

---

# Modeling and Analysis of Information Communities

---

**Peter Marbach**

Department of Computer Science  
University of Toronto

## Abstract

In this paper we propose a mathematical model to study communities in social networks where agents share/exchange information. We refer to such a social network as an information network. More precisely, we assume that there is a population of agents who are interested in sharing content. Agents differ in the type of content they are interested in, as well as their ability to produce content. Agents can form/join communities in order to maximize their utility for obtaining and producing content. We use the framework of Nash equilibrium to characterize the community structures that emerge under this model. In our analysis, we show that there always exists a Nash equilibrium. In addition, we characterize the community structure, as well as the content that is being produced, under a Nash equilibrium.

## 1 Introduction

Communities play an important role in social networks. However while there exists a large body of work that formally models and studies macroscopic properties of social networks such as the degree distribution and diameter, less work is available on mathematical models for microscopic properties of communities in social networks. Creating such a model is the topic of this paper. Ideally this model should be simple enough to allow a formal analysis, yet be expressive enough to provide insights into important microscopic properties of communities in social networks.

For our model we focus on a particular type of social networks, to which we refer to as *information networks*, where agents (individuals) share/exchange information. Sharing/exchanging of information is an important aspect of the social networks, both for social networks that we form in our everyday lives, as well as for online social networks such as for example Twitter.

To model communities in information networks, to which we refer to as *information communities*, we assume that all content (information) that is being shared is being produced by agents in the network. Furthermore, we assume that different agents have different abilities to produce (generate) content, and different interests in consuming (obtaining) content. Using this setup, we study the situation where agents form communities in order to share/exchange content. We formally model this situation through utility functions that captures the benefit that agents obtain from being in a particular information community, where we assume that agents obtain an utility both from consuming (obtaining) content as well as from producing content. Using this model we study the following questions: a) which community structures will emerge under this model, i.e. can we characterize the set of content producers/consumers that form a community, and b) which content will individual agents produce in a given community.

The model that we use to characterize information networks has three important components: how we model the space of content that is being produced and consumed in the, how we model agent's interests and ability to produce content, and how we measure the utility that agents obtain for consuming and producing content.

For our analysis we assume that each content item is being produced in an information network can be associated with a particular content type. We might think of a content type as a topic, or particular interest that agents have. In addition, we assume that the space of possible content types has a certain structure. In particular, we assume that there exists a measure of “closeness” (similarity) between content types that characterizes how strongly related two content types are. For example, “basketball” and “baseball” are both sports and one would assume that these two topics are stronger related with each other than for example “basketball” and “mathematics”. The assumption that there exists such a measure of closeness is an important feature of our model.

Another important aspect of the proposed model is how we characterize the agent’s interest, and agent’s ability to produce content. For our analysis, we assume that agents in a community might have different interests with respect to the content they are interested in, and different abilities to produce content. The model which type of content a given agent is interested in is based on the following intuition. We assume that each agent has a main interest, i.e. there exists a content type (topic) that the agent is interested in the most. However, we assume that that agents are not only interested in getting information for their main interest but also other type of content. Moreover, the further away a given content type is from their center of interest, the less interested they are in this content type. Or in other words, the less similar a given content type is with respect to the agent’s main interest, the less interested the agent is in this content type. For example, an agent that is interested in “baseball” might also be somewhat interested in “basketball”, but not necessarily in “mathematics”.

Similarly, we assume that agent’s have different abilities to produce content. Here we assume that agent’s have the highest ability to produce content for their center of interest. Or in other words, if an agent produces content items for the topic they are most interested in, then this content item is very likely to be relevant for this topic (content type). In addition, the further away (dissimilar) a given content type is from the agent’s center of interest, the less likely it is that a content item that the agent produces for this topic is indeed relevant.

Finally, we model the utilities that agents obtain for consuming and producing content as follows. We assume that for each content item that an agent consumes, the agent receives a reward if that content item is of interest to the agent. In addition, each content item that the agent consumes that agents pays a cost. This cost accounts for the effort/time required by an agent to consume a single content item (and decide whether it is of interest or not). Similarly, we assume that for each content item that an agent produces it receives a benefit that is equal to the number of agents that consume the content item, and find it to be of interest to them. In addition, each content item that an agent produces incurs a cost that is proportional to the number of agents that consume the content item. Under this model, an agent receives a high utility for content consumption if it receives a lot of content item that are of interest, and very few content items that are not of interest. Similarly, an agent receives a high utility for content production if most agents that consume the content items the agent produces find the content item of interest. These definitions imply that an information network is efficient (i.e. agents receive a high utility), if agents receive all content that is of interest to them, and no content that is not of interest. Of in other words, the information network has a structure such that the content each agent receives is content that is of interest to the agent.

In our analysis we focus on a particular content space, namely a one dimensional content space with the torus metrics. In addition, we focus on the limiting case of a dense agent population where we characterize the agent population through a density function. These assumptions simplify the analysis, and lead to results that are easier to interpret and obtain insights into the community structure in information networks.

An interesting outcome of the analysis is that the proposed model, albeit being very simple, indeed seems to be able to provide interesting insights into the microscopic structure of information communities. For example, the characterization of how content is being produced, i.e. which content each agent in a community produces, indeed matches what has been experimentally observed in real-life social networks. We discuss this connection in more details in Section 2 and 6.

In summary, the contributions of the paper are as follows. We propose a mathematical model to study information communities that is simple enough to allow a formal analysis. In particular, using a game-theoretic framework, we characterize the community structure that arises under a Nash equilibrium. Furthermore, the model is able to capture important microscopic properties of

information communities that have been observed in real-life social networks, such as for example the structure of content production in information communities.

Whereas this paper focuses on the analysis of the proposed model, there are several possible extension, as well as applications of the model to study additional properties of information networks such as content filtering and dissemination, as well as the connectivity structure within an information community. We discuss this research directions in more details in Section 7.

The rest of the paper is structured as follows. In Section 2 we discuss existing work that is most closely related to the model and analysis presented in this paper. In Sections 3 - 5 we define our mathematical model for information communities, and in Section 6 we present our result on community structures that emerge under a Nash equilibrium.

## 2 Related Work

We note that there is a large body work on the experimental and theoretical study of social networks and their properties; due to space constraints we highlight here only the work that is most relevant to the discussion presented in this paper.

There exists extensive work, both experimental and theoretical, on the macroscopic properties of social network graphs such as the small world phenomena, shrinking diameter, and power-law degree distribution. By now there are several mathematical models that describe well these properties; examples of such models are Kroenecker graphs [7] and geometric protean graphs [1]. The difference between this body of work and the model presented here is that these models on the social network graph a) do not explicitly model and analyze community structures, and b) focus on macroscopic properties of the social network graph rather than microscopic properties of communities in social networks.

There is also a large body of work on community detection algorithms (see for example [4] for a survey) including minimum-cut methods, hierarchical clustering, Givran-Newman algorithm, modularity maximization, spectral clustering, and many more. In this paper we are not so much concerned with detecting communities (clusters) in a social network (complex graph), but with modeling and characterizing the community structure that emerges in an information network. An interesting approach to community detection is taken in [3] by Chen et al. who use a game-theoretic approach to detect overlapping communities in social networks. In [3] it is assumed that there already exists an underlying social graph for the social network, and the utility that agents obtain when joining a community depends on a) which other agents joined the community and b) the underlying social graph. This is different from the approach in this paper where the utility depends on content production and consumption. In addition, the goal of the work in [3] is to develop an algorithm to detect overlapping communities, whereas in this paper we aim at characterizing the microscopic structure of information communities.

Related to the analysis in this paper is the work on content forwarding and filtering in social networks [2, 5, 6]. In particular the work by Zadeh, Goel and Munagala [2], and the work by Hegde, Massoulie, and Viennot [6]. In [2], Zadeh, Goel and Munagala consider the problem of information diffusion in social networks under a broadcast model where content forwarded (posted) by a user is seen by all its neighbors (followers, friends) in the social graph. For this model, the paper [2] studies whether there exists a network structure and filtering strategy that lead to both high recall and high precision. High recall means that all users receive all the content that they are interested in, and high precision means that all users only receive content they are interested in. The main result in [2] shows that this is indeed the case under suitable graph models such as for example Kroenecker graphs. In [6], Hegde, Massoulie, and Viennot study the problem where users are interested in obtaining content on specific topics, and study whether there exists a graph structure and filtering strategy that allows users to obtain all the content they are interested in. Using a game-theoretic framework (flow games), the analysis in [6] shows that under suitable assumptions there exists a Nash equilibrium, and selfish dynamics converge to a Nash equilibrium. The main difference between the model and analysis in [2, 6] and the approach in this paper is that model and analysis in [2, 6] does not explicitly consider and model community structures, and the utility obtained by users under the models in [2, 6] depends only on the content that agents receive, but not on the content agents produce.

There exists an interesting connection between the modeling assumption made by Zadeh, Goel and Munagala in [2] and by Hegde, Massoulié, and Viennot in [6], and a result obtained in this paper (Proposition 2, Section 5). Both papers [2, 6] make the modeling assumption that users produce content only on a small subset of content that they are interested in receiving. Zadeh, Goel and Munagala support this assumption in [2] through experimental results obtained on Twitter data that shows that Twitter users indeed tend to produce content on a narrower set of topics than they consume. Proposition 2 in Section 6 in this paper provides a formal validation/explanation for this assumption as it shows that under the proposed model it is optimal for agents (users) to produce content on a small subset of the content type that they are interested in consuming. This result illustrates that the proposed model is able to capture and explain important microscopic properties of information networks and communities.

### 3 Content Production and Consumption

In this section we present the mathematical model we use for our analysis. In particular, we describe the model that we use to characterize the content that agents produce and consume in an information network. In addition, we introduce how we characterize agents' interest in a particular type of content, as well as their abilities to produce a particular type of content.

For our analysis, we model the content that agents produce in an information community as follows. We assume that each content item that is being produced by an agent in the information network belongs to a particular content type. One might think of a content type as a topic, or particular interest that agents have. Furthermore, we assume that there exists a structure that relates the different content types with each other. In particular, we assume that there exists a measure of "closeness" between content types that characterizes how strongly related two content types are. For example "basketball" and "baseball" are both sports, and one would assume that these two topics are stronger related with each other than for example "basketball" and "mathematics".

More formally, we model the type of a content item as follows. We assume that the type of a content item is given by a point  $x$  in a metric space  $\mathcal{M}$ . The assumption that content types lie in a metric space allows for a natural way to compare the closeness between content types. In particular, the distance between two content types  $x, y \in \mathcal{M}$  is then given by the distance measure  $d(x, y)$ ,  $x, y \in \mathcal{M}$ , of the metric space  $\mathcal{M}$ .

Having defined the space  $\mathcal{M}$  of content types, we next describe agent's interest and ability to produce content. Recall that we assume that agents in a community might have different interests with respect to the content they are interested in, as well as different abilities to produce content.

We first model the content that a given agent is interested in. Our model is based on the following intuition. We assume that agents have a main interest, i.e. for each agent there exists a content type  $x \in \mathcal{M}$  that they are most interested in. Moreover, agents are interested in more than one content type, i.e. agents are not only interested in getting information for their main interest  $x \in \mathcal{M}$  but also other type of content. However, the further away a given content type is from their center of interest, the less interested they are in this content type.

To model this situation, we associate with each agent that consumes content a center of interest  $y \in \mathcal{M}$ . The center of interest of a given agent is the content type (topic) that an agent is most interested in. Given the center of interest  $y \in \mathcal{M}$  of an agent, the interest of agent  $y$  in content of type  $x \in \mathcal{M}$  is then given by

$$p(x|y) = f(d(x, y)),$$

where  $d(x, y)$  is the distance between the center of interest  $y$  and the content type  $x$ , and  $f : [0, \infty) \mapsto [0, 1]$  is a non-increasing function. The interpretation of the function  $p(x|y)$  is as follows: when agent  $y$  consumes (reads) a content item of type  $x$ , then it finds it interesting with probability  $p(x|y)$  given by

$$p(x|y) = f(d(x, y)).$$

As the function  $f$  is non-increasing, this model captures the intuition that agent an  $y \in \mathcal{M}$  is more interested in content that is close to its center of interest  $y$ .

We will refer in the following to an agent by its center of interest  $y \in \mathcal{M}$ , i.e. when we refer to an agent  $y \in \mathcal{M}$  we refer to the agent whose center of interest is equal to  $y$ .

Next consider a given agent that is producing content, and let  $y \in \mathcal{M}$  be the center of interest of the agent. The ability of agent  $y$  to produce content of type  $x \in \mathcal{M}$  is then given by

$$q(x|y) = g(d(x, y)),$$

where  $g : [0, \infty) \mapsto [0, 1]$  is a non-increasing function. The interpretation of this function is as follows. If agent  $y$  produces content of type  $x$ , then the content will be relevant to content type  $x$  with probability  $q(x|y)$  given by

$$q(x|y) = g(d(x, y)).$$

As the function  $g$  is non-increasing, this model captures the intuition that agent an  $y \in \mathcal{M}$  is better at producing content that is close to its center of interest  $y$ .

### 3.1 Assumptions on the functions $f$ and $g$

For our analysis we make additional assumptions on the functions  $f$  and  $g$  that we use to define the functions  $p(\cdot|y)$  and  $q(\cdot|y)$ . To state these assumptions, we introduce a few more definitions.

Throughout the paper, we use the following notation. Given a real-valued function  $f : \mathcal{M} \mapsto \mathbb{R}$  on a metric space  $\mathcal{M}$ , we define the support  $\text{supp}(f)$  by

$$\text{supp}(f) = \bar{A}$$

where

$$A = \{x \in \mathcal{M} | f(x) \neq 0\},$$

and  $\bar{A}$  is the closure of  $A$ .

Given real-valued function  $f : \mathcal{M} \mapsto \mathbb{R}$  on a metric space  $\mathcal{M}$ , we say that  $f$  is symmetric with respect to  $y \in \mathcal{M}$  if for  $x, x' \in \mathcal{M}$  such that

$$d(x, y) = d(x', y),$$

we have that

$$f(x) = f(x').$$

Using these definitions, we make the following assumptions for the function  $f$  and  $g$ .

**Assumption 1.** *The function  $f : [0, L] \mapsto [0, 1]$  is strictly decreasing and twice continuously differentiable on  $[0, L]$ , and has a bounded first derivative on  $[0, L]$ . Furthermore, the function  $f$  is locally strictly concave, i.e. there exists a constant  $b$ ,  $0 < b \leq L$ , such that*

$$f''(x) < 0, \quad x \in [0, b].$$

*The function  $g : [0, L] \mapsto [0, 1]$  is non-increasing on  $[0, L]$ , and strictly concave and twice continuously differentiable on its support  $\text{supp}(g)$  with*

$$g(0) > 0$$

and

$$g'(0) = 0.$$

The assumptions that the functions  $f$  and  $g$  are twice differentiable and (locally) strictly concave, are a technical assumptions used in the proofs of our results.

### 3.2 Content Space $\mathcal{R}$

In addition, we consider for our analysis a metric space that has a particular structure. More precisely, we consider a one dimensional metric space with the torus metric. The reason for using this structure is that it simplifies the analysis and allows us to obtain simple expressions for our results, that can easily be interpreted.

More formally, we consider in the following one-dimensional metric space for our analysis. The metric space is given by an interval  $\mathcal{R} = [-L, L] \in \mathbb{R}$ ,  $0 < L$ , with the torus metric, i.e. the distance between two points  $x, y \in \mathcal{R}$  is given by

$$d(x, y) = ||x - y|| = \min\{|x - y|, 2L - |x - y|\},$$

where  $|x|$  is the absolute value of  $x \in (-\infty, \infty)$ .

Note that we have that

$$||x - y|| \leq L, \quad x, y \in \mathcal{R}.$$

Furthermore, we have the following two properties,

1. for  $x, y \in \mathcal{R}$ , the addition of  $x$  and  $y$  is given by

$$x + y = \begin{cases} x + y, & x + y \in [-L, L]. \\ -2L + x + y, & x + y \geq L. \\ 2L + x + y, & x + y < -L. \end{cases}$$

2. we have that

$$x < y, \quad x, y \in \mathcal{R}.$$

if there exists a point  $b$ ,  $b \in (0, L)$ , such that

$$x = y - b.$$

Using the torus metric for the content space  $\mathcal{R}$  eliminates “border effects”, in the sense are no points that have a “special” position as it would be for example the case if we would an interval  $[-L, L]$  as the content space. This simplifies the analysis, and leads to simpler expressions for our results.

### 3.3 Limiting Case of a Dense Agent Population

Finally, we characterize the set of agents that participate in the information network, i.e. the set of agents consume and produce content in the information network. As we identify an agent by its center of interest  $y \in \mathcal{R}$ , the set of agents that consume content, as well as the set of agents that produce content, are then given by a subset of  $\mathcal{R}$ . There are two possible approaches to characterize these two sets of agents.

One approach is to assume that there exists a finite number of agents that produce, and consume, content. For example, the set of agents that consume content could be given by set of  $K$  agents  $y_k$ ,  $k = 1, \dots, K$  such that

$$y_k = -L + k\delta,$$

where  $\delta > 0$  is given by

$$\delta = \frac{2L}{K}.$$

This approach would result of a set  $K$  agents are “uniformly” spaced over the set  $\mathcal{R}$  with distance  $\delta$ .

Another approach is to consider the limiting case of a dense agent population. In particular, we obtain this limiting case by letting the above distance  $\delta$  between agents to approach 0, and consider the resulting density  $\mu(y)$ ,  $y \in \mathcal{R}$ , of agents in  $\mathcal{R}$ . We refer to this case as the continuous agent population model.

In this paper we consider the second approach and characterize the distribution of agents on  $\mathcal{R}$  by a density  $\mu(y)$ , where we make the following assumption on the density  $\mu(y)$ .

**Assumption 2.** *Both agents that consume content, and agent that produce content, are uniformly distributed on  $\mathcal{R}$  with density*

$$\mu(y) = 1, \quad y \in \mathcal{R}.$$

## 4 Information Communities

Having defined the content space, and the agent's interest and ability to produce content, we next model a community in an information network. Here we assume that a community is given by a set of users that produce content, and a set of users that consumes content.

More formally, for our analysis we assume that an information community  $C = (C_d, C_s)$  is given by a set  $C_d \subseteq \mathcal{R}$  of agents that consume content in the community  $C$ , and a set  $C_s \subseteq \mathcal{R}$  of agents that produce content in the community  $C$ . The subscript in  $C_d$  in  $C_s$  refer to "demand", and "supply", respectively. In the following we define more precisely how content is being produced and consumed, and the resulting utility functions for content consumption and production.

For a given a community  $C = (C_d, C_s)$ , we assume that each agent  $y \in C_s$  can decide how much effort they put into producing content of type  $x$ . We model this situation as follows. We let  $\beta_C(x|y)$ ,  $x \in \mathcal{R}$ , be the rate at which an agent  $y \in C_s$  generates content of type  $x$  in the community  $C$ . The function  $\beta_C(\cdot|y)$  is a non-negative function, and we have that

$$\beta_C(x|y) \geq 0, \quad x \in \mathcal{R}.$$

The total rate (over all agents  $y \in C_s$ ) at which content of type  $x$  is generated in the community  $C$  is then given by

$$\beta_C(x) = \int_{C_s} \beta_C(x|y) dy.$$

Recall that content of type  $x$  that is generated by agent  $y \in C_s$  is relevant to  $x$  with probability  $q(x|y)$ , and that total rate (over all agents  $y \in C_s$ ) at which relevant content is generated in  $C$  is equal to

$$Q_C(x) = \int_{C_s} \beta_C(x|y) q(x|y) dy.$$

We refer to  $Q_C(x)$  as the content supply function for community  $C$ .

Furthermore, we assume that each agent  $y \in C_d$  can decide on how much time it allocates to read (consume) content that is being produced in community  $C$ . Let  $\alpha_C(y)$ ,

$$0 \leq \alpha_C(y) \leq 1,$$

be the the fraction of time that agent  $y \in C_d$  allocates for consuming content in community  $C$ , and let the function  $P_C(x)$  be given by

$$P_C(x) = \int_{C_d} \alpha_C(y) p(x|y) dy.$$

We refer to  $P_C(x)$  as the content demand function of community  $C$ .

Using these definitions, we next characterize the the utilities that agents obtain in a given community  $C = (C_d, C_s)$  for consuming, and producing, content. We first define the utility for content consumption. For this, we assume that consuming a content item incurs a cost  $c$ . The cost  $c$  is a (per item) processing cost that reflects the effort/time required by an agent to read a single content item (and decide whether it is of interest or not). If the content item is of interest then the agent receives a reward equal to 1; otherwise the agent receives a reward equal to 0.

Using this cost and reward structure, we then obtain the following utility function for agents  $y \in C_d$ . Note that if the fraction of time that agent  $y \in C_d$  is consuming content in community  $C$  is equal to  $\alpha_C(y)$ ,

$$0 \leq \alpha_C(y) \leq 1,$$

then the (time-average) rate  $\mu_C(x|y)$  at which an agent  $y \in C_d$  receives rewards in community  $C$  is given by

$$\begin{aligned} \mu_C(x|y) &= \alpha_C(y) \int_{C_s} \beta_C(x|z) q(x|z) p(x|y) dz \\ &= \alpha_C(y) Q_C(x) p(x|y). \end{aligned}$$

The (time-average) rate at which an agent  $y \in C_d$  consumes content of type  $x$  in the community  $C$  is given by

$$\alpha_C(y) \beta_C(x),$$



and the (time-average) cost rate on agent  $y$  is given by

$$\alpha_C(y)\beta_C(x)c.$$

Combining these results, the (time-average) utility rate for content consumption  $U_C^{(d)}(y)$  of an agent  $y \in C_d$  in community  $C$  is given by

$$U_C^{(d)}(y) = \alpha_C(y) \int_{\mathcal{R}} [Q_C(x)p(x|y) - \beta_C(x)c] dx. \quad (1)$$

Similarly, the (time-average) utility rate for content production  $U_C^{(s)}(y)$  of an agent  $y \in C_s$  in community  $C$  is given by

$$\begin{aligned} U_C^{(s)}(y) &= \int_{x \in \mathcal{R}} \beta_C(x|y) \int_{z \in C_d} \alpha_C(z) [q(x|y)p(x|z) - c] dz dx \\ &= \int_{\mathcal{R}} \beta_C(x|y) [q(x|y)P_C(x) - \alpha_C c] dx, \end{aligned} \quad (2)$$

where

$$\alpha_C = \int_{I_C} \alpha_C(y) dy.$$

This utility rate has the following interpretation. Note that  $[q(x|y)p(x|z) - c]$  is the expected reward that agent  $z \in C_d$  receives from content of type  $x$  that is being produced by agent  $y \in C_s$ . Therefore in an economic setting  $[q(x|y)p(x|z) - c]$  is the amount (price) that agent  $z$  is willing to pay agent  $y$  for obtaining from agent  $y$  content of type  $x$ . In this sense, one interpretation of the utility rate of a content producer  $y \in C_s$  is that it reflects the revenue that  $y$  would obtain for the content that  $y$  produces in  $C$ . An alternative interpretation, and the one we adopt in this paper, is that the utility rate of content producer  $y \in C_s$  reflects the reputation, or “reputation score”, of agent  $y$  in the community  $C$ , i.e. it captures how beneficial the contributions of a content producer  $y$  are for the community  $C$ .

An interesting aspect of this model is that that total utility rates over all agents for content consumption and content production is the same, i.e. one can show that

$$\int_{C_d} U_C^{(d)}(y) dy = \int_{C_s} U_C^{(s)}(y) dy = \left( \int_{\mathcal{R}} P_C(x) Q_C(x) dx \right) - \alpha_C \beta_C c,$$

where

$$\beta_C = \int_{\mathcal{R}} \beta_C(x) dx.$$

#### 4.1 Interpretation of the Model

In our model we assume that there are agents that produce content as well as agents that consume content. There are two possible ways to interpret/apply this model. First we can assume that there exist two sets of agents: one set of agents that exclusively consumes content, and a second set of agents that exclusively produces content. In this model, the set of agents that produce and consume content are disjoint. Second, we can assume that there exists a single set of agents where each agent both consumes and produces content, and agents are best at producing the content type which they are most interested in.

For the model that we consider here, i.e. under the assumption that there exists an infinite number of agents as given by Assumption 2 both interpretation lead to identical results, i.e. our results in Section 6 are valid for both cases.

## 5 Community Structure and Nash Equilibrium

In the previous section we characterized the utility that agents obtain when consuming and producing content in a given community. In the following, we assume that there are several communities, and each agent can decide which community to join, and how much time/effort to allocate to produce and consume content in a community they join. Moreover, we assume that agents make these decisions in order to maximize their utility rates. For this situation we use a game-theoretic approach to characterize the community structures that emerge in an information network. In particular, we characterize the community structure that emerges using the concept of a Nash equilibrium. To do that we describe in the following how we define a community structure, and a Nash equilibrium, in an information network.

For our analysis, we characterize a community structure in an information network by a triplet  $(\mathcal{C}, \{\alpha_C(y)\}_{y \in \mathcal{R}}, \{\beta_C(\cdot|y)\}_{y \in \mathcal{R}})$  where  $\mathcal{C}$  is the set of communities  $C$  that exist in the structure, and

$$\alpha_C(y) = \{\alpha_C(y)\}_{C \in \mathcal{C}}$$

and

$$\beta_C(y) = \{\beta_C(\cdot|y)\}_{C \in \mathcal{C}}$$

indicate how individual agents consume and produce content in the different communities  $C = (C_d, C_s) \in \mathcal{C}$ . In the following, we require that for a given community structure  $(\mathcal{C}, \{\alpha_C(y)\}_{y \in \mathcal{R}}, \{\beta_C(\cdot|y)\}_{y \in \mathcal{R}})$ , we have for each community

$$C = (C_d, C_s) \in \mathcal{C}$$

that

$$\alpha_C(y) > 0, \quad y \in C_d,$$

and

$$\|\beta_C(y)\| > 0, \quad y \in C_s,$$

i.e. each agent  $y \in C_d$  allocates a strictly positive fraction of time  $\alpha_C(y)$  to community  $C$ , and each agent  $y \in C_s$  has a positive total rate  $\|\beta_C(y)\|$  in community  $C$ .

Furthermore, we require that the total fraction of time for content consumption over all communities is bounded by a constant  $E_p$ ,  $0 < E_p \leq 1$ , and we have that

$$\|\alpha_C(y)\| = \sum_{C \in \mathcal{C}} \alpha_C(y) \leq E_p, \quad y \in \mathcal{R},$$

where we use the convention that for a community  $C = (C_d, C_s)$  we have that

$$\alpha_C(y) = 0, \quad \text{if } y \notin C_d.$$

Similarly, we assume that the total content production rates of each agent can not exceed a given threshold  $E_q$ ,  $0 < E_q$ , and we have that

$$\|\beta_C(y)\| = \sum_{C \in \mathcal{C}} \|\beta_C(\cdot|y)\| \leq E_q, \quad y \in \mathcal{R},$$

where

$$\|\beta_C(\cdot|y)\| = \int_{\mathcal{R}} \beta_C(x|y) dx,$$

and we use the convention that for a community  $C = (C_d, C_s)$  we have that

$$\|\beta_C(\cdot|y)\| = 0, \quad \text{if } y \notin C_s.$$

These two requirements capture the intuition that each agent has limited resource (time) to consumer, and produce, content.

In the following, we are interested in a community structure where each agent belongs to at least one community. We refer to such a community structure as a covering community structure. More precisely, we use the following definition.

**Definition 1.** We call a community structure  $(\mathcal{C}, \{\alpha_C(y)\}_{y \in \mathcal{R}}, \{\beta_C(\cdot|y)\}_{y \in \mathcal{R}})$  a covering community structure if we have that

$$\cup_{C=(C_d, C_s) \in \mathcal{C}} C_d = \cup_{C=(C_d, C_s) \in \mathcal{C}} C_s = \mathcal{R}.$$

Using this definition, a Nash equilibrium is then given as follows. A community structure  $(\mathcal{C}^*, \{\alpha_{\mathcal{C}^*}(y)\}_{y \in \mathcal{R}}, \{\beta_{\mathcal{C}^*}(\cdot|y)\}_{y \in \mathcal{R}})$  is a Nash equilibrium if for all  $y \in \mathcal{R}$  we have that

$$\alpha_{\mathcal{C}^*}(y) = \arg \max_{\alpha_C(y): \|\alpha_C(y)\| \leq E_p} \sum_{C \in \mathcal{C}} \alpha_C(y) \int_{\mathcal{R}} [Q_C(x)p(x|y) - \beta_C(x)c] dx$$

and

$$\beta_{\mathcal{C}^*}(\cdot|y) = \arg \max_{\beta_C(\cdot|y): \|\beta_C(\cdot|y)\| \leq E_q} \sum_{C \in \mathcal{C}} \int_{\mathcal{R}} \beta_C(x|y)[q(x|y)P_C(x) - \alpha_C c] dx.$$

Note that under a Nash equilibrium each agent maximizes its own utility rate, i.e. a Nash equilibrium characterizes a community structure that would emerge in an information network where each agent tries to maximize its own utility.

In the following, we study whether there always exists a Nash equilibrium. Furthermore, if there exists a Nash equilibrium we want to characterize the community structure  $(\mathcal{C}^*, \{\alpha_{\mathcal{C}^*}(y)\}_{y \in \mathcal{R}}, \{\beta_{\mathcal{C}^*}(\cdot|y)\}_{y \in \mathcal{R}})$  under a Nash equilibrium.

For our analysis we make the following assumption on the relation of the processing cost  $c$  with respect to the functions  $f$  and  $g$  of Assumption 1.

**Assumption 3.** Let  $f$  and  $g$  be as given by Assumption 1, and let  $c$  be the processing cost as given in Eq. (1) and (2). Then we have that

$$f(0)g(0) - c > 0.$$

Note that if the assumption is not true then consumption utility under any community structure would either be equal to 0, or would be negative. As a result, this assumption excludes the (trivial) Nash equilibrium where agents do not join any information community.

## 6 Results

In this section we present the main results of our analysis of the above model. The proofs for the results are given in the appendix.

### 6.1 Existence of a Nash Equilibrium

Our first result establishes that there always exists a Nash equilibrium.

**Proposition 1.** *Let Assumption 1 - 3 be true, then there always exists a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_{\mathcal{C}^*}^*(y)\}_{y \in \mathcal{R}}, \{\beta_{\mathcal{C}^*}^*(\cdot|y)\}_{y \in \mathcal{R}})$  that is a covering community structure.*

Proposition 1 states that there always exists a Nash equilibrium that is a covering community structure, i.e. a community structure for which each agent is a member of at least one community. This result is an important result as it states that (under the above model) communities indeed emerge as an optimal structure to share content in an information network.

While Proposition 1 states that there always exists a Nash equilibrium, it does not provide a characterization of the community structure under a Nash equilibrium. In order to do that, we focus in the following on a particular family of Nash equilibria, and provide for this family a detailed characterization of the community structure under a Nash equilibrium.

To characterize the family of Nash equilibria we use the following definition.

**Definition 2.** We call a community  $C = (C_d, C_s)$  an interval community if we have that

$$C_d = C_s = I_C \subset \mathcal{R},$$

where  $I_C$  is an interval in  $\mathcal{R}$  with length  $|I_C|$  such that

$$0 < |I_C|.$$

The interval  $I_C$  can be closed or open, or be closed at one end and open at the other end.

For a given interval community  $C = (C_d, C_s) = (I_C, I_C)$ ,  $I_C \subset \mathcal{R}$ , we refer to  $|I_C|$  as the length of the interval, to

$$\text{mid}(I_C) = \frac{1}{|I_C|} \int_{I_C} y dy$$

as the midpoint of the interval, and to

$$L_C = \frac{|I_C|}{2}$$

as the half-length of the interval.

Given in interval community  $C = (I_C, I_C)$ , we refer to  $\text{mid}(I_C)$  as the center of interest of community  $C$ .

The next result states that there always exists a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  such that each community  $C \in \mathcal{C}^*$  is an interval community.

**Proposition 2.** Let Assumption 1 - 3 be true, then there exists a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  such that

(a) each community  $C = (C_d, C_s) \in \mathcal{C}^*$  is an interval community, i.e. we have that

$$C_d = C_s = I_C,$$

where  $I_C$  is an interval in  $\mathcal{R}$ .

(b) the set  $\{I_C\}_{C \in \mathcal{C}^*}$  is a set of mutually non-overlapping intervals that covers  $\mathcal{R}$ , i.e. we have that

$$I_C \cap I_{C'} = \emptyset, \quad C, C' \in \mathcal{C}^*, C \neq C',$$

and

$$\cup_{C \in \mathcal{C}^*} I_C = \mathcal{R}.$$

(c) there exists a constant  $L_I > 0$  such that

$$L_C = L_I, \quad C \in \mathcal{C}^*$$

where

$$2L_I < \min\{b, L\}$$

and  $b$  is the constant of Assumption 1.

Furthermore, for each community  $C = (I_C, I_C) \in \mathcal{C}^*$  we have that

(a)  $\alpha_C^*(y) = E_p, y \in I_C$ .

(b)  $\beta_C^*(x|y) = E_q \delta(x^*(y) - x)$ ,  $y \in I_C$  and  $x \in \mathcal{R}$ , where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_{I_C}(x).$$

and  $\delta$  is the Dirac delta function.

$$(c) U_C^{(d)}(u) > 0, y \in I_C.$$

$$(d) U_C^{(s)}(u) > 0, y \in I_C.$$

We provide the proof for Proposition 1 and 2 in Appendix I. Below we highlight some properties of the community structure given in Proposition 2.

Proposition 2 states that there always exists a Nash equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  consisting of interval communities that cover  $\mathcal{R}$ . Recall that in an interval community  $C = (C_d, C_s) = (I_C, I_C)$  we have that the content producers and content consumers cover the same interests, i.e. we have that

$$C_d = C_s = I_C.$$

This result captures that intuition that the content that is being produced and a consumed in a community should be “aligned” in the sense that the content producers in the community should have a high ability to produce content that is of interest to the content consumers.

Furthermore, Proposition 2 states that under a Nash equilibrium as given in Proposition 2, we have that in each community  $C = (I_C, I_C)$  a content producer  $y \in I_C$  focuses on producing a single type of content given by

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x).$$

As we show in Appendix C, exclusively producing content type  $x^*(y)$  indeed maximizes the content production utility of an agent. Moreover, experimental results suggest that this property indeed holds in real-life information networks. We will discuss this in more details in Subsection 6.6.

Finally, Proposition 2 states that the interval communities in Proposition 2 are non-overlapping, and each agent belongs to a single community. The fact that communities are non-overlapping seems not to be aligned with what one observes in real-life information networks. We will discuss this in more details in Section 7.

In the following we characterize in more details the properties of a Nash equilibrium as given in Proposition 2. In particular, we characterize the optimal content production rate of each agent under such a Nash equilibrium, the properties of the demand function and supply function in a community in under such a Nash equilibrium, as well as the resulting utility functions.

## 6.2 Optimal Content Type $x^*(y)$

Under the community structure given in Proposition 2, each content producer  $y \in I_C$  in a given community  $C \in \mathcal{C}^*$  produces the single content type given by

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_{I_C}(x).$$

In this subsection we characterize in more details the function  $x^*(y)$ ,  $y \in I_C$  for a given community  $C \in \mathcal{C}^*$ . In particular, we have the following result.

**Proposition 3.** *Let  $C = (C_d, C_s) = (I_C, I_C)$  be an interval community under a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2. Then the function  $x^*(y)$  given by*

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

is continuously differentiable and strictly increasing on  $I_C$ . Furthermore we have that

$$x^*(y) \in [y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y)), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)],$$

and

$$x^*(y) \in [\text{mid}(I_C), y] \cap \text{supp}(q(\cdot|y)), \quad y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C].$$

We provide a proof for Proposition 3 in Appendix C.

One important result of Proposition 3 is that the function  $x^*(y)$  is strictly increasing on  $I_C$ . This result implies that two different agents  $y, y' \in I_C$ ,  $y \neq y'$ , will produce different types of content, i.e. we have that

$$x^*(y) \neq x^*(y'), \quad y \neq y'.$$

This is an interesting result as it states that it is optimal for each agent to specialize on producing the content type  $x^*(y)$  that is not produced by any other agent in the community, i.e. agent  $y$  is the unique expert for content of type  $x^*(y)$  in the information community that  $y$  belongs to.

Another important aspect of Proposition 3 is that the function  $x^*(y)$  has the property that

$$x^*(y) \in [y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y)), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)],$$

and

$$x^*(y) \in [y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y)), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)].$$

This result states that each agent  $y \in I_C$  produces content that is either equal to their center of interest  $y$ , or closer to the center of interest  $\text{mid}(I_C)$  of the community  $C$  than their center of interest  $y$ . To get a more detailed understanding of how agents adapt the type of content that they produce towards the center of interest  $\text{mid}(I_C)$  of the community  $C$ , we next study the function  $\Delta^*(y)$  given by

$$\Delta^*(y) = \|y - x^*(y)\|, \quad y \in I_C.$$

The function  $\Delta^*(y)$  characterizes the absolute value of the “displacement” of the optimal content  $x^*(y)$  that agent  $y$  produces, and content  $y$  that the agent is best at producing which is equal to content type  $y$ . Or in other words, the function  $\Delta^*(y)$  characterizes by how much an agent  $y$  adapts its content  $x^*(y)$  towards the center of interest of the community  $C$ , i.e. by how much agent  $y$  produces content  $x^*(y)$  that is closer to the center of interest  $\text{mid}(I_C)$  of the community than its own center of interest given by  $y \in I_C$ .

In addition, the function  $\Delta^*(y)$  can be used to characterize the quality of the optimal content  $x^*(y)$  that agent  $y$  produces as we have that

$$q(x^*(y)|y) = g(\|y - x^*(y)\|) = g(\Delta^*(y)).$$

We have the following result for the function  $\Delta^*(y)$ .

**Proposition 4.** *Let  $C = (C_d, C_s) = (I_C, I_C)$  be an interval community under a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2. Then the function  $\Delta^*(y)$  given by*

$$\Delta^*(y) = \|y - x^*(y)\|, \quad y \in I_C,$$

is strictly decreasing on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C)]$ , and strictly increasing on  $[\text{mid}(I_C), \text{mid}(I_C) + L_C]$  with

$$\Delta^*(\text{mid}(I_C)) = 0.$$

We provide a proof for Proposition 4 in Appendix E.

Proposition 4 provides several interesting insights how agents adapt the type of content they produce towards the community  $C$  they belong to.

First, Proposition 4 states that

$$\Delta^*(y) > 0, \quad y \in I_C \setminus \{mid(I_C)\},$$

and hence

$$x^*(y) \neq y, \quad y \in I_C \setminus \{mid(I_C)\}.$$

This implies that agent  $y \in I_C \setminus \{mid(I_C)\}$  does not produce the type of content that it can produce with the highest quality  $q(x|y)$ , i.e. the content that is equal to their center of interest  $y$ . But agent  $y \in I_C \setminus \{mid(I_C)\}$  does adapt the content it produces towards the center of interest of community  $C$ , i.e.  $y$  produces content of type  $x^*(y)$  that is closer to the center of the community  $mid(I_C)$  than their own center of interest  $y$ .

Moreover, Proposition 4 states that  $\Delta^*(y)$  is strictly decreasing on  $[mid(I_C) - L_C, mid(I_C))$ , and strictly increasing on  $(mid(I_C), mid(I_C) + L_C]$ . This implies that the further away an agent is from the center of interest  $mid(I_C)$ , the more it will “adapt” the content it produces towards to the center of interest  $mid(I_C)$  of the community  $C$ .



### 6.3 Properties of Demand $P_C(x)$ and Supply Function $Q_C^*(x)$

We next we study the properties of the demand function  $P_C(x)$  given by

$$P_C(x) = \int_{I_C} \alpha_C(y)p(x|y)dy, \quad x \in \mathcal{R}, y \in I_C,$$

and the optimal supply function  $Q_C^*(x)$  given by

$$Q_C^*(x) = \int_{I_C} \beta_C^*(x|y)q(x|y)dy, \quad x \in \mathcal{R}, y \in I_C,$$

for an interval community  $C = (C_d, C_s) = (I_C, I_C) \in \mathcal{C}^*$  under a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2.

We first characterize the demand function  $P_C(x)$ .

**Proposition 5.** *Let  $C = (C_d, C_s) = (I_C, I_C)$  be an interval community under a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2. Then the demand function  $P_C(x)$  is given by*

$$P_C(x) = E_p \int_{y \in I_C} p(x|y)dy, \quad x \in \mathcal{R}, y \in I_C.$$

Furthermore, the function  $P_C(x)$  has the following properties,

- (a)  $P_C(x)$  is symmetric around  $y_0 = \text{mid}(I_C)$ .
- (b)  $P_C(x)$  is continuously differentiable on  $\mathcal{R}$ .
- (c)  $P_C(x)$  is strictly increasing on the interval  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ , and strictly decreasing on the interval  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ .
- (d)  $P_C(x)$  is twice continuously differentiable and strictly concave in  $x$  on the interval  $[y_0 - L_C, y_0 + L_C]$ .
- (e)  $\text{mid}(I_C) = \arg \max_{x \in \mathcal{R}} P_C(x)$

We provide a proof for Proposition 5 in Appendix B.

The results of Proposition 5 are intuitive. First, the result that

$$\text{mid}(I_C) = \arg \max_{x \in \mathcal{R}} P_C(x)$$

states that the content type that is the most popular is the content type that is identical to the center of interest  $\text{mid}(I_C)$  of the interval community  $C$ . Moreover, the result that  $P_C(x)$  is strictly increasing on  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ , and strictly decreasing  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ , implies that the further a content type  $x$  is from the center of interest  $\text{mid}(I_C)$  of community  $C$ , the less popular it is. Or in other words, Proposition 5 states that 'fringe' content, i.e. content that is far away from the center of interest of the community, is less in demand/popular compared with content that is close to the center of interest.

Next we characterize the properties of the optimal content supply function  $Q_C^*(x)$ .

**Proposition 6.** *Let  $C = (C_d, C_s) = (I_C, I_C)$  be an interval community under a Nash equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2. Then the optimal content supply function  $Q_C^*(x)$  is given by*

$$Q_C^*(x) = E_q \int_{I_C} \delta(x - x^*(y)) q(x^*(y)|y) dy, \quad x \in \mathcal{R}, y \in I_C,$$

where  $\delta(\cdot)$  is the Dirac delta function, and

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_C(x), \quad y \in I_C.$$

Furthermore we have that

a) the support  $\text{supp}(Q_C^*(x))$  of  $Q_C^*(x)$ ,  $x \in \mathcal{R}$  is given by

$$\text{supp}(Q_C^*(x)) = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*]$$

with

$$0 < L_C^* < L_C,$$

b)  $Q_C^*(x) = \beta_C^*(x) q(x|s(x))$ ,  $x \in I_C^*$ , where  $\beta_C^*(x)$ ,  $x \in I_C^*$ , is as defined in Lemma 14, and

c)  $Q_C^*(x)$  is continuous, and symmetric with respect to  $\text{mid}(I_C)$ , on  $I_C^*$ .

We provide a proof for Proposition 6 in Appendix F.

Proposition 6 states that

$$\text{supp}(Q_C^*(x)) = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*]$$

where

$$0 < L_C^* < L_C,$$

This result implies that the content type that is being produced by agents  $y \in C_s = I_C$  in the community  $C$  is a strict subset of  $I_C$ , i.e. the content that is being produced does not cover the center of interests of all agents  $y \in C_d = I_C$ . Or in other words, the content that is being produced is concentrated around to center of interest  $\text{mid}(I_C)$  of the community, and not spread out over the interval  $I_C$ . This implies that under a Nash equilibrium there is no overlap between the content that is being produced in different communities, i.e. each community produces a distinct set of content types.

Unlike the properties of the function  $P_C(x)$  as given by Proposition 5, the function  $Q^*(x)$  is not necessarily non-decreasing on  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ , and non-increasing  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ . Obtaining these properties would require additional assumptions on the functions  $f$  and  $g$ .

#### 6.4 Properties of Utility Functions $U_C^{(d)}(y)$ and $U_C^{(s)}(y)$

We next study the properties of the utility rate function for content consumption  $U_C^{(d)}(y)$ , and the utility rate function for content production  $U_C^{(s)}(y)$  for an interval community  $C = (C_d, C_s) = (I_C, I_C) \in \mathcal{C}^*$  under a Nash equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2.

We first study the properties of the utility rates for content consumption.

**Proposition 7.** *Let  $C = (C_d, C_s) = (I_C, I_C)$  be an interval community under a Nash equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2, and let  $Q_C^*(x)$  be the corresponding content supply function as given in Proposition 6. Then the maximal utility rate function for content consumption  $U_C^{(d)}(y)$  in community  $C$  by an agent  $y \in \mathcal{R}$  is given by*

$$U_C^{(d)}(y) = \max \left\{ 0, F_C^{(d)}(y) \right\}, \quad y \in \mathcal{R},$$

where

$$F_C^{(d)}(y) = E_p \left[ \int_{\mathcal{R}} p(x|y) Q_C^*(x) dx - 2L_C E_q c \right], \quad y \in \mathcal{R},$$

and we have that

$$\text{mid}(I_C) = \arg \max_{y \in \mathcal{R}} U_C^{(d)}(y).$$

Furthermore, the function  $U_C^{(d)}(y)$ ,  $y \in \mathcal{R}$ , has the properties that

a)  $U_C^{(d)}(y)$  is symmetric with respect to  $\text{mid}(I_C)$ , and we have that

$$\frac{d}{dy} U_C^{(d)}(\text{mid}(I_C)) = 0.$$

b)  $U_C^{(d)}(y)$  is strictly increasing on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C))$ , and we have that

$$\frac{d}{dy} U_C^{(d)}(y) > 0, \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)),$$

c)  $U_C^{(d)}(y)$  is strictly decreasing on  $(\text{mid}(I_C), \text{mid}(I_C) + L_C]$ , and we have that

$$\frac{d}{dy} U_C^{(d)}(y) < 0, \quad y \in (\text{mid}(I_C), \text{mid}(I_C) + L_C],$$

b)  $U_C^{(d)}(y)$  is non-decreasing on  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ ,

c)  $U_C^{(d)}(y)$  is non-increasing on  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ .

We provide a proof for Proposition 7 in Appendix G.

Note that the function  $U_C^{(d)}(y)$  in Proposition 7 is not only defined for agents  $y \in C_s = I_C$ , but for all agents  $y \in \mathcal{R}$ . For agents  $y \notin I_C$ , the function  $U_C^{(d)}(y)$  characterizes the maximal utility rate for content consumption that agent  $y$  would obtain in community  $C$ . Understanding the properties of the function  $U_C^{(d)}(y)$  is therefore important in establishing that a community  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given in Proposition 2 is indeed a Nash equilibrium.

The results of Proposition 7 are intuitive. First, the result that

$$\text{mid}(I_C) = \arg \max_{y \in \mathcal{R}} U_C^{(d)}(y)$$

states that the agent at the center of interest  $\text{mid}(I_C)$  of the interval community  $C$  obtains the highest utility for content consumption. Moreover, Proposition 7 states that the further away an agent  $y$  is from the center of interest  $\text{mid}(I_C)$  of the interval community  $C$ , the smaller a utility for content consumption agent  $y$  receives.

As the interval communities given in Proposition 2 are of equal length, Proposition 7 implies that the community structure  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2 is indeed a Nash equilibrium for agents that consume content. Indeed as the utility for content consumption decreases the further away an agent is from the center of interest  $mid(I_C)$  of an interval community, it is best for an agent  $y \in \mathcal{R}$  to consume content in the interval community  $C$  for which we have that

$$y \in I_C.$$

More specifically as the interval communities in Proposition 2 are of equal length, Proposition 7 implies that the agent “on the border” between two communities  $C$  and  $C'$ ,  $C, C' \in \mathcal{C}^*$ , obtains the same utility for content consumption in both communities  $C$  and  $C'$ , and hence is indifferent on which community to join. All the other agents  $y \in C$ , will obtain a higher utility in community  $C$  compared with community  $C'$ ; and vice versa.

We next study the properties of the utility rates for content production  $U_C^{(s)}(y)$ .

**Proposition 8.** *Let  $C = (C_d, C_s) = (I_C, I_C)$  be an interval community under a Nash equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2, and let  $P_C(x)$ ,  $x \in \mathcal{R}$ , be the corresponding content demand function as given in Proposition 5. Then the maximal utility rate function for content production  $U_C^{(s)}(y)$  for community  $C$  by an agent  $y \in \mathcal{R}$  is given by*

$$U_C^{(s)}(y) = \max \left\{ 0, F_C^{(s)}(y) \right\}, \quad y \in \mathcal{R},$$

where

$$F_C^{(s)}(y) = E_q \left[ q(x^*(y)|y) P_C(x^*(y)) - 2E_p L_C c \right],$$

and we have that

$$\text{mid}(I_C) = \arg \max_{y \in \mathcal{R}} U_C^{(s)}(y).$$

Furthermore, the function  $U_C^{(s)}(y)$ ,  $y \in \mathcal{R}$ , has the properties that

- a)  $U_C^{(s)}(y)$  is symmetric with respect to  $y_0 = \text{mid}(I_C)$  on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C) + L_C]$ , and we have that

$$\frac{d}{dy} U_C^{(s)}(\text{mid}(I_C)) = 0.$$

- b)  $U_C^{(s)}(y)$  is strictly increasing on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C))$ , and we have that

$$\frac{d}{dy} U_C^{(s)}(y) > 0, \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)).$$

- c)  $U_C^{(s)}(y)$  is strictly decreasing on  $(\text{mid}(I_C), \text{mid}(I_C) + L_C]$ , and we have that

$$\frac{d}{dy} U_C^{(s)}(y) < 0, \quad y \in (\text{mid}(I_C), \text{mid}(I_C) + L_C].$$

- b)  $U_C^{(s)}(y)$  is non-decreasing on  $(\text{mid}(I_C) - L, \text{mid}(I_C) - L_C]$ .

- c)  $U_C^{(s)}(y)$  is non-increasing on  $[\text{mid}(I_C) + L_C, \text{mid}(I_C) + L]$ .

We provide a proof for Proposition 8 in Appendix H.

Similar to Proposition 7, as we have that

$$\text{mid}(I_C) = \arg \max_{y \in \mathcal{R}} U_C^{(s)}(y)$$

Proposition 8 states that the agent at the center of interest  $\text{mid}(I_C)$  of the interval community  $C$  obtains the highest utility for content production. Moreover, Proposition 8 states that the further away an agent  $y$  is from the center of interest  $\text{mid}(I_C)$  of the interval community  $C$ , the smaller a utility for content production agent  $y$  receives.

As the interval communities given in Proposition 2 are of equal length, Proposition 8 implies that the community structure  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 8 is a Nash equilibrium for agents that produce content. Indeed as the utility for content production decreases the further away an agent is from the center of interest  $\text{mid}(I_C)$  of an interval community, it is best for an agent  $y \in \mathcal{R}$  to produce content in the interval community  $C$  for which we have that

$$y \in I_C.$$

More specifically, as the interval communities in Proposition 2 are of equal length, Proposition 8 implies that the agent ‘‘on the border’’ between two communities  $C$  and  $C'$ ,  $C, C' \in \mathcal{C}^*$ , obtains the same utility for content consumption in both communities  $C$  and  $C'$ , and hence is indifferent on which community to join. All the other agents  $y \in C$ , will obtain a higher utility in community  $C$  compared with community  $C'$ ; and vice versa.

By combining Proposition 7 and 8, we obtain that the community structure  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  given in Proposition 2 is indeed a Nash equilibrium, provided that the length of the intervals is not too long and all agents  $y \in C_s = I_C$  and  $y \in C_d = I_C$ ,  $C \in \mathcal{C}^*$  indeed obtain a positive utility in community  $C$ . We provide a formal presentation of this argument in Appendix I where we prove Proposition 2.

## 6.5 Maximal Community Size

Finally, we characterize the maximal size of an interval community under a Nash equilibrium as given by Proposition 2. Intuitively, in order for a community structure  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given in Proposition 2 to be a Nash equilibrium, the lengths of the intervals can not be too long, i.e. the length  $L_I$  given in Proposition 2 can not be too long, as otherwise the utilities of some of the agents would become negative. This intuition is indeed true. In particular we have the following results which shows that the maximal size depends on the content processing cost  $c$ , as well as the function  $f$  and  $g$  of Assumption 1.

**Proposition 9.** *If  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  is a Nash equilibrium as given by Proposition 2 and  $L_I$  is the length of the intervals  $\{I_C\}_{C \in C^*}$ , then we have that*

$$L_I < \max \left\{ l \in [0, L] \mid \int_0^l (g(0)f(x) - c)dx \geq 0 \right\},$$

where the functions  $f$  and  $g$  are as given by Assumption 1.

We provide a proof for Proposition 9 in Appendix J.

Proposition 9 states that the higher the processing cost  $c$ , the smaller  $g(0)$  and the faster the function  $f(x)$  decays, the smaller the size of a community will be. Note that a fast decay in  $f$  implies that agents are interested in a very narrow range of content topics; and hence the result of Proposition 9 is a property that one would expect for information communities.

## 6.6 Discussion

In this section we presented the results that we obtained for the the existence, and characteristics, of community structures that are a Nash equilibrium. An interesting aspect of the obtained result is that they indeed provide insights into properties of communities in real-life information networks. In addition, the obtained result provide insight into structure properties of information community, that might can be used to design new algorithms to analyze information networks and communities.

Proposition 2 states that each content producer in a community focuses on generating exactly one type of content, i.e we have that that an agent  $y \in C_s = I_C$  produces the content unique content type  $x^*(y)$  given by

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C.$$

This is an interesting results as this property/behavior has indeed been observed in real-life social networks. In particular, Zadeh, Goel and Munagala provide experimental results using data obtain from Twitter that shows Twitter users produce content on a very narrow set of topics, and consume content on a large set of topics [2]. The interpretation of this experimental result based on the analysis presented in this paper is that this behavior is indeed optimal, and that the topics for which a given Twitter user produces content is the optimal content for this user to produce. In particular, based on the result in Proposition 2, the topic(s) for which Twitter users produce content in a given community should be a very small subset of all the topics that are being consumed in this community. In this sense, the Proposition 2 provides a theoretical explanation for this property/behavior that has been observed in real-life social networks.

Proposition 3 and Proposition 4 state that agents adapt the content they produce towards center of interest of a community they belong to. This is an interesting result, as it provides insight regarding the issue of “homophily versus adaption” in social networks. One aspect of communities in social networks is that members of the community tend to have similar interests, and behave similarly. One question is whether is due to homophily, i.e. do individuals form a community because they have the same interest and behave similarly, or whether this is due to adaption, i.e. is due that through the interaction in a community its members develop similar interests and behave similarly. The results of this paper suggest that both mechanism are at work in information communities. In particular, homophily (similar interests) prompts individuals to join the same community. This is captured by the result of Proposition 3 which states the communities of a Nash equilibrium as given in Proposition 2 consists of interval communities, and hence the center of interests of agents in a given community belong to the same interval and are close together. Moreover, when an individual joins a community then adaptation prompts the individual to conform (adapt) away from their center of interest towards the center of interest of the community. This is captured by the result of Proposition 3 which states that an agent in community  $C$  does not produce content  $y$  for which they have the best ability, i.e. content type that is the center of interest of the agent, but agent  $y$  adapts the content type  $x^*(y)$  that is produces towards the center of interest  $mid(I_C)$  of the community.

Proposition 5 states that the content type that is the most popular is the content type that is identical to the the center of interest of the community. Moreover, it states that the further away a content type is from the center of interest of a community, the less popular it is in the community. This result is interesting as it confirms in a formal way the intuition that communities have a “center of interest”, i.e. there is a focus (center of interest) of the community and the closer a type of content (topic) is to that center of interest, the more popular (in demand) the content will be. While this property of a community may seem “intuitive obvious”, it is interesting to observe that this property is indeed recovered (confirmed) using the proposed model.

Proposition 6 implies that under a Nash equilibrium there is no overlap between the content that is being produced in different communities, i.e. each community produces a distinct set of content types. This is an interesting result as it suggest that each information community can be identified by unique “core content” that is only produced in this community. In particular, this result suggests that each real-life community has a “core interest” that is unique to the community, and hence identifies this community. Moreover, it seems that this property should be useful when trying to discover (identify) communities in information networks, i.e. one should be able to exploit this property to potential design more efficient algorithms to discover communities in information networks. How, and whether, this is possible is future research.

Finally, Proposition 7 and 8 state that the agent at the center of interest of a community obtain the highest utility rate for content consumption and production; and the further away an agent  $y$  is from the center of interest of the community, the smaller a utility it will receive. This result suggest that there is a “natural” way to rank members in a community, where a member is ranked higher if it receives a higher utility. In particular, if it is possible to observe the utilities that community members receive, then it is indeed possible to rank community members in this way. How, and when, it is possible to observe (estimate) utilities that community members receive, and how this information can be used to design algorithms to analysis information communities, is future research.



## 7 Conclusions

We proposed a mathematical model to study communities in information networks. The model is intentionally kept as simple as possible to allow for a formal analysis. We used the model to derive properties of the community structure and communities under a Nash equilibrium. In particular, we show that there always exists a Nash equilibrium. Moreover, we focused on Nash equilibria where that consist of interval communities, i.e. the center of interest of agents that belong to the community all lie within an interval on  $\mathcal{R}$ . For such Nash equilibria, we characterized the content type that agents produce in a given community, the content demand and supply functions, as well as the utility functions for content consumption and production. Finally, we derive a bound on the maximal size that an interval community under a Nash equilibrium can have. One interesting question is whether all Nash equilibria under the proposed model consist of interval communities of equal length as given in Proposition 2.

An interesting direction for future research is to investigate whether the model can be used to study additional properties of information communities. In particular, one direction of future research that we are working on is to investigate whether the proposed model can be used to study and characterize how content is being filtered and distributed within an information community.

Another interesting future research is to study whether the family of Nash equilibria considered in this paper are the only community structures that can emerge in a Nash equilibrium, or whether there are other Nash equilibria. Or in other words, to study whether a Nash equilibrium always consists of a set of interval communities of equal length, or whether it is possible to have a Nash equilibrium where the communities are of different sizes.

While the model captures important properties that have been observed in real-life information networks as discussed in the previous section, the model has limitations and fails to capture some properties of information communities. For example, it fails to capture that communities in information networks tend to have a hierarchical and overlapping structure. The reason for this is that the current model and analysis is limited to a one-dimensional metric space (for characterizing the content types) with homogeneous agents whose interest are uniformly distributed. Extending the analysis to more general settings of the is interesting future research.

## References

- [1] Anthony Bonato, Jeannette Janssen, and Pawel Pralat. Geometric protean graphs. *CoRR*, abs/1111.0207, 2011.
- [2] Reza Bosagh Zadeh, Ashish Goel, Kamesh Munagala, and Aneesh Sharma. On the precision of social and information networks. In *Proceedings of the first ACM conference on Online social networks*, pages 63–74. ACM, 2013.
- [3] Wei Chen, Zhenming Liu, Xiaorui Sun, and Yajun Wang. A game-theoretic framework to identify overlapping communities in social networks. *Data Mining and Knowledge Discovery*, 21(2):224–240, 2010.
- [4] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75 – 174, 2010.
- [5] Mangesh Gupte, MohammadTaghi Hajiaghayi, Lu Han, Liviu Iftode, Pravin Shankar, and Raluca M Ursu. News posting by strategic users in a social network. In *Internet and Network Economics*, pages 632–639. Springer, 2009.
- [6] Nidhi Hegde, Laurent Massoulié, and Laurent Viennot. Self-organizing flows in social networks. In *Structural Information and Communication Complexity*, pages 116–128. Springer, 2013.
- [7] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs: An approach to modeling networks. *J. Mach. Learn. Res.*, 11:985–1042, March 2010.
- [8] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

## A Feasible Communities

In this appendix we introduce additional definitions that we use in our analysis.

To do that, we first note that for a given a community  $C = (C_d, C_s)$  in a community structure  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  that is a Nash equilibrium, we have that

$$U_C^{(d)}(y) = \alpha_C(y) \int_{\mathcal{R}} [p(x|y)Q_C^*(x) - \beta_C^*(x)c] \geq 0, \quad y \in C_d,$$

where

$$Q_C^*(x) = \int_{C_s} \beta_C^*(x|y)q(x|y)dy$$

and

$$\beta_C^*(x) = \int_{C_s} \beta_C^*(x|y)dy,$$

as well as

$$U_C^{(s)}(y) = \int_{\mathcal{R}} \beta_C^*(x|y) [q(x|y)P_C(x) - \alpha_C c] \geq 0, \quad y \in C_s,$$

$$P_C(x) = \int_{C_d} \alpha_C^*(y)p(x|y)dy$$

and

$$\alpha_C^* = \int_{C_d} \alpha_C^*(y)dy,$$

i.e. the utilities of all agents  $y \in C_d$  and  $y \in C_s$  are non-negative. Indeed if this is not the case, and the utility obtained by an agent in the community is negative, then the agent is better off by not participating in the community at all which results in an utility rate equal to 0. Therefore, a necessary condition for a community  $C$  with allocation functions  $\{\alpha_C(y)\}_{y \in C_d}$  and  $\{\beta_C(\cdot|y)\}_{y \in C_s}$  to be part of a Nash equilibrium is that it is all agents obtain a non-negative utility in the community  $C$ . For our analysis we use a slightly stronger condition, and we require that utilities of all agents are strictly positive. We call such communities feasible communities. Note that all communities under the structure  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2 have the property that the utilities that agents receive are strictly positive.

More precisely, we have the following definition.

**Definition 3.** Consider a community  $C = (C_d, C_s)$  and allocations  $\{\alpha_C(y)\}_{y \in C_s}$  and  $\{\beta_C(\cdot|y)\}_{y \in C_s}$  such that

$$\alpha_C(y) > 0, \quad y \in C_d,$$

and

$$\int_{\mathcal{R}} \beta_C(x|y)dx > 0, \quad y \in C_s.$$

Then we call the community  $C$  a feasible community under  $\{\alpha_C(y)\}_{y \in C_d}$  and  $\{\beta_C(\cdot|y)\}_{y \in C_s}$ , if we have that

$$U_C^{(d)}(y) = \alpha_C(y) \int_{\mathcal{R}} [p(x|y)Q_C(x) - \beta_C(x)c] dx > 0, \quad y \in C_d,$$

where

$$Q_C(x) = \int_{C_s} \beta_C(x|y)q(x|y)dy$$

and

$$\beta_C(x) = \int_{C_s} \beta_C(x|y)dy,$$

as well as

$$U_C^{(s)}(y) = \int_{\mathcal{R}} \beta_C(x|y) [q(x|y)P_C(x) - \alpha_C c] dx > 0, \quad y \in C_s,$$

where

$$P_C(x) = \int_{C_d} \alpha_C(y)p(x|y)dy$$

and

$$\alpha_C = \int_{C_d} \alpha_C(y)dy.$$

## B Proof of Proposition 5

In this appendix we prove Proposition 5 which characterizes the properties of the demand function  $P_C(x)$  of an interval community  $C$  in a Nash equilibrium as given by Proposition 2.

Recall that for an interval community  $C = (C_d, C_s)$  with  $C_d = C_s = I_C \subset \mathcal{R}$ , the demand function  $P_C(x)$  given by

$$P_C(x) = \int_{I_C} \alpha_C(y) p(x|y) dy, \quad x \in \mathcal{R}.$$

Furthermore, recall that for a given interval  $I_C \subset \mathcal{R}$  we denote with  $L_C$  the half-length of the interval  $I_C$ , i.e. we have that

$$L_C = \frac{|I_C|}{2},$$

and we use  $\text{mid}(I_C)$  to denote the midpoint of an interval, i.e. we have that

$$\text{mid}(I_C) = \frac{1}{|I_C|} \int_{I_C} y dy.$$

Also recall that the metric space  $\mathcal{R}$  that we use in our analysis is given by an interval  $[-L, L] \in \mathbb{R}$ ,  $L > 0$ , with the torus metric as defined in Section 3.2.

We then have the following result.

**Lemma 1.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . If we have that that*

$$\alpha_C(y) = E_p, \quad y \in C_d,$$

and

$$L_C < \frac{L}{2},$$

then the demand function

$$P_C(x) = E_p \int_{I_C} p(x|y) dy, \quad x \in \mathcal{R}, y \in I_C,$$

has the properties that

- (a)  $P_C(x)$  is symmetric around  $y_0 = \text{mid}(I_C)$  on  $(-L, L)$ .
- (b)  $P_C(x)$  is continuously differentiable on  $\mathcal{R}$ .
- (c)  $P_C(x)$  is twice continuously differentiable and strictly increasing on  $[y_0 - L, y_0)$ , and strictly decreasing on the interval  $(y_0, y_0 + L)$ .
- (d)  $P_C(x)$  is strictly concave on  $[y_0 - L_C, y_0 + L_C]$ .
- (e)  $y_0 = \arg \max_{x \in \mathcal{R}} P_C(x)$

Note that Proposition 5 directly follows from Lemma 1 as the communities under a Nash equilibrium as given by Proposition 2 satisfy the assumptions made in Lemma 1.

*Proof.* In order to prove Lemma 1, without loss of generality we can assume that the interval  $I_C$  is given by

$$I_C = [-L_C, L_C], \quad L_C > 0.$$

Using the fact that by assumption we have that

$$\alpha_C(y) = E_p > 0, \quad y \in C_d = I_C,$$

it follows that

$$P_C(x) = E_p \int_{I_C} p(x|y) dy = E_p \int_{-L_C}^{L_C} f(|y - x|) dy, \quad (3)$$

where the function  $f$  is as given in Assumption 1.

We first show that  $P_C(x)$  is symmetric around  $y_0 = \text{mid}(I_C) = 0$ . For this we note that for

$$x \in [0, L)$$

we have that

$$\begin{aligned} \int_{-L_C}^{L_C} f(\|y - x\|)dy &= \int_{y=-L_C}^{L_C} f(\|-y + x\|)dy \\ &= - \int_{y=L_C}^{-L_C} f(\|y + x\|)dy \\ &= \int_{-L_C}^{L_C} f(\|y + x\|)dy. \end{aligned}$$

Combining the above result with Eq. (3) it follows that

$$P_C(x) = P_C(-x),$$

and hence  $P_C(x)$  is symmetric around  $y_0 = \text{mid}(I_C) = 0$  on  $(-L, L)$ .

Next we show that  $P_C(x)$  is strictly increasing on the interval  $[-L, 0)$ , and strictly decreasing on the interval  $(0, L)$ . For this we let

$$s = y - x, \quad y \in I_C, x \in (-L, L).$$

Using this definition, we obtain that

$$P_C(x) = E_p \int_{-L_C-x}^{L_C-x} f(\|s\|)ds$$

and

$$\frac{d}{dx} P_C(x) = -E_p f(\|L_C - x\|) + E_p f(\|L_C + x\|). \quad (4)$$

As the function  $\|L_C + x\|$ ,  $x \in \mathcal{R}$ , is continuous in  $x$ , and by Assumption 1 the function  $f$  is a continuous function on  $[0, L]$ , it follows that  $P_C(x)$  is continuously differentiable in  $x$  on  $\mathcal{R}$ .

Note that for  $x = 0$  we have that

$$\frac{d}{dx} P_C(x = 0) = 0.$$

Using the assumption that

$$L_C < \frac{L}{2},$$

we have that

$$\|L_C - x\| < \|L_C + x\|, \quad x \in (0, L).$$

Combining this result with Assumption 1 which states that the function  $f$  is strictly decreasing on  $[0, L]$ , we obtain that

$$f(\|L_C - x\|) > f(\|L_C + x\|), \quad x \in (0, L],$$

and

$$\frac{d}{dx} P_C(x) < 0, \quad x \in (0, L).$$

It then follows that  $P_C(x)$  is strictly decreasing on the interval  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ .

Using the same argument, we can show that

$$\frac{d}{dx} P_C(x) > 0, \quad x \in [-L, 0),$$

and  $P_C(x)$  is strictly increasing on the interval  $[-L, 0)$ .

Next we show that  $P_C(x)$  is twice continuously differentiable and strictly concave on  $[-L_C, L_C]$ . Using Eq. (4) and the assumption that

$$L_C < \frac{L}{2},$$

we obtain that

$$\frac{d^2}{dx^2} P_C(x) = E_p f'(|L_C - x|) + E_p f'(|L_C + x|), \quad x \in [-L_C, L_C].$$

By Assumption 1, we have that  $f$  is continuously differentiable on  $[0, L]$ , it follows that  $P_C(x)$  is twice continuously differentiable on  $[-L_C, L_C]$ . Furthermore, by Assumption 1 the function  $f$  is strictly decreasing on  $[0, L]$ , and we obtain that

$$\frac{d^2}{dx^2} P_C(x) = E_p f'(|L_C - x|) + E_p f'(|L_C + x|) < 0, \quad x \in [-L_C, L_C], \quad (5)$$

and  $P_C(x)$  is strictly concave on  $[-L_C, L_C]$ .

Finally, combining results the that  $P_C(x)$  is strictly increasing on  $(-L, 0)$  and strictly decreasing on  $(0, L)$  with the fact that  $P_C(x)$  is continuous, it follows that

$$\text{mid}(I_C) = 0 = \arg \max_{x \in \mathcal{R}} P_C(x).$$

□

## C Optimal Content Production

In this appendix we characterize the optimal content production rate under a Nash equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2. In particular, we show that it is optimal for an agent to produce a single content type  $x^*(y)$  as indicated by Proposition 3.

### C.1 Properties of the Optimal Content Production Rate $\beta_C^*(x|y)$

We start out by characterizing the optimal content production rate  $\beta_C^*(x|y)$  of an agent  $y$  in an interval community. More precisely, given an interval community  $C = (C_d, C_s)$  such that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ , we characterize how an agent  $y \in C_s = I_C$  optimally allocates its content production rate  $\beta_C(\cdot|y)$  in community  $C$ . Recall that the optimal allocation  $\beta_C^*(\cdot|y)$  of an agent  $y \in C_s$  in the community  $C$  is given by

$$\beta_C^*(\cdot|y) = \arg \max_{\beta(\cdot|y): \|\beta(\cdot|y)\| \leq \beta_C(y)} \int_{\mathcal{R}} \beta_C(x|y) [q(x|y)P_C(x) - \alpha_C c] dx, \quad (6)$$

where  $\beta_C(y) > 0$  denotes the maximal production rate of agent  $y$  can allocate to community  $C$ , and the functions  $P_C(x)$  and  $\alpha_C$  are given by

$$P_C(x) = \int_{I_C} \alpha_C(y) p(x|y) dy$$

and

$$\alpha_C = \int_{I_C} \alpha_C(y) dy.$$

Note that if for given agent  $y \in I_C$  we have that

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] \leq 0,$$

then the optimal production rate  $\beta_C^*(\cdot|y)$  of agent  $y$  is given by

$$\beta_C^*(x|y) = 0, \quad x \in \mathcal{R}.$$

As a result we focus in the following on agents  $y \in I_C$  for which we have that

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] > 0.$$

For this case we show that it is optimal for agent  $y$  to only produce content that maximizes the function  $q(x|y)P_C(x)$ . More precisely, we have the following result.

**Lemma 2.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C,$$

and for  $y \in C_s = I_C$  and  $\beta_C(y) > 0$ , let the function  $\beta_C^*(x|y)$ ,  $x \in \mathcal{R}$ , be given by

$$\beta_C^*(\cdot|y) = \arg \max_{\beta(\cdot|y): \|\beta(\cdot|y)\| \leq \beta_C(y)} \int_{\mathcal{R}} \beta_C(x|y) [q(x|y)P_C(x) - \alpha_C c] dx,$$

and let

$$b^* = \max_{x \in \mathcal{R}} q(x|y)P_C(x).$$

Finally, let the set  $A \subseteq \mathcal{R}$  be given by

$$A = \left\{ x \in \mathcal{R} \mid q(x|y)P_C(x) = b^* \right\}.$$

If for agent  $y \in I_C$  we have that

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] > 0,$$

then the optimal production rate  $\beta_C^*(\cdot|y)$  of agent  $y$  has the property that

$$\int_A \beta_C^*(x|y) dx = \int_{\mathcal{R}} \beta_C^*(x|y) dx.$$

*Proof.* We first observe that if for agent  $y \in I_C$  we have that

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] > 0,$$

then optimizing the function

$$\int_{\mathcal{R}} \beta_C(x|y) [q(x|y)P_C(x) - \alpha_C c] dx$$

is equivalent to optimizing the function

$$\int_{\mathcal{R}} \beta_C(x|y) q(x|y) P_C(x) dx,$$

as the term  $\alpha_C c$  does not depend on  $x$ . Furthermore, we observe that if for agent  $y \in I_C$  we have that

$$\max_{\mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] > 0, \quad y \in I_C,$$

then it is optimal for agent  $y \in I_C$  to choose an content production rate in community  $C$  such that

$$\int_{\mathcal{R}} \beta_C^*(x|y) dx = \beta_C(y),$$

i.e. the total rate allocated in community  $C$  is equal to the maximal allowed rate  $\beta_C(y)$  for agent  $y$ .

Using these two results, we prove the lemma by contradiction as follows. Suppose that for a given agent  $y \in I_C$  we have that

$$\max_{\mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] > 0, \quad y \in I_C.$$

Furthermore, suppose that for this agent  $y$  we have that  $\beta_C^*(x|y)$ ,  $x \in \mathcal{R}$ , is an optimal solution to the optimization problem

$$\beta_C^*(\cdot|y) = \arg \max_{\beta(\cdot|y): \|\beta(\cdot|y)\| \leq \beta_C(y)} \int_{\mathcal{R}} \beta_C(x|y) [q(x|y)P_C(x) - \alpha_C c] dx$$

such that

$$\int_A \beta_C^*(x|y) dx < \beta_C(y) = \int_{\mathcal{R}} \beta_C^*(x|y) dx.$$

Note that this implies that for the set

$$B = \mathcal{R} \setminus A$$

we have that

$$\int_B \beta_C^*(x|y) dx > 0.$$

As by Assumption 1 and by Lemma 1 the function  $q(x|y)P_C(x)$  is continuous with respect to  $x$  on  $\mathcal{R}$ , there exists a closed set  $B_0 \subseteq B$  such that

$$\max_{x \in B_0} [q(x|y)P_C(x)] = b' < b^*$$

and

$$\int_{B_0} \beta_C^*(x|y) dx = \epsilon > 0.$$

To obtain the contradiction, we consider the production rate function  $\beta'_C(\cdot|y)$ ,  $x \in \mathcal{R}$ , given as follows. We pick a point  $x^* \in A$  and set

$$\beta'_C(x|y) = \beta_C^*(x|y) + \epsilon \delta(x - x^*), \quad x \in A,$$

where  $\delta$  is the Dirac delta function. Furthermore, we set

$$\beta'_C(x|y) = 0, \quad x \in B_0,$$

and

$$\beta'_C(x|y) = \beta_C^*(x|y), \quad x \notin A \cup B_0.$$

Note that for the function  $\beta'_C(\cdot|y)$  we have that

$$\int_{\mathcal{R}} \beta'_C(x|y) dx = \int_{\mathcal{R}} \beta_C^*(x|y) dx.$$

Furthermore, comparing  $\beta_C^*(x|y)$  with  $\beta'_C(x|y)$  we obtain that

$$\begin{aligned} & \int_{\mathcal{R}} \beta'_C(x|y) q(x|y) P_C(x) dx - \int_{\mathcal{R}} \beta_C^*(x|y) q(x|y) P_C(x) dx \\ &= \int_{x \in A} \epsilon \delta(x - x^*) q(x|y) P_C(x) dx - \int_{B_0} \beta_C^*(x|y) q(x|y) P_C(x) dx \\ &\geq \epsilon b^* - \epsilon b' = \epsilon(b^* - b') > 0, \end{aligned}$$

where we used the fact that by construction we have that

$$q(x|y) P_C(x) \leq b', \quad x \in B_0,$$

and

$$\int_{B_0} \beta_C^*(x|y) dx = \epsilon.$$

This leads to a contradiction, i.e. as we have that

$$\int_{\mathcal{R}} \beta'_C(x|y) q(x|y) P_C(x) dx - \int_{\mathcal{R}} \beta_C^*(x|y) q(x|y) P_C(x) dx > 0,$$

it follows that  $\beta_C^*(x|y)$  is not an optimal rate allocation for agent  $y \in I_C$ . The result of the lemma then follows, i.e. if

$$\beta_C^*(\cdot|y) = \arg \max_{\beta(\cdot|y): \|\beta(\cdot|y)\| \leq \beta_C(y)} \int_{\mathcal{R}} \beta_C(x|y) (q(x|y) P_C(x) - \alpha_C c) dx,$$

then we have that

$$\int_{\mathcal{R}} \beta_C^*(x|y) dx = \int_A \beta_C^*(x|y) dx = \beta_C^*(y).$$

□

The following corollary is obtained immediately from Lemma 2.

**Corollary 1.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C$$

and

$$\beta_C(y) > 0, \quad y \in C_s = I_C.$$

If for agent  $y \in C_s = I_C$  we have that

$$\max_{x \in \mathcal{R}} [q(x|y) P_C(x) - \alpha_C c] > 0,$$

then the function

$$\beta_C^*(x|y) = \beta_C(y) \delta(x - x^*(y)), \quad x \in \mathcal{R},$$

where  $\delta$  is the Dirac delta function and

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_C(x), \quad y \in I_C,$$

is a solution to

$$\beta_C^*(\cdot|y) = \arg \max_{\beta(\cdot|y): \|\beta(\cdot|y)\| \leq \beta_C(y)} \int_{\mathcal{R}} \beta_C(x|y) [q(x|y) P_C(x) - \alpha_C c] dx.$$

In the next subsection we derive properties of the function

$$q(x|y) P_C(x), \quad x \in \mathcal{R}, y \in I_C,$$

that we then use to show that the optimization problem

$$\max_{x \in \mathcal{R}} q(x|y) P_C(x), \quad y \in I_C,$$

in Corollary 1 has a unique solution  $x^*(y)$ .



## C.2 Properties of the function $q(x|y)P_C(x)$

In this subsection, we characterize the properties of the function

$$q(x|y)P_C(x), \quad x \in \mathcal{R}, y \in I_C.$$

Our first result shows that for

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

we have that

$$q(x^*(y)|y)P_C(x^*(y)) > 0.$$

**Lemma 3.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . If we have that*

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C,$$

Then for

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

we have that

$$q(x^*(y)|y)P_C(x^*(y)) > 0.$$

*Proof.* As by assumption we have that

$$\alpha_C(y) = E_p > 0, \quad y \in C_d = I_C,$$

it follows from Lemma 1 that

$$P_C(x) > 0, \quad x \in \mathcal{R},$$

and from Assumption 1 that

$$q(y|y)P_C(y) = g(0)P_C(y) > 0, \quad y \in I_C.$$

Using this result, we obtain that

$$q(x^*(y)|y)P_C(x^*(y)) \geq q(y|y)P_C(y) > 0, \quad y \in I_C,$$

and the result of the lemma follows. □

Our next two results provides additional properties of the function  $q(x|y)P_C(x)$  that we use in our analysis.

**Lemma 4.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C,$$

and

$$L_C < \frac{L}{2}.$$

Then the function  $q(x|y)P_C(x)$  has the properties that

a) for  $y \in I_C$ , the function  $q(x|y)P_C(x)$  is continuously differentiable in  $x$  on  $\text{supp}(q(\cdot|y))$ .

b) for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)],$$

the function  $q(x|y)P_C(x)$  is strictly increasing in  $x$  on  $[\text{mid}(I_C) - L, y) \cap \text{supp}(q(\cdot|y))$ , and strictly decreasing in  $x$  on  $(\text{mid}(I_C), \text{mid}(I_C) + L) \cap \text{supp}(q(\cdot|y))$ .

c) for

$$y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C],$$

the function  $q(x|y)P_C(x)$  is strictly increasing in  $x$  on  $[\text{mid}(I_C) - L, \text{mid}(I_C)) \cap \text{supp}(q(\cdot|y))$ , and strictly decreasing in  $x$  on  $(y, \text{mid}(I_C) + L) \cap \text{supp}(q(\cdot|y))$ .

*Proof.* We first prove property a) of the function  $q(x|y)P_C(x)$ . By Assumption 1 we have that  $q(x|y)$  is continuously differentiable in  $x$  on  $\text{supp}(q(\cdot|y))$ . Furthermore, by Lemma 1, we have that the function  $P_C(x)$  is continuously differentiable on  $\mathcal{R}$ . It then follows that for  $y \in I_C$  we have that the function  $q(x|y)P_C(x)$  is continuously differentiable in  $x$  on  $\text{supp}(q(\cdot|y))$ .

Next we prove property b). By Assumption 1 the function  $g(x)$  is strictly decreasing on  $\text{supp}(g(\cdot)) \cap (0, L]$ , and by Lemma 1 the function  $P_C(x)$  is strictly increasing in  $x$  on  $[\text{mid}(I_C) - L, \text{mid}(I_C))$ , and strictly decreasing in  $x$  on  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ . It then follows that for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)],$$

we have that the function  $q(x|y)P_C(x)$  is strictly increasing in  $x$  on  $[\text{mid}(I_C) - L, y) \cap \text{supp}(q(\cdot|y))$ , and strictly decreasing in  $x$  on  $[\text{mid}(I_C), \text{mid}(I_C) + L) \cap \text{supp}(q(\cdot|y))$ .

Property c) of the function  $q(x|y)P_C(x)$  can be shown using the same argument as for property b).

The result of the lemma then follows.  $\square$

**Lemma 5.** Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C,$$

and

$$L_C < \frac{L}{2}.$$

Then for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)],$$

we have that the function  $q(x|y)P_C(x)$  is strictly concave in  $x$  on the interval

$$[y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y)).$$

Similarly for  $y \in I_C$  such that

$$y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C],$$

we have that the function  $q(x|y)P_C(x)$  is strictly concave in  $x$  on the interval

$$[\text{mid}(I_C), y] \cap \text{supp}(q(\cdot|y)).$$

*Proof.* We first consider the case where

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)].$$

By Assumption 1 we have that the function  $q(\cdot|y)$  is twice differentiable and strictly concave on the  $[y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y))$ , and by Lemma 1 we have that the function  $P_C(x)$  is twice continuously differentiable and strictly concave on the  $[y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y))$ . Furthermore, by Assumption 1 the function  $q(\cdot|y)$  is strictly decreasing on  $(y, \text{mid}(I_C)) \cap \text{supp}(q(\cdot|y))$ , and by Lemma 1 the function  $P_C(x)$  is strictly increasing on  $[y, \text{mid}(I_C)) \cap \text{supp}(q(\cdot|y))$ . It follows that for

$$x \in [y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y)),$$

we have that

$$\frac{d^2}{dx^2} q(x|y)P_C(x) = q(x|y)''P_C(x) + 2q'(x|y)P_C'(x) + q(x|y)P_C''(x) < 0.$$

This establishes the result for the case where  $y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)]$ .

The result for the case where

$$y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C],$$

can be obtained using the same argument. The result of the lemma then follows.  $\square$

The next proposition summarizes the properties of the function  $q(x|y)P_C(x)$  that we obtained in this appendix, and that we will frequently use in our analysis.

**Proposition 10.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C,$$

and

$$L_C < \frac{L}{2}.$$

Then for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)]$$

we have that the function  $q(x|y)P_C(x)$  is

- a) strictly increasing in  $x$  on  $[\text{mid}(I_C) - L, y) \cap \text{supp}(q(\cdot|y))$ .
- b) strictly concave in  $x$  on  $[y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y))$ .
- c) strictly decreasing in  $x$  on  $(\text{mid}(I_C), \text{mid}(I_C) + L) \cap \text{supp}(q(\cdot|y))$ .

Similarly, for

$$y \in [\text{mid}(I_C), \text{mid}(I_C)] + L_C$$

we have that the function  $q(x|y)P_C(x)$  is

- a) strictly increasing in  $x$  on  $[\text{mid}(I_C) - L, \text{mid}(I_C)) \cap \text{supp}(q(\cdot|y))$ .
- b) strictly concave in  $x$  on  $[\text{mid}(I_C), y] \cap \text{supp}(q(\cdot|y))$ .
- c) strictly decreasing in  $x$  on  $(y, \text{mid}(I_C) + L) \cap \text{supp}(q(\cdot|y))$ .

Finally, we derive one additional property for the function  $q(x|y)P_C(x)$ .

**Lemma 6.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C,$$

and

$$L_C < \frac{L}{2}.$$

Then the following is true. If for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)]$$

we have that

$$\sup \{x \in \mathcal{R} | q(x|y)P_C(x) > 0\} > \text{mid}(I_C),$$

then we have

$$q(\text{mid}(I_C)|y)P_C(\text{mid}(I_C)) > q(x|y)P_C(x), \quad x \in [y + L, \text{mid}(I_C) + L].$$

Similarly if for

$$y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C]$$

we have that

$$\inf \{x \in \mathcal{R} | q(x|y)P_C(x) > 0\} < \text{mid}(I_C),$$

then we have

$$q(\text{mid}(I_C)|y)P_C(\text{mid}(I_C)) > q(x|y)P_C(x), \quad x \in [\text{mid}(I_C) - L, y - L].$$

*Proof.* We prove the result for the case where

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)].$$

The result for the case where

$$y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C]$$

can be obtained using the same argument.

Let  $y$  be such that

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)].$$

We first show for this case that under the assumption that

$$2L_C < L,$$

it follows that when

$$\sup \{x \in \mathcal{R} | q(x|y)P_C(x) > 0\} > \text{mid}(I_C),$$

then we obtain that

$$q(x|y) < q(\text{mid}(I_C)|y), \quad x \in [y + L, \text{mid}(I_C) + L]. \quad (7)$$

In order to show this result, we have to show that

$$\|x - y\| > \|\text{mid}(I_C) - y\|, \quad x \in [y + L, \text{mid}(I_C) + L].$$

Note that the function  $|y - x|$  is decreasing in  $x$  on  $[y + L, \text{mid}(I_C) + L]$ , and therefore in order to show it result it suffices to show that

$$\|\text{mid}(I_C) + L - y\| > \|\text{mid}(I_C) - y\|.$$

As for  $y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)]$  we have that

$$\text{mid}(I_C) + L - y > L,$$

we have from the definition of the torus metric that

$$\|\text{mid}(I_C) + L - y\| = 2L - \text{mid}(I_C) - L + y = y - [\text{mid}(I_C) - y].$$

To see that

$$y - \lfloor \text{mid}(I_C) - y \rfloor > \|\text{mid}(I_C) - y\|,$$

note the following. As by assumption we have

$$2L_C < L,$$

it follows that

$$\begin{aligned} \|\text{mid}(I_C) - y\| &< L_C \\ &< L - L_C \\ &< L - \lfloor \text{mid}(I_C) - y \rfloor. \end{aligned}$$

Combining Eq. (7) with the fact that by Proposition 5 the function  $P_C(x)$  is strictly decreasing in  $x$  on  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ , we obtain that for  $y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)]$  we have that

$$q(x|y)P_C(x) < q(\text{mid}(I_C)|y)P_C(\text{mid}(I_C)), \quad x \in [y + L, \text{mid}(I_C) + L].$$

The result of the lemma then follows. □

### C.3 Properties of the Optimal Content $x^*(y)$

Using the result for the function  $q(x|y)P_C(x)$  that we obtained in the previous subsection, we characterize in this subsection the properties of the optimal content  $x^*(y)$  given by

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x).$$

Our first result follows directly from Proposition 10 and Lemma 6.

**Lemma 7.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C,$$

and

$$L_C < \frac{L}{2}.$$

Let

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C.$$

Then for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)]$$

we have that

$$x^*(y) \in [y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y)).$$

Similarly, for

$$y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C]$$

we have that

$$x^*(y) \in [\text{mid}(I_C), y] \cap \text{supp}(q(\cdot|y)).$$

Next we derive conditions under which optimization problem

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

has a unique solution.

**Lemma 8.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . If*

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C,$$

and

$$L_C < \frac{L}{2},$$

then there exists a unique solution  $x^*(y)$  to the optimization problem

$$\max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

which is the unique solution to the equation

$$q'(x|y)P_C(x) + q(x|y)P'_C(x) = 0, \quad x \in I_C \cap \text{supp}(g(\cdot|y)).$$

Furthermore,  $x^*(y)$  has the properties that

a)  $x^*(\text{mid}(I_C)) = \text{mid}(I_C)$ .

b) for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C))$$

we have

$$x^*(y) \in (y, \text{mid}(I_C)) \cap \text{supp}(q(\cdot|y)).$$

c) for

$$y \in (\text{mid}(I_C), \text{mid}(I_C) + L_C]$$

we have

$$x^*(y) \in (\text{mid}(I_C), y) \cap \text{supp}(q(\cdot|y)).$$

*Proof.* We prove the result for the case where

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)].$$

The result for the case where

$$y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C]$$

can be obtained using the same argument.

By Lemma 7, we have that

$$x^*(y) \in [y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y)), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)].$$

This implies that

$$x^*(\text{mid}(I_C)) = \text{mid}(I_C).$$

Furthermore, by Lemma 5 we have for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C))$$

that the function  $q(x|y)P_C(x)$  is strictly concave on the interval  $[y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y))$ , and it follows that the optimization problem

$$\max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)).$$

has an unique solution.

It remains to show that the unique solution  $x^*(y)$  is an interior solution, and we have that

$$x^*(y) \in (y, \text{mid}(I_C)) \cap \text{supp}(q(\cdot|y)), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C))$$



as well as that  $x^*(y)$  is the unique solution to the equation

$$q'(x|y)P_C(x) + q(x|y)P'_C(x) = 0, \quad x \in I_C \cap \text{supp}(g(\cdot|y)).$$

Note that for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C))$$

and

$$x \in [y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y))$$

we have that

$$\frac{d}{dx}q(x|y)P_C(x) = q'(x|y)P_C(x) + q(x|y)P'_C(x).$$

By Assumption 1 we have that

$$g(0) > 0,$$

and by the assumption that

$$\alpha_C(y) = E_p > 0, \quad y \in C_d = I_C,$$

have that

$$P_C(x) > 0, \quad x \in \mathcal{R}.$$

Furthermore by Assumption 1 we have have

$$g'(0) = 0$$

and by Lemma 1 we have

$$P'_C(y) > 0, \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)).$$

Combining these results, it then follows that for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C))$$

we have that

$$q'(y|y)P_C(y) + q(y|y)P'_C(y) = g(0)P'_C(y) > 0.$$

Hence we have that

$$x^*(y) > y.$$

Next we show that for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C))$$

we have that

$$x^*(y) < \text{mid}(I_C).$$

To do this, we consider two separate cases. For the first case, we assume that for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C))$$

we have

$$\sup \{x \in \mathcal{R} | q(x|y)P_C(x) > 0\} \geq \text{mid}(I_C).$$

For this case, we obtain for

$$y_0 = \text{mid}(I_C)$$

that

$$q'(y_0|y)P_C(y_0) + q(y_0|y)P'_C(y_0) = q'(y_0|y)P_C(y_0) < 0,$$

where we used the fact that by Assumption 1 we have have

$$q'(x|y) < 0, \quad x \in (y, y_0] \cap \text{supp}(q(\cdot|y)),$$

and by Lemma 1 we have

$$P'_C(\text{mid}(I_C)) = 0.$$

Hence we have for this case that

$$x^*(y) < \text{mid}(I_C).$$

For the second case, we assume that for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C))$$

we have

$$x_s = \sup \{x \in \mathcal{R} | q(x|y)P_C(x) > 0\} < \text{mid}(I_C).$$

For this case, we obtain that

$$q'(x_s|y)P_C(x_s) + q(x_s|y)P'_C(x_s) = q'(\text{mid}(I_C)|y)P_C(y) < 0,$$

where we used the fact that for

$$x_s = \sup \{x \in \mathcal{R} | q(x|y)P_C(x)\}$$

we have that

$$q(x_s|y) = 0$$

and

$$q'(x_s|y) < 0.$$

It follows that

$$x^*(y) < x_s.$$

The above results state that

$$x^*(y) > y$$

and

$$x^*(y) < \min\{y_0, x_s\},$$

and hence we have that  $x^*(y)$  is an interior solution, and we have that  $x^*(y)$  is the unique solution to the equation

$$q'(x|y)P_C(x) + q(x|y)P'_C(x) = 0, \quad x \in I_C \cap \text{supp}(g(\cdot|y)).$$

The result of the lemma then follows. □

The following proposition summarizes Lemma 2 to 8 that we obtained in this appendix.

**Proposition 11.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in C_d = I_C,$$

and

$$L_C < \frac{L}{2}.$$

If for agent  $y \in I_C$  we have that

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] > 0,$$

then the optimal content production rate allocation

$$\beta_C^*(\cdot|y) = \arg \max_{\beta(\cdot|y): |\beta(\cdot|y)| \leq \beta_C(y)} \int_{\mathcal{R}} \beta_C(x|y)(q(x|y)P_C(x) - \alpha_C c) dx$$

of agent  $y$  is given by

$$\beta_C^*(x|y) = \beta_C(y)\delta(x - x^*(y)), \quad x \in \mathcal{R},$$

where  $\delta$  is the Dirac delta function and  $x^*(y)$  is the unique solution to the optimization problem

$$\max_{x \in \mathcal{R}} q(x|y)P_C(x).$$

Furthermore,  $x^*(y)$  has the properties that

(a)  $q(x^*(y)|y) > 0, \quad y \in I_C.$

(b)  $x^*(y), y \in I_C$  is given by the unique solution to the equation

$$q'(x|y)P_C(x) + q(x|y)P_C'(x) = 0, \quad x \in I_C \cap \text{supp}(q(\cdot|y)).$$

(c) we have

$$x^*(\text{mid}(I_C)) = \text{mid}(I_C).$$

(d) we have

$$x^*(y) \in (y, \text{mid}(I_C)) \cap \text{supp}(q(\cdot|y)), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)),$$

and

$$x^*(y) \in (\text{mid}(I_C), y) \cap \text{supp}(q(\cdot|y)), \quad y \in (\text{mid}(I_C), \text{mid}(I_C) + L_C].$$

Our last result in this subsection characterizes the optimal solution  $x^*(y)$  for agents  $y$  outside the interval  $I_C$ . In particular, we have the following result.

**Lemma 9.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . If we have that*

$$\alpha_C(y) = E_p, \quad y \in I_C,$$

and

$$L_C < \frac{L}{2},$$

then the following is true for the function

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x).$$

For

$$y \in [\text{mid}(I_C) - L, \text{mid}(I_C) - L_C]$$

we have that

$$x^*(y) \leq x^*(\text{mid}(I_C) - L_C),$$

and for

$$y \in (\text{mid}(I_C) + L_C, \text{mid}(I_C) + L)$$

we have that

$$x^*(y) \geq x^*(\text{mid}(I_C) + L_C).$$

*Proof.* We prove the lemma for the case where

$$y \in [\text{mid}(I_C) - L, \text{mid}(I_C) - L_C].$$

The case where  $y \in (\text{mid}(I_C) + L_C, \text{mid}(I_C) + L)$  can be shown using the same argument.

Let  $y$  be such that

$$y \in [\text{mid}(I_C) - L, \text{mid}(I_C) - L_C],$$

and let

$$y_l = \text{mid}(I_C) - L_C.$$

Recall that by Lemma 8 we have that

$$x^*(y_l) \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)].$$

Furthermore, by the same argument as given in the proof of Lemma 8, we have that

$$x^*(y) \in [y, \text{mid}(I_C)] \cap \text{supp}(q(\cdot|y)), \quad y \in [\text{mid}(I_C) - L, \text{mid}(I_C) - L_C].$$

Therefore, in order to prove the result of the lemma it suffices to show that if

$$x^*(y) \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)], \quad y \in [\text{mid}(I_C) - L, \text{mid}(I_C) - L_C],$$

then we have that

$$x^*(y) \leq x^*(y_l).$$

To do this, we note that by the same argument as given in the proof of Lemma 5, we have that the function

$$q(x|y)P_C(x), \quad y \in [\text{mid}(I_C) - L, \text{mid}(I_C) - L_C],$$

is strictly concave in  $x$  on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C)]$ . It then follows that if for

$$y \in [\text{mid}(I_C) - L, \text{mid}(I_C) - L_C]$$

we have that

$$x^*(y) \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)],$$

then the optimization problem

$$x^*(y) \arg \max_{x \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)]} q(x|y)P_C(x), \quad y \in [\text{mid}(I_C) - L, \text{mid}(I_C) - L_C],$$

has a unique solution, given by

$$g'(x^*(y) - y)P_C(x^*(y)) + g(x^*(y) - y)P'_C(x^*(y)) = 0,$$

where  $g$  is the function of Assumption 1.

Therefore, in order to show that if

$$x^*(y) \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)], \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)],$$

then we have that

$$x^*(y) \leq x^*(y_l) = x^*(\text{mid}(I_C) - L_C),$$

it suffices to show that

$$g'(x^*(y_l) - y)P_C(x^*(y_l)) + g(x^*(y_l) - y)P'_C(x^*(y_l)) \leq 0.$$

Note that by Assumption 1 we have for

$$y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)]$$

that

$$g'(x^*(y) - y) \leq 0, \quad x^*(y) - y \in \text{supp}(g),$$

and by Lemma 1 we have that

$$P'_C(x^*(y)) \geq 0.$$

Furthermore, as the function  $g$  is by Assumption 1 decreasing and strictly concave on  $\text{supp}(g)$  and we have that

$$x^*(y_l) - y_l \leq x^*(y_l) - y,$$

it follows that

$$g(x^*(y_l) - y) \leq g(x^*(y_l) - y_l)$$

and

$$g'(x^*(y_l) - y) \leq g'(x^*(y_l) - y_l).$$

Combining these results with the fact that

$$g(x) \geq 0, \quad x \in \text{supp}(g),$$

and

$$P_C(x) \geq 0, \quad x \in \mathcal{R},$$

we obtain that

$$\begin{aligned} & g'(x^*(y_l) - y)P_C(x^*(y_l)) + g(x^*(y_l) - y)P'_C(x^*(y_l)) \\ & \leq g'(x^*(y_l) - y_l)P_C(x^*(y_l)) + g(x^*(y_l) - y_l)P'_C(x^*(y_l)) \end{aligned}$$

As by Proposition 11 we have that

$$g'(x^*(y_l) - y_l)P_C(x^*(y_l)) + g(x^*(y_l) - y_l)P'_C(x^*(y_l)) = 0,$$

we obtain that

$$g'(x^*(y_l) - y)P_C(x^*(y_l)) + g(x^*(y_l) - y)P'_C(x^*(y_l)) \leq 0.$$

The result of the lemma then follows. □

### D Proof of Proposition 3

In the previous appendix we showed that for a given interval community  $C = (C_d, C_s) = (I_C, I_C)$  such that

$$\alpha_C(y) = E_p > 0, \quad y \in I_C,$$

and

$$L_C < \frac{L}{2},$$

the optimization problem

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

has a unique solution.

In this subsection we study the properties of the function  $x^*(y)$ ,  $y \in I_C$ , i.e. we study the properties of the optimal content function  $x^*(\cdot)$  as a function of  $y \in I_C$ , and prove the results of Proposition 3.

We first show that the function  $x^*(y)$  is differentiable and strictly increasing on  $I_C$ .

**Lemma 10.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in I_C,$$

and

$$L_C < \frac{L}{2}.$$

Then the function  $x^*(y)$  given by

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

is continuously differentiable and strictly increasing on  $I_C$ .

*Proof.* Without loss of generality, we can assume that the interval  $I_C$  is given by

$$I_C = [-L_C, L_C].$$

We first consider the case where

$$y \in [0, L_C].$$

Recall that by Proposition 11 we have that  $x^*(y)$ ,  $y \in I_C$ , is the unique solution to the equation

$$q'(x|y)P_C(x) + q(x|y)P'_C(x) = 0, \quad x \in I_C \cap \text{supp}((q(\cdot|y))).$$

As by Proposition 11 we have that

$$x^*(y) \in [0, y] \cap \text{supp}(q(\cdot|y)), \quad y \in [0, L_C],$$

we can re-write  $x^*(y)$ ,  $y \in [0, L_C]$ , as the unique solution to

$$x^*(y) = \arg \max_{x \in [0, y]} g(y-x)P_C(x),$$

where the function  $g$  is as given in Assumption 1. It then follows that  $x^*(y)$ ,  $y \in [0, L_C]$ , is the unique solution to the equation

$$-g'(y-x)P_C(x) + g(y-x)P'_C(x) = 0, \quad y \in [0, L_C], x \in [0, y] \cap \text{supp}(g).$$

Let the function  $F^+(x, y)$  be given by

$$F^+(x, y) = -g'(y-x)P_C(x) + g(y-x)P'_C(x), \quad y \in [0, L_C], x \in [0, y].$$

Note that we have that

$$F^+(x^*(y), y) = 0, \quad y \in [0, L_C].$$

Furthermore we have that

$$\frac{d}{dx} F^+(x, y) = g''(y-x)P_C(x) - 2g'(y-x)P'_C(x) + g(y-x)P''_C(x), \quad y \in [0, L_C], x \in [0, y] \cap \text{supp}(g),$$

and

$$\frac{d}{dy}F^+(x, y) = -g''(y-x)P_C(x) + g'(y-x)P'_C(x), \quad y \in [0, L_C], x \in [0, y] \cap \text{supp}(g).$$

As by Assumption we have  $g(x)$  is strictly decreasing and concave in  $x$  for  $x \in [0, L_C] \cap \text{supp}(g)$ , and  $P_C(x)$  is strictly decreasing and concave in  $x$  for  $x \in [0, L_C]$ , it follows that for  $y \in [0, L_C]$  we have for

$$y \in [0, L_C] \text{ and } x \in [0, y] \cap \text{supp}(g)$$

that

$$\begin{aligned} g''(y-x)P_C(x) &< 0, \\ -g'(y-x)P'_C(x) &\leq 0, \end{aligned}$$

and

$$g(y-x)P''_C(x) < 0.$$

Combining the above results, it follows that

$$\frac{d}{dx}F^+(x, y) < 0, \quad y \in [0, L_C], x \in [0, y] \cap \text{supp}(g)$$

and

$$\frac{d}{dy}F^+(x, y) > 0, \quad y \in [0, L_C], x \in [0, y] \cap \text{supp}(g).$$

By Proposition 11 we have that

$$x^*(y) \in [0, y] \cap \text{supp}(g), \quad y \in [0, L_C],$$

and it follows that

$$\frac{d}{dx}F^+(x = x^*(y), y) < 0, \quad y \in [0, L_C], \quad (8)$$

and

$$\frac{d}{dy}F^+(x = x^*(y), y) > 0, \quad y \in [0, L_C]. \quad (9)$$

Using the implicit function theorem [8], we then obtain that the function  $x^*(y)$  is continuous and differentiable on  $[0, L_C]$  with

$$\frac{d}{dy}x^*(y) = \frac{-\frac{d}{dy}F^+(x = x^*(y), y)}{\frac{d}{dx}F^+(x = x^*(y), y)}, \quad y \in [0, L_C].$$

Combining this result with Eq. (8) and (9), we obtain that

$$\frac{d}{dy}x^*(y) = \frac{-\frac{d}{dy}F^+(x = x^*(y), y)}{\frac{d}{dx}F^+(x = x^*(y), y)} > 0,$$

and the function  $x^*(y)$  is strictly increasing on  $[0, L_C]$ .

Next we consider the case where

$$y \in [-L_C, 0].$$

For this case, let the function  $F^-(x, y)$  be given by

$$F^-(x, y) = g'(x-y)P_C(x) + g(x-y)P'_C(x), \quad y \in [-L_C, 0], x \in [y, 0].$$

Note that we have that

$$F^-(x^*(y), y) = 0, \quad y \in [-L_C, 0].$$

Using the same argument as for the case where  $y \in [0, L_C]$ , we obtain from the above results that the function  $x^*(y)$  is continuous and differentiable on  $[-L_C, 0]$  with

$$\frac{d}{dy}x^*(y) = \frac{-\frac{d}{dy}F^-(x = x^*(y), y)}{\frac{d}{dx}F^-(x = x^*(y), y)}, \quad y \in [-L_C, 0],$$

that we have that

$$\frac{d}{dy}x^*(y) = \frac{-\frac{d}{dy}F^-(x = x^*(y), y)}{\frac{d}{dx}F^-(x = x^*(y), y)} > 0.$$

It follows that the function  $x^*(y)$  is strictly increasing on  $[-L_C, 0]$ .

In order to show that the function  $x^*(y)$  is differentiable in  $I_C$ , we have to show that

$$\frac{-\frac{d}{dy}F^+(x = x^*(0), 0)}{\frac{d}{dx}F^+(x = x^*(0), 0)} = \frac{-\frac{d}{dy}F^-(x = x^*(0), 0)}{\frac{d}{dx}F^-(x = x^*(0), 0)}.$$

Indeed, as by Assumption 1 we have that

$$g'(0) = 0,$$

it follows from the above equations that

$$\frac{-\frac{d}{dy}F^+(x = x^*(0), 0)}{\frac{d}{dx}F^+(x = x^*(0), 0)} = \frac{-\frac{d}{dy}F^-(x = x^*(0), 0)}{\frac{d}{dx}F^-(x = x^*(0), 0)} = \frac{g''(0)P_C(0)}{g''(0)P_C(0) + g(0)P_C''(0)}.$$

It remains to prove that the function  $x^*(y)$  is continuously differentiable on  $I_C$ . By Assumption 1 we have that the function  $g(x)$ ,  $x \in [0, L]$ , is twice continuously differentiable. Furthermore, by Proposition 5, we have that the function  $P_C(x)$  is twice continuously differentiable on  $\mathcal{R}$ . It then follows that the derivative

$$\frac{d}{dy}x^*(y), \quad y \in I_C,$$

is continuous, and hence the function  $x^*(y)$ ,  $y \in I_C$ , is continuously differentiable on  $I_C$ .

The result of the lemma then follows. □

Proposition 3 then follows directly from Lemma 10 and Proposition 11.



## E Proof of Proposition 4

In this appendix we prove Proposition 4 which characterizes the the function  $\Delta^*(y)$  given by

$$\Delta^*(y) = \|y - x^*(y)\|, \quad y \in I_C.$$

Recall that the function  $\Delta^*(y)$  characterizes the absolute value of the “displacement” of the optimal content  $x^*(y)$  that agent  $y$  produces, and content  $y$  that the agent is best at producing which is equal to content type  $y$ .

We first establish the result of Proposition 4 that the function  $\Delta^*(y)$ ,  $y \in I_C$ , is symmetric with respect to  $\text{mid}(I_C)$ .

**Lemma 11.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in I_C,$$

and

$$L_C < \frac{L}{2}.$$

Then function

$$\Delta^*(y) = \|y - x^*(y)\|, \quad y \in I_C,$$

where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

is symmetric with respect to  $\text{mid}(I_C)$  on  $I_C$  with

$$\Delta^*(\text{mid}(I_C)) = 0.$$

*Proof.* Note that in order to show that  $\Delta^*(y)$  symmetric on  $I_C$  with respect to  $\text{mid}(I_C)$  for  $y \in I_C$ , we have to show that

$$\Delta^*(\text{mid}(I_C) + s) = \Delta^*(\text{mid}(I_C) - s), \quad s \in [0, L_C].$$

Note that from Proposition 11 we have that

$$\Delta^*(y) = \arg \max_{s \in [0, y]} g(s)P_C(y - s), \quad y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C],$$

and

$$\Delta^*(y) = \arg \max_{s \in [0, y]} g(s)P_C(y + s), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)].$$

Setting

$$F^{(+)}(y, s) = g(s)P_C(y - s), \quad y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C],$$

and

$$F^{(-)}(y, s) = g(s)P_C(y + s), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)],$$

we have that

$$\Delta^*(y) = \arg \max_{s \in [0, y]} F^{(+)}(y, s), \quad y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C],$$

and

$$\Delta^*(y) = \arg \max_{s \in [0, y]} F^{(-)}(y, s), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)].$$

Therefore in order to show that  $\Delta^*(y)$  symmetric with respect to  $\text{mid}(I_C)$  for  $y \in I_C$ , it suffices to show that

$$F^{(+)}(\text{mid}(I_C) - x, s) = F^{(-)}(\text{mid}(I_C) + x, s), \quad x \in [0, L_C], s \in [0, x].$$

Note that the conditions that  $x \in [0, L_C]$  and  $s \in [0, x]$  imply that

$$x - s \geq 0.$$

Using the the result from Proposition 5 that the function  $P_C(x)$  is symmetric with respect to  $mid(I_C)$ , we then obtain that for  $x \in [0, L_C]$  and  $s \in [0, x]$  that

$$\begin{aligned} F^{(+)}(mid(I_C) - x, s) &= g(s)P_C(mid(I_C) + x - s) \\ &= g(s)P_C(mid(I_C) - x + s) \\ &= F^{(-)}(mid(I_C) - x, s). \end{aligned}$$

It then follows that  $\Delta^*(y)$  is symmetric with respect to  $mid(I_C)$  on  $I_C$ .

Finally, from Proposition 11 which states that

$$x^*(mid(I_C)) = I_C,$$

we obtain

$$\Delta^*(mid(I_C)) = 0.$$

The result of the lemma then follows. □

We next establish the result of Proposition 4 that the function  $\Delta^*(y)$  is strictly decreasing on  $[mid(I_C) - L_C, mid(I_C)]$ , and strictly increasing on  $[mid(I_C), mid(I_C)] + L_C$ . In addition, we characterize the derivative of the function  $\Delta^*(y)$ ,  $y \in I_C$ , with respect to  $y$ , and show that

$$|\Delta_y^*(y)| < 1, \quad y \in I_C \setminus \{mid(I_C)\},$$

where

$$\Delta_y^*(y) = \frac{d}{dy} \Delta^*(y).$$

**Lemma 12.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in I_C,$$

and

$$L_C < \frac{L}{2}.$$

Then for the function  $\Delta^*(y)$  given by

$$\Delta^*(y) = \|y - x^*(y)\|, \quad y \in I_C,$$

where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_C(x), \quad y \in I_C,$$

we have that

- a)  $\Delta^*(y)$  is continuously differentiable on  $I_C \setminus \{mid(I_C)\}$ .
- b)  $\Delta^*(y)$  strictly decreasing on  $[mid(I_C) - L_C, mid(I_C)]$ , and strictly increasing on  $[mid(I_C), mid(I_C)] + L_C$ .
- c) the absolute value of the derivative  $\frac{d}{dy} \Delta^*(y)$  of the function  $\Delta^*(y)$  with respect to  $y$  is less than 1 on  $I_C \setminus \{mid(I_C)\}$ , i.e. we have that

$$|\Delta_y^*(y)| < 1, \quad y \in I_C \setminus \{mid(I_C)\},$$

where

$$\Delta_y^*(y) = \frac{d}{dy} \Delta^*(y).$$

*Proof.* Without loss of generality, we can assume that the interval  $I_C$  is given by

$$I_C = [-L_C, L_C].$$

By Lemma 11 we have that the function  $\Delta^*(y)$  is symmetric with respect to  $mid(I_C)$ , it suffices to prove part the lemma for the case where

$$y \in [0, L_C].$$

We first prove part a) and b) of the lemma.

Recall that  $x^*(y)$  is given by

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_C(x), \quad y \in I_C,$$

and that by Proposition 11 the necessary and sufficient condition for  $x^*(y)$ ,  $y \in [0, L_C]$ , to be the solution to the above optimization problem is given by

$$q'(x|y) P_C(x) + q(x|y) P_C'(x) = 0, \quad x \in [0, y].$$

By Proposition 11 which states that

$$x^*(y) \in [0, y] \cap \text{supp}(q(\cdot|y)), \quad y \in [0, L_C],$$

we can rewrite  $x^*(y)$  for

$$y \in [0, L_C]$$

as

$$x^*(y) = y - \Delta^*(y).$$

Using the result from Proposition 11 which states that  $x^*(y)$  is the unique solution to the optimization problem

$$\max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

we obtain that  $\Delta^*(y)$  is the unique solution to the optimization problem

$$\Delta^*(y) = \arg \max_{s \in [0, y] \cap \text{supp}(g)} g(s)P_C(y-s), \quad y \in [0, L_C].$$

Furthermore, from Lemma 5 we obtain that the necessary and sufficient condition for  $\Delta^*(y)$  to be the solution to this optimization problem is given by

$$g'(\Delta^*(y))P_C(y - \Delta^*(y)) - g(\Delta^*(y))P'_C(y - \Delta^*(y)) = 0, \quad \Delta^*(y) \in [0, y] \cap \text{supp}(g).$$

Let the function  $F(s, y)$  be given by

$$F(s, y) = g'(s)P_C(y-s) - g(s)P'_C(y-s), \quad y \in [0, L_C], s \in [0, y] \cap \text{supp}(g).$$

Note that we have that

$$F(\Delta^*(y), y) = 0, \quad y \in [0, L_C].$$

Furthermore we have that

$$\frac{d}{ds}F(s, y) = g''(s)P_C(y-s) - 2g'(s)P'_C(y-s) + g(s)P''_C(y-s), \quad y \in [0, L_C], s \in [0, y] \cap \text{supp}(g),$$

and

$$\frac{d}{dy}F(s, y) = g'(s)P'_C(y-s) - g(s)P''_C(y-s), \quad y \in [0, L_C], s \in [0, y] \cap \text{supp}(g).$$

As by Assumption 1 we have  $g(s)$  non-increasing and strictly concave in  $s$  for  $s \in [0, y] \cap \text{supp}(g)$ , and by Proposition 5 we have that the function  $P_C(x)$  is non-increasing and strictly concave for  $x \in [0, L_C]$ , it follows that for

$$y \in [0, L_C] \text{ and } s \in [0, y] \cap \text{supp}(g)$$

we have that

$$\begin{aligned} g''(s)P_C(y-s) &< 0, \\ g'(s)P'_C(y-s) &\geq 0, \end{aligned}$$

and

$$g(s)P''_C(y-s) < 0.$$

Combining the above results, it follows that

$$\frac{d}{ds}F(s, y) < 0, \quad y \in [0, L_C], s \in [0, y] \cap \text{supp}(g)$$

and

$$\frac{d}{dy}F(s, y) > 0, \quad y \in [0, L_C], s \in [0, y] \cap \text{supp}(g).$$

By Proposition 11 we have that

$$x^*(y) \in [0, y], \quad y \in [0, L_C],$$

and it follows that

$$\frac{d}{ds}F(s = \Delta^*(y), y) < 0, \quad y \in [0, L_C], \quad (10)$$

and

$$\frac{d}{dy}F(s = \Delta^*(y), y) > 0, \quad y \in [0, L_C]. \quad (11)$$

Using the implicit function theorem [8], we obtain that the function  $\Delta^*(y)$  is continuous and differentiable on  $[0, L_C]$  with

$$\frac{d}{dy}\Delta^*(y) = \frac{-\frac{d}{dy}F(s = \Delta^*(y), y)}{\frac{d}{ds}F(s = \Delta^*(y), y)}, \quad y \in [0, L_C].$$

As by Assumption 1 we have that the functions  $f$  is continuously differentiable, and the function  $g$  is continuously differentiable on  $\text{supp}(g)$ , it follows that the function  $\Delta^*(y)$  is continuously differentiable on  $[0, L_C]$ . Furthermore, using Eq. (10) and (11) we obtain that

$$\frac{d}{dy}\Delta^*(y) = \frac{-\frac{d}{dy}F(s = \Delta^*(y), y)}{\frac{d}{ds}F(s = \Delta^*(y), y)} > 0, \quad y \in [0, L_C],$$

and the function  $\Delta^*(y)$  is strictly increasing on  $[0, L_C]$ . As by Lemma 11 the function  $\Delta^*(y)$  is symmetric with respect to  $\text{mid}(I_C)$  on  $I_C$ , it follows that function  $\Delta^*(y)$  is continuously differentiable and strictly decreasing on  $[-L_C, 0]$ .

Note that the above argument only shows that the function  $\Delta^*(y)$  is continuously differentiable on the interval  $[0, L_C]$  and  $[-L_C, 0]$ , but does not show that the derivative is continuous at  $y = 0$ . Hence the above proof does not show that the derivative at  $y = 0$  exists, and we can only conclude that

$$\left| \frac{d}{dy}\Delta^*(y) \right| < 1, \quad y \in I_C \setminus \{\text{mid}(I_C)\}.$$

This establishes part a) and b) of the lemma.

It remains to prove part c) of the lemma. Recall that

$$\frac{d}{ds}F(s, y) = g''(s)P_C(y-s) - g'(s)P'_C(y-s) - \frac{d}{dy}F(s, y), \quad y \in [0, L_C], s \in [0, y] \cap \text{supp}(g),$$

and we can re-write the derivative of  $\Delta^*(y)$ ,  $y \in [0, L_C]$ , as

$$\frac{d}{dy}\Delta^*(y) = \frac{-\frac{d}{dy}F(s = \Delta^*(y), y)}{g''(\Delta^*(y))P_C(x^*(y)) - g'(\Delta^*(y))P'_C(x^*(y)) - \frac{d}{dy}F(s = \Delta^*(y), y)}, \quad y \in [0, L_C].$$

By Proposition 11 we have that

$$x^*(y) \in [0, y], \quad y \in [0, L_C].$$

Furthermore, from the above results we have that

$$g''(\Delta^*(y))P_C(x^*(y)) < 0, \quad y \in [0, L_C],$$

and

$$-g'(\Delta^*(y))P'_C(x^*(y)) \leq 0, \quad y \in [0, L_C],$$

as well as

$$-\frac{d}{dy}F(s = \Delta^*(y), y) < 0, \quad y \in [0, L_C].$$

Combining these results we obtain for  $y \in [0, L_C]$  that

$$g''(\Delta^*(y))P_C(x^*(y)) - g'(\Delta^*(y))P'_C(x^*(y)) - \frac{d}{dy}F(s = \Delta^*(y), y) < -\frac{d}{dy}F(s = \Delta^*(y), y) < 0,$$

and hence we have that

$$\left| \frac{d}{dy}\Delta^*(y) \right| < 1, \quad y \in [0, L_C].$$

□

Using the result of Lemma 12, we derive an additional result for the function  $x^*(y)$ ,  $y \in I_C$ , which characterizes the image of the function  $x^*(y)$ .

**Lemma 13.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , be a given constant, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval on  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in C_d,$$

and

$$L_C < \frac{L}{2}.$$

Then the image of the function  $x^*(y)$ ,

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

is given by an interval  $I_C^* \subset \mathcal{R}$  such that

$$\text{mid}(I_C^*) = \text{mid}(I_C)$$

and

$$L_C^* = \frac{|I_C^*|}{2} < L_C.$$

*Proof.* As the interval  $I_C$  is a connected set and by Lemma 10 the function  $x^*(y)$  is continuous on  $I_C$ , it follows that the image of  $x^*(y)$ ,  $y \in I_C$ , is a connected set and hence given by an interval  $I_C^*$ . That means that the closure of the image is given by a closed interval.

By Proposition 11, we have that

$$x^*(y) = y + \Delta^*(y), \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)],$$

and

$$x^*(y) = y - \Delta^*(y), \quad y \in [\text{mid}(I_C), \text{mid}(I_C) + L_C],$$

where  $\Delta^*$  is given by

$$\Delta^*(y) = ||y - x^*(y)||, \quad y \in I_C.$$

Furthermore by Lemma 10, we have that the function  $x^*(y)$  is strictly increasing on  $I_C$ . Therefore in order to prove the lemma, we have to show that

$$\Delta^*(\text{mid}(I_C) - L_C) = \Delta^*(\text{mid}(I_C) + L_C) > 0,$$

Note that by Lemma 12 we have that function  $\Delta^*(y)$  is symmetric with respect to  $\text{mid}(I_C)$ , and it follows that

$$\text{mid}(I_C^*) = \text{mid}(I_C)$$

and

$$\Delta^*(\text{mid}(I_C) - L_C) = \Delta^*(\text{mid}(I_C) + L_C).$$

Therefore, in order to prove the lemma it remains to show that

$$\Delta^*(\text{mid}(I_C) + L_C) > 0.$$

To obtain this result, note that by Lemma 12 the function  $\Delta^*(y)$  is strictly increasing on  $[\text{mid}(I_C), \text{mid}(I_C) + L_C]$ , and by Lemma 11 we have that

$$\Delta^*(\text{mid}(I_C)) = 0.$$

It then follows that

$$\Delta^*(\text{mid}(I_C) + L_C) > 0,$$

which completes the proof of the lemma.  $\square$

## F Proof of Proposition 6

In this appendix we prove Proposition 6. To do this, we characterize in the next subsection the density  $\beta_C^*(x)$  of the rate at which content of type  $x$  is being produced in an interval community.

### F.1 Properties of $\beta_C^*(x)$

Having characterized in Proposition 11 the optimal content production rate  $\beta_C^*(\cdot|y)$  of an agent in a given interval community  $C = (I_C, I_C)$ , we characterize in this subsection the resulting density  $\beta_C^*(x)$  of the rate at which of content of type  $x$  that is being produced in the community  $C$ , i.e. the density  $\beta_C^*(x)$  given by

$$\beta_C^*(x) = \int_{I_C} \beta_C^*(x|y)dy, \quad x \in \mathcal{R}.$$

We use the following definitions. Recall that by Lemma 13 we have for a given interval community  $C = (I_C, I_C)$ , that the closure  $I_C^*$  of the image of the function  $x^*(y), y \in I_C$  is given by an interval  $I_C^* \subset \mathcal{R}$  such that

$$\text{mid}(I_C^*) = \text{mid}(I_C)$$

and

$$L_C^* = \frac{|I_C^*|}{2} < L_C.$$

Using  $I_C^*$ , let the function

$$y = s(x), \quad x \in I_C^*,$$

be such that

$$x^*(y) = x,$$

i.e.  $s(x)$  identifies the agent  $y \in I_C$  for which it is optimal to produce content type  $x$ .

Finally, let

$$x_y^*(y) = \frac{d}{dy}x^*(y), \quad y \in I_C,$$

be the derivative of the function  $x^*(y)$  with respect to  $y$ . Note that by Lemma 10 this derivative is well-defined.

We then have the following two results for  $\beta_C^*(x), x \in I_C^*$ .

**Lemma 14.** *Let  $E_p, 0 < E_p \leq 1$ , and  $E_q, 0 < E_q$ , be given constants, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in I_C$$

and

$$L_C < \frac{L}{2},$$

as well as that

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] > 0, \quad y \in I_C,$$

where  $\alpha_C = 2E_p L_C$ . Assuming that

$$\beta_C(y) = E_q, \quad y \in I_C,$$

where  $\beta_C(y)$  is the bound on the total content production rate of agent  $y$  in community  $C$ , let  $\beta_C^*(\cdot|y)$  be the optimal content production rate of agent  $y \in I_C$  as given in Proposition 11, i.e. we have that

$$\beta_C^*(x|y) = E_q \delta(x - x^*(y)), \quad y \in I_C, x \in \mathcal{R},$$

where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C.$$

Then the density

$$\beta_C^*(x) = \int_{I_C} \beta_C(x|y)dy, \quad x \in \mathcal{R},$$

of the rate at which of content of type  $x$  that is being generated in community  $C$  under the optimal contention production allocations  $\beta_C^*(\cdot|y)$ ,  $y \in I_C$  is given by

$$\beta_C^*(x) = \frac{1}{x_y^*(s(x))} E_q, \quad x \in I_C^*,$$

where  $I_C^*$  is the image of the function  $x^*(y)$ ,  $y \in I_C$ .

*Proof.* By Proposition 3 we have that the function  $x^*(y)$  is strictly increasing on  $I_C$ , and hence it is a bijection on  $I_C$ . It then follows that the inverse function  $s(x)$  of  $x^*(y)$  is well defined on  $I_C^*$ . Furthermore by Lemma 13 we have that the closure  $I_C^*$  of the image of  $x^*(y)$ ,  $y \in I_C$ , is given

$$I_C^* = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*]$$

where

$$0 < L_C^* < L_C.$$

Without loss of generality, we can assume that  $\text{mid}(I_C) = 0$  and we have that

$$I_C = [-L_C, L_C]$$

and

$$I_C^* = [-L_C^*, L_C^*].$$

As we have that

$$\alpha_C(y) = E_p > 0, \quad y \in C_d$$

and

$$\int_{\mathcal{R}} \beta_C^*(x|y) dx = E_q > 0, \quad y \in C_d,$$

as well as

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] > 0, \quad y \in I_C,$$

it follows from Proposition 11 that the optimal production rate allocation for an agent  $y \in I_C$  is given by

$$\beta_C^*(x|y) = E_q \delta(x - x^*(y)), \quad x \in \mathcal{R},$$

where  $d$  is the Dirac delta function,  $x^*(y)$  is the unique solution to the optimization problem

$$\max_{x \in \mathcal{R}} q(x|y)P_C(x).$$

Combing these results with the fact that  $x^*(y)$  is strictly increasing on  $I_C$ , we obtain that

$$\int_{-L_C^*}^x \beta_C^*(\tau) d\tau = \int_{-L_C}^{s(x)} E_q dy$$

and

$$\beta_C^*(x) = \frac{d}{dx} \int_{-L_C}^{s(x)} E_q dy = E_q \frac{d}{dx} s(x).$$

Using well-known results for the derivative of a function [8], the density  $\beta_C^*(x)$  of the rate at which content of type  $x$  that is being generated in community  $C$  under the optimal contention production allocations  $\beta_C^*(\cdot|y)$ ,  $y \in I_C$ , is then given by

$$\beta_C^*(x) = E_q \frac{1}{x_y^*(y = s(x))}, \quad x \in I_C^*,$$

where  $x_y^*(y = s(x))$  is the derivative of the function  $x^*(y)$  with respect to  $y$  evaluated for  $y = s(x)$ . Note that by Proposition 3 the derivative  $x_y^*(y)$ ,  $y \in I_C$ , is well-defined.

This completes the proof of the lemma. □



**Lemma 15.** Let  $E_p$ ,  $0 < E_p \leq 1$ , and  $E_q$ ,  $0 < E_q$ , be given constants, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that

$$\alpha_C(y) = E_p, \quad y \in I_C$$

and

$$L_C < \frac{L}{2},$$

as well as that

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C c] > 0, \quad y \in I_C,$$

where  $\alpha_C = 2E_p L_C$ . Assuming that

$$\beta_C(y) = E_q, \quad y \in I_C,$$

where  $\beta_C(y)$  is the bound on the total content production rate of agent  $y$  in community  $C$ , let  $\beta_C^*(\cdot|y)$  be the optimal content production rate of agent  $y \in I_C$  as given in Proposition 11, i.e. we have that

$$\beta_C^*(x|y) = E_q \delta(x - x^*(y)), \quad y \in I_C, x \in \mathcal{R},$$

where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

and let  $I_C^*$  be the image of the function  $x^*(y)$ ,  $y \in I_C$ .

Then the density

$$\beta_C^*(x) = \int_{I_C} \beta_C(x|y)dy, \quad x \in \mathcal{R},$$

of the rate at which of content of type  $x$  that is being generated in community  $C$  under the optimal contention production allocations  $\beta_C^*(\cdot|y)$ ,  $y \in I_C$ , is continuous, and symmetric with respect to  $\text{mid}(I_C)$ , on  $I_C^*$ .

Furthermore, we have that

$$\beta_C^*(x) > E_q, \quad x \in I_C^*.$$

*Proof.* Without loss of generality, we can assume that the intervals  $I_C$  and  $I_C^*$  are given by

$$I_C = [-L_C, L_C]$$

and

$$I_C^* = [-L_C^*, L_C^*].$$

We first show that the function  $\beta_C^*(x)$ ,  $x \in I_C^*$ , is symmetric with respect to  $\text{mid}(I_C)$ . Let the function  $\Delta^*(y)$  be given by

$$\Delta^*(y) = ||y - x^*(y)||, \quad y \in I_C.$$

By Lemma 11 we have that the function  $\Delta^*(y)$  is symmetric with respect to  $\text{mid}(I_C) = 0$  on  $I_C$ , as well as strictly decreasing on  $[-L_C, 0)$  and strictly increasing on  $(0, L_C]$  with

$$\Delta^*(0) = 0.$$

It then follows that for the derivative

$$\Delta_y^*(y) = \frac{d}{dy} \Delta^*(y), \quad y \in I_C \setminus \{\text{mid}(I_C)\}$$

we have that

$$\Delta_y^*(-y) = -\Delta_y^*(y), \quad y \in I_C \setminus \{\text{mid}(I_C)\}.$$

Using the definition of  $\Delta^*(y)$ , we obtain that

$$x_y^*(y) = \begin{cases} 1 - \Delta_y^*(y), & y \in [0, L_C], \\ 1 + \Delta_y^*(y), & y \in [-L_C, 0]. \end{cases}$$

Therefore in order to show that  $\beta_C^*(x)$  is symmetric with respect to  $\text{mid}(I_C)$  on  $I_C^*$ , we have to show that

$$s(-x) = -s(x), \quad x \in I_C^*.$$

To do this, recall that

$$x^*(y) = \begin{cases} y - \Delta^*(y), & y \in [0, L_C], \\ y + \Delta^*(y), & y \in [-L_C, 0]. \end{cases}$$

Using the result that  $\Delta^*(y)$  is symmetric with respect to  $\text{mid}(I_C)$  on  $I_C$ , we have that

$$-x^*(y) = -y + \Delta^*(y) = -y + \Delta^*(-y) = x^*(-y), \quad y \in [0, L_C].$$

It then follows that

$$-x^*(y) = x^*(-y), \quad y \in I_C$$

and

$$s(-x^*(y)) = s(x^*(-y)) = -y, \quad y \in I_C.$$

As

$$s(x^*(y)) = y,$$

we obtain that

$$s(-x) = -s(x), \quad x \in I_C^*.$$

This proves that  $\beta_C^*(x)$  is symmetric with respect to  $\text{mid}(I_C)$  on  $I_C^*$ .

We next show that the function  $\beta_C^*(x)$  is continuous on  $I_C^*$ . Recall that we have that

$$\beta_C^*(x) = \frac{1}{x_y^*(s(x))} E_q, \quad x \in I_C^*,$$

where  $x_y^*(y)$ ,  $y \in I_C$ , is the derivative of the function

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_C(x), \quad y \in I_C.$$

By Proposition 10 we have that the function  $x^*(y)$  is continuously differentiable on  $I_C$  and that we have that

$$x_y^*(y) > 0, \quad y \in I_C.$$

This result implies that the function  $\beta_C^*(x)$  is continuous on  $I_C^*$ .

It remains to show that

$$\beta_C^*(x) > E_q, \quad x \in I_C^*.$$

From Lemma 12 we have that the function  $\Delta^*(y)$  is strictly decreasing on  $[-L_C, 0)$  and strictly increasing on  $(0, L_C]$ , and

$$|\Delta_y^*(y)| < 1, \quad y \in I_C \setminus \{\text{mid}(I_C)\}.$$

It then follows that

$$\frac{1}{1 - \Delta_y^*(s(x))} > 1, \quad x \in (\text{mid}(I_C), \text{mid}(I_C) + L_C^*]$$

and

$$\frac{1}{1 + \Delta_y^*(s(x))} > 1, \quad x \in [\text{mid}(I_C) - L_C^*, \text{mid}(I_C)),$$

and we obtain that

$$\beta_C^*(x) > E_q, \quad x \in I_C^* \setminus \{\text{mid}(I_C)\}.$$

It remains to show that for

$$y_0 = \text{mid}(I_C)$$

we have that

$$\beta_C^*(y_0) = \frac{1}{x_y^*(s(y_0))} E_q > E_q,$$

or

$$x_y^*(s(y_0)) < 1.$$

From the proof of Lemma 10, we have that

$$x_y^*(s(y_0)) = \frac{g''(0)P_C(0)}{g''(0)P_C(0) + g(0)P_C''(0)}.$$

From Assumption 1 we have that

$$g(0) > 0 \text{ and } g''(0) > 0,$$

and from Proposition 5 we have that

$$P_C(0) > 0 \text{ and } P_C''(0) > 0.$$

Combining these results, it follows that

$$x_y^*(s(y_0)) < 1.$$

This completes the proof of the lemma. □

## F.2 Properties of $Q_C^*(x)$

Having studied in the previous subsection the optimal content production rate allocation for a given interval community  $C = (I_C, I_C)$ , we prove in this subsection the results of Proposition 6 which characterizes the properties of the content supply function  $Q_C^*(x)$  given

$$Q_C^*(x) = \int_{y \in I_C} \beta_C^*(x|y)q(x|y)dy, \quad x \in \mathcal{R},$$

where  $\beta_C^*(\cdot|y)$  is the optimal content production rate of agent  $y$ .

Recall that by Lemma 13 we have for a given interval community  $C = (I_C, I_C)$ , that the image  $I_C^*$  of the function  $x^*(y)$ ,  $y \in I_C$  is given by an interval

$$I_C^* = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*] \subset I_C.$$

Furthermore, recall that the function  $s(x)$ ,  $x \in I_C^*$ , is such that for  $y = s(x)$  we have that

$$x^*(y) = x.$$

We have the following result.

**Lemma 16.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , and  $E_q$ ,  $0 < E_q$ , be given constants, and let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Furthermore assume that*

$$\alpha_C(y) = E_p, \quad y \in I_C$$

and

$$L_C < \frac{L}{2},$$

as well as that

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C] > 0, \quad y \in I_C,$$

where

$$\alpha_C = 2E_p L_C.$$

Assuming that

$$\beta_C(y) = E_q, \quad y \in I_C,$$

where  $\beta_C(y)$  is the bound on the total content production rate of agent  $y$  in community  $C$ , let  $\beta^*(\cdot|y)$  be the optimal content production rate of agent  $y \in I_C$  as given in Proposition 11, i.e. we have that

$$\beta_C^*(x|y) = E_q \delta(x - x^*(y)), \quad y \in I_C, x \in \mathcal{R},$$

where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C,$$

and let  $I_C^*$  be the image of the function  $x^*(y)$ ,  $y \in I_C$ .

The optimal content supply function  $Q_C^*(x)$  is given by

$$Q_C^*(x) = E_q \int_{y \in I_C} \delta(x - x^*(y))q(x^*(y)|y)dy, \quad x \in \mathcal{R}, y \in I_C,$$

and we have that

a) the support  $\text{supp}(Q_C^*(x))$  of  $Q_C^*(x)$ ,  $x \in \mathcal{R}$  is given by

$$\text{supp}(Q_C^*(x)) = I_C^* = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*]$$

where

$$0 < L_C^* < L_C.$$

b) the function  $Q_C^*(x)$  is given by

$$Q_C^*(x) = \begin{cases} \beta_C^*(x)q(x|s(x)), & x \in I_C^*, \\ 0, & x \notin I_C^*, \end{cases}$$

where

$$\beta_C^*(x) = \int_{I_C} \beta_C^*(x|y)dy, \quad x \in I_C^*.$$

c) the function  $Q_C^*(x)$  is continuous, and symmetric with respect to  $\text{mid}(I_C)$ , on  $I_C^*$ .

Note that Proposition 6 follows directly from Lemma 16 as the communities under a Nash equilibrium as given by Proposition 2 satisfy the assumptions made in Lemma 1. In particular as for an interval community  $C = (I_C, I_C) \in \mathcal{C}^*$  under a Nash equilibrium as given by Proposition 2 we have that

$$U_C^{(s)}(y) > 0, \quad y \in C_s = I_C,$$

and

$$\alpha_C(y) = E_p > 0, \quad y \in C_d = I_C,$$

it follows that

$$\max_{x \in \mathcal{R}} [q(x|y)P_C(x) - \alpha_C C] > 0, \quad y \in I_C,$$

where  $\alpha_C = 2E_p L_C$ .

Furthermore, by Proposition 11 we have for an interval community  $C = (I_C, I_C) \in \mathcal{C}^*$  under a Nash equilibrium as given by Proposition 2 that

$$\int_{\mathcal{R}} \beta_C^*(x|y) dx = E_q > 0, \quad y \in C_s = I_C.$$

*Proof.* Note that by Proposition 11, we have that

$$\beta_C^*(x|y) = E_q \delta(x - x^*(y)), \quad x \in \mathcal{R}, y \in I_C,$$

where  $\delta(\cdot)$  is the Dirac delta function, and

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in I_C.$$

Using the result that

$$E_q \delta(x - x^*(y))q(x|y) = E_q \delta(x - x^*(y))q(x^*(y)|y), \quad x \in \mathcal{R}, y \in I_C,$$

we obtain the result that

$$Q_C^*(x) = \int_{y \in I_C} \beta_C^*(x|y)q(x|y)dy = E_q \int_{y \in I_C} \delta(x - x^*(y))q(x^*(y)|y)dy, \quad x \in \mathcal{R}.$$

Let  $I_C^*$  be the image of the function  $x^*(y)$ ,  $y \in I_C$ . By Lemma 13 we have that  $I_C^*$  is given by an interval in  $\mathcal{R}$ , such that

$$\text{mid}(I_C^*) = \text{mid}(I_C)$$

and the half-length  $L_C^*$  of the interval  $I_C^*$  is strictly smaller than  $L_C$ , i.e. we have that

$$L_C^* = \frac{|I_C^*|}{2} < L_C.$$

Combining this result with fact that

$$\beta_C^*(x|y) = E_q \delta(x - x^*(y)), \quad y \in I_C, x \in \mathcal{R},$$

and Proposition 11 which states that

$$q(x^*(y)|y) > 0, \quad y \in I_C,$$

we obtain that

$$\text{supp}(Q_C^*(x)) = I_C^* = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*]$$

with

$$L_C^* < L_C.$$

Next we show that

$$Q_C^*(x) = \beta_C^*(x)q(x|s(x)), \quad x \in I_C^*.$$

Using the same argument as given in the proof of Lemma 14, we have that

$$\int_{-L_C^*}^x Q_C^*(\tau) d\tau = \int_{-L_C}^{s(x)} E_q q(x^*(y)|y) dy$$

and

$$Q_C^*(x) = \frac{d}{dx} \int_{-L_C}^{s(x)} E_q q(x^*(y)|y) dy = E_q q(x|s(x)) \frac{d}{dx} s(x).$$

As we have that

$$\beta_C^*(x) = E_q q(x|s(x)) \frac{d}{dx} s(x), \quad x \in I_C^*,$$

it follows that

$$Q_C^*(x) = \beta_C^*(x) q(x|s(x)), \quad x \in I_C^*,$$

and

$$Q_C^*(x) = 0, \quad x \notin I_C^*,$$

Next we show that the function  $Q_C^*(x)$  is continuous on  $I_C^*$ . As we have that

$$Q_C^*(x) = \beta_C^*(x) q(x|s(x)), \quad x \in I_C^*,$$

it follows that  $Q_C^*(x)$  is continuous on  $I_C^*$  if we can show that the functions  $\beta_C^*(x)$  and  $q(x|s(x))$  are continuous on  $I_C^*$ . By Lemma 15 we have that the function  $\beta_C^*(x)$  is continuous on  $I_C^*$ . By Assumption 1, we have that the function  $g(x)$ ,  $x \in [0, L]$ , is continuous in  $x$ . From Lemma 10 which states that the function  $x^*(y)$ ,  $y \in I_C$ , is continuous in  $y$ , we have that the function  $s(x)$ ,  $x \in I_C^*$ , is continuous in  $x$ . Finally, by Assumption 1 we have that the function  $g(x)$ ,  $x \in [0, L]$ , is continuous in  $x$ , and it follows that the function

$$q(x|s(x)) = g(|x - s(x)|), \quad x \in I_C^*,$$

is continuous in  $x$ . This establishes that the function  $Q_C^*(x)$  is continuous on  $I_C^*$ .

It remains to show that the function  $Q_C^*(x)$  is symmetric with respect to  $mid(I_C)$  on  $I_C^*$ . By Lemma 15, we have that  $\beta_C^*(x)$  is symmetric with respect to  $mid(I_C)$  on  $I_C^*$ , therefore in order to show that  $Q_C^*(x)$  is symmetric with respect to  $mid(I_C)$  on  $I_C^*$ , it suffices to show that  $q(x|s(x))$  is symmetric with respect to  $mid(I_C)$  on  $I_C^*$ . To do this, without loss of generality we can assume that the interval  $I_C$  is given by

$$I_C = [-L_C, L_C]$$

and the interval  $I_C^*$  is given by

$$I_C^* = [-L_C^*, L_C^*],$$

where

$$L_C^* < L_C.$$

Recall that

$$q(x|s(x)) = g(\Delta^*(s(x))).$$

Therefore in order to show that  $q(x|s(x))$  is symmetric with respect to  $mid(I_C) = 0$  on  $I_C^*$ , it suffices to show that

$$g(\Delta^*(s(-x))) = g(\Delta^*(s(x))), \quad x \in I_C^*.$$

By Lemma 11 the function  $\Delta^*(y)$  is symmetric with respect to  $mid(I_C) = 0$  on  $I_C$ , and from the proof of Lemma 15 we have that

$$s(-x) = -s(x), \quad x \in I_C^*.$$

It then follows that

$$g(\Delta^*(s(-x))) = g(\Delta^*(-s(x))) = g(\Delta^*(s(x))), \quad x \in I_C^*,$$

and we obtain that  $Q_C^*(x)$  is symmetric with respect to  $mid(I_C) = 0$  on  $I_C^*$ .

This completes the proof of the lemma. □

## G Proof of Proposition 7

In this appendix we prove Proposition 7 which characterize the utility rates  $U_C^{(d)}(y)$  for content consumption under a Nash Equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2.

To do this we consider the case where we are given a set  $C_s = I_C$  of content producers where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Note that this situation applies to the Nash equilibrium given by Proposition 2 where all communities are given by interval communities, i.e. the set of agents that produce (and consume) content in a community  $C$  is given by an interval  $I_C$  in  $\mathcal{R}$ . For this situation, we assume that we are given the content supply function  $Q_C(x)$ ,  $x \in \mathcal{R}$ , for the content produced by agents  $y \in C_s = I_C$ , and then characterize then the maximal utility of content consumption that a agent  $y \in \mathcal{R}$  can obtain under the content supply function  $Q_C(x)$ .

Our first result characterize the maximal content consumption utility function  $U_C^{(d)}(y)$ ,  $y \in \mathcal{R}$ , under the additional assumption that the content production rates  $\beta_C(x|y)$  of the agents in the set  $C_s = I_C$  have the property that

$$\int_{\mathcal{R}} \beta_C(x|y) dx = E_q > 0.$$

For this case, we have the following result.

**Lemma 17.** *Let  $E_q$ ,  $0 < E_q$ , be a given constant, and let  $C_s = I_C$  be a set of content producers where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Assume that*

$$\int_{\mathcal{R}} \beta_C(x|y) dx = E_q > 0, \quad y \in I_C,$$

where  $\beta_C(x|y)$ ,  $y \in I_C$  and  $x \in \mathcal{R}$ , is the rate allocation of an agent  $y \in I_C$ . Then the maximal utility rate for content consumption that an agent  $y \in \mathcal{R}$  obtains from the producers  $y \in I_C$  is given by

$$U_C^{(d)}(y) = \max \{0, F_C^{(d)}(y)\}, \quad y \in \mathcal{R},$$

where

$$F_C^{(d)}(y) = E_p \int_{\mathcal{R}} [p(x|y)Q_C(x) - 2L_C E_q c] dx, \quad y \in \mathcal{R},$$

and  $E_p$ ,  $0 < E_p \leq 1$ , is the maximal fraction of time that agent  $y$  can allocate to consuming content from the producers  $y \in C_s = I_C$ , and

$$Q_C(x) = \int_{I_C} \beta_C(x|y) dy, \quad x \in \mathcal{R}.$$

*Proof.* The result of this lemma follows directly from the definition of the utility rate for content consumption of an agent  $y \in \mathcal{R}$  given by

$$U_C^{(d)}(y) = \alpha_C(y) \int_{\mathcal{R}} [Q_C(x)p(x|y) - \beta_C(x|y)c] dx,$$

and the assumptions that

$$0 \leq \alpha_C(y) \leq E_p$$

and

$$\int_{\mathcal{R}} \beta_C(x|y) dx = E_q > 0, \quad y \in I_C.$$

In particular, under the assumption that

$$\int_{\mathcal{R}} \beta_C(x|y) dx = E_q > 0, \quad y \in I_C,$$

we have that

$$\int_{\mathcal{R}} \beta_C(x) = \int_{y \in I_C} \int_{\mathcal{R}} \beta_C(x|y) dx dy = 2L_C E_q.$$

□

Using the result of this lemma, we first characterize in the next the properties of the function  $F_C^{(d)}(y)$  which we then use then to characterize the properties of the utility rates for content consumption  $U_C^{(d)}(y)$ , and to prove Proposition 7.

### G.1 Properties of the function $F_C^{(d)}(y)$

In this subsection, we characterize the properties of the function  $F_C^{(d)}(y)$ ,  $y \in \mathcal{R}$ , for the special case where we have that

$$Q_C(x) = Q_C^*(x), \quad x \in \mathcal{R},$$

and  $Q_C^*(x)$  is as given in Proposition 6, i.e. we assume that the content supply function  $Q_C^*(x)$ ,  $x \in \mathcal{R}$ , has the same properties as the optimal supply function for an interval community in a Nash equilibrium as given by Proposition 2.

For this case, we have the following result.

**Lemma 18.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , and  $E_q$ ,  $0 < E_q$ , be given constants, and let  $I_C$  be an interval in  $\mathcal{R}$ . For*

$$L_C = \frac{|I_C|}{2},$$

assume that

$$2L_C < \min\{b, L\}$$

where  $b$  is the constant of Assumption 1. Furthermore, assume for the function  $Q_C^*(x)$ ,  $x \in \mathcal{R}$ , that

1.  $Q_C^*(x)$  is bounded on  $\mathcal{R}$ .
2. the support  $\text{supp}(Q_C^*)$  of  $Q_C^*(x)$  is given by

$$\text{supp}(Q_C^*) = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*]$$

where

$$0 < L_C^* < L_C,$$

3.  $Q_C^*(x)$  is non-negative, continuous, and symmetric with respect to  $\text{mid}(I_C)$ , on  $(\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*)$ .

Then the function  $F_C^{(d)}(y)$  given by

$$F_C^{(d)}(y) = E_p \int_{\mathcal{R}} [p(x|y)Q_C^*(x)dx - 2L_C E_q c] dx, \quad y \in \mathcal{R},$$

has the properties that

- a)  $F_C^{(d)}(y)$  is symmetric with respect to  $y_0 = \text{mid}(I_C)$  on  $\mathcal{R}$ , and we have for  $y = \text{mid}(I_C)$  that

$$\frac{d}{dy} F_C^{(d)}(y) = 0.$$

- b)  $F_C^{(d)}(y)$  is strictly increasing on  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ , and we have that

$$\frac{d}{dy} F_C^{(d)}(y) > 0, \quad y \in (\text{mid}(I_C) - L, \text{mid}(I_C)).$$

- c)  $F_C^{(d)}(y)$  is strictly decreasing on  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ , and we have that

$$\frac{d}{dy} F_C^{(d)}(y) < 0, \quad y \in (\text{mid}(I_C), \text{mid}(I_C) + L).$$

*Proof.* Let

$$I_C^* = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*].$$

Note that the assumptions on the function  $Q_C^*(x)$ ,  $x \in \mathcal{R}$ , made in the lemma imply that

$$0 < \int_{I_C^*} Q_C^*(x) dx.$$



Without loss of generality, we assume for our proof that the function  $Q_C^*(x)$  is non-negative, continuous, and symmetric with respect to  $\text{mid}(I_C)$ , on  $I_C^* = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*]$

We first show that  $F_C^{(d)}(y)$  is symmetric with respect to  $y_0 = \text{mid}(I_C)$  on  $\mathcal{R}$ . Note that we have that

$$p(\text{mid}(I_C) - x | \text{mid}(I_C) - y) = p(\text{mid}(I_C) + x | \text{mid}(I_C) + y) = f(\|x - y\|), \quad x, y \in \mathcal{R}.$$

Combining this result with the assumption that  $Q_C^*(x)$  is symmetric with respect to  $\text{mid}(I_C)$  on  $(\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*)$ , and letting

$$y_0 = \text{mid}(I_C),$$

we obtain

$$\begin{aligned} F_C^{(d)}(y_0 + y) &= E_p \int_{\mathcal{R}} [p(x|y_0 + y)Q_C^*(x)dx - 2L_C E_q c] dx \\ &= E_p \int_{-L_C^*}^{L_C^*} [p(y_0 + x|y_0 + y)Q_C^*(y_0 + x)dx - 2L_C E_q c] dx \\ &= E_p \int_{-L_C^*}^{L_C^*} [p(y_0 - x|y_0 - y)Q_C^*(y_0 - x)dx - 2L_C E_q c] dx \\ &= E_p \int_{\mathcal{R}} [p(x|y_0 - y)Q_C^*(x)dx - 2L_C E_q c] dx \\ &= F_C^{(d)}(y_0 - y), \quad y \in [0, L]. \end{aligned}$$

It then follows that  $F_C^{(d)}(y)$  is symmetric with respect to  $y_0 = \text{mid}(I_C)$  on  $\mathcal{R}$ .

Next we prove that the function  $F_C^{(d)}(y)$  is strictly increasing on  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ , and strictly decreasing on  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ . As  $F_C^{(d)}(y)$  is symmetric with respect to  $y_0 = \text{mid}(I_C)$  on  $\mathcal{R}$ , it suffices to show that  $F_C^{(d)}(y)$  is strictly increasing on  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ , and we have that

$$\frac{d}{dy} F_C^{(d)}(y) > 0, \quad y \in (\text{mid}(I_C) - L, \text{mid}(I_C)).$$

Therefore, we assume for the remainder of the proof that

$$y < \text{mid}(I_C).$$

We have that

$$\frac{d}{dy} F_C^{(d)}(y) = \int_{I_C^*} Q_C^*(x) \frac{d}{dy} p(x|y) dx,$$

where  $I_C^*$  is the support of  $Q_C^*(x)$ ,  $x \in \mathcal{R}$ . Recall that by assumption we have that

$$I_C^* = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*]$$

where

$$0 < L_C^* < L_C < \frac{L}{2}.$$

In the following, we consider three separate cases.

The first case that we consider is where we have that

$$y \in [\text{mid}(I_C) + L_C^* - L, \text{mid}(I_C) - L_C^*].$$

As we have that

$$0 < L_C^* < L_C < \frac{L}{2},$$

we obtain

$$\text{mid}(I_C) + L_C^* - L < \text{mid}(I_C) - L_C^*.$$

For this case we have that

$$\frac{d}{dy}p(x|y) > 0$$

as by Assumption 1 the function  $f$  is strictly decreasing on  $[0, L]$ . Combining this result we the fact that

$$\int_{I_C^*} Q_C^*(x)dx > 0,$$

it follows that

$$\frac{d}{dy}F_C^{(d)}(y) = \int_{I_C^*} Q_C^*(x) \frac{d}{dy}p(x|y)dx < 0, \quad y \in [\text{mid}(I_C) + L_C^* - L, \text{mid}(I_C) - L_C^*].$$

Next we consider the case where

$$y \in [\text{mid}(I_C) - L_C^*, \text{mid}(I_C)] = [y_0 - L_C^*, y_0].$$

For this case we obtain that

$$\begin{aligned} \frac{d}{dy} F_C^{(d)}(y) &= \int_{I_C^*} Q_C^*(x) \frac{d}{dy} p(x|y) dx \\ &= \int_{y_0 - L_C^*}^y Q_C^*(x) f'(y-x) dx - \dots \\ &\quad - \int_0^{y_0 - y} Q_C^*(y_0 + s) \left( f'(y_0 - y + s) + f'(y_0 - y - s) \right) ds - \dots \\ &\quad - \int_{-y}^{y_0 + L_C^*} Q_C^*(x) f'(x-y) dx \\ &= - \int_0^{y_0 - y} Q_C^*(y_0 + s) \left( f'(y - y_0 - s) + f'(y - y_0 + s) \right) ds + \dots \\ &\quad + \int_0^{y - y_0 + L_C^*} Q_C^*(y - s) \left( f'(s) - f'(2(y_0 - y) + s) \right) ds. \end{aligned}$$

By Assumption 1 the function  $f$  is strictly decreasing on  $[0, L]$ , and it follows that

$$f'(y - y_0 + s) + f'(y - y_0 - s) < 0, \quad s \in [0, y - y_0],$$

and

$$- \int_0^{y - y_0} Q_C^*(y_0 + s) \left( f'(y - y_0 + s) + f'(y - y_0 - s) \right) ds > 0.$$

Next we show that

$$\int_0^{y - y_0 + L_C^*} Q_C^*(s - y) \left( f'(s) - f'(2(y_0 - y) + s) \right) ds > 0.$$

To do this, we note that

$$2(y_0 - y) + s \leq 2L_C^* < 2L_C, \quad s \in [0, y - y_0 + L_C^*].$$

By Assumption 1 the function  $f$  is strictly concave on  $[0, b]$ . As by construction we have that

$$2L_C < b,$$

it follows that

$$f'(2(y_0 - y) + s) < f'(s), \quad s \in [0, y - y_0 + L_C^*]$$

and

$$\int_0^{y - y_0 + L_C^*} Q_C^*(y - s) \left( f'(s) - f'(2(y_0 - y) + s) \right) ds > 0.$$

Finally, we consider the case where

$$y \in (\text{mid}(I_C) - L, \text{mid}(I_C) + L_C^* - L) = [y_0 - L, y_0 + L_C^* - L).$$

Let

$$\Delta_L = y - (y_0 - L).$$

Recall that

$$\text{mid}(I_C) + L_C^* - L < \text{mid}(I_C) - L_C^*,$$

as by assumption we have that

$$0 < L_C^* < L_C < \frac{L}{2}.$$

We then obtain that

$$\begin{aligned} \frac{d}{dy} F_C^{(d)}(y) &= \int_{I_C^*} Q_C^*(x) \frac{d}{dy} p(x|y) dx \\ &= - \int_{y_0 - L_C^*}^{y_0 - \Delta_L} Q_C^*(x) f'(x - y) dx - \dots \\ &\quad - \int_0^{\Delta_L} Q_C^*(y_0 + s) \left( f'(y_0 - y - s) + f'(y_0 - y + s) \right) ds + \dots \\ &\quad + \int_{y_0 + \Delta_L}^{y_0 + L_C^*} Q_C^*(x) f'(x - y) dx \\ &= - \int_0^{L_C^* - \Delta_L} Q_C^*(y_0 - L_C^* + s) \left( f'(L - L_C^* + \Delta_L + s) - f'(L - L_C^* + \Delta_L + s) \right) ds - \dots \\ &\quad - \int_0^{\Delta_L} Q_C^*(y_0 + s) \left( f'(y_0 - y - s) + f'(y_0 - y + s) \right) ds \\ &= - \int_0^{\Delta_L} Q_C^*(y_0 + s) \left( f'(y_0 - y - s) + f'(y_0 - y + s) \right) ds. \end{aligned}$$

By Assumption 1 the function  $f(\cdot)$  is strictly decreasing on  $[0, L]$ , and it follows that

$$- \int_0^{\Delta_L} Q_C^*(y_0 + s) \left( f'(y_0 - y - s) + f'(y_0 - y + s) \right) ds > 0, \quad y \in (y_0 - L, y_0 + L_C^* - L).$$

It then follows that we have that

$$\frac{d}{dy} F_C^{(d)}(y) > 0, \quad y \in (\text{mid}(I_C) - L, \text{mid}(I_C)).$$

As by assumption the function  $Q_C^*(x)$  is symmetric with respect to  $mid(I_C)$ , and we have that

$$p(y_0 - s|y) = p(y_0 + s|y) = f(s), \quad s \in [0, L],$$

we obtain that derivative  $\frac{d}{dy}F_C^{(d)}(y)$  at  $y = y_0$  is equal to

$$\int_0^{L_C} Q_C^*(y_0 - s) [-f'(s) + f'(s)] ds = 0.$$

The result of the lemma then follows. □

## G.2 Properties of the Function $U_C^{(d)}(y)$

We next characterize the properties of the function

$$U_C^{(d)}(y) = \max \left\{ 0, F_C^{(d)}(y) \right\}, \quad y \in \mathcal{R},$$

where

$$F_C^{(d)}(y) = E_p \int_{\mathcal{R}} \left[ p(x|y) Q_C^*(x) dx - 2L_C E_q c \right] dx, \quad y \in \mathcal{R},$$

and  $Q_C^*(x)$  is as given in Proposition 6.

Compared with Lemma 18, we add for this analysis the assumption that

$$F_C^{(d)}(y) = E_p \int_{\mathcal{R}} \left[ p(x|y) Q_C^*(x) dx - 2L_C E_q c \right] dx > 0, \quad y \in I_C.$$

The assumption

$$U_C^{(d)}(y) > 0, \quad y \in I_C,$$

reflects the result of Proposition 2 which states that under a Nash equilibrium all agents have strictly positive utility rates. We then have the following result.

**Lemma 19.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , and  $E_q$ ,  $0 < E_q$ , be given constants, and let  $I_C$  be an interval in  $\mathcal{R}$ . For*

$$L_C = \frac{|I_C|}{2},$$

assume that

$$2L_C < \min\{b, L\}$$

where  $b$  is the constant of Assumption 1. Furthermore assume for the function  $Q_C^*(x)$ ,  $x \in \mathcal{R}$ , that

1.  $Q_C^*(x)$ ,  $x \in \mathcal{R}$ , is bounded on  $\mathcal{R}$ .
2. the support  $\text{supp}(Q_C^*)$  of  $Q_C^*(x)$  is given by

$$\text{supp}(Q_C^*) = [\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*]$$

where

$$0 < L_C^* < L_C,$$

3.  $Q_C^*(x)$  is non-negative, continuous, and symmetric with respect to  $\text{mid}(I_C)$ , on  $(\text{mid}(I_C) - L_C^*, \text{mid}(I_C) + L_C^*)$ .

Let

$$U_C^{(d)}(y) = \max \left\{ 0, F_C^{(d)}(y) \right\}, \quad y \in \mathcal{R},$$

where

$$F_C^{(d)}(y) = E_p \left[ \int_{\mathcal{R}} p(x|y) Q_C^*(x) dx - 2L_C E_q c \right], \quad y \in \mathcal{R}.$$

If we have that

$$U_C^{(d)}(y) > 0, \quad y \in I_C,$$

then the function  $U_C^{(d)}(y)$ ,  $y \in \mathcal{R}$ , has the properties that

- a)  $U_C^{(d)}(y)$  is symmetric with respect to  $y_0 = \text{mid}(I_C)$  on  $\mathcal{R}$ , and we have for  $y = \text{mid}(I_C)$  that

$$\frac{d}{dy} U_C^{(d)}(\text{mid}(I_C)) = 0.$$

- b)  $U_C^{(d)}(y)$  is strictly increasing on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C))$ , and we have that

$$\frac{d}{dy} U_C^{(d)}(y) > 0, \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)).$$

c)  $U_C^{(d)}(y)$  is strictly decreasing on  $(\text{mid}(I_C), \text{mid}(I_C) + L_C]$ , and we have that

$$\frac{d}{dy}U_C^{(d)}(y) < 0, \quad y \in (\text{mid}(I_C), \text{mid}(I_C) + L_C].$$

d)  $U_C^{(d)}(y)$  is non-decreasing on  $[\text{mid}(I_C) - L, \text{mid}(I_C))$ .

e)  $U_C^{(d)}(y)$  is non-increasing on  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ .

*Proof.* The result of the lemma follows directly from the result of Lemma 18 and the the fact that

$$U_C^{(d)}(y) = \max \left\{ 0, F_C^{(d)}(y) \right\}, \quad y \in \mathcal{R}.$$

□

### G.3 Proof of Proposition 7

The result of Proposition 7 follows from Lemma 19 as the setting of Proposition 7 satisfies all the assumptions made in the statement of Lemma 19.

In particular, Proposition 6 states the the content supply function  $Q_C^*(x)$  satisfies the assumptions made on  $Q_C^*(x)$  in Lemma 19. Moreover, Proposition 2 states that under a Nash equilibrium all agents receive a strictly positive utility, and we have that for a community  $C = (I_C, I_C) \in \mathcal{C}^*$  in a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2 that

$$U_C^{(d)}(y) > 0, \quad y \in I_C.$$

Finally, by Proposition 3 we have for a community  $C = (I_C, I_C) \in \mathcal{C}^*$  under a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2 we have that

$$\int_{\mathcal{R}} \beta_C^*(x|y) dx = E_q, \quad y \in C_s = I_C,$$

and

$$\alpha_C^*(y) = E_p, \quad y \in C_d = I_C.$$



## H Proof of Proposition 8

In this appendix we prove Proposition 8 which characterize the utility rates for content production under a Nash Equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2.

To do that, we use the same approach as in the previous subsection where we characterized the utility rates for content consumption. That is, we consider the case where we are given a set  $C_d = I_C$  of content consumers where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . In addition, we assume that we are given the content demand function  $P_C(x)$ ,  $x \in \mathcal{R}$ , for the content consumed by the agents  $y \in C_d = I_C$ , and then characterize then the maximal utility of content consumption that a agent  $y \in \mathcal{R}$  can obtain under the content demand function  $P_C(x)$ .

Our first result characterize the utility function under the additional assumption that the for the fraction of time  $\alpha_C(y)$  that agents in the set  $C_s = I_C$  allocate for content consumption has the property that

$$\alpha_C(y) = E_p > 0, \quad y \in I_C,$$

For this case, we have the following result.

**Lemma 20.** *Let  $E_p$ ,  $0 < E_q \leq 1$ , be a given constant, and let  $C_d = I_C$  be a set of content consumers where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ . Assume that*

$$\alpha_C(y) = E_p > 0, \quad y \in I_C,$$

where  $\alpha_C(y)$ ,  $y \in C_d = I_C$ , is the fraction of time an agent  $y$  consumes content. Then the maximal utility rate for content producer that an agent  $y \in \mathcal{R}$  obtains from the consumers  $y \in I_C$  is given by

$$U_C^{(s)}(y) = \max \left\{ 0, F_C^{(s)}(y) \right\}, \quad y \in \mathcal{R},$$

where

$$F_C^{(s)}(y) = E_q \left[ q(x^*(y)|y) P_C(x^*(y)) - 2L_C E_p c \right], \quad y \in \mathcal{R},$$

and  $E_q$ ,  $0 < E_q$ , is the maximal total content production rate that agents  $y \in \mathcal{R}$  can allocated to the set of agents  $C_d = I_C$ , and

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_C(x).$$

*Proof.* The result of this lemma follows directly from the definition of the utility rate for content production of an agent  $y \in \mathcal{R}$  given by

$$U_C^{(d)}(y) = \int_{\mathcal{R}} \beta_C^*(x|y) \left[ q(x|y) P_C(x) p(x|y) - \alpha_C c \right] dx,$$

where  $\beta_C^*(\cdot|y)$  is the optimal content production rate given by

$$\beta_C^*(\cdot|y) = \arg \max_{\beta(\cdot|y): \|\beta(\cdot|y)\| \leq \beta_C(y)} \int_{\mathcal{R}} \beta_C(x|y) \left[ q(x|y) P_C(x) - \alpha_C c \right] dx,$$

and the assumptions that

$$0 \leq \beta_C(y) \leq E_q$$

and

$$\alpha_C(y) = E_p > 0, \quad y \in I_C.$$

In particular, under the assumption that

$$\alpha_C(y) = E_p > 0, \quad y \in I_C,$$

we obtain that

$$\alpha_C = \int_{I_C} \alpha_C(y) dy = 2L_C E_p.$$

Furthermore, by Corollary 1 we have that

$$\beta_C^*(\cdot|y) = E_q \delta(x^*(y) - x), \quad y \in I_C,$$

where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_C(x),$$

is a solution to the optimization problem

$$\max_{\beta(\cdot|y): \|\beta(\cdot|y)\| \leq E_q} \int_{\mathcal{R}} \beta_C(x|y) \left[ q(x|y) P_C(x) - \alpha_C c \right] dx.$$

□

## H.1 Properties of the Function $F_C^{(s)}(y)$

Next we characterize the properties of the function  $F_C^{(s)}(y)$ ,  $y \in \mathcal{R}$ , for the special case where we have that the function  $P_C(x)$  is as given in Proposition 5, i.e. we assume that the content demand function  $P_C(x)$ ,  $x \in \mathcal{R}$ , has the same properties as the content demand function for an interval community in a Nash equilibrium as given by Proposition 2.

For this case, we have the following two results.

**Lemma 21.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , and  $E_q$ ,  $0 < E_q$ , be given constants, and let  $I_C$  be an interval in  $\mathcal{R}$ . For*

$$L_C = \frac{|I_C|}{2},$$

assume that

$$L_C < \frac{L}{2}$$

Furthermore, assume for the function  $P_C(x)$ ,  $x \in \mathcal{R}$ , that

- (a)  $P_C(x)$  is positive and twice continuously differentiable on  $\mathcal{R}$ .
- (b)  $P_C(x)$  is symmetric with respect to  $\text{mid}(I_C)$ .
- (c)  $P_C(x)$  is strictly increasing on the interval  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ , and strictly decreasing on the interval  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ .
- (d)  $P_C(x)$  is strictly concave in  $x$  on the interval  $[\text{mid}(I_C) - L_C, \text{mid}(I_C) + L_C]$ .

Then the function  $F_C^{(s)}(y)$  given by

$$F_C^{(s)}(y) = E_q[q(x^*(y)|y)P_C(x^*(y)) - 2L_C E_p c], \quad y \in \mathcal{R},$$

where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x),$$

has the properties that

- a)  $F_C^{(s)}(y)$  is symmetric with respect to  $y_0 = \text{mid}(I_C)$  on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C) + L_C]$ , and we have that

$$\frac{d}{dy} F_C^{(s)}(\text{mid}(I_C)) = 0.$$

- b)  $F_C^{(s)}(y)$  is strictly increasing on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C))$ , and we have that

$$\frac{d}{dy} F_C^{(s)}(y) > 0, \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)).$$

- c)  $F_C^{(s)}(y)$  is strictly decreasing on  $(\text{mid}(I_C), \text{mid}(I_C) + L_C]$ , and we have that

$$\frac{d}{dy} F_C^{(s)}(y) < 0, \quad y \in (\text{mid}(I_C), \text{mid}(I_C) + L_C].$$

*Proof.* Let

$$y_0 = \text{mid}(I_C).$$

We first show that  $F_C^{(s)}(y)$  is symmetric with respect to  $y_0 = y_0$  on  $[y_0 - L_C, y_0 + L_C]$ . For this we rewrite  $F_C^{(s)}(y)$  as

$$F_C^{(s)}(y) = E_q \left[ g(\Delta^*(y)) P_C(y - \Delta^*(y)) - 2L_C E_p c \right], \quad y \in [y_0, y_0 + L_C],$$

and

$$F_C^{(s)}(y) = E_q \left[ g(\Delta^*(y)) P_C(y + \Delta^*(y)) - 2L_C E_p c \right], \quad y \in [y_0 - L_C, y_0],$$

where

$$\Delta^*(y) = \|y - x^*(y)\|, \quad y \in [y_0 - L_C, y_0 + L_C].$$

By the same argument as given to prove Lemma 12, we have that the function  $\Delta^*(y)$  is symmetric with respect to  $y_0$  on  $[y_0 - L_C, y_0 + L_C]$ . Furthermore, by assumption we have that the function  $P_C(x)$  is symmetric with respect to  $y_0$  on  $\mathcal{R}$ . It then follows that for  $s \in [0, L_C]$  we have

$$\begin{aligned} F_C^{(s)}(y_0 + s) &= E_q[g(\Delta^*(y_0 + s))P_C(y_0 + s - \Delta^*(y_0 + s)) - 2L_C E_p c] \\ &= E_q[g(\Delta^*(y_0 - s))P_C(y_0 + s - \Delta^*(y_0 - s)) - 2L_C E_p c] \\ &= E_q[g(\Delta^*(y_0 - s))P_C(y_0 - s + \Delta^*(y_0 - s)) - 2L_C E_p c] \\ &= F_C^{(s)}(y_0 - s). \end{aligned}$$

It then follows that the function  $F_C^{(s)}(y)$  symmetric with respect to  $y = y_0$  on  $[y_0 - L_C, y_0 + L_C]$ .

Next we show that  $F_C^{(s)}(y)$  is strictly increasing on  $[y_0 - L_C, y_0)$ , and strictly decreasing on  $(y_0, y_0 + L_C]$ . As the function  $F_C^{(s)}(y)$  is symmetric with respect to  $y_0$  on  $[y_0 - L_C, y_0 + L_C]$ , it suffices to prove the result for

$$y \in (y_0, y_0 + L_C],$$

i.e. to show that

$$\frac{d}{dy} F_C^{(s)}(y) < 0, \quad y \in (y_0, y_0 + L_C].$$

We then obtain that

$$\frac{d}{dy} F_C^{(s)}(y) = \frac{d}{dy} q(x^*(y)|y)P_C(x^*(y)) = \frac{d}{dy} g(\Delta^*(y))P_C(y - \Delta^*(y)), \quad y \in (y_0, y_0 + L_C].$$

As by assumption the function  $P_C(x)$  is strictly concave on  $[y_0, y_0 + L_C]$ , as well as is strictly increasing on the interval  $(y_0 - L, y_0)$ , and strictly decreasing on the interval  $(y_0, y_0 + L)$ , it follows by the same argument as given to prove Lemma 10 that the optimization problem

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x), \quad y \in (y_0, y_0 + L_C],$$

has a unique solution. Furthermore, by the same argument as given to prove Lemma 12, we obtain that the function  $\Delta^*(y)$  is strictly increasing on  $[y_0, y_0 + L_C]$ .

Let

$$\Delta_y^*(y) = \frac{d}{dy} \Delta^*(y), \quad y \in (y_0, y_0 + L_C],$$

and let  $P'_C(x)$  be the derivative of  $P_C(x)$  with respect to  $x$ . By the same argument as given in the proof of Lemma 12, we have that the derivative  $\Delta_y^*(y)$  is well-defined on  $(y_0, y_0 + L_C]$ . Furthermore, by assumption the derivative  $P'_C(x)$  is well defined on  $\mathcal{R}$ .

Using these definitions, we obtain that

$$\begin{aligned} \frac{d}{dy} g(\Delta^*(y))P_C(y - \Delta^*(y)) &= g'(\Delta^*(y))P_C(y - \Delta^*(y))\Delta_y^*(y) - \dots \\ &\quad - g(\Delta^*(y))P'_C(y - \Delta^*(y))\Delta_y^*(y) + \dots \\ &\quad + g(\Delta^*(y))P'_C(y - \Delta^*(y)) \\ &= g'(\Delta^*(y))P_C(y - \Delta^*(y))\Delta_y^*(y) + \dots \\ &\quad + g(\Delta^*(y))P'_C(y - \Delta^*(y))\left[1 - \Delta_y^*(y)\right]. \end{aligned}$$

By Assumption 1 we have that

$$g'(\Delta^*(y)) < 0, \quad y \in (y_0, y_0 + L_C] \cap \text{supp}(g).$$

By the same argument as given to prove Lemma 12, we obtain that the function  $\Delta^*(y)$  is strictly increasing on  $(y_0, y_0 + L_C]$  and we have that

$$\Delta_y^*(y) > 0, \quad y \in (y_0, y_0 + L_C].$$

and

$$|\Delta_y^*(y)| < 1, \quad y \in (y_0, y_0 + L_C].$$

Furthermore, by assumption we have that

$$P'_C(y) < 0, \quad y \in (y_0, y_0 + L_C].$$

Combining these results, it follows that

$$1 - \Delta_y^*(y) > 0, \quad y \in (y_0, y_0 + L_C],$$

and

$$\begin{aligned} \frac{d}{dy} g(\Delta^*(y)) P_C(y - \Delta^*(y)) &= g'(\Delta^*(y)) P_C(y - \Delta^*(y)) \Delta_y^*(y) + \dots \\ &\quad + g(\Delta^*(y)) P'_C(y - \Delta^*(y)) [1 - \Delta_y^*(y)] \\ &< 0, \quad y \in (y_0, y_0 + L_C]. \end{aligned}$$

Finally, by Assumption 1 we have that

$$g'(0) = 0.$$

Furthermore, as by assumption we have the  $P_C(x)$  is symmetric with respect to  $\text{mid}(I_C)$  and  $P_C(x)$  is strictly concave on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C) + L_C]$ , it follows that

$$P'_C(\text{mid}(I_C)) = 0,$$

and by the same argument as given to prove Lemma 8 we have that

$$x^*(y_0) = y_0$$

and

$$\Delta^*(y_0) = 0.$$

Combining these results, it follows from the above expression of the derivative of  $F_C^{(s)}(y)$  that for  $y = y_0$  we have that

$$\frac{d}{dy} F_C^{(s)}(y) = 0.$$

The result of the lemma then follows. □

**Lemma 22.** Let  $E_p, 0 < E_p \leq 1$ , and  $E_q, 0 < E_q$ , be given constants, and let  $I_C$  be an interval in  $\mathcal{R}$ . For

$$L_C = \frac{|I_C|}{2},$$

assume that

$$L_C < \frac{L}{2}$$

Furthermore, assume for the function  $P_C(x)$ ,  $x \in \mathcal{R}$ , that

- (a)  $P_C(x)$  is positive and twice continuously differentiable on  $\mathcal{R}$ .
- (b)  $P_C(x)$  is symmetric with respect to  $\text{mid}(I_C)$ .
- (c)  $P_C(x)$  is strictly increasing on the interval  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ , and strictly decreasing on the interval  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ .

Then the function  $F_C^{(s)}(y)$  given by

$$F_C^{(s)}(y) = E_q \left[ q(x^*(y)|y)P_C(x^*(y)) - 2L_C E_p c \right], \quad y \in \mathcal{R},$$

where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x),$$

has the properties that

- a)  $F_C^{(s)}(y)$  is non-decreasing on  $[\text{mid}(I_C) - L, \text{mid}(I_C) - L_C]$ .
- b)  $F_C^{(s)}(y)$  is non-increasing on  $[\text{mid}(I_C) + L_C, \text{mid}(I_C) + L]$ .

*Proof.* We prove the result for  $y \in [\text{mid}(I_C) + L_C, \text{mid}(I_C) + L]$ , i.e. to prove that  $F_C^{(s)}(y)$  is non-increasing on  $[\text{mid}(I_C) + L_C, \text{mid}(I_C) + L]$ . The result that  $F_C^{(s)}(y)$  is non-decreasing on  $[\text{mid}(I_C) - L, \text{mid}(I_C) - L_C]$  can be obtained by the same argument.

Consider two agents  $y, y'$  such that

$$y, y' \geq y_0 + L_C$$

and

$$y' > y,$$

and show that for

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x)$$

and

$$x^*(y') = \arg \max_{x \in \mathcal{R}} q(x|y')P_C(x),$$

we have that

$$q(x^*(y)|y)P_C(x^*(y)) > q(x^*(y')|y')P_C(x^*(y')).$$

We prove this result by contradiction as follows. Suppose that the result is not true and we have that

$$q(x^*(y)|y)P_C(x^*(y)) \leq q(x^*(y')|y')P_C(x^*(y'));$$

in this case we show that there exists a value  $\hat{x}_y$  such that

$$q(\hat{x}_y|y)P_C(\hat{x}_y) > q(x^*(y')|y')P_C(x^*(y'))$$

which contradicts our assertion that

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x).$$

In particular, let  $\hat{x}_y$  be given by

$$\hat{x}_y = y - (y' - x^*(y')).$$

We then have that

$$q(x^*(y')|y')P_C(x^*(y')) < q(\hat{x}_y|y)P_C(\hat{x}_y).$$

To see that this is true note that

$$q(x^*(y')|y') = q(\hat{x}_y|y) = g(|y' - x^*(y')|)$$

and

$$P_C(x^*(y')) < P_C(\hat{x}_y),$$

as by assumption the function  $P_C(x)$  is strictly decreasing for  $x \geq \text{mid}(I_C)$ .

The result of the lemma then follows. □

We next characterize the properties of the function

$$U_C^{(s)}(y) = \max \left\{ 0, F_C^{(s)}(y) \right\}, \quad y \in \mathcal{R},$$

where

$$F_C^{(s)}(y) = E_q[q(x^*(y)|y)P_C(x) - E_p c], \quad y \in \mathcal{R},$$

where

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x).$$

Similar to the analysis in the previous subsection, we add for this analysis the assumption that

$$U_C^{(s)}(y) > 0, \quad y \in I_C.$$

We have the following result.

**Lemma 23.** *Let  $E_p$ ,  $0 < E_p \leq 1$ , and  $E_q$ ,  $0 < E_q$ , be given constants, and let  $I_C$  be an interval in  $\mathcal{R}$ . For*

$$L_C = \frac{|I_C|}{2},$$

assume that

$$L_C < \frac{L}{2}$$

Furthermore, assume for the function  $P_C(x)$ ,  $x \in \mathcal{R}$ , that

- (a)  $P_C(x)$  is positive and twice continuously differentiable on  $\mathcal{R}$ .
- (b)  $P_C(x)$  is symmetric with respect to  $\text{mid}(I_C)$ .
- (c)  $P_C(x)$  is strictly increasing on the interval  $(\text{mid}(I_C) - L, \text{mid}(I_C))$ , and strictly decreasing on the interval  $(\text{mid}(I_C), \text{mid}(I_C) + L)$ .
- (d)  $P_C(x)$  is strictly concave in  $x$  on the interval  $[\text{mid}(I_C) - L_C, \text{mid}(I_C) + L_C]$ .

Let

$$U_C^{(s)}(y) = \max \left\{ 0, F_C^{(s)}(y) \right\}, \quad y \in \mathcal{R},$$

where

$$F_C^{(s)}(y) = E_q[q(x^*(y)|y)P_C(x^*(y)) - E_p c], \quad y \in \mathcal{R},$$

and

$$x^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x).$$

If we have that

$$U_C^{(s)}(y) > 0, \quad y \in I_C,$$

then the function  $U_C^{(s)}(y)$ ,  $y \in \mathcal{R}$ , has the properties that

- a)  $U_C^{(s)}(y)$  is symmetric with respect to  $y_0 = \text{mid}(I_C)$  on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C) + L_C]$ , and we have that

$$\frac{d}{dy} U_C^{(s)}(\text{mid}(I_C)) = 0.$$

- b)  $U_C^{(s)}(y)$  is strictly increasing on  $[\text{mid}(I_C) - L_C, \text{mid}(I_C))$ , and we have that

$$\frac{d}{dy} U_C^{(s)}(y) > 0, \quad y \in [\text{mid}(I_C) - L_C, \text{mid}(I_C)).$$

- c)  $U_C^{(s)}(y)$  is strictly decreasing on  $(\text{mid}(I_C), \text{mid}(I_C) + L_C]$ , and we have that

$$\frac{d}{dy} U_C^{(s)}(y) < 0, \quad y \in (\text{mid}(I_C), \text{mid}(I_C) + L_C].$$

b)  $U_C^{(s)}(y)$  is non-decreasing on  $[mid(I_C) - L, mid(I_C) - L_C]$ .

c)  $U_C^{(s)}(y)$  is non-increasing on  $[mid(I_C), mid(I_C) + L]$ .

*Proof.* The result of the lemma follows directly from the result of Lemma 21 and 22, combined with the assumption that

$$U_C^{(s)}(y) > 0, \quad y \in I_C.$$

□



## H.2 Proof of Proposition 8

The result of Proposition 8 follows from Lemma 23 as the setting of Proposition 8 satisfies all the assumptions made in the statement of Lemma 23.

In particular, Proposition 5 states the the content demand function  $P_C(x)$  satisfies the assumptions made on  $P_C(x)$  in Lemma 23. Moreover, Proposition 2 states that under a Nash equilibrium all agents receive a strictly positive utility, and we have that for a community  $C = (I_C, I_C) \in \mathcal{C}^*$  in a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2 that

$$U_C^{(s)}(y) > 0, \quad y \in I_C.$$

Finally, by Proposition 3 we have for a community  $C = (I_C, I_C) \in \mathcal{C}^*$  under a Nash equilibrium  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2 we have that

$$\int_{\mathcal{R}} \beta_C^*(x|y) dx = E_q, \quad y \in C_s = I_C,$$

and

$$\alpha_C^*(y) = E_p, \quad y \in C_d = I_C.$$

## **I Proof of Proposition 2**

In this appendix we prove Proposition 2. To do this, we first derive three additional results. First we characterize the optimal community selection for a given agent, and show that it is optimal for agents to consume, or produce, content in a single community. Next, we provide sufficient conditions for an interval community to be a feasible community, i.e. a community where all agents obtain strictly positive utility rates.

## I.1 Optimal Community Selection for Content Consumers and Producers

In this subsection we characterize the optimal allocation for content consumption and production by an agent  $y \in \mathcal{R}$  in a given community structure.

First we characterize the optimal allocation for content consumption by a given agent  $y \in \mathcal{R}$ . More precisely, for a given community structure  $(\mathcal{C}, \{\alpha_C(y)\}_{y \in \mathcal{R}}, \{\beta_C(\cdot|y)\}_{y \in \mathcal{R}})$ , as well as a given agent  $y \in \mathcal{R}$ , we characterize the optimal allocation for content consumption

$$\alpha_C^*(y) = \{\alpha_C^*(y)\}_{C \in \mathcal{C}}$$

given by

$$\alpha_C^*(y) = \alpha_C^* = \arg \max_{\alpha_C(y): \|\alpha_C(y)\| \leq E_p} \sum_{C \in \mathcal{C}} \alpha_C(y) \int_{\mathcal{R}} [Q_C(x)p(x|y) - \beta_C(x)c] dx,$$

where  $E_p$  is the maximal fraction of time that agent can allocate to content consumption.

We have the following result.

**Lemma 24.** *Let  $(\mathcal{C}, \{\alpha_C(y)\}_{y \in \mathcal{R}}, \{\beta_C(\cdot|y)\}_{y \in \mathcal{R}})$  be a given community structure. Assume that for a given agent  $y \in \mathcal{R}$  we have that*

$$\max_{C \in \mathcal{C}} \int_{\mathcal{R}} [p(x|y)Q_C(x) - \beta_C(x)c] dx > 0,$$

where for the community  $C = (C_d, C_s) \in \mathcal{C}$  we have

$$Q_C(x) = \int_{y \in C_s} \beta_C(x|y)q(x|y)dy$$

and

$$\beta_C(x) = \int_{y \in C_s} \beta_C(x|y)dy.$$

Then an optimal content consumption rate allocation for this agent  $y$  is given by

$$\alpha_C^*(y) = \begin{cases} E_p, & C = C^*(y) \\ 0, & \text{otherwise,} \end{cases}$$

where

$$C^*(y) = \arg \max_{C \in \mathcal{C}} \int_{\mathcal{R}} [p(x|y)Q_C(x) - \beta_C(x)c] dx,$$

and  $E_p$ ,  $0 < E_p \leq 1$ , is the maximal fraction of time that agent can allocate to content consumption, i.e. we have that

$$\alpha_C^*(y) = \arg \max_{\alpha_C(y): \|\alpha_C(y)\| \leq E_p} \sum_{C \in \mathcal{C}} \alpha_C(y) \int_{\mathcal{R}} [Q_C(x)p(x|y) - \beta_C(x)c] dx.$$

Lemma 24 states that is optimal for an agent  $y \in \mathcal{R}$  to allocate all its content consumption to a single community in a community structure  $(\mathcal{C}, \{\alpha_C(y)\}_{y \in \mathcal{R}}, \{\beta_C(\cdot|y)\}_{y \in \mathcal{R}})$ , namely the community  $C^*(y)$  given by

$$C^*(y) = \arg \max_{C \in \mathcal{C}} \int_{\mathcal{R}} [p(x|y)Q_C(x) - \beta_C(x)c] dx.$$

*Proof.* Recall the definition of the utility rate for content consumption given by

$$U_C^{(d)}(y) = \alpha_C(y) \int_{\mathcal{R}} [Q_C(x)p(x|y) - \beta_C(x)c] dx.$$

The result of the lemma follows then immediately from the fact that the utility rate for content consumption is linear in  $\alpha_C(y)$ .  $\square$

Next we characterize the optimal allocation for content production of a given agent  $y \in \mathcal{R}$ . More precisely, for a given community structure  $(\mathcal{C}, \{\alpha_C(y)\}_{y \in \mathcal{R}}, \{\beta_C(\cdot|y)\}_{y \in \mathcal{R}})$ , and a given agent  $y \in \mathcal{R}$ , we characterize the optimal allocation for content production

$$\beta_C^*(y) = \{\beta_C^*(\cdot|y)\}_{C \in \mathcal{C}}$$

given by

$$\beta_C^*(\cdot|y) = \arg \max_{\beta_C(\cdot|y): \|\beta_C(\cdot|y)\| \leq E_q} \sum_{C \in \mathcal{C}} \int_{\mathcal{R}} \beta_C(x|y) [q(x|y)P_C(x) - \alpha_C c] dx,$$

where  $E_q$  is the maximal rate that agent can allocate to content production.

The next lemma shows that it is optimal for agent  $y$  to allocate all its content production rate in a given community structure  $(\mathcal{C}, \{\alpha_C(y)\}_{y \in \mathcal{R}}, \{\beta_C(\cdot|y)\}_{y \in \mathcal{R}})$  to a single community  $C \in \mathcal{C}$ .

**Lemma 25.** *Let  $(\mathcal{C}, \{\alpha_C(y)\}_{y \in \mathcal{R}}, \{\beta_C(\cdot|y)\}_{y \in \mathcal{R}})$  be a given community structure. Assume that for a given agent  $y \in \mathcal{R}$  we have that*

$$\max_{C \in \mathcal{C}} \left( q(x_C^*(y)|y)P_C(x_C^*(y)) - \alpha_C c \right) > 0,$$

where

$$x_C^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x),$$

$$P_C(x) = \int_{y \in \mathcal{C}_d} \alpha_C(y)p(x|y)dy,$$

and

$$\alpha_C = \int_{y \in I_C} \alpha_C(y)dy.$$

Then the optimal content production rate allocation for this agent  $y$  is given by

$$\beta_C^*(x|y) = \begin{cases} E_q \delta(x_C^*(y) - x), & C = C^*(y) \\ 0, & \text{otherwise,} \end{cases}$$

where

$$C^*(y) = \arg \max_{C \in \mathcal{C}} q(x_C^*(y)|y)P_C(x_C^*(y))$$

and  $E_q$  is the maximal rate that agent can allocate to content production, i.e. we have that

$$\beta_C^*(\cdot|y) = \arg \max_{\beta_C(\cdot|y): \|\beta_C(\cdot|y)\| \leq E_q} \sum_{C \in \mathcal{C}} \int_{\mathcal{R}} \beta_C(x|y) [q(x|y)P_C(x) - \alpha_C c] dx.$$

*Proof.* Note that if for a community  $C \in \mathcal{C}$  we have that

$$\max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right) \leq 0,$$

then the optimal utility that agent  $y$  can achieve in this community is equal to 0, and hence it is optimal for agent  $y$  to set  $\alpha_C(y) = 0$  for this community. As a result, to prove the lemma it suffices to consider communities  $C \in \mathcal{C}$  such that

$$\max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right) > 0.$$

Let  $\mathcal{C}^+$  be this set, i.e. we have that

$$\mathcal{C}^+ = \left\{ C \in \mathcal{C} \mid \max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right) > 0 \right\}.$$

By Proposition 11, we have that if agent  $y$  allocates the content production rate

$$\beta_C(y) = \int_{\mathcal{R}} \beta_C(x|y) dx$$

to community  $C$ , and we have that

$$\max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right) > 0,$$

then the optimal utility rate for content production for agent  $y$  in community  $C$  is given by

$$\beta_C(y) \max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right).$$

Let  $\beta_C^*(\cdot|y)$  be the rate allocation for content production as given in the lemma, and let  $\beta_C(\cdot|y)$  be another rate allocation function such that

$$\sum_{C \in \mathcal{C}^+} \beta_C(y) \leq E_q$$

and

$$\beta_C(y) = 0, \quad C \notin \mathcal{C}^+,$$

where

$$\beta_C(y) = \int_{\mathcal{R}} \beta_C(x|y) dx.$$

To prove the lemma, we have to show that the utility rate for content production of agent  $y$  is not higher under the allocation  $\beta_C(\cdot|y)$  compared with allocation  $\beta_C^*(\cdot|y)$ . Note that the optimal utility rate for agent  $y$  under  $\beta_C^*(\cdot|y)$  is given by

$$E_q \max_{x \in \mathcal{R}} \left( q(x|y)P_{C^*(y)}(x) - \alpha_{C^*(y)} c \right)$$

where

$$C^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y)P_C(x),$$

and the utility rate for agent  $y$  under  $\beta_C(\cdot|y)$  is given by

$$\sum_{C \in \mathcal{C}} \beta_C(y) \max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right).$$

Therefore in order to show that the utility rate for content production of agent  $y$  is not higher under the allocation  $\beta_C(\cdot|y)$  compared with allocation  $\beta_C^*(\cdot|y)$ , we have to show that

$$E_q \max_{x \in \mathcal{R}} \left( q(x|y)P_{C^*(y)}(x) - \alpha_{C^*(y)} c \right) - \sum_{C \in \mathcal{C}} \beta_C(y) \max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right) \geq 0.$$

Note that for this expression we have that

$$\begin{aligned} & E_q \max_{x \in \mathcal{R}} \left( q(x|y)P_{C^*(y)}(x) - \alpha_{C^*(y)} c \right) - \sum_{C \in \mathcal{C}} \beta_C(y) \max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right) = \\ & \left( E_q - \sum_{C \in \mathcal{C}} \beta_C(y) \right) \max_{x \in \mathcal{R}} \left( q(x|y)P_{C^*(y)}(x) - \alpha_{C^*(y)} c \right) + \dots \\ & + \sum_{C \in \mathcal{C}} \beta_C(y) \left[ \max_{x \in \mathcal{R}} \left( q(x|y)P_{C^*(y)}(x) - \alpha_{C^*(y)} c \right) - \max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right) \right]. \end{aligned}$$

By assumption we have that

$$\max_{C \in \mathcal{C}} \left( q(x_C^*(y)|y)P_C(x_C^*(y)) - \alpha_C c \right) > 0,$$

and by construction we have that

$$\sum_{C \in \mathcal{C}} \beta_C(y) \leq E_q.$$

This implies that

$$\left( E_q - \sum_{C \in \mathcal{C}} \beta_C(y) \right) \max_{x \in \mathcal{R}} \left( q(x|y)P_{C^*(y)}(x) - \alpha_{C^*(y)} c \right) \geq 0.$$

Furthermore, by assumption we have that

$$C^*(y) = \arg \max_{C \in \mathcal{C}} q(x_C^*(y)|y)P_C(x_C^*(y)),$$

and we obtain that

$$\sum_{C \in \mathcal{C}} \beta_C(y) \left[ \max_{x \in \mathcal{R}} \left( q(x|y)P_{C^*(y)}(x) - \alpha_{C^*(y)}c \right) - \max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right) \right] \geq 0.$$

Combining these two results, we obtain that

$$E_q \max_{x \in \mathcal{R}} \left( q(x|y)P_{C^*(y)}(x) - \alpha_{C^*(y)}c \right) - \sum_{C \in \mathcal{C}} \beta_C(y) \max_{x \in \mathcal{R}} \left( q(x|y)P_C(x) - \alpha_C c \right) \geq 0.$$

and the result of the lemma follows. □

## I.2 Existence of Feasible Communities

Consider a Nash equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  as given by Proposition 2. We then have that each interval community  $C = (I_C, I_C) \in C^*$  is a feasible community as defined in Definition 3, i.e. we have that

$$U_C^{(d)}(u) > 0, \quad y \in I_C$$

and

$$U_C^{(s)}(u) > 0, \quad y \in I_C,$$

and the utility rates of all agents in the community are strictly positive. Therefore, in order to prove Proposition 2, we have to show that feasible interval communities exist. The next result provides sufficient conditions for an interval community  $C = (I_C, I_C)$  to be a feasible community.

**Lemma 26.** *Let  $C = (C_d, C_s)$  be an interval community, i.e. we have that  $C_d = C_s = I_C$ , where  $I_C \subset \mathcal{R}$  is an interval in  $\mathcal{R}$ , and let*

$$l_0 = \sup\{l \in [0, L] | f(l)g(l) - c > 0\}.$$

If we have that

$$0 < 2L_C < l_0,$$

then for all rate allocations  $\{\alpha_C(y)\}_{y \in C_s}$  and  $\{\beta_C(\cdot|y)\}_{y \in C_s}$  such that

$$\alpha_C(y) > 0, \quad y \in C_d$$

and

$$\int_{\mathcal{R}} \beta_C(x|y) dx > 0, \quad y \in C_s,$$

the community  $C$  is a feasible community under  $\{\alpha_C(y)\}_{y \in C_s}$  and  $\{\beta_C(\cdot|y)\}_{y \in C_d}$ .

*Proof.* Without loss of generality, we can assume that the interval  $I_C$  is given by

$$I_C = [-L_C, L_C].$$

Recall that the utility rate for content producers  $y \in C_s = I_C$  is given by

$$U_C^{(s)}(y) = \int_{\mathcal{R}} \beta_C(x|y) \int_{z \in C_d = I_C} \alpha_C(z) [q(x|y)p(x|z) - c] dz dx.$$

By assumption we have that

$$g(2L_C) > 0,$$

and it follows

$$I_C \subseteq \text{supp}(q(\cdot|y)), \quad y \in C_s = I_C.$$

Furthermore by Assumption 1 we have that

$$q(x|y)p(x|z) > g(2L_C)f(2L_C), \quad x, y, z \in I_C.$$

Combining these two results, we obtain for

$$y \in C_d = I_C$$

that

$$\begin{aligned} U_C^{(s)}(y) &= \int_{x \in I_C} \beta_C(x|y) \int_{z \in I_C} \alpha_C(z) [q(x|y)p(x|z) - c] dz dx \\ &\geq \int_{x \in I_C} \beta_C(x|y) \int_{z \in I_C} \alpha_C(z) [g(2L_C)f(2L_C) - c] dz dx \\ &= [g(2L_C)f(2L_C) - c] \int_{x \in I_C} \beta_C(x|y) \int_{z \in I_C} \alpha_C(z) dz dx. \end{aligned}$$

By assumption we have that

$$\alpha_C(y) > 0, \quad y \in C_d = I_C$$

and

$$\int_{\mathcal{R}} \beta_C(x|y) dx > 0, \quad y \in C_s = I_C,$$

and we obtain that

$$\int_{z \in I_C} \alpha_C(z) dz > 0$$

and

$$\int_{x \in I_C} \beta_C(x|y) \int_{z \in I_C} \alpha_C(z) dz dx > 0, \quad y \in C_s = I_C.$$

Furthermore by assumption we have that

$$0 < 2L_C < l_0,$$

and it follows that

$$f(2L_C)g(2L_C) - c > 0.$$

Combing the above results, we obtain that

$$\left[ g(2L_C)f(2L_C) - c \right] \int_{x \in I_C} \beta_C(x|y) \int_{z \in I_C} \alpha_C(z) dz dx > 0$$

and

$$U_C^{(s)}(y) > 0, \quad y \in C_s = I_C.$$

Similarly, for the utility rate for content consumers  $y \in C_d = I_C$  we have that

$$\begin{aligned} U_C^{(d)}(y) &= \alpha_C(y) \int_{x \in I_C} \int_{z \in I_C} \beta_C(x|z) \left[ q(x|z)p(x|y) - c \right] dz dx \\ &\geq \alpha_C(y) \int_{x \in I_C} \int_{z \in I_C} \beta_C(x|z) \left[ g(2L_C)f(2L_C) - c \right] dz dx \\ &= \alpha_C(y) \left[ g(2L_C)f(2L_C) - c \right] \int_{x \in I_C} \int_{z \in I_C} \beta_C(x|z) dz dx. \end{aligned}$$

By the same argument as given above we obtain that

$$U_C^{(d)}(y) > 0, \quad y \in C_d = I_C,$$

and the result of lemma follows. □



### I.3 Proof of Proposition 2

In this subsection we prove Proposition 2. We do this by construction, i.e. we construct a Nash equilibrium  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  where

$$C^* = \{C^1, \dots, C^K\},$$

and the communities

$$C^k = (C_d^k, C_s^k), \quad k = 1, \dots, K,$$

are given by a set of mutually non-overlapping intervals  $\{I_k\}_{k=1, \dots, K}$  of equal length, i.e. we have that

$$C^k = (C_d^k, C_s^k) = (I_k, I_k), \quad k = 1, \dots, K,$$

with

$$|I_k| = |I_{k'}|, \quad k, k' = 1, \dots, K.$$

We construct such a Nash equilibrium as follows. By Assumption 1 the functions  $f$  and  $g$  are continuous, and it follows that under the condition

$$f(0)g(0) - c > 0$$

there exists a  $l_0 > 0$  such that

$$f(2l_0)g(2l_0) - c > 0.$$

Furthermore, there exists a  $l^*, 0 < l^* < l_0$ , and an integer  $K$  such that

$$Kl^* = L,$$

where  $L$  is the half-length of the set  $\mathcal{R} = [-L, L]$ . Let

$$b_k = 2kl^*, \quad k = 0, \dots, K$$

and let

$$I_k = [b_{k-1}, b_k), \quad k = 1, \dots, K.$$

Using these definitions, we construct a community structure  $(C^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  where

$$C^* = \{C^1, \dots, C^K\}$$

and

$$C^k = (C_d^k, C_d^k) = (I_k, I_k), \quad k = 1, \dots, K.$$

Furthermore, for  $k = 1, \dots, K$ , we let

$$\alpha_{C^k}^*(y) = E_p, \quad y \in I_k,$$

and

$$\beta_{C^k}^*(x|y) = E_q \delta(x_{C^k}^*(y) - x), \quad y \in I_k,$$

where

$$x_{C^k}^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_{C^k}(x)$$

and

$$P_{C^k}(x) = E_p \int_{y \in I_k} p(x|y) dy, \quad x \in \mathcal{R}.$$

Note that  $Q_{C^k}^*(x)$ ,  $k = 1, \dots, K$ , is then given by

$$Q_{C^k}^*(x) = E_q \int_{y \in I_k} E_q \delta(x_{C^k}^*(y) - x) q(x|y) dy, \quad x \in \mathcal{R}.$$

As we have that

$$0 < l^* \leq l_0,$$

it follows from Lemma 26 the the communities  $C^k$ ,  $k = 1, \dots, K$ , are feasible communities under the rate allocations  $\{\alpha_C^*(y)\}_{y \in \mathcal{R}}$  and  $\{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}}$ , and we have that

$$U_{C^k}^{(d)}(y) > 0, \quad y \in I_k, k = 1, \dots, K,$$

and

$$U_{C^k}^{(s)}(y) > 0, \quad y \in I_k, k = 1, \dots, K.$$

Furthermore, by Lemma 19 and 24, it follows that the above rate functions  $\alpha_{C^k}^*(y)$  and  $\beta_{C^k}^*(\cdot|y)$  are indeed the optimal rate functions for  $y \in I_k, k = 1, \dots, K$ , i.e. we have that

$$\alpha_C^*(y) = \arg \max_{\alpha_C(y): \|\alpha_C(y)\| \leq E_p} \sum_{C \in \mathcal{C}} \alpha_C(y) \int_{\mathcal{R}} \left[ Q_C^*(x) p(x|y) - E_q c \right] dx$$

and

$$\beta_C^*(\cdot|y) = \arg \max_{\beta_C(\cdot|y): \|\beta_C(\cdot|y)\| \leq E_q} \sum_{C \in \mathcal{C}} \int_{\mathcal{R}} \beta_C(x|y) \left[ q(x|y) P_C(x) - 2l^* E_p c \right] dx.$$

To see this, note the following. As we have that

$$U_{C^k}^{(d)}(y) > 0, \quad y \in I_k, k = 1, \dots, K,$$

it follows from Lemma 19 that

$$\begin{aligned} C^k &= \arg \max_{C \in \mathcal{C}^*} \int_{\mathcal{R}} \left[ p(x|y) Q_C^*(x) - \beta_C(x) c \right] dx \\ &= \arg \max_{C \in \mathcal{C}^*} \int_{\mathcal{R}} \left[ p(x|y) Q_C^*(x) - E_q c \right] dx, \quad y \in I_k, k = 1, \dots, K. \end{aligned}$$

From Lemma 24, we then obtain that is it optimal for agents  $y \in C_d^k = I^k$  to consume content exclusively in community  $C^k$  and set

$$\alpha_{C^k}(y) = E_p, \quad y \in C_d^k = I^k.$$

Similarly, as we have that

$$U_{C^k}^{(s)}(y) > 0, \quad y \in I_k, k = 1, \dots, K,$$

it follows from Lemma 23 that

$$\begin{aligned} C^k &= \arg \max_{C \in \mathcal{C}^*} \left( q(x_C^*(y)|y) P_C(x_C^*(y)) - \alpha_C c \right) \\ &= \arg \max_{C \in \mathcal{C}^*} \left( q(x_C^*(y)|y) P_C(x_C^*(y)) - 2l^* E_p c \right), \quad y \in C_s^k = I^k, k = 1, \dots, K, \end{aligned}$$

where

$$x_C^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_C(x), \quad C \in \mathcal{C}^*, y \in \mathcal{R}.$$

From Lemma 25 we then obtain that is it optimal for agents  $y \in C_s^k = I^k$  to produce content exclusively in community  $C^k$ , and set

$$\beta_C^*(x|y) = \begin{cases} E_q \delta(x_{C^k}^*(y) - x), & C = C^k \\ 0, & \text{otherwise,} \end{cases}$$

where by Lemma 10  $x_{C^k}^*(y)$  is the unique solution to the optimization problem

$$x_{C^k}^*(y) = \arg \max_{x \in \mathcal{R}} q(x|y) P_{C^k}(x), \quad y \in \mathcal{R}, k = 1, \dots, K.$$

It then follows that the community structure  $(\mathcal{C}^*, \{\alpha_C^*(y)_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)_{y \in \mathcal{R}}\})$  is a Nash equilibrium. Furthermore, by construction we have that

$$\cup_{k=1, \dots, K} I_k = \mathcal{R},$$

and  $(\mathcal{C}^*, \{\alpha_C^*(y)_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)_{y \in \mathcal{R}}\})$  is a covering community structure.

The result of the proposition then follows.

## J Proof for Proposition 9

In this appendix we prove Proposition 9.

Let  $(\mathcal{C}^*, \{\alpha_C^*(y)\}_{y \in \mathcal{R}}, \{\beta_C^*(\cdot|y)\}_{y \in \mathcal{R}})$  be a Nash equilibrium as given by Proposition 2, and consider a given community  $C \in \mathcal{C}^*$ . We then prove the proposition by characterizing the utility for content consumption  $U_C^{(s)}(y_0)$  for agent

$$y_0 = \text{mid}(I_C)$$

in this community. Note that we have that

$$U_C^{(s)}(y_0) = E_p E_q \int_{z \in I_C} [q(x^*(y_0)|y_0)p(x^*(y_0)|z) - c] dz.$$

By Proposition 3 we have that

$$x^*(y_0) = y_0 = \arg \max_{x \in \mathcal{R}} q(x|y_0)P_C(x).$$

It then follows

$$U_C^{(s)}(y_0) = E_p E_q \int_{z \in I_C} [g(0)f(\|x^*(y_0) - z\|) - c] dz = 2E_p E_q \int_0^{L_C} [g(0)f(x) - c] dx,$$

where  $L_C$  is the length of the interval  $I_C$ . By Proposition 2 we have that

$$U_C^{(s)}(y_0) > 0,$$

and we obtain that

$$\int_0^{L_C} [g(0)f(x) - c] dx > 0.$$

The result then follows combining the above result with the fact that by Assumption 1 the function  $f(x)$ ,  $x \in [0, L]$ , is strictly decreasing in  $x$ .