Sensitivity of Information Disclosed in Amazon Reviews

Federica Fornaciari, C. Ranganathan, and V.N. Venkatakrishnan

University of Illinois at Chicago

Chicago, IL (USA)

fforna3@uic.edu, ranga@uic.edu, venkat@uic.edu

Abstract—As online product reviews become ubiquitous, more individuals increasingly write and rely on them. In an effort to share their experiences and opinions about a product, do individuals share private and sensitive information online? This study addresses this critical issue by examining the extent of sensitive information disclosed in Amazon.com's product reviews. We crawled Amazon.com and gathered all online reviews posted for six products that pertained to weight loss, anti-aging, sex-related, fragrance, baby care and electronic goods. This resulted in 3,485 reviews, which were text-analyzed and mined using Linguistic Inquiry and Word Count (LIWC) analysis. Then, data processed through LIWC were further analyzed through descriptive statistics and discriminant analysis. We found that Amazon's reviewers disclose high levels of sensitive information in the following categories: family, humans, positive emotions, negative emotions, sadness, cognitive mechanisms, concerns related to work, achievements, leisure and money. Sensitive disclosure is also found to be a function of the type of reviewer and of the anonymization strategies adopted.

Keywords-Privacy; Identity; Users-Generated Content; Sensitive Information; Natural Laguage Processing.

I. INTRODUCTION

With the wide spread of social network sites (SNSs) and the gained popularity of user-generated content, individuals increasingly share information online. The information posted online has various degrees of sensitivity and reveals different layers of one's personal life, identity [1], and personality traits [2]. Unfortunately, the features of online platforms open up the possibility of privacy infringements [3]. Despite the increased risks of losing control over personal information online, many still share several layers of the self, motivated by concerns that include desire of publicity [4], search for sociality [4], and narcissism [4]. The goal of this paper is to investigate patterns of self-disclosure in Amazon's reviews to further understand the levels of sensitivity of information shared by reviewers, and to provide implications related to end-user privacy concerns.

Amazon is a pioneer in incorporating customer reviews in e-commerce sites. Building on a history of improvements, the current review system in Amazon provides users with features such as ease-of-use and flexibility. For instance, users may decide to review products using their real name – and authorize Amazon to verify it using the credit cards information on their profile. Everyone, not just purchasers, has the possibility to review items. These vary from less personal products - as technology and electronics - to more personal ones - as baby products, weight loss, and anti-aging. Finally, Amazon provides the most active reviewers with different badges (e.g., Top Reviewer, Hall of Fame) to acknowledge their role within the Amazon's community. Reviews, at times, reveal detailed information about reviewers and/or about their family and friends. Perceived anonymity and trust in the community may encourage one to disclose different levels of sensitive information. And yet, deanonymization, privacy infringements, and loss of control over information may harm one's reputation and dignity, and generate psychological distress [5].

Exploring the extent of sensitive information shared in Amazon's reviews, this study pays attention to a number of factors, evaluating their role in encouraging reviewers to disclose information about the self and about others. In particular, this study measures and compares the levels of sensitive disclosure for reviewers based on their anonymization strategies (use of real name vs. nickname; location disclosure). Also, this study addresses the relationship between one's status within the Amazon's community - measured through the use of Amazon badges - and the sensitivity of information disclosed. In sum, the current research project contributes to understanding some of the factors that may encourage sensitive disclosure.

The organization of this paper is as follows. Section II discusses relevant literature. Section III outlines the problem investigated and the methodological approach. Section IV presents the findings. Sections V-VII explore and discuss findings and limitations of the current study.

II. LITERATURE BACKGROUND

A. Risks and Opportunities of Self-disclosure Online

Despite initial dystopian views, most research shows that individuals who interact through social media are exposed to both risks and opportunities. In fact, research suggests that SNSs may facilitate ties creation and maintenance, online community formation [6], [7], identity development [4], psychological reassurance, and self-expression [8]. Self-disclosure online, though, also entails risks of privacy infringements and identity theft [9], [10], commercial use of personal information [11], damages to reputation [5] stalking, reinforcing stereotypes and discrimination [12].

Clearly, self-disclosure online is an increasingly popular activity as users "share their ideas, interests, emotions, experiences, and knowledge with other" on the Web [13, p. 234]. Research has explored the motivations that may

encourage one to share personal information online. Among the perceived benefits of disclosure scholars identify four main areas: cognitive needs (as information seeking), affective needs (as entertainment), social integrative needs (as forming communities), and personal integrative needs (as identity formation) [14].

Social media provide new stages for sociality [3]. Research shows that social media are designed for sharing and connecting rather than for protecting privacy [8]. As a consequence, self-disclosure online may challenge one's ability to control personal information and to manage private and public boundaries [10]. Self-presentation online is crafted to show different angles of the self to different audiences. One may connect to distinct spheres of sociality showing different facets of one's identity depending on one's envisioned, desired or perceived audience. In such a process of disclosure, one may consider surveillance, data-mining, and behavioral marketing as remote possibilities or acceptable tradeoffs to enjoy the benefits of socialization and online community that may stem from disclosure [8]. Unfortunately though, managing levels of accessibility for different viewers is challenging and time-consuming, and ability to do so is often a function of internet literacy [3].

Research investigated the dynamics of social media reframing old questions and introducing new ones. However, as of yet, there are not studies that investigate self-disclosure in consumers' reviews sites. The current study addresses this gap in the attempt to identify and evaluate the disclosure of sensitive information (and the possible implications for privacy) in online platforms supposedly dedicated to e-commerce.

B. Sensitivity of Information

Research thoroughly explored the relationship between sensitivity of information and willingness to disclose often showing a negative correlation between the two [15]. However, most research focused on the sensitivity of information explicitly requested or required by a site. The current study is novel in its attempt to develop a method to evaluate the sensitivity of information disclosed in the unstructured texts of consumers' reviews that do no necessarily encourage sensitive disclosure.

Personal information may have different levels of sensitivity. Research often relates information sensitivity to its level of intimacy. Previous research adopted a number of strategies to measure depth and breadth of self-disclosure in consumers' reviews and online forums [16].

The main contribution of this paper is to measure the extent of self-disclosure on Amazon reviews and analyze on its privacy implications. Our approach is a quantitative one that aims to measuring the extent of self-disclosure. The main assumption of the current study is that language may be used as a valuable indicator of the sensitivity of information disclosed. Such an assumption draws from abundant research published in cognitive psychology that suggests that the words may be reflective of one's social relationships, personality, social behavior, and cognitive style. The use of language is also a meaningful indicator to measure the disclosure of positive or negative emotions and

other psychological processes [13], [17], as well as personality traits [2]. Words used may also reveal a variety of sensitive information [18], [19], [20], [21].

Our methodology involves measuring the sensitivity of information disclosed using the Linguistic Inquiry and Word Count (LIWC) software. LIWC has built-in dictionaries used to count words and separate them in psychologically meaningful categories. LIWC allows to process large samples of text thus providing valuable quantitative insight. Thus, LIWC provides a unique analytic approach that allows studying the granularity of information disclosed. Over decades of use, LIWC has been tested for validity and reliability of results, and successfully implemented to analyze text in a large variety of categories [17], [18].

For the scope of this study, the degree of sensitivity of information disclosed was measured using the framework adopted in Tausczik and Pennebaker's work [17] and implemented through the software LIWC. In particular, this study used LIWC to measure the following: social processes (family, friends, humans), affective processes (swear, positive emotion, negative emotion, anxiety, anger, sadness, cognitive mechanisms), biological processes (health, sexual), and personal concerns (work, achievements, leisure, home, money, religion, death) [17], [18].

Alternative scalable methods to study sensitive information in large portions of text include opinion mining, sentiment analysis, and other forms of natural language processing. These methods allow one to investigate point of view and subjectivity as they emerge from textual analysis [22]. For example, opinion and sentiment analysis have been successfully implemented for fake reviews detection in Amazon [23] or to mine and classify opinions and emotions from reviews in the blogosphere [13]. Alternatively, natural language processing has enabled the study of personality traits in SNSs [2]. Even though opinion and sentiment mining are very powerful methodological approaches, they tend to focus on solving opinion-oriented classification problems. As a consequence, they were not considered suitable for the scope of this study.

C. Research Questions

In particular, data were collected to address the following research questions:

RQ1 - To what extent do Amazon's reviewers reveal sensitive information when reviewing a product?

RQ2 – Is there a relationship between the disclosure of sensitive information and the use of a real name?

RQ3 – Is there a relationship between the disclosure of sensitive information and the disclosure of one's location?

RQ4 – Is there a relationship between type of reviewer and sensitivity of information disclosed?

III. METHOD

A. Types of Products

Amazon includes a large number of products whose nature may encourage different degrees of disclosure. For the current research, we selected six products across the spectrum in the attempt to implement a study that would be

doable yet comprehensive, exploring a breadth of products that may prime individuals to disclose different kinds of sensitive information. The items selected pertained to the following categories: sex-related, weight loss, anti-aging, fragrance, baby care, and electronic.

B. Variables and Data Collection

Amazon reviews are public. This facilitated the data collection that was operated through a crawler launched in the Amazon website in November 25th, 2012. The data collection process generated 3,485 .txt files of review for six different products. The unit of analysis was the single review. Each file included the text of the review as well as the following variables: real name (y/n), top reviewer (y/n), hall of fame reviewer (y/n), vine voice (y/n), length of review, location (y/n), number of stars (1-5), and number of reviews posted by the reviewer.

Some of these variables are identified through badges that Amazon awards to its reviewers. In particular, the badges Top Reviewer and Hall of Fame identify, respectively, reviewers who provided the most helpful contributions recently and longitudinally. The badge Vine Voice is provided to reviewers who received a free product for review. Finally, for the purpose of the study, "location" was turned into a yes/no categorical variable to distinguish between those who disclosed a "realistic location" (that included both city and state) from those who did not disclose their location or that provided vague or unrealistic information (e.g., state only, country only, or phantasy names).

The texts of the reviews were processed through the software LIWC to measure multiple variables that could be used as indicator of sensitive information. In particular, based on existing literature, the current study considered the "level of sensitivity" of information as a multidimensional variable. LIWC allowed measuring the percentage of words belong in each of the following categories: social processes (family, friend, humans); affective processes (swear, positive emotion, negative emotion, anxiety, anger, sadness, cognitive mechanisms); biological processes (health, sexual); and personal concerns (work, achievements, leisure, home, money, religion, death). Afterwards, we merged the six result files created through LIWC to generate a comprehensive spreadsheet that could be inputted in SPSS for statistical analysis.

IV. DATA ANALYSIS

To describe our sample we run descriptive statistics for the whole 3,485 reviews. Descriptive statistics included frequencies for categorical variables (real name, top reviewer, hall of fame reviewer, vine voice, and location). They included means, standard deviations, and range for continuous variables (length of review, and number of reviews posted by the reviewer).

Our sample included a larger number of reviewers who did not disclose their real name (72.4%), or their location (72.4%). Most did not belong in the Hall of Fame (99.8%), in the Top Reviewer (98.8%) or in the Vine Voice (96.4%). The typical reviewer in our sample published 30 reviews

(SD = 216; range = 5675), whose average length was of 97 words (SD = 107.64; range = 2080). In addition, most reviews were positive. In particular in a scale from 1 (worst) to 5 (best) the average number of stars was M = 4.26, SD = 1.23.

A. RQ1 - To what extent do Amazon's reviewers expose sensitive information when reviewing a product?

To answer the first research question, we measured the percentages of use of words per category of sensitive information in our sample. Then, we compared these results with the average level of information disclosed derived from a study conducted by Pennebaker and colleagues [17]. The latter study is the outcome of a collection and analysis of words used across a variety of settings including: emotional writing, control writing, science articles, blogs, novels, and talking (aggregated sample, N = 721,726).

From the comparison, Amazon reviewers use significantly more words belonging in the following categories: family (overall mean of use = .48%), humans (.78%), positive emotions (5.12%), negative emotions (1.7%), sadness (.48%), cognitive mechanisms (17.03%), and concerns related to work (2.71%), achievements (3.31%), leisure (1.31%), and money (1.66%). Significance was measured at the 95% confidence level.

B. RQ2 - Is there a relationship between the disclosure of sensitive information and the use of a real name?

To address the second research question we analyzed the whole sample comparing the disclosure of sensitive information for those who used a real name badge against those who did not. As we ware exploring the relationship between a categorical variable (real name badge) and a multidimensional continuous variable (level of sensitivity), we measured the strength of the relationship using a *discriminant analysis* with a 95% level of confidence.

The discriminant analysis highlighted some significant difference in the word use between the real name group and the non-real name group. In particular, reviewers who disclosed their real name were significantly more likely to use words in the following categories: sadness (Wilk's Lambda = .996, F = 14.09), health (Wilk's Lambda = .994, F=20.60), and concerns related to achievements (Wilk's Lambda=.995, F=16.79). The real name group was less significantly likely to discuss leisure-related concerns (Wilk's Lambda = .985, F = 51.32). Unfortunately, our sample included a larger number of non-real name reviewers (non-real name N = 2524; real name N = 961). As a consequence, the differences found may be affected by the differences in the size of the groups compared.

Finally, we calculated the level of sensitive information aggregating the frequencies of words use for all the categories analyzed. Such an aggregated value was then used to conduct a second *discriminant analysis* at the 95% level of confidence. Interestingly, such an analysis revealed a significant difference (Wilk's Lambda = .999; sig. = .023) showing that, overall, reviewers who used their real names disclosed higher levels of sensitive information.

In sum, individuals who used real names tended to disclose more information involving sadness, health processes and concerns related to personal achievements. They were less likely to discuss leisure-related concerns.

C. RQ3 - Is there a relationship between the disclosure of sensitive information and the disclosure of one's location?

To answer the third research question we compared the disclosure of sensitive information for those who provided their location against those who did not (or disclosed a vague or unrealistic location). To measure the strength of such a relationship, we used *discriminant analysis* with a 95% level of confidence to test each category of sensitive information. Afterwards, we run a second *discriminant analysis*, still at the 95%, to test the aggregated disclosure of sensitive information.

From the first discriminant analysis, we found significant differences in information disclosure between reviewers who revealed their real location and those who did not. In particular, those who disclosed their location were significantly more likely to use words in the following categories: sadness (Wilk's Lambda = .997, F = 11.48), health (Wilk's Lambda = .994, F = 20.75), achievements (Wilk's Lambda = .998, F = 6.80), and religion (Wilk's Lambda = .997, F = 10.43). They were significantly less likely to use words that belong in the following categories: positive emotions (Wilk's Lambda = .998, F = 7.01), sexual concerns (Wilk's Lambda = .998, F = 8.24), and leisure (Wilk's Lambda = .987, F = 47.40). Similarly to RQ2, reviewers who disclosed their location (N = 962) were much less than those who did not (N = 2,523). The second *discriminant analysis* showed that those who disclosed their location were slightly more likely to share sensitive information. Yet, such a difference was not found to be significant.

In sum, and consistently with the findings related to RQ2, individuals who disclosed their location tended to use more words related to sadness, they discussed more health processes, and were more likely to tackle concerns related to personal achievements.

D. RQ4 - Is there a relationship between type of reviewer and sensitivity of information disclosed?

To address the fourth research question we run a number of *discriminant analyses* at the 95% level of confidence to evaluate the relationship between each of the categorical variables related to the "type of reviewer" (Hall of Fame, Top, Vine Voice) and the multilevel continuous variable "level of sensitivity." As a result, some statistically significant differences were found.

In particular, Hall of Fame reviewers were significantly less likely to disclose affective processes (sig = .048; F = 3.903). Vine Voice reviewers were significantly less likely to use words belonging in the following categories: family (sig. = .007; F = 7.19), friends (sig. = .031; F = 4.68), humans (sig. = .001; F = 10.2), affective processes (sig. = .000; F = 26.7), positive emotions (sig. = .000; F = 12.58), negative emotions (sig. = .000; F = 12.33), anger (sig. =

.002; F = 9.6), sadness (sig. = .031; F = 4.67), cognitive mechanisms (sig. = .000; F = 22.7), leisure (sig. = .000; F = 21.78), home (sig. = .048; F = 3.9), and money (sig. = .000; F = 20.43). Top reviewers were significantly less likely to use words belonging in the following categories: affective processes (sig. = .004; F = 8.22), positive emotions (sig. = .043; F = 4.1), leisure (sig. = .042; F = 4.1), and money (sig. = .031; F = 4.66). Verified Purchase reviewers were more likely to use words in the following categories: swear (sig. = .019; F = 5.5), affective processes (sig. = .000; F = 18.93), positive emotions (sig. = .000 F = 16.04), work (sig. = .000; F = 13.1), achievement (sig. = .012; F = 6.27), and leisure (sig. = .000; F = 52.24). They were less likely to use words belonging in the categories that follow: humans (sig. = .014; F = 6.06), and health (sig. = .004; F = 8.5).

Additionally, we run discriminant analysis at the 95% level of confidence to gauge the relationship between type of reviewers and aggregated level of sensitive information disclosed. Such an analysis revealed significant differences in the disclosure of sensitive information. In particular, the groups more likely to engage in such a disclosure were the following: non-Hall of Fame reviewers (Wilk's Lambda = .997; sig. = .002); non-Top reviewers (Wilk's Lambda = .996; sig. = .000); non-Voice Vine, (Wilk's Lambda = .993; sig. = .000).

Unfortunately, the sample analyzed included importantly larger number of "regular reviewers" (non belonging in the categories Hall of Fame, Top, or Vine Voice). Such a distribution may likely reflect the general composition of the Amazon community – where most reviewers are occasional and non-professional - yet the differences found in our analysis are likely affected by the differences in the size of the groups compared.

In sum, regular reviewers (as opposed to reviewers who are awarded the special badges identified in this study) were often more likely to disclose sensitive information. They consistently tended to share higher level of personal information belonging in many categories, perhaps as a way to increase their personal participation in the Amazon community. These results may be consistent with common sense expectations. Yet, further research is necessary to further explore them, as the sample used in the current study included a limited number of non-regular reviewers (e.g., Top, Hall of Fame).

V. DISCUSSION

As it emerged from our analysis, Amazon reviewers in the sample collected tend to reveal higher level of sensitive information, compared to the average [17], in the following categories: family, humans, affect, positive emotions, negative emotions, sadness, cognitive mechanisms, and concerns related to work, achievements, leisure and money. Such a finding may suggest that people who post reviews online do so to actively participate in the Amazon community. Perhaps, they feel to be part of a trusted social circle within which one feels relatively safe in the disclosure of sensitive information about the self and the others – maybe without considering that Amazon reviews are public. These reviewers, in fact, do not seem to post reviews for the

gratification of receiving specific badges (otherwise they would be more active reviewers). And yet, they share high levels of sensitive information - levels that increase for those who also disclose their real offline identity (name and location). Trust in the perceived community may be an important component of the equation. These findings seem to suggest that many experience Amazon as a venue built around people who show their humanity, their social connection, their affective processes, their emotions, and their concerns. Current findings also suggest that many consider Amazon as a platform for building community and sharing information about one's social circles. Doing so, users partly "reinvent Amazon" by mingling the affordances of online retailing websites with those of SNSs [6]. Amazon, similarly to most social media, becomes a platform structured around individuals, where users may become the center of personal communities that share interests and life experiences.

Consistent with research conducted to understand participation in SNSs, this study may suggest that a large component of Amazon reviewers behave as active members of a community and use the website as a platform to perform their identity. Previous research suggests that those who are active SNSs users tend to have lower privacy concerns [24]. Such a consideration may apply to the current study as well. Admittedly, though, the data we collected and analyzed are not sufficient to claim that individuals who post product reviews on Amazon are less concerned about their privacy. However, our findings show that individuals who decide to post reviews online are likely to talk about their personal experiences - as well as about their social relationships – often disclosing high levels of sensitive information. And levels of disclosure increase for non-anonymous reviewers.

As detailed in Section II, research shows that need of social capital and desire of community building are strong factors motivating individuals to disclose information. Similarly, high levels of self-disclosure in Amazon may be motivated by the desire to develop and maintain online community. Amazon makes it easy for its users to post and read reviews and comments, perhaps presenting itself as a network that fosters sociality and publicity, and thereby encouraging users to disclose rather than withhold information. Similarly to what research has pointed out for SNSs [9], Amazon's network seem to have a large utility for its users who can develop sense of belonging and participation. Importantly, such a potential may implicitly encourage users to disclose - thereby also increasing the commercial value of Amazon.

The importance of Amazon for community building begins to emerge from the data addressing the first three research questions. In particular, individuals who disclose their identity in Amazon (real name and/or location) tend to disclose more information about their social processes, their sadness, their biological and health processes and their concerns related to personal achievements. Findings related to RQ2 and RQ3 reveal fairly similar tendencies, suggesting that individuals who disclose more information about their offline identity (real name and location) are also those who

appear to need social support. In fact, they disclose personal concerns as to reveal their needs and their weaknesses (e.g., sadness and concerns). As research suggests, individuals who seek social capital are often willing to accept privacy risks [12].

Finally, data analyzed to answer the fourth research question emphasize that "normal reviewers" consistently tend to share higher levels of sensitive information thus increasing their personal participation in the Amazon community. Unfortunately, this finding was significantly limited by the fact that our sample included few reviewers belonging in the categories Top, Hall of Fame and Vine Voice. To address such a limitation, a future study could be conducted from a users-centered perspective (using the reviewer as unit of analysis - instead of the review as we did in the current study - and collecting reviews based on the use of badge). A comparison of equally sized groups of reviewers would allow a better assessment of these findings.

VI. Limitations

Even though this study provided a number of contributions to the understanding of disclosure in online reviews, its scope had some limitations. Needless to say, Amazon commercializes thousands of products that belong in a wide variety of categories. Despite the attempt to select a number of products that would provide multiple perspectives to render the variety of patterns of self-disclosure in Amazon, the sample was limited to six products and, likely, provided a partial representation of the population analyzed. Thus, results may not be generalized to the entirety of Amazon's reviews, or to other communities of consumer's reviews. Despite such a limitation though, we believe that the current study provided a valuable contribute to research tackling online disclosure and related privacy risks. The use of LIWC and its ability to capture the granularity of information, particularly contributed to this outcome.

Also, some could argue that the LIWC software limits its evaluation of sensitive information to the use of words, taking them outside of their context of delivery. As a consequence, one may suggest that LIWC fails to capture the nuances of language, and label words as belonging in a category they do not really belong in. Even though such a critique may provide some fundamental insight that one need to take into consideration when analyzing text, it is normally assumed that the analysis of large samples of text (ours included 3,845 reviews) would control for such a risk.

VII. Conclusions

In this paper, we examined the extent of sensitive information disclosed in Amazon.com's product reviews. This was done by crawling Amazon's pages and gathering all online reviews posted for six products that pertained to weight loss, anti-aging, sex-related, fragrance, baby care and electronic goods. This resulted in 3,485 reviews, which were text-analyzed and mined using LIWC analysis. We further analyzed the results of the text-analysis through descriptive statistics and discriminant analysis. We found that Amazon's reviewers disclose higher levels of sensitive

information in the following categories: family, humans, positive emotions, negative emotions, sadness, cognitive mechanisms, concerns related to work, achievements, leisure and money. In addition, occasional and non-professional reviewers provided higher level of sensitive information, perhaps as a way to increase their participation in the Amazon community.

The conclusions for our study raises several open questions: first, whether it would be possible, by using methods similar to ours, to provide usable warning indicators that inform end-users when they input privacy sensitive reviews. A more ambitious (but perhaps more usable) system would also provide deanonymizing suggestions in case the system finds certain reviews to be sensitive. Continuing to retain the high quality of reviews similar to those found in Amazon, while providing deanonymizing suggestions would be a challenge to current socio-technical systems.

VIII. ACKNOWLEDGMENT

IX. REFERENCES

[1] D. G. Weckowski and J. Małyszko, "On information exchange for virtual identities: survey and proposal," *ICSD 2013 - The Seventh International Conference on Digital Society,* March, 2013, pp. 59-64.

[2] F. Celli, "Unsupervised personality recognition for social network sites," *ICSD 2012 - The Sixth International Conference on Digital Society,* March, 2012, pp. 59-62.

[3] A. Marwick and D. Boyd, "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience," *New Media & Society,13(1),* pp. 114-133, 2011.

[4] A. L., Mendelson and Z. Papacharissi, "Look At Us: Collective Narcissism in College Student Facebook Photo Galleries," in *A Networked Self: Identity, Community, and Culture on Social Network Sites*, Z. Papacharissi, Ed. New York: Routledge, 2011.

[5] D. J. Solove, "The Future of Reputation: Gossip, Rumor, and Privacy on the Internet," Yale University Press, 2007.

[6] D. Boyd and N. Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication, 13*, pp. 210-230, 2008.

[7] M. Forestier, J. Velcin, D. A. Zighed, "Analyzing social roles using enriched social network on on-line sub-communities," *ICSD 2012 - The Sixth International Conference on Digital Society,* March, 2012, pp. 59-62.

[8] J. Jarvis, "Public Parts. How Sharing in the Digital Age Improves the Way We Work and Live'" New York: Simon & Schuster, 2011.

[9] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the Facebook," in *Proceedings of 6th Workshop on Privacy Enhancing Technologies*, P. Golle & G. Danezis, Eds. Cambridge, U.K.: Robinson College, 2006.

[10] H. Nissenbaum, "Privacy in Context. Technology, Policy, and the Integrity of Social Life," Stanford: Stanford Law Books, 2010.

[11] A. Odlyzko, "Privacy and the clandestine evolution of e-commerce," *ICEC 2007 - Proceedings of the ninth international conference on Electronic commerce*. New York, NY, USA: ACM, August, 2007, pp. 3-6.

[12] N. Ellison, C. Lampe, C. Steinfield, and J. Vitak, "With a Little Help From My Friends. How Social Network Sites Affect Social Capital Processes," in *A Networked Self: Identity, Community, and Culture on Social Network Sites*, In Z. Papacharissi, Ed. New York : Routledge, 2011.

[13] A. Baloglu and M. S. Aktas, "An Automated Framework for Mining Reviews from Blogosphere," *International Journal of Advances in Internet Technology, 3(3&4),* pp. 234-244, 2010.

[14] L. Leung. "User-generated content on the internet: an examination of gratifications, civic engagement and psychological empowerment," *New Media & Society, 11,* pp. 1327–1347, 2009.

[15] D. L. Mothersbaugh, W. K. Foxx, S. E. Beatty, and S. Wang, "Disclosure antecedents in an online service context: The role of sensitivity of information," *Journal of Service Research, 15(1),* pp. 76-98, 2012.

[16] Y. Moon, "Intimate exchanges: Using computers to elicit self‐disclosure from consumers," *Journal of Consumer Research, 26(4),* pp. 323-339, 2000.

[17] J. W. Pennebaker, R. J. Booth, and M. E. Francis, "Operator's Manual. Linguistic Inquiry and Word Count," 2007.

[18] Y. R. Tausczik and J.W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of Language and Social Psychology, 29(1),* pp. 24–54, 2010.

[19] G. W. Alpers, et al., "Evaluation of computerized text analysis in an Internet breast cancer support group," *Computers in Human Behavior, 21*, pp. 361-376, 2005.

[20] D. J. Houghton and A. N. Joinson, "Linguistic markers of secrets and sensitive self-disclosure in Twitter," *45th Hawaii International Conference on System Science (HICSS)*. IEEE, January, 2012, pp. 3480-3489.

[21] J. M. Smyth, "Written emotional expression: Effect sizes, outcome types, and moderating variables," *Journal of consulting and clinical psychology, 66(1),* pp. 174-184, 1998.

[22] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies, 5(1),* 2012.

[23] B. Liu, "Sentiment analysis and subjectivity," *Handbook of Natural Language Processing*, pp. 627–666, 2010.

[24] J. Fogel and E. Nehmad, "Internet social network communities: Risk taking, trust, and privacy concerns," *Computers in Human Behavior, 25*, pp. 153-160, 2009.