

Big Data Analytics of Corporate Internet Disclosures

Mohamed A. K. Basuony
Associate Professor of Accounting
Department of Accounting, School of Business
American University in Cairo
Cairo
Egypt

Ehab K. A. Mohamed
Professor of Accounting
Department of Accounting and Finance,
Faculty of Management Technology
German University in Cairo
Cairo
Egypt

Ahmed Elragal
Professor of Information Systems
Department of Computer Science, Electrical and Space Engineering
Luleå University of Technology
Luleå
Sweden

Khaled Hussainey
Accounting and Financial Management
Faculty of Business and Law,
Portsmouth Business School
University of Portsmouth
Portsmouth
UK

Biographical Details:

Mohamed A. K. Basuony is Associate Professor of Accounting at the School of Business at the American University in Cairo. Prior to that, he worked at the German University in Cairo and Faculty of Commerce, Ain Shams University. He received both his Bachelor with honours and Master Degree from the same university. He obtained his Ph.D. degree from Brunel University in the UK. His work experience includes twenty years teaching undergraduates as well as post-graduates in Egypt and the UK. His areas of research are in performance management, balanced scorecard, management and strategic control, corporate governance, and corporate social responsibility. He has published a number of papers in international refereed journals and presented papers at numerous international conferences; he has also published two books.

Ehab K. A. Mohamed is the Vice President for Students Affairs at the German University in Cairo. He is also the Dean and Professor of Accounting at the Faculty of Management Technology. Prior to joining GUC he worked for 10 years at Sultan Qaboos University, Oman. He graduated from Cairo University and received his M.Sc. & Ph.D. from Cass Business School, London. He is a Fellow of the Chartered Institute of Internal Auditors, UK. His areas of research are in auditing, fraud, performance measurement, business education, financial reporting, and corporate governance. He has published a number of papers in international refereed journals and presented papers at numerous international conferences; he has also published four books.

Ahmed Elragal is an Information Systems Professor at Luleå University of Technology. In 2001, he has obtained his PhD in Decision Support Systems (DSS) from the University of Plymouth, UK. His research interests include Big Data [Analytics]; Mobility Mining; & Enterprise Systems. He has over 50 research articles published in international outlets. He is a member of the editorial board of the I & M; IJBIR; and Associate Editor of the IJISPM. He has over 10 years of consulting experience where he has consulted leading multinationals including SAP and Teradata.

Khaled Hussainey is Professor of Accounting and Financial Management in Portsmouth Business School. He completed his Ph.D. in Accounting and Finance at the University of Manchester in 2004, and since then he has held academic positions at Plymouth University; Stirling University, Manchester University, and Ain Shams University, before joining the University of Portsmouth as a Professor of Accounting and Financial Management in August 2016. He has published more than 100 refereed papers in academic journals and international conferences proceedings. he has been awarded the prestigious 2007 Best Paper Award of the British Accounting Review for his paper "Loss firms' annual report narratives and share price anticipation of earnings" & the prestigious 2012 Best Paper Award of the Journal of Risk Finance for my paper "Revisiting the capital structure puzzle: UK evidence". He is currently a Co-Editor in-Chief of Journal of Financial Reporting and Accounting; an Associate Editor for Journal of Applied Accounting Research and International Journal of Accounting, Auditing and Performance Evaluation.

Abstract:

Purpose - This study investigates the extent and characteristics of corporate Internet disclosure via companies' websites as well via social media and networks sites in the four leading English speaking stock markets, namely Australia, Canada, the United Kingdom, and the United States.

Design/methodology/approach – A disclosure index comprising of a set of items that encompass two facets of online disclosure, namely company websites and social media sites, is used. The paper adopts a data science approach to investigate corporate Internet disclosure practices among top listed firms in Australia, Canada, UK, and USA.

Findings – The results reveal the underlying relations between the determining factors of corporate disclosure, i.e. profitability, leverage, liquidity, and firm size. Profitability in its own has no great effect on the degree of corporate Internet disclosure whether via company websites or social media sites. Liquidity has an impact on the degree of disclosure. Firm size and leverage appear to be the most important factors driving better disclosure via social media. American companies tend to be on the cutting edge of technology when it comes to corporate disclosure.

Practical implications - The paper provides new insights into corporate Internet disclosure that will benefit all stakeholders with an interest in corporate reporting. Social media is an influential means of communication that can enable corporate office to get instant feedback enhancing their decision making process.

Originality/value – To the best of our knowledge, this study is amongst few studies of corporate disclosure via social media platforms. We have adopted disclosure index incorporating social media as well as applying data science approach in disclosure in an attempt to unfold how accounting could benefit from data science techniques.

Keywords: corporate disclosure, social media, big data, data science, Internet, Australia, Canada, UK, USA

Paper type: Research paper

1. Introduction

Al-Htaybat and von Alberti-Alhtaybat (2017:870) argue, “Accountants must engage with different parts of the organisation and must be jointly proactive about Big Data and corporate reporting. As they are to provide significant amount of data to stakeholders and likewise are to collect and analyse Big Data, accountants need to engage with data scientists to jointly have the most meaningfully output”. They also argue that there is a need for a “... significant further empirical analyses of Big Data analytics currently or potentially in use in accounting in general and corporate reporting in particular, either through a detailed case study or through a quantitative survey to gather a broader range of insights’. Disclosure literature focuses only on corporate websites. However little literature examines disclosure practices via social media sites (SMS). One of these very few and latest studies is the study of Jung et al. (2018) which uses the social media platform Twitter as a setting to provide early evidence that firms are strategic in how they disseminate earnings news.

Nowadays, it is seldom to open any popular publication, online or in the physical world, and not run into a reference to data science, analytics, big data, or some combination thereof. Big data is characterized by three main qualities: volume, variety, and velocity. The volume of the data is mainly its size. Velocity refers to the proportion with which data is varying. Lastly, variety embraces the different formats and types of data, as well as the different kinds of uses and ways of analyzing the data. One of the reasons making big data really big is that it is coming from a greater variety of sources than ever before, including sensory data, IoT, logs, clickstreams, and social media. Using these sources for analytics means that common structured data is now joined by unstructured data, such as text and human language, and semi-structured data, such as eXtensible Markup Language (XML), JSON or Rich Site Summary (RSS) feeds. Therefore, variety is just as big as volume. Added to that, big data can be described by its velocity or speed. This is basically the frequency of data generation or the frequency of the data to be delivered. The leading edge of big data is streaming data, which is collected in real-time from the websites. Some have discussed the addition of “veracity” to the definition of big data. Veracity focuses on the quality of the data. This characterizes big data quality as good, bad, or undefined due to data inconsistency, incompleteness, ambiguity, latency, deception, and approximations (Elragal & Klischewski, 2017).

This motivates us to use data science analytics to understand the determinants of corporate disclosure practice in the largest stock markets in four English-speaking countries, namely Australia, Canada, the United Kingdom (UK), and the United States of America (USA). In addition to Corporate Internet Disclosures (CID), we also investigate corporate disclosure via social media and networks. To achieve this objective a disclosure index comprising of a set of items that capture the extent of online information disclosure is adopted and used based on that developed by Mohamed et al. (2017). These items encompass two facets of online disclosure, namely disclosure via companies own websites and disclosure via social media sites. The paper is structured as follows. Section 2 reviews the literature. Section 3 discusses the methodology.

Section 4 provides analyses and discussions on the nature and characteristics of CID by companies listed in the four stock markets. Section 5 concludes.

2. Literature Review and theoretical framework

The theoretical justifications for companies engaging in corporate Internet disclosure can be explained by the agency theory where corporate disclosure is considered as a mean to controlling the agency costs arising from conflicts of interests between insiders and outside shareholders. Voluntary disclosure can reduce the agency costs by facilitating debt supplier's assessment of a firm's ability to meet its debts (Jensen and Meckling, 1976). Therefore, it is suggested that firms that are highly leveraged are motivated to try and satisfy debt suppliers by disclosing reliable information on their website to make creditors more confident about the companies' ability to pay their debts. Though this disclosure results in extra costs, disseminating reliable information to debt suppliers reduces agency costs. Similarly, shareholders demand more information to assess the companies' performance (Oyelere et al., 2003; Xiao et al., 2004). Also, according to the agency theory managers of highly profitable companies disclose more information on the companies' website to attain personal benefits such as the maintaining their positions and justifying compensations (Wallace et al., 1994; Haniffa and Cooke, 2002). Furthermore, agency theory infers that large firms exhibit higher agency costs due to the information asymmetry between market participants (Jensen and Meckling, 1976). To reduce these agency costs, larger firms disclose a large flow of corporate information. Large corporations can disclose information on the Internet for lower costs as they have the resources to do so.

On the other hand, signalling theory suggests that profitable companies have an incentive to disclose more information, to signal the companies' profitability to investors and to raise capital at the lowest price (Oyelere et al., 2003; Marston and Poley, 2004). The signalling theory further suggests that companies within the same industry tend to adopt the same level of disclosure. When a company within an industry tends not to follow the same disclosure practices, including internet disclosures, as others in the same industry, then it may be interpreted as a signal that the company is hiding bad news (Craven and Marston, 1999). Moreover, according to the signalling theory managers who hire large auditing firms signal to the market that they are willing to provide quality disclosures (Healy and Palepu, 2001). International audit firms are more likely to facilitate the diffusion of innovative practices, such as the internet financial disclosure (Hail, 2002; Xiao et al., 2004). Whereas, according to the political cost hypothesis, large companies interest more financial analysts placing companies under greater pressure because they are more publicly visible (Boubaker et al., 2012).

The widely used theories at the organizational level for the usage of social media are the Technology-Organization-Environment framework, which serves as a foundation to identify the determinants of social media usage (see for example, Srivastava and Teo, 2009), and the Resource-Based View theory (Zhu and Kraemer, 2005), which is considered a foundation for linking social media usage and value. Our study focuses on the adoption of the concept of Enterprise Generated Content, which reflects the usage of social media at the organizational level by using only the framework.

3. Methodology

The sample consists of the top listed in the stock exchanges of the Australia, Canada, UK and U|SA; namely ASX400 for Australia, TSX300 for Canada, FTSE350 for the UK, and S&P500 for the USA. Manufacturing companies represent 40% of the total sample and the service companies 60%.

The items in the disclosure index cover the presentation format, general content, and financial content. The presentation format aspect is important because it promotes the understandability of information and improves the timeliness of disclosure of information (Debreceeny et al., 2002). The Total Disclosure Index (TDI) is divided into two sub-indices: (1) Total Website Disclosure Index (TWDI): 39 items covering disclosure via companies own websites, (2) Social Network and Media Disclosure Index (SNMDI): 37 items covering disclosure via social networks and media websites. This index utilizes a dichotomous scoring approach to capture the level of disclosure. An item is scored 1 if it is disclosed and 0 if otherwise. The disclosure index for each company was calculated by dividing the actual scores awarded by the maximum possible scores appropriate for the company. This was applied to the total disclosure index (TDI) and all its sub-indices. Table 1 summarizes the definition and measurement of variables included in this study. The first group consists of 9 indices of corporate disclosure, and the second group consists of the determinants of corporate disclosure and the last group includes the stock market as a control variable.

The data collection process starts by checking the existence of an official website for each company, then moves to the next stage that requires the collection of data from the official corporate website by searching for the information required. Starting with the 39 website attributes that are classified under three categories of information, namely: format, general content, and financial content. Moving to the 37 attributes that are related to social media and networks, we search for the icons related to the 10 social media and networks platforms. If the company has an account on one of these social media and networks sites, then we examine the nature of information disclosed and categorize the information under three categories, namely: non-financial information, financial information, and annual reports. The internet disclosure data were collected in the period between March and May 2015. The data for internet disclosure were collected by six research assistant divided into two groups to check and validate the accuracy and quality of data collected. A further check was carried out by the three main researchers (authors). Data for the determinants of CID are collected from the ORBIS database. This approach results in the collection of 82 observations for each company, a total of 122,180 observations for the whole sample of the study.

Insert table 1 here

A big data analytics approach i.e., a data science approach, is adopted in this paper. Data science is the study of the generalizable extraction of knowledge from data. Data science is a new method that is nowadays available for knowledge extraction and theory development, which was not previously practical due to the scarcity of data. Big data analytics techniques e.g., clustering, associations, etc. have made it feasible for a machine to ask and validate interesting questions humans might have not ever considered (Dhar, 2013). The notion of analytics has been around for a long time which dates back to as old as the field of statistics in the 18th century. However, there are still substantial noticeable differences (Agrawal and Dhar, 2014). Amongst them is the rapid pace at which economic as well as social transactions are

carried out. Added to that, the ability of the human race to deeply understand the data set available has expanded by many folds. Furthermore, access to inexpensive computational resources including analytical tools has helped to give boost to data science as well as paved the road for democratization of data science as a field.

Grouping similar records into separate, or overlapping, clusters is one of the fundamental tasks in data science. In cluster analysis, the purpose is to group homogeneous (similar) objects in the same cluster and for heterogeneous (dissimilar) objects in another different cluster. Simply defined, clustering is classifying unclassified data. Depending on the type of cases, clustering can be executed using central or pairwise techniques. Central clustering minimizes the average distance between an observation and its cluster centroid. Hence, clustering can be described by means of cluster centroids. Clustering can be used to achieve both objectives; description and prediction. In our case, we have used K-means, central clustering, in order to cluster the 200 companies into separate clusters. Mixed Euclidean measure was adopted. Results show that the companies fall into 4 clusters with different themes and focus. The k-means algorithm is one of the most frequently used central clustering techniques. Data is divided iteratively into k clusters by minimizing average squared Euclidean Distance between the cases, or observations, and its cluster centre.

In our analysis, we adopt K-Means clustering approach using RapidMiner as a tool; this tool is commonly used in the data science literature (Provot and Fawcett, 2013; Ristiki et al. 2015). We use RapidMiner as analytics tool, which is equipped with an integrated analytical environment for machine learning, data mining, text mining and predictive analytics. It is used for business and industrial applications as well as for research, education, training, rapid prototyping, and application development. RapidMiner supports end-to-end data mining process lifecycle. RapidMiner is a member of Gartner's Magic quadrant, 2015 report, for advanced analytics platforms. It is open source and that has been another reason to use it in our analysis.

4. Finding and Analysis

4.1 Descriptive Analysis

Table 2 shows that twitter is the dominant social network platform for all items closely followed by Facebook. For general information, the first rank is Twitter (48.1%) followed by Facebook (42.4%) and then in the third rank is (41.4%). As for financial information and annual report, Twitter is the most dominant out of the five platforms, and then the other two platforms LinkedIn and Facebook followed Twitter. The platform that shows relatively insignificant percentage is Instagram. It has a (1.1%) for financial information and (0.9%) for annual report. Furthermore, it has (7.9%) and (9.9%) for general information and usage respectively, which makes Instagram the least effective amongst the five platforms.

Insert table 2 here

Table 2 shows that Youtube has the highest percentages compared to the other three social media platforms. Youtube has (35.2%) for general information and (37.1%) for usage and (5.9%) and (3%) respectively for financial information and annual report. On the other hand,

Flicker shows the least percentage of (0.4%) for financial information and (0.2%) for annual report. The table also shows that the two most common social networks over the four countries are Twitter and Facebook followed by LinkedIn and Google+. Moreover, the lowest percentage of usage comes in Instagram, which even has a 3.1% of usage in UK. It also shows that a country like USA shows a widespread of using social networks; Twitter is the most common in USA for usage with a percentage of 76.4 followed by Facebook with 74.3 while the opposite for general information, where Facebook is ranked first with 72.3 and followed by Twitter with 67.1. For financial information, Twitter in UK is the most platform using financial information with 29.5% followed by LinkedIn in Australia with 15.5%. Moreover, a zero percent in Instagram for financial information for both Canada and Australia. Twitter is ranked first in Australia with 33.8% then LinkedIn with 33.5% and followed by Google+ with 24.1% for the usage. Canada is using Twitter and LinkedIn with 39.9% for both of them followed by Facebook with 31.5%. Instagram is the lowest one among the other four platforms in the usage, general and financial information for all four countries.

The percentage of annual reports (AR) reflects the corporates' financial disclosure. Australia is the first country amid the other three countries in using Twitter to disclose AR with a percentage of 12.3%, followed by UK with 11.5% and then Canada with 11.4%. USA is the only country among the four countries that disclose AR in its Instagram with 2.6% while other three countries have zero%. USA is coming first in disclosing AR in Facebook with 9.2% followed by Canada and UK with 7.6% and 5.1% respectively and the lowest one is Australia with 3.8%.

Table 2 shows that Youtube is the most prevailing platform over the other three platforms in all four countries to disclose general information and usage, with more than average percentages in USA followed by UK and Canada with 30% and 26.5% respectively and then Australia at the end with 20.3%. Australia does not use at all any of Flickr and Slideshare to disclose general and financial information or any AR. All four countries either very rarely use the four social media platforms to disclose their annual report (AR) or not at all. For example, the only social media platform in UK that discloses information about AR is YouTube with 2.3% while other three platforms are zero %. Moreover, USA is the only country among the four countries that all social media platforms disclose AR information where YouTube is the first with 4.6% followed by Blogs and Slideshare with 1.8% for both of them and then Flickr comes at the end with 0.4%.

4.2 Data Science Analysis

This section provides the results and analysis of applying data science approach to the data collected on the extent, nature, and characteristics of CID by the companies listed on the four stock indices.

Clusters Membership

Table 3 and figures 1 & 2 show the clustering analysis for the various corporate disclosure indices. As it can be seen the sample of 1,481 companies are distributed over 3 clusters (cluster 0, cluster 1, and cluster 2) where cluster 0 has the highest number of constituents of 835 companies representing 56% of the total sample. On the other hand, cluster 2 only represents 1% of the sample consisting of 19 companies. The clustering is based on the *k-means* algorithm with each object assigned to precisely one of a set of clusters. Thus, each cluster groups together objects that are similar to each other and dissimilar to the objects belonging to other clusters. This grouping is based on the attributes represented in the form of nine various disclosure indices and sub-indices.

Table 3 shows that most of the clusters include companies that have relatively high degree of disclosure via companies' own website. Disclosure of financial content is relatively high across all three clusters. Meanwhile, the highest level of disclosure via SMS is in cluster 0. It is clearly apparent that high score in the format disclosure index is associated with high disclosure via social media. It appears that companies treat the various social media platforms as a format of disclosure regardless of the content of what is being disclosed.

Insert table 3 here

Insert figures 1&2 here

Looking thoroughly at the characteristics of the three clusters it is evident that there is sharp contrast between cluster 0 and cluster 2. Cluster 0 is characterized by firms showing reasonably high degree of disclosure across all the various disclosure indices when compared the other clusters. The cluster particularly shows the highest degree of adopting social media platforms for disclosure purposes among the 3 clusters. On the contrary, cluster 2 has the lowest degree of utilizing social media for disclosure; though this cluster reveals reasonable disclosure practices among member companies via their own websites. Nonetheless, companies in cluster 0 seem to be utilizing their own websites as well SMS for various disclosure purposes and in a more varying formats than companies in clusters 1 & 2.

Table 3 and figure 3 show the cluster membership by country and industry type. The results reveal that the majority (90%) of US companies are located in cluster 0, a cluster that is characterized by high corporate website disclosure and reasonable disclosure via social media. On a sharp contrast, the majority (69%) of Australian companies, 52% of UK Companies and 50% of Canadian companies are in cluster 1, a cluster that show high corporate website disclosure that is mainly caused by low degree of Internet financial disclosure, however companies in this cluster exhibit relatively low level of disclosure via social media. Significantly, only one of the US companies is grouped in cluster 2, revealing that the vast majority of US companies tend to have high degree of disclosure via their own websites. This result is further explained by the fact that the highest proportion of members in cluster 0, the cluster showing good usage of social media platforms for disclosure purposes, are from the USA. This leads us to conclude that companies in the USA are usually early adopters of technology. Similarly, Canada has second highest proportionate representation in cluster 0 and the second lowest proportionate representation in cluster 2. These results illustrate that companies in the UK share a high level of corporate disclosure via their websites, while only 46% of companies in the UK show good degree of disclosure via SMS. Meanwhile, Australian companies lag behind in terms of adopting social media platforms for corporate disclosure with only 29% of the companies grouped in cluster 0.

Insert figure 3 here

These results are consistent with the results of the World Economic Forum (2016), where the USA is ranked fifth in the world in terms of networked readiness index, particularly in the area of big data and social media; the report ranks the four countries as follows: USA, UK, Canada, and Australia. More precisely, Sweden is ranked top of the world in terms of business usage of technology; the ranking of the four countries is USA, UK, Australia, and Canada (www.weforum.org/gitr). A further explanation is provided by the results of the We Are Social (2016) report on the global digital and social landscape; the report reveals that out of the four countries included in our study the USA and UK have the highest social media penetration (59%) followed closely by the Canada and Australia (58%) (www.wearesocial.sg). More importantly, the SEC has taken various initiatives to encourage companies listed in the USA to utilize social media platforms for corporate disclosure.

Furthermore, table 3 reveals that 59% of the service companies are grouped in cluster 0, while 52% of the manufacturing companies are grouped in the same cluster. Therefore, the majority of companies in both sectors appear to adopt social media for disclosure purposes. Nonetheless, companies in both sector show good disclosure via companies own websites. This is consistent with some of the finding in the general literature on CID (Debreceeny and Rahman, 2005; Mohamed and Basuony, 2014).

Clusters Profile

A further analysis is carried out to understand the common profile of companies included in each of the clusters. To do so, a number of variables are inserted against each of the companies in each of the clusters. These variables represent some of the determinants of CID from previous studies as discussed in the literature review section. Figures 4 and 5 show the profile of each cluster based on profitability, leverage, liquidity and firm size. The figures also illustrate the profile of companies in each of the clusters per country. As it can be seen, companies from the USA represents almost half (54%) of the members in clusters 0; this number represents 90% of the US companies listed on the S&P500. The remaining companies are grouped in cluster 1 (10%) and cluster 2 (1%). Examining the results thoroughly reveal that the US companies in cluster 0 are characterized by large firm size and relatively higher leverage than companies grouped in clusters 1 and 2. Hence, larger firms with higher degree of leverage seems to be disclosing more information via social media. The only firm that is grouped in cluster 2 shows significantly higher liquidity than those in cluster 0 and 1.

Meanwhile, 19% of members in cluster 0 are operating in the UK, with the rest of members in the cluster almost equally distributed over Australia (14%) and Canada (13%). Similar to those in the USA, companies that disclose more via social media in the UK, Canada and Australia are characterized by large firm size and higher leverage than companies from the same countries that are grouped in clusters 1 and 2. This result is consistent with the findings with previous studies (Debreceeny et al., 2002; Oyelere et al., 2003; Mohamed and Basuony, 2014).

The results can be justified by the fact that to reduce these agency costs, larger firms disclose a large flow of corporate information (Jensen and Meckling, 1976). In addition, large companies are more likely to use information technology than small ones to improve financial reporting to meet the greater demand for information (Xiao et al., 1996). The relative cost of

disclosing information may also be lower than smaller ones (Oyelere et al., 2003). Furthermore, according to the political cost hypothesis it is argued that larger firms face more demands for information from analysts than smaller firms (Boubaker et al., 2012, Basuony et al., 2014). Moreover, the results are consistent with the agency theory that voluntary disclosure can reduce the agency costs by facilitating debt supplier's assessment of a firm's ability to meet its debts (Jensen and Meckling, 1976).

Almost 53% of members in cluster 2 are operating in the Australia; this cluster shows the lowest level of disclosure via social media, though companies in this cluster have a good disclosure levels via companies' websites. Looking carefully at the characteristics of companies in cluster 2, it is evident that those firms are relatively small firms with high liquidity. Moreover, looking at the profile of companies in cluster 1, the cluster that is characterized by low degree of SM disclosure, it is evident that companies that have high liquidity tend to disclose less via social media in all countries. Though, these companies adopt reasonably high degree of disclosure via their own websites. According to the signalling theory it could be argued that companies with poor liquidity might have incentives to disclose more information to justify this unfavourable outcome to shareholders (Wallace et al., 1994). Nonetheless, the results provide new insight into the relationship between liquidity and online corporate disclosure as previous studies failed to find any significant association between liquidity and disclosure (Barako et al., 2006).

Insert figures 4&5 here

Meanwhile, ROA does not seem to have a significant impact in the allocation of companies over the four clusters in any of the four countries. This is consistent of the findings of previous studies (Oyelere et al., 2003; Marston and Polei, 2004; Xiao et al., 2004). However, companies that have negative ROA are only clustered in cluster 2. The results reveal that companies with negative ROA are unlikely to have a very high degree of disclosure via social media like those companies grouped in cluster 0, or have a very low degree of disclosure via social media like their counterparts in cluster 1. Those firms however, have disclosure practices via their own websites that do not differ significantly from form with better ROA.

The results above provide valuable new insights into the disclosure patterns of the companies listed in the four countries and the degree at which they utilize their own websites as well as social media sites. More importantly, the results reveal the underlying relations between the determining factors of corporate disclosure, i.e. profitability, leverage, liquidity, and firm size. These underlying relations between the 'independent' variables are not usually revealed in this multi-dimensional manner using traditional analysis such as regression as most of the previous studies on CID use. Previous studies on CID that mainly use regression as means of analysis offer inconclusive evidence on the relationship between the various variables and CID (Oyelere et al., 2003; Marston and Polei, 2004; Xiao et al., 2004).

The results also reveal stark differences between the four countries in the disclosure patterns. Companies in the USA appear to be early adopters of new information technologies in disclosing information to their stakeholders. In all four countries, the financial content disclosed via social media still lag behind what is disclosed via companies' websites. Social media is still seen more as a marketing tool, with companies using social media platforms to promote their product & services rather to disseminate information to their stakeholders. However, more and more companies are taking advantage of the power of social media and it is ability to help in disseminating information in a more timely and equitable manner. Further

research is needed to behavioural implications of big data's impact on stakeholders (Brown-Liburd et al, 2015).

The data science technique applied allows us to visualize the multi-dimensional relationships that exist in the disclosure practices among companies in the sample and lead these companies to group into the various clusters. In a nutshell, the results indicate that generally firm size and leverage have a positive relationship with the level of corporate disclosure via companies' websites as well as social media sites. Meanwhile, profitability in its own has no great impact on the degree of CID whether via companies' websites or social media sites. Liquidity has an impact on the degree of disclosure; however this impact seems to be a negative one in terms of the use of social media platforms for corporate disclosure. American companies tend to be on the cutting edge of technology when it comes to CID. It also seems that companies treat the various social platforms as a format of disclosure regardless of the substance of the content. The major importance of the findings is that when the analysis incorporates a larger sample,

5. Summary and Conclusions

This paper investigates the extent, nature and characteristics of CID practices among companies listed in the Australia, Canada, UK, and USA. The paper goes beyond the traditional studies that examine internet disclosure via the mean of corporate websites by extending the examination to the attributes and determinants of corporate disclosure using social networks and social media websites. As there is little empirical study on the use of social networks and media websites for corporate disclosure, this paper is an important contribution to filling the gap the literature. The paper provides new insights into CID across four major stock markets that will benefit all stakeholders with an interest in corporate reporting. Data has been collected and analysed on 1,490 companies listed in the four countries.

The results of this paper provide an in-depth multi-dimensional analysis of the nature, extent and characteristics of CID in Australia, Canada, UK, and USA. The findings reached by using data science could have not been revealed by adopting traditional regression analysis. The results show that companies' disclosure practices are not based on a single determinant; they are rather subject to variances based on the various combinations of several factors such as profitability, leverage, liquidity, firm size and industry type. The results reveal that companies with negative ROA are unlikely to disseminate great deal of information via social media. However, companies that are relatively large and have high leverage have good usage of social media as means of disclosure. In agreement with the agency theory, large companies that are highly leveraged tend to have better online disclosure. Moreover, ROA does not seem to have a sole direct impact of the degree of CID.

While some of these findings are consistent with the findings in the traditional literature on CID, using data science allows us to see that there are underlying relations between not only the determining factors of disclosure but also the various aspects of disclosure. It is clear that using social media for corporate disclosure is associated with the format of internet disclosure. Thus, it appears that at this early stage of using social media for corporate disclosure, companies are treating the various social media platforms are simply as part of having a better format on their website. This novelty issue is bound to change in the next couple of years and it is expected that there will be an increase and improvement of the content substance of what is being disclosed via social media to match what is on companies' websites.

Our research has important implications for both the accounting regulators and the users of CID. It appears that some firm-specific characteristics motivate companies to disclose more information on their website and social media. The users of CID should understand what motivates companies to report information voluntarily on their websites and on the social media. Regulators, therefore, may consider imposing additional mandatory requirements to protect these users from any misleading information; to encourage companies to disclose value-relevant information and to create better information environment for the users. As CID are not currently audited by external auditors, the users of CID may prefer to see external assurance by external auditors on the accuracy and credibility of the information provided via corporate websites as well as the social media. Finally, our research highlight that researchers need to pay attention to not only disclosure on social media but also incorporate big data analytics in understanding the corporate disclosure practices.

References:

- Al-Htaybat, K. and von Alberti-Alhtaybat, L. (2017) "Big Data and corporate reporting: impacts and paradoxes", *Accounting, Auditing & Accountability Journal*, Vol. 30 Issue: 4, pp.850-873
- Agrawal, R. and Dhar, V. (2014), "Big Data, data science, and analytics: The opportunity and challenges for IS research", *Information Systems Research*, Vol. 25 No.3, pp. 443-448.
- Barako, D., Hancock, P., and Izan, H. (2006), "Factors influencing voluntary corporate disclosure by Kenyan companies", *Corporate Governance: An International Review*, Vol. 14 No. 2, pp. 107-125.
- Basuony, M., Mohamed, E., and Elbayoumi, A. (2014), "Corporate Internet Disclosure in the Arabian Gulf: An Empirical Examination of Determinants and Attributes", *Journal of Modern Accounting and Auditing*, Vol. 10 No.7, pp. 747-763.
- Boubaker, S., Lakhali, F., and Nekhili, M. (2012), "The determinants of web-based corporate reporting in France", *Managerial Auditing Journal*, Vol. 27 No. 2, pp. 126–155.
- Brown-Liburd, H, Issa, H and Lombardi, D. (2015). "Behavioral Implications of Big Data's Impact on Audit Judgment and Decision Making and Future Research Directions. *Accounting Horizons*, 29 (2): 451-468.
- Craven, B. and Marston, C.. (1999) "Financial reporting on the Internet by leading UK companies", *European Accounting Review*, Vol. No. 82, pp. 321-333.
- Debreceeny, R. and Rahman, A. (2005), "Firm-specific determinants of continuous corporate disclosure", *The International Journal of Accounting*, Vol. 40, pp. 249-78.
- Debreceeny, R., Gray, G. and Rahman, A. (2002), "The determinants of Internet financial reporting", *Journal of Accounting and Public Policy*, Vol. 21 No. (4/5), pp. 371-394.
- Dhar, V. (2013). "Data science and prediction", *Communications of the ACM*, Vol. 56 No. 12, pp. 64-73.
- Elragal, A. and Klischewski, R., (2017) "Theory-driven or Process-driven Prediction? Epistemological Challenges of Big Data Analytics". *Journal of Big Data*, Vol. 4 No. 1, pp1-20

- Gartner (2016), "Magic quadrant for advanced analytics platforms", 09 February, ID:G00275788, <https://www.gartner.com/doc/reprints?id=1-2YC9GD6&ct=160209&st=sb>, accessed on 19/2/2016
- Hail, L. (2002), "The impact of voluntary corporate disclosures on the ex ante cost of capital for Swiss firms", *European Accounting Review*, Vol. 11 No. 11, pp. 741-73.
- Haniffa, R. and Cooke, T. (2005) 'The impact of culture and governance on corporate social reporting', *Journal of Accounting and Public Policy*, Vol. 24 No. 5, pp. 391-430.
- Healy, P. and Palepu, K. (2001), "Information asymmetry, corporate disclosure, and the capital markets: a review of the empirical disclosure literature", *Journal of Accounting and Economics*, Vol. 31, pp. 405-40.
- Jensen, M. and Meckling, W. (1976), "Theory of the firm: managerial behaviour, agency costs and ownership structure", *Journal of Financial Economics*, Vol. 3 No. 4, pp. 305-360.
- Jung, M., Naughton, J., Tahoun, A., and Wang, C. (2018), "Do Firms Strategically Disseminate? Evidence from Corporate Use of Social Media", *The Accounting Review*, Vol. 93, No. 4, pp. 1-57.
- Marston, C. and Polei, A. (2004), "Corporate Reporting on the Internet by German Companies", *International Journal of Accounting Information Systems*, Vol. 5, pp. 285-311.
- Mohamed, E. and Basuony, M. (2014), "Determinants and characteristics of voluntary internet disclosures in GCC countries", *The International Journal of Digital Accounting Research*, 14, pp. 57-91.
- Mohamed, E., Basuony, M., and Hussain, M., (2017) "Dissemination of Corporate Information via Social Media and Networks in Africa", *International journal of Corporate governance*, Vol. 8, No. 3/4, pp. 236-267.
- Oyelere, P., Laswad, F., and Fisher, R. (2003), "Determinants of Internet Financial Reporting by New Zealand Companies", *Journal of International Financial Management and Accounting*, Vol. 14, No. 1, pp. 26-63.
- Srivastava, S. and Teo, T. (2009), E-government, E-business, and National Economic Performance, *Communication of AIS*, Vol. 26, pp. 267-286.
- Wallace, R., Naser, K. and Mora, A. (1994), "The relationship between the comprehensiveness of corporate annual reports and firm characteristics in Spain", *Accounting & Business Research*, Vol. 25 No. 97, pp. 41-53.
- World Economic Forum (2014), The Global Information Technology Report 2014 – Rewards and Risks of Big Data, A report by the World Economic Report and Insead, Geneva, Switzerland. Accessed on 13 March 2015 from www.weforum.org/gitr
- Xiao, J., Dyson, R. and Powell, P. (1996), "The impact of information technology on corporate financial reporting: a contingency perspective", *British Accounting Review*, Vol. 28 No. 3, pp. 203-27.
- Xiao, J., Yang, H. and Chow, C. (2004), "The Determinants and Characteristics of Voluntary Internet-Based Disclosures by Listed Chinese Companies", *Journal of Accounting and Public Policy*, Vol. 23 No. 3, pp. 191-225.

Zhu, K. and Kraemer, K. (2005), "Post-Adoption Variations in Usage and Value of E-business by Organization: Cross-Country Evidence from the Retail Industry", *Information Systems Research*", Vol. 16, pp. 61-84.

Table 1
Definition and Measurement of Variables

Symbol	Variable Definition	Measurement
Panel A		
FDI	Format Disclosure Index	the actual scores awarded divided by the maximum possible (14)
GCDI	General Content Disclosure Index	the actual scores awarded divided by the maximum possible (13)
FCDI	Financial Content Disclosure Index	the actual scores awarded divided by the maximum possible (12)
TCDI	Total Content Disclosure Index	the actual scores awarded divided by the maximum possible (25)
TWDI	Total Website Disclosure Index	the actual scores awarded divided by the maximum possible (39)
SNMDI	Social Network and Media Network Disclosure Index	the actual scores awarded divided by the maximum possible (37)
TDI	Total Disclosure Index	the actual scores awarded divided by the maximum possible (76)
Panel B		
FrmSize	Firm Size	Natural log of total assets
ROA	Return on Assets	Net Income / Total Assets
Lvg	Leverage	Total liabilities/ Total Assets
Lqdy	Liquidity	Current assets/ current Liabilities
IndTyp	Industry Type	Manufacturing = 1; Services = 2
Panel C		
Market	Stock Market	FTSE 350 = 1; S&P 500 = 2; TSX300 = 3; ASX400 = 4

Table 2: Summary of Social Media and Networks across the Four Countries

Panel (B): Summary of Social Networks across the Countries														
		UK (N = 350)				USA (N = 502)			Canada (No = 238)			Australia (N = 400)		
		Tot. %	%	Mean	SD.	%	Mean	SD.	%	Mean	SD.	%	Mean	SD.
Facebook	Us	43.5	29.4	0.29	0.456	74.3	0.75	0.435	31.5	0.32	0.466	24.1	0.24	0.428
	Gin	42.4	29.1	0.29	0.455	72.3	0.72	0.448	31.5	0.32	0.466	23.1	0.23	0.422
	Fin	10.3	12.6	0.13	0.332	14.6	0.15	0.353	8	0.08	0.272	4.3	0.04	0.202
	AR	6.4	5.1	0.05	0.221	9.2	0.09	0.289	7.6	0.08	0.265	3.8	0.04	0.190
Twitter	Us	51.5	44.4	0.44	0.498	76.4	0.76	0.425	39.9	0.40	0.491	33.8	0.34	0.473
	Gin	48.1	44.3	0.44	0.497	67.1	0.67	0.470	39.9	0.40	0.491	32.5	0.33	0.469
	Fin	17.3	29.5	0.30	0.457	12.6	0.13	0.332	14.3	0.14	0.351	14.2	0.14	0.350
	AR	10.9	11.5	0.11	0.319	9.2	0.09	0.289	11.4	0.11	0.318	12.3	0.12	0.328
Instagram	Us	9.9	3.1	0.03	0.175	19.6	0.20	0.397	5.5	0.05	0.228	6.3	0.06	0.242
	Gin	7.9	2.9	0.03	0.167	14.4	0.14	0.351	4.6	0.05	0.210	6.3	0.06	0.242
	Fin	1.1	0.3	0	0.053	3.2	0.03	0.176	0	0.00	0.000	0	0.00	0.000
	AR	0.9	0	0	0	2.6	0.3	0.159	0	0.00	0.000	0	0.00	0.000
Linkedin	Us	41.8	31.1	0.31	0.464	56.9	0.57	0.496	39.9	0.40	0.491	33.5	0.34	0.473
	Gin	41.4	31.1	0.31	0.464	55.7	0.56	0.497	39.5	0.39	0.490	33.5	0.34	0.473
	Fin	14.1	14.6	0.15	0.353	14	0.14	0.347	11.4	0.11	0.318	15.5	0.16	0.362
	AR	8.5	6	0.06	0.238	7.4	0.07	0.262	10.5	0.11	0.307	10.8	0.11	0.310
Google+	Us	19.2	12.9	0.13	0.335	37.7	0.38	0.485	7.6	0.08	0.265	8.8	0.09	0.283
	Gin	17.5	12.3	0.12	0.329	32.7	0.33	0.470	7.2	0.07	0.259	8.8	0.09	0.283
	Fin	4.1	4.3	0.04	0.203	7.8	0.08	0.269	1.7	0.02	0.129	0.8	0.01	0.086
	AR	2.5	2	0.02	0.140	4.8	0.05	0.214	1.7	0.02	0.129	0.5	0.01	0.071

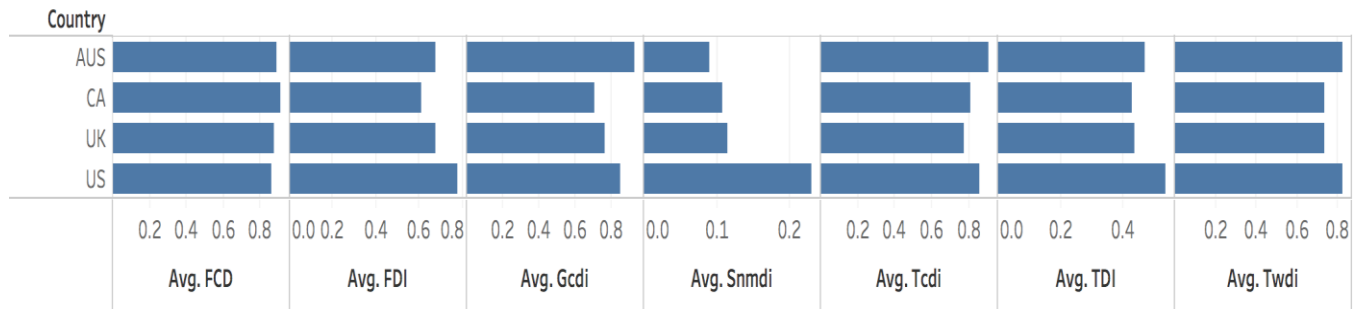
Panel (A): Summary of Social Media across the Four Countries														
		UK (N = 350)				USA (N = 502)			Canada (N = 238)			Australia (N = 400)		
		Tot. %	%	Mean	SD.	%	Mean	SD.	%	Mean	SD.	%	Mean	SD.
Blogs	Us	11.1	4	0.04	0.196	25.3	0.25	0.435	4.6	0.05	0.210	3.5	0.04	0.184
	Gin	10.2	3.7	0.04	0.189	23.0	0.23	0.422	4.6	0.05	0.210	3.3	0.03	0.178
	Fin	2.2	0.6	0.01	0.075	5.4	0.05	0.226	0	0.00	0.000	1	0.01	0.100
	AR	0.8	0	0.00	0.000	1.8	0.02	0.133	0	0.00	0.000	0.8	0.01	0.086
Flicker	Us	4.6	5.7	0.06	0.232	9.2	0.09	0.289	0.8	0.01	0.091	0	0.00	0.000
	Gin	3.8	5.7	0.06	0.232	6.8	0.07	0.252	0.8	0.01	0.091	0	0.00	0.000
	Fin	0.4	0.6	0.01	0.075	0.6	0.01	0.077	0.4	0.00	0.065	0	0.00	0.000
	AR	0.2	0	0.00	0.000	0.4	0.00	0.063	0.4	0.00	0.065	0	0.00	0.000
Youtube	Us	37.1	30	0.30	0.459	60.5	0.61	0.489	26.5	0.26	0.442	20.3	0.20	0.402
	Gin	35.2	29.7	0.30	0.458	55.5	0.56	0.497	26.1	0.26	0.440	19.8	0.20	0.399
	Fin	5.9	6.6	0.07	0.248	8.2	0.08	0.275	1.7	0.02	0.129	5	0.05	0.218
	AR	3.0	2.3	0.02	0.150	4.6	0.05	0.209	1.7	0.02	0.129	2.3	0.03	0.148
Slideshare	Us	4.3	1.4	0.01	0.119	11.2	0.11	0.315	1.3	0.01	0.112	0	0.00	0.000
	Gin	4.0	1.4	0.01	0.119	10.4	0.10	0.305	1.3	0.01	0.112	0	0.00	0.000
	Fin	2.1	0.9	0.01	0.093	5.2	0.05	0.222	0.8	0.01	0.091	0	0.00	0.000
	AR	0.7	0	0.00	0.000	1.8	0.02	0.133	0.8	0.01	0.091	0	0.00	0.000

Table 3: K-Means Clustering Results - Centroids

Panel (A): K-Means Clustering Results - Centroids								
Attribute	Cluster 0		Cluster 1		Cluster 2			
	N = 835	56%	N = 627	43%	N = 19	1%		
FDI	0.73		0.67		0.67			
GCDI	0.83		0.82		0.82			
FCD	0.88		0.88		0.86			
TCDI	0.85		0.84		0.81			
TWDI	0.80		0.78		0.76			
SNMDI	0.19		0.09		0.08			
TDI	0.50		0.44		0.43			
Panel (B): Clusters Membership by Country & Industry Type								
Cluster	Australia (n = 400)		Canada (n = 229)		UK (n = 350)		USA (n = 502)	
	No.	% of Sample	No.	% of Sample	No.	% of Sample	No.	% of Sample
Cluster 0	114	28.50%	111	49%	161	46%	449	90%
Cluster 1	276	69%	115	50%	184	53%	52	10%
Cluster 2	10	2.50%	3	1%	5	1%	1	n.s.
Cluster	Manufacturing (n = 594)			Services (n = 887)				
	No.	% of Sample		No.	% of Sample			
Cluster 0	309	52%		526	59%			
Cluster 1	281	47%		346	39%			
Cluster 2	4	1%		15	2%			

Figure 1
Indices by Country

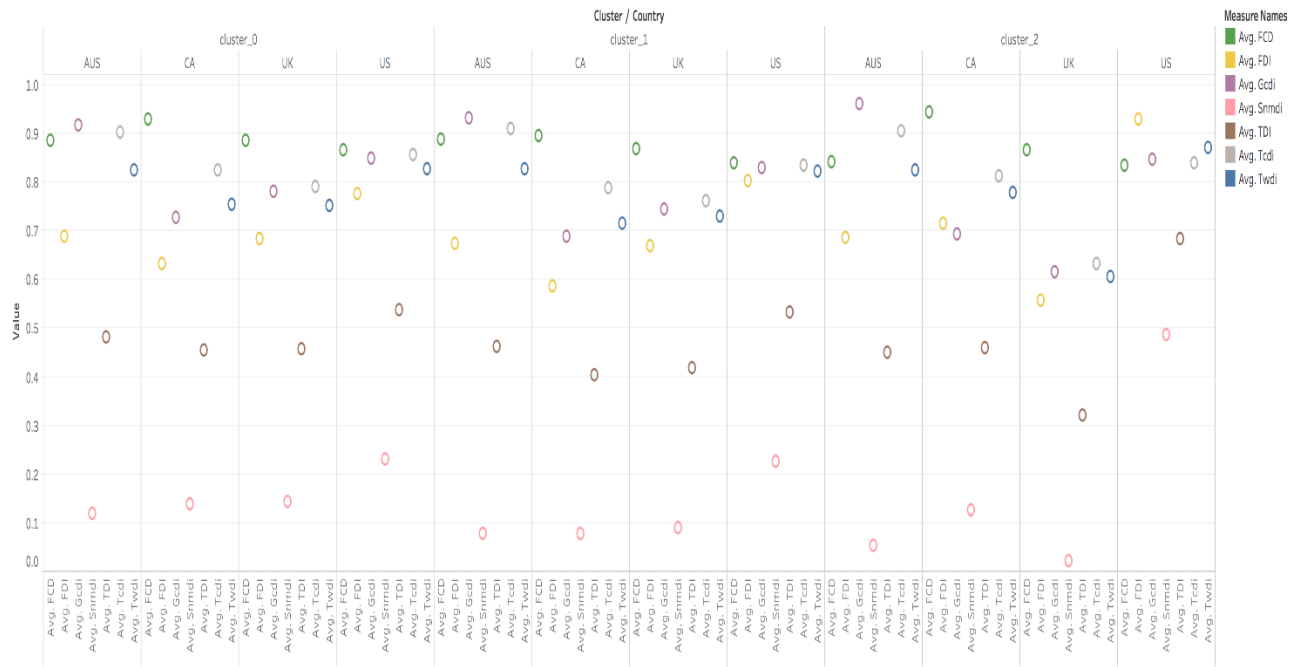
Country by Indexes



Average of FCD, average of FDI, average of Gcdi, average of Snmdi, average of Tcdi, average of TDI and average of Twdi for each Country.

Figure 2
Indices by Country

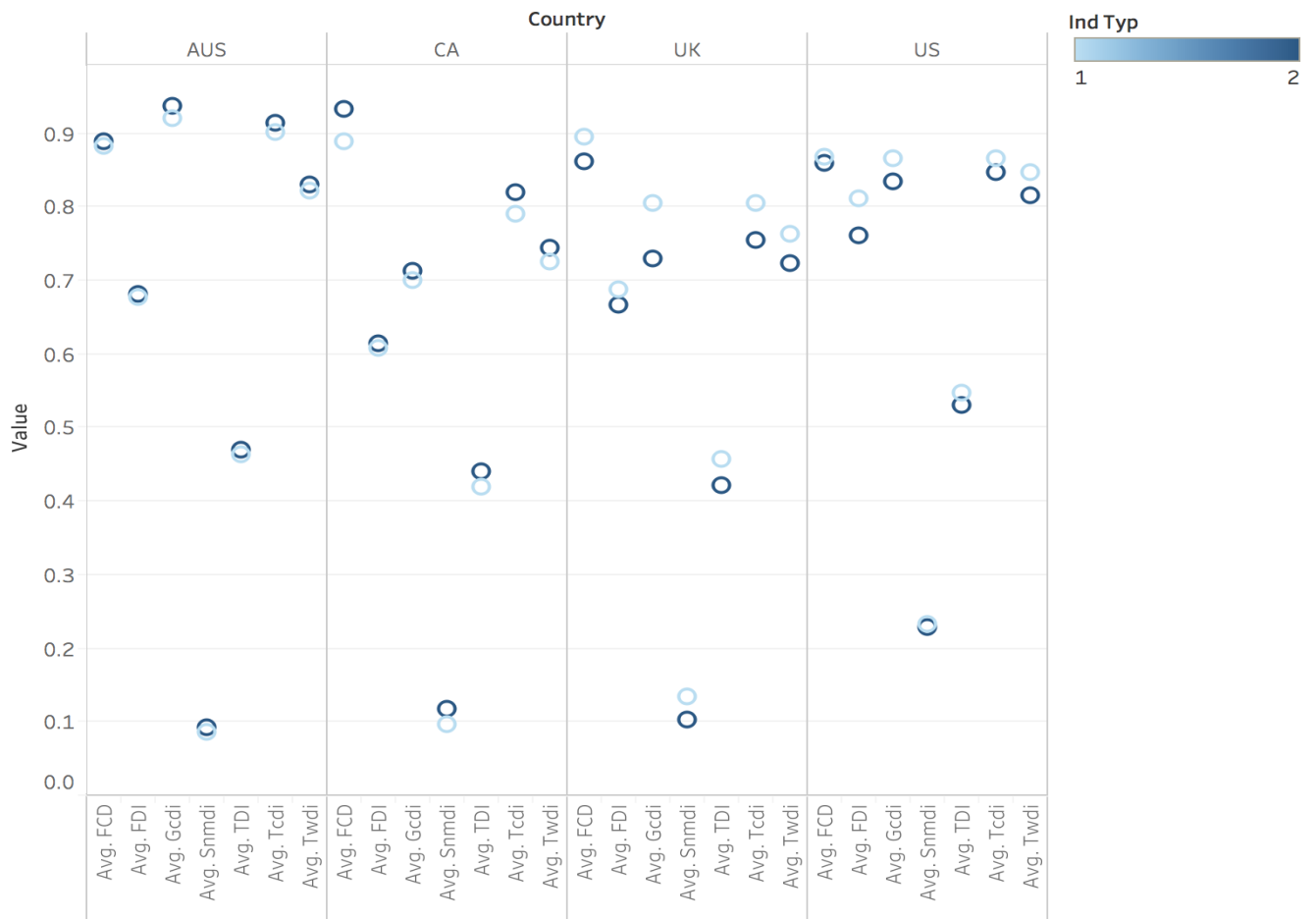
Cluster by Indexes



Avg. FCD, Avg. FDI, Avg. Gcdd, Avg. Snmdd, Avg. TDI, Avg. Tcdd and Avg. Twdd for each Country broken down by Cluster. Color shows details about Avg. FCD, Avg. FDI, Avg. Gcdd, Avg. Snmdd, Avg. TDI, Avg. Tcdd and Avg. Twdd.

Figure 3
Indices by Country and Industry

Country & Industry by Indexes



Avg. FCD, Avg. FDI, Avg. Gcdi, Avg. Snmndi, Avg. TDI, Avg. Tcdi and Avg. Twdi for each Country. Color shows details about Ind Typ. Details are shown for Ind Typ.

Figure 4
K-Means Clustering Results - Centroids

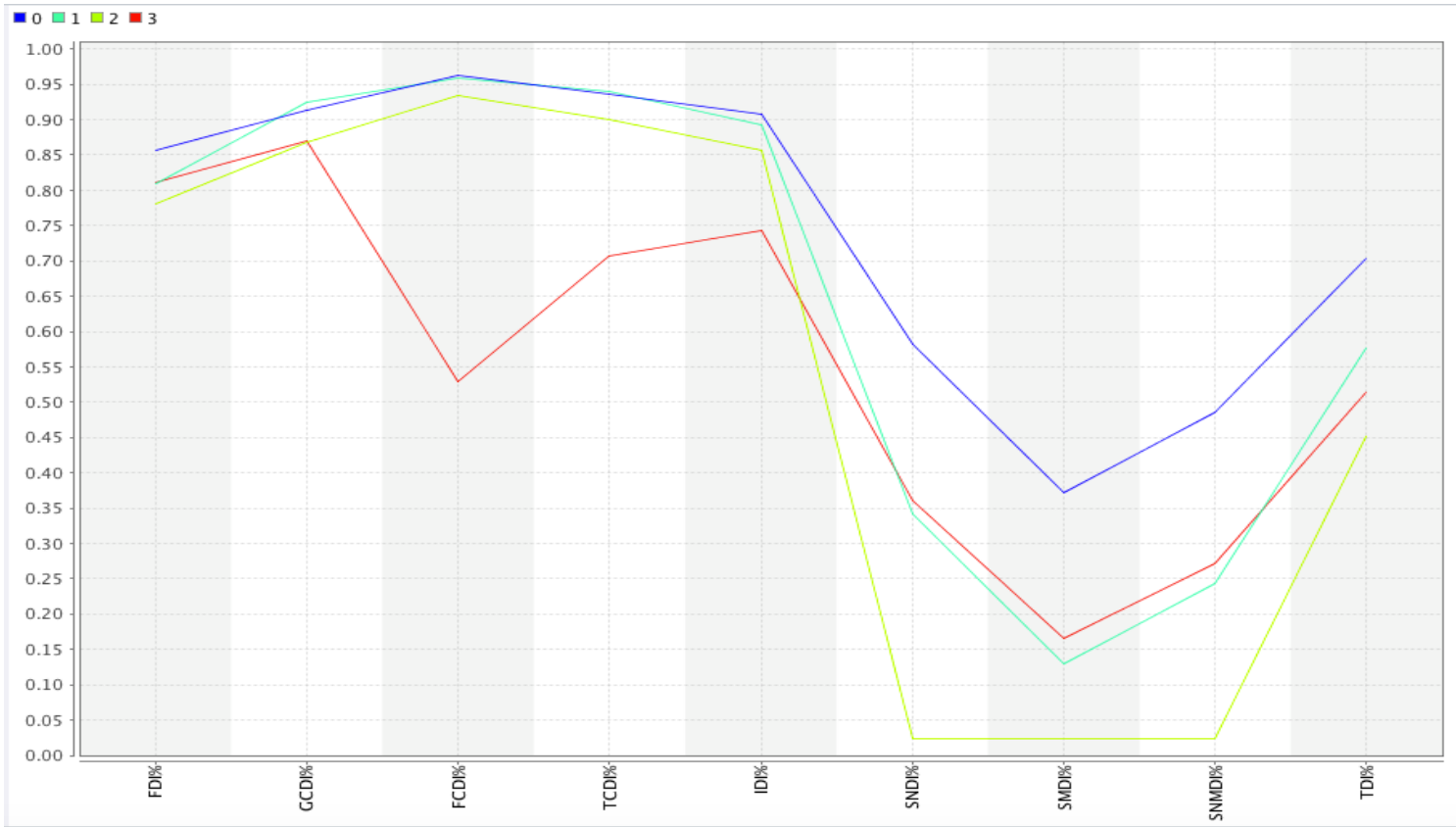
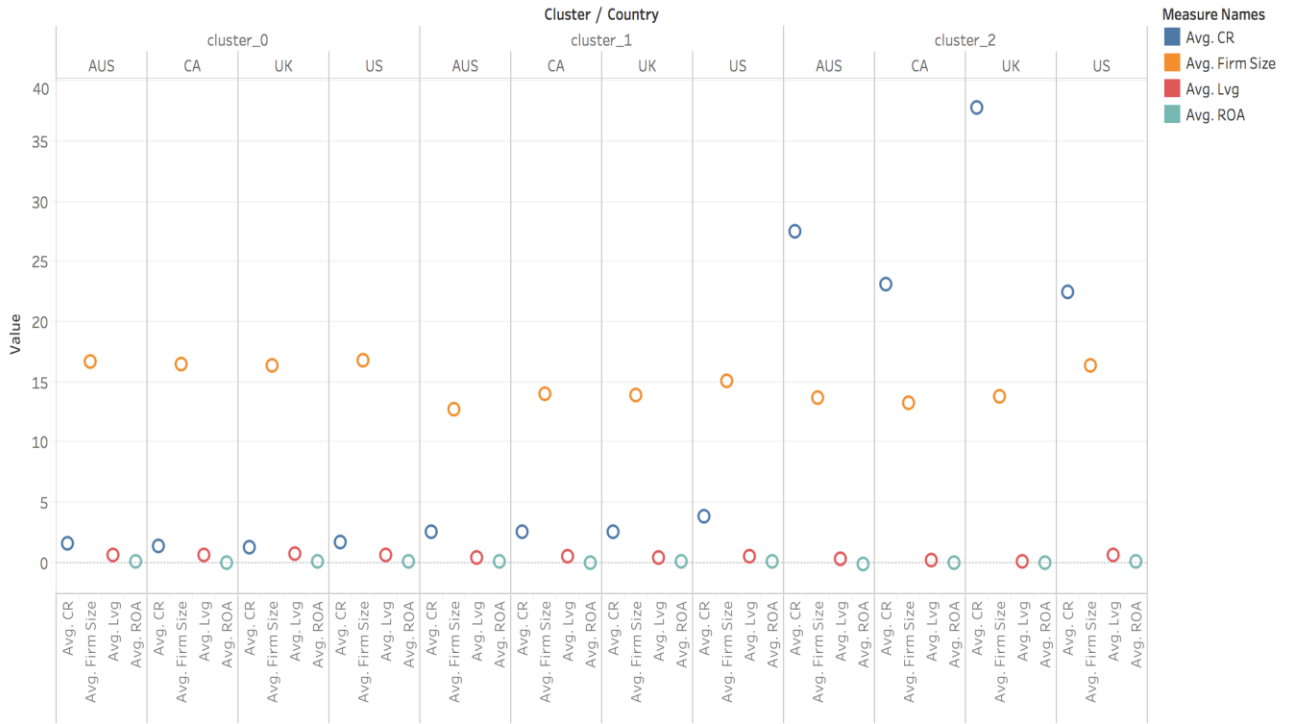


Figure 5
 Clusters Profile by Profitability, Leverage, Liquidity & Firm Size

Cluster, Country, Measures



Avg. CR, Avg. Firm Size, Avg. Lvg and Avg. ROA for each Country broken down by Cluster. Color shows details about Avg. CR, Avg. Firm Size, Avg. Lvg and Avg. ROA.