Information Production and Link Formation in Social Computing Systems

Yu Zhang, *Student Member, IEEE* and Mihaela van der Schaar, *Fellow, IEEE*

Abstract—Social computing provide a popular, cost-effective and scalable framework for building new engineering systems as well as improving the performance of numerous existing systems. However, the self-interest of agents of such systems generates intrinsic incentive problems. This work analyzes these incentive problems from several points of view. First, we analyze the trade-offs (of each individual agent) between the costs and benefits of producing information personally and forming links to collect information (from other agents), and the strategic implications of these trade-offs. A central point of the analysis is that information is assumed to be heterogeneous (rather than homogeneous as in previous analyses) and agents value this heterogeneity. The analysis has implications for the topology that emerges endogenously. For large populations, the implication is that the topology is necessarily of a core-periphery type: hub agents (at the core of the network) produce and share most of the information, while spoke agents (at the periphery of the network) derive most of their information from hub agents, producing little of it themselves. As the population becomes larger, the number of hub agents and the total amount of information produced grow in proportion to the total population. Our conclusions had been conjectured for many social computing systems but not been previously derived in any formal framework, and are in stark contradiction to the "law of the few" that had been established in previous work, under the assumption that information is homogeneous and part of the endowment of agents, rather than heterogeneous and produced.

Index Terms—Social computing, information production, network formation.

I. INTRODUCTION

SOCIAL computing is booming nowadays [1] - [3] and has recently led to a new trend: the integration of social computing techniques with communication networks to enable individuals with similar interests to connect to each other over a mobile or low-cost network infrastructure to share and disseminate user-generated information, such as multi-modal content, geographical information, event-related information, personal sensory information, etc. For instance, [4] and [5] propose mobile content distribution architectures relying on social relationship and providing economic incentives to offload the bandwidth-intensive task of content dissemination from service providers to the mobile subscribers in mobile and wireless networks. Distributed spectrum sensing in peer-to-peer cognitive networks is proposed in [6] to efficiently map the spectrum opportunities based on contributions from

individual cognitive terminals. The work of [7] formulates the location-aware information sharing and link formation in vehicular networks as a coalition game to improve the efficiency of road information dissemination among vehicles.

In all the aforementioned social computing applications, the decentralized strategic interactions of agents have decisive impact on the emerging network topologies as well as the stability and the efficiency of emerging networks and distributed systems. In this paper, we are focusing on two key features that characterize the strategic behavior of agents: information production and link formation. Information production refers to the agents' decisions to personally generating information. Here we use the abstract concept of "information" to represent any knowledge, multi-modal content, and other kinds of data and service that can be exchanged and shared in social computing systems. Several examples on information production would be the upload and creation of blogs, videos and photos on online platforms with user-generated content [1], the download and purchase of content from service provider in peer-to-peer networks [3] and content distribution networks [4], and the creation of local spectrum information sensed by individual cognitive terminals in peer-to-peer cognitive networks [6]. Link formation refers to all kinds of agents' decisions regarding exchanging the produced information with other agents, i.e. how agents create and dissolve links with each other to acquire information. For example, a peer node in P2P networks makes active decisions on which peers it should connect to and download content from; in peer-to-peer spectrum sensing, a cognitive terminal can request and exchange information about spectrum opportunities which have already been discovered by other terminals [6] [8]. Most of the existing works assume that agents decisions on these actions are exogenously determined by a system designer aiming to maximize the overall network utility. Hence, agents do not proactively make decisions but rather obediently follow the actions prescribed by the designer. While this assumption is realistic for traditional communication networks composed of obedient agents (e.g. sensor nodes), it does no longer hold in social computing systems, where networks are usually formed in an ad-hoc fashion by self-interested agents determining their own strategies in order to maximize their own utilities. Hence, a novel framework on studying the incentive of agents in the joint decision on information production and link formation according to their own self-interest is necessary. Such a framework does not only help to understand the agents' behavior in social computing systems but also facilitates the design of incentive protocols which encourage the agents to cooperate with each other in a way to optimize the system

efficiency. In the rest of this paper, we use "social computing system" and "network" interchangeably.

Self-interested link formation in social computing systems has been studied by microeconomics and network science researchers (see e.g. [9]), who analyze how the agents' self-interests lead to strategic link creation in a network and determine what network topologies arise. A simple model is proposed in [10], where the problem of link formation is formulated as a non-cooperative game among strategic agents. The agents can select which links to create with other agents in order to individually maximize their own utilities by trading off the potential rewards obtained from forming a link (e.g. information acquisition) against the incurred link creation cost (e.g. payment and maintenance costs for links). Analysis of this model can lead to useful predictions of what "equilibrium" topologies emerge. This basic model has subsequently been extended in various directions. For example, [11] studies the bilateral network formation in which the link creation requires mutual consent and the cost is two-sided. The network formation problem for the scenario where agents are heterogeneous is analyzed in [12], which shows that a strict equilibrium network is minimal and, conversely, every minimal network is a strict equilibrium for suitable costs and benefits [12].

Importantly, there are two key drawbacks that prevent these existing models from being successfully applied to social computing systems. First, most of these works assume that agents are endowed with exogenous amounts of information and focus solely on the strategic aspect of link formation. The fact that agents usually have the capability to self-produce information is neglected. Second, these works treat information collected from different agents as homogeneous, i.e. being equally valued and perfectly substitutable, while in social computing systems, information from different agents is heterogeneous with respect to location, devices used for gathering this information, content contained, etc., and thus it cannot be perfectly substitutable with each other. Meanwhile, given the information heterogeneity, an agent's benefit from information consumption no longer only depends on the total amount of information it consumes, but also on how many types of information and what amount of each type it acquires. Therefore, an agent normally has certain appreciation for information variety by valuing information produced by numerous sources higher than information produced by a single source [13].

Based on these characteristics of the agents, our work proposes a novel Information Production and Link Formation game (IPLF game) for social computing systems to jointly study agents' strategic behavior on information production and link formation. To deal with the agents' appreciation for information variety, we deploy the well-known Dixit-Stiglitz utility function [16] to capture the impact of information heterogeneity on the agents utilities. Using this formalism, we study what asymptotic equilibria emerge in networks of large sizes. Different from the idealized works in economics, e.g. [15], which assume perfectly substitutable information

(i.e. agents have no appreciation for variety) and predict the occurrence of "the law of the few" [15] [3], our analysis shows that when the size (population) of a network is sufficiently large, every (strict) non-cooperative equilibrium of the IPLF game consists of a hierarchical core-periphery structure with all agents belonging to one of the two types: hub agents producing large amounts of information have a large number of connections and serve as the major source of information sharing; and spoke agents produce and share limited amounts of information and mainly consume information acquired from hub agents. Importantly, as the network size grows, the population of hub agents as well as the total amount of information produced in the network grows proportionally to the network size. This shows that in social computing systems where agents appreciate information variety, the "law of the few" does not exist with the production and sharing no longer being dominated by a small group of powerful agents but rather being more distributed, and the network structure remaining flat but not scale-free [4].

Following the characterization on the equilibria in the IPLF game, we then analyze the efficiency of such equilibrium topologies and compare it with the social optimum by providing bounds on the Price of Stability and the Price of Anarchy. Stemming from this basic model, we further propose and investigate the IPLF game which incorporates indirect information sharing, which is a common feature in social computing systems that an agent cannot only consume and access information produced by its neighbors but also that produced by the neighbors of its neighbors.

The remainder of this paper is organized as follows. In Section II, we describe our basic model of the IPLF game. In Section III, we characterize the emerging non-cooperative equilibria. In Section IV, the model of the IPLF game with indirect information sharing is presented and discussed. We conclude in Section V and also outline future research topics.

II. BASIC MODEL

We consider information sharing in a social computing system where agents have the capability to produce information personally as well as form links with others in order to acquire information that they produce. The definition of "agent" in this paper is very general. It can represent individual nodes, devices, peers, and also autonomous sub-networks or groups of nodes cooperating with each other to form an integrated entity participating in the IPLF game. Examples of information being shared are multi-modal content, environmental sensory data in sensor networks, road information in vehicular networks, geographical information in location-based services, etc.

In all these different applications, the precise formulation of actions and utilities of the participating agents will depend on details of the application. For example, precisely how agents

---

[1]A strict equilibrium is a strategy profile in which each agent's utility is strictly larger than the resulting utility by choosing any other strategy.

[2]A minimal network is a topology where there is a unique path between any two agents in the network.

[3]"The law of the few" [15] states that under perfect information substitutability, only a small amount of agents in a network will produce information with the fraction of information producers in the total population goes to 0. Meanwhile, the total amount of information produced in the network is upper-bounded.

[4]A scale-free network is a network whose degree distribution follows a power law asymptotically. If a network is scale-free, the fraction of agents with large numbers of connections in the entire population decays fast to 0 as the network size grows [17].

can form links with others, whether there is any budget or practical constraints on connections such as physical locations and bandwidth limits, how does an agent benefit/cost from the presence of a link, etc. In this work, we use a stylized model to formulate the IPLF game without delving into the idiosyncrasies of any particular application, in order to capture the basic trade-offs and draw qualitative insights about the effects of self-interested behavior on the network structure of social computing systems. However, it should be noted that various alternative models are ready to be extended from this basic model for the analyses of particular applications, as discussed in Section V.

Let $N = \{1, 2, \ldots, n\}$ denote the set of agents in the system, where i and j represent typical members. Each agent i determines the amount of information it produces, which is called as its production level and is denoted by $x_i \in \mathbb{R}^+$. We assume a linear cost on information production as in [12] and hence, a cost of $c \in \mathbb{R}^+$ incurs to an agent for each unit of information it produces. Examples of the cost could be the energy consumption on spectrum sensing in P2P cognitive networks, subscription fees paid to service providers for content download in content distribution networks, etc. Besides the self-production, each agent also determines whether to create links with other peer agents in order to acquire information from them. Here we consider unilateral link formation as in [19], where links are created by the unilateral actions of agents and link costs are one-sided. An agent i's link formation action to another agent j is represented by a binary variable $g_{ij} = \{0, 1\}$. We set $g_{ij} = 1$ if agent i forms a link to agent $j \neq i$ and $g_{ij} = 0$ otherwise. For convenience, we set $g_{ii} = 0$ for all $i \in N$. For illustration purposes, we refer to the agent who forms a link and the agent with whom the link is formed as the creator and recipient of the link, respectively. Forming a link incurs a cost k, which abstracts all the costs and payments for creating and maintaining the link, as well as costs incurred in information exchange and transmission (e.g. uploading and downloading cost, energy consumption in transmission, etc.). Here we assume the value of k to be constant for each link. Meanwhile, since a link is formed unilaterally without the consent of the recipient, we assume that the creator undertakes the entire cost k (i.e. if there is some cost incurred for the recipient in receiving a link, the creator will also compensate the receiver by paying for it). The link formation decision by agent i is then represented by a row vector $\mathbf{g}_i = (g_{ij})_{j \in N, j \neq i} \in \{0, 1\}^{n-1}$. We assume that the information exchange across a link is undirected and thus, the connection status between agents i and j can be represented by a variable $\bar{g}_{ij} = max\{g_{ij}, g_{ji}\}$ which takes the value of 1 when they are connected and called as "neighbors" of each other. We also assume that $\bar{g}_{ii} = 1$. Since the link is undirected and we assume that each link has sufficient capacity to support the information transmission, neighboring agents can mutually access all information produced by each other

We consider a simultaneous move game where agents make decisions on information production and link formation at the same time. The set of strategies of agent i is thus denoted by $S_i = \mathbb{R}^+ \times \{0, 1\}^{n-1}$. A strategy profile in the IPLF game is written as $\mathbf{s} = (\mathbf{x}, \mathbf{g})$, where $\mathbf{x} = (x_i)_{i=1}^n$ and $\mathbf{g} = (\mathbf{g}_i)_{i=1}^n$ denote the information production and link for-

mation decisions of all agents, respectively. Let $\bar{\mathbf{g}} \triangleq [\bar{g}_{ij}]_{i,j \in N}$ denote the connectivity graph of the network, we also define $N_i(\mathbf{g}) \triangleq \{j | g_{ij} = 1\}$ as the set of agents to whom agent i forms a link and $N_i(\bar{\mathbf{g}}) \triangleq \{j | \bar{g}_{ij} = 1\}$ as the set of agents to whom agent i is connected, i.e. the neighbors of agent i.

Agents in the IPLF game benefit from consuming information available to them. In the basic model, we assume pure local externalities [15]: each agent only consumes information personally produced by itself and its neighbors [5], and cannot consume information produced by another agent who is more than one hop away from it on the connectivity graph $\bar{\mathbf{g}}$. As mention previously, existing works on strategic link formation assume that information produced by different agents is perfectly substitutable [10] [15], i.e. the total amount of information that an agent consumes fully determines its benefit. Therefore, an agent's benefit from information consumption is not affected by where the information comes from as long as the total amount of information consumed remains constant. This assumption, however, fails to capture agents' interests and benefit from consuming diverse information in social computing systems. For example, the spectrum information collected by different cognitive terminals at different locations should have different values in consumption and cannot be fully replaced. Hence, instead of assuming the perfect substitutability, we use the Dixit-Stiglitz preference model from [16] to capture agents' appreciation for information variety, under which an agent i's benefit from information consumption is given by

$$f_i(\mathbf{x}, \mathbf{g}) = v((x_i^\rho + \sum_{j \in N_i(\bar{\mathbf{g}})} x_j^\rho)^{1/\rho}) \tag{1}$$

for some function $v : \mathbb{R}^+ \to \mathbb{R}^+$ and some $\rho \in (0, 1)$. ρ measures an agent's appreciation for information variety. When $\rho < 1$, an agent obtains a higher benefit when it consumes information from a more diverse bundle of agents. It should be noted that when $\rho \to 1$, the agent's appreciation for information variety disappears and hence, the function (1) could also be used to model an agent's benefit when information is perfectly substitutable as in [10] and [15] by setting $\rho = 1$.

For the illustration purpose, we define $X_i \triangleq (x_i^\rho + \sum_{j \in N_i(\bar{\mathbf{g}})} x_j^\rho)^{1/\rho}$ as the amount of agent i's effective information. This could be interpreted as the amount of information that agent i really consumes after aggregating and processing all the information it acquires. The following two assumptions are imposed on the benefit function v:

Assumption 1: $v(\cdot)$ is a twice continuously differentiable, increasing, and strictly concave function.

Assumption 2: $v(\cdot)$ satisfies $v(0) = 0$, $v'(0) \triangleq \lim_{x \to 0^+} v(x) < \infty$, $v'(0) > \alpha$ where $\alpha > 0$ is a constant, and $\lim_{x \to \infty} v'(x) = 0$.

These two assumptions capture the diminishing marginal benefit in an agent's information consumption with respect to the amount of its effective information. That is, an agent's benefit increases with the amount of its effective information, while the rate of this increase decreases and approaches to 0.

---

[5] An analysis about indirect information sharing can be found in Section IV.

Hence, there always exists an upper bound on the amount of information that an individual agent is willing to acquire given the cost of information production and link formation.

Next, we analyze the effect of ρ on an agent's information consumption.

Lemma 1. Given a strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g})$, $(x_i^\rho + \sum_{j \in N_i(\bar{\mathbf{g}})} x_j^\rho)^{1/\rho} > (x_i^{\rho'} + \sum_{j \in N_i(\bar{\mathbf{g}})} x_j^{\rho'})^{1/\rho'}$ always holds for any $i \in N$ and any $0 < \rho < \rho' \leq 1$. ∎

Remark 1: For the proofs that are omitted in this paper, the complete proofs and computations can always be found in the online appendix [21].

This lemma can be proven by first showing that $(y^\rho + z^\rho)^{1/\rho} > (y^{\rho'} + z^{\rho'})^{1/\rho'}$ always holds for any $y, z > 0$ and $0 < \rho < \rho' \leq 1$ through the first-order partial derivatives of $(y^\rho + z^\rho)^{1/\rho}$ over ρ, which is always negative when $0 < \rho \leq 1$. Then this lemma follows as a straightforward extension.

Lemma 1 shows that given a strategy profile, an agent's level of appreciation for information variety increases as ρ becomes smaller, which is reflected in the increase on the amount of its effective information and the resulting benefit from information consumption.

With the above two assumptions, $f_i(\mathbf{x}, \mathbf{g})$ is also provably twice continuously differentiable, increasing and strictly concave in x_i, as shown in the following lemma.

Lemma 2. $f_i(\mathbf{x}, \mathbf{g}) = v((x_i^\rho + \sum_{j \in N_i(\bar{\mathbf{g}})} x_j^\rho)^{1/\rho})$ is twice continuously differentiable, increasing, and strictly concave in x_i on $\mathbb{R}^+$. ∎

Since $f_i(\mathbf{x}, \mathbf{g})$ is twice continuously differentiable, we further assume it to be sub-modular with the following assumption.

Assumption 3: $\frac{\partial^2 f_i}{\partial x_i \partial x_j}|_{x_i, x_j \in \mathbb{R}^+} < 0$.

Assumption 3 formalizes the substitutability among agents' information. Agent i's marginal benefit of production decreases against the amount of information produced by its neighbors. That is, the more information i acquires, the less incentive i has to produce information by itself.

To sum up, the utility of agent i is given by its benefit minus all incurred costs:

$$u_i(\mathbf{x}, \mathbf{g}) = v((x_i^\rho + \sum_{j \in N_i(\bar{\mathbf{g}})} x_j^\rho)^{1/\rho}) - cx_i - k|N_i(\mathbf{g})|. \quad (2)$$

We analyze the case of homogeneous agents in that v, c, k, and ρ are the same for all agents. We assume that $\alpha > c$ to ensure the network is socially valuable.

III. EQUILIBRIUM IN THE IPLF GAME

A. Equilibrium analysis for the basic model

We analyze the IPLF game as a non-cooperative one-shot game and consider pure strategies. Each agent maximizes its own utility given the strategies of others. A Nash equilibrium of the IPLF game is a strategy profile $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$ such that

$$u_i(\mathbf{s}_i^*, \mathbf{s}_{-i}^*) \geq u_i(\mathbf{s}_i, \mathbf{s}_{-i}^*), \forall \mathbf{s}_i \in \{0, 1\}^{n-1}, \forall i \in N. \quad (3)$$

Here, we use the convention that $\mathbf{s}_i$ and $\mathbf{s}_{-i}$ represents the strategies of agent i and all agents other than i, respectively. This section analyzes the equilibrium production and link formation behavior of agents. Given the definition of equilibrium

(3), an equilibrium production profile $\mathbf{x}^*$ satisfies the following equality for each agent:

$$v'(X_i^*)(X_i^*/x_i^*)^{1-\rho} = c, \ \forall i, \quad (4)$$

where $X_i^* = ((x_i^*)^\rho + \sum_{j \in N_i(\bar{\mathbf{g}})} (x_j^*)^\rho)^{1/\rho}$. That is, the marginal benefit of production should equal to the marginal cost and thus, an agent i has no incentive to produce more than x_i^* when the amount of its effective information is X_i^*.

We first derive the basic properties of agents' equilibrium behavior on information production and link formation. This is summarized in the following lemma. Although simple, these properties are important in characterizing the emerging equilibrium later.

Lemma 3. In any equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$ of the IPLF game, (i) $g_{ij}^* g_{ji}^* = 0$ for all $i, j \in N$; (ii) $x_i^* > 0$ for all $i \in N$; (iii) $x_i^* \leq \bar{x}$ for all $i \in N$ where $\bar{x}$ is the unique solution of the equation $v'(\bar{x}) = c$. ∎

We briefly explain Lemma 3.

(i) Since the information exchange across a link is undirected, if an agent i forms a link to another agent j it has already connected with, it incurs an additional cost of link formation but does not receive any extra information from j and hence, its utility strictly decreases by doing so. Hence in an equilibrium, there should be no redundant investment on link formation.

(ii) If an agent has no neighbors (i.e. no connections), it has a positive production level as $\alpha > c$. On the contrary, if an agent has neighbors, it also has a positive production level according to (4). Hence, an agent always has a positive production level and the information it acquires from neighbors can never fully replace its self-produced information.

(iii) According to Assumption 3, an agent has the largest marginal utility from production when it has no neighbors (i.e. it acquires no information from others). Hence, the highest production level could also be achieved in this case and is upper bounded due to the concavity of the benefit function (1).

The dependence of the value of $\bar{x}$ on c and v is neglected in the notation when no confusion is brought in. In the rest of this paper, $\bar{x}$ is referred as the maximum equilibrium production level.

The maximum equilibrium production level is critical in characterizing the relationship between the link formation cost and emerging equilibria of the IPLF game, as shown in the following theorem on the existence of the Nash equilibrium.

Theorem 1 (The Existence of Equilibrium). Pure strategy Nash equilibria always exist in the IPLF game and each equilibrium belongs to one of the following two types:

(i) each agent personally produces an amount $\bar{x}$ of information and no one forms any link;

(ii) each agent personally produces an amount strictly smaller than $\bar{x}$ and is connected with at least one other agent.

Proof Sketch: 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 directly. 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 i now becomes a vector $\mathbf{p}_i = \{p_{i1}, \ldots, p_{in}\}$, where $p_{ij} \in [0, 1]$ and $p_{ii} = 0$. We define the strength of a link to be $\bar{g}_{ij} = \bar{g}_{ji} = max\{p_{ij}, p_{ji}\}$. 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. We then show that each equilibrium of the IPLFM game has $p_{ij} \in \{0, 1\}$ for any $i, j \in N$, and thus is also an equilibrium of the IPLF game where the link formation choice is binary.

To prove the second part of the theorem, we classify all strategy profiles into two classes. The first class S^A contains strategy profiles with which there is no link in the network. The second class $S^B = S/S^A$ contains all other strategy profiles. Here S is the set of all strategy profiles. It can be shown that an equilibrium in S^A belongs to type (i) and an equilibrium in S^B should contain no isolated agent and belongs to type (ii). Otherwise, if a strategy profile in S^B contains isolated agents, then due to the fact that an isolated agent always has the largest production level, non-isolated agents will be attracted to connect with it, and which contradicts the definition of an equilibrium. ∎

Theorem 1 shows that in an equilibrium of the IPLF game, the network either is empty with all agents being isolated, or has no agent being isolated. In the rest of this section, we analyze the non-trivial case of strict equilibria with which the inequality (3) is strict, and we use equilibrium to mean strict pure strategy equilibrium when no confusion occurs. Without loss of generality, agents are ordered by their production levels in a strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g})$, i.e. $x_1 \geq x_2 \geq \cdots \geq x_n$. Given such ordering, there is always a positive integer $n_h(\mathbf{s}) \leq n$, such that $x_i = x_{n_h(\mathbf{s})}$ for all $i \leq n_h(\mathbf{s})$ and $x_j < x_{n_h(\mathbf{s})}$ for all $j > n_h(\mathbf{s})$. We call an agent $i \leq n_h(\mathbf{s})$ to be a high producer and an agent $j > n_h(\mathbf{s})$ to be a low producer. According to Lemma 3(iii), the production level of each high producer will never exceed $\bar{x}$ and is denoted as $\tilde{x}(\mathbf{x})$. To avoid the trivial case that the set of low producers is empty and all agents have the same production level, we assume that $n_h(\mathbf{s}) < n$ in this paper. A detailed analysis for strategy profiles where all agents produce the same amount of information can be found in the technical report [22].

By separating agents into two types, we are able to characterize in the following lemma the interactions between high and low producers at equilibrium.

Lemma 4. In an equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$, the following properties hold for $\mathbf{g}^*$: (i) For each $i \leq n_h(\mathbf{s}^*)$, $g_{ij}^* = 0$, $\forall j > n_h(\mathbf{s}^*)$; (ii) For each $i \leq n_h(\mathbf{s}^*)$, $\bar{g}_{i,i'}^* = 0$ for some $i' \leq n_h(\mathbf{s}^*)$ and $i \neq i'$; (iii) For each $j > n_h(\mathbf{s}^*)$, $g_{ji}^* = 1$, for some $i \leq n_h(\mathbf{s}^*)$.

Proof: The proof in Lemma 4 heavily relies on the use of contradictions. Here we only provide the proof of statement (i) due to the space limitation. The proofs for the other two statements follow similar ideas and can be found in the online appendix [21].

Suppose that there is an agent $j > n_h(\mathbf{s}^*)$ such that $g_{ij}^* = 1$. It is always true that i is connected with all other high producers, i.e. $\bar{g}_{ii'} = 1$, $\forall i' \leq n_h(\mathbf{s}^*)$ and $i' \neq i$. Otherwise, i could always strictly increase its payoff by switching its link with j to some other high producer that it does not connect with. If i is also connected with all low producers,

then according to Assumption 3, 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 $j' > n_h(\mathbf{s}^*)$ such that $\bar{g}_{ij'}^*$. Clearly, $g_{ij'j''}^* = 0$ for all $j'' > n_h(\mathbf{s}^*)$. 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 $j'' > n_h(\mathbf{s}^*)$ such that $g_{j''j'}^* = 1$. It can be concluded that $\bar{g}_{j''i}^* = 1$ 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 have that $x_{j'}^* \geq x_i^*$, which contradicts the fact that $x_{j'}^* < x_i^*$. Statement (i) follows. ∎

Lemma 4 provides important insights on how agents will interact with each other at equilibrium. Statement (i) and (ii) jointly characterize the redundancy on produced information at equilibrium. With a non-zero cost of link formation, the total amount of information produced by high producers will always be more than what each agent needs. Hence, each high producer neither has a full connection with all other high producers nor forms links with any low producer. Statement (iii) proves that a low producer always forms links with some high producers.

Given the agents' link formation behavior at equilibrium and the information redundancy, we can further determine, in the next lemma, that if a low producer forms links to other low producers in an equilibrium, it should have formed links to all high producers.

Lemma 5. In an equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$, $g_{jj'}^* = 1$, for some $j, j' > n_h(\mathbf{s}^*)$ only if $g_{ji}^* = 1$, $\forall i \leq n_h(\mathbf{s}^*)$. ∎

With Lemma 5, low producers can be further classified into two types in order to facilitate our analysis: (i) a low producer forming links to all high producers and to some of the low producers; (ii) a low producer forming links with some (or all) high producers and no low producer.

B. Asymptotic equilibria in the IPLF game

By analyzing the equilibrium properties of agents, we are then able to characterize the asymptotic equilibria when the network size grows. We prove in the following theorem that when the network size is sufficiently large, low producers of type (i) will disappear in any equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$, and there are only high producers and low producers of type (ii) left in the network. Given this, we can further characterize the equilibria topologies. Particularly, we show that each high producer produces an amount $\tilde{x}(\mathbf{x}^*)$ and each low producer (type (ii)) produces a smaller amount of information, denoted as $\underline{x}(\mathbf{x}^*)$. Meanwhile, each low producer forms links with $q(\mathbf{g}^*)$ high producers and does not form any link to low producers. Therefore, in a network at equilibrium, all links are formed towards high producers, which play the role of hubs for information sharing in the network. The network then exhibits a "core-periphery" structure with high producers form the core and low producers stay at the periphery.

Theorem 2 (Asymptotic Equilibria). Given c, k, ρ, and v, and when the network size n goes to infinity: (i) only two types of agents exist in any equilibrium $\mathbf{s}^* = (\mathbf{x}^*, \mathbf{g}^*)$: a hub agent is a high producer who produces an amount $\tilde{x}(\mathbf{x}^*)$ and form links only with other hub agents; a spoke agent is a low producer who produces an amount $\underline{x}(\mathbf{x}^*)$ and forms links with $q(\mathbf{g}^*)$ hub agents and no link to other spoke agent;

(ii) $\lim_{n\to\infty} \inf_{\mathbf{s}^* \in S_n^*} \{n_h(\mathbf{s}^*)\} = \infty$ and $\lim_{n\to\infty} \inf_{\mathbf{s}^* \in S_n^*} \{n_l(\mathbf{s}^*)\} = \infty$ where $n_l(\mathbf{s}^*) \triangleq n - n_h(\mathbf{s}^*)$ and S_n^* denotes the set of equilibrium strategy profiles when the network size is n. ■

A key observation in proving Theorem 2 is that because of the cost of link formation, an agent has to produce a sufficiently large amount of information, which is lower-bounded away from 0, in order to attract others to form links to it. Meanwhile, an agent has an upper bound on the amount of information that it would like to consume due to the concavity of the benefit function (1). As a result, there is an upper bound on the number of links that an agent would like to maintain with others. Therefore, when $n \to \infty$, the number of hub agents (i.e. high producers) goes to infinity as well to provide sufficient information for the consumption of all agents in the network, and low producers will become spoke agents by not mutually connecting with each other but only connecting with hub agents.

Theorem 2 proves that the core-periphery structure is a necessary condition for an equilibrium when n is sufficiently large, which is shown in Figure 1 with two examples of equilibrium topologies. Throughout this paper, we use $v(x) = ln(1 + x^\rho)^{1/\rho}$ as an exemplary utility function in the experiments. Both topologies exhibit the core-periphery structure composed of two rings. The inner ring (core) represents hub agents and the outer ring (periphery) represents spoke agents. Hence, our results are consistent with the empirical observations that a majority of users in social sharing services get most of their information from a relatively small group of hub agents [18]. By comparing Figure 1(a) and Figure 1(b), which portray equilibrium topologies with k changing from 2 to 5, we can see that the network becomes sparse as the link formation cost increases, when, instead of forming links with multiple hub agents, spoke agents get information only from hub agents in their vicinity and the network is virtually divided into many small sub-networks where each hub agent takes charge of the information provision to its local agents.

The core-periphery structure exhibits some similarity to the small-world networks [14], which is also predicted in [15] as the "law of the few". Nevertheless, different from the scale-free property where the fraction of hub agents diminishes to 0 as the network size grows to infinity, we show that under agents' appreciation for information variety, the population of hub nodes will grow proportionally to the network size, and their fraction in the total population is lower-bounded away from 0. Similar results have also been illustrated in [18], which reveals, through empirical measurement, that the probability of users who are followed by very large numbers of users on Twitter is above what a scale-free distribution would predict. This indicates that as more agents join the network, information production which is dominated by a small number of powerful producers can no longer satisfy the agents' desire
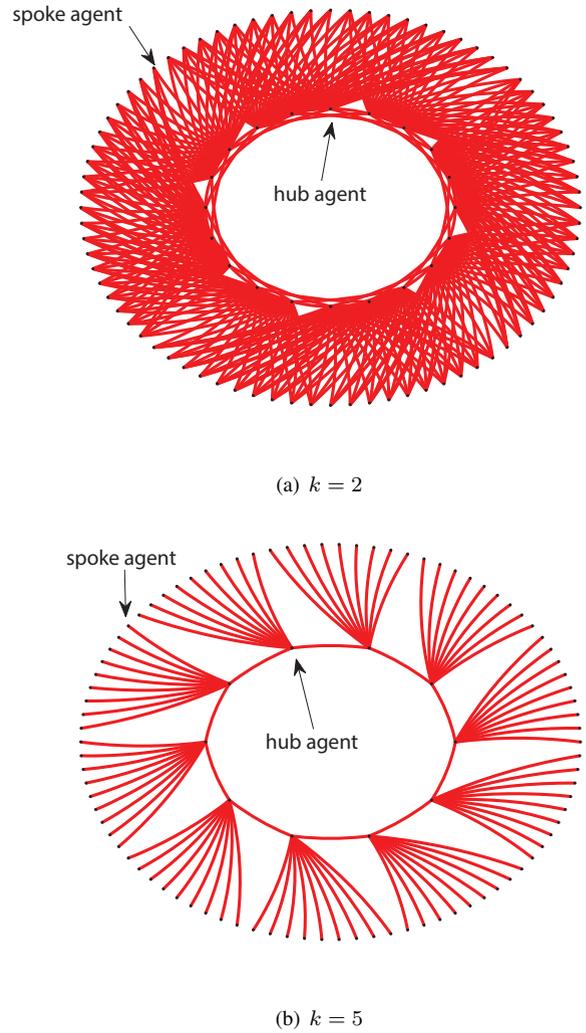

(a) $k = 2$

(b) $k = 5$

Fig. 1. Exemplary equilibria in a network with $n = 100$ agents ($c = 2$, $\rho = 0.8$)

for diverse information. As a result, new hub agents with different varieties of information will emerge to provide the information and stabilize the network.

Theorem 3. The number of hub agents at equilibrium grows at the same order as the entire population, i.e. $\inf_{\mathbf{s}^* \in S_n^*} \{n_h(\mathbf{s}^*)\}$ is $\Theta(n)$. More specifically, there exists a constant $\eta > 0$ such that

$$\lim_{n\to\infty} \inf_{\mathbf{s}^* \in S_n^*} \{n_h(\mathbf{s}^*)\}/n > \eta. \tag{5}$$

Proof: It has been proved in Theorem 2 that we can always find a sufficiently large value T such that spoke agents 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 $\inf_{\mathbf{s}^* \in S_n^*} \{n_h(\mathbf{s}^*)\}/n > \mu$, $\forall n < T$. Now look at the case when $n > T$. Due to the concavity of the benefit function $v(\cdot)$ and the fixed cost of link formation k, a spoke agent cannot connect to more than L_h hub agents, which upper-bounds the amount of information it receives from others at $L_h \bar{x}$. Hence, there exists a constant $\underline{x}$ such that the production level of a low producer $\underline{x}(\mathbf{x}^*)$ is no less than $\underline{x}$ in any equilibrium $\mathbf{s}^*$ when $n > T$. The total number of links that each high producer receives from low producers is
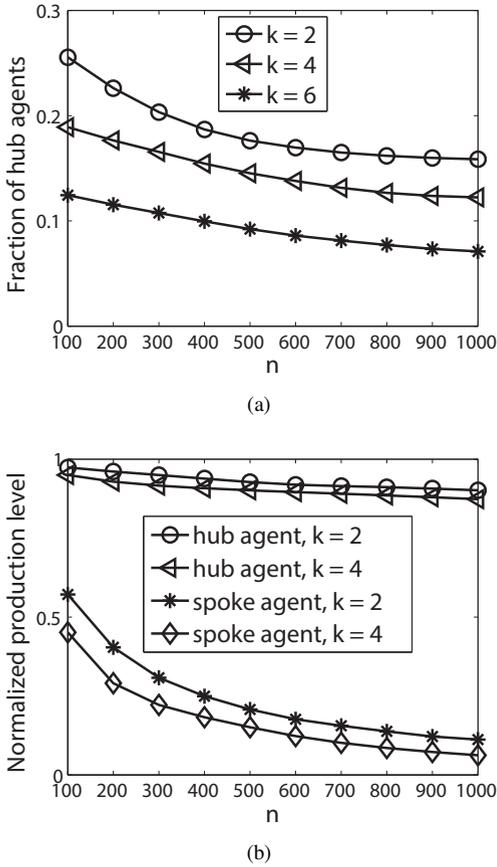
Fig. 2. (a) The fraction of hub agents changing against n; (b) The normalized production levels of agents changing against n ($c = 2$, $\rho = 0.8$)

thus also upper-bounded by a constant, denoted as H_l. Given the population of high producers, i.e. $n_h(\mathbf{s}^*)$, the population of low producers should be no more than $H_l n_h(\mathbf{s}^*)$, which further delivers a lower bound on $\inf_{\mathbf{s}^* \in S_n^*}\{n_h(\mathbf{s}^*)\}/n$ as $1/(1 + H_l)$, $\forall n < T$. We take $\eta = \min(\mu, \ 1/(1 + H_l))$ and hence Theorem 3 follows. ∎

Figure 2 plots how the fraction of high productions as well as $\tilde{x}(\mathbf{x}^*)$ and $\underline{x}(\mathbf{x}^*)$ at equilibrium change as n increases [6]. Figure 2(a) illustrates Theorem 3. The number of hub agents grows at a slower speed than n at the beginning and hence, $n_h(\mathbf{s}^*)/n$ monotonically decreases. However, when n is sufficiently large, the number of spoke agents that a hub agent can support reaches its upper bound. Hence, to support an equilibrium, more hub agents will emerge and $n_h(\mathbf{s}^*)/n$ stops decreasing, which indicates that $n_h(\mathbf{s}^*)$ starts to grow proportionally to n. Figure 2(b) shows how the production levels of agents change against n. For a better illustration, we plot the normalized production levels which are compared with their values at $n = 100$ on each curve. As it shows,

---

[6]It should be noted that for a given set of the network parameters k, c, ρ, and v, there are multiple equilibria. In the experiment, we run the experiment multiple rounds. At the beginning of each round, we assign a randomly chosen initial strategy profile $\mathbf{s}^{(0)}$ to all agents in the network. Then in each step t, agents sequentially update their individual strategies using best response dynamics with some inertia probability γ and the resulting strategy profile being $\mathbf{s}^{(t)}$ (i.e. an agent adapts its strategy to its best response with probability $1 - \gamma$ and remains to use its strategy adopted at step $t - 1$ with probability γ). Each round terminates when the strategy profile converges to some fixed point $\mathbf{s}$. The result plot in Figure 2 is averaged over all rounds of experiments.

the normalized production level of spoke agents drops more drastically than that of hub agents as n grows, since a spoke agent mainly relies on information acquired from others rather than self-production and will be more significantly influenced by the change on the network size.

As a direct result of Theorem 3, we quantify in the following corollary the amount of information generated in the network.

Corollary 1. The total amount of information produced in the network at equilibrium, i.e. $\sum_{i \in N} x_i^*$, grows at the same order as the population size n. That is, $\inf_{\mathbf{s}^* \in S_n^*}\{\sum_{i \in N} x_i^*\}$ and $\sup_{\mathbf{s}^* \in S_n^*}\{\sum_{i \in N} x_i^*\}$ are $\Theta(n)$. ∎

Therefore, as the network grows, it will become informationally richer with more agent-generated information available.

C. Equilibrium efficiency

In this section, we analyze the social welfare of the IPLF game to quantitatively study the efficiency of equilibria. The social welfare of the IPLF game is defined to be the sum of agents' individual utilities. For a strategy profile $\mathbf{s} = (\mathbf{x}, \mathbf{g})$, the social welfare is given by $W(\mathbf{x}, \mathbf{g}) \triangleq \sum_{i \in N} u_i(\mathbf{x}, \mathbf{g})$. A strategy profile $\mathbf{s}^{\#} = (\mathbf{x}^{\#}, \mathbf{g}^{\#})$ achieves the social optimum, denoted as $W^{\#}$, if and only if

$$W^{\#} \triangleq W(\mathbf{x}^{\#}, \mathbf{g}^{\#}) \geq W(\mathbf{x}, \mathbf{g}), \ \forall(\mathbf{x}, \mathbf{g}). \quad (6)$$

In the following proposition, an upper bound on $W^{\#}$ is given.

Proposition 1. There is a value $\bar{k}$ such that the social optimum $W^{\#} \in \left(n[v(\bar{x}) - c\bar{x}], \ n[v(n^{1/\rho}\hat{x}_n) - c\hat{x}_n] - nk/2 \right)$ when $k < \bar{k}$ and $W^{\#} = n[v(\bar{x}) - c\bar{x}]$ when $k > \bar{k}$. Here $\bar{x}$ is defined in Lemma 3(iii) and $\hat{x}_n$ is the solution of the equation $v'(n^{1/\rho}\hat{x}_n) = c/n^{(1-\rho)/\rho}$. ∎

When the social optimum is achieved in an empty network, there is no cost for link formation and thus, we have $W^{\#} = n[v(\bar{x}) - c\bar{x}]$. On the other hand, if the social optimum is achieved in a non-empty network, it is proven that each agent has at least one connection and hence, the minimum total link cost in the network is $nk/2$ while the maximum total utility from information consumption and production is $n[v(n^{1/\rho}\hat{x}_n) - c\hat{x}_n]$, which gives the corresponding upper bound of the social optimum.

In order to prove the existence of the threshold $\bar{k}$, 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 $k < k_1$. 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 $k > k_2$.

Next, we characterize the efficiency of the equilibrium strategy profile. The following proposition provides a lower bound on the social welfare that can be achieved by an equilibrium.

Proposition 2. The social welfare of any equilibrium strategy $\mathbf{s}^*$ is lower bounded by $n[v(\bar{x}) - c\bar{x}]$.

Proof: Since each agent can ensure a utility of $v(\bar{x}) - c\bar{x}$ by choosing not to form any link and produce an amount of $\bar{x}$, the utility it receives in an equilibrium is always no less than this. ∎

By combining Proposition 1 and 2, the following theorem provides an upper bound for the Price of Stability (PoS) in the
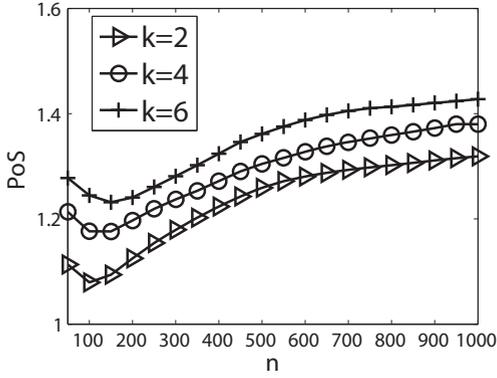
(a)
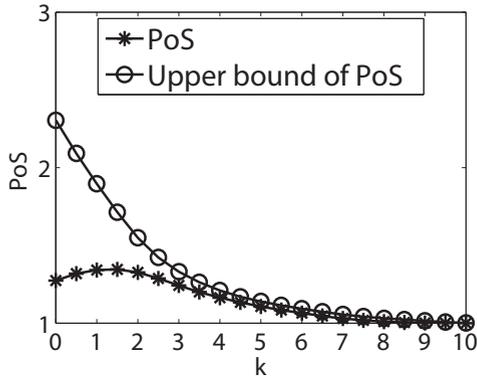
(b)

Fig. 3. (a) The PoS against the network size n; (b) The PoS and its lower bound against the link formation cost k ($c = 2$, $\rho = 0.8$, $n = 500$)

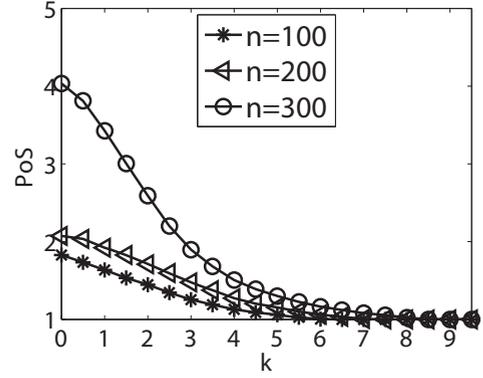Fig. 4. The Price of Stability with indirect information sharing ($c = 1$, $\rho = 0.7$)

IPLF game. It should be noted that since the lower bound in Proposition 2 applies to all equilibria in the IPLF game, this result also applies to the Price of Anarchy (PoA).

Theorem 4 (Upper Bound for PoS). The Price of Stability satisfies that $PoS \in (1, ([v(n^{1/\rho}\hat{x}_n) - c\hat{x}_n] - k/2)/[v(\bar{x}) - c\bar{x}])$ when $k < \bar{k}$ and $PoS = 1$ when $k > \bar{k}$. ∎

Figure 3 plots the PoS against n and k. As Figure 3(a) shows, the PoS monotonically increases with n when it is sufficiently large. This is due to the fact that the set of available strategy profiles in the IPLF game is enlarged along with n, whereas the set of equilibrium strategy profiles remains limited as each hub agent in an equilibrium can only support a limited number of spoke agents. However, the rate for such increase slows down as n becomes larger. In other words, the PoS saturates when n is sufficiently large. Figure 3(b) compares the PoS and our derived upper bound in Theorem 4. The upper bound is not tight when k is small and it is optimal to form as many links as possible in order to increase the information sharing efficiency. When k increases, the upper bound becomes tight, as the topologies produced by both the socially optimal profile and the equilibrium profile become sparse with their social welfare approaching $n[v(\bar{x}) - c\bar{x}]$.

IV. INDIRECT INFORMATION SHARING

In the basic model, an agent can only consume information produced by itself and its neighbors (i.e. agents it connects via links). In many practical scenarios, the information sharing

in social computing applications could be multi-hop. That is, an agent can also consume information which its neighbors acquired from other agents. In this section, we consider such indirect information sharing.

Several concepts are defined below before we formalize the IPLF game with indirect information sharing. Given a connectivity graph $\bar{g}$, we say that there is a path in g between two agents i and j if either $\bar{g}_{ij} = 1$ or there exist agents $j_1, \ldots, j_m$ distinct from i and j such that $\{\bar{g}_{ij_1} = \bar{g}_{j_1 j_2} = \cdots = \bar{g}_{j_m j} = 1\}$.

Definition 1 (Connected Network). A network is connected if there is a path between every pair of agents in its connectivity graph. A connected network is also called as a component.

Definition 2 (Minimally Connected Network). A network is minimally connected if there is a unique path between every pair of agents on its connectivity graph.

Given any two agents i and j, the distance between them, denoted as $d_{ij}(\bar{g})$ is defined as the number of hops (links) on the shortest path (the path contains the smallest number of links) between i and j in $\bar{g}$. If there is no path between i and j, then $d_{ij}(\bar{g}) = \infty$.

In the IPLF game with indirect information sharing, we assume that each agent can acquire the effective information from its neighbors, which already included the information that the neighbors acquire from other agents. Therefore, it is equivalent that each agent can acquire information produced by all agents who are connected with him via a path. However, if an agent receives certain information produced by another agent from multiple paths, it will automatically remove the redundancy and only keep one copy of the information in consumption. Let $N_i^l(\bar{g})$ denote the set of agents whose distance to agent i is l, the utility of agent i then can be formulated as follows:

$$w_i(\mathbf{x}, \mathbf{g}) = v((x_i^\rho + \sum_{l=1}^{n-1} \sum_{j \in N_i^l(\bar{g})} x_j^\rho)^{1/\rho}) - cx_i - k|N_i(\mathbf{g})|. \quad (7)$$

In this section, we will use all the terminologies (e.g. equilibrium, social optimum, etc.) that we defined for the IPLF game with direct information sharing in Section III wherever they apply. With indirect information sharing, it can be proved

that the IPLF game always has at least one equilibrium. The resulting equilibria are characterized in the following theorem.

Theorem 5. In the presence of indirect information sharing and when an agent's utility is given by (7), there are two values k_{max} and k_{min} for any given n, c, ρ, and v, such that

(i) When $k < k_{max}$, there exists a unique equilibrium where each agent has a production level $\bar{x}$ and no agent forms links;

(ii) When $k < k_{min}$, then each equilibrium preserves the following properties: (a) each agent produces an amount x_n which is the solution of $v(n^{1/\rho}x_n) = c$; (b) the network is minimally connected; (c) each agent has at least one connection to other agents;

(iii) When $k_{min} < k < k_{max}$, there are multiple equilibria each of which contains a minimally connected component with the rest of agents being isolated.

Proof Sketch: First, it is easy to show that for a component of size b in an equilibrium strategy profile $\mathbf{s}^*$, it is always minimally connected with each agent in it having the same production level. Then, we show that at any equilibrium, an agent's utility from information consumption and production monotonically increases with the size of component it is in. With this result, we prove that at any equilibrium, there is at most one component whose size is larger than 1. Suppose in an equilibrium strategy profile $\mathbf{s}^*$ 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 $\mathbf{s}^*$ 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 $\mathbf{s}^*$. It can be shown that agent j has the incentive to connect with the component C if and only if the link formation cost k is smaller than a value γ_b which is determined by the size b, which is proved to monotonically increases with b.

Summarizing all the above, it can be concluded that if $k < \gamma_1$, there is a unique equilibrium which contains a unique component and there is a path between any two agents; if $\gamma_1 < k < \gamma_{n-1}$, there are two equilibria where all agents are either isolated with each other or fully connected in a unique component; if $k > \gamma_{n-1}$, the network has a unique equilibrium where all agents are isolated with each other. Hence, we have $k_{max} = \gamma_{n-1}$ and $k_{min} = \gamma_1$. ∎

Corollary 2. Given c, ρ, and $v(\cdot)$, k_{max} monotonically increases with n, while k_{min} is constant. ∎

It should be noted that different from the core-periphery structure in the case of direct information sharing, where high producers are also the main source of information sharing and plays the role of hubs in the network, here the agents with the most connections form the hubs of the network and take the responsibility to share information. Next, we analyze the social optimum of the IPLF game with indirect information sharing. The proof is omitted due to its similarity to Theorem 5.

Theorem 6. In the presence of indirect information sharing and when an agent's utility is given by (7),

(i) There is a value $k_{max}^{opt} > k_{max}$ for any given n, c, ρ, and v, such that the social optimum is achieved in an empty network when $k > k_{max}^{opt}$;

(ii) There is a value $k_{min}^{opt} > k_{min}$ for any given n, c, ρ, and v, such that the social optimum is achieved in a minimally connected network when $k < k_{min}^{opt}$;

(iii) When $k_{min}^{opt} < k < k_{max}^{opt}$, the social optimum is achieved in a network which contains a minimally connected component with the rest of agents being isolated. ∎

Hence, the Price of Stability in the IPLF game with indirect information sharing, which is plotted in Figure 4, achieves 1 when $k > k_{max}^{opt}$ and is strictly larger than 1 when $k < k_{max}^{opt}$. With indirect information sharing, the topologies of the network are similar under both the socially optimal strategy profile and an equilibrium profile. Hence, the difference on the social welfare is highly influenced by how much information the agents produce in total. When k is small, the discrepancy between the total information productions of the socially optimal strategy profile and an equilibrium profile is higher than that when k is large, which gives a PoS monotonically decreasing against k.

V. CONCLUSION

In this paper, we investigate the problem of information production and network formation in social computing systems. Different from the existing literature, the agents' incentives for producing information themselves and for forming links to consume the information of others are jointly considered. Moreover, we determine rigorously how the agents' appreciation for information variety impacts their interactions and the emerging connectivity between them. We then analyze the efficiency of the emerging equilibrium topologies and compare their performances with that of the social optimum. We also study the IPLF game which incorporates indirect information sharing. Our analysis can be extended in several directions, among which we mention three. First, we assume in this model that agents are homogeneous in terms of their benefit functions and costs. The analysis on heterogeneous agents would represent an important future research direction which can hopefully provide new insights into the IPLF game, e.g. agents' heterogeneity could lead to the selection of equilibria and reduce the set of possible equilibrium topology as predicted in [12]. Second, alternative formulations on information transmission and link formation can be extended from our current model to encompass the features of various applications. Several examples include the bilateral link formation which requires the mutual consent of agents, the unilateral information transmission with which the information flow over a link is one-sided, and non-constant link formation cost with which the cost of establishing a link depends on the characteristics of the creator and the recipient as well as the amount of information (i.e. traffic) across this link. Finally, the design of effective incentive protocols (e.g. pricing schemes to subsidize or tax link formation) to stimulate the information production and sharing in order to reduce the Price of Stability and achieve the social optimum also forms an important future research direction.

REFERENCES

[1] M. Cha, H. Kwak, P. Rodriguz, Y. Ahn and S. Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," *Proc. 7th ACM SIGCOMM Conf. on Internet Measurement*, pp. 1-14, 2007.

[2] A. Kittur, E. Chi and B. Suh, "Crowdsourcing user studies with mechanical turk," *Proc. 26th Annual SIGCHI Conf. on Human Factors in Computing Systems*, 2008.

[3] J. Chakareski and P. Frossard, "Context-adaptive information flow allocation and media delivery in online social networks," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 4, 2010.

[4] M. Goemans, L. Li and M. Thottan, "Market sharing games applied tocontent distribution in ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1020-1033, 2006.

[5] S. Ioannidis, A. Chaintreau and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," *Proc. IEEE INFOCOM 2009*, pp. 1422-1430, 2009.

[6] M. Gandetto and C. Regazzoni, "Spectrum sensing: a distributed approach for cognitive terminals," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 546-557, 2007.

[7] W. Saad, Z. Han, A. Hjorungnes, D. Niyato and E. Hossain, "Coalition formation games for distributed cooperation among roadside units in vehicular networks," *J. Sel. Areas Commun.*, vol. 29, no. 1, pp. 48-60, 2010.

[8] T. Shu and M. Krunz, "Coordinated channel access in cognitive radio networks: a multi-level spectrum opportunity perspective," *INFOCOM 2009*, pp. 2976-2980, 2009.

[9] S. Goyal, "Connections: an introduction to the economics of networks," *Princeton University Press*, New Jersey, 2007.

[10] V. Bala and S. Goyal, "A non-cooperative model of network formation," *Econometrica*, vol. 68, no. 5, pp. 1181-1231, 2000.

[11] J. Corbo and D. Parkes, "The price of selfish behavior in bilateral network formation," *Proc. 24th Annual Symposium on Principles of distributed computing (PODC)*, 2005.

[12] A. Galeotti and S. Goyal, "Network formation with heterogeneous players," *Games and Economic Behavior*, vol. 54, no. 2, pp. 353-372, 2006.

[13] J. Musachhio, G. Schwartz and J. Walrand, "A two-sided market analysis of provider investment incentives with an application to the net-neutrality issue," *Review of Network Economics*, vol. 8, no. 1, pp. 3-25, 2009.

[14] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440-442, 1998.

[15] A. Galeotti and S. Goyal, "The Law of the Few," *The American Economic Review*, vol. 100, no. 4, pp. 1468-1492, 2010.

[16] A. Dixit and J. Stiglitz, "Monopolistic competition and optimum product diversity," *The American Economic Review*, vol. 67, no. 3, pp. 297-308, 1977.

[17] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509-512, 1999.

[18] H. Kwak, C. Lee, H. Park and S. Moon, "What is Twitter, a social network or a news media?" *Proc. 19th Int'l Conf. on World Wide Web (WWW)*, 2010.

[19] A. Fabrikant, A. Luthra, E. Maneva, C. Papadimitriou and S. Shenker, "On a network creation game," *Proc. 22nd Annual Symposium on Principles of distributed computing (PODC)*, 2003.

[20] J. Nash, "Non-cooperative games," *The Annals of Mathematics*, vol. 54, no. 2, pp. 286-295, 1951.

[21] Y. Zhang and M. van der Schaar, "Online appendix: Information production and link formation in social computing systems," *online appendix*, http://medianetlab.ee.ucla.edu/papers/link_formation_jsac_appendix.pdf.

[22] Y. Zhang, J. Park and M. van der Schaar, "Content production and link formation games," *UCLA Technical Report*, http://arxiv.org/abs/1109.4201,2011.

Yu Zhang (S'08) received the bachelors and masters degrees from Tsinghua University, Beijing, China, in 2006 and 2008, respectively. He is currently working towards the Ph.D degree with the Department of Electrical Engineering, University of California, Los Angeles.

Mihaela van der Schaar (SM'04-F'10) received the Ph.D degree from Eindhoven University of Technology, The Netherlands, in 2001. She is now a Professor with the Electrical Engineering Department, UCLA. Her research interests include engineering economics and game theory, strategic design, online reputation and social media, dynamic multi-user networks and system designs.