Isomorphism Certificates for Undirected Graphs

Michael Molloy and Laura Sedgwick [1]
Department of Computer Science
University of Toronto
Toronto, Canada

Abstract

For a graph G, we are interested in a minimum-size subgraph of G which, along with an unlabelled copy of G, provides enough information to reconstruct G. We perform a preliminary study of such subgraphs, focusing on the cases where G is a complete tree or a grid.

1 Introduction

A positive isomorphism certificate (PIC) of a graph, G, is a subgraph $D \subseteq G$ such that if one were given D along with an unlabelled copy of G, one could reconstruct G. D is a spanning subgraph, possibly with several isolated vertices, and so D provides the labels of the vertices of G. Typically, we wish to choose D to have the fewest edges possible.

For example, take G to be the n-star, $K_{1,n}$, where the universal vertex is labelled a and the other vertices are labelled $b_1, b_2, \ldots b_n$. Any spanning subgraph with two edges forms a PIC. Consider the subgraph with edgeset $\{ab_1, ab_2\}$. If we are given these edges, along with an unlabelled copy of $K_{1,n}$ we know immediately that vertex a must be the solitary universal vertex in the graph and the other vertices must all be adjacent to it. Thus we were given sufficient information to reconstruct the graph exactly. However, suppose we only had a subgraph with one edge, say ab_1. Then we might incorrectly reconstruct the graph so that b_1 was the universal vertex. Thus this subgraph does not provide sufficient information.

[1] The work presented in this paper is part of a Masters thesis written by the second author [4].

Isomorphism certificates were first studied on tournaments, and were introduced by Rubenstein [5] who conjectured that no tournament on $n > 5$ vertices has an isomorphism certificate of size less than $n-1$. (The adjective "positive" is redundant in the case of tournaments.) For further work on these and similar tournament certificates, see [1, 2, 3].

Our problem is set apart from a typical reconstruction problem by the fact that we are interested in reconstruction up to identity. Many well known reconstruction problems focus on reconstructing an isomorphic copy of a graph. Here, we are given the isomorphic copy of the graph and are interested in taking the problem one step further, by trying to recreate the original labelled graph.

The next section provides a few basic definitions. This is followed by two large sections presenting the main results from [4]. The first presents work on certificates for trees and the second for grids. We conclude with remarks on possible future work. Some details of the proofs are long but straightforward, and so for the sake of brevity we omit them. For complete details, we refer the reader to [4].

2 Definitions

We say that a graph G *contains* $D = (D', D'')$, written $G \succeq D$, provided D' is a spanning subgraph of G and D'' is a spanning subgraph of $\overline{G}$. We refer to the edges of D'' as non-edges of G.

We shall say that D is an *isomorphism certificate, (IC),* for G provided that:

1. $G \succeq D$.
2. If $G' \cong G$ and $G' \succeq D$ then $G' = G$.

In other words, given D along with an unlabelled copy of G, one can reconstruct G.

A *positive isomorphism certificate, (PIC)*, is an IC where $E(D'') = \emptyset$, a *negative isomorphism certificate, (NIC)*, of a graph G is an IC where $E(D') = \emptyset$ and a *mixed isomorphism certificate (MIC)* is an IC where $E(D'), E(D'') \neq \emptyset$. In this paper we shall focus on PIC's. For a discussion of the other types of certificates, see [4].

The *size* of an IC $D = \{D', D''\}$ is $|E(D')| + |E(D'')|$. An IC for a graph G is a *minimum IC* if G has no IC of smaller size. The *IC number* of a graph G, $IC(G)$, is the size of a minimum IC. These definitions are extended in the obvious manner to PIC's, NIC's and MIC's. The general problem that we are interested in is to determine these parameters for various graphs. In this paper, we will focus mainly on the PIC number.

Throughout this paper we shall relax the distinction between the pair $D = (D', D'')$ and the elements of the pair, D' and D''. For example, we shall often refer to the edge (u, v) as being in D where more properly $(u, v) \in D'$, and we often say that (u, v) is a *nonedge* of D if $(u, v) \in D''$. We refer to a vertex as being isolated in D when in fact we mean isolated in both the subgraphs induced by D' and D''. When it is understood that D is a PIC of G, then it is convenient to relax the distinction further, identifying D with D', and so we simply refer to D as being a subgraph of G.

3 Trees

In this section we focus on optimal PIC's for complete binary trees and complete k-ary trees. We also provide some discussion of NIC's and MIC's.

All trees in this section are rooted and we say that the root is at level zero of the tree. The vertices at level j are the children of the vertices at level $j-1$. The height of a tree is the length of the longest path from the root, i.e. the label of the last non-empty level. We denote the complete binary tree of height i by T_i, and we denote the complete k-ary tree of height i by T_i^k.

3.1 PIC number for T_i.

First we establish the PIC number T_i:

Theorem 1 *$PIC(T_i) = 2^{i+1} - 2i$.*

This is established for $i \leq 3$ in [4], and we omit the details here. Thus we assume $i \geq 4$.

To prove this theorem, the first step is to define a subgraph $D_i \subset T_i$ which we will prove to be a PIC.

To start, we recursively define $H_i \subset T_i$ as follows. Set a_0 to be the root of the tree, and for $j = 1, 2, 3$ let a_j, b_j be the left and right children of a_{j-1} respectively. Let c, d be the left and right children of b_1 respectively. To form H_i, we make the following modifications to T_i:

1. Delete the 2 edges descending from b_2 and the 2 edges descending from b_3.
2. Replace the edgeset of the subtree rooted at c by a copy of H_{i-2}.

To form D_i, we delete the two edges descending from b_1 from H_i. The case $i = 6$ is illustrated in Figure 3.1. Clearly D_i has $2^{i+1} - 2i$ edges.

As noted in [4], this PIC is not unique - T_i has other nonisomorphic PIC's of the same size.

Our proof consists of two steps: (i) proving that D_i is a PIC for T_i, and (ii) proving that no smaller PIC exists.

Step 1: D_i is a PIC for T_i.

We wish to show that if T' is isomorphic to T_i and if $T' \succeq D_i$ then $T' = T_i$. We will do so via the following series of claims. The claims have been stated and the proofs sketched. Many of the readers may choose to only read the claims, as the proofs are composed of a number of small details, some of which have been omitted. For more information, the reader can refer to [4].

Claim 2 *If $T' \succeq D_i$ and is isomorphic to T_i then a_2 is at level two of T'.*

Using the fact that a_2 is adjacent to the root of a complete binary subtree of height $i - 3 \geq 1$, namely that rooted by a_3, it follows that a_2 must be the parent of a_3 in T', and this implies that a_2 must be at level zero, one or two of T'.

Since $\deg(a_2) > 2$, it is not the root of T'.

If a_2 is at level one of T', then a_3 would have to be at level two of T', since it has degree 3 and so cannot be the root. Thus, all descendents of a_3 of height at most $i-1$ in T' are connected in D_i, and all of the level i descendents (i.e. leaf descendents) must be singleton components in D_i. However, since $i \geq 4$ there are at least 4 such descendents but only 2 singleton components in D_i. Therefore a_2 cannot be at level one of T', and so it must be at level two, as claimed.

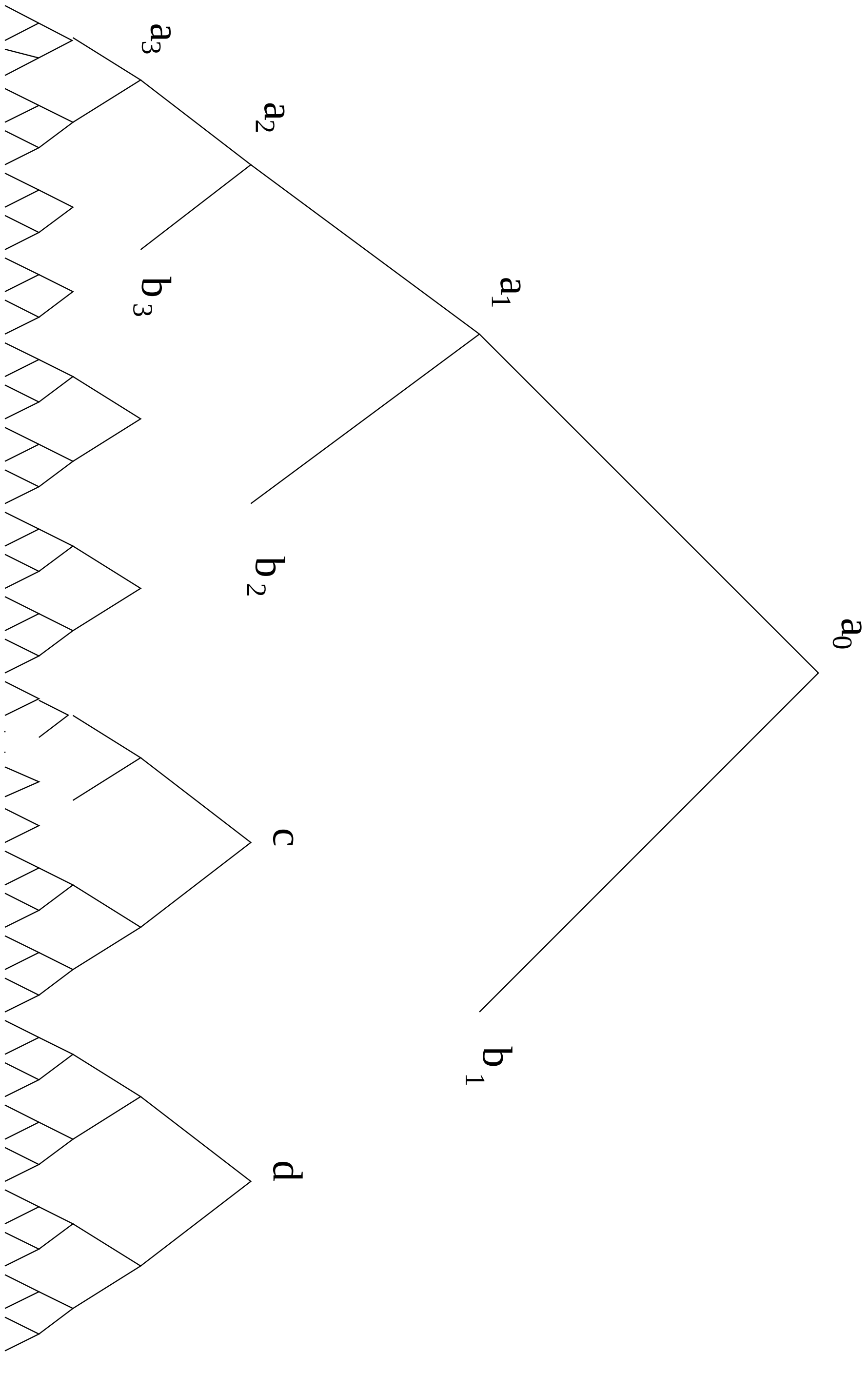

Figure 1: D_6 - PIC for T_6

Claim 3 *If $T' \succeq D_i$ and is isomorphic to T_i then a_1 is at level one of T' and a_3, b_3 are at level three of T'.*

Since a_2 is at level two of T', either a_1, a_3 or b_3 is its ancestor at level one of T'. A simple and straightforward argument based on the number of vertices of degree $0, 1$ and 2 in D_i yields that the root of T' must have degree 2 in D_i. Using this fact it is not hard to show that the ancestor of a_2 cannot be a_3 or b_3, thus establishing the claim. We omit the details.

Claim 4 *If $T' \succeq D_i$ and is isomorphic to T_i then a_0 is the root of T'.*

Since the root must have degree 2 in D_i (as mentioned earlier) and it must be adjacent to a_1 by Claim 3, the root must be a_0.

Claim 5 *If $T' \succeq D_i$ and is isomorphic to T_i then b_1 is at level one of T' and c, d are its descendents.*

That b_1 is at level one, follows immediately from Claim 4.

Note that d is the root of a subtree of height $i - 2$. It is straightforward to prove inductively that if v is the root of a subtree of height k in D_i then v must be at level at most $i - k$ in T', since v has two neighbours which are (by induction) at level at most $i - k + 1$ and at least one of them must be a child of v in T'. Therefore, d must be at level at most 2. The only remaining place for it is as a descendent of b_1.

Using the fact that c is adjacent to the root of a subtree of height $i - 3$, and is not part of that subtree, it is not hard to argue that c must be at level at most 2 and so is also a descendent of b_1.

The locations in T' of all descendents in T of a_1 are now easily determined. All that remains is to determine the location in T' of all descendents in T of c. This part follows from showing that for all $j \geq 1$, H_j is a PIC for T_j. It is straightforward to verify this for $j = 1, 2, 3$. In the case $j \geq 4$, note that since $D_j \subseteq H_j$, the preceding portion of this proof provides an inductive proof that H_j is a PIC for T_j as required. $\square$

Step 2: T_i does not have a smaller PIC.

Let D be any PIC for T_i . We will show that D contains all but at most $2i - 2$ of the $2^{i+1} - 2$ edges of T_i.

Claim 6 *D is missing at most two edges at any level. Furthermore, those two edges must have the same parent.*

It is easy to see that if any two non-siblings x, y at the same level in T_i do not have their ancestral edges in D, then by "exchanging" the subtrees rooted at x, y we create a tree $T' \neq T$ such that T' also contains D.

Claim 7 *If D is missing more than $2i-2$ edges, then the root of T_i has degree at least 1 in D.*

To prove Claim 7, we note that Claim 6 implies that if D is missing more than $2i-2$ edges, then at least one leaf of T_i is isolated in D. If the root of T_i were also isolated in D, then by exchanging the labels of the root and that leaf, we create a tree T' which is isomorphic to but not equal to T_i, and which also contains D. This contradicts the fact that D is a PIC.

Claims 6, 7 imply that D is missing at least $2i-1$ edges. Suppose that D is missing exactly $2i-1$ edges. Then exactly one of the 2 edges from the root is missing. Consider the two vertices at level one of T_i . One of them, say u, must be missing the edges to both of its children. If u is not connected to the root in D then u is an isolated vertex, while if u is connected to the root, then that edge is an isolated edge. In the first case we arrive at a contradiction in the same manner as in the proof of Claim 7, by considering exchanging the labels of that vertex and a leaf which is isolated in D. In the second case, by considering exchanging the labels of that vertex and the root, we arrive at a similar contradiction. $\square$

For the more general case of complete k-ary trees, we have the following:

Theorem 8 *For $i \geq 1, k \geq 3$, $PIC(T_i^k) = (\sum_{j=1}^{i} k^j) - (ki-2)$.*

The certificate and proof are very similar to those corresponding to Theorem 1. For details, see [4].

3.2 I.C.s and NIC's for T_i and T_i^k.

We conclude with a theorem which shows that the use of non-edges in an IC does not allow us to improve upon the size of our minimum PIC for T_i^k:

Theorem 9 *For $k \geq 2, i \geq 4$, $IC(T_i^k) = PIC(T_i^k)$.*

As any PIC is already an IC to prove this theorem we must only show that the $PIC(T_i^k)$ can not be improved upon by using a combination of edges and non-edges.

Outline of Proof: This theorem is proved by an argument similar to the one employed to prove Step 2 of Theorem 1. It can first be shown if more than k edges are cut between two levels then we must compensate for this fact with extra non-edges. We then make similar arguments showing that if any edges from the root are missing then our certificate requires extra non-edges. For the details, see [4].

4 Grids

This section focuses on determining the value of the PIC number for $m \times m$ grids. We have not been able to exactly identify this number in the general case, but we show that the PIC number for the $m \times m$ grid is asymptotic to the number of vertices in the grid. Smaller grids require a case intensive study in order to determine their PIC number and as these cases are of minimal interest we have ignored them.

Let G_m denote the $m \times m$ grid on m^2 vertices which is depicted in figure 2. We refer to the vertex in row i, column j as $g_{i,j}$. In this section, we will prove that $PIC(G_m) = m^2 - \mathrm{o}(m^2)$.

4.1 A PIC for G_m

For $m \geq 14$ let D_m denote the spanning subgraph of G_m with the following edges:

- $\{g_{4,j}g_{4,j+1} : 1 \leq j \leq m-1\}$
- $\{g_{i,4}g_{i+1,4} : 1 \leq i \leq m-1\}$
- $\{g_{i,m-3}g_{i+1,m-3} : 1 \leq i \leq m-1\}$
- $\{g_{i,j}g_{i,j+1} : 1 \leq i \leq m, m-4 \leq j \leq m-1\}$

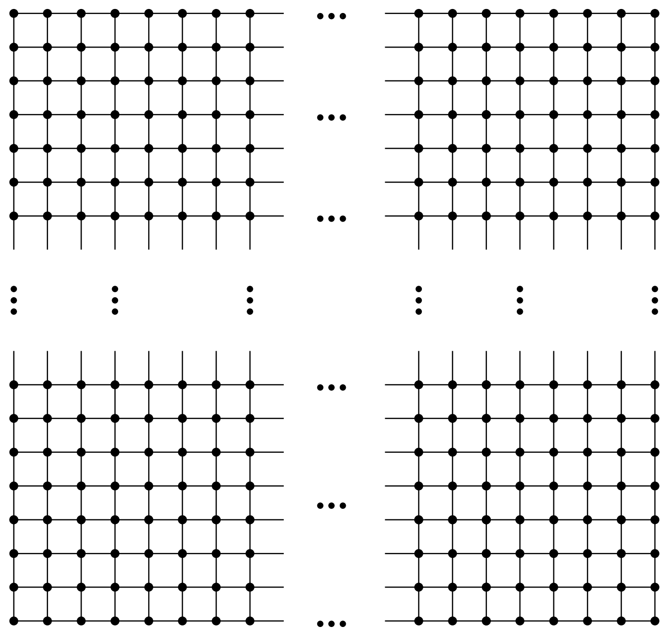

Figure 2: G_m

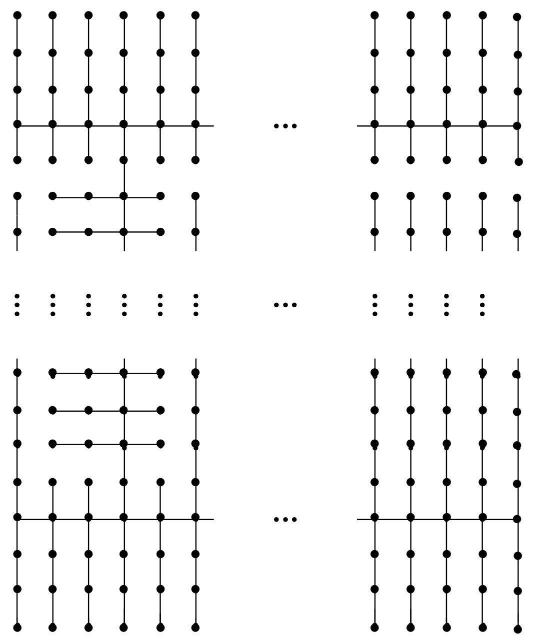

Figure 3: D_m

- $\{g_{i,j}g_{i,j+1} : 1 \leq i \leq m, i \neq 2,3,4,5, 1 \leq j \leq m-6\}$
- $\{g_{i,j}g_{i,j+1} : 2 \leq i \leq 5, 1 \leq j \leq 4\}$
- $\{g_{i,j}g_{i+1,j} : 2 \leq i \leq 4, 6 \leq j \leq m-5\}$

D_m is depicted in figure 3. We will prove that D_m is a valid PIC for G_m, thus showing:

Theorem 10 *For $m \geq 14$, $PIC(G_m) \leq m^2 - 1$.*

Proof Consider any graph G', $G' \cong G_m$ and $G' \succeq D_m$. We shall show $G' = G_m$.

Throughout this proof we denote by $a_{i,j}$ the vertex at row i, column j of G'. This vertex is equal to a vertex $g_{i',j'}$ for some i', j' and we write $a_{i,j} \equiv g_{i',j'}$.

The proof follows from a series of claims whose proofs have again been sketched. Refer to [4] for more details.

We begin by showing that the interior vertices in the fourth column of D_m must all lie in the same column of G'.

Claim 11 *For some k, $g_{i,4} \equiv a_{i,k}, 2 \leq i \leq m-1$.*

We prove this by considering an arbitrary vertex in column i of G_m, $g_{i,4}$ and its neighbour $g_{i-1,4}$. We can assume WLOG they are both in the same column of G', i.e. that $g_{i,4} \equiv a_{j,k}$ and $g_{i-1,4} \equiv a_{j-1,k}$. Now consider $g_{i+1,4}$. As it is adjacent to $g_{i,4}$ in D_m we know it is equivalent to one of $a_{j,k-1}, a_{j,k+1}, a_{j+1,k}$. However if it is either of the first two it must have a neighbour other than $g_{i,4}$ in common with $g_{i-1,4}$ which it does not. Therefore $g_{i+1,4} \equiv a_{j+1,k}$. Repeating this argument yields the claim.

We can apply the same argument to the fourth last column of D_m:

Claim 12 *For some l, $g_{i,m-3} \equiv a_{i,l}, 2 \leq i \leq m-1$.*

Our next step is to show the vertices in the fourth row of D_m, lying between columns 4 and $m-3$, are in the fourth row of G'.

Claim 13 $g_{4,j} \equiv a_{4,k+j-4}, 4 \leq j \leq m-3.$

Similar reasoning to that used in the proof of Claim 11 shows that these vertices are all in the same row. The fact that $g_{4,4} = a_{4,k}$ implies that this row is the fourth row.

Next, we fix the location of the fourth and fourth last columns. We begin by showing that the fourth column of D has index at least 4 in G'.

Claim 14 $k \geq 4.$

By claim 11, $g_{i,4} \equiv a_{i,k}, 2 \leq i \leq m-1$. For every $m-6 \leq i \leq m-2$, consider $g_{i,3}$ and $g_{i,5}$. Each of these vertices is adjacent to a vertex in column k and is not in column k. Thus we have identified the vertices in the set $\{a_{i,j} : i = k-1, k+1, m-6 \leq k \leq m-2\}$. So consider $g_{m-5,2}, g_{m-4,2}, g_{m-3,2}, g_{m-5,6}, g_{m-4,6}, g_{m-3,6}$. They must be equal to the six vertices in the set $\{a_{i,j} : i = k-2, k+2, m-5 \leq j \leq m-3\}$. Finally consider $g_{m-4,1}$ and $g_{m-4,7}$. They must be equal to $\{a_{m-4,k-3}, a_{m-4,k+3}\}$. Thus $k-3 \geq 1$ and $k \geq 4$.

The same argument yields a similar bound on l:

Claim 15 $l \leq m-3.$

Claim 13 implies that $l - k \geq m-7$. Therefore, we have the following:

Claim 16 $k = 4$ *and* $l = m-3.$

We are finally ready to focus on the vertices at the ends of columns 4 and $m-3$ in D. An argument very similar to that used for Claim 14 implies that these vertices are also the ends of the same columns in G', thus yielding:

Claim 17 $g_{i,4} \equiv a_{i,4}, 1 \leq i \leq m$ *and* $g_{i,m-3} \equiv a_{i,m-3}, 1 \leq i \leq m.$

At this point, we have fixed the locations of the vertices in columns 4 and $m-3$, as well as those in the fourth row extending between those columns. A simple counting argument establishes that all vertices in columns 1, 2, 3 of D must lie in columns 1, 2, 3 of G'. Similarly, those in columns $m-2, m-1, m$ of D lie in columns $m-2, m-1, m$ of G. After this, it is a simple matter to show that $G' = G$. For more details, see [4]. □

4.2 Lower bounds on the PIC number for G_m.

Here, we show that the bound in theorem 10 is asymptotically correct, i.e. that the PIC number of G_m is at least $m^2 - o(m^2)$. We prove this with the following general theorem:

Theorem 18 *Let $\mathcal{G}$ be a class of graphs such that for any graph $G \in \mathcal{G}$ on n vertices every non-identity automorphism of G has at most $n - \omega(n)$ fixed vertices, where $\omega(n)$ tends to ∞ with n. Then the PIC number for any graph $G \in \mathcal{G}$ on n vertices is at least $(1 - o(1))n$.*

Proof Fix any $\epsilon > 0$. We will show that for n sufficiently large, all PIC's for graphs on n vertices have at most ϵn components, and thus have at least $(1-\epsilon)n$ edges.

We will make use of the following trivial fact. The bound in this fact can be easily improved, but the stated bound will suffice for our purpose.

Fact: For any $j > 0$ there are at most $2^{\binom{j}{2}}$ non-isomorphic connected graphs with j vertices.

Proof of Fact: This follows immediately from the fact that there are exactly $2^{\binom{j}{2}}$ graphs with vertex set $\{1, ..., j\}$.

Now, choose n such that (i) $n > \frac{2}{\epsilon} \sum_{j<\frac{4}{\epsilon}} 2^{\binom{j}{2}}$ and (ii) for any graph $G \in \mathcal{G}$ on n vertices each automorphism of G leaves at most $n - \frac{8}{\epsilon}$ fixed vertices.

Now pick any $G \in \mathcal{G}$ which has n vertices, and let D be any PIC of G. Let c_j be the number of components with j vertices in D and let $C = \sum_{j \geq 1} c_j$ be the total number of components in D.

Suppose D has two isomorphic components C_1, C_2 of size x. Consider the bijection of $V(G)$ onto itself which maps each vertex of C_1 onto the corresponding vertex of C_2 and vice versa, and which leaves every other vertex fixed. This bijection produces a graph G' which also contains D, and so if D is a PIC, then this bijection must be an automorphism. Since every automorphism of G leaves at most $n - \frac{8}{\epsilon}$ fixed vertices, we have that $x \geq \frac{4}{\epsilon}$. Therefore, by our Fact, for each $j < \frac{4}{\epsilon}$, $c_j \leq 2^{\binom{j}{2}}$. Therefore, $\sum_{j<\frac{4}{\epsilon}} c_j \leq \sum_{j<\frac{4}{\epsilon}} 2^{\binom{j}{2}} \leq \frac{\epsilon}{2} n$.

Recall that our goal is to show $C \leq \epsilon n$. If $C > \epsilon n$, then $\sum_{j \geq \frac{4}{\epsilon}} c_j > \frac{\epsilon}{2} n$, and so

$$\sum_{j \geq 1} j c_j \geq \sum_{j \geq \frac{4}{\epsilon}} \frac{4}{\epsilon} c_j \geq \frac{4}{\epsilon} \sum_{j \geq \frac{4}{\epsilon}} c_j \geq \left(\frac{4}{\epsilon}\right) \left(\frac{\epsilon}{2} n\right) = 2n,$$

which contradicts the fact that $\sum_{j \geq 1} j c_j = n$. Therefore, the number of components of D is less then ϵn, and so $PIC(G) \geq (1 - \epsilon) n$. Since this is true for every $\epsilon > 0$ and n sufficiently large in terms of ϵ, our theorem follows. □

Remark: This proof technique is similar to that used by Alon to prove Proposition 3 of [5].

Theorem 18 yields our asymptotic bound on $PIC(G_m)$.

Theorem 19 $PIC(G_m) = m^2 - o(m^2)$.

Proof G_m has only 8 automorphisms (4 rotations each possibly combined with 1 reflection). Each of these leaves at most m fixed vertices. Therefore, the result follows from Theorem 18. □

5 Conclusions

The results presented in this paper only begin to scratch the surface of the many questions that may be asked - and hopefully someday answered - on certificates. Here we present just a few of these questions.

To further the collection of work on certificates for undirected graphs a natural approach would be to consider other specific graphs or classes of graphs. Is there anything that can be said about the IC number for planar graphs, trees, bipartite graphs etc. We are curious not only about bounds on the value of these IC numbers but also about the algorithmic complexity of determining them and of generating ICs for these graphs.

The reader might also be interested in considering more general questions. Is it possible to bound the IC number for arbitrary graphs as a function of some property of the graph? Is it possible to show that most graphs have IC numbers bounded above (or below) by a function of the number of vertices in the graph?

We hope that in the next few years these questions shall provide the combinatorics community with many fruitful hours of contemplation and discussion.

References

[1] N. Alon and M. Ruszinkó. *Short certificates for tournaments.* Elec. J. Comb. **4**, R12, (1997).

[2] P. Fishburn, J.H. Kim and P. Tetali. *Tournament certificates.* Manuscript.

[3] P. Fishburn, J.H. Kim and P. Tetali. *Score certificates for tournaments.* J. Graph Th. **24**, 117 - 138 (1997).

[4] L. Sedgwick. *Certificates for Undirected Graphs.* M.Sc. Thesis, Dept. of Computer Science, University of Toronto (1998). Available as CSRG Tech Report CSRG-435 at `www.cs.utoronto.ca/csri/reports.html`

[5] A. Rubenstein. *Why are certain properties of binary relations relatively more common in natural language?* Econometrica **64**, 343 - 356 (1996).