Excavating Business Intelligence from Social Media

E

Giannis Milolidakis
Technological Education Institute of Crete, Greece & Euromed Management, France

Demosthenes Akoumianakis
Technological Education Institute of Crete, Greece

Chris Kimble
Euromed Management, France & Université Montpellier II, France

Nikolas Karadimitriou
Technological Education Institute of Crete, Greece

INTRODUCTION

Social media technologies have changed the way individuals, teams and organizations communicate. Today, social networking sites, blogs, and forums, are used as channels of communication between family, friends, neighbours, and co-workers. Businesses are also facing new challenges. The volume of data in social media platforms can provide valuable insights on people's opinions that are difficult and costly to find elsewhere, it also offers new opportunities for business intelligence (BI).

In recent years, efforts by businesses to understand the environment in which they operate using new media channels are typically constrained to customer analytics. Indeed, many of these metrics are incorporated as services by social media to their corporate users. Although these measurements are useful to estimate the effectiveness of marketing campaigns or brand awareness, they appear to be insufficient to provide a deeper understanding of customers' behaviours. This is because much of the information retained by social networks is embedded in different artefacts such as unstructured text, videos, photos, and patterns of user actions. These are not uniform and do not afford standard manipulative actions. The situation is further complicated by the continuous evolution of social media, which constantly update their services and integrate new social media functions. This makes harder a one-off solution for collecting and analysing online data; something that could be achieved much more easily in a business supported forum or a bulletin board.

In this chapter, we provide a step towards a model for obtaining business intelligence through social media grounded in the concept of virtual settlements. The notion of a virtual settlement was introduced by Jones (1997) in an effort to understand virtual communities through what he termed as cyber-archaeological inquiries. Re-focusing the problem from virtual communities to more general cyber-phenomena allows us to approach social technologies as a kind of archaeological settlement whose excavation may facilitate improved BI. In this effort, we first motivate a method and then demonstrate the method's application to BI through a case study.

BACKGROUND

Transforming business relevant data into BI has been a long-standing aim of firms in different sectors of the industry (Chen, Chiang, & Storey, 2012; Wixom & Watson, 2010). In recent years, the portfolio of methods has expanded to account

Table 1. Recent efforts in utilizing social media data

Scholarship	Methods/Methodology	Social Media	Application
(Hogg, 2010)	- Social network analysis - Network topology	- Social network service (Essembly.com)	Identifying consumer preferences to help sellers design customized bundles of products.
(Xu, Guo, Li, Lau, & Liao, 2012)	- Social network analysis - Influence network - Web mining	- Users opinions Website supporting social networking functions (Epinions.com) - Micro-blogging platform (Twitter.com)	Analysing user opinions to discover influential user for marketing and enterprise reputation management.
(Castellanos et al., 2011)	- Natural Language Processing - Sentiment Analysis	- Micro-blogging platform (Twitter.com)	Conducting real-time and historical sentiment analysis to get customer insights on new marketing or sales campaign.
(Prestus & Bygstad, 2010)	- Text mining	- Social network service (Facebook.com)	Utilizing social media to manage customer relationships and customer communication.
(Zhang, Fuehres, & Gloor, 2011)	- Sentiment Analysis	- Micro-blogging platform (Twitter.com)	Predicting stock market indicators by measuring collective hope and fear.
(Bollen, Mao, & Zeng, 2011)	- Sentiment Analysis	- Micro-blogging platform (Twitter.com)	Predicting stock market indicators through mood tracker tools.
(Sobkowicz, Kaschesky, & Bouchard, 2012)	- Natural language processing - Semantic Web approaches	- Online discussion platform - Blogging platforms (Technorati.com)	Exploring ways that online content can be exploited used to inform decision makers and the potential impacts of policy initiatives.
(Abrahams, Jiao, Wang, & Fan, 2012)	- Text mining - Content analysis	- Online discussion platform (honda-tech.com)	Facilitate decision support for the vehicle defect discovery and classification process.
(Kavanaugh et al., 2012)	- Semantic analysis Visualization	- Social network service (Facebook.com) - Micro-blogging platform (Twitter.com) - Video-sharing platform (Youtube.com)	Leverage social media data by government officials to improve communication with citizens and managing crisis situations.
(Callarisa, García, Cardiff, & Roshchina, 2012)	- Quantitative assessment	- Travel related social network service (Tripadvisor.com)	Measuring brand equity using online customer reviews.

for social media such as social Web sites and social networking services. This is mainly due to the wealth of data hosted and made available with the end users' consent. Such data turn out to be useful in revealing not only cultural information about past and/or on-going incidents but also market trends, consumer behaviour and other business related aspects. Such information can be processed and analysed from various perspectives such as social network analysis (Scott, 1988), virtual ethnographic assessments (Harrison, 2009), data mining and information discovery (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). The capability of enterprises to appropriate such data stems from the social media platforms' openness, interoperability and use of third-party applications. An attempt to synthesize representative efforts in this direction is presented in Table 1.

The efforts presented above are indicative of the importance of social media for modern enterprises. However, they do not follow a common protocol and do not allow comparisons across cases. Moreover, they are frequently limited to what is 'traceable' using a certain tool at the expense of what is needed and most useful for an enterprise. To address these challenges, the present work links with efforts that conceive BI in Web 2.0 as a type of archaeological inquiry conducted in a virtual settlement; aiming to bring to the surface traces of digital remains of potential value for enterprises.

Archaeology and Cyber-Archaeology

The analogy between virtual spaces in computer-mediated communication and physical space in more traditional archaeological settlements was first introduced by Jones (1997) in an effort to develop a theory of what he termed cyber-archaeology. In his work, Jones referred to online infrastructures, the places where online communities can be found, as virtual settlements. Jones defined certain criteria that needed to be fulfilled for an online infrastructure to qualify as virtual a settlement:

1. A minimum level of interactivity.
2. A variety of communicators.
3. A virtual common-public-space where a significant portion of community interactions occur.
4. A minimum level of sustained stable membership.

If we accept these criteria, then most of today's popular social media platforms, such as Facebook or Twitter, would qualify as virtual settlements. Thus for Jones, virtual settlements, in the form of a suitable online infrastructure, are pre-requisites for the existence of virtual communities.

Subsequently, Jones' ideas have been used to study virtual communities in a variety settings, for example, Second Life (Harrison, 2009), blogs (Blanchard & Markus, 2004; Chin & Chignell, 2006), the blogosphere (Efimova & Hendrick, 2005) or blog spaces (Zhou & Davis, 2007), micro-blogging services such as the Twitter (Gruzd, Wellman, & Takhteyev, 2011), but also social networking services like Facebook (Akoumianakis, Milolidakis, Vlachakis, Karadimitriou, & Ktistakis, 2011; Zhao, Grasmuck, & Martin, 2008) and YouTube (Akoumianakis, Kafousis, Karadimitriou, & Tsiknakis, 2012).

Archaeological Excavation and Virtual Excavation for BI

Traditional archaeology studies human activities through an examination of the data that a previous generation has left behind, often in the form of the remains of human habitation. The perspective of the traditional archaeologist is that culture is enacted, that is, it is 'something that people do'. Consequently, the features of a living culture fade into memory and disappear when the people who enact it no longer exist, leaving only traces of its qualities attributes in the artefacts that they created (Lyman & Kahle, 1998). As Hodder (2003) explains, the material objects that humans create have a significance and meaning in accordance with their functions within the culture in which they were created.

According to Jones, virtual communities are also enacted structures where the specific features of these structures are influenced by certain attributes of the virtual settlement in which they exist. At first sight, the analogy between archaeology and cyber-archaeology seem compelling, but there are a number of significant differences in the two approaches, particularly when these ideas are applied to the excavation of virtual settlements for the purposes of uncovering information that is of use for business intelligence.

Firstly, the digital traces that are retained within a social media platform do not necessarily reflect a distant or past culture. If this were the case, then their practical value for business intelligence would be questionable. Thus, when virtual excavations are conducted for business intelligence, the participants have not vanished, as is the case of most traditional archaeological inquiries; such excavations set out to locate, identify and follow the activities of a group of active participants who constitute a source of useful information. Secondly, such virtual excavations

are more suited to understand the emergence of an on-going knowledge process rather than improving our understanding of what has happened in the past (Markus, Majchrzak, & Gasser, 2002). In a traditional archaeological inquiry excavation pre-dates classification of artefacts, as we shall see below, in a virtual excavation the order of these two activities is reversed as the online infrastructure that forms the basis for a virtual settlement is available and can be analysed prior to any excavation.

EXCAVATION-BASED BI

The rationale to approach BI through an excavation-based method is grounded on the belief that social media retain data that can bring to the surface 'hidden' knowledge of potential value to the enterprise and this knowledge, once revealed, can have implications on the firms' business strategy, marketing plans, and customer relations management. Nevertheless, the means through which social data are to be managed (i.e., extracted, filtered and transformed) so as to bring about the desirable benefits, remains a challenge. This is especially the case when relevant data reside on multiple and different genres of social media. Consequently, it is important, and in some sectors compelling, for enterprises to devise methods that are reliable and scalable to different media types. Key issues to be addressed in devising such methods include (a) what is relevant digital trace data and what artefacts leave the required traces; (b) where do they reside and what boundaries prohibit their access; (c) how are they retrieved and what sense is to be made.

The rest of this chapter elaborates on a model for appropriating virtual excavations of social media for BI purposes. Then, we present a case study in which Facebook pages and YouTube channels of three Greek telecommunication companies are treated as virtual settlements for extracting BI related information. Figure 1 presents the three distinct steps of our model and relates them to the theoretical streams of work that underpin each step.

Identifying Cultural Artefacts

A key feature of our model is the linking of the actions of a group of actors within a virtual settlement, which leaves a digital trace of their activities, to the context within which that activity took place. In this way, we are able to treat a particular digital trace as a cultural artefact related to a particular settlement.

In general, an artefact, in the form of a digital trace, may be considered a cultural artefact if its granularity allows extraction of information of cultural value. The first step in our method identifies which cultural artefacts should be deemed relevant to the BI objectives of an enterprise. As we are concerned with social media, the artefacts of primary interest fall into three main categories:

1. Artefacts that are philological or symbolic (e.g., posts and comments, 'like' buttons).
2. Artefacts that are linked with meeting places (e.g., Facebook walls, YouTube channels).
3. Artefacts that facilitate connections (e.g., user profiles, friend lists, subscriptions).

A feature of cultural artefacts that is important for our purposes relates to their affordances. Here the term affordance is used to denote the range of possibilities for action offered by an artefact (Gibson, 1979). It is necessary to consider affordances in order to identify possible clusters of related actions within a virtual settlement. It is also necessary to remember that different virtual settlements support different cultural artefacts, which leave different forms of digital traces. Thus, while some artefacts may appear to be common across different settlements, each settlement needs to be analysed separately to identify which cultural artefacts reside within it and which artefacts are of use for BI.

The Virtual Excavation Process

The second step is concerned with devising a protocol for managing digital trace data and transforming them from unstructured log entries

Figure 1. Excavation based model
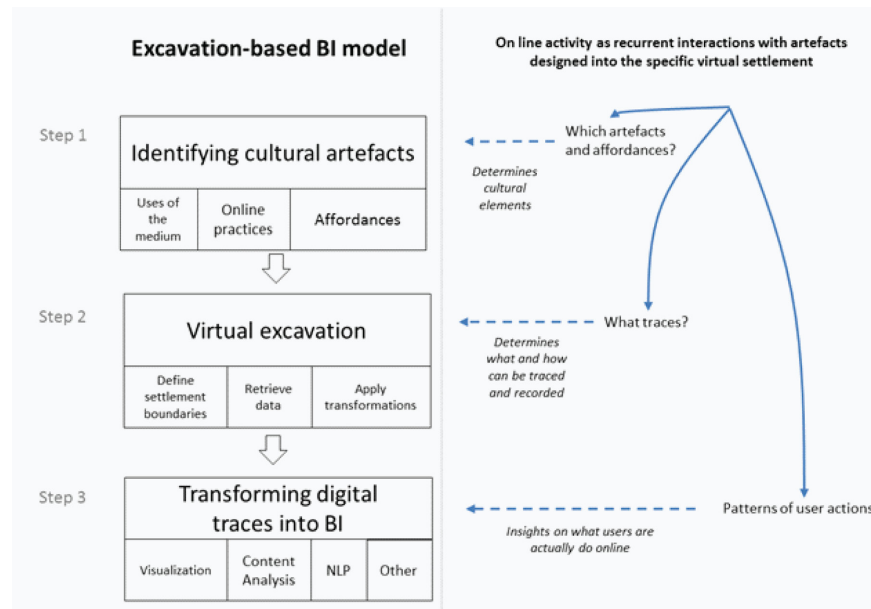
to structured data easily manipulated by the researcher. Such a protocol can be grounded on three basic steps:

1. Designate the settlements' thematic boundaries, thus designating which traces are relevant, and the tactics of boundary crossing (in case of multiple virtual settlements).
2. Make the necessary provisions for retrieving data from the virtual settlement by studying (possible) limitations of the public API (if exists) or alternative strategies to collect data.
3. Devise suitable transformation for the digital traces that remain to allow informative insights, pattern discovery, exploratory reviews, querying, etc.

Transforming Digital Traces into Business Intelligence

The final step is concerned with the repertoire of techniques to be used for making sense of digital traces for purposeful BI. Some widely used

approaches are social network analysis; natural language processing, sentiment analysis; text and Web mining; customer analytics. However, at the core of our method is the analysis of virtual traces through knowledge visualization (Eppler, 2011). Knowledge visualization is a technology flexible enough to allow a combination of methods; it is based on representations of physical referents and their relationships presented in a way that makes large data sets intelligible.

A CASE STUDY OF THREE GREEK TELECOMMUNICATION COMPANIES

Our case study focuses on three prepaid (or pay-as-you-go) mobile telecommunications brands that operate in Greece. An initial effort to investigate the use of Facebook for BI was reported in (Milolidakis, Akoumianakis, & Kimble, In press). Here, we expand our earlier empirical base by applying the method on another virtual settlement, YouTube.

Analysis of Virtual Settlements in Terms of Cultural Artefacts and Affordances

Amongst the functionalities of Facebook, fan pages constitute the top-level digital artefact through which users engage in interactions with the company but also other users. Such pages, frequently act as identity mechanisms inviting users to externalise opinion or state of mind. This information can be extracted from photos, videos, or textual announcements that the firm uploads on its fan page wall to reach their fans through each fan's news feed. These activities are traceable through dedicated APIs. Consequently, through such digital traces on Facebook, we can extract information on the brand's identity, intentions and cultural values, as well as the fans' opinion, concerns and pre-occupations.

YouTube is a video sharing platform where users post comments on videos or video responses, express positive (as in the case of Facebook) as well as negative feedback through dedicated buttons. YouTube users can maintain their own channel where uploaded videos and related activity can be displayed; they can also subscribe in channels and receive updates regarding new uploads. As all these are traceable, we can gain further insights on brand identity and the fans' opinions by analysing the brands' YouTube channels.

Excavation of Designated Virtual Settlements

Thematic Boundaries and Data Retrieval

Boundaries in the two virtual settlements under investigation are confined to the three Facebook fan pages and their YouTube channels. Data was captured using the public API of Facebook[1] and YouTube[2] platforms. The data extraction process was performed on August 24, 2012, while the data sample contains traces dated from 5 February 2009 to 23 August 2012 (see Table 2). Although some traces of interest could not be captured either

Table 2. Excavation identity

	Facebook	**YouTube**
Date of data excavation	24 August 2012	24 August 2012
Users	42,917	3,551
Activities	145,819	5,550
Oldest activity	5 February 2009 at 20:18 of type 'status update'	24 November 2009 of type 'video'
Most recent activity	23 August 2012 at 19:33 of type 'comment'	23 August 2012 of type 'video'

Type of activities captured	- Posts on wall (10,503)
	    o Status updates (9,238)
	       o Photos (539)
	       o Videos (353)
	       o Links (361)
	       o Questions (12)
	- Likes (88,030)
	    o On wall posts (64,023)
	    o On comments (24,007)
	- Comments (42,954)
	- Tags (4,332)
	    o On comments (4,125)
	       o On photos (207)
		- Videos (596)
		- Video responses (0)
		- Comments (4,954)

due to limitations of the platforms' API or due to privacy issues, our data sample is fairly complete and is adequate for our purposes.

Making Sense of Digital Traces for Business Intelligence

To make sense of digital trace data, we used knowledge visualizations implemented using Prefuse (Heer, Card, & Landay, 2005) - a java based toolkit. Details on the design and technical properties of these visualizations are reported elsewhere (Milolidakis, Akoumianakis, & Kimble, In press).

Active fans contributions and cross-poster's activity timeline: Active fans of a brand are users who have at least one activity of any kind and at least one wall post or a comment. In order to identify users that are active in more than one Facebook page, we rely on visualizations that blend users' activity in multiple Facebook pages to reveal cross page traffic. Both graphs of Figure

2 depict users and their connections. A closer look at this graph reveals that for brands 'B' and 'C', more than half of contributors with ten or more activities are cross posting. Such cross-posting reveals either users who have moved or are likely to be moving from one brand to another.

By visualizing the activity of the cross posters in timelines we were able to spot fans with cross-posting in two different brands during a short period. As an illustration, Figure 3 depicts a user's activity in two days. When clicking on a mark, a pop up window summarizes additional information about that activity. As the dialogs reveal (see Table 3), the fan has currently a prepaid plan with brand 'C' and is considering switching to brand 'A'.

YouTube videos and Facebook: A video that has been uploaded on YouTube is often shared on other social media. Here, we combine digital traces from both YouTube and Facebook. We extract remains of 596 videos originated from the brands' official YouTube channels, as well as remains of 715 YouTube videos that were traced in 915 wall posts or comments on Facebook. Through the graphs of Figure 4 and Figure 5 we visualized video-related activities. Each node represents a video and its size is proportional to the number of views as provided by YouTube. Figure 4 presents videos that were uploaded by the operator of each brand, while Figure 5 presents videos that were uploaded by the fans.

The visualizations reveal useful information. Firstly, Figure 4 reveals that videos posted on Facebook from the operators of brand 'A' and 'C' are, with a few exceptions, very popular videos (i.e. more than 1 million views) that originate from outside their official YouTube channel (i.e. videos from outer zones). Figure 5 confirms that very few videos posted by fans on each Facebook brand page originate from the brand's official YouTube channel. Furthermore, fans of brand 'A' have posted a significantly larger number of videos in relation to other two brands.

FUTURE RESEARCH

Virtual excavations provide a baseline for methods that are structured and organised around artefacts embedded in technology and the traces left behind by those artefacts. The potential of this method could extend far beyond tracing the user's online activities. Social media data is expected to continue attracting business attention as enterprises recognize their importance to understand both their clientele and their competition. Due to their size and the richness of their content, social media present a significant problem for those who wish to identify meaningful patterns. Business analytics, although useful in supporting decision making for the enterprise and driving business planning, tend to provide only a partial solution to the problem as they lack flexibility to be used in making sense of 'untypical' cyber-phenomena.

CONCLUSION

Virtual excavations should not be thought simply as a method for extracting data; rather they should be considered as a tool for understanding online practices of users through the analysis of the digital traces of their activity in a virtual settlement. In this chapter we have argued that social media can be subjected to a form of enquiry based on an archaeological model and have presented a model based around the notion of virtual excavations and that this method is a valid approach for obtaining business intelligence. Through a case study of three prepaid mobile brands of the Greek telecommunication industry we focused on (a) how cultural artefacts can be identified from the digital trace data created by their use; (b) where such digital traces might reside and what might inhibit their access; and (c) how are they retrieved and what sense is to be made of them. Our empirical evidence confirms earlier efforts (Milolidakis, Akoumianakis, & Kimble, In press) on the relevance and value of the method for enterprises. It also points out the method's wider implications when combined with transformational technologies such as knowledge visualization.

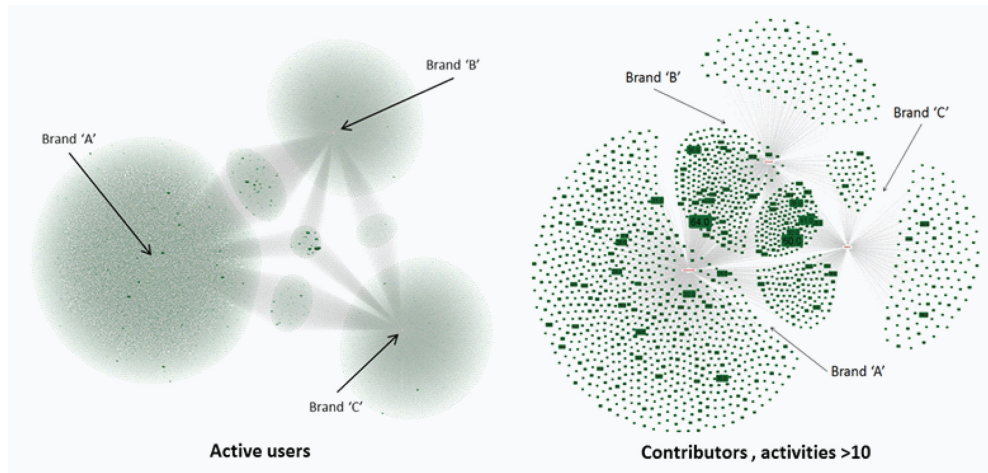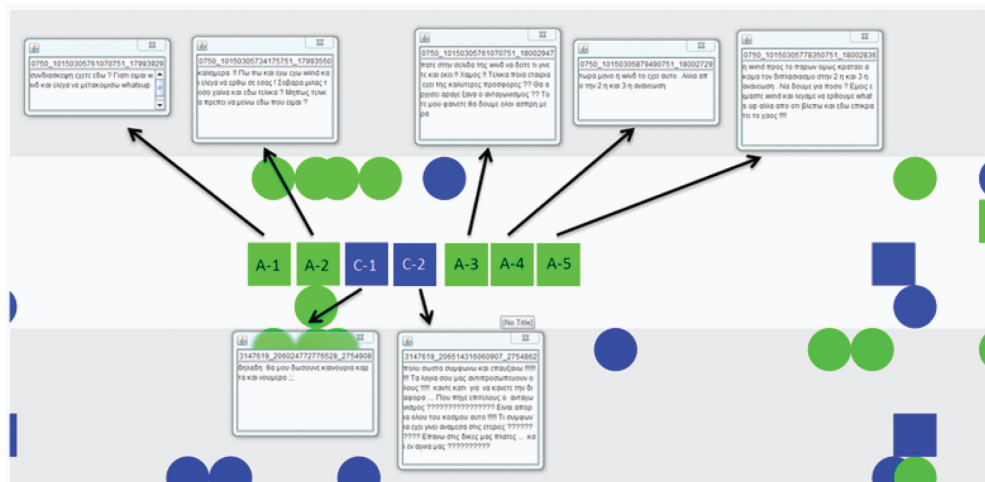*Figure 2. Active users on Facebook pages*

Figure 3. A cross-poster's timeline

Table 3. Comments of the fan of Figure 3 (translated from Greek)

Mark	Brand	Type of Activity	English Translation	Number of Likes
A-1	'A'	Comment	Everyone is gathered here ? Because I'm currently in [brand 'C'] and I'm thinking on moving on [brand 'A']	3
A-2	'A'	Comment	Good morning !! Wow I also have [brand 'C'] and I was thinking on coming to [brand 'A'] ! Really [,] that bad here after all? Should I stay where I am ?	0
C-1	'C'	Comment	So, they are going to give me new [SIM] card and number ???	0
C-2	'C'	Comment	[…] Do something to make a difference… At last, where is [market] competition ??????? […] What is the deal all companies have made??? […]	2
A-3	'A'	Comment	Go to [brand C] page to see what is also going on there […] Which company has the better offer after all?? When [market] competition will start?? Only then […]	1
A-4	'A'	Comment	Now only [brand 'C'] allows this. But only if you recharge for second and third time	0
A-5	'A'	Comment	[Brand 'C'] for now doubles on the second and third recharge. Let's see for how long? We are in [brand 'C'] and we thought to come to [brand 'A'] but as I can see here is also a mess	0

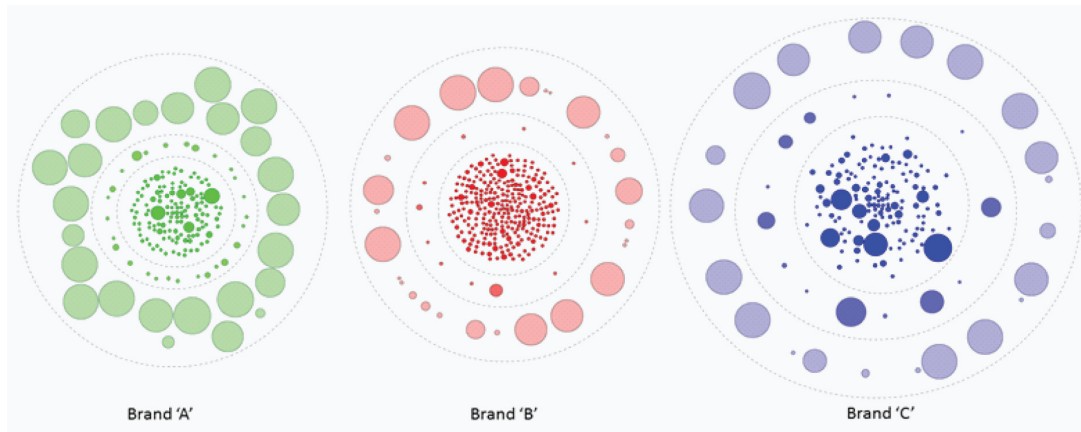*Figure 4. Videos on Facebook page and YouTube channel posted by operators*

Figure 5. Videos on Facebook page and YouTube channel posted by fans

REFERENCES

Abrahams, A. S., Jiao, J., Wang, G. A., & Fan, W. (2012). Vehicle defect discovery from social media. *Decision Support Systems*, *54*(1), 87–97. doi:10.1016/j.dss.2012.04.005

Akoumianakis, D., Kafousis, I., Karadimitriou, N., & Tsiknakis, M. (2012). Retaining and exploring online remains on YouTube. Presented at the 3-rd International Conference on Emerging Intelligent Data and Web Technologies, Bucharest, Romania.

Akoumianakis, D., Milolidakis, G., Vlachakis, G., Karadimitriou, N., & Ktistakis, G. (2011). Retaining and exploring digital traces. *International Journal of Virtual Communities and Social Networking*, *3*(4), 46–65. doi:10.4018/jvcsn.2011100104

Blanchard, A. L., & Markus, M. L. (2004). The experienced sense of a virtual community: Characteristics and processes. *ACM SIGMIS Database*, *35*(1), 65–79. doi:10.1145/968464.968470

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. doi:10.1016/j.jocs.2010.12.007

Callarisa, L., García, J. S., Cardiff, J., & Roshchina, A. (2012). Harnessing social media platforms to measure customer-based hotel brand equity. *Tourism Management Perspectives*, *4*(0), 73–79. doi:10.1016/j.tmp.2012.04.005

Castellanos, M., Dayal, U., Hsu, M., Ghosh, R., Dekhil, M., Lu, Y., et al. (2011). LCI: a social channel analysis platform for live customer intelligence. In *Proceedings of the 2011 international conference on Management of data* (pp. 1049-1058). New York, NY, USA: ACM. doi:10.1145/1989323.1989436.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *Management Information Systems Quarterly*, *36*(4), 1165–1188.

Chin, A., & Chignell, M. (2006). Finding evidence of community from blogging co-citations: a social network analytic approach. In *Proceedings of the IADIS International Conference on Web Based Communities* (pp. 191-200). San Sebastian, Spain.

Efimova, L., & Hendrick, S. (2005). In search for a virtual settlement: An exploration of weblog community boundaries. In Communities & Technologies, 5.

Eppler, M. J. (2011). What is an effective knowledge visualization? Insights from a review of seminal concepts. In *2011 15th International Conference on Information Visualisation (IV)* (pp. 349-354). Presented at the 2011 15th International Conference on Information Visualisation (IV). doi:10.1109/IV.2011.13.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gruzd, A., Wellman, B., & Takhteyev, Y. (2011). Imagining twitter as an imagined community. *The American Behavioral Scientist*, *55*(10), 1294–1318. doi:10.1177/0002764211409378

Harrison, R. (2009). Excavating Second Life. *Journal of Material Culture*, *14*(1), 75–106. doi:10.1177/1359183508100009

Heer, J., Card, S. K., & Landay, J. A. (2005). Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 421-430).

Hodder, I. (2003). The interpretation of documents and material culture. *Collecting and interpreting qualitative materials, 2*, 155-175.

Hogg, T. (2010). Inferring preference correlations from social networks. *Electronic Commerce Research and Applications*, *9*(1), 29–37. doi:10.1016/j.elerap.2009.04.006

Jones, Q. (1997). Virtual-communities, virtual settlements & cyber-archaeology: A theoretical outline. *Journal of Computer-Mediated Communication*, *3*(3), 35–49.

Kavanaugh, A. L., Fox, E. A., Sheetz, S. D., Yang, S., Li, L. T., & Shoemaker, D. J. et al. (2012). Social media use by government: From the routine to the critical. *Government Information Quarterly*, *29*(4), 480–491. doi:10.1016/j.giq.2012.06.002

Lyman, P., & Kahle, B. (1998). Archiving digital cultural artifacts. *D-Lib Magazine*, *4*(7).

Markus, M. L., Majchrzak, A., & Gasser, L. (2002). A design theory for systems that support emergent knowledge processes. *Management Information Systems Quarterly*, *26*(3), 179–212.

Milolidakis, G., Akoumianakis, D., & Kimble, C. (in press). Digital traces for business intelligence: A case study of mobile telecoms service brands in Greece. *Journal of Enterprise Information Management*.

Prestus, W., & Bygstad, B. (2010). Facebook as agile CRM? A business intelligence analysis of the airline ash crisis. In *Norsk Konferanse for Organisasjoners Bruk av Informasjonsteknologi*. NOKOBIT.

Scott, J. (1988). Social network analysis. *Sociology*, *22*(1), 109–127. doi:10.1177/0038038588022001007

Sobkowicz, P., Kaschesky, M., & Bouchard, G. (2012). Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly*, *29*(4), 470–479. doi:10.1016/j.giq.2012.06.005

Wixom, B., & Watson, H. (2010). The BI-based organization. *International Journal of Business Intelligence Research*, *1*(1), 13–28. doi:10.4018/jbir.2010071702

Xu, K., Guo, X., Li, J., Lau, R. Y. K., & Liao, S. S. Y. (2012). Discovering target groups in social networking sites: An effective method for maximizing joint influential power. *Electronic Commerce Research and Applications*, *11*(4), 318–334. doi:10.1016/j.elerap.2012.01.002

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter "I hope it is not as bad as I fear.". *Procedia-Social and Behavioral Sciences*, *26*, 55–62. doi:10.1016/j.sbspro.2011.10.562

Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, *24*(5), 1816–1836. doi:10.1016/j.chb.2008.02.012

Zhou, Y., & Davis, J. (2007). Discovering web communities in the blogspace. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS '07)*. IEEE Computer Society, Washington, DC, USA, (pp. 85-85). doi:10.1109/hicss.2007.177.

ADDITIONAL READING

Aral, S., Dellarocas, C., & Godes, D. (2013). Introduction to the special issue—Social media and business transformation: A framework for research. *Information Systems Research*, *24*(1), 3–13. doi:10.1287/isre.1120.0470

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*. doi:10.1016/j.bushor.2011.01.005

Stodder, D. (2012). *Customer Analytics in the Age of Social Media*. TDWI -The Data Warehousing Institute.

Zeng, D., Chen, H., Lusch, R., & Li, S.-H. (2010). Social Media Analytics and Intelligence. *IEEE Intelligent Systems*, *25*(6), 13–16. doi:10.1109/MIS.2010.151

KEY TERMS AND DEFINITIONS

Affordance: It is a quality of an object that denotes the range of possible actions offered by that object. Considering affordances is necessary to identify possibilities for action within a virtual settlement.

Cultural Artefact: Culture is viewed as 'enacted', that is, it is an emergent property of the activities of a certain group of actors. A cultural artefact is object that provides information about the activities of this group and the culture of its user.

Digital Trace: A digital trace constitutes a cultural artefact in a virtual settlement. It is some form of evidence of the activity of a user. Digital traces may be records stored in databases or log files which can be accessed by researchers though the use of API's or Web crawlers.

Social Media: Social media are specific type of virtual settlement which exists within Internet-based or cloud-based applications. Social media facilitate the co-creation and dissemination of the various forms of content that are generated by the users of that media.

Virtual Communities: A virtual community is a group of people who interact with each other within the boundaries of some virtual settlement where the characteristic of that community are, to some extent, shaped by the features that the surrounding on-line infrastructure makes available to them.

Virtual Excavation: A systematic acquisition of digital trace data from virtual settlements using the archaeological paradigm. Virtual excavations for business intelligence are aimed at locating and follow the activities of an active group of participants who constitute a source of useful business intelligence rather than improving our understanding of past events.

Virtual Settlement: A virtual space within an on-line infrastructure that supports some form of computer-mediated communication. It is analogous to the physical space within which a traditional archaeological settlement might be found. Like a traditional archaeological settlement, virtual settlements can be excavated using methods based on the archaeological paradigm.

ENDNOTES

[1] **Graph API:** http://developers.facebook.com/docs/reference/api/
[2] **YouTube API:** https://developers.google.com/youtube/