THE APPENDIX FOR THE PAPER:

"INFORMATION PRODUCTION AND LINK FORMATION IN SOCIAL COMPUTING SYSTEMS"

Yu Zhang, Mihaela van der Schaar

Appendix A

1) Proof of Lemma 1

First, we show that $(y^\rho + z^\rho)^{1/\rho} > (y^{\rho'} + z^{\rho'})^{1/\rho'}$ always holds for any $y, z > 0$ and $0 < \rho < \rho' \le 1$. This could be shown by taking the first-order partial derivatives of $(y^\rho + z^\rho)^{1/\rho}$ over ρ:

$$\begin{aligned}&\frac{\partial (y^\rho + z^\rho)^{1/\rho}}{\partial \rho} = (y^\rho + z^\rho)^{1/\rho}\left[-\frac{1}{\rho^2}\ln(y^\rho + z^\rho) + \frac{1}{\rho}\frac{1}{y^\rho + z^\rho}(y^\rho \ln y + z^\rho \ln z)\right] \\ &= \frac{1}{\rho}(y^\rho + z^\rho)^{1/\rho}\left[-\ln(y^\rho + z^\rho)^{1/\rho} + \frac{1}{y^\rho + z^\rho}(y^\rho \ln y + z^\rho \ln z)\right]\end{aligned}. \quad (1)$$

It is obvious that $\frac{1}{y^\rho + z^\rho}(y^\rho \ln y + z^\rho \ln z) < \max\left\{\ln y, \ln z\right\}$, while $\ln(y^\rho + z^\rho)^{1/\rho} > \max\left\{\ln y, \ln z\right\}$. Hence, the first-order partial derivative (1) is always negative.

The above analysis on two positive values y and z can be straightforwardly extended to n positive values and thus Lemma 1 follows. ■

2) Proof of Lemma 2

Taking the first-order derivative in x_i over f, we have that

$$\frac{\partial f}{\partial x_i} = v'\left(X_i\right)\left(\frac{X_i}{x_i}\right)^{1-\rho}. \quad (2)$$

Since $v'\left(X_i\right) > 0$, we have that $\frac{\partial f}{\partial x_i} > 0$ and $f\left(x_i, \boldsymbol{x}_{-i}, \boldsymbol{g}\right)$ is strictly increasing in x_i.

By taking the second-order derivative

$$\frac{\partial^2 f}{\partial x_i^2} = v''\left(X_i\right)\left(\frac{X_i}{x_i}\right)^{2(1-\rho)} + v'\left(X_i\right)(1-\rho)\left(\frac{X_i}{x_i}\right)^{-\rho}\frac{\left(\frac{X_i}{x_i}\right)^{1-\rho} x_i - X_i}{x_i^2}, \quad (3)$$

it can be verified that both terms in the RHS of (3) are smaller than 0. Hence, we have $\frac{\partial^2 f}{\partial x_i^2} < 0$ and thus $f\left(x_i\right)$ is twice differentiable and strictly concave in x_i. Hence, Lemma 2 follows. ■

3) Proof of Lemma 3

(i) Suppose that $\boldsymbol{s}^* = \left(\boldsymbol{x}^*, \boldsymbol{g}^*\right)$ with $g_{ij}^* g_{ji}^* = 1$, then agent i can strictly increase its utility by selecting $g_{ij} = 0$, which leads to a contradiction to the definition of an equilibrium. Hence, statement (i) follows.

(ii) According to Eq. (2), we have already shown that an agent i always produces a positive amount of information when it acquires a positive amount of information from its neighbors, i.e. $X_i > x_i$.

When agent i acquires no information from its neighbors, i.e. $X_i = x_i$, the first-order derivative in x_i over its benefit function becomes $\frac{\partial f}{\partial x_i} = v'\left(x_i\right)$. Since $v'\left(0\right) \geq \alpha > c$, agent i always has the incentive to produce at $x_i = 0$. Summing up, we have statement (2) follows.

(iii) Due to Assumption 3, agent i's marginal benefit of production monotonically decreases with the amount of information it acquires from others. Hence, it has the largest marginal benefit of production, i.e. the largest incentive to produce information, at every point of x_i when it acquires no information from others, i.e. $\left|N_i\left(\overline{g}\right)\right| = 0$. The corresponding utility function for agent i can be rewritten as

$$u_i\left(\boldsymbol{x}, \boldsymbol{g}\right) = v\left(x_i\right) - cx_i. \tag{4}$$

It stops producing new information when the marginal benefit of production equals to the marginal cost, i.e. at the point $x_i = \overline{x}$ where $v'\left(\overline{x}\right) = c$. Since $v\left(x\right)$ is strictly concave, $v'\left(x\right)$ is strictly decreasing and hence, $v'\left(\overline{x}\right) = c$ has a unique solution. Statement (iii) follows. ■

Appendix B

1) Proof of Theorem 1

We first prove the existence of Nash equilibrium. In general, it is difficult to show the existence of pure Nash equilibrium in network formation games. Hence, in this proof, we first consider the IPLF game where agents play mixed strategy, which is called as IPLFM. Particularly, in IPLFM, the link formation choice between two agents is not binary, but continuous. That is, the link formation strategy of an agent

agent i now becomes a vector $\boldsymbol{p}_i = \left\{p_{i1},\dots,p_{in}\right\}$, where $p_{ij} \in \left[0,1\right]$ and $p_{ii} = 0$. We define the strength of a link to be $\overline{g}_{ij} = \overline{g}_{ji} = \max\left\{p_{ij}, p_{ji}\right\}$. When $\overline{g}_{ij} = 1$, we say that the link between agents i and j is of full strength. The utility of agent i is then defined as

$$u_i\left(\boldsymbol{x},\boldsymbol{g}\right) = v\left(\left[x_i^\rho + \sum_{j\in N}\overline{g}_{ij}x_j^\rho\right]^{\frac{1}{\rho}}\right) - cx_i - k\sum_{j\in N}p_{ij}\,. \tag{5}$$

Since each agent plays a mixed strategy on both information production and link formation, it is always true that the IPLFM game has at least one equilibrium. In this rest of this proof, we show that each equilibrium of the IPLFM game has $p_{ij} \in \{0,1\}$ for any $i,j \in N$, which makes it also being an equilibrium of the IPLF game where the link formation choice is binary.

First, it is obvious that in any equilibrium of the IPLFM game, we have $p_{ij}^* p_{ji}^* = 0$ holds for any pair of agents i and j. Suppose there is a pair of agents i and j such that $p_{ij}^* \in (0,1)$ in an equilibrium. Now consider that agent i changes p_{ij}^* to $p_{ij}^* + \varepsilon$. The strength of the link between i and j increases if ε is positive, and decreases otherwise. To keep its total cost in information production and link formation constant, agent i also changes its production level to $x_i^* + \mu$ which satisfies the equality as $c(x_i^* + \mu) + (g_{ij}^* + \varepsilon)k = cx_i^* + g_{ij}^* k$, i.e. $\mu = -\frac{k}{c}\varepsilon$. After all these changes, agent i's benefit from information consumption becomes $v\left(\left[(x_i^*+\mu)^\rho + \sum_{l\in N, l\neq j}\overline{g}_{il}^*(x_l^*)^\rho + (g_{ij}^*+\varepsilon)(x_j^*)^\rho\right]^{\frac{1}{\rho}}\right)$. Since the total cost of agent i keeps unchanged, its utility increases if its benefit from information consumption increases, which is equivalent to the following inequality

$$(x_i^* - \frac{k}{c}\varepsilon)^\rho + \varepsilon(x_j^*)^\rho \ge (x_i^*)^\rho\,, \tag{6}$$

which can be further transformed into

$$x_i^* - \frac{k}{c}\varepsilon \ge ((x_i^*)^\rho - \varepsilon(x_j^*)^\rho)^{\frac{1}{\rho}}\,. \tag{7}$$

With simple manipulation, it can be proved that the LHS of (7) is a linear decreasing function of ε while the RHS is a concave decreasing function of ε. Since the LHS and the RHS equal at $\varepsilon = 0$, it is then obvious that there is a sufficiently small $\varepsilon > 0$ which makes (7) valid. Hence, the utility of agent i monotonically increases and we have that if $p_{ij}^* > 0$ in an equilibrium, then $p_{ij}^* = 1$ should always hold. Therefore, it is proved that each equilibrium of the IPLFM game has $p_{ij} \in \{0,1\}$ for any $i, j \in N$ and hence, the existence of Nash equilibrium in the IPLF game follows.

To prove the second part of the theorem, we classify all strategy profiles into two classes and analyze them separately. The first class S^A contains strategy profiles with which there is no link in the network, i.e. each agent is isolated. The second class $S^B = S / S^A$ contains all other strategy profiles. Here S is the set of all strategy profiles.

If there is a strategy profile $\boldsymbol{s} \in S^A$ that is an equilibrium, then each agent should produce $\overline{x}$ and hence statement (i) follows.

If there is a strategy profile $\boldsymbol{s}' \in S^B$ being an equilibrium. It should be noted that there is at least one link in the network under $\boldsymbol{s}'$. Now look at any two agents j and j' who are mutually connected with each other. According to Assumption 3 and Lemma 3(iii), both j and j' have production levels lower than $\overline{x}$. Suppose there is one agent i who is isolated under $\boldsymbol{s}'$, i.e. does not connect with any other agent in the network. Hence, i's production level is also $\overline{x}$. Without loss of generality, we assume that $g_{jj'} = 1$ in $\boldsymbol{s}'$. Then j can strictly increase its utility by switching the link from j' to i. Hence, there should be no isolated agent under $\boldsymbol{s}'$ and statement (ii) follows. ■

Appendix C

2) Proof of Lemma 4

(i) Suppose that there is an agent $j > n_h\left(\boldsymbol{s}^*\right)$ such that $g_{ij}^* = 1$. If there is also an agent $i' \le n_h\left(\boldsymbol{s}^*\right)$ such that $\overline{g}_{ii'}^* = 0$, then agent i can always strictly increase its utility by switching the link from j to i'. Therefore, we derive the fact that i is connected with all other high producers, i.e. $\overline{g}_{ii'}^* = 1,\ \forall i' \le n_h\left(s^*\right)\ and\ i' \ne i$. If i is also connected with all low producers, then it is obvious that each neighbor of agent j is also a neighbor of agent i, which indicates that the information that agent j

receives from its neighbors is no more than what is received by agent i. According to Assumption 3, if $x_j^* = x_i^*$, then agent i should have less incentive to produce then agent j. Hence, due to the concavity of the benefit function $v(\cdot)$ and the fact that all agents have the same incentive to produce at equilibrium (i.e. the partial derivative over $f_i(\cdot)$ in x_i equals to c), we should have $x_i^* \leq x_j^*$. Since this argument is valid for all agents other than i, it can be concluded that x_i^* is the smallest among all agents in the network, which contradicts the fact that i is a high producer. Hence, there is an agent $j' > n_h\left(\boldsymbol{s}^*\right)$ such that $\overline{g}_{ij'}^* = 0$. Clearly, $g_{j'j''}^* = 0$ for all $j'' > n_h\left(\boldsymbol{s}^*\right)$. Otherwise, j' can strictly increase its utility by switching the link from j'' to i.

Now we prove that each neighbor of agent j' is also a neighbor of agent i. Suppose there is an agent $j'' > n_h\left(\boldsymbol{s}^*\right)$ such that $g_{j''j'}^* = 1$. It can be concluded that $\overline{g}_{j''i}^* = 1$ also holds. Otherwise, j'' can strictly increase its utility by switching the link from j' to i. Regarding the fact that i is connected with all high producers, it implies that every agent who is a neighbor of agent j' is also a neighbor of agent i. Therefore, the amount of information that agent j' receives from its neighbors is no more than what is received by agent i. According to Assumption 3, we can conclude that $x_{j'}^* \geq x_i^*$, which contradicts the fact that $x_{j'}^* < x_i^*$. Hence, this statement follows.

(ii) Suppose an agent $i \leq n_h\left(\boldsymbol{s}^*\right)$ is connected with all $i' \leq n_h\left(\boldsymbol{s}^*\right)$, we can use an approach similar to the proof of statement (ii) to show a contradiction. Hence, for any agent $i \leq n_h\left(\boldsymbol{s}^*\right)$, there is an agent $i' \leq n_h\left(\boldsymbol{s}^*\right)$ with $\overline{g}_{ii'}^* = 0$ and this statement follows.

(iii) Suppose that there is a $j > n_h\left(\boldsymbol{s}^*\right)$ such that $g_{ji}^* = 0,\ \forall i \leq n_h\left(\boldsymbol{s}^*\right)$. This implies that agent j does not establish links with others. If, on the contrary, agent j establishes links, then any of these links is directed to some agent $j' > n_h\left(\boldsymbol{s}^*\right)$, but then j can strictly increase its utility by switching a link from j' to some $i \leq n_h\left(\boldsymbol{s}^*\right)$.

Using the same idea as statement (i), agent j also receives no link from others. As a result, agent j is isolated and produces an amount $\overline{x}$, which contradicts the fact of it being a low producer. Hence, this statement follows. ■

3) Proof of Lemma 5

Suppose there is an agent $j > n_h\left(\boldsymbol{s}^*\right)$ with $g_{jj'}^* = 1$ for some $j' > n_h\left(\boldsymbol{s}^*\right)$. Similar to Lemma 3, it implies that $\overline{g}_{ji}^* = 1$ holds for all $i \le n_h\left(\boldsymbol{s}^*\right)$. Otherwise, j can strictly increase its utility by switching the link from j' to i. Since it has already been proved in Lemma 3(ii) that a high producer never forms a link with a low producer, we have $\overline{g}_{ji}^* = 1$ for all $i \le n_h\left(\boldsymbol{s}^*\right)$. Hence, Lemma 4 follows. ■

Appendix C

1) Proof of Theorem 2

(i) Let $n_l\left(\boldsymbol{s}^*\right) = n - n_h\left(\boldsymbol{s}^*\right)$ denote the population of low producers in an equilibrium. Since each low producer forms links to at least one high producer, we should have the following inequality

$$v\left(\tilde{x}\left(\boldsymbol{x}^*\right)\right) > k, \tag{8}$$

in order to sustain a low producer's incentive to form a link. Hence, $\tilde{x}\left(\boldsymbol{x}^*\right)$ is lower bounded as follows:

$$\tilde{x}\left(\boldsymbol{x}^*\right) > \frac{k}{c}. \tag{9}$$

It is obvious that there exists an integer L_h such that

$$v\left(\left(L_h + 1\right)^{\frac{1}{\rho}} \frac{k}{c}\right) - v\left(L_h^{\frac{1}{\rho}} \frac{k}{c}\right) < k, \tag{10}$$

and hence a low producer has no incentive to form links to more than L_h high producers.

Next, suppose there is a constant upper bound $\overline{n}_h$ for the population of high producers in any equilibrium, which is independent of the value of n. That is $n_h\left(\boldsymbol{s}^*\right) < \overline{n}_h$, $\forall \boldsymbol{s}^*, n$.

We then further classify low producers into the following three classes:

(a) A low producer who only forms links to high producers and does not receive any link from other low producers;

(b) A low producer who only forms links to high producers and also receives links from other low producers;

(c) A low producer who forms links to all high producers and some low producers.

For notational convenience, the population sizes of the above three classes in an equilibrium are denoted as $n_{la}\left(\boldsymbol{s}^*\right)$, $n_{lb}\left(\boldsymbol{s}^*\right)$, and $n_{lc}\left(\boldsymbol{s}^*\right)$, respectively.

It is obvious that production level of a low producer of type (a) is bounded away from 0, denoted as x_{la}, since the total amount of information it acquires from others will not exceed a finite amount $L_h\overline{x}$. Therefore, each high producer can only be connected to a finite number of low producers of type (a) (otherwise, its production level will go arbitrarily to zero and be smaller than x_{la}, which leads to a contradiction). Given the fact that $n_h\left(\boldsymbol{s}^*\right)$ is upper bounded, it can be concluded that there is a constant upper bound $\overline{n}_{la}$ such that $n_{la}\left(\boldsymbol{s}^*\right) < \overline{n}_{la}$, $\forall \boldsymbol{s}^*, n$.

Now consider a low producer j who belongs to type (b). Let x_j^* denote its production level and we should also have $x_j^* > \frac{k}{c}$. Otherwise, no agent has the incentive to form links to it. Using the same argument as that for type (a), it can be concluded that there is a constant upper bound $\overline{n}_{lb}$ such that $n_{lb}\left(\boldsymbol{s}^*\right) < \overline{n}_{lb}$, $\forall \boldsymbol{s}^*, n$.

To analyze low producers of type (c), we further classify them into the following two sub-classes and denote their populations as $n_{lc1}\left(\boldsymbol{s}^*\right)$ and $n_{lc2}\left(\boldsymbol{s}^*\right)$, respectively.

(c1) a low producer of type (c) who produces an amount of information which is higher than $\frac{k}{c}$;

(c2) a low producer of type (c) who produces an amount of information which is lower than $\frac{k}{c}$.

It is obvious that $n_{lc1}\left(\boldsymbol{s}^*\right)$ is also upper-bounded. Meanwhile, we have that any two agents of type (c2) do not mutually form links to each other since the link formation cost exceeds the benefit they can obtain from mutual information sharing. Therefore, the maximum number of neighbors that an agent of type (c2) has is $n_h\left(\boldsymbol{s}^*\right) + n_{la}\left(\boldsymbol{s}^*\right) + n_{lb}\left(\boldsymbol{s}^*\right) + n_{lc1}\left(\boldsymbol{s}^*\right)$, which is upper-bounded. Using a similar argument as what we did for other types of agents, $n_{lc2}\left(\boldsymbol{s}^*\right)$ should also be upper-bounded. As a result, the total population $n_h\left(\boldsymbol{s}^*\right)+n_{la}\left(\boldsymbol{s}^*\right)+n_{lb}\left(\boldsymbol{s}^*\right)+n_{lc1}\left(\boldsymbol{s}^*\right)+n_{lc2}\left(\boldsymbol{s}^*\right)$ is upper-bounded by a constant number that is independent of the value of n, which leads to a contradiction.

Therefore, we conclude that there is no constant upper bound for $n_h\left(\boldsymbol{s}^*\right)$. Particularly, for any integer value m_h, we can always find a value of m such that $\inf_{\boldsymbol{s}^*\in S_n^*}\left\{n_h\left(\boldsymbol{s}^*\right)\right\}>m_h$ when $n>m$, where S_n^* denote the set of equilibrium strategy profiles when the network size is n. Hence, it can be concluded that $\inf_{\boldsymbol{s}^*\in S_n^*}\left\{n_h\left(\boldsymbol{s}^*\right)\right\}\to\infty$ when $n\to\infty$. Alternatively speaking, we can always find a sufficiently large n with which $n_h\left(\boldsymbol{s}^*\right)$ is also sufficiently large in any equilibrium such that each low producer only forms links to high producers because high producers have produced sufficient amount of information for any agent to consume. In this case, no low producer forms links to other low producers.

Consider two low producers j and j'. Let $d_j\left(\boldsymbol{s}^*\right)$ and $d_{j'}\left(\boldsymbol{s}^*\right)$ denote their degrees in equilibrium, i.e. the numbers of their neighbors, respectively. Without loss of generality, we assume that $d_j\left(\boldsymbol{s}^*\right)>d_{j'}\left(\boldsymbol{s}^*\right)$, or alternatively, $d_j\left(\boldsymbol{s}^*\right)\geq d_{j'}\left(\boldsymbol{s}^*\right)+1$. Since agent j has no incentive to delete its existing links formed to others, we have that

$$r_j\left(\boldsymbol{x}^*,\boldsymbol{g}^*\right)-r_{j'}\left(\boldsymbol{x}^*,\boldsymbol{g}^*\right)>\left(d_j\left(\boldsymbol{s}^*\right)-d_{j'}\left(\boldsymbol{s}^*\right)\right)k, \tag{11}$$

where $r_i\left(\boldsymbol{x},\boldsymbol{g}\right)\triangleq v\left(\left[x_i^\rho+\sum_{j\in N_i\left(\bar{g}\right)}x_j^\rho\right]^{1/\rho}\right)-cx_i$ is defined as agent i's *utility from information consumption and production*. Similarly, since agent j' has no incentive to form new links, we have that

$$r_j\left(\boldsymbol{x}^*,\boldsymbol{g}^*\right)-r_{j'}\left(\boldsymbol{x}^*,\boldsymbol{g}^*\right)>\left(d_j\left(\boldsymbol{s}^*\right)-d_{j'}\left(\boldsymbol{s}^*\right)\right)k. \tag{12}$$

Hence, there is a contradiction and each low producer should form links to the same number of high producers, denoted as $q\left(\boldsymbol{g}^*\right)$, and has the same production level, denoted as $\underset{\sim}{x}\left(\boldsymbol{x}^*\right)$. Hence, statement (i) of this theorem follows.

(ii) This statement is a direct byproduct from statement (i) and the proof is omitted here. ■

2) Proof of Theorem 3

It has been proved in Theorem 2 that we can always find a sufficiently large value T such that low producers will not mutually form links to each other in any equilibrium when the network size $n > T$.

It is obvious that we can find a constant value $\mu \in (0,1)$ such that $\inf_{s^* \in S_n^*} \{n_h(s^*)\} / n > \mu$, $\forall n < T$. Now look at the case when $n > T$. It can be learned from (10) that a low producer cannot connect to more than L_h high producers (hub agents), which upper-bounds the amount of information it receives from others at $L_h \overline{x}$. Hence, there exists a constant $\underline{x}$ such that the production level of a low producer $\underset{\sim}{x}(x^*)$ is no less than $\underline{x}$ in any equilibrium s^* when $n > T$. Using the similar argument as in Theorem 1, the total number of links that each high producer receives from low producers is also upper-bounded by a constant, denoted as H_l. Given the population of high producers, i.e. $n_h(s^*)$, the population of low producers should be no more than $H_l n_h(s^*)$, which further delivers a lower bound on $\inf_{s^* \in S_n^*} \{n_h(s^*)\} / n$ as $1/(1+H_l)$. We take $\eta = \min(\mu, 1/(1+H_l))$ and hence Theorem 2 follows. ■

3) Proof of Corollary 1

To prove this corollary, we only have to show either $\inf_{s^* \in S_n^*} \left\{\sum_{i \in N} x_i^*\right\}$ or $\sup_{s^* \in S_n^*} \left\{\sum_{i \in N} x_i^*\right\}$ is $\Omega(n)$. In this proof, we show that $\inf_{s^* \in S_n^*} \left\{\sum_{i \in N} x_i^*\right\}$ is $\Omega(n)$. The arguments for $\sup_{s^* \in S_n^*} \left\{\sum_{i \in N} x_i^*\right\}$ similarly follows and is omitted here. There is a straightforward upper bound on $\inf_{s^* \in S_n^*} \left\{\sum_{i \in N} x_i^*\right\}$ as $n\overline{x}$. A loose lower bound on $\inf_{s^* \in S_n^*} \left\{\sum_{i \in N} x_i^*\right\}$ could be calculated as $\frac{\eta k}{c} n$. Therefore, $\inf_{s^* \in S_n^*} \left\{\sum_{i \in N} x_i^*\right\}$ is $\Omega(n)$ and Corollary 1 follows.

■

Appendix D

1) Proof of Proposition 1

This proposition is proved by classifying the strategy profiles into two classes and analyzing them separately: S^A includes strategy profiles with which there is no link in the network and S^B includes strategy profiles with which there is at least one link in the network. For notational convenience, the optimal social welfares achieved in S^A and S^B are denoted as W^A and W^B, respectively.

First, we analyze S^A to find out the strategy profile which maximizes the social welfare in it. When there is no link in the network, i.e. each agent is isolated with others, the optimal utility is achieved when each agent maximizes its individual utility and produces an amount $\overline{x}$. The optimal social welfare achieved in S^A is then $W^A = n\left[v\left(\overline{x}\right) - c\overline{x}\right]$.

For W^B, we derive a bound for it in this proof by separately analyzing the highest sum utility from information consumption and production, i.e. $\sum_{i \in N} r_i\left(\boldsymbol{x}, \boldsymbol{g}\right)$, and the lowest sum link formation cost over all agents.

We first analyze the lowest sum link formation cost for W^B. We prove that (i) if $W^A < W^B$, then W^B is achieved by a strategy profile with which each agent (with the exception of one agent) at least one connection with others, i.e. no agent is isolated; (ii) if W^B is achieved by a strategy profile with which some agent is isolated, then $W^A > W^B$. Therefore, if W^B is the social optimum, the lowest sum link formation cost contained in it is at least $\frac{nk}{2}$.

First we consider the scenario that the maximum social welfare in S^B is achieved by a strategy profile $\boldsymbol{s}$ with which each non-isolated agent has only one neighbor. We prove that if the social welfare delivered by $\boldsymbol{s}$ is higher than W^A, then $\boldsymbol{s}$ contains at least $\frac{n}{2}$ links.

Suppose that there is at least one agent being isolated in $\boldsymbol{s}$. We select an isolated agent i. It is obvious that i should produce an amount $\overline{x}$ of information in $\boldsymbol{s}$ (otherwise the social welfare can always be improved by adjusting the production level of agent i).

Without loss of generality, we assume that n is even. Then there are at least two agents being isolated in $\boldsymbol{s}$. If the utility received by an isolated agent is higher than what received by a non-isolated agent, then the social welfare can be further increased by removing the existing links. Hence, the social welfare achieved by $\boldsymbol{s}$, which is W^B is strictly smaller than W^A, which satisfies case (ii). On the contrary, if the utility received by an isolated agent is lower than what received by a non-isolated agent, then the social welfare can be further increased by adding a link between two isolated agents, which leads to a contradiction that $\boldsymbol{s}$ achieves the largest social welfare in S^B.

To sum up the results so far, it can be concluded that if the maximum social welfare in S^B is achieved by a strategy profile $\boldsymbol{s}$ with which each non-isolated agent has only one neighbor, then $\boldsymbol{s}$ contains at least $\frac{n}{2}$ links if $W^B > W^A$.

Next, we consider the scenario when the maximum social welfare in S^B is achieved by a strategy profile $\boldsymbol{s}$ with which there is at least one non-isolated agent who has more than one neighbors. Similar to the above, we also assume that there is an agent i being isolated in $\boldsymbol{s}$.

Let j denote one representative agent who has more than one neighbors and let j' denote one of its neighbors. Without loss of generality, we assume that $g_{j'j} = 1$. Construct a strategy profile $\boldsymbol{s}'$ that is the same to $\boldsymbol{s}$ except the entries $g'_{j'j} = 0$ and $g'_{j'i} = 1$. It is obvious that any agent other than i, j, and j' receives the same utility in both $\boldsymbol{s}$ and $\boldsymbol{s}'$. The utility of j' in $\boldsymbol{s}'$, however, is higher than that in $\boldsymbol{s}$ since j' now connects to i, whose production level is higher than j. Similarly, the utility of agent i also increases but the utility of agent j decreases. Nevertheless, it is easy to show that the increase on agent i's utility outweighs the decrease on agent j's utility and hence, the social welfare achieved by $\boldsymbol{s}'$ is larger than that achieved by $\boldsymbol{s}$. This leads to a contradiction to our assumption on $\boldsymbol{s}$ to be the socially optimal strategy profile and we have the following conclusion:

If the maximum social welfare in S^B is achieved by a strategy profile $\boldsymbol{s}$ with which there is at least one non-isolated agent who has more than one neighbors, then no agent is isolated and $\boldsymbol{s}$ contains at least $\frac{n}{2}$ links.

Summing up, we have that if $W^A < W^B$, the strategy profile that achieves W^B contains at least $\frac{n}{2}$ links. Therefore, the lowest sum link formation cost contained in W^B in this case is $\frac{nk}{2}$.

Next, we analyze the highest sum utility from information consumption and production that can be achieved in S^B, i.e.

$$\begin{aligned} &\max_{\boldsymbol{x},\boldsymbol{g}} \left\{ \sum_{i \in N} r_i \left(\boldsymbol{x}, \boldsymbol{g} \right) \right\} \\ &s.t.\ \ x_i \geq 0, \forall i \in N \\ &\qquad g_{ij} \in \left\{ 0,1 \right\}, \forall i,j \in N \end{aligned} \quad . \tag{13}$$

It is obvious that the optimum of (13) is always achieved in a complete network, i.e. $\overline{g}_{ij} = 1, \forall i, j \in N$. Therefore, each agent has the same amount of effective information, denoted as X^{opt}, and has a degree of $n-1$. Let $\boldsymbol{s}^{opt}$ denote the strategy profile that achieves the optimum of (13). For an agent i, its production level in $\boldsymbol{s}^{opt}$, denoted as x_i^{opt} should then satisfy the following equation:

$$nv'\left(X^{opt}\right)\left(\frac{X^{opt}}{x_i^{opt}}\right)^{1-\rho} = c. \tag{14}$$

Since (14) is satisfied for all agents, it can be concluded that each agent has the same production level in $\boldsymbol{s}^{opt}$, denoted as x^{opt}, and $X^{opt} = n^{\frac{1}{\rho}} x^{opt}$. Eq. (14) then becomes

$$n^{\frac{1-\rho}{\rho}} v'\left(n^{\frac{1}{\rho}} x^{opt}\right) = c \tag{15}$$

whose solution is $\widehat{x}_n$. We then have the solution of (13) being $n\left[v\left(n^{1/\rho}\widehat{x}_n\right) - c\widehat{x}_n\right]$.

By combining the highest sum utility from information consumption and production and the lowest sum link formation cost, we can give an upper bound on W^B as being $n\left[v\left(n^{1/\rho}\widehat{x}_n\right) - c\widehat{x}_n\right] - \frac{nk}{2}$ when $W^B > W^A$. Hence, the first part of Proposition 1 follows.

In order to prove the existence of the threshold $\overline{k}$, it is sufficient to show that if the social optimum is achieved in a non-empty network when $k = k_1$, then it is also achieved in a non-empty network when $k < k_1$, which is always true since W^B is an decreasing function on k. Similarly, if the social optimum is achieved in an empty network when $k = k_2$, then it is also achieved in an empty network when $k > k_2$. ■

2) Proof of Proposition 2

Since each agent can receive a utility which is at least $v\left(\overline{x}\right) - c\overline{x}$ by choosing to produce $\overline{x}$ and form no links to other agents, the utility it receives in any equilibrium should always be no less than this. Hence this proposition follows. ■

3) Proof of Theorem 4

This is a straightforward outcome of Proposition 1 and 2. ■

Appendix E

4) Proof of Theorem 5

To prove this theorem, we first define the concept of a component.

Definition 3 (Component). A component in a network is a group of agents which satisfies the following properties:

(i) There is a path between any two agents in the group.

(ii) There is no path between any agent in the group and any agent outside the group.

Suppose there is a component of size b in an equilibrium strategy profile $\boldsymbol{s}^*$. Since each agent in this component has the same amount of effective information. It is easy to know that each of them should also have the same production level at equilibrium, denoted as y_b^*. According to the equilibrium condition, we have

$$v'\left(b^{\frac{1}{\rho}} y_b^*\right) b^{\frac{1-\rho}{\rho}} = c\,. \tag{16}$$

For illustration purpose, we consider an isolated agent as a particular component with size 1.

First, it is easy to see that in an equilibrium strategy profile, each agent in a component of size larger than 1 only forms at most one link. If there is an agent forms two links connecting with two different agents in the same component, it can always delete one of them with the amount of its effective information unchanged and its utility monotonically increased.

Then, we show that in any strategy profile, an agent's utility from information consumption and production, which is $v\left(b^{\frac{1}{\rho}} y_b^*\right) - c y_b^*$ monotonically increases with the size of component it is in. This can be seen by taking the first-order partial derivative of $v\left(b^{\frac{1}{\rho}} y_b^*\right) - c y_b^*$ over b, which is

$$\begin{aligned} & v'\left(b^{\frac{1}{\rho}} y_b^*\right)\left(b^{\frac{1}{\rho}} \frac{\partial y_b^*}{\partial b} + \frac{1}{\rho} b^{\frac{1}{\rho}-1} y_b^*\right) - c\frac{\partial y_b^*}{\partial b} \\ & = \frac{1}{\rho} b^{\frac{1}{\rho}-1} y_b^* v'\left(b^{\frac{1}{\rho}} y_b^*\right) + \left(b-1\right) c\frac{\partial y_b^*}{\partial b} > 0 \end{aligned}. \tag{17}$$

With this result, we can prove that in any equilibrium strategy profile, there is at most one component whose size is larger than 1. Suppose in an equilibrium strategy profile $\boldsymbol{s}^*$ where there are two components C_1 and C_2 of size b_1 and b_2. Without loss of generality, we assume that $1 < b_1 \leq b_2$. Then for an agent i in C_1 who forms a link, it is always beneficial to switching this link to any agents in C_2, which leads to a contradiction to the fact that $\boldsymbol{s}^*$ is an equilibrium.

Now consider another equilibrium where there is an isolated agent j and a component C whose size is $b > 1$. With a little abuse of notation, this equilibrium is also denoted as $\boldsymbol{s}^*$. Obviously, j has no incentive to form a link and connect itself with C if and only if the link formation cost $k > v\left(\left(b\left(y_b^*\right)^\rho + \left(z_b\right)^\rho\right)^{\frac{1}{\rho}}\right) - v\left(\overline{x}\right) - c\left(z_b - \overline{x}\right)$, where z_b is the solution of the following equation:

$$v'\left(\left(b\left(y_b^*\right)^\rho + \left(z_b\right)^\rho\right)^{\frac{1}{\rho}}\right)\left(\frac{\left(b\left(y_b^*\right)^\rho + \left(z_b\right)^\rho\right)^{\frac{1}{\rho}}}{z_b}\right)^{1-\rho} = c. \tag{18}$$

Here, $v\left(\left(b\left(y_b^*\right)^\rho + \left(z_b\right)^\rho\right)^{\frac{1}{\rho}}\right) - cz_b$ is the largest utility from information consumption and production that j can receive by forming a link with C and $v\left(\overline{x}\right) - c\overline{x}$ is its current utility. Similar to $v\left(b^{\frac{1}{\rho}} y_b^*\right) - cy_b^*$, we can prove that $v\left(\left(b\left(y_b^*\right)^\rho + \left(z_b\right)^\rho\right)^{\frac{1}{\rho}}\right) - cz_b$ also monotonically increases with b and hence, if an isolated agent has no incentive to form a link with one component, it also has no incentive to form a link with any component with smaller size. That is, let $\gamma_b = v\left(\left(b\left(y_b^*\right)^\rho + \left(z_b\right)^\rho\right)^{\frac{1}{\rho}}\right) - v\left(\overline{x}\right) - c\left(z_b - \overline{x}\right)$, it monotonically increases with b.

Summarizing all the above, it can be concluded that if $k < \gamma_1$, there is a unique equilibrium where there is a unique component in the network such that there is a path between any two agents; and if $\gamma_1 < k < \gamma_{n-1}$, there are multiple possible equilibria, each of which contains a minimally connected component whose size $b < n$ and the rest $n - b$ agents being isolatd. When $k > \gamma_{n-1}$, the network has a unique equilibrium where all agents are isolated with each other. Hence, we have $k_{\max} = \gamma_{n-1}$ and $k_{\min} = \gamma_1$. ■

5) *Proof of Corollary 2*

We have known from Theorem 5 that γ_b monotonically increases with b and hence, $k_{\max}$ monotonically increases with n. Meanwhile, $k_{\min} = \gamma_1$ is constant and does not change with n. ■

6) *Proof of Theorem 6*

Consider a strategy profile $\boldsymbol{s}^{\#}$ which achieves the social optimum. We look at a component C of size b in the network. Similar to Theorem 5, all agents in this component has the same production level in $\boldsymbol{s}^{\#}$, denoted as $y_b^{\#}$ which satisfies the following equality:

$$v'\left(b^{\frac{1}{\rho}} y_b^{\#}\right) b^{\frac{1}{\rho}} = c\,. \tag{19}$$

Due to the concavity of $v\left(\bullet\right)$, we have $y_b^{\#} > y_b^{*}$ when $b > 1$ and $y_b^{\#} = y_b^{*}$ when $b = 1$.

Now suppose there is an isolated agent i in $\boldsymbol{s}^{\#}$. The sum utility of i and agents in C are

$$v\left(\overline{x}\right) + bv\left(b^{\frac{1}{\rho}} y_b^{\#}\right) - c\left(\overline{x} + by_b^{\#}\right) - k\left(b-1\right). \tag{20}$$

If i forms a link to any agent in C, a new component of size $b+1$ is formed and the optimal sum utility of all agents in it is now

$$\left(b+1\right)v\left(\left(b+1\right)^{\frac{1}{\rho}} y_{b+1}^{\#}\right) - c\left(b+1\right)y_{b+1}^{\#} - kb\,. \tag{21}$$

The difference between (20) and (21) is

$$v\left(\overline{x}\right)+bv\left(b^{\frac{1}{\rho}}y_b^{\#}\right)-\left(b+1\right)v\left(\left(b+1\right)^{\frac{1}{\rho}}y_{b+1}^{\#}\right)-c\left(\overline{x}+by_b^{\#}-\left(b+1\right)y_{b+1}^{\#}\right)+k\,. \tag{22}$$

Since $\boldsymbol{s}^{\#}$ is optimal, we should have (22) larger than 0, i.e.

$$k\geq\left(b+1\right)v\left(\left(b+1\right)^{\frac{1}{\rho}}y_{b+1}^{\#}\right)-v\left(\overline{x}\right)-bv\left(b^{\frac{1}{\rho}}y_b^{\#}\right)-c\left(\left(b+1\right)y_{b+1}^{\#}-\overline{x}-by_b^{\#}\right). \tag{23}$$

Denote the RHS of (23) as κ_b, we prove that the first-order derivative of κ_b over b is larger than 0, as shown below:

$$\begin{aligned}\frac{\partial\kappa_b}{b}&=v\left(\left(b+1\right)^{\frac{1}{\rho}}y_{b+1}^{\#}\right)+\left(b+1\right)v'\left(\left(b+1\right)^{\frac{1}{\rho}}y_{b+1}^{\#}\right)\left(\frac{1}{\rho}\left(b+1\right)^{\frac{1}{\rho}-1}y_{b+1}^{\#}+\left(b+1\right)^{\frac{1}{\rho}}\frac{\partial y_{b+1}^{\#}}{\partial b}\right)\\&\quad-v\left(b^{\frac{1}{\rho}}y_b^{\#}\right)-bv'\left(b^{\frac{1}{\rho}}y_b^{\#}\right)\left(\frac{1}{\rho}b^{\frac{1}{\rho}-1}y_b^{\#}+b^{\frac{1}{\rho}}\frac{\partial y_b^{\#}}{\partial b}\right)\\&\quad-cy_{b+1}^{\#}-c\left(b+1\right)\frac{\partial y_{b+1}^{\#}}{\partial b}+cy_b^{\#}+cb\frac{\partial y_b^{\#}}{\partial b}\\&=v\left(\left(b+1\right)^{\frac{1}{\rho}}y_{b+1}^{\#}\right)-v\left(b^{\frac{1}{\rho}}y_b^{\#}\right)+\frac{1-\rho}{\rho}c\left(y_{b+1}^{\#}-y_b^{\#}\right)\end{aligned} .(24)$$

Because $v\left(\left(b+1\right)^{\frac{1}{\rho}}y_{b+1}^{\#}\right)-v\left(b^{\frac{1}{\rho}}y_b^{\#}\right)>0$ and $y_{b+1}^{\#}-y_b^{\#}>0$, we have $\frac{\partial\kappa_b}{b}>0$. Therefore, if an isolated agent connects with a component, the sum utility on information consumption and production has a larger increase if the component has a larger size.

Using (24), we can first show that there is at most one component whose size is larger than 1 in $\boldsymbol{s}^{\#}$. Suppose there are two components C_1 and C_2 with their sizes being $1<b_1\leq b_2$ without loss of generality. Since C_1 is minimally connected, then there is at least one agent, denoted as j, who does not receive any link formed by other agents. Hence, if agent j switches its link to C_2, it is no longer connected to C_1. After this switch, the sum link formation cost for agents in C_1 and C_2 does not change since the total number of links remains the same. Meanwhile, the change on the sum utility on information consumption and production is $\kappa_{b_2}-\kappa_{b_1-1}$ which is strictly larger than 0 since κ_b

monotonically increases with b and $b_2 > b_1 - 1$. Hence, it is always optimal to switch the links in C_1 to C_2 in order to increase the social welfare and there is at most one component whose size is larger than 1 in any socially optimal strategy profile $\boldsymbol{s}^{\#}$.

As a result, we can derive $k_{\max}^{opt} = \kappa_{n-1}$ and $k_{\min}^{opt} = \kappa_1$. When $k < k_{\min}^{opt}$, the social optimum is achieved in a minimally connected network and when $k > k_{\max}^{opt}$, the social optimum is achieved in an empty network. Meanwhile, when $k \in (\kappa_b, \kappa_{b+1})$ for any $b \in \{1, \ldots, n-2\}$, the social optimum is achieved in a network with a minimally connected component of $b+1$ agents and other $n-b-1$ isolated agents.

With simple computation, it can be shown that $\kappa_{n-1} > \gamma_{n-1}$ and $\kappa_1 > \gamma_1$. Hence, this theorem follows. ■

Appendix F

1) Bilateral link formation with continuous link strength

In this section, we analyze an alternative model where the linking choice between two agents is not binary, but continuous. That is, the linking strategy of an agent i now becomes a vector $\boldsymbol{p}_i = \left\{p_{i1}, \ldots, p_{in}\right\}$, where $p_{ij} \in \left[0,1\right]$ and $p_{ii} = 0$. We define the strength of a link to be $\overline{g}_{ij} = \overline{g}_{ji} = \max\left\{p_{ij} + p_{ji}, 1\right\}$. When $\overline{g}_{ij} = 1$, we say that the link between agents i and j is of full strength. This model possesses some similarity to the bilateral link formation since the strength of a link is now determined by the investments of two agents and the link formation cost is also two-sided. The utility of agent i is then defined as

$$u_i\left(\boldsymbol{x}, \boldsymbol{g}\right) = v\left(\left[x_i^{\rho} + \sum_{j\in N} \overline{g}_{ij} x_j^{\rho}\right]^{\frac{1}{\rho}}\right) - cx_i - k\sum_{j\in N} p_{ij} . \tag{25}$$

Proposition X. *In an equilibrium* $\boldsymbol{s}^* = \left(\boldsymbol{x}^*, \boldsymbol{g}^*\right)$, *if* $p_{ij}^* + p_{ji}^* > 0$, *then* $p_{ij}^* + p_{ji}^* = 1$.

Proof: Suppose $p_{ij}^* + p_{ji}^* \in \left(0,1\right)$ for some agents i and j and we assume $p_{ij}^* > 0$ without loss of generality. By taking the first-order partial derivative in p_{ij} over $y_i\left(\boldsymbol{x}, \boldsymbol{g}\right)$, we have that

$$v'\left(X_i^*\right)\left(X_i^* / x_j^*\right)^{1-\rho} x_j^* = \rho k . \tag{26}$$

According to (27), we then have $\left(x_j^*\right)^{\rho} / \left(x_i^*\right)^{1-\rho} = \rho k / c$. Since this analysis is symmetric for i and j, we have that $\left(x_j^*\right)^{\rho} / \left(x_i^*\right)^{1-\rho} = \left(x_i^*\right)^{\rho} / \left(x_j^*\right)^{1-\rho}$ and thus $x_i^* = x_j^*$. Hence, the equilibrium strategy profile should be symmetric where each agent produces the same amount of information. Otherwise, we can always find an agent who has the incentive to switch its link from one of its neighbors to some other agent who has a higher production level. Since we only consider asymmetric strategy in this paper, the above scenario cannot happen and thus we have $p_{ij}^* + p_{ji}^* = 1$ always holds. ■

Therefore, when the link formation cost is continuous and two-sided, each link still has full strength at equilibrium and there is no redundant investment on any link. Consequently, the analysis in the paper can be extended straightforwardly.