The Dynamics of (Not) Unfollowing Misinformation Spreaders

Anonymous Author(s)

ABSTRACT

1

2

3

4

5

7

8

9

10

11

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

Many studies explore how people "come into" misinformation exposure. But much less is known about how people "come out of" misinformation exposure. Do people organically sever ties to misinformation spreaders? And what predicts doing so? Over six months, we tracked the frequency and predictors of ~1M followers unfollowing ~5K health misinformation spreaders on Twitter. We found that misinformation ties are persistent. Monthly unfollowing rates are just 0.52%. Users are also 31% more likely to unfollow non-misinformation spreaders than they are to unfollow misinformation spreaders. Although generally infrequent, the factors most associated with unfollowing misinformation spreaders are (1) redundancy and (2) ideology. First, users initially following many spreaders, or who follow spreaders that tweet often, are most likely to unfollow later. Second, liberals are more likely to unfollow than conservatives. Overall, we observe strong persistence of misinformation ties. The fact that users rarely unfollow misinformation spreaders suggests a need for external nudges and the importance of preventing exposure from arising in the first place.

CCS CONCEPTS

• Human-centered computing → Social networks; Social media; Computer supported cooperative work; Social network analysis.

KEYWORDS

misinformation, social media, unfollowing, responsible web, twitter, social networks

ACM Reference Format:

Anonymous Author(s). 2018. The Dynamics of (Not) Unfollowing Misinformation Spreaders. In Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX). ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXXXXXXXXXXX

1 INTRODUCTION

Misinformation exposure affects key decisions like compliance with health regulations [11] and vaccine uptake intention [28]. Due to its importance, many studies explore how users become exposed to misinformation. But much less is known about when users choose to *unfollow* misinformation sources. Even some of the most *fundamental* empirical questions around misinformation unfollowing are still largely unanswered. For example, is unfollowing misinformation sources generally common or rare? And is misinformation

58

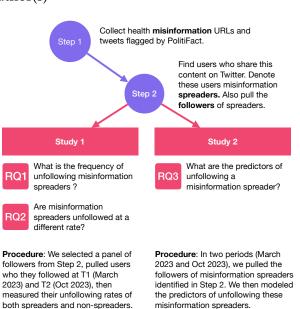


Figure 1: Graphical summary of studies.

exposure *self-reinforcing* or *self-correcting*? That is: Is high initial exposure predictive of *higher* or *lower* unfollowing rates?

Little is known about misinformation unfollowing—but understanding its frequency and predictors benefits researchers and platforms interested in a more responsible Web. First, regarding frequency: If unfollowing is rare, then this suggests content moderation and interventions to stop connections from forming in the first place are crucial for stopping exposure. But if unfollowing is common, then perhaps users organically reduce their misinformation exposure—reducing the content moderation burden. If unfollowing is common, this also complicates the interpretation of studies describing a 'snapshot' of misinformation exposure at a moment in time. Second, understanding the predictors of unfollowing misinformation spreaders can help both researchers and platforms design interventions to *further* increase unfollowing.

Here we provide the first large-scale account of misinformation unfollowing. We modeled the frequency and predictors of Twitter users (~1M followers, ~5K spreaders, ~3M edges) unfollowing health misinformation spreaders. We identified misinformation spreaders as Twitter users who shared content that PolitFact flagged as health misinformation. The study period ran from March 2023 to October 2023. We focus on health misinformation for two reasons. First, claims related to health are often more straightforwardly falsifiable than claims in other domains. Second, the cost of false health misinformation can be high.

Across two studies, we answer three research questions related to the frequency and predictors of unfollowing misinformation spreaders. See Figure 1 for a graphical overview.

115

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish,

to post on servers or to redistribute to lists, requires prior specific permission and/or a
 fee. Request permissions from permissions@acm.org.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY
 2018 Association for Computing Machinery.

⁵⁶ ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

⁵⁷ https://doi.org/XXXXXXXXXXXXX

Anon.

RQ1 (Study 1): How often are misinformation spreaders unfollowed? We track how often spreaders are unfollowed from the first time point to the second time point.
RQ2 (Study 1): Are misinformation spreaders unfollowed at a different rate than non-misinformation spreaders? We

compared how often a subset of followers unfollows spreaders vs non-spreaders.

RQ3 (Study 2): What predicts unfollowing a misinformation spreader? We model unfollowing as a function of initial exposure, ideology, edge characteristics, and platform activity.

2 RELATED WORK

Here we review related work on (1) factors perpetuating misinformation exposure and (2) predictors of unfollowing on social media.

2.1 Misinformation Exposure

Many studies explore how misinformation exposure arises in the first place. There is no uniform pathway. Factors perpetuating misinformation exposure can be grouped into two buckets: *individual* (e.g., ideology, cognitive reflection, attention) and *environmental* (e.g., algorithms, defaults).

2.1.1 Individual Factors. Misinformation exposure is driven by individual-level factors, such as selective exposure [8] cognitive reflection, [25], and inattention to accuracy [23].

Selective Exposure. People consume (mis)information consistent with their ideology. In a large study based on browsing data, conservatives consumed more untrustworthy news content than liberals overall [8] in the 2016 election. But for both Trump and Clinton supporters, users were more likely to visit untrustworthy websites consistent with their political ideology-with an especially large effect for Trump supporters [8]. Another large-scale study also supports selective exposure: people are much less likely to click on cross-ideology links in newsfeeds [1]. Moreover, selective exposure appears to be a stronger driver of misinformation consumption for users with extreme political ideologies. In 2016, consumption of fake news with respect to ideology was 'v-shaped': both extreme liberals and extreme conservatives consumed larger amounts of fake news [19]. Although fake news exposure decreased during the 2020 election, it was still the case that conservatives-and particularly, extreme conservatives-consumed more fake news than liberals [19]. Robertson et al. compared the amount of unreliable and partisan news users were exposed to in search results versus the amount of unreliable and partisan news they consumed. Consumption of the latter was significantly higher than the former. This also suggests users are seeking unreliable news (over and above what is being shown to them by platforms). Overall, there is large-scale evidence that selective exposure drives misinformation consumption.

Cognitive Reflection. Cognitive reflection is implicated in misinformation exposure and belief. The Cognitive Reflection Test (CRT) measures a person's tendency towards analytical thinking [7]. People who are less likely to engage in such thinking are more likely to consume and believe misinformation. In lab experiments, low-CRT participants are more likely to believe conspiracy theories and fake news [24, 25]. CRT correlates with on-platform behavior [20]: Twitter users lower in CRT are more likely to share low-quality news, and there are differences in Twitter accounts that low vs. high CRT users follow.

Inattention. Users may *share* misinformation because they are not paying attention to the accuracy of what they share [23, 26]. This line of research is principally concerned with *sharing* of misinformation but also has implications for *exposure* to misinformation. Users likely encounter more misinformation in their feeds because other users share content inattentively.

2.1.2 *Environmental Factors.* Misinformation exposure also may arise more incidentally through environmental factors. Here, we refer to both environments built for users (e.g., social media algorithms) and environments users build for themselves (e.g., defaults).

Algorithms. There is mixed evidence regarding the role of recommendation algorithms in promoting misinformation. Several auditing studies found consuming misinformation content leads to more misinformation content being recommended by algorithms [10, 21, 30]. These studies suggest recommendation algorithms may amplify *already-existing* misinformation exposure. Yet, the largest field experiment to date on this topic suggests algorithms *do not* promote misinformation on Facebook and Instagram [9]. In a different domain (extremist content on Youtube), Chen et al. also questions the significance of algorithmic effects and points to demand effects as a more important mechanism.

Defaults. Misinformation exposure can arise through the defaults that users set. Users curate a set of sources they regularly consume, what Kim calls a 'media repertoire'. This repertoire can serve as a default filter for news. For example, Flaxman et al. showed that most online news consumption was driven by users visiting their homepage. That is, the 'defaults' that users created (i.e., setting a homepage) strongly influenced the content that users saw. Analogously, the content from the people that a user chooses to follow online can be thought of as the 'default' content that the user sees. Combined with homophily, this can create ideologically segregated filter bubbles or echo chambers [3, 5, 29]. This is a different dynamic than selective exposure, where one is seeking out the information itself. Of course, the initial choosing of defaults is an *individual-level* decision. But this decision then creates a content *environment* that may expose one to misinformation.

2.2 Unfollowing on Social Media

Several variables have emerged as key predictors of unfollowing. A reciprocal tie between Twitter users A and B is associated with significantly lower odds of A unfollowing B [15, 16, 32]. Reciprocity may be a cause or effect of tie strength. Kwak et al. argues that reciprocal relationships on social media cause emotional closeness between two users since they see each other's posts. Alternatively, reciprocity may signal two users are already friends or acquaintances in an offline context [18]. Additionally, *redundancy*—either similar content to what the user follows or burst-tweeting—is a predictor of unfollowing [15, 17]. More relevant, Kaiser et al. found participants reported higher theoretical intentions to unfollow or

Table 1: How samples map to studies and research questions.

Sample Name	Description	Used in Study	Answers Research Question(s)
Initial Sample	Initial pull of spreaders and followers	-	-
Modeling Sample	Subset of the Initial Sample used for modeling predicting unfollowing	Study 2	RQ3
Rate Sample	Subset of the Modeling Sample used for rates of unfollowing	Study 1	RQ1, RQ2

block an imagined cross-party friend for posting misinformation, and Yoo et al. found liberals were more likely to report unfriending.

3 SPECIFIC HYPOTHESES

Considering unfollowing of misinformation spreaders in particular, we are specifically interested in two key predictors: T1 exposure (the number of spreaders a follower follows at time point 1) and partisan ideology. These variables represent network and individual-level variables, respectively.

3.1 Initial (T1) Exposure Effect on Unfollowing

It is not clear, based on the existing literature, if T1 exposure (the number of misinformation spreaders, hereafter 'spreaders', followed at the first time point) should be positively or negatively related to unfollowing. And yet, this empirical relationship is important since it can speak to the extent to which misinformation exposure is self-correcting or self-reinforcing:

- Reversion Hypothesis (H1): Higher exposure at T1 is associated with *higher* unfollowing at T2—meaning misinformation is partially *self-correcting*.
- Inertia Hypothesis (H2): Higher exposure at T1 is associated with *lower* unfollowing at T2—meaning misinformation is partially *self-reinforcing*.

The logic for the *reversion* hypothesis **(H1)** is that high T1 exposure would signal high redundancy. And if misinformation ties are redundant, the probability of each being unfollowed should increase. Additionally, if the follower came to be exposed to misinformation through some incidental mechanism and not an intentional one, then this makes regression to the mean a likely prediction. Both mechanisms would suggest *increased* unfollowing in T2 if T1 exposure is high.

The logic for the *inertia* hypothesis **(H2)** is that high misinformation exposure at T1 may be (1) a *consequence* of selective exposure due to extreme ideology [8] or (2) a *cause* of believing in the misinformation (since exposure¹ to misinformation increases its perceived accuracy [22]). Both mechanisms would suggest *decreased* unfollowing in T2 if T1 exposure is high.

3.2 Partisan Ideology Effect on Unfollowing

We hypothesized that ideology (left/right, moderate/extreme) would affect if a user unfollowed a misinformation spreader. Some evidence suggests liberals have a higher unfollowing/unfriending rate [33] so we hypothesized that **(H3)** liberals may be more likely to unfollow here (though political unfollowing is a different phenomenon). We also hypothesized that **(H4)** politically extreme users would be less likely to unfollow since ideological extremity is correlated with misinformation exposure [19]. But in 2020, *extreme liberals* decreased consumption of fake news [19]. And even in 2016, the top decile of conservatives had larger misinformation consumption than the top decile of liberals [8]. Consequently, we hypothesized that there would be a negative interaction effect (H5) between conservatism and ideological strength; an equivalent increase in ideological strength would reduce the probability of unfollowing more for conservatives than it would for liberals. We were also interested in (\mathbf{Q}) a possible interaction between T1 exposure and ideological strength such that the effect of T1 exposure on unfollowing might differ for ideologically moderate vs extreme users.

4 DATA

The data for Studies 1 and 2 come from an **Initial Sample** of misinformation spreaders and their followers, which we describe in Section 4.1. We then created two subsets of this **Initial Sample**—a **Rate Sample** used for Study 1 (rates of unfollowing) and a **Modeling Sample** used for Study 2 (modeling predictors of unfollowing). The creation of these samples is described in sections 4.2 and 4.3. See Table 1 for how samples relate.

The **Modeling Sample** is composed of users for whom we could obtain covariate information to include in our unfollowing model and who did not leave the platform between T1 and T2. The **Rate Sample** is an active, smaller sample of followers derived from the **Modeling Sample**. We pull the users who this sample follows at T1 and T2 so we can estimate unfollowing rates.

4.1 Initial Sample

To construct the **Initial Sample** of spreaders and followers, we: (1) identified health misinformation rumors from PolitiFact, (2) found users who shared URLs corresponding to these rumors ('spreaders'), and then (3) collected users who followed 'spreaders' ('followers').

Collecting Misinformation URLs and Tweets. In March 2023 we collected all health² misinformation³ rumors on PolitiFact since June 2021. We refer to the clean set of blog post URLs and tweets as Misinformation URLs and Misinformation Tweets, respectively. There were 84 such URLs and tweets.

Identifying Eligible Spreaders. We next identified 'Eligible Spreaders'. We conducted a search of Twitter to find all users who either (A) posted a Misinformation URL, (B) retweeted a post containing a Misinformation URL, (C) posted a Misinformation Tweet, or

³PolitiFact truth values: ['pants on fire', 'false', 'mostly false']

¹Though of course, it may be the case that not all misinformation spreaders share the same misinformation.

²PolitiFact categories: ['abortion', 'autism', 'coronavirus', 'drugs', 'disability', 'healthcare', 'health-check', 'public-health']

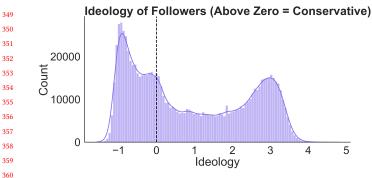


Figure 2: Distribution of the political ideology of followers from the Modeling Sample, measured using the method from Barberá. Most followers are conservative.

(D) retweeted a Misinformation Tweet. We denote the union of A, B, C, D as Eligible Spreaders.

Filtering Eligible Spreaders. We applied three filters to Eligible Spreaders. First, we removed any post or retweet of a post that contained a series of debunking words. We did this to remove users who were correcting misinformation as opposed to spreading it. Debunking words (Appendix Table 2), were generated by prompting GPT-3 to expand on a list of common debunking words. Second, we selected spreaders with follower and friend counts of over ten and a follower count of less than 20K. The lower bound was applied to make sure spreaders had some activity on the platform. The upper bound on followers was partially a resource constraint (since we had to pull all of the followers), but it also allowed us to understand 'regular' users and not celebrities. After the first two filters, there were 58 Misinformation URLs and Misinformation Tweets, since not all Misinformation URLs were ever tweeted. Some of these URLs had a disproportionate number of associated tweets. Consequently, in the third filter, we used a simple greedy algorithm to retrieve 5,600 misinformation spreaders⁴ while minimizing the number of spreaders that came from any specific story (Appendix Algorithm 1). These steps resulted in 5,613 misinformation spreaders (hereafter 'Spreaders'), and the rumor, or misinformation URL, with the largest number of associated spreaders made up 3.6% of the total.

Collecting Followers. We then pulled all followers of the spreaders at two time points: (T1) March 2023 and (T2) October 2023.

4.2 Modeling Sample

The participant pool for the **Modeling Sample** (used in Study 2) began from the followers and spreaders from the **Initial Sample**, and then two filters were applied. In the first filter, we restricted our analysis to only followers whose ideology could be estimated via data files provided by the first author of the 'Bayesian Ideal Point Estimation' [2] method. This is a method that estimates a user's ideology by who the user follows. Applying this restriction yielded 944,972 followers and 5,593 spreaders. As a robustness check (Appendix Figure 11), we show that relationships between *non-ideology*

406

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

1.0 1.0 0.5 0.0 10^{0} 10^{1} 10^{2} 10^{3}

Number of Spreaders Followed by Followers

Figure 3: Number of misinformation spreaders from the Modeling Sample followed at T1. Ideology is cut at zero using Barberá. Misinformation exposure is right-skewed.

variables and unfollowing hold for users whose ideology we could not estimate, suggesting the structural predictors of unfollowing do not differ for users whose ideology we could not estimate. In the second filter, we removed from our analysis any follower or spreader whose basic information (e.g., followers, tweet count, friend count) could not be pulled at either T1 or T2. This can happen for multiple reasons—voluntarily exiting the platform, getting banned, etc. It is important to remove edges where either the follower or spreader left the platform since including them would distort unfollowing rates. For the **Modeling Sample**, we also had to omit users whose accounts were protected since their information could not be pulled and hence could not be included in the model. After these two filters, there were 898,701 followers, 5,334 spreaders, and 3,376,785 edges. See Appendix Table 3 for variable explanations and descriptive statistics.

Most followers in the **Modeling Sample** are conservative (Figure 2), but there is a bimodal distribution similar to the 'v-shaped' distribution of fake news exposure in 2016 [19]. Misinformation exposure is right-skewed, and conservatives follow more misinformation spreaders than liberals (Figure 3). The followers in the top 10% for the number of misinformation spreaders followed at T1

Average z-score for top 10% of followers based on number of spreaders followed Abs Ideo Ideology Follower Tweet Count Follower Following Count Is Liberal

Figure 4: Characteristics of those Modeling Sample followers who are in the top 10% for following misinformation spreaders. Ideology is a continuous measure where positive is conservative. 'Recip' is the proportion of a follower's ties to spreaders that are reciprocated.

4

Log (N Spreaders Followed)

Anon.

407

408

409

410

411

412

413

414

415

416

417

418

450

451

452

453

454

455

456

457

458

459

460

461

462

463

 ⁴This number was based on a power analysis conducted for a relevant and concurrent
 project.

The Dynamics of (Not) Unfollowing Misinformation Spreaders

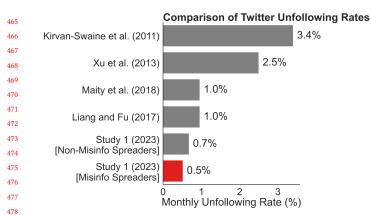


Figure 5: Comparing unfollowing rates across studies, misinformation spreaders are unfollowed relatively infrequently.

are generally (1) ideologically extreme and (2) conservatives (Figure 4), which tracks browsing studies on selective exposure [8, 19]. There is a high reciprocity rate (69%), driven by a few spreaders. In fact, the top 10% of spreaders by reciprocal ties are responsible for 66% of the reciprocal ties; the Gini coefficient of reciprocal ties (0.79) is even larger than that of followers (0.74). This could suggest either that the most active misinformation spreaders engage in frequent 'follow-back' behavior or that the most popular misinformation spreaders are embedded in close-knit communities where reciprocation rates are high.

4.3 Rate Sample

To estimate unfollowing rates of misinformation spreaders and non-spreaders (Study 1), we selected a panel of 2,500 followers (**Rate Sample**) from the **Modeling Sample** and pulled who these users followed at roughly⁵ the same time two points as in Study 2. The **Rate Sample** was constructed by applying several filters to followers in the **Modeling Sample**. First, we filtered followers in the **Modeling Sample** by activity level and network size (last tweet within 14 days, tweet count greater than 20, follower count between 20-20K, friends count between 20-20K) to ensure these users were active on Twitter. Second, we capped the number of friends at the 85th percentile (4,989) to further constrain network size. This was done due to resource constraints. Finally, 2,500 participants were randomly sampled from the remaining filtered pool to serve as a panel of egos.

5 STUDY 1: UNFOLLOWING OF SPREADERS AND NON-SPREADERS

5.1 Objectives

In Study 1, we analyze both (RQ1) the overall unfollowing rates of misinformation spreaders and (RQ2) if the unfollowing rate of misinformation spreaders differs from that of non-spreaders.

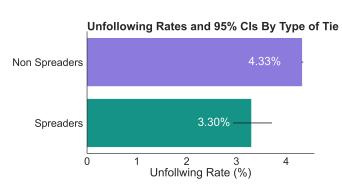


Figure 6: In Study 1, misinformation spreaders were unfollowed less than non-misinformation spreaders.

5.2 Participants

The participants started from the 2500 misinformation followers from the **Rate Sample**. See Table 1.

5.3 Methods

For each of these 2,500 follower 'egos', we pulled who they followed (their 'alters') at two time points—March 2023 (T1) and October 2023 (T2). We were interested in the proportion of (follower \Rightarrow alter) ties that were dissolved from T1 to T2 and if a tie is more likely to dissolve if the alter is a misinformation spreader. To avoid exit rates confounding unfollowing estimates—we might count a user as unfollowed if they instead deleted their account in T2—we removed all egos and alters who exited the platform because they were either suspended or their account was deleted.⁶ This process yielded 2,467 egos, 2,245,645 T1 alters, and 5,087,256 T1 (follower \Rightarrow alter) edges. 8,052 of the initial edges were with misinformation spreaders.

5.4 Results

5.4.1 RQ1: How often are misinformation spreaders unfollowed? Misinformation spreaders are very rarely unfollowed. The unfollowing rate of misinformation spreaders was 3.3% (95% CI = [2.93%, 3.72%]), or a monthly rate of 0.52% (95% CI = [0.46%, 0.58%]). We computed confidence intervals (CIs) using the Wilson [31] method for binomial proportions. We computed monthly unfollowing rates by dividing the total unfollowing rate by $\frac{\text{study duration in days}}{30}$. This unfollowing rate is substantially lower than Twitter unfollowing rates observed in prior studies⁷ (Figure 5) or what would be expected based on Kaiser et al.

5.4.2 RQ2: Are misinformation spreaders unfollowed at a different rate than non-spreaders? The unfollowing rate of non-misinformation spreaders (4.33%, 95% CI = [4.31%, 4.34%]) was 31% higher than the equivalent rate of misinformation spreaders (3.3%, 95% CI = [2.93%, 3.72%]). See Figure 6. Note that the 95% CIs for unfollowing rates

⁵Due to Twitter API instability, Study 2 started 11 days earlier and ended 4 days earlier than Study 1.

⁶We got this information by querying Twitter's Compliance endpoint and removing any accounts where the compliance status was 'suspended' or 'deleted'. ⁷Prior unfollowing studies often differ in some way (e.g: sample, exact measure, exclusion criteria) to our study, but they provide a *rough* baseline.

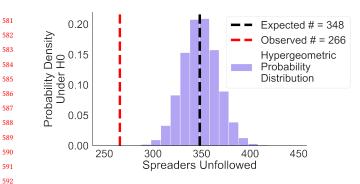


Figure 7: Fewer misinformation spreaders in Study 1 were unfollowed than would be expected under a null hypothesis of random unfollowing ($p = 1.6 \times 10^{-6}$). We modeled the null hypothesis with a hypergeometric distribution.

between the two groups do not overlap. (The wider confidence interval for the spreader unfollowing rate reflects the fact that there were fewer spreader edges in Study 1.)

We also tested the hypothesis that fewer spreaders were unfollowed than would be expected if unfollowing was random (or independent of whether the alter was a spreader). Under a null hypothesis of random unfollowing, we would expect the proportion of *severed* spreader ties (0.12%) to be the same as the *original* proportion of spreader ties (0.16%). But spreaders were a lower fraction of severed ties than initial ties. The null hypothesis of random unfollowing can be thought of as randomly 'sampling' users to unfollow without replacement. Then the probability of observing *k* spreader unfollows out of *n* total unfollows, given spreaders are $\frac{K}{N}$ proportion of initial ties, is described by the hypergeometric distribution with parameters (N = 5, 087, 256, K = 8, 052, n = 219, 979, k = 266):

$$P(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

and

$$P(X \le 266) = \sum_{i=0}^{266} P(X=i)$$

provides an exact p-value on the probability of observing 266 or fewer spreaders being unfollowed. That p-value is 1.6×10^{-6} (Figure 7). We conclude that fewer spreaders were unfollowed than one would expect if unfollowing was random.

6 STUDY 2: PREDICTORS OF UNFOLLOWING

6.1 Objectives

In Study 2, we answered what predicts a user unfollowing a misinformation spreader.

6.2 Participants

The participants were the spreaders and followers from the **Model**ing Sample. See Table 1.

6.3 Methods

We modeled a follower f unfollowing a misinformation spreader s using cluster-robust logistic regression. The data was at an edge level (follower $f \Rightarrow$ spreader s). We predicted if an edge that existed at T1 (March 2023) would be dissolved at T2 (October 2023). That is, we predicted if a follower would unfollow a spreader. To account for dependencies within spreaders, we report HC1 cluster-robust errors, clustered at the spreader level.⁸ See Appendix Table 4 for regression results. The covariates were the variables listed in Appendix Table 3, plus two-way interactions between our variables of interest: initial exposure (measured by number of spreaders followed at T1), ideological strength (measured by absolute value of Barberá's ideology measure), and an indicator for liberal ideology (equal to 1 if the Barberá measure is below 0). We used the *marginal effects* R package to compute average marginal effects (AME) from our logistic regression model.

6.4 Results

6.4.1 RQ3: What are the predictors of unfollowing a misinformation spreader? See Figure 8 for non-interaction AME estimates. See Appendix Table 4 for regression results.

Reciprocity. Reciprocity was the biggest predictor of unfollowing. A tie being reciprocal was associated with a significantly decreased probability of unfollowing, AME = -0.0134 (95% CI = [-0.0138, -0.0131]). The unfollowing rate of non-reciprocated misinformation-spreader ties (2.14%) was 2.4 times the unfollowing rate of reciprocated misinformation-spreader ties (0.89%).

Initial Exposure. The results were consistent with the reversion hypothesis (**H1**) and not the inertia hypothesis (**H2**): T1 exposure (AME = 0.0028, 95% CI = [0.0026, 0.0029]) and the spreader tweeting often from T1 to T2 (AME = 0.0022, 95% CI = [0.0021, 0.0023]) were associated with unfollowing. These can both be considered measures of 'redundancy'. There was a significant interaction between T1 exposure and partisan ideology, although the effects were directionally the same (see Figure 9). T1 exposure had a larger effect for liberals (AME = 0.0024, 95% CI = [0.0035, 0.0046]) than for conservatives (AME = 0.0024, 95% CI = [0.0023, 0.0025]). Interestingly, there was also a positive interaction between T1 exposure and ideological strength (Appendix Table 4). Overall, the reversion hypothesis (high exposure at T1 is associated with high unfollowing at T2) holds on average and separately for both liberals and conservatives—though this effect is roughly 1.7 times as strong for liberals.

Partisan Ideology. Misinformation unfollowing shows partisan asymmetries. Liberals are more likely to unfollow than conservatives. And at higher levels of ideological strength, this gap widens. First, **(H3)** was supported: liberals were more likely to unfollow than conservatives (AME = 0.0082, 95% CI = [0.0063, 0.0101]). Second, **(H4)** was technically refuted: Averaging across the sample, there was a near-zero, *positive* AME of ideological extremity on unfollowing (AME = 0.0004, 95% CI = [0.0001, 0.0006]). However, this near-zero, average effect masked large partisan asymmetries

⁸We also conducted a Bayesian hierarchical logistic regression with random intercepts for spreaders, but the model yielded a convergence error. Nonetheless, the results of the Bayesian model were very similar to the results of our cluster-robust logistic regression.

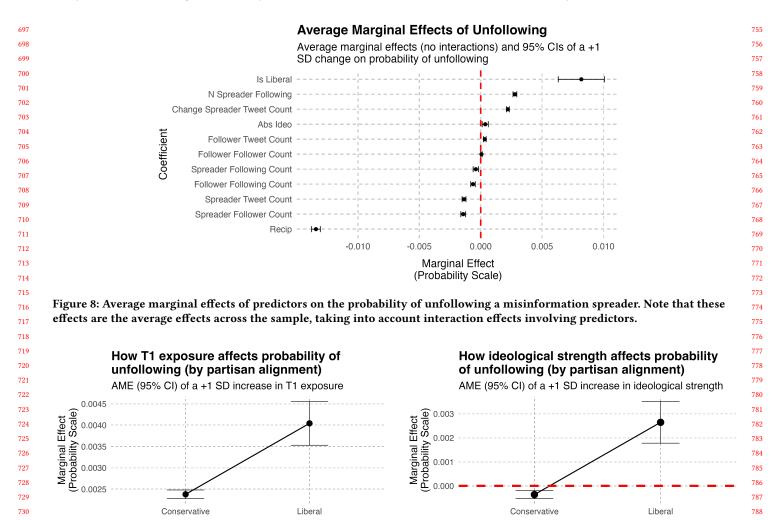


Figure 9: The effect of T1 exposure (how many spreaders a follower followed at T1) on the probability of unfollowing differed for liberals vs conservatives. T1 exposure has a *larger* effect on unfollowing for liberal users.

Ideology

(H5): For liberals, an increase in ideological strength was associated with an *increase* in unfollowing (AME = 0.0026, 95% CI = [0.0018, 0.0035]). Yet for conservatives, an increase in ideological strength was associated with a *decrease* in unfollowing (AME = -0.0004, 95% CI = [-0.0005, -0.0002]). That is: Extreme liberals are *more* likely to unfollow than moderate liberals, while extreme conservatives are *less* likely to unfollow than moderate conservatives. See Figure 10.

7 DISCUSSION

While much is known about misinformation exposure, little is known about misinformation unfollowing. We provided a largescale analysis of the frequency and predictors of unfollowing misinformation spreaders.

Regarding frequency, we found misinformation spreaders are rarely unfollowed. Monthly unfollowing rates are just 0.52%. Users

Figure 10: The effect of ideological strength on the probability of unfollowing differed for liberals vs conservatives. Ideological strength is associated with higher unfollowing for liberals and lower unfollowing for conservatives.

Ideology

are also 31% more likely to unfollow *non*-misinformation spreaders than they are to unfollow misinformation spreaders. The low overall unfollowing rate suggests a role for interventions or design changes that either prevent the initial formation or encourage the dissolution of these ties. Additionally, misinformation spreaders are unfollowed less than would be expected from [12], which asked laboratory participants their hypothetical intentions to unfollow an imagined misinformation spreader. This discrepancy has two implications. First, hypothetical exercises do not capture the realities of actual unfollowing behavior, highlighting the importance of large-scale studies 'in the field'. Second, there is room for effective interventions that move individual behaviors more toward individuals' stated goals regarding their information environments.

Although rare overall, some factors did (non-trivially) predict unfollowing. We found that reciprocity had a large downward effect on the probability of unfollowing. And likely because connections to misinformation spreaders were highly reciprocal, there
was a low overall unfollowing rate. This suggests that cost-effective
interventions to limit the spread of misinformation might target
non-reciprocated connections, which are easier to dissolve.

Considering initial exposure, we found more evidence for a rever-818 sion account than an inertia account: Overall, higher initial exposure 819 was associated with higher unfollowing and not lower unfollowing. 820 821 This could be driven by redundancy (indeed, spreaders tweeting of-822 ten⁹ were also predictive of unfollowing) or a regression to the mean effect if high initial exposure was incidental rather than intentional. 823 Of course, determining whether the higher rate of unfollowing 824 was due to redundancy or regression to the mean has important 825 implications for identifying the right messaging interventions to 826 encourage unfollowing. Highlighting information overload costs 827 828 might be more effective for the former, while messages that simply remind users of the information pollution in their networks 829 830 might be more effective for the latter. Future work (e.g., interviews of individuals who unfollow misinformation producers) can help 831 determine the mechanisms at play and inform messaging strategies. 832

We found partisan asymmetries. Liberals were more likely to 833 834 unfollow than conservatives. Future work could explore possible 835 mechanisms for this main effect. At more extreme ends of the spectrum, the gap between liberal and conservative unfollowing 836 widened: Extreme liberals are more likely to unfollow than moder-837 ate liberals, but extreme conservatives are less likely to unfollow 838 than moderate conservatives. Additionally, the reversion effect 839 is stronger for liberals. That is, initial exposure has a larger self-840 841 correction effect for liberals than for conservatives. It is worrying that extreme conservatives are both (1) more likely to consume 842 misinformation [8, 19] and (2) less likely to sever ties with those 843 who spread it. These two dynamics suggest that misinformation 844 exposure among conservatives is likely to stay at a high level. Our 845 findings suggest that future work that aims to dissolve connections 846 847 to misinformation spreaders should take into account the ideol-848 ogy of the user. On one hand, targeting left-leaning users is more likely to result in successful dissolution. On the other hand, external 849 interventions are more needed for conservative users. 850

To summarize, we observed high persistence of misinformation ties and asymmetries across partisan lines. The stability of misinformation ties also points to the importance of stopping misinformation exposure in the first place.

8 LIMITATIONS

This work has several limitations. First, our method of identifying misinformation was 'high precision' but 'low recall'; we cannot claim our results generalize to *all* spreaders of misinformation. Second, different dynamics may hold for misinformation spreaders with more than 20K followers. We were interested in 'ordinary' and 'non-celebrity' users. Third, we refer to cases where the edge between (User $A \Rightarrow$ User B) disappeared as *unfollowing*. But API limits prevent us from determining if User B *blocked* User A or manually removed User A as a follower. We suspect these cases are less common, and they are tie dissolutions, nonetheless. Fourth, our analysis concerns one stretch of time on one platform. It is

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

871

872

873

874

875

important to replicate findings across time and platforms. However, API restrictions make this increasingly difficult. Despite these limitations, this work provides a large-scale view of misinformation unfollowing.

REFERENCES

- Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. 2012. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*. ACM, Lyon France, 519–528. https://doi.org/10. 1145/2187836.2187907 1368 citations (Semantic Scholar/DOI) [2022-08-26].
- [2] Pablo Barberá. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political analysis* 23, 1 (2015), 76–91. Publisher: Cambridge University Press.
- [3] Andrei Boutyline and Robb Willer. 2017. The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks. *Political Psychology* 38, 3 (2017), 551–569. https://doi.org/10. 1111/pops.12337 287 citations (Semantic Scholar/DOI) [2023-007] _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/pops.12337.
- [4] Annie Y. Chen, Brendan Nyhan, Jason Reifler, Ronald E. Robertson, and Christo Wilson. 2023. Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels. *Science Advances* 9, 35 (Aug. 2023), eadd8080. https://doi.org/10.1126/sciadv.add8080 1 citations (Semantic Scholar/DOI) [2023-09-06] Publisher: American Association for the Advancement of Science.
- [5] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* 118, 9 (March 2021), e2023301118. https://doi.org/10.1073/pnas.2023301118 480 citations (Semantic Scholar/DOI) [2023-09-07] Publisher: Proceedings of the National Academy of Sciences.
- [6] Seth Flaxman, Sharad Goel, and Justin M. Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320. https://doi.org/10.1093/poq/nfw006 p0 Publisher: Oxford University Press.
- [7] Shane Frederick. 2005. Cognitive Reflection and Decision Making. Journal of Economic Perspectives 19, 4 (Dec. 2005), 25–42. https://doi.org/10.1257/ 089533005775196732 210 citations (Semantic Scholar/DOI) [2023-09-07].
- [8] Andrew Guess, Brendan Nyhan, and Jason Reifler. 2018. Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign. Report. Dartmouth College. https://apo.org.au/node/ 126961
- [9] Andrew M. Guess, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Edward Kennedy, Young Mie Kim, David Lazer, Devra Moehler, Brendan Nyhan, Carlos Velasco Rivera, Jaime Settle, Daniel Robert Thomas, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. 2023. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* 381, 6656 (July 2023), 398–404. https://doi.org/10.1126/science.abp9364 5 citations (Semantic Scholar/DOI) [2023-09-06] Publisher: American Association for the Advancement of Science.
- [10] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *Proceedings* of the ACM on Human-Computer Interaction 4, CSCW1 (May 2020), 48:1–48:27. https://doi.org/10.1145/3392854 114 citations (Semantic Scholar/DOI) [2023-08-25].
- [11] Roland Imhoff and Pia Lamberty. 2020. A Bioweapon or a Hoax? The Link Between Distinct Conspiracy Beliefs About the Coronavirus Disease (COVID-19) Outbreak and Pandemic Behavior. Social Psychological and Personality Science 11, 8 (Nov. 2020), 1110-1118. https://doi.org/10.1177/1948550620934692 Publisher: SAGE Publications Inc.
- [12] Ben Kaiser, Jerry Wei, Eli Lucherini, Kevin Lee, J. Nathan Matias, and Jonathan Mayer. 2021. Adapting security warnings to counter online disinformation. In 30th USENIX Security Symposium (USENIX Security 21). 1163–1180.
- [13] Johannes Kaiser, Cristian Vaccari, and Andrew Chadwick. 2022. Partisan Blocking: Biased Responses to Shared Misinformation Contribute to Network Polarization on Social Media. *Journal of Communication* 72, 2 (April 2022), 214–240. https://doi.org/10.1093/joc/jqac002 4 citations (Semantic Scholar/DOI) [2023-09-12].
- [14] Su Jung Kim. 2016. A repertoire approach to cross-platform media use behavior. New Media & Society 18, 3 (March 2016), 353–372. https://doi.org/10.1177/ 1461444814543162 95 citations (Semantic Scholar/DOI) [2023-09-07] Publisher: SAGE Publications.
- [15] Haewoon Kwak, Hyunwoo Chun, and Sue Moon. 2011. Fragile online relationship: a first look at unfollow dynamics in twitter. In *Proceedings of the SIGCHI*

925

926

927

⁹Though this may also be indicative of information overload.

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

986

- Conference on Human Factors in Computing Systems. ACM, Vancouver BC Canada, 1091–1100. https://doi.org/10.1145/1978942.1979104 161 citations (Semantic Scholar/DOI) [2022-05-08].
- [16] Hai Liang and King-wa Fu. 2017. Information overload, similarity, and redundancy: Unsubscribing information sources on twitter. *Journal of Computer-Mediated Communication* 22, 1 (2017), 1–17. https://doi.org/10.1111/jcc4.12178
- [17] Suman Kalyan Maity, Ramanth Gajula, and Animesh Mukherjee. 2018. Why Did They #Unfollow Me? Early Detection of Follower Loss on Twitter. In Proceedings of the 2018 ACM Conference on Supporting Groupwork (GROUP '18). Association for Computing Machinery, New York, NY, USA, 127–131. https://doi.org/10. 1145/3148330.3154514 3 citations (Semantic Scholar/DOI) [2022-06-21].
- [18] Parul Malik and Seungyoon Lee. 2020. Follow Me Too: Determinants of Transitive Tie Formation on Twitter. Social Media + Society 6, 3 (July 2020), 2056305120939248. https://doi.org/10.1177/2056305120939248 1 citations (Semantic Scholar/DOI) [2022-06-03] Publisher: SAGE Publications Ltd.
- [19] Ryan C. Moore, Ross Dahlke, and Jeffrey T. Hancock. 2023. Exposure to untrustworthy websites in the 2020 US election. *Nature Human Behaviour* 7, 7 (July 2023), 1096–1105. https://doi.org/10.1038/s41562-023-01564-2 6 citations (Semantic Scholar/DOI) [2023-09-06] Number: 7 Publisher: Nature Publishing Group.
- [20] Mohsen Mosleh, Gordon Pennycook, Antonio A. Arechar, and David G. Rand. 2021. Cognitive reflection correlates with behavior on Twitter. *Nature Communications* 12, 1 (Feb. 2021), 921. https://doi.org/10.1038/s41467-020-20043-0 70 citations (Semantic Scholar/DOI) [2023-09-05] Number: 1 Publisher: Nature Publishing Group.
- [21] Kostantinos Papadamou, Savvas Zannettou, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Michael Sirivianos. 2022. "It Is Just a Flu": Assessing the Effect of Watch History on YouTube's Pseudoscientific Video Recommendations. Proceedings of the International AAAI Conference on Web and Social Media 16 (May 2022), 723–734. https://doi.org/10.1609/icwsm.v16i1.19329 22 citations (Semantic Scholar/DOI) [2023-08-25].
- [22] Gordon Pennycook, Tyrone D. Cannon, and David G. Rand. 2018. Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology*. *General* 147, 12 (Dec. 2018), 1865–1880. https://doi.org/10.1037/xge0000465 559 citations (Semantic Scholar/DOI) [2023-01-08].
- [23] Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles, and David G. Rand. 2021. Shifting attention to accuracy can reduce misinformation online. *Nature* 592, 7855 (2021), 590–595. https://doi.org/10. 1038/s41586-021-03344-2
- [24] Gordon Pennycook and David G. Rand. 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188 (July 2019), 39–50. https://doi.org/10.1016/j.cognition. 2018.06.011 620 citations (Semantic Scholar/DOI) [2022-06-03].
- [25] Gordon Pennycook and David G. Rand. 2021. The Psychology of Fake News. Trends in Cognitive Sciences 25, 5 (May 2021), 388–402. https://doi.org/10.1016/ j.tics.2021.02.007 213 citations (Semantic Scholar/DOI) [2023-09-05] Publisher: Elsevier.
- [26] Gordon Pennycook and David G. Rand. 2022. Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications* 13, 1 (April 2022), 2333. https://doi.org/10.1038/s41467-022-30073-5 29 citations (Semantic Scholar/DOI) [2023-09-26] Number: 1 Publisher: Nature Publishing Group.
- [27] Ronald E. Robertson, Jon Green, Damian J. Ruck, Katherine Ognyanova, Christo Wilson, and David Lazer. 2023. Users choose to engage with more partisan news than they are exposed to on Google Search. *Nature* 618, 7964 (June 2023), 342–348. https://doi.org/10.1038/s41586-023-06078-5 5 citations (Semantic Scholar/DOI) [2023-08-25] Number: 7964 Publisher: Nature Publishing Group.
- [28] Jon Roozenbeek, Claudia R. Schneider, Sarah Dryhurst, John Kerr, Alexandra L. J. Freeman, Gabriel Recchia, Anne Marthe van der Bles, and Sander van der Linden. 2020. Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science* 7, 10 (2020), 201199. https://doi.org/10.1098/rsos.201199 Publisher: Royal Society.
- [29] Dominic Spohr. 2017. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review* 34, 3 (Sept. 2017), 150–160. https://doi.org/10.1177/0266382117722446 238 citations (Semantic Scholar/DOI) [2023-09-07] Publisher: SAGE Publications Ltd.
- [30] Lu Tang, Kayo Fujimoto, Muhammad (Tuan) Amith, Rachel Cunningham, Rebecca A. Costantini, Felicia York, Grace Xiong, Julie A. Boom, and Cui Tao.
 2021. "Down the Rabbit Hole" of Vaccine Misinformation on YouTube: Network Exposure Study. *Journal of Medical Internet Research* 23, 1 (Jan. 2021), e23262.
 https://doi.org/10.2196/23262 40 citations (Semantic Scholar/DOI) [2023-09-06]
 Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Label: Journal
 Medical Internet Research Publications Inc., Toronto, Canada.
- [31] Edwin B. Wilson. 1927. Probable Inference, the Law of Succession, and Statistical Inference. J. Amer. Statist. Assoc. 22, 158 (1927), 209–212. https://doi.org/10. 2307/2276774 Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

- [32] Bo Xu, Yun Huang, Haewoon Kwak, and Noshir Contractor. 2013. Structures of broken ties: exploring unfollow behavior on twitter. In *Proceedings of the 2013* conference on Computer supported cooperative work - CSCW '13. ACM Press, San Antonio, Texas, USA, 871. https://doi.org/10.1145/2441776.2441875 42 citations (Semantic Scholar/DOI) [2022-05-08].
- [33] Joseph Yoo, Margaret Yee Man Ng, and Thomas Johnson. 2018. Social networking site as a Political Filtering Machine: Predicting the Act of Political Unfriending and Hiding on Social Networking Sites. *The Journal of Social Media in Society* 7, 2 (Dec. 2018), 92–119. https://www.thejsms.org/index.php/JSMS/article/view/359 libs more likely to unfriend.

A APPENDIX

Table 2: These words were generated by prompting GPT-3 with 'Here are a list of words people use when they do not agree with a statement: lie, misinformation, debunked, false. What are other words?' We used the text-davinci-003 model with a temperature of 0.7.

Baseless	Disproved	Deceit	
Deceptive	Disputed	Distorted	
Delusion	Erroneous	Fabricated	
False	Falsehood	Fictitious	
Flawed	Hoax	Implausible	
Inaccurate	Incorrect	Lie	
Misinformation		Misrepresent	
Myth	No Evidence	Not True	
Refuted	Unfounded	Unreliable	
Unsubstantiate	ed Untrue	Unverified	
Fake	Fake News	Dubious	
nange Spreader Tweet Count - N Spreader Following - Spreader Following Count - Follower Follower Count -		Hall	
Follower Tweet Count -		Sample	
Follower Following Count -		All Followers	
		A	
Spreader Tweet Count -			
Spreader Tweet Count - Spreader Follower Count -			

Figure 11: We estimated a truncated version of our baseline specification (removing any ideology variables) on both the Modeling Sample and a full sample (i.e: including followers excluded from the Modeling Sample because their ideology could not be matched). Non-binary variables were z-scored. Coefficients are the effect of a +1SD change on the log odds of unfollowing. Coefficient estimates for the sample of all followers (red) were similar to those for the sample of followers whose ideology we could match (blue). Error bars are 95% CIs.

1040

1041

1042

1043

Table 3: Descriptive Statistics of Followers and Spreaders in the Modeling Sample. There are 898,701 followers, 5,334 spreaders, 3,376,785 edges. 'Is Reciprocal Tie' is a binary variable denoting if a tie is reciprocated. 'Is Liberal' is an indicator variable for when the ideology measure from Barberá is less than zero. The change in spreader tweet count was computed by subtracting the spreader's total count of tweets as of T2 from the equivalent tweet count in T1. In certain cases, the API returned fewer total tweets for a spreader in T2 than in T1. In these cases, we changed the 'Change Spreader Tweet Count' to zero.

Metric	Mean	SD	25%	50%	75%
Follower Tweet Count	18116	51193	566	3673	14984
Follower Follower Count	4311	208369	130	461	1510
Follower Following Count	2896	13362	628	1442	3107
N Spreader Following	4	13	1	1	3
Spreader Tweet Count	52311	97719	7084	20664	56644
Spreader Follower Count	1413	2389	161	536	1560
Spreader Following Count	1816	2223	387	962	2484
Change Spreader Tweet Count	4773	10009	131	1291	5018
Ideology (Positive Is Conservative)	1.65	1.57	0.06	2.29	3.0
Is Reciprocal Tie	0.69	0.46			
Is Liberal	0.24	0.43			

Table 4: Logistic regression with HC1 cluster robust errors at the spreader level and Bayesian hierarchical logistic regression with random intercepts for spreaders. Note: Non-binary variables are z-scored so coefficients can be interpreted as the accompanying change in the log odds of unfollowing with a +1SD increase in the predictor variable.

Dependent variable: Unfollowed			
(2)			
983*** (0.012)			
88*** (0.046)			
030*** (0.004)			
048^{***} (0.007)			
005** (0.002)			
150^{***} (0.017)			
.053* (0.028)			
206*** (0.028)			
72*** (0.005)			
223*** (0.013)			
067^{***} (0.007)			
313*** (0.035)			
66*** (0.021)			
061*** (0.005)			
916*** (0.018)			
0.492			
3376785			
)<			

Algorithm 1 The aim is to find the minimal number of spreaders coming from a single story that allows us to retain at least a target number of spreaders.	Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009
 Require: <i>T</i>, the target number of unique spreaders 1: Initialize λ, the maximum number of spreaders per story, to 1 2: Initialize <i>S</i>, the set of spreaders, by randomly sampling 1 user per story without replacement 	
 3: while S ≤ T do 4: Re-initialize S as the set of spreaders derived from randomly sampling λ spreaders per story without replacement 5: Increment λ by 1 	
6: end while	
	1
1	1