Minireview

Accuracy criterion for mean field approximations of Markov processes on hypergraphs[☆],[☆☆]

Dániel Keliger [a,b,*], Illés Horváth [c]

[a] Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary
[b] Alfréd Rényi Institute of Mathematics, Budapest, Hungary
[c] MTA-BME Information Systems Research Group, Hungary

ARTICLE INFO

ABSTRACT

We provide error bounds for the N-intertwined mean-field approximation (NIMFA) for local density-dependent Markov population processes with a well-distributed underlying network structure showing NIMFA being accurate when a typical vertex has many neighbors. The result justifies some of the most common approximations used in epidemiology, statistical physics and opinion dynamics literature under certain conditions. We allow interactions between more than 2 individuals, and an underlying hypergraph structure accordingly.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

1. Introduction

The analysis of stochastic population processes is an important topic in several disciplines, such as epidemiology, biology, economics or computer systems [1–5]. Such processes consist of a large number of interacting individuals (agents) that execute random actions based on the behavior of other individuals.

A widely-used framework is Markov population processes, where each individual is in a local state from a fixed, finite state space, and can change their state in a Markovian manner. For such models, the state space increases exponentially with the population size, making an exact analysis infeasible even for moderate population sizes, instead raising the question of good approximations as the next best thing.

The classical result of Kurtz [6,7] is based on two main assumptions: that each individual can observe the entire population, and that the Markovian transition rates of each individual depend on the observation in a density-dependent manner. The conclusion is that, as the number of individuals diverges, the evolution of the stochastic system converges to a deterministic mean-field limit. This limit is straightforward to compute numerically, and can serve as a good approximation of the stochastic system when the number of individuals is large. The mean-field limit of Kurtz is referred to as the *homogeneous mean-field approximation* in the present paper.

While the density-dependent Markov setting is flexible and covers many potential applications, the assumption that each individual can observe the entire population is very restrictive. In many population processes arising from real-life examples, individuals do not have full information about the entire population; instead, each individual can observe only a subset of the population. This information structure can be described by a network topology, where each individual has interactions only with its neighbors according to that topology.

The *N*-intertwined mean field approximation (NIMFA) [8] is a quenched mean-field approximation, where differential equations are considered for each individual based on their expected evolution. NIMFA is a deterministic process different from the homogeneous mean-field approximation that incorporates the network structure naturally, making it a potentially more accurate approximation. On the flip side, the computational complexity is considerably increased compared to the homogeneous mean-field approximation; nevertheless, it remains tractable for population sizes large enough to make it relevant for practical applications. Unfortunately, unlike for homogeneous systems, the justification for using NIMFA is poorly understood, mostly relying on numerical evidence [9,10] along with a few theoretical results [11–14].

In the present paper, we focus on a specific class of Markov processes dubbed *local density-dependent Markov population processes*, which preserves the density-dependent assumption of Kurtz, but allows an underlying network structure that dictates the environments observed by each individual. This setting covers many of the frequently used stochastic models, such as the SIS process in epidemiology [15–18], Glauber dynamics in statistical physics [19,20], or the voter model and majority vote in opinion dynamics [21,22]. We incorporate interactions between more than 2 vertices into the model with an underlying hypergraph structure accordingly to reflect on some recent developments in the theory of higher order interactions.

We provide general error bounds for NIMFA that are strong on well-distributed networks. Furthermore, under additional homogeneity assumptions, such as annealed or activity driven networks [23,24] we show these error bounds to be small, with the added benefit of further reducing the number of equations to other well-known approximations, like the *heterogeneous mean field approximation* [25]. Finally, we elaborate the on the argument given by K. Devriendt and P. Van Mieghem [26] and show that Szemerédi's regularity lemma [27] can be applied to reduce the number of equations (depending on a given ε error).

The rest of the paper is structured as follows. Section 2 introduces basic notation and setup for density-dependent Markov population processes along with examples of models used in the literature to illustrate these concepts and their applicability. Section 3 states the main results and also relates them to the recent work of Sridhar and Kar [12,13] and Parasnis et al. [14]. Section 4 discusses further reductions of NIMFA to more simple approximations used throughout the literature. Section 5 contains a summary of this paper along with the limitations of these results and possible directions for further research.

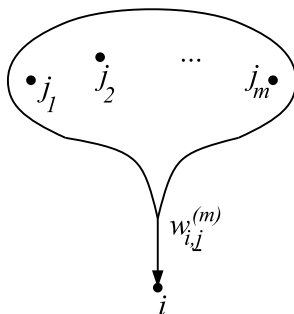Finally, proofs are contained in Section 6.

Fig. 1. Edge (hyperedge) with weight $w^{(m)}_{i,j_1,\ldots,j_m}$.

2. Setup

2.1. The underlying hypergraph

Let G be a finite hypergraph on N vertices. The vertex set is labeled $[N] = \{1, \ldots, N\}$. The hypergraph is not necessarily uniform; edges may contain up to $M + 1$ vertices. The edges are ordered, with the first vertex being special, and we will usually use the notation $(i, j_1, \ldots, j_m)$ for an edge where $1 \leq m \leq M$ and $i, j_1, \ldots, j_m \in [N]$. The idea behind the distinction of the first vertex in an edge is that $w^{(m)}_{i,j_1,\ldots,j_m}$ will describe the strength of connections where $j_1, \ldots, j_m$ have a joint effect on vertex i (see Fig. 1).

The $M = 1$ case corresponds to (directed) graphs.

We allow so-called *secondary loops* (abbreviated as s. loop), which are (i, j) edges with non-distinct vertices among $j_1, \ldots, j_m \in [N]$. Note that traditional loops for the $m = 1$ case are excluded from this definition.

We use the notation $[N]^m$ to denote the set of m-tuples, and $\underline{j}$ abbreviates $(j_1, \ldots, j_m)$.

For unweighted hypergraphs, adjacency indicators $a^{(m)}_{i,j_1,\ldots,j_m}$ (where $1 \leq m \leq M$ and $i, j_1, \ldots, j_m \in [N]$)

$$a^{(m)}_{i,\underline{j}} = \begin{cases} 1 & \text{if } i, j_1, \ldots, j_m \text{ are on the same hyperedge} \\ 0 & \text{else} \end{cases}$$

describe the connections between the vertices.

In-degrees for $1 \leq m \leq M$ are defined as

$$d^{(m)}(i) := \frac{1}{m!} \sum_{\underline{j} \in [N]^m} a^{(m)}_{i,\underline{j}}, \tag{1}$$

(where $m!$ is included to cancel the re-orderings of $\underline{j}$), and the average in-degree for each $1 \leq m \leq M$ is

$$\bar{d}^{(m)} := \frac{1}{N} \sum_{i=1}^{N} d^{(m)}(i).$$

In the literature, some normalization is usually assumed. In the present paper, we introduce normalized weights $w^{(m)}_{i,j_1,\ldots,j_m}$ and corresponding normalized in-degrees

$$\delta^{(m)}(i) := \sum_{\underline{j} \in [N]^m} w^{(m)}_{i,\underline{j}}.$$

representing the total weight of m-interactions effecting vertex $i \in [N]$. In the $M = 1$ case (classical graphs) we tend to omit the upper index (m) and write $w^{(m)}_{i,\underline{j}}$ simply as w_{ij}, and we also utilize the matrix notation $W = (w_{ij})_{i,j \in [N]}$.

We have two Conventions for the normalization.

Convention 1:

$$w^{(m)}_{i,\underline{j}} = \frac{a^{(m)}_{i,\underline{j}}}{m! \bar{d}^{(m)}}, \qquad \delta^{(m)}(i) = \frac{d^{(m)}(i)}{\bar{d}^{(m)}}. \tag{2}$$

Convention 2:

$$w^{(m)}_{i,\underline{j}} = \frac{a^{(m)}_{i,\underline{j}}}{m! d^{(m)}(i)}, \qquad \delta^{(m)}(i) = 1. \tag{3}$$

(The same $m!$ from (1) is now included in the conventions.)

For either convention, whenever the denominator would be 0, the numerator will also be 0, and $w_{i,\underline{j}}^{(m)}$ is simply set to 0 as well.

We set

$$w_{\max} = \max_{i,j,m} w_{i,\underline{j}}^{(m)}.$$

We are going to set regularity assumptions for the weights and degrees:

$$\delta^{(m)}(i) \le \delta_{\max}, \tag{4}$$

$$\sum_{\substack{j \in [N]^m \\ \underline{j} \text{ s. loop}}} w_{i,\underline{j}}^{(m)} \le R\sqrt{w_{\max}}. \tag{5}$$

For Convention 2, (4) always holds. For Convention 1, we need $d^{(m)}(i) \le \delta_{\max}\bar{d}^{(m)}$ (upper regularity of the hypergraph).

(5) always holds for $M = 1$. It also obviously holds if there are no secondary loops. In other cases, it is an actual restriction on the total weight of secondary loops.

Symmetry is in general not assumed, that is, the hypergraph may be directed.

For some results concerning classical graphs ($M = 1$) with Convention 2, the extra assumption is needed for *out*-degrees as well.

$$\delta^{\text{out}}(j) := \sum_{i \in [N]} w_{ij} \le \delta_{\max}^{\text{out}} \tag{6}$$

Assumption (4) and (6) can be understood as a weaker version of double stochasticity of W assumed in [12,13].

2.2. Local density dependent Markov population process

We define a Markov process on the hypergraph. Each vertex is in a state from a finite state space $\mathcal{S}$. $\xi_{i,s}(t)$ denotes the indicator that vertex i is in state s at time t; the corresponding vector notation is

$$\xi_i(t) = \left(\xi_{i,s}(t)\right)_{s \in \mathcal{S}}.$$

We also introduce the notation

$$\xi_{\underline{i},\underline{s}}^{(m)}(t) = \prod_{k=1}^{m} \xi_{i_k, s_k}(t),$$

where $\underline{i} = (i_1, \ldots, i_m)$ is an edge and $\underline{s} = (s_1, \ldots, s_m)$ is a collection of states ($s_k \in \mathcal{S}, k = 1, \ldots, m$). $\xi_{\underline{i},\underline{s}}^{(m)}(t)$ describes the indicator of vertices $i_1, \ldots, i_m$ being in states $s_1, \ldots, s_n$ at time t, respectively.

We define the *m-neighborhood* of vertex i corresponding to $\underline{s} = (s_1, \ldots, s_m)$ as

$$\phi_{i,\underline{s}}^{(m)}(t) = \sum_{\underline{j} \in \{N\}^m} w_{i,\underline{j}}^{(m)} \xi_{\underline{j},\underline{s}}^{(m)}(t). \tag{7}$$

Some explanation is in order. Let $\underline{s} = (s_1, \ldots, s_m)$ be fixed for now. According to (7), we consider all edges that include i and m other vertices; for each such edge, we check whether the m other vertices are exactly according to the configuration of states described by $\underline{s}$; if yes, their contribution to $\phi_{i,\underline{s}}^{(m)}(t)$ is $w_{i,\underline{j}}^{(m)}$, otherwise their contribution is 0.

The m-neighborhoods of i consist of $\phi_{i,\underline{s}}^{(m)}(t)$ for all possible configurations of states $\underline{s}$. The corresponding vector notation is

$$\phi_i^{(m)}(t) = \left(\phi_{i,\underline{s}}^{(m)}(t)\right)_{\underline{s} \in \mathcal{S}^m}, \tag{8}$$

and we may even write

$$\phi_i(t) = \left(\phi_i^{(m)}(t)\right)_{m=1}^{M} \tag{9}$$

for the entire neighborhood of i.

In (7), the normalized weights $w_{i,\underline{j}}^{(m)}$ are used; in case $w_{i,\underline{j}}^{(m)} = 0$ for some $\underline{j}$, the corresponding interaction is simply not present.

Each vertex may transition to another state in continuous time. The transition rates of a vertex may depend on all of its m-neighborhoods for $1 \le m \le M$; accordingly, the transition rate from s' to s is described by the function

$$q_{ss'} : \otimes_{m=1}^{M} \mathbb{R}^{\mathcal{S}^m} \to \mathbb{R}$$

for each $s' \ne s \in \mathcal{S}$.

We assume $q_{ss'}$ is locally Lipschitz, and we also require $q_{ss'}(\phi^{(1)}, \ldots, \phi^{(M)}) \geq 0$ for non-negative inputs.
For "diagonal" rates,

$$q_{ss} := -\sum_{s' \neq s} q_{s's}$$

corresponds to the total outgoing rate from state s.

The corresponding transition matrix is $Q = (q_{ss'})_{s,s' \in \mathcal{S}}$. We emphasize that in this convention $q_{ss'}$ refers to an $s \leftarrow s'$ transition and not an $s \rightarrow s'$ one. This ordering allows us to use column vectors and matrix multiplication from the left.

The dynamics of $(\xi_i(t))_{i=1}^N$ is a continuous-time Markov chain with state-space $\mathcal{S}^N$ where each vertex performs transitions according to the transition rates $q_{s's}$, independently from the others. After a transition, vertices update their neighborhood vectors $\phi_i(t)$. We call such dynamics *local-density dependent Markov processes*.

We define the process $(\xi_{i,s})_{i,s}$ formally via Poisson representation:

$$\xi_{i,s}(t) = \xi_{i,s}(0) + \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} \mathcal{N}_{i,ss'}\left(\mathcal{H}_{i,ss'}(t)\right) - \mathcal{N}_{i,s's}\left(\mathcal{H}_{i,s's}(t)\right), \tag{10}$$

$$\mathcal{H}_{i,ss'}(t) = \left\{(\tau, x) \in \mathbb{R}^2 \mid 0 \leq \tau \leq t, \ 0 \leq x \leq q_{ss'}(\phi_i(\tau))\xi_{i,s'}(\tau)\right\},$$

where for each choice of $1 \leq i \leq N$ and $s \neq s' \in \mathcal{S}$, $(\mathcal{N}_{i,ss'}(x, y) : x, y \geq 0)$ is a 2-dimensional Poisson-process with density 1, and the processes are independent for different (i, s, s') triples.

(10) is a cumulative formula counting all transitions of the vertex i to and from state s up to time t; $s \leftarrow s'$ transitions are generated using the Poisson points in the 2-dimensional domain $\mathcal{H}_{i,ss'}(t)$ which has area $\int_0^t q_{ss'}(\phi_i(\tau))\xi_{i,s'}(\tau)d\tau$, ensuring the proper transition rate for $s \leftarrow s'$ jumps at time τ. The second term of the sum corresponds to $s' \leftarrow s$ transitions in a similar manner.

2.3. N-intertwined mean field approximation

Although the state occupation probabilities of the population process can be described by the Chapman–Kolmogorov equations, the number of equations is $|\mathcal{S}|^N$, making it infeasible for numeric or analytic investigations even for moderate sized populations. To address this issue, several approximation schemes had been introduced in the literature with varying complexity.

This chapter discusses the quenched mean field approximation [8], also called the N-intertwined mean field approximation (NIMFA). NIMFA preserves all information regarding the graph structure and only neglects dynamical correlation between vertices. The goal is to derive state occupation probabilities for each vertex separately, resulting in a total of $|\mathcal{S}|N$ equations.

A possible intuition for NIMFA is as follows.

$$\frac{d}{dt}\mathbb{E}\left(\xi_i(t)\right) = \mathbb{E}\left[Q\left(\phi_i(t)\right)\xi_i(t)\right] \tag{11}$$

can be derived from (10). To close (11), we apply the approximation $\phi_i(t) \approx \mathbb{E}(\phi_i(t))$, which is reasonable when N is large and there is low correlation between vertices:

$$\mathbb{E}\left[Q\left(\phi_i(t)\right)\xi_i(t)\right] \approx \mathbb{E}\left[Q\left(\mathbb{E}\left(\phi_i(t)\right)\right)\xi_i(t)\right] = Q\left(\mathbb{E}\left(\phi_i(t)\right)\right)\mathbb{E}\left(\xi_i(t)\right).$$

Accordingly, the NIMFA approximation $z_i(t) = (z_{i,s}(t))_{s \in \mathcal{S}}$, $1 \leq i \leq N$ is the solution of the system

$$\frac{d}{dt}z_i(t) = Q\left(\zeta_i(t)\right)z_i(t),$$

$$\zeta_i(t) = \left(\zeta_i^{(m)}(t)\right)_{m=1}^M, \tag{12}$$

$$\zeta_i^{(m)}(t) = \left(\zeta_{i,\underline{s}}^{(m)}(t)\right)_{\underline{s} \in [N]^m} = \left(\sum_{\underline{j} \in \mathcal{S}^m} w_{i,\underline{j}}^{(m)} z_{\underline{j},\underline{s}}^{(m)}(t)\right)_{\underline{s} \in \mathcal{S}^m},$$

where $z_i(t)$ corresponds to $\xi_i(t)$ and $\zeta_i(t)$ corresponds to $\phi_i(t)$, and then the approximation used is

$$\mathbb{P}\left(\xi_{i,s}(t) = 1\right) = \mathbb{E}\left(\xi_{i,s}(t)\right) \approx z_{i,s}(t).$$

The following theorem ensures the existence and uniqueness of the solution of (12).

Theorem 1. *Let $\Delta^\mathcal{S}$ denote the set of probability vectors from $\mathbb{R}^\mathcal{S}$. For any initial condition $z_i(0) \in \Delta^\mathcal{S}$ for all i the ODE system* (12) *has a unique global solution such that $z_i(t) \in \Delta^\mathcal{S}$ for all i and $t > 0$ as well.*

2.4. Examples

In this section we give some examples for models covered by the formalism of Section 2.2.

The simplicial SIS model

We will use the simplicial SIS model, also referred to as the contact process as a running example.

In the $M = 1$ case (graphs) the setup is the following: Each vertex can be in one of two states: susceptible (S) and infected (I), hence the state space is $\mathcal{S} = \{S, I\}$. Infected vertices become susceptible at a constant rate $\gamma \geq 0$ while susceptible vertices receive the illness with rate proportional to number of its infected neighbors.

The number of infected neighbors of vertex $i \in [N]$ at time t equals to

$$\sum_{j=1}^{N} a_{ij} \xi_{j,I}(t)$$

as $a_{ij} \xi_{j,I}(t)$ the indicator of vertex j is connected to vertex i and that it is infected at time t. After normalizing it with $\bar{d}$ or $d(i)$ depending on our choice of convention 1 or 2 one gets

$$\sum_{j=1}^{N} w_{ij} \xi_{j,I}(t) = \phi_{i,I}(t).$$

Therefore, the transition rates takes the form $q_{SI}(\phi_i(t)) = \gamma$, $q_{IS}(\phi_i(t)) = \beta \phi_{i,I}(t)$ where $\beta \geq 0$ is a suitable constant factor. In matrix form:

$$Q(\phi_i(t)) = \begin{bmatrix} -\gamma & \gamma \\ \beta \phi_{i,I}(t) & -\beta \phi_{i,I}(t) \end{bmatrix}$$

For the SIS process NIMFA takes the form:

$$\frac{d}{dt} z_{i,I}(t) = -\gamma z_{i,I}(t) + \beta (1 - z_{i,I}(t)) \sum_{j=1}^{N} w_{ij} z_{j,I}(t).$$

Here we used $z_{i,S}(t) = 1 - z_{i,I}(t)$ which is also the reason why it enough to write the I components only.

The extension of the SIS model to hypergraphs is called the simplicial SIS model. The curing rate stays γ, however the infection dynamics is modified. A susceptible vertex can be infected via any $(m + 1)$-edge if all other m vertices are infected. The weighted sum of such edges $(m + 1)$-edges is

$$\sum_{\underline{j} \in [N]^m} w_{i,\underline{j}}^{(m)} \xi_{\underline{j},(I,\dots,I)}^{(m)}(t) = \phi_{i,(I,\dots,I)}^{(m)}(t).$$

The infection rates is sum of all the $1 \leq m \leq M$ with appropriate $\beta_1, \dots, \beta_M \geq 0$ factors:

$$q_{IS}(\phi_i(t)) = \sum_{m=1}^{M} \beta_m \phi_{i,(I,\dots,I)}^{(m)}(t).$$

For the simplicial SIS model NIMFA takes the form

$$\frac{d}{dt} z_{i,I}(t) = -\gamma z_{i,I}(t) + \left(1 - z_{i,I}(t)\right) \sum_{m=1}^{M} \beta_m \sum_{\underline{j} \in [N]^m} w_{i,\underline{j}}^{(m)} z_{\underline{j},(I,\dots,I)}^{(m)}(t),$$

Glauber dynamics

Glauber dynamics is a stochastic process whose stationary distribution coincides with the distribution given by a spin system, such as the Ising model [19].

There are two possible states: $\mathcal{S} = \{+, -\}$. Instead of the indicators

$$\xi_{i,+}(t), \; \xi_{i,-}(t)$$

it is customary use the sign variables

$$\sigma_i(t) := \xi_{i,+}(t) - \xi_{i,-}(t) = 2\xi_{i,+}(t) - 1.$$

In physical systems it is natural to assume w_{ij} is symmetric and $w_{ii} = 0$.
The dynamics is the following:

- At each time step, choose a vertex i uniformly.

- With probability $p_i(\sigma) = \frac{e^{\beta S_i(\sigma)}}{e^{\beta S_i(\sigma)}+1}$, vertex i switches to state $+$ (else $-$), where

$$S_i(\sigma) = \sum_{j=1}^{N} w_{ij}\sigma_j.$$

Note that $S_i(\sigma)$ arises from the reduction of the energy

$$H(\sigma) := -\frac{1}{2}\sum_{i<j} w_{ij}\sigma_i\sigma_j$$

when vertex i is flipped from $-$ to $+$. The stationary distribution is then given by the Gibbs measure

$$P(\sigma) := \frac{1}{Z}e^{-\beta H(\sigma)},$$
$$Z := \sum_{\sigma} e^{-\beta H(\sigma)}.$$

We modify the above dynamics. First, note that, in accordance with (7),

$$S_i(\sigma(t)) = \sum_{j=1}^{N} w_{ij}\left(\xi_{j,+}(t) - \xi_{j,-}(t)\right) = \phi_{i,+}(t) - \phi_{i,-}(t).$$

With a slight abuse of notation, we denote

$$S(\phi_i(t)) := \alpha\phi_{i,+}(t) - \gamma\phi_{i,-}(t),$$

allowing the dynamics to have a preferred state.

Furthermore, we turn to the continuous time version instead with transition rates given by

$$q_{+-}(\phi) = e^{\beta S(\phi)},$$
$$q_{-+}(\phi) = 1.$$

Since there are only two states, it is enough to consider the probabilities of occupying state $+$. For this, NIMFA gives the following system of ODEs:

$$\frac{d}{dt}z_{i,+}(t) = (1 - z_{i,+}(t))e^{\beta S(\zeta_i(t))} - z_{i,+}(t). \tag{13}$$

The equilibrium state is given by the fixed point problem

$$z_{i,+} = \frac{e^{\beta S(\zeta_i)}}{e^{\beta S(\zeta_i)} + 1}. \tag{14}$$

Assume $\alpha = 1$, $\gamma = -1$ as in the original setting and that the underlying weighted graph is regular: $\forall i \; \delta(i) = \sum_j w_{ij} = 1$. Than (14) reduces to

$$\sigma = \tanh\left(\frac{1}{2}\beta\sigma\right),$$
$$\forall i \; 2z_{i,+} - 1 = \sigma$$

giving back the classical mean field approximation of the Ising model on lattice. This is not surprising as both NIMFA and the classical mean field approach is based on the assumption of independence of vertices.

Based on [20], we can generalize the model for hypergraphs via extending $S(\phi)$ to

$$S(\phi_i(t)) := \sum_{m=0}^{M} \alpha_m\phi_{i,(+,\ldots,+)}^{(m)}(t) - \gamma_m\phi_{i,(-,\ldots,-)}^{(m)}(t)$$

allowing the system to lose even more energy when 3 or more neighbors have the same configuration on a hyper-edge.

The voter model

The voter model is a conceptually simple stochastic process modeling opinion dynamics [21]. In the most simple case, there are two possible states: $S = 0, 1$.

The dynamics can be described the following way: At each time step, we choose a vertex uniformly. Said vertex chooses an neighbor also uniformly, and copies its state. Similarly to the Glauber dynamics, we will study the continuous time version instead.

For vertex i, the ratio of neighbors sharing belief $s \in \{0, 1\}$ is

$$\frac{1}{d(i)} \sum_{j=1}^{N} a_{ij}\xi_{j,s}(t) = \phi_{i,s}(t)$$

with the choice of Convention 2. Hence, the transition rates take the form

$$q_{01}(\phi_i(t)) = \lambda\phi_{i,0}(t),$$
$$q_{10}(\phi_i(t)) = \lambda\phi_{i,1}(t) = \lambda\left(1 - \phi_{i,0}(t)\right).$$

Using $z_{i,1}(t) = 1 - z_{i,0}(t)$, NIMFA can be written as

$$\frac{d}{dt}z_{i,0}(t) = -\lambda(1 - \zeta_{i,0}(t))z_{i,0}(t) + \lambda\zeta_{i,0}(t)\left(1 - z_{i,0}(t)\right).$$

A modified majority rule model

Another popular model of opinion dynamics is the majority rule [21]. In this setting a group of $m + 1$ individuals are chosen who update their state simultaneously to the majority opinion. Ties are usually broke with either a random choice or setting a preferred opinion to win in this case, say opinion 1. For the sake of simplicity, we apply the latter approach.

Due to the continuous time setting we use, we modify the majority rule such that only one individual updates its opinion during a transition based on the state of the other vertices (not including its own opinion for the sake of simplicity).

As it is stated in [21], the hypergraph setting is more suitable for majority rule. We assume communities have a bounded size $M + 1$, while each individual can be a part of many, possibly overlapping communities.

$a_{i,j_1,\dots,j_m}^{(m)}$ is the indicator of vertices $i, j_1, \dots, j_m \in [N]$ being in a community. We assume symmetry in the indices and set $a_{i,j_1,\dots,j_m}^{(m)} = 0$ if there are duplicates. We use a slightly modified version of Convention 1:

$$w_{i,\underline{j}}^{(m)} = \frac{\alpha_m a_{i,\underline{j}}}{m!\bar{d}^{(m)}},$$

where α_m measures how much importance vertices put on communities of size $m + 1$. $w_{\max} = \max_m \frac{\alpha_m}{m!\bar{d}^{(m)}}$ can be small either due to vertices being part of many communities of size $m + 1$ on average or because they put less importance on said communicates.

Introduce the notation $|\underline{s}| = \sum_{l=1}^{m} s_l$. Vertex i in community $i, j_1, \dots, j_m$ changes its opinion to the majority of $j_1, \dots, j_m$ at rate $w_{i,\underline{j}}^{(m)}$. Therefore,

$$q_{01}(\phi_i(t)) = \sum_{m=0}^{M} \sum_{\underline{j}\in[N]^m} w_{i,\underline{j}}^{(m)} \mathbb{1}_{\{0 \text{ is the majority for } j_1,\dots,j_m\}}$$

$$= \sum_{m=0}^{M} \sum_{\underline{j}\in[N]^m} w_{i,\underline{j}}^{(m)} \sum_{|\underline{s}|<\frac{m}{2}} \prod_{l=1}^{m} \xi_{j_l,s_l}(t) = \sum_{m=0}^{M} \sum_{|\underline{s}|<\frac{m}{2}} \phi_{i,\underline{s}}^{(m)}(t),$$

$$q_{10}(\phi_i(t)) = \sum_{m=0}^{M} \sum_{|\underline{s}|\geq\frac{m}{2}} \phi_{i,\underline{s}}^{(m)}(t).$$

The NIMFA ODEs are

$$\frac{d}{dt}z_{i,0}(t) = (1 - z_{i,0}(t))\sum_{m=0}^{M} \sum_{|\underline{s}|<\frac{m}{2}} \zeta_{i,\underline{s}}^{(m)}(t) - z_{i,0}(t)\sum_{m=0}^{M} \sum_{|\underline{s}|\geq\frac{m}{2}} \zeta_{i,\underline{s}}^{(m)}(t).$$

3. Error bounds for NIMFA

In this section we are presenting our main results which bound the error arising from neglecting the dynamical correlation between vertices.

Recall that (11) was closed by assuming $\phi_i(t) \approx \mathbb{E}(\phi_i(t))$. We introduce an auxiliary process where the empirical neighborhood $\phi_i(t)$ is replaced by the approximate $\zeta_i(t)$ from (12):

$$\hat{\xi}_{i,s}(t) = \xi_{i,s}(0) + \sum_{\substack{s'\in\mathcal{S}\\s'\neq s}} \mathcal{N}_{i,ss'}\left(\mathcal{K}_{i,ss'}(t)\right) - \mathcal{N}_{i,s's}\left(\mathcal{K}_{i,s's}(t)\right),$$

$$\mathcal{K}_{i,ss'}(t) = \left\{(\tau, x) \in \mathbb{R}^2 \mid 0 \leq \tau \leq t, \ 0 \leq x \leq q_{ss'}\left(\zeta_i(\tau)\right)\hat{\xi}_{i,s'}(\tau)\right\}.$$

(15)

The process $\hat{\xi}_{i,s}(t)$ is an indicator process just like $\xi_{i,s}(t)$, so it takes 0 or 1 values, and $\sum_{s \in S} \hat{\xi}_{i,s}(t) = 1$ for any $i \in [N]$ and $t \geq 0$. However, assuming independent initial conditions, $\hat{\xi}_i(t)$ remain independent. Applying total expectation to (15) shows

$$\frac{d}{dt} \mathbb{E} \left(\hat{\xi}_i(t) \right) = Q \left(\zeta_i(t) \right) \mathbb{E} \left(\hat{\xi}_i(t) \right),$$

which, along with (12), implies that if $\mathbb{E} \left(\hat{\xi}_i(0) \right) = z_i(0)$, then $\hat{\xi}_i(t) - z_i(t)$ is a martingale and

$$\mathbb{E} \left(\hat{\xi}_i(t) \right) = z_i(t) \quad \forall t \geq 0. \tag{16}$$

Using the same background Poisson processes $\mathcal{N}_{i,ss'}$ provides a coupling between ξ and $\hat{\xi}$ that will be useful later on.

We aim to give an upper bound for $|\hat{\xi}(t) - \xi(t)|$, as well as for $|\hat{\xi}(t) - z(t)|$. We start with $|\hat{\xi}(t) - \xi(t)|$ by introducing the error terms

$$D_i^{(0)}(t) = \sup_{0 \leq \tau \leq t} \mathbb{E} \left(\sum_{s \in S} \left| \xi_{i,s}(\tau) - \hat{\xi}_{i,s}(\tau) \right| \right),$$

$$\tilde{D}_i^{(0)}(t) = \mathbb{E} \left(\sup_{0 \leq \tau \leq t} \sum_{s \in S} \left| \xi_{i,s}(\tau) - \hat{\xi}_{i,s}(\tau) \right| \right).$$

Apparently, the only difference between the two is the order in which we take the supremum in time. $\tilde{D}_i^{(0)}(t)$ is more strict as

$$D_i^{(0)}(t) \leq \tilde{D}_i^{(0)}(t).$$

Observe that $\sum_{s \in S} \left| \xi_{i,s}(\tau) - \hat{\xi}_{i,s}(\tau) \right|$ only has two possible values: 0 if $\xi_i(t) = \hat{\xi}_i(t)$, and 2 otherwise (as there will be two $s \in S$ indices where $\xi_{i,s}(t), \hat{\xi}_{i,s}(t)$ differs). This implies

$$\sup_{0 \leq \tau \leq t} \mathbb{P} \left(\xi_i(\tau) \neq \hat{\xi}_i(\tau) \right) = \frac{1}{2} D_i^{(0)}(t),$$

$$\mathbb{P} \left(\exists \, 0 \leq \tau \leq t : \xi_i(\tau) \neq \hat{\xi}_i(\tau) \right) = \frac{1}{2} \tilde{D}_i^{(0)}(t)$$

We also introduce error terms describing the environments arising from $\xi_i(t)$ and $\hat{\xi}_i(t)$:

$$D_i^{(m)}(t) = \sup_{0 \leq \tau \leq t} \mathbb{E} \left[\sum_{\underline{s} \in S^m} \left| \phi_{i,\underline{s}}^{(m)}(\tau) - \zeta_{i,\underline{s}}^{(m)}(\tau) \right| \right] \quad (1 \leq m \leq M),$$

$$\tilde{D}_i^{(m)}(t) = \mathbb{E} \left[\sup_{0 \leq \tau \leq t} \sum_{\underline{s} \in S^m} \left| \phi_{i,\underline{s}}^{(m)}(\tau) - \zeta_{i,\underline{s}}^{(m)}(\tau) \right| \right] \quad (1 \leq m \leq M).$$

Since the neighborhoods $\phi_i(t)$ and $\zeta_i(t)$ are constructed from the indicators $\xi_i(t)$ and $\hat{\xi}_i(t)$, it is reasonable to expect $\phi_i(t)$ and $\zeta_i(t)$ to be close to each other − as long as $\xi_i(t)$ and $\hat{\xi}_i(t)$ are also close. To avoid circular reasoning, we carry on handling these two types of errors together at the same time. This motivates the introduction of

$$D_{\max}^{(m)}(t) = \max_{i \in [N]} D_i^{(m)}(t),$$

$$D_{\max}(t) = \sum_{m=0}^{M} D_{\max}^{(m)}(t),$$

$$\tilde{D}_i(t) = \sum_{m=0}^{M} \tilde{D}_i^{(m)}(t).$$

The vector notation $\tilde{D}(t) = \left(\tilde{D}_i(t) \right)_{i \in [N]}$ will also be utilized.

Now we can go ahead to state the main results of the paper. The idea behind the statements is when the vertex weights are generally small (the network is well-distributed) then vertices has low correlation between each other, hence NIMFA is accurate.

Theorem 2 (*Main*). *Assume the initial conditions $\xi_i(0)$ are independent and* (16) *is satisfied. Then for every $t \geq 0$ there is a constant $C = C(t, \delta_{\max}, R)$ such that*

$$\max_i \sup_{0 \leq \tau \leq t} \mathbb{P}\left(\xi_i(\tau) \neq \hat{\xi}_i(\tau)\right) \leq \frac{1}{2}D_{\max}(t) \leq C\sqrt{w_{\max}}. \tag{17}$$

Furthermore, if we additionally assume $M = 1$ (having 1-uniform hypergraphs) then there exist constants $C_1 = C_1(\delta_{\max}), C_2 = C_2(\delta_{\max})$ such that for all $t \geq 0$

$$\left\|\tilde{D}(t)\right\| \leq C_1 \exp\left(C_2 \|W + I\| t\right) \|\mu\|,$$

$$\mu = \left(\sqrt{\sum_{j=1}^{N} w_{ij}^2}\right)_{i \in [N]}, \tag{18}$$

where the norm $\|\cdot\|$ is arbitrary, $W = \left(w_{ij}\right)_{i,j=1}^{N}$ and I is the identity matrix.

Remark 1. The reason why we have different results for $M > 1$ and $M = 1$ is technical in nature. The main observation is that in the $M = 1$ case $\hat{\xi}_{i,\underline{s}}(t) - z_{i,\underline{s}}(t)$ is a martingale making possible to take $\sup_{0 \leq \tau \leq t}$ inside the expectation via Doob's inequality. It is no longer the case for $M > 1$ where $\hat{\xi}_{i,\underline{s}}^{(m)}(t) - z_{i,\underline{s}}^{(m)}(t)$ is typically not a martingale itself.

(17) is a local result in the sense that it provides a uniform bound, ensuring that $\hat{\xi}_{i,s}(t)$ and $\xi_{i,s}(t)$ are close for all vertices i simultaneously. For example, in the SIS process it allows us to approximate infection probabilities for concrete individuals, not just global or mesoscopic ratios.

(18) will be elaborated on in Theorem 3.

In general, we cannot expect a similar local result for $\hat{\xi}_{i,s}(t)$ and $z_{i,s}(t)$ since $\hat{\xi}_{i,s}(t)$ is an indicator while $z_{i,s}(t)$ is a continuous variable. However, if we average out $\hat{\xi}_{i,s}(t)$ over a macroscopic set of vertices, a similar result will hold.

In (18) the use of ℓ^2 or ℓ^∞ is advised. Observe

$$\|W\|_\infty = \max_i \sum_j w_{ij} \leq \delta_{\max}$$

$$\|W\|_2 \leq \sqrt{\|W\|_1 \|W\|_\infty} = \sqrt{\left(\max_j \sum_i w_{ij}\right)\left(\max_j \sum_i w_{ij}\right)} \leq \sqrt{\delta_{\max}^{\text{out}} \delta_{\max}}, \tag{19}$$

making $\exp\left(C_2 \|W + I\| t\right)$ bounded in (18). Note that (19) is the only step where Assumption (6) regarding $\delta_{\max}^{\text{out}}$ is used.

As for $\|\mu\|$:

$$\|\mu\|_\infty = \max_{1 \leq i \leq N} \sqrt{\sum_{j=1}^{n} w_{ij}^2} \leq \max_{1 \leq i \leq N} \sqrt{w_{\max} \sum_{j=1}^{n} w_{ij}} \leq \sqrt{w_{\max} \delta_{\max}},$$

$$\|\mu\|_2 = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}^2}.$$

Convention 1 works well with the $O\left(\sqrt{w_{\max}}\right)$ error bound as $w_{\max} = \frac{1}{d}$ holds in that case suggesting vertices being close to independent when they have a lot of neighbors on average. Similarly to (17), it also gives a uniform error bound, making it possible to approximate the probabilities at the individual level. For Convention 2 on the other hand, $w_{\max} = \frac{1}{d_{\min}}$ is sensitive to even one vertex with a low degree. If we are not attached to uniform bounds in i, we can provide a more robust on for the error of a typical vertex, thus, it is possible to describe global or mesoscopic population statistics.

Let $\iota \sim U([N])$ the index of a randomly chosen vertex.

$$\mathbb{P}\left(\exists \tau \in [0, t] : \xi_\iota(\tau) \neq \hat{\xi}_\iota(\tau)\right) =$$

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{P}\left(\exists \tau \in [0, t] : \xi_i(\tau) \neq \hat{\xi}_i(\tau)\right) \leq$$

$$\frac{1}{2N} \sum_{i=1}^{N} \tilde{D}_i(t) \leq \sqrt{\frac{1}{4N} \sum_{i=1}^{N} \tilde{D}_i^2(t)} = O\left(\sqrt{\frac{1}{N} \|\mu\|_2^2}\right)$$

Observe

$$\frac{1}{N}\|\mu\|_2^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}^2 \tag{20}$$

is the squared and normalized Frobenius norm of the matrix W. We mention that such bound were used in [13] under more strict assumptions regarding W.

Note that for Convention 2

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{a_{ij}}{d^2(i)} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{d(i)}, \tag{21}$$

meaning the error is small when vertices typically have large degrees.

These observations along with Theorem 2 give the following result:

Theorem 3. *For $M = 1$ (directed, weighted graphs), there exist constants $C_1 = C_1(t, \delta_{max})$ and $C_2 = C_2(t, \delta_{max}, \delta_{max}^{out})$ such that*

$$\max_i \mathbb{P}\left(\exists \tau \in [0, t]: \xi_i(\tau) \neq \hat{\xi}_i(\tau)\right) \leq C_1 \sqrt{w_{max}}, \tag{22}$$

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{P}\left(\exists \tau \in [0, t]: \xi_i(\tau) \neq \hat{\xi}_i(\tau)\right) \leq C_2 \sqrt{\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}^2}. \tag{23}$$

So far, we have only accounted for the error between $\xi_i(t)$ and $\hat{\xi}_i(t)$, however, what we are actually interested in is the expectation $\mathbb{E}\left(\hat{\xi}_i(t)\right) = z_i(t)$, the solution of the ODE system given by NIMFA. Thankfully, $\left(\hat{\xi}_i(t)\right)_{i \in [N]}$ are independent, hence, their averages must concentrate around the mean:

Theorem 4. *Assume* (16) *holds with independent initial conditions. Then for any $t \geq 0$ and any $1 \leq K \leq N$,*

$$\mathbb{E}\left[\sup_{0 \leq \tau \leq t} \sum_{s \in \mathcal{S}} \left|\frac{1}{K}\sum_{i=1}^{K}\left(\hat{\xi}_{i,s}(\tau) - z_{i,s}(\tau)\right)\right|\right] \leq \frac{2|\mathcal{S}|}{\sqrt{K}}. \tag{24}$$

The most natural application of Theorem 4 is for $K = N$, but it is formulated in a way so that it can be applied to any convenient subset of vertices (the fact that the first K vertices are considered has no significance as the vertices can be reordered arbitrarily).

Together, Theorems 2–4 give an error bound for the NIMFA approximation (see Figs. 2 and 3).

Theorem 5. *Assume* (16) *holds with independent initial conditions. Then for any $t \geq 0$, there exists a constant $C = C(t, \delta_{\max}, R)$ such that*

$$\sup_{0 \leq \tau \leq t} \mathbb{E}\left(\sum_{s \in \mathcal{S}} \left|\frac{1}{N}\sum_{i=1}^{N}\left(\xi_{i,s}(\tau) - z_{i,s}(\tau)\right)\right|\right) \leq C\left(\sqrt{w_{\max}} + \frac{1}{\sqrt{N}}\right). \tag{25}$$

Furthermore, if we additionally assume $M = 1$, there exist constants $C_1 = C_1(t, \delta_{\max})$, $C_2 = C_2(t, \delta_{\max}, \delta_{max}^{out})$ such that

$$\mathbb{E}\left[\sup_{0 \leq \tau \leq t}\left(\sum_{s \in \mathcal{S}} \left|\frac{1}{N}\sum_{i=1}^{N}\left(\xi_{i,s}(t) - z_{i,s}(t)\right)\right|\right)\right] \leq C_1\left(\sqrt{w_{max}} + \frac{1}{\sqrt{N}}\right) \tag{26}$$

and

$$\mathbb{E}\left[\sup_{0 \leq \tau \leq t}\left(\sum_{s \in \mathcal{S}} \left|\frac{1}{N}\sum_{i=1}^{N}\left(\xi_{i,s}(t) - z_{i,s}(t)\right)\right|\right)\right] \leq C_2\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}^2} + \frac{1}{\sqrt{N}}\right) \tag{27}$$

where μ is the same as for Theorem 2.

Related works

In this section we compare our results to the recent independent work of Sridhar and Kar [12,13] and Parasnis et al. [14].
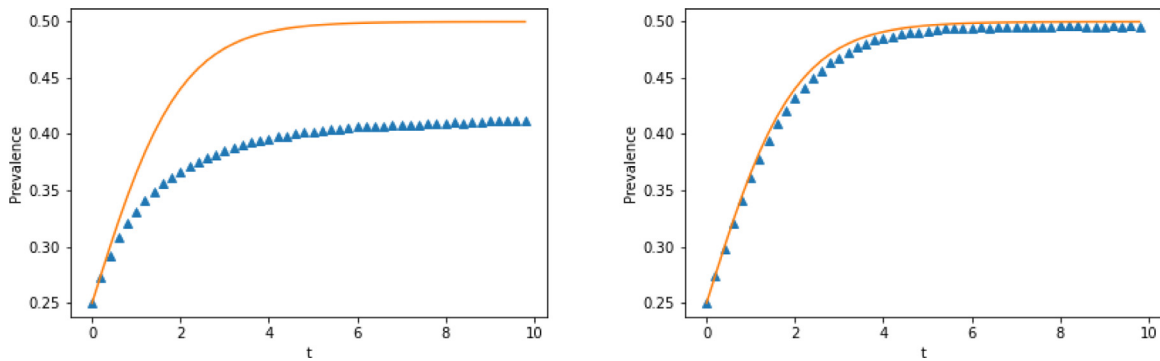
Fig. 2. The ratio of infected based on the average of 1000 simulations (triangles) compared to the estimate of NIMFA (solid line) on an $N = 1000$ vertex modified cycle graphs with the closest 10 (left) and 100 (right) neighbors being connected. ($\beta = 2, \gamma = 1$) As we increase the degrees NIMFA performs better.
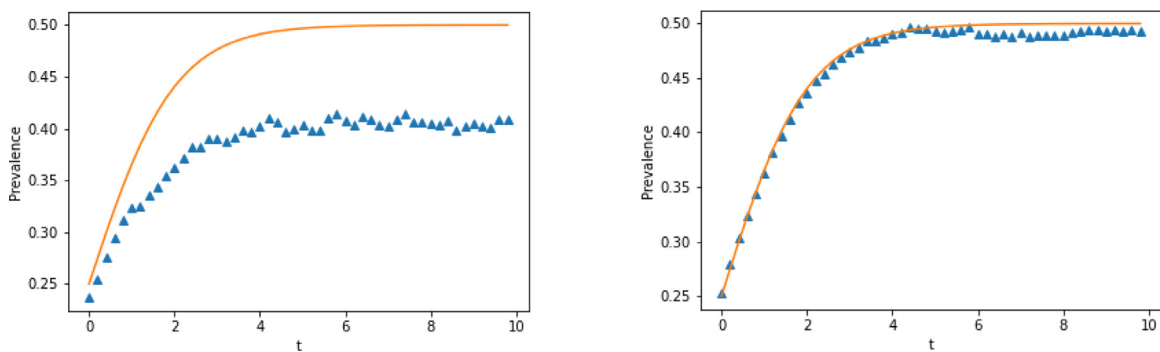
Fig. 3. The ratio of infected based on the average of 10 simulations (triangles) compared to the estimate of NIMFA (solid line) on an $N = 5000$ vertex modified cycle graphs with the closest 10 (left) and 100 (right) neighbors being connected. ($\beta = 2, \gamma = 1$) As we increase the degrees NIMFA performs better.

In [12] the authors describe how the state densities of certain related stochastic processes on weighted graphs with doubly symmetric matrix W can be approximated by a set of $O(N)$ ODEs analogous to NIMFA given that the normalized Frobenius norm $\frac{1}{N} \sum_{i=1} \sum_{j=1}^{N} w_{ij}^2$ is small and N is large.

Given the conclusions of Theorem 4.2 in [12] and Theorem 5 in the present paper are very similar in nature, it makes sense to compare the general setup, the conditions, the conclusions and the technique directly to those in the present paper.

Setup. Strictly speaking, the stochastic processes discussed in the present paper and in [12,13] are different. In our work, time is continuous while [12,13] start from discrete time steps then speed up time. This is a minor difference though, and with appropriate time scaling, the models in [12,13] and the present paper define essentially the same object.

Conditions. In the present paper, we require only that the normalized degrees are bounded. This is more general than the doubly stochastic W assumption of [12,13]. Specifically, our result also justifies Example 4.2 in [13].

Via (27), qualitatively the same type of error terms were retained in terms of the normalized Frobenius norm, but [12,13] provides an error probability bound that is exponential in N. In the present paper, we do not focus on this kind of large deviation bound in N.

[12,13] derive bounds for the global average. On the other hand, our results show more localized, uniform bounds in terms of vertices. This is made possible by the use of the auxiliary Markov processes $\hat{\xi}_i(t)$, allowing accurate predictions about individual vertices too, not just global averages. Our framework also allows higher order interactions, while [12,13] is restricted to order 2 interactions (graphs).

In [14] the authors study the SIR process in age-structured populations on time-varying networks. They show that when N and the rewiring rate is high the prevalence of the age groups can be described via an ODE system analogous to the metapopulation NIMFA model (34) in Section 4.2. Note that [14] applies to cases with fast, but finite rewiring rates as well, while our result only considers the idealized case of infinite rewiring rates.

4. Further reductions to NIMFA

This section relates NIMFA to other approaches from the literature. Although NIMFA is a major reduction of the exact Kolmogorov-equations, requiring only $O(N)$ ODEs to be solved, it can be still computationally prohibitive when the number of vertices is too large. Furthermore, NIMFA requires knowing both the full network structure and precise initial conditions for all vertices. We look at further reductions to (12) when additional structure is known for the network or initial conditions; several of these actually lead to other well-known models from the literature.

4.1. Homogeneous mean field approximation

The homogeneous mean field approximation (HMFA) assumes that the vertices are *well mixed*, meaning, every vertex interacts with every other with equal weights. Formally, this can be described by a complete hypergraph (with all loops and secondary loops):

$$w_{i,j}^{(m)} = \frac{1}{N^m}.$$

This definition may be generalized to include cases when $w_{i,j}^{(m)} = 0$ for certain m indices, e.g. $(M+1)$-uniform hypergraphs. For ease of notation, instead of modifying the definition of $w_{i,j}^{(m)}$, it is also possible to choose the rate functions $q_{ss'}(\phi)$ so that they do not depend on the appropriate $\phi^{(m)}$ coordinates, making the choice of $w_{i,j}^{(m)}$ irrelevant.

It is easy to see that for such networks, $w_{\max} = \frac{1}{N}$ and $\delta_{\max} = 1$. What remains to show is that (5) holds with some bounded R.

$$\sum_{\substack{\underline{j} \in [N]^m \\ \underline{j} \text{ iss. loop}}} w_{i,\underline{j}}^{(m)} = \frac{1}{N^m} \left| \left\{ \underline{j} \in [N]^m \middle| \underline{j} \text{ s. loop} \right\} \right| =$$

$$1 - \frac{1}{N^m} \left| \left\{ \underline{j} \in [N]^m \middle| \underline{j} \text{ not s. loop} \right\} \right| = 1 - \prod_{l=0}^{m-1} \left(1 - \frac{l}{N} \right) = \tag{28}$$

$$O\left(\frac{1}{N} \right) \ll \frac{1}{\sqrt{N}} = \sqrt{w_{\max}},$$

hence, R can be chosen arbitrarily small for large enough N.

Our goal now is to derive a small system of equations for

$$u(t) := \frac{1}{N} \sum_{i=1}^{N} z_i(t).$$

Our strategy is based on the observation that the neighborhood vectors $\zeta_i(t)$ are the same for all vertices.

$$\zeta_{i,\underline{s}}^{(m)}(t) = \frac{1}{N^m} \sum_{\underline{j} \in [N]^m} \prod_{l=1}^{m} z_{j_l, s_l}(t) = \prod_{l=1}^{m} \left(\frac{1}{N} \sum_{j_l=1}^{N} z_{j_l, s_l}(t) \right) =$$

$$\prod_{l=1}^{m} u_{s_l}(t) =: u_{\underline{s}}^{(m)}(t)$$

This results in the ODE system:

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = Q\left(U(t) \right) u(t),$$

$$U(t) = \left(u^{(m)}(t) \right)_{m=1}^{M}, \tag{29}$$

$$u^{(m)}(t) = \left(u_{\underline{s}}^{(m)}(t) \right)_{\underline{s} \in \mathcal{S}^m} = \left(\prod_{l=1}^{m} u_{s_l}(t) \right)_{\underline{s} \in \mathcal{S}^m}.$$

For example, the simplicial SIS model (29) takes the form

$$\frac{\mathrm{d}}{\mathrm{d}t} u_I(t) = -\gamma u_I(t) + (1 - u_I(t)) \sum_{m=1}^{M} \beta_m u_I^m(t).$$

which was used in [16].

In this setting, Theorem 5 shows the ratio of vertices in state $s \in \mathcal{S}$ can be approximated by $u_s(t)$ with $O\left(\frac{1}{\sqrt{N}} \right)$ error. The well known results of Kurtz [6,7] correspond to the $M = 1$ case.

Regular hypergraphs

Although (29) is both feasible for analytical and numerical investigations (due to its finite size) the assumption that the network structure is well-mixed is quite restrictive. However, as we will see, the well-mixed condition can be relaxed given uniform initial conditions.

We call a weighted hypergraph *regular* if

$$\forall\ 1 \le i \le N,\ 1 \le m \le M\ \ \delta^{(m)}(i) = 1. \tag{30}$$

Note that the value 1 is arbitrary and any other constant value would work with minor modifications to the rate functions $q_{ss'}$.

We note that (30) always holds for Convention 2 hypergraphs. For Convention 1, it holds when $d^{(m)}(i) = \bar{d}^{(m)}\ \forall 1 \le i \le N,\ 1 \le m \le M$ (that is, the hypergraph is regular in the usual sense).

Proposition 1. *Assume* (30) *and*

$$z_i(0) = u(0)\ \ \forall\ 1 \le i \le N$$

for some $u(0) \in \Delta^{\mathcal{S}}$. *Then the solution of* (12) *takes the form*

$$z_i(t) = u(t)\ \ \forall\ 1 \le i \le N$$

where $u(t)$ *satisfies* (29).

We mention that statements similar to Proposition 1 have appeared in the literature before in certain special cases [28, Proposition 3.18]. Combining Proposition 1 with Theorem 2 ensures the accuracy of the homogeneous mean field approximation on regular graphs with large degrees and homogeneous initial conditions disregarding any further network structure.

Proof of Proposition 1. Let $u(t)$ be the solution of (29). Set $z_i(t) = u(t)$. We have to show that $z_i(t)$ satisfies (12). The initial conditions are satisfied according to the assumption, and for the derivatives,

$$u_{\underline{s}}^{(m)}(t) = u_{\underline{s}}^{(m)}(t)\delta^{(m)}(i) = u_{\underline{s}}^{(m)}(t) \sum_{\underline{j} \in [N]^m} w_{i,\underline{j}}^{(m)} = \sum_{\underline{j} \in [N]^m} w_{i,\underline{j}}^{(m)} z_{\underline{j},\underline{s}}^{(m)}(t) = \zeta_{i,\underline{s}}^{(m)}(t),$$

$$\frac{\mathrm{d}}{\mathrm{d}t} z_i(t) = \frac{\mathrm{d}}{\mathrm{d}t} u(t) = Q\left(U(t)\right) u(t) = Q\left(\zeta_i(t)\right) z_i(t). \quad \square$$

4.2. Metapopulation models

As we saw in Section 4.1 , a way to reduce the number of equations is by grouping vertices together and representing them by a single averaged-out term. In practice, this approach will only work if the vertices grouped together are sufficiently homogeneous, which is typically not the case for the entire population. To mitigate this issue, we may introduce *communities*, inside which we assume homogeneity, then derive the dynamics between communities. This "higher resolution" may increase accuracy, at the cost of a larger ODE system.

In practice, the communities can be chosen by demographic and geographic criterion such as age and location. Alternatively, it is also possible to group vertices according to degree, or a third option is the use of community detection algorithms [29].

We present the general setup for metapopulation models first for graphs in Section 4.2.1, then for hypergraphs in Section 4.2.2.

For the SIS process on graphs similar results had been derived in [30].

4.2.1. Metapopulation models on graphs

First, assume $M = 1$. Divide the vertices into a partition $V_1, \ldots, V_K$ with size $|V_k| = N_k$ such that vertices inside a group are similar in some sense. The average weight between group V_k and V_l is

$$\tilde{w}_{kl} = \frac{\sum_{i \in V_k} \sum_{j \in V_l} w_{ij}}{N_k N_l}. \tag{31}$$

(In the idealized case of metapopulations, w_{ij} would have the same value $\tilde{w}_{kl}$ for each $i \in V_k, j \in V_l$ pair.)

Next we derive the dynamics for the averages

$$\bar{z}_k(t) := \frac{1}{N_k} \sum_{i \in V_k} z_i(t). \tag{32}$$

$\zeta_i(t)$ has the same value $\bar{\zeta}_k(t)$ for all $i \in V_k$:

$$\bar{\zeta}_k(t) = \zeta_i(t) = \sum_{j=1}^{N} w_{ij} z_j(t) = \sum_{l=1}^{K} \underbrace{N_l \tilde{w}_{kl}}_{\bar{w}_{kl}} \frac{1}{N_l} \sum_{j \in V_l} z_j(t) = \sum_{l=1}^{K} \bar{w}_{kl} \bar{z}_l(t). \tag{33}$$

Therefore, we can derive an ODE system for (32)

$$\frac{d}{dt}\bar{z}_k(t) = Q\left(\bar{\zeta}_k(t)\right)\bar{z}_k(t) \tag{34}$$

which is equivalent to (12) on the graph $\overline{\mathcal{G}}$ with vertex set $\{1, \ldots, K\}$ and weights $(\bar{w}_{kl})_{k,l=1}^{K}$.

4.2.2. Metapopulation models on hypergraphs

For the general metapopulation setting, we assume that for each $m = 1, \ldots, M$, the population is partitioned into local groups $V_1^{(m)}, \ldots, V_{K^{(m)}}^{(m)}$. The *type* of a vertex will be denoted by $k = \left(k^{(1)}, \ldots, k^{(M)}\right)$, which means that for each $m = 1, \ldots, M$, the given vertex is in the local group $V_{k^{(m)}}^{(m)}$. Vertices can be partitioned according to their type into $\prod_{m=1}^{M} K^{(m)}$ global groups.

We aim to define a hypergraph on the types, with weights consistent with the average of weights within each group. That said, with the above setup, this is easier to do using local groups for each $m = 1, \ldots, M$.

For a given m, $k^{(m)}$ and $\underline{l}^{(m)} = \left(l_1^{(m)}, \ldots, l_m^{(m)}\right)$, the *total local m-weight between $k^{(m)}$ and $\underline{l}^{(m)}$* is defined as

$$W_{k^{(m)}, \underline{l}^{(m)}}^{(m)} := \sum_{i \in V_{k^{(m)}}^{(m)}} \sum_{j_1 \in V_{l_1^{(m)}}^{(m)}} \cdots \sum_{j_m \in V_{l_m^{(m)}}^{(m)}} w_{i,\underline{j}}^{(m)}. \tag{35}$$

Then, using the notation

$$N_{\underline{l}^{(m)}} := \prod_{r=1}^{M} N_{l_r^{(m)}}^{(m)},$$

we define the weight of the edge containing the local groups $k^{(m)}, \underline{l}^{(m)}$ as

$$\tilde{w}_{k^{(m)}, \underline{l}^{(m)}}^{(m)} := \frac{W_{k^{(m)}, \underline{l}^{(m)}}^{(m)}}{N_{k^{(m)}} N_{\underline{l}^{(m)}}}. \tag{36}$$

Let $k(i) = \left(k^{(1)}(i), \ldots, k^{(M)}(i)\right)$ denote the type of i. For easier notation, we will often use $\iota \sim U([N])$, which is a random vertex independent from everything else. Then we define the average of $z_i(t)$ over type k as

$$\bar{z}_k(t) := \mathbb{E}\left(z_\iota(t) \mid k(\iota) = k\right) = \frac{1}{N_k} \sum_{i \in V_k} z_i(t). \tag{37}$$

In this case as well, $\zeta_i(t)$ has the same value for all $i \in V_k$; this common value will be denoted by $\bar{\zeta}_k(t)$. Let $\iota_1, \ldots, \iota_m$ denote i.i.d. copies of ι. Then

$$
\begin{aligned}
\bar{\zeta}_k^{(m)}(t) = \zeta_i^{(m)}(t) &= \sum_{\underline{j} \in [N]^m} w_{i,\underline{j}}^{(m)} z_{\underline{j}}^{(m)}(t) = \sum_{\underline{l}^{(m)}} \tilde{w}_{k, \underline{l}^{(m)}}^{(m)} \sum_{j_1 \in V_{l_1^{(m)}}^{(m)}} \cdots \sum_{j_m \in V_{l_m^{(m)}}^{(m)}} z_{\underline{j}}^{(m)}(t) \\
&= \sum_{\underline{l}^{(m)}} \underbrace{N_{\underline{l}^{(m)}} \tilde{w}_{k, \underline{l}^{(m)}}^{(m)}}_{:= \bar{w}_{k^{(m)}, \underline{j}^{(m)}}^{(m)}} \mathbb{E}\left(\prod_{r=1}^{m} z_{\iota_r}(t) \,\middle|\, k^{(m)}(\iota_1) = l_1^{(m)}, \ldots, k^{(m)}(\iota_m) = l_m^{(m)}\right) \\
&= \sum_{\underline{l}^{(m)}} \bar{w}_{k^{(m)}, \underline{j}^{(m)}}^{(m)} \prod_{r=1}^{m} \mathbb{E}\left(z_\iota(t) \mid k^{(m)}(\iota) = l_r^{(m)}\right).
\end{aligned} \tag{38}
$$

This means that the ODE system for (37) is formally the same as (34) (with the appropriate definition of $\bar{z}_k(t)$ and $\bar{\zeta}_k(t)$).

Note that $\bar{\zeta}_k(t)$ can also be expressed via $\bar{z}_k(t)$ as

$$
\begin{aligned}
\mathbb{E}\left(z_\iota(t) \mid k^{(m)}(\iota) = l_r^{(m)}\right) &= \mathbb{E}\left(\mathbb{E}\left(z_\iota(t) \mid k(i) = k\right) \mid k^{(m)}(\iota) = l_r^{(m)}\right) = \\
&\quad \mathbb{E}\left(\bar{z}_{k(\iota)}(t) \,\middle|\, k^{(m)}(\iota) = l_r^{(m)}\right),
\end{aligned}
$$

making (34) a closed system.

In the special case when the hypergraph is $(M+1)$-uniform, we can set $K^{(m)} = 1$ for all $m < M$ virtually making the local group $k^{(M)}$ and the global group k the same (apart from some 1's in the first $M-1$ components). In this case, Q only depends on $\bar{\zeta}^{(M)}(t)$ which can be expressed as

$$\bar{\zeta}_{k^{(M)}}^{(M)} = \sum_{\underline{l}^{(m)}} \bar{w}_{k^{(m)}, \underline{l}^{(m)}}^{(m)} \prod_{r=1}^{m} \bar{z}_{k^{(M)}(l_r)}(t).$$

4.3. Annealed networks

So far, we only focused on the dynamics of the Markov process neglecting the dynamics of the network itself. When there is a separation of scale between the speed of the Markov process and the changes to the network itself, two kinds of idealizations are typically used:

- *quenched networks*: the speed at which the network changes is much slower than the Markov process. In this case, the network is assumed constant in time.
- *annealed networks*: the speed at which the network changes is much faster than the Markov process. In this case, we consider the network changes averaged out for the interactions.

Annealed networks can be modeled by replacing connections $a_{i,j}^{(m)}$ in (2) and (3) with the average $\langle a \rangle_{i,j}^{(m)}$.

In this section, we present a setup for annealed networks generated via the configuration model [31]. Similar calculations can be made for other models that include e.g. degree correlation such as equation (93) in [32].

Once again, we start with the graph case.

In the configuration model the degrees $d(1), \ldots, d(N)$ are given beforehand, and vertex i receives $d(i)$ half-edges (*stubs*) initially. Then in each round, we choose two stubs at random to connect and form an edge, repeating this procedure until all stubs are paired.

Loops and multiple edges are possible, but their effect will be neglected. The expected connection between vertices i and j is

$$\langle a \rangle_{ij} = \frac{d(i)d(j)}{\bar{d}N}.$$

The degree of each vertex i indeed matches the prescribed $d(i)$ as

$$\sum_{j=1}^{N} \langle a \rangle_{ij} = \frac{d(i)}{\bar{d}} \frac{1}{N} \sum_{j=1}^{N} d(j) = d(i).$$

$\langle a \rangle_{ij}$ depends only on the degrees of i and j, so it can be interpreted as a metapopulation model where vertices are grouped according to their degree. (Note that here we also use the index $k = 0$ for isolated vertices if any.) The corresponding weights are

$$\tilde{w}_{kl} = \frac{kl}{\bar{d}^2 N},$$

for Convention 1, and

$$\tilde{w}_{kl} = \frac{l}{\bar{d}N}.$$

for Convention 2.

Let $q_k := \frac{kN_k}{\bar{d}N}$ denote the size biased degree distribution and introduce

$$\Theta(t) := \sum_{l=0}^{d_{\max}} q_l \bar{z}_l(t). \tag{39}$$

Using (33), $\bar{\zeta}_k(t)$ can be written as

$$\bar{\zeta}_k(t) = \frac{k}{\bar{d}} \Theta(t),$$

for Convention 1, and

$$\bar{\zeta}_k(t) = \Theta(t).$$

for Convention 2.

For example, the I component of the SIS process assuming Convention 1 is

$$\frac{\mathrm{d}}{\mathrm{d}t} \bar{z}_{k,I}(t) = -\gamma \bar{z}_{k,I}(t) + \frac{\beta}{\bar{d}} k \left(1 - \bar{z}_{k,I}(t) \right) \Theta_I(t),$$

$$\Theta_I(t) = \sum_{l=0}^{d_{\max}} q_l \bar{z}_{l,I}(t). \tag{40}$$

which is the Inhomogeneous Mean Field Approximation (IMFA) studied by Pastor-Satorras and Vespignani [25].

For Convention 1, to apply the results of the present paper, we need to assume upper regularity, i.e. $\delta_{\max} = \frac{d_{\max}}{\bar{d}}$ to be bounded. In many applications, the degree distribution converges to a fixed distribution, making $\bar{d}$ bounded; in such a setting, we accordingly require $d_{\max}$ to be bounded as well.

Assuming upper regularity,

$$w_{\max} = \frac{d_{\max}^2}{\bar{d}^2 N} = \frac{1}{N}\delta_{\max}^2$$

thus Theorem 5 actually provides an $O\left(\frac{1}{\sqrt{N}}\right)$ error bound.

As for Convention 2, $\delta_{\max} = 1$ holds as usual, and

$$w_{\max} = \frac{1}{N}\frac{d_{\max}}{\bar{d}}.$$

Unfortunately, one cannot relax the bound on $d_{\max}$ by using (23) instead of (22) as it requires bounds for the out-degrees:

$$\delta^{\text{out}}(j) = \sum_{i=1}^{N} w_{ij} = \sum_{k=0}^{d_{\max}} N_k \frac{d(j)}{\bar{d}N} = \frac{d(j)}{\bar{d}} \leq \frac{d_{\max}}{\bar{d}} \leq \delta_{\max}^{\text{out}}.$$

Now we turn to the hypergraph case $M > 1$. We generalize the notion of the configuration model in the following manner: For a fixed m, the m-degrees are given as $d^{(m)}(1), \ldots, d^{(m)}(N)$ and each vertex receives m-stubs based on their degree. In each round, we choose $m+1$ m-stubs at random to form an m-edge, then repeat this procedure until all of the stubs have been paired. This procedure is performed for each $1 \leq m \leq M$ independently.

For distinct $i, j_1, \ldots j_m$, the probability of connecting them in a given round is

$$\frac{d^{(m)}(i)\prod_{r=1}^{m} d^{(m)}(j_r)}{\binom{\bar{d}^{(m)}N}{m+1}} \approx \frac{(m+1)!d^{(m)}(i)\prod_{r=1}^{m} d^{(m)}(j_r)}{\left(\bar{d}^{(m)}N\right)^{m+1}}.$$

Since there are $\frac{\bar{d}^{(m)}N}{m+1}$ rounds in total, we set

$$\langle a\rangle_{i,\underline{j}}^{(m)} := \frac{m!d^{(m)}(i)\prod_{r=1}^{m} d^{(m)}(j_r)}{\left(\bar{d}^{(m)}N\right)^{m}}.$$

For the hypergraph case, we only examine Convention 1, for which

$$\tilde{w}_{k^{(m)},\underline{l}^{(m)}}^{(m)} = \frac{k^{(m)}}{\bar{d}^{(m)}}\frac{\prod_{r=1}^{m} l_r^{(m)}}{\left(\bar{d}^{(m)}N\right)^{m}}.$$

Once again, the resulting hypergraph can be interpreted as a metapopulation model, where the local groups are given according to the m-degrees of the vertices.

Clearly $\delta^{(m)}(i) = \frac{d^{(m)}(i)}{\bar{d}^{(m)}}$, so we make an upper regularity assumption in this case as well, from which $w_{\max} = O\left(\frac{1}{N}\right)$ follows.

For hypergraphs, we also need to check the condition (5).

$$\tilde{w}_{k^{(m)},\underline{l}^{(m)}}^{(m)} \leq \frac{\delta_{\max}^{m+1}}{N^m},$$

so (28) implies

$$\sum_{\substack{\underline{j}\in[N]^m \\ \text{jiss. loop}}} w_{i,\underline{j}}^{(m)} \leq C \sum_{\substack{\underline{j}\in[N]^m \\ \text{jiss. loop}}} \frac{1}{N^m} = O\left(\frac{1}{N}\right) \ll \sqrt{w_{\max}}, \tag{41}$$

hence arbitrarily small R can be used for large enough N.

The next step is to calculate $\bar{\zeta}_k(t)$ based on (34). Define

$$q_{k^{(m)}}^{(m)} := \frac{k^{(m)}N_{k^{(m)}}}{\bar{d}^{(m)}N},$$

the size-biased degree distribution of the m-vertices. Also define

$$\Theta^{(m)}(t) := \sum_{l=1}^{d_{\max}^{(m)}} q_l^{(m)} \mathbb{E}\left(z_\iota(t)|\, d^{(m)}(\iota) = l\right), \tag{42}$$

once again using the notation $\iota \sim U\left([N]\right)$.

Using (38),

$$\bar{\zeta}_k^{(m)}(t) = \sum_{\underline{l}^{(m)}} \bar{w}_{k^{(m)},\underline{j}^{(m)}}^{(m)} \prod_{r=1}^{m} \mathbb{E}\left(z_\iota(t)|\, d^{(m)}(\iota) = l_r^{(m)}\right)$$

$$= \frac{k^{(m)}}{\bar{d}^{(m)}} \sum_{\underline{l}^{(m)}} \prod_{r=1}^{m} q_{l_r}^{(m)} \mathbb{E}\left(z_\iota(t) | d^{(m)}(\iota) = l_r\right)$$

$$= \frac{k^{(m)}}{\bar{d}^{(m)}} \prod_{r=1}^{m} \sum_{l_r=1}^{d_{\max}^{(m)}} q_{l_r}^{(m)} \mathbb{E}\left(z_\iota(t) | d^{(m)}(\iota) = l_r\right)$$

$$= \frac{k^{(m)}}{\bar{d}^{(m)}} \left(\Theta^{(m)}(t)\right)^m.$$

Accordingly, e.g. the dynamics for the simplicial SIS model can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{z}_{k,I}(t) = -\gamma \bar{z}_{k,I}(t) + (1 - \bar{z}_{k,I}(t)) \sum_{m=1}^{M} \frac{\beta^{(m)}}{\bar{d}^{(m)}} \left(\Theta_I^{(m)}(t)\right)^m. \tag{43}$$

(43) was studied in [17] for the $(M+1)$-uniform case, where $\mathbb{E}\left(z_\iota(t) | d^{(M)}(\iota) = l\right)$ simplifies to $\bar{z}_k(t)$ as the global class k and the local class $k^{(M)}$ coincide.

4.4. Activity-driven networks

Activity-driven networks were introduced in [24].

Let $a_1, \ldots, a_K$ be positive numbers called *activities* and let $a(i)$ denote the activity of vertex i. Instead of a graphs structure, each vertex chooses a random vertex uniformly with rate $\beta a(i)$ and if they are an SI pair, the susceptible node becomes infected. Recoveries happen independently with rate γ.

The above model corresponds to an SIS process on the weighted graph

$$w_{ij} = \frac{a(i) + a(j)}{N}$$

since to form the (i, j) pair, either i or j needs to activate, and each vertex is chosen with probability $\frac{1}{N}$. The graph is a metapopulation model, with groups corresponding to the activity values.

We generalize this concept to allow higher order interactions. $a_1^{(m)}, \ldots, a_{K^{(m)}}^{(m)}$ are the possible m-activities and we assume that vertex i chooses m other vertices at random with rate $a^{(m)}(i)$. This results in a hypergraph with weights

$$w_{i,\underline{j}}^{(m)} = \frac{1}{N^m}\left(a_i^{(m)} + \sum_{r=1}^{m} a_{j_r}^{(m)}\right).$$

Assume the activity rates are bounded from above by some $a_{\max} < \infty$. Also, introduce

$$\bar{a}^{(m)} := \frac{1}{N}\sum_{i=1}^{N} a^{(m)}(i).$$

Then

$$\delta^{(m)}(i) = a_i^{(m)} + \frac{1}{N^m}\sum_{\underline{j}\in[N]^m}\sum_{r=1}^{m} a_{j_r}^{(m)} = a_i^{(m)} + \bar{a}^{(m)} \leq 2a_{\max}$$

so (4) is satisfied.

$w_{\max} \asymp \frac{1}{N}$ and (41) is applicable here as well satisfying (5), hence Theorem 2 applies.

$\bar{\zeta}_k(t)$ can also be expressed with the help of (38).

Proposition 2. *Let $\iota \sim U([N])$ a random index and $p_{k^{(m)}}^{(m)}$ be the ratio of vertices in the local group $k^{(m)}$. Also, define*

$$\psi^{(m)}(t) := \sum_{l=1}^{K^{(m)}} a_l^{(m)} p_l^{(m)} \mathbb{E}\left(z_\iota(t) | a^{(m)}(\iota) = l^{(m)}\right).$$

Then the neighborhood vectors have the form

$$\bar{\zeta}_k^{(m)}(t) = \left(a_{k^m}^{(m)} \mathbb{E}\left(z_\iota(t)\right) + \psi^{(m)}(t)\right) \mathbb{E}^{m-1}\left(z_\iota(t)\right).$$

The proof of Proposition 2 is given in Section 6.

For activity-driven networks, the simplicial SIS model takes the form

$$
\frac{\mathrm{d}}{\mathrm{d}t}\bar{z}_{k,I}(t) = -\gamma \bar{z}_{k,I}(t) + \left(1 - \bar{z}_{k,I}(t)\right) \cdot
$$
$$
\sum_{m=1}^{M} \beta_m \mathbb{E}^{m-1}\left(z_{\iota,I}(t)\right)\left(a_{km}^{(m)}\mathbb{E}\left(z_{\iota,I}(t)\right) + \psi_I^{(m)}(t)\right).
$$
(44)

[33] proves that (44) describes the large graph limit correctly when $M = 1$.

4.5. Dense graphs and Szemerédi's regularity lemma

We call a hypergraph dense if there is some $0 < p_0 \leq 1$ such that

$$
\bar{d}^{(m)} \geq p_0 N^m \quad \forall\, 1 \leq m \leq M.
$$
(45)

For Convention 1 graphs,

$$
\frac{1}{M!\,N} \leq w_{\max} \leq \frac{1}{p_0 N},
$$
$$
\delta_{\max} \leq \frac{1}{p_0}
$$

hold and (41) directly follows, satisfying the conditions for Theorem 2.

We focus on the graph case $M = 1$. We assume that the rate functions $q_{ss'}$ are *affine*, that is, they have the form

$$
q_{ss'}(\phi) = q_{ss'}^{(0)} + \sum_{r \in \mathcal{S}} q_{ss',r}^{(1)} \phi_r,
$$
(46)

where $q_{ss'}^{(0)}, \left(q_{ss',r}^{(1)}\right)_{r \in \mathcal{S}}$ are nonnegative constants. Many epidemiological models have this form, including the SIS process.

As it was pointed out in the preliminary work [26], Szemerédi's regularity lemma [27] provides a method to approximate (12) with a finite system up to arbitrary precision (for large enough N).

Roughly speaking, Szemerédi's regularity lemma states that any large enough dense graph can be partitioned into finitely many "boxes" (called an ε-regular partition) which have the same size (except one remainder box), and besides a few exceptional pairs the edge count between two boxes behaves as if coming from a randomly mixed graph, with error at most ε.

We denote an ε-regular partition by $V_0, V_1, \ldots, V_K$, where V_0 is the exceptional set.

$$
e(A, B) := \sum_{i \in A} \sum_{j \in B} a_{ij}
$$

refers to the number of edges between the vertex sets A, B with the convention that edges in $A \cap B$ are counted double.

We define the graph $\overline{\mathcal{G}}$ on vertices $(V_1, \ldots, V_K)$. (V_0 is neglected.)

The adjacency matrix is replaced by the edge density between $A, B \subseteq [N]$ defined as

$$
\rho(A, B) := \frac{e(A, B)}{|A| \cdot |B|}
$$
(47)

It is easy to see that $0 \leq \rho(A, B) \leq 1$.

The adjacency matrix counterpart for $\overline{\mathcal{G}}$ is simply the edge density between the $V_1, \ldots, V_K$ sets. For the average degree we further define

$$
p := \frac{\bar{d}}{N},
$$
(48)

$$
\kappa := \frac{|V_1|}{N} = \cdots = \frac{|V_K|}{N}
$$
(49)

where p is the global edge density of $\mathcal{G}$ and κ is the portion of vertices one box contains. The average degree in $\overline{\mathcal{G}}$ is $Kp \approx \frac{p}{\kappa}$, motivating the definition of the weights

$$
\bar{w}_{kl} := \frac{\kappa}{p} \rho(V_k, V_l).
$$
(50)

The corresponding solution of (12) on the graph $\overline{\mathcal{G}}$ with weights (50) is denoted by $(v_k(t))_{k=1}^{K}$ with initial condition

$$
v_k(0) = \frac{1}{|V_k|} \sum_{i \in V_k} z_i(0).
$$
(51)

Finally, we define

$$\bar{v}(t) := \sum_{k=1}^{K} \frac{|V_k|}{N} v_k(t) \tag{52}$$

and the average global density vector

$$\bar{z}(t) := \frac{1}{N} \sum_{i=1}^{N} z_i(t). \tag{53}$$

Theorem 6. $\forall T > 0, \varepsilon > 0, p_0 > 0 \exists K_{\max} \in \mathbb{Z}^+$ such that for any $\mathcal{G}$ simple graph with density parameter p_0 and $N \geq K_{max}$, there exists a partition $V_0, V_1, \ldots, V_K$ with $K \leq K_{\max}$ such that

- $|V_1| = \cdots = |V_K|$,
- $|V_0| \leq \varepsilon N$,
- $\sup_{0 \leq t \leq T} \|\bar{z}(t) - \bar{v}(t)\|_1 \leq \varepsilon$.

The proof is provided in Section 6.

Szemerédi's regularity lemma also guarantees that such a partition can be found in polynomial time [29].

We note that $K_{\max}$ may increase rapidly as $\varepsilon \to 0^+$ limiting the applicability of the approach. That said, for networks with extra community structure, this approach may still be useful.

5. Discussion

In this paper we examined the accuracy of the so called N-Intertwined Mean Field Approximation on hypergraphs. The idea of NIMFA is to assume vertices are independent from each other, then derive the dynamics of the occupation probabilities of each vertex. This leaves us with and ODE system of size $O(N)$ instead of an exponentially increasing system given by the exact Kolmogorov equations.

Our findings show that when the incoming weights are well distributed – for example, vertices typically have large degrees – then NIMFA gives an accurate approximation. Under additional assumptions we showed how the number of ODEs can be further reduced to give well-known approximation methods from the literature, such as the heterogeneous mean field approximation. Finally, we showed how Szemerédy's regularity lemma can be used to reduce the number of equations to constant order (depending only on the error desired) for large enough dense graphs.

These results have their limitations. The error bounds work poorly for truly sparse graphs (with bounded average degrees). Analyzing such systems probably requires qualitatively different approaches.

The upper regularity condition can be restrictive for certain applications. We conjecture that the results could be greatly generalized in this direction for degree distributions with fast decaying tails.

For the reduction for dense graph we applied the strong version of Szemerédy's lemma. The weak version of Szemerédy's lemma, however, has more desirable algorithmic properties and a smaller bound on the number of "boxes" one needs for a given ε. Extending the theorem in this direction might be beneficial for large, inhomogeneous, dense systems.

Finally, NIMFA has the disadvantage of requiring full knowledge of the network which is usually not possible in practice. Using metapopulation networks instead mitigates this problem, and also greatly reduces the number of equations required. This method, however, relies on the assumption that the metapopulation dynamics is close enough to the original one. Further research is needed to understand how well coarse graining performs in terms of preserving the network dynamics.

6. Proofs

6.1. General proofs

We state and prove a technical lemma first which will be used throughout other proofs.

Lemma 1. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ two sets of numbers such that $0 \leq |a_i|, |b_i| \leq 1$. Then

$$\left| \prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i \right| \leq \sum_{i=1}^{n} |a_i - b_i|.$$

Proof of Lemma 1. The proof is by induction on n. The statement is trivial for $n = 1$. For $n > 1$,

$$\left| \prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i \right| = \left| a_n \prod_{i=1}^{n-1} a_i - b_n \prod_{i=1}^{n-1} b_i \right| =$$

$$\left| (a_n - b_n) \prod_{i=1}^{n-1} a_i + b_n \left(\prod_{i=1}^{n-1} a_i - \prod_{i=1}^{n-1} b_i \right) \right| \leq$$

$$|a_n - b_n| \prod_{i=1}^{n-1} |a_i| + |b_n| \cdot \left| \prod_{i=1}^{n-1} a_i - \prod_{i=1}^{n-1} b_i \right| \leq$$

$$|a_n - b_n| + \left| \prod_{i=1}^{n-1} a_i - \prod_{i=1}^{n-1} b_i \right| \leq |a_n - b_n| + \sum_{i=1}^{n-1} |a_i - b_i| =$$

$$\sum_{i=1}^{n} |a_i - b_i| . \quad \square$$

Next we show that (12) exhibits a unique global solution.

Proof of Theorem 1. The right hand side of (12) is locally Lipschitz, so there is a unique local solution.

Instead of $q_{ss'}$, we use the modified rate functions

$$\hat{q}_{ss'}(\phi) := |q_{ss'}(\phi)| \tag{54}$$

$$\hat{q}_{ss}(\phi) = -\sum_{s' \neq s} \hat{q}_{s's}(\phi)$$

which are nonnegative for any input; note that $\hat{q}_{ss'}(\phi)\big|_{\phi \geq 0} = q_{ss'}(\phi)\big|_{\phi \geq 0}$.

The modified version of (12) is

$$\frac{\mathrm{d}}{\mathrm{d}t} \hat{z}_i(t) = \hat{Q}\left(\hat{\zeta}_i(t) \right) \hat{z}_i(t)$$

where $\hat{Q}(\phi) = \left(\hat{q}_{ss'}(\phi) \right)_{s,s' \in \mathcal{S}}$. The local solution uniquely exist in this case as well, and it either extends to a global solution or blows up at a finite time.

Assume that the local solution blows up at time t_0. Then $\hat{\zeta}_i(t)$ is well-defined for any $t < t_0$.

We construct an auxiliary time-inhomogeneous Markov process on $[0, t_0)$. The state space is $\mathcal{S}$ and the transition rates at time t are given by the matrix $\hat{Q}\left(\hat{\zeta}_i(t) \right)$. $p_s(t)$ denotes the probability of being in state $s \in \mathcal{S}$. The Kolmogorov equations have the form

$$\frac{\mathrm{d}}{\mathrm{d}t} p(t) = \hat{Q}\left(\hat{\zeta}_i(t) \right) p(t).$$

Since $\hat{Q}\left(\hat{\zeta}_i(t) \right)$ is continuous for $t < t_0$,

$$\max_{0 \leq \tau \leq t} \left\| \hat{Q}\left(\hat{\zeta}_i(\tau) \right) \right\|$$

exists and is finite.

Based on Grönwall's inequality,

$$\hat{z}_i(t) - p(t) = \hat{z}_i(0) - p(0) + \int_0^t \hat{Q}\left(\hat{\zeta}_i(u) \right) \left[\hat{z}_i(\tau) - p(\tau) \right] \mathrm{d}\tau,$$

$$\left\| \hat{z}_i(t) - p(t) \right\| = \left\| \hat{z}_i(0) - p(0) \right\| + \sup_{0 \leq u \leq t} \left\| \hat{Q}\left(\hat{\zeta}_i(\tau) \right) \right\| \int_0^t \left\| \hat{z}_i(\tau) - p(\tau) \right\| \mathrm{d}\tau,$$

$$\sup_{0 \leq \tau \leq t} \left\| \hat{z}_i(\tau) - p(\tau) \right\| \leq \left\| \hat{z}_i(0) - p(0) \right\| \exp\left(\sup_{0 \leq \tau \leq t} \left\| \hat{Q}\left(\hat{\zeta}_i(\tau) \right) \right\| \cdot t \right).$$

Choosing $p(0) = \hat{z}_i(0)$ shows that $\hat{z}_i(t) = p(t)$ for any $0 \leq t < t_0$ as well.

But $p(t)$ is a probability vector, that is, $\hat{z}_i(t) \in \Delta^{\mathcal{S}}$, which contradicts $\hat{z}_i(t)$ blowing up as $t \to t_0$, so the solution must be global.

Since the solution is on the simplex $\Delta^{\mathcal{S}}$, we have $\hat{q}_{ss'}\left(\hat{\zeta}_i(t) \right) = q_{ss'}\left(\hat{\zeta}_i(t) \right)$ (that is, the absolute values in (54) are not necessary). Therefore $\hat{z}_i(t)$ is a solution for the original Eq. (12) as well. Since the solution for (12) is unique, $\hat{z}_i(t) = z_i(t)$. This makes $z_i(t)$ a global solution with values on the simplex $\Delta^{\mathcal{S}}$. $\square$

6.2. Proof of Theorem 2

The strategy of the proof is to derive an inequality for $D_{\max}(t)$ and $\tilde{D}_i(t)$ such that Grönwall's inequality could be applied. In the first step, we are showing an inequality for the error of the indicators.

Lemma 2. *There exists $\tilde{C}_1 = \tilde{C}_1(\delta_{max})$ such that*

$$D^{(0)}_{max}(t) \le \tilde{C}_1 \int_0^t D_{max}(\tau)d\tau,$$

$$\tilde{D}^{(0)}_i(t) \le \tilde{C}_1 \int_0^t \tilde{D}_i(\tau)d\tau.$$

Proof of Lemma 2. $\oplus$ denotes symmetric difference.

$$\left| \xi_{i,s}(\tau) - \hat{\xi}_{i,s}(\tau) \right| \le$$

$$\sum_{\substack{s' \in \mathcal{S} \\ s' \ne s}} \left| \mathcal{N}_{i,ss'} \left(\mathcal{H}_{i,ss'}(\tau) \right) - \mathcal{N}_{i,ss'} \left(\mathcal{K}_{i,ss'}(\tau) \right) \right| + \left| \mathcal{N}_{i,ss'} \left(\mathcal{H}_{i,ss'}(\tau) \right) - \mathcal{N}_{i,ss'} \left(\mathcal{K}_{i,ss'}(\tau) \right) \right| \le$$

$$\sum_{\substack{s' \in \mathcal{S} \\ s' \ne s}} \mathcal{N}_{i,ss'} \left(\mathcal{H}_{i,ss'}(\tau) \oplus \mathcal{K}_{i,ss'}(\tau) \right) + \mathcal{N}_{i,s's} \left(\mathcal{H}_{i,s's}(\tau) \oplus \mathcal{K}_{i,s's}(\tau) \right) \le$$

$$\sum_{\substack{s' \in \mathcal{S} \\ s' \ne s}} \mathcal{N}_{i,ss'} \left(\mathcal{H}_{i,ss'}(t) \oplus \mathcal{K}_{i,ss'}(t) \right) + \mathcal{N}_{i,s's} \left(\mathcal{H}_{i,s's}(t) \oplus \mathcal{K}_{i,s's}(t) \right)$$

In the last step we used the fact that $\mathcal{H}_{i,ss'}(\tau) \oplus \mathcal{K}_{i,ss'}(\tau)$ is an increasing set in τ.

Since the right hand side does not depend on τ, it makes no difference whether we take $\sup_{0 \le \tau \le t}$ inside or outside of the expectation.

$$D^{(0)}_i(t) \le \tilde{D}^{(0)}_i(t) \le$$

$$\sum_{s \in \mathcal{S}} \sum_{\substack{s' \in \mathcal{S} \\ s' \ne s}} \mathbb{E} \left[\mathcal{N}_{i,ss'} \left(\mathcal{H}_{i,ss'}(t) \oplus \mathcal{K}_{i,ss'}(t) \right) + \mathcal{N}_{i,s's} \left(\mathcal{H}_{i,s's}(t) \oplus \mathcal{K}_{i,s's}(t) \right) \right]$$

The summations with respect to s and s' only contribute a constant factor $|\mathcal{S}|^2$ which will be neglected. Also, the same bound applies for $E\left[\mathcal{N}_{i,ss'} \left(\mathcal{H}_{i,ss'}(t) \oplus \mathcal{K}_{i,ss'}(t) \right) \right]$ and $E\left[\mathcal{N}_{i,s's} \left(\mathcal{H}_{i,s's}(t) \oplus \mathcal{K}_{i,s's}(t) \right) \right]$, so it is enough to keep track of only the first one, with a factor of 2.

The rate functions are Lipschitz-continuous on a compact domain due to assumption (4), so they are bounded; their maximum is denoted by q_{max}.

$$\mathbb{E} \left[\mathcal{N}_{i,ss'} \left(\mathcal{H}_{i,ss'}(t) \oplus \mathcal{K}_{i,ss'}(t) \right) \right] =$$

$$\mathbb{E} \left[\int_0^t \left| q_{ss'} \left(\phi_i(\tau) \right) \xi_{i,s'}(\tau) - q_{ss'} \left(\tilde{\phi}_i(\tau) \right) \hat{\xi}_{i,s'}(\tau) \right| d\tau \right] \le$$

$$\mathbb{E} \left[\int_0^t q_{max} \left| \xi_{i,s'}(\tau) - \hat{\xi}_{i,s'}(\tau) \right| + L_q \sum_{m=1}^M \sum_{\underline{r} \in \mathcal{S}^m} \left| \phi^{(m)}_{i,\underline{r}}(\tau) - \tilde{\phi}^{(m)}_{i,\underline{r}}(\tau) \right| d\tau \right] \le$$

$$\left(q_{max} + L_q \right) \int_0^t \sum_{m=0}^M D^{(m)}_i(\tau)d\tau \le \left(q_{max} + L_q \right) \int_0^t \sum_{m=0}^M \tilde{D}^{(m)}_i(\tau)d\tau$$

Setting $\tilde{C}_1 := 2 \left(q_{max} + L_q \right) |\mathcal{S}|^2$ yields

$$D^{(0)}_i(t) \le \tilde{D}^{(0)}_i(t) \le \tilde{C}_1 \int_0^t \sum_{m=0}^M D^{(m)}_i(\tau)d\tau \le \tilde{C}_1 \int_0^t \underbrace{\sum_{m=0}^M \tilde{D}^{(m)}_i(\tau)}_{=\tilde{D}_i(\tau)} d\tau. \quad \square$$

The second half of the proof of Theorem 2 involves estimating the difference between the neighbors $\phi_i(t)$ and $\zeta_i(t)$ via the differences of the indicators.

$\zeta_i(t)$ does not contain the indicators $\hat{\xi}_i(t)$ directly, only their expectation $z_i(t)$. To bridge this gap, we introduce "intermediate neighborhoods"

$$\hat{\phi}^{(m)}_{i,\underline{s}}(t) = \sum_{\underline{j} \in [N]^m} w^{(m)}_{i,\underline{j}} \hat{\xi}^{(m)}_{j,\underline{s}}(t).$$

Note that under (16) and independent initial conditions,

$$\mathbb{E}\left(\hat{\xi}_{\underline{i},\underline{s}}^{(m)}\right) = \mathbb{E}\left(\prod_{l=1}^{m}\hat{\xi}_{i_l,s_l}(t)\right) = \prod_{l=1}^{m}\mathbb{E}\left(\hat{\xi}_{i_l,s_l}(t)\right) = \prod_{l=1}^{m}z_{i_l,s_l}(t) = z_{\underline{i},\underline{s}}^{(m)}$$

for non-secondary loop $\underline{i}$ indices. Assumption (5) was made to ensure secondary loops have low total weight.

$$\left|\mathbb{E}\left(\hat{\phi}_{i,\underline{s}}^{(m)}(t)\right) - \zeta_{i,\underline{s}}^{(m)}(t)\right| = \left|\sum_{\underline{j}\in[N]^m} w_{i,\underline{j}}^{(m)}\left[\mathbb{E}\left(\hat{\xi}_{\underline{j},\underline{s}}^{(m)}(t)\right) - z_{\underline{j},\underline{s}}^{(m)}(t)\right]\right| =$$

$$\left|\sum_{\substack{\underline{j}\in[N]^m \\ \underline{j} \text{ s. loop}}} w_{i,\underline{j}}^{(m)}\left[\mathbb{E}\left(\hat{\xi}_{\underline{j},\underline{s}}^{(m)}(t)\right) - z_{\underline{j},\underline{s}}^{(m)}(t)\right]\right| \leq \sum_{\substack{\underline{j}\in[N]^m \\ \underline{j} \text{ s. loop}}} w_{i,\underline{j}}^{(m)} \leq R\sqrt{w_{\max}}. \tag{55}$$

The next lemma shows that $\hat{\phi}_i(t)$ and $\zeta_i(t)$ are close.

Lemma 3. *Assume* (16) *holds with independent initial conditions. Then there is a* $\tilde{C}_2 = \tilde{C}_2\left(\delta_{\max}, R\right)$ *such that for any* $1 \leq m \leq M$, $i \in [N]$

$$\sup_{0 \leq t} \mathbb{E}\left[\sum_{\underline{s}\in\mathcal{S}^m}\left|\hat{\phi}_{i,\underline{s}}^{(m)}(t) - \zeta_{i,\underline{s}}^{(m)}(t)\right|\right] \leq \tilde{C}_2\sqrt{w_{\max}}. \tag{56}$$

If we further assume $M = 1$*, there exists a* $\tilde{C}_3$ *such that for all* $t \geq 0$,

$$\mathbb{E}\left[\sup_{0 \leq t}\sum_{s\in\mathcal{S}}\left|\hat{\phi}_{i,s}(t) - \zeta_{i,s}(t)\right|\right] \leq \tilde{C}_3 \underbrace{\sqrt{\sum_{j=1}^{n}w_{ij}^2}}_{=\mu_i}. \tag{57}$$

Proof of Lemma 3. We start by applying (55).

$$\sup_{0 \leq t} \mathbb{E}\left[\sum_{\underline{s}\in\mathcal{S}^m}\left|\hat{\phi}_{i,\underline{s}}^{(m)}(t) - \zeta_{i,\underline{s}}^{(m)}(t)\right|\right] \leq$$

$$R|\mathcal{S}|^M\sqrt{w_{\max}} + \sup_{0 \leq t}\mathbb{E}\left[\sum_{\underline{s}\in\mathcal{S}^m}\left|\hat{\phi}_{i,\underline{s}}^{(m)}(t) - \mathbb{E}\left(\hat{\phi}_{i,\underline{s}}^{(m)}(t)\right)\right|\right].$$

The first term is of the desired form; we examine the second term.

$$\mathbb{E}\left[\sum_{\underline{s}\in\mathcal{S}^m}\left|\hat{\phi}_{i,\underline{s}}^{(m)}(t) - \mathbb{E}\left(\hat{\phi}_{i,\underline{s}}^{(m)}(t)\right)\right|\right] = \sum_{\underline{s}\in\mathcal{S}^m}\mathbb{E}\left(\left|\hat{\phi}_{i,\underline{s}}^{(m)}(t) - \mathbb{E}\left(\hat{\phi}_{i,\underline{s}}^{(m)}(t)\right)\right|\right) \leq$$

$$\sum_{\underline{s}\in\mathcal{S}^m}\sqrt{\mathbb{D}^2\left(\hat{\phi}_{i,\underline{s}}^{(m)}(t)\right)} = \sum_{\underline{s}\in\mathcal{S}^m}\sqrt{\sum_{\underline{j}\in[N]^m}\left(w_{i,\underline{j}}^{(m)}\right)^2\mathbb{D}^2\left(\hat{\xi}_{\underline{j},\underline{s}}^{(m)}(t)\right)} \leq$$

$$|\mathcal{S}|^M\sqrt{\sum_{\underline{j}\in[N]^m}\left(w_{i,\underline{j}}^{(m)}\right)^2} \leq |\mathcal{S}|^M\sqrt{\delta_{\max}w_{\max}}.$$

The bound is uniform in t, so it can be upgraded to $\sup_{0\leq t}$ for free, and (56) holds with $\tilde{C}_2 = \left(R + \sqrt{\delta_{\max}}\right)|\mathcal{S}|^M$.

Next we turn to (57).

$\hat{\xi}_{i,s}(t) - z_{i,s}(t)$ is a martingale, so

$$\hat{\phi}_{i,s}(t) - \zeta_{i,s}(t) = \sum_{j=1}^{N}w_{ij}\left[\hat{\xi}_{j,s}(t) - z_{j,s}(t)\right]$$

is also a martingale, and Doob's martingale inequality yields

$$\mathbb{E}\left[\sup_{0\leq\tau\leq t}\sum_{s\in\mathcal{S}}\left|\hat{\phi}_{i,s}(\tau)-\zeta_{i,s}(\tau)\right|\right]\leq\sum_{s\in\mathcal{S}}\mathbb{E}\left[\sup_{0\leq\tau\leq t}\left|\hat{\phi}_{i,s}(\tau)-\zeta_{i,s}(\tau)\right|\right]\leq$$

$$\sum_{s\in\mathcal{S}}\sqrt{\mathbb{E}\left[\sup_{0\leq\tau\leq t}\left|\hat{\phi}_{i,s}(\tau)-\zeta_{i,s}(\tau)\right|^2\right]}\leq 2\sum_{s\in\mathcal{S}}\sqrt{\mathbb{E}\left(\left|\hat{\phi}_{i,s}(t)-\zeta_{i,s}(t)\right|^2\right)}=$$

$$2\sum_{s\in\mathcal{S}}\sqrt{\mathbb{D}^2\left(\hat{\phi}_{i,s}(t)\right)}=2\sum_{s\in\mathcal{S}}\sqrt{\sum_{j=1}^{N}w_{ij}^2\mathbb{D}^2\left(\hat{\xi}_{j,s}(t)\right)}\leq\underbrace{2\left|\mathcal{S}\right|}_{=:\tilde{C}_3}\sqrt{\sum_{j=1}^{N}w_{ij}^2}.\quad\square$$

Next we show an upper bound for the differences of neighborhood vectors, which are captured by the values $D_{\max}^{(m)}(t)$.

Lemma 4. *Assume* (16) *and independent initial conditions. Then there exist constants* $\tilde{C}_4=\tilde{C}_5\left(\delta_{\max}\right)$ *such that for any* $t\geq 0$ *and* $1\leq m\leq M$

$$D_{\max}^{(m)}(t)\leq\tilde{C}_2\sqrt{w_{\max}}+\tilde{C}_4 D_{\max}^{(0)}(t).$$

where $\tilde{C}_2$ *comes from* Lemma 3.
   If we further assume $M=1$ *then*

$$\tilde{D}^{(1)}(t)\leq\tilde{C}_3\mu+W\tilde{D}^{(0)}(t).$$

where $\tilde{C}_3$ *comes from* Lemma 3.

Proof of Lemma 4. Using Lemma 3, we have

$$D_i^{(m)}(t)=\sup_{0\leq\tau\leq t}\mathbb{E}\left[\sum_{\underline{s}\in\mathcal{S}^m}\left|\phi_{i,\underline{s}}^{(m)}(\tau)-\zeta_{i,\underline{s}}^{(m)}(\tau)\right|\right]\leq$$

$$\tilde{C}_2\sqrt{w_{\max}}+\sup_{0\leq\tau\leq t}\mathbb{E}\left[\sum_{\underline{s}\in\mathcal{S}^m}\left|\phi_{i,\underline{s}}^{(m)}(\tau)-\hat{\phi}_{i,\underline{s}}^{(m)}(\tau)\right|\right]\leq$$

$$\tilde{C}_2\sqrt{w_{\max}}+\sum_{\underline{j}\in[N]^m}w_{i,\underline{j}}^{(m)}\left(\sup_{0\leq\tau\leq t}\mathbb{E}\left[\sum_{\underline{s}\in\mathcal{S}^m}\left|\xi_{\underline{j},\underline{s}}^{(m)}(\tau)-\hat{\xi}_{\underline{j},\underline{s}}^{(m)}(\tau)\right|\right]\right).$$

Lemma 1 provides

$$\left|\xi_{\underline{j},\underline{s}}^{(m)}(\tau)-\hat{\xi}_{\underline{j},\underline{s}}^{(m)}(\tau)\right|\leq\sum_{l=1}^{m}\left|\xi_{j_l,s_l}(\tau)-\hat{\xi}_{j_l,s_l}(\tau)\right|$$

$$\sup_{0\leq\tau\leq t}\mathbb{E}\left[\sum_{\underline{s}\in\mathcal{S}^m}\left|\xi_{\underline{j},\underline{s}}^{(m)}(\tau)-\hat{\xi}_{\underline{j},\underline{s}}^{(m)}(\tau)\right|\right]\leq\sup_{0\leq\tau\leq t}\mathbb{E}\left[\sum_{\underline{s}\in\mathcal{S}^m}\sum_{l=1}^{m}\left|\xi_{j_l,s_l}(\tau)-\hat{\xi}_{j_l,s_l}(\tau)\right|\right]\leq$$

$$\left|S\right|^M\sum_{l=1}^{m}\sup_{0\leq\tau\leq t}\mathbb{E}\left[\sum_{r\in\mathcal{S}}\left|\xi_{j_l,r}(\tau)-\hat{\xi}_{j_l,r}(\tau)\right|\right]\leq\left|S\right|^M\sum_{l=1}^{m}D_{j_l}^{(0)}(t)\leq M\left|\mathcal{S}\right|^M D_{\max}^{(0)}(t).$$

Putting the inequalities together yields

$$D_i^{(m)}(t)\leq\tilde{C}_2\sqrt{w_{\max}}+M\left|\mathcal{S}\right|^M D_{\max}^{(0)}(t)\underbrace{\sum_{\underline{j}\in[N]^m}w_{i,\underline{j}}^{(m)}}_{=\delta^{(m)}(i)}$$

$$D_{\max}^{(m)}(t)\leq\tilde{C}_2\sqrt{w_{\max}}+\underbrace{M\left|\mathcal{S}\right|^M\delta_{\max}}_{=:\tilde{C}_4}D_{\max}^{(0)}(t).$$

For the second part of Lemma 4, we once again use Lemma 3.

$$\tilde{D}_i^{(1)}(t)=\mathbb{E}\left[\sup_{0\leq\tau\leq t}\sum_{s\in\mathcal{S}}\left|\phi_{i,s}(\tau)-\zeta_{i,s}(\tau)\right|\right]\leq$$

$$\tilde{C}_3 \mu_i + \mathbb{E}\left[\sup_{0 \le \tau \le t} \sum_{s \in \mathcal{S}} \left|\phi_{i,s}(\tau) - \hat{\phi}_{i,s}(\tau)\right|\right] \le$$

$$\tilde{C}_3 \mu_i + \sum_{j=1}^{N} w_{ij}\left(\mathbb{E}\left[\sup_{0 \le \tau \le t} \sum_{s \in \mathcal{S}} \left|\xi_{j,s}(\tau) - \hat{\xi}_{j,s}(\tau)\right|\right]\right) = \tilde{C}_3 \mu_i + \sum_{j=1}^{N} w_{ij}\tilde{D}_j^{(0)}(t),$$

so

$$\tilde{D}^{(1)}(t) \le \tilde{C}_3 \mu + W\tilde{D}^{(0)}(t). \quad \square$$

With all the preparations done, we finally turn to proving Theorem 2.

Proof of Theorem 2. Using Lemmas 2 and 4 and Grönwall's inequality yields

$$D_{\max}(t) = D_{\max}^0(t) + \sum_{m=1}^{M} D_{\max}^{(m)}(t) \le$$

$$M\tilde{C}_2\sqrt{w_{\max}} + \left(M\tilde{C}_4 + 1\right)D_{\max}^{(0)}(t) \le$$

$$M\tilde{C}_2\sqrt{w_{\max}} + \left(M\tilde{C}_4 + 1\right)\int_0^t D_{\max}(\tau)\mathrm{d}\tau,$$

so

$$D_{\max}(t) \le \underbrace{M\tilde{C}_2 e^{\left(M\tilde{C}_4 + 1\right)t}}_{=:C}\sqrt{w_{\max}}.$$

Proving the second part is similar.

$$\tilde{D}(t) = \tilde{D}^{(0)}(t) + \sum_{m=1}^{M}\tilde{D}^{(m)}(t) \le \underbrace{M\tilde{C}_3}_{=:C_1}\mu + M\left(W + I\right)\tilde{D}^{(0)}(t) \le$$

$$C_1\mu + \underbrace{\tilde{C}_1 M}_{=:C_2}\int_0^t (W + I)\tilde{D}(\tau)\mathrm{d}\tau \Rightarrow$$

$$\left\|\tilde{D}(t)\right\| \le C_1\|\mu\| + C_2\|W + I\|\int_0^t \left\|\tilde{D}(\tau)\right\|\mathrm{d}\tau,$$

so

$$\left\|\tilde{D}(t)\right\| \le C_1 e^{C_2\|W+I\|t}\|\mu\|. \quad \square$$

6.3. Proof of Theorems 4 and 5

Proof of Theorem 4. For a fixed t and s, we apply Doob's inequality for the martingale $\frac{1}{K}\sum_{i=1}^{K}(\hat{\xi}_{i,s}(t) - z_{i,s}(t))$ and use independence to get

$$\mathbb{E}\left(\sup_{0 \le \tau \le t}\left|\frac{1}{K}\sum_{i=1}^{K}\left(z_{i,s}(\tau) - \hat{\xi}_{i,s}(\tau)\right)\right|\right) \le 2\mathbb{D}\left(\frac{1}{K}\sum_{i=1}^{K}\left(z_{i,s}(t) - \hat{\xi}_{i,s}(t)\right)\right) =$$

$$2\left(\frac{1}{K^2}\sum_{i=1}^{K}\underbrace{\mathbb{D}^2\left(z_{i,s}(t) - \hat{\xi}_{i,s}(t)\right)}_{\le 1}\right)^{1/2} \le \frac{2}{\sqrt{K}}, \tag{58}$$

and (24) follows by inserting $\sum_{s \in \mathcal{S}}$ on the left hand side at the cost of an $|\mathcal{S}|$ factor on the right hand side. The bound is uniform in t, so we can upgrade to $\sup_{0 \le t}$. $\quad \square$

Proof of Theorem 5. For (25), we consider $0 \le \tau \le t$ and use both Theorems 2 and 4:

$$\mathbb{E}\left(\sum_{s \in \mathcal{S}}\left|\frac{1}{N}\sum_{i=1}^{N}\left(\xi_{i,s}(\tau) - z_{i,s}(\tau)\right)\right|\right) \le$$

$$\sum_{s \in \mathcal{S}} \mathbb{E}\left(\left| \frac{1}{N} \sum_{i=1}^{N} \left(\xi_{i,s}(\tau) - \hat{\xi}_{i,s}(\tau) \right) \right| + \left| \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\xi}_{i,s}(\tau) - z_{i,s}(\tau) \right) \right| \right) \le$$

$$\frac{1}{N} \sum_{i=1}^{N} \underbrace{\sum_{s \in \mathcal{S}} \mathbb{E}\left| \left(\xi_{i,s}(\tau) - \hat{\xi}_{i,s}(\tau) \right) \right|}_{\le D_{\max}(t)} + \sum_{s \in \mathcal{S}} \underbrace{\mathbb{E}\left| \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\xi}_{i,s}(\tau) - z_{i,s}(\tau) \right) \right|}_{\le 2/\sqrt{N}} \le$$

$$D_{\max}(t) + \frac{2|\mathcal{S}|}{\sqrt{N}} \le C \left(\sqrt{w_{\max}} + \frac{1}{\sqrt{N}} \right).$$

The derivation of (26) is analogous to (25) with the exception of keeping the $\sup_{0 \le \tau \le t}$ inside the expectation and using (18) instead of (17).

For (27), we just note that

$$\mathbb{E}\left[\sup_{0 \le \tau \le t} \left(\sum_{s \in \mathcal{S}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\xi_{i,s}(\tau) - \hat{\xi}_{i,s}(\tau) \right) \right| \right) \right] \le \frac{1}{N} \|\tilde{D}(t)\|_1$$

$$\le \frac{1}{\sqrt{N}} \|\tilde{D}(t)\|_2 = O\left(\sqrt{\frac{1}{N} \|\mu\|_2^2} \right),$$

and the rest of the argument is essentially identical to the previous one. $\quad\square$

6.4. Proof of Proposition 2

Let $p_{k^m}^{(m)} := \frac{N_{k^{(m)}}}{N}$ denote the ratio of vertices in the local group k^m.

$$\bar{\zeta}_k^{(m)}(t) = \sum_{\underline{l}^{(m)}} \bar{w}_{k^{(m)}, \underline{j}^{(m)}}^{(m)} \prod_{r=1}^{m} \mathbb{E}\left(z_\iota(t) | k^{(m)}(\iota) = l_r^{(m)} \right)$$

$$= \sum_{\underline{l}^{(m)}} \left(\prod_{r=1}^{m} p_{l_r}^{(m)} \right) \left(a_{k^m}^{(m)} + \sum_{r=1}^{m} a_{l_r^{(m)}}^{(m)} \right) \prod_{r=1}^{m} \mathbb{E}\left(z_\iota(t) | k^{(m)}(\iota) = l_r^{(m)} \right)$$

$$= \sum_{\underline{l}^{(m)}} \left(a_{k^m}^{(m)} + \sum_{r=1}^{m} a_{l_r^{(m)}}^{(m)} \right) \prod_{r=1}^{m} p_{l_r}^{(m)} \mathbb{E}\left(z_\iota(t) | k^{(m)}(\iota) = l_r^{(m)} \right) \tag{59}$$

Observe

$$\sum_{l^m=1}^{K^{(m)}} p_{l_r}^{(m)} \mathbb{E}\left(z_\iota(t) | a^{(m)}(\iota) = l_r^{(m)} \right) =$$

$$\mathbb{E}\left(\mathbb{E}\left(z_\iota(t) | a^{(m)}(\iota) = l_r^{(m)} \right) \right) = \mathbb{E}\left(z_\iota(t) \right).$$

Also introduce

$$\psi^{(m)}(t) := \sum_{l=1}^{K^{(m)}} a_l^{(m)} p_l^{(m)} \mathbb{E}\left(z_\iota(t) | a^{(m)}(\iota) = l^{(m)} \right)$$

which is renaissance of an activity biased average.

We expand (59) based on the terms $a_{k^m}^{(m)} + \sum_{r=1}^{m} a_{l_r^{(m)}}^{(m)}$. For $a_{k^m}^{(m)}$

$$a_{k^m}^{(m)} \sum_{\underline{l}^{(m)}} \prod_{r=1}^{m} p_{l_r}^{(m)} \mathbb{E}\left(z_\iota(t) | k^{(m)}(\iota) = l_r^{(m)} \right) =$$

$$a_{k^m}^{(m)} \left(\sum_{l=1}^{K^{(m)}} p_l^{(m)} \mathbb{E}\left(z_\iota(t) | k^{(m)}(\iota) = l^{(m)} \right) \right)^m =$$

$$a_{k^m}^{(m)} \mathbb{E}^m \left(z_\iota(t) \right).$$

For the $a_{l_r^{(m)}}$ terms we have

$$\sum_{\underline{l}^{(m)}} a_{l_{r'}^{(m)}} \prod_{r=1}^{m} p_{l_r}^{(m)} \mathbb{E}\left(z_\iota(t)|\, k^{(m)}(\iota) = l_r^{(m)}\right) =$$

$$\underbrace{\sum_{l_{r'}=1}^{K^{(m)}} a_{l_{r'}^{(m)}} p_{l_{r'}}^{(m)} \mathbb{E}\left(z_\iota(t)|\, k^{(m)}(\iota) = l_{r'}^{(m)}\right)}_{\psi^{(m)}(t)} \sum_{\substack{l_r^{(m)}=1 \\ r \neq r'}}^{K^{(m)}} \prod_{\substack{r=1 \\ r \neq r'}}^{m} p_{l_r}^{(m)} \mathbb{E}\left(z_\iota(t)|\, k^{(m)}(\iota) = l_r^{(m)}\right) =$$

$$\psi^{(m)}(t) \left(\sum_{l=1}^{K^{(m)}} p_l^{(m)} \mathbb{E}\left(z_\iota(t)|\, k^{(m)}(\iota) = l^{(m)}\right) \right)^{m-1} = \psi^{(m)}(t) \mathbb{E}^{m-1}\left(z_\iota(t)\right).$$

Therefore, (59) reduces to

$$\bar{\zeta}_k^{(m)}(t) = a_{k^m}^{(m)} \mathbb{E}^m\left(z_\iota(t)\right) + \psi^{(m)}(t) \mathbb{E}^{m-1}\left(z_\iota(t)\right)$$
$$= \left(a_{k^m}^{(m)} \mathbb{E}\left(z_\iota(t)\right) + \psi^{(m)}(t) \right) \mathbb{E}^{m-1}\left(z_\iota(t)\right).$$

6.5. Proof of Theorem 6

Recall (47). We call the sets $X, Y \subset [N]$ ε-regular if for all $A \subseteq X$, $B \subseteq Y$ such that $|A| > \varepsilon |X|$, $|B| > \varepsilon |Y|$ one has

$$|\rho(A, B) - \rho(X, Y)| < \varepsilon.$$

We use Szemerédi's regularity lemma.

Lemma (*Szemerédi's Regularity Lemma*). For every $\varepsilon > 0$, $K_{\min} \in \mathbb{Z}^+$ there is a $K_{\max}$ such that if $N \geq K_{\max}$ there is a partition $V_0, V_1, \ldots, V_K$ such that

$$|V_0| < \varepsilon N,$$
$$|V_1| = \cdots = |V_K|,$$
$$K_{\min} \leq K \leq K_{\max}$$

and there are at most $\varepsilon \binom{K}{2}$ pairs of (V_k, V_l), $1 \leq k < l \leq K$ such that they are not ε-regular.

Fix a $\varepsilon' > 0$ and a $K_{\min}$ such that

$$K_{\min} > \frac{1}{\varepsilon'}.$$

This choice ensures that there are enough boxes such that most of the vertices are between boxes and not within them. This is a fairly common approach in the context of Szemerédi's regularity lemma [27].

Using Szemerédi's regularity lemma for ε', we obtain a partition denoted by $V_0, V_1, \ldots, V_K$.

For p and κ, as defined in (48) and (49), we have the following inequalities:

$$p = \frac{\bar{d}}{N} \geq \frac{p_0(N-1)}{N} \geq \frac{p_0}{2} > 0$$

$$1 = \sum_{k=0}^{K} \frac{|V_k|}{N} \geq \sum_{k=1}^{K} \frac{|V_k|}{N} = K\kappa \implies \kappa \leq \frac{1}{K} \leq \frac{1}{K_{\min}} < \varepsilon'$$

where we used $N \geq 2$.

Introduce the notations

$$\bar{z}_k(t) := \frac{1}{|V_k|} \sum_{i \in V_k} z_i(t),$$

$$\psi(t) := \sum_{k=1}^{K} \frac{|V_k|}{N} \|\bar{z}_k(t) - v_k(t)\|_1 = \kappa \sum_{k=1}^{K} \|\bar{z}_k(t) - v_k(t)\|_1.$$

If $V_0 = \emptyset$, we use the convention $z_0(t) \equiv 0$.

From (53) and (52), we have

$$\bar{z}(t) = \frac{1}{N} \sum_{i=1}^{N} z_i(t) = \sum_{k=0}^{K} \frac{|V_k|}{N} \frac{1}{|V_k|} \sum_{i \in V_k} z_i(t) = \sum_{k=0}^{K} \frac{|V_k|}{N} \bar{z}_k(t)$$

$$\|\bar{z}(t) - \bar{v}(t)\|_1 = \left\| \frac{|V_0|}{N} \bar{z}_0(t) + \sum_{k=1}^{K} \frac{|V_k|}{N} [\bar{z}_k(t) - v_k(t)] \right\|_1 \le$$

$$\frac{|V_0|}{N} \|\bar{z}_0(t)\|_1 + \sum_{k=1}^{K} \frac{|V_k|}{N} \|\bar{z}_k(t) - v_k(t)\|_1 \le \varepsilon' + \psi(t)$$

where in the last step we used $|V_0| < \varepsilon' N$ and

$$\|\bar{z}_0(t)\|_1 \le \frac{1}{|V_0|} \sum_{i \in V_0} \underbrace{\|z_i(t)\|_1}_{=1} = 1.$$

Going forward, it is enough to examine $\psi(t)$.

Next we calculate the derivative of $\bar{z}_k(t)$. As $M = 1$, (29) takes the form

$$\frac{d}{dt} z_{i,s}(t) = \sum_{s' \in \mathcal{S}} q_{ss'}(\zeta_i(t)) z_{i,s'}(t) =$$

$$\sum_{s' \in \mathcal{S}} q_{ss'}^{(0)} z_{i,s'}(t) + \sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{S}} q_{ss',r}^{(1)} \zeta_{i,r}(t) z_{i,s'}(t) =$$

$$\sum_{s' \in \mathcal{S}} q_{ss'}^{(0)} z_{i,s'}(t) + \sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{S}} q_{ss',r}^{(1)} \left[\sum_{j=1}^{N} \underbrace{\frac{a_{ij}}{\bar{d}}}_{w_{ij}^{(m)}} z_{i,s'}(t) z_{j,r}(t) \right]$$

$$\frac{d}{dt} \bar{z}_{k,s}(t) = \sum_{s' \in \mathcal{S}} q_{ss'}^{(0)} \bar{z}_{k,s'}(t) + \sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{S}} q_{ss',r}^{(1)} \left[\frac{1}{|V_k|} \sum_{i \in V_k} \sum_{j=1}^{N} \frac{a_{ij}}{\bar{d}} z_{i,s'}(t) z_{j,r}(t) \right]$$

Similarly,

$$\frac{d}{dt} v_{k,s}(t) = \underbrace{\sum_{s' \in \mathcal{S}} q_{ss'}^{(0)} v_{k,s'}(t) + \sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{S}} q_{ss',r}^{(1)} \sum_{l=1}^{K} \bar{w}_{kl} v_{k,s'}(t) v_{l,r}(t)}_{=: f_{k,s}(V(t))}$$

where $V(t) := (v_{k,s}(t))_{k \in [K], s \in \mathcal{S}}$ and $\overline{Z}(t) = (\bar{z}_{k,s}(t))_{k \in [K], s \in \mathcal{S}}$ analogously.

Next we show a Lipschitz-type inequality for $f_k = (f_{k,s})_{s \in \mathcal{S}}$.

$$\left| \bar{z}_{k,s'}(t) \sum_{l=1}^{K} \bar{w}_{kl} \bar{z}_{l,r}(t) - v_{k,s'}(t) \sum_{l=1}^{K} \bar{w}_{kl} v_{l,r}(t) \right| \le$$

$$\left| \bar{z}_{k,s'}(t) - v_{k,s'}(t) \right| \underbrace{\sum_{l=1}^{K} \bar{w}_{kl} \bar{z}_{l,r}(t)}_{\le \sum_{k=1}^{K} \bar{w}_{kl} \le \frac{2}{p_0 K} K = \frac{2}{p_0}} + \underbrace{v_{k,s'}(t)}_{\le 1} \sum_{l=1}^{K} \underbrace{\bar{w}_{kl}}_{\le \frac{2\kappa}{p_0}} \left| \bar{z}_{l,r}(t) - v_{l,r}(t) \right| \le$$

$$\frac{2}{p_0} \left(\left| \bar{z}_{k,s'}(t) - v_{k,s'}(t) \right| + \kappa \sum_{l=1}^{K} \left| \bar{z}_{l,r}(t) - v_{l,r}(t) \right| \right),$$

so

$$\left| f_{k,s}(\overline{Z}(t)) - f_{k,s}(V(t)) \right| \le q_{\max} \sum_{s' \in \mathcal{S}} \left| \bar{z}_{k,s'}(t) - v_{k,s'}(t) \right| +$$

$$\frac{2q_{\max}}{p_0} \sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{S}} \left(\left| \bar{z}_{k,s'}(t) - v_{k,s'}(t) \right| + \kappa \sum_{l=1}^{K} \left| \bar{z}_{l,r}(t) - v_{l,r}(t) \right| \right) =$$

$$q_{\max} \left(1 + \frac{2|\mathcal{S}|}{p_0} \right) \|\bar{z}_k(t) - v_k(t)\|_1 + \frac{2q_{\max} |\mathcal{S}|}{p_0} \psi(t).$$

Summation for $s \in \mathcal{S}$ results only in an extra $\mathcal{S}$ factor, so there exists a constant L_f such that

$$\left\| f_k(\overline{Z}(t)) - f_k(V(t)) \right\|_1 \le L_f \left(\|\bar{z}_k(t) - v_k(t)\|_1 + \psi(t) \right). \tag{60}$$

Next we look to replace the right hand side of $\frac{\mathrm{d}}{\mathrm{d}t}\bar{z}_{k,s}(t)$ with $f_{k,s}\left(\overline{Z}(t)\right)$. The corresponding error term is

$$g_{k,s}(t) := \sum_{s'\in\mathcal{S}}\sum_{r\in\mathcal{S}} q^{(1)}_{ss',r}\left[\frac{1}{|V_k|}\sum_{i\in V_k}\sum_{j=1}^N \frac{a_{ij}}{\bar{d}}z_{i,s'}(t)z_{j,r}(t) - \sum_{l=1}^K \bar{w}_{kl}\bar{z}_{k,s'}(t)\bar{z}_{l,r}(t)\right], \tag{61}$$

and from $\frac{\mathrm{d}}{\mathrm{d}t}\bar{z}_k(t) = g_k(t) + f_k\left(\overline{Z}(t)\right)$, we have

$$\bar{z}_k(t) = \bar{z}_k(0) + \int_0^t g_k(\tau)\mathrm{d}\tau + \int_0^t f_k\left(\overline{Z}(\tau)\right)\mathrm{d}\tau.$$

Using $\bar{z}_k(0) = v_k(0)$, $\psi(t)$ can be bounded from above by

$$\psi(t) = \kappa\sum_{k=1}^K \|\bar{z}_k(t) - v_k(t)\|_1 \le$$

$$t\cdot\sup_{0\le\tau\le t}\kappa\sum_{k=1}^K \|g_k(\tau)\|_1 + \int_0^t \kappa\sum_{k=1}^K \left\|f_k\left(\overline{Z}(\tau)\right) - f_k\left(V(\tau)\right)\right\|_1\mathrm{d}\tau \le$$

$$t\cdot\sup_{0\le\tau\le t}\kappa\sum_{k=1}^K \|g_k(\tau)\|_1 + L_f\int_0^t \kappa\sum_{k=1}^K \left(\|\bar{z}_k(\tau) - v_k(\tau)\|_1 + \psi(\tau)\right)\mathrm{d}\tau \le$$

$$t\cdot\sup_{0\le\tau\le t}\kappa\sum_{k=1}^K \|g_k(\tau)\|_1 + 2L_f\int_0^t \psi(\tau)\mathrm{d}\tau,$$

so from Grönwall's inequality,

$$\sup_{0\le t\le T}\psi(t) \le \left(T\cdot\sup_{0\le t\le T}\kappa\sum_{k=1}^K \|g_k(t)\|_1\right)e^{2L_f T}.$$

Therefore it is enough to show that $\sup_{0\le t\le T}\kappa\sum_{k=1}^K \|g_k(t)\|_1 = O\left(\varepsilon'\right)$, and with an appropriate choice of $\varepsilon = C\varepsilon'$ we can conclude

$$\sup_{0\le t\le T}\|\bar{z}(t) - \bar{v}(t)\|_1 \le \varepsilon.$$

$$\kappa\sum_{l=1}^K \|g_k(t)\|_1 =$$

$$\kappa\sum_{s\in\mathcal{S}}\sum_{k=1}^K\left|\sum_{s',r\in\mathcal{S}} q^{(1)}_{ss',r}\left[\frac{1}{|V_k|}\sum_{i\in V_k}\sum_{j=1}^N \frac{a_{ij}}{\bar{d}}z_{i,s'}(t)z_{j,r}(t) - \sum_{l=1}^K \bar{w}_{kl}\bar{z}_{k,s'}(t)\bar{z}_{l,r}(t)\right]\right| \le$$

$$\kappa q_{\max}\sum_{s,s',r\in\mathcal{S}}\sum_{k=1}^K\sum_{l=0}^K\left|\frac{1}{|V_k|}\sum_{i\in V_k}\sum_{j\in V_l}\frac{a_{ij}}{\bar{d}}z_{i,s'}(t)z_{j,r}(t) - \bar{w}_{kl}\bar{z}_{k,s'}(t)\bar{z}_{l,r}(t)\right|$$

$\sum_{s,s',r\in\mathcal{S}}(\dots)$ only contributes a factor of $|\mathcal{S}|^3$ which we can include in the constant factor along with $q_{\max}$. The remaining terms are

$$\kappa\sum_{k=1}^K\sum_{l=0}^K\left|\frac{1}{|V_k|}\sum_{i\in V_k}\sum_{j\in V_l}\frac{a_{ij}}{\bar{d}}z_{i,s'}(t)z_{j,r}(t) - \bar{w}_{kl}\bar{z}_{k,s'}(t)\bar{z}_{l,r}(t)\right|. \tag{62}$$

In the next step we shall get rid of the diagonal (k, l) terms and also the terms with $l = 0$. We have

$$\frac{1}{|V_k|}\sum_{i\in V_k}\sum_{j\in V_l}\frac{a_{ij}}{\bar{d}}z_{i,s'}(t)z_{j,r}(t) \le \frac{1}{|V_k|\bar{d}}\sum_{i\in V_k}\sum_{j\in V_k}1 = \frac{|V_l|}{\bar{d}} \le \frac{\varepsilon'}{p} \le \frac{2\varepsilon'}{p_0},$$

$$\bar{w}_{kl}\bar{z}_{k,s'}(t)\bar{z}_{l,r}(t) \le \frac{\kappa}{p} \le \frac{2\varepsilon'}{p_0},$$

so each term in the sum of (62) is $O(\varepsilon')$. There are $O(K)$ pairs which are either diagonal or $l = 0$, so their overall contribution to the sum is $O(\kappa K \varepsilon') = O(\varepsilon')$, hence we can neglect them and what we are left with is

$$\kappa \sum_{(k,l) \in \mathcal{I}} \left| \frac{1}{|V_k|} \sum_{i \in V_k} \sum_{j \in V_l} \frac{a_{ij}}{\bar{d}} z_{i,s'}(t) z_{j,r}(t) - \bar{w}_{kl} \bar{z}_{k,s'}(t) \bar{z}_{l,r}(t) \right|. \tag{63}$$

where $\mathcal{I} = \{(k,l) | k, l \in [K], k \neq l\}$.

In order to have an upper bound for (63) we want to use the properties of the ε'-regular partition. However, Szemerédi's regularity lemma uses subsets of $[N]$, or in other words, $0 - 1$ valued indicators of vertices compared to $z_{i,s}(t)$ which may take any value from $[0, 1]$.

To account for this problem, we introduce N independent homogeneous Markov processes taking values from $\mathcal{S}$. Each process makes Markov transitions according to the transition rate matrix $Q(\zeta_i(t))$ and its initial distribution is given by $(z_{i,s}(0))_{s \in \mathcal{S}}$. Let $\eta_{i,s}(t)$ be an indicator of the i'th such process is at state s at time t. We also apply the notations

$$\eta_i(t) = (\eta_{i,s}(t))_{s \in \mathcal{S}},$$

$$\bar{\eta}_k(t) := \frac{1}{|V_k|} \sum_{i \in V_k} \eta_i(t).$$

It is easy to see that $\mathbb{E}(\eta_i(t)) = z_i(t)$. Also, since $i \in V_k$ and $j \in V_l$, i and j are different for $k \neq l$, hence the corresponding processes are independent, so

$$z_{i,s'}(t) z_{j,r}(t) = \mathbb{E}(\eta_{i,s'}(t) \eta_{j,k}(t)),$$

$$\bar{z}_{k,s'}(t) \bar{z}_{l,r}(t) = \mathbb{E}(\bar{\eta}_{k,s'}(t) \bar{\eta}_{l,r}(t)).$$

Therefore, (63) can be bounded from above by

$$\mathbb{E}\left[\kappa \sum_{(k,l) \in \mathcal{I}} \left| \frac{1}{|V_k|} \sum_{i \in V_k} \sum_{j \in V_l} \frac{a_{ij}}{\bar{d}} \eta_{i,s'}(t) \eta_{j,r}(t) - \bar{w}_{kl} \bar{\eta}_{k,s'}(t) \bar{\eta}_{l,r}(t) \right| \right]. \tag{64}$$

The upper bound we aim to obtain does not depend on the artificial randomness just introduced, hence the expectation is ignored.

We make some algebraic manipulation to end up with edge densities needed for Szemerédi's regularity lemma. We use the notation

$$V_{k,s}(t) := \{ i \in V_k | \eta_{i,s}(t) = 1 \}.$$

Then

$$\frac{1}{|V_k|} \sum_{i \in V_k} \sum_{j \in V_l} \frac{a_{ij}}{\bar{d}} \eta_{i,s'}(t) \eta_{j,r}(t) = \frac{1}{|V_k| \bar{d}} e\left(V_{k,s'}(t), V_{l,r}(t) \right) =$$

$$\frac{|V_l|}{\bar{d}} \rho\left(V_{k,s'}(t), V_{l,r}(t) \right) \frac{|V_{k,s'}(t)|}{|V_k|} \frac{|V_{l,r}(t)|}{|V_l|} = \frac{\kappa}{p} \rho\left(V_{k,s'}(t), V_{l,r}(t) \right) \bar{\eta}_{k,s'}(t) \bar{\eta}_{l,k}(t).$$

By recalling (50), the inside of (64) can be rewritten as

$$\frac{\kappa^2}{p} \sum_{(k,l) \in \mathcal{I}} \left| \rho\left(V_{k,s'}(t), V_{l,r}(t) \right) - \rho\left(V_k, V_l \right) \right| \bar{\eta}_{k,s'}(t) \bar{\eta}_{l,r}(t). \tag{65}$$

Note that the summands of (65) are $O(1)$.

Using Szemerédi's lemma to (65) is relatively straightforward from now on. We still have to deal with non-ε'-regular k, l pairs, and pairs where either $|V_{k,s'}(t)| \leq \varepsilon' |V_k|$ or $|V_{l,r}(t)| \leq \varepsilon' |V_l|$. The former set of pairs are denoted by $\mathcal{I}_1$ and the latter by $\mathcal{I}_2$, and $\mathcal{I}_3 := \mathcal{I} \setminus (\mathcal{I}_1 \cup \mathcal{I}_2)$ denotes the non-problematic pairs.

Then from $|\mathcal{I}_1| \leq \varepsilon' \binom{K}{2} \leq \varepsilon' K^2$ we have

$$\frac{\kappa^2}{p} \sum_{(k,l) \in \mathcal{I}_1} \left| \rho\left(V_{k,s'}(t), V_{l,r}(t) \right) - \rho\left(V_k, V_l \right) \right| \bar{\eta}_{k,s'}(t) \bar{\eta}_{l,r}(t) = O\left(\varepsilon' \kappa^2 K^2 \right) = O\left(\varepsilon' \right).$$

$(k, l) \in \mathcal{I}_2$ is equivalent with $\bar{\eta}_{k,s'}(t) \leq \varepsilon'$ or $\bar{\eta}_{l,k}(t) \leq \varepsilon'$, yielding

$$\frac{\kappa^2}{p} \sum_{(k,l) \in \mathcal{I}_2} \left| \rho\left(V_{k,s'}(t), V_{l,r}(t) \right) - \rho\left(V_k, V_l \right) \right| \bar{\eta}_{k,s'}(t) \bar{\eta}_{l,r}(t) \leq$$

$$\frac{\varepsilon' \kappa^2}{p} \sum_{(k,l) \in \mathcal{I}_2} 1 = O\left(\varepsilon' \kappa^2 K^2\right) = O\left(\varepsilon'\right).$$

Finally, $(k, l) \in \mathcal{I}_3$ gives

$$\left| \rho\left(V_{k,s'}(t), V_{l,r}(t)\right) - \rho\left(V_k, V_l\right) \right| < \varepsilon' \Rightarrow$$

$$\frac{\kappa^2}{p} \sum_{(k,l) \in \mathcal{I}_3} \left| \rho\left(V_{k,s'}(t), V_{l,r}(t)\right) - \rho\left(V_k, V_l\right) \right| \bar{\eta}_{k,s'}(t) \bar{\eta}_{l,r}(t) \leq$$

$$\frac{\varepsilon' \kappa^2}{p} \sum_{(k,l) \in \mathcal{I}_2} 1 = O\left(\varepsilon' \kappa^2 K^2\right) = O\left(\varepsilon'\right).$$

This ensures that $\sup_{0 \leq t \leq T} \kappa \sum_{k=1}^{K} \|g_k(t)\|_1 = O\left(\varepsilon'\right)$ indeed holds, concluding the proof of Theorem 6. □

Declaration of competing interest

Data availability

No data was used for the research described in the article.

References

[1] D. Bruneo, M. Scarpa, A. Bobbio, D. Cerotti, M. Gribaudo, Markovian agent modeling swarm intelligence algorithms in wireless sensor networks, Perform. Eval. (2011).
[2] R. Bakhshi, L. Cloth, W. Fokkink, B.R. Haverkort, Mean-field framework for performance evaluation of push-pull gossip protocols, Perform. Eval. 68 (2) (2011) 157–179.
[3] R.A. Hayden, I. Horváth, M. Telek, Mean field for performance models with generally-distributed timed transitions, in: G. Norman, W. Sanders (Eds.), Quantitative Evaluation of Systems, in: Lecture Notes in Computer Science, vol. 8657, Springer International Publishing, 2014, pp. 90–105.
[4] G. Caravagna, Formal Modeling and Simulation of Biological Systems with Delays (Ph.D. thesis), Università di Pisa, 2011.
[5] R. Schlicht, G. Winkler, A delay stochastic process with applications in molecular biology, J. Math. Biol. 57 (5) (2008) 613–648.
[6] T. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, J. Appl. Probab. 7 (1970) 49–58.
[7] T.G. Kurtz, Strong approximation theorems for density dependent Markov chains, Stochastic Process. Appl. 6 (3) (1978) 223–240.
[8] P. Mieghem, The N-intertwined SIS epidemic network model, Computing 93 (2011) 147–169, http://dx.doi.org/10.1007/s00607-011-0155-y.
[9] C. Li, R. van de Bovenkamp, P. Van Mieghem, Susceptible-infected-susceptible model: A comparison of N-intertwined and heterogeneous mean-field approximations, Phys. Rev. E 86 (2012) 026116.
[10] D.H. Silva, S.C. Ferreira, W. Cota, R. Pastor-Satorras, C. Castellano, Spectral properties and the accuracy of mean-field approaches for epidemics on correlated power-law networks, Phys. Rev. Res. 1 (2019) 033024.
[11] P.L. Simon, I.Z. Kiss, On bounding exact models of epidemic spread on networks, 2017, https://arxiv.org/abs/1704.01726.
[12] A. Sridhar, S. Kar, Mean-field approximation for stochastic population processes in networks under imperfect information, 2021, https://arxiv.org/abs/2101.09644.
[13] A. Sridhar, S. Kar, On the accuracy of deterministic models for viral spread on networks, 2021, https://arxiv.org/abs/2104.04913.
[14] R. Parasnis, R. Kato, A. Sakhale, M. Franceschetti, B. Touri, Usefulness of the age-structured SIR dynamics in modelling COVID-19, 2022, https://arxiv.org/abs/2203.05111.
[15] P. Cisneros-Velarde, F. Bullo, Multi-group SIS epidemics with simplicial and higher-order interactions, 2020, https://arxiv.org/abs/2005.11404.
[16] I. Iacopini, G. Petri, A. Barrat, V. Latora, Simplicial models of social contagion, Nature Commun. 10 (2019) 2485.
[17] B. Jhun, M. Jo, B. Kahng, Simplicial SIS model in scale-free uniform hypergraph, 2019, https://arxiv.org/abs/1910.00375.
[18] Á. Bodó, G.Y. Katona, P.L. Simon, SIS epidemic propagation on hypergraphs, Bull. Math. Biol. 78 (2016) 713–735.
[19] R.J. Glauber, Time-dependent statistics of the Ising model, J. Math. Phys. 4 (2) (1963) 294–307.
[20] M. Ostilli, F. Mukhamedov, Continuous- and discrete-time Glauber dynamics. First- and second-order phase transitions in mean-field Potts models, EPL (Europhys. Lett.) 101 (6) (2013) 60008.
[21] J. Noonan, R. Lambiotte, Dynamics of majority rule on hypergraphs, Phys. Rev. E 104 (2021) 024316.
[22] N. Papanikolaou, G. Vaccario, E. Hormann, R. Lambiotte, F. Schweitzer, Consensus from group interactions: An adaptive voter model on hypergraphs, Phys. Rev. E 105 (2022) 054307.
[23] B. Guerra, J. Gómez-Gardeñes, Annealed and mean-field formulations of disease dynamics on static and adaptive networks, Physical Review E, Statistical, Nonlinear, and Soft Matter Physics 82 (2010) 035101.
[24] N. Perra, B. Gonçalves, R. Pastor-Satorras, A. Vespignani, Activity driven modeling of time varying networks, Sci. Rep. 2 (2012) http://dx.doi.org/10.1038/srep00469.
[25] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 3200–3203.
[26] K. Devriendt, P. Van Mieghem, Unified mean-field framework for susceptible-infected-susceptible epidemics on networks, based on graph partitioning and the isoperimetric inequality, Phys. Rev. E 96 (2017) 052314.
[27] E. Szemeredi, Regular partitions of graphs, Proc. C.N.R.S. Int. Colloq. 260 (1975) 11.
[28] I. Kiss, J. Miller, P. Simon, Mathematics of Epidemics on Networks, Vol. 46, Springer, 2017.
[29] N. Alon, R.A. Duke, H. Lefmann, V. Rödl, R. Yuster, The algorithmic aspects of the regularity lemma, in: Proceedings, 33rd Annual Symposium on Foundations of Computer Science, 1992, pp. 473–481.

[30] S. Bonaccorsi, S. Ottaviano, D. Mugnolo, F. De Pellegrini, Epidemic outbreaks in networks with equitable or almost-equitable partitions, SIAM J. Appl. Math. 75 (2015) 2421–2443, http://dx.doi.org/10.1137/140995829.

[31] M. Molloy, B. Reed, A critical point for random graphs with a given degree sequence, Random Struct. Algorithms 6 (2–3) (1995) 161–180.

[32] G.F. de Arruda, F.A. Rodrigues, Y. Moreno, Fundamentals of spreading processes in single and multilayer complex networks, Phys. Rep. 756 (2018) 1–59.

[33] L. Zino, A. Rizzo, M. Porfiri, An analytical framework for the study of epidemic models on activity driven networks, J. Complex Netw. 5 (6) (2017) 924–952.