

Science Forums

Cathy Su

19/10/2019

Executive Summary

Moderators at online forums are always interested in growing the participation while keeping a high quality discussion. Therefore, at times it becomes necessary to delete posts that might become problematic. The purpose of this study was to assess the variables that may influence which topics need to be closed on this online forum. It's been hypothesized that the type of post authors, topic of the post and XXX are factors in which discussions will need to be shut down. Discussion topics were randomly sampled from ScienceForums.Net (SFN). XXX studies passed the inclusion criteria and XXX variables represented potential contributing factors towards whether discussions will need to be closed. XXX proved to be the most consistent moderator of I-PA, suggesting that much of the discordance may be from motivational flux between initial intention and eventual behaviour. Anticipated regret and conscientiousness also had evidence as the moderators of I-PA. Perceived control/self-efficacy, planning, extraversion, habit and environmental proximity to recreation showed some evidence for moderation, while gender, agreeableness, openness, body mass index and ethnicity did not appear to moderate I-PA. The findings demonstrate that traditional intention theories may need augmentation to better account for the evidence present in I-PA discordance.

Introduction

Causal diagram

Methods

removal of outliers

Handling missing data

Calculation of additional variables

For the substantive questions we calculated some new variables as follows: * We extracted the 'year_started' of each topic as replacement for using the full 'startdate' which contains a lot more precise time than we need. * the 'proportion_deleted' is the number of deleted divided by total posts in the topic. * The 'post_rate' is the number of posts divided by the 'duration' of the topic. If 'duration' is zero, the 'post_rate' was also assigned zero.

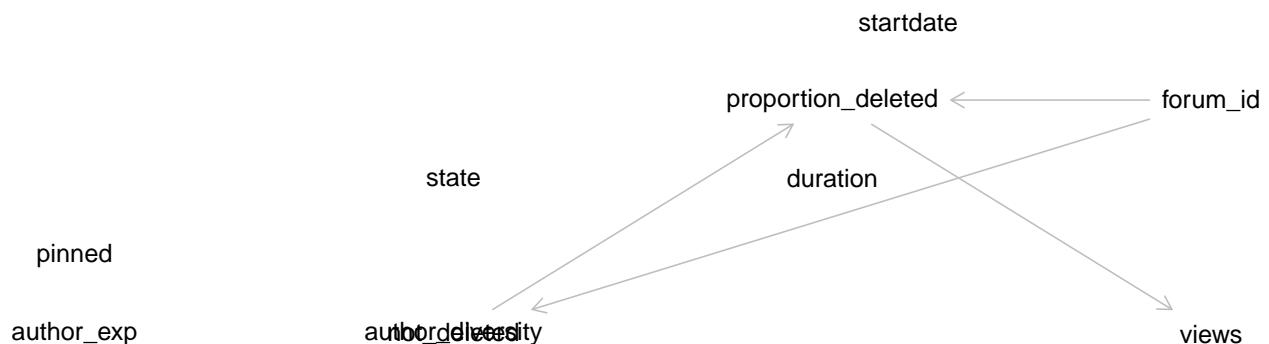


Figure 1: Causal diagram illustrates hypothesized relationships of experimental variables involved in relationship between testosterone and final group performance.

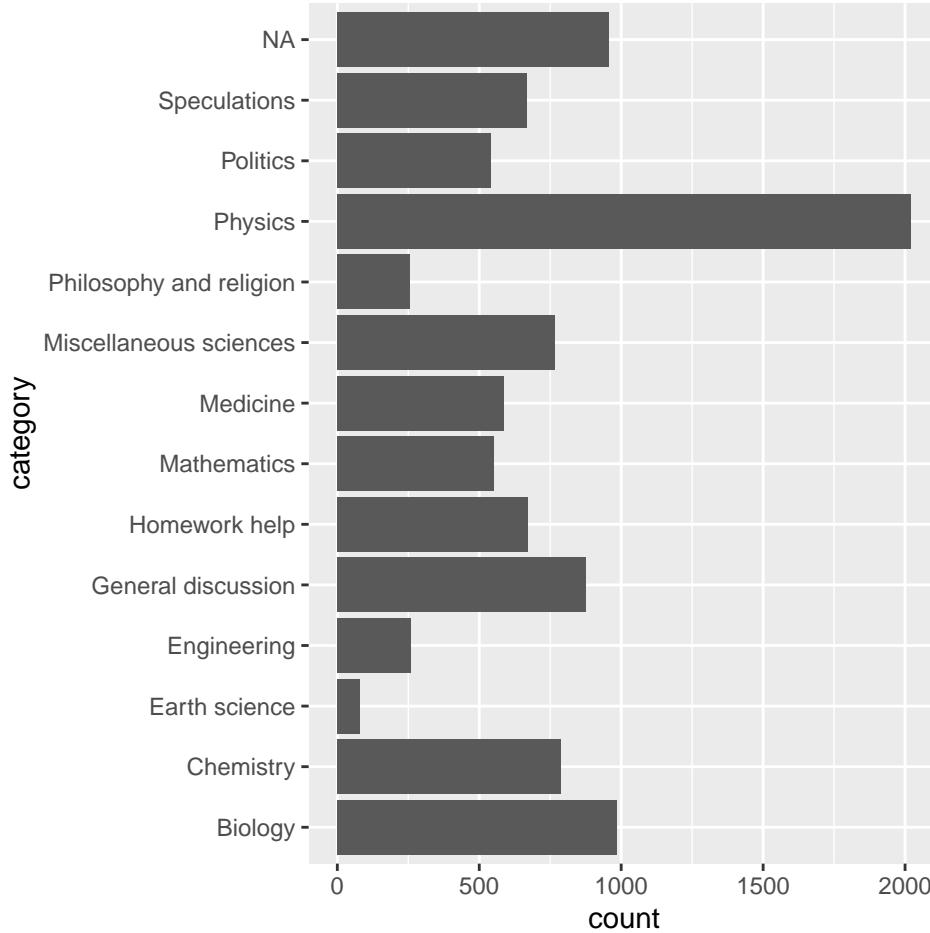


Figure 2: Breakdown of forums by topic, showing a portion of forums do not have any assigned topic.

Model building

Exploratory Data Analysis

Univariate variable distributions

Figure ?? shows the distribution of forum topics which fall into each category for the binary variables. This shows us that very few forum topics are deleted from view, closed, pinned or started by a banned author.



Additionally, although fo-

rum type is an important variable of interest, we found that there were many forum ids which were missing a categorization. This proportion is shown in the top bar of Figure 2.

We then examined the pairwise correlation of the continuous variables which is shown in Figure 3. There appears to be no collinearity and generally only weak correlation among these variables, but we can observe the following trends: * Looking across the diagonal, all of the continuous variables with the exception of ‘year_started’ have a significant right skew. * ‘posts’ and ‘views’ correlate positively with author diversity, author experience and topic duration but negatively with ‘year_started’, ‘proportion_deleted’ and ‘post_rate’.*

Q1. Relationship between views and posts

We saw in Figure 3 that views and posts are weakly positively correlated. In figure 4 we breaking this down by subforum for those topics which have a categorization.

Q2. Diverse discussions closed or deleted

Number of distinct authors who participated in the discussion.

Q3 discussions with proportionally more deleted posts tend to get closed more often

The univariate distributions of the group level variables is given across the diagonal in Figure 3. We saw that in particular, our diversity score appears bimodal. Although our score is calculated differently, (???) classified diversity score into two bins in their faultline analysis. This suggests that our diversity score is reasonable since it may also reflect some intrinsic bimodality present in the data.

We visualized the correlation matrix including the Pearson correlation coefficients (upper right half) between important variables in Figure 3. Based on the correlation coefficients, we do not need to remove variables based on collinearity. Additionally, it seems that

Right away we can make the following observations about the key variables:

- performance appears correlated with proportion of females and testosterone.
- testosterone appears correlated with cortisol, average age, proportion of females, time of day, performance and team size.
- diversity score appears correlated with team size.

Based on this we knew that in addition to final.performance, avg.log.testosterone, avg.log.cortisol and diversity.score we should consider whether to incorporate the four additional variables proportion.female, avg.age/age.variance, time.of.day, and team.size in our models.

Results

The results discussed by the original study (???) include that:

- considered in isolation, group diversity and testosterone are not significantly correlated with performance.
- when group diversity was 1

Model selection

Since the authors studied 2-way interactions, to choose the terms in the model we first performed model selection using best subsets and cross validation. We start off with all of the group level variables as depicted in Figure 3, adding based on the causal graph and the authors’ work the following interaction terms:

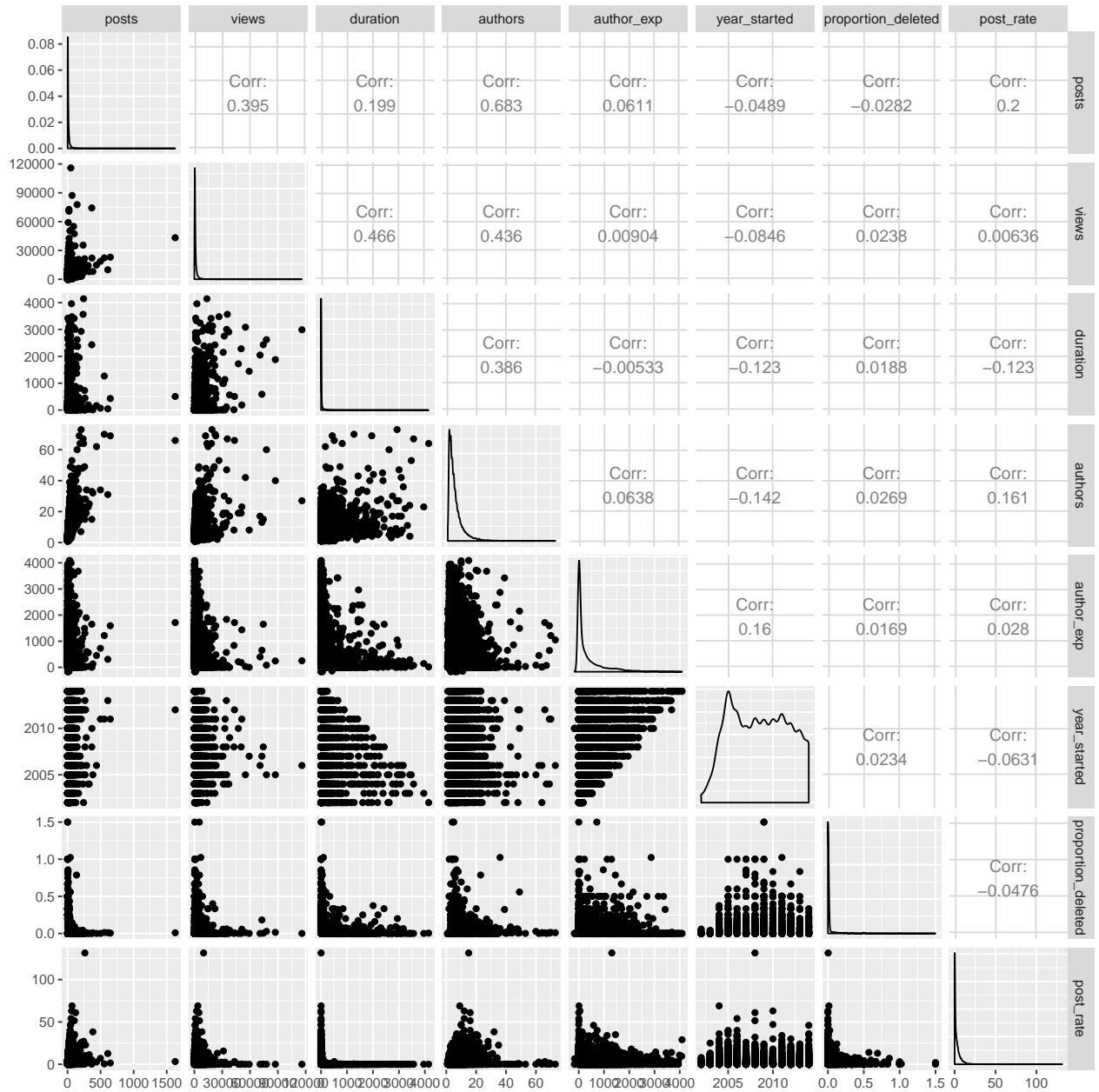


Figure 3: Pairwise correlations of important variables including their Pearson correlation coefficient. Significant correlations are marked by the corresponding number of asterisks.

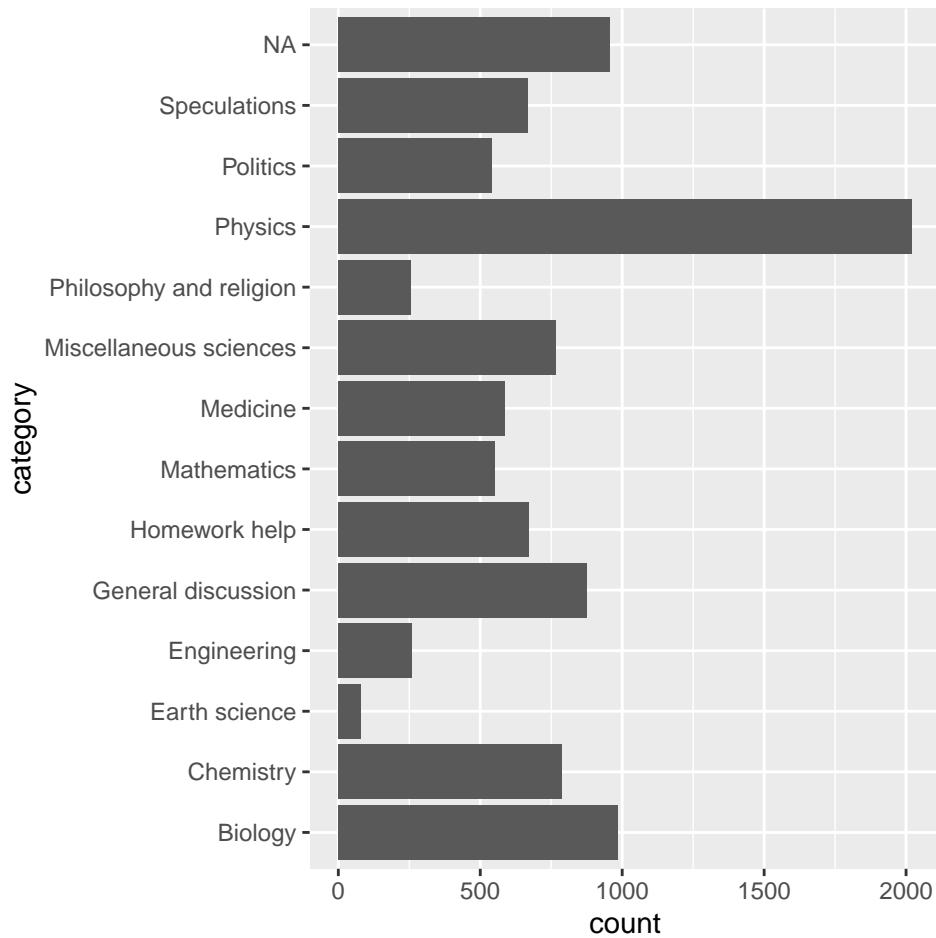


Figure 4: Breakdown of forums by topic, showing a portion of forums do not have any assigned topic.