Random subgraphs make identification affordable

Florent Foucaud[1], Guillem Perarnau[1] and Oriol Serra[1]

Abstract. An identifying code of a graph is a dominating set which uniquely determines all the vertices by their neighborhood within the code. Whereas graphs with large minimum degree have small domination number, this is not true for the identifying code number.

We show that every graph G with n vertices, maximum degree $\Delta = \omega(1)$ and minimum degree $\delta \geq c \log \Delta$, for some constant $c > 0$, contains a large spanning subgraph which admits an identifying code of size $O(n \log \Delta / \delta)$. The result is best possible both in terms of code size and in number of edges deleted. The proof is based on the study of random subgraphs of G using standard concentration tools and the local lemma.

1 Introduction

Consider any graph parameter that is not monotone with respect to graph inclusion. Given a graph G, a natural problem in this context is to study the minimum value of this parameter over all spanning subgraphs of G. In particular, how many edge deletions are sufficient in order to obtain from G a graph with optimal value of the parameter? Herein, we study this question with respect to the identifying code number of a graph, a well-studied non-monotone parameter.

An *identifying code* of a graph is a subset of vertices which is a dominating set C such that each vertex is uniquely determined by its neighborhood within C. More formally, each vertex x of $V(G) \setminus C$ has at least one neighbor in C (x is *dominated*) and for each pair u, v of vertices of G, $N[u] \cap C \neq N[v] \cap C$; u, v are *separated*. The minimum size of an identifying code in a graph G, denoted by $\gamma^{\text{ID}}(G)$, is the *identifying code number* of G. Identifying codes were introduced in [5], motivated by

[1] Universitat Politècnica de Catalunya, BarcelonaTech, C/ Jordi Girona 1-3, 08034 Barcelona, Spain. Email: florent.foucaud@ma4.upc.edu, guillem.perarnau@ma4.upc.edu, oserra@ma4.upc.edu

J. Nešetřil et al. (eds.), *The Seventh European Conference on Combinatorics, Graph Theory and Applications*

several applications. Generally speaking, if the graph models a facility or a computer network, identifying codes can be used to detect dangers in facilities [9] or failures in networks [5]. In this context, deleting edges to the underlying graph is particularly meaningful since it may represent the sealing of a door in a facility network, and the removal of a wire in a computer network.

Note that some graphs may not admit an identifying code, in particular when they have pairs of twin vertices (*i.e.* which have the same closed neighborhood). However any twin-free graph is easily seen to admit an identifying code (*e.g.* its vertex set). The identifying code number of a graph G on n vertices satisfies $\log_2(n+1) \leq \gamma^{\text{ID}}(G) \leq n$.

There are very dense graphs that have a huge identifying code number; sparse graphs, such as trees and planar graphs, also have a linear identifying code number [10]. On the other hand, one can also find sparse and dense graphs with identifying code number $O(\log n)$ [4,8].

It shall be observed from the previous facts that the identifying code number is not a monotone function with respect to the addition (or deletion) of edges. This motivates the following question:

Given any sufficiently dense graph, can we delete a small number of edges to get a spanning subgraph with a small identifying code? If the answer is positive, how many edges are sufficient (and necessary)?

In other words, we would like to study the minimum size of an identifying code among all spanning subgraphs of a given graph, and to determine the largest spanning subgraph with an asymptotically optimal identifying code.

Despite being dense, the random graph $G(n, p)$ (for $0 < p < 1$) has a logarithmic size identifying code, as with high probability,

$$\gamma^{\text{ID}}(G(n, p)) = (1 + o(1))\frac{2\log n}{\log(1/q)},$$

where $q = p^2 + (1-p)^2$ [4]. This suggests that in a dense graph, the lack of structure implies the existence of a small identifying code number. Hence, introducing some randomness to the structure of a dense graph having large identifying code number might decrease this number. Indeed, this intuition is used in this work.

By studying the behavior of a *random subgraph of a graph with large minimum degree* (see for example [1,6]), we prove the following:

Theorem 1.1. *For any graph G on n vertices (n large enough) with maximum degree $\Delta = \omega(1)$ and minimum degree $\delta \geq 66\log\Delta$, there exists* *a subset of edges $F \subset E(G)$ of size*

$$|F| = O(n \log \Delta) ,$$

such that

$$\gamma^{ID}(G \setminus F) = O\left(\frac{n \log \Delta}{\delta}\right) .$$

In particular, when the minimum degree is linear, $\delta = \Theta(n)$, this shows that it is enough to delete $O(n \log n)$ edges to get a logarithmic size identifying code. The next theorem shows that Theorem 1.1 cannot be improved much.

Theorem 1.2. *For any $d \geq 2$, there exists a d–regular graph G_n^d on n vertices with the following properties.*

1. *For any $M \geq 0$, there exists a constant $c > 0$ such that for any set of edges $F \subset E(G_n^d)$ satisfying $\gamma^{ID}(G_n \setminus F) \leq M\frac{n \log d}{d}$, $|F| \geq cn \log d$.*
2. *For any spanning subgraph H of G_n^d, $\gamma^{ID}(H) = \Omega\left(\frac{n \log d}{d}\right)$.*

When $\delta = \mathrm{Poly}(\Delta)$, Theorem 1.2 shows that Theorem 1.1 is tight, that is, we cannot hope for having a smaller identifying code by deleting any set of edges. Moreover, if Δ is bounded or $\delta \leq c' \log \Delta$, for some small constant $c' > 0$, there is no way to improve the size of the identifying code of G by deleting edges.

2 Methods and proofs

The complete proofs can be found in [3].

The proof of Theorem 1.1 focuses on the study of the random spanning subgraph $G(B, f)$ of G, where $B \subseteq V$ and $f : V(G) \to \mathbb{R}^+ \cup \{0\}$ is a function. Edges non incident to B are always present in $G(B, f)$, while each incident edge uv appears in $G(B, f)$ independently with probability $1 - p_{uv}$, where p_{uv} depends on $f(u)$, $f(v)$ and the degree of u and v in B, $d_B(u)$ and $d_B(v)$ respectively. The next lemma gives an exponential upper-bound on the probability that two vertices of $G(B, f)$ are not separated by B.

Lemma 2.1. *Given a graph G and a subset $B \subseteq V(G)$, consider the random subgraph $G(B, f)$. For every pair u, v of distinct vertices with $d_B(u) \geq d_B(v)$,*

$$\Pr\left(N_{G(B,f)}[u] \cap B = N_{G(B,f)}[v] \cap B\right) < e^{-3f(u)/16} .$$

The proof of Theorem 1.1 is structured in the following steps:

1. Select a set $C \subseteq V$ at random, where each vertex is selected independently with probability p. Using the Chernoff inequality, estimate the probability of the event A_C that C is small enough for our purposes. From C, construct the spanning subgraph $G(C, f)$ of G, with $f(u) = \min(66 \log \Delta, d_B(u))$.
2. Use Lovász Local Lemma and Lemma 2.1 to lower-bound the probability that the following events (whose disjunction we call A_{LL}) hold jointly: 1. in $G(C, f)$, each pair of vertices that are at distance at most 2 from each other are separated by C; 2. for each such pair and each neighbor of this pair in G, its degree within C in G is close to its expected value $d(v)p$. Show that with nonzero probability, A_C and A_{LL} hold jointly.
3. Find a dominating set D with $|D| = O(|C|)$; if A_{LL} holds, $C \cup D$ is an identifying code.
4. Show that, if A_C and A_{LL} hold, the expected number of deleted edges is small.

To prove Theorem 1.2, we first study the complete graph on n vertices. We combine the following two lemmata to get as a direct corollary Proposition 2.4.

Lemma 2.2. *For any $M \geq 0$, there exists a constant $c_0 > 0$ such that any graph G with $\gamma^{ID}(G) \leq M \log n$ contains at least $c_0 n \log n$ many edges.*

Lemma 2.3. *Let G be a graph and $\overline{G}$ its complement. If G and $\overline{G}$ are twin-free,*

$$\frac{1}{2} \leq \frac{\gamma^{ID}(\overline{G})}{\gamma^{ID}(G)} \leq 2 .$$

Proposition 2.4. *Let K_n be the complete graph on n vertices. For any $M \geq 0$, there exists a constant $c > 0$ such that for any set of edges $F \subset E(K_n)$ satisfying $\gamma^{ID}(K_n \setminus F) \leq M \log n$, $|F| \geq cn \log n$.*

Now, consider the graph G_n^d to be the disjoint union of cliques of order $d + 1$. Since each clique is a connected component, an asymptotically optimal identifying code for G_n^d must be also asymptotically optimal for each component. By Proposition 2.4, we must delete at least $\Omega(d \log d)$ edges from each clique to get an identifying code of size $O(\log d)$ in each component. Thus one must delete at least $\Omega(n \log d)$ edges from G_n^d to get an identifying code of size $O\left(\frac{n \log d}{d}\right)$, thus, proving Theorem 1.2.

3 Concluding remarks

1. In [2], the notion of a *watching system* has been introduced as a relaxation of identifying codes: in a watching system, code vertices ("watchers") are allowed to identify any subset of their closed neighborhood, and several watchers can be placed in one vertex. Hence, for any spanning subgraph G' of G and denoting by $w(G)$ the minimum size of a watching system of G, we have $w(G) \leq w(G') \leq \gamma^{\mathrm{ID}}(G')$. In particular, the watching number is a monotone parameter with respect to graph inclusion. From Theorem 1.1 we have:

Corollary 3.1. *Under the hypothesis of Theorem* 1.1*,*

$$w(G) = O\left(\frac{n \log \Delta}{\delta}\right).$$

2. Using Lemma 2.3, Theorem 1.1 can be adapted to the case where we want to *add* edges rather than deleting them.

3. Given a graph property $\mathcal{P}$, the *resilience* of G with respect to $\mathcal{P}$ is the minimum number of edges one has to delete to obtain a graph not satisfying $\mathcal{P}$. The resilience of monotone properties is well studied, in particular, in the context of random graphs [11]. Our result can be understood in terms of the resilience of the property $\mathcal{P}$, G does not admit an small identifying code. For any graph G satisfying the hypothesis of Theorem 1.1 and Theorem 1.2, the resilience with respect to $\mathcal{P}$ is at most $O(n \log \Delta)$, and this upper bound is attained.

4. The proof of Theorem 1.1 just provides an exponentially small lower bound on the probability that we can find the desired object. However, if we assume that $\Delta = n$, this probability can be shown to be $1 - o(1)$. In such a case our proof provides a randomized algorithm which constructs a good spanning subgraph and a small identifying code meeting the bounds on Theorem 1.1.

References

[1] N. ALON, *A note on network reliability*, In: "Discrete probability and algorithms", IMA Vol. Math. Appl. **72**, Springer, New York, 1995, 11–14.

[2] D. AUGER, I. CHARON, O. HUDRY and A. LOBSTEIN, *Watching systems in graphs: an extension of identifying codes*, Discrete Applied Mathematics, to appear.

[3] F. FOUCAUD, G. PERARNAU and O. SERRA, *Random subgraphs make identification affordable*, available in ArXiv e-prints, 2013.

[4] A. Frieze, R. Martin, J. Moncel, M. Ruszinkó and C. Smyth, *Codes identifying sets of vertices in random networks*, Discrete Mathematics **307** (9-10) (2007), 1094–1107.

[5] M. G. Karpovsky, K. Chakrabarty and L. B. Levitin, *On a new class of codes for identifying vertices in graphs*, IEEE Transactions on Information Theory **44** (1998), 599-611.

[6] M. Krivelevich, C. Lee and B. Sudakov, *Long paths and cycles in random subgraphs of graphs with large minimum degree*, ArXiv e-prints, 2012.

[7] A. Lobstein, *Watching systems, identifying, locating-dominating and discriminating codes in graphs: a bibliography*, http://www.infres.enst.fr/ lobstein/debutBIBidetlocdom.pdf

[8] J. Moncel, *On graphs on n vertices having an identifying code of cardinality* $\log_2(n+1)$, Discrete Applied Mathematics **154** (14) (2006), 2032–2039.

[9] S. Ray, R. Ungrangsi, F. De Pellegrini, A. Trachtenberg and D. Starobinski, *Robust location detection in emergency sensor networks*, IEEE Journal on Selected Areas in Communications **22** (6) (2004), 1016–1025.

[10] P. J. Slater and D. F. Rall, *On location-domination numbers for certain classes of graphs*, Congressus Numerantium **45** (1984), 97–106.

[11] B. Sudakov and V. H. Vu, *Local resilience of graphs*, Random Structures Algorithms **33** (4) (2008), 409–433.