Open Multi-Agent Systems: Gossiping with Deterministic Arrivals and Departures

Julien M. Hendrickx and Samuel Martin

Abstract— **This work is a first step towards the study of open multi-agent systems: systems that agents can join and leave, and where arrivals and departures happen on a time-scale comparable to that of the process running on the system. We study the behavior of the average pairwise gossip algorithm on such open systems, and provide an exact characterization of its evolution in terms of three scale-independent quantities that are shown to be solutions of a 3-dimensional linear dynamical system. We then focus on two particular cases: one where each departure is immediately followed by an arrival, and one where agents keep arriving without ever leaving the system, so that the number of agents grows unbounded.**

I. Introduction

Among the most frequently cited properties of multi-agent systems are their flexibility and scalability. In particular, they are deemed robust to the disappearance of agents and able to adapt to the arrival of new agents. Think for instance of a flock of birds, to the internet, or to an ad hoc network of mobile devices. Yet most results on multi-agent systems characterize their (often asymptotic) properties under the assumption that their composition remains unchanged. This apparent contradiction can be interpreted as an implicit assumption that the time-scale of the process considered is very different from that of the agents arrivals and departures But this assumption may not necessarily hold for very large systems; convergence speed typically decreases when the system size grows, while the frequency of arrivals departures would be expected to increase. In living systems, the birth frequency is indeed related to the population size. Similarly, in computer networks, the frequency of failures is proportional to the number of nodes in the network, and in many cases the same is true for the frequency of new connections. The assumption could also be questioned in chaotic environments where communications would be difficult and infrequent, resulting in a slow convergence rate, while agent malfunction would be more likely.

Hence we consider here *open multi-agent systems*, where agents keep arriving and/or leaving during the execution of the process considered.

Repeated arrivals and departures result in important differences in the analysis or the design of open multi-agent systems and cause several challenges:

State dimension: Every arrival results in an increase of the system state dimension, and every departure in a decrease of the system state dimension. Analyzing the evolution of the system state is therefore much more challenging than in "closed systems".

Absence of usual "convergence:" Being continuously perturbed by departures and arrivals, open systems will never asymptotically converge to a specific state. Rather, they may approach some form of steady state behavior, which can be characterized by some relevant descriptive quantities. As in classical control in the presence of perturbations, the choice of the measures is not neutral, and different descriptive quantities may behave in very different way. Think for example of H_1 or H_2 criteria in classical control.

Robustness and quality of the algorithms: The departure of an agent will almost inevitably affect the execution of any algorithm and may result in a loss of information, especially if the agent does not send a last message warning about its departure. Many algorithms designed to achieve a specific purpose, e.g. computing the shortest path between two nodes in a network, rely on the information held by all agents. Hence, these algorithms must first of all be robust to departures or arrivals. On the other hand, they do not need to be exact: Obtaining an exact result while the system composition keeps changing will indeed often be impossible. So a good algorithm should constantly produce results sufficiently close to the value to be computed. But obtaining that exact value in case the system composition were to stop changing may not be necessary, and this additional requirement could result in a loss of efficiency. This aspect is related to, but different from, the notion of self-stabilizing decentralized algorithms in computer science [1], [4]. Self-stabilizing algorithms must indeed produce an exact value once agents stop arriving or leaving, but are not required to achieve any specific performance while these arrivals and departures take place.

A. Contribution

We report here preliminary results of a ongoing study on open multi-agent systems. As a first step towards tackling the complexity of their analysis, we study the behavior of the discrete-time average pairwise gossip algorithm [2] with all-to-all (possible) communications, focusing on systems where departures and arrival take place at pre-determined times, see Section II for a complete definition.

We analyze the system evolution in terms of three "scale-independant" quantities, namely the expected mean $\mathbb{E}\bar{x}$, the

Julien Hendrickx is with the ICTEAM institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium. julien.hendrickx@uclouvain.be, Samuel Martin is with Université de Lorraine and CNRS, CRAN, UMR 7039, 2 Avenue de la Forêt et de Haye, 54518 Vandoeuvre-lès-Nancy, France samuel.martin@univ-lorraine.fr

This work was supported by the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Program, initiated by the Belgian Science Policy Office.

expected square mean $\mathbb{E}(\bar{x}^2)$ and the expected variance $\mathbb{E}(\overline{x^2} - \bar{x}^2)$ of the system state x. We show in Section III that these quantities can be characterized exactly, and that they evolve according to an associated 3-dimensional linear system.

In Section IV, we analyze in detail the case of systems with periodic replacements: a departure immediately followed by an arrival takes place every K time steps. We then focus in Section V on growing systems that agent keep joining without ever leaving. It will in particular be shown that periodic arrivals or replacements can result in a significant performance decrease in terms of variance.

B. Other works on open multi-agent systems

The possibility of agents joining or leaving the system has been recognized in computer science, and specific architectures have for example been proposed to deploy large-scale open multi-agent systems, see e.g. THOMAS project [3]. There also exist mechanisms allowing distributed computation processes to cope with the shut down of certain nodes or to take advantage of the arrival of new nodes.

Frameworks similar to open multi-agent arrivals have also been considered in the context of trust and reputation computation, motivated by the need to determine which arriving agents may be considered reliable, see e.g. the model FIRE [6]. However, the study of these algorithms behavior is mostly empirical.

Varying compositions were also studied in the context of self-stabilizing population protocols [1], [4], where interacting agents (typically finite-state machines) can undergo temporary or permanent failures, which can respectively represent the replacement or the departure of an agent. The objective in those works is to design algorithms that eventually stabilize on the desired answer if the system composition stops changing, i.e. once the system has become "closed".

Opinion dynamics models with arrivals and departures have also been empirically studied in [7], [9].

II. System Description

As explained in the introduction, we consider a multi-agent system whose composition evolves with time. We use integers to label the agents. We denote by $\mathcal{N}(t) \subset \mathbb{N}$ the set of agents present in the system at time t, and by $n(t)$ the number of agents present at time t, i.e. the cardinality of $\mathcal{N}(t)$. Each agent i holds a value $x_i(t) \in \Re$, and we make no assumptions about the values held at $t = 0$ by the agents initially present in the system.

We consider a discrete evolution of the time $t \in \mathbb{N}$. It is possible to interpret the discrete time t as a sampling of a continuous time variable. Samples then correspond to instants where an event occurred. We will comment later on this interpretation and on its implication on the scaling of different parameters. At each time t, one of three events may occur:

(a) Gossip: Two agents $i, j \in \mathcal{N}(t)$ are uniformly randomly and independently selected among the $n(t)$ agents present in the system (with in particular the possibility of selecting twice the same agent), and they update their values x_i, x_j by performing a pairwise average:

$$x_i(t+1) = x_j(t+1) = \frac{x_i(t) + x_j(t)}{2}. \qquad (1)$$

(b) Departure: One uniformly randomly selected agent $i \in \mathcal{N}(t)$ leaves the system, so that $\mathcal{N}(t+1) = \mathcal{N}(t) \setminus \{i\}$ and $n(t+1) = n(t) - 1$. This event may only occur if $n(t) > 0$.

(c) Arrival: One "new" agent $i \notin \mathcal{N}(s)$, $\forall s \leq t$, joins the system, so that $\mathcal{N}(t+1) = \mathcal{N}(t) \cup i$ and $n(t+1) = n(t)$. The initial value $x_i(t+1) \in \Re$ of the arriving agent is drawn independently from a constant distribution $\mathcal{D}$ with mean 0 and variance σ^2.

Note that all the random events above are assumed independent of each other. In addition, we will sometimes consider for simplicity a "replacement" event, which consists of the instantaneous combination of a departure and an arrival: an agent leaves the system and is instantaneously replaced.

Scale-independent quantities of interest

The aim of the study is to characterize the disagreement among agents, *i.e.* the distance to consensus. We say that consensus is reached asymptotically when

$$\lim_{t\to\infty} \max_{(i,j)\in\mathcal{N}(t)^2} |x_i(t) - x_j(t)| = 0. \qquad (2)$$

If the system dynamics does not include agent departures or arrivals, it is known that the gossip process we consider leads to consensus, see e.g. [2], [5]. The objective here is to understand how agent arrivals and departures impact the disagreement among agents. To do so, we study several quantities of interest. Because the system size may change significantly with time, we focus on scale-independent quantities, i.e. quantities whose values is independent of the size of the system. We consider in particular the empirical state mean and variance defined as

$$\begin{aligned} \bar{x} &= \tfrac{1}{n} \sum_{i\in\mathcal{N}} x_i, \\ \mathrm{Var}(x) &= \tfrac{1}{n} \sum_{i\in\mathcal{N}} (x_i - \bar{x})^2, \end{aligned} \qquad (3)$$

respectively, where references to time were removed to lighten the notation. Our study will focus on the evolution of $\mathbb{E}\mathrm{Var}(x)$, which will also require monitoring $\mathbb{E}(\bar{x})^2$. When new agents keep arriving it is impossible to achieve asymptotic consensus in the sense of (2), because the new agent's value will with high probability be different from the value of the agents already present in the system. The study of $\mathbb{E}\mathrm{Var}(x)$ will allow us to see how "far" the system will be from consensus. But we will see that in certain systems whose sizes grow unbounded, we may have $\lim_{t\to\infty} \mathbb{E}\mathrm{Var}(x) = 0$, corresponding to a form of "almost consensus".

III. Closed-form evolution of three scale-independent quantities

As explained in the previous section, we will analyze the evolution of the expected variance, the expected square mean, and to a lesser extend the expected mean. We show in this section that for a given sequence of gossips, arrivals and departures, these expected values evolve according to an associated 3-dimensional linear system. We first successively compute these variables after each type of event.

Stronger variations of the next lemma are available in different earlier works, see e.g. [2], [5]. We provide its proof in Appendix A for the sake of completeness.

Lemma 1 (Gossip): Suppose that a randomly selected pair of agents engage in a gossip averaging according to equation (1). Let x be the state of the system before that interaction, x' its state after the interaction, and n the number of agents. There holds

$$\begin{aligned} \mathbb{E}(\bar{x'}|x) &= \bar{x}, \\ \mathbb{E}(\overline{x'}^2|x) &= \bar{x}^2, \\ \mathbb{E}(\text{Var}(x')|x) &= (1-\tfrac{1}{n})\text{Var}(x). \end{aligned} \tag{4}$$

Lemma 2 (Departure): Suppose that a randomly selected agent departs from the system. Denote x the state before departure, x' the state after departure and n the number of agents before departure. Then, there holds

$$\begin{aligned} \mathbb{E}(\bar{x'}|x) &= \bar{x}, \\ \mathbb{E}(\overline{x'}^2|x) &= \tfrac{1}{(n-1)^2}\text{Var}(x)+\bar{x}^2, \\ \mathbb{E}(\text{Var}(x')|x)) &= (1-\tfrac{1}{(n-1)^2})\text{Var}(x). \end{aligned} \tag{5}$$

Lemma 3 (Arrival): Suppose that an agent arrives into the system. Denote x the state before arrival, x' the state after arrival and n the number of agents before arrival. Then, there holds

$$\begin{aligned} \mathbb{E}(\bar{x}'|x) &= \tfrac{n}{n+1}\bar{x}, \\ \mathbb{E}(\overline{x'}^2|x) &= \tfrac{n^2}{(n+1)^2}(\bar{x})^2+\tfrac{1}{(n+1)^2}\sigma^2 \\ \mathbb{E}(\text{Var}(x')|x) &= \tfrac{n}{(n+1)^2}(\bar{x})^2+\tfrac{n}{n+1}\text{Var}(x)+\tfrac{n}{(n+1)^2}\sigma^2 \end{aligned} \tag{6}$$

Notice that if a system undergoes a departure followed by an arrival, the number of agents n will not be the same when applying Lemmas 2 and 3. The previous lemmas show that the expected mean $\mathbb{E}\bar{x}$ evolves independently of the two other scale-independent quantities.

IV. Fixed size system with replacements

A. Description of the periodic evolution

As a first case study, we consider systems where the number of agents is mostly constant: an agent leaving the system is immediately replaced, and replacing a leaving agents is the only circumstance under which an agent joins the system. The number n of agents remains thus constant, except during the instantaneous replacements of the leaving agents. In addition, we suppose here that the timing of these events is periodic: exactly K gossip events take place between two replacements. Replacement occurs thus at all times $p(K+1), p\in\mathbb{N}$ and gossip events occur at times $p(K+1)+g, p\in\mathbb{N}, g\in\{1,\dots,K\}$. The evolution of the agent values for a typical realization of this system with $n=5, K=20$ is represented in Figure 1(A), and the evolution of the corresponding square mean value and variance is shown in Figure 1(B). One can see that this particular system periodically approaches a state of quasi-consensus before being perturbed each time by the replacement of an agent.

B. Recurrence Relation

The next proposition shows that the evolution of the expected mean, square mean and variance measured just after the replacements can be described by a time-independent stable linear iteration.

Proposition 1: Denote by x be the state of the system at time $p(K+1)+1$ for some $p\in\mathbb{N}$, a time just after replacement (i.e. one departure immediately followed by an arrival), and by x' the state vector at time $(p+1)(K+1)+1$, that is, K gossip iterations and one replacement later. There holds

$$\begin{pmatrix} \mathbb{E}(\bar{x}'^2|x) \\ \mathbb{E}(\text{Var}(x')|x) \end{pmatrix} = \frac{1}{n^2}\begin{pmatrix} (n-1)^2 & \rho \\ n-1 & (n^2-n-1)\rho \end{pmatrix}\begin{pmatrix} \bar{x}^2 \\ \text{Var}(x) \end{pmatrix} + \frac{\sigma^2}{n^2}\begin{pmatrix} 1 \\ n-1 \end{pmatrix} \tag{7}$$

and

$$\mathbb{E}(\bar{x}'|x) = \left(1-\frac{1}{n}\right)\bar{x}, \tag{8}$$

where we remind that $\rho=(1-\frac{1}{n})^K$ is the contraction ratio of K gossip iterations.
Moreover, the linear iterations (7) and (8) converge to a fixed points.

Proof: The proof of iterations (7) and (8) immediately follow from K applications of Lemma 1 followed by one application of Lemma 2 and one of Lemma 3. The convergence of (7) can be proved by observing that the matrix involved is irreducible and row sub-stochastic [8, Section 8.2]. The convergence of (8) is immediate. ■

The previous proposition can be read as a reccurence over the expected quantities instead of the conditional expected quantities by recalling that $\mathbb{E}(\mathbb{E}(X|Y))=\mathbb{E}(X)$. We see in (7) and (8) that the evolution of the expected mean is decoupled from that of the expected square mean and that of the variance. Moreover, the mean evolves exactly as if no gossip was taking place. On the other hand, the evolution of the variance and expected square mean are coupled, and cannot be described independently of each other.

C. Steady state regime

The periodic departures and arrivals prevent the system state x from converging. But we have seen in Proposition 1 that the expected mean, square mean and variance measured after the replacements do converge, as can be seen on an example in Figure 1(C). We provide in the next proposition

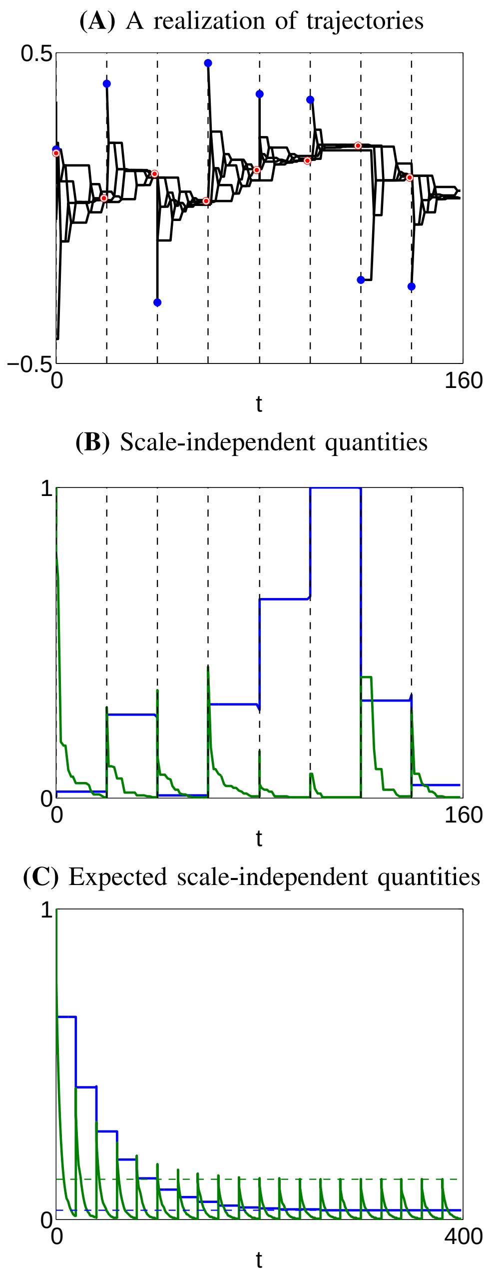

Fig. 1. Illustration of an open system with periodic replacements as considered in Section IV: The system contains $n = 5$ agents and replacements occur every $K = 20$ time-steps. Arriving agent values are drawn uniformly in $[-\frac{1}{2}, \frac{1}{2}]$ so that $\sigma^2 = \frac{1}{12}$. (A) shows the evolution with time of the agents values (in black) for a typical realization. Red circles highlight the departing agents while the blue circles correspond to the newly arrived agents. (B) shows the evolution of the variance (green) and square mean value (blue) for the same realization. (C) shows the evolution of the expected variance (green) and square mean value (blue) computed using Proposition 1 and Lemma 1. The curves in (B) and (C) have been normalized by their maximal values ($\max(\bar{x}^2) = 0.49$, $\max(\text{Var}(x)) = 0.13$, $\max(\mathbb{E}(\bar{x}^2)) = 0.34$, $\max(\mathbb{E}(\text{Var}(x))) = 0.11$).

a closed-form expression of their asymptotic values, which can be verified by direct computation.

Proposition 2: Iterations (7) and (8) asymptotically converge to fixed points whose coordinates are

$$\mathbb{E}\bar{x}|_{eq} = 0,$$

$$\mathbb{E}\text{Var}(x)|_{eq} = \frac{\sigma^2}{n}\frac{1-\frac{1}{n}}{(1-\frac{1}{2n}) - \rho(1-\frac{3}{2n})} \tag{9}$$

and

$$\mathbb{E}\bar{x}^2\Big|_{eq} = \frac{\sigma^2}{2n}\frac{1-\rho(1+\frac{2}{n})}{1+\frac{1}{2n} - \rho(1+\frac{3}{2n})}. \tag{10}$$

We remind that the iterations in Proposition 1 concern the quantities of interest immediately after the replacement, and so are thus the steady state values given in (9) and (10) in Proposition 2. Since the mean is not affected by the gossip iterations, the expected mean square is constant between two replacements and converges thus to the steady state value (10). The expected variance, on the other hand, decreases after each gossip iteration. It will thus not converge, but asymptotically approach a periodic behavior, taking the steady state value (11) immediately after the replacement, and decaying then geometrically with a rate ρ until the next replacement, as represented on an example in Figure 1(C).

1) Interpretation for extreme gossip/replacement ratios: We first consider the case where no gossip iteration takes place, which corresponds to taking $\rho = 1$ in iteration (7). The steady state values given in Proposition 2 become

$$\mathbb{E}\text{Var}(x)|_{eq} = \sigma^2\left(1-\frac{1}{n}\right), \quad \text{and} \quad \mathbb{E}\bar{x}^2\Big|_{eq} = \frac{\sigma^2}{n}, \tag{11}$$

which is indeed what one may expect, since $\rho = 1$ would correspond to taking n i.i.d. random values with mean 0 and variance σ^2.

If on the other hand we were to let $\rho \to 0$, corresponding to moving all agents values to their average before any replacement, we would have

$$\mathbb{E}\text{Var}(x)|_{eq} = \frac{\sigma^2}{n}\frac{n-1}{n-\frac{1}{2}}, \quad \text{and} \quad \mathbb{E}\bar{x}^2\Big|_{eq} = \frac{\sigma^2}{2n-1}. \tag{12}$$

The steady state value of $\mathbb{E}\bar{x}^2$ is lower than what we would have obtained by taking n independent random values. To shed some light on the specific value taken, we describe an alternative direct way of re-obtaining it. Observe that the average $\bar{x}(t)$ of the agents present in the system at time t can also be expressed as a weighted average of the initial values of all agents that are present at time t or have been present at some time before t. Let us denote by $\tilde{x}_s$ the initial value of the s^{th} agent that joined the system, and t_s the first time at which it is present. The weight of the $\tilde{x}_s$ in the average $\bar{x}(t_s)$ at time t_s is exactly $\frac{1}{n}$, as the arriving agent has exactly this value, and none of the other other $n-1$ agent values have been influenced by it. The weight of $\tilde{x}_{s-1}$, however, is smaller. That value was indeed distributed evenly between n

agents during the (perfectly averaging) gossip phase between the arrivals of $s-1$ and s, and one of these agents left the system prior to the arrival of s. Hence the weight of $\tilde{x}_{s-1}$ is $\frac{1}{n}(1-\frac{1}{n})$. Similarly, one can verify that the weight of $\tilde{x}_{s-2}$ is $\frac{1}{n}(1-\frac{1}{n})^2$, the weight of $\tilde{x}_{s-k}$ is $\frac{1}{n}(1-\frac{1}{n})^k$ (assuming there have been more than k arrivals since $t=0$.) In steady state, i.e., if the process has been running since $-\infty$, the average is thus

$$\bar{x}(t_s) = \sum_{k=-\infty}^{s} \frac{1}{n}\left(1-\frac{1}{n}\right)^{s-k} \tilde{x}_k = \sum_{k=0}^{\infty} \frac{1}{n}\left(1-\frac{1}{n}\right)^{k} x_{s-k}. \tag{13}$$

(During the transient situation, the infinite series is truncated and the remaining weight is distributed evenly between the values of the agents initially present in the system). One can verify that the weights in (13) sum to 1. Using $\mathbb{E}\tilde{x}_k = 0$ and $\mathbb{E}x_k^2 = \sigma^2$ (there is indeed asymptotically no influence of the agents initially present in the system and for which these assumptions were not made), and using the independence between the different $\tilde{x}_k$, we obtain

$$\begin{aligned}\mathbb{E}(\bar{x}(t_s))^2 &= \sum_{k=0}^{\infty}\left(\frac{1}{n}\left(1-\frac{1}{n}\right)^k\right)^2 \mathbb{E}x_{s-k}^2 \\ &= \frac{\sigma^2}{n^2}\sum_{s=0}^{\infty}\left(\left(1-\frac{1}{n}\right)^2\right)^k \\ &= \frac{\sigma^2}{n^2}\frac{1}{1-\left(1-\frac{1}{n}\right)^2} = \frac{\sigma^2}{2n-1},\end{aligned}$$

which corresponds to the steady state obtained in (12). The smaller expected square average in the presence of perfectly averaging gossip ($\rho = 0$) is thus explained by the influence of the initial values of all agents having been present in the system at some point. By contrast, these had no influence when no gossips are performed, i.e. when $\rho = 1$. A similar phenomenon occurs if $\rho \in (0,1)$, also leading to an expected square mean $\mathbb{E}\bar{x}^2 \leq \frac{\sigma^2}{n}$ due to influence of agents no longer present in the system, but its direct analysis is more complex. We finally note that the steady state expected variance in (12) is entirely explained by the arrival of the new agent, whose value is different from the common value of the $n-1$ agents already present.

2) Interpretation for large-scale systems: We now assume the number of agents n to be very large, while keeping K constant. To understand why K should be kept constant, suppose that our discrete time-instants t correspond to the sampling of a continuous time variable τ at the instants at which some events occur. A natural assumption, though clearly not the only possible one, is that the number of replacements per unit of time τ scales linearly with n. We can also expect the number of gossip iterations in which a given agent is involved per unit of time τ to be independent of the system size (at least for large n), so that the total number of gossip iterations taking place per unit of time τ would also scale linearly with n. As a result, the ratio between the number gossip iterations per unit of time and the number of replacements per unit of time, which is represented by K, should be independent of n.

Remember now that $\rho = (1-\frac{1}{n})^K$. Hence for large n and fixed K, we have $\rho = 1 - \frac{K}{n} + o(\frac{1}{n})$, and the steady state expected square mean (10) becomes

$$\begin{aligned}\mathbb{E}\bar{x}^2\big|_{eq} &= \frac{\sigma^2}{2n}\frac{1-\left(1-\frac{K}{n}+o(\frac{1}{n})\right)(1+\frac{2}{n})}{1+\frac{1}{2n}-\left(1-\frac{K}{n}+o(\frac{1}{n})\right)(1+\frac{3}{2n})} \\ &= \frac{\sigma^2}{2n}\frac{-\frac{2}{n}+\frac{K}{n}+o(\frac{1}{n})}{\frac{1}{2n}-\frac{3}{2n}+\frac{K}{n}+o(\frac{1}{n})} \\ &= \frac{\sigma^2}{2n}\frac{K-2+o(1)}{K-1+o(1)}.\end{aligned}$$

As a consequence,

$$\mathbb{E}\bar{x}^2\big|_{eq} \sim_{n\to+\infty} \frac{\sigma^2}{2n}\frac{K-2}{K-1},$$

which again is smaller than what would have been obtained by averaging n independent random variables with mean 0 and variance σ^2. This is again due to the influence of agents no longer present in the system, as in the case $\rho = 0$ in Section IV-C.1.

Similarly, for large n, the expected variance (9) becomes

$$\begin{aligned}\mathbb{E}\mathrm{Var}(x)|_{eq} &= \frac{\sigma^2}{n}\frac{1-\frac{1}{n}}{(1-\frac{1}{2n})-(1-\frac{K}{n}+o(\frac{1}{n}))(1-\frac{3}{2n})} \\ &= \frac{\sigma^2}{n}\frac{1-\frac{1}{n}}{1-\frac{1}{2n}-1+\frac{3}{2n}+\frac{K}{n}(1-\frac{3}{2n})+o(\frac{1}{n})} \\ &= \frac{\sigma^2}{n}\frac{1-\frac{1}{n}}{\frac{1}{n}+\frac{K}{n}+o(\frac{1}{n})} = \sigma^2\frac{1-\frac{1}{n}}{1+K+o(1)}.\end{aligned}$$

As a consequence,

$$\lim_{n\to+\infty}\mathbb{E}\mathrm{Var}(x)|_{eq} = \frac{\sigma^2}{1+K}.$$

For large n and constant K, the steady-state variance is thus inversely proportional to the number of gossip iterations taking place before an agent replacement. Moreover, suppose agent i is randomly selected among those present in the system at a given time. Then there holds at any time $\mathbb{E}x_i^2 = \mathbb{E}\bar{x}^2 + \mathbb{E}\mathrm{Var}(x)$. For large n, the steady state value of $\mathbb{E}x_i^2$ is thus mostly driven by the variance:

$$\lim_{n\to+\infty}\mathbb{E}x_i^2\big|_{eq} = \frac{\sigma^2}{(1+K)}. \tag{14}$$

Note that x_i could be considered as an estimate of the mean value of the distribution according to which the initial values of arriving agents are drawn, which in our case is 0. The limit (14) would then mean that for large n, a random agent estimate would be as accurate as the average of $K+1$ independent samples of that distribution, which can be considered a poor performance since one can verify that a randomly selected agent has (for large t) on average been involved in $2K$ gossip iterations. We will see later that this number of gossip iterations would indeed result in a much lower variance in comparable "closed systems".

Note finally that $\mathbb{E}\mathrm{Var}(x)\big|_{eq}$ above stands for the limit of the expected variance just after a replacement. Instead, the limit of the expected variance just before the next replacement will be $(1-\frac{1}{n})^K\frac{\sigma^2}{(1+K)}$ which converges to $\frac{\sigma^2}{(1+K)}$ when n is large for constant K

Since initial states have all been drawn independently from the same initial distribution, at any time it holds $\mathbb{E}x_i^2 = \mathbb{E}\overline{x^2}$, so that the expected square value is $\mathbb{E}x_i^2 = \mathbb{E}\bar{x}^2 + \mathbb{E}\text{Var}(x)$. Observe that for large n, the steady state of this value is mostly driven by the variance and

$$\lim_{n\to+\infty} \mathbb{E}x_i^2\big|_{eq} = \frac{\sigma^2}{(1+K)}.$$

The random agent estimate of the external distribution mean 0 is thus comparable to an estimate obtained by averaging $K+1$ independent samples.

V. Growing system without departure

We focus now on systems whose sizes grow unbounded because new agents keep joining while no agent ever leaves. Formally, the system is initially empty, and a new agent (with label n) joins the system at every time $t_n - 1$, for some sequence of times $1 = t_1 < t_2 < \dots$. As a consequence, the number of agents in the system is $n(t_n - 1) = n - 1$ and $n(t_n) = n$. Gossip steps take place at all times other than $t_n - 1, n \in \mathbb{N}$. As in the rest of this work, we assume that the initial value $x_n(t_n)$ of every agent is a random variable with $\mathbb{E}x_n(t_n) = 0, \mathbb{E}x_n(t_n)^2 = \sigma^2$. We let in addition $K_n = t_{n+1} - t_n - 1$ be the number of gossip steps taking place between the arrival of agent n and $n+1$, and will discuss later different possible dependence of K_n on n.

We focus on the values of the expected square mean and variance just after the arrivals of the agents, i.e. at times t_n. As in Section IV, their evolution can be described by a two-dimensional linear system. This system is here time-varying because n is not constant, but the absence of departures makes it triangular, and hence easier to analyze. It follows indeed from Lemmas 1 and 3 (or from a direct computation) that the expected square mean is $\mathbb{E}\bar{x}^2(t_n) = \frac{\sigma^2}{n}$, which decays thus to 0 when n grows as could be expected. The same lemmas imply that the expected variance satisfies

$$(n+1)\mathbb{E}\text{Var}(x(t_{n+1})) = n\mathbb{E}\text{Var}(x(t_n))\rho_n + \frac{n}{(n+1)}(\mathbb{E}\bar{x}^2(t_n) + \sigma^2),$$

with $\rho_n = (1-\frac{1}{n})^{K_n}$. Reintroducing $\mathbb{E}\bar{x}^2(t_n) = \frac{\sigma^2}{n}$ yields then

$$(n+1)\mathbb{E}\text{Var}(x(t_{n+1})) = n\mathbb{E}\text{Var}(x(t_n))\rho_n + \sigma^2. \qquad (15)$$

This recursion allows obtaining the following theorem characterizing the asymptotic variance, and proved in Appendix B.

Theorem 1: Consider the growing system without departure, and remember that K_n is the number of gossips between the arrival of agents n and $n+1$. Let $K \geq 1$.
(i) If $K_n = K$ for all $n \geq n_0$ for some n_0, then

$$\lim_{n\to\infty} \mathbb{E}\text{Var}(x(t_n)) = \frac{\sigma^2}{K+1}.$$

(ii) If $\lim_{n\to\infty} K_n = \infty$, then $\lim_{n\to\infty} \mathbb{E}\text{Var}(x(t_n)) = 0$.

Theorem 1(ii) shows that the system essentially converges to a consensus as soon as K_n grows unbounded, even if this growth is very slow, and even if the number K_n/n of gossips per agent between two consecutive arrivals tends to 0. Note, however, that each agent gets involved (with probability 1) in infinitely many gossips when $K_n \to \infty$. The expected number of gossips in which an agent has been involved at time t_n is indeed $2\frac{1}{n}\sum_{m=1}^n K_m$, which grows unbounded.

By contrast, in the case of a fixed $K_n = K$ agents have on average been involved in $2\frac{1}{n}\sum_{m=1}^n K_m = 2K$ gossips after any given arrival, which intuitively explains why the variance stays bounded away from 0. But the actual asymptotic value $\frac{\sigma^2}{K+1}$ obtained in Theorem 1(i) is remarkably high. As a basis for comparison, suppose we had first waited until the n agents were present in the system which would yield an expected variance $\sigma^2\frac{n-1}{n}$, and then performed the same number nK of gossip averaging operations between randomly selected pairs of nodes. It follows from nK application of Lemma 1 that the expected variance would then have been

$$\frac{n-1}{n}\sigma^2(1-\frac{1}{n})^{nK} \to_{n\to\infty} \sigma^2 e^{-K},$$

which is significantly lower than $\sigma^2/(K+1)$ (for $K=5$, the ratio of variance would be $\frac{e^{-5}}{1/6} \simeq 0.04$). The dynamics of the system composition deteriorates thus considerably the performances in terms of variance reduction.

Possible evolutions of K_n

Suppose that we interpret our discrete t as the sampling of a real continuous time variable τ at those times τ_t at which an event occurs, as in Section IV-C.2. It is again reasonable to assume the interaction rate of an agent to be independent of the system size, so that the total number of gossips per unit of time τ would grow linearly with n, as say $\lambda_g n$. Suppose first that the agents arrive at a fixed rate λ_a. In that case, the number of gossips between two arrivals would be linearly growing with n and $K_n = n\lambda_g/\lambda_a$. Theorem 1(ii) shows then that the variance would converge to 0.

But one could also imagine a linearly growing rate of arrivals $\lambda_r n$. This would for example be the case if the system attraction were growing with its size or if the arrivals resulted from some form of reproduction process. The number of gossip iterations between two arrivals would then be constant $K_n = K = (\lambda_g n)/(\lambda_r n)$, leading to a finite variance $\frac{\sigma^2}{1+K}$.

VI. Conclusions

In this paper, we argued that the possibility for an agent to leave or join the system is natural in many multi-agent systems, and should therefore be taken into account in their analysis or design. We have highlighted several challenges coming with the study of these open multi-agent systems. These include variations of the state dimension and the absence of the usual notion of state convergence. We focused on a simple open multi-agent system involving pairwise gossip averaging without communication restriction, and showed how the evolution of such a system can be characterized by studying relevant scale-independent quantities. We provided closed-form solution for the evolution of these quantities along with the expression of their steady state. The analysis was carried out for under two distinct scenarios: fixed system size with periodic agent replacements and systems whose size grows unbounded due to repeated agent arrivals and an absence of departure.

Interestingly, the steady state expected variance for large number of agents n were the same in the case of periodic replacements (see Section IV-C.2) and in the case of growing systems with periodic arrivals without departures (see Section V). Both tend indeed to $\frac{\sigma^2}{1+K}$ for large n, where K is the number of gossip iterations between two successive replacements or arrivals. K is moreover related to the expected number of gossip iterations in which an agent randomly selected in the system was involved, which in both cases is $2K$. The value of this variance is significantly higher than what would have been obtained in a comparable "closed system". Performing the same number of random gossip iterations per agent in a system without arrivals or departures and in which n agents have i.i.d. initial values would indeed lead to a variance close to $\sigma^2 e^{-K}$ for large n.

The methodology developed in this work is applicable to a much broader range of scenarios: A natural extension would be to have agents arriving or leaving at random times according to some prescribed distribution. One could also imagine that their arrivals and departures could be related to their age and/or to their position in some underlying interaction network. More generally, we believe that our approach could be adapted to other sorts of multi-agent systems involving more complex interactions.

References

[1] Dana Angluin, James Aspnes, Michael J Fischer, and Hong Jiang. Self-stabilizing population protocols. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 3(4):13, 2008.

[2] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE/ACM Transactions on Networking (TON)*, 14(SI):2508–2530, 2006.

[3] Carlos Carrascosa, Adriana Giret, Vicente Julian, Miguel Rebollo, Estefania Argente, and V Botti. Service oriented mas: an open architecture. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 1291–1292. International Foundation for Autonomous Agents and Multiagent Systems, 2009.

[4] Carole Delporte-Gallet, Hugues Fauconnier, Rachid Guerraoui, and Eric Ruppert. When birds die: Making population protocols fault-tolerant. In *International Conference on Distributed Computing in Sensor Systems*, pages 51–66. Springer, 2006.

[5] Fabio Fagnani and Sandro Zampieri. Randomized consensus algorithms over large scale networks. *IEEE Journal on Selected Areas in Communications*, 26(4):634–649, May 2008.

[6] Trung Dong Huynh, Nicholas R Jennings, and Nigel R Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.

[7] Gerardo Iñiguez, János Török, Taha Yasseri, Kimmo Kaski, and János Kertész. Modeling social dynamics in a collaborative environment. *EPJ Data Science*, 3(1):1, 2014.

[8] Carl D Meyer. *Matrix analysis and applied linear algebra*, volume 2. Siam, 2000.

[9] János Török, Gerardo Iñiguez, Taha Yasseri, Maxi San Miguel, Kimmo Kaski, and János Kertész. Opinions, conflicts, and consensus: modeling social dynamics in a collaborative environment. *Physical review letters*, 110(8):088701, 2013.

Appendix

A. Proofs of Lemmas 1, 2 and 3

Lemma 1 *Proof:* Let us first fix the nodes i, j involved in the gossip. Observe that $x'_i + x'_j = 2\frac{x_i+x_j}{2} = x_i + x_j$, and that $x'_k = x_k$ for all $k \neq i, j$. Hence $\bar{x}' = \bar{x}$, which implies $\mathbb{E}\bar{x}' = \bar{x}$ and $\mathbb{E}\bar{x}^2 = \bar{x}^2$. We also have

$$\begin{aligned}\mathrm{Var}(x') =& \frac{1}{n}\sum_{k=1}^{n}(x'_k - \bar{x})^2 \\ =& \frac{1}{n}\left(\sum_{k=1}^{n}(x_k-\bar{x})^2\right) - \frac{1}{n}(x_i-\bar{x})^2 - \frac{1}{n}(x_j-\bar{x})^2 \\ &+ \frac{2}{n}\left(\frac{x_i+x_j}{2} - \bar{x}\right)^2 \\ =& \mathrm{Var}(x) - \frac{1}{2n}(x_i-\bar{x})^2 - \frac{1}{2n}(x_j-\bar{x})^2 \\ &+ \frac{4}{n}(x_i-\bar{x})(x_j-\bar{x}). \end{aligned} \quad (16)$$

Observe now that $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ and $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \mathrm{Var}(x)$. Taking the expected value of (16) with respect to i and j yields then

$$\mathbb{E}(\mathrm{Var}(x')|x) = \left(1 - \frac{1}{n}\right)\mathrm{Var}(x).$$

■

Lemma 2. *Proof:* We remind that n denotes the number of agents before the departure. Suppose agent j leaves, then the new average is

$$\begin{aligned}\bar{x}' &= \frac{1}{n-1}\sum_{i=1, i\neq j}^{n} x_i = \frac{1}{n-1}\sum_{i=1}^{n} x_i - \frac{x_j}{n-1} \\ &= \frac{n}{n-1}\bar{x} - \frac{1}{n-1}x_j. \end{aligned} \quad (17)$$

Taking the expected value of this expression with respect to the randomly selected agent j yields

$$\mathbb{E}(\bar{x}'|x) = \frac{n}{n-1}\bar{x} - \frac{1}{n-1}\bar{x} = \bar{x}.$$

Exactly the same argument applied to $\overline{x^2} = \frac{1}{n}\sum_i x_i^2$ shows

$$\mathbb{E}(\overline{x'^2}|x) = \overline{x^2}, \quad (18)$$

which will be helpful in the sequel. Turning to the square mean, we obtain from (17)

$$\begin{aligned}\mathbb{E}((\bar{x}')^2|x) &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{n}{n-1}\bar{x} - \frac{1}{n-1}x_j\right)^2 \\ &= \frac{1}{n}\sum_{j=1}^{n}\left(\left(\frac{x_j}{n-1}\right)^2 - \frac{2nx_j\bar{x}}{(n-1)^2} + \left(\frac{n\bar{x}}{n-1}\right)^2\right) \\ &= \frac{x^Tx}{n(n-1)^2} - \frac{2n\bar{x}^2}{(n-1)^2} + \frac{n^2\bar{x}^2}{(n-1)^2} \\ &= \frac{1}{(n-1)^2}\left(\mathrm{Var}(x) + \bar{x}^2 - 2n\bar{x}^2 + n^2\bar{x}^2\right) \\ &= \frac{1}{(n-1)^2}\mathrm{Var}(x) + \bar{x}^2. \end{aligned}$$

Concerning the variance, it follows from this last equality and from (18) that

$$\begin{aligned}\mathbb{E}(\text{Var}(x')|x) &= \mathbb{E}(\overline{x'^2}|x) - \mathbb{E}((\bar{x}')^2|x) \\ &= \overline{x^2} - \frac{1}{(n-1)^2}\text{Var}(x) - \bar{x}^2 \\ &= \left(1 - \frac{1}{(n-1)^2}\right)\text{Var}(x).\end{aligned}$$

∎

Lemma 3. *Proof:* We remind that n is the number of agents prior to the arrival, and we label $n+1$ the arriving agent for simplicity, so that $x'_i = x_i$ for all $i \le n$. We begin again by computing the new average:

$$\begin{aligned}\bar{x}' &= \frac{1}{n+1}\left(x'_{n+1} + \sum_{i=1}^{n} x_i\right) \\ &= \frac{n}{n+1}\bar{x} + \frac{1}{n+1}x'_{n+1}. \qquad (19)\end{aligned}$$

Since $\mathbb{E}x'_{n+1} = 0$, we have $\mathbb{E}(\bar{x}'|x) = \frac{n}{n+1}\bar{x}$. By exactly the same reasoning but using $\mathbb{E}{x'_{n+1}}^2 = \sigma^2$ and $\overline{x^2} = \bar{x}^2 + \text{Var}(x)$, we also obtain

$$\begin{aligned}\mathbb{E}(\overline{x'^2}|x) &= \frac{n}{n+1}\overline{x^2} + \frac{1}{n+1}\sigma^2. \qquad (20) \\ &= \frac{n}{n+1}\bar{x}^2 + \frac{n}{n+1}\text{Var}(x) + \frac{1}{n+1}\sigma^2. \qquad (21)\end{aligned}$$

Turning to the square average, we obtain from (19)

$$\begin{aligned}\mathbb{E}((\bar{x}')^2|x) =& \frac{n^2}{(n+1)^2}(\bar{x}')^2 + \frac{n}{(n+1)^2}\bar{x}\mathbb{E}x'_{n+1} \\ &+ \frac{1}{(n+1)^2}\mathbb{E}(x'_{n+1})^2 \\ =& \frac{n^2}{(n+1)^2}(\bar{x}')^2 + 0 + \frac{1}{(n+1)^2}\sigma^2.\end{aligned}$$

The expression of the expected variance is again obtained by combining this last equality with (20):

$$\begin{aligned}\mathbb{E}(\text{Var}(x')|x) =& \mathbb{E}(\overline{x'^2}|x) - \mathbb{E}((\bar{x}')^2|x) \\ =& \frac{n}{n+1}\bar{x}^2 + \frac{n}{n+1}\text{Var}(x) + \frac{1}{n+1}\sigma^2 \\ &+ \frac{n^2}{(n+1)^2}(\bar{x}')^2 + \frac{1}{(n+1)^2}\sigma^2 \\ =& \frac{n}{(n+1)^2}\bar{x}^2 + \frac{n}{(n+1)}\text{Var}(x) + \frac{n}{(n+1)^2}\sigma^2.\end{aligned}$$

∎

B. Proof of Theorem 1

We prove the following proposition, which implies Theorem 1 when applied to $W_n = n\mathbb{E}Var(t_n)$.

Proposition 3: Let $n_0 \ge 2$. Let $\rho_n = (1 - \frac{1}{n})^{K_n}$, and consider the sequence defined by $W_1 = 0$ and

$$W_{n+1} = W_n\rho_n + \sigma^2. \qquad (22)$$

a) If $K_n \ge K$ for all $n \ge n_0$, then $\limsup_{n\to\infty} \frac{W_n}{n} \le \frac{\sigma^2}{1+K}$

b) If $K_n \le K$ for all $n \ge n_0$, then $\liminf_{n\to\infty} \frac{W_n}{n} \ge \frac{\sigma^2}{1+K}$.

Proof: We prove the statement (a) of the proposition. Statement (b) can be obtained in a similar way. It follows from (22) that

$$W_n = W_{n_0}\Pi_{m=n_0}^{n-1}\rho_m + \sigma^2 \sum_{s=n_0+1}^{n} \Pi_{m=s}^{n-1}\rho_m, \qquad (23)$$

with the convention $\Pi_{m=n}^{n-1}\rho_m = 1$. Using $\log(1-x)^k \le -kx$ (for $x < 1$), the assumption $K_n \ge K$ for all $n \ge n_0$, and the definition $\rho_n = (1 - \frac{1}{n})^{K_n}$ we obtain for $s \ge n_0$

$$\log\left(\Pi_{m=s}^{n-1}\rho_m\right) \le \sum_{m=s}^{n-1} -K\frac{1}{m} \qquad (24)$$

Observe that $\sum_{m=s}^{n-1}\frac{1}{m}$ is an upper approximation of the integral $\int_{x=s}^{n}\frac{1}{x}dx$, and hence

$$\frac{1}{s} + \frac{1}{s+1} + \cdots + \frac{1}{n-1} \ge \log(n) - \log(s) \ge \log\frac{n}{s}.$$

Reintroducing this in (24) yields

$$\Pi_{m=s}^{n-1}\rho_m \le \left(\frac{s}{n}\right)^K, \qquad \forall s \ge n_0,$$

and hence, noticing $W_n \ge 0$, (23) implies

$$W_n \le W_{n_0}\left(\frac{n_0}{n}\right)^K + \frac{\sigma^2}{n^K}\sum\sum_{s=n_0+1}^{n} s^K. \qquad (25)$$

Observing now that $\sum_{s=n_0+1}^{n} s^K$ is a lower approximation of $\int_{n_0+1}^{n+1} x^K dx$, we obtain

$$W_n \le W_{n_0}\left(\frac{n_0}{n}\right)^K + \frac{\sigma^2}{n^K}\frac{(n+1)^{K+1} - (n_0+1)^{K+1}}{K+1}. \qquad (26)$$

The first term in (26) decays to 0 when n grows, while the second one is bounded by $\frac{\sigma^2}{1+K}\frac{(n+1)^{K+1}}{n^K}$. Hence $\limsup \frac{W_n}{n} \le \frac{\sigma^2}{1+K}$.

Part (b) follows a parallel reasoning using the upper bound $\log(1 - \frac{1}{n})^K \ge -\frac{K}{n-1}$. ∎