive group of influencers might rather be that of the 12 highest than only 10. The nodes up to the 12th are companies. Those comprised between the 13th node and the knee of the eigenvector distribution constitute another group which is composed mainly of individuals (to see in the next subsection). Figure 8 demonstrates that increase qualifying the first group, and then the decrease qualifying the second group.

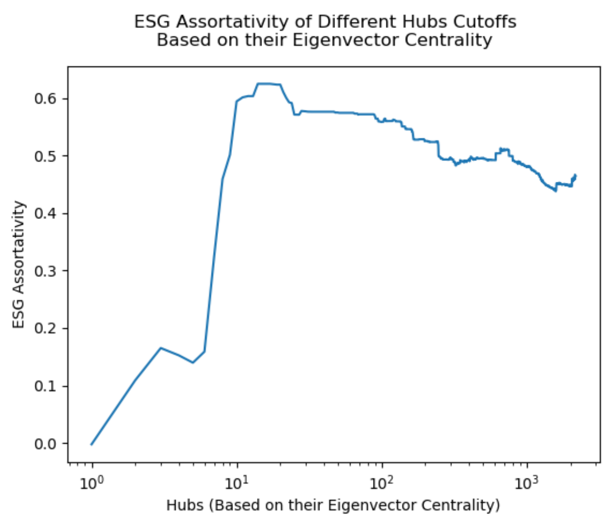


Figure 10. Cumulative ESG assortativity of influencers above the eigenvector centrality's distribution knee, sorted by their eigenvector centrality

The two additional influencers compared to the top 10 values with the highest eigenvector centralities are also important financial institutions, indicative of the meaningfulness of the group of influencers.

Eigenvector centrality rank	Company	Eigenvector Centrality	ESG score
1	Morgan Stanley	0.494	47
2	Citigroup Inc.	0.291	55
3	Deutsche Bank Aktiengesellschaft	0.259	59
4	JPMorgan Chase & Co.	0.179	49
5	Credit Suisse Group AG	0.106	53
6	The Goldman Sachs Group, Inc.	0.098	46
7	Blackstone Inc.	0.095	33
8	UBS Group AG	0.093	79
9	Bank of America Corporation	0.075	65
10	The Carlyle Group Inc.	0.071	29
11	HSBC Holdings plc	0.067	63
12	Barclays PLC	0.064	68

Table 7. Eigenvector centrality of 12 potential influencers based on ESG assortativity

4.2.4 Groups of influencers based on eigenvector centrality

As mentioned, another potential group of influencers is that above the knee of the eigenvector distribution pictured on figure 10. The knee is situated at the 0.0049 value, and the nodes above it to be considered part of the group of hubs represents 2.36% of all total nodes. The group would be composed 95.35% of individuals and 4.65% of companies, which is higher than the overall ratio of the bipartite graph of 86.85%. This reveals that there while the nodes with utmost centrality are companies, there is a large group of influencers who are individuals, as opposed to similar groups of influencers based on the degree distribution, where the nodes with the highest degrees will be companies.

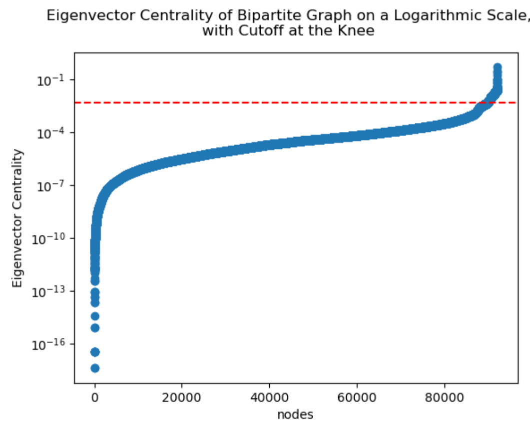


Figure 11. Cutoff at the knee of the eigenvector centrality distribution

The ESG assortativity in this group decreases as does the centrality, after which it will start to increase, making it so that this group of nodes has the lowest ESG assortativity (0.46) of all the groups of influencers. This is likely to be due to the higher presence of individuals in the mix, of which the ESG assortativity is more normally distributed. The degree assortativity is somewhat lower than that of the overall graph (-0.28 compared to -0.18), indicative of a sharp decrease in degree assortativity from the nodes with the highest 10 eigenvector centralities which, where the last 0.1 decrease in points are distributed across more than 97% of the graph. This may indicate that the highest degree disassortativity and the highest discrepancy in degree assortativity happens within this group.

4.2.5 Groups of influencers based on the degree distribution

Group of influencers based on 2 standard deviations above the mean

Another group of influencers was identified as any nodes two standard deviations above the mean of the degree distribution of the bipartite graph. The standard deviation was found to be of 9.79, and the group of hubs to be of 2.92% of total nodes in the graph, similar to that of the knee of the eigenvector centrality. The composition however differs, by which all but 3 nodes are companies. Individuals will therefore not be considered in this group of hubs. The eigenvector centralities for these hubs does not lead to conclusive insights about those nodes' centrality. The degree assortativity is the same as that of the influencers identified by the knee of the eigenvector distribution, invoking a parallel behaviour between nodes with highest degrees or highest eigenvector centralities.

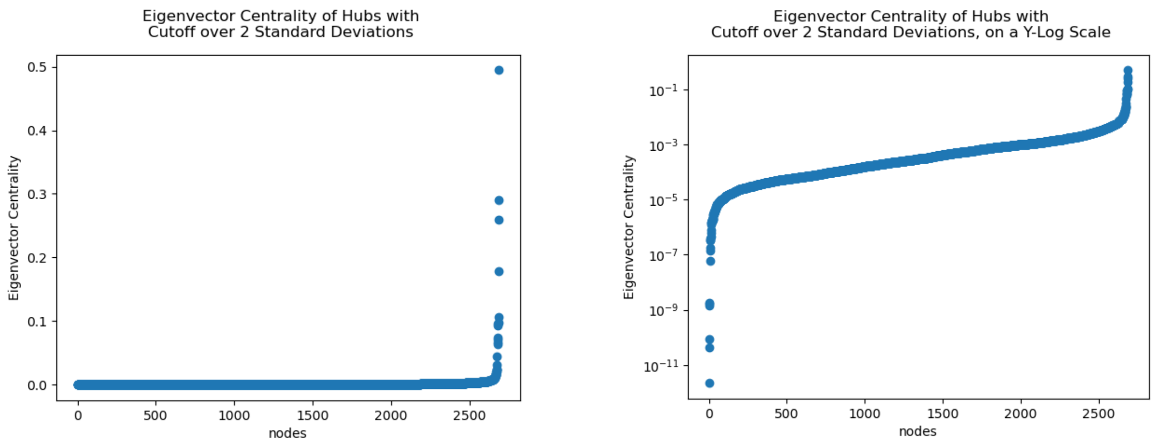


Figure 12. Eigenvector centrality distribution of hubs 2 standard deviations above the mean of the degree distribution

<u>Group of influencers based on x_{min}</u>

The x_{min} was computed on the bipartite graph's degree distributions pertaining to the individuals and the companies, which found respectively degree 7 and 60 as cutoffs, above which the distributions are considered to take a power law shape. The individuals' cutoff identifies 0.87% of individuals to be influencers (703) and to 4.06% of companies (459), arriving at similar numbers of influencers for each group. Interestingly, both groups sport the same ESG assortativity (0.28) and similar degree assortativity (-0.41 for companies and -0.43 for individuals). The degree assortativity is much lower than that of the original graph (more disassortative), closer to that of the 10 most influential hubs (quasi fully disassortative). The ESG assortativity however is more similar to the one computed on projections than on the bipartite. The ESG assortativity for this group was computed between entities in these groups and their neighbours within the graph.

4.2.6 Groups of influencers based on random graphs' maximum degrees

<u>Random bipartite graph</u>

As mentioned in the methodology section, random graphs were generated as bipartite and unipartite graphs. The degree distribution of graphs generated with density and edge probability followed a similar shape.

The graphs generated as a bipartite graph are seen to follow a power law distribution (figure 13), while the ones generated as unipartite follow a Poisson shape (figure 14), as expected.

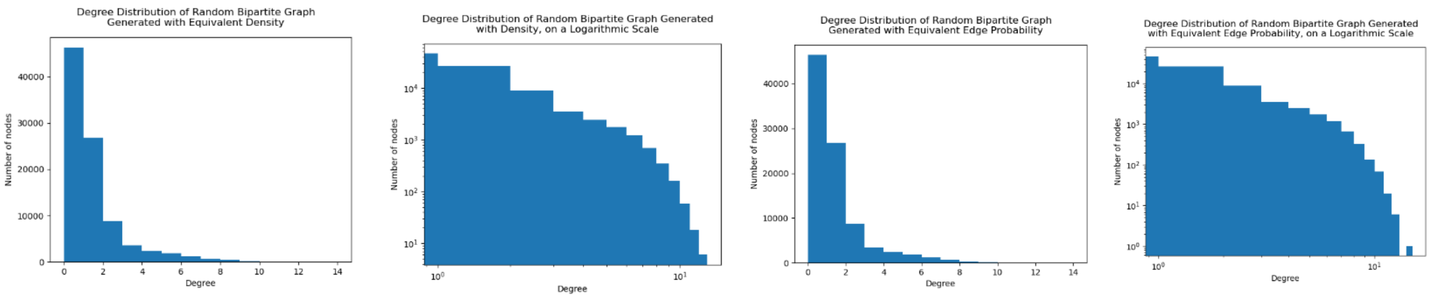


Figure 13. Random bipartite graphs generated as bipartite graphs

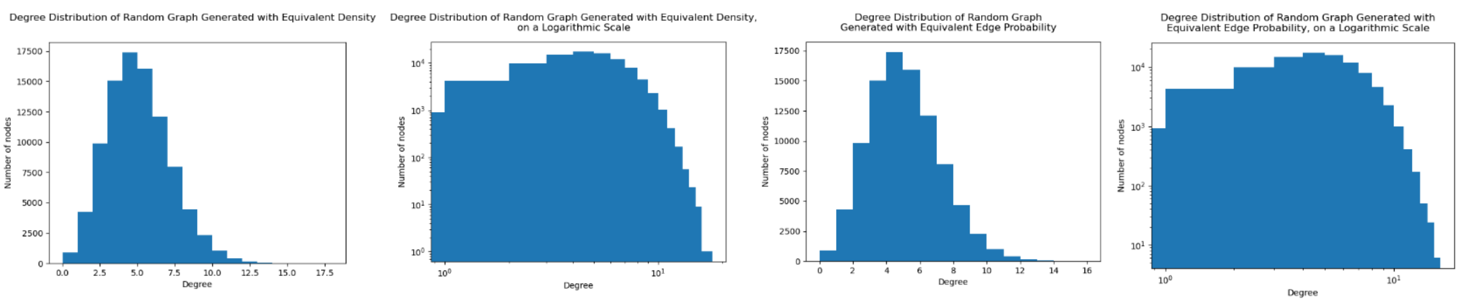


Figure 14. Random bipartite graphs generated as unipartite graphs

Apart from the degree distribution, the results for all four randomly generated networks are similar. The graphs generated as bipartite have a maximum degree that is slightly lower than that of the graphs generated as unipartite. The highest maximum degree would be that of the random unipartite graph generated with equivalent density (degree of 18). If the influencers are to be any node above this cutoff of the degree distribution of the original bipartite graph, influencers represent 4.46% of the total nodes. The lowest maximum degree would be that of the random bipartite graph generated with equivalent edge probability (degree of 13) (figure 14). If the influencers are to be any node above this cutoff of the degree distribution, influencers represent 5.7% of the total nodes. It must be noted that the random graphs do contain degrees of 1, which will have an impact on the maximum degree of the random graphs, and their differences in distribution with the original graphs.

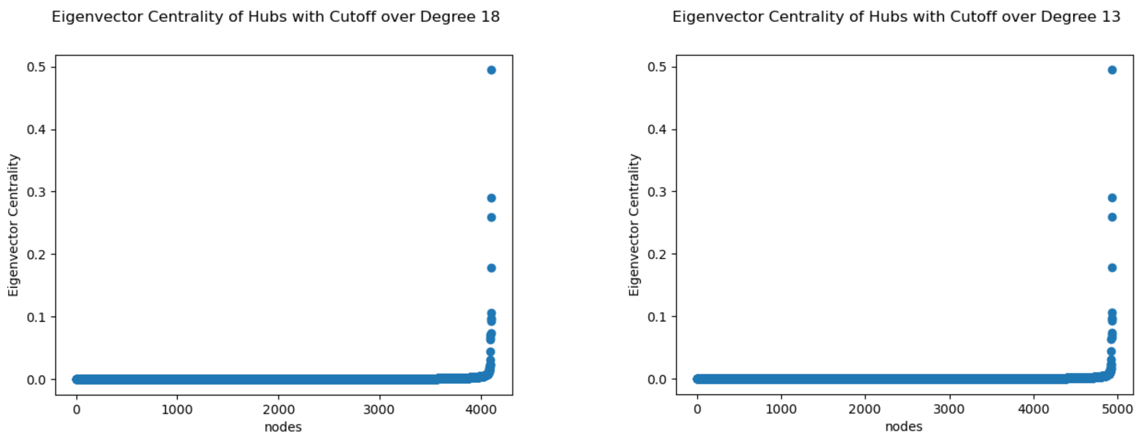


Figure 15. Eigenvector centralities of influencers generated with the lowest and highest cutoff of the random bipartite graphs

The maximum degree on these randomly generated graphs is much lower than that of the original graph which has a maximum degree of 557. The original graph therefore has a heavier right tail in its degree distribution, reflective of the nature of the nodes in that heavy tail, which can be considered as hubs. The degree assortativity of those hubs however is similar to that of the original graph (between -0.22 and -0.25 for all 4 graphs), indicating similar degree structure for most of that heavy tail as for the rest of the graph, as well as the ESG assortativity which is the same for all 4 graphs (0.57) and 0.02 points to that of that of the original graph.

<u>Random projections</u>

Random graphs were algo generated as equivalents of the projections. While still Poisson, the shape of the distribution has been affected by the number of nodes high density/edge probability provided, where for such dense graphs, the median degree has increased, resulting in no nodes with low degrees, and higher maximum degree. These distributions therefore reflect the effect of induced edges, redundancy, and transitivity in the original projections. The maximum degree of the random projections and the degree distribution of the random projections would therefore not be indicated for detecting meaningful groups of hubs, further reinforced by the % of resulting hubs, much higher than that of other methods (14.36% for companies and 17.56% for individuals).

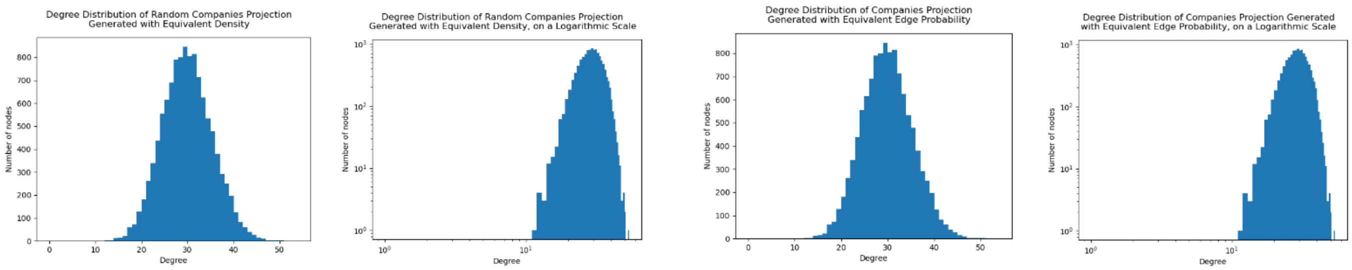


Figure 16. Random Companies projection

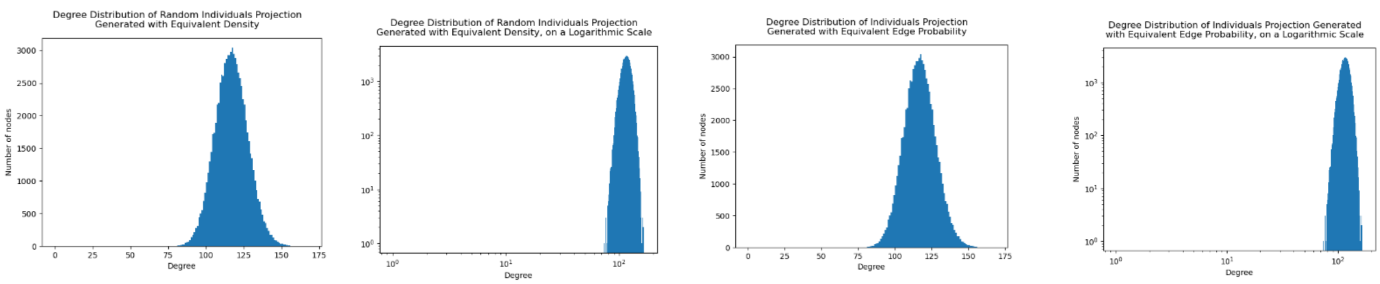


Figure 17. Random Individuals projection

4.3 Potential influencers based on hub dominance within communities

<u>Dominant hubs from the bipartite graphs</u>

On the entire bipartite graph, the Louvain algorithm found 34 communities of different sizes, as seen on figure 17, ranging from a size of 4 to 11099. As mentioned, the algorithm was run multiple times until reaching maximum modularity. The modularity returned was of 0.77, considered quite high. This suggests strong partitions between communities in the network, that is, that there significantly more edges within communities than between communities. This is indicative of meaningful communities as opposed to nodes groupings as an artifact of the partitioning process.

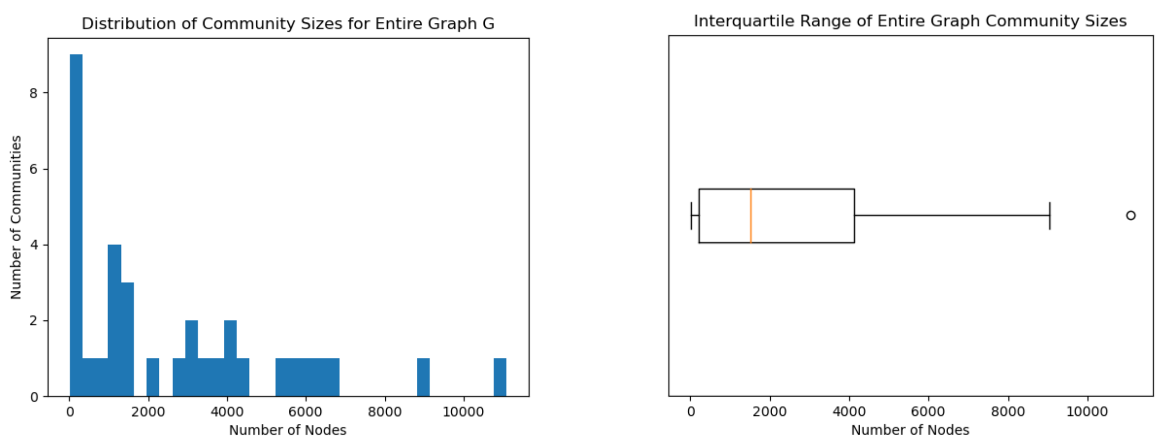


Figure 18. Community sizes of random bipartite graph

The nodes identified as the most dominant within their communities can be seen in table 8. Hub dominances are between 0 and 1, as expected. However, 23 out of 34 communities have a most dominant hub with a hub dominance of under 0.1, which is indicative of this measure being of mixed meaningfulness for identifying influencers. It is interesting to note that all the nodes with the highest hub dominance in the graph's communities are companies, not individuals. 50% of the dominant influencers are in the 90th to 100th percentiles of highest eigenvector values. Since the modularity of the community is high, the communities are meaningful, however, most the hub dominances may not be.

Community	Company with highest hub dominance	Hub dominance	Eigenvector centrality percentile
1	Deutsche Bank Aktiengesellschaft	0.038	100.00
2	PTT Public Company Limited	0.109	90.30
3	Westpac Banking Corporation	0.026	96.52
4	BNP Paribas SA	0.029	99.85
5	Walmart Inc.	0.015	97.66
6	Toyota Motor Corporation	0.015	95.98
7	Novartis AG	0.015	96.43
8	Commerzbank AG	0.03	97.71
9	State Bank of India	0.022	87.18
10	Orkla ASA	0.02	81.03
11	Cisco Systems, Inc.	0.023	97.67
12	Bank of China Limited	0.012	95.01
12	Hotai Motor Co.,Ltd.	0.714	26.12
13	Hotai Finance Co., Ltd.	0.714	2.43
13	UniCredit S.p.A.	0.058	98.20
14	Telecom Italia S.p.A.	0.058	96.54
15	Industrias Bachoco, S.A.B. de C.V.	0.857	15.35
16	Exelon Corporation	0.009	96.41
17	Brookfield Corporation	0.063	99.25
18	Anglo American Platinum Limited	0.061	92.50
19	Mari Petroleum Company Limited	0.049	31.74
20	Coca-Cola FEMSA, S.A.B. de C.V.	0.086	92.98
21	Blackstone Inc.	0.117	99.99
22	CapitaLand Limited	0.025	96.42
22	Gulfstream Natural Gas System, L.L.C.	0.667	9.52
23	Southeast Supply Header, LLC	0.667	9.52
24	First International Bank of Israel Ltd	0.062	81.37
25	HL Holdings Corporation	0.702	72.06
25	TOKAI Holdings Corporation	0.955	61.44
26	TOKAI Corp.	0.955	17.61
26	RPC, Inc.	0.438	33.78
27	Marine Products Corporation	0.438	33.78
28	Mega Financial Holding Co., Ltd.	0.028	30.33
29	Woori Financial Group Inc.	0.035	22.05
30	Banco do Brasil S.A.	0.029	94.06
31	Forterra, Inc.	0.191	40.64
31	Japan Investment Adviser Co., Ltd.	0.909	3.71
32	Japan Investment Adviser Co., Ltd., Investment Arm	0.909	3.70
33	Ayala Corporation	0.084	90.97
34	LifeVantage Corporation	0.5	34.29

Table 8. Eigenvector centralities of influencers (companies) identified by hub dominance within the entire graph

The highest hub dominance (0.955) is that of TOKAI Holdings Corporation, a Japanese company engaged in five main business segments: oil and gas, real estate, cable television, information and communication services, and water manufacture (Financial Times, 2023). This may be indicative of this company being tied to business leaders affiliated to a wide array of companies in its market. Since its eigenvector centrality is at 61.44 percentile, its influence might be strong in its market, and less so in the overall business network. This is reflective of a property of the Louvain algorithm and hub dominance which identify influencers within communities and do not reflect influencers that bridge communities together.

As of these findings, the meaningfulness of these influencers is uncertain, and will be further validated with statistical testing.

<u>Dominant hubs for the Companies projection</u>

16 communities were found on the Companies projection. The distribution of company communities' sizes was also rightly skewed, as observed on the original graph, but less evenly distributed and with outliers with larger sizes. The presence of larger outliers in the Companies projection could indicate the existence of influential companies that form major hubs or have more widespread connections. The relatively high modularity of 0.62 also reflects meaningful partitions, albeit less so than on the original graph.

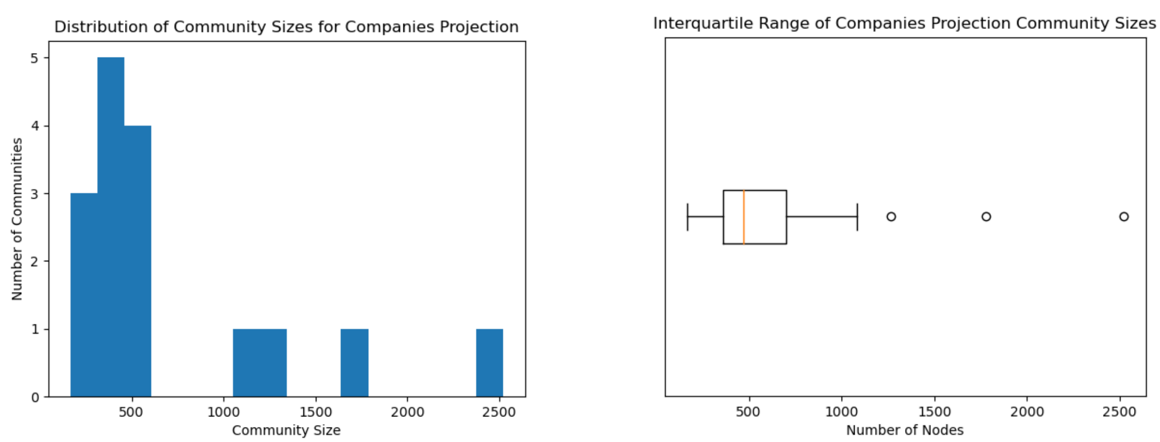


Figure 19. Community sizes of Companies projection

The hub dominance values of most important hubs in the projection are between 0.08 and 0.35. As hub dominance has been calculated as the extent to which a hub node is interconnected within its own community, this would indicate that the connections are more evenly distributed, and while there are influencers, they do not differ so much from the rest of the graph. This is validated by the high density of the projection and the similar findings of the 10 highest eigenvector centralities of the projection.

Community	Company with highest hub dominance	Hub dominance	Eigenvector centrality percentile
1	Morgan Stanley	0.132	99.98
2	BRF S.A.	0.192	86.09
3	Lloyds Banking Group plc	0.201	99.41
4	Deutsche Bank Aktiengesellschaft	0.15	99.99
5	Biogen Inc.	0.212	95.28
6	AUO Corporation	0.15	37.03
7	Bank Hapoalim B.M.	0.341	67.83
8	China Merchants Bank Co., Ltd.	0.093	85.18
9	ANZ Group Holdings Limited	0.224	96.62
10	Telia Company AB (publ)	0.245	92.34
11	Mitsubishi UFJ Financial Group, Inc.	0.085	96.31
12	Brookfield Corporation	0.231	98.2
13	ICICI Bank Limited	0.332	96.38
14	CIMB Group Holdings Berhad	0.137	87.24
15	Savola Group Company	0.156	62.39
16	POSCO Holdings Inc.	0.119	53.47

Table 9. Eigenvector centralities of influencers identified by hub dominance on the Companies projection

It must be noted while 75% of the dominant influencers are in the 90th to 100th percentiles of highest eigenvector values, only two companies have been identified as being the dominating influencers of their respective communities both for the communities created through the entire graph and through the Companies projection:

- Deutsche Bank Aktiengesellschaft, which returns hub domination in graph' community of 0.038 and in the Companies projection's community as 0.15.
- Brookfield Corporation which returns hub domination in graph' community of 0.063 and in the Companies projection's community as 0.231.

While their hub dominance is low, their presence as a dominating hub in both networks points to the importance of the node, as does their high eigenvector centrality. Indeed, these are nodes that might belong to clusters where their influence does not spread so strongly as other dominant hubs in other communities, but that will have a higher influence over the overall graph.

Dominant hubs for the Individuals projection

The communities for the Individuals projection have first been computed on individuals. Thereafter, the companies associated to the individuals were identified, for comparison of community sizes with the Companies projection, which had outliers with larger sizes. The distribution of the sizes is therefore displayed sizing the companies. Modularity in this case is also slightly higher than that of companies – 0.67. In this case, the distribution is also rightly skewed but more evenly so, due to individuals tending to cluster more consistently within communities or having diverse affiliations with various companies.

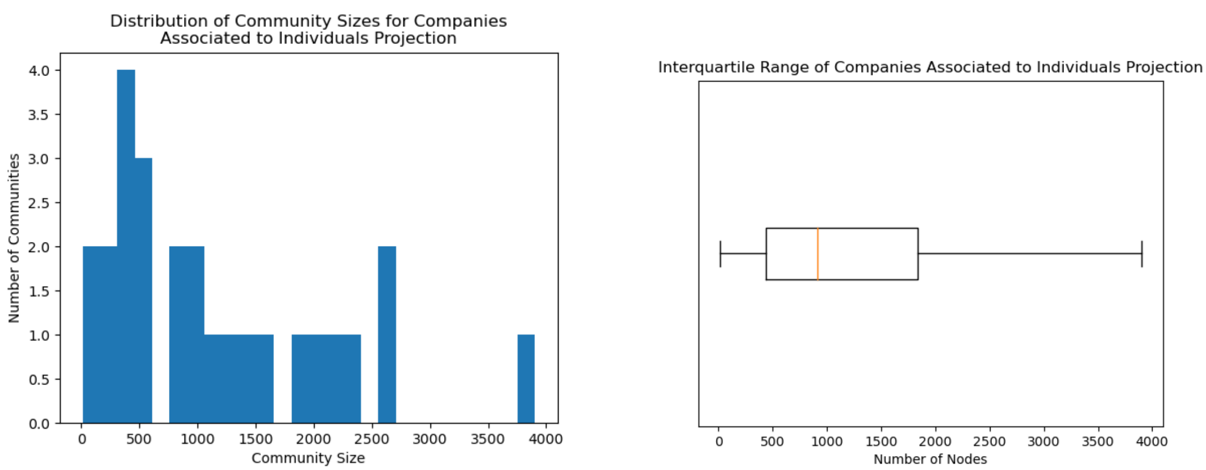


Figure 20. Community sizes of Individuals projection

Community	Individuals with highest hub dominance	Eigenvector centrality percentile
1	0.382	99.9
2	1.262	94.34
3	0.68	99.76
4	0.675	96.38
5	1.637	99.9
6	0.534	99.24
7	1.417	99.76
8	1.539	99.76

9	0.465	99.94
10	0.671	97.46
11	0.618	99.9
12	1.119	97.48
13	0.361	88.23
14	0.613	99.9
15	2.787	79.89
16	0.824	56.92
17	0.31	94.4
18	0.699	97.18
19	1.821	95.02
20	2.442	99.9
21	0.556	72.2
22	0.72	100
23	12.111	39.29
24	0.664	96.38
25	1.919	62.26
26	1.876	92.06

Table 10. Eigenvector centralities of influencers identified by hub dominance on the Individuals projection

The hub dominance of the Individuals projection denotes that some dominant influencers have a dominance that is orders of magnitude higher than the rest. Community 22 with its dominant hub with a dominance of 12.111 can illustrate this. A hub dominance of is exhibited when a node is connected to all other nodes, and a hub dominance over 1 suggests multi-link connections. A hub dominance of 12 would indicate that the node has many connections to the same other nodes. Community 22 is composed of 10 individuals that are associated to 164 companies. The dominant hub is an individual that works in two of those companies. While these are few companies, the number of edges might be explained by the fact that for one company, 75 individuals are part of the Community 22, and for the second company, there are 36. As those individuals have a presence in the other 10 companies represented by the community, it is reasonable to infer those two companies are strategically placed, which makes the community highly centralized around the dominant hub.

Hub dominance in the case of the Individuals' projection is therefore influenced by high transitivity, which has a high clustering effect around nodes. These results shed definitive light on the effect of transitivity of the Individuals' projection, and reveal that influence metrics are be overly affected by network structure and clustering, and cannot clearly define as to a node's importance. This is in line with the findings of Conyon and Muldoon, who found exaggerate clustering on their board members projections (Conyon *and al.*, 2004).

4.5 Mann-Whitney U Test

As presented in the last subsection, the potential groups of influencers that are build on projections present too heavy a bias as to be considered meaningful. Therefore, only groups of influencers based on the bipartite graph are considered for the Mann-Whitney U Test. The results from the Mann-Whitney U Test performed between the eigenvector centralities of companies within the bipartite graph and the eigenvector centralities on companies within the potential groups of hubs confirm that the groups are meaningful groups of hubs (table 11). The same can be said for the individuals (table 12). All results are deemed statistically significant (considering all *p*-values are below 0) and high absolute z-scores. The alternative hypothesis is therefore accepted for all groups of influencers:

$$H_1 = \textit{There is a difference between the eigenvector distribution of the bipartite graph and the eigenvector distribution of the group of influencers.}$$

	Mann-Whitney U Statistic	z score	p-value
Hubs with the 10 highest eigenvector centralities	99	-5.46	4.63E-08
Hubs with the 12 highest eigenvector centralities	173	-5.98	2.21E-09
Hubs with the highest eigenvector centralities, with cutoff at the knee	21543200	-38.11	0
Hubs with the highest degree, with cutoff at 2 Std Dev above the mean	6473002	-46.3	0
Hubs with the highest degree, with cutoff at xmin	381313.5	-31.02	2.48E-211
Hubs identified by random projections degree	2660435	-46.29	0
Hubs identified by random bipartite graph degree generated with equivalent density	14747300	-43.38	0
Hubs identified by random bipartite graph degree generated with equivalent probability	17808438	-41.01	0
Hubs identified by random unipartite graph degree generated with equivalent density	11115173	-44.99	0
Hubs identified by random unipartite graph degree generated with equivalent probability	13476610	-44.14	0
Hubs identified through hub dominance in communities based on the bipartite graph	145358	-3.9	9.63E-05
Hubs identified through hub dominance in communities based on the Companies projection	23471	-5.12507	2.97E-07

Table 11. Results from Mann-Whitney U test for the companies within the bipartite graph

	Mann-Whitney U Statistic	Z score	*p*-value
Hubs with the highest eigenvector centralities, with cutoff at the knee	739444	-10.83	2.52E-27
Hubs with the highest degree, with cutoff at 2 Std Dev above the mean	39854	-2.01	0.04
Hubs identified by random projections degree	141921518	-143.25	0
Hubs with the highest degree, with cutoff at xmin	9702030	-30.07	1.31E-198
Hubs identified by random projections degree	1.42E+08	-143.25	0
Hubs identified by random bipartite graph degree generated with equivalent density	128220	-4.09	4.36E-05
Hubs identified by random unipartite graph degree generated with equivalent density	43891.5	-2.52	0.01
Hubs identified by random bipartite graph degree generated with equivalent probability	314329	-7.03	2.13E-12
Hubs identified by random unipartite graph degree generated with equivalent probability	92125	-3.88	1.05E-04
Hubs identified through hub dominance in communities based on the Individuals projection	184364	-5.74	9.67E-09

Table 12. Results from Mann-Whitney U test for the individuals within the bipartite graph

Some remarks can be made about the results:

- The high z scores are possibly due to the elimination of hubs (highest eigenvector centralities) from the original distribution before comparison, as all groups of hubs contain at least the 10 largest centralities, leading to large differences between the distribution of the hubs and that of the original graph.
- Considering the large range of absolute z scores separately amongst companies and individuals, it is most possibly due to high difference in sample sizes, which makes comparison of differences between groups biased.
- However, the individuals have lower z scores than companies and larger sample sizes in most cases, indicating that there may be higher differences with the median of the distribution for the companies' hubs than the individuals.
- The lowest *p*-values belong to those distributions with the largest number of nodes. It is possible s those distributions are removes from the original eigenvector centrality distribution, the distributions' difference is clearer, and the test more significant.
- The negative z for all scores indicates that the original eigenvector distribution has a lower rank sum (or median) than the groups of hubs, as expected.

4.6 ESG Results

The Mann-Whitney test confirms that all groups of hubs are indeed meaningful as groups of influencers, therefore, the relationship between the eigenvector centralities and ESG scores will be evaluated for all groups.

4.6.1 Spearman correlation

A Spearman correlation was computed on the original bipartite graph, comparing the eigenvector centrality distribution to the distribution of ESG scores. The following correlation coefficients were found:

- For the bipartite graph – 0.345, with a *p*-value of 0, indicating a moderate correlation of statistical significance.
- For the companies within the bipartite graph – 0.4, with a *p*-value of 0, indicating a moderate correlation of statistical significance.
- For the individuals within the bipartite graph – 0.337, with a *p*-value of 0, indicating a moderate correlation of statistical significance.
- For the Companies projection – 0.418, with a *p*-value of 0, indicating a moderate correlation of statistical significance.
- For the Individuals projection – 0.335, with a *p*-value of 0, indicating a moderate correlation of statistical significance.

The test clearly indicates that as the eigenvector centrality increases, whether overall or solely for companies or individuals, so does the ESG score.

4.6.2 Visual Exploration

Visual explorations on different types of graphs confirm whether the ESG scores of those groups of influencers are higher than those of the rest of the entities.

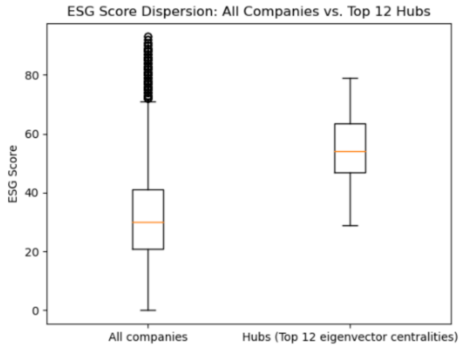


Figure 21: ESG distribution of hubs with the 12 highest eigenvector centralities

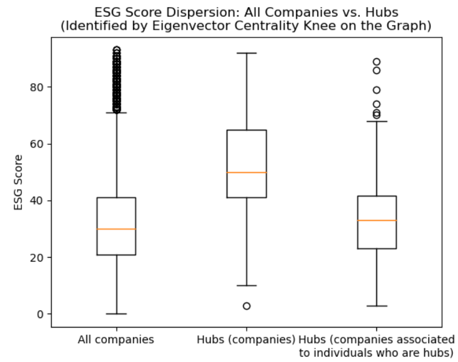


Figure 22: ESG distribution of hubs with the highest eigenvector centralities, with cutoff at the knee

1) The ESG distribution of the hubs with the 12 highest eigenvector centralities clearly indicates higher ESG scores than the original graph (figure 21).
2) The ESG distribution of hubs with the highest eigenvector centralities with cutoff at the knee indicates clearly that companies have higher ESG scores. However, the average ESG scores of the companies related to the individuals within that group follow the same distribution as the original graph. This result may be due to those individuals being affiliated to a variety of companies, in sufficient proportion and with sufficiently low scores as to skew the distribution.

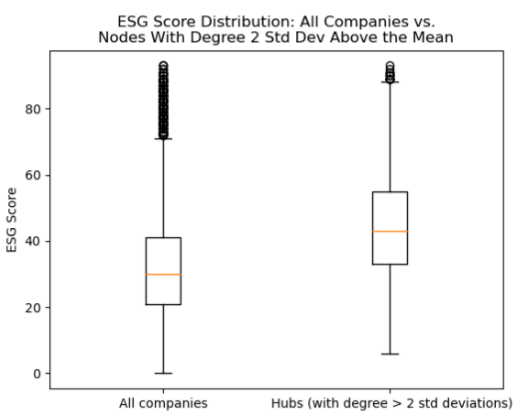


Figure 23: ESG distribution of hubs with the highest degree, with cutoff 2 standard deviations above the mean

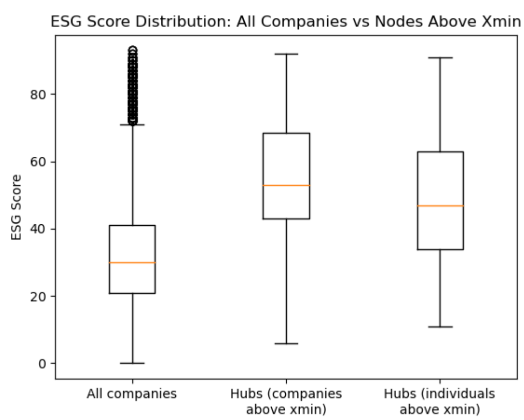


Figure 24: ESG distribution of hubs with the highest degree, with cutoff at xmin

3) The ESG distribution of hubs with the highest degree, with cutoff 2 standard deviations above the mean clearly indicates higher ESG scores than the original graph (figure 23).
4) The ESG distribution of hubs with the highest degree, with cutoff at xmin clearly indicates higher ESG scores than the original graph (figure 24), in the cases of both companies and individuals. As opposed to the individuals with the highest eigenvector centralities with cutoff at the knee, the individuals in this group have higher ESG scores, which may be due to smaller group size.

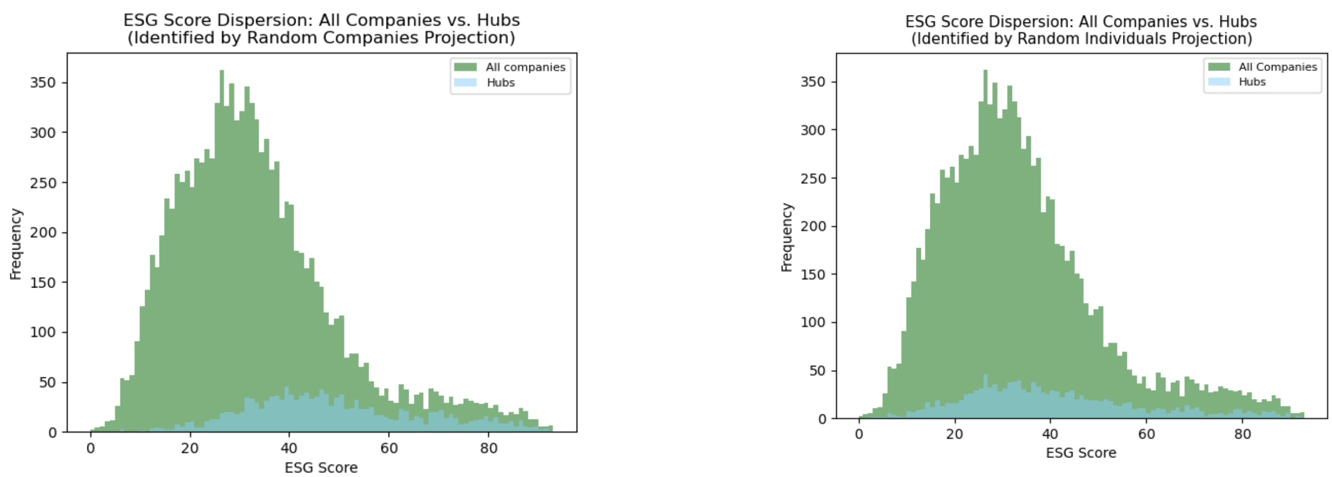


Figure 25. ESG distribution of hubs identified by random bipartite graph projections

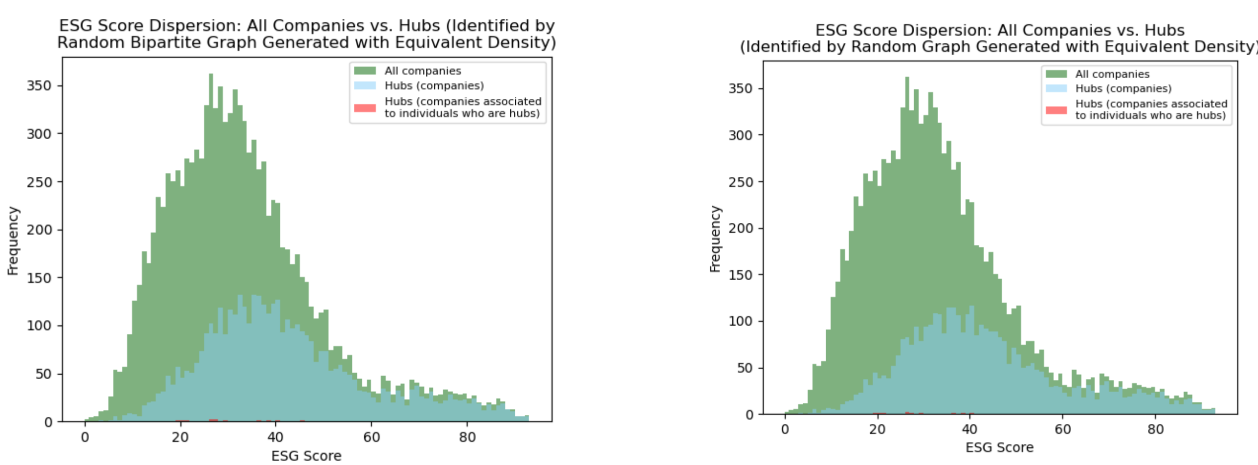
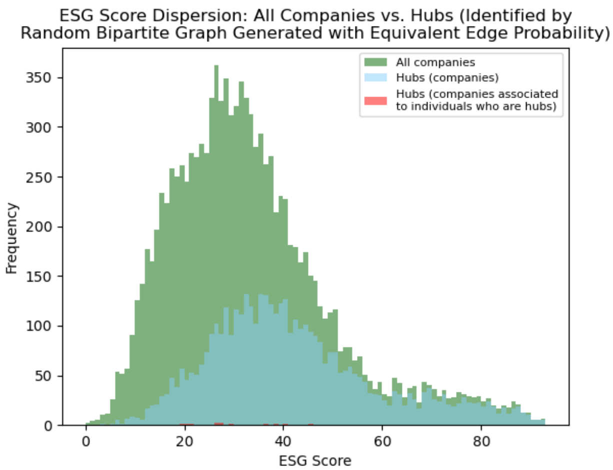


Figure 26. ESG distribution of hubs identified by random graph generated with equivalent density



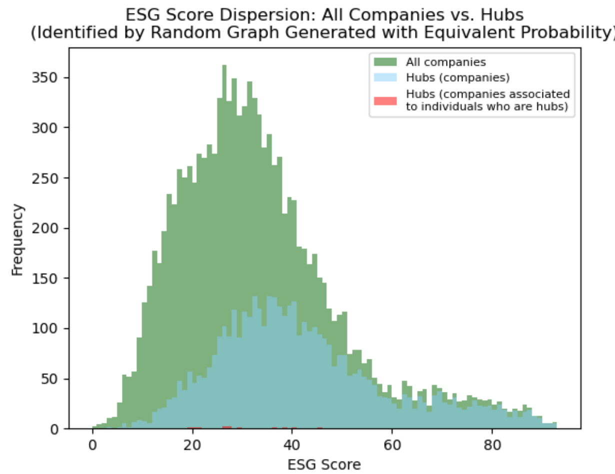


Figure 27. ESG distribution of hubs identified by random graph generated with equivalent edge probability

5) All ESG distributions for groups of hubs based on cutoffs identified by random graphs point to companies that are hubs having higher ESG scores than the overall graph. With respect to the individuals, on the projection, they have lower ESG scores, for similar reasons as the hubs with the highest eigenvector centralities with cutoff at the knee. As for the individuals on the bipartite graphs, not enough individuals are present to draw a conclusion.

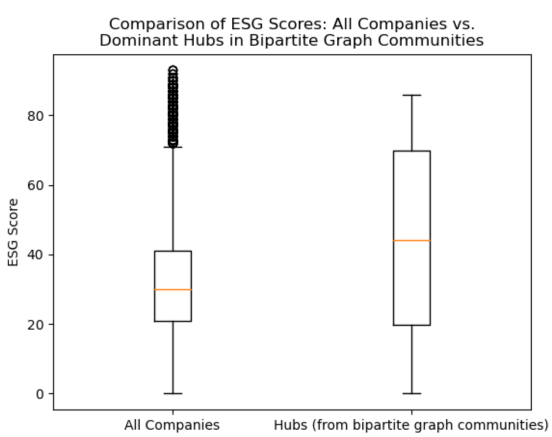


Figure 28: ESG distribution of dominant hubs within communities on the bipartite graph

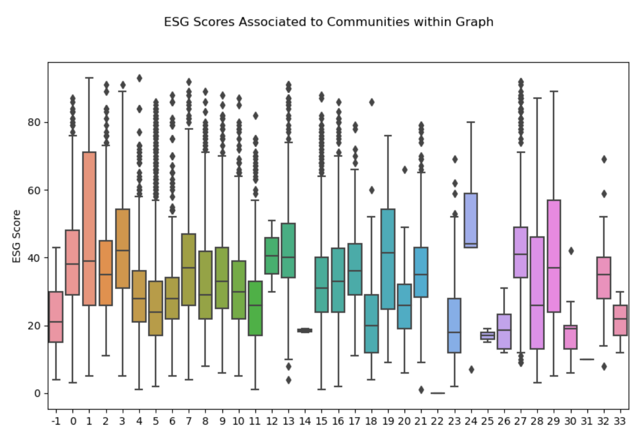


Figure 29: ESG distribution within the bipartite graph

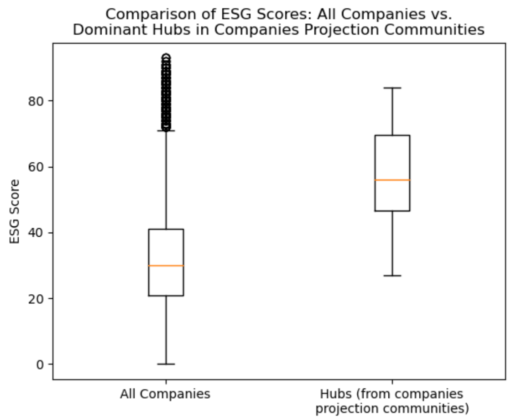


Figure 30: ESG distribution of dominant hubs within communities on the bipartite graph

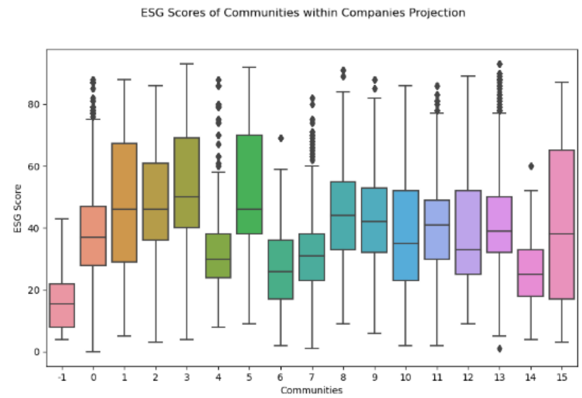


Figure 31: ESG distribution within the Companies projection

6) Most communities have a right skewed ESG distribution, similar to that of the original graph, whether for communities within the bipartite graph or the Companies projection. The hubs for on the bipartite graph do not have display a difference in ESG distribution than overall companies, whilst the ones within the Companies projection do. As the companies within the communities of the bipartite graph have been determined less biased due from transitivity, it can be inferred that hubs determined through communities will not necessarily display higher ESG scores than hubs identified on the overall graph.

Chapter 5: Conclusion

This paper has established that influencers who are vectors of influence within the overall business network indeed have higher ESG scores, specifically for companies. For the individuals, this is the case only for the ones with very highest degrees or eigenvector centrality. This has been determined by the Spearman correlation test and the ESG distributions for groups of hubs. Influencers within communities, however, do not have overall higher ESG scores. Further analysis on the structure of the communities is needed.

Another important insight from this paper is that that groups of influencers are better detected on the bipartite graph than on the graph's projections. Indeed, this study has found that due to induced links, transitivity, and assortativity, the structure of the bipartite projections has been heavily altered. This affects mostly groups of influencers that might be identified through eigenvector centrality of hub dominance. This is in line with the findings of other researchers who have also found bias in their business influencers' networks (Conyon *and al.*, 2004).

Multiple factors may have however impacted the scope of the work or results:

1) A main limitation of this study is that this paper tests only the most explicit influence an individual might have on a company – by sitting on the board or serving as a key executive. It does not test for influence outside of that explicit individual-company relationship, such as investments, shareholding, or any unofficial influence.
2) The lack of weight on the graph means that it does not hold the information on the strength of a relationship between a company and individual, which can affect a node's importance and potential influence.
3) The paper used ESG scores for its analysis. However, it is possible that this may introduce a bias within ESG correlation as some industry might have higher or lower average ESG scores than others, as well as specific influence within the network.

Future work on the subject may involve the following topics:

1) It would be interesting to study the positions of influence of financial industries within the network, as well as investment relationships to understand whether investments from influential financial companies with high ESG scores lead have an ESG impact on investees.
2) The use of historical data could offer insights into the before a relationship with an influential entity with a higher ESG score.
3) Propagation algorithms could be used to establish a causation effect between relationships with influential entities and ESG scores.
4) Exploring the nature of communities with statistical tests could offer more insights into the structure of the communities (indicative of whether they are based on geography, industry, or others) and further analysis of communities could test whether companies and individuals tend to cluster around similar ESG scores.

Bibliography

Albert, Réka, Hawoong Jeong and Albert-László Barabási (1999). « Diameter of the world-wide web », *Nature*, vol. 401, no 6749, p. 130-131.

Alstott, Jeff, Ed Bullmore and Dietmar Plenz (2014). « Powerlaw: A python package for analysis of heavy-tailed distributions », *PLoS ONE*, vol. 9, no 1, p. e85777.

Barabasi, Albert-László and Réka Albert (1999). « Emergence of scaling in random networks », *Science*, vol. 286, no 5439, p. 509-512.

Barabási, Albert-László and Pósfai Márton (2016). *Network science*, Cambridge University Press.

Boldi, Paolo and Sebastiano Vigna (2013). « Axioms for centrality », *arXiv pre-print server*.

Bonacich, Phillip (1987). « Power and centrality: A family of measures », *American journal of sociology*, vol. 92, no 5, p. 1170-1182.

Bonacich, Phillip (2007). « Some unique properties of eigenvector centrality », *Social Networks*, vol. 29, no 4, p. 555-564.

Borgatti, Stephen P (2005). « Centrality and network flow », *Social Networks*, vol. 27, no 1, p. 55-71.

Borgatti, Stephen P and Daniel S Halgin (2011). « Analyzing affiliation networks », *The Sage handbook of social network analysis*, vol. 1, p. 417-433.

Caporossi, Gilles and Eglantine Camby (2021). « Complex networks analysis ».

Conyon, Martin J. and Mark R. Muldoon (2004). « The small world network structure of boards of directors », *IO: Empirical Studies of Firms & Markets*.

Davis, Gerald F and Henrich R Greve (1997). « Corporate elite networks and governance changes in the 1980s », *American journal of sociology*, vol. 103, no 1, p. 1-37.

Diop, Issa Moussa, Chantal Cherifi, Cherif Diallo and Hocine Cherifi (2021). « Revealing the component structure of the world air transportation network », *Applied Network Science*, vol. 6, no 1.

Erdős, Paul and Alfréd Rényi (1960). « On the evolution of random graphs », *Publ. math. inst. hung. acad. sci*, vol. 5, no 1, p. 17-60.

Everett, M. G. and S. P. Borgatti (2013). « The dual-projection approach for two-mode networks », *Social Networks*, vol. 35, no 2, p. 204-210.

Financial Times (2023). « Tokai holdings corp. » https://markets.ft.com/data/equities/tearsheet/profile?s=3167:TYO.

Freeman, Linton C. (1978). « Centrality in social networks conceptual clarification », *Social Networks*, vol. 1, no 3, p. 215-239.

Governance & Accountability Institute, Inc. (2020). *90% of s&p 500 index® companies publish sustainability / responsibility reports in 2019*. Obtained on July 20th, 2023 https://www.ga-institute.com/ga-research-collection/sustainability-reporting-trends/2020-sp-500-flash-report.html

Guillaume, Jean-Loup and Matthieu Latapy (2004). « Bipartite structure of all complex networks », *Information Processing Letters*, vol. 90, no 5, p. 215-221.

Guillaume, Jean-Loup and Matthieu Latapy (2006). « Bipartite graphs as models of complex networks », *Physica A: Statistical Mechanics and its Applications*, vol. 371, no 2, p. 795-813.

Hadaj, Piotr, Dominik Strzałka, Marek Nowak, Małgorzata Łatka and Paweł Dymora (2022). « The use of plans and networkx in modeling power grid system failures », *Scientific Reports*, vol. 12, no 1, p. 17445.

Hagberg, Aric, Pieter Swart and Daniel Chult (2008). *Exploring network structure, dynamics, and function using networkx*,

Harjoto, Maretno Agus and Yan Wang (2020). « Board of directors network centrality and environmental, social and governance (esg) performance », *Corporate Governance: The International Journal of Business in Society*, vol. 20, no 6, p. 965-985.

Jeong, Hawoong, Sean P Mason, A-L Barabási and Zoltan N Oltvai (2001). « Lethality and centrality in protein networks », *Nature*, vol. 411, no 6833, p. 41-42.

Latapy, Matthieu, Clémence Magnien and Nathalie Del Vecchio (2008). « Basic notions for the analysis of large two-mode networks », *Social Networks*, vol. 30, no 1, p. 31-48.

Lizardo, Omar and Isaac Jilbert (2023). « Social networks: An introduction », *University of California, Los Angeles*.

Lohmann, Gabriele, Daniel S Margulies, Annette Horstmann, Burkhard Pleger, Joeran Lepsien, Dirk Goldhahn, *and al.* (2010). « Eigenvector centrality mapping for analyzing connectivity patterns in fmri data of the human brain », *PLoS ONE*, vol. 5, no 4, p. e10232.

Luthra, Raj , Fernando Salas, Abhishek Shirali and Edward Woodcock. (2022, November 29, 2022). Turning esg aspirations into results: The role of functional leaders, *McKinsey & Company*. https://www.mckinsey.com/capabilities/operations/our-insights/operations-blog/turning-esg-aspirations-into-results-the-role-of-functional-leaders

Mann, H. B. and D. R. Whitney (1947). « On a test of whether one of two random variables is stochastically larger than the other », *The Annals of Mathematical Statistics*, vol. 18, no 1, p. 50-60.

Mester, Attila, Andrei Pop, Bogdan-Eduard-Mădălin Mursa, Horea Greblă, Laura Dioşan and Camelia Chira (2021). « Network analysis based on important node selection and community detection », *Mathematics*, vol. 9, no 18, p. 2294.

Nacher, JC and T Akutsu (2011). « On the degree distribution of projected networks mapped from bipartite networks », *Physica A: Statistical Mechanics and its Applications*, vol. 390, no 23-24, p. 4636-4651.

Newman, M. (2010). *Networks: An introduction*, OUP Oxford.

Newman, M. E. J., D. J. Watts and S. H. Strogatz (2002). « Random graph models of social networks », *Proceedings of the National Academy of Sciences*, vol. 99, no suppl_1, p. 2566-2572.

Newman, Mark EJ (2003). « Mixing patterns in networks », *Physical Review E*, vol. 67, no 2, p. 026126.

Opsahl, Tore (2011). « Triadic closure in two-mode networks: Redefining the global and local clustering coefficients », *arXiv pre-print server*.

Ramaswamy, Vivek (2023). « A solution is in sight for the esg controversy », *Wall Street Journal, The*.

Raphaël, Tackx, Jean-Loup Guillaume and Fabien Tarissan (2015). *Revealing intricate properties of communities in the bipartite structure of online social networks*, vol. 2015,

Robins, Garry and Malcolm Alexander (2004). « Small worlds among interlocking directors: Network structure and distance in bipartite graphs », *Computational and Mathematical Organization Theory, suppl. Special Issue on Mathematical Representations and Models* ... vol. 10, no 1, p. 69-94.

S&P Global , Market Intelligence (2019). « Professional data », *Xpressfeed User Guide*.

S&P Global , Market Intelligence (2021). « S&p global esg scores », *Methodology Guide*.

Smith Judd, Mary (2023). *Gcs and esg practices, priorities and perspectives annual benchmark study*, ALM. Obtained from https://media2.mofo.com/v3/assets/blt5775cc69c999c255/blt48fdd314e5a67981/230621-mofo-gc-esg-report.pdf

Vasques Filho, Demival and Dion R. J. O'Neale (2018). « Degree distributions of bipartite networks and their projections », *Physical Review E*, vol. 98, no 2, p. 022307.

Vasques Filho, Demival and Dion R. J. O'Neale (2020). « Transitivity and degree assortativity explained: The bipartite structure of social networks », *Physical Review E*, vol. 101, no 5, p. 052305.

Wandelt, Sebastian, Xing Shi and Xiaoqian Sun (2020). « Complex network metrics: Can deep learning keep up with tailor-made reference algorithms? », *IEEE Access*, vol. PP, p. 1-1.

Wright, Joanna (2019). « Waters rankings 2019: Best alternative data provider: S&p global market intelligence », *WatersTechnology*.

Yang, K. C., B. Aronson, M. Odabas, Y. Y. Ahn and B. L. Perry (2022). « Comparing measures of centrality in bipartite patient-prescriber networks: A study of drug seeking for opioid analgesics », *PLoS ONE*, vol. 17, no 8, p. e0273569.

Appendices

Appendix A: Data collection methodology for public data for S&P's "Professional Data" package

Sources of Data

To ensure a comprehensive, consolidated Professional data set, S&P Capital IQ uses a wide variety of sources and techniques to collect its Professional universe:

U.S. and Canadian Companies

- ***Regulatory Filings:*** Forms 8-K, 10-Q, 10-K, DEF 14A, 40-F, 20F, quarterly and annual reports from various regulatory organizations, and Prospectus forms. These agencies include the U.S. Securities and Exchange Commission (SEC) via EDGAR, Canadian Securities Administrators (CSA) via SEDAR, Australian Securities & Investments Commission (ASIC) via ASX, and the Regulatory News Service (RNS), which is owned by the London Stock Exchange for U.K. companies.

- ***News Aggregators:*** S&P Capital IQ follows a variety of Newswires (e.g., Nexis®, Factiva Select, PE Week Wire, Globes online, AltAssets, and FSA Register), searching for news articles, press releases, and corporate announcements, in English as well as in select local languages.

- ***Company websites:*** Public, private, private equity, venture capital, investment firms and advisor firms are profiled directly from the company's own website. Changes on the professional and board pages on the company's websites are captured through web-monitoring.

- ***Company Surveys:*** S&P Capital IQ is regularly in touch with public companies as a means of gathering information.

- ***Key Developments:*** S&P Capital IQ's Executive Change Key Developments data serves as a source for interim data processing. As the Key Developments team scans thousands of sources daily, it passes all relevant executive change announcements on to the Professional data set team.

Table A. Sources of data for public data for S&P's "Professional Data" package (S&P Global 2019 : 11)