Community Structures in Information Networks, Technical Report

Martin Carrington and Peter Marbach

Department of Computer Science, University of Toronto, Toronto, Canada
{carrington,marbach}@cs.utoronto.edu

Abstract. We study community structures that emerge in an information network using the game-theoretic model proposed in [1]. In particular, we consider a particular family of community structures, and provide conditions under which there exists a Nash equilibrium within this family.

Keywords: information networks · community structure

1 Introduction

In this paper we consider a particular type of social network, which we refer to as an *information network*, where agents (individuals) share/exchange information. Sharing/exchanging of information is an important aspect of social networks, both for social networks that we form in our everyday lives, as well as for online social networks such as Twitter.

The work in [1] presents a model to study communities in information networks where agents produce (generate) content, and consume (obtain) content. Furthermore, the model allows agents to form communities in order to share/exchange content more efficiently, where agents obtain a certain utility for joining a given community. Using a game-theoretic framework, [1] characterizes the community structures that emerge in information networks as Nash equilibria. More precisely, [1] considers a particular family of community structures, and shows that (under suitable assumptions) there always exists a community structure that is a Nash equilibrium. One open question from [1] is whether the family of community structures considered includes all Nash equilibria, or whether there exist Nash equilibria that are not covered by the analysis in [1].

In this paper we address this question, and show that there do indeed exist Nash equilibria that are not covered by the analysis in [1]. One interesting, and important, characteristic is that the Nash equilibria that we derive in this paper have the property that some agents (individuals) are "excluded" from the community structure, i.e. do not participate in any of the information communities. If such Nash equilibria are to emerge in real-life (social) information networks, it would mean that some individuals are "marginalized". This is definitely an undesirable outcome that could come at great cost for the individuals that are "marginalized". As such, understanding when the Nash equilibria obtained in this paper do emerge in (social) information networks is an important question. We discuss this in more detail in Section 5.

The rest of the paper is organized as follows. In Section 2 we summarize the model presented in [1] that we use for our analysis. In Section 3 we define the family of community structures that we consider in this paper, and in Section 4 we present our results. Due to space constraints we refer to [1] for a review of related literature.

2 Background

In this section we review the model and results of [1]. Due to space constraints we keep the presentation of the model brief, and refer to [1] for a more detailed discussion of the model, and the results that were obtained in [1].

For our analysis we assume that each content item that is being produced in the information community is of a particular type. One might think of a content type as a topic, or an interest, that agents might have. Furthermore we assume that there exists a structure that relates different content types to each other. In particular, we assume there exists a measure of "closeness" between content types that characterizes how strongly related two content types are. For example, as "basketball" and "baseball" are both sports one would assume that the two topics are more closely related than "basketball" and "mathematics". To model this situation we assume that the type of a content item is given by a point x in a metric space, and the closeness between two content types $x, x' \in \mathcal{M}$ is then given by the distance measure $d(x, x')$, $x, x' \in \mathcal{M}$, for the metric space $\mathcal{M}$.

Having defined the set of content that can be produced in an information network, we next describe agents' interests in content as well as the agents' ability to produce content. To do this, we assume that there is a set $\mathcal{A}_d$ of agents that consume content, and a set $\mathcal{A}_s$ of agents that produce content, where the subscripts stand for "demand" and "supply". Furthermore, we associate with each agent that consumes content a center of interest $y \in \mathcal{M}$, i.e. the center of interest y of the agent is the content type (topic) that an agent is most interested in. The interest in content of type x of an agent with center of interest y is given by

$$p(x|y) = f(d(x, y)), \qquad x, y \in \mathcal{M}, \tag{1}$$

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 an agent with center of interest y consumes (reads) a content item of type x, then it finds it interesting with probability $p(x|y)$ as given by Eq. (1). As the function f is non-increasing, this model captures the intuition that the agent is more interested in content that is close to its center of interest y.

Similarly, given an agent that produces content, the center of interest y of the agent is the content type (topic) that the agent is most adept at producing. The ability of the agent to produce content of type $x \in \mathcal{M}$ is then given by

$$q(x|y) = g(d(x, y)), \tag{2}$$

where $g : [0, \infty) \mapsto [0, 1]$ is a non-increasing function.

In the following we identify an agent by its center of interest $y \in \mathcal{M}$, i.e. agent y is the agent with center of interest y. As a result we have that $\mathcal{A}_d \subseteq \mathcal{M}$ and $\mathcal{A}_s \subseteq \mathcal{M}$.

2.1 Information Community

We model an information community as follows. An information community $C = (C_d, C_s)$ consists of a set of agents that consume content $C_d \subseteq \mathcal{A}_d$ and a set of agents that produce content $C_s \subseteq \mathcal{A}_s$. Let $\beta_C(x|y)$ be the rate at which agent $y \in C_s$ generates content items of type x in community C. Let $\alpha_C(y)$ be the fraction of content produced in community C that agent $y \in C_d$ consumes. To define the utility for content consumption and production, we assume that when an agent consumes a single content item, it receives a reward equal to 1 if the content item is of interest and relevant, and pays a cost of $c > 0$ for consuming the item. The cost c captures the cost in time (energy) to read/consume a content item. Using this reward and cost structure, the utility rate ("reward minus cost") for content consumption of agent $y \in C_d$ is given by (see [1] for a detailed derivation)

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

where

$$Q_C(x) = \int_{y \in C_s} \beta_C(x|y) q(x|y) dy, \quad \text{and} \quad \beta_C(x) = \int_{y \in C_s} \beta_C(x|y) dy.$$

Similarly, the utility rate for content production of agent $y \in C_s$ is given by

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

where

$$P_C(x) = \int_{y \in C_d} \alpha_C(y) p(x|y) dy, \quad \text{and} \quad \alpha_C = \int_{z \in C_d} \alpha_C(z) dz.$$

As discussed in [1], the utility rate for content production captures how "valuable" the content produced by agent y is for the set of content consuming agents C_d in the community C.

2.2 Community Structure and Nash Equilibrium

Using the above definition of a community, a community structure that describes how agents organize themselves into communities is then given by a triplet $(\mathcal{C}, \{\alpha_{\mathcal{C}}(y)\}_{y \in \mathcal{A}_d}, \{\beta_{\mathcal{C}}(\cdot|y)\}_{y \in \mathcal{A}_s})$, where the set of communities $\mathcal{C}$ in this structure consists of communities C as defined in the previous section, and

$$\alpha_{\mathcal{C}}(y) = \{\alpha_C(y)\}_{C \in \mathcal{C}}, y \in \mathcal{A}_d, \quad \text{and} \quad \beta_{\mathcal{C}}(\cdot|y) = \{\beta_C(\cdot|y)\}_{C \in \mathcal{C}}, y \in \mathcal{A}_s,$$

are the consumption fractions and production rates, respectively, that agents allocate to the different communities $C \in \mathcal{C}$. We assume that the total consumption fractions and production rates of each agent are bounded by $E_p > 0$, and $E_q > 0$, respectively, i.e. we have that

$$\|\alpha_{\mathcal{C}}(y)\| = \sum_{C \in \mathcal{C}} \alpha_C(y) \leq E_p \leq 1, \qquad y \in \mathcal{A}_d,$$

and

$$\|\beta_{\mathcal{C}}(y)\| = \sum_{C \in \mathcal{C}} \|\beta_C(\cdot|y)\| \leq E_q, \qquad y \in \mathcal{A}_s,$$

where

$$\|\beta_C(\cdot|y)\| = \int_{x \in \mathcal{M}} \beta_C(x|y)dx.$$

We assume that agents form communities in order to maximize their utility rates, i.e. agents join communities, and choose allocations $\alpha_{\mathcal{C}}(y)$, and $\beta_{\mathcal{C}}(\cdot|y)$ to maximize their total consumption, and production utility rates, respectively.

A Nash equilibrium is then given by a community structure $(\mathcal{C}^*, \{\alpha^*_{\mathcal{C}}(y)\}_{y \in \mathcal{A}_d}, \{\beta^*_{\mathcal{C}}(\cdot|y)\}_{y \in \mathcal{A}_s})$ such that for all agents $y \in \mathcal{A}_d$ we have that

$$\alpha^*_{\mathcal{C}}(y) = \underset{\alpha_{\mathcal{C}}(y): \|\alpha_{\mathcal{C}}(y)\| \leq E_p}{\arg\max} \sum_{C \in \mathcal{C}} U^{(d)}_C(y),$$

and for all agents $y \in \mathcal{A}_s$, we have that

$$\beta^*_{\mathcal{C}}(\cdot|y) = \underset{\beta_{\mathcal{C}}(\cdot|y): \|\beta_{\mathcal{C}}(y)\| \leq E_q}{\arg\max} \sum_{C \in \mathcal{C}} U^{(s)}_C(y).$$

We call a Nash equilibrium a covering Nash equilibrium if for all agents $y \in \mathcal{A}_d$, we have that there exists at least one community $C \in \mathcal{C}$ such that $\alpha_C(y) > 0$, and for all agents $y \in \mathcal{A}_d$, we have that there exists at least one community $C \in \mathcal{C}$ such that $\|\beta_C(\cdot|y)\| > 0$.

2.3 Results

The above has been analyzed in [1] for the case of a specific metric space, and a specific family of information communities. In particular, the analysis in [1] was carried out for a one-dimensional metric space with the torus metric, and for discrete interval communities. Below we formally define the metric space and the family of interval communities that was considered in [1].

The analysis in [1] considered the following one-dimensional metric space. The metric space is given by an interval $\mathcal{R} = [-L, L) \subset \mathbb{R}$, $L > 0$, 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 $x \in \mathbb{R}$. Furthermore, the analysis in [1] assumes that

$$\mathcal{A}_d = \mathcal{A}_s = \mathcal{R},$$

i.e. for each content type $x \in \mathcal{R}$ there exists an agent in $\mathcal{A}_d$ who is most interested in content of type x, and there exists an agent in $\mathcal{A}_s$ who is most adept at producing content of type x.

In addition, the analysis in [1] considers a particular family $\mathcal{C}(L_C)$, $L_C > 0$, of community structures, given as follows. Let $N \geq 2$ be a a given integer, and let

$$L_C = \frac{L}{N},$$

where L is the half-length of the metric space $\mathcal{R} = [-L, L)$. Furthermore, let $\{m_k\}_{k=1}^N$ be a set of N evenly spaced points on the metric space $\mathcal{R} = [-L, L)$ given by

$$m_{k+1} = m_1 + 2L_C k, \qquad k = 1, ..., N-1.$$

The set $\mathcal{C} = \{C^k = (C_d^k, C_s^k)\}_{k=1}^N$ of communities in the community structure $\mathcal{C}(L_C)$ is then given by N communities $C^k = (C_d^k, C_s^k)$, and for each community C^k the set of content consuming agents C_d^k, and the set of content producing agents C_s^k, are given by the intervals

$$C_d^k = [m_k - L_C, m_k + L_C)$$

and

$$C_s^k = [m_k - L_C, m_k + L_C).$$

Furthermore, the allocations $\{\alpha_{\mathcal{C}}(y)\}_{y \in \mathcal{R}}$ and $\{\beta_{\mathcal{C}}(\cdot|y)\}_{y \in \mathcal{R}}$ are given by

$$\alpha_{C^k}(y) = \begin{cases} E_p & y \in C_d^k \\ 0 & \text{otherwise} \end{cases}, \qquad k = 1, ..., N,$$

and

$$\beta_{C^k}(\cdot|y) = \begin{cases} E_q \delta(x - x_y^*) & y \in C_s^k \\ 0 & \text{otherwise} \end{cases}, \qquad k = 1, ..., N,$$

where

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

The analysis in [1] shows that (under certain assumptions about the functions f and g that are used in Eq. (1) and Eq. (2)) there always exists a covering Nash equilibrium within the family $\mathcal{C}(L_C)$, $L_C > 0$, of community structures.

3 Community Structure $\mathcal{C}(L_C, l_d)$

In this section we consider a family of community structures that is more general than the family $\mathcal{C}(L_C)$, $L_C > 0$, of the previous section, and study whether there exists a Nash equilibrium within this family.

More precisely, we consider the following family $\mathcal{C}(L_C, l_d)$ of community structures. Let $N \geq 2$ be a a given integer. Furthermore, let

$$L_C = \frac{L}{N},$$

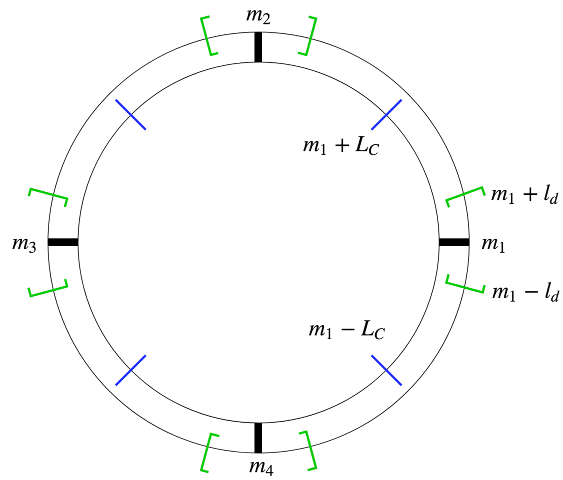

Fig. 1. The communities $\mathcal{C}$ for the case where $N = 4$ are illustrated. The metric space $\mathcal{R} = [-L, L)$ is shown as a ring to represent the torus (ring) meric. More precsiely ther are two rings: the outer ring represents the set of the content consumers $\mathcal{A}_d$, and the inner ring represents the set of content producers $\mathcal{A}_s$. The brackets on the outer ring bound the four consumption intervals C_d^k,, $k = 1, ..., 4$, and the lines on the inner ring bound the four production intervals C_s^k, $k = 1, ..., 4$.

where L is the half-length of the metric space $\mathcal{R} = [-L, L)$, and let l_d be such that

$$0 < l_d \leq L_C.$$

Finally, let $\{m_k\}_{k=1}^N$ be a set of N evenly spaced points on the metric space $\mathcal{R} = [-L, L)$ given by

$$m_{k+1} = m_1 + 2L_C k, \qquad k = 1, ..., N-1.$$

Given L_C and l_d as defined above, the set of communities $\mathcal{C} = \{C^k = (C_d^k, C_s^k)\}_{k=1}^N$ of the structure $\mathcal{C}(L_C, l_d)$ is then given by the intervals

$$C_d^k = [m_k - l_d, m_k + l_d), \qquad k = 1, ..., N.$$

and

$$C_s^k = [m_k - L_C, m_k + L_C), \qquad k = 1, ..., N.$$

Fig. 1 provides an illustration of these communities for the case of $N = 4$ communities.

Furthermore, the allocations $\{\alpha_{\mathcal{C}}(y)\}_{y \in \mathcal{R}}$ and $\{\beta_{\mathcal{C}}(\cdot|y)\}_{y \in \mathcal{R}}$ are given by

$$\alpha_{C^k}(y) = \begin{cases} E_p & y \in C_d^k \\ 0 & \text{otherwise} \end{cases}, \qquad k = 1, ..., N,$$

and

$$\beta_{C^k}(\cdot|y) = \begin{cases} E_q\delta(x - x_y^*) & y \in C_s^k \\ 0 & \text{otherwise} \end{cases}, \qquad k = 1, ..., N,$$

where

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

Note that when $l_d = L_C$, then the community structure $\mathcal{C}(L_C, l_d) = \mathcal{C}(L_C, L_C)$ is identical to the community structure $\mathcal{C}(L_C)$ of the previous section that was analyzed in [1]. In particular, in this case the community structure $\mathcal{C}(L_C, L_C)$ is again a covering community structure, i.e. all agents belong to at least one community in $\mathcal{C}(L_C, L_C)$. As a result, we will focus on community structures $\mathcal{C}(L_C, l_d)$ where

$$0 < l_d < L_C.$$

In this case the community structure $\mathcal{C}(L_C, l_d)$, $0 < l_d < L_C$, is no longer a covering community structure. In particular, the content consuming agents in the sets

$$D^k = [m_k + l_d, m_{k+1} - l_d), \qquad k = 1, ..., N-1,$$

and

$$D^N = [m_N + l_d, m_1 - l_d)$$

do not belong to any communities in $\mathcal{C}(L_C, l_d)$. On the other hand, note that all content producing agents $y \in \mathcal{R}$ do belong to at least one community C^k in the community structure $\mathcal{C}(L_C, l_d)$. In this sense, studying the existence of a Nash equilibrium within the family of community structures $\mathcal{C}(L_C, l_d)$ is studying whether there exists a Nash equilibrium from which some content consuming agents are excluded. We discuss the implications of such a Nash equilibrium in more detail in Section 5.

To study whether there exists a Nash equilibrium within the family $\mathcal{C}(L_C, l_d)$ of community structures as defined above, we use the following definitions. Let

$$x_y^*(l_d) = \arg\max_{x \in \mathcal{R}} q(x|y) \int_{-l_d}^{l_d} p(x|z)dz, \qquad y \in \mathcal{R}.$$

Furthermore, let the functions $G(y|L_C, l_d)$ and $H(y|L_C, l_d)$ be given by

$$G(y|L_C, l_d) = E_pE_q \int_{z=-L_C}^{L_C} p(x_z^*(l_d)|y)q(x_z^*(l_d)|z)dz - 2E_pE_qL_Cc, \qquad y \in \mathcal{R},$$

and

$$H(y|L_C, l_d) = E_pE_q q(x_y^*(l_d)|y) \int_{z=-l_d}^{l_d} p(x_y^*(l_d)|z)dz - 2E_pE_q l_d c, \qquad y \in \mathcal{R},$$

where $c > 0$ is the cost for consuming a single content item.

In addition, we make the following assumptions about the functions f and g that are used in Eq. (1) and Eq. (2).

Assumption 1. *The function* $f : [0, \infty) \mapsto [0, 1]$ *is given by*

$$f(x) = \max\{0, f_0 - ax\},$$

where $f_0 \in (0, 1]$ *and* $a > 0$. *The function* $g : [0, \infty) \mapsto [0, 1]$ *is given by*

$$g(x) = g_0,$$

where $g_0 \in (0, 1]$. *Furthermore, we have that*

$$f_0 g_0 > c. \tag{3}$$

We note that the condition given by Eq. (3) is a necessary condition for a Nash equilibrium to exist, i.e. it is shown in [1] that if this condition is not true, then there does not exist a Nash equilibrium.

4 Main Results

In this section we present the main results of our analysis. We first provide necessary and sufficient conditions for a community structure $\mathcal{C}(L_C, l_d)$ to be a Nash equilibrium.

Proposition 1. *Let the functions* f *and* g *be as given in Assumption 1. Furthermore, let* L_C^* *and* l_d^* *be such that*

$$0 < l_d^* < L_C^*, \quad \text{and} \quad L_C^* = \frac{L}{N},$$

where L *is the half-length of the metric space* $\mathcal{R} = [-L, L)$ *and* $N \geq 2$ *is an integer. Then the community structure* $\mathcal{C}(L_C^*, l_d^*)$ *is a Nash equilibrium if, and only if, we have that*

$$G(l_d^* | L_C^*, l_d^*) = 0, \quad \text{and} \quad H(L_C^* | L_C^*, l_d^*) \geq 0.$$

We provide a proof for Proposition 1 in Appendix C.

Our next result shows that there always exists a Nash equilibrium given that the half-length L of the metric space $\mathcal{R} = [-L, L)$ is large enough.

Proposition 2. *Let the functions* f *and* g *be as given in Assumption 1. If we have that*

$$L > 2\left[\frac{f_0}{a} - \frac{c}{a g_0}\right],$$

then there always exists a community structure $\mathcal{C}(L_C, l_d)$, $0 < l_d < L_C$, *that is a Nash equilibrium.*

We provide a proof for Proposition 2 in Appendix D.

Proposition 2 states that for functions f and g as given in Assumption 1, there always exists a Nash equilibrium in the family of community structures $\mathcal{C}(L_C, l_d)$ given that L is large enough, i.e. if we have that $L > 2\left[\frac{f_0}{a} - \frac{c}{a g_0}\right]$.

The next result provides a complete characterization of the values of L_C and l_d, $0 < l_d < L_C$, for which there exists a Nash equilibrium.

Proposition 3. *Let the functions f and g be as given in Assumption 1. Then the community structure $\mathcal{C}(L_C^*, l_d^*)$ with*

$$0 < l_d^* < L_C^*, \quad \text{and} \quad L_C^* = \frac{L}{N}$$

where $N \geq 2$ is an integer, is a Nash equilibrium if, and only if,

$$l_d^* = \frac{f_0}{a} - \frac{c}{a g_0}.$$

We provide a proof for Proposition 3 in Appendix E.

Note that the above result provides a complete characterization of the Nash equilibria within the family of community structures $\mathcal{C}(L_C, l_d)$. We discuss the interpretation of this result in more detail in the next section.

5 Conclusions

In this paper we show that there exists an additional family of Nash equilibria to the one identified in [1]. This result shows that there are more types of community structures that can emerge as Nash equilibria in (social) information networks than it may first appear. Studying whether the are additional Nash equilibria to the ones identified in [1] and in this paper, is interesting future research.

The Nash equilibria that we obtained have the property that some agents are excluded from the community structure, i.e. they do not belong to any of the communities. The reason for this is that these agents would have a negative utility in all of the communities that exist in the Nash equilibrium (see Appendix C for a formal derivation of this result). This means that these agents have the choice to either join a community where their utility would be negative, or not join any community at all (and obtain a utility of zero). Since in this situation agents are better off not joining any community, they are "marginalized". This outcome may come at a significant "social" cost to these agents. Studying this issue in depth is outside of the scope of this paper, but this is important and interesting future research. In particular, a natural question to ask in this context is whether, and how likely it is that the Nash equilibria that "marginalize" agents will indeed arise in information networks. This question can be studied formally by using the model in [1] to analyze the dynamics of community formation in information networks, and how the resulting dynamics can lead to the Nash equilibria that "marginalize" some agents. In addition, this question can be studied empirically to find out whether community structures that "marginalize" some individuals indeed occur in real-life information networks.

References

1. Marbach, P.: Modeling and Analysis of Information Communities. Arxiv:1511.08904, available at `http://www.cs.toronto.edu/~marbach/PUBL/info_communities.pdf`

Appendices

In Appendix A we derive the utility rate functions for agents in a community which is part of a community structure $\mathcal{C}(L_C, l_d)$. In Appendix B, we explain how these utility rate functions relate to the G and H functions defined in Section 2.2. We prove Propositions 1, 2, and 3 in Appendices C, D, and E, respectively.

A Utility Rate Functions in $\mathcal{C}(L_C, l_d)$ Under Assumption 1

In this section we derive the utility rate functions for agents in a community which is part of a community structure $\mathcal{C}(L_C, l_d)$, using the functions f and g given in Assumption 1.

Let $C^k = (C_d^k, C_s^k)$ be a community in a community structure $\mathcal{C}(L_C, l_d)$, with

$$C_d^k = [m_k - l_d, m_k + l_d)$$

and

$$C_s^k = [m_k - L_C, m_k + L_C).$$

Using the expressions for the utility rates given in Section 2.1, we obtain that the utility rate for content consumption of agent $y \in C_d^k$ is given by

$$U_{C^k}^{(d)}(y) = \alpha_{C^k}(y) \int_{x \in \mathcal{R}} [Q_{C^k}(x) p(x|y) - \beta_{C^k}(x) c] dx, \tag{4}$$

where

$$Q_{C^k}(x) = \int_{y \in C_s^k} \beta_{C^k}(x|y) q(x|y) dy$$

and

$$\beta_{C^k}(x) = \int_{y \in C_s^k} \beta_{C^k}(x|y) dy,$$

and that the utility rate for content production of agent $y \in C_s^k$ is given by

$$U_{C^k}^{(s)}(y) = \int_{x \in \mathcal{R}} \beta_{C^k}(x|y) [q(x|y) P_{C^k}(x) - \alpha_{C^k} c] dx, \tag{5}$$

where

$$P_{C^k}(x) = \int_{y \in C_d} \alpha_{C^k}(y) p(x|y) dy$$

and

$$\alpha_{C^k} = \int_{z \in C_d^k} \alpha_{C^k}(z) dz.$$

Now since C^k is a community in $\mathcal{C}(L_C, l_d)$, we have that the consumption and production allocations are given by

$$\alpha_{C^k}(y) = E_p, \qquad y \in C_d^k,$$

and

$$\beta_{C^k}(\cdot|y) = E_q \delta(x - x_y^*), \qquad y \in C_s^k,$$

respectively, where

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

Therefore, substituting these allocations into Equations 4 and 5, the functions simplify to

$$U_{C^k}^{(d)}(y) = E_p E_q \int_{z=m_k-L_C}^{m_k+L_C} p(x_z^*|y) q(x_z^*|z) dz - 2E_p E_q L_C c, \qquad y \in C_d^k, \tag{6}$$

and

$$U_{C^k}^{(s)}(y) = E_p E_q q(x_y^*|y) \int_{z=m_k-l_d}^{m_k+l_d} p(x_y^*|z) dz - 2E_p E_q l_d c, \qquad y \in C_s^k. \tag{7}$$

Lemma 1 characterizes x_y^*, for $y \in \mathcal{R}$.

Lemma 1. *Let f and g be given by Assumption 1. Let $C^k = (C_d^k, C_s^k)$ be a community in $\mathcal{C}(L_C, l_d)$, with*

$$C_d^k = [m_k - l_d, m_k + l_d)$$

and

$$C_s^k = [m_k - L_C, m_k + L_C).$$

Consider the optimization problem

$$\max_{x \in \mathcal{R}} q(x|y) P_{C^k}(x), \qquad y \in \mathcal{R}.$$

For the case where

$$0 < l_d \le \frac{f_0}{a},$$

we have that

$$x_y^* = m_k$$

is the unique solution to the optimization problem. For the case where

$$\frac{f_0}{a} < l_d \le L_C,$$

we have that x_y^* is a solution to the optimization problem if, and only if,

$$x_y^* \in \left[m_k + \frac{f_0}{a} - l_d, m_k + l_d - \frac{f_0}{a} \right].$$

Proof. In Assumption 1 we assume that $g(x) = g_0 \in (0,1]$, therefore the given optimization problem is equivalent to

$$\max_{x \in \mathcal{R}} P_{C^k}(x), \tag{8}$$

and any solution x^*_y is independent of y.

Without loss of generality, let $m_k = 0$. We begin by characterizing the function $P_{C^k}(x)$. By definition, we have

$$P_{C^k}(x) = E_p \int_{z=-l_d}^{l_d} \max\{0, f_0 - a\|z - x\|\}\, dz. \tag{9}$$

Setting $s = z - x$, we get that

$$P_{C^k}(x) = E_p \int_{s=-l_d-x}^{l_d-x} \max\{0, f_0 - a\|s\|\}\, ds.$$

Therefore

$$\frac{d}{dx} P_C(x) = -E_p \max\{0, f_0 - a\|l_d - x\|\} + E_p \max\{0, f_0 - a\|l_d + x\|\}.$$

There are four cases to consider. The case where

$$\|l_d + x\| < \frac{f_0}{a} \quad \text{and} \quad \|l_d - x\| < \frac{f_0}{a},$$

the case where

$$\|l_d + x\| \geq \frac{f_0}{a} \quad \text{and} \quad \|l_d - x\| \geq \frac{f_0}{a},$$

the case where

$$\|l_d + x\| < \frac{f_0}{a} \quad \text{and} \quad \|l_d - x\| \geq \frac{f_0}{a},$$

and the case where

$$\|l_d + x\| \geq \frac{f_0}{a} \quad \text{and} \quad \|l_d - x\| < \frac{f_0}{a}.$$

In the first case,

$$\frac{d}{dx} P_{C^k}(x) = E_p a \left[\|l_d - x\| - \|l_d + x\|\right],$$

which is negative when $x \in (0, L)$ and positive when $x \in [-L, 0)$. In the second case,

$$\frac{d}{dx} P_{C^k}(x) = 0.$$

In the third case,

$$\frac{d}{dx} P_{C^k}(x) = E_p(f_0 - a\|l_d + x\|) > 0.$$

In the fourth case,

$$\frac{d}{dx}P_{C^k}(x) = -E_p(f_0 - a\|l_d - x\|) < 0.$$

Note that we have the third case only when $x < 0$ and the fourth only when $x > 0$.

Therefore, for $x = 0$ we have

$$\frac{d}{dx}P_{C^k}(x) = 0,$$

for $x \in (0, L)$ we have

$$\frac{d}{dx}P_{C^k}(x) \begin{cases} = 0 & \|l_d + x\| \geq \frac{f_0}{a} \text{ and } \|l_d - x\| \geq \frac{f_0}{a} \\ < 0 & \text{otherwise} \end{cases},$$

and for $x \in [-L, 0)$ we have

$$\frac{d}{dx}P_{C^k}(x) \begin{cases} = 0 & \|l_d + x\| \geq \frac{f_0}{a} \text{ and } \|l_d - x\| \geq \frac{f_0}{a} \\ > 0 & \text{otherwise} \end{cases}.$$

Therefore, we have that $P_{C^k}(x)$ is non-decreasing on $[-L, 0)$ and non-increasing on $(0, L)$. If

$$\frac{f_0}{a} < l_d \leq L_C,$$

$P_{C^k}(x)$ is constant on

$$\left[\frac{f_0}{a} - l_d, l_d - \frac{f_0}{a}\right].$$

strictly decreasing on

$$\left[l_d - \frac{f_0}{a}, l_d\right],$$

and strictly increasing on

$$\left[-l_d, l_d - \frac{f_0}{a}\right].$$

If

$$0 < l_d \leq \frac{f_0}{a},$$

$P_{C^k}(x)$ is strictly decreasing on

$$[0, l_d]$$

and strictly increasing on

$$[-l_d, 0].$$

Therefore, for the case where

$$0 < l_d \leq \frac{f_0}{a},$$

we have that

$$x_y^* = 0$$

is the unique solution to the optimization problem

$$\max_{x \in \mathcal{R}} P_{C^k}(x), \tag{10}$$

and for the case where

$$\frac{f_0}{a} < l_d \leq L_C,$$

we have that x_y^* is a solution to the said optimization problem if, and only if,

$$x_y^* \in \left[\frac{f_0}{a} - l_d, l_d - \frac{f_0}{a}\right].$$

□

Let us consider the case where

$$\frac{f_0}{a} < l_d \leq L_C.$$

Substituting in the f and g functions given in Assumption 1, and using the result from Lemma 1 that

$$x_y^* \in \left[m_k + \frac{f_0}{a} - l_d, m_k + l_d - \frac{f_0}{a}\right], \qquad y \in \mathcal{R},$$

we get that Equations 6 and 7 simplify to

$$U_{C^k}^{(d)}(y|x_y^*) = E_p E_q g_0 \max\left\{0, f_0 - a\|y - x_y^*\|\right\}(2L_C) - 2E_pE_qL_Cc, \qquad y \in C_d^k, \tag{11}$$

and

$$U_{C^k}^{(s)}(y|x_y^*) = E_p E_q g_0 \int_{z=m_k-l_d}^{m_k+l_d} \max\left\{0, f_0 - a\|z - x_y^*\|\right\} dz - 2E_pE_ql_dc, \qquad y \in C_s^k. \tag{12}$$

Figure 2 illustrates C_d^k.

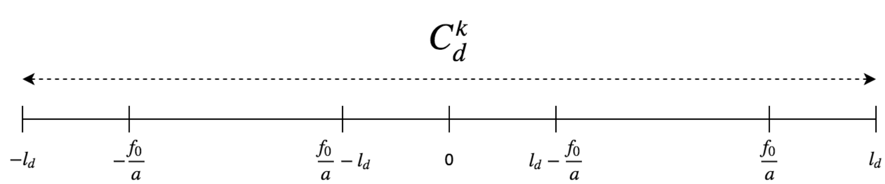

Fig. 2. The case where $m_k = 0$ and $l_d > \frac{f_0}{a}$ is illustrated.

Note that when $\|y - x_y^*\| \geq \frac{f_0}{a}$, i.e.

$$y \in \left[m_k - l_d, x_y^* - \frac{f_0}{a}\right] \bigcup \left[x_y^* + \frac{f_0}{a}, m_k + l_d\right],$$

we have that

$$U_{C^k}^{(d)}(y) = -2E_pE_qL_Cc.$$

And note that when $\|y - x_y^*\| < \frac{f_0}{a}$, i.e.

$$y \in \left[x_y^* - \frac{f_0}{a}, x_y^* + \frac{f_0}{a}\right],$$

we have that

$$U_{C^k}^{(d)}(y|x_y^*) = 2E_pE_qL_C\left[g_0(f_0 - a|y - x_y^*|) - c\right]. \tag{13}$$

Note that this consumption utility rate depends on x_y^*.

We now work out the integral in Equation 12. Note that

$$\max\left\{0, f_0 - a\|z - x_y^*\|\right\} \neq 0$$

only when

$$\|z - x_y^*\| = |z - x_y^*| < \frac{f_0}{a}$$

i.e. when

$$-\frac{f_0}{a} + x_y^* < z < \frac{f_0}{a} + x_y^*.$$

Since

$$x_y^* \in \left[m_k + \frac{f_0}{a} - l_d, m_k + l_d - \frac{f_0}{a}\right],$$

we have that

$$\frac{f_0}{a} + x_y^* \leq m_k + l_d$$

and that

$$-\frac{f_0}{a} + x_y^* \geq m_k - l_d.$$

Therefore,

$$\begin{aligned}\int_{z=m_k-l_d}^{m_k+l_d} \max\left\{0, f_0 - a\|z - x_y^*\|\right\} dz &= \int_{z=-\frac{f_0}{a}+x_y^*}^{\frac{f_0}{a}+x_y^*} (f_0 - a|z - x_y^*|)dz \\ &= \int_{z=-\frac{f_0}{a}}^{\frac{f_0}{a}} (f_0 - a|z|)dz \\ &= \frac{f_0^2}{a}.\end{aligned} \tag{14}$$

Therefore, we have that

$$U^{(s)}_{C^k}(y) = E_p E_q \left[g_0 \frac{f_0^2}{a} - 2 l_d c \right], \qquad y \in C_s^k.$$

Now let us turn to the case where

$$0 < l_d \leq \frac{f_0}{a}.$$

Substituting in the f and g functions given in Assumption 1, and using the result from Lemma 1 that

$$x_y^* = m_k, \qquad y \in \mathcal{R},$$

we get that Equations 6 and 7 simplify to

$$U^{(d)}_{C^k}(y) = E_p E_q g_0 \max\{0, f_0 - a\|y - m_k\|\}(2L_C) - 2E_p E_q L_C c, \qquad y \in C_d^k, \tag{15}$$

and

$$U^{(s)}_{C^k}(y) = E_p E_q g_0 \int_{z=m_k-l_d}^{m_k+l_d} \max\{0, f_0 - a\|z - m_k\|\}\, dz - 2E_p E_q l_d c, \qquad y \in C_s^k. \tag{16}$$

For

$$y \in C_d^k = [m_k - l_d, m_k + l_d),$$

we have

$$\|y - m_k\| = |y - m_k| \leq l_d.$$

And since we are assuming that

$$l_d \leq \frac{f_0}{a},$$

we have

$$|y - m_k| \leq \frac{f_0}{a}.$$

Therefore, Equation 15 can be written as

$$U^{(d)}_{C^k}(y) = 2E_p E_q L_C \left[g_0 (f_0 - a|y - m_k|) - c \right], \qquad y \in C_d^k.$$

Now we work out the integral in Equation 16. We have

$$\begin{aligned} \int_{z=m_k-l_d}^{m_k+l_d} \max\{0, f_0 - a\|z - m_k\|\}\, dz &= \int_{z=-l_d}^{l_d} (f_0 - a|z|) dz \\ &= 2 f_0 l_d - a l_d^2. \end{aligned}$$

Therefore, we have that

$$U^{(s)}_{C^k}(y) = E_p E_q l_d (2 f_0 g_0 - a g_0 l_d - 2c), \qquad y \in C_s^k.$$

Note that for the case where

$$\frac{f_0}{a} < l_d \leq L_C,$$

we have that

$$x_y^*, \qquad y \in \mathcal{R},$$

the solution to the optimization problem

$$\max_{x \in \mathcal{R}} q(x|y) P_{C^k}(x), \qquad y \in \mathcal{R},$$

is not unique. Therefore, in Equation 13, the consumption utility rate depends on x_y^*. However, the production utility rate does not depend on x_y^*, which is not surprising since by definition, x_y^* is such that it maximizes the production utility rate.

For agents $y \in C_d^k$, the function $U_{C^k}^{(d)}(y)$ given in Equation 6 gives the consumption utility rate of y. For agents $y \notin C_d^k$, that function gives the utility rate y would obtain if it allocated its entire consumption budget to C^k (i.e. if $\alpha_{C^k}(y) = E_p$).

Similarly, for agents $y \in C_s^k$, the function $U_{C^k}^{(s)}(y)$ given in Equation 7 gives the production utility rate of y. For agents $y \notin C_s^k$, that function gives the utility rate y would obtain if it allocated its entire production budget to C^k (i.e. $\|\beta_{C^k}(\cdot|y)\| = E_q$).

Therefore, we have that the utility rate functions, now defined on $\mathcal{R}$, are as follows.

For the case where

$$\frac{f_0}{a} < l_d \leq L_C,$$

for

$$y \notin \left[x_y^* - \frac{f_0}{a}, x_y^* + \frac{f_0}{a} \right],$$

we have

$$U_{C^k}^{(d)}(y) = -2 E_p E_q L_C c. \tag{17}$$

And for

$$y \in \left[x_y^* - \frac{f_0}{a}, x_y^* + \frac{f_0}{a} \right],$$

we have

$$U_{C^k}^{(d)}(y|x_y^*) = 2 E_p E_q L_C \left[g_0 (f_0 - a|y - x_y^*|) - c \right]. \tag{18}$$

And we have

$$U_{C^k}^{(s)}(y) = E_p E_q \left[g_0 \frac{f_0^2}{a} - 2 l_d c \right], \qquad y \in \mathcal{R}.$$

For the case where

$$0 < l_d \leq \frac{f_0}{a},$$

we have

$$U^{(d)}_{C^k}(y) = 2E_pE_qL_C\left[g_0(f_0 - a\|y - m_k\|) - c\right], \qquad \|y - m_k\| \le \frac{f_0}{a}, \tag{19}$$

and

$$U^{(d)}_{C^k}(y) = -2E_pE_qL_Cc, \qquad \|y - m_k\| > \frac{f_0}{a}. \tag{20}$$

And we have

$$U^{(s)}_{C^k}(y) = E_pE_ql_d(2f_0g_0 - ag_0l_d - 2c), \qquad y \in \mathcal{R}.$$

B $G(y|L_C, l_d)$ and $H(y|L_C, l_d)$ as Utility Functions

In the following lemma we show that $G(l_d|L_C, l_d)$ represents the consumption utility of agents at the ends of consumption intervals in the community structure $\mathcal{C}(L_C, l_d)$. Similarly, $H(l_d|L_C, l_d)$ represents the production utility of agents at the ends of production intervals in the community structure $\mathcal{C}(L_C, l_d)$.

Lemma 2. *Let $\mathcal{C}(L_C, l_d)$ be a community structure with*

$$0 < l_d < L_C,$$

and let $C^k = (C^k_d, C^k_s)$ be a community in $\mathcal{C}(L_C, l_d)$, with

$$C^k_d = [m_k - l_d, m_k + l_d)$$

and

$$C^k_s = [m_k - L_C, m_k + L_C).$$

We have that

$$G(l_d|L_C, l_d) = 0$$

if, and only if

$$U^{(d)}_{C^k}(m_k + l_d) = 0.$$

We also have that

$$H(L_C|L_C, l_d) \ge 0$$

if, and only if,

$$U^{(s)}_{C^k}(m_k + l_d) \ge 0.$$

Proof. Recall that the the functions $G(y|L_C, l_d)$ and $H(y|L_C, l_d)$ (defined in Section 2.2) are given by

$$G(y|L_C, l_d) = E_pE_q\int_{z=-L_C}^{L_C} p(x^*_z(l_d)|y)q(x^*_z(l_d)|z)dz - 2E_pE_qL_Cc, \qquad y \in \mathcal{R},$$

and

$$H(y|L_C, l_d) = E_pE_qq(x^*_y(l_d)|y)\int_{z=-l_d}^{l_d} p(x^*_y(l_d)|z)dz - 2E_pE_ql_dc, \qquad y \in \mathcal{R}.$$

Without loss of generality, let $m_k = 0$.

Then $G(y|L_C, l_d)$ corresponds exactly to the utility rate function $U^{(d)}_{C^k}(y)$ given in Equation 6. Similarly, $H(y|L_C, l_d)$ corresponds exactly to the utility rate function $U^{(s)}_{C^k}(y)$ given in Equation 7.

Therefore,

$$G(l_d|L_C, l_d) = 0$$

if, and only if,

$$U^{(d)}_{C^k}(l_d) = 0,$$

and

$$H(L_C|L_C, l_d) \geq 0$$

if, and only if,

$$U^{(s)}_{C^k}(l_d) \geq 0.$$

□

In Lemma 3, we characterize the utility rates at the ends of the consumption and production intervals in a community structure $\mathcal{C}(L_C, l_d)$.

Lemma 3. *Let f and g be given by Assumption 1. Let $C^k = (C^k_d, C^k_s)$ be a community in $\mathcal{C}(L_C, l_d)$, with*

$$C^k_d = [m_k - l_d, m_k + l_d)$$

and

$$C^k_s = [m_k - L_C, m_k + L_C).$$

Let

$$T = \frac{f_0}{a} - \frac{c}{ag_0}.$$

We have that

$$U^{(d)}_{C^k}(m_k + l_d) \begin{cases} > 0 & 0 < l_d < T \leq L_C \\ = 0 & l_d = T \leq L_C \\ < 0 & T < l_d \leq L_C \end{cases}. \tag{21}$$

Also, for $0 < l_d < 2T$, we have that

$$U^{(s)}_{C^k}(m_k + l_d) > 0.$$

Proof. Without loss of generality, let $m_k = 0$.

Let us consider the case where

$$\frac{f_0}{a} < l_d \leq L_C.$$

Recall that when

$$y \in \left[-l_d, x^*_y - \frac{f_0}{a}\right] \bigcup \left[x^*_y + \frac{f_0}{a}, l_d\right],$$

we have that

$$U^{(d)}_{C^k}(y) = -2E_pE_qL_Cc.$$

Therefore,

$$U^{(d)}_{C^k}(l_d) = -2E_pE_qL_Cc.$$

Also recall that

$$U^{(s)}_{C^k}(y) = E_pE_q\left[g_0\frac{f_0^2}{a} - 2l_dc\right], \qquad y \in \mathcal{R},$$

and therefore

$$U^{(s)}_{C^k}(l_d) = E_pE_q\left[g_0\frac{f_0^2}{a} - 2l_dc\right].$$

Now let us turn to the case where

$$0 < l_d \le \frac{f_0}{a}.$$

Recall that

$$U^{(d)}_{C^k}(y) = 2E_pE_qL_C\left[g_0(f_0 - a|y - m_k|) - c\right], \qquad y \in C^k_d.$$

Therefore, for $m_k = 0$ and $y = l_d$ we get

$$U^{(d)}_{C^k}(l_d) = 2E_pE_qL_C\left[g_0(f_0 - al_d) - c\right].$$

Also recall that

$$U^{(s)}_{C^k}(y) = E_pE_ql_d(2f_0g_0 - ag_0l_d - 2c), \qquad y \in \mathcal{R}.$$

Therefore,

$$U^{(s)}_{C^k}(l_d) = E_pE_ql_d(2f_0g_0 - ag_0l_d - 2c).$$

In summary, we have that

$$U^{(d)}_{C^k}(l_d) = \begin{cases} -2E_pE_qL_Cc & \frac{f_0}{a} < l_d \le L_C \\ 2E_pE_qL_C\left[g_0(f_0 - al_d) - c\right] & 0 < l_d \le \frac{f_0}{a} \end{cases}, \tag{22}$$

and that

$$U^{(s)}_{C^k}(l_d) = \begin{cases} E_pE_q\left[g_0\frac{f_0^2}{a} - 2l_dc\right] & \frac{f_0}{a} < l_d \le L_C \\ E_pE_ql_d(2f_0g_0 - ag_0l_d - 2c) & 0 < l_d \le \frac{f_0}{a} \end{cases}. \tag{23}$$

Let

$$T = \frac{f_0}{a} - \frac{c}{ag_0}.$$

Note that $U^{(d)}_{C^k}(l_d)$ is non-increasing in l_d and that $U^{(d)}_{C^k}(l_d) = 0$ only when $l_d = T$. Therefore,

$$U^{(d)}_{C^k}(l_d) \begin{cases} > 0 & 0 < l_d < T \leq L_C \\ = 0 & l_d = T \leq L_C \\ < 0 & T < l_d \leq L_C \end{cases} . \tag{24}$$

$U^{(s)}_{C^k}(l_d)$ is a continuous function, quadratic in l_d on $0 < l_c \leq \frac{f_0}{a}$ and linear on $\frac{f_0}{a} < l_d \leq L_C$. We have that

$$2f_0g_0 - ag_0l_d - 2c = 0$$

if, and only if, $l_d = 2T$. So if $2T \leq \frac{f_0}{a}$, then $U^{(s)}_{C^k}(l_d) > 0$ for

$$0 < l_d < 2T.$$

If $2T > \frac{f_0}{a}$, we have $U^{(s)}_{C^k}(l_d) > 0$ for

$$0 < l_d < \frac{g_0f_0^2}{2ac}.$$

But we always have that

$$\frac{g_0f_0^2}{2ac} \geq 2T = 2\left[\frac{f_0}{a} - \frac{c}{ag_0}\right],$$

since we get

$$(f_0g_0 - 2c)^2 \geq 0$$

after rearranging. Therefore, for $0 < l_d < 2T$, we have that

$$U^{(s)}_{C^k}(l_d) > 0.$$

In particular, $U^{(s)}_{C^k}(l_d) > 0$ whenever $U^{(d)}_{C^k}(l_d) \geq 0$.

□

C Proof of Proposition 1

We prove the "if" and "only if" parts of the proof in the following two Subsections.

C.1 $G(l_d^*|L_C^*, l_d^*) = 0$ and $H(L_C^*|L_C^*, l_d^*) \geq 0$.

Suppose that the community structure $\mathcal{C}(L_C^*, l_d^*)$ with

$$0 < l_d^* < L_C^*, \quad \text{and} \quad L_C^* = \frac{L}{N},$$

where $N \geq 2$ is an integer, is a Nash equilibrium. We need to show that

$$G(l_d^*|L_C^*, l_d^*) = 0, \quad \text{and} \quad H(L_C^*|L_C^*, l_d^*) \geq 0.$$

Let $C^k = (C_d^k, C_s^k)$ be a community in $\mathcal{C}(L_C^*, l_d^*)$. Without loss of generality, let

$$C_d^k = [-l_d^*, l_d^*)$$

and

$$C_s^k = [-L_C^*, L_C^*).$$

First we show that $H(L_C^*|L_C^*, l_d^*) \geq 0$. In Appendix A we show that the production utility rate is the same for all producing agents in C_s^k. Since the community structure is a Nash equilibrium, this utility rate must be nonnegative. In particular, by Lemma 2, we must have

$$H(L_C^*|L_C^*, l_d^*) \geq 0.$$

Now we want to show that $G(l_d^*|L_C^*, l_d^*) = 0$. We show this by contradiction.

Suppose that $G(l_d^*|L_C^*, l_d^*) > 0$. By Lemma 2, this means that $U_{C^k}^{(d)}(l_d) > 0$. Since, $U_{C^k}^{(d)}(y)$ is a continuous function of y, there must exist an agent

$$y_0 \in D^k = [l_d, 2L_C - l_d)$$

such that $U_{C^k}^{(d)}(y_0) > 0$. Agent y_0 has a consumption utility rate of zero, but if $U_{C^k}^{(d)}(y_0) > 0$, agent y_0 could have a positive consumption utility rate if it decided to consume the content in C^k. Therefore the community structure is not a Nash equilibrium.

Now suppose that $G(l_d^*|L_C^*, l_d^*) < 0$. By Lemma 2, this means that $U_{C^k}^{(d)}(l_d) < 0$. Since, $U_{C^k}^{(d)}(y)$ is a continuous function of y, there must exist an agent

$$y_0 \in C_d^k$$

such that $U_{C^k}^{(d)}(y_0) < 0$. This agent therefore has a negative utility rate, therefore, the community structure cannot be a Nash equilibrium.

C.2 $\mathcal{C}(L_C^*, l_d^*)$ is a Nash Equilibrium

Now we prove the other direction. Suppose that

$$G(l_d^* | L_C^*, l_d^*) = 0$$

and

$$H(L_C^* | L_C^*, l_d^*) \geq 0.$$

In order to show that $\mathcal{C}(L_C^*, l_d^*)$ is a Nash equilibrium, we must show that no content producer can increase its utility rate by changing communities and no content consumer can increase its utility rate by changing communities.

First consider the content producers. In Appendix A we show that the production utility rate is the same for all producing agents in a community in $\mathcal{C}(L_C^*, l_d^*)$. By symmetry, the production utility rate is the same for all agents in the N communities in $\mathcal{C}(L_C^*, l_d^*)$. And since we have

$$H(L_C^* | L_C^*, l_d^*) \geq 0,$$

by Lemma 2, that production utility rate is nonnegative. Therefore, the content producers have maximized their utility.

Now consider the content consumers. Let $C^k = (C_d^k, C_s^k)$ be a community in $\mathcal{C}(L_C^*, l_d^*)$. Without loss of generality, let

$$C_d^k = [-l_d^*, l_d^*)$$

and

$$C_s^k = [-L_C^*, L_C^*).$$

For any $l_d \in (0, L_C)$, according to Equations 17 to 20, we have that $U_{C^k}^{(d)}(y)$ is non-decreasing on the interval

$$[-L, x_y^*]$$

and non-increasing on the interval

$$[x_y^*, L].$$

Since

$$G(l_d^* | L_C^*, l_d^*) = 0,$$

by Lemma 2 we have that

$$U_{C^k}^{(d)}(-l_d^*) = U_{C^k}^{(d)}(l_d^*) = 0.$$

Therefore, agents consuming content from a community have a nonnegative consumption utility rate and can only get a nonpositive consumption utility rate in other communities (or a utility rate of 0 if they do not consume from any community). Also, agents that are not part of any community, i.e. agents in

$$\bigcup_{k=1}^{N} D^k$$

cannot increase their utility rate by consuming the content in any of the communities.

D Proof of Proposition 2

Let

$$l_d = \frac{f_0}{a} - \frac{c}{ag_0}.$$

Note that in Assumption 1 we assume that $f_0 g_0 > c$. Therefore

$$l_d > 0.$$

Let

$$L > 2l_d,$$
$$N = 2,$$

and

$$L_C = \frac{L}{N}.$$

Then we have that

$$0 < l_d < L_C.$$

Therefore, by Proposition 3, $\mathcal{C}(L_C, l_d)$ is a Nash equilibrium.

E Proof of Proposition 3

We make use of the following result in the proof.

Lemma 4. *For a community structure* $\mathcal{C}(L_C^*, l_d^*)$ *with*

$$0 < l_d^* < L_C^*, \quad \text{and} \quad L_C^* = \frac{L}{N}$$

where $N \geq 2$ *is an integer, we have that*

$$G(l_d^* | L_C^*, l_d^*) = 0$$

and

$$H(L_C^* | L_C^*, l_d^*) \geq 0$$

if and only if

$$l_d^* = \frac{f_0}{a} - \frac{c}{ag_0}. \tag{25}$$

Proof. Combining the results of Lemma 2 and Lemma 3, we get that

$$G(l_d^* | L_C^*, l_d^*) = 0$$

if and only if

$$l_d^* = \frac{f_0}{a} - \frac{c}{ag_0}, \tag{26}$$

and we also get that $H(l_s^* | L_C^*, l_d^*) \geq 0$ when Equation 26 holds. □

Now we prove Proposition 3. We prove the "if" and "only if" parts of the proof in the following two Subsections.

E.1 $\mathcal{C}(L_C^*, l_d^*)$ is a Nash Equilibrium

First we prove that if

$$l_d^* = \frac{f_0}{a} - \frac{c}{ag_0},$$

then the community structure $\mathcal{C}(L_C^*, l_d^*)$ with

$$0 < l_d^* < L_C^*, \quad \text{and} \quad L_C^* = \frac{L}{N},$$

where $N \geq 2$ is an integer, is a Nash equilibrium.

Note that $0 < l_d^*$ since we assume in Assumption 1 that $f_0 g_0 > c$.

Now, by Proposition 1, in order to show that this community structure $\mathcal{C}(L_C^*, l_d^*)$ is a Nash equilibrium, it is sufficient to show that

$$G(l_d^* | L_C^*, l_d^*) = 0$$

and

$$H(l_s^* | L_C^*, l_d^*) \geq 0.$$

If $l_d^* = \frac{f_0}{a} - \frac{c}{ag_0}$, then Lemma 4 gives us that result.

E.2 $l_d^* = \frac{f_0}{a} - \frac{c}{ag_0}$

Now we prove the other direction, i.e. we prove that if the community structure $\mathcal{C}(L_C^*, l_d^*)$ with

$$0 < l_d^* < L_C^*, \quad \text{and} \quad L_C^* = \frac{L}{N},$$

where $N \geq 2$ is an integer, is a Nash equilibrium, then we must have

$$l_d^* = \frac{f_0}{a} - \frac{c}{ag_0}.$$

By Proposition 1, if the community structure $\mathcal{C}(L_C^*, l_d^*)$ with

$$0 < l_d^* < L_C^*, \quad \text{and} \quad L_C^* = \frac{L}{N}$$

where $N \geq 2$ is an integer, is a Nash equilibrium, then

$$G(l_d^* | L_C^*, l_d^*) = 0$$

and

$$H(l_s^* | L_C^*, l_d^*) \geq 0.$$

Then by Lemma 4 we have that $l_d^* = T$.