

# STABLE LIMIT THEOREMS FOR EMPIRICAL PROCESSES UNDER CONDITIONAL NEIGHBORHOOD DEPENDENCE

Ji Hyung Lee and Kyungchul Song

*University of Illinois and University of British Columbia*

**ABSTRACT.** This paper introduces a new concept of stochastic dependence among many random variables which we call conditional neighborhood dependence (CND). Suppose that there are a set of random variables and a set of sigma algebras where both sets are indexed by the same set endowed with a neighborhood system. When the set of random variables satisfies CND, any two non-adjacent sets of random variables are conditionally independent given sigma algebras having indices in one of the two sets' neighborhood. Random variables with CND include those with conditional dependency graphs and a class of Markov random fields with a global Markov property. The CND property is useful for modeling cross-sectional dependence governed by a complex, large network. This paper provides two main results. The first result is a stable central limit theorem for a sum of random variables with CND. The second result is a Donsker-type result of stable convergence of empirical processes indexed by a class of functions satisfying a certain bracketing entropy condition when the random variables satisfy CND.

**KEY WORDS.** Conditional Neighborhood Dependence; Dependency Graphs; Markov Random Fields; Empirical Processes; Maximal Inequalities; Stable Central Limit Theorem

AMS MSC 2010: 60B10; 60F05; 60G57

JEL CLASSIFICATION: C12, C21, C31

---

*Date:* December 12, 2017.

We thank Denis Kojevnikov for his valuable comments. We also thank a referee for valuable suggestions for the improvement of the paper. All errors are ours. Corresponding address: Kyungchul Song, Vancouver School of Economics, University of British Columbia, 6000 Iona Drive, Vancouver, BC, Canada, V6T 1L4. Email address: kysong@mail.ubc.ca. Co-author address: Ji Hyung Lee, Department of Economics, University of Illinois, 1407 W. Gregory Dr., 214 David Kinley Hall, Urbana, IL 61801. Email address: jihyung@illinois.edu.

# 1. Introduction

Empirical processes indexed by a class of functions arise in many applications, in particular in developing asymptotic inference for nonparametric or semiparametric models and for goodness-of-fit specification tests. (See, e.g., Andrews (1994) and Chapter 3 of van der Vaart and Wellner (1996) for a review of applications of empirical process theory in statistics and econometrics.) While a predominant body of the literature on empirical process theory focuses on independent observations or time series observations, there is relatively little research on empirical processes with spatial or cross-sectional dependence. This paper aims to contribute to the literature by providing limit theorems for empirical processes which consist of random variables with a flexible, complex (cross-sectional) dependence structure.

In this paper, we introduce a new notion of stochastic dependence among a set of random variables. Suppose that we are given a set of random variables  $\{Y_i\}_{i \in N_n}$  indexed by a set  $N_n$ , where the set  $N_n$  is endowed with a neighborhood system so that each  $i \in N_n$  is associated with a subset  $\nu_n(i) \subset N_n \setminus \{i\}$  called the *neighborhood* of  $i$ . In this paper, we call the map  $\nu_n : N_n \rightarrow 2^{N_n}$  a *neighborhood system*. Given a neighborhood system  $\nu_n$  and a set of  $\sigma$ -fields  $\mathcal{M} \equiv (\mathcal{M}_i)_{i \in N_n}$ , we say that  $\{Y_i\}_{i \in N_n}$  is *conditionally neighborhood dependent* (CND) with respect to  $(\nu_n, \mathcal{M})$  if for any two non-adjacent subsets  $A$  and  $B$  of  $N_n$ ,  $((Y_i)_{i \in A}, (\mathcal{M}_i)_{i \in A})$  and  $((Y_i)_{i \in B}, (\mathcal{M}_i)_{i \in B})$  are conditionally independent given  $(\mathcal{M}_i)_{i \in \nu_n(A)}$ , where  $\nu_n(A)$  is the union of the neighborhoods of  $i \in A$  with the set  $A$  itself removed.

Our CND property is a generalization of both dependency graphs and Markov random fields with a global Markov property (Lauritzen, Dawid, Larsen, and Leimer (1990).) Dependency graphs were introduced by Stein (1972) in his study of normal approximation. (See Chen and Shao (2004) and Rinott and Rotar (1996) for a general local dependence notion that is different from ours.) A set of random variables have a graph as a dependency graph, if two sets of random variables are allowed to be dependent only when the two sets are adjacent in the graph. This dependence can be viewed as restrictive in many applications, as it requires that any random variables be independent even if their indices are indirectly connected in the graph. In contrast, CND random variables are allowed to be dependent even if they are not adjacent in a graph. The CND property captures the notion that “any two random variables are independent once we condition on the source of their joint dependence”. In this sense, the CND property is closely related to a Markov property in the literature of random fields. However, in contrast to the Markov property, the CND property does not require that the  $\sigma$ -fields  $\mathcal{M}_i$  be generated by  $Y_i$  itself.

This paper provides two main results. The first main result is a Berry-Esseen bound for a sum of CND random variables. Our bound is comparable to Berry-Esseen bounds established for a sum of random variables with a dependency graph in some generic situations in the literature. (Baldi and Rinott (1989), Chen and Shao (2004), and Penrose (2003) to name but a few.) This latter literature typically uses Stein’s method to establish the bound, but to the best of our knowledge, the existing proofs using Stein’s method for dependency graphs do not seem immediately extendable to a sum of CND random variables, due to a more flexible form of conditioning  $\sigma$ -fields involved in the CND property. In this paper, we use a traditional characteristic function-based method to derive a Berry-Esseen bound.

A typical form of a Berry-Esseen bound in these set-ups, including ours, involves the maximum degree of the neighborhood system, so that when the maximum degree is high, the bound is of little use. However, in many social networks observed, removing a small number of high-degree vertices tends to reduce the maximum degree of the neighborhood system substantially. Exploiting this insight, we provide a general version of a Berry-Esseen bound which uses conditioning on random variables associated with high degrees.

The second main result in this paper is a stable limit theorem for an empirical process indexed by a class of functions, where the empirical process is constituted by CND random variables. Stable convergence is a stronger notion of convergence than weak convergence, and is useful for asymptotic theory of statistics whose normalizing sequence has a random limit.

To obtain a stable limit theorem, we first extend the exponential inequality of Janson (2004) for dependency graphs to our set-up of CND random variables, and using this, we obtain a maximal inequality for an empirical process with a bracketing-entropy type bound. This maximal inequality is useful for various purposes, especially when one needs to obtain limit theorems that are uniform over a given class of functions indexing the empirical process. Using this maximal inequality, we establish the asymptotic equicontinuity of the empirical process which, in combination with the central limit theorem that comes from our previously established Berry-Esseen bound, gives a stable limit theorem. This enables stable convergence of the empirical process to a mixture Gaussian process.

As it turns out, our stable limit theorem for an empirical process requires that the maximum degree of the neighborhood system be bounded. However, in many real-life networks, the maximum degree can be substantial, especially when the networks behave like a preferential attachment network of Barabási-Albert. (Barabási and Albert (1999)) Thus, following the same spirit of extending the Berry-Esseen bound to the case conditional on high degree vertices, we extend the stable limit theory to a set-up where it relies only on those observations with relatively low degrees by conditioning on the random variables

associated with high degree vertices. This extension enables us to obtain a stable limit theorem for empirical processes when the maximum degree of the neighborhood system increases to infinity as the size of the system increases.

Stable convergence has been extensively studied in the context of martingale central limit theorems. (See, e.g. Hall and Heyde (1980).) See Häusler and Luschgy (2010) for stable limit theorems for Markov kernels and related topics. Recent studies by Kuersteiner and Prucha (2013) and Hahn, Kuersteiner, and Mazzocco (2016) established a stable central limit theorem for a sum of random variables having both cross-sectional dependence and time series dependence by utilizing a martingale difference array formulation of the random variables.

Markov-type cross-sectional dependence on a graph has received attention in the literature (Lauritzen (1996).) In particular, the pairwise Markov property of random variables says that two non-adjacent random variables are conditionally independent given all the other variables, and is captured by a precision matrix in a high dimensional Gaussian model. (See Meinshausen and Bühlmann (2008) and Cai, Liu, and Zhou (2016) for references.) This paper's CND property is stronger than the pairwise Markov property when the conditioning  $\sigma$ -fields,  $\mathcal{M}_i$ 's, are those that are generated by the random variables. However, the CND property encompasses the case where the latter condition does not hold, and thus includes dependency graphs as a special case, unlike Markov-type dependence mentioned before.

Wu (2005) introduced a dependence concept that works well with nonlinear causal processes. More recently, Jirak (2016) established a Berry-Esseen bound with optimal rate for nonlinear causal processes with temporal ordering. Chen and Wu (2016) considered a nonlinear spatial process indexed by a lattice in the Euclidean space. These models are distinct from ours. The major distinction of our approach is to model the stochastic process to be indexed by a generic graph, and model the dependence structure using conditional independence relations along the graph. Thus our approach works well with, for example, Markov random fields on an undirected graph. On the other hand, the models of this literature accommodate various temporal or spatial autoregressive processes. To the best of our knowledge, stable convergence of empirical processes indexed by a class of functions has not been studied under either dependency graphs or Markov random fields on a graph.

The remainder of the paper proceeds as follows. In Section 2, we formally introduce the notion of conditional neighborhood dependence (CND) and study its basic properties. In Section 3, we provide stable central limit theorems for a sum of CND random variables. We also present the stable convergence of an empirical process to a mixture Gaussian process. The mathematical proofs of the results are found in the appendix.

## 2. Conditional Neighborhood Dependence

### 2.1. Definition

Let  $\mathcal{N}$  be an infinite countable set. For each  $n = 1, 2, \dots$ , let  $N_n \subset \mathcal{N}$  be a finite set such that  $|N_n| = n$  and let  $2^{N_n}$  be the collection of all the subsets of  $N_n$ . We assume that  $N_n$  is a proper subset of  $N_{n+1}$  for each  $n \geq 1$ . We will call each element of  $N_n$  a *vertex*, and call any map  $\nu_n : N_n \rightarrow 2^{N_n}$  a *neighborhood system*, if for each  $i \in N_n$ ,  $i \notin \nu_n(i)$ .<sup>1</sup> Let us define for each  $A \subset N_n$ ,

$$\bar{\nu}_n(A) \equiv \left( \bigcup_{i \in A} \nu_n(i) \right) \cup A \text{ and } \nu_n(A) \equiv \bar{\nu}_n(A) \setminus A.$$

For  $\{i_1, \dots, i_m\} \subset N_n$ , we simply write  $\nu_n(i_1, \dots, i_m) = \nu_n(\{i_1, \dots, i_m\})$  and  $\bar{\nu}_n(i_1, \dots, i_m) = \bar{\nu}_n(\{i_1, \dots, i_m\})$ , suppressing the curly brackets. Let us call  $\bar{\nu}_n(A)$  *the  $\nu_n$ -closure of  $A$*  and  $\nu_n(A)$  *the  $\nu_n$ -boundary of  $A$* . The  $\nu_n$ -closure of  $A$  includes the vertices in  $A$  but the  $\nu_n$ -boundary around  $A$  excludes them.

If for each  $i, j \in N_n$ ,  $i \in \nu_n(j)$  implies  $j \in \nu_n(i)$ , we say that neighborhood system  $\nu_n$  is *undirected*. If there exists a pair  $i, j \in N_n$  such that  $i \in \nu_n(j)$  but  $j \notin \nu_n(i)$ , we say that neighborhood system  $\nu_n$  is *directed*.

It is often useful to compare different dependence structures governed by different neighborhood systems. When we have two neighborhood systems  $\nu_n$  and  $\nu'_n$  such that  $\nu_n(i) \subset \nu'_n(i)$  for each  $i \in N_n$  and  $\nu_n(j) \neq \nu'_n(j)$  for some  $j \in N_n$ , we say that  $\nu_n$  is *strictly finer than  $\nu'_n$*  and  $\nu'_n$  is *strictly coarser than  $\nu_n$* . When  $\nu_n(i) \subset \nu'_n(i)$  for each  $i \in N_n$ , we say that  $\nu_n$  is *weakly finer than  $\nu'_n$*  and  $\nu'_n$  is *weakly coarser than  $\nu_n$* .

Let us introduce the notion of dependence among a triangular array of  $\sigma$ -fields. Let  $(\Omega, \mathcal{F}, P)$  be a given probability space, and let  $\{\mathcal{F}_i\}_{i \in N_n}$  be a given triangular array of sub- $\sigma$ -fields of  $\mathcal{F}$ , indexed by  $i \in N_n$ . (Proper notation for the sub- $\sigma$ -field in the triangular array should be  $\mathcal{F}_{i_n}$ , but we suppress the  $n$  subscript for simplicity.) For any  $A \subset N_n$ , we let  $\mathcal{F}_A$  be the smallest  $\sigma$ -field that contains  $\mathcal{F}_i, i \in A$ , and  $\mathcal{F}_{-A}$  be the smallest  $\sigma$ -field that contains all the  $\sigma$ -fields  $\mathcal{F}_i$  such that  $i \in N_n \setminus A$ . When  $A = \emptyset$ , we simply take  $\mathcal{F}_A$  to be the trivial  $\sigma$ -field. We apply this notation to other triangular arrays of  $\sigma$ -fields, so that if  $\{\mathcal{M}_i\}_{i \in N_n}$  is a triangular array of sub- $\sigma$ -fields of  $\mathcal{F}$ , we similarly define  $\mathcal{M}_A$  and  $\mathcal{M}_{-A}$  for any  $A \subset N_n$ . For given two  $\sigma$ -fields, say  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , we write  $\mathcal{G}_1 \vee \mathcal{G}_2$  to represent the smallest  $\sigma$ -field that contains both  $\mathcal{G}_1$  and  $\mathcal{G}_2$ .

<sup>1</sup>Equivalently, one might view the neighborhood system as a graph by identifying each neighborhood of  $i$  as the neighborhood of  $i$  in the graph. However, it seems more natural to think of a stochastic dependence structure among random variables in terms of neighborhoods rather than in terms of edges in a graph.

Given a triangular array of  $\sigma$ -fields,  $\{\mathcal{M}_i\}_{i=1}^\infty$ , let us introduce a sub  $\sigma$ -field  $\mathcal{G}$  defined by

$$(2.1) \quad \mathcal{G} \equiv \bigcap_{n \geq 1} \bigcap_{i \in N_n} \mathcal{M}_i.$$

In many applications,  $\mathcal{G}$  is used to accommodate random variables with a common shock. For example, suppose that each  $\mathcal{M}_i$  is generated by a random vector  $(\varepsilon_i, U)$  from a set of random variables  $\{\varepsilon_i\}_{i \in N_n}$  that are conditionally independent given a common random variable  $U$ . Then we can take  $\mathcal{G}$  to be the  $\sigma$ -field generated by  $U$ . We will discuss examples of CND random vectors in a later section, after we study their properties.

Let us introduce the notion of dependence of an array of  $\sigma$ -fields that is of central focus in this paper.

**Definition 2.1.** (i) Given neighborhood system  $\nu_n$  on  $N_n$  and an array of  $\sigma$ -fields,  $\mathcal{M} \equiv \{\mathcal{M}_i\}_{i \in N_n}$ , we say that  $\sigma$ -fields  $\{\mathcal{F}_i\}_{i \in N'_n}$  for a given subset  $N'_n \subset N_n$  are *conditionally neighborhood dependent (CND) with respect to  $(\nu_n, \mathcal{M})$* , if for any  $A, B \subset N'_n$  such that  $A \subset N'_n \setminus \bar{\nu}_n(B)$  and  $B \subset N'_n \setminus \bar{\nu}_n(A)$ ,  $\mathcal{F}_A \vee \mathcal{M}_A$  and  $\mathcal{F}_B \vee \mathcal{M}_B$  are conditionally independent given  $\mathcal{M}_{\nu_n(A)}$ .

(ii) If  $\sigma$ -fields generated by random vectors in  $\{Y_i\}_{i \in N'_n}$  for a subset  $N'_n \subset N_n$  are CND with respect to  $(\nu_n, \mathcal{M})$ , we simply say that random vectors in  $\{Y_i\}_{i \in N'_n}$  are *CND with respect to  $(\nu_n, \mathcal{M})$* .

Conditional neighborhood dependence specifies only how conditional independence arises, not how conditional dependence arises. Conditional neighborhood dependence does not specify independence or dependence between  $Y_i$  and  $Y_j$  if  $j \in \nu_n(i)$  or  $i \in \nu_n(j)$ . Furthermore, conditional neighborhood dependence can accommodate the situation where the neighborhoods in the system  $\nu_n$  are generated by some random graph on  $N_n$ , as long as the random graph is  $\mathcal{G}$ -measurable. In such a situation, the results of this paper continue to hold with only minor modifications that take care of the randomness of  $\nu_n$ .

## 2.2. Monotonicity and Invariance to Conditioning

In general, a conditional independence property is not monotone in conditioning  $\sigma$ -fields. In other words, when  $X$  and  $Y$  are conditionally independent given a  $\sigma$ -field  $\mathcal{F}'$ , this does not imply conditional independence between  $X$  and  $Y$  given any sub- $\sigma$ -field of  $\mathcal{F}'$  or given any  $\sigma$ -field that contains  $\mathcal{F}'$ . However, CND partially obeys monotonicity in neighborhood systems. More specifically, the CND property with a finer neighborhood system implies more conditional independence restrictions than what the CND property

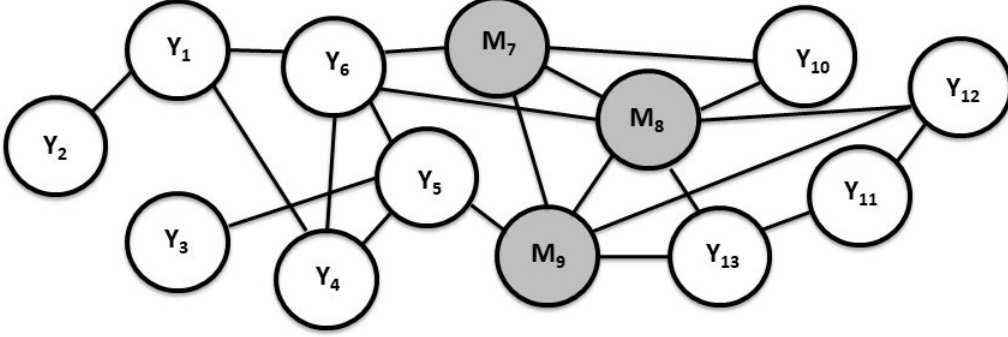


FIGURE 1. Conditional Neighborhood Dependence

Notes: Suppose we are given random variables  $(Y_1, \dots, Y_{13})$  and  $(M_1, \dots, M_{13})$ , where  $\mathcal{M}_i = \sigma(M_i)$ . The figure depicts the neighborhood system on  $N_n = \{1, 2, \dots, 13\}$ . If  $Y_1, \dots, Y_{13}$  are *CND* with respect to the neighborhood system as illustrated above and  $(\mathcal{M}_1, \dots, \mathcal{M}_{13})$ , it implies that  $Y_1, \dots, Y_6$  and  $Y_{10}, \dots, Y_{13}$  are conditionally independent given  $\mathcal{M}_7, \mathcal{M}_8, \mathcal{M}_9$ .

with a coarser neighborhood system implies. We introduce a lemma which makes precise this monotonicity property of *CND*. Suppose that we are given two neighborhood systems  $\nu_n$  and  $\nu'_n$  where  $\nu_n$  is weakly finer than  $\nu'_n$ . The following lemma shows that the *CND* property of a triangular array with respect to a given neighborhood system  $\nu'_n$  carries over to that with respect to  $\nu_n$  if  $\nu_n$  is undirected.

**Lemma 2.1.** *Given two neighborhood systems  $\nu_n$  and  $\nu'_n$ , suppose that  $\nu_n$  is weakly finer than  $\nu'_n$  and that a triangular array of  $\sigma$ -fields  $\{\mathcal{F}_i\}_{i \in N_n}$  is *CND* with respect to  $(\nu_n, \mathcal{M})$  for some  $\sigma$ -fields  $\mathcal{M} \equiv \{\mathcal{M}_i\}_{i \in N_n}$ . Suppose further that  $\nu_n$  is undirected. Then  $\{\mathcal{F}_i\}_{i \in N_n}$  is *CND* with respect to  $(\nu'_n, \mathcal{M})$ .*

**Proof:** Take any  $A, B \subset N_n$  such that  $A \subset N_n \setminus \bar{\nu}'_n(B)$  and  $B \subset N_n \setminus \bar{\nu}'_n(A)$ , so that  $A \subset N_n \setminus \bar{\nu}_n(B)$  and  $B \subset N_n \setminus \bar{\nu}_n(A)$  as well. Let  $D \equiv \nu'_n(A) \setminus \bar{\nu}_n(A)$ , which, by the undirectedness of  $\nu_n$ , implies that  $A \subset N_n \setminus \bar{\nu}_n(D)$ . Therefore,  $A \subset N_n \setminus \bar{\nu}_n(B \cup D)$  and  $B \cup D \subset N_n \setminus \bar{\nu}_n(A)$ . By the *CND* property,  $\mathcal{F}_A \vee \mathcal{M}_A$  and  $\mathcal{F}_{B \cup D} \vee \mathcal{M}_{B \cup D}$  are conditionally independent given  $\mathcal{M}_{\nu_n(A)}$ . By Lemma 4.2 of Dawid (1979), this implies that  $\mathcal{F}_A \vee \mathcal{M}_A$  and  $\mathcal{F}_B \vee \mathcal{M}_B$  are conditionally independent given  $\mathcal{M}_{\nu_n(A) \cup D}$ . This proves the lemma, because

$$\nu_n(A) \cup D = \nu'_n(A),$$

which follows due to  $\nu_n(A) \subset \nu'_n(A)$ ,  $\nu_n(A) \subset \bar{\nu}_n(A)$ , and  $\bar{\nu}_n(A) \setminus \nu_n(A) = A$ . ■

In the above lemma, the requirement that  $\nu_n$  be undirected cannot be eliminated. To see this, consider the following counterexample of the lemma when  $\nu_n$  is taken to be a directed neighborhood system.

**Example 1:** Let us take  $N_n = \{1, 2, 3, 4\}$  with  $n = 4$ . Let  $(\varepsilon_i)_{i \in N_n}$  be i.i.d. standard normal random variables. Let us take

$$Y_1 = \varepsilon_1, Y_2 = \varepsilon_2, Y_3 = \varepsilon_1 + \varepsilon_3 + \varepsilon_4, \text{ and } Y_4 = \varepsilon_4.$$

Let a neighborhood system  $\nu_n$  be given as  $\nu_n(i) = \emptyset$  for all  $i = 1, 2, 4$ , and  $\nu_n(3) = \{1, 4\}$ . Hence  $\nu_n$  is directed. Then we take  $\mathcal{M} = \mathcal{F} = (\mathcal{F}_i)_{i \in N_n}$ , where each  $\mathcal{F}_i$  denotes the  $\sigma$ -field generated by  $Y_i$ . Then it is not hard to see that  $(Y_i)_{i \in N_n}$  is CND with respect to  $(\nu_n, \mathcal{M})$ .

Now let us introduce another neighborhood system  $\nu'_n$  that is weakly coarser than  $\nu_n$ . Let  $\nu'_n(1) = \{3\}$ ,  $\nu'_n(2) = \{1\}$ ,  $\nu'_n(3) = \{1, 4\}$ , and  $\nu'_n(4) = \{3\}$ . Then we take  $A = \{1\}$  and  $B = \{4\}$  so that  $A \subset N_n \setminus \bar{\nu}'_n(B)$  and  $B \subset N_n \setminus \bar{\nu}'_n(A)$ . Note that  $\nu'_n(A) = \nu'_n(1) = \{3\}$ . Certainly  $Y_1$  and  $Y_4$  are not conditionally independent given  $Y_3$ , because  $Y_3$  involves both  $\varepsilon_1$  and  $\varepsilon_4$ . ■

As we shall see later, using further conditioning in the CND property, one may obtain a better normal approximation for a sum of CND random variables in some situations. Here we give a preliminary result which addresses the question of whether the CND property still holds after we increase the  $\sigma$ -fields in  $\mathcal{M}$  in a certain way. More precisely, suppose that  $Y \equiv \{Y_i\}_{i \in N_n}$  is CND with respect to  $(\nu_n, \mathcal{M})$ . Then let us choose  $N_n^* \subset N_n$ , and define a neighborhood system  $\nu_n^*$  on  $N_n$  such that for all  $A \subset N_n$ ,

$$(2.2) \quad \nu_n^*(A) \equiv \nu_n(A) \cap N_n^*.$$

Let  $\mathcal{M}^* \equiv \{\mathcal{M}_i^*\}_{i \in N_n}$ , where

$$(2.3) \quad \mathcal{M}_i^* \equiv \mathcal{M}_i \vee \mathcal{M}_{N_n \setminus N_n^*}.$$

Then we obtain the following result.

**Lemma 2.2.** *Suppose that  $Y \equiv \{Y_i\}_{i \in N_n}$  is CND with respect to  $(\nu_n, \mathcal{M})$ . Fix  $N_n^* \subset N_n$  and define  $\nu_n^*$  as in (2.2) and  $\mathcal{M}^* \equiv \{\mathcal{M}_i^*\}_{i \in N_n}$  as in (2.3). Then  $Y^* \equiv \{Y_i\}_{i \in N_n^*}$  is CND with respect to  $(\nu_n^*, \mathcal{M}^*)$ .*

**Proof:** Take  $A, B \subset N_n^*$  such that  $A \subset N_n^* \setminus \bar{\nu}_n(B)$  and  $B \subset N_n^* \setminus \bar{\nu}_n(A)$ . Note that for any  $A' \subset N_n^*$ ,  $N_n^* \setminus \bar{\nu}_n(A') = N_n^* \setminus \bar{\nu}_n^*(A')$ . Let

$$B^* \equiv B \cup ((N_n \setminus N_n^*) \setminus \bar{\nu}_n(A)).$$

By the CND of  $Y$ , we have that  $\sigma(Y_A) \vee \mathcal{M}_A$  and  $\sigma(Y_{B^*}) \vee \mathcal{M}_{B^*}$  are conditionally independent given  $\mathcal{M}_{\nu_n(A)}$ . Hence  $\sigma(Y_A) \vee \mathcal{M}_A$  and  $\sigma(Y_B) \vee \mathcal{M}_{B^*}$  are conditionally independent



given  $\mathcal{M}_{\nu_n(A)}$ . From (2.2), we have

$$\nu_n^*(A)^c \cap \nu_n(A) = \nu_n(A) \cap (N_n \setminus N_n^*).$$

Hence we can write

$$\nu_n(A) = \nu_n^*(A) \cup (\nu_n^*(A)^c \cap \nu_n(A)) = \nu_n^*(A) \cup (\nu_n(A) \cap (N_n \setminus N_n^*)),$$

so that  $\sigma(Y_A) \vee \mathcal{M}_A$  and  $\sigma(Y_B) \vee \mathcal{M}_{B^*}$  are conditionally independent when we condition on  $\mathcal{M}_{\nu_n^*(A) \cup (\nu_n(A) \cap (N_n \setminus N_n^*))}$ . This implies that  $\sigma(Y_A) \vee \mathcal{M}_A$  and  $\sigma(Y_B) \vee \mathcal{M}_B$  are conditionally independent given  $\mathcal{M}_{\nu_n^*(A) \cup (N_n \setminus N_n^*)}$  by the definition of  $B^*$ , and by the choice of  $A \subset N_n^*$ . ■

## 2.3. Examples

**2.3.1. Conditional Dependency Graphs.** Let  $G_n = (N_n, E_n)$  be an undirected graph on  $N_n$ , where  $E_n$  denotes the set of edges. Define  $\nu_n(i) = \{j \in N_n : ij \in E_n\}$ . Suppose that  $\{Y_i\}$  has  $G_n$  as a *conditional dependency graph*, i.e., for each  $A \subset N_n$ ,  $Y_A$  and  $Y_{N_n \setminus \bar{\nu}_n(A)}$  are conditionally independent given a  $\sigma$ -field  $\mathcal{C}$ . Dependency graphs were introduced by Stein (1972) and have received attention in the literature. (See, for example, Janson (1988), Baldi and Rinott (1989) and Rinott and Rotar (1996) for Berry-Esseen bounds for the sum of random variables with a dependency graph, and Janson (2004) for an exponential inequality. See Song (2015) for an application to permutation inference.) Then  $\{Y_i\}$  is CND with respect to  $(\nu_n, \mathcal{M})$ , where  $\mathcal{M}_i$  is taken to be  $\mathcal{C}$  for each  $i \in N_n$ . ■

**2.3.2. Functional Local Dependence.** Let  $G_n = (N_n, E_n)$  be a given directed graph, so that  $ij \in E_n$  represents an edge *from*  $i$  to  $j$ . The neighborhood  $N_n^O(i) = \{j \in N_n : ij \in E_n\}$  is the *out-neighborhood* of  $i$ , the set of vertices  $j$  that there is an edge from vertex  $i$  to vertex  $j$ . Similarly,  $N_n^I(i) = \{j \in N_n : ji \in E_n\}$  is the *in-neighborhood* of  $i$ . We define  $\bar{N}_n^O(i) = N_n^O(i) \cup \{i\}$  and  $\bar{N}_n^I(i) = N_n^I(i) \cup \{i\}$ . Suppose that  $Y_i$  is generated in the following way:

$$(2.4) \quad Y_i = \gamma_i(\varepsilon_{\bar{N}_n^I(i)}, \eta_i),$$

where  $\varepsilon_{\bar{N}_n^I(i)} = (\varepsilon_j)_{j \in \bar{N}_n^I(i)}$  and  $\varepsilon_i$ 's and  $\eta_i$ 's are independent across  $i$ 's and  $(\varepsilon_i)_{i \in N_n}$  and  $(\eta_i)_{i \in N_n}$  are independent from each other. Here the functions,  $\gamma_i$ 's, are nonstochastic.

This way of modeling local dependence among  $Y_i$ 's through base random variables,  $\varepsilon_i$ 's, is useful in many contexts of applications. In particular, each outcome  $Y_i$  may arise as a consequence of local interactions among individual variables, where locality is determined by a given graph  $G_n$ . (See Leung (2016) and Canen, Schwartz, and Song (2017) for applications in economics.)

Let us define a new (undirected) graph  $G'_n = (N_n, E'_n)$  such that  $ij \in E'_n$  if and only if  $\bar{N}_n^I(i) \cap \bar{N}_n^I(j) \neq \emptyset$ . In other words,  $i$  and  $j$  are adjacent in  $G'_n$ , if their in-neighborhoods overlap. Define  $N'_n(i) = \{j \in N_n : ij \in E'_n\}$  for each  $i \in N_n$ . Then it is not hard to see that  $\{Y_i\}_{i \in N_n}$  is CND with respect to  $(N'_n, \mathcal{M}')$ , for any triangular arrays of  $\sigma$ -fields  $\mathcal{M}'_i$ , as long as each  $\mathcal{M}'_i$  is the  $\sigma$ -fields generated by  $f_i((\varepsilon_j, \eta_j)_{j \in N_n})$  for any nonstochastic measurable map  $f_i$ . Note that we can take  $\mathcal{M}'_i$  to be trivial  $\sigma$ -fields in which case the functional local dependence is essentially equivalent to the dependency graph assumption.

However, functional local dependence has much richer implications than the dependency graph assumption alone, because it generates lots of conditional independence restrictions that are not implied by the dependency graph assumption alone. Using such restrictions, we can obtain conditional neighborhood dependence as follows: for each  $i \in N_n$ , let  $\mathcal{M}_i$  be the  $\sigma$ -field generated by  $(\varepsilon_i, \eta_i)$ . It is not hard to see that the triangular array  $\{Y_i\}_{i \in N_n}$  is CND with respect to  $(N_n^I, \mathcal{M})$ . Note that the neighborhood system  $N_n^I$  is weakly finer than  $N'_n$ , and hence when the graph  $G_n$  is undirected, by Lemma 2.1, the CND with respect to  $(N_n^I, \mathcal{M})$  expresses richer conditional independence restrictions than the CND with respect to  $(N'_n, \mathcal{M})$ .

The notion of functional local dependence is related to physical dependence of Wu (2005). The difference is that physical dependence in Wu (2005) is mainly intended for time series dependence, where  $Y_i$  involves only the past and present values of  $\varepsilon_i$ 's through a common function, whereas the functional local dependence captures local dependence through a system of neighborhoods of random variables. ■

**2.3.3. Markov Random Fields on a Neighborhood System.** Suppose that we have a triangular array of random vectors  $\{Y_i\}_{i \in N_n}$ ,  $Y_i \in \mathbf{R}^d$ , where there is an *undirected* neighborhood system  $\nu_n$  on  $N_n$ . A *path* in  $\nu_n$  between two vertices  $i$  and  $j$  is defined to be a sequence of distinct vertices  $i_1, i_2, \dots, i_m$  such that  $i_1 = i$  and  $i_m = j$  and  $i_t \in \nu_n(i_{t+1})$  for each  $t = 1, \dots, m - 1$ . Let us say that a set  $S \subset N_n$  *separates* sets  $A, B \subset N_n$ , if for any  $i \in A$  and any  $j \in B$ , every path between a vertex in  $A$  and a vertex in  $B$  intersects set  $S$ . Let us consider the following two notions of Markov properties (see Lauritzen (1996), p.32).

**Definition 2.2.** (i) We say that  $\{Y_i\}_{i \in N_n}$  satisfies the *pairwise Markov property* if for any two vertices  $i, j \in N_n$  such that  $i \notin \bar{\nu}_n(j)$  and  $j \notin \bar{\nu}_n(i)$ ,  $Y_i$  and  $Y_j$  are conditionally independent given  $Y_{N \setminus \{i, j\}}$ .

(ii) We say that  $\{Y_i\}_{i \in N_n}$  satisfies the *local Markov property* if for any vertex  $i \in N_n$  and any  $B \subset N_n \setminus \bar{\nu}_n(i)$ ,  $Y_i$  and  $Y_B$  are conditionally independent given  $Y_{\nu_n(i)}$ .

(iii) We say that  $\{Y_i\}_{i \in N_n}$  satisfies the *the global Markov property* if for any two subsets  $A$  and  $B$  of  $N_n$  which are separated by set  $S \subset N_n$ ,  $Y_A$  and  $Y_B$  are conditionally independent given  $Y_S$ .

Suppose that  $\mathcal{M}_i$  is the  $\sigma$ -field generated by  $Y_i$  for each  $i \in N_n$ . It is not hard to see that if  $\{Y_i\}_{i \in N_n}$  satisfies the global Markov property, it is *CND* with respect to  $(\nu_n, \mathcal{M})$ . And if  $\{Y_i\}_{i \in N_n}$  is *CND* with respect to  $(\nu_n, \mathcal{M})$ , it satisfies the local Markov property. Hence our notion of *CND* is an intermediate concept between the local and the global Markov properties.<sup>2</sup> Suppose that for each  $i \in N$ ,  $Y_i$  takes values in a finite set, say,  $\mathcal{Y}_i$ , and  $Y = (Y_i)_{i \in N_n}$  is a discrete random vector with a positive probability mass at each point in the Cartesian product  $\times_{i \in N} \mathcal{Y}_i$ . Then, the pairwise Markov property implies the global Markov property, and hence implies the *CND* property. (See p.119 of Koller and Friedman (2009).)

Note that Markov chains are not *CND* in general. For example, consider the set  $N_n = \{1, 2, \dots, n\}$  which represents time and a directed graph  $\nu_n$  on  $N_n$  such that  $\nu_n(i) = \{i-1\}$ . Let  $\{Y_i\}_{i \in N_n}$  be a Markov chain. Then the requirement from a *CND* property that  $Y_i$  and  $Y_{i+2}$  be conditionally independent given  $Y_{i-1}$  does not follow from the Markov chain property.

### 3. Stable Limit Theorems

#### 3.1. Stable Central Limit Theorems

**3.1.1. The Basic Result.** In this section, we give a Berry-Esseen bound conditional on  $\mathcal{G}$  for a triangular array of random variables that are *CND* with respect to  $(\nu_n, \mathcal{M})$ . Given a neighborhood system  $\nu_n$  on  $N_n$ , we define

$$d_{mx} \equiv \max_{i \in N_n} |\nu_n(i)| \text{ and } d_{av} \equiv \frac{1}{n} \sum_{i \in N_n} |\nu_n(i)|,$$

where  $|\nu_n(i)|$  denotes the cardinality of set  $\nu_n(i)$ . We call  $d_{mx}$  the *maximum degree* and  $d_{av}$  the *average degree* of neighborhood system  $\nu_n$ . We use  $d_{mx}$  and  $d_{av}$  to express the conditions for the neighborhood system  $\nu_n$ . For  $p = 1, \dots, 4$ , define

$$\mu_p \equiv \max_{i \in N_n} \left( \mathbf{E} \left[ \left| \frac{X_i}{\sigma_n} \right|^p \mid \mathcal{G} \right] \right)^{1/p},$$

<sup>2</sup> Lauritzen, Dawid, Larsen, and Leimer (1990) proposed Markov fields over directed acyclic graphs. They defined local Markov property and global Markov property and provided a sufficient condition under which both are equivalent.

where  $\sigma_n^2 \equiv \text{Var}(\sum_{i \in N_n} X_i | \mathcal{G})$ . Let

$$r_n^2 \equiv \frac{1}{\sigma_n^4} \mathbf{E} \left[ \left( \sum_{i \in N_n} \sum_{j \in \bar{\nu}_n(i)} \xi_{ij} \right)^2 \middle| \mathcal{G} \right],$$

where  $\xi_{ij} \equiv \mathbf{E}[X_i X_j | \mathcal{M}_{\nu_n(i,j)}] - \mathbf{E}[X_i X_j | \mathcal{G}]$ .

**Theorem 3.1.** *Suppose that a triangular array  $\{X_i\}_{i \in N_n}$  is *CND* with respect to  $(\nu_n, \mathcal{M})$ . Furthermore assume that  $\mathbf{E}[X_i | \mathcal{M}_{\nu_n(i)}] = 0$ , a.e. for each  $i \in N_n$ , and let*

$$(3.1) \quad \mathcal{A}_n \equiv \{\omega \in \Omega : nd_{mx} d_{av} \mu_3^3(\omega) \leq 1\}.$$

*Then there exists an absolute constant  $C > 0$  such that on the event  $\mathcal{A}_n$ , for each  $n \geq 1$ ,*

$$\sup_{t \in \mathbf{R}} \Delta_n(t; \mathcal{G}) \leq C \left( \sqrt{nd_{mx} d_{av} \mu_3^3} - \log(nd_{mx} d_{av} \mu_3^3) \sqrt{nd_{mx}^2 d_{av} \mu_4^4 + r_n^2} \right), \text{ a.e.,}$$

where

$$\Delta_n(t; \mathcal{G}) \equiv \left| P \left\{ \frac{1}{\sigma_n} \sum_{i \in N_n} X_i \leq t \middle| \mathcal{G} \right\} - \Phi(t) \right|$$

and  $\Phi$  is the distribution function of the standard normal distribution.

Since the conditional CDF is a.e. right continuous,  $\sup_{t \in \mathbf{R}} \Delta_n(t; \mathcal{G}) = \sup_{t \in \mathbf{Q}} \Delta_n(t; \mathcal{G})$ , a.e., for any countable dense subset  $\mathbf{Q}$  of  $\mathbf{R}$ . The use of the bound requires a good bound for  $r_n$ . Observe that  $\mathbf{E}[\xi_{ij} | \mathcal{G}] = 0$ . Therefore, when  $\xi_{ij}$ 's are locally dependent in a proper sense, we can expect that  $r_n^2$  is at most of the same order as the term  $nd_{mx}^2 d_{av} \mu_4^4$ . The following corollary gives a set of conditions under which this is true.

**Lemma 3.1.** *Suppose that the conditions of Theorem 3.1 hold. Furthermore,  $\mathcal{M}_i$ 's are conditionally independent given  $\mathcal{G}$ . Then*

$$r_n^2 \leq 8nd_{mx}^2 d_{av} \mu_4^4.$$

Focusing on a special case satisfying an additional condition below, we can obtain an improved version of Theorem 3.1.

**Condition A:** For any  $A, B \subset N_n$  such that  $A \subset N_n \setminus \bar{\nu}_n(B)$ ,  $\mathcal{M}_{\nu_n(A)}$  and  $\mathcal{F}_B$  are conditionally independent given  $\mathcal{G}$ .

Condition A accommodates conditional dependency graphs but excludes Markov random fields.

**Corollary 3.1.** *Suppose that a triangular array  $\{X_i\}_{i \in N_n}$  is *CND* with respect to  $(\nu_n, \mathcal{M})$ , and that Condition A holds. Suppose further that  $\mathbf{E}[X_i | \mathcal{G}] = 0$ , a.e. for each  $i \in N_n$ .*

Then there exists an absolute constant  $C > 0$  such that on the event  $\mathcal{A}_n$  defined in (3.1), for each  $n \geq 1$ ,

$$\sup_{t \in \mathbf{R}} \Delta_n(t; \mathcal{G}) \leq C \left( \sqrt{nd_{mx}d_{av}\mu_3^3} - \log(nd_{mx}d_{av}\mu_3^3) \sqrt{nd_{mx}^2d_{av}\mu_4^4} \right), \text{ a.e.,}$$

where we define  $\mu_p$  and  $\Delta_n(t; \mathcal{G})$  as in Theorem 3.1.

The improvement of the result due to Condition A is two fold. First, the condition  $\mathbf{E}[X_i | \mathcal{M}_{\nu_n(i,j)}] = 0$  is weakened to  $\mathbf{E}[X_i | \mathcal{G}] = 0$ . Second, the bound does not involve  $r_n^2$ . When it is cumbersome to compute a reliable bound for  $r_n^2$ , the above corollary can be useful.<sup>3</sup>

In the case of dependency graphs, there has been much research establishing a Berry-Esseen bound. When we confine our attention to the special case of  $|X_i| \leq 1/\sqrt{n}$ ,  $i \in N_n$ , and  $d_{mx} < D$  for all  $n \geq 1$  for some constant  $D > 0$ , our bound in Corollary 3.1 has the same  $n^{-1/2}$  rate as in Baldi and Rinott (1989) (Corollary 2), Chen and Shao (2004) (Theorem 2.7) and Penrose (2003) (Theorem 2.4), among others. These papers adopted Stein's method to obtain the bound. However, to the best of our knowledge, it is not straightforward to extend their results to our set-up of CND variables. The main reason is that the conditioning  $\sigma$ -field conditional on which two sets of random variables  $X_A$  and  $X_B$  are independent varies depending on the set  $A$ . Thus, for example, we cannot apply Equation (2.4) in Penrose (2003), p.31, in our context. In this paper, we resort to a more traditional Fourier analytic method in combination with Esseen's inequality.

The Berry-Esseen bound gives stable convergence of a sum of random vectors to a mixture normal distribution. More specifically, suppose that  $\{X_i\}_{i \in N_n}$  is a triangular array of random variables such that  $\mathbf{E}[X_i | \mathcal{M}_{\nu_n(i)}] = 0$ , a.e., for each  $i \in N_n$ , and for each  $t \in \mathbf{R}$ ,

$$\Delta_n(t; \mathcal{G}) \rightarrow_P 0, \text{ as } n \rightarrow \infty.$$

Then for each  $U \in \mathcal{G}$ , and for each uniformly continuous and bounded map  $f$  on  $\mathbf{R}$ , we have

$$\mathbf{E} \left[ f \left( \frac{1}{\sigma_n} \sum_{i \in N_n} X_i \right) 1_U \right] \rightarrow \mathbf{E}[f(\mathbb{Z})1_U], \text{ as } n \rightarrow \infty,$$

where  $\mathbb{Z}$  is a standard normal random variable that is independent of  $U$  and  $1_U$  denotes the indicator of event  $U$ .

<sup>3</sup>In the special case of conditional dependency graphs, one can follow the proof of Theorem 2.4 of Penrose (2003) to obtain a slightly improved bound that does not have the logarithmic factor. It appears that this improvement is marginal in many applications. For example, the quantity  $\sqrt{nd_{mx}^2d_{av}\mu_4^4}$  is asymptotically dominated by  $\sqrt{nd_{mx}d_{av}\mu_3^3}$ , when  $\mu_p^p = O_P(n^{-p/2})$  and  $d_{mx}$  increases with  $n$  slower than the rate  $n^{1/2}$ .

### 3.1.2. Conditional Neighborhood Dependence Conditional on High-Degree Vertices.

Let us extend Theorem 3.1 by considering normal approximation conditioning on random variables associated with high degree vertices.

Let  $N_n^* \subset N_n$  be a given subset, and  $\nu_n^*$  the neighborhood system on  $N_n$  given as in (2.2). Let

$$(3.2) \quad d_{mx}^* \equiv \max_{i \in N_n^*} |\nu_n^*(i)|, \text{ and } d_{av}^* \equiv \frac{1}{n^*} \sum_{j \in N_n^*} |\nu_n^*(j)|,$$

where  $n^* \equiv |N_n^*|$ . Hence,  $d_{mx}^*$  and  $d_{av}^*$  are the maximum and average degrees of the restriction of  $\nu_n$  to  $N_n^*$ . Moreover, define  $\mathcal{M}^* = (\mathcal{M}_i^*)_{i \in N_n}$ , where  $\mathcal{M}_i^* = \mathcal{M}_i \vee \mathcal{M}_{N_n \setminus N_n^*}$ , and write  $\mathcal{M}_{\nu_n^*(i)}^* = \mathcal{M}_{\nu_n^*(i)} \vee \mathcal{M}_{N_n \setminus N_n^*}$ . We have in mind choosing  $N_n^*$  so that the set  $N_n \setminus N_n^*$  consists only of high-degree vertices in the neighborhood system  $\nu_n$ . Then by Lemma 2.2, if  $(X_i)_{i \in N_n}$  is CND with respect to  $(\nu_n, \mathcal{M})$ , then  $(X_i)_{i \in N_n^*}$  is CND with respect to  $(\nu_n^*, \mathcal{M}^*)$ . The main idea is that if the difference between the two sums

$$(3.3) \quad \sum_{i \in N_n^*} (X_i - \mathbf{E}[X_i | \mathcal{M}_{\nu_n^*(i)}^*]) \text{ and } \sum_{i \in N_n} (X_i - \mathbf{E}[X_i | \mathcal{M}_{\nu_n(i)}])$$

is asymptotically negligible, we can use the Berry-Esseen bound for the first sum using Theorem 3.1 and deal with the remainder term that comes from the difference.

Now let us present an extended version of Theorem 3.1. Let

$$\mu_p^* \equiv \max_{i \in N_n^*} \left( \mathbf{E} \left[ \left| \frac{X_i}{\sigma_n^*} \right|^p \middle| \mathcal{G}_n^* \right] \right)^{1/p}, \text{ and } \tilde{\mu}_p^* \equiv \max_{i \in N_n \setminus N_n^*} \left( \mathbf{E} \left[ \left| \frac{X_i}{\sigma_n^*} \right|^p \middle| \mathcal{G}_n^* \right] \right)^{1/p},$$

where  $\sigma_n^{*2} \equiv \text{Var}(\sum_{i \in N_n^*} X_i | \mathcal{G}_n^*)$ , and

$$(3.4) \quad \mathcal{G}_n^* \equiv \mathcal{G} \vee \mathcal{M}_{N_n \setminus N_n^*}.$$

Note that the domain of the maximum in the definition of  $\mu_p^*$  is  $N_n^*$  whereas that of  $\tilde{\mu}_p^*$  is  $N_n \setminus N_n^*$ . We also define for  $i, j \in N_n^*$ ,

$$r_n^{*2} \equiv \frac{1}{\sigma_n^{*4}} \mathbf{E} \left[ \left( \sum_{i \in N_n^*} \sum_{j \in \bar{\nu}_n(i) \cap N_n^*} \xi_{ij} \right)^2 \middle| \mathcal{G}_n^* \right].$$

**Theorem 3.2.** *Suppose that a triangular array  $\{X_i\}_{i \in N_n}$  is CND with respect to  $(\nu_n, \mathcal{M})$  and  $\mathbf{E}[X_i | \mathcal{M}_{\nu_n(i)}] = 0$ , a.e. for each  $i \in N_n$  as in Theorem 3.1, and let for  $1 \leq r \leq 4$  and  $\varepsilon_n \geq 0$ ,*

$$(3.5) \quad \mathcal{A}_{n,r}(\varepsilon_n) \equiv \{\omega \in \Omega : n^* d_{mx}^* d_{av}^* \mu_3^{*3} \leq 1 \text{ and } (n - n^*) \tilde{\mu}_r^* + \rho_r^* \leq \varepsilon_n\},$$

where

$$\rho_r^* \equiv \left( \mathbf{E} \left[ \left[ \frac{1}{\sigma_n^*} \sum_{i \in N_n^*} \mathbf{E}[X_i | \mathcal{M}_{\nu_n^*(i)}^*] \right]^r \right] \right)^{1/r}.$$

Then there exists an absolute constant  $C > 0$  such that on the event  $\mathcal{A}_n(\varepsilon_n)$  with any constant  $\varepsilon_n > 0$ , for each  $t \in \mathbf{R}$ ,  $n \geq 1$ , and for each  $1 \leq r \leq 4$ ,

$$\begin{aligned} \sup_{t \in \mathbf{R}} \Delta^*(t; \mathcal{G}_n^*) &\leq C \sqrt{n^* d_{mx}^* d_{av}^* \mu_3^{*3}} - C \log(n^* d_{mx}^* d_{av}^* \mu_3^{*3}) \sqrt{n^* d_{mx}^{*2} d_{av}^* \mu_4^{*4} + r_n^{*2}} \\ &+ C \varepsilon_n^{r/(r+1)}, \text{ a.e.,} \end{aligned}$$

where

$$\Delta^*(t; \mathcal{G}_n^*) \equiv \left| P \left\{ \frac{1}{\sigma_n^*} \sum_{i \in N_n} X_i \leq t | \mathcal{G}_n^* \right\} - \Phi(t) \right|.$$

Compared to Theorem 3.1, the bound involves an additional term. This additional term arises because the sum may not be centered around zero when we condition on  $\mathcal{M}_{\nu_n^*(i)}^*$ . If  $|X_i| \leq 1/\sqrt{n}$  and  $(\sigma_n^*)^2 \geq c$  for some  $c > 0$ , we have  $(n - n^*)\tilde{\mu}_r^* = O((n - n^*)/\sqrt{n})$ . Furthermore, if  $X$  has  $\nu_n$  as a conditional dependency graph, we have  $\rho_r^* = 0$  (because  $\mathcal{M}_i = \mathcal{G}$  for all  $i \in N_n$  in this case) and hence as long as

$$(n - n^*)/\sqrt{n} \rightarrow 0,$$

as  $n \rightarrow \infty$ , the third term in the bound vanishes, and the same CLT as in Theorem 3.1 is restored. In this case, if  $(n^*/n)^{1/2}(((n - n^*)/\sqrt{n})^{4/5} + (d_{mx}^* d_{av}^*/\sqrt{n^*})^{1/2})$  converges to zero faster than  $(d_{mx} d_{av}/\sqrt{n})^{1/2}$ , Theorem 3.2 has an improved rate over Theorem 3.1. Such an approximation captures the situation where the neighborhood system  $\nu_n$  has a very small fraction of very high degree vertices.

Such an improvement can still arise generally, even if  $X$  does not have  $\nu_n$  as a conditional dependency graph. To see this, note that for each  $i \in N_n \setminus \bar{\nu}_n(N_n \setminus N_n^*)$ ,  $X_i$  is conditionally independent of  $\mathcal{M}_{N_n \setminus N_n^*}$  given  $\mathcal{M}_{\nu_n(i)}$ , we have

$$\begin{aligned} \mathbf{E}[X_i | \mathcal{M}_{\nu_n^*(i)}^*] &= \mathbf{E}[X_i | \mathcal{M}_{(\nu_n(i) \cap N_n^*) \cup (N_n \setminus N_n^*)}] \\ &= \mathbf{E}[X_i | \mathcal{M}_{\nu_n(i) \cup (N_n \setminus N_n^*)}] = \mathbf{E}[X_i | \mathcal{M}_{\nu_n(i)}] = 0. \end{aligned}$$

Hence

$$\rho_r^* = \left( \mathbf{E} \left[ \left[ \frac{1}{\sigma_n^*} \sum_{i \in \bar{\nu}_n(N_n \setminus N_n^*) \cap N_n^*} \mathbf{E}[X_i | \mathcal{M}_{\nu_n^*(i)}^*] \right]^r \right] \right)^{1/r}.$$

Suppose that  $|X_i| \leq 1/\sqrt{n}$ ,  $(\sigma_n^*)^2 \geq c$  for some  $c > 0$ ,  $n - n^* = O(1)$ , and  $d_{av} = O(1)$  and  $d_{av}^* = O(1)$ . Then we have

$$(3.6) \quad \rho_r^* = O(|\bar{\nu}_n(N_n \setminus N_n^*)|/\sqrt{n}) = O(d_{mx}/\sqrt{n}).$$

The rate in Theorem 3.2 improves on that in Theorem 3.1 because with  $p > 1$ ,  $(\rho_p^*)^{4/5} = O((d_{mx}/\sqrt{n})^{4/5}) = o((d_{mx}/\sqrt{n})^{1/2})$ .

As we shall see later in Section 3.2.5, the approach of CND conditional on high-degree vertices is useful for obtaining stable central limit theorem for empirical processes when the random variables are CND with respect to a neighborhood system having a maximum degree  $d_{mx}$  increasing to infinity as  $n \rightarrow \infty$ .

## 3.2. Empirical Processes and Stable Convergence

**3.2.1. Stable Convergence in Metric Spaces.** Let us first introduce some preliminary results about stable convergence in metric spaces. Let  $(\mathbb{D}, d)$  be a given metric space and define  $\mathcal{B}(\mathbb{D})$  to be the Borel  $\sigma$ -field of  $\mathbb{D}$ . Let  $(\Omega, \mathcal{F}, P)$  be the given probability space, where  $\mathcal{G}$  is a sub  $\sigma$ -field of  $\mathcal{F}$ . Recall that a map  $L : \mathcal{B}(\mathbb{D}) \times \Omega \rightarrow [0, 1]$  called a *Markov kernel* if for each  $\omega \in \Omega$ ,  $L(\cdot, \omega)$  is a Borel probability measure on  $\mathcal{B}(\mathbb{D})$  and for each  $B \in \mathcal{B}(\mathbb{D})$ ,  $L(B, \cdot)$  is  $\mathcal{F}$ -measurable. Following the notation in Häusler and Luschgy (2010), let us define the *marginal* of  $L$  on  $\mathcal{B}(\mathbb{D})$  as

$$PL(B) \equiv \int L(B, \omega) dP(\omega).$$

For a given finite collection  $\{h_1, \dots, h_k\} \subset \mathcal{H}$ , the finite dimensional projection of a Markov kernel  $L$  is defined to be a Markov kernel  $L_{h_1, \dots, h_k} : \mathcal{B}(\mathbf{R}^k) \times \Omega \rightarrow [0, 1]$  such that for any  $B \in \mathcal{B}(\mathbf{R}^k)$ ,

$$L_{h_1, \dots, h_k}(B, \omega) \equiv L(\{y \in \mathbb{D} : (y(h_1), \dots, y(h_k)) \in B\}, \omega).$$

In the spirit of the Hoffman-Jorgensen approach, we consider the following definition of stable convergence for empirical processes. (See Berti, Pratelli, and Rigo (2012), p.2., for a similar definition.)

**Definition 3.1.** Suppose that we are given a sub  $\sigma$ -field  $\mathcal{G}$ , a sequence of  $\mathbb{D}$ -valued stochastic processes  $\zeta_n$ , a Markov kernel  $L$  on  $\mathcal{B}(\mathbb{D}) \times \Omega$ , and a  $\mathbb{D}$ -valued Borel measurable random element  $\zeta$  that has a Markov kernel  $L$ .

Suppose that for each  $U \in \mathcal{G}$ , and each bounded Lipschitz functional  $f$  on  $\mathbb{D}$ ,

$$\mathbf{E}^*[f(\zeta_n)1_U] \rightarrow \int_U \int_{\mathbb{D}} f(y)L(dy, \omega)dP(\omega),$$



as  $n \rightarrow \infty$ , where  $\mathbf{E}^*$  denotes the outer-expectation and  $1_U$  is the indicator function of the event  $U$ . Then we say that  $\zeta_n$  converges to  $L$ ,  $\mathcal{G}$ -stably, (or equivalently,  $\zeta_n$  converges to  $\zeta$ ,  $\mathcal{G}$ -stably), and write

$$\zeta_n \rightarrow L, \mathcal{G}\text{-stably (or equivalently, } \zeta_n \rightarrow \zeta, \mathcal{G}\text{-stably.)}$$

Stable convergence according to Definition 3.1 implies weak convergence (in the sense of Hoffman-Jorgensen). When  $\zeta_n$  is Borel measurable, the above definition is equivalent to the weak convergence of Markov kernels and many of the existing results on stable convergence carry over. However, this equivalence does not extend to the case of  $\zeta_n$  being non-measurable, because there is no proper definition of Markov kernels for nonmeasurable stochastic processes. Nevertheless the above definition can still be useful when one needs to deal with random norming, as shown in the following lemma which generalizes Theorem 1' in Aldous and Eagleson (1978) and part of Theorem 3.18 of Häusler and Luschgy (2010).

**Lemma 3.2.** *Suppose that  $\zeta_n, \zeta, \xi_n, \xi$  are  $\mathbb{D}$ -valued random variables, and that  $\zeta_n \rightarrow \zeta$ ,  $\mathcal{G}$ -stably, where  $\zeta$  and  $\xi$  are Borel measurable, and let  $P_\zeta$  denote the distribution of  $\zeta$ . Then the following holds.*

(i) *If  $P^*\{d(\xi_n, \xi) > \varepsilon\} \rightarrow 0$  as  $n \rightarrow \infty$  for each  $\varepsilon > 0$ , and  $\xi$  is  $\mathcal{G}$ -measurable, then*

$$(\zeta_n, \xi_n) \rightarrow (\zeta, \xi), \mathcal{G}\text{-stably,}$$

where  $P^*$  denotes the outer probability.

(ii) *If  $f : \mathbb{D} \rightarrow \mathbb{D}$  is  $P_\zeta$ -a.e. continuous, then  $f(\zeta_n) \rightarrow f(\zeta)$ ,  $\mathcal{G}$ -stably.*

The first result is a stable-convergence analogue of Cramér-Slutsky lemma. The second result is a continuous mapping theorem.

**3.2.2. Stable Convergence of an Empirical Process.** Suppose that  $\{Y_i\}_{i \in N_n}$  is a given triangular array of  $\mathbf{R}$ -valued random variables which is GND with respect to  $(\nu_n, \mathcal{M})$ . Let  $\mathcal{H}$  be a given class of real measurable functions on  $\mathbf{R}$ , having a measurable envelope  $H$ . Then, we consider the following empirical process:

$$\{\mathbb{G}_n(h) : h \in \mathcal{H}\},$$

where, for each  $h \in \mathcal{H}$ ,

$$\mathbb{G}_n(h) \equiv \frac{1}{\sqrt{n}} \sum_{i \in N_n} (h(Y_i) - \mathbf{E}[h(Y_i) | \mathcal{M}_{\nu_n(i)}]).$$

The empirical process  $\nu_n$  takes a value in  $l^\infty(\mathcal{H})$ , the collection of bounded functions on  $\mathcal{H}$  which is endowed with the sup norm so that  $(l^\infty(\mathcal{H}), \|\cdot\|_\infty)$  forms the metric space  $(\mathbb{D}, d)$

with  $\|h\|_\infty \equiv \sup_{y \in \mathbf{R}} |h(y)|$ . In this section, we explore conditions for the class  $\mathcal{H}$  and the joint distribution of the triangular array  $\{Y_i\}_{i \in N_n}$  which delivers the stable convergence of the empirical process. Stable convergence in complete separable metric spaces can be defined as a weak convergence of Markov kernels. (See Häusler and Luschgy (2010).) However, this definition does not extend to the case of empirical processes taking values in  $\mathbb{D}$  that is endowed with the sup norm, due to non-measurability.

Weak convergence of an empirical process to a Gaussian process is often established in three steps. First, we show that the class of functions is totally bounded with respect to a certain pseudo-metric  $\rho$ . Second, we show that each finite dimensional projection of the empirical process converges in distribution to a multivariate normal random vector. Third, we establish the asymptotic  $\rho$ -equicontinuity of the empirical process.

Let  $\rho$  be a given pseudo-metric on  $\mathcal{H}$  such that  $(\mathcal{H}, \rho)$  is a totally bounded metric space. Then we define

$$U_\rho(\mathcal{H}) \equiv \{y \in \mathbb{D} : y \text{ is uniformly } \rho\text{-continuous on } \mathcal{H}\}.$$

The following theorem shows that we can take a similar strategy in proving the stable convergence of an empirical process to a Markov kernel. The structure and the proof of the theorem is adapted from Theorem 10.2 of Pollard (1990).

**Theorem 3.3.** *Suppose that the stochastic process  $\zeta_n \in \mathbb{D}$  is given, where  $(\mathcal{H}, \rho)$  is a totally bounded metric space. Suppose that the following conditions hold.*

- (i) *For each finite set  $\{h_1, \dots, h_k\} \subset \mathcal{H}$ ,  $(\zeta_n(h_1), \dots, \zeta_n(h_k)) \rightarrow L_{h_1, \dots, h_k}$ ,  $\mathcal{G}$ -stably, where  $L_{h_1, \dots, h_k}$  is a Markov kernel on  $\mathcal{B}(\mathbf{R}^k) \times \Omega$ .*
- (ii)  *$\zeta_n$  is asymptotically  $\rho$ -equicontinuous on  $\mathcal{H}$ , i.e., for each  $\varepsilon > 0, \eta > 0$ , there exists  $\delta > 0$  such that*

$$\limsup_{n \rightarrow \infty} P^* \left\{ \sup_{h, h' \in \mathcal{H}: \rho(h, h') < \delta} |\zeta_n(h) - \zeta_n(h')| > \eta \right\} < \varepsilon,$$

where  $P^*$  denotes the outer probability.

Then there exists a Markov kernel  $L$  on  $\mathcal{B}(\mathbb{D}) \times \Omega$  such that the following properties are satisfied.

- (a) *The finite dimensional projections of  $L$  are given by Markov kernels  $L_{h_1, \dots, h_k}$ .*
- (b)  *$PL(U_\rho(\mathcal{H})) = 1$ .*
- (c)  *$\zeta_n \rightarrow L$ ,  $\mathcal{G}$ -stably.*

Conversely, if  $\zeta_n$  converges to Markov kernel  $L$  on  $\mathcal{B}(\mathbb{D}) \times \Omega$ ,  $\mathcal{G}$ -stably, where  $PL(U_\rho(\mathcal{H})) = 1$ , then (i) and (ii) are satisfied.

It is worth noting that for stable convergence of empirical processes, the conditions for the asymptotic equicontinuity and the totally boundedness of the function class with respect to a pseudo-norm are as in the standard literature on weak convergence of empirical processes. The only difference here is that the convergence of finite dimensional distributions is now replaced by the stable convergence of finite dimensional projections.

**3.2.3. Maximal Inequality.** This subsection presents a maximal inequality in terms of bracketing entropy bounds. The maximal inequality is useful primarily for establishing asymptotic  $\rho$ -equicontinuity of the empirical process but also for many other purposes. First, we begin with a tail bound for a sum of CND random variables. Janson (2004) established an exponential bound for a sum of random variables that have a dependency graph. The following exponential tail bound is crucial for our maximal inequality. The result below is obtained by slightly modifying the proof of Theorem 2.3 of Janson (2004).

**Lemma 3.3.** *Suppose that  $\{X_i\}_{i \in N_n}$  is a triangular array of random variables that take values in  $[-M, M]$  and are CND with respect to  $(\nu_n, \mathcal{M})$ , with  $\mathbf{E}[X_i | \mathcal{M}_{\nu_n(i)}] = 0$ , and let  $\sigma_i^2 \equiv \text{Var}(X_i | \mathcal{M}_{\nu_n(i)})$  and  $V_n \equiv \sum_{i \in N_n} \mathbf{E}[\sigma_i^2 | \mathcal{G}]$  with  $\mathcal{G}$  as defined in (2.1).*

*Then, for any  $\eta > 0$ ,*

$$(3.7) \quad P \left\{ \left| \sum_{i \in N_n} X_i \right| \geq \eta | \mathcal{G} \right\} \leq 2 \exp \left( - \frac{\eta^2}{2(d_{mx} + 1)(2(d_{mx} + 1)V_n + M\eta/3)} \right), \text{ a.e.},$$

*for all  $n \geq 1$ .*

*Furthermore, if Condition A holds and the condition  $\mathbf{E}[X_i | \mathcal{M}_{\nu_n(i)}] = 0$  is replaced by  $\mathbf{E}[X_i | \mathcal{G}] = 0$  and the  $\sigma$ -fields  $\mathcal{M}_{\nu_n(i)}$  in  $\sigma_i$ 's are replaced by  $\mathcal{G}$ , then the following holds: for any  $\eta > 0$ ,*

$$(3.8) \quad P \left\{ \left| \sum_{i \in N_n} X_i \right| \geq \eta | \mathcal{G} \right\} \leq 2 \exp \left( - \frac{8\eta^2}{25(d_{mx} + 1)(V_n + M\eta/3)} \right), \text{ a.e.},$$

*for all  $n \geq 1$ .*

The bound in (3.8) is the one obtained by Janson (2004) for the case of dependency graphs. From this, the following form of maximal inequality for a finite set immediately follows from Lemma A.1 of van der Vaart (1996).

**Corollary 3.2.** *Suppose that  $\{Y_i\}_{i \in N_n}$  is a triangular array of random variables that are CND with respect to  $(\nu_n, \mathcal{M})$ . Let for each  $h \in \mathcal{H}$ ,  $V_n(h) \equiv n^{-1} \sum_{i \in N_n} \mathbf{E}[\sigma_i^2(h) | \mathcal{G}]$  and  $\sigma_i^2(h) \equiv \text{Var}(h(Y_i) | \mathcal{M}_{\nu_n(i)})$  with  $\mathcal{G}$  as defined in (2.1).*

Then there exists an absolute constant  $C > 0$  such that

$$\begin{aligned} & \mathbf{E} \left[ \max_{1 \leq s \leq m} |\mathbb{G}_n(h_s)| \middle| \mathcal{G} \right] \\ & \leq C(d_{mx} + 1) \left( \frac{J}{\sqrt{n}} \log(1 + m) + \sqrt{\log(1 + m) \max_{1 \leq s \leq m} V_n(h_s)} \right), \text{ a.e.}, \end{aligned}$$

for any  $n \geq 1$  and any  $m \geq 1$  with a finite subset  $\{h_1, \dots, h_m\}$  of  $\mathcal{H}$  such that for some constant  $J > 0$ ,  $\max_{1 \leq s \leq m} \sup_{x \in \mathbf{R}} |h_s(x)| \leq J$ .

Let us now elevate the above inequality to a maximal inequality over function class  $\mathcal{H}$ . Recall that we allow the random variables  $Y_i$ 's to be idiosyncratically distributed across  $i$ 's.

We define the following semi-norm on  $\mathcal{H}$ :

$$\bar{\rho}_n(h) \equiv \sqrt{\frac{1}{n} \sum_{i \in N_n} \mathbf{E}[h^2(Y_i)]}.$$

We denote  $N_{[]}(\varepsilon, \mathcal{H}, \bar{\rho}_n)$  to be the  $\varepsilon$ -bracketing number of  $\mathcal{H}$  with respect to  $\bar{\rho}_n$ , i.e., the smallest number  $J$  of the brackets  $[h_{L,j}, h_{U,j}]$ ,  $j = 1, \dots, J$ , such that  $\bar{\rho}_n(h_{U,j} - h_{L,j}) \leq \varepsilon$ . The following lemma establishes the maximal inequality in terms of a bracketing entropy bound.

**Lemma 3.4 (Maximal Inequality).** *Suppose that  $\{Y_i\}_{i \in N_n}$  is a triangular array of random variables that are *CND* with respect to  $(\nu_n, \mathcal{M})$ . Suppose further that the class  $\mathcal{H}$  of functions have an envelope  $H$  such that  $\bar{\rho}_n(H) < \infty$ . Then, there exists an absolute constant  $C > 0$  such that for each  $n \geq 1$ ,*

$$\mathbf{E}^* \left[ \sup_{h \in \mathcal{H}} |\mathbb{G}_n(h)| \right] \leq C(1 + d_{mx}) \int_0^{\bar{\rho}_n(H)} \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{H}, \bar{\rho}_n)} d\varepsilon.$$

The bracketing entropy bound in Lemma 3.4 involves the maximum degree  $d_{mx}$ . Hence the bound is useful only when the neighborhood system  $\nu_n$  does not have a maximum degree increasing with  $n$ .

**3.2.4. Stable Central Limit Theorem.** First, let us say that a stochastic process  $\{\mathbb{G}(h) : h \in \mathcal{H}\}$  is a  $\mathcal{G}$ -mixture Gaussian process if for any finite collection  $\{h_1, \dots, h_m\} \subset \mathcal{H}$ , the distribution of random vector  $[\mathbb{G}(h_1), \mathbb{G}(h_2), \dots, \mathbb{G}(h_m)]$  conditional on  $\mathcal{G}$  is a multivariate normal distribution. Also, we call a Markov kernel  $\mathcal{K}$  a  $\mathcal{G}$ -mixture Gaussian Markov kernel associated with a given  $\mathcal{G}$ -mixture Gaussian process  $\mathbb{G}$  if for any  $h_1, \dots, h_m \in \mathcal{H}$ , the conditional distribution of  $[\mathbb{G}(h_1), \mathbb{G}(h_2), \dots, \mathbb{G}(h_m)]$  given  $\mathcal{G}$  is given by the finite dimensional projection  $\mathcal{K}_{h_1, \dots, h_m}$  of  $\mathcal{K}$ . Let us summarize the conditions as follows.

**Assumption 3.1.** (a) There exists  $C > 0$  such that for all  $n \geq 1$ ,

$$(1 + d_{mx}) \int_0^{\bar{\rho}_n(H)} \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{H}, \bar{\rho}_n)} d\varepsilon < C.$$

(b) For any  $h_1, h_2 \in \mathcal{H}$ ,

$$\mathbf{E}[\mathbb{G}_n(h_1)\mathbb{G}_n(h_2)|\mathcal{G}] \rightarrow_P K(h_1, h_2|\mathcal{G}),$$

for some  $K(\cdot, \cdot|\mathcal{G})(\omega) : \mathcal{H} \times \mathcal{H} \rightarrow \mathbf{R}, \omega \in \Omega$ , which is positive semidefinite a.e., and non-constant at zero.

(c) For each  $h \in \mathcal{H}$ ,  $-n^{-1/2}r_n(h) \log(n^{-1/2}d_{mx}d_{av}) \rightarrow_P 0$ , as  $n \rightarrow \infty$ , where

$$r_n^2(h) \equiv \mathbf{E} \left[ \left( \frac{1}{\sqrt{n}} \sum_{i \in N_n} \sum_{j \in \bar{\nu}_n(i)} \xi_{ij}(h) \right)^2 \middle| \mathcal{G} \right],$$

with

$$\begin{aligned} \xi_{ij}(h) &\equiv \mathbf{E}[(h(X_i) - \mathbf{E}[h(X_i)|\mathcal{M}_{\nu_n(i)}])(h(X_j) - \mathbf{E}[h(X_j)|\mathcal{M}_{\nu_n(j)}])|\mathcal{M}_{\nu_n(i,j)}] \\ &\quad - \mathbf{E}[(h(X_i) - \mathbf{E}[h(X_i)|\mathcal{M}_{\nu_n(i)}])(h(X_j) - \mathbf{E}[h(X_j)|\mathcal{M}_{\nu_n(j)}])|\mathcal{G}]. \end{aligned}$$

(d) For each  $h \in \mathcal{H}$ ,  $\rho(h) \equiv \lim_{n \rightarrow \infty} \bar{\rho}_n(h)$  exists and satisfies that whenever  $\rho(h_n) \rightarrow 0$  as  $n \rightarrow \infty$ ,  $\bar{\rho}_n(h_n) \rightarrow 0$  as  $n \rightarrow \infty$  as well.

The following result gives a Donsker-type stable convergence of empirical processes.

**Theorem 3.4.** *Suppose that  $\{Y_i\}_{i \in N_n}$  is a triangular array of random variables that are  $CND$  with respect to  $(\nu_n, \mathcal{M})$ , satisfying Assumption 3.1. Suppose further that there exists  $C > 0$  such that for each  $n \geq 1$ ,  $\max_{i \in N_n} \mathbf{E}[H(Y_i)^4] < C$ , where  $H$  is an envelope of  $\mathcal{H}$ .*

*Then  $\nu_n$  converges to a  $\mathcal{G}$ -mixture Gaussian process  $\mathbb{G}$  in  $l^\infty(\mathcal{H})$ ,  $\mathcal{G}$ -stably, such that for any  $h_1, h_2 \in \mathcal{H}$ ,  $\mathbf{E}[\mathbb{G}(h_1)\mathbb{G}(h_2)|\mathcal{G}] = K(h_1, h_2|\mathcal{G})$ , a.e.*

*Furthermore, we have  $PK(U_\rho(\mathcal{H})) = 1$ , where  $\mathcal{K}$  is the  $\mathcal{G}$ -mixture Gaussian Markov kernel associated with  $\mathbb{G}$ .*

The fourth moment condition  $\max_{i \in N_n} \mathbf{E}[H(Y_i)^4] < \infty$  is used to ensure the convergence of finite dimensional distributions using Theorem 3.1. It is worth noting that Assumption 3.1(a) essentially requires that the maximum degree  $d_{mx}$  to be bounded. It is interesting that this condition was not required for the CLT in Theorem 3.1. This stronger condition for the maximum degree is used to establish the asymptotic equicontinuity of the process  $\{\mathbb{G}_n(h) : h \in \mathcal{H}\}$ . When the neighborhood system  $\nu_n$  is generated according to a model of stochastic graph formation, this condition is violated for many existing

models of graph formation used, for example, for social network modeling. In the next section, we utilize the approach of conditioning on high-degree vertices to weaken this condition.

### 3.2.5. Conditional Neighborhood Dependence Conditional on High-Degree Vertices.

As mentioned before, Assumption 3.1(a) requires that  $d_{mx}$  be bounded. Following the idea of conditioning on high degree vertices as in Theorem 3.2, let us explore a stable convergence theorem that relaxes this requirement. As we did prior to Theorem 3.2, we choose  $N_n^* \subset N_n$  to be a given subset and let  $d_{mx}^*$  and  $d_{av}^*$  be as defined in (3.2).

First, write

$$(3.9) \quad \mathbb{G}_n(h) = \mathbb{G}_n^*(h) + R_n^*(h) + \rho_n^*(h),$$

where

$$\begin{aligned} \mathbb{G}_n^*(h) &\equiv \frac{1}{\sqrt{n}} \sum_{i \in N_n^*} (h(Y_i) - \mathbf{E}[h(Y_i) | \mathcal{M}_{\nu_n^*(i)}^*]) \\ R_n^*(h) &\equiv \frac{1}{\sqrt{n}} \sum_{i \in N_n \setminus N_n^*} (h(Y_i) - \mathbf{E}[h(Y_i) | \mathcal{M}_{\nu_n(i)}]), \text{ and} \\ \rho_n^*(h) &\equiv \frac{1}{\sqrt{n}} \sum_{i \in N_n^*} (\mathbf{E}[h(Y_i) | \mathcal{M}_{\nu_n(i)}] - \mathbf{E}[h(Y_i) | \mathcal{M}_{\nu_n^*(i)}^*]). \end{aligned}$$

Note that

$$\mathbf{E}^* \left[ \sup_{h \in \mathcal{H}} |R_n^*(h)| \right] \leq \frac{2(n - n^*)}{\sqrt{n}} \max_{i \in N_n} \sqrt{\mathbf{E}[H^2(Y_i)]}.$$

Since  $\{Y_i\}_{i \in N_n^*}$  is *CND* with respect to  $(\nu_n^*, \mathcal{M}^*)$  as defined in (2.2) and (2.3), we can apply the previous results to  $\mathbb{G}_n^*(h)$ . This gives the following extension of the maximal inequality in Lemma 3.4. Since the maximal inequality is often of independent interest, let us state it formally.

**Lemma 3.5** (Maximal Inequality). *Suppose that  $\{Y_i\}_{i \in N_n}$  is a triangular array of random variables that are *CND* with respect to  $(\nu_n, \mathcal{M})$ . Suppose further that the class  $\mathcal{H}$  of functions have an envelope  $H$  such that  $\bar{\rho}_n(H) < \infty$ .*

*Then, there exists an absolute constant  $C > 0$  such that for all  $n \geq 1$ ,*

$$\begin{aligned} \mathbf{E}^* \left[ \sup_{h \in \mathcal{H}} |\mathbb{G}_n(h)| \right] &\leq \frac{C\sqrt{n^*}(1 + d_{mx}^*)}{\sqrt{n}} \int_0^{\bar{\rho}_n(H)} \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{H}, \bar{\rho}_n)} d\varepsilon \\ &\quad + C \left( \frac{n - n^*}{\sqrt{n}} \max_{i \in N_n} \sqrt{\mathbf{E}[H^2(Y_i)]} + \mathbf{E}^* \left[ \sup_{h \in \mathcal{H}} \rho_n^*(h) \right] \right). \end{aligned}$$

If  $\|H\|_\infty < C$  and  $n - n^* = O(1)$ , the second term in the bound is  $O(d_{mx}/\sqrt{n})$  similarly as we derived in (3.6). Thus this bound is an improvement over Lemma 3.4, whenever  $O(d_{mx}^*) = o(d_{mx})$  as  $n \rightarrow \infty$ . Let us turn to the Donsker-type stable convergence of an empirical process. We modify Assumption 3.1 as follows.

**Assumption 3.2.** (a) There exists  $C > 0$  such that for all  $n \geq 1$ ,

$$(1 + d_{mx}^*) \int_0^{\bar{\rho}_n(H)} \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{H}, \bar{\rho}_n)} d\varepsilon < C.$$

(b) For any  $h_1, h_2 \in \mathcal{H}$ ,

$$\mathbf{E}[\mathbb{G}_n(h_1)\mathbb{G}_n(h_2)|\mathcal{G}_n^*] \rightarrow_P K(h_1, h_2|\mathcal{G}^*),$$

for some  $K(\cdot, \cdot|\mathcal{G}^*)(\omega) : \mathcal{H} \times \mathcal{H} \rightarrow \mathbf{R}, \omega \in \Omega$ , which is positive semidefinite a.e, and non-constant at zero, and for some sub  $\sigma$ -field  $\mathcal{G}^*$  of  $\mathcal{F}$ , where  $\mathcal{G}_n^*$  is as defined in (3.4).

(c) For each  $h \in \mathcal{H}$ ,  $-n^{*-1/2}r_n^*(h) \log(n^{*-1/2}d_{mx}^*d_{av}^*) \rightarrow_P 0$ , as  $n \rightarrow \infty$ , where

$$r_n^{*2}(h) \equiv \mathbf{E} \left[ \left( \frac{1}{\sqrt{n}} \sum_{i \in N_n^*} \sum_{j \in \bar{\nu}_n(i) \cap N_n^*} \xi_{ij}(h) \right)^2 \middle| \mathcal{G}_n^* \right].$$

(d) For each  $h \in \mathcal{H}$ ,  $\rho(h) \equiv \lim_{n \rightarrow \infty} \bar{\rho}_n(h)$  exists and satisfies that whenever  $\rho(h_n) \rightarrow 0$  as  $n \rightarrow \infty$ ,  $\bar{\rho}_n(h_n) \rightarrow 0$  as  $n \rightarrow \infty$  as well.

(e)  $\mathbf{E}^* [\sup_{h \in \mathcal{H}} \rho_n^*(h)] \rightarrow 0$ , as  $n \rightarrow \infty$ .

While Condition (a) essentially requires that  $d_{mx}^*$  be bounded, Condition (c) allows  $d_{mx}$  to increase to infinity as  $n \rightarrow \infty$ . The condition in (b) that  $K(\cdot, \cdot|\mathcal{G}^*)$  be non-constant at zero requires that

$$n - n^* \rightarrow \lambda, \text{ as } n \rightarrow \infty,$$

for some  $\lambda \in [0, \infty)$ . Thus the number of the high degree vertices  $(n - n^*)$  selected when we set  $N_n^* \subset N_n$  should be bounded as  $n \rightarrow \infty$ . In combination with (e), this implies that we have  $(n - n^*)/\sqrt{n} + \mathbf{E}^* [\sup_{h \in \mathcal{H}} \rho_n^*(h)] \rightarrow 0$  as  $n \rightarrow \infty$ , which makes it suffice to focus on  $\mathbb{G}_n^*(h)$  in (3.9) for a stable limit theorem. We obtain the following extended version of Theorem 3.4.

**Theorem 3.5.** *Suppose that  $\{Y_i\}_{i \in N_n}$  is a triangular array of random variables that are *CND* with respect to  $(\nu_n, \mathcal{M})$ , satisfying Assumption 3.2. Suppose further that there exists  $C > 0$  such that for each  $n \geq 1$ ,  $\max_{i \in N_n} \mathbf{E}[H(Y_i)^4] < C$ , where  $H$  is an envelope of  $\mathcal{H}$ .*

*Then  $\nu_n$  converges to a  $\mathcal{G}^*$ -mixture Gaussian process  $\mathbb{G}^*$  in  $l^\infty(\mathcal{H})$ ,  $\mathcal{G}^*$ -stably, such that for any  $h_1, h_2 \in \mathcal{H}$ ,  $\mathbf{E}[\mathbb{G}^*(h_1)\mathbb{G}^*(h_2)|\mathcal{G}^*] = K(h_1, h_2|\mathcal{G}^*)$ , a.e.*

Furthermore, we have  $PK(U_\rho(\mathcal{H})) = 1$ , where  $\mathcal{K}$  is the  $\mathcal{G}^*$ -mixture Gaussian Markov kernel associated with  $\mathbb{G}^*$ .

If we take  $N_n^*$  to be identical to  $N_n$ , Theorem 3.5 is reduced to Theorem 3.4. However, Theorem 3.5 shows that approximation of the distribution of an empirical process by a mixture Gaussian process is possible even if  $d_{mx} \rightarrow \infty$  as  $n \rightarrow \infty$ .

## 4. Appendix: Mathematical Proofs

To simplify the notation, we follow van der Vaart (1996) and write  $a_n \lesssim b_n$  for any sequence of numbers, whenever  $a_n \leq Cb_n$  for all  $n \geq 1$  with some absolute constant  $C > 0$ . The absolute constant can differ across different instances of  $\lesssim$ .

For any positive integer  $k$  and  $\mathbf{i} \equiv (i_1, \dots, i_k) \in N_n^k$  and a triangular array of random variables  $\{X_i\}_{i \in N_n}$ , we define

$$(4.1) \quad X(\mathbf{i}) \equiv \prod_{r=1}^k X_{i_r}.$$

The following lemma is useful for the proofs of various results.

**Lemma 4.1.** *Suppose that  $\nu_n$  is a neighborhood system on  $N_n$  and  $\{X_i\}_{i \in N_n}$  is a triangular array of random variables that are *CND* with respect to  $(\nu_n, \mathcal{M})$ , where  $\mathcal{M} = \{\mathcal{M}_i\}_{i \in N_n}$ . Furthermore, for given positive integer  $k$ , let  $\mathbf{i} = (i_1, \dots, i_k) \in N_n^k$  be such that it has two partitioning subvectors  $I_{k,1}(\mathbf{i})$  and  $I_{k,2}(\mathbf{i})$  of  $\mathbf{i}$  such that the entries of  $I_{k,1}(\mathbf{i})$  are from  $N_n \setminus \bar{\nu}_n(I_{k,2}(\mathbf{i}))$  and the entries of  $I_{k,2}(\mathbf{i})$  are from  $N_n \setminus \bar{\nu}_n(I_{k,1}(\mathbf{i}))$ . Then,*

$$\mathbf{E}[X(\mathbf{i})|\mathcal{M}_{\nu_n(\mathbf{i})}] = \mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{M}_{\nu_n(I_{k,1}(\mathbf{i}))}]\mathbf{E}[X(I_{k,2}(\mathbf{i}))|\mathcal{M}_{\nu_n(I_{k,2}(\mathbf{i}))}].$$

Suppose further that Condition A holds. Then,

$$\mathbf{E}[X(\mathbf{i})|\mathcal{G}] = \mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{G}]\mathbf{E}[X(I_{k,2}(\mathbf{i}))|\mathcal{G}].$$

**Proof:** By the choice of  $\mathbf{i}$ , we have

$$(4.2) \quad \begin{aligned} \nu_n(\mathbf{i}) &= \nu_n(I_{k,1}(\mathbf{i})) \cup \nu_n(I_{k,2}(\mathbf{i})), \\ \nu_n(I_{k,1}(\mathbf{i})) &\subset \nu_n(\mathbf{i}), \text{ and } \nu_n(I_{k,2}(\mathbf{i})) \subset \nu_n(\mathbf{i}). \end{aligned}$$

To see the second statement, note that whenever  $i \in \nu_n(I_{k,1}(\mathbf{i}))$ , we have  $i \notin I_{k,1}(\mathbf{i})$  and  $i \notin I_{k,2}(\mathbf{i})$ , and there must exist  $j \in I_{k,1}(\mathbf{i})$  such that  $i \in \nu_n(j)$ . Since  $I_{k,1}(\mathbf{i}) \subset \mathbf{i}$ , we find that  $i \in \nu_n(\mathbf{i})$ .



As for the first statement of the lemma, we write

$$\begin{aligned}
\mathbf{E}[X(\mathbf{i})|\mathcal{M}_{\nu_n(\mathbf{i})}] &= \mathbf{E}[\mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{M}_{\nu_n(\mathbf{i})}, X(I_{k,2}(\mathbf{i}))]X(I_{k,2}(\mathbf{i}))|\mathcal{M}_{\nu_n(\mathbf{i})}] \\
&= \mathbf{E}[\mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{M}_{\nu_n(I_{k,1}(\mathbf{i}))\cup\nu_n(I_{k,2}(\mathbf{i}))}, X(I_{k,2}(\mathbf{i}))]X(I_{k,2}(\mathbf{i}))|\mathcal{M}_{\nu_n(\mathbf{i})}] \\
&= \mathbf{E}[\mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{M}_{\nu_n(I_{k,1}(\mathbf{i}))}]X(I_{k,2}(\mathbf{i}))|\mathcal{M}_{\nu_n(\mathbf{i})}] \\
&= \mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{M}_{\nu_n(I_{k,1}(\mathbf{i}))}]\mathbf{E}[X(I_{k,2}(\mathbf{i}))|\mathcal{M}_{\nu_n(\mathbf{i})}] \\
&= \mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{M}_{\nu_n(I_{k,1}(\mathbf{i}))}]\mathbf{E}[X(I_{k,2}(\mathbf{i}))|\mathcal{M}_{\nu_n(I_{k,2}(\mathbf{i}))}].
\end{aligned}$$

The second equality follows by (4.2). The third equality follows because  $\nu_n(I_{k,2}(\mathbf{i})) \setminus \bar{\nu}_n(I_{k,1}(\mathbf{i}))$  and  $I_{k,2}(\mathbf{i})$  are outside  $\bar{\nu}_n(I_{k,1}(\mathbf{i}))$  and  $X(I_{k,1}(\mathbf{i}))$  is conditionally independent of  $(\mathcal{M}_{\nu_n(I_{k,2}(\mathbf{i})) \setminus \bar{\nu}_n(I_{k,1}(\mathbf{i}))}, X(I_{k,2}(\mathbf{i})))$  given  $\mathcal{M}_{\nu_n(I_{k,1}(\mathbf{i}))}$  by the CND property of  $\{X_i\}$ . The fourth equality follows because  $\nu_n(I_{k,1}(\mathbf{i})) \subset \nu_n(\mathbf{i})$ . The fifth equality uses the fact that  $\nu_n(I_{k,1}(\mathbf{i})) \setminus \bar{\nu}_n(I_{k,2}(\mathbf{i}))$  is outside of  $\bar{\nu}_n(I_{k,2}(\mathbf{i}))$  and the CND property of  $\{X_i\}_{i \in N_n}$ .

Let us turn to the second statement of the lemma and now assume that Condition A holds. We write

$$\begin{aligned}
\mathbf{E}[X(\mathbf{i})|\mathcal{G}] &= \mathbf{E}[\mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{M}_{\nu_n(I_{k,1}(\mathbf{i}))}, X(I_{k,2}(\mathbf{i}))]X(I_{k,2}(\mathbf{i}))|\mathcal{G}] \\
&= \mathbf{E}[\mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{M}_{\nu_n(I_{k,1}(\mathbf{i}))}]X(I_{k,2}(\mathbf{i}))|\mathcal{G}] \\
&= \mathbf{E}[X(I_{k,1}(\mathbf{i}))|\mathcal{G}]\mathbf{E}[X(I_{k,2}(\mathbf{i}))|\mathcal{G}].
\end{aligned}$$

The second equality follows because  $I_{k,2}(\mathbf{i})$  is outside  $\bar{\nu}_n(I_{k,1}(\mathbf{i}))$  and by the CND property. The third equality follows by Condition A, i.e.,  $\mathcal{M}_{\nu_n(I_{k,1}(\mathbf{i}))}$  and  $X(I_{k,2}(\mathbf{i}))$  are conditionally independent given  $\mathcal{G}$ . ■

Let us present the proof of Theorem 3.1. Recall the notation in the theorem, and define  $X_i^\sigma \equiv X_i/\sigma_n$ ,  $W^\sigma \equiv \sum_{i \in N_n} X_i^\sigma$ , and  $W_i^\sigma \equiv \sum_{j \in \bar{\nu}_n(i)} X_j^\sigma$ . Let us define for  $t \in \mathbf{R}$  and  $\omega \in \Omega$ ,

$$\varphi_n(t)(\omega) \equiv \mathbf{E}[\exp(itW^\sigma)|\mathcal{G}](\omega),$$

where  $\mathbf{i} \equiv \sqrt{-1}$ . Note that  $\varphi_n$  is uniformly continuous on  $\mathbf{R}$  almost surely, and since  $\mathbf{E}[(W^\sigma)^2|\mathcal{G}] = 1$ , a.e.,  $\varphi_n$  is twice continuously differentiable almost surely. (See Yuan and Lei (2016).)

**Lemma 4.2.** *Suppose that the conditions of Theorem 3.1 hold. Then for each  $t \in \mathbf{R}$ ,*

$$|\varphi'_n(t) + t\varphi_n(t)| \lesssim \left( nt^2 d_{mx} d_{av} \mu_3^3 + t \sqrt{nd_{mx}^2 d_{av} \mu_4^4 + r_n^2} \right), \text{ a.e.}$$

**Proof:** First, as in the proof of Theorem 1 in Jenish and Prucha (2009), we decompose

$$(it - W^\sigma) \exp(itW^\sigma) = h_{1,t}(W^\sigma) - h_{2,t}(W^\sigma) - h_{3,t}(W^\sigma),$$

where

$$\begin{aligned} h_{1,t}(W^\sigma) &\equiv ite^{itW^\sigma} \left( 1 - \sum_{j \in N_n} X_j^\sigma W_j^\sigma \right), \\ h_{2,t}(W^\sigma) &\equiv e^{itW^\sigma} \sum_{j \in N_n} X_j^\sigma (1 - itW_j^\sigma - e^{-itW_j^\sigma}), \text{ and} \\ h_{3,t}(W^\sigma) &\equiv e^{itW^\sigma} \sum_{j \in N_n} X_j^\sigma e^{-itW_j^\sigma}. \end{aligned}$$

Now, let us consider

$$\begin{aligned} \mathbf{E}[h_{1,t}(W^\sigma)^2 | \mathcal{G}] &\leq t^2 \mathbf{E} \left[ \left( 1 - \sum_{i \in N_n} X_i^\sigma W_i^\sigma \right)^2 | \mathcal{G} \right] \\ &= t^2 \mathbf{E} \left[ \left( \sum_{i \in N_n} \sum_{j \in \bar{\nu}_n(i)} X_i^\sigma X_j^\sigma \right)^2 | \mathcal{G} \right] - t^2, \end{aligned}$$

because

$$\sum_{i \in N_n} \sum_{j \in \bar{\nu}_n(i)} \mathbf{E} [X_i^\sigma X_j^\sigma | \mathcal{G}] = 1,$$

by the definition of  $\sigma_n^2$ .

Define  $\sigma_{ij} \equiv \mathbf{E}[X_i^\sigma X_j^\sigma | \mathcal{G}]$  and  $\sigma_{ij}^* \equiv \mathbf{E}[X_i^\sigma X_j^\sigma | \mathcal{M}_{\nu_n(i,j)}]$ . Note that for  $\{i, j\}, \{k, l\} \subset N_n$  such that  $\{k, l\} \subset N_n \setminus \bar{\nu}_n(i, j)$  and  $\{i, j\} \subset N_n \setminus \bar{\nu}_n(k, l)$ , we have by Lemma 4.1,

$$\begin{aligned} (4.3) \quad \mathbf{E}[X_i^\sigma X_j^\sigma X_k^\sigma X_l^\sigma | \mathcal{M}_{\nu_n(i,j,k,l)}] &= \mathbf{E}[X_i^\sigma X_j^\sigma | \mathcal{M}_{\nu_n(i,j)}] \mathbf{E}[X_k^\sigma X_l^\sigma | \mathcal{M}_{\nu_n(k,l)}] \\ &= \sigma_{ij}^* \sigma_{kl}^*. \end{aligned}$$

Let

$$\tilde{\Delta}_n \equiv \left( \sum_{i \in N_n} \sum_{j \in \bar{\nu}_n(i)} \sigma_{ij} \right)^2 - \left( \sum_{i \in N_n} \sum_{j \in \bar{\nu}_n(i)} \sigma_{ij}^* \right)^2.$$

Then we can write

$$\begin{aligned} (4.4) \quad 1 &= \left( \sum_{i \in N_n} \sum_{j \in \bar{\nu}_n(i)} \sigma_{ij} \right)^2 \\ &= \left( \sum_{i \in N_n} \sum_{j \in \bar{\nu}_n(i)} \sigma_{ij}^* \right)^2 + \tilde{\Delta}_n = \sum' \sigma_{ij}^* \sigma_{kl}^* + \sum'' \sigma_{ij}^* \sigma_{kl}^* + \tilde{\Delta}_n, \end{aligned}$$

where the sum  $\sum'$  is over  $(i, j, k, l)$  such that  $i \in N_n, j \in \bar{v}_n(i), k \in N_n, l \in \bar{v}_n(k)$  and either  $\{k, l\} \cap \bar{v}_n(i, j) \neq \emptyset$  or  $\{i, j\} \cap \bar{v}_n(k, l) \neq \emptyset$ , and the sum  $\sum''$  is over  $(i, j, k, l)$  such that  $i \in N_n, j \in \bar{v}_n(i), k \in N_n, l \in \bar{v}_n(k)$  and  $\{k, l\} \subset N_n \setminus \bar{v}_n(i, j)$  and  $\{i, j\} \subset N_n \setminus \bar{v}_n(k, l)$ . This implies that

$$\begin{aligned} & \mathbf{E} \left[ \left( \sum_{i \in N_n} \sum_{j \in \bar{v}_n(i)} X_i^\sigma X_j^\sigma \right)^2 \middle| \mathcal{G} \right] - 1 \\ &= \sum' \mathbf{E}[X_i^\sigma X_j^\sigma X_k^\sigma X_l^\sigma | \mathcal{G}] + \sum'' \mathbf{E}[X_i^\sigma X_j^\sigma X_k^\sigma X_l^\sigma | \mathcal{G}] - 1 \\ &= \sum' \mathbf{E}[X_i^\sigma X_j^\sigma X_k^\sigma X_l^\sigma | \mathcal{G}] + \sum'' \mathbf{E}[\sigma_{ij}^* \sigma_{kl}^* | \mathcal{G}] - 1 \\ &= \sum' (\mathbf{E}[X_i^\sigma X_j^\sigma X_k^\sigma X_l^\sigma | \mathcal{G}] - \mathbf{E}[\sigma_{ij}^* \sigma_{kl}^* | \mathcal{G}]) - \mathbf{E}[\tilde{\Delta}_n | \mathcal{G}]. \end{aligned}$$

The second equality is by (4.3) and the third equality is by (4.4). The leading sum is bounded by  $Cnd_{mx}^2 d_{av} \mu_4^4$ , because the number of the terms in the sum  $\sum'$  is bounded by  $C_1 nd_{mx}^2 d_{av}$  for some constant  $C_1 > 0$ .

Let us focus on  $\tilde{\Delta}_n$ . We write  $\tilde{\Delta}_n = R_n(2 - R_n)$ , by using (4.4), where

$$R_n \equiv \sum_{i \in N_n} \sum_{j \in \bar{v}_n(i)} (\sigma_{ij} - \sigma_{ij}^*).$$

Since  $\mathbf{E}[R_n | \mathcal{G}] = 0$ , we have

$$\mathbf{E}[\tilde{\Delta}_n | \mathcal{G}] = -\mathbf{E}[R_n^2 | \mathcal{G}] = -r_n^2.$$

Hence,

$$\mathbf{E}[h_{1,t}(W^\sigma)^2 | \mathcal{G}] \lesssim t^2 (nd_{mx}^2 d_{av} \mu_4^4 + r_n^2).$$

Let us turn to  $h_{2,t}(W^\sigma)$ . Using series expansion of  $\exp(-itx)$  (e.g. see (3.2) of Tikhomirov (1980)), we bound

$$\mathbf{E}[h_{2,t}(W^\sigma) | \mathcal{G}] \leq \frac{t^2}{2} \sum_{i \in N_n} \mathbf{E}[|X_i^\sigma| (W_i^\sigma)^2 | \mathcal{G}] = \frac{t^2}{2} \sum_{i \in N_n} \sum_{j, k \in \bar{v}_n(i)} \mathbf{E}[|X_i^\sigma X_j^\sigma X_k^\sigma| | \mathcal{G}].$$

Using arithmetic-geometric mean inequality, we can bound the last term by  $Ct^2 nd_{mx} d_{av} \mu_3^3$ .

Finally, let us turn to  $h_{3,t}(W^\sigma)$ . We write  $\mathbf{E}[h_{3,t}(W^\sigma) | \mathcal{G}]$  as

$$\sum_{i \in N_n} \mathbf{E}[X_i^\sigma \exp(it(W^\sigma - W_i^\sigma)) | \mathcal{G}] = \sum_{i \in N_n} \mathbf{E} \left[ X_i^\sigma \exp \left( it \sum_{j \in N_n \setminus \bar{v}_n(i)} X_j^\sigma \right) \middle| \mathcal{G} \right].$$

The last conditional expectation is equal to

$$\begin{aligned} & \mathbf{E} \left[ \mathbf{E} \left[ X_i^\sigma \exp \left( it \sum_{j \in N_n \setminus \bar{\nu}_n(i)} X_j^\sigma \right) \middle| \mathcal{M}_{\nu_n(i)} \right] \middle| \mathcal{G} \right] \\ &= \mathbf{E} \left[ \mathbf{E} [X_i^\sigma | \mathcal{M}_{\nu_n(i)}] \mathbf{E} \left[ \exp \left( it \sum_{j \in N_n \setminus \bar{\nu}_n(i)} X_j^\sigma \right) \middle| \mathcal{M}_{\nu_n(i)} \right] \middle| \mathcal{G} \right] = 0. \end{aligned}$$

The first equality follows by CND and the second equality follows because  $\mathbf{E} [X_i^\sigma | \mathcal{M}_{\nu_n(i)}] = 0$ . Hence, it follows that

$$\mathbf{E}[h_{3,t}(W^\sigma) | \mathcal{G}] = 0.$$

Since we have

$$\varphi_n'(t) + t\varphi_n(t) = -i(\mathbf{E}[(it - W^\sigma) \exp(itW^\sigma) | \mathcal{G}]),$$

by collecting the results for  $h_{1,t}(W^\sigma)$ ,  $h_{2,t}(W^\sigma)$ , and  $h_{3,t}(W^\sigma)$ , we obtain the desired result. ■

**Proof of Theorem 3.1:** For each  $t \in \mathbf{R}$ ,

$$\varphi_n'(t) = -t\varphi_n(t) - i\mathbf{E}[(it - W^\sigma) \exp(itW^\sigma) | \mathcal{G}] \equiv -t\varphi_n(t) + \gamma_n(t), \text{ say.}$$

Taking integral of both sides, we obtain the following expression:

$$\varphi_n(t) = \exp\left(-\frac{t^2}{2}\right) \left[ 1 + \int_0^t \gamma_n(u) \exp\left(\frac{u^2}{2}\right) du \right]$$

or

$$(4.5) \quad \left| \varphi_n(t) - \exp\left(-\frac{t^2}{2}\right) \right| \leq \exp\left(-\frac{t^2}{2}\right) \int_0^t |\gamma_n(u)| \exp\left(\frac{u^2}{2}\right) du.$$

Note that for all  $t \geq 0$ ,

$$\int_0^t u^2 \exp(u^2/2) du \leq \exp(t^2/2)t,$$

and  $\int_0^t u \exp(u^2/2) du = \exp(t^2/2) - 1$ . Applying Lemma 4.2, the last term in (4.5) for  $t > 0$  is bounded by  $C(ta_n + (1 - \exp(-t^2/2))b_n)$ , where

$$a_n \equiv nd_{mx}d_{av}\mu_3^3, \text{ and } b_n \equiv \sqrt{nd_{mx}^2d_{av}\mu_4^4 + r_n^2},$$

for some absolute constant  $C > 0$ . Hence for any  $T \geq 1$ ,

$$\int_{[-T,T]} \left| \frac{\varphi_n(t) - e^{-t^2/2}}{t} \right| dt \leq Ca_n \int_{[-T,T]} dt + Cb_n \int_{[-T,T]} \left| \frac{1 - e^{-t^2/2}}{t} \right| dt.$$

The last sum is bounded by  $2CTa_n + 2C \log(T)b_n$ . Therefore, by Esseen's inequality (see e.g. Theorem 1.5.2 of Ibragimov and Linnik (1971), p.27), we obtain the following bound on the event  $\mathcal{A}_n$ ,

$$\Delta_n(t; \mathcal{G}) \lesssim (Ta_n + \log(T)b_n + T^{-1}) \lesssim (a_n^{1/2} - \log(a_n)b_n),$$

by taking  $T = a_n^{-1/2}$ . ■

**Proof of Lemma 3.1:** For  $(i, j)$  and  $(i', j')$  such that either  $\{i, j\} \cap \bar{v}_n(i', j') = \emptyset$  or  $\{i', j'\} \cap \bar{v}_n(i, j) = \emptyset$ ,  $\mathbf{E}[\xi_{ij}\xi_{i'j'}|\mathcal{G}] = \mathbf{E}[\xi_{ij}|\mathcal{G}]\mathbf{E}[\xi_{i'j'}|\mathcal{G}] = 0$ . Let  $A$  be the set of  $((i, j), (i', j'))$  such that  $i \in N_n, j \in \bar{v}_n(i), i' \in N_n$  and  $j' \in \bar{v}_n(i')$ . Then

$$\mathbf{E} \left[ \left( \sum_{i \in N_n} \sum_{j \in \bar{v}_n(i)} \xi_{ij} \right)^2 \middle| \mathcal{G} \right] \leq \sigma_n^4 \mu_4^4 \sum_{j=1}^4 A_j,$$

where  $A_1$  is the number of  $((i, j), (i', j')) \in A$  such that either  $i \in \bar{v}_n(i')$  or  $i' \in \bar{v}_n(i)$ ;  $A_2$  is the number of  $((i, j), (i', j')) \in A$  such that either  $i \in \bar{v}_n(j')$  or  $j' \in \bar{v}_n(i)$ ;  $A_3$  is the number of  $((i, j), (i', j')) \in A$  such that either  $j \in \bar{v}_n(i')$  or  $i' \in \bar{v}_n(j)$ ;  $A_4$  is the number of  $((i, j), (i', j')) \in A$  such that either  $j \in \bar{v}_n(j')$  or  $j' \in \bar{v}_n(j)$ . Thus, it is not hard to see that

$$\sum_{j=1}^4 A_j \leq 8nd_{mx}^2 d_{av},$$

completing the proof. ■

**Proof of Corollary 3.1:** Similarly as in the proof of Lemma 4.2, we decompose

$$(it - W^\sigma) \exp(itW^\sigma) = h_{1,t}(W^\sigma) - h_{2,t}(W^\sigma) - h_{3,t}(W^\sigma).$$

The treatment of  $h_{2,t}(W^\sigma)$  and  $h_{3,t}(W^\sigma)$  is the same as that of the proof of Lemma 4.2. The difference lies in the treatment of  $h_{1,t}(W^\sigma)$ . Using Condition A and Lemma 4.1, we note that for  $\{i, j\}, \{k, l\} \subset N_n$  such that  $\{k, l\} \subset N_n \setminus \bar{v}_n(i, j)$  and  $\{i, j\} \subset N_n \setminus \bar{v}_n(k, l)$ ,

$$(4.6) \quad \mathbf{E}[X_i^\sigma X_j^\sigma X_k^\sigma X_l^\sigma | \mathcal{G}] = \mathbf{E}[X_i^\sigma X_j^\sigma | \mathcal{G}]\mathbf{E}[X_k^\sigma X_l^\sigma | \mathcal{G}] = \sigma_{ij}\sigma_{kl}.$$

Following the same argument in the proof of Lemma 4.2, we find that

$$\mathbf{E} \left[ \left( \sum_{i \in N_n} \sum_{j \in \bar{v}_n(i)} X_i^\sigma X_j^\sigma \right)^2 \middle| \mathcal{G} \right] - 1 = \sum' (\mathbf{E}[X_i^\sigma X_j^\sigma X_k^\sigma X_l^\sigma | \mathcal{G}] - \sigma_{ij}\sigma_{kl}),$$

which is bounded by  $Cnd_{mx}^2 d_{av}\mu_4^4$ . Hence in the proof of Lemma 4.2, we do not need to deal with  $\tilde{\Delta}_n$ . Following the proofs of Lemma 4.2 and Theorem 3.1 for the rest of the terms, we obtain the desired result. ■

**Lemma 4.3.** *Suppose that  $X$  and  $Y$  are random variables such that  $\mathbf{E}[Y] = 0$  and  $(\mathbf{E}[|Y|^r])^{1/r} \leq M$  for some constants  $r, M > 0$ , and  $F$  is the CDF on  $\mathbf{R}$  with density function  $f$ . Then for any  $t \in \mathbf{R}$ ,*

$$\begin{aligned} |P\{X + Y \leq t\} - F(t)| &\leq |P\{X \leq t + q\} - F(t + q)| \\ &\quad + |P\{X \leq t - q\} - F(t - q)| + 4(\bar{c}M)^{r/(1+r)}, \end{aligned}$$

where  $\bar{c} \equiv \sup_{z \in \mathbf{R}} f(z)$  and  $q \equiv (Mr/\bar{c})^{1/(1+r)}$ .

**Proof:** First, note that for any  $\varepsilon > 0$ ,

$$(4.7) \quad |P\{X + Y \leq t\} - F(t)| \leq |P\{X + Y \leq t, |Y| \leq \varepsilon\} - F(t)| + P\{|Y| > \varepsilon\}.$$

As for the probability inside the absolute value above, we note that

$$\begin{aligned} P\{X + Y \leq t, |Y| \leq \varepsilon\} &\geq P\{X + \varepsilon \leq t, |Y| \leq \varepsilon\} \\ &\geq P\{X + \varepsilon \leq t\} - P\{|Y| > \varepsilon\}. \end{aligned}$$

Also, observe that

$$P\{X + Y \leq t, |Y| \leq \varepsilon\} \leq P\{X - \varepsilon \leq t\}.$$

Hence

$$\begin{aligned} &|P\{X + Y \leq t, |Y| \leq \varepsilon\} - F(t)| \\ &\leq \max\{|P\{X + \varepsilon \leq t\} - P\{|Y| > \varepsilon\} - F(t)|, |P\{X - \varepsilon \leq t\} - F(t)|\} \\ &\leq |P\{X \leq t - \varepsilon\} - F(t - \varepsilon)| + |P\{X \leq t + \varepsilon\} - F(t + \varepsilon)| \\ &\quad + 2 \sup_{z \in \mathbf{R}} f(z)\varepsilon + P\{|Y| > \varepsilon\}. \end{aligned}$$

From (4.7),

$$\begin{aligned} |P\{X + Y \leq t\} - F(t)| &\leq |P\{X \leq t - \varepsilon\} - F(t - \varepsilon)| + |P\{X \leq t + \varepsilon\} - F(t + \varepsilon)| \\ &\quad + 2 \sup_{z \in \mathbf{R}} f(z)\varepsilon + 2P\{|Y| > \varepsilon\}. \end{aligned}$$

Using Markov's inequality, we bound the last term by  $2\varepsilon^{-r}\mathbf{E}[|Y|^r] \leq 2\varepsilon^{-r}M^r$ . Taking  $\varepsilon = q$ , we obtain the desired result. ■

**Proof of Theorem 3.2:** We write  $\Delta^*(t; \mathcal{G})$  as

$$(4.8) \quad \left| P \left\{ \frac{1}{\sigma_n^*} \sum_{i \in N_n^*} X_i + R_n^* \leq t \mid \mathcal{G} \right\} - \Phi(t) \right|,$$

where  $R_n^* \equiv \frac{1}{\sigma_n^*} \sum_{i \in N_n \setminus N_n^*} X_i$ . We write

$$R_n^* = \sum_{i \in N_n \setminus N_n^*} \frac{X_i}{\sigma_n^*} = \sum_{i \in N_n \setminus N_n^*} \xi_i, \text{ say.}$$

Now, choose  $1 \leq r \leq 4$  and write

$$\mathbf{E} \left[ \left| \sum_{i \in N_n \setminus N_n^*} \xi_i \right|^r \middle| \mathcal{G} \right] \leq \left( \sum_{i \in N_n \setminus N_n^*} (\mathbf{E}[|\xi_i|^r | \mathcal{G}])^{1/r} \right)^r \leq (|N_n \setminus N_n^*| \tilde{\mu}_r^*)^r.$$

Hence on the event  $\mathcal{A}_{n,r}(\varepsilon_n)$ ,

$$(\mathbf{E}[|R_n^*|^r | \mathcal{G}])^{1/r} + \rho_r^* \leq \varepsilon_n.$$

Define for brevity,

$$S_{n,\sigma}^* \equiv \frac{1}{\sigma_n^*} \sum_{i \in N_n^*} (X_i - \mathbf{E}[X_i | \mathcal{M}_{\nu_n^*}^*(i)]).$$

By Lemma 4.3, the term (4.8) is bounded by (for any  $r > 0$ )

$$(4.9) \quad \begin{aligned} & |P \{S_{n,\sigma}^* \leq t + q_n | \mathcal{G}\} - \Phi(t + q_n)| \\ & + |P \{S_{n,\sigma}^* \leq t - q_n | \mathcal{G}\} - \Phi(t - q_n)| + 4(\phi(0)\varepsilon_n)^{r/(r+1)}, \end{aligned}$$

where  $\phi$  denotes the density of  $N(0, 1)$  and

$$q_n = \varepsilon_n^{r/(r+1)} / \phi(0)^{1/(r+1)}.$$

By Lemma 2.2,  $\{X_i\}_{i \in N_n^*}$  is CND with respect to  $(\nu_n^*, \mathcal{M}^*)$ , we apply Theorem 3.1 to the leading two terms in (4.9) to obtain their bound as

$$C \sqrt{n^* d_{mx}^* d_{av}^* \mu_3^{*3}} - C \log(n^* d_{mx}^* d_{av}^* \mu_3^{*3}) \sqrt{n^* d_{mx}^{*2} d_{av}^* \mu_4^{*4} + r_n^{*2}}$$

for some constant  $C > 0$ , delivering the desired result. ■

**Lemma 4.4.** *Suppose that  $(\mathbb{D}, d)$  is a given metric space and for each  $n \geq 1$ ,  $\xi_n, \zeta_n, \zeta$  are  $\mathbb{D}$ -valued random variables. If for each  $\varepsilon > 0$ ,  $P^*\{d(\xi_n, \zeta_n) > \varepsilon\} \rightarrow 0$  and  $\zeta_n \rightarrow \zeta$ ,  $\mathcal{G}$ -stably, as  $n \rightarrow \infty$ , then*

$$\xi_n \rightarrow \zeta, \mathcal{G}\text{-stably.}$$

**Proof:** First note that  $\zeta_n \rightarrow \zeta$ ,  $\mathcal{G}$ -stably if and only if for all event  $U \in \mathcal{G}$  and any closed set  $F \in \mathcal{B}(\mathbb{D})$ ,

$$(4.10) \quad \limsup_{n \rightarrow \infty} P^*\{\zeta_n \in F\} \cap U \leq P\{\zeta \in F\} \cap U.$$

(This can be shown following the proof of Theorem 1.3.4 (iii) of van der Vaart and Wellner (1996).) Using this and following the same arguments in the proof of Lemma 1.10.2 of van der Vaart and Wellner (1996), we deduce the lemma. ■

**Proof of Lemma 3.2:** (i) Since  $P^*\{d(\xi_n, \xi) > \varepsilon\} \rightarrow 0$ , we have  $P^*\{\tilde{d}((\zeta_n, \xi_n), (\zeta_n, \xi)) > \varepsilon\} \rightarrow 0$ , where  $\tilde{d}$  is a metric on  $\mathbb{D} \times \mathbb{D}$  defined as  $\tilde{d}((f_1, f_2), (g_1, g_2)) = d(f_1, g_1) + d(f_2, g_2)$  for  $f_1, f_2, g_1, g_2 \in \mathbb{D}$ . Furthermore, note that  $(\zeta_n, \xi) \rightarrow (\zeta, \xi)$ ,  $\mathcal{G}$ -stably, because  $\zeta_n \rightarrow \zeta$ ,  $\mathcal{G}$ -stably, and  $\xi$  is  $\mathcal{G}$ -measurable. Now the desired result follows by Lemma 4.4.

(ii) Note that  $\zeta'_n \rightarrow \zeta'$ ,  $\mathcal{G}$ -stably, if and only if for any event  $U \in \mathcal{G}$  and any open set  $G \in \mathcal{B}(\mathbb{D})$ ,

$$(4.11) \quad \liminf_{n \rightarrow \infty} P_*\{\zeta'_n \in G\} \cap U \geq P\{\zeta' \in G\} \cap U,$$

where  $P_*$  denotes the inner probability. Using this and following the same arguments in the proof of Theorem 3.27 of Kallenberg (1997) for the continuous mapping theorem for weak convergence, we can complete the proof of (ii). ■

For the proof of Theorem 3.3, we use the following lemma.

**Lemma 4.5.** *If  $f : \mathbb{D} \mapsto \mathbf{R}$  is bounded and continuous, and  $K \subset \mathbb{D}$  is compact, then for every  $\epsilon > 0$  there exists  $\tau > 0$  such that, if  $x \in K$  and  $y \in \mathbb{D}$  with  $\|x - y\| < \tau$ , then*

$$|f(x) - f(y)| < \epsilon.$$

**Proof of Theorem 3.3:** First, let us suppose that (i) and (ii) hold. To see that the marginal  $PL$  of  $L$  is a tight Borel law, note that the stable finite dimensional convergence of  $\zeta_n$  implies the convergence of the finite dimensional distributions of  $\zeta_n$ . Combining this with the asymptotic  $\rho$ -equicontinuity and using Theorems 1.5.4 and 1.5.7 of van der Vaart and Wellner (1996), we obtain that  $PL$  is a tight Borel law. The fact that  $PL$  is concentrated on  $U_\rho(\mathcal{H})$  follows from Theorem 10.2 of Pollard (1990).

Now let us show the  $\mathcal{G}$ -stable convergence of  $\zeta_n$ . We follow the arguments in the proof of Theorem 2.1 of Wellner (2005). Let  $\zeta \in U_\rho(\mathcal{H})$  be a random element whose distribution is the same as  $PL$ . Since  $(\mathcal{H}, \rho)$  is totally bounded, for every  $\delta > 0$  there exists a finite set of points  $\mathcal{H}_\delta$  that is  $\delta$ -dense in  $\mathcal{H}$  i.e.  $\mathcal{H} \subset \cup_{h \in \mathcal{H}_\delta} B(h, \delta)$  where  $B(h, \delta)$  is the open ball with center  $h$  and radius  $\delta$ . Thus, for each  $h \in \mathcal{H}$ , we can choose  $\pi_\delta(h) \in \mathcal{H}_\delta$  such that  $\rho(\pi_\delta(h), h) < \delta$ . Define

$$\zeta_{n,\delta}(h) = \zeta_n(\pi_\delta(h)), \text{ and } \zeta_\delta(h) = \zeta(\pi_\delta(h)) \text{ for } h \in \mathcal{H}.$$



By the  $\mathcal{G}$ -stable convergence of the finite dimensional projection of  $\zeta_n$ , we have for each  $U \in \mathcal{G}$  and for each bounded and continuous functional  $f : \ell^\infty(\mathcal{H}) \mapsto \mathbf{R}$ ,

$$(4.12) \quad \mathbf{E}^* [f(\zeta_{n,\delta})1_U] \rightarrow \mathbf{E} [f(\zeta_\delta)1_U].$$

Furthermore, the a.e. uniform continuity of the sample paths of  $\zeta$  implies that

$$(4.13) \quad P \left\{ \limsup_{\delta \rightarrow 0} \sup_{h \in \mathcal{H}} |\zeta(h) - \zeta_\delta(h)| = 0 \right\} = 1.$$

For each bounded and continuous functional  $f : \ell^\infty(\mathcal{H}) \mapsto \mathbf{R}$ , and for each  $U \in \mathcal{G}$  such that  $P(U) > 0$ ,

$$\begin{aligned} & |\mathbf{E}^* [f(\zeta_n)1_U] - \mathbf{E} [f(\zeta)1_U]| \\ & \leq |\mathbf{E}^* [(f(\zeta_n) - f(\zeta_{n,\delta}))1_U]| + |\mathbf{E} [(f(\zeta_{n,\delta}) - f(\zeta_\delta))1_U]| + |\mathbf{E} [(f(\zeta_\delta) - f(\zeta))1_U]|. \end{aligned}$$

The last two absolute values vanish as  $n \rightarrow \infty$  and then  $\delta \rightarrow 0$  by (4.12) and by (4.13) combined with the Dominated Convergence Theorem. We use the asymptotic  $\rho$ -equicontinuity of  $\zeta_n$  and Lemma 4.5 and the fact that  $PL$  is a tight law, and follow standard arguments to show that the leading difference vanishes as  $n \rightarrow \infty$  and then  $\delta \rightarrow 0$ . (See the proof of Theorem 2.1 of Wellner (2005) for details.)

Since the  $\mathcal{G}$ -stable convergence of  $\zeta_n$  implies that of its finite dimensional distributions, and the weak convergence of  $\zeta_n$ , the converse can be shown using the standard arguments. (Again, see the proof of Theorem 2.1 of Wellner (2005).) ■

**Proof of Lemma 3.3:** The proof follows that of Theorems 2.3 and 3.4 of Janson (2004). In particular (3.8) follows from Theorem 2.3. However, for (3.7), we need to modify the proof of Theorem 3.4 because  $\sigma_i^2$ 's are not necessarily  $\mathcal{G}$ -measurable, and hence the equations (3.9) and (3.10) on page 241 do not necessarily follow.

First, without loss of generality, we set  $M = 1$ . Following the proof of Theorem 3.4 in Janson (2004) (see (3.7) there), we obtain that for any  $c \geq 0$ ,

$$(4.14) \quad \mathbf{E}[\exp(cX_i)|\mathcal{M}_{\nu_n(i)}] \leq \exp(\sigma_i^2 g(c)),$$

where  $g(c) = e^c - 1 - c$ . Let  $N_j^* \subset N_n$ ,  $j = 1, \dots, J$  be disjoint subsets which partition  $N_n$  such that for any  $i_1, i_2 \in N_j^*$ ,  $i_1 \neq i_2$ ,  $i_1 \notin \nu_n(i_2)$  and  $i_2 \notin \nu_n(i_1)$ . Fix  $u \geq 0$ ,  $p_j \geq 0$ ,  $j = 1, \dots, J$  such that  $\sum_{j=1}^J p_j = 1$  and  $w_j \in [0, 1]$ ,  $j = 1, \dots, J$  such that  $\sum_{j \in N_n : i \in N_j^*} w_j = 1$  for all  $i \in N_n$ . Then using Lemma 4.1 and (4.14) and following the same argument in (3.8)

of Janson (2004),

$$\begin{aligned}
\mathbf{E} \left[ \exp \left( u \sum_{i \in N_n} X_i \right) \middle| \mathcal{G} \right] &\leq \sum_{j=1}^J p_j \mathbf{E} \left[ \prod_{i \in N_j^*} \mathbf{E} \left[ \exp \left( \frac{w_j u}{p_j} X_i \right) \middle| \mathcal{M}_{\nu_n(i)} \right] \middle| \mathcal{G} \right] \\
&\leq \sum_{j=1}^J p_j \mathbf{E} \left[ \exp \left( \sum_{i \in N_j^*} \sigma_i^2 g \left( \frac{w_j u}{p_j} \right) \right) \middle| \mathcal{G} \right] \\
&\leq \sum_{j=1}^J p_j \mathbf{E} \left[ \exp \left( \sum_{i \in N_n} \sigma_i^2 g \left( \frac{w_j u}{p_j} \right) \right) \middle| \mathcal{G} \right],
\end{aligned}$$

because  $g(\cdot) \geq 0$ . The last term above is bounded by

$$(4.15) \quad \sum_{j=1}^J p_j \mathbf{E} \left[ \exp \left( (\kappa_n + V_n) g \left( \frac{w_j u}{p_j} \right) \right) \middle| \mathcal{G} \right]$$

where

$$\kappa_n = \log \mathbf{E} \left[ \exp \left( \left| \sum_{i \in N_n} \sigma_i^2 - V_n \right| \right) \middle| \mathcal{G} \right].$$

As for  $\kappa_n$ , note that

$$(4.16) \quad \kappa_n \leq \log \mathbf{E} [\exp(|V_n|) \middle| \mathcal{G}] = V_n,$$

because  $V_n$  is  $\mathcal{G}$ -measurable. Let  $W = \sum_{j=1}^J w_j$  and take  $p_j = w_j/W$  to rewrite (4.15) as

$$\mathbf{E} [\exp((\kappa_n + V_n)g(uW)) \middle| \mathcal{G}].$$

Hence we have for each  $t \geq 0$ ,

$$P \left\{ \sum_{i \in N_n} X_i > t \middle| \mathcal{G} \right\} \leq \mathbf{E} [\exp((\kappa_n + V_n)g(uW) - ut) \middle| \mathcal{G}].$$

If we take

$$u = \frac{1}{W} \log \left( \frac{t}{(\kappa_n + V_n)W} + 1 \right)$$

and let  $\varphi(x) \equiv (1+x) \log(1+x) - x$ , the last bound becomes

$$\begin{aligned}
\exp \left( -(\kappa_n + V_n) \varphi \left( \frac{t}{(\kappa_n + V_n)W} \right) \right) &\leq \exp \left( -\frac{t^2}{2W(W(\kappa_n + V_n) + t/3)} \right) \\
&\leq \exp \left( -\frac{t^2}{2W(2WV_n + t/3)} \right),
\end{aligned}$$

where the first inequality follows by the inequality:  $\varphi(x) \geq x^2/(2(1+x/3))$ ,  $x \geq 0$ , and the last inequality follows by (4.16). Now, as in the proof of Theorem 2.3 of Janson (2004),

the rest of the proof can be proceeded by taking  $\{(N_j^*, w_j)\}_{j=1}^J$  as a minimal fractional proper cover of  $N_n$ . ■

**Proof of Lemma 3.4:** We adapt the proof of Theorem A.2 of van der Vaart (1996) to accommodate the CND property of  $\{Y_i\}_{i \in N_n}$ . Fix  $q_0$  so that

$$2^{-q_0} \leq \bar{\rho}_n(H) \leq 2^{-q_0+1}$$

and for each  $q \geq q_0$ , construct a nested sequence of partitions  $\mathcal{H} = \cup_{i=1}^{N_q} \mathcal{H}_{q_i}$  such that

$$(4.17) \quad \frac{1}{n} \sum_{j \in N_n} \mathbf{E}^* \left( \sup_{h, g \in \mathcal{H}_{q_i}} |h(Y_j) - g(Y_j)|^2 \right) < 2^{-2q} \text{ for every } i = 1, \dots, N_q.$$

By the definition of  $\bar{\rho}_n(h)$  and the bracketing entropy,  $N_q$  can be taken to satisfy

$$\log N_q \leq \sum_{r=q_0}^q \log (1 + N_{\square}(2^{-r}, \mathcal{H}, \bar{\rho}_n)).$$

Choose for each  $q$  a fixed element  $h_{q_i}$  from each  $\mathcal{H}_{q_i}$  and set  $\pi_q h = h_{q_i}$  and  $\Delta_q h = (\sup_{h, g \in \mathcal{H}_{q_i}} |h - g|)^*$ , whenever  $h \in \mathcal{H}_{q_i}$ , where  $(h)^*$  defines the minimal measurable cover of  $h$  (Dudley (1985).) Then  $(\bar{\rho}_n(\Delta_q h))^2 < 2^{-2q}$  from (4.17), and  $\pi_q h$  and  $\Delta_q h$  run through a set of  $N_q$  functions as  $h$  runs through  $\mathcal{H}$ . Define for each fixed  $n$  and  $q \geq q_0$ , the following numbers and indicator functions:

$$\begin{aligned} \alpha_q &= 2^{-q} / \sqrt{\log N_{q+1}}, \\ A_{q-1} h &= 1 \{ \Delta_{q_0} h \leq \sqrt{n} \alpha_{q_0}, \dots, \Delta_{q-1} h \leq \sqrt{n} \alpha_{q-1} \}, \\ B_q h &= 1 \{ \Delta_{q_0} h \leq \sqrt{n} \alpha_{q_0}, \dots, \Delta_{q-1} h \leq \sqrt{n} \alpha_{q-1}, \Delta_q h > \sqrt{n} \alpha_q \}, \\ B_{q_0} h &= 1 \{ \Delta_{q_0} h > \sqrt{n} \alpha_{q_0} \}. \end{aligned}$$

Because the partitions are nested,  $A_q h$  and  $B_q h$  are constant in  $h$  on each of the partitioning sets  $\mathcal{H}_{q_i}$  at level  $q$ . Now decompose  $h = (h - \pi_{q_0} h) + \pi_{q_0} h \equiv I + II$ , say, with  $I \equiv h - \pi_{q_0} h = (h - \pi_{q_0} h)$ . Then we can write

$$\begin{aligned} I &= (h - \pi_{q_0} h) B_{q_0} h + \sum_{q=q_0+1}^{\infty} (h - \pi_q h) B_q h + \sum_{q=q_0+1}^{\infty} (\pi_q h - \pi_{q-1} h) A_{q-1} h \\ &\equiv I_a + I_b + I_c, \text{ say.} \end{aligned}$$

We analyze the empirical process at each of  $I_a$ ,  $I_b$  and  $I_c$ .

**Control of  $I_a$ :** Let us bound  $\mathbf{E}^*[\sup_{h \in \mathcal{H}} |\mathbb{G}_n((h - \pi_{q_0}h)B_{q_0}h)|]$  by

$$\mathbf{E}^* \left( \sup_{h \in \mathcal{H}} \frac{1}{\sqrt{n}} \sum_{i \in N_n} (|h - \pi_{q_0}h|B_{q_0}h)(X_i) + \sup_{h \in \mathcal{H}} \frac{1}{\sqrt{n}} \sum_{i \in N_n} \mathbf{E}[|h - \pi_{q_0}h|B_{q_0}h)(X_i)|\mathcal{M}_\nu(i)] \right).$$

Since  $|h - \pi_{q_0}h|B_{q_0}h \leq 2H1\{2H > \sqrt{n}\alpha_{q_0}\} \leq 4H^2/(\sqrt{n}\alpha_{q_0})$ , we bound the last expression by

$$8\alpha_{q_0}^{-1}\bar{\rho}_n(H)^2 \leq 8 \left( 2^{-q_0} / \sqrt{\log N_{q_0+1}} \right)^{-1} 2^{-2q_0+2} \leq 32 \cdot 2^{-q_0} \sqrt{\log N_{q_0+1}},$$

due to our choice of  $q_0$  satisfying that  $\bar{\rho}_n(H)^2 \leq 2^{-2q_0+2}$ .

**Control of  $I_c$ :** For  $I_c = \sum_{q=q_0+1}^{\infty} (\pi_q h - \pi_{q-1}h) A_{q-1}h$ , there are at most  $N_q - 1$  functions  $\pi_q h - \pi_{q-1}h$  and at most  $N_{q-1} - 1$  functions  $A_{q-1}h$ . Since the partitions are nested, the function  $|\pi_q h - \pi_{q-1}h| A_{q-1}h$  is bounded by  $\Delta_{q-1}h A_{q-1}h \leq \sqrt{n}\alpha_{q-1}$ . Applying Corollary 3.2 (with  $J = \sqrt{n}\alpha_{q-1}$  and  $m = N_q - 1$ ) to  $\mathbb{G}_n(\pi_q h - \pi_{q-1}h)A_{q-1}h$ ,

$$\begin{aligned} & \mathbf{E} \left[ \max_{h \in \mathcal{H}} |\mathbb{G}_n(\pi_q h - \pi_{q-1}h) A_{q-1}h| \middle| \mathcal{G} \right] \\ & \lesssim (d_{mx} + 1) \left( \alpha_{q-1} \log N_q + \sqrt{(\log N_q) \max_{1 \leq s \leq N_{q-1}} V_n(h_s)} \right) \end{aligned}$$

where  $\{h_s : s = 1, \dots, N_{q-1}\} \equiv \{(\pi_q - \pi_{q-1})h A_{q-1}h : h \in \mathcal{H}\}$ . From the law of the iterated conditional expectations and Jensen's inequality,

$$\begin{aligned} \mathbf{E}^* \left[ \sup_{h \in \mathcal{H}} |\mathbb{G}_n I_c| \right] & \lesssim (d_{mx} + 1) \sum_{q=q_0+1}^{\infty} \left( \sqrt{(\log N_q) \mathbf{E}^* \left[ \max_{1 \leq s \leq N_{q-1}} V_n(h_s) \right]} + 2^{-q} \sqrt{\log N_q} \right) \\ & \lesssim 2(d_{mx} + 1) \sum_{q=q_0+1}^{\infty} 2^{-q} \sqrt{\log N_q}, \end{aligned}$$

where for the last inequality, we used (4.17) so that

$$(4.18) \quad \mathbf{E}^* \left[ \max_{1 \leq s \leq N_{q-1}} V_n(h_s) \right] \leq \mathbf{E}^* \left( \max_{1 \leq s \leq N_{q-1}} \left( \frac{1}{n} \sum_{i \in N_n} \mathbf{E} [h_s^2(Y_i) | \mathcal{M}_i] \middle| \mathcal{G} \right) \right) \lesssim 2^{-q+1}.$$

**Control of  $I_b$  and  $II$ :** The proof of these parts are the same as that of Theorem A.2 of van der Vaart (1996) except that we use  $\bar{\rho}_n(\cdot)$  instead of  $\|\cdot\|_{P,2}$  so that we have  $\mathbf{E}^* \|\mathbb{G}_n I_b\|_{\mathcal{H}} \lesssim \sum_{q=q_0+1}^{\infty} 2^{-q} \sqrt{\log N_q}$  and  $\mathbf{E}^* \|\mathbb{G}_n II\|_{\mathcal{H}} \lesssim 2^{-q_0} \sqrt{\log N_{q_0}}$ .

Now collecting the results for  $I_a$ ,  $I_b$ ,  $I_c$  and  $II$ , we have

$$\begin{aligned} \mathbf{E}^* \left[ \sup_{h \in \mathcal{H}} |\mathbb{G}_n(h)| \right] &\lesssim (d_{mx} + 1) \sum_{q=q_0+1}^{\infty} 2^{-q} \sqrt{\log N_q} \\ &\lesssim (d_{mx} + 1) \int_0^{\bar{\rho}_n(H)} \sqrt{1 + \log N_{\square}(\varepsilon, \mathcal{H}, \bar{\rho}_n)} d\varepsilon, \end{aligned}$$

giving the required result. ■

**Proof of Theorem 3.4:** We prove conditions for Theorem 3.3. Let us first consider the convergence of finite dimensional distributions. Without loss of generality, we consider the CLT for  $\mathbb{G}_n(h)$  for some  $h \in \mathcal{H}$  such that  $\text{Var}(\mathbb{G}_n(h)|\mathcal{G}) > 0$ , a.e.. Assumption 3.1 (a) together with the moment condition for the envelope  $H$  implies that  $d_{mx} < C$  for all  $n \geq 1$  for some  $C > 0$ . We apply Theorem 3.1 to obtain the convergence of finite dimensional distributions. By the CND property of  $(Y_i)_{i \in N_n}$  and Assumption 3.1(b), Condition (i) in Theorem 3.3 is satisfied.

Let us prove asymptotic  $\rho$ -equicontinuity (with  $\rho(h, h) \equiv \rho(h) \equiv \lim_{n \rightarrow \infty} \bar{\rho}_n(h)$ .) Define  $\mathcal{H}_{n,\delta} \equiv \{h - g : \bar{\rho}_n(h - g) \leq \delta, h, g \in \mathcal{H}\}$ . Then, by Lemma 3.4,

$$\mathbf{E}^* \left[ \sup_{h \in \mathcal{H}_{n,\delta}} |\mathbb{G}_n(h)| \right] \lesssim (1 + d_{mx}) \int_0^\delta \sqrt{1 + \log N_{\square}(\varepsilon, \mathcal{H}_{n,\delta}, \bar{\rho}_n)} d\varepsilon.$$

By noting that  $\mathcal{H}_{n,\delta}$  is contained in  $\mathcal{H} - \mathcal{H}$  and by Assumption 3.1(a), the last bound vanishes as  $\delta \rightarrow 0$  for each  $n \geq 1$ . Thus the asymptotic  $\rho$ -equicontinuity follows, by the condition for  $\rho(h)$ . ■

**Proof of Theorem 3.5:** By Assumption 3.2, we have

$$\frac{n - n^*}{\sqrt{n}} \max_{i \in N_n} \sqrt{\mathbf{E}[H^2(Y_i)]} + \mathbf{E}^* \left[ \sup_{h \in \mathcal{H}} \rho_n^*(h) \right] \rightarrow 0,$$

as  $n \rightarrow \infty$ . Hence the desired result follows applying Theorem 3.4 to  $\mathbb{G}_n^*$  in (3.9). ■

## References

- ALDOUS, D. J., AND G. K. EAGLESON (1978): “On Mixing and Stability of Limit Theorems,” *Annals of Probability*, 6, 325–331.
- ANDREWS, D. W. K. (1994): “Empirical Process Methods in Econometrics,” in *Handbook of Econometrics*, pp. 2247–2294. Elsevier.
- BALDI, P., AND Y. RINOTT (1989): “On Normal Approximations of Distributions in Terms of Dependency Graphs,” *Annals of Probability*, 17, 1646–1650.

- BARABÁSI, A.-L., AND R. ALBERT (1999): “Emergence of Scaling in Random Networks,” *Science*, 286, 509–512.
- BERTI, P., L. PRATELLI, AND P. RIGO (2012): “Limit Theorems for Empirical Processes Based on Dependent Data,” *Electronic Journal of Probability*, 9, 1–18.
- CAI, T. T., W. LIU, AND H. H. ZHOU (2016): “Estimating Sparse Precision Matrix: Optimal Rates of Convergence and Adaptive Estimation,” *Annals of Statistics*, 44, 455–488.
- CANEN, N., J. SCHWARTZ, AND K. SONG (2017): “Estimating Local Interactions Among Many Agents Who Observe Their Neighbors,” *Working Paper*.
- CHEN, L., AND W. B. WU (2016): “Stability and Asymptotics for Autoregressive Processes,” *Electronic Journal of Statistics*, 10, 3723–3751.
- CHEN, L. H. Y., AND Q.-M. SHAO (2004): “Normal Approximation Under Local Dependence,” *Annals of Probability*, 32, 1985–2028.
- DAWID, P. A. (1979): “Conditional Independence in Statistical Theory,” *Journal of the Royal Statistical Society, B*, 41, 1–31.
- DUDLEY, R. M. (1985): “An Extended Wichura Theorem, Definitions of Donsker Class, and Weighted Empirical Distributions,” in *Probability in Banach Spaces V*, ed. by A. Beck, R. Dudley, M. Hahn, J. Kuelbs, and M. Marcus, pp. 141–178. Springer.
- HAHN, J., G. KUERSTEINER, AND M. MAZZOCCO (2016): “Central Limit Theory for Combined Cross-Section and Time Series,” *arXiv:1610.01697 [stat.ME]*.
- HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. Academic Press, New York, USA.
- HÄUSLER, E., AND H. LUSCHGY (2010): *Stable Convergence and Stable Limit Theorems*. Springer Science+Business Media, New York, USA.
- IBRAGIMOV, L. A., AND Y. V. LINNIK (1971): *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.
- JANSON, S. (1988): “Normal Convergence by Higher Semiinvariants With Applications to Sums of Dependent Random Variables and Random Graphs,” *Annals of Probability*, 16, 305–312.
- (2004): “Large Deviations for Sums of Partly Dependent Random Variables,” *Random Structures and Algorithms*, 24, 234–248.
- JENISH, N., AND I. R. PRUCHA (2009): “Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields,” *Journal of Econometrics*, 150, 86–98.
- JIRAK, M. (2016): “Berry-Esseen Theorems Under Weak Dependence,” *Annals of Probability*, 44, 2024–2063.
- KALLENBERG, O. (1997): *Foundations of Modern Probability*. Springer, New York.
- KOLLER, D., AND N. FRIEDMAN (2009): *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, Cambridge, Massachusetts.

- KUERSTEINER, G. M., AND I. R. PRUCHA (2013): “Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity,” *Journal of Econometrics*, 174, 107–126.
- LAURITZEN, S. L. (1996): *Graphical Models*. Springer, New York.
- LAURITZEN, S. L., A. P. DAWID, B. N. LARSEN, AND H.-G. LEIMER (1990): “Independence Properties of Directed Markov Fields,” *Networks*, 20, 491 – 505.
- LEUNG, M. P. (2016): “Treatment and Spillover Effects under Network Interference,” *Working Paper*.
- MEINSHAUSEN, N., AND P. BÜHLMANN (2008): “High-Dimensional Graphs and Variable Selection With the LASSO,” *Annals of Statistics*, 34(3), 1436–1462.
- PENROSE, M. (2003): *Random Geometric Graphs*. Oxford University Press, Oxford, UK.
- POLLARD, D. (1990): *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 2, Institute of Mathematical Statistics, Hayward, USA.
- RINOTT, Y., AND V. ROTAR (1996): “A Multivariate CLT for Local Dependence with  $n^{1/2} \log n$  Rate and Applications to Multivariate Graph Related Statistics,” *Journal of Multivariate Analysis*, 56, 333–350.
- SONG, K. (2015): “Measuring the Graph Concordance of Locally Dependent Observations,” *arXiv:1504.03712v2 [stat.ME]*.
- STEIN, C. (1972): “A Bound for the Error in the Normal Approximation to the Distribution of a Sum of Dependent Random variables,” *Proceedings in the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 2, 583–602.
- TIKHOMIROV, A. N. (1980): “On the Convergence Rate in the Central Limit Theorem for Weakly Dependent Random Variables,” *Theory of Probability and Its Applications*, 25, 790–809.
- VAN DER VAART, A. W. (1996): “New Donsker Classes,” *Annals of Statistics*, 24, 2128–2140.
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer, New York, USA.
- WELLNER, J. A. (2005): *Empirical Processes: Theory and Applications*. Special Topics Course Notes at Delft Technical University.
- WU, W. B. (2005): “Nonlinear System Theory: Another Look at Dependence,” *Proceedings of National Academy of Science USA*, 102, 14150–14154.
- YUAN, D., AND L. LEI (2016): “Some Results Following from Conditional Characteristic functions,” *Communications in Statistics - Theory and Methods*, 45, 3706–3720.