Social media monitoring and understanding: an integrated mixed methods approach for the analysis of social media

Guendalina Graffigna* and Giuseppe Riva

Department of Psychology,
Università Cattolica del Sacro Cuore,
Largo Gemelli, 1, 20123 Milan, Italy
E-mail: guendalina.graffigna@unicatt.it
E-mail: giuseppe.riva@unicatt.it
*Corresponding author

Abstract: It is increasingly apparent that social media is a unique and important resource to explore spontaneous relations and discourses among individuals. Data analysis in social media, however, is difficult because approaches have not been streamlined. In this paper, we move from the discussion of present perspectives in the analysis of social media to then discuss the methodological strategies available and their limits. We finally discuss the value of choosing a mixed methods approach to the analysis of social media, exemplifying its heuristic contribution by a real research case: the one-year monitoring of Italian women spontaneous social media discourses on the hypoactive sexual desire disorder (HSDD).

Keywords: social media interaction; online spontaneous discourses; mixed methods; qualitative research; quantitative research; buzz metrics; hypoactive sexual desire disorder; HSDD; social media analysis; social media listening.

Reference to this paper should be made as follows: Graffigna, G. and Riva, G. (xxxx) ‘Social media monitoring and understanding: an integrated mixed methods approach for the analysis of social media’, Int. J. Web-Based Communities, Vol. X, No. Y, pp.xxx–xxx.

Biographical notes: Guendalina Graffigna received her PhD in Social Psychology at Catholic University of the Sacred Hearth of Milan (Italy) where she actually is Assistant Professor. At the same University she teaches ‘Qualitative Methodology’ and she is Associate Director for the II level Masters degrees ‘Mobile marketing and social research methods’ and member of the steering committee of the PhD programme in Psychology at the Catholic University of Milan, Italy. She is also Editorial Manager for the journal Micro & Macro Marketing and Associate Director of the Study and Training Centre of ASSIRM (Italian Association of Marketing Research Instistutes).

Giuseppe Riva is Professor (tenure position) of General Psychology and Communication Psychology, and member of the steering committee of the PhD programme in Psychology at the Catholic University of Milan, Italy. His research work led to the development of a new scientific discipline – Cyberpsychology – integrating innovative research ranging from usability and ergonomics, clinical psychology and cognitive science, to mobile devices and simulation apparatus. He is currently the President of the International Association of CyberPsychology, Training, and Rehabilitation and member of the steering committee of the Society for Computers in Psychology.
1 Introduction

The emergence of social media has changed the way people interact with computers and mobile devices (Qualman, 2010), as underlined by Turkle (2011, p.xlii): “online connections were first conceived as a substitute for face-to-face contact… We discovered the network – the world of connectivity to be uniquely suited to the overworked and overscheduled life it makes possible. And now we look to the network to defend us against loneliness even as we use it control the intensity of our connections”. In fact, social media are virtual spaces where individuals experience an optimal experience (Mauri et al., 2011) and can create, meet and discuss about their experiences, values and opinions (Espinoza and Juvonen, 2011; Gonzales and Hancock, 2011). On one side, these virtual spaces can be the fertile ground for interpersonal sharing of knowledge, for the formation of friendships and even close relationships because of the shared interests and values of their users (Howard, 2002; Smith, 2010; Bargh and McKenna, 2004). On the other side, social media allow users to negotiate presentations of self in their profiles. The extent to which profiles are authentic or playful varies across sites and serves as important identity signal (Boyd and Ellison, 2007).

For this variability and heterogeneity, the study of social media is a complicated task. In order to contribute to the methodological debate about social media analysis, in this paper we will describe a mixed-methods research approach aimed at interlacing different qualitative and quantitative analytical strategies in order to provide a more grounded and complete understanding of dialogical and interactive dynamics of social media. To motivate our approach we will refer to a real case of research – the one-year monitoring of Italian women spontaneous social media discourses on the hypoactive sexual desire disorder (HSDD). The research case reported is not aimed to be a sound evaluation of the methodological approach proposed, but rather an example to illustrate the heuristic value and the potential limits of our approach.

1.1 Social media: the arena interpersonal knowledge and practice sharing

Spontaneous interactions which take place in social media are complex and multifactor processes capable of influencing in an effective but discrete way subjects belonging to a social network (Turkle, 2011). Internet users, indeed, involved in multiple relationships and characterised by a fluid identity due to their concurrent belonging to different online communities, are naturally involved in multiple conversations and thus, potentially influencers over their interlocutors: the articulation of the web in several places of social aggregation (i.e., chat rooms, social networks) or in communities of practices (i.e., forums, blogs, online communities) is the fertile ground for interpersonal sharing of knowledge, information, gossip, evaluations, and for comparing experiences and orienting interpersonal decision making processes (Ellonen et al., 2010).

These features make internet the most powerful and discreet medium for social discourses and word of mouth diffusion (See-To and Ho, 2014). Differently from face-to-face exchanges, in the web people, thanks to anonymity, feel less inhibited; this makes internet users more willing to share personal information, to be honest and courteous with their interlocutors (Roed, 2003; Bickart and Schindler, 2001). Furthermore, online discourses have greater propagation speed, are fully available and powerful (direction one to many), without omitting the absence of social pressure typical of face-to-face interactions (Phelps et al., 2004; DiMaggio et al., 2001; Howard et al.,
Social media monitoring and understanding

2000; Kraut et al., 2002). Finally, potential influence of social media discourses is due to people shared perception that interpersonal communication is more ‘trustable’ than formal media communication regarding the same topic (Bickart and Schindler, 2001; Brooks, 1957), although the comments posted in an official website might be considered more persuasive and trustable than those appeared on the same topics in spontaneous online portals.

In this regard, social media have also totally changed the health communication world. Social media exchanges about heath today help patients to become really protagonists of their care management thanks to the possibility of sharing and constructing knowledge about their illness, their care (Nambisan and Nambisan, 2009) and their identity (Miles, 2009). Social media thus represent a useful knowledge and support resource for individuals in the different sphere of their human existence and they contribute to enhance individuals’ coping in their daily life (Gaspar et al., 2014).

1.2 The cognitive and narrative nature of social media

At the psychological level, social media can be experienced by individuals in many different ways, so each interaction is a unique situation (Turner and Shah, 2011). Users can explore blogs and social networks, present themselves and communicate through them in many different ways despite accessing the resources through the same technologies. Furthermore, social media are a social and cognitive space (Riva et al., 2006) through which relationships are created and cognition is elaborated. For example, as recently demonstrated by Chou and Edge (2012) the use of social networks influences people’s perceptions of others’ lives via the availability heuristic, where users tend to base judgment on examples easily recalled, and the correspondence bias, where they tend to attribute the positive content presented on the social network, rather than situational factors, to others’ personalities, especially those of strangers.

Also, Elphinston and Noller (2011) show the excessive use of social media influences romantic relationships, resulting in problems such as jealousy and dissatisfaction. Trognon (1992, p.117) clearly explained that far from being circumscribed by experimental laboratory settings, cognitive activities are routine daily activities. In other words, it is in everyday life, and in conversational interaction especially, that we put our cognitive skills to practical use. Finally, the social media experience is always situated in a specific context, requiring detailed analysis of the social context to understand the interaction (Riva and Galimberti, 2001; Riva, 2001). These aspects clearly underline the situated, cognitive and narrative nature of the social media experience. Further, we can define it as a process by which a group of social actors in a given situation negotiate the meaning of the various situations that arise between them within an electronic environment (Riva, 2001). But how can we study the experience of social media?

1.3 Experiences and pitfalls in the study of social media

Due to these psycho-social functions, the analysis of social media is particularly useful because it allows (Gaspar et al., 2014)

1 to potentially access a wide sample of individuals and experiences in the real time expression of their life
G. Graffigna and G. Riva

2 to get insight about different moments and phases of the exchanges spontaneously occurring online and thus to capture the dynamics of such exchanges and interpersonal interactions

3 to cast light on social influences on individuals’ attitudes and behaviours in a given social context.

However, in existing approaches, there is not a clear path for analysing social media interactions (see Figure 1).

**Figure 1** The analysis of strengths and limitations of traditional approaches to the analysis of social media (see online version for colours)

<table>
<thead>
<tr>
<th><strong>Strengths</strong></th>
<th><strong>Weaknesses</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast and cheap web crawling</td>
<td>Content analysis often detached from its linguistic/discursive context → loss of ecological understanding of the dialogical context</td>
</tr>
<tr>
<td>Descriptive statistics of web productivity and readership (how many users do write and read about x?)</td>
<td>Fragmentation and loss of the 'whole picture' → scarce systematisation</td>
</tr>
<tr>
<td>Possibility to measure volume and trends of online discourses</td>
<td>Crawling at the 'margins' of the web</td>
</tr>
<tr>
<td>Balancing the measurements and standardisation of quantitative tools</td>
<td>Is 'many' = one or more actors?</td>
</tr>
<tr>
<td>... with the flexibility and intelligence of qualitative research</td>
<td>Loss of real pragmatic relevance of online discourse analysis</td>
</tr>
<tr>
<td>... in an integrated and systematic approach</td>
<td>IMIME approach</td>
</tr>
</tbody>
</table>

Marketing researchers usually utilise quantitatively procedures, such as content mapping and user metrics (analytics) (Sterne, 2010). Buzz metrics tools (see for instance, Sponder, 2012) are able to dimension the volume and frequency of 'post'/spontaneous discourses related to specific topics/key words, track their trends in a given period of time, detect the web-sources where these track come from. These tools have the advantage of managing a great amount of textual data in a short period of time, to provide descriptive statistics of detected topics and of their characteristics (frequency, source, actors involved). However, these analyses provide less emphasis on interpretation and do not explain how users operate in culturally grounded contexts. Furthermore, automated and computational approaches to social media analysis and sentiment detection may be superficial and incur in errors (Balahur et al., 2014). Thus, also supporters of quantitative and software based approaches to the analysis of social media, underline the limits of such analysis and its failure in the detection of ecological, really situates and informative nuances social media exchanges.

Another very well established approach to quantitatively analyse social media discourses is the social network analysis strategy (Haythornthwaite, 1996) that focuses, measures and describes paths of information exchange and relationship among web-users in social media, by allowing the description of groups dynamics (Fu et al., 2012). This
approach is particularly useful to track directions and volumes of online discourses, but with a loss of contextual cues. As Gonzalez-Bailon et al. (2010, p.2) affirmed: “not all networks allow the same flows of information, and users might form different networks even when using the same internet technology”. In our opinion, thus, social network analysis is not enough when trying to catch full insight on user generated contents in social media.

Social researchers usually employ qualitative methods to guarantee a more depth and grounded understanding of social media discourses. For instance, nethnography (Kozinets, 2010) is a branch of ethnography (Spradley, 1979) that analyses the spontaneous conversations and behaviours of internet users to provide useful insights in the area. Other scholars prefer to base their social media investigation on conversational and discourse analysis (see for instance, Herring, 2001) in order to describe interactive and rhetorical features of spontaneous discourses that take place online and their interlacement with individuals’ attitudes and behaviours (Talbot, 2007). Moreover, other researchers claim that content analysis should be the starting point of any social media analysis. For instance in the introduction to the second edition of their classical Analyzing Media Messages book, Riffe (2009, p.8) and colleagues underlined the importance of content analysis: “research during the past 35 years suggests that the effects of mass media are contingent on a variety of factors and conditions. Despite increasing interest in what people do with media messages and how of if they learn from them, content analysis remained an important means of categorizing all forms of contents.”

However, although qualitative approaches are relevant, offering ecological insights on the form and content of spontaneous discourses that take place online (Graffigna and Olson, 2009), they are often considered impressionistic and subjective, and do not provide grounds for generalising across cases (Howard, 2002). As a matter of fact, qualitative analysis carried out so far on the Web 2.0 appears still poorly systematic and not really able to support interventions and strategic decisions. Furthermore qualitative approaches to social media analysis have the limit to only focus on small portions of social media discourses: often one or two web-communities/social networks a priori selected by the researcher. Thus, in our perspective, qualitative research approaches may be an important contribution only in a more advanced phase of social media analysis, after that a preliminary exploration and measurement of the whole Web 2.0 discursive arena has been performed and when cues for selecting the communities/social networks exchanges more relevant for further analysis have been detected.

On the basis of these consideration and according to other scholars (see for instance, Rohm et al., 2013), from our perspective the complexity and fluidity of social media require an integrated and mixed method approach of study. In particular, we consider it important to base social media analysis on the synergy and systematisation of different quantitative and qualitative analytic strategies, corroborated by the use of dedicated software. In the following paragraphs, we will discuss our perspective by introducing a real case of research.
2 The integrated mixed methods approach: methodological characteristics and heuristic value

To overcome the drawbacks of current methods, we suggest the value of opting for a four-step integrated mixed methods (IMiMe) approach. This approach integrates different qualitative analytic traditions (i.e., interpretive analysis, discourse analysis, thematic analysis) with quantitative automated analyses performed by software for the treatment of textual data (Radian6, Atlas.ti and T-lab.) in order to answer four main knowledge objectives:

1. How much do social media users talk about the topic?
2. Where do they talk about it?
3. How do they talk about it?
4. And what do they say?

In this paper, we shall describe the methodological features and goals of each analytical phase by then exemplifying the results achievable drawing on a real research case. In particular the exemplificative research case reported in this paper is a one-year study monitoring of spontaneous social media discourses of Italian women on HSDD: a sexual dysfunction characterised by the lack or absence of sexual fantasies and desire for sexual activity. We deliberately chose to focus on this exemplificative case due to the sensitiveness and methodological complexity of the topic under investigation: HSDD online discourses are not easily circumscribable in a priori chosen portions of the web; furthermore they imply a depth and ecological understanding of psycho-social and interactive characteristics of the exchange in which these discourses are generated. In this paper we shall report some exemplificative results in order to better clarify the methodological features, the potentiality and the limits of the IMiMe approach. This study is an exemplifying research case but doesn’t aim to be a sound evaluation of the IMiMe approach.

2.1 Step 1: assessing productivity and readership

Firstly, the IMiMe approach requires the analysis of two quantitative indicators related to the volumes and the trends of social media discourses on the topic of interest. In particular an index of ‘productivity’ (‘How much do people talk about the topic?’) and of ‘readership’ (‘How much do people look for the topic?’) on the selected topic were analysed. The index of ‘productivity’ may be calculated with a web crawler [e.g., Radian6 – http://www.radian6.com, but other similar crawler can be used; for its use in social media analysis see, for example Laine and Frühwirth (2010)] on the basis of specific combination of key words. The choice of key words and of their logic combination with Boolean operators is a sensitive part of the analysis and should be done based on the deep evaluation of the research objectives and of the thematic domain of the monitoring. For instance, in our exemplificative study, we used the following key words: HSDD; women sexual desire decrease; pink Viagra; women sexual desire.

This crawling should be restricted to the language and country domain relevant for the monitoring. For instance in our case the search was restricted to only Italian web
Social media monitoring and understanding

This first step of analysis helps in tracking the trends of discourses on the topic in a given period of time (i.e., day by day, month by month, over the year) and in identifying interesting peaks in the discussion about specific topics. Furthermore, the analysis allows to compare trends of discourse productivity among different topics/key words.

For example, in our case we found that women tended to speak more about a potential solution to the problem (Viagra Rosa is the colloquial way in which speakers referred to a potential drug devoted to treat this problem) rather than the problem itself (HSDD), and that there was a peak of ‘discourse productivity’ in the month of June.

The level of ‘productivity’ is corroborated by a second quantitative index to track trends of ‘readership’ of user generated contents about the topic under investigation. It is a shared knowledge that the influencing power of social media is not only related to those who ‘leave their tracks’ on the web but especially to people who connect to the web to search for information and answers to their doubts, even without leaving any comment or message. This index may be calculated in an easy way on the basis of Google Insights (http://www.google.com/insights/search/): an online tool that allows the comparison of search volume patterns across specific regions, categories and time frames.

For instance in our research case, we furthered the retrieved peak of productivity related to the month of June, by assessing statistics of web search on the same topic in that period using Google Insights. We found a peak in the search volume for this topic in the first days of June, about one week before the peak of discourse productivity in social media. To better understand these findings, we assessed the topic coverage on TV and radio, finding that at the end of May these media commented the potential risk of a new drug for the treatment of HSDD, while no mention of the drug or disorder was found in previous months. These results suggest an interesting dynamic: news casting on traditional media in the short term, one day to one week, increases audience’s interest and web searches, leading to an increase in social media discourses on the topic in the medium term, one week to one month.

2.2 Step 2: topography of social media sources

This second step of the IMiMe analysis is inspired by the principles of conversational analysis applied to internet mediated communication (Galimberti and Cilento Ibarra, 2009; Galiberti et al., 2001). This qualitative ‘topography’ of social media sources helps in detecting the conversational context in which online discourses appeared. This analysis gives important additional information for the interpretation of spontaneous discourses’ contents. For instance, a strong positive or negative affirmation might change its value/strength depending on the type of context where it is published (cfr. Figure 2): it is a matter of evidence that a strong negative comment posted in an ‘entertaining context’ could have an ironic enunciative function, that lowers its negative strength; on the contrary the same message published in a serious context, frequented by users to receive answers to their doubts (‘pragmatic tools’), would have a stronger negative influence on its readers.
Thanks to the systematic ‘crawling’ conducted in the previous analysis and to additional manual searches, a panel of social media sources in which online discourse about the topic of interest takes place may be constructed. The web sources included in the panel may be profiled according to their dialogical characteristics and to the ‘fruition needs’ they seemed to answer. On the basis of this analysis in particular, a conceptual map (‘topography’) can be developed (see Figure 2):

- The horizontal axis of the map describes the main need that drives internet use: this axis opposes website mainly searched to receive advices and solutions to one’s own problem (positive pole) to website mainly frequented for entertainment needs or curiosity (negative pole).

- The vertical axis describes the attitude of internet users towards the contents fruited in the internet: it articulates website which mainly sustain a unidirectional fruition of information (positive pole), to website primarily devoted to the sharing and co-production of contents (negative pole).

This topographic map identifies four web ‘territories’:

- In the right-high quadrant are located the ‘expert websites’. This is the place of ‘scientific discourses’, where information with a high scientific level are reported, often as direct declarations of experts (i.e., journalists, physicians, politicians). Here the interaction among web-users is low and mainly related to Q&A with experts.

- In the left-high quadrant are located the ‘generalist websites’. Those are websites offering general information, not deeply discussed and reported in a ‘lightly’ conversational style. Also in this case the interaction/discussion among web-users is low and people tend to visit these websites to only catch information.
• In the right-low quadrant are located the ‘pragmatic websites’. Those are the websites searched by individuals in the aim of retrieving pragmatic and concrete advices for solving their problems. Here the interaction among participants is high and people engage in the sharing of concrete practices and suggestions for the daily life management.

• Finally, in the left-low quadrant are located the ‘entertaining websites’. These websites are the ‘chat place’ where web-users search for entertainment occasions and gossips to share. The level of discourses is light, often ironic.

Furthermore, this topography of social media sources maybe related to quantitative indexes measuring the web relevance of each portion of the map, and thus of each detected website. These indexes provide insights about the probability that detected online discourses may be viewed by internet users in a common web search. Among the various web-relevance indexes that may be used, the simpler ones are: the value of backlink, that measure the ‘popularity’ of the different social media sources calculated on the basis of the number of links in ‘entrance’ to the website and the value of page rank, that indicates the position of a certain website in the rank of Google searches [see for a detailed explanation, Cho et al. (1998)].

To sum up, this step of the analysis allows a better understanding of contextual features of the social media discourses selected, and thus allows to better ‘weight’ the rhetoric and content features of each retrieved message. This analysis is innovative because it allows the consideration of the whole exchange in which a message indexed with the given key words appears, and thus to get insights on the relational, dialogical and thematic dynamics that influenced and produced it.

In our research, this analysis showed a prevalence of ‘pragmatic’ sources (e.g., that offer practical advices to solve the problem), which produced the 59% of the whole messages retrieved in the monitoring. This evidence may indicate that women were interested not only in advice from leading resources, but also in the practical suggestions from others who have had the same experiences. A slightly different trend was evident in the month of June, where there was an increase of productivity in the ‘expert’ site cluster. ‘Pragmatic’ sources resulted also those with the highest web relevance, in term of a high average number of backlinks (in average the websites classified in this cluster counted 15.491.331.81 backlinks and an average page rank of 5/10).

2.3 Step 3: discourse analysis

This analysis is aimed at systematically describing the argumentative characteristics of online discourses. This discursive analysis, inspired by the principle of computer mediated discourse analysis (Herring, 2001) may be performed using a specific software, Atlas.ti (http://www.atlasti.com/index.html) that helps the researcher to uncover and systematically analyse complex phenomena hidden in text and multimedia data (Kelle, 1995). The program provides tools that let the user locate, code, and annotate findings in primary data material, to weigh and evaluate their importance, and to visualise complex relations between them.

This analysis requires the ‘a priori’ definition of discursive indicators/variables, relevant according to the research objectives. In this analytical step, messages detected in the first crawling phase are carefully read and analysed according to their rhetoric and linguistic features, in order to describe their illocutorian (Trognon, 1992) function in the
social media discourse and to interpret the implicit communicative aims of their sender. Discursive indicators adopted in this analytical phase may be related to the ‘genre’ of the writing (i.e., comedy, drama, ironic essay), to the roles assumed by senders/speakers (i.e., experts, lay) and to the representational and emotive tone of the discourse (i.e., problematising, descriptive, minimising).

For instance, in our research case we focused this analysis on the four discursive indicators:

1. the connotation of the speakers involved in the discourse [‘experts’ (i.e., doctors, psychologists, journalists) vs. ‘lay’ (i.e., consumers, patients)]
2. the discursive representation of the problem (i.e., ‘psychological/emotional nature’ vs. ‘organic/pathological nature’)
3. the discursive representation of the problem severity [i.e., ‘low’ (transitory problem) vs. ‘high’ (problem that needs a specific treatment)].

In particular, in our case, this discursive analysis showed that spontaneous online discourse about HSDD involved mainly lay actors, such as consumers, patients, and caregivers, rather than experts, such as doctors, journalists, opinion leaders. Furthermore, the analysis showed that in online spontaneous discourse, HSDD was mainly described as an emotional or psychological problem, transitory in nature, so it was perceived as less relevant. Finally, the problem seemed to be individualistic, evaluated as a problem of both partners. Such representations of HSDD were consistent throughout the year in social media discourses.

2.4 Step 4: semantic mapping

A fundamental step in our IMiMe approach is the content analysis of messages retrieved in the first step of crawling [please see for instance, Krippendorff (2012)] for further details on how to perfume a qualitative content analysis). In the IMiMe framework, we suggest to perform this analytical step with the support of the dedicated software for thematic analysis: T-LAB (http://www.tlab.it/en/presentation.php). T-LAB uses text-driven automatic analysis approach, based on statistical algorithms, which allows meaningful patterns of words and themes to emerge [see for exemplificative application of this software-based analysis: Gambetti and Graffigna (2010) and Gilardi and Lozza (2009)]. In particular, the software allows several strategies of content analysis.

For the IMiMe purposes, the following analysis are the most useful: thematic analysis of elementary contexts; specificities analysis; correspondences analysis.

1. **Thematic analysis of elementary contexts**: such an analysis, based on cluster and correspondence analysis, is aimed at obtaining and exploring a representation of corpus contents (i.e., the posts retrieved related to the topic of interest) through few and significant thematic clusters, each of which consists of a set of elementary contexts (i.e., sentences and paragraphs) characterised by the same patterns of key words and is described through the lexical units (i.e., words and lemmas) and the variables most characteristic of the context units of which it is composed. This analysis is particularly useful to obtain a synthetic representation of main themes covered in the retrieved online discourses.
Specificities analysis: this analysis indicates the most typical lemmas (over-used lemmas) and those which are typically absent (under-used lemmas) in a text subset. Thus Specificities Analysis underlines semantic features that differentiate a text subset from the rest of the corpus. CHI-square test defines analysis outcomes significance. This analysis is useful, for instance, to compare thematic characteristics of spontaneous discourses detected in different websites, or in different months.

Correspondences analysis: this analysis describes the relations between semantic units (lemmas) and the variable that divides the corpus (i.e., the different websites, the different period of online discourses monitoring). This analysis outcomes consist in a matrix reporting lemmas in row and variable levels in columns. Correspondences analysis casts light on relations between the data gathering techniques and their specific discussion contents. The analysis also detects factors that explain variance among variables. The meaning of these factors is interpreted on the basis of the threshold values of their constitutive elements (namely the lemmas and the four level of the considered variable).

Figure 3  The semantic mapping of web sources retrieved in the HSDD analysis

Due to limited space, we report only the results from the correspondence analysis performed on our exemplificative research case. The thematic variance of this discursive corpus may be explained by a semantic map (see Figure 3), where:
12 G. Graffigna and G. Riva

a the horizontal axis opposes contents related to ‘intervention’ (i.e., words such as: solving the problem, drug, psychological support, partner) to contents more linked to diagnosis of the problem (i.e., words such as problem, disease, mind, psycho, depression, hormone).

b the vertical axis opposes contents which support the use of a specific drug to treat the problem (i.e., words such as: research, studies, efficacy, solving the problem, drug) with contents more focused on the problems of pharmacological treatment of HSDD (i.e., words such as: low efficacy, approval, side effects, to impede).

The software also permits projection of the different portions of the corpus on this semantic map to better visualise and interpret their thematic focus and to describe their differences. Portions of texts were detected by variables such as, in this research case, the months in which the discourses appeared and their sources/clusters of web sources.

The semantic map showed a prevalence of discourses focused on HSDD diagnosis in the first part of the year (January, February, March and August, and in the ‘pragmatic tools’), but the only discourses related to the new drug treatment for HSDD were detected in November, June and July. Particularly in June, discourses mainly focused on the discussion of side effects of the new drug.

3 Discussions

As we have seen in the literature review, studying social media is quite complicated. The social media can be experienced in many different ways. Also, they are a social and cognitive space. Finally, their experience is always situated in a specific context, even when the user is alone. These points clearly underline the situated nature of the social media experience, defined as a process by which a group of social actors in a given situation negotiate the meaning of the various situations arising between them in an electronic environment. Actual approaches to the analysis of social media appear to us partial, only able to capture some of their linguistic or interactional characteristic. In particular the traditional methodological divide between quantitative and qualitative approaches to the analysis of social media should be overcome from our perspective, in order to guarantee a more sound, ecological and insightful approach to the interpretation of social media discourses.

In this frame, we claim for methodological approaches to social media able to conjointly answer to four main and complementary objectives:

1 How much do social media users talk about the topic?

2 Where do they talk about it?

3 How do they talk about it?

4 What do they say?

To answer these questions in this paper, we proposed a mixed methods research framework, the four-step IMiMe approach, allowing parallel and interconnected analytical strategies to be set up and run in the same domain. The basis of this approach is the integration of the results of each component within analysis. Further, it takes advantage of both qualitative analytic traditions like interpretive analysis, discourse
analysis, and thematic analysis and also quantitative automated analyses performed by software dedicated to the treatment of textual data (Radian6, Atlas.ti and T-lab).

The IMiMe approach is intended to be a research framework, to sensitise and guide researchers to a wider and mixed methods lens to analyse social media. From our perspective the integration of four different analyses provides a more complete understanding of social media dynamics on specific social topic. Furthermore, this integrate and mixed methods approach to the analysis of social media unveils the multimedia convergence of people’s interactions, offering interesting insight on the spread of news and online word-of-mouth dynamics. Since social media discourses are chaotic and complex in their nature, a multi-methods approach can help researchers better interpret these phenomena by offering several perspectives to deepen the analysis. Analysis of online social discourses not only means measuring their volume and trends over time, but also to deeply analyse dialogical dynamics involved in the discourses in order interpret the potential influence on internet users attitudes, cognition and behaviour. On the other hand, to only qualitatively describe a priori determined portions of social media discourse can be partial and misleading, not systematising the whole picture of social media interactions happening on the web on a given topic of interest. From our perspective, the complexity and fluidity of social media discourses today require to overcoming the traditional divide between quantitative and qualitative analytical sensitiveness, in the aim of better interlacing them for achieving wider and more complete analysis.

4 Conclusions

In conclusion, although needing further experience and verification, in this paper we claimed that a mixed methods approach to social media analysis provides a balance between the systematisation of quantitative measurement with the ecological and situated nature of qualitative research strategies. In a mixed and multi methods approach the researcher may use the data collected on one phase of the analysis to tune and define the characteristics in the other levels. IMiMe thus seems to us particularly useful to go deep to understand how users operate in culturally-grounded contexts. Specifically, from our perspective IMiMe can address three main priorities:

- **systematisation**: online discourses are by definition fluid and mutable, so studying them requires research approaches capable of balancing a sensitive interpretation with more quantitative analytical strategies able to measure the occurrences under investigation
- **ecology**: studying online discourses serves to detect real conversations that take place in the social media, necessitating ‘listening strategies’ that not only to map these conversations but also catch their discursive and argumentative characteristics
- **creativity**: finally, the analysis of online discourse not only should be aimed at producing information, but most importantly at generating new insights and creative solutions able to orient future interventions.

As with all research methods, however, the proposed framework is only a tool for simplifying and understanding complex data, and not a substitute for insight, clear thinking, and intimate knowledge of the subject matter.
5 Limits and future research directions

Although the IMiMe approach was useful in our case, further studies are needed to explore the applicability and heuristic value of this approach with other topics such as consumption or other domains of social media research. Furthermore, there is a need for a better assessment and definition of the tools and software to be used for achieving the different level of analysis implied in the approach, for example those used to determine the index of social media productivity and readership, and those for systematic analysis of discursive and thematic features of social media.

References


**Notes**

1 Extracts of this study results have been previously discussed in a workshop on social media analysis (Thinking Qualitatively Workshop Series, Edmonton Canada) held by the first author on July 2013.

2 The study was carried out in 2011 and only considered Italian websites and Italian text. The goal of the study was to focus on the whole spectrum of social media discourses linked to the key words: HSDD, the name of the new drug, women sexual disorders. A panel of 209 websites were considered and deeply analysed. A total amount of 1,553 posts indexed with these key words were retrieved and analysed according to the four strategies involved in the IMiMe approach.