Rank of the vertex-edge incidence matrix of r-out hypergraphs

Colin Cooper
Department of Informatics
King's College
London WC2B 4BG
England

Alan Frieze[*]
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA15213
U.S.A.

July 6, 2021

Abstract

We consider a space of sparse Boolean matrices of size $n \times n$, which have finite co-rank over $GF(2)$ with high probability. In particular, the probability such a matrix has full rank, and is thus invertible, is a positive constant with value about 0.2574 for large n.

The matrices arise as the vertex-edge incidence matrix of 1-out 3-uniform hypergraphs The result that the null space is finite, can be contrasted with results for the usual models of sparse Boolean matrices, based on the vertex-edge incidence matrix of random k-uniform hypergraphs. For this latter model, the expected co-rank is linear in the number of vertices n, [4], [6].

For fields of higher order, the co-rank is typically Poisson distributed.

1 Introduction

For positive integers $r \geq 1$, $s \geq 3$, let $\boldsymbol{M}(s, r, n)$ be the space of $n \times rn$ matrices with entries generated in the following manner. For each $i = 1, ..., n$ there are r columns $C_{i,j},\ j = 1, ..., r$. Each column $C_{i,j}$ has a unit entry in row i, and $s-1$ other unit entries, in rows chosen randomly with replacement from $[n]$, or without replacement from $[n] - \{i\}$, all other entries

[*] Research supported in part by NSF Grant DMS1661063

in the column being zero. In general we consider the arithmetic on entries in the matrix, (and thus the evaluation of linear dependencies), to be over $GF(2)$. If so, in the "with replacement case", if two unit entries coincide the entry is set to zero. When $r = 1$, the matrix consists of an identity matrix plus $s-1$ random units in each column.

When $s = 2$, and entries are chosen without replacement, M is the vertex-edge incidence matrix of the random graph $G_{r-\text{out}}(n)$. This model of random graphs has been extensively studied, and is known to be r-connected for $r \geq 2$, Fenner and Frieze [7], to have a perfect matching for $r \geq 2$, Frieze [8], and to be Hamiltonian for $r \geq 3$, Bohman and Frieze [3]. By considering $s \geq 3$ we are considering r-out, s-uniform hypergraphs. This is closer to a model of random matrices considered in Cooper, Frieze and Pegden, [6], where the columns are chosen independently from all columns with s ones. A more general paper by Coja-Oghlan et al., [4], gives the limiting rank in this latter model for a wide range of assumptions on the distribution of non-zero entries in the rows and columns. The fundamental difference between the r-out model of random matrices, and those of [4], [6] is the presence of an $n \times n$ identity matrix as a sub-matrix in the without replacement case.

Of particular interest is the case $r = 1$ which gives $n \times n$ Boolean matrices. We will show that over $GF(2)$, for $r = 1, s = 3$, the linear dependencies among the rows of M (*dependencies* for short) are w.h.p. either small (bounded in expectation) or large (of size about $n/2$), and the distributions of these dependencies are somewhat entangled. For $r = 1, s = 3$, define a Poisson parameter ϕ for small dependencies. The value of ϕ differs marginally in summation range between the "with replacement" ϕ_R, and "without replacement" models $\phi_{\overline{R}}$ as follows:

$$\phi_R = \sum_{\ell \geq 1} \frac{1}{\ell} (2e^{-2})^\ell \sum_{j=0}^{\ell-1} \frac{\ell^j}{j!}, \qquad \phi_{\overline{R}} = \sum_{\ell \geq 2} \frac{1}{\ell} (2e^{-2})^\ell \sum_{j=0}^{\ell-2} \frac{\ell^j}{j!}. \tag{1}$$

The numeric values are $\phi_R \approx 0.5215$, and $\phi_{\overline{R}} \approx 0.1151$, where $a \approx b$ means approximately equal.

Let π be the probability distribution given by

$$\pi(k) = \begin{cases} \prod_{j=1}^{\infty} \left(1 - \left(\frac{1}{2}\right)^j\right) & k = 0. \\ \frac{\prod_{j=k+1}^{\infty}\left(1-\left(\frac{1}{2}\right)^j\right)}{\prod_{j=1}^{k}\left(1-\left(\frac{1}{2}\right)^j\right)} \left(\frac{1}{2}\right)^{k^2} & k \geq 1, \end{cases} \tag{2}$$

and define

$$P(\sigma, \lambda) = \frac{\phi^\sigma}{\sigma!} e^{-\phi} \sum_{\ell \geq \lambda} \pi(\lambda) \binom{\ell}{\lambda} \left(\frac{1}{2^\sigma}\right)^j \left(1 - \frac{1}{2^\sigma}\right)^{\ell-\lambda}. \tag{3}$$

Here as in the rest of the paper, σ indicates the dimension of the space induced by *small* dependencies and λ indicates the dimension of the space induced by *large* dependencies.

Theorem 1. *Let M be chosen u.a.r. from $\boldsymbol{M}(3,1,n)$. Let $d \geq 0, d = O(1)$ be integer. The limiting probability that, over $GF(2)$, the matrix M has co-rank d, is given by*

$$\lim_{n\to\infty} \mathbb{P}(\text{co-rank}(M) = d) = \sum_{\sigma=0}^{d} P(\sigma, d-\sigma). \tag{4}$$

In particular,

$$\mathbb{P}(\text{rank}(M) = n) \sim P(0,0) = e^{-\phi}\pi(0) = e^{-\phi}\prod_{j=1}^{\infty}\left(1 - \left(\frac{1}{2}\right)^{j}\right).$$

Theorem 1 differs from previous results on sparse random Boolean matrices on several counts. In particular, the co-rank (dimension of the null space) is finite, and the matrix is invertible with probability $e^{-\phi}\pi(0)$, where $\pi(0) \approx 0.2888$. The problem can be seen as an instance of the change in rank, if any, arising from small perturbations of the identity matrix.

The finite co-rank given in Theorem 1 can be contrasted with results for the edge-vertex incidence matrix of random hypergraphs, ([4], [6]), where the expected co-rank is linear in the number of vertices n, and the probability of a full rank matrix is exponentially small.

The joint distribution of co-rank given by (3) is a curious mixture of a Poisson with parameter ϕ given in (1), and the distribution $\pi(\cdot)$ given in (2). This distribution was previously observed for dense matrices in [10] (see below for a full definition). This mixture arises due to a gap property in the size of the dependencies (small or large), which we next explain. The negative correlation between the two types of dependency is characterized by the binomial.

Let $\boldsymbol{x} \in \{0,1\}^n$ be a *dependency* if $\boldsymbol{x}M = 0$. Let $|\boldsymbol{x}| = |\{j : x_j = 1\}|$. We say that a set of rows $D \subseteq [n]$ is a dependency if $D = \{j : x_j = 1\}$ for some dependency $\boldsymbol{x}$. An ℓ-dependency is one where $|\boldsymbol{x}| = \ell$ or $|D| = \ell$. The following theorem summarises the type of dependency we can expect:

Theorem 2. *W.h.p. either (i) a dependency $\boldsymbol{x}$ is* small *i.e. $|\boldsymbol{x}| \leq \omega$ where $\omega \to \infty$ slowly or (ii) $\boldsymbol{x}$ is* large *i.e. $|\boldsymbol{x}| = n/2 + O(\sqrt{n \log n})$.*

This in itself is rather interesting. One does not expect to find this gap in the size of dependencies $\boldsymbol{x}$. Estimating the interaction between small and large dependencies is the problem we solve. A dependency $\boldsymbol{x}$ is *fundamental* if there is no other dependency $\boldsymbol{y} \neq \boldsymbol{x}$ such that $\boldsymbol{y} \leq \boldsymbol{x}$, componentwise. We will prove in Section 2 that the number Z of fundamental small dependencies is asymptotically distributed as $Po(\phi)$ i.e. Poisson with mean ϕ.

Equation (4) arises from the fact that $P(\sigma, \lambda)$ is the limiting probability that M that the small dependencies span a space of dimension σ and the large dependencies span a space of dimension λ.

For the model of random matrices over $GF(2)$ in which the entries $m_{i,j}$ are i.i.d Bernoulii random variables with $\mathbb{P}(m_{i,j}=1)=p$, the distribution of dimension d of the null space is given by $\pi(d)$ of (2) for a wide range of p. This result was obtained for $p=1/2$ by Kovalenko et al., [10], and extended to the range $\min(p(n),1-p(n)) \geq (\log n + c(n))/n$, (where $c(n) \to \infty$ slowly) by Cooper [5].

Finally we consider some other related cases. For $r=1$ and $s=2$, M has expected rank $\sim n-(\log n)/2$. This is because the expected number of components in a random mapping is $\sim (1/2)\log n$, (see e.g., [9]). Note: For s even, the rows of M add to zero modulo 2. The following theorem will be immediate from the proof of Theorem 1.

Theorem 3. *If $r \geq 2$ and $s=2,\ 3$, then M has rank $n^* = n - \mathbb{1}_{\{s=2\}}$, w.h.p.*

Results for other finite fields follow easily from the analysis over $GF(2)$. We use the non-standard notation $GF(t)$ for a finite field of order t, rather than the usual $GF(q)$; and for brevity we consider only the 'without replacement' case. We consider three simple models with u.a.r. entries from a distribution $\{f_i\}$ over the non-zero elements i of $GF(t)$. Because M has 3 entries in each column, there are more cases for $GF(3)$.

Model 1: The field is $GF(3)$, and all three non-zero entries in a column are 1.

Model 2: The diagonal entries are 1, and the two other non-zero entries in each column are drawn u.a.r. from the distribution $\{f_i\}$.

Model 3: All three non-zero entries in each column are drawn u.a.r. from the uniform distribution $\{f_i\}$.

For Model 2, let $\gamma = f_{t-1}$, $\alpha = \sum f_i f_{t-i-1}$. For Model 3, let $\gamma = \sum_i f_i f_{t-i}$, $\alpha = \sum_{i+j+k=0} f_i f_j f_k$. Let ϕ_t be given by

$$\phi_t = \sum_{\ell \geq 2} \frac{1}{\ell} \left(2\gamma e^{-2}\right)^{\ell} \sum_{i=0}^{\ell-2} \frac{\ell^i}{i!}. \tag{5}$$

Theorem 4. *The following asymptotic results hold over $GF(t)$.*

1. *Model 1: If $t=3$ the limiting probability that M has rank $n-1$ is 1.*

2. *Models 2 and 3: If $t \geq 3$ then provided $\alpha < 2\gamma \leq 1$,*

$$\mathbb{P}(\text{rank}(M) = n-d) \sim \frac{\phi_t^d}{d!}\, e^{-\phi_t}.$$

In the simplest case where the entries are sampled uniformly from the non-zero elements of $GF(t)$, the theorem holds for either model with $\gamma = 1/(t-1)$.

Notation: Apart from $O(\cdot), o(\cdot), \Omega(\cdot)$ as a function of $n \to \infty$, we use the notation $A_n \sim B_n$ if $\lim_{n\to\infty} A_n/B_n = 1$. The symbol $a \approx b$ indicates approximate numerical equality due e.g., to decimal truncation. The notation $\omega(n)$ describes a function tending to infinity as $n \to \infty$. The expression *with high probability* (w.h.p.), means with probability $1 - o(1)$, where the $o(1)$ is a function of n, which tends to zero as $n \to \infty$.

Outline of the proof for $GF(2)$ with $r = 1, s = 3$

Because the proofs are rather technical, *we give a detailed proof in the "with replacement" model*, and indicate separately in Section 9 why these results are also valid in the "without replacement" model. The difference in the range of summation indices for $\phi_{\overline{R}}$ is explained in detail in Section 9.2.

We refer to the rows of M as $M_i, i \in [n]$ and to the columns as $C_j, j \in [n]$. By a set of rows S, we mean the set of rows $M_i, i \in S$. A set of rows with indices L is linearly dependent (zero-sum) if $\sum_{i\in L} M_i = 0(\text{mod } 2)$. A linear dependence L is *small* if $|L| \leq \omega$, where $\omega = \omega(n)$ is a function tending slowly to infinity with n. A linear dependence L is *large* if $|L| = (n/2)(1 + O(\sqrt{\log n/n}))$. As part of our proof, we show that w.h.p. there are no other sizes of dependency. A set of zero-sum rows L is *fundamental* if L contains no smaller zero-sum set and is disjoint from all other zero-sum sets. The zero-sum sets of size about $n/2$ are not disjoint. We count k-sequences of large dependencies with a property we call *simple.* Many of the problems with the proofs arise because the large dependencies are not disjoint, and are conditioned by the simultaneous presence of small linear dependencies in M.

We next outline the main steps in the proof of Theorem 1.

1. In Section 2 we prove that the number Z of small fundamental dependencies has factorial moments $\mathbf{E}(Z)_k \sim \phi^k$, where ϕ is given by (1). Thus Z is asymptotically Poisson distributed and
$$\mathbb{P}\left(M \text{ has } i \text{ small fundamental linear dependencies} \sim \frac{\phi^i}{i!}e^{-\phi}\right).$$
2. For $M \in \boldsymbol{M}(3, 1, n)$ w.h.p. any fundamental sets of zero-sum rows of M are either small (of size $\ell \leq \omega$) or large (of size $\ell = (n/2)(1 + O(\sqrt{\log n/n})))$. This is proved in Section 3.

3. In Section 5 we discuss *simple* sequences of large dependencies, and in Section 6 we estimate the moments of these sequences and determine their interaction with small dependencies.

4. We estimate the number of simple sequences, conditional on the the number of small fundamental dependencies. This leads to an approximate set of linear equations whose solution completes the proof of Theorem 1.

2 Small linear dependencies in $GF(2)$: with replacement

Notation For $1 \leq k \leq \omega$, where $\omega \to \infty$ arbitrarily slowly with n, let $X_k(M)$ or $Y_k(M)$ denote the number of index sets of k-dependencies in M. A k-dependency is *small* if $k \leq \omega$ and we use Y_k when $k \leq \omega$ and use X_k when $k \sim n/2$. We will show that for other values of k, $X_k = 0$ w.h.p. We also use $Z_d, d \leq \omega$ to denote the number d of fundamental (minimal) dependent sets among the rows of M.

We first consider dependencies with $s = o(n^{1/2})$ rows. For $L \subseteq [n]$, let $\mathcal{F}(S)$ denote the event that the rows corresponding to S are dependent. Let Y_s denote the number of s-set dependencies.

Lemma 5. *If* $|S| = s = o(n^{1/2})$ *then*

$$\mathbb{P}(\mathcal{F}(S)) \sim \left(\frac{2s}{n}\right)^s e^{-s}. \tag{6}$$

If $\omega \to \infty$, $\omega \leq s = o(n^{1/2})$ *then* $Y_s = 0$ *w.h.p.*

Proof. Suppose that $s = o(n^{1/2})$ and $S = [s]$. Then,

$$\begin{aligned}\mathbb{P}(\mathcal{F}(S)) &= \left(2\left(\frac{s}{n}\right)\left(\frac{n-s}{n}\right)\right)^s \left(\left(\frac{s}{n}\right)^2 + \left(\frac{n-s}{n}\right)^2\right)^{n-s} \\ &\sim \left(\frac{2s}{n}\right)^s e^{-2s}, \qquad \text{using } s = o(\sqrt{n}). \end{aligned} \tag{7}$$

Explanation: $2\left(\frac{s}{n}\right)\left(\frac{n-s}{n}\right)$ is the probabilty that exactly one of the two random choices in a column of S lies in a row of S. $\left(\frac{s}{n}\right)^2 + \left(\frac{n-s}{n}\right)^2$ is the probabilty that neither of the two random choices in a column of $[n] \setminus S$ lies in a row of S.

This verifies (6). It follows that

$$\mathbf{E}(Y_s) \sim \binom{n}{s}\left(\frac{2s}{n}\right)^s e^{-2s} \sim \frac{(2s)^s e^{-2s}}{s!},$$

As $\mathbf{E}Y_{s+1}/\mathbf{E}(Y_s) \sim 2/e$ we have that $\mathbf{E}Y_\omega = e^{-\Omega(\omega)}$ and so w.h.p. there are no dependencies with $\omega \leq s = o(n^{1/2})$. □

The next lemma deals with small fundamental dependencies. For $S \subseteq [n]$, let $\mathcal{F}^*(S)$ denote the event that the rows corresponding to S deprise a fundamental dependency. Let

$$\kappa_s = \frac{(s-1)!}{s^s}\sigma_s, \tag{8}$$

where

$$\sigma_s = \sum_{j=0}^{s-1} \frac{s^j}{j!}.$$

Lemma 6. $\mathbb{P}(\mathcal{F}^*(S) \mid \mathcal{F}(S)) = \kappa_s$.

Proof. The rows of the dependency S consist of an $s \times s$ sub-matrix $M_{S,S}$ and a zero $(s \times n - s)$ sub-matrix. For $i \in S$, if $M_{i,i} = 1$, then wh.p. there is a unique entry $M_{j,i} = 1$ which gives rise to an *edge* (i, j). If $M_{i,i} = 0$ we regard this as a loop (i, i). Thus $M_{S,S}$ is the incidence matrix of a random functional digraph D_S, and S is fundamental iff the underlying graph of D_S is connected. For $s \geq 1$, $\mathbb{P}(D_S \text{ is connected}) = \kappa_s$ (see e.g., [1] or [9]). □

We now prove

Lemma 7. *Small fundamental dependent sets of M are pairwise disjoint, w.h.p.*

Proof. Let S, T be two small fundamental zero-sum row sets with a non-trivial intersection $C = S \cap T$ and differences $A = S \backslash T,\ B = T \backslash S$, where $A, B \neq \emptyset$. As the functional digraphs D_S, D_T are connected At least one column of $A \cup B$ must contain two random ones. The probability of this is at most

$$\sum_{k=2}^{2\omega} \binom{n}{k} k \left(\frac{\omega}{n}\right)^{k-1} \left(\frac{k}{n}\right)^2 = o(1). \tag{9}$$

□

Given this lemma we can now prove

Lemma 8. *The number Z of small fundamental dependent sets among the rows of M is asymptotically Poisson distributed with parameter ϕ_R, and thus*

$$\mathbb{P}(Z = d) \sim \frac{\phi_R^d}{d!} e^{-\phi_R}. \tag{10}$$

Proof. Fix $S \subseteq [n]$ and let $S_1, \ldots, S_d$ be a partition of S with $|S_i| = s_i, i = 1, 2, \ldots, d$. Let $P(s_1, \ldots, s_d)$ be the probability that each $S_i, i = 1, 2, \ldots, d$ is a fundamental set, given that S is a dependency. Thus,

$$P(s_1, \ldots, s_d) = \frac{(s_1)^{s_1} \cdots (s_d)^{s_d}}{s^s} \prod_{i=1,\ldots,d} \mathbb{P}(D_{S_i} \text{ connected}) = \frac{1}{s^s} \prod_{i=1}^{d} (s_i - 1)! \sigma_{s_i}.$$

Explanation: the factor $\frac{(s_1)^{s_1} \cdots (s_d)^{s_d}}{s^s}$ is the conditional probability that the random choices for columns with index in S_i are in rows with index in S_i.

Thus, using (6), we see that

$$\begin{aligned}
\mathbf{E}(Z)_d \sim & \sum_{s \geq 1} \frac{(2s)^s}{s!} e^{-2s} \sum_{s_1 + \cdots + s_d = s} \binom{s}{s_1, \ldots, s_d} P(s_1, \ldots, s_d) & (11) \\
= & \sum_{s \geq 1} \sum_{s_1 + \ldots + s_d = s} \prod_{i=1}^{k} (2e^{-2})^{s_i} \frac{1}{s_i} \sigma_{s_i} \\
= & \left(\sum_{s \geq 1} \frac{1}{s} (2e^{-2})^s \sigma_s \right)^d \\
= & \phi_R^d. & (12)
\end{aligned}$$

Thus, by the method of moments, the number of small disjoint fundamental zero-sum sets Z tends tend to a Poisson distribution with parameter ϕ_R. □

3 Large zero-sum sets: First moment calculations

Define an index set J_a as follows,

$$J_a = \{n/2 - \sqrt{an \log n} \leq \ell \leq n/2 + \sqrt{an \log n}\} \text{ and } \overline{J}_a = [n] \setminus J_a,\ a \geq 0. \tag{13}$$

Lemma 9. *(***Large linearly dependent sets.)** *Let X_ℓ denote the number of ℓ-dependencies among the rows of M.*

(i) $\sum_{\ell\in J_1} \mathbf{E}X_\ell \sim 1$.

(ii) Let $D = [n]\backslash([\omega]\cup J_1)$*, where* $\omega \to \infty$ *arbitrarily slowly with* n*. Then* $\sum_{\ell\in D} \mathbf{E}X_\ell = o(1)$.

Proof. From (7), the expected number of dependencies of size ℓ is

$$\mathbf{E}X_\ell = \binom{n}{\ell}\left(2\left(\frac{\ell}{n}\right)\left(\frac{n-\ell}{n}\right)\right)^{\ell}\left(\left(\frac{\ell}{n}\right)^2 + \left(\frac{n-\ell}{n}\right)^2\right)^{n-\ell}.$$

We next approximate the expression for $\mathbf{E}X_\ell$. We note the following expansion.

$$(1+x)\log(1-x^2)+(1-x)\log(1+x^2) = -2\left(x^3 + \frac{x^4}{2} + \frac{x^7}{3} + \sum_{k\geq 4} \mathbb{1}_{\{k \text{ even}\}}\frac{x^{2k}}{k}\left(1+\frac{kx^3}{k+1}\right)\right). \tag{14}$$

We write $\mathbf{E}X_\ell = \binom{n}{\ell}\Phi_\ell^n$, $\ell = (n/2)(1+\varepsilon)$, where

$$\begin{aligned}
\Phi_\ell =& \left(\frac{1-\varepsilon^2}{2}\right)^{\frac{(1+\varepsilon)}{2}}\left(\left(\frac{1+\varepsilon}{2}\right)^2 + \left(\frac{1-\varepsilon}{2}\right)^2\right)^{\frac{(1-\varepsilon)}{2}}\\
=&\frac{1}{2}(1-\varepsilon^2)^{\frac{(1+\varepsilon)}{2}}(1+\varepsilon^2)^{\frac{(1-\varepsilon)}{2}}\\
=&\frac{1}{2}\exp\left\{\frac{1}{2}\left((1+\varepsilon)\log(1-\varepsilon^2)+(1-\varepsilon)\log(1+\varepsilon^2)\right)\right\}\\
=&\frac{1}{2}\exp\left\{-\left(\varepsilon^3+\frac{\varepsilon^4}{2}+\frac{\varepsilon^7}{3}+\sum_{k\geq 4}\mathbb{1}_{\{k \text{ even}\}}\varepsilon^{2k}\left(\frac{1}{k}+\frac{\varepsilon^3}{k+1}\right)\right)\right\}\\
=&\frac{1}{2}\exp\left\{-\left(\varepsilon^3+\frac{\varepsilon^4}{2}+\varepsilon_7\right)\right\},
\end{aligned}\tag{15}$$

where $|\varepsilon_7| \leq 2|\varepsilon|^7/3$ for sufficiently small ε.
Also for $\ell = (n/2)(1+\varepsilon)$, $|\varepsilon| < 1$,

$$\binom{n}{\ell} = \left(1+O\left(\frac{1}{n}\right)\right)\ \frac{2^n}{\sqrt{2\pi n(1-\varepsilon^2)}}\ \exp\left(-n\left(\frac{\varepsilon^2}{2}+\frac{\varepsilon^4}{12}+\varepsilon_6\right)\right), \tag{16}$$

where $|\varepsilon_6| \leq |\varepsilon|^6/10$.

Case 1: $\ell \in J_1$ **.** From (16) with $|\varepsilon| = 2\sqrt{(\log n)/n}$ we have

$$\frac{1}{2^n}\sum_{\ell\notin J_1}\binom{n}{\ell} = O(1/n^{5/2}),$$

so that

$$\frac{1}{2^n}\sum_{\ell\in J_1}\binom{n}{\ell}=1-O(1/n^{5/2}).$$

Using (15), for $\ell\in J_1$, ${\Phi_\ell}^n=e^{\Theta(n\varepsilon^3)}/2^n$. Then, as $n\varepsilon^3=O(\log^{3/2}n/\sqrt{n})$,

$$\sum_{\ell\in J_1}\mathbf{E}X_\ell=\sum_{\ell\in J_1}\binom{n}{\ell}\frac{1}{2^n}e^{\Theta(n\varepsilon^3)}=1+o(1).$$

For future reference, we note that for $|\varepsilon|<c<1$,

$$\begin{aligned}\mathbf{E}X_\ell=&\binom{n}{\ell}\frac{1}{2^n}\exp\left\{-n\left(\varepsilon^3+\frac{\varepsilon^4}{2}+\varepsilon_7\right)\right\}\\=&\frac{(1+o(1))}{\sqrt{2\pi n(1-\varepsilon^2)}}\exp\left\{-n\left(\frac{\varepsilon^2}{2}+\varepsilon^3+\frac{\varepsilon^4}{2}+\frac{\varepsilon^4}{12}+\varepsilon_6+\varepsilon_7\right)\right\}\\=&\frac{(1+o(1))}{\sqrt{2\pi n(1-\varepsilon^2)}}\exp\left\{-\frac{n\varepsilon^2}{2}\left((1+\varepsilon)^2+\frac{\varepsilon^2}{6}+\varepsilon_6+\varepsilon_7\right)\right\}.\end{aligned}\tag{17}$$

Case 2: $\ell\in D$. Write $D=[n]\setminus([\omega]\cup J_1)$ as $D=D_1\cup D_2\cup D_3$ where $D_1=\{\omega,\dots,3n/10\}$, $D_2=\{7n/10,\dots,n\}$ and $D_3=D\setminus(D_1\cup D_2)$. Thus, for $\ell\in D_3$, $\ell=(n/2)(1+\varepsilon)$ where $-2/5\le\varepsilon\le-\sqrt{(2\log n)/n}$ or $\sqrt{(2\log n)/n}\le\varepsilon\le 2/5$.

Case $\ell\in D_1$. For sufficiently large n, Stirling's approximation implies that

$$\binom{n}{\ell}\le\frac{n^n}{\ell^\ell(n-\ell)^{n-\ell}},$$

so for some constant C (in both with and without replacement models)

$$\mathbf{E}X_\ell\le\frac{Cn^n}{\ell^\ell(n-\ell)^{n-\ell}}\left(2\left(\frac{\ell}{n}\right)\left(\frac{n-\ell}{n}\right)\right)^\ell\left(\left(\frac{\ell}{n}\right)^2+\left(\frac{n-\ell}{n}\right)^2\right)^{n-\ell}.\tag{18}$$

Continuing with this expression, using $\ell=\lambda n$ for $\lambda<1/2$,

$$\begin{aligned}\mathbf{E}X_\ell\le&C\left(\frac{2^\lambda}{\lambda^\lambda(1-\lambda)^{1-\lambda}}\lambda^\lambda(1-\lambda)^\lambda(\lambda^2+(1-\lambda)^2)^{1-\lambda}\right)^n\\=&C\left(2^\lambda(1-\lambda)^\lambda\left(1-\lambda+\frac{\lambda^2}{1-\lambda}\right)^{1-\lambda}\right)^n\\\le&C\left(2^\lambda(1-\lambda)^\lambda e^{-\lambda(1-\lambda)+\lambda^2}\right)^n\\=&C\left(2(1-\lambda)e^{-1+2\lambda}\right)^{\lambda n}\\=&C[g(\lambda)]^{\lambda n}.\end{aligned}$$

The function $g(\lambda)$ is strictly concave and has a unique maximum at $\lambda = 1/2$ with $g(1/2) = 1$. For $\lambda \leq 3/10$, $g(\lambda) \leq g(3/10) = (7/5)e^{-2/5} < 1$ so that

$$\sum_{\ell \in D_1} \mathbf{E} X_\ell \leq C \sum_{\ell \in D_1} g(3/10)^\ell = o(1).$$

Case $\ell \in D_2$. Referring to (17), the function $h(\varepsilon) = (\varepsilon^2/2)((1+\varepsilon)^2 + \varepsilon^2/6 + \varepsilon_6 + \varepsilon_7)$ satisfies $h(\varepsilon) > 2/25$ for $\varepsilon \geq 2/5$, and so

$$\sum_{\ell \in D_2} \mathbf{E} X_\ell \leq \sum_{\ell \in D_2} e^{-\Omega(n)} = o(1).$$

Case $\ell \in D_3$. For $\sqrt{(2\log n)/n} \leq |\varepsilon| \leq \sqrt{(25 \log n)/n}$, the function $h(\varepsilon) \geq (1-o(1))(\log n)/n$. Let D_{3a} be the values of ℓ in this range

$$\sum_{\ell \in D_{3a}} \mathbf{E} X_\ell = O(\sqrt{n \log n})/n^{1-o(1)}) = o(1/n^{1/3}).$$

Let $D_{3b} = D_3 \setminus D_{3a}$. Then $\varepsilon^2/2 \geq (25/2)(\log n)/n$, and $(1+\varepsilon)^2 + \varepsilon^2/6 + \varepsilon_6 + \varepsilon_7 > 9/25$. Referring to (17),

$$\sum_{\ell \in D_{3b}} \mathbf{E} X_\ell = O(n)/n^4 = o(1/n^3).$$

□

4 Higher moments of large zero-sum sets: Background

Let $A \Delta B$ denote the symmetric set difference $(A \cup B) \setminus (A \cap B)$ of the sets A and B. Suppose that, over $GF(2)$, the rows $M[i], i \in A$ indexed by A are zero-sum, thus $\boldsymbol{z}_A = \sum_{i \in A} M[i] = \mathbf{0}$. Let B be another set such that $\boldsymbol{z}_B = \mathbf{0}$. We can write $\boldsymbol{z}_A = \boldsymbol{z}_{A \setminus B} + \boldsymbol{z}_{A \cap B}$ and $\boldsymbol{z}_B = \boldsymbol{z}_{B \setminus A} + \boldsymbol{z}_{A \cap B}$. Adding these two sets of rows modulo 2 has the effect of canceling the intersection $A \cap B$. Thus (i) $\boldsymbol{z}_A + \boldsymbol{z}_B = 0$, whether $\boldsymbol{z}_{A \cap B}$ is itself zero-sum or not; and (ii) $\boldsymbol{z}_A + \boldsymbol{z}_B = \boldsymbol{z}_{A \Delta B}$.

Recall that a set of zero-sum rows is fundamental if it contains no smaller zero-sum set of rows. For small sets we were able to count fundamental dependencies directly. We have to adopt an alternative strategy for large zero-sum sets. We use an approach similar to the one given in [5]. We count *simple* sequences of large linearly dependent row sets $B = (B_1, ..., B_k)$, $k \geq 1$ constant, and where $|B_i| \in J_1$ so that $|B_i| \sim n/2$. A k-tuple of large dependent sets $B = (B_1, ..., B_k)$ is simple, if for all sequences $(j_1 < j_2 < ... < j_l)$ and $(1 \leq l \leq k)$ the set differences satisfy

$$|B_{j_1} \Delta B_{j_2} \Delta \cdots \Delta B_{j_l}| \in J_1 \tag{19}$$

For any given matrix M there is a largest k such that $B_1, ..., B_k$ are simple. In which case, we say k is *maximal* and $B_1, ..., B_k$ is a *maximal simple sequence*.

Let $V(M) = \{\emptyset\} \cup \{B : B \text{ is zero-sum in } M\}$, then $(V(M), \Delta)$ is a vector space over GF_2 under the convention that $0 \cdot B = \emptyset,\ 1 \cdot B = B$. In $V(M)$ a simple sequence $(B_1, ..., B_k)$ is an ordered basis for a subspace S of dimension k.

Given $k = O(1)$ linearly dependent sets of rows with index sets $B_1, \cdots, B_k$, there are 2^k intersections of these sets and their complements. For each $\boldsymbol{x} = (x_1, \cdots, x_k)$, $\boldsymbol{x} \in \{0,1\}^k$ we let $I_{\boldsymbol{x}} = \cap_{i=1,...,k} B_i^{(x_i)}$ where $B_i^{(0)} = \overline{B}_i = [n] \setminus B_i$ and $B_i^{(1)} = B_i$. The index sets $I_{\boldsymbol{x}}$ are disjoint by definition and their union (including $\boldsymbol{x}_0 = (0, \cdots, 0)$) is $[n]$.

Next let $B(\boldsymbol{x}) = \Delta_{i:x_i=1} B_i$ for $\boldsymbol{x} \in \{0,1\}^k$. Let $K = 2^k - 1$. Define a $K \times K$ matrix $U = U[x, y]$, $x, y \in \{0,1\}^k$, $\boldsymbol{x}, \boldsymbol{y} \neq 0$, by $U(\boldsymbol{x}, \boldsymbol{y}) = 1$ iff $I_{\boldsymbol{y}} \subseteq B(\boldsymbol{x})$.

$$\text{Row index } \boldsymbol{x} = (x_1, x_2, \dots, x_k) \text{ is the indicator vector for } B(\boldsymbol{x}) = \Delta_{i:x_i=1} B_i,$$

$$\text{Column index } \boldsymbol{y} = (y_1, y_2, \dots, y_k) \text{ is the indicator vector for } I_{\boldsymbol{y}} = \bigcap_{i=1,...,k} B_i^{(y_i)},$$

Thus $B(\boldsymbol{x})$ is the union of the sets $I_{\boldsymbol{y}}$ where $y_i = 1$ for an odd number of the given sets B_i such that $x_i = 1$. This follows inductively by generating B_1, $B_1 \Delta B_2$, $(B_1 \Delta B_2) \Delta B_3$ etc in the given order. It follows that $U(\boldsymbol{x}, \boldsymbol{y}) = 1$ iff $x_i = y_i = 1$ for an odd number of indices i, and thus, over $GF(2)$,

$$U(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{k} x_i y_i. \tag{20}$$

Our aim is to use U, treated as a real matrix to show that w.h.p. $|I_{\boldsymbol{x}}| \sim n/2^k$ for every $\boldsymbol{x}$. We do this by observing that given the characterisation $U(\boldsymbol{x}, \boldsymbol{y}) = 1_{I_{\boldsymbol{y}} \subseteq B(\boldsymbol{x})}$, the vector $(|I_{\boldsymbol{x}}|, \boldsymbol{x} \in \{0,1\}^k, \boldsymbol{x} \neq 0$ is the solution $\boldsymbol{z}$ over the reals of an equation

$$U\boldsymbol{z} = \boldsymbol{b} \text{ where } \boldsymbol{b} \sim \frac{n}{2}\mathbf{1}, \tag{21}$$

assuming that $B = (B_1, ..., B_k)$ is simple. To prove that $|I_{\boldsymbol{x}}| \sim n/2^k$, we prove the properties of U listed in Lemma 10 below.

Equation (20) implies that by arranging the rows and column indices of U in the same order, U will be symmetric. We will choose an ordering such the first k rows and columns correspond to $x_i = e_i, i = 1, 2, \dots, k$ where $e_1 = (1, 0, \dots, 0)$ etc. After this we let Q be the $k \times K$ matrix with column indices x made up of the first k rows. Thus row i represents $B_i, i = 1, ..., k$ and U contains a $k \times k$ identity matrix in the first k rows and columns.

The row indexed by $\boldsymbol{x} = (x_1, ..., x_k)$ is the linear combination $\sum_{i=1}^k x_i \boldsymbol{r}_i$ of the rows of Q, and corresponds to $B(\boldsymbol{x})$ in the vector space $V(M)$ given above.

Lemma 10. *The $K \times K$ matrix U has the following properties:*

(i) The matrix U symmetric.

(ii) Every row or column of U has 2^{k-1} non-zero entries.

(iii) Any two distinct rows of U have 2^{k-2} common non-zero entries.

(iv) The matrix U is non-singular when the entries are taken to be over the real numbers, and the matrix $S = UU^\top = U^2 = 2^{k-2}(I + J)$ is symmetric, with inverse $S^{-1} = (1/2^{k-2})(I - J/2^k)$; where J is the all-ones matrix.

Proof. (i) This follows immediately from (20), and the above construction.

(ii) Fix $\boldsymbol{x}$ and assume that $x_1 = 1$. There are 2^{k-1} choices for the values of $y_i, i = 2, 3, \ldots, k$. Having made such a choice, there are two choices for y_1, exactly one of which will give $\sum_{i=1}^k x_i y_i = 1$.

(iii) Fix $\boldsymbol{x}, \boldsymbol{x}'$ and think of rows $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{x} + \boldsymbol{x}'$ as non-empty subsets of $[2^k]$. Then we have $|\boldsymbol{x}| = |\boldsymbol{x}'| = |\boldsymbol{x} \setminus \boldsymbol{x}'| + |\boldsymbol{x}' \setminus \boldsymbol{x}| = 2^{k-1}$, by (iii). Thus $|\boldsymbol{x}| + |\boldsymbol{x}'| - (|\boldsymbol{x} \setminus \boldsymbol{x}| + |\boldsymbol{x}' \setminus \boldsymbol{x}|) = 2|\boldsymbol{x} \cap \boldsymbol{x}'| = 2^{k-1}$.

(iv) That the matrix U is non-singular over the real numbers, uses an argument given in [2] (pages 11-13). Let $S = UU^\top$. Let $\boldsymbol{u}, \boldsymbol{v}$ be distinct rows of U, then $\boldsymbol{u} \cdot \boldsymbol{u} = 2^{k-1}$ and $\boldsymbol{u} \cdot \boldsymbol{v} = 2^{k-2}$. Thus $S = 2^{k-2}(I + J)$, where J is the all-ones matrix. The reader can check that $S^{-1} = \frac{1}{2^{k-2}}(I - \frac{1}{2^k}J)\ 2^{k-1}$ which implies that U is invertible too. □

The definition of a simple k-tuple $(B_1, ..., B_k)$ requires that all sets B_i be large and their set differences to be distinct and of size $\sim n/2$. Let $(|B_1|, \ldots, |B_k|) \sim (n/2)\mathbf{1}$ be the vector of these set sizes. Over the reals, solving (21) gives the sizes of the subsets $I_{\boldsymbol{x}}$.

Lemma 11. *Let $(B_1, ..., B_k)$ be a simple sequence. Then for all $\boldsymbol{x} \in \{0, 1\}^k$,*

$$|I_{\boldsymbol{x}}| = \frac{n}{2^k}\left(1 \pm 2^k\sqrt{\frac{\log n}{n}}\right). \tag{22}$$

Proof. Let $i = 1, ..., K$ index the rows of U and let $B(\boldsymbol{x})$ be the set corresponding to the row $\boldsymbol{x}$ of U. Let $U\boldsymbol{x} = \boldsymbol{b}$ where $b_{\boldsymbol{x}} = 2|B(\boldsymbol{x})|/n = 1 + \varepsilon_i$, where $|\varepsilon_i| \leq 2\sqrt{\log n/n}$. The matrix $S = U^2$, so $S\boldsymbol{x} = U\boldsymbol{b} = \boldsymbol{c}$ where $c_i = 2^{k-1}(1 + \delta_x)$ where $\delta_{\boldsymbol{x}} = \sum \varepsilon_j/2^{k-1}$, the summation being over a 2^{k-1}-subset of rows $\boldsymbol{x}$ of U. Thus, as J is $K \times K$ where $K = 2^k - 1$,

$$\boldsymbol{x} = S^{-1}\boldsymbol{c} = \frac{1}{2^{k-2}}\left(I - \frac{1}{2^k}J\right)\ 2^{k-1}(\mathbf{1} + \boldsymbol{\delta}) = \frac{1}{2^{k-1}}\mathbf{1} + \boldsymbol{\eta},$$

where $|\boldsymbol{\eta}| \leq 2^k\sqrt{\log n/n}$. It follows that w.h.p. the solution to (21) satisfies $|I_{\boldsymbol{x}}| = (n/2^k)(1 \pm 2^k\sqrt{\log n/n})$ for all $\boldsymbol{x} \in \{0,1\}^k$. □

Remark 12. *The proof of Lemma 11 implies that if we have a $K \times K$ matrix in which (ii), (iii) of Lemma 10 are satisfied asymptotically, then*

5 Simple sequences of large zero-sum sets.

Let $B_1, B_2, \ldots, B_k$ be a simple sequence. In row M_i of the matrix M, there is a 1 in the diagonal entry $M_{i,i}$. As $s = 3$ there needs to be two (random) 1's in column C_i chosen in a way to ensure the linear dependence of $B_1, \ldots, B_k$. The following lemma describes where these non-zeros must be placed.

Lemma 13. *$B_1, \cdots, B_k$ are dependencies if and only if the following holds for all $i \in [n]$: suppose that $i \in I_{\boldsymbol{x}}$, $\boldsymbol{x} = (x_1, ..., x_k)$. Suppose that the two random non-zeros $e_1(i), e_2(i)$ in column i are in $I_{\boldsymbol{u}}, I_{\boldsymbol{v}}$ respectively. Then we must have $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v} (\mathrm{mod}\ 2)$.*

Proof. Consider $1 \leq m \leq k$. If $x_m = 0$ then $i \notin B_m$ and so either none or both of j, j' are in B_m, and so zero or two unit entries in this column are in B_m. We must therefore have either $u_m = v_m = 0$ or $u_m = v_m = 1$ and $x_m = u_m + v_m$. If $x_m = 1$ then exactly one of $e_1(i), e_2(i)$ are in B_m and so $u_m = 1, v_m = 0$, or vice versa. Thus in all cases $x_m = u_m + v_m$. □

The main result of this section is the following.

Lemma 14. *Let $k \geq 1$ be a positive integer, and let X_k count the number of simple k-sequences of large dependencies. Then $\mathbf{E}(X_k) \sim 1$.*

Proof. We have to estimate the expected number of simple sequences $(B_1, ..., B_k)$ of large dependencies. By (22) of Lemma 11 the index sets $I_{\boldsymbol{x}}$ have size $|I_{\boldsymbol{x}}| = (n/2^k)(1+O(\sqrt{\log n/n}))$. Let $K = 2^k - 1$ as above, and let

$$\Omega = \left\{ \boldsymbol{h} = (h_0, h_1, ..., h_K) : h_i \text{ satisfies (22)}, \sum_{i=0}^{K} h_i \in J_1 \right\}.$$

Then we define $\Phi(\boldsymbol{h}, k)$ by

$$\begin{aligned} \mathbf{E}(X_k) =& \sum_{\boldsymbol{h}\in\Omega} \binom{n}{h_0, h_1, \ldots, h_K} \prod_{\boldsymbol{x}\neq 0} \left(2 \sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\} \\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} \frac{h_{\boldsymbol{u}}}{n}\frac{h_{\boldsymbol{v}}}{n} \right)^{h_{\boldsymbol{x}}} \left(\sum_{\boldsymbol{u}} \left(\frac{h_{\boldsymbol{u}}}{n}\right)^2 \right)^{h_0} && (23) \\ =& \sum_{\boldsymbol{h}\in\Omega} \binom{n}{h_0, h_1, \ldots, h_K} \Phi(\boldsymbol{h}, k). && (24) \end{aligned}$$

Explanation of (23). Let $h_{\boldsymbol{x}} = |I_{\boldsymbol{x}}|$. The multinomial coefficient $\binom{n}{h_0,h_1,\ldots,h_K}$ counts the number of choices for the subsets $I_{\boldsymbol{x}}$. In the product, in order for $B_1, \ldots, B_k$ to be zero-sum, for $\boldsymbol{x} \neq 0$ we need to cancel the diagonal entries $M_{j,j} = 1$ of $j \in I_x$ within the columns indexed by $I_{\boldsymbol{x}}$. This is achieved by putting one entry in rows $I_{\boldsymbol{u}}$ and one in rows $I_{\boldsymbol{v}}$ where $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{x}$. The last factor counts the choices for the entries of columns indexed by I_0 over the row index sets $I_{\boldsymbol{u}}$, either zero or two in an index set, in order to preserve the zero-sum property.

Set $h_{\boldsymbol{x}} = (n/2^k)(1 + \varepsilon_{\boldsymbol{x}})$ where $|\varepsilon_{\boldsymbol{x}}| = O(\sqrt{\log n/n})$. We note that $\sum_{\boldsymbol{x}} \varepsilon_{\boldsymbol{x}} = 0$, implies that

$$\sum_{\boldsymbol{x}} h_{\boldsymbol{x}} \varepsilon_{\boldsymbol{x}} = \frac{n}{2^k} \sum_{\boldsymbol{x}} (\varepsilon_{\boldsymbol{x}} + \varepsilon_{\boldsymbol{x}}^2) = \frac{n}{2^k} \sum_{\boldsymbol{x}} \varepsilon_{\boldsymbol{x}}^2 \text{ and } \sum_{\boldsymbol{x}} h_{\boldsymbol{x}} \varepsilon_{\boldsymbol{x}}^2 = \frac{n}{2^k} \sum_{\boldsymbol{x}} \varepsilon_{\boldsymbol{x}}^2 + O\left(\frac{\log^{3/2} n}{n^{1/2}}\right).$$

And then Stirling's approximation implies that

$$\begin{aligned}\binom{n}{h_0, h_1, \ldots, h_K} &\sim \frac{n^n \sqrt{2\pi n}}{\prod_{\boldsymbol{x} \in \{0,1\}^k} ((n/2^k)(1 + \varepsilon_{\boldsymbol{x}}))^{h_{\boldsymbol{x}}} (\sqrt{2\pi n/2^k})^{2^k}} \\ &= 2^{kn} \exp\left\{ -\sum_{\boldsymbol{x} \in \{0,1\}^k}^{K} h_{\boldsymbol{x}} \left(\varepsilon_{\boldsymbol{x}} - \frac{\varepsilon_{\boldsymbol{x}}^2}{2} \right) + O(\log n) \right\} \\ &= 2^{kn} \exp\left\{ -\frac{n}{2^{k+1}} \sum_{\boldsymbol{x} \in \{0,1\}^k}^{K} \varepsilon_{\boldsymbol{x}}^2 + O(\log n) \right\} = 2^{kn} n^{O(1)}.\end{aligned}$$

In addition, by considering random 2^k-colorings of $[n]$ we see from the Chernoff bounds that

$$\sum_{\boldsymbol{h} \in \Omega} \binom{n}{h_0, h_1, \ldots, h_K} = 2^{kn}(1 - O(n^{-2^k/3})). \tag{25}$$

With respect to (23), using $\sum_{\boldsymbol{x}} \varepsilon_{\boldsymbol{x}} = 0$, we see that

$$\begin{aligned}\left(\sum_{\boldsymbol{u} \in \{0,1\}^k} \left(\frac{h_{\boldsymbol{u}}}{n} \right)^2 \right)^{h_0} &= \left(\sum_{\boldsymbol{u}} \frac{1}{2^{2k}} (1 + 2\varepsilon_{\boldsymbol{u}} + \varepsilon_{\boldsymbol{u}}^2) \right)^{h_0} \\ &= \left(\frac{1}{2^k} \right)^{h_0} \left(1 + \frac{1}{2^k} \sum_{\boldsymbol{u}} \varepsilon_{\boldsymbol{u}}^2 \right)^{h_0} \\ &= \left(\frac{1}{2^k} \right)^{h_0} \exp\left\{ \frac{n}{2^k} (1 + \varepsilon_0) \log \left(1 + \sum_{\boldsymbol{u}} \frac{\varepsilon_{\boldsymbol{u}}^2}{2^k} \right) \right\} \\ &= \left(\frac{1}{2^k} \right)^{h_0} \exp\left\{ \frac{n}{2^{2k}} \sum_{\boldsymbol{u}} \varepsilon_{\boldsymbol{u}}^2 + O\left(\frac{\log^{3/2} n}{n^{1/2}} \right) \right\}.\end{aligned} \tag{26}$$

If $\boldsymbol{x} \neq 0$ then each index $\boldsymbol{z}$ occurs exactly once in $\sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\}\\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} (\varepsilon_{\boldsymbol{u}}+\varepsilon_{\boldsymbol{v}})$ and so $\sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\}\\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} (\varepsilon_{\boldsymbol{u}}+\varepsilon_{\boldsymbol{v}}) = \sum_{\boldsymbol{z}} \varepsilon_{\boldsymbol{z}} = 0$. Therefore,

$$\begin{aligned}
\left(2 \sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\}\\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} \frac{h_{\boldsymbol{u}}}{n}\frac{h_{\boldsymbol{v}}}{n}\right)^{h_{\boldsymbol{x}}} &= \left(2 \sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\}\\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} \frac{1}{2^{2k}}(1+\varepsilon_{\boldsymbol{u}}+\varepsilon_{\boldsymbol{v}}+\varepsilon_{\boldsymbol{u}}\varepsilon_{\boldsymbol{v}})\right)^{h_{\boldsymbol{x}}} \\
&= \left(\frac{1}{2^k}\right)^{h_{\boldsymbol{x}}} \left(1+\frac{1}{2^k}\sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\}\\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} 2\varepsilon_{\boldsymbol{u}}\varepsilon_{\boldsymbol{v}}\right)^{h_{\boldsymbol{x}}} \\
&= \left(\frac{1}{2^k}\right)^{h_{\boldsymbol{x}}} \exp\left\{\frac{n}{2^k}(1+\varepsilon_{\boldsymbol{x}})\log\left(1+2\sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\}\\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} \frac{\varepsilon_{\boldsymbol{u}}\varepsilon_{\boldsymbol{v}}}{2^k}\right)\right\} \\
&= \left(\frac{1}{2^k}\right)^{h_{\boldsymbol{x}}} \exp\left\{\frac{n}{2^k}\sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\}\\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} \frac{2\varepsilon_{\boldsymbol{u}}\varepsilon_{\boldsymbol{v}}}{2^k} + O\left(\frac{\log^{3/2} n}{n^{1/2}}\right)\right\}.
\end{aligned}$$

Note that

$$\Lambda = \sum_{\boldsymbol{x}\neq 0}\sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\}\\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} 2\varepsilon_{\boldsymbol{u}}\varepsilon_{\boldsymbol{v}} = \sum_{\boldsymbol{u}} \varepsilon_{\boldsymbol{u}} \sum_{\substack{\boldsymbol{x}+\boldsymbol{u}\\ \boldsymbol{x}\neq 0}} \varepsilon_{\boldsymbol{x}+\boldsymbol{u}} = \sum_{\boldsymbol{u}} \varepsilon_{\boldsymbol{u}} \sum_{\boldsymbol{v}\neq\boldsymbol{u}} \varepsilon_{\boldsymbol{v}},$$

gives

$$\Lambda + \sum_{\boldsymbol{u}} \varepsilon_{\boldsymbol{u}}^2 = \left(\sum_{\boldsymbol{u}} \varepsilon_{\boldsymbol{u}}\right)^2 = 0.$$

Thus using $\sum_{\boldsymbol{x}} h_{\boldsymbol{x}} = n$,

$$\begin{aligned}
\Phi(\boldsymbol{h},k) &= \left(\frac{1}{2^k}\right)^{\sum_{\boldsymbol{x}} h_{\boldsymbol{x}}} \exp\left\{\frac{n}{2^{2k}}\left(\sum_{\boldsymbol{u}}\varepsilon_{\boldsymbol{u}}^2 + \sum_{\boldsymbol{x}\neq 0}\sum_{\substack{\{\boldsymbol{u},\boldsymbol{v}\}\\ \boldsymbol{u}+\boldsymbol{v}=\boldsymbol{x}}} 2\varepsilon_{\boldsymbol{u}}\varepsilon_{\boldsymbol{v}}\right) + O\left(\frac{\log^{3/2} n}{n^{1/2}}\right)\right\} \\
&= \frac{1}{2^{kn}} e^{O(\log^{3/2} n/\sqrt{n})}. \qquad (27)
\end{aligned}$$

It follows from (24), (25) and (27) above that

$$\mathbf{E}(X_k) = 1 + O\left(\frac{\log^{3/2} n}{\sqrt{n}}\right) = 1 + o(1). \qquad (28)$$

□

6 Conditional expected number of small zero-sum sets

Let $(B_1,\ldots,B_k)$ be a fixed sequence of subsets of $[n]$ with $|B_i|\in J_1$ for $i=1,2,\ldots,k\leq\omega$. Let $\mathcal{B}$ be the event

$$\mathcal{B}=\{(B_1,...B_k)\text{ is a simple sequence of large row dependencies}\}\,. \tag{29}$$

We need to understand the conditioning imposed by this event $\mathcal{B}$. Suppose that $|I_{\boldsymbol{x}}|=h_{\boldsymbol{x}}\sim n/2^k$ for $\boldsymbol{x}\in\{0,1\}^k$.

Lemma 15. *Given $\mathcal{B}$ and $i\in I_{\boldsymbol{x}}$, the distribution of the row indices k,ℓ of the other two non-zeros in column i is as follows: if $\boldsymbol{x}\neq 0$ then choose $\boldsymbol{u},\boldsymbol{v}$ such that $\boldsymbol{x}=\boldsymbol{u}+\boldsymbol{v}\bmod 2$ with probability*

$$p(\boldsymbol{u},\boldsymbol{v})=\frac{2h_{\boldsymbol{u}}h_{\boldsymbol{v}}}{\sum_{\boldsymbol{y}+\boldsymbol{z}=\boldsymbol{x}}h_{\boldsymbol{y}}h_{\boldsymbol{z}}},$$

and then randomly choose $k\in I_{\boldsymbol{u}},\ell\in I_{\boldsymbol{v}}$. If $\boldsymbol{x}=0$ then choose $\boldsymbol{u}$ with probability

$$p(\boldsymbol{u},\boldsymbol{u})=\frac{h_{\boldsymbol{u}}^2}{\sum_{\boldsymbol{y}\in\{0,1\}^k}h_{\boldsymbol{y}}^2},$$

and then randomly choose $k,\ell\in I_{\boldsymbol{u}}$.

Proof. This follows from the fact that the non-zeros in each column are independently chosen with replacement and from the condition given in Lemma 13. □

Let $(S_j,\ s_j=|S_j|\leq\omega,\ j=1,2,\ldots,\ell\leq\omega)$ be a sequence of pairwise disjoint small subsets of $[n]$ and $S=\bigcup_{j=1}^{\ell}S_j$ and $s=|S|$. We define the events

$$\mathcal{S}_j=\{S_j\text{ is a small zero-sum row set}\}\text{ for }j=1,2,\ldots,\ell\text{ and }\mathcal{S}=\bigcap_{j=1}^{\ell}\mathcal{S}_j.$$

$$\mathcal{S}_j^*=\{S_j\text{ is a small fundamental zero-sum row set}\}\text{ for }j=1,2,\ldots,\ell\text{ and }\mathcal{S}^*=\bigcap_{j=1}^{\ell}\mathcal{S}_j^*.$$

Lemma 16.

$$\mathbb{P}(\mathcal{S}^*\mid\mathcal{B})\sim\mathbb{P}(\mathcal{S}^*). \tag{30}$$

Proof. Let $I_{\boldsymbol{x}}$, $\boldsymbol{x}\in\{0,1\}^k$, be as defined in Section 4. Let $S_{\boldsymbol{x}}=S\cap I_{\boldsymbol{x}}$ and $J_{j,\boldsymbol{x}}=S_j\cap I_{\boldsymbol{x}}$ and $\ell_{j,\boldsymbol{x}}=|J_{j,\boldsymbol{x}}|$ for $i=1,2,\ldots,m$ and $J_{\boldsymbol{x}}=\bigcup_{j=1}^{m}J_{j,\boldsymbol{x}}$ and $\ell_{\boldsymbol{x}}=|J_{\boldsymbol{x}}|$. Let $J_{\boldsymbol{0}}=I_{\boldsymbol{0}}\setminus S$ and $\ell_{\boldsymbol{0}}=|S_{\boldsymbol{0}}|$. We now consider the probability that column i is consistent with $\mathcal{S}$. We let $h_{\boldsymbol{x}}=|I_{\boldsymbol{x}}|$ and $s_{\boldsymbol{x}}=|S_{\boldsymbol{x}}|$ for $\boldsymbol{x}\in\{0,1\}^k$.

Case 1: $i \in I_{\mathbf{0}} \setminus J_{\mathbf{0}}$**.** For each column $i \in I_{\mathbf{0}} \setminus J_{\mathbf{0}}$, the task here is to estimate the probability that the two non-zeros $e_1(i), e_2(i)$ are in rows consistent with the occurrence of $\mathcal{S}$. Because $i \in I_{\mathbf{0}}$ and $\mathcal{B}$ occurs, we know from Lemma 13 that $e_1(i), e_2(i) \in I_{\boldsymbol{u}}$ for some $\boldsymbol{u} \in \{0,1\}^k$. For $\mathcal{S}$ to occur, we require that zero or two of $e_1(i), e_2(i)$ fall in $J_{\boldsymbol{u}}$, an event of conditional probability $(1 - s_{\boldsymbol{u}}/h_{\boldsymbol{u}})^2 + (s_{\boldsymbol{u}}/h_{\boldsymbol{u}})^2$.

Let $E_{\boldsymbol{u}}$ denote the number of non-zero pairs from $I_{\mathbf{0}} \setminus J_{\mathbf{0}}$ falling in $J_{\boldsymbol{u}}$. Then the conditional probability that the non-zeros of $I_{\mathbf{0}} \setminus S_{\mathbf{0}}$ are consistent with $\mathcal{S}$ is given by

$$\mathbb{P}(I_{\mathbf{0}} \setminus S_{\mathbf{0}} \text{ is consistent } \mathcal{S} \mid \mathcal{B}) = \mathbf{E}\left(\prod_{\boldsymbol{u}} \left(1 - 2\frac{s_{\boldsymbol{u}}}{h_{\boldsymbol{u}}} + 2\left(\frac{s_{\boldsymbol{u}}}{h_{\boldsymbol{u}}}\right)^2\right)^{E_{\boldsymbol{u}}}\right) \tag{31}$$

Given $\mathcal{B}$, we see that $E_{\boldsymbol{u}}$ is distributed as $Bin(h_{\mathbf{0}} - s_{\mathbf{0}}, p(\boldsymbol{u}, \boldsymbol{u}))$, and has expectation

$$\mathbf{E}(E_{\boldsymbol{u}}) = (h_{\mathbf{0}} - s_{\mathbf{0}}) \frac{h_{\boldsymbol{u}}^2}{h_{\mathbf{0}}^2 + h_1^2 + \cdots + (h_{2^k-1})^2} \sim \frac{h_{\mathbf{0}}}{2^k}.$$

The Chernoff bounds imply that $E_{\boldsymbol{u}}$ is concentrated around its mean $(h_{\mathbf{0}} - s_{\mathbf{0}})p(\boldsymbol{u}, \boldsymbol{u}) \sim \frac{N}{2^k}$, where $N = n/2^k$. Thus,

$$\left| E_{\boldsymbol{u}} - \frac{h_{\mathbf{0}}}{2^k} \right| \leq n^{2/3} \quad \text{with probability at least } 1 - e^{-\Omega(n^{1/3})}. \tag{32}$$

Going back to (31) and using (32) we have

$$\mathbb{P}(I_{\mathbf{0}} \setminus S_{\mathbf{0}} \text{ is consistent with the occurrence of } \mathcal{S} \mid \mathcal{B}) \sim \prod_{\boldsymbol{u}} \left(1 - \frac{2s_{\boldsymbol{u}}}{N}\right)^{N/2^k} \sim \exp\left\{-2\sum_{\boldsymbol{u}} \frac{s_{\boldsymbol{u}}}{2^k}\right\} = e^{-s/2^{k-1}}. \tag{33}$$

Case 2: $i \in I_{\boldsymbol{x}} \setminus J_{\boldsymbol{x}}$**,** $\boldsymbol{x} \neq \mathbf{0}$**.** Given $\mathcal{B}$, and $i \in I_{\boldsymbol{x}}$, suppose that the non-zeros $e_1(i), e_2(i)$ of column i lie in $I_{\boldsymbol{u}}, I_{\boldsymbol{x}+\boldsymbol{u}}$ respectively, $\boldsymbol{u} \in \{0,1\}^k$. The probability of this is $p(\boldsymbol{u}, \boldsymbol{x} + \boldsymbol{u})$. The number $E_{\boldsymbol{x}}(\boldsymbol{u}, \boldsymbol{x} + \boldsymbol{u})$ of such pairs of non-zeros in $I_{\boldsymbol{u}}, I_{\boldsymbol{x}+\boldsymbol{u}}$ has distribution $Bin((h_{\boldsymbol{x}} - s_{\boldsymbol{x}})p(\boldsymbol{u}, \boldsymbol{x} + \boldsymbol{u}))$, and expectation asymptotic to $(h_{\boldsymbol{x}} - s_{\boldsymbol{x}})/2^{k-1}$.

The rows of $S_1, \ldots, S_\ell$ have to be zero-sum in this column, so either exactly one non-zero falls in each of $S_{j,\boldsymbol{u}}, S_{j,\boldsymbol{x}+\boldsymbol{u}}$ for some $1 \leq j \leq \ell$ or exactly one non-zero falls in each of

$I_{\boldsymbol{u}} \setminus S_{\boldsymbol{u}}, I_{\boldsymbol{x}+\boldsymbol{u}} \setminus S_{\boldsymbol{x}+\boldsymbol{u}}$. The probability of this is

$$\begin{aligned}
P(\boldsymbol{u}, \boldsymbol{x}+\boldsymbol{u}) = &\mathbf{E}\left(\left(\sum_{j=1}^{\ell} \frac{s_{j,\boldsymbol{u}}}{h_{\boldsymbol{u}}} \frac{s_{j,\boldsymbol{x}+\boldsymbol{u}}}{h_{\boldsymbol{x}+\boldsymbol{u}}} + \frac{h_{\boldsymbol{u}} - s_{\boldsymbol{u}}}{h_{\boldsymbol{u}}} \frac{h_{\boldsymbol{x}+\boldsymbol{u}} - s_{\boldsymbol{x}+\boldsymbol{u}}}{h_{\boldsymbol{x}+\boldsymbol{u}}}\right)^{E_{\boldsymbol{x}}(\boldsymbol{u},\boldsymbol{x}+\boldsymbol{u})}\right) \\
&\sim \left(\sum_{j=1}^{\ell} \frac{s_{j,\boldsymbol{u}} s_{j,\boldsymbol{x}+\boldsymbol{u}}}{N^2} + \frac{N - s_{\boldsymbol{u}}}{N} \frac{N - s_{\boldsymbol{x}+\boldsymbol{u}}}{N}\right)^{(N - s_{\boldsymbol{x}})/2^{k-1}} \\
&\sim e^{-(s_{\boldsymbol{u}} + s_{\boldsymbol{x}+\boldsymbol{u}})/2^{k-1}}.
\end{aligned}$$

For a given $\boldsymbol{x}$ there are 2^{k-1} unordered pairs $S_{\boldsymbol{u}}, S_{\boldsymbol{x}+\boldsymbol{u}}$, so

$$\mathbb{P}(I_{\boldsymbol{x}} \setminus S_{\boldsymbol{x}} \text{ is consistent with } \mathcal{S}) \sim \exp\left\{-\frac{1}{2^{k-1}} \sum_{\{u,\boldsymbol{x}+\boldsymbol{u}\}} (s_{\boldsymbol{u}} + s_{\boldsymbol{x}+\boldsymbol{u}})\right\} = e^{-s/2^{k-1}}. \tag{34}$$

(In the sum in (34) $s_{\boldsymbol{u}} + s_{\boldsymbol{x}+\boldsymbol{u}}$ and $s_{\boldsymbol{x}+\boldsymbol{u}} + s_{\boldsymbol{u}}$ contribute as one term.) Thus

$$\mathbb{P}(I_{\boldsymbol{x}} \setminus S_{\boldsymbol{x}} \text{ is consistent with } \mathcal{S}, \forall \boldsymbol{x} \neq \mathbf{0}) \sim e^{-(2^k-1)s/2^{k-1}}. \tag{35}$$

Case 3: $i \in S_{j,\boldsymbol{x}} \subseteq I_{\boldsymbol{x}}$, $\boldsymbol{x} \neq 0$. For $i \in S_{j,\boldsymbol{x}}$, one non-zero needs to be in S_j, and the other to avoid S_j. Let $\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{u}$. Suppose that the pair $e_1(i)$, $e_2(i)$ fall in $I_{\boldsymbol{u}}, I_{\boldsymbol{u}+\boldsymbol{x}}$. The probability this happens is

$$P_j(\boldsymbol{u}, \boldsymbol{v}) \sim \frac{1}{2^{k-1}} \left(\frac{s_{j,\boldsymbol{u}}}{h_{\boldsymbol{u}}} \frac{h_{\boldsymbol{v}} - s_{j,\boldsymbol{v}}}{h_{\boldsymbol{v}}} + \frac{s_{j,\boldsymbol{v}}}{h_{\boldsymbol{v}}} \frac{h_{\boldsymbol{u}} - s_{j,\boldsymbol{u}}}{h_{\boldsymbol{u}}}\right). \tag{36}$$

The events $\{\boldsymbol{u}, \boldsymbol{x}+\boldsymbol{u}\}$ are disjoint and exhaustive, so for a given $i \in S_{j,\boldsymbol{x}}$ the probability $p(i,j)$ of success (i.e. the S_j-indexed rows of column i sum to zero) is

$$\begin{aligned}
p(i,j) = \sum_{\{\boldsymbol{u},\boldsymbol{u}+\boldsymbol{x}\}} P_j(\boldsymbol{u}, \boldsymbol{u}+\boldsymbol{x}) \sim \frac{1}{2^{k-1}} &\sum_{\boldsymbol{u},\boldsymbol{v}=\boldsymbol{x}+\boldsymbol{u}} \left(\frac{s_{j,\boldsymbol{u}}}{N} \frac{N - s_{j,\boldsymbol{v}}}{N} + \frac{s_{j,\boldsymbol{v}}}{N} \frac{N - s_{j,\boldsymbol{u}}}{N}\right) \\
&\sim \frac{s_j}{N 2^{k-1}} \left(1 + O\left(\frac{\omega}{N}\right)\right).
\end{aligned} \tag{37}$$

Every column of $S_{j,\boldsymbol{x}}$ has to succeed or S_j is not a small zero-sum set. Thus

$$\mathbb{P}(S_{\boldsymbol{x}} \text{ succeeds}) \sim \left(\frac{s_j(1 + O(s/N))}{N 2^{k-1}}\right)^{s_{j,\boldsymbol{x}}}.$$

As $\sum s_{j,\boldsymbol{x}} = s_j$,

$$\mathbb{P}(S_{\boldsymbol{x}} \text{ succeeds } \forall \boldsymbol{x}) \sim \left(\frac{s_j}{N 2^{k-1}}\right)^{s_j - s_{j,\mathbf{0}}}. \tag{38}$$

Case 4: $i \in S_{j,\mathbf{0}} \subseteq I_{j,\mathbf{0}}$**.** In the case that $\boldsymbol{x} = \mathbf{0}$, and $S_{j,\mathbf{0}} \subseteq I_{j,\mathbf{0}}$, the non-zeros in a column of $S_{j,\mathbf{0}}$ must both fall in the same index set $I_{\boldsymbol{u}}$; one in $S_{j,\boldsymbol{u}}$ and one in $I_{\boldsymbol{u}} \setminus S_{j,\boldsymbol{u}}$. Thus $P(\boldsymbol{u}, \boldsymbol{u})$ is now summed over all $I_{\boldsymbol{u}}$, a total of 2^k such sets. For $i \in S_{j,\mathbf{0}}$, the probability $p(i)$ of success is

$$p(i) = \sum_{\{\boldsymbol{u},\boldsymbol{u}\}} P(\boldsymbol{u}, \boldsymbol{u}) \sim \frac{1}{2^k} \sum_{\boldsymbol{u}} \left(2 \frac{s_{j,\boldsymbol{u}}}{N} \frac{N - s_{j,\boldsymbol{u}}}{N} \right) \sim \frac{s_j}{N 2^{k-1}} \left(1 + O\left(\frac{\omega}{N} \right) \right).$$

The final term is the same as in (37), and we obtain

$$\mathbb{P}(S_{j,\mathbf{0}} \text{ succeeds}) \sim \left(\frac{s_j}{N 2^{k-1}} \right)^{s_{j,\mathbf{0}}} \tag{39}$$

Using (33), (35), (38) and (39), we obtain

$$\mathbb{P}(\mathcal{S} \mid \mathcal{B}) \sim \prod_{j=1}^{m} \left(\frac{s_j}{N 2^{k-1}} \right)^{s_j} e^{-(2^k - 1)s/2^{k-1}} e^{-s/2^{k-1}} = \prod_{j=1}^{m} \left(\frac{2 s_j}{n} \right)^{s_j} e^{-2s}, \tag{40}$$

after using (7). This completes the proof of $\mathbb{P}(\mathcal{S} \mid \mathcal{B}) \sim \mathbb{P}(\mathcal{S})$. To replace $\mathcal{S}$ by $\mathcal{S}^*$ we just need to let $K_j, j = 1, 2, \ldots, m$ denote the set of i in Case 3 where $i \in S_{j,\boldsymbol{x}}$. We see from (36) that the positions of the non-zeros in the columns K_j are asymptotically uniform over S_j. This is because each $k \in J_{j,\boldsymbol{u}}$ is chosen with probability asymptotic to $\frac{1}{s_{j,\boldsymbol{u}}} \cdot \frac{s_{j,\boldsymbol{u}}}{h_{\boldsymbol{u}}}$ and similarly for $k \in J_{j,\boldsymbol{x}+\boldsymbol{u}}$. In which case, the conditional probability that S_j is fundamental is obtained by multiplying by κ_{s_j}. This completes the proof of the lemma. □

We can now use inclusion-exclusion to prove

Lemma 17. *Let Σ_σ be the event that there are exactly σ disjoint small fundamental dependencies. Then,*

$$\mathbb{P}(\Sigma_\sigma \mid \mathcal{B}) \sim \frac{\phi_R^\sigma e^{-\phi_R}}{\sigma!} \sim \mathbb{P}(\Sigma_\sigma).$$

Proof. Let

$$T_\ell = \frac{1}{\ell!} \sum_{1 \leq s_1, \ldots, s_\ell \leq \omega} \sum_{|S_i| = s_i, i=1,\ldots,\ell} \mathbb{P}\left(\bigcap_{i=1}^{\ell} \mathcal{S}_i^* \middle| \mathcal{B} \right) \sim \frac{1}{\ell!} \sum_{1 \leq s_1, \ldots, s_\ell \leq \omega} \sum_{|S_i| = s_i, i=1,\ldots,\ell} \mathbb{P}\left(\bigcap_{i=1}^{\ell} \mathcal{S}_i^* \right) \sim$$
$$\frac{1}{\ell!} \sum_{1 \leq s_1, \ldots, s_\ell \leq \omega} \binom{n}{s_1, \ldots, s_\ell} \prod_{i=1}^{\ell} \left(\frac{2 s_i}{n} \right)^{s_i} e^{-2 s_i} \kappa_{s_i} \sim \frac{1}{\ell!} \sum_{1 \leq s_1, \ldots, s_\ell \leq \omega} \prod_{i=1}^{\ell} \frac{(2 s_i)^{s_i}}{s_i!} e^{-2 s_i} \kappa_{s_i}$$
$$\sim \frac{1}{\ell!} \left(\sum_{s=1}^{\infty} \frac{(2e^{-2})^s}{s} \sigma_s \right)^{\ell} \sim \frac{\phi_R^\ell}{\ell!}.$$

The first approximation follows from Lemma 16 and the second from (7), (8).

Using Inclusion-Exclusion, we have

$$\mathbb{P}(\Sigma_\sigma \mid \mathcal{B}) = \sum_{\ell \geq \sigma} (-1)^{k-\sigma} \binom{\ell}{\sigma} T_\ell \sim \sum_{\ell \geq \sigma} (-1)^{\ell-\sigma} \binom{\ell}{\sigma} \frac{\phi_R^\ell}{\ell!} = \frac{\phi_R^\sigma e^{-\phi_R}}{\sigma!}.$$

Lemma 10 gives us the unconditional probability. □

Let X_k count the number of simple k-sequences as in Lemma 14.

Lemma 18. *If $\sigma = O(1)$ then $\mathbf{E}(X_k \mid \Sigma_\sigma) \sim 1$.*

Proof.

$$\begin{aligned}
\mathbf{E}(X_k \mid \Sigma_\sigma) &= \sum_{\mathcal{B}=(B_1,\dots,B_k)} \mathbb{P}(\mathcal{B} \mid \Sigma_\sigma) \\
&= \sum_{\mathcal{B}=(B_1,\dots,B_k)} \frac{\mathbb{P}(\Sigma_\sigma \mid \mathcal{B})\mathbb{P}(\mathcal{B})}{\mathbb{P}(\Sigma_\sigma)} \\
&= \sum_{\mathcal{B}=(B_1,\dots,B_k)} \frac{\mathbb{P}(\mathcal{B})}{\mathbb{P}(\Sigma_\sigma)} \sum_{\ell \geq \sigma} (-1)^{\ell-\sigma} \binom{k}{\sigma} T_\ell \\
&= \sum_{\mathcal{B}=(B_1,\dots,B_k)} \frac{\mathbb{P}(\mathcal{B})}{\mathbb{P}(\Sigma_\sigma)} \sum_{\ell \geq \sigma} (-1)^{\ell-\sigma} \binom{k}{\sigma} \frac{1}{\ell!} \sum_{1 \leq s_1,\dots,s_\ell \leq \omega} \sum_{|S_i|=s_i, i=1,\dots,\ell} \mathbb{P}\left(\bigcap_{i=1}^{\ell} \mathcal{S}_i^* \middle| \mathcal{B} \right) \\
&\sim \sum_{\mathcal{B}=(B_1,\dots,B_k)} \frac{\mathbb{P}(\mathcal{B})}{\mathbb{P}(\Sigma_\sigma)} \sum_{\ell \geq \sigma} (-1)^{\ell-\sigma} \binom{k}{\sigma} \frac{1}{\ell!} \sum_{1 \leq s_1,\dots,s_\ell \leq \omega} \sum_{|S_i|=s_i, i=1,\dots,\ell} \mathbb{P}\left(\bigcap_{i=1}^{\ell} \mathcal{S}_i^* \right) \\
&\sim \sum_{\mathcal{B}=(B_1,\dots,B_k)} \frac{\mathbb{P}(\mathcal{B})}{\mathbb{P}(\Sigma_\sigma)} \mathbb{P}(\Sigma_\sigma) \\
&= \mathbf{E}(X_k) \sim 1.
\end{aligned}$$

□

7 Joint distribution of small and large dependencies

7.1 $P_n(0,d)$: the case of no small fundamental dependencies.

Let $P_n(0,d)$ be the probability that $M \in \boldsymbol{M}(n)$ has no small fundamental dependencies and the maximum number of large simple dependencies is d. Let $\pi(d)$ be given by (2). The purpose of this section is to prove the following.

$$P_n(0,d) \sim \pi(d)\, e^{-\phi}. \tag{41}$$

Let V be the vector space generated by the dependencies. Let $\mathcal{L}_\lambda$ be the event that the dimension of V is λ. Let

$$p(0,\lambda) = \mathbb{P}(\Sigma_0 \wedge \mathcal{L}_\lambda) \text{ and } p(0) = \mathbb{P}(\Sigma_0).$$

Lemma 19. *For $0 \leq \lambda = O(1)$, $p(0,\lambda) \sim P(0,\lambda)$ where $P(0,\lambda) = \pi(\lambda)\, e^{-\phi}$.*

Proof. For $0 \leq k = O(1)$, we have from Lemma 18 that

$$1 \sim \mathbf{E}(X_k \mid \Sigma_0) = \sum_{\lambda \geq k} \mathbf{E}(X_k \mid \Sigma_0 \wedge \mathcal{L}_\lambda) p(\lambda \mid 0), \tag{42}$$

where $p(\lambda \mid 0) = p(0,\lambda)/p(0)$. Let $\mathcal{H}$ be the event that there exists a set of dependent rows H where $\omega \leq |H| \notin J_1$. Then we have

$$\begin{aligned}\mathbf{E}(X_k \mid \Sigma_0 \wedge \mathcal{L}_\lambda) &= \mathbf{E}(X_k \mid \Sigma_0, \wedge\mathcal{L}_\lambda \wedge \neg\mathcal{H})\mathbb{P}(\neg\mathcal{H}) + \mathbf{E}(X_k \mid \Sigma_0 \wedge \mathcal{L}_\lambda \wedge \mathcal{H})\mathbb{P}(\mathcal{H}) \\ &\sim \prod_{i=0}^{k-1}(2^\lambda - 2^i). \end{aligned} \tag{43}$$

Justification for (43): Given $\Sigma_0 \wedge \mathcal{L}_\lambda \wedge \neg\mathcal{H}$ there are 2^λ vectors in V. Choosing i members of a simple sequence generates a subspace of dimension i, and we eliminate 2^i vectors from consideration as the next member of the sequence. Given $\neg\mathcal{H}$ the number of simple sequences is given by the RHS of (43). Equation (43) then follows from $\mathbb{P}(\mathcal{H}) = o(1)$.

It follows from (43) that for $\lambda \geq 0$,

$$1 \sim \sum_{\lambda=k}^{\infty} \frac{p(0,\lambda)}{p(0)} \prod_{i=0}^{k-1}(2^\lambda - 2^i). \tag{44}$$

The asymptotic solution of (44) is given by the following lemma.

Lemma 20. *For $\lambda \geq 0$, the solutions to*

$$1 = \sum_{\lambda=k}^{\infty} q_\lambda \prod_{i=0}^{k-1} (2^\lambda - 2^i), \qquad k \geq 0. \tag{45}$$

are given by $q_\lambda = \pi(\lambda)$ of (2).

Proof. Gaussian coefficients are defined as

$$\begin{bmatrix} \lambda \\ k \end{bmatrix}_z = \frac{\prod_{i=1}^{k} (z^{\lambda-i+1} - 1)}{\prod_{i=1}^{k} (z^i - 1)}. \tag{46}$$

Using (46) with $z = 2$, equation (45) can be rewritten as

$$1 = 2^{\binom{k}{2}} \prod_{i=1}^{k} (2^i - 1) \sum_{\lambda=k}^{\infty} q_\lambda \begin{bmatrix} \lambda \\ k \end{bmatrix}_2. \tag{47}$$

Put $\psi_k = 1/\left(2^{\binom{k}{2}} \prod_{i=1}^{k} (2^i - 1)\right)$, we see that q_λ is the solution to

$$\sum_{\lambda=k}^{\infty} \begin{bmatrix} \lambda \\ k \end{bmatrix}_2 q_\lambda = \psi_k, \qquad k \geq 0. \tag{48}$$

Fix $\delta \geq 0$, multiply equation $k \geq \delta$ in (48) by $(-1)^{k-\delta} 2^{\binom{k-\delta}{2}} \left[\begin{smallmatrix} k \\ \delta \end{smallmatrix}\right]_2$, and sum these equations over $k \geq \delta$. This gives

$$\begin{aligned} \sum_{k=\delta}^{\infty} (-1)^{k-\delta} 2^{\binom{k-\delta}{2}} \begin{bmatrix} k \\ \delta \end{bmatrix}_2 \psi_k &= \sum_{k=\delta}^{\infty} \sum_{\lambda=k}^{\infty} (-1)^{k-\delta} \begin{bmatrix} k \\ \delta \end{bmatrix}_2 2^{\binom{k-\delta}{2}} \begin{bmatrix} \lambda \\ k \end{bmatrix}_2 q_\lambda && (49) \\ &= \sum_{k=\delta}^{\infty} \sum_{\lambda=k}^{\infty} (-1)^{k-\delta} \begin{bmatrix} \lambda-\delta \\ k-\delta \end{bmatrix}_2 2^{\binom{k-\delta}{2}} \begin{bmatrix} \lambda \\ \delta \end{bmatrix}_2 q_\lambda \\ &= \sum_{\lambda=\delta}^{\infty} \begin{bmatrix} \lambda \\ \delta \end{bmatrix}_2 q_\lambda \sum_{k=\delta}^{\lambda} (-1)^{k-\delta} \begin{bmatrix} \lambda-\delta \\ k-\delta \end{bmatrix}_2 2^{\binom{k-\delta}{2}} && (50) \\ &= q_\delta. && (51) \end{aligned}$$

Explanation: (50) **to** (51)**:** Gaussian coefficients satisfy the identity

$$(1+x)(1+zx)\cdots(1+z^{r-1}x) = \sum_{\ell=0}^{r} \begin{bmatrix} r \\ \ell \end{bmatrix}_z z^{\binom{\ell}{2}} x^\ell. \tag{52}$$

To prove the last summation on the right hand side of (50) is zero for $\lambda > \delta$, use (52) with $x = -1, z = 2$, $\ell = k - \delta$ and $r = \lambda - \delta$. This gives $\sum_{\ell=0}^{\lambda-\delta} \left[\begin{smallmatrix}\lambda-\delta\\ \ell\end{smallmatrix}\right]_2 2^{\binom{\ell}{2}}(-1)^{\ell} = 0$ for $\lambda > \delta$. For $z < 1$, taking the limit of (52) gives

$$\prod_{\ell=0}^{\infty}(1 + z^{\ell}x) = \sum_{\ell=0}^{\infty} \frac{z^{\binom{\ell}{2}} x^{\ell}}{\prod_{i=1}^{\ell}(1 - z^{i})}. \tag{53}$$

Replacing δ by λ in equation (49), we see that the solution q_λ to (45) is

$$\begin{aligned} q_\lambda &= \sum_{k=\lambda}^{\infty} \frac{(-1)^{k-\lambda} 2^{\binom{k-\lambda}{2}-\binom{k}{2}}}{\prod_{i=0}^{\lambda-1}(2^{\lambda-i}-1)\prod_{i=\lambda}^{k-1}(2^{k-i}-1)} \\ &= \frac{\left(\frac{1}{2}\right)^{\lambda^2}}{\prod_{i=1}^{\lambda}\left(1-\left(\frac{1}{2}\right)^{i}\right)} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell}\left(\frac{1}{2}\right)^{\binom{\ell}{2}}\left(\frac{1}{2}\right)^{(1+\lambda)\ell}}{\prod_{i=1}^{\ell}\left(1-\left(\frac{1}{2}\right)^{i}\right)} && (54) \\ &= \left(\frac{1}{2}\right)^{\lambda^2} \frac{\prod_{i=\lambda+1}^{\infty}\left(1-\left(\frac{1}{2}\right)^{i}\right)}{\prod_{i=1}^{\lambda}\left(1-\left(\frac{1}{2}\right)^{i}\right)} = \pi(\lambda), && (55) \end{aligned}$$

where $\pi(\lambda)$ is given in (2). To get from (54) to (55), use (53) with $z = 1/2$ and $x = (-1/2^{\lambda+1})$. □

The $p(0, \lambda)$ only satisfy (45) asymptotically and so to prove the lemma, we show that for large K,

$$\sum_{\substack{\lambda \geq K \\ \sigma \geq 0}} q_\lambda \leq \varepsilon, \tag{56}$$

where $\varepsilon > 0$ is arbitrarily small. Now,

$$\prod_{i=0}^{k-1}(2^{\lambda} - 2^{i}) = 2^{k\lambda}\prod_{i=0}^{k-1}\left(1 - \frac{1}{2^{\lambda-i}}\right) \geq 2^{k\lambda}\left(1 - \sum_{i=0}^{k-1}\frac{1}{2^{\lambda-i}}\right) \geq 2^{(k-1)\lambda}.$$

It follows that

$$\sum_{\substack{\lambda \geq K \\ \sigma \geq 0}} q_\lambda \leq 2^{-K(K-1)}.$$

Thus (56) holds if $K \geq \sqrt{2\log_2 1/\varepsilon}$. □

7.2 $P_n(1,d)$: the case of one small fundamental dependency.

Introduce the notation $P_n([m,d])$ for the probability that $M \in \boldsymbol{M}(n)$ has exactly m small fundamental dependencies and the maximum number of large simple dependencies is d. Thus there are small dependencies $D_1,...,D_m$ and (not necessarily unique) large dependencies $B_1,...,B_d$ corresponding to M having a null space of dimension $m+d$. In the case $m=0$, it follows from (41) that $P_n(0,d) \sim \pi(d)\, e^{-\phi}$.

Before considering $P_n(m,d)$, we explain the basic principle by deriving $P_n(1,d)$. The general case will follow from the recursive application of this.

Let $M \in \boldsymbol{M}([n])$ and let L be a fixed set of rows, $|L| = \ell$. We write

$$M = \begin{pmatrix} S_L & R \\ C & M' \end{pmatrix}.$$

Here S_L is $\ell \times \ell$, R is $\ell \times (n-\ell)$, C is $(n-\ell)\times \ell$ and M' has rows and columns indexed by $[n]-L$.

The event $R = \mathbf{0}$, is dependent only on the columns of $[n]-L$ in M. Provided $\ell = o(n^{1/2})$, $R = \mathbf{0}$ has probability

$$\mathbb{P}(R=\mathbf{0}) = \left(1 - \frac{\ell}{n}\right)^{2(n-\ell)} \sim e^{-2\ell}.$$

Given $R = \mathbf{0}$, M' is a uar element of $\boldsymbol{M}([n]-L)$. This follows directly from M is a uar element of $\boldsymbol{M}([n])$. Each column of M has 2 random entries, and these are not in the rows of L. At this point

$$M = \begin{pmatrix} S_L & 0 \\ C & M' \end{pmatrix}. \tag{57}$$

Independently of what happens in the columns of $[n]-L$ in M, the event that within the columns of L the sub-matrix S_L is the vertex-edge incidence matrix of a connected random mapping D_L, and thus each column of the sub-matrix C has one uar entry, is (see Section 2)

$$P(D_L) \sim \left(\frac{2\ell}{n}\right)^{\ell} \cdot \frac{(\ell-1)!}{\ell^\ell}\sigma_\ell.$$

The probability $P_{n-\ell}(0,k) \sim e^{-\phi}\pi(k)$ that M' has no small dependencies and k large ones is given by (41) above. Let $P^*(j,k;1)$ be the probability that exactly j of the k large dependencies of M' remain as dependencies after adding back the sub-matrix C. To maintain continuity of exposition, the analysis of this event is deferred until Section 7.4. Equation (63) of Section 7.4 with $m=1$, gives

$$P^*(j,k;1) \sim \binom{k}{j}\left(\frac{1}{2}\right)^j \left(1-\frac{1}{2}\right)^{k-j} = \binom{k}{j}\left(\frac{1}{2}\right)^k.$$

Let

$$P_n(1,j],L) = \mathbb{P}(M \text{ has 1 small fundamental dependency } L \text{ and } j \text{ large dependencies}),$$

Thus using (41), and the above

$$P_n(1,j],L) \sim \left(\frac{2}{n}\right)^{\ell} (\ell-1)!\sigma_\ell \cdot e^{-2\ell} \cdot \sum_{k\geq j} \binom{k}{j} \left(\frac{1}{2}\right)^{k} \pi(k) e^{-\phi}.$$

The probability that L is dependent, but $R \neq \mathbf{0}$ is $O(\ell^2/n)$. The events that L is the unique fundamental dependency are exclusive and exhaustive, so $\mathbb{P}([1,j])$ is the sum of these. Thus, summing over $\binom{n}{\ell}$ for $\ell \geq 1$ gives

$$P_n(1,j]) \sim \phi\, e^{-\phi} \cdot \sum_{k\geq j} \binom{k}{j} \left(\frac{1}{2}\right)^{k} \cdot \pi(k).$$

7.3 The general case of $\mathrm{null}(M) = d$, with m small fundamental dependencies

The matrix M' in (57) is a uar element of $\boldsymbol{M}([n]-L)$, and we can repeat the above construction with M' instead of M. We remove a set of columns L' and conditional on $R' = \mathbf{0}$, the sub-matrix M'' is a uar element of $\boldsymbol{M}([n]-L-L')$. In this way we can obtain the probability $\mathbb{P}([2,j])$ of two small and j large dependencies, and so on.

To systematize this, let $M_0 = M, L_0 = L, R_0 = R, n_0 = n, \ell_0 = |L|$ and let $M_1 = M', n_1 = n_0 - \ell_0$. Thus, M_0 is a uar element of $\boldsymbol{M}(n_0)$ and with some relabelling of $[n]-L$, M_1 is a uar element of $\boldsymbol{M}(n_1)$, etc.

In this way, we remove a sequence $(L_0, L_1, ..., L_{m-1})$ of column sets, of total size at most $m\omega$. As $n - m\omega \sim n$, the result (41) holds in $\boldsymbol{M}(n_m)$ with the same asymptotic probability. Taking the subspace $\boldsymbol{M}([0,k], n_m)$ of $M(n_m)$, we work back to the subspace of $\boldsymbol{M}$ with small fundamental dependencies $L_0, ..., L_m$ and $j \leq k$ large dependencies, and thus to $P_n(m,j])$, the probability of $\boldsymbol{M}([m,j],n)$.

Summarizing the necessary generalizations, we have

$$\mathbb{P}(R(L_j) = \mathbf{0},\ j = 0, ..., m-1) \sim \prod_{j=0}^{m-1} e^{-2\ell_j}, \tag{58}$$

$$P(D_{L_j}, j = 0, ..., m-1) \sim \prod_{j=0}^{m-1} \left(\frac{2\ell_j}{n_j}\right)^{\ell_j} \cdot \frac{(\ell_j - 1)!}{\ell_j^{\ell_j}} \sigma_{\ell_j}, \tag{59}$$

$$P^*(j, k; m) \sim \binom{k}{j} \left(\frac{1}{2^m}\right)^j \left(1 - \frac{1}{2^m}\right)^{k-j}, \tag{60}$$

$$P_{n-\ell}(0, k]) \sim e^{-\phi}\, \pi(k). \tag{61}$$

The last line is (41). For continuity of exposition, the proof of (60) is deferred until Theorem 21 in Section 7.4 below.

The dependency of probability in (60) on the sizes $\ell_j \le \omega, j = 0, ..., m-1$, is hidden in the $(1 + o(1))$ term in the asymptotic notation. We multiply (58) by (59), and add over all distinct sets of removed columns $(L_0, ..., L_{m-1})$, and noting that each entry is repeated m times in such sequences, we obtain a quantity $\Psi(m)$ given by

$$\begin{aligned}
\Psi(m) &\sim \frac{1}{m!} \sum_{\ell \ge 1} \sum_{\ell = \ell_0 + \cdots + \ell_{m-1}} \binom{n}{\ell_0, \ldots, \ell_{m-1}} \prod_{j=0}^{m-1} \left(\mathbb{P}(R(L_j) = \mathbf{0}) \cdot P(D_{L_j})\right) \\
&\sim \frac{1}{m!} \sum_{\ell \ge 1} \sum_{\ell = \ell_0 + \cdots + \ell_{m-1}} \prod_{j=0}^{m} (2e^{-2})^{\ell_j} \frac{1}{\ell_j} \sigma_{\ell_j} \\
&= \frac{\phi^m}{m!}.
\end{aligned}$$

Thus, multiplying $\Psi(m)$ by (60) and (61), and summing over $k \ge j$ large dependencies,

$$\begin{aligned}
P_n(m, j]) &\sim \frac{\phi^m}{m!}\, e^{-\phi} \sum_{k \ge j} P^*(j, k; m) \cdot \pi(k) \\
&\sim \frac{\phi^m}{m!}\, e^{-\phi} \sum_{k \ge j} \pi(k) \binom{k}{j} \left(\frac{1}{2^m}\right)^j \left(1 - \frac{1}{2^m}\right)^{k-j}.
\end{aligned} \tag{62}$$

Finally, the probability that null$(M) = d$ is

$$\mathbb{P}(\text{null}(M) = d) = \sum_{m=0}^{d} P_n(m, d-m]),$$

which completes the proof of Theorem 1.

7.4 Going back from M' to M. Change in dimension of null space.

Write $M = \begin{pmatrix} S_L & 0 \\ C & M' \end{pmatrix}$ as given in (57). In this section we prove the following theorem.

Theorem 21. *Suppose that the $(n-L)\times(n-L)$ sub-matrix M' of M has no small dependencies, and k large simple dependencies, and the $L\times L$ sub-matrix S_L of M has m small fundamental dependencies of total size L. Then the probability $P^*(j,k;m)$ that the maximum number of large simple dependencies in M is j is given by*

$$P^*(j,k;m) = \binom{k}{j}\left(\frac{1}{2^m}\right)^j\left(1-\frac{1}{2^m}\right)^{k-j}. \tag{63}$$

Before proceeding with the proof of Theorem 21, we give an outline of the proof structure. Each columns of the sub-matrix C has a unique random non-zero entry in the rows of M'. On average about $\ell/2$ of these non-zeros fall in the rows of any large dependency B of M'. To extend B to a dependency A of M, we may need to include some rows of S_L in A to cancel any non-zeros of C which fall in the rows of B.

Thus in general $A\cap L \neq \emptyset$, and some rows of A were deleted to give B. If M' has k large dependencies $B_1,...,B_k$, then any extension of these sets needs to preserve and extend the intersection structure $I'_{\boldsymbol{x}},\ x\in\{0,1\}^k$ in M' to M. If $j\le k$ of the sets B_i extend successfully then the final intersection structure will be given by $I_y, y\in\{0,1\}^j$. The interaction of this structure with L is the one described in Section 6 and summarized by (40). The extensions are not unique. If A is a large dependency, and L is small, then $A\Delta L$ is large. It was exactly this problem which obliged us to construct our proofs in this way.

Proof of Theorem 21

Suppose M' has k large dependencies $B_1,...,B_k$ but no small dependencies. In this case there is a well defined vector space of dimension k spanned by $B_1,...,B_k$. Assume the m small dependencies $D_j, j=0,...,m-1$ occupy the first L columns. The matrix M can be written as follows.

$$M = \begin{pmatrix} D_0 & 0 & 0 & \cdots & 0 & 0 \\ C_{0,1} & D_1 & 0 & \cdots & 0 & 0 \\ C_{0,2} & C_{1,2} & D_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ C_{0,m-1} & C_{1,m-1} & C_{2,m-1} & \cdots & D_{m-1} & 0 \\ C_{0,m} & C_{1,m} & C_{2,m} & \cdots & C_{m-1,m} & M' \end{pmatrix}.$$

Let $|D_j| = \ell_j$ where $L = (\ell_0 + \cdots + \ell_{m-1})$. Each $(n_j - \ell_j) \times \ell_j$ sub-matrix $C_j = (C_{j,j+1}, ..., C_{j,m})^\top$ has exactly one random one in each column. The probability any of these ones fall in any $C_{j,i}$ where $j+1 \leq i \leq m-1$ for $j = 0, ..., m-1$ is $O(\omega^3/n)$. Conditional on this not occurring, the non-zero entry in each column is u.a.r. in $n' = n - L$. Tidying up, and writing $C'_j = C_{j,m}$ we have

$$M = \begin{pmatrix} D_0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & D_{m-1} & 0 \\ C'_0 & C'_1 & \cdots & C'_{m-1} & M' \end{pmatrix} = \begin{pmatrix} D & 0 \\ C & M' \end{pmatrix}. \tag{64}$$

Let us write $B_r \diamond D_s$ and say B_r agrees with D_s if there exists a set of row indices $J_{r,s}$, a subset of the row indices of D_s, such that $B_r \cup J_{r,s}$ is a dependency of M. We will be able to extend B_r to a dependency $A_r = B_r \cup (\bigcup_{s=1}^m J_{r,s})$ if and only if $B_r \diamond D_s, s = 1, 2, \ldots, m$. In which case we say that $\mathcal{B}_r$ occurs.

For $i \in D_s$, column i has a unit entry in row i, and if the random unit entries are in rows t, t'. We use the notation $e_1(i) = t \in D_s$, $e_2(i) = t' \notin D_s$. Let X_s be the set of column indices associated with the cycle of D_s.

Lemma 22.

(a) $B_r \diamond D_s$ if and only if $|\{i \in X_s : e_2(i) \in B_r\}|$ is even.

(b) Over the random choices, $e_2(i), i \in D_s, s = 1, 2, \ldots, m$, $\mathbb{P}(B_r \diamond D_s, s = 1, 2, \ldots, m) \sim 1/2^m$.

(c) Suppose that $Y \subseteq [r]$ is arbitrary. Then $\mathbb{P}(\mathcal{B}_{r+1} \mid \mathcal{B}_i, i \in Y, \neg\mathcal{B}_i, i \notin Y) \sim \mathbb{P}(\mathcal{B}_{r+1})$. I.e. the occurrence of $\mathcal{B}_{r+1}$ is asymptotically independent of the occurrence or non-occurrence of the events in $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_r$.

Proof. (a) Suppose the vertices of the cycle of D_s are labelled $1, ..., \ell$, with edges $(1,2), ..., (\ell-1, \ell), (\ell, 1)$. Let $(i, i+1)$ be such an edge, where $i, i+1 \in D_s$ and thus $i+1 = e_1(i)$. Then let $x_i = 1$ if $e_2(i) \in B_r$. We introduce variables $y_i, z_i, i = 1, 2, \ldots, \ell$ and these will be used to define the index set of rows $J_{r,s}$, if this is possible. We interpret $y_i = 1$ to mean $i \in J_{r,s}$ and $z_i = 1$ to mean that $e_1(i) \in J_{r,s}$. For $B_r \cup J_{r,s}$ to be a dependency we need $x_i + y_i + z_i = 0$ for $i = 1, 2, \ldots, \ell$. For consistency we need $y_{i+1} = z_i$ for $i = 1, 2, \ldots, \ell$ where $y_{\ell+1} = y_1$. This leads to the equations $y_i + y_{i+1} = x_i, i = 1, 2, \ldots, \ell$. These equations are feasible if and only if

$$x_1 + x_2 + \cdots + x_\ell = 0. \tag{65}$$

If (65) holds there are exactly two possible choices for the y_i. Choosing an arbitrary value in $\{0, 1\}$ for y_1, determines $y_i, i = 2, ..., \ell$ and thus $J_{r,s} = \{i : y_i = 1\}$.

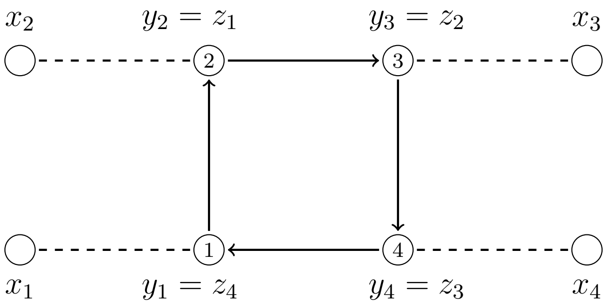

Figure 1: Example: Cycle $(1, 2, 3, 4)$ with labelling. The edges $(i, e_1(i))$ are drawn solid, and edges $(i, e_2(i))$ dashed.

We deal with the attached trees by working backwards from the cycle to the leaves. Suppose that vertex i is not in the cycle and that the values x_j, y_j, z_j have been determined for its parent j its tree. We are forced to take $z_i = y_j$ and then y_i is determined from $x_i+y_i+z_i = 0$. Each time we find that $y_i = 1$, we add i to $J_{r,s}$.

(b) This follows from the fact that $|B_i| \sim n/2$ and $|L| = o(n^{1/2})$ and $e_2(i)$ is chosen randomly from a set of size $n-\omega$. Furthermore, the values $e_2(i), i \in D_{s_1}$ are independent of the choices $e_2(i), i \in D_{s_2}$ if $s_1 \neq s_2$.

(c) Let $I_{\boldsymbol{x}} = (P_1, P_2, \dots, P_{2^r})$ be the partition of $[n-L]$ induced by $B_1, B_2, \dots, B_r$ as described in Section 6. Each part of the partition contains $\sim n/2^r$ rows. The occurrence of $\mathcal{B}_i, i \in Y, \neg\mathcal{B}_i, i \notin Y$ is determined by the the allocation of the $e_2(i)$ into each part. As such, if $e_2(i)$ lies in some part P_t, $1 \leq t \leq 2^r$ then it is distributed uniformly over P_t. Each part of $I_{\boldsymbol{x}}$ is split into two asymptotically equal parts by B_{r+1}. In one "half" we will $x_i = 0$ and in the other "half" we will have $x_i = 1$, where x_i is computed with respect to B_{r+1}. It follows that (65) holds with probability $\sim 1/2$. □

Lemma 22(b) implies that $\mathbb{P}(\mathcal{B}_r) \sim 1/2^m$ and then (c) implies that the probability $P^*(j, k, m)$ of j sets surviving out of k is asymptotic to

$$P^*(j, k; m) = \binom{k}{j} \left(\frac{1}{2^m}\right)^j \left(1 - \frac{1}{2^m}\right)^{k-j}. \tag{66}$$

8 Further comments: Rank over $GF(t)$, and $GF(2)$ for $r \geq 2, s = 2, 3$: Proof of Theorem 3

8.1 Rank over $GF(2)$ for $r \geq 2, s = 2, 3$

Case $r = 2, s = 2$. An $n \times 2n$ matrix of this type has even column sum and row rank $n^* = n - 1$ w.h.p.

Borrowing from [9] Theorem 16.5, for $r = 1$, the expected number of fundamental zero-sum sets of size ℓ is

$$\mathbf{E}X_\ell = \binom{n}{\ell}\left(\frac{\ell-1}{n-1}\right)^\ell\left(\frac{n-1-\ell}{n-1}\right)^{n-\ell} \cdot \frac{1}{(\ell-1)^\ell}\sum_{k=2}^{\ell}(k-1)!k\ell^{\ell-k-1} \sim e^{-\ell}\frac{1}{\ell}\sum_{j=0}^{\ell-2}\frac{\ell^j}{(\ell)_j}.$$

As the last sum tends to $e^\ell/2$ we have $\mathbf{E}X_\ell \leq 1/\ell$. If L is zero-sum, so is $[n] - L$. For $r = 2$ the total expected number of ℓ-dependencies, $2 \leq \ell \leq n-2$ is at most

$$4\sum_{\ell=2}^{n/2}\mathbf{E}X_\ell\left(\frac{\ell-1}{n-1}\right)^\ell \sim 4\sum_{\ell=2}^{n/2}\left(\frac{\ell-1}{n}\right)^\ell\frac{1}{\ell} = O\left(\frac{1}{n^2}\right).$$

Case $r = 2, s = 3$. It follows from the proofs that an $n \times 2n$ matrix of this type has full row rank w.h.p., as the 'second matrix' cancels the constant number of dependencies in the first (if any).

8.2 Rank over $GF(t)$, $t > 2$: Proof of Theorem 4

The proof of Theorem 4 is greatly simplified by the w.h.p. lack of large dependencies.

Case I: The sum of all rows. Let $W(M)$ be an indicator that $\sum_{i=1}^n \boldsymbol{r}_i = \mathbf{0}$, i.e.that the rows of M sum to zero. Then with arithmetic over $GF(t)$,

$$\mathbf{E}W = \begin{cases} \left(\sum_i f_i f_{t-1-i}\right)^n & \text{Model 1,2} \\ \left(\sum_{i+j+k=0} f_i f_j f_k\right)^n & \text{Model 3} \end{cases}.$$

Thus unless $t = 3$ and $f_1 = 1$ (Model 1), $\mathbf{E}W \to 0$ as $n \to \infty$.

Case II: The sum of ℓ rows. Let L be a set of row indices of size ℓ. For a given column i where $i \in L$, for the rows of L to be dependent, one of two events must occur. Either there is a unique random entry in the rows of L which cancels the entry $M_{i,i}$ in row i (Model 2, $\gamma = f_{t-1}$; Model 3, $\gamma = \sum f_i f_{t-i}$). Or there are 3 entries in the column which sum to zero (Model 2, $\alpha = \sum f_i f_{t-i-1}$; Model 3, $\alpha = \sum_{i+j+k=0} f_i f_j f_k$). For a column i, where $i \in [n] - L$ there must either be no random entries, or two random entries adding to zero, with probability $\beta = \sum f_i f_{t-i}$. Thus

$$\mathbf{E}X_\ell = \binom{n}{\ell}\left(2\gamma\frac{\ell}{n}\left(\frac{n-\ell}{n}\right) + \alpha\left(\frac{\ell}{n}\right)^2\right)^\ell \left(\beta\left(\frac{\ell}{n}\right)^2 + \left(\frac{n-\ell}{n}\right)^2\right)^{n-\ell}. \tag{67}$$

The sum of ℓ rows, $\ell \leq \omega$.
From (67) above, using the methods of Section 2 we find $\mathbf{E}Y_\ell$ is given by

$$\mathbf{E}Y_\ell \sim \frac{(2\gamma\ell)^\ell}{\ell!}e^{-2\ell}.$$

Extracting the moments of the fundamental dependencies Z from $\mathbf{E}Y_\ell$ as in Section 2 gives ϕ_t, as given by (5).

The sum of ℓ rows, $\omega < \ell = o(n)$.
As $\beta, \gamma \leq 1$ then $\sum \mathbf{E}X_{\ell>\omega} \to 0$. This follows by comparison with the analysis in Section 3.

The sum of ℓ rows, $\ell = cn$.
Let $\ell = cn$, then

$$\begin{aligned}\mathbf{E}X_{cn} =& O(1)\left(\frac{(2\gamma c(1-c) + \alpha c^2)^c}{c^c}\ \frac{(\beta c^2 + (1-c)^2)^{1-c}}{(1-c)^{(1-c)}}\right)^n \\ =& O(1)\left(D^c G^{1-c}\right)^n.\end{aligned}$$

We prove that, provided $1 \geq 2\gamma > \alpha$, then $D(c) < 1,\ G(c) < 1$ for $c \in (0,1)$, and thus $\mathbf{E}X_{cn} \to 0$.

Firstly $D(0) = 2\gamma \leq 1$, and $D(c) = 2\gamma - (2\gamma - \alpha)c$ which is monotone decreasing in c. Secondly $G(0) = 1$, $G(1) = 1$, and $G'(c) = 0$ at $c = 1 \pm \sqrt{\beta/(\beta+1)}$. Let $\widehat{c} = 1 - \sqrt{\beta/(\beta+1)}$, then $G(\widehat{c}) = 2\sqrt{\beta(\beta+1)} - 2\beta$. As $2\sqrt{\beta(\beta+1)} - 2\beta < 1$, $G(c)$ is a minimum at $\widehat{c}$.

9 Appendix A. Converting between the with and without replacement models

9.1 $\mathbf{E}Y_\ell$ for ℓ small.

Regarding (70), let

$$A=\left(\frac{(\ell-1)(n-\ell)}{(n-1)_2}\right)^{\ell}\left(\left(\frac{(\ell)_2}{(n-1)_2}\right)+\left(\frac{(n-1-\ell)_2}{(n-1)_2}\right)\right)^{n-\ell}.$$

Then

$$\begin{aligned}A&=\left(\frac{(\ell-1)(n-\ell)}{n^2}\right)^{\ell}\left(\left(\frac{\ell}{n}\right)^2+\left(\frac{n-\ell}{n}\right)^2\right)^{n-\ell}\\&\times\left(\frac{n^2}{(n-1)_2}\right)^{n}\left(1-\frac{3(n-\ell)+\ell-2}{\ell^2+(n-\ell)^2}\right)^{n-\ell}.\end{aligned}$$

However

$$\left(\frac{n^2}{(n-1)_2}\right)^{n}=(1+O(1/n))e^3, \tag{68}$$

and for $\ell=o(n)$

$$B=\left(1-\frac{3(n-\ell)+\ell-2}{\ell^2+(n-\ell)^2}\right)^{n-\ell}=(1+O(\ell/n))e^{-3}, \tag{69}$$

which proves equivalence as $A\sim 1$.

$\mathbf{E}X_\ell$ for ℓ large. Note from (69) that B is less than one for any feasible ℓ, and if $\ell=(n/2)(1+o(1)$ then $B=(1+O(1/n))e^{-2}$. Also for any $\ell\to\infty$,

$$(\ell-1)^{\ell}=(\ell)^{\ell}\left(\frac{\ell-1}{\ell}\right)^{\ell}=(1+O(1/\ell))(\ell)^{\ell}e^{-1}.$$

$\mathbf{E}(X)_k$ for $\ell\sim n/2$. Referring to (24), in the with-replacement model we have

$$\Phi(\boldsymbol{h},k)=\prod_{x\neq 0}\left(2\sum_{\substack{\{u,v\}\\u+v=x}}\frac{h_{\boldsymbol{u}}}{n}\frac{h_{\boldsymbol{v}}}{n}\right)^{h_{\boldsymbol{x}}}\left(\sum_{\boldsymbol{u}}\left(\frac{h_{\boldsymbol{u}}}{n}\right)^2\right)^{h_0}$$

The equivalent to $\Phi(\boldsymbol{h},k)$ in the without-replacement model is

$$\begin{aligned}\Psi(\boldsymbol{h},k) =&\prod_{x\neq 0}\left(2\left(\sum_{\substack{\{u,v\}\neq\{x,0\}\\ u+v=x}}\frac{h_{\boldsymbol{u}}h_{\boldsymbol{v}}}{(n-1)_2}+\frac{(h_{\boldsymbol{x}}-1)h_0}{(n-1)_2}\right)\right)^{h_{\boldsymbol{x}}}\left(\frac{(h_0-1)_2}{(n-1)_2}+\sum_{u\neq 0}\left(\frac{(h_{\boldsymbol{u}})_2}{(n-1)_2}\right)\right)^{h_0}\\ =&\Phi(\boldsymbol{h},k)\left(\frac{n^2}{(n-1)_2}\right)^n\prod_{x\neq 0}\left(1-\frac{h_0}{\sum h_{\boldsymbol{u}}h_{\boldsymbol{v}}}\right)^{h_{\boldsymbol{x}}}\left(1-\frac{\sum_{\boldsymbol{u}}h_{\boldsymbol{u}}+2h_0-2}{\sum h_{\boldsymbol{u}}^2}\right)^{h_0}\\ =&\Phi(\boldsymbol{h},k)\cdot C.\end{aligned}$$

As $h_i=(1+o(1))n/2^k$, and $(1-h_0/\sum h_{\boldsymbol{u}}h_{\boldsymbol{v}})^{h_{\boldsymbol{x}}}\sim e^{-2/2^k}$ we have

$$\prod_{x\neq 0}\left(1-\frac{h_0}{\sum h_{\boldsymbol{u}}h_{\boldsymbol{v}}}\right)^{h_{\boldsymbol{x}}}\sim(e^{-2/2^k})^{2^k-1}=e^{-2+1/2^{k-1}},$$

and

$$\left(1-\frac{\sum_{\boldsymbol{u}}h_{\boldsymbol{u}}+2h_0-2}{\sum h_{\boldsymbol{u}}^2}\right)^{h_0}\sim\left(1-\frac{2^k+2}{n}\right)^{n/2^k}=e^{-1-1/2^{k-1}}.$$

Combining this with (68) gives

$$C\sim e^3e^{-2+1/2^{k-1}}e^{-1-1/2^{k-1}}=1.$$

9.2 Without replacement

Let $S=\{2\leq\ell\leq\omega\}$ where $\omega\to\infty$ slowly with n. For $\ell\in S$, let $Y_\ell(M)$ be the number of index sets of zero-sum rows of size ℓ in M. Similarly to (7)

$$\mathbf{E}Y_\ell=\binom{n}{\ell}\left(2\frac{(\ell-1)(n-\ell)}{(n-1)_2}\right)^\ell\left(\left(\frac{(\ell)_2}{(n-1)_2}\right)+\left(\frac{(n-1-\ell)_2}{(n-1)_2}\right)\right)^{n-\ell}. \tag{70}$$

Assuming that $\ell=o(\sqrt{n})$ then

$$\mathbf{E}Y_\ell=\frac{(2(\ell-1))^\ell}{\ell!}e^{-2\ell}(1+o(1)).$$

If L is zero-sum then the sub-matrix $M_{L,L}$ is the incidence matrix of a random functional digraph D_L with no fixed points, in which case there are $\ell-1$ off-diagonal entries in any column of $M_{L,L}$ and we exclude cycles of size one. The probability that the underlying graph of D_L is connected is

$$\mathbb{P}(D_L\text{ connected})=\frac{(\ell-1)!}{(\ell-1)^\ell}\sum_{j=0}^{\ell-2}\frac{\ell^j}{j!}.$$

References

[1] B. Bollobas, *Random Graphs*, 2nd edition. Cambridge University Press (2001).

[2] R. Brualdi and H. Ryser, *Combinatorial Matrix Theory.* Cambridge University Press. (1991).

[3] T. Bohman and A.M. Frieze, Hamilton cycles in 3-out, *Random Structures and Algorithms* 35, 393-417, (2009).

[4] A. Coja-Oghlan, A. Err, P. Gao, S. Hetterich, M. Rolvien, The rank of sparse random matrices, SODA 2020, 579-591, (2020).

[5] C. Cooper, On the rank of random matrices, *Random Structures and Algorithms* 16, 209-232, (2000).

[6] C. Cooper, A.M. Frieze and W. Pegden, On the rank of a random binary matrix, SODA 2019, 946-955, (2019).

[7] T. Fenner and A.M. Frieze, On the connectivity of random m-orientable graphs and digraphs, *Combinatorica* 2, 347-359, (1982).

[8] A.M. Frieze, Maximum matchings in a class of random graphs, *Journal of Combinatorial Theory B* 40, 196-212, (1986).

[9] A.M. Frieze and M. Karoński, *Introduction to Random Graphs*, Cambridge University Press, (2016).

[10] I. N. Kovalenko, A. A. Levitskya and M. N. Savchuk, *Selected Problems in Probabilistic Combinatorics.* Naukova Dumka, Kyiv (1986) (in Russian).