Consensus of Interacting Particle Systems on Erdös-Rényi Graphs

Grant Schoenebeck* Fang-Yi Yu†

Abstract

Interacting Particle Systems—exemplified by the voter model, iterative majority, and iterative $k-$majority processes—have found use in many disciplines including distributed systems, statistical physics, social networks, and Markov chain theory. In these processes, nodes update their "opinion" according to the frequency of opinions amongst their neighbors.

We propose a family of models parameterized by an update function that we call Node Dynamics: every node initially has a binary opinion. At each round a node is uniformly chosen and randomly updates its opinion with the probability distribution specified by the value of the update function applied to the frequencies of its neighbors' opinions.

In this work, we prove that the Node Dynamics converge to consensus in time $\Theta(n \log n)$ in complete graphs and dense Erdös-Rényi random graphs when the update function is from a large family of "majority-like" functions. Our technical contribution is a general framework that upper bounds the consensus time. In contrast to previous work that relies on handcrafted potential functions, our framework systematically constructs a potential function based on the state space structure.

1 Introduction

We propose the following stochastic process—that we call ***Node Dynamics***—on a given network of n agents parameterized by an update function $f : [0, 1] \to [0, 1]$. In the beginning, each agent holds a binary "opinion", either red or blue. Then, in each round, an agent is uniformly chosen and updates its opinion to be red with probability $f(p)$ and blue with probability $1-f(p)$ where p is the fraction of its neighbors with the red opinion.

Node dynamics generalizes processes of interest in many different disciplines including distributed systems, statistical physics, social networks, and even biology.

Voter Model: In the voter model, at each round, a random node chooses a random neighbor and updates to its opinion. This corresponds to the Node Dynamics with

$$f(x) = x.$$

This models has been extensively studied in mathematics [15, 22, 27, 28], physics [6, 9], and even in social networks [8, 34, 35, 36, 14]. A key question studied is how long it takes the dynamics to reach consensus on different network typologies.

Iterative majority: In the iterative majority dynamics, in each round, a randomly chosen node updates to the opinion of the majority of its neighbors. This corresponds to the Node Dynamics where

$$f(x) = \begin{cases} 1 & \text{if } x > 1/2; \\ 1/2 & \text{if } x = 1/2; \\ 0 & \text{if } x < 1/2. \end{cases}$$

Typically works about Majority Dynamics study 1) when the dynamics converge, how long it takes the dynamics to converge, and whether they converge to the original majority opinion—that is, does majority dynamics successfully aggregate the original opinion [25, 7, 23, 31, 37].

Iterative k-majority: In this dynamics, in each round, a randomly chosen node collects the opinion of k randomly chosen (with replacement) neighbors and updates to the opinion of the majority of those k opinions. This corresponds to the Node Dynamics where

$$f(x) = \sum_{\ell=\lceil k/2 \rceil}^{k} \binom{k}{\ell} x^{\ell}(1-x)^{n-\ell}.$$

A synchronized variant of this dynamics is proposed as a protocol for stabilizing consensus: collection of n agents initially hold a private opinion and interact with the goal of agreeing on one of the choices, in the presence of $O(\sqrt{n})$-*dynamic adversaries* which can adaptively change the opinions of up to $O(\sqrt{n})$ nodes at every round. In the synchronized variant of this dynamics, Doerr et al. [17] prove 3-majority reaches "stabilizing almost" consensus on the complete graph in the presence of $O(\sqrt{n})$-dynamic adversaries. Many works extend this result beyond binary opinions [16, 13, 5, 1].

*University of Michigan, schoeneb@umich.edu. He is gratefully supported by the National Science Foundation under Career Award 1452915 and AitF Award 1535912.

†University of Michigan, fayu@umich.edu. He is gratefully supported by the National Science Foundation under AitF Award 1535912.

Iterative ρ-noisy majority model: [20, 21] In this dynamics, in each round, a randomly chosen node updates the majority opinion of its neighbors with probability $1 - \rho$ and uniformly at random with probability ρ.

$$f(x) = \begin{cases} 1 - \rho/2 & \text{if } x > 1/2; \\ 1/2 & \text{if } x = 1/2; \\ \rho/2 & \text{if } x < 1/2. \end{cases}$$

Genetic Evolution Model: In biological systems, the chance of survival of an animal can depend on the frequencies of its kin and foes in the network [3, 29]. Moreover, this frequency depending dynamics is also known to model the dynamics for maintaining the genetic diversities of a population [24, 32].

Our Contribution We focus on a large set of update functions f that are symmetric, smooth, and satisfy a property well call "majority-like", intuitively meaning that agents update to the majority opinion strictly more often than the fraction of neighbors holding the majority opinion. We obtain tight bounds for the consensus time—the time that it takes the system to reach a state where each node has an identical opinion—on Erdös-Rényi random graphs.

Our main technical tool is a novel framework for upper bounding the hitting time for a general discrete-time homogeneous Markov chain $(\mathcal{X}, P)$, including non-reversible and even reducible Markov chains. This framework decomposes the problem so that we only need to upper bound two sets of parameters for all $x \in \mathcal{X}$—the reciprocal of the probability of decreasing the distance to target $1/p^+(x)$ and the ratio of the probability of decreasing the distance to the target and the probability of increasing the distance to the target: $p^-(x)/p^+(x)$. Our technique can give much stronger bounds than simply lower bounding $p^-(x)$ and upper bounding $p^+(x)$.

Once we apply this decomposition to our consensus time problem, the problem becomes very manageable. We show the versatility of our approach by extending the results to a variant of the stabilizing consensus problem, where we show that all majority-like dynamics convergence quickly to the "stabilizing almost" consensus on the complete graph in the presence of adversaries.

A large volume of literature is devoted to bounding the hitting time of different Markov process and achieving fast convergence. The techniques typically employed are (1) showing the Markov chain has fast mixing time [30], (2) reducing the dimension of the process into small set of parameters (e.g. the frequency of each opinion) and using a mean field approximation and concentration property to control the behavior of the process [5], or (3) using handcrafted potential functions [31].

Our results fill in a large gap that these results do not adequately cover. Mixing time is not well-defined in non-reversible or reducible Markov chains, and so does not apply to Markov chains with multiple absorption states, like in the consensus time question we study. Reducing the dimension and using a mean field approximation fails for two reasons. First, summarizing with a small set of parameters is not possible when the process of interest has small imperfections (like in a fixed Erdös-Rényi graph). Second, the mean-field of our dynamics has unstable fixed points; in such cases the mean field does not serve as a useful proxy for the Markov process. Handcrafting potential functions also runs into several problems: the first is that because we consider dynamics on random graphs, the dynamic is not a priori well specified; so there is no specific dynamic to handcraft a potential function for. Secondly, we wish to solve the problem for a large class of update functions f, and so cannot individually hand-craft a potential function for each one. Typically, the potential function is closely tailored to the details of the process.

Additional Related Work Our model is similar to that of Schweitzer and Behera [33] who study a variety of update functions in the homogeneous setting (complete graph) using simulations and heuristic arguments. However, they leave a rigorous study to future work.

2 Preliminaries

2.1 Node Dynamics Given an undirected graph $G = (V, E)$ let $\Gamma(v)$ be the neighbors of node v and $\deg(v) = |\Gamma(v)|$.

We define a ***configuration*** $x^{(G)} : V \to \{0, 1\}$ to assign the "color" of each node $v \in G$ to be $x^{(G)}(v)$ so that $x^{(G)} \in \{0, 1\}^n$. We will usually suppress the superscript when it is clear. We will use uppercase (e.g., $X^{(G)}$) when the configuration is a random variable. Moreover we say v is ***red*** if $x(v) = 1$ and is ***blue*** if $x(v) = 0$. We then write the set of red vertices as $x^{-1}(1)$. We say that a configuration x is ***in consensus*** if $x(\cdot)$ is the constant function (so all nodes are red or all nodes are blue). Given a node v in configuration x we define $r_x(v) = \frac{|\Gamma(v) \cap X^{-1}(1)|}{\deg(v)}$ to be its fraction of red neighbors.

DEFINITION 2.1. *An **update function** is a mapping $f : [0, 1] \mapsto [0, 1]$ with the following properties:*

Monotone $\forall x, y \in [0, 1]$, *if $x < y$, then $f(x) \le f(y)$.*

Symmetric $\forall t \in [0, 1/2]$, $f(1/2 + t) = 1 - f(1/2 - t)$.

Absorbing *$f(0) = 0$ and $f(1) = 1$.*

We define node dynamics as follows:

DEFINITION 2.2. *A **node dynamics** $\mathrm{ND}(G, f, X_0)$ with an undirected graph $G = (V, E)$, update function f and initial configuration X_0 is a stochastic process over configurations at time t, $\{X_t\}_{t\geq 0}$ where X_0 is the initial configuration. The dynamics proceeds in rounds. At round t, a node v is picked uniformly at random, and we update*

$$X_t(v) = \begin{cases} 1 & \text{with probability } f(r_{X_{t-1}}(v)) \\ 0 & \text{otherwise} \end{cases}$$

This formulation is general enough to contain many well known dynamics such as the aforementioned voter model, iterated majority model, and 3-majority dynamics.

Note that in some of the original definitions the nodes syncronously update; whereas, to make our presentation more cohesive, we only consider asynchronous updates.

In this paper, we will focus on the interaction between the update function f and geometric structure of G. More specifically, we are interested in the consensus time defined as following.

DEFINITION 2.3. *The **consensus time** of node dynamics $\mathrm{ND}(G, f, X_0)$ is a random variable $T(G, f, X_0)$ denoting the first time step that* ND *is in a consensus configuration. The **(maximum) expected consensus time** $\mathrm{ME}(G, f)$ is the maximum expected consensus time over any initial configuration, $\mathrm{ME}(G, f) = \max_{X_0} \mathbb{E}[T(G, f, X_0)]$.*

Now we define some properties of functions.

DEFINITION 2.4. *Given positive M_1, M_2, a function $f : I \subseteq \mathbb{R} \mapsto \mathbb{R}$ is called M_1-**Lipschitz** in $I \subseteq \mathbb{R}$ if for all $x, y \in I$,*

$$|f(x) - f(y)| \leq M_1|x - y|.$$

*Moreover, f is M_2-**smooth** in $I \subseteq \mathbb{R}$ if for all $x, y \in I$,*

$$|f'(x) - f'(y)| \leq M_2|x - y|.$$

2.2 Potential Theory for Markov Chains Let $\mathcal{M} = (X_t, P)$ be a discrete time-homogeneous Markov chain with finite state space Ω and transition matrix P. For $x, z \in \Omega$, we define $\tau_a(x)$ to be the ***hitting time*** for a with initial state x:

$$\tau_a(x) \triangleq \min\{t \geq 0 : X_t = a, X_0 = x\},$$

and $\tau_A(x)$ to be the hitting time to a set of state $A \subseteq \Omega$:

$$\tau_A(x) \triangleq \min\{t \geq 0 : X_t \in A, X_0 = x\}.$$

By the Markov property, the expected hitting time can be written as linear equation.

$$\mathbb{E}_{\mathcal{M}}[\tau_A(x)] = \begin{cases} 1 + \sum_{y\in\Omega} P_{x,y}\mathbb{E}_{\mathcal{M}}[\tau_A(y)] & \text{if } x \notin A, \\ 0 & \text{if } x \in A \end{cases}$$

Due to the memory-less property of Markov chain, sometimes it is useful to analyze its first step. Let's consider a general measurable function $w : \Omega \mapsto \mathbb{R}$. If the Markov chain starts at state $X = x$, the next state is the random variable X', then the average change of $w(X)$ in one transition step is given by

$$(\mathcal{L}w)(x) \triangleq \mathbb{E}_{\mathcal{M}}[w(X') - w(X)|X = x] = \sum_{y\in\Omega} P_{x,y}w(y) - w(x)$$

To reduce the notation we will use $\mathbb{E}_{\mathcal{M}}[w(X')|X]$ to denote the expectation of the measurable function $w(X')$ given the previous state at X.

DEFINITION 2.5. *Given Markov chain $\mathcal{M}$ with state space Ω, $D \subsetneq \Omega$, and two real-valued functions ψ, ϕ with domain Ω, we define the **Poisson equation** as the problem of solving the function $w : \Omega \mapsto \mathbb{R}$ such that*

$$\mathcal{L}w(x) = -\phi(x) \text{ where } x \in D,$$
$$w(x) = \psi(x) \text{ where } x \in \partial D.$$

where the $\partial D \triangleq \cup_{x\in D} \operatorname{supp} p(x, \cdot) \setminus D$ is the exterior boundary of D w.r.t the Markov chain.

Note that solving the expected hitting time of set A is a special case of the above problem by taking $D = \Omega \setminus A$, $\phi(x) = 1$ and $\psi(x) = 0$. The next fundamental theorem shows that super solutions to an associated boundary value problem provide upper bounds for the Poisson equation in Definition 2.5.

THEOREM 2.1. (MAXIMUM PRINCIPLE [19]) *Given Markov Chain $\mathcal{M}$ with state space Ω, $D \subsetneq \Omega$, and two real-valued functions ψ, ϕ with domain Ω, suppose $s : \Omega \mapsto \mathbb{R}$ is a non-negative function satisfying*

$$\mathcal{L}s(x) \leq -\phi(x) \text{ where } x \in D,$$
$$s(x) \geq \psi(x) \text{ where } x \in \partial D.$$

Then $s(x) \geq w(x)$ for all $x \in D$.

COROLLARY 2.1. (SUPER SOLUTION FOR HITTING TIME) *Given Markov Chain $\mathcal{M}$ with state space Ω and a set of states $A \subsetneq \Omega$, suppose $s_A : \Omega \mapsto \mathbb{R}$ is a non-negative function satisfying*

$$\mathcal{L}s_A(x) \leq -1 \text{ where } x \notin A, \quad (2.1)$$
$$s_A(x) \geq 0 \text{ where } x \in A.$$

*Then $s_A(x) \geq \mathbb{E}_{\mathcal{M}}[\tau_A(x)]$ for all $x \notin A$. Moreover we call s_A a **potential function** for short.*

2.3 Concentration Inequalities

THEOREM 2.2. ([4]) *Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a finite set of N real numbers, that $X_1, \ldots, X_n$ denote a random sample without replacement from $\mathcal{X}$ and that $Y_1, \ldots, Y_n$ denote a random sample with replacement from $\mathcal{X}$. If $f : \mathbb{R} \mapsto \mathbb{R}$ is continuous and convex, then*

$$\mathbb{E}f\left(\sum_{i=1}^{n} X_i\right) \leq \mathbb{E}f\left(\sum_{i=1}^{n} Y_i\right)$$

THEOREM 2.3. (A CHERNOFF BOUND [18]) *Let $X \triangleq \sum_{i=1}^{n} X_i$ where X_i for $i \in [n]$ are independently distributed in $[0, 1]$. Then for $0 < \epsilon < 1$*

$$\mathbb{P}[X > (1+\epsilon)\mathbb{E}X] \leq \exp\left(-\frac{\epsilon^2}{3}\mathbb{E}X\right)$$
$$\mathbb{P}[X < (1-\epsilon)\mathbb{E}X] \leq \exp\left(-\frac{\epsilon^2}{2}\mathbb{E}X\right)$$

If a bounded function g on a probability space (X, P) which is Lipschitz for most of the measure in X, then the following theorem prove a concentration property of g by using union bound and Azuma inequality.

THEOREM 2.4. (BAD EVENTS [18]) *Consider a random object $X = (X_1, \ldots, X_n)$ with probability measure $\mathbb{P}$. Let E be an event in the space of X. Fix a real-valued function g with domain X which is bounded, $m \leq g(X) \leq M$. Let c_i be the maximum effect of changing the ith input coordinate of g conditioned both inputs being in E:*

$$\sup_{x,x' \in E, \forall j \neq i \, x_j = x'_j} |g(x) - g(x')| \leq c_i.$$

Then,

$$\mathbb{P}\left[g(X) > \mathbb{E}[g(X)] + t + (M-m)\mathbb{P}[\neg E] \middle| E\right]$$

is bounded above by $\exp\left(-2t^2/\left(\sum_i c_i^2\right)\right)$.

We say a sequence of events $\{A_n\}_{n\geq 1}$ happens *with high probability* if $\lim_{n\to\infty} \mathbb{P}[A_n] = 1$ that is $\mathbb{P}[A_n] = 1 - o(1)$.

2.4 Erdös-Rényi Random Graphs

Here we present the definition of Erdös-Rényi random graphs and show several well-known properties of them that we need.

DEFINITION 2.6. (ERDÖS-RÉNYI RANDOM GRAPH) *$G_{n,p}$ is a random undirected graph on node set $V = [n]$[1] where each pair of nodes is independently connected with a fixed probability p. We further use $\mathcal{G}_{n,p}$ to denote this random object.*

[1] $[n] = \{1, 2, \ldots, n\}$

Let A_G be the *adjacency matrix of* G, so $(A_G)_{i,j} = 1$ if $v_i \sim v_j$ and 0 otherwise, and $\bar{A} = \mathbb{E}_{\mathcal{G}_{n,p}}[A_G]$, so $\bar{A}_{i,j} = p$ if $i \neq j$ and 0 otherwise. Let $deg(v)$ be the degree of node v.

DEFINITION 2.7. *The* weighted adjacency matrix of undirected graph G *is defined by*

$$M_G(i,j) = \begin{cases} \frac{1}{\sqrt{deg(v_i)deg(v_j)}} & \text{if } (A_G)_{i,j} = 1; \\ 0 & \text{otherwise} . \end{cases}$$

DEFINITION 2.8. (EXPANSIVENESS [12]) *For $\lambda \in [0, 1]$, we say that a undirected graph G is a λ-expander if if $\lambda_k(M_G) \leq \lambda$ for all $k > 1$ where $\lambda_k(M_G)$ is the k-th largest eigenvalue.*

THEOREM 2.5. (SPECTRAL PROFILE OF $G_{n,p}$ [11]) *For $G_{n,p}$, we denote I as identity matrix and J as the matrix that has ones. If $G_{n,p}$ has $p = \omega(\frac{\log n}{n})$, then with probability at least $1 - 1/n$, for all k*

$$|\lambda_k(M_G) - \lambda_k(\bar{M})| = O\left(\sqrt{\log n/(np)}\right)$$

where $(\bar{M})_{i,j} = \frac{1}{n-1}$ if $i \neq j$ and $(\bar{M})_{i,i} = 0$.

Because the spectrum of $\bar{M}$ is $\{1, -1/(n-1)\}$ where $-1/(n-1)$ has multiplicity $n-1$, we can have the following corollary

COROLLARY 2.2. *If $p = \omega(\frac{\log n}{n})$, the $G \sim \mathcal{G}_{n,p}$ is $O\left(\sqrt{\frac{\log n}{np}}\right)$-expander with probability $1 - O(1/n)$,*

Let $e(S,T)$ denote the number of edges between S and T (double counting edges from $S \cap T$ to itself), and let $vol(S)$ count the number of edges adjacent to S. The following lemma relates the number of edges between two sets of nodes in an expander to their expected number in a random graph.

LEMMA 2.1. (IRREGULAR MIXING LEMMA [10]) *If G is a λ-expander, then for any two subsets $S, T \subseteq V$:*

$$\left| e(S,T) - \frac{vol(S)vol(T)}{vol(G)} \right| \leq \lambda\sqrt{vol(S)vol(T)}$$

Finally, let $E(\delta_d; v)$ denote the event that the degree of some fixed node v is between $(1-\delta_d)np$ and $(1+\delta_d)np$ and let $E(\delta_d) = \cap_{v\in V} E(\delta_d; v)$ a ***nearly uniform degree event***. By applying theorem 2.3 we yields the following lemma.

LEMMA 2.2. (UNIFORM DEGREE) *For any $v \in V$, if $G \sim \mathcal{G}_{n,p}$*

$$\mathbb{P}[E(\delta_d; v)] \leq 2\exp\left(-\delta_d^2 np/3\right) \tag{2.2}$$

Furthermore, by union bound

$$\mathbb{P}[E(\delta_d)] \leq 2n\exp\left(-\delta_d^2 np/3\right) \tag{2.3}$$

3 Warm-up: Majority-liked Update Function on Complete Graph

In this section we consider majority-like node dynamics on the complete graph K_n with n nodes in which every pair of nodes has an edge (no self-loops). We use this as a toy example to give intuition for dense Erdös-Rényi graphs even though we will obtain better bounds later.

THEOREM 3.1. *Let $\mathcal{M} = \mathrm{ND}(K_n, f, X_0)$ be a node dynamic over the complete graph K_n with n nodes. If the update function f satisfies $\forall x : 1/2 < x < 1$ then $x \leq f(x)$, then the maximum expected consensus time of a node dynamic over K_n is*

$$\mathrm{ME}(K_n, f) = O\left(n^2\right).$$

A standard method of proving fast convergence is to guess a potential function of each state and prove the expectation decreases by 1 after every step—this is just an application of corollary 2.1.

As a warm-up, we will prove theorem 3.1 by guessing a potential function and applying corollary 2.1.

Proof. [theorem 3.1] Given a configuration x, define $Pos(x) \triangleq |x^{-1}(1)|$ then for all red nodes v where $x^{(K_n)}(v) = 1$ have $r_x(v) = \frac{Pos(x)-1}{n-1}$; otherwise $r_x(v) = \frac{Pos(x)}{n-1}$.

Because the node dynamics $\mathcal{M}$ is on the complete graph, $\mathcal{M}$ is *lumpable* with respect to partition $\{\Sigma_l\}_{0\leq l\leq n}$ where $\Sigma_l = \{x \in \Omega : Pos(x) = l\}$ such that for any subsets Σ_i and Σ_j in the partition, and for any states x, y in subset Σ_i,

$$\sum_{z\in\Sigma_j} P(x, z) = \sum_{z\in\Sigma_j} P(y, z)$$

Furthermore, inspired by an analysis of Voter Model [2] we consider $\psi : [n] \mapsto \mathbb{R}$ as

$$\psi(k) = (n-1)\big[k(H(n-1) - H(k-1)) + (n-k)(H(n-1) - H(n-k-1))\big]$$

where $H(k) \triangleq \sum_{\ell=1}^{k} \frac{1}{\ell}$, and define the potential function as

$$\phi(x) = \psi(Pos(x)) \tag{3.4}$$

The proof of the following claim is deferred to the full version, and here we just give some intuition as to why this potential function for the voter model works. The sequence $(Pos(X_t))_{t\geq 0}$ can be seen as a random walk on $0, 1, \ldots, n$ with *drift*[2]. Moreover the drift depends on $f(Pos(x)) - Pos(x)$. For voter model $f(x) = x$, there is no drift. For majority-like function because there is a positive drift toward n when $Pos(x) > n/2$; and a negative drift toward 0 when when $Pos(x) < n/2$. Informally the drift is always *helping* and thus the potential function for voter models works.

[2]The formal definition of drift is in Equation (4.8).

CLAIM 3.1. *Our definition of ϕ satisfies the inequalities* (2.1)*: Given Markov Chain $\mathcal{M} = \mathrm{ND}(K_n, f, X_0)$ in theorem 3.1, ϕ defined in* (3.4) *are non-negative and satisfy*

$$\mathcal{L}\phi(x) \leq -1 \text{ where } x \neq 0^n, 1^n,$$
$$\phi(x) \geq 0 \text{ where } x = 0^n, 1^n.$$

Combining claim 3.1 and corollary 2.1, we have

$$\mathbb{E}_{\mathcal{M}}[T(K_n, f, x)] \leq \phi(x).$$

By direct computation, if $0 \leq k < n$, $\psi(k+1) - \psi(k) = (n-1)(H(n-k-1) - H(k))$. Therefore, the maximum $\psi(k)$ happens at $k = \lfloor n/2 \rfloor$,

$$\mathrm{ME}(K_n, f) \leq \psi(\lfloor n/2 \rfloor) \leq (\ln 2)n^2,$$

and completes our proof.

4 Smooth Majority-like Update Function on Dense G_{np}

In this section, we consider the smooth Majority-liked update function defined as follows:

DEFINITION 4.1. *We call an update function f a* smooth majority-like update function *if it satisfies $\forall x : 1/2 < x < 1$ then $x < f(x)$ and the following technical conditions hold:*

Lipschitz *There exists M_1 such that f is M_1-Lipschitz in $[0, 1]$.*

Condition at $1/2$ *There exists an open interval $I_{1/2}$ containing $1/2$ and constants $1 < \check{M}_1$, $0 < \check{M}_2$ such that f is $\check{M}_2$-smooth in $I_{1/2}$ and $1 < \check{M}_1 \leq f'(1/2)$.*

Condition at 0 **and** 1 *There exists intervals $I_0 \ni 0$, $I_1 \ni 1$ and a constant $\hat{M}_1 < 1$ such that $\forall x \in I_0$, $f(x) \leq \hat{M}_1 x$ and $\forall x \in I_1$, $1 - f(x) \leq \hat{M}_1(1 - x)$.*

Intuitively, the majority-like update function should be "smooth" and not tangent with $y = x$. The following figure shows an example of smooth majority-like update function. Now we are ready to state our main theorem.

THEOREM 4.1. *Let $\mathcal{M} = \mathrm{ND}(G, f, X_0)$ be a node dynamic over $G \sim G_{np}$ with $p = \Omega(1)$, and let f be a smooth majority-like function. Then the expected consensus time of a node dynamic over G is*

$$\mathrm{ME}(G, f) = O(n \log n)$$

with high probability.

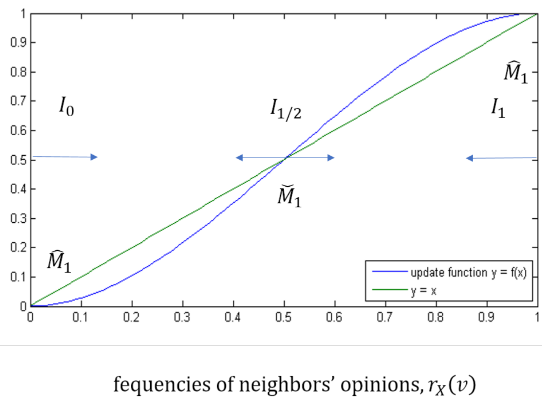

Figure 1: An example of smooth majority-like update function.

This theorem shows the fast convergence rate of this process. Note that there is some chance of getting a disconnected graph $G \sim \mathcal{G}_{n,p}$ which results in a reducible Markov chain $\mathcal{M}$ which cannot converge from some initial configurations. Therefore, we can only ask for the fast convergence result *with high probability.*

We note that, the technical conditions exclude interactive majority updates, which we leave for future work.

4.1 Proof Overview Here we will first outline the structure of the proof. In section 4.2 we propose a paradigm for proving an upper bound for the hitting time when the state space has special structure. In section 4.3, we use the result in section 4.2 to prove theorem 4.1.

where large literature have devote to different process. Most of them achieve fast convergence result by using handcraft potential function or showing the Markov chain has fast mixing time. However it is not easy to find clever potential function for any process, and the fast mixing time is not well defined in reducible Markov chain. Recall that the expected consensus time is

$$\tau(x) \triangleq \mathbb{E}_{\mathcal{M}}[T(G, f, x)]$$

which is exactly the hitting time of states 0^n and 1^n. However in contrast to section 3, finding a clever potential function is much harder here. We prove theorem 4.1 using that *the expected hitting can be formulated as a system of linear equations* (2.1) and by explicitly estimating an upper bound of this system of linear equations. Moreover, following the intuition in section 3, the Markov chain $\mathcal{M}$ can be *nearly* characterized by one parameter $Pos(x)$ when the node dynamics is on a graph that is close to the complete graph. We exploit this structure of our Markov chain and *construct* a potential function for Equations (2.1).

4.2 A Framework for Upper Bounding the Hitting Time We want to upper bound the hitting time from arbitrary state x to $\{0^n, 1^n\}$ denoted as $\tau(x)$ of a given time-homogeneous Markov chain $\mathcal{M} = (\Omega, P)$ with finite state space $\Omega = \{0,1\}^n$ where $P(x, y) > 0$ only if the states x, y only differ by one digit, $|x - y| \leq 1$.

We let $Pos(x)$ be the *position* of state $x \in \Omega$:

$$Pos(x) \triangleq |x^{-1}(1)|, \text{ and } pos(x) \triangleq Pos(x)/n \tag{4.5}$$

and the *bias* of x as

$$Bias(x) \triangleq |n/2 - Pos(x)|, \text{ and } bias(x) \triangleq Bias(x)/n \tag{4.6}$$

Note that the $Bias(x) = n/2$ if and only if $x = 0^n, 1^n$.

Suppose that $\mathcal{M}$ can be "almost" characterized by one parameter $Bias(x)$. Informally, we want the transitions at states x and y to be similar if $Bias(x) = Bias(y)$. Therefore with the notion of first step analysis we define $\{(p_G^+(x), p_G^-(x))\}_{x\in\Omega}$ where

$$\begin{aligned} p_G^+(x) &= \mathbb{P}_{\mathcal{M}}[Bias(X') = Bias(X) + 1 | X = x], \\ p_G^-(x) &= \mathbb{P}_{\mathcal{M}}[Bias(X') = Bias(X) - 1 | X = x]. \end{aligned} \tag{4.7}$$

Moreover, we call $p_G^+(x)$ the ***exertion*** and define the ***drift*** of state x as follows

$$D(x) \triangleq \mathbb{E}_{\mathcal{M}}[Bias(X') - Bias(X) | X = x]. \tag{4.8}$$

It is easy to see $D(x) = p^+(x) - p^-(x)$.

Since $\mathcal{M}$ can be almost characterized by one parameter, $Bias(x)$, $\mathcal{M}$ is almost lumpable with respect to the partition induced by $Bias(\cdot)$. The following lemma gives us a scheme for constructing an upper bound for the hitting time:

Lemma 4.1. (Pseudo-lumpability lemma) *Let $\mathcal{M} = (\Omega, P)$ have finite state space $\Omega = \{0,1\}^n$ with even n*[3] *and $P(x, y) > 0$ only if the states x and y differ in at most one coordinate and*

$$\begin{aligned} d_0 &= \max_{x: Bias(x) = 0} \frac{1}{p^+(x)} \\ d_l &= \max_{x: Bias(x) = l} \frac{1}{p^+(x)} + \max_{x: Bias(x) = l} \left(\frac{p^-(x)}{p^+(x)}\right) d_{l-1} \end{aligned} \tag{4.9}$$

where $0 < l < n/2$, and $\{(p^+(x), p^-(x)\}_{x\in\Omega}$ are as defined in (4.7). *Then the maximum expected hitting*

[3] To avoid cumbersome notion of parity we only consider n to be even here.

time from state x to $\{0^n, 1^n\}$ can be bounded as follows:

$$\max_{x\in\Omega} \mathbb{E}_{\mathcal{M}}[\tau(x)] \leq \sum_{0\leq \ell < n/2} d_\ell$$

where $\tau(x)$ denotes the hitting time from state x to $\{0^n, 1^n\}$.

Remark 4.1. *At first glance it appears this lemma "couples" the process $\mathcal{M}$ with a birth-and-death chain [26], but is actually stronger as the following example illustrates. We define an unbiased random walk where the self transition probability of nodes differs. For all $x \in \{0,1\}^n \setminus \{0^n, 1^n\}$ let $p^+(x) = p^-(x) = \frac{1}{2+x_1}$, and 0^n and 1^n are absorbing states. This lemma yield a polynomial time upper bound because $1/p^+(x) = 3$ and $p^-(x)/p^+(x) = 1$. On the other hand, consider a birth-and-death chain on $\{0, 1, \ldots, n/2\}$, such that $P(k, k+1) = \min_{x\in\Omega:Bias(x)=k} p^+(x)$ and $P(k, k-1) = \max_{x\in\Omega:Bias(x)=k} p^-(x)$. Because $P(k, k+1) = 1/3$ and $P(k, k-1) = 1/2$ for all $0 < k < n/2$, the corresponding birth-and-death chain takes exponential time to reach $n/2$.*

lemma 4.1 can be derived from corollary 2.1 and is proven in appendix A. Intuitively, to get a potential function $s(x)$ for hitting time $\tau(x)$, we order the states in terms of the value of $Bias(\cdot)$, and take the process as a non-uniform random walk on $[n]$. Then we recursively estimate $s(x)$ for each x in increasing order of $Bias(x)$.

To use lemma 4.1, to upper bound $\tau(x)$ we need to prove for *every* configurations $x \in \Omega$

1. An upper bound for $1/p^+(x)$.

2. An upper bound for $p^-(x)/p^+(x)$.

In theorem 4.2 we give a framework that uses the upper bounds for $1/p^+(x)$ and $p^-(x)/p^+(x)$ to obtain upper bounds for expected hitting time. To have some intuition about the statement of the theorem, observe that if the drift $D(x)$ is bounded below by some positive constant both $1/p^+(x)$ and $p^-(x)/p^+(x)$ have nice upper bounds. However, this bound fails when the drift is near zero or even negative. Taking our node dynamics on dense G_{np} as an example, when the states have either very small or very large $bias(x)$ the drift $D(x)$ can be very close to zero or even negative. The drift near 1/2 is close to 0 because the effects of red and blue largely cancel each other. The drift near the extreme point is small because there are very few nodes outside the majority.

As a result, we partition the states into subsets, and take addition care on the sets of states with small drift. In theorem 4.2 we partition the states into $\Sigma^{\mathrm{s}}, \Sigma^{\mathrm{m}}, \Sigma^{\mathrm{l}}$ according to the bias as follows:

$$\begin{aligned} \Sigma^{\mathrm{s}} &= \{x \in \Omega : bias(x) < \check{\epsilon}\} \\ \Sigma^{\mathrm{m}} &= \{x \in \Omega : \check{\epsilon} \leq bias(x) \leq 1/2 - \hat{\epsilon}\} \\ \Sigma^{\mathrm{l}} &= \{x \in \Omega : 1/2 - \hat{\epsilon} < bias(x)\} \end{aligned} \tag{4.10}$$

The small constants $\hat{\epsilon}$ and $\check{\epsilon}$ depend on the process.

Theorem 4.2. *Given $\mathcal{M} = (\Omega, P)$ defined in lemma 4.1, if there exist constants $\hat{\epsilon}$ and $\check{\epsilon}$ defining the partition $\Sigma^{\mathrm{s}}, \Sigma^{\mathrm{m}}$, and Σ^{l} and some constants $p^+, A_1, A_2, A_3, B_1 > 0$, and $0 < r, A_2, A_3 < 1$ such that*

$$p^+ < p^+(x) \leq 1 \text{ if } x \in \Sigma^{\mathrm{s}}, \Sigma^{\mathrm{m}} \tag{4.11}$$

$$r < \frac{p^+(x)}{(1/2 - bias(x))} \leq 1 \text{ if } x \in \Sigma^{\mathrm{l}} \tag{4.12}$$

and

$$\frac{p^-(x)}{p^+(x)} \leq 1 + A_1 \left(\frac{B_1}{\sqrt{n}} - bias(x)\right) \text{ if } x \in \Sigma^{\mathrm{s}} \tag{4.13}$$

$$\frac{p^-(x)}{p^+(x)} \leq 1 - A_2 \text{ if } x \in \Sigma^{\mathrm{m}} \tag{4.14}$$

$$\frac{p^-(x)}{p^+(x)} \leq 1 - A_3 \text{ if } x \in \Sigma^{\mathrm{l}}, \tag{4.15}$$

the maximum expected hitting time is

$$\max_{x\in\Omega} \mathbb{E}_{\mathcal{M}}[\tau(x)] = O(n \log n)$$

where $\tau(x)$ is the hitting time from state x to $\{0^n, 1^n\}$.

The proof of theorem 4.2, it is rather straightforward using lemma 4.1, and carefully constructing the potential function from the recursive Equation (4.9).

4.3 Proof of Theorem 4.1 In this section, we will use theorem 4.2, to prove an $O(n \log n)$ time bound by exploiting properties of our process. Specifically, let our node dynamic $\mathcal{M} = \mathrm{ND}(G, f, X_0)$ be a node dynamic over G sampled from $\mathcal{G}_{n,p}$, it is sufficient to prove an upper bound for $1/p^+_G(x)$ and $p^-_G(x)/p^+_G(x)$. Note that we use subscripts to emphasize the dependency of the graph G.

To apply theorem 4.2, we partition the states into three groups Σ^s, Σ^m, and Σ^l defined in (4.10). The constants $\check{\epsilon}$ and $\hat{\epsilon}$ depend on the update function f and the probability of an edge p and will be specified later. Figure 4.3 illustrates the partitions of the states.

The following lemma upper bounds $1/p^+_G(x)$:

Lemma 4.2. (lower bound for $p^+_G(x)$) *Given node process $\mathcal{M}$ on G, if G λ-expander with nearly uniform*

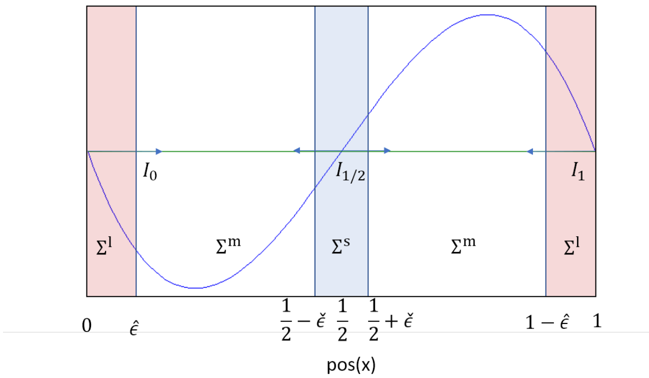

Figure 2: An illustration of partition in section 4.3.

degree $E(\delta_d)$*,* $\delta_d < 1$ *and* $\lambda^2 < \frac{1-\delta_d}{1+\delta_d} \cdot \min\{\frac{\hat{\epsilon}}{18}, \frac{(1/2-\hat{\epsilon})^2}{2}\}$*, then for* $p^+ \triangleq \frac{\hat{\epsilon}}{2} f\left(\frac{\hat{\epsilon}}{2}\right)$*,*

$$p^+ < p_G^+(x) \leq 1 \text{ if } x \in \Sigma^s \cup \Sigma^m \tag{4.16}$$

$$\frac{1}{4} < \frac{p_G^+(x)}{(1/2-bias(x))} \leq 1 \text{ if } x \in \Sigma^l. \tag{4.17}$$

This lemma is proved by apply mixing lemma 2.1 to show that the probability of increasing bias is (1) larger than some constant for $x \in \Sigma^s \cup \Sigma^m$ lemma B.1 and (2) proportional to the size of minority in Σ^l in lemma B.2. The proof details are in appendix B.

The second part follows from the following lemma:

LEMMA 4.3. (UPPER BOUND FOR $p_G^-(x)/p_G^+(x)$) *Given node process* $\mathcal{M}$ *on* G*, if* $G \sim \mathcal{G}_{n,p}$*, then there exist positive constant* A_1, A_2, A_3, B_1*, and* $0 < A_2, A_3 < 1$ *such that, with high probability,*

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 + A_1\left(\frac{B_1}{\sqrt{n}} - bias(x)\right) \text{ if } x \in \Sigma^s \tag{4.18}$$

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 - A_2 \text{ if } x \in \Sigma^m \tag{4.19}$$

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 - A_3 \text{ if } x \in \Sigma^l \tag{4.20}$$

Instead of bounding $p_G^-(x)/p_G^+(x)$ directly, the drift $D_G(x) \triangleq p_G^+(x) - p_G^-(x)$ is more natural to work with. Taking the complete graph as example, $D_G(x) = f(pos(x)) - pos(x)$. Therefore instead of proving an upper bound of $p_G^-(x)/p_G^+(x)$ directly, we prove an lower bound for the drift in Appendix B (lemma B.3, B.4, B.5, and B.8). Combining with lemma 4.2, these gives us an desired upper bound for $p_G^-(x)/p_G^+(x)$.

Proof. [theorem 4.1] By corollary 2.2 $G \sim \mathcal{G}_{n,p}$ is a $O\left(\sqrt{\frac{\log n}{np}}\right)$-expander with high probability. Thus, we can apply lemma 4.2 and 4.3 to theorem 4.2, which finishes the proof.

5 The Stabilizing Consensus Problem

The consensus problem in the presence of an adversary (known as Byzantine agreement) is a fundamental primitive in the design of distributed algorithms.

For the stabilizing-consensus problem—a variant of the consensus problem, Doerr et al. [17] proves synchronized 3-majority converges fast to an almost stable consensus on a complete graph in the presence of $O(\sqrt{n})$-dynamic adversaries which, at every round, can adaptively change the opinions of up to $O(\sqrt{n})$ nodes.

Here we consider an asynchronous protocol for this problem:

DEFINITION 5.1. *Given a complete network of n anonymous nodes with update function* f*, and* $F \in \mathbb{N}$*. In the beginning configuration, each node holds a binary opinion specified by* $x_0(\cdot)$*. In each round:*

1. *An adaptive dynamic adversary can arbitrarily corrupt up to* F *agents, and change the reports of their opinions in this run (the true opinion of these nodes is restored and will be reported once the adversary stops corrupting them).*
2. *A randomly chosen node updates its opinion according to node dynamics. (If the chosen node is corrupted by adversary in that run, the adversary can arbitrarily update the opinion of the chosen node.)*

DEFINITION 5.2. (n^γ-ALMOST CONSENSUS) *We say a complete network of n anonymous nodes reaches an* n^γ*-almost consensus if all but* $O(n^\gamma)$ *of the nodes support the same opinion.*

Our analysis in section 4 can be naturally extended to the stabilizing consensus problem and proves all majority-liked update functions (definition 4.1) are stabilizing almost consensus protocols and have the same convergence rate.

THEOREM 5.1. *Given* n *nodes, fixed* $\gamma > 1/2$*,* $F = O(\sqrt{n})$*, and initial configuration* $X_0 \in \{0,1\}^n$*, the node dynamic* $\mathrm{ND}(K_n, f, x_0)$ *on a complete graph with update function* f *reaches an* n^γ*-almost consensus in the presence of any* F*-corrupt adversary within* $O(n \log n)$ *rounds with high probability.*

REMARK 5.1. *The goal of this section is not to promote majority-liked node dynamics as a state-of-art protocol for the stabilizing consensus problem, but to show the versatile power of the our framework of proving convergence time in section 2.1. Additionally we modify the* *formulation of the problem here to make our presentation more cohesive.*

Let the random process with the presence of some fixed F-dynamic adversary $\mathcal{A}_F$ defined in theorem 5.1 be denoted $\mathcal{X}(\mathcal{A}_F) = (X_t)_{t\geq 0}$. Observe that our framework in section 4.2 only works for Markov chain, but with the presence of adaptive adversary the process is no longer a Markov chain. As a result, we "couple" this process with a nice Markov chain $\mathcal{Y}(F) = (Y_t)_{t\geq 0}$, and use the Markov chain as a proxy to understand the original process.

The proof has two parts: we first define the proxy Markov chain $\mathcal{Y}(F)$ and prove an upper bound of almost consensus time by using the tools in section 4.2. Secondly, we construct a monotone coupling between $\mathcal{Y}(F)$ and $\mathcal{X}(\mathcal{A}_F)$ to prove $\mathcal{X}(\mathcal{A}_F)$ also converges to almost consensus fast.

5.1 Upper Bounding the Expected Almost Consensus Time for $\mathcal{Y}(F)$. With the notation defined in section 4, we define $\mathcal{Y}(F)$. Informally, we construct $\mathcal{Y}(F)$ as a pessimistic version of $\mathrm{ND}(K_n, f, X_0)$ with the presence of an adversary: at every round the adversary tries to push the state toward the unbiased configuration, and it always corrupts F nodes with the minority opinion.

Initially, $Y_0 = X_0$. At time t if we set $y = Y_t$, Y_{t+1} is uniformly sampled from

$$\begin{aligned} &\{y' \in \Omega : \exists i \in [n], \forall j \neq i, y'_j = y_j\} \\ &\cap \{y' \in \Omega : Bias(y') = Bias(y) + 1\} \end{aligned} \tag{5.21}$$

with probability $\max\{f\left(\frac{1}{2} + bias(y)\right)\left(\frac{1}{2} - bias(y)\right) - \frac{(M_1+1)F}{n}, 0\}$, or uniformly sampled from

$$\begin{aligned} &\{y' \in \Omega : \exists i \in [n], \forall j \neq i, y'_j = y_j\} \\ &\cap \{y' \in \Omega : Bias(y') = Bias(y) - 1\} \end{aligned} \tag{5.22}$$

with probability $\min\{f\left(\frac{1}{2} - bias(y)\right)\left(\frac{1}{2} + bias(y)\right) + \frac{(M_1+1)F}{n}, 1\}$; otherwise Y_{t+1} stays the same: $Y_{t+1} = y$.

Recall that the time to reach an n^γ-almost consensus is the hitting time to the set of states

$$A_\gamma \triangleq \{y \in \Omega : bias(y) > 1/2 - n^{-(1-\gamma)}\},$$

and we use $T_\gamma(z)$ to denote the hitting time to a set of state A_γ.

Lemma 5.1. *The expected n^γ-almost consensus time of the Markov chain $\mathcal{Y}$ is $\max_y \mathbb{E}_{\mathcal{Y}(F)}[T_\gamma(y)] = O(n \log n)$.*

This lemma is very similar to theorem 4.1 and we defer the proof to the full version.

5.2 Monotone Coupling Between $\mathcal{Y}(F)$ And $\mathcal{X}(\mathcal{A}_F)$. To transfer the upper bound of $\mathcal{Y}(F)$ to $\mathcal{X}(\mathcal{A}_F)$, we need to build a "nice" coupling between them which is characterized as follow:

Definition 5.3. (Monotone Coupling) *Let X, Y be two random variables on some partially ordered set $(\Sigma, \geq)$. Then a* monotone coupling *between X and Y is a measure $(\tilde{X}, \tilde{Y})$ on $\Sigma \times \Sigma$ such that*

- *The marginal distributions $\tilde{X}$ and X have the same distribution;*
- *The marginal distributions $\tilde{Y}$ and Y have the same distribution;*
- $\mathbb{P}_{(\tilde{X},\tilde{Y})}[\tilde{X} \geq \tilde{Y}] = 1$.

Note that the function $bias(\cdot)$ induces a natural total order $\leq_{bias}$ of our state space $\Omega = \{0,1\}^n$ such that for $x, y \in \Omega$, $x \leq_{bias} y$ if and only if $bias(x) \leq bias(y)$. We can also define a partial order over sequences of states: given two sequences $(X_t)_{t\geq 0}$, $(Y_t)_{t\geq 0}$ we call $(X_t)_{t\geq 0} \leq_{bias} (Y_t)_{t\geq 0}$ if $\forall t \geq 0\ X_t \leq_{bias} Y_t$. We use calligraphic font to represent the whole random sequence, e.g. $\mathcal{Z} = (Z_t)_{t\geq 0}$.

Lemma 5.2. *There exists a monotone coupling $(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}})$ between $\mathcal{X}(\mathcal{A}_F)$ and $\mathcal{Y}(F)$ under the partial order $\leq_{bias}$*

The proof of this lemma is straightforward, and we defer the proof to the full version.

5.3 Proof of Theorem 5.1

Proof. [theorem 5.1] We call an event A *increasing* if $x \in A$ implies that any $y \geq x$ is also in A. Observe that $A_\gamma := \{y \in \Omega : bias(y) > 1/2 - n^{-(1-\gamma)}\}$ is increasing with respect to $\leq_{bias}$. Therefore given a random sequence $\mathcal{Z} = (Z_t)_{t\geq 0}$

$$\mathbb{P}_{\mathcal{Z}}[T_\gamma(z) > \tau] = \mathbb{P}_{\mathcal{Z}}\left[\max_{t\leq\tau} bias(Z_t) \leq 1/2 - n^{-(1-\gamma)}\right]$$

By lemma 5.2, for fixed $\tau > 0$ and initial configuration $z \in \Omega$:

$$\begin{aligned} &\mathbb{P}_{\mathcal{X}(\mathcal{A}_F)}[T_\gamma(z) > \tau] \\ &= \mathbb{P}_{\mathcal{X}}\left[\max_{t\leq\tau} bias(X_t) \leq 1/2 - n^{-(1-\gamma)}\right] \\ &= \mathbb{P}_{(\tilde{\mathcal{X}},\tilde{\mathcal{Y}})}\left[\max_{t\leq\tau} bias(\tilde{X}_t) \leq 1/2 - n^{-(1-\gamma)}\right] \\ &= \mathbb{P}_{(\tilde{\mathcal{X}},\tilde{\mathcal{Y}})}\left[\max_{t\leq\tau} bias(\tilde{X}_t) \leq 1/2 - n^{-(1-\gamma)}, \tilde{\mathcal{X}} \geq_{bias} \tilde{\mathcal{Y}}\right] \\ &\leq \mathbb{P}_{(\tilde{\mathcal{X}},\tilde{\mathcal{Y}})}\left[\max_{t\leq\tau} bias(\tilde{Y}_t) \leq 1/2 - n^{-(1-\gamma)}\right] \\ &= \mathbb{P}_{\mathcal{Y}(F)}[T_\gamma(z) > \tau]. \end{aligned}$$

On the other hand, applying Markov's inequality

$$\mathbb{P}_{\mathcal{Y}(F)}[T_\gamma(z) > \tau] \leq \frac{E_{\mathcal{Y}(F)}[T_\gamma(z)]}{\tau},$$

and by lemma 5.1, $\mathbb{P}_{\mathcal{Y}(F)}[T_\gamma(z) > \tau]$ can be arbitrary small by taking $\tau = O(n \log n)$ which finishes the proof.

6 Future Work

This work leaves several open questions. The most glaring question is whether we can analyze the maximum consensus time of iterative majority dynamic on Erdös Rényi random graphs. Another question is if we prove the upper bounds of consensus time on *sparse* Erdös Rényi random graphs, or even general expander graphs? On the other hand, can we prove lower bound for the consensus time on these graphs?

To answer these questions, we may need a more refined understanding of "expansiveness" in graph theory, because the conventional notion for expansion/pseudo-randomness of graph can tell us the *average* connection between the nodes in a set and its complement. However for node dynamics, we need to know: given a set of nodes how *every* nodes in the set connects to the set's complement.

References

[1] Mohammed Amin Abdullah and Moez Draief. Global majority consensus by local majority polling on graphs of a given degree sequence. *Discrete Applied Mathematics*, 180:1–10, 2015.

[2] David Aldous et al. Interacting particle systems as stochastic social dynamics. *Bernoulli*, 19(4):1122–1149, 2013.

[3] J Antonovics and P Kareiva. Frequency-dependent selection and competition: empirical approaches. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 319(1196):601–613, 1988.

[4] Rémi Bardenet, Odalric-Ambrym Maillard, et al. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015.

[5] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale, and Luca Trevisan. Stabilizing consensus with many opinions. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 620–635. Society for Industrial and Applied Mathematics, 2016.

[6] Eli Ben-Naim, Laurent Frachebourg, and Paul L Krapivsky. Coarsening and persistence in the voter model. *Physical Review E*, 53(4):3078, 1996.

[7] Itai Benjamini, Siu-On Chan, Ryan ODonnell, Omer Tamuz, and Li-Yang Tan. Convergence, unanimity and disagreement in majority dynamics on unimodular graphs and random graphs. *Stochastic Processes and their Applications*, 126(9):2719–2733, 2016.

[8] Claudio Castellano, Daniele Vilone, and Alessandro Vespignani. Incomplete ordering of the voter model on small-world networks. *EPL (Europhysics Letters)*, 63(1):153, 2003.

[9] Claudio Castellano, Vittorio Loreto, Alain Barrat, Federico Cecconi, and Domenico Parisi. Comparison of voter and glauber ordering dynamics on networks. *Physical review E*, 71(6):066107, 2005.

[10] Fan Chung and Ron Graham. Quasi-random graphs with given degree sequences. *Random Structures & Algorithms*, 32(1):1–19, 2008.

[11] Fan Chung and Mary Radcliffe. On the spectra of general random graphs. *the electronic journal of combinatorics*, 18(1):P215, 2011.

[12] Fan RK Chung. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

[13] Colin Cooper, Robert Elsässer, and Tomasz Radzik. The power of two choices in distributed voting. In *International Colloquium on Automata, Languages, and Programming*, pages 435–446. Springer, 2014.

[14] Colin Cooper, Tomasz Radzik, Nicolás Rivera, and Takeharu Shiraga. Fast plurality consensus in regular expanders. *arXiv preprint arXiv:1605.08403*, 2016.

[15] J Theodore Cox and David Griffeath. Diffusive clustering in the two dimensional voter model. *The Annals of Probability*, pages 347–370, 1986.

[16] James Cruise and Ayalvadi Ganesh. Probabilistic consensus via polling and majority rules. *Queueing Systems*, 78(2):99–120, 2014.

[17] Benjamin Doerr, Leslie Ann Goldberg, Lorenz Minder, Thomas Sauerwald, and Christian Scheideler. Stabilizing consensus with the power of two choices. In *Proceedings of the twenty-third annual ACM symposium on Parallelism in algorithms and architectures*, pages 149–158. ACM, 2011.

[18] Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[19] Andreas Eberle. Markov processes. *Lecture Notes at University of Bonn*, 2009.

[20] Glenn Ellison. Learning, local interaction, and coordination. *Econometrica: Journal of the Econometric Society*, pages 1047–1071, 1993.

[21] Reza Gheissari and Anna Ben Hamou. Aimpl: Markov chain mixing times, available at http://aimpl.org/markovmixing.

[22] Richard A Holley and Thomas M Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663, 1975.

[23] Yashodhan Kanoria, Andrea Montanari, et al. Majority dynamics on trees and the dynamic cavity method. *The Annals of Applied Probability*, 21(5): 1694–1748, 2011.

[24] Motoo Kimura and George H Weiss. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4):561, 1964.

[25] Paul L Krapivsky and Sidney Redner. Dynamics of majority rule in two-state interacting spin systems. *Physical Review Letters*, 90(23):238701, 2003.

[26] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[27] Thomas M Liggett. Coexistence in threshold voter models. *The Annals of Probability*, pages 764–802, 1994.

[28] Thomas M Liggett et al. Stochastic models of interacting systems. *The Annals of Probability*, 25 (1):1–29, 1997.

[29] Jane Molofsky, Richard Durrett, Jonathan Dushoff, David Griffeath, and Simon Levin. Local frequency dependence and global coexistence. *Theoretical population biology*, 55(3):270–282, 1999.

[30] Andrea Montanari and Amin Saberi. Convergence to equilibrium in local interaction games and ising models. *arXiv preprint arXiv:0812.0198*, 2008.

[31] Elchanan Mossel, Joe Neeman, and Omer Tamuz. Majority dynamics and aggregation of information in social networks. *Autonomous Agents and Multi-Agent Systems*, 28(3):408–429, 2014.

[32] F James Rohlf and Gary D Schnell. An investigation of the isolation-by-distance model. *The American Naturalist*, 105(944):295–324, 1971.

[33] Frank Schweitzer and Laxmidhar Behera. Nonlinear voter models: the transition from invasion to coexistence. *The European Physical Journal B-Condensed Matter and Complex Systems*, 67(3): 301–318, 2009.

[34] Vishal Sood and Sidney Redner. Voter model on heterogeneous graphs. *Physical review letters*, 94 (17):178701, 2005.

[35] Krzysztof Suchecki, Victor M Eguiluz, and Maxi San Miguel. Conservation laws for the voter model in complex networks. *EPL (Europhysics Letters)*, 69(2):228, 2005.

[36] Krzysztof Suchecki, Víctor M Eguíluz, and Maxi San Miguel. Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution. *Physical Review E*, 72(3): 036132, 2005.

[37] Omer Tamuz and Ran J Tessler. Majority dynamics and the retention of information. *Israel Journal of Mathematics*, 206(1):483–507, 2015.

A Proofs in Section 4.2

Proof. [lemma 4.1] Define $s : \Omega \mapsto \mathbb{R}$ as follows

$$s(x) = \sum_{\ell=Bias(x)}^{n/2-1} d_\ell, \text{ where } Bias(x) < n/2$$

$$s(x) = 0, \text{ where } Bias(x) = n/2$$

Note that the value of s only depends on the bias of each state and for $x, y \in \Omega$ with $Bias(x) = Bias(y)$ we have $s(x) = s(y)$, so we can abuse the notion and consider potential function with integral domain $s : [0, n/2-1] \mapsto \mathbb{R}$ such that $s(l) \triangleq s(x)$ for some x such that $Bias(x) = l$.

To prove s is a valid super solution of τ, by corollary 2.1 it is sufficient for us to show that

$$\mathcal{L}s(x) \leq -1 \text{ where } Bias(x) < n/2, \quad \text{(A.1)}$$

$$s(x) \geq 0 \text{ where } Bias(x) = n/2. \quad \text{(A.2)}$$

For the Equation (A.1), if $Bias(x) = \ell$ and $0 < \ell < n/2$,

$$\mathcal{L}s(x) = \sum_{y\in\Omega} P_{x,y}s(y) - s(x) = \sum_{y\in\Omega} P_{x,y}\big(s(Bias(y)) - s(Bias(x))\big)$$

By the definition of $\mathcal{M}$, $P(x, y) > 0$ only if the states x, y differ by at most one digit, we only need to consider the states y such that $|Bias(y)-\ell| \leq 1$, by the defintion of s,

$$\begin{aligned}\mathcal{L}s(x) &= \sum_{y:Bias(y)=\ell+1} P_{x,y}\big(s(\ell+1) - s(\ell)\big) \\ &+ \sum_{y:Bias(y)=\ell-1} P_{x,y}\big(s(\ell-1) - s(\ell)\big) \\ &= -\left(\sum_{y:Bias(y)=\ell+1} P_{x,y}\right) d_\ell + \left(\sum_{y:Bias(y)=\ell-1} P_{x,y}\right) d_{\ell-1} \\ &= -\mathbb{P}_{\mathcal{M}}[Bias(X') = \ell+1|X = x]d_\ell \\ &+ \mathbb{P}_{\mathcal{M}}[Bias(X') = \ell-1|X = x]d_{\ell-1}\end{aligned}$$

by the definition of $p^+(x)$ and $p^-(x)$

$$\begin{aligned}\mathcal{L}s(x) &= -p^+(x)d_\ell + p^-(x)d_{\ell-1} \\ &\leq -p^+(x)\left(\frac{1}{p^+(x)} + \frac{p^-(x)}{p^+(x)}d_{\ell-1}\right) + p^-(x)d_{\ell-1} = -1\end{aligned}$$

where the last equality comes from the definition of d_ℓ On the other hand, if $Bias(x) = 0$

$$\begin{aligned}\mathcal{L}s(x) &= \sum_{y:Bias(y)=1} P_{x,y}\big(s(1) - s(0)\big) \\ &= -\mathbb{P}_{\mathcal{M}}[Bias(X') = 1|X = x]d_0 = -p^+(x)d_0 \leq -1.\end{aligned}$$

Equation (A.2) automatically holds by the definition of s.

Therefore, applying corollary 2.1 we have $\max_{x\in\Omega} \tau(x) \leq \max_{x\in\Omega} s(x) = \sum_{\ell=0}^{n/2-1} d_\ell$.

The proof of theorem 4.2, which is rather straightforward but tedious, uses lemma 4.1 and a careful estimation of the potential function from the recursive equation (4.9).

is rather straightforward but tedious by use of lemma 4.1, and estimation of the potential function from the recursive Equation (4.9) carefully.

Proof. [theorem 4.2] With the help of lemma 4.1, we only need to give an upper bound the recursive equations (4.9). With the condition in the statements, suppose we prove the following equations: There exists some positive constant C_1, C_2, C_3, C_4, D_1 such that

$$d_\ell \leq C_1\sqrt{n} \text{ where } \ell < D_1\left\lceil\sqrt{n}\right\rceil, \tag{A.3}$$

$$d_\ell \leq C_2\frac{n}{\ell} \text{ where } D_1\left\lceil\sqrt{n}\right\rceil \leq \ell \leq \check{\epsilon}n, \tag{A.4}$$

$$d_\ell \leq C_3 \text{ where } \check{\epsilon}n < \ell \leq (1/2-\hat{\epsilon})n, \tag{A.5}$$

$$d_\ell \leq C_4\frac{n}{n/2-\ell} \text{ where } (1/2-\hat{\epsilon})n < \ell < n/2 \tag{A.6}$$

Supposing the above inequalities are true, by lemma 4.1 we can complete the proof as follows:

$$\begin{aligned}&\max_{x\in\Omega}\mathbb{E}_{\mathcal{M}}[T(G, f, x)] \leq \sum_{\ell=0}^{n/2-1} d_\ell \\ &\leq \sum_{\ell=0}^{D_1\lceil\sqrt{n}\rceil-1} C_1\sqrt{n} + \sum_{\ell=D_1\lceil\sqrt{n}\rceil}^{\check{\epsilon}n} C_2\frac{n}{\ell} \\ &+ \sum_{\ell=\check{\epsilon}n+1}^{(1/2-\hat{\epsilon})n} C_3 + \sum_{\ell=(1/2-\hat{\epsilon})n+1}^{n/2-1} C_4\frac{n}{n/2-\ell} \\ &\leq D_1\left\lceil\sqrt{n}\right\rceil \cdot C_1\sqrt{n} + C_2 n \sum_{\ell=D_1\lceil\sqrt{n}\rceil}^{\check{\epsilon}n} \frac{1}{\ell} + \\ &C_3(1/2-\hat{\epsilon}-\check{\epsilon})n + C_4 n\sum_{\ell=1}^{\hat{\epsilon}n-1}\frac{1}{\ell} \\ &= O(n\ln n)\end{aligned}$$

Now we are going to use induction to prove Equations (A.3), (A.4), (A.5), and (A.6).

Equation (A.3)**:** We first use induction to prove the following inequality: If $A(n) = \frac{1}{p^+} + \frac{\sqrt{n}}{p^+A_1B_1}$ and $B(n) = \frac{\sqrt{n}}{p^+A_1B_1}$, for all $\ell, 0 \leq \ell \leq D_1\left\lceil\sqrt{n}\right\rceil$

$$d_\ell \leq A(n)\left(1 + \frac{A_1B_1}{\sqrt{n}}\right)^\ell - B(n) \tag{A.7}$$

Because for all constant D_1 there exists some constant $C_1 > 0$ such that $A(n)\left(1 + \frac{A_1B_1}{\sqrt{n}}\right)^\ell - B(n) \leq C_1\sqrt{n}$ for all $\ell \leq D_1\left\lceil\sqrt{n}\right\rceil$, the Equation (A.3) is proven once the Equation (A.7) is true. Now, let's prove (A.7).

For $\ell = 0$, applying Equation (4.11) to Equation (4.9), we have

$$d_0 = \max_{x\in\Omega:Bias(x)=0}\frac{1}{p^+(x)}$$

(because $\{x \in \Omega : Bias(x) = 0\} \subset \Sigma^s \cup \Sigma^m$)

$$\leq \max_{x\in\Sigma^s\cup\Sigma^m}\frac{1}{p^+(x)}$$

(by Equation (4.11)) $\leq 1/p^+$

(by the definition of A and B)

$$= A - B$$

Suppose $d_{\ell-1} \leq A\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^{\ell-1} - B$ for some $1 < \ell < D_1\lceil\sqrt{n}\rceil$. Since $\ell < D_1\lceil\sqrt{n}\rceil - 1 < \check{\epsilon}n$, $\{x \in \Omega : Bias(x) = \ell\} \subset \Sigma^s$ and we can apply equation (4.13) and (4.11) to equation (4.9) and have

$$d_\ell \leq \frac{1}{p^+} + \left(1 + A_1\left(\frac{B_1}{\sqrt{n}} - \ell\right)\right) d_{\ell-1} \leq \frac{1}{p^+} + \left(1 + \frac{A_1B_1}{\sqrt{n}}\right) d_{\ell-1} \tag{A.8}$$

By induction hypothesis, and definition of B we have

$$\begin{aligned}
&\leq \frac{1}{p^+} + \left(1+\frac{A_1B_1}{\sqrt{n}}\right)\left(A\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^{\ell-1} - B\right)\\
&\leq A\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^{\ell} - B - \left(\frac{A_1B_1}{\sqrt{n}}B - \frac{1}{p^+}\right)\\
&\leq A\left(1+\frac{A_1B_1}{\sqrt{n}}\right)^{\ell} - B
\end{aligned}$$

Equation (A.4)**:** We use induction again to prove Equation (A.4) holds for $D_1\lceil\sqrt{n}\rceil \leq \ell \leq \check{\epsilon}n$.

For $\ell = D_1\lceil\sqrt{n}\rceil$, we already have $d_\ell \leq C_1\sqrt{n} \leq C_2\frac{n}{D_1\lceil\sqrt{n}\rceil}$ so if we take $C_2 \geq C_1D_1$

$$d_\ell \leq C_1\sqrt{n} \leq C_2\frac{n}{D_1\lceil\sqrt{n}\rceil} = C_2\frac{n}{\ell}.$$

Suppose $d_{\ell-1} \leq C_2\frac{n}{\ell-1}$ for some $D_1\lceil\sqrt{n}\rceil < \ell < \check{\epsilon}n$. Because $\{x \in \Omega : Bias(x) = \ell\} \subset \Sigma^s$, by equation (A.8) and induction hypothesis we have

$$\begin{aligned}
d_\ell &\leq \frac{1}{p^+} + \left(1 + A_1\left(\frac{B_1}{\sqrt{n}} - \ell\right)\right) d_{\ell-1}\\
&= \frac{1}{p^+} + \left(1 - \left(\frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)\right) d_{\ell-1}
\end{aligned}$$

and

$$\begin{aligned}
d_\ell &\leq \frac{1}{p^+} + \left(1 - \left(\frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)\right) C_2\frac{n}{\ell-1}\\
&= \frac{C_2n}{\ell} + \left(\frac{1}{p^+} + \left(1 - \frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)\frac{C_2n}{\ell-1} - \frac{C_2n}{\ell}\right).
\end{aligned}$$

Therefore equation (A.4) is proven if $\frac{1}{C_2p^+} + \left(1 - \frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)\frac{n}{\ell-1} - \frac{n}{\ell} \leq 0$. By taking $C_2 \geq \frac{2}{p^+A_2}$ and $D_1 \geq 4B_1$ and $D_1^2 \geq 4/A_1$ we have

$$\frac{A_1}{2} \leq A_1 - \frac{A_1B_1}{D_1} - \frac{1}{D_1^2}$$

(because $\ell > D_1\lceil\sqrt{n}\rceil$)

$$\begin{aligned}
&\leq \frac{n}{\ell}\left(\frac{A_1\ell - A_1B_1\sqrt{n}}{n} - \frac{1}{\ell}\right)\\
&\leq \frac{n}{\ell-1}\left(\frac{A_1\ell - A_1B_1\sqrt{n}}{n} - \frac{1}{\ell}\right)\\
&= \frac{n}{\ell} - \frac{n}{\ell-1}\left(1 - \frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)
\end{aligned}$$

Because $C_2 \geq \frac{2}{p^+A_2}$, we have $\frac{1}{C_2p^+} \leq \frac{A_1}{2}$ and using the above inequality we get

$$\frac{1}{C_2p^+} \leq \frac{n}{\ell} - \frac{n}{\ell-1}\left(1 - \frac{A_1\ell - A_1B_1\sqrt{n}}{n}\right)$$

which completes proving Equation (A.4). Finally, by the Equation (A.4)

$$d_{\check{\epsilon}n} \leq C_2\frac{n}{\check{\epsilon}n} = C_2/\check{\epsilon}. \tag{A.9}$$

Equation (A.5)**:** We use induction to prove d_ℓ is bounded above by some constant C_3 for all ℓ such that $\check{\epsilon}n < \ell \leq (1/2 - \hat{\epsilon})n$.

For $\ell = \check{\epsilon}n + 1$, because $\{x \in \Omega : Bias(x) = \check{\epsilon}n + 1\} \subset \Sigma^m$, we can apply (4.19) and (4.16) into Equation (4.9) and have

$$d_\ell \leq \frac{1}{p^+} + (1 - A_2)\, d_{\ell-1} \tag{A.10}$$

(by Equation (A.9))

$$\begin{aligned}
&\leq \frac{1}{p^+} + (1 - A_2)\, C_2/\check{\epsilon}\\
&\leq A_2\frac{1}{p^+A_2} + (1 - A_2)\, C_2/\check{\epsilon}
\end{aligned}$$

Because $0 \leq A_2 < 1$ if we take $C_3 = \max\{\frac{1}{p^+A_2}, C_2/\check{\epsilon}\}$ the base case of (A.3) is true. Suppose $d_{\ell-1} \leq C_3$ for some $\check{\epsilon}n < \ell < (1/2 - \hat{\epsilon})n$, because $\{x \in \Omega : Bias(x) = \ell\} \subset \Sigma^m$ we can use (A.10) and

$$\begin{aligned}
d_\ell &= \frac{1}{p^+} + (1 - A_2)\, d_{\ell-1} \leq \frac{1}{p^+} + (1 - A_2)\, C_3\\
&\leq C_3 - \left(A_2C_3 - \frac{1}{p^+}\right) \leq C_3,
\end{aligned}$$

because $C_3 = \max\{\frac{1}{p^+A_2}C_2/\check{\epsilon}\}$

This finishes the proof of Equation (A.5).

Equation (A.6)**:** Because $\{x \in \Omega : Bias(x) = \ell\} \subset \Sigma^l$ for all $(1/2 - \hat{\epsilon})n < \ell < n/2$ we can apply (4.20) and (4.17) into Equation (A.9), and get

$$d_\ell \leq \frac{4}{1/2 - \ell/n} + (1 - A_3)d_{\ell-1} = \frac{4n}{n/2 - \ell} + (1 - A_3)d_{\ell-1}$$

Recursively applying this relation, d_ℓ is upper bounded by

$$\sum_{j=(1/2-\hat{\epsilon})n+1}^{l} \frac{4n(1-A_3)^{\ell-j}}{n/2-j} + (1-A_3)^{\ell-(1/2-\hat{\epsilon})n} d_{(1/2-\hat{\epsilon})n}$$

because of Equation (A.5) this is at most

$$\leq \sum_{j=(1/2-\hat{\epsilon})n+1}^{l} \frac{4n(1-A_3)^{\ell-j}}{n/2-j} + C_3$$

(taking $i = \ell - j$)

$$\leq 4n \sum_{i=0}^{\ell-(1/2-\hat{\epsilon})n-1} \frac{(1-A_3)^i}{n/2-\ell+i} + C_3$$

$$\leq \frac{4n}{n/2-\ell} \sum_{i=0}^{\ell-(1/2-\hat{\epsilon})n-1} (1-A_3)^i \frac{n/2-\ell}{n/2-\ell+i} + C_3$$

Because $(1-A_3)^i \frac{n/2-\ell}{n/2-\ell+i} \leq (1-A_3)^i$ and taking $C_4 \geq 2/A_3 + \hat{\epsilon} C_3$, d_ℓ is bounded above by

$$\frac{4n}{n/2-\ell} \sum_{i=0}^{\infty} (1-A_3)^i + C_3 = \frac{4}{A_3} \frac{n}{n/2-\ell} + C_3 \leq \frac{C_4 n}{n/2-\ell}$$

B Exertion and Drift: Proof of Lemma 4.2 and 4.3

In this section, we want to control the exertion $p_G^+(x)$ and drift $p_G^+(x) - p_G^-(x)$ of the process $\mathcal{M}$ on graph $G \sim \mathcal{G}_{n,p}$. To achieve these upper bounds, we prove several properties of dense Erdös-Rényi graphs which might seem ad-hoc, but there is a common thread under these lemmas: concentration phenomena in dense Erdös-Rényi graph. Our main tools are the spectral property of random graph and several variants of Chernoff bounds.

B.1 Exertion and Lemma 4.2 We partition the lemma 4.2 into lemma B.1 and B.2, and use the mixing lemma 2.1 to show all configurations have $p_G^+(x)$ close to that of the complete graph if G is a good expander.

Lemma B.1. (Exertion of Σ^s, Σ^m) *If G λ-expander with nearly uniform degree $E(\delta_d)$, $\delta_d < 1$ and $\lambda^2 < \frac{1-\delta_d}{1+\delta_d} \cdot \frac{\hat{\epsilon}}{18}$, for all x with $Bias(x) < 1/2 - \hat{\epsilon}$,*

$$\frac{\hat{\epsilon}}{2} f\left(\frac{\hat{\epsilon}}{2}\right) < p_G^+(x) \leq 1.$$

Proof. Let's consider a fixed configuration x where $\hat{\epsilon} \leq pos(x) < 1/2$ and where the number of red nodes is less than the number of blue nodes'. We can partition the V into three sets of vertices $S_x, T_x, U_x \subset V$ such that

$$S_x = \{v \in V : x(v) = 0, r_x(v) < \frac{\hat{\epsilon}}{2}\}, \tag{B.11}$$

$$T_x = \{v \in V : x(v) = 0, r_x(v) \geq \frac{\hat{\epsilon}}{2}\}, \text{ and} \tag{B.12}$$

$$U_x = \{v \in V : x(v) = 1\}. \tag{B.13}$$

Observe that U_x is the set of red nodes in configuration x, and $S_x \cup T_x$ is the set of blue nodes so $|S_x \cup T_x| = Pos(x) \geq \hat{\epsilon} n$. Moreover by the definition of $\mathcal{M}$ with update function f, the definition in (B.12), and the monotone property of f, the probability a node $v \in T_x$ becomes red in the next step, given v is chosen and the current configuration is x, is greater than $f(\frac{\hat{\epsilon}}{2})$. As a result, every node in T_x has a constant probability to change if chosen, and

$$p_G^+(x) \geq \frac{|T_x|}{n} \cdot f\left(\frac{\hat{\epsilon}}{2}\right)$$

Therefore, if we prove the following inequality

$$|S_x| < \frac{\hat{\epsilon}}{2}|V| \tag{B.14}$$

then the size of set T_x is greater than $\frac{\hat{\epsilon}}{2}|V|$, and we have $p_G^+(x) \geq \frac{\hat{\epsilon}}{2} f\left(\frac{\hat{\epsilon}}{2}\right)$ which finishes the proof.

Now it is sufficient for us to prove equation (B.14). By the definition in (B.11) we can upper bound the number of edges between S_x and U_x, $e(S_x, U_x)$, and use mixing lemma 2.1 to upper bound the size of S_x.

First, since the degree of nodes are nearly uniform, the volume of S_x, and U_x can be bounded

$$(1-\delta_d)np|S_x| \leq vol(S_x) \leq (1+\delta_d)np|S_x| \tag{B.15}$$

$$(1-\delta_d)np|U_x| \leq vol(U_x) \leq (1+\delta_d)np|U_x|, \tag{B.16}$$

and by the definition of S_x in (B.11) the number of edges between S_x and U_x can be bounded as follows:

$$e(S_x, U_x) \leq \frac{\hat{\epsilon}}{2} \cdot vol(S_x) \leq \frac{\hat{\epsilon}}{2} \cdot (1+\delta_d)np|S_x| \tag{B.17}$$

Applying mixing lemma 2.1 on sets S_x and U_x, and we have

$$\left| e(S_x, U_x) - \frac{vol(S_x)vol(U_x)}{vol(G)} \right| \leq \lambda \sqrt{vol(S_x)vol(U_x)}$$

$$\frac{vol(S_x)vol(U_x)}{vol(G)} - e(S_x, U_x) \leq \lambda \sqrt{vol(S_x)vol(U_x)}$$

(by equation (B.17))

$$\frac{vol(S_x)vol(U_x)}{vol(G)} - \frac{\hat{\epsilon}}{2} vol(S_x) \leq \lambda \sqrt{vol(S_x)vol(U_x)}$$

(B.18)

$$\left(\frac{vol(U_x)}{vol(G)} - \frac{\hat{\epsilon}}{2}\right) \sqrt{vol(S_x)} \leq \lambda \sqrt{vol(U_x)}$$

For the left hand side, because the degree of G is near uniform, we can approximate the ratio of $\frac{vol(U_x)}{vol(G)}$ by the ratio of $\frac{|U_x|}{|G|}$ as follows

$$\left(\frac{vol(U_x)}{vol(G)} - \frac{\hat{\epsilon}}{2}\right)\sqrt{vol(S_x)}$$
$$\geq \left(\frac{(1-\delta_d)|U_x|}{(1+\delta_d)|V|} - \frac{\hat{\epsilon}}{2}\right)\sqrt{vol(S_x)}$$

Because $pos(x) < 1/2$, this is

$$\geq \left(\frac{(1-\delta_d)/2}{(1+\delta_d)} - \frac{\hat{\epsilon}}{2}\right)\sqrt{vol(S_x)}$$
$$\geq \frac{1}{2}\left(\frac{1-\delta_d}{1+\delta_d} - \hat{\epsilon}\right)\sqrt{vol(S_x)}$$
$$\geq \frac{1}{3}\sqrt{vol(S_x)}. \tag{B.19}$$

For the right hand side, we can upper bound the volume of U_x by

$$vol(U_x) \leq vol(V) \leq (1+\delta_d)n^2 p. \tag{B.20}$$

Applying equations (B.19) and (B.20) into equation (B.18) yields

$$\frac{1}{9}vol(S_x) \leq \lambda^2 vol(U_x)$$
$$vol(S_x) \leq 9\lambda^2(1+\delta_d)n^2 p$$
$$(1-\delta_d)np|S_x| \leq 9\lambda^2(1+\delta_d)n^2 p$$
$$|S_x| \leq 9\lambda^2\frac{1+\delta_d}{1-\delta_d}n = o(n)$$

which is smaller than $\frac{\hat{\epsilon}}{2}n$ because $\lambda^2 < \frac{1-\delta_d}{1+\delta_d}\cdot\frac{\hat{\epsilon}}{18}$.

Lemma B.2. (Exertion of Σ^l) *If G is a λ-expander with nearly uniform degree $E(\delta_d)$, $\delta_d < 1$ and $\lambda^2 < \frac{1-\delta_d}{1+\delta_d}\cdot\frac{(1/2-\hat{\epsilon})^2}{2}$, for all x with $bias(x) > 1/2 - \hat{\epsilon}$,*

$$\frac{1}{4}(1/2 - bias(x)) < p_G^+(x) \leq (1/2 - bias(x)).$$

Proof. Without lose of generality, we consider the configuration x where $pos(x) < \hat{\epsilon}$.

The proof of the upper bound is straightforward. Suppose $H_v = \{v$ changes from blue to red in the step given v is chosen and the configuration is $x\}$

$$p_G^+(x) = \mathbb{P}_{\mathcal{M}}[Bias(X_1) = Bias(x) + 1 | X_0 = x]$$
$$= \frac{1}{n}\sum_{v\in V}\mathbb{P}_{\mathcal{M}}[H_v]$$
$$\leq \frac{1}{n}\sum_{v\in V}\mathbb{I}[v \text{ is blue}] = pos(x) = (1/2 - bias(x)).$$

For the lower bound, similar to lemma B.1, given a configuration, we partition the set of nodes V into three sets S'_x, T'_x, U'_x

$$S'_x = \{v \in V : x(v) = 1, r_x(v) \geq \frac{1}{2}\}, \tag{B.21}$$
$$T'_x = \{v \in V : x(v) = 1, r_x(v) < \frac{1}{2}\}, \tag{B.22}$$
$$U'_x = \{v \in V : x(v) = 1\} = S'_x \cup T'_x \tag{B.23}$$

To show a lower bound for $p_G^+(x)$, it is sufficient to show that the fraction of red nodes $T'_x \subset V$ is large and has constant probability to change to blue if selected to update. Because the probability that node $v \in T'_x$ becomes blue in the next step given v is chosen with configuration x is $f(1 - r_x(v))$, by the definition in (B.22) and by the monotone property of f

$$f(1 - r_x(v)) \geq f\left(\frac{1}{2}\right) \geq 1/2.$$

Suppose

$$|S'_x| < \frac{1}{2}Pos(x) \tag{B.24}$$

then the size of set T'_x is greater than $\frac{1}{2}Pos(x)$, and we have a lower bound for $p_G^+(x)$: $\frac{1}{2}\frac{|T'_x|}{n} \geq \frac{1}{4}pos(x)$ which finishes the proof

Now it is sufficient for us to prove equation (B.24). By the definition in (B.21) we can upper bound the number of edges between S'_x and U'_x, $e(S'_x, U'_x)$, and use mixing lemma 2.1 to upper bound the size of S'_x.

First, since the degree of nodes are nearly uniform, the volume of S_x, and U_x can be bounded

$$(1-\delta_d)np|S'_x| \leq vol(S'_x) \leq (1+\delta_d)np|S'_x| \tag{B.25}$$
$$(1-\delta_d)np|U'_x| \leq vol(U'_x) \leq (1+\delta_d)np|U'_x|, \tag{B.26}$$

and by the definition of S'_x in (B.11) the number of edges between S'_x and U'_x can be bounded as follows

$$e(S'_x, U'_x) \geq \frac{1}{2}\cdot vol(S'_x) \tag{B.27}$$

Applying mixing lemma 2.1 on sets S'_x and U'_x, we have

$$\left|e(S'_x, U'_x) - \frac{vol(S'_x)vol(U'_x)}{vol(G)}\right| \leq \lambda\sqrt{vol(S'_x)vol(U'_x)}$$
$$e(S'_x, U'_x) \leq \frac{vol(S'_x)vol(U'_x)}{vol(G)} + \lambda\sqrt{vol(S'_x)vol(U'_x)}$$

By equation (B.27)

$$\frac{1}{2}\cdot vol(S'_x) \leq \frac{vol('S_x)vol(U'_x)}{vol(G)} + \lambda\sqrt{vol(S'_x)vol(U'_x)}$$

Reorganizing the last inequality we have,

$$vol(S'_x) \le \left(\frac{\lambda}{\frac{1}{2} - \frac{vol(U'_x)}{vol(G)}} \right)^2 vol(U'_x)$$

Because $\frac{1}{2} - \frac{vol(U'_x)}{vol(G)} = \frac{1}{2} - pos(x) > 1/2 - \hat{\epsilon}$,

$$vol(S'_x) \le \frac{\lambda^2}{(1/2 - \hat{\epsilon})^2} vol(U'_x).$$

Finally by equations (B.25) and (B.26) and taking δ_d small enough

$$|S'_x| \le \frac{\lambda^2}{(1/2 - \hat{\epsilon})^2} \cdot \frac{1+\delta_d}{1-\delta_d} vol(U'_x) < \frac{1}{2} vol(U'_x)$$

The last inequality holds because $\lambda^2 < \frac{1-\delta_d}{1+\delta_d} \cdot \frac{(1/2-\hat{\epsilon})^2}{2}$.

B.2 Drift and Lemma 4.3 In this section, we want to prove lemma 4.3 As discussed in section 4.3, we will prove lower bounds for drift $D_G(x)$ in Σ^s, Σ^m and Σ^l separately, and use the lower bound for $p_G^+(x)$ in lemma 4.2 to prove lemma 4.3.

B.2.1 Drift in Σ^s and Σ^m The high level idea is to use a serial of triangle inequalities: Given a configuration $x \in \Omega$:

1. The drift $D_G(x)$ is close to its expectation $\mathbb{E}_{\mathcal{G}_{n,p}}[D_G(x)]$;
2. The expectation $\mathbb{E}_{\mathcal{G}_{n,p}}[D_G(x)]$ is close to the drift on complete graphs $D_{K_n}(x)$; and
3. The drift on the complete graph $D_{K_n}(x)$ is lower bounded by its $bias(x)$.

The third part is easy because when $pos(x) > 1/2$ the drift $D_{K_n}(x)$ is

(B.28)
$$D_{K_n}(x) = p^+_{K_n}(x) - p^-_{K_n}(x) = f(pos(x)) - pos(x)$$

and equations (4.18), (4.19), and (4.20) can be obtained by the definition of f.

Therefore, our strategy for states in Σ^s, Σ^m is to argue the value of $\{D_G(x)\}_{x\in\Omega}$ is close to $\{D_{K_n}(x)\}_{x\in\Omega}$ with high probability. The first part is proved in lemmas B.3 and B.4, and the second part is fulfilled in lemma B.5. Informally lemma B.3 shows $\mathbb{E}_{\mathcal{G}_{n,p}}[D_G(x)] - D_{K_n}(x) = O(1/n)$, and lemma B.4 shows $\mathbb{E}_{\mathcal{G}_{n,p}}[D_G(x)] - D_{K_n}(x) = O(\sqrt{(\log n)/n})$.

Before digging into the lemmas, let's rewrite $D_G(x)$.

Without lost of generality if $pos(x) > 1/2$,

$$\begin{aligned} D_G(x) &= \mathbb{E}_{\mathcal{M}}[Pos(X')|X = x] - Pos(x) \\ &= \frac{1}{n} \sum_{v \in V} \mathbb{P}_{\mathcal{M}}[X'(v) = 1 | v \text{ is chosen}, X = x] - pos(x) \\ &= \frac{1}{n} \sum_{v \in V} f(r_x(v)) - pos(x), \end{aligned}$$

and by symmetry of $\mathcal{G}_{n,p}$ we can fixed arbitrary node $v \in V$ and have

$$\mathbb{E}_{\mathcal{G}_{n,p}}[D_G(x)] = \mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))] - pos(x). \tag{B.29}$$

Lemma B.3. (Expected Drift in Σ^s) *If $x \in \Sigma^s$ where $bias(x) < \check{\epsilon}$ then there exists constant $K_1 > 0$ such that for large enough n,*

$$\mathbb{E}_{\mathcal{G}_{n,p}}[D_G(x)] \ge f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right) - \frac{K_1}{n}.$$

Lemma B.4. (Expected Drift in Σ^m) *If $x \in \Sigma^m$ where $\check{\epsilon} \le bias(x) \le 1/2 - \hat{\epsilon}$ then there exists constant $K_2 > 0$ such that for large enough n, $\mathbb{E}_{\mathcal{G}_{n,p}}[D_G(x)]$ is greater than*

$$f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right) - K_2\sqrt{\frac{\log n}{n}}.$$

Lemma B.5. (Small noise in Σ^s and Σ^m) *For all $x \in \Sigma^s \cup \Sigma^m$, $D_G(x)$ there exists a constant $L > 0$ such that when n large enough*

$$D_G(x) - \mathbb{E}_{\mathcal{G}_{n,p}}[D_G(x)] > -\frac{L}{\sqrt{n}}$$

happens with high probability over the randomness of $\mathcal{G}_{n,p}$.

Properties of $r_x(v)$. Due to equation (B.29), to prove lemma B.3 and B.4, it is sufficient for us to analyze $\{\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))]\}_{x\in\Omega}$ is close to $\{f(pos(x))\}_{x\in\Omega}$ for some fixed node $v \in V$. We use the principal of deferred decisions—*We reveal the randomness of the graph $G \sim \mathcal{G}_{n,p}$ after fixing node v and configuration x*, and apply a union bound over all configurations $x \in \Omega$.

Fixing a configuration and node v, let's consider a bin with $Pos(x)$ red balls and $n - Pos(x)$ blue balls, if we sample k balls without replacement, the expected number of red balls among those k ball is $pos(x) \cdot k$, and this random number has the same distribution as the random variable $r_x(v) \cdot k$ if $G \sim \mathcal{G}_{n,p}$ is conditioned on the degree of v being k.

We define $E_x(\delta_r; v)$ to be the event

$$E_x(\delta_r; v) \triangleq \{G : |r_x(v) - pos(x)| \leq \delta_r pos(x)\}. \tag{B.30}$$

Since $r_x(v) \cdot k$ can be seen as a sample without replacement, a standard argument combining theorems 2.3 and 2.2 upper bounds the probability of it deviating from expectation by the one of sampling with replacement.

$$\mathbb{P}_{\mathcal{G}_{n,p}}[E_x(\delta_r; v) | deg(v) = k] \leq 2\exp\left(-\frac{\delta_r^2 k\, pos(x)}{3}\right) \tag{B.31}$$

Proof of the Lemmas As discussed below equation (B.29), we want to prove the difference between $\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))]$ and $f(pos(x))$ is of order $O(1/n)$. However, in contrast to the $O(\sqrt{(\log n)/n})$ error in lemma B.4 we need a smoothness property of f around $1/2$ to derived this stronger result. The following two lemmas prove some basic results about smooth functions and conditional variance.

LEMMA B.6. *Given $I \subseteq \mathbb{R}$, and X is a random variable with support in I and expectation $\mathbb{E}X$, if $g : \mathbb{R} \mapsto \mathbb{R}$ is M_2-smooth in I, then*

$$\left|\mathbb{E}[g(X)] - g(\mathbb{E}X)\right| \leq \frac{M_2}{2}\left(\mathbb{E}X^2 - (\mathbb{E}X)^2\right).$$

LEMMA B.7. *Given a real-valued random variable X and $\epsilon > 0$ such that $\mathbb{P}[\mathbb{E}X - \epsilon \leq X \leq \mathbb{E}X + \epsilon)] > 0$, we have*

$$\mathrm{Var}[X | \mathbb{E}X - \epsilon \leq X \leq \mathbb{E}X + \epsilon] \leq \mathrm{Var}[X]$$

Proof. [lemma B.6] Let $h(t) \triangleq g(\mathbb{E}X + t(X - \mathbb{E}X))$. Because g is smooth, we use the fundamental theorem of Calculus, and have

$$\mathbb{E}[g(X)] - g(\mathbb{E}X) = \mathbb{E}_X[g(X) - g(\mathbb{E}X)] = \mathbb{E}\left[\int_0^1 h'(t)dt\right]$$
$$= \mathbb{E}\left[\int_0^1 g'\big(\mathbb{E}X + t(X - \mathbb{E}X)\big)(X - \mathbb{E}X)dt\right]$$

Because g is M_2-smooth, we have for all $a, a+b \in I$ $g'(a) - M_2|b| \leq g'(a+b) \leq g'(a) + M_2|b|$ and by taking $a = \mathbb{E}X$ and $b = t(X - \mathbb{E}X)$

$$\begin{aligned}
&\mathbb{E}_X[g(X)] - g(\mathbb{E}X) \\
&\leq \mathbb{E}_X\left[\int_0^1 \left(g'(\mathbb{E}X) + M_2 t(X - \mathbb{E}X)\right)(X - \mathbb{E}X)dt\right] \\
&= \mathbb{E}_X\left[g'(\mathbb{E}X)(X - \mathbb{E}X) + \frac{M_2}{2}(X - \mathbb{E}X)^2\right] \\
&= \frac{M_2}{2}\mathbb{E}_X\left[(X - \mathbb{E}X)^2\right]
\end{aligned}$$

The lower bound $-\frac{M_2}{2}\mathbb{E}_X\left[(X - \mathbb{E}X)^2\right] \leq \mathbb{E}_X[g(X)] - g(\mathbb{E}X)$ is can be derived similarity.

Proof. [lemma B.7] Let A the event that X is in the interval $[\mathbb{E}X - \epsilon, \mathbb{E}X + \epsilon]$

$$\begin{aligned}
&\mathrm{Var}[X | (1-\delta_r)\mathbb{E}X \leq X \leq (1+\delta_r)\mathbb{E}X] && \text{(B.32)}\\
&= \mathrm{Var}[X|A] \\
&= \mathbb{E}\left[\big(X - \mathbb{E}[X|A]\big)^2 | A\right] \\
&\leq \mathbb{E}\left[\big(X - \mathbb{E}[X]\big)^2 | A\right] && \text{(B.33)}
\end{aligned}$$

The last inequality is true, because for all z, $\mathbb{E}\left[(Z - z)^2\right] \geq \mathbb{E}\left[(Z - \mathbb{E}Z)^2\right]$.

On the other hand, $\mathrm{Var}[X]$ is equal to

$$\mathbb{E}\left[(X - \mathbb{E}[X])^2 \middle| A\right]\Pr[A] + \mathbb{E}\left[(X - \mathbb{E}[X])^2 \middle| \neg A\right](1 - \Pr[A])$$

Because $|X - \mathbb{E}[X]| \geq \epsilon$ conditioned on $\neg A$ and $\epsilon \geq |X - \mathbb{E}[X]|$ if A happens,

$$\mathbb{E}\left[(X - \mathbb{E}[X])^2 \middle| A\right] \leq \mathrm{Var}[X] \tag{B.34}$$

The proof is completed by combining (B.33), and (B.34).

Proof. [lemma B.3] Following equation (B.29), our goal is to derive a better approximation of

$$|\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))] - f(pos(x))|$$

We take $\check{\epsilon}$ small enough so that $[1/2 - 2\check{\epsilon}, 1/2 + 2\check{\epsilon}] \subseteq I_{1/2}$, and so by the definition the update function f is $\check{M}_2$-smooth function in $[1/2 - 2\check{\epsilon}, 1/2 + 2\check{\epsilon}]$. Moreover we take constants δ_r, δ_d to that $\delta_r \leq \check{\epsilon}$ and $\delta_d < 1$.

Let $\mathcal{E}$ be the event $E_x(\delta_r; v) \wedge E(\delta_d; v)$ defined in equation (B.30) and lemma 2.2 respectively. Informally, if $\mathcal{E}$ happens that means the value of $r_x(v)$ is close to its expectation $pos(x)$ and the degree is nearly uniform. Therefore we can decompose $\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))] - f(pos(x))$ as follows

$$\begin{aligned}
&|\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))] - f(pos(x))| \\
&\leq |\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))|\mathcal{E}] - f(pos(x))| + \mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}] \\
&\leq \left|\mathbb{E}_{\mathcal{G}_{n,p}}\left[f(r_x(v))|\mathcal{E}\right] - f\big(\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|\mathcal{E}]\big)\right| \\
&\quad + \left|f\big(\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|\mathcal{E}]\big) - f(pos(x))\right| + \mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}]
\end{aligned}$$

Now we want to give upper bounds for these three terms.

For the first term, $|\mathbb{E}_{\mathcal{G}_{n,p}}\left[f(r_x(v))|\mathcal{E}\right] - f\big(\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|\mathcal{E}]\big)|$ only depends on the random variable $r_x(v)|\mathcal{E}$. By the definition of $\mathcal{E}$ the random variable $r_x(v)|\mathcal{E}$ has support in $[(1-\delta_r)pos(x), (1+\delta_r)pos(x)]$ and $pos(x) \in [1/2 - \check{\epsilon}, 1/2 + \check{\epsilon}]$. Therefore the support of $r_x(v)|\mathcal{E}$ is in $I_{1/2}$ and

$$\begin{aligned}
&\left|\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))|E] - f\left(\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|\mathcal{E}]\right)\right| \\
&\leq \frac{\check{M}_2}{2}\mathrm{Var}_{\mathcal{G}_{n,p}}[r_x(v)|\mathcal{E}] \\
&\leq \frac{\check{M}_2}{2}\mathrm{Var}_{\mathcal{G}_{n,p}}[r_x(v)|E(\delta_d)].
\end{aligned}$$

The first inequality is true because f is smooth in $I_{1/2}$ and lemma B.6. The second comes from lemma B.7. Now we want to upper bound $\mathrm{Var}[r_x(v)|E(\delta_d)]$. Recall that we observed that the random variable $k \cdot r_x(v)|deg(v) = k$ can be seen as sampling balls from bin with a $pos(x)$ fraction of red balls without replacement. Because the variance is a convex function by theorem 2.2, the value of $\mathrm{Var}[r_x(v)|deg(v) = k]$ is upper bounded by the variance of sampling k balls from the same bin with replacement, $\frac{pos(x)(1-pos(x))}{k}$. As a result,

$$\mathrm{Var}_{\mathcal{G}_{n,p}}[r_x(v)|E(\delta_d)] \leq \frac{pos(x)(1-pos(x))}{(1-\delta_d)np}.$$

Because $x \in \Sigma^s$, $1/2 - \check{\epsilon} < pos(x) < 1/2 + \check{\epsilon}$, and δ_d is some constant independent of n

(B.35)
$$|\mathbb{E}_{\mathcal{G}_{n,p}}\left[f(r_x(v))|\mathcal{E}\right] - f\left(\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|\mathcal{E}]\right)| = \frac{1/4 - \check{\epsilon}^2}{(1-\delta_d)p} \cdot \frac{1}{n}.$$

For the second term, because the update function f is Lipschitz, it is sufficient to prove an upper bound for $|\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|E] - pos(x)|$. Note that in the properties of $r_x(v)$ we show that $\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)] = pos(x)$. By the law of total probability we have, $|\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)] - \mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|E]| \leq |\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|\neg E] - \mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|\mathcal{E}]|\mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}]$ which is less than $\leq 2\mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}]$ because $0 \leq r_x(v) \leq 1$. Therefore we have

$$|\mathbb{E}_{\mathcal{G}_{n,p}}[r_x(v)|\mathcal{E}] - pos(x)| \leq 2\mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}]. \tag{B.36}$$

For the last term, $\mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}]$ we just use a union bound:

$$\mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}] = \mathbb{P}_{\mathcal{G}_{n,p}}[\neg E_x(\delta_r; v) \cup \neg\mathcal{E}(\delta_d; v)]$$
$$\leq \mathbb{P}_{\mathcal{G}_{n,p}}[\neg E_x(\delta_r)|E(\delta_d; v)] + \mathbb{P}_{\mathcal{G}_{n,p}}[\neg E(\delta_d; v)]$$

(B.37)
$$\leq 2\exp\left(-\frac{1}{3}\delta_r^2(1-\delta_d)np \cdot pos(x)\right) + 2\exp\left(\frac{-\delta_d^2 np}{3}\right).$$

Equation (B.37) is derived from equation 2.2 and equation (B.31). Because $p > 0$ and $pos(x) \geq 1/2 - \check{\epsilon}$ are constants when $x \in \Sigma^s$ for large enough n we have

$$\mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}] \leq \frac{1}{n}. \tag{B.38}$$

Recall that δ_d, δ_r are constants independent of n. Combining equations (B.35), (B.36), and (B.38), we finish the proof with $K_1 = \frac{1/4 - \check{\epsilon}^2}{(1-\delta_d)p} + 3$.

For lemma B.4, we want to prove the difference between $\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))]$ and $f(pos(x))$ is of order $O(\sqrt{\log n/n})$ which is much weaker than lemma B.3, and we only need to use the Lipschitz properties of update function f, and concentration phenomenon for $r_x(v)$ shown in equation (B.31).

Proof. [lemma B.4] Let $\mathcal{E}$ be the event of $E_x(\delta_r; v) \wedge E(\delta_d; v)$ defined in equation (B.30) and lemma 2.2 respectively. Informally, if $\mathcal{E}$ happens that means the value of $r_x(v)$ is close to expectation $pos(x)$ and the degree is nearly uniform. Therefore we can decompose $\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))] - f(pos(x))$ as follows

$$\begin{aligned} &|\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))] - f(pos(x))| \\ &\leq |\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))|\mathcal{E}] - f(pos(x))| + \mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}] \end{aligned} \tag{B.39}$$

The first term $|\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))|\mathcal{E}] - f(pos(x))|$ Since the update function f is Lipschitz with Lipschitz constant M_1, if the event $\mathcal{E}$ happens $|r_x(v) - pos(x)| \leq \delta_r pos(x)$ and,

$$\begin{aligned} &|f(r_x(v)) - f(pos(x))| \leq M_1 \cdot |r_x(v) - pos(x)| \\ &\leq M_1 \cdot \delta_r pos(x) \leq M_1\delta_r \end{aligned}$$

By taking $\delta_r = A\sqrt{\log n/n}$ for some constant A which will be specified later we have

$$|\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))|\mathcal{E}] - f(pos(x))| = M_1 A\sqrt{\frac{\log n}{n}} \tag{B.40}$$

For the second term by equation (B.37), $\mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}]$ is smaller than

$$2\exp\left(-\frac{\delta_r^2(1-\delta_d)np \cdot pos(x)}{3}\right) + 2\exp\left(\frac{-\delta_d^2 np}{3}\right)$$

because $pos(x) \geq \hat{\epsilon} = \Omega(1)$ when $x \in \Sigma^m$. If δ_d is some small constant and $\delta_r = A\sqrt{\log n/n}$, then by taking A large enough $\mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}]$ is smaller than $2\exp\left(-\frac{A^2(1-\delta_d)p \cdot pos(x)}{3}\log n\right) + 2\exp\left(\frac{-\delta_d^2 np}{3}\right)$ Therefore

$$\mathbb{P}_{\mathcal{G}_{n,p}}[\neg\mathcal{E}] \leq \sqrt{\frac{\log n}{n}} \tag{B.41}$$

Combining equation (B.40) and (B.41) into equation (B.39), and have

$$|\mathbb{E}_{\mathcal{G}_{n,p}}[f(r_x(v))] - f(pos(x))| \leq (M_1 A + 1)\sqrt{\frac{\log n}{n}}$$

and the proof is completed by taking $K_2 = (M_1 A + 1)$.

Proof. [lemma B.5] Given a fixed configuration $x \in A$, random variable $D_G(x)$ has expectation $D(x)$ with randomness over $\mathcal{G}_{n,p}$. Assuming the following claim which we will later prove:

Claim B.1. *If δ_d is some fixed constant, there exists some constant $K > 0$ such that for all $t > 1/\sqrt{n}$, then*

$$\mathbb{P}_{\mathcal{G}_{n,p}}[D_G(x) - \mathbb{E}_{\mathcal{G}_{n,p}} D_G(x) < -Kt|E(\delta_d)] \leq \exp\left(-n^2 t^2\right)$$

By taking $t = \sqrt{2 \ln 2 / n}$

(B.42)
$$\mathbb{P}_{\mathcal{G}_{n,p}}\left[D_G(x) - \mathbb{E}_{\mathcal{G}_{n,p}} D_G(x) < -\frac{K\sqrt{2\ln 2}}{\sqrt{n}} | E(\delta_d)\right] \geq 1 - \exp(-2n\ln 2) = 1 - \frac{1}{4^n}.$$

Apply a union bound over all configurations $x \in \Omega = \{0,1\}^n$ we will derived a high probability result with $L = K\sqrt{2\ln 2}$

$$\mathbb{P}_{\mathcal{G}_{n,p}}\left[\forall x, D_G(x) - \mathbb{E}_{\mathcal{G}_{n,p}} D_G(x) \geq -\frac{K\sqrt{2\ln 2}}{\sqrt{n}}\right]$$
$$\geq \mathbb{P}_{\mathcal{G}_{n,p}}\left[\forall x, D_G(x) - \mathbb{E}_{\mathcal{G}_{n,p}} D_G(x) \geq -\frac{K\sqrt{2\ln 2}}{\sqrt{n}} | E(\delta_d)\right] - \mathbb{P}_{\mathcal{G}_{n,p}}[\neg E(\delta_d)].$$

By union bound, it is greater than

$$1 - 2^n \mathbb{P}_{\mathcal{G}_{n,p}}\left[D_G(x) - \mathbb{E}_{\mathcal{G}_{n,p}} D_G(x) \geq -\frac{K\sqrt{2\ln 2}}{\sqrt{n}} | E(\delta_d)\right] - \mathbb{P}_{\mathcal{G}_{n,p}}[\neg E(\delta_d)].$$

By equation (B.42), this is lower bounded by

$$1 - 2^n \cdot 4^{-n} - \mathbb{P}_{\mathcal{G}_{n,p}}[\neg E(\delta_d)] = 1 - o(1).$$

Therefore, it is sufficient for us to prove claim B.1.

Following the analysis in equation (B.29), if $pos(x) > 1/2$,

$$D_G(x) = \frac{1}{n}\sum_{v \in V} f(r_x(v)) - pos(x)$$

Now we think of $\{f(r_{G,x}(v))\}_{x\in\Omega, v\in V}$ as a set real-valued functions with input G indexed by x and v. Similarly we think $\{D_G(x)\}_{x\in\Omega}$ as a set of real-valued functions with input G. We will apply theorem 2.4 with event $E(\delta_d)$ to prove claim B.1 which consists of two parts: showing the maximum effect/Lipschitz constant c_i is small, and showing the event $E(\delta_d)$ happens with high probability so that it does not change the expectation too much.

For the first part, recall that the update function f is M_1-Lipschitz. Because given x, v if the degree of the node v is k then adding/removing a single edge in G changes the value of $r_{G,x}(v)$ by at most $1/k$, $r_{G,x}(v)$ is $1/k$-Lipschitz. Therefore the Lipschitz constants of $\{f(r_{G,x}(v))\}_{x\in\Omega, v\in V}$ are uniformly bounded by $O(M_1/k) = O(1/k)$. Moreover fixing x if every node have degree at least k, adding/removing a single edge in G only affects two endpoints, and changes the value of $\frac{1}{n}\sum_{v\in V} f(r_x(v))$ by at most $O(\frac{1}{nk})$.

As a result, if $E(\delta_d)$ happens, every node has nearly uniform degree with constant δ_d. For all G, G' in $E(\delta_d)$ that differ in just the presence of a single edge e, we can take $c_e = \max_{G,G'} |D_G(x) - D_{G'}(x)|$ and

(B.43)
$$c_e = O\left(\frac{1}{n \min_{v\in V} deg(v)}\right) = O\left(\frac{1}{n^2}\right)$$

Therefore, there exists some constant $\xi > 0$ such that $\sum_e c_e^2 = \xi/n^2$ and $0 \leq D_G(x) \leq 1$, so we can apply theorem 2.4 and

$$\mathbb{P}_{\mathcal{G}_{n,p}}\left[D_G(x) - \mathbb{E}_{\mathcal{G}_{n,p}} D_G(x) < -t' - \mathbb{P}_{\mathcal{G}_{n,p}}[\neg E(\delta_d)] | E(\delta_d)\right] \leq \exp\left(-\frac{2}{\xi} n^2 t'^2\right)$$

Note that by equation (2.3) when δ_d is some fixed constant and n is large enough $\mathbb{P}_{\mathcal{G}_{n,p}}[\neg E(\delta_d)] \leq 1/\sqrt{n}$, and we finish the proof of equation (B.1) by taking $K \geq \sqrt{\xi/2} + 1$ and $t \geq 1/\sqrt{n}$.

B.2.2 Drift in Σ^l Here we consider the phase of the process when the fraction of red nodes is almost 1. The laziness $1/p_G^+(x)$ should be roughly the inverse of the fraction of blue nodes and increases as the bias increases. As a result to prove equation (4.20) we need to give a better lower bound for the drift $D_G(x)$.

LEMMA B.8. *There exists small enough constants $\delta_d > 0$, $\hat{\epsilon} > 0$, and $K_3 > 0$. If G has nearly uniform degree, $E(\delta_d)$, such that $D_G(x) \geq K_3(1/2 - bias(x))$ for all $x \in \Sigma^l$.*

The following proof is basically a counting argument: when $x \in \Sigma^l$ the number of red nodes is so small for any node to ahve a majority of red neighbors.

Proof. Without lose of generality, we only consider configurations x where $pos(x) < \hat{\epsilon}$ and $pos(x) = 1/2 - bias(x)$. Given p, δ_d we can take $\hat{\epsilon}$ small enough such that $\frac{\hat{\epsilon}}{(1-\delta_d)p} \in I_0$. Because there are at most $\hat{\epsilon}n$ red nodes and for all $v \in V$ $deg(v) \geq (1-\delta_d)np$, we have $r_x(v) \in I_0$ and by the property of update function

(B.44)
$$f(r_x(v)) \leq \hat{M}_1 \cdot r_x(v) < r_x(v)$$

If we define $R_x = \{u \in V : x(u) = 1\}$ to be the set of red nodes, by similarly to equation (B.29)[4] we have

$$p_G^+(x) - p_G^-(x) = pos(x) - \frac{1}{n}\sum_{v\in V} f(r_x(v)).$$

[4] In contrast to equation (B.29) where $pos(x) > 1/2$, here $pos(x) < 1/2$

By the equation (B.44), this is greater than

$$\begin{aligned}&\geq pos(x) - \frac{1}{n}\sum_{v\in V}\hat{M}_1 r_x(v)\\&= pos(x) - \frac{\hat{M}_1}{n}\sum_{v\in V}\frac{e(S_x,v)}{deg(v)}\\&\geq pos(x) - \frac{\hat{M}_1}{n}\frac{e(S_x,V)}{\min_{v\in V} deg(v)}.\end{aligned}$$

The last is true because $deg(v) \geq \min_{v\in V} deg(v)$ and $\sum_v e(S_x,v) = e(S_x,V)$. Because $\sum_v e(S_x,v) \leq |S_x|\max_{u\in S_x} deg(u)$, and $\frac{|S_x|}{n} = pos(x)$,

$$\begin{aligned}p_G^+(x) - p_G^-(x) &\geq \left(1 - \frac{\hat{M}_1 \max_{u\in S_x} deg(u)}{\min_{v\in V} deg(v)}\right) pos(x)\\&\geq \left(1 - \frac{1+\delta_d}{1-\delta_d}\hat{M}_1\right) pos(x)\\&> K_3 pos(x) = K_3(1/2 - bias(x))\end{aligned}$$

The last inequality is true by taking δ_d small enough and $0 < K_3 \leq 1 - \frac{1+\delta_d}{1-\delta_d}\hat{M}_1$.

B.3 Proof of Lemma 4.3

Proof. We prove each equation in turn.

Equation (4.18). First, for drift $D_G(x) = p_G^+(x) - p_G^-(x)$ we apply the idea illustrated at the beginning of section B.2.1. For the first and second steps, we have

$$\begin{aligned}&p_G^+(x) - p_G^-(x) - \left(f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right)\right)\\&= p_G^+(x) - p_G^-(x) - (p^+(x) - p^-(x)) + (p^+(x) - p^-(x))\\&- \left(f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right)\right)\end{aligned}$$

By lemma B.3 and B.5, with high probability, this is greater than

$$\geq -\frac{K_1}{n} - \frac{L}{\sqrt{n}} \geq -\frac{K_1 + L}{\sqrt{n}}$$

For the last step, because the update function f satisfies $f'(1/2) \geq \check{M}_1 > 1$, we can take $\check{\epsilon}$ small enough such that for all h such that $0 \leq h < \check{\epsilon}$,

$$f\left(\frac{1}{2} + h\right) - f\left(\frac{1}{2}\right) \geq \frac{\check{M}_1 + 1}{2}h$$

As a result, with high probability we have for all $x \in \Sigma^s$ where $bias(x) < \check{\epsilon}$

$$p_G^-(x) - p_G^+(x) \leq -\frac{\check{M}_1 - 1}{2}bias(x) + \frac{K_1 + L}{\sqrt{n}} \tag{B.45}$$

On the other hand, by equation (4.16), we have

$$1 \leq \frac{1}{p_G^+(x)} < \frac{1}{\frac{\hat{\epsilon}}{2}f\left(\frac{\hat{\epsilon}}{2}\right)} \tag{B.46}$$

Multiplying equation (B.45) by equation (B.46) we have,

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 + -\frac{\check{M}_1 - 1}{\hat{\epsilon}f\left(\frac{\hat{\epsilon}}{2}\right)}bias(x) + \frac{K_1 + L}{\frac{\hat{\epsilon}}{2}f\left(\frac{\hat{\epsilon}}{2}\right)}\frac{1}{\sqrt{n}}$$

which finishes the proof of equation (4.18) by taking $A_1 = \frac{\check{M}_1 - 1}{\hat{\epsilon}f\left(\frac{\hat{\epsilon}}{2}\right)}$, and $B_1 = \frac{2(K_1+L)}{\check{M}_1 - 1}$ which are positive constants.

For equation (4.19). First, for drift $D_G(x) = p_G^+(x) - p_G^-(x)$ using argument similar to the proof of (4.18), we have with high probability by lemmas B.4 and B.5, for all $x \in \Sigma^m$, $p_G^+(x) - p_G^-(x)$ is greater than

$$f\left(\frac{1}{2} + bias(x)\right) - \left(\frac{1}{2} + bias(x)\right) - \frac{K_2}{\sqrt{n}} - \frac{L}{n}. \tag{B.47}$$

Recalled that the update function f is Lipschitz and $\forall 0 < h < 1/2, f(1/2 + h) > 1/2 + h$, we can define its minimum over a compact set $[1/2 + \check{\epsilon}, 1 - \hat{\epsilon}]$

$$0 < \delta_f \triangleq \min_{\check{\epsilon}\leq h\leq 1/2-\hat{\epsilon}} f\left(\frac{1}{2} + h\right) - \left(\frac{1}{2} + h\right) \tag{B.48}$$

Combining equations (B.47), and (B.48), for large enough n we have with high probability for all $x \in \Sigma^m$

$$p_G^-(x) - p_G^+(x) \leq -\delta_f + \frac{K_2}{\sqrt{n}} + \frac{L}{n} \leq -\frac{\delta_f}{2} \tag{B.49}$$

Multiplying equation (B.49) by equation (B.46) we have,

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 - \frac{\delta_f}{\hat{\epsilon}f\left(\frac{\hat{\epsilon}}{2}\right)}$$

which finishes the proof of equation (4.19) by taking $A_2 = \frac{\delta_f}{\hat{\epsilon}f\left(\frac{\hat{\epsilon}}{2}\right)}$ and $0 < A_2 < 1$.

For equation (4.20), by lemmas B.8 and 4.2, we have $p_G^-(x) - p_G^+(x) \leq -K_3(1/2 - bias(x))$, and $\frac{1}{4}(1/2 - bias(x)) \leq p_G^+(x)$. Therefore

$$\frac{p_G^-(x)}{p_G^+(x)} \leq 1 - 4K_3$$

This finishes the proof by taking $A_3 = 4K_3$.