9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

"I Heard It on the (Silk) Road": Structural and Cultural Pathways to Novelty Introduction in an Online Community

Keywords: online communities, novelty, innovation, networks, text mining

Extended Abstract

Nowadays, online communities pioneer a wide range of highly relevant novelties affecting our physical, as well as digital, lives. For instance, the now widely used word "LOL" originated in an online posting board in the mid 1980-s; other well-known expressions, such as "FAQ" and "selfie" have a similar origin. Despite the importance of innovations generated in online communities, our knowledge of the features of individuals who originally introduce these novelties remains limited. Taking stock of a diverse, burgeoning set of scholarly traditions, we focus on users' structural and cultural embeddedness [1] within their community as key antecedents of the introduction of an innovation into the community itself. We study this phenomenon in a particular context: the Silk Road forum, an online community complementing a pioneering illegal marketplace of the same name situated on the Dark Web [2].

Because we focus on online communities that build their norms of expression and store shared knowledge through text, we operationalize novelty introduction as lexical innovation – the introduction of new words into the community. We then proceed to survey how users' *structural embeddedness* – their position within the community network – affects their innovative potential. More importantly, we explore users' boundary-spanning roles – that is, brokerage between different sub-communities – and especially how these roles relate to the diversity of knowledge shared in these sub-communities. Finally, we evaluate users' *cultural embeddedness* — the extent to which they are compliant with community norms of linguistic expression – as an antecedent of their innovative potential.

We consider a word to be an innovation if it appears in a time period t without having been used in any of the previous periods [3]. Exploiting the fact that users in the Silk Road communities were linked by their posting across forum threads, we compute measures of network centrality and clustering by constructing a bipartite network in which one category of nodes are users, and the other are forum threads [4]. To measure user' access to diverse types of knowledge shared in different communities, we first build an LDA (Latent Dirichlet Allocation) model [5] of all forum texts and obtain distributions of LDA topics across forum threads written in various communities [6]. To obtain a measure of the diversity of topics a user has had access to (also: *thread similarity*) we calculate the average pairwise cosine similarity of the vectors of topic distributions of all the threads in which they have posted. Finally, to operationalize the extent to which a user was culturally embedded within community norms of linguistic expression, we build a word2vec model [7] of the community language on a sub sample of texts and calculate the average log likelihood of the user's posts given the community language model [8].

Exploiting both network techniques and advanced text mining methods on a rich dataset of textual interactions in an online forum allows us to advance our understanding of creative processes and sociolinguistic innovation in online communities. We derive two main findings

from our analysis: first, we find that the effect of being structurally embedded in the community is moderated by diversity of the content one is exposed to (See Figure 1). This means that individuals who occupy a less embedded position in the network (like brokers), benefit from observing different content - in line with the so-called "recombination" hypothesis - but more deeply embedded individuals benefit from seeing similar content - a mechanism of "deepening". Second, we show that being culturally embedded in this community has an inverse U-shape effect over the likelihood of introducing a novelty. That is, our results provide evidence for the tension between the norm-following that is required in order to introduce innovations that are meaningful for the community, and the norm-breaking that is needed to be creative and "think outside of the box". Finally, we introduce two methodologically innovative measures: first, we use word-embedding models to capture previously unexplored nuance and depth of language and thus characterize adherence to community cultural norms; second, we use LDA models to operationalize the diversity of content to which one is exposed – a sound and rich approach that can be applied in future research to further understand how network structure and the content traversing it intermesh.

References

[1] S. Zukin and P. DiMaggio, *Structures of Capital: The Social Organization of Economic Life*. S. Zukin and P. DiMaggio, Eds. New York, NY, Melbourne, Australia: Cambridge University Press., 1990.

[2] G. Branwen, *Darknet Market Archives (2013-2015)*, 2021. [Online]. Available: `https://www.gwern.net/DNM-archives`.

[3] C. Danescu-Niculescu-Mizil, R. West, D. Jurafsky, J. Leskovec, and C. Potts, "No country for old members," in *Proceedings of the 22nd international conference on World Wide Web - WWW '13*, New York, NY: ACM Press, 2013, pp. 307–318. DOI: `10.1145/2488388.2488416`.

[4] M. Latapy, C. Magnien, and N. D. Vecchio, "Basic notions for the analysis of large two-mode networks," en, *Social Networks*, vol. 30, no. 1, pp. 31–48, Jan. 2008. DOI: `10.1016/j.socnet.2007.04.006`. (visited on 12/05/2022).

[5] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012. DOI: `10.1145/2133806.2133826`.

[6] A. T. J. Barron, J. Huang, R. L. Spang, and S. DeDeo, "Individuals, institutions, and innovation in the debates of the French Revolution," en, *Proceedings of the National Academy of Sciences*, vol. 115, no. 18, pp. 4607–4612, 2018. DOI: `10.1073/pnas.1717729115`.

[7] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *ICLR Workshop Papers.*, arXiv: 1301.3781, 2013. [Online]. Available: `http://arxiv.org/abs/1301.3781`.

[8] M. Taddy, "Document Classification by Inversion of Distributed Language Representations," en, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Beijing, China: Association for Computational Linguistics, 2015, pp. 45–49. DOI: `10.3115/v1/P15-2008`.

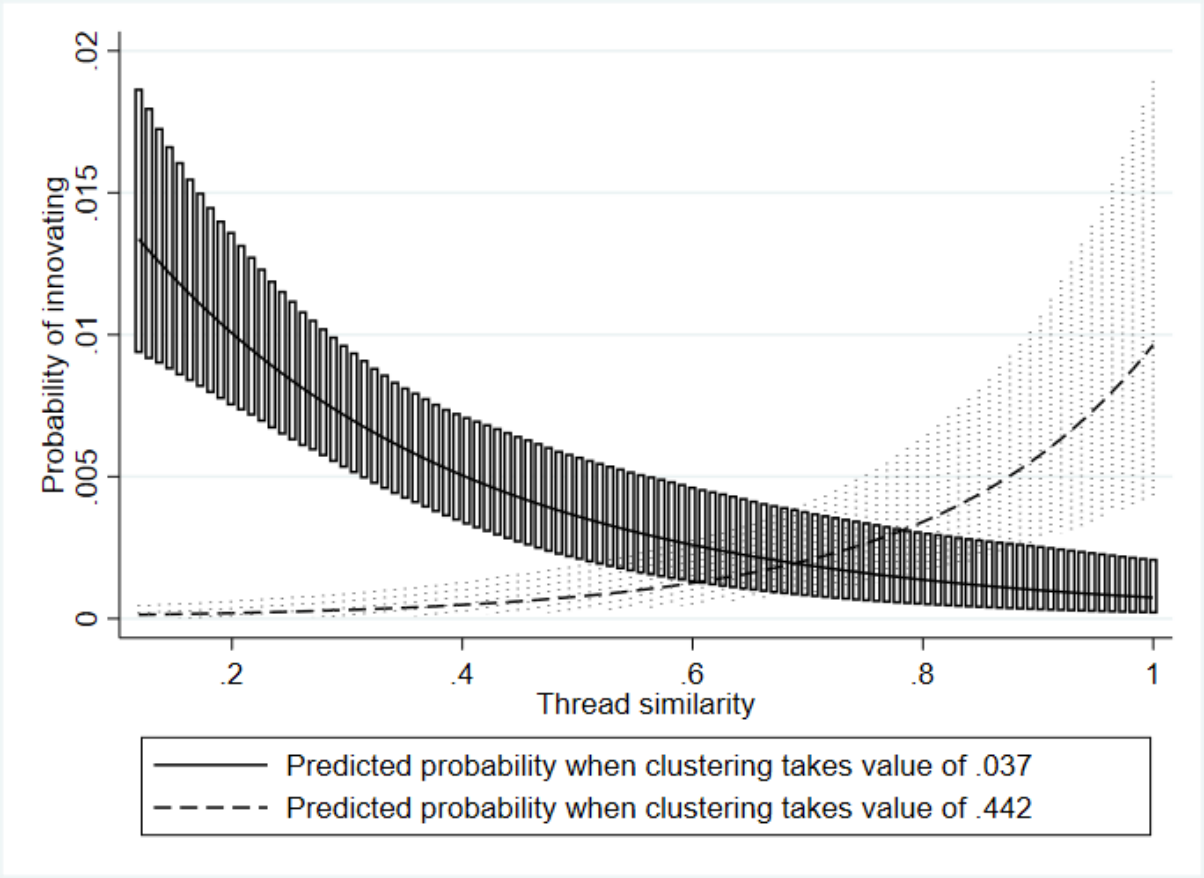

Figure 1: **Interaction with thread similarity at high and low values of clustering.** In this figure we plot the interaction effect with thread similarity (lower values indicate access to more diverse knowledge) at high (95th percentile) and low (5th percentile) values of clustering, corresponding, respectively, to highly clustered individuals and brokers. The figure shows that individuals with low clustering values (solid line) are more likely to innovate if they post across dissimilar threads and, furthermore, that the less similar the threads they post in, the better. On the contrary, individuals who are highly clustered (dashed line) benefit, in terms of innovativeness, by posting in similar threads. In fact, for high levels of thread similarity, clustered individuals become more likely to introduce an innovation than brokers.