NET Institute*

www.NETinst.org

Working Paper #16-08

September 2016

**The Impact of Curation Algorithms on
Social Network Content Quality and Structure**

Ron Berman

The Wharton School, University of Pennsylvania

Zsolt Katona

Haas School of Business, UC Berkeley

* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, <http://www.NETinst.org>, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.

The Impact of Curation Algorithms on Social Network Content Quality and Structure *

Ron Berman[†]

Zsolt Katona[‡]

Abstract

Curation algorithms are selection and ranking algorithms on social media that help consumers experience better content. These algorithms have been blamed in the past few years for the creation of “filter bubbles” and other phenomena in social media.

We analyze a platform with producers and consumers of content to understand the impact of curation algorithms on the amount of friends each consumer has and the quality of content created by each producer. Our model takes into account both vertical and horizontal differentiation and analyzes three different types of algorithms.

We find that without algorithmic curation, the number of friends a consumer has and the quality of content on the platform are strategic complements. When algorithmic curation is introduced, the resulting process makes consumers less selective in their choice of whom to follow. In equilibrium, producers of content receive lower payoffs because they enter into a prisoner’s dilemma like contest. The quality of content on the platform may increase if the marginal cost of producing this quality is high enough but not too high. Both of these effects may result theoretically in more diverse content consumed by consumers, but in equilibrium we find that a few of the algorithms may reduce the horizontal distance of matched content, which may result in a filter bubble. We identify an algorithm that focuses on filtering low quality items that results in higher quality of content as well as higher diversity under specific conditions.

Keywords: Social Media, Filtering, Ranking, Filter Bubble, Algorithmic Curation, Game Theory.

*We would like to thank the NET Institute (www.netinst.org) for its generous support of our project.

[†]Assistant Professor of Marketing, The Wharton School, email: ronber@wharton.upenn.edu

[‡]Associate Professor of Marketing, Haas School of Business, email: zskatona@berkeley.edu

1 Introduction

When social media platforms emerged, one of their appeals to consumers was that by telling the platform who their friends are, the online experience of users will not only be more social, but also that the content consumed and generated will be more tailor made and customized to fit the users' tastes. The main assumption was that by connecting to homophilous friends, consumers will be exposed to content which is more in their domain of interest as well as more personal and relevant to them. If in the past, for example, national newspapers did not delve into local matters while local newspapers did give a person content which is more related to her surroundings, social platforms were meant to be the ultimate source for relevant and personalized content.

A second appeal of social platforms was that users can also become creators and publishers of content, and share it easily with their friends or other interested parties. By having readers interact with this content, its creators could now gain utility (e.g through fame, reputation or just intrinsic enjoyment) from this process without having to compete with other creators through standard distribution channels.

The adoption of these platforms and the ease of use of their content creation technology, however, has led to an increase in the volume of content available for consumption online as well as the speed with which it is created and changed. Statistics published by leading platforms¹ indicate that over 48,000 photos were published on Instagram every minute in 2015, while Twitter users posted over 347,000 tweets per minute in the same time frame. This large volume of content has many times led to information overload on active users as well as to desensitization to new ideas (Rodriguez et al. 2014).

To help users manage this overload, social platforms often introduce algorithms whose goal is to select or rank the best content for users to consume. In the case of Facebook's newsfeed, for example, the platform selects stories that "are influenced by your connections and activity on Facebook. This helps you to see more stories that interest you from friends you interact with the most."² Recently Instagram introduced a ranking algorithm "so you'll see the moments you care about first."³ The problem the platform was trying to solve was that "On average, people miss 70

¹See, e.g. <http://www.cio.com/article/2915592/social-media/7-staggering-social-media-use-by-the-minute-stats.html>

²<https://www.facebook.com/help/327131014036297/>

³<http://blog.instagram.com/post/145322772067/160602-news>

percent of their feeds. It's become harder to keep up with all the photos and videos people share as Instagram has grown.”⁴

Although the introduction of these algorithms had the goal of increasing user activity and satisfaction, this was not always the case. In April 2016 it emerged that users of Facebook have been sharing less original personal content over time and opted to post and share more professionally produced content such as news and other information.⁵

A second issue often attributed to the adoption of filtering algorithms is the creation of “filter bubbles”,⁶ which is a state in which ideas and concepts match the consumer’s beliefs and are limited in their diversity. Early work by Van Alstyne and Brynjolfsson (2005) has shown that increased connectivity between online users can either fragment or better connect communities depending on the content and consumption preferences of the community members. As a result, “cyber-balkanization” may arise online. Recently, the issue has become so prevalent that The Onion, a satirical website, published an article with the sarcastic title “Horrible Facebook Algorithm Accident Results In Exposure To New Ideas”.⁷

In this paper we analyze the effects that selection and filtering algorithms have on social network structure, the quality of content produced by network members and the diversity of content consumed. This process, which we call *algorithmic curation* may have several possible effects on consumers of content and creators of content. On the consumption side, consumers need to decide who to follow (or friend) on social networks, and prior to the introduction of algorithmic curation, had to manually curate their friends lists taking into account the expected quality of content produced by these friends, as well as the topics they choose to post on. In addition, given the content created by their friends, consumers needed to decide how and how much of the content to consume, and consequentially which of these items merit interaction such as sharing the content or “liking” it. On the creation side, content producers needed to decide how much to invest in the quality of each content item, as well as could actively choose the topics to write on and the frequency of posting content items.

When algorithmic curation is introduced to a platform, it may have a profound impact on the

⁴Ibid.

⁵<http://www.bloomberg.com/news/articles/2016-04-07/facebook-said-to-face-decline-in-people-posting-personal-content>

⁶The term was coined in Pariser (2011).

⁷<http://www.theonion.com/article/horrible-facebook-algorithm-accident-results-expos-53841>

decisions of consumers and creators of content, which we call *receivers* and *senders*, respectively. For receivers, the number of friends to connect to and the number of items to read depends on their expectation about utility provided from posts by friends. If algorithmic curation is able to select better content that will be shared with receivers, they may become less selective in their choice of whom to friend, and as a result their friends lists will become longer. For senders, if algorithmic curation increases the number of friends a receiver has, an individual sender may now face stronger competition in receiving attention for her content. In addition, the sender may have a harder time passing the bar of quality required for her content to pass a filtering algorithm, and a harder time passing the bar of quality a receiver is seeking from content she interacts with. All in all, algorithmic curation may cause an increase or a decrease in the quality of content on the platform, which may result in more selective or less selective friending by receivers.

Our goal is to analyze these effects and to determine the impact of algorithmic curation on the equilibrium outcome of the platform. To achieve this goal we set up an analytical model with three players - a platform, receivers and senders. Receivers in our model receive utility when they consume costly content items that were produced by senders they follow. These items have two stochastic components to their utility which depend on the distance of a sender from a receiver in terms of their tastes, as well as the quality of the content item which may fluctuate between items produced by the same sender. To control and maximize the utility they receive from items, receivers can decide whom to friend as well as how many content items to read from the content available to them. We use a standard homophily assumption and assume that receivers prefer to connect with their closer friends first, and that they share a similar taste compared to more distant senders. This is similar to the idea of strong and weak ties that is commonly analyzed in the network literature.

Senders in our model receive utility when their content is liked by receivers, a process which depends on the utility they provide to the receiver, but also on the competition they face and on how selective receivers are when they interact with content. A sender can decide to increase the average quality of content it produces at some cost, an action which may give her an advantage compared to the other competing items produced by other senders. When receivers connect to more senders, the amount of potential likes a sender may receive increases, but on the other hand the competition for each of those likes increases as well.

Finally, we compare three types of algorithms a platform may enact to a benchmark case when no curation is in effect. The Perfect Algorithm (PA) selects for each receiver only the content which is above a utility threshold for that receiver. This algorithm is “perfect” in the sense that given a receiver-sender matched content item, the algorithm can observe the true utility the receiver will receive and can filter out the lowest utility generating items. The Quality Algorithm (QA) selects the content that passes a minimum quality bar, but can’t tell for each receiver if the item is close to her taste or far from it. This algorithm can be seen as the platform encouraging the creation of high quality items overall, and focusing less on the specific match to each consumer. The third algorithm, which we call the Distance Algorithm (DA), selects for each consumer the content which is close enough to her, but cannot tell whether each item is of high or low quality.

In our analysis, we focus on several aspects of the equilibrium including the quality of content created and the number of friends receivers connect to. We first consider the decision of the receiver to connect to friends and select content items to read given the average quality of items produced by senders as well as the effect of the algorithm on the utility of items the receiver has access to. Second, we look at the decision of a sender to set different quality levels given the chances her content will pass the different selection algorithms and will be liked by a receiver. Third, we analyze how much of the content is filtered in equilibrium in the different cases, and in a sense get a measure of how much content is “wasted” because of the algorithm. Fourth, we compare the average taste mismatch of content items consumed by receivers to understand the patterns of content diversity emerging from using these algorithms.

Our results show that in the benchmark case when algorithmic curation is not employed, the quality of items and the number of friends a receiver has are strategic complements. The quality and number of friends decrease when the cost of consuming content or producing quality increases, as expected. When a receiver sets a higher bar for giving a “like”, or when the receiver decides to put more weight on comparing a content item to the other items before giving a “like”, we find that friendships as well as the average quality of content increases. In that sense, by being more selective before giving “likes”, receivers can influence senders to increase the quality of their content, which leads to an increase in the number of friends receivers have. When the platform operates a curation algorithm, and in comparison to the benchmark case, we find that in general curation increases the number of friends a receiver is willing to connect to because the platform provides assurances

that the receiver will not experience too negative utility. Both the perfect algorithm and distance algorithms provide full assurance for the receiver to connect to all senders. In the case of the quality algorithm, the process provides a partial assurance that causes more connectivity by senders until the average quality produced is high enough, in which case receivers connect to all senders.

The equilibrium level of content quality created under all three curation algorithms is higher than the benchmark case when the cost of investing in quality is low enough. The quality increases in the selectivity of the receivers for giving likes as well as in the cost of content consumption. This is in contrast to the benchmark case where an increase in content consumption costs reduces the quality and connectivity on the platform. The reason for this is that the algorithms are able to use the consumption cost as an anchor for filtering, increasing the filtering threshold if the cost goes up. As a result consumers do not need to consume less items, which allows senders to keep increasing quality profitably. When the cost of producing quality is higher, however, the equilibrium quality level decreases. When comparing the three algorithms, we find that for low levels of quality creation costs, the two partial algorithms (distance and quality) lead to lower quality levels than the perfect algorithm due to a lower level of filtering. However, when the costs reach a certain threshold we find that both partial algorithms result in higher quality levels than the perfect algorithm. The reason is that when quality becomes expensive enough so that content quality reaches the filtering threshold the partial filtering algorithms provide a larger incentive to senders to remain above it.

We also find that content consumption costs interact with the algorithm and quality levels. In the absence of algorithmic filtering, higher costs lead to lower quality levels. This is driven by how senders internalize the content consumption cost. When readers have a high cost, they consume less content and also make fewer friends. But this leads to lower incentives to senders through depressing demand for content. On the other hand, filtering makes sure that receivers get content that they like, hence they do not hold back on connecting to senders. But as receivers' costs increase the filtering gets more selective, intensifying competition between senders thereby leading to higher quality levels.

The payoff of the senders decreases under all curation algorithms when compared to the no filtering case. This results from several effects. First, curation imposes a cost on the algorithms in the form of a bar they need to pass to receive a like. This increases the average cost for every item who receives a like because more content items are produced and do not receive likes. Second,

since curation increases the number of friends a receiver connects to, it creates more competition for senders which lowers their utility. Third, since in equilibrium a higher level of quality arises, senders need to spend more than compared to the no curation case.

We make several contributions to academic research as well as a few practical recommendations. On the analysis of possible causes for filter bubbles, we show that not all filtering algorithms are created equal. When platforms use algorithms that focus on consumers such as the distance algorithm and the perfect algorithms, filter bubbles may arise, and the average distance of content to a consumer is lower than the average distance when no filtering is used. This result is flipped with the quality algorithm which increases the average distance between receivers and the content they consume. The intuition behind this result is subtle. Perfect and distance algorithms focus on utilities received by consumers as well as the distance between two platform members. Since consumers in this case will increase the number of friends they have, and as a result the average distance of content sent to them, the platform will have to counteract this effect by imposing stronger filters in equilibrium, which ends up in lower distances overall experienced by the receivers. In contrast, a quality algorithm also provides enough assurance for receivers to connect to more friends, but the platform in this case cannot control the distance of content consumed resulting in more diverse content.

The implications of our research apply to a growing stream of practices by firms employing algorithms to improve user experience, sometimes resulting in surprising reduction in user satisfaction in the long run, or resulting in other phenomena such as dissatisfaction of advertisers because of lowered profits or increased polarization in content and communication patterns. Our second contribution is providing an explanation for these phenomena using a model that endogenizes the actions receivers and senders take before consuming and producing content in social media. Since network connectivity, the amount of filtering as well as the emerging quality may all change as a result of applying a filtering algorithm, our analysis gives predictions regarding the resulting equilibrium which in many cases leads to increased polarization despite increased connectivity. By focusing not only on the network structure but also on the horizontal and vertical aspects of content consumption, we are also able to show the effects of competition on content producers when different filtering algorithms are employed, and provide platforms with recommendations on which actions they can take to influence the creation of higher quality content or increased connectivity.

The remainder of the paper is structured as following: Section 2 overviews related work on social algorithms, recommendation systems and the impact of filtering and ranking algorithm. Section 3 introduces our three player model as well as the different types of algorithmic curation. Section 4 analyzes the receivers’ decision problem and resulting connectivity with exogenous senders, and Section 5 analyzes and endogenizes the senders’ choice and receivers “like” activities. Finally, Section 6 discusses limitations of our model and concludes. The proofs are relegated to the Appendix throughout the paper.

2 Related Literature

The literature analyzing the impact of selection and filtering algorithms has initially focused on how to design better algorithms, famous examples of which are Shardanand and Maes (1995) and Linden et al. (2003). Recently, more focus has been put on the economics and impact of such algorithms on consumers and the diversity of consumed products (See, *e.g.*, Latzer et al. (2016) for a survey).

One stream of this recent research has focused on the impact of recommendation systems on product consumption trying to determine whether the diversity and volume of products consumed increases after a recommendation system is introduced. Oestreicher-Singer and Sundararajan (2012) and Hosanagar et al. (2013), for example, find in an observational and experimental setting that the introduction of recommendation systems tends to increase the diversity of products bought by customers, as well as to increase their consumption volume.

A second stream of research focuses on the impact of news and media aggregators on the network structure and quality of content created online. In the works of Athey and Mobius (2012) and Chiou and Tucker (2015), the authors find in an empirical analysis that news aggregators serve as a complement to news websites, and allowing news aggregators to include excerpts of content from news sites benefits news sites in the long run as the number of their visitors increase, and in addition visitors are able to visit and explore more niche content. Dellarocas et al. (2013) and Roos et al. (2015) analyze the ability of websites to link directly and strategically to one another. They find that on one hand, websites that create content can benefit from aggregators helping visitors find their content, but on the other hand the same aggregators appropriate advertising profits as well as cause more competition for the content creator websites. In this sense, the findings are similar to

ours that a filtering platform that helps consumers find better content causes more linking within the network which results in stronger competition for content creators.

Recent work surveyed in Lazer (2015) looks to analyze empirically the impact of social algorithms. Specifically, research by Facebook published in Bakshy et al. (2015) has shown that algorithmic curation does limit somewhat the cross transfer of political ideas between the virtual isles on social media, but the filtering effect of the algorithm has a smaller impact than the preferences of consumers on whether to consume or not consume the content. Additional work by Nguyen et al. (2014) and Flaxman et al. (2016) found evidence for filter bubbles, while some of the work found evidence for a lowered average distance between the preferences of consumers, but also an increase in the dispersion of content consumed. None of this research has looked at the endogenous choice of consumers that determines their social media links or the resulting quality of content generated, and its relation to the impact on the distance to content consumed.

Su et al. (2016) resembles our study in that it focuses on analyzing the impact of a recommendation system on the network structure of a social media platform. In their analysis, the authors analyze Twitter’s “Who to follow” system that gives users on the platform suggestions on which other users to follow. They show, using empirical data and an analytical model, that popular users on the platform will “get richer” and will gain more followers, but that in general the entire network will have more connections created. Their analysis does not look at the strategic impact on the quality of content created on the network, nor the impact of different filtering algorithms and how they will impact the network structure. The complementary work in Iyer and Katona (2015) analyzes the incentives of social media members to become senders based on the structure of the network they face and the resulting amount of competition. The authors find that increased connectivity increases competition among senders, a finding which is similar to ours. This results in higher message intensity, but in equilibrium results in a lower amount of senders creating content. In our model we endogenize the network structure as well as analyze different possible algorithms for curation and their impact. Our findings on the impact on distance and quality of content are also an addition to the literature.

3 Model

We have three types of players in the market. The first one is the social networking platform who provides the infrastructure for consumers to befriend each other and share content. The second and third types are both consumers some of whom are senders who produce content and some of whom are receivers who consume content. There are N consumers and each of them can be a receiver or a sender. Let α denote the proportion of consumers who take on the role of a sender. Thus, we have αN senders and $(1 - \alpha)N$ receivers in the market.

3.1 Receivers

Receivers are identical, and are evenly distributed in the social network. Let R be a prototypical receiver. Receiver R consumes content items, where an item posted by sender i yields utility $U_i = Q_i - D_i$. The quality of the item Q_i is distributed uniformly in the $[q_i, q_i + 1]$ interval, where q_i is the baseline quality level set by sender i . The distance D_i between the sender i and receiver R is also a random variable and is distributed with pdf:

$$g_D(x) = \begin{cases} 2 - \varphi & 0 \leq x < \frac{1}{2} \\ \varphi & \frac{1}{2} \leq x < 1 \end{cases} \quad (1)$$

where φ measures the connectivity of a receiver with $0 < \varphi < 1$. Receivers set φ by deciding how many consumers to connect to, where the extreme case of $\varphi = 1$ means that the receiver is connected to all of the senders in the market. We assume that receivers and senders are uniformly distributed throughout the social network, therefore the number of senders the receiver is connected to is $\varphi \alpha N$.

This setup has two main features. First, the receiver does not exactly observe the distance from a given friend, but has an idea about the distribution of distances. This captures the fact that social distance is hard to observe, might also depend on a certain topic, and can fluctuate over time. Second, the expected distance becomes higher as an individual selects more friends. This is a natural assumption given that people tend to become friends with individuals that are similar to them - a phenomenon referred to as homophily. The specific distribution we use was selected to facilitate parsimony but is not required for the results to hold. The above formulation does have the nice feature of resembling the idea of strong and weak ties with fixed amount of possibilities

for strong ties.

When a receiver consumes content, she decides to read J content items, each item from one of J random friends, where $J < \varphi\alpha N$, the number of friends the receiver has among the senders. The cost of reading each item is c and the utility of content from a random friend J is $U_j = Q_j - D_j$ as defined before, with $j = 1 \dots J$.

Consequently, receiver R 's payoff is:

$$\pi_R = \sum_{i=1}^J (U_i - c) = \sum_{i=1}^J (U_i) - Jc \quad (2)$$

Receivers also express their liking of content which gives utilities to senders. When a piece of content consumed by a receiver is above a certain threshold, the receiver likes it. The details of this process are described in the section that details the utilities of the senders.

3.2 The platform and curation

Consumers cannot be sure about the utility they get from a piece of content until they consume it. Due to the variation in content quality and fit, they often decide to consume content that turns out to yield low or potentially negative utility ex-post. We model algorithmic curation as a process that filters content in order to maximize consumer payoff.

We assume that an algorithm assigns a score to each piece of content based on the sender, receiver, and the content quality. Let S_i denote this score variable. The algorithm filters all content below some threshold score t , where t is set by the social network platform. A perfect algorithm is able to exactly measure the payoff the receiver will derive from a piece of content in advance and simply assigns the utility as the score. In reality, platforms use algorithms that are imperfect. We will compare the following algorithms:

- The perfect algorithm (PA) observes all relevant information and can perfectly determine receiver utility, setting the score equal to the receiver utility, that is $S_i = U_i$.
- The quality algorithm (QA) does not observe social distance nor takes it into account. It observes quality perfectly and sets the score to the quality of a piece of content, that is $S_i = Q_i$.
- The distance algorithm (DA) ignores quality and sets the score equal to the distance between the sender of a specific content and the receiver. In this case, the algorithm filters above a

certain distance threshold. Thus, here we set $S_i = -D_i$, because $-D_i > t$ is equivalent to $D_i < -t$.

We assume that the objective of the platform is to maximize consumer utility. This is consistent with the idea of the platform profiting from advertising, thus aiming to maximize the time consumers spend on the site.

3.3 Senders

Senders derive utility from receivers liking their content. In order to receive more likes sender i can increase her content quality by setting a higher q_i , resulting in a higher random quality $Q_i \approx U[q_i, q_i + 1]$. Higher quality comes at a cost of kq_i , making the sender's utility

$$\pi_S^i = \#(likes) - kq_i \quad (3)$$

In order for a sender to receive a like from a receiver, she needs her item to be noticed by a receiver, as well as to be deemed good enough to merit receiving a like. We assume that a receiver picks one item from the J items she reads to decide whether it has passed a utility threshold that merits receiving a like, u_l . There are two forces that operate and may determine how the threshold utility is chosen by the receiver. On one hand, a receiver may decide to give a like to items which are good in an absolute sense and pass a fixed utility value we denote r , which we assume is higher than the cost of consuming a content item c . On the other hand, the receiver may decide to give an item a like only if it is relatively good compared to other content items. In this case the receiver will give the item a like only if the utility it generated is higher than the expected utility from a content item read by the receiver. In this sense, if other senders increase their quality, the receiver will expect higher utility items, and sender i will face competition and it will be harder for her to receive a like for an item.

We define $\beta \in [0, 1]$ as the weight the receiver puts on the relative threshold, which results in u_l , the minimum utility threshold for a like:

$$u_l = (1 - \beta)r + \beta \mathbb{E} \left[\frac{\sum_{j=1}^J U_j}{J} \right] \quad (4)$$

Summing up the likes from all receivers and adding the cost, we can write the sender's payoff:

$$\pi_S^i = \varphi(1 - \alpha)N \frac{\Pr(U_i > u_l \text{ and } S_i > t)}{J} - kq_i \quad (5)$$

3.4 Timing

The timing of the game is as follows. First, receivers pick φ to determine the number of friends they have and senders choose their quality levels q_i simultaneously. Second, the platform chooses the threshold t for their algorithm. Finally, receivers decide how many content items to read and like some of them, at which point payoffs are realized.

4 Analysis of Receivers and the Platform

We first analyze the receivers' decision making process and how the platform's curation algorithm affects these decisions. Therefore, in this section we assume that senders are passive and they all set a baseline quality of $q_i = q$.

4.1 Benchmark

We first analyze a benchmark case where the platform does not use any filtering algorithm. A prototypical receiver's expected payoff is

$$\mathbb{E}(\pi_R) = J \cdot (\mathbb{E}(U_j|\varphi) - c) \quad (6)$$

Note that the receiver cannot distinguish between senders and content utility before reading, hence the receiver either reads content from no friend or from all friends. Given that $\mathbb{E}(U_j|\varphi) = q + \frac{1-\varphi}{4}$, the receiver will not read any content if $c - q > \frac{1-\varphi}{4}$, resulting in $J = 0$. Otherwise, the receiver will set $J = \varphi\alpha N$ and read content from all friends.

In the first stage, when the receiver decides on φ , the receiver solves:

$$\max_{\varphi} \mathbb{E}(\pi_R) = \max_{\varphi} \varphi\alpha N \left(q - c + \frac{1-\varphi}{4} \right) \quad (7)$$

The interior solution is $\varphi_{NA}^* = \frac{1+4q-4c}{2}$ which yields that if $c - q > \frac{1}{4}$ receivers will not connect to anyone and set $\varphi_{NA}^* = 0$. When $c - q < -\frac{1}{4}$, receivers will connect to everyone and set $\varphi_{NA}^* = 1$.

Proposition 1. *The number of friends receivers choose is increasing in the content quality, but is decreasing in the cost of content consumption.*

A particular example that we will use for comparison between the cases is that of $c = q$, when we get $\varphi_{NA}^* = 1/2$.

4.2 Perfect Algorithm

We now analyze the receiver's decisions under the perfect algorithm (PA) and the decision of the platform regarding the filtering threshold. Starting at the last stage, recall that the platform intends to maximize receiver payoff. With a given t threshold, the payoff is

$$\mathbb{E}(\pi_R) = J \cdot \Pr(U_j > t) \mathbb{E}(U_j - c | \varphi, U_j > t) \quad (8)$$

As in the benchmark case, the receiver cannot distinguish between different content pieces before reading so she either reads none or all. Thus the platform simply maximizes $\Pr(U_j > t) \mathbb{E}(U_j - c | \varphi, U_j > t)$, which reaches its maximum at $t^* = c$. The intuition is fairly straightforward: given that the platform can perfectly observe the utility that the receiver will get, it filters out all content that would decrease the receiver's total payoff. Solving for the receiver's optimal friendship choice, we derive the following proposition.

Proposition 2. *Under the perfect algorithm, consumers set $\varphi_{PA}^* = 1$, friending all senders in the network. The proportion of content filtered out by the algorithm increases in content consumption cost and decreases in content quality.*

The proposition tells us that the algorithm increases consumer utility by filtering out content that could have a negative impact on total consumer utilities. Interestingly, the algorithm also incentivizes receivers to connect to more friends than without an algorithm. In particular, consumers now have an incentive to connect to all senders. The reason is that consumers can now afford to risk getting content from friends who are at a larger social distance, because the algorithm offers insurance against mismatching content.

To illustrate our results better, we present the solution in the case of $c = q$ (the solution of the general case is presented in the Appendix). For $c = q$, we get

$$\Pr(U_j > c) \mathbb{E}(U_j - c | \varphi, U_j > c) = \frac{7 - 3\varphi}{24} \quad (9)$$

Thus, in this case, the receiver maximizes

$$\max_{\varphi} \mathbb{E}(u_R) = \max_{\varphi} \varphi \alpha N \frac{7 - 3\varphi}{24} \quad (10)$$

yielding a corner solution of $\varphi_{PA}^* = 1$ which is clearly higher than the benchmark case of $\varphi_{NA}^* = 1/2$.

Another good way to measure the algorithm's behavior and compare the different cases is to calculate the level of filtering, which is the probability that a piece of content from a random friend is hidden from the receiver. Naturally, with no filtering, this probability is 0. In case of the perfect algorithm, we get $p_{PA}^* = \Pr(U_j > c) = \frac{1+\varphi_{PA}^*}{4} = \frac{1}{2}$ in the focal case of $c = q$.

4.3 Quality Algorithm

In case of the quality algorithm, the platform maximizes

$$\mathbb{E}(\pi_R) = J \cdot \Pr(Q_j > t) \mathbb{E}(U_j - c | \varphi, Q_j > t) \quad (11)$$

As we derive in the proof, $\Pr(Q_j > t) \mathbb{E}(U_j - c | \varphi, Q_j > t)$ reaches its maximum at $t^* = c + \frac{1+\varphi}{4}$. As long as $q < t^* < q + 1$, we obtain

$$\Pr(Q_j > t^*) \mathbb{E}(U_j - c | \varphi, Q_j > t^*) = \frac{(4q - 4c + 3 - \varphi)^2}{32} \quad (12)$$

Solving the receiver's maximization problem, we get

$$\varphi_{QA}^* = \frac{4q - 4c + 3}{3}, \quad p_{QA}^* = \frac{1}{2} + \frac{2c - 2q}{3} \quad (13)$$

for $0 \leq c - q \leq 3/4$, yielding the following result.

Proposition 3. *Under the quality algorithm, consumers connect to all friends if $c \leq q$. When $c > q$, consumers connect to fewer friends than under the perfect algorithm, but to more friends than without a filtering algorithm. Consumer utility and the level of filtering are lower than under the perfect algorithm.*

The proposition demonstrates that the quality algorithm is a less effective solution than the perfect algorithm. It results in a lower level of filtering and results in a lower consumer utility than the perfect algorithm. However, the quality algorithm does not always perform strictly worse. In case of $c \leq q$ the receiver sets $\varphi_{QA}^* = 1$ with a level of filtering of $p_{QA}^* = 1/2$. For $c > q$, the number of friends consumers connect to starts to decrease and reaches 0 at $c = q + 3/4$. The level of filtering also increases as the cost grows, offering a more selective experience to consumers, but due to the imperfect filtering it cannot keep up with the reduction in expected utility, thereby decreasing consumer incentives to connect to friends.

4.4 Distance Algorithm

In case of the distance algorithm, the platform maximizes

$$\mathbb{E}(\pi_R) = J \cdot \Pr(-D_j > t) \mathbb{E}(U_j - c | \varphi, -D_j > t) = J \cdot \Pr(D_j < -t) \mathbb{E}(U_j - c | \varphi, D_j < -t) \quad (14)$$

As we derive in the proof, $\Pr(D_j < -t) \mathbb{E}(U_j - c | \varphi, D_j < -t)$ reaches its maximum at $t^* = c - q - \frac{1}{2}$. Solving the receivers maximization problem, we get the following result.

Proposition 4. *Under the distance algorithm, consumers set $\varphi_{DA}^* = 1$, connecting to all senders in the network. The proportion of content filtered out by the algorithm is higher than the proportion filtered out by the perfect algorithm iff $c > q$.*

Interestingly, despite its imperfection, the distance algorithm gives enough insurance to senders so that they connect to all senders. Since content from senders at a large distance is filtered out, the risk of connecting to everybody is not high enough to discourage receivers from connecting to many senders, some of them at potentially high distances.

4.5 Average distance

By employing curation algorithms social networks found themselves in a controversy regarding the diversity of content each consumer reads. In our model, social distance captures the difference in opinions between consumers, hence we can calculate the average distance of content that receivers read. Let $\bar{d}$ denote the expected value of a random piece of content that a receiver encounters in equilibrium. We obtain the following results.

Corollary 1. *There exists a $\hat{\Delta} \approx 0.35$ such that if $\frac{1}{2} > q - c > \hat{\Delta}$ then $\bar{d}_{QA} \geq \bar{d}_{NA} > \bar{d}_{DA} > \bar{d}_{PA}$. If $\hat{\Delta} > q - c > -\frac{1}{4}$ then $\bar{d}_{QA} \geq \bar{d}_{NA} > \bar{d}_{PA} > \bar{d}_{DA}$.*

The results clearly show the different ways the three algorithms affect content available to readers. An immediate implication of the corollary is that under both the perfect algorithm and distance algorithm, the average distance of content experienced by receivers is smaller than when no curation is used for any quality level q . Even though both of these algorithms effectively increase connectivity to the maximum, the fact that they filter content yields in equilibrium lower exposure to diverse ideas by receivers. The intuition is that when q is fixed, increasing the number of friends shifts the distribution of content facing receivers farther towards higher distances, and as a result

the curation algorithms effectively focus on reducing this distance to counteract the effect which results in lower diversity of ideas. Although it is quite intuitive that focusing on decreasing the distance to consumers will lower the average distance in equilibrium, it is interesting that focusing on improving the user's experience and utility also has a similar effect.

The quality algorithm, on the other hand, improves the diversity of content experienced by users compared to having no curation. The result is counterintuitive - when quality q is fixed and the algorithm cannot observe distance, filtering on quality will only affect the average distance through the amount of connectivity in the network. Since the quality algorithm can achieve increased connectivity when q is higher than c , the quality algorithm has a potential to increase the average distance of content items read by consumers and thus increase the diversity of content being consumed.

5 Analysis of Senders

In this section we analyze the senders' behavior. The decision they make is the quality level that they set to shift the distribution of message quality available to receivers. As senders are identical we are looking for a symmetric equilibrium where each sender sets the same level of quality. Throughout the section, we assume that $\alpha = 1/2$ for the sake of parsimony, implying that the number of senders equals the number of receivers. We start with the benchmark case of no curation.

5.1 No curation

When examining sender i 's decision, we denote the quality level set by her by q_i , while all other senders set q_{-i} . Let $G(\cdot)$ denote the CDF of the utility distribution $U_i = Q_i - D_i$ when $q_i = 0$ and let $g(\cdot)$ denote its PDF.⁸ In the absence of curation, sender i maximizes her expected payoff as follows:

$$\max_{q_i} \mathbb{E} \pi_S^i = \max_{q_i} \frac{\varphi N}{2} \cdot \frac{1 - G(u_l - q_i)}{J} - k q_i. \quad (15)$$

The first order condition then becomes:

$$-g(u_l - q_i) \left(\frac{\partial u_l}{\partial q_i} - 1 \right) = k \quad (16)$$

⁸Exact formulas for the CDF and PDF are given at the beginning of the Appendix.

Using the fact that $\frac{\partial u_l}{\partial q} = \beta/J$, and that $u_l = r(1 - \beta) + \beta q^* + \beta \frac{1-\varphi}{4}$ when all senders set q^* , we obtain, that in equilibrium

$$g\left((r - q^*)(1 - \beta) + \beta \frac{1-\varphi}{4}\right) = k / \left(1 - \frac{2\beta}{\varphi N}\right) = \frac{kN\varphi}{N\varphi - 2\beta} \quad (17)$$

Since the second order conditions of the maximization problem hold only on the increasing parts of $g(\cdot)$, we find the best response quality of the sender, q^* , to the share of friends each receiver has φ :

$$q^*(\varphi) = \begin{cases} r + \frac{\beta}{1-\beta} \frac{1-\varphi}{4} + \frac{1}{1-\beta} - \frac{kN}{(1-\beta)(N\varphi-2\beta)} & 0 \leq k \leq \frac{\varphi}{2} - \frac{\beta}{N} \\ r + \frac{\beta}{1-\beta} \frac{1-\varphi}{4} + \frac{1}{(1-\beta)(2-\varphi)} - \frac{kN}{(1-\beta)(N\varphi-2\beta)} \frac{\varphi}{2-\varphi} & \frac{\varphi}{2} - \frac{\beta}{N} \leq k \end{cases} \quad (18)$$

We assume $k < \bar{k} \leq 1$ to make sure that the local maximum obtained above is also a global maximum.⁹

It is apparent that the best-response quality increases in r , the threshold set by the receiver, and it naturally decreases in k , the cost of quality. To avoid encountering high degree polynomials in the next steps of the analysis we study the case of large N values, in the rest of this section by assuming that $N \rightarrow \infty$. We then obtain

$$q^*(\varphi) \xrightarrow{N \rightarrow \infty} \begin{cases} r + \frac{\beta}{1-\beta} \frac{1-\varphi}{4} + \frac{1}{1-\beta} - \frac{k}{(1-\beta)\varphi} & 0 \leq k \leq \frac{\varphi}{2} \\ r + \frac{\beta}{1-\beta} \frac{1-\varphi}{4} + \frac{1-k}{(1-\beta)(2-\varphi)} & \frac{\varphi}{2} \leq k \leq \bar{k} \end{cases} \quad (19)$$

To derive the equilibrium, we use the formula from the previous section, where we determined that the receiver sets $\varphi = \frac{1+4q-4c}{2}$. Substituting this into (19), we get a quadratic equation for φ with either a corner solution or a single solution:

$$\varphi^* = \begin{cases} 1 & 0 \leq k \leq \frac{3+\beta}{4} + (1-\beta)(r-c) \\ \frac{4(r-c)(1-\beta)+5-2\beta-\sqrt{D}}{4-2\beta} & \frac{3+\beta}{4} + (1-\beta)(r-c) \leq k \leq \bar{k} \end{cases} \quad (20)$$

where $D = 8(1-\beta)(r-c)(2(1-\beta)(r-c) - 3 + 2\beta) + 4\beta(1+\beta) + 16(2-\beta)k - 23$

As it is apparent from the above formula, the strategic interaction between quality and connectivity drives up the number of friends receivers pick to the maximum level if costs are below a certain threshold. As it becomes costlier to produce quality messages the connectivity declines. Comparative statics analysis reveals the following:

⁹The formula for $\bar{k} \leq 1$ is given in the Appendix.

Proposition 5. *Both q^* and φ^* increase in r , β and decrease in c and k .*

Proposition 5 shows that increasing the minimum utility r required for a like will increase the number of friends a receiver chooses to connect to, as well as will increase the quality of content produced by the senders. In this sense, being more selective helps the receivers and improves the quality of content on the platform. This results in the receivers being less careful about selecting their friends, since giving likes more carefully helps align the senders to produce higher quality content. A similar effect happens when increasing β , in which case the receiver puts a stronger pressure on senders to beat the expected quality delivered by all senders, which causes them to increase their quality of content produced.

When the cost of consuming content c is increasing, the receiver decreases the amount of friends it connects to, because this will yield a lower distance from the senders on average, and will give a higher probability of consuming content whose utility passes the minimal cost facing the receiver. As a result the sender will face less competition in the market, and the quality will decrease. Finally, when k increases, senders face a higher cost of producing quality and will lower the quality they choose to provide.

5.2 Perfect Algorithm

As we have shown in Section 4, the perfect algorithm filters out any messages that yield negative payoff to the sender. Hence if the difference of the realized quality and the realized distance is less than the cost of consuming a message, the receiver does not see or read the message. Each player is thus potentially being filtered out with a probability depending on the quality level chosen by the sender. The number of messages the receiver chooses from changes depending on the realized qualities, and distances yielding an a priori random number of messages to choose from. The probability that a given player who sets quality q is filtered out is $G(c - q)$. We thus obtain the payoff for a given sender i , when everybody else sets q_{-i} as follows.

$$\max_{q_i} \frac{\varphi N}{2} \cdot (1 - G(u_l - q_i)) \sum_{L=0}^{J-1} \left[\frac{1}{L+1} \binom{J-1}{L} (1 - G(c - q_{-i}))^L G(c - q_{-i})^{J-1-L} \right] - k q_i. \quad (21)$$

The first order condition then becomes:

$$-\frac{\varphi N}{2} \cdot g(u_l - q_i) \left(\frac{\partial u_l}{\partial q_i} - 1 \right) \sum_{L=0}^{J-1} \left[\frac{1}{L+1} \binom{J-1}{L} (1 - G(c - q_{-i}))^L G(c - q_{-i})^{J-1-L} \right] = k. \quad (22)$$

As before, we assume that $N \rightarrow \infty$ to obtain a tractable formula, resulting in

$$\frac{g(u_l - q_i)}{(1 - G(c - q_{-i}))} = k. \quad (23)$$

Comparing to the benchmark case, it is apparent that senders have a higher incentive to invest in quality for a fixed utility threshold, u_l . This is driven by the reduced competition due to filtering. On the other hand, the utility threshold u_l potentially increases as low quality messages are filtered out, resulting in receivers becoming more selective. To fully understand the combination of these forces, we determine the equilibrium quality level. As we have shown in Section 4, receivers always set $\varphi^* = 1$, regardless of quality. Substituting $\varphi = 1$ and $q_i = q_{-i} = q^*$ into (23), we obtain

$$g(u_l - q^*) = k(1 - G(c - q^*)). \quad (24)$$

Calculating u_l yields $u_l = (1 - \beta)r + \beta \left(\frac{1+2c-2q^*}{3} + q^* \right)$ when $c > q^*$ and $u_l = (1 - \beta)r + \beta \left(\frac{2(c-q^*)^3 + 3(c-q^*)^2 - 1}{3(c-q^*)^2 + 6(c-q^*) - 3} + q^* \right)$ otherwise. Given that we assumed small c values, we have to use the latter formula for our equilibrium solution. Furthermore, this also yields that we will use $G(x) = x^2/2 + x + 1/2$ on the right-hand side of (24). Hence the equilibrium quality is obtained as the solution of

$$1 + (1 - \beta)(r - q^*) + \beta \frac{2(c - q^*)^3 + 3(c - q^*)^2 - 1}{3(c - q^*)^2 + 6(c - q^*) - 3} = k \frac{1 - 2(c - q^*) - (c - q^*)^2}{2} \quad (25)$$

The above equation is a quartic in q^* , leading to an excessively complex closed-form solution. Thus, we resort to analyzing special cases and conducting comparative statics using the implicit function theorem. We obtain the following results:

Proposition 6.

1. q^* increases in r , β , c and decreases in k .
2. Senders set a higher quality and obtain a lower payoff under the perfect algorithm than without filtering.

The first part of the proposition shows that r , β and k affect the amount of effort spent on quality in a similar way to the case of no curation. More demanding readers generally entice senders to produce higher quality as long as the cost is not high.

Interestingly, the first part also reveals that increasing receiver's cost has the opposite effect. This parameter is an important component of our model as the basic idea behind curation is to

prevent readers from spending valuable time on content yielding low utility. As our results show, curation has a double positive effect on receivers. In addition to the baseline effect of filtering our low quality content, it also incentivizes senders to create higher quality content. The higher the cost the higher the pressure on senders, since the probability of being filtered out increases with c . Therefore, senders have to spend more on quality to avoid being hidden from receivers.

The second part of our results establishes a comparison between the case of no filtering and perfect filtering. For similar reasons as above filtering increases quality efforts via the threat of filtering out senders. This threat makes them increase quality over and above what they would set when the only incentive is to pass a reader's like threshold. The result also shows a potentially unintended consequence: together with the increased effort comes a reduced payoff for senders. While a higher quality results in more likes in total, the effort is not worth it. But interestingly, filtering makes it more appealing for senders to spend more on quality because they expect fewer competitors for a given receiver due to curation. That is, senders are not driven to spend extra by the direct effect of filtering, but the indirect, competitive effect akin to a prisoner's dilemma.

5.3 Quality Algorithm

As the quality algorithm filters out any items with quality less than t^* , senders can increase their base quality q_i to make sure their items are not being filtered by the platform. Using the result from Section 4, we notice that receivers set $\varphi = 1$ when $q \geq c$. We will therefore assume $q^* \geq c$ in equilibrium and verify this result holds following the analysis.

Suppose the platform sets $t^* = c + \frac{1+\varphi}{4} = c + \frac{1}{2}$ as the threshold for filtering quality, then the probability that an item from any senders passes filtering is $Pr(Q > t^*) = \min(q + 1 - t^*, 1)$. The expected utility received from such items is:

$$\begin{aligned} \mathbb{E}[U_i|Q > t^*] &= \mathbb{E}[Q - D|Q > t^*] = \mathbb{E}[Q|Q > t^*] - \mathbb{E}[D] = \\ &= \frac{q + 1 + t^*}{2} - \frac{1 + \varphi}{4} = \frac{2q + 2c + 1}{4} \implies u_l = (1 - \beta)r + \beta \frac{2q + 2c + 1}{4} \end{aligned} \quad (26)$$

In order to receive a like, sender i 's item needs to pass the quality filter as well as the utility threshold u_l set by the receiver. The probability of item i not being filtered for quality and being

eligible for receiving a like when $t^* > u_l$ is:

$$P(q_i) = Pr(Q_i > t^*) - Pr(U_i < u_l \text{ and } Q_i > t^*) = \begin{cases} (q_i + 1 - t^*)(1 - \frac{2u_l + 1 - t - q_i}{2}) & q_i < u_l \\ (q_i + 1 - t^*) - \frac{(u_l + 1 - t)^2}{2} & u_l \leq q_i < t^* \\ 1 - \frac{(u_l + 1 - q_i)^2}{2} & q_i \geq t^* \end{cases} \quad (27)$$

The sender maximizes:

$$\max_{q_i} \frac{P(q_i)}{\min(q_{-i} + 1 - t^*, 1)} - kq_i \quad (28)$$

Given $P(q_i)$, for low values of $q_i < u_l$, the profit of the sender may first decrease and then increase. It then reaches a part linear in q_i when $t^* \geq q_i \geq u_l$, and finally achieves a local maximum when $q_i \geq t^*$. When k is not too high¹⁰, the profit of the senders is positive and increases in q^* for $q^* \geq t^*$. In addition, the profit of sender i increases in q_i when $u_l \leq q_i < t^*$. The local maximum will therefore become a global maximum for low values of k , and will be in the range of $q^* \geq t^*$, which means the maximization problem reduces to solving $q^* = 1 - k + u_l$. Plugging in u_l and solving for q^* we receive:

$$q^* = \max \left(r + \frac{\beta + 2\beta c - 4k + 4}{4 - 2\beta}, c + \frac{1}{2} \right) \quad (29)$$

Analyzing the equilibrium quality level we find that:

Proposition 7.

- q^* increases in r , c and decreases in k . q^* increases in β for low values of k and r .
- Senders set a higher quality and obtain a lower payoff under the quality algorithm than without filtering if and only if $k > \beta + (1 - \beta)(1/2 + r - c)$.

The first part of the proposition finds similar effects for r , c , β and k as in the perfect algorithm case. The second part shows that quality filtering results in higher equilibrium quality levels for low quality cost k , but when the cost of quality increases, the quality is lower than under the case of no curation. This implies that for low values of k senders set a quality level that effectively makes their content always pass the quality filter. The intuition is that the quality filter puts a

¹⁰ $k < \frac{1}{4}(2 - \beta)(\beta + (12c + 2r)(1 - \beta))$

sharp bound below which the senders face no competition, and as such, moving from setting a quality level below t^* to above t^* has a larger increase compared to the case of no curation. This results in a bigger incentive for each sender to pass the threshold as compared to before. When the cost of quality becomes high enough, however, this tradeoff reverses and the benefit of passing the threshold is no longer viable. As a result the senders will set their quality level to the threshold, while in the NA case the quality is set higher.

5.4 Distance Algorithm

As with the perfect algorithm, we take advantage of the result derived in Section 4 that receivers set $\varphi^* = 1$ regardless of quality. In case of the distance algorithm, the probability of not being filtered out by the platform when setting quality q is simply $Pr(-D > c - q - 1/2) = Pr(D < 1/2 - c + q) = 1/2 - c + q$ when $c - 1/2 \leq q \leq c + 1/2$. Furthermore,

$$\begin{aligned} \mathbb{E}[U_i | D < 1/2 - c + q] &= \mathbb{E}[Q - D | D < 1/2 - c + q] = \mathbb{E}[Q] - \mathbb{E}[D | D < 1/2 - c + q] = \\ &\left(q + \frac{1}{2}\right) - \min\left(1, \frac{1}{2} - c + q\right) = \max\left(c, q - \frac{1}{2}\right) \implies u_l = (1 - \beta)r + \beta \max\left(c, q - \frac{1}{2}\right) \end{aligned} \quad (30)$$

We also need to determine the probability of the joint event of passing the u_l threshold and not being filtered out. This can be calculated as $P(q) = Pr(D < 1/2 - c + q) - Pr(D < 1/2 - c + q \cup Q - D < u_l) = 1/2 - c + q - \frac{(1/2 - c + u_l)^2}{2}$ when $q \leq c + 1/2$ and $P(q) = 1/2 - (u_l - q) - (u_l - q)^2/2$ otherwise. As before, the utility function of sender i is

$$\max_{q_i} \frac{P(q_i)}{\min(1/2 - c + q_{-i}, 1)} - kq_i. \quad (31)$$

Since the payoff function is linear in q_i for $q_i \leq c + 1/2$ we get that senders set either $q_i = 0$ or $q_i \geq c + 1/2$ depending on whether $k > \frac{1}{1/2 - c + q_{-i}}$ or not. Thus, when k is not too high the solution will be in the range where $q_i > c + 1/2$, yielding that in equilibrium $1 + u_l - q^* = k$. Substituting u_l yields

$$1 + (1 - \beta)(r - q^*) - \frac{\beta}{2} = k. \quad (32)$$

When k is too high for such an interior solution to be maintained, we get a corner solution with $q_i = c + 1/2$. Thus, the equilibrium quality level is

$$q^* = \max\left(r + \frac{2 - 2k - \beta}{2 - 2\beta}, c + \frac{1}{2}\right). \quad (33)$$

Analyzing this solution we get the following results.

Proposition 8.

1. q^* increases in r , β , c and decreases in k .
2. Senders set a higher quality and obtain a lower payoff under the distance algorithm than without filtering if and only if $k > 1 - (1 - \beta)(1/2 + c - r)$.

The comparative statics in the first part show a similar pattern as all the other curation algorithms. The second part reveals a similarly intriguing pattern as in the case of quality filtering. When costs are low, the distance algorithm results in the same or lower quality than no curation (strictly lower for $\beta > 0$). But when costs are somewhat higher, quality levels reach a generally lower level where filtering becomes a harsh penalty for reducing quality. Therefore, senders will set a high enough quality level so that the distance filtering threshold becomes so high that they never get filtered out.

6 Discussion and Conclusion

Through using our model that includes creators of content (senders), consumers of content (receivers) and an intermediating platform, our analysis has shown that algorithmic curation can alter the structure of the network as well as the quality and diversity of content on a social network.

The analysis of the two sided market required a few simplifying assumptions that facilitated both parsimony of exposition as well as the ability to perform more specific analysis of different aspects of the resulting content distribution. Specifically, there are several other possibilities to describe the receiver consumption decision and “like” process, including more sophisticated search or satisficing processes which are left for future work. Another assumption we have made is that the market has homogeneous receiver and sender marginal costs. A promising avenue for future work would be to examine how senders and receivers with heterogeneous costs are affected by the different algorithms.

When platforms use algorithmic curation, we have shown that receivers will become less selective in choosing whom to follow, while senders in general will have to increase the quality of their content when the cost of producing quality is high, though they will decrease it when the cost is low. Figure 1 illustrates how the different parameters impact the resulting equilibrium quality of content on the platform. The top plots illustrate that when the receiver shifts weight to the expected content

quality in order to give a like, the platform's quality increases. The bottom plots illustrate how the equilibrium quality increases in c , the receiver's cost of consuming content, but this increase is more pronounced under the distance algorithm.

Figure 1: Equilibrium quality q^* as a function of the marginal cost of quality k . Each figure details the parameters used to plot the results.

Furthermore, we find that a few algorithms (such as the PA and DA) may increase the distance of matched content between receivers and senders and create filter bubbles, while other algorithms (such as the QA) may decrease this distance. As such, if a platform wishes to avoid the emergence of such bubbles, one solution would be to focus on quality filtering and recommendation algorithms, and less on taste matching algorithms.

This non-monotonic impact of the different algorithms may explain why in some cases it is observed that introducing algorithms encourages higher quality items on networks and more pluralistic content consumed by different receivers, while in other cases the effects are the opposite.

An interesting implication of our finding is that by making it easier for senders to create high quality content, platforms can alter the incentives of receivers to connect to senders as well as the resulting equilibrium diversity of content. One way platforms can approach this problem is by supplying tools that allow producers easier creation and editing of content.

References

- Susan Athey and Markus Mobius. The impact of news aggregators on internet news consumption: The case of localization. 2012.
- Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.
- Lesley Chiou and Catherine Tucker. Content aggregation by platforms: The case of the news media. *NBER Working Paper*, (w21404), 2015.
- Chrysanthos Dellarocas, Zsolt Katona, and William Rand. Media, aggregators, and the link economy: Strategic hyperlink formation in content networks. *Management Science*, 59(10):2360–2379, 2013.
- Seth Flaxman, Sharad Goel, and Justin Rao. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, page nfw006, 2016.
- Kartik Hosanagar, Daniel Fleder, Dokyun Lee, and Andreas Buja. Will the global village fracture into tribes? recommender systems and their effects on consumer fragmentation. *Management Science*, 60(4):805–823, 2013.
- Ganesh Iyer and Zsolt Katona. Competing for attention in social communication markets. *Management Science*, 2015.
- Michael Latzer, Katharina Hollnbuchner, Natascha Just, and Florian Saurwein. The economics of algorithmic selection on the internet. In *Handbook on the Economics of the Internet*, chapter 19,

- pages 395–425. Edward Elgar Publishing, 2016. URL http://EconPapers.repec.org/RePEc:elg:eechap:14700_19.
- David Lazer. The rise of the social algorithm. *Science*, 348(6239):1090–1091, 2015.
- Greg Linden, Brent Smith, and Jeremy York. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80, 2003.
- Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*, pages 677–686. ACM, 2014.
- Gal Oestreicher-Singer and Arun Sundararajan. Recommendation networks and the long tail of electronic commerce. *MIS Quarterly*, 36(1):65–84, 2012.
- Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.
- Manuel Gomez Rodriguez, Krishna Gummadi, and Bernhard Schoelkopf. Quantifying information overload in social media and its impact on social contagions. In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.
- Jason MT Roos, Carl F Mela, and Ron Shachar. The effect of links and excerpts on internet news consumption. *Available at SSRN 2678938*, 2015.
- Upendra Shardanand and Pattie Maes. Social information filtering: algorithms for automating “word of mouth”. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210–217. ACM Press/Addison-Wesley Publishing Co., 1995.
- Jessica Su, Aneesh Sharma, and Sharad Goel. The effect of recommendations on network structure. In *Proceedings of the 25th International Conference on World Wide Web*, pages 1157–1167. International World Wide Web Conferences Steering Committee, 2016.
- Marshall Van Alstyne and Erik Brynjolfsson. Global village or cyber-balkans? modeling and measuring the integration of electronic communities. *Management Science*, 51(6):851–868, 2005.

Appendix

PRELIMINARIES: We first derive a number of properties of the utility distribution we use throughout the proofs.

Let $Q \approx U[0, 1]$ and $D = (1 - \varphi)D_s + \varphi D_l$, where $D_s \approx U[0, 1/2]$ and $D_l \approx U[0, 1]$ with Q, D_s, D_l independent. Let $g(u)$ denote the PDF and $G(u)$ denote the cdf of $U = Q - D$. Then we have

$$\mathbb{E}[U] = \frac{1 - \varphi}{4}, \quad (34)$$

$$g(u) = \begin{cases} \varphi(1 + u) & -1 \leq u < -\frac{1}{2} \\ 1 + (2 - \varphi)u & -\frac{1}{2} \leq u < 0 \\ 1 - \varphi u & 0 \leq u < \frac{1}{2} \\ (2 - \varphi)(u - 1) & \frac{1}{2} \leq u \leq 1 \end{cases} \quad (35)$$

$$G(u) = \begin{cases} \frac{1}{2}\varphi(u + 1)^2 & -1 \leq u < -\frac{1}{2} \\ \frac{1}{4}(\varphi(1 - 2u^2) + (2u + 1)^2) & -\frac{1}{2} \leq u < 0 \\ \frac{1}{4}(\varphi(1 - 2u^2) + 4u + 1) & 0 \leq u < \frac{1}{2} \\ \frac{1}{2}(\varphi(1 - u)^2 - 2(u - 2)u) & \frac{1}{2} \leq u \leq 1 \end{cases} \quad (36)$$

$$\Pr(U > t)\mathbb{E}(U|U > t) = \begin{cases} \frac{1}{24}(6 - \varphi(2 + 12t^2 + 8t^3)) & -1 \leq u < -\frac{1}{2} \\ \frac{1}{24}(7 - 12t^2 - 16t^3 - \varphi(3 - 8t^3)) & -\frac{1}{2} \leq u < 0 \\ \frac{1}{24}(7 - 12t^2 - \varphi(3 - 8t^3)) & 0 \leq u < \frac{1}{2} \\ \frac{1}{24}(8 - 24t^2 + 16t^3 - \varphi(4 - 12t^2 + 8t^3)) & \frac{1}{2} \leq u \leq 1 \end{cases} \quad (37)$$

PROOF OF PROPOSITION 1: The payoff function $\alpha N \varphi \left(q - c + \frac{1 - \varphi}{4} \right)$ is quadratic in φ and reaches its maximum in $\varphi^* = \frac{1 + 4q - 4c}{2}$. The latter is clearly increasing in q and decreasing in c . $\square$

PROOF OF PROPOSITION 2: The expected payoff $\mathbb{E}(\pi_R) = J \cdot \Pr(U_j > t)\mathbb{E}(U_j - c|\varphi, U_j > t)$ can be written as

$$\mathbb{E}(\pi_R) = J \int_t^{q+1} (x - c)g(x - q)dx \quad (38)$$

By increasing t the integration interval decreases, but the integrand does not change. Hence, the integral is maximized when the integrand is all non-negative, which occurs at $t^* = c$.

To calculate the optimal φ , we first need to calculate $J \cdot \Pr(U_j > c) \mathbb{E}(U_j - c | \varphi, U_j > c)$ as a function of φ . To use the results obtained in (37), we transform

$$E(\varphi) := \Pr(U_j > c) \mathbb{E}(U_j - c | \varphi, U_j > c) = \Pr(U_j > c) \mathbb{E}(U_j | \varphi, U_j > c) - c \Pr(U_j > c) = \\ \Pr(U > c - q) \mathbb{E}(U | U > c - q) - (c - q)(1 - G(c - q)) \quad (39)$$

since the U_j used here is shifted by q to the right compared to the U in (37). As apparent from (37), $E(\varphi)$ is linear in φ with a negative slope, therefore $\varphi E(\varphi)$ is maximized at

$$\arg \max_{\varphi} \varphi E(\varphi) = \frac{E(0)}{2(E(0) - E(1))} = \begin{cases} \frac{1-4(c-q)}{2} & c - q < -1 \\ \frac{3-12(c-q)}{2-12(c-q)-12(c-q)^2-4(c-q)^3} & -1 \leq c - q < -\frac{1}{2} \\ \frac{7-18(c-q)+12(c-q)^2+8(c-q)^3}{6-12(c-q)+8(c-q)^3} & -\frac{1}{2} \leq c - q < 0 \\ \frac{7-18(c-q)+12(c-q)^2}{6-12(c-q)+8(c-q)^3} & 0 \leq c - q < \frac{1}{2} \\ 1 & \frac{1}{2} \leq c \leq 1 \end{cases} \quad (40)$$

Examining the above function reveals that it never falls below 1, therefore we get the corner solution of $\varphi^* = 1$. $\square$

PROOF OF PROPOSITION 3: In case of the QA, the expected payoff can be written as

$$\mathbb{E}(\pi_R) = J \cdot \Pr(Q_j > t) \mathbb{E}(Q_j - D_j - c | \varphi, Q_j > t) = J \cdot \Pr(Q_j > t) (\mathbb{E}(Q_j - \mathbb{E}(D_j + c) | \varphi, Q_j > t)) = \\ J \cdot \Pr(U_j > t) \left(\mathbb{E} \left(Q_j - \frac{1+\varphi}{4} - c | \varphi, Q_j > t \right) \right) = J \int_t^{q+1} \left(x - \frac{1+\varphi}{4} - c \right) \cdot 1 dx \quad (41)$$

As in the case of the perfect algorithm, the integral is maximizes when the integrand is non-negative, that is at $t^* = \frac{1+\varphi}{4} + c$. To obtain the optimal φ , we calculate the integral

$$\int_{\frac{1+\varphi}{4}+c}^{q+1} \left(x - \frac{1+\varphi}{4} - c \right) \cdot 1 dx = \frac{(4q - 4c + 3 - \varphi)^2}{32} \quad (42)$$

yielding the optimal φ derived before the Proposition. $\square$

PROOF OF PROPOSITION 4: Similarly to the previous case, the expected payoff can be written as

$$\mathbb{E}(\pi_R) = J \int_{-1}^t \left(-x + q + \frac{1}{2} - c \right) \cdot ((1 - 2\varphi)I(x > -1/2) + \varphi I(x < -1/2)) dx \quad (43)$$

which is maximized at $t = c - q - \frac{1}{2}$. At this threshold the expected payoff is

$$\mathbb{E}(\pi_R) = \begin{cases} \frac{1-4(c-q)-4\varphi}{4} & c - q < -\frac{1}{2} \\ \frac{2-\varphi}{2} (c - q - \frac{1}{2})^2 & -\frac{1}{2} \leq c - q < 0 \\ \frac{1}{4} \left(1 - 4(c - q) - \varphi \left(1 - 2(c - q + \frac{1}{2})^2 \right) \right) & 0 \leq c - q < \frac{1}{2} \\ 0 & \frac{1}{2} \leq c - q \end{cases} \quad (44)$$

Examining the above function reveals that it never falls below 1, therefore we get the corner solution of $\varphi^* = 1$. $\square$

PROOF OF COROLLARY 1: When there is no curation, the average distance from which a receiver reads a message is simply the expected distance $\frac{1+\varphi}{4}$ with $\varphi_{NA}^* = \frac{1}{2} + 2(q - c)$, that is

$$\bar{d}_{NA} = \begin{cases} \frac{1}{2} & c - q < -\frac{1}{4} \\ \frac{3}{8} + \frac{q-c}{2} & -\frac{1}{4} \leq c - q \end{cases}$$

In case of the perfect algorithm, $\varphi_{PA} = 1$, but messages with utility below c are filtered out.

$$\bar{d}_{PA} = \begin{cases} \frac{1}{2} & c - q < -1 \\ \frac{1-3(c-q)+(c-q)^3}{3-6(c-q)-3(c-q)^2} & -1 \leq c - q < 0 \\ \frac{1+q-c}{3} & 0 \leq c - q < 1 \end{cases} \quad (45)$$

Under the quality algorithm, the distance is independent from the filtering, thus the expected distance $\frac{1+\varphi}{4}$ with $\varphi_{QA}^* = 1 + \frac{4}{3(q-c)}$, that is

$$\bar{d}_{QA} = \begin{cases} \frac{1}{2} & c - q < 0 \\ \frac{1}{2} + \frac{q-c}{3} & 0 \leq c - q \end{cases} \quad (46)$$

In case of the distance algorithm, $\varphi_{DA}^* = 1$ and messages with distance above $1/2 + q - c$ are filtered out. Hence the average distance is simply

$$\bar{d}_{DA} = \begin{cases} \frac{1}{2} & c - q < -\frac{1}{2} \\ \frac{1}{4} + \frac{q-c}{2} & -\frac{1}{2} \leq c - q \end{cases} \quad (47)$$

$\square$

PROOF OF PROPOSITION 5: Before examining comparative statics, we provide details on how we obtained the solution in equation 20. A sketch is already presented in the main text. To make sure that the first order condition results in a local maximum, we need to make sure that the second order condition associated with the first order condition presented in (16) is negative. Since g is a triangle function with a peak at 0 this is ensured if $u_l - q_i < 0$.

We also have to make sure that the solution of (16) yields not only a local, but also a global maximum. The marginal cost crosses the marginal revenue in at most two points, one being identified by the FOC. As the SOC reveals the marginal cost exceeds the marginal revenue after this point, so the payoff declines above this point. However, it is possible that the marginal cost function crosses the marginal revenue once more, from above for small values of q_i , creating a potential for $q_i = 0$ corner solution. To avoid this, we assume that $k < \bar{k}$, where $\bar{k} = g(u_l^*)$ is the marginal revenue at $q_i = 0$. It is apparent from the definition that $\bar{k} < 1$ and by checking the exact value at $\varphi = 1$, we see that $\bar{k} > \frac{3+\beta}{4} + (1-\beta)(r-c)$ ensuring that we obtain a non-trivial solution in (20) as long as β and r are not too high.

For the comparative statics results we differentiate the φ^* function given in (20) with respect to r, β, c and k and obtain the stated signs. To obtain the results on equilibrium quality, we differentiate $q^* = \frac{2\varphi^*-1}{4} + c$. $\square$

PROOF OF PROPOSITION 6: Before proving comparative statics, we detail how we obtained u_l after equation (24). Since senders are indistinguishable and set the same quality in equilibrium, we have

$$u_l = (1 - \beta)r + \beta\mathbb{E}(U_i | U_i > c) \quad (48)$$

as messages with utility below c are filtered out. If U denotes the baseline utility obtained from a message with $q = 0$, we get

$$\mathbb{E}(U_i | U_i > c) = \mathbb{E}(U + q | U > c - q) = \mathbb{E}(U | U > c - q) + q \quad (49)$$

Finally, we derive

$$\mathbb{E}(U + q | U > t) = \frac{1}{G(t)} \int_t^1 xg(x)dx = \frac{1}{G(t)} \left(\int_t^0 x(1+x)dx + \int_{\max(0,t)}^1 x(1-x)dx \right) \quad (50)$$

To prove part 1 of the proposition, we apply the implicit function theorem to (25) to obtain the stated results.

For part two, we use the φ_{NA}^* solution obtained in the previous Section in (20) and plug $q_{NA}^* = \frac{2\varphi_{NA}^* - 1}{4} + c$ into (25). We get that the LHS is larger than the RHS in this point, indicating that the marginal revenue is above the marginal cost. Since the marginal cost is constant, $q_{NA}^* < q_{PA}^*$ follows. $\square$

PROOF OF PROPOSITION 7: Part 1 can be verified by examination of the quality obtained in 29. For part 2, the quality under no curation case:

$$q_{NA}^* = \begin{cases} \frac{1-k}{1-\beta} + r & 0 \leq k \leq \frac{1}{4}(\beta + 4\beta c - 4c - 4\beta r + 4r + 3) \\ \frac{1}{4} \left(\frac{2(-2\beta - \sqrt{D} + 4(1-\beta)(r-c) + 5)}{4-2\beta} - 1 \right) + c & \frac{1}{4}(\beta + 4\beta c - 4c - 4\beta r + 4r + 3) \leq k \end{cases} \quad (51)$$

for $D = 4\beta(\beta + 1) + 8(1 - \beta)(r - c)(2\beta + 2(1 - \beta)(r - c) - 3) + 16(2 - \beta)k - 23$

Comparing expression 51 to 29 we find that q_{QA}^* is higher when $k > \beta + (\frac{1}{2} + r - c)(1 - \beta)$. $\square$

PROOF OF PROPOSITION 8: Part 1 can be verified by simple examination of the equilibrium quality obtained 33.

Proving part 2 requires a comparison between the equilibrium quality obtained 33 and the one obtained in Proposition 5. $\square$