Self-organizing Networks of Information Gathering Cognitive Agents

Ahmed M. Alaa, Member, IEEE, Kartik Ahuja, and Mihaela van der Schaar, Fellow, IEEE

Abstract

In many scenarios, networks emerge endogenously as cognitive agents establish links in order to exchange information. Network formation has been widely studied in economics, but only on the basis of simplistic models that assume that the value of each additional piece of information is constant. In this paper we present a first model and associated analysis for network formation under the much more realistic assumption that the value of each additional piece of information depends on the type of that piece of information and on the information already possessed: information may be complementary or redundant. We model the formation of a network as a non-cooperative game in which the actions are the formation of links and the benefit of forming a link is the value of the information exchanged minus the cost of forming the link. We characterize the topologies of the networks emerging at a Nash equilibrium (NE) of this game and compare the efficiency of equilibrium networks with the efficiency of centrally designed networks. To quantify the impact of information redundancy and linking cost on social information loss, we provide estimates for the Price of Anarchy (PoA); to quantify the impact on individual information loss we introduce and provide estimates for a measure we call Maximum Information Loss (MIL). Finally, we consider the setting in which agents are not endowed with information, but must produce it. We show that the validity of the well-known “law of the few” depends on how information aggregates; in particular, the “law of the few” fails when information displays complementarities.

Index Terms

Cognitive networking, cognitive agents, information networks, network formation, self-organizing networks.

The authors are with the Department of Electrical Engineering, University of California Los Angeles (UCLA), Los Angeles, CA, 90095, USA (e-mail: ahmedmalaa@ucla.edu, ahujak@ucla.edu, mihaela@ee.ucla.edu). This work was funded by the Office of Naval Research (ONR).
Self-organizing Networks of Information
Gathering Cognitive Agents

I. INTRODUCTION

Many emerging networks are formed endogenously, by the actions of self-organizing, cognitive agents. Examples of such networks are: dynamic spectrum management by wireless users [1], social networks overlaid on technological networks [2] [3], device-to-device communication, vehicular networks [5], Internet-of-Things (IoT) [6], and smart sensor networks [7]. In many of these networks, users connect to each other in order to gather information about the environment. For instance, secondary users exchange information about spectrum occupancy in cognitive radio networks [8] [9], autonomous rescue robots exchange environmental sensory information [6] [10], etc.

Strategic network formation was first studied in the economics literature. Some of this literature [11]-[14] asks which networks are stable (according to some criteria) and hence more likely to persist and be observed. A (smaller) literature asks which networks emerge as the result of some specific dynamic process [15] [16]. In all these works, simplistic models of information are assumed: the value of each additional piece of information is constant [11] [12]. However, in cognitive communications, information possessed by different agents can be redundant or complementary. For instance, secondary users in a multi-band cognitive radio system may be interested in gathering information about spectrum occupancy for bands that they do not sense by communicating with other users who do sense these bands [8]; sensors deployed over a correlated random field [17]-[21] may be interested in gathering complementary measurements about some set of physical processes of interest, etc.

This paper studies cognitive information networks (CIN), in which agents self-organize to gather/exchange and produce information. We define a CIN as any network that forms endogenously by the actions of agents who aim at gathering information about a state of the world. This state of the world can be spectrum occupancy information and primary user activity in a multi-band cognitive radio system, location information provided by anchors of wireless networks, a set of messages sent by information sources in a multicast network, or blogs, videos, and data exchanged by users of social-physical networks. Agents are cognitive since they perceive information possessed by other agents, reason about which links to establish, how much information to produce, and then take information production and link formation decisions which result in an endogenously-formed network topology.

In this paper, we present a first model and analysis for self-organizing networks of cognitive information gathering agents. We assume that agents in a CIN benefit only from gathering non-redundant information about the state of the world. We assume that agents possess different amounts of information, and they form links with each other in order to gather information and maximize their knowledge of the state of the world. Information possessed by different agents may be correlated (redundant), and link formation is costly. Hence, agents should cognitively select which agents to link with. We formulate this problem as a non-cooperative network formation game. Using
information-theoretic measures for the value of the information possessed by each agent, we aim at characterizing the emerging stable network topologies at Nash Equilibrium (NE). Throughout our analysis, we focus on two classes of linking cost scenarios: homogeneous link formation cost and heterogeneous link formation costs. In the former, connecting to any agent entails the same cost, while in the later, the link cost is recipient-dependent. We show that the networks that emerge at equilibrium are minimally connected; thus, agents tend to minimize the overall cost of constructing the network. With homogeneous link costs, equilibrium leads to a network in which each component is a star. Moreover, we show how information redundancy affects the link cost ranges at which the network becomes connected or disconnected: when the link costs are homogeneous, redundancies in agents’ information decreases the Price-of-Anarchy (PoA); in contrast, information redundancy can induce costly anarchy in connected networks with heterogeneous link costs.

Finally, we consider a setting in which each agent will not only decide which links to form, but also the amount of information to produce and we provide a characterization for the emerging NE. We show that when the number of agents is large, the fraction of agents producing information at equilibrium depends on the amount of redundancy in the agents’ information. When the agents produce strongly correlated information, the fraction of information producers is small and tends to zero as the number of agents tends to infinity: most agents get the information they need from a small set of agents. On the other hand, when agents have uncorrelated information, the number of information producers can grow at the same rate of total number of agents. Thus, such networks violate what Galeotti and Goyal [22] call the “law of the few”. In addition, we quantify the total amount of information produced in an asymptotically large network and identify scenarios in which the amount of information produced at equilibrium grows with the number of agents.

The paper aims at studying and understanding the self-organization of cognitive agents, what network topologies emerge and how efficient they are. Thus, we adopt an abstract model for CINs without delving into the idiosyncratic details of specific applications. However, our results provide insights into practical CINs such as distributed channel access in interweave cognitive radio systems, self-organizing sensor networks, cooperative localization [23], and joint source and network coding [24] [25].

The rest of the paper is organized as follows. In Section II, we formalize the network formation game among agents in a CIN. Section III characterizes the emerging stable networks when the link formation costs are homogeneous, and the efficiency of such networks are investigated. Section IV analyzes the network topology and equilibrium efficiency for the case of heterogeneous link costs. The joint information production and link formation game is studied in Section V. Finally, conclusions are drawn in Section VI.

II. Basic Model

In this section, we discuss the problem setting and propose a basic model to formulate the endogenous formation game emerging among cognitive agents.
A. Information model

Let $\mathcal{N} = \{1, 2, 3, ..., N\}$ be the set of agents in the CIN. Each agent $i$ possesses exogenous information in the form of a discrete random variable $X_i$ and aims to form links with other agents to maximize its utility, which is defined as the benefit from the total information it possesses minus the linking cost. The formation of links is costly; thus, an agent has to trade off the benefits of the information it obtains from another agent versus the cost it needs to pay for connecting with that agent. The amount of information in $X_i$ is quantified by the entropy function $H(X_i)$. In addition, the random variables of all agents may be correlated, which indicates that some agents may possess similar information that is redundant to that of the other agents. The common information between agent $i$ and $j$ is captured by the mutual information $I(X_i; X_j)$.

The information possessed by the set of agents $\mathcal{N}$ is captured by an entropic vector $\vec{H} \in \mathcal{H}$, where $\mathcal{H}$ is a $2^N - 1$ dimensional Euclidean space, which corresponds to the positive orthant of the real coordinate space $\mathbb{R}_+^{2^N - 1}$. A vector $\vec{H}$ is said to be entropic if there exists a random variable tuple $(X_1, X_2, ..., X_N)$, where associated with any subset $\mathcal{V}$ of $\mathcal{N}$, there is a joint entropy $H(X_\mathcal{V})$ that is an element of $\vec{H}$, where $X_\mathcal{V} = \{X_i | i \in \mathcal{V}\}$ [26]. The elements of $\vec{H}$ represent the joint entropies between all possible subsets of random variables. The vector $\vec{H} \in \Gamma^*_N$, where $\Gamma^*_N \subset \mathcal{H}$ is the set of all entropic vectors of order $N$, and is known as the entropic region [26]. We denote by $\vec{\mathcal{H}}$ the set of entropic vector having $H(X_1, X_2, ..., X_N) = \sum_{i=1}^N H(X_i)$, where $\vec{\mathcal{H}} \subset \Gamma^*_N$. The set of entropic vectors in $\vec{\mathcal{H}}$ correspond to the case where the information have no redundancies, thus the model reduce to the aggregation models in [11] [12] [14].

The entropic vector can be constructed as follows. Given the set of agents $\mathcal{N}$ and a corresponding set of random variables $\mathcal{X} = \{X_1, X_2, ..., X_N\}$, we construct the set $\mathcal{V} = \mathcal{P}(\mathcal{X}) \setminus \{\phi\}$, where $\mathcal{P}(\mathcal{X})$ is the power set of $\mathcal{X}$. If $\mathcal{V} = \{v_1, v_2, ..., v_{|\mathcal{V}|}\}$, then the entropic vector is given by $\vec{H} = (H(X_{v_i}))_{i=1}^{|\mathcal{V}|}$, where $|\mathcal{V}| = 2^N - 1$, and $H(X_{v_i})$ is the joint entropy between all random variables in the set $v_i$. For instance, if we have 3 agents in the network, then $\mathcal{V} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}$, and the entropic vector $\vec{H}$ is given by $(H(X_1), H(X_2), H(X_3), H(X_{1,2}), H(X_{1,3}), H(X_{2,3}), H(X_{1,2,3}))^T$, where $H(X_{1,2}) = H(X_1, X_2)$. We denote a single element in the entropic vector as $\vec{H}(v) = H(X_v)$. The mutual information between the random variables possessed by any two subsets $\mathcal{V}$ and $\mathcal{U}$ of agents is given by [27]

$$I(X_{\mathcal{V}}; X_{\mathcal{U}}) = H(X_{\mathcal{V}}) + H(X_{\mathcal{U}}) - H(X_{\mathcal{V}}, X_{\mathcal{U}}).$$

The total amount of information in the network is given by the joint entropy of the random variables of individual agents $H(\mathcal{X}) = H(X_1, X_2, X_3, ..., X_N)$, where $H(\mathcal{X}) \in \vec{H}$.

The mutual information between any two agents $i$ and $j$ is given by $I(X_i; X_j) = H(X_i) - H(X_i|X_j)$, where $H(X_i|X_j)$ is the conditional entropy which represents the additional information attained by agent $j$ from connecting to $i$, i.e. the amount of extra information that $j$ gets when getting the information of $i$. If this benefit is low, it means that $I(X_i; X_j)$ is high, i.e. $X_i$ and $X_j$ are highly correlated, and vice versa. Note that mutual information is symmetric, i.e. $I(X_i; X_j) = H(X_i) - H(X_i|X_j) = H(X_j) - H(X_j|X_i)$. The total amount of redundant information is given by $\sum_{i=1}^N H(X_i) - H(X_1, X_2, ..., X_N)$, which corresponds to the Kullback Leibler (KL)
divergence between the joint probability mass function (pmf) \( p(X_1, X_2, ..., X_N) \) and \( \Pi_{i=1}^{N} p(X_i) \), where \( p(X_i) \) is the marginal distribution of the random variable of agent \( i \). Let \( p = p(X_1, X_2, ..., X_N) \) and \( q = \Pi_{i=1}^{N} p(X_i) \), the KL divergence for these distributions can be computed as follows [27]

\[
D(p||q) = \sum_{i=1}^{N} H(X_i) - H(X_1, X_2, ..., X_N).
\]

The KL divergence corresponds to the distance between an entropic vector and the corresponding entropic vector in the set \( \mathcal{H} \) where all random variables are independent. Throughout the paper, we use \( H(X_i) \) to denote \( H(X / \{X_i\}) \).

B. Network formation game

Agents benefit from gathering information by linking to other agents. The link formation strategy adopted by agent \( i \) is denoted by a tuple \( g_i = (g_{ij})_{j \in \{1, ..., N\} \setminus \{i\}} \in \{0, 1\}^{N-1} \); \( g_{ij} = 1 \) if agent \( i \) forms a link with agent \( j \) and \( g_{ij} = 0 \) otherwise. We assume unilateral link formation where an agent decides to form a link and solely bears the cost of link formation\(^1\). A strategy profile \( g \) is defined as the collection of strategies of all agents, i.e. \( g := (g_i)_{i=1}^{N} \in \mathcal{G} \), where \( \mathcal{G} \) is a finite space. When agent \( i \) forms a link with agent \( j \), it incurs a cost of \( c_{ij} \). We define the topology of the network as \( \mathcal{T} = \{ (i, j) \in \mathcal{N} \times \mathcal{N} | \max\{g_{ij}, g_{ji}\} = 1 \} \). All connected agents exchange information bilaterally; thus \( \mathcal{T} \) is an undirected graph. Information is shared between agents that are indirectly connected and agents do not benefit from receiving multiple versions of the same information from the same agent. Such model is suitable for networks with multi-hop relaying where information is forwarded from one node to another [28]. We write \( i \rightarrow j \) to indicate that agent \( j \) is reachable by agent \( i \) either directly or indirectly. Define the set of agents that \( i \) form links with (set of neighbors) as \( \mathcal{N}_i(g) = \{ j | g_{ij} = 1 \} \), and the set of agents reachable by agent \( i \) as \( \mathcal{R}_i(g) = \{ j | i \rightarrow j \} \). Throughout the paper, we adopt the following definitions. A component \( C \) is a set of agents such that \( i \rightarrow j, \forall i, j \in C \), and \( i \not\rightarrow j, \forall i \in C \) and \( j \not\in C \), i.e. two agents in two different components cannot share information. A component is minimally connected if each agent \( i \in C \) is connected to each agent \( j \in C \) via a unique path. The utility function of agent \( i \) is given by

\[
u_i(g) = f(\sum_{j \in \mathcal{N}_i(g)} c_{ij}) - \sum_{j \in \mathcal{R}_i(g)} c_{ij},
\]

where the function \( f(\cdot) \) represents the benefit of agent \( i \) from the information it gathers. We assume that the agents benefit from acquiring information increases, while the marginal benefit decreases, with the increase of the amount of information gathered. That is, in a sensor network setting, the benefit of a sensor node from collecting information saturates if it is connected to a large number of sensors; thus, \( f(\cdot) \) is twice continuously differentiable, increasing, and concave with \( f(0) = 0 \). Note that the total information acquired by \( i \) in (2) can be written in terms of the conditional entropies based on the chain rule as [27]

\[
H(X_i \cup \mathcal{R}_i(g)) = H(X_i) + \sum_{k=1}^{\mathcal{R}_i(g)} H(X_j | X_{j_1}, X_{j_2}, ..., X_{j_k}),
\]

\(^1\)Other link formation models, such as link formation with bilateral consent, can be used with an appropriate solution concept such as pairwise stability.
where \( R_i(g) = \{ j_1, j_2, \ldots, j_{|R_i(g)|} \} \), which implies that agents benefit by acquiring new information conditioned on its own information and the information it acquires from other connections. Moreover, the aggregate information can be expressed in terms of the mutual information as

\[
H(X_{i \cup R_i(g)}) = H(X_i) + H(X_{R_i(g)}) - I(X_i; X_{R_i(g)}),
\]

where the term \( H(X_{R_i(g)}) \) represents the net information that agent \( i \) acquires after connecting to the agents in \( N_i(g) \), where the term \( I(X_i; X_{R_i(g)}) \) captures the redundancy between the information of agent \( i \) and the information it acquires from the set \( R_i(g) \). Let \( u = (u_1, u_2, \ldots, u_N) \). Throughout the paper, we denote the network formation game by \( G^N(N, G, u, \vec{H}) \). We assume a complete information scenario, where all agents have knowledge of the entropic vector \( \vec{H} \), the strategy space \( G \) and the utilities of all agents \( u \).

C. Stability concept and network efficiency

The link formation game is formulated as a non-cooperative simultaneous move game and we focus on the Nash Equilibrium (NE) as the solution concept. The NE is defined as follows

\[
u_i(g_i^*, g_{-i}^*) \geq u_i(g_i, g_{-i}^*), \forall g_i \in \{0, 1\}^{N-1}, \forall i \in N,
\]

where \( g_{-i}^* \) is the strategies of all users other than \( i \). A strict NE is obtained by making the inequality in (3) strict. The game can have multiple NE defined as \( G^* = \{ g^*_i | \forall u_i(g_i^*, g_{-i}^*) \geq u_i(g_i, g_{-i}^*), \forall g_i \in \{0, 1\}^{N-1} \} \). In the following Theorem, we show that there exists at least one network in the NE, i.e. \( G^* \neq \phi \).

\textbf{Theorem 1: (The Existence of Nash Equilibrium)} A pure strategy NE always exists for \( G^N(N, G, u, \vec{H}) \).

\textbf{Proof} See Appendix A.

The social welfare of the network formation game is defined as the sum of agents’ individual utilities. For a strategy profile \( g \), the social welfare is defined as

\[
U(g) := \sum_{i \in N} u_i(g).
\]

A strategy profile \( \tilde{g} \) is called socially optimal if it maximizes the social welfare (achieves the social optimum \( \tilde{U} \)), i.e.

\[
\tilde{U} := U(\tilde{g}) \geq U(g), \forall g \in G.
\]

When there are multiple equilibria, we use two metrics to assess the equilibrium efficiency. First, we adopt the Price of Anarchy (PoA) to quantify the impact of the agents’ selfish behavior on the social welfare. The PoA is defined as the ratio between the social optimum and the lowest social welfare achieved at equilibrium, i.e.

\[
\text{PoA} = \frac{\tilde{U}}{\min_{g^* \in G^*} U(g^*)}.
\]

In addition, we analyze the impact of the agents’ selfish behavior on the information gathering process by defining a novel metric that we term the Maximum Information Loss (MIL). The MIL is defined as the maximum difference
\[ MIL = \max_i \left( \sup_{g_i \in G^*} H(X_i \cup X_{R_i(g_i^*)}) - \inf_{g_i \in G^*} H(X_i \cup X_{R_i(g_i^*)}) \right). \] (7)

between the amount of information gathered by any agent at two different equilibria as shown in (7). Unlike the PoA, the MIL quantifies the maximum information loss without considering the link cost. In addition, while the PoA considers the welfare of all agents, the MIL quantifies the highest information loss incurred by an agent in the worst case.

### III. Nash Equilibrium Analysis for Homogeneous Link Costs

In this section, we assume that the cost of forming a link between any two agents \( i \) and \( j \) is given by \( c_{ij} = c, \forall i, j \in N \). The goal of this section is to answer the following question: given an entropic vector \( \tilde{H} \), what are the network topologies \( T \) that can emerge at an NE of the game \( G^N \) when the link costs are homogeneous? We start with the following motivating example to identify different factors that affect the equilibria of \( G^2 \).

#### A. Motivating example for two-agents interaction: does information redundancy matter?

Consider a simple network with only two agents (\( N = 2 \)) possessing random variables \( X_1 \) and \( X_2 \). We aim at characterizing the equilibria of \( G^2 = \langle \{1, 2\}, G, u, \tilde{H} \rangle \). The strategy of agent 1 is simply a linking decision \( g_{12} \in \{0, 1\} \), while for agent 2, the strategy is \( g_{21} \in \{0, 1\} \). We write \( G^2 \) in normal form in Table I, where the row player is agent 2 and the column player is agent 1. Each cell displays the utilities of agents 1 and 2 respectively. Assume that the link cost is the same for both agents and equal to \( c \). It can be easily shown that the payoffs of agent 1 are given by \( u_1(g_{12} = 1, g_{21} = 1) = u_1(g_{12} = 1, g_{21} = 0) = f(H(X_1, X_2)) - c \), \( u_1(g_{12} = 0, g_{21} = 1) = f(H(X_1, X_2)) \), and \( u_1(g_{12} = 0, g_{21} = 0) = f(H(X_1)) \).

**TABLE I: Two agent network formation game in normal form**

| \( g_{12} = 1 \) | \( g_{12} = 0 \) |
|------------------|------------------|
| \( g_{21} = 1 \) | \( u_1(g_{12} = 1, g_{21} = 1) \), \( u_2(g_{12} = 1, g_{21} = 1) \) | \( u_1(g_{12} = 0, g_{21} = 1) \), \( u_2(g_{12} = 0, g_{21} = 1) \) |
| \( g_{21} = 0 \) | \( u_1(g_{12} = 1, g_{21} = 0) \), \( u_2(g_{12} = 1, g_{21} = 0) \) | \( u_1(g_{12} = 0, g_{21} = 0) \), \( u_2(g_{12} = 0, g_{21} = 0) \) |

Fig. 1 depicts the entropic region \( \Gamma_2^* \) of the two random variables \( X_1 \) and \( X_2 \). Note that for \( N = 2 \), we have \( \Gamma_2^* = \Gamma_2 \), thus the entropic region can be easily constructed by applying the three Shannon inequalities \( H(X_1) \leq H(X_1, X_2) \), \( H(X_2) \leq H(X_1, X_2) \), and \( H(X_1) + H(X_2) \geq H(X_1, X_2) \). The intersection of these three hyperplanes in \( \mathcal{H} \) results in the polyhedral cone depicted in Fig. 1. For a given entropic vector, the KL divergence \( D(p(X_1, X_2)||p(X_1)p(X_2)) \) represents the distance between this vector and the hyperplane \( \tilde{H} \), which represents
the set of entropic vectors for which $X_1$ and $X_2$ are independent, which correspond to the aggregation models in [11] [12]. If $D(p(X_1, X_2)||p(X_1)p(X_2)) = 0$, then the entropic vector lies on $\tilde{H}$.

![Fig. 1: The entropic region $\Gamma_2^*$ for 2 random variables.](image)

The equilibria of this game depend on both the link cost and the entropic vector, which corresponds to the amount of information redundancy. For an arbitrary entropic vector, the game has two possible equilibria $g^* = (g_{12} = 1, g_{21} = 0)$ and $g^* = (g_{12} = 0, g_{21} = 1)$ if $c \leq f(H(X_1, X_2)) - f(\max\{H(X_1, H(X_2))\})$. Assume that $H(X_1) > H(X_2)$. Therefore, the network has a unique equilibrium $g^* = (g_{12} = 0, g_{21} = 1)$ when $f(H(X_1, X_2)) - f(H(X_1)) \leq c \leq f(H(X_1, X_2)) - f(H(X_2))$, and a unique equilibrium $g^* = (g_{12} = 0, g_{21} = 0)$ when $c \geq f(H(X_1, X_2)) - f(H(X_2))$. On the other hand, if we fix the link cost and the entropies $H(X_1)$ and $H(X_2)$, we observe that the equilibria change by changing the KL divergence. For instance, the network has two equilibria $g^* = (g_{12} = 1, g_{21} = 0)$ and $g^* = (g_{12} = 0, g_{21} = 1)$ when $c \leq f(H(X_1) + H(X_2) - D(p||q)) - f(\max\{H(X_1), H(X_2)\})$. Thus, as the entropic vector becomes closer to the hyperplane $\tilde{H}$, i.e. $D(p||q)$ decreases, the cost threshold for which these two equilibria emerge increases. Thus, the characterization of the NE is sensitive to the amount of information redundancy $D(p||q)$, even if we fix the individual entropies $H(X_1)$ and $H(X_2)$. Note that the strategy profile $g = (g_{12} = 1, g_{21} = 1)$ never emerges as an NE since any of the two agents can break the link it forms and get a strictly higher utility. In the next subsection, we aim at obtaining a generic characterization for the NE of $G^N$.

**B. Characterization of the NE for $G^N$**

We start by providing a key feature of information networks in the following proposition.

**Proposition 1:** (Network minimality) All network components are minimally connected in NE.

**Proof** See Appendix B.

Proposition 1 implies that agents in each component will form the minimal number of links possible to get all the information from each other. This results from indirect information sharing within each network component, i.e. if there exists a path to an agent then there is no extra benefit in making a direct link to that agent since all the information from that agent is already accessible.
Next, we characterize the connectivity of the network as a function of the link cost in the following Lemma.

**Lemma 1: (Network connectivity regions)**

(i) If \( c \leq c_l \), with \( c_l = f(H(X)) - f(\min_i H(X_i)) \), then, at every NE (a) the network is minimally connected (the network has one component) and (b) the amount of information possessed by each agent is \( H(X) \) (all information is shared).

(ii) If \( c \geq c_u \), where \( c_u = f(H(X)) - f(\min_i H(X_i)) \), then there is a unique NE which is strict. At this equilibrium, the network is fully disconnected and the amount of information possessed by each agent \( i \) is \( H(X_i) \) (no information is shared).

**Proof** See Appendix C. ■

From the above Lemma, we can see that two factors affect the connectivity of a network: the amount of information possessed by each agent and the redundancies among the agents’ information. Based on the result of Lemma 1, we define three regions for the connectivity of the NE networks based on the link cost as follows:

- **Connected agents region \( (K_C) \):** The network has a single component when the link cost is \( c \leq c_l \).
- **Isolated agents region \( (K_I) \):** The network has \( N \) components when the link cost is \( c \geq c_u \).
- **Mixed region \( (K_M) \):** Depending on the entropic vector, the network can have different number of components ranging from 1 to \( N \) when the link cost is \( c_l \leq c \leq c_u \).

While the connectivity regions describe the impact of link cost on network topology, they also have informational significance. For instance, the amount of information possessed by any agent in the \( K_C \) region is \( H(X) \), while in the \( K_I \) region, no agent \( i \) gathers any extra information other than its own intrinsic information \( H(X_i) \). On the other hand, agents in the \( K_M \) region can end up gathering different amounts of information as there are potentially multiple equilibria with different topologies and connectedness. Thus, the \( K_M \) region plays an important role in determining the network efficiency as we will show in the next subsection. It is interesting to study the impact of the information redundancy on the width of the \( K_M \) region versus link cost, i.e. the range of costs over which \( K_M \) exists. Note that if this width is small, the network topology becomes very sensitive to the link cost in the sense that small increase in the cost can move the network from the connected phase to the fully disconnected phase, which is a drastic topological and informational phase transition.

Several key questions arise in this context: How sensitive is the network to small changes in the link cost? How sharp can a transition from one topological phase to another be? What is the impact of the agents information on these transitions? In the following, we develop a qualitative framework for analyzing the impact of information on the connectivity regions of the NE networks and its sensitivity to link cost changes. In this paper, we say that a phase transition happens if \( \mathbb{P}(A | c \leq c_l) = 1 \), and \( \mathbb{P}(A | c \geq c_u) = 0 \), where \( A \) is some property of the network. In this section, we say that a phase transition occurs if \( \mathbb{P}(|g^*| = N - 1 | c \leq c_l) = 1 \), and \( \mathbb{P}(|g^*| = N - 1 | c \geq c_u) = 0 \), where \( |g^*| \) is the total number of links in the network in an NE. Thus, \( A \) is the property that the network is connected. Thus, a phase transition occurs in a CIN as the cost varies from \( c_l \) to \( c_u \). Such a phase transition is not only topological, but also informational as it corresponds to the amount of information gathered by agents at each
phase. Next, we quantify the sharpness of this transition in terms of the width of the $K_M$ region in the following definition.

**Definition 1:** Range of the $K_M$ region $\Psi$ - is the cost range over which the $K_M$ region exists, i.e. $\Psi = c_u - c_l$.

For a CIN, the function $\Psi$ corresponds to the cost range for which the network can have multiple equilibria with different topologies. If $\Psi = 0$, the $K_M$ region is empty. We can define $\Psi$ mathematically as

$$\Psi = f(\min_i H(X_i)) - f(\min_i H(X_i)),$$

(8)

An important question is how does information redundancy affect the transition between phases of connectivity? In the following Theorem, we show that $\Psi$ decreases with the increase of information redundancy. We denote by $D_{-i}(p||q)$ the KL-divergence of the information possessed by all agents except $i$. Therefore, we have $D_{-i}(p||q) = \sum_{j=1, j\neq i}^{N} H(X_j) - H(X_{-i})$, which corresponds to the total information redundancy in the information possessed by a set of $N-1$ agents. Note that $D_{-i}(p||q)$ is an information redundancy metric that quantifies the total amount of information redundancy for a subset of $N-1$ agents. We need this metric since the sharpness of phase transitions depends on the information possessed by $N-1$ agents only.

**Theorem 2:** (Information redundancy induces sharpness in phase transition) The range of the $K_M$ region $\Psi$ is a monotonically decreasing function of $D_{-i}(p||q)$, $i = \arg\min_k H(X_{-k})$.

**Proof** The Theorem says that if $D_{-i,1}(p||q) > D_{-i,2}(p||q)$, then $\Psi(D_{-i,1}(p||q)) \leq \Psi(D_{-i,2}(p||q))$. From (8), we know that $\Psi$ can be written as $\Psi = f\left(\sum_{j=1, j\neq i}^{N} H(X_j) - D_{-i}(p||q)\right) - f(H(X_i))$, where $i = \arg\min_k H(X_k)$. Since all individual entropies are fixed, then $\Psi$ is only dependent on $D_{-i}(p||q)$. Note that $f(x)$ is monotonically increasing function of $x$, then $f\left(\sum_{j=1, j\neq i}^{N} H(X_j) - D_{-i}(p||q)\right)$ is monotonically decreasing function of $D_{-i}(p||q)$. Therefore, $\Psi$ decreases with increasing $D_{-i}(p||q)$.

Fig. 2: Connectivity regions on the link cost-KL divergence plane for three different entropic vectors
Theorem 2 says that increasing the correlation in the random variables possessed by the set of \( N - 1 \) agents having the minimum joint entropy decreases the cost range separating the \( \mathcal{K}_C \) and \( \mathcal{K}_I \) regions. Thus, the sharpness of phase transition increases as the network can drastically move from the connected phase to the empty phase for a small change of the link cost. But, is it possible that the \( \mathcal{K}_M \) region becomes empty? Next, we give the necessary and sufficient condition for which the \( \mathcal{K}_M \) region becomes empty.

*Proposition 2:* The \( \mathcal{K}_M \) region is empty if and only if there exists a single agent whose entropy is strictly higher than all other agents, and the remaining \( N - 1 \) agents possess identical information.

The proposition follows directly from (8). This proposition implies that when all agents rely only on a single agent to get information, then the benefit from connecting to that agent solely determines whether the network should be connected or not. This can be the case in an interweave cognitive radio system when one secondary user carries out spectrum sensing and others connect to it, or in a sensor network where only one sensor is close to the sensed physical process and acts as a single source node. For such a system, the \( \mathcal{K}_M \) region will always be empty; whether the system belongs to the fully connected or fully disconnected region depends only on the linking cost. On the other hand, if agents are more heterogeneous, i.e. more than one agent are highly informative, the \( \mathcal{K}_M \) region is always not empty. To illustrate these results, we consider the following example.

*Illustrative example 1:* To illustrate the impact of information redundancy on the phase transition, consider the following example. Assume that we have a 3-agent CIN, with \( H(X_1) > H(X_2) \), and \( H(X_2) = H(X_3) \). Assume that agent 1 has information that is uncorrelated to that of agents 2 and 3. Thus, we have \( D(p||q) = I(X_2;X_3) \).

In Fig. 2, we plot the connectivity regions in the cost-KL divergence plane for three values of agent 1 entropy \( H^3(X_1) > H^2(X_1) > H^1(X_1) \), where the corresponding connectivity regions for setting \( H(X_1) = H^i(X_1) \) are denoted by \( \{ \mathcal{K}_C^i, \mathcal{K}_I^i, \mathcal{K}_M^i \} \). An exemplary utility function of \( f(x) = \log(1+x) \) is used. It is clear that, in agreement with Theorem 2, the width of the \( \mathcal{K}_M \) region decreases as the information redundancy increases. Note that in the \( \mathcal{K}_M \) region, we might have both a connected and a fully disconnected network as two possible equilibria. When agents 2 and 3 information are fully correlated, we have a sharp threshold of \( c = f(H^i(X_1)) \), below which we have a connected network, and above which we have a fully disconnected network, and the \( \mathcal{K}_M \) region is empty. The intuition behind this is that since agents 2 and 3 are fully correlated, they only benefit from connecting to agent 1. Thus, it is the benefit from getting agent’s 1 information that solely determines the cost at which the network would be connected or not. If agents 2 and 3 information are not correlated, they add value to the network, and the cost thresholds becomes dependent on their information as well.

While Lemma 1 focuses on the impact of link cost on the connectivity of the network, it does not provide a complete characterization for an NE network. In the next Theorem, we give the necessary and sufficient conditions for the emergence of an arbitrary CIN topology in NE.

*Theorem 3:* A network in which the components are precisely \( \{ \mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_K \} \) can be supported in a NE if and only if the following relationships between the cost and the value of information are satisfied.
Fig. 3: Impact of information heterogeneity on the phase transitions.

1) \( f(H(X_{C_i})) - \min\{f(H(X_{C_i}\backslash\{j\})), f(H(X_j))\} \geq c, \forall i \in \{1, 2, ..., K\}, j \in C_i. \)
2) \( f(H(X_{C_i\cup C_j})) - f(H(X_{C_i})) \leq c, \forall i, j \in \{1, 2, ..., K\}. \)

**Proof** See Appendix D.

From Theorem 3 we know that, at NE, the network is generally composed of multiple components and each component is minimally connected. Each component possess a set of random variables that are jointly highly correlated to the joint random variables possessed by other components. Condition (1) in Theorem 3 implies that different components have no incentives to connect to other components, while condition (2) shows that each agent in a component benefits from a link to some other agent in that component. Note that due to indirect information sharing, many equilibria can exists with highly variant topologies. In the subsequent Theorem, we refine the equilibrium notion used, and we determine the topologies emerging in a strict NE.

**Theorem 4:** A network is a strict NE if and only if the following conditions are simultaneously satisfied

- All conditions stated in Theorem 3 are satisfied.
- For each component \( C \) of size \( M > 1 \), there exists a set \( \zeta \subseteq C \) with \( |\zeta| \geq M - 1 \) such that \( \zeta = \{j \mid f(H(X_C)) - f(H(X_{C\backslash\{j\}})) > c\} \).

- Each non-singleton component forms a core-sponsored star topology.

**Proof** See Appendix E.

This Theorem states that for homogeneous link formation costs, each network component of size \( M \) comprises a single agent bearing the cost of getting connected to \( M - 1 \) other agents. Such networks exhibit a core-periphery structure, i.e. a single agent at the core is connected to a set of \( M - 1 \) periphery agents. The conditions in Theorem 4 state that periphery agents must be high entropy agents. This is because the benefit obtained by connecting to a periphery agent \( j \) at equilibrium must exceed the cost, i.e. \( f(H(X_C)) - f(H(X_{C\backslash\{j\}})) > c \). The intuition behind this condition is as follows. For an agent to be a periphery agent, it must have both high entropy and low redundancy.
with the information possessed by other component members such that core agents have an incentive to form a
link with it. Fig. 4 depicts an exemplary topology of a CIN at strict NE for various link formation cost ranges.
In the next subsection, we study the efficiency of the NE networks and compare the self-organized CINs to those
designed by a network planner.

C. Equilibrium efficiency analysis

In this subsection, we investigate the equilibrium efficiency of \( G^N \) with homogeneous link costs by quantifying
the PoA and MIL. We aim at studying the impact of information redundancy on both metrics. The goal of this
section is to compare the networks emerging by the actions of agents and those designed by a central planner. In
the following Lemma, we quantify the PoA of CINs with homogeneous link costs.

\textit{Lemma 2:} For a CIN with homogeneous link costs, the Price-of-Anarchy satisfies

\[
\text{PoA} = \frac{N f(H(X)) - (N - 1)c}{\sum_{i=1}^{N} f(H(X_i))}, \forall \left( \overline{H}, c \right) \in K_M.
\]

\textbf{Proof} We know that in the \( K_C \) region, all the NE networks are connected. Thus, the social welfare of any network
in \( K_C \) is given by \( U(g^*) = N f(H(X)) - (N - 1)c \). The socially optimal network in \( K_C \) is the one with a social
welfare of \( \tilde{U} \), where \( \tilde{U} = \max_{g \in G} U(g) \). Since \( f(H(X)) - f(H(X_i)) > c, \forall i \) in the \( K_C \) region, then it is clear
that a connected network maximizes the social welfare. Therefore, \( \tilde{U} = U(g^*) \) and the PoA = 1 in the \( K_C \) region.
Next, we focus on the \( K_I \) region. In this region, any connection will result a negative payoff for any agent who
forms a link since \( c > f(H(X)) - f(\min_i H(X_i)) \). Thus, the social optimal is a fully disconnected network, which
is also the unique (strict) NE, and the PoA = 1 in the \( K_I \) region. For the \( K_M \) region, the maximum PoA will
occur if a fully disconnected network is an equilibrium and a connected network is a social optimum. In what follows, we show that this is indeed possible. Consider the case when \( f(H(X_i, X_j)) - f(H(X_i)) < c, \forall i, j \), and \( f(H(\mathcal{X})) - f(H(X_i)) > c, \forall i \). In this case, agents do not get immediate benefit from forming links to individual agents, thus a fully disconnected network is an NE since not forming a link is a best response for all agents when all other agents do not form a link. Therefore, the PoA in the \( \mathcal{K}_M \) region is upper bounded by the social welfare of a connected network and that of a fully disconnected network, i.e. \( \text{PoA} \leq \frac{N(f(H(\mathcal{X})) - (N-1)c)}{\sum_{i=1}^N f(H(X_i))} \). \[ \square \]

This Lemma shows that all NE networks in the \( \mathcal{K}_C \) and \( \mathcal{K}_I \) regions are socially optimal. NE networks in the \( \mathcal{K}_M \) region may not be socially optimal but an upper bound on the PoA exists. This bound is a function of both the entropic vector and link cost. In the following Theorem, we investigate the impact of information redundancy on the PoA.

**Theorem 5:** For a CIN with homogeneous link costs in the \( \mathcal{K}_M \) region, the upper bound on the Price-of-Anarchy is a monotonically decreasing function of the amount of total information redundancy \( D(p|q) \).

**Proof** The PoA can be written as \( \text{PoA} \leq \frac{N(f(\sum_{i=1}^N H(X_i) - D(p|q)) - (N-1)c)}{\sum_{i=1}^N f(H(X_i))} \). Note that the benefit function \( f(x) \) is monotonically increasing in \( x \). Thus, as \( D(p|q) \) increases, \( f(\sum_{i=1}^N H(X_i) - D(p|q)) \) decreases, and the PoA decreases consequently. Therefore, we have \( \frac{\partial \text{PoA}}{\partial D(p|q)} < 0 \).

This Theorem says that the more redundancy in the information possessed by the agents, the lesser is the price of agents’ selfish behavior. While the social welfare captures the sum utilities, it does not quantify the individual losses by agents. In the next corollary, we quantify the MIL for different connectivity regions.

**Corollary 1:** For a CIN with homogeneous link cost, the MIL satisfies

\[
\text{MIL} = 0, \forall (\widehat{H}, c) \in \mathcal{K}_C \cup \mathcal{K}_I,
\]

and

\[
\text{MIL} \leq H(\mathcal{X}) - \min_i H(X_i), \forall (\widehat{H}, c) \in \mathcal{K}_M.
\]

**Proof** In the \( \mathcal{K}_C \) region, we know that all NE networks are connected. Thus, \( \sup_{g^* \in \mathcal{G}} H(X_i)_{i \in R_i(g^*)} = \inf_{g^* \in \mathcal{G}} H(X_i)_{i \in R_i(g^*)} = H(\mathcal{X}) \), and \( \text{MIL} = 0 \). Similarly, in the \( \mathcal{K}_I \) region, we have \( \sup_{g^* \in \mathcal{G}} H(X_i)_{i \in R_i(g^*)} = \inf_{g^* \in \mathcal{G}} H(X_i)_{i \in R_i(g^*)} = \min_i H(X_i) \), thus \( \text{MIL} = 0 \). In the \( \mathcal{K}_M \) region, and using similar arguments to that used in proof of Theorem 5, the MIL is maximized if both a connected and a fully disconnected network are equilibria. In this case, \( \sup_{g^* \in \mathcal{G}} H(X_i)_{i \in R_i(g^*)} = H(\mathcal{X}) \), and \( \inf_{g^* \in \mathcal{G}} H(X_i)_{i \in R_i(g^*)} = \min_i H(X_i) \). Thus, \( \text{MIL} \leq H(\mathcal{X}) - \min_i H(X_i) \).

Fig. ?? depicts the PoA for a 3-agent CIN with an entropic vector satisfying \( H(X_1, X_2, X_3) = H(X_1) + H(X_2, X_3) \). It is clear that the PoA is greater than 1 only in the \( \mathcal{K}_M \) region. In addition, the PoA decreases as the KL divergence increases, since the value of information in the network decreases, which means that the best equilibrium (connected network) achieves a smaller social welfare while the welfare of the worst equilibrium (fully disconnected network) is independent of the KL divergence. The PoA also decreases as the link cost increases. It
is clear from Fig. ?? that when $D(p|q) = 4$, the network exhibit a phase transition with an empty $K_M$ region, i.e. the network changes from a connected to a fully disconnected network if the cost exceeds a certain threshold. Thus, for $D(p|q) = 4$ the network is robust to efficiency loss for all values of link cost as the $K_M$ region becomes empty. Thus, while information redundancy reduces the total amount of information of the network, it increases the robustness of the network to loss of social welfare and guides the self-organized network to social optimality. Fig. ?? depicts the MIL upper bound for the same network. It is clear that unlike the PoA, the MIL upper bound does not depend on the link cost, but decreases monotonically with the increasing information redundancy. In the next section, we extend our analysis to the case when link costs are heterogeneous, and focus on the impact of heterogeneous link costs on the network efficiency.

IV. NASH EQUILIBRIUM ANALYSIS FOR HETEROGENEOUS LINK COSTS

In this section, we extend the analysis done in the previous section for the game $G^N$, but assuming that the cost of link formation is exclusively recipient-dependent, i.e. $c_{ij} = c_j, \forall i$. We focus on studying the impact of link cost heterogeneity on the topologies, connectivity, and efficiency of the CINs in NE. It is easy to show that Proposition 1 applies to the case of heterogeneous link costs, thus we know that all network components in NE are minimally connected.

A. Characterization of the NE for $G^N$

The following proposition relates the link costs to the connectivity of the NE networks.

Proposition 3:
(i) If $c_i < f(H(X)) - f(H(X-i)), \forall i$, then, at every NE (a) the network is minimally connected (the network has one component) and (b) the amount of information possessed by each agent is $H(X)$ (all information is shared).
(ii) If $f(H(X)) - f(\min_j H(X-j)) < \min_{k \in N \setminus \{i\}} c_k$, where $i = \arg \min_j H(X-j)$, then there is a unique NE which is strict. At this equilibrium, the network is fully disconnected and the amount of information possessed by each agent $i$ is $H(X_i)$ (no information is shared).

Proof This can be proven straightforwardly using the same arguments in the proof of Lemma 1. ■

This proposition shows that the network topology is highly dependent on the heterogeneity of the agents as it depends both on the heterogeneous costs and heterogeneous information of agents. Also the case when all NE networks are connected corresponds to the $K_C$ region in the homogeneous cost scenario, while the case when the NE is a fully disconnected network corresponds to the $K_I$ region. An appropriate definition for the connectivity regions for the heterogeneous cost case is given by (9), (10), and (11).

In the following Theorem, we give a generic characterization for the networks in NE.

Theorem 6: A network in which the components are precisely $\{C_1, C_2, ..., C_K\}$ can be supported in a NE if and only if the following relationships between the cost and the value of information are satisfied
Proof This can be proven following the same idea for the proof of Theorem 3.

Note that unlike the homogeneous cost scenario, we cannot characterize the connectivity versus a single value for link cost. Thus, the concepts of phase transition do not apply in this scenario. In the next subsection, we analyze the efficiency of the NE networks.

B. Equilibrium Efficiency Analysis

In this subsection, we quantify the impact of the costs heterogeneity on the network efficiency. Unlike the case of the homogeneous link costs, we show that information redundancy induces costly anarchy when the link costs are recipient-dependent. In the following Lemma, we quantify the PoA for the $\mathcal{K}_C$ and $\mathcal{K}_I$ regions.

**Lemma 3:** For a CIN with heterogeneous link costs, the PoA satisfies

\[
\text{PoA} = \begin{cases} 1, & \forall \left( \vec{H}, c = (c_1, c_2, ..., c_N) \right) \in \mathcal{K}_C \,, \forall \left( \vec{H}, c \right) \in \mathcal{K}_I \, : \frac{N f(H(\mathcal{X})) - (N-1) \min_k c_k}{N f(H(\mathcal{X})) - \sum_{i=1}^N c_j + \min_k c_k} \,, \forall \left( \vec{H}, c \right) \in \mathcal{K}_C \end{cases}
\]

and

\[
\text{PoA} \leq \frac{N f(H(\mathcal{X})) - (N-1) \min_k c_k}{\sum_{i=1}^N H(X_i)} \, : \forall \left( \vec{H}, c \right) \in \mathcal{K}_M.
\]

**Proof** See Appendix F.

Thus, unlike in the homogeneous cost scenario, not all NE networks in the $\mathcal{K}_C$ region are socially optimal. In fact, any NE network other than a periphery-sponsored star with the agent having the lowest link cost residing in the core, is not socially optimal. Hence, how does the information redundancy impact the PoA in this case? The following Theorem answers this question.

**Theorem 7:** For a CIN with recipient-dependent link costs in the $\mathcal{K}_C$ region, the Price-of-Anarchy increases with the increase of information redundancy $D(p|q)$.
Proof In the $K_C$ region, the PoA can be written as $\text{PoA} = \frac{N f(\sum_{i=1}^{N} H(X_i) - D(p||q)) - (N-1) \min_k c_k}{N f(\sum_{i=1}^{N} H(X_i) - D(p||q)) - \sum_{j=1}^{N} c_j + \min_k c_k}$. Let $\bar{c} = \sum_{j=1}^{N} c_j - \min_k c_k$. Taking the first derivative with respect to $D(p||q)$, we get

$$\frac{\partial \text{PoA}}{\partial D(p||q)} = \frac{f'(\sum_{i=1}^{N} H(X_i) - D(p||q))(N\bar{c} - (N - 1) \min_k c_k)}{\left(N f(\sum_{i=1}^{N} H(X_i) - D(p||q)) - \sum_{j=1}^{N} c_j + \min_k c_k\right)^2}.$$ 

It is clear that $N\bar{c} - (N - 1) \min_k c_k$ is strictly positive, and by definition of the benefit function, we know that $f'(x) > 0$. Therefore, $\frac{\partial \text{PoA}}{\partial D(p||q)} > 0$. $\blacksquare$

Thus, in stark contrast with the results obtained for the homogeneous cost CINs, Theorem 7 says that information redundancy induces costly anarchy for a network in $K_C$ region. This results from the heterogeneity of the link formation costs, which promotes the anarchy in the network as agents are no longer indifferent to the links they form as in the homogeneous cost scenario. As a matter of fact, some agents may end up forming expensive links and getting the same amount of information that they could have gathered by forming a cheaper link. When information redundancy increases, the value of the information gathered by agents decreases, thus, anarchy costs more and the PoA increases. Contrarily, in the $K_M$ region, the upper bound on PoA decreases as the information redundancy increases in a similar manner to the homogenous link costs scenario. Unlike the PoA, the MIL is not sensitive to cost heterogeneity since it only cares about the informational losses. It can be easily shown that the MIL in recipient-dependent CINs behaves in the same way as in the homogeneous cost scenario. In the next section, we tackle the problem of joint information production and link formation in CINs.

V. JOINT INFORMATION PRODUCTION AND LINK FORMATION GAMES IN CINs

In the network formation game so far, we have assumed that agents in a CIN are gifted with an exogenously determined entropic vector. Nevertheless, in many practical CINs, agents decide the amount of information to “produce” given some production cost. In this section, we focus on a CIN where each agent jointly decides the amount of information to produce and the links to form.

A. Game formulation

When agents choose what information to produce a crucial aspect is how information aggregates. [22] assumes that information aggregates simply by addition; this will be the case only if the value of each additional piece of information is constant; thus, there are no complementarities nor redundancies. [13] assumes a specific functional form, the Dixit-Stiglitz function; this captures informational complementarities and redundancies in a very special way. In this paper, we consider two modes of aggregation that seem more natural and are suggested by the formulation of information in terms of entropy.

The information production decision taken by $N$ agents in a CIN corresponds to the selection of a point inside the entropic region $\Gamma_N$. Correlations between the random variables of different agents are exogenously
determined by external factors, e.g. geographical locations of sensors. To capture information redundancy, we define an aggregation function $F_{\mathcal{H}} : \mathbb{R}_+^N \rightarrow \mathbb{R}$, that maps the entropies of a set of agents to a joint entropy of these agents, i.e. \( H(X_1, X_2, \ldots, X_N) = F_{\mathcal{H}}(H(X_1), H(X_2), \ldots, H(X_N)) \). It is clear that the range of the function $F_{\mathcal{H}}(.)$ should belong to $\Gamma^*_N$. Throughout this section, we study two different aggregation functions: the first is the one corresponding to independent random variables $H(X_1, X_2, \ldots, X_N) = \sum_{i=1}^N H(X_i)$, and the second is the one corresponding to strongly correlated random variables $H(X_1, X_2, \ldots, X_N) = \max\{H(X_1), H(X_2), \ldots, H(X_N)\}$. Both aggregation functions provide insights on how information redundancy affects the information production decisions at equilibrium.

In real-world networks, the aggregation function captures the informational relationships between different agents in a CIN. For instance, in a sensor network where sensors are deployed over a correlated random field [18], the information production decision can be thought of as the precision at which a sensor quantizes its measurements. Larger precision corresponds to larger value for the entropy. However, no matter what precision a sensor uses, its measurements will be correlated to that of another nearby sensor. Thus, the joint entropy of the two sensors would be governed not only by the precision they decide, but also by the redundancy in their information that is determined exogenously by their geographical locations and the nature of the physical process that they sense. The aggregation function captures such exogenous factors, and based on it, the behavior of cognitive agents is determined.

In the information production and link formation game, the strategy of an agent $i$ is denoted by $s_i = (H(X_i), g_i)$. A strategy profile of the game is written as $s = (H(X_1), H(X_2), \ldots, H(X_N), g)$, and the strategy space is $\mathcal{S}$. We denote the joint information production and link formation game by $\hat{\mathcal{G}}^N = \langle \mathcal{N}, \mathcal{S}, \mathcal{U} \rangle$. Thus, different from $\mathcal{G}^N$, agents do not observe an entropic vector, but they decide the entropic vector based on knowledge of the aggregation function.

The utility function of agent $i$ is given by

\[
u_i(s) = f(H(X_i)) - kH(X_i) - |\mathcal{N}_i|c,
\]

where $k$ is the cost of producing one unit of information, $|\mathcal{N}_i|$ is the number of agents which agent $i$ form links with, and $H(X_{i \cup \mathcal{R}_i(g)})$ is determined by $F_{\mathcal{H}}$ given the production levels of all agents. We adopt the NE as a solution concept. Thus, a strategy profile $s^*$ is an NE profile if no agent benefits from unilaterally forming a link, breaking a link, or altering the amount of information it produces. The set of NE profiles is denoted by $\mathcal{S}^*$. Finally, we denote by $\bar{H}$ the maximum amount of information that each agent can produce at equilibrium, thus $\bar{H}$ can be obtained by solving $f(\bar{H}) = k$ [22]. In the following subsection, we revisit the motivating example of the two agents interaction in order to understand the cognitive behavior of agents in $\hat{\mathcal{G}}^2$.

B. Motivating example for two-agents interaction: To produce or not to produce?

Consider a simple CIN with only two agents ($N = 2$) who are playing the game $\hat{\mathcal{G}}^2$. We aim at characterizing the equilibria of $\hat{\mathcal{G}}^2 = \langle \{1, 2\}, \mathcal{S}, \mathcal{U} \rangle$, and investigate the impact of $F_{\mathcal{H}}$, $k$, and $c$ on the cognitive behavior of the agents. Specifically, we are interested in identifying scenarios in which one agent may decide not to produce any
information and fully rely on the other. Let us focus on agent 1. The utility function of this agent is given by

$$u_1(s) = f(H(X_1 \cup \mathcal{R}_1(g))) - kH(X_1) - g_{12}c,$$

where $\mathcal{R}_1(g) = \emptyset$ if $g_{12} = g_{21} = 0$, and $\mathcal{R}_1(g) = 2$ otherwise. The best response of agent 1 is given by

$$u_1(s^*) = \max_{g_{12}, H(X_1)} \left( f(H(X_1 \cup \mathcal{R}_1(g))) - kH(X_1) - g_{12}c \right).$$

Note that the decision of agent 1 depends on the value of $H(X_1 \cup \mathcal{R}_1(g))$, which is determined by $F_H$. For 2 agents, the entropic vector is $\mathbf{H} = [H(X_1), H(X_2), H(X_1, X_2)]$. The function $F_H$ maps the information production decisions $H(X_1)$ and $H(X_2)$ to $H(X_1, X_2)$. Thus, we have $H(X_1, X_2) = F_H(H(X_1), H(X_2))$. We focus on two different aggregation functions $F_H(H(X_1), H(X_2)) = H(X_1) + H(X_2)$ and $F_H(H(X_1), H(X_2)) = \max\{H(X_1), H(X_2)\}$.

- For $F_H(H(X_1), H(X_2)) = H(X_1) + H(X_2)$:

  In this case, the information of agents 1 and 2 are not redundant, which means that the random variables $X_1$ and $X_2$ are independent. Thus, $F_H$ maps the production profile of both agents to a point in the set $\tilde{\mathcal{H}}$. This
reduces to the aggregation function used in [22]. Fig. 5 plots \( F_{\mathcal{H}} \) in the entropic space \( \mathcal{H} \), which corresponds to the upper surface of the convex cone \( \Gamma_2 \). Assume that the link cost is given by \( c > k\bar{H} \). In this case, we have a unique equilibrium in which \( g_{12}^* = g_{21}^* = 0 \), and \( H^*(X_1) = H^*(X_2) = \bar{H} \). Thus, we have a fully disconnected network with both agents producing information. This means that when the link cost is very high, every agent decides to produce information and not to get information from the other. Now assume that \( c < k\bar{H} \). It is easy to show that \( g_{12}^*g_{21}^* = 0, g_{12}^* = 1 \) or \( g_{21}^* = 1 \), and \( H^*(X_1) + H^*(X_2) = \bar{H} \). Thus, when the link cost is low, agents generally produce some of the information they need and get some other information from the other agent. However, one possible equilibrium has one agent producing an amount \( \tilde{H} \) of information with the other forming a link with it.

- \( F_{\mathcal{H}}(H(X_1), H(X_2)) = \max\{H(X_1), H(X_2)\} \): Agents may possess fully correlated information in which the joint entropy is always bounded by the entropy of one of them. Fig. 6 plots \( F_{\mathcal{H}} \) which corresponds to the lower surface of the convex cone \( \Gamma_2 \). In this case, it is never beneficial for any agent to form a link and produce a positive amount of information simultaneously. For \( c > k\bar{H} \), we have a unique equilibrium comprising a fully disconnected network with each agent producing \( \tilde{H} \). For \( c < k\bar{H} \), we have only one agent producing positive amount of information in every equilibrium.

Thus, information redundancy influences the agents’ choice for information production. When the information contains no redundancies, there exist many equilibria in which both agents produce positive amount of information when \( c < k\bar{H} \). However, for \( c < k\bar{H} \), when agents have strongly correlated information, every equilibrium has only one agent producing information. Thus, redundancy discourages information sharing between agents and reduces the number of agents producing information when the link cost is low. When \( c > k\bar{H} \), we always have a disconnected network with all agents producing information for both aggregation functions. However, the total amount of information in the network when the random variables of both agents are independent is \( H(X_1, X_2) = 2\bar{H} \), while when the information of both agents are fully correlated (i.e., \( H(X_1, X_2) = \max\{H(X_1), H(X_2)\} \)), we have \( H(X_1, X_2) = \bar{H} \). In the next subsection, we generalize these results to the \( \tilde{G}^N \) game.

**C. Characterization of the NE for \( \tilde{G}^N \) and asymptotic information production behavior**

In this subsection, we characterize the NE for the \( \tilde{G}^N \) game. We study the equilibria for the two aggregation functions \( \tilde{F}_{\mathcal{H}}(H(X_1), H(X_2), ..., H(X_K)) = \sum_{i=1}^{K} H(X_i) \), and \( \tilde{F}_{\mathcal{H}}(H(X_1), H(X_2), ..., H(X_K)) = \max\{H(X_1), H(X_2), ..., H(X_N)\} \). In the following Theorem, we obtain some properties of the equilibria of \( \tilde{G}^N \) when the aggregation function is \( \tilde{F} \).

**Theorem 8:** For the aggregation function \( \tilde{F} \) we have:

(1) If \( c > k\bar{H} \), then for an \( \tilde{s}^* \) where the network is fully disconnected and every agent produces the individually optimal amount of information \( (H^*(X_i) = \bar{H}) \).

(2) If \( c < k\bar{H} \), then \( s^* \) is an equilibrium if and only if: (1) the CIN is minimally connected, (2) the total amount of information is \( H(X) = \bar{H} \), and (3) if any agent \( i \) forms a link in the network \( (g_{ij}^* = 1, i, j \in N) \), then the
cost of linking should be less than the cost of producing the amount of information obtained by forming a link $c \leq kH^*(X_{-i})$.

**Proof** See Appendix G. ■

Condition (1) results from indirect information sharing among connected agents. In addition, the network has a total information of $\bar{H}$ since all agents perfectly share the information they produce, which results in condition (2). Finally, condition (3) says that the cost of linking should be less than the cost of producing the amount of information obtained via linking. In the following Theorem, we characterize the equilibrium when the aggregation function is $F$.

**Theorem 9:** If $c > k\bar{H}$, then for an aggregation function $F$, there exists a unique equilibrium $s^*$ where $g^*_i = 0$, and $H^*(X_i) = \bar{H}$, $\forall i, j \in \mathcal{N}$. If $c < k\bar{H}$, then $s^*$ is an equilibrium if and only if: (1) the CIN is minimally connected, (2) there exists exactly one agent $i$ with $H^*(X_i) = \bar{H}$, and $H^*(X_{-i}) = 0$, (3) all agents with zero information production form exactly one link.

**Proof** See Appendix H. ■

Theorem 9 states that when agents’ information is strongly correlated, information production is monopolized by exactly one agent. That is, unlike the case of uncorrelated information, agents do not distribute the production of information among multiple agents who produce complementary information. Thus, we conclude that information redundancy can have significant impact on the information production behavior at equilibrium.

Several questions arise in networks where cognitive agents take joint information production and link formation decisions: what is the fraction of agents producing information at equilibrium when the number of agents increase? What is the asymptotic total amount of information in the network? In this subsection, we aim at addressing these questions and characterizing the asymptotic informational behavior of agents in a CIN.

In the rest of this subsection, we investigate the asymptotic behavior of two basic qualities: the fraction of agents producing information at equilibrium, and the total amount of information in the network. Denote the set of agents producing information at equilibrium by $\mathcal{I}(s) = \{ i \mid i \in \mathcal{N}, \text{ and } H^*(X_i) > 0 \}$. [22] show that if every additional piece of information has the same value and there is no indirect information sharing, then in equilibrium, information as produced by only a small subset of agents and that fraction becomes vanishingly small as the network size grows, i.e. $\lim_{N \to \infty} \sup_{s^* \in \mathcal{S}} \frac{\mathcal{I}(s^*)}{N} = 0$. [22] calls this “the law of the few”. In the next corollary, we characterize the fraction of information producers and the total amount of information in the network in $\bar{G}^\infty$.

**Corollary 2:** In the $\bar{G}^N$ game, when $c > k\bar{H}$, we have

$$\lim_{N \to \infty} \frac{\mathcal{I}(s^*)}{N} = 1,$$

for both $F_H$ and $E_H$. For $E_H$, the total amount of information in the network in $G^\infty$ is given by

$$\lim_{N \to \infty} H(X_1, X_2, ..., X_N) = \bar{H},$$
while for $\tilde{F}_H$ we have
\[
\lim_{N \to \infty} H(X_1, X_2, \ldots, X_N) = \infty. 
\]

**Proof** From Theorem 8, we know that when $c > k\bar{H}$, then we have a unique equilibrium $s^*$ for both $F_H$ and $\tilde{F}_H$ in which $g_{ij}^* = 0, \forall i, j \in \mathcal{N}$, and $H^*(X_i) = \bar{H}$. Thus, we have $H^*(X_i) > 0, \forall i \in \mathcal{N}$, and $\frac{|I(s^*)|}{N} = 1$, which applies when the number of agents in the CIN grows to infinity, hence (13) follows. Next, we focus on the total amount of information in the network. For $E_H$, we have $H(X_1, X_2, \ldots, X_N) = \max\{\bar{H}, \bar{H}, \ldots, \bar{H}\} = \bar{H}$, and (14) follows. Finally, for $\tilde{F}_H$, we have $H(X_1, X_2, \ldots, X_N) = \sum_{i=1}^{N} \bar{H} = N\bar{H}$, and (15) follows.

Corollary 2 says that when the link cost is very high, the network is fully disconnected and every agent produces the information it needs. Thus, when the network is asymptotically large, every agent is an information producer no matter what the amount of information redundancy is. The number of agents producing information is always $N$. While the number of information producers does not depend on $F_H$, it is clear that the total amount of information in the network depends on the amount of redundancy. When the agents information are strongly correlated, the total amount of information is always bounded by $\bar{H}$. On the other hand, when agents have uncorrelated information, the total amount of information in the network is unbounded in asymptotically large networks. In the next corollary, we study the case of $c < k\bar{H}$.

**Corollary 3:** In the $\tilde{G}^N$ game, when $c < k\bar{H}$, the fraction of information producers for $E_H$ is given by
\[
\lim_{N \to \infty} \sup_{s^* \in \mathcal{S}^*} \frac{|I(s^*)|}{N} = 0, 
\]
while for $\tilde{F}_H$ we have
\[
\lim_{N \to \infty} \sup_{s^* \in \mathcal{S}^*} \frac{|I(s^*)|}{N} = 1. 
\]

For both $E_H$ and $\tilde{F}_H$, the total information in the network is
\[
\lim_{N \to \infty} H(X_1, X_2, \ldots, X_N) = \bar{H}. 
\]

**Proof** We start by deriving (16). From Theorem 9, we know that for $\tilde{F}_H$, every equilibrium has only one information producer. When the number of agents grows to infinity, we will still have one information producer and $\frac{|I(s^*)|}{N} = 0$. In order to prove (17), one needs to find one network in equilibrium for $\tilde{F}_H$ in which, for arbitrary $N$, we have $N$ information producers. Consider this network for $N$ agents. Assume that $H(X_i) = \frac{\bar{H}}{N}, \forall i \in \mathcal{N}$, and the network has a single component which is periphery-sponsored star network. It is clear that for this network, $|I(s)| = N$. We want to show that this network is an NE by showing that every agents strategy is best response to all others. It is easy to see that since $c < k\bar{H}$, each periphery agent has no incentive to break its link with the core since $\frac{\Delta}{N} - k\bar{H} > c$ when $N$ is asymptotically large. Moreover, no agent has incentive to alter its information production profile since the total information in the network is $\sum_{i=1}^{N} \frac{\bar{H}}{N} = \bar{H}$. Thus, $s$ is an NE. Since this applies to any $N$, (17) follows. Finally, since the network is always connected in any equilibrium, then (18) directly follows. ■
This corollary states that the law of the few introduced in [22] does not apply in the case of indirect information sharing. In other words, there exists an equilibrium for some information aggregation function in which all agents in the network are information producers, and production is no longer dominated by a small set of hub agents. Moreover, it is clear that this property depends on the amount of information redundancy. If the agents’ information are strongly correlated, the law of the few applies and information production is dominated by a small fraction of agents in every equilibrium. Fig 7 depicts the equilibria for 8-agents when the aggregation function is $\tilde{H}$ and $\tilde{F}$.

Similar to the results obtained in section III, we note that information redundancy induces phase transitions in the network. Let the network property $A$ for the phase transition be the property that the number of agents producing information in any equilibrium is infinite, i.e. $\inf_{s^* \in S} \frac{T(s)}{N} = 1$. It is clear that when the aggregation function is $E_H$, we have a sharp phase transition, i.e. $\Psi = 0$, and a cost threshold of $c = k\bar{H}$. When $c < k\bar{H}$, only one agent produces information, thus $\lim_{N \to \infty} \frac{T(s)}{N} = 0$. Contrarily, when $c > k\bar{H}$, all agents produce information, and we have $\lim_{N \to \infty} \frac{T(s)}{N} = 1$. 

Fig. 7: Exemplary equilibria for different aggregation functions. The law of the few does not apply when the agents’ information has no redundancies.

Fig. 8: Phase transition in CINs with strongly correlated information structure.
VI. CONCLUSIONS

In this work, we present a first model for the endogenous formation of networks by cognitive agents who aim at gathering and producing information. Using Nash Equilibrium (NE) as a solution concept, we formulated a non-cooperative network formation game where agents get informational benefits by forming costly links with each other. We show that the information possessed by the cognitive agents affects the network topology, efficiency, and information production behavior. We show the impact of information redundancy on the topologies of NE networks, and its impact on the network efficiency in terms of the Price-of-Anarchy (PoA) and Maximum Information Loss (MIL). Finally, we consider the asymptotic behavior of a network where each agent both produces information and forms links with other agents. For such networks, we study the impact of information redundancy on the number of agents producing information at equilibrium. The validity of the law of few depends on how information aggregates.

APPENDIX A

PROOF OF THEOREM 1

We know that if we allow mixed strategies, then every game with a finite number of players in which each player can choose from finitely many pure strategies has at least one Nash equilibrium. Assume that agent $i$ adopts a mixed strategy $\Delta_i = (p_{i1}, p_{i2}, \ldots, p_{iN})$, where $p_{ij}$ is the probability that agent $i$ forms a link with agent $j$. The utility of agent $i$ in this case is obtained by averaging over all possible networks as follows

$$u_i(\Delta_i) = \sum_{j=1}^{2^{N-1}-1} w_j f(H(X_i \cup X_{\alpha_j})) - \sum_{t=1}^{N-1} p_{it}c,$$

where $\alpha_j$ is an element of the power set of $\mathcal{N}/\{i\}$, and $w_j$ is the probability of the emergence of a certain network based on the mixed strategies. For instance, in a 2 agent network, the utility function of agent 1 is given by

$$u_1(\Delta_1) = (p_1(1-p_2) + p_2(1-p_1) + p_1p_2) f(H(X_1, X_2))$$

$$+ (1-p_1) (1-p_2) f(H(X_1)) - p_{12} c.$$ 

In this case, $w_1 = p_1(1-p_2) + p_2(1-p_1) + p_1p_2$ and $w_2 = (1-p_1)(1-p_2)$. Now assume we induce a perturbation $\epsilon$ to the mixed strategy of agent $i$ by modifying $p_{ik}$ to $p_{ik} + \epsilon$ for a certain $k$, where $\epsilon \in [-p_{ik}, 1-p_{ik}]$. We call this modified strategy $\Delta_i^\epsilon$. Note that we can write any $w_j$ in (A.1) in the form of $w_j = \tilde{w}_j p_{ik} + \bar{w}_j (1-p_{ik})$. This results in a perturbed utility $u_i^\epsilon(\Delta_i^\epsilon)$ as follows

$$u_i^\epsilon(\Delta_i^\epsilon) = \sum_{j=1}^{2^{N-1}-1} \left( \tilde{w}_j (p_{ik} + \epsilon) f(H(X_i \cup X_{\alpha_j})) \right)$$

$$+ \bar{w}_j (1-\epsilon-p_{ik}) f(H(X_i \cup X_{\alpha_j})) - (p_{ik} + \epsilon)c - \sum_{l=1, l\neq k}^{N-1} p_{il}c. \quad (A.2)$$

which can be rearranged as

$$u_i^\epsilon(\Delta_i^\epsilon) =$$
Let \( \delta = \sum_{j=1}^{N-1} (\tilde{w}_j - \tilde{w}_j) f \left( H(X_i \cup X_{\alpha_j}) \right) - c \). It can be easily observed that $\frac{\partial u^*_i(\Delta_i)}{\partial \epsilon} > 0$ if $\delta > 0$, and $\frac{\partial u^*_i(\Delta_i)}{\partial \epsilon} < 0$ otherwise. Thus, if $\delta > 0$, agent $i$ can always increase its utility by increasing $\epsilon$ and setting $\epsilon = 1 - p_{ik}$, i.e. playing a pure strategy $p_{ik} = 1$. On the other hand, if $\delta < 0$, agent $i$ can always increase its utility by decreasing $\epsilon$ such that $p_{ik} = 0$. Thus, for all $k \in \mathcal{N}/\{i\}$, agent $i$ selects a strategy $p_{ik} \in \{0, 1\}$. Due to symmetry, this applies to all agents in $\mathcal{N}$. Therefore, it follows that a pure strategy NE always exists.

**Appendix B**

**Proof of Proposition 1**

If the component $C$ is not minimally connected, then it has at least one cycle as there exist agents $i$ and $j$ that are connected via two paths $p_{ij,1}$ and $p_{ij,2}$, such that any of the two paths is not a subset of the other. For such component at NE, assume that agent $v$ is on path $p_{ij,1}$ and agent $w$ is on path $p_{ij,2}$. Note that all the agents receive the same amount of total information $H(C)$. We know that there indeed exists links: $g^*_{xv}$ (or $g^*_{vx}$) and $g^*_{wy}$ (or $g^*_{yw}$), where agent $x \in p_{ij,1}$ and agent $y \in p_{ij,2}$. Now focus on any link of them, say $g^*_{wy} = 1$. We observe that agent $w$ can break this link and still receive the same benefit by gathering the same amount of information from path $p_{ij,1}$, thus receiving a strictly higher utility function as it will not pay the cost for the link with agent $y$, which contradicts the fact that $g^*$ is an NE. Thus, a single path exists between any two agents.

**Appendix C**

**Proof of Lemma 1**

If there exists an agent in which other agents have an incentive to connect to even if they possess all other information in the network, then the network is indeed connected at any equilibrium. This is satisfied if and only if the linking cost satisfies $c < f(H(\mathcal{X})) - f(H(X_i))$ for some agent $i$ in $\mathcal{N}$, i.e. the marginal benefit from connecting to that agent is always more than the link cost irrespective to the current connections of the agent forming the link. Thus, we must have $c \leq \max_i f(H(\mathcal{X})) - f(H(X_i))$. Hence, part (i) of the Lemma follows.

If no agent have an incentive to form any link, then the network is fully disconnected. From the monotonicity property of the entropy, we know that if agent $i$ has no incentive to connect to a set $\mathcal{V}$ of agents, then it has no incentive to connect to a set $\mathcal{U}$ if $\mathcal{U} \subseteq \mathcal{V}$. Thus, if agent $i$ has no incentive to connect to the set $\mathcal{N}/\{i\}$ via a single link, then it has no incentive to form any link in the network. This occurs if $c > f(H(\mathcal{X})) - f(H(X_i))$. If this condition is satisfied for all agents, then the network is indeed disconnected, and part (ii) of the Theorem follows.

**Appendix D**

**Proof of Theorem 3**

For the network to be in NE, no agent should have an incentive to unilaterally deviate by forming a new link or breaking a link. Focus on a certain component $C_i$. Inside this component, each agent should either have an
incentive to form at least one link, or other agents should have an incentive to connect to it. Otherwise, this agent can be removed from the component while strictly increasing the utility of some agent. Thus, we must have either \( f(H(X_{C_j})) - f(H(X_j)) > c \) or \( f(H(X_{C_j})) - f(H(X_{C_i/f(j)})) > c \) for all agents \( j \) in \( C_i \). This should apply to all components in the network. Hence, condition (1) follows.

Now focus on the interaction between different components of the network. If any agent in component \( C_i \) benefits from forming a link to any agent in component \( C_j \), then the network is not NE since in this case an agent in \( C_i \) can strictly increase its utility by unilateral deviation. Hence, we should have \( f(H(X_{C_i/f(j)})) - f(H(X_{C_i})) \leq c \) for any two components in the network. Thus, condition (2) follows.

**Appendix E**

**Proof of Theorem 4**

Under strict NE, a non-singleton component \( C \) has two agents \( i \) and \( j \) such that \( g_{ij}^* = 1 \). Now assume that \( g_{kj}^* = 1 \) for some agent \( k \in C \). It is clear that \( k \) can achieve the same utility by deleting its link with \( j \) and connecting to \( i \). This contradicts with the fact that \( g^* \) is a strict NE. Thus, \( g_{kj}^* = 0 \). Using a similar argument, it can be shown that \( g_{ki}^* = 0 \). Thus, we conclude that \( g_{ik}^* = 1 \). This is true for all agents \( k \in C \), which implies that a single core agent \( i \) forms links with all other agents in \( C \). Therefore, the core agent \( i \) should strictly increase its utility for each of the \( M - 1 \) links it forms. The benefit of agent \( i \) from forming a link with agent \( j \) given that \( i \) is connected to all other agents in \( C \) is given by \( f(H(X_{C})) - f(H(X_{C/f(j)})) - c \). This should be positive for all periphery agents \( j \) in \( \zeta_C \), because otherwise core agent \( i \) can break some of the links in the component. Thus, for agent \( i \in C \) to be a core agent, and for agents \( j \in C \) to reside in the periphery, we must have \( f(H(X_{C})) - f(H(X_{C/f(j)})) > c, \forall j \neq i \). Note that conditions (1) and (2) in Theorem 3 should also be satisfied for the network to be at NE, while the feasibility of organizing each component as a core-sponsored star guarantees that the network is at strict NE. Thus, strict NE exists if there exists an NE with the set \( \zeta \) has a cardinality not less than \( M - 1 \) for all components, i.e. a single core agent can sponsor each component.

**Appendix F**

**Proof of Lemma 3**

For a connected network in the \( K_C \) region, the utility of agent \( i \) is given by \( u_i(g^*) = f(H(\mathcal{X})) - \sum_{m \in N_i(g^*)} c_m \). The social welfare is given by \( U(g^*) = \sum_{i \in \mathcal{N}} u_i(g^*) \). Since we know from Proposition 1 that the network is minimally connected at equilibrium, then it has exactly \( N - 1 \) links. Therefore, we have \( U(g^*) = N \int (H(\mathcal{X})) - \sum_{j \in \mathcal{J}} c_j \), where \( \mathcal{J} \) is the set of links in the network designated by the index of link recipient, and \( \mathcal{J} = N - 1 \). It is obvious that the social optimal topology is a periphery-sponsored star with the agent \( k = \arg \min_j c_j \) residing in the core of the star. The social welfare of such topology is \( U(g^*) = N \int (H(\mathcal{X})) - (N - 1) \min_j c_j \). Note that this is also an NE equilibrium as each agent does not benefit from breaking its link with the core agent and linking to any other periphery agents. Now, we aim at identifying the equilibrium with the worst social welfare. Assume that the link costs are arranged ascendingly as \( c_1 < c_2 < c_3 < \ldots < c_{N-1} < c_N \). We know that the network is
minimally connected, thus total costs of link formation is given by $\sum_{j \in J} c_j$. What are the elements of the set $J$ such that the total cost is maximized and the network is at equilibrium? Note that for the socially optimal profile, $J = \{1, 1, 1, ..., 1\}$, with a cardinality of $N - 1$. Now assume a line network with $g^*_{i,i+1} = 1, \forall 1 \leq i < N$. Thus, we have $g^*_{12} = g^*_{23} = ... = g^*_N N = 1$. Thus, $J = \{2, 3, ..., N - 1, N\}$. It can be easily shown that this line network is stable, since no agent $i$ can break its link with agent $i + 1$ and increase its utility. For instance, if $i$ breaks its link with $i + 1$, it must connect to any agent $j > i + 1$ to receive the same amount of information but at a higher cost. It can be also shown that this is the worst equilibrium. This is because for a connected network, only one agent $i$ connects to the agent $N$ with the highest link cost, and others can connect to $i$ and get the information of $N$ via indirect sharing. The same applies to agent $N - 1$, where one agent connects to it, and others share information by connecting to that agent. Thus, to maximize the total link cost and maintain equilibrium, only one link is formed with each agent except the one with the minimum link cost. Thus, the social welfare in this case is $NH(\mathcal{X}) - \sum_{j=1}^{N} c_j + \min_k c_k$, and the PoA formula follows.

For the $K_I$ region, it is clear that the PoA = 1 since we have a unique equilibrium. Finally, for $K_M$ region, the proof is the same as that of Lemma 2.

APPENDIX G

PROOF OF THEOREM 8

We start with the case of $c > k\bar{H}$. Assume that there exists a link in $g^*$ with $g^*_{ij} = 1$. In this case, agent $i$ can always better off by breaking this link and producing an amount $\bar{H}$ of information. This applies to any agent $i$ in $\mathcal{N}$. Thus, we have a unique equilibrium with $g^*_{ij} = 0$, and $H^*(X_i) = \bar{H}, \forall i, j \in \mathcal{N}$. Now focus on the case of $c < k\bar{H}$. We show that if $s$ satisfies (1), (2), and (3), then $s$ is an NE. The minimality of each network component can be easily proved using Proposition 1. Now we show that the connected network is an NE. Since in a connected network an agent has to produce an amount $\bar{H}$ of information, which is not optimal since $c < k\bar{H}$. Thus, no agent in a connected network has incentive to break its link and part (1) follows. Since the network is minimally connected, then each agent obtains all the total amount of information $H(\mathcal{X})$. If $H(\mathcal{X}) = \bar{H}$, then no agent in the component has incentive to alter their information production profile because all agents benefit only from obtaining an amount $\bar{H}$ of information. Thus, part (2) is proved. Finally, if $c \leq kH^*(X_{-i}), \forall i$, then no agent in the network has incentive to break the link it forms and produce an amount $H^*(X_{-i})$ of information on its own. Thus, $s$ is a Nash equilibrium.

We now prove the converse. Let $s$ be an equilibrium. Assume that the network has two components $C_1$ and $C_2$. It is clear that total amount of information in each component must be $\bar{H}$ at equilibrium, thus, any agent with positive amount of information production in one component will better off by not producing any information and forming a link to the other component. Thus, the network is connected in NE and part (1) follows. Due to indirect information sharing, part (2) is directly concluded. Finally, if $s$ is an equilibrium and $g_{ij} = 1$, then this should be optimal for agent $i$, thus $c \leq kH^*(X_{-i}), \forall iN$. 

March 17, 2015 DRAFT
APPENDIX H

PROOF OF THEOREM 9

The case of $c > kH$ is exactly the same as in Theorem 8 and the proof will be similar to that in Appendix G. Now focus on the case of $c < kH$. We show that if $s$ satisfies (1), (2), and (3), then $s$ is an NE. Part (1) follows from Proposition 1 and the proof of Theorem 8. Now assume that only one agent in the network produces $H$ information and all others do not produce any information and only form links in the network. In this case, the agent producing information does not better off by producing any amount of information other than $H$. In addition, the agents forming links do not better off by forming new links or breaking their links and producing information since $c < kH$. Since there are $N - 1$ agents forming links, then the network is connected, and no agent benefits from forming an extra link in the network, which concludes part (2).

We now prove the converse. Let $s$ be an equilibrium. Due to indirect information sharing, part (1) follow straightforwardly. Assume that we have two agents with $H^*(X_i) > H^*(X_j) > 0$, then agent $j$ can always better off by setting $H^*(X_j) = 0$ since the aggregate information of $i$ and $j$ is $H^*(X_j)$. Therefore, the agent with maximum information production has to set $H^*(X_i) = H$, and all others do not produce information and form a link in the network since $c < kH$. Finally, since $s$ is an equilibrium, agents act optimally, thus each agent from the set of $N - 1$ non-producers forms exactly one link in the network.

REFERENCES

[1] F. Khozeimeh, and S. Haykin, “Brain-Inspired Dynamic Spectrum Management for Cognitive Radio Ad Hoc Networks,” IEEE Trans. Wireless Commun., vol. 11, no. 10, pp. 3509-3517, Oct. 2012.
[2] K.-C. Chen, M. Chiang, and H. V. Poor, “From Technological Networks to Social Networks,” IEEE J. Sel. Areas Commun., vol. 31, no. 9, pp. 548-572, Sep. 2013.
[3] O. Yagan, D. Qian, J. Zhang, D. Cochran, “Conjoining speeds up information diffusion in overlaying social-physical networks”, IEEE J. Sel. Areas Commun., vol. 31, no. 6, pp. 1038-1048, 2013.
[4] Z. Yuan, J. B. Song, and Z. Han, “Interference Aware Routing Using Network Formation Game in Cognitive Radio Mesh Networks,” IEEE J. Sel. Areas Commun., vol. 31, no. 11, pp. 2494-2503, Nov. 2013.
[5] W. Saad, Z. Han, A. Hjorungnes, D. Niyato, E. Hossain, “Coalition Formation Games for Distributed Cooperation Among Roadside Units in Vehicular Networks,” IEEE J. Sel. Areas Commun., vol. 29, no. 1, pp. 48-60, Jan. 2011.
[6] A. P. Athreya and P. Tague, “Network Self-Organization in the Internet of Things,” IEEE International Workshop on Internet-of-Things Networking and Control, 2013.
[7] P. K. Biswas, and S. Phoha, “Self-Organizing Sensor Networks for Integrated Target Surveillance,” IEEE Trans. Computers, vol. 55, no. 8, pp. 1033-1047, Aug. 2006.
[8] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and Tamer Basar,”Coalitional Games for Distributed Collaborative Spectrum Sensing in Cognitive Radio Networks,” IEEE INFOCOM, pp. 2114-2122, 2009.
[9] T. Shu and M. Krunz.”Coordinated Channel Access in Cognitive Radio Networks: A Multi-level Spectrum Opportunity Perspective,” IEEE INFOCOM, pp. 2976-2980, 2009.
[10] H. Sugiyama, T. Tsujioka, and M. Murata, “Autonomous chain network formation by multi-robot rescue system with ad hoc networking,” 2010 IEEE International Workshop on Safety Security and Rescue Robotics (SSRR), Bremen, 26-30 July, 2010.
[11] V. Bala and S. Goyal, “A non-cooperative model of network formation,” Econometrica, vol. 68, no. 5, pp. 1181-1231.
[12] M. Jackson and A. Watts, “On the Formation of Interaction Networks in Social Coordination Games,” Games and Economic Behavior, vol. 41, pp. 265-291, 2002.
[13] Y. Zhang and M. van der Schaar, “Information Production and Link Formation in Social Computing Systems,” IEEE J. Sel. Areas Commun., vol. 30, no. 11, pp. 2136-2145, Dec. 2012.

[14] Y. Zhang and M. van der Schaar, “Strategic Networks: Information Dissemination and Link Formation Among Self-Interested Agents,” IEEE J. Sel. Areas Commun., vol. 31, no. 6, pp. 1115-1123, Jun. 2013.

[15] J. Xu, S. Zhang, and M. van der Schaar, “Network Dynamics with Incomplete Information and Learning,” Allerton 2014.

[16] Y. Song and M. van der Schaar, “Dynamic Network Formation with Incomplete Information,” accepted and to appear in Economic Theory, 2015.

[17] J. Barros and S. D. Servetto, “Network Information Flow With Correlated Sources,” IEEE Trans. Inform. Theory, vol. 52, no. 1, Jan. 2006, pp. 155-170.

[18] Y. Sung, H. V. Poor, and H. Yu, “How Much Information Can One Get From a Wireless Ad Hoc Sensor Network Over a Correlated Random Field?,” IEEE Trans. Inform. Theory, vol. 55, no. 6, Jun. 2009, pp. 2827-2847.

[19] H. Gupta, V. Navda, S. R. Das, V. Chowdhary, “Efficient Gathering of Correlated Data in Sensor Networks,” ACM Transactions on Sensor Networks (TOSN), vol. 4, no. 1, Jan. 2008.

[20] A. Giridhar and P. R. Kumar, “Computing and Communicating Functions Over Sensor Networks,” IEEE J. Sel. Areas Commun., vol. 23, no. 4, pp. 755-764, Apr. 2005.

[21] H. Luo, Y. Liu, and S. K. Das, “Routing Correlated Data with Fusion Cost in Wireless Sensor Networks,” IEEE Trans. Mobile Comp., vol. 5, no. 11, pp. 1620-1632, Nov. 2006.

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

[23] Z. Zhao, R. Zhang, X. Cheng, L. Yang, and B. Jiao, Network Formation Games for the Link Selection of Cooperative Localization in Wireless Networks,” IEEE ICC, pp. 4577-4582, 2014.

[24] A. Ramamoorthy, “Minimum Cost Distributed Source Coding Over a Network,” IEEE Trans. Inform. Theory, vol. 57, no. 11, Jan. 2011.

[25] A. Lee, M. Medard, K. Z. Haigh, S. Gowen, and P. Rubel, “Minimum-Cost Subgraphs for Joint Distributed Source and Network Coding,” Proc. of Workshop on Network Coding, Theory and Applications, 2007.

[26] Z. Zhang and R. W. Yeung, “On Characterization of Entropy Function via Information Inequalities,” IEEE Trans. Inform. Theory, vol. 44, no. 4, pp. 1440-1452, Jul. 1998.

[27] T. M. Cover and J. A. Thomas, “Elements of Information Theory,” New York, Wiley, 2006.

[28] G. Quer, F. Librino, L. Canzian, L. Badia, and M. Zorzi, “Inter-Network Cooperation Exploiting Game Theory and Bayesian Networks,” IEEE Trans. Commun., vol. 61, no. 10, pp. 4310 - 4321, Oct. 2013.

[29] M. Krivelevich and B. Sudakov, “The phase transition in random graphs ?a simple proof”, http://arxiv.org/abs/1201.6529, 2012.