# Optimally Improving Cooperative Learning in a Social Setting

**Shahrzad Haddadan** [1] **Cheng Xin** [2] **Jie Gao** [2]

## Abstract

We consider a cooperative learning scenario where a collection of networked agents with individually owned classifiers dynamically update their predictions, for the same classification task, through communication or observations of each other's predictions. Clearly if highly influential vertices use erroneous classifiers, there will be a negative effect on the accuracy of all the agents in the network. We ask the following question: how can we optimally fix the prediction of a few classifiers so as maximize the overall accuracy in the entire network. To this end we consider an aggregate and an egalitarian objective function. We show a polynomial time algorithm for optimizing the aggregate objective function, and show that optimizing the egalitarian objective function is NP-hard. Furthermore, we develop approximation algorithms for the egalitarian improvement. The performance of all of our algorithms are guaranteed by mathematical analysis and backed by experiments on synthetic and real data.

## 1. Introduction

With the breakthrough of AI technologies and the availability of big data, we are witnessing a flourish of AI models that are owned by different entities and trained by using private or proprietary data, even for generic purposes such as voice recognition, natural language processing or image segmentation. These models do not stay in isolation. There is a natural opportunity for these models to interact with each other and collectively improve their performance.

One of the concrete application scenarios is in cybersecurity, where security agents in a network collectively detect anomalous traffic patterns that are potentially associated with cyber attacks. These security agents may be affiliated with different entities in the network. They may or may not be able to directly share their collected data due to privacy or other logistic reasons but can share their beliefs on whether the network is under attack or not, and if so, which type of attack. In such a scenario, improving the model of one agent by collecting more data, recruiting domain experts to provide high quality labels of observation data, or retraining using a more powerful model can help to improve the quality of prediction locally. As these security agents stay on a network, they naturally have the opportunity to exchange their assessments. The quality improvement at one agent spills to other agents in the network.

Another application scenario is in online social networks. With generative AI, it is now a lot easier to create fake contents such as images and videos. Consider an online social network in which agents share pictures and comment on them, e.g., Instagram. Assume that some of these pictures may be AI generated. While everyone can have his or her opinion on whether a wide-spreading picture is real or fake, the participants who are more skilled in image generation or who have access to powerful models can help the rest of the network to discern real ones and fake ones.

It is not new to utilize data and models in a distributed setting. With the explosion of personal data from smartphones and wearable devices and the increasing awareness of data privacy, decentralized learning, as opposed to users sharing their data to a central enterprise, has gained popularity in recent years. Federated learning and gossip learning are such examples. However in this scenarios, the decentralized agents are still tightly coupled – they use the same model architecture and sometimes exchange gradient or model parameters directly. In our work, we consider a loosely coupled, cooperative setting where agents share a general task requirement and they select individual model parameters and architecture, train their models on their private data. Such agent autonomy is a necessity when the agents are not affiliated with the same ownership. In addition, we consider the agents sharing their *predictions* with other agents preferably those within proximity or with established trust relationships. We call this model *cooperative learning in a social setting*. Essentially each agent has her own view of the world and through exchanging predictions with others, collectively we hope to improve the accuracy for all agents.

[1]Rutgers Business School, Piscataway, NJ, USA [2]Department of Computer Science, Rutger University, Piscataway, NJ, USA. Correspondence to: Shahrzad Haddadan <shaddadan@business.rutgers.edu>, Jie Gao <jg1555@rutgers.edu>.

*Proceedings of the $41^{st}$ International Conference on Machine Learning*, Vienna, Austria. PMLR 235, 2024. 

When only predictions from other agents are shared, it is up to the individual agent to decide how to update their models. On a first-order basis, we can assume for now that the outcome of such information sharing can be approximated by a weighted linear combination of the agent $u$'s own assessment and the predictions from other neighbors. The weights can be either fixed or time-varying, e.g., based on the trust level of neighbors. This leads to two natural models, DeGroot (1974) and Friedkin & Johnsen (1990), originally proposed in modeling opinion dynamics in social network. In DeGroot model, each agent's prediction is a simple weighted linear combination of neighbors' predictions. When the weight coefficients are fixed (or when the updates are frequently enough for time-varying weights), all agents converge to global consensus. In FJ model, each agent incorporates in the update step a vector of personal assessment (which can be guided by the difference of local data distribution). The model still converges, and each agent arrives at possibly different predictions, reflecting adaptation to individual local data distributions.

The generic framework captures how a group of networked agents build their respective models collectively. When new training data is introduced to one agent in the system, the other agents indirectly receive the benefit of it. Therefore it is an interesting question to analyze the collective benefit and also ask the optimization question of *where to inject new training data to maximize the collective benefit*. This is the research question we focus on in this paper.

### 1.1. Related Work

**Decentralized learning** Training a single high-quality global model using decentralized data and computation has been studied extensively in decentralized optimization (e.g., for kernel methods (Colin et al., 2016), PCA (Fellus et al., 2015), stochastic gradient descent (Blot et al., 2016), multi-armed bandit (Lazarsfeld & Alistarh, 2023) and generalized linear models (He et al., 2018)). Federated learning (Konečný et al., 2015; 2016a;b; McMahan et al., 2017; Krishnan et al., 2024) uses a client-server architecture and considers multiple local models, trained using respective local data, with model parameters aggregated and shared through a central global model. Gossip learning (Ormándi et al., 2013; He et al., 2018; Hegedűs et al., 2019; 2016; Giaretta & Girdzijauskas, 2019), on the other hand, does not assume a central node. Instead, each node updates its own local parameters via training and then its updated parameters are shared by information exchange with other nodes in the network. This setting removes the single point of failure in the system and thus is more robust and scalable, without compromised performance (Hegedűs et al., 2016; 2021). These gossip learning protocols consider the exchange of local models (or crucial parameters such as local gradients) directly. This requires that all agents participating in gossip learning use the same type of models, which is a limitation. It is also known that models or gradients can reveal knowledge of the training data (thus raising concerns to data privacy, see a recent survey here (Zhang et al., 2023)).

**Social learning** The study of learning and decision making in a social network has been studied for longer than a decade. In these works, agents predict the status of the world, and based on their prediction they take an action to maximize a utility function. In a social setting these decisions are not made in a void, as each agent observes the prediction of her neighbors or their actions, and henceforth updates her prediction. These models are vastly studied by economists who are interested in understanding whether the agents' decisions converge to the same value (consensus), how fast is the rate of convergence, whether an equilibrium exists, and if the consensus leads agents to an optimal decision (learning) (Acemoglu et al., 2011; Arieli et al., 2021; Golub & Jackson, 2010; Golub & Sadler, 2016; Rahimian & Jadbabaie, 2017; Hązła et al., 2019; Eckles et al., 2019; Bindel et al., 2015).

Given a social network, various works tackle optimization problems in which an algorithm makes minimal changes in the network to maximize the improvement of a desirable property. For instance, various works consider the problem of maximizing information diffusion by seeding information in a number of selected source vertices (Kempe et al., 2003; Seeman & Singer, 2013; Eckles et al., 2019; Garimella et al., 2017) or by adding links (Borgs et al., 2014; D'Angelo et al., 2019). Some works optimally insert links into a network to maximize the information flow between two groups of nodes (Cinus et al., 2023; Zhu et al., 2021; Adriaens et al., 2023; Haddadan et al., 2021; 2022; Santos et al., 2021), and others optimaly alter innate opinion of users in the FJ model to reduce polarization or disagreements (Tsioutsiouliklis et al., 2022; Abebe et al., 2018; Musco et al., 2018).

Our work bridges the above lines of work. We study a framework in which agents are performing learning tasks and exchange their predictions until each makes a final decision. Unlike prior decentralized learning works, our agents do not necessarily use the same model nor they share any parameters, they solely exchange their predictions. In this sense, our framework falls into the context of social learning. However, instead of studying problems such as existence of consensus or convergence rates, we focus on the problem of optimally injecting information to a selected subset of agents to improve the overall betterness in the network. Another difference with social learning framework is that we do not consider one fixed model of information exchange, in contrast, we assume a general linear model for information exchange. Therefore, our methods are applicable to any linear model whether it describes users of a social network, each evaluating the truthfulness of an online content, or whether it describes intelligent systems in which agents

cooperate for a classification task.

### 1.2. Summary of Contributions

Motivated by prior works on decentralized and social learning, we consider a new framework called *cooperative learning in a social setting*. In this framework, we consider the problem of *optimally selecting $k$ agents and improving their innate predictions to maximize the overall network improvement*. We consider both an *aggregate* objective function and an *egalitarian* one, and assume different levels of access to the model's parameters which are (1) the joint probability distribution of the classifiers forming agents' innate predictions, denoted by $\pi$, and (2) the expressed social influence matrix, denoted by $\bar{W}$.

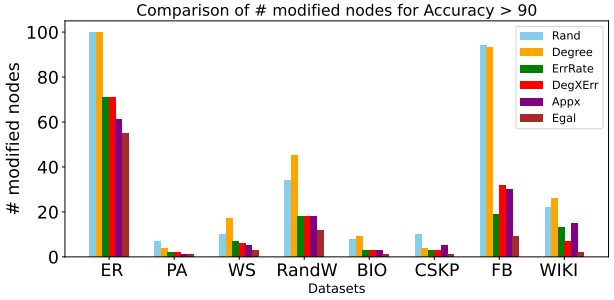
1. We provide a polynomial time algorithm for the aggregate improvement problem. This algorithm uses only the innate error rates and $\bar{W}$, and it does not need any additional knowledge of $\pi$; see Section 4.1.

2. We show that solving the egalitarian improvement problem is hard: it is NP-hard to solve it exactly even if both parameters are entirely known. We also show that if $\bar{W}$ and only the innate error rates are available, without any further assumption even finding an approximation solution is hard; see Theorems A.3 and A.4 .

3. We provide two approximation algorithms for the egalitarian improvement problem. The first algorithm, EgalAlg is a greedy algorithm and needs full access to $\pi$. The second one, $\mathsf{EgalAlg}^{(\mathrm{appx})}$, approximates the greedy choice and only needs access to the innate error rates, but it assume that the innate predictions are pairwise independent. We show that with some modifications $\mathsf{EgalAlg}^{(\mathrm{appx})}$ works under the assumption that the vertices have group dependency; see Section 4.2.

4. We compare the two algorithms for egalitarian improvement by running experiments on real and synthetic graphs and compare their performance to four benchmark methods. Our experiments show that by modifying only a few vertices, we succeed in increasing the accuracy of a high percentage of network's agents; see Figure 1.

Figure 1: Comparison of # modified nodes for Accuracy $> 90\%$ on different dataset (lower is better).




## 2. Models and Problem Definition

Consider a set $\Omega$ whose elements are labeled as $+1$ or $-1$, i.e., there exists a function $y : \Omega \to \{-1, +1\}$ such that for any $a \in \Omega$, $y(a) \in \{-1, +1\}$ is the (true) label of $a$. Consider a set of agents $V = \{v_1, v_2, \ldots, v_n\}$ who lack access to the true labels. Given $a \in \Omega$, each agent $v_i$ uses a classifier $\hat{y}_i$ to predict the label of $a$, i.e., we have $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ where for any $i$, $\hat{y}_i : \Omega \to \{-1, +1\}$. The innate error rate of each agent $v_i$ is defined as:

$$\mathsf{err}(v_i) = \mathop{\mathbb{E}}_{a\sim\Omega}\left[\mathbf{1}\left(\hat{y}_i(a) \neq y(a)\right)\right] = \mathop{\mathbb{P}}_{a\sim\Omega}\left(\hat{y}_i(a) \neq y(a)\right).$$

We assume that this error is independent of true label. This assumption merely makes our analysis simpler and without it our results can be generalized with trivial modifications.

These predictions form the *innate assessment* of the agents, which we denote by $z^{(0)}(i, a)$, for arbitrary $i$ and $a$; thus, $z^{(0)}(i, a) = \hat{y}_i(a)$. After, agents communicate with each other through a *time varying averaging process*. At each point of time, each agent $v_i$ communicates with a subset of other agents $\mathcal{N}^{(t)}(v_i)$, and her assessment will update as:

$$z^{(t)}(i, a) = C_i \cdot \left( \sum_{v_j \in \mathcal{N}^{(t)}(v_i)} W^{(t)}_{i,j} z^{(t-1)}(j, a) + \alpha_i \hat{y}_i(a) \right),$$

where $W^{(t)}_{i,j}$ denotes the influence of $v_j$ on $v_i$ at time $t$, $\alpha_i$ is $v_i$'s stubbornness, $C_i$ is a normalizing constant and $z^{(t-1)}(i, a)$ and $z^{(t)}(i, a)$ respectively denote $v_i$'s assessment before the $t$th round of communication and after it.

Formally, given $a \in \Omega$, let $\hat{\mathbf{y}}(a) = (\hat{y}_1(a), \hat{y}_2(a), \ldots \hat{y}_n(a))$ and $\mathbf{z}^{(0)}(a) = \hat{\mathbf{y}}(a)$. Assume having a sequence of $n \times n$ matrices $\left(W^{(t)}\right)_{t=1}^{\infty}$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and $\boldsymbol{C^{(t)}} = (C_1^{(t)}, C_2^{(t)}, \ldots, C_n^{(t)})$. For any $t \geq 1$ we have:

$$\mathbf{z}^{(t)}(a) = \boldsymbol{C^{(t)}} \odot \left( W^{(t)} \mathbf{z}^{(t-1)}(a) + \boldsymbol{\alpha} \odot \hat{\mathbf{y}}(a) \right), \quad (1)$$

where $\odot$ denotes the Hadamard product which is defined to be the vector (matrix) obtained from the pairwise product of the elements in two other vectors (matrices). [1] Let $\mathbf{z}^*(a) = (z^*(1, a), z^*(2, a), \ldots, z^*(n, a))$ be the vector that the above process converges to i.e.,

$$\mathbf{z}^*(a) = \lim_{t\to\infty} \mathbf{z}^{(t)}(a).$$

We call $\mathbf{z}^*(a)$ the *the expressed prediction* vector. We assume that this limit has the following closed form:

$$\exists \bar{W} \in \mathbb{R}^{+^{n\times n}} \text{ s. t. } \forall a \in \Omega,\ \mathbf{z}^*(a) = \bar{W}\hat{\mathbf{y}}(a)^{\mathsf{T}}. \quad (2)$$

[1] Let $A, B \in \mathbb{R}^{n\times m}$, and $C = A \odot B$. We have that $C \in \mathbb{R}^{n\times m}$ and $C_{ij} = A_{ij} \cdot B_{ij}$

$\mathbb{R}^+$ being the set of all non-negative real numbers. We call $\bar{W}$ the *expressed (social) influence* matrix.

The above assumption is indeed proven to hold true in many well studied models. To name a few:

**DeGroot Model (DeGroot, 1974)** Given a row-stochastic matrix $W \in \mathbb{R}^{+n\times n}$, DeGroot Model evolves as:

$$\mathbf{z}^{(t)}_{\text{DeGroot}}(a) = W\mathbf{z}^{(t-1)}_{\text{DeGroot}}(a),\ \mathbf{z}^{(0)}_{\text{DeGroot}}(a) = \hat{\mathbf{y}}(a),$$

It is known that

$$\mathbf{z}^{*}_{\text{DeGroot}}(a) = \Pi\hat{\mathbf{y}}(a)\ ,$$

where $\Pi$ is a block diagonal matrix with blocks $\Pi_1, \Pi_2, \ldots, \Pi_k$, and each block corresponds to a strongly connected component of the graph corresponding to $W$. In each block all rows are identical.

**Friedkin–Johnsen (FJ) Model (Friedkin & Johnsen, 1990)** Given an arbitrary matrix with non-negative weights $W$, for each $i \in [n]$ let $C_i = 1/(1 + \sum_{j\in\mathcal{N}(i)} W_{i,j})$. FJ model evolves as follows:

$$\mathbf{z}^{(t)}_{\text{FJ}}(a) = \boldsymbol{C} \odot \left(W\mathbf{z}^{(t-1)}_{\text{FJ}}(a) + \hat{\mathbf{y}}(a)\right),\ \mathbf{z}^{(0)}_{\text{FJ}}(a) = \hat{\mathbf{y}}(a).$$

It is known that if $W$ is symmetric the FJ model converges to the following closed form:

$$\mathbf{z}^{*}_{\text{FJ}}(a) = (I + L)^{-1}\hat{\mathbf{y}}(a),$$

where $L$ is the combinatorial Laplacian of the indirected graph corresponding to $W$ and $I$ denotes the identity matrix.

Both DeGroot and FJ model use a constant matrix $W$ in their evolution. The following shows a time varying evolution:

**Finite time Models** Assume a finite sequence of row stochastic matrices $W^{(1)}, W^{(2)}, \ldots, W^{(T)}$, for any $t > T$, let $W^{(t)}$ be the identity matrix and let $\boldsymbol{\alpha}$ be all zeros vector. It is straightforward to see that the general equation of Equation (1) will converge to

$$\mathbf{z}^{*}_{\text{Finite}}(a) = \bar{W}\hat{\mathbf{y}}(a),\ \bar{W} = \prod_{i=1}^{T} W^{(i)}\ .$$

### 2.1. Statement of Problems

Assume any time varying averaging model which converges to a close form as Equation (2). Thus, for a given $a \in \Omega$ and $v_i \in V$ the *expressed* prediction of $v_i$ on $a$ is equal to

$$z^*(i,a) = \sum_{j=1}^{n} \bar{W}_{ij}\hat{y}_j(a)\ .$$

Note that the value of $z^*(i,a)$ is a function of the expressed social influence matrix $\bar{W}$ as well as the quality of all agents' classifiers which is formulated in the joint probability distribution of $\hat{\mathbf{y}}$. If the true label $y(a)$ equals $+1$, the positive values of $z^*(i,a)$ are preferable, otherwise negative values of $z^*(i,a)$ are better. In other words, we would like $y(a)$ and $z^*(i,a)$ to have the same sign. We define $\mathcal{Z}(i,a) : V \times \Omega \to [-1,1]$ as follows:

$$\mathcal{Z}(i,a) = y(a)\cdot z^*(i,a)\ .$$

The larger values for $\mathcal{Z}(i,a)$ correspond to the fact that agent $v_i$ makes a good prediction on an object $a$. If $\mathcal{Z}(i,a) < 0$ we consider this prediction *faulty*.

**Improving quality of selected classifiers** Assume that we have a tool to improve the quality of classifiers. Formally, let $\varphi \in (0,1]$ be a given constant. For any arbitrary agent $v_i$ we may improve $\hat{y}_i$ to $\tilde{y}_i : \Omega \to [0,1]$ as follows:

$$\forall a \in \Omega,\ \tilde{y}_i(a) = (1-\varphi)\hat{y}_i(a) + \varphi y(a)\ . \tag{3}$$

We would like to improve the quality of a selected subset of agents' classifiers (innate predictions) to maximize the overall quality of expressed predictions among all agents. Formally we are interested in selecting a subset $S$ of $k$ agents, i.e., $S \subseteq V$, $|S| = k$ and improve the quality of the classifier as follows:

$$\forall a \in \Omega,\ \tilde{y}_i(a) = \begin{cases} (1-\varphi)\hat{y}_i(a) + \varphi y(a) & \text{if } v_i \in S \\ \hat{y}_i(a) & \text{if } v_i \notin S \end{cases} \tag{4}$$

Let

$$z^*_{\text{new}}(i,a) = \sum_{j=1}^{n} \bar{W}_{ij}\tilde{y}_j(a)\ \&\ \mathcal{Z}_{\text{new}}(i,a) = y(a)\cdot z^*_{\text{new}}(i,a)\ .$$

In the first problem that we study, $S$ is selected to maximize an *aggregate* objective function:

$$\mathcal{G}^{(\text{agg})}(S) \triangleq \mathop{\mathbb{E}}_{a\sim\Omega}\left[\sum_{i=1}^{n} \mathcal{Z}_{\text{new}}(i,a) - \mathcal{Z}(i,a)\right] \tag{5}$$

The above objective function is great, but it has the shortcoming of any other aggregate objective function: it is possible that one agent benefits enormously from it at the cost of many other agents getting extremely little.

Therefore, we also study the following *egalitarian* objective function in which we count the expected *number* of agents whose faulty predictions will improve.

$$\mathcal{G}^{(\text{egal})}(S) \triangleq \mathop{\mathbb{E}}_{a\sim\Omega}\left[\sum_{i=1}^{n} \mathbf{1}(\mathcal{Z}(i,a)<0 \wedge \mathcal{Z}(i,a)<\mathcal{Z}_{\text{new}}(i,a))\right].$$

We are now ready to formally state the problems.

*Problem* 1 (Aggregate improvement through improving $k$ selected agents). What is an optimal way to select a subset $S \subseteq V$ and update their innate predictions as Equation (4) to maximize the following objective function:

$$\text{OPT}^{(\text{agg})} = \max_{S\subseteq V;|S|=k} \mathcal{G}^{(\text{agg})}(S)\ .$$

*Problem* 2 (Egalitarian improvement through improving $k$ selected agents). What is an optimal way to select a subset $S \subseteq V$ and update their innate predictions as Equation (4) to maximize the following objective function:

$$\mathrm{OPT}^{(\mathrm{egal})} = \max_{S \subseteq V; |S|=k} \mathcal{G}^{(\mathrm{egal})}(S) \, .$$

The above process has two main parameters: a joint probability distribution $\pi : \Omega \times \{-1,+1\}^V \to [0,1]$, where for any $a \in \Omega$ and $\vec{b} \in \{-1,+1\}^n$, $\pi(a,\vec{b}) = \mathbb{P}(\wedge_{i=1}^n \hat{y}_i(a) = b_i)$ as well as $\bar{W}$ which is the expressed influence matrix. While in most applications $\bar{W}$ is either available in closed form (e.g., for DeGroot, FJ or finite models) or it can be approximated using iterative methods, our access to $\pi$ depends on assumptions which may vary depending on our application. In fact, we present our algorithms assuming different access levels to the joint probability distribution $\pi$.

## 3. Summary of Results

In this section, we present a summary of our main results. Let us first list the assumptions we make on the access level to $\pi$ and $\bar{W}$:

**Assumption 3.1** (Best scenario)**.** Assume having *complete* knowledge of $\pi$.

The above assumption is reasonable if $\Omega$ is a small finite set. For instance in cases where $\Omega$ can be partitioned to a few types and the agents make the same predictions on any element of the same type, e.g., see (DeMarzo et al., 2003; Hązła et al., 2023; Gaitonde et al., 2021)

Clearly, it is possible that the above assumption does not hold true. However, the algorithm needs to have *some knowledge* of the probability distribution of innate predictions.

**Assumption 3.2** (some knowledge of $\pi$)**.** Assume having some knowledge of classifiers' dependence/independence and the innate error rates.

**Assumption 3.3** (minimum knowledge of $\pi$)**.** Assume having only knowledge of innate error rates $\{\mathsf{err}(v_i)\}_{v_i \in V}$.

The summary of our main results is that Problem 1 is easy, i.e., we show an exact polynomial time algorithm for it assuming minimum knowledge of $\pi$. On the other hand, we show that Problem 2 is hard, i.e., we show that exactly solving it is NP-complete even assuming full knowledge of $\pi$. Note that this hardness also holds for the cases in which we have less knowledge of $\pi$. Later, we show greedy algorithms for approximately solving it under different assumptions.

Initially, in all of our results we assume access to the closed form of $\bar{W}$. We then show that the guarantees still hold with a slight change when an approximation of $\bar{W}$ is given.

**Assumption 3.4** (knowledge to an approximation of $\bar{W}$)**.** Assume having knowledge of $\widehat{\bar{W}}$ such that

$$\left| \widehat{\bar{W}} - \bar{W} \right|_1 \leq \epsilon \, ,$$

where $|\cdot|_1$ is the $\ell_1$ norm and $\epsilon$ a precision parameter.

### 3.1. Optimizing the aggregate objective function

**Theorem 3.5.** *There is an algorithm with run-time complexity $\Theta\left(n^2\right)$ which given $\bar{W}$ and $\{\mathsf{err}(v_i)\}_{i=1}^n$ as input parameters outputs $S$ such that $\mathcal{G}^{(\mathrm{agg})}(S) = \mathrm{OPT}^{(\mathrm{agg})}$.*

*Remark* 3.6. Let ALG be the algorithm whose performance guarantees are presented in Theorem 3.5. Let $S$ be the output of ALG when $\bar{W}$ is given to it as input parameter, and let $S'$ be the output if $\widehat{\bar{W}}$ is given. We have:

$$\mathcal{G}^{(\mathrm{egal})}(S') \geq \mathcal{G}^{(\mathrm{egal})}(S) - 2k\epsilon \, .$$

We present this algorithm in Section 4.1 and Appendix A.1.

### 3.2. Optimizing the egalitarian objective function

We now present our results about Problem 2.

We call a matrix with no negative entry *non-negative*. In DeGroot model if $W$ is non-negative, $\bar{W}$ is also non-negative and in FJ model if $W$ is non-negative and symmetric, $\bar{W}$ is also non-negative (Chebotarev & Shamis, 2006) .

In this section we assume that $\bar{W}$ is non-negative.

**Theorem 3.7.** *Under Assumption 3.1 and assuming $|\Omega|$ is polynomial in $n$ , Problem 2 is NP-hard.*

Since Problem 2 is NP-hard, we concentrate on finding an approximation algorithm for it in different scenarios.

**Approximate solution with full access to $\pi$.** Assume that $\Omega$ is finite and for any $a \in \Omega$ and $b \in \{-1,+1\}^n$ we can evaluate the probability of the event $\bigwedge_{i=1}^n \hat{y}_i(a) = b(i)$.

**Theorem 3.8.** *There is a greedy algorithm with runtime $\Theta\left(|\Omega| n^2 k\right)$ which by receiving $\pi$ and $\bar{W}$ as input parameters outputs $S$ satisfying*

$$\mathcal{G}^{(\mathrm{egal})}(S) \geq (1 - 1/e)\mathrm{OPT}^{(\mathrm{egal})}.$$

A pseudocode of our algorithm, EgalAlg is presented in Appendix A.3, an overview of main ideas is presented Section 4.2. Note that the runtime of EgalAlg grows linearly in $|\Omega|$. Later, we present EgalAlg$^{(\mathrm{appx})}$whose complexity *does not* grow with $|\Omega|$ (see Section 4.2). Therefore, even if $\pi$ is fully known, by employing EgalAlg$^{(\mathrm{appx})}$, we may prefer to suffice to a lower quality approximation to gain better time complexity.

**Approximate solution with minimum information** While the previous result shows a constant approximation, the following theorem shows that by only having the innate error rates $\{\mathsf{err}(v_j)\}_{j=1}^n$ and $\bar{W}$ we are not able to achieve any good approximation.

**Theorem 3.9.** *Any solution to Problem 2 which only uses $\bar{W}$ and innate error rates $\{\mathsf{err}(v_j)\}_{j=1}^n$ makes an error which can be as large as $\Omega(n)$.*

We restate and prove the above theorem in Theorem A.4.

Motivated by the above results, we now present our results when more knowledge about the classifiers are available. For instance, in addition to knowing the error rates we assume that the classifiers are pairwise independent.

Using these assumptions we design $\mathsf{EgalAlg}^{(\mathrm{appx})}$ and show theoretical guarantees for its performance.

**Approximate solution assuming independence.** Assume that the classifiers $\{\hat{y}_i\}_{v_i \in V}$ are pairwise independent. We have:

**Theorem 3.10.** *There is a greedy algorithm, $\mathsf{EgalAlg}^{(\mathrm{appx})}$, with run-time $\Theta(n^3 k)$ which by receiving $\{\mathsf{err}(v_i)\}_{v_i \in V}$ and $\bar{W}$ as input parameters outputs $S$ satisfying*

$$\mathcal{G}^{(\mathrm{egal})}(S) \geq [(1-1/e) - \Delta^{\mathrm{ind}}] \cdot \mathrm{OPT}^{(\mathrm{egal})} ,$$

*where $\Delta^{\mathrm{ind}}$ is a parameter depending on the network. If the network is* nicely structured *$\Delta^{\mathrm{ind}} = o(1)$; see Section 4.2.1 .*

We generalize the assumption of pairwise independence to group dependence as follows:

**Approximate solution assuming group dependence.** Assume that some agents belong to opposing groups $\mathcal{R}$ and $\mathcal{B}$ and some agents are colorless; they are in $\mathcal{W}$. The classifiers of the agents in $\mathcal{W}$ are independent, and the classifiers of $\mathcal{R}$ and $\mathcal{B}$ agents have positive intra-correlation and negative inter-correlation as described in Definition 4.7. This model describes a situation when have a classification task that can be influenced by an exogenous factor, e.g, their position or political leaning. Clearly by setting $V = \mathcal{W}$ we will have the previous model. In this case we have:

**Theorem 3.11.** *There is a greedy algorithm, $\mathsf{EgalAlg}^{(\mathrm{appx})}$, with run-time $\Theta(n^3 k)$ which by receiving* individual and group *error rates and $\bar{W}$ as input parameters outputs $S$ satisfying*

$$\mathcal{G}^{(\mathrm{egal})}(S) \geq [(1-1/e) - \Delta^{\mathrm{gr}}] \cdot \mathrm{OPT}^{(\mathrm{egal})} ,$$

*where $\Delta^{\mathrm{gr}} \geq \Delta^{\mathrm{ind}}$ is a parameter depending on the network and the dominance of colors $\mathcal{R}$ and $\mathcal{B}$ on other agents. Not surprisingly, $\Delta^{\mathrm{gr}}$ becomes closer to $\Delta^{\mathrm{ind}}$ as the number of colorless agents increases. If the network is* nicely structured *and* nicely colored *$\Delta^{\mathrm{gr}} = o(1)$; See Section 4.2.2 .*

The following remark holds ture both under pairwise Independence and group dependency:

*Remark* 3.12. Let $S$ be the output of $\mathsf{EgalAlg}^{(\mathrm{appx})}$ when $\bar{W}$ is given to it as input parameter, and let $S'$ be the output if $\widehat{\bar{W}}$ is given. We have:

$$\mathcal{G}^{(\mathrm{egal})}(S') \geq \mathcal{G}^{(\mathrm{egal})}(S)(1 - 8k\epsilon) .$$

## 4. Algorithms

In this section we present our algorithms. All of the algorithms are greedy. We provide an exact solution for Problem 1 in Section 4.1. In Section 4.2, because of the NP-harness of Problem 2, we present a *constant* approximation algorithm for it; we call this algorithm $\mathsf{EgalAlg}$ . We then present $\mathsf{EgalAlg}^{(\mathrm{appx})}$ which has a lower time complexity but worst approximation guarantees assuming pairwise independence of agents. In this case, our approximation ratio depends on the network structure. In Section 4.2.2, we modify $\mathsf{EgalAlg}^{(\mathrm{appx})}$ so that it works under a milder assumption formalized as *group dependency*.

Pseudocodes for our egalitarian algorithms are presented in Appendix A.3

### 4.1. The aggregate objective function

For any vertex $v_j$ let's define the influence score and its approximation by $\mathrm{Inf}(v_j) = \sum_{i=1}^n \bar{W}_{ij}\mathsf{err}(v_i)$ . The following lemma is proven in Appendix A.1.

**Lemma 4.1.** *Let $U = \{u_1, u_2, \ldots, u_k\}$ be top-$k$ vertices with highest value of $\sum_{u_j \in U} \mathrm{Inf}(u_j)$. We have that:*

$$\mathcal{G}^{(\mathrm{agg})}(U) = \mathrm{OPT}^{(\mathrm{agg})} .$$

**Proof of Theorem 3.5 and Remark 3.6** With the above lemma, we design an algorithm that for all $v_j$s calculates their influence score, and outputs the top-$k$. The complexity of such algorithm is $\Theta(n^2 + n\log n + k) = \Theta(n^2)$. In Appendix A.1 we also prove Remark 3.6.

### 4.2. The egalitarian objective function

Optimizing the egalitarian function is NP-hard (See Theorem A.3). We show that $\mathcal{G}^{(\mathrm{egal})} : 2^V \to [0, n]$ is monotonic and sub-modular (See Lemmas A.5 and A.6). Thus, a greedy algorithm will provides a $(1-1/e)$ approximation (Nemhauser & Wolsey, 1978).

We now concentrate on obtaining the greedy choice. Formally, we define the function $\mathsf{gr}(S) : 2^V \to V$ as follows:

$$\mathsf{gr}(S) = \operatorname*{argmax}_{u \in V} \mathcal{G}^{(\mathrm{egal})}(S \cup \{u\}) - \mathcal{G}^{(\mathrm{egal})}(S) . \quad (6)$$

The following lemma is proven in Appendix A.3.1:

**Lemma 4.2.** *For any $S \subseteq V$ we have:*

$$\mathsf{gr}(S) = \operatorname*{argmax}_{u \in V} \sum_{\substack{i=1:n \\ \bar{W}_{iu} \neq 0}} \Delta\mathcal{G}_i(S,u) \;, \tag{7}$$

*where $\Delta\mathcal{G}_i(S,u)$ is defined to be*[2]

$$\mathop{\mathbb{P}}_{a\sim\Omega}\left(\mathcal{Z}(i,a) \leq 0 \wedge \left(\bigwedge_{\substack{v_j \in S \\ \bar{W}_{ji}\neq 0}} y(a)=\hat{y}_j(a)\right) \wedge y(a)\neq\hat{y}_u(a)\right)$$

**Proof of Theorem 3.8** Our proposed algorithm, EgalAlg starts with $S = \emptyset$. Iteratively, $\mathsf{gr}(S)$ is added to $S$ until $|S| = k$. Assume that $\Omega$ is finite and we have access to $\pi$. Using Lemma 4.2 we obtain the greedy choice as follows: for any $a \in \Omega$ and $v_i \in V$, we evaluate the validity of the event $\left(\bigwedge_{\substack{v_j\in S \\ \bar{W}_{ji}\neq 0}} y(a)=\hat{y}_j(a)\right) \wedge y(a)\neq\hat{y}_u(a)$. If this event is valid, we calculate $\mathcal{Z}(i,a)$ and verify $\mathcal{Z}(i,a) < 0$ which takes $n$ steps. We find best $u$ by using Equation (7) and iterating over all choices of $i \in V$ and $a \in \Omega$. The total runtime for $k$ iterations is $\Theta(|\Omega|n^2k)$.

#### 4.2.1. Independent classifiers

In this section we consider the case where the classifiers $\{\hat{y}_i\}_{v_i\in V}$ are pairwise independent which falls into the scenario in which we have *some* knowledge of $\pi$ (Assumption 3.2). In this case we may estimate the greedy choice as a function of $\{\mathsf{err}(v_i)\}_{i=1}^n$ and $\bar{W}$.

Let's state the main result of this section and then we present the steps that lead us to the selection of the greedy choice:

**Theorem 4.3.** *Let $S$ be the output of a greedy algorithm which starts by taking $S = \emptyset$ and for $k$ steps keeps updating $S$ to $S \cup \{g\}$ where*

$$g = \operatorname*{argmax}_{u} \sum_{\substack{i=1:n \\ \bar{W}_{iu}\neq 0}} \widehat{\Delta\mathcal{G}}_i(S,u) \;,$$

*and $\widehat{\Delta\mathcal{G}}_i(S,u)$ is defined in Lemma 4.6, we have:*

$$\mathcal{G}^{(\mathrm{egal})}(S) \geq [(1-1/e) - \Delta^{\mathrm{ind}}] \cdot \mathrm{OPT}^{(\mathrm{egal})} \;,$$

*with $\Delta^{\mathrm{ind}} = \Theta(|\mathcal{A}|)$ and $\mathcal{A}$ is the set of ambiguous vertices.*

**Ambiguous vertices** Consider the partitioning of $V$ with $V^+$ as low error vertices and $V^-$ as high error vertices:

$$V^+ = \{v_j \mid \mathsf{err}(v_j) \leq 1/2\} \;\&\; V^- = \{v_j \mid \mathsf{err}(v_j) > 1/2\}$$

[2] We use $\wedge$ and $\vee$ to denote respectively the logical operations conjunction and disjunction which are *"and"* and *"or"*.

with respect to this partition we define the following vectors whose elements are in $[0,1]$:

$$\mathcal{E}^+ = (1-2\mathsf{err}(v_j))_{v_j\in V^+} \;\&\; \mathcal{E}^- = (2\mathsf{err}(v_j)-1)_{v_j\in V^-}$$

For any arbitrary vertex $v_i \in V$, low error and high error vertices contribute in the value of $\mathcal{Z}(i,a)$ through the following coefficient vectors:

$$\bar{W}_i^+ = (\bar{W}_{ij})_{v_j\in V^+} \quad \&\; \bar{W}_i^- = (\bar{W}_{ij})_{v_j\in V^-}$$

The ambiguous vertices are those who *are not* dominated by neither $V^+$ or $V^-$. Formally,

**Definition 4.4.** [Ambiguous vertices] Let $\bar{W}_i = (\bar{W}_{i1}, \bar{W}_{i2}, \ldots, \bar{W}_{in})$, $|\cdot|_2$ denote the $\ell_2$ norm and $\langle\cdot,\cdot\rangle$ dot product. We call a vertex $v_i \in V$ *ambiguous* if it satisfies:

$$\left|\frac{\langle \bar{W}_i^+, \mathcal{E}^+\rangle}{|\bar{W}_i|_2} - \frac{\langle \bar{W}_i^-, \mathcal{E}^-\rangle}{|\bar{W}_i|_2}\right| \leq 4\sqrt{\log n} \;.$$

A network is *nicely structured* if it has no ambiguous vertex.

If a vertex is *non-ambiguous* we can estimate the gain associated to it very precisely:

**Lemma 4.5.** *Let $\Delta\mathcal{G}_i(S,u)$ and $\widehat{\Delta\mathcal{G}}_i(S,u)$ be as defined respectively as in Lemma 4.2 and Lemma 4.6. If a vertex is* non-ambiguous *we have:*

$$\left|\Delta\mathcal{G}_i(S,u) - \widehat{\Delta\mathcal{G}}_i(S,u)\right| \leq o(n^{-1}) \;.$$

We now present the following lemma related to the approximation of greedy choice. All the proofs and details are presented in Appendix A.4.

**Lemma 4.6.** *Let $\Delta\mathcal{G}_i(S,u)$ be as Lemma 4.2. Let $\widehat{\Delta\mathcal{G}}_i(S,u) : 2^V \times V \to [0,1]$ be defined as follows:*

$$\widehat{\Delta\mathcal{G}}_i(S,u) \triangleq \mathbf{1}\left(\Psi_i(S,u) < 0\right)\mathsf{err}(u) \prod_{\substack{v_j\in S \\ \bar{W}_{ij}\neq 0}} (1-\mathsf{err}(v_j)) \;.$$

*We have:*

$$\left|\widehat{\Delta\mathcal{G}}_i(S,u) - \Delta\mathcal{G}_i(S,u)\right| \leq \exp\left(-\frac{\Psi_i(S,u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right) \;,$$

*where*

$$\Psi_i(S,u) = -\bar{W}_{iu} + \sum_{v_j\in S} \bar{W}_{ij} + \sum_{\substack{j=1:n \\ j\notin S\cup\{u\}}} \bar{W}_{ij}[1-2\mathsf{err}(j)] \;.$$

**Proof of Theorem 3.10 and Remark 3.12** A complete pseudocode of our proposed algorithm, $\mathsf{EgalAlg}^{(\mathrm{appx})}$, is presented in Appendix A.3 (Algorithm 2). It is easy to see that the runtime is dominated by $\Theta(n^3k)$. Note that the correctness of Theorem 3.10 is directly concluded from Theorem 4.3 by setting $|\mathcal{A}| = 0$. We present the proof of Theorem 4.3 and Remark 3.12 in Appendix A.4.2.

#### 4.2.2. GROUP DEPENDENT CLASSIFIERS

We now consider a case when agents are either *red*, *blue* or none (*white*). The agents who are red or blue agent either all follow a group decision, or they independently follow an individual decision. The group decision of the blue agents is always the opposite of the group decision of red agents.

**Definition 4.7.** [Group Dependence] Assume that the set of agents $V$ can be partitioned as $V = \mathcal{R} \cup \mathcal{B} \cup \mathcal{W}$. We assume a set of classifiers $\hat{y}_1^{\text{indv}}, \hat{y}_2^{\text{indv}}, \ldots, \hat{y}_n^{\text{indv}} : \Omega \to \{-1, +1\}$ which are pairwise independent. Furthermore, we assume two group classifiers $\hat{y}_{\mathcal{R}}^{\text{gr}}, \hat{y}_{\mathcal{B}}^{\text{gr}} : \Omega \to \{-1, +1\}$ which satisfy:

$$\forall a \in \Omega,\ \hat{y}_{\mathcal{R}}^{\text{gr}}(a) \neq \hat{y}_{\mathcal{B}}^{\text{gr}}(a)\ .$$

Given a constant $\rho \in [0, 1]$, these classifiers construct $\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n\}$ as follows:

With probability $\rho$ the red and blue agents follow their groups' decision, i.e.,

$$\forall v_i \in \mathcal{R},\ \hat{y}_i(a) = \hat{y}_{\mathcal{R}}^{\text{gr}}(a) \wedge \forall v_i \in \mathcal{B},\ \hat{y}_i(a) = \hat{y}_{\mathcal{B}}^{\text{gr}}(a)$$

And the white agents independently follow their individual decisions, i.e, for all $v_i \in \mathcal{W}$, $\hat{y}_i(a) = \hat{y}_i^{\text{indv}}(a)$ .

Alternatively, with probability $1 - \rho$ , all agents independently use their individual classifiers. i.e,

$$\forall v_i \in V,\ \hat{y}_i(a) = \hat{y}_i^{\text{indv}}(a)\ .$$

In the above setting we use the following notation: For any $v_j \in V$ we define $\mathsf{err}^{\text{indv}}(v_j) = \mathbb{P}\left(\hat{y}_j^{\text{indv}}(a) \neq y(a)\right)$, and

$$\mathsf{err}(\mathcal{R}) = \mathbb{P}\left(\hat{y}_{\mathcal{R}}^{\text{gr}}(a) \neq y(a)\right)\ \&\ \mathsf{err}(\mathcal{B}) = \mathbb{P}\left(\hat{y}_{\mathcal{B}}^{\text{gr}}(a) \neq y(a)\right)$$

It is immediate from the definition that $1 - \mathsf{err}(\mathcal{B}) = \mathsf{err}(\mathcal{R})$.

In this setting, the estimation of $\Delta\mathcal{G}_i(S, u)$ is more involved and is presented in Section 4.2.2. In this case, our greedy algorithm uses $\{\mathsf{err}^{\text{indv}}(v_j)\}_{j=1}^n$, $\mathsf{err}(\mathcal{R})$, $\mathsf{err}(\mathcal{B})$ and $\bar{W}$ or its approximation. The final result follows:

**Theorem 4.8.** *Let $S$ be the output of a greedy algorithm which approximates greedy choice as defined in Lemma A.12. We have:*

$$\mathcal{G}^{(\text{egal})}(S) \geq [(1 - 1/e) - \Delta^{\text{gr}}] \cdot \text{OPT}^{(\text{egal})}\ ,$$

*where $\Delta^{\text{gr}} = \Theta(\rho|\mathcal{A}^{\mathcal{W}}| + (1 - \rho)|\mathcal{A}|)$, $\mathcal{A}$ is the set of ambiguous vertices defined before and $\mathcal{A}^{\mathcal{W}}$ is the set of $\mathcal{W}$-*ambiguous *vertices.*

***$\mathcal{W}$*-Ambiguous vertices.** As in Section 4.2.2, we partition $\mathcal{W}$ to low error vertices $\mathcal{W}^+$ and high error vertices $\mathcal{W}^-$. Similarly we define $\mathcal{E}^{\mathcal{W}^+}$, $\mathcal{E}^{\mathcal{W}^-}$ and for any $v_i \in V$, $\bar{W}_i^{\mathcal{W}+}$ and $\bar{W}_i^{\mathcal{W}-}$; for details see Appendix A.5.4. We define:

**Definition 4.9.** [ $\mathcal{W}$-Ambiguous vertices ] Let $\bar{W}_i^{\mathcal{W}} = (\bar{W}_{ij})_{v_j \in \mathcal{W}}$, and $|\cdot|_2$ be the $\ell_2$ norm and $\langle \cdot, \cdot \rangle$ be dot product.

We call an agent $v_i \in V$, $\mathcal{W}$-*ambiguous* if it satisfies

$$\left| \frac{\langle \bar{W}_i^{\mathcal{W}+}, \mathcal{E}^{\mathcal{W}^+} \rangle}{|\bar{W}_i^{\mathcal{W}}|_2} - \frac{\langle \bar{W}_i^{\mathcal{W}-}, \mathcal{E}^{\mathcal{W}^-} \rangle}{|\bar{W}_i^{\mathcal{W}}|_2} \right| \leq 4\sqrt{\log n} + \Delta\bar{W}_i\ ,$$

where $\Delta\bar{W}_i = \left| \sum_{v_j \in \mathcal{R}} \bar{W}_{ij} - \sum_{v_j \in \mathcal{B}} \bar{W}_{ij} \right|$ . A *network* is *nicely colored* if no vertex is $\mathcal{W}$-ambiguous.

**Proof of Theorem 3.11 and Remark 3.12** Pseudocode and details are presented in Appendices A.3, A.5 and A.5.5. Theorem 3.11 can be directly concluded from Theorem 4.8 by setting $|\mathcal{A}| = |\mathcal{A}^{\mathcal{W}}| = 0$.

## 5. Experiments

In this section, we empirically validate the effectiveness of our proposed methods for Problem 2 through a series of meticulously designed experiments. We test our algorithms EgalAlg and EgalAlg$^{(\text{appx})}$ (pseudocodes in Appendix A.3) benchmarked against random selections together with three heuristic methods: selecting nodes based on degree (Degree), innate error rate (ErrRate), and the product of degree and error rate (DegXErr).

**Synthetic Datasets** Synthetic data is generated with three components: a random graph $G$, a weight matrix $\bar{W}$ and initial opinions $\hat{\mathbf{y}}$. To generate $G$, we employ Erdős-Rényi model (ER) (Erdös & Rényi, 1959), Barabási-Albert model (PA) (Barabási & Albert, 1999) and Watts-Strogatz model (WS) (Watts & Strogatz, 1998). We generate the weight matrix $\bar{W}$ using the FJ model. Each $\hat{y}_i(a)$ is sampled from Bernoulli distribution with a randomly chosen $p_i$.

**Real-World Graphs** We also evaluate our methods on four diverse real network datasets (Rossi & Ahmed, 2015), BIO (Duch & Arenas, 2005; Bader et al., 2012), CSPK (Bader et al., 2013), FB (Rozemberczki et al., 2019), WIKI (Leskovec et al., 2010). Here we also apply finite step FJ model to construct weight matrix $\bar{W}$ and randomly sample $\hat{\mathbf{y}}$ from Bernoulli distributions.

We define our accuracy Acc to be the achieved egalitarian gain, normalized by its upper bound. For each dataset, we progressively increase $k$. Obviously, as $k$ increases, the accuracy should increase. Therefore, we fix the threshold value $k = \log(n)$ for different datasets and compare the corresponding Acc of different methods. We also report the number of modified nodes $k$ required to achieve certain levels of accuracy. See Table 1 for details. Figure 2 shows two of these results on real and synthetic graphs and more are presented in Figure 3. More details about experiments can be found in Appendix B. Code of our experiments is available through link [3].

[3] https://github.com/jackal092927/social_learning_public

Table 1: Comparison of experiments on five methods Egal=EgalAlg, Appx=EgalAlg$^{(\text{appx})}$, Rand=Random selection.

| | | Datasets | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Score** | **Method** | ER (128) | PA (128) | WS (128) | RandW (128) | BIO (297) | CSPK (39) | FB (620) | WIKI (890) |
| Acc@ k=log(n) | Rand | 0.11 | 0.88 | 0.53 | 0.18 | 0.80 | 0.63 | 0.35 | 0.48 |
| | Degree | 0.08 | 0.96 | 0.42 | 0.12 | 0.78 | 0.84 | 0.36 | 0.49 |
| | ErrRate | 0.22 | **1.00** | 0.76 | 0.47 | 0.96 | 0.94 | 0.53 | 0.54 |
| | DegXErr | 0.18 | **1.00** | 0.89 | 0.37 | 0.96 | **1.00** | 0.72 | 0.78 |
| | Appx | 0.18 | **1.00** | 0.87 | 0.41 | 0.94 | 0.84 | 0.62 | 0.64 |
| | Egal | **0.27** | **1.00** | **1.00** | **0.58** | **1.00** | **1.00** | **0.88** | **0.96** |
| #k @ Acc>90% | Rand | >100 | 7 | 10 | 34 | 8 | 10 | 94 | 22 |
| | Degree | >100 | 4 | 17 | 45 | 9 | 4 | 93 | 26 |
| | ErrRate | 71 | 2 | 7 | 18 | 3 | 3 | 19 | 13 |
| | DegXErr | 71 | 2 | 6 | 18 | 3 | 3 | 32 | 7 |
| | Appx | 61 | **1** | 5 | 18 | 3 | 5 | 30 | 15 |
| | Egal | **55** | **1** | **3** | **12** | **1** | **1** | **9** | **2** |
| #k @ Acc>75% | Rand | 83 | 8 | 16 | 61 | 15 | 14 | 37 | 55 |
| | Degree | 83 | 5 | 28 | 64 | 14 | 6 | 20 | 54 |
| | ErrRate | 46 | 3 | 10 | 31 | 5 | 4 | 14 | 26 |
| | DegXErr | 51 | 3 | 8 | 36 | 4 | 3 | 8 | 16 |
| | Appx | 47 | **2** | 9 | 35 | 6 | 7 | 11 | 39 |
| | Egal | **36** | **2** | **4** | **19** | **2** | **2** | **4** | **3** |

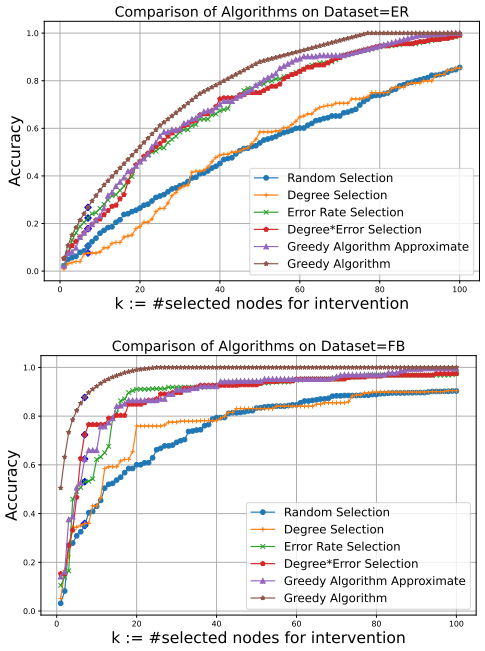



Figure 2: Algorithms performance on ER (top) and WIKI (bottom).

From the results we can conclude that, in summary, all six algorithms can be categorized into four tiers: Tier1={EgalAlg}, Tier2={EgalAlg$^{(\text{appx})}$}, Tier3={DegXErr, ErrRate}, Tier4={Degree, Rand}. We rank the efficiency of these methods as: Tier1$\gg$Tier2$\geq$Tier3$\gg$Tier4. Our greedy algorithm EgalAlg in general performs best on all datasets. In some datasets the EgalAlg$^{(\text{appx})}$ algorithm outperforms algorithms in Tier3 & 4 when $k$ is very small, however, the performances of Tier2 & 3 algorithms quickly become similar as $k$ increases. Tier4 algorithms are always the slowest in improving our egalitarian objective function. On almost all datasets, our greedy algorithm can achieve more than 80% accuracy within only $\log n$ nodes selected to intervene, and it beats all the other baselines. Our greedy approximation algorithm can achieve more than 70% accuracy. On all datasets, it beats our two baselines in Tier4 and on some data sets it beats all the baselines of Tier3 & 4.

## Conclusion

Given a network in which agents cooperatively perform a classification task, we analyse the problem of optimally choosing $k$ vertices and improving their innate predictions to maximize the overall network improvement.

**Limitations and Future Directions** In this paper, our modeling relies on a few simplifications which may pose limitations in the applicability of the methods, and they may be addressed in future works:

1. We assume that the social planner is capable of improving every agent's innate prediction equally through Equation (3). In reality, this improvement may depend on the *agent*, $i$, as well as the *data point*, $a$.

2. Our analyses are valid when the social influence graph has non-negative weights and, in the current form, they do not generalize to graphs with negative weights, e.g., signed graphs.

We believe that overcoming any of the above limitations would be an interesting extension of our work, and we propose them as future directions.

## Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

## Acknowledgements

Xin and Gao would like to acknowledge NSF support through CCF-2118953, CCF-2208663, DMS-2311064, DMS-2220271, and IIS-2229876.

## References

Abebe, R., Kleinberg, J., Parkes, D., and Tsourakakis, C. E. Opinion dynamics with varying susceptibility to persuasion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1089–1098, 2018.

Acemoglu, D., Dahleh, M. A., Lobel, I., and Ozdaglar, A. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.

Adriaens, F., Wang, H., and Gionis, A. Minimizing hitting time between disparate groups with shortcut edges. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, pp. 1–10, 2023.

Arieli, I., Sandomirskiy, F., and Smorodinsky, R. On social networks that support learning. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, EC '21, pp. 95–96, 2021.

Bader, D. A., Meyerhenke, H., Sanders, P., and Wagner, D. *Graph Partitioning and Graph Clustering*. American Mathematical Society, 2012.

Bader, D. A., Meyerhenke, H., Sanders, P., and Wagner, D. (eds.). *Graph Partitioning and Graph Clustering, 10th DIMACS Implementation Challenge Workshop, Georgia Institute of Technology, Atlanta, GA, USA, February 13-14, 2012. Proceedings*, volume 588 of *Contemporary Mathematics*, 2013. American Mathematical Society. ISBN 978-0-8218-9869-7. URL http://dblp.uni-trier.de/db/conf/dimacs/dimacs2012.html.

Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

Bindel, D., Kleinberg, J., and Oren, S. How bad is forming your own opinion? *Games and Economic Behavior*, 92: 248–265, 2015.

Blot, M., Picard, D., Cord, M., and Thome, N. Gossip training for deep learning. In *9th NIPS Workshop on Optimization for Machine Learning*, November 2016.

Borgs, C., Brautbar, M., Chayes, J., and Lucier, B. Maximizing social influence in nearly optimal time. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pp. 946–957. SIAM, 2014.

Chebotarev, P. and Shamis, E. On proximity measures for graph vertices. arXiv math/0602073, 2006.

Cinus, F., Gionis, A., and Bonchi, F. Rebalancing social feed to minimize polarization and disagreement. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, pp. 369–378, 2023.

Colin, I., Bellet, A., Salmon, J., and Clémençon, S. Gossip dual averaging for decentralized optimization of pairwise functions. In Balcan, M. F. and Weinberger, K. Q. (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1388–1396. PMLR, 2016.

D'Angelo, G., Severini, L., and Velaj, Y. Recommending links through influence maximization. *Theoretical Computer Science*, 764:30–41, 2019.

DeGroot, M. H. Reaching a consensus. *J. Am. Stat. Assoc.*, 69(345):118–121, 1974.

DeMarzo, P. M., Vayanos, D., and Zwiebel, J. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly journal of economics*, 118(3):909–968, 2003.

Duch, J. and Arenas, A. Community identification using extremal optimization. *Physical Review E*, 72:027104, 2005.

Eckles, D., Esfandiari, H., Mossel, E., and Rahimian, M. A. Seeding with costly network information. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, pp. 421–422, 2019.

Erdös, P. and Rényi, A. On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290, 1959.

Fellus, J., Picard, D., and Gosselin, P.-H. Asynchronous gossip principal components analysis. *Neurocomputing*, 169:262–271, December 2015.

Friedkin, N. E. and Johnsen, E. C. Social influence and opinions. *J. Math. Sociol.*, 15(3-4):193–206, January 1990.

Gaitonde, J., Kleinberg, J., and Tardos, É. Polarization in geometric opinion dynamics. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 499–519, July 2021.

Garimella, K., Gionis, A., Parotsidis, N., and Tatti, N. Balancing information exposure in social networks. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, volume 30, pp. 4666–4674, 2017.

Giaretta, L. and Girdzijauskas, Š. Gossip learning: Off the beaten path. In *2019 IEEE International Conference on Big Data (Big Data)*, pp. 1117–1124, December 2019.

Golub, B. and Jackson, M. O. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149, 2010.

Golub, B. and Sadler, E. Learning in social networks. *The Oxford Handbook of the Economics of Networks*, 2016.

Haddadan, S., Menghini, C., Riondato, M., and Upfal, E. RePBubLik: Reducing polarized bubble radius with link insertions. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, WSDM '21, pp. 139–147, 2021.

Haddadan, S., Menghini, C., Riondato, M., and Upfal, E. Reducing polarization and increasing diverse navigability in graphs by inserting edges and swapping edge weights. *Data Mining and Knowledge Discovery*, 36(6): 2334–2378, 2022.

Hązła, J., Jin, Y., Mossel, E., and Ramnarayan, G. A geometric model of opinion polarization. *Mathematics of Operations Research*, 49(1):251–277, 2023.

He, L., Bian, A., and Jaggi, M. COLA: Decentralized linear learning. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 4541–4551, 2018.

Hegedűs, I., Berta, Á., Kocsis, L., Benczúr, A. A., and Jelasity, M. Robust decentralized low-rank matrix decomposition. *ACM Trans. Intell. Syst. Technol.*, 7(4):1–24, May 2016.

Hegedűs, I., Danner, G., and Jelasity, M. Gossip learning as a decentralized alternative to federated learning. In *Distributed Applications and Interoperable Systems*, pp. 74–90. Springer International Publishing, 2019.

Hegedűs, I., Danner, G., and Jelasity, M. Decentralized learning works: An empirical comparison of gossip learning and federated learning. *J. Parallel Distrib. Comput.*, 148:109–124, February 2021.

Hązła, J., Jadbabaie, A., Mossel, E., and Rahimian, M. A. Reasoning in Bayesian opinion exchange networks is PSPACE-hard. In Beygelzimer, A. and Hsu, D. (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 1614–1648. PMLR, 25–28 Jun 2019.

Kempe, D., Kleinberg, J., and Tardos, E. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pp. 137–146, 2003.

Konečný, J., McMahan, B., and Ramage, D. Federated optimization: Distributed optimization beyond the datacenter. arXiv 1511.03575, November 2015.

Konečný, J., Brendan McMahan, H., Ramage, D., and Richtárik, P. Federated optimization: Distributed machine learning for on-device intelligence. arXiv 1610.02527, October 2016a.

Konečný, J., Brendan McMahan, H., Yu, F. X., Richtárik, P., Suresh, A. T., and Bacon, D. Federated learning: Strategies for improving communication efficiency. arXiv 1610.05492, October 2016b.

Krishnan, S., Jose Anand, A., Srinivasan, R., Kavitha, R., and Suresh, S. *Handbook on Federated Learning: Advances, Applications and Opportunities*. CRC Press, January 2024.

Lazarsfeld, J. and Alistarh, D. The power of populations in decentralized learning dynamics. arXiv 2306.08670, June 2023.

Leskovec, J., Huttenlocher, D., and Kleinberg, J. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1361–1370. ACM, 2010.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. y. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Singh, A. and Zhu, J. (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282. PMLR, 2017.

Musco, C., Musco, C., and Tsourakakis, C. E. Minimizing polarization and disagreement in social networks. In *Proceedings of the 2018 World Wide Web Conference*, pp. 369–378, 2018.

Nemhauser, G. L. and Wolsey, L. A. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research*, 3(3):177–188, 1978.

Ormándi, R., Hegedűs, I., and Jelasity, M. Gossip learning with linear models on fully distributed data. *Concurr. Comput.*, 25(4):556–571, February 2013.

Rahimian, M. A. and Jadbabaie, A. Bayesian learning without recall. *IEEE Transactions on Signal and Information Processing over Networks*, 3(3):592–606, 2017.

Rossi, R. A. and Ahmed, N. K. The network data repository with interactive graph analytics and visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp. 4292–4293, 2015.

Rozemberczki, B., Davies, R., Sarkar, R., and Sutton, C. GEMSEC: Graph embedding with self clustering. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 65–72. ACM, 2019.

Santos, F. P., Lelkes, Y., and Levin, S. A. Link recommendation algorithms and dynamics of polarization in online social networks. *Proceedings of the National Academy of Sciences*, 118(50):e2102141118, 2021.

Seeman, L. and Singer, Y. Adaptive seeding in social networks. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 459–468, 2013.

Tsioutsiouliklis, S., Pitoura, E., Semertzidis, K., and Tsaparas, P. Link recommendations for PageRank fairness. In *Proceedings of the ACM Web Conference 2022*, pp. 3541–3551, 2022.

Watts, D. J. and Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

Zhang, R., Guo, S., Wang, J., Xie, X., and Tao, D. A survey on gradient inversion: Attacks, defenses and future directions. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pp. 5678–5685, July 2023.

Zhu, L., Bao, Q., and Zhang, Z. Minimizing polarization and disagreement in social networks via link recommendation. *Advances in Neural Information Processing Systems*, 34: 2072–2084, 2021.

# A. Additional material: proofs and details of algorithms

Let $\Omega^+ = \{a \in \Omega \mid y(a) = +1\}$ and $\Omega^- = \{a \in \Omega \mid y(a) = -1\}$. The following are some useful lemmas that we will use throughout:

**Lemma A.1.** *Let $v_j \in V$ be an arbitrary agent. We have:*

$$\mathop{\mathbb{E}}_{a\sim\Omega^+}[\hat{y}_i(a)] = 1 - 2\mathsf{err}(v_i), \quad \mathop{\mathbb{E}}_{a\sim\Omega^-}[\hat{y}_i(a)] = 2\mathsf{err}(v_i) - 1\,.$$

*Proof.*

$$\begin{aligned}\mathop{\mathbb{E}}_{a\sim\Omega^+}[\hat{y}_i(a)] &= +1\cdot\mathbb{P}\left(\hat{y}_i(a) = +1\right) - 1\cdot\mathbb{P}\left(\hat{y}_i(a) = -1\right)\\ &= +1\cdot\mathbb{P}(\hat{y}_i(a) = y(a)) - 1\cdot\mathbb{P}(\hat{y}_i(a) \neq y(a))\\ &= (1 - \mathsf{err}(v_i)) - \mathsf{err}(v_i)\\ &= 1 - 2\mathsf{err}(v_i)\,.\end{aligned}$$

Similarly we have that

$$\mathop{\mathbb{E}}_{a\sim\Omega^-}[\hat{y}_i(a)] = +1\cdot\mathbb{P}\left(\hat{y}_i(a) \neq y(a)\right) - 1\cdot\mathbb{P}\left(\hat{y}_i(a) = y(a)\right) = \mathsf{err}(v_i) - (1 - \mathsf{err}(v_i)) = 2\mathsf{err}(v_i) - 1\,.$$

□

**Proposition A.2.** *Given any non-negative $\bar{W}$ used in equation (2) and $S$ on which we improve $\hat{y}_j$ to $\tilde{y}_j$, we have*

$$\forall v_i \in V,\ \mathcal{Z}_{\text{new}}(i,a) > \mathcal{Z}(i,a) \iff \exists v_j \in S, y(a)\hat{y}_j(a) = -1 \wedge \bar{W}_{ij} > 0\,.$$

*Proof.* By looking at Equation (4), it is evident that if $\hat{y}_j(a) = y(a)$ then $\tilde{y}_j(a) = \hat{y}_j(a)$. Thus, improving $v_j$ will only improve prediction on $a \in \Omega$ iff $\hat{y}_j(a) \neq y(a)$ or equivalently $\hat{y}_j(a)y(a) = -1$. Note that for any other $v_i \in V$, $\tilde{y}_j(a)$ appears in $\mathcal{Z}_{\text{new}}(i,a)$ with coefficient $\bar{W}_{ij}$. Since the matrix is non-negative we either have $\mathcal{Z}_{\text{new}}(i,a) = \mathcal{Z}(i,a)$ or $\mathcal{Z}_{\text{new}}(i,a) > \mathcal{Z}(i,a)$. Therefore, we will have $\mathcal{Z}_{\text{new}}(i,a) > \mathcal{Z}(i,a)$ iff $\hat{y}_j(a)y(a) = -1$ and $\bar{W}_{ij} \neq 0$. □

### A.1. Missing proofs from Section 4.1: Analysis of aggregate optimization

***Proof of Lemma 4.1.*** Let $\Omega^+ = \{a \in \Omega \mid y(a) = +1\}$ and $\Omega^- = \{a \in \Omega \mid y(a) = -1\}$. We have that:

$$\begin{aligned}\mathcal{G}^{(\text{agg})}(S) &= \mathop{\mathbb{E}}_{a\sim\Omega}\left[\sum_{i=1}^{n}\mathcal{Z}_{\text{new}}(i,a) - \mathcal{Z}(i,a)\right]\\ &= \mathop{\mathbb{E}}_{a\sim\Omega^+}\left[\sum_{i=1}^{n} z^*_{\text{new}}(i,a) - z^*(i,a)\right]\mathbb{P}(a\in\Omega^+) + \mathop{\mathbb{E}}_{a\sim\Omega^-}\left[\sum_{i=1}^{n} z^*(i,a) - z^*_{\text{new}}(i,a)\right]\mathbb{P}(a\in\Omega^-)\\ &= \sum_{i=1}^{n}\left(\mathop{\mathbb{E}}_{a\sim\Omega^+}[z^*_{\text{new}}(i,a) - z^*(i,a)]\,\mathbb{P}(a\in\Omega^+) + \mathop{\mathbb{E}}_{a\sim\Omega^-}[z^*(i,a) - z^*_{\text{new}}(i,a)]\,\mathbb{P}(a\in\Omega^-)\right)\,. \qquad (8)\end{aligned}$$

Note that:

$$z^*(i,a) - z^*_{\text{new}}(i,a) = \sum_{j=1}^{n}\bar{W}_{ij}(\hat{y}_j(a) - \tilde{y}_j(a)) = \sum_{j\in S}\bar{W}_{ij}(\hat{y}_j(a) - \tilde{y}_j(a)) = \varphi\sum_{j\in S}\bar{W}_{ij}[\hat{y}_j(a) - y(a)]$$

Thus,

$$z^*(i,a) - z^*_{\text{new}}(i,a) = \begin{cases}\varphi\sum_{j\in S}\bar{W}_{ij}[\hat{y}_j(a) + 1] & \text{if } a\in\Omega^-\\ \varphi\sum_{j\in S}\bar{W}_{ij}[\hat{y}_j(a) - 1] & \text{if } a\in\Omega^+\end{cases}$$

Using linearity of expectation and plugging in Lemma A.1 we have:

$$\mathbb{E}\left[z^*(i,a) - z^*_{\text{new}}(i,a)\right] = \begin{cases} \varphi \sum_{j\in S} \bar{W}_{ij}[2\mathsf{err}^{(0)}(v_i) - 1 + 1] & \text{if } a \in \Omega^- \\ \varphi \sum_{j\in S} \bar{W}_{ij}[1 - 2\mathsf{err}^{(0)}(v_i) - 1] & \text{if } a \in \Omega^+ \end{cases}$$

Therefore,

$$\mathop{\mathbb{E}}_{a\sim\Omega^-}\left[z^*(i,a) - z^*_{\text{new}}(i,a)\right] = 2\varphi\, \mathsf{err}^{(0)}(v_i) \sum_{j\in S} \bar{W}_{ij} = \mathop{\mathbb{E}}_{a\sim\Omega^+}\left[z^*_{\text{new}}(i,a) - z^*(i,a)\right] .$$

Plugging in the above in Equation (8) we obtain:

$$\mathcal{G}^{(\text{agg})}(S) = 2\varphi \sum_{j\in S}\sum_{i=1}^{n} \bar{W}_{ij}\mathsf{err}^{(0)}(v_i)[\mathbb{P}(a\in\Omega^+) + \mathbb{P}(a\in\Omega^-)] = 2\varphi \sum_{j\in S}\sum_{i=1}^{n} \bar{W}_{ij}\mathsf{err}^{(0)}(v_i) .$$

This means by picking $k$ vertices with highest values of $\text{Inf}(j) = \sum_{i=1}^{n} \bar{W}_{ij}\mathsf{err}^{(0)}(v_i)$ we will obtain the optimal solution for Problem 1. In order to find these values, we first need to find all the values of $\text{Inf}(j)$ for all $j \in V$. Which takes $\Theta(n^2)$ number of steps. Then we have to find top $k$ elements among these values, which will take $\Theta(kn)$. □

***Proof of Remark 3.6.*** Let $\widehat{\text{Inf}}(j) = \sum_{i=1}^{n} \widehat{\bar{W}}_{i,j}\mathsf{err}^{(0)}(v_j)$. For all $j$, we have that

$$\left|\text{Inf}(j) - \widehat{\text{Inf}}(j)\right| = \sum_{i=1}^{n} \left|\bar{W}_{ij} - \widehat{\bar{W}}_{ij}\right| \mathsf{err}^{(0)}(v_j) \le \sum_{i=1}^{n} \left|\bar{W}_{ij} - \widehat{\bar{W}}_{ij}\right|$$

Note that the right-hand side is the $\ell_1$ norm of the $j$th column of $\bar{W} - \widehat{\bar{W}}$, let's denote the $j$th column of these matrix respectively by $\bar{W}_{\cdot j}$ and $\widehat{\bar{W}}_{\cdot j}$. Since the $\ell_1$ norm of a matrix is defined the be the maximum over $\ell_1$ norm of all of its columns we have that

$$\forall j \;\; \left|\text{Inf}(j) - \widehat{\text{Inf}}(j)\right| \le \left|\bar{W}_{\cdot j} - \widehat{\bar{W}}_{\cdot j}\right|_1 \le \epsilon .$$

In the previous theorem we showed that

$$\mathcal{G}^{(\text{agg})}(S) = 2\varphi \sum_{j\in S} \text{Inf}(j) .$$

Since size of $S$ is $k$ and $\varphi \le 1$ the total error is bounded by $2k\epsilon$ . □

### A.2. Missing proof from Section 4.2: Analysis of egalitarian optimization

#### A.2.1. HARDNESS RESULTS

**Theorem A.3.** *There is a polynomial time reduction from the set cover problem to Problem 1.*

*Proof.* Consider an arbitrary $S \subseteq V$. We use Proposition A.2 to simplify $\mathcal{G}^{(\text{egal})}(S)$. For any $v_i \in V$, we define: $\Omega_{v_j} \triangleq \{a \in \Omega \mid y(a)\hat{y}_j(a) = -1\}$. For any $a \in \Omega$, we denote its probability by $\mu_a$. For a given $S$ we have:

$$\begin{aligned}
\mathcal{G}^{(\text{egal})}(S) &= \mathop{\mathbb{E}}_{a\sim\Omega}\left[\sum_{i=1}^{n} \mathbf{1}(\mathcal{Z}(i,a) < 0 \wedge \mathcal{Z}(i,a) < \mathcal{Z}_{\text{new}}(i,a))\right] \\
&= \sum_{a\in\Omega}\sum_{i=1}^{n} \mu_a \mathbf{1}(\mathcal{Z}(i,a) < 0) \cdot \mathbf{1}\left(\bigvee_{v_j\in S} \left(a \in \Omega_{v_j} \wedge \bar{W}_{ij} > 0\right)\right) \\
&= \sum_{(a,v_i)\in\Omega\times V} \mu_a \mathbf{1}(\mathcal{Z}(i,a) < 0) \cdot \mathbf{1}\left(\bigvee_{v_j\in S} \left(a \in \Omega_{v_j} \wedge \bar{W}_{ij} > 0\right)\right) \qquad (9)
\end{aligned}$$

Using the above simplification, we now construct the following instance of the weighted set cover problem:

Consider a bipartite graph where one part is $U = \{(v_i, a) \in V \times \Omega \mid \mathcal{Z}(i,a) < 0\}$ and the other part is $S = V$. There is an edge between any $v_j \in V$ to a pair $(v_i, a) \in U$ iff $a \in \Omega_{v_j} \wedge \bar{W}_{ij} > 0$ and the weight of each pair $(a, v_i)$ is $\mu_a$. Under the new definition Equation (9) is equivalent to

$$\mathcal{G}^{(\text{egal})}(S) = \sum_{(a,v_i)\in U} \mu_a \cdot \mathbf{1}\left(\bigvee_{v_j \in S} \left(a \in \Omega_{v_j} \wedge \bar{W}_{ij} > 0\right)\right)$$

Note that

$$\mathbf{1}\left(\bigvee_{v_j \in S} \left(a \in \Omega_{v_j} \wedge \bar{W}_{ij} > 0\right)\right) = 1 \iff \text{there is an edge between } v_j \text{ and } (a, v_i) \text{ and } v_j \in S\ .$$

The reduction is polynomial in sizes of $\Omega$ and $V$. Since we assume that $\Omega$ has polynomial size, it is a polynomial time reduction. This completes the proof. □

***Proof of Theorem 3.7.*** The proof follows from Theorem A.3 and the fact that set cover is NP-hard. □

**Theorem A.4.** *Consider Problem 2 and assume $k = 1$. There exist two networks with the same number of agents, same $\bar{W}$ and same error rates $\{\mathsf{err}(v_j)\}_{j=1}^n$. In these network, only the joint probability distributions $\pi_1$ and $\pi_2$ are different. There are subsets $V_1, V_2 \subseteq V$ such that $V_1 \cap V_2 = \emptyset$. In the first network we have that for any $u \in V_1, \mathcal{G}^{(\text{egal})}(u) = \Theta(n)$ and for any $u \notin V_1$ we have $\mathcal{G}^{(\text{egal})}(u) = \Theta(1)$. In the second network for any $u \in V_2$ will have $\mathcal{G}^{(\text{egal})}(u) = \Theta(n)$ and any $u \notin V_2$ will satisfy $\mathcal{G}^{(\text{egal})}(u) = \Theta(1)$.*

*Proof.* Let $V = \{u_1, u_2, u_3, u_4\} \cup \{v_1, v_2, v_3, \ldots, v_{2n}\}$. We define $\bar{W}$ to be the following matrix:

$\bar{W}_{u_1,v_j} = \bar{W}_{u_2,v_j} = 1$ for all $j = 1 : n$ , and $\bar{W}_{u_3,v_j} = \bar{W}_{u_4,v_j} = 1$ for all $j = n+1 : 2n$.

For each vertex in $V$ we also have a self loop of weight 1, i.e., $\bar{W}_{u_i u_i} = \bar{W}_{v_j v_j} = 1$ for all $i, j$.

The error rates of these agents are as follows: $\mathsf{err}(v_j) = 0$ for all $j = 1 : 2n$ and $\mathsf{err}(u_i) = 1/2$ for all $j = 1 : 4$.

In the first network the error of $\hat{y}_{u_1}(a)$ is negatively correlated with $\hat{y}_{u_2}(a)$ and $\hat{y}_{u_3}$ is positively correlated with $\hat{y}_{u_4}$ as:

$$\mathbb{P}\left(\hat{y}_{u_1}(a) \neq \hat{y}_{u_2}(a)\right) = 1 \quad \& \ \mathbb{P}\left(\hat{y}_{u_3}(a) = \hat{y}_{u_4}(a)\right) = 1$$

Both $\hat{y}_{u_1}$ and $\hat{y}_{u_2}$ are independent from $\hat{y}_{u_3}$ and $\hat{y}_{u_4}$.

Let $V_1 = \{u_3, u_4\}$. We now show that $\mathcal{G}^{(\text{egal})}(u_3) = \mathcal{G}^{(\text{egal})}(u_4) = n$ and for any $u \notin V_1$, $\mathcal{G}^{(\text{egal})}(u) \in \{0, 1\}$.

From Lemma 4.2 we conclude that for any vertex $u$ we have:

$$\mathcal{G}^{(\text{egal})}(u) = \sum_{i;\bar{W}_{iu}\neq 0} \mathbb{P}_{a\sim\Omega}\left(\mathcal{Z}(i,a) \leq 0 \wedge y(a) \neq \hat{y}_u(a)\right)$$

Thus, it is immediate that for each $v_j$, we have $\mathcal{G}^{(\text{egal})}(v_j) = 0$.

Consider $u_1$

$$\mathcal{G}^{(\text{egal})}(u_1) = \sum_{i=1:n} \mathbb{P}_{a\sim\Omega}\left(\mathcal{Z}(i,a) \leq 0 \wedge y(a) \neq \hat{y}_{u_1}(a)\right) + \mathbb{P}_{a\sim\Omega}\left(\mathcal{Z}(u_1,a) \leq 0 \wedge y(a) \neq \hat{y}_{u_1}(a)\right)$$

Since $u_1$ is connected to $v_1, \dots v_n$, the last summand is 0 if $n > 1$, and it is $1/2$ if $n = 1$. In any case it is a constant. We now look at the first summand.

$$\begin{aligned} first\ summand &= \sum_{i=1:n} \mathop{\mathbb{P}}_{a\sim\Omega}((\hat{y}_{u_1}(a) + \hat{y}_{u_2}(a) + \hat{y}_i(a)) \cdot y(a) \leq 0 \wedge y(a) \neq \hat{y}_u(a)) \\ &= \sum_{i=1:n} \mathop{\mathbb{P}}_{a\sim\Omega}\left(y(a)^2 \leq 0 \wedge y(a) \neq \hat{y}_u(a)\right) = 0 \end{aligned}$$

The last equation follows from the fact that always $\hat{y}_{u_1}(a) \neq \hat{y}_{u_2}(a)$, and $\hat{y}_i(a) = y(a)$.

Similarly we can show that $\mathcal{G}^{(\text{egal})}(u_2) = \{0, 1/2\}$.

For $u_3$, let $c$ be a constant which is $c \in \{0, 1/2\}$. We have:

$$\begin{aligned} \mathcal{G}^{(\text{egal})}(u_3) &= \sum_{i=n+1:2n} \mathop{\mathbb{P}}_{a\sim\Omega}(\mathcal{Z}(i,a) \leq 0 \wedge y(a) \neq \hat{y}_{u_3}(a)) + \mathop{\mathbb{P}}_{a\sim\Omega}(\mathcal{Z}(u_3,a) \leq 0 \wedge y(a) \neq \hat{y}_{u_3}(a)) \\ &= \sum_{i=1:n} \mathop{\mathbb{P}}_{a\sim\Omega}((\hat{y}_{u_3}(a) + \hat{y}_{u_4}(a) + \hat{y}_i(a)) \cdot y(a) \leq 0 \wedge y(a) \neq \hat{y}_u(a)) + c \\ &= \sum_{i=1:n} \mathop{\mathbb{P}}_{a\sim\Omega}((y(a) - 2y(a)) \cdot y(a) \leq 0)\, \mathsf{err}(u_3) + c \\ &= n/2 + c\,. \end{aligned}$$

Similarly we have $\mathcal{G}^{(\text{egal})}(u_4) \in \{n/2, (n+1)/2\}$.

Therefore, both $u_3$ and $u_4$ can be the optimal choice for this network. And any other choice will have an error of magnitude $\Theta(n)$.

In the second network, we make the following change:

$$\mathbb{P}\left(\hat{y}_{u_1}(a) = \hat{y}_{u_2}(a)\right) = 1 \quad \& \ \mathbb{P}\left(\hat{y}_{u_3}(a) \neq \hat{y}_{u_4}(a)\right) = 1$$

We still let both $\hat{y}_{u_1}$ and $\hat{y}_{u_2}$ are independent from $\hat{y}_{u_3}$ and $\hat{y}_{u_4}$.

Using a similar analysis we can show that

$$\mathcal{G}^{(\text{egal})}(u_1) = \mathcal{G}^{(\text{egal})}(u_2) = n/2 + 1 \ \& \ \forall u \in V,\ u \neq u_1, u_2 \implies \mathcal{G}^{(\text{egal})}(u) \in \{0, 1/2\}\,.$$

□

### A.2.2. MONOTONICITY AND SUBMODULARITY OF $\mathcal{G}^{(\text{egal})}$

**Lemma A.5.** *Assume $S' \subseteq S \subseteq V$, we have: $\mathcal{G}^{(\text{egal})}(S') \leq \mathcal{G}^{(\text{egal})}(S)$.*

*Proof.* From Proposition A.2 that for any arbitrary $S \subseteq V$ we have:

$$\mathcal{G}^{(\text{egal})}(S) = \sum_{i=1}^{n} \mathbb{P}\left(y(a)z^*(i,a) \leq 0 \wedge \bigvee_{\substack{v_j \in S \\ \bar{W}_{ij} \neq 0}} (y(a) \neq \hat{y}_j(a))\right)$$

For $S' \subseteq S$, we split the event $\bigvee_{\substack{v_j \in S \\ \bar{W}_{ij} \neq 0}} (y(a) \neq \hat{y}_j(a))$ to the two following non-intersecting events:

$$\bigvee_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} (y(a)\neq \hat{y}_j(a)) = \underbrace{\left(\bigvee_{\substack{v_j\in S'\\ \bar{W}_{ij}\neq 0}} (y(a)\neq \hat{y}_j(a))\right)}_{E_{i1}} \vee \underbrace{\left(\bigvee_{\substack{v_j\in S\setminus S'\\ \bar{W}_{ij}\neq 0}} (y(a)\neq \hat{y}_j(a)) \wedge \bigwedge_{\substack{v_j\in S'\\ \bar{W}_{ij}\neq 0}} (y(a)= \hat{y}_j(a))\right)}_{E_{i2}}$$

Since $E_1$ and $E_2$ are non-intersecting we have:

$$\mathcal{G}^{(\text{egal})}(S) = \sum_{i=1}^{n} \mathbb{P}(y(a)z^*(i,a) < 0 \wedge E_{i1}) + \sum_{i=1}^{n} \mathbb{P}(y(a)z^*(i,a) < 0 \wedge E_{i2})\,.$$

Note that $\mathcal{G}^{(\text{egal})}(S') = \sum_{i=1}^{n} \mathbb{P}(y(a)z^*(i,a) < 0 \wedge E_{i1})$. Therefore, we conclude the premise. □

**Lemma A.6.** *Consider arbitrary $S \subseteq V$ and $u, v \in V \setminus S$. We have:*

$$\mathcal{G}^{(\text{egal})}(S\cup\{u,v\}) + \mathcal{G}^{(\text{egal})}(S) \leq \mathcal{G}^{(\text{egal})}(S\cup\{u\}) + \mathcal{G}^{(\text{egal})}(S\cup\{v\})\,.$$

*Proof.* Like previous lemma, we split the events of RHS and LHS to non-intersecting smaller events.

Consider the following events:

$$E_{FFF}(i) \equiv \left(\bigvee_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} \hat{y}_i(a)\neq y(a)\right) \wedge \left(\hat{y}_u(a)\neq \hat{y}(a)\wedge \bar{W}_{iu}\neq 0\right) \wedge \left(\hat{y}_v(a)\neq y(a)\wedge \bar{W}_{iv}\neq 0\right)$$

$$E_{TFF}(i) \equiv \left(\neg\bigvee_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} \hat{y}_i(a)\neq y(a)\right) \wedge (\hat{y}_u(a)\neq \hat{y}(a)\wedge \bar{W}_{iu}\neq 0) \wedge \left(\hat{y}_v(a)\neq y(a)\wedge \bar{W}_{iv}\neq 0\right)$$

$$E_{TTF}(i) \equiv \left(\neg\bigvee_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} \hat{y}_i(a)\neq y(a)\right) \wedge \neg\left(\hat{y}_u(a)\neq \hat{y}(a)\wedge \bar{W}_{iu}\neq 0\right) \wedge \left(\hat{y}_v(a)\neq y(a)\wedge \bar{W}_{iv}\neq 0\right)$$

$$E_{TFT}(i) \equiv \left(\neg\bigvee_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} \hat{y}_i(a)\neq y(a)\right) \wedge \left(\hat{y}_u(a)\neq \hat{y}(a)\wedge \bar{W}_{iu}\neq 0\right) \wedge \neg\left(\hat{y}_v(a)\neq y(a)\wedge \bar{W}_{iv}\neq 0\right)$$

and similar definitions for $E_{TTT}(i)$, $E_{FTT}(i)$, $E_{FTF}(i)$ and $E_{FFT}(i)$.

$$\mathcal{G}^{(\text{egal})}(S\cup\{u,v\}) = \sum_{i=1}^{n} \sum_{(X,Y,Z)\in\{T,F\}^3\setminus\{(F,F,F)\}} \mathbb{P}\left(y(a)z^*(i,a)\leq 0\wedge E_{X,Y,Z}(i)\right),$$

$$\mathcal{G}^{(\text{egal})}(S) = \sum_{i=1}^{n} \sum_{(Y,Z)\in\{T,F\}^2} \mathbb{P}\left(y(a)z^*(i,a)\leq 0\wedge E_{T,Y,Z}(i)\right),$$

$$\mathcal{G}^{(\text{egal})}(S\cup\{u\}) = \sum_{i=1}^{n} \sum_{(X,Y,Z)\in\{T,F\}^3\setminus\{(F,F,F),(F,F,T)\}} \mathbb{P}\left(y(a)z^*(i,a)\leq 0\wedge E_{X,Y,Z}(i)\right)$$

and finally

$$\mathcal{G}^{(\text{egal})}(S \cup \{v\}) = \sum_{i=1}^{n} \sum_{(X,Y,Z)\in\{T,F\}^3\setminus\{(F,F,F),(F,T,F)\}} \mathbb{P}\left(y(a)z^*(i,a) \leq 0 \wedge E_{X,Y,Z}(i)\right)$$

By counting the number of appearances of each term on the RHS and LHS we may conclude the premise. □

### A.3. Pseudocode of the greedy algorithms

In this section we present our algorithms for egalitarian improvement.

**Overview of algorithms** The first algorithm EgalAlg has access to $\pi$ and finds the greedy choice accurately.

The second algorithm $\mathsf{EgalAlg}^{(\text{appx})}$ receives parameters $mode$, $\vec{\mathbf{err}}$ and $\bar{W}$ as input parameters. If we assume pairwise independence of classifiers, $mode = \text{ind}$ and $\vec{\mathbf{err}}$ contains agents' error rates. If we assume group dependency $mode = \text{gr}$ and $\vec{\mathbf{err}}$ contains the agent's individual error rates as well as $\mathsf{err}(\mathcal{R})$ and $\mathsf{err}(\mathcal{B})$. Depending on the mode of the algorithm, $\mathsf{EgalAlg}^{(\text{appx})}$calls subsequent procedures $\mathsf{EstGain}^{\text{ind}}$ and $\mathsf{EstGain}^{\text{gr}}$ for estimating the greedy choice.

The pseudocodes are as follows and analysis is presented in subsequent subsections:

**Algorithm 1** $\mathsf{EgalAlg}\left(\pi, \bar{W}\right)$

$S = \emptyset$
**for** $i = 1 : k$ **do**
  $maxval = 0$
  **for** $j = 1 : n$ **do**
    $\Delta\mathcal{G}(S, v_j) = 0$
    **for** $a \in \Omega$ **do**
      **for** $\ell = 1 : n$ **do**
        $E = \mathcal{Z}(i,a) \leq 0 \wedge \left(\bigwedge_{\substack{v_j\in S \\ \bar{W}_{ji}\neq 0}} y(a){=}\hat{y}_j(a)\right) \wedge y(a){\neq}\hat{y}_u(a)$
        **if** $E \equiv T$ and $\bar{W}_{\ell j} \neq 0$ **then**
          $\Delta\mathcal{G}(S, v_j) = \Delta\mathcal{G}(S, v_j) + \pi(E)$
        **end if**
      **end for**
    **end for**
    **if** $\Delta\mathcal{G}(S, v_j) \geq maxval$,
      $g = v_j$
  **end for**
  $S = S \cup \{g\}$
**end for**

**Algorithm 2** $\mathsf{EgalAlg}^{(\mathrm{appx})}\big(mode, \vec{\mathbf{err}}, \bar{W}\big)$

$S = \emptyset$
**for** $i = 1 : k$ **do**
  $maxval = 0$
  **for** $j = 1 : n$ **do**
    $\widehat{\Delta\mathcal{G}}^{\mathrm{indv}}(S, v_j) = \mathsf{EstGain}^{\mathrm{ind}}(S, v_j, \{\mathsf{err}(v_i)\}_{i=1}^n, \bar{W})$ (Algorithm 3)
    $\widehat{\Delta\mathcal{G}}(S, v_j) = \widehat{\Delta\mathcal{G}}^{\mathrm{indv}}(S, v_j)$
    **if** mode = gr **then**
      $\widehat{\Delta\mathcal{G}}^{\mathrm{gr}}(S, v_j) = \mathsf{EstGain}^{\mathrm{gr}}(S, v_j, \{\mathsf{err}(v_i)\}_{i=1}^n, \mathsf{err}(\mathcal{R}), \mathsf{err}(\mathcal{B}), \bar{W})$ (Algorithm 4)
      $\widehat{\Delta\mathcal{G}}(S, v_j) = \rho\widehat{\Delta\mathcal{G}}^{\mathrm{gr}}(S, v_j) + (1-\rho)\widehat{\Delta\mathcal{G}}^{\mathrm{indv}}(S, v_j)$
    **end if**
    **if** $\widehat{\Delta\mathcal{G}}(S, v_j) \geq maxval$,
      $g = v_j$
  **end for**
  $S = S \cup \{g\}$
**end for**

**Algorithm 3** $\mathsf{EstGain}^{\mathrm{ind}}(S, u, \{\mathsf{err}(v_j)\}_{j=1}^n, \bar{W})$

$\widehat{\Delta\mathcal{G}}(S, u) = 0$
**for** $\ell = 1 : n$ **do**
  $\mathsf{tru}_\ell(S) = 1$
  **for** $v_m \in S$ **do**
    **if** $\bar{W}_{\ell m} \neq 0$,
      $\mathsf{tru}_\ell(S) = \mathsf{tru}_\ell(S) \times (1 - \mathsf{err}(v_m))$
  **end for**
  **if** $\Psi_\ell(S, u) < 0$ and $\bar{W}_{\ell u} \neq 0$
  $\widehat{\Delta\mathcal{G}}(S, u)$ += $\mathsf{err}(u) \cdot \mathsf{tru}_\ell(S)$
**end for**

**Algorithm 4** $\mathsf{EstGain}^{\mathrm{gr}}(S, u, \{\mathsf{err}(v_j)\}_{j=1}^{n}, \mathsf{err}(\mathcal{R}), \mathsf{err}(\mathcal{B}), \bar{W})$

$\widehat{\Delta\mathcal{G}}(S,u) = 0$
**boolean** $Case1 = T$
**boolean** $Case2R = F$
**boolean** $Case2B = F$
**boolean** $Case3 = F$
**for** $\ell = 1 : n$ **do**
  $\mathsf{tru}_\ell(S) = 1$
  **for** $v_m \in S$ **do**
    **if** $\bar{W}_{\ell m} \neq 0$ **then**
      **if** $v_m \in \mathcal{W}$ **then** $\mathsf{tru}_\ell(S) = \mathsf{tru}_\ell(S) \times (1 - \mathsf{err}(v_m))$
      **if** $(v_m \in \mathcal{R}) \wedge Case1 = T$ **then**
        $Case1 = F$
        $Case2R = T$
      **end if**
      **if** $(v_m \in \mathcal{B}) \wedge Case1 = T$ **then**
        $Case1 = F$
        $Case2B = T$
      **end if**
      **if** $(v_m \in \mathcal{R} \wedge Case2B = T)$ or $(v_m \in \mathcal{B} \wedge Case2R = T)$ **then**
        $Case2B = Case2R = F$
        $Case3 = T$
      **end if**
    **end if**
  **end for**
  **if** $Case1 \wedge u \in \mathcal{W}$ **then**
    $\Delta\mathcal{G}(S,u) += \mathsf{tru}_\ell(S)\mathsf{err}(u)[\mathbf{1}(\Psi_i(\mathcal{B} \cup S), \mathcal{R} \cup \{u\})\mathsf{err}(\mathcal{R}) + \mathbf{1}(\mathcal{R} \cup S), \mathcal{B} \cup \{u\})\mathsf{err}(\mathcal{B})]$
  **end if**
  **if** $Case1 \wedge u \in \mathcal{R}$ **then**
    $\Delta\mathcal{G}(S,u) += \mathsf{tru}_\ell(S)\mathsf{err}(u)\mathbf{1}\Psi_i(\mathcal{R} \cup S, \mathcal{B} \cup \{u\})$
  **end if**
  **if** $Case1 \wedge u \in \mathcal{B}$ **then**
    $\Delta\mathcal{G}(S,u) += \mathsf{tru}_\ell(S)\mathsf{err}(u)\mathbf{1}\Psi_i(\mathcal{B} \cup S, \mathcal{R} \cup \{u\})$
  **end if**
  **if** $Case2B \wedge u \in \mathcal{W}$ **then**
    $\Delta\mathcal{G}(S,u) += \mathsf{err}(u)\mathsf{err}(\mathcal{R})\mathsf{tru}(S)\mathbf{1}\Psi(\mathcal{B} \cup S, \mathcal{R} \cup \{u\})$
  **end if**
  **if** $Case2R \wedge u \in \mathcal{W}$ **then**
    $\Delta\mathcal{G}(S,u) += \mathsf{err}(u)\mathsf{err}(\mathcal{B})\mathsf{tru}(S)\mathbf{1}\Psi(\mathcal{R} \cup S, \mathcal{B} \cup \{u\})$
  **end if**
  **if** $Case2B \wedge u \in \mathcal{R}$ **then**
    $\Delta\mathcal{G}(S,u) += \mathsf{err}(\mathcal{R})\mathsf{tru}(S)\mathbf{1}\Psi(\mathcal{B} \cup S, \mathcal{R})$
  **end if**
  **if** $Case2R \wedge u \in \mathcal{B}$ **then**
    $\Delta\mathcal{G}(S,u) += \mathsf{err}(\mathcal{B})\mathsf{tru}(S)\mathbf{1}\Psi(\mathcal{R} \cup S, \mathcal{B})$
  **end if**
**end for**
**return** $\Delta\mathcal{G}(S,u)$

### A.3.1. FINDING THE GREEDY CHOICE

Remember from the main text that

$$\Delta\mathcal{G}_i(u,S) = \mathop{\mathbb{P}}_{a\sim\Omega}\left(\mathcal{Z}(i,a)\leq 0 \wedge \left(\bigwedge_{\substack{v_j\in S\\ \bar{W}_{ji}\neq 0}} y(a)=\hat{y}_j(a)\right) \wedge y(a)\neq\hat{y}_u(a)\right) .$$

***Proof of Lemma 4.2***. It is immediate from Proposition A.2 that for any arbitrary $S\subseteq V$ we have:

$$\mathcal{G}^{(\text{egal})}(S) = \sum_{i=1}^{n}\mathbb{P}\left(y(a)z^*(i,a)\leq 0 \wedge \bigvee_{v_j\in S}(\bar{W}_{ij}\neq 0 \wedge y(a)\neq\hat{y}_j(a))\right)$$

Writing the above for $S\cup\{u\}$ and simplifying we obtain:

$$\begin{aligned}
\mathcal{G}^{(\text{egal})}(S\cup\{u\}) =& \sum_{i=1}^{n}\mathbb{P}\left(y(a)z^*(i,a)\leq 0 \wedge \bigvee_{v_j\in S\cup\{u\}}(\bar{W}_{ij}\neq 0 \wedge y(a)\neq\hat{y}_j(a))\right)\\
=& \sum_{i=1}^{n}\mathbb{P}\left(y(a)z^*(i,a)\leq 0 \wedge \bigvee_{v_j\in S}(\bar{W}_{ij}\neq 0 \wedge y(a)\neq\hat{y}_j(a))\right)\\
&+\mathbb{P}\left(y(a)z^*(i,a)\leq 0 \wedge \bigwedge_{v_j\in S}(\bar{W}_{ij}=0 \vee y(a)=\hat{y}_j(a)) \wedge (\bar{W}_{iu}\neq 0 \wedge y(a)\neq\hat{y}_u(a))\right)\\
=& \mathcal{G}^{(\text{egal})}(S) + \sum_{\substack{i=1:n\\ \bar{W}_{iu}\neq 0}}\mathbb{P}\left(y(a)z^*(i,a)\leq 0 \wedge \left(\bigwedge_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} y(a)=\hat{y}_j(a)\right) \wedge y(a)\neq\hat{y}_u(a)\right)\\
=& \mathcal{G}^{(\text{egal})}(S) + \sum_{\substack{i=1:n\\ \bar{W}_{iu}\neq 0}}\Delta\mathcal{G}_i(S,u)
\end{aligned}$$

□

The following lemma is a middle step for approximation of the greedy choice:

**Lemma A.7.** *Let* $\Delta\mathcal{G}_i(S,u)$ *be*

$$\Delta\mathcal{G}_i(S,u) = \mathop{\mathbb{P}}_{a\sim\Omega}\left(\mathcal{Z}(i,a)\leq 0 \wedge \left(\bigwedge_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} y(a)=\hat{y}_j(a)\right) \wedge y(a)\neq\hat{y}_u(a)\right)$$

*We have that*

$$\Delta\mathcal{G}_i(S,u) = \mathbb{P}\left(T_i(S)\wedge F(u)\right)\Gamma_i(S,u) ,$$

*where*

$$\Gamma_i(S,u) = \mathbb{P}\left(a \in \Omega^+\right) \mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) < -\sum_{v_j\in S}\bar{W}_{ji} + \bar{W}_{iu}\right)$$
$$+ \mathbb{P}\left(a \in \Omega^-\right) \mathop{\mathbb{P}}_{a\in\Omega^-}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) > \sum_{v_j\in S}\bar{W}_{ji} - \bar{W}_{iu}\right) .$$

*and $T_i(S)$ and $F(u)$ are the following events:*

$$T_i(S) \doteq \bigwedge_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} \hat{y}_j(a) = y(a), \qquad F(u) \doteq \hat{y}_u(a) \neq y(a)$$

*Proof.* Let $E$ denote the event

$$E \triangleq \left(\bigwedge_{\substack{v_j\in S\\ \bar{W}_{ji}\neq 0}} y(a) = \hat{y}_j(a)\right) \wedge y(a) \neq \hat{y}_u(a) .$$

We may write

$$\mathbb{P}\left(y(a)z^*(i,a) \leq 0 \wedge E\right) = \mathbb{P}\left(y(a)z^*(i,a) \leq 0 \mid E\right)\mathbb{P}(E)$$

In order find $\mathbb{P}\left(y(a)z^*(i,a) \leq 0 \mid E\right)$, we split the probability based on the true label of $a$:

If $a \in \Omega^+$ we have:

$$\mathbb{P}\left(y(a)z^*(i,a) \leq 0 \mid E\right) = \mathbb{P}\left(z^*(i,a) \leq 0 \mid \left(\bigwedge_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} \hat{y}_j(a) = +1\right) \wedge \hat{y}_u(a) = -1\right)$$

Note that

$$z^*(i,a) = \sum_{j=1}^{n} \bar{W}_{ij}\hat{y}_j(a)$$

thus,

$$\mathbb{P}\left(y(a)z^*(i,a) \leq 0 \mid E\right) = \mathbb{P}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) + \sum_{v_j\in S}\bar{W}_{ij} - \bar{W}_{iu} \leq 0\right)$$

Similarly, if $a \in \Omega^-$ we have:

$$\mathbb{P}\left(y(a)z^*(i,a) \leq 0 \mid E\right) = \mathbb{P}\left(z^*(i,a) \geq 0 \mid \left(\bigwedge_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} \hat{y}_j(a) = -1\right) \wedge \hat{y}_u(a) = +1\right)$$

Since

$$z^*(i,a) = \sum_{j=1}^{n} \bar{W}_{ij}\hat{y}_j(a)$$

we have:

$$\mathbb{P}\left(y(a)z^*(i,a) \geq 0 \mid E\right) = \mathbb{P}\left(\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a) - \sum_{v_j \in S} \bar{W}_{ij} + \bar{W}_{iu} \geq 0\right)$$

Rearranging and putting together, we obtain the premise.

□

### A.4. Missing material from Section 4.2.1: estimating greedy choice assuming independence

***Proof of Lemma 4.6.*** Let

$$\mathsf{tru}_i(S) = \mathbb{P}\left(\bigwedge_{\substack{v_j \in S \\ \bar{W}_{ij} \neq 0}} \hat{y}_j(a) = y(a)\right) .$$

Using independence and from Lemma A.7 we can write

$$\Delta\mathcal{G}_i(S,u) = \mathsf{err}(u)\mathsf{tru}_i(S)\Gamma_i(S,u) \tag{10}$$

In Lemma A.8 which follows this proof, we show that by taking

$$\widehat{\Gamma}_i(S,u) = \begin{cases} 0 & \text{if } \Psi_i(S,u) \geq 0 \\ 1 & \text{otherwise} \end{cases}$$

we have

$$\left|\widehat{\Gamma}_i(S,u) - \Gamma_i(S,u)\right| \leq \exp\left(-\frac{\Psi_i(S,u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right) \leq \exp\left(-\frac{\Psi_i(u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right),$$

Plugging in this approximation in Equation (10) we obtain:

$$\widehat{\Delta\mathcal{G}_i}(S,u) = \begin{cases} \mathsf{err}(u)\mathsf{tru}_i(S) & \text{if } \Psi_i(S,u) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Note that since the classifiers are independent we have

$$\mathsf{tru}_i(S) = \mathbb{P}\left(\bigwedge_{\substack{v_j \in S \\ \bar{W}_{ij} \neq 0}} \hat{y}_j(a) = y(a)\right) = \prod_{\substack{v_j \in S \\ \bar{W}_{ij} \neq 0}} \mathbb{P}\left(\hat{y}_j(a) = y(a)\right) = \prod_{\substack{v_j \in S \\ \bar{W}_{ij} \neq 0}} \left(1 - \mathsf{err}(v_j)\right) .$$

We have $\mathsf{tru}_i(S), \mathsf{err}(u) \leq 1$, thus the error of approximating $\Delta\mathcal{G}_i(S,u)$ using Equation (11) is at most $\exp\left(-\frac{\Psi_i(u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right)$.

□

**Lemma A.8.** *Assume that all the all the classifiers $\hat{y}_1(a), \hat{y}_2(a), \ldots, \hat{y}_n(a)$ are pairwise independent. We can estimate $\Gamma_i(S,u)$ as follows:*

$$\widehat{\Gamma}_i(S,u) = \begin{cases} 0 & \textit{if } \Psi_i(S,u) > 0 \\ 1 & \textit{otherwise} \end{cases}$$

*where,*

$$\Psi_i(S,u) = \sum_{v_j \in S} \bar{W}_{ij} + \sum_{\substack{j=1:n \\ j \notin S \cup \{u\}}} \bar{W}_{ij}[1 - 2\mathsf{err}(j)] - \bar{W}_{iu}$$

*The error of this estimation is bounded by:*

$$\left|\widehat{\Gamma}_i(S,u) - \Gamma_i(S,u)\right| \leq \exp\left(-\frac{\Psi_i(S,u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right) \leq \exp\left(-\frac{\Psi_i(u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right),$$

*where*

$$\Psi_i(u) = \sum_{j=1}^n \bar{W}_{ij}[1 - 2\mathsf{err}(j)] - 2\mathsf{err}(u)\bar{W}_{iu}\ .$$

*Proof.* Let's remember the definition of $\Gamma_i(S,u)$ from Lemma A.7:

$$\begin{aligned}\Gamma_i(S,u) = &\mathbb{P}\left(a \in \Omega^+\right) \mathop{\mathbb{P}}_{a \in \Omega^+}\left(\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a) < -\sum_{v_j \in S} \bar{W}_{ij} + \bar{W}_{iu}\right) \\ &+ \mathbb{P}\left(a \in \Omega^-\right) \mathop{\mathbb{P}}_{a \in \Omega^-}\left(\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a) > \sum_{v_j \in S} \bar{W}_{ij} - \bar{W}_{iu}\right)\ . \end{aligned} \tag{12}$$

Assume first that $\Psi_i(S,u) > 0$. We use the Hoeffding bound Theorem A.20 to estimate

$$\mathop{\mathbb{P}}_{a \in \Omega^+}\left(\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a) < -\sum_{v_j \in S} \bar{W}_{ij} + \bar{W}_{iu}\right) \quad \& \mathop{\mathbb{P}}_{a \in \Omega^-}\left(\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a) > \sum_{v_j \in S} \bar{W}_{ij} - \bar{W}_{iu}\right) \tag{13}$$

In order to bound the first probability in Equation (13), note that $\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a) < -\sum_{v_j \in S} \bar{W}_{ij} + \bar{W}_{iu}$ iff :

$$\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a) < \mathbb{E}\left[\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a)\right] - \mathbb{E}\left[\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a)\right] - \sum_{v_j \in S} \bar{W}_{ij} + \bar{W}_{iu}$$

furthermore, from Lemma A.1 we have $a \in \Omega^+$ implies:

$$\mathbb{E}\left[\sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}\hat{y}_j(a)\right] = \sum_{\substack{j=1:n \\ v_j \notin S \cup \{u\}}} \bar{W}_{ij}[1 - 2\mathsf{err}(j)]\ .$$

Thus,

$$\begin{aligned}
&\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j(a) < -\sum_{v_j\in S}\bar{W}_{ij}+\bar{W}_{ui}\right)\\
&=\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j(a) < \mathbb{E}\left[\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j(a)\right] - \sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}[1-2\mathsf{err}(j)] - \sum_{v_j\in S}\bar{W}_{ij}+\bar{W}_{iu}\right)\\
&=\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j(a) < \mathbb{E}\left[\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j(a)\right] - \Psi_i(S,u)\right)
\end{aligned}$$

Since $\hat{y}_i(a)$s are pairwise independent, and $\Psi_i(S,u)>0$ we may use the Hoeffding bound to obtain:

$$\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j(a) < -\sum_{v_j\in S}\bar{W}_{ij}+\bar{W}_{iu}\right) \le \exp\left(-\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}^2}\right).$$

The second probability in Equation (13) may be bounded similarly as follows:

$$\begin{aligned}
&\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}\hat{y}_j(a) > \sum_{v_j\in S}\bar{W}_{ji}-\bar{W}_{ui}\right)\\
&=\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}\hat{y}_j(a) > \mathbb{E}\left[\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}\hat{y}_j(a)\right] - \sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}[2\mathsf{err}(j)-1] + \sum_{v_j\in S}\bar{W}_{ji}-\bar{W}_{ui}\right)\\
&=\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}\hat{y}_j(a) > \mathbb{E}\left[\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}\hat{y}_j(a)\right] + \Psi_i(S,u)\right)
\end{aligned}$$

Again, under pairwise independence and $\Psi_i(S,u)>0$ the above probability is bounded as:

$$\mathop{\mathbb{P}}_{a\in\Omega^-}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}\hat{y}_j(a) > \sum_{v_j\in S}\bar{W}_{ji}-\bar{W}_{ui}\right) \le \exp\left(-\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2}\right).$$

Putting together, we obtain: If $\Psi_i(S,u) > 0$:

$$\begin{aligned}\Gamma_i(S,u) =& \mathbb{P}\left(a \in \Omega^+\right) \mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ji}\hat{y}_j(a) < -\sum_{v_j\in S}\bar{W}_{ji} + \bar{W}_{ui}\right)\\ &+ \mathbb{P}\left(a \in \Omega^-\right) \mathop{\mathbb{P}}_{a\in\Omega^-}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ji}\hat{y}_j(a) > \sum_{v_j\in S}\bar{W}_{ji} - \bar{W}_{ui}\right)\\ \leq& \exp\left(\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2}\right)[\mathbb{P}(a\in\Omega^+) + \mathbb{P}(a\in\Omega^-)] = \exp\left(\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2}\right)\end{aligned}$$

Note that $\Gamma_i(S,u) \geq 0$. Therefore, if $\Psi_i(S,u) > 0$, we define $\widehat{\Gamma}_i(S,u) = 0$ and we will have:

$$0 \leq \Gamma_i(S,u) - \widehat{\Gamma}_i(S,u) \leq \exp\left(-\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2}\right) .$$

Let's now find a lower bound on $\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2}$ which is independent of $S$. We have:

$$\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2 \leq \sum_{i=1}^{n}\bar{W}_{ji}^2 .$$

and

$$\begin{aligned}\Psi_i(S,u) &= \sum_{v_j\in S}\bar{W}_{ij} + \sum_{\substack{j=1:n\\ j\notin S\cup\{u\}}}\bar{W}_{ij}[1-2\mathsf{err}(j)] - \bar{W}_{iu}\\ &\geq \sum_{\substack{j=1:n\\ j\neq u}}\bar{W}_{ij}[1-2\mathsf{err}(j)] - \bar{W}_{iu}\\ &= \sum_{j=1}^{n}\bar{W}_{ij}[1-2\mathsf{err}(j)] - 2\mathsf{err}(u)\bar{W}_{iu} \;= \Psi_i(u) .\end{aligned}$$

Thus,

$$0 \leq \Gamma_i(S,u) - \widehat{\Gamma}_i(S,u) \leq \exp\left(-\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2}\right) \leq \exp\left(-\frac{\Psi_i(u)^2}{\sum_{j=1}^{n}\bar{W}_{ji}^2}\right) .$$

Assume now that $\Psi_i(S,u) < 0$. In this case we write the first probability in Equation (13) as:

$$\begin{aligned}
&\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) < -\sum_{v_j\in S}\bar{W}_{ij} + \bar{W}_{iu}\right) = 1 - \mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) \geq -\sum_{v_j\in S}\bar{W}_{ij} + \bar{W}_{iu}\right)\\
=&1 - \mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) \geq \mathbb{E}\left[\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a)\right] - \sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}[1-2\mathsf{err}(j)] - \sum_{v_j\in S}\bar{W}_{ji} + \bar{W}_{iu}\right)\\
=&1 - \mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) \geq \mathbb{E}\left[\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a)\right] - \Psi_i(S,u)\right)
\end{aligned}$$

Since $\Psi_i(S,u) < 0$, we have $-\Psi_i(S,u) > 0$ using the pairwise independence of the classifiers, we employ the Hoeffding bound and obtain that:

$$\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) < -\sum_{v_j\in S}\bar{W}_{ij} + \bar{W}_{iu}\right) \geq 1 - \exp\left(-\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}^2}\right).$$

Similarly we have :

$$\mathop{\mathbb{P}}_{a\in\Omega^-}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) > \sum_{v_j\in S}\bar{W}_{ij} - \bar{W}_{ui}\right) \geq 1 - \exp\left(\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}^2}\right).$$

Putting together, we obtain: If $\Psi_i(S,u) < 0$:

$$\begin{aligned}
\Delta\mathcal{G}_i(S,u) =& \mathbb{P}\left(a\in\Omega^+\right)\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ji}\hat{y}_j(a) < -\sum_{v_j\in S}\bar{W}_{ji} + \bar{W}_{ui}\right)\\
&+ \mathbb{P}\left(a\in\Omega^-\right)\mathop{\mathbb{P}}_{a\in\Omega^-}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ji}\hat{y}_j(a) > \sum_{v_j\in S}\bar{W}_{ji} - \bar{W}_{ui}\right)\\
\geq&[1-\exp\left(-\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2}\right)][\mathbb{P}(a\in\Omega^+) + \mathbb{P}(a\in\Omega^-)] = 1-\exp\left(-\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2}\right)
\end{aligned}$$

Note that $\Gamma_i(S,u) \leq 1$. Therefore, if $\Psi_i(S,u) < 0$, we define $\widehat{\Gamma}_i(S,u) = 1$ and we will have:

$$0 \leq \widehat{\Gamma}_i(S,u) - \Gamma_i(S,u) \leq \exp\left(-\frac{\Psi_i(S,u)^2}{\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ji}^2}\right) \leq \exp\left(-\frac{\Psi_i(u)^2}{\sum_{j=1}^n\bar{W}_{ji}^2}\right).$$

This completes our proof. □

### A.4.1. MISSING MATERIAL FROM SECTION 4.2.1 RELATED TO AMBIGUOUS VERTICES

**Lemma A.9.** *Let $V^+$ be those agents with error less than $1/2$ and $V^-$ be those agents with error greater than $1/2$. i.e,*

$$V^+ = \{v_j \mid \mathsf{err}(v_j) \leq 1/2\} \quad \& \, V^- = \{v_j \mid \mathsf{err}(v_j) > 1/2\}$$

*and consider the following vectors*

$$\bar{W}_i^+ = (\bar{W}_{ij})_{v_j \in V^+} \quad \& \, \mathcal{E}^+ = (1 - 2\mathsf{err}(v_j))_{v_j \in V^+}$$

*and*

$$\bar{W}_i^- = (\bar{W}_{ij})_{v_j \in V^-} \quad \& \, \mathcal{E}^- = (2\mathsf{err}(v_j) - 1)_{v_j \in V^-}$$

*and*

$$\bar{W}_i = (\bar{W}_{i1}, \bar{W}_{i2}, \ldots, \bar{W}_{in})$$

*We have that:*

$$\begin{aligned}
\exp\left(-\frac{\Psi_i(u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right) &\leq \exp\left(-\frac{1}{4}\left(\frac{\langle \bar{W}_i^+, \mathcal{E}^+\rangle}{\left|\bar{W}_i\right|_2} - \frac{\langle \bar{W}_i^-, \mathcal{E}^-\rangle}{\left|\bar{W}_i\right|_2} - 2\right)^2\right) \\
&\leq \exp\left(-\frac{1}{4}\left(M \cdot \frac{\left|\bar{W}_i^+\right|_2}{\left|\bar{W}_i\right|_2} \cdot \left|\mathcal{E}^+\right|_2 - \frac{\left|\bar{W}_i^-\right|_2}{\left|\bar{W}_i\right|_2} \cdot \left|\mathcal{E}^-\right|_2 - 2\right)^2\right)
\end{aligned}$$

*where*

$$M = \frac{\max \bar{W}_i^+}{\min \bar{W}_i^+} \cdot \frac{1}{\min \mathcal{E}_i^+} \, .$$

*Proof.* We show the above equation by finding an lower bound for $\frac{\Psi_i(u)^2}{\sum_{i=1}^n \bar{W}_{ij}^2}$.

Consider an arbitrary vector $X$the $\ell_2$ norm is defined as:

$$|X|_2 = \sqrt{\sum_{x_j \in X} x_j^2}$$

Note that all of the above vectors only have positive elements.

We have:

$$\begin{aligned}
\frac{\Psi_i(u)^2}{\sum_{i=1}^n \bar{W}_{ij}^2} &= \left(\frac{\Psi_i(u)}{\left|\bar{W}_i(V)\right|_2}\right)^2 \\
&= \left(\frac{\sum_{v_i \in V^+} \bar{W}_{ij}(1 - 2\mathsf{err}(v_j))}{\left|\bar{W}_i(V)\right|_2} - \frac{\sum_{v_i \in V^-} \bar{W}_{ij}(2\mathsf{err}(v_j) - 1)}{\left|\bar{W}_i(V)\right|_2} - \frac{2\mathsf{err}(u)\bar{W}_{iu}}{\left|\bar{W}_i(V)\right|_2}\right)^2 \\
&= \left(\frac{\langle \bar{W}_i^+, \mathcal{E}^+\rangle}{\left|\bar{W}_i\right|_2} - \frac{\langle \bar{W}_i^-, \mathcal{E}^-\rangle}{\left|\bar{W}_i\right|_2} - \frac{2\mathsf{err}(u)\bar{W}_{iu}}{\left|\bar{W}_i\right|_2}\right)^2 \qquad (14)
\end{aligned}$$

where in the last equation $\langle\rangle$ denotes dot product.

Using Pólya-Szegö's inequality we have:

$$\frac{\langle \bar{W}_i^+, \mathcal{E}^+\rangle}{\left|\bar{W}_i\right|_2} \geq \frac{\left|\bar{W}_i^+\right|_2}{\left|\bar{W}_i\right|_2} \cdot \frac{\max \bar{W}_i^+}{\min \bar{W}_i^+} \cdot \frac{|\mathcal{E}^+|_2}{\min \mathcal{E}^+}$$

Using Cauchy Schwarz we have

$$\frac{\langle \bar{W}_i^-, \mathcal{E}^- \rangle}{\left|\bar{W}_i\right|_2} \leq \frac{\left|\bar{W}_i\right|_2}{\left|\bar{W}_i^-\right|_2} \cdot \left|\mathcal{E}^-\right|_2$$

Therefore, letting $M = \frac{\max \bar{W}_i^+}{\min \bar{W}_i^+} \cdot \frac{1}{\min \mathcal{E}_i^+}$

we have:

$$\frac{\Psi_i(u)^2}{\sum_{i=1}^n \bar{W}_{ij}^2} \geq \left( M \cdot \frac{\left|\bar{W}_i^+\right|_2}{\left|\bar{W}_i\right|_2} \cdot \left|\mathcal{E}^+\right|_2 - \frac{\left|\bar{W}_i^-\right|_2}{\left|\bar{W}_i\right|_2} \cdot \left|\mathcal{E}^-\right|_2 - \frac{2\mathsf{err}(u)\bar{W}_{iu}}{\left|\bar{W}_i\right|_2} \right)^2$$

The premise may be concluded from the fact that $\frac{\mathsf{err}(u)\bar{W}_{iu}}{\left|\bar{W}_i\right|_2} \leq 1$.

□

***Proof of Lemma 4.5.*** From Lemma 4.6 we now that

$$\left|\widehat{\Delta\mathcal{G}_i}(S,u) - \Delta\mathcal{G}_i(S,u)\right| \leq \exp\left(-\frac{\Psi_i(u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right) ,$$

and from Lemma A.9 we have:

$$\exp\left(-\frac{\Psi_i(u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right) \leq \exp\left(-\frac{1}{4}\left(M \cdot \frac{\left|\bar{W}_i^+\right|_2}{\left|\bar{W}_i\right|_2} \cdot \left|\mathcal{E}^+\right|_2 - \frac{\left|\bar{W}_i^-\right|_2}{\left|\bar{W}_i\right|_2} \cdot \left|\mathcal{E}^-\right|_2 - 2\right)^2\right)$$

Putting together and assume that $v_i$ is not ambiguous. We have that:

$$\begin{aligned}\left|\widehat{\Delta\mathcal{G}_i}(S,u) - \Delta\mathcal{G}_i(S,u)\right| &\leq \exp\left(-\frac{\Psi_i(u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right) \leq \exp\left(-\frac{1}{4}\left(M \cdot \frac{\left|\bar{W}_i^+\right|_2}{\left|\bar{W}_i\right|_2} \cdot \left|\mathcal{E}^+\right|_2 - \frac{\left|\bar{W}_i^-\right|_2}{\left|\bar{W}_i\right|_2} \cdot \left|\mathcal{E}^-\right|_2 - 2\right)^2\right) \\ &\leq \exp\left(-\frac{1}{4}\left(3\sqrt{\log n} - 2\right)^2\right) \leq \exp\left(-\frac{1}{4}(5\log n)\right) = o(n^{-1}) .\end{aligned}$$

□

### A.4.2. MISSING MATERIAL FROM SECTION 4.2.1: PROOF OF THE MAIN THEOREMS

In this subsection we present a pseudocode of our algorithm under the assumption that the agents are pairwise independent.

The following theorem bounds the error of this algorithm:

**Theorem A.10.** *Assume that $S$ is the output of Algorithm 2. We have that:*

$$\mathcal{G}^{(\mathrm{egal})}(S) \geq [(1-1/e) - \Delta^{\mathrm{ind}}] \cdot \mathrm{OPT}^{(\mathrm{egal})} ,$$

*where*

$$\Delta^{\mathrm{ind}} = \sum_{i=1}^n \exp\left(-\frac{\Psi_i(u)^2}{\sum_{i=1}^n \bar{W}_{ij}^2}\right) .$$

*Proof.* The proof immediately follows from the fact that we have a submodular and monotone function and that the error greedy choice taken as

$$g = \operatorname*{argmax}_u \sum_{i=1}^n \widehat{\Delta\mathcal{G}_i}(S,u) ,$$

and that for any $v_i$ we have

$$\left|\widehat{\Delta\mathcal{G}_i}(S,u) - \Delta\mathcal{G}_i(S,u)\right| \leq \exp\left(-\frac{\Psi_i(u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right) ,$$

□

***Proof of Theorem 4.3 and Theorem 3.10*** . Using the above theorem Theorem 4.3 and Theorem 3.10 can be directly concluded just by noticing that for any non-ambiguous vertex the error induced on the greedy choice is at most $o(n^{-1})$. Thus, in total all of non-ambiguous vertices together will have an error of $o(1)$. The ambiguous each can have an error as large as 1. □

**Assuming having access to approximation $\widehat{\bar{W}}$.** If we only have access to $\widehat{\bar{W}}$, we may run $\mathsf{EgalAlgo}^{\mathrm{ind}}$ using $\widehat{\bar{W}}$. In this case the approximation guarantee can be concluded from the following lemma which is similar to Lemma A.8:

**Lemma A.11.** *Assume that all the all the classifiers $\hat{y}_1(a), \hat{y}_2(a), \ldots, \hat{y}_n(a)$ are pairwise independent. We can estimate $\Gamma_i(S,u)$ (See Equation (12)) as follows:*

$$\widehat{\Gamma}_i(S,u) = \begin{cases} 0 & \text{if } \widehat{\Psi}_i(S,u) > 0 \\ 1 & \text{otherwise} \end{cases}$$

*where,*

$$\widehat{\Psi}_i(S,u) = \sum_{v_j \in S} \widehat{\bar{W}}_{ij} + \sum_{\substack{j=1:n \\ j \notin S \cup \{u\}}} \widehat{\bar{W}}_{ij}[1 - 2\mathsf{err}(j)] - \widehat{\bar{W}}_{iu}$$

*The error of this estimation is bounded by:*

$$\left|\widehat{\Gamma}_i(S,u) - \Gamma_i(S,u)\right| \leq \exp\left(-\frac{\Psi_i(u)^2}{4\sum_{i=1}^n \bar{W}_{ij}^2}\right)(1+\epsilon) ,$$

*where*

$$\Psi_i(u) = \sum_{j=1}^n \bar{W}_{ij}[1 - 2\mathsf{err}(j)] - 2\mathsf{err}(u)\bar{W}_{iu} .$$

*Proof.* Similar to the proof of Lemma A.8 we may bound the two terms of $\Gamma_i(S,u)$ as follows:

$$\begin{aligned} &\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n \\ v_j \notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) < -\sum_{v_j\in S}\bar{W}_{ij} + \bar{W}_{ui}\right) \\ \leq &\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n \\ v_j \notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) < \mathbb{E}\left[\sum_{\substack{j=1:n \\ v_j \notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a)\right] - \widehat{\Psi}_i(S,u) - 2\epsilon\right) \end{aligned}$$

Similarly to the previous case if $\widehat{\Psi}_i(S,u) > 0$ we may use Hoeffding bound to bound the above probability as:

$$\exp\left(-\frac{(\widehat{\Psi}_i(u) - 2\epsilon)^2}{\sum_{j=1}^n \bar{W}_{ij}^2}\right) \leq \exp\left(-\frac{(\Psi_i(u) - 4\epsilon)^2}{\sum_{j=1}^n \bar{W}_{ij}^2}\right)$$

Let's now bound RHS:

$$\exp\left(-\frac{(\Psi_i(u)-4\epsilon)^2}{\sum_{j=1}^n \bar{W}_{ij}^2}\right) \le \exp\left(-\frac{\Psi_i(u)^2}{\sum_{j=1}^n \bar{W}_{ij}^2} + 8\epsilon\frac{\Psi_i(u)}{\sum_{j=1}^n \bar{W}_{ij}^2}\right) \le \exp\left(-\frac{\Psi_i(u)^2}{\sum_{j=1}^n \bar{W}_{ij}^2} + 8\epsilon\right)$$
$$\le \exp\left(-\frac{\Psi_i(u)^2}{\sum_{j=1}^n \bar{W}_{ij}^2}\right)\exp(8\epsilon) \le (1+8\epsilon)\exp\left(-\frac{\Psi_i(u)^2}{\sum_{j=1}^n \bar{W}_{ij}^2}\right)$$

If $\widehat{\Psi}_i(S,u) < 0$ we may write:

$$\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) > -\sum_{v_j\in S}\bar{W}_{ij} + \bar{W}_{ui}\right)$$
$$\le \mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a) > \mathbb{E}\left[\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}} \bar{W}_{ij}\hat{y}_j(a)\right] - \widehat{\Psi}_i(S,u) + 2\epsilon\right)$$

and with a similar argument we obtain the same bound. Bounding the probability in the case where $a\in\Omega^-$ can be done similarly. And the rest of the proof holds similarly to proof of Lemma A.8. □

***Proof of Remark 3.12 independent case*** . The remark is a direct conclusion of the fact that the error of the greedy choice is bounded by

$$\sum_{i=1}^n \left|\widehat{\Gamma}_i(S,u) - \Gamma_i(S,u)\right| \le (1+\epsilon)\sum_{i=1}^n \exp\left(-\frac{\Psi_i(u)^2}{4\sum_{j=1}^n \bar{W}_{ij}^2}\right) .$$

□

### A.5. Missing proofs from Section 4.2.2: Estimating greedy choice assuming group dependence

Assume the assumption presented in Definition 4.7 and remember the following definition from previous sections

$$\Delta\mathcal{G}_i(S,u) = \mathop{\mathbb{P}}_{a\sim\Omega}\left(\mathcal{Z}(i,a)\le 0 \wedge \left(\bigwedge_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}} y(a)=\hat{y}_j(a)\right)\wedge y(a)\neq\hat{y}_u(a)\right),$$

and our goal is to estimate $\Delta\mathcal{G}_i(S,u)$ for all $S\subseteq V$, $u\in V$ and $i\in[n]$.

We may write

$$\Delta\mathcal{G}_i(S,u) = \Delta\mathcal{G}_i^{\text{gr}}(S,u)\rho + \Delta\mathcal{G}_i^{\text{indv}}(S,u)(1-\rho) ,$$

, where the super-scripts show whether individual or group decisions have been made. Estimation of $\Delta\mathcal{G}_i^{\text{indv}}(S,u)$ can be obtained using Lemma 4.6 and by calling $\mathsf{EstGain}^{\text{ind}}$ of Algorithm 2. Estimation of $\Delta\mathcal{G}_i^{\text{gr}}(S,u)$ can be done through a series of lemmas that we present here, and it depends on the whether vertices of $S_i$, defined as $S_i \triangleq \{v_j\in S \mid \bar{W}_{ij}\neq 0\}$, intersects $\mathcal{R}$, $\mathcal{B}$ or both. A pseudocode is presented at Algorithms 2 and 4. Our analysis is presented in Appendices A.5.1 to A.5.3. The following lemma summarizes these results:

**Lemma A.12.** *Assume the model presented in Definition 4.7. Having a set $S$ let $g$ be the vertex which maximizes the following function*

$$g = \mathop{\text{argmax}}_{u\in V}\ \rho\sum_{\substack{i=1:n\\ \mathcal{W}_{ij}\neq 0}} \widehat{\Delta\mathcal{G}}_i^{\text{gr}}(S,u) + (1-\rho)\sum_{\substack{i=1:n\\ \mathcal{W}_{ij}\neq 0}} \widehat{\Delta\mathcal{G}}_i^{\text{indv}}(S,u)$$

*Where $\widehat{\Delta\mathcal{G}}_i^{\mathrm{gr}}(S,u)$ may be obtained from Lemmas A.14, A.16 and A.17, and $\widehat{\Delta\mathcal{G}}_i^{\mathrm{indv}}(S,u)$ may be obtained from Lemma 4.6. We have that:*

$$\mathcal{G}^{(\mathrm{egal})}\left(S\cup\{\mathsf{gr}(S)\}\right)-\mathcal{G}^{(\mathrm{egal})}\left(S\cup\{g\}\right)\leq\rho\sum_{i=1}^{n}\exp\left(-\frac{(\Psi_i^{\mathcal{W}}(u)-\Delta\bar{W}_i)^2}{4\sum_{v_j\in\mathcal{W}}\bar{W}_{ij}^2}\right)+(1-\rho)\sum_{i=1}^{n}\exp\left(-\frac{(\Psi_i(u))^2}{4\sum_{v_j\in V}\bar{W}_{ij}^2}\right)$$

We will use the following definitions and results throughout the section.

$$\Delta\mathcal{G}_i^{\mathrm{gr}}(S,u)=\underset{a\sim\Omega}{\overset{\mathrm{gr}}{\mathbb{P}}}\left(\mathcal{Z}(i,a)\leq 0\wedge\mathcal{T}_i(S)\wedge\mathcal{F}(u)\right)\ ,$$

where $\mathcal{T}_i(S)$ and $\mathcal{F}(u)$ are the following events:

$$\mathcal{T}_i(S)=\bigwedge_{v_j\in S_i}y(a)=\hat{y}_j(a)\quad\&\ \mathcal{F}(u)=y(a)\neq\hat{y}_u(a)\ .$$

for any $v_i$ and $X\subseteq V$ we denote:

$$\bar{W}_i(X)=\sum_{v_j\in X}\bar{W}_{ij}$$

In addition, for any $X,Y\subseteq V$ we define,

$$\Psi_i(X,Y)=\sum_{v_j\in V\setminus(X\cup Y)}\left[1-2\mathsf{err}^{\mathrm{indv}}(v_j)\right]+\bar{W}_i(X)-\bar{W}_i(Y)\ .$$

We will use the following lemma throughout:

**Lemma A.13.** *Let's define*

$$\Gamma_i^{+}(X,Y)=\underset{a\in\Omega^{+}}{\mathbb{P}}\left(\sum_{v_j\in V\setminus(X\cup Y)}\bar{W}_{ij}\hat{y}_j^{\mathrm{indv}}(a)\leq\bar{W}_i(Y)-\bar{W}_i(X)\right)\ ,$$

*and*

$$\Gamma_i^{-}(X,Y)=\underset{a\in\Omega^{-}}{\mathbb{P}}\left(\sum_{v_j\in V\setminus(X\cup Y)}\bar{W}_{ij}\hat{y}_j^{\mathrm{indv}}(a)\geq\bar{W}_i(X)-\bar{W}_i(Y)\right)\ .$$

*We may estimate the above probabilities as*

$$\widehat{\Gamma}_i^{+}(X,Y)=\widehat{\Gamma}_i^{-}(X,Y)=\begin{cases}0 & \textit{if }\ \Psi_i(X,Y)\geq 0\\ 1 & \textit{if }\ \Psi_i(X,Y)<0\end{cases}$$

*We have that:*

$$\left|\Gamma_i^{+}(X,Y)-\widehat{\Gamma}_i^{+}(X,Y)\right|\leq\exp\left(-\frac{(\Psi_i(X,Y))^2}{\sum_{v_j\in V\setminus(X\cup Y)}\bar{W}_{ij}^2}\right)\ ,$$

*and*

$$\left|\Gamma_i^{-}(X,Y)-\widehat{\Gamma}_i^{-}(X,Y)\right|\leq\exp\left(-\frac{(\Psi_i(X,Y))^2}{\sum_{v_j\in V\setminus(X\cup Y)}\bar{W}_{ij}^2}\right)\ .$$

*Proof.* Assume $a \in \Omega^+$ in this case we have that

$$\mathop{\mathbb{E}}_{a\in\Omega^+}\left[\hat{y}_j^{\text{indv}}(a)\right] = 1 - 2\mathsf{err}^{indv}(v_j)$$

Thus,

$$\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \leq \bar{W}_i(Y) - \bar{W}_i(X) \iff$$

$$\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \leq \mathbb{E}\left[\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)\right] - \left(\mathbb{E}\left[\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)\right] - \bar{W}_i(Y) + \bar{W}_i(X)\right) \iff$$

$$\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \leq \mathbb{E}\left[\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)\right] - \Psi_i(X,Y)$$

If $\Psi_i(X,Y) > 0$ we may use the Hoeffding bound to obtain:

$$\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \leq \bar{W}_i(Y) - \bar{W}_i(X)\right) \leq \exp\left(-\frac{(\Psi_i(X,Y))^2}{\sum_{v_j\in V\setminus(X\cup Y)}\bar{W}_{ij}^2}\right).$$

If $-\Psi_i(X,Y) > 0$, we write:

$$\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \geq \bar{W}_i(Y) - \bar{W}_i(X) \iff$$

$$\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \geq \mathbb{E}\left[\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)\right] - \Psi_i(X,Y)$$

Thus,

$$\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \geq \bar{W}_i(Y) - \bar{W}_i(X)\right) \leq \exp\left(-\frac{(\Psi_i(X,Y))^2}{\sum_{v_j\in V\setminus(X\cup Y)}\bar{W}_{ij}^2}\right)$$

Therefore,

$$\begin{aligned}
&\mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \leq \bar{W}_i(Y) - \bar{W}_i(X)\right)\\
=&1 - \mathop{\mathbb{P}}_{a\in\Omega^+}\left(\sum_{v_j\in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \geq \bar{W}_i(Y) - \bar{W}_i(X)\right)\\
\geq&1 - \exp\left(-\frac{(\Psi_i(X,Y))^2}{\sum_{v_j\in V\setminus(X\cup Y)}\bar{W}_{ij}^2}\right).
\end{aligned}$$

Putting together we have:

$$\left|\Gamma_i^+(X,Y) - \widehat{\Gamma}_i^+(X,Y)\right| \leq \exp\left(-\frac{(\Psi_i(X,Y))^2}{\sum_{v_j\in V\setminus(X\cup Y)}\bar{W}_{ij}^2}\right),$$

Similarly if $a \in \Omega^-$ we have:

$$\mathop{\mathbb{E}}_{a\in\Omega^-}\left[\hat{y}_j^{\text{indv}}(a)\right] = 2\mathsf{err}^{\text{indv}}(v_j) - 1$$

thus,

$$\sum_{v_j \in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \geq \bar{W}_i(X) - \bar{W}_i(Y) \iff$$

$$\sum_{v_j \in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a) \geq \mathbb{E}\left[\sum_{v_j \in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)\right] + \left(-\mathbb{E}\left[\sum_{v_j \in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)\right] + \bar{W}_i(X) - \bar{W}_i(Y)\right)$$

and

$$-\mathbb{E}\left[\sum_{v_j \in V\setminus(X\cup Y)} \bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)\right] + \bar{W}_i(X) - \bar{W}_i(Y) = -\sum_{v_j \in V\setminus(X\cup Y)} \bar{W}_{ij}[2\mathsf{err}^{\text{indv}}(v_j) - 1] + \bar{W}_i(X) - \bar{W}_i(Y)$$
$$= \Psi_i(X, Y)\,.$$

The rest of the proof follows similarly to the previous case. □

### A.5.1. Case 1. Colorless $S_i$

Assume that $S_i \subseteq \mathcal{W}$,

**Lemma A.14.** *If $S_i \subseteq \mathcal{W}$ and $u \in \mathcal{W}$, for any arbitrary $v_i$ we may take:*

$$\widehat{\Delta\mathcal{G}}_i^{\text{gr}}(S, u)$$
$$= \begin{cases} \mathsf{err}^{\text{indv}}(u) \cdot \prod_{\substack{v_j \in S \\ \bar{W}_{ij}\neq 0}} (1 - \mathsf{err}^{\text{indv}}(v_j)) & \text{if } \Psi_i(\mathcal{R}\cup S, \mathcal{B}\cup\{u\}) < 0 \;\&\Psi_i(\mathcal{B}\cup S, \mathcal{R}\cup\{u\}) < 0 \\ \mathsf{err}^{\text{indv}}(u) \cdot \prod_{\substack{v_j \in S \\ \bar{W}_{ij}\neq 0}} (1 - \mathsf{err}^{\text{indv}}(v_j)) \cdot \mathsf{err}(\mathcal{R}) & \text{if } \Psi_i(\mathcal{R}\cup S, \mathcal{B}\cup\{u\}) > 0 \;\&\Psi_i(\mathcal{B}\cup S, \mathcal{R}\cup\{u\}) < 0 \\ \mathsf{err}^{\text{indv}}(u) \cdot \prod_{\substack{v_j \in S \\ \bar{W}_{ij}\neq 0}} (1 - \mathsf{err}^{\text{indv}}(v_j)) \cdot \mathsf{err}(\mathcal{B}) & \text{if } \Psi_i(\mathcal{R}\cup S, \mathcal{B}\cup\{u\}) < 0 \;\&\Psi_i(\mathcal{B}\cup S, \mathcal{R}\cup\{u\}) > 0 \\ 0 & \text{otherwise} \end{cases}$$

*and we have:*

$$\left|\widehat{\Delta\mathcal{G}}_i^{\text{gr}}(S, u) - \Delta\mathcal{G}_i^{\text{gr}}(S, u)\right| \leq \exp\left(-\frac{(\Psi_i^{\mathcal{W}}(u) - \Delta\bar{W}_i)^2}{4\sum_{v_j\in\mathcal{W}} \bar{W}_{ij}^2}\right)$$

*Proof.*

$$\underset{a\sim\Omega}{\overset{\text{gr}}{\mathbb{P}}}(\mathcal{Z}(i,a) \leq 0 \wedge \mathcal{T}_i(S) \wedge \mathcal{F}(u)) = \underset{a\sim\Omega^+}{\overset{\text{gr}}{\mathbb{P}}}(z(i,a) \leq 0 \wedge \mathcal{T}_i(S) \wedge \mathcal{F}(u))\,\mathbb{P}(a \in \Omega^+)$$
$$+ \underset{a\sim\Omega^-}{\overset{\text{gr}}{\mathbb{P}}}(z(i,a) \geq 0 \wedge \mathcal{T}_i(S) \wedge \mathcal{F}(u))\,\mathbb{P}(a \in \Omega^-)\,.$$

We have:

$$\underset{a\sim\Omega^+}{\overset{\text{gr}}{\mathbb{P}}}(z(i,a) \leq 0 \wedge \mathcal{T}_i(S) \wedge \mathcal{F}(u)) = \underset{a\sim\Omega^+}{\overset{\text{gr}}{\mathbb{P}}}(z(i,a) \leq 0 \mid \mathcal{T}_i(S) \wedge \mathcal{F}(u)) \underset{a\sim\Omega^+}{\overset{\text{gr}}{\mathbb{P}}}(\mathcal{T}_i(S) \wedge \mathcal{F}(u)) \;\&$$
$$\underset{a\sim\Omega^-}{\overset{\text{gr}}{\mathbb{P}}}(z(i,a) \leq 0 \wedge \mathcal{T}_i(S) \wedge \mathcal{F}(u)) = \underset{a\sim\Omega^-}{\overset{\text{gr}}{\mathbb{P}}}(z(i,a) \leq 0 \mid \mathcal{T}_i(S) \wedge \mathcal{F}(u)) \underset{a\sim\Omega^-}{\overset{\text{gr}}{\mathbb{P}}}(\mathcal{T}_i(S) \wedge \mathcal{F}(u))$$

Note that since $S \subseteq \mathcal{W}$ and $u \in \mathcal{W}$ we have:

$$\underset{a\sim\Omega^-}{\overset{\text{gr}}{\mathbb{P}}}(\mathcal{T}_i(S) \wedge \mathcal{F}(u)) = \underset{a\sim\Omega^+}{\overset{\text{gr}}{\mathbb{P}}}(\mathcal{T}_i(S) \wedge \mathcal{F}(u)) = \mathsf{err}^{\text{indv}}(u) \cdot \prod_{\substack{v_j \in S \\ \bar{W}_{ij}\neq 0}} (1 - \mathsf{err}^{\text{indv}}(v_j))\,.$$

Furthermore,

$$
\begin{aligned}
&\overset{\text{gr}}{\underset{a\sim\Omega^+}{\mathbb{P}}}\left(z(i,a)\le 0 \mid \mathcal{T}_i(S)\wedge\mathcal{F}(u)\right)\\
=&\overset{\text{gr}}{\underset{a\sim\Omega^+}{\mathbb{P}}}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j(a)+\sum_{v_j\in S}\bar{W}_{ij}-\bar{W}_{iu}\le 0\right)\\
=&\underset{a\sim\Omega^+}{\mathbb{P}}\left(\sum_{\substack{v_j\in\mathcal{W}\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)+\sum_{v_j\in\mathcal{R}}\bar{W}_{ij}\hat{y}_j^{\text{gr}}(a)+\sum_{v_j\in\mathcal{B}}\bar{W}_{ij}\hat{y}_j^{\text{gr}}(a)+\sum_{v_j\in S}\bar{W}_{ij}-\bar{W}_{iu}\le 0\right)\\
=&\underset{a\sim\Omega^+}{\mathbb{P}}\left(\sum_{\substack{v_j\in\mathcal{W}\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)+\sum_{v_j\in\mathcal{R}}\bar{W}_{ij}-\sum_{v_j\in\mathcal{B}}\bar{W}_{ij}+\sum_{v_j\in S}\bar{W}_{ij}-\bar{W}_{iu}\le 0\right)\mathsf{err}(\mathcal{B})\\
&\quad+\underset{a\sim\Omega^+}{\mathbb{P}}\left(\sum_{\substack{v_j\in\mathcal{W}\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)+\sum_{v_j\in\mathcal{B}}\bar{W}_{ij}-\sum_{v_j\in\mathcal{R}}\bar{W}_{ij}+\sum_{v_j\in S}\bar{W}_{ij}-\bar{W}_{iu}\le 0\right)\mathsf{err}(\mathcal{R})\\
=&\Gamma_i^+(\mathcal{R}\cup S,\mathcal{B}\cup\{u\})\mathsf{err}(\mathcal{B})+\Gamma_i^+(\mathcal{B}\cup S,\mathcal{R}\cup\{u\})\mathsf{err}(\mathcal{R})\,.
\end{aligned}
$$

Similarly we have that:

$$
\begin{aligned}
&\overset{\text{gr}}{\underset{a\sim\Omega^-}{\mathbb{P}}}\left(z(i,a)\ge 0 \mid \mathcal{T}_i(S)\wedge\mathcal{F}(u)\right)\\
=&\overset{\text{gr}}{\underset{a\sim\Omega^-}{\mathbb{P}}}\left(\sum_{\substack{j=1:n\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j(a)-\sum_{v_j\in S}\bar{W}_{ij}+\bar{W}_{iu}\ge 0\right)\\
=&\underset{a\sim\Omega^-}{\mathbb{P}}\left(\sum_{\substack{v_j\in\mathcal{W}\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)+\sum_{v_j\in\mathcal{R}}\bar{W}_{ij}\hat{y}_j^{\text{gr}}(a)+\sum_{v_j\in\mathcal{B}}\bar{W}_{ij}\hat{y}_j^{\text{gr}}(a)-\sum_{v_j\in S}\bar{W}_{ij}+\bar{W}_{iu}\ge 0\right)\\
=&\underset{a\sim\Omega^-}{\mathbb{P}}\left(\sum_{\substack{v_j\in\mathcal{W}\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)-\sum_{v_j\in\mathcal{R}}\bar{W}_{ij}+\sum_{v_j\in\mathcal{B}}\bar{W}_{ij}-\sum_{v_j\in S}\bar{W}_{ij}+\bar{W}_{iu}\ge 0\right)\mathsf{err}(\mathcal{B})\\
&\quad+\underset{a\sim\Omega^-}{\mathbb{P}}\left(\sum_{\substack{v_j\in\mathcal{W}\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)-\sum_{v_j\in\mathcal{B}}\bar{W}_{ij}+\sum_{v_j\in\mathcal{R}}\bar{W}_{ij}-\sum_{v_j\in S}\bar{W}_{ij}+\bar{W}_{iu}\ge 0\right)\mathsf{err}(\mathcal{R})\\
=&\Gamma_i^-(\mathcal{R}\cup S,\mathcal{B}\cup\{u\})\mathsf{err}(\mathcal{B})+\Gamma_i^-(\mathcal{B}\cup S,\mathcal{R}\cup\{u\})\mathsf{err}(\mathcal{R})\,.
\end{aligned}
$$

Therefore, $\mathbb{P}^{\text{gr}}_{a\sim\Omega}\left(\mathcal{Z}(i,a)\le 0\wedge\mathcal{T}_i(S)\wedge\mathcal{F}(u)\right)$ is equal to

$$
\begin{aligned}
\mathsf{err}^{\text{indv}}(u)\cdot\prod_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}}(1-\mathsf{err}^{\text{indv}}(v_j))\cdot\big(&[\Gamma_i^+(\mathcal{R}\cup S,\mathcal{B}\cup\{u\})\mathsf{err}(\mathcal{B})+\Gamma_i^+(\mathcal{B}\cup S,\mathcal{R}\cup\{u\})\mathsf{err}(\mathcal{R})]\,\mathbb{P}(a\in\Omega^+)\\
&+[\Gamma_i^-(\mathcal{R}\cup S,\mathcal{B}\cup\{u\})\mathsf{err}(\mathcal{B})+\Gamma_i^-(\mathcal{B}\cup S,\mathcal{R}\cup\{u\})\mathsf{err}(\mathcal{R})]\,\mathbb{P}(a\in\Omega^-)\big)
\end{aligned}
$$

We may now employ Lemma A.13 to estimate the above probabilities.

By replacing the estimations $\widehat{\Gamma}_i^+$ and $\widehat{\Gamma}_i^-$ in the above formula we obtain:

$$\widehat{\Delta\mathcal{G}}_i^{\mathrm{gr}}(S,u)$$
$$= \begin{cases} \mathsf{err}^{\mathrm{indv}}(u)\cdot\prod_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}}(1-\mathsf{err}^{\mathrm{indv}}(v_j)) & \text{if } \Psi_i(\mathcal{R}\cup S,\mathcal{B}\cup\{u\})<0 \;\&\Psi_i(\mathcal{B}\cup S,\mathcal{R}\cup\{u\})<0 \\ \mathsf{err}^{\mathrm{indv}}(u)\cdot\prod_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}}(1-\mathsf{err}^{\mathrm{indv}}(v_j))\cdot\mathsf{err}(\mathcal{R}) & \text{if } \Psi_i(\mathcal{R}\cup S,\mathcal{B}\cup\{u\})>0 \;\&\Psi_i(\mathcal{B}\cup S,\mathcal{R}\cup\{u\})<0 \\ \mathsf{err}^{\mathrm{indv}}(u)\cdot\prod_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}}(1-\mathsf{err}^{\mathrm{indv}}(v_j))\cdot\mathsf{err}(\mathcal{B}) & \text{if } \Psi_i(\mathcal{R}\cup S,\mathcal{B}\cup\{u\})<0 \;\&\Psi_i(\mathcal{B}\cup S,\mathcal{R}\cup\{u\})>0 \\ 0 & \text{otherwise} \end{cases}$$

□

**Lemma A.15.** *If $S_i\subseteq\mathcal{W}$ and $u\in\mathcal{R}$, for any arbitrary $v_i$ we may take:*

$$\widehat{\Delta\mathcal{G}}_i^{\mathrm{gr}}(S,u)=\begin{cases}\mathsf{err}^{\mathrm{indv}}(u)\cdot\prod_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}}(1-\mathsf{err}^{\mathrm{indv}}(v_j)) & \textit{if } \Psi_i(\mathcal{B}\cup S,\mathcal{R}\cup\{u\})<0\\ 0 & \textit{otherwise}\end{cases}$$

*If $u\in\mathcal{B}$, for any arbitrary $v_i$ we may take:*

$$\widehat{\Delta\mathcal{G}}_i^{\mathrm{gr}}(S,u)=\begin{cases}\mathsf{err}^{\mathrm{indv}}(u)\cdot\prod_{\substack{v_j\in S\\ \bar{W}_{ij}\neq 0}}(1-\mathsf{err}^{\mathrm{indv}}(v_j)) & \textit{if } \Psi_i(\mathcal{R}\cup S,\mathcal{B}\cup\{u\})<0\\ 0 & \textit{otherwise}\end{cases}$$

*and in both cases we have:*

$$\left|\widehat{\Delta\mathcal{G}}_i^{\mathrm{gr}}(S,u)-\Delta\mathcal{G}_i^{\mathrm{gr}}(S,u)\right|\leq\exp\left(-\frac{(\Psi_i^{\mathcal{W}}(u)-\Delta\bar{W}_i)^2}{4\sum_{v_j\in\mathcal{W}}\bar{W}_{ij}^2}\right).$$

*Proof.* Like previous lemma we have:

$$\begin{aligned}&\mathop{\mathbb{P}}_{a\sim\Omega^+}^{\mathrm{gr}}(z(i,a)\leq 0\mid\mathcal{T}_i(S)\wedge\mathcal{F}(u))\\ &=\mathop{\mathbb{P}}_{a\sim\Omega^+}\left(\sum_{\substack{v_j\in\mathcal{W}\\ v_j\notin S\cup\{u\}}}\bar{W}_{ij}\hat{y}_j^{\mathrm{indv}}(a)+\sum_{v_j\in\mathcal{R}}\bar{W}_{ij}\hat{y}_j^{\mathrm{gr}}(a)+\sum_{v_j\in\mathcal{B}}\bar{W}_{ij}\hat{y}_j^{\mathrm{gr}}(a)+\sum_{v_j\in S}\bar{W}_{ij}-\bar{W}_{iu}\leq 0\right)\end{aligned}$$

If $u\in\mathcal{R}$ since we are conditioning on $\mathcal{F}(u)$ the above probability is equal to:

$$\mathop{\mathbb{P}}_{a\sim\Omega^+}^{\mathrm{gr}}(z(i,a)\leq 0\mid\mathcal{T}_i(S)\wedge\mathcal{F}(u))=\Gamma_i^+(\mathcal{B}\cup S,\mathcal{R}\cup\{u\}).$$

Similarly

$$\mathop{\mathbb{P}}_{a\sim\Omega^+}^{\mathrm{gr}}(z(i,a)\geq 0\mid\mathcal{T}_i(S)\wedge\mathcal{F}(u))=\Gamma_i^-(\mathcal{B}\cup S,\mathcal{R}\cup\{u\}).$$

If $u\in\mathcal{B}$ we have

$$\mathop{\mathbb{P}}_{a\sim\Omega^+}^{\mathrm{gr}}(z(i,a)\leq 0\mid\mathcal{T}_i(S)\wedge\mathcal{F}(u))=\Gamma_i^+(\mathcal{R}\cup S,\mathcal{B}\cup\{u\}).$$

Similarly

$$\underset{a\sim\Omega^+}{\overset{\text{gr}}{\mathbb{P}}}\left(z(i,a)\geq 0 \mid \mathcal{T}_i(S)\wedge\mathcal{F}(u)\right)=\Gamma_i^-\left(\mathcal{R}\cup S,\mathcal{B}\cup\{u\}\right).$$

Putting together, and employing Lemma A.13 we get the premise.

□

### A.5.2. Case 2. monochromatic $S_i$

**Lemma A.16.** *Let $\mathcal{C}$ be either $\mathcal{R}$ or $\mathcal{B}$ and $\bar{\mathcal{C}}$ be the other color. Assume $S_i\cap\mathcal{C}\neq\emptyset$ and $S_i\cap\bar{\mathcal{C}}=\emptyset$. In this case, for any $u\in\mathcal{C}$, we have $\forall i,\Delta\mathcal{G}_i^{\text{gr}}(S,u)=0$.*

*For $u\in\bar{\mathcal{C}}$ we may use the following estimation:*

$$\widehat{\Delta\mathcal{G}}_i^{\text{gr}}(S,u)=\begin{cases}\mathsf{err}(\bar{\mathcal{C}})\prod_{v_j\in S_i\cap\mathcal{W}}(1-\mathsf{err}^{\text{indv}}(v_j)) & \text{if}\quad \Psi_i(\mathcal{C}\cup S,\bar{\mathcal{C}})<0\\ 0 & \text{otherwise}\ .\end{cases}$$

*and for $u\in\mathcal{W}$ we may use the following estimation:*

$$\widehat{\Delta\mathcal{G}}_i^{\text{gr}}(S,u)=\begin{cases}\mathsf{err}(u)\mathsf{err}(\bar{\mathcal{C}})\prod_{v_j\in S_i\cap\mathcal{W}}(1-\mathsf{err}^{\text{indv}}(v_j)) & \text{if}\quad \Psi_i(\mathcal{C}\cup S,\bar{\mathcal{C}}\cup\{u\})<0\\ 0 & \text{otherwise}\ .\end{cases}$$

*The above estimations satisfy:*

$$\left|\widehat{\Delta\mathcal{G}}_i^{\text{gr}}(S,u)-\Delta\mathcal{G}_i^{\text{gr}}(S,u)\right|\leq\exp\left(-\frac{(\Psi_i^{\mathcal{W}}(u)-\Delta\bar{W}_i)^2}{4\sum_{v_j\in\mathcal{W}}\bar{W}_{ij}^2}\right).$$

*Proof of Lemma A.16.* Assume that $S_i$ intersects only with one color $\mathcal{C}$ and its intersection with the other color $\bar{\mathcal{C}}$ is empty.

If $u\in\mathcal{C}$, then $\mathbb{P}(\mathcal{T}_i(S)\wedge\mathcal{F}(u))=0$. Thus, $\forall i\ \Delta\mathcal{G}_i^{\text{gr}}(S,u)=0$.

If $u\in\bar{\mathcal{C}}$:

We have that $\mathbb{P}(\mathcal{T}_i(S)\wedge\mathcal{F}(u))=\mathsf{err}(\bar{\mathcal{C}})\cdot\prod_{v_j\in S_i\cap\mathcal{W}}(1-\mathsf{err}^{\text{indv}}(v_j))$ . Furthermore,

$$\begin{aligned}\underset{a\in\Omega^+}{\overset{\text{gr}}{\mathbb{P}}}\left(z^*(i,a)\leq 0 \mid \mathcal{T}_i(S)\wedge\mathcal{F}(u)\right)&=\mathbb{P}\left(\sum_{v_j\in\mathcal{W}\setminus S_i}\bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)+\bar{W}(\mathcal{C})-\bar{W}(\bar{\mathcal{C}})+\bar{W}(S\cap\mathcal{W})\leq 0\right)\\&=\Gamma_i^+\left(\mathcal{C}\cup S,\bar{\mathcal{C}}\right).\end{aligned}$$

Similarly,

$$\underset{a\in\Omega^-}{\overset{\text{gr}}{\mathbb{P}}}\left(z^*(i,a)\geq 0 \mid \mathcal{T}_i(S)\wedge\mathcal{F}(u)\right)=\Gamma_i^-\left(\mathcal{C}\cup S,\bar{\mathcal{C}}\right).$$

Putting together and employing Lemma A.13 we conclude the first part of the premise.

If $u\in\mathcal{W}$: We have that $\mathbb{P}(\mathcal{T}_i(S)\wedge\mathcal{F}(u))=\mathsf{err}(u)\mathsf{err}(\bar{\mathcal{C}})\cdot\prod_{v_j\in S_i\cap\mathcal{W}}(1-\mathsf{err}^{\text{indv}}(v_j))$. Furthermore,

$$\begin{aligned}\underset{a\in\Omega^+}{\overset{\text{gr}}{\mathbb{P}}}\left(z^*(i,a)\leq 0 \mid \mathcal{T}_i(S)\wedge\mathcal{F}(u)\right)&=\mathbb{P}\left(\sum_{v_j\in\mathcal{W}\setminus S_i}\bar{W}_{ij}\hat{y}_j^{\text{indv}}(a)+\bar{W}_i(\mathcal{C})-\bar{W}_i(\bar{\mathcal{C}})+\bar{W}_i(S\cap\mathcal{W})-\bar{W}_{iu}\leq 0\right)\\&=\Gamma_i^+\left(\mathcal{C}\cup S,\bar{\mathcal{C}}\cup\{u\}\right).\end{aligned}$$

Similarly,

$$\underset{a\in\Omega^-}{\overset{\text{gr}}{\mathbb{P}}}\left(z^*(i,a)\geq 0 \mid \mathcal{T}_i(S)\wedge\mathcal{F}(u)\right)=\Gamma_i^-\left(\mathcal{C}\cup S,\bar{\mathcal{C}}\cup\{u\}\right)\ .$$

Putting together and employing Lemma A.13 we conclude the second part of the premise. □

### A.5.3. CASE 3. BICHROMATIC $S_i$

**Lemma A.17.** *If $S_i\cap\mathcal{R}\neq\emptyset$ and $S_i\cap\mathcal{B}\neq\emptyset$, we have $\forall i\ \Delta\mathcal{G}_i^{\text{gr}}(S,u)=0$.*

*Proof.* Note that $\mathbb{P}(\mathcal{T}_i(S))=0$. Thus, we conclude the premise. □

### A.5.4. MISSING MATERIAL FROM SECTION 4.2.2: $\mathcal{W}$-AMBIGUOUS VERTICES.

Let's first present the definition of $\mathcal{W}$-ambiguous vertices in detail:

Consider the following partitioning of white vertices to low and high error parts:

$$\mathcal{W}^+=\{v_j\mid \mathsf{err}(v_j)\leq 1/2\}\ \ \&\ \mathcal{W}^-=\{v_j\mid \mathsf{err}(v_j)>1/2\}$$

with respect to this partition we define the following vectors:

$$\mathcal{E}^{\mathcal{W}^+}=(1-2\mathsf{err}(v_j))_{v_j\in\mathcal{W}^+}\quad\&\quad \mathcal{E}^{\mathcal{W}^-}=(2\mathsf{err}(v_j)-1)_{v_j\in\mathcal{W}^-}$$

$$\text{and}\qquad \bar{W}_i^{\mathcal{W}+}=(\bar{W}_{ij})_{v_j\in\mathcal{W}^+}\quad\&\quad \bar{W}_i^{\mathcal{W}-}=(\bar{W}_{ij})_{v_j\in\mathcal{W}^-}$$

**Definition A.18** ( $\mathcal{W}$-Ambiguous vertices )**.** Let $\bar{W}_i^{\mathcal{W}}=(\bar{W}_{ij})_{v_j\in\mathcal{W}}$, and $|\cdot|_2$ be the $\ell_2$ norm and $\langle\cdot,\cdot\rangle$ be dot product. We call an agent $v_i\in V$, *$\mathcal{W}$-ambiguous* if it satisfies

$$\left|\frac{\langle\bar{W}_i^{\mathcal{W}+},\mathcal{E}^{\mathcal{W}^+}\rangle}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}-\frac{\langle\bar{W}_i^{\mathcal{W}-},\mathcal{E}^{\mathcal{W}^-}\rangle}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}\right|\leq 4\sqrt{\log n}+\Delta\bar{W}_i\ ,$$

where $\Delta\bar{W}_i=\left|\sum_{v_j\in\mathcal{R}}\bar{W}_{ij}-\sum_{v_j\in\mathcal{B}}\bar{W}_{ij}\right|$ . A *network* is *nicely colored* if no vertex is $\mathcal{W}$-ambiguous.

We now show that if a vertex is *not* ambiguous w.r.t. $\mathcal{W}$ the group gain associated to it can be estimated with high precision:

**Lemma A.19.** *Let $\widehat{\Delta\mathcal{G}_i}^{\text{gr}}(S,u)$ be as defined in Lemmas A.14, A.16 and A.17. If a vertex is* non-ambiguous *w.r.t $\mathcal{W}$ we have:*

$$\left|\Delta\mathcal{G}_i^{\text{gr}}(S,u)-\widehat{\Delta\mathcal{G}_i}^{\text{gr}}(S,u)\right|\leq o(n^{-1})\ .$$

*Proof.* Note that from Lemmas A.14, A.16 and A.17 we have that for any $v_i$:

$$\left|\Delta\mathcal{G}_i^{\text{gr}}(S,u)-\widehat{\Delta\mathcal{G}_i}^{\text{gr}}(S,u)\right|\leq\exp\left(-\frac{(\Psi_i^{\mathcal{W}}(u)-\Delta\bar{W}_i)^2}{4\sum_{v_j\in\mathcal{W}}\bar{W}_{ij}^2}\right)\ .$$

Note that we have:

$$\begin{aligned}\frac{(\Psi_i^{\mathcal{W}}(u)-\Delta\bar{W}_i)^2}{\sum_{v_j\in\mathcal{W}}\bar{W}_{ij}^2}&=\left(\frac{\Psi_i^{\mathcal{W}}(u)-\Delta\bar{W}_i}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}\right)^2\\&=\left(\frac{\langle\bar{W}_i^{\mathcal{W}+},\mathcal{E}^{\mathcal{W}+}\rangle}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}-\frac{\langle\bar{W}_i^{\mathcal{W}-},\mathcal{E}^{\mathcal{W}-}\rangle}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}-\frac{2\mathsf{err}(u)\bar{W}_{iu}}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}-\frac{\Delta\bar{W}_i}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}\right)^2\\&\geq\left(\frac{\langle\bar{W}_i^{\mathcal{W}+},\mathcal{E}^{\mathcal{W}+}\rangle}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}-\frac{\langle\bar{W}_i^{\mathcal{W}-},\mathcal{E}^{\mathcal{W}-}\rangle}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}-\frac{\Delta\bar{W}_i}{\left|\bar{W}_i^{\mathcal{W}}\right|_2}-2\right)^2\end{aligned}$$

Assuming that the vertex is not ambiguous w.r.t. whites we have:

$$\left| \frac{\langle \bar{W}_i^{\mathcal{W}+}, \mathcal{E}^{\mathcal{W}+} \rangle}{\left|\bar{W}_i^{\mathcal{W}}\right|_2} - \frac{\langle \bar{W}_i^{\mathcal{W}-}, \mathcal{E}^{\mathcal{W}-} \rangle}{\left|\bar{W}_i^{\mathcal{W}}\right|_2} \right| > 4\sqrt{\log n} + \Delta \bar{W}_i$$

which implies

$$\frac{(\Psi_i^{\mathcal{W}}(u) - \Delta \bar{W}_i)^2}{\sum_{v_j \in \mathcal{W}} \bar{W}_{ij}^2} = \left( \frac{\Psi_i^{\mathcal{W}}(u) - \Delta \bar{W}_i}{\left|\bar{W}_i^{\mathcal{W}}\right|_2} \right)^2 \geq (3\sqrt{\log n})^2 \geq 5 \log n$$

From which we conclude that the error is bounded as:

$$\exp\left( -\frac{(\Psi_i^{\mathcal{W}}(u) - \Delta \bar{W}_i)^2}{\sum_{v_j \in \mathcal{W}} \bar{W}_{ij}^2} \right) \leq \exp(-5/4 \log n) = o(n^{-1}) \,.$$

□

#### A.5.5. MISSING MATERIAL FROM SECTION 4.2.2: PROOF OF THE MAIN THEOREMS

***Proof of Theorem 4.8 and Remark 3.12.*** Note that the error of the greedy choice approximation is either generated from approximation of $\sum_{i=1}^n \widehat{\Delta \mathcal{G}}_i^{\text{gr}}(S,u)$ or $\sum_{i=1}^n \widehat{\Delta \mathcal{G}}_i^{\text{indv}}(S,u)$. The error of $\widehat{\Delta \mathcal{G}}_i^{\text{indv}}(S,u)$ may be bounded similar to the independent case. To see the bound the error of $\sum_{i=1}^n \widehat{\Delta \mathcal{G}}_i^{\text{gr}}(S,u)$ note that the error induced by all *non-ambiguous* vertices is $o(n \times n^{-1}) = o(1)$. The error of each ambiguous vertex is at most one, therefore, we conclude the result.

**Approximation when having access $\widehat{\mathcal{W}}$.** Follows similarly as in Lemma A.11 □

### A.6. Concentration Bounds

**Theorem A.20.** *Let $X_1, X_2, \ldots X_k$ be independent random variables in range $[-1, 1]$ with means $\mu_1, \ldots, \mu_k$ and let $w_i$s be weight coefficients . Taking $\mu = \sum_{i=1}^k w_i \mu_i$. We have that for any $\varepsilon > 0$:*

$$\mathbb{P}\left( \sum_{i=1}^k w_i X_i \geq \mu + \varepsilon \right) \leq \exp\left( -\frac{\varepsilon^2}{4 \sum_{i=1}^k w_i^2} \right),$$

*and*

$$\mathbb{P}\left( \sum_{i=1}^k w_i X_i \leq \mu - \varepsilon \right) \leq \exp\left( -\frac{\varepsilon^2}{4 \sum_{i=1}^k w_i^2} \right) \,.$$

## B. Additional experiments and details of set up

Our proposed problem and methods in general do not assume any prior knowledge of the given datasets. Our algorithms are deterministic with no parameter to tune. It has potential to be applied to any problems as long as the objective function of the problems of interests related to our proposed objective function. In the following, we include the parameter settings of the problems in our experiments. These parameters are not tuned for our methods. We just fix the parameters to make sure the problems are nontrivial enough. We use the same parameters for all methods in our experiments.

We let $|\Omega| = 3$ and for initial opinion $\hat{\mathbf{y}}$, we randomly generate for each agent $v_i$ a random vector $(\hat{y}_i(a) : a \in \Omega)$ with each entry sampled from Bernoulli distribution with probability $p_i$ of $\hat{y}_i(a) = +1$ sampled uniformly from $[0.3, 0.9]$. For random graph generators, some of the key parameters are fixed as follows:

- Number of nodes: 128
- Erdős-Rényi graph: Probability for edge creation $p = 0.005$.
- Barabási-Albert preferential attachment model: Number of edges to attach from a new node to existing nodes $m = 5$

- Watts-Strogatz small-world graph: Each node is joined with its $k = 5$ nearest neighbors in a ring topology; The probability of rewiring each edge $p = 0.25$

In practice, we found that the infinite FJ model with weight matrix converging to $(I + L)^{-1}$ makes the problem less interesting than general case in practice. The reason is that all entries of the matrix $(I + L)^{-1}$ are positive and non-zero. With such weight matrix, we found even a random selection algorithm can converge very fast. To avoid such simple cases, here we apply a finite $t$-step FJ model. We fix $t = 3$ to ensure the induced matrix $\bar{W}$ is sparse enough.

Besides randomly generated graphs from tree classical models (ER, PA and WS), we also test our algorithms on random generated matrix $\bar{W}$ directly, denoted as RandomW. Each entry of $\bar{W}$ is independently sampled from a uniform distribution on $[0, 1]$. We let the sparsity of $\bar{W}$ to be around $0.95$. Each row of $\bar{W}$ is normalized thus sum up to be $1$.

The statistics of real and synthetic datasets are summarized in Table 2. Experiment results are illustrated in Figure 3.

We let the set of faulty prediction be $\mathcal{Z}^f \triangleq \mathbb{E}_{a\in\Omega}[\sum_{i=1}^{n} \mathbf{1}\mathcal{Z}(i, a) < 0]$ , which serves as an upperbound on the egalitarian improvement. We define the accuracy Acc as:

$$\mathrm{Acc} = \frac{\mathcal{G}^{(\mathrm{egal})}(S)}{\mathcal{Z}^f} \ .$$

We calculate all expected values by taking averages over $a \in \Omega$.

Table 2: Table of statistics of datasets. Sparsity of $\bar{W}$ represents the percentage of zero entries in $\bar{W}$.

| | size | sparsity of $\bar{W}$ | faulty prediction $\mathcal{Z}^f$ |
|---|---|---|---|
| ER | 128 | 0.98 | 74.67 |
| PA | 128 | 0.37 | 32.0 |
| WS | 128 | 0.74 | 38.67 |
| BIO | 297 | 0.45 | 55.33 |
| CSPK | 39 | 0.75 | 14.67 |
| FB | 620 | 0.75 | 161.33 |
| WIKI | 890 | 0.66 | 238.67 |
| RandomW | 128 | 0.95 | 37.67 |

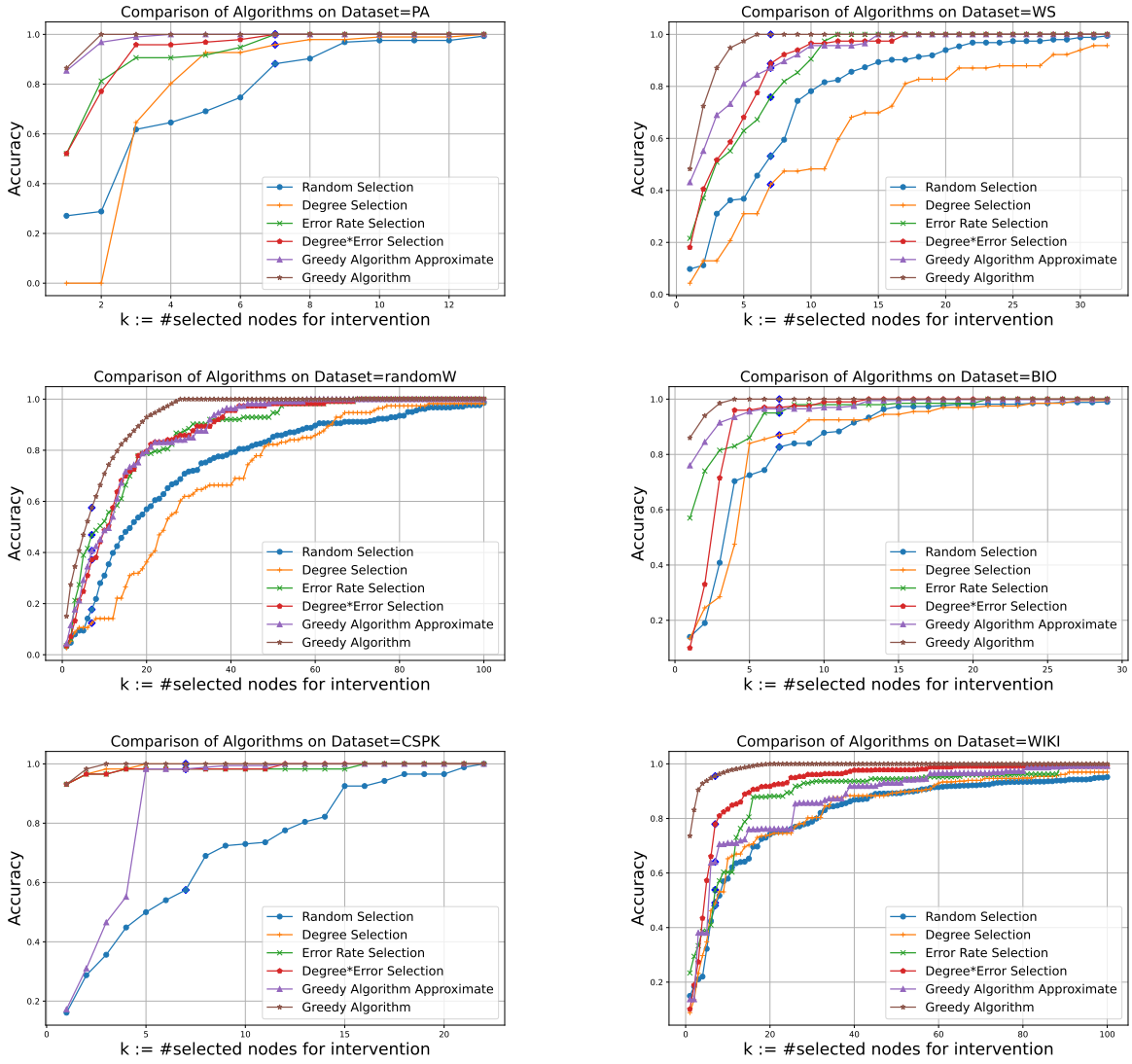



Figure 3: More experimental results on different datasets.