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

Ron Berman† Zsolt Katona‡

Abstract

Curation algorithms are selection and ranking algorithms that social media platforms use to improve user experience. This paper analyzes the impact of curation algorithms on the number of friends consumers connect to and the quality of content created by producers. The model takes into account both vertical and horizontal differentiation and analyzes three different types of algorithms.

The results show that without algorithmic curation, the number of friends an individual has and the quality of content on the platform are strategic complements. Introducing algorithmic curation makes consumers less selective in their follower lists when content quality is low. In equilibrium, producers of content receive lower payoffs because they enter into a contest leading to a prisoner’s dilemma. The quality of content on the platform may increase if the marginal cost of producing this quality is high enough. Both of these effects may result theoretically in more diverse content consumption, but in equilibrium we find that a perfect filtering algorithm may reduce the horizontal distance of matched content resulting 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 the promise of personalized content. By telling the platform who their friends are, the online experience of users became not only more social, but the content consumed and generated was also tailored and customized to fit the users’ tastes. The main assumption was that by connecting to homophilous friends, consumers will be exposed to content which feels more personal and within their domain of interest. If in the past, for example, national newspapers did not delve into local matters while local newspapers did, social platforms were meant to be the ultimate source for relevant and personalized content.

A second appeal of social platforms was that users could also become content creators, while sharing it easily with their friends and followers. Creators of content could thus gain utility (through fame, reputation or intrinsic enjoyment) by having readers interact with the content without having to compete with other creators over limited publishing space.

The adoption of these platforms and accessibility of their content creation technology has led to a dramatic increase in the frequency and volume of available content. Statistics published by leading platforms\footnote{See, e.g. http://www.cio.com/article/2915592/social-media/7-staggering-social-media-use-by-the-minute-stats.html} 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 \cite{Rodriguez2014}.

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.”\footnote{https://www.facebook.com/help/327131014036297/} Recently Instagram introduced a ranking algorithm “so you’ll see the moments you care about first.”\footnote{http://blog.instagram.com/post/145322772087/160602-news} The problem the platform was trying to solve was that “On average, people miss 70 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—the 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 community preferences. 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 and creators of content. On the consumption side, consumers need to decide whom to follow (or friend) on social networks. Prior to the introduction of algorithmic curation, consumers manually curated their friends lists while taking into account the topics and quality of content their will encounter. Moreover, consumers needed to decide how much content to consume and subsequently which of these items merit interaction such as sharing or “liking” them. On the production side, content producers needed to decide how much to invest in the quality of each content item, as well as pick the topics and the frequency of posting content items.

When algorithmic curation is introduced to a platform, it may have a profound impact on the decisions of consumers and creators, which we call *receivers* and *senders*, respectively. For receivers, if algorithmic curation is able to improve their experience, they may become less selective in their friend selection resulting in longer friends lists. For senders, if algorithmic curation increases

⁴Ibid.
⁶The term was coined in Pariser (2011).
network connectivity, an individual sender may now face stronger competition in receiving attention for her content. Moreover, the sender may have a harder time passing the filtering bar of the algorithm, and a harder time passing the bar of quality that elicits a positive interaction from a receiver. 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.

To analyze these effects we set up an analytical model with three types of players—a platform, receivers and senders. Receivers derive utility from consuming 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 among items produced by the same sender. To control and maximize the utility they receive from items, receivers decide whom to friend as well as how many content items to consume.

Senders in our model receive utility when their content is liked by receivers. This process depends on the utility they provide to the receiver, but also on the competition they face from other senders. In response, senders can decide to increase the average quality of content they produce.

We compare three types of algorithms available to platforms with 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 that it can observe the true utility the receiver will obtain from content. The Quality Algorithm (QA) selects the content that passes a minimum quality bar, but cannot tell for each receiver if the item is close to her taste or far from it. In this way, the platform encourages high quality content, and focuses less on the individual match to each consumer. The third algorithm, which we call the Distance Algorithm (DA), selects for each consumer the content from the closest friends to her, but cannot tell whether each item is of high or low quality.

Our analysis focuses on the equilibrium quality of content and number of friends receivers follow. We first consider the decision of the receiver to follow senders and read content items. Second, we look at the decision of senders to set a quality level 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. Finally, 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 first results show that when the platform uses a curation algorithm, receivers are in general incentivized to connect to more senders because the platform provides assurances that the receiver will not experience too low utility. The perfect algorithm provides more assurance than the quality algorithm for the receiver and elicits higher connectivity. 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. Interestingly the distance algorithm is an exception as it does not increase connectivity. The fact that the platform filters horizontally differentiated senders does not give the receivers any incentive to increase connectivity which remains unchanged compared to the no algorithm case.

Examining the filtering process and the resulting connectivity in the network allows us to study the diversity of messages an individual receives. We do so by comparing the average social distance between a sender and a receiver of a consumed message. When the average quality of messages is low compared to consumption costs, both the perfect algorithm and the quality algorithm increase the average distance. This is a result of the increased connectivity achieved by these algorithms compared to the no algorithm case, which leads to connections to individuals at a higher distance. Even though some messages may come from individuals far away in the social space, they will pass through the filters if their quality is high enough. The distance algorithm does not affect the average distance since it does not change connectivity and the filtering threshold is high enough that it is not binding.

Intriguingly, when the average quality of messages is high, the perfect algorithm leads to the lowest average distance. The intuition is that above a certain level of quality, receivers have connected to all senders under all algorithms and except for the perfect algorithm, none of them filter any content, resulting in higher average distances. The perfect algorithm, however, still filters content with decently high quality, trying to compensate for the increased distances. This yields lower average distance observed by the receiver, a characteristic of a filter bubble.

As apparent, the popular argument that filtering always causes these bubbles does not hold in general. Our results show that only the perfect algorithm decreases the average distance and
only when messages are of high quality in a highly connected network. In the more typical setting of relatively low quality messages, filtering algorithms tend to increase the average distance and increase content diversity.

Analyzing the full model with endogenized sender behavior reveals that the equilibrium level of content quality created under the quality and perfect algorithms is higher than the benchmark case when the cost of investing in quality is high. When the costs are low, senders produce high enough quality that no filtering is needed and the algorithms are not effective. When the cost of quality is high, however, the algorithms effectively threaten the senders to increase their quality or face filtering. This allows the platform to provide receivers content that maintains a higher level of quality compared to the no algorithm case.

Interestingly content consumption costs have an opposite effect in the benchmark case of no algorithm and under the perfect algorithm. In the benchmark case an increase in content consumption costs reduces the quality and connectivity on the platform, whereas under the perfect algorithm it increases both. The reason for this is that the algorithm is able to use the consumption cost as an anchor for filtering, increasing the filtering threshold if the cost go up. As a result consumers do not need to consume less items, which allows senders to keep increasing quality profitably. The perfect algorithm thus transfers some of the consumption costs to senders and pressures them to create high quality content.

As a consequence, the payoff of the senders is higher under the perfect algorithm than under the other algorithms only for lower values of quality costs. When the cost of quality is high enough, the senders under the perfect algorithm face a highly competitive situation which requires them to maintain higher quality although their profits are eroded. This is a combination of 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 payoff. Third, since in equilibrium a higher level of quality arises, senders need to spend more on quality than in the case of no curation.

An interesting outcome of this analysis demonstrates that the social welfare under a curation algorithms is higher than without an algorithm only when quality is relatively cheap to produce.
Although the algorithms may reduce the payoff for senders, they more than compensate with increased consumer surplus, as long as the platform makes it cheap for senders to invest in quality. On the other hand, when quality costs are high, the social welfare declines when using an algorithm by placing too much pressure on content creators. While we do not directly study the business model and revenues of the platform, the social welfare analysis provides insights as the platform arguably profits most when the sum of payoffs in the ecosystem is the highest. In essence, the platforms face a conflicting choice. Either increase welfare and create a bubble, or increase diversity but lower welfare.

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, increased polarization in communication patterns, and dissatisfaction of advertisers because of lowered profits. A second contribution provides 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—in many cases it leads to increased polarization despite 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 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. Section 6 discusses the implications of curation on social welfare. Finally, Section 7 discusses limitations to 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
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 recommen-
dation 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 that 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 $\alpha$ denote the proportion of consumers who take on the role of a sender. Thus, we have $\alpha N$ 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. When receiver $R$ consumes a content item posted by sender $j$, the item yields utility $U_j = Q_j - D_j$. The quality of the item $Q_j$ is distributed uniformly in the $[q_j, q_j + 1]$ interval, where $q_j$ is the baseline quality level set by sender $j$. The distance $D_j$ between the sender $j$ and receiver $R$ is also a random variable and is distributed uniformly between $[0, 2\varphi]$, where $\varphi$ measures the connectivity of a receiver with $0 < \varphi < 1$. Receivers set $\varphi$ 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 network. When connecting to senders, receivers will take into account the expected utility of the content generated by each sender. If the receiver will be indifferent between connecting and not connecting with a sender, we assume they decide to not connect with a sender. This is equivalent to the assumption that connecting to senders has a small cost associated with it.\footnote{We elected not to model this cost to maintain the model’s parsimony.} 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 main assumption is that there are cases when receives will receive negative expected utility if they connect to too many senders and if the quality of content produced is not high enough to compensate for that.

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 $j = 1 \ldots J$.

Consequently, receiver $R$'s payoff is:

$$\pi_R = \sum_{j=1}^{J} (U_j - c) = \sum_{j=1}^{J} (U_j) - Jc \tag{1}$$

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 $S_j$ 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 $S_j = U_j$.

- The quality algorithm (QA) does not observe social distance nor takes it into account. It observers quality perfectly and sets the score to the quality of a piece of content, that is $S_j = Q_j$.

- The distance algorithm (DA) ignores quality and sets the score equal to the distance between the sender of a specific item and the receiver. In this case, the algorithm filters above a certain distance threshold. Thus, here we set $S_j = -D_j$, because $-D_j > t$ is equivalent to $D_j < -t$.

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 $j$ can increase her content quality by setting a higher $q_j$, resulting in a higher random quality $Q_j \approx U[q_j, q_j + 1]$. Higher quality comes at a cost of $kq_j$, making the sender’s utility

$$\pi_j^S = \#(likes) - kq_j$$

(2)
Receivers set $\varphi$. Senders set $q_i$.

Platform filters items with $S_j < t$.

Receivers read $J$ items. Pick one and “like” it if utility is higher than $r$.

Platform picks algorithm and sets $t$.

Payoffs are realized.

Figure 1: Timing of actions by the senders, receivers and the platform in the social network.

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 reserve utility $r$ that merits receiving a like. We assume the threshold $r$ is larger than the cost of reading an item $c$, which means a sender may receive a like only if they generated net positive utility for the receiver. Since the receiver picks an item randomly from her set of friends, having a larger set of friends means that the sender will face stronger competition receiving likes from receivers with more friends. In addition, if receivers have a high bar for liking content, senders will have a tougher time generating likes.

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

$$\pi_S^j = \varphi (1 - \alpha) N \frac{\Pr(U_j > r \text{ and } S_j > t)}{J} - kq_j$$

(3)

3.4 Timing

Figure 1 illustrates the timing of the game. First, receivers pick $\varphi$ to determine the number of friends they have and senders choose their quality levels $q_j$ simultaneously. Second, the platform chooses the threshold $t$ for their algorithm. When the content is produced by the senders, the platform filters some of the items based on the algorithm chosen. 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 \( q_j = q \).

4.1 Benchmark

We begin with the case where the platform does not use any filtering algorithm. A prototypical receiver’s expected payoff is

\[
\mathbb{E}(\pi_R) = J \cdot (\mathbb{E}(U_j|\varphi) - c)
\]

Note that the receiver cannot distinguish between senders and the expected utility is the same for all messages before reading, hence the receiver either reads content from no friend or from all friends. Given that \( \mathbb{E}(U_j|\varphi) = q + \frac{1}{2} - \varphi \), the receiver will not read any content if \( c - q > 1/2 - \varphi \), resulting in \( J = 0 \). Otherwise, the receiver will set \( J = \varphi N \) and read content from all friends.

In the first stage, when the receiver decides on \( \varphi \), the receiver solves:

\[
\max_{\varphi} \mathbb{E}(\pi_R) = \max_{\varphi} \varphi N \left( q - c + \frac{1}{2} - \varphi \right)
\]

The interior solution is \( \varphi^*_N = \frac{q-c}{2} + \frac{1}{4} \). Furthermore, if \( c - q > 1/2 \) receivers will not connect to anyone and set \( \varphi^*_N = 0 \), whereas if \( c - q < -3/2 \), receivers will connect to everyone and set \( \varphi^*_N = 1 \). The following proposition summarizes the main implications for the benchmark case: content quality and content consumption costs have opposite effects on the number of friends a receiver chooses. When quality is high, receivers are more open to friendships expecting higher content. However, a high consumption cost has the opposite effect, reducing the number of connections in the network.

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 \( \varphi^*_N = 1/4 \). In this case the majority of friends the receiver will connect to will be her closer friends, with distances lower than 1/2.
4.2 Perfect Algorithm

We now analyze the case of the perfect algorithm (PA) which fully observes and filters based on both message quality and sender distance. Starting at the last stage, recall that the platform intends to maximize receiver payoff. With a given $t$ threshold, the payoff is

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

(6)

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 $\Pr(U_j > t) \mathbb{E}(U_j - c, U_j > t)$, 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:

- When $-1 < c - q < 1$, receivers set $\varphi^* = \frac{1+q-c}{2}$. When $c - q \leq -1$ receivers connect to all senders in the network. When $c - q \geq 1$ receivers do not connect to any sender.

- 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 may have an incentive to connect to all senders when the cost is low enough compared to quality. 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

$$
\Pr(U_j > c) \mathbb{E}(U_j - c | \varphi, U_j > c) = \begin{cases} 
\frac{1}{6} \left( 4\varphi^2 - 6\varphi + 3 \right) & \varphi < \frac{1}{2} \\
\frac{1}{12\varphi} & \varphi \geq \frac{1}{2}
\end{cases}
$$

(7)
Thus, in this case, the receiver maximizes

$$\max_\varphi \mathbb{E}(u_R) = \max_\varphi \varphi\alpha N \Pr(U_j > c)\mathbb{E}(U_j - c|\varphi, U_j > c)|_{c=q}$$  \hspace{1cm} (8)$$
yielding the solution of $\varphi^*_P = 1/2$ which is clearly higher than the benchmark case of $\varphi^*_N = 1/4$. As opposed to (5), the benefit from increasing the number of friends has a positive quadratic term in the top line of (7), making it worthwhile to set a higher $\varphi$ than in the NA case.

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 $p^*_P = \Pr(U_j \leq c) = 1/2$ in the focal case of $c = q$.

It is interesting to contrast this result to the no-algorithm result. Although the receivers double their number of friends with the perfect algorithm, half the content is filtered in this focal case. The ratio of the amount of content consumed under the no-algorithm case to the perfect algorithm is 1, which means that the perfect algorithm does not effectively increase content consumption when $c = q$. However, the algorithm does make a difference in the average distance of a consumed message as we derive later in Section 4.5.

### 4.3 Quality Algorithm

In case of the quality algorithm (QA), the platform maximizes

$$\mathbb{E}(\pi_R) = J \cdot \Pr(Q_j > t)\mathbb{E}(U_j - c|\varphi, Q_j > t)$$  \hspace{1cm} (9)$$
As we derive in the proof, $\Pr(Q_j > t)\mathbb{E}(U_j - c|\varphi, Q_j > t)$ reaches its maximum at $t^* = c + \varphi$. As long as $q < t^* < q + 1$, we obtain

$$\Pr(Q_j > t^*)\mathbb{E}(U_j - c|\varphi, Q_j > t^*) = \frac{1}{2}(c - q + \varphi - 1)^2$$  \hspace{1cm} (10)$$
Solving the receiver’s maximization problem, we get

$$\varphi^*_Q = \frac{q - c + 1}{3}, \quad p^*_Q = \frac{2(c - q) + 1}{3}$$  \hspace{1cm} (11)$$
for $-1/2 < c - q < 1$, yielding the following result.

**Proposition 3.** Under the quality algorithm:
Receivers connect to more senders than without a filtering algorithm if \( c - q > -1/2 \). When \( c - q \leq -1/2 \) the quality algorithm does not filter any items and the receivers connect to the same number of friends as with the no algorithm case.

Receivers connect to fewer friends and receiver utility is lower than under the perfect algorithm.

The proposition demonstrates that the quality algorithm is generally a less effective solution than the perfect algorithm in increasing consumer utility and connectivity. However, consumers receive enough assurance from the algorithm to connect to more friends than without an algorithm as long as \( c - q \) is not too low. The perfect algorithm is more efficient than the quality algorithm in both increasing connectedness as well as increasing consumer utility. 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 (DA), the platform maximizes

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

(12)

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 receiver’s maximization problem, we get the following result.

**Proposition 4.** Under the distance algorithm:

- **Receivers set** \( \varphi^*_{DA} = \frac{2-c}{2} + \frac{1}{4}, \) when \(-3/2 < c - q < 1/2\). When \( c - q \leq -3/2 \) receivers connect to all senders in the network. When \( c - q \geq 1/2 \) receivers do not connect to any sender.

- **No content is filtered out by the algorithm in equilibrium.**

Interestingly, the distance algorithm does not encourage more connectedness in the network than in the benchmark NA case. The reason is that since the algorithm is set to filter out all content with some distance above a threshold, there is no value of adding friends with higher distances, which results in no actual content being filtered out in equilibrium.
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 \( \tilde{d} \) denote the expected value of a random piece of content that a receiver encounters in equilibrium. We obtain the following results.

**Corollary 1.** When \( c - q \) is low, \( \tilde{d}_{QA} = \tilde{d}_{NA} = \tilde{d}_{DA} > \tilde{d}_{PA} \). For intermediate values of \( c - q \), \( \tilde{d}_{PA} > \tilde{d}_{QA} \geq \tilde{d}_{NA} = \tilde{d}_{DA} \), and for high values of \( c - q \), \( \tilde{d}_{QA} = \tilde{d}_{PA} > \tilde{d}_{NA} = \tilde{d}_{DA} \).

The results clearly show the different ways the three algorithms affect content available to readers. An immediate implication of the corollary is that when the quality is high (\( c - q \) is low), the average distance of content seen by the receiver under all algorithms is higher than that of the perfect algorithm, resulting in more diverse content being displayed. The intuition is that above a certain level of quality, receivers have connected to all senders under all algorithms and except for the perfect algorithm, none of them filter any content, resulting in higher average distances. The perfect algorithm, however, still filters content with quite high quality, trying to compensate for the increased distances, yielding lower average distance observed by the receiver.

When quality is low, it is interesting that focusing on improving the user’s experience and utility increases average distances and content diversity compared to the no algorithm case. This is a result of the increased connectivity achieved by these algorithms compared to the no algorithm case, which results in shifting the distribution of distances further and allowing the receiver to experience more diverse content.

Figure 2 compares the impact of changes in \( c - q \) on the connectivity of the receivers under the different algorithms (top panels) and the average distance of content consumed by the receiver with the different algorithms (bottom panels). To emphasize the differences between the algorithms we graph the absolute connectivity and average distances (left panels) and the difference of the PA and QA cases from the NA cases (right panels).

In the top panels we can see that both the perfect and quality algorithms generally increase connectivity compared to the new algorithm, and that this effect is stronger when the quality of content \( q \) is low, or the cost of consuming content \( c \) is high. Despite the increased connectivity, the bottom panels show that the filtering effect of the algorithms decreases the difference in average
distances compared to the differences in connectivity, and in the case of the perfect algorithm, this may result in shorter average distances to senders, a characteristic of a filter bubble.

The pattern we find sheds lights on which algorithms are responsible for filter bubbles. The popular argument that filtering always causes these bubbles does not hold in general. Our results show that only the perfect algorithm decreases the average distance and only under special circumstances: when quality is high compared to consumption costs and when consumers connect to many senders to start with. In the typical situations—with relatively low quality senders—filtering algorithms tend to increase the average distance between a sender and receiver of a consumed message.

5 Analysis of Senders and Equilibrium

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 $j$’s decision, we denote the quality level set by her as $q_j$, while all other senders set $q_{-j}$. Let $G(\cdot)$ denote the CDF of the utility distribution $U_j = Q_j - D_j$ when $q_j = 0$ and let $g(\cdot)$ denote its PDF. In the absence of curation, sender $j$ maximizes her expected payoff as follows:

$$\max_{q_j} \mathbb{E}(\pi^i_j) = \max_{q_j} \frac{\varphi N}{2} \cdot \frac{1 - G(r - q_j)}{J} - kq_j. \quad (13)$$

When receivers find consuming messages worthwhile they set $J = \varphi N/2$, hence the first order condition becomes:

$$g(r - q_j) = k \quad (14)$$

$^9$Exact formulas for the CDF and PDF are given at the beginning of the Appendix.
Figure 2: Receiver connectivity \( \varphi \) (top panels) and average distance \( \bar{d} \) (bottom panels) as a function of \( c - q \). The left panels display the functional values, while the right panels display the difference from the no algorithm case.
Since the second order conditions of the maximization problem hold only on the increasing parts of $g(\cdot)$, we find that the best response quality of a sender, $q^*$, to the share of friends each receiver has, $\varphi$, is

$$q^*(\varphi) = r + 2\varphi(1 - k)$$  \hspace{1cm} (15)

We also assume $0 \leq k < k_{NA}$ to make sure that the local maximum obtained above is also a global maximum.\footnote{The value is fully derived in the Appendix.}

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 derive the equilibrium, we use the formula from the previous section, where we determined that the receiver sets $\varphi = \frac{2(r-c)+1}{4}$. Substituting this into (15), we get a quadratic equation for $\varphi$ with the following solutions:

$$\varphi^* = \begin{cases} 
1 & 0 < k \leq \frac{2(r-c)+1}{4} \\
\frac{1+2(r-c)}{4k} & \frac{2(r-c)+1}{4} < k < k_{NA}
\end{cases}$$  \hspace{1cm} (16)

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. A comparative statics analysis reveals the following:

**Proposition 5.** Both $q^*$ and $\varphi^*$ increase in $r$ 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.

When the cost of consuming content $c$ is increasing, the receiver decreases the amount of friends she 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.

It is interesting to note that the positive interaction between quality and connections hurts senders in general. While more connections provide senders more opportunity for senders to reach an audience, the limited rewards from a single receiver cancels this benefit out. At the same time, increased connectivity results in potentially larger social distances forcing senders to work harder and to invest more in quality in order to overcome the distances.

5.2 Perfect Algorithm

As we have shown in Section 4, the perfect algorithm filters out any messages that yield negative payoff to a receiver. 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 message 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 to like 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 sender who sets quality $q$ is filtered out is $G(c - q)$. We thus obtain the payoff for a given sender $j$, when everybody else sets $q_{-j}$ as follows:

$$\max_{q_j} \frac{\varphi N}{2} \cdot (1 - G(r - q_j)) \sum_{L=0}^{J-1} \left[ \frac{1}{L+1} \binom{J-1}{L} (1 - G(c - q_{-j}))^L G(c - q_{-j})^{J-1-L} \right] - kq_j. \quad (17)$$

The first order condition is thus

$$\frac{\varphi N}{2} \cdot g(r - q_j) \sum_{L=0}^{J-1} \left[ \frac{1}{L+1} \binom{J-1}{L} (1 - G(c - q_{-j}))^L G(c - q_{-j})^{J-1-L} \right] = k. \quad (18)$$

To obtain a tractable formula we assume that $N \to \infty$, and conduct all further analysis in this section in the limit. The first order condition becomes

$$\frac{g(r - q_j)}{(1 - G(c - q_{-j}))} = k. \quad (19)$$

Compared to the benchmark case, it is apparent that senders have a higher incentive to invest in quality for a fixed like threshold, $r$. This is driven by the reduced competition due to filtering. On the other hand, the filtering threshold increases in the cost $c$ and potentially lowers incentive to
invest in quality. To fully understand the combination of these forces, we determine the equilibrium quality level. As we show in the Appendix, senders will always set \( q_j^* \geq r \). Substituting \( q_j = q_{-j} = q^* \) into (19), and solving for the equilibrium, we obtain the equilibrium quality\(^{11}\)

\[
q^* = \begin{cases} 
2(1 - k) + r & 0 < k \leq \frac{r-c}{2} \\
\frac{1+(c+2)k-k\sqrt{2k(c+2k-r)+1}}{k} & \frac{r-c}{2} < k < \bar{k}_{PA}
\end{cases}
\]

(20)

Analyzing the result above we obtain the following results:

**Proposition 6.**

- \( \varphi^* = 1 \) and the receivers connect to all senders.
- \( q^* \) increases in \( r \) and in \( c \) and decreases in \( k \).
- For low values of \( k \) senders set the same quality and obtain the same payoff as without filtering.
- For intermediate values of \( k \), senders set a higher quality and obtain a higher payoff than without filtering.
- For high values of \( k \), senders set a higher quality than without an algorithm, but obtain a lower payoff when \( r - c \) is low.

The first part of the proposition states that receivers connect in equilibrium to all senders. Given the full insurance the platform provides against low utility items, there is no risk for the receivers now and they might as well connect to all senders. That is, the result we obtained by just examining receivers is reinforced by the presence of active senders. The perfect algorithm induces receivers to connect to more senders and this incentivizes senders to invest in content quality, further increasing the willingness of senders to establish connections.

The second part of the proposition shows that \( r \) 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 second part also reveals that increasing receivers’ cost has the opposite effect compared to without an algorithm. 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

\(^{11}\)The value of \( \bar{k}_{PA} \) is derived in the Appendix
low utility. As our results show, curation has a double positive effect on receivers. In addition to
the baseline effect of filtering out 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 effort on quality to avoid
being hidden from receivers.

The remaining results establish a comparison between the case of no filtering and perfect filter-
ing. While qualities and sender payoffs are equal between the two cases when the cost of quality
is low, filtering increases quality efforts when $k$ is high enough as discussed above. This threat of
being filtered out makes senders 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 for high values
of $k$ when $r - c$ is low enough. 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 by the indirect, competitive effect
akin to a prisoner’s dilemma.

5.3 Quality Algorithm

As the quality algorithm filters out any message with quality less than $t^*$, senders can increase their
base quality $q_j$ to make sure their items are not being filtered by the platform. Likewise, because
receivers only like content above the threshold $r$, a sender can increase their quality $q_j$ to generate
items with utility that passes this threshold.

Using the result from Section 4, suppose the platform sets $t^* = c + \varphi$ as the threshold for
filtering quality, then the probability that an item from any senders passes filtering is $Pr(Q > t^*) = \min(q + 1 - t^*, 1)$. In order to receive a like, sender $j$’s item needs to pass the quality filter
as well as the utility threshold $r$ set by the receiver. We denote as $P(q_j) = Pr(Q_j > t^*, U_j > r)$
the probability of item $j$ not being filtered for quality and being eligible for receiving a like. The
sender then maximizes

$$\max_{q_j} \frac{P(q_j)}{\min(q - j + 1 - t^*, 1)} - kq_j$$

The full derivation is available in the Appendix.
Examining the profit function and its first and second order conditions, the profit is locally maximized only when $q_j \geq t^*$. As we show in the Appendix, when $r$ is high, the qualities coincide with the case of the no algorithm. For intermediate values of $r$, however, solving for the first order condition when assuming that in equilibrium $q_j = q^*$, we obtain\(^{13}\)

\[
q^* = \begin{cases} 
  r + 2\varphi(1 - k) & 0 < k \leq \frac{1}{2} + r - c \\
  c + \varphi & \frac{1}{2} + r - c < k \leq \overline{k}_{QA}
\end{cases}
\] (22)

Analyzing the equilibrium quality level and connectivity, we find that

\[
\varphi_{QA}^* = \begin{cases} 
  1 & 0 < k \leq \frac{2(r - c) + 1}{4} \\
  \frac{1 + 2(r - c)}{4k} & \frac{2(r - c) + 1}{4} < k \leq \frac{1}{2} + r - c \\
  \frac{1}{2} & \frac{1}{2} + r - c < k \leq \overline{k}_{QA}
\end{cases}
\] (23)

The following proposition summarizes the results.

**Proposition 7.** Under the quality algorithm:

- **No filtering takes place in equilibrium.**
- **When $k$ is low, the equilibrium behaves as in the no algorithm case.** $q^*$ increases in $r$ and decreases in $k$ and $c$.
- **When $k$ is high, senders set a higher quality in equilibrium, yet obtain a lower payoff compared to without filtering.**

The first part of the proposition finds that quality filtering is effective in the sense that senders always set a high enough quality to make sure their content is not filtered. This effectively increases the cost of quality facing the senders.

The second part finds similar effects for $r$, $k$, and $c$ as in the no algorithm case, and generally shows that when quality is cheap, quality filtering is not necessary.

The third part shows that quality filtering results in higher equilibrium quality levels for high quality cost $k$ but a lower profit for the senders. In such cases, quality filtering acts as a threat for the senders to maintain a higher quality even if the cost is too high, and as a result, senders

\(^{13}\)The value of $\overline{k}_{QA}$ is derived in the Appendix.
maintain a fixed quality level above the level they would have set without filtering. This enhanced quality causes higher connectivity in equilibrium, but results in lower payoff for senders because of additional costs faced by the senders. The intuition is that the quality filter puts a 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.

5.4 Distance Algorithm

As the results in Section 4 have shown, using a distance algorithm does not cause receivers to change their connectivity, which results in no filtering taking place for any value of $q$ that may be set by the senders.

Consequentially, since senders can only affect the quality of the content they create, when facing a distance filtering algorithm senders have no incentive to invest in quality differently than in the case of no filtering. The conclusion is that the results of Proposition 5 apply to this case as well.

5.5 Comparison

The previous sections have shown that the cost of content consumption $c$, the utility threshold $r$ for receiving likes and the cost of producing quality $k$ all interact and may impact which algorithm generates higher connectivity, higher quality and higher sender profits in equilibrium. In this section we compare the effects of the different algorithms.

Figure 3 illustrates the equilibrium values of $q^*$ and $\varphi^*$ as functions of $k$ when fixing $c = 0.35$ and $r = 0.4$. Comparing the top and bottom left panels, we notice the correlation between the quality and connectivity, which is a result of the strategic complementarity between quality and connectivity. We also see that when $k$ is low, the filtering algorithms do not make a difference as senders elect to produce high quality content which in turn entices receivers to connect to all senders. For high values of $k$, the quality algorithm leads to somewhat higher quality and connectivity levels compared to the case of no algorithm or the distance algorithm, but the perfect algorithm generally achieves higher qualities and full connectivity. The right hand side panels illustrate how increases

---

14 These values were chosen because they allow a wide range of $k$ values where the equilibrium is not a corner solution, but the same general pattern holds for all parameter values.
in $k$ cause an increase in the difference of both quality and connectivity as compared to the no algorithm case.

Focusing on filter-bubbles, we can also compare the equilibrium average distance $\overline{d}$ of content receivers consume. For the NA and QA cases, the average distance is equal to the equilibrium connectivity $\varphi^*$. When comparing the NA and PA cases, we arrive at the following result:

**Corollary 2.** There exists a $\frac{r - c^2}{2} + \frac{1}{4} < k < \min(k_{PA}, k_{NA})$ s.t. if $0 < k < k$ then $\overline{d}_{NA} > \overline{d}_{PA}$. When $k > k$, $\overline{d}_{NA} < \overline{d}_{PA}$.

The corollary shows that the findings of corollary 1 hold in equilibrium, and that for low values of $k$, the perfect algorithm will create a filter-bubble where receivers are exposed to less diverse content compared to without an algorithm.

Finally, in Figure 4 we illustrate the profits achieved by the senders in the conditions of Figure 3. We see that the PA profit declines continuously in the cost of quality $k$, while under the no algorithm case and quality algorithm the profit is initially lower than the PA profit, but as it declines slower with $k$, for high values of $k$ it is essentially higher. This is a result of the higher quality created by the perfect algorithm, which eventually costs too much and yields lower profit. In summary, these figures illustrate our main results well. Filtering algorithms force senders to create higher quality content especially in the case of the perfect algorithm. However, unless content creation costs are low, this is a burden for senders, lowering their profits and hindering their potential willingness to enter the market. To better understand the effects of algorithms on the entire ecosystem provided by the platform, we examine the overall social welfare in the next section.

### 6 Social Welfare

Throughout the analysis we have focused on the connectivity of the receivers, quality of content and profit of senders as the algorithms the platform uses have direct implications on those. Although in this paper we do not explicitly model the profit goal of the platform, it is intuitive that increased welfare may attract more receivers and senders to the social network in the long run, resulting in higher profit for the platform.

To compare the social welfare between the three different cases we notice that when the

---

15The distance algorithm case is equivalent to the no algorithm case and we omit it.
Figure 3: Sender equilibrium quality $q^*$ connectivity (top panels) and receiver equilibrium connectivity $\varphi^*$ (bottom panels) as a function of $k$. The left panels display the functional values, while the right panels display the difference from the no algorithm case. $\bar{k} \approx 0.727$ is the maximal value of $k$ under which all senders are profitable and generate quality above 0 when $c = 0.35$ and $r = 0.4$. 
market is fully symmetric, the social welfare is proportional to the sum of profit of one sender and the expected utility of one receiver:

\[ SW = \pi^*_S + q^* + \frac{1}{2}d - c \]  

(24)

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 receivers and senders to participate in the platform. Using these results we are able to compare equilibrium social welfare when we focus on low values of \( r \), as they allow a wide range of quality costs \( k \) for which senders will participate in the platform under every algorithm.

The following proposition summarizes the result

**Proposition 8.** For low values of \( r \):

- The social welfare with the quality algorithm is equal to without an algorithm for low values of \( k \). When \( r - c + \frac{1}{2} < k < \frac{2}{2c-2r+3} \) it is higher under the quality algorithm, and when \( \frac{2}{2c-2r+3} < k \leq \min(\bar{k}_{NA}, \bar{k}_{QA}) \) it is higher without an algorithm.
• When $0 < k < \frac{r - c}{2}$, the social welfare under the perfect algorithm is higher than the social welfare without an algorithm.

• When $\frac{r - c}{2} \leq k < \min(\overline{k}_{NA}, \overline{k}_{PA})$ there exists a range of $r$ and $c$ with a corresponding $k$ threshold such that when $k > k$, the social welfare with the perfect algorithm is lower than with no algorithm.

• $k$ is less than $\frac{2}{2r - 2r + 3}$ when $r < r - c < 1$ with $r \approx 0.033$. Otherwise, $k > \frac{2}{2r - 2r + 3}$.

The proposition shows that the quality algorithm in general is unable to generate more social welfare than not using an algorithm at all, except for a small range of intermediate values of $k$. The intuition behind this result stems from the fact that although the quality of content items may increase under the quality algorithm resulting in higher consumer utility, in equilibrium this results in lower sender profit, and higher connectivity which yields higher average distances to the receiver. Combining these three effects together results in lower social welfare. The intermediate value is unique because there is a range where quality and connectivity are higher (and thus improve welfare), but the cost of quality is still not high enough to generate a decrease in welfare, resulting in more welfare for the quality algorithm case.

When examining the perfect algorithm, we see a more nuanced result. For low sender cost levels, the perfect algorithm is able to increase quality without penalizing senders in their payoff, resulting in higher social welfare. When $k$ is high, on the other hand, the payoff of senders under the perfect algorithm drops below that of the NA case, which lowers social welfare below the NA case. Combined with the increased connectivity and higher average distances generated under the perfect algorithm, this may result in lower social welfare.

The last item compares the point where the social welfare for the quality algorithm goes below the no algorithm, to the point of the perfect algorithm. For a wide range of $r - c$ values, we see that the social welfare difference becomes negative for lower values of $k$ with the perfect algorithm compared to the quality algorithm. This means that for high values of $k$, the quality algorithm may have an advantage in improving social welfare.

An implication of this result is that applying curation algorithms is effective 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 Insta-
Figure 5: Social welfare as a function of $k$ when $c = 0.35$ and $r = 0.4$. The left panel displays the values, while the right panel displays the difference from the no algorithm case. $\bar{k} \approx 0.727$ is the minimal value of $k$ under which all senders are profitable and generate quality above 0 when $c = 0.35$ and $r = 0.4$.

gram 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 to content by receivers.

Figure 5 illustrates the three cases. It is clear that for most of the range of $k$ the perfect algorithm generates more welfare 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 very 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.

7 Discussion and Conclusion

The implementation of curation algorithms by social platforms raises questions about their impact in the long run on content diversity experienced by users. If filter bubbles are indeed a result of using curation algorithms, platforms should carefully consider which algorithms best serve their
users. As an example, strong curation may allow “fake news” to spread since they match the preconceived notion of some readers, but they never reach other readers who are able to flag them as fake.

Despite the important implications, arguments surrounding these algorithms tend to be simplistic and ignore many facets of this complicated ecosystem. Our model that includes creators of content (senders), consumers of content (receivers) and an intermediating platform, has allowed us to show that algorithmic curation can alter the structure of the network as well as the quality and diversity of content on a social network.

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. Furthermore, we find that algorithms such as the PA may decrease the distance of matched content between receivers and senders and create filter bubbles, while other algorithms (such as the QA) may increase this distance on average. 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.

Our analysis of the two-sided market required simplifying assumptions and is of course not without limitations. These simplifications facilitated both parsimony of exposition as well as the ability to perform more specific analysis of different aspects of the resulting content distribution. Specifically, we have abstracted away from the explicit goal of the platform beyond focusing on specific aspects of consumer utility. Moreover, there are other possibilities to describe the receiver consumption decision and the “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.

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. The resulting higher quality may increase connectivity, but may potentially reduce senders payoffs and diversity of content consumed by receivers. One way platforms can approach this problem is by supplying tools that allow producers easier creation and editing of content, while focusing their algorithms on encouraging overall higher quality content without trying to match consumer tastes.

References


Appendix

PRELIMINARIES: We first derive a number of properties of the utility distribution we use throughout the proofs. Let $Q \sim U[0, 1]$, $D \sim U[0, \phi]$ and $g(), G()$ be the pdf and cdf of $U = Q - D$. We have:
\[ E[U] = \frac{1}{2} - \varphi. \]  

(25)

\[
g(u) = \begin{cases} 
1 + \frac{u}{2\varphi} & -2\varphi \leq u < \min(1 - 2\varphi, 0) \\
\frac{1}{2\varphi} & \min(1 - 2\varphi, 0) \leq u < 0 \\
1 & 0 \leq u < \max(0, 1 - 2\varphi) \\
\frac{1-u}{2\varphi} & \max(0, 1 - 2\varphi) \leq u \leq 1 
\end{cases}
\]  

(26)

\[
G(u) = \begin{cases} 
\frac{u + \frac{u^2}{4\varphi} + \varphi}{4\varphi} & -2\varphi \leq u < \min(1 - 2\varphi, 0) \\
1 + \frac{2u-1}{4\varphi} & \min(1 - 2\varphi, 0) \leq u < 0 \\
u + \varphi & 0 \leq u < \max(0, 1 - 2\varphi) \\
1 - \frac{(1-u)^2}{4\varphi} & \max(0, 1 - 2\varphi) \leq u \leq 1 
\end{cases}
\]  

(27)

\[
\Pr(U > t)E(U|U > t) = \begin{cases} 
\frac{1}{6} \left( -\frac{t^3}{\varphi} - 3t^2 + 4\varphi^2 - 6\varphi + 3 \right) & -2\varphi \leq t < \min(1 - 2\varphi, 0) \\
\frac{1-3t^2}{12\varphi} & \min(1 - 2\varphi, 0) \leq t < 0 \\
\frac{1}{6} \left( -3t^2 + 4\varphi^2 - 6\varphi + 3 \right) & 0 \leq t < \max(0, 1 - 2\varphi) \\
\frac{(t-1)^2(2t+1)}{12\varphi} & \max(0, 1 - 2\varphi) \leq t \leq 1 
\end{cases}
\]  

(28)

**Proof of Proposition 1**  
The payoff function \( \varphi\alphaN \left( q - c + \frac{1-2\varphi}{2} \right) \) is quadratic in \( \varphi \) and reaches its maximum at \( \varphi^* = \frac{2-q}{2} + \frac{1}{4} \). The latter is clearly increasing in \( q \) and decreasing in \( c \). □

**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

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

(29)

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 \( \varphi \), we first need to calculate \( J \cdot \Pr(U_j > c)\mathbb{E}(U_j - c|\varphi, U_j > c) \) as a
function of \( \varphi \). To use the results obtained in (28), we transform

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

(30)
since the \( U_j \) used here is shifted by \( q \) to the right compared to the \( U \) in (28). The resulting function is increasing for every \( \varphi < \frac{1-\tilde{t}}{2} \) when \( \tilde{t} = c - q \). Hence, the receiver sets \( \varphi^{\ast}_{PA} = \frac{1+q-c}{2} \).

To prove the second part of the proposition, the probability of content being filtered by the algorithm is \( \Pr(U_j \leq c | \varphi^{\ast}) = \Pr(U \leq \tilde{t} | \varphi^{\ast}) = G(\tilde{t} | \varphi^{\ast}) \) when \( \tilde{t} = c - q \). We get

\[
\frac{dG(\tilde{t}, \varphi^{\ast})}{d\tilde{t}} = g(\tilde{t} | \varphi^{\ast}) + \frac{\partial G(\tilde{t}, \varphi^{\ast})}{\partial \varphi^{\ast}} \frac{\partial \varphi^{\ast}}{\partial \tilde{t}}
\]

(31)

This expression is always positive for \(-1 \leq \tilde{t} \leq 1\), and since \( \tilde{t} \) is increasing in \( c \) and decreasing in \( q \), this proves the result.

\[\square\]

**Proof of Proposition 3** In case of the QA, the expected payoff can be written as

\[
\mathbb{E}(\pi_R) = J \cdot \Pr(Q_j > t) \mathbb{E}(Q_j - D_j - c | \varphi, Q_j > t) = J \cdot \Pr(Q_j > t) (\mathbb{E}(Q_j - \mathbb{E}(D_j + c) | \varphi, Q_j > t)) = J \cdot \Pr(U_j > t) (\mathbb{E}(Q_j - \varphi - c | \varphi, Q_j > t)) = J \int_{t}^{q+1} (x - \varphi - c) \cdot 1dx
\]

(32)

As in the case of the perfect algorithm, the integral is maximized when the integrand is non-negative, that is at \( t^{\ast} = \varphi + c \). To obtain the optimal \( \varphi \), we calculate the integral

\[
\int_{\varphi+c}^{q+1} (x - \varphi - c) \cdot 1dx = \frac{1}{2}(c - q + \varphi - 1)^2
\]

(33)

Plugging the integral into the receiver’s utility function, we find that it has a local maximum at \( \varphi^{\ast} = \frac{1}{3}(1 - c + q) \), and a local minimum at \( 1 - c + q \). We require that \( c + \varphi^{\ast} < q + 1 \) to make sure the algorithm does not filter all content, which means the local maximum is also a global maximum. In addition, since the algorithm can filter for qualities between \( q \) and \( q + 1 \), when \( c - q < -\frac{1}{2} \), the algorithm does not filter at all, and the receiver experiences the no algorithm case.

Comparing this value to the value found for the PA algorithm in proposition 2, we find that \( \varphi^{\ast}_{PA} > \varphi^{\ast}_{QA} \) when \( 1 > c - q > -\frac{1}{2} \). The other items are straightforward comparisons to the values found for the other algorithms.

\[\square\]
Proof of Proposition 4: Similarly to the previous case, the expected payoff can be written as

$$E(\pi_R) = J \Pr(D_j < -t)E(U_j - c|\varphi, D_j < -t)$$

$$= J \int_{0}^{-t} \left(q + \frac{1}{2} - x - c\right) \cdot g_D(x) dx$$

When $g_D()$ is the PDF of $D_j$. The expected payoff is maximized at $t = c - q - \frac{1}{2}$. At this threshold the expected payoff is

$$E(\pi_R) = \begin{cases} \varphi \left(q - c - \varphi + \frac{1}{2}\right) & c - q \leq \frac{1}{2} - 2\varphi \\ \frac{1}{16}(2q - 2c + 1)^2 & \frac{1}{2} - 2\varphi < c - q \leq \frac{1}{2} \end{cases}$$

The function is maximized at $\varphi^* = \frac{2(q-c)+1}{4}$ for $-\frac{3}{2} < c - q < \frac{1}{2}$. When $c - q \leq -\frac{3}{2}$, $\varphi^* = 1$ and when $c - q \geq \frac{1}{2}$, $\varphi^* = 0$.

Using this result, the amount of content filtered by the algorithm amounts to zero, because the receiver has no benefit of adding senders with distance that will be filtered later.

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

$$d_{NA} = \begin{cases} 1 & c - q \leq -\frac{3}{2} \\ \frac{q-c}{2} + \frac{1}{4} & -\frac{3}{2} < c - q \leq \frac{1}{2} \\ 0 & c - q > \frac{1}{2} \end{cases}$$

In case of the perfect algorithm messages with utility below $c$ are filtered out. Hence, the average distance encountered by a customer is $E_D[D|U_j > c]$ with $\varphi = \varphi_{PA}^*$. Calculation results in:

$$d_{PA} = \begin{cases} 1 & c - q \leq -2 \\ \frac{1}{3} \left( \frac{1}{c-q} - \frac{5}{c-q+4} + q - c + 4 \right) & -2 < c - q \leq -1 \\ \frac{1+3(q-c)+3(q-c)^2}{3(1+2(q-c))} & -1 < c - q \leq 0 \\ \frac{q-c+1}{3} & 0 < c - q \leq 1 \\ 0 & c - q > 1 \end{cases}$$
Under the quality algorithm, the distance is independent from the filtering, thus the expected
distance $\varphi$ with $\varphi_{QA} = \frac{q-c+1}{3}$ when $1 > c - q > -1/2$ and $\varphi_{NA}$ otherwise, that is

$$d_{QA} = \begin{cases} 
1 & c - q \leq \frac{-3}{2} \\
\frac{q-c}{2} + \frac{1}{4} & \frac{-3}{2} < c - q \leq \frac{-1}{2} \\
\frac{q-c+1}{3} & \frac{-1}{2} < c - q \leq 1 \\
0 & c - q > 1 
\end{cases}$$

(39)

In case of the distance algorithm, the expected distance is the same as under the no algorithm case.

Comparison of these values shows that $d_{QA} \geq d_{NA} = d_{DA}$ for for every $1 \geq c - q \geq -2$ with
strict inequality when $1 > c - q > -\frac{1}{2}$ the applicable range of values. In addition, $d_{PA} > d_{NA}$ if
and only if $A < c - q < 1$ where $A \approx -1.263$.

**Proof of Proposition 5.** Before examining comparative statics, we provide details on how
we obtained the solution in equation (16). A sketch is already presented in the main text. To ensure
that the first order condition results in a local maximum, we need to check that the second order
condition associated with the first order condition presented in (14) is negative. This is satisfied
when $q_j \geq r$ and $r + 2\varphi \geq q_j \geq r + 2\varphi - 1$. We also have to make sure that the solution of (14)
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 $q_j$, creating a potential for $q_j = 0$ corner solution. To avoid this, we assume that
$k < k_{NA}$, where in equilibrium the profit of the sender is positive when $k < k_{NA}$. Solving for the
positive profit results in $k_{NA} = \min(\frac{8c}{2(r+c)} - 2, 1)$ ensuring that we obtain a non-trivial solution
in (16) as long as $r < 1$.

For the comparative statics results we differentiate the $\varphi^*$ function given in (16) with respect to
$r, c$ and $k$ and obtain the stated signs. To obtain the results on equilibrium quality, we differentiate
$q^* = r + 2\varphi^*(1 - k)$.

**Proof of Proposition 6.** Before proving comparative statics, we detail how we obtained
$q^*(\varphi)$. From the point of view of receiver $j$, since the marginal cost of quality is fixed at $k$, her
best response will fall in the range where the marginal revenue \( g(r - q_j) \) is decreasing in \( q_j \). Examination of \( g(r - q_j) \) shows this happens only when \( q_j \geq r \) and \( r + 2\varphi \geq q_j \geq r + 2\varphi - 1 \). Adding the condition \( k < \tilde{k}_{PA} \) with

\[
\tilde{k}_{PA} = \min \left( 2 \sqrt{\frac{r(r + 4)((c - r)^2 - 8)}{(c^2 - r(r + 4))^2(c(c + 8) - r(r + 4) + 16)^2}}, \frac{(c - r)(c(r + 2) - (r - 2)(r + 4))}{(c^2 - r(r + 4)) (c(c + 8) - r(r + 4) + 16)} \right),
\]

(40)
similarly to the previous section is sufficient to ensure a global maximum. Solving the FOC \( g(r - q) - k(1 - G(c - q)) = 0 \) along with the solution \( \varphi^* = \frac{1 + q - c}{2} \) from Proposition 2 and the above conditions, shows the only solution is the one we report. The comparative statics in part two follow in a straightforward manner from \( q^* \).

For parts three to five, the equilibrium qualities are:

\[
q^*_{NA} = \begin{cases} 
q \leq r(k - 2) + 1 & 0 < k \leq \frac{r - c}{2} + \frac{1}{4} \\
\frac{1}{4}(-2c(k - 2) - 2(k + 2)r + k + 2) & \frac{r - c}{2} + \frac{1}{4} \leq k < \tilde{k}_{NA}
\end{cases}
\]

(41)

\[
q^*_{PA} = \begin{cases} 
q \leq r(k - 2) + 1 & 0 < k \leq \frac{r - c}{2} \\
\frac{1}{4}(2c(k - 2) - 2(k + 2)r + k + 2) & \frac{r - c}{2} \leq k < \tilde{k}_{PA}
\end{cases}
\]

(42)

The senders’ equilibrium profits are:

\[
\pi^N_{SA} = \begin{cases} 
k^2 - k(r + 2) + 1 & 0 < k \leq \frac{r - c}{2} + \frac{1}{4} \\
\frac{1}{4}(2c(k - 2) - 2(k + 2)r + k + 2) & \frac{r - c}{2} + \frac{1}{4} \leq k < \tilde{k}_{NA}
\end{cases}
\]

(43)

\[
\pi^N_{PA} = \begin{cases} 
k^2 - k(r + 2) + 1 & 0 < k \leq \frac{r - c}{2} \\
\frac{1}{4}(2c(k - 2) - 2(k + 2)r + k + 2) & \frac{r - c}{2} \leq k < \tilde{k}_{PA}
\end{cases}
\]

(44)

For \( 0 < k \leq \frac{r - c}{2} \) and \( \frac{r - c}{2} < k \leq \frac{r - c}{2} + \frac{1}{4} \) the results follow from a direct comparison. When \( \frac{r - c}{2} + \frac{1}{4} < k \leq \min(\tilde{k}_{NA}, \tilde{k}_{PA}) \), we find that \( \pi^N_{SA} > \pi^N_{PA} \) when:

\[
k > \tilde{k}(r, c) = \frac{16c^3 + c^2(34 - 48r) + \sqrt{-(c - r)^2 - 8}^2 (4c^2 - 8c(r - 1) + 4(r - 2)r - 33)}{(13c - 13r + 34)(5c - 5r - 2)} - 2 - 2 \frac{4c(r(12r - 17) - 8) + 2r(17 - 8r + 16) - 36}{(13c - 13r + 34)(5c - 5r - 2)}
\]

(45)

(46)

(47)
Solving for $r$ and $c$ in the condition $\frac{r - c}{2} + \frac{1}{4} < \hat{k}(r, c) \leq \min(\hat{k}_{NA}, \hat{k}_{PA})$, yields that for $r < \frac{1}{17}(53 - 8\sqrt{33})$ there exists a range of values of $0 < c < r$ s.t. the profits under the NA case are higher than under the PA case. For example, when $r < 0.21$, this condition holds for any $c < r$.

\[\square\]

**Proof of Proposition 7** The probability of receiving a like, $P(Q_j)$ depends on the values of $r$ and $t$ and can be broken up into three cases:

When $r \geq t$:

\[
P(Q_j) = \begin{cases} 
0 & q_j \leq r - 1 \\
\frac{(q_j - r + 1)^2}{4 \varphi} & r - 1 < q_j \leq \min(r, r + 2 \varphi - 1) \\
\frac{2q_j - 2r + 1}{4 \varphi} & r < q_j \leq r + 2 \varphi - 1 \\
q_j - r - \varphi + 1 & r + 2 \varphi - 1 < q_j \leq r \\
1 - \frac{(q_j - r - 2 \varphi)^2}{4 \varphi} & \max(r, r + 2 \varphi - 1) < q_j \leq r + 2 \varphi \\
1 & q_j > r + 2 \varphi 
\end{cases}
\]  

(48)

When $r + 2 \varphi \geq t > r$:

\[
P(Q_j) = \begin{cases} 
0 & q_j \leq t - 1 \\
\frac{(q_j - t + 1)(q_j - 2r + t + 1)}{4 \varphi} & t - 1 < q_j \leq \min(r + 2 \varphi - 1, t) \\
\frac{2q_j - 2r + 1}{4 \varphi} & t < q_j \leq r + 2 \varphi - 1 \\
q_j - \frac{(r - t)^2}{4 \varphi} & r + 2 \varphi - 1 < q_j \leq t \\
1 - \frac{(q_j - r - 2 \varphi)^2}{4 \varphi} & \max(t, r + 2 \varphi - 1) < q_j \leq r + 2 \varphi \\
1 & q_j > r + 2 \varphi 
\end{cases}
\]  

(49)

The case $t > r + 2 \varphi$ is not possible since $t^* = c + \varphi$ and $r \geq c$, and thus we omit it.

When using each one of these functions in the profit function, we notice (similar to the NA case) that the marginal revenue may cross the marginal cost up to two times, in which case one point will be a local maximum, and the other a local minimum. In addition, since both functions have linear parts, it is possible to have a discontinuity in the first order from a positive marginal profit to a negative one, depending on the value of $k$.
The second order condition of Equation 21 is only negative on the concave parts of the functions, which always happen in the range \( q_j \geq t \). In addition, on the linear parts the maximum is achieved at the right edge of the range, which is always the beginning of a concave part of the function. The best response of the receiver is therefore in the range \( q_j \geq t \), which means that in equilibrium \( P(Q_{-j} > t) = 1 \), and effectively the content of senders is not filtered because of low quality.

Using this simplification, we find that when \( r \geq t \), the solution to the first order condition coincides with the NA result, but when \( r + 2\varphi > t \geq r \), the solution has a discontinuity, yielding the one appearing in 22:

\[
q(\varphi) = \begin{cases} 
  r + 2\varphi(1 - k) & 0 < k \leq \frac{1}{2} + r - c \\
  c + \varphi & \frac{1}{2} + r - c < k \leq k_{QA}
\end{cases}
\]  

(50)

With \( k_{QA} = \min(1, \frac{-4c^2+c(8r+4)-4r(r+1)+7}{8c+4}) \)

Using this result and the result from section 4.3 that \( \varphi^*_NA = \frac{q-c+1}{3} \) when \( c - q > \frac{1}{2} \) and \( \frac{2(q-c)+1}{4} \) when \( c - q \leq -\frac{1}{2} \) we find the feasible equilibrium values \( q^*_QA \) and \( \varphi^*_QA \) which adhere to the constraint that the profit is positive for some positive value of \( k \).

The result is:

\[
\varphi^*_QA = \begin{cases} 
  1 & 0 < k \leq \frac{2(r-c)+1}{4} \\
  \frac{1+2(r-c)}{4k} & \frac{2(r-c)+1}{4} < k \leq \frac{1}{2} + r - c \\
  \frac{1}{2} & \frac{1}{2} + r - c < k \leq k_{QA}
\end{cases}
\]  

(51)

The comparative statics follow directly from the equilibrium values of \( q^* \).

\[ \square \]

**Proof of Corollary 2.** As mentioned in the text, \( d_{NA} = \varphi^*_NA \) and \( d_{QA} = \varphi^*_QA \). For the perfect algorithm case:

\[
d_{PA} = \begin{cases} 
  \frac{1}{3} \left( \frac{1}{c+2k-r-2} - \frac{5}{c+2k-r+2} - c + 2k + r + 6 \right) & 0 < k \leq \frac{r-c}{2} \\
  \frac{1}{3} \left( 1 - \frac{2c(2k+4k^2-2k+1+3k+1)}{2k+4k^2-2k+1-ck+kr-1} \right) & \frac{r-c}{2} < k < k_{PA}
\end{cases}
\]  

(52)

The result follows from direct comparison of the values when \( r \leq 1 \).

\[ \square \]

**Proof of Proposition 8.** The social welfare functions are:

\[
SW_{NA} = \begin{cases} 
  k^2 - k(r + 4) + r + 2 & 0k \leq \frac{r-c}{2} + \frac{1}{4} \\
  -2c(4k+4k^2-2k+1+3k+1) & \frac{r-c}{2} + \frac{1}{4} < k < k_{NA}
\end{cases}
\]  

(53)

39
The proof follows when assuming $r < 1$ for the first two items. For the third item it is possible to show a boundary on $k$ when $1 > r > r$ and $r \approx 0.39$. A similar bound can be found for values of $c$.

For the fourth item, the social welfare difference $SW_{PA} - SW_{NA}$ at $k = \frac{2}{2c-2r+3}$ equals:

$$SW_{PA} = \begin{cases} \frac{1}{3} \left( \frac{5}{c+2k-r+2} + \frac{1}{-c-2k+r+2} + c + k(3k - 3r - 10) + 2r + 3 \right) & 0 < k \leq \frac{r-c}{2} \\ -3k^3(c(c+4)-(r-2)(r+6))+2k^2(3c(-c+r-2)+3r+8)+(2c(9-5y)+2ry-6r+3y-3)-8y+8 & \frac{r-c}{2} < k \leq \bar{k}_{PA} \end{cases}$$

and $y = \sqrt{2k(c + 2k - r) + 1}$

$$SW_{QA} = \begin{cases} k^2 - k(r + 4) + r + 2 & 0 < k \leq \frac{r-c}{2} + \frac{1}{4} \\ -2c((k-4)k+1)+k(k-2(k+2)r)+2r+1 & \frac{r-c}{2} + \frac{1}{4} < k \leq r - c + \frac{1}{2} \\ \frac{1}{2} (-4c^2 + c(-8k + 8r + 4) - 4k - 4r(r + 1) + 7) + c < k < \bar{k}_{QA} \end{cases}$$

When we denote $x = r - c$ and $y = \sqrt{12(x-2)x + 25}$. This function is decreasing in $x$, and is negative when $1 > x = r - c > \bar{r}$ where $r \approx 0.033$. \hfill \Box