

Collective dynamics of opposing groups with stochastic communication

Shi Jin* and Ruiwen Shu†

July 3, 2019

Abstract

We propose models describing the collective dynamics of two *opposing* groups of individuals with *stochastic* communication. Individuals from the same group are assumed to align in a stochastic manner, while individuals from different groups are assumed to anti-align. Under reasonable assumptions, we prove the large time behavior of *separation*, in the sense that the variation inside a group is much less than the distance between the two groups. The separation phenomena are verified by numerical simulations.

1 Introduction

Collective dynamics [2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 16, 17] is an important phenomenon in social and biological sciences. It describes the dynamics of a large group of agents resulted from their communications or interactions. For example, in opinion dynamics, for a group of N people, the opinion of the i -th person at time t is modeled by a vector $\mathbf{z}_i(t) \in \mathbb{R}^d$, whose time evolution follows the first order dynamics

$$\dot{\mathbf{z}}_i = \frac{1}{N} \sum_{i' \neq i} \psi_{i,i'}(\mathbf{z}_i, \mathbf{z}_{i'}) (\mathbf{z}_{i'} - \mathbf{z}_i), \quad i = 1, \dots, N. \quad (1.1)$$

Here $\psi_{i,i'}(\mathbf{z}_i, \mathbf{z}_{i'})$ is the interaction kernel for a pair of individuals i and i' , describing how their communication affects their opinions. One usually assumes

$$\psi_{i,i'}(\mathbf{z}_i, \mathbf{z}_{i'}) = \psi(|\mathbf{z}_i - \mathbf{z}_{i'}|) \geq 0. \quad (1.2)$$

This means (1) people in the group are *indistinguishable*, i.e., the interaction law does not depend on their indices but their distances; (2) people tend to *align* their opinions with each other via communication. Under the assumption (1.2), Motsch and Tadmor [16] proved *consensus* of opinions $\lim_{t \rightarrow \infty} \max_{i,i'} |\mathbf{z}_i - \mathbf{z}_{i'}| = 0$ for lower bounded interaction kernel, and *formation of clusters* for compactly supported interaction kernel under certain assumptions on initial data.

However, there are realistic situations where the assumption (1.2) may not be true:

*School of Mathematical Sciences, Institute of Natural Sciences, MOE-LSEC and SHL-MAC, Shanghai Jiao Tong University, Shanghai 200240, China. Email: shijin-m@sjtu.edu.cn. SJ was supported by NSF grants DMS-1522184, DMS-1819012 and DMS-1107291; RNMS KI-Net and NSFC grant No. 31571071.

†University of Maryland, College Park, 4176 Campus Drive, College Park, MD 20742, USA. Email: rshu@cscamm.umd.edu. RS was supported in part by NSF grants DMS16-13911, RNMS11-07444 (KI-Net) and ONR grant N00014-1812465.

- The communication of different pairs of people may be drastically different. For example, people with similar backgrounds or within the same community are more likely to align their opinions, while people from opposing groups may tend to *anti-align* their opinions, i.e., make one's opinion farther from the others'.
- Not every pair of people are necessarily communicating with each other all the time, and their chance of communication may be time dependent. In fact, the communication among people has intrinsic feature of *stochasticity*: the best one can say is that a pair of people communicate with certain probability.

In this work, we propose an opinion dynamics model in the framework of (1.1) to capture both issues mentioned above. To model the anti-alignment between opposing groups, we consider two groups of people, with opinions $\{\mathbf{x}_i(t)\}_{i=1}^{N_1}$ and $\{\mathbf{y}_j(t)\}_{j=1}^{N_2}$, satisfying the first order model¹

$$\begin{cases} \dot{\mathbf{x}}_i = \frac{1}{N_1} \sum_{i' \neq i} \psi_{i,i'}^+(t)(\mathbf{x}_{i'} - \mathbf{x}_i) - \frac{1}{N_2} \sum_j \psi_{i,j}^-(t)(\mathbf{y}_j - \mathbf{x}_i), \\ \dot{\mathbf{y}}_j = \frac{1}{N_2} \sum_{j' \neq j} \psi_{j,j'}^+(t)(\mathbf{y}_{j'} - \mathbf{y}_j) - \frac{1}{N_1} \sum_i \psi_{j,i}^-(t)(\mathbf{x}_i - \mathbf{y}_j). \end{cases} \quad (1.3)$$

Here $\psi_{i,i'}^+ = \psi_{i',i}^+ \geq 0$ and $\psi_{j,j'}^+ = \psi_{j',j}^+ \geq 0$ describe the alignment of opinions between people in the same group, and $\psi_{i,j}^- = \psi_{j,i}^- \geq 0$ describes the anti-alignment between people in opposing groups.

We propose two scenarios to describe the stochasticity of communication. In the first scenario, we assume that whether a pair of people are communicating is random but do not change with time. To be precise, we assume $\psi_{i,i'}^+$ and $\psi_{j,j'}^+$ are constants, given randomly by the Bernoulli distribution

$$\psi_{i,i'}^+ \sim B(p) = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p \end{cases}, \quad \psi_{j,j'}^+ \sim B(p), \quad (1.4a)$$

for some fixed $0 < p < 1$, representing the rate of communication between people in the same group. We assume similarly for $\psi_{i,j}^-$:

$$\psi_{i,j}^- \sim B(q), \quad (1.4b)$$

for some fixed $0 < q < 1$, representing the rate of communication between people in opposing groups. The random variables $\psi_{i,i'}^+$, $\psi_{j,j'}^+$ and $\psi_{i,j}^-$ are assumed to be mutually independent for different indices i, j, i', j' .

In the second scenario, we assume that whether a pair of people are communicating is random, and change at time $k\tau$, $k \in \mathbb{Z}$, where $\tau > 0$ is a fixed time step. To be precise we assume $\psi_{i,i'}^+$ and $\psi_{j,j'}^+$ are functions of time, being constant on each interval $[k\tau, (k+1)\tau)$, $k = 0, 1, \dots$, given randomly by

$$\psi_{i,i'}^+|_{[k\tau, (k+1)\tau)} = \psi_{i,i'}^{+,k} \sim B(p), \quad \psi_{j,j'}^+|_{[k\tau, (k+1)\tau)} = \psi_{j,j'}^{+,k} \sim B(p), \quad (1.5a)$$

and similarly for $\psi_{i,j}^-$:

$$\psi_{i,j}^-|_{[k\tau, (k+1)\tau)} = \psi_{i,j}^{-,k} \sim B(q). \quad (1.5b)$$

¹From now on, summations in i, i' always run from 1 to N_1 , and j, j' from 1 to N_2 . $\sum_{i' \neq i}$ means summation over the index i' which is not equal to i .

In this model, after time τ , the communication functions will be resampled, though by the same distribution.

Due to the alignment effect within a group and anti-alignment effect between different groups, it is natural to expect that the opinions of different groups will reach *separation*, i.e., the disagreement of opinions between any two people from different groups is much larger than that inside the same group. In our main results (Theorems 2.2 and 2.3) we prove exactly such large time behavior of (1.3) with large probability for both scenarios, under reasonable assumptions on parameters and initial data. This forms sharp contrast with the large time behavior of *consensus* proved in [16] for global alignment interactions.

In this paper we assume that $\psi_{i,i'}^+$, $\psi_{j,j'}^+$ and $\psi_{i,j}^-$ are independent of \mathbf{x}_i and \mathbf{y}_j . This suits more the era of modern social media by which people can communicate easily regardless of the spatial distance between them. Although this assumption may seem restrictive, we believe that our main result, which is based on energy estimates, can be generalized to some cases with $\psi_{i,i'}^+$, $\psi_{j,j'}^+$ and $\psi_{i,j}^-$ depending on \mathbf{x}_i and \mathbf{y}_j . This is left as future work.

We remark that our model (1.3) is closely related to the stochastic block model (SBM) [13]. We refer to [1] and references therein for a thorough review of SBM, and some numerical experiments of a model we study in [14]. In its simplest version, SBM is a random graph, whose vertices are two groups with sizes N_1 and N_2 respectively, and the probability of the appearance of an edge depends on whether the two vertices are in the same group. Therefore the time evolution problem of its adjacency matrix is similar to (1.3) with (1.4), but instead of having anti-alignment, here the interaction between different groups is still alignment, but with a communication rate smaller than that within the same group. In fact, as mentioned in [1], this evolution problem from SBM may not have a large time behavior of separation (i.e., the sign of components of the dominating eigenvector may not distinguish the two groups), and the main goal of studying the stochastic block model is to recognize the two groups (usually called *clustering*) using other methods. Although our model (1.3) does not serve as a method for clustering, its large time behavior is interesting for its own sake, from the perspective of opinion and other collective dynamics.

Our model (1.3) with (1.5) can be viewed as a randomized version of a deterministic model, with coefficients $\psi_{i,i'}^+ = \psi_{j,j'}^+ = p$, $\psi_{i,j}^- = q$, in the same spirit as the deterministic bi-flocking model in [9] and its applications to the data clustering problem [14]. Since it is straightforward to see the separation for this deterministic model, our separation result for (1.5) provides evidence that randomizing an interacting particle system may preserve its large time behavior with large probability, and therefore suggests that RBM could be effective to capture certain large time behavior of interacting particle systems.

This paper is organized as follows: in Section 2 we introduce some basic notations and state our main results. In Section 3 we give some probabilistic descriptions of the coefficient matrices given by (1.4) and (1.5), which are essential to the proof of the main results. In Section 4 we prove our main results. In Section 5 we give some numerical results verifying our main results. The paper is concluded in Section 6.

2 Notations and main results

We first give a precise description of the *separation* of locations (i.e., opinions of people) of two groups. Define the mean locations

$$\bar{\mathbf{x}} = \frac{1}{N_1} \sum_i \mathbf{x}_i, \quad \bar{\mathbf{y}} = \frac{1}{N_2} \sum_j \mathbf{y}_j, \quad (2.1)$$

and the deviation of locations

$$\hat{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}, \quad \hat{\mathbf{y}}_j = \mathbf{y}_j - \bar{\mathbf{y}}. \quad (2.2)$$

Define the variance of locations within each group:

$$\text{var}(\mathbf{x}) = \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2, \quad \text{var}(\mathbf{y}) = \frac{1}{N_2} \sum_j |\hat{\mathbf{y}}_j|^2. \quad (2.3)$$

Then the relative size between $|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2$ and $\text{var}(\mathbf{x}) + \text{var}(\mathbf{y})$ is an indicator of group separation in the sense of L^2 : if $|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2$ is much larger than $\text{var}(\mathbf{x}) + \text{var}(\mathbf{y})$, then the two groups are well-separated in an average sense.

We also concern with the separation in the L^∞ sense, meaning that the two groups are completely separated by some hyperplane. If each $|\hat{\mathbf{x}}_i|$ and $|\hat{\mathbf{y}}_j|$ is much smaller than $|\bar{\mathbf{x}} - \bar{\mathbf{y}}|$, then the two group are completely separated by the perpendicular bisector of the segment connecting $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$.

Definition 2.1. *We say an event \mathcal{A}_N depending a large parameter N happens with large probability if*

$$\lim_{N \rightarrow \infty} P(\mathcal{A}_N) = 1. \quad (2.4)$$

We first state our separation result for the first scenario, i.e., (1.3) with (1.4).

Theorem 2.2. *Consider (1.3) with parameters given by (1.4). Assume N_1 and N_2 are comparable, in the sense that there exists constant $\kappa > 0$ such that*

$$\frac{1}{\kappa} N_1 \leq N_2 \leq \kappa N_1. \quad (2.5)$$

Then for sufficiently large $N = \min\{N_1, N_2\}$ (in terms of p, q and κ), there exist constants $\lambda_-(N), \lambda_+(N)$, with

$$\lambda_-(N) \leq \frac{C(p, q, \alpha)}{N^{1-\alpha}}, \quad \lambda_+(N) \geq c(p, q, \alpha) N^{1-\alpha}, \quad \text{for any } 0 < \alpha < 1, \quad (2.6)$$

such that, with large probability, if the initial data of (1.3) satisfies

$$\lambda(0) := \frac{\text{var}(\mathbf{x}(0)) + \text{var}(\mathbf{y}(0))}{|\bar{\mathbf{x}}(0) - \bar{\mathbf{y}}(0)|^2} \leq \lambda_+, \quad (2.7)$$

then there holds the L^2 separation

$$\lambda(t) = \frac{\text{var}(\mathbf{x}(t)) + \text{var}(\mathbf{y}(t))}{|\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2} \leq \lambda_- + (\lambda(0) - \lambda_-) e^{-\mu t}, \quad \mu = p/2, \quad (2.8)$$

and L^∞ separation

$$\tilde{\lambda}(t) := \frac{\max_i |\hat{\mathbf{x}}_i(t)|^2 + \max_j |\hat{\mathbf{y}}_j(t)|^2}{|\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2} \leq C(p, q, \kappa, \alpha) (\lambda_- + \tilde{\lambda}(0) e^{-\mu t}). \quad (2.9)$$

The meaning of this theorem is as follows: for large N , with large probability, if initially the means of the two groups are not too close (compared to their variances), then for large time the two groups separate at exponential rate, in both L^2 and L^∞ sense.

The proof of this theorem is based on L^2 estimates on $|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2$ and $\text{var}(\mathbf{x}) + \text{var}(\mathbf{y})$, using certain good properties of the coefficient matrices which hold with large probability (see Section 3). Then the L^2 separation is proved by an ODE stability argument (see Lemma 4.1). Using the L^2 result, we analyze the evolution of a single $\hat{\mathbf{x}}_i$ and obtain the L^∞ separation.

Next we state our separation result for the second scenario, i.e., (1.3) with (1.5).

Theorem 2.3. *Consider (1.3) with parameters given by (1.5), and assume (2.5). Then for any fixed $\Lambda > 0$, if $N = \min\{N_1, N_2\}$ is sufficiently large (in terms of Λ, τ, p, q and κ), then with large probability, if the initial data of (1.3) satisfies*

$$\tilde{\lambda}(0) := \frac{\max_i |\hat{\mathbf{x}}_i(0)|^2 + \max_j |\hat{\mathbf{y}}_j(0)|^2}{|\bar{\mathbf{x}}(0) - \bar{\mathbf{y}}(0)|^2} \leq \Lambda, \quad (2.10)$$

then there holds the L^∞ separation at some large time $T = T(\Lambda, \tau, p, q, \kappa)$, in the sense that there exists a constant unit vector \mathbf{v} such that

$$\max_i (\mathbf{x}_i(t) \cdot \mathbf{v}) < \min_j (\mathbf{y}_j(t) \cdot \mathbf{v}), \quad (2.11)$$

for all $t \geq T$.

This theorem gives the L^∞ separation under slightly stronger assumption (2.10) on the initial data, compared to (2.7). To prove this theorem, we basically apply Theorem 2.2 to (1.3) with (1.5) on the time interval $[0, K\tau]$ for some fixed large K (the legitimacy of this application will be justified in the proof). Then the L^∞ separation result (2.9) will imply (2.11) at time $T = K\tau$. Then a simple geometric argument shows that (2.11) will hold for all $t > T$ once it holds for $t = T$.

With the parameters given by (1.5), we do not expect quantitative separation results like (2.8) or (2.9) to hold for all large time. In fact, there is always a positive probability such that the ‘good properties’ of coefficient matrices do not hold on a long time interval $[K_1\tau, K_2\tau]$, with $(K_2 - K_1)$ being arbitrarily large. Quantitative estimates like (2.8) or (2.9) may break down on such time intervals since they originate from a bi-stable ODE argument (see Lemma 4.1).

3 Probabilistic descriptions of coefficient matrices

Before proceeding to the proof, we first need some good properties of typical coefficient matrices given randomly by Bernoulli distributions. For simplicity of notation, we will consider the case of (1.4), and clearly the same results apply to the case of (1.5) for each fixed k .

3.1 Description of $\{\psi_{i,i'}^+\}$ and $\{\psi_{j,j'}^+\}$

For the alignment coefficient matrices $\{\psi_{i,i'}^+\}$ and $\{\psi_{j,j'}^+\}$, we need to control the L^2 contraction rate of the attraction part. We define the *Fiedler number* $F_1(\psi^+)$ of $\{\psi_{i,i'}^+\}$ as the second

smallest eigenvalue of the matrix $A = (a_{i,i'})$ given as²

$$a_{i,i'} = \begin{cases} -\frac{1}{N_1}\psi_{i,i'}^+, & i \neq i', \\ \frac{1}{N_1}\sum_{i'' \neq i} \psi_{i,i''}^+, & i = i', \end{cases} \quad (3.1)$$

and similarly define $\mathbf{F}_2(\psi^+)$ for $\{\psi_{j,j'}^+\}$.

The following lemma of Juhasz [15] says that the Fiedler numbers of $\{\psi_{i,i'}^+\}$ and $\{\psi_{j,j'}^+\}$ are close to p with large probability, for large N .

Lemma 3.1 (Theorem 2 of [15]). *Let $\{\psi_{i,i'}^+\}$ and $\{\psi_{j,j'}^+\}$ be given randomly by (1.4a). For any $\delta > 0$ and $\epsilon > 0$, there exists $N_0 = N_0(p, \delta, \epsilon)$, such that for any $N = \min\{N_1, N_2\} > N_0$,*

$$P\left(|\mathbf{F}_1(\psi^+) - p| > \delta \text{ or } |\mathbf{F}_2(\psi^+) - p| > \delta\right) < \epsilon. \quad (3.2)$$

3.2 Description of $\{\psi_{i,j}^-\}$

For the anti-alignment coefficient matrix $\{\psi_{i,j}^-\}$, we first introduce the following notations: Define the 'row (column) mean' of $\{\psi_{i,j}^-\}$

$$\Psi_i^- := \frac{1}{N_2} \sum_j \psi_{i,j}^-, \quad \Psi_j^- := \frac{1}{N_1} \sum_i \psi_{i,j}^-, \quad (3.3)$$

and the overall mean

$$\bar{\Psi}^- := \frac{1}{N_1 N_2} \sum_{i,j} \psi_{i,j}^- = \frac{1}{N_1} \sum_i \Psi_i^- = \frac{1}{N_2} \sum_j \Psi_j^-. \quad (3.4)$$

Define the 'maximal row (column) deviation'

$$\hat{\Psi}_i^- = \Psi_i^- - \bar{\Psi}^-, \quad \hat{\Psi}_j^- = \Psi_j^- - \bar{\Psi}^-, \quad D(\Psi^-) = \max\{\max_i |\hat{\Psi}_i^-|, \max_j |\hat{\Psi}_j^-|\}. \quad (3.5)$$

Then we have the following lemma, which says that the row (column) mean is almost constant with large probability, for large N .

Lemma 3.2. *Assume (2.5) is satisfied for some $\kappa > 0$. Let $\{\psi_{i,j}^-\}$ be given randomly by (1.4b). For any $0 < \alpha < 1$ and $\epsilon > 0$, there exists $N_0 = N_0(q, \kappa, \alpha, \epsilon)$, such that for any $N = \min\{N_1, N_2\} > N_0$,*

$$P\left(|\Psi_i^- - q| \geq \frac{1}{N^{(1-\alpha)/2}} \text{ for some } i\right) < \epsilon. \quad (3.6)$$

This estimate is expected in view of the central limit theorem, since Ψ_i^- , as an average of $\{\psi_{i,j}^-\}_{j=1}^{N_2}$, approximately behaves like a normal distribution with mean q and variance $O(1/N_2)$, and thus a deviation larger than $O(1/\sqrt{N_2})$ should have small probability.

²Notice that 0 is the smallest eigenvalue of A with eigenvector $(1, \dots, 1)^T$.

Proof of Lemma 3.2. For a fixed i , $\{\psi_{i,j}^-\}_{j=1}^{N_2}$ are a set of N_2 i.i.d. random variables with Bernoulli distribution $B(q)$. Therefore for any $z > 0$ with zN_2 being an integer,

$$P(\Psi_i^- \geq z) \leq \sum_{n=zN_2}^{N_2} a_n, \quad a_n := \binom{N_2}{n} q^n (1-q)^{N_2-n}. \quad (3.7)$$

Notice that

$$\frac{a_{n+1}}{a_n} = \frac{(N_2 - n)q}{(n+1)(1-q)} \leq 1, \quad \text{if } n \geq qN_2, \quad (3.8)$$

therefore for $z \geq q$, using the estimate $n^n e^{-n+1} \leq n! \leq (n+1)^{n+1} e^{-n}$,

$$\begin{aligned} P(\Psi_i^- \geq z) &\leq N_2 a_{zN_2} = \frac{N_2 \cdot N_2!}{(zN_2)!((1-z)N_2)!} q^{zN_2} (1-q)^{(1-z)N_2} \\ &\leq \frac{N_2 \cdot (N_2+1)^{N_2+1} e^{-N_2}}{(zN_2)^{zN_2} e^{-zN_2+1} \cdot ((1-z)N_2)^{(1-z)N_2} e^{-(1-z)N_2+1}} q^{zN_2} (1-q)^{(1-z)N_2} \\ &= \frac{N_2(N_2+1)}{e^2} \left[\left(\frac{(N_2+1)q}{zN_2} \right)^z \left(\frac{(N_2+1)(1-q)}{(1-z)N_2} \right)^{1-z} \right]^{N_2} \\ &\leq \frac{N_2(N_2+1)}{e} \left[\left(\frac{q}{z} \right)^z \left(\frac{1-q}{1-z} \right)^{1-z} \right]^{N_2} \\ &= \frac{N_2(N_2+1)}{e} \exp \left[N_2 \left(z(\log q - \log z) + (1-z)(\log(1-q) - \log(1-z)) \right) \right]. \end{aligned} \quad (3.9)$$

Notice that by writing $z = q + \delta$ and using Taylor expansion for small δ ,

$$\begin{aligned} &z(\log q - \log z) + (1-z)(\log(1-q) - \log(1-z)) \\ &= (q + \delta) \left(\frac{-\delta}{q} + \frac{\delta^2}{2q^2} \right) + (1-q - \delta) \left(\frac{\delta}{1-q} + \frac{\delta^2}{2(1-q)^2} \right) + O_q(\delta^3) = -\frac{\delta^2}{2q(1-q)} + O_q(\delta^3). \end{aligned} \quad (3.10)$$

For $0 < \alpha < 1$ and N_2 large, we can take

$$\delta_1 = \frac{1}{N_2^{(1-\alpha)/2}}, \quad \delta = \delta_1 - \frac{1}{N_2} \left\{ (q + \delta_1)N_2 \right\} \leq \delta_1, \quad (3.11)$$

where $\{\cdot\}$ means taking the decimal part, and guarantee that zN_2 is an integer, and $\delta \geq \frac{9}{10}\delta_1$ for large N_2 . Then we conclude

$$P \left(\Psi_i^- \geq q + \frac{1}{N_2^{(1-\alpha)/2}} \right) \leq P(\Psi_i^- \geq q + \delta) \leq \frac{N_2(N_2+1)}{e} \exp \left(-\frac{N_2^\alpha}{4q(1-q)} \right), \quad (3.12)$$

for large N_2 . By similar argument for the event $\Psi_i^- \geq q - \frac{1}{N_2^{(1-\alpha)/2}}$, we finally obtain

$$P \left(\left| \Psi_i^- - q \right| \geq \frac{1}{N_2^{(1-\alpha)/2}} \right) \leq N_2^2 \exp \left(-\frac{N_2^\alpha}{4q(1-q)} \right), \quad (3.13)$$

for large N_2 .

Since Ψ_i^- are independent for different i , we have

$$\begin{aligned}
& P\left(|\Psi_i^- - q| \geq \frac{1}{N^{(1-\alpha)/2}} \text{ for some } i \in \{1, \dots, N_1\}\right) \\
&= 1 - \prod_{i=1}^{N_1} P\left(|\Psi_i^- - q| < \frac{1}{N^{(1-\alpha)/2}}\right) \\
&\leq 1 - \left(1 - N_2^2 \exp\left(-\frac{N_2^\alpha}{4q(1-q)}\right)\right)^{N_1} \\
&\leq N_1 N_2^2 \exp\left(-\frac{N_2^\alpha}{4q(1-q)}\right) \leq \kappa N_2^3 \exp\left(-\frac{N_2^\alpha}{4q(1-q)}\right),
\end{aligned} \tag{3.14}$$

and the conclusion follows since the last quantity converges to 0 as $N = \min\{N_1, N_2\} \rightarrow \infty$. \square

Remark 3.3. We can define similar notations for $\psi_{i,i'}^+$:

$$\Psi_i^+ := \frac{1}{N_1} \sum_{i' \neq i} \psi_{i,i'}^+, \quad \bar{\Psi}^+ = \frac{1}{N_1} \sum_i \Psi_i^+, \tag{3.15}$$

and

$$\hat{\Psi}_i^+ = \Psi_i^+ - \bar{\Psi}^+, \quad D(\Psi^+) = \max_i |\hat{\Psi}_i^+|, \tag{3.16}$$

and then the application of Lemma 3.2 to $\{\psi_{i,i'}^+\}$ with (1.5) gives

$$P\left(|\Psi_i^+ - p| > \frac{1}{N^{(1-\alpha)/2}} \text{ for some } i\right) < \epsilon, \tag{3.17}$$

for $N > N_0(p, \alpha, \epsilon)$.

4 Proof of Theorems 2.2 and 2.3

In this section we prove Theorems 2.2 and 2.3. We first prove a lemma which will be used in the proof of Theorem 2.2.

Lemma 4.1. Let $f(t), g(t)$ be positive functions satisfying the inequalities

$$\begin{aligned}
\dot{f} &\geq A_{11}f - A_{12}g, \\
\dot{g} &\leq A_{21}f - A_{22}g,
\end{aligned} \tag{4.1}$$

with $A_{12}, A_{21} > 0$. If the coefficients satisfy

$$\Delta := (A_{11} + A_{22})^2 - 4A_{21}A_{12} > 0, \tag{4.2}$$

and initially

$$\frac{g(0)}{f(0)} < \lambda_+, \quad \lambda_+ := \frac{(A_{11} + A_{22}) + \sqrt{\Delta}}{2A_{12}}, \tag{4.3}$$

then

$$\frac{g(t)}{f(t)} \leq \lambda_- + \frac{g(0)}{f(0)} e^{-\mu t}, \quad \lambda_- := \frac{2A_{21}}{(A_{11} + A_{22}) + \sqrt{\Delta}}, \quad \mu := A_{12}(\lambda_+ - \frac{g(0)}{f(0)}). \tag{4.4}$$

Proof. Let $\lambda(t) = g(t)/f(t)$, and it satisfies

$$\dot{\lambda} = \frac{1}{f}\dot{g} - \frac{g}{f^2}\dot{f} \leq A_{12}\lambda^2 - (A_{11} + A_{22})\lambda + A_{21} = A_{12}(\lambda - \lambda_-)(\lambda - \lambda_+). \quad (4.5)$$

If $\lambda(0) < \lambda_+$ as assumed, then $\dot{\lambda} < 0$ as long as $\lambda_- < \lambda(t) < \lambda_+$, and thus $\lambda(t)$ cannot go above $\lambda(0)$. Therefore the conclusion follows from a comparison with the linear ODE $\dot{\lambda} = -\mu(\lambda - \lambda_-)$. \square

Proof of Theorem 2.2. Fix $\alpha > 0$ small. Lemmas 3.1 and 3.2 imply that, if N is large, then with large probability, there holds

$$\begin{aligned} \mathbf{F}(\psi^+) &:= \min\{\mathbf{F}_1(\psi^+), \mathbf{F}_2(\psi^+)\} \geq p - \frac{p}{12}, \\ |\bar{\Psi}^- - q| &\leq \min\left\{\frac{q}{24}, \frac{p}{24}\right\}, \quad D(\Psi^-) \leq \min\left\{\frac{1}{N^{(1-\alpha)/2}}, \frac{p}{24}\right\}, \\ |\bar{\Psi}^+ - p| &\leq \frac{p}{24}, \quad D(\Psi^+) \leq \min\left\{\frac{1}{N^{(1-\alpha)/2}}, \frac{p}{24}\right\}. \end{aligned} \quad (4.6)$$

Therefore it suffices to prove the large time behavior (2.8) and (2.9) for coefficient matrices satisfying (4.6).

STEP 1: L^2 estimate for $|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2$.

The time evolution of $\bar{\mathbf{x}}$ is given by

$$\begin{aligned} \dot{\bar{\mathbf{x}}} &= -\frac{1}{N_1 N_2} \sum_{i,j} \psi_{i,j}^-(\mathbf{y}_j - \mathbf{x}_i) = -\frac{1}{N_2} \sum_j \Psi_j^- \mathbf{y}_j + \frac{1}{N_1} \sum_i \Psi_i^- \mathbf{x}_i \\ &= \bar{\Psi}^- (\bar{\mathbf{x}} - \bar{\mathbf{y}}) + \frac{1}{N_1} \sum_i \hat{\Psi}_i^- \hat{\mathbf{x}}_i - \frac{1}{N_2} \sum_j \hat{\Psi}_j^- \hat{\mathbf{y}}_j, \end{aligned} \quad (4.7)$$

where the symmetry of $\psi_{i,i'}^+$ is used in the first equality. Therefore by subtracting its counterpart for $\bar{\mathbf{y}}$ and conducting energy estimate,

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 &= 2\bar{\Psi}^- |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 + 2(\bar{\mathbf{x}} - \bar{\mathbf{y}}) \cdot \left(\frac{1}{N_1} \sum_i \hat{\Psi}_i^- \hat{\mathbf{x}}_i - \frac{1}{N_2} \sum_j \hat{\Psi}_j^- \hat{\mathbf{y}}_j \right) \\ &\geq (2\bar{\Psi}^- - c_1) |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 - \frac{1}{c_1} \left| \frac{1}{N_1} \sum_i \hat{\Psi}_i^- \hat{\mathbf{x}}_i - \frac{1}{N_2} \sum_j \hat{\Psi}_j^- \hat{\mathbf{y}}_j \right|^2 \\ &\geq (2\bar{\Psi}^- - c_1) |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 - \frac{2D(\Psi^-)^2}{c_1} \left(\frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 + \frac{1}{N_2} \sum_j |\hat{\mathbf{y}}_j|^2 \right) \\ &= (2\bar{\Psi}^- - c_1) |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 - \frac{2D(\Psi^-)^2}{c_1} (\text{var}(\mathbf{x}) + \text{var}(\mathbf{y})), \end{aligned} \quad (4.8)$$

where we used Cauchy-Schwarz in the first inequality, with $c_1 > 0$ to be chosen. This gives an exponential growth estimate for $|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2$, up to an error term of $(\text{var}(\mathbf{x}) + \text{var}(\mathbf{y}))$ with small coefficient.

STEP 2: L^2 estimate for $(\text{var}(\mathbf{x}) + \text{var}(\mathbf{y}))$.

Subtracting (4.7) from the first equation of (1.3) gives

$$\begin{aligned}\dot{\hat{\mathbf{x}}}_i &= \frac{1}{N_1} \sum_{i' \neq i} \psi_{i,i'}^+(\mathbf{x}_{i'} - \mathbf{x}_i) - \frac{1}{N_2} \sum_j \psi_{i,j}^-(\mathbf{y}_j - \mathbf{x}_i) - \dot{\hat{\mathbf{x}}} \\ &= \frac{1}{N_1} \sum_{i' \neq i} \psi_{i,i'}^+(\hat{\mathbf{x}}_{i'} - \hat{\mathbf{x}}_i) - \frac{1}{N_2} \sum_j \psi_{i,j}^-(\hat{\mathbf{y}}_j - \hat{\mathbf{x}}_i) - \Psi_i^-(\bar{\mathbf{y}} - \bar{\mathbf{x}}) - \dot{\hat{\mathbf{x}}}.\end{aligned}\quad (4.9)$$

Then

$$\begin{aligned}\frac{d}{dt} \left(\frac{1}{2N_1} \sum_i |\hat{\mathbf{x}}_i|^2 \right) &= \frac{1}{N_1^2} \sum_{i,i':i' \neq i} \psi_{i,i'}^+(\hat{\mathbf{x}}_{i'} - \hat{\mathbf{x}}_i) \cdot \hat{\mathbf{x}}_i - \frac{1}{N_1 N_2} \sum_{i,j} \psi_{i,j}^-(\hat{\mathbf{y}}_j - \hat{\mathbf{x}}_i) \cdot \hat{\mathbf{x}}_i \\ &\quad - \frac{1}{N_1} \sum_i \hat{\Psi}_i^-(\bar{\mathbf{y}} - \bar{\mathbf{x}}) \cdot \hat{\mathbf{x}}_i - (\bar{\Psi}^-(\bar{\mathbf{y}} - \bar{\mathbf{x}}) + \dot{\hat{\mathbf{x}}}) \cdot \frac{1}{N_1} \sum_i \hat{\mathbf{x}}_i \\ &= -\frac{1}{2N_1^2} \sum_{i,i':i' \neq i} \psi_{i,i'}^+ |\hat{\mathbf{x}}_{i'} - \hat{\mathbf{x}}_i|^2 - \frac{1}{N_1 N_2} \sum_{i,j} \psi_{i,j}^-(\hat{\mathbf{y}}_j - \hat{\mathbf{x}}_i) \cdot \hat{\mathbf{x}}_i \\ &\quad - (\bar{\mathbf{y}} - \bar{\mathbf{x}}) \cdot \frac{1}{N_1} \sum_i \hat{\Psi}_i^- \hat{\mathbf{x}}_i \\ &= -I + II + III,\end{aligned}\quad (4.10)$$

where we used $\sum_i \hat{\mathbf{x}}_i = 0$.

Notice that I can be written as

$$\begin{aligned}\frac{1}{2N_1^2} \sum_{i,i':i' \neq i} \psi_{i,i'}^+ |\hat{\mathbf{x}}_{i'} - \hat{\mathbf{x}}_i|^2 &= \frac{1}{N_1^2} \sum_{k=1}^d \left(\sum_i \left(\sum_{i' \neq i} \psi_{i,i'}^+ \right) |\hat{x}_i^{(k)}|^2 - \sum_{i,i':i' \neq i} \psi_{i,i'}^+ \hat{x}_i^{(k)} \hat{x}_{i'}^{(k)} \right) \\ &= \frac{1}{N_1} \sum_{k=1}^d (\hat{x}^{(k)})^T A \hat{x}^{(k)},\end{aligned}\quad (4.11)$$

where $\hat{x}^{(k)} = (\hat{x}_1^{(k)}, \dots, \hat{x}_{N_1}^{(k)})^T$ with $\hat{x}_i^{(k)}$ denoting the k -th component of $\hat{\mathbf{x}}_i$, and A is as given in (3.1). Thus I is bounded below by

$$\frac{1}{2N_1^2} \sum_{i,i':i' \neq i} \psi_{i,i'}^+ |\hat{\mathbf{x}}_{i'} - \hat{\mathbf{x}}_i|^2 \geq F_1(\psi^+) \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2, \quad (4.12)$$

since the vector $\hat{x}^{(k)}$ is orthogonal to the eigenvector $(1, \dots, 1)^T$ of the symmetric matrix A with the smallest eigenvalue 0.

II is controlled by

$$\begin{aligned}& -\frac{1}{N_1 N_2} \sum_{i,j} \psi_{i,j}^-(\hat{\mathbf{y}}_j - \hat{\mathbf{x}}_i) \cdot \hat{\mathbf{x}}_i \\ &= -\frac{1}{N_1 N_2} \sum_{i,j} \psi_{i,j}^- \hat{\mathbf{y}}_j \cdot \hat{\mathbf{x}}_i + \frac{1}{N_1} \sum_i \Psi_i^- |\hat{\mathbf{x}}_i|^2 \\ &\leq \frac{1}{N_1 N_2} \sum_{i,j} \psi_{i,j}^- \left(\frac{1}{2} |\hat{\mathbf{y}}_j|^2 + \frac{1}{2} |\hat{\mathbf{x}}_i|^2 \right) + (\bar{\Psi}^- + D(\Psi^-)) \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 \\ &= \frac{1}{N_2} \sum_j \Psi_j^- \frac{1}{2} |\hat{\mathbf{y}}_j|^2 + \frac{1}{N_1} \sum_i \Psi_i^- \frac{1}{2} |\hat{\mathbf{x}}_i|^2 + (\bar{\Psi}^- + D(\Psi^-)) \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 \\ &= (\bar{\Psi}^- + D(\Psi^-)) \left(\frac{3}{2} \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 + \frac{1}{2} \frac{1}{N_2} \sum_j |\hat{\mathbf{y}}_j|^2 \right).\end{aligned}\quad (4.13)$$

III is controlled by

$$-(\bar{\mathbf{y}} - \bar{\mathbf{x}}) \cdot \frac{1}{N_1} \sum_i \hat{\Psi}_i^- \hat{\mathbf{x}}_i \leq c|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 + \frac{1}{4c} \frac{1}{N_1} \sum_i |\hat{\Psi}_i^-|^2 |\hat{\mathbf{x}}_i|^2 \leq c|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 + \frac{D(\Psi^-)^2}{4c} \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2, \quad (4.14)$$

where $c > 0$ is a constant to be chosen.

Using (4.11), (4.13) and (4.14) in (4.10), we obtain

$$\begin{aligned} & \frac{d}{dt} \left(\frac{1}{2N_1} \sum_i |\hat{\mathbf{x}}_i|^2 \right) \\ & \leq -\mathbf{F}_1(\psi^+) \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 + (\bar{\Psi}^- + D(\Psi^-)) \left(\frac{3}{2} \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 + \frac{1}{2} \frac{1}{N_2} \sum_j |\hat{\mathbf{y}}_j|^2 \right) \\ & \quad + c|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 + \frac{D(\Psi^-)^2}{4c} \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 \\ & \leq -\mathbf{F}_1(\psi^+) \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 + (\bar{\Psi}^- + \frac{p}{6} + D(\Psi^-)) \left(\frac{3}{2} \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 + \frac{1}{2} \frac{1}{N_2} \sum_j |\hat{\mathbf{y}}_j|^2 \right) \\ & \quad + \frac{D(\Psi^-)^2}{p} |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2, \end{aligned} \quad (4.15)$$

by choose $c = D(\Psi^-)^2/p$. Similarly,

$$\begin{aligned} & \frac{d}{dt} \left(\frac{1}{2N_2} \sum_j |\hat{\mathbf{y}}_j|^2 \right) \\ & \leq -\mathbf{F}_2(\psi^+) \frac{1}{N_2} \sum_j |\hat{\mathbf{y}}_j|^2 + (\bar{\Psi}^- + \frac{p}{6} + D(\Psi^-)) \left(\frac{1}{2} \frac{1}{N_1} \sum_i |\hat{\mathbf{x}}_i|^2 + \frac{3}{2} \frac{1}{N_2} \sum_j |\hat{\mathbf{y}}_j|^2 \right) \\ & \quad + \frac{D(\Psi^-)^2}{p} |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2. \end{aligned} \quad (4.16)$$

Summing them together, we obtain

$$\frac{1}{2} \frac{d}{dt} (\text{var}(\mathbf{x}) + \text{var}(\mathbf{y})) \leq - \left(\mathbf{F}(\psi^+) - 2(\bar{\Psi}^- + \frac{p}{6} + D(\Psi^-)) \right) (\text{var}(\mathbf{x}) + \text{var}(\mathbf{y})) + \frac{2D(\Psi^-)^2}{p} |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2. \quad (4.17)$$

This gives an exponential decay estimate for $(\text{var}(\mathbf{x}) + \text{var}(\mathbf{y}))$, up to an error term of $|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2$ with small coefficient.

STEP 3: ODE stability analysis to get L^2 separation.

Now we claim that (4.8) and (4.17), with (4.6), imply the L^2 separation result (2.8). In fact, we apply Lemma 4.1 with

$$f = |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2, \quad g = \text{var}(\mathbf{x}) + \text{var}(\mathbf{y}), \quad (4.18)$$

and

$$A_{11} = 2(2\bar{\Psi} - c_1), \quad A_{12} = \frac{4D(\Psi^-)^2}{c_1}, \quad A_{21} = \frac{4D(\Psi^-)^2}{p}, \quad A_{22} = 2 \left(\mathbf{F}(\psi^+) - 2(\bar{\Psi}^- + \frac{p}{6} + D(\Psi^-)) \right). \quad (4.19)$$

Using (4.6), one checks that with the choice $c_1 = p/24$,

$$\begin{aligned} \frac{\Delta}{4} &= \frac{1}{4}(A_{11} + A_{22})^2 - A_{21}A_{12} = \left(\mathbb{F}(\psi^+) - c_1 - \frac{p}{3} - 2D(\Psi^-) \right)^2 - \frac{16D(\Psi^-)^3}{pc_1} \\ &\geq \left(p - \frac{p}{24} - \frac{p}{24} - \frac{p}{3} - \frac{p}{12} \right)^2 - \frac{1000}{p^2 N^{3(1-\alpha)/2}} \geq \frac{p^2}{4} - \frac{1000}{p^2 N^{3(1-\alpha)/2}}, \end{aligned} \quad (4.20)$$

which is clearly positive if N is large enough. Therefore Lemma 4.1 applies, and gives the large time behavior (2.8) under the initial assumption (2.7), with λ_{\pm} and μ given as in the lemma.

Then we estimate the asymptotic behavior of λ_{\pm} and μ for large N . In fact,

$$(A_{11} + A_{22}) + \sqrt{\Delta} \geq A_{11} + A_{22} \geq p, \quad (4.21)$$

$$A_{12}, A_{21} = O\left(\frac{1}{N^{1-\alpha}}\right), \quad (4.22)$$

for large N . This gives the L^2 asymptotic behavior (2.6). The fact that the decay rate $\mu = A_{12}(\lambda_+ - \frac{g(0)}{f(0)})$ is independent of N follows by further requiring $\frac{g(0)}{f(0)} < \lambda_+/2$ (with λ_+ as given by Lemma 4.1) and then

$$\mu \geq A_{12}\lambda_+/2 = \left((A_{11} + A_{22}) + \sqrt{\Delta} \right) / 2 \geq p/2. \quad (4.23)$$

STEP 4: L^∞ separation.

Finally we prove the L^∞ separation (2.9). We rewrite (4.9) as

$$\begin{aligned} \dot{\hat{\mathbf{x}}}_i &= \frac{1}{N_1} \sum_{i' \neq i} \psi_{i,i'}^+(\hat{\mathbf{x}}_{i'} - \hat{\mathbf{x}}_i) - \frac{1}{N_2} \sum_j \psi_{i,j}^-(\hat{\mathbf{y}}_j - \hat{\mathbf{x}}_i) - \hat{\Psi}_i^-(\bar{\mathbf{y}} - \bar{\mathbf{x}}) \\ &\quad - \frac{1}{N_1} \sum_{i'} \hat{\Psi}_{i'}^- \hat{\mathbf{x}}_{i'} + \frac{1}{N_2} \sum_j \hat{\Psi}_j^- \hat{\mathbf{y}}_j \\ &= -(\Psi_i^+ - \Psi_i^-) \hat{\mathbf{x}}_i + S, \end{aligned} \quad (4.24)$$

where the source term

$$S := \frac{1}{N_1} \sum_{i' \neq i} \psi_{i,i'}^+ \hat{\mathbf{x}}_{i'} - \frac{1}{N_2} \sum_j \psi_{i,j}^- \hat{\mathbf{y}}_j - \hat{\Psi}_i^-(\bar{\mathbf{y}} - \bar{\mathbf{x}}) - \frac{1}{N_1} \sum_{i'} \hat{\Psi}_{i'}^- \hat{\mathbf{x}}_{i'} + \frac{1}{N_2} \sum_j \hat{\Psi}_j^- \hat{\mathbf{y}}_j, \quad (4.25)$$

can be estimated by

$$\begin{aligned} |S|^2 &\leq 5 \left(\left| \frac{1}{N_1} \sum_{i' \neq i} \psi_{i,i'}^+ \hat{\mathbf{x}}_{i'} \right|^2 + \left| \frac{1}{N_2} \sum_j \psi_{i,j}^- \hat{\mathbf{y}}_j \right|^2 + \left| \hat{\Psi}_i^-(\bar{\mathbf{y}} - \bar{\mathbf{x}}) \right|^2 + \left| \frac{1}{N_1} \sum_{i'} \hat{\Psi}_{i'}^- \hat{\mathbf{x}}_{i'} \right|^2 + \left| \frac{1}{N_2} \sum_j \hat{\Psi}_j^- \hat{\mathbf{y}}_j \right|^2 \right) \\ &\leq 5(\text{var}(\mathbf{x}) + \text{var}(\mathbf{y}) + D(\Psi^-)^2 |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 + \text{var}(\mathbf{x}) + \text{var}(\mathbf{y})) \\ &\leq 10(\text{var}(\mathbf{x}) + \text{var}(\mathbf{y}) + D(\Psi^-)^2 |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2), \end{aligned} \quad (4.26)$$

by using the fact that $\psi_{i,i'}^+$ and $\psi_{i,j}^-$ are at most 1. The smallness of $|S|^2$ (compared with $|\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2$, in view of the L^2 estimates (2.8)) enables us to estimate the decay of $|\hat{\mathbf{x}}_i|^2$. In fact, multiplying

(4.24) by $\hat{\mathbf{x}}_i$ and using $\Psi_i^+ - \Psi_i^- \geq \bar{\Psi}^+ - \bar{\Psi}^- - D(\Psi^+) - D(\Psi^-)$ gives

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} |\hat{\mathbf{x}}_i|^2 &\leq -(\bar{\Psi}^+ - \bar{\Psi}^- - D(\Psi^+) - D(\Psi^-)) |\hat{\mathbf{x}}_i|^2 + |S \cdot \hat{\mathbf{x}}_i| \\
&\leq -\left(p - \frac{p}{24} - q - \frac{p}{24} - \frac{p}{12} - \frac{p}{3}\right) |\hat{\mathbf{x}}_i|^2 + \frac{3|S|^2}{4p} \\
&\leq -\left(\frac{p}{2} - q\right) |\hat{\mathbf{x}}_i|^2 + \frac{10}{p} (\text{var}(\mathbf{x}) + \text{var}(\mathbf{y}) + D(\Psi^-)^2 |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2) \\
&\leq -\left(\frac{p}{2} - q\right) |\hat{\mathbf{x}}_i|^2 + \frac{10}{p} (\lambda_- + \lambda(0)e^{-\mu t} + \frac{1}{N^{1-\alpha}}) |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 \\
&\leq -\left(\frac{p}{2} - q\right) |\hat{\mathbf{x}}_i|^2 + C(\lambda(0)e^{-\mu t} + \frac{1}{N^{1-\alpha}}) |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2,
\end{aligned} \tag{4.27}$$

by using (4.6) and (2.8), and C denotes a constant³ independent of N . By taking i as the index with maximal $|\hat{\mathbf{x}}_i|^2$ and conducting similar estimate for $|\hat{\mathbf{y}}_j|^2$, we obtain

$$\frac{1}{2} \frac{d}{dt} (\max_i |\hat{\mathbf{x}}_i|^2 + \max_j |\hat{\mathbf{y}}_j|^2) \leq -\left(\frac{p}{2} - q\right) (\max_i |\hat{\mathbf{x}}_i|^2 + \max_j |\hat{\mathbf{y}}_j|^2) + C(\lambda(0)e^{-\mu t} + \frac{1}{N^{1-\alpha}}) |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2. \tag{4.28}$$

Integrating in time gives

$$\begin{aligned}
\max_i |\hat{\mathbf{x}}_i(t)|^2 + \max_j |\hat{\mathbf{y}}_j(t)|^2 &\leq e^{-(p-2q)t} (\max_i |\hat{\mathbf{x}}_i(0)|^2 + \max_j |\hat{\mathbf{y}}_j(0)|^2) \\
&\quad + C \int_0^t e^{-(p-2q)(t-s)} (\lambda(0)e^{-\mu s} + \frac{1}{N^{1-\alpha}}) |\bar{\mathbf{x}}(s) - \bar{\mathbf{y}}(s)|^2 ds.
\end{aligned} \tag{4.29}$$

Next we notice that using (2.8) in (4.8) (with a different choice $c_1 = q/24$) gives an exponential growth estimate

$$\begin{aligned}
\frac{d}{dt} |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 &\geq (4\bar{\Psi}^- - \frac{q}{12}) |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 - \frac{96D(\Psi^-)^2}{q} (\text{var}(\mathbf{x}) + \text{var}(\mathbf{y})) \\
&\geq \left(4q - \frac{q}{6} - \frac{q}{12} - \frac{96D(\Psi^-)^2}{q} (\lambda_- + \lambda(0)e^{-\mu t})\right) |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 \\
&\geq 2q |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2,
\end{aligned} \tag{4.30}$$

if $t \geq t_1$, with t_1 being independent of N (which follows from $D(\Psi^-)^2 \leq 1/N^{1-\alpha}$ and $\lambda(0) \leq \lambda_+ = O(N^{1-\alpha})$). Therefore

$$|\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2 \geq e^{2q(t-s)} |\bar{\mathbf{x}}(s) - \bar{\mathbf{y}}(s)|^2, \quad \forall t \geq s \geq t_1. \tag{4.31}$$

Using this in (4.29), we obtain

$$\begin{aligned}
&\max_i |\hat{\mathbf{x}}_i(t)|^2 + \max_j |\hat{\mathbf{y}}_j(t)|^2 \\
&\leq C_1 e^{-(p-2q)t} + C \int_{t_1}^t e^{-(p-2q)(t-s)} (\lambda(0)e^{-\mu s} + \frac{1}{N^{1-\alpha}}) |\bar{\mathbf{x}}(s) - \bar{\mathbf{y}}(s)|^2 ds \\
&\leq C_1 e^{-(p-2q)t} + C \int_{t_1}^t e^{-(p-2q)(t-s)} (\lambda(0)e^{-\mu s} + \frac{1}{N^{1-\alpha}}) e^{-2q(t-s)} ds \cdot |\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2 \\
&\leq C_1 e^{-(p-2q)t} + C \left(\lambda(0)e^{-\mu t} + \frac{1}{N^{1-\alpha}} \right) |\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2,
\end{aligned} \tag{4.32}$$

³In the rest of this proof, different C may denote different constants which are all independent of N .

with

$$C_1 = \max_i |\hat{\mathbf{x}}_i(0)|^2 + \max_j |\hat{\mathbf{y}}_j(0)|^2 + C \int_0^{t_1} e^{(p-2q)s} (e^{-\mu s} + \frac{1}{N^{1-\alpha}}) |\bar{\mathbf{x}}(s) - \bar{\mathbf{y}}(s)|^2 ds. \quad (4.33)$$

Therefore

$$\frac{\max_i |\hat{\mathbf{x}}_i(t)|^2 + \max_j |\hat{\mathbf{y}}_j(t)|^2}{|\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2} \leq \frac{C_1}{|\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2} e^{-(p-2q)t} + C \left(\lambda(0) e^{-\mu t} + \frac{1}{N^{1-\alpha}} \right). \quad (4.34)$$

Finally we notice that the second inequality in (4.30) implies that

$$\frac{d}{dt} |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2 \geq -C_2 |\bar{\mathbf{x}} - \bar{\mathbf{y}}|^2, \quad (4.35)$$

for $0 \leq t \leq t_1$, with C_2 independent of N . Therefore, for $0 \leq s \leq t_1$,

$$|\bar{\mathbf{x}}(t_1) - \bar{\mathbf{y}}(t_1)|^2 \geq e^{-C_2(t_1-s)} |\bar{\mathbf{x}}(s) - \bar{\mathbf{y}}(s)|^2, \quad (4.36)$$

which implies

$$|\bar{\mathbf{x}}(s) - \bar{\mathbf{y}}(s)|^2 \leq e^{C_2 t_1} |\bar{\mathbf{x}}(t_1) - \bar{\mathbf{y}}(t_1)|^2 \leq e^{C_2 t_1} e^{-2q(t-t_1)} |\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2, \quad (4.37)$$

where we used (4.31) in the last inequality. It follows that

$$\begin{aligned} \frac{C_1}{|\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2} e^{2qt} &= \frac{\max_i |\hat{\mathbf{x}}_i(0)|^2 + \max_j |\hat{\mathbf{y}}_j(0)|^2}{|\bar{\mathbf{x}}(0) - \bar{\mathbf{y}}(0)|^2} \cdot \frac{|\bar{\mathbf{x}}(0) - \bar{\mathbf{y}}(0)|^2}{|\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2} e^{2qt} \\ &\quad + C \int_0^{t_1} e^{(p-2q)s} (e^{-\mu s} + \frac{1}{N^{1-\alpha}}) \frac{|\bar{\mathbf{x}}(s) - \bar{\mathbf{y}}(s)|^2}{|\bar{\mathbf{x}}(t) - \bar{\mathbf{y}}(t)|^2} ds \cdot e^{2qt} \\ &\leq e^{C_2 t_1 + 2qt_1} \tilde{\lambda}(0) + C e^{C_2 t_1 + 2qt_1 + pt_1} \leq C(1 + \tilde{\lambda}(0)). \end{aligned} \quad (4.38)$$

Therefore the first term on RHS of (4.34) behaves like $C(1 + \tilde{\lambda}(0))e^{-pt}$ which decays faster than $e^{-\mu t}$. Therefore (4.34) gives the L^∞ separation (2.9) since $\lambda(0) \leq \tilde{\lambda}(0)$. \square

Remark 4.2. *In fact, (4.6) are the only conditions on the coefficient matrices used in this proof. Therefore, the proof applies to (1.3) (not necessarily with (1.4)) on $[0, T]$, as long as the coefficient matrices satisfies (4.6) for any $t \in [0, T]$. This will be the starting point of the proof of Theorem 2.3 given below.*

Proof of Theorem 2.3. Recall that Theorem 2.3 assumes the coefficient matrices are given by the random piecewise constants (1.5). We will fix the choice of $\alpha \in (0, 1)$. Fix $K > 0$ to be chosen, it is clear that with large probability, (4.6) holds for $[0, T]$, $T = K\tau$. In fact, for any $\epsilon > 0$, for any fixed k , if $N \geq N_0(\epsilon, p, q)$, then (by Lemmas 3.1 and 3.2) with probability $1 - \epsilon$, (4.6) holds for $[k\tau, (k+1)\tau]$. By the independence of the coefficient matrices for different k ,

$$P\left(\text{(4.6) holds for } [0, K\tau]\right) \geq (1 - \epsilon)^K \rightarrow 1, \quad \text{as } \epsilon \rightarrow 0, \quad (4.39)$$

which gives (4.6) for $t \in [0, T]$ with large probability.

The assumption (2.7) on the initial data also holds for large N , since $\lambda(0) \leq \tilde{\lambda}(0) \leq \Lambda \leq \lambda_+(N)$. Therefore, by Remark 4.2, the estimate (2.9) holds for (1.3) with (1.5) with large probability, on the fixed time interval $[0, T]$. Since $\lambda_- \sim 1/N^{1-\alpha}$ and $\tilde{\lambda}(0) \leq \Lambda$, by taking

N and K large enough (in terms of $(\Lambda, \tau, p, q, \kappa)$, with K being independent of N), one can guarantee that the RHS of (2.9) at time $T = K\tau$ is no more than $1/16$. Therefore, with large probability, if N is sufficiently large, then

$$\tilde{\lambda}(T) \leq \frac{1}{16}, \quad T = K\tau, \quad K = K(\Lambda, \tau, p, q, \kappa). \quad (4.40)$$

Notice that

$$|\hat{\mathbf{x}}_i \cdot (\bar{\mathbf{y}} - \bar{\mathbf{x}})| \leq |\hat{\mathbf{x}}_i| \cdot |\bar{\mathbf{y}} - \bar{\mathbf{x}}| \leq \sqrt{\tilde{\lambda}} |\bar{\mathbf{y}} - \bar{\mathbf{x}}|^2, \quad (4.41)$$

and similar for \mathbf{y} . Thus by (4.40), for any i and j ,

$$\begin{aligned} (\mathbf{y}_j(T) - \mathbf{x}_i(T)) \cdot \frac{\bar{\mathbf{y}}(T) - \bar{\mathbf{x}}(T)}{|\bar{\mathbf{y}}(T) - \bar{\mathbf{x}}(T)|^2} &= \hat{\mathbf{y}}_j(T) \cdot \frac{\bar{\mathbf{y}}(T) - \bar{\mathbf{x}}(T)}{|\bar{\mathbf{y}}(T) - \bar{\mathbf{x}}(T)|^2} - \hat{\mathbf{x}}_i(T) \cdot \frac{\bar{\mathbf{y}}(T) - \bar{\mathbf{x}}(T)}{|\bar{\mathbf{y}}(T) - \bar{\mathbf{x}}(T)|^2} + 1 \\ &\geq 1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2} > 0. \end{aligned} \quad (4.42)$$

Therefore (2.11) holds at time T , with the vector⁴ $\mathbf{v} = \frac{\bar{\mathbf{y}}(T) - \bar{\mathbf{x}}(T)}{|\bar{\mathbf{y}}(T) - \bar{\mathbf{x}}(T)|}$.

Finally we show that (2.11) at time T implies (2.11) for all time $t \geq T$. In fact, (2.11) at time T implies the existence of a constant c such that

$$\mathbf{x}_i(T) \cdot \mathbf{v} < c, \quad \mathbf{y}_j(T) \cdot \mathbf{v} > c, \quad \forall i, j. \quad (4.43)$$

Taking dot product of (1.3) with \mathbf{v} gives

$$\begin{aligned} \dot{x}_i &= \frac{1}{N_1} \sum_{i' \neq i} \psi_{i,i'}^+(t)(x_{i'} - x_i) + \frac{1}{N_2} \sum_j \psi_{i,j}^-(t)(y_j + x_i), \\ \dot{y}_j &= \frac{1}{N_2} \sum_{j' \neq j} \psi_{j,j'}^+(t)(y_{j'} - y_j) + \frac{1}{N_1} \sum_i \psi_{j,i}^-(t)(x_i + y_j), \end{aligned} \quad (4.44)$$

where $x_i := -(\mathbf{x}_i \cdot \mathbf{v} - c)$, $y_j := \mathbf{y}_j \cdot \mathbf{v} - c$ satisfy $x_i(T) > 0$, $y_j(T) > 0$. This ODE system clearly propagates the positiveness of $\{x_i\}$ and $\{y_j\}$ if all the coefficients $\psi_{i,i'}^+$, $\psi_{i,j}^-$, $\psi_{j,j'}^+$ are nonnegative. Therefore

$$\mathbf{x}_i(t) \cdot \mathbf{v} < c, \quad \mathbf{y}_j(t) \cdot \mathbf{v} > c, \quad \forall i, j, \quad (4.45)$$

for all $t \geq T$, and (2.11) follows. □

5 Numerical results

In this section we provide a few numerical results verifying Theorems 2.2 and 2.3.

For both models ((1.3) with (1.4) and (1.3) with (1.5)), we take the spatial dimension to be one, $N_1 = N_2 =: N$, and

$$p = 0.3, \quad q = 0.2, \quad (5.1)$$

while for (1.3) with (1.5), we take the time step $\tau = 1$ for changing communication matrix. We take the initial data $\{x_i(0)\}$ and $\{y_j(0)\}$ to be i.i.d. uniformly distributed random variables in

⁴This means that at time T the two groups are separated by the perpendicular bisector of the segment connecting $\bar{\mathbf{x}}(T)$ and $\bar{\mathbf{y}}(T)$.

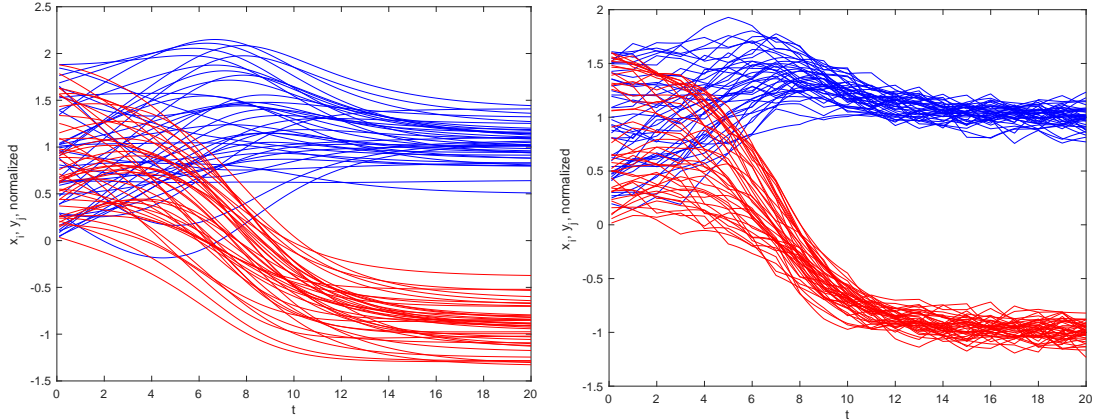


Figure 1: Simulation for (1.3), with $N_1 = N_2 = 40$. Left: with (1.4); right: with (1.5). The horizontal axis is time, and the vertical axis is $x_i(t)$ and $y_j(t)$, normalized. Red and blue curves represent the individuals in the first and second groups, respectively.

$[0, 1]$, and compute the solution at $T = 20$ to (1.3) exactly by using matrix exponentials. Figure 1 shows one simulation for each model with $N = 40$. For the sake of demonstration⁵, at each t we normalize the vector $(x_1(t), \dots, x_N(t), y_1(t), \dots, y_N(t))$ to have ℓ^2 norm equal to $\sqrt{2N}$, and this does not change the dynamics since (1.3) is linear.

One can see that for both models, the two groups completely separate from each other. Also, after getting to a certain extent of separation, $\text{var}(x)$ and $\text{var}(y)$ no longer shrink, compared to $|\bar{x} - \bar{y}|^2$. This is exactly as predicted by Theorem 2.2: the separation indicator $\lambda(t)$, as defined in (2.8), decays to some positive constant λ_- but does not necessarily converge to zero.

To further investigate the limiting behavior of $\lambda(t)$ as t getting large for (1.3) with (1.4), we compute this model with randomly sampled initial data and coefficient matrices for $n_{test} = 10000$ times, for various values of N . The sample average of $\lambda(T)$ is computed by discarding $n_{discard} = 100$ samples with largest $\lambda(T)$ which may contain the cases with small probability for which the separation claimed in Theorem 2.2 may fail. The relation of N and the sample average of $\lambda(T)$ is plotted in Figure 2. It is clear that the latter behaves like $O(1/N)$ for large N , which means that the estimate $\lambda_- \leq C/N^{1-\alpha}$, $\forall 0 < \alpha < 1$, given in (2.6), is optimal.

6 Conclusion

In this paper we propose collective dynamical models describing the interaction of two opposing groups with stochastic communication, where inter-group and intra-group communication functions have opposite signs. Under reasonable assumptions we prove the *separation* of opinions of the two groups. Compared with existing results for collective dynamics, our results suggest:

- Anti-alignment between opposing groups and alignment within the same group may lead to separation of two clusters, instead of consensus. This may happen even if the initial

⁵Without this normalization, $x_i(t)$ and $y_j(t)$ may exhibit exponential growth in time, which makes it hard to demonstrate the extent of separation.

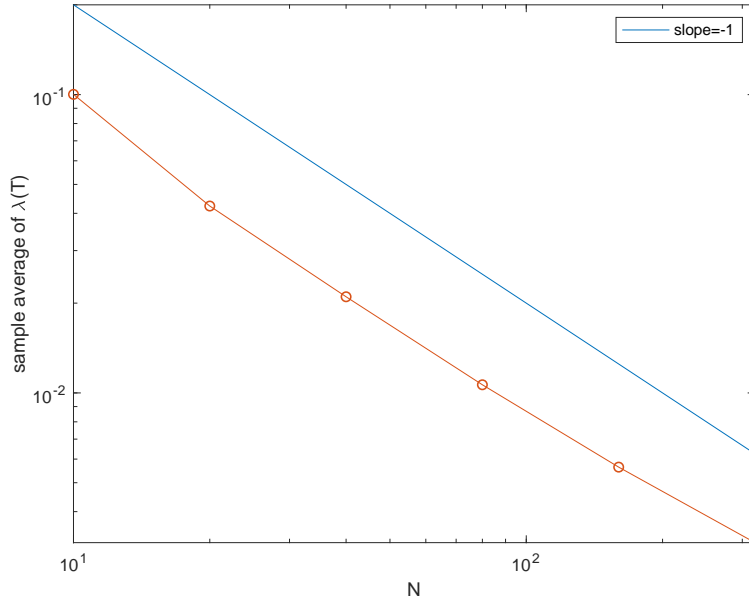


Figure 2: Sample average of $\lambda(T)$, $T = 20$, for (1.3) with (1.4), for various N . $n_{test} = 10000$ samples are taken, while $n_{discard} = 100$ samples with largest $\lambda(T)$ are discarded. The horizontal axis is N , the circles are sample averages of $\lambda(T)$, and the straight line is slope -1.

data is *well-mixed*.

- With stochastic communication, the expected large time behavior may still be observed with large probability.

There are a few future topics to study:

- Generalizing our results to cases where the coefficients depend on the locations \mathbf{x}_i and \mathbf{y}_j .
- Generalizing our results to models with more than two groups. Numerical examples in [14] have shown the separation for three groups.
- Second-order Cucker-Smale models [5, 6, 11] with anti-alignment and possibly stochastic communication. Some partial results in this direction are already obtained by [9] in deterministic setting.

References

- [1] E. Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.
- [2] N. Bellomo, M. Herrero, and A. Tosin. On the dynamics of social conflicts: Looking for the black swan. *Kinetic Related Models*, 6:459–479, 2013.
- [3] A. Bertozzi, J. A. Carrillo, and T. Laurent. Blow-up in multidimensional aggregation equations with mildly singular interaction kernels. *Nonlinearity*, 22:683–710, 2009.

- [4] V. D. Blondel, J. M. Hendricks, and J. N. Tsitsiklis. On Krause’s multi-agent consensus model with state-dependent connectivity. *IEEE Trans. Automat. Control*, 54(2586-2597), 2009.
- [5] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Trans. Automat. Control*, 52(5):852–862, 2007.
- [6] F. Cucker and S. Smale. On the mathematics of emergence. *Jpn. J. Math.*, 2(1):197–227, 2007.
- [7] P. Degond and S. Motsch. Continuum limit of self-driven particles with orientation interaction. *Math. Model Methods Appl. Sci.*, 18:1193–1215, 2008.
- [8] B. Duering, P. Markowich, J. F. Pietschmann, and M. T. Wolfram. Boltzmann and Fokker-Planck equations modeling opinion formation in the presence of strong leaders. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 465:3687–3708, 2012.
- [9] D. Fang, S.-Y. Ha, and S. Jin. Emergent behaviors of the Cucker-Smale ensemble under attractive-repulsive couplings and Rayleigh frictions. *Math. Model Methods Appl. Sci.*, to appear.
- [10] S.-Y. Ha, T. Ha, and J.-H. Kim. On the complete synchronization of the Kuramoto phase model. *Physica D: Nonlinear Phenomena*, 239(17):1692–1700, 2010.
- [11] S.-Y. Ha and E. Tadmor. From particle to kinetic and hydrodynamic descriptions of flocking. *Kinetic and Related Models*, 1(3):415–435, 2008.
- [12] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence: Models, analysis and simulation. *J. Artificial Soc. Social Simul.*, 5, 2002.
- [13] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.
- [14] S. Jin, L. Li, and J.-G. Liu. Random Batch Methods (RBM) for interacting particle systems. *preprint*.
- [15] F. Juhasz. The asymptotic behavior of Fiedler’s algebraic connectivity for random graphs. *Discrete Mathematics*, 96:59–63, 1991.
- [16] S. Motsch and E. Tadmor. Heterophilious dynamics enhances consensus. *SIAM Review*, 56(4):577–621, 2014.
- [17] G. Toscani. Kinetic models of opinion formation. *Commun. Math. Sci.*, 4:481–496, 2006.