



I can't believe she wrote that on amazon!

Patterns of self-disclosure in amazon's reviews

by Federica Fornaciari¹, C. Ranganathan², and V.N. Venkatakrisnan³

¹ Department of Communication, ² Department of Information and Decision Sciences, ³ Department of Computer Science, University of Illinois at Chicago



1. motivations

- As online product reviews become ubiquitous, many individuals write and rely on them. Amazon is the biggest online retailers in the world.
- When reviewing products, one may share **private** and **sensitive information** about the self or about other people. Consequent privacy risks include de-anonymization, identity theft, loss of reputation, and psychological harm.
- The current study addresses this critical issue by examining the **extent of sensitive information disclosed** in amazon.com's product reviews.
- For example, here's part of a sensitive review written by a user who reveals her **real name** and **location** and who **posts a video of her son**:

"this is my video review of **my 5-month-old son** playing with (name of product). Little (name of the kid) in this video **is turning 4** at the end of the month [...]. Our daughter had 6 teeth **when we adopted her at 16 months old** [...]."

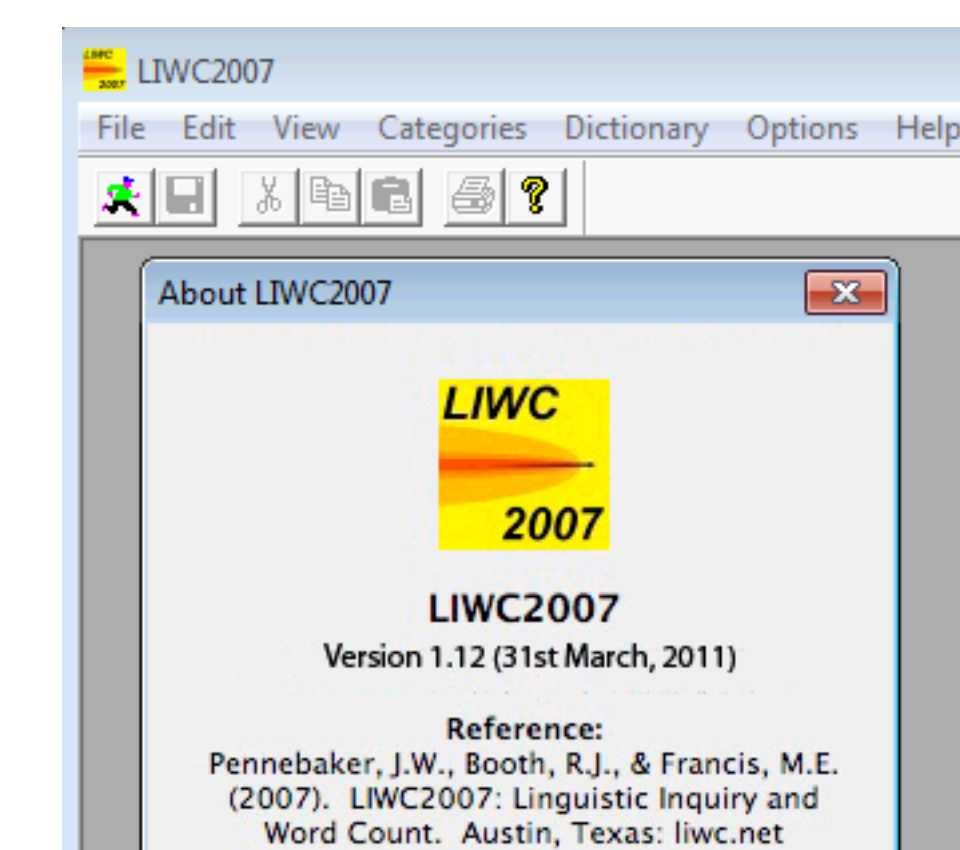
2. theoretical context

When engaging in self-disclosure online, one weighs the potential for **publicity** and **sociality** against the risks for **privacy infringements**:

- social media facilitate **ties creation and maintenance**, online **community** formation, **identity** development, psychological **reassurance**, and **self-expression** (Ellison et al,2011).
- self-disclosure online, though, also entails risks of **privacy infringements** and **identity theft** (Acquisti & Gross, 2006), **commercial use** of personal information, and damages to **reputation**.

Sensitivity of information is positively correlated to "the potential loss associated with the disclosure of that information" (Mothersbaugh et al., 2012).

We measured information sensitivity using the purchased software **Linguistic Inquiry Word Count** for textual analysis (Pennebaker et al., 2007; Tausczik & Pennebaker, 2010).



3. amazon.com

Amazon developed the following "**badges**" to characterize reviewers:

- hall of fame** - reviewers who have been highly ranked in terms of helpfulness, in a longitudinal basis
- voice vine** - reviewers who received a free product for review
- top reviewer** - reviewers highly ranked in terms of helpfulness, in a product-by-product basis

Additional badges include:

- verified purchase**, **real name**, and **number of reviews posted**

Amazon includes different "**genres**" of products, as the following:

- baby, beauty, collectibles, electronics, jewelry, music, sports and outdoor, and many more...

Different "**genres**" of products may encourage a reviewer to reveal different types of information, more or less personal, and more or less sensitive.



4. RQs & design

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

Is there a **relationship** between **disclosure of sensitive information** and:

RQ2 - use of a **real name**?

RQ3 - disclosure of one's **location**?

RQ4 - **genre** of the **product** reviewed?

RQ5 - type of **reviewer**?

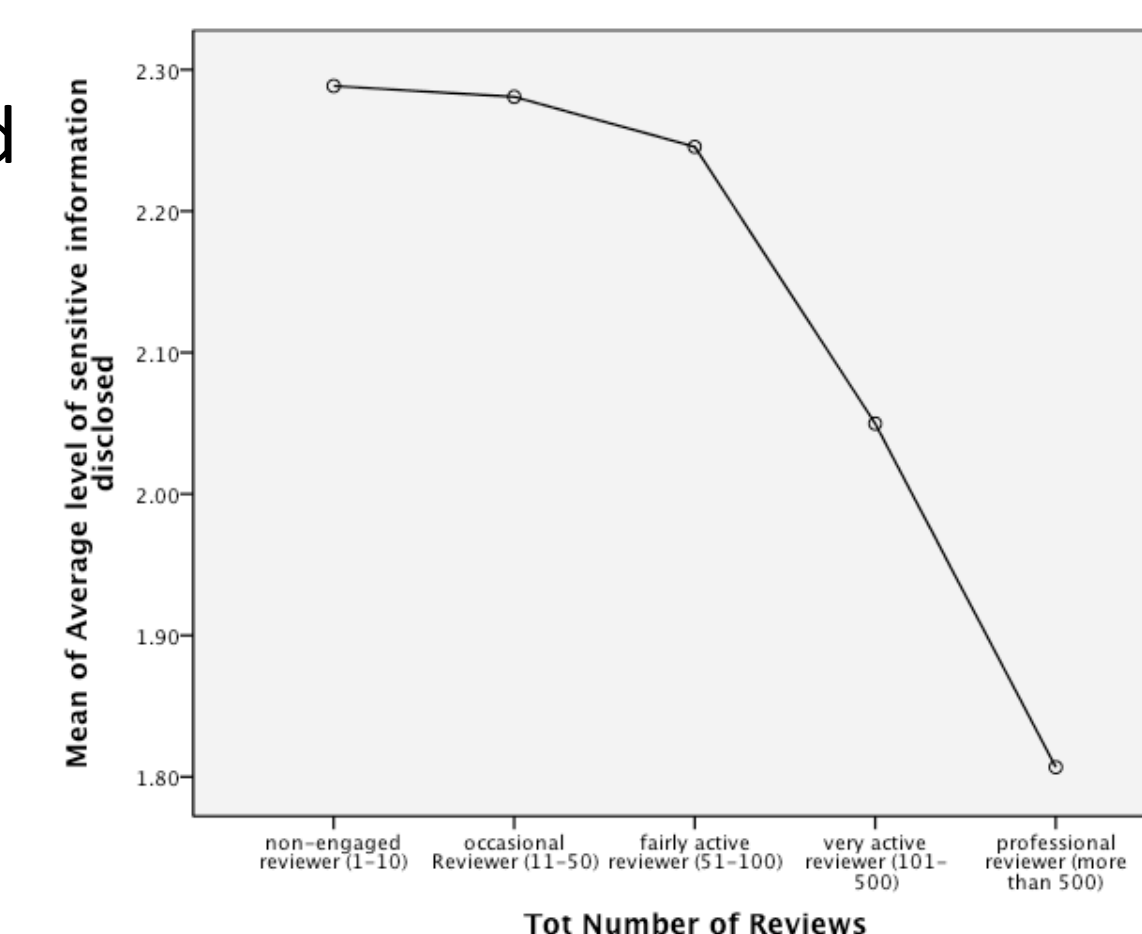
- We collected **3,485 reviews** for **6 products**: anti-aging, baby, electronic, fragrance, sex-related and weight loss.
- We analyzed the reviews using the software LIWC. LIWC has built-in dictionaries that count words and separate them in **psychologically meaningful categories**. For this study, we considered the use of words in the following categories: **pronouns**, **social processes**, **affective processes**, **biological processes**, and **personal concerns**. Each of these categories includes a number of sub-categories.
- Then, we processed the data through SPSS implementing **descriptive statistics**, **discriminant analysis** and **ANOVA techniques**.

5. results

- Compared to average levels of disclosure (Pennebaker et al., 2007), Amazon's reviewers share significantly higher levels of sensitive information related to **family**, **humans**, **positive emotions**, **negative emotions**, **sadness**, **cognitive mechanisms**, **concerns related to work**, **achievements**, **leisure** and **money**.

- Those who use a **real name** or disclose their **location** reveal significantly higher levels of personal information about **sadness**, **health processes**, and concerns related to **personal achievements**, and less about **leisure related concerns**. Perhaps they do so seeking social and psychological support.

- The **sensitivity of information** disclosed is a function of the "**genre**" of the product reviewed.
- Non-professional** and **occasional reviewers** disclose higher levels of sensitive information (see graph). This suggests their higher engagement in the Amazon community.



6. conclusion & references

Further research could investigate ways to provide:

- usable **warning indicators** that inform end-users when they input privacy sensitive reviews (for example, as pop-up windows)
- appropriate **de-anonymizing suggestions** in case the system finds certain reviews to be sensitive

References

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

Ellison, N. B., Vitak, J., Steinfield, C., Gray, R., and Lampe, C. 2011b. Negotiating privacy concerns and social capital needs in a social media environment. In S. Trepte and L. Reinecke, Eds. Privacy Online, Berlin: Heidelberg.

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

Pennebaker, J. W., Booth, R. J. and Francis, M. E. 2007. Operator's Manual. Linguistic Inquiry and Word Count: LIWC2007..

Tausczik, Y.R., Pennebaker, J.W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Soc Psychology, 29(1), 24-54.

