

Meta-commentary This document is meant to serve as an example for future proposal documents (our re-branding of "critiques").

In a nutshell, the problem and proposed data/approach, and how it relates to the reading We propose identifying texts that *engage* users, inspired by Mishne and Glance's discussion of how comments indicate user engagement. Concretely, we propose predicting which posts on Slashdot generate many direct replies, or serve as the root of large subtrees. These "structural" (as opposed to textual features) can serve as gold-standard labels, and thus obviate the need to do hand annotation. Moreover, it has been proposed that various structural properties ("fatness", "depth", or "h-index" of comment trees) indicate an engaging, fruitful debate, with possible implications for theories in political science [citations omitted].

Applications One potential application of a successful system for identifying thought-provoking texts is highlighting such content for users, as a means of overcoming information overload (and rewarding the creators of such content). A system that identifies texts that are merely provocative in the sense of goading or irritating others could also be used for filtering such content, or perhaps warning the authors that they might want to edit their language (perhaps there are cases where the author is unaware that their text is inflammatory).

Texts that are engaging could provide an opportunity for studying how people express agreement and disagreement [citation], which seems like another possible project direction.

Further discussion A distinction that does seem important to make is the difference from engaging *topics*. True, finding "hot-button", controversial topics may be interesting, and some have approached this problem [citation omitted], but identifying such topics might devolve into simple lexicon lookup or just looking at activity rates. (No guarantees this assertion is true, though!). At any rate, we propose looking at texts that are engaging while controlling for topic.

To control for topic and get useful negative (i.e., less-provocative posts), it might be useful to compare posts within the same thread, since they will be on the same topic, more or less.

Now, large fan-out could be due to "thought-provokingness" or "irritatingness", as mentioned above. It might be useful to try studying this distinction, trying to decide whether commenters are trying to persuade others, or just express their opinions; one could perhaps make some headway into this by looking at if/when the commenters revisit the site, if such timestamps and information were available (a bit problematic to interpret "negative" feedback here), or try to look at text similarity to see if the poster is at least referring to the issue at hand. Another idea that came up after class was to consider posts that generate large fan-out with many of the children being marked as "Interesting" — this might help distinguish thought-provoking from goading posts.

We acknowledge that restricting our consideration to something like Slashdot could be problematic; for instance, on review sites, disagreement might take a rather different form.