Social Media Content Analysis in the Higher Education Sector: From Content to Strategy

Luciana Oliveira, ISCAP / IPP Polytechnic of Porto, Porto, Portugal & INESC TEC, University of Porto, Porto, Portugal

Álvaro Figueira, CRACS, INESC TEC, University of Porto, Porto, Portugal

ABSTRACT

Social media has become one of the most prolific fields for interchange of multidisciplinary expertise. In this paper, computer science, communication and management are brought together for the development of a sound strategic content analysis, in the Higher Education Sector. The authors present a study comprised of two stages: analysis of SM content and corresponding audience engagement according to a weighted scale, and a classification of content strategies, which builds on different noticeable articulations of editorial areas among organizations. Their approach is based on an automatic classification of content according to a pre-defined editorial model. The proposed methodology and research results offer academic and practical findings for organizations striving on social media.

Keywords:
Content Analysis, Editorial Model, Higher Education Sector, Social Media, Strategy, Text Categorization

INTRODUCTION

The undeniable growth of social media environments has been introducing profound changes in society and in the communication management landscape. Though social media impacts are still subject of research in a wide variety of fields, in what organizations are concerned, two main aspects are consistently revealed throughout literature. The first relates to the (each day not so) new empowered role of millions of social media users, co-creators, active voices and active influencers, which organizations fail to understand and engage with. The second aspect relates to the fact that organizations are still "reluctant or unable to develop strategies and allocate resources to engage effectively" (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011, p. 242) in these environments, thus appear to be ignoring or mismanaging "the opportunities and challenges presented by creative consumers".

DOI: 10.4018/IJWP.2015040102

In fact, organizations rush into social media environments following the trend to create a social presence in multiple channels and/or to explore their potential (Berthon et al., 2012; Fournier, Avery, Fournier, & Avery, 2011; Gensler, Völckner, Liu-Thompkins, & Wiertz, 2013; Habibi, Laroche, & Richard, 2014; Hanna et al., 2011; Kaplan & Haenlein, 2010; Kirtiş & Karahan, 2011), without previously defining a clear strategic approach, which should include, for instance, clear insights on their target audience and a content editorial plan and calendar, in order to foster the achievement of the overall business objectives.

Following the trend to join social media environments, reaching for and aiming at mediatization (Zerfass, Fink, & Linke, 2011), organizations are also lacking monitoring practices and benchmarking frameworks.

As a result, an editorial model for the development of social media strategies was constructed and its articulation should provide the basis for persistent monitoring and well suited benchmarking practices. It also provides sector positioning assessment and measurement of the return on investment (ROI) on social media. The application of this model is supported by a text-mining and categorization methodology, which is also detailed, and which allowed to reveal the sector's investments allocation in communication areas and a three-fold overall classification of social media strategies being pursued. Since this is a particularly complex challenge for atypical markets, namely organizations providing higher education services, presented research focuses mainly on social media communication management in public Higher Polytechnic Portuguese Education Institutions (HPPEI).

This paper is organized as follows: in the next two sections the concept of social media strategy in Higher Education (HE) is introduced and related work on social media content analysis in this sector is presented. The proposed editorial model for the development and evaluation of social media strategies is then explained, followed by the research design and scope, which supports the two final sections of results obtained in content analysis and on the classification of content strategies. The paper is concluded with discussion on its main contributions, methodology applicability and limitations.

SOCIAL MEDIA STRATEGIES IN HEI

The Higher Education Sector (HES) has become a global competitive environment, where the decreasing number of student enrolments and the decreasing financial government support has been raising unforeseen external challenges. In fact, Higher Education Institutions (HEI) have been pushed to seek for additional financial funding and to devote a large amount of efforts to ensure brand awareness and distinctiveness. In this context, the role of social media has become a fertile field, and at least for that latter one, it has become essential.

Acknowledging a broad range and diversity of organizational publics raises distinct communication and management needs. Social media strategies aimed at organizations with such a diversity of stakeholders' expectations, service distinctiveness, societal expected intervention and corresponding external pressure are lacking research. Thus, it is important to research, reveal, systematize and bring forward modes of intervention that can provide the balance between their institutional and transactional needs, in order to ensure their survival and competitive potential.

According to this perspective, the concept of social media strategy proposed by Lardi (2013), needs to be aligned and framed in a holistic integrated organizational communication management model, such as the one proposed by Cornelissen (2014), Moss (2011), Kunsch (2003) and Robinson (2011), for instance. According to Moss and DeSanto (2011), a well-articulated communication strategy should provide a clear indication of the overall direction, purpose and intended outcome

of the communication function, expressed in terms of a unifying big idea that will run through and help integrate all communication activities, and like all other functional strategies, it needs to be aligned with and to support the higher organizational strategic management.

RELATED WORK

Despite the popularity that social media has been gaining in academic research, there are very few contributions related to social media content analysis in the public Higher Education Sector (HES) and, to the best of our knowledge, none related to a strategic editorial approach to content analysis. Lack of research is also stated by Qi and Mackie (2014), who report on a wider spectrum of social media research on educational purposes.

Chauhan and Pillai (2013) report on content strategy as tool to assess brand awareness and customer engagement in leading HEI, in India. Authors present several remarkable contributions which are essential as a ground basis for the research presented in this paper, namely: content is key to social media performance, structure / presentation and informative potential is key to obtain audience engagement; content strategy may constitute a crisis management and prevention tool; motivational drivers, non-transactional behavior and emotional messages are highlighted as important components of content and engagement drivers; and the need for organizations to assume the role of community moderators is stated.

Also, although Chauhan and Pillai (2013) use rigorous method (Netnography) and a sophisticated statistical content analysis, data collection and categorization are exclusively conducted by humans, which is not a viable nor a cost effective solution for organizations seeking persistent social media monitoring. An alternative methodology, in the field of HEI, is proposed by Lai and To (2015), who suggest a grounded theory based on computer-aided lexical analysis with statistical and graphical methods to identify the key dimensions of the topics, while minimizing human errors, as well as coding and categorization biases. In this case, research is based on identifying keywords' frequency through lexical software and applying exploratory factor analysis in order to grouping them into several factors, thus resulting on a map of concept clusters.

The main limitation of the text mining techniques proposed by Lai and To (2015), methodology is that, when considered isolated / not integrated in other more robust and informative processes / technologies, they are not comprehensive, thus the authors suggest this integration and the manipulation of the techniques by, for instance, reputation managers. Social media content analysis supported by text mining techniques has been growing increasingly popular in several areas of knowledge (Andersen, 2007; He, Zha, & Li, 2013; Hollebeek, Glynn, & Brodie, 2014; Rebholz-Schuhmann, Kirsch, & Couto, 2005; Russell, 2013; Saxton & Waters, 2014; Zappavigna, 2012). This work is no exception to that trend, thus machine learning and six text mining classifiers were used to categorize social media content according to the proposed editorial model.

The integration of text mining techniques in research design is presented in the following section. Research scope is also defined.

AN EDITORIAL MODEL FOR CONTENT STRATEGIES

The construction of a social media strategy should be supported by the definition of the most relevant content areas of focus, which should derive from a set of guiding principles, that are sector specific, namely: the relevance of HEI mission towards society, the specifics of the educational service, a holistic approach to communication management as well as the two-way symmetrical and dialogical nature of social media environments and the spread of user-generated content by

empowered users. These principles are aimed at establishing the most basic level of understanding / perception towards the specific characteristics of the educational sector. The Social Media strategy, as a corporate development function, requires the implementation of context driven communication processes, aimed at one or more public profiles (or 'personas').

In order to address the above mentioned principles an editorial model was developed, and is composed of the seven main content areas of focus for the HES (Table 1), which can be articulated by organizations in the planning, development and assessment of their social media strategy. The editorial model identifies seven main areas of message purpose and is aimed at answering the question: what is the main purpose of this message?

Though this paper is not aimed at fully describing the underlying conceptual approach that supports each of these content areas, it is relevant to highlight how the organizational publics are framed in this model.

The proposed model aims to offer a very wide perspective of content domains for HEI's communication managers / practitioners, propelling a holistic strategic analysis and planning. However, the best outcome of this process is a fine tuned balance of the content areas that are key to each organization's individual corporate strategy. Considering Mintzberg's (1994) terminology, the proposed model serves as a tool for strategic planning (analysis), which also requires strategic thinking (synthesis), when considering the key elements of the organizations' mission that will be nurtured by social media. Therefore, this model comprehends the existence of several optimal balances, in terms of diversity and intensity, of content areas.

Consequently, the design of a social media strategy should be as balanced as possible, though aimed at the specific publics of each of the social media channels in which the organization chooses to engage in. Though literature refers potential students, current students, and alumni

Table 1. Social media editorial model for HPPEI

Education	Research	Society	Identity	Administration	Relationship	Information
- Promotes higher education courses (educational offer) - Promotes complementary training (internal or external)	- Informs on and / or calls for participation in: congresses, seminars and other scientific meetings - Promotes / informs on internal and external research results / awards - Promotes / informs on internal and external publications (papers, articles, books, proceedings, etc.)	-Promotes / informs on organizational partnerships and contracts and patents - Promotes employability, streaming placement offers and career opportunities -Promotes / informs on knowledge / technology transfer - Promotes other organizations' initiatives / performance (partners and other relevant stakeholders) - Promotes demonstrations, exhibitions and showcases, conducted by students or faculty (emphasis on competencies and societal integration)	- Institutional events (celebrations, awards and tributes, graduation ceremonies, etc.) - Students, faculty and staff honorable mentions - Institutional promotion, advertising (identity, image, reputation) -Corporate Social Responsibility initiatives - Institutional clipping - Participation / representation in fairs and exhibitions	- Informs on deadlines and administrative processes - Informs on procedures and admissions - Promotes and informs on support services (goals, contacts, working hours, etc.)	- Fosters conversation - Requires opinions - Introduces current internal, external, societal or academic issues propelling audience involvement - Boosts emotional connection between organization and publics (greetings, humor, sympathy, motivation, etc.)	- Streams external relevant information, news and regulations related to academic areas, political and societal issues (economic and social impact) - Informs on recreational and cultural initiates with no particular connection to schools' scientific areas (concerts, sports events, etc.)

as the main social media publics for the HES (Qi & Mackie, 2014) multiple communication approaches, within the same organizational communication plan, may be necessary, in order to address other publics' profiles.

Finally, the proposed editorial model can be used by organizations that are at different social media maturity stages: planning to adopt, experimenting without a plan, integrating/trying to strategically integrate social media, etc. Our intent is that the editorial model serves both the emergent and the deliberate social media strategies (Mintzberg, 1987), that is: strategies that are entirely pre-planned and deliberate paths or strategies that emerge from unexpected opportunistic developments or by trial and error.

RESEARCH DESIGN AND SCOPE

Research was conducted on the total population of Higher Polytechnic Portuguese Education Institutions (HPPEI), using a quantitative content analysis methodology (Riff, Lacy, & Fico, 2014, p. 3).

The full list of HPPEI was retrieved from the DGES – Direção Geral de Ensino Superior (General Higher Educational Management service) website. A total of 116 agents was initially considered, which included polytechnic schools integrated into polytechnic institutes and polytechnic schools integrated into universities. The number of agents was then reduced to 94 in order to include only the schools providing educational services, disregarding the polytechnic institutes and universities (managing entities).

The scope definition also included the measurement of HPPEI's social media channels adoption rates, in order to include the most relevant channels, and only those that had been in use at least since the 1st of September 2013, with the intent of extending the analysis to a full school year. In order to support research in reliable sources, the study considered only the social media networks mentioned on the HPPEI's official websites. This method aimed at ensuring that the social media webpages under analysis were actually managed by the HPPEI, instead of other internal or external stakeholders, such as students, employees (administrative or faculty) or alumni on their own.

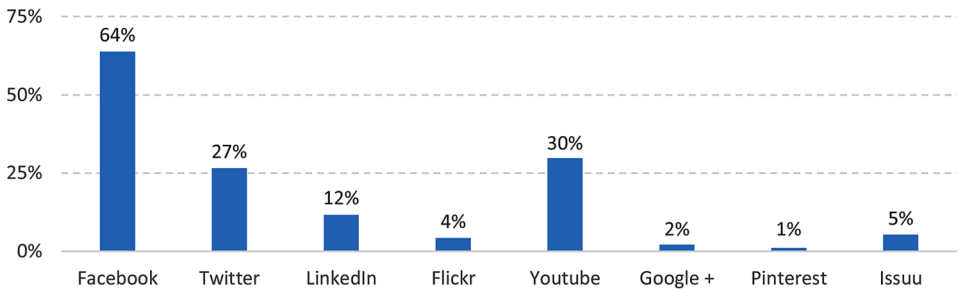
The overall criteria for scope definition was defined as follows:

- Include the most relevant social media network/channel, according to which Facebook proved to be the most representative social media website, with an adoption rate of 64% in the sector, as illustrated in Figure 1;
- Include only Facebook pages, discarding Facebook profiles (3 profiles) that organizations were using to interact with their audience, in order to assure metrics' cohesiveness;
- Include only Facebook pages created before 1st September 2013, in order to sustain research in a set of agents that could provide a full school year of consistent data;
- Include Facebook pages managed by organizations providing the actual educational service, disregarding all top-level corporate / political affiliation organizations, in order to assure audience homogeneity in research.

The above criteria were used during the preliminary analysis conducted on the aforementioned 94 agents' websites and social media pages, resulting on a set of 43 agents which met these requirements.

In Figure 1 it is also possible to observe the total diversity of social media tools used by HPPEI. After analyzing each one of these channels it was possible to verify that, though 30%

Figure 1. Channel usage for social media



of the agents have Youtube accounts, they are mostly used for video storage and publication, but not for interaction with users. The same applies to Issuu, Pinterest and Flickr. Twitter accounts register very low levels of tweets, most of which include the exact same content that is publish on Facebook pages (abbreviated Facebook posts), which leads us to conclude that this network is either still in early adoption stages, in this Portuguese sector, or its potential is still being assessed by HEI.

Data Retrieval

The following stage consisted of retrieving and classifying all messages posted by HPPEI on Facebook. The aforementioned criteria, lead us to 43 eligible HEI. To retrieve Facebook posts from all these HEI two methods were used in parallel: 1) an in-house made system, specially built for the purpose, using the Java language, and the available Facebook API, and; 2) a third-party software ("PowerQuery") able to retrieve information from social networks into an Excel format.

Both systems were fed with the 43 agent's Facebook Page Id's and the following post fields were retrieved: PostId, Message, Link, Name, Description, Caption, #Likes, #Comments and #Shares. The two systems retrieved the same number of posts (15444), during the entire school year, which consolidated the confidence about the size of the returning set.

Categorization of Messages

The next phase consisted on performing the classification of the 15444 posts according to the previously defined editorial model.

Published research concerning text mining is not scarce. Several text classifiers have been proposed over the last three decades with excellent performing results. Curiously, nowadays this is again a very hot research topic due to new special properties of texts being examined. We are referring to texts posted in social networks, which are typically very short, have abundance of smileys, inclusion of links, many punctuation signs, etc. These special characteristics make text classification, again, a difficult task, and new techniques are being explored. In the last decade we have observed that research has been incorporating these new features and creating, consistently, better classification models, especially for classification under supervised training.

On a first stage, we presented 350 posts to be classified by a human specialist, which roughly took 1.5 hours to complete. Using this value, we can extrapolate that classifying the whole set would take more than one week of office time work. Clearly, the "human classification" is not a viable solution for a large-enough set of posts. If we used multiple specialists to do it in parallel, the time would be proportionally reduced but, we would face another problem: the consistency of

the classification. Moreover, in an era of exponential expansion of digital social networks, human classification of posts is clearly a far too heavy endeavor to be made by hand. As a consequence, the search for an automatic classification technique of the posts was needed.

For this step we decided to use an ensemble of six of the most promising, and prominent, classifiers: Support Vector Machines (Crammer & Singer, 2002), Random Forests (Breiman, 2001), LogiBoost (Friedman, Hastie, & Tibshirani, 2000), K-Nearest Neighbours (Altman, 1992), 1992), MultiLayer Perceptrons (Freund & Schapire, 1999) and Deep Neural Networks (Collobert & Weston, 2008) and (Bengio, Courville, & Vincent, 2013). All the algorithms were used through public and open source libraries ("caret" and "h2o"), available for the R programming language, which we used to combine the results.

A second stage consisted on running the set of 350 manually classified posts through the classifiers for training, and then computing the respective accuracy of the automatic classification. The result of using a 10-fold cross-validation proved the techniques achieved results above 68% of accuracy. As a second step, we gave the classifiers a bigger set of 512 manually classified posts (by the same human specialist) for retraining, and we recomputed the new accuracy. We got an improvement of only 3% on average for the six algorithms, from an increase of 46% of the sample size. Therefore, we didn't continue to increase the training set to avoid over-fitting.

Hence, we then run the whole set of 15409 posts on our six trained classifiers to obtain a predicted category for each post, by each classifier. Finally, we used the mode of these six algorithms as the final result, i.e., we used the prevailing category of the six-set as the final predictive category.

As a final data preparation step we had to discard 120 posts that had no mode, and 185 posts that were published out of the research pre-study time-frame. Hence, this retrieval and classification procedure returned 15104 classified posts.

RESULTS

Results presented in this section derive from statistical analysis conducted on the data obtained, according to the above mentioned research design. Inferences are also derived, according to which we analyze content as an end in itself, as an assumed consequence or as evidence of the social and organizational context (Riff et al., 2014).

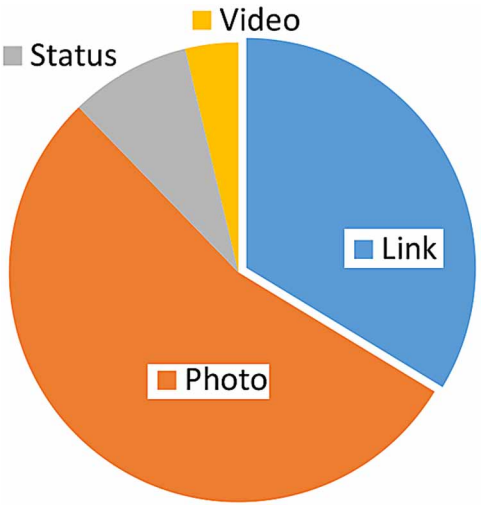
Two complementary analysis are presented in the following sections: a sector content analysis, in which we analyze content type and corresponding audience engagement, and a classification of content strategies, which builds on different noticeable articulations of editorial areas among organizations.

Content Analysis

In this section we focus on content type and corresponding engagement. Facebook 'likes,' 'shares' and 'comments' are used as manifest variables for publics' engagement according to a weighted scale.

Our resulting database is composed by the posts produced by the 43 HEI, forming a matrix of 27 columns (attributes) and 15104 rows (records). Each record represents a "message" posted by some specific HEI. Analyzing our collected posts, we count the number of posts of each type (Facebook distinguishes between four types of posts: 'link,' 'photo,' 'status' and 'video'). As is illustrated in Figure 2, the post type 'photo' is the dominant one, and 'link' is the second one, both by a large difference to the other types.

Figure 2. Distribution of post type in the sector



In Table 2 we deepen the analysis and present a cross-checking view between the available post types, and the corresponding resulting engagement from each post type.

We observe that the number of 'likes' per post is always bigger than any of the other two engagement variables ('shares' and 'comments'). On the other hand, the number of 'shares' per post is always bigger then the number of 'comments' per post.

According to the above results, the best posting type to create engagement is the 'video,' while the second best is the 'photo,' followed by the 'link.' Moreover, the 'video' is almost consistently the best post type to get any type of engagement (apart from the 'comments' on photos).

Finally, we observe that the best engagement ratio is achieved through 'likes' on posts of type 'video.' The second best is through 'likes' on posts of type 'photo,' and the worst is 'comments' on 'links.'

It is, however, curious to observe that each HEI (on average) posts 'photos' one order of magnitude (2461) above the number of 'videos' (211). Therefore, not using the most promising mean to generate more engagement.

We can also refer to "totals" amongst all HEI to find that, during a one-year period, on average, each HEI makes 358 posts and gets 3407 responses (which is about one order of magnitude higher).

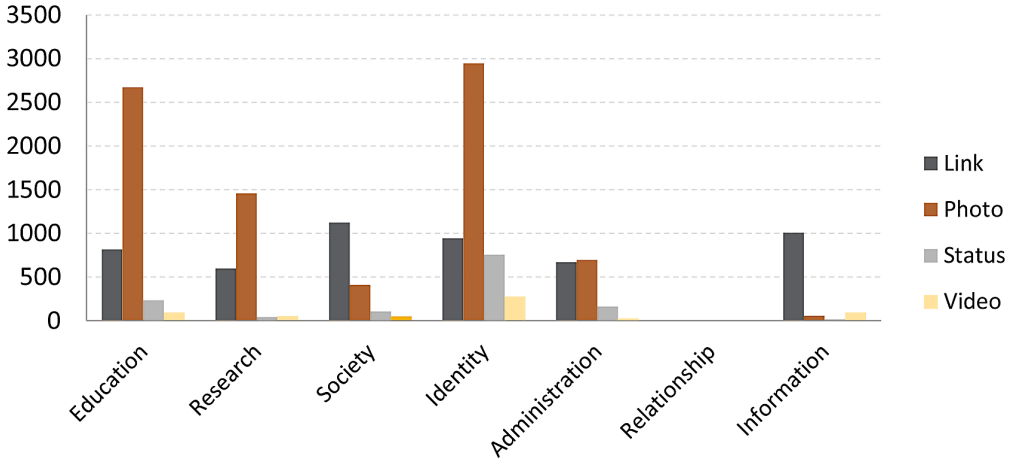
Performing an analysis based on editorial areas (as categorized by the ensemble of classifiers), we get the distribution depicted in Figure 3.

The editorial area 'Identity' is clearly the most prone area for posting. Every type of post has more posts in 'Identity' than in the other editorial areas. 'Education' and 'Research' are the second and third areas with more posts. We observe that in 'Society' and 'Information,' the dominant post type is the 'link,' whereas in other areas is the 'photo.' We note that the editorial area 'Relationship' is almost non-existent. This might be an indicator of organizations failing their attempts to take full advantage of social media potential concerning the development of social conversations and meaningful, emotional and bonding relationships. Cultural issues, lack of technological and / or communication competencies, unavailable human resources or teams dedicated to social media management could represent some of the reasons for this failure. Engagement and post type per editorial area are detailed in Table 3.

Table 2. Post type and corresponding engagement (heat map)

Post type		#Interactions' type			Engagement ratios				Interaction per post
Type	Total Posts	Likes	Shares	Comments	Like Ratio	Share Ratio	Comment Ratio	Tot. Eng.	
Link	5190	20573	6300	549	4.0	1.2	0.1	27427	5.3
Photo	8331	90579	12037	3201	10.9	1.4	0.4	105829	12.7
Status	1311	3218	693	292	2.5	0.5	0.2	4206	3.2
Video	576	7073	1783	193	12.3	3.1	0.3	9064	15.7

Figure 3. Posts type per editorial area



In Table 3 it is possible to observe that the most frequent post type is the pair ('Identity,' 'photo'), followed by ('Education,' 'photo'). Apart from the editorial area 'Relationship,' the least frequent pair is ('Administration,' 'video'), ('Research,' 'status'), ('Research,' 'video') and ('Society,' 'video'). It is also possible to observe that the editorial area most likely to create engagement is "Identity", for all types of response.

Table 3. Engagement per editorial area and per post type (heat map)

Editorial Area	Engagement			Post type			
	Likes	Shares	Comments	Link	Photo	Status	Video
Education	28710	5989	867	816	2673	233	90
Research	13088	2018	392	596	1457	41	49
Society	6899	2808	156	1125	409	105	49
Identity	53863	6517	2174	943	2948	756	275
Administration	13241	2047	489	670	697	161	22
Relationship	69	0	0	0	5	0	0
Information	4748	1316	118	1007	58	14	89

Performing a cross-check analysis between the editorial areas and the post type versus the resulting engagement, we get a very similar result for each editorial area, as depicted in Figure 4 (the top part of the graph was cut off due to a very high number of likes when compared to shares and comments, thus maximum values are presented).

Essentially, there are two relations that stand across all editorial areas of the graphic:

- The 'like' is the prominent type of interaction. It is particularly noticeable when combined with posts type 'photo' and 'link.'
- Typically, considering "totals", it is possible to establish the following order of expected engagement return (from the most engaging type to the least one): 'photo,' 'link,' 'video' and, 'status.'
- If, instead, ratios of engagement per post are considered, then the order (from the most engaging type to the least one) would be: 'video,' 'photo,' 'link' and, 'status.'

It is important to distinguish the three types of engagement, since receiving / making a 'like' in a post is not the same as receiving / making a 'share' or a 'comment.' In order to address this question, a survey was conducted on a total population of 190 individuals, which included students, teachers and communication professionals. Individuals were given a scale of 1 to 10 to classify 'likes,' 'comments' and 'shares' according to the following question: "When an organization evaluates the quality of interactions with its Facebook fans, which do you believe that should be the value (weight) assigned to each of the following interactions?". The mode of classifications obtained resulted in: one 'like' equals 4 points, one 'comment' equals 8 points and a 'share' equals 10 points.

The above weights – 'like' (0.4), 'comment' (0.8) and, 'share' (1) – were incorporated in a *score function* to calculate the 'weighted engagement' and to verify which are the editorial areas more prone to receive interaction (Figure 5).

As illustrated in Figure 5, even giving a weighted punctuation based on the perception of the type of engagement, the areas with more engagement are still 'Identity' followed by 'Education.' Curiously, 'Research' and 'Administration' have almost the same engagement value and 'Society,' followed by 'Information,' are the editorial areas with less engagement ('Relationship' remains almost non-existent).

Classification of Content Strategies

The previous classification allowed to obtain a firm insight on the editorial areas in which HPPEI have been investing, in relation to each agent's total communication effort (i.e., total number of Facebook messages, for all agents in a one year period). This classification also allowed to perform a sector analysis, based on benchmarking methods (Camp, 2013). This is one of the most efficient ways to determine the sector's tendency and to identify potential best practices and / or low performance, providing helpful understandings for adjusting / improving an organization's social media strategy.

In this analysis focus relies on revealing the sector tendency in terms of communication areas of focus and the overall classification of the social media strategies in three main types – centralized, decentralized and hybrid (Oliveira & Figueira, 2015). Examples of decentralized strategies are shown in Figure 6, as being the type of strategy mainly identified in the sector.

The sector tendency analysis reveals that HPPEI invest heavily on communicating their identity ('Identity'), in order to create and manage a positive internal and external image (Figure 6). In fact, this is particularly relevant in the HES. The projection of the organizational identity

Figure 4. Number of posts per type and area, versus engagement

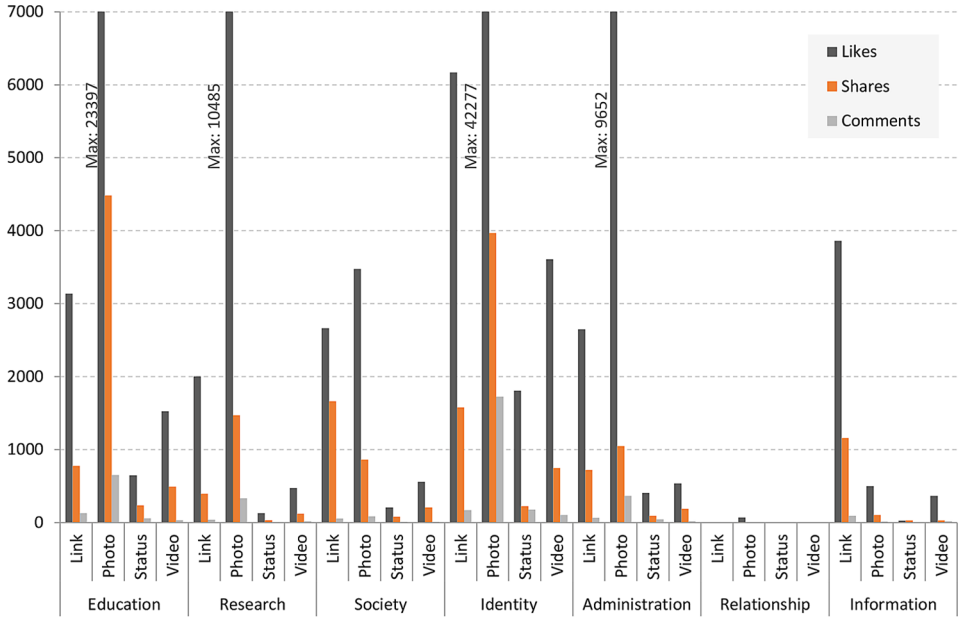
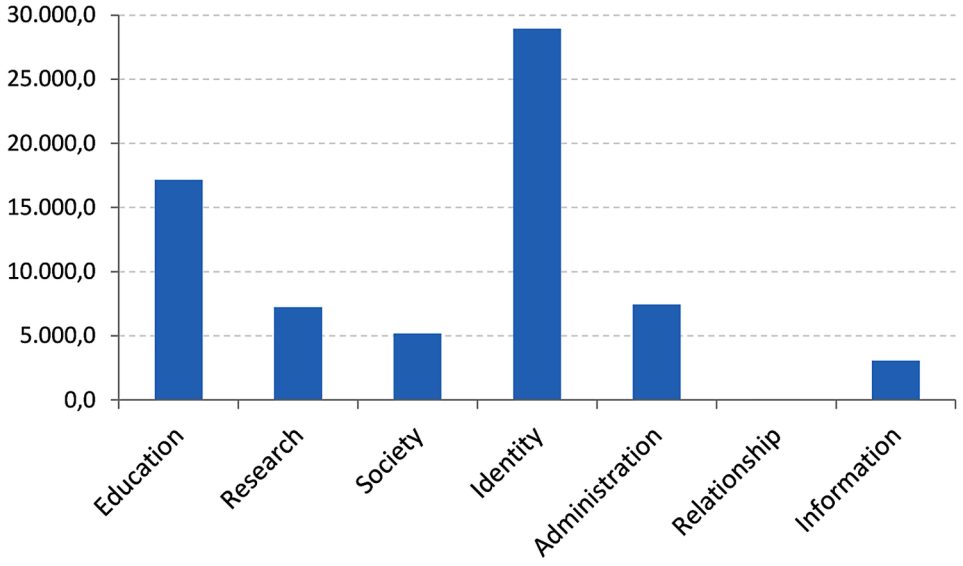
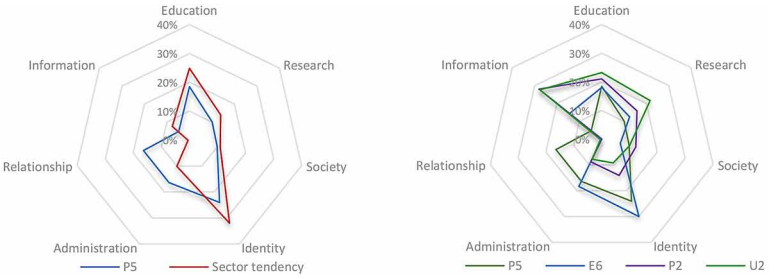


Figure 5. Total engagement score per editorial area



is key to maintain and protect a strong reputation, particularly for organizations dealing with different expectations and demands from stakeholders, since it determines its distinctiveness. HPPEI's (and HEI in general) reputation is one of the main factors impacting students' and parents' choice of the educational service provider, being able to provide HEI with a "first-choice"

Figure 6. (a) Sector tendency; (b) Decentralized strategies (Oliveira & Figueira, 2015)



status (Kotler & Fox, 1985). The organizational reputation also serves as an indicator for the underlying quality of the services provided and of its performance, thus being extremely relevant for the development of (commercial and institutional) partnerships' and to the extension of its societal intervention.

The second most relevant area of focus of HPPEI's social media strategies is 'Education,' which consists on advertising and providing information on the core educational service: higher education courses and complementary training (internal or external). Whilst 'Identity' serves the institutional dimension of organizational communication, 'Education' serves the transactional needs supported by the marketing communications dimension. This area is intimately connected to the organization's instrumental survival (profit), since no HEI can survive without students enrolling their courses.

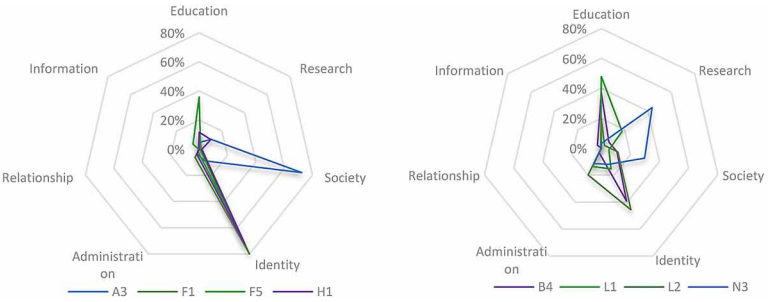
Communication related to 'Research,' 'Society,' 'Information' and 'Administration' is not as relevant as 'Identity' and 'Education' in the overall social media strategies being implemented. This may be an indicator of the, previously identified, growing competitive environment these organizations have been facing, leading them to focus mainly on reputation management and ensuring economic survival.

However, it is the dialogical communication focused on the building of relationships that appears completely neglected in the sector ('Relationship'), as stated in the previous section. Among the 43 agents considered in the study, only one of the agents (P5) has been investing in the development of relationships per se with stakeholders, as shown in Figure 6. (Kilgour, Sasser, and Larke (2015)) identify this need as social transformation processes occurring in organizations that are trying to transform their brand messages into social sources of relationships, rather than sources of commercial information. For this reason, it would be relevant to monitor this agent, examining engagement strategies and withdrawing good practices.

Analyzing the highest and lowest standard deviations among each agents' editorial areas allowed to identify the 3 main types of strategies being pursued in social media, as illustrated in Figure 6 and Figure 7:

- Decentralized strategies (Figure 6-b) focus on several areas with approximate amounts of investment in each one, presenting lower standard deviations;
- Centralized strategies present higher standard deviations, revealing investments in one or two focus areas, neglecting all remaining areas (Figure 7-a);
- Hybrid strategies invest disproportionally in more than one focus area, with very high investments in two or three areas and low investment in the remaining ones, as illustrated in Figure 7-b.

Figure 7. (a) Centralized strategies; (b) Hybrid strategies (Oliveira & Figueira, 2015)



This classification is not aimed at determining which type of social media strategy best suits the HES. Research is not aimed at such a narrow understanding, at least at this point. In fact, it wouldn't be totally accurate to state that decentralized strategies would result in more efficient social media strategies, though covering a higher number of editorial areas tends to reveal an increased potential by diversifying messages and addressing wider audience expectations. Ultimately, one must not neglect that the social media strategy needs to be aligned with the overall organizational goals, which should also serve as an efficiency measurement indicator. Therefore, the hybrid strategies disproportional communication investments may, in fact, reveal diversified balances of fundamental sets of editorial areas being pursued in social media, providing valuable insights.

According to this perspective, centralized social media strategies could denote that, either these agents have assigned a very narrow and specific objective to their content strategy, or they are still unable to perceive other possible of content areas, stakeholders' expectations and/or social media channels potential.

In any case, it is possible to withdraw examples and good practice among communication areas, even from agents pursuing centralized strategies. For instance, agent A3 (in Figure 7-a) focuses only the editorial area 'Society.' Monitoring this agent could be relevant for agents presenting communication deficits in this area and aiming at improving it. This is to say that a sector benchmarking analysis and corresponding withdrawal of comparison outputs and good practice needs to be built upon a previous assessment of each HEI current social media approach, in order to identify possible weakness and strengths to be corrected / fostered.

CONCLUSION

Much of the contributions in the domain of social media content analysis have been centered in depicting content as an end in itself or conducted as far as to relate content type with the amount of engagement, without an instrumental view of the strategic value of content. This is evident in literature and in reports published by social media monitoring applications. Research and results presented in this paper consist on one of the first attempts to present an editorial model for a specific sector, namely the Higher Education Sector, which, as explained, is not a transactional straightforward context; and, to connect it with an automatic classification system.

Presented research is also not without limitations. Although the editorial model is potentially generalizable to all HEI, such generalization may have geographic boundaries, namely because of the social, economic and politic context. Therefore, research on the applicability of the editorial model in other countries is required. The classification mechanism, based on an ensemble

of classifiers, is also prone to failures. Nevertheless, there is a positive high correlation between accuracy and the length of the message.

However, albeit these constraints, the structure of the proposed methodology may be applicable to other sectors, with suiting editorial models, which need to be developed and tested.

ACKNOWLEDGMENT

This work is financed by the ERDF – European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project «Reminds/ UTAP-ICDT/EEI-CTP/0022/2014»

REFERENCES

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*(3), 175–185.

Andersen, P. (2007). *What is Web 2.0?: ideas, technologies and implications for education* (Vol. 1). UK: JISC Bristol.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on* Pattern Analysis and Machine Intelligence, *35*(8), 1798–1828.

Berthon, P. R., Pitt, L. F., Plangger, K., Shapiro, D., Berthon, P. R., Pitt, L. F., & Shapiro, D. et al. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business Horizons*, *55*(3), 261–271. doi:10.1016/j.bushor.2012.01.007

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324

Camp, R. C. (2013). Benchmarking: the search for industry best practices that lead to superior performance: Milwaukee, Wis.: Quality Press; Quality Resources.

Chauhan, K., & Pillai, A. (2013). Role of content strategy in social media brand communities: A case of higher education institutes in India. *Journal of Product and Brand Management*, *22*(1), 40–51. doi:10.1108/10610421311298687

Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. *Paper presented at theProceedings of the 25th international conference on Machine learning*. doi:10.1145/1390156.1390177

Cornelissen, J. (2014). Corporate communication: A guide to theory and practice. *Sage (Atlanta, Ga.)*.

Crammer, K., & Singer, Y. (2002). On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, *2*, 265–292.

Fournier, S., Avery, J., Fournier, S., & Avery, J. (2011). The uninvited brand. *Business Horizons*, *54*(3), 193–207. doi:10.1016/j.bushor.2011.01.001

Freund, Y., & Schapire, R. E. (1999). Large margin classification using the perceptron algorithm. *Machine Learning*, *37*(3), 277–296. doi:10.1023/A:1007662407062

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *Annals of Statistics*, *28*(2), 337–407. doi:10.1214/aos/1016218223

Gensler, S., Völckner, F., Liu-Thompkins, Y., & Wiertz, C. (2013). Managing Brands in the Social Media Environment. *Journal of Interactive Marketing*, *27*(4), 242–256. doi:10.1016/j.intmar.2013.09.004

Habibi, M. R., Laroche, M., & Richard, M.-O. (2014). Brand communities based in social media: How unique are they? Evidence from two exemplary brand communities. *International Journal of Information Management*, *34*(2), 123–132. doi:10.1016/j.ijinfomgt.2013.11.010

Hanna, R., Rohm, A., Crittenden, V. L., Hanna, R., Rohm, A., & Crittenden, V. L. (2011). We're all connected: The power of the social media ecosystem. *Business Horizons*, *54*(3), 265–273. doi:10.1016/j.bushor.2011.01.007

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, *33*(3), 464–472. doi:10.1016/j.ijinfomgt.2013.01.001

Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, *28*(2), 149–165. doi:10.1016/j.intmar.2013.12.002

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. doi:10.1016/j.bushor.2009.09.003

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241–251. doi:10.1016/j.bushor.2011.01.005

Kilgour, M., Sasser, S. L., & Larke, R. (2015). The social media transformation process: Curating content into strategy. *Corporate Communications: An International Journal*, *20*(3), 326–343. doi:10.1108/CCIJ-07-2014-0046

Kirtiş, A. K., & Karahan, F. (2011). To be or not to be in Social Media arena as the most cost-efficient marketing strategy after the global recession. *Procedia: Social and Behavioral Sciences*, *24*, 260–268. doi:10.1016/j.sbspro.2011.09.083

Kotler, P., & Fox, K. F. (1985). *Strategic marketing for educational institutions*. Prentice-Hall.

Kunsch, M. M. K. (2003). *Planejamento de relações públicas na comunicação integrada*: Summus editorial.

Lai, L. S., & To, W. (2015). Content Analysis of Social Media: A grounded Theory Approach. *Journal of Electronic Commerce Research*, *16*(2), 138.

Lardi, K. (2013). The Social Media Strategy Framework. In K. Lardi & R. Fuchs (Eds.), Social Media Strategy: Step-by-step guide to building your social business (1st ed., pp. 17-23). Zurique: Zurich: vdf Hochschulvlg.

Mintzberg, H. (1987). The strategy concept I: Five Ps for strategy. *California Management Review*, *30*(1), 11–24. doi:10.2307/41165263

Mintzberg, H. (1994). The fall and rise of strategic planning. *Harvard Business Review*, *72*(1), 107–114.

Moss, D. (2011). *Strategy-Making and Planning in the Communications Context. In Public Relations: A Managerial Perspective* (pp. 111–142). Sage Publications.

Moss, D., & DeSanto, B. (2011). Public relations: A managerial perspective. *Atlanta, Ga.: Sage*.

Oliveira, L., & Figueira, Á. (2015). Benchmarking Analysis of Social Media Strategies in the Higher Education Sector. *Procedia Computer Science*, *64*, 779–786. doi:10.1016/j.procs.2015.08.628

Qi, B., & Mackie, L. (2014). Utilising Social Media Technology to Raise Brand Awareness in Higher Education. *Paper presented at the WEBIST*.

Rebholz-Schuhmann, D., Kirsch, H., & Couto, F. (2005). Facts from text—is text mining ready to deliver? *PLoS Biology*, *3*(2), e65. doi:10.1371/journal.pbio.0030065 PMID:15719064

Riff, D., Lacy, S., & Fico, F. (2014). *Analyzing media messages: Using quantitative content analysis in research*. Routledge.

Robinson, R. A. (2011). Business-to-Business Public Relations Agency Practice. In *Public Relations: A Managerial Perspective* (p. 222).

Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O'Reilly Media, Inc.

Saxton, G. D., & Waters, R. D. (2014). What do stakeholders like on Facebook? Examining public reactions to nonprofit organizations' informational, promotional, and community-building messages. *Journal of Public Relations Research*, *26*(3), 280–299. doi:10.1080/1062726X.2014.908721

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web*. A&C Black.

Zerfass, A., Fink, S., & Linke, A. (2011). Social media governance: Regulatory frameworks as drivers of success in online communications. *Paper presented at the14th International Public Relations Research Conference*, Miami, USA.

Luciana Oliveira holds an MSc in Multimedia and is a PhD candidate in Communication. Currently she is an Assistant Lecturer at ISCAP / IPP in scientific area of Consultancy and Organizational Communication. She has been a member of several European/international projects. Her research interests include: Social Media, Social Networks, Communication and Technology Mediated Education. Álvaro Figueira graduated in "Mathematics Applied to Computer Science" in Faculty of Sciences (UP) in 1995. He got his MSc in "Foundations of Advanced Information Technology" from Imperial College, London, in 1997, and his PhD in Computer Science from University of Porto, in 2004. Prof. Figueira is currently an Assistant Professor with tenure at Faculty of Sciences in University of Porto. His research interests are in the areas of web mining, community detection, e-learning and web-based learning and standards in education. Prof Figueira is also a researcher of the Associated Laboratory, INESC TEC, where he has been leading international projects involving University of Texas at Austin, University of Porto, University of Coimbra and University of Aveiro, regarding the automatic detection of relevance in social networks.