

Supplementary Information for “Easy-access on-line social media metrics are associated with misinformation sharing activity”

Júlia Számely, Alessandro Galeazzi, Júlia Koltai, and Elisa Omodei

Robustness checks

Distributions of factuality scores

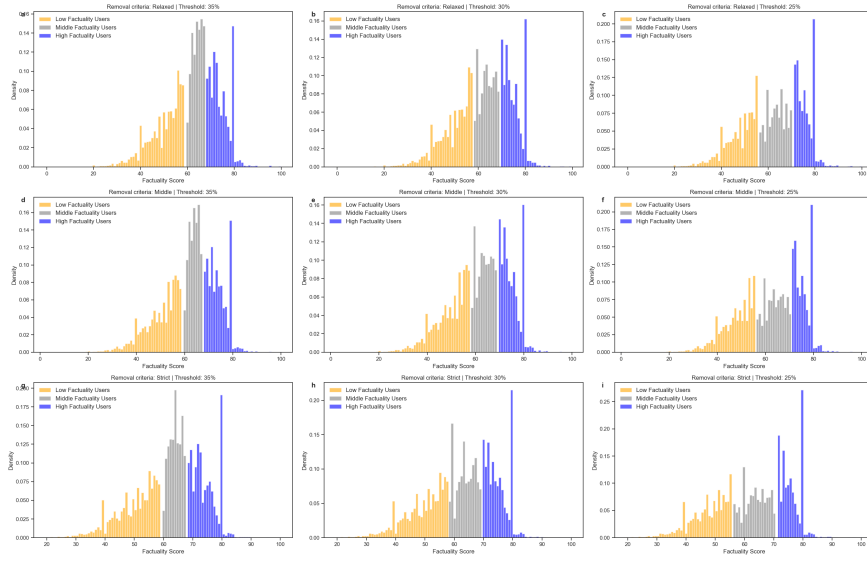


Figure 1: Distribution of factuality scores across the nine groups, defined by varying levels of regular user and bot filtering criteria (Relaxed, Middle, Strict) and low/high factuality group cutoffs (35/35, 30/30, 25/25).

Distributions of follower count

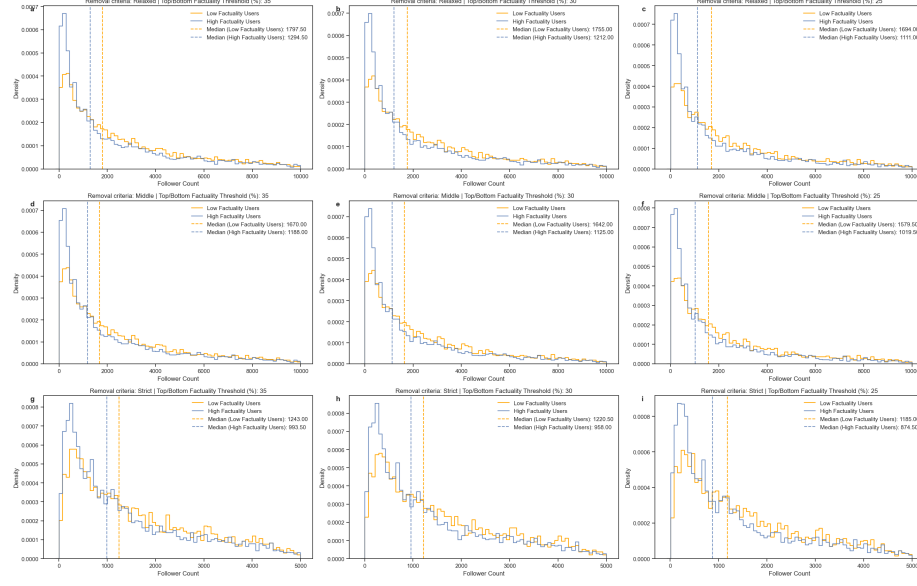


Figure 2: Low factuality users tend to have higher follower count than high factuality users. Comparing the distributions of follower count between low (orange curve) and high factuality (blue curve) users, we find that low factuality users have a significantly higher number of followers compared to high factuality users ($p\text{-value} < 0.0001$) in each of the tested datasets. The median value for low factuality users (orange dotted line) is higher than for high factuality users (blue dotted line) in each of the tested datasets.

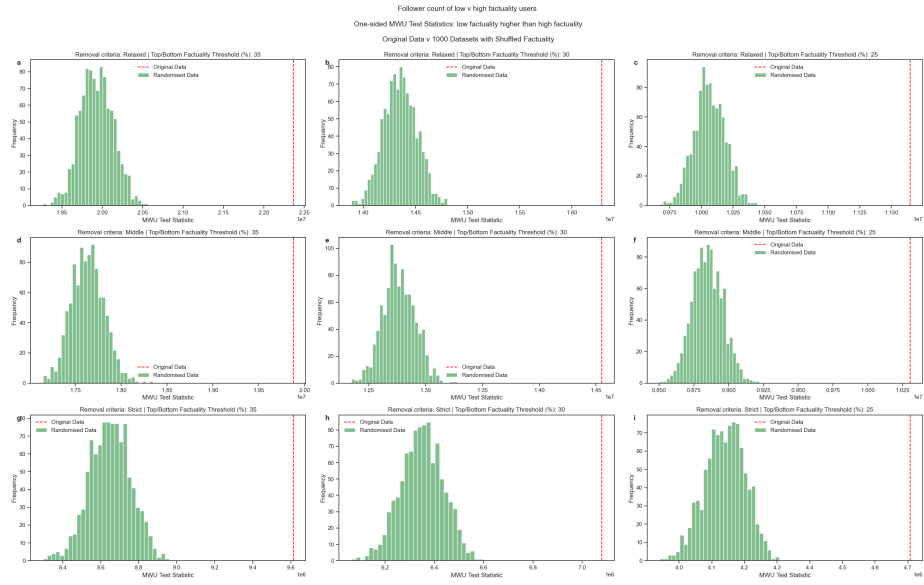


Figure 3: Low factuality users tend to have higher follower count than high factuality users. The MWU test score obtained from the empirical data (red dotted line) is higher than all the MWU test scores calculated on the 1,000 shuffled datasets (green bars) in each of the tested datasets.

Distributions of tweet count

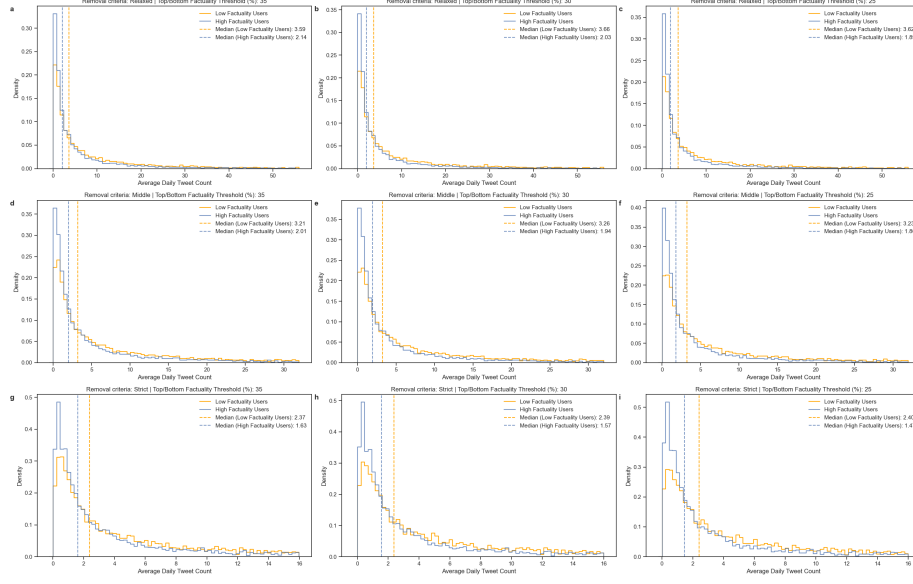


Figure 4: **Low factuality users tend to have higher tweet count than high factuality users.** Comparing the distributions of tweet count between low (orange curve) and high factuality (blue curve) users, we find that low factuality users have a significantly higher number of tweets compared to high factuality users ($p\text{-value} < 0.0001$) in each of the tested datasets. The median value for low factuality users (orange dotted line) is higher than for high factuality users (blue dotted line) in each of the tested datasets.

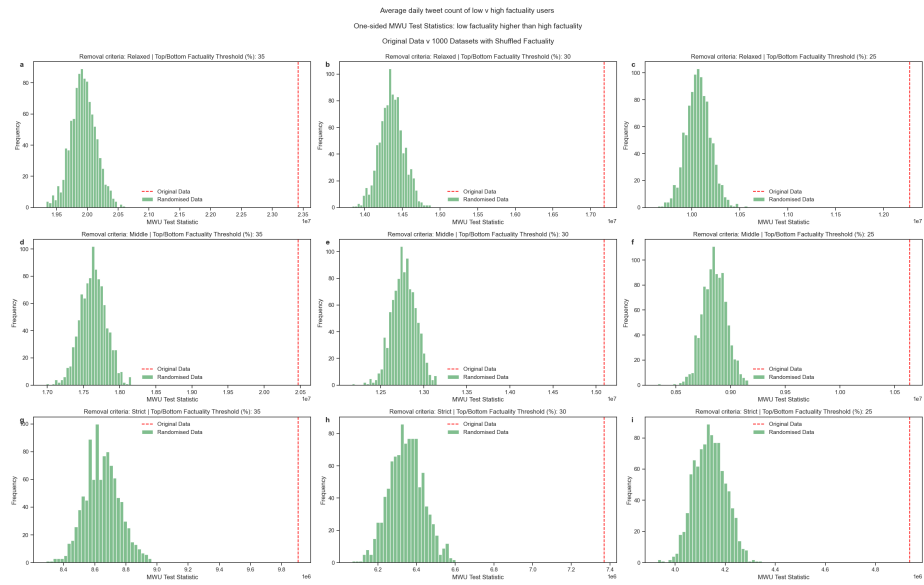


Figure 5: **Low factuality users tend to have higher tweet count than high factuality users.** Panel b: The MWU test score obtained from the empirical data (red dotted line) is higher than all the MWU test scores calculated on the 1,000 shuffled datasets (green bars) in each of the tested datasets.

Distributions of followed account count

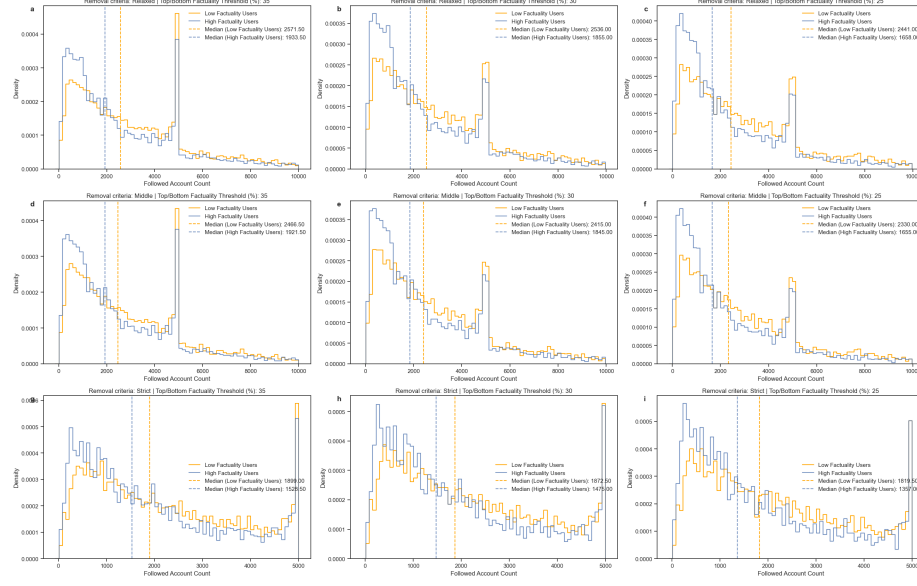


Figure 6: **Low factuality users tend to have higher followed account count than high factuality users.** Comparing the distributions of followed account count between low (orange curve) and high factuality (blue curve) users, we find that low factuality users have a significantly higher number of followed accounts compared to high factuality users ($p\text{-value} < 0.0001$) in each of the tested datasets. The median value for low factuality users (orange dotted line) is higher than for high factuality users (blue dotted line) in each of the tested datasets. The peak around 5000 is due to an X policy that limits the number of new followed accounts until the user obtains more followers.

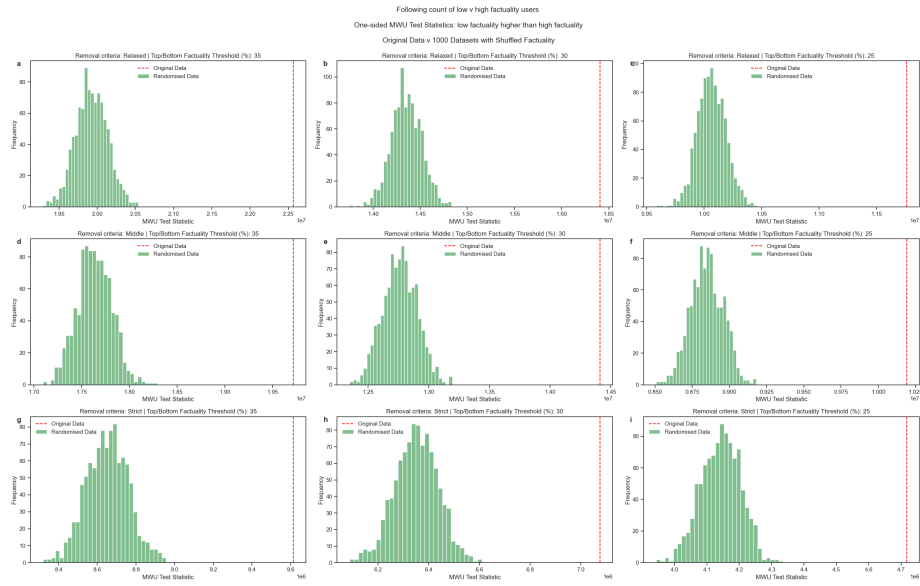


Figure 7: Low factuality users tend to have higher followed account count than high factuality users. The MWU test score obtained from the empirical data (red dotted line) is higher than all the MWU test scores calculated on the 1,000 shuffled datasets (green bars) in each of the tested datasets.

Distributions of account age

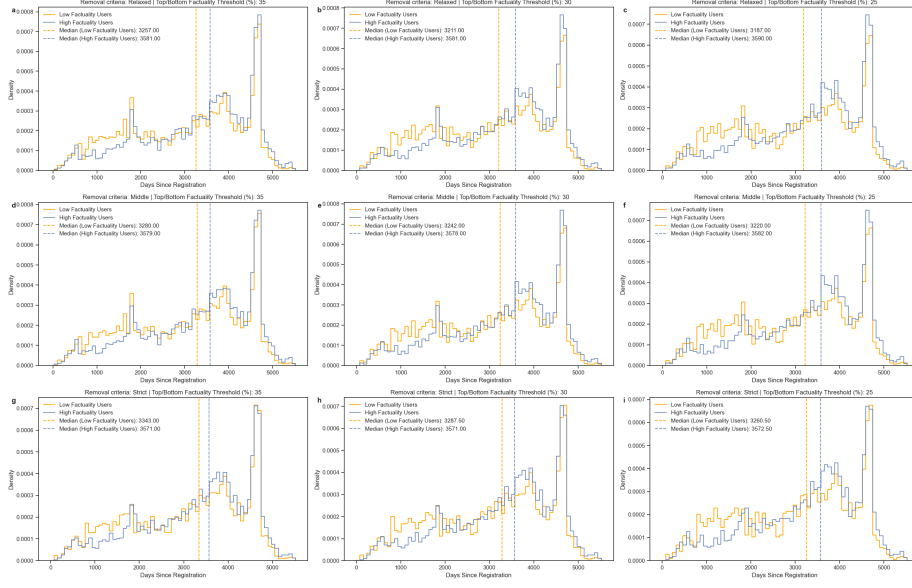


Figure 8: **Low factuality users tend to have lower number of days since registration than high factuality users.** Comparing the distributions of the number of days since registration between low (orange curve) and high factuality (blue curve) users, we find that low factuality users have a significantly lower number of days compared to high factuality users ($p\text{-value} < 0.0001$) in each of the tested datasets. The median value for low factuality users (orange dotted line) is lower than for high factuality users (blue dotted line) in each of the tested datasets.

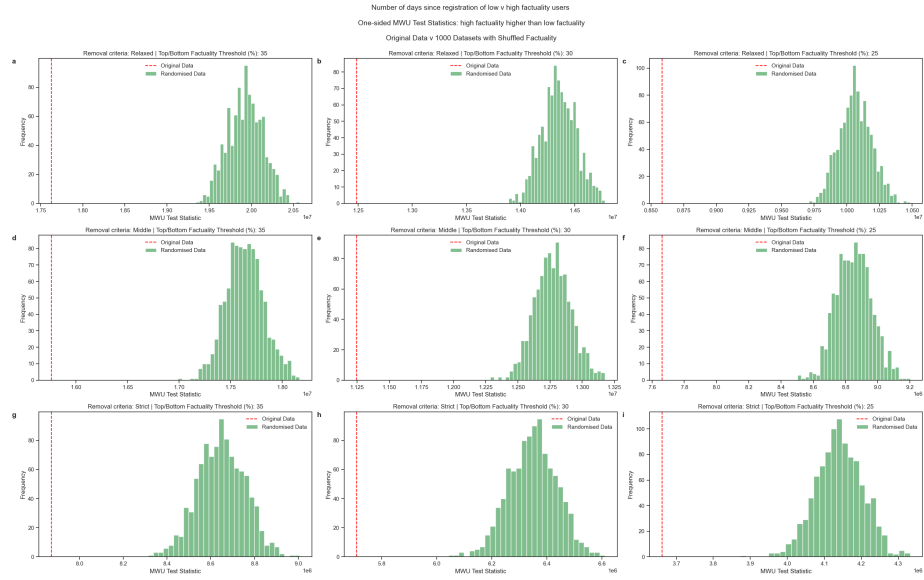


Figure 9: Low factuality users tend to have lower number of days since registration than high factuality users. The MWU test score obtained from the empirical data (red dotted line) is lower than all the MWU test scores calculated on the 1,000 shuffled datasets (green bars) in each of the tested datasets.

Regression results

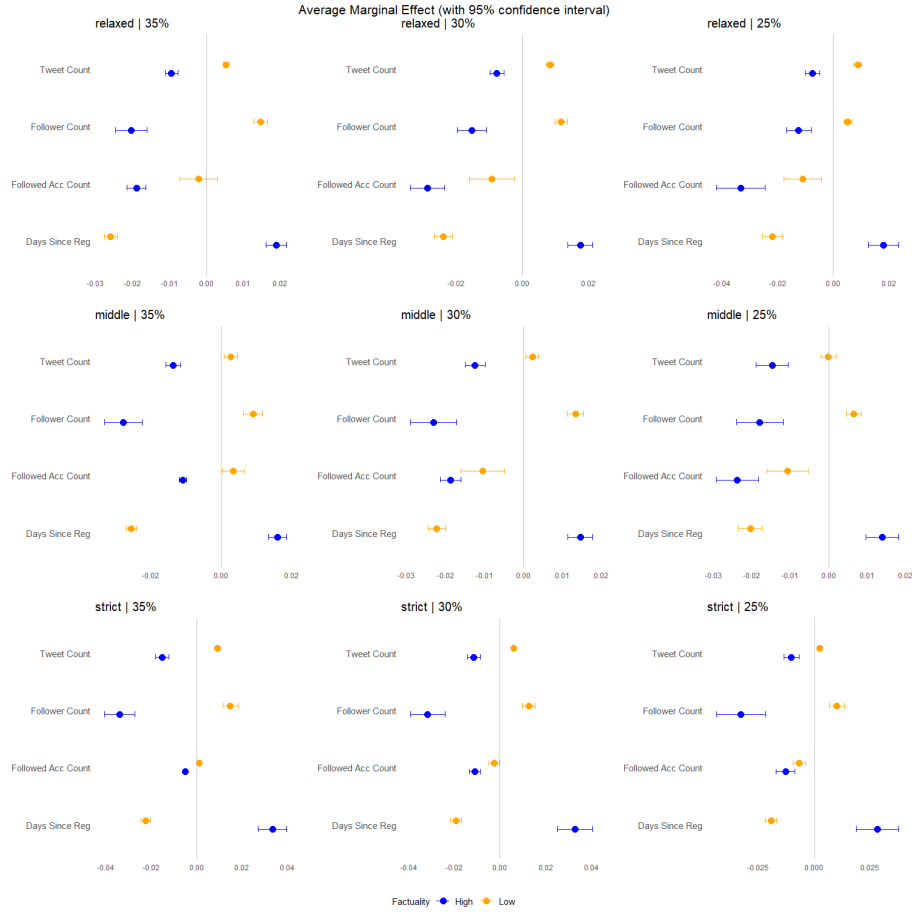


Figure 10: Average marginal effects of social network characteristics on factuality for the 9 datasets for checking robustness of results.

Interactions

Through plotting the interactions of follower count with all other metrics, certain differences in the relationships between these metrics and factuality levels across users with varying follower counts have been observed. Specifically, our analysis reveals that in case of users with fewer followers the effect of the followed account count diminishes on high factuality, whereas those with a larger follower base show an increased likelihood of high factuality with more followed accounts (Figure 11, panel a). Thus, the effect of followed account count on high factuality is nonexistent among those with smaller follower base, while positive among those with larger follower base. For users in the low factuality group, we observe a similar trend across both follower count categories: as users follow more accounts, they become less likely to belong to the low factuality group. This effect is slightly stronger for users with a higher follower count. Among those with fewer followed accounts, popular users are more likely to have low factuality, while in case of those who follow many accounts, there is no significant difference in their likelihood of low factuality by popularity.

In relation to average daily tweet count our analysis (Figure 11, panel b) shows that as users increase their tweet activity, they are less likely to belong to the high factuality category and more likely to belong to the low factuality category, regardless of their follower count. Nevertheless, in the low factuality group, this relationship is less pronounced among users with fewer followers.

Regarding the relationship between factuality and days since registration (Figure 11, panel c) we show that as users spend more time registered on the platform, they are more likely to belong to the high factuality category and less likely to belong to the low factuality category, and this is consistent in direction and magnitude across users, regardless of their follower count. However, the positive effect on high factuality is slightly stronger among users with higher follower counts. For the low factuality group, we observe a negative relationship with days since registration, with a marginally stronger effect among those with more followers. It shows that among younger users popular ones are more likely to have lower factuality compared to less popular ones. However, among older users there is no difference in the likelihood of factuality between more and less popular users.

In examining average daily tweet count across different levels of followed account count and its association with factuality levels (Figure 11, panel d) we find that among those with few accounts followed the average tweet count does not affect the likelihood of belonging to the high factuality group. However, an increase in tweet activity is associated with an increased likelihood of belonging to the high factuality group among those who follow more accounts. For the low factuality group, tweet count is negatively associated with low factuality and this effect is stronger among those who follow more accounts. There is no significant difference in the likelihood of belonging to the low factuality group among those who tweet infrequently, however, among those who tweet frequently, users following few accounts are way more likely to have low factuality than users who follow more accounts. Figure 11, panel d corresponds to Figure 6, panel a

in the main text.

Examining the interaction between days since registration and tweet count in relation to factuality levels reveals consistent trends across users with different average daily tweet counts (Figure 11, panel e). Our analysis demonstrates that as the registration period lengthens, users are more likely to exhibit higher factuality levels, and less likely to exhibit lower factuality levels, irrespective of their tweet activity.

Our findings regarding the interaction between days since registration and followed account count (Figure 11, panel f) indicate that among users who follow fewer accounts, those who have been registered on the platform for a longer duration are more likely to be in the high factuality category, while among those who follow more accounts, the trend is the opposite: the longer they are on the platform, the less likely that they belong to the high factuality group. The relationship between days since registration and low factuality does not vary based on followed account count. The relationship is consistently negative between days since registration and low factuality, indicating that longer registration periods are associated with a decreased likelihood of low factuality, however, this effect is somewhat more pronounced among users who follow fewer accounts. Figure 11, panel f corresponds to Figure 6, panel b in the main text.

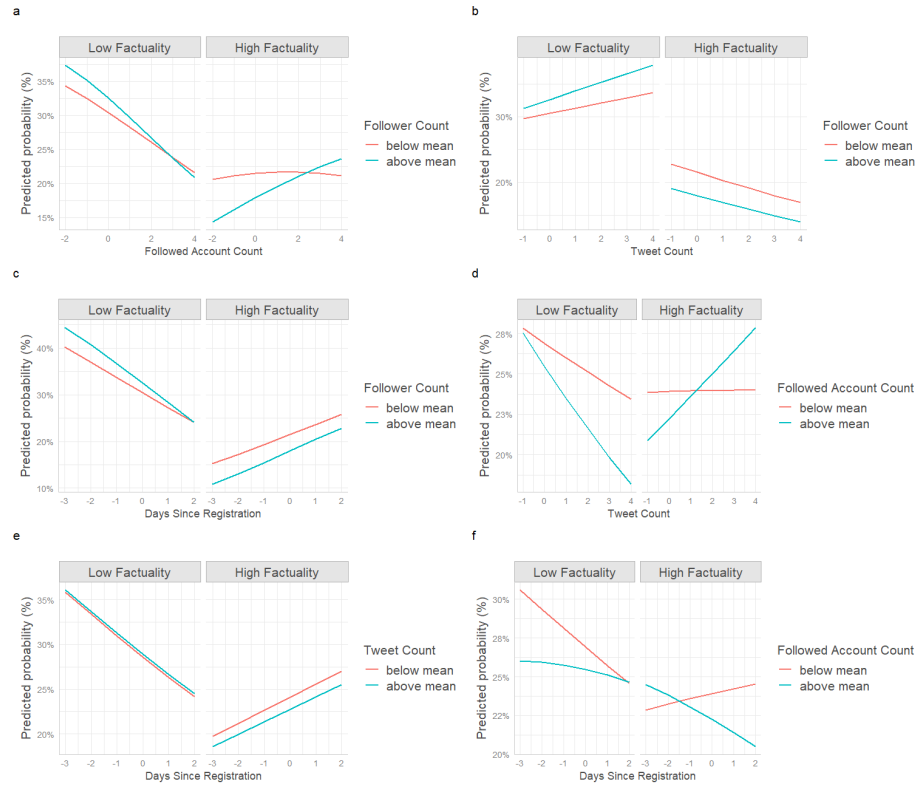


Figure 11: Interactions between various social network metrics in relation to belonging to low and high factuality groups.

Pseudo-R (McFaddens R)

Pseudo-R (McFaddens R) obtained from shuffled datasets against those derived from the real data to demonstrate that the Pseudo-R obtained from our regression exceeds that expected by random chance.

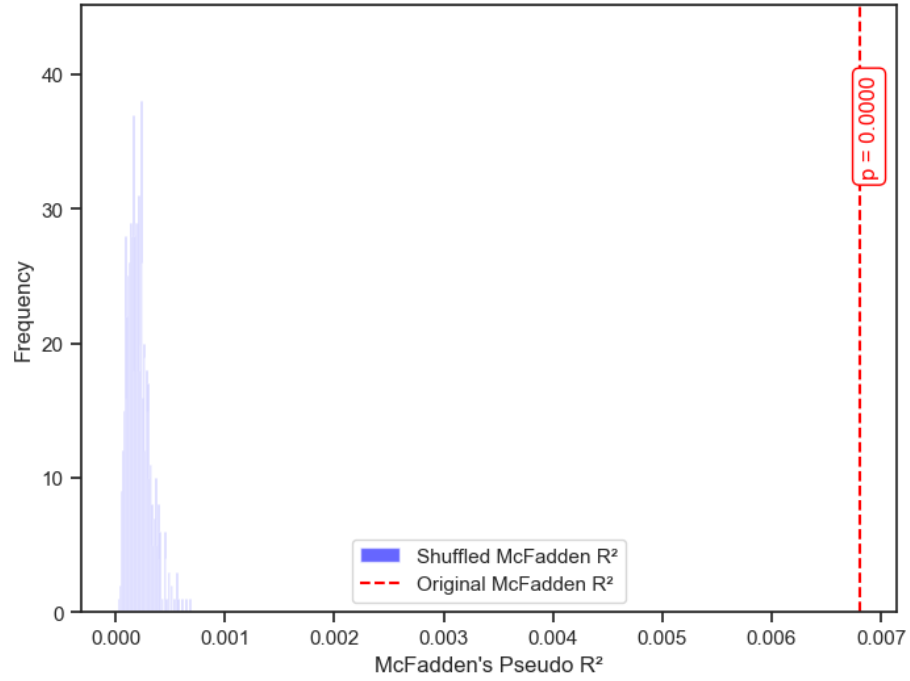


Figure 12: Pseudo-R (McFaddens R) using the four social network metrics on the real dataset (red dotted line) compared to the accuracy scores obtained on the 1000 randomised datasets (blue bars).