



**HAL**  
open science

## Digital traces for business intelligence: A case study of mobile telecoms service brands in Greece

Giannis Milolidakis, Demosthenes Akoumianakis, Chris Kimble

### ► To cite this version:

Giannis Milolidakis, Demosthenes Akoumianakis, Chris Kimble. Digital traces for business intelligence: A case study of mobile telecoms service brands in Greece. *Journal of Enterprise Information Management*, 2013, 27 (1), pp.66-98. 10.1108/JEIM-09-2012-0061 . halshs-00954440

**HAL Id: halshs-00954440**

**<https://shs.hal.science/halshs-00954440>**

Submitted on 2 Mar 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Digital traces for business intelligence: A case study of mobile telecoms service brands in Greece

Giannis Milolidakis, Demosthenes Akoumianakis and Chris Kimble

## Abstract

**Purpose** – Data from social media (SM) has grown exponentially and created new opportunities for businesses to supplement their business intelligence (BI). However, there are many different platforms all of which are in a constant state of evolution. The purpose of this paper is to describe a generic methodology for the gathering of data from SM and transforming it into valuable BI.

**Design/methodology/approach** – The approach taken is termed virtual excavation and builds on the similarities between the manipulation of technological artefacts virtual communities using various forms of SM and the excavation and analysis of physical artefacts found in archaeological settlements.

**Findings** – The paper reports on a case study using this technique that looks at the Facebook fan pages of three mobile telecommunications service providers in Greece. The paper identifies many of the standard BI indicators as well as demonstrating that additional information relating to cross-page use can be collected by looking at how users manipulate artefact such as the “like” button in Facebook.

**Research limitations/implications** – Although the methodology is widely applicable, the paper only reports on the analysis of one platform, Facebook, and is heavily reliant on visualization tools. Future work will examine different platforms and different tools for analysis.

**Practical implications** – The paper discusses some of the ways in which this approach could be used and suggests some areas in which it might be applied.

**Originality/value** – The approach of using virtual excavations to extract BI from virtual communities in online SM offers a systematic approach for dealing with a variety of information from a variety of different media that is not found in techniques based on information systems or management science.

Keywords Virtual communities, Social media, Business intelligence, Facebook pages, Virtual excavations

## 1 Introduction

Business Intelligence (BI) has existed as a term since 1958 when Hans Peter Luhn (Luhn, 1958) used it to describe an automatic system for information dissemination by utilizing data-processing machines for abstracting, encoding and archiving any kind of documents of an organization. In this early conception of BI, the term business was considered as the “collection of activities carried on for whatever purpose” while the notion of intelligence was defined as “the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal”. The overall objective of the system proposed by Luhn was to support organizational actions, speedily and efficiently, by providing appropriate information. Since such early work, BI research has grown and expanded in both range and scope including, from a technical viewpoint, advances in computer-aided decision and data mining systems, and from a managerial perspective, ways of getting the right information to the right people at the right time. Consequently, BI is broadly defined as the process of taking items of data, analysing them, and condensing their essence into the basis of business actions, enabling management to gain new insights and thereby contributing to their business decisions (Davenport, 2006; Gilad and Gilad, 1985; Li, 2005; Pirttimäki, 2007a; Tyson, 1986; Watson and Wixom, 2007). Today, BI embraces a variety of activities, including social media analysis, to allow businesses to gain deeper understanding of their organization and build competitive advantage.

Social media (SM) refers to a variety of genres of technology with a wide range of capabilities. For example, blogging platforms allow users to maintain a log of their activities, opinion or thoughts, triggering other users' feedback by following or commenting on the content. Micro-blogging services, such as Twitter and Foursquare, make provisions for users to publish short messages regarding news stories, the user's current location, or nearby events and share them with their connections/followers. Social networking platforms such as Facebook allow the construction and maintenance of social relationships among users with common interests or activities. As SM become increasingly available across a wide variety of different devices and platforms, new business opportunities arise and large amounts of social data are generated with the potential to benefit business enterprises.

With respect to BI and SM, a recent survey aimed at businesses and IT executives (Stodder, 2012) indicates that 26% of respondents' organizations are currently analysing data from SM; 22% are planning to analyse such data within one year; and 16% are going to do so within two years. From the same report, the top three objectives of organizations pursuing BI with SM are:

- (a) gaining a deeper understanding of their customers (56%);
- (b) identifying paths to buying decisions (31%); and
- (c) monitoring and measuring sentiment drivers (30%).

It is worth noting that all three objectives relate to gaining a deeper understanding of customers through SM. The above indicates not only the increasing importance of SM for modern organizations but also the compelling need to analyse what is embedded in SM and how this can be used for the benefit of business. The literature indicates that the features embedded in SM include user profiling mechanisms, provisions for creating and sharing user-generated content, tools for communicating, making and sustaining connections (Kim et al., 2010). However, the possibility of appropriating these features for BI has been only partially explored. For example, (Chau and Xu, 2012) concentrate on a multi-method framework for blogs, which may not be applicable to other types of systems. The complexity of the task stems from the fact that there are a variety of systems (e.g., blogging and micro-blogging platforms, pages in social networking services, channels in video-sharing services) that contain a variety of data (e.g., popular opinion on products, services, market trends and information about competitors). Moreover, such information is typically embedded in a number of different artefacts, such as unstructured text, videos, patterns of user actions, etc., which are not uniform and do not afford pre-determined and standard manipulative actions. Complexity also results from constant evolution of SM where new functions (e.g., new versions of public APIs, improved search mechanisms, etc) are introduced in order to update and improve existing services. This makes it harder to find a one-off solution for collecting and analysing data contained in SM; something that could be much more easily achieved in the more stable environment of a traditional business supported forum or a bulletin board. Consequently, there is a need for a methodology to guide both the efficient gathering of the right data at the right time, and their processing and transformation into valuable BI. We believe that our research will contribute to this by describing a method grounded in a conception of viewing SM as virtual settlements.

The notion of a virtual settlement was introduced by Jones (1997) in an effort to understand virtual communities through what he termed 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 the discovery of BI. Specifically, we argue that SM such as social web sites and social networking services host a wealth of data, which are made available with the end users' consent, that reveal not only cultural information about past and / or on-going online communities but also about market trends, consumer behaviour and other business related

insights. Depending on the theoretical footing and research strand, such information can be processed and analysed from various perspectives such as social network analysis (Scott, 1988), ethnographic assessments (Harrison, 2009), data mining or information discovery (Fayyad et al., 1996). This article describes a systematic method for conducting virtual excavations in online computer-mediated settlements and demonstrates the application of this method to BI through a case study of three mobile telecommunication services in Greece.

The article is structured as follows. The section “Theoretical Motivation & Research Objectives” provides a critical appraisal of current conceptions and prominent efforts to advance BI by exploiting SM. Our analysis reveals not only the need for systematizing inquiries but also shortcomings in the type and scope of available perspectives. Building on this, the “Methodology” section describes our method, emphasizing the features that make virtual excavation a useful metaphor for analytical insight. The method is then validated using a case study covering the telecommunications sector in Greece. The article concludes with a critical discussion on the findings, the implications and the limitations of current work and directions for our future work

## **2 Theoretical Motivation & Research Objectives**

Striving for BI has been a long-standing aim of firms in different sectors of industry (Wixom and Watson, 2010). This is evidenced not only by investments made, but also by the intensity of research that characterizes the field. This section offers a synthesis of past and on-going research efforts in the area of BI, which provides the motivation and focus of our present work.

### **2.1 BI and recent advancements**

Since it was first coined in 1958, the term BI has acquired a number of different connotations (Pirttimäki, 2007b). Gilad and Gilad (1985), referred to BI as a process that produces relevant and reliable information for a company’s strategic goals and objectives. Ghoshal and Kim (1986) recognized BI as an essential competitive tool for the collection and analysis of information on markets, new technologies, customers, competitors and broad social trends. Later, Vedder, et al (1999) suggested a dualistic approach referring to BI as both a process and a product. According to Vedder et al, BI as a process is “the set of legal and ethical methods a company use to harness information” and as a product, it is the “information about competitors’ activities from public and private sources”. In her doctoral dissertation, Pirttimäki (2007a) examined several different views of BI describing it as a multifaceted concept that refers to processes, technologies, methods, information products and tools to support managing business information and making faster and better decisions. Computerized systems for BI are commonly referred to as decision making tools (Power, 2007) while data mining techniques (Cabena et al., 1998) are often used to reveal hidden information from massive amount of data. In spite of their high added value, BI efforts are not without their critics. It is often claimed that BI delivers reports and analytical insights ex-post to an event. In other words, by the time BI results are available, key decisions and/or commitments have already been made. Others criticize the high investments made in infrastructure, storage and processing of data, while others point to the lack of systematic and methodological frames of reference to integrate BI processes into organizational routines, which would improve BI’s economic efficiency and efficacy.

Increased recognition of these problems has led researchers to consider a variety of improvements and extensions in the traditional toolset of BI. For instance, efforts concentrate on real-time or near real-time BI (Azvine et al., 2005; Watson et al., 2006) to reduce the latency between operational data and the analysis over those data. Similarly,

cloud based BI solutions aim to address the increased cost of building and maintaining data warehouses (Gurjar and Rathore, 2013). Pervasive BI (McKendrick, 2008; Vesset and McDonough, 2009) and BI 2.0 (Nelson, 2010; Trujillo and Maté, 2012) take advantage of the capabilities of web 2.0 technologies to bring BI to as many users of the enterprise through interactive web applications. Finally, the current discussion of the future of BI (Chen et al., 2012) is concerned with BI and mobile devices (BI 3.0), in particular exploiting data produced by the increasing use of ubiquitous computing and technologies such as RFID, barcodes, and radio tags. As this work concentrates on BI using SM, the following section will focus on more detailed account of work in this area.

## **2.2 BI and related attempts on SM**

SM have become the focus of increased attention in the BI field because most users' daily activity on the Internet takes place in these platforms (Kaplan and Haenlein, 2010). Recent research has examined the impact of SM in decision-making (Meredith and O'Donnell, 2011; Power and Phillips-Wren, 2011); how SM drives business performance in terms of marketing and brand awareness (Callarisa et al., 2012; de Vries et al., 2012); communication and reputation management (Seebach et al., 2012); product development and crowdsourcing (Piller et al., 2012; Tran et al., 2012); human resource management (Kluemper and Rosen, 2009; Roberts and Roach, 2009) and the use of SM to increase enterprises' intellectual capital (Sussan, 2012; Vasileiadou and Missler-Behr, 2011). In addition, SM platforms' openness and interoperability make it easy for third-party applications to exploit their features and the user-generated content they contain for business-related purposes. Nevertheless, the challenges involved are numerous and concern not only technical problems but also issues related to ownership, security and privacy (Rosenblum, 2007). Table 1 summarizes some indicative examples of recent works that have recognized the value of such data and have striven to realize the potential benefits of SM data for the enterprise.

[Insert table 1 near here]

**Table 1: Recent works utilizing SM data**

A review of these studies quickly confirms that a number of different strategies are available for appropriating social data (Kietzmann et al., 2011). A popular approach relies on SM monitoring software (Laine and Frühwirth, 2010; Mayeh et al., 2012) such as Lithium, Jive, Infegy, Radian, TweetFeel, Twitter Sentiment, Jodange and ScoutLabs. Typically, these are deployed by large enterprises with the intention of 'listening' to user opinions and needs (Rappaport, 2011). Another strategy exploits the concept of business-sponsored virtual communities (Füller et al., 2006). These are cyber-structures created by an enterprise to establish a channel for bi-directional interaction with its clientele. These may be used to raise awareness, keep customers informed or stimulate users' involvement in design processes. The choice of infrastructure deemed appropriate for this purpose may vary from social web sites (Hagel, 1999), dedicated practice-oriented toolkits (Akoumianakis et al., 2009), to networking services such as Facebook and YouTube (Mangold and Faulds, 2009). BI tactics also vary. The literature reports on the use of instruments such as design contests, lead user workshops, consumer opinion management, toolkits for user innovation (Von Hippel, 2001), co-design toolkits (Franke and Shah, 2003; Franke and von Hippel, 2003) and communities for customer co-creation (Piller et al., 2012). In all cases, the common ground is the desire to gather data which will bring about competitive advantage, particularly in cases such as firms in highly innovative and knowledge-intensive sectors of the industry, where consumer co-creation, product differentiation and product innovation are valued. More recently, BI efforts have been re-directed from the online exchanges of outsiders to the enterprises' internal staff. This perspective has gained considerable attention from middle and higher level management seeking 'inside-the-company' improvements (DiMicco et al., 2008; Jackson et al., 2007; Kim et al., 2008). An example of this type of BI infrastructure is Socialcast.com – a

social networking platform for employees that mimics the way Facebook users engage in conversation.

### **2.3 SM as Virtual Settlements**

In many ways, the established approaches to BI using SM have come out of the need for enterprises to appropriate the social web and the increased availability of user-generated content. The efforts reviewed earlier are indicative of the importance of BI for modern enterprises. However, they do not follow a common protocol and do not allow for comparisons across cases. Moreover, they are frequently limited to what is 'traceable' using a certain tool at the expense of what is needed and is most useful for an enterprise. To address these challenges, the present work links to efforts that conceive of BI using SM as a type of archaeological inquiry conducted in a virtual settlement. This perspective aims to bring to the surface digital remains that are of potential value to enterprises.

The analogy between computer-mediated communication spaces and archaeological settlements was initially introduced by Jones (1997) in his effort to develop a theory of a cyber-archaeology. In its original formulation the concept of a virtual settlement was conceived as a construct for studying virtual communities through their online cultural artefacts. Jones referred to online infrastructures – places where online communities can be found – as virtual settlements. Jones' view is based on the conception of virtual communities as socially enacted structures whose formation and sustainability are subject to the following criteria: (a) a minimum level of interactivity; (b) a variety of participants; (c) a virtual common-public-space where a significant portion of community interactions occur; and (d) a sustained, stable minimum level of membership. The above establishes the guiding principles for separating virtual settlements from virtual communities. Thus, it is the virtual settlement and not the virtual environment, which is a pre-requisite for an online community. Jones' ideas have been influential amongst social scientists and his cyber-archaeological perspective has been applied to a variety of SM, including Second Life (Harrison, 2009), blogs (Blanchard and Markus, 2004; Chin and Chignell, 2006), the blogosphere (Efimova and Hendrick, 2005) or blogspaces (Zhou and Davis, 2007), micro-blogging services such as Twitter (Gruzd et al., 2011) as well as social networking services like Facebook (Hewitt and Forte, 2006; Zhao et al., 2008) and YouTube (Akoumianakis et al., 2012).

These efforts indicate that SM can be subjected to a virtual archaeological inquiry to solicit insights on both the historical traces of users' past activity and on emerging trends and digital culture. Nevertheless, the diversity in the techniques used and the level of fidelity of the results, point to a need for a structured and organized method. The key issues that need to be address by such a method are:

1. What is the phenomenon under investigation?
2. What traces (or artefacts) can be used to provide evidence of this phenomenon?
3. Are these traces of users' activity readily available, and where and how they can be collected?
4. What type of analysis is required?

To gain further insight, it is worth reminding ourselves that traditional archaeological inquiries provide the means for studying human activities through the data that a previous society has left behind. As direct contact with the creators of such data is lacking, archaeologists are primarily concerned with remains that provide information about their creators' lives and culture. Consequently, the baseline of such an inquiry is rooted in a normative perspective that views culture as something that people do at a particular point in time, but which fades into memory and disappears leaving only the artefacts it created (Lyman and Kahle, 1998). Consequently, the significance and meaning of the material objects that humans create can

be interpreted in relation to the function that they had within the culture in which they were created (Hodder, 2003).

Virtual excavations of SM sites however differ in several ways. Firstly, the digital objects do not necessarily reflect a distant or past culture. In fact, if this were the case, then their practical value for BI would be questionable. Secondly, in virtual excavations cultural participants do not vanish as in the case of traditional archaeological inquiry. Indeed, virtual excavations for BI should be targeted to locating, identifying and following the activities of current participants, as these constitute a live source of useful information. Finally, virtual excavations are more aligned to facilitating an emergent knowledge process (Markus et al., 2002) rather than an improved understanding of the past. The above has consequences for the theoretical framing of virtual excavations as well as the methods to be used. In what follows, we briefly describe how these implications are dealt with the present work.

## **2.4 Research focus**

In our analysis so far, we have argued that SM retain data which can bring to the surface 'hidden' knowledge that can be of potential value to an enterprise, and that this knowledge, once revealed, can have implications for a firms' business strategy. Nevertheless, the means through which this data is managed (i.e., extracted, filtered and transformed), remains a challenge. This is particularly the case when the data resides in multiple SM. We believe that an approach based on the concept of virtual excavations provides one solution to this problem. The key issues that must be addressed in devising such an approach are:

1. What is relevant digital trace data and what artefacts leave the required traces?
2. Where do they reside and what boundaries prohibit their access?
3. How are they retrieved and what sense can be made of them?

Although there is work that attempts to address these issues, the results have been sub-optimal as they fail to systematize BI using SM and do not provide a scalable and customizable technique. This is primarily due to constraints stemming from the perspective adopted and their underlying philosophy. Thus, for example, information systems researchers frequently emphasize the technicalities of SM and devise methods (i.e., algorithms, crawling strategies, screening methods, etc.) which are only applicable to certain platforms; consequently, cross-media comparisons cannot be made. Similarly, management scientists and organizational researchers tend to neglect the technicalities of different SM platforms, concentrate on aggregate artefacts (e.g. blogs) and extract embedded intelligence using text processing methods (He et al., 2013) or document visualization techniques (Abbasi and Chen, 2008). This level of analysis is useful, but is disassociated from the meaning of actions, such as pressing a 'like' button, which demonstrate a different form of interaction between users.

The above are some of the factors that have driven our desire to develop a more systematic and organized treatment of SM that can be used for BI purposes. To this end, the present work constitutes part of an on-going attempt to advance BI through virtual excavations of SM. In this vein, our baseline is that by extracting and processing digital trace data, we can foster new BI capabilities, just as archaeologists create new understandings of the past by excavating physical remains.

## **3 Methodology**

To test the concept of virtual excavations as a method for extracting BI, we undertook a case study of three prepaid (pay-as-you-go) mobile telecommunications service brands that operate in Greece using their fan pages on Facebook. For anonymity, we will refer to the

three brands as 'A', 'B' and 'C'. According to (Socialbakers.com, 2012) the telecommunication sector is the most socially devoted sector with a 60.4% response rate for their Facebook fan pages. The same source also acknowledges that the three Facebook fan pages of this study are ranked in the 6th, 55th and 56th position among Greek Facebook fan pages. Table 2 depicts data regarding the fan base and the activity of the Facebook pages. The figures displayed were retrieved from Facebook on August 20, 2012. It should be noted that the three telecommunication companies offer not only prepaid plans but also regular plans; our focus is only on the Facebook pages for the prepaid plans.

[Insert table 2 near here]

**Table 2: Facebook fan pages**

Prior to launching the virtual excavation, a detailed analysis of the virtual settlement was undertaken to identify relevant artefacts and the associated actions that create the digital traces that are retained and made available through the platform's public API. This analysis confirmed Facebook's provisions for creating and sharing user-generated contents in specific forms and under certain conditions. For our purposes, Facebook fan pages constitute the top-level digital artefact through which users engage in interactions with the company and with other users. Such pages frequently act as identity mechanisms, inviting users to externalize their opinions or state of mind.

The 'glue' bringing together different people is the page's wall. This is as much an online meeting point as a newsfeed aggregator, acting as the host of a wealth of useful information. In the case of the three brands in our study, such information may untangle the company's cultural values as revealed by the presence or absence of an eco-friendly profile, its commitment to sustainable development, policy declarations, etc. This information can be found in digital objects such as photos, videos or textual announcements that the firm uploads on its fan page wall and reach their fans through each fan's news feed. First-time visitors can also declare positive feedback to a page by hitting a button that is labelled with a thumbs-up icon and the word 'like'; this will automatically make the user a fan of the web page with all contents published by the page being displayed in their news feed.

The functionality of the 'like' button can also be found in other artefacts of Facebook such as comments, photos, videos, etc., thus making it possible for users to externalize intentions by sharing posts, declaring 'likes' and commenting in the page's Facebook wall in a manner that is visible to other users. Such micro-interactions can lead to aggregate formations, such as friends lists, that are useful cultural bindings and reveal the users' social circles. In light of the above, it is argued that digital remains on Facebook designate on the one hand the brand's identity, intentions and cultural values, and on the other hand, the fans' opinions, concerns and pre-occupations. Such digital remains are manifested as traces of the users' actions with artefacts such as fan pages, wall posts, friend lists, posts, comments, etc.

Table 3 summarizes what were deemed to be relevant artefacts and what sorts of digital traces can be compiled. Depending on purpose and scope of BI inquiry, the type and range of artefacts may be extended or narrowed, which will impact on the digital trace data sample.

[Insert table 3 near here]

**Table 3: Cultural Artefacts of Facebook**

### **3.1 Data samples and instruments**

The data on which the present research is grounded was extracted on August 24, 2012 and covers a period of three and a half years including interactions dated from the 5th February 2009 to 23rd August 2012. Table 4 presents the identity of the data sample. It is worth noting that several of the traces suggested by our analysis (see Table 3) could not be captured



either due to limitations of the platform's API or due to privacy issues. Nevertheless, our data sample is considered fairly complete and adequate for the intended purposes.

[Insert table 4 near here]  
**Table 4: Excavation identity**

Due to a specificity of the Greek internet where users frequently communicate through 'Greeklsh' – a transliteration of Greek using the Latin alphabet – we used a 'Greeklsh' to Greek text transformation engine (Chalamandaris et al., 2006) to properly visualize our data set through specific tools and count word frequencies. No further screening or filtering of the data was performed. In terms of qualifying users as fans of a page, it was assumed that Facebook users declaring a positive feedback on a page using the 'like' button are fans of the page. We were not able to extract who these users were using the Facebook API, we could only access the number of total users for each page as presented in Table 2. All Facebook fan pages considered in this study were open, which means that every Facebook user is allowed to create or 'like' content (e.g. wall posts, comments) that appears on the pages' walls. As we were not able to access information about whether or not active fans (i.e. fans where that performed activities in a page's wall) were actual fans of the pages (i.e. to have hit the page's 'like' button), for reasons of simplicity we decided to consider active fans as actual fans of the pages.

The data extracted was stored in a relational database representing details such as type of trace, owner, time of creation etc. This allowed us to calculate metrics such as operator response rates and times to active fans' questions. Nevertheless, our primary interest has been to examine more complex cyber-phenomena that qualify the parties engaged, the types of artefacts chosen to convey contributions and cross-page activities of contributors. In terms of analytical instruments, our work relies on the use of a set of visualizations crafted specifically to support the cyber-phenomena under investigation. As discussed in the following section, a variety of visualizations were used to derive useful insights on aggregate patterns of activity as well as factors determining the relative success of different brands during certain periods.

## **4 Findings**

This section discusses the most prominent findings leading to ascertaining the added value of our approach. In doing so, we do not aim for an exhaustive treatment of all possible BI related issues due to space limitations. Thus, the focus is limited on a relevant subset of key findings by following a chronological account of the activities undertaken, starting with the role of the fan page operators and moving to an analysis of interaction patterns revealed.

### **4.1 Contributors, type of contributions, response rates and times**

Our data sample represents online activities undertaken by three types of contributors, namely page operators (n=3); active fans (n=42,914) and casual contributors (n=16,485). The operators are the users acting as representatives of the three brands and as the administrators of each page. Active fans are Facebook users other than page operators who contribute (in whatever way we could capture using the API) to at least one of the three pages. Contributors are Facebook users who create content through wall posts or comments.

Attempting to understand the attention paid by brands to their social networking fans, it was deemed appropriate to examine the operators' response patterns in the fans' wall posts and the associated time differences in response time. Such measurements are typical BI metrics widely acknowledged by SM monitoring tools (Socialbakers.com, 2012). Nevertheless, these metrics are not enough to establish informative indications of the quality and the rate of the

responses nor how much attention is paid by the operators to users' questions or complaints. Deeper and more involved analysis of the choice of artefacts used as well as their design affordances can provide further useful insights. In terms of artefacts used, there was an overwhelming preference for comments, replies to comments and 'like' buttons for expressing positive opinion. Moreover, comments may be intertwined in structured sequences of 'comment-reply' depicting comment replies to the operator's response. As an illustration of the above, Figure 1 depicts two instances of users' wall posts with comment replies created by the fan page operators. In one case (wall post A) the operator's comment has been 'liked' by the creator of the wall post while in the other case (wall post B) there is a comment reply with the phrase 'Thank you very much' in Greek.

[Insert figure 1 near here]

**Figure 1: Operators responses on users' wall posts**

Tracing and analysing these exchanges (i.e., counting comment replies) and taking account of the owner (i.e. who replies) as well as their contents (i.e. whether they contain expressions such as 'thank you' either in Greek (i.e. /ευχαριστώ/) or in English) can be useful. From a total of 10,503 Facebook wall posts traced, (A=7,619; B=1,572; C=1,312), 936 (8.91%) were posted by the operator of the fan page, while 9,567 (91.09%) were posted by fans. In an effort to qualify the posts representing questions, content analysis was undertaken. A post was classified as a question if it contained at least one occurrence of a question mark (i.e. the symbols ';' in Greek and '?' in English) in the body of the post. Special markers such as smiles that combine the characters ';' and ')' were found in 73 wall posts (e.g., approximately 0.70% of total wall posts) and those wall posts were not qualified as questions. From the total of fan wall posts, 5,641 (58.96%) had question marks in their content. These wall posts were made by 3,592 distinct users (A=2,870; B=474; C=347). Operators responded to 4,466 (46.68%) of total fan wall posts. For wall posts qualified as questions, operators responded in 3,405 (60.36%) of them. The number of wall post creators in each brand does not sum up to the number of distinct users reported, as users were found to cross post in more than one fan page. Table 5 provides further details and summarizes our findings by brand.

[Insert table 5 near here]

**Table 5: Response rates by operators**

The average operator response time (i.e. time passed from the creation of the wall post to the first operator response) to a fan question was 924 minutes (A=909; B=934; C=998). More specifically, there were 1,288 (37.83%) fan questions that had an operator response in less than an hour; while in 249 (7.31%) questions the responses were made in less than five minutes. In 553 (16.24%) fan questions, the operator response time was more than a day. Table 6 summarizes the response times.

[Insert table 6 near here]

**Table 6: Response times**

Fan questions had in total 14,172 comments, with 4,473 (31.59%) made by operators, while 9,699 (68.44%) by the fans. Out of the total fan comments, 5,016 were left by the creators of the question. Those comments were made on 2,794 different questions. More specifically, fans left 3,759 comments on 2,090 of their own questions with a response by the operator. Gathering these 3,759 comments, extracting their contents and visualizing them can reveal popular words, with either positive or negative connotation. We have used word cloud visualizations (see Figure 2) with the following colour coding scheme. Green designates the words 'thank', 'ευχαριστώ' (i.e. 'thank you') and 'ωραία' (i.e. 'nice') allowing for spelling errors and related abbreviation such as 'thnk', 'thx', etc. Red is used for the words 'πρόβλημα', 'ντροπή', 'κακό' that stand for 'problem', 'shame' and 'bad' in English, respectively. Needless

to say that there could be any other combination of words that justifies a certain cause. Moreover, the results of measuring frequencies of word occurrence for a selection of words implying negative or positive meaning are summarized in Table 7.

[Insert figure 2 near here]  
**Figure 2: Word cloud**

[Insert table 7 near here]  
**Table 7: Word frequencies**

From the 3,759 comments posted by fans on their own questions and responded by the operator, we counted how many of them had a gratitude expression. A total of 1,118 comments were found to contain a gratitude expression made by the question creator. These comments were made in 1,031 fan questions that represent the 49.33% of the questions that had a response by an operator. In 945 (91.66%) questions, gratitude comments were made after the first operator response while in 86 (8.34%) questions the gratitude comments made before the first operator response.

Focusing on 'likes', a total of 525 fan questions with a response by the operator (i.e., 15.42% of total questions with operator responses) were 'liked' by the creator of each question (see Table 8).

[Insert table 8 near here]  
**Table 8: Fan questions, comments and 'likes'**

In light of the above, we can argue that the telecommunication companies considered in this study are evidently paying attention to their social networking fans. This is evident not only from the response rates, which amount to almost 50% of the questions posted by the fans, but also by the timing of the operator's response which in some cases was immediate (i.e. a few seconds) resulting in an average response time of 15.4 hours. Brand 'A' was found to score the lowest response rate (56.86%) in contrast to the highest response rate (73.32%) of brand 'C'. It turns out that brand 'A' is the most promptly responded with a median response time of approximately 3.3 hours and a minimum response time of 16 seconds. On the other hand, the less promptly responded brand appears to be brand 'C' with a median response time of 10.7 hours and 19.04% of responses made after a day. The above metrics show the type of results that can be extracted by the traditional SM monitoring tools.

What it is not clear from these results so far is to what extent fans were satisfied with the responses received by the operators of the fan pages. To this end, it should be noted that although Facebook was not built to support transactions of the kind that modern customer relationship management (CRM) systems are built to provide, yet it is used as a CRM system (Prestus and Bygstad, 2010). Sentiment analysis (Pang and Lee, 2008) of textual content, as useful as it may be in some cases, can lead to obscure findings. We do not claim that a full-scale sentiment analysis was performed in this study. Nevertheless, it is evident that the word 'thank' either in Greek or English was found to have more appearances on brand 'A' with 4.92% of total words that were identified on the comments posted by the creator of the question (on questions that had a response by an operator). This equates to approximately one occurrence every 20 words while brand 'B' was found to have one occurrence every 23 words and brand 'C' to have one occurrence every 33 words. By focusing on the gratitude expressions and the ratios that are being identified, it is revealed that brand 'A' has the lower percentage (47.93%) of questions with a gratitude expression when the operator responded; while brand 'B' scores almost 56%.

Additional information of consumer satisfaction may be found in non-textual information (chronological data) or the users' choice of interaction elements such as 'like' buttons. By

counting how many operator responses to fan questions had been 'liked' by the creator of the question, Brand 'B' receives 'likes' in 20% of their responses where brand 'A' is limited to 14.10% despite the fact that it is the brand that responds most promptly. Further investigation on the questions with gratitude expressions showed that for brand 'B', in 97.55% of questions the comment with the gratitude expression was submitted after the last operator response. In contrast, brand 'A' had almost 10% of those questions having a comment with a gratitude expression submitted before the operator had submitted a response. From the above, it can be argued that what is 'hidden' in 'banal' manipulative actions (i.e., 'like' button presses) or other micro-interactions between the users can be used to reveal further information of value for the enterprise such as the level of consumer satisfaction, and better understand the engagement of users in such transactions.

## 4.2 Active fans and contributors

Active fans are users having at least one activity of any kind and contributors are users posting at least one wall post or comment. Our analysis was aimed at tracking such fans with activities in more than one page, which should be more likely to yield useful comparisons between brands, and detecting events that created impact for fans and active contributors.

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. Figure 3 depicts such a visualization using our data set. In this graph, users are represented with green nodes and Facebook pages with red nodes. User nodes are connected to the page nodes in the case of at least one activity performed by the user in that page. The size of the user nodes can be adjusted to meet the specific properties of the user (e.g. the users presence in different groups, the number of the activities they performed for the connected pages). When the mouse is over a node, the edges connecting that node to other nodes are highlighted (in red) and a number appears at the middle of the edge, indicating the number of distinct actions of the user within the connected group. A force directed layout (Barnes and Hut, 1986) has been applied in order to position users according to pages where they are active.

[Insert figure 3 near here]

**Figure 3: Active users on Facebook pages (Top graph represents the entire data set while the bottom a zoom area)**

In light of the above, Figure 3 in the outer circle reveals three concentrations of users around the page nodes of the three brands. Each of these clouds indicates activity on a single page. In the inner parts of the graph, there are activity aggregations representing contributors to brand 'A' and 'B', brand 'A' and 'C' and brand 'B' and 'C' respectively. Finally, one group of nodes at the centre of the graph indicates users that have activities in all three Facebook pages.

Further insights are obtained by studying the users' contribution in time. For this purpose, the data set was reconstructed as shown in Figure 4. The graph depicts timelines of user activities (wall posts, comments, 'likes') ordered by time of creation. Each line of the graph (differentiated by white and grey stripes) represents one user. The coloured marks in each user's line indicate an interaction with a Facebook page's wall. Green marks indicate activity on the wall of brand 'A'; red marks indicate activity on the wall of brand 'B'; while blue marks are activities on the wall of brand 'C'. Within the timeline of a user we can find two different shapes of marks. The round marks indicate activity of type 'like' while the square marks depict wall posts or comments. Round marks placed at the top of the user's timeline represent 'likes' on wall posts while those at the bottom depict 'likes' on comments. Similarly, square marks placed topmost indicate wall posts while the square marks at the bottom

indicate comments. When clicking on a mark, a pop up window summarizes additional information about that activity. Traversing the graph from left to right provides a chronological history of users' activity with the most recent placed to the right of the timeline. It is worth mentioning that for 'likes', Facebook API doesn't allow access to creation time. In order to overcome this problem, we have considered as creation time of the 'like', the creation time of the corresponding artefact (e.g. comment, wall posts) on which the 'like' was performed.

[Insert figure 4 near here]

**Figure 4: Timeline of users' activities**

A total of 42,914 active fans had 134,090 interactions on the walls of the three brands. In this sample, the operators' activities have been excluded. Of the total activity captured, 59.41% was on the page of brand 'A', 21.64% was on the page of brand 'B' and 18.95% was on the page of brand 'C'. 56.15% (24,095) of active fans made just one activity while 4.95% (2,123) made ten or more activities. 38.41% (16,485) of total active fans contributing with at least one wall post or a comment. These users (i.e. contributors) made 66.61% (89,320) of total activities (excluding operators' activity). That is 35,545 comments; 9,567 wall posts; 41,010 likes and 3,198 tags on photos or wall posts. 26.30% (4,335) of total contributors made just one activity while 10.96% (1,806) made ten or more activities. Table 9 summarizes these data for each brand.

[Insert table 9 near here]

**Table 9: Active fans and contributors**

Regarding cross posting in multiple pages (see Table 10), 94.46% (40,535) of the total number of active fans had activities on only one Facebook wall; 4.94% (2,120) fans had activities on two different walls and only 0.60% (258) of fans had activities on all three walls. If we only focus on contributors, then 89.49% (14,752) of the total contributors had activities related to just one page; 9.11% (1,502) had activities in two pages and 1.40% (231) had activities on three pages.

[Insert table 10 near here]

**Table 10: Cross-posting fans**

Only 38.41% of total active fans fall in the category of contributors, indicating that 61.59% of total users contribute only 'likes'. These users, although they may be considered as the audience of the Facebook fan page and their collective activity may play an important role, are characterized by a relatively low degree of engagement with the brands. The same applies to the 56.15% of active fans that have only created one activity. As further study of these casual contributors was not deemed necessary, subsequent analysis focused on the activities of the most active contributors (i.e. those having 10 or more activities).

Figure 5 presents the contributors with ten or more activities of any type. In this graph, the size of the node depicts the total number of wall posts of the user. A closer look at Figure 5 reveals that for brand 'B' and brand 'C', more than half of contributors with ten or more activities are cross posting while the greatest part of contributors that have activity in brand 'A', are not cross posting. In order to gain insights into customer opinions and preferences between brands, we further narrowed most active contributors to those cross posting in two or more pages. These users are more likely to make comparisons between brands and their activity will allow us to spot events that had more impact for them.

[Insert figure 5 near here]

**Figure 5: Contributors that had ten or more activities of any type**

Figure 6 presents timelines of the contributions by users with ten or more activities of any type in two or more pages. Careful inspection of this graph reveals areas of particular interest. Four of them have been annotated for the purposes of this discussion. Areas A and C reveal segments of the timeline activity of two users, highlighting points in time that users appear to switch their activities from one brand's wall to another. Figure 7 and Figure 8 untangle user activities in Area A and Area C in terms of type and content while Table 11 and Table 12 provide a translation in English of the content and further details about these activities. Apparently, these two users had substantial activity in a short period in two different walls. Looking closer at the first fan's activity (see Figure 7), it becomes apparent that at that time the user was actively engaged with brand 'A' and brand 'C'. It is worth noticing that the activity presented there, was performed in a period of three days. The fan created 'likes', wall posts and comments. The fan's posts on the wall of brand 'A' reveal a frustration about the signal coverage. This is depicted not only through wall posts (A-3, A-5) and comments (A-1) but also through the user's 'likes' of comments that express the same frustration (A-2). On the wall of brand 'C' the fan is reproducing advertisements (C-3) of brand 'C' and gives information to other users (C2). The second fan was rather frustrated about the lack of competitiveness among brands. The fan was found to be seeking information about moving from brand 'C' to brand 'A' but was informed by other users that brand 'A' is confronted with similar issues.

[Insert figure 6 near here]

**Figure 6: Timeline of contributors that had ten or more activities of any type on two or more pages**

[Insert figure 7 near here]

**Figure 7: Zoom on area A of Figure 6**

[Insert figure 8 near here]

**Figure 8: Zoom on area C of Figure 6**

[Insert table 11 near here]

**Table 11: Translated content from area A**

[Insert table 12 near here]

**Table 12: Translated content from area C**

Finally, examining events that created impact on Facebook for these three brands we focused on two areas of the timeline visualization of Figure 6. Area B depicts activities that were produced during a three-day contest for brand 'A'. Users were asked by the operator to create 'likes' or comments in three wall posts – one for each day – in order to take part in the contest. To a similar extent, the activity of area D was created when the operator of brand 'B' posted one wall post asking the users to mention their friends (i.e. tag) in a comment in order to take part to the contest. Brand 'A' in its contest did not use the tagging functionality, as tagging was not available by Facebook at that time. From the analysis of those wall posts, it became evident that brands take advantage of the currently available functionalities (i.e. affordances of artefacts) offered by the virtual environment to create a buzz around their brand.

## 5 Discussion

In this section we discuss our findings, mainly at the level of the method, and focus on the value of an approach based on virtual excavations for BI. In particular, we focus on the three key issues identified in the paper's research focus section, namely (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 of

them. Following this, we highlight some of the broader implications for our work, as well as its limitations and the directions it might take in the future.

### **5.1 Virtual excavations and BI**

Virtual excavations, as presented in this article, should not be understood as merely web analytics. Their capabilities extend far beyond tracing the user's online activities and revealing the user's browsing and purchasing patterns. The full benefit of virtual excavations are realized when they are structured and organized around aggregate artefacts embedded in a technology. Virtual excavations constitute an approach for understanding online practice through its digital remains. This is achieved through the ability to interrelate different digital traces, which otherwise might appear as banal interactions, to create a more accurate picture of the context in which these interactions took place. For example, as we saw in the findings section, the sequence in which operator responses (e.g. likes and comments) are performed or frequent and recurrent posts by the same user of certain types of messages (e.g. including the prefix 'http://' or gratitude expressions), can be useful in making sense of the user's original intentions. In this sense, virtual excavations can provide not only retrospective accounts of historical data but also act as predictive analytics, simulating and assessing potential future outcomes grounded in the parameters of reliable historical trends. For example, our case study provided valuable insights into the effects of certain service strategies associated with the three brands. It revealed that brands that are quick to diversify through unlimited plans experienced high end user concentration, which then led them to initiate strategies to protect market share and customer base. Similarly, it would be possible to create projections of how external factors, such a legislation, marketing campaigns or television programmes, could influence consumers in their search for competitive services. We believe that virtual excavations for BI should be seen as part of an emergent knowledge process (Markus et al., 2002) that results from an unpredictable set of actors, unclear requirements and deliberations that have no pre-determined structure or sequence. This situation is precisely why methods are needed to guide and organize what is to be traced and where and how sense is to be made of it.

### **5.2 Implications**

We will now briefly examine some of the implications of our work. Firstly, a pre-requisite for virtual excavations is the capability to locate and trace the appropriate set of digital remains. These need not be limited to the type offered by web analytics tools. Our experience shows that there is added value to be gained by a targeted consideration of the affordances of artefacts embedded in technology, the manipulative actions they invite, and the traces these actions leave behind. Such a focus, amongst other things, allows us to ascribe meaning to otherwise banal, keystroke level activities and guides our analysis and exploration of the data set. In our work, we have attempted this for Facebook (see Table 3), which has led not only to envisioning appropriate visualizations but also to highlighting shortcomings in the API that restricts traces of certain types. Thus, it is argued that approaches capitalizing on virtual excavations provide a structured method to untangle the hidden knowledge embedded in massive data sets originated from SM, which is achievable in a timely and purposeful manner and is flexible enough to account for the variety of hosting services and data formats i.e., textual, tabular, images, videos, channels, 'likes', 'tweets', etc.

This brings to the forefront the second implication of the present work, which relates to identifying where the relevant artefacts reside and what might prohibit their access. In an increasingly interconnected social web, relevant artefacts are spread across different virtual settlements, with variable capabilities and public APIs. At present crossing boundaries of virtual settlements is not straightforward. Although there are provisions such as sharing widgets, they are not sufficient to bring to the surface the 'hidden' knowledge that characterizes user communities. At the same time, recent improvements in semantic APIs

only partially capture online traffic. The complications arising from this type of constraint were prominent in the analysis of the case presented in this article (see section on limitations and future work below).

Finally, it should be noted that the visualizations we used in the work described in this article provide only one approach to the way we might make sense of digital remains. They were chosen because visualizations constitute a powerful transformational technology. Nevertheless, they require a substantial investment to create and use since the tools that are available need adjustment and, in some cases, substantial refinements, extensions and enhancements before that can be used. Additionally, in order to deal with the sheer volume of data, computational efficiency is also an important factor. Consequently, if visualizations are to be used, it is important for researchers to be well informed about the technological implications of what they wish to do and to make good use of appropriate visualization toolkits.

### **5.3 Limitations and future work**

Our research has highlighted a number of limitations with this approach; with some of them being pursued as part of our on-going activities. Firstly, our current work addresses a single platform, namely Facebook. However, our attempts to generate useful BI would benefit from investigations covering multiple virtual settlements and platforms. Specifically, in our reference case, the three brands maintain presence in video-sharing platforms such as YouTube to share TV adverts and other information. YouTube provides an open API and it would be of value to apply the method described in this article to produce complementary data relating to designing contests, inviting suggestions on adverts and so on. However, our initial intention to undertake such a cross settlement excavation, taking into account the presence of the three brands in both Facebook and YouTube, was plagued by problems associated with lack of interoperability, structural mismatches between the two platforms and an inability to form complete mappings between aggregate artefacts and their affordances.

Secondly, our experiences indicate that there are cases where it may not be possible to extract the full range of digital trace data that is needed. This can be due to either the restricted design affordances of certain artefacts (e.g. a user can be tagged in a comment or a photo) or the limitations of the APIs. For example, in our case study, it was not possible to access the time 'like' button traces are created. This is further complicated in cross-settlement analyses where the features of the various APIs that are available may pose challenges.

Finally, regarding the techniques used to make sense of digital traces, the present work relies primarily on visualizations. These however could be complemented by, for example, social network analysis, observation or interviewing strategies. Social network analysis might bring insights into patterns of relationships and interactions amongst users and lead to discovering the underlying social structure in user networks while observation or interviewing might provide contextual insights into user's behaviour, both of which would complement an analysis based on techniques of visualization.

### **5.4 Concluding remarks**

The work we discuss in this article examines several scenarios to illustrate the notion of digital traces, how they are retained and the way in which they may obtain added value for enterprises. It also hints at the methodological challenges that need to be addressed when designing and conducting virtual excavations. These relate to defining the virtual settlement and its boundaries, what traceable remains are useful and how they are revealed as well as what techniques may be employed to add value to bytes of code. Our on-going work concentrates on extending virtual excavations and the associated techniques needed to



facilitate insights into boundary spanning in virtual settings. Specific challenges in this direction include the development of mechanisms for locating digital remains across virtual settlements and aggregating them in a manner that will lead to establishing richer insights on a community's presence and practice across virtual settings and boundaries. Amongst other things, this is expected to test affordances that enable or constrain the interoperability of virtual settlements and the mechanics through which it is attained.

## 6 References

- Abbasi, A., Chen, H., 2008. CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication. *MIS Quarterly* 32, 811–837.
- Abrahams, A.S., Jiao, J., Wang, G.A., Fan, W., 2012. Vehicle defect discovery from social media. *Decision Support Systems* 54, 87–97.
- 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., Stefanakis, D., Akrivos, A., Vellis, G., Kotsalis, D., Plemenos, A., Vidakis, N., 2009. Virtual Operations in Common Information Spaces: Boundary Objects and Practices, in: *Leveraging Knowledge for Innovation in Collaborative Networks*. pp. 207–216.
- Azvine, B., Cui, Z., Nauck, D.D., 2005. Towards real-time business intelligence. *BT Technol J* 23, 214–225.
- Barnes, J., Hut, P., 1986. A hierarchical  $O(N \log v)$  force-calculation algorithm. *nature* 324.
- Blanchard, A.L., Markus, M.L., 2004. The experienced sense of a virtual community: Characteristics and processes. *ACM SIGMIS Database* 35, 65–79.
- Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of Computational Science*.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A., 1998. *Discovering data mining: from concept to implementation*. Prentice Hall Upper Saddle River, NJ.
- 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, 73–79.
- Castellanos, M., Dayal, U., Hsu, M., Ghosh, R., Dekhil, M., Lu, Y., Zhang, L., Schreiman, M., 2011. LCI: a social channel analysis platform for live customer intelligence, in: *Proceedings of the 2011 International Conference on Management of Data, SIGMOD '11*. ACM, New York, NY, USA, pp. 1049–1058.
- Chalamandaris, A., Protopapas, A., Tsiakoulis, P., Raptis, S., 2006. All Greek to me! An automatic Greeklissh to Greek transliteration system, in: *Proceedings of the 5th Intl. Conference in Language Resources and Evaluation*. pp. 1226–1229.
- Chau, M., Xu, J., 2012. Business Intelligence in Blogs: Understanding Consumer Interactions and Communities. *MIS quarterly* 36, 1189–1216.
- Chen, H., Chiang, R.H.L., Storey, V.C., 2012. Business intelligence and analytics: From big data to big impact. *MIS Quarterly* 36, 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*. San Sebastian, Spain, pp. 191–200.
- Davenport, T.H., 2006. *Competing on analytics*. Harvard Business School Publishing Corporation, Boston.
- De Vries, L., Gensler, S., Leeflang, P.S.H., 2012. Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing* 26, 83–91.
- DiMicco, J., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B., Muller, M., 2008. Motivations for social networking at work, in: *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*. pp. 711–720.

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

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. *AI magazine* 17, 37.

Franke, N., Shah, S., 2003. How communities support innovative activities: an exploration of assistance and sharing among end-users. *Research policy* 32, 157–178.

Franke, N., Von Hippel, E., 2003. Satisfying heterogeneous user needs via innovation toolkits: the case of Apache security software. *Research Policy* 32, 1199–1215.

Füller, J., Bartl, M., Ernst, H., Mühlbacher, H., 2006. Community based innovation: How to integrate members of virtual communities into new product development. *Electronic Commerce Research* 6, 57–73.

Ghoshal, Kim, S.K., 1986. Building Effective Intelligence Systems for Competitive Advantage. *Sloan Management Review* 28, 49–58.

Gilad, B., Gilad, T., 1985. A systems approach to business intelligence. *Business Horizons* 28, 65–70.

Gruzd, A., Wellman, B., Takhteyev, Y., 2011. Imagining twitter as an imagined community. *American Behavioral Scientist* 55, 1294–1318.

Gurjar, Y.S., Rathore, V.S., 2013. Cloud Business Intelligence - Is What Business Need Today. *International Journal of Recent Technology and Engineering* 1, 81–86.

Hagel, J., 1999. Net gain: Expanding markets through virtual communities. *Journal of Interactive Marketing* 13, 55–65.

Harrison, R., 2009. Excavating Second Life. *Journal of Material Culture* 14, 75–106.

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*.

Hewitt, A., Forte, A., 2006. Crossing boundaries: Identity management and student/faculty relationships on the Facebook. Poster presented at CSCW, Banff, Alberta.

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, 29–37.

Jackson, A., Yates, J.A., Orlikowski, W., 2007. Corporate Blogging: Building community through persistent digital talk, in: *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference On*. pp. 80–80.

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

Kaplan, A.M., Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons* 53, 59–68.

Kavanaugh, A.L., Fox, E.A., Sheetz, S.D., Yang, S., Li, L.T., Shoemaker, D.J., Natsev, A., Xie, L., 2012. Social media use by government: From the routine to the critical. *Government Information Quarterly* 29, 480–491.

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*.

Kim, S.T., Lee, C.K., Hwang, T., 2008. Investigating the influence of employee blogging on IT workers' organisational citizenship behaviour. *International Journal of Information Technology and Management* 7, 178–189.

Kim, W., Jeong, O.R., Lee, S.W., 2010. On social Web sites. *Information Systems* 35, 215–236.

Kluemper, D.H., Rosen, P.A., 2009. Future employment selection methods: evaluating social networking web sites. *Journal of managerial Psychology* 24, 567–580.

Laine, M.O.J., Frühwirth, C., 2010. Monitoring Social Media: Tools, Characteristics and Implications, in: *Tyrväinen, P., Jansen, S., Cusumano, M.A. (Eds.), Software Business, Lecture Notes in Business Information Processing. Springer Berlin Heidelberg*, pp. 193–198.

Li, H., 2005. Applications of data warehousing and data mining in the retail industry, in: *Services Systems and Services Management, 2005. Proceedings of ICSSSM'05. 2005 International Conference On*. pp. 1047–1050.

- Luhn, H.P., 1958. A business intelligence system. *IBM Journal of Research and Development* 2, 314–319.
- Lyman, P., Kahle, B., 1998. Archiving digital cultural artifacts. *D-Lib Magazine* 4.
- Mangold, W.G., Faulds, D.J., 2009. Social media: The new hybrid element of the promotion mix. *Business horizons* 52, 357–365.
- Markus, M.L., Majchrzak, A., Gasser, L., 2002. A Design Theory for Systems That Support Emergent Knowledge Processes. *MIS Quarterly* 26, 179–212.
- Mayeh, M., Scheepers, R., Valos, M., 2012. Understanding the role of social media monitoring in generating external intelligence, in: *ACIS 2012: Location, location, location: Proceedings of the 23rd Australasian Conference on Information Systems 2012*. pp. 1–10.
- McKendrick, J., 2008. Pervasive Business Intelligence Means “BI for the Masses”. *Informatica*.
- Meredith, R., O'Donnell, P., 2011. A Framework for Understanding the Role of Social Media in Business Intelligence Systems. *Journal of Decision Systems* 20, 263–282.
- Nelson, G.S., 2010. Business Intelligence 2.0: Are we there yet, in: *Proceedings of the SAS Global Forum*.
- Pang, B., Lee, L., 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2, 1–135.
- Piller, F., Vossen, A., Ihl, C., 2012. From Social Media to Social Product Development: The Impact of Social Media on Co-Creation of Innovation. *Die Unternehmung* 65.
- Pirttimäki, V., 2007a. Business intelligence as a managerial tool in large Finnish companies.
- Pirttimäki, V., 2007b. Conceptual analysis of business intelligence. *SA Journal of Information Management* 9.
- Power, D.J., 2007. A Brief History of Decision Support Systems [WWW Document]. *DSSResources.COM*. URL <http://dssresources.com/history/dsshhistory.html> version 4.0 (accessed 10.17.12).
- Power, D.J., Phillips-Wren, G., 2011. Impact of Social Media and Web 2.0 on Decision-Making. *Journal of Decision Systems* 20, 249–261.
- 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)*.
- Rappaport, S.D., 2011. *Listen First!: Turning Social Media Conversations Into Business Advantage*. John Wiley & Sons.
- Roberts, S.J., Roach, T., 2009. Social Networking Web Sites and Human Resource Personnel. *Business Communication Quarterly*.
- Rosenblum, D., 2007. What anyone can know: The privacy risks of social networking sites. *IEEE Security & Privacy* 40–49.
- Scott, J., 1988. Social network analysis. *Sociology* 22, 109–127.
- Seebach, C., Beck, R., Denisova, O., 2012. Sensing Social Media for Corporate Reputation Management: A Business Agility Perspective. *ECIS 2012 Proceedings*. Barcelona, Spain
- 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, 470–479.
- Socialbakers.com, 2012. Socially Devoted [WWW Document]. *Socialbakers.com*. URL <http://sociallydevoted.socialbakers.com/> (accessed 8.23.12).
- Stodder, D., 2012. Customer Analytics in the Age of Social Media, *TDWI Best Practice Report Q3*. TDWI -The Data Warehousing Institute.
- Sussan, F., 2012. Consumer interaction as intellectual capital. *Journal of Intellectual Capital* 13, 81–105.
- Tran, A., Ul Hasan, S., Park, J., 2012. Crowd Participation Pattern in the Phases of a Product Development Process that Utilizes Crowdsourcing. *Industrial Engineering & Management Systems* 11, 266–275.
- Trujillo, J., Maté, A., 2012. Business Intelligence 2.0: A General Overview. *Business Intelligence* 98–116.

- Tyson, K.W.M., 1986. Business intelligence: putting it all together. Leading edge publications, Lombard, IL.
- Vasileiadou, E.M., Missler-Behr, M., 2011. Virtual Embeddedness and Social Media as a Basis for the Relational Capital Management of new Ventures. *The Electronic Journal of Knowledge Management* 9, 188–203.
- Vedder, R.G., Vanecek, M.T., Guynes, C.S., Cappel, J.J., 1999. CEO and CIO perspectives on competitive intelligence. *Commun. ACM* 42, 108–116.
- Vesset, D., McDonough, B., 2009. Improving Organizational Performance Management Through Pervasive Business Intelligence. IDC.
- Von Hippel, E., 2001. Perspective: User toolkits for innovation. *Journal of product innovation management* 18, 247–257.
- Watson, H.J., Wixom, B.H., 2007. The current state of business intelligence. *Computer* 40, 96–99.
- Watson, H.J., Wixom, B.H., Hoffer, J.A., Anderson-Lehman, R., Reynolds, A.M., 2006. Real-Time Business Intelligence: Best Practices at Continental Airlines. *Information Systems Management* 23, 7–18.
- Wixom, B., Watson, H., 2010. The BI-Based Organization. *International Journal of Business Intelligence Research* 1, 13–28.
- 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*.
- 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.
- Zhao, S., Grasmuck, S., Martin, J., 2008. Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior* 24, 1816–1836.
- Zhou, Y., Davis, J., 2007. Discovering web communities in the blogspace, in: *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference On*. pp. 85–85.

**Table 1: Recent works utilizing social media data**

Titles and scholarship	Social Media	Application
"Inferring preference correlations from social networks", (Hogg 2010)	- Social network service (Essembly.com)	Identifying consumer preferences to help sellers design customized bundles of products
"Discovering target groups in social networking sites: An effective method for maximizing joint influential power", (K. Xu et al. 2012)	- Users opinions website supporting social networking functions (Epinions.com) - Micro-blogging platform (Twitter.com)	Analysing user opinions to discover the user groups with max joint influential power for marketing and enterprise reputation management
"LCI: a social channel analysis platform for live customer intelligence", (Castellanos et al. 2011)	- Micro-blogging platform (Twitter.com)	Conducting real-time and historical sentiment analysis for inferring human emotions to get instant customer insight on events like a new marketing or sales campaign
"Facebook as Agile CRM? A Business Intelligence Analysis of the Airline Ash Crisis", (Prestus & Bygstad 2010)	- Social network service (Facebook.com)	Utilizing social media to manage customer relationships effectively (as an agile CRM tool) and to supplement customer communication
"Predicting Stock Market Indicators Through Twitter 'I hope it is not as bad as I fear'", (X. Zhang et al. 2011)	- Micro-blogging platform (Twitter.com)	Predicting stock market indicators by measuring collective hope and fear
"Twitter mood predicts the stock market", (Bollen et al. 2011)	- Micro-blogging platform (Twitter.com)	Predicting stock market indicators through mood tracker tools
"Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web", (Sobkowicz et al. 2012)	- Online discussion platform - Blogging platforms (Technorati.com)	Exploring ways where online content from social media can be exploited to inform decision makers about constituent opinions, emerging trends, and on the feasibility and potential impacts of policy initiatives
"Vehicle defect discovery from social media", (Abrahams et al. 2012)	- Online discussion platform (honda-tech.com)	Facilitate decision support for the vehicle defect discovery and classification process
"Social media use by government: From the routine to the critical", (Kavanaugh et al. 2012)	- 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
"Harnessing social media platforms to measure customer-based hotel brand equity", (Callarisa et al. 2012)	- Travel related social network service (Tripadvisor.com)	Measuring brand equity using online customer reviews
"Social media competitive analysis and text mining: A case study in the pizza industry", (He et al. 2013)	- Social network service (Facebook.com) - Micro-blogging platform (Twitter.com)	Create actionable competitive intelligence using web mining for the restaurant industry

**Table 2: Facebook fan pages**

Brand	Fan base	Talking about this	Position across Greek fan pages	Joined Facebook
A	268,369	2,166	6	04-Feb-09
B	42,404	223	55	05-Jun-09
C	42,057	302	56	04-Feb-11

**Table 3: Cultural Artefacts of Facebook**

Cultural artefact of practice	General Description	Commonly used for	Is manifested through	Action possibilities for users	Affordances	Traces produced during its use
Page	A Facebook page – as advertised by Facebook – is a place for businesses, organizations and brands to share stories, content (e.g. videos, photos) and connect with people. Pages can be further customized to support additional content by adding applications	Dissemination of ideas; linking people around a point of interest (e.g. a brand)	Filling a page profile	Becoming a fan; view content; create a wall post (i.e. post updates); view milestones	Affords member list; hosts applications such as discussion board, reviews and customized applications	Members list, wall posts, comments, 'likes'
Wall	The wall is a place within the Facebook page where users can share information. Information is shared through wall posts. The wall displays all related with the page user's activity. The wall is a feature existing also in user profiles	Presenting and publishing content; public space of the page; newsfeed aggregator	The page or the user profile	View and create wall post	Aggregates posts	Wall posts, comments, 'likes'
Wall post	A post on the wall can host a short message that is shared among the pages fans. A wall post can be commented, "liked" or shared. A post can be text optionally combined with photos, videos and links or questions	Information container; media container; expressing opinion; seeking information	The wall	View; view creator; view and post comments; view and post 'like'; share	Allows discussion; can be "liked" or shared by a user	Wall posts, comments, 'likes', views

Cultural artefact of practice	General Description	Commonly used for	Is manifested through	Action possibilities for users	Affordances	Traces produced during its use
Comment	A short text message	Facilitates discussion; opinion expression; seeking information	A dialog placed under a wall post, a photo, a video, etc.	View; view creator; view and post like	Allows discussion; can be "liked" or shared by a user	Comments, 'likes'
'Like'	'Like' declares a feedback (probably positive). It can be performed in wall posts, comments, photographs, events, etc. In the past, like button served as a share mechanism	Sharing content; a positive indication	A button labelled with a thumbs up icon and the text 'like'	View; view creator	Allows a connection of the user to an artefact (e.g. wall post, comment, etc.). This action becomes visible in users profile	'Likes'
User Profile	A page dedicated to presenting an actual human being. Same as a fan pages, contains a wall where activities of the user are represented	Representing a user	Filling user profiles	View, add friend, send message, poke, etc.	Allowing self-disclosure; web presence	Friend list, wall posts, profile image, etc.
Tag	Tagging is a function of Facebook allowing people to be linked with other people or pages with something they post such as a photo, a link, etc. Initially, it was used to annotate people that appeared in a photo but later this functionality could be used with wall posts and comments.	Afford connection of users with specific artefacts	The character '@' before the name of a user on text, or a cross pointing to the position to be tagged in photos	View	Allows a connection of the user to an artefact	Tags



**Table 4: Excavation identity**

Date of data excavation	24 August 2012
Facebook users	42,917
Facebook activities	145,819
Oldest Facebook activity	5 February 2009 at 20:18 of type 'status update'
Most recent Facebook activity	23 August 2012 at 19:33 of type 'comment'
Type of Facebook activities captured through API (number of activities counted)	<ul style="list-style-type: none"><li>• Posts on wall (10,503)<ul style="list-style-type: none"><li>○ Status updates (9,238)</li><li>○ Photos (539)</li><li>○ Videos (353)</li><li>○ Links (361)</li><li>○ Questions (12)</li></ul></li><li>• Likes (88,030)<ul style="list-style-type: none"><li>○ On wall posts (64,023)</li><li>○ On comments (24,007)</li></ul></li><li>• Comments (42,954)</li><li>• Tags (4,332)<ul style="list-style-type: none"><li>○ On comments (4,125)</li><li>○ On photos (207)</li></ul></li></ul>

**Table 5: Response rates by operators**

	Brand 'A'		Brand 'B'		Brand 'C'	
	n	(%)	n	(%)	n	(%)
Wall posts made by operators (% of total page's wall posts)	225	2.95	387	24.62	324	24.70
Wall posts made by fans (% of total page's wall posts)	7,394	97.05	1,185	75.38	988	75.30
Total wall posts (% of total page's wall posts)	7,619	100	1,572	100	1,312	100
Wall post created by fans that had responses by the operator (% of wall posts made by fans)	3,243	43.86	636	53.67	587	59.41
Questions i.e. wall posts made by fans having question marks (% of wall posts made by fans)	4,379	59.22	696	58.73	566	57.29
Questions that had at least one operator response (% of total fan questions)	2,490	56.86	500	71.84	415	73.32
Number of different fans that created question	2,870	-	474	-	347	-

**Table 6: Response times**

	Brand 'A'		Brand 'B'		Brand 'C'	
	n	(%)	n	(%)	n	(%)
Less than one minute (% of total fan questions with an operator response)	15	0.60	1	0.20	0	0.00
Within five minutes (% of total fan questions with an operator response)	229	9.20	12	2.40	8	1.93
Within an hour (% of total fan questions with an operator response)	1025	41.16	171	34.20	92	22.17
Less than a day (% of total fan questions with an operator response)	2109	84.70	407	81.40	336	80.96
More than a day (% of total fan questions with an operator response)	381	15.30	93	18.60	79	19.04
Average response time in minutes	910	-	934	-	998	-
Median response time in minutes	150	-	273	-	642	-
Interquartile range	941		1,090		1,131	
Minimum response time in minutes	0.26	-	0.96	-	1.35	-
Maximum response time in minutes	21,238	-	18,676	-	10,867	-

**Table 7: Word frequencies**

	Brand 'A'		Brand 'B'		Brand 'C'	
	n	(%)	n	(%)	n	(%)
'Ευχαριστώ' (i.e. thank you) (% of total word count)	804	4.92	165	4.31	109	2.95
'Ωραία' (i.e. 'nice') (% of total word count)	28	0.17	13	0.34	4	0.11
'Πρόβλημα' (i.e. 'problem') (% of total word count)	65	0.40	8	0.21	10	0.27
'Ντροπή' (i.e. 'shame') (% of total word count)	2	0.01	4	0.10	1	0.03
'Κακό' (i.e. 'bad') (% of total word count)	11	0.07	4	0.10	0	0.00
Total words count	16,336	-	3,831	-	3,697	-

**Table 8: Fan questions, comments and 'likes'**

	Brand 'A'		Brand 'B'		Brand 'C'	
	n	(%)	n	(%)	n	(%)
Questions with a 'like' made by question creator on the operator response (% of total questions with operator responses)	351	14.10	100	20.00	74	17.83
Comments made by fans on questions(% of total comments on questions)	8,132	71.19	836	55.62	731	58.67
Comments made by operators on questions (% of total comments on questions)	3,291	28.81	667	44.38	515	41.33
Total comments on questions (% of total comments on questions)	11,423	100	1,503	100	1,246	100
Questions that had comments by the question creator (% of total questions)	2,231	50.95	325	46.70	238	42.05
Questions that had comments both by the question creator and the operator(% of total question)	1,594	36.40	293	42.10	203	35.87
Questions with gratitude expression by their creator and response by an operator (% of total question that had responses by both the creator and the operator)	764	47.93	163	55.63	104	51.23
Questions with gratitude expression by their creator after the first response by the operator (% of total question with gratitude expression by their creator and response by an operator)	692	90.58	159	97.55	94	90.38
Questions with gratitude expression by their creator before the first response by the operator (% of total question with gratitude expression by their creator and response by an operator)	72	9.42	4	2.45	10	9.62

**Table 9: Active fans and contributors**

	Brand 'A'		Brand 'B'		Brand 'C'	
	n	(%)	n	(%)	n	(%)
Active fans (% of total fans of the page)	27,587	10.28	9,267	21.85	8,697	20.68
Activity of active fans(% of total activity in all pages)	79,667	59.41	29,001	21.64	25,412	18.95
Active fans with just one activity (% of total active fans of the page)	16,211	58.76	4,094	44.18	5,531	63.60
Active fans with ten or more activities(% of total active fans of the page)	1,258	4.56	410	4.42	358	4.12
Contributors (% of total active fans of the page)	11,275	40.87	4,671	50.40	1,671	19.21
Activity of contributors (% of total page activities)	53,663	67.36	21,106	72.75	12,476	49.09
Contributors with just one activity (% of total active fans of the page)	3,631	32.20	684	14.64	439	26.27
Contributors with ten or more activities (% of total active fans of the page)	1,110	9.84	344	7.36	242	14.48
Median activities of contributors	2	-	3	-	3	-
Interquartile range	3	-	2	-	5	-

**Table 10: Cross-posting fans**

	Active fans (% of total active fans)	Contributors (% of total contributors)	Contributors with $\geq 10$ activities (% of total contributors)
Having activities in just one page	40,535 (94.46)	14,752 (89.49)	231 (1.40)
Having activities in two pages	2,120 (4.94)	1,502 (9.11)	433 (2.63)
Having activities in three pages	258 (0.60)	231 (1.40)	151 (0.92)

**Table 11: Translated content from area A**

Mark	Brand	Type of activity	English translation	Number of comments	Number of likes
A-1	'A'	Comment	ARE YOU GOING TO DO SOMETHING FOR SIGNAL [coverage]? TELL US JUST TO KNOW AND NOT HOPE FOR SOMETHING MORE. GIFTS AND OFFERS ARE EASY TO GIVE, FAST AND ALMOST FREE BY YOUR SPONSORS. WHAT ABOUT SIGNAL [coverage] AFTER ALL! ALL CITY OF THESALONIKI SPEAKS FROM BALCONIES	-	6
A-2	'A'	Like on comment	[...] Also about the signal that is mentioned by others, they are right. In many places in Greece, I have confirmed it. [Brand 'C'] has very good prices that's true, but doesn't give web mail so I will have to decide to lose one of my two numbers or to have two devices like john wayne and get twice radiation, or to buy device that accepts 2 SIM. [...]	-	2
C-1	'C'	Comment	Finally, [Brand 'C'] you are welcome to Facebook!!! Welcome!	-	5
A-3	'A'	Wall post	WE ARE LOOKING FOR SIGNAL [coverage] WITH BINOCULARS! LEAVE SILLY EXPERIMENTS WITH CHINESE HARDWARE [referring to signal enhancers] AND DO SOMETHING [...]	2	4
C-2	'C'	Comment	* the offer with free double time on the first recharge of each month until 30/06 is valid for [Brand 'C'] that received a related SMS	-	2
C-3	'C'	Wall post	Join the CLUB! Come to [Brand 'C'] EASY, FAST and of course COMPLETELY FREE! How? Call free [...]	0	2
A-4	'A'	Wall post	[Brand 'B'] continues to give 1500' + 1500 SMS until 30/06 for those activated before 02/02. But we here in [Brand 'A'] that have activated since last year we don't have it anymore. Thus, we are dupes and they are smart with [Brand 'B'] to have until 30/06?	0	5
A-5	'A'	Wall post	9 OUT OF 10 THAT LEAVE [Brand 'A'] IS BECAUSE THEY DO NOT HAVE ANY SIGNAL [coverage], AND FEEL VERY NICE! YOU?	0	2



**Table 12: Translated content from area C**

Mark	Brand	Type of activity	English translation	Number of comments	Number of likes
A-1	'A'	Comment	Everyone is gathered here ? Because I'm currently in [Brand 'C'] and I'm thinking of 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 1: Operators responses on users' wall posts





Figure 3: Active users on Facebook pages (Top graph represents the entire data set while the bottom a zoom area)

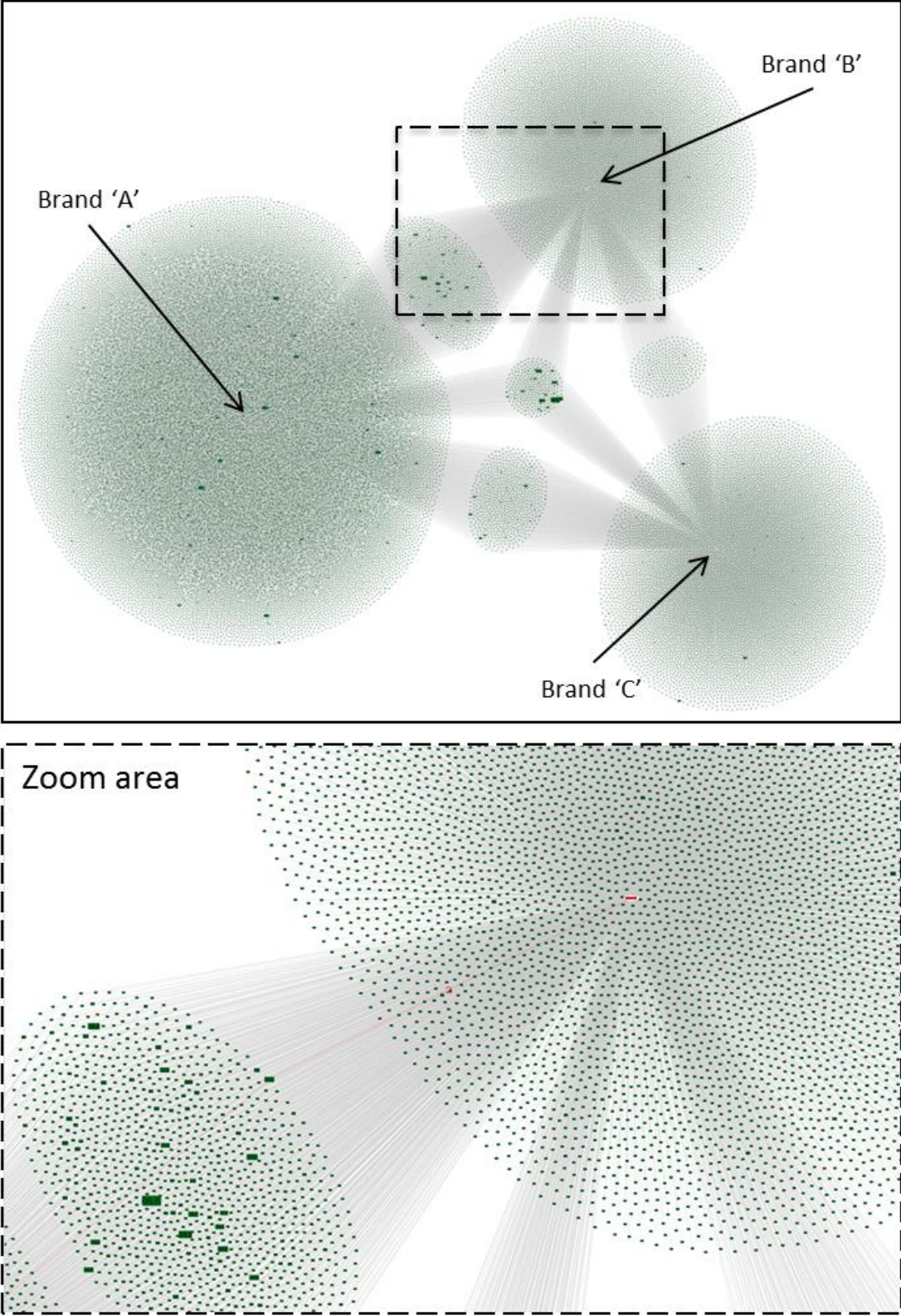


Figure 4: Timeline of users' activities

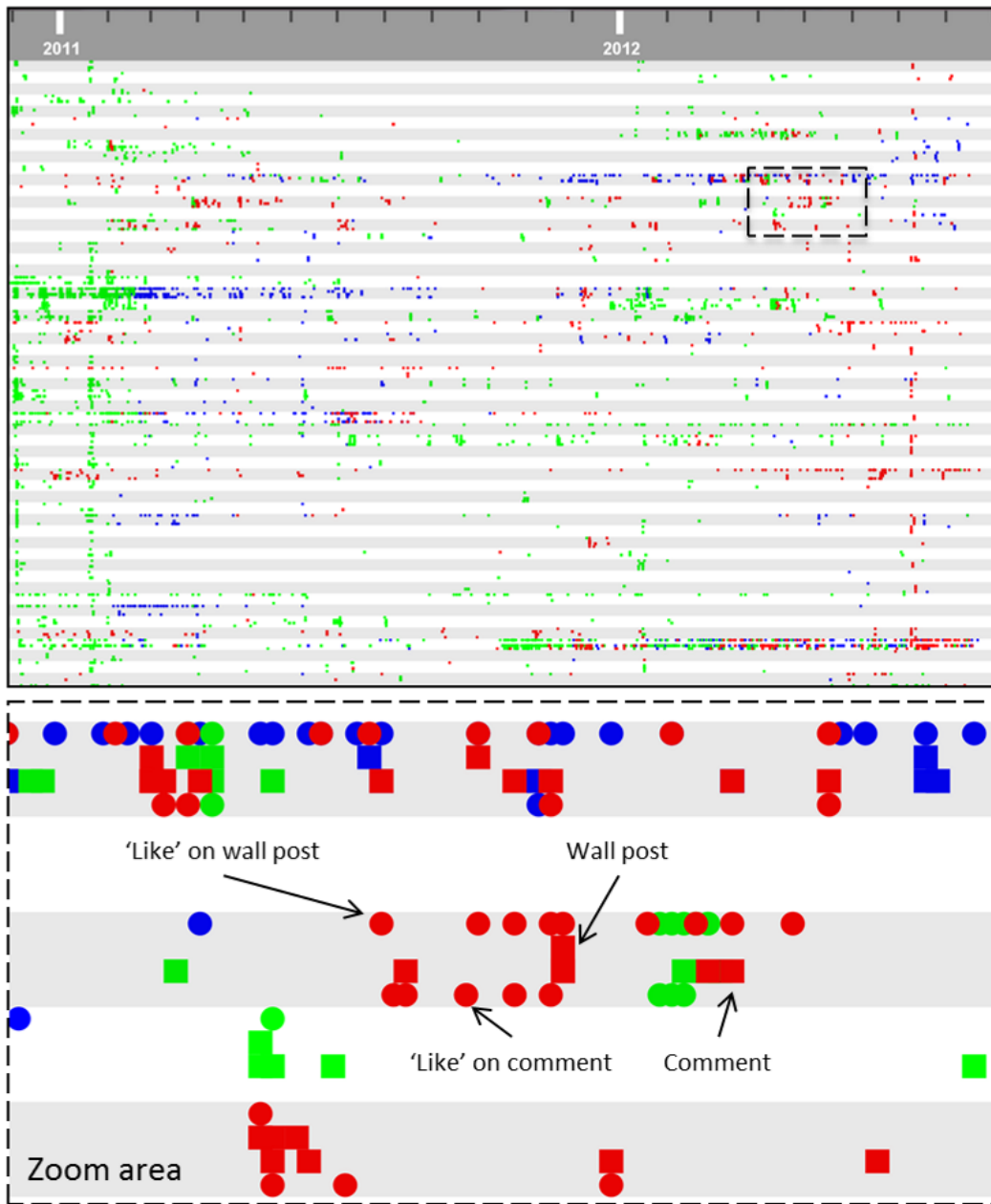
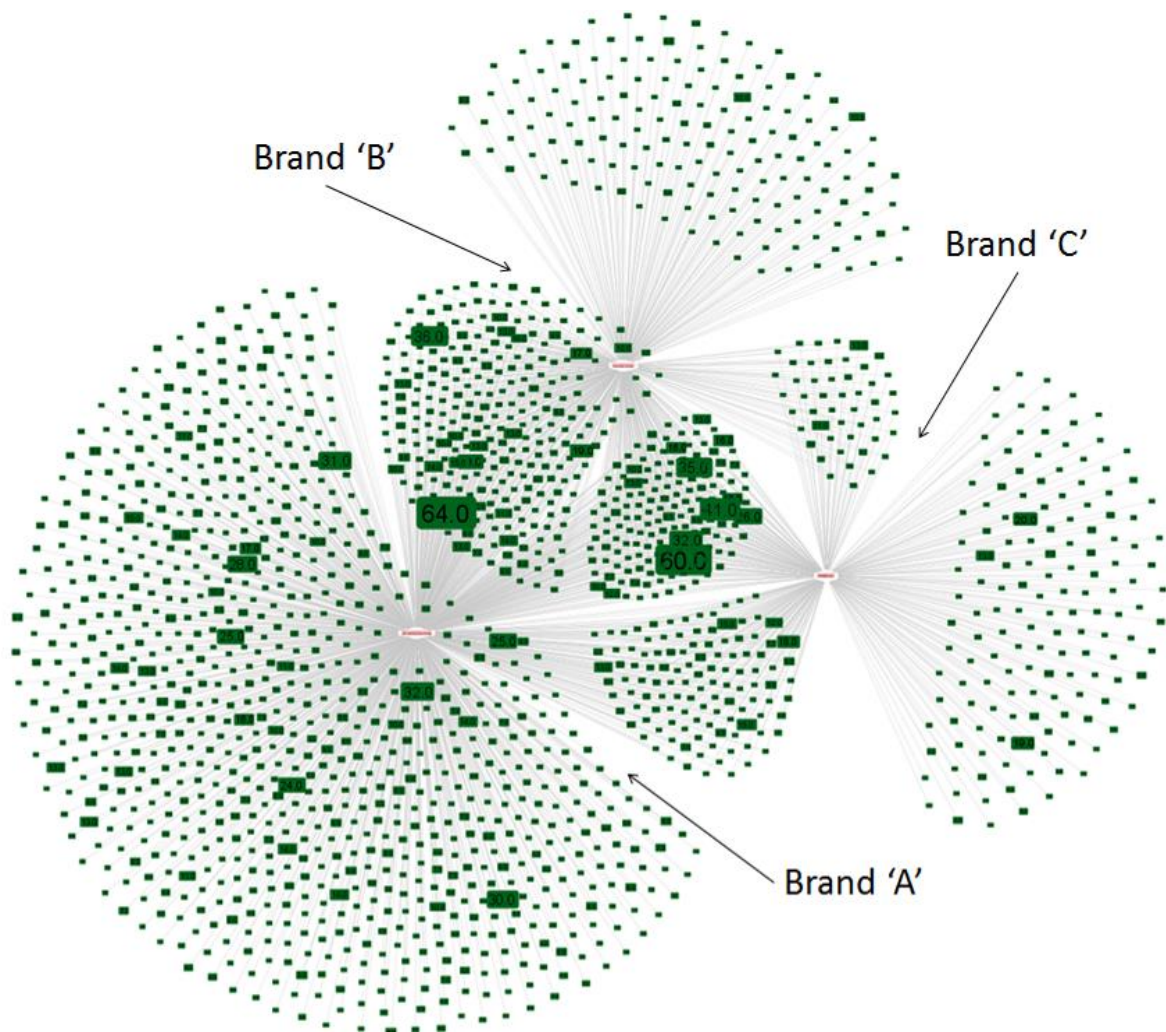


Figure 5: Contributors that had ten or more activities of any type



**Figure 6: Timeline of contributors that had ten or more activities of any type on two or more pages**



Figure 7: Zoom on area A of Figure 6

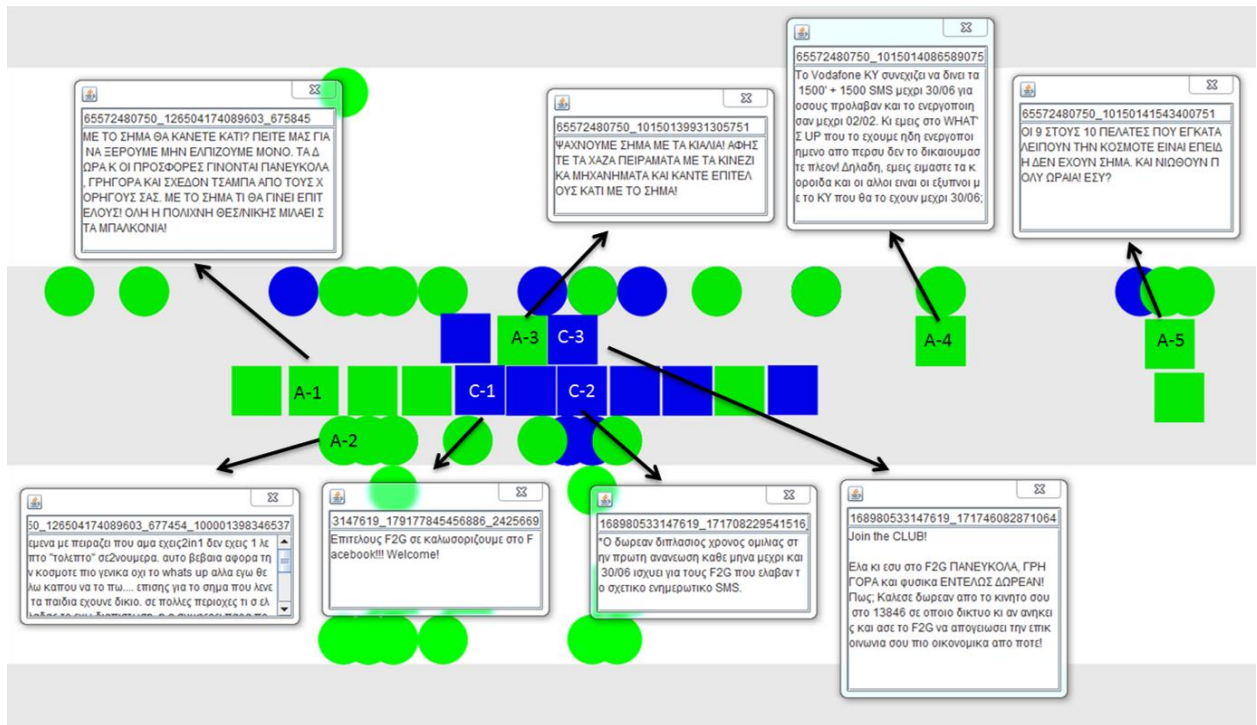




Figure 8: Zoom on area C of Figure 6

