Online Appendices

A. Robustness Checks of the Main Results From the First Social-Nudge Experiment

A.1. Analyzing All Providers Who Were Sent at Least One Social Nudge in the Experiment

For analyses reported in the main text, we focused on providers who had never received any social nudges before the first social nudge experiment (as explained in Section 3), in order to estimate how social nudges change behavior when a platform starts to implement the social nudge function. In this section, we report the production-boosting and diffusion effects of social nudges among all providers whose followers sent them at least one social nudge during our experiment ($N = 1,946,118$), as a robustness check.

Using regression specification (1), we predicted the number of videos a provider uploaded (*Number of Videos Uploaded*) and the number of social nudges sent by a provider to other providers (*Number of Social Nudges Sent*) on the first reception day. As shown in Table 10, receiving social nudges boosted the number of videos upload on the first reception day by 9.53% (0.0222 standard deviations; $p < 0.0001$), and increased the number of social nudges sent to other providers by 13.92% (0.0323 standard deviations; $p < 0.0001$). Therefore, the immediate effects of receiving social nudges are qualitatively unchanged if we examine all providers whose followers sent them at least one social nudge during our experiment.

Table 10 Effects of Social Nudges Among All Providers Who Were Sent at Least One Social Nudge in the First Social-Nudge Experiment

Outcome Variable	Number of Videos Uploaded on the First Reception Day	Number of Social Nudges Sent on the First Reception Day
	(1)	(2)
Treatment	0.0222****	0.0323****
	(0.0014)	(0.0014)
Relative Effect Size	9.53%	13.92%
Observations	1,946,118	1,946,118

Note: Number of Videos Uploaded and Number of Social Nudges Sent were standardized to have a unit standard deviation before entering the regressions. Columns (1)–(2) include all providers whose followers sent them at least one social nudge during our first social-nudge experiment. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

A.2. Predicting Production Within 24 Hours Following the First Nudge

In the main text, we examined videos providers uploaded on the first reception day, defined as the calendar date the first social nudge was sent to them during the experiment. As a robustness

Table 11 Effects of Social Nudges on Content Production Within 24 Hours Following the First Nudge

Outcome Variable	Number of Videos Uploaded	Upload Incidence
	Within 24 Hours Following the First Nudge	
	(1)	(2)
Treatment	0.0297****	0.0131****
	(0.0020)	(0.0006)
Relative Effect Size	12.45%	13.22%
Observations	993,676	993,676

Notes: Number of Videos Uploaded Within 24 Hours Following the First Nudge was standardized to have a unit standard deviation before entering the regression. Columns (1)–(2) include all providers in the sample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

check, we tracked the number of videos each provider uploaded during 24 hours since the first social nudge. We constructed two outcome variables: *Number of Videos Uploaded Within 24 Hours Following the First Nudge* and *Upload Incidence Within 24 Hours Following the First Nudge*. We predicted these outcome variables using regression specification (1).

As shown in Table 11, treatment providers boosted the number of videos uploaded within 24 hours following the first social nudge by 12.45% (0.0297 standard deviations; p<0.0001; column (1)) and increased their likelihood of uploading any video within 24 hours following the first social nudge by 13.22% (1.31 percentage points; p<0.0001; column (2)). These results suggest that the positive effect of social nudges on content production is robust in terms of this alternative time frame.

B. The Second Social-Nudge Experiment as a Replication

We conducted another experiment to replicate the effects of social nudges on production and diffusion observed in the first field experiment (Sections 4 and 5). The replication experiment lasted from 5pm on September 14, 2018 to the end of September 20, 2018. It lasted longer than the main experiment and targeted a nonoverlapping but smaller sample of providers than the main experiment.[1] Providers targeted by the replication experiment were randomly assigned into either the treatment condition or the control condition. Similar to our main experiment (Section 3), our analyses of the replication experiment focused on providers who satisfied two criteria: (1) at least one of their followers sent them a social nudge during the experimental period, and (2) they had never received any social nudges before the experiment. Our final analysis sample consisted of 678,090 qualified providers, among whom 338,415 were in the treatment condition and 339,675 were in the control condition.

[1] We first randomly sampled a portion of providers to be included in the main experiment. Then among the remaining providers, we randomly sampled a smaller portion of providers to be involved in the replication experiment.

B.1. Direct Effects of Social Nudges on Content Production (Replicated)

We first examined the production boosting effect of receiving social nudges over time in the second experiment. Specifically, using regression specification (1), we predicted the number of videos uploaded each day from the first reception day on until the first day when the difference between two conditions was not statistically significant. We report the estimation results in Table 12, which shows that the effect sizes observed in the second experiment are comparable to the effect sizes observed in the main experiment (Table 4). Therefore, our results on the direct production-boosting effect of social nudges are robust.

Table 12 Over-Time Direct Effects of Social Nudges on Content Production (Replicated)

Outcome Variable	Number of Videos Uploaded			
	On Day 1 (First Reception Day)	On Day 2	On Day 3	On Day 4
	(1)	(2)	(3)	(4)
Treatment	0.0228****	0.0107****	0.0083***	0.0033
	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Relative Effect Size	11.83%	7.79%	3.96%	
Observations	678,090	678,090	678,090	678,090

Notes: Number of Videos Uploaded was standardized to have a unit standard deviation before entering the regressions. The unit of analysis for all columns was a provider on Day t relative to the first reception day, where $t = 1$ means the first reception day. Columns (1)–(4) include all providers in our sample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

B.2. Effects of Social Nudges on Nudge Diffusion (Replicated)

Next we tested the diffusion effect of receiving social nudges over time in the second experiment. Specifically, for each day t starting from the first reception day, we predicted the number of social nudges sent on that day using regression specification (1) until the first day when the difference between two conditions was not statistically significant. We report the estimation results in Table 13, which shows that the effect sizes observed in the second experiment are comparable to the effect sizes observed in the main experiment (Table 6). Therefore, our results on the diffusion of social nudges are robust.

C. Additional Analyses about the Direct Effects of Social Nudges

C.1. Addressing an Alternative Explanation about Control Providers' Resentment

As explicated in Section 1, we expected social nudges to boost content production because providers receiving social nudges might feel more valued by others and thus more motivated to supply effort. However, one potential alternative explanation for our observed difference in video production between treatment and control providers is that through other ways beyond the message center, control providers realized that their followers sent them social nudges but they could not receive

Table 13 Over-Time Effects of Social Nudges on Nudge Diffusion (Replicated)

Outcome Variable	Number of Social Nudges Sent				
	On Day 1 (First Reception Day)	On Day 2	On Day 3	On Day 4	On Day 5
	(1)	(2)	(3)	(4)	(5)
Treatment	0.0325****	0.0215****	0.0084***	0.0057*	0.0039
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Relative Effect Size	16.25%	14.16%	5.78%	4.02%	
Observations	678,090	678,090	678,090	678,090	678,090

Notes: Number of Social Nudges Sent was standardized to have a unit standard deviation before entering the regressions. The unit of analysis for all columns was a provider on Day t relative to the first reception day, where $t = 1$ refers to the first reception day. Columns (1)–(5) include all providers in our sample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

these social nudges, which might make control providers feel resentful toward the platform and thus reduce their production. As mentioned in Section 3, the only way for users to directly communicate with each other on the platform is through the private-message function. It is plausible that during our experiment followers privately messaged providers after sending them social nudges, which led control providers to realize that they were blocked from viewing nudges. To address this alternative explanation, we conducted two sets of additional analyses.

First, we examined how private messages influenced control providers. If control providers knew via private messages that their followers sent them a social nudge but they were not allowed to see the nudge and if this created resentment, we should expect that receiving private messages from followers who sent them social nudges during the experiment negatively impacted control providers' content production. To test this possibility, we used the DiD method. This analysis included two observations per control provider, with one observation corresponding to the first reception day and one observation corresponding to the day before the experiment. For each observation of provider i, her content production equaled the number of videos uploaded on the corresponding day (either the first reception day or the day before the experiment). The DiD regression specification is formulated as below

$$\begin{aligned} \textit{Outcome Variable}_{it} = {} & \beta_0 + \beta_1 \textit{Private Messages Incidence}_i + \beta_2 \textit{First Reception Day}_{it} \\ & + \beta_3 \textit{Private Messages Incidence}_i * \textit{First Reception Day}_{it} + \epsilon_{it} \end{aligned} \tag{12}$$

whereby $\textit{Private Messages Incidence}_i$ was a binary variable that equaled one if the follower who sent provider i the first social nudge in the experiment (i.e., provider i's first social-nudge sender) also sent any private messages to i between the start date of the experiment and provider i's first reception day (including both ends) and zero otherwise[2]; and $\textit{First Reception Day}_{it}$ was a binary

[2] To protect user privacy, Platform O could not share the content of private messages with us. Thus, we could not use content analysis to identify whether each provider's first social-nudge sender told the provider about the nudge in their private communications, but instead we used whether a provider received private messages from their first social-nudge sender as a proxy, since receiving such private messages was the only plausible channel for control providers to find out the blocking of social nudges.

Table 14 The Role of Private Messages in Content Production Among Control Providers

Panel A: DiD Analysis about Private Messages Among Control Providers	
Outcome Variable	Number of Videos Uploaded
	(1)
Private Messages Incidence	0.1603****
	(0.0116)
First Reception Day	0.1258****
	(0.0020)
Private Messages Incidence * First Reception Day	0.2248****
	(0.0203)
Observations	993,400

Panel B: Comparison of Two Subsamples Based on Private Message Incidence		
Outcome Variable	Number of Videos Uploaded	
Subsample	*Providers Who Received Any Private Messages From the First Social-Nudge Sender*	*Providers Who Received No Private Messages From the First Social-Nudge Sender*
	(1)	(2)
Treatment	0.0710***	0.0235****
	(0.0192)	(0.0020)
Relative Effect Size	14.15%	12.36%
Observations	28,142	965,534

Notes: Number of Videos Uploaded was standardized to have a unit standard deviation before entering the regressions. Panel A includes all control providers in our sample, with each control provider contributing two observations. Standard errors in Panel A are clustered at the provider level. Column (1) in Panel B includes treatment and control providers who received any private messages from their first social-nudge sender between the start date of the experiment and the first reception day, and column (2) in Panel B includes treatment and control providers who did not receive any private messages from their first social-nudge sender in this period. Robust standard errors reported in the parentheses in Panel B. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

variable that equaled one if an observation corresponded to the first reception day and zero if the observation corresponded to the day before the experiment. We clustered standard errors by provider.

As shown in Table 14 Panel A, since the coefficient on the interaction between $Private\ Messages\ Incidence_i$ and $First\ Reception\ Day_{it}$ is positive ($p < 0.0001$), we have no evidence to suggest that receiving private messages from followers who sent them social nudges during the experiment would reduce control providers' content production.

Second, we split the whole provider sample in our experiment into two subsamples based on whether each provider's first social-nudge sender sent any private messages to the provider between the start date of the experiment and provider i's first reception day (including both ends). Within each subsample, we compared the *Number of Videos Uploaded* on the first reception day between treatment and control conditions using regression specification (1).

As shown in Table 14 Panel B, no matter whether a provider got any private messages from their first social-nudge sender, receiving social nudges increased treatment providers' content production, relative to control providers' (both p-values < 0.001). The relative effect size is very similar among providers who got private messages from their first social-nudge sender (14.15% as shown in column (1)) and among providers who did not get private messages from their first social-nudge sender (12.36% as shown in column (2)).

Altogether, these results do not support the alternative explanation: it is unlikely that communication from followers via private messages led control providers to find out they could not view social nudges, elicited resentment, and thus reduced their motivation to produce videos. In addition, note that all broadcasters selected into our analysis sample had not received any social nudges before the experiment. Thus, it is unlikely for providers in our analysis to naturally realize that they did not receive social nudges during the experiment without any hints from followers.

C.2. Role of Likes and Comments

As mentioned in Section 3, when viewers watch a video, they can mark that they like the video and leave comments below the video. Since receiving social nudges could immediately boost video production (see Section 4.1), nudge recipients may also immediately receive more likes and comments due to the increased number of videos uploaded. Such positive feedback from viewers may in turn motivate nudge recipients to produce more videos going forward. This raises the question of whether and to what extent the immediate increase in likes and comments due to receiving social nudges drives the observed over-time effect of social nudges on nudge recipients' content production (as shown in Section 4.3).

To answer this question, we first tested whether receiving social nudges led the recipient to obtain more likes and comments. For each provider on her first reception day, we calculated the number of likes and comments she obtained that day (*Number of Likes on the First Reception Day* and *Number of Comments on the First Reception Day*, respectively). We winsorzied these two variables at the 95th percentile of their respective nonzero values because they were highly skewed (due to a small number of providers being too popular). Using regression specification (1), we predicted these two outcome variables. As shown in Table 15 Panel A, treatment providers obtained more likes than control providers on the first reception day by 0.0112 standard deviations, or 4.52% ($p < 0.0001$; column (1)); treatment providers obtained more comments than control providers on the first reception day by 0.0108 standard deviations, or 5.40% ($p < 0.0001$; column (2)). Hence, treatment providers obtained more likes and comments after they received social nudges, relative to control providers.

We next tested how much the immediate increase in likes and comments due to the receipt of social nudges contributed to the effects of receiving social nudges on content production during the few days after the first reception day. In one series of regressions, we predicted the number of videos uploaded the day following the first reception day using regression specification (1), and we

Table 15 Effects of Social Nudges on Content Production With or Without Controlling for the Role of Likes and Comments

Panel A		
Outcome Variable	Number of Likes on the First Reception Day	Number of Comments on the First Reception Day
	(1)	(2)
Treatment	0.0112****	0.0108****
	(0.0020)	(0.0020)
Relative Effect Size	4.52%	5.40%
Observations	993,676	993,676

Panel B				
Outcome Variable	Number of Videos Uploaded Following the First Reception Day			
	(1)	(2)	(3)	(4)
Treatment	0.0129****	0.0091****	0.0094****	0.0090****
	(0.0020)	(0.0019)	(0.0019)	(0.0019)
Number of Likes on the First Reception Day		0.3374****		0.2299****
		(0.0009)		(0.0018)
Number of Comments on the First Reception Day			0.3225****	0.1253****
			(0.0009)	(0.0018)
Relative Effect Size	5.29%	3.74%	3.87%	3.68%
Observations	993,676	993,676	993,676	993,676

Panel C				
Outcome Variable	Number of Videos Uploaded on the Second Day Following the First Reception Day			
	(1)	(2)	(3)	(4)
Treatment	0.0065**	0.0061**	0.0063**	0.0061**
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Number of Likes the Day Following the First Reception Day		0.1209****		0.0853****
		(0.0010)		(0.0020)
Number of Comments the Day Following the First Reception Day			0.1152****	0.0408****
			(0.0010)	(0.0020)
Relative Effect Size	2.54%	2.40%	2.46%	2.42%
Observations	993,676	993,676	993,676	993,676

Notes: All continuous variables were standardized to have a unit standard deviation before entering the regressions. All columns in Panels A, B, and C include all providers in the sample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

compared the regression results with or without controlling for the number of likes or comments a provider obtained the day before (i.e., *Number of Likes on the First Reception Day* and *Number of Comments on the First Reception Day*). According to Table 15 Panel B, without controlling for likes or comments on the first reception day, we find that receiving social nudges boosted the number of videos uploaded by 0.0129 standard deviations (5.29%, $p < 0.0001$; column (1) of Panel A) on the day following the first reception day. This effect reduced but remained statistically significantly when we controlled for the number of likes a provider got on her first reception day (column (2)), the number of comments she got on her first reception day (column (3)), or both (column (4)).

Further, we predicted the *number of videos uploaded* on the *second day* following the first reception day using regression specification (1), and we compared the regression results with or without controlling for the number of likes or comments a provider obtained the day before (i.e., on the day following the first reception day). As shown in Table 15 Panel C, the positive effect of receiving social nudges on content production two days later reduced only slightly when we added these control variables.

Altogether, these findings indicate that getting more likes and comments after treatment providers uploaded more videos in response to social nudges contributed to some extent to the over-time effect of social nudges on content production. However, increased likes and comments are not the only reason, neither are the primary reason why the effect of receiving social nudges on content production lasted for days, since we observe only a slight to moderate decrease in the magnitude of the production-boosting effect of social nudges after the first reception day when we controlled for likes and comments providers obtained earlier. This suggests that receiving social nudges per se is sufficient to motivate video production a few days, even without additional positive feedback providers receive due to their increased production along the way[3].

C.3. Do Social Nudges Cannibalize Likes/Comments?

Sending social nudges is a new way for viewers to interact with providers and express appreciation of their videos on top of the existing mechanisms including likes and (positive) comments. A natural question is whether viewers will mark fewer likes and leave fewer comments once they begin to use social nudges. In other words, will social nudges cannibalize the use of likes and comments? We address this question in two aspects. First, from the perspective of providers, we find that receiving social nudges did not lead providers to receive fewer likes and comments; if anything, we causally document that social nudges brought providers more likes and comments in the next couple of days (see Online Appendix C.2). Second, from the perspective of viewers, we tested whether sending social nudges decreased the usage of likes and comments, and we report this analysis in this subsection.

When we investigated this question, Platform O no longer stored detailed data about likes and comments that took place around our experimental period; thus, we leveraged observational data in December 2021. We first randomly sampled $10,000,000$ viewers who logged onto Platform O between December 1, 2021 and December 7, 2021. For each viewer in the sample, we calculated

[3] We conducted these analyses because it is *theoretically* interesting to tease apart whether the lingering effect of receiving social nudges on content production is driven by providers obtaining an increased amount of positive feedback on their videos or by providers feeling motivated by social nudges per se. But *practically speaking*, we believe the feedback mechanism is meaningful because increased likes and comments are *consequences* of the initial boost in content production in response to social nudges. Thus, we do not distinguish these mechanisms when we calculate the global impact of social nudges in Section 6.

Table 16 Users Who Sent Social Nudges Did Not Reduce the Usage of Likes/Comments

Outcome Variable	Number of Likes Marked (1)	Number of Comments Left (2)
Incidence of Sending Social Nudges	0.2594^{****}	0.1836^{****}
	(0.0023)	(0.0023)
Post	-0.0057^{****}	-0.0008
	(0.0006)	(0.0004)
Incidence of Sending Social Nudges * Post	0.0556^{****}	0.0323^{****}
	(0.0010)	(0.0010)
Observations	1,412,164	1,412,164

Notes: Both outcome variables are standardized to have a unit deviation. Robust standard errors clustered at the viewer level are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

whether she sent at least one social nudge in that period (*Incidence of Sending Social Nudges*), and obtained a host of features, including gender, age, her residential community type (i.e., countryside, town, or urban area), the tier of city she lives in (e.g., first tier, second tier, etc.), the number of followers she had on November 30, 2021, the number of users she was following on November 30, 2021, whether she had uploaded any video by November 30, 2021, how long she was active on Platform O in the previous week (November 24—November 30, 2021), how long she watched videos on Platform O in the previous week, and whether she was banned from social interactions (such as marking likes and sending comments) on Platform O in the previous week. Via Coarsened Exact Matching (CEM, e.g., Iacus et al. 2012), we constructed two matched groups: the "treatment" group who sent at least one social nudge in the week of December 1—December 7, 2021 (353,041 viewers; *Incidence of Sending Social Nudges* = 1), and the "control" group who did not send any social nudge that week (353,041 viewers; *Incidence of Sending Social Nudges* = 0). For each viewer i in the matched sample and for each outcome variable, we measured two observations, with one observation corresponding to the week of December 1—December 7, 2021 (*Post* = 1), and the other observation corresponding to the prior week (November 24—November 30, 2021; *Post* = 0). The DiD regression specification is formulated as

$$\begin{aligned} Outcome\ Variable_{it} = {} & \beta_0 + \beta_1 Incidence\ of\ Sending\ Social\ Nudges_i + \beta_2 Post_{it} \\ & + \beta_3 Incidence\ of\ Sending\ Social\ Nudges_i * Post_{it} + \epsilon_{it} \end{aligned} \tag{13}$$

where the number of likes marked by viewer i during week t ($Number\ of\ Likes\ Marked_{it}$) and the number of comments left by viewer i during week t ($Number\ of\ Comments\ Left_{it}$) serve as the outcome variables. We clustered standard errors by viewers.

As shown in Table 16, the coefficient on the interaction term between $Incidence\ of\ Sending\ Social\ Nudges_i$ and $Post_{it}$ is positive and significant for both outcome variables (both p-values < 0.0001). This suggests that viewers who sent any social nudges marked more likes and left more comments, compared to viewers who did not send social nudges. Thus, combining the identification strategies of matching and DiD, we find no evidence for cannibalization of social nudges on likes and comments.

C.4. Effects of Social Nudges Across Providers With Different Baseline Productivity

The scant prior literature that has examined the causal effects of peer recognition without financial incentives has not provided a clear answer to the question of whether peer recognition can boost recipients' production (Restivo and van de Rijt 2014, Gallus et al. 2020). In a field experiment involving top 10% of providers to Wikipedia, Restivo and van de Rijt (2014) found that peer recognition increased only the most productive 1% of content providers but not providers ranked at the 91st–99th percentile. To test whether the production-boosting effect of social nudges can generalize to providers with different levels of baseline productivity, we divided the providers in our sample into three subsamples: providers whose historical production—the number of videos uploaded during the week prior to the experiment—was (1) below or at the 90th percentile of the distribution of historical production across all providers in the sample ("*low-productivity providers*"; ignored by Restivo and van de Rijt (2014)), (2) in the 91st–99th percentile range ("*medium-productivity providers*"; comparable to the definition of less-productive providers in Restivo and van de Rijt (2014)), and (3) at the 100th percentile ("*high-productivity providers*"; comparable to the most productive 1% providers in Restivo and van de Rijt (2014)). For each subsample, we separately estimated the effect of receiving social nudges on the day of nudges being sent. That is, we predicted the number of videos uploaded on the first reception day using regression specification (1).

As shown in Table 17, the number of videos uploaded on the first reception day was boosted by 19.37% among low-productivity providers (0.0220 standard deviations; $p < 0.0001$; column (1)), by 6.91% among medium-productivity providers (0.0577 standard deviations; $p < 0.0001$; column (2)), and by 7.67% among high-productivity providers (0.2145 standard deviations; $p < 0.05$; column (3)). Overall, these results suggest that not only the most productive 1% of providers but also providers whose historical production was in the 0th-99th percentile range were also motivated by receiving social nudges.

C.5. Comparison Between Social Nudges and Platform-Initiated Nudges

Receiving a social nudge and its implied recognition may make providers feel more valued, thus motivating them to produce new videos. However, such information communicated via social nudges from neighbors may not be passed on by nudges from the platform to encourage production. To explore whether social nudges have effects beyond regular nudges sent from companies, we leveraged another randomized field experiment that tested the effects of receiving nudges from

Table 17 Direct Effects of Social Nudge Across Providers With Different Historical Production Levels

Outcome Variable	Number of Videos Uploaded on the First Reception Day		
Subsample	*Low-Productivity Providers*	*Medium-Productivity Providers*	*High-Productivity Providers*
	(1)	(2)	(3)
Treatment	0.0220****	0.0577****	0.2145*
	(0.0015)	(0.0124)	(0.0926)
Relative Effect Size	19.37%	6.91%	7.67%
Observations	901,286	83,838	8,552

Notes: Number of videos uploaded was standardized to have a unit standard deviation before entering the regressions. Each column includes the providers in the corresponding subsample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

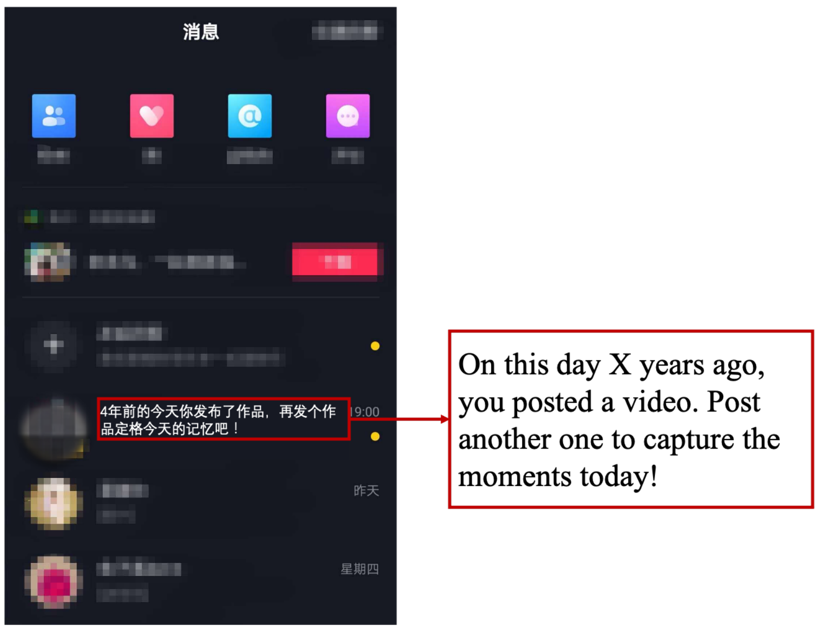

Figure 3 A Platform-Initiated Nudge[4]

Platform O. We refer to this experiment as the *platform-initiated nudge experiment*. The platform-initiated nudge experiment randomly targeted a subset of users on Platform O, no matter whether they were targeted by the social-nudge experiments. We note that comparing the results of our main social-nudge experiment versus the platform-initiated nudge experiment does not causally estimate the difference between social nudges and platform-initiated nudges. Specifically, since the two experiments were conducted in different time periods and providers were not randomly assigned to receive one of these two types of nudges, the providers included in the two experiments were not exactly comparable. As described below, we sought to construct samples from the two experiments that were as comparable to each other as possible.

[4] In order to protect Platform O's identity, similar to how we dealt with Figure 1, we created Figure 3 by modifying the app interface of a widely-used video-sharing platform. Platform O has a similar app interface to Figure 3.

Experiment Design and Data. The platform-initiated nudge experiment was conducted between 9AM on July 22, 2019 and 5AM on August 30, 2019. Half of the providers were randomly assigned to the treatment condition, and the other half to the control condition. During the experiment, the platform identified providers who published a video one or more years ago exactly on the same date. For these providers, Platform O created a message that read, "On this day X years ago, you posted a video. Post another one to capture the moments today!" where "X" was filled in with the actual number of years that had elapsed.[5] The only factor that we manipulated between treatment and control providers was that Platform O actually sent out the aforementioned message to treatment providers on that date, but not to control providers. Therefore, control providers could not receive any platform-initiated nudges. Messages about platform-initiated nudges were displayed in the message center, the same as social nudges (see Figure 1 (b)).

We first selected treatment and control providers who were qualified to be sent at least one message from Platform O during our experiment. For these providers, we defined "the first reception day" as the day when they first became qualified to be sent Platform O's nudge message. The sample selection criteria and the definition of the first reception day here match our approach in the social-nudge experiments.

The Effects of Receiving Platform-Initiated Nudges on Production. We first examined the effects of receiving platform-initiated nudges on content production, both on the first reception day and in the next few days. Consistent with our analytical strategy for the social-nudge experiments (Section 4.3), we examined how receiving nudges from Platform O affected providers' production each day between treatment and control providers among the full sample of providers from the first reception day on until the first day when the difference between conditions was not statistically significant. Specifically, for each day t starting from the first reception day (where t equals $1, 2, \cdots$ and $t = 1$ refers to the first reception day itself), we predicted the number of videos uploaded on that day using regression specification (1).

We report the regression results in Table 18 Panel A. On the first reception day, the platform-initiated nudge treatment lifted the number of videos uploaded by 0.0105 standard deviations ($p < 0.0001$), which amounts to a 5.55% increase relative to the average in the control condition, as shown in column (1). On Day 2 (the day following the first reception day), the number of video uploaded was higher in the treatment condition than in the control condition by 0.0026 standard deviations, or 1.37% ($p < 0.0001$; column (2)); on Day 3 (the second day from the first reception

[5] To avoid disturbing providers, Platform O sent out a maximum of two messages to each provider in one week. Specifically, on the first day of each week during the experiment, for provider i, Platform O identified the dates during that week on which provider i uploaded any video exactly one or more years ago. If more than two dates satisfied the criterion, Platform O picked the dates on which the video uploaded exactly one or more years ago had the highest or second highest views (among all videos uploaded in the same week one or more years ago).

Table 18 Comparison of Social Nudges and Platform-Initiated Nudges

Panel A: Direct Effects of Platform-Initiated Nudges on Content Production				
Outcome Variable	Number of Videos Uploaded			
	On Day 1 (First Reception Day)	On Day 2	On Day 3	On Day 4
	(1)	(2)	(3)	(4)
Treatment	0.0105^{****}	0.0026^{****}	0.0019^{**}	0.0011
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Relative Effect Size	5.55%	1.37%	0.99%	
Observations	11,043,476	11,043,476	11,043,476	11,043,476

Panel B: Comparison of Social Nudges and Platform-Initiated Nudges Using an Overlapping Sample of Providers		
Outcome Variable	Number of Videos Uploaded On Day 1 (First Reception Day)	
	Platform-Initiated Nudges	*Social Nudges*
Treatment	0.0152	0.0216^{***}
	(0.0093)	(0.0063)
Relative Effect Size		12.35%
Observations	63,467	63,467

Notes: Number of videos uploaded was standardized to have a unit standard deviation before entering the regressions. Panel A includes all providers who satisfied sample selection criteria for the platform-initiated nudge experiment. The unit of analysis in Panel A was a provider on Day t relative to the first reception day, where $t = 1$ means the first reception day. Panel B includes providers who were selected for both the social-nudge experiment and the platform-initiated nudge experiment. The unit of analysis in Panel B was a provider on her first reception day. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

day), the increase was 0.0019 standard deviations, or 0.99% ($p < 0.01$; column (3)). The effect of receiving platform-initiated nudges on the nudge recipient's production was not significant on Day 4 (the third day after the first reception day; column (4)).

Next we compared the production-boosting effects of social nudges and platform-initiated nudges. Figure 4 displays the relative effect sizes of these two kinds of nudges, as well as the corresponding 95% confidence intervals.[6] The effect of receiving platform-initiated nudges on *Number of Videos Uploaded* was generally below that of receiving social nudges. In particular, on the reception day (i.e., $t = 1$), receiving social nudges increased the number of uploaded videos by 13.21% ($p < 0.0001$), more than twice as large as the increase of 5.55% ($p < 0.0001$) engendered by receiving platform-initiated nudges; on the day following the first reception day (i.e., $t = 2$), receiving social nudges increased the number of uploaded videos by 5.29% ($p < 0.0001$), almost three times as large as the increase of 1.37% ($p < 0.0001$) brought by receiving platform-initiated nudges; on the second day following the first reception day (i.e., $t = 3$), receiving social nudges increased the number of uploaded videos by 2.54% ($p < 0.01$), almost twice as large as the increase of 0.99% ($p < 0.01$) from by receiving platform-initiated nudges.[7]

[6] The upper (lower) bound of each 95% confidence interval in Figure 4 equaled the upper (lower) bound of the 95% confidence interval of the corresponding regression coefficient on treatment (based on raw data) divided by the average of the outcome variable in the control condition.

[7] Notably, during the four days since the first reception day (including the first reception day itself) for which we reported the day-by-day effects of both nudges, most (88%) providers were sent only one social nudge in the social

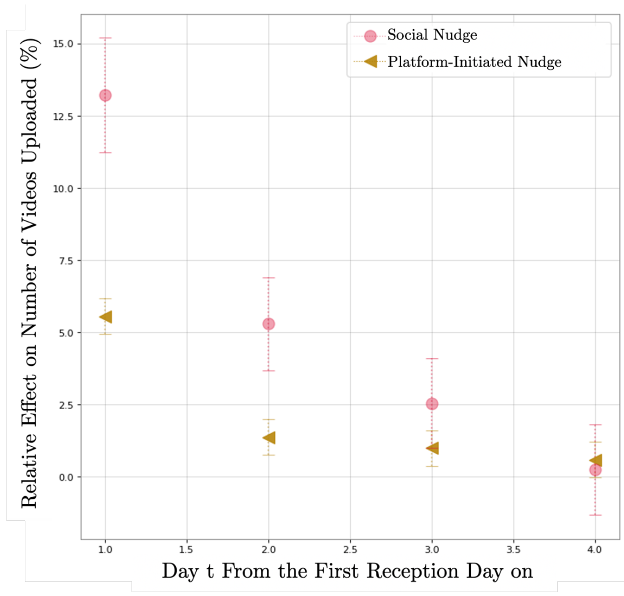

Figure 4 Comparing the Relative Effect Size of Receiving a Platform-Initiated Nudge Versus Social Nudge Over Time

Note: The error bars represent 95% confidence intervals.

In addition, as a robustness check, we analyzed only providers who were included in both the platform-nudge experiment and our first social-nudge experiment ($N = 63,467$) to as cleanly estimate the difference between these two kinds of nudges as possible. Among these overlapping providers, we re-analyzed the direct effects of receiving social nudges or platform-initiated nudges. Since the effects of social nudges and platform-initiated nudges on the recipients' production among these overlapping providers were no longer statistically significant after the first reception day, we focused on comparing these effects on the first reception day. As shown in Table 18 Panel B, the effect of receiving platform-initiated nudges on the number of videos uploaded on the first reception day was not significant ($p = 0.10$; column (1))[8], while receiving social nudges significantly boosted the number of videos uploaded on the first reception day by 0.0216 standard deviations, or 12.35% ($p < 0.001$; column (2)). These results further provide suggestive evidence that social nudges boosted providers' production to a larger extent than platform-initiated nudges.

Nevertheless, platform-initiated nudges like what Platform O tested can still be quite useful to platforms, as it is not an easy job to develop interventions that improve online platforms' operational performance. For instance, at Google and Bing, more than 10,000 online field experiments

nudge experiment, and most (89%) providers were sent only one platform-initiated nudge in the platform-initiated nudge experiment. Hence, the production-boosting effects of both social nudges and platform-initiated nudges were mostly driven by one nudge, making the comparison fair.

[8] Even if we put aside whether the estimated effect was statistically significant, receiving platform-initiated nudges was estimated to increase the number of videos uploaded on the first reception day by 0.0152 standard deviations (or 4.54%) among the overlapping sample of providers, which was still lower than the effect of receiving social nudges on the number of videos uploaded on the first reception day (i.e., 0.0216 standard deviations or 12.35%).

are conducted each year, only 10%–20% of which identify interventions with positive effects.[9] As shown above, though less effective than social nudges in boosting production, platform-initiated nudges tested on Platform O did increase production, and the magnitude of their production-boosting effect is actually comparable with the effect sizes of other subtle interventions that recent research tested on online platforms via randomized field experiments. For example, presenting users with different kinds of performance feedback on a Chinese mobile-app-based recipe crowdsourcing platform increased recipe postings by 1.8%-4.8% (Huang et al. 2019), and displaying the number of people who had applied for a job on LinkedIn increased the job application rate by 3.5% (Gee 2019).

D. Proofs for the Social Network Model

This section presents the proofs of our technical results for the social network model, as well as the details for the estimation strategy discussed in Section 6.

D.1. On Condition $\mathcal{C}$

Lemma 2 *If $(\alpha_d, \mathbf{D})$ satisfies Condition $\mathcal{C}$, the matrix $\mathbf{I}-(1/(1-\alpha_d))\mathbf{D})$ is then invertible with*

$$\left(\mathbf{I}-\frac{1}{1-\alpha_d}\cdot\mathbf{D}\right)^{-1}=\lim_{k\to+\infty}\mathbf{M}(k)=\lim_{k\to+\infty}\left(\mathbf{I}+\sum_{i=1}^{k}\frac{1}{(1-\alpha_d)^i}\mathbf{D}^i\right). \tag{14}$$

Proof of Lemma 2. By condition $\mathcal{C}$, $\lim\limits_{k\to+\infty}\mathbf{M}(k)$ exists, which we denote as $\mathbf{M}$. Therefore,

$$\lim_{k\to+\infty}\frac{1}{(1-\alpha_d)^k}\cdot\mathbf{D}^k=\mathbf{0}_{|E|\times|E|},\ \text{ where } \mathbf{0}_{|E|\times|E|} \text{ is the 0 matrix of dimension } |E|\times|E|.$$

Furthermore, we have that

$$\left(\mathbf{I}-\frac{1}{1-\alpha_d}\cdot\mathbf{D}\right)\mathbf{M}=\lim_{k\to+\infty}\left(\mathbf{I}-\frac{1}{1-\alpha_d}\cdot\mathbf{D}\right)\mathbf{M}(k)=\lim_{k\to+\infty}\left(\mathbf{I}-\frac{1}{(1-\alpha_d)^{k+1}}\cdot\mathbf{D}^{k+1}\right)=\mathbf{I},$$

where the second inequality follows from the identity $(\mathbf{I}-\mathbf{A})(\mathbf{I}+\mathbf{A}+\mathbf{A}^2+\cdots+\mathbf{A}^k)=\mathbf{I}-\mathbf{A}^{k+1}$ for any square matrix $\mathbf{A}$. Therefore, $\mathbf{I}-(1/(1-\alpha_d))\mathbf{D}$ is invertible and its inverse is given by Equation (14). □

Remark 1 *The proof of Lemma 2 also implies that, under Condition $\mathcal{C}$, $\mathbf{I}-\mathbf{D}$, is invertible with*

$$(\mathbf{I}-\mathbf{D})^{-1}=\mathbf{I}+\sum_{k=1}^{+\infty}\mathbf{D}^k\geq\mathbf{0}_{|E|\times|E|}. \tag{15}$$

[9] See `https://hbr.org/2017/09/the-surprising-power-of-online-experiments` for details.

Condition $\mathcal{C}$ is sometimes hard to verify in practice, so we leverage the concept of matrix norm to provide an easy-to-verify sufficient condition. Specifically, we denote ℓ_q-norm of matrices by $||\cdot||_q$ for any $q\in[1,+\infty]$, which is the operator norm defined through $||\mathbf{A}||_q=\sup_{\boldsymbol{z}:||\boldsymbol{z}||_q\le 1}||\mathbf{A}\boldsymbol{z}||_q$ for any squared matrix $\mathbf{A}$ and $\boldsymbol{z}$ with appropriate dimensions (Horn and Johnson 2012). Also, we say that the $(\alpha_d,\mathbf{D})$ satisfies $\mathcal{C}_q(\delta)$ for some $\delta\in(0,1)$, provided that $||(1/(1-\alpha_d))\mathbf{D}||_q\le\delta$. We note that if $(\alpha_d,\mathbf{D})$ satisfies $\mathcal{C}_q(\delta)$ for some $\delta\in(0,1)$, the inverse of $\mathbf{I}-(1/(1-\alpha_d))\mathbf{D}$ is given by Equation (14) (see, e.g., Corollary 5.6.16 and Corollary 5.6.17 of Horn and Johnson 2012), which also implies that Condition $\mathcal{C}$ is satisfied.

For Platform O, based on our estimates of $\mathbf{D}$ and α_d (see Section 6.2, Table 8 in particular), we quantify that $(\alpha_d,\mathbf{D})$ satisfies $\mathcal{C}_\infty(0.6)$.[10] Therefore, Condition $\mathcal{C}$ is satisfied for Platform O.

D.2. Proof of Theorem 1

Let us assume throughout the proof that $(\alpha_d,\mathbf{D})$ satisfies Condition $\mathcal{C}$. Also recall that the system of the social network model is defined by Equations (3) and (4).

Let us denote by $\boldsymbol{y}^*(t):=\mathbb{E}[\boldsymbol{y}(t)]$ and $\boldsymbol{x}^*(t):=\mathbb{E}[\boldsymbol{x}(t)]$. Since $\epsilon_e^y(t)$ and $\epsilon_e^x(t)$ are the random errors with a zero mean, it then follows from Equation (4) that $\boldsymbol{y}^*(t)=\boldsymbol{\mu}+\sum_{1\le s\le t-1}\alpha_d^{t-s}\mathbf{D}\boldsymbol{y}^*(s)+\mathbf{D}\boldsymbol{y}^*(t)$, or equivalently

$$\boldsymbol{y}^*(t)=(\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu}+\sum_{1\le s\le t-1}\alpha_d^{t-s}\mathbf{D}\boldsymbol{y}^*(s)\right), \tag{16}$$

where $(\mathbf{I}-\mathbf{D})^{-1}$ is well-defined by Equation (15). We also note that $\boldsymbol{y}^*(1)=(\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu}$. Similarly, by Equation (3), we have:

$$x_i^*(t)=\sum_{1\le s\le t-1}\alpha_p^{t-s}\sum_{e\in E:e_d=i}p_e y_e^*(s). \tag{17}$$

We now show that the sequence $\{\boldsymbol{y}^*(t):t\ge 1\}$ is componentwise increasing and bounded, so it converges to a limit. We relegate the proof of the following lemma after the proof of Theorem 1.

Lemma 3 *Regarding $\{\boldsymbol{y}^*(t):t\ge 1\}$, the following statements hold:*

(a) *For each $e\in E$, $y_e^*(t)$ is increasing in t ($t\in\mathbb{Z}_+$).*

(b) *For each $e\in E$ and each $t\in\mathbb{Z}_+$, $y_e^*(t)\le\mathcal{BE}_e(\mathbf{D},\boldsymbol{\mu})$.*

Lemma 3 implies that the limit of the sequence $\{\boldsymbol{y}^*(t):t\ge 1\}$ exists and is finite, which we denote as $\boldsymbol{y}^*$. Note that, by Equation (16),

$$\boldsymbol{y}^*(t+1)=(\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu}+\sum_{1\le s\le t}\alpha_d^{t+1-s}\mathbf{D}\boldsymbol{y}^*(s)\right) \tag{18}$$

[10] To protect the sensitive data, we cannot report the tightest value of δ.

and

$$\alpha_d \boldsymbol{y}^*(t) = (\mathbf{I}-\mathbf{D})^{-1}\left(\alpha_d\boldsymbol{\mu} + \sum_{1\le s\le t-1} \alpha_d^{t-s+1}\mathbf{D}\boldsymbol{y}^*(s)\right). \tag{19}$$

Taking the difference between Equation (18) from Equation (19) leads to $\boldsymbol{y}^*(t+1) - \alpha_d\boldsymbol{y}^*(t) = (\mathbf{I}-\mathbf{D})^{-1}\left((1-\alpha_d)\boldsymbol{\mu} + \alpha_d\mathbf{D}\boldsymbol{y}^*(t)\right)$. Taking $t\to+\infty$ on both sides leads to $\boldsymbol{y}^* - \alpha_d\boldsymbol{y}^* = (\mathbf{I}-\mathbf{D})^{-1}\left((1-\alpha_d)\boldsymbol{\mu}+\alpha_d\mathbf{D}\boldsymbol{y}^*\right)$. Reorganizing the terms, we have $\boldsymbol{y}^* = \boldsymbol{\mu} + (1/(1-\alpha_d))\mathbf{D}\boldsymbol{y}^*$. Since $\mathbf{I}-(1/(1-\alpha_d))\mathbf{D}$ is invertible by Lemma 2, $\boldsymbol{y}^* = (\mathbf{I}-(1/(1-\alpha_d))\mathbf{D})^{-1}\boldsymbol{\mu} = \mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$.

Finally, it remains to show that $\lim_{t\to\infty}\mathbb{E}[x(t)] = x^*$ and $x^* = \boldsymbol{\eta}^T\boldsymbol{y}^*$. With the same argument as the proof for $\boldsymbol{y}^*$, we have, by Equation (17), $\lim_{t\to\infty}\mathbb{E}[\boldsymbol{x}(t)] = \boldsymbol{x}^*$ where $x_i^* = \frac{1}{1-\alpha_p}\sum_{e\in E:e_d=i} p_e y_e^*$, for all $i\in V$. Therefore,

$$x^* = \sum_{i\in V} x_i^* = \sum_{e\in E}\frac{1}{1-\alpha_p}p_e y_e^* = \boldsymbol{\eta}^T\boldsymbol{y}^*.$$

To avoid repetition, we omit the details and conclude the proof of Theorem 1. □

We now give the proof of Lemma 3.

Proof of Lemma 3. We prove **Part (a)** by induction. Note that $\boldsymbol{y}^*(2) = (\mathbf{I}-\mathbf{D})^{-1}(\boldsymbol{\mu}+\alpha_d\mathbf{D}\boldsymbol{y}^*(1))$. Because $y_e^*(1)\ge 0$ for all $e\in E$, we have, by Equation (15),

$$\boldsymbol{y}^*(2) = (\mathbf{I}-\mathbf{D})^{-1}(\boldsymbol{\mu}+\alpha_d\mathbf{D}\boldsymbol{y}^*(1)) \ge (\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu} = \boldsymbol{y}^*(1).$$

Therefore, the base case holds.

Next, we show that if $y_e^*(s)\ge y_e^*(s-1)$ for all $e\in E$ and $2\le s\le t$, then $y_e^*(t+1)\ge y_e^*(t)$ for all $e\in E$. By Equation (16), we have

$$\begin{aligned}
\boldsymbol{y}^*(t+1)-\boldsymbol{y}^*(t) =& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu}+\sum_{1\le s\le t}\alpha_d^{t+1-s}\mathbf{D}\boldsymbol{y}^*(s)\right) - (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu}+\sum_{1\le s\le t-1}\alpha_d^{t-s}\mathbf{D}\boldsymbol{y}^*(s)\right)\\
=&(\mathbf{I}-\mathbf{D})^{-1}\left(\alpha_d^t\mathbf{D}\boldsymbol{y}^*(1)+\sum_{s=1}^{t-1}\alpha_d^s\mathbf{D}(\boldsymbol{y}^*(t+1-s)-\boldsymbol{y}^*(t-s))\right)\\
\ge&\mathbf{0}_{|E|},
\end{aligned}$$

where the inequality follows from the inductive hypothesis, $y_e^*(1)\ge 0$ for all $e\in E$, and $(\mathbf{I}-\mathbf{D})^{-1}\ge \mathbf{0}_{|E|\times|E|}$ (see Equation (15)). Therefore, **Part (a)** follows immediately from the standard induction argument.

We prove **Part (b)** by induction. Observe that, by Equations (14)-(15) and $0<\alpha_d<1$,

$$\boldsymbol{y}^*(1) = (\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu} = \left(\mathbf{I}+\sum_{k=1}^{+\infty}\boldsymbol{D}^k\right)\boldsymbol{\mu} \le \left(\mathbf{I}+\sum_{k=1}^{+\infty}\frac{1}{(1-\alpha_d)^k}\boldsymbol{D}^k\right)\boldsymbol{\mu} = \mathcal{BE}(\mathbf{D},\boldsymbol{\mu}).$$

The base case holds.

Next, we show that if $\boldsymbol{y}^*(s) \leq \mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$ for all $1 \leq s \leq t$, it holds that $\boldsymbol{y}^*(t+1) \leq \mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$. By Equations (15)-(16) and the inductive hypothesis, we have

$$\begin{aligned}
\boldsymbol{y}^*(t+1) =& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \sum_{1\leq s\leq t} \alpha_d^{t+1-s}\mathbf{D}\boldsymbol{y}^*(s)\right) \\
\leq& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \sum_{1\leq s\leq t} \alpha_d^{t+1-s}\mathbf{D}\cdot\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})\right) \\
=& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \frac{\alpha_d(1-\alpha_d^t)}{1-\alpha_d}\mathbf{D}\cdot\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})\right) \\
\leq& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \frac{\alpha_d}{1-\alpha_d}\mathbf{D}\cdot\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})\right),
\end{aligned} \tag{20}$$

where the second inequality follows from $0 < \alpha_d^t < 1$. Expanding $\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$ by Equation (14), we have

$$\begin{aligned}
(\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \frac{\alpha_d}{1-\alpha_d}\mathbf{D}\cdot\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})\right) &= (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \frac{\alpha_d}{1-\alpha_d}\mathbf{D}\cdot\left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{1}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu}\right) \\
&= (\mathbf{I}-\mathbf{D})^{-1}\left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{\alpha_d}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu}.
\end{aligned} \tag{21}$$

Furthermore, invoking Equation (14), we evaluate $(\mathbf{I}-\mathbf{D})\,\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$ as follows:

$$\begin{aligned}
(\mathbf{I}-\mathbf{D})\,\mathcal{BE}(\mathbf{D},\boldsymbol{\mu}) &= (\mathbf{I}-\mathbf{D})\left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{1}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu} \\
&= \left(\mathbf{I} + \sum_{k=1}^{+\infty}\left(\frac{1}{(1-\alpha_d)^k} - \frac{1}{(1-\alpha_d)^{k-1}}\right)\mathbf{D}^k\right)\boldsymbol{\mu} \\
&= \left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{\alpha_d}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu}.
\end{aligned} \tag{22}$$

Therefore,

$$\mathcal{BE}(\mathbf{D},\boldsymbol{\mu}) = (\mathbf{I}-\mathbf{D})^{-1}\left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{\alpha_d}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu}. \tag{23}$$

Combining Equations (20), (21), and (23) immediately yields our desired inequality that $\boldsymbol{y}^*(t+1) \leq \mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$. This proves the induction step. We conclude the proof of Lemma 3. □

Proof of Corollary 1. Since $(\alpha_d,\mathbf{D})$ satisfies Condition $\mathcal{C}$,

$$\lim_{k\uparrow+\infty}\tilde{\boldsymbol{y}}(k) = \lim_{k\uparrow+\infty}\widetilde{\mathcal{BE}}(\mathbf{D},\boldsymbol{\mu},k) = \lim_{k\uparrow+\infty}\mathbf{M}(k)\boldsymbol{\mu} = \mathcal{BE}(\mathbf{D},\boldsymbol{\mu}) = \boldsymbol{y}^*,$$

where the third equality follows from 2 and the fourth from Theorem 1. Hence, we have

$$\lim_{k\uparrow+\infty}\tilde{x}(k) = \lim_{k\uparrow+\infty}\boldsymbol{\eta}^T\tilde{\boldsymbol{y}}(k) = \boldsymbol{\eta}^T\boldsymbol{y}^* = x^*.$$

Since $\mathbf{D} \geq \mathbf{0}$, $\mathbf{M}(k)$ is componentwise increasing in k, so is $\tilde{\boldsymbol{y}}(k) = \mathbf{M}(k)\boldsymbol{\mu}$. Therefore, $\tilde{x}(k) = \boldsymbol{\eta}^T\tilde{\boldsymbol{y}}(k)$ increasing in k as well. □

Algorithm 1 Approximate Global Effect of Social Nudges

Down Sampling: Uniformly randomly sample a subset of nodes $\tilde{V} \subset V$. Find the set of edges that point to a node in $\tilde{V}$, $\tilde{E} := \{e \in E : e_d \in \tilde{V}\}$), and the set of edges that originate from a node in $\tilde{V}$, $\tilde{L} := \{\ell \in E : \ell_o \in \tilde{V}\}$).

Parameter Initialization: For each $e \in \tilde{E}$, estimate μ_e and p_e. For each $\ell \in \tilde{L}$, estimate p_ℓ. For each $e \in \tilde{E}, \ell \in \tilde{L}, e_d = \ell_o$, estimate $d_{e\ell}$. Estimate α_d and α_p.

Direct Effect of Social Nudges on Content Production: Estimate

$$\hat{w}_0 := \sum_{i \in \tilde{V}} \sum_{e \in \tilde{E}, e_d = i} \frac{\mu_e p_e}{1 - \alpha_p}.$$

Indirect Effect of Social Nudges on Content Production: Estimate:

$$\hat{w}_1 := \sum_{i \in \tilde{V}} \sum_{e \in \tilde{E}, \ell \in \tilde{L}, e_d = \ell_o = i} \frac{\mu_e d_{e\ell} p_\ell}{(1 - \alpha_p)(1 - \alpha_d)}$$

Total Production Boost on the Entire Population: Scaling the estimates back to V:

$$\hat{w} := \frac{|V|}{|\tilde{V}|}(\hat{w}_0 + \hat{w}_1)$$

D.3. Estimating the Global Effect of Social Nudges

In this section, we present an approximation algorithm to estimate the global effect of social nudges on production boost. Then, we show that this algorithm generates consistent estimate for $\tilde{x}(1)$. Based on the two approximations (confining the diffusion radius to 1 and downsampling the user nodes to $\tilde{V}$) introduced in Section 6.2, Algorithm 1 provides a detailed global effect estimation procedure. We are now ready to prove that Algorithm 1 produces an unbiased estimate for $\tilde{x}(1)$.

Proposition 1 *Algorithm 1 yields an unbiased estimate for $\tilde{x}(1)$.*

Proof of Proposition 1. As shown in Section 6.2, the total production boost can be theoretically approximated as $\tilde{x}(1)$ by including all providers on the entire social network (the node set V), and is practically approximated as $\hat{w}$ according to our Algorithm 1 by sampling a subset of providers from V (i.e., $\tilde{V}$). Now we prove that $\hat{w}$ is an unbiased estimate for $\tilde{x}(1)$.

First, reorganizing the sum by nodes rather than edges, we have

$$\begin{aligned}\tilde{x}(1) &= \boldsymbol{\eta}^\top \left(\mathbf{I} + \frac{1}{(1-\alpha_d)}\mathbf{D}\right)\boldsymbol{\mu} = \sum_{e \in E}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell \in E:\ \ell_d = e_o} \frac{\mu_e d_{\ell e} p_e}{(1-\alpha_p)(1-\alpha_d)}\right) \\ &= \sum_{e \in E}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell \in E:\ e_d = \ell_o} \frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right) = \sum_{i \in V}\left(\sum_{e \in E:\ e_d = i}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell \in E:\ e_d = \ell_o} \frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right).\end{aligned}$$

Note that $\hat{w}$ is defined as

$$\hat{w} = \frac{|V|}{|\tilde{V}|}\sum_{i \in \tilde{V}}\left(\sum_{e \in \tilde{E}:\ e_d = i}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell \in \tilde{L}:\ e_d = \ell_o} \frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right).$$

where $\tilde{E} := \{e \in E : e_d \in \tilde{V}\}$ and $\tilde{L} := \{\ell \in E : \ell_o \in \tilde{V}\}$. Clearly, we can replace $\tilde{E}$ and $\tilde{L}$ with E respectively in the definition of w. Moreover, for each node i in V, we use a binary random variable $s_i \in \{0,1\}$ to denote whether node i is selected in the sample $\tilde{V}$. Then, we can write w as

$$\hat{w} = \frac{|V|}{|\tilde{V}|}\sum_{i\in\tilde{V}}\left(\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right)$$
$$= \frac{|V|}{|\tilde{V}|}\sum_{i\in V}\left(s_i\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right).$$

Since we uniformly and randomly sample the node set $\tilde{V}$, we have, for each $i \in V$, $\mathbb{E}[s_i] = \frac{|\tilde{V}|}{|V|}$. Hence,

$$\mathbb{E}[\hat{w}] = \frac{|V|}{|\tilde{V}|}\sum_{i\in V}\left(\mathbb{E}[s_i]\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right)$$
$$= \frac{|V|}{|\tilde{V}|}\sum_{i\in V}\left(\frac{|\tilde{V}|}{|V|}\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right)$$
$$= \sum_{i\in V}\left(\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right) = \tilde{x}(1).$$

This concludes the proof. □

D.4. Production Boost from Nudges Sent by New Users

In this section, we quantity the total production boost attributed to the organic nudges sent by new users. See Online Appendix F for notations and modeling details.

Proposition 2 *The additional production boost per period from the organic nudges sent by new users in N is given by:*

$$\Delta x^* := \bar{x}^* - x^* = \bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_{E'}) = \sum_{e\in E'} \nu_e, \tag{24}$$

provided that $(\alpha_d, \bar{\mathbf{D}})$ satisfies Condition $\mathcal{C}$. Here $\bar{x}^*$ and x^* are the vectors of production boost in the long-run steady state for the network with and without the set of new users, respectively.

Proof of Proposition 2. Define $\bar{\boldsymbol{\mu}}_E \in \mathbb{R}^{|\bar{E}|}$ with $\bar{\mu}_e = \mu_e$ for $e \in E$ and $\bar{\mu}_e = 0$ for $e \in E'$ (recall that $\bar{E} = E \cup E'$ and $E \cap E' = \emptyset$). Hence, $\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}}_E + \bar{\boldsymbol{\mu}}_{E'}$. By Equation (5), $\mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{v}})$ is linear in vector $\bar{\boldsymbol{v}} \in \mathbb{R}^{|\bar{E}|}$. Therefore,

$$\bar{x}^* = \bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}) = \bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_E + \bar{\boldsymbol{\mu}}_{E'}) = \bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_E) + \bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_{E'}).$$

Hence, to prove Equation (24), it suffices to show that

$$x^* = \bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_E),$$

or, equivalently by Theorem 1,

$$\bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_E) = \boldsymbol{\eta}^T \mathcal{BE}(\mathbf{D}, \boldsymbol{\mu}). \tag{25}$$

Since N is the set of new users, no one has followed them yet. Therefore, there is no edge whose destination is the origin of any edge in E', i.e., for any $e \in E'$ and any $\ell \in \bar{E}$, we have $\ell_d \neq e_o$. Therefore, by definition, $d_{\ell e} = 0$ for any $e \in E'$ and any $\ell \in \bar{E}$. This implies that $(\bar{\mathbf{D}}\bar{\boldsymbol{\mu}}_E)_e = (\mathbf{D}\boldsymbol{\mu})_e$ for all $e \in E$ and $(\bar{\mathbf{D}}\bar{\boldsymbol{\mu}}_E)_e = 0$ for $e \in E'$. Therefore, by a standard induction argument, we have, for any $k \geq 1$,

$$\left(\frac{1}{(1-\alpha_d)^k}\bar{\mathbf{D}}^k \bar{\boldsymbol{\mu}}_E\right)_e = \left(\frac{1}{(1-\alpha_d)^k}\mathbf{D}^k \boldsymbol{\mu}\right)_e \text{ if } e \in E; \ \left(\frac{1}{(1-\alpha_d)^k}\bar{\mathbf{D}}^k \bar{\boldsymbol{\mu}}_E\right)_e = 0 \text{ if } e \in E'. \tag{26}$$

Because $\bar{\boldsymbol{\eta}}_e = \boldsymbol{\eta}_e$ for all $e \in E$, Equation (26) further implies that, for all $k \geq 1$,

$$\frac{1}{(1-\alpha_d)^k}\bar{\boldsymbol{\eta}}^T \bar{\mathbf{D}}^k \bar{\boldsymbol{\mu}}_E = \frac{1}{(1-\alpha_d)^k}\boldsymbol{\eta}^T \mathbf{D}^k \boldsymbol{\mu}. \tag{27}$$

Furthermore, because $\bar{\mu}_e = 0$ for $e \in E'$, we have

$$\bar{\boldsymbol{\eta}}^T \cdot \mathbf{I} \cdot \bar{\boldsymbol{\mu}}_E = \bar{\boldsymbol{\eta}}^T \bar{\boldsymbol{\mu}}_E = \boldsymbol{\eta}^T \boldsymbol{\mu} = \boldsymbol{\eta}^T \cdot \mathbf{I} \cdot \boldsymbol{\mu}. \tag{28}$$

Therefore, Definition 1 implies that:

$$\begin{aligned}\bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_E) =& \bar{\boldsymbol{\eta}}^T \left(\mathbf{I} + \sum_{k=1}^{+\infty} \frac{1}{(1-\alpha_d)^k}\bar{\mathbf{D}}^k\right)\bar{\boldsymbol{\mu}}_E = \bar{\boldsymbol{\eta}}^T \cdot \mathbf{I} \cdot \bar{\boldsymbol{\mu}}_E + \sum_{k=1}^{+\infty}\frac{1}{(1-\alpha_d)^k}\bar{\boldsymbol{\eta}}^T \bar{\mathbf{D}}^k \bar{\boldsymbol{\mu}}_E \\ =& \boldsymbol{\eta}^T \cdot \mathbf{I} \cdot \boldsymbol{\mu} + \sum_{k=1}^{+\infty}\frac{1}{(1-\alpha_d)^k}\boldsymbol{\eta}^T \bar{\mathbf{D}}^k \boldsymbol{\mu} = \boldsymbol{\eta}^T\left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{1}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu} = \boldsymbol{\eta}^T \mathcal{BE}(\mathbf{D}, \boldsymbol{\mu}),\end{aligned} \tag{29}$$

where the third equality follows from Equations (27) to (28). Therefore, equality (25) holds. Finally, the last equality of (24) follows immediately from the linearity of the BCE measure $\mathcal{BE}(\bar{\mathbf{D}}, \cdot)$ in $\boldsymbol{v}$ and the definition of ν_e (Equation (9)). We have concluded the proof of Proposition 2. □

E. Social Network Model Estimation Details

In this section, we present the estimation details for the global effect of social nudges, including the parameters needed in Algorithm 1: μ_e ($e \in E$), p_e ($e \in E$), $d_{e\ell}$ ($e, \ell \in E$ and $e_d = \ell_o$), α_p, and α_d. We also present two robustness checks.

E.1. Estimation of μ_e

Due to Platform O's rule that a user can send no more than one social nudge to another user each day, estimated μ_e (and $d_{e\ell}$) falls between 0 and 1. In this case, the parameter μ_e measures the expected probability of e_o sending a social nudge to e_d per day when e_o has not received nudges from her followers recently. We estimate μ_e by taking advantage of the fact that providers in the control group of our social nudge experiment cannot receive nudges (and thus cannot be motivated to send more nudges out because of receiving nudges themselves) during the experiment. We sampled 5 million edges uniformly at random from all edges whose origin was in the control condition of our social-nudge experiment. Here, we do not require the origins of these edges to satisfy the selection criteria of our analysis sample mentioned in Section 3 since we use this random edge sample to represent the overall edges on Platform O. Our goal is to train a prediction model to estimate μ_e for each $e \in E$.

We fit the logistic regression model (30) to predict μ_e, i.e., the probability that $\textit{Social-Nudge Incidence}_e = 1$. We select features based on the commonly recognized characteristics in the network economics literature (see, e.g., Jackson 2010) such as the degrees of a node in V (measured by the number of followers and the number of followings the node has) and the strength of an edge in E (measured by whether e_o and e_d has a bi-directional relationship, i.e., whether there exists $e' \in E$ such that $e'_o = e_d$ and $e'_d = e_o$). Among a large set of network-based features that we explore, our final logistic regression model includes features that satisfy two criteria: (1) the coefficient on the feature is statistically significant, and (2) the combination of selected features maximizes the performance of the logistic regression model. Specifically, the final retained features include (1) whether e_o's number of followers was greater than the median value across all origin nodes in the sample (*Large Number of Followers for* e_o), (2) whether e_o's number of followings was greater than the median value across all origin nodes in the sample (*Large Number of Following for* e_o), (3) whether e_d was also following e_o ($\textit{Two-Way Tie}_e$), and (4) the baseline productivity (which equals the average number of videos uploaded per day across the 30 days before the experiment) of e_d (*Baseline Productivity of* e_d)[11].

$$\begin{aligned}
&\log\left(\frac{\mathbb{P}(\textit{Social-Nudge Incidence}_e = 1)}{1-\mathbb{P}(\textit{Social-Nudge Incidence}_e = 1)}\right)\\
=&\beta_0 + \beta_1 \textit{Large Number of Followers for } e_o + \beta_2 \textit{Large Number of Following for } e_o \\
&+ \beta_3 \textit{Two-Way Tie}_e + \beta_4 \textit{Baseline Productivity of } e_d + \epsilon_e
\end{aligned} \tag{30}$$

Table 19 reports the estimated coefficients (β_i) and the standard errors of the estimates. We implement a five-fold cross validation to evaluate the performance of this logistic regression model, which has a 99.99% average accuracy and a 0.78 Area Under Curve (AUC), suggesting qualified prediction performance. For all the edges in $\tilde{E}$, we can estimate the probability that e_o will nudge e_d in a given period by Equation (31).

$$\begin{aligned}
\frac{1}{\mu_e} =& 1 + \exp(-(\beta_0 + \beta_1 \textit{Large Number of Followers for } e_o + \beta_2 \textit{Large Number of Following for } e_o \\
&+ \beta_3 \textit{Two-Way Tie}_e + \beta_4 \textit{Baseline Productivity of } e_d))
\end{aligned} \tag{31}$$

[11] The correlation between the baseline productivity and social-nudge incidence is -0.0021.

Table 19 The Results of a Logistic Regression Model Predicting Social-Nudge Incidence

	Coefficient (1)	Standard Error (2)
Intercept	-9.9943****	0.1204
Large Number of Followers for e_o	1.4398****	0.1309
Large Number of Following for e_o	-0.8518****	0.1013
Two-Way Tie$_e$	1.0309****	0.1048
Baseline Productivity of e_d	-0.3977****	0.0951

Note: *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

E.2. Estimation of p_e and α_p

Recall that the parameter p_e ($e \in E$) measures the immediate positive effect of receiving *one* social nudge from e_o on provider e_d's production, i.e., the production boosting effect in the same time period when the nudge is sent. The time discounting factor α_p indicates that receiving one social nudge from e_o boosts provider e_d's production by $p_e\alpha_p^t$ in the t^{th} period after e_d receives the nudge. To cleanly estimate p_e and α_p, from the analysis sample of our social-nudge experiment (as defined in Section 3), we identify $962,120$ providers who were sent only *one* social nudge on their first reception day (accounting for 97% of the analysis sample). Those providers were not sent any nudges prior to the experiment (per the selection criteria of our analysis sample), so they were sent one social nudge for the first time on their first reception day.

Since we jointly estimate parameters p_e and α_p, we focus on estimating p_e as the average treatment effect. That is, p_e is independent of the edge $e \in E$. Specifically, we first estimate the coefficient on treatment (i.e, β_1) in regression specification (1) for each day t since the first reception day (where $t = 1$ refers to the first reception day itself) until β_1 becomes statistically insignificant on a given day t. The dependent variable examined here is *Number of Videos Uploaded*$_{it}$. On Day 4 (i.e., three days after the first reception day), β_1 is no longer statistically significant, so we use the estimates of β_1 from Day 1 to Day 3. The regression results are reported in Table 20.

We denote $p(t)$ as the regression coefficient on treatment estimated using raw data without standardization for Day t ($t = 1, 2, 3$). In Table 20, we report the corresponding regression coefficient on treatment using standardized data to protect Platform O's sensitive information. We jointly estimate (p_e, α_p) by minimizing the sum of squared errors in the following nonconvex program:

$$\min_{(p_e,\alpha_p)} \left\{ \sum_{t=1}^{3} \epsilon_t^2 \;\middle|\; p(t) = p_e \alpha_p^{t-1} + \epsilon_t, t = 1, 2, 3 \right\}. \tag{32}$$

Solving (32) yields p_e and α_p, which we report in Table 8 column (1).

Table 20 Over-Time Direct Effects of Receiving One Social Nudge on Content Production

Panel A: Main Experiment				
Outcome Variable	Number of Videos Uploaded			
	on Day 1 (First Reception Day)	on Day 2	on Day 3	on Day 4
	(1)	(2)	(3)	(4)
Treatment	0.0263^{****}	0.0125^{****}	0.0077^{****}	0.0003
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Relative Effect Size	13.71%	5.28%	3.12%	
Observations	962,120	962,120	962,120	962,120

Panel B: Replication Experiment					
Outcome Variable	Number of Videos Uploaded				
	on Day 1 (First Reception Day)	on Day 2	on Day 3	on Day 4	on Day 5
	(1)	(2)	(3)	(4)	(5)
Treatment	0.0237^{****}	0.0183^{****}	0.0102^{****}	0.0060^{*}	0.0028
	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)
Relative Effect Size	12.68%	8.56%	5.01%	2.98%	
Observations	655,001	655,001	655,001	655,001	655,001

Note: Number of Videos Uploaded was standardized to have a unit deviation before entering the regressions. Panel A includes providers who were sent only one social nudge on their first reception day in the main experiment. Panel B includes providers who were sent only one social nudge on their first reception day in the replication experiment. The unit of analysis for all columns was a provider on Day t, where $t = 1$ refers to the first reception day. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

E.3. Estimation of $d_{e\ell}$ and α_d

The estimation of $d_{e\ell}$ and α_d follows a similar approach to that of (p_e, α_p). The parameter $d_{e\ell}$ measures the increase in e_d's probability of sending a social nudge to ℓ_d on the day of receiving *one* social nudge from e_o ($e_d = \ell_o$). By definition, $d_{e\ell} = 0$ if $e_d \neq \ell_o$. The parameter α_d quantifies the time-discounting factor of such effect, such that receiving one social nudge from e_o boosts the number of nudges provider e_d would send to ℓ_d by $d_{e\ell}\alpha_d^t$ in the t^{th} period after e_d receives the nudge. We focus on the subset of providers from the analysis sample of our social-nudge experiment who (1) were sent only *one* social nudge on their first reception day and (2) were following at least one user the day before the main experiment. We estimate the diffusion effect of a social nudge by comparing the number of nudges providers sent per following relationship between the treatment and control conditions on and after their first reception day, adopting regression specification (1) for each day since the first reception day.

Consistent with the estimation of p_e and α_p, we jointly estimate parameters $d_{e\ell}$ and α_d. Due to the joint estimation, we focus on estimating $d_{e\ell}$ as the average treatment effect. That is, $d_{e\ell}$ is independent of the edge $e, \ell \in E$. Specifically, we first estimate the coefficient on treatment (i.e., β_1) in regression specification (1) for each day t since the first reception day (where $t = 1$ refers to the first reception day) until β_1 becomes statistically insignificant on a given day t. The dependent variable examined here is *Number of Social Nudges Sent per Edge*$_{it}$. It equals the number of social nudges sent by provider $i \in V$ on day t since the first reception day divided by her number of following. Starting from Day 3, β_1 is no longer statistically significant, so we use the estimates of β_1 from Day 1 to Day 2. The regression results are reported in Table 21.

We denote $d(t)$ as the regression coefficient estimated using raw data without standardization for Day t $(t=1,2)$. In Table 21, we report the corresponding regression coefficient on treatment using standardized data to protect Platform O's sensitive information. We jointly estimate $(d_{e\ell}, \alpha_d)$ by minimizing the sum of squared errors in the following nonconvex program:

$$\min_{(d_e,\alpha_d)} \left\{ \sum_{t=1}^{2} \epsilon_t^2 \;\middle|\; d(t) = d_{e\ell}\alpha_d^{t-1} + \epsilon_t, t = 1,2 \right\}. \tag{33}$$

Solving (33) yields $d_{e\ell}$ and α_d, which we report in Table 8 column (1).

Table 21 Over-Time Diffusion Effects of Receiving One Social Nudge

Panel A: Main Experiment			
Outcome Variable	Number of Social Nudges Sent per Edge		
	on Day 1 (First Reception Day)	on Day 2	on Day 3
	(1)	(2)	(3)
Treatment	0.0080***	0.0049*	0.0025
	(0.0021)	(0.0021)	(0.0021)
Relative Effect Size	10.55%	8.06%	
Observations	947,730	947,730	947,730
Panel B: Replication Experiment			
Outcome Variable	Number of Social Nudges Sent per Edge		
	on Day 1 (First Reception Day)	on Day 2	on Day 3
	(1)	(2)	(3)
Treatment	0.0082***	0.0050*	0.0010
	(0.0025)	(0.0025)	(0.0025)
Relative Effect Size	10.89%	9.06%	
Observations	640,920	640,920	640,920

Notes: Number of Social Nudges Sent per Edge was standardized to have a unit standard deviation before entering the regressions. Panel A includes providers who were sent only one social nudge on their first reception day in the main experiment and were following at least one user the day before the main experiment. Panel B includes providers who were sent only one social nudge on their first reception day in the replication experiment and were following at least one user the day before the replication experiment. The unit of analysis for all columns was a provider on Day t, where $t=1$ refers to the first reception day. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

E.4. Robustness Checks

As the first robustness check, we re-sample $\tilde{V}$ and re-estimate the global effect of social nudges using parameters estimated from our main social-nudge experiment. As shown in Table 22, the estimation results based on the new sample of $\tilde{V}$ are very similar to the results based on the sample reported in Section 6.2, confirming that our estimates reported in Table 9 are robust.

As the second robustness check, we re-estimate (p_e, α_p) and $(d_{e\ell}, \alpha_d)$ based on the same method described above but use data from the second social-nudge replication experiment. Regarding the

Table 22 Estimation of the Global Effect of Social Nudges

	The Estimation Result			
	Reported in Section 6.2 (1)		Based on Another Sample of $\tilde{V}$ (2)	
Direct Effect	130.08	*One Day:* 47.55 *Beyond One Day:* 82.53	132.85	*One Day:* 48.56 *Beyond One Day:* 84.29
Indirect Effect	10.59		10.87	
Global Effect	140.67		143.72	
The Ratio of Indirect Effect to Direct Effect	8.14%		8.19%	

parameters p_e and α_p, we estimate the coefficient on treatment (i.e., β_1) in regression specification (1) for each day t where outcome variable is *Number of Videos Uploaded*$_{it}$ since the first reception day (where $t = 1$ refers to the first reception day itself) until β_1 becomes statistically insignificant on a given day t. For the replication experiment, on day 5, β_1 is no longer statistically significant. So we use the estimates of β_1 from Day 1 to Day 4 to jointly estimate p_e and α_p. The regression results using standardized data are presented in Panel B of Table 20. The corresponding solution to the nonconvex program (32) yields p_e and α_p, which we report in Table 8 column (2) and are consistent with the estimates derived from the first experiment.

Regarding the parameters $d_{e\ell}$ and α_d, we estimate the coefficient on treatment (i.e, β_1) in regression specification (1) for each day t where outcome variable is *Number of Social Nudges Sent per Edge*$_{it}$ since the first reception day (where $t = 1$ refers to the first reception day itself) until β_1 becomes statistically insignificant on a given day t. Starting from Day 3, β_1 is no longer statistically significant, so we use the estimates of β_1 from Day 1 to Day 2 to jointly estimate $d_{e\ell}$ and α_d. The regression results are presented in Panel B of Table 21. The corresponding solution to the nonconvex program (33) yields $d_{e\ell}$ and α_d, which we report in Table 8 column (2) and are consistent with the estimates derived from the first experiment.

In addition, we apply Algorithm 1 and the data from the second social-nudge experiment to estimate the global effect of social nudges. We report the estimation results in Table 9 column (2). Compared to the naïve estimation, including the over-time accumulation of the direct boosting effect of social nudges on recipients' production leads to a 200% (i.e., $(146.06 - 48.65)/48.65$) increase, and considering nudge diffusion leads to an additional 25% (i.e., 12.24/48.65) increase. The indirect production boost from social-nudge diffusion accounts for at least 8.38% (i.e., 12.24/146.06) of the direct production effect. All of these results are fairly consistent with our estimation results based on data from the first social-nudge experiment (see Table 9).

F. Operational Problems About Social Network Model

In this section, we study two operational problems with our social network model: (1) optimal seeding, and (2) provider recommendation for new users. We solve the problems leveraging the SNI developed in Section 6.3.

F.1. Optimal Seeding

Here we provide details about how to solve the optimal seeding application as presented in Section 6.3. Assuming that user e_o will, on average, send more nudges to e_d if the platform encourages her to do so, we denote that for each $e \in K$, the average number of social nudges sent per day will increase by a relative effect of δ_μ after e_o receives the motivation from the platform (i.e., from μ_e to $\mu_e(1+\delta_\mu)$).[12]

It is straightforward to derive that the global effect increment of social nudges with respect to the selected edges, K, is $\boldsymbol{\eta}^T \mathcal{BE}(\mathbf{D}, \boldsymbol{\mu}_K)\delta_\mu$, where $\boldsymbol{\mu}_K \in \mathbb{R}^{|E|}$ represents a vector with an entry of edge $e \in K$ (resp. $e \notin K$) equal to μ_e (resp. 0). Such production boost can be reasonably approximated by

$$\Delta(K, 1) := \boldsymbol{\eta}^T \cdot \widetilde{\mathcal{BE}}(\mathbf{D}, \boldsymbol{\mu}_K, 1)\delta_\mu = \delta_\mu \cdot \sum_{e \in K} \tilde{\nu}_e(1). \tag{34}$$

Therefore, it is (approximately) "optimal" to select n edges in E with the highest (approximate) SNIs, i.e., the n edges with the largest $\tilde{\nu}_e(1)$'s.

Table 23 Global Effect of Social Nudges for Different Strategies

	Setting	Δ_o (1)	Δ_r (2)	Ξ (3)
(1)	$\lvert K\rvert = 0.1\lvert E\rvert$, $\delta_\mu = 100\%$	54.77	13.78	297%
(2)	$\lvert K\rvert = 0.1\lvert E\rvert$, $\delta_\mu = 10\%$	5.48	1.38	297%

The platform may adopt the random strategy, which randomly targets a subset of edges $K \subset E$ and encourages the users to nudge more on these edges. This random strategy is the most straightforward and simplest way to stimulate social nudges sent on a platform. We benchmark the approximately "optimal" strategy (i.e., to target the edges with the highest $\tilde{\nu}_e(1)$'s) against the random strategy and compare their performances in production boost. We define Δ_r (resp. Δ_o) as the additional production boost under the random (resp. "optimal") strategy, and $\Xi := (\Delta_o - \Delta_r)/\Delta_r \times 100\%$ as the relative improvement of the "optimal" strategy over the random strategy. We evaluate these two strategies based on the same sample of $\tilde{V}$ as the one used to generate the

[12] Our method of deriving the optimization strategy can be easily carried over to a setting where the average number of social nudges sent per day will increase by an absolute effect of δ_μ after e_o receives one push from the platform (i.e., from μ_e to $\mu_e + \delta_\mu$) for each $e \in K$.

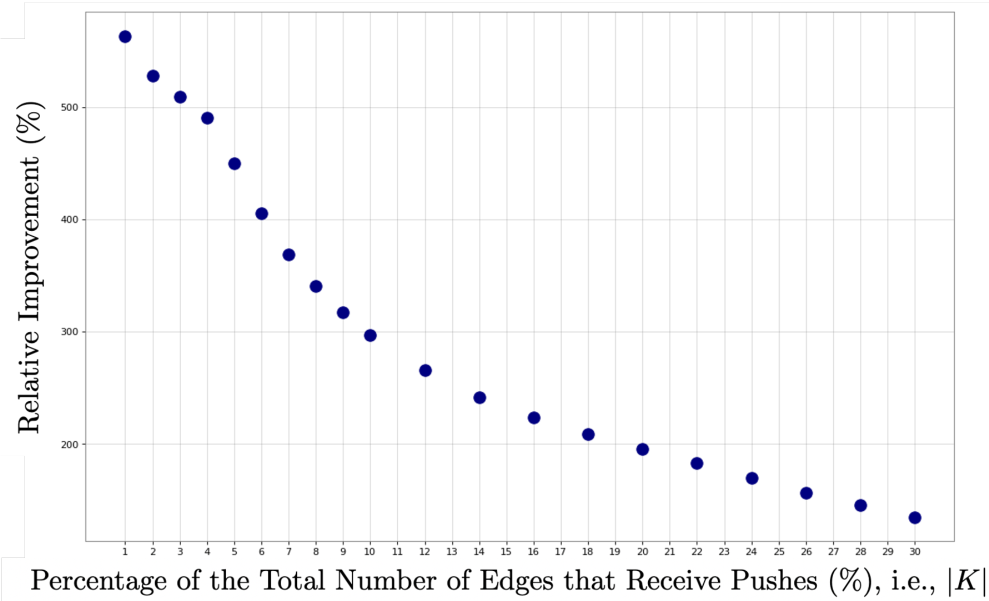

Figure 5 **Relative Improvements of the Optimal Strategy Over the Random Strategy as $|K|$ Changes**

global effect estimates in Table 9 column (1). We also examine how Ξ changes according to δ_μ and the size of the targeted edges $n = |K|$. The simulation results are reported in Table 23 and Figure 5. The primary observation is that the "optimal" strategy based on (approximate) SNIs substantially outperforms the random strategy, regardless of the effectiveness of the platform's encouragement for users to send additional nudges (i.e., the magnitude of δ_μ). In particular, this relative edge is most prominent if the constraint on the number of targeted edges is tighter (i.e., n is smaller).

F.2. Content Provider Recommendation for New Users

Below, we provide details about the provider recommendation problem for newly registered platform users. The platform will first construct a provider list based on the basic demographic features (e.g., age, gender and location) of each new user, together with her interests in specific content categories reported upon registration. Next, the platform decides the ranking of the provider list, following which it sequentially recommends the content providers to the new user.

Recall that the existing social network of Platform O is denoted by $G = (V, E)$, where V is the set of existing nodes (users), and E is the set of existing edges (following relationships). We denote the set of newly registered users as N. For each new user $i \in N$, denote the set of existing providers this user chooses to follow as U_i and the associated set of new following relationships as $E_i := \{(i,u) : u \in U_i\}$. Therefore, the new social network with those new users can be written as $\bar{G} := \left(\bar{V}, \bar{E}\right)$, where $\bar{V} := V \cup N$ is the set of all users and $\bar{E} := E \cup \left(\bigcup_{i\in N} E_i\right)$ is the set of all edges. The production boost vector $\bar{\boldsymbol{\eta}} := (p_e/(1-\alpha_p) : e \in \bar{E})$, the organic nudge vector $\bar{\boldsymbol{\mu}} := (\mu_e : e \in \bar{E})$, and the diffusion matrix $\bar{\mathbf{D}} := (d_{\ell e} : (\ell, e) \in \bar{E}^2)$ with respect to the new social network can be defined accordingly. By Theorem 1, as long as $\bar{\mathbf{D}}$ satisfies Condition $\mathcal{C}$, the total number of nudges and the total production boost per period are given by $\bar{\boldsymbol{y}}^* = \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}})$ and $\bar{x}^* := \bar{\boldsymbol{\eta}}^T \bar{\boldsymbol{y}}^*$, respectively. Define $E' := \bigcup_{i \in N} E_i$ as the set of new edges and $\bar{\boldsymbol{\mu}}_{E'} \in \mathbb{R}^{|\bar{E}|}$ as a vector with an entry of edge $e \in E'$ (resp. $e \notin E'$) equal to μ_e (resp. 0). The additional production boost attributed to the social nudges sent by the new users is given by $\Delta x^* = \bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_{E'}) = \sum_{e \in E'} \nu_e$.[13] Such additional production boost can be reasonably approximated by

$$\Delta \tilde{x}^*(1) := \bar{\boldsymbol{\eta}}^T \cdot \widetilde{\mathcal{BE}}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_{E'}, 1) = \sum_{e \in E'} \tilde{\nu}_e(1) = \sum_{i \in N} \left(\sum_{e \in E_i} \tilde{\nu}_e(1) \right), \tag{35}$$

provided that $\tilde{\nu}_e(1)$ is a reasonable approximation of ν_e for each $e \in E'$. Note that Equation (35) also implies the content provider recommendation of each new user can be optimized separately.

Let us now consider the content provider recommendation problem for the new user $i \in N$. To begin with, Platform O identifies a content provider list for the new user based on her features, which we denote as $M_i \subset V$. Then, the platform selects $V_i \subset M_i$ with $|V_i| = m$ and recommends the providers in V_i to the user in a sequential manner. Similar to the optimal seeding problem, the platform avoids overly interfering its users, so the total number of recommended providers to each new user, m, is generally not too large but at the magnitudes of a few dozens. Denote the probability that a new user will follow the j-th provider recommended to her as c_j. For simplicity, we assume that c_j is only dependent on the ranking of the provider in the list, i.e. j, but independent of the identity of him. Because a new user will have a higher chance to follow the provider recommended to her earlier, it holds that $c_1 \geq c_2 \geq \cdots \geq c_m$. The platform's recommendation strategy for user i can be summarized as the provider list V_i together with a bijection $\pi : \{1, 2..., m\} \to V_i$, where $\pi(j)$ refers to the provider ranked in the j-th position. By Equation (35), under the recommendation strategy (V_i, π), the (approximate) additional production boost from the social nudges sent by new user i is given by

$$\sum_{j=1}^{m} c_j \tilde{\nu}_{(i, \pi(j))}(1), \tag{36}$$

where $(i, \pi(j))$ is the edge in E_i, representing that new user i follows the j-th recommended provider and, thus, $c_j \tilde{\nu}_{(i,\pi(j))}(1)$ is the expected additional production boost by recommending the j-th provider. For any $i' \in M_i$, we call $\tilde{\nu}_{(i,i')}(1)$ the induced (approximate) SNI of provider i'. It is clear from Equation (36) that the (approximate) "optimal" strategy is to select m providers in M_i with the highest induced (approximate) SNIs and the rank them in the descending order of the induced (approximate) SNI.

Similar to optimal seeding, we would compare the SNI-based provider recommendation with the benchmark random recommendation, which recommends the content providers based on a random

[13] See Proposition 2 in Online Appendix D.4 for the formal result.

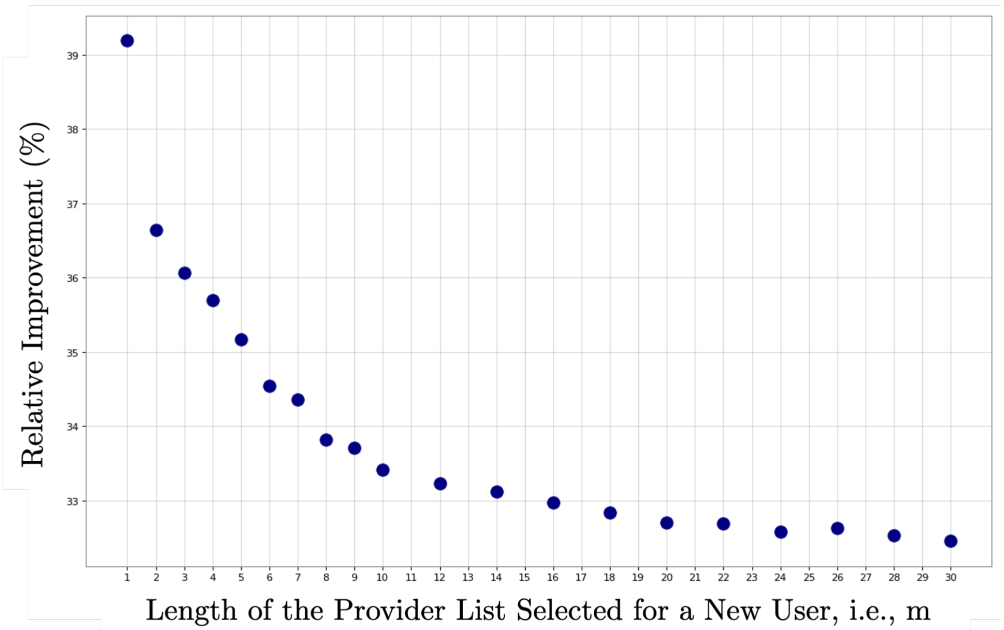

Figure 6 **Relative Improvements of the Optimal Strategy Over the Random Strategy as m Changes**

permutation of M_i. To evaluate the edge of our "optimal" strategy over the random strategy, we randomly sample $1,000$ new users on Platform O and examine the provider list of varying lengths m. We quantify the performance metrics Δ_o (production boost of the "optimal" strategy), Δ_r (production boost of the random strategy), and $\Xi = (\Delta_o - \Delta_r)/\Delta_r \times 100\%$ (the relative improvement of the "optimal" strategy over the random strategy) for recommending providers to new users and report the results in Figure 6. Consistent with the optimal seeding problem, the "optimal" provider recommendation strategy based on SNIs outperforms the random strategy, especially when the provider list length m is small.

In sum, we show that our social network model could help the platform optimize, among others, its seeding and provider recommendation strategies. The platform's (approximately) optimal strategy prescribed by our social-nudge indices is much more effective in boosting total production than the simple random strategy.

G. Data

In this section, we report in several tables the distributional information of the features and outcome variables studied in this paper, correlations between variables, and the distributional information of the degrees of the sample network used in Section 6. We provide a guideline of the organization of these tables, which is intended to help readers locate the information of interest. Specifically, in Tables 25–36, for each variable mentioned in the paper (including the main body and appendices), we first standardize it to have a unit deviation if it is a continuous variable (as explained in Section 3), and then report the quantiles (i.e., 1%, 25%, 50%, 75%, and 99%) in the full sample, as well as mean and standard deviation in each condition (i.e., treatment or control). In Table 37, for variables mentioned in the main body of the paper, we report moderate to large correlations between variables (i,e, cases where the absolute value of the correlation coefficient is no less than 0.30). In Table 38, we report the quantiles of the in-degrees and out-degrees of the sample of nodes used to calculate the global effect of social nudges in Section 6 (i.e., $\tilde{V}$). Like other variables, we are not allowed to reveal the distributional information of the raw data on the in-degrees or out-degrees. In order to maintain the relative magnitude between the in-degrees and out-degrees of these nodes, we first scale the in-degrees and out-degrees by a fixed constant (rather than their respective standard deviation, which differs between in-degrees and out-degrees), and then calculate the quantiles.

Table Index	Focus	Statistics
Table 25, Table 26	The first social-nudge experiment	Quantiles
Table 27	The second social-nudge experiment	Quantiles
Table 28	The platform-initiated nudge experiment	Quantiles
Table 29	A matched viewer sample for examining cannibalization	Quantiles
Table 30	Providers who were sent only one social nudge	Quantiles
Table 25, Table 31	The first social-nudge experiment	Mean and standard deviation in each condition
Table 33	The second social-nudge experiment	Mean and standard deviation in each condition
Table 34	The platform-initiated nudge experiment	Mean and standard deviation in each condition
Table 35	A matched viewer sample for examining cannibalization	Mean and standard deviation in each condition
Table 36	Providers who were sent only one social nudge	Mean and standard deviation in each condition
Table 37	All variables studied in the main body	Correlation
Table 38	Sample network	Quantiles

Table 24 Summary of Tables Related to Data Disclosure

Statistics Prior to the Experiment						
Variable	Location	1%	25%	50%	75%	99%
Female (Binary)[1]	Table 1	0.0000	0.0000	1.0000	1.0000	1.0000
Number of Followers	Table 1	0.0000	0.0026	0.0090	0.0289	0.6083
Number of Following	Table 1	0.0000	0.1749	0.4680	1.1108	4.3647
Number of Uploaded Videos	Table 1	0.0000	0.0000	0.0000	0.3528	4.5866
Number of Days with Videos Uploaded	Table 1	0.0000	0.0000	0.0000	0.8214	4.1070
Historical Like Rate	Table 3	0.0000	0.3182	0.6869	1.3090	4.3511
Statistics Related to the First Reception Day						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded	Table 2	0.0000	0.0000	0.0000	0.0000	3.6644
Upload Incidence (Binary)	Table 2	0.0000	0.0000	0.0000	0.0000	1.0000
Number of Videos Uploaded Conditional on Uploading Anything[2]	Table 2	1.8322	1.8322	1.8322	3.6644	12.8253
Two-Way Tie (Binary)	Table 2	0.0000	0.0000	0.0000	1.0000	1.0000
Total Views	Table 3	0.0000	0.0000	0.0000	0.0000	5.2715
Complete View Rate	Table 3	0.0000	0.1274	0.6607	1.4905	4.2946
Like Rate	Table 3	0.0000	0.4039	0.9596	1.6964	4.2857
Comment Rate	Table 3	0.0000	0.0000	0.4852	1.1716	4.3539
Following Rate	Table 3	0.0000	0.0000	0.0000	0.0000	4.0533
Number of Social Nudges Sent	Table 5	0.0000	0.0000	0.0000	0.0000	4.9896
Number of Likes on the First Reception Day	Table 15	0.0000	0.0000	0.0000	0.0000	7.4641
Number of Comments on the First Reception Day	Table 15	0.0000	0.0000	0.0000	0.0000	6.2778
Number of Videos Uploaded Among Low-Productivity Providers[3]	Table 17	0.0000	0.0000	0.0000	0.0000	3.6644
Number of Videos Uploaded Among Medium-Productivity Providers[4]	Table 17	0.0000	0.0000	0.0000	1.8322	7.3287
Number of Videos Uploaded Among High-Productivity Providers[5]	Table 17	0.0000	0.0000	1.8322	3.6644	18.3218

Table 25 Summary Statistics About the First (Main) Social-Nudge Experiment (I)

[1]Note: Hereafter, we label all the binary variables. For each binary variable, we maintained its raw values, rather than standardizing it to have a unit deviation.
[2]Note: By definition, *Number of Videos Uploaded Conditional on Uploading Anything* includes the nonzero values of *Number of Videos Uploaded*. We standardized the *Number of Videos Uploaded* across all providers involved in the first social-nudge experiment, and then extracted its nonzero standardized values to construct *Number of Videos Uploaded Conditional on Uploading Anything*.
[3,4,5]Note: For these three variables, we first standardized Number of Videos Uploaded across all providers involved in the first social-nudge experiment, and then extracted the standardized values for low-productivity, medium-productivity, and high-productivity providers, respectively.

Variable	Location	1%	25%	50%	75%	99%
Statistics Beyond the First Reception Day						
Number of Videos Uploaded on Day 2	Table 4	0.0000	0.0000	0.0000	0.0000	4.2770
Number of Videos Uploaded on Day 3	Table 4	0.0000	0.0000	0.0000	0.0000	3.8533
Number of Videos Uploaded on Day 4	Table 4	0.0000	0.0000	0.0000	0.0000	3.9571
Number of Social Nudges Sent on Day 2	Table 6	0.0000	0.0000	0.0000	0.0000	6.0318
Number of Social Nudges Sent on Day 3	Table 6	0.0000	0.0000	0.0000	0.0000	6.0384
Statistics About All Providers Who Were Sent at Least One Nudge in the First Social-Nudge Experiment (Including Those Who Received Nudges Prior to It)						
Number of Videos Uploaded on the First Reception Day	Table 10	0.0000	0.0000	0.0000	0.0000	4.9110
Number of Social Nudges Sent on the First Reception Day	Table 10	0.0000	0.0000	0.0000	0.0000	6.3290
Statistics Within 24 Hours Following the First Nudge						
Number of Videos Uploaded Within 24 Hours Following the First Nudge	Table 11	0.0000	0.0000	0.0000	0.0000	4.2495
Upload Incidence Within 24 Hours Following the First Nudge (Binary)	Table 11	0.0000	0.0000	0.0000	0.0000	1.0000
Statistics About Control Providers Used to Explore Whether Being Blocked from Social Nudges Caused Reactance						
Number of Videos Uploaded	Table 14	0.0000	0.0000	0.0000	0.0000	4.5216
Private Messages Incidence (Binary)	Table 14	0.0000	0.0000	0.0000	0.0000	1.0000
First Reception Day (Binary)	Table 14	0.0000	0.0000	0.5000	1.0000	1.0000
Statistics Broken Down by Whether Providers Received Any Private Message From their First Social-Nudge Sender						
Number of Videos Uploaded Among Providers Who Received Private Messages From the First Social-Nudge Sender	Table 14	0.0000	0.0000	0.0000	0.0000	7.3287
Number of Videos Uploaded Among Providers Who Received No Private Messages From the First Social-Nudge Sender	Table 14	0.0000	0.0000	0.0000	0.0000	3.6644

Table 26 Summary Statistics About the First (Main) Social-Nudge Experiment (II)

Statistics During the Experimental Period						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1	Table 12	0.0000	0.0000	0.0000	0.0000	3.7439
Number of Videos Uploaded on Day 2	Table 12	0.0000	0.0000	0.0000	0.0000	4.8366
Number of Videos Uploaded on Day 3	Table 12	0.0000	0.0000	0.0000	0.0000	5.0023
Number of Videos Uploaded on Day 4	Table 12	0.0000	0.0000	0.0000	0.0000	4.8774
Number of Social Nudges Sent on Day 1	Table 13	0.0000	0.0000	0.0000	0.0000	5.1553
Number of Social Nudges Sent on Day 2	Table 13	0.0000	0.0000	0.0000	0.0000	3.5356
Number of Social Nudges Sent on Day 3	Table 13	0.0000	0.0000	0.0000	0.0000	4.2500
Number of Social Nudges Sent on Day 4	Table 13	0.0000	0.0000	0.0000	0.0000	4.4280
Number of Social Nudges Sent on Day 5	Table 13	0.0000	0.0000	0.0000	0.0000	4.4868

Table 27 Summary Statistics About the Second (Replication) Social-Nudge Experiment (I)

Statistics During the Second Social-Nudge Experiment						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1	Table 18	0.0000	0.0000	0.0000	0.0000	4.6657
Number of Videos Uploaded on Day 2	Table 18	0.0000	0.0000	0.0000	0.0000	4.7433
Number of Videos Uploaded on Day 3	Table 18	0.0000	0.0000	0.0000	0.0000	4.7279
Number of Videos Uploaded on Day 4	Table 18	0.0000	0.0000	0.0000	0.0000	4.7239
Statistics About Providers Who Were Both in the First Social-Nudge and Platform-Initiated Nudge Experiments						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1 of the Platform-Initiated Nudge Experiment	Table 18	0.0000	0.0000	0.0000	0.0000	6.2209
Number of Videos Uploaded on Day 1 of the First Social-Nudge Experiment	Table 18	0.0000	0.0000	0.0000	0.0000	5.4965

Table 28 Summary Statistics About the Platform-Initiated Nudge Experiment (I)

Statistics About Data Used in the Cannibalization Analysis						
Variable	Location	1%	25%	50%	75%	99%
Number of Likes Marked	Table 16	0.0000	0.0095	0.0476	0.2667	5.2484
Number of Comments Left	Table 16	0.0000	0.0000	0.0098	0.0881	3.1047
Incidence of Sending Social Nudges (Binary)	Table 16	0.0000	0.0000	0.5000	1.0000	1.0000
Post (Binary)	Table 16	0.0000	0.0000	0.5000	1.0000	1.0000

Table 29 Summary Statistics About the Cannibalization Analysis (I)

Statistics During the First Social-Nudge Experiment						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1	Table 20	0.0000	0.0000	0.0000	0.0000	3.6644
Number of Videos Uploaded on Day 2	Table 20	0.0000	0.0000	0.0000	0.0000	4.2770
Number of Videos Uploaded on Day 3	Table 20	0.0000	0.0000	0.0000	0.0000	3.8533
Number of Videos Uploaded on Day 4	Table 20	0.0000	0.0000	0.0000	0.0000	3.9571
Number of Social Nudges Sent per Edge on Day 1	Table 21	0.0000	0.0000	0.0000	0.0000	1.4113
Number of Social Nudges Sent per Edge on Day 2	Table 21	0.0000	0.0000	0.0000	0.0000	1.2236
Number of Social Nudges Sent per Edge on Day 3	Table 21	0.0000	0.0000	0.0000	0.0000	0.9718
Statistics During the Second Social-Nudge Experiment						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1	Table 20	0.0000	0.0000	0.0000	0.0000	3.7439
Number of Videos Uploaded on Day 2	Table 20	0.0000	0.0000	0.0000	0.0000	4.8366
Number of Videos Uploaded on Day 3	Table 20	0.0000	0.0000	0.0000	0.0000	5.0023
Number of Videos Uploaded on Day 4	Table 20	0.0000	0.0000	0.0000	0.0000	3.2516
Number of Videos Uploaded on Day 5	Table 20	0.0000	0.0000	0.0000	0.0000	5.0074
Number of Social Nudges Sent per Edge on Day 1	Table 21	0.0000	0.0000	0.0000	0.0000	1.4468
Number of Social Nudges Sent per Edge on Day 2	Table 21	0.0000	0.0000	0.0000	0.0000	0.9906
Number of Social Nudges Sent per Edge on Day 3	Table 21	0.0000	0.0000	0.0000	0.0000	0.5149

Table 30 Summary Statistics About Providers Who Were Sent Only One Social Nudge During a Given Experiment (I)

Statistics Prior to the Experiment						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Female (Binary)	492,599	0.5134	0.4998	492,182	0.5138	0.4998
Number of Followers	496,976	0.0622	1.0383	496,700	0.0605	0.9602
Number of Following	496,976	0.8485	1.0008	496,700	0.8480	0.9992
Number of Uploaded Videos	496,976	0.3674	0.9851	496,700	0.3693	1.0147
Number of Days with Videos Uploaded	496,976	0.5057	0.9977	496,700	0.5078	1.0023
Historical Like Rate	430,522	0.9688	1.0003	430,771	0.9689	0.9997

Statistics Related to the First Reception Day						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded	496,976	0.2241	1.0537	496,700	0.1980	0.9430
Upload Incidence (Binary)	496,976	0.0770	0.2666	496,700	0.0677	0.2511
Number of Videos Uploaded Conditional on Uploading Anything[1]	38,281	2.9099	2.5688	33,602	2.9267	2.2715
Two-Way Tie (Binary)	496,976	0.4577	0.4982	496,700	0.4529	0.4978
Total Views	496,976	0.1816	1.0211	496,700	0.1645	0.9784
Complete View Rate	38,154	0.9526	1.0004	33,480	0.9519	0.9995
Like Rate	38,154	1.1530	0.9963	33,480	1.1703	1.0042
Comment Rate	38,154	0.7934	1.0032	33,480	0.8002	0.9964
Following Rate	38,154	0.2372	1.0152	33,480	0.2331	0.9824
Number of Social Nudges Sent	496,976	0.2412	1.0214	496,700	0.2087	0.9779
Number of Likes on the First Reception Day	496,976	0.2595	1.0071	496,700	0.2483	0.9928
Number of Comments on the First Reception Day	496,976	0.2104	1.0106	496,700	0.1996	0.9893
Number of Videos Uploaded Among Low-Productivity Providers[2]	450,919	0.1358	0.7745	450,367	0.1137	0.6519
Number of Videos Uploaded Among Medium-Productivity Providers[3]	41,791	0.8931	1.8235	42,047	0.8354	1.7681
Number of Videos Uploaded Among High-Productivity Providers[4]	4,266	3.0133	4.5146	4,286	2.7987	4.0303

Statistics Beyond the First Reception Day						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 2	496,976	0.2568	1.0042	496,700	0.2439	0.9958
Number of Videos Uploaded on Day 3	496,976	0.2608	1.0077	496,700	0.2543	0.9922
Number of Videos Uploaded on Day 4	496,976	0.2514	0.9989	496,700	0.2508	1.0011
Number of Social Nudges Sent on Day 2	496,976	0.1910	1.0158	496,700	0.1770	0.9838
Number of Social Nudges Sent on Day 3	496,976	0.1813	1.0037	496,700	0.1785	0.9963

Table 31 Summary Statistics About the First (Main) Social-Nudge Experiment (III)

[1,2,3,4]Note: See the notes in Table 25 for how we constructed these variables.

Statistics Among All Providers Who Were Sent at Least One Nudge in the First Social-Nudge Experiment						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on the First Reception Day	973,236	0.2551	1.0310	972,882	0.2329	0.9678
Number of Social Nudges Sent on the First Reception Day	973,236	0.2643	1.0246	972,882	0.2320	0.9745
Statistics Within 24 Hours Following the First Nudge						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded Within 24 Hours Following the First Nudge	496,976	0.2685	1.0339	496,700	0.2388	0.9646
Upload Incidence Within 24 Hours Following the First Nudge (Binary)	496,976	0.1121	0.3155	496,700	0.0990	0.2987
Statistics Based on Private-Message Incidence						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded Among Providers Who Received Any Private Messages From the First Social-Nudge Sender	15,073	0.5729	1.6714	13,069	0.5019	1.5337
Number of Videos Uploaded Among Providers Who Received No Private Messages From the First Social-Nudge Sender	481,903	0.2132	1.0265	483,631	0.1898	0.9204

Table 32 Summary Statistics About the First (Main) Social-Nudge Experiment (IV)

Experimental Statistics						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1	338,415	0.2152	1.0034	339,675	0.1924	0.9965
Number of Videos Uploaded on Day 2	338,415	0.2385	1.0175	339,675	0.2213	0.9822
Number of Videos Uploaded on Day 3	338,415	0.2190	0.9913	339,675	0.2107	1.0086
Number of Videos Uploaded on Day 4	338,415	0.2009	0.9727	339,675	0.1976	1.0265
Number of Social Nudges Sent on Day 1	338,415	0.2323	1.0264	339,675	0.1998	0.9727
Number of Social Nudges Sent on Day 2	338,415	0.1732	1.0301	339,675	0.1517	0.9689
Number of Social Nudges Sent on Day 3	338,415	0.1533	1.0141	339,675	0.1449	0.9858
Number of Social Nudges Sent on Day 4	338,415	0.1467	1.0127	339,675	0.1410	0.9872
Number of Social Nudges Sent on Day 5	338,415	0.1438	1.0096	339,675	0.1398	0.9903

Table 33 Summary Statistics About the Second (Replication) Social-Nudge Experiment (II)

Statistics During the Experimental Period						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1	5,522,864	0.2004	1.0080	5,520,612	0.1899	0.9919
Number of Videos Uploaded on Day 2	5,522,864	0.1945	1.0057	5,520,612	0.1919	0.9943
Number of Videos Uploaded on Day 3	5,522,864	0.1929	1.0022	5,520,612	0.1910	0.9978
Number of Videos Uploaded on Day 4	5,522,864	0.1921	1.0001	5,520,612	0.1910	0.9999
Statistics Among Providers Who Were Both in the First Social-Nudge and Platform-Initiated Nudge Experiments						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1 of the Platform-Initiated Nudges Experiment	31,752	0.3924	1.3493	31,715	0.3753	1.2725
Number of Videos Uploaded on Day 1 of the Social-Nudges Experiment	31,870	0.2601	1.0957	31,597	0.2315	1.0027

Table 34 Summary Statistics About the Platform-Initiated Nudge Experiment (II)

Statistics about the Observational Data Used in the Cannibalization Analysis						
	Post = 0 &					
	Incidence of Sending Social Nudges = 1			*Incidence of Sending Social Nudges = 0*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Likes Marked	353,041	0.5308	1.1520	353,041	0.2715	0.7699
Number of Comments Left	353,041	0.2907	1.2175	353,041	0.1071	0.6511
	Post = 1 &					
	Incidence of Sending Social Nudges = 1			*Incidence of Sending Social Nudges = 0*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Likes Marked	353,041	0.5807	1.1925	353,041	0.2658	0.7578
Number of Comments Left	353,041	0.3222	1.2725	353,041	0.1063	0.6588

Table 35 Summary Statistics About the Cannibalization Analysis (II)

Statistics in the First Social-Nudge Experiment						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1	481,376	0.2101	1.0192	480,744	0.1848	0.9071
Number of Videos Uploaded on Day 2	481,376	0.2395	0.9615	480,744	0.2275	0.9550
Number of Videos Uploaded on Day 3	481,376	0.2446	0.9731	480,744	0.2372	0.9513
Number of Videos Uploaded on Day 4	481,376	0.2349	0.9558	480,744	0.2347	0.9618
Number of Social Nudges Sent per Edge on Day 1	474,188	0.0811	0.9868	473,542	0.0733	0.9590
Number of Social Nudges Sent per Edge on Day 2	474,188	0.0649	1.0103	473,542	0.0601	0.9448
Number of Social Nudges Sent per Edge on Day 3	474,263	0.0559	0.9710	473,613	0.0536	0.9483
Statistics in the Second Social-Nudge Experiment						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1	326,817	0.2009	0.9634	328,184	0.1783	0.9430
Number of Videos Uploaded on Day 2	326,817	0.2216	0.9784	328,184	0.2041	0.9353
Number of Videos Uploaded on Day 3	326,817	0.2050	0.9572	328,184	0.1952	0.9617
Number of Videos Uploaded on Day 4	326,817	0.1883	0.9155	328,184	0.1829	0.9082
Number of Videos Uploaded on Day 5	326,817	0.2018	0.9646	328,184	0.1991	0.9654
Number of Social Nudges Sent per Edge on Day 1	319,822	0.0813	0.9882	321,098	0.0733	0.9516
Number of Social Nudges Sent per Edge on Day 2	319,822	0.0574	0.9484	321,098	0.0527	0.9703
Number of Social Nudges Sent per Edge on Day 3	319,925	0.0444	0.9922	321,194	0.0427	0.9646

Table 36 Summary Statistics About Providers Who Were Sent Only One Social Nudge During a Given Experiment (II)

Variables		Correlation
Number of Uploaded Videos in Prior One Week	Number of Days with Videos Uploaded in Prior One Week	0.8145
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 1	0.3617
Number of Uploaded Videos in Prior One Week	Upload Incidence on Day 1	0.3206
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 1 Conditional on Uploading Anything	0.3167
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 2	0.4259
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 3	0.4555
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 4	0.4380
Number of Days with Videos Uploaded in Prior One Week	Number of Videos Uploaded on Day 1	0.3258
Number of Days with Videos Uploaded in Prior One Week	Upload Incidence on Day 1	0.3673
Number of Days with Videos Uploaded in Prior One Week	Number of Videos Uploaded on Day 2	0.3813
Number of Days with Videos Uploaded in Prior One Week	Number of Videos Uploaded on Day 3	0.3945
Number of Days with Videos Uploaded in Prior One Week	Number of Videos Uploaded on Day 4	0.3830
Number of Days with Videos Uploaded in Prior One Week	Total Views	0.3190
Number of Videos Uploaded on Day 1	Upload Incidence on Day 1	0.7558
Number of Videos Uploaded on Day 1	Number of Videos Uploaded on Day 1 Conditional on Uploading Anything	1.0000
Number of Videos Uploaded on Day 1	Number of Videos Uploaded on Day 2	0.3529
Number of Videos Uploaded on Day 1	Total Views	0.5805
Upload Incidence on Day 1	Total Views	0.6197
Number of Videos Uploaded on Day 1 Conditional on Uploading Anything	Number of Videos Uploaded on Day 2	0.3377
Number of Videos Uploaded on Day 2	Number of Videos Uploaded on Day 3	0.4188
Number of Videos Uploaded on Day 2	Number of Videos Uploaded on Day 4	0.3015
Number of Videos Uploaded on Day 3	Number of Videos Uploaded on Day 4	0.4245
Like Rate	Comment Rate	0.4769
Like Rate	Historical Like Rate	0.5857
Comment Rate	Historical Like Rate	0.4253

Table 37 Correlation Between Variables When the Absolute Value of Correlation Is Greater Than 0.30

Notes: We only calculate the correlations between variables mentioned in the main text (Tables 1–6). These variables were measured among the full sample of providers in the first (main) social-nudge experiment. Any pair of variables that is not shown in this table has an absolute value of correlation coefficient below 0.30. For the variables that do not have values among some providers (e.g., *Number of Videos Uploaded on Day 1 Conditional on Uploading Anything*), we ignore the providers who do not values. That is, when we calculate the correlation between two variables, we only consider the providers who have values in both variables.

Statistics About the In-degrees and Out-degrees of the Sample Network									
Variable	0.1%	1%	5%	25%	50%	75%	95%	99%	99.9%
In-degrees	0.0556	0.0556	0.0556	0.0556	0.2222	0.6667	8.0000	49.6111	619.9022
Out-degrees	0.0556	0.0556	0.0556	0.1667	0.7222	2.7222	17.6667	46.8333	53.9444

Table 38 Summary Statistics About the Sample Network