Maximal Quasi-Bicliques with Balanced Noise Tolerance: Concepts and Co-clustering Applications

Jinyan Li* Kelvin Sim† Guimei Liu‡ Limsoon Wong§

Abstract

The rigid all-versus-all adjacency required by a maximal biclique for its two vertex sets is extremely vulnerable to missing data. In the past, several types of *quasi-bicliques* have been proposed to tackle this problem, however their noise tolerance is usually unbalanced and can be very skewed. In this paper, we improve the noise tolerance of maximal quasi-bicliques by allowing every vertex to tolerate up to the same number, or the same percentage, of missing edges. This idea leads to a more natural interaction between the two vertex sets—a balanced most-versus-most adjacency. This generalization is also non-trivial, as many large-size maximal quasi-biclique subgraphs do not contain any maximal bicliques. This observation implies that direct expansion from maximal bicliques may not guarantee a complete enumeration of all maximal quasi-bicliques. We present important properties of maximal quasi-bicliques such as a bounded closure property and a fixed point property to design efficient algorithms. Maximal quasi-bicliques are closely related to co-clustering problems such as documents and words co-clustering, images and features co-clustering, stocks and financial ratios co-clustering, etc. Here, we demonstrate the usefulness of our concepts using a new application—a bioinformatics example—where prediction of true protein interactions is investigated.

Keywords: Maximal quasi-bicliques, maximal bicliques, balanced noise tolerance, co-clustering applications, prediction of missing protein-protein interactions.

1 Introduction

Maximal bicliques, also known as maximal complete bipartites, are a classical concept in graph theory [9, 2, 16]. A biclique in a graph consists of two disjoint vertex subsets between which every vertex is adjacent to all vertices in the other subset, exhibiting a type of *all-versus-all* interaction (connection). A biclique H is maximal in a graph G if and only if there is no other biclique in G that contains H. This strict all-versus-all interaction requirement makes maximal bicliques extremely vulnerable to missing data—if any edge in a maximal biclique is missing, the resulting subgraph is not a maximal biclique any more. Also, this strong requirement prevents the discovery of bipartite subgraphs that exhibit a *most-versus-most* interaction between the two vertex subsets, which may be a more natural interaction in many real-life situations.

We generalize here the concept of maximal bicliques by relaxing the all-versus-all interaction requirement to meet demands from real-life applications such as information retrieval, biological and financial data mining problems where missing and noisy data are common. Specifically, we introduce two types of most-versus-most interactions, and we call such subgraphs *maximal quasi-bicliques*.

Our first idea is to permit every vertex in the two vertex subsets to *disconnect* from up to μ number of vertices in the other subset, but at the same time this vertex must be adjacent to at least μ number of them. The second idea is to allow each vertex in the two vertex subsets to disconnect from up to a small fraction ($\delta\%$) of vertices in the other subset. Both ideas emphasize a balanced noise tolerance. "Balanced" is in the sense that "all" vertices are freed to accommodate missing edges or disconnections up to the same level of degree. Thus, our maximal quasi-bicliques can avoid skewed distribution of missing edges, which cannot be achieved by those quasi-bicliques proposed in the literature [1, 4, 28, 18, 10, 25]. For example, the ϵ-bicliques [18] allow only one side of the vertices to tolerate the same degree of missing edges, thus this concept achieves only an asymmetric form of noise tolerance. The α-quasi-bicliques [28] allow only some of the vertices in the graph to have the same degree of noise tolerance, and thus some other vertices may have a very low connectivity. Our previous definition

*School of Computer Engineering, Nanyang Technological University. Email: jyli@ntu.edu.sg.

†Institute for Infocomm Research, A*STAR (Agency for Science, Technology and Research), Singapore. Email: shsim@i2r.a-star.edu.sg.

‡School of Computing, National University of Singapore. Email: liugm@comp.nus.edu.sg.

§School of Computing, National University of Singapore. Email: wongls@comp.nus.edu.sg.

of maximal ϵ-quasi-bicliques as proposed in [25] causes the problem that a set of fully isolated vertices can form exponential number of maximal quasi-bicliques. Other works such as eigenvector-based quasi-bicliques [4], and dense bipartite cliques [1, 10] all require only a minimum global density of the edges—thus some specific vertex may, again, have a very low connectivity, even a possible zero-connectivity.

It is interesting that our newly defined maximal quasi-bicliques in a graph G sometimes do not contain any maximal biclique of G. This indicates that maximal quasi-biclique subgraphs of a graph can be categorized into two non-overlapping types: those that contain a maximal biclique and those do not contain any. Therefore, the introduction of the concept of maximal quasi-bicliques can bring up novel subgraph patterns that can never be covered by the definition of maximal bicliques, or never be extended to. Hence, the change from the all-versus-all to most-versus-most is a non-trivial generalization for maximal bicliques.

As a maximal quasi-biclique in a graph G does not always contain a maximal biclique of G, a direct expansion from all maximal bicliques may not give a full answer for enumerating all maximal quasi-bicliques of a graph. Enumerating all maximal quasi-bicliques from a graph is thus technically non-trivial, even when all maximal bicliques are known. For example, algorithms reported in [24, 2, 14, 8] for enumerating maximal bicliques are not directly usable. The algorithm shown in [28] enumerates only α-quasi-bicliques that are subgraphs generated from maximal bicliques by adding their maximal α-extensions. So, it is not usable either. We study some properties of maximal quasi-bicliques such as a bounded closure property and a fixed point theorem, and make use of them to develop a new algorithm to enumerate all maximal quasi-bicliques from large graphs.

Maximal bicliques and maximal quasi-bicliques are closely related to many co-clustering applications. For example, they are related to documents-and-words, stocks-and-financial ratios, or images-and-features co-clustering in the information retrieval field [6, 7, 22, 18, 25]. They are also related to web community mining [23, 20] and many bioinformatics studies such as interacting protein groups' discovery [13, 4, 19], disease and genes co-clustering [12], and phylogenetic tree construction [28, 24, 8]. The computational problem there is to find a maximal number of object entities that are contained in a maximal number of attribute entities, given the binary containment relation of all the entities. As our definition generalizes the concept of maximal bicliques, and it is more comprehensive than the maximal quasi-bicliques used before, a straightforward application of our newly proposed concept is for finding balanced noise-tolerance co-clusterings from the binary containment relation databases. To show the high potential of our maximal quasi-bicliques, we suggest a new application—the prediction of true interactions for proteins in a cell.

Contribution of this paper:

- Two types of maximal quasi-bicliques that have balanced noise tolerance are introduced. These new maximal quasi-bicliques can effectively overcome the main limitations of existing maximal quasi-bicliques—the skewed noise tolerance that often leads to some edges with very low connectivity.
- Our maximal quasi-bicliques are not a trivial generalization of the classical maximal bicliques because there exist many large size maximal quasi-bicliques that cannot be directly expanded from any maximal biclique. We present proof and examples to verify this. We also present a bounded-closure property and a fixed point theory for our maximal quasi-bicliques.
- Our comprehensive experimental results obtained from benchmark and real-life graph data sets show that Mishra's ϵ-bicliques [18], the most related maximal quasi-bicliques, can easily suffer from skewed noise tolerance, and that Mishra's algorithm is significantly slower than our algorithm. We also introduce a new co-clustering application for bioinformatics where protein interactions can be predicted by using our maximal quasi-bicliques.

In the following sections, we review some basic definitions for maximal bicliques, and then formally define our maximal quasi-bicliques. We elaborate these new ideas with examples, properties, and theorems in Section 4. To find our maximal quasi-bicliques, we present a modified version of the *completeQB* [25] algorithm in Section 5. Detailed literature work review are presented in Section 6. Finally, we report experimental results and give a conclusion.

2 Background on Maximal Bicliques

An *undirected graph* G is a pair $\langle V, E\rangle$, where V is a set of vertices and $E \subseteq V \times V$[1] is a set of edges between the vertices. Two vertices are *adjacent* or *connected* if there is an edge between them. The *neighbourhood* $\beta(v, G)$ of a vertex v in $G = \langle V, E\rangle$ is the set of vertices adjacent to v, denoted $\beta(v, G) = \{u \mid \{u, v\} \in E\}$. The neighbourhood $\beta(X, G)$ of a set of vertices X in

[1] As the graph is undirected, for convenience, we define $V \times V$ as the set $\{\{u, v\} \mid u \in V, v \in V\}$.

$G = \langle V, E\rangle$ is the set of vertices adjacent to every vertex in X; that is, $\beta(X, G) = \bigcap_{v\in X} \beta(v, G) = \{u \mid u \in V,$ and $X \subseteq \beta(u, G)\}$.

A graph can be equivalently described by its adjacency matrix. Let $G = \langle V, E\rangle$ be a graph with $V = \{v_1, v_2, ..., v_p\}$. The *adjacency matrix* $\mathbf{A}$ of G is the $p \times p$ matrix defined by

$$\mathbf{A}[i, j] = \begin{cases} 1 & \text{if } \{v_i, v_j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

A graph $G' = \langle V', E'\rangle$ is a *subgraph* of a graph $G = \langle V, E\rangle$ if $V' \subseteq V$ and $E' \subseteq E$. If $V' \subset V$ or $E' \subset E$, we say G' is a proper subgraph of G. If G' is a subgraph of G, we say G is a superset graph of G', or G contains G'.

A graph $G = \langle V, E\rangle$ is a *bipartite* if its vertex set V can be partitioned into two disjoint nonempty sets V_1 and V_2, and every edge in E connects a vertex in V_1 and a vertex in V_2. So, there is no edge in E connecting two vertices within V_1 or two vertices within V_2. A bipartite G is often denoted as $G = \langle V_1, V_2, E\rangle$.

A bipartite $G = \langle V_1, V_2, E\rangle$ is called a *biclique* if, for every $v_1 \in V_1$ and $v_2 \in V_2$, there is an edge between v_1 and v_2. Thus the edge set E of a biclique $G = \langle V_1, V_2, E\rangle$ is completely determined by the two vertex sets V_1 and V_2. So we can omit the edge set and denote a biclique G simply as $G = \langle V_1, V_2\rangle$.

DEFINITION 2.1. (MAXIMAL BICLIQUE) *Let G' be a biclique subgraph of a graph G. If there does not exist any other biclique subgraph G'' of G such that G' is a proper subgraph of G'', then G' is a maximal biclique of G.*

Our previous work [14] has shown that efficiently listing a complete set of maximal bicliques from a graph G is equivalent to the mining of all closed patterns [21, 26, 11, 27] of the adjacency matrix of G. An earlier algorithm, called a concensus algorithm [2], can be also used to enumerate all maximal bicliques from a large graph. However, it is much slower than the method based on closed pattern mining as described in [14, 15].

3 New Concepts: μ-tolerance and δ%-tolerance Maximal Quasi-Bicliques

As mentioned, due to the strict all-versus-all requirement, maximal bicliques have zero-tolerance for missing and/or noise data. Here, we define *maximal quasi-bicliques* to generalize maximal bicliques, so that our maximal quasi-bicliques can accommodate missing data in a balanced manner. Our new concept has two ways for noise tolerance: (1) tolerating some absolute number of missing edges; and (2) tolerating a percentage of missing edges.

DEFINITION 3.1. (QUASI-BICLIQUE SUBGRAPH) *A bipartite subgraph $H = \langle V_1, V_2, E_H\rangle$ of a graph $G = \langle V, E\rangle$ is a* quasi-biclique *subgraph of G if $E_H = (V_1 \times V_2) \cap E$.*

Note that E_H is fully determined given V_1, V_2, and E. Thus we often abbreviate a quasi-biclique subgraph as $H = \langle V_1, V_2\rangle$ when the context G is clear.

DEFINITION 3.2. (**μ-tolerance maximal quasi-biclique**) *Let $H = \langle V_1, V_2\rangle$ be a quasi-biclique subgraph of G. Let μ be a small integer number. Then H is a* μ-tolerance quasi-biclique *subgraph of G if for each $v \in V_i$, $i = 1$ or 2,*

(i) v is disconnected from at most μ number of vertices in V_j, $j \neq i$, and

(ii) v is adjacent to at least μ number of vertices in V_j.

H is a μ-tolerance maximal quasi-biclique *subgraph of G if there is no other μ-tolerance quasi-biclique subgraph $H' = \langle V_1', V_2'\rangle$ of G such that $V_1' \supseteq V_1$ and $V_2' \supseteq V_2$.*

The condition (ii) "v is adjacent to at least μ number of vertices in V_j" is important. This requirement can prevent μ-tolerance maximal quasi-bicliques from some skewness of missing edges. For example, given a graph G consisting of n vertices but no edges, let $\mu = 2$, then any two disjoint pairs of vertices of G can form a μ-tolerance maximal quasi-biclique if the condition (ii) is not required. Observe that all these μ-tolerance maximal quasi-bicliques are useless. However, under our Definition 3.2, there is no such μ-tolerance maximal quasi-biclique in G. This is a subtle and critical difference between our definition and the one that we previously proposed in [25] for maximal quasi-bicliques, as the condition (ii) is not required by [25].

DEFINITION 3.3. (**δ%-tolerance maximal quasi-biclique**) *Let $H = \langle V_1, V_2\rangle$ be a quasi-biclique subgraph of G. Let δ% be a small percentage value. Then H is a* δ%-tolerance quasi-biclique *subgraph of G if for each $v \in V_i$, $i = 1$ or 2, v is disconnected from at most δ% number of vertices in $V_j, j \neq i$. H is a* δ%-tolerance maximal quasi-biclique *subgraph of G if there is no other δ%-tolerance quasi-biclique subgraph $H' = \langle V_1', V_2'\rangle$ of satisfying $V_1' \supseteq V_1$ and $V_2' \supseteq V_2$.*

Example. Figure 1(a) shows a graph G consisting of 6 vertices and 5 edges. The quasi-biclique subgraph $\langle\{v_1, v_2\}, \{v_5, v_6\}\rangle$ is a μ-tolerance maximal quasi-biclique subgraph of G for $\mu = 1$. However, it is not a δ%-tolerance maximal quasi-biclique for $\delta\% = 20\%$. Two 20%-tolerance maximal quasi-bicliques are $\langle\{v_1\}, \{v_4, v_5, v_6\}\rangle$ and $\langle\{v_6\}, \{v_1, v_2, v_3\}\rangle$, which are also maximal biclique subgraphs of G.

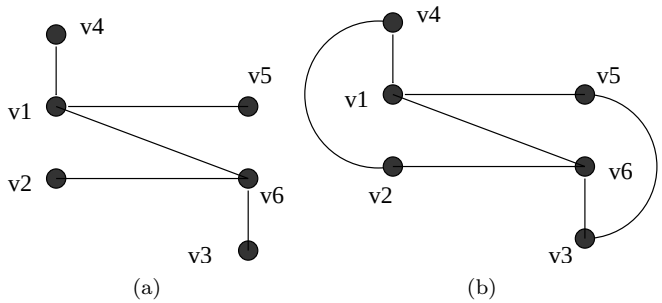


Figure 1: (a) Maximal quasi-bicliques contained in a graph G; (b) maximal bicliques all contained in a maximal quasi-biclique.

4 Properties

We give counter-examples to prove that a maximal quasi-biclique subgraph in a graph G can contain none of the maximal biclique subgraphs of G. We then present a *bounded closure* property for μ-tolerance maximal quasi-bicliques. We also prove that δ%-tolerance maximal quasi-bicliques do not have this property. We further show that both μ-tolerance and δ%-tolerance maximal quasi-bicliques have a *fixed point* property. All these properties are useful for discovering quasi-bicliques from a graph.

PROPOSITION 4.1. *A μ-tolerance or a δ%-tolerance maximal quasi-biclique subgraph H of a graph G does not always contain a maximal biclique subgraph of G.*

Proof. We use the graph G in Figure 1(a) again. It is known that $H = \langle\{v_1, v_2\}, \{v_5, v_6\}\rangle$ is a μ-tolerance maximal quasi-biclique subgraph of G for $\mu = 1$. It is also a δ%-tolerance maximal quasi-biclique for $\delta\% = 50\%$. The two and only maximal bicliques of G are: $M_1 = \langle v_1, \{v_4, v_5, v_6\}\rangle$ and $M_2 = \langle v_6, \{v_1, v_2, v_3\}\rangle$. Observe that neither M_1 nor M_2 is contained in H. So, this proposition is true.

In fact, this happens quite often in benchmark graphs. For example, for the c-fat200-1 and c-fat500-1 graph from the Second DIMACS Challenge benchmarks [2], many maximal quasi-biclique subgraphs do not contain any maximal biclique subgraphs. Detailed results are presented in Section 7.

PROPOSITION 4.2. *Let $H = \langle V_1, V_2\rangle$ be a maximal biclique of a graph G. Then (i) there exists at least one maximal quasi-biclique that contains H; and (ii) the total number of μ-tolerance maximal quasi-bicliques can be less than that of maximal bicliques for some μ.*

[2] Available at `ftp://dimacs.rutgers.edu/pub/challenge/graph/benchmarks/clique/`

Proof. For (i), by definition, $H = \langle V_1, V_2\rangle$ is a μ-tolerance or δ%-tolerance quasi biclique for any μ and δ. If no vertex in G can be added into V_1 or V_2, then H itself is a maximal quasi-biclique. So, there exists at least one maximal quasi-biclique that contains H.

For (ii), we prove by using an example. In Figure 1(b), there is one and only one μ-tolerance ($\mu = 2$) maximal quasi-biclique in graph $G = \langle\{v_1, v_2, v_3\}, \{v_4, v_5, v_6\}\rangle$, which is G itself. However, there are at least 8 maximal bicliques in this graph such as $\langle\{v_1\}, \{v_4, v_5, v_6\}\rangle$, $\langle\{v_1, v_2, v_3\}, \{v_6\}\rangle$, $\langle\{v_1, v_2\}, \{v_4, v_6\}\rangle$, etc. So, this graph contains less number of maximal quasi-bicliques than that of maximal bicliques.

This proposition says that every maximal biclique of a graph is contained in at least one maximal quasi-biclique, but the total number of maximal quasi-bicliques is not necessarily bigger than maximal bicliques, as multiple maximal bicliques can be contained in the same maximal quasi-biclique. By Propositions 4.1 and 4.2, we see that the mining of maximal quasi-bicliques is a technically non-trivial task, even when all maximal bicliques are known in advance.

Next we present a bounded closure property for μ-tolerance quasi-bicliques. We then use this property to design a depth-first algorithm to enumerate μ-tolerance quasi-bicliques.

THEOREM 4.1. *Let $H = \langle V_1^h, V_2^h\rangle$ be a μ-tolerance quasi-biclique of a graph $G = \langle V, E\rangle$ satisfying $|V_1^h|, |V_2^h| > 2\mu$. Let $V_1 \subseteq V_1^h$, $V_2 \subseteq V_2^h$, and $|V_1|, |V_2| \geq 2\mu$. Then $K = \langle V_1, V_2, E^k\rangle$ is also a μ-tolerance quasi-biclique of G where $E^k = (V_1 \times V_2) \cap E$.*

Proof. For all $v \in V_i$, $i = 1$ or 2, then $v \in V_i^h$. Then v is disconnected from at most μ number of vertices in $V_j^h, j \neq i$. This implies that v is disconnected from at most μ number of vertices in V_j.

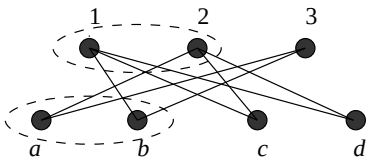


Figure 2: No closure property in δ%-tolerance quasi-bicliques.

On the other hand, since v is disconnected from at most μ number of vertices in V_j, then v is adjacent to at least $(|V_j| - \mu)$ number of vertices in V_j. Observe that $(|V_j| - \mu) \geq \mu$. Therefore, $K = \langle V_1, V_2, E^k \rangle$ is a μ-tolerance quasi-biclique.

It is interesting to define a minimal quasi-biclique:

DEFINITION 4.1. *[μ-tolerance minimal quasi-biclique] Let $H = \langle V_1, V_2 \rangle$ be a μ-tolerance quasi-biclique subgraph of G. H is a* μ-tolerance minimal quasi-biclique *subgraph of G if there is no other μ-tolerance quasi-biclique subgraph $H' = \langle V_1', V_2' \rangle$ of G such that $V_1 \supseteq V_1'$ and $V_2 \supseteq V_2'$.*

Note that the maximum vertex size (for $|V_1|$ and $|V_2|$) of μ-tolerance minimal quasi-bicliques in a graph G is 2μ due to Theorem 4.1. However the minimum vertex size (for $|V_1|$ and $|V_2|$) of μ-tolerance minimal quasi-bicliques is μ by definition. Quasi-bicliques between them may be μ-tolerance, may be not.

By Theorem 4.1, it follows that if a subgraph $H = \langle V_1, V_2 \rangle$, where $|V_1|, |V_2| \geq 2\mu$, is not a μ-tolerance quasi-biclique, then all its superset graphs are not a μ-tolerance quasi-biclique either. This bounded closure property is used in our *μ-completeQB* algorithm, presented in Section 5, for enumerating all μ-tolerance maximal quasi-bicliques.

We next use a counter-example to explain that this bounded closure property (Theorem 4.1) is not true for δ%-tolerance maximal quasi-bicliques.

Example. For the graph G in Figure 2, the quasi-biclique $H = \langle \{1,2,3\}, \{a,b,c\} \rangle$ is a 40%-tolerance quasi-biclique. It is also maximal for $\delta\% = 40\%$. However, the quasi-biclique $H' = \langle \{1,2\}, \{a,b\} \rangle$ is not a 40%-tolerance quasi-biclique. But the quasi-biclique $\langle \{1\}, \{b\} \rangle$ or $\langle \{2\}, \{a\} \rangle$ is a 40%-tolerance quasi-biclique.

Therefore, a non-δ%-tolerance quasi-biclique does not mean that its superset quasi-bicliques are not δ%-tolerance; it is possible that some of them are. So a level-wise depth-first search for δ%-tolerance maximal quasi-bicliques has to be exhaustive, which should be avoided.

We introduce a fixed point theorem for determining quasi-bicliques when a quasi-biclique is a maximal δ%-tolerance quasi-bicliques in a graph. Before presenting the theorem, we define an extended version of the neighbourhood for a set of vertices.

DEFINITION 4.2. **(Extended neighborhood of a vertex set)** *Let X be a nonempty set of vertices in a graph G. The μ-neighborhood $\beta_\mu(X, G)$ of X in G is the set of vertices in G adjacent to at least $(|X| - \mu)$ number of vertices in X. That is,*

$$\beta_\mu(X, G) = \{v \mid |\beta(v, G) \cap X| \geq (|X| - \mu)\}$$

Similarly, the δ%-neighbourhood $\beta_{\delta\%}(X, G)$ of X in G is the set of vertices in G adjacent to at least $(1-\delta\%)$ of vertices in X. That is,

$$\beta_{\delta\%}(X, G) = \left\{ v \;\middle|\; \frac{|\beta(v, G) \cap X|}{|X|} \geq (1 - \delta) \right\}$$

A quasi-biclique $(\beta_\mu(X, G), X)$ may not be μ-tolerance. However, some subset of $\beta_\mu(X, G)$ together with X can form a quasi-biclique. Suppose $W \subseteq \beta_\mu(X, G)$ such that $\langle W, X \rangle$ is a μ-tolerance quasi-biclique. If W is maximal, we denote $W = \beta_\mu^*(X, G)$. Similarly, we define $\beta_{\delta\%}^*(X, G)$.

THEOREM 4.2. *Let X and Y be two disjoint subsets of vertices of a graph G. Let $\mu \geq 0$ be a small integer number, and $0 \leq \delta\% < 1$ be a percentage value. Then the quasi-biclique $H = \langle X, Y \rangle$ is μ-tolerance maximal if and only if*

$$X = \beta_\mu^*(Y, G), \text{ and } Y = \beta_\mu^*(X, G)$$

And it is δ%-tolerance maximal if and only if

$$X = \beta_{\delta\%}^*(Y, G), \text{ and } Y = \beta_{\delta\%}^*(X, G)$$

Proof. We first prove the left-to-right direction. Suppose $H = \langle X, Y \rangle$ is a μ-tolerance maximal quasi-biclique. Then $Y \subseteq \beta_\mu(X, G)$. As there is no more vertex in G that can be added in Y, it is maximal in $\beta_\mu(X, G)$. Therefore, $Y = \beta_\mu^*(X, G)$. We can also prove $X = \beta_\mu^*(Y, G)$ in a similar fashion.

We now prove the right-to-left direction by contradiction. Suppose $H = \langle X, Y \rangle$ is not a maximal quasi-biclique. Then w.l.o.g., a new vertex can be added in X or Y such that the new subgraph is a maximal quasi-biclique. We denote the new subgraph as $\langle X \cup \{a\}, Y \rangle$. Then $(X \cup \{a\}) = \beta_\mu^*(Y, G)$. This is contradictory to the assumption that $X = \beta_\mu^*(Y, G)$. Therefore, H is a maximal quasi-biclique.

This theorem induces a function $f(H)$ for finding δ%-tolerance maximal quasi-bicliques. That is, starting from a quasi-biclique H, f can be iteratively applied to expand it until a fixed point is reached, which is guaranteed to be a δ%-tolerance maximal quasi-bicliques containing H. Thus this theorem is helpful to find all δ%-tolerance maximal quasi-bicliques that contain a group of pre-specified vertices. Without this theorem, such a task needs an exhaustive search. We also note that enumerating all δ%-tolerance maximal quasi-bicliques from a given graph under a certain δ is still a challenging computational problem that we will be working on in future.

5 Our Algorithm

As maximal quasi-bicliques of a small vertex size are prone to random patterns, we are more interested in those whose vertex size at each side exceeds a pre-specified threshold $ms \geq \mu$. By Theorem 4.1, if $ms \geq 2\mu$, then maximal ϵ-quasi bicliques [25] (where $\epsilon = \mu$) are actually our μ-tolerance maximal quasi-bicliques. So, we can use the *completeQB* algorithm [25] as a subroutine to mine one part of our μ-tolerance maximal quasi-bicliques. Maximal ϵ-quasi bicliques have balanced noise tolerance, but they can still be easily skewed because they do not require the condition (ii) as of our Definition 3.2. Details of the *CompleteQB* algorithm can be found in [25]. If $ms < 2\mu$, we conduct a depth-first search for μ-tolerance maximal quasi-bicliques whose vertex size is between ms and 2μ. Algorithm 1 presents the pseudo codes of our modified algorithm.

Lines 4-15 are necessary per Theorem 4.1 and Definition 4.1. For non-bipartite graphs, our algorithm needs a pre-process step to transform them into equivalent bipartite graphs; we also need a post process to remove duplicate maximal quasi-bicliques. This graph transformation is as follows: Let $G = (V, E)$ be the general graph. We transform it to a bipartite graph $G' = (V, V', E')$ where $V' = \{v'_i \mid v_i \in V\}$ and $E' = \{(v_i, v'_j) \mid (v_i, v_j) \in E\}$. Note that G' has double the number of vertices and number of edges. Also, for any $A, B \subseteq V$, $\langle A, B\rangle$ is a biclique of G if and only if $\langle A, B'\rangle$ is a biclique of G'.

6 Related Work on Quasi-bicliques

In a conceptual conjunctive clustering research work, Mishra *et al.* [18] defined a concept called ϵ-bicliques. A bipartite subgraph $H = \langle V_1, V_2\rangle$ is an ϵ-biclique if every vertex in V_1 is adjacent to at least $(1 - \epsilon)$ of the vertices in V_2. Though this asymmetry of the "quasi" between V_1 and V_2 is a special need for the conceptual conjunctive co-clustering between the objects and attributes, ϵ-bicliques may not be well applied to other applications such as web community co-clustering, interacting protein groups co-clustering, documents and words co-clustering, and etc. This is because "every vertex in V_1 is adjacent to at least $(1-\epsilon)$ of the vertices in V_2" does not mean every vertex in V_2 is adjacent to at least $(1-\epsilon)$ of the vertices in V_1. However, in many real-life applications, every vertex is equally treated, and balanced and symmetrical noise tolerance for every vertex is commonly required.

Our definition of δ%-tolerance maximal quasi-bicliques is related to but quite different from α-quasi-bicliques [28]. An α-quasi-biclique is an ordered pair $\langle X_B \cup X_E, Y_B \cup Y_E\rangle$, where $\langle X_B, Y_B\rangle$ is a maximal biclique and $\langle X_E, Y_E\rangle$ is its maximal α-extension. An α-extension of $\langle X, Y\rangle$ is an ordered pair (X_e, Y_e) where $X_e \subseteq \overline{X}$, $Y_e \subseteq \overline{Y}$, and $X_e \cap Y_e = \emptyset$, such that every vertex in X_e and in Y_e is adjacent to at least α% of the vertices in Y and in X, respectively. So, all α-quasi-bicliques [28] are exactly an expansion from a maximal biclique. As many of our δ%-tolerance maximal quasi-biclique do not contain any maximal biclique, semantically, α-quasi-bicliques and our δ%-tolerance maximal quasi-biclique can have a big area of non-overlapping.

Algorithm 1 Algorithm μ-*CompleteQB*

Input:

A bipartite graph $G = (V_1, V_2, E)$; ms—the minimum size threshold for the vertex sets; μ—a small integer number;

Description:

1: **if** $ms \geq 2\mu$ **then**
2: Use $completeQB(G, ms, \epsilon = \mu)$ to mine and output all maximal ϵ-quasi bicliques;
3: Return;
4: **if** $ms < 2\mu$ **then**
5: Let $ms_1 = 2\mu$;
6: Use $completeQB(G, ms_1, \epsilon = \mu)$ to mine and output maximal ϵ-quasi bicliques;
7: Extract all ms-size μ-tolerance quasi-bicliques by exhaustive search;
8: Let $\{G_1, \ldots, G_n\}$ be these quasi bicliques;
9: **for all** G_i **do**
10: Expanding G_i to become bigger μ-tolerance quasi-biclique by depth-first search till at most level $(2\mu - 1)$;
11: Let M be the set of all these expanded μ-tolerance quasi-bicliques;
12: **for all** $H \in M$ **do**
13: **if** H is not a subgraph for any of maximal quasi-bicliques output by Step 6 **then**
14: output H as a μ-tolerance maximal quasi-biclique;
15: Return;

They also differ in that the α percentage is only locally in terms of X_B or of Y_B, whereas the δ% in our definition is globally in terms of $X_B \cup X_E$ or $Y_B \cup Y_E$. One more difference is that an α-quasi-biclique may not be a δ%-tolerance maximal quasi-biclique for any δ.

Our difference common to both ϵ-bicliques [18] and α-quasi bicliques [28] is that they are limited to only bipartite graphs. In contrast, our definitions work for non-bipartite graphs as well.

Our previous work by Sim *et al.* [25] introduced a type of quasi-bicliques called maximal ϵ-quasi bicliques, which tolerates noisy/missing data for co-clustering stocks and financial ratios. Those maximal bicliques have balanced noise tolerance for every vertex just like maximal quasi-bicliques introduced here, but those maximal bicliques [25] can be easily skewed because they do not require the condition (ii) as of Definition 3.2 in this work. As mentioned above, this normally leads to many maximal quasi-bicliques that actually do not contain any edge, namely useless bipartite subgraphs. Also in [25], the percentage-tolerance maximal quasi-bicliques were not introduced.

An eigenvector-based bioinformatics research work [4] proposed a quasi-bipartite sub-structure to analyse topological structure of protein-protein interaction networks. However, it is not clear how the quasi-bipartites are defined in [4]. Instead, a quasi-bipartite is just roughly described as two disjoint protein groups between which high level connectivity is expressed. All the quasi-bipartites are determined by eigenvectors with a negative eigenvalue of the adjacency matrix of the graph.

The density-based quasi-bicliques, usually called dense bipartite cliques, include those defined in [1, 10] where all bicliques require only a minimum global density of the edges—thus some specific vertex may, as said in Introduction, have a very low connectivity, even a possible zero-connectivity. All these are skewed distribution of missing edges.

Besson *et al.* introduced DR-bi-sets [3] as a new fault-tolerant pattern type alternative to formal concept discovery. DR-bi-sets are similar to maximal quasi-bicliques. However, we differ at: (a) Our Definition 3.2 (μ-tolerance maximal quasi-bicliques) requires two bounds: one is a upper bound for noise tolerance, the other is a lower bound for necessary connectivity of a vertex. The two bounds together ensure the mathematical soundness of the definition. It is theoretically important to raise the second bound. However, Besson *et al.* did not introduce the second bound for their DR-bi-sets; (b) Our Proposition 4.1 says that a maximal quasi-biclique H of a graph G sometimes may not contain any maximal biclique of G. It indicates that maximal quasi-biclique is a non-trivial generalization to the classical maximal biclique, bringing up novel subgraphs. Besson *et al.* did not touch this point in their paper. (c) Our δ-tolerance maximal quasi-bicliques (Definition 3.3) and the related fixed point theorem (Theorem 4.2) were not studied by Besson *et al.* This theorem induces a function $f(H)$ for finding δ-tolerance maximal quasi-bicliques. Specifically, starting from a quasi-biclique H, $f(H)$ can be iteratively applied to expand it until a fixed point is reached, which is guaranteed to be a δ-tolerance maximal quasi-biclique containing H. Thus this theorem is helpful to find all δ-tolerance maximal quasi-bicliques that contain a group of pre-specified vertices. Without this theorem, such a task needs an exhaustive search.

7 Experimental Results

We evaluate our maximal quasi-bicliques in three aspects: (1) We conduct experiments to demonstrate that there are many large-size maximal quasi-bicliques from benchmark graphs that do not contain any maximal bicliques; (2) We show that Mishra's ϵ-bicliques [18] may have very skewed noise tolerance and missing edge distribution, we also compare the efficiency of our algorithm with Mishra's algorithm; (3) We introduce a new bioinformatics application based on the idea of maximal quasi-bicliques.

7.1 Many Maximal Quasi-bicliques Contain No Maximal Biclique In this section, we report our experimental results obtained from two benchmark graphs of the DIMACS Challenge. See the first two rows of Table 1 for their edge connectivity and density information.

Datasets	#vertices	#edges	edge density
c-fat500-1	500	4459	0.0357
c-fat200-1	200	3235	0.163
yeast-p2p	4904	17440	0.00145

Table 1: A protein interaction graph and 2 DIMACS benchmark graphs.

As shown in Table 2, about 81% of maximal quasi-bicliques in c-fat500-1, whose vertex set sizes are ≥ 5, do not contain any maximal bicliques, though there are 278352 maximal bicliques in c-fat500-1. A similar phenomenon can be observed for the other graph c-fat200-1.

We also found many large-size maximal quasi-bicliques. For example, in the c-fat500-1 graph, there are 917 μ-tolerance ($\mu = 1$) maximal quasi-bicliques

graph	min. size (ms) for $\lvert V_1\rvert$ and $\lvert V_2\rvert$	# max quasi-bicliques that contain a max biclique	# max quasi-bicliques that do not contain a max biclique (%)
c-fat200-1	6	12301	2143 (15%)
	5	45047	132263 (75%)
c-fat500-1	6	168758	6305 (4%)
	5	390174	1753973 (81%)

Table 2: The number of μ-tolerance maximal quasi-bicliques that contain a maximal biclique subgraph for $\mu = 1$, and the number of those that do not contain any maximal biclique subgraphs.

whose two vertex set sizes are both ≥ 8. However, there is no maximal biclique in this graph whose two vertex set sizes are both ≥ 8. Furthermore, none of the 917 maximal quasi-bicliques contains any of the 278352 maximal bicliques. This means that none of the above large-size maximal quasi-bicliques can be obtained by expanding any of the maximal bicliques. So, our definition indeed non-trivially generalizes the concept of maximal bicliques, bringing up many novel subgraph patterns of large size.

These benchmark examples highlight again that: The concept of α-quasi-bicliques [28] is semantically very different from our quasi-bicliques, as any α-quasi-biclique must contain a maximal biclique, but ours are not necessarily to contain any.

7.2 Skewness and Efficiency Comparison with Mishra's ϵ-bicliques We implemented the *Approximate Maximum Biclique Algorithm* [18] to mine ϵ-bicliques, and we used a modified version of *CompleteQB*, as shown in Algorithm 1, to mine μ-tolerance maximal quasi-bicliques. We applied both of them to the benchmark graph of the DIMACS Challenge c-fat200-1, and also to a real dataset, the yeast protein-protein interaction (ppi) dataset downloaded from DIP [3]. The yeast ppi dataset is represented in the form of an undirected graph. See Table 1 for its edge connectivity and density information.

For constrasting the skewness, we compared just 1000 ϵ-bicliques [18] and 1000 μ-tolerance maximal quasi-bicliques mined from each graph, as the Approximate Maximum Biclique Algorithm does not mine the complete set of ϵ-bicliques.

To measure the skewness of a vertex in a ϵ-biclique and in a μ-tolerance maximal quasi-biclique, we calculate the vertex's missing edges to its opposite vertex set in percentage. We set $\mu = 1$, $ms = 5$ for c-fat200-1 and $\mu = 1$, $ms = 3$ for yeast ppi dataset as parameter settings to generate 1000 μ-tolerance maximal quasi-bicliques from each dataset. Due to the definition of μ-tolerance maximal quasi-biclique, the missing edges of vertices in μ-tolerance maximal quasi-bicliques from c-fat200-1 and yeast ppi dataset are bounded by $\mu/ms = 1/5 = 20\%$ and $\mu/ms = 1/3 = 33.33\%$ respectively, thus the skewness of the missing edges is prevented.

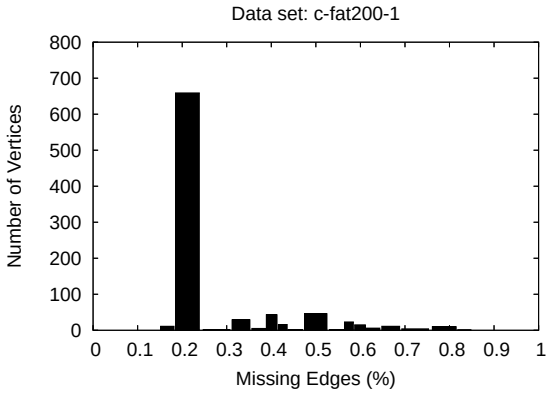


Figure 3: Distribution of the missing edges of vertices in ϵ-bicliques. Here, ϵ-bicliques are mined from c-fat200-1.

For ϵ-bicliques, we set $\epsilon = 20\%$, $ms = 5$ for c-fat200-1 and $\epsilon = 33.33\%$, $ms = 3$ for yeast ppi dataset to mine ϵ-bicliques. The distribution of the missing edges of all vertices in the 1000 ϵ-bicliques has a long tail, as expected. See Figure 3 for detailed distribution information for the ϵ-bicliques from c-fat200-1, and Figure 4 for the distribution information from the yeast ppi dataset.

Observe that many vertices' missing edges exceed $\epsilon = 20\%$ and $\epsilon = 33.33\%$, implying that these vertices are skewed. There are even some vertices which are 100% *not* connected to any vertex in the opposite side of its ϵ-biclique, as shown in Figure 4. This skewness is mainly attributed to the asymmetrical quasi tolerance allowed in ϵ-bicliques.

Figure 5(a) and 5(b) show two examples of skewed

[3] http://dip.doe-mbi.ucla.edu/

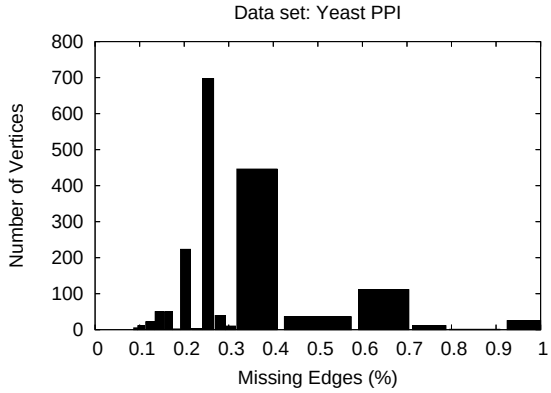


Figure 4: Distribution of the missing edges of vertices in ϵ-bicliques. Here, ϵ-bicliques are mined from yeast ppi dataset.

ϵ-bicliques discovered. In Figure 5(a), the skewed ϵ-biclique is mined from c-fat200-1. Its node v_1 is highly skewed as it has 80% missing edges to its opposite vertex set. In Figure 5(b), the skewed ϵ-biclique is mined from yeast ppi dataset. Its nodes v_1 and v_2 are highly skewed as both have 50% missing edges to its opposite vertex set.

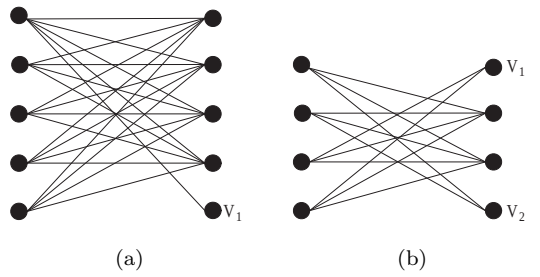


Figure 5: (a) A skewed ϵ-biclique in c-fat200-1. (b) A skewed ϵ-biclique in yeast ppi dataset.

We next examine the efficiency of the Approximate Maximum Biclique Algorithm and our μ-*CompleteQB* Algorithm, by comparing their running time for mining the ϵ-bicliques and μ-tolerance maximal quasi-bicliques respectively. The Approximate Maximum Biclique Algorithm needs to set another 3 parameters $\hat{m}$, m and t for mining ϵ-bicliques. The algorithm uses these parameters to determine the number of vertices to be picked to find ϵ-bicliques within them. Details of the parameters is explained in [18].

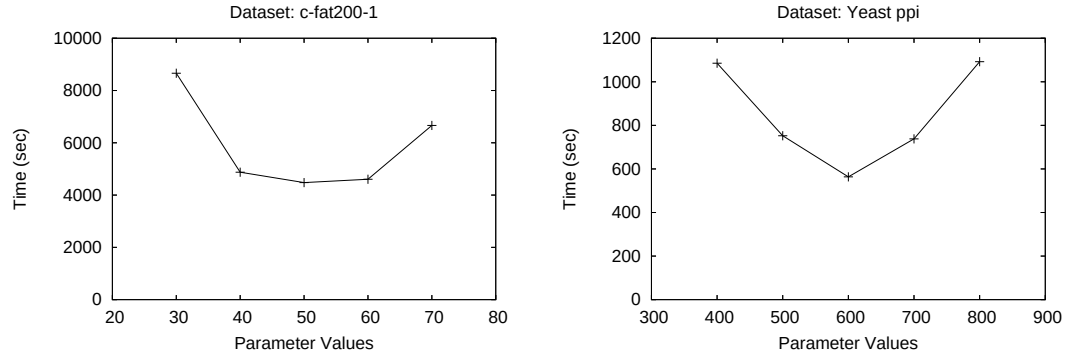
Initially, we used the settings recommended in [18] to mine the ϵ-bicliques, but the algorithm could not complete after running 12 hours on either of c-fat200-1 or the yeast ppi dataset. Hence, we find the optimal parameter settings for the datasets in a heuristic way. To simplify the process, we set $\hat{m} = m = t$. Figure 6(a) and 6(b) present the running time taken to mine 1000 ϵ-bicliques on c-fat200-1 and yeast ppi datasets respectively under different parameter settings for $\hat{m}$, m and t. We can see that the optimum settings for c-fat200-1 is at $\hat{m} = m = t = 50$, and for yeast ppi dataset is at $\hat{m} = m = t = 600$. We do not present the running time for the μ-*CompleteQB* Algorithm under these Figures because μ-*CompleteQB* does not require these parameters and it only took 26 seconds and 2 seconds to mine 1000 μ-tolerance maximal quasi-bicliques from c-fat200-1 and yeast ppi dataset respectively.

7.3 Protein Interaction Prediction by Co-clustering Proteins Co-clustering problem refers to an unsupervised learning process that identifies two disjoint vertex sets of a graph G between which every pair of vertices are adjacent to each other. So, in fact it is equivalent to the mining of maximal bicliques from G. Note that co-clustering is different from co-partitioning problem as in co-partitioning problems, subsets of nodes in one side of a bipartite graph are required to be non-overlapping.

Very often in co-clustering applications, the graph G is specialized and represented by a bipartite graph $\langle V_1, V_2 \rangle$ where V_1 and V_2 are of two different kinds of vertices. For example, in the documents-and-words co-clustering problem [6, 7], V_1 usually represents a set of documents but V_2 represents a set of words. This is similar for images-and-features co-clustering problems [22] and stocks-and-ratios co-clustering problems [25]. This situation can be also found in a bioinformatics research problem—the reconstruction of the supertree of Life [28, 24, 8]. Its computational problem is to find a maximal number of genes that are contained in a maximal number of species (taxa), given the binary containment relation of all the genes and taxa.

While for other co-clustering applications, all vertices in the graph G are of the same kind. This includes the web community co-clustering problem [23, 20], and the discovery of interacting protein group pairs [13, 4, 19]. For these situations, the graph G is usually represented as a general graph.

As our newly defined maximal quasi-bicliques are capable of a balanced noise tolerance and possess a most-versus-most connection, they are useful for finding enlarged co-clusterings for these applications. The enlarged co-clusterings can in turn strengthen the association between the two vertex sets, and thus can improve the quality of the applications. We demonstrate the usefulness of our maximal quasi-bicliques by a new



(a) (b)

Figure 6: Running time of the Approximate Maximum Biclique Algorithm on the c-fat200-1 and yeast ppi datasets.

application—prediction of missing interactions in a protein interaction network.

The background of this bioinformatics problem is as follows. Using laboratory experiments to determine the interactions is very expensive and time-consuming; the current known and validated interaction data is also far from complete. So, it is important to use computational methods to make prediction of protein interactions. Our idea is based on the observation that certain protein groups in a protein network interact with one another in a way like a maximal biclique subgraph in a graph [4, 13, 19]. So, if a maximal quasi-biclique subgraph H is identified from a protein interaction graph, then the missing edges in H are most likely to be true interactions. Here, a protein is represented by a vertex, and an interaction between two proteins is represented by an edge.

We apply this idea to a widely accepted protein interaction graph, called MIPS CYGD dataset, which is the whole-genome protein interaction network of yeast [17]. This dataset contains 15,456 protein interactions and 4554 distinct proteins. We removed duplicated interactions and self-interactions from the dataset. The remaining number of interactions is 12,319, and we use this cleaned dataset to conduct the experiments.

We consider quasi-bicliques whose vertex sets contain at least 10 vertices each, as recommended by [4, 13, 19]. A total of 34026 μ-tolerance maximal quasi-bicliques ($\mu = 1$) are discovered. And there are 809 distinct missing edges in these μ-tolerance maximal quasi-bicliques.

The question now is that whether these 809 missing edges are potentially true interactions. As there is no explicit biological information to verify our prediction, we evaluate the interactions using a very recent bioinformatics method [5]: *A pair of proteins in a true biological interaction are much more likely to be similar in function than random protein pairs.* We found that 148 pairs (18%) of the 809 pairs of potentially interacting proteins have a common annotated function in MIPS for both proteins in the pair. This rate is much higher than for random protein pairs. Actually, we generated 100 sets of 809 random protein pairs using the function-annotated yeast proteins in MIPS. The average number of the random pairs that are homogeneous in function is only 42.5 out of 809, namely 5%, with a standard deviation of 6.73. As 18% is 3.6 folds higher than 5%, our prediction results are far better than the random results. In fact, the 148 potentially interacting proteins can be ranked according to the concept of "guilt by association of common interacting partners" [5]; that is, two proteins are more likely to interact if a larger proportion of their interaction partners are actually shared. A simple way to do such a ranking is $S(u, v, G) = |\beta(u, G) \cap \beta(v, G)|/|\beta(u, G) \cup \beta(v, G)|$, where u and v are the pair in question, and G is the original protein interaction graph. Under this ranking, as $S(u, v, G)$ increases from 0 to 0.4, the fraction among our predicted interaction proteins that share a common function increases from 18% (148/809) to 78% (11/14), nearly 15 folds higher than that of random pairs under the same $S(u, v, G)$ thresholds.

We also validate our prediction in a second way. We randomly remove 30% of the edges from the interaction graph of the MIPS CYGD dataset. Then we consider those quasi-bicliques having at least 7 vertices in each of their vertex sets, and discover μ-tolerance maximal quasi-bicliques ($\mu = 1$). The missing edges in these maximal quasi-bicliques were predicted as true interactions. We found that about 50% of these predicted interactions are actually those that we have removed—i.e., the true interactions. The balance of the predicted interactions may also be true interactions, but we are unable to confirm due to the incompleteness of the MIPS

CYGD dataset. Nevertheless, based on our experiments with respect to function homogeneity above, we believe that at least 78% of the remaining predictions are most likely true.

These preliminary results shows high potential of maximal quasi-bicliques in the prediction of missing protein interactions. There are many interesting related problems for future research. For example, why small-size quasi-bicliques cannot be used for interaction prediction (any biological evidence), what is the trade off between the vertex set size and μ, the coverage, the sensitivity, and precision. The contribution here is that we show the potential of using maximal quasi-bicliques in a new application area of co-clustering.

8 Conclusion

We have proposed two types of most-versus-most connections—μ-tolerance and δ%-tolerance—to refine maximal quasi-bicliques for balanced noise tolerance of missing edges. We have made the observation that a maximal quasi-biclique of a graph does not always contain a maximal biclique subgraph of this graph. This observation suggests that the expansion approach should be avoided to enumerate maximal quasi-bicliques. We have discussed a bounded closure property and a fixed point theorem, which are useful for designing efficient algorithms for enumerating maximal quasi-bicliques. We have also presented a new co-clustering application where true protein interactions can be identified through our quasi-bicliques.

As for future work, we note that the noise tolerance by our maximal quasi-bicliques is fixed at level μ (or δ%) for every vertex. We can extend the definition such that vertices from different sides are allowed to tolerate different level of missing data. Also, the current definition does not avoid those bipartite subgraphs with small sizes. So, an important sub-problem is how to directly enumerate maximal quasi-bicliques with large sizes which are statistically stronger than the small ones. We also plan to further investigate the prediction of protein interactions, and other co-clustering problems.

Acknowledgement

This work has been supported by NTU (Nanyang Technological University) Tier-1 start-up grant (ref. no. RG66/07) to the first author of this paper. The third and last authors also wish to acknowledge the support of an MOE AcRF Tier 1 grant and an SERC PSF grant.

References

[1] J. Abello, M. G. C. Resende, and S. Sudarsky. Massive quasi-clique detection. In *LATIN '02: Proceedings of the 5th Latin American Symposium on Theoretical Informatics*, pages 598–612, London, UK, 2002. Springer-Verlag.

[2] G. Alexe, S. Alexe, Y. Crama, S. Foldes, P. L. Hammer, and B. Simeone. Consensus algorithms for the generation of all maximal bicliques. *Discrete Applied Mathematics*, 145(1):11–21, 2004.

[3] J. Besson, C. Robardet, and J.-F. Boulicaut. Mining a new fault-tolerant pattern type as an alternative to formal concept discovery. In *Proceedings of the 14th International Conference on Conceptual Structures ICCS'06, LNCS 4068*, pages 144–157, 2006.

[4] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen. Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucleic Acids Research*, 31 (9):2443–2450, 2003.

[5] H. N. Chua, W. K. Sung, and L. Wong. Exploiting indirect neighbors and topological weight to predict protein function from protein-protein interactions. *Bioinformatics*, 22:1623–1630, 2006.

[6] I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of The 7th ACM SIGKDD International Conference on Know ledge Discovery and Data Mining (KDD-2001)*, pages 269–274, 2001.

[7] I. S. Dhillon, S. Mallela, and D. S. Modha. Information-theoretic co-clustering. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003)*, pages 89–98, 2003.

[8] A. C. Driskell, C. Ane, J. G. B. M. M. McMahon, B. C. OMeara, and M. J. Sanderson. Prospects for building the tree of life from large sequence databases. *Science*, 306:1172–1174, 2004.

[9] D. Eppstein. Arboricity and bipartite subgraph listing algorithms. *Information Processing Letters*, 51:207–211, 1994.

[10] D. Gibson, R. Kumar, and A. Tomkins. Discovering large dense subgraphs in massive graphs. In *VLDB '05: Proceedings of the 31st international conference on Very large data bases*, pages 721–732. VLDB Endowment, 2005.

[11] G. Grahne and J. Zhu. Fast algorithms for frequent itemset mining using fp-trees. *IEEE Transactions on Knowledge and Data Engineering*, 17(10):1347–1362, October 2005.

[12] D. Hanisch, A. Zien, R. Zimmer, and T. Lengauer. Co-clustering of biological networks and gene expression data. *Bioinformatics*, 18 Suppl 1:S145–54, 2002.

[13] H. Li, J. Li, and L. Wong. Discovering motif pairs at interaction sites from protein sequences on a proteome-wide scale. *Bioinformatics*, 22:989–996, 2006.

[14] J. Li, H. Li, D. Soh, and L. Wong. A correspondence between maximal complete bipartite subgraphs and closed patterns. In *The 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, pages 146–156, 2005.

[15] J. Li, G. Liu, H. Li, and L. Wong. Maximal biclique subgraphs and closed pattern pairs of the adjacency matrix: A one-to-one correspondence and mining algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 19 (12):1625–1637, 2007.

[16] K. Makino and T. Uno. New algorithms for enumerating all maximal cliques. In *Proceedings of the 9th Scandinavian Workshop on Algorithm Theory (SWAT 2004)*, pages 260–272. Springer-Verlag, 2004.

[17] H. W. Mewes, D. Frishman, and U. Guldener. MIPS: A database for genomes and protein sequences. *Nucleic Acids Res*, 30(1):31–34, 2002.

[18] N. Mishra, D. Ron, and R. Swaminathan. A new conceptual clustering framework. *Machine Learning*, 56 (1-3):115–151, 2004.

[19] J. L. Morrison, R. Breitling, D. J. Higham, , and D. R. Gilbert. A lock-and-key model for proteinprotein interactions. *Bioinformatics*, 22 (16):20122019, 2006.

[20] T. Murata. Discovery of user communities from web audience measurement data. In *Proceedings of The 2004 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2004)*, pages 673–676, 2004.

[21] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Efficient mining of association rules using closed itemset lattices. *Journal of Information Systems*, 24:25–46, 1999.

[22] G. Qiu. Image and feature co-clustering. In *17th International Conference on Pattern Recognition (ICPR 2004)*, pages 991–994, 2004.

[23] J. E. Rome and R. M. Haralick. Towards a formal concept analysis approach to exploring communities on the world wide web. In *International Conference on Formal Concept Analysis*, pages 33–48, 2005.

[24] M. J. Sanderson, A. C. Driskell, R. H. Ree, O. Eulenstein, and S. Langley. Obtaining maximal concatenated phylogenetic data sets from large sequence databases. *Molecular Biology and Evolution*, 20(7):1036–1042, 2003.

[25] K. S. Sim, J. Li, V. Gopalkrishnan, and G. Liu. Mining maximal quasi-bicliques to co-cluster stocks and financial ratios for value investment. In *Proceedings of the 2006 IEEE International Conference on Data Mining (ICDM'06)*, pages 1059–1063, 2006.

[26] T. Uno, M. Kiyomi, and H. Arimura. Lcm ver. 3: Collaboration of array, bitmap and prefix tree for frequent itemset mining. In *Proc. of the ACM SIGKDD Open Source Data Mining Workshop on Frequent Pattern Mining Implementations*, 2005.

[27] J. Wang, J. Han, and J. Pei. CLOSET+: Searching for the best strategies for mining frequent closed itemsets. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03), Washington, DC, USA*, pages 236–245, 2003.

[28] C. Yan, J. G. Burleigh, and O. Eulenstein. Identifying optimal incomplete phylogenetic data sets from sequence databases. *Molecular Phylogenetics and Evolution*, 35(3):528–535, 2005.