CONJURA · OPEN PROBLEM

Short Certificates That a Random Bipartite Graph Has No Small Non-Expanding Set

Constant right degree, sets of logarithmic size

Status. Statement: AI-written from Boaz Barak, *The Complexity of Public-Key Cryptography*, Cryptology ePrint Archive 2017, not yet checked by a human. Nothing here is settled: the survey poses the existence of such a certificate as a question and takes no position on the answer, reporting only that no algorithm is known for the unbalanced expansion problem in this parameter range, and that a worst-case **NP**-hardness of approximation result at matching parameters would be evidence for a negative answer — while saying it does not know whether such a result is likely to hold.

Category. *Public-Key Cryptography & Assumptions.*

Conjecture (informal). *Take a bipartite graph in which every vertex on the right has exactly d neighbours on the left, with d a fixed constant, and the graph is otherwise random. Call a set of right vertices non-expanding if its neighbourhood on the left is smaller than the set itself. In a random graph of this kind, small non-expanding sets do not occur; the survey's main candidate for a public-key encryption scheme resting on a combinatorial problem plants one of logarithmic size and keeps it as the private key, and its security requires that a planted graph cannot be told apart from a random one. The question is whether the absence of such a set can be proved rather than merely believed: is there a polynomial-size string that a polynomial-time checker can read alongside the graph and be convinced by, which exists for almost every random graph and can never exist for a graph that does have a small non-expanding set? Such a certificate would be a nondeterministic distinguisher, and it would show that this supposedly combinatorial assumption carries the same kind of structure as the lattice assumptions it was meant to be an alternative to. The natural route is that a non-expanding set gives a short linear dependency among the neighbourhood vectors of the graph over the two-element field, so the absence of any short dependency would already certify that no small non-expanding set exists — and certifying that a random sparse binary matrix has no short dependency among its columns is something nobody knows how to do.*

1 The setting

The Applebaum–Barak–Wigderson cryptosystem [ABW10] is the survey's main example of an attempt to base public-key encryption on the average-case difficulty of a combinatorial problem, a classification the survey itself flags as not well defined. Its public key is an (n, m, d) graph with constant right degree d in which a set S of $k = O(\log n)$ right vertices has been planted so that S has only $k - 1$ left-neighbours; the private key is S. Encryption applies Goldreich's local map [Gol11, App15] to a random input, and decryption works precisely because the k output bits indexed by S depend on only $k - 1$ input bits, so they take at most 2^{k-1} values. Security requires, among other things, that a planted graph cannot be distinguished from a uniformly random one — the un-

balanced expansion problem. Expansion problems in graphs have been studied for decades [HLW06], and no algorithm is known in this parameter range.

A word on the orientation of the planted set, because the source is internally inconsistent about it. Figure 6 on page 19 plants "a set S of right vertices of size $k = O(\log n)$ such that $|\Gamma_H(S)| = k - 1$ where $\Gamma_H(S)$ denotes the set of left-neighbors of S in H", and the discussion on page 20 — which writes one linear equation per right vertex and says the equations indexed by S are linearly dependent — agrees with it. The prose bullet on page 19 states the orientation the other way round, planting a set of *left* vertices of size k with at most $k - 1$ neighbours on the right. Everything below follows Figure 6 and page 20, since that is the orientation under which the scheme in Figure 6 decrypts and under which the linear-dependency remark is correct.

Whether this assumption is really unlike the lattice and coding assumptions is what the survey is worried about. All known lattice-based public-key schemes can be broken given an oracle for a problem in **NP** ∩ **coNP** [AR05]; that is a form of computational structure which generic one-way function candidates do not have, and the survey treats it as the mark of an assumption that has said too much about itself. The analogue here is whether the "no" side of the expansion problem has short proofs. A certificate as below is a nondeterministic distinguisher: it exists for almost every random graph and can never exist for a planted graph, since a planted graph is not k-good. Its existence would therefore place the unbalanced expansion assumption in the same structured class as the lattice assumptions, and would undercut the claim that the [ABW10] scheme is a genuinely different foundation for public-key encryption.

There is one obvious handle, and it is also the warning. As the survey observes on page 20, identify each right vertex j with the vector in $\mathbb{F}_2^n$ indicating $\Gamma_H(j)$; a non-expanding set S then yields $|S|$ vectors supported inside a coordinate subspace of dimension $|S| - 1$, hence a linear dependency among at most k of the m vectors. The implication runs one way only. Absence of any dependency among at most k columns *implies* k-goodness, so a certificate that no $\leq k$ columns are dependent is a sound — though possibly incomplete — certificate of k-goodness; the converse fails, since k vectors summing to zero over $\mathbb{F}_2$ can have a support union of size up to dk. The

survey's phrase, that "this problem can be thought of as the task of looking for a short linear dependency", is offered as evidence that the problem may be algebraic rather than combinatorial, and not as a certification route. Reading the dependency question as a lower bound on the minimum distance of a random low-density code is this page's own gloss and not the survey's: pages 19–21 mention neither codes nor minimum distance. The certificates that Feige, Kim and Ofek [FKO06] constructed for the non-satisfiability of random 3CNF formulas — the same kind of object, and the ones the survey invokes when it says that the very-low-noise variant of the [ABW10] pseudorandomness assumption becomes structured — live at high constraint density, whereas the graphs here are sparse and the sets in question have only logarithmically many vertices.

The parameter range matters twice over. The question is non-vacuous only where a random graph is itself k-good with high probability: the survey notes that as k grows, at some point depending on m/n a non-expanding set appears in a random graph with high probability and the distinguishing assumption becomes unconditionally true. In that regime the certification question below is vacuous, because the property to be certified no longer holds for a typical graph. Below that point, brute force settles k-goodness in roughly n^k time, quasi-polynomial for $k = O(\log n)$, which is also the ceiling on the security the scheme can offer; a certificate would have to be findable in genuinely polynomial time to be an attack, but its mere existence is what decides the structural question. The survey's own suggestion for evidence against existence is a worst-case **NP**-hardness of approximation result for the unbalanced expansion problem at these parameters, and it says explicitly that it does not know whether such a result should be expected. Note also that no unconditional refutation is in reach: proving that no polynomial-size certificate exists for an average-case property of this kind would be a separation far beyond current techniques, so the affirmative direction stated below is the attackable one.

Progress to date. The survey gives context and no resolution. Expansion problems have a long literature but no algorithm is known at these parameters [HLW06]. The one observation the survey contributes is the linear-dependency reformulation of page 20 quoted above, which is the natural route to a certificate; the minimum-distance reading of that route is this page's gloss, and in any case

the question asks only whether a certificate *exists*, not whether one can be computed efficiently. Finally, the survey records that in the very-low-noise regime of the same cryptosystem the companion pseudorandomness assumption does become structured, because it admits a non-constructive short certificate of the Feige–Kim–Ofek type [FKO06] — evidence that certificates of this general shape are sometimes available and worth hunting for.

2 Notation and parameters

Write $[n] = \{1, \dots, n\}$. An (n, m, d) *graph* is a bipartite graph H with left vertex set $[n]$, right vertex set $[m]$, in which every right vertex has exactly d neighbours on the left. For $j \in [m]$ write $\Gamma_H(j) \subseteq [n]$ for the set of left-neighbours of j, and for $S \subseteq [m]$ write

$$\Gamma_H(S) \;:=\; \bigcup_{j \in S} \Gamma_H(j).$$

Let $\mathcal{G}_{n,m,d}$ denote the distribution on (n, m, d) graphs in which the sets $\Gamma_H(1), \dots, \Gamma_H(m)$ are drawn independently and uniformly from the d-element subsets of $[n]$.

All asymptotics are as $n \to \infty$, with $m = m(n)$ and $k = k(n)$ functions of n and d a constant.

- n: number of left vertices of the bipartite graph, and the security parameter.
- m: number of right vertices. The survey fixes no bound; that m exceeds n follows from the [ABW10] pseudorandomness assumption rather than from any stated constraint.
- d: right degree; every right vertex has exactly d left-neighbours. A constant, and the survey says no more about it.
- k: size bound on the non-expanding sets to be excluded; $k = O(\log n)$, the range used by the cryptosystem this comes from.

3 The conjecture

Definition 1 (Small non-expanding sets). Let H be an (n,m,d) graph and $k \geq 1$. A set $S \subseteq [m]$ is *non-expanding* if $|\Gamma_H(S)| \leq |S| - 1$. The graph H is *k-good* if it has no non-expanding set of size at most k, that is, if

$$|\Gamma_H(S)| \;\geq\; |S| \qquad \text{for every } S \subseteq [m] \text{ with } 1 \leq |S| \leq k.$$

In particular a k-good graph has no set of exactly k right vertices with only $k-1$ left-neighbours, which is the structure planted in the private key of the cryptosystem described in Section 1.

Definition 2 (Nondeterministic certification of k-goodness). A *certification scheme* for k-goodness at parameters (n,m,d,k) consists of a polynomial p and a deterministic polynomial-time algorithm V that takes an (n,m,d) graph H and a string $w \in \{0,1\}^{p(n+m)}$ and outputs a bit. It is *sound* if for every (n,m,d) graph H and every w,

$$V(H,w) = 1 \implies H \text{ is } k\text{-good},$$

and *complete on average* if

$$\Pr_{H \leftarrow \mathcal{G}_{n,m,d}} \left[\exists\, w \in \{0,1\}^{p(n+m)} : V(H,w) = 1 \right] \;=\; 1 - o(1).$$

Soundness is required for all graphs; completeness only on average over $\mathcal{G}_{n,m,d}$.

Conjecture 1 (short certificates for the absence of small non-expanding sets). *Let d be a constant, and let $m = m(n)$ and $k = k(n) = O(\log n)$ be the parameters used by the [ABW10] cryptosystem, for which the property to be certified holds almost always, that is,*

$$\Pr_{H \leftarrow \mathcal{G}_{n,m,d}} \left[H \text{ is } k\text{-good} \right] \;=\; 1 - o(1).$$

Then for every such (d,m,k) there exists a certification scheme for

k-goodness at parameters (n, m, d, k) that is both sound and complete on average.
Refuting the statement means exhibiting constants and functions (d, m, k) in the above range for which no sound certification scheme is complete on average.

Remark 1 (readings fixed by this write-up). Four choices above are the page author's rather than the survey's. First, the direction: the survey poses the existence of a certificate as a question and takes no position, and the affirmative direction is stated here because an unconditional negative answer is not attackable. Second, the random graph model: the survey says only "a random (n, m, d) graph" and never fixes the distribution, and the independent-uniform-d-subsets model of Section 2 is the natural reading but not the only one. Third, the $1 - o(1)$ completeness threshold and the requirement that soundness hold in the worst case are a formalisation of the survey's phrase "a nondeterministic procedure to certify". Fourth, the survey asks the question at "parameters matching those used by [ABW10]" without restating them; Conjecture 1 therefore quantifies over that range as the survey names it, and imposes no bound of its own on d or on m.

4 Bibliography

[ABW10] B. Applebaum, B. Barak, A. Wigderson. *Public-key cryptography from different assumptions.* Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, pages 171–180, 2010.

[App15] B. Applebaum. *The cryptographic hardness of random local functions — survey.* IACR Cryptology ePrint Archive, 2015:165, 2015.

[AR05] D. Aharonov, O. Regev. *Lattice problems in NP ∩ coNP.* J. ACM, 52(5):749–765, 2005; preliminary version in FOCS '04.

[FKO06] U. Feige, J. H. Kim, E. Ofek. *Witnesses for non-satisfiability of dense random 3CNF formulas.* 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), pages 497–508, 2006.

[Gol11] O. Goldreich. *Candidate one-way functions based on expander graphs.* Studies in Complexity and Cryptography, pages 76–87, 2011; original version ECCC TR00-090, 2000.

[HLW06] S. Hoory, N. Linial, A. Wigderson. *Expander graphs and their applications.* Bulletin of the American Mathematical Society, 43(4):439–561, 2006.