The Annals of Applied Probability
1997, Vol. 7, No. 4, 1080–1105

ON COUPLING CONSTRUCTIONS AND RATES IN THE CLT FOR DEPENDENT SUMMANDS WITH APPLICATIONS TO THE ANTIVOTER MODEL AND WEIGHTED *U*-STATISTICS

By Yosef Rinott[1] and Vladimir Rotar[2]

University of California, San Diego and Russian Academy of Sciences

This paper deals with rates of convergence in the CLT for certain types of dependency. The main idea is to combine a modification of a theorem of Stein, requiring a coupling construction, with a dynamic set-up provided by a Markov structure that suggests natural coupling variables. More specifically, given a stationary Markov chain $\mathrm{X}^{(t)}$, and a function $U = U(\mathrm{X}^{(t)})$, we propose a way to study the proximity of U to a normal random variable when the state space is large.

We apply the general method to the study of two problems. In the first, we consider the antivoter chain $\mathrm{X}^{(t)} = \{X_i^{(t)}\}_{i \in \mathscr{V}}$, $t = 0, 1, \dots,$ where $\mathscr{V}$ is the vertex set of an n-vertex regular graph, and $X_i^{(t)} = +1$ or -1. The chain evolves from time t to $t+1$ by choosing a random vertex i, and a random neighbor of it j, and setting $X_i^{(t+1)} = -X_j^{(t)}$ and $X_k^{(t+1)} = X_k^{(t)}$ for all $k \neq i$. For a stationary antivoter chain, we study the normal approximation of $U_n = U_n^{(t)} = \sum_i X_i^{(t)}$ for large n and consider some conditions on sequences of graphs such that U_n is asymptotically normal, a problem posed by Aldous and Fill.

The same approach may also be applied in situations where a Markov chain does not appear in the original statement of a problem but is constructed as an auxiliary device. This is illustrated by considering weighted U-statistics. In particular we are able to unify and generalize some results on normal convergence for degenerate weighted U-statistics and provide rates.

1. Introduction and results.

1.1. *Background and motivation.* Consider a random quantity X and a real-valued function of X, $W = W(\mathrm{X})$. For example, X may be the state of a particle system, and W may count the number of particles in a given set or the number of sites having a particular value or the sum of the values in some or all sites. We are interested in conditions under which the distribution of W is close to normal. Henceforth we standardize W, assuming that

$$EW = 0, \qquad EW^2 = 1. \tag{1.1}$$

Received June 1996; revised May 1997.

[1]Research supported in part by NSF Grant DMS-95-04616.

[2]Research supported in part by the International Division of NSF Grant DMS-95-04616 and by the grant of the Russian Foundation for Basic Research 96-01-01229.

AMS 1991 *subject classifications*. Primary 60F05, 60K35; secondary 62E20, 60J10.

Key words and phrases. Stein's method, random graphs, distance regularity, Markov chains.

The main idea of this paper is to combine a suitable version of a theorem of Stein requiring a coupling construction with a dynamic set-up that suggests natural coupling variables. We start with Stein's method. Given a random variable (r.v.) W, Stein's framework is based on the construction of another variable W' (coupling) such that the pair (W, W') is exchangeable (i.e., their joint distribution is symmetric) and

$$E(W'|W) = (1-\lambda)W \tag{1.2}$$

for some positive $\lambda < 1$. Theorem 1.1 shows that a measure of proximity of W to normality may be provided in terms of this exchangeable pair and λ, with a good approximation requiring $W' - W$ to be sufficiently small.

We first quote a slightly modified but equivalent version of a theorem of Stein [(1986), page 35].

THEOREM 1.1. *For a function h: $\mathbb{R} \to \mathbb{R}$, set $\Phi h = \int_{-\infty}^{\infty} h(z)\Phi(dz)$ where Φ is the standard normal measure. Then for any exchangeable pair (W, W') satisfying (1.1), (1.2) and any continuously differentiable bounded function h,*

$$\begin{aligned} |Eh(W) - \Phi h| \leq {} & \frac{1}{\lambda}(\sup|h - \Phi h|)\sqrt{\mathrm{Var}\{E[(W'-W)^2|W]\}} \\ & + \frac{1}{4\lambda}(\sup|h'|)\, E|W'-W|^3, \end{aligned} \tag{1.3}$$

where h' is the derivative of h. Also, for all real x,

$$\begin{aligned} \sup_x |P(W \leq x) - \Phi(x)| \leq {} & \frac{2}{\lambda}\sqrt{\mathrm{Var}\{E[(W'-W)^2|W]\}} \\ & + \frac{1}{(2\pi)^{1/4}}\sqrt{\frac{1}{\lambda}E|W'-W|^3}. \end{aligned} \tag{1.4}$$

In concrete models below, the size or dimension of X is designated by a number n. For example, n may be the number of sites in a particle system or the number of summands in a particular sum. In other words, $\mathsf{X} = \mathsf{X}(n)$ where n is an integer parameter, and we will consider the asymptotic behavior of $W(\mathsf{X})$ for large n. Nevertheless, when it does not lead to a misunderstanding, we may omit n.

Let us turn to the dynamic setup and its relation to coupling constructions. First, consider the case when, for n fixed, the distribution of the quantity $\mathsf{X} = \mathsf{X}(n)$ is the stationary distribution of an ergodic Markov chain, say, $\{\mathsf{X}^{(t)}\} = \{\mathsf{X}^{(t)}(n)\}$, where $t = 0, 1, \ldots$ denotes the time variable. In certain applications, the stationary distribution is not given explicitly, which seems to make the study of the distribution of W difficult. However, under certain conditions, the combination of the Markov structure and the approach mentioned leads to a derivation of a normal approximation for W even without calculating the unknown stationary distribution, by the natural choice $(W, W') = (W(\mathsf{X}^{(t)}), W(\mathsf{X}^{(t+1)}))$. Exchangeability of this pair clearly holds if the chain $\{\mathsf{X}^{(t)}\}$ is stationary and reversible. Another condition for exchangeability is given in Lemma 1.1. As indicated above, this coupling should be expected to yield good rates if for the chain under consideration the r.v. $W(\mathsf{X}^{(t)})$ does not change much in one step, that is, if the difference $W(\mathsf{X}^{(t+1)}) - W(\mathsf{X}^{(t)})$ is small. We will demonstrate this approach by considering the so-called antivoter particle system [see, e.g., Liggett (1985) and references therein, and Aldous and Fill (1994) and below for details].

Furthermore, the above approach may also be useful if the original problem is not described in terms of a Markov chain. In this case, one may try to construct a suitable Markov chain, and use it as above in defining W'. Constructions of this type for certain problems were proposed by Stein (1986), and Diaconis (1977, 1989 and private communication). We will illustrate this possibility by considering weighted U-statistics, including the classical case of nondegenerate U-statistics, as well as certain types of degenerate U-statistics.

Returning to Stein's method, in Theorem 1.2 below, we will improve Theorem 1.1 in the following directions.

1. The last term in (1.4) may be crude: even when W is the normalized sum of n independent variables, it leads to a bound of the order $n^{-1/4}$; see Stein (1986). We improve this term when $|W' - W|$ is bounded. The latter condition holds in many situations in which Stein's method is useful, for example, when W counts vertices, edges or subgraphs having a certain property determined by some random process on a finite graph.
2. The improved rates apply not only to indicators of half lines as in (1.4) but to a broad class of nonsmooth functions h.
3. Theorem 1.2 extends the range of applications of this approach by replacing (1.2) by a weaker condition, allowing (1.2) to hold only approximately. The discussion of weighted nondegenerate U-statistics demonstrates the utility of this extension.

1.2. *A general theorem.* Following Bhattacharya and Ranga Rao (1986), and Rinott and Rotar (1996) (see both for further references), we define for a given function $h(x)$, $x \in \mathbb{R}$ and $\varepsilon > 0$,

$$h_\varepsilon^+(x) = \sup\{h(x+y): |y| \le \varepsilon\}, \qquad h_\varepsilon^-(x) = \inf\{h(x+y): |y| \le \varepsilon\},$$
$$\tilde{h}(x;\varepsilon) = h_\varepsilon^+(x) - h_\varepsilon^-(x).$$

Let $\mathscr{H}$ be a class of measurable functions on a real line such that the following hold:

1. All functions in $\mathscr{H}$ are uniformly bounded in absolute value by a constant assumed to be 1 without loss of generality;
2. For any real numbers c and d and for any $h \in \mathscr{H}$, the function $h(cx+d)$ belongs to $\mathscr{H}$;
3. For any $\varepsilon > 0$ and any $h \in \mathscr{H}$, the functions h_ε^+, h_ε^- are also in $\mathscr{H}$, and

$$\int \tilde{h}(x;\varepsilon)\Phi(dx) \le a\varepsilon \tag{1.5}$$

for some constant a which depends only on the class $\mathscr{H}$.

We assume without loss of generality that

$$a \geq \sqrt{2/\pi}. \tag{1.6}$$

The indicators of all half lines, or the indicators of all intervals compose classes which satisfy these conditions (with $a = \sqrt{2/\pi}$ and $a = 2\sqrt{2/\pi}$, respectively). As a less trivial example one may consider the indicators of finite or countable unions of disjoint intervals, such that in each union the distance between any two intervals is not less than a fixed positive constant. (Say, as in the set $\bigcup_{k=0}^{\infty}[2k, 2k+1]$.)

THEOREM 1.2. *Let* (W, W') *be exchangeable and (1.1) hold. Define the r.v.* $R = R(W)$ *by*

$$E(W'|W) = (1-\lambda)W + R, \tag{1.7}$$

where λ *is a number satisfying* $0 < \lambda < 1$. *Then*

$$\begin{aligned} \delta &:= \sup\{|Eh(W) - \Phi h|: h \in \mathscr{H}\} \\ &\leq \frac{6}{\lambda}\sqrt{\mathrm{Var}\{E[(W'-W)^2|W]\}} + 19\frac{\sqrt{ER^2}}{\lambda} + 6\sqrt{\frac{a}{\lambda}E|W'-W|^3}, \end{aligned} \tag{1.8}$$

where a *is the constant from (1.5).*

Also, if

$$|W' - W| \leq A \tag{1.9}$$

for a constant A, *then*

$$\delta \leq \frac{12}{\lambda}\sqrt{\mathrm{Var}\{E[(W'-W)^2|W]\}} + 37\frac{\sqrt{ER^2}}{\lambda} + 48\frac{aA^3}{\lambda} + 8\frac{aA^2}{\sqrt{\lambda}}. \tag{1.10}$$

REMARKS. For differentiable bounded functions h, a bound like (1.3) plus $\sqrt{ER^2}/\lambda$ with appropriate constants holds, generalizing (1.3) to the case that $R \neq 0$.

The constants in (1.8), (1.10) are not the best possible; more careful calculations would yield better ones. Note also that from (1.1) and (1.7) it follows that

$$E(W'-W)^2 = 2\lambda - 2E[WR(W)]. \tag{1.11}$$

The relations (1.11) and (1.9) imply

$$\frac{A^2}{\lambda} \geq 2 - 2\frac{E[WR(W)]}{\lambda}, \tag{1.12}$$

and hence the last term in (1.10) is not bigger than the third term plus $16aA^2\sqrt{|E(WR)|}/\lambda$. (This is obvious if either $A/\sqrt{\lambda} > 1$ or $\sqrt{|E(WR)|}/\sqrt{\lambda} \geq 1/2$. The latter inequality follows readily from (1.12) if $A/\sqrt{\lambda} \leq 1$.) In particular, if $R = 0$, the last term influences only the constant at the third one.

The bound (1.8) is a rather direct generalization of (1.4). The proof of (1.10), to be given in Section 4, is more complicated. It follows the main outline of

Stein's approach and makes use of some techniques developed in Rinott and Rotar (1996).

Let us discuss the conditions of Theorem 1.2. First, consider the exchangeability of (W, W') when the distribution of X coincides with the stationary distribution of a stationary ergodic Markov chain $\{\mathsf{X}^{(t)}\}$, and $(W, W') = (W(\mathsf{X}^{(t)}), W(\mathsf{X}^{(t+1)}))$ in distribution. By stationarity, $W(\mathsf{X}^{(t+1)})$ and $W(\mathsf{X}^{(t)})$ have the same marginal distributions; however, they may not be exchangeable. In this setup, if $W(\mathsf{X}^{(t)})$ is indeed close to normal when the size n of X is large, one may expect the pair $(W(\mathsf{X}^{(t)}), W(\mathsf{X}^{(t+1)}))$ to be close to bivariate normal. The latter distribution implies exchangeability, and therefore it is natural to expect this property at least asymptotically in n. In this context, the exchangeability condition appears to be in a sense almost necessary and certainly natural.

If the original Markov chain $\{\mathsf{X}^{(t)}\}$ is stationary and reversible, then the pair $(W(\mathsf{X}^{(t)}), W(\mathsf{X}^{(t+1)}))$ is clearly exchangeable. If the chain is not reversible (e.g, in the case of the antivoter chain; see below), the following simple lemma may be useful.

LEMMA 1.1. *Let* $\mathsf{X}^{(t)}$ *be a stationary process,* $T(\mathsf{X}^{(t)})$ *assume nonnegative integer values, and suppose* $T(\mathsf{X}^{(t+1)}) - T(\mathsf{X}^{(t)}) = +1$, 0 *or* -1. *Set* $W = f(T(\mathsf{X}^{(t)}))$, $W' = f(T(\mathsf{X}^{(t+1)}))$ *where* f *is a measurable function. Then* (W, W') *is an exchangeable pair.*

The proof is given in Section 4. Here note only that although $T(\mathsf{X}^{(t)})$ may not be a Markov chain, it is sufficiently similar to a birth and death chain, and the exchangeability of $(T(\mathsf{X}^{(t)}), T(\mathsf{X}^{(t+1)}))$ will be shown to be akin to the reversibility of birth and death chains.

Next, consider condition (1.7). If the pair (W, W') is close to bivariate normal, then the linearity of the conditional expectation as a function of W should hold approximately, indicating that (1.7) is a natural condition in the present set-up, and one may expect the remainder term R to be small. In fact, R can be viewed as a remainder term in the expansion of the conditional expectation of $W' - W$, centered at W.

Finally, we discuss the main term on the right-hand sides of (1.8) and (1.10) involving $\mathrm{Var}\{E[(W' - W)^2|W]\}$. When (W, W') are jointly normal, it is easy to verify that $E[(W' - W)^2|W] = \lambda^2 W^2 + \text{constant}$, and $\mathrm{Var}\{E[(W' - W)^2|W]\} = \lambda^4 \mathrm{Var}\{W^2\}$, which implies that the first term in the bounds of Theorems 1.1 and 1.2 has the order λ. This indicates that λ should be small, and then, if W is close to normal, one may indeed expect $\mathrm{Var}\{E[(W' - W)^2|W]\}$ to be small.

We turn now to applications.

1.3. *The antivoter model.* The antivoter model was introduced on infinite lattices by Matloff (1977); see also Liggett (1985) and references there. Donnelly and Welsh (1984) and Aldous and Fill (1994) consider the case of finite graphs. We describe a discrete time version of the antivoter chain on a finite graph, which for our goal reflects all the essential features of the usual continuous time model.

Consider an n-vertex r-regular graph G. To each vertex i of the graph we associate at time t a random variable $X_i^{(t)}$ which takes values $+1$ or -1. Set $\mathsf{X}^{(t)} = \{X_i^{(t)}\}$, and define a Markov chain by the following transition rule: at each time t, a vertex, say i, is chosen at random (all vertices equally likely), and then another vertex, say j, is chosen at random from $\mathscr{N}_i$, the set of neighbors of i with respect to the graph. Then $X_i^{(t+1)}$ is set to equal $-X_j^{(t)}$, and $X_k^{(t+1)} = X_k^{(t)}$ for all $k \neq i$. In words, the chain evolves by a random vertex looking at a random neighbor, and setting its value to the opposite of its neighbor.

Henceforth we assume that G is neither bipartite nor an n-cycle. This assumption implies that the set of all $2^n - 2$ configurations in which not all $X_i^{(t)}$ are identical is irreducible, and the support of the stationary distribution is that set; see Donnelly and Welsh (1984) and Aldous and Fill (1994). We assume also the chain $\mathsf{X}^{(t)}$ to be stationary, so sometimes we will omit the index t writing X_i for $X_i^{(t)}$, and so on.

Set $U = U_n(\mathsf{X}) = \sum_{i=1}^n X_i$, and $\sigma^2 = \sigma_n^2 = \operatorname{Var} U_n$. In many situations σ_n^2 has the order of n. We shall quote some facts about it below and in Section 2, along with a discussion on the asymptotic normality of U_n for suitable sequences of graphs. The latter issue was posed as an open problem by Aldous and Fill (1994).

Throughout this paper, the letter C stands for any universal constant, perhaps different in different formulas or on two sides of an inequality.

Let $W = W_n = U_n/\sigma_n$. Using the Markov structure of the model for the coupling construction as described in Section 1.2, we obtain the following theorem.

THEOREM 1.3. *For any n-vertex r-regular graph G, and any function $h \in \mathscr{H}$,*

$$\sup\{|Eh(W) - \Phi h|: h \in \mathscr{H}\} \le C\left(\frac{\sqrt{\operatorname{Var} Q}}{r\sigma^2} + a\frac{n}{\sigma^3}\right), \tag{1.13}$$

where

$$Q = Q_n = \sum_{i=1}^n \sum_{j \in \mathscr{N}_i} X_i X_j, \tag{1.14}$$

and a is the constant from (1.5).

So, for regular graphs the problem is reduced to the variance of Q, that is, to mixed moments of the fourth order under the stationary distribution of the chain. For some simple graphs, $\operatorname{Var} Q$ may be calculated or bounded in a direct way, leading to a CLT with rates for W_n; see examples below and in Sections 2.2 and 2.3. However, for more complicated graphs, it may not be easy to compute good bounds for $\operatorname{Var} Q$. We describe below a bound on $\operatorname{Var} Q$

for certain graphs. A more detailed discussion will be provided in Section 2. Our goal here is not to exhaust the subject but rather to demonstrate that, at least under certain conditions, Var Q and the resulting rate of convergence in Theorem 1.3 may be tractable.

To simplify notations, assume G to be connected, and denote by $\partial(i, j)$ the *distance* between vertices i, j, that is, the number of edges traversed in the shortest walk joining these vertices. Denote by d the *diameter*, that is, the maximal value of the distance function.

Suppose now that G is a *distance regular* graph, which means the following [see, e.g., Brouwer, Cohen and Neumaier (1989), Biggs (1993)]. For all $m = 1, \dots, d$ and any pair of vertices (i, j) with $\partial(i, j) = m$, the number of vertices $k \in \mathscr{N}_i$ for which $\partial(k, j) = m - 1$, as well as the number of those for which $\partial(k, j) = m + 1$, do not depend on the position of the vertices (i, j) but only on the distance m. Then, since we always assume the graph is regular, the number of $k \in \mathscr{N}_i$ for which $\partial(k, j) = m$ also depends only on m. Denote the last number by a_m. Note that, if r is the degree of the graph, a_m/r is the probability that a random neighbor of i has the same distance to j as i. Let $\alpha = \min\{a_m/r;\ m = 1, \dots, d\}$.

LEMMA 1.2. *For any G as defined above,*

$$\text{Var}\, Q \le C(2/\alpha)^d n^2 r. \tag{1.15}$$

Theorem 1.3 and Lemma 1.2 easily imply the proposition.

PROPOSITION 1.1. *Let G_n be a sequence of distance regular graphs having diameters d_n, characteristics $\alpha = \alpha_n$, as defined above, and degrees r_n. Let $r_n \to \infty$ and for some absolute strictly positive constants $\bar{d}, \bar{\alpha}, \bar{\sigma}$,*

$$d_n \le \bar{d}, \qquad \alpha_n \ge \bar{\alpha}, \qquad \sigma_n^2 \ge \bar{\sigma}^2 n. \tag{1.16}$$

Then the distribution of W_n converges to the standard normal distribution as $n \to \infty$.

EXAMPLE. Consider the *Hamming graph* which is a graph whose vertices are the k^d d-dimensional vectors with elements from a finite set of size k, two being adjacent when they differ in just one coordinate [see, e.g., Biggs (1993)].

For $d = 2$ it can be represented as a $k \times k$ matrix of vertices, and two distinct vertices are neighbors if there is a row or a column containing both.

The graph defined is distance regular with $n = k^d$, $r = d(k-1)$. Its diameter equals d, and $a_m = m(k-2),\ m = 1, \dots, d$. Thus $a_m/r = m(k-2)/[d(k-1)] > 1/(2d)$ for $k \ge 3$. We show in Section 2 that σ_n^2 has the order of n, and asymptotic normality of W_n, with $n = k^d$, d fixed and $k \to \infty$, follows from Proposition 1.1.

Condition (1.16) is not necessary. In Section 2 we will show that (1.16) may be weakened and consider other examples. In the same section we prove all assertions concerning the antivoter model.

1.4. *Weighted U-statistics.* In this section we consider weighted U-statistics, including classical U-statistics as a special case. The models below may serve as an illustration of the second approach mentioned in Section 1.1, when a Markov chain does not appear in the original statement of a problem but is constructed as an auxiliary device.

Let $\mathsf{X} = (X_1, \dots, X_n)$, where the X_i's are i.i.d. r.v.'s. Consider the U-statistic

$$U = U_n(\mathsf{X}) = \sum_{\mathsf{j}} w(\mathsf{j})\psi(X_{j_1}, \dots, X_{j_k}), \tag{1.17}$$

where k is a fixed integer less than or equal to n; the summation is over all $\mathsf{j} = (j_1, \dots, j_k)$ such that $j_1 \neq \cdots \neq j_k$ and $1 \leq j_p \leq n$, for $p = 1, \dots, k$; the function ψ is symmetric under permutation of its arguments, and $w(\mathsf{j})$ is a symmetric nonnegative weight function. Both ψ and the weights w may also depend on n, but as a rule this will be suppressed in the notations.

For an interesting application of weighted U-statistics, see, for example, Nowicki and Wierman (1988) who studied the asymptotic distribution of the number U_n of subgraphs of a random n-vertex graph which are isomorphic to a given graph. For this problem, Barbour, Karoński and Ruciński (1989) obtained sharp rates of normal convergence of $h(U_n)$ for smooth functions h, using a different variant of Stein's method.

Set $\Psi_{\mathsf{j}} = \psi(X_{j_1}, \dots, X_{j_k})$ and assume throughout this paper that

$$E\Psi_{\mathsf{j}} = 0, \qquad E\Psi_{\mathsf{j}}^2 = 1. \tag{1.18}$$

The symbol $C(k)$ below denotes a constant, perhaps different in different formulas, depending only on k.

Let $\mathsf{X}^* = (X_1^*, \dots, X_n^*)$ be an independent replica of X. Select in X one coordinate, say i, at random, and replace it by X_i^*. Denote the resulting vector by X'. Formally, $\mathsf{X}' = (X_1, \dots, X_{I-1}, X_I^*, X_{I+1}, \dots, X_n)$, where I is a r.v. taking values $1, \dots, n$ with equal probabilities, and independent of all other r.v.'s. Let $U' = U_n' = U_n(\mathsf{X}')$. Sometimes below we omit the index n. The above construction is similar to a coupling proposed by Stein in certain examples. Note also that the transition from X to X' may be viewed as the one-step evolution of a reversible Markov chain $\mathsf{X}^{(t)}$ for which the joint distribution of $(\mathsf{X}^{(t)}, \mathsf{X}^{(t+1)})$ coincides with that of $(\mathsf{X}, \mathsf{X}')$.

Set $W = W_n = U_n/\sigma_n$, and $W' = W_n' = U_n'/\sigma_n$, where $\sigma_n^2 = EU_n^2$.

For an integer i and $\mathsf{j} \ni i$, denote by $\Psi_{i,\mathsf{j}}^*$ the random variable which is the result of replacing in $\psi(X_{j_1}, \dots, X_{j_k})$ the r.v. X_i by X_i^*. By construction

$$E(U' - U \mid \mathsf{X}, \mathsf{X}^*) = \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathsf{j} \ni i} w(\mathsf{j})(\Psi_{i,\mathsf{j}}^* - \Psi_{\mathsf{j}}). \tag{1.19}$$

We will see later that this relation allows us to verify conditions (1.2) or (1.7).

The main result of this section concerns the degenerate case (see below for details), but first we touch on the following.

Nondegenerate statistics. This is the case when

$$P(E\{\psi(X_1, \dots, X_k) \mid X_1\} = 0) < 1, \tag{1.20}$$

and one may hope for normal convergence without strong conditions on ψ.

First note that (1.20) implies the existence of a $b_0 > 0$ depending on ψ and the distribution of X_i such that $\sigma_n^2 \geq b_0^2 \sum_{i=1}^n v_i^2$, where $v_i = v_i^{(n)} = \sum_{\mathrm{j} \ni i} w(\mathrm{j})$ [see, e.g., Lee (1990)].

As will be shown in Section 3, if (1.20) holds, the condition (1.2) is not fulfilled, but the version (1.7) with a remainder is true with $\lambda = 1/n$ and an R for which

$$ER^2 \leq C(k) \frac{E\Psi^2}{n^2} \frac{\sum_{i, m; i \neq m} v_{im}^2}{b_0^2 \sum_i v_i^2}, \tag{1.21}$$

where the r.v. $\Psi = \psi(X_1, ..., X_k)$, and $v_{im} = v_{im}^{(n)} = \sum_{\mathrm{j} \ni i, m} w(\mathrm{j})$. In particular, for complete statistics, that is, for the case $w(\mathrm{j}) \equiv 1$, one has $v_i^{(n)} \geq C(k) n^{k-1}$, $v_{im}^{(n)} \leq n^{k-2}$, and $ER^2 \leq C(k) E\Psi^2 / b_0^2 n^3$. This means that the second term in (1.7) is smaller in order than the linear part λW, making it possible to consider U statistics by the method under discussion. As we will see, under mild conditions the same is true for weighted statistics.

For complete statistics, the CLT in the case (1.20) is well elaborated; see, for example, the book by Lee (1990), which contains much of the literature on the subject. To our knowledge, weighted statistics are not as well investigated; in particular, the accuracy of the normal approximation has not been described yet. Using the approach of this paper, we will consider in Section 3 the case of bounded ψ, and will prove the following.

Let $\hat{v}_n = \max_{1 \leq i \leq n} v_i$, $\bar{v}_n^2 = (1/n) \sum_{i=1}^n v_i^2$, and $\beta_i^2(n) = (\sum_{m=1, m \neq i}^n v_{im}^2) / (\sum_{m=1, m \neq i}^n v_{im})^2$. Note that $(1/n) \leq \beta_i^2(n) \leq 1$, and $\beta_i^2(n) = O(1/n)$ if all v_{im} have "the same order."

PROPOSITION 1.2. *Let ψ be bounded uniformly in n, and (1.20) hold. Then for a constant L depending only on ψ, k and the distribution of X_i,*

$$\sup\{|Eh(W) - \Phi h| \colon h \in \mathscr{H}\} \leq L \left\{ \left(\frac{\sum_{1 \leq i \leq n} v_i^2 \beta_{in}^2}{\sum_{1 \leq i \leq n} v_i^2} \right)^{1/2} + a \frac{1}{\sqrt{n}} \left(\frac{\hat{v}_n}{\bar{v}_n} \right)^3 \right\}, \tag{1.22}$$

where a is the same constant as in (1.5).

It is easy to see that, say, for weights such that all v_i's are of the same order and the same is true for v_{im}'s (which is obviously the case for complete statistics), the right-hand side of (1.22) is $O(1/\sqrt{n})$.

Suppose now (1.20) does not hold. In this case, the characteristics R and λ depend on the "degree of degeneracy." In particular, if

$$E\{\psi(X_1, \dots, X_k) \mid X_1, \dots, X_{k-1})\} = 0 \quad \text{a.s.}, \tag{1.23}$$

then $E\{\Psi^*_{i,\mathrm{j}} \mid \mathsf{X}\} = 0$, and consequently (1.19) implies (1.2), or in other words $R = 0$, with $\lambda = k/n$. (The coefficient k arises since in the double sum in (1.19) each j is counted k times.)

The gap between (1.23) and (1.20) is filled as follows: if for some $l \leq k$, one has $P(E\{\psi(X_1, \dots, X_k) \mid X_1, \dots, X_l)\} = 0) < 1$ but $E\{\psi(X_1, \dots, X_k) \mid X_1, \dots, X_{l-1})\} = 0$ a.s., then (1.7) holds with $\lambda = l/n$ and, under mild conditions, with a "small" R. We omit the details of this intermediate case and turn to the following.

Degenerate U-statistics. For simplicity, we restrict ourselves to the case $k = 2$, and consider the r.v.:

$$U = U_n(\mathsf{X}) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \psi(X_i, X_j), \tag{1.24}$$

where w_{ij} are nonnegative weights, $w_{ij} = w_{ji}$, $w_{ii} = 0$. Assume

$$E(\psi(X_1, X_2) \mid X_2) = 0 \quad \text{a.s.}, \tag{1.25}$$

which, in view of (1.18), implies in particular that $\sigma_n^2 := EU_n^2 = \sum_{i,j=1}^n w_{ij}^2$.

In the case (1.25), U_n may be asymptotically normal for either of the following reasons: the weights w_{ij} are different for different (i, j) in a way that ensures a weak dependence between the summands in (1.24) or $\psi = \psi_n$ depends on n in a specific way (see below for details).

The former factor was investigated in many papers; see, for example, Janson (1984), O'Neil and Redner (1993), and references there. In particular, O'Neil and Redner (1993) showed the following.

Let the weight function w_{ij} be uniformly bounded, and H_n be the maximum number of nonzero weights in each collection $w_{i1}, \dots, w_{in}$. Suppose there are constants k, K, and α, $0 \leq \alpha < 1$, such that

$$0 < kn^{1+\alpha} \leq \sigma_n^2 \quad \text{and} \quad H_n \leq Kn^{\alpha}. \tag{1.26}$$

Then U_n is asymptotically normal.

The latter of the mentioned factors was considered in Hall (1984) where for complete statistics, that is, for the case $w_{ij} \equiv 1$, the following assertion was proved. Let $\psi = \psi_n$ and $\gamma_n(x, y) = E\{\psi_n(X_1, x)\psi_n(X_1, y)\}$. Set $\Psi_n = \psi_n(X_1, X_2)$, $\Gamma_n = \gamma_n(X_1, X_2)$. Then U_n is asymptotically normal under the condition

$$n^{-1} E\Psi_n^4 + E\Gamma_n^2 \to 0. \tag{1.27}$$

Below we unify both factors and provide some rates of convergence. To this end, set $\mathsf{w}_i = (w_{i1}, \dots, w_{in})$, and define

$$D_n = \frac{\sum_{k,l=1}^n (\mathsf{w}_k \cdot \mathsf{w}_l)^2}{(\sum_{k=1}^n |\mathsf{w}_k|^2)^2},$$

where $|\cdot|$ and $\cdot$ denote length and dot product, respectively. The characteristic D_n plays an essential role in an assertion below. One may say that D_n

measures the extent to which the summands in (1.24) are dependent. For a complete statistic, that is, when $w_{ij} \equiv 1$, one has $D_n = 1$, while, say, in the case (1.26) $D_n \to 0$. This will be proved in the end of Section 3.

First, consider a corollary of Theorem 1.4. This corollary is informative if the quantities $|\mathrm{w}_i|$ do not differ much in order of magnitude, or in other words if the statistic is not "too asymmetric." Set $M_n = (\max_{i\le n} |\mathrm{w}_i|)/(\min_{i\le n} |\mathrm{w}_i|)$.

PROPOSITION 1.3. *There exists an absolute constant C such that*

$$\delta_n := \sup\{|Eh(W_n) - \Phi h|\colon h \in \mathscr{H}\} \le C M_n^{3/2}\left(a\frac{E\Psi_n^4}{\sqrt{n}} + D_n E\Gamma_n^2\right)^{1/2}. \tag{1.28}$$

In particular, if M_n and $E\Psi_n^4$ are uniformly bounded, for W_n to be asymptotically normal it is sufficient that $D_n E\Gamma_n^2 \to 0$. The last condition reflects the influence of both mentioned factors.

Proposition 1.3 is a corollary of the following more precise but somewhat less transparent assertion. Let

$$F_n = \frac{\sum_{i=1}^n |\mathrm{w}_i|^4}{(\sum_{i=1}^n |\mathrm{w}_i|^2)^2}.$$

Note that $n^{-1} \le F_n \le 1$, and $F_n \le M_n^4/n$, so if M_n is uniformly bounded, $F_n = O(1/n)$.

THEOREM 1.4. *There exists an absolute constant C such that*

$$\delta_n \le C\big(aE\Psi_n^4(nF_n^3)^{1/4} + D_n E\Gamma_n^2\big)^{1/2}. \tag{1.29}$$

In the case (1.26), we have $D_n = o(1)$ as mentioned above, and it is easy to verify that $F_n = O(1/n)$. Therefore, the above-mentioned result of O'Neil and Redner (1993) follows from Theorem 1.4 save that, in order to obtain rates, we require the finiteness of the fourth moments. The bound (1.29) includes the same characteristic $E\Gamma_n^2$ as in (1.27); however, Hall's result does not follow completely from (1.29) due to the first term in the brackets in (1.29), which for complete statistics has the order $n^{-1/2}E\Psi_n^4$. This may be connected with the fact that we deal with rates; providing rates leads to some crudeness in calculations.

Hall applied his theorem to prove the asymptotic normality of the squared error of some nonparametric density estimators with certain bandwidths. Our result provides rates for this problem.

The rest of the paper is structured as follows. In Section 2 we prove Theorem 1.3 and other results from Section 1.3 and consider distance regular graphs in more detail. Section 3 contains the proofs of assertions concerning U-statistics. In Section 4 we prove Theorem 1.2 and Lemma 1.1.

2. The antivoter model: proofs and additional examples.

In this section we elaborate on and prove the results presented in Section 1.3 and discuss some conditions on sequences of graphs for asymptotic normality of W_n.

2.1. *Proof of Theorem* 1.3. For a fixed time t, set $\mathsf{X} = \mathsf{X}^{(t)}$ and $\mathsf{X}' = \mathsf{X}^{(t+1)}$. We apply Theorem 1.2 to the pair $(W, W') = (W(\mathsf{X}), W(\mathsf{X}'))$. In the present case, the Markov chain $\mathsf{X}^{(t)}$ is not reversible. However, exchangeability of the pair (W, W') follows from Lemma 1.1 and the relation $U(\mathsf{X}) = 2T(\mathsf{X}) - n$, where $T = T(\mathsf{X})$ counts the number of vertices i where $X_i = 1$.

Next we verify (1.7). Set

$$a(\mathsf{X}) = \text{the number of edges } \{i, j\} \text{ with } X_i = X_j = 1,$$

$$b(\mathsf{X}) = \text{the number of edges } \{i, j\} \text{ with } X_i = X_j = -1,$$

$$c(\mathsf{X}) = \text{the number of edges } \{i, j\} \text{ with } X_i \neq X_j.$$

Observe that for a regular graph of a degree r

$$T(\mathsf{X}) = [2a(\mathsf{X}) + c(\mathsf{X})]/r, \qquad n - T(\mathsf{X}) = [2b(\mathsf{X}) + c(\mathsf{X})]/r.$$

Instead of W and W', we consider the corresponding unstandardized variables U and U'. It is not hard to see that

$$P(U' - U = -2 \mid \mathsf{X}) = 2a(\mathsf{X})/(rn), \qquad P(U' - U = 2 \mid \mathsf{X}) = 2b(\mathsf{X})/(rn). \tag{2.1}$$

Therefore

$$E[(U' - U) \mid \mathsf{X}] = 4b(\mathsf{X})/(rn) - 4a(\mathsf{X})/(rn) = 2(n - 2T)/n = -2U/n.$$

Dividing through by σ_n and taking expectation conditioned on W, (1.7) follows with $\lambda = 2/n$ and $R = 0$. Also, (1.9) holds with $A = 2/\sigma_n$, since for any n, $|U' - U|$ takes one of the values 2 or 0 only. It follows from the discussion after (1.10) that apart from the first term on the right-hand side of (1.10), the remaining terms do not exceed aCn/σ_n^3.

We now compute the first term on the r.h.s. of (1.10). By (2.1) we have

$$E[(U' - U)^2 \mid \mathsf{X}] = 4[2a(\mathsf{X})/(rn) + 2b(\mathsf{X})/(rn)].$$

Next, note the relations

$$2a(\mathsf{X}) + 2b(\mathsf{X}) + 2c(\mathsf{X}) = rn, \qquad Q = 2a(\mathsf{X}) + 2b(\mathsf{X}) - 2c(\mathsf{X}),$$

which imply

$$4[a(\mathsf{X}) + b(\mathsf{X})] = Q + rn. \tag{2.2}$$

It follows readily that $E[(U' - U)^2 \mid \mathsf{X}] = 2(Q + rn)/(rn)$ and

$$\operatorname{Var} E[(W' - W)^2 \mid \mathsf{X}] = C \operatorname{Var}\left[\frac{Q}{rn\sigma_n^2}\right] = \frac{C}{(rn)^2\sigma_n^4} \operatorname{Var} Q.$$

Since W is a function of X, $\operatorname{Var} E[(W' - W)^2 \mid W] \le \operatorname{Var} E[(W' - W)^2 \mid \mathsf{X}]$, and (1.13) follows. □

We obtained (1.13) from Theorem 1.2. Obviously, it applies to step functions h. If $\sigma_n^2 \ge Cn$, which is the case for many graphs, the right-hand side of (1.13) becomes $(C/rn)\sqrt{\operatorname{Var} Q} + C/\sqrt{n}$. Note that for step functions, (1.4) would only imply the bound $(C/rn)\sqrt{\operatorname{Var} Q} + Cn^{-1/4}$.

2.2. *Bounds on* Var Q. The bound (1.13) contains two quantities: σ_n and Var Q. For the former we will use the following lemma from Aldous and Fill (1994) concerning r-regular graphs.

LEMMA 2.1. *For a set of vertices A, let $\bar{\kappa}(A)$ be the number of edges with either both ends in A or both in A^c, and set $\kappa = \inf_A \bar{\kappa}(A)$, where the infimum is taken over all subsets of vertices. Then*

$$\frac{2\kappa}{r} \le \sigma_n^2 \le n.$$

The main aim of the discussion below is to suggest ways to compute Var Q (or upper bounds for it) for certain graphs and find conditions for asymptotic normality of W for natural sequences of graphs.

Donnelly and Welsh (1984) computed correlations of the type needed here in terms of the following dual process, called the annihilating random walk. Assume that particles are located at some of the vertices of the graph G. Let B denote the set of vertices which are initially occupied by particles. At times $t = 1, 2, \ldots,$ a random particle is selected, and it moves to a randomly chosen neighboring vertex. If the latter vertex is already occupied, the two particles are annihilated, and the vertex becomes empty. For a set B of even cardinality, let $\tau(B)$ denote the total number of steps taken by all particles until all are annihilated, and as before, let X_i denote the value at vertex i in a stationary antivoter chain on G. We need the following fact, which is a special case of Theorem 1 of Donnelly and Welsh (1984); see also Griffeath (1979).

LEMMA 2.2. *We have*

$$\begin{aligned} P(X_i X_j = 1) &= P(\tau(\{i, j\}) \text{ is even}),\\ P(X_i X_j X_k X_l = 1) &= P(\tau(\{i, j, k, l\}) \text{ is even}) \end{aligned} \tag{2.3}$$

for any distinct vertices i, j, k, l.

Assume now that G is *distance regular*, with degree r, and diameter d. In addition to the notations of Section 1.3, for each pair of vertices (i, j) with $\partial(i, j) = m$, let

b_m be the number of vertices k in $\mathscr{N}_i$ which satisfy $\partial(k, j) = m + 1$;
c_m be the number of vertices k in $\mathscr{N}_i$ which satisfy $\partial(k, j) = m - 1$.

The number a_m is defined as in Section 1.3.

The numbers $\{a_m, b_m, c_m\}$ in some arrangement, sometimes as a matrix, are called the *intersection array* of G; see Biggs (1993), Brouwer, Cohen and Neumaier (1989) for details. The latter book is devoted to distance regular graphs and provides a host of examples.

Note that for any distance regular graph, $\tau(\{i, j\})$ will have the same distribution if we allow only the particle initially at i to move and count the number of steps until it reaches j. Also, a walk from i to j can be viewed as a sequence of steps; each step brings the particle a unit distance closer, or farther away, or maintains the distance to its target at j. It is easy to see that for a distance regular graph, the probability of such a particular sequence depends only on $\partial(i, j)$, and therefore the distribution of $\tau(\{i, j\})$ depends only on $\partial(i, j)$.

Set $p_m = P(\tau(\{i, j\})$ is even) where $m = \partial(i, j)$, $m = 1, \dots, d$, and $\varepsilon_m = |p_m - 1/2|$. Then by Lemma 2.2, $|E(X_i X_j)| = |2p_m - 1| = 2\varepsilon_m$. Observe now that

$$\varepsilon_m \le \varepsilon_1. \tag{2.4}$$

Indeed, for a particle moving from i to j with $\partial(i, j) = m$, let P be the probability that, when entering the neighborhood of j for the first time, the particle has already made an even number of steps. Then $p_m = p_1 P + (1 - p_1)(1 - P)$, which easily implies (2.4). Thus for $i \neq j$,

$$|E(X_i X_j)| \le 2\varepsilon_1. \tag{2.5}$$

Next, consider an annihilating random walk with four initial particles. Since the particles move one at a time, after annihilation of the first pair, the second pair will be at some positive distance and considerations similar to the above imply that

$$|E(X_i X_j X_k X_l)| = |2P(\tau(\{i, j, k, l\}) \text{ is even}) - 1| \le 2\varepsilon_1. \tag{2.6}$$

From (2.5) and (2.6) we have

$$\operatorname{Var} Q \le EQ^2 \le C\{nr + n^2 r^2 \varepsilon_1\}, \tag{2.7}$$

where C is an absolute constant. So, it remains to provide an appropriate bound for ε_1. The following lemma gives such a bound in terms of the *intersection array* matrix

$$L = \frac{1}{r}\begin{pmatrix} a_1 & b_1 & 0 & \dots & \dots & 0 & 0 \\ c_2 & a_2 & b_2 & 0 & 0 & \dots & 0 \\ 0 & c_3 & a_3 & b_3 & 0 & \dots & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & \dots & \dots & 0 & c_d & a_d \end{pmatrix}. \tag{2.8}$$

LEMMA 2.3.

$$\varepsilon_1 = \frac{1}{2r}\left|(I + L)^{-1}_{11}\right|. \tag{2.9}$$

PROOF. By the formula of total probability, we readily obtain $p_m = [c_m(1 - p_{m-1}) + a_m(1 - p_m) + b_m(1 - p_{m+1})]/r$, $m = 1, \dots d$, where $p_0 = 1$ and $p_{d+1} = 0$. Note that $c_m + a_m + b_m = r$, $m = 1, \dots, d$ (where $b_d = 0$), and in particular $a_1 + b_1 = r - 1$, since $c_1 = 1$. In matrix notation we then have $(I + L)\mathrm{p} = \mathrm{v}$, where p is the d-vector whose components are $p_1, \dots, p_d$, and v is a d-vector whose first entry is $1 - 1/r$ and the remaining entries are equal to 1. It follows that $(I+L)(\mathrm{u}-\mathrm{p})=\mathrm{s}$, where u is a d-vector with all entries equal to 1/2, and the entries of s are all 0 save the first entry which is $1/(2r)$. The matrix L is strictly substochastic, hence its spectral radius is strictly smaller than 1 and thus $(I+L)$ is invertible. We have $\mathrm{u}-\mathrm{p}=(I+L)^{-1}\mathrm{s}$, and in particular $1/2-p_1=(1/2r)(I+L)^{-1}_{11}$. □

We remark briefly that a similar approach may be attempted with weaker assumptions on the graph. In more general cases, the matrix L should be replaced by the adjacency matrix of the graph. See Donnelly and Welsh (1984) for such results.

For certain graphs the right-hand side of (2.9) may be estimated directly, yielding a bound on Var Q.

EXAMPLE. The *complete k-partite graph* is a union of k subgraphs, such that there are no edges between pairs of vertices in the same subgraph, and all other pairs of vertices are connected. If each subgraph has l vertices, this is a distance regular graph with the number of vertices $n=kl$, diameter $d=2$, and degree $r=(k-1)l$. The intersection array is determined by $b_1=l-1$, $c_2=r$, and hence $a_2=0$.

Direct calculation shows that for any $k\geq 3$, the absolute value of the determinant of $I+L$ is bounded below by 1, and also $|(I+L)^{-1}_{11}|\leq 1$. So, by Lemma 2.3 and (2.7), Var $Q\leq Cn^2r$.

To evaluate σ_n, we apply Lemma 2.1. Let $k\geq 3$, and $A=A_1\cup A_2\cup\cdots\cup A_k$, where A_i is a set of vertices from the ith subgraph, $i=1,\dots,k$. The number of edges with either both ends in A or both in A^c equals $\sum_{1\leq i<j\leq k}\{|A_i|\cdot|A_j|+(l-|A_i|)(l-|A_j|)\}$, and it is easy to see that each term in the latter sum is bounded below by $l^2/2$. It follows that $\bar{\kappa}(A)\geq\binom{k}{2}(l^2/2)$, and by Lemma 2.1 we obtain $\sigma_n^2\geq n/2$.

The above bounds and Theorem 1.3 imply that, as $k\to\infty$, or $l\to\infty$ (or both), W_n converges to normal in distribution.

The following lemma may be useful when $|(I+L)^{-1}_{11}|$ cannot be calculated explicitly.

LEMMA 2.4. *If $a_i/r\geq\alpha$ for all $i=1,\dots,d$ and some $\alpha>0$, then $|(I+L)^{-1}_{11}|\leq(2/\alpha)^d$.*

PROOF. We can write $I+L=(\alpha+1)I+D$ where D is a tridiagonal substochastic matrix. The eigenvalues of such a D (and L) are real and their absolute values are no larger than 1 [see, e.g., Horn and Johnson (1985), Problem 8.3.7 and Theorem 8.1.22]. Therefore the eigenvalues of $I+L$ are all in the interval $[\alpha,2]$. The same applies to every principal submatrix of $I+L$. This shows that α^d is a positive lower bound on the determinant of $I+L$, and 2^{d-1} is an upper bound on any principal minor of order $d-1$. It follows that $|(I+L)^{-1}_{11}|\leq 2^{d-1}\alpha^{-d}$. □

Lemmas 2.4 and 2.3 and inequality (2.7) imply Lemma 1.2 from Section 1.3, and hence Proposition 1.1. We applied Proposition 1.1 in that section to the Hamming graph. To complete the discussion of that example, it remains to show that $\sigma_n^2 \geq \bar{\sigma}^2 n$ for some $\bar{\sigma} > 0$. To this end, one may again apply Lemma 2.1.We omit the calculations, which are similar to those used for the previous example.

Next we supplement the above conditions on the graph structure by one more which is different in nature.

2.3. *A simple condition on the graph structure.* For the complete graph, direct calculations yield the bounds Var $Q \leq Cn^3$ and $\sigma_n^2 \geq cn$ where C and c are universal constants. Together with (1.13) this leads to the rate $O(1/\sqrt{n})$ in the CLT for W_n.

The condition discussed below applies to graphs which are in some sense close to being complete but need not be distance regular. We consider rather simple graphs; however, even for those, the stationary distribution does not seem to be computable explicitly.

Consider a regular graph of a degree r, and set $\mu(j) = \sum_{i \in \mathcal{N}_j} |(\mathcal{N}_i - \{j\})\Delta(\mathcal{N}_j - \{i\})|$, where $\mathcal{N}_i - \{j\}$ is the set of all vertices in $\mathcal{N}_i$ except for the vertex j, Δ denotes symmetric difference of sets and $|\cdot|$ their cardinality.

Note that for any complete graph, as well as for any collection of nonconnected complete subgraphs, $\mu(j) = 0$ for all j. In general, $|\mu(j)| < 2r^2$.

LEMMA 2.5. *For any regular graph of degree r,*

$$(2.10) \qquad \left| EQ + \frac{nr}{2r-1} \right| \leq \frac{1}{2r-1} \sum_{j=1}^{n} \mu(j), \qquad \mathrm{Var}\{Q\} \leq 16nr^2 + 2n \sum_{j=1}^{n} \mu(j).$$

EXAMPLE. Consider a collection of k nonconnected complete subgraphs each of degree r (i.e., each having $r+1$ vertices). In each subgraph, choose a pair of vertices and disconnect them; call one vertex chosen "left", and the other "right." Starting with the first subgraph, we connect each right vertex with the left vertex of the next subgraph, and the right vertex of the last subgraph with the left vertex of the first subgraph, obtaining a "cycle" of subgraphs.

It is easy to calculate that for the graph constructed, any "chosen" vertex j has $|\mu(j)| \leq (r-1) + (r-1) + 2(r-1) = 4(r-1)$, and for any "nonchosen" vertex $\mu(j) \leq 2$. Thus $\sum_{j=1}^{n} \mu(j) \leq 2k(4r-4) + k(r-1)2 \leq 10kr < 10n$, since $n = k(r+1)$.

So, in this case $\mathrm{Var}\{Q\} \leq 16nr^2 + 20n^2$.

Similarly, for any regular graph of degree r , which may be divided into subgraphs such that in each subgraph no more than d_0 vertices have connections with vertices from the other subgraphs, and apart from these vertices all others are connected with each vertex of the same subgraph,

$$\mathrm{Var}\{Q\} \leq C(d_0)(nr^2 + n^2),$$

where the constant $C(d_0)$ depends only on d_0.

The above bounds together with Theorem 1.3 again imply asymptotic normality of W_n with the rate $O(1/\sqrt{n}+1/r)$, provided $\sigma_n^2 \geq \bar{\sigma}^2 n$ and $r = r_n \to \infty$.

REMARK. Instead of the numbers $\mu(j)$, consider the random variables $\nu(j) = \sum_{i\in\mathcal{N}_j} \sum_{k\in\mathcal{N}_i-\{j\}} X_k - (r-1)\sum_{i\in\mathcal{N}_j} X_i$. It can be shown with some elementary calculations that $|\nu(j)| \leq \mu(j)$ everywhere, and therefore $E|\nu(j)| \leq \mu(j)$. The bounds in (2.10) remain valid if $\mu(j)$ is replaced by $E|\nu(j)|$ for all j. Such bounds are less explicit but essentially more precise. Below we prove this last version.

PROOF OF LEMMA 2.5. Let $\{\mathrm{X}^{(t)}\}$ be the Markov chain described in Section 1.3, and a pair of vectors (X, Y) have the same joint distribution as $(\mathrm{X}^{(t)}, \mathrm{X}^{(t+1)})$. Observe that

$$E(Y_iY_j|\mathrm{X}) = \frac{1}{nr}\sum_{k\in\mathcal{N}_i}(-X_kX_j) + \frac{1}{nr}\sum_{k\in\mathcal{N}_j}(-X_kX_i) + \frac{n-2}{n}X_iX_j. \tag{2.11}$$

Therefore

$$EQ = E\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}E(Y_iY_j|\mathrm{X}) = -\frac{2}{nr}E\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}\sum_{k\in\mathcal{N}_j}X_kX_i + \frac{n-2}{n}EQ.$$

This leads to

$$\begin{aligned}
EQ &= -\frac{1}{r}E\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}\sum_{k\in\mathcal{N}_j}X_kX_i = -\frac{1}{r}E\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}\biggl(1 + \sum_{k\in\mathcal{N}_j-\{i\}}X_kX_i\biggr)\\
&= -n - E\sum_{i=1}^n X_i\biggl(\frac{1}{r}\sum_{j\in\mathcal{N}_i}\sum_{k\in\mathcal{N}_j-\{i\}}X_k\biggr)\\
&= -n - \frac{r-1}{r}\sum_{i=1}^n\sum_{j\in\mathcal{N}_i}E(X_iX_j) - \frac{1}{r}\sum_{i=1}^n E[X_i\nu(i)]\\
&= -n - \frac{r-1}{r}EQ - \frac{1}{r}\sum_{i=1}^n E[X_i\nu(i)],
\end{aligned}$$

which easily implies the first bound in (2.10).

The computation of EQ^2 is similar but more tiresome. We provide only a sketch. If the numbers i_1, i_2, i_3, i_4 are all distinct,

$$\begin{aligned}
&E(Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}|\mathrm{X})\\
&\quad= \frac{1}{nr}\biggl(\sum_{k\in\mathcal{N}_{i_1}}(-X_kX_{i_2}X_{i_3}X_{i_4}) + \sum_{k\in\mathcal{N}_{i_2}}(-X_kX_{i_1}X_{i_3}X_{i_4})\biggr)\\
&\qquad + \frac{1}{nr}\biggl(\sum_{k\in\mathcal{N}_{i_3}}(-X_kX_{i_1}X_{i_2}X_{i_4}) + \sum_{k\in\mathcal{N}_{i_4}}(-X_kX_{i_1}X_{i_2}X_{i_3})\biggr)\\
&\qquad + \frac{n-4}{n}(X_{i_1}X_{i_2}X_{i_3}X_{i_4}).
\end{aligned}\tag{2.12}$$

Let $\Sigma' = \sum_{i=1}^{n} \sum_{j \in \mathscr{N}_i} \sum_{m \in \mathscr{N}_i, m \neq j}$, and let Σ'' be the sum over all pairs (i, j), (m, l) such that $j \in \mathscr{N}_i$, $l \in \mathscr{N}_m$, and all indices are distinct. We have $EQ^2 = 2nr + 4\Sigma' E(Y_j Y_m) + \Sigma'' E(Y_i Y_j Y_m Y_l)$.

Noting that $|\Sigma' E(Y_j Y_m)| \leq nr^2$, and using (2.12) and the last expression for EQ^2, one can conclude by calculations similar to the above for EQ, that $\Sigma'' E(X_i X_j X_m X_l) \leq -(n-2)EQ + 10nr^2 - ((r-1)/r)EQ^2 + n \sum_{m=1}^{n} E|\nu(m)|$.

The last inequality implies $EQ^2 \leq (r/(2r-1))(16nr^2 + n\sum_{m=1}^{n} E|\nu(m)|) + (2r/(2r-1))EQ - (nr/(2r-1))EQ$, and therefore, since $EQ < 0$ (see below) and $|EQ| \leq nr$,

$$\operatorname{Var}\{Q\} \leq 16nr^2 + n \sum_{m=1}^{n} E|\nu(m)| + |EQ| \left| \frac{nr}{2r-1} + EQ \right|.$$

The second bound in (2.10) now follows from the first. □

The relation $EQ < 0$ follows from $n/2 > \operatorname{Var} U = 2E[a(\mathrm{X}) + b(\mathrm{X})]/r = (EQ + rn)/(2r)$, where the first two facts can be found in Aldous and Fill (1994) and the third follows from (2.2).

2.4. *Some additional remarks on dependency structures.* In the antivoter model, the largest among the covariances $\operatorname{Cov}(X_i, X_j)$ tends to occur when i and j are neighbors with respect to the graph. These correlations are negative; see Donnelly and Welsh (1984). However there is no reason to expect that in general a stronger notion of negative dependence, such as *negative association* as defined by Joag-Dev and Proschan (1983), holds.

In a more general context, consider $\mathrm{X}(n) = \{X_i(n)\}_{i \in V}$, a real-valued stationary process (so that time is not indicated), where $\mathscr{V}$ is the vertex set of a graph G with n vertices, (e.g., a spin system on a graph, the antivoter model, etc.). Let $U(n) = \sum_{i=1}^{n} X_i(n)$, and $\sigma_n^2 = \operatorname{Var} U(n)$. Conditions for asymptotic normality of $U(n)$ assuming positive or negative association of $\{X_i(n)\}$ and related, somewhat weaker, assumptions can be found in Newman (1984); see also references therein. For example, one may derive from these results that if the system $\mathrm{X}(n)$ is positively or negatively associated and if $\sum_{i=1}^{n} E|X_i(n)|^3/[\sum_{i=1}^{n} \operatorname{Var} X_i(n)]^{3/2} \to 0$, then the distribution of $[U(n) - EU(n)]/\sigma_n$ converges to a normal distribution, provided

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{i=1}^{n} \sum_{j:\, j \neq i} \operatorname{Cov}(X_i(n), X_j(n)) = 0. \tag{2.13}$$

Newman applied results of this type to the Ising model under conditions which guarantee the required association property.

Returning to the antivoter model, note that if σ_n^2 is of order larger than $n^{2/3}$, Theorem 1.3 shows that asymptotic normality of W as $n \to \infty$ holds provided $(\sqrt{\operatorname{Var} Q}/r\sigma_n^2) \to 0$. The latter condition is equivalent to

$$\lim_{n\to\infty} \frac{1}{r^2 \sigma_n^4} \sum_{i} \sum_{j \in \mathscr{N}_i} \sum_{k} \sum_{l \in \mathscr{N}_k} \operatorname{Cov}(X_i(n) X_j(n), X_k(n) X_l(n)) = 0. \tag{2.14}$$

Since in this case association does not hold, it is not surprising that moments of order higher than 2 appear. Here, and in other examples, moments of order 4 arise.

Preliminary calculations show that our approach may be applicable to certain particle systems. The present approach may yield results (and rates) for such systems when association (or the FKG condition) does not hold.

3. *U*-statistics: proofs.

We start with the general scheme. As in Section 1.4, denote by I a r.v. taking values $1, \dots, n$ with equal probabilities and independent of all other r.v.'s. Then

$$U' - U = \sum_{\mathrm{j} \ni I} w(\mathrm{j})(\Psi^*_{I,\mathrm{j}} - \Psi_{\mathrm{j}}). \tag{3.1}$$

Note that (3.1) implies (1.19).

To arrive at (1.7), we use the Hoeffding (1948) representation (more precisely H-projection) and define the functions $g(x) = E\psi(X_1, X_2, \dots, X_{k-1}, x)$ and $\theta(x_1, \dots, x_k) = \psi(x_1, \dots, x_k) - g(x_1) - \cdots - g(x_k)$. Set $\Theta_{\mathrm{j}} = \theta(X_{j_1}, \dots, X_{j_k})$ for $\mathrm{j} = (j_1, \dots, j_k)$, and for $i \in \mathrm{j}$ denote by $\Theta^*_{i,\mathrm{j}}$ the random variable which is the result of replacing in $\theta(X_{j_1}, \dots, X_{j_k})$ the r.v. X_i by X_i^*. Proceeding from (1.19) it is straightforward to calculate that

$$\begin{aligned} &E(U' - U \mid \mathrm{X}, \mathrm{X}^*) \\ &\quad = -\frac{1}{n}U + \frac{1}{n}\Biggl\{\sum_{i=1}^n \sum_{\mathrm{j} \ni i} w(\mathrm{j})\Theta^*_{i,\mathrm{j}} - (k-1)\sum_{\mathrm{j}} w(\mathrm{j})\Theta_{\mathrm{j}} \\ &\qquad + \sum_{i=1}^n \sum_{\mathrm{j} \ni i} w(\mathrm{j}) g(X_i^*)\Biggr\}. \end{aligned} \tag{3.2}$$

Since $E\{g(X_i^*) \mid \mathrm{X}\} = 0$, (3.2) implies (1.7) with $\lambda = 1/n$ and

$$R = \frac{1}{n\sigma_n} E\Biggl\{\sum_{i=1}^n \sum_{\mathrm{j} \ni i} w(\mathrm{j})\Theta^*_{i,\mathrm{j}} - (k-1)\sum_{\mathrm{j}} w(\mathrm{j})\Theta_{\mathrm{j}} \mid U\Biggr\}. \tag{3.3}$$

PROOF OF PROPOSITION 1.2. Considering this to be a preliminary result, we give only a sketch of the proof.

As was noted in Section 1.4 in the case (1.20), $\sigma_n^2 \ge b_0^2 \sum_{i=1}^n v_i^2$ for a $b_0 > 0$. Furthermore, if the sets of indices j and j′ do not have more than one common index, Θ_{j} and $\Theta_{\mathrm{j}'}$ are uncorrelated. The same also holds for pairs $(\Theta^*_{i,\mathrm{j}}, \Theta^*_{i',\mathrm{j}'})$ and $(\Theta^*_{i,\mathrm{j}}, \Theta_{\mathrm{j}'})$. Note also that $|E\Theta^*_{i,\mathrm{j}}\Theta^*_{i',\mathrm{j}'}| \le CE\Psi^2$, and the same holds for the products of the other r.v.'s mentioned above. This implies that for a constant $C(k)$,

$$ER^2 \le C(k)\frac{1}{n^2\sigma_n^2} \sum_{i,m;\, i \ne m} \sum_{\mathrm{j} \ni i,m} \sum_{\mathrm{j}' \ni i,m} w(\mathrm{j})w(\mathrm{j}')E\Psi^2 = C(k)\frac{E\Psi^2}{n^2\sigma_n^2} \sum_{i,m;\, i \ne m} v_{im}^2$$

and (1.21) follows.

Let now $|\psi| \le b$ for some $b > 0$. Then, by (3.1),

$$|W' - W| \le \frac{2b \max_{1\le i\le n} v_i}{\sigma_n} \le \frac{2b \max_{1\le i\le n} v_i}{b_0\sqrt{\sum_{1\le i\le n} v_i^2}}. \tag{3.4}$$

To apply Theorem 1.2, it remains to estimate $\mathrm{Var}\{E(W'-W)^2 \mid W)\}$. Let $\eta_{i,\mathrm{j}} = \Psi^*_{i,\mathrm{j}} - \Psi_{\mathrm{j}}$. Then

$$\mathrm{Var}\{E(W'-W)^2 \mid W)\} \le \mathrm{Var}\{E(W'-W)^2 \mid X, X^*)\} \le \frac{1}{n^2\sigma_n^4}\,\mathrm{Var}\, K, \tag{3.5}$$

where $K = \sum_i (\sum_{\mathrm{j}\ni i} w(\mathrm{j})\eta_{i,\mathrm{j}})^2$.

The calculation of Var K is somewhat tiresome but straightforward. For the case $k = 2$, similar calculations will be provided in the proof of Theorem 1.4 below. So, we omit the details, noting only that the leading term in the expansion for EK^2 happens to be less than the square of the leading term in EK, and the remaining leading term is less than $C(k)b^4(\sum_{1\le i,\, m\le n,\, i\ne m} v_{im}^2)(\sum_{1\le i\le n} v_i^2)$.

Combining all bounds above and Theorem 1.2, one easily comes to (1.22). □

PROOF OF THEOREM 1.4. Let $\eta_{ij} = \eta_{ijn} = \psi_n(X_i^*, X_j) - \psi_n(X_i, X_j)$. In this particular case, (3.1) takes the form

$$U' - U = 2\sum_{j=1,\, j\ne I}^{n} w_{Ij}\eta_{Ij}, \tag{3.6}$$

and therefore

$$E(U' - U \mid \mathrm{X}, \mathrm{X}^*) = \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1,\, j\ne i}^{n} w_{ij}\eta_{ij}. \tag{3.7}$$

In view of (1.25) this implies

$$E(U' - U \mid U) = -\frac{2}{n}U.$$

Thus in this case, $R = 0$ and $\lambda = 2/n$.

Observe now that $E\eta_{12}^4 \le CE\Psi_n^4$ and by (1.18), $\sigma_n^2 = \sum_{i,j=1}^n w_{ij}^2 = \sum_{i=1}^n |\mathrm{w}_i|^2$. Making use of these relations, (3.6) and (1.25), we have

$$\begin{aligned} E(W'-W)^4 &= \frac{16}{n\sigma_n^4} E\sum_{i=1}^{n}\biggl(\sum_{j=1}^{n} w_{ij}\eta_{ij}\biggr)^4 \\ &= \frac{16}{n\sigma_n^4}\sum_{i=1}^{n}\biggl(\sum_{j=1}^{n} w_{ij}^4 E\eta_{ij}^4 + \sum_{j,\,k=1;\,k\ne j}^{n} w_{ij}^2 w_{ik}^2 E\eta_{ij}^2\eta_{ik}^2\biggr) \qquad (3.8) \\ &\le \frac{16}{n\sigma_n^4} E\eta_{12}^4 \sum_{i,j,k=1}^{n} w_{ij}^2 w_{ik}^2 \le \frac{C}{n\sigma_n^4} E\Psi_n^4 \sum_{i=1}^{n} |\mathrm{w}_i|^4 = \frac{CE\Psi_n^4}{n} F_n. \end{aligned}$$

We get from (3.8) for the last term in (1.8) that

$$\sqrt{E|W' - W)|^3/\lambda} \le C(E\Psi_n^4)^{3/8}(nF_n^3)^{1/8}. \tag{3.9}$$

Furthermore, by (3.6),

$$\mathrm{Var}\{E(W'-W)^2 \,|\, \mathrm{X}, \mathrm{X}^*)\} \le \frac{16}{n^2\sigma_n^4} \mathrm{Var}\, K, \tag{3.10}$$

where

$$K = \sum_{i=1}^n \biggl(\sum_{j=1}^n w_{ij}\eta_{ij} \biggr)^2.$$

In order to compute Var K note first that in the present degenerate case,

$$EK = \sum_{i=1}^n \sum_{j=1}^n w_{ij}^2 E\eta_{ij}^2 = (E\eta_{12}^2)\sigma_n^2. \tag{3.11}$$

Next consider EK^2. In the bound below we replace fourth order moments such as $E\{\eta_{12}^2\eta_{13}^2\}$, or $E\{\eta_{12}^2\eta_{13}\eta_{23}\}$ and similar ones by $E\eta_{12}^4$. Similar moments, in which there is an index of η which appears only once, always vanish in the degenerate case. We obtain readily

$$\begin{aligned} EK^2 &\le (E\eta_{12}^2)^2 \biggl(\sum_{i,\, j,\, k,\, l;\, k\ne i,\, j;\, l\ne i,\, j} w_{ij}^2 w_{kl}^2 \biggr) \\ &\quad + CE\eta_{12}^4 \biggl(\sum_{i,\, j,\, k;\, k\ne i,\, j} w_{ij}^2 w_{ik} w_{jk} \biggr) \\ &\quad + C|E\{\eta_{13}\eta_{14}\eta_{23}\eta_{24}\}| \biggl(\sum_{i,\, j,\, k,\, l} w_{ik}w_{il}w_{jk}w_{jl} \biggr). \end{aligned} \tag{3.12}$$

The first term is less than the square of (3.11). The second is bounded by

$$\begin{aligned} &CE\psi_n^4(X_1, X_2) \sum_i \sum_j w_{ij}^2 \biggl(\sum_k w_{ik} w_{jk} \biggr) \\ &\quad \le CE\Psi_n^4 \biggl(\sum_i \sum_j w_{ij}^4 \biggr)^{1/2} \biggl(\sum_i \sum_j (\mathrm{w}_i \cdot \mathrm{w}_j)^2 \biggr)^{1/2} \\ &\quad \le CE\Psi_n^4 \biggl(\sum_i \sum_j w_{ij}^4 \biggr)^{1/2} \biggl(\sum_i |\mathrm{w}_i|^2 \biggr) \\ &\quad = CE\Psi_n^4 \biggl(\sum_i \sum_j w_{ij}^4 \biggr)^{1/2} \sigma_n^2 = CE\Psi_n^4 F_{1n}^{1/2} \sigma_n^4, \end{aligned}$$

where the last equality defines F_{1n}.

The third term in (3.12) does not exceed

$$CE\gamma_n^2(X_1, X_2) \sum_{k=1}^n \sum_{l=1}^n (\mathrm{w}_k \cdot \mathrm{w}_l)^2.$$

Combining all these bounds, for the first term in (1.8), we have

$$\frac{1}{\lambda}\sqrt{\mathrm{Var}\{E(W'-W)^2\,|\,W)\}} \le \frac{1}{\lambda}\sqrt{\mathrm{Var}\{E(W'-W)^2\,|\,\mathrm{X},\mathrm{X}^*)\}}$$
$$\le C\{D_n E\Gamma_n^2 + F_{1n}^{1/2} E\Psi_n^4\}^{1/2}.$$

Note that for $w_{ij} \equiv 1$ in the brackets in the right-hand side, we have exactly the expression from (1.27), but we should also take into account the bound in (3.9). It remains to apply (1.8) and to compare all the terms in the final bound. When doing this, one should use that $F_{1n} \le F_n$ and $F_n \ge 1/n$. We omit simple calculations. □

In conclusion we prove that $D_n \to 0$ under condition (1.26). Indeed, setting $\bar{w} = \max_{i,j} w_{ij}$, we have $(\mathrm{w}_k \cdot \mathrm{w}_l) \le \bar{w}^2 K n^\alpha$ and

$$\sum_{k\neq l} (\mathrm{w}_k \cdot \mathrm{w}_l)^2 \le \bar{w}^2 K n^\alpha \sum_{k,l}\sum_j w_{kj} w_{lj} = \bar{w}^2 K n^\alpha \sum_{l,j} w_{lj}\Big(\sum_k w_{jk}\Big)$$
$$\le \bar{w}^3 (K n^\alpha)^2 \sum_{l,j} w_{lj} = \bar{w}^3 (K n^\alpha)^2 \sum_l \Big(\sum_j w_{lj}\Big)$$
$$\le \bar{w}^4 (K n^\alpha)^3 n = \bar{w}^4 K^3 n^{1+3\alpha}.$$

Hence, $D_n = O(n^{\alpha-1}) = o(1)$ for $\alpha < 1$.

4. Proofs of Theorem 1.2 and Lemma 1.1.

PROOF OF THEOREM 1.2.

LEMMA 4.1. *Let P be a probability measure on $\mathbb{R}$. For any bounded measurable function h and $t \in (0,1)$, define*

$$h_t(x) = \int h(x+ty)\Phi(dy).$$

Then

$$\sup\left\{\left|\int h\,d(P-\Phi)\right| : h \in \mathscr{H}\right\} \le 2.8 \sup\left\{\left|\int h_t\,d(P-\Phi)\right| : h \in \mathscr{H}\right\} + 4.7at, \tag{4.1}$$

where a is the same as in (1.5).

This lemma is close to Lemma 2.11 in Götze (1991). To prove (4.1) it suffices to apply a standard smoothing inequality [see Bhattacharya and Ranga Rao (1986), Lemma 11.4, page 95] and to use the closure property of $\mathscr{H}$ w.r.t. the operations $h_\varepsilon^+, h_\varepsilon^-$, and the fact that for any signed measure μ,

$$\int h\,d(\mu * \Phi_t) = \int h_t\,d\mu,$$

where $\Phi_t(x) = \Phi(x/t)$. We omit the details. □

Fix for now $t \in (0,1)$ and set

$$f(x) = \frac{1}{\phi(x)} \int_{-\infty}^{x} (h_t(u) - \Phi h_t)\phi(u)\,du, \tag{4.2}$$

where $\phi(\cdot)$ is the standard normal density. Differentiation of f yields the well-known relation [Stein (1972), (1986)]

$$f'(x) - xf(x) = h_t(x) - \Phi h_t. \tag{4.3}$$

Exchangeability of (W, W') and (1.7) imply

$$\begin{aligned} 0 &= E\{(W' - W)[f(W') + f(W)]\} \\ &= 2E\{f(W)(W' - W)\} \\ &\quad + E\{(W' - W)[f(W') - f(W)]\} \\ &= -2\lambda E\{Wf(W)\} + 2E\{f(W)R(W)\} + E\{(W' - W)[f(W') - f(W)]\}, \end{aligned}$$

and hence

$$E\{Wf(W)\} = \frac{E\{(W' - W)[f(W') - f(W)]\}}{2\lambda} + \frac{E\{f(W)R(W)\}}{\lambda}. \tag{4.4}$$

Together with (4.3) this implies

$$\begin{aligned} Eh_t(W) - \Phi h_t \;&=\; Ef'(W) - \frac{E\{(W' - W)[f(W') - f(W)]\}}{2\lambda} - \frac{E\{f(W)R(W)\}}{\lambda} \\ &= \frac{1}{2\lambda} E\{f'(W)[2\lambda - 2E[WR(W)] - (W' - W)^2]\} \\ &\quad + \frac{1}{\lambda} E\{f'(W)E[WR(W)] - f(W)R(W)\} \\ &\quad - \frac{1}{2\lambda} E\{(W' - W)[f(W') - f(W) - (W' - W)f'(W)]\} \\ &:= J_1 + J_2 + J_3. \end{aligned}$$

Since $|h| \le 1$,

$$|f| \le \sqrt{2\pi} \le 2.6, \;\; |f'| \le 4 \tag{4.5}$$

[see, e.g., Stein (1986)].

From (4.5) and (1.11) it follows that

$$\begin{aligned} |J_1| &\le \frac{2}{\lambda} E|E\{2\lambda - 2E[WR(W)] - (W' - W)^2 \mid W\}| \\ &\le \frac{2}{\lambda} \sqrt{E\{2\lambda - 2E[WR(W)] - E[(W' - W)^2 \mid W]\}^2} \\ &= \frac{2}{\lambda} \sqrt{\mathrm{Var}\{E[(W' - W)^2 | W]\}}. \end{aligned} \tag{4.6}$$

In view of (4.5) and (1.1),

$$\begin{aligned}|J_2| &\le \frac{4}{\lambda}E|WR(W)| + \frac{2.6}{\lambda}E|R(W)| \\ &\le \frac{1}{\lambda}\Big[4\sqrt{EW^2ER^2(W)} + 2.6\sqrt{ER^2(W)}\Big] = \frac{6.6}{\lambda}\sqrt{ER^2(W)}.\end{aligned} \tag{4.7}$$

When providing (4.8) below, we again use (4.5), take into account that, in view of (4.3), $f''(x) = f(x) + xf'(x) + h_t'(x)$ and denote by τ a r.v. which is uniformly distributed on [0,1] and independent of all other r.v.'s under consideration. Setting $V = W' - W$ we have

$$\begin{aligned}|2\lambda J_3| &= |E\{(W'-W)[f(W') - f(W) - (W'-W)f'(W)]\}| \\ &= |E\{V^3(1-\tau)f''(W+\tau V)\}| \\ &= |E\{V^3(1-\tau)[f(W+\tau V) + (W+\tau V)f'(W+\tau V) + h_t'(W+\tau V)]\}| \\ &\le 1.3E|V|^3 + 2E|V^3(|W| + |W'|)| + |E\{V^3(1-\tau)h_t'(W+\tau V)\}| \\ &\le 5.3A^3 + |E\{V^3(1-\tau)h_t'(W+\tau V)\}|.\end{aligned} \tag{4.8}$$

[In the last inequality we used (1.9) and (1.1).]

Using $h_t'(x) = -(1/t)\int h(x+ty)\phi'(y)\,dy$, and $\int \phi'(x)\,dx = 0$, we have

$$\begin{aligned}&|E\{V^3(1-\tau)h_t'(W+\tau V)\}| \\ &\quad= \frac{1}{t}\left|E\left\{V^3(1-\tau)\int h(W+\tau V+ty)\phi'(y)\,dy\right\}\right| \\ &\quad= \frac{1}{t}\left|E\left\{V^3(1-\tau)\int[h(W+\tau V+ty) - h(W+\tau V)]\phi'(y)\,dy\right\}\right| \\ &\quad\le \frac{A^3}{t}\int E\{(1-\tau)|h(W+\tau V+ty) - h(W+\tau V)|\}|\phi'(y)|\,dy \\ &\quad\le \frac{A^3}{t}\int E\{(1-\tau)[h^+_{A+t|y|}(W) - h^-_{A+t|y|}(W)]\}|\phi'(y)|\,dy \\ &\quad= \frac{A^3}{2t}\int E\tilde{h}(W, A+t|y|)|\phi'(y)|\,dy.\end{aligned}$$

Set $\delta = \sup\{|Eh(W) - \Phi h|: h \in \mathscr{H}\}$, and denote by Z a standard normal random variable. From the last inequality, the closure property of $\mathscr{H}$ w.r.t. the operations h_ε^+, h_ε^- and (1.5), it follows that

$$\begin{aligned}&|E\{V^3(1-\tau)h_t'(W+\tau V)\}| \\ &\quad\le \frac{A^3}{2t}\left\{\int\big[E\tilde{h}(W, A+t|y|) - E\tilde{h}(Z, A+t|y|)\big]|\phi'(y)|\,dy\right. \\ &\qquad\qquad\left. + \int E\tilde{h}(Z, A+t|y|)|\phi'(y)|\,dy\right\} \\ &\quad\le \frac{A^3}{2t}\left\{2\delta\int|\phi'(y)|\,dy + \int a(A+t|y|)|\phi'(y)|\,dy\right\} \\ &\quad\le \frac{A^3}{2t}\{2\delta + aA\} + \frac{aA^3}{2}.\end{aligned} \tag{4.9}$$

Denoting $\sup\{|Eh_t(W) - \Phi h_t|: h \in \mathscr{H}\}$ by δ_t, and collecting (4.6)–(4.9), we obtain

$$\delta_t \le \frac{2}{\lambda}\sqrt{\mathrm{Var}\{E[(W'-W)^2|W]\}} + \frac{6.6}{\lambda}\sqrt{ER^2(W)} + \frac{5.3}{2\lambda}A^3 + \frac{1}{4\lambda}aA^3 + \frac{1}{4\lambda}\frac{(2\delta + aA)A^3}{t}. \tag{4.10}$$

From (4.1) we have

$$\delta \le 2.8\delta_t + 4.7at.$$

It remains to substitute (4.10) in the last inequality and minimize the right-hand side of the resulting inequality in t. Eventually, using in particular, (1.6), we obtain

$$\delta \le \frac{5.6}{\lambda}\sqrt{\mathrm{Var}\{E[(W'-W)^2|W]\}} + \frac{18.5}{\lambda}\sqrt{ER^2(W)} + 10\frac{aA^3}{\lambda} + 3.7\frac{aA^2}{\sqrt{\lambda}} + 5.2\sqrt{\frac{a\delta A^3}{\lambda}}. \tag{4.11}$$

The latter inequality can be solved in δ to yield (1.10). We again omit straightforward calculations. □

PROOF OF LEMMA 1.1. The lemma can be restated as follows: *Let* $\{T^{(t)}\}$ *be a stationary nonnegative, integer valued process satisfying* $T^{(t+1)} - T^{(t)} = +1$, 0 *or* -1. *Then* $(T^{(t)}, T^{(t+1)})$ *is an exchangeable pair. In particular, for any measurable* f, *if* $W = f(T^{(t)})$, $W' = f(T^{(t+1)})$, *then* (W, W') *is an exchangeable pair.*

For integers i, j in the range of $T^{(t)}$, set $\pi_i = P(T^{(t)} = i)$ and $p_{ij} = P(T^{(t+1)} = j \mid T^{(t)} = i)$. These probabilities do not depend on t. Let $\boldsymbol{\pi}$ denote the row vector whose components are π_i, and let P denote the matrix whose entries are p_{ij}; $\boldsymbol{\pi}$ and P may have an infinite dimension. By stationarity the equation $\boldsymbol{\pi}P = \boldsymbol{\pi}$ holds (although $T^{(t)}$ need not be a Markov chain). The same equation arises in birth and death chains, and it is well known that if it has a solution, then it is unique, it can be written explicitly and it satisfies $\pi_i p_{ij} = \pi_j p_{ji}$ (which implies reversibility for birth and death chains). Here, the latter relation is equivalent to $P(T^{(t)} = i,\ T^{(t+1)} = j) = P(T^{(t)} = j,\ T^{(t+1)} = i)$, implying that $(T^{(t)}, T^{(t+1)})$ is an exchangeable pair. □

REFERENCES

ALDOUS, D. and FILL, J. A. (1994). Reversible Markov chains and random walks on graphs. Unpublished manuscript.

BARBOUR, A. D., KAROŃSKI, M. and RUCIŃSKI, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* 47 125–145.

BHATTACHARYA, R. N. and RANGA RAO, R. (1986). *Normal Approximation and Asymptotic Expansion*. Krieger, Melbourne, FL.

BIGGS, N. (1993). *Algebraic Graph Theory*. Cambridge Univ. Press.

BROUWER, A. E., COHEN, A. M. and NEUMAIER, A. (1989). *Distance-Regular Graphs*. Springer, Berlin.

DIACONIS, P. (1977). The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* 5 72–81.

DIACONIS, P. (1989). An example of Stein's method. Technical report, Dept. Statistics, Stanford Univ.

DONNELLY, P. and WELSH, D. (1984). The antivoter problem: random 2-colourings of graphs. In *Graph Theory and Combinatorics* (B. Bollobás, ed.) 133–144. Academic Press, New York.

GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* 19 724–739.

GRIFFEATH, D. (1979). *Additive and Cancellative Interacting Particle Systems. Lecture Notes in Math.* 724. Springer, Berlin.

HALL, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* 14 1–16.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* 19 293–325.

HORN, R. A. and JOHNSON, C. A. (1985). *Matrix Analysis*. Cambridge Univ. Press.

JANSON, S. (1984). The asymptotic distribution of incomplete U-statistics. *Z. Wahrsch. Verw. Gebiete* 66 495–505.

JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* 11 286–295.

LEE, A. J. (1990). *U-statistics: Theory and Practice*. Dekker, New York.

LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer, New York.

MATLOFF, N. (1977). Ergodicity conditions for a dissonant voting model. *Ann. Probab.* 5 371–386.

NEWMAN, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. In *Inequalities in Statistics and Probability* (Y. L. Tong, ed.) 127–140. IMS, Hayward, CA.

NOWICKI, K. and WIERMAN, J. C. (1988). Subgraph counts in random graph using incomplete U-statistics methods. *Discrete Math.* 72 299–310.

O'NEIL, K. and REDNER, R. A. (1993). Asymptotic distributions of weighted U-statistics of degree 2. *Ann. Probab.* 21 1159–1169.

RINOTT, Y. and ROTAR, V. (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate, and applications to multivariate graph related statistics. *J. Multivariate Anal.* 56 333–350.

STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 2 583–602. Univ. California Press, Berkeley.

STEIN, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, CA.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093
E-MAIL: yrinott@ucsd.edu

CENTRAL ECONOMIC–MATHEMATICAL INSTITUTE
RUSSIAN ACADEMY OF SCIENCES
MOSCOW 117418
RUSSIA
AND
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093
E-MAIL: rotar@cemi.rssi.ru