

# Cluster-Adaptive Network A/B Testing: From Randomization to Estimation

**Yang Liu**

*Institute of Statistics and Big Data  
Renmin University of China  
Beijing, 100872, China*

YANGLIU2022@RUC.EDU.CN

**Yifan Zhou**

*Department of Statistics  
George Washington University  
Washington, DC 22202, USA*

YZHOU92@GWU.EDU

**Ping Li**

*VecML Inc.  
Bellevue, WA 98004, USA*

PINGLI98@GMAIL.COM

**Feifang Hu**

*Department of Statistics  
George Washington University  
Washington, DC 22202, USA*

FEIFANG@GWU.EDU

**Editor:** Edoardo Maria Airoidi

## Abstract

The performance of A/B testing in both online and offline experimental settings hinges on mitigating network interference and achieving covariate balancing. These experiments often involve an observable network with identifiable clusters, and measurable cluster-level and individual-level attributes. Exploiting these inherent characteristics holds potential for refining experimental design and subsequent statistical analyses. In this article, we propose a novel cluster-adaptive network A/B testing procedure, which contains a cluster-adaptive randomization (CLAR) and a cluster-adjusted estimator (CAE) to facilitate the design of the experiment and enhance the performance of ATE estimation. The CLAR sequentially assigns clusters to minimize the Mahalanobis distance, which further leads to the balance of the cluster-level covariates and the within-cluster-averaged individual-level covariates. The cluster-adjusted estimator (CAE) is tailored to offset biases caused by network interference. The proposed procedure has the following two folds of the desirable properties. First, we show that the Mahalanobis distance calculated for the two levels of covariates is  $O_p(m^{-1})$ , where  $m$  represents the number of clusters. This result justifies the simultaneous balance of the cluster-level and individual-level covariates. Under mild conditions, we derive the asymptotic normality of CAE and demonstrate the benefit of covariate balancing on improving the precision for estimating ATE. The proposed A/B testing procedure is easy to calculate, consistent, and achieves higher accuracy. Extensive numerical studies are conducted to demonstrate the finite sample property of the proposed network A/B testing procedure.

**Keywords:** Network interference, Covariate balance, Spillover effect, Adaptive design, Graph cluster randomization.

## 1. Introduction

Network A/B testing (Gui et al., 2015) refers to the problem of conducting randomized controlled experiments in the presence of network interference (Manski, 2000), particularly when units are parts of a network. While A/B testing serves as a prominent tool in decision-making within IT industries (Kohavi et al., 2013; Larsen et al., 2024), the presence of network-induced interference challenges the fundamental assumption of *Stable Unit Treatment Value* (SUTVA) inherent in conventional A/B testing (Imbens and Rubin, 2015; Rubin, 1974), rendering the estimation of average treatment effects (ATE) more intricate. This phenomenon, known as *peer effect* or *network interference*, poses significant challenges for network A/B testing.

In various applications, networks often display distinctive cluster structures (Gui et al., 2015; Holtz et al., 2020; Ugander et al., 2011, 2013), which can be leveraged to mitigate the bias stemming from interference and to enhance efficiency in estimating ATE. These clusters typically comprise subsets of units exhibiting denser internal connections than external ones. When interference is localized, cluster formation can be facilitated through community detection algorithms (Fortunato, 2010; Newman, 2006) offering the perspective of cluster-level randomization to counteract the effects of interference and increase the comparability of the results (Eckles et al., 2017; Gui et al., 2015; Ugander and Yin, 2023; Ugander et al., 2013). This strategy, referred to as graph-cluster randomization (Ugander and Yin, 2023; Ugander et al., 2013), emerges as a promising approach to deal with network interference. After the implementation of graph-cluster randomization, the local interference approximated by the *neighborhood treatment response assumption* (NTRA, see Eckles, Karrer, and Ugander, 2017; Leung, 2022a; Ugander, Karrer, Backstrom, and Kleinberg, 2013) can be used to construct a Horvitz-Thompson estimator for an unbiased ATE estimation. These three key elements, i.e., the graph cluster randomization, the neighborhood treatment assumption, and the Horivitz-Thompson’s estimator, establish a basis for the consistent and efficient estimation of the ATE in the presence of network interference. For additional recent development of the research on the experiments involving with networks, see Yu et al. (2022), Biswas and Airoidi (2018), Karwa and Airoidi (2018), Pouget-Abadie et al. (2019), Shi et al. (2023), Chen et al. (2023), Ogburn et al. (2024), Gao and Ding (2023), Imai et al. (2021), Forastiere et al. (2022) and the references therein.

Notwithstanding the strengths of the aforementioned approach, effective exploitation of cluster information has the potential to enhance the performance of network A/B testing. For instance, cluster-level covariates and individual-level covariates can be included in the randomization to enhance the balance of treatment arms. The covariate balance is crucial to ensure the credibility of the experiment (Fisher, 1935; Kohavi et al., 2013), particularly when connected units within a cluster demonstrate similar treatment responses due to shared covariate attributes, such as hobbies and education status. Additionally, network features of a cluster, such as the average degree, and the number of nodes, may also influence the evaluation of ATE. Relying solely on randomization for balancing covariates may result in substantial covariate imbalance, ultimately affecting the performance of A/B testing procedure in ATE estimation (Hu and Hu, 2012; Morgan and Rubin, 2012; Rosenberger and Sverdlov, 2008).

Regarding the estimation stage, the weights used by the Horvitz-Thompson estimator are contingent on the cluster structure determined by the treated neighbor and the chosen randomization scheme. Certain cluster network configurations, for instance, units connecting multiple clusters, may yield exceedingly large weights (small probabilities) for specific units, regardless of randomization, while allocating relatively small weights (large probabilities) to units solely connected within their own cluster. This situation can make the Horvitz-Thompson estimator sensitive to the outcome values, introducing challenges for ATE estimation. Therefore, it is essential to have alternatively consistent estimator in the presence of network interference.

In this paper, we present a refined A/B testing procedure that effectively leverages cluster information in both randomization and estimation phases. In the randomization step, we introduce a *cluster-adaptive randomization* (CLAR) that sequentially minimizes the Mahalanobis distance, improving the balance of the treatment arms. Furthermore, we consider using both cluster-level covariates and within-cluster averages of individual-level covariates to obtain a finer covariate balancing. In the estimation step, we introduce a concept of informative units to identify appropriate units for the inclusion in the estimation step based on NTRA criteria and network information. Based on the informative units, we propose a *cluster-adjusted estimator* (CAE) to achieve consistent ATE estimation. Given that our approach integrates cluster information across randomization and estimation phases, we term it the *cluster-adaptive network A/B testing procedure*.

Our main contribution can be summarized in the following three folds. First, the proposed CLAR endeavors to achieve finer multisourced covariates balancing. The implications of using distinct sources of covariates are explored theoretically and numerically, underscoring the important role of cluster-level covariate balance in enhancing the efficiency for estimating ATE. Second, the proposed CAE is computationally straightforward and does not rely on weights to obtain the consistency for estimating ATE. Third, we integrate the frameworks of network interference (Aronow and Samii, 2017; Athey et al., 2018; Basse and Airoldi, 2018b; Eckles et al., 2017; Forastiere et al., 2021; Leung, 2022b; Liu et al., 2022; Ugander et al., 2013; Wang et al., 2023; Zhou et al., 2024) and covariate-adaptive randomization (Hu et al., 2023; Hu and Hu, 2012; Ma et al., 2024) to theoretically analyze our cluster-adaptive network A/B testing procedure. Techniques such as Markov chain theory (Meyn and Tweedie, 2013) and Stein’s lemma (Chen et al., 2010; Chin, 2018; Ross, 2011) are used to investigate the theoretical property of our proposed procedure. The derived results under a general outcome model and a simple additive model indicate that the proposed A/B testing procedure delivers not only consistent but also more precise ATE estimation. These findings further emphasize the importance of covariate balancing within the context of network A/B testing.

The rest of this paper is organized as follows. In Section 2, we present the framework of network A/B testing. The cluster-adaptive network A/B testing procedure is presented in Section 3. The theoretical properties of our proposed procedure are studied in Section 4. Numerical studies with a hypothetical network and a real data example are conducted in Section 5 and Section 6 to demonstrate the finite sample properties of our proposed procedure, respectively. The conclusion remarks are given in Section 7. Proofs of main results, additional numerical studies, and detailed explanations of numerical studies are relegated to the Appendices A, B and C, respectively.

## 2. Network Interference in A/B testing

Consider a network consisting of  $n$  interconnected units (nodes), represented by an undirected graph  $G = (V, E)$  with a symmetric  $n \times n$  adjacency matrix  $\mathbf{A}$ , where the  $(i, j)$ -th element is denoted by  $A_{ij}$ . Let  $\mathcal{N}_i = \{i' : A_{ii'} = 1 \text{ and } i' \neq i\}$  be the index set of the neighbors of the  $i$ -th unit and let  $d_i$  denote the size of  $\mathcal{N}_i$ , i.e., the degree of the  $i$ -th node. Suppose the  $n$  units are partitioned into  $m$  disjoint clusters (sets of nodes),  $C_1, \dots, C_m$ , with  $c_j$  denoting the number of nodes in the  $j$ -th cluster. Let  $\tilde{\mathcal{C}} = \sigma(C_1, \dots, C_m)$  represent the sigma algebra generated by the index sets of the clusters and  $\mathcal{A} = \sigma(\mathbf{A})$ . In some applications, the cluster may correspond to specific covariate information, such as the units' current education institution or their geographical location. In other examples, the label of the clusters may arise from applying community detection algorithms to the adjacency matrix  $\mathbf{A}$  (Leung, 2023; Newman and Girvan, 2004; Raghavan et al., 2007; Ugander et al., 2013). These algorithms, including label propagation and modularity maximization, can identify units that are more densely connected within a cluster than with those outside it. For a more comprehensive review of the community detection algorithms, we refer the readers to Newman (2006) and Fortunato (2010) as well as the references therein. Given that cluster labels can emerge from either of these scenarios, we assume the clusters are known and observed prior to the experiment design.

Consider an experiment with two treatments, treatment 1 and treatment 2. For  $1 \leq i \leq n$ , let  $T_i$  denote the treatment assignment for the  $i$ -th unit, such that  $T_i = 1$  indicates the assignment to treatment 1, while  $T_i = 0$  represents the assignment to treatment 2. Define  $\mathbf{T} = (T_1, \dots, T_n)^\top$  as the  $n$ -vector of the individual-level treatment assignments and  $\mathcal{T}$  as the domain of  $\mathbf{T}$ , respectively. Following the potential outcome framework (Aronow and Samii, 2017; Athey et al., 2018; Forastiere et al., 2021; Rubin, 1974), let  $Y_i(\mathbf{T})$  denote the potential outcome of the  $i$ -th subject given  $\mathbf{T}$  and  $Y_i = \sum_{\mathbf{t} \in \mathcal{T}} \mathbb{I}(\mathbf{T} = \mathbf{t}) Y_i(\mathbf{t})$  represent the observed outcome of the  $i$ -th unit, respectively.

The objective of A/B testing often involves determining whether all units should be assigned the new treatment (Gui et al., 2015; Larsen et al., 2024). In other words, it aims to compare the outcomes with  $\mathbf{T} = \mathbf{1}$  and  $\mathbf{T} = \mathbf{0}$ , where  $\mathbf{1}$  and  $\mathbf{0}$  denote the  $n$ -vectors of ones and zeros, respectively. The estimand of interest in this context is the global average treatment effect (ATE), i.e.,

$$\tau(\mathbf{1}, \mathbf{0}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i(\mathbf{1}) - Y_i(\mathbf{0})].$$

In practice, assigning all units to a single treatment is often unfeasible, making it impossible to directly observe both  $Y_i(\mathbf{1})$  and  $Y_i(\mathbf{0})$  simultaneously. However, careful design of A/B testing can yield an observed outcome  $Y_i$  that approximates either  $Y_i(\mathbf{1})$  or  $Y_i(\mathbf{0})$ . Units providing sufficient information to estimate  $\tau(\mathbf{1}, \mathbf{0})$  are referred to informative units, defined as follows:

**Definition 2.1 (Informative Units)** *Given a treatment assignment  $\mathbf{t} \in \mathcal{T}$ , the  $i$ -th unit is an informative unit, if either  $Y_i = Y_i(\mathbf{1})$  or  $Y_i = Y_i(\mathbf{0})$ ; and the  $i$ -th unit is noninformative, if  $Y_i \neq Y_i(\mathbf{1})$  and  $Y_i \neq Y_i(\mathbf{0})$ . The set of informative units in  $C_j$  is denoted by  $\text{Inf}_j$ .*

When no interference is assumed, the established *Stable Unit Treatment Value Assumption* (SUTVA) describes a scenario where all  $n$  units in the experiment are informative (Imbens and Rubin, 2015).

**Assumption 2.1 (SUTVA)** *For any  $\mathbf{t}, \mathbf{t}' \in \mathcal{T}$ , if the  $i$ -th elements of  $\mathbf{t}$  and  $\mathbf{t}'$  satisfy  $t_i = t'_i$ , then  $Y_i(\mathbf{t}) = Y_i(\mathbf{t}')$ .*

SUTVA posits that  $Y_i$  will remain unaffected by  $T_{i'}$  for  $i' \neq i$ . This can be interpreted as  $Y_i = T_i Y_i(\mathbf{1}) + (1 - T_i) Y_i(\mathbf{0})$ , forming the basis for the estimation of  $\tau(\mathbf{1}, \mathbf{0})$ .

However, in cases involving interference, a unit’s outcome may be influenced by treatment assignments of other units within the network. In the absence of interference assumptions, estimating  $\tau(\mathbf{1}, \mathbf{0})$  might be impossible (Basse and Airoidi, 2018a). To address this problem, the *Neighborhood Treatment Response Assumption* (NTRA) has been introduced (Eckles et al., 2017; Forastiere et al., 2021; Gui et al., 2015; Hong and Raudenbush, 2006; Ugander et al., 2013). The subsequent NTRA expands SUTVA by accounting for a specific form of local interference.

**Assumption 2.2 (NTRA)** *For any  $\mathbf{t}, \mathbf{t}' \in \mathcal{T}$ , if the  $k$ -th elements of  $\mathbf{t}$  and  $\mathbf{t}'$  satisfy that*

$$\text{if } t_k = t'_k \quad \text{for all } k \in \mathcal{N}_i \cup \{i\}, \quad \text{then } Y_i(\mathbf{t}) = Y_i(\mathbf{t}').$$

Under NTRA, network interference solely affects a unit’s outcome through its connected neighbors. Therefore, the outcome of a unit remains unaffected by treatment changes for units not in its neighborhood. Consequently,  $Y_i = Y_i(\mathbf{1})$  if  $t_k = 1$  for all  $k \in \mathcal{N}_i \cup \{i\}$ , and  $Y_i = Y_i(\mathbf{0})$  if  $t_k = 0$  for all  $k \in \mathcal{N}_i \cup \{i\}$ . NTRA implies that  $Y_i$  is informative for estimating  $\tau(\mathbf{1}, \mathbf{0})$  when the  $i$ -th unit, along with its neighbors, is uniformly assigned to the same treatment group.

NTRA captures the mechanism of interference observed in numerous networks. In the experiment described in Section 6, students tend to be influenced more by their connected peers than by those they are unfamiliar with. Consequently, a student might experience a spillover effect if their peers belong to different treatment groups, a situation consistent with NTRA. In addition to the social networks similar to the one presented in Section 6, spatial experiments offer another example where NTRA is applicable (Leung, 2022b). In experiments related to food delivery apps, randomization is often performed on geographical units. The food delivery services are usually localized, which may not be influenced by distant businesses; however, they could be affected by nearby stores or customers, reflecting the principles of NTRA. If it is necessary to take into account a high-order spillover effect, the NTRA considered here can be extended to the high-order NTRA discussed in Athey et al. (2018). Assumption 2.2 is presented for the sake of simplicity.

Even when NTRA is applicable, the method used for randomization remains crucial, as it may help increase the number of informative units. In numerous real-world networks, nodes within the same cluster tend to be more densely connected than those in different clusters (Newman, 2006). This observation together with NTRA highlights the benefits of the graph cluster randomization, i.e., applying complete randomization at the cluster-level, to estimate  $\tau(\mathbf{1}, \mathbf{0})$ , because it increases the probabilities that the units are informative (Eckles et al., 2017). For example, let  $\text{In}_j = \{i \in C_j : \text{if for all } i' \text{ satisfying } A_{ii'} = 1, i' \in C_j\}$  represent the set of units in the  $j$ -th cluster whose neighbors also belong to the  $j$ -th cluster.

If all units in a cluster are assigned the same treatment, NTRA dictates that all units in  $\text{In}_j$  are informative with probability one. If the units are connected with fewer numbers of different clusters, their probability of being informative will also increase. However, it may be inevitable to rely on randomization to determine whether the units in  $\text{C}_j \cap \text{In}_j^c$  are informative.

### 3. Cluster-Adaptive Network A/B Testing

#### 3.1 Cluster-Adaptive Randomization

Let  $\mathbf{X}_{i,\text{IN}}$  represent the  $p$ -individual-level covariates of the  $i$ -th unit and  $\mathbf{X}_{j,\text{CL}}$  denote the  $q$ -cluster-level covariates of the  $j$ -th cluster. In the randomization step, we can use the information of the baseline covariates to enhance the performance of the randomization procedure. We designate  $\boldsymbol{\xi}_j$  as the covariates used in randomization. For instance, to incorporate both the cluster-level and individual-level covariates in randomization, we might consider  $\boldsymbol{\xi}_j = (\mathbf{X}_{j,\text{CL}}^\top, \bar{\mathbf{X}}_{j,\text{IN}}^\top)$ , where  $\bar{\mathbf{X}}_{j,\text{IN}} = c_j^{-1} \sum_{i \in \text{C}_j} \mathbf{X}_{i,\text{IN}}$  represents the within cluster average of the individual-level covariates. If only the cluster-level covariates are considered, it suffices to set  $\boldsymbol{\xi}_j = \mathbf{X}_{j,\text{CL}}$ .

An important component of cluster-adaptive randomization (CLAR) is the imbalance measure, reflecting the differences in covariates between the two treatment arms. Let  $Z_j$  denote the treatment assignment for the  $j$ -th cluster, that is, if  $Z_j = t$  then  $T_i = t$  for all  $i \in \text{C}_j$  and  $t \in \{0, 1\}$ . In addition, let  $m_1 = \sum_{j=1}^m Z_j$  and  $m_2 = m - m_1$  represent the numbers of clusters assigned to treatments 1 and 2, respectively. We denote the sample averages of the baseline covariates for treatment 1, treatment 2, and all clusters as  $\bar{\boldsymbol{\xi}}_{1,m} = m_1^{-1} \sum_{j=1}^m Z_j \boldsymbol{\xi}_j$ ,  $\bar{\boldsymbol{\xi}}_{2,m} = m_2^{-1} \sum_{j=1}^m (1 - Z_j) \boldsymbol{\xi}_j$ , and  $\bar{\boldsymbol{\xi}}_m = m^{-1} \sum_{i=1}^m \boldsymbol{\xi}_j$ , respectively. We adopt the Mahalanobis distance as the imbalance measure, which can be calculated as follows:

$$\begin{aligned} M_{2j} &= (\bar{\boldsymbol{\xi}}_{1,2j} - \bar{\boldsymbol{\xi}}_{0,2j})^\top \text{cov} [\bar{\boldsymbol{\xi}}_{1,2j} - \bar{\boldsymbol{\xi}}_{0,2j}]^{-1} (\bar{\boldsymbol{\xi}}_{1,2j} - \bar{\boldsymbol{\xi}}_{0,2j}) \\ &\propto \frac{j}{2} \cdot (\bar{\boldsymbol{\xi}}_{1,2j} - \bar{\boldsymbol{\xi}}_{0,2j})^\top S_m^{-1} (\bar{\boldsymbol{\xi}}_{1,2j} - \bar{\boldsymbol{\xi}}_{0,2j}), \end{aligned} \quad (1)$$

where  $S_m = (m - 1)^{-1} \sum_{j=1}^m (\boldsymbol{\xi}_j - \bar{\boldsymbol{\xi}}_m)(\boldsymbol{\xi}_j - \bar{\boldsymbol{\xi}}_m)^\top$  represents the sample covariance matrix calculated for the  $m$  clusters. Using the Mahalanobis distance offers various advantages: First, it is an affine invariant measure that standardizes and encapsulates the imbalance for each covariate. Consequently, a lower value of the imbalance measure suggests balanced covariates across the two treatment arms. Moreover, minimizing the Mahalanobis distance increases the comparability of the two treatment arms, thus enhancing the accuracy of the estimation of ATE (Ma et al., 2024; Morgan and Rubin, 2012; Qin et al., 2024).

The CLAR is described in Algorithm 1. This procedure sequentially assigns a pair of clusters, ensuring that only one of the two clusters is allocated to treatment 1. This pairwise assignment results in two possible assigned outcomes and, consequently, two corresponding imbalance score values:  $M_{2j}^{(1)}$  and  $M_{2j}^{(2)}$ . The procedure assigns a higher probability to the assignment that leads to the smaller imbalance value, thereby maintaining the scale of the imbalance measure. To maintain randomness and mitigate imbalance, it is generally recommended to opt for a sufficiently large value of  $\rho$  but avoid values that are very close to 1, for example, within the range of  $[0.7, 0.9]$ .

---

**Algorithm 1** Cluster-Adaptive Randomization (CLAR)

---

```

1: Input: baseline covariates  $\{\boldsymbol{\xi}_j\}_{j=1}^m$ ; probability of the biased coin  $1/2 < \rho < 1$ ;
2: Compute  $\mathbf{S}_m$  based on  $\{\boldsymbol{\xi}_j\}_{j=1}^m$ ;
3: Assign  $Z_1 \sim \text{Bernoulli}(1/2)$  and set  $Z_2 = 1 - Z_1$ ;
4: for  $j = 2$  to  $\lceil m/2 \rceil$  do
5:   if  $2j \leq m$  then
6:     Let  $M_{2j}^{(1)}$  and  $M_{2j}^{(2)}$  be the pseudo imbalance scores computed by  $\{\boldsymbol{\xi}_l\}_{l=1}^{2j}$ ;
7:     Compute  $M_{2j}^{(1)}$  from (1) by assuming  $(Z_{2j-1}, Z_{2j}) = (0, 1)$ ;
8:     Compute  $M_{2j}^{(2)}$  from (1) by assuming  $(Z_{2j-1}, Z_{2j}) = (1, 0)$ ;
9:     if  $M_{2j}^{(1)} = M_{2j}^{(2)}$  then
10:      Assign  $Z_{2j-1} \sim \text{Bernoulli}(1/2)$  and set  $Z_{2j} = 1 - Z_{2j-1}$ ;
11:     else
12:       if  $M_{2j}^{(1)} < M_{2j}^{(2)}$  then
13:        Assign  $Z_{2j-1} \sim \text{Bernoulli}(1 - \rho)$  and set  $Z_{2j} = 1 - Z_{2j-1}$ ;
14:       else
15:        Assign  $Z_{2j-1} \sim \text{Bernoulli}(\rho)$  and set  $Z_{2j} = 1 - Z_{2j-1}$ ;
16:     else
17:      Assign  $Z_{2j-1} \sim \text{Bernoulli}(1/2)$ .

```

---

It is worth noting that CLAR is a sequential randomization procedure, which may be potentially useful for real-world A/B testing problems. The random component of CLAR ensures a valid comparison of the two treatment arms (Imbens and Rubin, 2015; Rosenberger and Lachin, 2015; Rosenberger et al., 2019), which prohibits the deliberate assignment of the treatment by the experimenter. Second, the sequential nature of this procedure enables the integration of CLAR with the ramping process that monitors the validity of the experiment (Kohavi et al., 2020; Xu et al., 2015). This may also be useful for extending the proposed cluster-adaptive network A/B testing procedure for sequential monitoring purposes (Lan and DeMets, 1989; Zhu and Hu, 2019).

**Remark 3.1** *As demonstrated in Section 4, using the Mahalanobis distance as the imbalance measure, stems from the linear relationship between  $Y_i$  and  $\boldsymbol{\xi}_j$ . If the relationship between  $Y_i$  and  $\mathbf{X}_{j,\text{CL}}$  is known and is not linear, we have room for a further improvement by adapting the Mahalanobis distance. For instance, if  $\mathbb{E}[Y_i(\mathbf{0})|\mathbf{X}_{j,\text{CL}}] = \mu_0 + \mathbf{X}_{j,1,\text{CL}}^2\beta_1 + \epsilon_i$  and  $\mathbb{E}[Y_i(\mathbf{1})|\mathbf{X}_{j,\text{CL}}] = \mathbb{E}[Y_i(\mathbf{0})|\mathbf{X}_{j,\text{CL}}] + \mu_1 - \mu_0$  for  $i \in C_j$ , then we can modify (1) by using the difference-in-quadratic-covariate-means instead of the difference-in-covariate-means.*

### 3.2 Cluster-Adjusted Estimator

When randomization is implemented at the cluster-level, NTRA highlights that whether a unit is informative depends on its position in the network and the observed network structure. For instance, if NTRA holds and the  $i$ -th unit satisfies  $(\{i\} \cup \mathcal{N}_i) \subset C_j$ , then the  $i$ -th unit is informative with probability 1. However, if  $i \in C_j$  and  $(\{i\} \cup \mathcal{N}_i) \not\subset C_j$ , the  $i$ -th unit is informative only if  $Z_{j'} = Z_j$  for all  $j'$  where  $i' \in C_{j'} \cap \mathcal{N}_i$ ; in other words, when all neighbors of the  $i$ -th unit are assigned to the same treatment as the  $i$ -th unit. The inherent

cluster structure may result in denser connections within clusters, leading to a majority of units falling into the first category (being informative with probability 1), while fewer units fall into the second category (being informative with probability smaller than 1), thereby ensuring an adequate number of informative units. To maintain clarity, we refer to the first type of nodes as the *inner nodes* and the second type as the *outer nodes*.

The Horvitz-Thompson estimator is commonly used to achieve an unbiased estimate for  $\tau(\mathbf{1}, \mathbf{0})$ . It is important to note that the weights used by the Horvitz-Thompson estimator are partially determined by the network positioning of the nodes. Due to their lower likelihood of being informative, the outer nodes tend to carry larger weights compared to the inner nodes. However, practical scenarios might involve a smaller proportion of outer nodes, leading to a large variance of the Horvitz-Thompson estimate, affected by the higher weight values of the outer nodes. Therefore, it would be advantageous to develop an estimator that is less influenced by network exposure and is simpler to calculate.

Let  $n_j$  denote the number of units in  $\text{Inf}_j$  and  $\mathcal{M} = \{j : n_j > 0\}$  denote the set of clusters with a positive number of informative units. Furthermore, denote the sample average of the observed outcomes in  $\text{Inf}_j$  by  $\hat{Y}_j = n_j^{-1} \sum_{i \in \text{Inf}_j} Y_i$ . We propose the cluster-adjusted estimator (CAE) as follows:

$$\hat{\tau}_{\text{CAE}} = \frac{1}{m_1^*} \sum_{j \in \mathcal{M}} Z_j \hat{Y}_j - \frac{1}{m_0^*} \sum_{j \in \mathcal{M}} (1 - Z_j) \hat{Y}_j,$$

where  $m_1^* = \sum_{j \in \mathcal{M}} Z_j$  and  $m_0^* = \sum_{j \in \mathcal{M}} (1 - Z_j)$ . The proposed CAE is a difference in two-steps means estimator that averages the outcomes on both the individual-level and the cluster-level. The two steps averages can effectively reduce both bias and variance, especially when units within the same cluster share common characteristics.

## 4. Theoretical Properties of Cluster-Adaptive Network A/B Testing

In this section, we investigate the theoretical properties of the proposed cluster-adaptive network A/B testing procedure. We also demonstrate the efficiency gain from improving the covariate balancing and the potential impact of the network structure on the subsequent estimation.

### 4.1 The Balance Property of Cluster-Adaptive Randomization

To derive the theoretical properties of CLAR, we introduce the following assumption.

**Assumption 4.1** 1. Given  $\mathcal{A}$  and  $\tilde{\mathcal{C}}$ , there exists  $\lambda_{\text{IN}}, \tilde{\lambda}_{\text{IN}} \geq 0$  such that  $m^{-1} \sum_{j=1}^m c_j^{-1} = \lambda_{\text{IN}} + o(1)$ , and  $m^{-1} \sum_{j=1}^m n_j^{-1} = \tilde{\lambda}_{\text{IN}} + o(1)$ .

2. Given  $\mathcal{A}$  and  $\tilde{\mathcal{C}}$ ,  $\{\mathbf{X}_{j,\text{CL}}\}_{j=1}^m$  and  $\{\mathbf{X}_{i,\text{IN}}\}_{i=1}^n$  are i.i.d. copies of  $\mathbf{X}_{\text{CL}}$  and  $\mathbf{X}_{\text{IN}}$  with means  $\boldsymbol{\mu}_{\text{CL}}$  and  $\boldsymbol{\mu}_{\text{IN}}$  and positive definite variance-covariance matrices  $\Sigma_{\text{CL}}$  and  $\Sigma_{\text{IN}}$ , respectively. In addition,  $\mathbf{X}_{\text{CL}}$  and  $\mathbf{X}_{\text{IN}}$  are independent and there exists a well defined random vector  $\boldsymbol{\xi}$ , such that  $\boldsymbol{\xi}_j = (\mathbf{X}_{j,\text{CL}}^\top, \tilde{\mathbf{X}}_{j,\text{IN}}^\top)^\top$  is i.i.d.  $\boldsymbol{\xi}$  with mean  $\boldsymbol{\mu}_{\boldsymbol{\xi}} = (\boldsymbol{\mu}_{\text{CL}}^\top, \boldsymbol{\mu}_{\text{IN}}^\top)^\top$  and variance-covariance matrix

$$\Sigma_{\boldsymbol{\xi}\boldsymbol{\xi}} = \begin{pmatrix} \Sigma_{\text{CL}} & \mathbf{0} \\ \mathbf{0} & \lambda_{\text{IN}} \Sigma_{\text{IN}} \end{pmatrix}.$$

3. Let  $\boldsymbol{\eta} = \Sigma_{\boldsymbol{\xi}\boldsymbol{\xi}}^{-1/2}(\boldsymbol{\xi} - \boldsymbol{\mu}_{\boldsymbol{\xi}})$  satisfying  $\mathbb{E}[\|\boldsymbol{\eta}\|^a] < \infty$  for a given  $a > 2$ , where  $\Sigma_{\boldsymbol{\xi}\boldsymbol{\xi}}^{-1/2}$  is the Cholesky square root of  $\Sigma_{\boldsymbol{\xi}\boldsymbol{\xi}}^{-1}$ .
4. Let  $\boldsymbol{\eta}'$  be a i.i.d. copy of  $\boldsymbol{\eta}$ . Denote  $\Gamma_{\boldsymbol{\eta}}(y)$  as the joint distribution function of  $(\boldsymbol{\eta}, \boldsymbol{\eta}')$  and suppose there exists  $k_c \in \mathbb{Z}^+$  and  $0 \leq c_\gamma \leq 1$  such that

$$\Gamma_{\boldsymbol{\eta}}^{k_c^*}(A) \geq c_\gamma \int_A \gamma^{k_c^*}(y) dy \quad \text{for any Borel set } A,$$

where  $\Gamma_{\boldsymbol{\eta}}^{k^*}(\cdot)$  is the  $k$ -th convolution of  $\Gamma_{\boldsymbol{\eta}}(\cdot)$ ,  $\gamma(\cdot)$  is a density function with  $\inf_{y \in O} \gamma(y) > 0$  for an open set  $O$ , and  $\gamma^{k^*}(\cdot)$  is the  $k$ -th convolution of  $\gamma(\cdot)$ .

Assumption 4.1 requires that the cluster covariates  $\{\boldsymbol{\xi}_j\}_{j=1}^m$  are independent and identically distributed, which is in a similar manner to the assumptions made for random geometric graphs (Penrose, 2003). The moment condition, i.e., Assumption 4.1.3, serves the purpose to facilitate the study of the convergence rate of  $M_m$ . Assumption 4.1.4 is further introduced to ensure the Markov chain induced by  $M_m$  is positive Harris recurrent (Ma et al., 2024; Meyn and Tweedie, 2013). This particular property aids in deriving the asymptotic normality of CAE under CLAR. It is important to note that if  $\{\boldsymbol{\xi}_j\}_{j=1}^m$  is dependent but stationary, Assumption 4.1.2 may be adapted as the assumption used in Kojevnikov et al. (2021). In addition, Assumption 4.1.4 may be extended for the convolution of the joint density kernel of the stationary process. We defer the elaboration of these extended theoretical results as potential avenues for future research.

**Theorem 4.1** *Suppose Assumption 4.1 holds. If  $\{\boldsymbol{\xi}_j\}_{j=1}^m$  are used in CLAR, then  $M_m = O_p(m^{-1})$ .*

The proof of Theorem 4.1 utilizes the drift condition (Meyn and Tweedie, 2013) for the Markov chain. See Appendix A.1 for technical details. Theorem 4.1 justifies the usage of CLAR for balancing covariates, since it establishes the order of  $M_m$ . In particular, when using complete randomization at the cluster-level,  $M_m$  follows a  $\chi_{p+q}^2$  distribution (Morgan and Rubin, 2012; Qin et al., 2024). On the contrary, under CLAR,  $M_m$  converges to zero in probability at the rate of  $m^{-1}$ . Consequently, CLAR exhibits superior performance in contrast to complete randomization, ensuring a simultaneous balance for both cluster-level and individual-level covariates when  $\boldsymbol{\xi}_j = (\mathbf{X}_{j,\text{CL}}^\top, \mathbf{X}_{j,\text{IN}}^\top)$ .

## 4.2 Theoretical Properties of Cluster-Adjusted Estimator

For  $\mathbf{t} \in \{\mathbf{1}, \mathbf{0}\}$ , let  $\hat{\mu}_j(\mathbf{t}) = n_j^{-1} \sum_{i \in \text{Inf}_j} \mathbb{E}[\hat{Y}_j(\mathbf{t}) | \mathbf{X}_j, \tilde{\mathbf{C}}, \mathcal{A}]$ ,  $\mathbf{X}_j = (\mathbf{X}_{j,\text{CL}}, \mathbf{X}_{j,\text{IN}})$ , and  $\hat{\mathbf{X}}_{j,\text{IN}} = n_j^{-1} \sum_{i \in \text{Inf}_j} \mathbf{X}_{i,\text{IN}}$ . We introduce the following decomposition for the cluster-level potential outcomes:

$$\hat{Y}_j(\mathbf{t}) - \hat{\mu}_j(\mathbf{t}) = \hat{f}_j(\mathbf{t}, \mathbf{X}_j) + \hat{\epsilon}_j(\mathbf{t}), \quad (2)$$

where  $\hat{f}_j(\mathbf{t}, \mathbf{X}_j) = n_j^{-1} \sum_{i \in \text{Inf}_j} \{\mathbb{E}[Y_i(\mathbf{t}) | \mathbf{X}_j, \tilde{\mathbf{C}}, \mathcal{A}] - \mathbb{E}[Y_i(\mathbf{t})]\}$  represents the cluster-level conditional mean,  $\hat{\epsilon}_j(\mathbf{t}) = n_j^{-1} \sum_{i \in \text{Inf}_j} \{Y_i(\mathbf{t}) - \mathbb{E}[\hat{Y}_j(\mathbf{t}) | \mathbf{X}_j, \tilde{\mathbf{C}}, \mathcal{A}]\}$  corresponds to the cluster-level average of the errors.

Furthermore, consider the  $i$ -th unit being an outer nodes, whether it is informative depends on its connected clusters' treatment assignment. This form of dependency makes the derivation of the theoretical properties of CAE notably intricate. However, Assumption 2.2 (NTRA) indicates that

$$Y_i \mathbb{I}\{i \in \text{Inf}_j\} = [Z_j Y_i(\mathbf{1}) + (1 - Z_j) Y_i(\mathbf{0})] \mathbb{I}\{i \in \text{Inf}_j\},$$

and  $\hat{Y}_j = n_j^{-1} \sum_i^n Y_i \mathbb{I}\{i \in \text{Inf}_j\} = Z_j \hat{Y}_j(\mathbf{1}) + (1 - Z_j) \hat{Y}_j(\mathbf{0})$ . As a result, conditioning on  $\{\mathbb{I}\{i \in \text{Inf}_j\}, i \in C_j\}$ , the values of  $\{\hat{Y}_j(\mathbf{1}), \hat{Y}_j(\mathbf{0})\}$  do not rely on  $Z_{j'}$  for  $j' \neq j$ . This insight enables us to work with the filtrations  $\tilde{\mathcal{F}}_0 = \mathcal{A} \otimes \tilde{\mathcal{C}} \otimes \mathcal{I}$  and  $\tilde{\mathcal{F}}_k = \mathcal{A} \otimes \tilde{\mathcal{C}} \otimes \mathcal{I} \otimes \Xi_{2k} \otimes \mathcal{Z}_{2k} \otimes \mathcal{Y}_{2k}$ , where  $\mathcal{I} = \sigma(\text{Inf}_j, 1 \leq j \leq m)$  encapsulates the information of the informative units by hypothetically assuming that such information was available before randomization.

**Assumption 4.2** *Given  $\mathcal{A}$  and  $\tilde{\mathcal{C}}$ ,  $n_j = |\text{Inf}_j| > 0$  for all  $1 \leq j \leq m$ .*

We introduce Assumption 4.2 to simplify the theoretical derivations. As shown in Section 5, the proposed cluster-adaptive network A/B testing still offers advanced performance when this assumption does not hold.

**Assumption 4.3** *Under Assumptions 4.1, the following hold.*

1. (a)  $\mathbb{E}[Y_i(\mathbf{1}) | \tilde{\mathcal{F}}_0] = \mathbb{E}[Y_i(\mathbf{1})]$  and  $\mathbb{E}[Y_i(\mathbf{0}) | \tilde{\mathcal{F}}_0] = \mathbb{E}[Y_i(\mathbf{0})]$  for  $1 \leq i \leq n$ ; and (b)  $\bar{\mu}_j(\mathbf{1}) = c_j^{-1} \sum_{i \in C_j} \mathbb{E}[Y_i(\mathbf{1})] = \mu(\mathbf{1})$  and  $\bar{\mu}_j(\mathbf{0}) = c_j^{-1} \sum_{i \in C_j} \mathbb{E}[Y_i(\mathbf{0})] = \mu(\mathbf{0})$ , for  $1 \leq j \leq m$ .
2. Let  $\mathbf{Inf}_j$  denote the set of all possible values of  $\text{Inf}_j$ . Denote the  $i$ -th element of  $\mathbf{Inf}_j$  by  $\text{Inf}_j^{(i)}$ , and the associated sample averages of the mean outcomes in  $\text{Inf}_j$  and  $\text{Inf}_j^c \cap C_j$  by  $\hat{\mu}_{j, \text{Inf}_j^{(i)}}(\mathbf{1})$ ,  $\hat{\mu}_{j, \text{Inf}_j^{(i)}}^c(\mathbf{1})$ ,  $\hat{\mu}_{j, \text{Inf}_j^{(i)}}(\mathbf{0})$ ,  $\hat{\mu}_{j, \text{Inf}_j^{(i)}}^c(\mathbf{0})$ , respectively. Then

$$\begin{aligned} \Psi_{\text{Diff, Inf}} &= \max_j \left\{ c_j^{-1} (c_j - n_j) \max_{\text{Inf}_j^{(i)} \in \mathbf{Inf}_j, \mathbf{t} \in \{\mathbf{1}, \mathbf{0}\}} \left[ \left| \hat{\mu}_{j, \text{Inf}_j^{(i)}}(\mathbf{t}) - \hat{\mu}_{j, \text{Inf}_j^{(i)}}^c(\mathbf{t}) \right| \right] \right\} \\ &= o \left( \max \left\{ m^{-1/2}, m^{-(2-\lambda_2)/2} \right\} \right). \end{aligned}$$

3. Given  $\mathcal{A}$  and  $\tilde{\mathcal{C}}$ ,  $\{\hat{f}_j(\mathbf{t}, \mathbf{X}_j)\}_{j=1}^m$  are independent and satisfy  $\mathbb{E}[\hat{f}_j(\mathbf{t}, \mathbf{X}_j)^4] < \infty$ . There exist  $\sigma_{\hat{f}}^2, \tilde{\sigma}_{\hat{f}}^2 > 0$  such that

$$\begin{aligned} m^{-1} \sum_{j=1}^m \mathbb{V} \left[ \hat{f}_j(\mathbf{1}, \mathbf{X}_j) - \hat{f}_j(\mathbf{0}, \mathbf{X}_j) \middle| \tilde{\mathcal{F}}_0 \right] &= \sigma_{\hat{f}}^2 + o_p(1), \\ \text{and } m^{-1} \sum_{j=1}^m \mathbb{V} \left[ \hat{f}_j(\mathbf{1}, \mathbf{X}_j) + \hat{f}_j(\mathbf{0}, \mathbf{X}_j) \middle| \tilde{\mathcal{F}}_0 \right] &= \tilde{\sigma}_{\hat{f}}^2 + o_p(1). \end{aligned}$$

4.  $\epsilon_{i_1}(\mathbf{t})$  and  $\epsilon_{i_2}(\mathbf{t})$  are independent conditional on  $\mathbf{A}$  and the event  $\mathbf{A}_{i_1 i_2} = \max \mathbf{A}_{i_1 j} \mathbf{A}_{j i_2} = 0$  for  $i_1 \neq i_2$  and  $1 \leq i_1, i_2 \leq n$ .

5. There exists  $\lambda_1 > 0$  such that

$$\sum_{j=1}^m \mathbb{E} \left[ |\hat{\epsilon}_j(\mathbf{1})|^3 + |\hat{\epsilon}_j(\mathbf{0})|^3 \mid \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m \right] \asymp m^{\lambda_1}$$

and

$$\sum_{j=1}^m \mathbb{E} \left[ \hat{\epsilon}_j^4(\mathbf{1}) + \hat{\epsilon}_j^4(\mathbf{0}) \mid \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m \right] \asymp m^{\lambda_1}.$$

6. Define

$$\sigma_{m,\text{Ind}}^2 = \sum_{j=1}^m \left\{ Z_j \mathbb{V}[\hat{\epsilon}_j(\mathbf{1}) \mid \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m] + (1 - Z_j) \mathbb{V}[\hat{\epsilon}_j(\mathbf{0}) \mid \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m] \right\},$$

$$\sigma_{m,\text{Bet}}^2 = \sum_{j_1 \neq j_2} \left\{ Z_{j_1} Z_{j_2} \text{Cov} \left[ \hat{\epsilon}_j(\mathbf{1}), \hat{\epsilon}_j(\mathbf{0}) \mid \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m \right] \right. \\ \left. + (1 - Z_{j_1})(1 - Z_{j_2}) \text{Cov} \left[ \hat{\epsilon}_j(\mathbf{0}), \hat{\epsilon}_j(\mathbf{0}) \mid \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m \right] \right\},$$

and  $\sigma_{m,\epsilon}^2 = \sigma_{m,\text{Ind}}^2 + \sigma_{m,\text{Bet}}^2$ . Then, there exist  $\sigma_\epsilon^2 > 0$  and  $2/3\lambda_1 < \lambda_2 < 2$  such that  $\mathbb{E} \left[ (m^{-\lambda_2} \sigma_{m,\epsilon}^2 - \sigma_\epsilon^2)^2 \mid \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m \right] = o(1)$ .

7. Let  $N(j)$  denote the index set of the clusters that are connected to  $C_j$  and assigned with the same treatment as  $C_j$ . Given  $\tilde{\mathcal{F}}_0$ , there exists  $\lambda_4 > 0$ , such that  $\Psi_{\text{max,CL}} = \max_j \#\{N(j)\} = O(m^{\lambda_3})$ , where  $\lambda_3 = \{[4^{-1}(3\lambda_2 - 2\lambda_1)] \wedge [3^{-1}(2\lambda_2 - \lambda_1)]\} - \lambda_4$ .

Assumption 4.3.1 ensures the consistency of the CAE. First, Assumption 4.3.1.(a) requires that once all units are assigned with the same treatment, the outcome of one unit will not be affected by the network and thus ensures the consistency of CAE. Furthermore, the ATE can also be written as  $\tau(\mathbf{1}, \mathbf{0}) = n^{-1} \sum_{j=1}^m c_j [\bar{\mu}_j(\mathbf{1}) - \bar{\mu}_j(\mathbf{0})]$ , Assumption 4.3.1.(b) is introduced to ensure the identification of the mean effects  $n^{-1} \sum_{i=1}^n \mathbb{E}[Y_i(\mathbf{1})]$  and  $n^{-1} \sum_{i=1}^n \mathbb{E}[Y_i(\mathbf{0})]$ , which is analogous to the usual random effect model that treats the between-cluster heterogeneity as a zero-mean random effect.

Assumption 4.3.2 is introduced to restrict the impact of network interference on the estimation of CAE. The size of the bias depends both on the within cluster heterogeneity and  $c_j - n_j$ . This indicates that if the network has within cluster homogeneous structures, such properties may improve the performance of the CAE in reducing the bias caused by network interference.

Assumptions 4.3.3–4.3.7 are similar to the conditions considered in Leung (2020), except that our focuses on demonstrating the impact of  $\{\mathbf{X}_j\}_{j=1}^m$ . Because Assumption 4.1 assumes that  $\{\mathbf{X}_j\}_{j=1}^m$  are independent and  $\hat{f}_j(\mathbf{t}, \mathbf{X}_j)$  is a function of  $\mathbf{X}_j$ , we also assume that  $\{\hat{f}_j(\mathbf{t}, \mathbf{X}_j)\}_{j=1}^m$  are independent, but allowing  $\hat{f}_j(\mathbf{t}, \mathbf{X}_j)$  to be not identically distributed. Assumption 4.3.3 further characterizes the behavior of  $\{\hat{f}_j(\mathbf{1}, \mathbf{X}_j), \hat{f}_j(\mathbf{0}, \mathbf{X}_j)\}_{j=1}^m$ . Finally, Assumption 4.3.5–4.3.7 are introduced to demonstrate the impact of the rate of network dependency on the asymptotic normality of CAE.

**Theorem 4.2** *Suppose Assumptions 2.2, 4.1, 4.2, and 4.3 hold.  $\hat{\tau}_{\text{CAE}}$  is consistent, i.e.,*

$$\begin{aligned} \hat{\tau}_{\text{CAE}} - \tau(\mathbf{1}, \mathbf{0}) &= m^{-1} \left\{ \sum_{j=1}^m (2Z_j - 1) [\hat{f}_j(\mathbf{1}, \mathbf{X}_j) + \hat{f}_j(\mathbf{0}, \mathbf{X}_j)] \right. \\ &\quad \left. + \sum_{j=1}^m [\hat{f}_j(\mathbf{1}, \mathbf{X}_j) - \hat{f}_j(\mathbf{0}, \mathbf{X}_j)] + 2 \sum_{j=1}^m [Z_j \hat{\epsilon}_j(\mathbf{1}) - (1 - Z_j) \hat{\epsilon}_j(\mathbf{0})] \right\} + \Psi_{\text{Diff,Inf}} \\ &= o_p(1). \end{aligned} \quad (3)$$

Furthermore, there exists  $\sigma_{m,\text{CAE}}^2 = m^{-1}(\sigma_{\text{Design}}^2 + \sigma_f^2 + 4m^{\lambda_2-1}\sigma_\epsilon^2)$  with  $\sigma_{\text{Design}}^2 \geq 0$  such that

$$\sigma_{m,\text{CAE}}^{-1} \{ \hat{\tau}_{\text{CAE}} - \tau(\mathbf{1}, \mathbf{0}) \} \xrightarrow{\text{D}} \mathcal{N}(0, 1).$$

If complete randomization is used, then  $\sigma_{\text{Design}}^2 = \tilde{\sigma}_f^2$ .

**Remark 4.1**  $\lambda_2$  describes the impact of the network structure on the asymptotic behavior of the error term in (3). Let  $\text{Ed}_{j_1, j_2}$  denote the number of edges connecting clusters  $\mathbf{C}_{j_1}$  and  $\mathbf{C}_{j_2}$ , and let  $\Psi_{\text{max,IN}} = \max_i \#\mathcal{N}_i$  represent the maximum size of the neighborhood  $\mathcal{N}_i$ . If we assume  $\text{Cov}[\epsilon_i(\mathbf{t}), \epsilon_{i'}(\mathbf{t})] < \infty$ , then

$$\sigma_{m,\text{Ind}}^2 \lesssim O \left( (1 + \Psi_{\text{max,IN}}) \sum_{j=1}^m n_j^{-1} \right) \quad (4)$$

$$\text{and } \sigma_{m,\text{Bet}}^2 \lesssim O \left( \sum_{j_1 \neq j_2} n_{j_1}^{-1} n_{j_2}^{-1} \mathbb{I}\{Z_{j_1} = Z_{j_2}\} \text{Ed}_{j_1, j_2} \right). \quad (5)$$

Here,  $\Psi_{\text{max,IN}}$  captures the level of inter-dependency within the clusters. The number of correlated clusters considered in (5) is bounded by  $m\Psi_{\text{max,CL}}$ . If  $\Psi_{\text{max,IN}}$ ,  $\Psi_{\text{max,CL}}$ , and  $\max_{j_1, j_2}(\text{Ed}_{j_1, j_2})$  are all bounded, then a larger value of  $n_j$  implies smaller values of  $\sigma_{m,\text{Ind}}^2$  and  $\sigma_{m,\text{Bet}}^2$ . This further highlights the efficiency gained through the proposed CAE.

In practice, the rates of  $\Psi_{\text{max,IN}}$ ,  $\Psi_{\text{max,CL}}$ , and  $\max_{j_1, j_2}(\text{Ed}_{j_1, j_2})$  could depend on  $n$  and  $m$ . Assumptions 4.3.4–4.3.6 establish the conditions under which the rate of this dependency does not affect the asymptotic normality of the third term in (3).

$\sigma_{\text{Design}}^2$  defined in Theorem 4.2 reflects the impact of the chosen randomization scheme on the subsequent estimation of ATE with CAE. The difference  $\tilde{\sigma}_f^2 - \sigma_{\text{Design}}^2$  illustrates the efficiency gain resulting from the improved covariate balancing. In the general case where the form of  $\hat{f}_j(\mathbf{t}, \mathbf{X})$  is not explicitly assumed,  $\sigma_{\text{Design}}^2$  may not have a close form. To unveil the impact of CLAR on the theoretical properties of CAE, we study a specific scenario in which the outcome follows a linear-in-means model.

**Corollary 4.1** *Under the conditions of Theorem 4.2, consider the following outcome model:*

$$Y_i(\mathbf{t}) = t_i\mu_1 + (1 - t_i)\mu_0 + d_i^{-1} \left( \alpha_1 \sum_{k \in \mathcal{N}_i} t_k + \alpha_0 \sum_{k \in \mathcal{N}_i} (1 - t_k) \right) + \mathbf{X}_{j,\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}} \mathbb{I}\{i \in C_j\} + \mathbf{X}_{i,\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}} + \epsilon_i,$$

where  $\epsilon_i$  are i.i.d.  $\mathcal{N}(0, \sigma_e^2)$ . Then  $\hat{f}_j(\mathbf{X}_j) = \hat{f}_j(\mathbf{1}, \mathbf{X}_j) = \hat{f}_j(\mathbf{0}, \mathbf{X}_j) = (\mathbf{X}_{j,\text{CL}} - \boldsymbol{\mu}_{\text{CL}})^\top \boldsymbol{\beta}_{\text{CL}} + (\hat{\mathbf{X}}_{j,\text{IN}} - \boldsymbol{\mu}_{\text{IN}})^\top \boldsymbol{\beta}_{\text{IN}}$ ,  $\hat{\epsilon}_j = \hat{\epsilon}_j(\mathbf{1}) = \hat{\epsilon}_j(\mathbf{0}) = n_j^{-1} \sum_{i \in \text{Inf}_j} \epsilon_i$ ,  $\sigma_{\mathbf{f}}^2 = 0$ ,  $\lambda_2 = 1$ ,  $\sigma_\epsilon^2 = \tilde{\lambda}_{\text{IN}} \sigma_e^2$  and the following hold.

1. Under complete randomization,  $\sigma_{\text{Design}}^2 = 4\{\mathbb{V}[\mathbf{X}_{\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}}] + \tilde{\lambda}_{\text{IN}} \mathbb{V}[\mathbf{X}_{\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}}]\}$  and

$$\sigma_{m,\text{CAE}}^2 = 4m^{-1} \left\{ \tilde{\lambda}_{\text{IN}} \sigma_e^2 + \mathbb{V}[\mathbf{X}_{\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}}] + \tilde{\lambda}_{\text{IN}} \mathbb{V}[\mathbf{X}_{\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}}] \right\}.$$

2. Suppose CLAR is implemented with  $\boldsymbol{\xi}_j = \mathbf{X}_{j,\text{CL}}$ , then  $\sigma_{\text{Design}}^2 = 4\tilde{\lambda}_{\text{IN}} \mathbb{V}[\mathbf{X}_{\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}}]$ , and

$$\sigma_{m,\text{CAE}}^2 = 4m^{-1} \left\{ \tilde{\lambda}_{\text{IN}} \sigma_e^2 + \tilde{\lambda}_{\text{IN}} \mathbb{V}[\mathbf{X}_{\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}}] \right\}.$$

3. Suppose CLAR is implemented with  $\boldsymbol{\xi}_j = \bar{\mathbf{X}}_{j,\text{IN}}$ . Let  $\zeta_m = m^{-1/2} \sum_{j=1}^m (2Z_j - 1)(\hat{\mathbf{X}}_{j,\text{IN}} - \bar{\mathbf{X}}_{j,\text{IN}})^\top \boldsymbol{\beta}_{\text{IN}}$  and there exists  $\sigma_{\text{diff}}^2 > 0$ , such that  $\mathbb{V}[\zeta_m] = \sigma_{\text{diff}}^2 + o(1)$ . Then,  $\sigma_{\text{Design}}^2 = 4\{\mathbb{V}[\mathbf{X}_{\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}}] + \sigma_{\text{diff}}^2\}$ , and

$$\sigma_{m,\text{CAE}}^2 = 4m^{-1} \left\{ \tilde{\lambda}_{\text{IN}} \sigma_e^2 + \sigma_{\text{diff}}^2 + \mathbb{V}[\mathbf{X}_{\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}}] \right\}.$$

4. Suppose CLAR is implemented with  $\boldsymbol{\xi}_j = \mathbf{X}_j = (\mathbf{X}_{j,\text{CL}}^\top, \bar{\mathbf{X}}_{j,\text{IN}}^\top)^\top$ , then  $\sigma_{\text{Design}}^2 = 4\sigma_{\text{diff}}^2$ , and

$$\sigma_{m,\text{CAE}}^2 = 4m^{-1} \left\{ \tilde{\lambda}_{\text{IN}} \sigma_e^2 + \sigma_{\text{diff}}^2 \right\},$$

where  $\sigma_{\text{diff}}^2$  is defined in 3.

**Remark 4.2** We give the explicit expressions for  $\sigma_{m,\text{CAE}}^2$  in Corollary 4.1 when  $\hat{f}_j(\mathbf{t}, \mathbf{X}_j)$  is linear in  $\mathbf{X}_j$ . This corollary reveals that  $\sigma_{\text{Design}}^2$  depends on the chosen randomization procedure and the selected covariates in the implementation of CLAR. The reduced values of  $\sigma_{m,\text{CAE}}^2$  when comparing the three randomization schemes underscore the efficiency gains achieved from balancing covariates with CLAR. Moreover, including  $\{\bar{\mathbf{X}}_{j,\text{IN}}\}_{j=1}^m$  might not entirely mitigate the variance of CAE associated with individual-level covariates due to disparities between  $\bar{\mathbf{X}}_{j,\text{IN}}$  and  $\hat{\mathbf{X}}_{j,\text{IN}}$ . The effectiveness of balancing  $\{\bar{\mathbf{X}}_{j,\text{IN}}\}_{j=1}^m$  in enhancing the efficiency of CAE could depend on the value of  $\tilde{\lambda}_{\text{IN}}$ , which is related with the fractions of the informative units  $\{n_j/c_j\}_{j=1}^m$ .

## 5. Numerical Studies

In this section, we evaluate the finite sample properties of our proposed procedure with hypothetical networks. The cluster sizes  $\{c_j\}_{j=1}^m$  are generated from a discrete power-law distribution, with a parameter 4, and the parameter representing the minimum value of a cluster, 12 (Kolaczyk, 2009). The clusters are generated from the small-world model (Watts and Strogatz, 1998) to mimic the cluster structure in real applications. Finally, we generate the edges that connect different clusters from a Rényi random graph with  $r \times n$  edges, where  $r \in \{0.2, 0.4, \dots, 2\}$  represents the pre-specified *reconnecting probability* characterizing the portion of edges that connects different clusters. Therefore, when the value of  $r$  increases, the magnitude of the interference increases, resulting in more difficult situations to evaluate the ATE.

The following model is assumed to generate the outcome:

$$\begin{aligned}
 Y_i &= T_i \mu_1 + (1 - T_i) \mu_0 + \alpha_1 d_i^{-1} \sum_{k \in \mathcal{N}_i} T_k + \alpha_0 d_i^{-1} \sum_{k \in \mathcal{N}_i} (1 - T_k) \\
 &\quad + \sum_{j=1}^m \mathbf{X}_{j,\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}} \mathbb{I}\{i \in C_j\} + \mathbf{X}_{i,\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}} + \epsilon_i^*, \\
 \epsilon_i^* &= \epsilon_i + \sum_{j=1}^m A_{ij} \epsilon_j,
 \end{aligned} \tag{6}$$

where  $\mu_1 = 2$  and  $\mu_0 = 1$  are the direct effects and  $\alpha_1 = 2$  and  $\alpha_0 = 1$  are the spillover effects. Therefore, the ATE is  $\tau(\mathbf{1}, \mathbf{0}) = \mu_1 - \mu_0 + \alpha_1 - \alpha_0 = 2$ . The cluster-level covariates  $\mathbf{X}_{j,\text{CL}} = (X_{j,1,\text{CL}}, X_{j,2,\text{CL}})^\top$  and the individual-level covariates  $\mathbf{X}_{i,\text{IN}} = (X_{i,1,\text{IN}}, X_{i,2,\text{IN}}, X_{i,3,\text{IN}})^\top$  are generated as follows:  $X_{j,1,\text{CL}}$ , the scaled cluster size  $c_j/\mathbb{E}[c_j]$ ;  $X_{j,2,\text{CL}}$ , the density of the  $j$ -th cluster;  $X_{i,1,\text{IN}}$ , the indicator of being an outer node;  $X_{i,2,\text{IN}}$ , the number of edges connecting with nodes in  $C_{j'}$  if  $i \in C_j$  and  $j \neq j'$ ; and  $X_{i,3,\text{IN}}$ , the number of edges connecting nodes in  $C_j$  if  $i \in C_j$ . The associated effects of the cluster-level and individual-level covariates are  $\boldsymbol{\beta}_{\text{CL}} = (1, 0.8)^\top$  and  $\boldsymbol{\beta}_{\text{IN}} = (1, 0.5, 0.5)^\top$ . The random errors  $\epsilon_i$  are i.i.d.  $N(0, 2^2)$  and are independent of  $\{\mathbf{X}_{j,\text{CL}}\}_{j=1}^m$  and  $\{\mathbf{X}_{i,\text{IN}}\}_{i=1}^n$ .

We conduct 100 experiments with 100 networks, each containing  $m$  clusters, for  $m \in \{50, 100, 200\}$ . In the randomization step, we compare the following four design schemes: 1) complete randomization (CR); 2) cluster-adaptive randomization (CLAR) with cluster-level covariates (CL),  $\boldsymbol{\xi}_j = \mathbf{X}_{j,\text{CL}}$  (CLAR-CL); 3) CLAR with individual-level covariates (Ind),  $\boldsymbol{\xi}_j = \bar{\mathbf{X}}_{j,\text{IN}}$  (CLAR-Ind); and 4) CLAR with both cluster and individual-levels covariates (Both),  $\boldsymbol{\xi}_j = (\mathbf{X}_{j,\text{CL}}^\top, \bar{\mathbf{X}}_{j,\text{IN}}^\top)^\top$  (CLAR-Both). In the estimation step, we compare the cluster-adaptive estimator (CAE) with the difference-in-means (DIM) estimator.

### 5.1 Evaluation of Covariates Balance under the Four Randomization Schemes

We first evaluate the performance of the four randomization schemes on balancing covariates. Figure 1 indicates that the inclusion of the covariates in CLAR makes the distribution of the Mahalanobis distance concentrated at zero. The performance in the balance of the covariates improves as  $m$  increases and as more covariates are included in CLAR. Therefore, the use of CLAR generally improves the balance of the covariates in terms of the Mahalanobis distance.

CLUSTER-ADAPTIVE NETWORK A/B TESTING

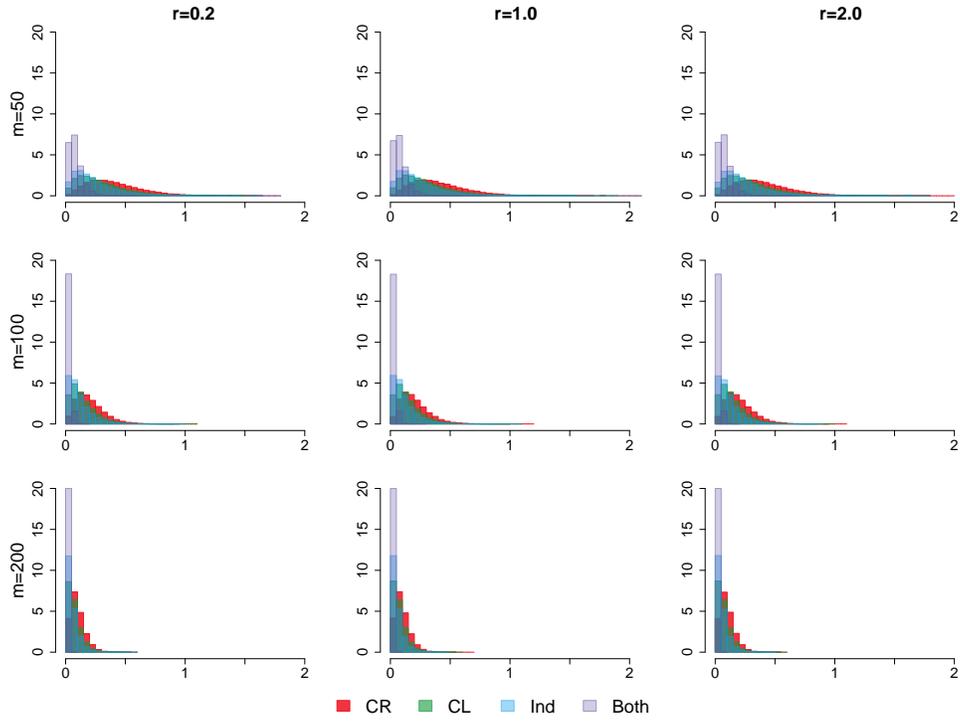


Figure 1: Histograms of the Mahalanobis distance  $M_m$  under different randomization schemes and  $r \in \{0.2, 1.0, 2.0\}$ .

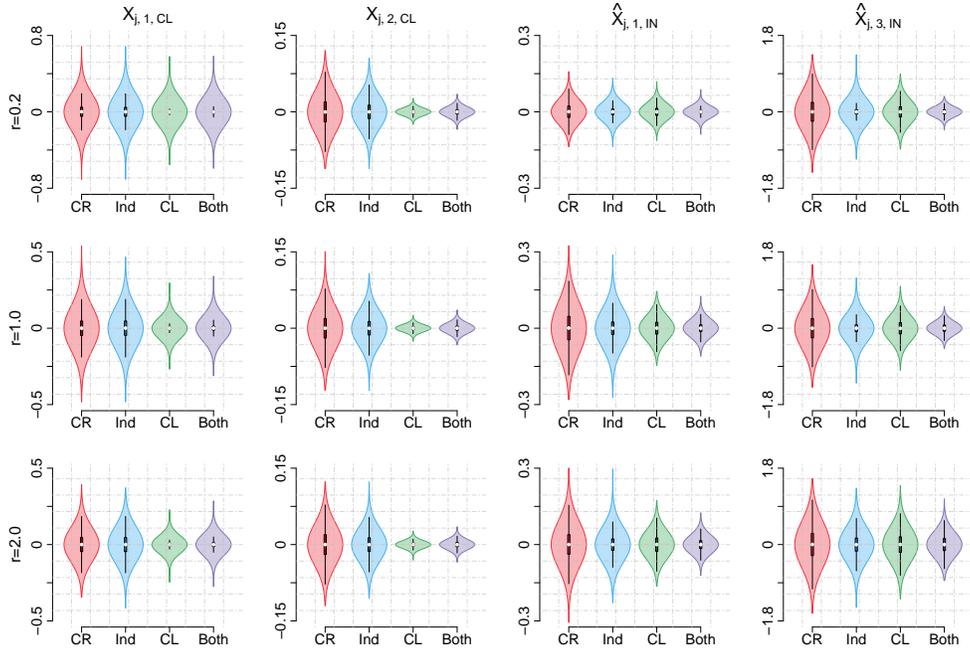


Figure 2: Violin plots of the difference-in-covariate-means for  $X_{j,1,CL}$ ,  $X_{j,2,CL}$ ,  $\hat{X}_{j,1,CL}$ , and  $\hat{X}_{j,3,IN}$  under  $m = 200$ .

The performance of CAE may be affected by the balance of covariates with respect to  $\mathbf{X}_{j,\text{CL}}$  and  $\hat{\mathbf{X}}_{j,\text{IN}}$ . We then assess the finite sample properties of the four randomization schemes on balancing ( $X_{j,1,\text{CL}}, X_{j,2,\text{CL}}, \hat{X}_{j,1,\text{IN}}, \hat{X}_{j,2,\text{IN}}$ ) in Figure 2. The distributions of the difference-in-covariates mean of the cluster-level covariates are more concentrated at zero when these covariates are included in CLAR. On the other hand, the distributions under CR or CLAR-Ind are more spread out. Note that  $\hat{X}_{j,1,\text{IN}}$  and  $\hat{X}_{j,2,\text{IN}}$  cannot be directly balanced by CLAR, because their values still depend on the treatment assignments of other clusters. Figure 2 shows that CLAR-Both and CLAR-CL have better balancing performance for  $\hat{\mathbf{X}}_{j,\text{IN}}$  than CLAR-Ind. It suggests that the inclusion of  $\mathbf{X}_{\text{CL}}$  in CLAR is necessary to achieve the performance for balancing  $\mathbf{X}_{j,\text{CL}}$  and  $\hat{\mathbf{X}}_{j,\text{IN}}$ .

### 5.2 Comparison of Different Network A/B Testing Approaches

This section evaluates the impact of balancing cluster-level and individual-level covariates, and the network interference on the estimation of  $\tau(\mathbf{1}, \mathbf{0})$ . Table 1 compares the performance of the DIM estimator and CAE under the four randomization schemes for selected values of  $r$ . Figures 3 and 4 further evaluate the bias and the standard deviation for different network A/B testing approaches.

Table 1: Evaluation of treatment effect with different network A/B testing approaches.

$m$	Design	$r = 0.2$						$r = 1.8$					
		$\hat{\tau}_{\text{DIM}}$			$\hat{\tau}_{\text{CAE}}$			$\hat{\tau}_{\text{DIM}}$			$\hat{\tau}_{\text{CAE}}$		
		Bias	SD	RMSE									
50	CR	-0.138	1.472	1.478	-0.011	0.952	0.952	-0.764	1.099	1.339	-0.017	1.407	1.407
	CL	-0.121	1.303	1.309	0.004	0.679	0.679	-0.755	0.878	1.158	-0.004	1.018	1.018
	Ind	-0.132	1.337	1.343	-0.013	0.724	0.724	-0.751	0.918	1.186	0.005	1.126	1.126
	Both	-0.128	1.320	1.327	-0.002	0.663	0.663	-0.750	0.871	1.150	0.003	0.992	0.992
100	CR	-0.129	0.806	0.817	-0.005	0.663	0.663	-0.755	0.952	1.215	-0.006	0.983	0.983
	CL	-0.129	0.609	0.623	-0.002	0.477	0.477	-0.752	0.771	1.077	-0.002	0.713	0.713
	Ind	-0.122	0.680	0.690	0.001	0.496	0.496	-0.757	0.842	1.132	-0.006	0.789	0.789
	Both	-0.126	0.625	0.637	0.001	0.440	0.440	-0.751	0.786	1.087	-0.006	0.683	0.683
200	CR	-0.127	1.174	1.181	0.001	0.471	0.471	-0.748	1.177	1.395	0.002	0.698	0.698
	CL	-0.131	1.107	1.115	-0.003	0.343	0.343	-0.742	1.099	1.326	0.004	0.504	0.504
	Ind	-0.125	1.130	1.137	-0.000	0.347	0.347	-0.749	1.135	1.360	-0.001	0.556	0.556
	Both	-0.128	1.121	1.128	-0.003	0.301	0.301	-0.749	1.115	1.343	0.001	0.477	0.477

Figure 3 shows that the proposed CAE is consistent, whereas the DIM estimator is generally biased in the presence of interference. The bias for CAE is close to zero under different values of  $r$ . However, the bias for the DIM estimator increases as  $r$  increases. It is worth noting that the degrees of bias under different randomization schemes are close for the same estimator. This suggests that a proper adjustment in the estimation step is essential to achieve consistency in the presence of network interference.

Figure 4 demonstrates the impact of balancing covariates on the subsequent estimation for  $\tau(\mathbf{1}, \mathbf{0})$ . Although the standard deviation tends to be larger for the network with a higher degree of interference, the curves of the standard deviation of CAE under CLAR are below the curve under CR. This illustrates that the balancing covariates can improve the efficiency of estimating  $\tau(\mathbf{1}, \mathbf{0})$  with CAE.

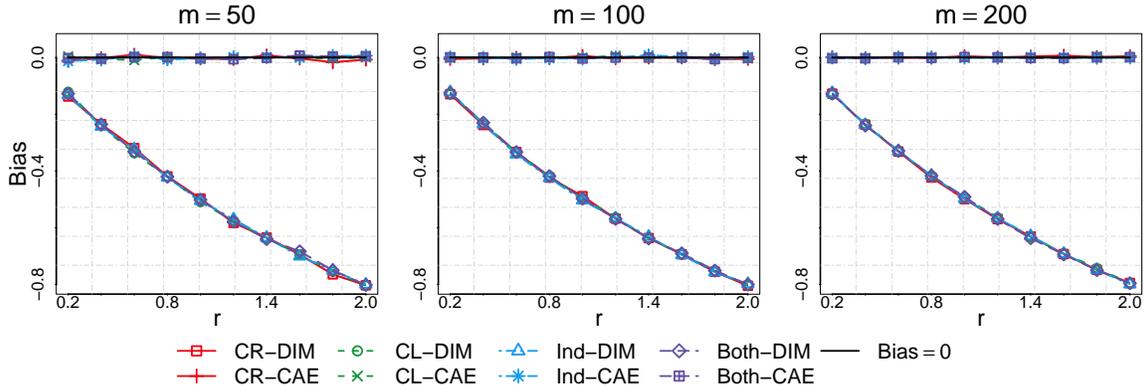


Figure 3: Bias for evaluating the ATE under different network A/B testing approaches.

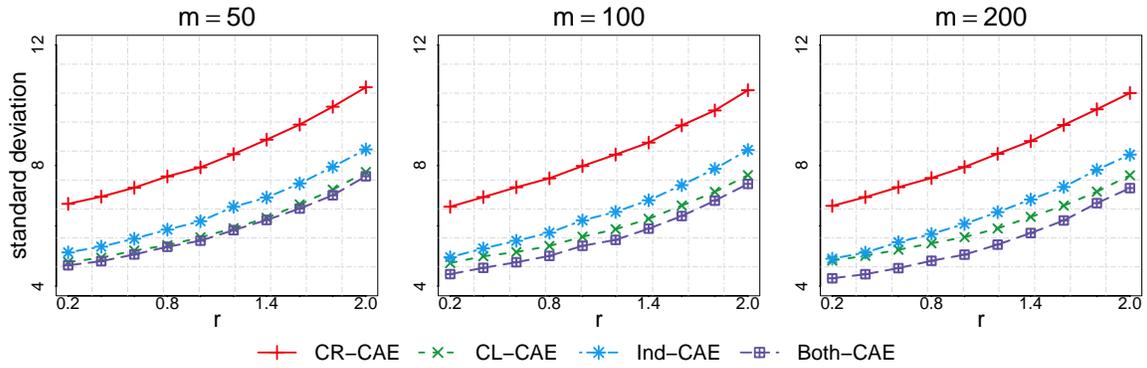


Figure 4: Standard deviations of  $\sqrt{m^*}(\hat{\tau}_{CAE} - \tau(\mathbf{1}, \mathbf{0}))$  under the four randomization schemes.

Figure 5 further evaluates the distribution of CAE under the four randomization schemes. The distribution of CAE under the design with better balance properties tends to be more concentrated and has a smaller variance. Moreover, the CAEs under CLAR-CL and CLAR-Both have similar performance in terms of the standard deviation and RMSE and are better than the CAE under CLAR-Ind and CR. This may indicate that the balance of cluster-level covariates is important for improving the efficiency of CAE.

To demonstrate the robust property of the CAE against the value of  $m^*$ , we report different fractions of clusters and units included in CAE in Table 2. The values in Table 2 provides the evidence that Assumption 4.2 can be relaxed. Table 2 shows that  $m \neq m^*$  holds when  $r \geq 1.0$ , indicating the violation of Assumption 4.2. Furthermore, the fraction of informative units per cluster reduces as  $r$  increases. In such situations, we can still observe that the use of CLAR can reduce the standard deviation of CAE. For instance, as shown in Table 1, when  $m = 200$  and  $r = 1.8$  the standard deviation of CAE under CLAR-CL, CLAR-Ind, and CLAR-Both are 0.504, 0.556 and 0.477, respectively, and are smaller than the value under CR, i.e., 0.698. This further demonstrates the desirable properties of our proposed procedure.

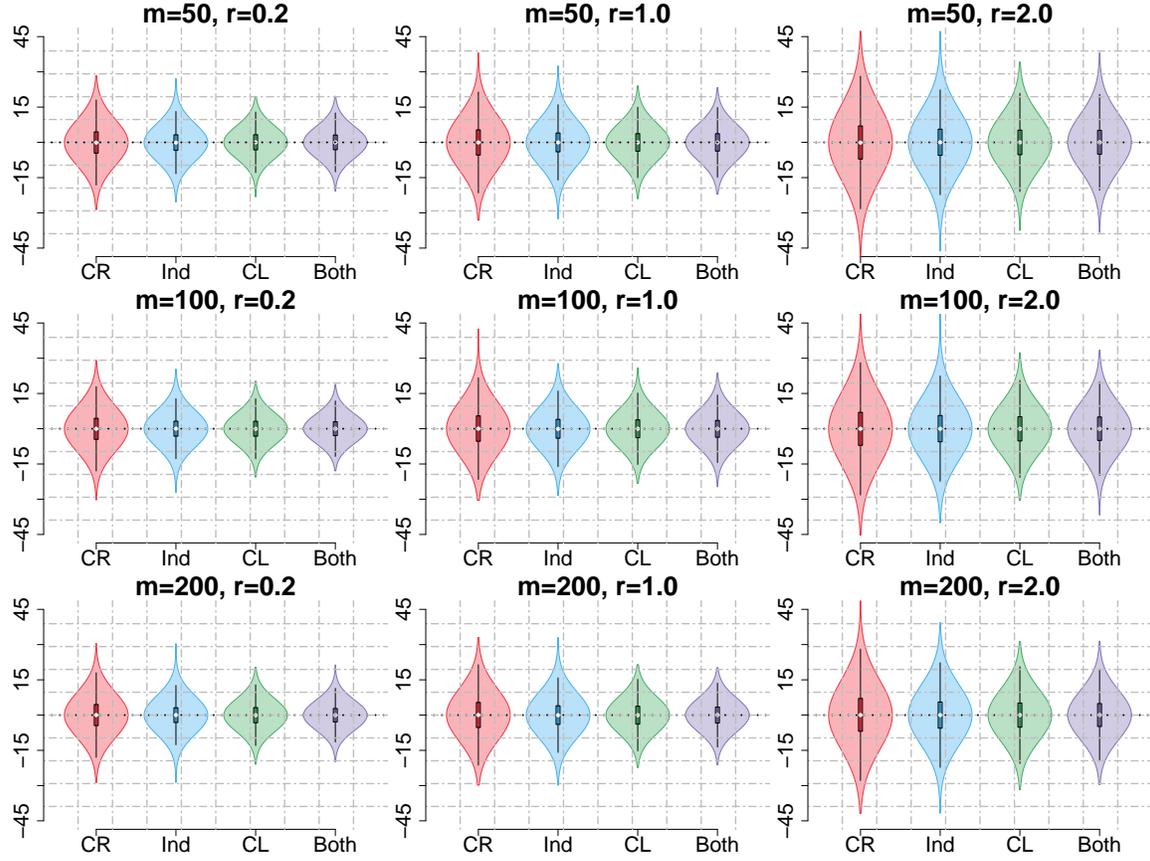

 Figure 5: Distributions of  $\sqrt{m^*}(\hat{\tau}_{\text{CAE}} - \tau(\mathbf{1}, \mathbf{0}))$  under the four randomization schemes.

 Table 2: Average fractions of clusters and samples included in CAE:  $\rho_1 = m^{-1}m^*$ ,  $\rho_2 = n^{-1}\mathbb{E}[\sum_j n_j]$ , and  $\rho_3 = \sum_{j=1}^m c_j^{-1}\mathbb{E}[n_j]$ .

Fraction(%)		$r$				
		0.2	0.6	1.0	1.4	1.8
CR	$\rho_1$	100.00	100.00	99.80	98.13	92.84
	$\rho_2$	82.10	55.41	37.46	25.39	17.19
	$\rho_3$	82.05	55.29	37.36	25.28	17.11
CL	$\rho_1$	100.00	100.00	99.83	98.33	93.13
	$\rho_2$	81.87	54.87	36.77	24.67	16.48
	$\rho_3$	81.86	54.84	36.77	24.64	16.47
Ind	$\rho_1$	100.00	100.00	99.81	98.23	92.92
	$\rho_2$	81.93	55.06	37.02	24.92	16.74
	$\rho_3$	81.88	54.95	36.92	24.82	16.66
Both	$\rho_1$	100.00	100.00	99.83	98.31	93.10
	$\rho_2$	81.86	54.89	36.79	24.69	16.51
	$\rho_3$	81.84	54.84	36.77	24.65	16.49

In Appendix B, we present an example to evaluate the performance of our proposed procedure when the clusters are generated from stochastic block models. These results are consistent with our findings presented in this section.

## 6. Real Data Example

In this section, a real-world social network drawn from a large-scale field experiment is presented to further demonstrate the property of the proposed procedure. The experiment was designed to study an anti-conflict intervention adjusting the social norm of community members in 56 US middle schools in New Jersey (Paluck et al., 2016). Through social connections, some of the students may attract more students’ attention and spread perceptions of conflict as less socially normative. Pre-randomization surveys were conducted to measure social connections in schools and to map the complete social network. This dataset is available at <https://www.icpsr.umich.edu/web/civicleads/studies/37070/versions/V2>. We redesign and analyze this study to detect the impact of the intervention if all students receive the treatment, i.e., the “all versus nothing” treatment effect. Since students in the same grade are more likely to be connected to each other, 146 clusters were formed by stratifying the schools and grades of the students. Figure 6 presents an illustrative graph of a school, where students in four grades are labeled with four colors.

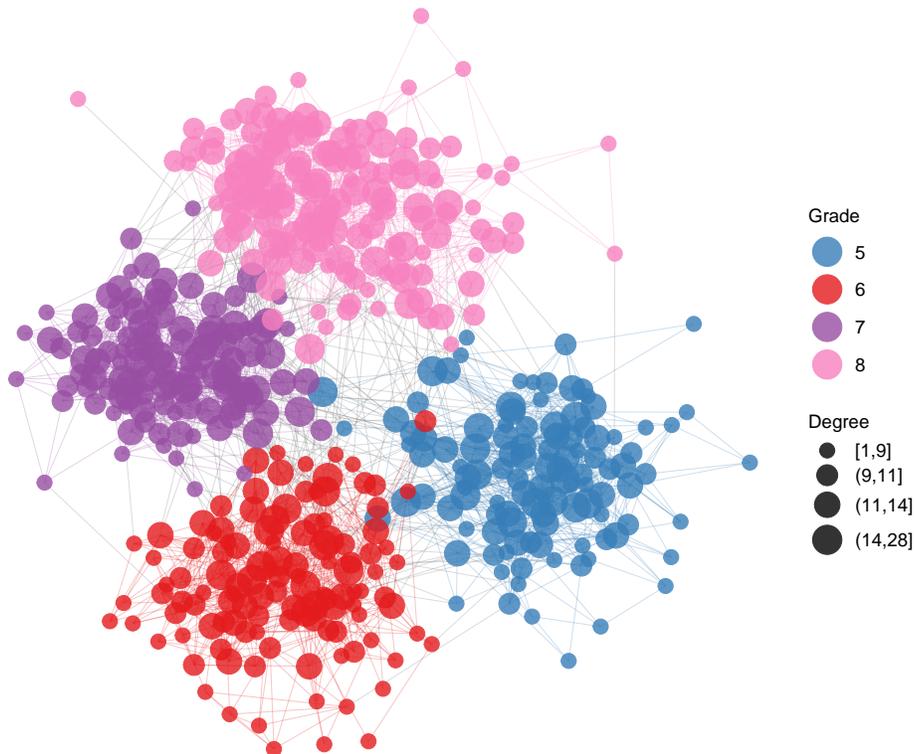


Figure 6: The network graph within one school as a subgraph of the full network dataset. There are four clusters (grades) in this school represented by four colors.

The outcome variable is the attitude toward conflict (ATC) score, which summarizes the questions evaluating the anti-conflict attitude of a student. The formula for calculating ATC is presented in Appendix C. The synthetic data of 146 clusters and 19,929 students are generated according to the following fitted linear model:

$$\begin{aligned} \text{ATC}_i &= 3.05T_i + 2.82(1 - T_i) + 0.51d_i^{-1} \sum_{k \in \mathcal{N}_i} T_k + 0.33d_i^{-1} \sum_{k \in \mathcal{N}_i} (1 - T_k) \\ &\quad - 0.31\text{Grade}_{i,1} - 0.78\text{Grade}_{i,2} - 1.23\text{Grade}_{i,3} \\ &\quad + \sum_{j=1}^m \mathbf{X}_{j,\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}} \mathbb{I}\{i \in C_j\} + \mathbf{X}_{i,\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}} + \epsilon_i, \end{aligned} \quad (7)$$

where  $\mathbf{X}_{i,\text{IN}} = (X_{i,1,\text{IN}}, \dots, X_{i,5,\text{IN}})^\top$ ,  $\mathbf{X}_{j,\text{CL}} = (X_{j,1,\text{CL}}, \dots, X_{j,4,\text{CL}})^\top$ ,  $\boldsymbol{\beta}_{\text{IN}} = (-0.29, -0.41, 0.09, 0.06, -0.06)^\top$ , and  $\boldsymbol{\beta}_{\text{CL}} = (-0.43, 0.21, -0.18, -3.43)^\top$ , and  $\tau(\mathbf{1}, \mathbf{0}) = 0.41$ . Detailed information of the covariates and data are presented in Appendix C.

To perform network A/B testing, we consider the following four approaches in the randomization step: 1. complete randomization (CR); 2. CLAR with cluster-level covariates (CLAR-CL); 3. CLAR with individual-level covariates (CLAR-Ind); and 4. CLAR with both cluster-level and individual-level covariates (CLAR-Both). For estimating ATE, we compare the difference-in-means (DIM) estimator with CAE. All of the simulation studies are based on 10,000 replications.

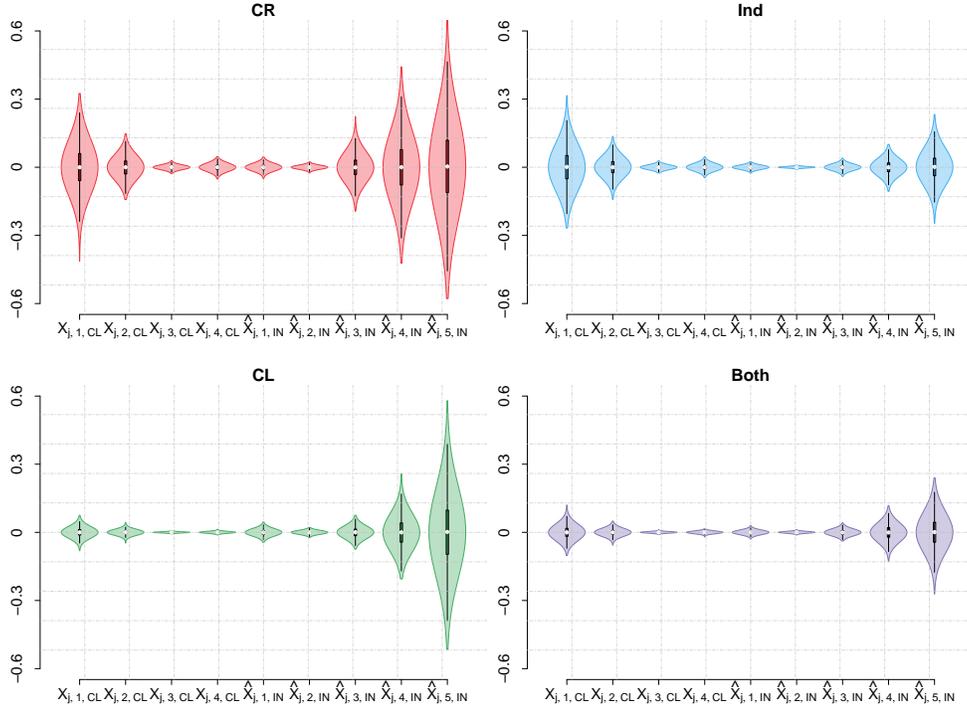


Figure 7: Distribution of difference-in covariate means for  $\mathbf{X}_{j,\text{CL}}$  and  $\hat{\mathbf{X}}_{j,\text{IN}}$  under the four randomization schemes.

Figure 7 evaluates the balance properties of the four randomization schemes. Compared to CR, the use of covariates in CLAR generally improves its performance on covariates balancing. Although CLAR-Ind achieves finer balance for the individual-level covariates, it may not generate sufficient balance for the cluster-level covariates. Similarly, the balance of cluster-level covariates may not imply the adequate balance for the individual-level covariates. On the other hand, CLAR-Both ensures the balance for both of the two levels of covariates simultaneously.

Table 3: Bias, standard deviation (SD) and MSE for different A/B testing approaches.

	$\hat{\tau}_{\text{DIM}} (10^{-2})$				$\hat{\tau}_{\text{CAE}} (10^{-2})$			
	$ \text{Bias}/\tau(\mathbf{1}, \mathbf{0}) $	Bias	SD	RMSE	$ \text{Bias}/\tau(\mathbf{1}, \mathbf{0}) $	Bias	SD	RMSE
CR	4.921	-2.018	8.677	8.908	0.844	-0.346	8.308	8.315
CL	5.046	-2.069	7.582	7.859	1.065	-0.437	7.581	7.594
Ind	5.075	-2.081	8.429	8.682	0.741	-0.304	7.866	7.872
Both	5.033	-2.061	7.500	7.778	0.982	-0.403	7.071	7.082

Table 3 investigates the properties of different network A/B testing procedures for estimating  $\tau(\mathbf{1}, \mathbf{0})$ . The DIM estimator generally has a larger bias than CAE regardless of randomization schemes. Combining with the conclusion drawn from Figure 7, Table 3 also indicates that CAE under the randomization scheme with finer covariate balancing properties achieves higher efficiency, that is, CAE under CLAR-Both has the minimal standard deviation and the minimal RMSE.

## 7. Conclusions

In this article, we introduce a novel cluster-adaptive network A/B testing procedure aimed at enhancing the balance of covariates and improving the efficiency of ATE estimation. Our approach combines the CLAR for the randomization step and the CAE for the estimation step. The primary goal is to achieve finer covariate balance, leading to enhanced comparable treatment arms, with an extra gain of an improved accuracy on the estimation of ATE. Moreover, we establish the convergence rate of the imbalance measure, which provides a foundation for understanding the effectiveness of our procedure in achieving covariate balancing. The asymptotic properties of CAE under CLAR are derived, highlighting the efficiency gain obtained from the enhanced covariate balance. Our theoretical findings, together with numerical simulations, underscore the superiority of our approach over purely relying on complete randomization for the estimation of ATE.

The use of informative units in estimation leads to an inevitable loss of balancing ability with CLAR as we demonstrated in Section 5 and Corollary 4.1. This issue can be partially solved with the rerandomized-adaptive randomization proposed in Liu et al. (2022). Consequently, the resulting covariate balancing may lead to a further improved ATE estimation.

Our proposed network A/B testing procedure is particularly suitable for networks characterized by sparsely connected clusters. However, in cases where the observed network has less densely connected clusters, Assumption 4.3 might not hold. This situation could also hinder the identifiability of the "all versus nothing" average treatment effect even with neighborhood treatment assumptions and graph-cluster randomization. In such instances,

an alternative estimand, such as the overall treatment effect (Forastiere et al., 2021; Sävje et al., 2021), might be more appropriate. It would be interesting to study the potential gain of balancing individual-level network characteristics in estimating the treatment effect for this type of problem.

Assumption 4.3 depends largely on the assumption of a homogeneous treatment effect, which may not always hold in practical scenarios. When the covariates of a cluster interact with the treatment, the effect estimated from different clusters may vary. Moreover, the outer nodes linked to distinct clusters could experience various spillover effects. An potentially useful approach to address this issue and to estimate such heterogeneous effects within the context of network interference may involve combining our proposed A/B testing procedure with the tree-based method presented in Bargagli-Stoffi et al. (2020). This problem is left as an important direction for future research.

Performing statistical inference for the treatment effect is also of great importance in A/B testing (Ma et al., 2015, 2020). It would be straightforward to conduct the randomization test (Athey et al., 2018; Basse et al., 2019; Pouget-Abadie et al., 2019) using our proposed network A/B testing procedure. Alternatively, a Wald type of inference could be developed based on the asymptotic properties of the CAE. This approach requires an accurate estimate of the standard error of the CAE. It would be desirable to extend the framework proposed in this paper to obtain a valid statistical inference for the network A/B testing problem.

In practice, the presence of influential unobserved covariates poses a persistent challenge, potentially affecting the estimation of the ATE (Rosenbaum et al., 2010; Rosenberger and Lachin, 2015). From a design perspective, it is always important to ensure the comparability of the treatment arm with respect to both of the observed and unobserved covariates. Additionally, the impact of unobserved covariates on the subsequent ATE inference are essential for developing robust inference against varied unobserved covariates settings. Recent work by Liu and Hu (2022) studies the properties of a class of covariate-adaptive randomization for balancing unobserved covariates. Liu and Hu (2023) further extends their work to study the subsequent statistical inference. It is desirable to extend their work to address the challenges posed by unobserved covariates within the context of network A/B testing.

## Acknowledgments

The authors would like to thank the Associate Editor and the three anonymous Referees for the insightful comments that lead to a much improved version of the paper. The work of Ping Li and Feifang Hu was partially conducted at Baidu, and they thank Baidu Cognitive Computing Lab for the support. Yang Liu’s research was supported by the National Natural Science Foundation of China (grant nos. 12301324). This research was also supported by Public Computing Cloud, Renmin University of China.

## Appendix A. Proof of the Main Results

This section provides the proof and supplementary lemmas for our theoretical results. We prove the balance properties of CLAR in Section A.1. The asymptotic properties of CAE are investigated in Section A.2.

### A.1 Cluster-Adaptive Randomization

When applying CLAR to balance covariates, we consider: (1) using cluster-level covariates  $\boldsymbol{\xi}_j = \mathbf{X}_{j,\text{CL}}$ ; (2) using individual-level covariates by setting  $\boldsymbol{\xi}_j = \bar{\mathbf{X}}_{j,\text{IN}}$ ; or (3) using both of the two levels of covariates  $\boldsymbol{\xi}_j = (\mathbf{X}_{j,\text{CL}}, \bar{\mathbf{X}}_{j,\text{IN}}^\top)^\top$ . Because (1) and (2) can be understood as special cases of (3), we will only prove our theoretical results with  $\boldsymbol{\xi}_j = (\mathbf{X}_{j,\text{CL}}, \bar{\mathbf{X}}_{j,\text{IN}}^\top)^\top$ .

By LLN and Assumption 4.1, it follows that

$$\bar{\boldsymbol{\xi}} = m^{-1} \sum_{j=1}^m \boldsymbol{\xi}_j = \boldsymbol{\mu}_\xi + o_p(1), \quad \text{and} \quad S_m = \Sigma_{\xi\xi} + o_p(1),$$

where  $\Sigma_{\xi\xi}$  is positive definite. Then, (1) is equivalent to

$$\begin{aligned} \tilde{M}_{2k} &= \frac{k}{2} (\bar{\boldsymbol{\xi}}_{1,2k} - \bar{\boldsymbol{\xi}}_{0,2k})^\top \Sigma_{\xi\xi}^{-1} (\bar{\boldsymbol{\xi}}_{1,2k} - \bar{\boldsymbol{\xi}}_{0,2k}) \\ &= \frac{1}{2k} \Lambda_k^\top \Lambda_k, \end{aligned} \tag{8}$$

where  $\Lambda_k = \sum_{j=1}^k \Delta_j = \sum_{j=1}^k (2Z_{2j-1} - 1)(\boldsymbol{\eta}_{2j-1} - \boldsymbol{\eta}_{2j})$  and  $\boldsymbol{\eta}_j = \Sigma_{\xi\xi}^{-1/2} (\boldsymbol{\xi}_j - \boldsymbol{\mu}_\xi)$ . Furthermore, (8) implies that

$$\begin{aligned} \tilde{M}_{2k}^{(0)} - \tilde{M}_{2k}^{(1)} &= \frac{2}{k} (\Lambda_{k-1} + \Delta_k)^\top (\Lambda_{k-1} + \Delta_k) - \frac{2}{k} (\Lambda_{k-1} - \Delta_k)^\top (\Lambda_{k-1} - \Delta_k) \\ &= \frac{8}{k} \Delta_k^\top \Lambda_{k-1}, \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E}[\tilde{M}_{2k}^{(0)} - \tilde{M}_{2k}^{(1)} | \Lambda_{k-1}, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}] \\ &= -\frac{8}{k} (2\rho - 1) \text{sign} \left\{ (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k})^\top \Lambda_{k-1} \right\} (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k})^\top \Lambda_{k-1}, \end{aligned}$$

for  $\rho > 1/2$ . Analyzing the signs of  $(\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k})^\top \Lambda_{k-1}$  yields that the assignment of a pair of treatments  $(Z_{2k-1}, Z_{2k})$  by CLAR is always in the direction to minimizing  $|\Lambda_{k-1}|$  to zero. Therefore,  $\{\Lambda_k\}_{k=1}^\infty$  is a Markov chain under CLAR.

We study the properties of  $\{\Lambda_k\}_{k=1}^\infty$  in the following lemmas, which are the building blocks for the proof of Theorem 4.1 and Theorem 4.2.

**Lemma A.1** *Suppose Assumption 4.1 holds and  $\mathbb{E}[\|\boldsymbol{\eta}\|^a] < \infty$  for a given  $a > 2$ . Then  $\mathbb{E}[\|\Lambda_m\|^2] = O(m^{\frac{1}{a-1}})$  and  $\Lambda_m = O_p(m^{\frac{1}{2(a-1)}}) = o_p(m^{\frac{1}{2}})$ .*

**Proof** We follow Ma et al. (2024) to derive this lemma. For a vector  $\mathbf{u} = (u_1, \dots, u_d)^\top \in \mathbb{R}^d$ , let  $\|\mathbf{u}\|^2 = \mathbf{u}^\top \mathbf{u} = \sum_{k=1}^d u_k^2$ . By Taylor expansion and some calculations, we have

$$\|\mathbf{u} + \mathbf{v}\|^a - \|\mathbf{u}\|^a \leq a(\mathbf{u}^\top \mathbf{v})\|\mathbf{u}\|^{a-2} + J_a(\|\mathbf{v}\|^a + \|\mathbf{v}\|^2\|\mathbf{u}\|^{a-2}) \quad (9)$$

for  $a > 2$ ,  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$ , and  $J_a$  is a constant depending on  $a$ . Setting  $\mathbf{u} = \Lambda_k$  and  $\mathbf{v} = \Delta_k = (2T_{2k-1} - 1)(\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k})$  arrives

$$\begin{aligned} \|\Lambda_k\|^a - \|\Lambda_{k-1}\|^a &\leq a(\Lambda_{k-1}^\top \Delta_k)\|\Lambda_{k-1}\|^{a-2} \\ &\quad + J_a \{ \|\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k}\|^a + \|\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k}\|^2 \|\Lambda_{k-1}\|^{a-2} \}. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{E}[\|\Lambda_k\|^a | \mathcal{F}_{k-1}] - \|\Lambda_{k-1}\|^a &\leq -a(2q-1)\mathbb{E} \left[ \left| (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k})^\top \Lambda_{k-1} \right| \middle| \mathcal{F}_{k-1} \right] \|\Lambda_{k-1}\|^{a-2} \\ &\quad + 2J_a \{ \mathbb{E}[\|\boldsymbol{\eta}\|^a] + \mathbb{E}[\|\boldsymbol{\eta}\|^2] \|\Lambda_{k-1}\|^{a-2} \}. \end{aligned} \quad (10)$$

To derive a sharper bound for (10), set  $\tilde{\Lambda}_k = \|\Lambda_k\|^{-1}\Lambda_k$ , such that  $\tilde{\Lambda}_k^\top \tilde{\Lambda}_k = \|\Lambda_k\|^{-2}\|\Lambda_k\|^2 = 1$ . Then, we have

$$\begin{aligned} \mathbb{E} \left[ \left| (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k})^\top \Lambda_{k-1} \right| \middle| \mathcal{F}_{k-1} \right] &= \|\Lambda_k\| \mathbb{E} \left[ \left| (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k})^\top \tilde{\Lambda}_{k-1} \right| \middle| \mathcal{F}_{k-1} \right] \\ &\geq \|\Lambda_k\| \inf_{\tilde{\Lambda}_k^\top \tilde{\Lambda}_k = 1} \mathbb{E} \left[ \left| (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k})^\top \tilde{\Lambda}_{k-1} \right| \middle| \mathcal{F}_{k-1} \right] \\ &= 2\|\Lambda_k\| \end{aligned}$$

provided that  $\mathbb{V}[\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k}] = 2\mathbf{I}$  and its minimal eigenvalue is 2, where  $\mathbf{I}$  is the identity matrix and  $\boldsymbol{\eta}_{2k-1}$  and  $\boldsymbol{\eta}_{2k}$  are independent. This suggests that

$$\begin{aligned} \mathbb{E}[\|\Lambda_k\|^a | \mathcal{F}_{k-1}] - \|\Lambda_{k-1}\|^a &\leq -a2(2q-1)\|\Lambda_{k-1}\|^{a-1} \\ &\quad + 2J_a \{ \mathbb{E}[\|\boldsymbol{\eta}\|^a] + \mathbb{E}[\|\boldsymbol{\eta}\|^2] \|\Lambda_{k-1}\|^{a-2} \}. \end{aligned} \quad (11)$$

Since  $f(x) = -c_1x^{a-1} + c_2x^{a-2} + c_3$  for  $x \geq 0$  has a unique upper bound determined by the constants  $c_1, c_2, c_3$  and  $\|\Lambda_k\| \geq 0$ , there exists  $J'_a$  such that

$$\begin{aligned} \mathbb{E}[\|\Lambda_k\|^a | \mathcal{F}_{k-1}] - \|\Lambda_{k-1}\|^a &\leq J'_a, \\ \text{and hence } \mathbb{E}[\|\Lambda_k\|^a] &\leq J'_a + \mathbb{E}[\|\Lambda_{k-1}\|^a] \leq kJ'_a. \end{aligned}$$

For  $a = 2$ , it follows from (11) that

$$\mathbb{E}[\|\Lambda_k\|^2] \leq \mathbb{E}[\|\Lambda_{k-1}\|^2] - 4(2\rho-1)\mathbb{E}[\|\Lambda_{k-1}\|] + 2J_2\mathbb{E}\|\boldsymbol{\eta}\|^2,$$

and thus  $\mathbb{E}[\|\Lambda_k\|^2] \leq \mathbb{E}[\|\Lambda_{k-1}\|^2]$  when  $\mathbb{E}[\|\Lambda_{k-1}\|] > (4(2\rho-1))^{-1}2J_2\mathbb{E}\|\boldsymbol{\eta}\|^2$ . Consequently, choose  $\tilde{k}$  to be the largest  $j$  for  $1 \leq j \leq k$  satisfying

$$\mathbb{E}[\|\Lambda_{k-1}\|] \leq (4(2\rho-1))^{-1}2J_2\mathbb{E}\|\boldsymbol{\eta}\|^2,$$

then we have

$$\mathbb{E} [\|\Lambda_k\|^2] \leq \mathbb{E} [\|\Lambda_{\tilde{k}+1}\|^2] \leq \mathbb{E} [\|\Lambda_{\tilde{k}}\|^2] + 2J_2\mathbb{E}\|\boldsymbol{\eta}\|^2. \quad (12)$$

Finally, for  $a > 2$ , setting  $s = a - 1$  and  $t = \frac{a-1}{a-2}$  such that  $s^{-1} + t^{-1} = 1$ , then Hölder inequality implies that

$$\begin{aligned} \mathbb{E} [\|\Lambda_{\tilde{k}}\|^2] &= \mathbb{E} \left[ \|\Lambda_{\tilde{k}}\|^{\frac{a-2}{a-1}} \cdot \|\Lambda_{\tilde{k}}\|^{\frac{1}{a-1}+1} \right] \\ &\leq \left\{ \mathbb{E} \left[ \|\Lambda_{\tilde{k}}\|^{\frac{a-2}{a-1} \cdot t} \right] \right\}^{\frac{1}{t}} \left\{ \mathbb{E} \left[ \|\Lambda_{\tilde{k}}\|^{(\frac{1}{a-1}+1)s} \right] \right\}^{\frac{1}{s}} \\ &= (\mathbb{E}\|\Lambda_{\tilde{k}}\|)^{\frac{1}{t}} \left\{ \mathbb{E} [\|\Lambda_{\tilde{k}}\|^a] \right\}^{\frac{1}{s}} \leq (\tilde{k})^{\frac{1}{a-1}} J'_a + 2J_2\mathbb{E}\|\boldsymbol{\eta}\|^2. \end{aligned} \quad (13)$$

Plugging (13) in (12) gives

$$\mathbb{E}[\|\Lambda_m\|^2] \leq m^{\frac{1}{a-1}} J'_a + 2J_2\mathbb{E}\|\boldsymbol{\eta}\|^2 = O(m^{\frac{1}{a-1}}).$$

This completes the proof for this lemma. ■

We next prove Theorem 4.1 in the following lemma.

**Lemma A.2** *Under the condition of Lemma A.1,  $M_m = O_p(m^{-1})$ .*

**Proof** First, assume  $m = 2k$ , and plugging the result of Lemma A.1 in (8), we have

$$\tilde{M}_{2k} = \frac{\Lambda_k^\top \Lambda_k}{2k} = o_p(k^{1/2})o_p(k^{1/2}) \cdot \frac{1}{2k} = o_p(1).$$

Next, assume  $m = 2k + 1$ , then we have

$$\begin{aligned} \tilde{M}_{2k+1} &= \frac{2k+1}{4} (\bar{\boldsymbol{\xi}}_{1,2k+1} - \bar{\boldsymbol{\xi}}_{0,2k+1}) \Sigma_{\boldsymbol{\xi}\boldsymbol{\xi}}^{-1} (\bar{\boldsymbol{\xi}}_{1,2k+1} - \bar{\boldsymbol{\xi}}_{0,2k+1}) \\ &= \frac{1}{2k} \{\Lambda_k + (2Z_{2k+1} - 1)\boldsymbol{\eta}_{2k+1}\}^\top \{\Lambda_k + (2Z_{2k+1} - 1)\boldsymbol{\eta}_{2k+1}\} + o_p(k^{-1}) \\ &= \frac{1}{2k} \left( \Lambda_k^\top \Lambda_k + 2(2Z_{2k+1} - 1)\boldsymbol{\eta}_{2k+1}^\top \Lambda_k + \boldsymbol{\eta}_{2k+1}^\top \boldsymbol{\eta}_{2k+1} \right) + o_p(1) \\ &= \frac{\Lambda_k^\top \Lambda_k}{2k} + O_p(k^{-1}) = o_p(1), \end{aligned}$$

provided that  $2Z_{2k+1} - 1$  equals either 1 or -1 and  $\boldsymbol{\eta}_{2k+1} = O_p(1)$ , when  $k$  goes to infinity. ■

The following lemma ensures that the Poisson equation associated with the chain  $\{\Lambda_k\}_{k=1}^\infty$  admits a proper root, which provides the foundation for deriving property of CAE under CLAR.

**Lemma A.3** *Under Assumption 4.1,  $\{\Lambda_k\}_{k=1}^\infty$  is a positive Harris recurrent Markov chain.*

**Proof** It is obvious that  $\{\Lambda_k\}_{k=1}^\infty$  is a Markov chain with period 4. Let  $\mathbf{P}_\lambda(x, A)$  denote the transition probability measure of  $\{\Lambda_k\}_{k=1}^\infty$ . It follows from (11) that

$$\begin{aligned} \mathbf{P}_\lambda \|\Lambda\|^a - \|\Lambda\|^a &\leq -a(2\rho - 1)\|\Lambda\|^{a-1} + 2J_a \left\{ \mathbb{E}[\|\boldsymbol{\eta}\|^a] + \mathbb{E}[\|\boldsymbol{\eta}\|^2] \|\Lambda\|^{a-2} \right\} \\ &\leq 1 - a(\rho - 1/2)\|\Lambda\|^{a-1} + b\mathbb{I}\{\|\Lambda\| \leq c\} \end{aligned} \quad (14)$$

for some  $b, c > 0$ . Then  $\{\Lambda_k\}_{k=1}^\infty$  is a positive Harris recurrent Markov chain, if  $\{\Lambda_k\}_{k=1}^\infty$  is  $\psi$ -irreducible and  $\{\|\Lambda\| \leq c\}$  is petite for every  $c > 0$  by Theorems 11.3.4, 14.2.3 and 14.3.6 of Meyn and Tweedie (2013). To see this, we follow Section 4.2 of Meyn and Tweedie (2013) and define the resolvent  $K_{1/2}(x, A)$  as

$$K_{1/2}(x, A) = \sum_{k=0}^{\infty} \mathbf{P}_\lambda^k(x, A) 2^{-(k+1)}.$$

The key is to show that there exists  $\phi(\cdot) > 0$  such that

$$K_{1/2}(x, A) \geq \int_{A-x} \phi(y) dy \quad \text{for any Borel set } A, \quad (15)$$

then  $\{\Lambda_k\}_{k=1}^\infty$  is a  $\psi$ -irreducible  $T$ -chain by Proposition 5.5.5 and Theorem 6.0.1 of Meyn and Tweedie (2013).

To begin with, note that

$$\begin{aligned} &\mathbb{P}\{\Lambda_k \in A \mid \Lambda_{k-1} = x, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}\} \\ &= \mathbb{P}\{\Lambda_k + (2Z_{2k-1} - 1)(\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k}) \in A \mid \Lambda_{k-1}, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}\} \\ &= \mathbb{P}\{\Lambda_k + (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k}) \in A \mid \Lambda_{k-1}, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}, Z_{2k-1} = 1, Z_{2k} = 0\} \\ &\quad \times \mathbb{P}\{Z_{2k-1} = 1, Z_{2k} = 0 \mid \Lambda_{k-1}, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}\} \\ &+ \mathbb{P}\{\Lambda_k - (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k}) \in A \mid \Lambda_{k-1}, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}, Z_{2k-1} = 1, Z_{2k} = 0\} \\ &\quad \times \mathbb{P}\{Z_{2k-1} = 0, Z_{2k} = 1 \mid \Lambda_{k-1}, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}\}, \end{aligned}$$

let  $\Gamma_\boldsymbol{\eta}(y)$  denote the joint distribution of  $\boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}$ , then it follows from taking the integral on both sides of the above equation that

$$\begin{aligned} &\mathbb{P}\{\Lambda_k \in A \mid \Lambda_{k-1}\} \\ &\geq q \wedge (1 - q) \mathbb{P}\{\Lambda_k + (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k}) \in A \mid \Lambda_{k-1} = x, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}\} \\ &= q \wedge (1 - q) \mathbb{I}\{\Lambda_{k-1} = x\} \Gamma_\boldsymbol{\eta}(A - x), \end{aligned}$$

and

$$\begin{aligned} &\mathbb{P}\{\Lambda_k \in A \mid \Lambda_{k-1} = x, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}\} \\ &\geq \rho \wedge (1 - \rho) \mathbb{P}\{\Lambda_k - (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k}) \in A \mid \Lambda_{k-1} = x, \boldsymbol{\eta}_{2k-1}, \boldsymbol{\eta}_{2k}\} \\ &= \rho \wedge (1 - \rho) \mathbb{I}\{\Lambda_{k-1} = x\} \Gamma_{-\boldsymbol{\eta}}(A - x). \end{aligned}$$

By the Chapman-Kolmogorov equations, for some  $n_c > 0$  and  $n_0 > 0$ , we have

$$\begin{aligned}
 \mathbf{P}_\lambda^{2n_c k}(x, A) &= \mathbb{P} \{ \Lambda_{2n_c k + n_0} \in A \mid \Lambda_{n_0} = x \} \\
 &\geq \mathbb{P} \{ \Lambda_{n_c k + n_0} \in A \mid \Lambda_{n_0} = x \} \mathbb{P} \{ \Lambda_{2n_c k + n_0} \in A \mid \Lambda_{n_c k + n_0} \in A \} \\
 &\geq \{ \rho \wedge (1 - \rho) \}^{2n_c k} \Gamma_{\boldsymbol{\eta}}^{n_c k} * \Gamma_{-\boldsymbol{\eta}}^{n_c k}(A - x) \\
 &\geq \{ \rho \wedge (1 - \rho) \}^{2n_c k} c_\gamma^{2k} \int_{A-x} \tilde{\gamma}^{k*}(y) dy
 \end{aligned}$$

where  $\Gamma_{\boldsymbol{\eta}}^k(y)$  is the  $k$ -convolution of  $(\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top$ ,  $\Gamma_{\boldsymbol{\eta}} * \Gamma_{-\boldsymbol{\eta}}(y)$  is the convolution of  $(\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top$  and  $(-\boldsymbol{\eta}_1^\top, -\boldsymbol{\eta}_2^\top)^\top$ , and  $\tilde{\gamma}(y) = \gamma(y) * \gamma(-y)$  is convolution of  $\gamma(y)$  and  $\gamma(-y)$ . Combining the above equation with the definition of the resolvent, we have

$$\begin{aligned}
 K_{1/2}(x, A) &= \sum_{k=0}^{\infty} \mathbf{P}_\lambda^k(x, A) 2^{-(k+1)} \\
 &\geq 2^{-1} \sum_{k=0}^{\infty} \mathbf{P}_\lambda^{2n_c k}(x, A) 2^{-2n_c k} \\
 &\geq \int_{A-x} 2^{-1} \sum_{k=0}^{\infty} \left\{ [\rho \wedge (1 - \rho)]^{2n_c} c_\gamma^2 \right\}^k \tilde{\gamma}^{k*}(y) dy = \int_{A-x} \phi(y) dy,
 \end{aligned}$$

by setting  $\phi(y) = 2^{-1} \sum_{k=0}^{\infty} \left\{ [\rho \wedge (1 - \rho)]^{2n_c} c_\gamma^2 \right\}^k \tilde{\gamma}^{k*}(y) dy$ . This completes the proof for this lemma.  $\blacksquare$

## A.2 Cluster-Adjusted Estimator

We first derive the decomposition of  $\hat{\tau}_{\text{CAE}}$ . First, notice that

$$\hat{\tau}_{\text{CAE}} - \tau(\mathbf{1}, \mathbf{0}) = B_{m,1} + B_{m,2} + B_{m,3} + B_{m,4},$$

where

$$\begin{aligned}
 B_{m,1} &= \frac{1}{m_1} \sum_{j=1}^m Z_j \left( \hat{Y}_j(\mathbf{1}) - \hat{\mu}_j(\mathbf{1}) \right) - \frac{1}{m_0} \sum_{j=1}^m (1 - Z_j) \left( \hat{Y}_j(\mathbf{0}) - \hat{\mu}_j(\mathbf{0}) \right), \\
 B_{m,2} &= \frac{1}{m_1} \sum_{j=1}^m Z_j \left( \hat{\mu}_j(\mathbf{1}) - \hat{\mu}_j^c(\mathbf{1}) \right) \frac{c_j - n_j}{c_j} - \frac{1}{m_0} \sum_{j=1}^m (1 - Z_j) \left( \hat{\mu}_j(\mathbf{0}) - \hat{\mu}_j^c(\mathbf{0}) \right) \frac{c_j - n_j}{c_j}, \\
 B_{m,3} &= \sum_{j=1}^m \left\{ \left( \frac{Z_j}{m_1} - \frac{1}{m} \right) \bar{\mu}_j(\mathbf{1}) - \left( \frac{1 - Z_j}{m_0} - \frac{1}{m} \right) \bar{\mu}_j(\mathbf{0}) \right\}, \\
 B_{m,4} &= \sum_{j=1}^m \left( \frac{1}{m} - \frac{c_j}{n} \right) [\bar{\mu}_j(\mathbf{1}) - \bar{\mu}_j(\mathbf{0})],
 \end{aligned}$$

$$\hat{\mu}_j(\mathbf{t}) = n_j^{-1} \sum_{i \in \text{Inf}_j} \mathbb{E}[Y_i(\mathbf{t})], \text{ and } \hat{\mu}_j^c(\mathbf{t}) = e_j^{-1} \sum_{i \in \text{Inf}_j^c \cap C_j} \mathbb{E}[Y_i(\mathbf{t})] \text{ for } \mathbf{t} \in \{\mathbf{0}, \mathbf{1}\}.$$

It is easy to check from Assumption 4.3.1 that  $B_{m,3} = B_{m,4} = 0$ . We next derive the bias term  $B_{m,2}$ . It follows from the definition of  $\Psi_{\text{Diff,Inf}}$  in Assumption 4.3.2 that

$$\begin{aligned} & \frac{1}{m_1} \sum_{j=1}^m Z_j (\hat{\mu}_j(\mathbf{1}) - \hat{\mu}_j^c(\mathbf{1})) \frac{c_j - n_j}{c_j} - \frac{1}{m_0} \sum_{j=1}^m (1 - Z_j) (\hat{\mu}_j(\mathbf{0}) - \hat{\mu}_j^c(\mathbf{0})) \frac{c_j - n_j}{c_j} \\ & \leq \frac{1}{m_1} \sum_{j=1}^m \left| \frac{c_j - n_j}{c_j} \right| \cdot |\hat{\mu}_j(\mathbf{1}) - \hat{\mu}_j^c(\mathbf{1})| + \frac{1}{m_0} \sum_{j=1}^m \left| \frac{c_j - n_j}{c_j} \right| \cdot |\hat{\mu}_j(\mathbf{0}) - \hat{\mu}_j^c(\mathbf{0})| \\ & \lesssim \frac{2}{m} \sum_{j=1}^m \left| \frac{c_j - n_j}{c_j} \right| \cdot \{ |\hat{\mu}_j(\mathbf{1}) - \hat{\mu}_j^c(\mathbf{1})| + |\hat{\mu}_j(\mathbf{0}) - \hat{\mu}_j^c(\mathbf{0})| \} \leq 2\Psi_{\text{Diff,Inf}}. \end{aligned}$$

We next derive  $B_{m,1}$  as follows:

$$\begin{aligned} B_{m,1} &= \frac{1}{m_1} \sum_{j=1}^m Z_j (\hat{Y}_j(\mathbf{1}) - \hat{\mu}_j(\mathbf{1})) - \frac{1}{m_0} \sum_{j=1}^m (1 - Z_j) (\hat{Y}_j(\mathbf{0}) - \hat{\mu}_j(\mathbf{0})) \\ &= m^{-1} \left( 1 + \frac{D_m}{m} \right)^{-1} \left( 1 - \frac{D_m}{m} \right)^{-1} \left\{ R_{m,1} - \frac{D_m}{m} \cdot R_{m,2} \right\}, \end{aligned} \quad (16)$$

$$\begin{aligned} \text{where } R_{m,1} &= \sum_{j=1}^m (2Z_j - 1) \left[ (\hat{Y}_j(\mathbf{1}) + \hat{Y}_j(\mathbf{0})) - (\hat{\mu}_j(\mathbf{1}) + \hat{\mu}_j(\mathbf{0})) \right] \\ &\quad + \sum_{j=1}^m \left[ (\hat{Y}_j(\mathbf{1}) - \hat{Y}_j(\mathbf{0})) - (\hat{\mu}_j(\mathbf{1}) - \hat{\mu}_j(\mathbf{0})) \right], \end{aligned}$$

$$\begin{aligned} \text{and } R_{m,2} &= \sum_{j=1}^m (2Z_j - 1) \left[ (\hat{Y}_j(\mathbf{1}) - \hat{Y}_j(\mathbf{0})) - (\hat{\mu}_j(\mathbf{1}) - \hat{\mu}_j(\mathbf{0})) \right] \\ &\quad + \sum_{j=1}^m \left[ (\hat{Y}_j(\mathbf{1}) + \hat{Y}_j(\mathbf{0})) - (\hat{\mu}_j(\mathbf{1}) + \hat{\mu}_j(\mathbf{0})) \right]. \end{aligned}$$

Note that under either CLAR or CR,  $m^{-1/2}D_m = O_p(1)$  and thus  $(1 + m^{-1}D_m)^{-1}(1 - m^{-1}D_m)^{-1} = 1 + o_p(1)$ . By Lemmas A.4, A.5 and A.6 derived below, it follows that  $(m\sigma_{m,\text{CAE}})^{-1}R_{m,2} = O_p(1)$ . The above two results combing with (16) give

$$\sigma_{m,\text{CAE}}^{-1} (\hat{\tau}_{\text{CAE}} - \tau(\mathbf{1}, \mathbf{0})) = (m\sigma_{m,\text{CAE}})^{-1} (R_{m,1} + mB_{m,2}) + o_p(1). \quad (17)$$

To derive the normality of the CAE, we can further decompose  $R_{m,1}$  and  $R_{m,2}$  as follows:

$$\begin{aligned} R_{m,1} &= R_{m,1,1} + R_{m,1,2} + 2R_{m,1,3}, \\ R_{m,2} &= R_{m,2,1} + R_{m,2,2} + 2R_{m,2,3}, \end{aligned}$$

where  $R_{m,1,1} = \sum_{j=1}^m (2Z_j - 1)[\hat{f}_j(\mathbf{1}, \mathbf{X}_j) + \hat{f}_j(\mathbf{0}, \mathbf{X}_j)]$ ,  $R_{m,2,1} = \sum_{j=1}^m (2Z_j - 1)[\hat{f}_j(\mathbf{1}, \mathbf{X}_j) - \hat{f}_j(\mathbf{0}, \mathbf{X}_j)]$ ,  $R_{m,1,2} = \sum_{j=1}^m [\hat{f}_j(\mathbf{1}, \mathbf{X}_j) - \hat{f}_j(\mathbf{0}, \mathbf{X}_j)]$ ,  $R_{m,2,2} = \sum_{j=1}^m [\hat{f}_j(\mathbf{1}, \mathbf{X}_j) + \hat{f}_j(\mathbf{0}, \mathbf{X}_j)]$ ,  $R_{m,1,3} = \sum_{j=1}^m [Z_j \hat{\epsilon}_j(\mathbf{1}) - (1 - Z_j) \hat{\epsilon}_j(\mathbf{0})]$ , and  $R_{m,2,3} = \sum_{j=1}^m [Z_j \hat{\epsilon}_j(\mathbf{1}) + (1 - Z_j) \hat{\epsilon}_j(\mathbf{0})]$ .

Based on the decomposition given above, we derive the properties of  $R_{m,1,1}$ ,  $R_{m,1,2}$ ,  $R_{m,1,3}$ ,  $R_{m,2,1}$ ,  $R_{m,2,2}$ , and  $R_{m,2,3}$  in the following lemmas. Then we prove Theorem 4.2 and Corollary 4.1 in Sections A.2.1 and A.2.2, respectively.

**Lemma A.4** *Suppose Assumptions 2.2, 4.1, 4.2, and 4.3 hold. Under CLAR, there exists  $\sigma_{\text{Design}}^2 > 0$  and  $\tilde{\sigma}_{\text{Design}}^2 > 0$  such that*

$$m^{-1/2} (R_{m,1,1}, R_{m,1,2}) \xrightarrow{\mathbf{D}} (\zeta_{1,1}, \zeta_{1,2})^\top \quad \text{and} \quad m^{-1/2} (R_{m,2,1}, R_{m,2,2}) \xrightarrow{\mathbf{D}} (\zeta_{2,1}, \zeta_{2,2})^\top .$$

Here,  $\zeta_{1,1}$  and  $\zeta_{1,2}$  are independent and satisfy that  $\zeta_{1,1} \sim \mathcal{N}(0, \sigma_{\text{Design}}^2)$  and  $\zeta_{1,2} \sim \mathcal{N}(0, \sigma_f^2)$ , and  $\zeta_{2,1}$  and  $\zeta_{2,2}$  are independent and satisfy  $\zeta_{2,1} \sim \mathcal{N}(0, \tilde{\sigma}_{\text{Design}}^2)$  and  $\zeta_{2,2} \sim \mathcal{N}(0, \tilde{\sigma}_f^2)$ .

**Proof**

Note that the derivation for the joint distribution of  $(R_{m,2,1}, R_{m,2,2})^\top$  is an analog to that of  $(R_{m,1,1}, R_{m,1,2})^\top$ . Therefore, we only derive the joint distribution of  $(R_{m,1,1}, R_{m,1,2})^\top$  for the simplicity of presentation. Without loss of generality, we assume that  $m$  is even. Let  $r_{j,1,1} = (2Z_j - 1)[\hat{f}_j(\mathbf{1}, \mathbf{X}_j) + \hat{f}_j(\mathbf{0}, \mathbf{X}_j)]$  and  $r_{j,1,2} = \hat{f}_j(\mathbf{1}, \mathbf{X}_j) - \hat{f}_j(\mathbf{0}, \mathbf{X}_j)$ . By Assumption 4.3.1,  $\mathbb{E}[\hat{f}_j(\mathbf{t}, \mathbf{X}_j) | \tilde{\mathcal{F}}_k] = 0$  for  $\mathbf{t} \in \{\mathbf{1}, \mathbf{0}\}$ , it follows that

$$\begin{aligned} \mathbb{E} \left[ r_{2k-1,1,1} + r_{2k,1,1} | \tilde{\mathcal{F}}_{k-1} \right] &= -(2\rho - 1) \mathbb{E} \left[ \left\{ \left( \hat{f}_{2k-1}(\mathbf{1}, \mathbf{X}_{2k-1}) - \hat{f}_{2k}(\mathbf{1}, \mathbf{X}_{2k}) \right) \right. \right. \\ &+ \left. \left. \left( \hat{f}_{2k-1}(\mathbf{0}, \mathbf{X}_{2k-1}) - \hat{f}_{2k}(\mathbf{0}, \mathbf{X}_{2k}) \right) \right\} \text{sign} \left\{ (\boldsymbol{\eta}_{2k-1} - \boldsymbol{\eta}_{2k})^\top \Lambda_{k-1} \right\} \middle| \tilde{\mathcal{F}}_{k-1} \right] \\ &:= h(\Lambda_{k-1}), \end{aligned}$$

$$\text{and} \quad \mathbb{E} \left[ r_{2k-1,1,2} + r_{2k,1,2} | \tilde{\mathcal{F}}_{k-1} \right] = 0.$$

Note that  $h(\Lambda_{k-1}) < \max_k \mathbb{E} |\hat{f}_{2k-1}(\mathbf{1}, \mathbf{X}_{2k-1}) - \hat{f}_{2k}(\mathbf{1}, \mathbf{X}_{2k})| + \max_k \mathbb{E} |\hat{f}_{2k-1}(\mathbf{0}, \mathbf{X}_{2k-1}) - \hat{f}_{2k}(\mathbf{0}, \mathbf{X}_{2k})| \leq \infty$ . By Lemma A.3,  $\{\Lambda_k\}_{k=1}^\infty$  is a positive Harris recurrent Markov chain satisfying  $h(-\Lambda) = -h(\Lambda)$ . Let  $\pi_\lambda(\cdot)$  and  $\mathbf{P}_\lambda(x, A)$  denote the invariant probability measure and the transition probability measure of  $\{\Lambda_k\}$ . Furthermore, let  $\mathbf{P}_\lambda[h(\Lambda)] = \int h(y) d\mathbf{P}_\lambda(\Lambda, dy)$  and consider the following Poisson equation:

$$\hat{h}(\Lambda) - \mathbf{P}_\lambda \hat{h}(\Lambda) = h(\Lambda) - \pi_\lambda(\Lambda). \quad (18)$$

It follows from (14) and Theorem 14.3.7 of Meyn and Tweedie (2013) that  $\pi_\lambda(\|\Lambda\|^{a-1}) \leq b/[a(q-1/2)c_0]$  and thus  $\max_n \mathbb{E}[\|\Lambda\|^{a-1}] < \infty$ . Take  $a = 2$ , and it follows from Theorem 17.4.2 of Meyn and Tweedie (2013) with  $V = \|\Lambda\|^2$  that the Poisson equation (18) admits a solution satisfying the bound  $|\hat{h}(\Lambda)| \leq K(\|\Lambda\|^2 + 1)$  for some  $K > 0$ .

By the symmetry of  $Z_j$  and  $1 - Z_j$ , the transition probabilities for  $\{\Lambda_k\}$  equal to that of  $\{-\Lambda_k\}$ . Hence,  $\pi_\lambda(\cdot)$  is symmetric,  $\pi_\lambda(h) = 0$ ,  $\hat{h}(\Lambda) - \mathbf{P}_\lambda \hat{h}(\Lambda) = h(\Lambda)$ , and  $\hat{h}(-\Lambda) = -\hat{h}(\Lambda)$ . Let  $\Delta r_{k,1,1} = (r_{2k-1,1,1} + r_{2k,1,1}) - \mathbb{E}[r_{2k-1,1,1} + r_{2k,1,1} | \tilde{\mathcal{F}}_{k-1}]$ ,  $\Delta r_{k,1,2} = \hat{h}(\Lambda_k) - \mathbb{E}[\hat{h}(\Lambda_k) | \tilde{\mathcal{F}}_{k-1}]$ , and  $\Delta r_{k,1,3} = [\hat{f}_{2k-1}(\mathbf{1}, \boldsymbol{\xi}) - \hat{f}_{2k-1}(\mathbf{0}, \boldsymbol{\xi})] + [\hat{f}_{2k}(\mathbf{1}, \boldsymbol{\xi}) - \hat{f}_{2k}(\mathbf{0}, \boldsymbol{\xi})]$ , then it follows that

$$\begin{aligned} \sum_{j=1}^m r_{j,1,1} &= \sum_{k=1}^{m/2} \{\Delta r_{k,1,1} + \Delta r_{k,1,2}\} + \hat{h}(\Lambda_0) - \hat{h}(\Lambda_{m/2}), \\ \sum_{j=1}^m r_{j,1,2} &= \sum_{k=1}^{m/2} \Delta r_{k,1,3}, \end{aligned}$$

where  $\{\Delta r_{k,1,1} + \Delta r_{k,1,2} + \Delta r_{k,1,3}\}_{k=1}^{m/2}$  is a sequence of zero-mean martingale differences. Since  $\pi_\lambda(V^2) = \pi_\lambda(\|\Lambda\|^4) < \infty$ , it follows from Theorem 17.5.3 of Meyn and Tweedie (2013) and the ergodic theorem that there exists  $\sigma_{\text{Design}}^2 > 0$  such that

$$\begin{aligned} \frac{1}{m} \sum_{k=1}^{m/2} \mathbb{E} \left[ (\Delta r_{k,1,1} + \Delta r_{k,1,2})^2 | \tilde{\mathcal{F}}_{k-1} \right] &:= \sigma_{\text{Design}}^2 + o_p(1), \\ \text{and } \frac{1}{m} \sum_{k=1}^{m/2} \mathbb{E} \left[ \Delta r_{k,1,3}^2 | \tilde{\mathcal{F}}_{k-1} \right] &= \frac{1}{m} \sum_{j=1}^m \mathbb{V}[f_j(\mathbf{1}, \boldsymbol{\xi}) - \hat{f}_j(\mathbf{1}, \boldsymbol{\xi}) | \tilde{\mathcal{F}}_0] = \sigma_f^2. \end{aligned}$$

It is easy to verify that  $\{2Z_{2k-1} - 1, 2Z_{2k} - 1, \Lambda_{k-1}, \boldsymbol{\xi}_{2k-1}, \boldsymbol{\xi}_{2k}\} \stackrel{\text{D}}{=} \{-(2Z_{2k-1} - 1), -(2Z_{2k} - 1), -\Lambda_{k-1}, \boldsymbol{\xi}_{2k-1}, \boldsymbol{\xi}_{2k}\}$  under  $\pi_\lambda$ , so that

$$\mathbb{E}_{\pi_\lambda}[(\Delta r_{k,1,1} + \Delta r_{k,1,2})\Delta r_{k,1,3}] = -\mathbb{E}_{\pi_\lambda}[(\Delta r_{k,1,1} + \Delta r_{k,1,2})\Delta r_{k,1,3}],$$

and thus  $\mathbb{E}_{\pi_\lambda}[(\Delta r_{k,1,1} + \Delta r_{k,1,2})\Delta r_{k,1,3}] = 0$  and

$$\frac{1}{m} \sum_{k=1}^{m/2} \mathbb{E}[(\Delta r_{k,1,1} + \Delta r_{k,1,2})\Delta r_{k,1,3} | \tilde{\mathcal{F}}_{k-1}] = 0 + o_p(1).$$

Therefore, the Lemma A.4 follows from Theorem 17.5.3 of Meyn and Tweedie (2013) and the martingale central limit theorem (Hall and Heyde, 1980, Theorem 3.2).  $\blacksquare$

**Lemma A.5** *Suppose Assumptions 2.2, 4.1, 4.2, and 4.3 hold. Under CR,*

$$m^{-1/2} (R_{m,1,1}, R_{m,1,2}) \xrightarrow{\text{D}} (\zeta_{1,1}, \zeta_{1,2})^\top \quad \text{and} \quad m^{-1/2} (R_{m,2,1}, R_{m,2,2}) \xrightarrow{\text{D}} (\zeta_{2,1}, \zeta_{2,2})^\top.$$

Here  $\zeta_{1,1}$  and  $\zeta_{1,2}$  are independent and satisfy  $\zeta_{1,1} \sim \mathcal{N}(0, \tilde{\sigma}_f^2)$  and  $\zeta_{1,2} \sim \mathcal{N}(0, \sigma_f^2)$ , and  $\zeta_{2,1}$  and  $\zeta_{2,2}$  are independent and satisfy  $\zeta_{2,1} \sim \mathcal{N}(0, \sigma_f^2)$  and  $\zeta_{2,2} \sim \mathcal{N}(0, \tilde{\sigma}_f^2)$ .

**Proof** The proof of this lemma simply follows from the proof of the Lemma A.4 and the fact that  $\{Z_j\}_{j=1}^m$  and  $\{\mathbf{X}_j\}_{j=1}^m$  are independent.  $\blacksquare$

To deal with the network correlated errors, define the following indicator

$$G_{i_1, i_2} = \mathbb{I} \left\{ A_{i_1, i_2} + \max_{i_3} A_{i_1, i_3} A_{i_2, i_3} + \mathbb{I}\{i_1 \neq i_2, \text{ and } T_{i_1} = T_{i_2}\} \right\},$$

for  $1 \leq i_1, i_2 \leq n$ . The properties of  $R_{m,1,3}$  and  $R_{m,2,3}$  are studied in the following lemma.

**Lemma A.6** *Suppose Assumptions 2.2, 4.1, 4.2, and 4.3 hold. There exists a  $a > 0$  such that conditioning on  $\mathcal{Z}_m, \Xi_m$  and  $\tilde{\mathcal{F}}_0$ , for  $k = 1, 2$ ,*

$$\begin{aligned} \mathbf{d}_w \left( \tilde{R}_{m,k,3}, \mathcal{N}(0, 1) \right) &\leq \frac{\Psi_{\text{max,CL}}^2}{\sigma_{m,\epsilon}^3} \sum_{j=1}^m \left\{ \mathbb{E} |\hat{\epsilon}_j(\mathbf{1})|^3 + \mathbb{E} |\hat{\epsilon}_j(\mathbf{0})|^3 \right\} \\ &+ \frac{\sqrt{26} \Psi_{\text{max,CL}}^{3/2}}{\sqrt{\pi} \sigma_{m,\epsilon}^2} \sqrt{\sum_{j=1}^m \left\{ \mathbb{E} [\hat{\epsilon}_j^4(\mathbf{1})] + \mathbb{E} [\hat{\epsilon}_j^4(\mathbf{0})] \right\}} = O(m^{-a}), \end{aligned} \quad (19)$$

where  $\tilde{R}_{m,k,3} = \sigma_{m,\epsilon}^{-1} R_{m,k,3}$  and  $\mathbf{d}_w(X, Y)$  represents the Wasserstein distance between  $X$  and  $Y$ . Then, conditioning on  $\mathcal{Z}_m, \Xi_m$  and  $\tilde{\mathcal{F}}_0$ ,  $\tilde{R}_{m,k,3} \xrightarrow{\mathbf{D}} \mathcal{N}(0, 1)$  for  $k = 1, 2$ .

**Proof** The proof contains two steps. In the first step, we show that  $\mathbb{V}[\tilde{R}_{m,k,3} | \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m] = \sigma_{m,\epsilon}^2$ , for  $k = 1, 2$ . Then we apply Steins' method (Ross, 2011) to derive the bound of the Wasserstein distance between  $\tilde{R}_{m,k,3}$  and  $\mathcal{N}(0, 1)$ .

We shall first calculate the variance of  $R_{m,1,3}$  and  $R_{m,2,3}$ . Denote  $\mathbb{E}_m[\cdot] = \mathbb{E}[\cdot | \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m]$ ,  $\mathbb{V}_m[\cdot] = \mathbb{V}[\cdot | \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m]$ , and  $\text{Cov}_m[\cdot] = \mathbb{Cov}[\cdot | \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m]$ . It follows from Assumption 4.3.4 that

$$\begin{aligned} \mathbb{E}_m \left[ \sum_{j=1}^m Z_j \hat{\epsilon}_j(\mathbf{1}) \right] &= \mathbb{E}_m \left[ \sum_{j=1}^m (1 - Z_j) \hat{\epsilon}_j(\mathbf{0}) \right] = 0, \\ \mathbb{V}_m \left[ \sum_{j=1}^m Z_j \hat{\epsilon}_j(\mathbf{1}) \right] &= \sum_{j=1}^m Z_j \mathbb{E}_m [\hat{\epsilon}_j^2(\mathbf{1})] + \sum_{j_1 \neq j_2} Z_{j_1} Z_{j_2} \mathbb{E}_m [\hat{\epsilon}_{j_1}^2(\mathbf{1}) \hat{\epsilon}_{j_2}^2(\mathbf{1})], \\ \mathbb{V}_m \left[ \sum_{j=1}^m (1 - Z_j) \hat{\epsilon}_j(\mathbf{0}) \right] &= \sum_{j=1}^m (1 - Z_j) \mathbb{E}_m [\hat{\epsilon}_j^2(\mathbf{0})] + \sum_{j_1 \neq j_2} (1 - Z_{j_1})(1 - Z_{j_2}) \mathbb{E}_m [\hat{\epsilon}_{j_1}^2(\mathbf{0}) \hat{\epsilon}_{j_2}^2(\mathbf{0})], \end{aligned}$$

and

$$\begin{aligned} \text{Cov}_m \left[ \sum_{j=1}^m Z_j \hat{\epsilon}_j(\mathbf{1}), \sum_{j=1}^m (1 - Z_j) \hat{\epsilon}_j(\mathbf{0}) \right] &= \sum_{j=1}^m Z_j (1 - Z_j) \text{Cov}_m [\hat{\epsilon}_j(\mathbf{1}), \hat{\epsilon}_j(\mathbf{0})] \\ &+ \sum_{j_1 \neq j_2} Z_{j_1} (1 - Z_{j_2}) n_{j_1}^{-1} n_{j_2}^{-1} \sum_{i_1 \in \text{Inf}_{j_1}} \sum_{i_2 \in \text{Inf}_{j_2} \cap \mathcal{N}_{i_1}} \text{Cov}_m [\epsilon_{i_1}(\mathbf{1}), \epsilon_{i_2}(\mathbf{0})] G_{i_1, i_2} = 0, \end{aligned}$$

provided that  $Z_j(1 - Z_j) = 0$ , the exclusion of the nodes in  $\{i \in C_j : \forall i' \in C_{j'} \text{ with } j' \neq j, A_{i, i'} = 1, Z_j \neq Z_{j'}\}$ , and  $G_{i_1, i_2} = 0$  for  $i_1 \in \text{Inf}_{j_1}$  and  $i_2 \in \text{Inf}_{j_2}$  by Assumption 2.2. By Assumption 4.3.6, the conditional variance of  $R_{m,1,3}$  and  $R_{m,2,3}$  are

$$\begin{aligned} \mathbb{V}_m [R_{m,1,3}] &= \mathbb{V}_m [R_{m,2,3}] \\ &= \sum_{j=1}^m Z_j \mathbb{E}_m [\hat{\epsilon}_j^2(\mathbf{1})] + \sum_{j=1}^m (1 - Z_j) \mathbb{E}_m [\hat{\epsilon}_j^2(\mathbf{0})] \\ &\quad + \sum_{j_1 \neq j_2} Z_{j_1} Z_{j_2} \mathbb{E}_m [\hat{\epsilon}_{j_1}(\mathbf{1}) \hat{\epsilon}_{j_2}(\mathbf{1})] + \sum_{j_1 \neq j_2} (1 - Z_{j_1})(1 - Z_{j_2}) \mathbb{E}_m [\hat{\epsilon}_{j_1}(\mathbf{0}) \hat{\epsilon}_{j_2}(\mathbf{0})] \\ &= \sigma_{m, \text{Ind}}^2 + \sigma_{m, \text{Bet}}^2 = \sigma_{m, \epsilon}^2. \end{aligned}$$

Next we derive the bound for the Wasserstein distance. Let

$$\tilde{R}_{m,1,3}^{(-j)} = \sigma_{m,\epsilon}^{-1} \sum_{j' \notin \mathcal{N}(j)} \{Z_{j'} \hat{\epsilon}_{j'}(\mathbf{1}) + (1 - Z_{j'}) \hat{\epsilon}_{j'}(\mathbf{0})\}$$

and  $g_h(x)$  be the solution of the differential equation:

$$g'_h(x) - x g_h(x) = h(x) - \Phi(h),$$

where  $\Phi(x)$  is the cumulative distribution function of a standard normal random variable. By Assumption 4.3.4, it follows that  $\tilde{R}_{m,1,3}^{(-j)}$  and  $\{\hat{\epsilon}_j(\mathbf{1}), \hat{\epsilon}_j(\mathbf{0})\}$  are conditionally independent given  $\mathcal{Z}_m$  and thus

$$\begin{aligned} & \mathbb{E}_m \left[ \sigma_{m,\epsilon}^{-1} r_{j,1,3} \cdot g \left( \tilde{R}_{m,1,3}^{(-j)} \right) \right] = \mathbb{E}_m \left[ \sigma_{m,\epsilon}^{-1} \{Z_j \hat{\epsilon}_j(\mathbf{1}) + (1 - Z_j) \hat{\epsilon}_j(\mathbf{0})\} g \left( \tilde{R}_{m,1,3}^{(-j)} \right) \right] \\ & = \sigma_{m,\epsilon}^{-1} \{Z_j \mathbb{E}_m [\hat{\epsilon}_j(\mathbf{1})] + (1 - Z_j) \mathbb{E}_m [\hat{\epsilon}_j(\mathbf{0})]\} \mathbb{E}_m \left[ g \left( \tilde{R}_{m,1,3}^{(-j)} \right) \right] \\ & = 0. \end{aligned}$$

It immediately follows that

$$\begin{aligned} & \mathbb{E}_m [R_{m,1,3} \cdot g(R_{m,1,3})] \\ & = \mathbb{E}_m \left[ \sigma_{m,\epsilon}^{-1} \sum_{j=1}^m r_{j,1,3} \left\{ g(\tilde{R}_{m,1,3}) - g(\tilde{R}_{m,1,3}^{(-j)}) - (\tilde{R}_{m,1,3} - \tilde{R}_{m,1,3}^{(-j)}) g'(\tilde{R}_{m,1,3}) \right\} \right] \\ & + \mathbb{E}_m \left[ \sigma_{m,\epsilon}^{-1} \sum_{j=1}^m r_{j,1,3} (\tilde{R}_{m,1,3} - \tilde{R}_{m,1,3}^{(-j)}) g'(\tilde{R}_{m,1,3}) \right], \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E}_m \left[ \left| g'(\tilde{R}_{m,1,3}) - \tilde{R}_{m,1,3} g'(\tilde{R}_{m,1,3}) \right| \right] \\ & \leq \left| \mathbb{E}_m \left[ \frac{1}{\sigma_{m,\epsilon}} \sum_{j=1}^m r_{j,1,3} \left\{ g(\tilde{R}_{m,1,3}) - g(\tilde{R}_{m,1,3}^{(-j)}) - (\tilde{R}_{m,1,3} - \tilde{R}_{m,1,3}^{(-j)}) g'(\tilde{R}_{m,1,3}) \right\} \right] \right| \quad (20) \end{aligned}$$

$$+ \left| \mathbb{E} \left[ g'(\tilde{R}_{m,1,3}) \left( 1 - \frac{1}{\sigma_{m,\epsilon}} \sum_{j=1}^m r_{j,1,3} (\tilde{R}_{m,1,3} - \tilde{R}_{m,1,3}^{(-j)}) \right) \right] \right| \quad (21)$$

By Taylor expansion, Assumption 4.3.5, the triangle inequality, and the arithmetic-geometric mean inequality, (20) can be bounded above by

$$\begin{aligned} & \frac{\|g''\|}{2\sigma_{m,\epsilon}} \sum_{j=1}^m \mathbb{E}_m \left[ \left| r_{j,1,3} (\tilde{R}_{m,1,3} - \tilde{R}_{m,1,3}^{(-j)}) \right| \right] \\ & \leq \frac{1}{\sigma_{m,\epsilon}^3} \sum_{j_1=1}^m \sum_{j_2, j_3 \in \mathbf{N}(j_1)} \mathbb{E}_m |r_{j_1,1,3} r_{j_2,1,3} r_{j_3,1,3}| \\ & \leq \frac{\Psi_{\max, \text{CL}}^2}{\sigma_{m,\epsilon}^3} \sum_{j=1}^m \mathbb{E}_m \left[ |Z_j \hat{\epsilon}_j(\mathbf{1}) + (1 - Z_j) \hat{\epsilon}_j(\mathbf{0})|^3 \right] \\ & \leq \frac{\Psi_{\max, \text{CL}}^2}{\sigma_{m,\epsilon}^3} \sum_{j=1}^m \mathbb{E}_m \left[ |\hat{\epsilon}_j(\mathbf{1})|^3 + |\hat{\epsilon}_j(\mathbf{0})|^3 \right]. \quad (22) \end{aligned}$$

Following the same derivation of the proof of Theorem 3.5 of Ross (2011), (21) can be bounded above by

$$\begin{aligned} & \frac{\|g'\|}{\sigma_{m,\epsilon}^2} \mathbb{E}_m \left| \sigma_{m,\epsilon}^2 - \sum_{j_1=1}^m r_{j_1,1,3} \sum_{j_2 \in \mathbf{N}(j_1)} r_{j_2,1,3} \right| \leq \frac{\sqrt{26}\Psi_{\max,\text{CL}}^{3/2}}{\sqrt{\pi}\sigma_{m,\epsilon}^2} \sqrt{\sum_{j=1}^m \mathbb{E} [r_{j,1,3}^4]} \\ & \leq \frac{\sqrt{26}\Psi_{\max,\text{CL}}^{3/2}}{\sqrt{\pi}\sigma_{m,\epsilon}^2} \sqrt{\sum_{j=1}^m \mathbb{E} [\hat{\epsilon}_j^4(\mathbf{1}) + \hat{\epsilon}_j^4(\mathbf{0})]}, \end{aligned} \quad (23)$$

where the last inequality follows from the fact that  $Z_j(1 - Z_j) = 0$ . Then, plugging (22) and (23) into (20) and (21), respectively, yields the first part of (19).

Moreover, Assumption 4.3.4–4.3.7 imply that

$$\begin{aligned} & \frac{\Psi_{\max}^2}{\sigma_{m,\epsilon}^3} \sum_{j=1}^m \mathbb{E}_m \left[ |\hat{\epsilon}_j(\mathbf{1})|^3 + |\hat{\epsilon}_j(\mathbf{0})|^3 \right] \asymp m^{2\lambda_3 + \lambda_1 - \frac{3\lambda_2}{2}}, \\ & \frac{\sqrt{26}\Psi_{\max}^{3/2}}{\sqrt{\pi}\sigma_{m,\epsilon}^2} \sqrt{\sum_{j=1}^m \mathbb{E}_m \left[ \hat{\epsilon}_j^4(\mathbf{1}) + \hat{\epsilon}_j^4(\mathbf{0}) \right]} \asymp m^{\frac{3\lambda_3}{2} + \frac{\lambda_1}{2} - \lambda_2}. \end{aligned}$$

Based on Assumption 4.3.7, when  $\lambda_2 < 2\lambda_1$ ,  $\lambda_3 = \frac{3\lambda_2 - 2\lambda_1}{4} - \lambda_4$ , and

$$m^{2\lambda_3 + \lambda_1 - \frac{3\lambda_2}{2}} + m^{\frac{3\lambda_3}{2} + \frac{\lambda_1}{2} - \lambda_2} = m^{-2\lambda_4} + m^{-\frac{3\lambda_4}{2} - \frac{2\lambda_1 - \lambda_2}{8}};$$

when  $\lambda_2 > 2\lambda_1$ ,  $\lambda_3 = \frac{2\lambda_2 - \lambda_1}{3} - \lambda_4$ , and

$$m^{2\lambda_3 + \lambda_1 - \frac{3\lambda_2}{2}} + m^{\frac{3\lambda_3}{2} + \frac{\lambda_1}{2} - \lambda_2} = m^{-\frac{\lambda_2 - 2\lambda_1}{6} - 2\lambda_4} + m^{-\frac{3\lambda_4}{2}};$$

and when  $\lambda_2 = 2\lambda_1$ ,  $\lambda_3 = \lambda_1 - \lambda_4$ , and

$$m^{2\lambda_3 + \lambda_1 - \frac{3\lambda_2}{2}} + m^{\frac{3\lambda_3}{2} + \frac{\lambda_1}{2} - \lambda_2} = m^{-2\lambda_4} + m^{-\frac{3\lambda_4}{2}}.$$

Then there exists  $a > 0$  such that (19) holds for  $\tilde{R}_{m,1,3}$ . In a similar manner, we can also show that (19) holds for  $\tilde{R}_{m,2,3}$ . Finally, the asymptotic distribution of  $\tilde{R}_{m,k,3}$  is an immediate consequence of (19). This completes the proof for this lemma.  $\blacksquare$

Note that the above lemma is not affected by the randomization procedure, it holds for both CR, and CLAR. Based on the above lemmas, we derive the asymptotic independence of  $(R_{m,1,1}, R_{m,1,2}, R_{m,1,3})$  and the asymptotic independence of  $(R_{m,2,1}, R_{m,2,2}, R_{m,2,3})$  as follows.

**Lemma A.7** *Suppose the Assumptions of Lemmas A.4 and A.5 hold, then*

$$\begin{aligned} & (m^{-1/2}R_{m,1,1}, m^{-1/2}R_{m,1,2}, m^{-\lambda_2/2}R_{m,1,3})^\top \xrightarrow{\mathbf{D}} (\zeta_{1,1}, \zeta_{1,2}, \zeta_{1,3})^\top \\ & \text{and } (m^{-1/2}R_{m,2,1}, m^{-1/2}R_{m,2,2}, m^{-\lambda_2/2}R_{m,2,3})^\top \xrightarrow{\mathbf{D}} (\zeta_{2,1}, \zeta_{2,2}, \zeta_{2,3})^\top, \end{aligned}$$

where  $\zeta_{1,1}$ ,  $\zeta_{1,2}$ , and  $\zeta_{1,3}$  are independent satisfying  $\zeta_{1,1} \sim \mathcal{N}(0, \sigma_{\text{CLAR}}^2)$ ,  $\zeta_{1,2} \sim \mathcal{N}(0, \sigma_f^2)$  and  $\zeta_{1,3} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ ; and  $\zeta_{2,1}$ ,  $\zeta_{2,2}$ , and  $\zeta_{2,3}$  are independent satisfying  $\zeta_{2,1} \sim \mathcal{N}(0, \tilde{\sigma}_{\text{CLAR}}^2)$ ,  $\zeta_{2,2} \sim \mathcal{N}(0, \tilde{\sigma}_f^2)$  and  $\zeta_{2,3} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ .

**Proof** The derivation for  $(m^{-1/2}R_{m,2,1}, m^{-1/2}R_{m,2,2}, m^{-\lambda_2/2}R_{m,2,3})^\top$  is analogous to that for  $(m^{-1/2}R_{m,1,1}, m^{-1/2}R_{m,1,2}, m^{-\lambda_2/2}R_{m,1,3})^\top$ , and is omitted. For simplicity of notation, let  $R_{m,1,1}^\dagger = m^{-1/2}R_{m,1,1}$ ,  $R_{m,1,2}^\dagger = m^{-1/2}R_{m,1,2}$ , and  $R_{m,1,3}^\dagger = m^{-\lambda_2/2}R_{m,1,3}$ . In Lemmas A.4, A.5, and A.6, we have derived the marginal normality of  $m^{-1/2}R_{m,1,1}$ ,  $m^{-1/2}R_{m,1,2}$ , and  $m^{-\lambda_2/2}R_{m,1,3}$ , for CR and CLAR. It suffices to show that the three components are asymptotically independent. Recall that  $\mathbb{E}_m[\cdot] = \mathbb{E}[\cdot | \tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m]$ , then for  $a_1, a_2, a_3 \in \mathbb{R}^3$ , we have

$$\begin{aligned} & \mathbf{P} \left( R_{m,1,1}^\dagger \leq a_1, R_{m,1,2}^\dagger \leq a_2, R_{m,1,3}^\dagger \leq a_3 \right) \\ &= \mathbb{E} \left[ \mathbb{I} \left\{ R_{m,1,1}^\dagger \leq a_1 \right\} \mathbb{I} \left\{ R_{m,1,2}^\dagger \leq a_2 \right\} \right. \\ & \quad \left. \times \left( \mathbb{E}_m \left[ \mathbb{I} \left\{ R_{m,1,3}^\dagger \leq a_3 \right\} \right] - \mathbf{P}_m \left( \zeta_3 \leq \sigma_{m,\epsilon}^{-1} m^{\frac{\lambda_2}{2}} \cdot a_3 \right) \right) \right] \end{aligned} \quad (24)$$

$$\begin{aligned} & + \mathbb{E} \left[ \mathbb{I} \left\{ R_{m,1,1}^\dagger \leq a_1 \right\} \mathbb{I} \left\{ R_{m,1,2}^\dagger \leq a_2 \right\} \right. \\ & \quad \left. \times \mathbf{P}_m \left( \zeta_3 \leq \sigma_{m,\epsilon}^{-1} m^{\frac{\lambda_2}{2}} \cdot a_3 \right) \mathbb{I} \left\{ \left| m^{-\lambda_2} \sigma_{m,\epsilon}^2 - \sigma_\epsilon^2 \right| > \epsilon \right\} \right] \end{aligned} \quad (25)$$

$$\begin{aligned} & + \mathbb{E} \left[ \mathbb{I} \left\{ R_{m,1,1}^\dagger \leq a_1 \right\} \mathbb{I} \left\{ R_{m,1,2}^\dagger \leq a_2 \right\} \left\{ \mathbf{P}_m \left( \zeta_3 \leq \sigma_{m,\epsilon}^{-1} m^{\frac{\lambda_2}{2}} \cdot a_3 \right) - \mathbf{P} \left( \sigma_\epsilon \xi_3 \leq a_3 \right) \right\} \right. \\ & \quad \left. \times \mathbb{I} \left\{ \left| m^{-\lambda_2} \sigma_{m,\epsilon}^2 - \sigma_\epsilon^2 \right| \leq \epsilon \right\} \right] \end{aligned} \quad (26)$$

$$+ \mathbb{E} \left[ \mathbb{I} \left\{ R_{m,1,1}^\dagger \leq a_1 \right\} \mathbb{I} \left\{ R_{m,1,2}^\dagger \leq a_2 \right\} \mathbb{I} \left\{ \left| m^{-\lambda_2} \sigma_{m,\epsilon}^2 - \sigma_\epsilon^2 \right| \leq \epsilon \right\} \right] \mathbf{P} \left( \sigma_\epsilon \xi_3 \leq a_3 \right) \quad (27)$$

$$\rightarrow \mathbf{P} \left( \xi_{1,1} \leq a_1 \right) \mathbf{P} \left( \xi_{1,2} \leq a_2 \right) \mathbf{P} \left( \xi_{1,3} \leq a_3 \right), \quad (28)$$

where  $\xi_3 \sim \mathcal{N}(0, 1)$  and  $\mathbf{P}_m(\cdot < a)$  is the conditional CDF given  $\tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m$ . Conditioning on  $\tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m$ ,  $\mathbb{I} \left\{ R_{m,1,1}^\dagger \leq a_1 \right\}$  and  $\mathbb{I} \left\{ R_{m,1,2}^\dagger \leq a_2 \right\}$  are fixed. (24) converges to zero by dominated convergence theorem and Lemma A.6. In addition,  $\mathbf{P}_m \left( \left| m^{-\lambda_2} \sigma_{m,\epsilon}^2 - \sigma_\epsilon^2 \right| > \epsilon \right)$  converges to zero and  $\mathbf{P}_m \left( \left| m^{-\lambda_2} \sigma_{m,\epsilon}^2 - \sigma_\epsilon^2 \right| \leq \epsilon \right)$  converges to one by 6. of Assumption 4.3. Therefore, (25) also converges to zero. Furthermore, because  $m^{-\lambda_2} \sigma_{m,\epsilon}^2$  converges to  $\sigma_\epsilon^2$  in probability given  $\tilde{\mathcal{F}}_0, \mathcal{Z}_m, \Xi_m$ , it follows that

$$\mathbf{P}_m \left( \zeta_3 \leq \sigma_{m,\epsilon}^{-1} m^{\frac{\lambda_2}{2}} a_3 \right) \rightarrow \mathbf{P} \left( \sigma_\epsilon \xi_3 \leq a_3 \right).$$

Then (26) converges to zero by the dominated convergence theorem. Finally, (27) converges to (28) by Lemma A.4 or Lemma A.5 and  $\mathbf{P}_m \left( \left| m^{-\lambda_2} \sigma_{m,\epsilon}^2 - \sigma_\epsilon^2 \right| \leq \epsilon \right)$  converges to one. The proof of this lemma is now complete.  $\blacksquare$

**Lemma A.8** *Under the Assumptions of Lemmas A.4 and A.5,  $(m\sigma_{m,\text{CAE}})^{-1}R_{m,1} \xrightarrow{\mathbf{D}} \mathcal{N}(0,1)$ , and  $(m\tilde{\sigma}_{m,\text{CAE}})^{-1}R_{m,2} \xrightarrow{\mathbf{D}} \mathcal{N}(0,1)$ , where  $\tilde{\sigma}_{m,\text{CAE}}^2 = m^{-1}(\sigma_{\text{Design}}^2 + \sigma_f^2 + 4m^{\lambda_2-1}\sigma_\epsilon^2)$ .*

**Proof** We prove Lemma A.8 by taking the value of  $\lambda_2$  into consideration for  $R_{m,1}$ . The proof for  $R_{m,2}$  is similar and is omitted.

First, when  $\lambda_2 = 1$ , it follows from Lemmas A.4, A.5, and A.6 that

$$(m\sigma_{m,\text{CAE}})^{-1}R_{m,1} = (\sigma_{\text{Design}}^2 + \sigma_f^2 + 4\sigma_\epsilon^2)^{-1/2}m^{-1/2}R_{m,1} \xrightarrow{\mathbf{D}} \mathcal{N}(0,1).$$

Second, if  $\lambda_2 < 1$ , Lemmas A.4, A.5, and A.6 imply that

$$\begin{aligned} \frac{\sigma_{\text{Design}}^2 + \sigma_f^2}{\sigma_{\text{Design}}^2 + \sigma_f^2 + 4m^{\lambda_2-1}\sigma_\epsilon^2} &\xrightarrow{a.s.} 1, \\ \frac{R_{m,1,1} + R_{m,1,2}}{\sqrt{m(\sigma_{\text{Design}}^2 + \sigma_f^2)}} &\xrightarrow{\mathbf{D}} \mathcal{N}(0,1), \\ \text{and } \frac{R_{m,1,3}}{\sqrt{m(\sigma_{\text{Design}}^2 + \sigma_f^2)}} &= o_p(1). \end{aligned}$$

since  $\mathbb{V}[m^{-1/2}R_{m,1,3}] = o(1)$ . These results further imply that

$$\begin{aligned} (m\sigma_{m,\text{CAE}})^{-1}R_{m,1} &= (\sigma_{\text{Design}}^2 + \sigma_f^2 + 4m^{\lambda_2-1}\sigma_\epsilon^2)^{-1}m^{-1/2}(R_{m,1,1} + R_{m,1,2} + R_{m,1,3}) \\ &= \sqrt{\frac{\sigma_{\text{Design}}^2 + \sigma_f^2}{\sigma_{\text{Design}}^2 + \sigma_f^2 + 4m^{\lambda_2-1}\sigma_\epsilon^2}} \frac{R_{m,1,1} + R_{m,1,2} + 2R_{m,1,3}}{\sqrt{m(\sigma_{\text{Design}}^2 + \sigma_f^2)}} \\ &\xrightarrow{\mathbf{D}} \mathcal{N}(0,1). \end{aligned}$$

Finally, when  $1 < \lambda_2 < 2$ , we have

$$\begin{aligned} \frac{4m^{\lambda_2-1}\sigma_\epsilon^2}{\sigma_{\text{Design}}^2 + \sigma_f^2 + 4m^{\lambda_2-1}\sigma_\epsilon^2} &\xrightarrow{a.s.} 1, \\ \frac{R_{m,1,1} + R_{m,1,2}}{\sqrt{4m^{\lambda_2}\sigma_\epsilon^2}} &= o_p(1) \\ \text{and } \frac{R_{m,1,3}}{\sqrt{m^{\lambda_2}\sigma_\epsilon^2}} &\xrightarrow{\mathbf{D}} \mathcal{N}(0,1). \end{aligned}$$

It follows that

$$\begin{aligned} (m\sigma_{m,\text{CAE}})^{-1}R_{m,1} &= (\sigma_{\text{Design}}^2 + \sigma_f^2 + 4m^{\lambda_2-1}\sigma_\epsilon^2)^{-1}m^{-1/2}(R_{m,1,1} + R_{m,1,2} + R_{m,1,3}) \\ &= \sqrt{\frac{4m^{\lambda_2-1}\sigma_\epsilon^2}{\sigma_{\text{Design}}^2 + \sigma_f^2 + 4m^{\lambda_2-1}\sigma_\epsilon^2}} \frac{R_{m,1,1} + R_{m,1,2} + 2R_{m,1,3}}{\sqrt{4m^{\lambda_2}\sigma_\epsilon^2}} \\ &\xrightarrow{\mathbf{D}} \mathcal{N}(0,1). \end{aligned}$$

The proof for Lemma A.8 is completed. ■

## A.2.1 PROOF FOR THEOREM 4.2

We first derive the consistency of CAE. Notice that  $\lambda_2 < 2$  by 6. of Assumption 4.3, it follows that

$$\sigma_{m,\text{CAE}}^2 = m^{-1}\{\sigma_{\text{Design}}^2 + \sigma_{\mathbf{f}}^2\} + m^{\lambda_2-2}\sigma_{\epsilon}^2 \rightarrow 0,$$

and thus  $m^{-1}R_{m,1}$  converges to zero in  $\mathcal{L}_2$ . Furthermore, 7. of Assumption 4.3 indicates that  $B_{m,2} \lesssim \Psi_{\text{Diff,Inf}} = o(1)$ . Putting the two pieces together yields

$$\hat{\tau}_{\text{CAE}} - \tau(\mathbf{1}, \mathbf{0}) = m^{-1}R_{m,1} + B_{m,2} + o_p(1) = o_p(1).$$

Moreover, it follows from Lemma A.7 that

$$\begin{aligned} \sigma_{m,\text{CAE}}^{-1}B_{m,2} &\lesssim \sigma_{m,\text{CAE}}^{-1}\Psi_{\text{Diff,Inf}} \\ &= (\sigma_{\text{Design}}^2 + \sigma_{\mathbf{f}}^2 + 4m^{\lambda_2-1}\sigma_{\epsilon}^2)^{-1/2}m^{1/2}\Psi_{\text{Diff,Inf}}. \end{aligned}$$

When  $\lambda_2 \leq 1$ ,  $\max\{m^{-1/2}, m^{-(2-\lambda_2)/2}\} = m^{-1/2}$  and 6. of Assumption 4.3 yields  $\sigma_{m,\text{CAE}}^{-1}B_{m,2} \lesssim m^{1/2}\Psi_{\text{Diff,Inf}} = o_p(1)$ . When  $\lambda_2 > 1$ ,  $\max\{m^{-1/2}, m^{-(2-\lambda_2)/2}\} = m^{-(2-\lambda_2)/2}$  and similarly  $\sigma_{m,\text{CAE}}^{-1}B_{m,2} \lesssim [m^{-(\lambda_2-1)}(\sigma_{\text{Design}}^2 + \sigma_{\mathbf{f}}^2) + 4\sigma_{\epsilon}^2]^{-1/2}m^{(2-\lambda_2)/2}\Psi_{\text{Diff,Inf}} = o_p(1)$ . It follows from (17) and Lemma A.8 that

$$\begin{aligned} \sigma_{m,\text{CAE}}^{-1}\{\hat{\tau}_{\text{CAE}} - \tau(\mathbf{1}, \mathbf{0})\} &= (m\sigma_{m,\text{CAE}})^{-1}R_{m,1} + o_p(1) \\ &\xrightarrow{\mathbf{D}} \mathcal{N}(0, 1). \end{aligned}$$

This completes the proof for Theorem 4.2.

## A.2.2 PROOF FOR COROLLARY 4.1

Notice that the model considered in Corollary 4.1 satisfies Assumption 4.3. Moreover,  $R_{m,1,1}$ ,  $R_{m,1,2}$ , and  $R_{m,1,3}$  can be written as

$$\begin{aligned} R_{m,1,1} &= 2 \sum_{j=1}^m (2Z_j - 1) \hat{f}_j(\mathbf{X}_j) \\ &= 2 \sum_{j=1}^m (2Z_j - 1) \{(\mathbf{X}_{j,\text{CL}} - \boldsymbol{\mu}_{\text{CL}})^\top \boldsymbol{\beta}_{\text{CL}} + (\hat{\mathbf{X}}_{j,\text{IN}} - \boldsymbol{\mu}_{\text{IN}})^\top \boldsymbol{\beta}_{\text{IN}}\}, \end{aligned}$$

$R_{m,1,3} = \sum_{j=1}^m (2Z_j - 1) \hat{\epsilon}_j$ , and  $R_{m,1,2} = 0$ .

Under CR, simple calculations and 1. and 2. of Assumption 4.1 yield

$$\begin{aligned} \mathbb{V}[m^{-1/2}R_{m,1,1}|\mathcal{A}, \tilde{\mathbf{C}}] &= 4 \left\{ \mathbb{V}[\mathbf{X}_{\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}}] + \sum_{j=1}^m n_j^{-1} \mathbb{V}[\mathbf{X}_{\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}}] \right\} \\ &= 4 \left\{ \mathbb{V}[\mathbf{X}_{\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}}] + \tilde{\lambda}_{\text{IN}} \mathbb{V}[\mathbf{X}_{\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}}] \right\} + o(1), \\ \mathbb{V}[m^{-1/2}R_{m,1,3}|\mathcal{A}, \tilde{\mathbf{C}}] &= \sigma_e^2 \sum_{j=1}^m n_j^{-1} = \tilde{\lambda}_{\text{IN}} \sigma_e^2 + o(1), \end{aligned}$$

provided that  $\{Z_j\}_{j=1}^m$  and  $\{\mathbf{X}_{j,\text{CL}}, \hat{\mathbf{X}}_{j,\text{IN}}, \hat{\epsilon}_j\}_{j=1}^m$  are independent. It then follows from Theorem 4.2 that  $\sigma_{\hat{\epsilon}}^2 = 0$ ,

$$\begin{aligned}\sigma_{\text{Design}}^2 &= 4 \left\{ \mathbb{V}[\mathbf{X}_{\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}}] + \tilde{\lambda}_{\text{IN}} \mathbb{V}[\mathbf{X}_{\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}}] \right\}, \\ \sigma_{m,\epsilon}^2 &= m \tilde{\lambda}_{\text{IN}} \sigma_e^2,\end{aligned}$$

and hence  $\lambda_2 = 1$  and  $\sigma_\epsilon^2 = \tilde{\lambda}_{\text{IN}} \sigma_e^2$ . Note that in all the randomization schemes, the convergence rate of  $R_{m,1,3}$  is the same. Therefore, the rate of convergence of CAE under the outcome model assumed in Corollary 4.1 is  $m^{1/2}$ .

Next, suppose  $\boldsymbol{\xi}_j = \mathbf{X}_{j,\text{CL}}$  is used in CLAR. It follows from Lemma A.1 that

$$\begin{aligned}m^{-1/2} \sum_{j=1}^m (2Z_j - 1)(\boldsymbol{\xi}_j - \boldsymbol{\mu}_\xi) &= m^{-1/2} \sum_{j=1}^m (2Z_j - 1)(\mathbf{X}_{j,\text{CL}} - \boldsymbol{\mu}_{\text{CL}}) \\ &= O_p(m^{-1/2}) = o_p(1),\end{aligned}$$

and hence

$$m^{-1/2} R_{m,1} = m^{-1/2} (\tilde{R}_{m,1,1}^1 + R_{m,1,3}) + o_p(1),$$

where  $\tilde{R}_{m,1,1}^1 = 2 \sum_{j=1}^m (2Z_j - 1)(\hat{\mathbf{X}}_{j,\text{IN}} - \boldsymbol{\mu}_{\text{IN}})^\top \boldsymbol{\beta}_{\text{IN}}$ . When  $\boldsymbol{\xi}_j = \mathbf{X}_{j,\text{CL}}$ , because  $\{\mathbf{X}_{j,\text{CL}}\}_{j=1}^m$  and  $\{\mathbf{X}_{i,\text{IN}}\}_{i=1}^n$  are independent,  $\{\mathbf{X}_{i,\text{IN}}\}_{i=1}^n$  and  $\{Z_j\}_{j=1}^m$  are also independent. This implies that

$$\mathbb{V}[m^{-1/2} \tilde{R}_{m,1,1}^1 | \tilde{\mathcal{C}}, \mathcal{A}] = \tilde{\lambda}_{\text{IN}} \mathbb{V}[\mathbf{X}_{\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}}] + o(1),$$

and hence  $\sigma_{\text{Design}}^2 = 4 \tilde{\lambda}_{\text{IN}} \mathbb{V}[\mathbf{X}_{\text{IN}}^\top \boldsymbol{\beta}_{\text{IN}}]$  and  $\sigma_{m,\epsilon}^2 = m \tilde{\lambda}_{\text{IN}} \sigma_e^2$ .

Assume that  $\boldsymbol{\xi}_j = \bar{\mathbf{X}}_{j,\text{IN}}$  is used in CLAR, then Lemma A.1 implies that

$$\begin{aligned}m^{-1/2} \sum_{j=1}^m (2Z_j - 1)(\boldsymbol{\xi}_j - \boldsymbol{\mu}_\xi) &= m^{-1/2} \sum_{j=1}^m (2Z_j - 1)(\bar{\mathbf{X}}_{j,\text{IN}} - \boldsymbol{\mu}_{\text{IN}}) \\ &= O_p(m^{-1/2}) = o_p(1),\end{aligned}$$

and

$$m^{-1/2} R_{m,1} = m^{-1/2} (\tilde{R}_{m,1,1}^2 + R_{m,1,3}) + o_p(1),$$

where  $\tilde{R}_{m,1,1}^2 = 2 \sum_{j=1}^m (2Z_j - 1) \{ (\mathbf{X}_{j,\text{CL}} - \boldsymbol{\mu}_{\text{CL}})^\top \boldsymbol{\beta}_{\text{CL}} + (\hat{\mathbf{X}}_{j,\text{IN}} - \bar{\mathbf{X}}_{j,\text{IN}})^\top \boldsymbol{\beta}_{\text{IN}} \}$ . By the proof of Lemma A.4, there exists  $\sigma_{\text{diff}}^2 > 0$  such that

$$\mathbb{V} \left[ \sum_{j=1}^m (2Z_j - 1) (\hat{\mathbf{X}}_{j,\text{IN}} - \bar{\mathbf{X}}_{j,\text{IN}})^\top \boldsymbol{\beta}_{\text{IN}} \middle| \tilde{\mathcal{C}}, \mathcal{A} \right] = m \sigma_{\text{diff}}^2 + o(1).$$

The independence between  $\{\mathbf{X}_{j,\text{CL}}\}_{j=1}^m$  and  $\{\mathbf{X}_{i,\text{IN}}\}_{i=1}^n$  implies that under CLAR with  $\boldsymbol{\xi}_j = \bar{\mathbf{X}}_{j,\text{IN}}$ , we have

$$\mathbb{V} \left[ \sum_{j=1}^m (2Z_j - 1) \{ (\mathbf{X}_{j,\text{CL}} - \boldsymbol{\mu}_{\text{CL}})^\top \boldsymbol{\beta}_{\text{CL}} \} \middle| \tilde{\mathcal{C}}, \mathcal{A} \right] = m \mathbb{V}[\mathbf{X}_{\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}}].$$

Therefore,  $\sigma_{\text{Design}}^2 = 4\{\mathbb{V}[\mathbf{X}_{\text{CL}}^\top \boldsymbol{\beta}_{\text{CL}}] + \sigma_{\text{diff}}^2\}$  and  $\sigma_{m,\epsilon}^2 = m\tilde{\lambda}_{\text{IN}}\sigma_e^2$ .

Finally, when  $\boldsymbol{\xi}_j = (\mathbf{X}_{j,\text{CL}}^\top, \bar{\mathbf{X}}_{j,\text{IN}}^\top)^\top$  are used in CLAR, Lemma A.1 implies that

$$m^{-1/2} \sum_{j=1}^m (2Z_j - 1)(\mathbf{X}_{j,\text{CL}} - \boldsymbol{\mu}_{\text{CL}}) = o_p(1),$$

and  $m^{-1/2} \sum_{j=1}^m (2Z_j - 1)(\mathbf{X}_{j,\text{IN}} - \boldsymbol{\mu}_{\text{IN}}) = o_p(1),$

hold simultaneously. This further implies that

$$m^{-1/2} R_{m,1} = m^{-1/2} (\tilde{R}_{m,1,1}^3 + R_{m,1,3}),$$

where  $R_{m,1,1}^3 = \sum_{j=1}^m (2Z_j - 1)(\hat{\mathbf{X}}_{j,\text{IN}} - \bar{\mathbf{X}}_{j,\text{IN}})^\top \boldsymbol{\beta}_{\text{IN}}$ . Then it follows from Lemma A.4 that  $\sigma_{\text{Design}}^2 = 4\sigma_{\text{diff}}^2$  and  $\sigma_{m,\epsilon}^2 = m\tilde{\lambda}_{\text{IN}}\sigma_e^2$ . This completes the proof for this corollary.

## Appendix B. Additional Results for Stochastic Block Model

In this section, we present additional simulation studies to further demonstrate the property of our proposed procedure with the stochastic block model. The cluster sizes  $\{c_j\}_{j=1}^m$  are also generated from the discrete power-law distribution, with a parameter 4, and the parameter representing the minimum value of a cluster, 12. Then, we generate Rényi random graphs with  $0.4 \times n$  edges for the within cluster edges and  $r \times n$  edges for the between cluster edges. Here,  $r \in \{0.2, 0.4, \dots, 2\}$  is the prespecified *reconnecting probability* characterizing the portion of edges that connects different clusters.

The outcome model (6) and the parameter setting used in Section 5 are also considered in this section. In the randomization step, we also compare the four randomization schemes: 1) complete randomization (CR); 2) CLAR with the cluster-level covariates (CLAR-CL),  $\boldsymbol{\xi}_j = \mathbf{X}_{j,\text{CL}}$  (CL); 3) CLAR with the individual-level covariates (CLAR-Ind),  $\boldsymbol{\xi}_j = \bar{\mathbf{X}}_{j,\text{IN}}$  (IN); and 4) CLAR with both cluster-level and individual-level covariates,  $\boldsymbol{\xi}_j = (\mathbf{X}_{j,\text{CL}}, \bar{\mathbf{X}}_{j,\text{IN}}^\top)^\top$  (Both). In the estimation step, CAE is compared with the difference-in-means (DIM) estimator.

The performance of the four randomization schemes on the balance of covariates is evaluated in Section B.1. We further compare the performance of different network A/B testing approaches in Section B.2.

### B.1 Evaluation of Covariates Balance under the Four Randomization Schemes

The balance properties of the four randomization schemes are evaluated from the following two aspects. We first assess the Mahalanobis distance in Figure 8. The marginal covariate balance represented by the difference in covariate means is presented in Figure 9.

CLUSTER-ADAPTIVE NETWORK A/B TESTING

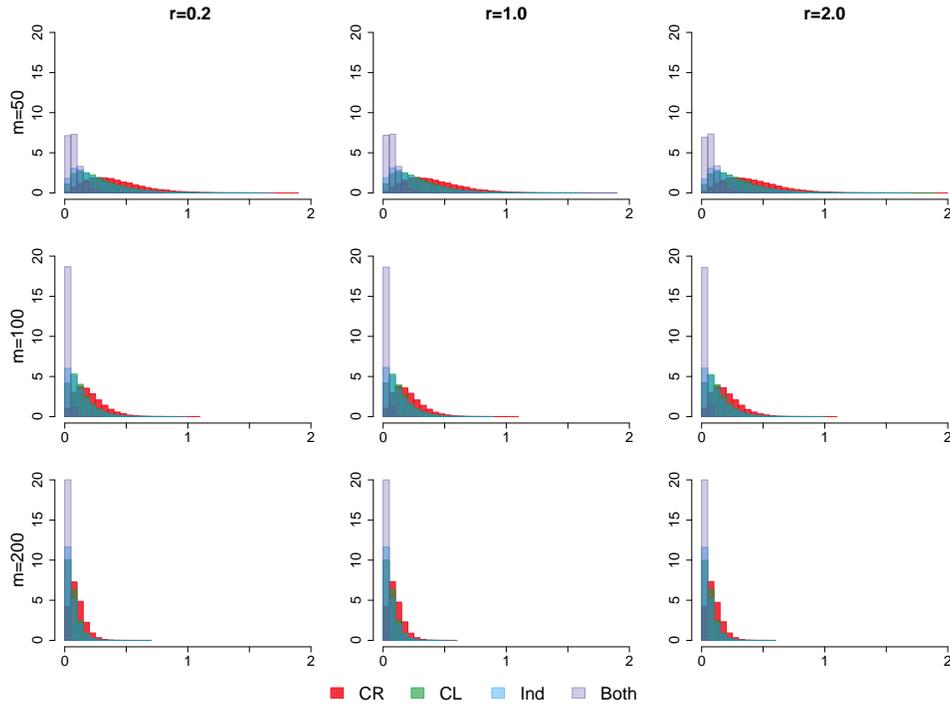


Figure 8: Histogram of the Mahalanobis distance  $M_m$  under different randomization schemes and  $r \in \{0.2, 1.0, 2.0\}$ .

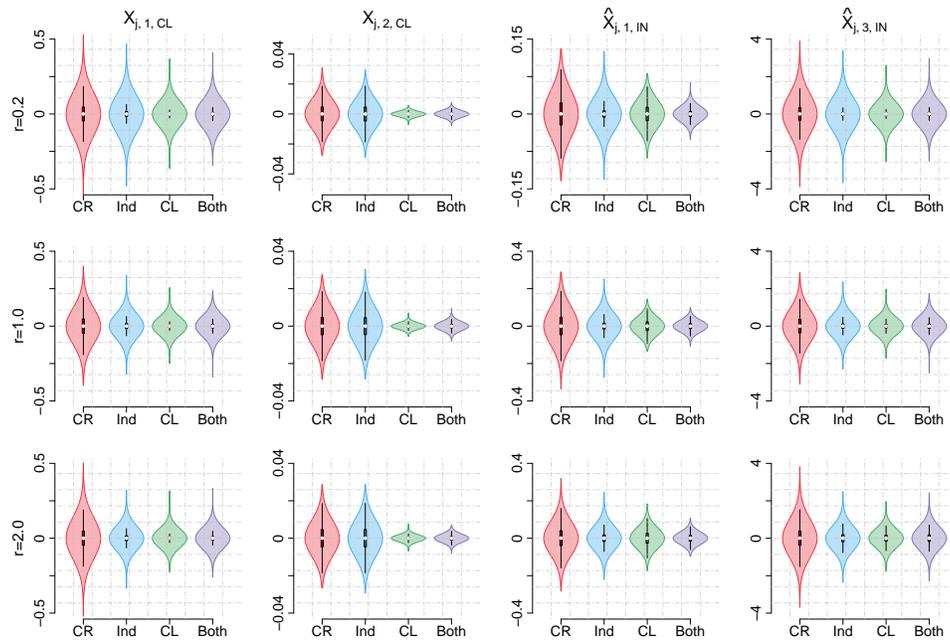


Figure 9: Violin plots of the difference-in-covariate-means for  $X_{j,1,CL}$ ,  $X_{j,2,CL}$ ,  $\hat{X}_{j,1,CL}$ , and  $\hat{X}_{j,3,IN}$  under  $m = 200$ .

The conclusions drawn from Figures 8 and 9 are consistent with Section 5.1. The Mahalanobis distance under CLAR-Both is the most concentrated around zero. The distributions of the Mahalanobis distance under CLAR-CL and CLAR-Ind are more spread out than the distribution under CLAR-both but are more concentrated around zero than the distribution under CR. Furthermore, the use of CLAR generally improves the covariate imbalance of  $(X_{j,1,CL}, X_{j,2,CL}, \hat{X}_{j,1,IN}, \hat{X}_{j,1,IN})^\top$ . The distributions of the difference in covariate means of the four random variables are more concentrated at zero under CLAR.

## B.2 Comparison of Different Network A/B Testing Approaches

Table 4: Evaluation of treatment effect with different network A/B testing approaches.

$m$	Design	$r = 0.2$						$r = 1.8$					
		$\hat{\tau}_{DIM}$			$\hat{\tau}_{CAE}$			$\hat{\tau}_{DIM}$			$\hat{\tau}_{CAE}$		
		Bias	SD	RMSE									
50	CR	-0.105	8.110	8.111	0.001	1.509	1.509	-0.692	4.746	4.797	-0.007	1.805	1.806
	CL	-0.209	7.768	7.771	-0.012	0.950	0.951	-0.697	4.306	4.363	0.001	1.165	1.165
	Ind	-0.131	7.888	7.889	-0.001	1.136	1.136	-0.706	4.383	4.440	-0.003	1.306	1.306
	Both	-0.177	7.738	7.740	-0.008	1.061	1.061	-0.654	4.362	4.411	0.006	1.242	1.242
100	CR	-0.107	5.756	5.757	-0.001	0.989	0.989	-0.669	4.803	4.849	-0.000	1.323	1.323
	CL	-0.116	5.497	5.499	-0.004	0.533	0.533	-0.669	4.366	4.417	-0.002	0.789	0.789
	Ind	-0.123	5.554	5.555	-0.003	0.659	0.659	-0.691	4.451	4.504	-0.001	0.920	0.920
	Both	-0.163	5.491	5.493	-0.005	0.587	0.587	-0.675	4.370	4.422	-0.001	0.826	0.826
200	CR	-0.075	5.300	5.300	-0.001	0.709	0.709	-0.661	4.942	4.986	-0.006	0.944	0.944
	CL	-0.132	5.044	5.046	-0.003	0.339	0.339	-0.634	4.610	4.653	0.001	0.531	0.531
	Ind	-0.110	5.060	5.061	-0.002	0.449	0.449	-0.650	4.677	4.722	-0.003	0.630	0.630
	Both	-0.098	4.998	4.999	-0.000	0.363	0.363	-0.694	4.587	4.639	-0.003	0.546	0.546

The finite sample properties of different A/B testing approaches are evaluated in Table 4. In particular, we evaluate the biases of DIM estimator and CAE in Figure 10. The standard deviation and the distribution of CAE are further evaluated in Figures 11 and 12, respectively. We also report the average fraction of the samples included in CAE in Table 5.

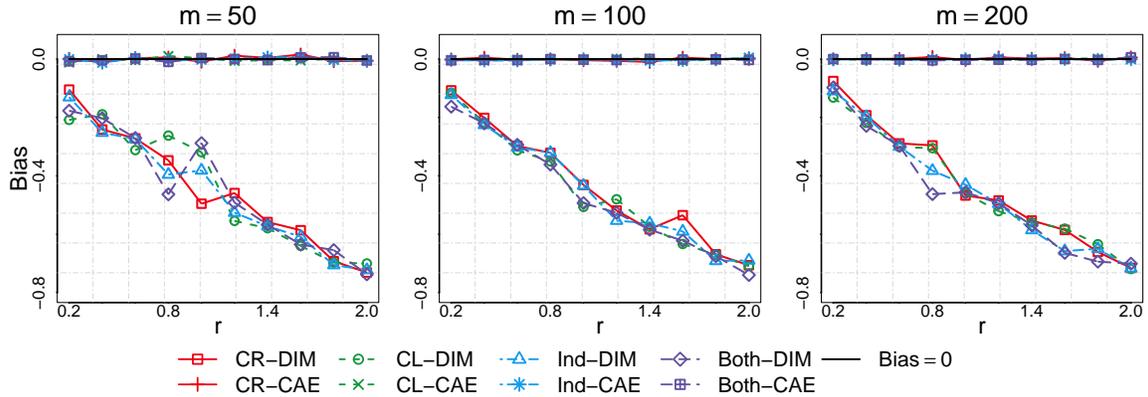


Figure 10: Bias for evaluating the ATE under different network A/B testing approaches.

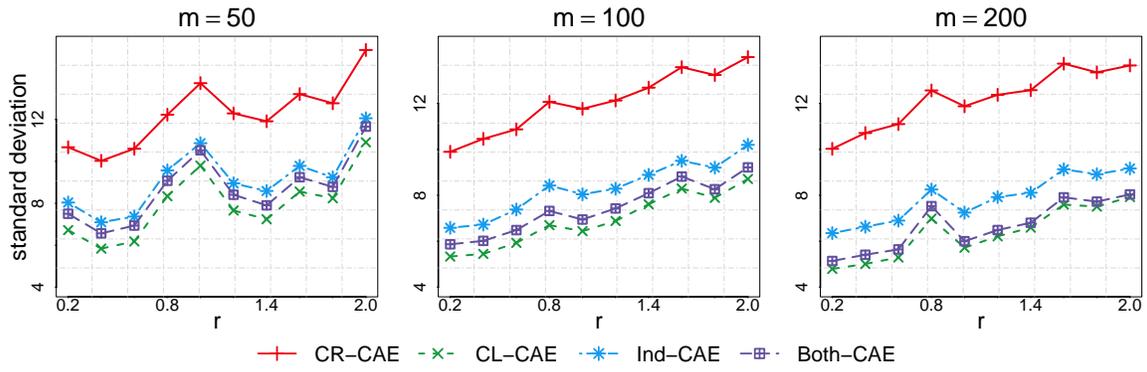


Figure 11: Standard deviation of  $\sqrt{m^*}(\hat{\tau}_{\text{CAE}} - \tau(\mathbf{1}, \mathbf{0}))$  under four randomization schemes.

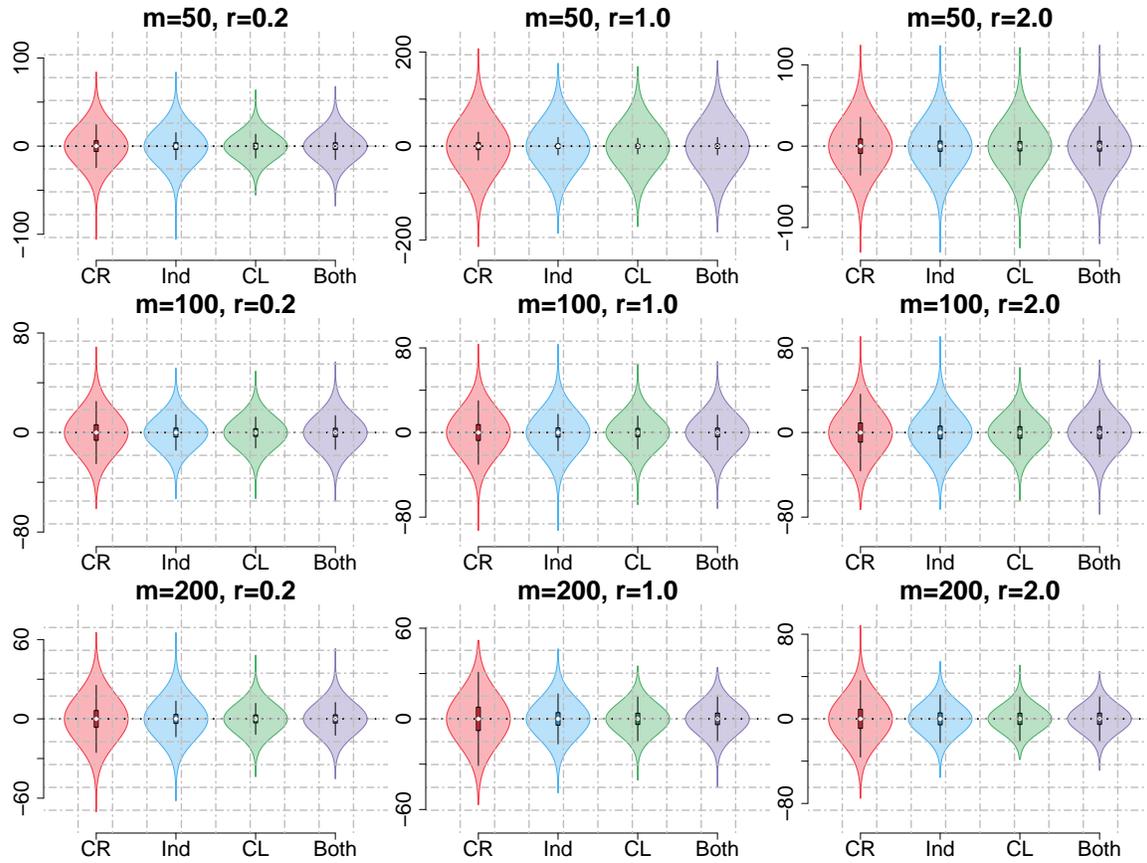


Figure 12: Distributions of  $\sqrt{m^*}(\hat{\tau}_{\text{CAE}} - \tau(\mathbf{1}, \mathbf{0}))$  under the four randomization approaches.

Table 5: Average fractions of clusters and samples included in CAE:  $\rho_1 = m^{-1}m^*$ ,  $\rho_2 = n^{-1}\mathbb{E}[\sum_j n_j]$ , and  $\sum_{j=1}^m c_j^{-1}\mathbb{E}[n_j]$ .

Fraction(%)		$r$				
		0.2	0.6	1.0	1.4	1.8
CR	$\rho_1$	100.00	100.00	99.79	98.11	92.87
	$\rho_2$	82.10	55.42	37.43	25.35	17.22
	$\rho_3$	82.06	55.34	37.32	25.25	17.14
CL	$\rho_1$	100.00	100.00	99.83	98.28	93.17
	$\rho_2$	81.88	54.90	36.74	24.63	16.51
	$\rho_3$	81.88	54.90	36.72	24.61	16.50
IND	$\rho_1$	100.00	100.00	99.81	98.24	93.09
	$\rho_2$	81.91	55.01	36.87	24.77	16.67
	$\rho_3$	81.89	54.98	36.83	24.73	16.64
Both	$\rho_1$	100.00	100.00	99.82	98.27	93.14
	$\rho_2$	81.87	54.90	36.75	24.64	16.52
	$\rho_3$	81.86	54.89	36.72	24.62	16.50

The performance of different network A/B testing approaches is also consistent with our findings presented in Section 5.2. The conclusion can be drawn in the following four folds. First, the CAE is consistent, whereas the DIM estimator suffers from the bias caused by network interference. Second, the standard deviation of the CAE is smaller than that of the DIM estimator, resulting in a higher efficiency. Third, the covariate balance generally improves the performance of the CAE. For instance, the CAE under CLAR has a higher efficiency than the CAE under CR in terms of a smaller standard deviation. Finally, Table 5 further supports the usage of our proposed cluster-adaptive network A/B testing procedure. It indicates that the advanced performance of the proposed procedure is not affected by the violation of the Assumption 4.2.

### Appendix C. Experimental Details for Paluck et al. (2016)

This section provides the experimental details for Paluck et al. (2016). The attitude toward conflict score (ATC) is calculated by summarizing the survey questions presented in Table 6. The answers to these questions are all “yes” or “no”, which are labeled as dummy variables with values zero and one. Then ATC can be calculated via the following formula,

$$\begin{aligned}
 \text{ATC} = & \left\{ -\text{CILW2} - \text{CFLW2} + \text{CSCAW2} + \text{CLHCW2} - \text{CBNPW2} - \text{CMOSW2} \right. \\
 & \left. + \text{FLIBW2} - \text{FLSHW2} - \text{FLSDW2} + \text{ADSCW2} + \text{ADNPW2} + \text{ADTSW2} \right\} \\
 & - \left\{ -\text{CIL} - \text{CFL} + \text{CSCA} + \text{CIHC} - \text{CBNP} - \text{CMOS} \right. \\
 & \left. + \text{FLIB} - \text{FLSH} - \text{FLSD} + \text{ADSC} + \text{ADNP} + \text{ADTS} \right\}.
 \end{aligned}$$

The covariates considered in the outcome model are listed as follows: 1)  $X_{i,1,\text{IN}}$ : HOSN0, that is, the indicator for the question that friends say I have a really nice house, if the value

Table 6: Questions and variable names in the surveys.

Question	Variable name	
	Wave 1	Wave 2
I have a lot of conflict	CIL	CILW2
My friends have a lot of conflict	CFL	CFLW2
We can change the conflict	CSCA	CSCAW2
I would be interested in invitation to change	CIHC	CLHCW2
Bullying is not a problem	CBNP	CBNPW2
Sometimes you have to be mean	CMOS	CMOSW2
I feel I belong	FLIB	FLIBW2
I have stayed home from school	FLSH	FLSHW2
I have been bothered by feeling sad and down	FLSD	FLSDW2
Teachers and rules help solve student conflicts	ADSC	ADSCW2
Teachers don't let kids get picked on	ADNP	ADNPW2
I can talk to an adult at this school	ADTS	ADTSW2

is yes; 2)  $X_{i,2,\text{IN}}$ : HOSN0-NA, the indicator for the question that friends say I have a really nice house, if the value is NA; 3)  $X_{i,3,\text{IN}}$ : ETHW, that is, the indicator of the ethnicity of the respondent, if the value is white; 4)  $X_{i,4,\text{IN}}$ : the indicator of outer node; 5)  $X_{i,5,\text{IN}}$ : the number of edges connecting other clusters; 6)  $X_{j,1,\text{CL}}$ : the cluster size  $c_j$ ; 7)  $X_{j,2,\text{CL}}$ : the portion of white students in  $C_j$ ; 8)  $X_{i,3,\text{CL}}$ : the portion of male students in  $C_j$ ; and 9)  $X_{i,\text{CL},4}$ : the density of  $C_j$ .

## References

- Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.
- Susan Athey, Dean Eckles, and Guido W Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018.
- Falco J Bargagli-Stoffi, Costanza Tortù, and Laura Forastiere. Heterogeneous treatment and spillover effects under clustered network interference. *arXiv preprint arXiv:2008.00707*, 2020.
- Guillaume W Basse and Edoardo M Airoldi. Limitations of design-based causal inference and A/B testing under arbitrary and network interference. *Sociological Methodology*, 48(1):136–151, 2018a.
- Guillaume W Basse and Edoardo M Airoldi. Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika*, 105(4):849–858, 2018b.
- Guillaume W Basse, Avi Feller, and Panos Toulis. Randomization tests of causal effects under interference. *Biometrika*, 106(2):487–494, 2019.

- Niloy Biswas and Edoardo M Airoldi. Estimating peer-influence effects under homophily: Randomized treatments and insights. In Complex Networks IX: Proceedings of the 9th Conference on Complex Networks CompleNet 2018, pages 323–347. Springer, 2018.
- Louis HY Chen, Larry Goldstein, and Qi-Man Shao. Normal approximation by Stein’s method. Springer, New York, 1st edition, 2010.
- Qianyi Chen, Bo Li, Lu Deng, and Yong Wang. Optimized covariance design for ab test on social network under interference. In A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, Advances in Neural Information Processing Systems, volume 36, pages 37448–37471. Curran Associates, Inc., 2023.
- Alex Chin. Central limit theorems via stein’s method for randomized experiments under interference. arXiv preprint arXiv:1804.03105, 2018.
- Dean Eckles, Brian Karrer, and Johan Ugander. Design and analysis of experiments in networks: Reducing bias from interference. Journal of Causal Inference, 5(1):20150021, 2017.
- R.A. Fisher. Design of Experiments. Oliver and Boyd, Edinburgh, 1935.
- Laura Forastiere, Edoardo M Airoldi, and Fabrizia Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. Journal of the American Statistical Association, 116(534):901–918, 2021.
- Laura Forastiere, Fabrizia Mealli, Albert Wu, and Edoardo M Airoldi. Estimating causal effects under network interference with bayesian generalized propensity scores. Journal of Machine Learning Research, 23(289):1–61, 2022.
- Santo Fortunato. Community detection in graphs. Physics Reports, 486(3-5):75–174, 2010.
- Mengsi Gao and Peng Ding. Causal inference in network experiments: regression-based analysis and design-based properties. arXiv preprint arXiv:2309.07476, 2023.
- Huan Gui, Ya Xu, Anmol Bhasin, and Jiawei Han. Network A/B testing: From sampling to estimation. In Proceedings of the 24th International Conference on World Wide Web, pages 399–409, 2015.
- Peter Hall and Christopher C Heyde. Martingale Limit Theory and Its Application. Academic press, New York, 1st edition, 1980.
- David Holtz, Ruben Lobel, Inessa Liskovich, and Sinan Aral. Reducing interference bias in online marketplace pricing experiments. arXiv preprint arXiv:2004.12489, 2020.
- Guanglei Hong and Stephen W Raudenbush. Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. Journal of the American Statistical Association, 101(475):901–910, 2006.
- Feifang Hu, Xiaoqing Ye, and Li-Xin Zhang. Multi-arm covariate-adaptive randomization. Science China Mathematics, 66(1):163–190, 2023.

- Yanqing Hu and Feifang Hu. Asymptotic properties of covariate-adaptive randomization. The Annals of Statistics, 40(3):1794–1815, 2012.
- Kosuke Imai, Zhichao Jiang, and Anup Malani. Causal inference with interference and noncompliance in two-stage randomized experiments. Journal of the American Statistical Association, 116(534):632–644, 2021.
- Guido W Imbens and Donald B Rubin. Causal Inference in Statistics, Social, and Biomedical Sciences. Cambridge University Press, New York, 2015.
- Vishesh Karwa and Edoardo M Airoidi. A systematic investigation of classical causal inference strategies under mis-specification due to network interference. arXiv preprint arXiv:1810.08259, 2018.
- Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. Online controlled experiments at large scale. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1168–1176, 2013.
- Ron Kohavi, Diane Tang, and Ya Xu. Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing. Cambridge University Press, New York, 2020.
- Denis Kojevnikov, Vadim Marmer, and Kyungchul Song. Limit theorems for network dependent random variables. Journal of Econometrics, 222(2):882–908, 2021.
- Eric D. Kolaczyk. Statistical Analysis of Network Data. Springer, New York, 2009.
- KK Gordon Lan and David L DeMets. Changing frequency of interim analysis in sequential monitoring. Biometrics, 45(3):1017–1020, 1989.
- Nicholas Larsen, Jonathan Stallrich, Srijan Sengupta, Alex Deng, Ron Kohavi, and Nathaniel T Stevens. Statistical challenges in online controlled experiments: A review of a/b testing methodology. The American Statistician, 78(2):135–149, 2024.
- Michael P Leung. Treatment and spillover effects under network interference. Review of Economics and Statistics, 102(2):368–380, 2020.
- Michael P Leung. Causal inference under approximate neighborhood interference. Econometrica, 90(1):267–293, 2022a.
- Michael P Leung. Rate-optimal cluster-randomized designs for spatial interference. The Annals of Statistics, 50(5):3064–3087, 2022b.
- Michael P Leung. Network cluster-robust inference. Econometrica, 91(2):641–667, 2023.
- Yang Liu and Feifang Hu. Balancing unobserved covariates with covariate-adaptive randomized experiments. Journal of the American Statistical Association, 117(538):875–886, 2022.
- Yang Liu and Feifang Hu. The impacts of unobserved covariates on covariate-adaptive randomized experiments. The Annals of Statistics, 51(5):1895–1920, 2023.

- Yang Liu, Yifan Zhou, Ping Li, and Feifang Hu. Adaptive A/B test on networks with cluster structures. In International Conference on Artificial Intelligence and Statistics, pages 10836–10851. PMLR, 2022.
- Wei Ma, Feifang Hu, and Lixin Zhang. Testing hypotheses of covariate-adaptive randomized clinical trials. Journal of the American Statistical Association, 110(510):669–680, 2015.
- Wei Ma, Yichen Qin, Yang Li, and Feifang Hu. Statistical inference for covariate-adaptive randomization procedures. Journal of the American Statistical Association, 115(531):1488–1497, 2020.
- Wei Ma, Ping Li, Li-Xin Zhang, and Feifang Hu. A new and unified family of covariate adaptive randomization procedures and their properties. Journal of the American Statistical Association, 119(545):151–162, 2024.
- Charles F Manski. Economic analysis of social interactions. Journal of Economic Perspectives, 14(3):115–136, 2000.
- Sean P Meyn and Richard L Tweedie. Markov Chains and Stochastic Stability. Cambridge University Press, Cambridge, United Kingdom, 3rd edition, 2013.
- Kari Lock Morgan and Donald B Rubin. Rerandomization to improve covariate balance in experiments. The Annals of Statistics, 40(2):1263–1282, 2012.
- Mark EJ Newman. Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23):8577–8582, 2006.
- Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. Physical Review E, 69(2):026113, 2004.
- Elizabeth L Ogburn, Oleg Sofrygin, Ivan Diaz, and Mark J Van der Laan. Causal inference for social network data. Journal of the American Statistical Association, 119(545):597–611, 2024.
- Elizabeth Levy Paluck, Hana Shepherd, and Peter M Aronow. Changing climates of conflict: A social network experiment in 56 schools. Proceedings of the National Academy of Sciences, 113(3):566–571, 2016.
- Mathew Penrose. Random Geometric Graphs, volume 5. Oxford University Press, New York, 2003.
- Jean Pouget-Abadie, Guillaume Saint-Jacques, Martin Saveski, Weitao Duan, Souvik Ghosh, Ya Xu, and Edoardo M Airoldi. Testing for arbitrary interference on experimentation platforms. Biometrika, 106(4):929–940, 2019.
- Yichen Qin, Yang Li, Wei Ma, Haoyu Yang, and Feifang Hu. Adaptive randomization via mahalanobis distance. Statistica Sinica, 34(1):353–375, 2024.
- Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. Physical Review E, 76(3):036106, 2007.

- Paul R Rosenbaum, P Rosenbaum, and Briskman. Design of observational studies, volume 10. Springer, New York, 2010.
- William F Rosenberger and John M Lachin. Randomization in Clinical Trials: Theory and practice. John Wiley & Sons, Hoboken, N.J, 2nd edition, 2015.
- William F Rosenberger and Oleksandr Sverdlov. Handling covariates in the design of clinical trials. Statistical Science, 23(3):404–419, 2008.
- William F Rosenberger, Diane Uschner, and Yanying Wang. Randomization: The forgotten component of the randomized clinical trial. Statistics in Medicine, 38(1):1–12, 2019.
- Nathan Ross. Fundamentals of Stein’s method. Probability Surveys, 8:210 – 293, 2011.
- Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5):688, 1974.
- Fredrik Sävje, Peter M Aronow, and Michael G Hudgens. Average treatment effects in the presence of unknown interference. The Annals of Statistics, 49(2):673–701, 2021.
- Chengchun Shi, Xiaoyu Wang, Shikai Luo, Hongtu Zhu, Jieping Ye, and Rui Song. Dynamic causal effects evaluation in a/b testing with a reinforcement learning framework. Journal of the American Statistical Association, 118(543):2059–2071, 2023.
- Johan Ugander and Hao Yin. Randomized graph cluster randomization. Journal of Causal Inference, 11(1):20220014, 2023.
- Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow. The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503, 2011.
- Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 329–337, 2013.
- Jialu Wang, Ping Li, and Feifang Hu. A/B testing in network data with covariate-adaptive randomization. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, Proceedings of the 40th International Conference on Machine Learning, volume 202 of Proceedings of Machine Learning Research, pages 35949–35969. PMLR, 23–29 Jul 2023.
- Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. Nature, 393(6684):440–442, 1998.
- Ya Xu, Nanyu Chen, Addrian Fernandez, Omar Sinno, and Anmol Bhasin. From infrastructure to culture: A/B testing challenges in large scale social networks. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, page 2227–2236, 2015.

Christina Lee Yu, Edoardo M Airoidi, Christian Borgs, and Jennifer T Chayes. Estimating the total treatment effect in randomized experiments with unknown network structure. Proceedings of the National Academy of Sciences, 119(44):e2208975119, 2022.

Zhixin Zhou, Ping Li, and Feifang Hu. Adaptive randomization in network data. Electronic Journal of Statistics, 18(1):47–76, 2024.

Hongjian Zhu and Feifang Hu. Sequential monitoring of covariate-adaptive randomized clinical trials. Statistica Sinica, 29(1):265–282, 2019.