On the Size and Recovery of Submatrices of Ones in a Random Binary Matrix

Xing Sun *and Andrew Nobel †

21 May, 2008

Abstract

Binary matrices, and their associated submatrices of 1s, play a central role in the study of random bipartite graphs and in core data mining problems such as frequent itemset mining (FIM). Motivated by these connections, this paper addresses several statistical questions regarding submatrices of 1s in a random binary matrix with independent Bernoulli entries. We establish a three-point concentration result, and a related probability bound, for the size of the largest square submatrix of 1s in a square Bernoulli matrix, and extend these results to non-square matrices and submatrices with fixed aspect ratios. We then consider the noise sensitivity of frequent itemset mining under a simple binary additive noise model, and show that, even at small noise levels, large blocks of 1s leave behind fragments of only logarithmic size. As a result, standard FIM algorithms, which search only for submatrices of 1s, cannot directly recover such blocks when noise is present. On the positive side, we show that an error-tolerant frequent itemset criterion can recover a submatrix of 1s against a background of 0s plus noise, even when the size of the submatrix of 1s is very small.

Tentatively Accepted at the Journal of Machine Learning Research

A preliminary version of some of these results appeared in "Significance and Recovery of Block Structures in Binary Matrices with Noise", X. Sun and A.B. Nobel, Proceedings of the 19th Annual COLT meeting, H.U. Simon and G. Lugosi eds., Springer, 2006.

*Merck & Co., Inc., Whitehouse Station, New Jersey, USA 08889. Email xing_sun@merck.com

†Department of Statistics and Operation Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3260, Email nobel@email.unc.edu

1 Introduction

In many situations, the data obtained from a standard numerical experiment can be represented by a rectangular matrix, whose columns correspond to subjects or samples, and whose rows correspond to variables or features measured for each subject. In a number of important cases, the measured features can take one of two values, and the resulting data can be represented as a binary matrix. Prominent examples include data mining tasks such as frequent pattern mining, single nucleotide polymorphism (SNP) data obtained from inbred strains having two allelic variants, and quantized versions of continuous measurements.

The initial analysis of large data sets (typically involving many features and small to moderate numbers of samples) is often exploratory, reflecting the increasing use of such data for hypothesis generation, as well as more traditional hypothesis testing. In unsupervised settings, exploratory analysis seeks to identify patterns or other regularities in the observed data that may point to useful (and potentially unknown) associations between variables, samples or both.

The most common form of exploratory analysis is clustering. Clustering algorithms divide the available samples or variables into disjoint groups so that objects in the same group are, in a suitable sense, close together, while objects in different groups are far apart. A natural extension of standard clustering, usually called biclustering or subspace clustering, looks directly for associations between sets of samples and sets of variables. These associations are represented by submatrices of the data matrix.

In the case of binary matrices, the simplest submatrices of interest are constant, with all entries equal to 1. Submatrices of this sort play a key role in data mining applications, and arise naturally in the study of bipartite graphs (see the discussion below). Motivated in part by these connections, this paper considers the extremal properties of submatrices of 1s in a random binary matrix, and considers the recovery of such submatrices in the presence of noise. More specifically, our analyses are based on a model in which the entries of the principal matrix, and the noise, respectively, are independent Bernoulli(p) random variables. We provide significance bounds for the size of submatrices of 1s under the Bernoulli null hypothesis, and use these to establish limits on the performance of standard data mining methods in the presence of Bernoulli noise. In the same context, we establish several results on the precise asymptotic size of maximal submatrices of 1s, extending to the setting of bipartite graphs earlier work of Bollobás and Erdős [6] and Matula [19] on the size of maximal cliques in random graphs. Lastly, we establish finite sample and asymptotic results concerning the recovery of all-1s submatrices in the presence of noise.

1.1 Overview

Connections between binary matrices, frequent itemset mining, and bipartite graphs are discussed in the next section. Section 3 is devoted to the size of the largest square submatrix of 1s in a random binary matrix. Extensions to non-square matrices are described in Section 4. Section 5 contains a short simulation study that supports our theoretical bounds in a non-asymptotic setting. Section 6 is devoted to the noise sensitivity of frequent itemset mining and the recoverability of block structures in the presence of noise.

2 Motivation and Background

An $m \times n$ binary matrix is an indexed family $X = \{x_{i,j} : i \in [m], j \in [n]\}$ where $x_{i,j} \in \{0,1\}$ and $[k]$ denotes the set $\{1, \ldots, k\}$. A submatrix of X is a sub-family $U = \{x_{i,j} : i \in A, j \in B\}$ where $A \subseteq [m]$ and $B \subseteq [n]$; the Cartesian product $C = A \times B$ will be called the index set of U, and we will write $U = X[C]$. When no ambiguity will arise, the index set C itself will be referred to as a submatrix of X.

2.1 Frequent Itemset Mining

Frequent itemset mining (FIM) [3, 2], also known as market basket analysis, is a central problem in the field of Data Mining. Generalizations such as bi-clustering and subspace clustering [28, 7, 1, 8] remain active areas of research. A discussion of FIM and related methods can be found in [18, 13, 11].

In the frequent itemset problem, the available data is described by a list $S = \{s_1, \ldots, s_n\}$ of items and a set $T = \{t_1, \ldots, t_m\}$ of transactions. Each transaction t_i consists of a subset of the items in S. If S contains the items available for purchase at a store, then t_i represents a record of the items purchased during the ith transaction, without multiplicity. The goal of FIM is to identify every (maximal) set of items that appear together in more than k transactions, where $k \geq 1$ is a threshold that quantifies "frequent". The data for the FIM problem can readily be represented by an $m \times n$ binary matrix X, with entry $x_{i,j} = 1$ if transaction t_i contains item s_j, and $x_{i,j} = 0$ otherwise. In this form the FIM problem can be stated as follows: given X and $k \geq 1$, find every submatrix of 1s in X having at least k rows, and report the associated set of columns. If the threshold k is allowed to vary, then FIM algorithms essentially seek to find every maximal submatrix of 1s in the data matrix X.

The ongoing application of FIM to large data sets for the purposes of exploratory and related analyses raises a number of natural statistical questions, which we address below in the general setting of random binary matrices. One natural question is how to assign a nominal significance value to the discovery of a moderately sized submatrix of 1s in a large data matrix, accounting for the obvious issue of multiple comparisons arising in this case. Another question is how standard FIM methods perform in the presence of noise, a common feature of many high-throughput measurement technologies. The third question is how one can recover a submatrix of 1s embedded in a larger matrix of 0s when noise is present.

2.2 Bipartite Graphs

Binary matrices are in one to one correspondence with bipartite graphs. An $m \times n$ binary matrix X can be viewed as the adjacency matrix of a graph $G = (V, E)$, where the vertex set V of G is the disjoint union of two sets V_1 and V_2, with $|V_1| = m$ and $|V_2| = n$, corresponding to the rows and columns of X, respectively. There is an edge $(i, j) \in E$ between vertices $i \in V_1$ and $j \in V_2$ if and only if $x_{i,j} = 1$; there are no edges between vertices in V_1 or vertices in V_2. A submatrix U of X with index set $C = A \times B$ corresponds to the subgraph G' of G induced by the vertex set $A \cup B$. If every entry of U is equal to one, then there is an edge (i, j) between every pair of vertices $i \in A$ and $j \in B$, and G' is then a complete bipartite subgraph of G. Thus maximal submatrices of 1s in X correspond to bicliques in G. This connection is the basis for the biclustering algorithm of Tanay *et al.* [29].

It is known (*c.f.* [10, 14, 22]) that the problem of finding a biclique with the largest number of edges in a given bipartite graph G is NP-complete, and thus the same is true of the general frequent itemset problem with no restriction on the threshold k. Several approximate methods [14, 20] have been proposed for finding large bicliques in bipartite graphs in polynomial time. Mishra *et al.* [20] show that the results provided by their randomized algorithm overlap a large fraction of the largest bicliques with high probability.

Our interest here is in assessing the significance and extremal size of maximal bicliques in random bipartite graphs. We do not address the question of how to search for such bicliques, and refer the interested reader to the papers above and the references therein for more details.

3 Largest Submatrices of 1s: Square Case

In this section we study the size of the largest square submatrix of 1s in a square binary matrix whose entries are independent Bernoulli(p) random variables. Non-square matrices and submatrices are considered in Section 4.

Definition: Let $Z = \{z_{i,j} : i, j \geq 1\}$ be an infinite array of independent binary random variables with $P(z_{i,j} = 1) = p = 1 - P(z_{i,j} = 0)$, where the probability $p \in (0, 1)$ is fixed. For $n \geq 1$, let $Z_n = \{z_{i,j} : 1 \leq i, j \leq n\}$.

Thus Z_n is an $n \times n$ binary random matrix comprising the "upper left corner" of the collection $\{z_{i,j}\}$. This definition allows us to make almost-sure type statements concerning the asymptotic behavior of functions of Z_n.

Definition: Given a binary matrix X, let $M(X)$ be the largest k such that there exists a $k \times k$ submatrix of 1s in X. Note that $M(X)$ is invariant under row and column permutations of X.

From a statistical point of view, the random matrix Z_n follows a simple null model under which the observed binary data matrix has no special structure, and $M(\cdot)$ acts as a natural test statistic with which to detect departures from the null. Our analysis begins with a bound on the probability that $M(Z_n)$ exceeds a fixed integer $k \geq 1$. We follow a standard first moment argument (*c.f.* [4]).

Fix n for the moment, and for each $1 \leq k \leq n$ let U_k be the number of $k \times k$ submatrices of ones in Z_n. Then, letting $S = \{C = A \times B : A, B \subseteq [n], |A| = |B| = k\}$, we may write

$$U_k \;=\; \sum_{C \in S} I\{\text{all entries of } Z_n[C] \text{ are } 1\} \tag{1}$$

from which it follows that

$$EU_k \;=\; |S| \cdot P(\text{all entries of } Z_n[C] \text{ are } 1) \;=\; \binom{n}{k}^2 p^{k^2}.$$

By Markov's inequality and the previous display,

$$P\left(M(Z_n) \geq k\right) \;=\; P(U_k \geq 1) \;\leq\; EU_k \;=\; \binom{n}{k}^2 p^{k^2}. \tag{2}$$

We wish to identify an integer k_n for which EU_{k_n} is approximately equal to one. For values $k > k_n$ the rightmost expression in (2) provides an effective means for bounding the probability on the left. Note that $EU_n = p^{n^2} < 1$, and $EU_1 = n^2 p > 1$ when n is sufficiently large. Moreover, it is clear from the definition that $U_{k+1} \leq U_k$, so that EU_k is non-increasing in k. Using the Stirling approximation of the rightmost expression in (2), define

$$\phi_n(s) \;=\; (2\pi)^{-\frac{1}{2}}\, n^{n+\frac{1}{2}}\, s^{-s-\frac{1}{2}}\, (n-s)^{-(n-s)-\frac{1}{2}}\, p^{\frac{s^2}{2}}, \qquad s \in (0,n). \tag{3}$$

The quantity $\phi_n(k)$ is an approximation of $(EU_k)^{1/2}$: the ratio $\phi_n(k)/(EU_k)^{1/2}$ is bounded away from zero and infinity, independent of n, k, and tends to one if k and $n-k$ tend to infinity with n. Let $s(n)$ be any positive real root of the equation

$$1 = \phi_n(s). \tag{4}$$

The next lemma shows that $s(n)$ is unique and grows as logarithmically with n.

Lemma 1. *When n is sufficiently large, the equation (4) has a unique root $s(n)$ satisfying $\log_b n < s(n) < 2\log_b n$, where $b = p^{-1}$.*

Using the bounds of Lemma 1 and some technical but straightforward calculations, one may obtain a simple asymptotic expression for $s(n)$.

Lemma 2. *The root $s(n)$ defined by (4) has the form*

$$s(n) \;=\; 2\,\log_b n \;-\; 2\,\log_b \log_b n + C + o(1) \tag{5}$$

where $b = p^{-1}$ and $C = 2\log_b e - 2\log_b 2$.

The proofs of Lemmas 1 and 2 can be found in Section 7.1. Let $k(n) = \lceil s(n) \rceil$ be the least integer greater than or equal to $s(n)$. The next proposition provides an upper bound on $P(M(Z_n) \geq k)$ for $k > k(n)$. Its proof appears in Section 7.2.

Proposition 1. *For each $\epsilon > 0$, when n is sufficiently large, $P\left(M(Z_n) \geq k(n) + r\right) \leq n^{-2\,r}\,(\log_b n)^{2r+\epsilon}$.*

One may obtain a cruder bound, on the probability that $M(Z_n)$ is at least $2\log_b n + r$, in a simpler fashion by noting that

$$EU_k \;=\; \binom{n}{k}^2 p^{k^2} \;\leq\; \frac{n^{2k}}{k!^2} e^{-k^2 \log b} \;\leq\; \frac{e^{2k\ln n - k^2 \ln b}}{k^2} \;\leq\; n^{-2r} \tag{6}$$

when $k \geq 2\log_b n + r$. Both the upper bound of Proposition 1 and the definition of $s(n)$ are based on the inequality (2), which follows from a simple union bound on the probability that $M(Z_n)$ is at least k. The union bound is typically quite loose, but it is sufficiently strong in this context to ensure that, for large n, the random variable $M(Z_n)$ is close to the threshold $s(n)$. Indeed, it follows from Proposition 1 and the first Borel Cantelli Lemma that, with probability one, $M(Z_n)$ is eventually less than $s(n) + 1$. Using a more involved second moment argument, one can establish a corresponding lower bound on $M(Z_n)$. Together these bounds yield the following result.

Theorem 1. *Given any $\epsilon > 0$, with probability one, $s(n) - 1 - \epsilon < M(Z_n) < s(n) + \epsilon$ when n is sufficiently large.*

It follows from Theorem 1 that for large n the size of the largest square submatrix of 1s in Z_n can take one of at most two integer values in an interval of width $1 + 2\epsilon$ containing the number $s(n)$. Indeed, it is shown in the proof of Theorem 1 that there is a sequence of integers $\{r(n)\}$ close to $\{s(n)\}$ such that, with probability one, when n is sufficiently large $M(Z_n) \in \{r_n - 1, r(n)\}$. Thus $M(Z_n)$ exhibits two-point concentration and does not possess a limiting continuous distribution.

The proof of Theorem 1 is given in Section 8. The outline of the proof follows arguments of Bollobás and Erdős [6], who studied the size of the largest clique $cl(G_n)$ in a random graph G_n with n vertices, where each edge is included independently with probability p. They showed that for a deterministic function $c(n)$, equal to $s(n)$ up to the constant and lower order terms, eventually almost surely $|cl(G_n) - c(n)| < 3/2$. Matula [19] independently established a similar result. See these references or Bollobás [5] for more details. Theorem 1 extends these results to balanced bicliques in balanced bipartite random graphs. (Unbalanced bipartite graphs are considered in the next section.)

Dawande *et al.* [8] used first and second moment arguments to show (in our terminology) that $P(\log_b n \leq M(Z_n) \leq 2\log_b n) \to 1$ as n tends to infinity. Improving these results, Park and Szpankowski [21] showed that $P((1+\epsilon)\log_b n \leq M(Z_n) \leq (2-\epsilon)\log_b n)$ tends to 1 as n tends to infinity for any fixed $0 < \epsilon < 1$. Koyutürk *et al.* [15] studied the problem of finding dense patterns in binary data matrices. They used a Chernoff type bound for the binomial distribution to assess whether an individual submatrix has an enriched fraction of ones, and employed the resulting test as the basis for a heuristic search for significant bi-clusters. However, the effects of multiple testing are not considered in their assessments of significance. Tanay *et al.* [28] assessed the significance of bi-clusters in a real-valued matrix using likelihood-based weights, a normal approximation and a standard Bonferroni correction to account for the multiplicity of submatrices. Use of the normal approximation for individual submatrices leads to subtoptimal bounds in non-Gaussian settings.

3.1 Smallest Maximal Submatrix of 1s

Square submatrices of 1s will occur by chance in a random binary matrix. The largest such submatrix has approximately $2 \log_b n - 2 \log_b \log_b n$ rows. Conversely, one may ask about the size of the *smallest* maximal square submatrix of 1s. (A square submatrix of 1s is maximal if there is no larger square submatrix of 1s that properly contains it.)

Definition: Let $L(Z_n)$ be the smallest k such that there exists at least one $k \times k$ maximal submatrix of 1's in Z_n.

Theorem 1 implies that $L(Z_n) \leq 2 \log_b n$. An analysis based on second moment arguments similar to those used in the proof of Theorem 1 yields the following, tighter bound. The proof can be found in [27].

Theorem 2. *With probability one,*

$$\lim_{n \to \infty} \frac{L(Z_n)}{\log_b n} = 1.$$

Bollobás and Erdős [6] establish a related result on the size of the smallest clique in a random graph. However their proof can not be directly extended to obtain the theorem above. Indeed, an extension of their argument provides a lower bound on the size of the smallest square submatrix of 1s that is not properly contained within a rectangular submatrix of 1s, and the resulting bound is necessarily larger than the one in Theorem 2.

4 Non-Square Matrices

In this section we consider the case where the primary matrix and the target submatrices of 1s may be rectangular, but maintain fixed row/column aspect ratios as the size of the primary matrix grows. Natural analogs of Proposition 1 and Theorem 1 are obtained in this setting. For $m, n \geq 1$ define the random matrix $Z(m, n) = \{z_{i,j} : i \in [m], j \in [n]\}$.

Definition: Let $\alpha > 0$ and $\beta > 0$ be aspect ratios for the primary matrix and target submatrices, respectively. Define $M_n(Z : \alpha, \beta)$ to be the largest k such that $Z(\lceil \alpha n \rceil, n)$ contains a $\lceil \beta k \rceil \times k$ submatrix of 1s.

The asymptotic behavior of $M_n(Z : \alpha, \beta)$ is the same as that of $M_n(Z : \alpha^{-1}, \beta^{-1})$, so we assume in what follows that $\beta \geq 1$. The analysis of $M_n(Z : \alpha, \beta)$ proceeds along the same lines as that of $M(Z_n)$. Investigating the value of k for which the expected number of $\lceil \beta k \rceil \times k$ submatrices of 1s in $Z(\lceil \alpha n \rceil, n)$ is equal to 1, we arrive at the function

$$s(n, \alpha, \beta) \;=\; \frac{1+\beta}{\beta} \log_b n - \frac{1+\beta}{\beta} \log_b \left(\frac{1+\beta}{\beta} \log_b n \right) + \log_b \alpha + C(\beta) + o(1), \qquad (7)$$

where $b = p^{-1}$ and $C(\beta) = \beta^{-1}((1+\beta)\log_b e - \beta \log_b \beta)$ depends only on β.

Note that the aspect ratio α of the primary matrix appears only in the constant term of $s(n, \alpha, \beta)$, and therefore plays only a minor role in what follows. The proofs of Proposition 2 and Theorem 3 below are similar to their analogs in the square case, with additional notation and work required to handle the two aspect ratios, and are omitted. Detailed arguments can be found in [27].

Proposition 2. *Fix aspect ratios $\alpha > 0$, $\beta \geq 1$. For every $\epsilon > 0$, when n is sufficiently large* $P\left(M_n(Z : \alpha, \beta) \;\geq\; \lceil s(n, \alpha, \beta) \rceil + r\right) \;\leq\; 2n^{-(\beta+1)\,r} \left(\log_b n\right)^{(\beta+1+\epsilon)r}$.

Remark: When the aspect ratio α of the primary matrix is fixed, it does not play an essential role in the asymptotic behavior of $M_n(Z : \alpha, \beta)$, which is dominated by higher order factors involving only the aspect ratio β of the target submatrices. It is natural then to consider a situation in which the aspect ratio α of the primary matrix can increase with n. This might model, for example, the scaling and cost structure of a given high-throughput technology over time. In the case where $\alpha(n) = n^\gamma$ for some $\gamma > 0$, the proof of Proposition 2 can be modified to show that

$$P\left(M_n(Z : n^\gamma, \beta) \;\geq\; \left(\gamma + \frac{\beta+1}{\beta} \right) \log_b n \right) \;\leq\; 2n^{-(\beta+1)\,r} \left(\log_b n\right)^{(\beta+1+\epsilon)r}.$$

On the other hand, one can readily show that if $\beta \geq 1$ is fixed and m grows exponentially with n, then $Z(m, n)$ will contain a $\lceil \beta n \rceil \times n$ submatrix of 1's with probability bounded away from zero. For fixed aspect ratios α and β one may obtain an asymptotic concentration result for $M_n(Z : \alpha, \beta)$ analogous to Theorem 1.

Theorem 3. *For fixed $\alpha > 0$ and $\beta \geq 1$, with probability one $|M_n(Z : \alpha, \beta) - s(n, \alpha, \beta)| \leq \frac{5}{2}$ when n is sufficiently large.*

Theorem 3 implies that $Z(\alpha n, n)$ contains a submatrix of 1s having aspect ratio β and area $(\beta+1)\log_b^2 n$, the latter increasing with β. Park and Szpankowski [21] establish a related result, showing that if we do not restrict β, the aspect ratio of the submatrices, then with high probability the submatrix of 1s in $Z(m, n)$ with the largest area is of size $O(n) \times \ln b$ or $\ln b \times O(n)$.

Table 1: Distribution of observed $\hat{M}(Z_n)$ based on simulation

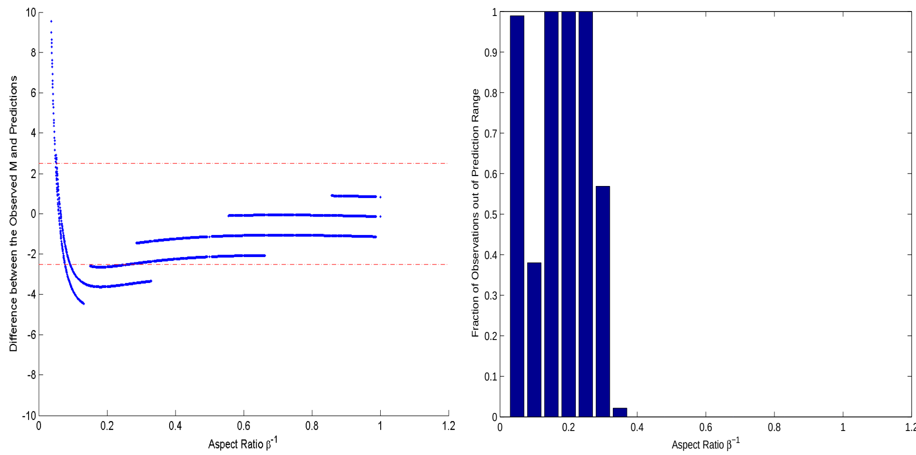
p	n	$s(n)$	k	Proportion of $\hat{M} = k$
0.2	40	3.55	3	85.75%
			4	14.25%
	80	4.58	4	97%
			5	3%
0.3	40	4.78	4	50.5%
			5	49.5%
	80	5.64	5	85%
			6	15%
0.35	40	5.22	4	63.75%
			5	36%
			6	0.25%
	80	6.21	5	7.75%
			6	90.75%
			7	1.50%

5 Simulation Study

The results of the previous sections hold when n is sufficiently large. In order to assess their validity for moderate values of n, we carried out a simple simulation study. For $n = 40$ and $n = 80$ we generated 400 $n \times n$ random binary matrices with $p = .2$, $p = .3$ and $p = .35$ respectively. Then we applied the FP-growth algorithm [12] to identify all maximal submatrices of ones. For each maximal submatrix of ones we recorded the length of its shorter side, and let $\hat{M}$ be the maximum among these lengths. Thus $\hat{M}$ is the side length of the largest square submatrix of 1's in the generated random matrix. We recorded the values of $\hat{M}$ over all simulations and compared these values to the corresponding bounds. Table 1 summarizes the results. Note that in each simulation $-1.5 < \hat{M} - s(n) < 1$.

In order to check the theoretical bounds on $M_n(Z : 1, \beta)$ with $\beta \geq 1$, we considered the 400 random 80×80 matrices with p=0.3 used to evaluate the result for square submatrices above. For each such matrix, we identified all maximal rectangular submatrices of 1s, and recorded the length of both their longer and shorter sides. For each $\beta \geq 1$ we defined

Figure 1: Difference between observed $\hat{M}(\beta)$ and its predicted value from theory.



$\hat{M}(\beta)$ to be the largest k such that at least one $\lceil \beta k \rceil \times k$ or $k \times \lceil \beta k \rceil$ submatrix of 1's was observed. The difference between $\hat{M}(\beta)$ and $s(80, 1, \beta)$ was calculated and is displayed in Figure 1. The x-axes in both panels are equal to $1/\beta$. The y-axis in the left panel is the difference between $\hat{M}(\beta)$ and $s(80, 1, \beta)$, and the y-axis in the right panel is the proportion of simulations which are inconsistent with the theoretical predictions of Theorem 3. Note that even for the moderate matrix size $n = 80$, the theoretical predictions are very accurate when the aspect ratio β is less than 2.5. In these cases, all the observed size lengths are within the range of predicted values.

6 Fragmentation and Recovery in the Presence of Noise

In this section we shift our attention from submatrices of 1s in Z_n to a setting in which Z_n plays the role of binary noise. Formally, we study the additive model

$$Y_n = X_n \oplus Z_n, \tag{8}$$

where each matrix is of dimension $n \times n$. The operation $\oplus$ is the standard exclusive-or: $0 \oplus 0 = 1 \oplus 1 = 0$ and $0 \oplus 1 = 1 \oplus 0 = 1$. The matrix $X_n = \{x_{i,j}\}$ is a non-random binary matrix that contains the "true" values of interest, in the absence of noise, and Z_n is a random binary matrix that acts as noise, with intensity $p \in (0, 1)$. The matrix $Y_n = \{y_{i,j} = x_{i,j} \oplus z_{i,j}\}$ represents the observed binary data. Thus the effect of the noise is to randomly flip some of the values of X in Y. The model (8) is the binary version of the standard additive noise model common in statistical inference.

6.1 Noise Sensitivity

Much of the data to which data mining methods are applied is obtained by high-throughput technologies or the automated collection of information from diverse sources with varying levels of reliability. The resulting data sets are often subject to moderate levels of error and noise. Noise can also arise when binary data are obtained by thresholding continuous data, as is sometimes done in microarray analyses. Whatever its source, noise can potentially have serious consequences for frequent itemset methods if they are applied in a direct way to identify submatrices of 1s.

Indeed, this conclusion is already apparent from Theorem 1. If each entry of the target matrix X_n is zero, then $Y_n = Z_n$ and the largest $k \times k$ submatrix of ones in Y_n has $k \approx 2 \log_b n$ with $b = p^{-1}$. At the other extreme, if every entry of X_n is equal to one, then the entries of Y_n are independent Bernoulli$(1-p)$ random variables, and in this case the largest square submatrix of ones in Y has side-length $k \approx 2 \log_{b'} n$ with $b' = (1-p)^{-1}$. The next result extends this reasoning to any underlying target matrix X_n.

Proposition 3. *Fix $0 < p < 1/2$. Let $\{X_n\}$ be any sequence of $n \times n$ square binary matrices, and let $Y_n = X_n \oplus Z_n$. For each $\epsilon > 0$, eventually almost surely $(2-\epsilon) \log_b n < M(Y_n) \leq 2 \log_{b'} n$, where $b = p^{-1}$ and $b' = (1-p)^{-1}$.*

Proof of Proposition 3: Fix n and let $\tilde{W}_n = \{\tilde{w}_{i,j}\}$ be an $n \times n$ binary matrix with independent entries, defined on the same probability space as $\{z_{i,j}\}$, such that

$$\tilde{w}_{i,j} = \begin{cases} \text{Bern}\left(\frac{1-2p}{1-p}\right) & \text{if } x_{ij} = y_{ij} = 0 \\ 1 & \text{if } x_{ij} = 0, y_{ij} = 1 \\ y_{i,j} & \text{if } x_{ij} = 1 \end{cases} \tag{9}$$

where the Bernoulli variable in the first condition is independent of $\{z_{i,j}\}$. Define $\tilde{Y}_n = Y_n \vee \tilde{W}_n$ to be the entry-wise maximum of Y_n and $\tilde{W}_n$. Then clearly $M(Y_n) \leq M(\tilde{Y}_n)$, as any submatrix of ones in Y_n must also be present in $\tilde{Y}_n$. Moreover, the variables $\tilde{y}_{i,j}$ are i.i.d. with $P(\tilde{y}_{i,j} = 1) = 1 - p$, so that we may regard $\tilde{Y}_n$ as a Bern$(1-p)$ noise matrix. It then follows from Theorem 1 that $M(Y_n) \leq 2 \log_{b'} n$ eventually almost surely. To obtain the other inequality, define

$$\hat{w}_{i,j} = \begin{cases} \text{Bern}\left(\frac{p}{1-p}\right) & \text{if } x_{ij} = y_{ij} = 1 \\ 0 & \text{if } x_{ij} = 1, y_{ij} = 0 \\ y_{i,j} & \text{if } x_{ij} = 0 \end{cases} \tag{10}$$

and let $\hat{Y}_n = Y_n \wedge \hat{W}_n$ be the entry-wise maximum of Y_n and $\hat{W}_n$. It is easy to verify that $M(Y_n) \geq M(\hat{Y}_n)$, and that the entries in $\hat{Y}_n$ are i.i.d. Bern(p). Theorem 1 then implies that $M(Y_n) \geq (2-\epsilon)\log_b n$ eventually almost surely.

Proposition 3 can be interpreted as follows. No matter what type of block structures might exist in X, in the presence of random noise these structures leave behind only logarithmic fragments in the observed data. Under the additive noise model (8), block structures in X cannot be recovered directly by methods such as frequent itemset mining that look for maximal submatrices of ones without errors.

6.2 Recovery

In light of Proposition 3, it is natural to consider methods for identifying submatrices of 1s that may be contaminated with a certain fraction of 0s. These submatrices correspond, in the data mining and bipartite graph settings, to approximate frequent itemsets and approximate bicliques, respectively. A number of different error-tolerant frequent itemset mining algorithms have been proposed in the literature [24, 23, 30, 26, 17, 16]. Most are based on criteria that require the average of the identified submatrices to be greater than a user specified threshold τ. One can readily adapt the first moment argument to obtain significance bounds for submatrices with a large fraction of 1s; details can be found in [27].

Here we consider the simple problem of recovering a (potentially small) submatrix C of 1s embedded in a matrix of 0s from a single noisy observation. Proposition 3 shows that one cannot recover C directly using standard frequent itemset mining; instead we consider the Approximate Frequent Itemset (AFI) algorithm developed in Liu *et al.* [17].

Definition: Given a binary matrix U with index set C, let $F(U) = |C|^{-1}\sum_{(i,j)\in C} u_{i,j}$ be the fraction of ones in U, or equivalently, the average of the entries of U.

Let u_{i*} and u_{*j} denote the rows and columns, respectively, of a given submatrix U.

Definition: Let $\tau \in [0,1]$ be fixed. A submatrix U of a binary matrix Y is a τ-*approximate frequent itemset* (τ-AFI) if each of its rows satisfies $F(u_{i*}) \geq \tau$ and each of its columns satisfies $F(u_{*j}) \geq \tau$. Define $\mathrm{AFI}_\tau(Y)$ to be the collection of all τ-AFIs in Y.

The definition above comes from Liu *et al.* [17], who presented an algorithm for identifying AFIs in binary matrices.

Let X_n be an $n \times n$ binary matrix that consists of an $l \times l$ submatrix of ones having index set C^*, with all other entries equal to 0. (The rows and columns of C^* need not be contiguous.) Suppose that $Y_n = X_n \oplus Z_n$, where Z_n has noise level $p \in (0, 1/2)$. We wish to recover the index set C^* of the target submatrix from Y_n.

To this end, assume that the noise level p is unknown, but that there is a known upper bound p_0 such that $p < p_0 < 1/2$, and let $\tau = 1 - p_0$ be an associated error threshold. We estimate C^* by the index set of the largest square τ-AFI in the observed matrix Y_n. More precisely, let $\hat{\mathcal{C}}$ be the family of index sets of square submatrices $U \in \mathrm{AFI}_\tau(Y_n)$, and let

$$\hat{C} \; = \operatorname*{argmax}_{C \in \hat{\mathcal{C}}} |C|$$

be the index set of any maximal sized submatrix in $\hat{\mathcal{C}}$. (The set $\hat{\mathcal{C}}$ contains 1×1 submatrices with entry 1, so it is non-empty whenever Y_n is not identically 0.) Note that $\hat{\mathcal{C}}$ and $\hat{C}$ depend only on the observed matrix Y_n. Let the ratio

$$\Lambda \; = \; |\hat{C} \cap C^*| / |\hat{C} \cup C^*|$$

measure the overlap between the estimated index set $\hat{C}$ and the true index set C^*. Clearly $0 \le \Lambda \le 1$, and values of Λ close to one indicate better overlap. The proof of the next theorem is given in Section 9.

Theorem 4. *When n is sufficiently large, for any $0 < \alpha < 1$ such that $8\alpha^{-1}(\log_b n + 2) \le l$ we have*

$$P\left(\Lambda \le \frac{1-\alpha}{1+\alpha}\right) \; \le \; \Delta_1(l) \, + \, \Delta_2(\alpha, l). \tag{11}$$

Here $\Delta_1(l) = 2le^{-\frac{3l(p-p_0)^2}{8p}}$ and $\Delta_2(\alpha, l) = 2n^{-\frac{1}{4}\alpha l + 2\log_b n}$, with $b = \exp\{3(1 - 2\,p_0)^2/8p\}$.

Remarks: The second term $\Delta_2(\alpha, l)$ is less than $2n^{-4/\alpha}$ and is the dominant term in the probability upper bound if $l/\ln(n)$ is large. The logarithmic base b is derived from an upper bound on the tails of the binomial distribution, and is always larger than $\tilde{b} = \exp\{3(1 - 2\,p_0)^2/8p_0\}$. By a crude bound, $\Delta_1(l) \le \tilde{\Delta}_1(l) := e^{-\sqrt{l}}$ when l is sufficiently large. Thus, by replacing b with $\tilde{b}$ and $\Delta_1(l)$ with $\tilde{\Delta}_1(l)$, one obtains a probability bound that does not depend on the unknown parameter p.

As a corollary of Theorem 4, we can also get results in an asymptotic setting. Suppose that $\{X_n : n \ge 1\}$ is a sequence of square binary matrices, and that X_n contains an $l_n \times l_n$ submatrix C_n^* of 1s with all other entries equal to 0. Let $Y_n = X_n \oplus Z_n$, and let Λ_n measure the overlap between C_n^* and the estimate $\hat{C}_n$ produced by the AFI-based recovery method above. The following result follows from Theorem 4 and the Borel Cantelli lemma.

Corollary 1. *If $l_n \geq 8\psi(n)(\log_b n+2)$ where $\psi(n) \to \infty$ as $n \to \infty$, then eventually almost surely*

$$\Lambda_n \;\geq\; \frac{1-\psi(n)^{-1}}{1+\psi(n)^{-1}} \to 1.$$

Reuning-Scherer studied several recovery problems in [25]. In the case considered here, he calculated the fraction of 1s in every row and every column of Y, and then selected those rows and columns for which these fractions exceeded an appropriate threshold. His algorithm is easily seen to be consistent when $l \geq n^\alpha$ for $\alpha > 1/2$. However, it is easy to show using the central limit theorem that individual row and column sums alone are not sufficient to recover C^* when $l \leq n^\alpha$ for $\alpha < 1/2$. In the latter case, one gains considerable power by directly considering submatrices, and as the result above demonstrates, one can consistently recover C_n^* if $l_n/\ln(n) \to \infty$.

7 Proofs of Preliminary Results

7.1 Proofs of Lemmas 1 and 2

Proof of Lemma 1: Differentiating $\log_b(\phi_n(s))$ yields

$$\frac{\partial \log(\phi_n(s))}{\partial s} = \frac{1}{2(n-s)} + \log_b(n-s) - s - \log_b s - \frac{1}{2s},$$

which is negative when $\log_b n < s < 2\log_b n$. A routine calculation shows that for $0 < s \leq \log_b n$,

$$\begin{aligned}
\log_b \phi_n(s) \;&=\; (n+\frac{1}{2})\log_b n - (s+\frac{1}{2})\log_b s - (n-s+\frac{1}{2})\log_b(n-s) - \frac{s^2}{2} - \frac{1}{2}\log_b 2\pi \\
&\geq\; s\left(\log_b(n-\log_b n) - \frac{s}{2} - \log_b\log_b n\right) - \frac{1}{2}\log_b s - \frac{1}{2}\log_b 2\pi \;\; > 0
\end{aligned}$$

when n is sufficiently large. Similarly, for $2\log_b n \leq s < n$,

$$\begin{aligned}
\log_b \phi_n(s) \;&\leq\; s\left(\log_b(n-s) - \frac{s}{2} - \log_b s\right) - \frac{1}{2}\log_b s - \frac{1}{2}\log_b 2\pi + 2s + \frac{s\log_b s}{2} \\
&\leq\; s\left(2 - \frac{\log_b s}{2}\right) - \frac{1}{2}\log_b s - \frac{1}{2}\log_b 2\pi \;\; < \;\; 0
\end{aligned}$$

when n is sufficiently large. Thus for sufficiently large n, there exists a unique solution $s(n)$ of the equation $\phi_n(s) = 1$ with $s(n) \in (\log_b n, 2\log_b n)$. ■

Proof of Lemma 2: Taking logarithms of both sides of the equation $\phi_n(s) = 1$ and rearranging terms yields

$$\frac{1}{2}\log_b \frac{n}{n-s} + n\log_b \frac{n}{n-s} - (s+\frac{1}{2})\log_b s + s\log_b(n-s) - \frac{s^2}{2} = \frac{\log_b 2\pi}{2}.$$

Lemma 1 implies that $s(n)$ belongs to the interval $(\log_b n, 2\log_b n)$, so we consider the above equation in the case that $n >> s$. Dividing both sides of the equation by s yields

$$\log_b(n-s) - \frac{s}{2} - \log_b s = -\log_b e + O(\frac{\log_b s}{s}),$$

which can be rewritten as

$$\log_b n - \frac{s}{2} - \log_b \log_b n = \log_b \frac{s}{\log_b n} - \log_b \frac{n-s}{n} - \log_b e + O(\frac{\log_b s}{s}). \tag{12}$$

For each n, define $R(n)$ via the equation

$$s(n) \;=\; 2\log_b n - 2\log_b \log_b n + R(n).$$

Plugging this expression into (12), it follows that $R(n) = 2\log_b e - 2\log_b 2 + o(1)$, and the result follows from the uniqueness of $s(n)$. ■

7.2 Proof of Proposition 1

To establish the bound with r independent of n, it suffices to consider a sequence r_n that changes with n in such a way that $1 \le r_n \le n$. Fix n for the moment, let $l = k(n) + r_n$, and let U_l be be the number of $l \times l$ submatrices of 1s in Z_n. Then by Markov's inequality and Stirling's approximation,

$$P(M(Z_n) \ge r) \;=\; P(U_l \ge 1) \;\le\; E(U_l) \;=\; \binom{n}{l}^2 p^{l^2} \;\le\; 2\phi_n^2(l). \tag{13}$$

A straightforward calculation using the definition of $\phi_n(\cdot)$ shows that one can decompose the rightmost term above as follows:

$$2\phi_n^2(l) \;=\; 2\,\phi_n^2(k(n))\, p^{r\cdot k(n)}\, [\, A_n(r)\, B_n(r)\, C_n(r)\, D_n(r)\,]^2,$$

where

$$A_n(r) \;=\; \left(\frac{n-r-k(n)}{n-k(n)}\right)^{-n+r+k(n)+\frac{1}{2}} \qquad B_n(r) \;=\; \left(\frac{r+k(n)}{k(n)}\right)^{-k(n)-\frac{1}{2}}$$

$$C_n(r) \;=\; \left(\frac{n-k(n)}{r+k(n)}\, p^{\frac{k(n)}{2}}\right)^{r} \qquad D_n(r) \;=\; p^{\frac{r^2}{2}}$$

Note that $p^{r \cdot k(n)} = o(n^{-2r}(\log_b n)^{2r+\epsilon})$ for any fixed $\epsilon > 0$, and that $\phi_n^2(k(n)) \leq 1$ by the monotonicity of $\phi_n(\cdot)$ and the definition of $k(n)$. Thus it suffices to show that $A_n(r) \cdot B_n(r) \cdot C_n(r) \cdot D_n(r) = O(1)$ when n is sufficiently large. To begin, note that for any fixed $\delta \in (0, 1/2)$, when n is sufficiently large,

$$C_n(r)^{\frac{1}{r}} = \frac{n - k(n)}{r + k(n)} p^{\frac{k(n)}{2}} \leq \frac{n}{k(n)} p^{\frac{k(n)}{2}} \leq \frac{n}{(2-\delta)\log_b n} \frac{\frac{2+\delta}{2}\log_b n}{n},$$

which is less than one. Note that $B_n(r) \leq 1$. It only remains to show $A_n(r) \cdot D_n(r) = O(1)$. Simple calculations yield that $\ln A_n(r) \leq r$. Consequently, $\ln A_n(r) \cdot D_n(r) \leq r - \frac{r^2}{2}$, which is bounded from above. ■

8 Proof of Theorem 1

The proof of Theorem 1 is established via a sequence of technical lemmas. Modifying our earlier notation slightly, let $U_k(n)$ denote the number of $k \times k$ submatrices of 1s in Z_n. In what follows ϵ is a fixed positive number less than $\frac{1}{2}$. Our argument parallels that outlined in Bollobás [5]. We begin with the following definition.

Definition: For each $k \geq 1$, let n'_k be the least integer n such that

$$EU_k(n) \geq k^{3+\epsilon}, \tag{14}$$

and let n_k be the largest integer n such that

$$EU_k(n) \leq k^{-3-\epsilon}. \tag{15}$$

Note that n_k and n'_k exist for sufficiently large $k \geq 1$, as $EU_k(k) = p^{k^2} \leq k^{-3-\epsilon}$, $EU_k(n)$ is monotone increasing in n, and $EU_k(n) \to \infty$ as $n \to \infty$.

Lemma 3. *Let n_k and n'_k be defined as above.*

a. When k is sufficiently large, $n'_k < n_{k+1}$.

b. When k is sufficiently large, $n'_k - n_k < C_1 \frac{n_k \log k}{k}$ for some constant $C_1 > 2$.

c. $\lim_{k\to\infty} \frac{n_{k+2} - n_{k+1}}{n_{k+1} - n_k} = b^{\frac{1}{2}}$.

Proof of (a): It follows from the definition of n_k that

$$\binom{n_k}{k} p^{\frac{k^2}{2}} \leq k^{-\frac{(3+\epsilon)}{2}} \quad \text{and} \quad \binom{n_k + 1}{k} p^{\frac{k^2}{2}} \geq k^{-\frac{(3+\epsilon)}{2}}. \tag{16}$$

Rearranging terms in the first inequality, and noting that $(n_k-k)!/n_k! \le (n_k-k)^{-k}$ we obtain, in turn, the inequalities

$$\frac{k^{\frac{(3+\epsilon)}{2}}}{k!\; b^{\frac{k^2}{2}}} \le \frac{1}{(n_k-k)^k} \quad \text{and} \quad n_k \le b^{\frac{k}{2}} \left[\frac{k!}{k^{\frac{(3+\epsilon)}{2}}}\right]^{\frac{1}{k}} + k.$$

Rearranging the terms in the second inequality of (16), one may establish by a similar argument the inequalities

$$k^{\frac{(3+\epsilon)}{2}} \;\ge\; b^{\frac{k^2}{2}} \frac{k!}{(n+1)^k} \quad \text{and} \quad n_k \;\ge\; b^{\frac{k}{2}} \left(\frac{k!}{k^{\frac{3+\epsilon}{2}}}\right)^{\frac{1}{k}} - 1.$$

Combining the two bounds on n_k above, yields

$$b^{\frac{k}{2}} \left(k!\, k^{-\frac{3+\epsilon}{2}}\right)^{\frac{1}{k}} - 1 \;\le\; n_k \;\le\; b^{\frac{k}{2}} \left(k!\, k^{-\frac{(3+\epsilon)}{2}}\right)^{\frac{1}{k}} + k \tag{17}$$

and the asymptotic relation

$$n_k \;=\; b^{\frac{k}{2}} (k!)^{\frac{1}{k}} + o(k\, b^{\frac{k}{2}}). \tag{18}$$

From the definition of n_k', one can establish in a similar fashion the inequalities

$$b^{\frac{k}{2}} \left(k!\; k^{\frac{3+\epsilon}{2}}\right)^{\frac{1}{k}} \;\le\; n_k' \;\le\; b^{\frac{k}{2}} \left(k!\; k^{\frac{(3+\epsilon)}{2}}\right)^{\frac{1}{k}} + k + 1. \tag{19}$$

and the asymptotic relation

$$n_k' = b^{\frac{k}{2}} (k!)^{\frac{1}{k}} + o(k\, b^{\frac{k}{2}}). \tag{20}$$

The asymptotic expressions for n_k and n_k' ensure that $n_k' < n_{k+1}$ when k is sufficiently large.

Proof of (b): It follows from inequalities (17) and (19) that, when k is sufficiently large,

$$\begin{aligned}
n_k' - n_k \;&\le\; b^{\frac{k}{2}} \left(k!\; k^{\frac{(3+\epsilon)}{2}}\right)^{\frac{1}{k}} + k + 1 - \left[b^{\frac{k}{2}} \left(k!\, k^{-\frac{3+\epsilon}{2}}\right)^{\frac{1}{k}} - 1\right] \\
&\le\; b^{\frac{k}{2}} \left(k!\, k^{-\frac{3+\epsilon}{2}}\right)^{\frac{1}{k}} (k^{\frac{3+\epsilon}{k}} - 1) + k + 2 \\
&\le\; (n_k+1)(k^{\frac{3+\epsilon}{k}} - 1) + k + 2 \\
&<\; n_k\, C_1 \frac{\log k}{k}.
\end{aligned} \tag{21}$$

for some constant $C_1 > 2$. The third inequality above is a consequence of (17), while the last inequality follows from the fact that $x - 1 < 2\ln x$ for x close to 1.

Proof of (c): It follows from equations (18) and (20) that

$$\frac{n_{k+1}}{n_k} = b^{\frac{1}{2}} + o(1) \quad \text{and} \quad \frac{n_{k+2}}{n_{k+1}} = b^{\frac{1}{2}} + o(1).$$

Therefore, as k tends to infinity,

$$\frac{n_{k+2}-n_{k+1}}{n_{k+1}-n_k} = \frac{\frac{n_{k+2}}{n_{k+1}}-1}{1-\frac{n_k}{n_{k+1}}} \to b^{\frac{1}{2}}. \tag{22}$$

This completes the proof of Lemma 3. ■

We now continue the analysis of $U_k(n)$. The second moment argument used below requires bounds on the ratio

$$g(U_k(n)) := \mathrm{Var}(U_k(n))/(EU_k(n))^2$$

which arises in a standard Chebyshev bound on the tails of $U_k(n)$. Letting

$$S = \{C = A \times B : A, B \subseteq [n], |A| = |B| = k\}$$

be the family of index sets of $k \times k$ submatrices, we see that

$$U_k(n)^2 \;=\; \sum_{C,C' \in S} I\{\text{each entry of } Z_n[C] \text{ and } Z_n[C'] \text{ is } 1\}$$

From the last display one may readily derive that

$$EU_k(n)^2 \;=\; \sum_{l=1}^{k} \binom{n}{k}\binom{k}{l}\binom{n-k}{k-l} \sum_{r=1}^{k} \binom{n}{k}\binom{k}{r}\binom{n-k}{k-r} \cdot p^{2\,k^2-lr},$$

where the indices k and l indicate the number of rows and columns, respectively, that the submatrices C and C' have in common. As $EU_k(n) = \binom{n}{k}^2 p^{k^2}$, we find that

$$g(U_k) \;=\; \sum_{l=0}^{k}\sum_{r=0}^{k} \frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} b^{lr} - 1,$$

where $b = p^{-1}$. Recall that $0 < \epsilon < \frac{1}{2}$ is fixed.

Lemma 4. *There exists a constant $C_0 > 0$ such that $g(U_k(n)) \leq C_0 k^{-1-\epsilon}$ for every sufficiently large k and every $n'_k \leq n \leq n_{k+1}$.*

Proof of Lemma 4: To begin, note that

$$\begin{aligned}
g(U_k(n)) \;&=\; \sum_{l=0}^{k}\sum_{r=0}^{k} \frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} (b^{lr}-1) \\
&=\; \sum_{l=1}^{k}\sum_{r=1}^{k} \frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} (b^{lr}-1) \\
&<\; \sum_{l=1}^{k}\sum_{r=1}^{k} \frac{\binom{k}{l}\binom{n-k}{k-l}}{\binom{n}{k}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} b^{lr} \leq \left(\sum_{r=1}^{k} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} (b^{r^2/2}) \right)^2,
\end{aligned}$$

where the last inequality follows from the fact that $b^{lr} \leq b^{\frac{l^2+r^2}{2}}$. Thus it suffices to show that

$$\sum_{r=1}^{k} h(r) = O(k^{-1/2-\epsilon/2}) \;\text{ where }\; h(r) := \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} b^{r^2/2}. \tag{23}$$

If $n \geq n'_k$, then by inequality (19), $n \geq b^{\frac{k}{2}}\left(k!\, k^{\frac{3+\epsilon}{2}}\right)^{\frac{1}{k}}$, which implies that $k \leq 2\log_b n$. Similarly, inequality (17) implies that if $n \leq n_{k+1}$ then $k \geq (2-\eta)\log_b n$ for some fixed $0 < \eta < 1/2$. Finally, it follows from the assumption that $n \geq n'_k$ and the definition of n'_k that $\binom{n}{k} p^{\frac{k^2}{2}} = \sqrt{EU_k(n)} \geq \sqrt{EU_k(n'_k)} \geq k^{3/2+\epsilon/2}$. Using these inequalities, one can upper bound $h(1)$, $h(k-1)$ and $h(k)$ as follows:

$$\begin{aligned}
h(1) &= \frac{\binom{k}{1}\binom{n-k}{k-1}}{\binom{n}{k}} b^{1/2} = \frac{b^{1/2}k^2(n-k)!(n-k)!}{(n-2k+1)!n!} < \frac{b^{1/2}k^2}{n-k} = O(k^2 b^{-k/2}),\\
h(k-1) &= \frac{k(n-k)}{\binom{n}{k}} b^{\frac{k^2}{2}-k+\frac{1}{2}} \leq \frac{knb^{\frac{1}{2}-k}}{\sqrt{EU_k(n)}} = O\left(k^{-1/2-\epsilon/2}\, b^{-k(1-\eta)/(2-\eta)}\right)\\
h(k) &= \frac{b^{\frac{k^2}{2}}}{\binom{n}{k}} = \frac{1}{\sqrt{EU_k(n)}} \leq k^{-3/2-\epsilon/2},
\end{aligned}$$

In order to establish inequality (23), it therefore suffices to verify that when k is sufficiently large,

$$h(r) \leq h(1) + h(k-1) \tag{24}$$

for any $1 < r < k-1$. To this end, note that

$$\frac{h(r+1)}{h(r)} = \frac{(k-r)^2 b^{r+\frac{1}{2}}}{(r+1)\,(n-2k+r+1)}.$$

When $r \leq \frac{1}{3}k$, the inequality $k \leq 2\log_b n$ implies that

$$\frac{h(r+1)}{h(r)} \leq \frac{bk^2 b^{\frac{k}{3}}}{n-2k+r+1} \leq \frac{bk^2 n^{\frac{2}{3}}}{n-2k+r+1} < 1.$$

When $\frac{2}{3}k \leq r < k-1$ the inequality $k \geq (2-\eta)\log_b n$ with $0 < \eta < 1/2$ implies that

$$\frac{h(r+1)}{h(r)} \geq \frac{b^{\frac{2k}{3}}}{k(n+r+1)} \geq \frac{n^{\frac{2(2-\eta)}{3}}}{k(n+r+1)} > 1.$$

In order to show inequality (24), it now suffices to show that $h(r)$ is log-convex for all integer $r \in [\lceil\frac{k}{3}\rceil - 1, \lceil\frac{2k}{3}\rceil]$. Since for $r \in [\lceil\frac{k}{3}\rceil - 1, \lceil\frac{2k}{3}\rceil]$,

$$\ln h(r) = \ln h(\lceil\frac{k}{3}\rceil - 1) + \sum_{i=0}^{r-\lceil\frac{k}{3}\rceil} \ln \frac{h(\lceil\frac{k}{3}\rceil + i)}{h(\lceil\frac{k}{3}\rceil + i - 1)}.$$

it is equivalent to show that $\frac{h(r+1)}{h(r)}$ is monotone increasing. To verify this, note that the derivative $\partial[h(r+1)/h(r)]/\partial r$ is equal to

$$b^{\frac{2r+1}{2}}\left[\frac{-2(k-r)(r+1)(n-2k+r+1)-(k-r)^2 2(2r+n-2k+2)}{(r+1)^2(n-2k+r+1)^2}+\frac{(k-r)^2\ln b}{(r+1)(n-2k+r+1)}\right]$$

$$= \frac{b^{\frac{2r+1}{2}}(k-r)}{(r+1)(n-2k+r+1)}\left[\frac{-2(r+1)(n-2k+r+1)-(k-r)(2r+n-2k+2)}{(r+1)(n-2k+1)}+(k-r)\ln b\right].$$

When k is sufficiently large and $n \gg k > r$, the sum of the leading terms on the last expression above is

$$-2n(r+1)-(k-r)n+(k-r)(r+1)n\ln b \;=\; n(-r^2\ln b+kr\ln b-k-r+(k-r)\ln b-2).$$

By plugging in $r=\frac{k}{3}$ and $r=\frac{2k}{3}$, it is not hard to check that this quadratic form in r is positive for any $r \in [\lceil\frac{k}{3}\rceil-1, \lceil\frac{2k}{3}\rceil]$ when k is sufficiently large, and the desired monotonicity follows. ■

Lemma 5. *With probability one, when k is sufficiently large, $M(Z_n)=k$ whenever $n'_k \le n \le n_{k+1}$.*

Proof of Lemma 5: By the definition of n_{k+1} and Markov's inequality, when $n \le n_{k+1}$,

$$P(M(Z_n)>k) \;\le\; E(U_{k+1}(n)) \;\le\; E(U_{k+1}(n_{k+1})) \;\le\; (k+1)^{-3-\epsilon}. \tag{25}$$

Moreover, Chebyshev's inequality and Lemma 4 together imply that for $n'_k \le n \le n_{k+1}$,

$$P(M(Z_n)<k)=P(U_k(n)=0) \;\le\; \frac{Var(U_k(n))}{(EU_k(n))^2} \;\le\; C_0\cdot k^{-1-\epsilon}. \tag{26}$$

As $M(Z_n)$ is monotone increasing with n, the previous bounds yield

$$\begin{aligned}\sum_{k\ge1}P\left(\bigcup_{n=n'_k}^{n_{k+1}}\{M(Z_n)\neq k\}\right) &\le \sum_{k\ge1}P\left(M(Z_{n'_k})<k\right)+\sum_{k\ge1}P\left(M(Z_{n_{k+1}})\ge k\right)\\ &\le \sum_{k\ge1}\left(C_0\cdot k^{-1-\epsilon}+\frac{1}{k^{3+\epsilon}}\right)<\infty.\end{aligned}$$

and the result follows from the Borel-Cantelli lemma. ■

Proof of Theorem 1: From Lemma 5 we may deduce that with probability one $M(Z_n)$ is eventually equal to one of two possible consecutive integers, whose values depend only on n. It follows from their definition that $n_k < n'_k$, and by Lemma 3 both integers tend to infinity as k tends to infinity. Therefore for every k greater than or equal to some k_0 we have

$$\ldots < n_k < n'_k < n_{k+1} < n'_{k+1} < \ldots.$$

Thus for all $n \geq n_{k_0}$ there exists a unique integer k (depending on n) such that $n'_k \leq n \leq n_{k+1}$ or $n_k < n < n'_k$. In the former case, Lemma 5 implies that $M(Z_n) = k$ when n is sufficiently large. In the latter case, Lemma 5 and the monotonicity of $M(Z_n)$ in n imply that

$$k - 1 = M(Z_{n_k}) \leq M(Z_n) \leq M(Z_{n'_k}) = k, \tag{27}$$

when n is sufficiently large, so that $M(Z_n)$ can take one of at most two possible values, $k-1$ and k.

It remains to connect $M(Z_n)$ and $s(n)$. To begin, let n be such that $n'_k \leq n \leq n_{k+1}$ for some $k \geq k_0$. Then by definition of n_{k+1} and $s(n)$,

$$(1 + o(1))\, \phi_n(k+1) \;=\; (EU_{k+1}(n))^{1/2} \;\leq\; (EU_{k+1}(n_{k+1}))^{1/2} \;\leq\; k^{-3/2-\epsilon/2} \;<\; 1 \;=\; \phi_n(s(n)).$$

As $\phi_n(k)$ is monotone decreasing in k, we conclude that $s(n) < k+1$ when n is sufficiently large. Similarly,

$$(1 + o(1))\, \phi_n(k) \;=\; (E\, U_k(n))^{1/2} \;\geq\; (E\, U_k(n'_k))^{1/2} \;\geq\; k^{3/2+\epsilon/2} \;>\; 1 \;=\; \phi_n(s(n)),$$

which implies $s(n) > k$. Thus, with probability one, when n is sufficiently large

$$n'_k \leq n \leq n_{k+1} \text{ implies } k < s(n) < k+1 \text{ and } M(Z_n) = k. \tag{28}$$

Suppose now that $n_k \leq n \leq n'_k$. Then $s(n_k) \leq s(n) \leq s(n'_k)$ and the arguments above show that $s(n_k) < k$ and $s(n'_k) > k$. We establish that $s(n'_k) - s(n_k) = o(1)$. To this end, note that

$$0 < s(n'_k) - s(n_k) \;=\; 2\log_b \frac{n'_k}{n_k} - 2\log_b \frac{\log_b n'_k}{\log_b n_k} + o(1) \;\leq\; 2\log_b \frac{n'_k}{n_k} + o(1)$$

as $\frac{\log_b n'_k}{\log_b n_k} > 1$. It therefore suffices to show that $\log_b \frac{n'_k}{n_k} = o(1)$, but this follows from part (b) of Lemma 3. Putting the bounds above together with Lemma 5, we find that with probability one, when n is sufficiently large

$$n_k \leq n \leq n'_k \text{ implies } k - \epsilon < s(n) < k + \epsilon \text{ and } M(Z_n) \in \{k-1, k\}. \tag{29}$$

Combining relations (28) and (29) yields the desired bound on $M(Z_n)$. ■

9 Proof of Theorem 4

The following lemmas are used in the proof of Theorem 4. Lemma 6 shows that $|\hat{C}|$ is greater than or equal to $|C^*|$ with high probability, and Lemma 9 shows that $\hat{C}$ can only contain a small proportion of entries outside C^*. Lemma 7 and Lemma 8 are used in the proof of Lemma 9.

Lemma 6. *Under the conditions of Theorem 4,* $P\left(|\hat{C}| < l^2\right) \leq \Delta_1(l)$.

Proof of Lemma 6: Let $u_{1*}, ..., u_{l*}$ be the rows of C^* in Y, and let V be the number of rows satisfying $F(u_{i*}) < \tau = 1 - p_0$. By the union bound and a standard bound [9] on the tail of the binomial distribution, $P(V \geq 1) \leq l \cdot e^{-\frac{3l(p-p_0)^2}{8p}}$. The same inequality holds for the number V' of columns u_{*j} of C^* such that $F(u_{*i}) < 1 - p_0$. Since $\{|\hat{C}| < l^2 = |C^*|\} \subset \{C^* \notin AFI_\tau(Y)\} \subset \{V \geq 1\} \cup \{V' \geq 1\}$, we have

$$\begin{aligned} P\{|\hat{C}| < l^2\} &\leq P(V \geq 1) + P(V' \geq 1) \\ &\leq 2le^{-\frac{3}{8p}l(p-p_0)^2} = \Delta_1(l). \quad \blacksquare \end{aligned}$$

Lemma 7. *Given* $0 < \tau_0 < 1$*, if there exists a* $k \times r$ *binary matrix* V *such that* $F(V) \geq \tau_0$*, then there exists a* $v \times v$ *submatrix* U *of* V *such that* $F(U) \geq \tau_0$*, where* $v = \min\{k, r\}$.

Proof of Lemma 7: Without loss of generality, assume $v = k \leq r$. Order the columns of V in descending order of the number of 1s they contain. If U contains the first v columns in this order, then $F(U) \geq \tau_0$. $\blacksquare$

Lemma 8. *Let* $1 < \gamma < 2$*. Let* W *be a binary matrix, and let* R_1 *and* R_2 *be two square submatrices of* W *such that (i)* $|R_2| = k^2$*, (ii)* $|R_1 \backslash R_2| > k^\gamma$ *and (iii)* $R_1 \in AFI_\tau(W)$*. Then when* k *is sufficiently large there exists a square submatrix* $D \subset R_1 \backslash R_2$ *such that* $|D| \geq k^{2\gamma-2}/16$ *and* $F(D) \geq \tau$.

Proof of Lemma 8: The result is clearly true if $R_1 \cap R_2 = \emptyset$, so we assume that R_1 and R_2 overlap After suitable row and column permutations, $R_1 \backslash R_2$ can be expressed either as a single maximal rectangular submatrix W_1, or as the union of two overlapping maximal rectangular $W_1 \cup W_2$. (A submatrix W of $R_1 \backslash R_2$ is maximal if there is no other submatrix of $R_1 \backslash R_2$ that contains it.)

Case 1: R_1 **and** R_2 **overlap on an edge.** Suppose that the difference $R_1 \backslash R_2$ can be expressed as a single rectangular submatrix W_1. Let l_1 and l_2 be the side lengths of W_1. In this case, the side length of the square submatrix R_1 must be less than k, and consequently $\max(l_1, l_2) \leq k$. Since $|R_1 \backslash R_2| \geq k^\gamma$, it follows that $\min(l_1, l_2) \geq k^{\gamma-1}$. As $R_1 \in AFI_\tau(W)$ we have $F(W_1) \geq \tau$. By Lemma 7, there exists a $v \times v$ submatrix D of W_1 such that $F(D) \geq \tau$ and $v \geq \min(l_1, l_2) \geq k^{\gamma-1}$.

Case 2: R_1 **and** R_2 **overlap on a corner.** Suppose $R_1 \backslash R_2$ is the union $W_1 \cup W_2$ of two maximal rectangular submatrices. Then clearly $\max(|W_1|, |W_2|) \geq \frac{|R_1 \backslash R_2|}{2}$. Without loss of generality, we assume that $|W_1| \geq |W_2|$. As $R_1 \in AFI_\tau(W)$, $F(W_1) \geq \tau$, and it suffices by Lemma 7 to show that the length of the shorter side of W_1 is greater than $k^{\gamma-1}/4$.

Let $l_1 \leq l_2$ be the side lengths of W_1 and suppose for the moment that $l_1 < k^{\gamma-1}/4$. Then $l_2 > \frac{|R_1 \backslash R_2|}{2k^{\gamma-1}/4}$ and $|R_1| = l_2^2 \geq \frac{|R_1 \backslash R_2|^2}{k^{2\gamma-2}/4}$, and it follows that

$$|R_1 \backslash R_2| \;\geq\; |R_1| - |R_2| \;>\; \frac{|R_1 \backslash R_2|^2}{k^{2\gamma-2}/4} - k^2.$$

Dividing both sides of the previous inequality by $|R_1 \backslash R_2|$ and using the assumption $|R_1 \backslash R_2| \geq k^\gamma$ yields

$$1 \;>\; \frac{|R_1 \backslash R_2|}{k^{2\gamma-2}/4} - \frac{k^2}{|R_1 \backslash R_2|} \;\geq\; 4k^{(2-\gamma)} - k^{(2-\gamma)} \;=\; 3k^{(2-\gamma)}.$$

When k is sufficiently large, this yields a contradiction and completes the proof. ■

Lemma 9. *Let $\mathcal{A}$ be the collection of $C \in \hat{\mathcal{C}}$ such that $|C| \geq l^2$ and $\frac{|C \cap C^{*c}|}{|C|} \geq \alpha$, where $\alpha \in (0,1)$ satisfies $l \geq 8\alpha^{-1}(\log_b n + 2)$. Let A be the event that $\mathcal{A} \neq \emptyset$. If n is sufficiently large,*

$$P(A) \;\leq\; \Delta_2(\alpha, l).$$

Proof of Lemma 9: Recall that $|C^*| = l^2$. If $C \in \mathcal{A}$ then $C \in \mathrm{AFI}_{1-p_0}(Y)$ and

$$|C \backslash C^*| \;=\; |C| \;\cdot\; \frac{|C \cap C^{*c}|}{|C|} \;\geq\; l^2 \;\cdot\; \alpha \;=\; l^\gamma$$

where $\gamma \;=\; 2 + \log_l \alpha$. Thus, by Lemma 8 there exists a $v \times v$ submatrix D of $C \backslash C^*$ such that $F(D) \;\geq\; 1 - p_0$ and $v \;\geq\; \frac{\alpha l}{4}$. It follows that

$$\max_{c \in \hat{\mathcal{C}}} M^\tau(C \cap C^{*c}) \;\geq\; v \;\geq\; \frac{\alpha l}{4},$$

where $\tau = 1 - p_0$ and $M^\tau(X)$ is size of the largest square submatrix with average greater than τ in a given matrix X.

Let $W = W(Y, C^*)$ be an $n \times n$ binary random matrix, with $w_{ij} = y_{ij}$ if $(i,j) \notin C^*$, and $w_{ij} \sim \mathrm{Bern}(p)$ otherwise. Then it is clear that

$$M^\tau(W) \;\geq\; \max_{c \in \hat{\mathcal{C}}} M^\tau(C \cap C^{*c}) \;\geq\; \frac{\alpha l}{4}.$$

When n is sufficiently large and $l \geq 8\alpha^{-1}(\log_b n + 2)$, we can bound $P(A)$ as follows

$$\begin{aligned} P(A) \quad &\leq \quad P(\max_{c \in \hat{\mathcal{C}}} M^\tau(C \cap C^{*c}) \;\geq\; \frac{\alpha l}{4}) \\ &\leq \quad P(M^\tau(W) \geq \frac{\alpha l}{4}) \;\leq\; 2n^{-(\alpha l/4 - 2\log_{b'} n)}, \end{aligned} \tag{30}$$

where $b' = e^{\frac{3(1-p_0-p)^2}{8p}}$. Note that the last inequality follows from a first moment argument similar to that in the proof of Proposition 1 and a standard inequality for the tails of the binomial distribution(*c.f.* Problem 8.3 of [9]). As $p_0 > p$, $b < b'$, and consequently one can bound the right hand side of inequality (30) by $\Delta_2(\alpha, l)$. For detailed proof of inequality (30), please refer to Proposition 3.3.1 in [27]. ■

Proof of Theorem 4: Let E be the event that $\{\Lambda \leq \frac{1-\alpha}{1+\alpha}\}$. It is clear that E can be expressed as the union of two disjoint events E_1 and E_2, where

$$E_1 = \{|\hat{C}| < |C^*|\} \cap E \text{ and } E_2 = \{|\hat{C}| \geq |C^*|\} \cap E$$

One can bound $P(E_1)$ by $\Delta_1(l)$ via Lemma 6.

It remains to bound $P(E_2)$. By the definition of Λ, the inequality $\Lambda \leq \frac{1-\alpha}{1+\alpha}$ can be rewritten equivalently as

$$1 + \frac{|\hat{C} \cap C^{*c}|}{|\hat{C} \cap C^*|} + \frac{|\hat{C}^c \cap C^*|}{|\hat{C} \cap C^*|} \geq \frac{1+\alpha}{1-\alpha}.$$

When $|\hat{C}| \geq |C^*|$, one can verify that $|\hat{C} \cap C^{*c}| \geq |\hat{C}^c \cap C^*|$, which implies that

$$1 + \frac{|\hat{C} \cap C^{*c}|}{|\hat{C} \cap C^*|} + \frac{|\hat{C}^c \cap C^*|}{|\hat{C} \cap C^*|} \leq 1 + 2\frac{|\hat{C} \cap C^{*c}|}{|\hat{C} \cap C^*|}.$$

Therefore, $E_2 \subset E_2^*$, where

$$\begin{aligned} E_2^* &= \{|\hat{C}| \geq |C^*|\} \cap \left\{1 + 2\frac{|\hat{C} \cap C^{*c}|}{|\hat{C} \cap C^*|} \geq \frac{1+\alpha}{1-\alpha}\right\} \\ &\subset \{|\hat{C}| \geq l^2\} \cap \left\{1 + 2\frac{|\hat{C} \cap C^{*c}|}{|\hat{C} \cap C^*|} \geq \frac{1+\alpha}{1-\alpha}\right\}. \end{aligned}$$

Notice that $1 + 2\frac{|\hat{C} \cap C^{*c}|}{|\hat{C} \cap C^*|} \geq \frac{1+\alpha}{1-\alpha}$ implies $\frac{|\hat{C} \cap C^{*c}|}{|\hat{C}|} \geq \alpha$. Therefore, by Lemma 9, $P(E_2^*) \leq \Delta_2(\alpha, l)$. ■

Acknowledgements: The authors would like to thank Professor Robin Pemantle for helpful discussions about, and comments on, early versions of this work, and an anonymous referee whose comments led to a simpler proof, and improved statement, of Theorem 1. The work described in this paper was supported in part by NSF grant DMS 0406361.

References

[1] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, 94-105, 1998.

[2] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. Verkamo. Fast discovery of association rules. In U. M. Fayyad, U. and et. al, editors, *Advances in Knowledge Discove and Data Mining*, AAAI Press, Chapter 12, 307-328, 1996.

[3] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, 207-216, 1993.

[4] N. Alon, J. Spencer. *The Probabilistic Method*. John Wiley. New York, 1991.

[5] B. Bollobás. *Random Graphs*. 2nd ed., Cambridge University Press, 2001.

[6] B. Bollobás, P. Erdős. Cliques in Random Graphs. In *Mathematical Proceedings of the Cambridge Philosophy Society*, 80: 419-427, 1976.

[7] Y. Cheng and G. M. Church. Biclustering of expression data. In *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology*, 93-103, 2000.

[8] M. Dawande, P. Keskinocak, J. Swaminathan, and S. Tayur. On bipartite and multipartite clique problems. *Journal of Algorithms*, 41: 388-403, 2001.

[9] L. Devroye, L. Gyorfi, and G, Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, New York, 1996.

[10] M. R. Garey and D. S. Johnson. *Computers and Intractability, A guide to the theory of NP-completeness*. Freeman, San Francisco, 1979.

[11] B. Goethals. Survey on Frequent Pattern Mining. w w w. a d r e m. u a.a c.b e/ g o e t h a l s/ s o f t w a r e/ s u r v e y. p d f.

[12] J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, 1-12, 2000.

[13] D. J. Hand, H. Mannila and P. Smyth. *Principles of Data Mining*. MIT Press, 2001.

[14] D. S. Hochbaum. Approximating clique and biclique problems. *Journal of Algorithms*, 29(1): 174-200, 1998.

[15] M. Koyutürk, W. Szpankowski and A. Grama. Biclustering Gene-Feature Matrices for Statistically Significant Dense Patterns. In *Proceedings of the 8th Annual International Conference on Research in Computational Molecular Biology*, 480-484,2004.

[16] J. Liu, S. Paulsen, X. Sun, W. Wang, A.B. Nobel, and J. Prins. Mining approximate frequent itemsets in the presence of noise: algorithm and analysis. In *Proceedings of the SIAM International Conference on Data Mining*, 405-416, 2006.

[17] J. Liu, S. Paulsen, W. Wang, A. B. Nobel, and J. Prins. Mining Approximate Frequent Itemsets from Noisy Data. In *Proceedings of the IEEE International Conference on Data Mining*, 721-724, 2005.

[18] S. Madeira and A. Oliveira. Biclustering Algorithms for Biological Data Analysis: A Survey. *IEEE Transactions on Computational Biology and Bioinformatics*, 1(1): 24-45, 2004.

[19] D. Matula. The largest clique size in a random graph. *Technical Report, Southern Methodist University,* CS 7608, 1976.

[20] N. Mishra, D. Ron and R. Swaminathan. A New Conceptual Clustering Framework. *Machine Learning*. 56(1-3): 115-151, 2004.

[21] G. Park and W. Szpankowski. Analysis of Biclusters with Applications to gene Expression Data. In *Proceedings of Conference on Analysis of Algorithms*, 267-274, 2005.

[22] R. Peeters. The maximum edge biclique problem is NP-complete. *Discrete Applied Mathematics*, 131(3): 651-654, 2003.

[23] J. Pei, G. Dong, W. Zou, and J. Han. Mining Condensed Frequent-Pattern Bases. *Knowledge and Information Systems*, 6(5), 2002.

[24] J. Pei, A. K. Tung, and J. Han. Fault-tolerant frequent pattern mining: Problems and challenges. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, 2001.

[25] J. D. Reuning-Scherer. Mixture Models for Block Clustering. *Ph.D. Thesis, Yale university,* 1997.

[26] J. K. Seppänen, and H. Mannila. Dense Itemsets. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 683-688, 2004.

[27] X. Sun. Significance and recovery of block structures in binary and real-valued matrices with noise. *Ph.D. Thesis, UNC Chapel Hill*, 2007.

[28] A. Tanay, R. Sharan, and R. Shamir. Dicovering statistically significant biclusters in gene expression data. *Bioinformatics*, 18: 136-144, 2002

[29] A. Tanay, R. Sharan and R. Shamir. Biclustering Algorithms: A Survey. *Handbook of Computational Molecular Biology*, Chapman & Hall/CRC, Computer and Information Science Series, in press, 2005.

[30] C. Yang, U. Fayyad, and P. S. Bradley. Efficient discovery of error-tolerant frequent itemsets in high dimensions. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2001.