

Online Appendix - Curation Algorithms and Filter Bubbles in Social Networks

A Robustness to Timing

We test the robustness of our model to the timing of the algorithm by letting the platform set β_Q , β_D and t simultaneously, followed by the receiver setting φ .

Using the function $E\varphi$ from Equation 52, we find two candidates for the optimal connectivity φ set by the receivers: $\varphi^* = \frac{\beta_Q q + \beta_Q - t}{2\beta_D}$ and $\varphi^* = \frac{\beta_Q(1+q-2c)+t}{4\beta_Q-2\beta_D}$.

Plugging back into the receiver's utility and maximizing over t , β_Q and β_D , the maximum is achieved when $\beta_Q^* = \beta_D^*$ and when $t^* = \beta_Q^* \cdot c$. The two potential solutions for φ^* presented above also coincide when the receiver's utility is maximized.

When $\beta_Q^* = 1$, we find that the PA algorithm from Section 7.1 sets exactly the same set of parameters and achieves the same profit.

B Convex Consumption Cost

In this section, we change the main model to allow for a convex content consumption cost function. Instead of assuming a constant c cost for each consumer piece of content, we introduce the cost function $C()$, where $C(J)$ denotes the total cost of reading J items. We assume that $C()$ is increasing and convex enough for an interior solution to exist. A simple example of a convex cost function that will satisfy this is $C(J) = cJ^2/2$. All other specification of the model remain the same as those of the main model presented in Section 3. We next analyze how the optimal filtering and friendship decisions change.

B.1 Benchmark

When there is no filtering algorithm at work, we start by deriving the optimal number of content items to be consumed, J . Given φ , the consumer maximizes

$$\mathbb{E}(\pi_R) = J \cdot (\mathbb{E}(U_j|\varphi)) - C(J) = J \cdot (q + 1/2 - \varphi) - C(J) \quad (1)$$

subject to $J < \varphi\alpha N$ in the last stage. Let $J^*(\varphi)$ denote the optimum which is given by the solution of $C'(J) = q + 1/2 - \varphi$ when φ is large enough or $J^*(\varphi) = \varphi\alpha N$ when φ is small. In the stage before, the consumer maximizes $J^*(\varphi)(q + 1/2 - \varphi) - C(J^*(\varphi))$ to obtain the optimal φ^* . We obtain the following result:

Lemma A1. *Consumers always choose to read all content available in equilibrium: $J^*(\varphi^*) = \varphi^*\alpha N$*

Proof. Assume on the contrary that $J^*(\varphi^*) < \varphi^*\alpha N$. Decreasing φ to $\varphi' = J^*(\varphi^*)$ would lead to a strictly higher payoff, a contradiction. \square

Using the lemma, we can simplify the maximization problem and simply solve

$$\max_{\varphi} \mathbb{E}(\pi_R) = \varphi \cdot (q + 1/2 - \varphi) - C(\varphi) \quad (2)$$

The optimal φ^* is given by the solution of $q + 1/2 = C'(\varphi) + 2\varphi$, yielding the following:

Proposition A1. *The friendship proportion chosen by consumers is increasing in quality and decreasing in the cost of content consumption. In the case of quadratic content consumption costs of $C(J) = (c/2)J^2$, we have $\varphi^* = \min\left(1, \frac{q+1/2}{c+2}\right)$.*

This result is similar to that in the basic model's benchmark with the only difference being the functional form.

B.2 Filtering

We now move to examine the case of a filtering algorithm at work. Compared to the basic model, there is an additional consideration. Since the receiver cost for each consumed message is different, and the receiver's utility is affected by it, a good algorithm will take that into account. Hence, we assume that all filtering algorithms track the number of items consumed by users and are informed about the marginal cost of reading each successive item. In case of the Perfect Algorithm, for

example, this implies that the algorithm sets the threshold at $t = c_i$, where c_i is the marginal cost of reading message i . Similarly, the threshold is $c_i + \varphi$ in case of the quality algorithm and $c_i - q - 1/2$ for the distance algorithm.

We first establish that similar to the benchmark, consumers always choose to read all content available in equilibrium.

Lemma A2. *Under any algorithm, $J^*(\varphi^*) = \varphi^* \alpha N$*

Proof. The proof is the same as without an algorithm. If there was an optimal φ^* with $J^* < \alpha \varphi^* N$, the consumer would be better off decreasing φ , because it decreases the expected distance of all messages, while not decreasing any utility for the messages read. \square

Next, we examine the consumer payoff under filtering. For example, in the case of the perfect algorithm, the expected payoff from item one is similar to the benchmark case:

$$\Pr(U_1 > c_1) \mathbb{E}(U_1 - c_1 | \varphi, U_1 > c_1) \quad (3)$$

For the second item arriving, it is slightly more complicated because if the first item has been filtered out the algorithm uses the c_1 for the threshold since the user has not read a single item.

$$\begin{aligned} & \Pr(U_1 > c_1) \Pr(U_2 > c_2) \mathbb{E}(U_2 - c_2 | \varphi, U_2 > c_2) + \\ & (1 - \Pr(U_1 > c_1)) \Pr(U_2 > c_1) \mathbb{E}(U_2 - c_1 | \varphi, U_2 > c_1) \end{aligned} \quad (4)$$

In general, for the j th arriving item, the payoff is $\sum_{i=1}^j PrF_{j-1}^{j-i} \Pr(U_j > c_i) \mathbb{E}(U_j - c_i | \varphi, U_j > c_i)$ where PrF_j^h denotes the probability that h items have been filtered out of the first j arriving pieces of content. Hence the total payoff from all content is

$$\begin{aligned} & \sum_{j=1}^{\varphi \alpha N} \sum_{i=1}^j PrF_{j-1}^{j-i} \Pr(U_j > c_i) \mathbb{E}(U_j - c_i | \varphi, U_j > c_i) = \\ & = \sum_{i=1}^{\varphi \alpha N} \sum_{j=i}^{\varphi \alpha N} PrF_{j-1}^{j-i} \Pr(U_j > c_i) \mathbb{E}(U_j - c_i | \varphi, U_j > c_i) \end{aligned} \quad (5)$$

Proposition A2. *The optimal friendship proportion choices satisfy $\varphi_{NA}^* = \varphi_{DA}^* \leq \varphi_{QA}^* \leq \varphi_{PA}^*$ and all of these quantities are increasing in q .*

Proof. First, we show that $\varphi_{NA}^* = \varphi_{DA}^*$. This is similar to the basic model: distance filtering does not filter any message in equilibrium since the user does not have any incentive to add more

friends whose messages would be filtered out anyway. Given no filtering, the optimal φ is the same as under NA. Next, we prove $\varphi_{NA}^* \leq \varphi_{QA}^*$, by showing that the consumer payoff under the quality algorithm has a non-negative derivative when setting $\varphi = \varphi_{NA}^*$, yielding that the optimal φ under QA must be at least as high as under NA. Increasing φ has two effects: i) it adds an additional message to be potentially received by the user and ii) it reduces the expected payoff from messages received even without the increase (due to the increased distance). The derivative is a sum of this (typically) positive (i) and negative (ii) component. Clearly, the positive effect (i) is larger under QA than under NA because the algorithm filters out low utility messages, hence the expected benefit from an extra message is higher. Furthermore the negative effect (ii) is smaller under the quality algorithm. Under NA the derivative is simply -1 for each message that arrives. For the QA the derivative is larger than -1 (smaller in absolute value) for each message because filtering decreases the sensitivity to a shifted distance distribution. Since the positive component of the derivative is higher under QA than under NA and negative is smaller in absolute value, and gives that the total derivative is 0 under NA, it must be positive under QA. We can show that $\varphi_{QA}^* \leq \varphi_{PA}^*$ in the same fashion as PA is a better filtering mechanism, leading to a higher payoff and less dependence on φ for each arriving message. Finally, we show that the optimal φ is increasing in q for both the QA and PA. In these cases, the addition of an extra message (effect i) clearly increases with q while the decrease in utility from existing messages (ii) does not change with q . Hence, increasing q , leads to a higher optimal φ . \square

C Cost and Quality Heterogeneity

In our model all receivers and senders are symmetric resulting in a very regular and symmetric equilibrium outcome. In practice there is likely much more heterogeneity that we do not account for in our basic model. Here, we relax these assumption to demonstrate how our model could be adapted to a more asymmetric setting. We assume that each receiver has a different consumption cost c_i drawn uniformly from the $U(0, 2c)$ interval. Furthermore, we assume that there are K equal sized segments with $\alpha N/K$ senders in each. Qualities are different between segments and equal within segments such that every sender in segment k has quality q_k . Although we only model receiver actions here and consider sender qualities exogenous, the difference in qualities could be a result of different sender costs between the different segments.

As in our basic model, receivers will make decisions on how much content to consume and how many senders to connect to. Here, we assume that each receiver will make a separate connectivity decision for each sender segment. The connectivity of receiver i into segment k is denoted by φ_{ik} .

Building on the results of our main model, we can derive the ‘in-degree’ of each segment and each sender. We denote by $^{in}d_k$ the number of receivers that connect to a sender with quality q_k in segment k . This can be calculated as $(1 - \alpha)N \int_0^{2c} \varphi(x - q_k) \frac{1}{2^c} dx$ where $\varphi(x - q)$ is the connectivity derived in Section 4 for a receiver with cost x when quality is q for any of the four algorithms. For the sake of simplicity, we derive the degrees for $c = 1$. In the no algorithm, distance and perfect algorithm cases we get

$$\frac{^{in}d_k^{NA}}{(1 - \alpha)N} = \frac{^{in}d_k^{DA}}{(1 - \alpha)N} = \begin{cases} 1 & q_k \geq_k \frac{7}{2} \\ 1 - \frac{(7/2 - q_k)^2}{8} & \frac{7}{2} > q_k \geq_k \frac{3}{2} \\ \frac{(q_k + 1/2)^2}{8} & \frac{3}{2} > q_k \geq_k -\frac{1}{2} \\ 0 & -\frac{1}{2} > q_k \end{cases} \quad \frac{^{in}d_k^{PA}}{(1 - \alpha)N} = \begin{cases} 1 & q_k \geq_k 3 \\ 1 - \frac{(3 - q_k)^2}{8} & 3 > q_k \geq_k 1 \\ \frac{(q_k + 1)^2}{8} & 1 > q_k \geq_k -1 \\ 0 & -1 > q_k \end{cases}$$

while in the quality algorithm case we obtain

$$\frac{^{in}d_k^{q_k A}}{(1 - \alpha)N} = \begin{cases} 1 & q_k \geq_k \frac{7}{2} \\ 1 - \frac{(7/2 - q_k)^2}{8} & \frac{7}{2} > q_k \geq_k \frac{5}{2} \\ 1 - \frac{(7/2 - q_k)^2}{8} + \frac{(5/2 - q_k)^2}{24} & \frac{5}{2} > q_k \geq_k \frac{3}{2} \\ \frac{(q_k + 1/2)^2}{8} + \frac{8q_k - 4q_k^2 - 1}{48} & \frac{3}{2} > q_k \geq_k 1 \\ \frac{(q_k + 1)^2}{12} + \frac{(q_k - 1/2)^2}{24} & 1 > q_k \geq_k \frac{1}{2} \\ \frac{(q_k + 1)^2}{12} & \frac{1}{2} > q_k \geq_k -1 \\ 0 & -1 > q_k \end{cases}$$

Analyzing the functions, we see that $^{in}d_k^{NA} \leq ^{in}d_k^{QA} \leq ^{in}d_k^{PA}$. That is, just as the overall connectivity in our basic model, the perfect algorithm produces the highest in-degree for a sender followed by the quality algorithm and the no algorithm (and distance algorithm).

All the three functions are S-shaped in q_k . Higher quality senders therefore get more incoming connections. This shows a more realistic representation of a social network where some nodes are more popular. The reasons for popularity can be multiple-fold, but a common argument is that

some senders produce better quality due to lower costs. Heterogeneous costs therefore lead to skewed degree distribution with a number of more popular nodes. Due to the S-shaped function the degrees can be even more skewed than the qualities.

The out-degrees in our heterogeneous model also vary due to the content consumption costs. Receivers with lower cost will connect to more senders (from each segment). Depending on the dispersion of these costs the degree-distribution can be more or less skewed.

Finally, we note that skewed degrees and less than full connectivity are likely to hold in the full equilibrium with endogenous quality choices. Consider an example where all but one segment of senders has prohibitively high production costs leading to $q_k = 0$ for $\forall k \neq H$. Segment H has very low or zero production cost, leading to an arbitrarily high q_H . In this case (with receiver costs distributed as $U(0, 2)$), every receiver connects to all senders in segment H , but only to a portion of senders in the rest of the network ($1/4$ for NA , $1/3$ for QA , $1/2$ for PA and these numbers could even be lower if we assumed that zero quality investment leads to less than $1/2$ expected quality). Since there is only one high quality segment out of many, the overall connectivity level is strictly less than 1 for all algorithms. Although this is an extreme example, we speculate that the equilibrium quality outcome depends continuously on the sender production costs parameter space, leading to less extreme interior solutions with proportional connectivity.

D Distance Heterogeneity

In this section we focus on cases where the preference for distances of content items may depend on the location of the senders. This allows us to analyze markets with inherent polarization between segments of the population, as well as test our model for robustness of results when costs of content consumption are non-linear.

D.1 Segmented Senders

Suppose the senders can be segmented into two segments, a close segment (1) and a far segment (2). The distance of content generated by segment 1 is distributed $U[0, 2\varphi]$ if $\varphi \leq 1/2$ and $U[0, 1]$ if $\varphi > 1/2$, while the content in segment 2 is distributed $U[1, 2\varphi]$ if $\varphi > 1/2$. The content a receiver sees is generated randomly by the users in these segments. Hence, the distances a receiver observes is a mixture of the distances from each segment weighted by the size of each segment. If they set

$\varphi \leq 1/2$, they observe content from segment 1 with probability 1. If they set $\varphi > 1/2$, they observe content from segment 1 with probability $\frac{1/2}{\varphi}$ and from segment 2 with probability $\frac{\varphi-1/2}{\varphi}$.

A straightforward analysis of the resulting mixture CDF of the distance distribution leads to the following result:

Lemma A3. *The mixture distribution of distances from the two segments has the same CDF as the distribution $U[0, 2\varphi]$ for any $0 < \varphi \leq 1$.*

Notably, Lemma A3 generalizes to any number of segments with this mixture structure.¹ Given segments numbered $k = 1, \dots, m$ each having a distance distribution $U[\frac{2(k-1)}{m}, \min(\frac{2k}{m}, 2\varphi)]$, the mixture distribution of distances where each sender has a probability of $\frac{\min(k/m, \varphi) - (k-1)/m}{\varphi}$ of appearing in the receiver's content feed equals $U[0, 2\varphi]$. Hence, the results of the paper hold for any finite segmentation of the market. Specifically, the results hold if each sender produces content with a non-random distance, but a-priori the receiver would not know from which receiver they will read messages before logging onto the social network.

D.2 Polarized Receivers

A second analysis we perform looks at the case of a polarized market, where receivers may have strong preferences not to read content of a specific type. For example, consumers may be unhappy if they read content that does not match their political ideology. Using the same two segment model introduced above, we now assume that the cost of reading an item from segment 1 is c_1 and reading an item from segment 2 costs the receiver c_2 , with $c_2 > c_1$. Performing the analysis in the benchmark (NA), PA and QA cases, we arrive at the following result.

Proposition A3. *When the cost $c_2 > c_1$:*

- $\varphi_{PA}^* \geq \varphi_{QA}^* \geq \varphi_{NA}^*$.
- *The utility of the receiver is highest in the PA case, intermediate in the QA case, and lowest in the NA case.*
- *There is an intermediate range of qualities for which the receiver does not connect to additional senders if quality increases, and only connects to all senders in segment 1.*

¹The proof stems from a straightforward sum of piece-wise segment CDFs.

Proof. In the benchmark case (NA), the expected utility of the receiver is:

$$\mathbb{E}[\pi_R^{NA}] = \begin{cases} \varphi\alpha N \frac{1+2(q-\varphi-c_1)}{2} & 0 < \varphi \leq 1/2 \\ \frac{\alpha N(c_2-c_1+\varphi(1+2q-2\varphi-2c_2))}{2} & 1/2 < \varphi \leq 1 \end{cases} \quad (6)$$

The optimal connectivity φ_{NA}^* equals:

$$\varphi_{NA}^* = \begin{cases} 0 & c_1 > q + 1/2 \\ \frac{1-2(c_1-q)}{4} & q - 1/2 < c_1 \leq q + 1/2 \\ 1/2 & c_1 \leq q - 1/2 < c_2 \\ \frac{1-2(c_2-q)}{4} & q - 3/2 < c_2 \leq q - 1/2 \\ 1 & c_2 \leq q - 3/2 \end{cases} \quad (7)$$

When using the perfect algorithm (PA), the platform will filter content from segment 1 if it has utility above c_1 and from segment 2 if it has utility above c_2 . The resulting expected utility of the receiver is:

$$\mathbb{E}[\pi_R^{PA}] = \begin{cases} \varphi\alpha N \int_0^{2\varphi} \int_{\max(c_1+D,q)}^{\max(c_1+D,q+1)} \frac{Q-D-c_1}{2\varphi} dQdD & 0 < \varphi \leq 1/2 \\ \varphi\alpha N \left(\int_0^1 \int_{\max(c_1+D,q)}^{\max(c_1+D,q+1)} \frac{Q-D-c_1}{2\varphi} dQdD + \int_1^{2\varphi} \int_{\max(c_2+D,q)}^{\max(c_2+D,q+1)} \frac{Q-D-c_2}{2\varphi} dQdD \right) & \frac{1}{2} < \varphi \leq 1 \end{cases} \quad (8)$$

Solving for the optimal connectivity φ_{PA}^* results in:

$$\varphi_{PA}^* = \begin{cases} 0 & c_1 > q + 1 \\ \frac{1+q-c_1}{2} & q < c_1 \leq q + 1 \\ 1/2 & c_1 < q < c_2 \\ \frac{1+q-c_2}{2} & q - 1 < c_2 \leq q \\ 1 & c_2 \leq q - 1 \end{cases} \quad (9)$$

Finally, as the quality algorithm does not observe individual distances of senders, it cannot infer the cost of consuming a content item for each receiver. When $\varphi < 1/2$, the QA algorithm will therefore filter items with quality lower than $c + \varphi$ as before, however, when $\varphi > 1/2$, the algorithm will filter items with quality below the expected distance plus the expected cost, $t_{QA}^* = \frac{c_1-c_2}{2\varphi} + c_2 + \varphi$, resulting in the profit:

$$\mathbb{E}[\pi_R^{QA}] = \begin{cases} \varphi\alpha N \int_0^{2\varphi} \int_{\max(c_1+\varphi,q)}^{\max(c_1+\varphi,q+1)} \frac{Q-D-c_1}{2\varphi} dQdD & 0 < \varphi \leq 1/2 \\ \varphi\alpha N \left(\int_0^1 \int_{\max(t_{QA}^*,q)}^{\max(t_{QA}^*,q+1)} \frac{Q-D-c_1}{2\varphi} dQdD + \int_1^{2\varphi} \int_{\max(t_{QA}^*,q)}^{\max(t_{QA}^*,q+1)} \frac{Q-D-c_2}{2\varphi} dQdD \right) & \frac{1}{2} < \varphi \leq 1 \end{cases} \quad (10)$$

The maximization over φ results in:

$$\varphi_{QA}^* = \begin{cases} 0 & c_1 > q + 1 \\ \frac{1+q-c_1}{3} & q - 1/2 < c_1 \leq q + 1 \\ 1/2 & c_1 < q - 1/2 < c_2 \\ \frac{1}{4} + \frac{q-c_2}{2} & q - 3/2 < c_2 \leq q - 1/2 \\ 1 & c_2 \leq q - 3/2 \end{cases} \quad (11)$$

A comparison of the resulting connectivity shows that $\varphi_{PA}^* \geq \varphi_{QA}^* \geq \varphi_{NA}^*$. Plugging in the equilibrium connectivity into the utility of the receivers, we find that $\pi_{R,PA}^* \geq \pi_{R,QA}^* \geq \pi_{R,NA}^*$. \square

The first two items of Proposition A3 shows that our results are robust to market polarization where consumers have a strict preference for content from senders with closer distances to them. As before, the perfect algorithm is the most efficient in increasing connectivity and receiver utility, followed by the quality algorithm.

The last item shows that because there is a gap between the costs of consuming content from segment 1 vs. 2, there will be a range of qualities for which receivers will not increase connectivity even if senders increase the quality of their content, until quality is high enough. The intuition is that once the receiver has connected to all senders in segment 1, connecting to additional senders will lower utility (because of the increased cost), unless quality is high enough for this additional connectivity.

An implication of this result is that when consumers have polarized preferences, although curation algorithms increase connectivity, they do so only in a weak sense. For certain ranges of qualities of content, algorithms will have no impact. Moreover, even if the polarizing content provider increases the quality of their content to attract more readers, many receivers will still prefer to not follow them and ignore their content, until there is a dramatic improvement in quality. As the proof of Proposition A3 shows, there is a range of qualities where connectivity is “sticky” at 1/2 for the QA and NA cases. In this range, the perfect algorithm drives connectivity of either 1/2 or higher, and hence produces a lower hurdle for senders to increase connectivity in a polarized market.

E Impressions vs. Likes

When the utility of the sender depends on the number of impressions receivers consume, their utility is:

$$\frac{\varphi N}{2} Pr(s \geq t) - kq \quad (12)$$

When no curation algorithm is in play, $Pr(s \geq t) = 1$, and the sender's optimal quality to set is $q_{NA}^* = 0$. As a result, the equilibrium connectivity will be $\varphi_{NA}^* = \frac{1-2c}{4}$ when $c < 1/2$, and $\varphi_{NA}^* = 0$ when $c \geq 1/2$. The maximum connectivity is $\varphi_{NA}^* = 1/4$.

When the quality algorithm is being used by the social network, we can use the value of t^* developed for the receiver side to calculate:

$$Pr(s_{QA} \geq t_{QA}^*) = \begin{cases} 2\varphi & c + \varphi \leq q \\ 2\varphi(q - c + 1 - \varphi) & 0 < c - q + \varphi < 1 \\ 0 & c - q + \varphi \geq 1 \end{cases} \quad (13)$$

Plugging into the sender's profit function, it can be written as:

$$\pi_S^{QA} = \begin{cases} -kq & c - q + \varphi \geq 1 \\ q(N\varphi^2 - k) + N\varphi^2(1 - \varphi - c) & 0 < c - q + \varphi < 1 \\ N\varphi^2 - kq & q \geq c + \varphi \end{cases} \quad (14)$$

Hence, if $N\varphi^2 - k < 0$, then $q_{QA}^* = 0$, and otherwise, $q_{QA}^* = c + \varphi$ (which is the filtering threshold for the QA algorithm), as long as $N\varphi^2 - k(c + \varphi) > 0$ (in which case q_{QA}^* is again zero). Using the results of the equilibrium levels of φ_{QA}^* from Section 4.3, we find that $q_{QA}^* = c + 1/2$ when $k \leq \frac{N}{4(c+1/2)}$ and zero otherwise. Correspondingly, $\varphi_{QA}^* = 1/2$ when $k \leq \frac{N}{4(c+1/2)}$ and $\frac{1-c}{3}$ when k is larger.

When the perfect algorithm is being used, the only part of the first order condition which may result in a maximum yields:

$$\frac{N}{2}\varphi(c - q + 2\varphi) = k \quad (15)$$

Using the receiver's side result for φ_{PA} , we arrive at the solution $q_{PA}^* = c - \frac{2k}{N} + 2$ and $\varphi_{PA}^* = 1$.

The results replicate the results found in the full model where senders benefit from receiving "likes". The quality and connectivity under PA is higher than under QA which in turn is higher than under NA.

F Numerical Solution to the Platform's Algorithm Selection

Finding the numerical solution to the platform's optimization problem is not straightforward for two reasons:

1. An algorithm is described by a two dimensional parameter (β_Q, β_D) .
2. Given an algorithm, it is not always possible to find the profit maximizing quality the Sender will set without numerically searching for equilibria (instead of solving first order conditions).

To simplify the task, we first prove the following Lemma:

Lemma A4. *The algorithm parameters β_D and β_Q can be represented by a single parameter $\beta = \frac{\beta_D}{\beta_Q}$.*

Proof. The score $s = \beta_Q \cdot Q - \beta_D \cdot D$ only plays a role in the filtering when the platform determines if $s \geq t$ and whether an item passes the filter or not.

As the proof of Proposition 9 shows, there are two possible values of t the platform will set: $\beta_Q(c + \varphi) - \beta_D\varphi$ and $\frac{\beta_Q(2\beta_Dc + (\beta_Q - \beta_D)q) + 2\beta_D\varphi(\beta_Q - \beta_D)}{\beta_D + \beta_Q}$.

By replacing β_D with $\beta_D = \beta\beta_Q$, the first condition can be written as:

$$s \geq t \iff \beta_Q \cdot Q - \beta_D \cdot D \geq \beta_Q(c + \varphi) - \beta_D\varphi \quad (16)$$

$$\iff Q - \beta D \geq c + \varphi - \beta\varphi \quad (17)$$

while the second can be written as:

$$s \geq t \iff \beta_Q \cdot Q - \beta_D \cdot D \geq \frac{\beta_Q(2\beta_Dc + (\beta_Q - \beta_D)q) + 2\beta_D\varphi(\beta_Q - \beta_D)}{\beta_D + \beta_Q} \quad (18)$$

$$\iff Q - \beta D \geq \frac{(2\beta c + (1 - \beta)q) + 2\beta\varphi(1 - \beta)}{\beta + 1} \quad (19)$$

Hence, every algorithm where $\beta_Q \neq 0$ can be represented using a single parameter β . Because we have seen that setting $\beta_Q = 0$ applies the distance algorithm, which is equivalent to not using an algorithm, we can focus only on the cases of $\beta_Q \neq 0$. \square

The lemma says that two filtering algorithms represented by (β_Q, β_D) and (β'_Q, β'_D) are equivalent in equilibrium outcomes if $\frac{\beta'_D}{\beta'_Q} = \frac{\beta_D}{\beta_Q}$.

Hence, in our numerical analysis we focus on finding the optimal value of β (the ratio between β_D and β_Q that maximizes the receiver's utility in equilibrium). To resolve the second issue limiting

calculation of an equilibrium, we focus on the case of $r = c = 0$ which allows calculating the symmetric first order condition the Senders face. When $r = c = 0$, and when $0 \leq \beta \leq 1$, the symmetric first order condition of the Sender depends on whether $\varphi < 1$ or $\varphi = 1$. When $\varphi < 1$ it can be written as:

$$\begin{cases} (-2 + q - \beta q + 2k(1 + q))/(-3 + \beta) = 0 & 1 + \beta + 2\beta q \leq 2q \\ ((k(1 - 2q)^2)/\beta + 2(-3 + \beta)(-2 + q - \beta q) - 2k(5 + 6q + 4q^2) + \\ \beta k(5 + 4q(2 + q)))/(2(-3 + \beta)^2) = 0 & (-2 + \beta)q \leq -1 \text{ and } 1 + \beta + 2\beta q > 2q \\ \frac{6+k+(-2+\beta)\beta(-2+5k)-12q-4kq+4\beta(4-3k+\beta(-1+2k))q+4(-1+\beta)^2kq^2}{2(-3+\beta)^2\beta} = 0 & \textit{otherwise} \end{cases} \quad (20)$$

and when $\varphi = 1$ the first order condition is:

$$\begin{cases} 2 - 2k - q = 0 & 1 + \beta \leq q \\ 2 - q + (k(\beta^2 + (-1 + q)^2 - 2\beta(1 + q)))/(2\beta) = 0 & q < 1 + \beta \text{ and } q > 1 \\ \frac{2+k-2\beta k+2(-1+\beta-k)q}{2\beta} = 0 & q < 1 \text{ and } (\beta + \beta q < 1 \text{ and } ((\beta = q \text{ and } \beta + 2\beta q = 1) \text{ or } (\beta + 2\beta q > 1 \text{ and } \beta \geq q))) \\ \frac{2+k-2\beta k+2(-1+\beta-k)q}{2\beta} = 0 & q < 1 \text{ and } (\beta + 2\beta q > 1 \text{ and } \beta \geq q \text{ and } (\beta + q > 1 \text{ or } \beta + \beta q > 1)) \\ \frac{2+2(-1+\beta)q+k(\beta^2+(-1+q)^2-2\beta(1+q))}{2\beta} = 0 & \textit{otherwise} \end{cases} \quad (21)$$

Using these closed forms and the receiver's equilibrium connectivity and utility that can be found in the proof for Proposition 9, we can calculate, for each value of β the equilibrium receiver's utility, and thus find the optimal value the platform will set for β .

G Endogenous Sender versus Receiver Decision

An alternative strategy to increasing the quality of her content is for a sender to increase the amount of messages she sends. Senders may also choose to become receivers if competition is so strong that the utility from receiving surpasses that of sending.

We note that this scenario might not necessarily be more realistic than the one with an exogenous share of senders, as many times on social media content producers are professional organizations who cannot switch to being receivers. To analyze this scenario, we allow consumers to be either a sender or a receiver in each content consumption/creation opportunity. We operationalize this decision by letting consumers set α endogenously while simultaneously deciding on φ and q . Effectively, we are considering consumers as mixing between sending and receiving. In this case, consumers will set α that makes them indifferent between reading and sending one content item.² Following the

²This setup can be easily extended to assuming there is an "exchange-rate" between the utility from sending one

derivation in Section 4 the expected utility from reading one item $Pr(S_j > t)\mathbb{E}[U_j - c|S_j > t]$ does not depend on α .³ Consequently, the decision of the platform on the filtering threshold t as well as the decision on the connectivity φ will not change.

Section 6 shows, however, that the utility from sending one item depends on α in two ways: a *higher* α (1) *increases* competition for likes and (2) *decreases* the number of potential readers for an item. Taking these effects into account, the consumers solve the following optimization problem:

$$\varphi^* = \arg \max_{\varphi} \mathbb{E}[\pi_R] = \arg \max_{\varphi} JPr(S_j > t)\mathbb{E}[U_j - c|S_j > t] \quad (22)$$

$$q^* = \arg \max_{q_j} \mathbb{E}[\pi_S] = \arg \max_{q_j} \varphi(1 - \alpha)N\mathbb{E}\left[\frac{Pr(U_j > r \text{ and } S_j > t)}{J}\right] - kq_j \quad (23)$$

subject to:

$$\mathbb{E}[\pi_S] = Pr(S_j > t)\mathbb{E}[U_j - c|S_j > t] \quad (24)$$

Our analysis focuses on the case of $r = c = 0$, and yields the following result:

Proposition A4. *When consumers set φ , q and α simultaneously:*

- q and φ increase with k , while α decreases with k .
- $\alpha_{PA}^* > \alpha_{NA}^* > \alpha_{QA}^*$ for low values of k and $\alpha_{PA}^* > \alpha_{NA}^* = \alpha_{QA}^*$ for higher values of k .
- There is a $0 < \tilde{k} < 1$ such that receiver utility is highest under the quality algorithm when $0 < k < \tilde{k}$ and is highest under the perfect algorithm when $k > \tilde{k}$.

Proof. Denote $a = \frac{1-\alpha}{\alpha}$. For each algorithm, we solve the first order conditions of the receiver over φ and the sender over q , with the restriction that:

$$\mathbb{E}[\pi_s] = Pr(S_j > t)\mathbb{E}[U_j|S_j > t] \quad (25)$$

We then complete the solution for the range of parameters when $\varphi = 1$, $q = 0$, or when the algorithms do not filter content and yield a corner solution.

item to reading one item. We assume a rate of 1:1 for parsimony.

³The NA case can be modeled as setting $t = -\infty$.

The NA solution results in:

$$\varphi_{NA}^* = \begin{cases} 1/4 & 0 < k \leq \frac{1}{3} \\ \frac{k}{4(1-2k)} & \frac{1}{3} < k \leq \frac{4}{9} \\ 1 & \frac{4}{9} < k \leq 1 \end{cases} \quad q_{NA}^* = \begin{cases} 0 & 0 < k \leq \frac{1}{3} \\ \frac{3k-1}{2(1-2k)} & \frac{1}{3} < k \leq \frac{4}{9} \\ \frac{4}{5-\sqrt{9-8k}} & \frac{4}{9} < k \leq 1 \end{cases}$$

$$a_{NA}^* = \begin{cases} 1/3 & 0 < k \leq \frac{1}{3} \\ \frac{k^2}{1-2k} & \frac{1}{3} < k \leq \frac{4}{9} \\ k + \frac{\sqrt{9-8k}+3}{4} & \frac{4}{9} < k \leq 1 \end{cases}$$

The PA solution results in:

$$\varphi_{PA}^* = \begin{cases} 1/2 & 0 < k \leq \frac{1}{3} \\ \frac{1}{12} (3k + \sqrt{9k(k+2)} - 3 + 3) & \frac{1}{3} < k \leq \frac{7}{6} \end{cases} \quad q_{PA}^* = \begin{cases} 0 & 0 < k \leq \frac{1}{3} \\ \frac{1}{6} (3k + \sqrt{9k(k+2)} - 3 - 3) & \frac{1}{3} < k \leq \frac{7}{6} \end{cases}$$

$$a_{PA}^* = \begin{cases} 1/3 & 0 < k \leq \frac{1}{3} \\ \frac{1}{6} k (3k + \sqrt{9k(k+2)} - 3) & \frac{1}{3} < k \leq \frac{7}{6} \end{cases}$$

The QA solution is:

$$\varphi_{QA}^* = \begin{cases} 1/2 & 0 < k \leq \frac{2}{5} \\ \frac{k}{4(1-2k)} & \frac{2}{5} < k \leq \frac{4}{9} \\ 1 & \frac{4}{9} < k \leq 1 \end{cases} \quad q_{QA}^* = \begin{cases} 1/2 & 0 < k \leq \frac{2}{5} \\ \frac{3k-1}{2(1-2k)} & \frac{2}{5} < k \leq \frac{4}{9} \\ \frac{4}{5-\sqrt{9-8k}} & \frac{4}{9} < k \leq 1 \end{cases}$$

$$a_{QA}^* = \begin{cases} \frac{4(k+1)}{7} & 0 < k \leq \frac{2}{5} \\ \frac{k^2}{1-2k} & \frac{2}{5} < k \leq \frac{4}{9} \\ k + \frac{\sqrt{9-8k}+3}{4} & \frac{4}{9} < k \leq 1 \end{cases}$$

A simple analysis shows that:

1. $\alpha_{PA}^* > \alpha_{NA}^* > \alpha_{QA}^*$ for low values of k and $\alpha_{PA}^* > \alpha_{NA}^* = \alpha_{QA}^*$ for higher values of k .
2. $q_{QA}^* > q_{NA}^* = q_{PA}^*$ for low values of k , $q_{QA}^* > q_{NA}^* > q_{PA}^*$ for intermediate k 's and $q_{QA}^* = q_{NA}^* > q_{PA}^*$ for high k 's.
3. $\varphi_{PA}^* = \varphi_{QA}^* > \varphi_{NA}^*$ for low values of k , $\varphi_{PA}^* > \varphi_{QA}^* > \varphi_{NA}^*$ for intermediate k 's and $\varphi_{NA}^* = \varphi_{QA}^* > \varphi_{PA}^*$ for high k 's.

To compare receiver utilities, we calculate $\frac{1}{1+a} \varphi \mathbb{E}[U_j | \varphi]$ for each algorithm. The results then

follow from a simple comparison:

$$\mathbb{E}[\pi_{R,NA}^*] = \begin{cases} \frac{3}{64} & k \leq \frac{1}{3} \\ \frac{k^2}{16(k-1)^2(1-2k)} & \frac{1}{3} < k \leq \frac{4}{9} \\ \frac{2(\sqrt{9-8k}+3)}{(5-\sqrt{9-8k})(4k+\sqrt{9-8k}+7)} & \frac{4}{9} < k \leq 1 \end{cases} \quad (26)$$

$$\mathbb{E}[\pi_{R,PA}^*] = \begin{cases} \frac{1}{14} & k \leq 1/3 \\ -\frac{k}{k(\sqrt{9k(k+2)-3}+7)-2\sqrt{9k(k+2)-3}-3k^2+6}} & \frac{1}{3} < k \leq 1 \end{cases} \quad (27)$$

$$\mathbb{E}[\pi_{R,QA}^*] = \begin{cases} \frac{175}{96(4k+11)} & k \leq 2/5 \\ \frac{2(2-5k)k-1}{96(k-1)^2(2k-1)} & \frac{2}{5} < k \leq \frac{4}{9} \\ \frac{8(2k(\sqrt{9-8k}-15)-33\sqrt{9-8k}+117)}{3(5-\sqrt{9-8k})^3(4k+\sqrt{9-8k}+7)} & \frac{4}{9} < k \leq 1 \end{cases} \quad (28)$$

□

Proposition A4 shows, as before, that quality and connectivity are strategic complements on the platform. Moreover, as the proof shows, the quality algorithm achieves the highest quality among the three algorithms, as expected. When k increases, sending becomes more expensive, which lowers the incentive to send. That process, however, leads to more receivers, which increases the incentive to send. In equilibrium, the availability of more receivers as k increases drives an increase in quality and connectivity.

Among all three cases, the perfect algorithm encourages the most sending in equilibrium. Because this algorithm filters the most content, it effectively lowers competition for likes, and increases the incentive to send content. This increased sending, however, means there are less potential receivers who can like a content item. The outcome of this tradeoff is that in equilibrium, the perfect algorithm yields low receiver utility (compared to the quality algorithm, that encourages higher quality), until k is quite high, in which case the qualities are all low and the perfect algorithm gives the receiver the highest utility through filtering.

H User Surplus

Although we do not explicitly model the profit goal of the platform in this paper, it is intuitive that increased user surplus may attract more receivers and senders to the social network in the

long run, resulting in higher profits for the platform.

To compare the surplus between the three different cases,⁴ we notice that when $\alpha = 1/2$, the surplus is proportional to the sum of the equilibrium expected utilities of one sender and one receiver:

$$US = \mathbb{E}[\pi_S^* + \pi_R^*] \quad (29)$$

Moreover, the results in the previous sections have shown that the different algorithms have different upper limits on the maximum cost of quality \bar{k} for which it is profitable for the senders to provide positive quality on the platform. Using these results we are able to compare equilibrium surplus when we focus on low values of r . The following proposition summarizes the result:

Proposition A5. *When $3/4 > r = c > 0$:*

- *User surplus of the quality algorithm is equal to the user surplus without an algorithm when $k < 1/2$. It is higher with a quality algorithm when $\frac{1}{2} < k \leq \min(\frac{7}{8r+4}, 1)$. The surplus without an algorithm is higher when $\frac{7}{8r+4} < k \leq 1$.*
- *When $0 < k < \min(\sqrt{\frac{1}{r^2+4r}}, 1)$, the surplus with the perfect algorithm is higher than the surplus without an algorithm. When $\sqrt{\frac{1}{r^2+4r}} \leq k < 1$, the surplus is higher without an algorithm than with the perfect algorithm.*
- *There is a range of r values and corresponding range of k values for which the quality algorithm generates the highest surplus.*
- *When $\frac{3}{8} < r < \frac{3}{4}$ and $\frac{7}{8r+4} < k < 1$, not using an algorithm generates the highest surplus.*

Proof. When $3/4 > r = c > 0$, $0 < k < 1$, and N is large, the surplus functions are:

$$US_{NA} = \begin{cases} \frac{3}{2} - 2k & 4k \leq 1/4 \\ \frac{1}{16k^2} & 1/4 < k < 1 \end{cases} \quad (30)$$

$$US_{PA} = \begin{cases} \frac{9k^3 - 4\sqrt{4k^2+1}k^2 + 2\sqrt{4k^2+1} - 2}{6k^3} & k \leq 2/3 \\ \frac{-6\sqrt{4k^2+1} + k(-15\sqrt{4k^2+1} + 31k + 15) + 6}{12k^2} & \frac{2}{3} \leq k < \min(\sqrt{\frac{1}{r^2+4r}}, 1) \\ \frac{1}{12}(1-r)^3 & \frac{1}{\sqrt{r(r+4)}} \leq k < 1 \end{cases} \quad (31)$$

⁴The distance algorithm case is equivalent to the no algorithm case and we omit it.

$$US_{QA} = \begin{cases} \frac{3}{2} - 2k & k \leq \frac{1}{4} \\ \frac{1}{16k^2} & \frac{1}{4} < k \leq \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} < k \leq \frac{7}{8r+4} \\ \frac{2}{27}(1-r)^3 & \frac{7}{8r+4} < k < 1 \end{cases} \quad (32)$$

The proof follows from a simple comparison of the functions. \square

The proposition shows that the quality algorithm in general is unable to generate more surplus than not using an algorithm at all, except for a range of intermediate values of k . The intuition behind this result stems from the fact that when k is low, the market is in full connectivity and the quality algorithm is unable to encourage more quality. When k is very high, however, it becomes unprofitable for the sender to increase quality above zero because of the filtering, resulting in lower user surplus. Only for intermediate values of k is the quality algorithm able to encourage higher quality that results in increased surplus.

Examining the perfect algorithm, we see a more nuanced result. For low sender cost levels, the perfect algorithm increases quality without hurting senders too much, resulting in a higher surplus. When k is high, on the other hand, the payoff of senders under the perfect algorithm is lower than not investing in quality at all, which lowers surplus below the NA case.

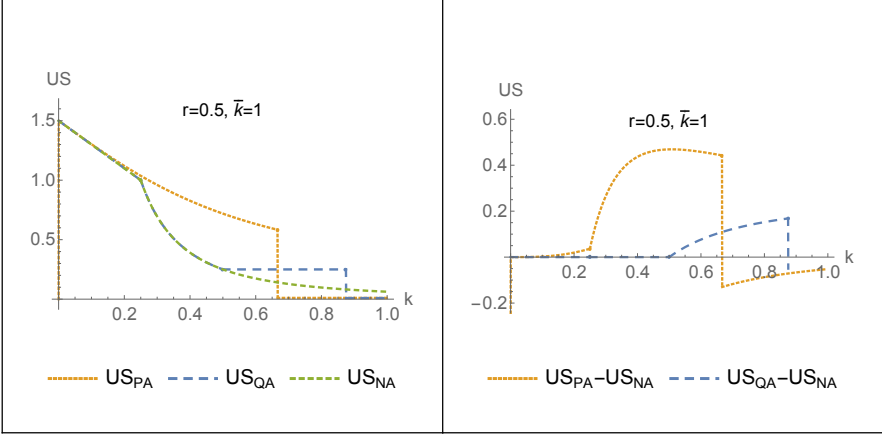
The last two points stem from the fact that the upper threshold of k for which a sender sets positive quality is lowest for PA, highest for NA and intermediate for QA. Consequently senders will provide lower quality with the PA when compared to the QA and NA above some k threshold, resulting in lower user surplus. For a wide range of intermediate r values, we see that the quality algorithm generates the most surplus, while for high values of r , the highest surplus is generated without an algorithm, because of this impact on the upper k threshold.

An important implication of this result is that using curation increases surplus only if combined with a strategy to reduce the cost of creating high quality content for platforms. This may explain, for example, why some platforms which have added novel content creation features (such as Instagram filters or Facebook's 360° photos), are able to maintain a large following and high levels of content creation despite applying algorithms that limit the exposure of content by receivers.

Figure A1 illustrates the three cases. It is clear that for most of the range of k the perfect

algorithm generates more surplus than the quality and the no algorithm cases, a result which is reversed for very high values of k . The positive difference between the quality and the no algorithm cases is also evident, but as can be seen, it is small, mostly because the effect of the increased quality by the quality algorithm comes into play for high values of k , where the difference created in the equilibrium qualities and connectivity is small.

Figure A1: User Surplus



Left: function values; right: difference from the no algorithm case.

$$r = c = 1/2.$$

I Popularity Filtering

We examine the case in which the platform cannot directly observe quality, but instead uses the popularity of a given post to infer it. This is potentially a more realistic representation of how filtering is operationalized as responses to content are often used as proxies for quality. There are many complicating factors when examining feedback-based systems, from the exploration-exploitation tradeoff to dynamic issues such as herding. We assume most of these away and examine popularity filtering in a mostly rigid and static setting to be able to compare with the rest of our results. For the sake of parsimony in this extension we only analyze the receiver actions and treat sender actions as exogenous.

To filter based on an item’s popularity, we assume the platform does not observe the quality or distance of an item from a receiver, but first observes the liking behavior of a large sample of

M receivers. The platform can make inference on the item's quality based on the liking behavior in those M instances. As each receiver consumes J messages and picks one to like at random, the share of "likes" will be between 0 and $1/J$, depending on the utility the item provides. The platform will then use these inferences for filtering. We specifically assume that no filtering is done by the platform until the scoring phase has been completed.

Since the platform does not observe the distance, it can only use the information from likes to infer quality. Observing a like tells the platform that the utility the user obtained was at least r . Upon observing a sufficiently large amount of likes with different unobserved distances, the platform infers the posterior distribution of quality from the percentage of observations with a like. If $1/J$ share of the observations show a like then quality is uniformly distributed between $r + 2\varphi$ and $q + 1$. If none of them have a like, then quality is distributed $\sim U[q, r]$. When a positive $0 < \rho < 1$ proportion of $1/J$ of the observations indicate a like then the quality is $r + 2\varphi\rho$. That is, when quality falls in the $[r, r + 2\varphi]$ interval the platform can perfectly infer quality given a sufficiently high number of observations.

The filtering rule for a given receiver is only dependent on ρ . When $\rho = 0$, the expected quality is $(q + r)/2$ as long as $r \geq q$ while the expected distance is φ as usual. Hence content is filtered out if and only if $r/2 - \varphi - c < 0$. When $\rho > 0$, the quality is perfectly inferred from ρ as $Q = r + 2\varphi\rho$ and content is filtered if and only if $r + \varphi(2\rho - 1) - c < 0$. As a consequence a content item is never filtered out if $\rho > 1/2$. Deriving the expected receiver utility from the above and maximizing, we obtain the following results:

Proposition A6. *If M is sufficiently high and $1/2 > c - q > -1/2$ then:*

$$\varphi_{POPA}^* = \begin{cases} \frac{q-c+1}{3} = \varphi_{QA}^* & \text{if } c - q \leq r - q \leq \frac{1+2(c-q)}{3} \\ \frac{q-c}{2} + \frac{1+r}{4} > \varphi_{QA}^* & \text{if } \frac{1+2(c-q)}{3} \leq r - q < 2(c - q) - \frac{1-\sqrt{5-8(c-q)}}{2} \\ \frac{q-c}{2} + \frac{1}{4} = \varphi_{NA}^* & \text{if } 2(c - q) - \frac{1-\sqrt{5-8(c-q)}}{2} \leq r - q \leq 1 \end{cases} \quad (33)$$

- Receiver utility is decreasing whereas the average distance is first increasing then decreasing in r .
- When $c - q \leq -1/2$ or $c - q \geq 1/2$, we always have $\varphi_{POPA}^* = \varphi_{NA}^*$.

Proof. We derive the expected utility of a consumer from a received message in three cases and multiply with φ to get the total from all messages. In case (i) when $\varphi \geq r - c + q$ the expected

utility is the same as in the quality algorithm case because all messages with quality below r are filtered out in both cases. Above r the filtering rule is the same. Therefore, the total expected utility is $\varphi(1-\varphi-c+q)^2/$ (as long as $q < \varphi+c < q+1$). In case (ii) when $r-c+q > \varphi \geq r/2-c+q$, the filtering threshold becomes r and the expected quality of an unfiltered message is $(r+q+1)/2$, therefore the total expected utility from messages is $\varphi(q+1-r)(r+q+1-2c-2\varphi)/2$ (as long as $q < r < q+1$). Finally, in case (iii) when $\varphi < r/2-c+q$, all messages are unfiltered and we calculate the expected utilities distinguishing between whether the quality is above or below r . Above r , we get the same formula as in case (ii). When quality is below r , the expected quality is $(r+q)/2$ and the total expected utility becomes $\varphi(r-q)(r+q-2c-2\varphi)/2 + \varphi(q+1-r)(r+q+1-2c-2\varphi)/2$. Overall the utility function can be pieced together from the above three cases with case (i) being cubic and cases (ii) and (iii) being quadratic each with a unique maximum. When $r \leq (1+2c+q)/3$ then case (i)'s maximum dominates. Above this threshold we need to compare the local maxima of cases (ii) and (iii). They are equal when $r = 2c - q - (1 - \sqrt{5 - 8c + 8q})/2$. Below this threshold case (ii)'s local maximum is higher, above, case (iii)'s is higher. Calculating the maximum values shows that they are decreasing in r . Finally, since the average distance of a received (and unfiltered) message is simply φ , we obtain the stated result. \square

The results indicate that the liking threshold impacts receiver utility and connectivity substantially, driven by the information content of a like. When $r = c$, the platform can essentially perfectly observe quality⁵ and the outcome is equivalent to that of the quality algorithm. As r increases the perfect inference breaks down. This does not immediately lead to a reduction in receiver utility as the average distance is high enough that low quality messages below r would be filtered out even if their quality were perfectly observed. Hence an imperfect observation does not have an impact.

However, as r reaches a higher threshold the utility will start to decrease. Surprisingly, this leads to an *increase* in connectivity. The reason is that imperfect quality observations make the platform *more conservative* for low utilities and the platform will filter out messages that it would not have had, had it observed their quality perfectly. This incentivizes receivers to reach out further and set a higher φ .

As r becomes even higher the positive feedback can no longer be sustained and the outcome

⁵For qualities above $r+2\varphi$ the platform only knows that quality is higher than this threshold, but these messages will never be filtered out

falls back to the level of the no algorithm case. Hence connectivity drops and receiver utility is reduced further.

In summary, we see that connectivity is non-monotonic in the liking threshold. First it matches the quality algorithm, then it further increases before dropping back to the no algorithm levels. While receiver utility is decreasing, the average distance of messages received is also non-monotonic and it is the highest in the intermediate range for a non fully informative quality signal.

Overall, the results in this extension tell us that popularity-based filtering can be an alternative to direct observation of quality information and may even have a side-effect of increasing connectivity above the levels of what we get in the quality algorithm. There are two important factors regarding the informativeness of consumer actions. One is that the platform has to observe a high enough number of consumer responses for each content item. In our model this can be ensured by having a high share of receivers, e.g., a low α . This ensures a sufficiently high number of observations without significantly impacting receiver utility due to the low overall percentage of messages in the scoring phase. The other one is that a consumer action has to be a good signal of quality. If, for example, the platform observes threshold-type signals akin to likes, the threshold has to be in the relevant range for filtering in order to provide the most information about expected consumer utilities.