

----- REVIEW 1 -----

PAPER: 2

TITLE: Analyzing the most important part of the World Wide Web

AUTHORS: Ionut Trestian, Chunjing Xiao and Aleksandar Kuzmanovic

OVERALL EVALUATION: 1 (weak accept)

----- REVIEW -----

This paper is very interesting and does confront a topic which has a great deal of data, however difficult, which would constitute a good web observatory. The paper overall fits well into the theme of Web Observatories as it focuses on looking at how people interact with the Web and their behaviors.

There are, however, a few things that could be improved or expanded upon which would strengthen this work. The general grammar of the paper could be improved and tightened up in the first instance. Some passages, especially the abstract, are very difficult to understand as a result of missing punctuation and detract from the paper considerably. However, these issues can be resolved with better copy editing.

Although the methods for collecting are described in the footnotes, I think, given the nature of the data being collected, more needs to be said in the paper itself about how the privacy of users was addressed in this project.

The authors need to expand their analysis of the data as well to delve more into what behaviors relating to the pornographic sites might mean. To say that we found that people do X, and therefore X is what people do does not go far enough in explaining what the behaviors mean in a sociological context.

----- REVIEW 2 -----

PAPER: 2

TITLE: Analyzing the most important part of the World Wide Web

AUTHORS: Ionut Trestian, Chunjing Xiao and Aleksandar Kuzmanovic

OVERALL EVALUATION: 1 (weak accept)

----- REVIEW -----

Interesting investigation, well written paper. A comparison with youtube in sections 5 and 6 would have been interesting, though, similar to the comparison of search behavior in section 3.

----- REVIEW 3 -----

PAPER: 2

TITLE: Analyzing the most important part of the World Wide Web

AUTHORS: Ionut Trestian, Chunjing Xiao and Aleksandar Kuzmanovic

OVERALL EVALUATION: 0 (borderline paper)

----- REVIEW -----

This paper makes too many leaps in its methodology and analysis. I want to see many more details if such rather loose words are to be tied to the figures. Some examples of looseness:

- (end of section 3): jump from greater expressed precision in query to this being a user need rather than a poor performance from the {search,recommender} engine.
- (s4) on what basis do you know the site is popular in Singapore etc.? Given that you claim significance in regional/cultural for this later.
- being much more careful about asserting links between e.g. length or fidelity of video and the content without giving a method for how the content was sampled.
- explanation of ratings by "cultural difference" isn't explained, just stated (section 6.1 -- and it's not clear what hypothesis is being investigated for ratings).
- given early statement of purpose in comparison to non-porn sites there is no comparison in the rating and sentiment analysis sections.
- poor (to non-existent) comparison with the cited work in section 7.

Recommend acceptance as a short paper on the condition that:

- 1) The authors clearly set out why the porn observatory is relevant and important to the wider Web Observatory community? Does the dataset illustrate weaknesses or strengths in the tools used, or patterns that have not been observed elsewhere? What lessons are learnt -- in terms of Web Observatories -- from studying this data?
- 2) Keep the analysis strictly to the data -- there is plenty to report in terms of the structure, linkage, access stats of the data etc. Avoid layering weak interpretation about behaviour, content, culture, location etc. over these -- it's fine for this to be future work applying rigorous methodologies to explain patterns in the data (indeed this would be good, interesting, work).
- 3) Needs a good pass to tidy up language, punctuation, etc.

Other comments:

- I question the opening assertion of "important" vs. perhaps "popular".
- "fascinating cultural differences"
- [4] is weak. no attribution on linked website I could see. Cite needs attaching to all previous sentences?
- implicit contradiction between "tremendous [...] hundreds of research papers" and the opening "few studies". Somewhat clarified in next para.
- cite required for "user behaviour related to search engines" (s3)
- section 3 needs to make much better reference to the figures/tables.

Some other comments:

- wayward comma 2nd sentence opening intro para
- porn sites should be footnotes rather than cites
- "not so comprising"