The Pennsylvania State University
The Graduate School
College of Engineering

GREEDY ALGORITHM FOR APPROXIMATING
MAXIMUM INDUCED MATCHING

A Thesis in
Computer Science and Engineering

by

Zhenyao Li

©Zhenyao Li

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2014
The thesis of Zhenyao Li is approved and reviewed by the following:

Piotr Berman  
Associate Professor of Computer Science and Engineering  
Thesis Advisor

Sofya Raskhodnikova  
Associate Professor of Computer Science and Engineering

Lee Coraor  
Head of Department of Computer Science and Engineering

Signatures are on file in the Graduate School.
Abstract

An induced matching in a graph $G = (V, E)$ is $M \subseteq E$ such that it is a matching and also the edge set of an induced subgraph of $G$. The goal in the Maximum Induced Matching (MIM) problem is to maximize the size of $M$. This problem can be modelled as a special case of Set Packing, or Maximum Independent Set, and, like these problems, it is very hard to approximate, which motivates our focus on restricted classes of graphs. This work improves on the work of Duckworth et al. who showed that MIM is APX-hard even for bipartite 3-regular graphs, while a simple linear time greedy algorithm gives an approximation ratio of $d - 1$ for $d$-regular graph. We improve this ratio for 3-regular graphs from 2 to $\frac{5}{3}$, also using a linear time greedy algorithm. We believe that our new methodology can be applied to other classes of graphs as well. More specifically, we conjecture that for 4-regular graphs it gives an approximation ratio of $\frac{7}{3}$, improving on 3.

We also provide an improved lower bound on approximability of MIM in 3-regular graphs.

Keywords: Graph theory, combinatorial problems, approximation algorithms, induced matchings, greedy algorithms.
## List of Figures

1. Marked edges form an induced matching. . . . . . . . . . . . . 3
2. Greedy algorithm DMZ-ALG of Duckworth *et al.* . . . . . . . 4
3. Algorithm BL-ALG for *d*-regular-MIM. To complete the de-
   scription, we need an implementation of an *X*-predictor *Y*. . . 6
4. *Θ*- and *Θ*-graphs. . . . . . . . . . . . . . . . . . . . . . . . . . 7
5. Diagram for Case 5 . . . . . . . . . . . . . . . . . . . . . . . . . . 9
6. Diagram for Case 6 . . . . . . . . . . . . . . . . . . . . . . . . . . 10
7. Two gadgets $\Gamma_u, \Gamma_v$ (solid edges), contact edges (dashed) and
   with extra edges (thin or very short thick) added to form $\Gamma'_u, \Gamma'_v$. 14
1 Introduction

In this thesis, we study the Maximum Induced Matching problem. The input to the problem is an undirected graph \((V, E)\). For an edge set \(A \subseteq E\), we use \(V(E)\) to denote its node set, the union of edges in \(A\). For a node set \(S \subseteq V\), we define \(E(S) = \{e \in E : e \subseteq S\}\), the set of edges induced by \(S\). A set of edges \(M\) is an induced matching if \(|V(M)| = 2|M|\) and \(E(V(M)) = M\). The goal in MIM is to find an induced matching of maximum size (see an example in Figure 1.) This problem was introduced by Stockmeyer and Vazirani [1] who motivated it as a risk-free marriage problem: find the maximum number of married couples such that each married person is compatible with no married person other than his/her spouse. Another motivation stems from the Strong Edge Coloring problem, where adjacent edges should be given different colors. Additionally, two edges that are both adjacent to a third edge should also be given different colors. Subject to these constrains, we minimize the number of colors. Equivalently, edges of each color form an induced matching. As in similar coloring problems, if we can approximate the maximum single color set with ratio \(F\), we can also approximate the minimum number of colors with ratio \(F \ln |V|\). (Here, \(F\) is an upper bound on \(|S_A|/|S^*|\), where \(S_A\) is the solution found by the algorithm, and \(S^*\) is an optimum solution). In the last 30 years MIM was investigated by a number of people. Duckworth et al. [2] provide a recent extensive bibliography.

To us, the problem offers an interesting case study on greedy algorithms. Greedy algorithms are typically fast and scalable (meaning that the running time is linear or proportional to the time of sorting), and for some problems they offer the best known approximation ratios. However, sometimes we need to replace the obvious choice criteria with more insightful ones. One could offer a life advice that even if we agree to be greedy (i.e., make our selections fast and never change them), we do not have to be stupid. But this advise is not always correct. For example, for the Set Cover (and Set Packing) problem, the most obvious choice criteria—picking the largest (or the smallest) set suffices to obtain the best known approximation. So, it is a genuinely open problem: does it pay to try to be clever? It is morally reassuring if we can show that it does.

Our results concern restricted versions of MIM: \(d\)-MIM (restricted to graphs of degree \(d\), i.e., where every node has at most \(d\) neighbors) and \(d\)-regular-MIM (restricted to \(d\)-regular graphs, where every node has exactly \(d\) neighbors).
For $d$-regular-MIM, Duckworth et al. [2] provided a very simple greedy algorithm with approximation ratio $d - 1$. For $d=3$, we improve this ratio to $5/3 < 2$. They also provided characterizations of inapproximability, namely that one should not expect (in the sense that we explain in Section 5) polynomial time approximation algorithms for 3-MIM with a ratio better than $1 + 1/475$ and for 3-regular-MIM with a ratio better than $1 + 1/K$, where $K \approx 1250$ is implicit. We improve these provably difficult ratios to $1 + 1/288$ and $1 + 1/482$, respectively.

2 The algorithm of Duckworth et al.

The algorithm of Duckworth et al. [2] is based on the observation that MIM is a special case of Set Packing.

For each node $u$ we will consider a star of $u$, the set of edges that contain that node; in turn, for each edge $e$ we consider the union of stars of its endpoints and we call it a “double star”. More formally,

**Definition 1** For $v \in V$ we define a “star of $v$” as $S_v = \{e \in E : v \in e\}$. For $\{u, v\} \in E$ we define a “double star” of $\{u, v\}$ as $D_{\{u,v\}} = S_u \cup S_v$.

**Observation 1** Double stars have the following two properties:

1. $M \subseteq E$ is an induced matching (IM) iff $D_e \cap D_f = \emptyset$ for every two different edges $e, f \in M$;

2. if $d$ is the maximum node degree then $|D_e| \leq 2d - 1$ for every $e \in E$.

According to item 1 in Observation 1, MIM is equivalent to finding a maximum set packing, namely, a packing of sets of the form $D_e$. Moreover, according to item 2 of Observation 1, when we are solving an instance of $d$-MIM, the sizes of these sets are bounded by $2d - 1$, which allows us to apply heuristics for packing sets of bounded size, like the ones of Hurkens and Shriver [3]. So far, no better algorithms for $d$-MIM were presented. In this work, we show how to approximate 3-regular-MIM better by using heuristics and upper bound arguments that are specific to this problem.

Analyzing the approximation ratio of algorithms for MIM in $d$-regular graphs requires an upper bound on the size of induced matchings. Because $|D_e| = 2d - 1$, a packing of $D_e$ sets has at most $|E|/(2d - 1)$ sets. This is
the bound used by Duckworth et al. [2]. They described a greedy algorithm that repeats selecting an edge from $H$, the current set of edges.

After a selection of edge $e$, the greedy algorithm removes from the current set every edge $h$ that is in conflict with $e$, i.e., such that $D_e \cap D_h \neq \emptyset$.

**Definition 2** For each edge $e \in H$, the edge set $E_e = \{h \in H : D_e \cap D_h \neq \emptyset\}$.

With this definition, we can formulate greedy algorithm DMZ-ALG (see Figure 2).

In the analysis of the approximation ratio of DMZ-ALG we need another definition.

**Definition 3** When $H \subseteq E$ is implicit, $\deg(u) = |\{u,v \in H : v \in V\}|$.

**Lemma 1** $E_e$ has these two properties:

1. $E_e = \{h \in H : h \cap V(D_e) \neq \emptyset\}$;
2. $|E_{\{u,v\}}| \leq d(\deg(u) + \deg(v) - 2) + 1$.

**Proof.** Item 1 is obvious. To show Item 1, assume that $D_h \cap D_e \neq \emptyset$, then we have edge $f \in D_h \cap D_e$. Because $f \in D_e$, $f \subset V(D_e)$. Because $f \in D_h$, $f \cap h \neq \emptyset$. This implies that $h \cap V(D_e) \neq \emptyset$. Concerning Item 2, we use the fact that $E - H$ is a union of sets of the form $E_{f'}$, where $f'$s are edges already selected. Thus, if we have $\{u,v\} \in H$, while $\{v,w\} \in E - H$ then $\{v,w\} \in E_{f'}$ for some selected $f$. This $f$ does not contain $w$, otherwise $\{u,v\}$
would also be removed. This means that there is a path \( (u, v, w, x, y) \) where \( f = \{x, y\} \). This implies that every edge that contains \( w \) would either be removed before the selection of \( f \) or during that selection.

We conclude that \( E_{\{u,v\}} \) consists of edge \( \{u, v\} \) and the union of star sets \( S_w \) for those \( w \notin e \) where either \( \{w, u\} \in H \) or \( \{w, v\} \in H \). Hence, the number of those star sets is \( \deg(u) + \deg(v) - 2 \).

It suffices to analyze DMZ-ALG for one connected component. Let \( m \) be the number of edges in that connected component.

In the first iteration DMZ-ALG selects an edge \( e \) that has both nodes of degree \( d \), and by Lemma 1, \( |E_e| \leq d(2d - 2) + 1 \). In subsequent iterations, we always can select \( e \) that contains a node \( u \) with \( \deg(u) \leq d - 1 \), hence \( |E_e| \leq d(2d - 3) + 1 = (2d - 1)(d - 1) \) (again, by Lemma 1.)

Summarizing, we have a lower bound \( BL \) on the number of selected edges and an upper bound \( BU \) on the size of the maximum induced matching:

\[
BL \geq \frac{m - d}{(2d - 1)(d - 1)}, \quad BU \leq \frac{m}{2d - 1}. \tag{1}
\]

Because \( BL \) and \( BU \) are integer, we can show that \( (d - 1)BL \geq BU \). If not, let \( m > d \) be the smallest integer such that some \( BL, BU \) satisfy (1) while \( (d - 1)BL < BU \). Subtracting \( (2d - 1)(d - 1) \) from \( m \) decreases \( BL \) by 1 and \( BU \) by \( d - 1 \), hence \( m \) would still have that property unless \( m - d \leq (2d - 1)(d - 1) \). But then \( BU \leq d - 1 \), while \( BL \geq 1 \), a contradiction. (This analysis is slightly tighter than in [2].)
3 New upper bound and selection criteria

These ideas work for 3-regular-MIM but we formulate them in more general terms, because we hope that they also work for 4-regular-MIM and perhaps they can be extended for any \( d \)-regular-MIM.

**Definition 4** For an edge set \( M \), set \( X_M = E - \bigcup_{e \in M} D_e \).

**Definition 5** Let \( Y(A) \) be an integer function defined on edge sets. Integer function \( Y \) is an \( X \)-estimator if \( |A \cap X_M| \geq Y(A) \) for every IM \( M \).

**Observation 2** If \( M \) is an IM, then \( |M| = \frac{|E| - |X_M|}{2d - 1} \).

With the previous upper bound, \( \frac{|E|}{2d - 1} \), a way to guarantee that a greedy algorithm satisfies approximation ratio \( \rho \) was that when the selection of \( e \) removes \( E_e \) set of edges we have \( |E_e| \leq (2d - 1)\rho \). With \( X \)-predictor it suffices to have

\[
|E_e| - Y(E_e) \leq (2d - 1)\rho.
\]

Our last idea is to count both edges and nodes, rather than just edges. Because our input is a \( d \)-regular graph, we have this identity:

\[
|E| = \frac{|E| + d|V|}{3}.
\]

Now to guarantee ratio \( \rho \) it suffices to select edges so that

\[
\frac{|E_e| + d|R_e| - 3Y(E_e)}{3} \leq (2d - 1)\rho \equiv |E_e| + d|R_e| - 3Y(E_e) \leq 3(2d - 1)\rho.
\]

**Definition 6** Set \( U \subseteq V \) consists of nodes that were not removed yet, i.e., \( U = \{ u \in V : \deg(u) > 0 \} \), \( R_e \subseteq U \) is the set of nodes that would be removed by \( e \), i.e., with all incident \( H \)-edges in \( E_e \).

Let \( U_i = \{ u \in U : \deg(u) = i \} \), and \( N_i \) is the set of neighbors of nodes in \( U_i \). We also define \( \sigma(u) \), the sum of degrees of the neighbors of \( u \).

The resulting algorithm, BL-Alg, is shown in Figure 3. To analyze its approximation ratio in 3-regular graphs we need additional definitions.

**Definition 7** \( R_e \) is the union of three disjoint parts: \( e \), \( R_e^i = V_{D(e)} \) and the remainder \( R_e^o \), where \( R_e^o \) consists of nodes that are not in \( e \cup R_e^i \) but have neighbors only in \( R_e^i \).
\begin{algorithm}
\begin{algorithmic}
\STATE $H \leftarrow E$
\STATE $U \leftarrow V$
\STATE $M \leftarrow \emptyset$
\WHILE{$H \neq \emptyset$}
\STATE $e \leftarrow$ a member of $H$ minimizing $c(e) = |E_e| + d|R_e| - 3Y(E_e)$
\STATE insert $e$ to $M$
\STATE remove $E_e$ from $H$
\STATE remove $R_e$ from $U$
\ENDWHILE
\end{algorithmic}
\end{algorithm}

Figure 3: Algorithm BL-Alg for $d$-regular-MIM. To complete the description, we need an implementation of an $X$-predictor $Y$.

The following fact is used to compute $Y(A)$.

\textbf{Lemma 2} Assume that $M$ is an induced matching and $C$ is a set of edges of a 4-cycle. If $C \cap M \neq \emptyset$ then $C \cap X_M \neq \emptyset$.

\textbf{Proof.} Let $e \in M \cap C$ and $e'$ be the edge in $C$ that is disjoint with $e$. Nodes of $e'$ cannot be in $V_M$ because they are adjacent to $e$, thus $e' \in X_M$. \(\square\)

\section{3-regular graphs}

In this section we will prove the approximation ratio $\rho = 5/3$. As we have shown in the previous section, it suffices that we are always able to select an edge $e$ such that

\[|E_e| + d|R_e| - 3Y(E_e) \leq 3(2d-1)\rho = c(e) = |E_e| + 3|R_e| - 3Y(E_e) \leq 25\]

We will show that (2) holds with two exceptions. The first selection in a connected component has $c(e) \leq 37$. In the subsequent selections, it may happen that we select $e$ with $c(e) = 27 > 25$, but then the subsequent selection has $c(e') = 7$.

For 3-regular graphs we use the following lemma to compute $Y(E_e)$.

\textbf{Lemma 3} A $\Theta$-graph is a set of 7 edges $T \subseteq H$ of which 6 edges of the cycle $u, v, w, z, y, x$, and the 7th edge is $\{v, y\}$ (see Figure 4). A set of edges obtained from $T$ by inserting edge $\{u, z\}$ is a $\Theta'$-graph.
1. If $T$ is a $\Theta$-graph then $T \cap X_M \neq \emptyset$.
2. If $T'$ is a $\Theta'$-graph then $|T' \cap X_M| \geq 2$.

Proof. Property of $T$ follows from Lemma 2 if $M \cap T \neq \emptyset$. Otherwise $v$ and $y$ are not in $V_M$ because they belong to three edges of $T$, hence $\{v, y\} \in X_M$. Now we prove the property of $T'$. Each edge $e \in T'$ belongs to two 4-cycles, thus if $e \in M$ then two edges of $T'$ that are opposite to $e$ in these 4-cycles are in $X_M$ by Lemma 2. If $T' \cap M = \emptyset$ then $\{v, y\}, \{u, z\} \in X_M$.

We define

$$Y(A) = \begin{cases} 2 & \text{if } A \text{ contains a } \Theta', \\ 1 & \text{if } A \text{ contains a } \Theta, \\ 0 & \text{otherwise} \end{cases}$$

The following case analysis is a major part of the proof of Theorem 1.

Case 0: $|E_e| \leq 5$. Because $E_e$ is connected, $|R_e| \leq |E_e|$, hence $c(e) \leq |E_e| + 3(|E_e| + 1) = 4|E_e| + 3 \leq 23$.

Case 0 settles the situations when $v \in U_1$, $e = \{u, v\} \in H$ and $\sigma(v) \leq 5$. In particular, this happens when $\deg(u) = 2$: $\sigma(u) \leq 1 + 3$, and when $u$ two neighbors in $U_1$: $\sigma(u) \leq 1 + 1 + 3$. Thus later we assume that each $u \in N_1$ has degree 3 and exactly one neighbor in $U_1$.

Case 1: $U_1 \neq \emptyset$. We select $u \in N_1$ with the least $\sigma(u)$. Assume $u \in N_1$ has neighbors $v_0, v_1, v_2$, and $v_0 \in U_1$; our selection is $e = \{u, v_0\}$.

In this case $c(e) \leq |E_e| + 3|R_e|$ while $|E_e| = \sigma(u) \leq 1 + 3 + 3 = 7$; moreover $R_e = \{u, v_0, v_1, v_2\} \cup R_e^0$, hence $c(e) \leq 7 + 3 \times 4 + 3|R_e^0|$ and it suffices to show $|R_e^0| \leq 2$. Assume by the way of contradiction that $|R_e^0| \geq 3$. Note that $R_e^0$ consists of nodes with all neighbors in set $\{v_1, v_2\}$; if the sum of degrees of nodes in $R_e^0$ is 3, then $R_e^0 \subset U_1$, and either $v_1$ or $v_2$ has two neighbors in $U_1$, a contradiction (Case 0). So this sum is 4 and $R_e^0$ consists of two nodes of
degree 1 and one node of degree 2, so both $v_1$ and $v_2$ have one neighbor of
degree 1 and one of degree 2 and $\sigma(v_1) = 6$, again a contradiction ($v_1 \in N_1$
and $\sigma(v_1) < \sigma(u)$).

**Case 2:** two nodes in $U_2$ are adjacent. We select this edge, say $e = \{u, v\}$. Let
$w, x$ be the other neighbors of $u, v$. $|E_e| \leq 7$ because $E_e$ consists of $e$
and 6 edges incident to $R_e^o = \{w, x\}$; $R_e^o$ consists of nodes with all neighbors
in $R_e^o$, the the sum of degrees of nodes of $R_e^o$ is at most 4 and $|R_e^o| \leq 2$. We
conclude that $c(e) \leq 7 + 3(2 + 2 + 2) = 25$.

**Case 3:** a cycle in $H$ has 3 nodes $u, v_0, v_1$, and $u \in U_2$. We select $e = \{u, v_0\}$,
then $R_e^o = \{v_1, w\}$ where $w$ is the non-cycle neighbor of $v_0$, and the sum of
degrees in $R_e^o$ is at most 3; because $U_1 = \emptyset$ we can conclude that $|R_e^o| \leq 1$
and $c(e) \leq |E_e| + 3(2 + |R_e^i| + |R_e^o|) \leq 7 + 3(2 + 2 + 1) = 22$.

**Case 4:** $e = (u, v), u \in U_2$, $e$ is in a 4-cycle of $H$-edges. We select such $e$
so $\sigma(v)$ is minimal. $R_e^o = \{w_0, w_1, w_2\}$ where $w_0$ is the second neighbor of $u$,
$w_1$ is the fourth node of the 4-cycle and $w_2$ is the neighbor of $v$ outside the
cycle.

$$c(e) = |E_e| + 3(2 + |R_e^i|) + 3|R_e^o| - 3Y(E_e)$$
$$= 9 + 15 + 3|R_e^o| - 3Y(E_e)$$
$$= 24 + 3(|R_e^o| - Y(E_e))$$

It suffices to show that $|R_e^o| \leq Y(E_e)$. It is obvious when $R_e^0 = \emptyset$.

For $|R_e^o| > 1$ we will show a contradiction. Observe that the sum of
degrees in $R_e^o$ is at most 4: we can use at most two edges incident to $w_2$, and
at most one edge incident to $w_0, w_1$. Thus all these edges are used to connect
$R_e^o$ to two nodes of degree 2. This implies that $\sigma(v) = 8$ and $\sigma(w_0) = 7$, but
$\sigma(v) \leq \sigma(w_0)$.

For $|R_e^o| > 0$ we will show that a $\Theta$ is contained in $E_e$ and thus $Y(E_e) \geq 1$.
Consider $x \in R_e^o$. If $x$ is adjacent to both $w_1$ and $w_2$, we have a $\Theta$
with 6-cycle $(x, w_1, w_0, u, v, w_2)$. Other possibilities lead to contradictions. Note
that $x \in U_2$ because the neighbors of $x$ are either $w_0, w_1$ or $w_0, w_2$. If $x$
is adjacent to $w_0, w_1$ we have a 3-cycle, hence Case 3. If $x$ is adjacent to $w_0, w_2$
then either $w_2 \in U_2$ and we have Case 2, or $w_2 \in U_3$ and

$$\sigma(w_0) = \text{deg}(u) + \text{deg}(x) + \text{deg}(w_1) < \text{deg}(u) + \text{deg}(w_2) + \text{deg}(w_1) = \sigma(v).$$

**Case 5:** node $u$ has two neighbors in $U_2$, $v_0$ and $v_1$. Let $w_2$ be the third
neighbor of $u$ and $w_0, w_1$ other neighbors of $v_0, v_1$. 

8
Let \( e = \{u, v_0\} \). If \( R^u_e = \emptyset \) then \( c(e) < 25 \): \( R^u_e = \{v_1, w_0, w_2\} \); \( E_e \) consists of \( e \) and edges incident to \( R^u_e \), so \( |E_e| = 9 \) and thus \( c(e) \leq 9 + 3(2 + 3) = 24 \). Thus suppose that \( x_e \in R^u_e \). If \( x_e \) is adjacent to \( v_1 \), then \( x_e \) is equal to \( w_1 \) and it is also adjacent to \( w_2 \). Thus \( x_e \in U_2 \) and it is adjacent to \( w_0, w_2 \).

Next we try to select \( f = \{u, v_1\} \); \( c(f) < 25 \) unless some \( x_f \in U_2 \) is adjacent to \( w_1 \) and \( w_2 \) (by repeating the reasoning).

If \( c(e) > 25 \) and \( c(f) > 25 \), our last attempt is to select \( g = \{w_0, x_e\} \): note that \( w_0 \) has two neighbors in \( U_2 \), namely \( v_0, x_0 \). Let \( y \) be the third neighbor of \( w_0 \); \( R^y_g = \{y, v_0, w_2\} \), and \( g \) is a good selection unless some \( x_g \in U_2 \) is adjacent to both \( y \) and \( w_2 \). Note that we already know all neighbors of \( w_2 \) in \( U_2 \): \( u, x_e \) and \( x_f \); hence \( x_g = x_f \); we also know both neighbors of \( x_f \), namely \( w_1 \) and \( w_2 \), hence \( y = w_2 \). Thus \( c(g) > 25 \) unless \( w_0 \) and \( w_1 \) are adjacent. In that case we have a connected component of \((U, H)\) with \( 8 \) nodes, \( u, w_0, w_1, w_2 \in U_3 \) and \( v_0, v_1, x_0, x_1 \in U_2 \), and with \( 10 \) edges: \( 8 \) edges incident to nodes in \( U_2 \) plus \( \{u, w_2\} \) and \( \{w_0, w_1\} \). Every selection \( e' \) in that component eliminates \( 9 \) nodes and \( 6 \) edges, hence \( c(e') = 27 \), but it also leave \( 2 \) nodes connected with an edge \( e'' \), so the next selection is \( e'' \) with \( c(e'') = 7 \).
Thus two selections together cost $34 < 2 \times 25$.

**Case 6**: the remaining case. Assume that $u \in U_2$, $v_0, v_1$ are the neighbors of $u$ (in $U_3$, otherwise Case 2), and that the other neighbors of $v_0, v_1$ are $w_0, w_1$ and $w_2, w_3$, respectively. Because we do not have Case 5, $|\{w_0, w_1, w_2, w_3\}| = 4$. Let $e = \{u, v_0\}$ and $f = \{u, v_1\}$ (see Figure 6).

Consider selecting $e$, note that $R_e^i = \{v_1, w_0, w_1\}$ and $|E_e| \leq 10$, so $c(e) \leq 10 + 5 \times 3 + 3(|R_e^i| - Y(E_e))$, hence this is a good selection unless $|R_e^i| > Y(E_e)$. Therefore assume that $x_e \in R_e^i$. If $x_e \in U_2$, we have one of the previous cases: if $x_e$ is adjacent to $v_1$ we have Case 2, otherwise $x_e$ is adjacent to $w_0, w_1$ and we have Case 5. In the remaining situation, $x_e \in U_3$ and it is adjacent to all three nodes of $R_e^i$, $w_0, w_1, v_1$, and w.l.o.g. we can assume that $x_e = w_2$, so we have edges $\{w_0, w_2\}$, $\{w_1, w_2\}$.

We repeat the reasoning for $f$, and a prior case occurs or $|R_e^f| = 0$ unless one of $w_0, w_1$ is connected to both $w_2, w_3$. In that situation we have three edges between $w_0, w_1$ and $w_2, w_3$, and thus a matching of two edges, say, $\{w_0, w_2\}$ and $\{w_1, w_3\}$. This matching together with paths $(w_0, v_0, w_1)$ and $(w_2, v_1, w_3)$ forms a 6-cycle, and the third edge between $w_0, w_1$ and $w_2, w_3$, together with this 6-cycle, creates a $\Theta$-graph contained in $E_e$. Therefore $Y(E_e) \geq 1$.

Thus $e$ is a good selection unless $|R_e^i| \geq 2$, so some $x \neq x_e = w_2$ belongs to $R_e^i$. By the above argument, $x$ belongs to the pair $w_2, w_3$, so $x = w_3$ and as $x$ is connected to both $w_0, w_1$, we have all 4 edges between $w_0, w_1$ and

![Figure 6: Diagram for Case 6.](image-url)
$w_2, w_3$. This new edge, together with $\Theta$ we have just described, forms a $\Theta'$, hence $Y(E_e) = 2$, while $|R_e^o| \leq 2$.

**Theorem 1** For every $\varepsilon > 0$ there exists a linear time greedy algorithm that approximates 3-regular-MIM problem with ratio $5/3 + \varepsilon$.

**Proof.** Algorithm BL-ALG can be implemented in polynomial time in 3-regular graphs. Function $c(e)$ is determined by $E_e$, which is a subset of (at most) 13 edges that belong to $E_e$ at the start of the algorithm. Thus it takes $O(1)$ time to compute each $c(e)$. When we select an edge, we remove $O(1)$ edges and nodes, hence we need to recompute only $O(1)$ values of $c(e)$. Moreover, the range of values of $c(e)$ is also $O(1)$, thus selecting an edge that minimizes $c(e)$ takes $O(1)$ time.

To simplify the analysis of the approximation ratio, we can make the initial edge selection as follows. We arbitrarily select node $u_0$ and we remove it from $U$, and the three incident edges form $H$. Then the former neighbors of $u_0$ belong to $U_2$ and all edge selections in that component will be done in the presence of at least one node in $U_1 \cup U_2$. As one of the Cases 1-6 has to apply, every selection will satisfy $c(e) \leq 25$ (after averaging the costs when we select two edges in Case 5).

When BL-ALG starts in a component, we have some $2n$ nodes and $3n$ edges. Later, there will be $t_i$ iterations with $Y(E_e) = i$ used in the computation of $c(e)$ of the selected edge, which means that this run of BL-ALG discovers an edge-disjoint collection of $t_1 \Theta$-graphs and $t_2 \Theta'$-graphs. When the algorithm terminates in a connected component we have $H = U = \emptyset$. Thus we guarantee the selection of at least

$$\frac{3n - 3 + 3(2n - 1) - 3(t_1 + 2t_2)}{25} \geq \frac{3|E| - |X_{M^*}|}{5} - \frac{6}{25} = \frac{6}{5}|M^*| - \frac{6}{25}. $$

One can see that in sufficiently large connected components, of size $\Omega(1/\varepsilon)$, this gives approximation ratio $5/3 + \varepsilon$. In smaller components we can apply an exponential time exact algorithm. $\square$
5 Improved lower bound

Duckworth et al. show that it is NP-hard to approximate 3-MIM with a ratio better than $1 + 1/474$, and that it is similarly hard to approximate 3-regular-MIM with a constant ratio. We improve their analyse.

Definition 8 A combinatorial maximization problem $\mathcal{P}$ has a set of instances $\mathcal{I}$, a size function $\text{Size} : \mathcal{I} \to \mathbb{N}$, and instance $I$ has a set $\text{Valid}(I)$ of valid solutions, and an objective function $\text{Value} : \mathcal{I} \to \mathbb{N}$. We assume that instances have positive sizes and the values of solutions are positive.

An optimum solution for $I$ is $S \in \text{Valid}(I)$ that maximizes $\text{Value}(S)$. We denote it $S^*_I$.

An randomized algorithm provides $\rho$-approximation for $\mathcal{P}$ if on each instance $I$ of $\mathcal{P}$ it runs in time that is polynomial in $\text{Size}(I)$ and with probability at least $2/3$ finds $S \in \text{Valid}(I)$ such that $\rho \text{Value}(S) \geq \text{Value}(S^*_I)$.

Clearly, MIM is a combinatorial maximization problem, the size of an instance can be the number of edges, valid solutions are non-empty induced matching, and the value of a solution is its size.

For a number of combinatorial maximization problems it was proven that, for sufficiently low $\rho$, a $\rho$-approximation algorithm is unlikely to exists, more precisely, its existence would imply that every NP-problem can be solved in randomized polynomial time. Tools to obtain such results are special types of reductions, translations of problems to other problems. Such reductions may be quite involved, and involve randomization. In other cases we use gadget reductions that are simple and deterministic and utilize the results obtained in an involved fashion.

Here we will use an already known result and simple gadget reductions. The known result in question has the form of

$(\alpha - \beta)$-Gap Property: there exists $\varepsilon_0 > 0$ such that for every positive $\varepsilon < \varepsilon_0$ there exists a randomized polynomial time translation of SAT instances of size $n'$ to instances of $\mathcal{P}$ of size $n = p(n')$ such that satisfiable SAT instances are translated into instances that have valid solutions with value at least $(\beta - \varepsilon)n$, while unsatisfiable instances are translated into instances where no valid solution has value larger than $(\alpha + \varepsilon)n$.

If a problem has such a property, then a $\rho$-approximation with $\rho < \beta/\alpha$ could be used to decide SAT instances in randomized polynomial time.

The cornerstone of this approach is the work of Håstad [6] who had shown the $(1/2,1)$ gap property for satisfiability of equations modulo 2 with three
variables per equation. In this problem an instance of size \( n \) is a set of \( n \) equations of the form \( x + y + z = b \), and we want to find an assignment of 0-1 values to the variables that would maximize the number of true equations in the set.

Chlebk and Chlebková [4] used that result to show \((\frac{94}{194}; \frac{95}{194})\)-gap property for 3-regular-MIS, Maximum Independent Set problem restricted to 3-regular graphs. One can use this result as follows:

Given instance \((V, E)\) of 3-regular-MIS,

- translate each node \( u \) into a gadget \( \Gamma_u \) and
- connect the gadgets with additional edges.

Then we can show that an IM \( M \) for the resulting instance can be replaced with IM \( M' \) such that

- \( |M'| \geq |M| \);
- \( M' \) does not contain any additional edges (that connect node gadgets);
- each gadget contains at least \( k \) and at most \( k + 1 \) edges of \( M' \);
- gadgets that contain \( k + 1 \) edges of \( M' \) cannot be connected.

Moreover, given an independent set \( I \subset V \) we can define \( M_I \) such for every \( u \in V \) we have \( |\Gamma_u \cap M| \geq k \) and \( |\Gamma_u \cap M| = k + 1 \) for \( u \in I \).

If a node gadget has \( a \) nodes then this proves \((\frac{194k+94}{194a}; \frac{194k+95}{194a})\)-gap property for a version of MIM, and thus hardness of approximation for a ratio lower than \( \frac{194k+95}{194k+94} = 1 + \frac{1}{194k+94} \).

For 3-MIM, Duckworth et al. describe gadgets with 5 nodes and \( k = 1 \). Consider an instance \((V, E)\) of 3-regular-MIS. It is well-known that in a \( d \)-regular graph the set of edges is a disjoint union of \( d \) perfect matchings, so we assume that \( E = E_0 \cup E_1 \cup E_2 \), where \( E_i \)’s are those perfect matching.

\( \Gamma_u \) is a path of 4 edges and 5 nodes, \((a_0^u, c_0^u, a_1^u, c_1^u)\). We say that \( c \)’s and \( a \)’s are contact nodes and auxiliary nodes.

The additional edges are introduced as follows: if \( \{u, v\} \in E_0 \) we add edge \( \{c_0^u, c_0^v\} \), and if \( \{u, v\} \in E - E_0 \) we add edge \( \{c_1^u, c_1^v\} \).

Consider an IM \( M \) for the resulting instance, we need to show how to replace \( M \) with \( M' \) that properly corresponds to an independent set in \((V, E)\).

We will show how to replace edges that contain nodes of some node gadget \( \Gamma_u \).

Suppose that \( e \in M \) and \( c_0^u \in e \). We can replace \( e \) with \( f = \{a_0^u, c_0^v\} \). Because \( a_0^u \) has only one neighbor, \( D_f \subseteq D_e \) and \( M \) is still an IM.

Next, suppose that \( e \in M \) and \( c_1^u \in e \); we can replace \( e \) with \( f = \{a_2^u, c_1^v\} \). This is clearly fine if \( e = f \), otherwise \( e \) is of the form \( \{c_1^u, c_1^v\} \) and no edge of \( M \) can contain \( a_2^u \). As a result, the only possible edge containing \( a_1^u \) is
Figure 7: Two gadgets $\Gamma_u, \Gamma_v$ (solid edges), contact edges (dashed) and with extra edges (thin or very short thick) added to form $\Gamma'_u, \Gamma'_v$.

$\{c_u^u, a_u^u\}$, but because of our first type of replacement, such edges are not in $M$, and no edge of $M$ contain $a_u^u$. Therefore the replacement does not conflict with other edges of $M$.

After the replacements, all edges of $M$ are inside node gadget, and it is easy to see that $\Gamma_u$ can contain at most two edges of $M$, and if it does, then these edges are $\{a_0^u, c_0^u\}$ and $\{a_2^u, c_1^u\}$. Therefore gadgets that contain two edges of $M$ cannot be connected, and $I_M = \{u \in V : |M \cap \Gamma_u| = 2\}$ is independent.

One can observe that the gadget relies strongly on the auxiliary nodes having less than 3 neighbors. Duckworth et al. describe how to modify the gadget to bring the degree of every node up to 3, this approach increases $k$ from 1 to 6, which proves the hardness of the approximation ratio of $1 + 1/(6 \times 194 + 94)$. Instead, we show how to obtain $k = 2$.

This reduction starts by defining $\Gamma_u$ for each $u \in V$ as before, but to describe the modification, we pair the gadgets of $u$ and $v$ for each $\{u, v\} \in E_0$.

To obtain $\Gamma'_u$ and $\Gamma'_v$ we insert 8 nodes as shown in Figure 7, the left 4 to $\Gamma_u$ and the right 4 to $\Gamma_v$, and we add 14 edges as shown.

It suffices to show that we can replace edges in an IM $M$ so that among the new edges, for each pair of gadget $M$ contains exactly two (one per gadget,
thus increasing $k$ from 1 to 2), marked as very short thick lines in Figure 7. If this is indeed the case, the short thick lines do not conflict with any edges of $\Gamma$ and we can finish the normalizing replacements as before, with the same conclusions.

Consider first the auxiliary structure $S$ that consists of 8 added nodes, together with all 13 edges that are incident to those 8 nodes. We will show that $S$ cannot contain more than 2 edges of $M$: a cycle of $x$ nodes can contain two edges of $M$ only if $c > 5$, thus the only way the upper 6 nodes (and 8 edges) of $S$ can contain two edges of $S$ is as shown (or symmetric), in particular, when those edges contain both left-most and rightmost upper nodes, and in this case no edge of $M$ can contain the lower nodes of $S$. It is also to see that only one edge that contains a lower node of $S$ can be in $S$, and if it does, we can have only one upper edge.

As a result, we can continue our analysis as if $S$ did not exist and $a_1^u, a_1^v$ were nodes of degree 2.

Next, suppose $e = \{a_0^u, c_0^v\} \in M$. If we use it we eliminate nodes $c_0^u, c_0^v, a_2^u, a_2^v$. As a result, nodes $a_1$ are also eliminated, because we have eliminated all their neighbors. We can replace $e$ with $\{c_0^u, c_0^v\} \in M$ and we would eliminate a proper subset of nodes eliminated by $e$.

Finally, suppose $e = \{a_0^u, a_2^v\} \in M$. If we use it we eliminate nodes $c_0^u, a_0^v, c_1^u, a_1^v$. As a result, node $c_0^v$ is already eliminated as we eliminated all its neighbors. We can replace $e$ with $\{a_0^v, a_2^v\}$.

We can conclude the following

**Theorem 2** Assuming that there exists an NP-problems that cannot be solved in random polynomial time, there exists no randomized polynomial time algorithm that approximates 3-MIM with a ratio lower than $1 + 1/288$ and 3-regular-MIM with a ratio lower than $1 + 1/482$.

6 Conclusions and open problems

We believe that all results in this paper can be improved and merit further work.

We conjecture that one can eliminate $e$ from the statement of Theorem 1 by making a more careful first edge choice and proving it by additional case analysis. It is also worth attempting to extend Theorem 1 for the case of 4-regular-MIM and other similar cases.
References


