Community Structures in Information Networks

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 a game-theoretic approach. In particular, we consider a particular family of community structures, and provide conditions under which there exists a Nash equilibrium within this family.

Keywords: Social networks · 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

K. Avrachenkov et al. (Eds.): GameNets 2019, LNICST 277, pp. 119–127, 2019.
https://doi.org/10.1007/978-3-030-16989-3_9

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.

The rest of the paper is organized as follows. In Sect. 2 we summarize the model presented in [1] that we use for our analysis. In Sect. 3 we define the family of community structures that we consider in this paper, and in Sect. 4 we present our results.

Due to space constraints we refer to [1] for a review of related literature, and we only point to the work on content forwarding and filtering in social networks by Zadeh, Goel and Munagala [2], and the work by Hegde, Massoulie, and Viennot [3], as they are most closely related to the analysis presented in this paper. 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 Kronecker graphs. In [3], 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 [3] 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 [2,3] and the approach in this paper is that model and analysis in [2,3] does not explicitly consider and model community structures, and the utility obtained by users under the models in [2,3] depends only on the content that agents receive, but not on the content agents produce.

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

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

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

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

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

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

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

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

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 model has been analyzed in [1] for the case of a specific metric space, and a specific family of information communities. More precisely, the analysis in [1] considered the one-dimensional metric space 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 given integer, and let

$$L_C = \frac{L}{N}, \tag{3}$$

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

$$m_{k+1} = m_1 + 2L_C k, \qquad k = 1, \ldots, N - 1. \tag{4}$$

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

$$C^k_d = [m_k - L_C, m_k + L_C) \text{ and } C^k_s = [m_k - L_C, m_k + L_C).$$

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

$$\alpha_{C^k}(y) = \begin{cases} E_p & y \in C^k_d \\ 0 & \text{otherwise} \end{cases}, \qquad k = 1, \ldots, N, \tag{5}$$

and

$$\beta_{C^k}(\cdot|y) = \begin{cases} E_q \delta(x - x_y^*) & y \in C_s^k \\ 0 & \text{otherwise} \end{cases}, \qquad k = 1, \ldots, N, \tag{6}$$

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 Eqs. (1) and (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 given integer and let $L_C = \frac{L}{N}$ as given by Eq. (3). In addition, let $\{m_k\}_{k=1}^N$ be a set of N evenly spaced points on the metric space $\mathcal{R} = [-L, L)$ as given by Eq. (4).

Given L_C, l_d, and m_k, $k = 1, \ldots, N$, 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

$$C_d^k = [m_k - l_d, m_k + l_d) \text{ and } C_s^k = [m_k - L_C, m_k + L_C).$$

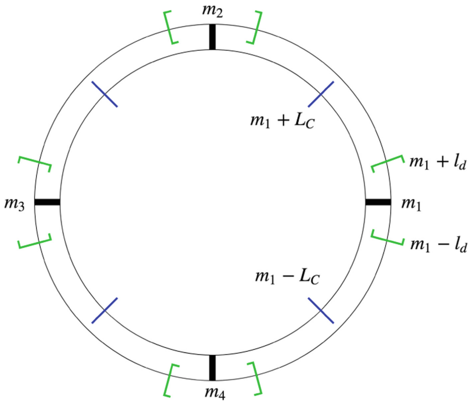

Fig. 1. The communities $\mathcal{C}$ for the case where $N = 4$ are illustrated. The metric space $\mathcal{R} = [-L, L)$ is shown as a ring to represent the torus (ring) metric. More precisely there are two rings: the outer ring represents the set of the content consumers $\mathcal{A}_d$, and the inner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

Figure 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 $\{\beta_{\mathcal{C}}(\cdot|y)\}_{y\in\mathcal{R}}$ of the community structure $\mathcal{C}(L_C, l_d)$ are as given by Eq. (5), and Eq. (6), respectively.

Note that for $l_d = L_C$, 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 we have that $l_d < L_C$. 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

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

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 Sect. 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

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

and

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

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 Eqs. (1) and (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*

$$f_0 g_0 > c. \tag{7}$$

We note that the condition given by Eq. (7) 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. Due to space constraints, we state the results without proofs, the proofs can be found in [4]. 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*

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

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.*

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,

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

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]. 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 [4] 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.

References

1. Marbach, P.: Modeling and analysis of information communities. arXiv:1511.08904
2. Zadeh, R.B., Goel, A., Munagala, K., Sharma, A.: On the precision of social and information networks. In: Proceedings of the First ACM Conference on Online Social Networks, pp. 63–74. ACM (2013)
3. Hegde, N., Massoulié, L., Viennot, L.: Self-organizing flows in social networks. In: Moscibroda, T., Rescigno, A.A. (eds.) SIROCCO 2013. LNCS, vol. 8179, pp. 116–128. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-03578-9_10
4. Carrington, M., Marbach, P.: Community structures in information networks. Technical report. http://www.cs.toronto.edu/~marbach/PUBL/gamenets_2019.pdf