Text Mining and Dimension Reduction Methods of Exploring Isomorphism in Corporate Communications

Atsuho Nakayama, Adriana Paliwoda Matiolańska and Emilia Smolak-Lozano

Abstract This study analyzes the isomorphism pressures within the context of sustainability by exploring Twitter communication in the energy sector. Social Media, in particular Twitter, provide a new opportunity to explore the linguistic dimension in corporate communications. This paper proposes the use of Social Media linguistic data (tweets, including hashtags and keywords) and a triangulated method (text mining, web mining, linguistic and content analysis) to examine trends in tweets for individual companies. Based on the institutional theory of organizational communication, this study examines the relationship between the concept of sustainability and isomorphism, the latter of which leads to the adoption of similar models and attitudes among

Atsuho Nakayama
Faculty of Urban Liberal Arts, Division of Business Administration and Division of Economics
Tokyo Metropolitan University, 1-1 Minami-Ohsawa, Hachioji-shi, 192-0397, Japan
✉ atsuho@tmu.ac.jp

Adriana Paliwoda Matiolańska
Department of Trade and Market Institutions, Cracow University of Economics, Poland
✉ paliwoda@uek.krakow.pl

Emilia Smolak-Lozano
Department of Audiovisual Communication and Advertising, University of Malaga
Avda. Cervantes, 2. 29071 Malaga, Spain
✉ emilia.smolak@gmail.com

Archives of Data Science, Series A
(Online First)
KIT Scientific Publishing
Vol. 7, No. 1, 2020

DOI: 10.5445/KSP/1000098012/09
ISSN 2363-9881

organizations. It applies text mining and correspondence methods within R software for energy sector tweets in English from 2016. The results reveal a tendency among energy companies to follow similar patterns in Twitter communication on sustainability. This provides insight into the mechanisms that lead to isomorphism in organizational communication.

1 Introduction

Recently, there has been an increasing academic focus on the interactive and communicative constructions of institutions to understand how organizations sustain institutional pressures. Rhetorical commitments that create narrative dynamics in organizational communication are central to institutional diffusion and change. Understanding how organizations maintain cross-institutional pressures over time requires more empirical research, especially with respect to the idea of sustainability.

After the concept of sustainability was introduced, it began to dominate modern discourse by establishing a common language between the state, businesses, and ecologists or organizations dealing with nature conservation, including ecological social movements (Macnaghten and Urry, 2005). Burns (2011) points out that the sustainability revolution has already taken place on normative-cognitive, practical and institutional levels. However, the necessity of consensus can produce similar forms and practices among organizations. Therefore, the institutional perspective also attempts to reveal why social interactions are often repeatable and similar. Furthermore, it questions how the existence of isomorphism manifests itself in corporate communication.

Based on a scientific tradition that regards communication as constitutive of an organization, this paper discusses the communicative mechanisms that present the concept of sustainability standards through new digital communication channels. Additionally, drawing on the institutional theory of organizational communication, we analyze the relationship between the concept of sustainability and isomorphism that leads to the adoption of similar models and attitudes among organizations, in particular within digital communications. We focus on the semantic analysis of sustainability discourse among large energy corporations. In particular, we analyze the semantic features of corporate communications practice (Public Relations (PR) strategy) that help to position a brand or firm and manage its reputation (Sethi, 1979; Uzunoğlu et al., 2017; Verčič

et al., 2018). This study extends the growing body of institutional research by studying communication patterns via Social Media among energy companies. We elucidate isomorphism pressures within the context of sustainability by exploring Twitter communication in the energy sector using text mining and correspondence methods in R software (Fellows, 2018).

2 Prior Literature

Institutional theory is an important tool for understanding the adoption and diffusion of socially responsible behavior and sustainability initiatives (Jennings and Zandbergen, 1995; Ashworth et al., 2009). It has created new insights into organizations, accompanied by research tools enabling more in-depth analyses, as well as a shift of focus (Peters, 2012). The literature linking institutional theory with the management of social issues, especially corporate social responsibility, has used institutional theory as an explanation for the adoption of sustainable or socially responsible practices (Glover et al., 2014; Brammer et al., 2012; Misani, 2010; Campbell, 2007).

Institutional studies have drawn attention to pressures that arise in size, geographical, and industry-based fields to explain these patterns, and this work is continuously under development. Factors such as political environment, governance structure and cultural differences play an important role in the responsible behavior of firms (Matten and Moon, 2008; Maignan and Ralston, 2002). Adopting legitimate practices increases an organization's survival probability (Lamin and Zaheer, 2012). However, when companies interact, they tend to imitate and influence each other's activities. The isomorphic process explains why firms trend toward similarity in behavior and practices due to being communally intertwined (Joseph and Taplin, 2012), and the strategies, forms, and practices of the isomorphic process are explained by organizational imitation.

Lieberman and Asaba (2006) point out that imitation is a frequent and popular behavior of firms and industries. In the literature, we can find several theories that explain mimetic phenomena from different viewpoints. Social approaches (Dacin, 1997; Deephouse, 1996; DiMaggio and Powell, 1983) focusing on legitimacy reasons in adoption and imitation, as well as economic, competitive variants (Ketokivi and Schroeder, 2004; Haunschild and Miner, 1997; Scott, 2001; Hannan and Freeman, 1997) emphasizing economically motivated adoption and mimicking of practices, may justify isomorphism and the adoption

of similar practices among firms. Companies attempt to mimic the practices of other companies to benefit from their experience (Hutzschenreuter et al., 2011), to maintain their market position or to neutralize competitor threats (Lieberman and Asaba, 2006). Mimetic pressures come as a result of uncertainty and ambiguities in goals, whereas normative pressures are a result of professionalization and common control (Heugens and Lander, 2009; DiMaggio and Powell, 1983). Furthermore, the institutional theory of organizational communication points out that institutional rules are communicatively constructed and tend to be reproduced in organizational communication (Lammers and Barbour, 2006; Cornelissen et al., 2015). The media shapes social reality (Livingstone, 2009). As such, the realities created through the media regarding the way organizations work, their products, values and preferences serve as a common basis for building organizational reality (Sinha et al., 2012; Anderson and Warren, 2011).

Institutions are constituted by discourses (Maguire and Hardy, 2009), along with legitimization (Vaara and Tienari, 2008). Throughout their communications, they may reinterpret the norms, values, and beliefs of stakeholders (Lindblom, 2010; Sethi, 1975, 1979; Dowling and Pfeffer, 1975). Communications studies remind us of the constitutive and formative role of communication for organizations, and generally point out the distinction between "talk" and "action", while recognizing talk as an action in its own right (Ashcraft et al., 2009). Additionally, the messages generated by institutional pressures support the creation of a conceptual and empirical link between the macro-world of institutions and the micro-world of organizational communication.

Matten and Moon (2008) point out the "contextual construct" role of the corporation in the sustainable debate. This allows organizations – like other social entities – to emerge and act (Couldry, 2012). In line with this premise, managing sustainability is an outcome of the response to external pressure and criticism from civil society (Gomez et al., 2015), and creating consensus on issues of institutional sustainability leads to similarity in the form of practice among organizations, including corporate communication. Furthermore, digital communication brings new possibilities of dialogical and meaningful connection with stakeholders in this area (Uzunoğlu et al., 2017; Abitol and Lee, 2017; Verčič et al., 2018) and creates a field for exploration.

Twitter is an example of a Social Media channel which, when used strategically, allows for interactive, real-time and engaging conversation with key stakeholders that benefits organizational public relations (Frame and Brachotte, 2015; Saffer

et al., 2013; Smitko, 2019). Studies on media and business relations show that corporations are very involved in shaping not only their brands and images but also the standards and ideals of a good business. Sahlin-Andersson and Engwall (2002) argued that we should analyze contemporary enterprises as forms of "edited corporation", where corporations are partly shaped by images and ideals that revolve around them. Within organizational studies, isomorphism in communication is generally analyzed via sustainability/corporate social responsibility (CSR) reporting (Higgins et al., 2018; Parente et al., 2014), ISO (International Organization for Standardization) standards (Castka and Balzarova, 2008) or value chain sustainable practices (Yawar and Kauppi, 2018; Amaeshi et al., 2008). Digital communication is not well recognized. Therefore, our research attempts to fill this gap by initiating studies in this area. We adopted the following research questions:

1. How do the energy sector's most prominent companies construct discourse about sustainability via corporate communication strategy on Twitter?
2. To what degree does this discourse follow similar patterns of Twitter communication strategy among companies from the energy sector?

3 Analysis

This study analyzes semantic features of the Twitter communication strategies of energy sector companies. As the research is centered on sustainability discourse, the tweets were pre-selected and limited to those that construct an online conversation on this topic, focusing on social and environmental change. Next, we focus on semantic homophiles that can help to increase our understanding of semantic aspects that shape online corporate communication to determine how semantics shapes the discourse as part of the online conversation strategy.

We propose the use of social media data and a method to analyze Tweet trend in each company. The energy sector tweets in English were subjected to text mining processes for statistical linguistic analysis using R. Text mining can reveal and visualize latent structures and content in unstructured or weakly structured text data in a given collection of documents. Text mining techniques involve linguistic, statistical, and machine-learning algorithms using the word cloud package of R software. The URLs for tweets (http:) and links to tweets were removed to avoid

affecting the top word count statistics and to keep the word clustering objective and balanced. Similarly, to maintain fresh linguistic content the program also eliminated white spaces, numbers, and stop words. The procedure transformed the text to lower-case and preserved the originals. Emoticons and other symbols were not considered. The analysis was run with punctuation removed.

The goal of this study was to represent the topic of a document based on the frequency of semantically significant terms. Document-term matrices were calculated by text mining the tweet data. Then, by aggregating the document-term matrix for each company, we calculated data representing word frequency. We analyzed the data by correspondence analysis (CA) (Greenacre, 1984), because CA provides visual indicators of the contributions of words and documents based on axes. CA is a popular method for analyzing contingency tables, and is based on SVD (Singular Value Decomposition). CA can graphically represent the results in the statistical interpretation based on the importance of the axis.

This study included the energy sector's most prominent companies (based on Platts Top 250 Ranking) and their Twitter accounts. As such, we selected the most powerful energy companies that constitute the global leaders of the industry. First, in an explorative qualitative study carried out in November and December 2016, we pre-identified 16 hashtags and keywords related to the issues of sustainability and social change as primary areas of CSR activity in the energy sector (sustainable/sustainability, climate change, emission, clean energy, environment, low carbon, social, communities, gas flaring, spills, recycle, innovation, energy, future energy), in Polish, Spanish, French, and English. These hashtags were applied to monitor the energy sector companies' Tweets (50 Twitter accounts in total in Table 1 and 2) using online monitoring tools (e.g., Brand24) in 2016 (the last complete last year to which Twitter grants access) and pre-established keywords and hashtags. As a result, 3,043 tweets (4 % of all tweets) containing at least one of the monitored keywords and/or hashtags were identified (we classified these as "sustainability-oriented tweets") and subsequently organized into one database (data, tweets, link, company, keyword or hashtags). Using R software, an exploratory linguistic analysis based on statistical text mining techniques was performed on 2,081 tweets in English to determine the narrative contextualization of communication by tweets.

Table 1: International energy companies and their names on Twitter.

International Energy Company	Name on Twitter	Remarks
Marathon Petroleum	@MarathonPetrol_	No related tweets
Enterprise Product Partners	@EProd_Careers	
OJSC Rosneft Oil&Co	@RosneftEN	
Tokyo Electronic Power Co	@TEPCO_English	
OJSC Lukoil Oil Co	@lukoilengl	
RWE AG	@RWE_Group	
Electricite de France	@EDFofficiel	French
Sasol Ltd	@SasolLTD	
Coal India Ltd	@CoalIndiaHQ	
Tenaga Nasional Berhard	@Tenaga_Nasional	
PG&E Corp	@PGE4Me	
ENGIE SA	@ENGIEgroup	
Canadian Natural Resources Ltd	@CNRLCareers	
National Grid plc	@nationalgriduk	
American Electric Power Co, Inc	@AEPnews	
Enbridge	@Enbridge	
Indian Oil Corporation	@IndianOilcl	
Cenovu s Energy	@cenovus	
E.ON	@EON_SE_en	
Cameco	@camecocommunity	
Bharatpetroleum	@BPCLimited	
Exxon Mobile	@exxonmobil	
Chevron	@Chevron	
Royal Dutch Shell	@Shell	
Phillips 66	@Phillips66Co	
ConocoPhillips	@conocophillips	
Valero Energy	@ValeroEnergy_	No related tweets
SINOPEC	@SinopecNews	
NESTE	@nestecorp	
GAZPROM	@GazpromEN	
Encana	@encana	
Suncor	@suncorenergy	
Hess	@HessCorporation	
Iberdrola	@iberdrola	Spanish
Gas Natural	@GNF_es	Spanish
BP	@BP_plc	
Oil |& Natural Gas Corporation	@ONGC_	
Ecopetrol	@ECOPETROL_SA	Spanish
TOTAL S.A.	@Total	
Statoil	@Statoil	
Next Era Energy	@NextEraEnergyR	
Enel	@EnelGroup	
ENI	@eni	
EDP-Energias de Portugal S.A. Portugal	@energiapt	No related tweets

Table 2: Polish energy companies and their names on Twitter.

Polish Energy Company	Name on Twitter	Remarks
PKN ORLEN	@PKN_ORLEN	
Tauron	@TauronPE	
LOTOS	@GrupaLOTOS	
PGE	@RzecznikPGE	
PSE S.A.	@PSETalk	No related tweets
ENEA	@ENEA_Consulting	No related tweets

This analysis revealed that "energy", "innovation", "future", "sustainability", and "sustainable" were the most frequently used expressions and keywords among sustainability-oriented tweets. The term "energy", as the basic sector description, dominated other words more closely related to social change. "Innovation" and "future" were more frequent than "sustainability" or "sustainable", used here as general expressions. The 10 most frequently used words belong to the symbolic dimension, confirming the observed tendencies.

The frequency statistics of text usage indicate a brand/corporate-oriented PR strategy ("energy", "sustainability", "innovation"), albeit without unique branding, in online conversations or a proactive approach toward sustainable change in the dialogue with stakeholders on Twitter. To provide deeper insight, a word cloud was generated. Figure 1 shows the results obtained from word cloud analysis. A word matrix was constructed using a frequency threshold of 15 occurrences, below which keywords were removed from the word cloud library. R analysis of the word cloud demonstrated that the narrative content of the CSR digital strategy is auto-referential, indicating the prevailing approach to symbolic communication on the one hand (universal terms generally used in CSR/sustainability-oriented communication strategies) and corporate/branding topics on the other. This is more evident in the tweet content than in the trending topics.

Table 3: The 10 most frequently used words.

Word	Frequency	Word	Frequency
energy	1365	bpstats	127
innovation	263	enel	126
future	179	new	118
sustainability	156	environment	108
sustainable	134	wil	99

The text mining method explores common terms and their associations. The main word that is over-represented in the Twitter discourse about the sustainability of the energy sector is “energy”. It is worth noting that the explorative analysis yields more relevant results, as it detects a considerably higher number of keywords and expressions. The PR strategy of the energy sector is based on the general and symbolic use of universal expressions such as “sustainable”, “sustainability”, “future”, “innovation”, “development”, “environment”, “climate”, and “change”. In lower proportions, but still dominating over the social change discourse, the following expressions are used: “bpstats”, “CEO”, “release”, “conference”, “project”, and “startup”. This reveals a brand/corporate-oriented narrative and traditional CSR vision (reporting, events, CEO presence). The PR strategy employs the sustainability discourse in a universal symbolic manner, and includes general terms related to the sector’s future in terms of new development (“innovation”, “startup”, “technology”). However, these general terms are not necessarily related directly to CSR issues because they are not employed in the specific context of social change.

Figure 1: The result of word cloud.

“Innovation” and “sustainability”, as both keywords and trending topics, are the most common expressions employed in the discourse on Twitter. However, they are symbolic in meaning, once again avoiding a clear association with specific

initiatives that would reflect a real change in corporate CSR strategy beyond branding purposes.

In general, the cloud illustrates a rather dispersed narrative with strong autoreferential ("energy") and corporate future orientation ("innovation"), accompanied by branding efforts ("conference", "projects", "auto-references", "stats", "CEOs") with a symbolic CSR/sustainability communication approach ("climate change", "future", "new", "globe", etc.). The brand- and corporate-related expressions can be identified at a relatively high frequency: "Enel", "Suncor and BP stats", "Tepco and Indianoil" (mainly as keywords in the tweets' content). A corporate cluster can also be identified, represented by "stats", "startups", "release", and "CEO", as well as the most commonly applied terms of "innovation", "future", and "technology".

This indicates a strong brand focus on PR in the energy sector, with a clear symbolic orientation toward the companies' and sector's own futures. Regarding sustainability issues, this narrative is rather symbolic. Social change-oriented words such as "research", "impact", or "challenge" can be observed, but they are outnumbered by corporate words in sustainability-oriented tweets. Terms such as "talks", "top", "best", "story", "news", "industry", "business", "company", and "economy" comprise a significant proportion of the word cloud, and are used widely within the sector.

Collectively, these tweet data reflect a strong presence of the corporate PR strategy. It should be noted ENI and ENEL made efforts to create their own narratives concerning social change issues within the sustainability discourse: "eniday", "theenimodel", or "enelgreenpower". These expressions illustrate a mixed PR strategy in CSR communication on Twitter. This strategy mentions social progress initiatives, but includes a strong brand and corporate association.

In contrast, the social change narrative is used marginally, although it is more evident within the content of tweets than within trending topics: "emissions", "communities", "progress", "support", "transition", "renewable", "education", "need", "safety", "clean", "green", "cities", "wind", "major save", "carbon", and "social" were among the terms used. These words appear in tweet content less than the corporate, future and general energy discourse. Furthermore, as the word cloud shows, they are used less and in a more dispersed manner. In the majority of cases, the use of these words does not imply a clear social change context.

The narrative omits or includes minimal use of terms related to social change, such as: “low carbon”, “wind”, “environmental”, “help”, “model”, “value”, “people”, and “commitment/committed”. There was no statistically significant presence of texts concerning wind power, COP21/COP22, or solar innovation. Generally, the discourse regarding social change is highly fragmented and dispersed around many different social and environmental topics, and is sometimes nearly insignificant within the overall PR strategy.

As the next step, CA was performed by using the “corresp” function in the “MASS” package in R (Venables and Ripley, 2002). A document-term matrix was calculated by text mining using the 2,081 English tweets. Then, by aggregating the document-term matrix for each company, we calculated data representing company-specific word frequency. We excluded companies with few tweets and words with low frequencies. We included words that were used at least 15 times and excluded companies with fewer than 50 tweets. Our research goal was to determine the topic of a document based on the frequency of semantically significant terms, so companies with few tweets and words with low frequencies were not very important.

A contingency table containing 13 companies (Chevron, Indian Oil Corporation, TOTAL S.A., ENGIE SA, Royal Dutch Shell, Suncor, Enbridge, E.ON, NESTE, BP, ENI, Exxon Mobile, Enel) and 164 words was analyzed by CA. If a table has r rows and c columns, the number of dimensions in the correspondence analysis result is r minus 1 or c minus 1, whichever is less. Our data contain 13 rows and 164 columns. Thus, the number of dimensions in our analysis result was 12. The number of dimensions to retain in the results can be determined by examining the eigenvalues, which correspond to the amount of information retained by each dimension. Dimensions are listed in decreasing order according to the amount of variance explained (dimension 1 explains the most variance, followed by dimension 2, etc.). The contribution of dimensions shows the of sum of eigenvalues retained by each dimension. The contribution of dimensions 1 to 12 were 16.18, 16.07, 10.93, 9.13, 8.39, 8.02, 7.33, 6.56, 5.70, 5.09, 3.53, and 3.07, respectively. If the data were random, the expected value of the eigenvalue for each dimension would be 1/(number of rows−1) = 1/12 = 8.33 %. Any axis with a contribution larger than the percentages should be considered important for the interpretation of the data (Bendixen, 1995). Thus, we could use dimensions 1, 2, 3,4 and 5 for the interpretation of the data.

To obtain sufficient information from the results of analysis, the contribution of the dimensions should be 70 to 80.

Nevertheless, CA is aimed at displaying the rows and columns in a lower-dimensional vector space. As such, we adopted a two-dimensional representation that is easy to understand visually and is more comprehensible than the multidimensional spatial nature of this study. When fewer than the maximum number of possible dimensions are used, it is important how well the row and column categories are represented in the lower-dimensional representation. If the quality of representation of the two-dimensional result is sufficient, one can examine plots of the row points and column points to learn which categories of the row variable are similar, which categories of the column variable are similar, and which row and column categories are similar to each other. In CA, one must imagine a line from each object to the center of the map and assess the angle when pairs of lines meet, when interpreting associations between row categories, column categories, or row and column categories. If the angle between two arrows is acute, there is a strong association between the corresponding row categories, column categories, or row and column categories. If the angle is very small, it suggests these objects are associated. Ninety-degree angles indicate no relationship ,and angles near 180 degrees indicate negative associations. The closer an arrow is to an axis, the greater is the contribution of its object on that axis relative to the other axis. If an arrow is halfway between two axes, its object contributes to both equally. If an arrow is far from the center of the map, it suggests a relatively strong association with the dimension. Short arrows indicate either a weak association or no association.

Figure 2 shows the two-dimensional configuration of words. Dimension 1 is represented by the horizontal axis and dimension 2 by the vertical axis. Along Dimension 1, we can observe that “release”, “news”, and “suncor” are furthest away from the origin. Horizontal Dimension 1 of the solution reflects the difference between the topics regarding this corporate content. Placed in the upper half are “bpstats”, “icymi”, “outlook”, “growth”, “demand”, “want”, “think”, “keep”, “story”’, “next”, “know”, and “year”. These words suggest a focus on the future (“next”, “year”, and “growth”) and express that the company builds a story or narrative and expresses its desires (“keep”, “want”, and “demand”). Placed in the lower half are “head”, “starace”, “francesco”, “enel”, “enelgreenpower”, and “talk”. These words express a corporate orientation based on stats and brand names, including sustainable adjectives (e.g., “enelgreenpower”) and the name of a CEO.

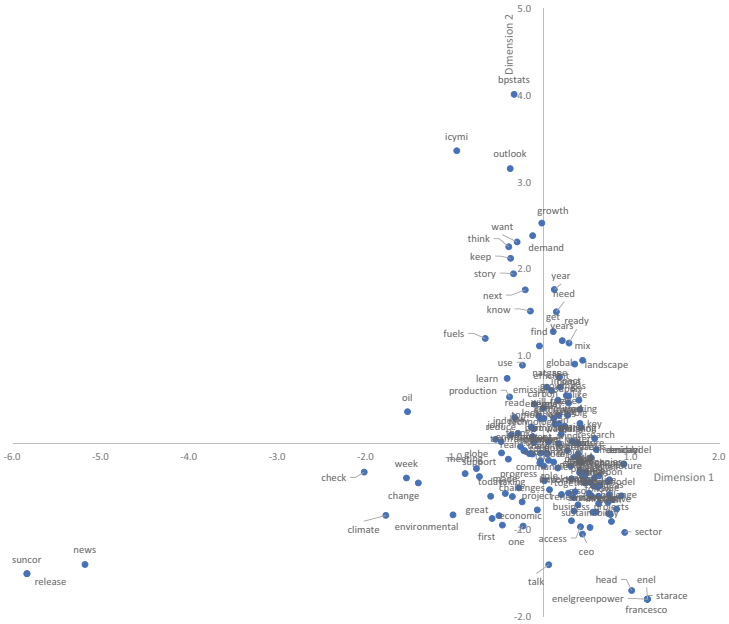

Figure 2: Two-dimensional configuration of words.

Figure 3 shows the two-dimensional configuration of companies. The arrows corresponding to “Suncor”, “BP”, and “ENEL” are far from the center of the configuration, suggesting differences in tweet content between Suncor, ENEL, and BP. Suncor is placed in the lower left part of the configuration. Enel is placed in the upper half ,and BP is placed the lower half along the vertical axis. Suncor, Enel, and BP posted different topics on Twitter compared to other companies. If the angle between these three companies and others is very small, it suggests that the topics of the companies’ tweets are related. If there is a short arrow next to a company, it suggests that there is a weak association or no association. If a company’s arrow is farther from the center of the map, it suggests that the company has a relatively strong association with the dimension. On the other hand, if a company’s arrow is nearer to the center of the map, it suggests that the company has a relatively weak association with the dimension.

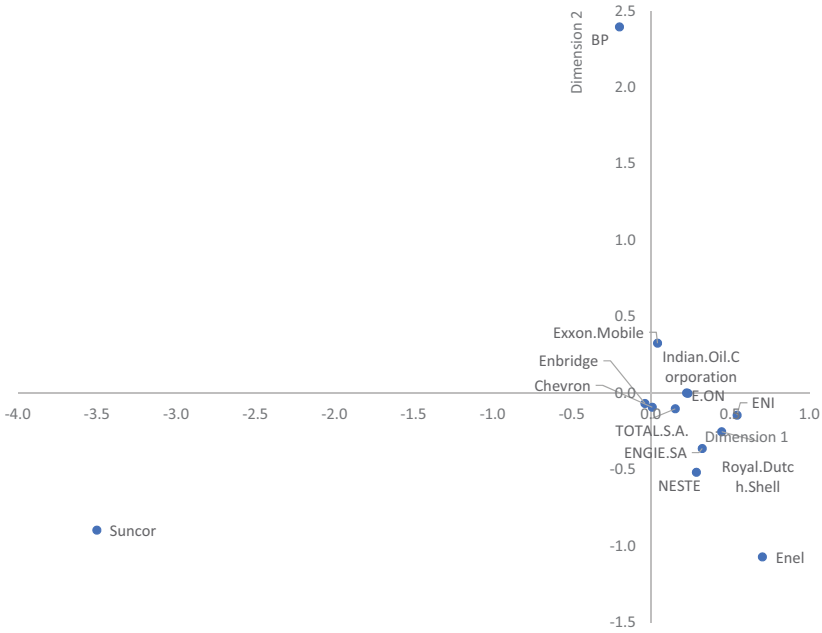

Figure 3: Two-dimensional configuration of companies.

The results indicate that these three companies are trying to differentiate the nature of their Twitter communication regarding sustainability issues. Among the three companies, BP focuses the most on social and environmental issues. The angles between "Enel", "ENGIE.SA"', and "NESTE" are small, so the tweet content of these companies is similar. Furthermore, the arrows for "ENGIE.SA" and "NESTE" are shorter than the arrows for "Enel", indicating a weaker association. "Exxon.Mobile", "ENGIE.SA", and "NESTE" are located farther from the center, indicating that their Twitter topics differ from those of companies closer to the center. "Enbridge", "Chevron", "TOTAL.S.A.", "E.ON", "Indian.Oil.Corporation", and "Royal.Dutch.Shell" are located close to the center, and these companies posted about similar topics on Twitter. However, the associations between the characteristics of their Twitter topics are weak. This is due to the high level of diversity of words used in their topics, as well as the fact that their topics do not always relate to the same words.

It is important to look beyond the distances between companies and words when interpreting associations between them. Let us imagine a line connecting a

company and word with the origin on the graph. The sharper the angle between the company and word, the stronger the relationship between them. In such cases, the company might frequently post about topics using the same words. The topics of sustainability-related tweets from "Suncor", "Enel", and "BP" are very different from those of other companies in the energy sector. The angles between "Suncor" and words such as "news" and "release" are the smallest, which means that Suncor uses these words most frequently, denoting a clear corporate orientation. Additionally, Suncor uses words such as "climate", "check", "change", "week", and "environmental". "Enel" often uses words like "francesco", "starace", "elgreenpower", and "head", as well as "sector", "sustainability", "ceo", "access", and "business". Therefore, it can be concluded that ENGIE SA, NESTE, and Enel have common communication tendencies. This may indicate that all three companies aim to position themselves as sources of clean energy, albeit with a strong level of corporate messaging as well. However, ENGIE SA and NESTE have weaker associations than those of Enel, since ENGIE SA and NESTE are located closer to the center of the graph.

BP frequently uses words such as "bpstats", "icymi", "outlook", "growth", "demand", "want", "think", "keep", "story", "next", and "know". These words have small angles with BP and are placed far from the center, so they are strongly associated with BP's communication strategy. Exxon Mobile and BP have similar associations. However, those of Exxon Mobile are weaker than those of BP, since Exxon Mobile was located closer to the center. Other companies located close to the center of the graph use various words that are also located near the center. However, the associations between these companies and words are weak because the lengths of the arrows from the center are short.

4 Conclusion

These findings reveal that the top firms in the energy industry generally implement a passive, ceremonial, and symbolic style of communication via Social Media; in this case, on Twitter. Applying R text mining by means of word clustering verifies its utility in the analysis of textual datasets generated from online social network conversations. The use of word clustering in the present study shows that the statistical linguistic analysis assessment provides more significant results than more traditional content analysis methods, such as keyword or hashtag monitoring.

The dimension reduction method shows that the entire discourse is passive and defensive rather than performative or proactive, as it embraces mainly corporate issues. With respect to sustainability/CSR topics, the discourse is very general and one-dimensional, referring to such universal themes as future-oriented, renewable or sustainable. It refers to a corporate perspective and the sector's future. As shown by the data, the Twitter narrative of the most important energy companies is wide-ranging in terms of the diversity of topics that guide online conversations on the social network. However, more profound references, or clear and practical examples of such issues are absent. The online discourse on sustainability in the energy sector does not include specific expressions that would textually demonstrate genuine social and environmental activity by the companies. In spite of the broad spectrum of topics covered by the energy sector's tweets, the tweets have a symbolic and corporate/brand-oriented character as well as a sectorial approach, underlying the brand statements on corporate positions with a tactical rather than a strategic perspective.

Based on the data analyzed, the CSR communication profile is oriented toward the central sustainability concept as well as innovation and the future, as shown by the abundance of various keywords and hashtags referring to related issues. There is mutual isomorphism among companies within the energy sector in their online communication strategies. We observe that companies tend to imitate one another's Twitter communications regarding the market and corporate positioning. This indicates that, in certain situations, companies' Twitter discourses reflect similar communication patterns.

A strategic approach to sustainability may act as another influencing factor. However, further studies in this area are needed to confirm the initial results, including both case studies and broader analyses that take into account critical events affecting business communication. In general, the results of the analysis indicate deeper similarities in communication as a consequence of the companies' perceptions of the external environment (including the competitors). Isomorphism may result from seeking an effective narrative in the face of a crisis, or from the company's positioning of itself within the market. This pattern may be a subject for future research: To identify the degree of connection between the strategic orientation of the companies with respect to sustainability and the semantic content of their Social Media communications.

Nevertheless, sustainability-related content is less prevalent than corporate and sectorial talk, which is marked by the use of the word "energy" as

both a keyword and a hashtag. Therefore, we can conclude that Twitter communication focused on the sustainability of the energy sector tends toward homogeneity, and follows symbolic approaches to legitimization as a result of isomorphic pressures. The results reveal an isomorphic trend among energy companies to follow similar Twitter communication patterns on sustainability. This study shows that organizations are susceptible to mimetic isomorphism in their Twitter communication, especially when constructing their discourse and narrative under formal and informal pressures. Future studies should re-examine the criteria for selecting words for analysis and assess the use of alternative dimensional reduction models.

Acknowledgements We express our gratitude to the anonymous referees for their valuable reviews. This work was supported by a Grant-in-Aid for Scientific Research (C) (No. 16K00052) from the Japan Society for the Promotion of Science.Data was collected thanks to financial support of Polish National Science Center NCN (grant No 2012/05/D/HS4/01177).

References

Abitol A, Lee SY (2017) Messages on CSR-Dedicated Facebook Pages: What Works and What Doesn't. Public Relations Review 43(4):796–808. DOI: 10.1016/j.pubrev.2017.05.002.

Amaeshi KM, Osuji OK, Nnodim P (2008) Corporate Social Responsibility in Supply Chains of Global Brands: A Boundaryless Responsibility? Clarifications, Exceptions and Implications. Journal of Business Ethics 81(1):223–234. DOI: 10.1007/s10551-007-9490-5.

Anderson A, Warren L (2011) The Entrepreneur as Hero and Jester: Enacting the Entrepreneurial Discourse. International Small Business Journal 29(6):589–609. DOI: 10.1177/0266242611416417.

Ashcraft KL, Kuhn TR, Cooren F (2009) Constitutional Amendments: "Materializing" Organizational Communication. The Academy of Management Annals 3(1):1–64. DOI: 10.5465/19416520903047186.

Ashworth R, Boyne G, Delbridge R (2009) Escape from the Iron Cage? Organizational Change and Isomorphic Pressures in the Public Sector. Journal of Public Administration Research and Theory 19(1):165–187. DOI: 10.1093/jopart/mum038.

Bendixen MT (1995) Compositional Perceptual Mapping Using Chi-Squared Trees Analysis and Correspondence Analysis. Journal of Marketing Management 11(6):571–581. DOI: 10.1080/0267257X.1995.9964368.

Brammer S, Jackson G, Matten D (2012) Corporate Social Responsibility and Institutional Theory: New Perspectives on Private Governance. Socio-Economic Review 10(1):3–28. DOI: 10.1093/ser/mwr030.

Burns TR (2011) The Sustainability Revolution: A Societal Paradigm Shift Ethos, Innovation, Governance Transformation. Sociologisk Forskning 48(3):93–108, Sveriges Sociologförbund (Swedish Sociological Association). ISSN: 0038-0342, URL: http://www.jstor.org/stable/41698145.

Campbell JL (2007) Why Would Corporations Behave in Socially Responsible Ways? An Institutional Theory of Corporate Social Responsibility. Academy of Management Review 32(3):946–967. DOI: 10.5465/amr.2007.25275684.

Castka P, Balzarova MA (2008) ISO 26000 and Supply Chains – On the Diffusion of the Social Responsibility Standard. International Journal of Production Economics 111(2):274–286. DOI: 10.1016/j.ijpe.2006.10.017.

Cornelissen JP, Durand R, Fiss PC, Lammers JC, Vaara E (2015) Putting Communication Front and Center In Institutional Theory and Analysis. Academy of Management Review 40(1):10–27. DOI: 10.5465/amr.2014.0381.

Couldry N (2012) John Wiley & Sons, Hoboken (USA). ISBN: 978-0-745639-20-8.

Dacin MT (1997) Isomorphism in Context: the Power and Prescription of Institutional Norms. The Academy of Management Journal 40(1):46–81. DOI: 10.5465/257020.

Deephouse DL (1996) Does isomorphism Legitimate? The Academy of Management Journal 39(4):1024–1039. DOI: 10.2307/256722.

DiMaggio PJ, Powell WW (1983) The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. American Sociological Review 48(2):147–160. DOI: 10.2307/2095101.

Dowling J, Pfeffer J (1975) Organizational Legitimacy: Social Values and Organizational Behavior. The Pacific Sociological Review 18(1):122–136. DOI: 10.2307/1388226.

Fellows I (2018) Word Clouds. URL: https://cran.r-project.org/web/packages/wordcloud/. R package.

Frame A, Brachotte G (2015) Le Tweet Stratégique: Use of Twitter as a PR Tool by French Politicians. Public Relations Review 40(2):278–287. DOI: 10.1016/j.pubrev.2014.11.005.

Glover JL, Champion D, Daniels KJ, Dainty AJD (2014) An Institutional Theory Perspective on Sustainable Practices Across the Dairy Supply Chain. International Journal of Production Economics 152:102–111. DOI: 10.1016/j.ijpe.2013.12.027.

Gomez LM, Chalmeta R, Sosa-Varela JC (2015) Usage and Importance of Social Media for Corporate Communication and Stakeholder Dialogue. In: Thriving in a New World Economy, Developments in Marketing Science: Proceedings of the Academy of Marketing Science, pp. 56–59. Springer, Cham (Switzerland), Plangger K (ed.). DOI: 10.1007/978-3-319-24148-7_18.

Greenacre MJ (1984) Theory and Applications of Correspondence Analysis. Academic Press, Cambridge (USA). ISBN: 978-0-122990-50-2.

Hannan MT, Freeman J (1997) The Population Ecology of Organizations. American Journal of Sociology 82(5):929–964. DOI: 10.1086/226424.

Haunschild R, Miner AS (1997) Modes of Interorganizational Imitation: The Effects of Outcome Salience and Uncertainty. Administrative Science Quarterly 42(3):472–500. DOI: 10.2307/2393735.

Heugens PMAR, Lander MW (2009) Structure! Agency! (And Other Quarrels): A Meta-Analysis of Institutional Theories of Organization. The Academy of Management Journal 52(1):61–85. DOI: 10.5465/AMJ.2009.36461835.

Higgins C, Stubbs W, Milne M (2018) Is Sustainability Reporting Becoming Institutionalised? The Role of an Issues-Based Field. Journal of Business Ethics 147(2):309–326. DOI: 10.1007/s10551-015-2931-7.

Hutzschenreuter T, Lewin AY, Dresel S (2011) Governance Modes for Offshoring Activities: A Comparison of US and German Firms. International Business Review 20(3):291–313. DOI: 10.1016/j.ibusrev.2011.01.007.

Jennings PD, Zandbergen PA (1995) Ecologically Sustainable Organizations: An Institutional Approach. Academy Management Review 20(4):1015–1052. DOI: 10.2307/258964.

Joseph C, Taplin R (2012) Local Government Website Sustainability Reporting: A Mimicry Perspective. Social Responsibility Journal 8(3):363–372. DOI: 10.1108/17471111211247938.

Ketokivi MA, Schroeder RG (2004) Strategic, Structural Contingency and Institutional Explanations in the Adoption of Innovative Manufacturing Practices. Journal of Operations Management 22(1):63–89. DOI: 10.1016/j.jom.2003.12.002.

Lamin A, Zaheer S (2012) Wall Street vs. Main Street: Firm Strategies for Defending Legitimacy and Their Impact on Different Stakeholders. Organzation Science 23(1):47–66. DOI: 10.1287/orsc.1100.0631.

Lammers JC, Barbour JB (2006) An Institutional Theory of Organizational Communication. Communication Theory 16(3):356–377. DOI: 10.1111/j.1468-2885.2006.00274.x

Lieberman MB, Asaba S (2006) Why Do Firms Imitate Each Other? Academy of Management Review 31(2):366–385. DOI: 10.5465/amr.2006.20208686.

Lindblom CK (2010) The Implications of Organizational Legitimacy for Corporate Social Performance and Disclosure. In: Gray R, Bebbington J, Gray S (eds.), Social and Environmental Accounting, SAGE Publications, Los Angeles (USA), SAGE Library in Accounting and Finances, Vol. 2. ISBN: 978-1-848601-69-7.

Livingstone S (2009) On the Mediation of Everything: ICA Presidential Address 2008. Journal of Communication 59(1):1–18. DOI: 10.1111/j.1460-2466.2008.01401.x

Macnaghten P, Urry J (2005) Alternatywne przyrody. Nowe myślenie o przyrodzie i społeczeństwie. Baran B (ed.), Wydawnictwo Naukowe Scholar, Warszawa (Poland). ISBN: 83-7383-161-4.

Maguire S, Hardy C (2009) Discourse and Deinstitutionalization: The Decline of DDT. The Academy of Management Journal 52(1):148–178. DOI: 10.5465/AMJ.2009.36461993.

Maignan I, Ralston DA (2002) Corporate Social Responsibility in Europe and the U.S: Insights from Businesses' Self-Presentations. Journal of International Business Studies 33(3).

Matten D, Moon J (2008) "Implicit" and "Explicit" CSR: A Conceptual Framework for a Comparative Understanding of Corporate Social Responsibility. Academy of Management Review 33(2):404–424. DOI: 10.5465/amr.2008.31193458.

Misani N (2010) The Convergence of Corporate Social Responsibility Practices. Management Research Review 33(7):734–748. DOI: 10.1108/01409171011055816.

Parente PHN, Mota AF, Cabral ACdA, Santos SMd, Brandão IdF (2014) Disclosure and Institutional Sustainability Theory: An Investigation in the Electricity Sector. RACE- Revista de Administraçâo, Contabilidade e Economia 13(3):889–917. URL: https://portalperiodicos.unoesc.edu.br/race/article/view/4202.

Peters BG (2012) Institutional Theory in Political Science: The New Institutionalism, 3rd edn. Continuum, London (United Kingdom). ISBN: 978-1-441130-42-6.

Saffer AJ, Sommerfeldt EJ, Taylor M (2013) The Effects of Organizational Twitter Interactivity on Organization-Public Relationships. Public Relations Review 39(3):213–215. DOI: 10.1016/j.pubrev.2013.02.005.

Sahlin-Andersson K, Engwall L (2002) Sahlin-Andersson K, Engwall L (eds.), Stanford University Press, Stanford (USA). ISBN: 978-0-804741-97-2.

Scott RW (2001) Institutions and Organizations, 2nd edn. Foundations for Organizational Science, SAGE Publications, Los Angeles (USA). ISBN: 978-0-761920-01-4.

Sethi SP (1975) Dimensions of Corporate Social Performance: An Analytical Framework for Measurement and Evaluation. California Management Review 17(3):58–64, Schools of Business Administration, University of California, Berkley (USA). DOI: 10.2307/41162149.

Sethi SP (1979) A Conceptual Framework for Environmental Analysis of Social Issues and Evaluation of Business Response Patterns. Academy of Management Review 4(1):63–74. DOI: 10.2307/257404.

Sinha V, S.Subramanian K, Bhattacharya S, Chaudhuri K (2012) The Contemporary Framework on Social Media Analytics as an Emerging Tool for Behavior Informatics. HR Analytics And Business Process Management 17(2):65–84.

Smitko K (2019) Donor Engagement Through Twitter. Public Relations Review 38(4):633–635. DOI: 10.1016/j.pubrev.2012.05.012.

Uzunoğlu E, Türkel S, Akyar BY (2017) Engaging Consumers Through Corporate Social Responsibility Messages on Social Media: An Experimental Study. Public Relations Review 43(5):989–997. DOI: 10.1016/j.pubrev.2017.03.013.

Vaara E, Tienari J (2008) A Discursive Perspective on Legitimation Strategies in Multinational Corporations. Academy of Management Review 33(4):985–993. DOI: 10.5465/amr.2008.34422019.

Venables WN, Ripley BD (2002) Modern Applied Statistics with S, 4th edn. Springer, New York. ISBN 0-387-95457-0.

Verčič AT, Sriramesh K, Verčič D (2018) Introduction to the Special Section: CSR in Hypermodern Times. Public Relations Review 44(4):431–432. DOI: 10.1016/j.pubrev.2018.08.006.

Yawar SA, Kauppi K (2018) Understanding the Adoption of Socially Responsible Supplier Development Practices Using Institutional Theory: Dairy Supply Chains in India. Journal of Purchasing & Supply Management 24(2):164–176. DOI: 10.1016/j.pursup.2018.02.001.