NETWORK FORMATION GAMES BASED ON CONDITIONAL INDEPENDENCE GRAPHS

Sergio Barbarossa[1], *Paolo Di Lorenzo*[1], *Mihaela van der Schaar*[2]

[1] Sapienza University of Rome, DIET, Via Eudossiana 18, 00184 Rome, Italy.
[2] University of California Los Angeles (UCLA) EE Dept., 420 Westwood Plaza, Los Angeles, USA
E-mail: {sergio, dilorenzo}@infocom.uniroma1.it, mihaela@ee.ucla.edu

ABSTRACT

The goal of this paper is to propose a network formation game where strategic agents decide whether to form or sever a link with other agents depending on the net balance between the benefit resulting from the additional information coming from the new link and the cost associated to establish the link. Differently from previous works, where the benefits are functions of the distances among the involved agents, in our work the benefit is a function of the mutual information that can be exchanged among the agents, conditioned to the information already available before setting up the link. An interesting result of our network formation game is that, under certain conditions, the final network topology tends to match the topology of the Markov graph describing the conditional independencies among the random variables observed in each node, at least when the cost of forming a link is small.

***Index Terms*—** Network formation game, Markov graphs

1. INTRODUCTION AND RELATED WORK

The problem addressed in this work is how networks of strategic agents form and how the network formation mechanism is related to the statistical properties of the observations available to each agent. The strategic network formation literature studies how networks form as a result of the strategic choices of agents deciding whether they should link with other agents by trading off the benefits associated with connecting to other agents versus the costs incurred by forming links, i.e., links are strategic choices made by the agents. Such network formation games have been investigated in the economics literature since the seminal works of Jackson and Wolinsky [1], Bala and Goyal [2], Jackson and Watts [3], and Ballester et al. [4]. All these works assumed complete information available at all the agents. Only recently, the work of Song and van der Schaar studied the incomplete information case and showed the effect of partial information on the final topology [5]. Moreover, most of these works assumed that agents are homogeneous, but heterogeneity is clearly an important fact when the goal of the agents is to produce, disseminate and consume (diverse) information. Information networks (i.e. networks where the goal of the agents is to strategically produce, disseminate and consume information) have been formalized and studied for the first time in Zhang and van der Schaar [6]-[7]. In these works, agents are considered to be heterogeneous, i.e. they produce different kinds of information (different types of music or news reports) and want to consume/disseminate diverse information.

Differently from all previous works, where the benefit depends on the distance among the nodes, expressed in number of hops along the geodesic path connecting them, in this paper the benefit is a function of the additional information (in the Shannon sense) that an agent can acquire through new links, which depends on the statistical relations among the variables involved. In particular, we will focus on Markov graphs [8], even if our approach is not restricted to hold only in such a case. There are then two graphs to keep in mind: the conditional independence graph and the real network composed of the links formed by the agents. An interesting result is that, if the cost is sufficiently low, the topology of the network formed by strategic agents tends to coincide with the topology of the Markov graph. We consider, as examples of application, the minimization of prediction variance and the maximization of mutual information. Some numerical results corroborate our findings, including the case of imperfect knowledge of the statistical parameters.

2. OBSERVATION MODEL

Given a group of n agents, we denote by X_i the observation of agent i, with $i = 1, \ldots, n$. We assume that the observations X_i are instantiations of *statistically dependent* random variables whose statistical dependencies is represented through the conditional independence graph, defined as follows [8]: The *conditional independence graph* associated to a set of random variables $X_1, X_2, \ldots, X_n$, is the undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, with $\mathcal{V}$ and $\mathcal{E}$ denoting the sets of n vertexes and the set of edges, respectively, with the property that a link ij is not in the edge set $\mathcal{E}$ if and only if X_i is statistically independent of X_j, conditioned to all other variables. We use the symbol $X_i \perp\!\!\!\perp X_j | X_A$ to indicate that X_i is statistically independent of X_j, conditioned to the variables whose indices

The work of S. Barbarossa and P. Di Lorenzo has been supported by TROPIC Project, Nr. 318784.

belong to the set A. We also denote by $\mathcal{N}_i$ the boundary of i (i.e., the set of direct neighbors of i) and by $\overline{\mathcal{N}}_i$ the complement of $\mathcal{N}_i \cup i$ in $\mathcal{V}$. A random vector $\boldsymbol{X} := (X_1, \ldots, X_n)$ with graph $\mathcal{G}$ satisfies the *local Markov property* if, for every vertex i, it holds $X_i \perp\!\!\!\perp X_{\overline{\mathcal{N}}_i} | X_{\mathcal{N}_i}$. Alternatively, the *pairwise Markov property* states that, for all non-adjacent vertices i and j, we have $X_i \perp\!\!\!\perp X_j | X_A$, where A is the set of all nodes in $\mathcal{V}$ except i and j. Interestingly, these two properties are equivalent [8]. The Markov graph might be disconnected. In such a case, the graph will be composed of separated clusters, denoted by $\mathcal{I}_k$, with $k = 1, 2, \ldots, K$, where K is the number of clusters, of statistically independent variables. In the next section, we will show how these properties are going to affect the final topology of our network formation game.

3. NETWORK FORMATION GAME

Let $N = \{1, 2, \ldots, n\}$ be the set of agents, with $n \geq 3$. Each agent, let us say agent i, possesses some local information resulting from the observation of a random variable X_i. Its goal is to establish links with other agents if this can bring additional information, taking into account the cost associated to forming a link. The strategy adopted by an agent i is denoted by a tuple $\boldsymbol{g}_i = \{g_{ij}\}_{j \in N/i} \in \{0, 1\}^{n-1}$, where $g_{ij} = 1$ if agent i forms a link with agent j and $g_{ij} = 0$, otherwise. The decision to form a link is the result of a trade-off between the benefit resulting from forming a link and the cost associated to establishing a link. The topology of the network is then represented by an adjacency matrix whose rows are the tuple's $\boldsymbol{g}_i$, for $i = 1, \ldots, n$. The set of all strategies is denoted by $\boldsymbol{g} := (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n)$. We denote with $\mathbb{N}_i$ the set of neighbors of agent i in the adjacency graph and with $\mathbb{C}_i$ the connected component (excluding node i) which node i belongs to. We use a different symbol here for the neighborhood $\mathbb{N}_i$ of agent i in the network, to distinguish it from the neighborhood $\mathcal{N}_i$ of the random variable X_i in the conditional independence graph. For any player i, we call the nodes in $\mathcal{N}_i$ the *Markov neighbors* of i; conversely, the nodes in $\mathbb{N}_i$ are simply the (network) neighbors of i. The utility of agent i is expressed as:

$$u_i(\boldsymbol{g}) = b_i(\mathbb{C}_i) - \sum_{k \in \mathbb{N}_i} c_{ik}, \tag{1}$$

where $b_i(\mathbb{C}_i)$ is the benefit for player i to belong to component $\mathbb{C}_i$, while c_{ik} is the cost associated to establishing a direct (one hop) link between nodes i and k[1].

In most of the literature on this subject, the benefits are typically functions of the distance (expressed in terms of number of hops) between node i and each other node in the subset $\mathbb{C}_i$. In this work, we significantly depart from this assumption, as in many real circumstances, what really counts is the extra information acquired through new links (either direct or indirect), irrespective of the number of hops between the nodes. What we propose instead, is a class of benefit functions that depend on the statistical dependencies among the random variables associated to each node. More specifically, the benefit for node i to add a link to node j is an increasing function of the mutual information $I(X_i; X_{j \cup \mathbb{C}_j} | \mathbb{C}_i)$ between X_i and X_j (plus, possibly, $X_{\mathbb{C}_j}$, i.e., all variables connected to X_j), *conditioned* to the variables $X_{\mathbb{C}_i}$, which i is already connected with. To capture the "Markov" character of our proposed game, we assume that the function $b_i(\mathbb{C}_i)$, for any $j \neq i$ and $j \notin \mathbb{C}_i$, satisfies the following properties:

P.1) if $\mathbb{C}_i \equiv \mathcal{N}_i$ and $j \notin \mathcal{N}_i$, $\Rightarrow b_i(\mathbb{C}_i \cup j) = b_i(\mathbb{C}_i)$;
P.2) if $\mathbb{C}_i \neq \mathcal{N}_i$ and $j \in \mathcal{N}_i$, $\Rightarrow b_i(\mathbb{C}_i \cup j) > b_i(\mathbb{C}_i)$
P.3) if $i \in \mathcal{I}_\ell$ and $j \in \mathcal{I}_k$, with $\mathcal{I}_\ell \cap \mathcal{I}_k = \emptyset$,
$\Rightarrow \quad b_i(\mathbb{C}_i \cup j) = b_i(\mathbb{C}_i), \forall \mathbb{C}_i$.

In words, P.1) states that if node i is already connected with *all* its Markov neighbors, then it is not worthy to add any further neighbor; P.2) states that, whenever a Markov neighbor is not part of the connected component which player i belongs to, it is always beneficial to add such a node in the neighborhood of i; finally, P.3) states that there is no benefit in connecting two statistical independent sets of nodes. It is worth adding that properties P.1 and P.2 do not rule out the possibility that, whenever $\mathbb{C}_i$ does not include *all* Markov neighbors of i, it might be be beneficial for player i to include some non-Markov neighbors. Also, since the decision depends on the cost/benefit net balance, it might also happen that a Markov neighbor is not added, because its addition incurs a cost greater than the benefit.

We consider pure (not mixed) link formation strategies, where each agent maximizes its own utility given the strategies of the others. The choice of a strategy $\boldsymbol{g}_i$ from player i has an impact on the benefits of, potentially all, other players, as the inclusion or elimination of a link may affect the topology of the connected components $\mathbb{C}_k$ associated to (possibly all) other players. Given this interplay among the players, it is then of primary importance to study the conditions ensuring the existence, and possibly uniqueness, of equilibrium points and to devise network formation strategies with provable convergence properties. Before doing that, it is worth emphasizing that, when dealing with network formation games, standard notions like the Nash Equilibrium are not well suited because the creation or elimination of a link implicitly involves some form of mutual consent between the players involved. A more suitable concept, able to capture the need for coordination is *pairwise stability* [1]. A network is pairwise stable with respect to the allocation rules $u_i(\boldsymbol{g})$ if:

(i) $\forall ij \in \boldsymbol{g}, u_i(\boldsymbol{g}) \geq u_i(\boldsymbol{g} - ij)$ and $u_j(\boldsymbol{g}) \geq u_j(\boldsymbol{g} - ij)$;
(ii) $\forall ij \notin \boldsymbol{g}$, if $u_i(\boldsymbol{g} + ij) > u_i(\boldsymbol{g})$ then $u_j(\boldsymbol{g} + ij) < u_j(\boldsymbol{g})$.

In words, these conditions state that the addition of a new link requires mutual consent of the players involved. Con-

[1] In the rest of the paper, we will assume the same cost for all the links, for simplicity, but the theory holds also for different costs.

versely, every player has the discretion to unilaterally eliminate a link, without asking for the consent of the other player involved.

The network formation game starts with an empty network and proceeds through the following stages. At each stage t, a randomly chosen agent i, characterized by a current strategy $\boldsymbol{g}_i^t$ and connected component $\mathbb{C}_i^t$, performs the following steps:
S.1 i picks up at random an agent $j \notin \mathbb{C}_i^t$; i and j add the link ij if $u_i(\boldsymbol{g}^t + ij) - u_i(\boldsymbol{g}^t) \geq 0$ and $u_j(\boldsymbol{g}^t + ij) - u_j(\boldsymbol{g}^t) \geq 0$;
S.2 if link ij is added, agent i checks if it is still convenient to keep its links with its immediate neighbors $\mathbb{N}_i^t$; agent i is allowed to sever a link ik with an older neighbor k if $u_i(\boldsymbol{g} - ik) > u_i(\boldsymbol{g})$;
S.3 nodes i and j update the list of their subsets: $\boldsymbol{g}_i^t \to \boldsymbol{g}_i^{t+1}; \boldsymbol{g}_j^t \to \boldsymbol{g}_j^{t+1}$.
S.4 repeat until no topology changes occur.

To fully specify the game, for each link ij under test, we need to specify how to compute the marginal benefit $u_i(\boldsymbol{g}^t + ij) - u_i(\boldsymbol{g}^t)$. We consider the two extreme cases:
G.S) $u_i(\boldsymbol{g}^t + ij) - u_i(\boldsymbol{g}^t) = b_i(\mathbb{C}_i^t \cup j) - b_i(\mathbb{C}_i^t) - c$;
G.C) $u_i(\boldsymbol{g}^t + ij) - u_i(\boldsymbol{g}^t) = b_i(\mathbb{C}_i^t \cup \mathbb{C}_j^t) - b_i(\mathbb{C}_i^t) - c$.

In the former case, in forming the link ij, agent i gets only the benefit coming from j; in the latter case, i gets the benefit from the overall connected component $\mathbb{C}_j$, which agent j belongs to. The latter case is the most complex to implement, but it typically outperforms the former case.

The game stops when all players do not have any incentive to modify their own neighborhood. If this situation arises, the network is in a pairwise stable condition. In principle, of course, there is no guarantee of convergence of this kind of game. Nevertheless, for the class of functions fulfilling properties P.1, P.2, and P.3, we can show that the above game converges to a pairwise stable equilibrium. More specifically, we prove the following

Theorem: Let us assume that the random variables $X_1, \ldots, X_n$ collected by the n agents are described by a Markov conditional independence graph $\mathcal{G}$. We denote by b_{min} and b_{max} the lower and upper bounds on the marginal benefit: $b_{min} := \min_{i,j \in \mathcal{N}_i} [b_i(\mathbb{C}_i \cup j) - b_i(\mathbb{C}_i)]$;
$b_{max} := \max_{i,j \in \mathcal{N}_i} [b_i(\mathbb{C}_i \cup j) - b_i(\mathbb{C}_i)]$. Then, if $c < b_{min}$, game G.S converges to a topology coinciding with the Markov graph whereas G.C converges to a topology composed by clusters $\mathbb{I}_k$, with $k = 1, \ldots, K$, where each $\mathbb{I}_k$ is a spanning tree of $\mathcal{I}_k$, with possibly less (and different) edges than $\mathcal{I}_k$;
If $c > b_{max}$, the final topology is fully disconnected.
Proof. Let us start with game G.S. At each stage of the game, agent i is confronted with three possible situations:
i) $j \in \mathcal{N}_i$: In this case, because of P.2, if $c < b_{min}$, including ij is beneficial for both i and j, then, link ij is added;
ii) $j \notin \mathcal{N}_i$, but $j, i \in \mathcal{I}_k$: In this case, the link ij could be added or not, depending on the net benefits;
iii) $j \in \mathcal{I}_k$ and $i \in \mathcal{I}_\ell$, with $\mathcal{I}_\ell \cap \mathcal{I}_k = \emptyset$: In this case, the link ij is not added, because of P.3).
From the above, it follows that, sooner or later, all Markov neighbors will be included in each agent's list (plus, possibly, some non-Markov neighbor, in case ii)). But, as soon as the connected component $\mathbb{C}_i$ will include *all* Markov neighbors of i, agent i will drop, by a unilateral decision, all its non-Markov neighbors, because of P.1.). The final result is then a network topology coinciding with the Markov graph.

Let us consider now game G.C. In this case, by construction, the two sets $\mathbb{C}_i$ and $\mathbb{C}_j$ are disjoint, i.e., $\mathbb{C}_i \cap \mathbb{C}_j \equiv \emptyset$, because otherwise $\mathbb{C}_i$ and $\mathbb{C}_j$ will be part of the same connected component. This is not possible, because, at each stage, agent i checks only nodes outside its own connected component. We have now two possibilities:
i) $i \in \mathcal{I}_i$ and $\mathbb{C}_j \cap \mathcal{I}_i \equiv \emptyset$: In this case, link ij is not going to be added, because of P.3;
i) $i \in \mathcal{I}_i$ and $\mathbb{C}_j \cap \mathcal{I}_i \neq \emptyset$: In this case, link ij can be added, depending on the cost/benefit balance.
The game stops as soon as $\mathbb{C}_i$ contains all Markov neighbors of node i. Differently from game G.S, $\mathbb{C}_i$ might include links between non-Markov neighbors and might exclude some direct links between Markov neighbors, but the game will end up with the same clusters as the Markov graph, with possibly different edges within each cluster. No loops will appear, because of step **S.2**. Conversely, if $c > c_{max}$, the cost for establishing a link is so high that there is never an advantage in adding a link. Hence, the final topology is the fully disconnected one. ■

4. APPLICATIONS

Prediction: Let us suppose that the agents are sensors; $b_i(X_i|\mathbb{N}_i)$ could measure the precision with which X_i can be predicted by using the observations collected by the nodes linked to agent i. We denote by $\sigma^2_{i.\mathbb{C}_i}$ the estimation variance of X_i, based on the observations collected by the agents belonging to $\mathbb{C}_i$. Let us consider game G.S first. In such a case, we define the function $b_i(X_i|\mathbb{C}_i)$ as:

$$b_i(X_i|\mathbb{C}_i) = \log\left(\frac{1}{\sigma^2_{i.\mathbb{C}_i}}\right). \quad (2)$$

In such a case, step S.1 of the game forms a link if the following inequality is satisfied: $\sigma^2_{i.\mathbb{C}_i}/\sigma^2_{i.\mathbb{C}_i \cup j} > e^c$. Introducing the partial correlation coefficient [8], [9] $\rho_{ij.\mathbb{C}_i}$ between the variables i and j, conditioned to the set of variables $X_{\mathbb{C}_i}$, the test in step **S.1** can be rewritten as:

$$\rho^2_{ij.\mathbb{C}_i} > 1 - e^{-c}. \quad (3)$$

The meaning of the test is clear: The link ij is formed if the partial correlation between X_i and X_j, conditioned to the current neighbors of i, is sufficiently large. The higher is the cost, the larger the partial correlation must be. It is also straightforward to check that this test satisfies all properties P.1 $\div$ P.3. In fact, $\rho_{ij.\mathbb{C}_i} = 0$ if $\mathbb{C}_i \equiv \mathcal{N}_i$ (P.1) or if i and j belong to separated components (P.3); conversely, $\rho_{ij.\mathbb{C}_i} \neq 0$ if $j \in \mathcal{N}_i$.

Mutual information: Let us consider now the case where the goal of each agent is to maximize the additional information received when forming a new link, conditioned to the information that is already available. In this case, we can define the function $b_i(X_i|\mathbb{N}_i)$ as the entropy of X_i, conditioned to its neighbors. In the case of jointly Gaussian random variables, the inequality to be checked to decide whether to form a new link with agent j is

$$-\frac{1}{2}\log\left(1-\rho^2_{ij.\mathbb{C}_i}\right) > c \tag{4}$$

or, equivalently $\rho^2_{ij.\mathbb{C}_i} > 1 - e^{-2c}$. Interestingly, at least in the Gaussian case, this game has exactly the same form as the previous game and then it fulfills the same properties.

Numerical results: As a numerical test, we considered a sensor network composed of 6 agents observing a Gauss-Markov Random Field (GMRF). The covariance matrix is generated as a random definite positive symmetric matrix, having a sparse inverse. Fig. 1 shows the sum of prediction error variances, over all the network nodes, obtained by running games G.S and G.C, vs. the cost per link. The benchmark is given by the sum of the minimum variances obtained by a centralized system making predictions over all data (grey curve in the bottom). The blue line and the red stars represent the case of perfect knowledge of the covariance matrix. We can check a number of interesting results. First of all, the overall variance coincides with the social optimum when the cost is zero and increases as the cost increases, because some useful links are dropped. The value reaches an upper bound coinciding with the case in which the network becomes fully disconnected and each agent cannot benefit of any exogenous information, because of the high costs. Secondly, the performance of G.S and G.C are quite close to each other. However, it is worth clarifying that, for intermediate costs, the results in the two cases do not necessarily coincide. The more or less similarity is dictated by the structure of the covariance matrix. What really distinguishes games G.S and G.C is the overall cost paid to form the network, as reported in Fig. 2 showing the total cost vs. the cost per link. We can see in fact from Fig. 2 that, at low costs, game G.C pays a lower cost because it makes possible for every node to reach every other useful node through less direct links than game G.S. Finally, in Figs. 1 and 2 we report also what happens when the covariance matrix is unknown, but it is estimated from a finite set of N_s data. The presence of errors induces two kinds of error: an error in the final topology and an error in the value of the variance. To distinguish between these two errors, we report here only the effect

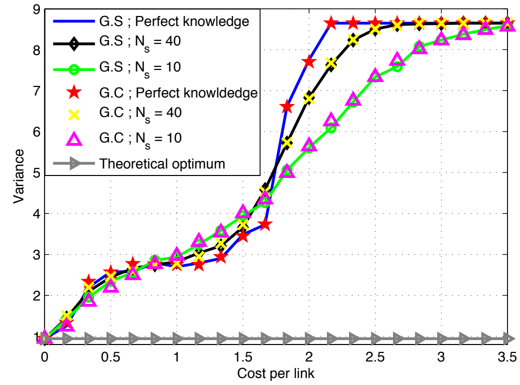

Fig. 1: Total estimation variance vs. cost.

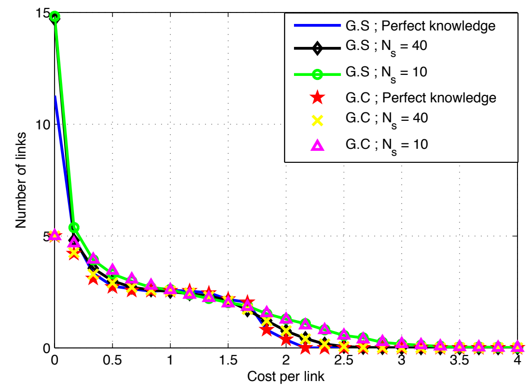

Fig. 2: Total cost vs. cost per link.

of errors in the network topology on the achievable variance. Interestingly, the results show that the presence of errors can even make the prediction variance lower, for intermediate cost values. This happens because, at low costs, in the presence of errors, agents tend to form more links than necessary, as the precision matrix will tend to be full. The price paid for such a behavior is a higher cost, as shown in Fig. 2.

5. CONCLUSION

In this paper we have proposed a network formation game where the link formation mechanism is based on the statistical properties of the observations available to each agent. The network formation game leads to a final topology that tends to match the independence graph, when the cost associated to each link is negligible, and to prune such a graph when the cost increases. Interestingly enough, in the presence of errors in the knowledge of the statistical parameters, the players inherently tend to form more links than necessary. This is indeed a robust behavior, paid by a higher cost.

6. REFERENCES

[1] M.O. Jackson and A. Wolinsky, "A strategic model of social and economic networks," *Journal of Economic Theory*, vol. 71, pp. 44–74, 1996.

[2] V. Bala and S. Goyal, "A noncooperative model of network formation," *Econometrica*, vol68, no. 5, pp. 1181–1229, 2000.

[3] M.O. Jackson and A. Watts, "The evolution of social and economic networks," *Journal of Economic Theory*, vol. 106, pp. 265–295, 2002.

[4] C. Ballester, A. Calvo-Armengol and Y. Zenou, "Who's who in networks. Wanted: The key player," *Econometrica*, vol. 74, no. 5, pp. 1403–1417, 2006.

[5] Y. Song and M. van der Schaar, "Dynamic Network Formation with Incomplete Information," in *Proc. of XVI Southwest Economic Theory Conference*, Irvine, March 2014.

[6] Y. Zhang and M. van der Schaar, "Information Production and Link Formation in Social Computing Systems," in *IEEE J. Sel. Areas Commun.*, Special issue on Economics of Communication Networks and Systems, vol. 30, no. 10, pp. 2136–2145, Dec. 2012.

[7] Y. Zhang and M. van der Schaar, "Strategic Networks: Information Dissemination and Link Formation Among Self-interested Agents," in *IEEE J. Sel. Areas Commun.* - Special issue on Network Science, vol. 31, no. 6, pp. 1115–1123, June 2013.

[8] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, John Wiley Sons, 1990.

[9] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol.2, Inference and relationship - 4th ed. London, Griffin, 1979.

[10] A. Anandkumar, L. Tong, A. Swami, "Detection of Gauss-Markov Random Fields With Nearest-Neighbor Dependency," *IEEE Trans. on Information Theory*, Vol. 55, no. 2, pp. 816–827, Feb. 2009.