Exploring the Influence of Reddit Mechanics on Content Popularity Factors and the Counterinfluence of User Activity on Content Popularity Factors and Reddit Mechanics

A Research Paper Submitted to the Department of Engineering and Society
Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia – Charlottesville, Virginia

In the Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Siddharth Nanda

Fall, 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature _____ Date: _November 20, 2020_

Siddharth Nanda

Approved _____ Date: 11/30/20

Richard Jacques, Associate Professor of STS, Department of Engineering and Society

Introduction

The internet greatly expanded the ability for individuals to communicate and forever changed the mechanics by which people both interact and view interactions. Social media websites, websites that enable online communication of information, ideas, and other content through virtual communities (Merriam-Webster, n.d.) (Alexa Internet, Inc., n.d.), have now become the primary method by which individuals interact with each other and consume information on the internet. Analysis of these websites from a technical and sociological perspective could help provide insight into the evolution of the mechanics of interactions between individuals and perhaps offer a view of the next stage of these mechanics. This paper investigates the influence of the mechanics of a social media website, Reddit, on the content popularity factors of the website and the counterinfluence of user activity on content popularity factors and the website's interaction mechanics; answering the question: "How do Reddit's mechanics cause certain factors to be more important than others in creating popular content on Reddit and how does activity on Reddit influence those same factors and mechanics?".

The results of the corresponding technical research project, which was focused on predicting the popularity of content on Reddit, will be used to support conjectures made through two logical STS frameworks. The first of these two frameworks is "Configuration and Script" (N.E.J. Oudshoorn, 2003): which focuses on the process of defining the users, setting constraints on their actions, and how the design of objects conceptualizes a method of their use. This framework will be used to understand the usage of the website, how its creators intended to have users interact on the website, the intended usage of site mechanics, and how users ultimately interpret the design to enable their unique forms of usage.

The other framework, Actor-Network Theory, will be used to form an understanding of users and website mechanics as "actors" within a "network" where actors (both living and non-living) can exert influence on each other. This framework can be used to understand elements of Reddit culture and how Reddit users affect each other and the development of the site itself. Both frameworks will help contextualize results presented in the technical research to ultimately answer the STS question. Before attempting to answer the question or analyze the results of previous research, it's important to understand Reddit and its mechanics.

Understanding Reddit

Reddit is the 6th most popular website in the US (Alexa Internet, Inc., n.d.) and a social media website that has somewhat of an unconventional structure compared to its popular peers. Founded in 2005 by UVA students Steve Huffman and Alexis Ohanian, Reddit focuses user interactions around user submitted content and has a variety of mechanics for users to share thoughts and opinions on content. Content such as links, text, or video can be shared through two means: the creation of an individual discussion space for the content ("thread"), or a reply to some existing content ("comment"). Virtual communities are organized through self-selecting groups ("subreddits"), which a user can join or leave at will, and subreddits serve as a collection of threads of a related topic. Users are heavily encouraged not only to interact with content via creation, but also through voting: indicating their responses to posted content (both threads and comments) simply by clicking an up-arrow (indicating a positive response) or down-arrow (negative response) available on all content. Reddit displays the cumulative value of all upvotes and downvotes as a score on every piece of content; it also presents various methods for the user to peruse content based on time and score (top scored posts over the past day/week/month, rising content, new content). For content that elicits a strong reaction from a user, Reddit offers the

option for users to purchase "awards" for content which are then displayed prominently. The pricing of these awards varies greatly: and the more expensive the award given, the more benefits the recipient of the award enjoys. Reddit is also unique in that users are anonymous and are encouraged to use a username rather than their real names unlike a social media website like Facebook or Twitter. However, this does not prevent the rise of "influencers" (users which generate popular content with far above average level of community interaction) thanks to a mechanic known as karma. Karma is a user metric depending on the overall popularity of a user's content. Reddit's communities are moderated by users themselves and each subreddit can have unique rules and restrictions as determined by its moderators (for example, requiring a certain amount of karma generated from comments in the subreddit to create new posts). The typical Reddit user can choose to interact with others simply by observing content they're interested in (browsing a subreddit, its threads, and the comments contained within those threads, and voting on that content), by actively interacting within a community (joining a subreddit, posting threads to it, and commenting on threads), or through a combination of the described observation/interaction mechanics. While the individual mechanics of the website may seem simple to understand, Reddit has a reputation of being difficult to interact with due to: the interaction between its mechanics (such as subreddits, threads, comments, upvotes, awards, and karma), its diverse subcultures, and user generated exclusivity via strong voting mechanics.

The Technical Research

In the technical research project conducted alongside this STS research, "Predicting Comment Popularity within Reddit Communities through Text and Metadata based Multiclass Classification Models", machine learning classifiers were built and utilized to predict content popularity and identify content popularity factors. Comment data was gathered from a certain

number of threads on selected subreddits using an original, extensible data-collection tool built around the Python Reddit API Wrapper. The data collected included: comment text, the time a comment was created, the time difference in when the comment was created and when the comment or thread it was replying to was created (time difference from parent comment), how far down a comment was in a chain of comments (the depth level of the comment), the score of the comment, and the number of awards the comment had received. Certain characteristics of content were derived from the collected data including: the comment length in words, the comment length in characters, and the AFINN sentiment score of the comment (how positive or negative of a tone the comment took). Machine Learning methodologies (essentially automated statistical methodologies) were used to create various classification models from the characteristics of the data collected that could be used to classify data into score groups (created by scaling data then stratifying them into popularity categories). Models created solely from either the text data itself or solely from metadata (such as length, time difference, depth, awards, etc.) both demonstrated acceptably high, as well as comparable, accuracy in the classification tasks (predicting the score group the input data belonged to). Analysis of feature importance (content popularity factors) saw strong importance of the time difference in when the comment was created and when the comment or thread it was replying to was created, comment length (both character and word), and sentiment score of the comment (how positive or negative of a tone the comment took).

Contextualizing Results via Configuration and Script

As stated in the introduction, "Configuration and Script" is an STS framework that focuses on the process of defining the users, setting constraints on their actions, and how the design of objects conceptualizes a method of their use. Interpreting the results of the technical

research project under this framework can help build understanding of the intended usage of Reddit via it's unique mechanics (configuration) and how users interpret the provided usage mechanics (script): answering the first part of the research question "How do Reddit's mechanics cause certain factors to be more important than others in creating popular content on Reddit". In the section above which describes Reddit, several mechanics and usage paths were detailed: the "intended configuration" of the website's designers. The technical research project can be viewed as a study in how the mechanics (commenting, subreddits) influence user behavior (voting, commenting). It's possible to assert that the commenting mechanic, which produces far more original content on Reddit than the posting of threads, influences content popularity factors due to user variation in commenting (execution of "scripts") showing a strong correlation with voting. The content popularity factors, features, on which the score classifiers were built all illustrate different scripts that the users followed: with several factors deriving from variation in user usage of the commenting mechanic (time difference in replies, sentiment). With this in mind, it's possible to assert that *repeated usage of popular "scripts" defines a culture*. This also illustrate how Reddit's mechanics are not entirely responsible for the creation of popular content on Reddit as other mechanics such as giving awards had much lower feature importance. This leads to the second half of the STS question, "how does activity on Reddit influence those same factors and mechanics?", which can be better analyzed via the Actor-Network Theory framework.

Contextualizing Results via Actor-Network Theory

As stated in the introduction, Actor-Network Theory is a framework that can be used in this context to represent users and website mechanics as "actors" within a "network" where actors (both living and non-living) can exert influence on each other. Interpreting the results of

the technical research project under this framework can help build understanding of how "actors" (users) affect each other (users and mechanics) within the network (Reddit) and help answer the second half of the STS question: "how does activity on Reddit influence those same factors and mechanics?". The results of the technical research project can also provide commentary on how users have influenced content popularity factors through their influence on actors (mechanics). For example, the karma mechanic serves as a core motivation for users creating content and incentivizes quick commenting due to higher levels of interaction with the comments section earlier in a thread's lifetime (time difference factor). The low level of importance of giving awards to content popularity might've seen the introduction of new, cheaper award options to incentives spending on Reddit. With sentiment seeing high feature popularity, moderators can choose to disincentivize negative or positive content depending on a subreddit's culture to drive higher interaction. These assertions allow for a substantiative conclusion: activity on Reddit can influence content popularity and Reddit mechanics simply by illustrating popularity and driving interaction, a major goal of social media. With this in mind, it's possible to assert that *when networks share a goal, actors influence each other to assert a culture.*

Conclusion

The technical research project conducted alongside this STS research, "Predicting Comment Popularity within Reddit Communities through Text and Metadata based Multiclass Classification Models", was integral in helping answer the question of "How do Reddit's mechanics cause certain factors to be more important than others in creating popular content on Reddit and how does activity on Reddit influence those same factors and mechanics?". Contextualizing the results of the project via Configuration and Script allowed for additional understanding into how variation of user interaction with mechanics result in certain content

popularity factors to be more important than others; while contextualizing via Actor-Network Theory helped reveal how activity on Reddit can influence content popularity and Reddit mechanics simply by illustrating popularity and driving interaction. This research has value due to the proliferation of online hive minds in recent years, such as online far-right groups. Online hive minds are dangerous and cause real world chaos as online popularity translates to increasingly strongly held real-world beliefs. This research not only helped identify popularity factors (the technical project), but analyzed their development and breaking down hive minds will require research such as this alongside research that aims to understand mechanics to disincentivize the formation of these dangerous communities and the externalities of those proposed mechanics.

References

Alexa Internet, Inc. (n.d.). *Top Sites in United States*. Retrieved from Alexa:

> https://www.alexa.com/topsites/countries/US

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In

> W. E. Law, *Shaping Technology/Building Society: Studies in Sociotechnical Change* (pp.

> 225-258). Cambridge, Massachusetts : MIT Press.

Merriam-Webster. (n.d.). *social media*. Retrieved from Merriam-Webster: https://www.merriam-

> webster.com/dictionary/social%20media

N.E.J. Oudshoorn, T. P. (2003). Introduction: How users and non-users matter. In T. P. N.E.J.

> Oudshoorn, *How users matter. The co-construction of users and technology* (pp. 1-25).

> Cambridge, Massachusetts: MIT Press.