Electronic Journal of Statistics
Vol. 15 (2021) 2336–2359
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1839

On a complete and sufficient statistic for the correlated Bernoulli random graph model

Donniell E. Fishkind

Department of Applied Mathematics and Statistics,
Johns Hopkins University,
Baltimore, MD 21218
e-mail: def@jhu.edu

Avanti Athreya

Department of Applied Mathematics and Statistics,
Johns Hopkins University,
Baltimore, MD 21218
e-mail: dathrey1@jhu.edu

Lingyao Meng

Department of Applied Mathematics and Statistics,
Johns Hopkins University,
Baltimore, MD 21218
e-mail: lmeng2@jhu.edu

Vince Lyzinski

Department of Mathematics,
University of Maryland, College Park,
College Park, MD 20742
e-mail: vlyzinsk@umd.edu

Carey E. Priebe

Department of Applied Mathematics and Statistics,
Johns Hopkins University,
Baltimore, MD 21218
e-mail: cep@jhu.edu

Abstract: Inference on vertex-aligned graphs is of wide theoretical and practical importance. There are, however, few flexible and tractable statistical models for correlated graphs, and even fewer comprehensive approaches to parametric inference on data arising from such graphs. In this paper, we consider the correlated Bernoulli random graph model (allowing different Bernoulli coefficients and edge correlations for different pairs of vertices), and we introduce a new variance-reducing technique—called *balancing*—that can refine estimators for model parameters. Specifically, we construct a disagreement statistic and show that it is complete and sufficient; balancing can be interpreted as Rao-Blackwellization with this disagreement statistic. We show that for unbiased estimators of functions of model parameters, balancing generates uniformly minimum variance unbiased estimators (UMVUEs). However, even when unbiased estimators for model parameters do *not* exist—which, as we prove, is the case with both the heterogeneity correlation and the total correlation parameters—balancing is still useful, and lowers mean squared error. In particular, we demonstrate how balancing can improve the efficiency of the alignment strength estimator for the total correlation, a parameter that plays a critical role in graph matchability and graph matching runtime complexity.

MSC2020 subject classifications: Primary 62F10, 62B05.
Keywords and phrases: Rao-Blackwell, Lehmann-Scheffé, total correlation (graph), alignment strength, graph matching.

Received March 2020.

1. Overview

Paired random graphs with a natural alignment between their vertex sets arise in a wide variety of application domains; for example, the interaction dynamics of the same set of users across two social media platforms, or a pair of connectomes (brain graphs) as imaged from two different subjects of the same species. Given a pair of such graphs, the problem of *graph matching*—that is, optimally aligning the two vertex sets in order to minimize edge disagreements, usually with the purpose of obtaining the natural alignment—has a rich mathematical history, and graph matching now plays a fundamental role in algorithms for machine learning and pattern recognition; see the excellent surveys in [2, 5, 13].

The *correlated Bernoulli random graph model*, described in Section 2, is the focus of our work in this paper. It is a versatile model used to describe two graphs that are correlated with each other across a natural alignment between their vertex sets. The model allows for different probabilities of adjacency for different pairs of vertices, and allows for different edge correlations between different pairs of vertices across the natural alignment. This model is simple enough to be theoretically and computationally tractable, yet it is rich enough to successfully describe real data, and it has been profitably employed in this capacity; see, for example, [3, 10, 11].

The contributions of this paper fall into two groups, the second group utilizing the machinery of the first group.

Our first group of contributions: In the context of a correlated Bernoulli random graph model, we introduce a "smoothing" procedure, called *balancing*, which reduces the mean-squared error for any estimator of a function of model parameters; specifically, for any estimator S of a function of model parameters $g(\theta)$, the balanced estimator $\overline{S}$ has the same bias as S, but has lower variance. Indeed, under a nondegeneracy condition, we prove in Theorem 4 that if S is an unbiased estimator of $g(\theta)$ then $\overline{S}$ is the UMVU estimator of $g(\theta)$; this is because $\overline{S}$ is a Rao-Blackwellization of S via the *disagreement statistic* $\mathcal{H}$, and we prove in our main result Theorem 2 that $\mathcal{H}$ is complete and sufficient, under the nondegeneracy condition. We also prove in Theorem 3, under the nondegeneracy condition, that if S is an unbiased estimator of $g(\theta)$ then any statistic T is also an unbiased estimator of $g(\theta)$ if and only if $\overline{S} = \overline{T}$.

These results should not be taken for granted; in Example 4 we illustrate that even knowing, hence fixing, the mean of the adjacency probabilities creates a violation of the nondegeneracy condition, and the conclusions of the above theorems then will indeed fail, in general.

Our second group of contributions of this paper focuses on very recent advances in [4] regarding the correlated Bernoulli random graph model. Specifically, the paper [4] introduced and showed the importance of a novel model parameter called *total correlation*, which combines inter- and intra-graph contributions to a unified measure of the correlation between the pair of graphs. The authors empirically demonstrated—in broad families within the model—that graph matching complexity and matchability are each functions of total correlation. They also proved that the statistic called *alignment strength* is a strongly consistent estimator of total correlation.

Our second group of contributions: In the context of a correlated Bernoulli random graph model, the alignment strength statistic $\mathfrak{str}$ was shown in [4] to be a strongly consistent estimator of total correlation ϱ_T between the pair of graphs; however, we point out here that $\mathfrak{str}$ is **not a balanced statistic**, hence, as noted above, the mean squared error in estimating ϱ_T is reduced by using $\overline{\mathfrak{str}}$ instead. We then prove (in Theorems 7, 8, 9) that there do not exist unbiased estimators for several correlation parameters, including ϱ_T. Empirical experiments in Section 5 suggest that balancing the numerator and denominator of $\mathfrak{str}$ separately, which we call the *modified alignment strength* $\mathfrak{str}'$, often has less bias than $\overline{\mathfrak{str}}$ as an estimator of ϱ_T, always has less variance than $\overline{\mathfrak{str}}$, and we conjecture that $\mathfrak{str}'$ always has less mean square error than $\overline{\mathfrak{str}}$ in estimating ϱ_T.

The organization of this paper is as follows. The correlated Bernoulli random graph model, important functions of the parameters, and important statistics are described in Section 2. Our main results are stated in Section 3 and proved in Section 4. Empirical demonstrations are in Section 5.

2. Correlated Bernoulli random graphs

We begin by describing the correlated Bernoulli random graph model. It consists of a pair of random graphs; without loss of generality these graphs are on the same vertex set. (Indeed, the natural alignment between their vertex sets is a bijection, and the associated one-to-one correspondence can be thought of as an identification.) For simplicity of further notation, let us suppose that the N (= number-of-vertices-choose-two) pairs of vertices are arbitrarily ordered.

Define the set $\mathcal{R} := \{(p_1, p_2, \ldots, p_N, \varrho_1, \varrho_2, \ldots, \varrho_N) : p_1, p_2, \ldots, p_N, \varrho_1, \varrho_2, \ldots, \varrho_N \in [0,1]\}$.

Definition 1 (Correlated Bernoulli Random Graph Model)**.** The parameter space for the correlated Bernoulli random graph model, denoted Θ, is any particular subset of $\mathcal{R}$, possibly a proper subset. For each $(p_1, p_2, \ldots, p_N, \varrho_1, \varrho_2, \ldots, \varrho_N) \in \Theta$, the pair of random graphs are described as follows. For each $i = 1, 2, \ldots, N$, the indicator random variable X_i for adjacency of the ith pair of vertices in the first graph and the indicator random variable Y_i for adjacency of the ith pair of vertices in the second graph are each marginally distributed Bernoulli(p_i), and the Pearson correlation coefficient of X_i, Y_i is ϱ_i (assuming that p_i is not 0 or 1, in which case the value of ϱ_i is irrelevant). Other than these dependencies, the random variables $X_1, X_2, \ldots, X_N, Y_1, Y_2, \ldots, Y_N$ are independent.

It is not hard to see that that the choice of these parameters uniquely specifies the joint distribution of the two graphs (see Appendix A of [4]). Indeed, we can sample from the distribution in the following manner. For each $i = 1, 2, \ldots, N$ independently, sample $X_i \sim$ Bernoulli(p_i), then conditioned on the value x_i of X_i, sample $Y_i \sim$ Bernoulli$(\varrho_i x_i + (1 - \varrho_i)p_i)$. For all $i = 1, 2, \ldots, N$, define

$$\begin{aligned} q_{i,1} &:= p_i^2 + \varrho_i p_i(1 - p_i), \\ q_{i,0} &:= (1 - p_i)^2 + \varrho_i p_i(1 - p_i), \\ q_{i,\star} &:= (1 - \varrho_i)p_i(1 - p_i); \end{aligned}$$

these are, respectively, the probability that $X_i = Y_i = 1$, the probability that $X_i = Y_i = 0$, and the probability that $[X_i = 1$ and $Y_i = 0]$.

Let X and Y denote the random vectors whose ith components are respectively X_i and Y_i, for all $i = 1, 2, \ldots, N$; thus, in effect, X and Y are like the adjacency matrices representing the respective graphs. Let $\mathcal{X} := \{(x, y) : x, y \in \{0, 1\}^N\}$ denote the sample space for the correlated Bernoulli random graph model; in particular, x and y respectively are possible realizations of the adjacency vectors X and Y.

Note that if $\varrho_1 = \varrho_2 = \cdots = \varrho_N = 1$ then almost surely the two graphs are isomorphic, and if $\varrho_1 = \varrho_2 = \cdots = \varrho_N = 0$ then the two graphs are independent, meaning that the collection of random variables $X_1, X_2, \ldots, X_N, Y_1, Y_2, \ldots, Y_N$ is independent.

Define $\mathcal{R}^o := \{(p_1, p_2, \ldots, p_N, 0, 0, \ldots, 0) : p_1, p_2, \ldots, p_N \in \mathbb{R}\}$. The parameter space Θ will be called *nondegenerate* if $\Theta \cap \mathcal{R}^o$ has an interior point, relative to $\mathcal{R}^o$; i.e. there exists $z \in \Theta \cap \mathcal{R}^o$ and real number $\epsilon > 0$ such that $\Theta \cap \mathcal{R}^o$ contains all points in $\mathcal{R}^o$ that are less than ϵ distant from z. Nondegeneracy of Θ will play a critical role here; it is an assumption explicitly required for most of the theorems in this paper. Furthermore, when this condition is assumed, it is not merely for ease of exposition or analysis; indeed, we will demonstrate that absence of this condition, when it is assumed, can falsify the conclusions of the theorems that assume this condition.

Remark 1. The results in this paper provide machinery for improved estimation in the context of the correlated Bernoulli random graph model, which is a versatile and currently popular random graph model utilized heavily in the study of graph matching and similar disciplines. Nonetheless, the way the correlated Bernoulli random graph model is defined here—and the nature of the results in this paper and their proofs—render these results also expressible more broadly in terms of random correlated Bernoulli vectors X, Y, for the pair of random vectors X, Y defined above, without underlying graph structure. Thus, in particular, with this broader perspective, we do not need to restrict the number of components N (for X and Y) to be $\binom{n}{2}$, where n is the number of vertices in an underlying graph. Indeed, we can consider N as any positive integer.

2.1. Important statistics and functions of the parameters

The most important statistic in this paper, the *disagreement vector* statistic $\mathcal{H}$, is defined first. This statistic is foundational for the first group of our results; in Theorem 2 we will show, under the nondegeneracy condition, that $\mathcal{H}$ is complete and sufficient.

Definition 2. The (vector-valued) *disagreement vector* statistic $\mathcal{H} : \mathcal{X} \to \{0,\star,1\}^N$ is defined as follows: For all $(x,y) \in \mathcal{X}$, the vector $\mathcal{H}(x,y) \in \{0,\star,1\}^N$ is such that, for each $i = 1,2,\ldots,N$, the ith component of $\mathcal{H}(x,y)$ is equal to 1 if $x_i = y_i = 1$, is equal to 0 if $x_i = y_i = 0$, and is equal to $\star$ if $x_i \neq y_i$. For all $h \in \{0,\star,1\}^N$, the preimage of h, which is the set $\mathcal{H}^{-1}(h)$, is denoted as $\mathcal{X}_h$, and is called a *disagreement class*. Note that $\mathcal{X}$ is partitioned into the disjoint union $\mathcal{X} = \bigcup_{h\in\{0,\star,1\}^N} \mathcal{X}_h$.

The following definitions are key for the second group of our results.

The *Bernoulli parameter mean* μ and the *Bernoulli parameter variance* σ^2 are defined as

$$\mu := \frac{1}{N}\sum_{i=1}^{N} p_i, \qquad \sigma^2 := \frac{1}{N}\sum_{i=1}^{N}(p_i-\mu)^2.$$

The *empirical density of* X, denoted $\mathfrak{d}_X$, the *empirical density of* Y, denoted $\mathfrak{d}_Y$, and the *combined empirical density*, denoted $\mathfrak{d}_{X,Y}$, are statistics $\mathcal{X} \to \mathbb{R}$ that are respectively defined as

$$\mathfrak{d}_X := \frac{1}{N}\sum_{i=1}^{N} X_i, \qquad \mathfrak{d}_Y := \frac{1}{N}\sum_{i=1}^{N} Y_i, \qquad \mathfrak{d}_{X,Y} := \frac{1}{2}(\mathfrak{d}_X + \mathfrak{d}_Y).$$

Clearly, all three of these statistics are unbiased estimators of the parameter μ. Then, we define the statistics $\mathfrak{d}_{X\cap Y}, \mathfrak{d}_{X\cup Y} : \mathcal{X} \to \mathbb{R}$ as

$$\mathfrak{d}_{X\cap Y} := \frac{1}{N}\sum_{i=1}^{N} X_i Y_i, \qquad \mathfrak{d}_{X\cup Y} := \mathfrak{d}_X + \mathfrak{d}_Y - \mathfrak{d}_{X\cap Y}.$$

Note that for all $(x,y) \in \mathcal{X}$, we have that $\mathfrak{d}_{X\cap Y}(x,y) := \frac{|\{i\,:x_i=1 \text{ and } y_i=1\}|}{N}$, and we also have that $\mathfrak{d}_{X\cup Y}(x,y) := \frac{|\{i\,:x_i=1 \text{ or } y_i=1\}|}{N}$.

Next, the *disagreement enumeration statistic* $\Delta : \mathcal{X} \to \mathbb{R}$ is

$$\Delta := \sum_{i=1}^{N}(X_i - Y_i)^2;$$

in particular, for all $(x, y) \in \mathcal{X}$, $\Delta(x, y)$ is the number of components at which x and y disagree. Clearly, we have, for all $\theta \in \Theta$, $\mathbb{E}(\Delta) = 2\sum_{i=1}^{N}(1-\varrho_i)p_i(1-p_i)$. For all $h \in \{0, \star, 1\}^N$ and $(x, y) \in \mathcal{X}_h$, we have $2^{\Delta(x,y)} = |\mathcal{X}_h|$.

The *heterogeneity correlation* ϱ_H is a parameter defined by

$$\varrho_H := \frac{\sigma^2}{\mu(1-\mu)}.$$

In the case where μ is 0 or 1 then any convention may be adopted for defining ϱ_H (but it must be a value between 0 and 1). It is not hard to show that **a)** it holds that $0 \leq \varrho_H \leq 1$, and **b)** it holds that $\varrho_H = 1$ if and only if each of the p_i's are either 0 or 1, and **c)** it holds that $\varrho_H = 0$ if and only if all of the p_i's are equal to each other (of course, statements **b)** and **c)** aren't to be applied to the case where μ is 0 or 1).

Define the *total correlation* parameter ϱ_T as

$$\varrho_T := 1 - \frac{\sum_{i=1}^{N}(1-\varrho_i)p_i(1-p_i)}{N\mu(1-\mu)}.$$

In the case where μ is 0 or 1 then any convention may be adopted for defining ϱ_T (but it must be a value between 0 and 1). Note that in the case where all ϱ_i are equal, say to the value ϱ_E, then $(1-\varrho_T) = (1-\varrho_E)(1-\varrho_H)$. It is always the case that $0 \leq \varrho_T \leq 1$. In [4] it was empirically demonstrated—for the correlated graphs in broad families within our model—that graph matching complexity as well as graph matchability are each functions of total correlation, hence the importance of total correlation.

The *alignment strength* $\mathfrak{str} : \mathcal{X} \to \mathbb{R}$ is an important statistic defined as

$$\mathfrak{str} := 1 - \frac{\Delta/N}{\mathfrak{d}_X\,(1-\mathfrak{d}_Y) + (1-\mathfrak{d}_X)\,\mathfrak{d}_Y}. \tag{1}$$

In the case that x and y are both all zeros or both all ones then any convention may be adopted for defining $\mathfrak{str}$ (but it should be a value between 0 and 1). The definition of alignment strength $\mathfrak{str}$ arose in [4] in a natural way. Specifically, 1 minus the alignment strength is the ratio of disagreements between the two graphs —through the natural alignment— divided by the average number of disagreements over all vertex bijections between the two graphs; see there for more details. In [4] it is proven under mild conditions that $\mathfrak{str}$ is a strongly consistent estimator of ϱ_T.

An equivalent formula for alignment strength is

$$\mathfrak{str} = \frac{\mathfrak{d}_{X\cap Y} - \mathfrak{d}_X\mathfrak{d}_Y}{\mathfrak{d}_{X,Y} - \mathfrak{d}_X\mathfrak{d}_Y}; \tag{2}$$

it follows immediately using the easily-derived identities mentioned later in Equations (8) through (12).

3. The results

As mentioned in Section 1, our main results, which will be listed in this section, can be divided into two groups.

The first group of our results: In the context of correlated Bernoulli random graphs, we begin with Theorem 2, which asserts, under a nondegeneracy condition, that the disagreement vector statistic $\mathcal{H}$ is complete and sufficient; using this, given any estimator of a function of model parameters, we describe a way to refine ("balance") the estimator, reducing the mean squared error. Indeed, under the nondegeneracy condition, given any unbiased estimator of a function of model parameters, we characterize all unbiased estimators (Theorem 3) and the UMVU estimator (Theorem 4). The second group of our results: Theorems 7, 8, and 9 show that there are no unbiased estimators of various graph correlation measures, including total correlation ϱ_T; however, not only does balancing alignment strength improve alignment strength's mean squared error in estimating ϱ_T, but balancing numerator and denominator separately is seen empirically to be a further improvement.

Our first result is the following theorem.

Theorem 2. *If the parameter space Θ is nondegenerate, then the disagreement vector $\mathcal{H}$ is a complete and sufficient statistic.*

Theorem 2 is proved in Section 4.2.4.

Given any statistic $S : \mathcal{X} \to \mathbb{R}$, define the statistic $\overline{S} : \mathcal{X} \to \mathbb{R}$ as follows. For all $(x, y) \in \mathcal{X}$ and $h \in \{0, \star, 1\}^N$ such that $(x, y) \in \mathcal{X}_h$, define $\overline{S}(x, y) := \frac{1}{|\mathcal{X}_h|} \sum_{(x',y') \in \mathcal{X}_h} S(x', y')$; in particular, $\overline{S}$ is a constant function on $\mathcal{X}_h$. We say that $\overline{S}$ is the *balanced* variant of S; the *balancing* of S means the substituting of $\overline{S}$ in place of S when performing an estimation or inference task. If $S = \overline{S}$ then we say S is a *balanced* statistic. Of course, S is balanced if and only if S is a constant function on $\mathcal{X}_h$ for each $h \in \{0, \star, 1\}^N$. The following are our main results.

Theorem 3. *Suppose the parameter space Θ is nondegenerate, and a statistic $S : \mathcal{X} \to \mathbb{R}$ is an unbiased estimator of $g(\theta)$, where $g : \Theta \to \mathbb{R}$. Then a statistic $T : \mathcal{X} \to \mathbb{R}$ is an unbiased estimator of $g(\theta)$ if and only if $\overline{T} = \overline{S}$. In particular, $\overline{S}$ is an unbiased estimator of $g(\theta)$.*

Theorem 4. *Suppose the parameter space Θ is nondegenerate, and a statistic $S : \mathcal{X} \to \mathbb{R}$ is an unbiased estimator of $g(\theta)$, where $g : \Theta \to \mathbb{R}$. Then there exists a UMVU estimator of $g(\theta)$ and, in fact, $\overline{S}$ is the UMVU estimator of $g(\theta)$.*

We prove Theorem 3 in Sections 4.1 and 4.2, and we prove Theorem 4 in Section 4.3; it is essentially a consequence of Theorem 2. The key idea in proving Theorem 4 is that $\overline{S}$ is the composition of some function with $\mathcal{H}$ —a complete and sufficient statistic by Theorem 2—and thus the Lehmann-Scheffe Theorem dictates that $\overline{S}$ is UMVU. Indeed, $\overline{S}$ is the Rao-Blackwellization of S conditioning on $\mathcal{H}$. Section 4.3 spells out the details.

(An excellent reference for Rao-Blackwell theory are the original papers [12, 1], and an excellent reference for the Lehmann-Scheffe Theorem are the original papers [8, 9].)

Remark 5. Fundamentally, since balancing is Rao-Blackwellization, it is a regularization technique that reduces an estimator's variance without changing its mean. Indeed, suppose the parameter space Θ is nondegenerate, and a statistic $S : \mathcal{X} \to \mathbb{R}$ is any estimator, whether biased or unbiased, of $g(\theta)$. Then $\overline{S}$ has the minimum possible variance among all estimators with the same expected value as S. (This, in turn, implies that $\overline{S}$ has the minimum mean squared error among all such estimators.) To see this, define $\ell(\theta) := \mathbb{E}(S)$ for all $\theta \in \Theta$, and then consider S, $\overline{S}$, and ℓ in Theorems 3 and 4.

Example 1. The disagreement enumeration statistic $\Delta : \mathcal{X} \to \mathbb{R}$ is clearly a balanced statistic, since it is a constant function on each $\mathcal{X}_h$. Hence, by Theorem 4, when Θ is nondegenerate, Δ is the UMVU estimator of its expected value $\mathbb{E}(\Delta) = 2\sum_{i=1}^{N}(1-\varrho_i)p_i(1-p_i)$.

Example 2. When $N > 1$, the statistic $\mathfrak{d}_X(1-\mathfrak{d}_Y) + (1-\mathfrak{d}_X)\mathfrak{d}_Y$ is **NOT** a balanced statistic; indeed, consider $h = [\star, \star, \dots, \star]^T$, and consider $(x', y') \in \mathcal{X}_h$ such that x' is all zeros and y' is all ones, and consider $(x'', y'') \in \mathcal{X}_h$ such that the first $\lfloor \frac{N}{2} \rfloor$ entries of x'' are all zeros and of y'' are all ones, and the last $\lceil \frac{N}{2} \rceil$ entries of x'' are all ones and of y'' are all zeros—the statistic $\mathfrak{d}_X(1-\mathfrak{d}_Y)+(1-\mathfrak{d}_X)\mathfrak{d}_Y$ at (x', y') has the value 1, and at (x'', y'') has a value approaching $\frac{1}{2}$, hence the statistic is not constant on $\mathcal{X}_h$, hence is not balanced.

Example 3. When $N > 1$, the alignment strength statistic $\mathfrak{str}$ is **NOT** a balanced statistic. This is because in Example 1 we have that the numerator of $\mathfrak{str}$ in Equation (1) is balanced, and in Example 2 we have that the denominator of $\mathfrak{str}$ in Equation (1) is not balanced; hence $\mathfrak{str}$ is not a constant function on all $\mathcal{X}_h$, and is thus not balanced.

It is important to note that the claims in Theorems 3 and 4 may fail without the assumption of nondegeneracy for the parameter space Θ, as highlighted in the next example. (In particular, this points to the non-triviality of Theorems 3 and 4.)

Example 4. If the value of μ is known, hence fixed, then Θ is contained in a particular hyperplane intersecting $\mathcal{R}^o$, and Θ is degenerate; in this scenario, we will show in Section 4.8 that unbiasedness of estimators is not characterized as described in Theorem 3 and, often, there do not exist UMVU estimators for functions of the model parameters, even when there exist unbiased estimators.

The following corollary is an immediate consequence of Theorem 4 and the fact that sums and products of constant functions (on respective $\mathcal{X}_h$) are constant (on respective $\mathcal{X}_h$).

Corollary 6. *Suppose the parameter space Θ is nondegenerate, and a statistic $S : \mathcal{X} \to \mathbb{R}$ is an unbiased estimator of $g(\theta)$, where $g : \Theta \to \mathbb{R}$, and a statistic $S' : \mathcal{X} \to \mathbb{R}$ is an unbiased estimator of $g'(\theta)$, where $g' : \Theta \to \mathbb{R}$. Then, for any*

$a, b \in \mathbb{R}$, $a\overline{S} + b\overline{S}'$ is UMVUE for $ag(\theta) + bg'(\theta)$. Indeed, $a\overline{S} + b\overline{S}'$ is balanced, $\overline{S} \cdot \overline{S}'$ is balanced, and (if $\overline{S}'$ is nonzero) $\overline{S}/\overline{S}'$ is balanced.

The next set of theorems are applications of the above theorems—and the methodologies of their proofs—to unbiasedness and efficiency of statistics for estimating various graph correlation parameters, particularly total correlation.

Theorem 7. *Suppose the parameter space Θ is nondegenerate, and $N > 1$. There does not exist an unbiased estimator of the heterogeneity correlation ϱ_H.*

Theorem 8. *Suppose the parameter space Θ is nondegenerate, and $N > 1$. There does not exist an unbiased estimator of the total correlation ϱ_T.*

We prove Theorems 7 and 8 in Section 4.5.

In the following negative result on estimating edge correlation, besides the assumption of a nondegenerate parameter space, we have additional assumptions that all pairs of vertices share the same edge correlation parameter (ie. Θ is restricted so that, for all $\theta \in \Theta$, it holds that $\varrho_1 = \varrho_2 \cdots = \varrho_N$), and we also assume that this edge correlation parameter is not always zero. Specifically:

Theorem 9. *Suppose that the following three conditions hold:*

- ***a)*** *The parameter space Θ is nondegenerate.*
- ***b)*** *The parameter space Θ is such that edge correlations are component-uniform, meaning that there exists a function $\varrho_E : \Theta \to \mathbb{R}$ such that, for all $\theta \in \Theta$ and all $i = 1, 2, \ldots, N$, $\varrho_i(\theta) = \varrho_E(\theta)$.*
- ***c)*** *The parameter space Θ is not a subset of $\mathcal{R}^o$, i.e. there exists $\theta \in \Theta$ such that $\varrho_E(\theta) \neq 0$.*

Then there does not exist an unbiased estimator of ϱ_E.

We prove Theorem 9 in Section 4.4.

Remark 10. Suppose the parameter space Θ is nondegenerate. As mentioned in Example 3, alignment strength $\mathfrak{str} = \frac{\mathfrak{d}_{X\cap Y} - \mathfrak{d}_X \mathfrak{d}_Y}{\mathfrak{d}_{X,Y} - \mathfrak{d}_X \mathfrak{d}_Y}$ is **NOT** balanced when $N > 1$. (So, the bias in estimating ϱ_T is the same for $\mathfrak{str}$ as for $\overline{\mathfrak{str}}$, but the variance of $\overline{\mathfrak{str}}$ is less than the variance of $\mathfrak{str}$.) Next, define the modified alignment strength $\mathfrak{str}' := \frac{\overline{\mathfrak{d}_{X\cap Y} - \mathfrak{d}_X \mathfrak{d}_Y}}{\mathfrak{d}_{X,Y} - \mathfrak{d}_X \mathfrak{d}_Y}$; by Corollary 6, $\mathfrak{str}'$ is balanced. We will empirically show in Section 5 that $\mathfrak{str}'$ is often superior to $\overline{\mathfrak{str}}$ as an estimator of ϱ_T. Also, in Section 4.6 we will prove the following clean formulas:

$$\mathfrak{str}' = \frac{\mathfrak{d}_{X\cap Y} - \mathfrak{d}_X \mathfrak{d}_Y + \frac{1}{4}\left[\frac{\Delta}{N^2} - (\mathfrak{d}_X - \mathfrak{d}_Y)^2\right]}{\mathfrak{d}_{X,Y} - \mathfrak{d}_X \mathfrak{d}_Y + \frac{1}{4}\left[\frac{\Delta}{N^2} - (\mathfrak{d}_X - \mathfrak{d}_Y)^2\right]} = \frac{\mathfrak{d}_{X\cap Y} - \mathfrak{d}^2_{X,Y} + \frac{\Delta}{4N^2}}{\mathfrak{d}_{X,Y}(1 - \mathfrak{d}_{X,Y}) + \frac{\Delta}{4N^2}}. \tag{3}$$

When x and y are both all zeros or all ones then any convention for defining $\mathfrak{str}'$ is acceptable, provided that it is between 0 and 1. Note that $\mathfrak{str}'$ can have less variance than $\overline{\mathfrak{str}}$ (and in general it does) without violating Theorem 4, since the expected values of $\mathfrak{str}'$ and $\overline{\mathfrak{str}}$ can be different.

Sometimes balancing a statistic—even at one sample space point—requires averaging an exponential number of values. Remark 10 is notable for its simple expressions for the balanced statistics comprising $\mathfrak{str}'$, and in Section 5 a linear time algorithm is given for computing $\overline{\mathfrak{str}}$.

The following result is proved in Section 4.7; it follows from Corollary 6 and Remark 10.

Corollary 11. *Suppose the parameter space Θ is nondegenerate and also suppose that $\Theta \subseteq \mathcal{R}^o$. Then the statistic $\mathfrak{d}_{X,Y}(1-\mathfrak{d}_{X,Y}) - \frac{1}{2N}\left(1-\frac{1}{2N}\right)\Delta$ is UMVUE for σ^2.*

These are the results in this paper, and they will be proven next in Section 4.

4. Proof of the results in Section 3

We begin by proving Theorem 3 and Theorem 4; the proofs of Theorem 2 and Theorem 9 will be built on the methodology of the forward direction of the proof of Theorem 3. The rest of the results in Section 3 will also be shown.

4.1. Proof of the reverse direction of Theorem 3

The reverse direction of Theorem 3 can be equivalently formulated in the following way. Suppose the parameter space Θ is nondegenerate, and a statistic $S : \mathcal{X} \to \mathbb{R}$ is an unbiased estimator of $g(\theta)$, where $g : \Theta \to \mathbb{R}$. If the statistic $T : \mathcal{X} \to \mathbb{R}$ satisfies the condition that for all $h \in \{0, \star, 1\}^N$ $\quad \sum_{(x,y)\in\mathcal{X}_h} T(x,y) = \sum_{(x,y)\in\mathcal{X}_h} S(x,y)$ then T is an unbiased estimator of $g(\theta)$.

Proving the reverse direction of Theorem 3 is quite straightforward. For each $h \in \{0, \star, 1\}^N$, the elements of $\mathcal{X}_h$ are equiprobable. In particular, for all $\theta \in \Theta$ it holds that

$$
\begin{aligned}
\mathbb{E}(T) &= \sum_{(x,y)\in\mathcal{X}} \mathbb{P}(x,y)T(x,y) = \sum_{h\in\{0,\star,1\}^N} \sum_{(x,y)\in\mathcal{X}_h} \mathbb{P}(x,y)T(x,y) \\
&= \sum_{h\in\{0,\star,1\}^N} \left[\left(\prod_{i:h_i=1} q_{i,1} \right) \left(\prod_{i:h_i=0} q_{i,0} \right) \left(\prod_{i:h_i=\star} q_{i,\star} \right) \sum_{(x,y)\in\mathcal{X}_h} T(x,y) \right] \\
&= \sum_{h\in\{0,\star,1\}^N} \left[\left(\prod_{i:h_i=1} q_{i,1} \right) \left(\prod_{i:h_i=0} q_{i,0} \right) \left(\prod_{i:h_i=\star} q_{i,\star} \right) \sum_{(x,y)\in\mathcal{X}_h} S(x,y) \right] \\
&= \mathbb{E}(S) = g(\theta).
\end{aligned}
$$

Thus T is an unbiased estimator of $g(\theta)$. □

4.2. Proof of the forward direction of Theorem 3 and of Theorem 2

The proof of the forward direction of Theorem 3 involves notation that is complex at first glance, and the core ideas may be challenging to follow when presented all at once in full generality. Our expositional strategy is as follows. After proving the basic preliminary Lemma 12 in Section 4.2.1, we proceed to first prove the forward direction of Theorem 3 in the special cases where $N = 1, 2$ in Section 4.2.2, so that the notation, reasoning, and strategy are crystal clear, and then in Section 4.2.3 we prove the forward direction of Theorem 3 in full and clear generality. Then we will show in Section 4.2.4 that the machinery of Section 4.2.3 proves Theorem 2 (Completeness of $\mathcal{H}$).

4.2.1. Preliminaries

We begin with a technical lemma, Lemma 12. Two polynomials in a single variable that are equal as functions at infinitely many points are, by interpolation theory, equal algebraically (meaning that the two polynomials have the same coefficients as each other). However, for two polynomials in more than one variable, this may fail. For example, consider the polynomial $p_1^2 - p_2$ and the zero polynomial, in the two variables p_1 and p_2. These two polynomials agree as functions on a parabola, but they are not equal algebraically. However, if two polynomials of any degree agree as functions on an open neighborhood then they are equal algebraically. Formally:

Lemma 12. *Suppose that Θ is nondegenerate, and $g, \tilde{g} : \Theta \cap \mathcal{R}^o \to \mathbb{R}$ are two polynomials in the variables $p_1, p_2, \ldots, p_N$ such that for all $\theta \in \Theta \cap \mathcal{R}^o$ it holds that $g(\theta) = \tilde{g}(\theta)$. Then the coefficients of the polynomial g are identical to the respective coefficients of the polynomial $\tilde{g}$.*

The proof of Lemma 12 is a straightforward induction on the maximum degree of the polynomials g and $\tilde{g}$, considering sequential partial derivatives. An equivalent formulation of Lemma 12 can be found in the classical textbook *Algebra* of Serge Lang [7], Chapter IV, Corollary 1.6.

4.2.2. Proof of the forward direction of Theorem 3 for the particular cases where $N = 1, 2$, by way of illustration for the general case

The forward direction of Theorem 3 can be formulated as follows. Suppose the parameter space Θ is nondegenerate, and the two statistics $S, T : \mathcal{X} \to \mathbb{R}$ are each unbiased estimators of $g(\theta)$, where $g : \Theta \to \mathbb{R}$. Then for all $h \in \{0, \star, 1\}^N$ it holds that $\sum_{(x,y)\in\mathcal{X}_h} T(x, y) = \sum_{(x,y)\in\mathcal{X}_h} S(x, y)$.

To best illustrate, we begin with a proof for the case where $N = 1$. Taking the expectation for parameters $\theta \in \Theta \cap \mathcal{R}^o$, we see that $\mathbb{E}(S) = (1-p_1)^2 S(0,0) + p_1(1-p_1)\big(S(0,1) + S(1,0)\big) + p_1^2 S(1,1)$. In particular, g needs to be a quadratic polynomial in the single variable p_1 on $\Theta \cap \mathcal{R}^o$, say $g(p_1) := g^{(0)} p_1^0 + g^{(1)} p_1^1 + g^{(2)} p_1^2$ where $g^{(0)}$, $g^{(1)}$, and $g^{(2)}$ are fixed coefficients. By the nondegeneracy of Θ and Lemma 12, we can uniquely represent polynomials as vectors with respective entries being the coefficients of p_1^0, p_1^1, and p_1^2, respectively. Thus $(1 - p_1)^2$ is represented as $\left[\begin{smallmatrix} 1 \\ -2 \\ 1 \end{smallmatrix}\right]$, and $p_1(1 - p_1)$ is represented as $\left[\begin{smallmatrix} 0 \\ 1 \\ -1 \end{smallmatrix}\right]$, and p_1^2

is represented as $\left[\begin{smallmatrix}0\\0\\1\end{smallmatrix}\right]$, and g is represented as $\left[\begin{smallmatrix}g^{(0)}\\g^{(1)}\\g^{(2)}\end{smallmatrix}\right]$. In particular, S being an unbiased estimator for g on $\Theta \cap \mathcal{R}^o$ means that

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} S(0,0) \\ S(0,1)+S(1,0) \\ S(1,1) \end{bmatrix} = \begin{bmatrix} g^{(0)} \\ g^{(1)} \\ g^{(2)} \end{bmatrix}$$

Denote the left hand side matrix as A, that is $A = \left[\begin{smallmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -1 & 1 \end{smallmatrix}\right]$; since A is invertible, and T has to satisfy the above equation as well, we therefore have that $\left[\begin{smallmatrix} S(0,0) \\ S(0,1)+S(1,0) \\ S(1,1) \end{smallmatrix}\right] = \left[\begin{smallmatrix} T(0,0) \\ T(0,1)+T(1,0) \\ T(1,1) \end{smallmatrix}\right]$, which precisely says that for all $h \in \{0, \star, 1\}^N$ it holds that $\sum_{(x,y)\in\mathcal{X}_h} T(x,y) = \sum_{(x,y)\in\mathcal{X}_h} S(x,y)$, and the case where $N=1$ is proven.

By further way of illustration, we next prove the case where $N=2$. Taking the expectation for parameters $\theta \in \Theta \cap \mathcal{R}^o$, we have

$$\begin{aligned}
\mathbb{E}(S) = \quad & (1-p_1)^2(1-p_2)^2 \left[S(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right],\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right])\right] \\
& + \quad (1-p_1)^2 p_2(1-p_2)\Big[S(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right],\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]) + S(\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right],\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right])\Big] \\
& + \quad p_1(1-p_1)p_2(1-p_2)\Big[S(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right],\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]) + S(\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right],\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]) \\
& \qquad + S(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right],\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]) + S(\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right],\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right])\Big] \quad + \quad \cdots \quad + \quad \cdots \\
= & (1-p_1)^2(1-p_2)^2 \sum_{(x,y)\in\mathcal{X}_{\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]}} S(x,y) \; + \; (1-p_1)^2 p_2(1-p_2) \sum_{(x,y)\in\mathcal{X}_{\left[\begin{smallmatrix}0\\\star\end{smallmatrix}\right]}} S(x,y) \\
& + \quad (1-p_1)^2 p_2^2 \sum_{(x,y)\in\mathcal{X}_{\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]}} S(x,y) \quad + \quad p_1(1-p_1)(1-p_2)^2 \sum_{(x,y)\in\mathcal{X}_{\left[\begin{smallmatrix}\star\\0\end{smallmatrix}\right]}} S(x,y) \\
& + \quad p_1(1-p_1)p_2(1-p_2) \sum_{(x,y)\in\mathcal{X}_{\left[\begin{smallmatrix}\star\\\star\end{smallmatrix}\right]}} S(x,y) \; + \; p_1(1-p_1)p_2^2 \sum_{(x,y)\in\mathcal{X}_{\left[\begin{smallmatrix}\star\\1\end{smallmatrix}\right]}} S(x,y) \\
& + \quad p_1^2(1-p_2)^2 \sum_{(x,y)\in\mathcal{X}_{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]}} S(x,y) \quad + \quad p_1^2 p_2(1-p_2) \sum_{(x,y)\in\mathcal{X}_{\left[\begin{smallmatrix}1\\\star\end{smallmatrix}\right]}} S(x,y) \\
& + \quad p_1^2 p_2^2 \sum_{(x,y)\in\mathcal{X}_{\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]}} S(x,y).
\end{aligned}$$

Note in particular that g would have to be a polynomial in the two variables p_1, p_2, with its monomials each consisting of a constant, denoted $g^{(k_1,k_2)}$, times $p_1^{k_1} p_2^{k_2}$, where $k_1, k_2 \in \{0,1,2\}$. By the nondegeneracy of Θ and Lemma 12, g can be uniquely represented by the vector of coefficients ordered lexicographically (i.e. dictionary order) by superscript:
$[g^{(0,0)}, g^{(0,1)}, g^{(0,2)}, g^{(1,0)}, g^{(1,1)}, g^{(1,2)}, g^{(2,0)}, g^{(2,1)}, g^{(2,2)}]^T$. Indeed, all other polynomials with monomials each consisting of a constant times $p_1^{k_1} p_2^{k_2}$, where $k_1, k_2 \in \{0,1,2\}$, will also be similarly represented (uniquely) by the vector

of coefficients ordered lexicographically by superscript. For example, in the matrix on the left hand side below, the columns respectively are the vectors (of lexicographically ordered monomial coefficients) representing the respective polynomials $(1-p_1)^2(1-p_2)^2$, $(1-p_1)^2p_2(1-p_2)$, $(1-p_1)^2p_2^2$, ..., which are the respective probabilities of $(x,y)\in\mathcal{X}_h$ for h's lexicographically ordered ("$\star$" has the value $\frac{1}{2}$) as: $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}0\\\star\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}\star\\0\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}\star\\\star\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}\star\\1\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}1\\\star\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$. Now, S being an unbiased estimator of g on $\Theta\cap\mathcal{R}^o$ means precisely that the following linear system holds:

$$\left[\begin{array}{rrr|rrr|rrr} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ \hline -2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 4 & -2 & 0 & -2 & 1 & 0 & 0 & 0 & 0\\ -2 & 2 & -2 & 1 & -1 & 1 & 0 & 0 & 0\\ \hline 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0\\ -2 & 1 & 0 & 2 & -1 & 0 & -2 & 1 & 0\\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{array}\right] \left[\begin{array}{c} \sum_{(x,y)\in\mathcal{X}_{[0,0]^T}} S(x,y)\\ \sum_{(x,y)\in\mathcal{X}_{[0,\star]^T}} S(x,y)\\ \sum_{(x,y)\in\mathcal{X}_{[0,1]^T}} S(x,y)\\ \hline \sum_{(x,y)\in\mathcal{X}_{[\star,0]^T}} S(x,y)\\ \sum_{(x,y)\in\mathcal{X}_{[\star,\star]^T}} S(x,y)\\ \sum_{(x,y)\in\mathcal{X}_{[\star,1]^T}} S(x,y)\\ \hline \sum_{(x,y)\in\mathcal{X}_{[1,0]^T}} S(x,y)\\ \sum_{(x,y)\in\mathcal{X}_{[1,\star]^T}} S(x,y)\\ \sum_{(x,y)\in\mathcal{X}_{[1,1]^T}} S(x,y) \end{array}\right]$$

$$= \left[\begin{array}{c} g^{(0,0)}\\ g^{(0,1)}\\ g^{(0,2)}\\ \hline g^{(1,0)}\\ g^{(1,1)}\\ g^{(1,2)}\\ \hline g^{(2,0)}\\ g^{(2,1)}\\ g^{(2,2)} \end{array}\right] \tag{4}$$

Observe that the left-hand-side matrix above is the Kronecker product $A\otimes A$, where A is the lower triangular matrix with diagonals all ones mentioned in the proof of the case where $N=1$. Note that $A\otimes A$ is thus lower triangular with diagonals all ones, thus has nonzero determinant and is invertible. Since T solves the same linear system (above, Equation (4)) as S does, we conclude —by multiplying both sides of the equation above by the inverse of $A\otimes A$— that for all $h\in\{0,\star,1\}^N$ it holds that $\sum_{(x,y)\in\mathcal{X}_h} T(x,y)=\sum_{(x,y)\in\mathcal{X}_h} S(x,y)$, and the case where $N=2$ is now also proven.

4.2.3. Proof of the forward direction of Theorem 3, the general case

With the proofs of the cases $N=1,2$ as illustration, we now prove the result for arbitrary N. Let $\otimes^N A$ denote the N-fold Kronecker product $A\otimes A\otimes\cdots\otimes A$. Next, let $\overrightarrow{S/\mathcal{H}}$ denote the vector whose components are respectively

$\sum_{(x,y)\in\mathcal{X}_h} S(x,y)$ for each of the $h \in \{0,\star,1\}^N$, ordered lexicographically according to h. Restricting to parameters $\theta \in \Theta \cap \mathcal{R}^o$, g is a polynomial in the variables $p_1, p_2, \ldots, p_N$, with its monomials each consisting of a constant, denoted $g^{(k_1,k_2,\ldots,k_N)}$, times $p_1^{k_1} p_2^{k_2} \cdots p_N^{k_N}$, where $k_1, k_2, \ldots, k_N \in \{0,1,2\}$. By the nondegeneracy of Θ and Lemma 12, we have that g can be uniquely represented by the column vector of monomial coefficients ordered lexicographically by the powers of $p_1, p_2, \ldots, p_N$; denote this vector $\vec{g}$.

We claim that S being an unbiased estimator for $g(\theta)$ on $\Theta \cap \mathcal{R}^o$ means precisely that S satisfies the linear system

$$[\otimes^N A] \ \cdot \ \overrightarrow{S/\mathcal{H}} \ = \ \vec{g}. \tag{5}$$

This can be verified directly by noting that for each $h \in \{0,\star,1\}^N$ and $(x,y) \in \mathcal{X}_h$, the probability of (x,y) is given by (and is simplified with elementary algebra)

$$\begin{aligned} &\prod_{i=1}^{N} \left\{ \begin{array}{cc} (1-p_i)^2 & \text{if } h_i = 0 \\ p_i(1-p_i) & \text{if } h_i = \star \\ p_i^2 & \text{if } h_i = 1 \end{array} \right\} \\ &\quad = \sum_{(k_1,k_2,\ldots,k_N)\in\{0,1,2\}^N} \left(\prod_{j=1}^{N} A_{k_j+1\,,\,2\cdot h_j+1} \right) \cdot \left(\prod_{j=1}^{N} p_j^{k_j} \right) \end{aligned} \tag{6}$$

where, in the subscript of $A_{k_j+1\,,\,2\cdot h_j+1}$, "$\star$" has the value $\frac{1}{2}$, meaning that when h_j is $\star$ then $2\cdot$"$\star$"$+1$ is defined to be 2. With nondegeneracy of Θ and Lemma 12, we have uniqueness of polynomial coefficients, and Equation (6) directly yields Equation (5).

Since $\otimes^N A$ is a lower triangular matrix with all diagonals being ones, it is an invertible matrix. Now, T has to satisfy Equation (5) as well; multiplying both sides of the equation by the inverse of $\otimes^N A$ yields that $\overrightarrow{S/\mathcal{H}} = \overrightarrow{T/\mathcal{H}}$, which precisely says that for all $h \in \{0,\star,1\}^N$ it holds that $\sum_{(x,y)\in\mathcal{X}_h} T(x,y) = \sum_{(x,y)\in\mathcal{X}_h} S(x,y)$, and the forward direction of Theorem 3 is now proved. $\square$

4.2.4. Proof of Theorem 2

Theorem 2 states that if the parameter space Θ is nondegenerate, then $\mathcal{H}$ is a complete and sufficient statistic. Sufficiency of $\mathcal{H}$ is immediate, since, conditioning on $\mathcal{H}$ being any $h \in \{0,\star,1\}^N$, we have that (X,Y) is distributed discrete uniform on $\mathcal{X}_h$ (its support). Also note that the probability of any sample point is a function of $\mathcal{H}$ and the model parameters, which implies that $\mathcal{H}$ is sufficient.

We now use the machinery of the previous Section 4.2.3 to prove the rest of Theorem 2; that if the parameter space Θ is nondegenerate, then $\mathcal{H}$ is a complete statistic. Completeness of $\mathcal{H}$ here means that, for any $f : \{0,\star,1\}^N \to \mathbb{R}$, if $\mathbb{E}(f(\mathcal{H})) = 0$ for all $\theta \in \Theta$ then f is the zero function.

For any $f : \{0,\star,1\}^N \to \mathbb{R}$, let $\vec{\mathrm{f}}$ denote a column vector whose respective entries are $|\mathcal{X}_h| \cdot f(h)$ for the respective $h \in \{0,\star,1\}^N$ ordered lexicographically.

(I.e., the first component of $\vec{\mathrm{f}}$ is $|\mathcal{X}_h| \cdot f(h)$ for the h that is all zeros, the second component of $\vec{\mathrm{f}}$ is $|\mathcal{X}_h| \cdot f(h)$ for the h that is all zeros except last entry is $\star$, etc.) By the nondegeneracy of Θ and Equation (5), we have that $\mathbb{E}(f(\mathcal{H})) = 0$ for all $\theta \in \Theta \cap \mathcal{R}^o$ would mean that $[\otimes^N A] \cdot \vec{\mathrm{f}} = \vec{0}$; by the invertibility of $[\otimes^N A]$, we would have that $\vec{\mathrm{f}} = \vec{0}$, hence f is the zero function. This proves the completeness of $\mathcal{H}$. ☐

Note that proof of the forward direction of Theorem 3 is based on the injectivity of $[\otimes^N A]$ as a function, and proof of the completeness of $\mathcal{H}$ in Theorem 2 is based on $[\otimes^N A]$ having a trivial nullspace, so these two results are equivalent.

4.3. Proof of Theorem 4

Theorem 4 states that if the parameter space Θ is nondegenerate, and a statistic $S : \mathcal{X} \to \mathbb{R}$ is an unbiased estimator of $g(\theta)$, for $g : \Theta \to \mathbb{R}$, then there exists a UMVU estimator of $g(\theta)$ and, in fact, the balanced statistic $\overline{S}$ is the UMVU estimator of $g(\theta)$.

We prove this as follows, under the assumption that Θ is nondegenerate. Recall first that $\overline{S}$ is an unbiased estimator of $g(\theta)$ by Theorem 3. Now, note that $\overline{S}$ is a constant function on $\mathcal{X}_h$ for each $h \in \{0, \star, 1\}^N$; thus there exists a function $\Phi : \{0, \star, 1\}^N \to \mathbb{R}$ such that $\overline{S}$ is the function composition $\Phi \circ \mathcal{H}$. Since $\mathcal{H}$ is a complete and sufficient statistic by Theorem 2, the Lehmann-Scheffe Theorem asserts that the composition $\overline{S} = \Phi \circ \mathcal{H}$ is UMVUE for $g(\theta)$. ☐

It is structurally interesting to note that $\overline{S}$ is the Rao-Blackwellization of S when conditioning on the complete and sufficient statistic $\mathcal{H}$. Indeed, for any $h \in \{0, \star, 1\}^N$, the Rao Blackwellization $\mathbb{E}(S|\mathcal{H} = h)$ is precisely the mean of the values of S on $\mathcal{X}_h$, which is precisely the statistic $\overline{S}$.

4.3.1. Another proof of Theorem 4, by first-principles

In this section, we mention a "first-principles" proof of Theorem 4, besides the Lehmann-Scheffe proof methodology in the previous Section 4.3.

For any statistic $T : \mathcal{X} \to \mathbb{R}$ which is an unbiased estimator of $g(\theta)$, we compute the variance of T, for any particular $\theta \in \Theta$, as:

$$\begin{aligned}
&\mathrm{Var}(T) \\
&= \sum_{(x,y)\in\mathcal{X}} \mathbb{P}(x,y)\big(T(x,y) - g(\theta)\big)^2 \\
&= \sum_{h\in\{0,\star,1\}^N} \sum_{(x,y)\in\mathcal{X}_h} \mathbb{P}(x,y)\big(T(x,y) - g(\theta)\big)^2 \\
&= \sum_{h\in\{0,\star,1\}^N} \left[\left(\prod_{i:h_i=1} q_{i,1}\right)\left(\prod_{i:h_i=0} q_{i,0}\right)\left(\prod_{i:h_i=\star} q_{i,\star}\right) \sum_{(x,y)\in\mathcal{X}_h} \big(T(x,y) - g(\theta)\big)^2\right].
\end{aligned}$$

In particular, $\mathrm{Var}(T)$ can be minimized over such unbiased T by, for all $h \in \{0,\star,1\}^N$, minimizing $\sum_{(x,y)\in\mathcal{X}_h}\big(T(x,y)-g(\theta)\big)^2$ subject to the constraint that $\sum_{(x,y)\in\mathcal{X}_h} T(x,y) = \sum_{(x,y)\in\mathcal{X}_h} S(x,y)$; this is because of Theorem 3. Treating the $T(x,y)$ as variables, this convex optimization problem has a global minimizer, when the objective gradient is equivalued (by the KKT conditions) hence the minimum variance is achieved when $T = \overline{S}$, independent of $\theta \in \Theta$, and Theorem 4 is shown. $\square$

4.4. Proof of Theorem 9

Theorem 9 states that if the following three conditions hold:

- **a)** The parameter space Θ is nondegenerate.
- **b)** The parameter space Θ is such that the edge correlations are component-uniform, meaning that there exist a function $\varrho_E : \Theta \to \mathbb{R}$ such that, for all $\theta \in \Theta$ and all $i = 1, 2, \ldots, N$, $\varrho_i(\theta) = \varrho_E(\theta)$.
- **c)** The parameter space Θ is not a subset of $\mathcal{R}^o$, i.e. there exists $\theta \in \Theta$ such that $\varrho_E(\theta) \neq 0$.

Then there does not exist an unbiased estimator of $\varrho_E(\theta)$.

Suppose, by way of contradiction, that statistic $S : \mathcal{X} \to \mathbb{R}$ is an unbiased estimator of $\varrho_E(\theta)$. For $\theta \in \Theta \cap \mathcal{R}^o$, where $\varrho_E \equiv 0$, we have by Equation (5) that $[\otimes^N A] \cdot \overrightarrow{S/\mathcal{H}} = \vec{0}$, since Θ is nondegenerate. By the invertibility of $\otimes^N A$ we thus have that $\overrightarrow{S/\mathcal{H}} = \vec{0}$, which, by the reasoning in Section 4.1, implies that $\mathbb{E}(S) = 0$ for all $\theta \in \Theta$, which is a contradiction because there exists $\theta \in \Theta$ where $\mathbb{E}(S) = \varrho_E(\theta) \neq 0$. $\square$

4.5. Proof of Theorems 7 and 8

Theorems 7 and 8 state that if the parameter space Θ is nondegenerate and $N > 1$ then there does not exist an unbiased estimator of the heterogeneity correlation ϱ_H nor of the total correlation ϱ_T.

We will just focus on $\Theta \cap \mathcal{R}^o$; on this set it is easy to see that $\varrho_T = \varrho_H$. Thus, by the development in Section 4.2 and the nondegeneracy of Θ, we will have proved Theorems 7 and 8 if we show that, on $\Theta \cap \mathcal{R}^o$, $\varrho_H := \frac{\sigma^2}{\mu(1-\mu)}$ is not a polynomial in the variables $p_1, p_2, \ldots, p_N$. By way of contradiction, suppose that, on $\Theta \cap \mathcal{R}^o$, ϱ_H is a polynomial in the variables $p_1, p_2, \ldots, p_N$.

Let $(\tilde{p_1}, \tilde{p_2}, \ldots, \tilde{p_N}, 0, 0, \ldots, 0)$ be an interior point of $\Theta \cap \mathcal{R}^o$, relative to $\mathcal{R}^o$; such a point exists by the nondegeneracy of Θ. Consider fixing the values of p_i to be $\tilde{p_i}$ for each $i = 2, 3, \ldots, N$, and varying only p_1. This results in ϱ_H, σ^2, and μ being polynomials in a single variable. Denote this variable by t instead of p_1 for ease of notation, and these respective polynomials are thus denoted $\varrho_H(t)$, $\sigma^2(t)$, and $\mu(t)$. Let $\mathcal{I}$ be a real, open interval containing $\tilde{p_1}$, such that for all $t \in \mathcal{I}$ we have $(t, \tilde{p_2}, \ldots, \tilde{p_N}, 0, 0, \ldots, 0) \in \Theta \cap \mathcal{R}^o$; a nontrivial such $\mathcal{I}$ exists by the nondegeneracy of Θ.

Using basic algebra, $\sigma^2(t)$ is quadratic in t, and the coefficient of t^2 in $\sigma^2(t)$ is $\left(\frac{1}{N} - \frac{1}{N^2}\right)$, and $\mu(t)(1-\mu(t))$ is quadratic in t, and the coefficient of t^2 in $\mu(t)(1-\mu(t))$ is $-\frac{1}{N^2}$. Now, by definition, $\sigma^2(t) = \mu(t)(1-\mu(t))\varrho_H(t)$, and the coefficients of the respective powers of t on the left hand side are respectively equal to the coefficients of the powers of t on the right hand side, since $\mathcal{I}$ is an interval (and invoking polynomial interpolation theory). This implies that polynomial $\varrho_H(t)$ can't have positive degree, and thus is constant. However, this constant is nonnegative (indeed, it has been pointed out in Section 2.1 that $0 \le \varrho_H \le 1$), which means that the coefficient of t^2 in (the left hand side) $\sigma^2(t)$ is positive, but the coefficient of t^2 in (the right hand side) $\mu(t)(1-\mu(t))\varrho_H(t)$ is nonnegative times negative, which is nonpositive. By the contradiction, we have thus proved Theorems 7 and 8. $\square$

4.6. Proof of Equation (3) in Remark 10

Recall that $\mathfrak{str} = \frac{\mathfrak{d}_{X\cap Y} - \mathfrak{d}_X \mathfrak{d}_Y}{\mathfrak{d}_{X,Y} - \mathfrak{d}_X \mathfrak{d}_Y}$, and also recall the definition of the modified alignment strength $\mathfrak{str}' := \frac{\overline{\mathfrak{d}_{X\cap Y} - \mathfrak{d}_X \mathfrak{d}_Y}}{\overline{\mathfrak{d}_{X,Y} - \mathfrak{d}_X \mathfrak{d}_Y}}$. The main goal of this section is to prove Equation (3) in Remark 10; namely, we show that

$$\mathfrak{str}' = \frac{\mathfrak{d}_{X\cap Y} - \mathfrak{d}_X\mathfrak{d}_Y + \frac{1}{4}\left[\frac{\Delta}{N^2} - (\mathfrak{d}_X - \mathfrak{d}_Y)^2\right]}{\mathfrak{d}_{X,Y} - \mathfrak{d}_X\mathfrak{d}_Y + \frac{1}{4}\left[\frac{\Delta}{N^2} - (\mathfrak{d}_X - \mathfrak{d}_Y)^2\right]} = \frac{\mathfrak{d}_{X\cap Y} - \mathfrak{d}_{X,Y}^2 + \frac{\Delta}{4N^2}}{\mathfrak{d}_{X,Y}(1-\mathfrak{d}_{X,Y}) + \frac{\Delta}{4N^2}}. \tag{7}$$

In order to do this, we will appeal to the following identities:

$$\mathfrak{d}_X + \mathfrak{d}_Y = \mathfrak{d}_{X\cap Y} + \mathfrak{d}_{X\cup Y} \tag{8}$$

$$\mathfrak{d}_{X,Y} = \frac{\mathfrak{d}_{X\cap Y} + \mathfrak{d}_{X\cup Y}}{2} \tag{9}$$

$$N \cdot \mathfrak{d}_{X\cap Y} + \Delta = N \cdot \mathfrak{d}_{X\cup Y} \tag{10}$$

$$\mathfrak{d}_{X\cap Y} = \mathfrak{d}_{X,Y} - \frac{\Delta}{2N} \tag{11}$$

$$\mathfrak{d}_{X\cup Y} = \mathfrak{d}_{X,Y} + \frac{\Delta}{2N} \tag{12}$$

Equation (8) holds by simple inclusion-exclusion, Equation (9) follows directly from Equation (8), Equation (10) is combinatorially trivial, and Equations (11) and (12) follow from Equations (9) and (10).

It is trivial to see that $\mathfrak{d}_{X\cap Y}$ and $\mathfrak{d}_{X,Y}$ are balanced, so we need only compute $\overline{\mathfrak{d}_X\mathfrak{d}_Y}$. Indeed, for any $h \in \{0, \star, 1\}^N$ and and any $(x,y) \in \mathcal{X}_h$, we have the following (using the identities in Equations (8) through (12), and combinatorial symmetry, and well-known identities involving binomial coefficients):

$$\overline{\mathfrak{d}_X\mathfrak{d}_Y}(x,y) = \frac{1}{2^{\Delta(x,y)}} \sum_{(x',y')\in\mathcal{X}_h} \mathfrak{d}_X(x')\mathfrak{d}_Y(y')$$

$$= \frac{1}{2^{\Delta(x,y)}} \sum_{i=0}^{\Delta(x,y)} \binom{\Delta(x,y)}{i} \frac{N\mathfrak{d}_{X\cap Y}(x,y)+i}{N} \;\; \frac{N\mathfrak{d}_{X\cap Y}(x,y)+\Delta(x,y)-i}{N}$$
$$= \frac{\mathfrak{d}_{X\cap Y}(x,y)\mathfrak{d}_{X\cup Y}(x,y)}{2^{\Delta(x,y)}} \sum_{i=0}^{\Delta(x,y)} \binom{\Delta(x,y)}{i}$$
$$+ \frac{\Delta(x,y)}{2^{\Delta(x,y)}N^2} \sum_{i=0}^{\Delta(x,y)} \binom{\Delta(x,y)}{i} i \;-\; \frac{1}{2^{\Delta(x,y)}N^2} \sum_{i=0}^{\Delta(x,y)} \binom{\Delta(x,y)}{i} i^2$$
$$= \mathfrak{d}_{X\cap Y}(x,y)\mathfrak{d}_{X\cup Y}(x,y) \;+\; \frac{\Delta(x,y)}{2^{\Delta(x,y)}N^2}\left(\Delta(x,y)2^{\Delta(x,y)-1}\right)$$
$$- \frac{1}{2^{\Delta(x,y)}N^2}\left(\Delta(x,y)+\Delta^2(x,y)\right)2^{\Delta(x,y)-2}$$
$$= \left(\mathfrak{d}_{X,Y}(x,y) - \frac{\Delta(x,y)}{2N}\right)\left(\mathfrak{d}_{X,Y}(x,y) + \frac{\Delta(x,y)}{2N}\right)$$
$$+ \frac{\Delta^2(x,y)}{2N^2} - \frac{\Delta(x,y)+\Delta^2(x,y)}{4N^2}$$
$$= \mathfrak{d}^2_{X,Y}(x,y) - \frac{\Delta(x,y)}{4N^2}. \tag{13}$$

Thus, by Equation (13) and the definition $\mathfrak{d}_{X,Y} = \frac{\mathfrak{d}_X+\mathfrak{d}_Y}{2}$ we obtain that

$$\overline{\mathfrak{d}_{X\cap Y} - \mathfrak{d}_X\mathfrak{d}_Y} = \mathfrak{d}_{X\cap Y} - \overline{\mathfrak{d}_X\mathfrak{d}_Y} = \mathfrak{d}_{X\cap Y} - \mathfrak{d}^2_{X,Y} + \frac{\Delta}{4N^2}$$
$$= \mathfrak{d}_{X\cap Y} - \mathfrak{d}_X\mathfrak{d}_Y + \frac{1}{4}\left[\frac{\Delta}{N^2} - (\mathfrak{d}_X - \mathfrak{d}_Y)^2\right] \text{ and}$$
$$\overline{\mathfrak{d}_{X,Y} - \mathfrak{d}_X\mathfrak{d}_Y} = \mathfrak{d}_{X,Y} - \overline{\mathfrak{d}_X\mathfrak{d}_Y} = \mathfrak{d}_{X,Y}(1-\mathfrak{d}_{X,Y}) + \frac{\Delta}{4N^2}$$
$$= \mathfrak{d}_{X,Y} - \mathfrak{d}_X\mathfrak{d}_Y + \frac{1}{4}\left[\frac{\Delta}{N^2} - (\mathfrak{d}_X - \mathfrak{d}_Y)^2\right]; \tag{14}$$

from this we have that Equation (7), i.e. Equation (3), is proven, as desired. □

4.7. Proof of Corollary 11

Corollary 11 states that if the parameter space Θ is nondegenerate and also $\Theta \subseteq \mathcal{R}^o$ then the statistic $\mathfrak{d}_{X,Y}(1-\mathfrak{d}_{X,Y}) - \frac{1}{2N}\left(1-\frac{1}{2N}\right)\Delta$ is UMVUE for σ^2. We prove this now.

Because Θ is nondegenerate, we pointed out in Example 1 that $\Delta : \mathcal{X} \to \mathbb{R}$ is the UMVU estimator of $2\sum_{i=1}^N (1-\varrho_i)p_i(1-p_i)$; here where $\Theta \subseteq \mathcal{R}^o$, we thus have that $\frac{\Delta}{2N} : \mathcal{X} \to \mathbb{R}$ is the UMVU estimator for $\frac{1}{N}\sum_{i=1}^N p_i(1-p_i) = \mu(1-\mu) - \sigma^2$.

From Equation (14) and Theorem 4 we have that $\mathfrak{d}_{X,Y}(1-\mathfrak{d}_{X,Y}) + \frac{\Delta}{4N^2}$ is the UMVU estimator of $\mathbb{E}(\mathfrak{d}_{X,Y} - \mathfrak{d}_X\mathfrak{d}_Y)$, which is equal to $\mu(1-\mu)$ since by hypothesis $\mathfrak{d}_X$ and $\mathfrak{d}_Y$ are independent.

Finally, by Corollary 6, we have that $\mathfrak{d}_{X,Y}(1-\mathfrak{d}_{X,Y})+\frac{\Delta}{4N^2}-\frac{\Delta}{2N}$ is the UMVU estimator for $\mu(1-\mu)-[\mu(1-\mu)-\sigma^2]$, and the result is shown. $\square$

4.8. Necessity of Θ nondegeneracy assumption in Theorems 3, 4

In the statement of Theorems 3 and 4 we assume that the parameter space Θ is nondegenerate. In this section we show that the claims of these theorems can fail if this condition is not satisfied.

Specifically, we will focus on a scenario in which the value of μ is known, in which case the parameter space is reduced to parameter tuples that have the prescribed value μ. This restricts the parameter space Θ to a particular hyperplane, which makes Θ degenerate; we will show that the claims of Theorems 3 and 4 then fail, in general.

For simplicity, in this entire section, let us take $N=2$, set $\varrho_1=\varrho_2=0$, and suppose that $\mu : 0<\mu<1$ is known; other than this, we allow $0<p_1<1$ and $0<p_2<1$. Here, $\frac{p_1+p_2}{2}=\mu$ yields $p_2=2\mu-p_1$. Denote $\delta := \min\{\mu, 1-\mu\}$, and denote $p := p_1$; the parameter space is reduced to single variable p on the interval $(\mu-\delta,\mu+\delta)$. There are $2^4=16$ points in $\mathcal{X}$; for each $(x,y)\in\mathcal{X}$, the probability of (x,y) is given by

$$\phi_{(x,y)}(p) := \left\{\begin{array}{ll} p & \text{if } x_1=1 \\ 1-p & \text{if } x_1=0 \end{array}\right\} \times \left\{\begin{array}{ll} p & \text{if } y_1=1 \\ 1-p & \text{if } y_1=0 \end{array}\right\}$$
$$\times \left\{\begin{array}{ll} 2\mu-p & \text{if } x_2=1 \\ 1-2\mu+p & \text{if } x_2=0 \end{array}\right\} \times \left\{\begin{array}{ll} 2\mu-p & \text{if } y_2=1 \\ 1-2\mu+p & \text{if } y_2=0 \end{array}\right\},$$

which is a polynomial of degree 4.

In Section 4.2, consider the linear system in Equation (4); that 9-by-9 linear system—describing statistic S being an unbiased estimator of g—now becomes a 5-by-9 linear system over here. This is because the columns of the left hand side matrix $A\otimes A$ and also the right hand side of the linear system, which there were each 9-vectors consisting of the coefficients of particular polynomials in two variables, can each now—in the reduced parameter space—be expressed as polynomials of degree 4 in a single variable, thus with five coefficients instead of 9 coefficients. As a 5-by-9 linear system, there is a nontrivial nullspace, and linear system solutions describing unbiasedness are no longer unique, which implies that there will exist a statistic T also an unbiased estimator of g such that it does not hold for all $h\in\{0,\star,1\}^N$ that $\sum_{(x,y)\in\mathcal{X}_h} T(x,y)=\sum_{(x,y)\in\mathcal{X}_h} S(x,y)$. This completes our demonstration that the claim of Theorem 3 may fail in the absence of the nondegeneracy assumption for Θ.

Next, we illustrate that the claim of Theorem 4 may fail without the nondegeneracy assumption for Θ. The disagreement enumeration statistic Δ is, by definition, unbiased for its expected value $\mathbb{E}(\Delta)=2\sum_{i=1}^N(1-\varrho_i)p_i(1-p_i)$, and we pointed out in Example 1 that Δ is the UMVU estimator of $\mathbb{E}(\Delta)=2\sum_{i=1}^N(1-\varrho_i)p_i(1-p_i)$ if Θ is nondegenerate. Nonetheless, in general, we will see that there is no UMVU estimator of $\mathbb{E}(\Delta)=2\sum_{i=1}^N(1-\varrho_i)p_i(1-p_i)$ when μ is fixed.

Indeed, consider $N = 2$, any fixed value of μ, and let the parameter space be parameterized by p exactly as we did above in this section. We will next formulate a quadratic program to find an unbiased estimator for $\mathbb{E}(\Delta)$ which, for any given value of p, has least variance among the unbiased estimators. It will turn out that there is a unique solution to this quadratic program. Then, if the solution differs for two different values of p then there does not exists a UMVU estimator. Indeed, we performed computations, and found that this occurred when $\mu = .25$, as one example of many.

We now describe how to compute (in the scenario of this section) the unbiased estimator which, for any given p, has least variance among the unbiased estimators.

Let the points in $\mathcal{X}$ be ordered in any specified way, say $z_1, z_2, \ldots, z_{16}$. Define the matrix $M \in \mathbb{R}^{5\times 16}$ wherein, for all i, j, the entry M_{ij} is the coefficient of p^{i-1} in the polynomial $\phi_{z_j}(p)$ (where $\phi_{z_j}(p)$ is as defined earlier in this section). For any statistic $S : \mathcal{X} \to \mathbb{R}$, let S be expressed as a vector $\vec{S} \in \mathbb{R}^{16}$ wherein, for all $i = 1, 2, \ldots, 16$, we define $\vec{S}_i := S(z_i)$. A function on the reduced parameter space $g : (\mu - \delta, \mu + \delta) \to \mathbb{R}$ can only have an unbiased estimator if g is a polynomial in the variable p of degree at most 4; this is because g would need to be a linear combination of the ϕ's. Say that $\vec{g} \in \mathbb{R}^5$ is the vector wherein for all $i = 1, 2, \ldots, 5$, we define $\vec{g}_i$ to be the coefficient of p^{i-1} in g. Because $(\mu - \delta, \mu + \delta)$ is a nontrivial interval (indeed, we just need at least 5 points) and by the uniqueness of interpolating polynomials, we have that the unbiased estimators S of any particular g are precisely the solutions $\vec{S}$ of the linear system $M\vec{S} = \vec{g}$.

Suppose that there exists an unbiased estimator of g. Among unbiased estimators of g, to find one of minimum variance **for any specific value of** $p \in (\mu - \delta, \mu + \delta)$, we proceed as follows. Let the vector of sample point probabilities for the respective 16 sample space points be denoted $\vec{\varpi} \in \mathbb{R}^{16}$; we have $\vec{\varpi}^T = [p^0, p^1, p^2, p^3, p^4] \cdot M$, which is a positive vector since $p \in (\mu - \delta, \mu + \delta)$; finding a (globally) unbiased estimator with (specifically for p) minimum variance is equivalent to the quadratic, convex optimization problem $\min \sum_{i=1}^{16} \vec{\varpi}_i \vec{S}_i^2$ such that $\vec{S}$ satisfies $M\vec{S} = \vec{g}$ (we minimize the second moment for the estimator, since the first moment is fixed). Define a bijective change of variables where new variables $\vec{S}' \in \mathbb{R}^{16}$ are such that for all $i = 1, 2, \ldots, 16$ we have $\vec{S}'_i := \sqrt{\vec{\varpi}_i} \vec{S}_i$, and define $M' \in \mathbb{R}^{5\times 16}$ such that, for all i, j, we have $M'_{ij} = \frac{1}{\sqrt{\vec{\varpi}_j}} M_{ij}$. Now this minimum variance problem is equivalent to $\min \|\vec{S}'\|_2$ such that $\vec{S}'$ satisfies $M'\vec{S}' = \vec{g}$. Classical generalized inverse theory guarantees a unique solution $\vec{S}' = M'^{\dagger}\vec{g}$ (the symbol $\dagger$ denotes the Moore-Penrose Generalized Inverse of the matrix), which corresponds to statistic S wherein $S(z_i) = \frac{(M'^{\dagger}\vec{g})_i}{\sqrt{\vec{\varpi}_i}}$ for each $i = 1, 2, \ldots, 16$, which is unique as having minimum variance (for the particular value of p) among the (globally) unbiased estimators.

This concludes the description of the way we computed, in the scenario of this section, the unbiased estimator which, for a fixed value of p, has least variance. (An excellent reference for matrix analysis in general, with theory of generalized inverses, is [6].)

5. Simulation experiments: comparing $\mathfrak{str}$, $\overline{\mathfrak{str}}$, and $\mathfrak{str}'$

As we mentioned earlier, in [4] it was empirically demonstrated—for correlated graphs in broad families within our model—that graph matching complexity as well as graph matchability are each functions of total correlation, and it was also proved in [4] that alignment strength $\mathfrak{str}$ is a strongly consistent estimator of ρ_T. The specific formulation/definition of alignment strength $\mathfrak{str}$ arose in a very natural way; see [4]. Nonetheless, $\mathfrak{str}$ suffers from a deficiency; in Example 3 we pointed out that $\mathfrak{str}$ is not balanced. The balanced statistic $\overline{\mathfrak{str}}$ reduces the variance, keeping the expected value unchanged. In this section we will empirically demonstrate that another balanced statistic, denoted $\mathfrak{str}'$, is often superior to $\overline{\mathfrak{str}}$ in estimating ρ_T. Note that there is no contradiction to Theorem 4, which asserts that, assuming the parameter space is nondegenerate, $\overline{\mathfrak{str}}$ is UMVUE for $\mathbb{E}(\mathfrak{str})$; indeed, $\mathfrak{str}'$ can be biased as an estimator of $\mathbb{E}(\mathfrak{str})$. Which can be a good thing; we will see that $\mathfrak{str}'$ frequently has less bias than $\overline{\mathfrak{str}}$ in the estimation of ρ_T, and in all of these experiments here $\mathfrak{str}'$ has less variance than $\overline{\mathfrak{str}}$.

But we first make a computationally helpful observation about computing the value of $\overline{\mathfrak{str}}$.

In general, when given an arbitrary statistic $S : \mathcal{X} \to \mathbb{R}$, the computation of the value of $\overline{S}$, even for just one particular sample space point $(x, y) \in \mathcal{X}$, can require exponential time; indeed, there are $2^{\Delta(x,y)}$ values to average. In the case of computing $\overline{\mathfrak{str}}$, this computation can be greatly simplified as follows. Given any particular $h \in \{0, \star, 1\}^N$ and any particular $(x, y) \in \mathcal{X}_h$ (such that not both x and y are all zeros, and not both x and y are all ones), we have by Equation (2) and Equation (10), that

$$\begin{aligned}
&\overline{\mathfrak{str}}(x,y)\\
&:= \frac{1}{2^{\Delta(x,y)}} \sum_{(x',y')\in\mathcal{X}_h} \mathfrak{str}(x',y')\\
&= \frac{1}{2^{\Delta(x,y)}} \sum_{i=0}^{\Delta(x,y)} \binom{\Delta(x,y)}{i} \frac{\mathfrak{d}_{X\cap Y}(x,y) - \left(\mathfrak{d}_{X\cap Y}(x,y) + \frac{i}{N}\right)\left(\mathfrak{d}_{X\cap Y}(x,y) + \frac{\Delta(x,y)}{N} - \frac{i}{N}\right)}{\mathfrak{d}_{X,Y}(x,y) - \left(\mathfrak{d}_{X\cap Y}(x,y) + \frac{i}{N}\right)\left(\mathfrak{d}_{X\cap Y}(x,y) + \frac{\Delta(x,y)}{N} - \frac{i}{N}\right)}
\end{aligned}$$

The above provides a linear time algorithm for computing $\overline{\mathfrak{str}}$ for any $(x, y) \in \mathcal{X}$, although this computation is much more involved then the very simple formula for $\mathfrak{str}'$ as given in Remark 10.

Now we are prepared to do a simulation experiment to compare the variances of $\mathfrak{str}$, and $\overline{\mathfrak{str}}$, and $\mathfrak{str}'$. The expected values of $\mathfrak{str}$ and $\overline{\mathfrak{str}}$ are of course the same, so there is no difference between their biases in estimating ϱ_T. However, the expected values of $\mathfrak{str}$ and $\mathfrak{str}'$ are not the same, so we want to also compare their biases in estimating ϱ_T.

In the first set of experiments, we did 200 independent replicates of the following experiment. We realized $p_1, p_2, p_3, p_4, p_5, p_6, \varrho_1, \varrho_2, \varrho_3, \varrho_4, \varrho_5, \varrho_6$ (which correspond to the six pairs of vertices in a vertex set with four vertices) independently from a Uniform$(0, 1)$ distribution; the first five such experiments' values were:

p_1	p_2	p_3	p_4	p_5	p_6	ϱ_1	ϱ_2	ϱ_3	ϱ_4	ϱ_5	ϱ_6
0.6892	0.7224	0.4795	0.8985	0.4022	0.7043	0.8429	0.9852	0.8006	0.3118	0.5768	0.5751
0.7482	0.1499	0.6393	0.1182	0.6207	0.7295	0.8988	0.6088	0.7388	0.0553	0.9440	0.0100
0.4505	0.6596	0.5447	0.9884	0.1544	0.2243	0.9390	0.2537	0.1417	0.7538	0.8715	0.8094
0.0838	0.5186	0.6473	0.5400	0.3813	0.2691	0.8154	0.1326	0.4379	0.1319	0.5076	0.6088
0.2290	0.9730	0.5439	0.7069	0.1611	0.6730	0.0014	0.5450	0.3504	0.3559	0.7888	0.4799

Then, for these realized parameters, we computed (exactly, by enumerating the sample space and sample point probabilities) the values of $\mathbb{E}(\mathfrak{str})$, $\mathbb{E}(\mathfrak{str}')$, ϱ_T, $\text{Var}(\mathfrak{str})$, $\text{Var}(\overline{\mathfrak{str}})$, and $\text{Var}(\mathfrak{str}')$. Of course, $\mathbb{E}(\mathfrak{str}) = \mathbb{E}(\overline{\mathfrak{str}})$. The first five experiments' outcomes were:

$\mathbb{E}(\mathfrak{str})$	$\mathbb{E}(\mathfrak{str}')$	ϱ_T	$\text{Var}(\mathfrak{str})$	$\text{Var}(\overline{\mathfrak{str}})$	$\text{Var}(\mathfrak{str}')$
0.6851	0.6857	0.7516	0.1219	0.1214	0.1206
0.6835	0.6843	0.7093	0.0885	0.0879	0.0870
0.6827	0.6833	0.7011	0.0745	0.0740	0.0734
0.4310	0.4339	0.4697	0.1345	0.1318	0.1291
0.5619	0.5635	0.5789	0.1073	0.1059	0.1043

In every one of these 200 experiments, we had that $\text{Var}(\mathfrak{str}) > \text{Var}(\overline{\mathfrak{str}}) > \text{Var}(\mathfrak{str}')$. In 199 of these 200 experiments we had $\mathbb{E}(\mathfrak{str}) < \mathbb{E}(\mathfrak{str}') < \varrho_T$.

We then repeated the 200 experiments, except that $\varrho_i = 0$ for all $i = 1, 2, 3, 4, 5, 6$. The first five experiments' outcomes were:

$\mathbb{E}(\mathfrak{str})$	$\mathbb{E}(\mathfrak{str}')$	ϱ_T	$\text{Var}(\mathfrak{str})$	$\text{Var}(\overline{\mathfrak{str}})$	$\text{Var}(\mathfrak{str}')$
0.3278	0.3320	0.3234	0.1222	0.1182	0.1149
0.4965	0.4986	0.4918	0.1052	0.1033	0.1014
0.4240	0.4269	0.4169	0.1098	0.1072	0.1048
0.1260	0.1335	0.1333	0.1433	0.1354	0.1307
0.5204	0.5225	0.5177	0.1094	0.1076	0.1056

Again, in every one of these 200 experiments, we had that $\text{Var}(\mathfrak{str}) > \text{Var}(\overline{\mathfrak{str}}) > \text{Var}(\mathfrak{str}')$. However, in only 41 of the 200 experiments was the bias of $\mathfrak{str}'$ less than that of $\mathfrak{str}$, meaning that $|\mathbb{E}(\mathfrak{str}') - \varrho_T| < |\mathbb{E}(\mathfrak{str}) - \varrho_T|$.

We then repeated the first 200 experiments, except that $p_i = \frac{1}{2}$ for all $i = 1, 2, 3, 4, 5, 6$. The first five experiments' outcomes were:

$\mathbb{E}(\mathfrak{str})$	$\mathbb{E}(\mathfrak{str}')$	ϱ_T	$\text{Var}(\mathfrak{str})$	$\text{Var}(\overline{\mathfrak{str}})$	$\text{Var}(\mathfrak{str}')$
0.6866	0.6872	0.7392	0.0989	0.0983	0.0976
0.3062	0.3108	0.3496	0.1375	0.1330	0.1293
0.3776	0.3812	0.4257	0.1367	0.1333	0.1301
0.3745	0.3781	0.4221	0.1384	0.1349	0.1316
0.6384	0.6393	0.6919	0.1095	0.1086	0.1075

Again, in every one of these 200 experiments, we had that $\text{Var}(\mathfrak{str}) > \text{Var}(\overline{\mathfrak{str}}) > \text{Var}(\mathfrak{str}')$. In all these 200 experiments we had $\mathbb{E}(\mathfrak{str}) < \mathbb{E}(\mathfrak{str}') < \varrho_T$. (In all of the above 600 experiments, we adopted the convention that the statistics have the value 0 at the two sample points (x, y) where x and y are all zeros and where they are all ones. Indeed, we saw empirically that this choice had a negligible numerical impact on the experiments here.)

Remarkably, in all of the many tens of thousands of experiments that we conducted, for many different parameter values, we always found that the mean squared error in estimating ϱ_T, denote it $MSE(\cdot, \varrho_T)$, was lower for the modified alignment strength $\mathfrak{str}'$ than for the balanced alignment strength $\overline{\mathfrak{str}}$. Based on these computations, we conjecture the following.

Conjecture 13. *For all N and $\theta \in \mathcal{R}$, it holds that $MSE(\mathfrak{str}', \varrho_T) \leq MSE(\overline{\mathfrak{str}}, \varrho_T)$.*

6. Summary and future directions

Our setting is the correlated Bernoulli random graph model for the production of a pair of correlated random graphs, wherein different pairs of vertices are allowed different probabilities of adjacency, and inter-graph edge correlations are allowed to be different for different pairs of vertices. This is a broad and useful model. Our main results come in two groups.

The first group of results: We introduce a "balancing" procedure to lower the mean squared error for any statistic used to estimate any function of the model parameters; it is essentially a Rao-Blackwellization procedure utilizing the disagreement vector statistic, which we prove is complete and sufficient. Indeed, given any unbiased estimator of any function of the model parameters, we neatly characterize all unbiased estimators, as well as the UMVUE estimator for this function of the model parameters. With these tools, we obtain the second group of results, which involve estimating the total correlation parameter, which is of current interest in the theory of Graph Matching, and has been recently shown to play a critical role in matchability and also in graph matching runtime complexity (when graph matching is solved exactly via integer programming) [4].

Future steps would be to extend our results in this paper to broader random graph models and to settle Conjecture 13.

Acknowledgments

This material is based on research sponsored by the Air Force Research Laboratory and Defense Advanced Research Projects Agency (DARPA) under agreement number FA8750-20-2-1001 and FA8750-18-2-0035. This work is also supported in part by the D3M program of DARPA. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory and DARPA or the U.S. Government.

The authors also gratefully acknowledge the support of NIH grant BRAIN U01-NS108637. The authors also thank an anonymous referee for very helpful and thoughtful comments that added much to the quality of the exposition.

References

[1] D. Blackwell, Conditional expectation and unbiased sequential estimation, *The Annals of Mathematical Statistics* **18** (1947), pp 105–110. MR0019903
[2] D. Conte, P. Foggia, C. Sansone, M. Vento, Thirty years of graph matching in pattern recognition, *Int. J. Pattern Recognit. Artif. Intell.* **18** (2004), pp 265–298.
[3] D.E. Fishkind, S. Adali, H.G. Patsolic, L. Meng, D. Singh, V. Lyzinski, C.E. Priebe, Seeded graph matching, *Pattern Recognition* **87** (2019), pp 203–215.
[4] D.E. Fishkind, L. Meng, A. Sun, C.E. Priebe, V. Lyzinski, Alignment strength and correlation for graphs, *Pattern Recognition Letters* **125** (2019), pp 295–302.
[5] P. Foggia, G. Perncannella, M. Vento, Graph matching and learning in pattern recognition in the last 10 years, *Int. J. Pattern Recognit. Artif. Intell.* **28** (2014). MR3190010
[6] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press (1985); 2nd edition (2013). MR2978290
[7] S. Lang, *Algebra*, Springer, 3rd Edition (2005). MR1878556
[8] E.L. Lehmann and H. Scheffe, Completeness, similar regions and unbiased estimation: Part I, *Sankhya: The Indian Journal of Statistics* **10** (1950), pp 305–340. MR0039201
[9] E.L. Lehmann and H. Scheffe, Completeness, similar regions and unbiased estimation: Part II, *Sankhya: The Indian Journal of Statistics* **15** (1955), pp 219–236. MR0072410
[10] V. Lyzinski, D.E. Fishkind, M. Fiori, J.T. Vogelstein, C.E. Priebe, G. Sapiro, Graph matching: Relax at your own risk, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38** (2016), pp 60–73.
[11] V. Lyzinski, D.E. Fishkind, C.E. Priebe, Seeded graph matching for correlated Erdos-Renyi graphs, *Journal of Machine Learning Research* **15** (2014), pp 3513–3540. MR3291408
[12] C.R. Rao, Information and accuracy attainable in the estimation of statistical parameters, *Bulletin of the Calcutta Mathematical Society* **37** (1945), pp 81–91. MR0015748
[13] J. Yan, X.C. Yin, W. Lin, C. Deng, H. Zha, X. Yang, A short survey of recent advances in graph matching, In: *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, ACM (2016), pp 167–174.