Perfect Completion and Deletion in Random Graphs

Assaf Natanzon *   Ron Shamir * †

Abstract

In the Perfect Completion problem one wishes to add the fewest possible edges to a graph in order to obtain a perfect graph. How large can the size of the added edge set be compared to the size of the edge set of the original graph? We show that with high probability the smallest perfect graph containing a random graph $G(n, n^{0.6})$ has $\Omega(n^{1.8})$ edges. We also show that with high probability the minimum edge set that must be deleted from a random graph $G(n, 0.5)$ has $\Omega(n^2)$ edges.

1 Introduction

Edge Completion/Deletion problems call for making minimum changes to the edge set of an input graph in order to obtain a graph with a desired property. Let $\mathcal{F}$ be a family of graphs, and let $G = (V, E)$. In the *$\mathcal{F}$-completion problem* one seeks an edge set F such that $G(V, E \cup F) \in \mathcal{F}$ and $|F|$ is minimum. In the *$\mathcal{F}$-deletion problem* we seek a set $F \subseteq E$ s.t. $(V, E \setminus F) \in \mathcal{F}$ and $|F|$ is minimum. The deletion problem is equivalent to finding a maximum subgraph of G with property $\mathcal{F}$. In the completion problem, we seek a minimum supergraph of G with property $\mathcal{F}$.

Completion and deletion problems are fundamental in graph theory. Already in 1979, Garey and Johnson mentioned 18 different types of vertex and edge completion/deletion problems [5, Section A1.2]. Completion and deletion problems have applications in several fields, including molecular biology and numerical algebra. For basic graph theoretic terminology, see e.g. [7].

Most of the results obtained so far concerning (the decision versions of) deletion/completion problems are NP-completeness results. Chain Completion and Chordal Completion were shown to be NP-complete by Yannakakis [15]. As noted in [6], the NP-completeness of Interval Completion and Unit Interval Completion also follows from [15]. Interval Completion was directly shown to be NP-complete in [5, problem GT35] and [9]. Deletion problems on interval graphs and unit interval graphs were proved to be NP-complete in [6]. For comparability graphs, completion was shown to be NP-complete by Hakimi and Young [8] and deletion was shown to be NP-complete by Yannakakis [16]. Natanzon et al [12] have shown the NP-hardness of several edge modification problems including comparability editing, perfect editing, deletion and completion. Approximation algorithms exist for variants of chordal completion [1, 11] and interval completion [13].

A natural combinatorial question arises when dealing with completion problems. How large can the completion set size be compared to the original graph size? In this paper we study lower bounds for sizes of the modified edge sets with respect to perfect completion and deletion. For a graph $G = (V, E)$ we denote $|V| = n, |E| = m$. For a family of graphs $\mathcal{F}$ define $\Phi_{\mathcal{F}}(m, n) = max_G min\{|\bar{E}| : G(V, E \cup \bar{E}) \in \mathcal{F}\}$. For a family of graphs $\mathcal{F}$ define $\Psi_{\mathcal{F}}(m, n) = max_G min\{|\bar{F}| : G(V, E \setminus \bar{F}) \in \mathcal{F}\}$.

*Department of Computer Science, Tel Aviv University, Tel-Aviv, Israel. Email: {natanzon,shamir}@math.tau.ac.il.

†corresponding author. Supported in part by the Israel science Foundation founded by Israeli Academy of Sciences and Humanities.

Chung and Mumford [4] have shown that the smallest chordal supergraph containing the random graph $G(n, \frac{k}{n})$ has $\Omega(n^2)$ edges, and that the smallest chordal supergraph containing a planar graph contains $O(nlog(n))$ edges.

We show here that the smallest perfect graph containing a random graph $G(n, n^{-0.6})$ has $\Omega(n^{1.8})$ edges and that the minimum perfect deletion set of a random graph $G(n, 0.5)$ has $\Omega(n^2)$ edges. We also give a simple example of a graph with n vertices and $O(n^{1.5})$ edges whose smallest chordal supergraph contains $O(n^2)$ edges.

Our lower bounds for Perfect Completion and Deletion imply the same lower bounds for all sub-families of the perfect graphs, such as chordal graphs, comparability graphs, interval graphs and many more.

2 Background

We start with some background and definitions. For more background on random graphs see [2].

Let n be a positive integer, and let $0 \leq p \leq 1$. A random graph $G(n, p)$ is a probability space over the set of graphs on the vertex set $\{1, \ldots, n\}$ determined by $Pr[(i, j) \in G] = p$, independently for each pair (i, j).

We cite below some basic bounds on large deviations, which will be used later.

Theorem 2.1 *Chernoff bounds [3]: Let $p \in [0, 1]$, $X_1, \ldots, X_n$ be mutually independent random variables with $Pr[X_i = 1] = p$ and $Pr[X_i = 0] = 1 - p$, and define $X = X_1 + \ldots + X_n$. Then:*

1. *$Pr[X - np > a] < e^{a - pnln(1+\frac{a}{pn}) - aln(1+\frac{a}{pn})}$.*
2. *If $p = \frac{1}{2}$ then $Pr[|X - \frac{n}{2}| > a] < 2e^{\frac{-2a^2}{n}}$*
3. *When $p = o(1)$, $a = O(pn)$, and $a > \frac{2pn}{3}$ we have a simple bound:*

$$Pr[X - np > a] \leq Pr[X - np > \frac{2pn}{3}] < e^{\frac{-2pn}{27}}.$$

Theorem 2.2 *Chebychev Inequality [14]: For any positive λ*

$$Pr\left[|X - E[X]| \geq \lambda\sqrt{Var[X]}\right] \leq \frac{1}{\lambda^2}$$

An event A on a family of random graphs $G(n, p)$ happens *with high probability* if $\lim_{n \to \infty} Pr(A) = 1$.

3 Perfect Completion

A graph is called *perfect* if every induced subgraph has a chromatic number which equals its maximum clique size (cf. [10, 7]). In this section we study the random graph $G(n, \frac{1}{n^{0.6}})$. We shall show that with high probability G has $O(n^{1.4})$ edges but the smallest perfect graph that contains it has $\Omega(n^{1.8})$ edges.

Claim 3.1 *For a random graph $G(n, \frac{1}{n^{0.6}})$, with high probability $m = O(n^{1.4})$.*

Proof: Let m denote the random variable of the number of edges in G. Define a random variable e_i for each edge i where $e_i = 1$ if $i \in E$ and $e_i = 0$ otherwise. Then $m = \sum_{i=1}^{\binom{n}{2}} e_i$. Since the variables are independent and $Pr(e_i = 1) = \frac{1}{n^{0.6}}$, $E(m) = \binom{n}{2} * \frac{1}{n^{0.6}}$. By Theorem 2.1 (3) (taking, say, $a = n^{1.4}$), $Pr[m > E(m) + n^{1.4}] = Pr[m > \frac{1}{n^{0.6}}\binom{n}{2} + n^{1.4}] < e^{-\frac{2n^{1.4}}{27}}$. So with high probability $m = O(n^{1.4})$. ■

Claim 3.2 *With high probability the number of labeled* 5*-cycles in* G *is* $\Theta(n^2)$.

Proof: The number of possible labeled 5-cycles in a graph is $\frac{4!}{2}\binom{n}{5}$ (we have to choose the 5 vertices of the cycle and order them). For every 5-cycle we will define an indicator X_i, $0 < i \leq \frac{4!}{2}\binom{n}{5}$ with $X_i = 1$ if that 5-cycle is present in G and $X_i = 0$ otherwise. Define $X = \sum X_i$. By linearity of expectation we have

$$E[X] = \frac{4!}{2}\binom{n}{5}p^5$$

i.e., $E[X] = \Theta(n^2)$.

$$Var[X] = \sum Var[X_i] + \sum_{i \neq j} Cov[X_i, X_j]$$

since X_i is an indicator

$$Var[X_i] = p_i(1 - p_i) \leq p_i = E[X_i]$$

we have

$$Var[X] \leq E[X] + \sum_{i \neq j} Cov[X_i, X_j].$$

For indices i, j, we shall write $i \sim j$ if $i \neq j$ and the events X_i, X_j are not independent. If $i \sim j$

$$Cov[X_i, X_j] = E[X_iX_j] - E[X_i]E[X_j] \leq E[X_iX_j] = Pr[X_i \wedge X_j]$$

and therefore

$$Var[X] \leq E[X] + \sum_{i \sim j} Pr[X_i \wedge X_j]$$

In our case $i \sim j$ if X_i and X_j are different and have common edges.

X_i and X_j may have one, two or three common edges. There are $O(n^8)$ sets with one common edge and the probability for each such set is p^9, there are $O(n^7)$ sets with two common edges and the probability for each such set is p^8 and there are $O(n^6)$ sets with three common edges and the probability for each such set is p^7 .
Therefore

$$Var[X] = O(n^2) + O(n^8p^9) + O(n^7p^8) + O(n^6p^7) = O(n^{2.6}).$$

Using Chebychev Inequality (Theorem 2.2), with $\lambda = \sqrt{n}$

$$Pr\left[|X - E[X]| > \sqrt{n}\sqrt{Var[X]}\right] \leq \frac{1}{n}. \tag{3.1}$$

Hence, with probability larger than $1 - \frac{1}{n}$,

$$|X - E[X]| \leq \sqrt{nVar[X]}$$

Since $Var(X) = O(n^{2.6})$, for n sufficiently large $Var(X) < Cn^{2.6}$ for some constant C. Since $E(X) = \Theta(n^2)$, for n sufficiently large, $An^2 < E(X) < Bn^2$ for some constants A, B. Since w.h.p. $|X - E[X]| \leq \sqrt{nVar[X]} \leq \sqrt{nCn^{2.6}}$, it follows that for n sufficiently large $An^2 - \sqrt{C}n^{1.8} < X < Bn^2 + \sqrt{C}n^{1.8}$. Therefore w.h.p. $X = \Theta(n^2)$. ■

Claim 3.3 *With high probability the degree of every vertex in G is at most $2n^{0.4}$.*

Proof: For $2 \leq i \leq n$ let $X_i = 1$ if $(1, i) \in E$ and $X_i = 0$ otherwise. Define $X = \sum_{i=1}^{n} X_i$ by Theorem 2.1 (3) we have

$$Pr(X - n^{0.4} > n^{0.4}) < e^{\frac{-2n^{0.4}}{27}} < \frac{1}{n^2}$$

(the last inequality is true for large enough n). Therefore the probability that there is a vertex with degree higher than $2n^{0.4}$ is less than $\frac{1}{n}$. ■

Claim 3.4 *With high probability there is no pair of vertices $i, j \in V$ for which there exist $\Omega(n^{0.2})$ pairs of vertices $k, l \neq i, j$ such that $(i, k), (k, l), (l, j) \in E$.*

Proof: For a pair of vertices i, j let us look at the set $B_{i,j} = Adj(i) \cup Adj(j) \setminus \{i, j\}$. Since with high probability the degree of every vertex is less than $2n^{0.4}$, with high probability $|B_{i,j}| < 4n^{0.4}$ for all i, j. We now argue that the probability that $|E(G_{B_{i,j}})| = \Omega(n^{0.2})$ is less than $\frac{1}{n^3}$. To see this let us look at the graph $G(t = 4n^{0.4}, n^{-0.6})$. With probability larger than $1 - \frac{1}{n^3}$, G has $\Theta(\binom{t}{2} n^{-0.6}) = \Theta(n^{0.2})$ edges (the proof is similar to that of Lemma 3.1. The number of pairs k, l such that $(i, k), (k, l), (l, j) \in E$ is less than the number of edges in $G_{B_{i,j}}$, therefore with probability less than $\frac{1}{n}$ there is a pair i, j with $\Omega(n^{0.2})$ such paths. ■

Let $i, j, k \in V$. We say that k is a *common neighbor* of i and j if $(k, i) \in E$ and $(k, j) \in E$.

Claim 3.5 *Let $X_{i,j}$ be the number common neighbors vertices of vertices i and j in V. With high probability $X_{i,j} < 20$ for all i, j.*

Proof: Let $X_{i,j}^k = 1$ if $(i, k) \in E$ and $(j, k) \in E$, and define $X_{i,j} = \sum X_{i,j}^l$,

$$Pr[X_{i,j} > C] = \sum_{i=C}^{n} \binom{n}{i} n^{-1.2i} (1 - n^{-1.2})^{(n-i)} \leq \sum_{i=C}^{n} n^i n^{-1.2i} =$$

$$= \sum_{i=C}^{n} n^{-0.2i} < n * n^{-0.2C} < \frac{1}{n^3}.$$

Therefore, with high probability for all i, j is $X_{i,j} < C$. ■

Claim 3.6 *With high probability there is no pair of vertices $i, j \in V$ that are contained in $\Omega(n^{0.2})$ 5-cycles such that i, j are not neighbors in those 5-cycles.*

Proof: Suppose i, j are contained in a 5-cycle such that i, j are not neighbors in this 5-cycle. Then there exist three vertices k, l, m such that $(i, k), (j, k) \in E$ and $(i, l), (l, m), (j, m) \in E$. By the previous claims there are at most $C \leq 20$ such k's and $O(n^{0.2})$ such pairs of l, m. Therefore there are $O(n^{0.2})$ such 5-cycles. ■

Theorem 3.7 *For the random graph $G(n, n^{-0.6})$, with high probability G has $O(n^{1.5})$ edges and the smallest perfect graph containing G has $\Omega(n^{1.8})$ edges.*

Proof: Consider a minimal perfect graph G' containing G. Every 5-cycle in G' has a chord. From Claim 3.2 G has $\Theta(n^2)$ 5-cycles (not necessarily induced), and from Claim 3.6 an edge can be a chord of $O(n^{0.2})$ 5-cycles. Therefore G' has $\Omega(\frac{n^2}{n^{0.2}}) = \Omega(n^{1.8})$ edges, so we added $\Omega(n^{1.8})$ edges. ■

Corollary 3.8 *For every subfamily $\mathcal{F}$ of perfect graphs, $\Phi_{\mathcal{F}}(\Theta(n^{1.4}), n) = \Omega(n^{1.8})$.*

4 Perfect Deletion

In this chapter we study the random graph $G(n, \frac{1}{2})$. We shall show that with high probability G has $\Theta(n^2)$ edges and the minimum perfect deletion set also contains $\Theta(n^2)$ edges. Let m be the number of edges in G.

Claim 4.1 *For the random graph $G(n, \frac{1}{2})$, with high probability $m = \Theta(n^2)$.*

Proof: Define a random variable e_i for each edge, $e_i = 1$ if $i \in E$ and $e_i = 0$ otherwise. Then $m = \sum_{i=1}^{\binom{n}{2}} e_i$. Since the variables are independent, and $p(e_i = 1) = \frac{1}{2}$, by Theorem 2.1 (2) $Pr[|m - \frac{\binom{n}{2}}{2}| > n\sqrt{n}] < e^{-\frac{n^3}{\binom{n}{2}}} < e^{-n}$. Therefore with high probability $\frac{\binom{n}{2}}{2} - n^{1.5} < m < \frac{\binom{n}{2}}{2} + n^{1.5}$. ■

Claim 4.2 *Let X be the number of induced 5-cycles in G. With high probability $X = \Theta(n^5)$.*

Proof: We assign an indicator X_i for every set of 5 vertices. Let $X_i = 1$ if the graph induced by these five vertices is an induced 5-cycle, and $X_i = 0$ otherwise. It can be easily checked that $Pr[X_i = 1] = \frac{4!}{2*2^{10}} = \frac{3}{256}$.

$X = \sum_{i=1}^{\binom{n}{5}} X_i$, and by linearity of expectation we have $E[X] = \frac{3}{256}\binom{n}{5} = \Theta(n^5)$.

$$Var[X] \leq E[X] + \sum_{i \sim j} Pr[X_i \wedge X_j]$$

In our case $i \sim j$ if X_i and X_j are different and have at least 2 common vertices. There are $O(n^8)$ sets with two common vertices, $O(n^7)$ sets with three common vertices and $O(n^6)$ set with four common vertices. The probability $Pr[X_i \wedge X_j]$ is less than 1, therefore $Var[X] = O(n^8)$. Using Equation (3.1), since $E[X] = \Theta(n^5)$, with probability larger than $1 - \frac{1}{n}$, $|X - E[X]| < \sqrt{nVar(X)} = O(n^{4.5})$ *i.e.*, $X = \Theta(n^5)$.

■

Claim 4.3 *Any edge in E belongs to $O(n^3)$ induced 5-cycles.*

Proof: Let $e_i \in E$. Each 5-cycle containing e_i as an edge contains only 3 more vertices and there is a constant number of possible 5-cycles on each set of 5 vertices. ■

Theorem 4.4 *With high probability the random graph $G(n, \frac{1}{2})$ has $\Theta(n^2)$ edges and minimum perfect deletion set of size $\Theta(n^2)$.*

Proof: Let us look at a minimum perfect deletion set of the random graph $G(n, \frac{1}{2})$. We have to remove an edge from every induced 5-cycle. Since there are $\Theta(n^5)$ induced 5-cycles and each edge can participate in at most $O(n^3)$ induced 5-cycles, we have to remove $\Omega(\frac{n^5}{n^3}) = \Omega(n^2)$ edges. ■

Corollary 4.5 *For every subfamily $\mathcal{F}$ of perfect graphs, $\Psi_{\mathcal{F}}(\Theta(n^2), n) = \Omega(n^2)$.*

5 Concluding Remarks

In the *$\mathcal{F}$ -editing problem* one seeks a set of non-edges F and a set of edges K such that $G(V, E \cup F \setminus K) \in \mathcal{F}$ and $|F \cup K|$ is minimum. As a side remark, we show that there may be substantial gaps between the sizes of minimum deletion , completion and editing in a graph. Let $G_1 = (V_1, E_1)$ where $V_1 = (v_1, v_2, w_1, ..., w_n)$ and $E_1 = \{(v_i, w_j)|i = 1, 2, j = 1, ..., n\}$. G_1 has a chordal completion set of size 1 (the edge (v_1, v_2)), but every chordal deletion set contains at least $n - 1$ edges: Otherwise there are at least two vertices w_i, w_j such that v_1, w_i, v_2, w_j induce a C_4. Let $G_2 = (V_2, E_2)$ where $V_2 = (v_1, ..., v_n, w_1, ..., w_n)$ and $E_2 = \{(v_i, v_j), (w_i, w_j)|1 \le i, j \le n, i \ne j\} \cup \{(v_i, w_j)|1 \le i, j \le n - 1\} \cup \{(v_n, w_n)\}$. G_2 has a chordal deletion set of size 1 (the edge (v_n, w_n)). In contrast and every chordal completion set of G_2 contains at least $n - 1$ edges, since otherwise there exist $i, j \ne n$ such that v_n, v_i, w_j, w_n induce a C_4. Finally, let $G = G_1 \cup G_2$. G has a minimum deletion and completion set of sizes at least n and an editing set of size 2.

A graph is called *chordal* (or *triangulated*) if every cycle of length four or more in it contains a chord. We will build a family of graphs $G = (V, E)$ with $\Theta(n^{1.5})$ edges that have a minimum chordal completion of size $\Theta(n^2)$. Though the gap is smaller than that proved in [4], we include the example here since it is constructive and elementary.

A bipartite graph $G = (P, Q, E)$ is called a *chain graph* if the vertices of one of the parts, say P, can be ordered v_1, v_2,..,v_k so that $Adj(v_1) \supseteq Adj(v_2) \supseteq ... \supseteq Adj(v_k)$. Let $G = (V, W, E)$ be the following bipartite graph: $V = (v_1, v_2, \ldots, v_n)$, $W = (w_1, w_2, \ldots, w_n)$ and $E = \{e_i = (v_i, w_i)|1 \le i \le n\}$. In words, G is a perfect matching on a $2n$-vertex graph, or nK_2. Using the nested neighborhood property of chain graphs, it is easy to show that the smallest chain graph containing G has $\Theta(n^2)$ edges.

For a bipartite graph $G = (P, Q, E)$ we define a new graph $C(G) = (N, E')$, where $N = P \cup Q$, $E' = E \cup \{(u, v)|v \in P\} \cup \{(u, v) \mid u, v \in Q\}$ (*i.e.,* we replace the independent set on each part by a clique.) Yannakakis [15] showed that G is a chain graph iff $C(G)$ is triangulated.

We are now ready to construct our concrete family of examples. Let $G = (V, E)$ be the following graph: $V = \cup_{i=1}^{n} W_i$ where

$$W_i = \{w_i^1, w_i^2, \ldots w_i^n\}$$

and

$$E = \{(w_i^j, w_i^k)|1 \le i, j, k \le n\} \cup \{(w_j^i, w_k^i)|1 \le i, j, k \le n\},$$

In particular, for each i, j the subgraph induced on $W_i \cup W_j$ is isomorphic to $C(nK_2)$. G has n^2 vertices and

$$|E| = n^3 + n\binom{n}{2} = \Theta(n^3).$$

Using the observations above, one can show that a smallest chordal completion set of G contains $\Theta(n^4)$ edges.

Acknowledgments

We thank Haim Kaplan for many useful remarks.

References

[1] A. Agrawal, P. Klein, and R. Ravi. Cutting down on fill using nested dissection: provably good elimination orderings. In A. George, J. R. Gilbert, and J. W. H. Liu, editors, *Graph Theory and Sparse Matrix Computation*, pages 31–55. Springer, 1993.

[2] N. Alon, J. Spencer, and P. Erdos. *The Probablistic Method.* Wiley, New York, 1991.

[3] H. Chernoff. A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[4] F. R. K. Chung and D. Mumford. Chordal completions of planar graphs. *Journal of Combinatorial Theory series B*, 62:96–106, 1994.

[5] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman and Co., San Francisco, 1979.

[6] P. W. Goldberg, M. C. Golumbic, H. Kaplan, and R. Shamir. Four strikes against physical mapping of DNA. *Journal of Computational Biology*, 2(1):139–152, 1995.

[7] M. C. Golumbic. *Algorithmic Graph Theory and Perfect Graphs.* Academic Press, New York, 1980.

[8] S. Louis Hakimi, Edward F. Schmeichel, and Neal E. Young. Orienting graphs to optimize reachability. *Information Processing Letters*, 63(5):229–235, September 1997.

[9] T. Kashiwabara and T. Fujisawa. An NP-complete problem on interval graphs. In *IEEE International Symposium on Circuits and Systems (12th)*, pages 82–83. Institute of Electrical and Electronics Engineers; Piscataway, N.J., 1979.

[10] L. Lovász. A characterization of perfect graphs. *J. Combin. Theory*, pages 95–98, 1972.

[11] A. Natanzon, R. Shamir, and R. Sharan. A polynomial approximation algorithm for the minimum fill-in problem. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC-98)*, pages 41–47, New York, May 23–26 1998. ACM Press.

[12] A. Natanzon, R. Shamir, and R. Sharan. Complexity classification of some edge modification problems. In *Proc. 25th Int. Workshop (WG '99), Graph-Theoretic Concepts in Computer Science.* Springer-Verlag, 1999.

[13] R. Ravi, A. Agrawal, and P. Klein. Ordering problems approximated: single processor scheduling and interval graph completion. In *Proc. ICALP 1991*, pages 751–762. Springer, 1991. LNCS 510.

[14] P. L. Tche'byshef. Des valeurs moyennes. *Journal de Mathematiques pures et appliquees*, pages 177–184, 1867.

[15] M. Yannakakis. Computing the minimum fill-in is NP-complete. *SIAM J. Alg. Disc. Meth.*, 2, 1981.

[16] M. Yannakakis. Edge deletion problems. *SIAM J. Computing*, 10(2):297–309, 1981.