Exploring the nature and frequency of Social Presence in a Voluntary Online Learning Platform

Aatish, Ali, Harshita, Nishitha, Zeenar

Introduction

The study focuses on investigating the extent and occurrence of Social Presence in a voluntary online learning platform provided by Zooniverse to engage the public in citizen science. The question for the inquiry is: What is the nature and frequency of Social Presence, as defined by the Community of Inquiry framework, in the voluntary online discussions on the Zooniverse platform?

The paper describes the theoretical framework of Community of Inquiry with a special focus on Social Presence that underpins our study. The paper highlights the current literature that uses the Social Presence theory and situates our inquiry questions in the current literature. In addition, it discusses the research design that we propose to undertake to find the answer to the research question and explains the significance of our study in relation to its practical and theoretical contribution.

Theoretical Framework

The research is built on the Community of Inquiry framework developed by Garrison, Anderson and Archer, 1999). The key assumption of the model is that learning occurs in a social environment with the interaction of three key elements i.e., social presence, cognitive presence and teaching presence. In their seminal paper, Garrison, Anderson, and Archer (1999) used the text-based data in computer-mediated conversations in learning settings to investigate the indicators for the three kinds of presences (social, cognitive and teaching).

Building on this work, Rourke, Anderson, Garrison, and Archer (1999) explained social presence as "the ability of learners to project themselves socially and affectively into a community of inquiry". They developed a coding scheme for social presence. They distributed social presence into three categories including: affective, interactive and cohesive. The definitions and examples of the three elements that constitutes the social presence are as follows: (i) Affective responses are described as expression of feelings through written communication. Expression of emotions, humor and self-disclosure are three key representations of expression of feelings in written communication. (ii) Interactive responses are defined as the acknowledgement of other posts, quoting others, asking questions, referring explicitly to others' comments, complimenting and commenting on others' comments and expressing the agreement/disagreement with another person. The features of interaction are emulated in face-to-face interactions through gestures, facial expressions and verbal interpersonal communication. These interactions are required for people to feel comfortable, welcomed and be able to express their ideas. (iii) Cohesive responses are defined as feelings of an 'ingroup' or 'quorum' developed by using phrases that reflect the sense of community. Such interactions may include referring participants by their names, using greetings and salutations and using group words such as 'we' and 'us' etc (Rourke, et. al., 1999)

Literature Review

Much of the earlier social presence research focused on how modalities with varying levels of immersion afford different levels of presence. Research on the impact of modality on social presence to date most often compares (1) Computer-mediated Communication (CMC) with Face-to-face (F2F) communication, (2) text-based CMC with other different audiovisual modalities, and (3) immersive virtual environments with non-immersive virtual environments.

Historically, F2F interaction has been considered to be the hallmark of social presence, therefore, a considerable amount of research compares F2F

communication with CMC to determine how successful 'virtual world' is at establishing a social presence. Most of these studies found that communicators experience lower levels of social presence during CMC compared to FtF conversations. Rourke, Anderson, Garrison and Archer developed the coding scheme for coding the social presence in a text-based online discussion conversation, using Garrison et. al. (1999) Community of Inquiry framework. The social presence was then identified with the cognitive presence, inciting the finding that the social presence has a positive relationship with the cognitive presence.

Researchers have investigated social presence in online class environments for decades (Lowenthal & Dunlap, 2020; Akcaoglu and Lee, 2018). For example, Akcaoglu and Lee (2016) explored the effect of students' group size on their perceptions of social presence in graduate-level online courses. They found that students in small group discussion (versus whole class discussion) perceive a higher level of social presence in terms of sociability, social space, and group cohesion. A recent study by Akcaoglu and Lee (2018) revealed the use of Facebook groups as supplemental social spaces in asynchronous online courses positively impacts students' perceptions of social presence and their connection with other classmates and instructor.

Several studies have also tried to investigate the motivations of the participants and the relationship between the motivation, participation and learning of the participants in the citizen science projects. Curtis (2015) found that a high level of social interaction was a powerful motivator and can help sustain participation. Price and Lee (2013) showed that participants' improvements in scientific literacy were related to social participation in the program. Jennett, et. al. (2015) showed that, in relation to motivation and learning, there is a virtuous cycle where a volunteer improves their knowledge and skills by doing the tasks and sharing it in a community of peers helps to increase their self-confidence and their desire to share. The community helps them become more competent in both performing the task and assuming new roles in the community. Jackson, Osterlund, Crowston, Harandi, and Trouille (2020) explored the changes in the volunteer's engagement with the resource in relationship to the change in project and complexity of the tasks assigned to them. They found that novice volunteers draw their understanding from the subject authority as well as community members in Zooniverse platforms. However, as they become experts in their journeys they rely substantially on their agency and the communal responses and rely less on subject-authority for feedback and advice. It is evident that, studies clearly vouch for the importance of social presence in building learning, motivation and agency in citizen science projects, however, it would be interesting to describe the nature of the social presence by identifying the occurrence and extent to which the affective, interactive and cohesive features of social presence are found in the posts during voluntary participation in citizen science project.

Significance

Literature indicates that in citizen science projects having social interaction improves individual's participation (Curtis, 2015), keep up their motivation to participate (Price and Lee, 2013), improvements in scientific literacy and eventually contributes to the knowledge and self-confidence (Jennett, et. al., 2015). It also improves individual's agency and command on the topic and willingness to participate in and learn from the communal interactions (Jackson, et. al., 2020). Considering the importance of social presence, we studied the association between the constructs in social presence theoretical framework proposed by Garrison et. al. (1999).

Considering the importance of social presence in continuing the learning and for the development of a pool of citizen scientists, it is important to investigate how people undertake social interactions in a text-based environment developed to contribute to the citizen science projects. This paper intends to contribute to the literature by analyzing text-based interactions among volunteers in a citizen science web portal, Zooniverse.

Research Design

Our design for the research consisted of two phases:

- Annotating a sample of worker comments for the type of social presence using the Community of Inquiry framework.

- Building classifiers for each category using the sample and pre-trained models.

For the first part, we collected a sample of 400 comments of the workers of the Gravity Spy project from the Zooniverse discussion forums using web scraping tools, along with other comment related information like author, topic, section and whether the comment is a reply or not. We used the text part of the comments along with the URLs for the analysis. Two of the team members provided annotations for each of the comments and for each of the 11 categories of social presence as defined by the modified version of the Community of Inquiry framework, and we calculated the intercoder agreement for each category to measure the quality of the data. Disagreements were resolved through a discussion and for cases where no agreement was reached, we used the annotations of the coder we had more confidence in to decide the label for the record.
For the second part, we built classifiers for each of the sub-categories as they are not mutually exclusive. Two kinds of classifiers were used depending on the linguistic characteristics of the sub-categories – keywords-based and machine learning-based. The descriptions of the classifiers for each sub-category are given in Table 2 (Appendices). These classifiers were then used to make predictions on a larger collection containing 90,501 comments (~85% of all the comments for Gravity Spy) to predict social presence in the discussions and the results were analyzed using statistical tools. To explore the association between the three subcategories to see if the model proposed by Rourke et. al., (1999) holds true in this case we used chi-square test and for investigating the occurrence of social presence features, descriptive percentages and frequencies were used.

Results

Krippendorff's alpha values for annotations provided by the coders, and the precision, recall and F1 scores for the sub-categories for which a machine learning model was used to create the classifiers are provided in Table 3 (Appendices). All the intercoder agreement values were above 0.65 which shows substantial level of agreement between the coders. The training dataset was heavily imbalanced with more than 85% of the data in the negative class for all the sub-categories and this, combined with the small size of the dataset, affected the performance of the models.

The predictions using the classifiers showed that ~37% of the comments had the presence of at least one of the categories of social presence (Figure 1, Appendices). To understand the nature of the social presence and the association between the three categories of social presence including affective, interactive, and cohesive, we conducted three separate chi-square tests of association on the same discussion threads retrieved from Zooniverse platform. The test explored if the three sub-categories are associated with each other, as the theory purports. All three components of social presence were nominal variables, i.e., representing presence or absence of the category in a comment. The results are as follows:

i. The affective and interactive components of social presence are significantly associated, χ^2 (1, N=90501) = 6951.10, $p < 0.001$. 37.8% of the comments showing interactive presence also show affective presence (Table 4, Appendices).
ii. The affective and cohesive components are significantly associated with each other, χ^2 (1, N= 90501) =13179.07, $p < 0.001$. 72.5% of the comments showing cohesive presence also show affective presence (Table 5, Appendices).
iii. The interactive and cohesive components also depict significant association, $\chi 2$ (1, N= 90501) = 15834.23, $p < 0.001$. 91.5% of the comments showing cohesive presence also show interactive presence (Table 6, Appendices).

The category-wise distribution showed that the frequency of the "interactive" category was the highest (~27% of all comments) (Figure 2, Appendices) amongst the three subcategories (affective, interactive and cohesive). While interacting with each other, volunteers asked questions, made explicit reference to each other's messages, continued the conversation thread, engaged with others' comments by expressing agreement and/or disagreement, and complimented or criticized each other's opinion.

At the sub-category level, the frequency of people asking questions was the highest (~17%

of all comments) (Figure 3 in Appendices) followed by expression of emotions and self-disclosure.

Discussion

Importance of social space and interactions cannot be denied, especially in the age of covid-19, where there is reduced face-to-face communications and people are living in physically distanced worlds. Our study explored the social interactions in the virtual platform, developed to engage citizens in science projects.

While studying the social interaction in the environment, we found that around 37% of the discussion posts were evidencing the social interactions. Most posts used the interactive features, followed by affective and cohesive features of social interaction. This implies that individuals communicate with each other using more interactive features of the conversation frequently by continuing a thread, replying to each other, and agreeing or disagreeing with each other's' posts.

It is interesting to note that all the three features of social presence stipulated by Garrison et. al. (ibid, 1999) i.e., interactive, affective and cohesive were found to be significantly associated, providing an empirical evidence for the confluence of categories and sub-categories.

Within the interactive category and amongst 11 subcategories, 'asking questions' was the most frequent social function (Figure 3 in Appendices) evident in the discussion posts. While doing annotations, we realized that through asking questions, volunteers specified a concept that they were not clear about, clarified any misconceptions, generated discussion on a topic and/or investigated about a technical glitch they were facing in the system.
Surprisingly, asking questions is followed by expression of emotions and self-disclosure. This implies that volunteers rely more on expressing their emotions while discussing a topic, giving feedback to each other and/or building on each other's comment and also disclose features of their identity (such as their past experience, their skill level in the Zooniverse platform, the expenditure of their time during the project, etc.). The expression of emotions, being a fundamental human characteristic, brings 'life' to the discussions in a virtual environment. In the discussions, we noticed people expressing the emotions of surprise, wonder, smile, raising hands, showing appreciation, walk/running, thanking and appreciation gestures, "googley eyes", amongst others. These emotions are conceived as an alternate to gestures and postures and illustrate the content in the textual messages (Sternbergh, 2014). Self-disclosure, however, might be helpful for individuals because that helps them to relate with the challenges that they are facing, encourage them to make the leap through to higher levels of participation in citizen science work (McKenna and Green, 2002).

Although the use of cohesive words (such as us, we, everyone, etc.) is less evident in the Zooniverse platform, use of other features of social presence may make participants feel that they belong to the community. It may develop a sense that they are working towards contributing to the scientific developments through engaging in the citizen science projects (Jackson, et. al. 2020).

Conclusion

We found a substantial degree of social presence in the comments of the workers of the Gravity Spy project, the majority of which was interactive in nature. The workers showed social presence most commonly by asking questions, expressing emotion and disclosing information about themselves. There was a significant degree of correlation between the 3 categories of social presence (affective, interactive and cohesive) showing that the posts display social presence in multiple and in interconnected ways.

For further analysis, we can extend the data from the Zooniverse platform to other citizen science projects to explore the nature of social presence across projects. Moreover, interviewing the volunteers about the nature of their social presence could help us explain why particular individuals interact in certain ways or capitalize more on one social function than another. Furthermore, we can look into how use of certain features by volunteer peers, guide the way they interact with each other. Additionally, it would also be important to describe the other two theoretical constructs of the community of inquiry framework (cognitive and teaching presence) and explore the association between the presence of three categories.

References

Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. The internet and higher education, 2(2-3), 87-105.

Akcaoglu, M., & Lee, E. (2016). Increasing social presence in online learning through small group discussions. The international review of research in open and distributed learning, 17(3).

Akcaoglu, M., & Lee, E. (2018). Using Facebook groups to support social presence in online learning. Distance Education, 39(3), 334-352.

Curtis, V. (2015). 'Online citizen science projects: an exploration of motivation, contribution and participation'. Ph.D. thesis. The Open University. URL: http://oro.open.ac.uk/42239/.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. The internet and higher education, 2(2-3), 87-105.

Heldal, I., Schroeder, R., Steed, A., Axelsson, A. S., Spant, M., & Widestrom, J. (2005, March). Immersiveness and symmetry in copresent scenarios. In IEEE Proceedings. VR 2005. Virtual Reality, 2005. (pp. 171-178). IEEE. Garrison, D. R.,

Jennett, J., Kloetzer, L., Schneider, D., Iacovides, I., Cox, A. L., Gold, M., Fuchs, B., Eveleigh, A., Mathieu, K., Ajani, Z. and Talsi, Y. (2016). 'Motivations, learning and creativity in online citizen science'. JCOM 15 (03), A05.

Lowenthal, P. R., & Dunlap, J. C. (2020). Social presence and online discussions: a mixed method investigation. Distance Education, 41(4), 490-514.

McKenna, K.Y.A., & Green, A.S. (2002). Virtual group dynamics. Group Dynamics 6:116–127.

Price, C. A. and Lee, H.-S. (2013). 'Changes in participants' scientific attitudes and epistemological beliefs during an astronomical citizen science project'. Journal of Research in Science Teaching 50 (7), pp. 773–801. DOI: 10.1002/tea.21090.

Sternbergh, A. (Nov 17, 2014). Smile, you're speaking emoji: The rapid evolution of a wordless tongue. New York Magazine. Retrieved from https://nymag.com/intelligencer/2014/11/emojis-rapid-evolution.html.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. The Journal of Distance Education/Revue de l'ducation Distance, 14(2), 50-71.

Rains, S. A., Akers, C., Pavlich, C. A., Tsetsi, E., & Appelbaum, M. (2019). Examining the quality of social support messages produced face-to-face and in computer-mediated communication: The effects of hyperpersonal communication. Communication Monographs, 86(3), 271-291. doi:10.1080/03637751.2019.1595076.

Traphagan, T. W., Chiang, Y. H. V., Chang, H. M., Wattanawaha, B., Lee, H., Mayrath, M. C., ... & Resta, P. E. (2010). Cognitive, social and teaching presence in a virtual world and a text chat. Computers & Education, 55(3), 923-936.

Yamada, M., Kaneko, K., & Goda, Y. (2016, September). Social presence visualizer: Development of the collaboration facilitation module on CSCL. In International Conference on Collaboration Technologies (pp. 174-189). Springer, Singapore.

Appendices

Category	Indicators	Definition	Example
Affective	Expression of emotions	Conventional expression of emotions or unconventional expression of emotion, includes repetitious punctuation, conspicuous, capitalization, emoticons	"I just can't stand it when….!!!" "ANYBODY OUT THERE!"
	Use of humor	Teasing, cajoling, irony, understatement, sarcasm	The banana crop in Edmonton is looking good this year
	Self-disclosure	Presents details of life outside of class, or expresses vulnerability; Disclosing the work, background, or individual identity characteristics to make a point	"Where I work, this is what we do…." "I just don't understand this question"
Interactive	Continuing a thread	Use reply feature of software, rather than starting a new thread	Software dependent, e.g. "Subject: Re" or "Branch form"
	Referring explicitly to others' messages	Using software features to quote others entire message (such as link or @) Cutting and pasting selection of others' messages, using quotation marks (active voice) Direct references to contents of others' posts (passive voice)	Software dependent, e.g. "Martha write:" or text prefaced by less-than symbol </ "In your message you talked about Moore's distinction between…. "
	Asking questions	Students ask questions of other students or the moderator	"anyone else had experience with WEBCT?"
	Complimenting, expressing, appreciation	Complimenting others or content of others' messages	"I really like your interpretation of the reading"
	Expressing agreement or disagreement	Expressing agreements with others or content of others' messages	"I was thinking the same thing. You really hit the nail on the head."
Cohesive	Vocatives	Addressing or referring to participants by name	"I think John made a good point." "John, what do you think?"

	Addresses or refers to the group using inclusive pronouns	Addresses the group as we, us, our, group	"Our textbook refers to…" "I think we veered off track"
	Phatics, salutations	Communication that serves a social function; greetings, closures	"Hi All", "that's it for now" "We're having the most beautiful weather here"

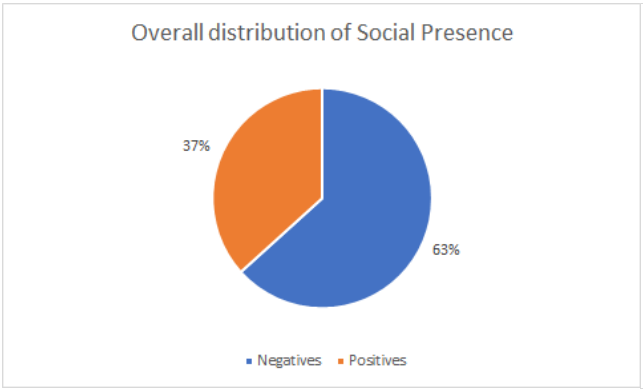
Table 1: Table showing categories and the sub-categories with their definitions and examples

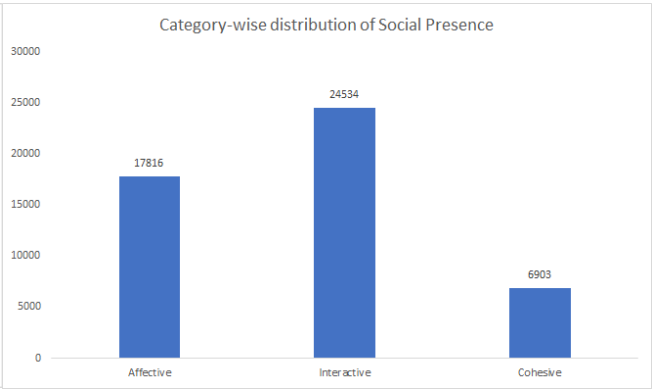
Category	Classifier
Expression of emotions	Pre-trained BERT model fine-tuned on the *Twitter US Airline* data (https://www.kaggle.com/crowdflower/twitter-airline-sentiment) for detecting the presence of sentiment or emotion + Presence of emojis
Use of humor	Pre-trained BERT model fine-tuned on the *200k Short Texts for Humor Detection* data (https://www.kaggle.com/moradnejad/200k-short-texts-for-humor-detection) for detecting the presence of sentiment or emotion + Presence of humor indicating emojis
Self-disclosure	Pre-trained BERT model fine-tuned on the collected sample
Continuing a thread	Use the scraped data to check if the comment is a reply or not
Referring explicitly to others' messages	Detect presence of certain URLs present in comment which are links to other messages
Asking questions	Pre-trained BERT model fine-tuned on the collected sample
Complimenting, expressing appreciation or critique	Pre-trained BERT model fine-tuned on the collected sample
Expressing agreement or disagreement	Detect presence of words and phrases indicating agreement or disagreement from a predefined list
Vocatives	Detect presence of user mentions (user handles beginning with '@')
Addresses or refers to the group using inclusive pronouns	Detect presence of group inclusive pronouns from a predefined list
Phatics, salutation	Detect presence of salutation words from a predefined list

Table 2: Descriptions of the classifiers for each category

Sub-categories	Krippendorff's alpha	Precision	Recall	F1
Expression of emotions	0.7129	0.66	0.75	0.68
Use of humor	0.6870	0.64	0.63	0.64
Self-disclosure	0.6810	0.47	0.7	0.56
Asking questions	0.8021	0.73	0.79	0.76
Complimenting, expressing appreciation or criticism	0.7225	0.58	0.78	0.67

Table 3: Table showing the intercoder agreements and the accuracy values for the categories where ML models were used





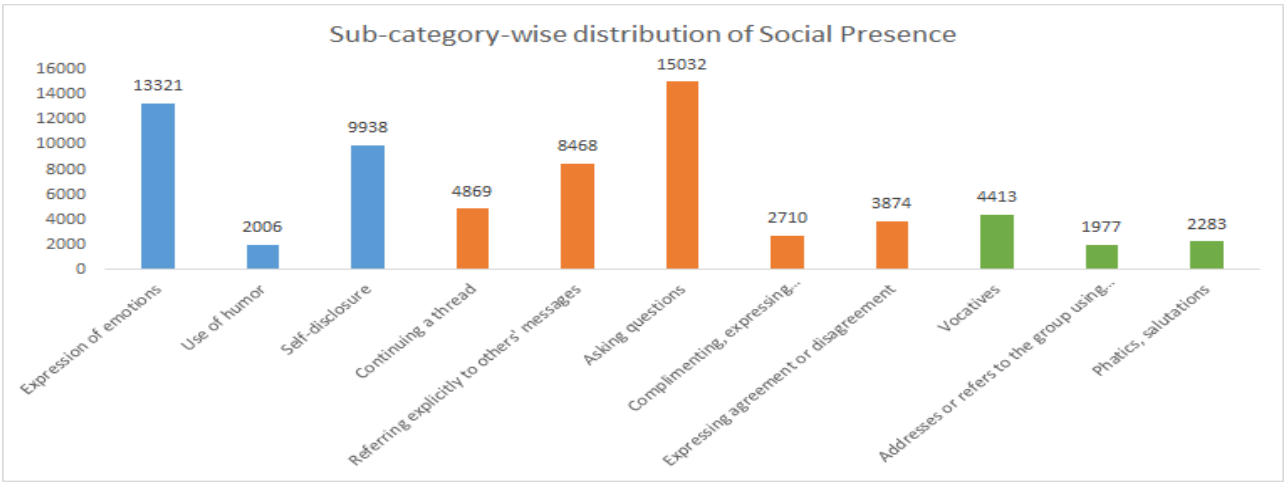


Figure 1 (top left) : Pie-chart showing the distribution of Social Presence in all the comments; Figure 2 (top right) : Bar plot showing the distribution of the categories of Social Presence in all the comments; Figure 3 (bottom) : Bar plot showing the distribution of the sub-categories of Social Presence in all the comments.

		Interactive	
		Absent	Present
Affective	Absent	87.0%	62.2%
	Present	13.0%	37.8%
$\chi 2(1, N=90501) = 6951.10, p < 0.001$			

Table 4: Chi-square crosstabulation of affective and interactive

		Cohesive	
		Absent	Present
Affective	Absent	84.7%	27.5%
	Present	15.3%	72.5%
$\chi 2(1, N= 90501) =13179.07, p < 0.001$			

Table 5: Chi-square crosstabulation of affective and cohesive

		Cohesive	
		Absent	Present
Interactive	Absent	78.2%	8.2%
	Present	21.8%	91.8%
$\chi 2(1, N= 90501) = 15834.23, p < 0.001$			

Table 6: Chi-square crosstabulation of interactive and cohesive