Models for the diffusion of beliefs in social networks

Christophe Chamley[†], Anna Scaglione[*] and Lin Li[*]
Boston University, University of California Davis

I. INTRODUCTION

Signal processing is very much tied to extracting information and making inferences from physical phenomena. The traditional modalities, that our field is given credit for advancing, are speech, images, video, communication signals, remote sensing and a number of biomedical sensors that digital and array processing methods enable. More recently, Brain Machine Interfaces (BMI) have also become a research focus of signal processing researchers. We continue to fill the gap between human signals and computers, leading to today's highly computerized social landscape.

Controlling the medium through which people inform each other and communicate, opens the door to learn what they value and influence their beliefs, doing in an automated fashion what humans have been competing to do throughout history, gaining prestige to gain power. For the architects of these networks it is perhaps time to more closely consider their applications.

How do individuals influence each other decisions? Why leaders emerge in societies or fall out of favor? How do societies steer their beliefs and values? In the last century mathematical models developed in social sciences and economic theory attempted to capture macroscopic trends emerging from individuals communications and their decision making process.

There are excellent review articles and books that summarize the theories used to infer and interpret social swarming phenomena, and some have more closely inspired the models we review [1]–[3]. As we will discuss in Section II, the aim of these theories is to define a conceptual framework to examine them mathematically, to gain insights on what may happen under plausible simplifying assumptions. Often, these mathematical models are separated from the physical modalities of these interactions because the

The authors research was supported by NSF CCF-1011811.

emphasis is on the analysis of the mechanisms at work rather than the accurate quantitative prediction of results.

These highly stylized models, which often postulate simple mechanisms for the sake of mathematical tractability, can provide remarkable insights on the macroscopic trends observed in reality. In practice, the type and media used for social interactions depend on technologies humans have devised over the centuries to engage each other, facilitate transactions, share information, and last but not least, to make correct decisions. Society structures are very much influenced by the advent of these technologies, since they affect how easy it is to access what is useful, as well as to strategically place oneself to promote selfish interests. Humans invent new methods and forge relationship, alliances and institutions to support these technologies and in response to them. This is why a *link* in 1400th century Florence is not quite the same as a link between humans, social groups or corporations today. These technologies, in order to work, have to be adopted collectively. Hence, ultimately, their success depends on humans embracing them to pursue their selfish purposes more efficiently, accurately and effectively.

What makes this time especially ripe for new theoretical ideas is that in the digital age, not only interactions happen at lightning speed, but the information about them is efficiently encoded in bits, which also leave a long lasting digital trace behind. This trace can be stored and replicated at very modest cost. A dominant position in social media amounts to have a dominant position in influencing what people value. Managing these digital interaction, not surprisingly, is the technological race that is deciding the fortunes and the dismay of Internet companies. They are competing for becoming the central carrier of information useful to the large majority of humans, so as to gain information about those transactions, that would allow them, in turn, to influence the largest number of people. A more recent development is that these information carriers today follow and sense us in many moments of our life, through our Smartphones. As discussed in Section V, technologies are being developed and perfected to provide accurate information about individuals and social contexts. Individuals release information in exchange for receiving help with their decision making, and are lured to do so by a portfolio of tools that is hard to resist even by the most private person.

Compared to the many excellent overviews that exist in this field, from books [1]–[3] to recent comprehensive overviews [4], a goal of this paper is to compare basic models used in economics to gain insight on social behavior with information fusion problems that have been widely investigated in our field. We first review, in Section II, basic models that have been used in economics to capture opinion diffusion among rational agents. We contrast these findings to similar sensor fusion problems studied in information theory and in signal processing, where the limitations in the learning performance are the consequence of physical constraints in communicating, rather than self-interest. In Section III, we discuss a more general information sharing model, where agents are allowed to have interactions over an arbitrary network. In the same section, we draw analogies between these theories and studies on network diffusion algorithms in signal processing. In Section IV-A, we briefly review the analysis of network formation under strategic models, motivating the emergence of communication graph between self-interested individuals with specific utilities that balance the costs and benefits of sharing information with others. The last section of the paper suggests possible gaps in understanding how to more directly use and directly value the wealth of information that computer networks are harvesting about humans economic decision systems for economic analysis.

II. Opinion diffusion among rational agents

Each individual's opinion about a fact (weather, weight of an animal, suitability of a political candidate) is in general subject to errors. In the averaging of individuals' opinions, errors cancel each other and by the law of large numbers, the average opinion of a large number of individuals is a much more precise estimator. This property was well recognized already in 1785 when the marquis de Condorcet in [5] advocated the virtues of voting by a large number of people, and of democracy for the aggregation of individual informations. The insight of Condorcet in the aggregation of the information of a large number of people motivated, in 1907, the article "Vox populi" by Francis Galton [6], who provided a colorful demonstration, averaging the guesses of several people on the weight of an ox in a market and coming remarkably close to its true value, thus demonstrating the concept that the averaging of opinions leads to accurate estimates.

Condorcet was also well aware about the perturbation that are induced by the observation of others' opinions or actions on that process of aggregation. Taking the example of voting, there is a critical difference between voting by individuals who just express their preference based on their individual information and individuals who observe the votes of others in a process of sequential decisions (*e.g.*, votes), and are therefore influenced by their observation before choosing their action. That difference is key for the mechanisms of social learning.

Modeling mathematically these phenomena includes basic ingredients that are familiar to researchers in signal processing. Assume that the social learning is about a number θ that is the realization of a random variable in a set Θ. That set can have a finite number of element, say 0 for "sunshine" and 1 for "rain", or a number in a continuum as the weight of an ox, as in the experiment of Galton. The agents involved in the social learning process have a mental model relating their evidence to the unknown parameter θ.

Let us assume that this evidence takes the form of a signal s_i. In the experiment of Galton, the signal is the guess of an observer the ox, as based solely on the sight of the ox.

The action of the agent is denoted by a, in a set $\mathcal{A}$. That set can have a finite number of elements or be a continuum. When the agent takes or does not carry an umbrella, the set has two elements. When the agent expresses an opinion about the weight of an ox, $\mathcal{A}$ is the continuum of real positive numbers. The opinion expressed by an agent is an action that when observed by other agents, carries a signal that feeds into the process of social learning. That message may be his information but in general that is not the case: the action may be motivated by other factors besides the private information, that action may be restricted to a limited set of possibilities or its observation by others may be garbled. The literature has investigated a number of structures and has focused on the relation between these structures and whether social learning converges to the true value of θ, and if so, how fast.

The standard model is the sequential model with complete recall where agents are put in an exogenous sequence. Each agent's information is his private information and the *history* of past actions by agents who preceded him in the sequence. Agents are rational in two senses: (i) in the formulation of their *belief* on θ (*i.e.,* their subjective probability distribution on θ), they use Bayes' rule in combining their private information with the public information of the history of past actions; (ii) they chose the action that maximizes the expected value, under their belief, of a payoff function that depends only on the true value θ, their own action, and some individual characteristics.

Define here a belief as the the probability that an event occurs, (*e.g.* θ is in some set Θ). As in statistical signal processing models, a core property of rational learning in the standard sequential model is that the public belief is a martingale. A belief on an event can be viewed as the price of a stock, or a bet, that delivers one dollar if the event eventually turns out to occur. The price changes as new information comes in, but the expected value of the price change is zero. As in efficient financial markets, if that price would be expected to, say, increase, agents would readjust upwards the price today. That relation between rational learning, martingales and efficient financial market actually provides the key idea for a proof of the convergence theorem of bounded martingales. This theorem implies that in the standard social learning among rational agents, there cannot be perpetual random cycles of beliefs. That property sets the rational learning apart from other processes such as adaptive learning. Beliefs converge, but they may converge to a wrong value, or they may converge to the truth but the convergence may be slow.

Two generic models of individual information have been used with particular success in the literature. The first is the Gaussian model in which the prior distribution of θ is Gaussian $\theta \sim \mathcal{N}(\theta_0, \sigma_0^2)$, and each individual signal is the true value θ plus an additive Gaussian noise (AGN) $s_i \sim \mathcal{N}(\theta, \sigma_i^2)$. That model is appropriate in some context such as the distribution of the ox weight, even if it includes negative outcomes, assuming that σ_0 and σ_i are small with respect to θ_0. The model conveniently generates a simple linear learning rule, but we should be aware of at least two important additional features: first any new signal reduces the variance of the distribution of θ and thus the uncertainty about θ. Second, private signals are unbounded. In this way, whatever the precision of public information, there are some private signals that can "overrule" the consensus. These features play a critical role in social learning. The first one implies that the as the variance of the public distribution of θ increases, individuals who know that distribution reduce the weight of their own private information for their action and that action is therefore less informative for others. This property slows down the process of social learning. The second feature implies that an agent's action always carries an information (that may be vanishingly small) to others. It follows that the social learning is never "stuck" and converges (possibly slowly) to the true value of θ_0.

In the second model of private information, private signals take values in a finite set. Without loss of generality, one may assume the signals to be binary with value 0 and 1. For example if there are two values of θ, 0 and 1, we may have $P(s_i = \theta|\theta) = p$, where p is a parameter that measures the precision of the signal. Such a signal is bounded. Moreover, the variance is not monotonically decreasing with the number of signals. For example, if the probability that $\theta = 1$ is near 1, a signal 0 reduces that probability and therefore increases the variance of the belief and the uncertainty of agents.

These two models can obviously be expanded, depending on setting, and their use should be adapted to the context. In the experiment of Galton, the first model is relevant. The second model is more appropriate when agents make a zero-one decision (*e.g.* taking out an umbrella depending on the probability of rain), or when new information may increase the variance of the public belief. Because the Gaussian model has been so successful both in economics and signal processing, we discuss it first in the context where agents reveal their private information. We will consider binary signals later.

A. The Gaussian model

We assume here that in the Gaussian model that was previously introduced, the messages of agents are just their signals: they speak the truth. We relax this assumption later. The model is illustrated in Fig. 1.

In this model, it is well known that the optimal fusion rule is a weighted average. Specifically, let $s = (s_1, \dots, s_n)$ and $\gamma_i = \sigma_0^2/\sigma_i^2$ be the signal to noise ratio of the signal at node i. The Minimum Mean

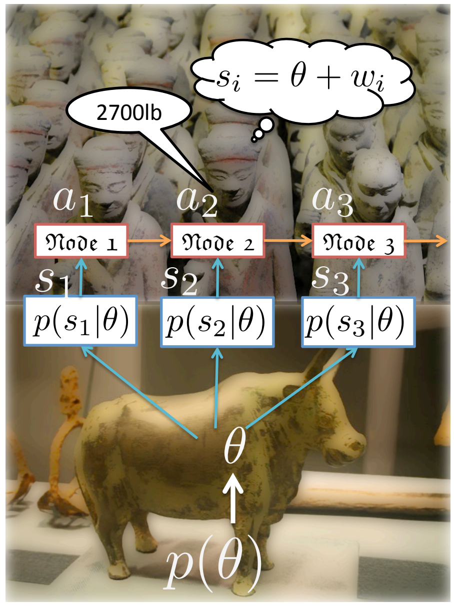

Fig. 1: A group of social agents processing information about the weight of an ox.

Squared Error (MMSE) estimate can be expressed as

$$E\{\theta|s\} = \frac{\sum_{i=1}^{n} \gamma_i s_i}{\sum_{k=0}^{n} \gamma_k}.$$

To see how the error variance decreases as n increases, one can show that the MMSE reduction at the nth stage is measured by the following factor:

$$\rho_n = \left(\sum_{k=0}^{n} \gamma_k \right)^{-1} = \frac{\rho_{n-1}}{1 + \gamma_n \rho_{n-1}}.$$

In fact, the simple weighted averaging procedure that leads to the calculation of $E\{\theta|s\}$ provides estimates with an error variance that is $MMSE_n = \rho_n \sigma_0^2$, at the fraction $\rho_n < 1$ of the original variance σ_0^2. The accuracy of the estimate $\theta_n^\star = E\{\theta|s\}$ grows therefore linearly with the number of estimates that are averaged. This estimate can be updated recursively, and each iteration can be viewed as a message passed from one social agent to the rest of the group. Let $h_n = (a_1, \ldots, a_n)$ the *history* vector of the actions a_i, assuming that the agents indexes correspond to the order of their actions. Denoting by $\tau_n = \rho_n / \rho_{n-1}$, with simple algebra one can show that the estimate (and optimal action) of the nth agent is the convex combination $\theta_n^\star = (1 - \tau_n) s_n + \tau_n \theta_{n-1}^\star$. After the nth guess, the belief model $p_n(\theta|h_{n-1}, s_n)$ is therefore updated to be $\theta \sim \mathcal{N}(\theta_n^\star, \sigma_0^2 \rho_n)$. An interesting aspect of this basic model is that agents do not need to see the history of all the previous actions, i.e., they only need to receive the estimate from the previous agent, as shown in the serial communication model in Fig. 1. In the sequential rule, the coefficient τ_n, representing the MMSE improvement resulting from the update, tends asymptotically to one, and it can be interpreted as the amount of trust placed on the historical aggregate estimate compared to the private information added in the mix; this phenomenon models how new information is trusted less and less, capturing (and justifying) the opinion hardening of the experienced old-timers and the coefficient τ_n can be viewed as the trust coefficient that the node places on the history compared to its own private information. This illustrative example is especially helpful to grasp how powerful social learning can be in mathematical terms and also to understand the strong tie that exists between economic modeling and estimation theory principles that are also the foundations of signal processing. The main difference between economic and signal processing modeling lies in what limits this ideal learning scenario.

In fact, in signal processing the ideal $\text{MMSE}_n = \rho_n \sigma_0^2$ is unattainable due to technical limitations, such as the finite precision of computers and limited communication resources that are available to acquire the data. For example, the finite precision in the representation of sensors information s_i through quantized values has been the subject of intense investigation in both signal processing (see e.g. [7]) and information theory, where the problem is referred to as the *multi-terminal source coding problem* [8]. In [9] the authors considered a sequential decoding problem resembling the Gaussian model discussed above and illustrated in Fig. 1. As customary in information theoretic analyses, the underlying assumption that the nodes are exposed to infinitely long sequences of independent private observations regarding an independent and identically distributed (i.i.d.) sequence of θ (in practice this works as a bound for the case of a single guess). For example, the people in the market need to guess the weight of a large number of oxes and now their guesses have to be written in a card, so that a limited number of digits can be allotted to describe each guess. Let $\rho_n = MMSE_n/\sigma_0^2$ now represent the MSE reduction when the bits encoding the $n-1$ signals are strictly not greater than R_{n-1} per ox weight. If the previous $n-1$ agent can communicate only to its immediate neighbor (this is the case of *serial network*) and the message can contain up to R_{n-1} bits per ox weight, then the MSE reduction factor $\rho_n = MMSE_n/\sigma_0^2$ obeys the following recursive equation [9]:

$$\rho_n = \frac{\rho_{n-1}}{1+\gamma_n \rho_{n-1}} + \frac{\sigma_0^2}{1+\gamma_n}\left(\frac{1-\rho_{n-1}}{\gamma_n+\rho_{n-1}}\right)2^{-R_{n-1}}.$$

It is interesting to compare this with the original MSE reduction factor with uncoded transmission, which is identical as $R_{n-1} \to \infty$. It is not difficult to verify that, unless the rates are allowed to grow to infinity (at approximately a logarithmic rate with n), the distortion will no longer improve as $O(1/n)$. Losses arise also if the agents have access to all prior history of *actions* which in this case are the digital codes produced by the previous participants. This corresponds to the setup called *parallel network* in [9] where the shared knowledge can be viewed as that of an information *hub*. In this case:

$$\rho_n = \frac{\rho_{n-1}}{1+\gamma_n \rho_{n-1}} + \left(\frac{\sigma_0^2 \rho_{n-1}^2}{\gamma_n + \rho_{n-1}}\right) 2^{-R_{n-1}}.$$

Asymptotically, if all guesses have to be stored in a finite computer memory, with a total amount $R_{tot} = \sum_{i=1}^{n} R_i$ of bits available per guess then, as $n \to \infty$ the agents will have a vanishing rate allowed to submit each guess and the attainable ρ_n in the limit becomes the solution of the equation:

$$R_{tot} = \frac{n}{2\sigma_0^2}\left(\frac{1}{\rho} - 1\right) - \frac{1}{2}\log \rho.$$

The information theoretic bound for the attainable accuracy under rate constraints may be relevant in the future when collecting the wisdom of crowds in digital form in computer clouds, will be faced with storage limits. These models can capture actual limits for wide organizations that are required to accrue information over a dispersed area through the support of information technology in real time. In this case, rate bounds can be directly tied to the power expenditure necessary to communicate reliably R_n bits of information per second, which in a contention free channel affected exclusively by thermal additive noise is given by the celebrated Shannon formula $R_n < W_n \log(1 + \frac{P_n}{N_0 W_n})$.

While the cost of digital telemetry, digital communications and computation has been at the center-stage in the literature on sensor networking, in the vast majority of social sciences, the analysis has been typically divorced from technological limitations, yet losses compared to the optimal estimate are the norm rather than the exception. Vives in [10] included a cost for the quality of private information explicitly in his analysis, that rises for s_i that are less noisy. However, nobody accounted for the cost of sharing information with others in digital form, which is an interesting issue to consider as high-speed trading communities increasingly rely on computer networks to share information and compute actions. There is a relatively recent trend in computer science [11] that focuses on recasting game theoretic analyses under computation constraints which leaves out communication constraints. The next section reviews a model that has been used in economics to capture the lack of perfect information about the other agents beliefs.

B. Private efficiency and social inefficiency in social learning

Imperfections in social network models typically arise because the agents act selfishly and do not take into account the benefit of their private information for others. Suppose that the optimal investment level of a firm depends on the future price θ and that, with a proper normalization, the loss function of the firm

n that invests a_i is $-E_i\{(a_i - \theta - \eta_i)^2\}$, where the expectation E_i is computed according to the firm's belief about θ, and η_i is a parameter that is specific to the firm and cannot be observed by others. That representation of the payoff of the firm captures the general property that one cannot observe exactly the decision process of agents.

In the Gaussian model, the firm chooses $a_i = E_i\{\theta\} + \eta_i$ where $E_i\{\theta\}$ is computed as in the previous section and is a linear function of s_i with parameters that are publicly known. The action a_i is a message on s_i that is garbled by the idiosyncratic noise η_i. That simple model that is derived from Vives in [10] enables us to understand a key issue in the inefficiency of social learning. Assume first that each agent chooses his action in ignorance of others. His action reveals his private information with a noise. When an observer of history aggregates the information that is provided by the observation of the actions of n agents, the precision of his belief on θ (the inverse of the variance), increases like nA where A is inversely related to the variance of the observation noise η_i. Assume now that each agent i observes the actions of all agents j with $j < i$. In that case, Vives has shown that the precision of the belief from history increases with n like $n^{1/3}$. In other terms, the convergence of the variance toward 0 is much slower in the process of social learning, like $1/n^{1/3}$, instead of $1/n$ when agents ignore others. The slow learning when individuals observe others has a simple interpretation. In the determination of $E_i\{\theta\}$, agent i takes a linear combination between the estimate from history and the estimate from his private information, as we have seen previously. The weight of history grows with n since there is more information and the weight of his private signal s_i decreases. Because of the noise η_i that has a constant variance, the impact of the private signal becomes gradually dwarfed by the noise and the action a_i becomes gradually less informative about the signal s_i. The process is equivalent to an increase of the variance of η_i relative to a constant impact of s_i on a_i.

In the present model, selfish agents do not take into account the information that their action convey to others. Simple intuition shows that a small subsidy that induces agents to amplify their difference with the consensus belief is Pareto efficient. That subsidy has only a second-order welfare impact on each agent i who is at his optimum but it has a first-order effect on the welfare the agents j with $j > i$.

When the noise η_i vanishes or when the decision model about agents is perfectly known, there is no inefficiency in social learning although agents take an action that is different from their belief. A rational observer who knows perfectly the decision model of the self-interested agent is able to recover perfectly the private information of the agent. There is no informational difference between the message of the action and the private information of the agent.

The inefficiency of social learning takes a remarkable form in the present model but that inefficiency is generic in all processes of social learning. Selfish agents who observe others use that information and tend to imitate others, rationally. That imitation tends to reduce the information that they convey to others by their own action.

Social learning converges here slowly, but it converges to the truth because individual signals are Gaussian and unbounded. There is always some probability that some agents have a signal with a sufficiently high absolute value to convey a message on his private information. In models where signals are discrete, such as in the binary model, private signals are bounded and social learning may stop altogether in finite time. The variance of the public belief on θ may remained bounded below by a strictly positive number.

C. Pathologies in social learning: informational cascades and herds

In the previous discussion, the imperfection of the individuals' messages arises from the observation noise on the actions of others. A particularly strong form of message imperfection is the restriction of the action set $\mathcal{A}$ to discrete values. The communication between agents is restricted as if they could speak only two words. Even under such restriction, one can show that when agents are informed by an unbounded signal, such as for example the case where s_i is the output of an additive Gaussian noise channel, social learning while slow, still converges to the truth. However, when agents are privately informed by a bounded signal, such as for example the signal emerging from a discrete memoryless binary channel, then social learning may not converge to the truth. More remarkably, it may converge to a distribution on θ such that after a finite time, agents who choose between two actions choose the inferior one given the true value of θ. That is the celebrated property of the *informational cascades* of Bikhchandani, Hirshleifer and Welch (BHW) [12] (see also [13]). The property is easy to explain.

In the BHW model, the parameter θ is randomly determined between the values of 1 ("good state") and 0 ("bad state"). The utility of each agent is $u_i = (\theta - c)a_i$, with $0 < c < 1$ and $a_i \in [0,1]$. The utility mathematically captures the fact that a profit $1 - c$ is realized only if $\theta = 1$ and the action $a_i = 1$ predicts this outcome correctly, an action $a_i = 0$ incurs in no cost and no benefit either way, while if the agent decides in error $a_i = 1$ when $\theta = 0$ the utility is negative, and equal to $-c$. Without loss of generality, take $c = 1/2$. Private signals are such that $P(s_i = \theta|\theta) = q$. Let μ_n the probability of the good state at the beginning of round n, conditional on the history of past actions. From Bayes' rule, if $\mu_n > q$, agent n takes the action $a_n = 1$ even if he has a bad signal ($s_n = 0$): the weight of history dominates his private signal. In that case, his private signal has no impact on his action and that action reveals no information. The public belief remains unchanged and the process repeats itself for all future agents.

There is an informational cascade and all agents take the same action, in this case invest, even if the true state is bad. Such a situation must occur in finite time because of the martingale convergence theorem: μ_n is a martingale and must converge but it cannot converge in the interval $(q, 1-q)$ because in such an interval, private actions are signal revealing and make μ_n jump. One can exclude the convergence to q or $1-q$ for almost all values of (μ_0, q). Therefore, the limit is outside of $(q, 1-q)$ in which case an informational cascade takes place. That limit is reached in a finite number of steps because of the discrete Bayes updating.

In an informational cascade, all agents take the same action and a *herd* must take place. The reverse proposition is not true, as emphasized by Smith and Sørensen (2000) [14]. Modify the BHW model such that agents have a Gaussian signal $s_i = \theta + w_i$ where $\theta \in \{0, 1\}$. Private signals are now unbounded. The public belief μ_n which is a martingale, converges and the limit cannot be in the interval $(0, 1)$, following an argument that is similar to the case with binary signals. One can show that social learning converges to the truth and μ_n converges to θ. However, with probability 1, a herd eventually takes place in finite time! Because of the martingale convergence theorem, one cannot have an infinite subsequence of "contrarian agents" who act against the consensus: such agents by going against the consensus reveal that there information is stronger than the consensus and thus induce a jump in the consensus. Such jumps cannot occur an infinite number of times because of the convergence of the martingale.

It is interesting to compare the information cascades, with the optimal rate of learning and with the rate of learning that characterizes a set of sensors that shares information with and without communication limitations. The interesting question is what happens when sensors can share only a binary decision, emulating the scenario in the BHW model of rational agents that are limited to perform a binary action; in the context of engineering design the goal of the action is to bear as much useful information as possible to a fusion center, which collects the common knowledge, or to the next sensor.

In this case the engineering literature on the topic of decentralized detection is relevant, and the interesting results that emerge are that the learning rate decreases but does not stop to grow, unlike the BHW model. This problem was first posed by Tenney and Sandell in '81 [15] and their results were later generalized in [16] for a parallel architecture with a fusion center and in [17] for a serial network configuration. In the parallel configuration [16], each one of the n agents evaluates a finite-valued function (also known as the decision rule) $a_i \in \{1, \cdots, D\}$ of its private observation s_i using a likelihood ratio test and then transmits it to the fusion center. Based on these incomplete information $\{a_1, \cdots, a_n\}$, the fusion center determines one true state out of the two hypotheses using the maximum a posteriori (MAP) rule, as illustrated in Fig. 2. The objective of the fusion center is to minimize the error probability, under the assumption that the private observations are conditionally i.i.d, with a known conditional distribution $p(s_i|\theta = j)$, $j = 0, 1$. For the case when $D = 2$, it has been shown that there exists an optimal decision rule such that the probability of error decays exponentially fast with the number of the agents n. Hence, given the log-likelihood ratio $\lambda_i = \frac{p(s_i|\theta=1)}{p(s_i|\theta=0)}$ and the optimal threshold τ, it is asymptotically optimal to have all the agents use the same decision rule, in the form

$$a_i = \begin{cases} 0, & \text{if } \lambda_i \leq \tau \\ 1, & \text{otherwise} \end{cases}$$

Interestingly, when there are $M > 2$ hypotheses, a single decision rule for all the agents will no longer lead to an asymptotically exponential rate of learning. In fact, the optimal learning is attained by having at most $M(M-1)/2$ number of distinct decision rules [16]. Moreover, a tree structure for aggregation of sensor decisions is studied in [16].

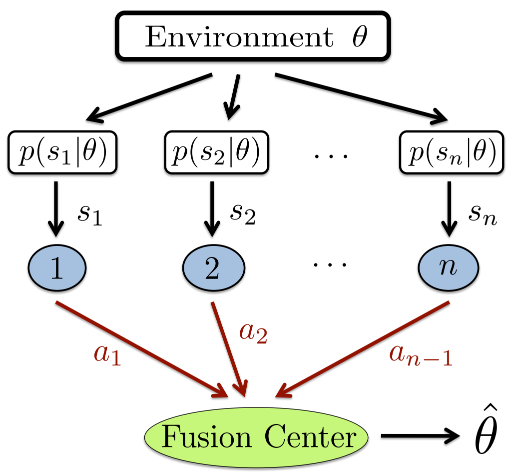

Fig. 2: Parallel Configuration

Consider now the case when the agents are in a serial network configuration (often known as the tandem network) [17]. Suppose that the agents are indexed consecutively with $i = 1$ being the first agent and $i = n$ being the last agent. Each agent observes a random signal s_i and evaluates a binary decision $a_i \in \{0, 1\}$ using a likelihood ratio test. The decision is then passed to the next agent, but the history of previous decisions $h_{i+1} = (a_1, \cdots, a_{i-1})$ is not visible (see Fig. 3). Again, the goal is to design an optimal strategy that minimizes the probability of error. Under the assumption that the observations s_i are conditionally independent, it has been found that the probability of error is bounded away from zero if and only if the absolute log-likelihood is bounded almost surely. In contrast, when the log-likelihood is unbounded, it was proved in [17] that there exists a set of decision rules such that yields optimal rate of learning. Moreover, the set of optimal decision rules takes the form:

$$a_i = \begin{cases} 0, & \text{if } \lambda_i \leq \tau_{i,n}(1) \\ 1, & \text{if } \lambda_i > \tau_{i,n}(0) \\ a_{i-1}, & \text{otherwise} \end{cases}$$

where $\tau_{i,n}(1) \leq \tau_{i,n}(0)$ are the thresholds whose values depend on a_{i-1}, i.e., the decision of the previous agent and are given as a result of a rather complicated optimization. There are two interesting observations to make. First, notice from the construction of the decision rules that it may be optimal for the agent to simply copy his neighbor action/decision. Second, the rate of learning might stop, if the log-likelihood is bounded almost surely, which is similar to the herding phenomenon observed in the BHW model. In general, the rate of learning is much slower than exponential. In fact, authors in [17] proved that it is always sub-exponential.

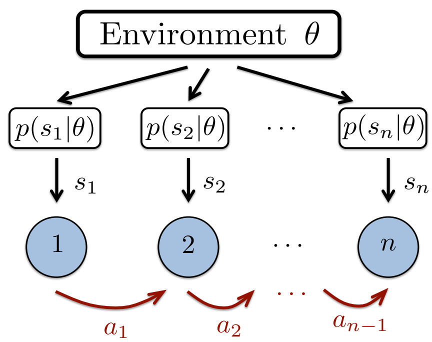

Fig. 3: Tandem Network

Another interesting insight is that the key difference between the decentralized detection and the BHW model is the driving force behind every decision that each agent makes (or the action that each agent takes). In the BHW model, agents act selfishly in the sense that they do not make collective decisions on behalf of the group because every agent is associated to an specific utility function which need to be maximized when making a decision. In contrast, in the decentralized detection models, agents in fact work together to achieve a common goal, that is, to maximize the rate of learning based on their collective knowledge. Nevertheless, lack of learning in the serial network example may occur, irrespective of the agents best intentions toward each other.

III. Opinion diffusion in general networks

Social learning is hardly obtained by accumulating all historical values sequentially. Agreement often emerges organically, by averaging opinions among subsets of neighbors [18], [19]. Their selection is often random and opportunistic and agents act multiple times, often in asynchronous random events. In fact, in recent years the analysis of social phenomena has increasingly emphasized all these aspects and, among them, special attention has been placed in understanding the role of the network structure as the scaffold that supports exchanges of information, rewards and costs. What we will discuss in this section are models for the formation of these networks, but for the time being let us assume that the limitations that exist in communicating with peers are given and that we are interested in analyzing their effect on the process of opinion formation.

Smith and Sørensen (2008) [20] analyze limited memory in the sequential model by assuming that agents sample randomly a limited number of past observations. The analysis is much more difficult than in the model with complete recall because the history of actions does not generate a public belief that evolves as a martingale. The boundedness of private information has a strong impact on the properties of the social learning. It is intuitive that the weight of history increases with the sample size. Hence, quite remarkably, welfare may fall in the sample size. More recent studies can be found in [4], who found conditions where social learning can happen in unstructured communication settings.

A standard method to capture the limitations of observations is to represent the social network as a graph $G = (E, V)$ where agents are the collection of n vertexes V and relationships are represented through a set of edges E. Node i is can only exchange information with a subset $\mathcal{N}_i \subset V$ and the cardinality $d_i = |\mathcal{N}_i|$ is called the node degree. The connectivity of the graph is often represented through an $n \times n$ adjacency matrix A whose elements $A_{ij} = 1$ iff $ij \in E$. Other useful matrices associated to the graph G are the incidence matrix and the Laplacian matrix $L = diag(d_1, \dots d_n) - A$.

It is not difficult to recognize that the interest is in characterizing how estimates emerge and coalesce in networks is parallel to the research trend in signal processing focusing on Consensus Gossiping (CG) algorithms [21]. The motivation behind the research on CG algorithm is that sensor deployments are vast and irregular. Rather than trying to forward information to a fusion center, researchers focused on providing means to compute these queries via near-neighbors communications. This is not unlike human societies that have created money so that information about the value of goods and their best allocation can emerge through swarming of self-interested agents, rather than from the control of a central market operator. CG algorithms offer similar advantages in sensor networking. In local communications normally messages are assumed to have infinite precision, but there is a fixed cost per message to account for and links may have also an associated cost.

One of the most widely studied algorithms is Average Consensus Gossiping (ACG), whose goal is to compute and distribute the value of the average of a certain initial state variable $x_{i,0}, i = 1, \dots, n$ among all agents. We discuss some of its main characteristics that we will see later relate closely to the analysis of opinion diffusion in arbitrary networks introduced by DeGroot in [22].

In its simplest instantiation, ACG attains this goal by having each node update its local state variable as the following convex combination of its local state and of its neighbors states:

$$x_{i,k} = x_{i,k-1} - \epsilon \sum_{j \in \mathcal{N}_i} (x_{i,k-1} - x_{j,k-1}),$$

where $0 < \epsilon < 1/\max(d_i)$. If the network has a path connecting every two agents, asymptotically the entire network converges to $x_{i,\infty} = \frac{1}{n}\sum_{i=1}^{n} x_{i,0}$ which is the average of all initial states. Stacking all the variables in one vector $x_k = (x_{1,k}, \ldots, x_{n,k})$ one can see that these interactions correspond to perform the linear mapping $x_k = (I - \epsilon L)x_{k-1}$ of the sensors states, where L is the graph Laplacian introduced before. The graph Laplacian defined above is symmetric and has always a null vector represented by the constant vector $\mathbf{1}$. It can be shown that the matrix $I - \epsilon L - \mathbf{1}\mathbf{1}^T/n$ is a contraction. Hence, at each iteration the accuracy of ACG improves exponentially, with a rate that is equal to $1 - \epsilon\lambda_2(L)$, where $\lambda_2(L)$ is the second smallest eigenvalue of the graph Laplacian L. This simple analysis relates directly the ability to attain consensus in the graph to a specific measure of connectivity that best capture the rate of diffusion on G. In fact, the quantity $\lambda_2(L)$ is the so called Fiedler eigenvalue, or algebraic connectivity of the graph, that has central importance in spectral graph theory [21], [23]. Furthermore, the nullity $\nu(L)$ represents the number of connected components in the graph. Interestingly, if $p = \nu(L) > 1$ then the network will be partitioned in p groups in consensus.

Considering again Galton market experiment, we indicated that the agents estimating the weight of the ox in the market could have been gossiping about their estimates with a near neighbor, weighting differently the opinions they express depending on how trustworthy they were. We showed that the optimal fusion rule is equivalent to the following convex combination of the current estimate with the local private information $\theta^\star_n = (1 - \tau_n)s_n + \tau_n\theta^\star_{n-1}$ and also indicated that, as the sequence of actions gets longer, the trust placed on the private information diminishes.

To capture these effects in a more general social graph, a popular model proposed by DeGroot in [22] associates to $G = (E, V)$ an $n \times n$ (doubly stochastic) trust matrix T, where the ij element $\tau_{ij} > 0$ if $ij \in V$ and is zero otherwise. The elements τ_{ij} in [22] are part of the model of the agent opinion formation process. In this case agents average their estimate with others repeatedly, forming at every round of gossip k a convex combination of their respective opinions. The model generically refers to the quantities that are averaged as beliefs $x_{i,k}$. At each iteration, node i updates his belief as:

$$x_{i,k} = \sum_{j \in \mathcal{N}_i} \tau_{ij} x_{i,k-1}.$$

For the specific choice of $T = I - \epsilon L$ the DeGroot model is equivalent to the ACG discussed above and, hence, the reader should have no trouble guessing how the analysis goes: if the eigenvalue $\lambda(T) = 1$ has multiplicity p then there will be p groups in agreement. For $p = 1$ the entire group can obtain an estimate equal to $\theta_n^\star$ asymptotically, and the rate of convergence is the spectral radius $\rho(T - \mathbf{1}\mathbf{1}^T/n)$ which for $T = I - \epsilon L$ is equivalent to $1 - \epsilon\lambda_2(L)$.

One can immediately note that the DeGroot model analysis is not concerned with establishing how close the agents will come to the right decision but rather if they all will eventually agree (see Fig. 4 for an illustration). Also, the agents are exchanging directly their belief variables. In this sense, the framework of Bayesian rational agents driven by selfish intents is lost.

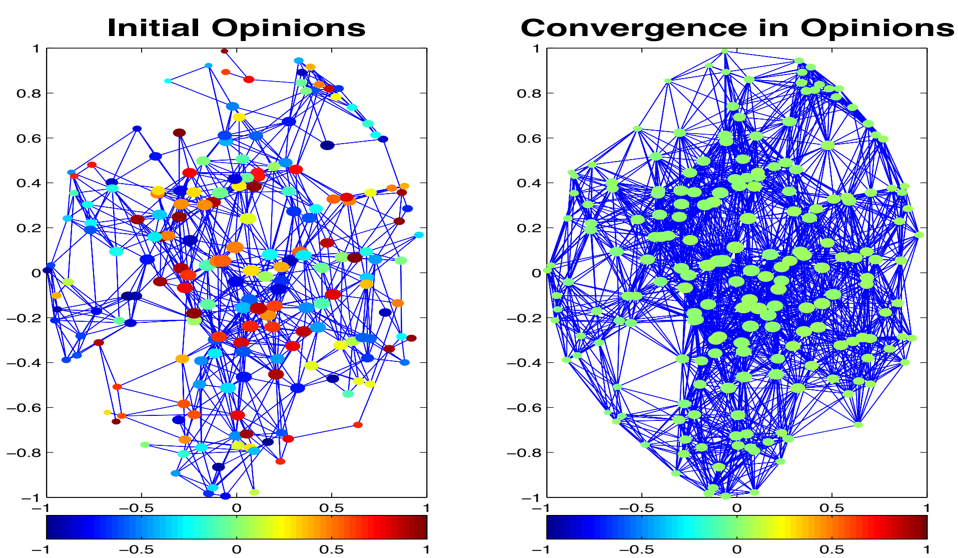

Fig. 4: Each color on the random graph nodes represents a different belief value. On the right side we see all nodes with the same color after 300 interactions illustrating the convergence in opinion using DeGroot Model.

There are many interesting variations of the DeGroot model. Among them worth mentioning are the so called *bounded confidence models*. The *Hegselmann-Krause* (HK) model [24], [25] is perhaps the closest one. Opinions are probabilities $x_{i,k} = P(\theta = 1|h_{i,k}, s_i)$ and are updated synchronously, similar to the DeGroot model, through a convex combination of the neighbors beliefs. However, the HK model introduces a confidence level (or a threshold) τ to model the lack of influence among agents whose beliefs are too distant, which is captured by modifying the trust matrix as function of the network state $T(x_k)$ to exclude every edge ij such that $|x_{i,k} - x_{j,k}| > \tau$. Specifically, let $\mathcal{T}_{i,k} = \{j \in \mathcal{N}_i : \ |x_{i,k} - x_{jk}| \leq \tau\}$ be the set of neighbors whose absolute opinion distances are less than τ after the k^{th} interaction, i.e. the set of *trusted neighbors* plus the node itself. Let $|\mathcal{T}_{i,k}| \geq 1$ be its cardinality. Then, after exchanging beliefs with their neighbors, individual opinions are updated synchronously according to the following rule:

$$x_{i,k} = \frac{1}{|\mathcal{T}_{i,k}|} \sum_{j \in \mathcal{T}_{i,k}} x_{j,k-1}.$$

If at any point $|\mathcal{T}_{i,k}| = 1$ then node i has no trustworthy neighbors and maintains his or her own opinion. Though similar to the DeGroot model, a significant difference is that the trust matrix in the HK model depends on the state and, therefore, the dynamics are non-linear. In particular, the set of neighbors with which agent i updates its belief may change with k.

An extension of the HK model introduced in [24] includes asymmetric confidence intervals $[-\tau_l, \tau_r]$ and another related extension, called the heterogeneous HK (htHK) model [26], introduces different confidence bounds $\bar{\tau} = [\tau_1, \cdots, \tau_n]$.

A randomized version of these models, which resembles asynchronous ACG algorithms is by *Deffuant-Weisbuch*. Their model, introduced in [27], [28], analyzes random pair-wise interactions between agents whose opinion distance is smaller than a threshold, instead of considering a synchronous update of all the state at each iteration, i.e.

$$\begin{aligned} x_{i,k} &= x_{i,k-1} + \mu(x_{j,k-1} - x_{i,k-1}) \qquad (1) \\ x_{j,k} &= x_{j,k-1} + \mu(x_{i,k-1} - x_{j,k-1}), \end{aligned}$$

where $\mu \in (0, 0.5]$ is called the mixing parameter. The interesting aspect of all bounded confidence models is that they exhibit a phase transition from herding to polarized opinions. This is illustrated in Fig. 5 and Fig. 6. In the first figure on the left we can see that over time two distinct belief emerge and they are held by two groups of agents, while on the right a sample of the same initial random network model converges to a consensus in belief. The only difference between the two is the value of the threshold that defines the bounded confidence model. The second figure (Fig. 6) shows numerically that the transition from consensus to a society split in groups that are polarized occurs at a certain value of the bounded confidence threshold τ, with high probability. One can observe from more extensive numerical tests that the phase transition is sensitive to the distribution of the initial beliefs but it is rather insensitive to the actual topology of the graph G, as long as it is connected.

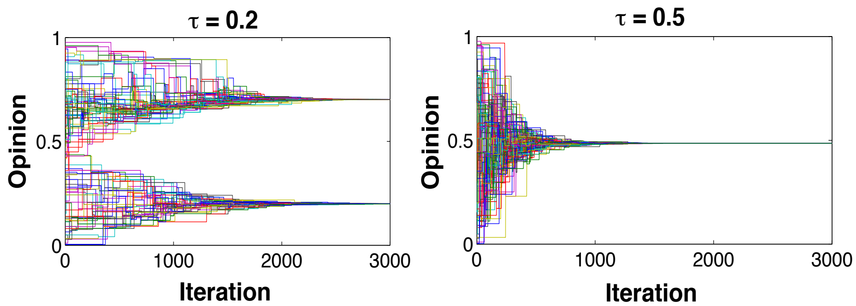

Fig. 5: Opinion evolution with $n = 100$ agents in a connected random geometric graph network, on an unit area with radius 0.6, for two different values of the confidence threshold τ.

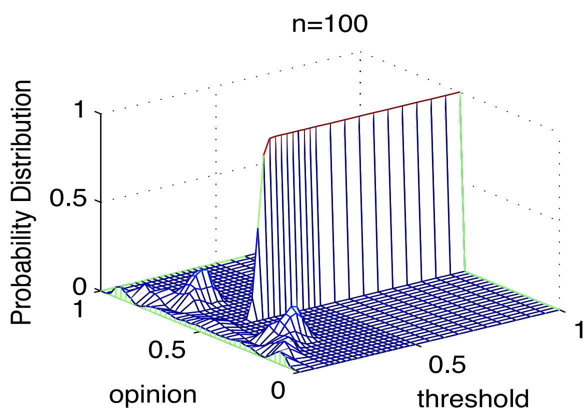

Fig. 6: Histogram of opinion distribution with $\tau = 0.5$ if $|x_{i,k} - x_{j,k}| < \tau$ for k sufficiently large to approach the limit.

IV. Network formation

In the previous section we introduced a graph that constrains the interactions among agents. For social agents there is both self-interest in being exposed to information from other agents as well as a potential cost that can represent the effort in engaging them in the interaction. How can one model the social forces that forge a certain graph? This question is posed in an interesting class of studies, briefly highlighted in Section IV-A, which considers the graph G of relationships as the outcome of strategic partnerships. What is also interesting about these studies is that they depart from the classical game theoretic model of equilibrium since the establishment of a link ij between two nodes requires a coordinated action from the two edges. The question often posed in these studies is if these strategic alliances lead to the greatest social welfare.

If, however, we look at optimizing the network structure from the engineering perspective then the graphical models that emerge come directly from the statistical model of the information, and they capture the intrinsic dependency of the observations made by the agents and their latent decision variables. This topic is fleshed out in Section IV-B.

A. Strategic network formation

In strategic network formation models, each node receives a certain utility $u_i(E)$ from a given graph $G = (V, E)$ and each partnership contributes to this utility by changing benefits and a costs to the two nodes i and j that agree to become connected by an edge ij. In global terms, the *efficiency* of a graph G can be measured either by its sum utility, $\sum_{i=1}^{n} u_i(E)$, which is highest for the most efficient graph, or by considering if the utility vector $u(E) = (u_1(E), \ldots, u_n(E))$ dominates all other possible graph utilities vectors, in the sense that $\forall E', \forall i \in \{1, n\} : u_i(E) \geq u_i(E')$, in which case the graph is called *Pareto efficient*. A formulation for this problem that has been popular in both economics and computer science is as follows (see Chapter 6 of [2] for a thorough review). Each node extracts a utility from a graph that is larger if there is a lower distance from all other nodes, with a cost to maintain links that grows with the nodal degree. For example, an interesting special case analyzed by Fabrikant in [29] is when the utility of graph G to node i takes the form:

$$u_i(E) = \sum_{j=1}^{n} -\ell_{ij}(E) - cd_i(E)$$

where $\ell_{ij}(E)$ is the length of the shortest path in E connecting node i and node j. What are the efficient graphs and what graphs are stable? Note that the notion of stability here is considering what Jackson in [2] calls *pairwise* stability, which for a certain graph G happens if:

- For all $ij \in E$, $u_i(E \geq u_i(E - ij)$ and $u_j(E) \geq u_j(E - ij)$
- For all $ij \notin E$ if $u_i(E + ij) > u_i(E)$ then $u_j(E + ij) < u_j(E)$

In other words, all the edges present make the nodes connected better off and no node makes a sacrifice.

Looking at Fabrikant utility model, consider that the cost of not being connected to anybody is infinite. This automatically excludes graphs that have isolated components from being efficient or pairwise stable in this model. The interesting result is that the utility model above has a unique efficient network and only two possible outcomes:

- For $c < 1$ the complete network is efficient. The cost for adding a link is always lower that the benefit of being one hop distance. The complete network is also pairwise stable.

- For $c > 1$ a star encompassing all nodes is the efficient configuration. One node is taking a large burden in cost so that all the rest can have a 2 hop distance from each other. This is a pairwise stable solution but not the only one.

Fabricant also proved in [29] that the diameter of any pairwise stable network is at most $2\sqrt{c}+1$ and that comparing the total cost $-\sum_{j=1}^{n} u_i(E)$ of these stable solutions with the total cost of the star configuration, which is $2(n-1)(n-1+c)$, their ratio (called by Papadimitriou the *price of anarchy* [30]) is no larger than $17\sqrt{c}$.

There are many other utility functions and variations on this model, like for example the model introduced by Bala and Goyal [31] where the terms $-\ell_{ij}(E)$ are replaced by $\delta^{\ell_{ij}(E)}$, with $\delta < 1$, and the network is directed. In some cases the phase transitions as the parameters that define costs and benefits change that are very similar, others that instead lead to more complex topologies than the star or the complete graph. There are also studies on the dynamic formation of networks, and we refer the reader to [2] for more details.

In concluding this section we have seen that there is a wealth of mathematical tools to examine opinion diffusion and also to predict the structures that are favored in these social interactions. The theory, however, is divorced from the specific technological tools that are in some cases enhancing in some cases influencing, in some cases replacing the agents decision making engine. These important trends are outlined in Section V.

Before that, in the next section we introduce a totally different approach to select graphs to exchange beliefs that is popular in machine learning and signal processing. In the context of decentralized sensors systems, these models provide systematic guidelines to structure both communications and computations aimed at making optimal Bayesian inferences [32]. In fact, the sensors communications can be organized to emulate the implementation of the algorithm known as Belief Propagation [33], which requires passing special messages over specific graphs, called Bayesian Networks, derived from the statistical model of the information. Similarities that may exist between Belief Propagation and the processing of correlated private information in social networks, can provide insight on a society ability to merge all evidence successfully and behave as a rational swarm. This is another interesting direction to explore at the intersection between artificial intelligence and socio-economic studies, which is discussed next.

B. Bayesian Networks and Belief Propagation

Pearl's work [33] pioneered the use of graphical representations to perform Bayesian inference via message passing. The graphs represent the conditional dependency existing between the random data acquired and between the data and latent variables in an observation model; the latent variables include the unknown information θ that is the object of the inference. Considering a set of random variables $\vec{\theta} = (\theta_1, \ldots, \theta_n)$, it is said their distribution factorizes according to a certain Bayesian Network[1] (BN) graph $G = (V, E)$, if the joint distribution of these random variables can be expressed as:

$$p(\theta_1, \ldots, \theta_n) = \prod_{i=0}^{n} p(\theta_i | \mathrm{Pa}(\theta_i))$$

where $\mathrm{Pa}(\theta_i)$ is a subset of $\{\theta_1, \ldots, \theta_n\}/\{\theta_i\}$ whose elements are called *parents* of node θ_i; in the BN graphs vertexes correspond to random variables and the parents have directed links towards all their children, forming the set of edges E of G. It is important to remark that this type of factorization of $p(\theta_1, \ldots, \theta_n)$ is always possible, because of the chain rule of probability. However, the factorization is truly beneficial in computing marginals when $\mathrm{Pa}(\theta_i)$ are small. These graphical representation are the scaffold for the celebrated algorithm called *Belief Propagation* (sometimes referred to as the *sum product* algorithm) which leverages on this factorization especially when the resulting graphs are sparse.

Let us offer a concrete example. The case examined in Section II-A is far too simple to appreciate the power of this machinery. In fact, the distribution of interest to make the inference is $p(\theta|s_1, \ldots, s_n)$ and, since the data are conditionally independent given $p(s_1, \ldots, s_n|\theta) = \prod_{i=1}^{n} p(s_i|\theta)$ the unknown variable θ is the *parent* of all private information and the structure of the BN is rather trivial. There is no need of marginalizing the distribution, all it is needed is to collect the beliefs $p(s_i|\theta)$. A more interesting case arises when $s_i = \theta_i + w_i$ where $(\theta_1, \ldots, \theta_n) \sim \mathcal{N}(\vec{\theta}_0, \Sigma_0)$; for simplicity, assume w_i are i.i.d. zero mean Gaussian noise random variables all with variance σ^2 (this is also called a Gauss-Markov model). In this case the agents are interested in making an inference on their own θ_i and this requires marginalizing the other agents θ_j. When the observations s_i are conditionally independent given θ_i, s_i have each θ_i as their single parent and are, therefore, simple appendices of the graphical structure of the latent variables $(\theta_1, \ldots, \theta_n)$. The complexity of the BN associated to a Gauss-Markov model is tied to the sparsity of the inverse of their covariance, or Fisher Information Matrix (FIM) $J = \Sigma_0^{-1} + \frac{1}{\sigma^2} I$. In fact, let $h = \frac{1}{\sigma^2}(s_1, \ldots, s_n)$. It can be shown [34] that:

$$p(\theta_1, \ldots, \theta_n | s_1, \ldots, s_n) \propto exp\left\{ -\frac{1}{2}\vec{\theta} J \vec{\theta}^T + \vec{\theta} \vec{h}^T \right\},$$

which is the general way the joint PDF of Gaussian vectors is written when studying the BN of the associated Gauss-Markov models. The expression of the joint PDF in a graphical model is subject to the

[1]BN are sometimes referred to as *graphical models*, and expanded in the so called *factor graphs*.

following general factorization (see Hammersley and Clifford [35]):

$$p(\theta_1, \ldots, \theta_n | s_1, \ldots, s_n) \propto \prod_{i \in V} \psi_i(\theta_i) \prod_{ij \in E} \psi_{ij}(\theta_i, \theta_j)$$

which, in the case of the Gauss-Markov model we used as an example, has the following factors:

$$\psi_i(\theta_i) = \exp\left\{-\frac{1}{2} A_i \theta_i^2 + \beta_i \theta_i\right\}, \quad \psi_{ij}(\theta_i, \theta_j) = \exp\left\{-\frac{1}{2}(\theta_i, \theta_j) B_{ij} (\theta_i, \theta_j)^T\right\}.$$

The choice of the parameters A_i and B_{ij} values is not unique. They need to be selected so that

$$\vec{\theta} J \vec{\theta}^T = \sum_{i \in V} A_i \theta_i^2 + \sum_{ij \in E} (\theta_i, \theta_j) B_{ij} (\theta_i, \theta_j)^T,$$

where $G = (V, E)$ is the BN that one associates with the Gauss-Markov model, which is sparse as long as J is sparse, given that B_{ij} have to capture the contributions coming from all off-diagonal elements of J. In general, Belief Propagation is based on the fact that, given the Hammersley and Clifford factorization of the PDF, the variable elimination process of the PDF satisfies certain fixed point equations for the so called *message functions* when the BN is a tree [33]. The elimination of a generic variable i is associated

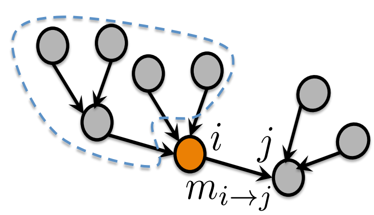

Fig. 7: Message passing over a BN tree. The dashed line encircles the parents of node i. The message $m_{i \to j}$ is passed from node i to node j to eliminate the variables in the BN subtree rooted at i.

with the following fixed-point equations for the message functions $m_{i \to j}(\theta_j)$ (where $i \to j$ is a parent to child link):

$$m_{i \to j}(\theta_j) = \int \psi_{ij}(\theta_i, \theta_j) \psi_i(\theta_i) \prod_{k \to i, k \in Pa(\theta_i)} m_{k \to i}(\theta_i) d\theta_i.$$

Combining the fixed point message functions at each node, one obtains:

$$p_i(\theta_i | s_1, \ldots, s_n) \propto \psi_i(\theta_i) \prod_{k \to j, k \in \mathrm{Pa}(\theta_i)} m_{k \to i}(\theta_i),$$

which means that, eventually, each node learns a function proportional to the belief that incorporates all the evidence in the network and can make an optimum inference in the Bayesian sense. In other words, the message function, $m_{i \to j}(\theta_j)$, from neighbor i to j is the function that results from saturating all of the variables in the parents $\mathrm{Pa}(\theta_i)$ subtree rooted at node i (see Fig. 7). The implementation of belief propagation is obtained by initializing these message functions to, for example, 1 for every value of θ_j, and then iteratively exchanging messages that are re-computed using the equation above, until the algorithm reaches its fixed points.

For random variables that are discrete and with finite sample spaces, the integrals are replaced by sums, and $m_{i\to j}(\theta_j)$ are evaluated over discrete points. For Gaussian Markov fields like the one we considered to show the BN representation, rather than exchanging message functions with continuous support, the information exchange can also be considerably simplified but, to avoid too long of a digression, we direct the reader to [36] and [34].

The analysis of belief propagation over BNs is very insightful since one can point to the difficulty of a certain inference problem just by exploring the complexity of the associated BN. In this sense this approach reveals what stochastic models are best suited for decentralized decision making. Typically those whose BN are not *loopy* and with sparse dependencies work best, and are also those for which the fixed points are unique. Wide experimentation with these algorithms shows also that the message passing methodology in belief propagation has a considerable degree of resilience, even when the conditions for exact convergence are violated.

Overall, this brief introduction should be sufficient to draw some relationship between graphical models and social networks. *In principle they are not related, since BN representations are agnostic about who controls what the data, where they become available to make the decision, centrally or at a set of separate agents, and BN do not allude to actual communications or social interactions; in other words, they are merely a mathematical tool to organize efficiently the calculation of marginal probabilities and make inferences. However, as remarked by several authors considering data aggregation in decentralized sensor networks [32], BN do suggest a systematic methodology to organize communications among distributed observers that have private information on phenomena with specific statistical dependency, so that the network can perform an inference on latent variables that is optimal in the Bayesian sense.* The existence of graphical models that support the computation of optimum inferences is intriguing since it may suggest potential pressure that may affect the network formation in a social setting, coming from the inherent statistical structure of the collective information model. This pressure could manifest itself in the agents empirical observation that forging certain information pathways, choosing certain hierarchies, and processing specific messages that resemble those in belief propagation, all lead to faster and more accurate local decisions for the parties involved.

V. The marriage of social and information networks

The marriage between social and information network has been an enormously successful and prolific one, especially in recent years. For a large fraction of society, decision making, social and personal expression are increasingly relying on computer networks. On the one hand, Internet companies compete on introducing innovation that allows them to get more up-close and personal with their users everyday choices. This phenomenon is discussed next.

A. *The Internet telescope on social behavior*

The great socio-economic implications of providing Internet information services to a large base of consumers are easily apparent. At the same time, from looking at the evolution of Internet technologies and applications, it seems that no amount of information about the customer is too much information and the competition to grab this information is fierce. In fact, to further help this competition, the present trend is towards establishing a standard for what is called *context aware computing*, i.e. a universal model to capture service usage information from the consumer activity, as part of the IP Multimedia Subsystem (IMS) Next Generation Networking (NGN) architecture. This standardization process would unleash further advances in the sector by facilitating the design of components (hardware and software) that would help generate Internet applications seamlessly integrated to work on computers and Smartphones.

While users are lured by the convenience of adopting technology to enhance their decision making, the design and support of these sophisticated applications is clearly motivated by the rich information they provide on their users, and the ability to influence their decisions in essentially two ways: 1) targeted advertising; 2) recommendation systems [37].

Perhaps the greatest advances in the sector today are tied to the additional information Smartphones can offer, by essentially *stalking* consumers in an increasingly invasive fashion, in exchange for help on sorting out tasks and entertaining them in their daily activities. Starting from the first context-aware application, a mobile tour guide called *Active Badge* [38], in the last ten years, there has been a growing trend towards refining the network support for the generation of this side information [39]. The so called *location based services* [40], available as Smartphones applications, accrue data about their users contexts. At first, the only *context* variable recorded had been the user position (the GPS location). Now a growing research area is devoted to capturing a richer record to discern the user activity, using gyroscopes, microphones and cameras, to refine the notion of what is the context in which certain actions are performed (e.g. is the user walking or still, is this a crowded area or a quiet place, etc.). More importantly, computer networking is now exploring social, group and community contexts using indicators of what are the specific interest groups the consumer is engaged in as indicators of social context [41] and from them there is a growing interest in developing *context-aware* recommendation systems [42]. Recent proposals to refine social context information go as far as suggesting to modify 802.11 network protocols to essentially make Smartphones directly capable of establishing node to node communications and support ad-hoc networking activities with peers with specific applications [43], [44]. These proposals build on a considerable body of research on Delay Tolerant Networks (DTN) started with [45], that has focused on the design of mechanisms to route information among mobile terminals using their random encounters. Applied on Smartphones, DTN protocols would allow collaborative activities that could enhance communications among individuals engaged in a variety of social endeavors while, in exchange, giving direct measures of the surrounding phones one use has been engaged with.

Where does the theory we described fit in? In the previous sections, in reviewing the basic mathematical elements of social models, we have determined that the private information s_i, the actions available $a_i \in \mathcal{A}_i$, the agent utility function $u_i(\theta, a_i)$ and, last but not least, the social relationships, are the necessary ingredients to forecast human behavior. The examples we examined use simple tractable selections for all of these mathematical entities but to predict real trends one would have to fit models using informative data. How to best harvest private information from willing participants, in exchange for some desirable service, is the real quest for Internet companies of various scales. It is clear that observing the specific selection of the consumer (the agent actions), in response to Internet search results and context, provides a basic glimpse on the association between private information, beliefs and actions taken by the agents, which in the theoretical models is fully described by their utility functions. By personalizing search engines, offering additional services, companies have the ability to collect evidence over many instances on a specific individual selections. This represents a huge leap compared to what marketing activities and studies had available prior to the Internet. Review systems, surveys and rating systems (e.g. the *like* button on Facebook) and online shopping provide considerable evidence on the agent utility. Information on the agent context offers an additional window onto the agent s_i that can improve the ability to discern the social agent belief, and how that belief is translated into an action. Given that social settings help shape agents opinions, having access to social context is also very important, and tracking social trends can help fill the gaps in records that are generally very incomplete on the agents preferences, given the wide number of options available for their actions. Finally, access to information about the belief of many agents should give information about the variable θ examined in their decision. For example fusing information about several agents purchase history, or subscription to blogs, could inform about the true value of an asset or the importance of a political decision.

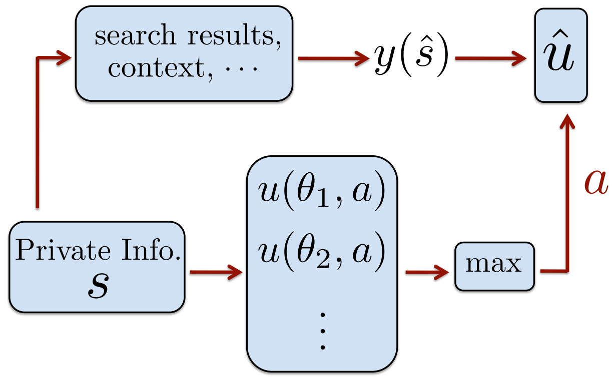

Fig. 8: Information about context and queries from the agent along with the agent decisions inform indirectly on the individual economic choice system.

But a theory that defines a clear trade-off between the costs and benefits of such services seems to be missing. Companies that work in this space, perhaps lacking a theory to focus their efforts on Internet applications that would give the best information for the lowest cost, appear to be restlessly experimenting with new products. In addition, even though the mathematical models we reviewed and the context data accrued by Internet service data are related, a considerable amount of the data mining that is discussed in the literature is either entirely based on heuristics or it explores statistically cause and effect relationships in a rigorous Bayesian framework, but it does not make an explicit attempt of estimating the variables and functions involved in the theoretical social models we described. In a sense, what is missing is a framework to estimate rigorously the human economic choice system, relating context evidence with the system model and ranking the quality of certain contextual information. This is a sensing and system identification problem whose formulation may turn out to be outright impossible, but yet, that is well worth thinking about. The research on *influence diagrams* interpreted as graphical models for individual decision biases [46], [47], are examples of how one can systematically approach the problem of modeling agents behavior. The problem also hinges on the unnerving issue of *privacy*, as the orwellian dystopia (albeit implemented at the hands of private enterprises in most countries) increasingly resembles reality.

VI. Conclusion

Our paper compared models for social learning and opinion diffusion in the context of economics and social science with those used in signal processing over networks of sensors, explaining how learning emerges or fails to emerge in both scenarios. It also critically discusses how the advent of the Internet and of Smartphones is generating a wealth of data and enhancing the decision making capabilities of social agents in ways that were never conceivable before. The paper also argues that more engineering research is needed to advance modeling and inference algorithms and enhance even further the cyber interactions of social agents. This area has tremendous potential as well as carries tremendous risks. Consumers loose their privacy and can be influenced in undesired ways. Hence, a critical consideration to make in expanding our understanding on this subject is to what extent it is safe to increase the ability of hardware and software to capture contextual data about the customers.

REFERENCES

[1] C.P. Chamley, *Rational Herds: Economic Models of Social Learning*, Cambridge University Press, 2003.

[2] M.O. Jackson, *Social and Economic Networks*, Princeton University Press, 2008.

[3] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010.

[4] D. Acemoglu, M.A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," in *Review of Economic Studies*, 2010, vol. 78, pp. 1201–1236.

[5] Jean-Antoine-Nicolas de Caritat (marquis de Condorcet), *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralitée des voix*, De l'Imprimerie royale, 1785.

[6] F. Galton, "Vox populi," *Nature*, vol. 75, pp. 450–451, 1907.

[7] R. Viswanathan and P.K. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54 –63, jan 1997.

[8] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed., 1977.

[9] S.C. Draper and G.W. Wornell, "Side information aware coding strategies for sensor networks," in *IEEE Journal Selected Areas in Communications*, 2004.

[10] X. Vives, "How fast do rational agents learn?," in *Review of Economic Studies*, 1993, vol. 60, pp. 329–347.

[11] Christos H. Papadimitriou and Tim Roughgarden, "Computing equilibria in multi-player games," in *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, Philadelphia, PA, USA, 2005, SODA '05, pp. 82–91, Society for Industrial and Applied Mathematics.

[12] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," in *Journal of Political Economy*, 1992, vol. 100, pp. 992–1026.

[13] S. Lohmann, "Dynamics of informational cascades: The monday demonstrations in leipzig, east germany, 1989-1991," in *World Politics*, vol. 47, pp. 42–101.

[14] P. Sørensen and L. Smith, "Pathological outcomes of observational learning," in *Econometrica*, 2000, vol. 68, pp. 371–398.

[15] R.R. Tenney and N.R. Sandell, "Detection with distributed sensors," in *IEEE Transactions on Aeroespace and Electronic Systems*, 1981, vol. 17.

[16] J.N. Tsitsiklis, "Decentralized detection by a large number of sensors," in *Math. Contr.. Signals. Syst.*, 1988, vol. 1, pp. 167–182.

[17] W.-P. Tay, J.N. Tsitsiklis, and M.Z. Win, "On the sub-exponential decay of detection error probabilities in long tandems," in *IEEE Trans. Inf. Theory*, 2008, pp. 4767–4771.

[18] V. Bala and S. Goyal, "Learning from neighbours," in *Review of Economic Studies*, 1998, vol. 65, pp. 595–621.

[19] Glenn Ellison and Drew Fudenberg, “Word-of-mouth communication and social learning,” *The Quarterly Journal of Economics*, vol. 110, no. 1, pp. 93–125, 1995.

[20] Lones Smith and Peter Norman Sorensen, “Rational social learning by random sampling,” 2008.

[21] A.G. Dimakis, S. Kar, J.M.F. Moura, M.G. Rabbat, and A. Scaglione, “Gossip algorithms for distributed signal processing,” *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847 –1864, nov. 2010.

[22] M.H. DeGroot, “Reaching a consensus,” in *Journal of American Statistcal Association*, 1974, vol. 69, pp. 118–121.

[23] M. Fiedler, “Algebraic connectivity of graphs,” in *Czechoslovak Mathematical Journal*, 1973, vol. 23, pp. 298–305.

[24] R. Hegselmann and U. Krause, “Opinion dynamics and bounded confidence models, analysis and simulations,” in *Jounral of Artificial Societies and Social Simulation*, 2002, vol. 5.

[25] V.D. Blondel, J.M. Hendrickx, and J.N. Tsitsiklis, “On Krause’s multi-agent consensus model with state-dependent connectivity,” in *IEEE Trans. Auto. Control*, 2009, vol. 54, pp. 2586–2597.

[26] A. Mirtabatabaei and F. Bullo, “On opinion dynamics in heterogeneous networks,” in *American Control Conference (ACC)*, 2011.

[27] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, “Mixing beliefs among interacting agents,” in *Adv. Compl. Syst.*, 2000, vol. 3, pp. 87–98.

[28] G. Weisbuch, “Bounded confidence and social networks,” in *The European Physical Journal B - Condensed Matter and Complex Systems*, 2004, vol. 38, pp. 339–343.

[29] Alex Fabrikant, Ankur Luthra, Elitza Maneva, Christos H. Papadimitriou, and Scott Shenker, “On a network creation game,” in *Proceedings of the twenty-second annual symposium on Principles of distributed computing*, New York, NY, USA, 2003, PODC ’03, pp. 347–351, ACM.

[30] Christos Papadimitriou, “Algorithms, games, and the internet,” in *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, New York, NY, USA, 2001, STOC ’01, pp. 749–753, ACM.

[31] V. Bala and S. Goyal, “A noncooperative model of network formation,” in *Econometrica*, 2000, vol. 68, pp. 1181–1229.

[32] Mujdat Cetin, Lei Chen, John Fisher, Alexander Ihler, Randolph Moses, and Alan Willsky, “Distributed fusion in sensor networks: A graphical models perspective,” *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 42–55, Jul 2006.

[33] Judea Pearl, *Probabilistic inference in intelligent systems*, Morgan Kaufmann, 1988.

[34] Alan S. Willsky Dmitry M. Malioutov, Jason K. Johnson, “Walk-sums and belief propagation in gaussian graphical models,” *Journal of Machine Learning Research*, vol. 7, pp. 2031–2064, 2006.

[35] J. Hammersley and P. Clifford, *Unpublished manuscript*, Markov fields on finite graphs and lattices, 1971.

[36] P. Rusmevichientong and B. Van Roy, “An analysis of belief propagation on the turbo decoding graph with gaussian densities,” *IEEE Trans. Information Theory*, vol. 48, 2001.

[37] A. Ansari, S. Essegaier, and R. Kohli, “Internet recommendation systems,” in *Journal of Marketing Research*, 2000, vol. 37, pp. 363–375.

[38] R. Want, A. Hopper, V. Falcão, and J. Gibbons, “The active badge location system,” in *ACM Trans. Inf. Syst.*, 1992, vol. 10, pp. 91–102.

[39] A.K. Dey, G.D. Abowd, and D. Salber, “A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications,” in *Journal of Human-Computer Interaction*, 2001, vol. 16, pp. 97–166.

[40] B.N. Schilit, N. Adams, and R. Want, “Context-aware computing applications,” in *IEEE Workshop on Mobile Computing Systems and Applications*, 1994.

[41] A.S. Pentland, "Socially aware computation and communication," in *Journal of Computer*. 2005, pp. 33–40, IEEE Computer Society Press.

[42] G. Adomavicius, A. Tuzhilin, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender Systems Handbook, chapter Context-aware Recommender Systems*, Springer, 2010.

[43] E. Miluzzo, N.D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S.B. Eisenman, X. Zheng, and A.T. Campbell, "Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application," in *Proceedings of the 6th ACM conference on Embedded network sensor systems*, 2008, pp. 337–350.

[44] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: social-based forwarding in delay tolerant networks," in *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, 2008, pp. 241–250.

[45] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, 2003, pp. 27–34.

[46] Fahiem Bacchus and Adam Grove, "Graphical models for preference and utility," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, San Francisco, CA, USA, 1995, UAI'95, pp. 3–10, Morgan Kaufmann Publishers Inc.

[47] Daphne Koller and Brian Milch, "Multi-agent influence diagrams for representing and solving games," *Games and Economic Behavior*, vol. 45, no. 1, pp. 181 – 221, 2003, ¡ce:title¿First World Congress of the Game Theory Society¡/ce:title¿.