An invariance principle-based concentration result for large-scale stochastic pairwise interaction network systems

Giacomo Como, *Member, IEEE,* Fabio Fagnani, *Member, IEEE,* and Sandro Zampieri, *Fellow, IEEE*

***Abstract*—We study stochastic pairwise interaction network systems whereby a finite population of agents, identified with the nodes of a (directed) graph, update their states in response to both individual mutations and pairwise interactions with their neighbors. The considered class of systems includes the main epidemic models —such as the SIS, SIR, and SIRS models—, certain social dynamics models —such as the voter and anti-voter models—, as well as evolutionary dynamics on graphs. Since these stochastic systems fall into the class of finite-state Markov chains, they always admit stationary distributions. We analyze the asymptotic behavior of the stationary distributions of stochastic pairwise interaction network systems in the limit as the population size grows large, while the interaction network maintains certain mixing properties. Our approach relies on the use of Lyapunov-type functions to obtain concentration results on these stationary distributions. Notably, our results are not limited to fully mixed population models, as they do apply to a much broader spectrum of interaction network structures, including, e.g., Erdös-Rényi random graphs.**

***Index Terms*—Stochastic network systems, pairwise interaction systems, concentration of stationary distributions, equilibrium selection, stochastically stable states.**

I. Introduction

Pairwise Interaction Network models (PIN models) constitute a class of network systems whereby a finite —possibly very large— population of agents update their states asynchronously, according to stochastic rules that account for both individual mutations and pairwise interactions between neighbor agents [1]–[3]. The class of PIN models encompasses the main microscopic epidemic models —such as SIS, SIR, or SIRS models over networks [4]—, certain social dynamics models —such as the voter, anti-voter, and the Axelrod models [5]–[7]— as well as evolutionary dynamics on graphs [8], [9].

Despite the significant attention that PIN models have received over the years, their theoretical understanding on general interaction patterns still presents considerable gaps. With some notable exceptions —including the analysis of the extinction time in network SIS epidemic models [10] or of size of the population that eventually becomes infected in network SIR epidemic models— the majority of the literature on the analysis of epidemic models has concentrated on mean-field models, studying the epidemic as a system of ordinary differential equations (ODEs) [11]–[14]. In fact, under proper technical assumptions, Kurtz's theorem [15] ensures that the solutions of such mean-field ODEs well approximate the evolution of PIN models on fully mixed populations (i.e., when the interaction pattern is the complete graph) of increasing size over bounded time horizons, see, e.g., [16, Chapter 10] or [17, Chapter 5]. Recent works explore extensions of these asymptotic results for special classes of dynamics on certain sequences of sparse or dense graphs [18]–[21], as well as to continuous state spaces [22].

While the Kurtz theorem provides a deterministic approximation of the transient dynamic behavior of stochastic interaction network systems on fully mixed populations in the large-scale limit, the infinite-horizon behavior of such systems is best captured by characterizing their stationary distributions for finite population sizes and possibly analyzing their convergence times to stationarity [16, Chapter 11]. When the considered stochastic network system lacks ergodicity due, e.g., to the existence of absorbing configurations (such as the all-susceptible configuration in the aforementioned SIS and SIRS epidemic models), a common approach consists in introducing a (small) noise term that makes the system ergodic and characterizing the dependence of the stationary distribution on the noise level [23]–[25]. Explicit analytical expressions are rare and essentially limited to reversible systems [26], [27]. However, it is often possible to study the asymptotic behavior of such stationary distributions in the large population (and vanishing noise) limit [16, Chapter 12]. By studying this (double) limit, one can often show that the probability mass of every sequence of invariant distributions of these network systems concentrates on either a single or a small subset of population states, thus leading to equilibrium selection of so-called stochastically stable population states [28], [29].

Notably, for models over complete graphs, the set of such stochastically stable population states has been proved to be a subset of the so-called Birkhoff center of the corresponding mean-field ODE, defined as the closure of the set of its recurring points: c.f. [30], [31] and [16, Theorem 12.6.2]. While such Birkhoff center includes equilibrium points, cycles, and more complex limit sets of the mean-field ODE, regardless of their stability, in some special cases this analysis has been refined to rule out unstable equilibrium points of the mean-field ODE [16, Theorem 12.6.4]. Observe that the results in [30], [31] and [16, Theorem 12.6.2] can be applied whenever the result of Kurtz's theorem can be extended, such as in the special classes of dynamics and sequences of graphs mentioned above [18]–[21].

In this paper, we focus on PIN models over general interaction patterns and show that some of the results obtained for PIN models on fully mixed populations can actually be extended to far more general classes of interaction patterns

G. Como and F. Fagnani are with the Department of Mathematical Sciences "G.L. Lagrange," Politecnico di Torino, 10129 Torino, Italy (e-mail: {giacomo.como; fabio.fagnani}@polito.it). G. Como is also with the Department of Automatic Control, Lund University, 22100 Lund, Sweden. S. Zampieri is with the Department of Information Engineering, University of Padua, via Gradenigo 14/b, Padova, Italy (e-mail: zampi@dei.unipd.it).

This work was partially supported by the Research Project PRIN 2022 "Extracting Essential Information and Dynamics from Complex Networks" (Grant Agreement number 2022MBC2EZ) funded by the Italian Ministry of University and Research.

beyond complete graphs. Our approach bypasses the need to generalize Kurtz's finite-horizon results in order to study the infinite-horizon behavior. In fact, our results are achieved by developing Lyapunov function arguments tailored to the PIN models, enabling us to establish a concentration result for their stationary probability distributions. Specifically, we demonstrate that, if the mean-field ODE associated with the PIN model admits a class-$\mathcal{C}^2$ global Lyapunov function V, then, in the large-scale limit, every sequence of stationary distributions of the PIN models concentrates on the set of zeros of the time-derivative $\dot{V}$ of such Lyapunov function. This holds true under the assumption that the interaction pattern of the PIN model satisfies a specific topological property, referred to as asymptotic total mixing. As we shall illustrate, in addition to complete graphs, random graphs such as the Erdös-Rényi model (under the assumption that the average degree grows faster than the cubic root of the square graph order) also enjoy this property. As a consequence, our approach applies to PIN models on a significantly broader spectrum of interaction patterns beyond fully mixed populations. We then generalize our results to account for cases where a class-$\mathcal{C}^2$ mean-field Lyapunov function is available only for suitable perturbations of the considered PIN model, but not necessarily for the original one, as is the case for the SIRS epidemic model.

We wish to point out that, while the use of Lyapunov functions for Markov chains is standard in the context of stochastic stability (c.f. the celebrated Foster's theorem [32], [33] for positive recurrence of irreducible Markov chains with infinite state space), the technical novelty of our approach resides in using mean-field Lyapunov functions to prove concentration results for PIN models. Indeed, the latter are finite-state Markov chains, so that existence of stationary distributions is always guaranteed, and their uniqueness is implied by irreducibility [34], whereas the interest is rather on concentration of their stationary distributions in the large-scale limit. In this sense, rather than stability analysis, our results are to be considered as a sort of invariance principle for large-scale stochastic PIN models.

The rest of the paper is organized as follows. The last paragraph of this Introduction gathers some notational conventions. In Section II, PIN models are introduced. In Section III, we first present the important concepts of limit drift and mean-field Lyapunov function (Section II-B) as well as of asymptotically totally mixing network (Section III-A). Then, we formally state our main results on concentration properties of PIN models in the large-scale limit (Section III-B). In Section IV, various examples are discussed, including forgetful PIN models (Section IV-A), binary PIN models (Section IV-B), as well as the SIRS model that does not fit in either of the previous two categories (Section IV-C). The main technical contributions are presented in Section V, where first the mean drift of PIN models is analyzed (Section V-A), then concentration results are developed for PIN models on finite networks (Section V-B), and finally such results are applied in order to prove the large-scale limit concentration results (Section V-C). Section VI gathers some concluding remarks while the Appendices contain the proofs of some technical results, including the proof that the Erdös-Rényi random graph is asymptotically totally mixing.

Notation For a finite set $\mathcal{A}$, its cardinality is denoted by $|\mathcal{A}|$. For two finite sets $\mathcal{A}$ and $\mathcal{B}$, and another, not necessarily finite, set $\mathcal{F}$, we denote by $\mathcal{F}^{\mathcal{A}}$ and $\mathcal{F}^{\mathcal{A}\times\mathcal{B}}$ the spaces of vectors z of dimension $|\mathcal{A}|$ and of matrices Z of dimension $|\mathcal{A}|\times|\mathcal{B}|$, with entries z_i in $\mathcal{F}$ indexed by the elements i of $\mathcal{A}$ and Z_{ij} in $\mathcal{F}$ indexed by the pairs (i,j) in $\mathcal{A}\times\mathcal{B}$, respectively. For a vector a in $\mathbb{R}^{\mathcal{A}}$ the symbols $\|a\|_\infty$, $\|a\|_2$, and $\|a\|_1$ stand for the standard l_∞, l_2, and l_1 norms. For a matrix A symbol $\|A\|_\infty$ means the ∞-norm of its vectorization, namely the maximum absolute value of its entries. The identity matrix is denoted by I, the all-one vector by $\mathbf{1}$, while δ^j stands for the vector with all entries equal to zero except for the j-th that is equal to 1. For a vector z in $\mathcal{F}^{\mathcal{A}}$ and some i in $\mathcal{A}$, z_{-i} denotes the vector in $\mathcal{F}^{\mathcal{A}\setminus\{i\}}$ obtained from z by removing its i-th entry. The simplex of probability vectors over a finite set $\mathcal{A}$ is denoted by

$$\mathcal{P}(\mathcal{A}) = \left\{z \in \mathbb{R}_+^{\mathcal{A}} : \, ||z||_1 = 1\right\}$$

where $\mathbb{R}_+$ denotes the set on non-negative real numbers.

Throughout the paper, we consider functions having as domain the simplex $\mathcal{P}(\mathcal{A})$ and taking values over $\mathbb{R}^q$. Such functions are defined to be differentiable of class $\mathcal{C}^k$, whenever it is possible to extend them to some open set $\mathcal{O} \subseteq \mathbb{R}^{\mathcal{A}}$ such that $\mathcal{O} \supseteq \mathcal{P}(\mathcal{A})$ and such extension is differentiable of class $\mathcal{C}^k$ in the classical sense. We notice that for a differentiable map $V : \mathcal{P}(\mathcal{A}) \to \mathbb{R}$, the gradient in general depends on the particular extension of V, but its projection on the hyperplane $\{x \in \mathbb{R}^{\mathcal{A}} \;:\; \mathbf{1}'x = 0\}$ does not. Throughout the paper, we denote with the symbol $\nabla V(\theta)$ such projection intended as a row vector. In particular, if x in $\mathbb{R}^{\mathcal{A}}$ is such that $\mathbf{1}'x = 0$, then, $\nabla V(\theta)x$ is a well defined scalar that does not depend on the specific extension of V. (C.f. also [16, Appendix B.3].)

A (finite, directed) graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is the pair of a nonempty finite set of nodes $\mathcal{V}$ and of a set of links $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$. We shall always assume that there are no self-loops, i.e., that $(u,u) \notin \mathcal{E}$ for every u in $\mathcal{V}$. We shall refer to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ as undirected if $(u,v) \in \mathcal{E}$ implies that $(v,u) \in \mathcal{E}$ and as nonempty if $\mathcal{E} \neq \emptyset$. A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $(u,v) \in \mathcal{E}$ for every pair of nodes $u \neq v$, is called *complete*.

II. Problem setting

A. Pairwise interaction network models

We consider stochastic *pairwise interaction network (PIN) models*, whereby a nonempty finite set of agents $\mathcal{V}$ are engaged in repeated interactions on a nonempty graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The nodes in $\mathcal{V}$ represent the agents and the directed links in $\mathcal{E}$ represent the allowed pairwise interactions. Precisely, the presence of a link (u,v) in $\mathcal{E}$ directed from its tail node u to its head node $v \neq u$ indicates a direct influence of the state of agent v on that of agent u. We shall refer to the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ as the *interaction pattern* and to $n = |\mathcal{V}|$ and $m = |\mathcal{E}|$ as the *order* and the *size*, respectively, of the PIN model.

Every agent u in $\mathcal{V}$ is endowed with a time-varying *state* $X_u(t)$, taking values in a nonempty finite state set $\mathcal{A}$, at every discrete time instant $t = 0, 1, 2, \ldots$. We stack all the agents' states in a vector $X(t) = (X_u(t))_{u\in\mathcal{V}}$, to be referred to as the system *configuration*, taking values in the configuration space $\mathcal{X} = \mathcal{A}^{\mathcal{V}}$. The PIN model then generates a discrete-time Markov chain $X(t)$ on the configuration space $\mathcal{X}$, whereby, for every time instant $t = 0, 1, 2, \ldots$, the next configuration $X(t+1)$ differs from the current one $X(t)$ in at most one entry, corresponding to an agent u in $\mathcal{V}$, which modifies her action from $X_u(t)$ to $X_u(t+1)$ as a result of either an individual mutation or of a pairwise interaction with a neighbor agent in the interaction pattern $\mathcal{G}$, as described below.

More precisely, at every time instant $t = 0, 1, 2, \ldots$, we have that:

- a pairwise interaction takes place with probability ρ, while an individual mutation takes place with probability $1-\rho$;
- given that an individual mutation takes place, an agent u is chosen uniformly at random from the set $\mathcal{V}$ and gets activated: given its current state $X_u(t) = i$ in $\mathcal{A}$, agent u modifies it to a new state $X_u(t+1) = j$ in $\mathcal{A}$ with conditional probability P_{ij}, while the rest of the agents keep their states unaltered, so that $X_{-u}(t+1) = X_{-u}(t)$;
- given that a pairwise interaction takes place, a directed link (u, v) is activated, chosen uniformly at random from the set $\mathcal{E}$: given that agent u —corresponding to the tail node of the activated link— is in state $X_u(t) = i$ in $\mathcal{A}$, and agent v —corresponding to the head node of the activated link— is in state $X_v(t) = \ell$ in $\mathcal{A}$, agent u modifies its state to a new state $X_u(t+1) = j$ in $\mathcal{A}$ chosen with conditional probability $\phi_{ij}(\ell)$, while once again the rest of the agents keep their states unaltered, so that $X_{-u}(t+1) = X_{-u}(t)$.

Throughout, it is always implicitly assumed that

$$0 \leq \rho \leq 1, \qquad \sum_{j \in \mathcal{A}} P_{ij} = 1, \qquad \sum_{j \in \mathcal{A}} \phi_{ij}(\ell) = 1,$$

for every two states i and ℓ in $\mathcal{A}$. We shall refer to the row-stochastic matrix P in $\mathbb{R}_+^{\mathcal{A}\times\mathcal{A}}$ as the *mutation transition matrix* and to the map $\phi : \mathcal{A} \to \mathbb{R}_+^{\mathcal{A}\times\mathcal{A}}$ as the *interaction transition tensor*. We shall refer to the quadruple $(\mathcal{A}, \rho, P, \phi)$ as the *parameters* of the considered PIN model.

A probability vector μ in $\mathcal{P}(\mathcal{X})$ over the configuration space $\mathcal{X}$ is referred to as a *stationary distribution* for a PIN model if, whenever its initial configuration $X(0)$ has probability distribution $\mathbb{P}(X(0) = \mathbf{x}) = \mu_{\mathbf{x}}$ for all $\mathbf{x}$ in $\mathcal{X}$, the PIN model configuration $X(t)$ has probability distribution $\mathbb{P}(X(t) = \mathbf{x}) = \mu_{\mathbf{x}}$ for all $\mathbf{x}$ in $\mathcal{X}$ and for every time instant $t = 0, 1, 2, \ldots$. Observe that, since every PIN model is a Markov chain on the finite configuration space $\mathcal{X}$, it always admits at least one stationary distribution μ in $\mathcal{P}(\mathcal{X})$.

A PIN model is referred to as *ergodic* if, for every two configurations $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{X}$, there exists a sequence of node and link activations and of corresponding spontaneous mutations and pairwise interactions, respectively, all having positive probability, that lead the PIN configuration from $X(0) = \mathbf{x}$ to $X(t) = \mathbf{y}$ in a finite number of time steps $t \geq 0$. It is well known that the stationary distribution μ of an ergodic PIN model (and in fact of every ergodic finite-state Markov chain [34]) is unique and assigns positive probability $\mu_{\mathbf{x}} > 0$ to every configuration $\mathbf{x}$ in $\mathcal{X}$. Notice that ergodicity is ensured, for instance, provided that the mutation probability $1-\rho > 0$ is positive and the mutation transition matrix P is irreducible.

Remark 1: Observe that the stationary distribution of a PIN model may not be unique if the PIN model is not ergodic. Although our theory is applicable to all such PIN models, the results that will be presented hold particular significance for ergodic PIN models. As mentioned in Section I, certain PIN models of interest that are not themselves ergodic can be analyzed by introducing a perturbation in the mutation kernel that makes them ergodic and then by taking the limit as such perturbation vanishes. Such vanishing noise limit approach has been adopted, e.g., in the theory of random perturbations of dynamical systems [35] and then applied within the theory of learning in games [23]–[25].

Example 1 (SIS epidemic models): In classical SIS epidemic models, the state of an agent can be either susceptible (S) or infected (I). A susceptible agent can become infected if it interacts with an infected agent, while infected agents can spontaneously recover and become susceptible. A PIN model that encompasses classical SIS epidemic models is described below.

- the agents' state set is $\mathcal{A} = \{0, 1\}$, where 0 indicates the susceptible state and 1 indicates the infected state;
- the interaction transition tensor is
$$\phi(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \phi(1) = \begin{pmatrix} 1 - \phi_{01}(1) & \phi_{01}(1) \\ 0 & 1 \end{pmatrix},$$
where $\phi_{01}(1)$ is the probability of contagion transmission, given that a pairwise interaction takes place between a susceptible and an infected individual;
- the mutation transition matrix P is a 2×2 stochastic matrix in which P_{10} is the probability of spontaneous recovery and P_{01} the probability of spontaneous infection, given that a mutation takes place.

Notice that, in the standard case when $P_{01} = 0$, the PIN model is not ergodic and the only stationary distribution is a delta distribution supported on the configuration where all agents are susceptible, which is the unique absorbing configuration.

Remark 2: Continuous-time PIN models can also be considered, as defined in the following. For an interaction pattern $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let every agent v in $\mathcal{V}$ and every link (u, v) in $\mathcal{E}$ be equipped with an independent Poisson clock, such that the rate of every agent's clock is $\omega > 0$ and the rate of every link's clock is $\varpi > 0$. Then, define a continuous time Markov chain $\overline{X}(\tau)$ on the configuration space $\mathcal{X}$ as follows. If the clock of agent u ticks at some time $\tau \geq 0$, then agent u updates her state from $\overline{X}_u(\tau^-)$ to $\overline{X}_u(\tau)$ according to the conditional probability $\mathbb{P}(\overline{X}_u(\tau) = j | \overline{X}(\tau^-) = \mathbf{x}) = P_{\mathbf{x}_u j}$ for every state j in $\mathcal{A}$, and configuration $\mathbf{x}$ in $\mathcal{X}$, while the rest of the agents keep their states unaltered, so that $\overline{X}_{-u}(\tau) = \overline{X}_{-u}(\tau^-)$. On the other hand, if the clock of link (u, v) ticks at time τ, then agent u modifies her state $\overline{X}_u(\tau)$ with conditional probability distribution $\mathbb{P}(\overline{X}_u(\tau) = j | X(\tau^-) = \mathbf{x}) = \phi_{\mathbf{x}_u j}(\mathbf{x}_v)$, while once again the rest of the agents keep their states unaltered, so that $\overline{X}_{-u}(\tau) = \overline{X}_{-u}(\tau^-)$. Observe that two clocks tick at the same time with probability zero, as we have assumed them to be independent. It then follows that the so-called jump chain [34, Section 2.6] of the continuous-time Markov chain $\overline{X}(\tau)$ is the discrete-time PIN model $X(t)$ with interaction pattern $\mathcal{G}$ and parameters $(\mathcal{A}, \rho, P, \phi)$, where the pairwise interaction probability is

$$\rho = \frac{m\varpi}{n\omega + m\varpi} \,,$$

where $n = |\mathcal{V}|$ and $m = |\mathcal{E}|$. In particular, a probability distribution μ in $\mathcal{P}(\mathcal{X})$ is stationary for $\overline{X}(\tau)$ if and only if it is stationary for $X(t)$.

B. Limit drift and mean-field Lyapunov functions

In the rest of the paper we shall study concentration properties of stationary distributions of PIN models. Specifically, let $\theta : \mathcal{X} \to \mathcal{P}(\mathcal{A})$ be the function mapping a configuration $\mathbf{x}$ in $\mathcal{X}$ into its *type* $\theta(\mathbf{x})$ in $\mathcal{P}(\mathcal{A})$ whose entries

$$\theta_i(\mathbf{x}) := \frac{1}{n}|\{v \in \mathcal{V} : \mathbf{x}_v = i\}| \,, \qquad i \in \mathcal{A} \,,$$

represent the empirical frequencies of the different states in the configuration $\mathbf{x}$. We shall then focus on concentration properties of the *type process*

$$\Theta(t) := \theta(X(t)) \,, \qquad t = 0, 1, \ldots \,. \qquad (1)$$

Notice that, for every discrete time instant $t = 0, 1, 2, \ldots$, the random variable $\Theta(t)$ takes values in the compact simplex $\mathcal{P}(\mathcal{A})$ of probability vectors over the state space $\mathcal{A}$. In particular, we shall focus on the asymptotic behavior of the distribution of $\Theta(t)$ in the large-scale limit, i.e., as the order n of the PIN model grows large while its parameters $(\mathcal{A}, \rho, P, \phi)$ remain fixed.

Remark 3: In fact, the type $\Theta(t)$ associated with a PIN model with state space $\mathcal{A}$ and order n takes values in a finite subset

$$\mathcal{P}_n(\mathcal{A}) = \{\theta(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\} \,,$$

of the simplex $\mathcal{P}(\mathcal{A})$. Now, observe that, when the interaction pattern of a PIN model is the complete graph, the type process $\Theta(t) = \theta(X(t))$ is itself a Markov chain with state space $\mathcal{P}_n(\mathcal{A})$. Since the cardinality

$$|\mathcal{P}_n(\mathcal{A})| = \binom{n + |\mathcal{A}| - 1}{|\mathcal{A}| - 1} \leq (n+1)^{|\mathcal{A}|} \,,$$

grows polynomially fast in the order n while the cardinality $|\mathcal{X}| = |\mathcal{A}|^n$ of the configuration space $\mathcal{X}$ grows exponentially fast, the special case of complete interaction pattern provides a formidable reduction of complexity. In contrast, when the interaction network is not a complete graph, the type process $\Theta(t) = \theta(X(t))$ is in general not Markovian, while the process $X(t)$ always is. This creates a specific technical challenge that will be addressed in the rest of the paper.

We now define the notions of *limit drift* and *mean-field Lyapunov function* for a PIN model with a general interaction pattern.

Definition 1: The *limit drift* of a PIN model with parameters $(\mathcal{A}, \rho, P, \phi)$ is the function

$$\overline{D} : \mathcal{P}(\mathcal{A}) \to \mathbb{R}^{\mathcal{A}} \,,$$

mapping a probability vector θ into the vector

$$\overline{D}(\theta) := \left((1-\rho)P^\top + \rho \sum_{\ell \in \mathcal{A}} \theta_\ell \phi(\ell)^\top - I \right) \theta \,. \qquad (2)$$

Definition 2: A *mean-field Lyapunov function* of a PIN model with limit drift $\overline{D}$ is a differentiable map

$$V : \mathcal{P}(\mathcal{A}) \to \mathbb{R} \,,$$

such that

$$\nabla V(\theta) \cdot \overline{D}(\theta) \leq 0 \,, \qquad \forall \theta \in \mathcal{P}(\mathcal{A}) \,. \qquad (3)$$

Example 1 (continued): We compute now the limit drift of the SIS epidemic model introduced earlier and propose two possible mean-field Lyapunov functions. We use multiple times below the fact that $\theta_0 = 1 - \theta_1$.

From definition (3), we compute as follows:

$$\begin{aligned}
\overline{D}(\theta_0, \theta_1) = \;& -\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + (1-\rho) \begin{pmatrix} 1 - P_{01} & P_{10} \\ P_{01} & 1 - P_{10} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \\
& + \rho \left(\theta_0 \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + \theta_1 \begin{pmatrix} 1 - \phi_{01}(1) & 0 \\ \phi_{01}(1) & 1 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \right) \\
= \;& \begin{pmatrix} (1-\rho)(-P_{01}\theta_0 + P_{10}\theta_1) - \rho\phi_{01}(1)\theta_0\theta_1 \\ (1-\rho)(P_{01}\theta_0 - P_{10}\theta_1) + \rho\phi_{01}(1)\theta_0\theta_1 \end{pmatrix} .
\end{aligned}$$

As expected, the two entries $\overline{D}_0(\theta_0, \theta_1)$, $\overline{D}_1(\theta_0, \theta_1)$ of the limit drift sum up to zero. By letting $b := \rho\phi_{01}(1)$, that is the contagion probability, $c := (1-\rho)P_{10}$, that is the recovery probability, and $\alpha := (1-\rho)P_{01}$, that is the spontaneous infection probability, we get that

$$\overline{D}_1(\theta_0, \theta_1) = \alpha\theta_0 - c\theta_1 + b\theta_0\theta_1 = \alpha + (b - c - \alpha)\theta_1 - b\theta_1^2 \,.$$

A mean-field Lyapunov function for the SIS model is then

$$V_1(\theta_0, \theta_1) = -\left(\alpha\theta_1 + \frac{1}{2}(b - c - \alpha)\theta_1^2 - \frac{1}{3}b\theta_1^3 \right) \,,$$

since

$$\nabla V_1(\theta) \cdot \overline{D}(\theta) = -\left(\alpha + (b - c - \alpha)\theta_1 - b\theta_1^2\right)^2 \leq 0 \,.$$

Another possible mean-field Lyapunov function is

$$V_2(\theta_0, \theta_1) = \theta_1 - (z'' - z') \ln(\theta_1 - z'') \,,$$

where $z' < 0 < z''$ are the roots of the polynomial

$$p(z) = \alpha + (b - c - \alpha)z - bz^2 \,. \qquad (4)$$

Indeed, we have

$$\nabla V_2(\theta) \cdot \overline{D}(\theta) = -b(\theta_1 - z'')^2 \leq 0 \,.$$

The functions $\nabla V_1(\theta) \cdot \overline{D}(\theta)$ and $\nabla V_2(\theta) \cdot \overline{D}(\theta)$ are plotted in Figure 1, in the special case when $\alpha = 0.05$, $b = 1$, and $c = 0.3$.

In the special case of fully mixed populations, i.e., when the interaction pattern is the complete graph, Kurtz's theorem [15] (see also [17]) ensures that the type process $\Theta(\lfloor n\tau \rfloor)$ of a PIN model on a complete interaction pattern converges in

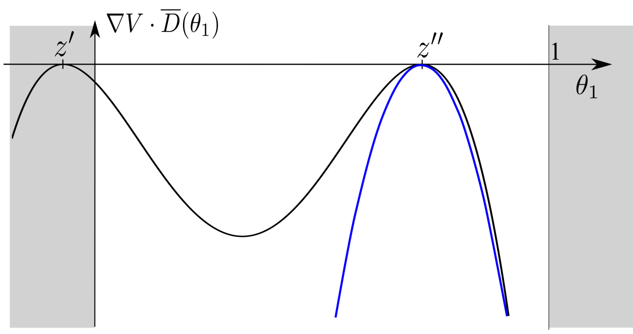

Fig. 1: Graphs of the functions $\nabla V_1(\theta) \cdot \overline{D}(\theta)$ (black) and $\nabla V_2(\theta) \cdot \overline{D}(\theta)$ (blue) for the two mean-field Lyapunov functions $V_1(\theta)$ and $V_2(\theta)$ proposed in Example 1. They are plotted as functions of the variable θ_1 in the special case when $\alpha = 0.05$, $b = 1$, and $c = 0.3$.

probability as the order n grows large to the solution $\theta(\tau)$ of the mean-field ODE[1]

$$\dot{\theta} = \overline{D}(\theta)\,, \tag{5}$$

over bounded time horizons $\tau \in [0, T]$. Hence, by Definition 2, a mean-field Lyapunov function $V(\theta)$ for a PIN model is a differentiable Lyapunov function for the mean-field ODE (5).

Notice that, if $V(\theta)$ is a mean-field Lyapunov function, then the celebrated Barbashin-Krasovskii-LaSalle invariance principle for deterministic dynamical systems [36, Theorem 4.4] implies that, for every initial condition, the solution of the mean-field ODE (5) converges, as t grows large, to a subset of the set[2]

$$\mathcal{Z} = \{\theta \in \mathcal{P}(\mathcal{A}) : \nabla V(\theta) \cdot \overline{D}(\theta) = 0\}\,. \tag{6}$$

In particular, this implies that the so-called Birkhoff center of the mean-field ODE (5), i.e., the closure $\overline{\mathcal{R}}$ of the set $\mathcal{R}$ of its recurrent points [16, Section 12.6.1], is a subset of $\mathcal{Z}$.

Moreover, [16, Theorem 12.6.2] implies that every sequence of stationary probability distributions of the type process $\Theta(t)$ of the PIN model of order n concentrates on the Birkhoff center $\overline{\mathcal{R}}$ of the mean-field ODE (5).[3] This leads to the following result for PIN models on fully mixed populations.

Theorem 0: Consider a sequence of stochastic PIN models with the same parameters $(\mathcal{A}, \rho, P, \phi)$ on complete interaction patterns of order $n \geq 2$. For every $n \geq 2$, let μ_n be a stationary distribution. Assume that the limit drift $\overline{D}(\theta)$ is Lipschitz-continuous on $\mathcal{P}(\mathcal{A})$ and that there exists a mean-field Lyapunov function $V : \mathcal{P}(\mathcal{A}) \to \mathbb{R}$. Then,

$$\lim_{n\to+\infty} \mu_n\{\mathbf{x} \in \mathcal{X} : \nabla V(\theta(\mathbf{x})) \cdot \overline{D}(\theta(\mathbf{x})) > -\delta\} = 1\,, \tag{7}$$

for every $\delta > 0$.

[1]Note that, since the limit drift $\overline{D}(\theta)$ is Lipschitz-continuous on $\mathcal{P}(\mathcal{A})$, the Cauchy problem associated to the mean-field ODE (5) is well-posed for every initial condition $\theta(0)$ in $\mathcal{P}(\mathcal{A})$.

[2]By this it is meant that there exists a nonempty subset $\mathcal{K} \subseteq \mathcal{Z}$ — specifically, the largest invariant subset of $\mathcal{Z}$— such that $\mathrm{dist}(\theta(t), \mathcal{K}) \to 0$ as t grows large, where $\mathrm{dist}(\theta, \mathcal{K}) = \inf\{||\theta - \tilde{\theta}|| : \tilde{\theta} \in \mathcal{K}\}$ denotes the distance of a point θ in $\mathcal{P}(\mathcal{A})$ from the set $\mathcal{K}$.

[3]Precisely, this means that, $\mu_n(\mathcal{O}) \to 1$ for every open set $\mathcal{O} \subseteq \mathcal{P}(\mathcal{A})$ such that $\overline{\mathcal{R}} \subseteq \mathcal{O}$.

The previous theorem shows that every stationary distribution of the PIN model over a large scale complete interaction pattern admitting a mean-field Lyapunov function tends to concentrate on the configurations $\mathbf{x}$ for which $\theta(\mathbf{x}) \in \mathcal{Z}$, where $\mathcal{Z}$ is defined in (6). When applied to Example 1 this means that, when the population is large and the interaction pattern is complete, $\theta(\mathbf{x})$ tends to concentrate on the value z'' that is the positive root of the polynomial $\alpha+(b-c-\alpha)z-bz^2$. This is in agreement with the predictions of the mean-field ODE of the SIS model whose solutions all converge to z''.

In the following sections, we shall develop a novel methodology that leads to a completely independent proof of Theorem 0. Our approach does not rely on Kurtz's theorem nor on its extensions and leads to a remarkable generalization of Theorem 0 to PIN models over non complete interaction patterns.

III. Main results

In this section, we present our two main results regarding concentration of stationary distributions of stochastic PIN models in the large-scale limit on non-necessarily complete interaction patterns.

A. Asymptotically totally mixing networks

We start by introducing an index in graph theory that will play a crucial role in our analysis. This index, called total mixing gap, evaluates how close a graph is from the complete graph.

Definition 3: Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a nonempty graph. For two subsets of nodes $\mathcal{S}, \mathcal{U} \subseteq \mathcal{V}$, let

$$\mathcal{E}_{\mathcal{S}\mathcal{U}} = \{(s, u) \in \mathcal{E} \;:\; s \in \mathcal{S},\, u \in \mathcal{U}\}\,,$$

be the set of directed links pointing from nodes in $\mathcal{S}$ to nodes in $\mathcal{U}$. Define

$$W^{(1)}_{\mathcal{G}} := \max_{\mathcal{U}\subseteq\mathcal{V}} \left| \frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{m} - \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)} \right|,$$

and

$$W^{(2)}_{\mathcal{G}} := \max_{\substack{\mathcal{S},\mathcal{U}\subseteq\mathcal{V}\\ \mathcal{S}\cap\mathcal{U}=\emptyset}} \left| \frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)} \right|,$$

where $n = |\mathcal{V}|$, and $m = |\mathcal{E}|$. Then the *total mixing gap* of $\mathcal{G}$ is defined as

$$W_{\mathcal{G}} := \max\left\{W^{(1)}_{\mathcal{G}}, W^{(2)}_{\mathcal{G}}\right\}. \tag{8}$$

Clearly, for every nonempty graph $\mathcal{G}$, we have that $W_{\mathcal{G}} \geq 0$. On the other hand, since $|\mathcal{E}_{\mathcal{U}\mathcal{U}}| \leq m$ and $|\mathcal{U}| \leq n$, then

$$W^{(1)}_{\mathcal{G}} = \max_{\mathcal{U}\subseteq\mathcal{V}} \left| \frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{m} - \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)} \right| \leq \max\{1, 1\} = 1\,.$$

Moreover, since $|\mathcal{E}_{\mathcal{S}\mathcal{U}}| \leq m$ and $|\mathcal{S}||\mathcal{U}| \leq n^2/4$ for every $\mathcal{S}, \mathcal{U} \subseteq \mathcal{V}$ such that $\mathcal{S} \cap \mathcal{U} = \emptyset$, we get that

$$\begin{aligned} W^{(2)}_{\mathcal{G}} &= \max_{\substack{\mathcal{S},\mathcal{U}\subseteq\mathcal{V}\\ \mathcal{S}\cap\mathcal{U}=\emptyset}} \left| \frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)} \right| \\ &\leq \max\{1, n/(4(n-1))\} = 1\,, \end{aligned}$$

where the last equality follows from the fact that $n \leq 2(n-1)$ for $n \geq 2$. Hence,

$$0 \leq W_{\mathcal{G}} \leq 1\,, \tag{9}$$

for every nonempty graph $\mathcal{G}$. We now provide estimates of the total mixing gap for some graphs, showing that the inequality (9) is tight.

Example 2: In the special case when $\mathcal{G}$ is the complete graph, it is clear that $W_{\mathcal{G}} = 0$. In fact, the converse is also true: if $W_{\mathcal{G}} = 0$, then $W_{\mathcal{G}}^{(2)} = 0$ and so for every two nodes $s \neq u$ in $\mathcal{V}$ we have that

$$\left| \frac{|\mathcal{E}_{\{s\}\{u\}}|}{m} - \frac{1}{n(n-1)} \right| \leq W_{\mathcal{G}} = 0\,,$$

and since $|\mathcal{E}_{\{s\}\{u\}}| \in \{0,1\}$ and $m > 0$ (because the graph is nonempty), it must be that $m = n(n-1)$, i.e., $\mathcal{G}$ is the complete graph.

Example 3: For the star graph with $n-1$ undirected links connecting a single node s to all other nodes $\mathcal{U} := \mathcal{V} \setminus \{s\}$ we get

$$\begin{array}{rcl} W_{\mathcal{G}} & \geq & W_{\mathcal{G}}^{(1)} \\ & \geq & \left| \dfrac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{m} - \dfrac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)} \right| \\ & = & \left| \dfrac{0}{2(n-1)} - \dfrac{(n-1)(n-2)}{n(n-1)} \right| \quad = \quad 1 - \dfrac{2}{n}\,, \end{array}$$

that converges to 1 as the order n grows large.

Example 4: Consider a toroidal grid graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node set $\mathbb{Z}_h^2$ for some even integer $h \geq 2$ (where $\mathbb{Z}_h$ stands for the set of integers modulo h) and link set

$$\mathcal{E} = \{(i,j) \in \mathbb{Z}_h^2 \times \mathbb{Z}_h^2 : \, |i_1 - j_1| + |i_2 - j_2| = 1\}\,.$$

Let $\mathcal{U}^* = \{i \in \mathbb{Z}_h^2 : \, 0 \leq i_1 < h/2\}$ and $\mathcal{S}^* = \mathbb{Z}_h^2 \setminus \mathcal{U}^*$. Then, $n = h^2$, $m = 4n$, $|\mathcal{U}^*| = n/2$ and $|\mathcal{E}_{\mathcal{U}^*\mathcal{S}^*}| = 4h$. Hence, we get that

$$W_{\mathcal{G}} \geq W_{\mathcal{G}}^{(2)} \geq \left| \frac{|\mathcal{E}_{\mathcal{S}^*\mathcal{U}^*}|}{m} - \frac{|\mathcal{S}^*||\mathcal{U}^*|}{n(n-1)} \right| = \left| \frac{4h}{4n} - \frac{n^2/4}{n(n-1)} \right|\,,$$

that converges to $1/4$ as the order $n = h^2$ grows large.

The following definition captures a fundamental asymptotic property of sequences of graphs.

Definition 4: A sequence of graphs $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ is an *asymptotically totally mixing (ATM)* network if

$$\lim_{n \to +\infty} W_{\mathcal{G}_n} = 0\,.$$

Notice that a sequence $\mathcal{K}_n$ of complete graphs with increasing order n is an ATM network since $W_{\mathcal{K}_n} = 0$ for all $n \geq 2$, as shown in Example 2. In contrast, a sequence of star graphs $\mathcal{G}_n$ with increasing order n is not an ATM network, since in this case $W_{\mathcal{G}_n} \to 1$ as n grows large, as shown in Example 3, nor is a sequence of toroidal grids with increasing order n, since in that case $\liminf_n W_{\mathcal{S}_n} \geq 1/4$ as n grows large, as shown in Example 4.

In fact, an important example of an ATM network is provided by the Erdös-Rényi random graphs. These are defined as follows: for a positive integer n and a parameter p in $(0,1]$, let $\mathcal{G}(n,p)$ be the random undirected graph of order n whereby pairs of distinct nodes are connected with probability p, independently from one another. Then, we have the following result.

Proposition 1: If $np^3 \to +\infty$ as n grows large, then the sequence of Erdös-Rényi random graphs $\mathcal{G}(n,p)$ is ATM with probability 1.

Proof: See Appendix B. ∎

B. Concentration of PIN models on ATM interaction networks

We are now in a position to formally state the two main technical results of this paper. The first one provides the extension of Theorem 0 to PIN models on ATM interaction networks.

Theorem 1: Consider a sequence of PIN models with the same parameters $(\mathcal{A}, \rho, P, \phi)$ on an ATM interaction network $\mathcal{G}_n$. For every $n \geq 2$, let μ_n be a stationary distribution. Denote by $\overline{D}$ the limit drift and assume there exists a mean-field Lyapunov function $V : \mathcal{P}(\mathcal{A}) \to \mathbb{R}$ of class-$\mathcal{C}^2$. Then

$$\lim_{n \to +\infty} \mu_n \left\{ \mathbf{x} \in \mathcal{X} : \, \nabla V(\theta(\mathbf{x})) \cdot \overline{D}(\theta(\mathbf{x})) > -\delta \right\} = 1\,, \tag{10}$$

for every $\delta > 0$.

Proof: See Section V-C. ∎

As illustrated in the next section, there are cases when Theorem 1 cannot be directly applied, e.g., since one cannot find a class-$\mathcal{C}^2$ mean-field Lyapunov function for the PIN model. In this case, we sharpen our technique by resorting to an approximate mean-field Lyapunov function. This result will then be applied to the SIRS epidemic model in Section IV.

Definition 5: Consider a family of PIN models with parameters $(\mathcal{A}, \rho, P_\alpha, \phi_\alpha)_{\alpha \geq 0}$, and let $\overline{D}_\alpha(\theta)$, for $\alpha \geq 0$, be their limit drift. An *approximate mean-field Lyapunov function* for such a family is a family of differentiable functions

$$V_\alpha : \mathcal{P}(\mathcal{A}) \to \mathbb{R}\,, \qquad \alpha > 0\,,$$

such that there exist two nonnegative-valued functions $h : \mathcal{P}(\mathcal{A}) \to \mathbb{R}_+$ and $\zeta : \mathbb{R}_+ \to \mathbb{R}_+$ where h is continuous on $\mathcal{P}(\mathcal{A})$ and

$$\lim_{\alpha \to 0} \zeta(\alpha) = 0\,, \tag{11}$$

such that

$$\nabla V_\alpha(\theta) \cdot \overline{D}_\alpha(\theta) \leq -h(\theta) + \zeta(\alpha)\,, \qquad \forall \theta \in \mathcal{P}(\mathcal{A})\,. \tag{12}$$

We can now state the following result, generalizing Theorem 1.

Theorem 2: Consider a family of PIN models with parameters $(\mathcal{A}, \rho, P_\alpha, \phi_\alpha)_{\alpha \geq 0}$ on an ATM interaction network $\mathcal{G}_n$. Assume that there exists a class-$\mathcal{C}^2$ approximate mean-field Lyapunov function V_α as in Definition 5. For every $\alpha > 0$ and $n \geq 2$, let $\mu_n^{(\alpha)}$ be a stationary distribution. Then,

$$\lim_{\alpha \to 0} \liminf_{n \to +\infty} \mu_n^{(\alpha)} \{ \mathbf{x} \in \mathcal{X} : \, h(\theta(\mathbf{x})) < \delta \} = 1\,, \tag{13}$$

for every $\delta > 0$.

Proof: See Section V-C. ∎

Remark 4: Observe that, if a class-$\mathcal{C}^2$ function $V_0 : \mathcal{P}(\mathcal{A}) \to \mathbb{R}_+$ existed such that both $V_\alpha(\theta) \xrightarrow{\alpha\to 0} V_0(\theta)$ and $\nabla V_\alpha(\theta) \xrightarrow{\alpha\to 0} \nabla V_0(\theta)$ uniformly on $\mathcal{P}(\mathcal{A})$, then taking the limit of both sides of (12) would yield

$$\nabla V_0(\theta) \cdot \overline{D}_0(\theta) \le -h(\theta) \le 0\,, \qquad \forall \theta \in \mathcal{P}(\mathcal{A})\,.$$

In this case, $V_0(\theta)$ would be a mean-field Lyapunov function for the PIN model with parameters $(\mathcal{A}, \rho, P_0, \phi_0)$ and Theorem 1 could be applied directly. However, there are families of PIN models such as the SIRS epidemic model described in Example 7 of Section IV-C, which admit approximate mean-field Lyapunov function V_α that does not even admits a limit as α vanishes. Theorem 2 proves useful in addressing exactly these cases.

IV. Examples of applications

In this section, we introduce some classes of PIN models and apply the concentration results of the previous section.

A. Forgetful PIN models

We first consider the special case of PIN models where the entries of the interaction transition tensor $\phi_{ij}(\ell)$ do not depend on the current state i. Precisely, we assume that there exists a row-stochastic matrix R such that

$$\phi_{ij}(\ell) = R_{\ell j}\,, \qquad \forall i, j, \ell \in \mathcal{A}\,.$$

Notice that the limit drift for these models has the form

$$\overline{D}(\theta) = (S^\top - I)\theta\,, \qquad \forall \theta \in \mathcal{P}(\mathcal{A})\,, \tag{14}$$

where

$$S = (1-\rho)P + \rho R\,, \tag{15}$$

is row-stochastic, i.e., the matrix associated with the linear map $\overline{D}(\theta)$ is the transpose of a Laplacian matrix. These models are referred to as forgetful PIN models and arise in evolutionary game dynamics [8], [9].

Assume that the row-stochastic matrix $S = (1-\rho)P + \rho R$ is irreducible. It then follows from the Perron-Frobenius theorem that S admits a unique invariant probability vector $\pi = S^\top \pi$ in $\mathcal{P}(\mathcal{A})$. For such models over ATM interaction networks, Theorem 1 yields a concentration of the type process on the vector π. Precisely, we have the following result.

Proposition 2: Consider a sequence of forgetful PIN models with the same parameters $(\mathcal{A}, \rho, P, \phi)$ over an ATM interaction network $\mathcal{G}_n$. Assume that the stochastic matrix S in (15) is irreducible and let $\pi = S^\top \pi$ in $\mathcal{P}(\mathcal{A})$ be its unique invariant probability vector. Let μ_n be a stationary distribution for all $n \ge 2$. Then,

$$\lim_{n\to+\infty} \mu_n\{\mathbf{x} \in \mathcal{X} \;:\; \|\theta(\mathbf{x}) - \pi\|_2 < \delta\} = 1\,.$$

for every $\delta > 0$.

Proof: We first construct a quadratic function $V(\theta)$ that is positive definite with respect to π and such that (3) holds true. To this aim, we consider the Laplacian matrix $L := I - S^\top$ and we recall some basic well-known properties:

- all eigenvalues λ of L are such that $\mathbb{Re}(\lambda) \ge 0$;
- the only eigenvalue on the imaginary axis is 0 and this is simple with left and right eigenvectors being, respectively, $\mathbf{1}$ and π.

As a consequence, the spectrum of the matrix

$$A = L + \pi \mathbf{1}^\top$$

lies in the open half-plane $\mathbb{Re}(\lambda) > 0$ and, consequently, the Lyapunov equation

$$\Pi A + A^\top \Pi = I\,, \tag{16}$$

admits a solution Π in $\mathbb{R}^{\mathcal{A}\times\mathcal{A}}$ that is symmetric and positive definite. Consider now the function $V : \mathcal{P}(\mathcal{A}) \to \mathbb{R}$, defined as

$$V(\theta) = (\theta - \pi)^\top \Pi (\theta - \pi)\,, \qquad \forall \theta \in \mathcal{P}(\mathcal{A})\,. \tag{17}$$

Clearly, $V(\theta)$ is of class $\mathcal{C}^2$. Moreover, it follows from (14) that

$$\begin{aligned}
\nabla V(\theta) \cdot \overline{D}(\theta) &= -2(\theta - \pi)^\top \Pi L \theta \\
&= -2(\theta - \pi)^\top \Pi (L + \pi \mathbf{1}^\top)(\theta - \pi) \\
&= -(\theta - \pi)^\top \left(\Pi A + A^\top \Pi\right)(\theta - \pi) \\
&= -||\theta - \pi||_2^2 \le 0\,.
\end{aligned}$$

This shows that (3) is satisfied, so that $V(\theta)$ is a mean-field Lyapunov function for the considered PIN model. The result now follows from Theorem 1. ■

Explicit examples of forgetful PIN models come from evolutionary game theory, as for instance the following best response learning model.

Example 5: Consider a 2-player symmetric game with set of actions $\mathcal{A}$ and payoff matrix U in $\mathbb{R}^{\mathcal{A}\times\mathcal{A}}$ (i.e., U_{ij} is the payoff obtained by the first player when she plays i and her opponent plays j). We assume the agents to interact in a pairwise fashion and to choose best response actions with respect to the above specified 2-player game. Precisely, the jump interaction probabilities are given by

$$R_{\ell j} = \begin{cases} |\mathcal{B}(\ell)|^{-1} & \text{if } j \in \mathcal{B}(\ell) \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{B}(\ell) = \underset{k\in\mathcal{A}}{\operatorname{argmax}}\, U_{k\ell}$ is the best response set.

B. Binary PIN models

We now analyze in detail general PIN models with binary state space $\mathcal{A} = \{0, 1\}$, encompassing the SIS model proposed in Example 1. Notice that in this case seven parameters are sufficient to specify the model: the interaction probability ρ, two mutation probabilities P_{01} and P_{10}, and the four elements of the interaction transition tensor $\phi_{01}(0)$, $\phi_{01}(1)$, $\phi_{10}(0)$, and $\phi_{10}(1)$. In particular, the limit drift takes the following form

$$\begin{aligned}
\overline{D}_1(\theta) = {} & (1-\rho)\left(P_{01}\theta_0 - P_{10}\theta_1\right) \\
& +\rho\left(\phi_{01}(0)\theta_0^2 + (\phi_{01}(1) - \phi_{10}(0))\theta_0\theta_1 - \phi_{10}(1)\theta_1^2\right).
\end{aligned}$$

By substituting $\theta_0 = 1 - \theta_1$, we get that

$$\overline{D}_1(\theta) = p(\theta_1)$$

where p is the quadratic polynomial

$$p(z) = a_0 + a_1 z + a_2 z^2 \tag{18}$$

with coefficients

$$\begin{aligned} a_0 &= (1-\rho)P_{01} + \rho\phi_{01}(0) \\ a_1 &= -(1-\rho)\left(P_{01} + P_{10}\right) + \rho(-2\phi_{01}(0) + \phi_{01}(1) - \phi_{10}(0)) \\ a_2 &= \rho\left(\phi_{01}(0) - \phi_{01}(1) + \phi_{10}(0) - \phi_{10}(1)\right) . \end{aligned} \tag{19}$$

Notice that

$$\begin{aligned} p(0) &= (1-\rho)P_{01} + \rho\phi_{01}(0) \geq 0 \,, \\ p(1) &= -(1-\rho)P_{10} - \rho\phi_{10}(1) \leq 0 \,, \end{aligned}$$

so that there always exists at least one zero of $p(z)$ in the interval $[0, 1]$. When $p(0) > 0$ and $p(1) < 0$ (this happens for instance when $\rho < 1$ and the two mutation terms P_{01} and P_{10} are both nonzero) such zero is unique and is denoted by z^*. The next result shows that, under suitable assumptions, z^* is exactly where concentration of the component 1 of the type process takes place.

Proposition 3: Consider a sequence of binary PIN models with the same parameters $(\mathcal{A}, \rho, P, \phi)$ over an ATM interaction network $\mathcal{G}_n$. Let μ_n be a stationary distribution of the PIN model for all $n \geq 2$. If the polynomial $p(z)$ defined in (18) admits a unique zero z^* in $[0, 1]$, then

$$\lim_{n\to+\infty} \mu_n\{\mathbf{x} \in \mathcal{X} \ : \ |\theta_1(\mathbf{x}) - z^*| < \delta\} = 1 \,,$$

for every $\delta > 0$.

Proof: Consider $V : \mathcal{P}(\mathcal{A}) \to \mathbb{R}$ defined by

$$V(\theta) = -\left(a_0\theta_1 + \frac{1}{2}a_1\theta_1^2 + \frac{1}{3}a_2\theta_1^3\right) \,, \qquad \forall\theta \in \mathcal{P}(\mathcal{A}) \,.$$

Notice that

$$\nabla V(\theta) \cdot \overline{D}(\theta) = -p(\theta_1)^2 \,.$$

Therefore,

$$\{\theta \in \mathcal{P}(\mathcal{A}) \ : \ \nabla V(\theta) \cdot \overline{D}(\theta) = 0\} = \{(1 - z^*, z^*)\}$$

and the result now follows from Theorem 1. ■

Example 1 (continued): For the SIS epidemic model introduced in Example 1, the polynomial in (18) coincides with (4). When $\alpha > 0$, $b > 0$, and $c > 0$, then, by Proposition 3, the stationary distribution of the associated PIN model over an ATM interaction network concentrates around the set of configurations $\mathbf{x}$ such that $\theta_1(\mathbf{x}) = z^*(\alpha)$, where $z^*(\alpha)$ is the root of $p(z)$ in $[0, 1]$, in which we highlighted its dependence on α. Notice that

$$\lim_{\alpha\to 0} z^*(\alpha) = z^* = \max\{1 - c/b, 0\} \,. \tag{20}$$

Considering the limit also in α we further get

$$\lim_{\alpha\to 0^+} \lim_{n\to+\infty} \mu_n^{(\alpha)}\{\mathbf{x} \ : \ |\theta_1(\mathbf{x}) - z^*| < \delta\} = 1 \,,$$

for every $\delta > 0$. We notice that z^* coincides with the asymptotically stable equilibrium of the corresponding mean-field ODE (5). In particular, this shows that the well known bifurcation phenomenon according which for $c > b$ the epidemics dies out while for $c < b$ it becomes endemic at level $1 - c/b$, holds true not only for complete interaction patterns, but also for general ATM interaction networks, as for instance Erdös-Rényi random graphs.

Other popular examples of PIN models falling into this family are reported below.

Example 6 (Voter/Anti-voter models): In binary opinion dynamics models, the two states 0 and 1 represent two different opinions. In the *noisy voter model* [37], [38], agents modify their opinion in pairwise interactions by copying the opinion of the agent they meet or because of a mutation (equal for the two states). This is represented by the following choices of the parameters:

$$\phi_{01}(1) = \phi_{10}(0) = 1 \,, \qquad \phi_{01}(0) = \phi_{10}(1) = 0 \,,$$

and $P_{01} = P_{10} = 1$. In this case, we have

$$\overline{D}_1(\theta) = (1-\rho)(\theta_0 - \theta_1).$$

In the *anti-voter* model [2], [6] no mutation is present, namely $\rho = 1$, and agents take the opposite opinion of the agent they meet. Precisely,

$$\phi_{01}(1) = \phi_{10}(0) = 0 \,, \qquad \phi_{01}(0) = \phi_{10}(1) = 1 \,.$$

In this case, we have

$$\overline{D}_1(\theta) = \theta_0^2 - \theta_1^2.$$

Notice that in both examples we have $z^* = 1/2$. As for Example 1, this result matches the asymptotic behavior of the corresponding mean-field ODEs. We notice how, for the noisy voter model, the concentration result does not depend on the mutation probability $1 - \rho$ as long as $1 - \rho > 0$. When $\rho = 1$ the system is no longer ergodic, invariant distributions live on the two extreme points 0 and 1 (as they are absorbing states of the chain) for every value of n.

C. Non-binary models

Finally, we present an example that does not fit in any of the two categories above.

Example 7 (SIRS epidemic model): In the SIRS epidemic model we have $\mathcal{A} = \{0, 1, 2\}$ where $\mathbf{x}_v = 0$ means the individual v is susceptible, $\mathbf{x}_v = 1$ means that the individual v is infected and $\mathbf{x}_v = 2$ means that the individual v is recovered. The non-zero off-diagonal terms in P are P_{01}, P_{12}, and P_{20}, the conditional probability of, respectively, spontaneous infection, spontaneous recovery, and spontaneous return to susceptibility. As in the SIS model, the only off diagonal term of $\phi(\ell)$ is $\phi_{01}(1)$ the conditional probability of contagion transmission. It is convenient to set the following variables:

$$b := \rho\phi_{01}(1) \,, \quad c := (1-\rho)P_{12} \,, \quad d := (1-\rho)P_{20} \,,$$

$$\alpha := (1-\rho)P_{01} \,.$$

We assume that $\alpha > 0$ and moreover that $b > c > 0$ and $d > 0$. We denote by $\overline{D}_\alpha(\theta)$ its limit drift. It can be seen that

$$\begin{aligned}\overline{D}_{1,\alpha}(\theta) &= b(1-\theta_1-\theta_2)\theta_1 - c\theta_1 + \alpha(1-\theta_1-\theta_2)\\ &= (\theta_1+\alpha/b)[b(1-\theta_1-\theta_2)-c] + \alpha c/b\\ &= -b(\theta_1+\alpha/b)[(\theta_1-z_1^*)+(\theta_2-z_2^*)] + \alpha c/b\,,\\ \overline{D}_{2,\alpha}(\theta) &= c\theta_1 - d\theta_2 = c(\theta_1-z_1^*) - d(\theta_2-z_2^*)\,,\end{aligned}$$

where

$$z_1^* = \frac{d(b-c)}{b(d+c)}, \quad z_2^* = \frac{c(b-c)}{b(d+c)}\,. \tag{21}$$

Now, consider the family of functions

$$V_\alpha(\theta_1,\theta_2) := \theta_1 - (z_1^*+\alpha/b)\ln(\theta_1+\alpha/b) + \frac{b}{2c}(\theta_2-z_2^*)^2\,.$$

Notice that these functions are of class $\mathcal{C}^2$ on $\mathcal{P}(\mathcal{A})$ for $\alpha > 0$ and have partial derivatives

$$\frac{\partial V_\alpha}{\partial\theta_1}(\theta) = \frac{\theta_1-z_1^*}{\theta_1+\alpha/b}, \qquad \frac{\partial V_\alpha}{\partial\theta_2}(\theta) = \frac{b}{c}(\theta_2-z_2^*)\,. \tag{22}$$

Straightforward computations now yield

$$\begin{aligned}&\nabla V_\alpha(\theta)\cdot\overline{D}_\alpha(\theta)\\ &\quad= -b\,(\theta_1-z_1^*)^2 - \frac{bd}{c}\,(\theta_2-z_2^*)^2 + \frac{\alpha c}{b}\frac{\theta_1-z_1^*}{\theta_1+\frac{\alpha}{b}}\\ &\quad\le -b\,(\theta_1-z_1^*)^2 - \frac{bd}{c}\,(\theta_2-z_2^*)^2 + \frac{\alpha c}{b}\,,\end{aligned}$$

where the last inequality follows form the fact that the function

$$g(z) = \frac{z-z_1^*}{z+\alpha/b}$$

is increasing in $z \in [0,1]$ and hence $g(z) \le 1$.

We are then in a position to apply Theorem 2 by taking

$$h(\theta) = -b\,(\theta_1-z_1^*)^2 - \frac{bd}{c}\,(\theta_2-z_2^*)^2\,, \quad \zeta(\alpha) = \alpha c/b\,,$$

and to conclude that (13) holds true. This essentially says that the invariant measure $\mu_n^{(\alpha)}$ converges, as n grows large and as α tends to zero, to a probability measure whose projection by $\theta(\cdot)$ coincides with the Dirac delta distribution centered in

$$\theta^* = (1-z_1^*-z_2^*, z_1^*, z_2^*)\,,$$

where z_1^* and z_2^* are defined in (21).

V. Proofs of the main concentration results

In this section, we present all technical results and, in particular, we prove Theorems 1 and 2. We first derive general approximation results that are valid for any PIN model on arbitrary interaction patterns. We then restrict to ATM interaction networks and we establish the two large scale concentration results.

A. Mean drift

Consider a PIN model with interaction pattern $\mathcal{G} = (\mathcal{V},\mathcal{E})$ and parameters $(\mathcal{A},\rho,P,\phi)$, and let $\mathcal{X}$ be its configuration space. A key quantity in the analysis of the PIN model is its *mean drift*, defined as the function $D : \mathcal{X} \to \mathbb{R}^{\mathcal{A}}$ mapping a configuration $\mathbf{x}$ in $\mathcal{X}$ into the vector

$$D(\mathbf{x}) := n\mathbb{E}\left[\theta(X(t+1)) - \theta(X(t))|X(t) = \mathbf{x}\right]\,. \tag{23}$$

Entries $D_i(\boldsymbol{x})$ represent the conditional expected variation of the number of agents in a given state i, when the current configuration of the PIN model is $X(t) = \mathbf{x}$. Notice that, since $\sum_{i\in\mathcal{A}}\theta_i(\mathbf{x}) = 1$ for every configuration $\mathbf{x}$ in $\mathcal{X}$, we have that

$$\begin{aligned}0 &= n\mathbb{E}\left[\sum_{i\in\mathcal{A}}(\theta_i(X(t+1))-\theta_i(X(t)))\,|X(t)=\mathbf{x}\right]\\ &= \sum_{i\in\mathcal{A}} D_i(\mathbf{x})\,,\end{aligned}$$

i.e., $D(\mathbf{x})$ is always a zero-sum vector.

It also proves useful to introduce the notion of *boundary* of a configuration $\mathbf{x}$ in $\mathcal{X}$, defined as the vector $\xi(\mathbf{x})$ of the empirical frequencies of the pairs of states in $\mathbf{x}$ that are connected by links of $\mathcal{G}$. Formally, let $\xi : \mathcal{X} \to \mathcal{P}(\mathcal{A}\times\mathcal{A})$ be the function mapping a configuration $\mathbf{x}$ in $\mathcal{X}$ to $\xi(\mathbf{x})$ in $\mathcal{P}(\mathcal{A}\times\mathcal{A})$ with entries

$$\xi_{ij}(\mathbf{x}) := \frac{1}{m}|\{(u,v)\in\mathcal{E} \,:\, \mathbf{x}_u = i,\ \mathbf{x}_v = j\}|\,, \qquad i,j\in\mathcal{A}\,, \tag{24}$$

where we recall that $m = |\mathcal{E}|$ is the number of edges of the interaction pattern $\mathcal{G}$. Clearly, if $\mathcal{G}$ is undirected, then $\xi_{ij}(\mathbf{x}) = \xi_{ji}(\mathbf{x})$ for every two states i and j in $\mathcal{A}$ and every configuration $\mathbf{x}$ in $\mathcal{X}$.

Finally, we introduce the linear operator $\mathcal{Q} : \mathcal{P}(\mathcal{A}\times\mathcal{A}) \to \mathbb{R}^{\mathcal{A}}$ mapping a boundary ξ in $\mathcal{P}(\mathcal{A}\times\mathcal{A})$ into a zero-sum vector $\mathcal{Q}(\xi)$ in $\mathbb{R}^{\mathcal{A}}$ with entries

$$(\mathcal{Q}(\xi))_i := \sum_{\ell\in\mathcal{A}}\sum_{j\in\mathcal{A}}\xi_{j\ell}\phi_{ji}(\ell) - \sum_{\ell\in\mathcal{A}}\xi_{i\ell}\,, \qquad i\in\mathcal{A}\,. \tag{25}$$

Then, we have the following result.

Lemma 1: The mean drift of a PIN model with interaction network $\mathcal{G} = (\mathcal{V},\mathcal{E})$ and parameters $(\mathcal{A},\rho,P,\phi)$ satisfies

$$D(\mathbf{x}) = (1-\rho)(P^\top - I)\theta(\mathbf{x}) + \rho\mathcal{Q}(\xi(\mathbf{x})) \tag{26}$$

for every configuration $\mathbf{x}$ in $\mathcal{X}$.

Proof: Since the configuration $X(t)$ of a PIN model can change at most one of its entries at a time, we have that either

$$\theta(X(t+1)) - \theta(X(t)) = 0\,,$$

(when no change occurs) or

$$\theta(X(t+1)) - \theta(X(t)) = n^{-1}(\delta^j - \delta^i)\,,$$

for some $i \neq j$ in $\mathcal{A}$ (when an agent modifies its state from i to j). Call E_{ij} the latter event and notice that, because of the updating mechanism, we have that

$$\mathbb{P}(E_{ij}\,|\,X(t)=\mathbf{x}) = (1-\rho)\theta_i(\mathbf{x})P_{ij} + \rho\sum_{\ell\in\mathcal{A}}\xi_{i\ell}(\mathbf{x})\phi_{ij}(\ell)\,.$$

Thus,

$$\begin{aligned}
D_i(\mathbf{x}) &= \sum_{j\neq i}\mathbb{P}(E_{ji}\,|\,X(t)=\mathbf{x}) - \sum_{j\neq i}\mathbb{P}(E_{ij}\,|\,X(t)=\mathbf{x}) \\
&= (1-\rho)\sum_{j\neq i}\theta_j(\mathbf{x})P_{ji} + \rho\sum_{\ell\in\mathcal{A}}\sum_{j\neq i}\xi_{j\ell}(\mathbf{x})\phi_{ji}(\ell) \\
&\quad -(1-\rho)\sum_{j\neq i}\theta_i(\mathbf{x})P_{ij} - \rho\sum_{\ell\in\mathcal{A}}\sum_{j\neq i}\xi_{i\ell}(\mathbf{x})\phi_{ij}(\ell) \\
&= (1-\rho)\left(P^\top\theta(\mathbf{x})-\theta(\mathbf{x})\right)_i + \rho\left(\mathcal{Q}(\xi(\mathbf{x}))\right)_i ,
\end{aligned}$$

which yields the result. ■

Lemma 1 states that the mean drift $D(\mathbf{x})$ of a PIN model depends on the configuration $\mathbf{x}$ through both its type $\theta(\mathbf{x})$ and its boundary $\xi(\mathbf{x})$. This is a consequence of the particular structure of PIN models that allow for both individual mutations and pairwise interactions on the graph $\mathcal{G}$, but not, e.g., on higher order interactions.

Below, we discuss in general how the mean drift and the limit drift are related. We start with the following result.

Lemma 2: Consider a PIN model with interaction pattern $\mathcal{G}=(\mathcal{V},\mathcal{E})$ and parameters $(\mathcal{A},\rho,P,\phi)$. Let $D(\mathbf{x})$ be its mean drift and $\overline{D}(\theta)$ be its limit drift. Then,

$$||D(\mathbf{x})-\overline{D}(\theta(\mathbf{x}))||_1 \leq 2\rho||\xi(\mathbf{x})-\theta(\mathbf{x})\theta(\mathbf{x})^\top||_1 ,$$

for every configuration $\mathbf{x}$ in $\mathcal{X}$.

Proof: First, from definitions (2) and (25) we can rewrite the limit drift as

$$\overline{D}(\theta) = (1-\rho)(P^\top - I)\theta + \rho\mathcal{Q}(\theta\theta^\top)\,. \tag{27}$$

We now fix a configuration $\mathbf{x}$ in $\mathcal{X}$ and write θ for $\theta(\mathbf{x})$ and ξ for $\xi(\mathbf{x})$. From expressions (26) and (27), and the definition of $\mathcal{Q}$ we obtain

$$\begin{aligned}
||D(\mathbf{x})-\overline{D}(\theta)||_1 &= \rho||\mathcal{Q}(\xi)-\mathcal{Q}(\theta\theta^\top)||_1 \\
&= \rho\sum_{i\in\mathcal{A}}|(\mathcal{Q}(\xi-\theta\theta^\top)_i| \\
&\leq \rho\sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{A}}\sum_{\ell\in\mathcal{A}}|\xi_{j\ell}-\theta_j\theta_\ell|\phi_{ji}(\ell) \\
&\quad +\rho\sum_{i\in\mathcal{A}}\sum_{\ell\in\mathcal{A}}|\xi_{i\ell}-\theta_i\theta_\ell| \\
&= 2\rho||\xi-\theta\theta^\top||_1\,.
\end{aligned}$$

This proves the claim. ■

Lemma 2 provides an upper bound on the 1-norm distance between the mean drift $D(\mathbf{x})$ in a configuration $\mathbf{x}$ and the corresponding limit drift $\overline{D}(\theta(\mathbf{x}))$ of a PIN model. This is no more than twice the l_1-distance between the boundary $\xi(\mathbf{x})$ of $\mathbf{x}$ and the product distribution $\theta(\mathbf{x})\theta(\mathbf{x})^\top$. It is then clear that the limit drift $\overline{D}(\theta(\mathbf{x}))$ provides a good approximation of the mean drift $D(\mathbf{x})$ when the boundary $\xi(\mathbf{x})$ is close to $\theta(\mathbf{x})\theta(\mathbf{x})^\top$ for every configuration $\mathbf{x}$ in $\mathcal{X}$.

A special case when this occurs is when the interaction pattern $\mathcal{G}$ is the complete graph. First notice that, in this case, the boundary $\xi(\mathbf{x})$ of a configuration $\mathbf{x}$ is completely determined by its type $\theta(\mathbf{x})$. Specifically, for a complete graph with n nodes, we have that $\xi(\mathbf{x})=\xi^{(n)}(\theta(\mathbf{x}))$, where

$$\xi^{(n)}_{ij}(\theta) := \frac{n\theta_i\theta_j}{n-1} - \frac{\theta_i\delta^i_j}{n-1} = \theta_i\theta_j + \frac{\theta_i(\theta_j-\delta^i_j)}{n-1}\,. \tag{28}$$

As a consequence, also the mean drift $D(\mathbf{x})$ depends on the type $\theta(\mathbf{x})$ only: this is consistent with the observation made in Remark 3 that, for complete interaction networks, the type $\theta(X(t))$ is itself a Markov chain. Lemma 2 and expression (28) imply that, for complete interaction networks

$$||D(\mathbf{x})-\overline{D}(\theta(\mathbf{x}))||_1 \leq \frac{4\rho}{n-1}\,, \qquad \forall \mathbf{x}\in\mathcal{X}\,, \tag{29}$$

so that the mean drift and the limit drift coincide in the large-scale limit.

The following result can then be interpreted as a generalization of (29).

Lemma 3: For every configuration $\mathbf{x}$ in $\mathcal{X}$

$$||D(\mathbf{x})-\overline{D}(\theta(\mathbf{x}))||_1 \leq 2\rho|\mathcal{A}|^2 W_{\mathcal{G}} + \frac{4\rho}{n-1}\,. \tag{30}$$

Proof: First, it follows from definitions (8), (24), and (28) that

$$W_{\mathcal{G}} = \max_{\mathbf{x}\in\mathcal{X}}\left|\left|\xi(\mathbf{x})-\xi^{(n)}(\theta(\mathbf{x}))\right|\right|_\infty\,. \tag{31}$$

Then, for every configuration $\mathbf{x}$ in $\mathcal{X}$, we have that

$$\begin{aligned}
||D(\mathbf{x})-\overline{D}(\theta(\mathbf{x}))||_1 &\leq 2\rho||\xi(\mathbf{x})-\theta(\mathbf{x})\theta(\mathbf{x})^\top||_1 \\
&\leq 2\rho||\xi(\mathbf{x})-\xi^{(n)}(\theta(\mathbf{x}))||_1 \\
&\quad +2\rho||\xi^{(n)}(\theta(\mathbf{x}))-\theta(\mathbf{x})\theta(\mathbf{x})^\top||_1 \\
&\leq 2\rho|\mathcal{A}|^2 W_{\mathcal{G}} + \frac{4\rho}{n-1}\,,
\end{aligned} \tag{32}$$

where the first inequality follows from Lemma 2, the second one is a consequence of the triangle inequality, and the last one is implied by (31) and (28). ■

Lemma 3 states that the mean drift $D(\mathbf{x})$ is close to the limit drift $\overline{D}(\theta(\mathbf{x}))$ when the order n is large and the total mixing gap of the interaction pattern $W_{\mathcal{G}}$ is small.

B. Concentration results for finite networks

Below, we shall use the notation $\mathbb{E}_\mu$ to indicate the expectation with respect to a stationary distribution μ. The following technical result, which is a simple consequence of stationarity, will prove to be very useful in our subsequent derivations.

Lemma 4: Consider a PIN model and let μ in $\mathcal{P}(\mathcal{X})$ be one of its stationary distributions. Moreover, let $D(\mathbf{x})$ be its mean drift. Then, for every class-$\mathcal{C}^2$ function $V:\mathcal{P}(\mathcal{A})\to\mathbb{R}$, we have that

$$|\mathbb{E}_\mu[\nabla V(\theta(X(t)))\cdot D(X(t))]| \leq \frac{K}{n}\,,$$

where K is a non-negative constant only depending on V.

Proof: See Appendix A. ■

Lemma 4 allows us to prove the following result for PIN models in stationarity.

Proposition 4: Consider a PIN model. Let μ in $\mathcal{P}(\mathcal{X})$ be one of its stationary distributions and let $D(\mathbf{x})$ be its mean drift. Assume there exists a class-$\mathcal{C}^2$ function $V:\mathcal{P}(\mathcal{A})\to\mathbb{R}$ such that

$$\nabla V(\theta(\mathbf{x}))\cdot D(\mathbf{x}) \leq -F(\mathbf{x})+\epsilon\,, \qquad \forall\mathbf{x}\in\mathcal{X}\,, \tag{33}$$

for some $F : \mathcal{X} \to \mathbb{R}_+$ and $\epsilon \geq 0$. Then, there exists a constant $K > 0$ only depending on V such that

$$\mu\left(\{\mathbf{x} \in \mathcal{X} \,:\, F(\mathbf{x}) < \delta\}\right) \geq 1 - \frac{K}{n\delta} - \frac{\epsilon}{\delta}\,, \tag{34}$$

for every $\delta > 0$.

Proof: It follows from Lemma 4 and inequality (33) that

$$\begin{aligned}\mathbb{E}_\mu\left[F(X(t))\right] &\leq \left|\mathbb{E}_\mu\left[\nabla V(\theta(X(t))) \cdot D(X(t))\right]\right| \\ &\quad + \mathbb{E}_\mu\left[\nabla V(\theta(X(t))) \cdot D(X(t)) + F(X(t))\right] \\ &\leq \frac{K}{n} + \epsilon\,.\end{aligned}$$

Now, by applying the Markov inequality to the nonnegative random variable $F(X(t))$, we get

$$\mu\left(\{X(t) \in \mathcal{X} \,:\, F(X(t)) \geq \delta\}\right) \leq \frac{\mathbb{E}_\mu[F(X(t))]}{\delta} \leq \frac{K}{n\delta} + \frac{\epsilon}{\delta}\,,$$

thus proving the claim. ∎

Remark 5: Proposition 4 implies that, when inequality (33) holds with $\epsilon = 0$ and n grows large, the stationary distributions tend to concentrate where $F(\mathbf{x})$ is close to zero. In particular when

$$\nabla V(\theta(\mathbf{x})) \cdot D(\mathbf{x}) \leq 0\,, \qquad \forall \mathbf{x} \in \mathcal{X}\,, \tag{35}$$

we have that the stationary distributions tend to concentrate where $\nabla V(\theta(\mathbf{x})) \cdot D(\mathbf{x})$ is close to zero. Inequality (35) states that $V(\theta(X(t)))$ is non-increasing in expectation for the process $X(t)$ in stationarity. Therefore, in this case, we can interpret V as a Lyapunov function for $X(t)$. Notice that the mean drift $D(\mathbf{x})$ depends not only on the parameters $(\mathcal{A}, \rho, P, \phi)$ of the PIN model but also on the interaction pattern $\mathcal{G}$. Hence, in general, the search for a Lyapunov function V satisfying (35) is hard since it must consider also the interaction pattern $\mathcal{G}$. This essentially limits the applicability of Proposition 4 for $\varepsilon = 0$ to the case of a complete interaction pattern.

In the following, we shall focus on the application of Proposition 4 with $\varepsilon > 0$, using class-$\mathcal{C}^2$ functions $V : \mathcal{P}(\mathcal{A}) \to \mathbb{R}$ that in general do not satisfy (35), but instead are such that

$$\nabla V(\theta) \cdot \overline{D}(\theta) \leq \zeta\,, \qquad \forall \theta \in \mathcal{P}(\mathcal{A})\,, \tag{36}$$

for some nonnegative constant $\zeta \geq 0$, where we recall the $\overline{D}(\theta)$ is the limit mean drift as defined in (2). Notice that the limit mean drift $\overline{D}(\theta)$ does not depend on the interaction pattern $\mathcal{G}$ of the considered PIN model, but only on its parameters. We are now ready to formulate our main concentration result. Given a class-$\mathcal{C}^2$ function $V : \mathcal{P}(\mathcal{A}) \to \mathbb{R}$ and constant $\delta > 0$, define

$$\mathcal{X}_\delta = \{\mathbf{x} \in \mathcal{X} \,:\, \nabla V(\theta(\mathbf{x})) \cdot \overline{D}(\theta(\mathbf{x})) > -\delta\}\,. \tag{37}$$

Theorem 3: Consider a PIN model with parameters $(\mathcal{A}, \rho, P, \phi)$ and interaction pattern $\mathcal{G}$. Let μ in $\mathcal{P}(\mathcal{X})$ be one of its stationary distributions and let $\overline{D}(\theta)$ be its limit mean drift. Assume that there exists a class-$\mathcal{C}^2$ function $V : \mathcal{P}(\mathcal{A}) \to \mathbb{R}$ such that (36) holds for some $\zeta \geq 0$. Then, there exists a constant $C > 0$ only depending on V such that, for every $\delta > \zeta$ it holds

$$\mu\left(\mathcal{X}_{\delta-\zeta}\right) \geq 1 - \frac{C}{n\delta} - 2|\mathcal{A}|^2 ||\nabla V||_\infty \frac{W_\mathcal{G}}{\delta} - \frac{\zeta}{\delta}\,, \tag{38}$$

where

$$||\nabla V||_\infty := \max_{\theta \in \mathcal{P}(\mathcal{A})} ||\nabla V(\theta)||_\infty\,.$$

Proof: We can write

$$\nabla V(\theta(\mathbf{x})) \cdot D(\mathbf{x}) \leq \nabla V(\theta(\mathbf{x})) \cdot \overline{D}(\theta(\mathbf{x})) + \overline{\epsilon} \tag{39}$$

where

$$\begin{aligned}\overline{\epsilon} &:= \max_{\mathbf{x} \in \mathcal{X}} \left|\nabla V(\theta(\mathbf{x})) \cdot (D(\mathbf{x}) - \overline{D}(\theta(\mathbf{x})))\right| \\ &\leq ||\nabla V||_\infty \max_{\mathbf{x} \in \mathcal{X}} ||D(\mathbf{x}) - \overline{D}(\theta(\mathbf{x}))||_1 \\ &\leq ||\nabla V||_\infty \left(2|\mathcal{A}|^2 W_\mathcal{G} + \frac{4}{n-1}\right)\,,\end{aligned} \tag{40}$$

and where the second inequality follows from Lemma 3 and the fact that $\rho \leq 1$. We define

$$F(\mathbf{x}) := -\nabla V(\theta(\mathbf{x})) \cdot \overline{D}(\theta(\mathbf{x})) + \zeta \geq 0\,,$$

and we rewrite inequality (39) as

$$\nabla V(\theta(\mathbf{x})) \cdot D(\mathbf{x}) \leq -F(\mathbf{x}) + \zeta + \overline{\epsilon}\,. \tag{41}$$

We now apply Proposition 4. Concentration inequality (34) together with definition (37) and inequality (40) yield

$$\begin{aligned}\mu\left(\mathcal{X}_{\delta-\zeta}\right) &= \mu\left(\{\mathbf{x} \in \mathcal{X} \,|\, F(\mathbf{x}) < \delta\}\right) \\ &\geq 1 - \frac{K}{n\delta} - \frac{4||\nabla V||_\infty}{(n-1)\delta} - 2|\mathcal{A}|^2 ||\nabla V||_\infty \frac{W_\mathcal{G}}{\delta} - \frac{\zeta}{\delta} \\ &\geq 1 - \frac{C}{n\delta} - 2|\mathcal{A}|^2 ||\nabla V||_\infty \frac{W_\mathcal{G}}{\delta} - \frac{\zeta}{\delta}\,,\end{aligned}$$

with $C = K + 8||\nabla V||_\infty$. ∎

Notice that the right-hand side of inequality (38) is the sum of various terms. The first one depends only on the size of the interaction network and is negligible for large scale networks, the second one depends on the distance of the interaction network from a totally mixing one, while the last one depends on how far the function V is from being a mean-field Lyapunov function for the PIN model.

A particularly relevant special case of Theorem 3 is when (36) holds true with $\zeta = 0$, i.e., when

$$\nabla V(\theta) \cdot \overline{D}(\theta) \leq 0\,, \qquad \forall \theta \in \mathcal{P}(\mathcal{A})\,. \tag{42}$$

In fact, a function V satisfying (3) can be interpreted as a Lyapunov function for the deterministic mean-field ODE (5). Furthermore, in the special case of complete interaction pattern, we have $\mathcal{W}_\mathcal{G} = 0$ so that Theorem 3 has the following corollary.

Corollary 1: Consider a PIN model with complete interaction pattern and let μ in $\mathcal{P}(\mathcal{X})$ be one of its stationary distributions. Assume that there exists a class-$\mathcal{C}^2$ function

$V : \mathcal{P}(\mathcal{A}) \to \mathbb{R}$ satisfying (3). Then, there exists a constant C only depending on V such that

$$\mu(\mathcal{X}_\delta) \geq 1 - \frac{C}{n\delta} \tag{43}$$

for every $\delta > 0$.

C. Proofs of Theorems 1 and 2

Theorems 1 and 2 are a direct consequence of Theorem 3 applied to ATM interaction networks. Detailed proofs are below.

Proof of Theorem 1: It follows from Theorem 3 with $\zeta = 0$ and Definition 4 that

$$\mu_n(\mathcal{X}_\delta) \geq 1 - \frac{C}{n\delta} - |\mathcal{A}|^2 ||\nabla V||_\infty \frac{W_{\mathcal{G}_n}}{\delta} \overset{n\to+\infty}{\longrightarrow} 1\,,$$

for every $\delta > 0$. ■

Proof of Theorem 2 : Define

$$\mathcal{X}_\delta^{(0)} := \{\mathbf{x} \in \mathcal{X} :\, h(\theta(\mathbf{x})) < \delta\}\,,$$

and, for every $\alpha > 0$,

$$\mathcal{X}_\delta^{(\alpha)} := \{\mathbf{x} \in \mathcal{X} :\, \nabla V_\alpha(\theta(\mathbf{x})) \cdot \overline{D}_\alpha(\theta(\mathbf{x})) > -\delta\}\,,$$

As Theorem 2 assumes inequality (12), we can apply Theorem 3 and obtain that for every $\alpha > 0$ and $\delta > \zeta(\alpha)$ the following estimation holds

$$\begin{aligned}
\mu_n^{(\alpha)}\left(\mathcal{X}_\delta^{(0)}\right) &\geq \mu_n\left(\mathcal{X}_{\delta-\zeta(\alpha)}^{(\alpha)}\right) \\
&\geq 1 - \frac{C}{n\delta} - |\mathcal{A}|^2 ||\nabla V||_\infty \frac{W_{\mathcal{G}_n}}{\delta} - \frac{\zeta(\alpha)}{\delta} \\
&\overset{n\to+\infty}{\longrightarrow} 1 - \frac{\zeta(\alpha)}{\delta}\,,
\end{aligned}$$

where the limit relation follows from the ATM assumption (see Definition 4). This yields

$$\liminf_{n\to+\infty} \mu_n^{(\alpha)}\left(\mathcal{X}_\delta^{(0)}\right) \geq 1 - \frac{\zeta(\alpha)}{\delta}\,.$$

It then follows from (11) that

$$\lim_{\alpha\to 0} \liminf_{n\to+\infty} \mu_n^{(\alpha)}\left(\mathcal{X}_\delta^{(0)}\right) \geq 1 - \lim_{\alpha\to 0} \frac{\zeta(\alpha)}{\delta} = 1\,,$$

thus proving (13). ■

Remark 6: It is worth pointing out that Theorem 3 is more general than Theorem 1 and Theorem 2, as it applies to single PIN models with a given interaction pattern $\mathcal{G}$ of finite order n. In particular, Theorem 3 can also be applied in a context of a sequence of PIN models on ATM interaction networks $\mathcal{G}_n$ with parameters $(\mathcal{A}, \rho_n, P_n, \phi_n)$ that depend on the order n and converge to some given limit parameters $(\mathcal{A}, \rho, P, \phi)$ as n grows large.

VI. Conclusion

In this paper, we have proposed a tool that provides estimates on the invariant distribution of Markov chains resulting from network systems in which agents change their state either by a mutation or by an interaction with their neighbor agents. The key ingredient enabling such estimates is the existence of a Lyapunov-type function for the drift associated with these systems. This method is intrinsically robust, and this allows for its application in situations where the model is only partially known and uncertainty is present. We demonstrate the robustness of the method by proving that the estimates on the invariant distributions valid for complete interaction networks also hold true for perturbed versions of these graphs, such as for instance Erdös-Rényi random graphs. Indeed, leveraging on this fact, we have been able to show, for instance, that the asymptotic behavior of the SIS epidemic models on Erdös-Rényi random graphs coincides with the asymptotic behavior of the SIS models on complete interaction networks. Another example of the method's robustness is shown in Theorem 2 and its application to SIRS epidemic models. In this case, we prove that the method provides useful estimates of the invariant distributions even when the functions $V(\theta)$ are only approximately mean-field Lyapunov functions of the network system.

However, the applicability of this robust convergence result extends far beyond the examples treated in this paper. On one hand, it can be applied to other families of Markov chains resulting from multi-agent systems, as those associated with evolutionary population games. On the other hand, its application can be extended to encompass different types of uncertainty, such as cases where the behavior of the agents is approximately described by a transition probability, or even contexts where a fraction of the agents have a completely unknown behavior.

Finally, even if the setting of this work is purely system theoretic and no control is considered, our approach can in principle lead to the study of control problems for Markov chains, extending concepts and tools from classical control theory for ODEs systems. These kinds of extensions are the subject of our present investigations.

References

[1] P. Donnelly and D. Welsh. Finite particle systems and infection models. *Mathematical Proceedigs of the Cambridge Philosophical Society*, 94:167–182, 1983.

[2] T. M. Liggett. *Interacting Particle Systems*. Springer-Verlag, 1985.

[3] T. M. Liggett. *Stochastic Interacting Systems: Contact, Voter, and Exclusion processes*. Springer Berlin, 1999.

[4] C. Nowzari, V. M. Preciado, and G. J. Pappas. Analysis and control of epidemics: a survey of spreading processes on complex networks. *IEEE Control Systems Magazine*, 36(1):26–46, 2016.

[5] J. T. Cox. Coalescing random walks and voter model consensus times on the torus in $\mathbb{Z}^d$. *Annals of Probability*, 17:133–166, 1989.

[6] P. Donnelly and D. Welsh. The antivoter problem: random 2-colourings of graphs. *Graph Theory and Combinatorics*, pages 133–144, 1984.

[7] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, 2009.

[8] E. Lieberman, C. Hauert, and M.A. Nowak. Evolutionary dynamics in graphs. *Nature*, 433:312–316, 2005.

[9] M. A. Nowak. *Evolutionary Dyanamics*. Harvard University Press, 2006.

[10] A. Ganesh, L. Massoulié, and D. Towsley. The effect of network topology on the spread of epidemics. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 1455–1466, 2005.
[11] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, 115(772):700–721, 1927.
[12] A. Lajmanovich and J.A. Yorke. A deterministic model for gonorrhea in a nonhomogeneous population. *Mathematical Biosciences*, 28(3-4):221–236, 1976.
[13] W. Mei, S. Mohagheghi, S. Zampieri, and F. Bullo. On the dynamics of deterministic epidemic propagation over networks. *Annual Reviews in Control*, 44:116–128, 2017.
[14] M. Alutto, L. Cianfanelli, G. Como, and F. Fagnani. On the dynamic behavior of the network SIR epidemic model. *IEEE Transactions on Control of Network Systems*, 2025.
[15] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7(1):49–58, 1970.
[16] W. H. Sandholm. *Population Games and Evolutionary Dynamics*. MIT Press, 2010.
[17] M. Draief and L. Massoulié. *Epidemics and Rumors in Complex Networks*. Cambridge University Press, Cambridge, UK, 2010.
[18] S. Janson, M. Luczak, and P. Windridge. Law of large numbers for the SIR epidemic on a random graph with given degrees. *Random Structures & Algorithms*, 45(4):726–763, 2014.
[19] W.S. Rossi, G. Como, and F. Fagnani. Threshold models of cascades in large-scale networks. *IEEE Transactions on Network Science and Engineering*, 6(2):158–172, 2019.
[20] D. Keliger, I. Horváth, and B. Takács. Local-density dependent Markov processes on graphons with epidemiological applications. *Stochastic Processes and Their Applications*, 148:324–352, 2022.
[21] C. Ravazzi, G. Como, M. Garetto, E. Leonardi, and A. Tarable. Asynchronous semianonymous dynamics over large-scale networks. *SIAM Journal on Applied Dynamical Systems*, 22(2):1300–1343, 2023.
[22] G. Como and F. Fagnani. Scaling limits for continuous opinion dynamics systems. *Annals of Applied Probability*, 21(4):1537–1567, 2011. Submitted.
[23] M. Kandori, G. J. Mailath, and R. Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 1993.
[24] G. Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071, 1993.
[25] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1998.
[26] L. E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.
[27] L. E. Blume. *The Economy As An Evolving Complex System II*, chapter Population Games. CRC Press, 1997.
[28] H. P. Young. The evolution of conventions. *Econometrica*, 61:57–84, 1993.
[29] M. Benaïm and J. Weibull. Deterministic approximation of stochastic evolution in games. *Econometrica*, 71:873–903, 2003.
[30] M. Benaïm. Recursive algorithms, urn processes, and the chainig number of chain recurrent sets. *Ergodic Theory and Dynamical Systems*, 18(1):53–87, 1998.
[31] M. Benaïm and J.Y. Le Boudec. On mean field convergence and stationary regime. https://arxiv.org/abs/1111.5710, 2011.
[32] F. G. Foster. On the stochastic matrices associated with certain queuing processes. *The Annals of Mathematical Statistics*, 24(3):355–360, 1953.
[33] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Springer Verlag, 1993.
[34] J. R. Norris. *Markov chains*. Cambridge University Press, 1998.
[35] M. I. Freidlin and A. D. Wentzell. *Random Perturbations and Dynamical Systems*. Springer Berlin, Heidelberg, 2012.
[36] H. K. Khalil. *Nonlinear Systems*. Prentice Hall, 3rd edition, 2002.
[37] B. L. Granovsky and N. Madras. The nosiy voter model. *Stochastic Processes and Their Applications*, 55:23–43, 1995.
[38] A. Carro, R. Toral, and M. San Miguel. The noisy voter model on complex networks. *Scientific Reports*, 6(24775), 2016.
[39] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley, 2008.

Appendix A
Proof of Lemma 4

The fact that μ is stationary implies that

$$\mathbb{E}_\mu[V(\theta(X(t+1))) - V(\theta(X(t)))] = 0\,. \tag{44}$$

Since V is $\mathcal{C}^2$, for every θ and h in $\mathbb{R}^{\mathcal{A}}$, there exists β in $[0,1]$ such that

$$V(\theta+h) = V(\theta) + \nabla V(\theta)\cdot h + \frac{1}{2}h^\top \Delta V(\theta+\beta h)h\ .$$

If we apply this expansion to both sides and use (44) letting

$$\theta := \theta(X(t))\,, \qquad h := \theta(X(t+1)) - \theta(X(t))\,,$$

we obtain

$$\mathbb{E}_\mu[\nabla V(\theta)h] = -\frac{1}{2}\mathbb{E}_\mu[h^\top \Delta V(\theta+\beta h)h]\,. \tag{45}$$

Notice now that, for every $\mathbf{x}$ in $\mathcal{X}$,

$$\begin{aligned}\mathbb{E}_\mu[h|X(t)=\mathbf{x}] &= \mathbb{E}_\mu\big[\big(\theta(X(t+1)) - \theta(X(t))\big)|X(t)=\mathbf{x}\big]\\ &= \frac{1}{n}D(\mathbf{x})\,,\end{aligned}$$

so that the left-hand side of (45) is equal to

$$\begin{aligned}\mathbb{E}_\mu\left[\nabla V(\theta)\cdot h\right] &= \mathbb{E}_\mu\left[\mathbb{E}[\nabla V(\theta)\cdot h|X(t)]\right]\\ &= \mathbb{E}_\mu\left[\nabla V(\theta)\cdot \mathbb{E}[h|X(t)]\right]\\ &= \frac{1}{n}\mathbb{E}_\mu\left[\nabla V(\theta)\cdot D(X(t))\right]\ .\end{aligned} \tag{46}$$

From (45) and (46), we obtain

$$\begin{aligned}\left|\mathbb{E}_\mu\left[\nabla V(\theta)\cdot D(X(t))\right]\right| &\le \frac{n}{2}E_\mu\left[|h^\top \Delta V(\theta+\beta h)h|\right]\\ &\le \frac{n}{2}C_V E_\mu\left[h^\top h\right]\ ,\end{aligned} \tag{47}$$

where $C_V = \max_\theta ||\Delta V(\theta)||_2$ and $||\cdot||_2$ denotes the induced 2-norm of a matrix. Since $\|\theta(\mathbf{y})-\theta(\mathbf{x})\|_2^2 \le 2/n^2$ for every $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{X}$ such that $\mathbb{P}(X(t+1)=\mathbf{y}|X(t)=\mathbf{x}) \ne 0$, we have that

$$E_\mu\left[h^\top h\right] \le \frac{2}{n^2}\,,$$

which, together with (47) implies the claim. ■

Appendix B
Proof of Proposition 1

We start by recalling the following fundamental large-deviations result, known as the Chernov bound.

Lemma 5 (Theorem A.1.4 in [39]): Let Y be a random variable with binomial distribution with parameters l and p. Then, for every $\alpha > 0$

$$\mathbb{P}\left(Y \ge lp(1+\alpha)\right) \le \exp(-2lp^2\alpha^2)\,, \tag{48}$$

and

$$\mathbb{P}\left(Y \le lp(1-\alpha)\right) \le \exp(-2lp^2\alpha^2)\,. \tag{49}$$

Proposition 5: Consider the Erdös-Rényi random graph $\mathcal{G}(n,p)$ with $n \ge 2$ nodes and link probability p in $]0,1]$. Then,

$$\mathbb{P}(W_{\mathcal{G}(n,p)} \ge \eta) \le 3^{n+2}\exp(-n^2p^3\eta^3/36)\,, \tag{50}$$

for every $0 < \eta \le 1$.

Proof: For a subset of nodes $\mathcal{U} \subseteq \mathcal{V}$, let $Y_{\mathcal{U}} = |\mathcal{E}_{\mathcal{U}\mathcal{U}}|/2$ be the number of undirected links connecting pairs of nodes in $\mathcal{U}$ in the Erdös-Rényi random graph $\mathcal{G}(n,p)$. Then $Y_{\mathcal{U}}$ is a random variable with binomial distribution with parameters $\binom{|\mathcal{U}|}{2}$ and p. Observe that (48) implies that

$$\begin{aligned}\mathbb{P}\bigg(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{|\mathcal{U}|(|\mathcal{U}|-1)} \geq p(1+\alpha)\bigg) &= \mathbb{P}\bigg(Y_{\mathcal{U}} \geq \binom{|\mathcal{U}|}{2}p(1+\alpha)\bigg) \\ &\leq \exp(-|\mathcal{U}|(|\mathcal{U}|-1)p^2\alpha^2)\,, \end{aligned} \tag{51}$$

while (49) implies that

$$\begin{aligned}\mathbb{P}\bigg(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{|\mathcal{U}|(|\mathcal{U}|-1)} \leq p(1-\alpha)\bigg) &= \mathbb{P}\bigg(Y_{\mathcal{U}} \leq \binom{|\mathcal{U}|}{2}p(1-\alpha)\bigg) \\ &\leq \exp(-|\mathcal{U}|(|\mathcal{U}|-1)p^2\alpha^2)\,. \end{aligned} \tag{52}$$

First, let $\mathcal{Z} = \{\mathcal{U} \subseteq \mathcal{V}\}$ the set of all nodes subsets. For every η such that $0 < \eta \leq 1$, define

$$\mathcal{Z}_\eta = \left\{\mathcal{U} \subseteq \mathcal{V}:\ \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)} < \frac{p\eta}{2}\right\}, \qquad \overline{\mathcal{Z}}_\eta = \mathcal{Z} \setminus \mathcal{Z}_\eta\,,$$

$$\chi_\eta = \mathbb{P}\left(\max_{\mathcal{U}\in\mathcal{Z}_\eta}\left|\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{m} - \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)}\right| \geq \eta\right)\,,$$

and

$$\overline{\chi}_\eta = \mathbb{P}\left(\max_{\mathcal{U}\in\overline{\mathcal{Z}}_\eta}\left|\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{m} - \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)}\right| \geq \eta\right)\,.$$

On the one hand, notice that for every node set $\mathcal{U}$ in $\mathcal{Z}_\eta$, we have that $|\mathcal{E}_{\mathcal{U}\mathcal{U}}| \leq |\mathcal{U}|(|\mathcal{U}|-1) < p\eta n(n-1)/2$, and $m \leq n(n-1)$, so that

$$\left|\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{|\mathcal{U}|(|\mathcal{U}|-1)} - \frac{m}{n(n-1)}\right| \leq \max\left\{\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{|\mathcal{U}|(|\mathcal{U}|-1)}, \frac{m}{n(n-1)}\right\} \leq 1\,,$$

and

$$\begin{aligned}\left|\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{m} - \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)}\right| &= \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{m}\left|\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{|\mathcal{U}|(|\mathcal{U}|-1)} - \frac{m}{n(n-1)}\right| \\ &\leq \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{m} \\ &< p\eta\frac{n(n-1)}{2m}\,.\end{aligned}$$

It follows that

$$\begin{aligned}\chi_\eta &= \mathbb{P}\left(\max_{\mathcal{U}\in\mathcal{Z}_\eta}\left|\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{m} - \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)}\right| \geq \eta\right) \\ &\leq \mathbb{P}\left(\frac{m}{n(n-1)} \leq \frac{p}{2}\right) \\ &\leq \exp(-n(n-1)p^2/4)\,, \\ &\leq \exp(-n^2p^2/8)\,,\end{aligned} \tag{53}$$

where the second inequality follows from (52) with $\mathcal{U} = \mathcal{V}$ and $\alpha = 1/2$ and the last one from the fact that $(n-1) \geq n/2$ for $n \geq 2$.

On the other hand, for every node set $\mathcal{U}$ in $\overline{\mathcal{Z}}_\eta$, let

$$p_{\mathcal{U}}^+ := \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{m} - \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)} \geq \eta\right)$$

$$p_{\mathcal{U}}^- := \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{m} - \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)} \leq -\eta\right)\,,$$

Observe that, for every $\mathcal{U}$ in $\overline{\mathcal{Z}}_\eta$,

$$\begin{aligned}p_{\mathcal{U}}^+ &= \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|n(n-1)}{m|\mathcal{U}|(|\mathcal{U}|-1)} \geq 1 + \eta\frac{n(n-1)}{|\mathcal{U}|(|\mathcal{U}|-1)}\right) \\ &\leq \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|n(n-1)}{m|\mathcal{U}|(|\mathcal{U}|-1)} \geq 1 + \eta\right) \\ &\leq \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|n(n-1)}{m|\mathcal{U}|(|\mathcal{U}|-1)} \geq \frac{1+\eta/3}{1-\eta/3}\right) \\ &\leq \mathbb{P}\left(\left\{\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{|\mathcal{U}|(|\mathcal{U}|-1)} \geq p(1+\eta/3)\right\} \right. \\ &\qquad\qquad \left. \cup \left\{\frac{|\mathcal{E}_{\mathcal{V}\mathcal{V}}|}{n(n-1)} \leq p(1-\eta/3)\right\}\right) \\ &\leq \mathbb{P}\bigg(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{|\mathcal{U}|(|\mathcal{U}|-1)} \geq p(1+\eta/3)\bigg) \\ &\qquad\qquad + \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{V}\mathcal{V}}|}{n(n-1)} \leq p(1-\eta/3)\right) \\ &\leq \exp(-|\mathcal{U}|(|\mathcal{U}|-1)p^2\eta^2/9) + \exp(-n(n-1)p^2\eta^2/9) \\ &\leq \exp(-n(n-1)p^3\eta^3/18) + \exp(-n(n-1)p^2\eta^2/9) \\ &\leq 2\exp(-n(n-1)p^3\eta^3/18) \\ &\leq 2\exp(-n^2p^3\eta^3/36)\,,\end{aligned} \tag{54}$$

where the first inequality follows from the fact that $|\mathcal{U}| \leq n$, the second from the fact that

$$\frac{1+\eta/3}{1-\eta/3} = 1 + \frac{2\eta/3}{1-\eta/3} \leq 1 + \eta\,, \qquad \forall \eta\ : 0 \leq \eta \leq 1\,,$$

the fourth one from a union bound argument, the fifth one by applying (51) and (52) with $\alpha = \eta/3$, respectively, the sixth one since $|\mathcal{U}|(|\mathcal{U}|-1) \geq n(n-1)p\eta/2$ for every $\mathcal{U}$ in $\overline{\mathcal{Z}}_\eta$, the seventh since $p\eta/2 \leq 1$, and the last one since $n \geq 2$. Analogously, for every $\mathcal{U}$ in $\overline{\mathcal{Z}}_\eta$, we can argue that

$$\begin{aligned}p_{\mathcal{U}}^- &= \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|n(n-1)}{m|\mathcal{U}|(|\mathcal{U}|-1)} \leq 1 - \eta\frac{n(n-1)}{|\mathcal{U}|(|\mathcal{U}|-1)}\right) \\ &\leq \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|n(n-1)}{m|\mathcal{U}|(|\mathcal{U}|-1)} \leq 1 - \eta\right) \\ &\leq \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|n(n-1)}{m|\mathcal{U}|(|\mathcal{U}|-1)} \leq \frac{1-\eta/2}{1+\eta/2}\right) \\ &\leq \mathbb{P}\left(\left\{\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{|\mathcal{U}|(|\mathcal{U}|-1)} \leq p(1-\eta/2)\right\} \right. \\ &\qquad\qquad \left. \cup \left\{\frac{|\mathcal{E}_{\mathcal{V}\mathcal{V}}|}{n(n-1)} \geq p(1+\eta/2)\right\}\right) \\ &\leq \mathbb{P}\bigg(\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{|\mathcal{U}|(|\mathcal{U}|-1)} \leq p(1-\eta/2)\bigg) \\ &\qquad\qquad + \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{V}\mathcal{V}}|}{n(n-1)} \geq p(1+\eta/2)\right) \\ &\leq \exp(-|\mathcal{U}|(|\mathcal{U}|-1)p^2\eta^2/4) + \exp(-n(n-1)p^2\eta^2/4) \\ &\leq \exp(-n(n-1)p^3\eta^3/8) + \exp(-n(n-1)p^2\eta^2/4) \\ &\leq 2\exp(-n(n-1)p^3\eta^3/8) \\ &\leq 2\exp(-n^2p^3\eta^3/16)\,.\end{aligned} \tag{55}$$

Now, define for every $\mathcal{U}$ in $\overline{\mathcal{Z}}_\eta$

$$p_{\mathcal{U}} := \mathbb{P}\left(\left|\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|}{2m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)}\right| \geq \eta\right) .$$

A union bound argument together with inequalities (54) and (55) imply that

$$p_{\mathcal{U}} \leq p_{\mathcal{U}}^+ + p_{\mathcal{U}}^- \leq 4\exp(-n^2p^3\eta^3/36)\,, \qquad \forall \mathcal{U} \in \overline{\mathcal{Z}}_\eta\,.$$

By applying the union bound once more, using the above estimation, and noting that $|\overline{\mathcal{Z}}_\eta| \leq |\mathcal{Z}| = 2^n$, we get

$$\begin{aligned}\overline{\chi}_\eta &= \mathbb{P}\left(\max_{\mathcal{U}\in\overline{\mathcal{Z}}_\eta}\left|\frac{|\mathcal{E}_{\mathcal{U}\mathcal{U}}|}{2m} - \frac{|\mathcal{U}|(|\mathcal{U}|-1)}{n(n-1)}\right| \geq \eta\right)\\ &\leq \sum_{\mathcal{U}\in\overline{\mathcal{Z}}_\eta} p_{\mathcal{U}}\\ &\leq 4|\mathcal{Z}_\eta|\exp(-n^2p^3\eta^3/36)\\ &\leq 4\cdot 2^n\exp(-n^2p^3\eta^3/36)\,.\end{aligned} \tag{56}$$

Now, let

$$\mathcal{W} = \{(\mathcal{S},\mathcal{U}) : \mathcal{S}\subseteq\mathcal{V}, \mathcal{U}\subseteq\mathcal{V}, \mathcal{S}\cap\mathcal{U} = \emptyset\}\,.$$

Then, for $(\mathcal{S},\mathcal{U})$ in $\mathcal{W}$, the number $|\mathcal{E}_{\mathcal{S}\mathcal{U}}|$ of directed links from nodes in $\mathcal{S}$ to nodes in $\mathcal{U}$ in $\mathcal{G}(n,p)$, that is equal to the number of undirected links connecting nodes in $\mathcal{S}$ with nodes in $\mathcal{U}$, is a binomial random variable of parameters $|\mathcal{S}||\mathcal{U}|$ and p, so that, for every $\alpha \geq 0$,

$$\mathbb{P}(|\mathcal{E}_{\mathcal{S}\mathcal{U}}| \geq |\mathcal{S}||\mathcal{U}|p(1+\alpha)) \leq \exp(-2|\mathcal{S}||\mathcal{U}|p^2\alpha^2)\,. \tag{57}$$

For every η such that $0 < \eta \leq 1$, define

$$\mathcal{W}_\eta := \left\{(\mathcal{S},\mathcal{U})\in\mathcal{W} : \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)} \leq \frac{p\eta}{2}\right\},\quad \overline{\mathcal{W}}_\eta = \mathcal{W}\setminus\mathcal{W}_\eta\,,$$

$$\psi_\eta = \mathbb{P}\left(\max_{(\mathcal{S},\mathcal{U})\in\mathcal{W}_\eta}\left|\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)}\right| \geq \eta\right)\,,$$

and

$$\overline{\psi}_\eta = \mathbb{P}\left(\max_{(\mathcal{S},\mathcal{U})\in\overline{\mathcal{W}}_\eta}\left|\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)}\right| \geq \eta\right)\,.$$

Notice that, for every $(\mathcal{S},\mathcal{U})$ in $\mathcal{W}_\eta$,

$$|\mathcal{E}_{\mathcal{S}\mathcal{U}}| \leq |\mathcal{S}||\mathcal{U}| \leq n(n-1)p\eta/2\,,$$

so that, using also the fact that $m \leq n(n-1)$, we get

$$\left|\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)}\right| \leq \frac{|\mathcal{S}||\mathcal{U}|}{m} \leq \frac{n(n-1)p\eta}{2m} = \binom{n}{2}\frac{p\eta}{m}\,.$$

Therefore,

$$\begin{aligned}\psi_\eta &= \mathbb{P}\left(\max_{(\mathcal{S},\mathcal{U})\in\mathcal{W}_\eta}\left|\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)}\right| \geq \eta\right)\\ &\leq \mathbb{P}\left(m \leq \binom{n}{2}p\right)\\ &\leq \exp(-n^2p^2/8)\,,\end{aligned} \tag{58}$$

where the second inequality follows from (52) with $\alpha = 1/2$. On the other hand, for $(\mathcal{S},\mathcal{U})$ in $\overline{\mathcal{W}}_\eta$, define

$$p_{\mathcal{S}\mathcal{U}}^+ := \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)} \geq \eta\right)$$
$$p_{\mathcal{S}\mathcal{U}}^- := \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)} \leq -\eta\right)\,,$$

and observe that, following the same steps as in (54), we obtain that

$$\begin{aligned}p_{\mathcal{S}\mathcal{U}}^+ &= \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|n(n-1)}{m|\mathcal{S}||\mathcal{U}|} \geq 1 + \eta\frac{n(n-1)}{|\mathcal{S}||\mathcal{U}|}\right)\\ &\leq \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|n(n-1)}{m|\mathcal{S}||\mathcal{U}|} \geq 1+2\eta\right)\\ &\leq \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|n(n-1)}{m|\mathcal{S}||\mathcal{U}|} \geq \frac{1+\eta/2}{1-\eta/2}\right)\\ &\leq \mathbb{P}\Big(\Big\{|\mathcal{E}_{\mathcal{S}\mathcal{U}}| \geq p|\mathcal{S}||\mathcal{U}|(1+\eta/2)\Big\}\\ &\qquad\qquad \cup\Big\{m \leq n(n-1)p(1-\eta/2)\Big\}\Big)\\ &\leq \mathbb{P}\left(|\mathcal{E}_{\mathcal{S}\mathcal{U}}| \geq p|\mathcal{S}||\mathcal{U}|(1+\eta/2)\right)\\ &\qquad +\mathbb{P}\left(m \leq n(n-1)p(1-\eta/2)\right)\\ &\leq \exp(-|\mathcal{S}||\mathcal{U}|p^2\eta^2/2) + \exp(-n^2p^2\eta^2/8)\\ &\leq \exp(-n^2p^3\eta^3/8) + \exp(-n^2p^2\eta^2/8)\\ &\leq 2\exp(-n^2p^3\eta^3/8)\,,\end{aligned} \tag{59}$$

where the first inequality follows from the fact that $|\mathcal{S}||\mathcal{U}| \leq |\mathcal{S}|(n-|\mathcal{S}|) \leq n^2/4$, the second one since

$$\frac{1+\eta/2}{1-\eta/2} = 1 + \frac{\eta}{1-\eta/2} \leq 1+2\eta\,, \qquad \forall\eta :\ 0\leq\eta\leq 1\,,$$

the fourth one from a union bound argument, the fifth one by applying (57) and (52) with $\alpha = \eta/2$, the sixth one since

$$|\mathcal{S}||\mathcal{U}| > n^2p\eta/4\,,$$

for every $(\mathcal{S},\mathcal{U}) \notin \mathcal{W}_\eta$, and the last one since $p\eta \leq 1$.

Similarly, we get

$$\begin{aligned}p_{\mathcal{S}\mathcal{U}}^- &= \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|n(n-1)}{m|\mathcal{S}||\mathcal{U}|} \leq 1 - \eta\frac{n(n-1)}{|\mathcal{S}||\mathcal{U}|}\right)\\ &\leq \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|n(n-1)}{m|\mathcal{S}||\mathcal{U}|} \leq 1-2\eta\right)\\ &\leq \mathbb{P}\left(\frac{|\mathcal{E}_{\mathcal{S}\mathcal{U}}|n(n-1)}{m|\mathcal{S}||\mathcal{U}|} \leq \frac{1-\eta}{1+\eta}\right)\\ &\leq \mathbb{P}\Big(\Big\{|\mathcal{E}_{\mathcal{S}\mathcal{U}}| \leq p|\mathcal{S}||\mathcal{U}|(1-\eta)\Big\}\\ &\qquad\qquad \cup\Big\{m \geq n(n-1)p(1+\eta)\Big\}\Big)\\ &\leq \mathbb{P}\left(|\mathcal{E}_{\mathcal{S}\mathcal{U}}| \leq p|\mathcal{S}||\mathcal{U}|(1-\eta)\right)\\ &\qquad +\mathbb{P}\left(m \geq n(n-1)p(1+\eta)\right)\\ &\leq \exp(-2|\mathcal{S}||\mathcal{U}|p^2\eta^2) + \exp(-n^2p^2\eta^2/2)\\ &\leq \exp(-n^2p^3\eta^3/2) + \exp(-n^2p^2\eta^2/2)\\ &\leq 2\exp(-n^2p^3\eta^3/8)\,.\end{aligned} \tag{60}$$

For every $(\mathcal{S},\mathcal{U})$ in $\overline{\mathcal{W}}_\eta$, let

$$p_{\mathcal{SU}} := \mathbb{P}\left(\left|\frac{|\mathcal{E}_{\mathcal{SU}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)}\right| \geq \eta\right) .$$

A union bound argument together with inequalities (59) and (60) imply that

$$p_{\mathcal{SU}} \leq p^+_{\mathcal{SU}} + p^-_{\mathcal{SU}} \leq 4\exp(-n^2p^3\eta^3/8) .$$

By now applying one more time the union bound, the above estimation, and noting that $|\overline{\mathcal{W}}_\eta| \leq |\mathcal{W}| = 3^n$, we get

$$\begin{aligned} \overline{\psi}_\eta &= \mathbb{P}\left(\max_{(\mathcal{S},\mathcal{U})\in\overline{\mathcal{W}}_\eta}\left|\frac{|\mathcal{E}_{\mathcal{SU}}|}{m} - \frac{|\mathcal{S}||\mathcal{U}|}{n(n-1)}\right| \geq \eta\right) \\ &\leq \sum_{(\mathcal{S},\mathcal{U})\in\overline{\mathcal{W}}_\eta} p_{\mathcal{SU}} \\ &\leq 4|\overline{\mathcal{W}}_\eta|\exp(-n^2p^3\eta^3/8) \\ &\leq 4\cdot 3^n\exp(-n^2p^3\eta^3/8) . \end{aligned} \tag{61}$$

Then, estimations (53), (56), (58), and (61) imply that

$$\begin{aligned} \mathbb{P}\left(W_{\mathcal{G}(n,p)} \geq \eta\right) &= \chi_\eta + \overline{\chi}_\eta + \psi_\eta + \overline{\psi}_\eta \\ &\leq 2\exp(-n^2p^2/8) \\ &+ 4\cdot 2^n\exp(-n^2p^3\eta^3/36) \\ &+ 4\cdot 3^n\exp(-n^2p^3\eta^3/8) \\ &\leq (4\cdot(3^n+2^n)+2)\exp(-n^2p^3\eta^3/36) \\ &\leq 3^{n+2}\exp(-n^2p^3\eta^3/36) , \end{aligned}$$

thus proving the claim. ■

We can now easily prove Proposition 1 from Proposition 5.

Proof of Proposition 1: If $np^3 \to +\infty$, then we have that, for every $\eta > 0$, the expression $3^{n+2}\exp(-n^2p^3\eta^3/36)$ vanishes faster than exponentially as n grows large. Hence, by (50),

$$\sum_{n\geq 2}\mathbb{P}(W_{\mathcal{G}(n,p)} \geq \eta) \leq \sum_{n\geq 2} 3^{n+2}\exp(-n^2p^3\eta^3/36) < +\infty ,$$

and the Borel-Cantelli lemma implies that

$$\limsup_{n\to+\infty} W_{\mathcal{G}(n,p)} \leq \eta ,$$

with probability 1. As this holds true for every $\eta > 0$, we conclude that $W_{\mathcal{G}(n,p)} \to 0$ with probability 1 as n grows large. ■

Giacomo Como (M'12) Giacomo Como is a Professor at the Department of Mathematical Sciences, Politecnico di Torino, Italy. He is also a Senior Lecturer at the Automatic Control Department, Lund University, Sweden. He received the B.Sc., M.S., and Ph.D. degrees in Applied Mathematics from Politecnico di Torino, Italy, in 2002, 2004, and 2008, respectively. He was a Visiting Assistant in Research at Yale University in 2006–2007 and a Postdoctoral Associate at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology in 2008–2011. Prof. Como currently serves as Senior Editor for the *IEEE Transactions on Control of Network Systems*, and as Associate Editor for *Automatica* and the *IEEE Transactions on Automatic Control*. He served as Associate Editor for the *IEEE Transactions on Network Science and Engineering* (2015-2021) and for the *IEEE Transactions on Control of Network Systems* (2016-2022). He was the IPC chair of the IFAC Workshop NecSys'15 , a semiplenary speaker at the International Symposium MTNS'16, and the chair of the IEEE-CSS Technical Committee on Networks and Communications (2019-2024). He is a recipient of the 2015 George S. Axelby Outstanding Paper Award. His research interests are in dynamics, information, and control in network systems with applications to cyber-physical systems, infrastructure networks, and social and economic networks.

Fabio Fagnani received the Laurea degree in Mathematics from the University of Pisa and the Scuola Normale Superiore, Italy, in 1986 and the Ph.D. degree in Mathematics from the University of Groningen, The Netherlands, in 1991. From 1991 to 1998, he was an Assistant Professor of Mathematical Analysis at the Scuola Normale Superiore. In 1997, he was a Visiting Professor at the Massachusetts Institute of Technology, US. Since 1998, he has been a Professor of Mathematical Analysis at Politecnico of Torino, Italy, where he was the Coordinator of the PhD program in Mathematics for Engineering Sciences form 2006 to 2012 and the Head of the Department of Mathematical Sciences from 2012 to 2019. He was an Associate Editor for the *IEEE Transactions on Automatic Control*, the *IEEE Transactions on network Science and Engineering*, and the *IEEE Transactions on Control of network Systems*. His current research topics are on cooperative algorithms and dynamical systems over networks, inferential distributed algorithms, and opinion dynamics.

Sandro Zampieri received the Laurea degree in Electrical Engineering and the Ph.D. degree in System Engineering from the University of Padova, Italy, in 1988 and 1993, respectively. Since 2002 he has been Full Professor of Automatic Control at the Department of Information Engineering of the University of Padova. He has been the head of the Department of Information Engineering from 2014 until 2018. In 1991–1992, 1993 and 1996 he was Visiting Scholar at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology. He has held visiting positions also at the Department of Mathematics of the University of Groningen and at the Department of Mechanical Engineering of the University of California at Santa Barbara. Prof. Zampieri was the general chair of the 1st IFAC Workshop on Estimation and Control of Networked Systems 2009, program chair of the 3rd IFAC Workshop on Estimation and Control of Networked Systems 2012, and publication chair of the IFAC World Congress 2011. He was the chair of the IFAC technical committee "Networked systems" on 2005–2008. He was one of the recipients of the 2016 IEEE Transactions on Control of Network Systems Best Paper Award. He has been an IEEE Fellow since 2022. His research interests include networked control, control of complex systems and distributed control and estimation with applications to the smart grids.