A Privacy Preserving Randomized Gossip Algorithm via Controlled Noise Insertion

Filip Hanzely[a] **Jakub Konečný**[b] **Nicolas Loizou**[c,*]

Peter Richtárik[a,c,d] **Dmitry Grishchenko**[e]

[a] *King Abdullah University of Science and Technology, KSA*
[b] *Google, USA*
[c] *University of Edinburgh, UK*
[d] *Moscow Institute of Physics and Technology, Russia*
[e] *Université Grenoble Alpes, France*

Abstract

In this work[2] we present a randomized gossip algorithm for solving the average consensus problem while at the same time protecting the information about the initial private values stored at the nodes. We give iteration complexity bounds for the method and perform extensive numerical experiments.

1 Introduction

In this paper we consider the average consensus (AC) problem. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected connected network with node set $\mathcal{V} = \{1, 2, \dots, n\}$ and edges $\mathcal{E}$ such that $|\mathcal{E}| = m$. Each node $i \in \mathcal{V}$ "knows" a private value $c_i \in \mathbb{R}$. The goal of AC is for every node of the network to compute the average of these values, $\bar{c} \overset{\text{def}}{=} \frac{1}{n}\sum_i c_i$, in a distributed fashion. That is, the exchange of information can only occur between connected nodes (neighbours).

The literature on distributed protocols for solving the average consensus problem is vast and has long history [18, 19, 1, 8]. In this work we focus on one of the most popular class of methods for solving the average consensus, the randomized gossip algorithms and propose a gossip algorithm for protecting the information of the initial values c_i, in the case when these may be sensitive. In particular, we develop and analyze a privacy preserving variant of the randomized pairwise gossip algorithm ("randomly pick an edge $(i, j) \in \mathcal{E}$ and then replace the values stored at vertices i and j by their average") first proposed in [2] for solving the average consensus problem. While we shall not formalize the notion of privacy preservation in this work, it will be intuitively clear that our methods indeed make it harder for nodes to infer information about the private values of other nodes, *which might be useful in practice.*

1.1 Related Work on Privacy Preserving Average Consensus

The introduction of notions of privacy within the AC problem is relatively recent in the literature, and the existing works consider two different ideas. In [7], the concept of differential privacy [3] is used to protect the output value $\bar{c}$ computed by all nodes. In this work, an exponentially decaying Laplacian noise is added to the consensus computation. This notion of privacy refers to protection of the *final average*, and formal guarantees are provided. A different line of work with a more stricter

*Corresponding author (n.loizou@sms.ed.ac.uk)

[2]The full-length paper, which includes a number of additional algorithms and results, is available in [6].

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

goal is the design of privacy-preserving average consensus protocols that guarantee protection of the *initial values* c_i of the nodes [16, 14, 15]. In this setting each node should be unable to infer a lot about the initial values c_i of any other node. In the existing works, this is mainly achieved with the clever addition of noise through the iterative procedure that guarantees preservation of privacy and at the same time converges to the exact average. We shall however mention, that none of these works address any specific notion of privacy (no clear measure of privacy is presented) and it is still not clear how the formal concept of differential privacy [3] can be applied in this setting.

1.2 Main Contributions

In this work, we present the first randomized gossip algorithm for solving the Average Consensus problem while at the same time protecting the information about the initial values. To the best of our knowledge, this work is the first which combines the *gossip-asynchronous framework* with the privacy concept of protection of the initial values. Note that all the previously mentioned privacy preserving average consensus papers propose protocols which work on the synchronous setting (all nodes update their values simultaneously).

The convergence analsysis of proposed gossip protocol (Algorithm 1) is dual in nature. The dual approach is explained in detail in Section 2. It was first proposed for solving linear systems in [5, 12] and then extended to the concept of average consensus problems in [10, 13]. The dual updates immediately correspond to updates of the primal variables, via an affine mapping.

Algorithm 1 is inspired by the works of [14, 15], and protects the initial values by inserting noise in the process. Broadly speaking, in each iteration, each of the sampled nodes first adds a noise to its current value, and an average is computed afterward. Convergence is guaranteed due to the correlation in the noise across iterations. Each node remembers the noise it added last time it was sampled, and in the following iteration, the previously added noise is first subtracted, and a fresh noise of smaller magnitude is added. Empirically, the protection of initial values is provided by first injecting noise into the system, which propagates across the network, but is gradually withdrawn to ensure convergence to the true average.

2 Technical Preliminaries

Primal and Dual Problems Consider solving the (primal) problem of projecting a given vector $c = x^0 \in \mathbb{R}^n$ onto the solution space of a linear system:

$$\min_{x\in\mathbb{R}^n} \{P(x) \overset{\text{def}}{=} \tfrac{1}{2}\|x - x^0\|^2\} \quad \text{subject to} \quad \mathbf{A}x = b, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $x^0 \in \mathbb{R}^n$. We assume the problem is feasible, i.e., that the system $\mathbf{A}x = b$ is consistent. With the above optimization problem we associate the dual problem

$$\max_{y\in\mathbb{R}^m} D(y) \overset{\text{def}}{=} (b - \mathbf{A}x^0)^\top y - \tfrac{1}{2}\|\mathbf{A}^\top y\|^2. \tag{2}$$

The dual is an unconstrained concave (but not necessarily strongly concave) quadratic maximization problem. It can be seen that as soon as the system $\mathbf{A}x = b$ is feasible, the dual problem is bounded. Moreover, all bounded concave quadratics in $\mathbb{R}^m$ can be written in the as $D(y)$ for some matrix $\mathbf{A}$ and vectors b and x^0 (up to an additive constant).

With any dual vector y we associate the primal vector via an affine transformation, $x(y) = x^0 + \mathbf{A}^\top y$. It can be shown that if y^* is dual optimal, then $x^* = x(y^*)$ is primal optimal. Hence, any dual algorithm producing a sequence of dual variables $y^t \to y^*$ gives rise to a corresponding primal algorithm producing the sequence $x^t \overset{\text{def}}{=} x(y^t) \to x^*$. See [5, 12] for the correspondence between primal and dual methods.

Randomized Gossip Setup: Choosing A. In the gossip framework we wish $(\mathbf{A}, b)$ to be an *average consensus (AC)* system.

Definition 1. *([10]) Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with $|\mathcal{V}| = n$ and $|\mathcal{E}| = m$. Let $\mathbf{A}$ be a real matrix with n columns. The linear system $\mathbf{A}x = b$ is an "average consensus (AC) system" for graph $\mathcal{G}$ if $\mathbf{A}x = b$ iff $x_i = x_j$ for all $(i, j) \in \mathcal{E}$.*

Algorithm 1 Privacy Preserving Gossip Algorithm via Controlled Noise Insertion

Input: vector of private values $c \in \mathbb{R}^n$; initial variances $\sigma_i^2 \in \mathbb{R}_+$ and variance decrease rate ϕ_i such that $0 \leq \phi_i < 1$ for all nodes i.
Initialize: Set $x^0 = c$; $t_1 = t_2 = \cdots = t_n = 0$, $v_1^{-1} = v_2^{-1} = \cdots = v_n^{-1} = 0$.
for $t = 0, 1, \ldots k-1$ **do**

1. Choose edge $e = (i,j) \in \mathcal{E}$ uniformly at random
2. Generate $v_i^{t_i} \sim N(0, \sigma_i^2)$ and $v_j^{t_j} \sim N(0, \sigma_j^2)$
3. Set $w_i^{t_i} = \phi_i^{t_i} v_i^{t_i} - \phi_i^{t_i - 1} v_i^{t_i - 1}$ and $w_j^{t_j} = \phi_j^{t_j} v_j^{t_j} - \phi_j^{t_j - 1} v_j^{t_j - 1}$
4. Update the primal variable: $x_i^{t+1} = x_j^{t+1} = \frac{x_i^t + w_i^{t_i} + x_j^t + w_j^{t_j}}{2}$, $\forall\, l \neq i,j:\ x_l^{t+1} = x_l^t$
5. Set $t_i = t_i + 1$ and $t_j = t_j + 1$

end
return x^k

In the rest of this paper we focus on a specific AC system; one in which the matrix $\mathbf{A}$ is the incidence matrix of the graph $\mathcal{G}$ (see Model 1 in [5]). In particular, we let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be the matrix defined as follows. Row $e = (i,j) \in \mathcal{E}$ of $\mathbf{A}$ is given by $\mathbf{A}_{ei} = 1$, $\mathbf{A}_{ej} = -1$ and $\mathbf{A}_{el} = 0$ if $l \notin \{i,j\}$. Notice that the system $\mathbf{A}x = 0$ encodes the constraints $x_i = x_j$ for all $(i,j) \in \mathcal{E}$, as desired. It is also known that randomized Kaczmarz method [17, 4, 11] applied to Problem 1 is equivalent to randomized gossip algorithm (see [10, 13, 9] for more details).

3 Private Gossip via Controlled Noise Insertion

In this section, we present the Gossip algorithm with Controlled Noise Insertion. As mentioned in the introduction, the approach is similar to the technique proposed in [14, 15]. Those works, however, address only algorithms in the synchronous setting, while our work is the first to use this idea in the asynchronous setting. Unlike the above, we provide finite time convergence guarantees and allow each node to add the noise differently, which yields a stronger result.

In our approach, each node adds noise to the computation independently of all other nodes. However, the noise added is correlated between iterations for each node. We assume that every node owns two parameters — the initial magnitude of the generated noise σ_i^2 and rate of decay of the noise ϕ_i. The node inserts noise $w_i^{t_i}$ to the system every time that an edge corresponding to the node was chosen, where variable t_i carries an information how many times the noise was added to the system in the past by node i. Thus, if we denote by t the current number of iterations, we have $\sum_{i=1}^n t_i = 2t$.

In order to ensure convergence to the optimal solution, we need to choose a specific structure of the noise in order to guarantee the mean of the values x_i converges to the initial mean. In particular, in each iteration a node i is selected, we subtract the noise that was added last time, and add a fresh noise with smaller magnitude: $w_i^{t_i} = \phi_i^{t_i} v_i^{t_i} - \phi_i^{t_i-1} v_i^{t_i-1}$, where $0 \leq \phi_i < 1, v_i^{-1} = 0$ and $v_i^{t_i} \sim N(0, \sigma_i^2)$ for all iteration counters $k_i \geq 0$ is independent to all other randomness in the algorithm. This ensures that all noise added initially is gradually withdrawn from the whole network.

After the addition of noise, a standard Gossip update is made, which sets the values of sampled nodes to their average. Hence, we have $\lim_{t\to\infty} \mathbb{E}\left[\left(\bar{c} - \frac{1}{n}\sum_{i=1}^n x_i^t\right)^2\right] = 0$, as desired.

It is not the purpose of this paper to define any quantifiable notion of protection of the initial values formally. However, we note that it is likely the case that the protection of private value c_i will be stronger for bigger σ_i and for ϕ_i closer to 1.

We now provide results of dual analysis of Algorithm 1.

Theorem 2. *Let us define* $\rho \overset{def}{=} 1 - \frac{\alpha(\mathcal{G})}{2m}$ *and* $\psi^t \overset{def}{=} \frac{1}{\sum_{i=1}^n \left(d_i \sigma_i^2\right)} \sum_{i=1}^n d_i \sigma_i^2 \left(1 - \frac{d_i}{m}\left(1 - \phi_i^2\right)\right)^t$, *where* $\alpha(\mathcal{G})$ *stands for algebraic connectivity of* $\mathcal{G}$ *and* d_i *denotes the degree of node* i. *Then for all*

$k \geq 1$ *we have the following bound*

$$\mathbb{E}\left[D(y^*) - D(y^k)\right] \leq \rho^k \left(D(y^*) - D(y^0)\right) + \frac{\sum \left(d_i \sigma_i^2\right)}{4m} \sum_{t=1}^{k} \rho^{k-t} \psi^t.$$

Note that ψ^t is a weighted sum of t-th powers of real numbers smaller than one. For large enough t, this quantity will depend on the largest of these numbers. This brings us to define M as the set of indices i for which the quantity $1 - \frac{d_i}{m}\left(1 - \phi_i^2\right)$ is maximized: $M = \arg\max_i \left\{1 - \frac{d_i}{m}\left(1 - \phi_i^2\right)\right\}$. Then for any $i_{\max} \in M$ we have

$$\psi^t \approx \frac{1}{\sum_{i=1}^n \left(d_i \sigma_i^2\right)} \sum_{i \in M} d_i \sigma_i^2 \left(1 - \frac{d_i}{m}\left(1 - \phi_i^2\right)\right)^t = \frac{\sum_{i \in M} d_i \sigma_i^2}{\sum_{i=1}^n \left(d_i \sigma_i^2\right)} \left(1 - \frac{d_{i_{\max}}}{m}\left(1 - \phi_{i_{\max}}^2\right)\right)^t,$$

which means that increasing ϕ_j for $j \notin M$ will not substantially influence convergence rate. Note that as soon as we have

$$\rho > 1 - \frac{d_i}{m}\left(1 - \phi_i^2\right) \tag{3}$$

for all i, the rate from theorem 2 will be driven by ρ^k (as $k \to \infty$) and we will have $\mathbb{E}\left[D(y^*) - D(y^k)\right] = \tilde{O}\left(\rho^k\right)$. One can think of the above as a threshold: if there is i such that ϕ_i is large enough so that the inequality (3) does not hold, the convergence rate is driven by $\phi_{i_{\max}}$. Otherwise, the rate is not influenced by the insertion of noise. Thus, in theory, we do not pay anything in terms of performance as long as we do not hit the threshold. One might be interested in choosing ϕ_i so that the threshold is attained for all i, and thus $M = \{1, \ldots, n\}$. This motivates the following result:

Corollary 3. *Let us choose* $\phi_i \overset{def}{=} \sqrt{1 - \frac{\gamma}{d_i}}$ *for all* i*, where* $\gamma \leq d_{\min}$*. Then*

$$\mathbb{E}\left[D(y^*) - D(y^k)\right] \quad \leq \quad \left(1 - \min\left(\frac{\alpha(\mathcal{G})}{2m}, \frac{\gamma}{m}\right)\right)^k \left(D(y^*) - D(y^0) + \frac{\sum_{i=1}^n \left(d_i \sigma_i^2\right)}{4m} k\right).$$

As a consequence, $\phi_i = \sqrt{1 - \frac{\alpha(\mathcal{G})}{2d_i}}$ *is the largest decrease rate of noise for node* i *such that the guaranteed convergence rate of the algorithm is not violated.*

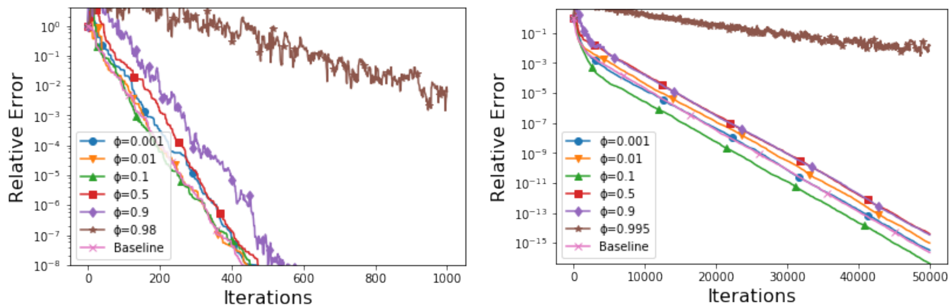
4 Experiments

In this section we present a preliminary experiment (for more experiments see Section C, in the Appendix) to evaluate the performance of the Algorithm 1 for solving the Average Consensus problem. The algorithm has two different parameters for each node i. These are the initial variance $\sigma_i^2 \geq 0$ and the rate of decay, ϕ_i, of the noise.

In this experiment we use two popular graph topologies the cycle graph (ring network) with $n = 10$ nodes and the random geometric graph with $n = 100$ nodes and radius $r = \sqrt{\log(n)/n}$.

In particular, we run Algorithm 1 with $\sigma_i = 1$ for all i, and set $\phi_i = \phi$ for all i and some ϕ. We study the effect of varying the value of ϕ on the convergence of the algorithm.

In Figure 1 we see that for small values of ϕ, we eventually recover the same rate of linear convergence as the Standard Pairwise Gossip algorithm (Baseline) of [2]. If the value of ϕ is sufficiently close to 1 however, the rate is driven by the noise and not by the convergence of the Standard Gossip algorithm. This value is $\phi = 0.98$ for cycle graph, and $\phi = 0.995$ for the random geometric graph in the plots we present.



(a) (b)

Figure 1: Convergence of Algorithm 1, on the cycle graph (left) and random geometric graph (right) for different values of ϕ. The "Relative Error " on the vertical axis represents the $\frac{\|x^t-x^*\|^2}{\|x^0-x^*\|^2}$

References

[1] Dimitri P Bertsekas and John N Tsitsiklis. *Parallel and distributed computation: numerical methods*, volume 23. Prentice hall Englewood Cliffs, NJ, 1989.

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

[3] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[4] Robert M Gower and Peter Richtárik. Randomized iterative methods for linear systems. *SIAM Journal on Matrix Analysis and Applications*, 36(4):1660–1690, 2015.

[5] Robert M Gower and Peter Richtárik. Stochastic dual ascent for solving linear systems. *arXiv:1512.06890*, 2015.

[6] F. Hanzely, J. Konečný, N. Loizou, P. Richtárik, and D. Grishchenko. Privacy preserving randomized gossip algorithms. *arXiv:1706.07636*, 2017.

[7] Zhenqi Huang, Sayan Mitra, and Geir Dullerud. Differentially private iterative synchronous consensus. In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, pages 81–90. ACM, 2012.

[8] David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*, pages 482–491. IEEE, 2003.

[9] Nicolas Loizou, Michael Rabbat, and Peter Richtárik. Provably accelerated randomized gossip algorithms. *arXiv:1810.13084*, 2018.

[10] Nicolas Loizou and Peter Richtárik. A new perspective on randomized gossip algorithms. In *4th IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2016.

[11] Nicolas Loizou and Peter Richtárik. Linearly convergent stochastic heavy ball method for minimizing generalization error. *NIPS Optimization for Machine Learning Workshop*, 2017.

[12] Nicolas Loizou and Peter Richtárik. Momentum and stochastic momentum for stochastic gradient, newton, proximal point and subspace descent methods. *arXiv:1712.09677*, 2017.

[13] Nicolas Loizou and Peter Richtárik. Accelerated gossip via stochastic heavy ball method. *arXiv:1809.08657*, 2018.

[14] Nicolaos E Manitara and Christoforos N Hadjicostis. Privacy-preserving asymptotic average consensus. In *Control Conference (ECC), 2013 European*, pages 760–765. IEEE, 2013.

[15] Yilin Mo and Richard M Murray. Privacy preserving average consensus. *IEEE Transactions on Automatic Control*, 62(2):753–765, 2017.

[16] Erfan Nozari, Pavankumar Tallapragada, and Jorge Cortés. Differentially private average consensus: obstructions, trade-offs, and optimal algorithm design. *Automatica*, 81:221–231, 2017.

[17] Thomas Strohmer and Roman Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *Journal of Fourier Analysis and Applications*, 15(2):262–278, 2009.

[18] John N Tsitsiklis. Problems in decentralized decision making and computation. Technical report, DTIC Document, 1984.

[19] John N Tsitsiklis, Dimitri Bertsekas, and Michael Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986.

Appendix

A Proof of Theorem 2

Before going into proofs, recall that duality mapping was defined as

$$x^t \stackrel{\text{def}}{=} x(y^t) = x^0 + \mathbf{A}^\top y^t. \tag{4}$$

The following lemma is connecting primal variables with dual suboprimality.

Lemma 4. *[5] (Relationship between convergence measures) Suppose that x is primal variable corresponding to the dual variable y as defined in* (4)*. Dual suboptimality can be expressed as the following:*

$$D(y^*) - D(y) = \tfrac{1}{2}\|\bar{c}\mathbf{1} - x\|^2. \tag{5}$$

A.1 Iteration Decrease

We first start with the lemma providing us with the expected decrease in dual suboptimality for each iteration.

Lemma 5. *Let d_i denote the number of neighbours of node i. Then,*

$$\begin{aligned}\mathbb{E}\left[D(y^*) - D(y^{t+1})\right] \leq \left(1 - \frac{\alpha(\mathcal{G})}{2m}\right)\mathbb{E}\left[D(y^*) - D(y^t)\right] + \frac{1}{4m}\sum_{i=1}^{n} d_i \sigma_i^2 \mathbb{E}\left[\phi_i^{2t_i}\right] \\ - \frac{1}{2m}\sum_{e\in\mathcal{E}} \mathbb{E}\left[\left(\phi_i^{t_i-1} v_i^{t_i-1} x_j^t + \phi_j^{t_j-1} v_j^{t_j-1} x_i^t\right)\right].\end{aligned} \tag{6}$$

Proof. Firstly we will compute increase in the dual function value in iteration t:

$$\begin{aligned}D(y^{t+1}) - D(y^t) &= \frac{1}{2}\|\bar{c}\mathbf{1} - x^t\|^2 - \frac{1}{2}\|\bar{c}\mathbf{1} - x^{t+1}\|^2 \\ &= \frac{1}{2}\left((\bar{c} - x_j^t)^2 + (\bar{c} - x_i^t)^2 - (\bar{c} - x_j^{t+1})^2 - (\bar{c} - x_i^{t+1})^2\right) \\ &= -\bar{c}\left(x_j^{t+1} + x_i^{t+1} - x_j^t - x_i^t\right) + \frac{1}{2}\left(\left(x_j^t\right)^2 + \left(x_i^t\right)^2 - \left(x_j^{t+1}\right)^2 - \left(x_i^{t+1}\right)^2\right) \\ &= -\bar{c}\left(w_j^{t_j} + w_i^{t_i}\right) + \frac{1}{2}\left(\left(x_j^t\right)^2 + \left(x_i^t\right)^2\right) \\ &\quad -\frac{1}{4}\left(\left(x_j^t + x_i^t\right)^2 + 2\left(x_j^t + x_i^t\right)\left(w_j^{t_j} + w_i^{t_i}\right) + \left(w_j^{t_j} + w_i^{t_i}\right)^2\right) \\ &= \frac{1}{4}\left(x_j^t - x_i^t\right)^2 - \left(w_i^{t_i} + w_j^{t_j}\right)\left(\bar{c} + \frac{1}{2}\left(x_j^t + x_i^t\right)\right) - \frac{1}{4}\left(w_i^{t_i} + w_j^{t_j}\right)^2.\end{aligned} \tag{7}$$

Now we want to estimate the expectation of this gap. Our main goal is to find $\mathbb{E}\left[D(y^{t+1}) - D(y^t)\right]$. There are 3 terms in (7). Since the expectation is linear, we will evaluate the expectations of the 3 terms separately and merge them at the end.

Taking the expectation over the choice of edge and inserted noise in iteration t we obtain

$$\mathbb{E}\left[\frac{1}{4}\left(x_i^t - x_j^t\right)^2 | x^t\right] = \frac{1}{4m}\sum_{e\in\mathcal{E}}\left(x_i^t - x_j^t\right)^2. \tag{8}$$

Thus we have

$$
\begin{aligned}
&\mathbb{E}\left[D(y^{t+1})-D(y^t)-\frac{1}{4}\left(x_i^t-x_j^t\right)^2 \mid x^t\right]\\
&\quad\overset{(8)}{=}\mathbb{E}\left[D(y^{t+1})-D(y^t)\mid x^t\right]-\frac{1}{4m}\sum_{e\in\mathcal{E}}\left(x_i^t-x_j^t\right)^2\\
&\quad=\mathbb{E}\left[D(y^*)-D(y^t)\mid x^t\right]+\mathbb{E}\left[D(y^{t+1})-D(y^*)\mid x^t\right]-\frac{1}{4m}\sum_{e\in\mathcal{E}}\left(x_i^t-x_j^t\right)^2\\
&\quad\overset{(5)}{=}\frac{1}{2}\|\bar{c}\mathbf{1}-x^t\|^2-\mathbb{E}\left[\frac{1}{2}\|\bar{c}\mathbf{1}-x^{t+1}\|^2\mid x^t\right]-\frac{1}{4m}\sum_{e=(i,j)\in\mathcal{E}}\left(x_i^t-x_j^t\right)^2\\
&\quad\overset{(16)}{\leq}\frac{1}{2}\|\bar{c}\mathbf{1}-x^t\|^2-\mathbb{E}\left[\frac{1}{2}\|\bar{c}\mathbf{1}-x^{t+1}\|^2\mid x^t\right]-\frac{\alpha(\mathcal{G})}{4m}\|\bar{c}\mathbf{1}-x^t\|^2\\
&\quad=\left(1-\frac{\alpha(\mathcal{G})}{2m}\right)\frac{1}{2}\|\bar{c}\mathbf{1}-x^t\|^2-\mathbb{E}\left[\frac{1}{2}\|\bar{c}\mathbf{1}-x^{t+1}\|^2\mid x^t\right]\\
&\quad\overset{(5)}{=}\left(1-\frac{\alpha(\mathcal{G})}{2m}\right)\left(D(y^*)-D(y^t)\right)-\mathbb{E}\left[D(y^*)-D(y^{t+1})\mid y^t\right].
\end{aligned}
$$

Taking the full expectation of the above and using tower property, we get

$$
\mathbb{E}\left[D(y^{t+1})-D(y^t)-\frac{1}{4}\left(x_i^t-x_j^t\right)^2\right]\leq\left(1-\frac{\alpha(\mathcal{G})}{2m}\right)\mathbb{E}\left[D(y^*)-D(y^t)\right]-\mathbb{E}\left[D(y^*)-D(y^{t+1})\right]. \tag{9}
$$

□

Now we are going to take the expectation of the second term of (7). We will use the "tower rule" of expectations in the form $\mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[X\mid Y,Z\right]\mid Y\right]\right]=\mathbb{E}\left[X\right]$, where X,Y,Z are random variables. In particular, we get

$$
\mathbb{E}\left[-\left(w_i^{t_i}+w_j^{t_j}\right)\left(\bar{c}+\frac{1}{2}\left(x_j^t+x_i^t\right)\right)\right]=\mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[-\left(w_i^{t_i}+w_j^{t_j}\right)\left(\bar{c}+\frac{1}{2}\left(x_j^t+x_i^t\right)\right)\mid e^t,x^t\right]\mid x^t\right]\right].
$$

In the equation above, e^t denotes an edge selected at in the iteration t. Let us first calculate the inner most expectation on the right hand side of the above identity:

$$
\begin{aligned}
&\mathbb{E}\left[-\left(w_i^{t_i}+w_j^{t_j}\right)\left(\bar{c}+\frac{1}{2}\left(x_j^t+x_i^t\right)\right)\mid e^t,x^t\right]\\
&\quad\overset{(*^1)}{=}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}-\phi_i^{t_i}v_i^{t_i}-\phi_j^{t_j}v_j^{t_j}\right)\left(\bar{c}+\frac{1}{2}\left(x_j^t+x_i^t\right)\right)\mid e^t,x^t\right]\\
&\quad\overset{(*^2)}{=}\mathbb{E}\left[\overbrace{\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)}^{\text{constant}}\overbrace{\left(\bar{c}+\frac{1}{2}\left(x_j^t+x_i^t\right)\right)}^{\text{constant}}\mid e^t,x^t\right]\\
&\qquad+\mathbb{E}\left[\left(-\phi_i^{t_i}v_i^{t_i}-\phi_j^{t_j}v_j^{t_j}\right)\overbrace{\left(\bar{c}+\frac{1}{2}\left(x_j^t+x_i^t\right)\right)}^{\text{constant}}\mid e^t,x^t\right]\\
&\quad=\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\left(\bar{c}+\frac{1}{2}\left(x_j^t+x_i^t\right)\right)\\
&\qquad+\mathbb{E}\left[\left(-\phi_i^{t_i}v_i^{t_i}-\phi_j^{t_j}v_j^{t_j}\right)|e,x^t\right]\left(\bar{c}+\frac{1}{2}\left(x_j^t+x_i^t\right)\right)\quad(\to 0)\\
&\quad\overset{L.6}{=}\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\left(\bar{c}+\frac{1}{2}\left(x_j^t+x_i^t\right)\right)\\
&\quad=\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\bar{c}+\frac{1}{2}\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\left(x_j^t+x_i^t\right),
\end{aligned}
$$

where $(*^1)$ means definition of $w_i^{t_i}$ and $(*^2)$ means linearity of expectation.

Now we take the expectation of the last expression above with respect to the choice of an edge at t-th iteration. We obtain

$$\begin{aligned}
&\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\overline{c}+\frac{1}{2}\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\left(x_j^t+x_i^t\right)|x^t\right]\\
&\overset{(*^2)}{=}\frac{1}{2}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\left(x_j^t+x_i^t\right)|x^t\right]+\cancelto{0}{\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\overline{c}|x^t\right]}\\
&\overset{L.6}{=}\frac{1}{2}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\left(x_j^t+x_i^t\right)|x^t\right]\\
&\overset{(*^3)}{=}\frac{1}{2m}\sum_{e\in\mathcal{E}}\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)\left(x_j^t+x_i^t\right)\\
&=\frac{1}{2m}\sum_{e\in\mathcal{E}}\left(\phi_i^{t_i-1}v_i^{t_i-1}x_i^t+\phi_j^{t_j-1}v_j^{t_j-1}x_j^t\right)+\frac{1}{2m}\sum_{e\in\mathcal{E}}\left(\phi_i^{t_i-1}v_i^{t_i-1}x_j^t+\phi_j^{t_j-1}v_j^{t_j-1}x_i^t\right)\\
&\overset{(*^4)}{=}\frac{1}{2m}\sum_{i=1}^{n}d_i\phi_i^{t_i-1}v_i^{t_i-1}x_i^t+\frac{1}{2m}\sum_{e\in\mathcal{E}}\left(\phi_i^{t_i-1}v_i^{t_i-1}x_j^t+\phi_j^{t_j-1}v_j^{t_j-1}x_i^t\right),
\end{aligned}$$

where $(*^3)$ means definition of expectation and $(*^4)$ means change of the summation order.

A.2 Lemmas about noise insertion

Lemma 6. *Suppose that we run Algorithm 4 for t iterations and t_i denotes the number of times that some edge corresponding to node i was selected during the algorithm.*

1. *$v_i^{t_i}$ and t_j are independent for all (i.e., not necessarily distinct) i, j.*
2. *$v_i^{t_i}$ and $\phi_j^{t_j}$ are independent for all (i.e., not necessarily distinct) i, j.*
3. *$w_i^{t_i}$ and $w_j^{t_j}$ have zero correlation for all $i \neq j$.*
4. *x_j^t and $\phi_i^{t_i}v_i^{t_i}$ have zero correlation for all (i.e., not necessarily distinct) i, j.*

Proof.

1. Follows from the definition of v_i^t.
2. Follows from the definition of v_i^t.
3. Note that we have $w_i^{t_i} = \phi_i^{t_i}v_i^{t_i} - \phi_i^{t_i-1}v_i^{t_i-1}$ and $w_j^{t_j} = \phi_j^{t_j}v_j^{t_j} - \phi_j^{t_j-1}v_j^{t_j-1}$. Clearly, $v_i^{t_i}$ and $w_j^{t_j}$ have zero correlation. Similarly $v_i^{t_i-1}$ and $w_j^{t_j}$ have zero correlation. Thus, $w_i^{t_i}$ and $w_j^{t_j}$ have zero correlation.
4. Clearly, x_j^t is a function initial state and all instances of random variables up to the iteration t. Thus, $v_i^{t_i}$ is independent to x_j^t from the definition. Thus, x_j^t and $\phi_i^{t_i}v_i^{t_i}$ have zero correlation.

□

Lemma 7.

$$\mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}x_i^t\right]=\frac{1}{2}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right]. \tag{10}$$

Proof.

$$
\begin{aligned}
&\mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}x_i^t\right]\\
&\overset{(*^5)}{=} \mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}\left(\left(x_i^t - \frac{\phi_i^{t_i-1}v_i^{t_i-1}}{2}\right) + \frac{\phi_i^{t_i-1}v_i^{t_i-1}}{2}\right)\right]\\
&\overset{(*^2)}{=} \mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}\left(x_i^t - \frac{\phi_i^{t_i-1}v_i^{t_i-1}}{2}\right)\right] + \frac{1}{2}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right]\\
&\overset{(*^6)}{=} \mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}\left(\frac{x_i^{t_i-1} + x_l^{t_l^0} + w^{t_i-1} + w^{t_l^0} - \phi_i^{t_i-1}v_i^{t_i-1}}{2}\right)\right] + \frac{1}{2}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right]\\
&\overset{(*^1)}{=} \mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}\left(\frac{x_i^{t_i-1} + x_l^{t_l^0} + \phi_i^{t_i-1}v_i^{t_i-1} - \phi_i^{t_i-2}v_i^{t_i-2} + \phi_l^{t_l^0}v_l^{t_l^0} - \phi_l^{t_l^0-1}v_l^{t_l^0-1} - \phi_i^{t_i-1}v_i^{t_i-1}}{2}\right)\right]\\
&\qquad + \frac{1}{2}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right]\\
&= \mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}\left(\frac{x_i^{t_i-1} + x_l^{t_l^0} + \phi_l^{t_l^0}v_l^{t_l^0} - \phi_l^{t_l^0-1}v_l^{t_l^0-1} - \phi_i^{t_i-2}v_i^{t_i-2}}{2}\right)\right] + \frac{1}{2}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right]\\
&\overset{L.6}{=} \cancelto{0}{\mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}\right]}\mathbb{E}\left[\left(\frac{x_i^{t_i-1} + x_l^{t_l^0} + \phi_l^{t_l^0}v_l^{t_l^0} - \phi_l^{t_l^0-1}v_l^{t_l^0-1} - \phi_i^{t_i-2}v_i^{t_i-2}}{2}\right)\right]\\
&\qquad + \frac{1}{2}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right]\\
&= \frac{1}{2}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right],
\end{aligned}
$$

where in step $(*^5)$ we add and subtracting $\frac{\phi_i^{t_i-1}v_i^{t_i-1}}{2}$. In the step $(*^6)$ we denote by l a node such that that the noise $\phi_i^{t_i-1}v_i^{t_i-1}$ was added to the system when the edge (i, l) was chosen (we do not consider $t_i = 0$ since in this case the Lemma 7 trivially holds).

□

Taking the expectation with respect to the algorithm we obtain

$$
\begin{aligned}
&\mathbb{E}\left[-\left(w_i^{t_i} + w_j^{t_j}\right)\left(\bar{c} + \frac{1}{2}\left(x_j^t + x_i^t\right)\right)\right]\\
&\overset{(*^7)}{=} \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[-\left(w_i^{t_i} + w_j^{t_j}\right)\left(\bar{c} + \frac{1}{2}\left(x_j^t + x_i^t\right)\right)|x^t, e\right]|x^t\right]\right]\\
&= \mathbb{E}\left[\frac{1}{2m}\sum_{i=1}^n d_i\phi_i^{t_i-1}v_i^{t_i-1}x_i^t + \frac{1}{2m}\sum_{e\in\mathcal{E}}\left(\phi_i^{t_i-1}v_i^{t_i-1}x_j^t + \phi_j^{t_j-1}v_j^{t_j-1}x_i^t\right)\right]\\
&\overset{(*^2)}{=} \frac{1}{2m}\sum_{i=1}^n d_i\mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}x_i^t\right] + \frac{1}{2m}\sum_{e\in\mathcal{E}}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}x_j^t + \phi_j^{t_j-1}v_j^{t_j-1}x_i^t\right)\right]\\
&\overset{L.7}{=} \frac{1}{4m}\sum_{i=1}^n d_i\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right] + \frac{1}{2m}\sum_{e\in\mathcal{E}}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}x_j^t + \phi_j^{t_j-1}v_j^{t_j-1}x_i^t\right)\right], \quad (11)
\end{aligned}
$$

where $(*^7)$ means tower rule.

Lemma 8.

$$\mathbb{E}\left[\left(\phi_i^{t_i}v_i^{t_i} + \phi_j^{t_j}v_j^{t_j}\right)^2|x^t, e^t\right] = \sigma_i^2\phi_i^{2t_i} + \sigma_j^2\phi_j^{2t_j}. \quad (12)$$

Proof. Since we have $\mathbb{E}\left[\left(\phi_i^{t_i}v_i^{t_i} + \phi_j^{t_j}v_j^{t_j}\right)|x^t, e^t\right] = 0$, and also for any random variable X: $\mathbb{E}\left[X^2\right] = \mathbb{V}(X) + \mathbb{E}[X]^2$, we only need to compute the variance:

$$\mathbb{V}\left(\phi_i^{t_i}v_i^{t_i} + \phi_j^{t_j}v_j^{t_j}\right) = \mathbb{V}\left(\phi_i^{t_i}v_i^{t_i}\right) + \mathbb{V}\left(\phi_j^{t_j}v_j^{t_j}\right) = \left(\phi_i^{t_i}\right)^2\mathbb{V}\left(v_i^{t_i}\right) + \left(\phi_j^{t_j}\right)^2\mathbb{V}\left(v_j^{t_j}\right).$$

□

Taking an expectation of the third term of (7) with respect to lastly added noise, the expression $\mathbb{E}\left[\left(w_i^{t_i}+w_j^{t_j}\right)^2|x^t,e^t\right]$ is equal to

$$\begin{aligned}
&\overset{(*^1)}{=} \mathbb{E}\left[\left(\phi_i^{t_i}v_i^{t_i}+\phi_j^{t_j}v_j^{t_j}-\phi_i^{t_i-1}v_i^{t_i-1}-\phi_j^{t_j-1}v_j^{t_j-1}\right)^2|x^t,e^t\right]\\
&\overset{(*^2)}{=} \mathbb{E}\left[\left(\phi_i^{t_i}v_i^{t_i}+\phi_j^{t_j}v_j^{t_j}\right)^2|x^t,e^t\right]-2\mathbb{E}\left[\overbrace{\left(\phi_i^{t_i}v_i^{t_i}+\phi_j^{t_j}v_j^{t_j}\right)}^{\mathbb{E}\left[\phi_i^{t_i}v_i^{t_i}+\phi_j^{t_j}v_j^{t_j}\right]=0}\overbrace{\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)}^{\text{constant}}|x^t,e^t\right]\\
&\quad+\overbrace{\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)^2|x^t,e^t\right]}^{\text{constant}}\\
&\overset{(12)}{=} \sigma_i^2\phi_i^{2t_i}+\sigma_j^2\phi_j^{2t_j}+\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)^2.
\end{aligned}$$

Taking the expectation over e^t we obtain:

$$\begin{aligned}
\mathbb{E}\left[\left(w_i^{t_i}+w_j^{t_j}\right)^2|x^t\right] &\overset{(*^7)}{=} \mathbb{E}\left[\mathbb{E}\left[\left(w_i^{t_i}+w_j^{t_j}\right)^2|e^t,x^t\right]|x^t\right]\\
&= \mathbb{E}\left[\sigma_i^2\phi_i^{2t_i}+\sigma_j^2\phi_j^{2t_j}+\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)^2|x^t\right]\\
&\overset{(*^3)}{=} \frac{1}{m}\sum_{e\in\mathcal{E}}\left(\sigma_i^2\phi_i^{2t_i}+\sigma_j^2\phi_j^{2t_j}\right)+\frac{1}{m}\sum_{e\in\mathcal{E}}\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)^2\\
&\overset{(*^4)}{=} \frac{1}{m}\sum_{i=1}^{n}d_i\sigma_i^2\phi_i^{2t_i}+\frac{1}{m}\sum_{e\in\mathcal{E}}\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)^2.
\end{aligned}$$

Finally, taking the expectation with respect to the algorithm we get

$$\begin{aligned}
&\mathbb{E}\left[\left(w_i^{t_i}+w_j^{t_j}\right)^2\right]\\
&\overset{(*^7)}{=} \mathbb{E}\left[\mathbb{E}\left[\left(w_i^{t_i}+w_j^{t_j}\right)^2|x^t\right]\right]\\
&\overset{(*^2)}{=} \frac{1}{m}\sum_{i=1}^{n}d_i\sigma_i^2\mathbb{E}\left[\phi_i^{2t_i}\right]+\frac{1}{m}\sum_{e\in\mathcal{E}}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}+\phi_j^{t_j-1}v_j^{t_j-1}\right)^2\right]\\
&=\frac{1}{m}\sum_{i=1}^{n}d_i\sigma_i^2\mathbb{E}\left[\phi_i^{2t_i}\right]+\frac{1}{m}\sum_{e\in\mathcal{E}}\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2+\left(\phi_j^{t_j-1}v_j^{t_j-1}\right)^2+2\phi_i^{t_i-1}v_i^{t_i-1}\phi_j^{t_j-1}v_j^{t_j-1}\right]\\
&\overset{(*^4)}{=} \frac{1}{m}\sum_{i=1}^{n}d_i\sigma_i^2\mathbb{E}\left[\phi_i^{2t_i}\right]+\frac{1}{m}\sum_{i=1}^{n}d_i\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right]+\frac{2}{m}\sum_{e\in\mathcal{E}}\mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}\phi_j^{t_j-1}v_j^{t_j-1}\right]\\
&\overset{L.6}{=} \frac{1}{m}\sum_{i=1}^{n}d_i\sigma_i^2\mathbb{E}\left[\phi_i^{2t_i}\right]+\frac{1}{m}\sum_{i=1}^{n}d_i\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right]+\frac{2}{m}\sum_{e\in\mathcal{E}}\cancelto{0}{\mathbb{E}\left[\phi_i^{t_i-1}v_i^{t_i-1}\right]}\cancelto{0}{\mathbb{E}\left[\phi_j^{t_j-1}v_j^{t_j-1}\right]}\\
&=\frac{1}{m}\sum_{i=1}^{n}d_i\sigma_i^2\mathbb{E}\left[\phi_i^{2t_i}\right]+\frac{1}{m}\sum_{i=1}^{n}d_i\mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}\right)^2\right]. \qquad (13)
\end{aligned}$$

Combining (7) with (9), (11) and (13) we obtain

$$
\begin{aligned}
\mathbb{E}\left[D(y^*) - D(y^{t+1})\right] &\leq \left(1 - \frac{\alpha(\mathcal{G})}{2m}\right)\mathbb{E}\left[D(y^*) - D(y^t)\right] - \frac{1}{4m}\sum_{i=1}^{n} d_i \mathbb{E}\left[\left(\phi_i^{t_i-1} v_i^{t_i-1}\right)^2\right] \\
&\quad - \frac{1}{2m}\sum_{e\in\mathcal{E}} \mathbb{E}\left[\left(\phi_i^{t_i-1} v_i^{t_i-1} x_j^t + \phi_j^{t_j-1} v_j^{t_j-1} x_i^t\right)\right] \\
&\quad + \frac{1}{4m}\sum_{i=1}^{n} d_i \sigma_i^2 \mathbb{E}\left[\phi_i^{2t_i}\right] + \frac{1}{4m}\sum_{i=1}^{n} d_i \mathbb{E}\left[\left(\phi_i^{t_i-1} v_i^{t_i-1}\right)^2\right] \\
&= \left(1 - \frac{\alpha(\mathcal{G})}{2m}\right)\mathbb{E}\left[D(y^*) - D(y^t)\right] + \frac{1}{4m}\sum_{i=1}^{n} d_i \sigma_i^2 \mathbb{E}\left[\phi_i^{2t_i}\right] \\
&\quad - \frac{1}{2m}\sum_{e\in\mathcal{E}} \mathbb{E}\left[\left(\phi_i^{t_i-1} v_i^{t_i-1} x_j^t + \phi_j^{t_j-1} v_j^{t_j-1} x_i^t\right)\right],
\end{aligned}
$$

which concludes the proof.

A.3 Proof of Theorem 2

Lemma 9. *After t iterations of algorithm 4 we have*

$$
\mathbb{E}\left[\phi_i^{2t_i}\right] = \left(1 - \frac{d_i}{m}\left(1 - \phi_i^2\right)\right)^t. \tag{14}
$$

Proof.

$$
\begin{aligned}
\mathbb{E}\left[\phi_i^{2t_i}\right] &= \sum_{j=0}^{t} \mathbb{P}(t_i = j)\phi_i^{2j} = \sum_{j=0}^{t} \binom{t}{j}\left(\frac{m-d_i}{m}\right)^{t-j}\left(\frac{d_i}{m}\phi_i^2\right)^j \\
&= \left(\frac{m-d_i}{m} + \frac{d_i}{m}\phi_i^2\right)^t = \left(1 - \frac{d_i}{m}\left(1-\phi_i^2\right)\right)^t.
\end{aligned}
$$

□

Lemma 10. *Random variables $\phi_i^{t_i-1} v_i^{t_i-1}$ and x_j^t are nonegatively correlated, i.e.*

$$
\mathbb{E}\left[\phi_i^{t_i-1} v_i^{t_i-1} x_j^t\right] \geq 0. \tag{15}
$$

Proof. Denote $R_{i,j}$ to be a random variable equal to 1 if the noise $w_i^{t_i}$ was added to the system when edge (i, j) was chosen and equal to 0 otherwise. We can rewrite the expectation in the following way:

$$
\begin{aligned}
\mathbb{E}\left[\phi_i^{t_i-1} v_i^{t_i-1} x_j^t\right] &= \overbrace{\mathbb{E}\left[\phi_i^{t_i-1} v_i^{t_i-1} x_j^t \mid R_{i,j} = 1\right]}^{\geq 0} \mathbb{P}(R_{i,j} = 1) \\
&\quad + \overbrace{\mathbb{E}\left[\phi_i^{t_i-1} v_i^{t_i-1} x_j^t \mid R_{i,j} = 0\right]}^{0} \mathbb{P}(R_{i,j} = 0) \geq 0.
\end{aligned}
$$

The inequality $\mathbb{E}\left[\phi_i^{t_i-1} v_i^{t_i-1} x_j^t \mid R_{i,j} = 1\right] \geq 0$ holds due to the fact that $\phi_i^{t_i-1} v_i^{t_i-1}$ was added to x_j with the positive sign. □

Combining (6) with Lemmas 9 and 10 we obtain

$$\begin{aligned}
\mathbb{E}\left[D(y^*) - D(y^{t+1})\right] &\leq \left(1 - \frac{\alpha(\mathcal{G})}{2m}\right)\mathbb{E}\left[D(y^*) - D(y^t)\right] + \frac{1}{4m}\sum_{i=1}^{n} d_i\sigma_i^2 \mathbb{E}\left[\phi_i^{2t_i}\right] \\
&\quad - \frac{1}{2m}\sum_{e\in\mathcal{E}} \mathbb{E}\left[\left(\phi_i^{t_i-1}v_i^{t_i-1}x_j^t + \phi_i^{t_j-1}v_j^{t_j-1}x_i^t\right)\right] \\
&\overset{(15)}{\leq} \left(1 - \frac{\alpha(\mathcal{G})}{2m}\right)\mathbb{E}\left[D(y^*) - D(y^t)\right] + \frac{1}{4m}\sum_{i=1}^{n} d_i\sigma_i^2 \mathbb{E}\left[\phi_i^{2t_i}\right] \\
&\overset{(14)}{=} \left(1 - \frac{\alpha(\mathcal{G})}{2m}\right)\mathbb{E}\left[D(y^*) - D(y^t)\right] + \frac{1}{4m}\sum_{i=1}^{n} d_i\sigma_i^2 \left(1 - \frac{d_i}{m}\left(1 - \phi_i^2\right)\right)^t \\
&= \left(1 - \frac{\alpha(\mathcal{G})}{2m}\right)\mathbb{E}\left[D(y^*) - D(y^t)\right] + \frac{\sum\left(d_i\sigma_i^2\right)}{4m}\psi^t.
\end{aligned}$$

The recursion above gives us inductively the following

$$\mathbb{E}\left[D(y^*) - D(y^k)\right] \leq \rho^k\left(D(y^*) - D(y^0)\right) + \frac{\sum\left(d_i\sigma_i^2\right)}{4m}\sum_{t=1}^{k}\rho^{k-t}\psi^t,$$

which concludes the proof of the theorem.

A.4 Proof of Corollary 3

Note that we have

$$\begin{aligned}
\psi^t &= \frac{1}{\sum_{i=1}^n d_i\sigma_i^2}\sum_{i=1}^{n} d_i\sigma_i^2\left(1 - \frac{d_i}{m}\left(1 - \left(1 - \frac{\gamma}{d_i}\right)\right)\right)^t \\
&= \frac{1}{\sum_{i=1}^n d_i\sigma_i^2}\sum_{i=1}^{n} d_i\sigma_i^2\left(1 - \frac{\gamma}{m}\right)^t \quad = \quad \left(1 - \frac{\gamma}{m}\right)^t.
\end{aligned}$$

In view of Theorem 2, this gives us the following:

$$\begin{aligned}
\mathbb{E}\left[D(y^*) - D(y^k)\right] &\leq \left(D(y^*) - D(y^0)\right)\rho^k + \frac{\sum\left(d_i\sigma_i^2\right)}{4m}\sum_{t=1}^{k}\rho^{k-t}\psi^t \\
&\leq \left(D(y^*) - D(y^0)\right)\rho^k + \frac{\sum\left(d_i\sigma_i^2\right)}{4m}\sum_{t=1}^{k}\rho^{k-t}\left(1 - \frac{\gamma}{m}\right)^t \\
&\leq \left(D(y^*) - D(y^0)\right)\rho^k + \frac{\sum\left(d_i\sigma_i^2\right)}{4m}k\max\left(\rho, 1 - \frac{\gamma}{m}\right)^k \\
&\leq \left(D(y^*) - D(y^0) + \frac{\sum\left(d_i\sigma_i^2\right)}{4m}k\right)\max\left(\rho, 1 - \frac{\gamma}{m}\right)^k.
\end{aligned}$$

B Technical Lemmas

Lemma 11. *Let $x \in \mathbb{R}^n$ such that $\frac{1}{n}\sum_i x_i = \bar{c}$. Then*

$$\frac{1}{2}\|\bar{c}\mathbf{1} - x\|^2 \leq \frac{1}{2\alpha(\mathcal{G})}\sum_{e=(i,j)\in\mathcal{E}}(x_i - x_j)^2. \tag{16}$$

Proof.

$$\begin{aligned}
\frac{1}{2}\|\bar{c}\mathbf{1} - x\|^2 &\overset{(18)}{=} \frac{1}{4n}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x_i)^2 \quad = \quad \frac{1}{2n}\sum_{(i,j)}(x_j - x_i)^2 \\
&\overset{(19)}{\leq} \frac{\beta(\mathcal{G})}{2n}\sum_{e=(i,j)\in\mathcal{E}}(x_i - x_j)^2 \quad \overset{\text{Lemma 15}}{=} \quad \frac{1}{2\alpha(\mathcal{G})}\sum_{e=(i,j)\in\mathcal{E}}(x_i - x_j)^2
\end{aligned}$$

□

Lemma 12.

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n}(x_j - x_i)\right)^2 = \frac{n}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x_i)^2 \tag{17}$$

Proof. Using simple algebra we have

$$\begin{aligned}
\sum_{i=1}^{n}\left(\sum_{j=1}^{n}(x_j - x_i)\right)^2 &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n}x_j - nx_i\right)^2 \\
&= \sum_{i=1}^{n}\left(\left(\sum_{j=1}^{n}x_j\right)^2 + n^2x_i^2 - 2nx_i\left(\sum_{j=1}^{n}x_j\right)\right) \\
&= n\left(\sum_{j=1}^{n}x_j\right)^2 + n^2\sum_{i=1}^{n}x_i^2 - 2n\left(\sum_{j=1}^{n}x_j\right)^2 \\
&= n^2\sum_{i=1}^{n}x_i^2 - n\left(\sum_{i=1}^{n}x_i\right)^2.
\end{aligned}$$

Manipulating right hand side of (17) we obtain

$$\begin{aligned}
\frac{n}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x_i)^2 &= \frac{n}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(x_j^2 + x_i^2 - 2x_ix_j\right) \\
&= n^2\sum_{i=1}^{n}x_i^2 - n\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j \quad = \quad n^2\sum_{i=1}^{n}x_i^2 - n\left(\sum_{i=1}^{n}x_i\right)^2.
\end{aligned}$$

Clearly, LHS and RHS of (17) are equal. □

Lemma 13. *For any $x \in \mathbb{R}^n$ we have :*

$$\frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x_i)^2 \quad = \quad \|\bar{c}\mathbf{1} - x\|^2 \tag{18}$$

Proof. In order to show (18) it is enough to notice that

$$\begin{aligned}
\|\bar{c}\mathbf{1} - x\|^2 &= \sum_{i=1}^{n}(\bar{c} - x_i)^2 \quad = \quad \sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j=1}^{n}x_j - x_i\right)^2 \\
&= \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{1}{n}(x_j - x_i)\right)^2 \quad \overset{(17)}{=} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{n}(x_j - x_i)^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_j - x_i)^2.
\end{aligned}$$

□

Lemma 14. *The eigenvalues of $\tilde{\mathbf{L}} = n\mathbf{I} - \mathbf{1}\mathbf{1}^\top$ are $\{0, n, n, \dots, n\}$*

Proof. Clearly, $\tilde{\mathbf{L}}\mathbf{1} = 0$. Consider some vector x such that $\langle x, \mathbf{1}\rangle = 0$. Then, $\tilde{\mathbf{L}}x = n\mathbf{I}x - \mathbf{1}\mathbf{1}^\top x = nx + \mathbf{1}\mathbf{1}^\top x = nx$ thus x is an eigenvector corresponding to eigenvalue n. Thus, we can pick $n-1$ linearly independent eigenvectors of $\tilde{\mathbf{L}}$ corresponding to eigenvalue n, which concludes the proof. □

With graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ we now associate a certain quantity, which we shall denote $\beta = \beta(\mathcal{G})$. It is the smallest nonnegative number β such that the following inequality[3] holds for all $x \in \mathbb{R}^n$:

$$\sum_{(i,j)} (x_j - x_i)^2 \leq \beta \sum_{(i,j)\in\mathcal{E}} (x_j - x_i)^2. \tag{19}$$

Lemma 15. $\beta(\mathcal{G}) = \frac{n}{\alpha(\mathcal{G})}$.

Proof. The Laplacian matrix of $\mathcal{G}$ is the matrix $\mathbf{L} = \mathbf{A}^\top \mathbf{A}$. We have $\mathbf{L}_{ii} = d_i$ (degree of vertex i), $\mathbf{L}_{ij} = \mathbf{L}_{ji} = -1$ if $(i,j) \in \mathcal{E}$ and $\mathbf{L}_{ij} = 0$ otherwise. A simple computation reveals that for any $x \in \mathbb{R}^n$ we have

$$x^\top \mathbf{L} x = \sum_{e=(i,j)\in\mathcal{E}} (x_i - x_j)^2.$$

Let $\tilde{\mathbf{A}}$ be the $n(n-1)/2 \times n$ matrix corresponding to the complete graph $\tilde{\mathcal{G}}$ on $\mathcal{V}$. Let $\tilde{\mathbf{L}} = \tilde{\mathbf{A}}^\top \tilde{\mathbf{A}}$ be its Laplacian. We have $\tilde{\mathbf{L}}_{ii} = n-1$ for all i and $\tilde{\mathbf{L}}_{ij} = -1$ for $i \neq j$. So, $\tilde{\mathbf{L}} = n\mathbf{I} - \mathbf{1}\mathbf{1}^\top$. Then

$$x^\top \tilde{\mathbf{L}} x = n\|x\|^2 - \left(\sum_{i=1}^n x_i\right)^2 = \sum_{(i,j)} (x_i - x_j)^2.$$

Inequality (19) can therefore be recast as follows:

$$x^\top (n\mathbf{I} - \mathbf{1}\mathbf{1}^\top) x \leq x^\top \beta(\mathcal{G}) \mathbf{L} x, \qquad x \in \mathbb{R}^n.$$

Let $\beta = \beta(\mathcal{G})$. Note that both $\tilde{\mathbf{L}}$ and $\beta\mathbf{L}$ are Hermitian thus have real eigenvalues and there exist an orthonormal basis of their eigenvectors. Suppose that $\{x_1, \dots x_n\}$ are eigenvectors of βL corresponding to eigenvalues $\lambda_1(\beta\mathbf{L}), \lambda_2(\beta\mathbf{L}) \dots, \lambda_n(\beta\mathbf{L})$. Without loss of generality assume that these eigenvectors form an orthonormal basis and $\lambda_1(\beta\mathbf{L}) \geq \dots \geq \lambda_n(\beta\mathbf{L})$

Clearly, $\lambda_n(\beta\mathbf{L}) = 0$, $x_n = \mathbf{1}/\sqrt{n}$, and $\lambda_{n-1}(\beta\mathbf{L}) = n$. Lemma 14 states that eigenvalues of $\tilde{\mathbf{L}}$ are $\{0, n, n, \dots, n\}$.

One can easily see that eigenvector corresponding to zero eigenvalue of $\tilde{\mathbf{L}}$ is x_n. Note that eigenvectors $x_1, \dots, x_{n-1}$ generate an eigenspace corresponding to eigenvalue n of $\tilde{\mathbf{L}}$.

Consider some $x = \sum_{i=1}^n c_i x_i$, $c_i \in \mathbb{R}$ for all i. Then we have

$$x^\top \tilde{\mathbf{L}} x = \sum_{i=1}^n \lambda_i\left(\tilde{\mathbf{L}}\right) c_i^2 \leq \sum_{i=1}^n \lambda_i\left(\beta\mathbf{L}\right) c_i^2 = x^\top \beta \mathbf{L} x,$$

which concludes the proof. □

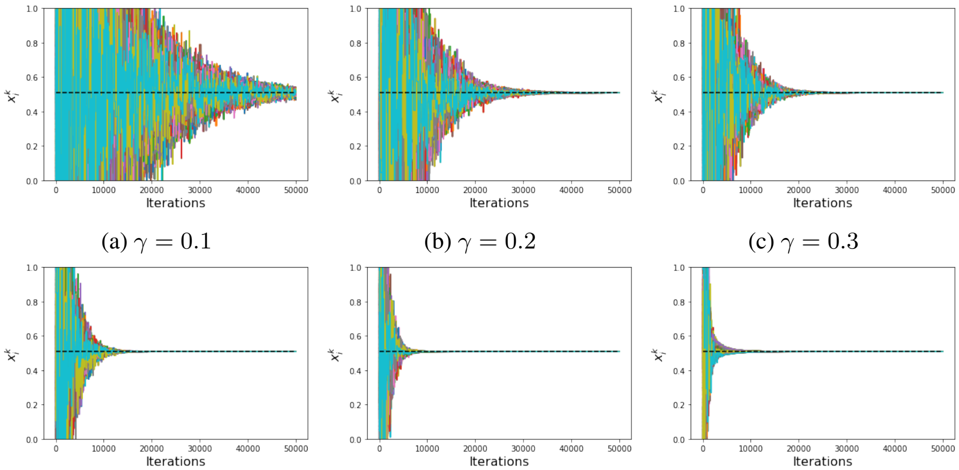
C More Experiments

C.1 Variance 1 and different decay rates

In this subsection, we perform a similar experiment to the one of Section 4, but now the values ϕ_i are not all the same. We rather control them by the choice of γ through $\phi_i \stackrel{\text{def}}{=} \sqrt{1 - \frac{\gamma}{d_i}}$(see Corollary 3). Note that by decreasing γ, we increase ϕ_i, and thus smaller γ means the noise decays at a slower rate. Here, due to the regular structure of the cycle graph, we present only results for the random geometric graph.

We remark that we again see the existence of a threshold predicted by theory, beyond which the convergence is dominated by the inserted noise. Otherwise, we recover the rate of the Standard Gossip algorithm (Baseline Method).

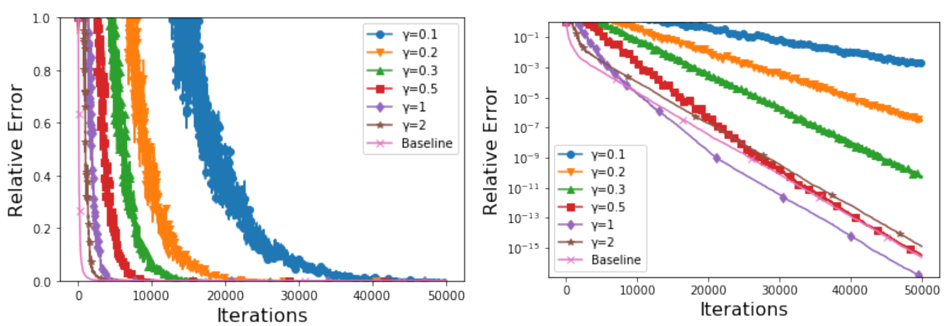
[3] We write $\sum_{(i,j)}$ to indicate sum over all *unordered* pairs of vertices. That is, we do not count (i,j) and (j,i) separately, only once. By $\sum_{(i,j)\in\mathcal{E}}$ we denote a sum over all edges of $\mathcal{G}$. On the other hand, by writing $\sum_i \sum_j$, we are summing over all (unordered) pairs of vertices twice.



(a) $\gamma = 0.1$ (b) $\gamma = 0.2$ (c) $\gamma = 0.3$

(d) $\gamma = 0.5$ (e) $\gamma = 1$ (f) $\gamma = 2$

Figure 2: Trajectories of the values of x_i^t for Controlled Noise Insertion run on the random geometric graph for different values of ϕ_i, controlled by γ.



(a) Normal Scale (b) Logarithmic Scale

Figure 3: Convergence of the Controlled Noise Insertion run on the random geometric graph for different values of ϕ_i, controlled by γ.

C.2 Impact of varying ϕ_i

In this experiment, we demonstrate the practical utility of letting the rate of decay ϕ_i to be different on each node i. In order to do so, we run the experiment on the random geometric graph and compare the settings investigated in the previous two experiments — the noise decay rate driven by ϕ, or by γ.

In first place, we choose the values of ϕ_i such that that the two factors in Corollary 3 are equal. For the particular graph we used, this corresponds to $\gamma \approx 0.17$ with $\phi_i = \sqrt{1 - \frac{\alpha(\mathcal{G})}{2d_i}}$. Second, we make the factors equal, but with constraint of having ϕ_i to be equal for all i. This corresponds to $\phi_i \approx 0.983$ for all i.

The performance for a large number of iterations is displayed in the left side of Figure 4. We see that the above two choices indeed yield very similar practical performance, which also eventually matches the rate predicted by theory. For a complete comparison, we also include the performance of the Standard Gossip algorithm (Baseline).

The important message is conveyed in the histogram in the right side of Figure 4. The histogram shows the distribution of the values of ϕ_i for different nodes i. The minimum of these values is what we needed in the case of identical ϕ_i for all i. However, most of the values are significantly higher. This means, that if we allow the noise decay rates to depend on the number of neighbours, we are able to increase the amount of noise inserted, without sacrificing practical performance. This is beneficial, as more noise will likely be beneficial for any formal notion of protection of the initial values.

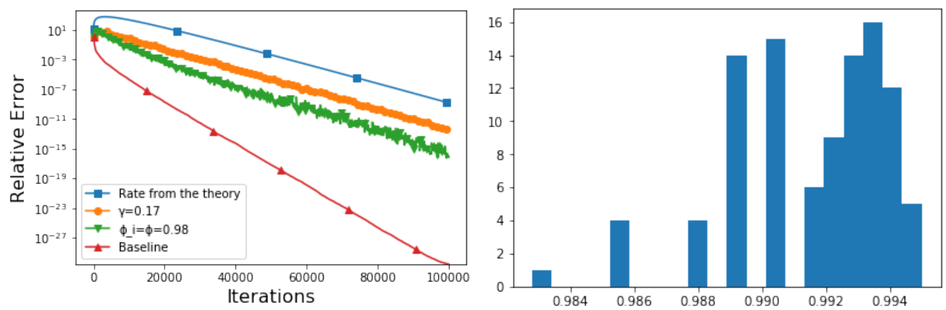


Figure 4: Left: Performance of the noise oracle with noise decrease rate chosen according to Corollary 3. Right: Histogram of of distribution of ϕ_i