Codes Identifying Sets of Vertices in Random Networks*

Alan Frieze†
Carnegie Mellon University
alan@random.math.cmu.edu

Ryan Martin‡
Iowa State University
rymartin@iastate.edu

Julien Moncel§
ERTé "maths à modeler"
julien.moncel@imag.fr

Miklós Ruszinkó¶
Computer and Automation Institute
Hungarian Academy of Sciences
ruszinko@lutra.sztaki.hu

Cliff Smyth‖
Massachusetts Institute of Technology
csmyth@math.mit.edu

July 19, 2006

Abstract

In this paper we deal with codes identifying sets of vertices in random networks; that is, $(1, \leq \ell)$-identifying codes. These codes enable us to detect sets of faulty processors in a multiprocessor system, assuming that the maximum number of faulty processors is bounded by a fixed constant ℓ. The $(1, \leq 1)$-identifying codes are of special interest. For random graphs we use the model $\mathcal{G}(n,p)$, in which each one of the $\binom{n}{2}$ possible edges exists with probability p. We give upper and lower bounds on the minimum cardinality of a $(1, \leq \ell)$-identifying code in a random graph, as well as threshold functions for the property of admitting such a code. We

*2000 Mathematics Subject Classification 94C12, 05C80, 94B60

†Research partially supported by NSF grant CCR-0200945

‡Corresponding author. Research partially supported by NSA grant H98230-05-1-0257

§Research partially supported by EURODOC grant number 03 00553 01. Part of this research was done while the author was at Université Joseph Fourier

¶Research supported in part by OTKA Grants T038198, T046234. This work relates to Department of the Navy Grant N00014-04-1-4034 issued by the Office of Naval Research International Field Office. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein.

‖Research partially supported by NSF VIGRE grant DMS-9819950. Part of this research was done while the author was at Carnegie Mellon University.

derive existence results from probabilistic constructions. A connection between identifying codes and superimposed codes is also established.

1 Codes identifying sets of vertices

1.1 Motivation

Identifying codes were defined in [10] to model fault diagnosis in multiprocessor systems. In these systems, it may happen that some of the processors become faulty, in some sense that depends on the purpose of the system. We wish to detect and replace such processors, so that the system can work properly. We assume that our hardware is of such a quality that, at any time, at most ℓ of the processors of the system are faulty, where ℓ is a fixed constant. Let us assume that each processor `p` of the system is able to run a procedure `test(p)`, which checks its own state as well as the state of its neighboring processors `N(p)`. This procedure returns only binary information; *e.g.*, `0` if `p` or a processor of its neighborhood `N(p)` is faulty, and `1` otherwise. This information is returned to a central controller, which is considered not to be part of the system. Note that the procedure doesn't reveal the identity of the faulty processor: If `test(p)` outputs `0`, then all we can say is that `p` and/or some of its surrounding processors in `N(p)` is faulty. We wish to devise a subset of processors `C` such that:

(i) If all the processors of `C` return `1` then none of the processors of the network is faulty,

(ii) If at least one, but at most ℓ of the processors are malfunctioning, then the central controller is able to locate them using `C`.

1.2 Formal definition of identifying codes

We model our multiprocessor system by a simple, undirected graph $G = (V, E)$, whose vertices are processors and whose edges are links between these processors. For a vertex $v \in V$, let us denote $N[v]$ to be the closed neighborhood of v: $N[v] = N(v) \cup \{v\}$. Let $C \subseteq V$ be a subset of vertices of G, and for all subsets of at most ℓ vertices $X \subseteq V$, let us denote

$$I(X, C) := \bigcup_{x \in X} N[x] \cap C.$$

If all the $I(X, C)$'s are distinct then we say that C *separates* the sets of at most ℓ vertices of G. Since, $I(\emptyset, C) = \emptyset$, so for all nonempty X of size at most ℓ, $I(X, C)$ is nonempty so we say that C *covers* the sets of at most ℓ vertices of G. We say that C is a code identifying sets of at most ℓ vertices of G if and only if C covers and separates the sets of at most ℓ vertices of G. The dedicated terminology [12, 13] for such codes is $(1, \leq \ell)$-*identifying codes*. Here let ℓ-*ID code* state for $(1, \leq \ell)$-identifying code, and simply *ID code* state for $(1, \leq 1)$-identifying code.

The sets $I(X, C)$ are called *identifying sets* of the corresponding X's. The most investigated case is the one with $\ell = 1$: In this case, C is an ID code if and only if for all vertices v in G, the shadows $N[v] \cap C$ are all different and non-empty.

Clearly, the set of vertices C corresponding to a set of processors $\mathtt{C}$ is an ℓ-ID code of G if and only if $\mathtt{C}$ satisfies both conditions (i) and (ii) of Section 1.1.

Whereas $C = V$ is trivially always a covering code, not every graph has an ℓ-ID code. For example, if G contains two vertices u and v such that $N[u] = N[v]$, then G can have no ID code, since for any subset of vertices C we have $N[u] \cap C = N[v] \cap C$. In fact, a graph admits an ℓ-ID code if and only if for every pair of subsets $X \neq Y$, $|X|, |Y| \leq \ell$, we have $N[X] \neq N[Y]$, where $N[X]$ denotes $\bigcup_{x \in X} N[x]$. In the case where G admits an ℓ-ID code, then $C = V$ is always a ℓ-ID code of G, hence we are usually interested in finding an ℓ-ID code of minimum cardinality.

1.3 Networks having no known structure

Without making any assumption on the structure of the network, we would like to know how this problem behaves as the size of the network grows. If we just draw links independently at random between processors, what is the probability that the resulting network admits an ℓ-ID code? What is the asymptotic expected value of an ℓ-ID code in such a network?

To handle these kinds of questions, we investigate ℓ-ID codes in random graphs. We use the model $\mathcal{G}(n, p)$, in which each one of the $\binom{n}{2}$ possible edges exists independently with probability p, with p possibly being a function of n. We will use the standard notation $G_{n,p}$ to denote a labelled random graph of the probability space $\mathcal{G}(n, p)$. For a given graph G with n vertices and m edges, the probability that $G_{n,p} = G$ is $p^m(1-p)^{\binom{n}{2}-m}$.

We say that a property Π holds for almost every graph in $\mathcal{G}(n, p)$ (or Π holds with high probability) if and only if the probability

$$\Pr(G_{n,p} \text{ has the property } \Pi)$$

tends to 1 as n tends to infinity. Similarly, Π holds for almost no graph in $\mathcal{G}(n, p)$ if and only if $\Pr(G_{n,p}$ has the property $\Pi)$ tends to 0 as n tends to infinity. We refer the reader to [2], [9] for a complete introduction to random graphs.

In this paper $\log x$ denotes the logarithm in base e. The notation ω, o, O, Θ, $\sim$ are used in the conventional sense, *i.e.*, for sequences $f(n)$ and $g(n)$ we have $f(n) = \omega(g(n))$ if $f(n)/g(n) \to +\infty$, $f(n) = o(g(n))$ if $f(n)/g(n) \to 0$, $f(n) \sim g(n))$ if $g(n)(1 - o(1)) \leq f(n) \leq g(n)(1 + o(1))$, $f(n) = O(g(n))$ if $f(n) \leq cg(n)$ holds for every n with appropriate constant c, and $f(n) = \Theta(g(n))$ if $c_1 f(n) \leq g(n) \leq c_2 f(n)$ holds for every n with appropriate constants c_1 and c_2.

1.4 Outline of the paper

In Section 2.1, we deal with the cardinality of ID codes in random graphs and in Section 2.2, the threshold properties for such a code to exist. Section 3 deals with ℓ-ID codes ($\ell \geq 2$) in random graphs. In general, Section 3.2 reinforces a connection between ℓ-ID and ℓ-superimposed codes, first observed in [10], which is of independent interest. In Section 4 we present some questions arising from this work.

2 ID codes in random graphs

In the following theorem we determine the exact asymptotic behavior of the cardinality $c = c(G)$ of a minimum ID code in not too sparse and not too dense random graphs.

2.1 Minimum cardinality of an ID code

In this section let

$$q = p^2 + (1-p)^2.$$

Theorem 1 *Let $p, (1-p) \geq 4 \log\log n / \log n$. Then for almost every graph in $\mathcal{G}(n,p)$, we have $c(G_{n,p}) \sim \frac{2\log n}{\log(1/q)}$, i.e., for every fixed $\epsilon > 0$,*

$$\lim_{n\to\infty} \Pr\left(\left| c(G_{n,p}) \cdot \left(\frac{2\log n}{\log(1/q)} \right)^{-1} - 1 \right| \geq \epsilon \right) = 0.$$

To see the upper bound for c we need the following proposition.

Proposition 1 *Let C be a subset of vertices of cardinality c of $G_{n,p}$. The probability that C is not an ID code of $G_{n,p}$ is bounded by:*

$$\Pr(C \text{ is not an ID code}) \leq \binom{c}{2} pq^{c-2} + c(n-c)pq^{c-1} + \binom{n-c}{2} q^c.$$

Proof : Indeed, let C be a subset of vertices of cardinality c. For each pair of distinct vertices $x \neq y$, let us denote $A_{x,y}(C)$ the event $\{B(x) \cap C = B(y) \cap C\}$. The probability that C is not a separating code is

$$\Pr(\cup_{x\neq y} A_{x,y}(C)) \leq \sum_{x\neq y} \Pr(A_{x,y}(C))$$

If $x \in C$ and $y \in C$, then $\Pr(A_{x,y}(C)) = pq^{c-2}$; if $x \in C$ and $y \notin C$, or $x \notin C$ and $y \in C$, then $\Pr(A_{x,y}(C)) = pq^{c-1}$; and if $x \notin C$ and $y \notin C$, then $\Pr(A_{x,y}(C)) = q^c$. Thus,

$$\Pr(\cup_{x\neq y} A_{x,y}(C)) \quad \leq \quad \binom{c}{2} pq^{c-2} + c(n-c)pq^{c-1} + \binom{n-c}{2} q^c$$

□

The upper bound is now straightforward, i.e.,

Lemma 1 *Let ϵ have the property that $n^\epsilon \to +\infty$ (that is, $\epsilon = \omega\left((\log n)^{-1}\right)$), and p such that p and $1-p$ are $\omega\left((\log n)^{-1}\right)$. Then almost every graph in $\mathcal{G}(n,p)$ admits an ID code of cardinality less than or equal to*

$$\frac{(2+\epsilon)\log n}{\log(1/q)}.$$

Proof : By Proposition 1 we have :

$$\Pr(C \text{ is not an ID code}) \leq \binom{c}{2} pq^{c-2} + c(n-c)pq^{c-1} + \binom{n-c}{2} q^c$$

for every subset $C \subseteq V$ of cardinality $c = c(n)$. It is easy to see that if both p and $1-p$ are $\omega\left((\log n)^{-1}\right)$, then for $c = n$ this quantity tends to 0 (see Lemma 4), which proves that for such a p almost every graph in $\mathcal{G}(n,p)$ admits an ID code. Now let $c = \frac{(2+\epsilon)\log n}{\log(1/q)}$. Since both p and $1-p$ are $\omega\left((\log n)^{-1}\right)$, then we have $c = o(n)$. We can rewrite this probability,

$$\Pr\left(C \text{ is not an ID code}\right) \leq (n-c)^2 q^c \left[1 + \frac{2c}{n-c}\frac{p}{q} + \frac{c^2}{(n-c)^2}\frac{p}{q^2}\right].$$

Since the term in between brackets tends trivially to 1, it remains to show that $(n-c)^2 q^c$ tends to 0:

$$\begin{aligned}
(n-c)^2 q^c &= \exp\left\{2\log(n-c) + c\log q\right\} \\
&\leq \exp\left\{2\log(n-c) - (2+\epsilon)\log n\right\} \\
&\leq n^{-\epsilon} \\
&= o(1).
\end{aligned}$$

□

Clearly, in any graph the cardinality of an ID code is at least $\lceil \log_2(n+1) \rceil$ (easy to see – the identifying sets $I(x,C)$ are nonempty distinct subsets of 2^C). Therefore, the minimum cardinality of an ID code of a random graph is almost surely $\Theta(\log n)$. In order to determine that the exact value of the constant is 2; that is to say the upper bound of Theorem 1 is asymptotically tight, we will use Suen's inequality, first introduced in [15] and revised in [8]. It is also cited in [1]. This has a similar setup to that of the Lovász Local Lemma [4, 1] in that it uses a so-called dependency graph. Our definitions and notation will come from [8].

Let $\mathcal{I}$ be an index set of events. We consider events A_i for $i \in \mathcal{I}$ with indicator variable I_i. The indicator I_i has $\mathbb{E}[I_i] = p_i$ for $i \in \mathcal{I}$ and $X = \sum_{i \in \mathcal{I}} I_i$. There is a *dependency graph* with vertex set $\mathcal{I}$ and $i \sim j$ so that if there are

any subsets $J_1, J_2 \subset \mathcal{I}$ with no edge between any $i \in J_1$ and any $j \in J_2$, then any Boolean combination of $\{A_i : i \in J_1\}$ and any Boolean combination of $\{A_j : j \in J_2\}$ are independent of each other. Let the notation $k \sim \{i, j\}$ mean that vertex k is adjacent to either i or j or both.

- $\mu := \sum_{i \in \mathcal{I}} p_i$
- $\Delta := \sum_{\{\{i,j\}: i \sim j\}} \mathbb{E}(I_i I_j)$
- $\delta := \max_{i \in \mathcal{I}} \sum_{j \sim i} p_j$

We combine these results in one statement.

Theorem 2 (Suen) *With the above setup,*

$$\Pr(X = 0) \leq \exp\left\{-\mu + \Delta e^{2\delta}\right\}$$

Now we are ready to get the claimed lower bound.

Lemma 2 *Let p have the property that*

$$2p(1-p) \geq \theta = \frac{4 \log\log n}{\log n}$$

and let $\epsilon = \frac{3 \log\log n}{\log n}$. With high probability, there exists no ID code of cardinality less than

$$\frac{(2-\epsilon)\log n}{\log(1/q)}.$$

Proof : First, we fix a set C of cardinality $c := \left\lfloor \frac{(2-\epsilon)\log n}{\log(1/q)} \right\rfloor$. This implies that $n^{\epsilon-2} \leq q^c \leq n^{\epsilon-2}/q \leq 2n^{\epsilon-2}$. We use Suen's inequality to bound the probability that C is an ID code. In order to apply Theorem 2, we let $\mathcal{I}$ be the set of pairs of vertices in $V \setminus C$. The event A_i is the event that both vertices represented by i have the same intersection set in C. Thus $X \neq 0$ implies that C is not an ID code (but not vice-versa).

We put $i \sim j$ if and only if the corresponding pairs of vertices have a nonempty intersection. Therefore, $p_i = q^c$ for all $i \in \mathcal{I}$. Also, whenever $i \sim j$,

$$\mathbb{E}[I_i I_j] = \left(p^3 + (1-p)^3\right)^c.$$

We have $|\mathcal{I}| = \binom{n-c}{2}$, $|\{\{i,j\} : i \sim j\}| = 3\binom{n-c}{3}$ and $|\{j : j \sim i\}| = 2(n-c-2)$ for all $i \in \mathcal{I}$. This gives

- $\mu = \binom{n-c}{2} q^c$,
- $\Delta = 3\binom{n-c}{3}\left(p^3 + (1-p)^3\right)^c$, and
- $\delta = 2(n-c-2)q^c$.

$$
\begin{aligned}
&\Pr\left(C \text{ is an ID code}\right)\\
&\quad\leq\quad \exp\left\{-\binom{n-c}{2}q^c + 3\binom{n-c}{3}(p^3+(1-p)^3)^c e^{4nq^c}\right\}\\
&\quad\leq\quad \exp\left\{-\binom{n-c}{2}q^c\left(1-n\left(1-\frac{1-q}{2q}\right)^c e^{4nq^c}\right)\right\}\\
&\quad\leq\quad \exp\left\{-\frac{n^\epsilon}{5}\left(1-n\exp\left\{-\frac{1-q}{2q}\frac{(2-\epsilon)\log n}{\log 1/q}+O(n^{\epsilon-1})\right)\right\}\right\}
\end{aligned}
$$

Now the function $\frac{x-1}{x\log x}$ decreases in the range $[0,1]$ and so using Boole's inequality to bound the probability that any ID code exists,

$$
\begin{aligned}
&\Pr\left(\text{There exists an ID code of cardinality } c\right)\\
&\quad\leq\quad \binom{n}{c}\exp\left\{-\frac{n^\epsilon}{5}\left(1-n\exp\left\{\frac{\theta}{2(1-\theta)}\frac{(2-\epsilon)\log n}{\log(1-\theta)}+O(n^{\epsilon-1})\right)\right\}\right\}\\
&\quad\leq\quad \binom{n}{c}\exp\left\{-\frac{n^\epsilon}{5}\left(1-n^{-\epsilon/2+\theta/2+O(\theta^2)}\right)\right\}
\end{aligned}
$$

Now $\binom{n}{c} = e^{O((\log n)^3/\log\log n)}$ and $n^\epsilon = \Omega((\log n)^3)$ and so

$$\Pr\left(\text{There exists an ID code of cardinality } c\right) = o(1).$$

This finishes the proof of Theorem 1. □

For the most studied $p = 1/2$ random graph model the answer is asymptotically as follows:

Corollary 1 *$c(G_{n,1/2}) \sim 2\log_2 n$, with high probability.*

Note that any ID code in any n-vertex graph is of size at least $\lceil \log_2(n+1)\rceil$ (see [10]).

2.2 Threshold probabilities admitting ID codes

In order to deal with threshold functions, we need two fundamental results of Erdős and Rényi [5, 6], that we give here as stated in [1, Theorems 5.3 and 5.4]. These theorems describe very accurately the threshold functions for the number of connected components which are trees in a random graph.

Theorem 3 *Let us denote by X the random variable equal to the number of isolated vertices of $G_{n,p}$.*

(i) *If $pn - \log n \to -\infty$ then for every $L \in \mathbb{R}$ we have $\Pr(X \geq L) \to 1$.*

(ii) *If $pn - \log n \to x$ for some $x \in \mathbb{R}$ then X tends to the Poisson distribution with mean $\lambda := e^{-x}$, that is $\Pr(X = r)$ tends to $e^{-\lambda}\frac{\lambda^r}{r!}$ for all $r \geq 0$.*

(iii) *If* $pn - \log n \to +\infty$ *then* $X = 0$ *for almost every graph in* $\mathcal{G}(n,p)$.

Theorem 4 *For any* $k \geq 2$, *denote by* T_k *the random variable equal to the number of components of* $G_{n,p}$ *which are trees of order* k.

(i) *If* $p = o(n^{-\frac{k}{k-1}})$ *then* $T_k = 0$ *for almost every graph in* $\mathcal{G}(n,p)$.

(ii) *If* $n^{\frac{k}{k-1}}p \to z$ *for some constant* $z > 0$ *then* T_k *tends to the Poisson distribution with mean* $\lambda := z^{k-1}\frac{k^{k-2}}{k!}$, *that is* $\Pr(T_k = r)$ *tends to* $e^{-\lambda}\frac{\lambda^r}{r!}$ *for all* $r \geq 0$.

(iii) *If* $pn^{\frac{k}{k-1}} \to +\infty$ *and* $pkn - \log n - (k-1)\log\log n \to -\infty$ *then for every* $L \in \mathbb{R}$ *we have* $\Pr(T_k \geq L) \to 1$.

(iv) *If* $pkn - \log n - (k-1)\log\log n \to x$ *for some* $x \in \mathbb{R}$ *then* T_k *tends to the Poisson distribution with mean* $\frac{e^{-x}}{k \times k!}$.

(v) *If* $pkn - \log n - (k-1)\log\log n \to +\infty$ *then* $T_k = 0$ *for almost every graph in* $\mathcal{G}(n,p)$.

As a consequence of Proposition 1, if $p \neq 1$ is constant then almost every graph in $\mathcal{G}(n,p)$ has an ID code. But if we let p be a function of n this does not remain necessarily true. For instance, if $p = p(n)$ is too large, then $G_{n,p}$ contains almost surely two universal vertices (*i.e.* vertices adjacent to all other ones), which prevents $G_{n,p}$ from having an ID code. On the other hand, if p is so small that there are almost surely no edges, then $G_{n,p}$ has an ID code; but for a small-but-not-too-small p, there are almost surely isolated edges in $G_{n,p}$, which prevent it from having an ID code. We show that isolated edges and universal vertices are the only obstructions for having an ID code. To summarize, the situation is the following:

Theorem 5 *For any* $\epsilon > 0$ *we have:*

- *If* $p = o(n^{-2})$ *then almost every graph in* $\mathcal{G}(n,p)$ *has an ID code (almost surely, this is unique – the entire vertex set),*
- *if* $pn^2 \to +\infty$ *and* $p \leq \frac{1}{2n}(\log n + (1-\epsilon)\log\log n)$ *then almost no graph in* $\mathcal{G}(n,p)$ *has an ID code,*
- *if* $\frac{1}{2n}(\log n + (1+\epsilon)\log\log n) \leq p \leq 1 - \frac{1}{n}(\log n + \epsilon\log\log n)$ *then almost every graph in* $\mathcal{G}(n,p)$ *has an ID code,*
- *if* $p \geq 1 - \frac{1}{n}(\log n - \epsilon\log\log n)$ *then almost no graph in* $\mathcal{G}(n,p)$ *has an ID code.*

Note that the non-trivial interval of existence of a code is asymmetric, since its lower bound is asymptotically $\frac{\log n}{2n}$, while its upper bound is roughly $1 - \frac{\log n}{n}$. This comes from the fact that we need to separate all pairs of *adjacent* vertices. Indeed, provided they are covered, two non-adjacent vertices are automatically

separated. Intuitively, in a dense graph (*i.e.* when p tends to 1) any two vertices are adjacent with high probability, thus we need to separate a lot of pairs of vertices. But when p tends to 0, most of the vertices of $G_{n,p}$ are non-adjacent, thus only a few number of pairs of vertices have to be considered.

We split the proof of this theorem in the following four propositions.

Proposition 2 *If $p = o(n^{-2})$, then almost every graph in $\mathcal{G}(n,p)$ has an ID code.*

Proof : This follows from the fact that for such a p almost surely $G_{n,p}$ has no edge, hence has V as a unique ID code. □

Proposition 3 *For any $\epsilon > 0$, if $pn^2 \to +\infty$ and $p \leq \frac{1}{2n}(\log n + (1-\epsilon)\log\log n)$ then almost no graph in $\mathcal{G}(n,p)$ has an ID code.*

Proof : This follows from Theorem 4 (iii) applied with $k = 2$: For such a p almost every graph in $\mathcal{G}(n,p)$ has a connected component which is a tree on two vertices, *i.e.* an isolated edge. A graph having an isolated edge has no ID code. □

Proposition 4 *For any $\epsilon > 0$, if $p \geq \frac{1}{2n}(\log n + (1+\epsilon)\log\log n)$ and $p \leq 1 - \frac{1}{n}(\log n + \epsilon \log\log n)$ then almost every graph in $\mathcal{G}(n,p)$ has an ID code.*

Proof : The vertex set of $G_{n,p}$ is an ID code if and only if it is a separating code. From Proposition 1, the probability that the vertex set of $G_{n,p}$ is not a separating code is less or equal to $\binom{n}{2} p \left(p^2 + (1-p)^2\right)^{n-2} = f^n(p)$. The function $f^n : x \mapsto \binom{n}{2} x \left(x^2 + (1-x)^2\right)^{n-2}$ increases from 0 to some $\alpha_n = \Theta\left(n^{-\frac{1}{2}}\right)$, then decreases to some $\beta_n = \frac{1}{2} - \Theta\left(n^{-\frac{1}{2}}\right)$, and then increases again. Since $n^{-\frac{1}{2}}$ tends to 0 more slowly than $\frac{1}{2n}(\log n + (1+\epsilon)\log\log n)$, for n large, the maximum of $f^n(p)$ on the interval

$$\left[\frac{1}{2n}(\log n + (1+\epsilon)\log\log n), 1 - \frac{1}{n}(\log n + \epsilon\log\log n)\right]$$

is attained for $p = \frac{1}{2n}(\log n + (1+\epsilon)\log\log n)$ or $p = 1 - \frac{1}{n}(\log n + \epsilon\log\log n)$. It then suffices to check that

$$f^n\left(\frac{1}{2n}(\log n + (1+\epsilon)\log\log n)\right)$$

and

$$f^n\left(1 - \frac{1}{n}(\log n + \epsilon\log\log n)\right)$$

both tend to 0 as n tends to infinity, which is straightforward. □

Proposition 5 *For any $\epsilon > 0$, if $p \geq 1 - \frac{1}{n}(\log n - \epsilon \log\log n)$ then almost no graph in $\mathcal{G}(n,p)$ has an ID code.*

Proof : We use the fact that the number of universal vertices (*i.e* vertices which are neighbors of all other ones) in $\mathcal{G}(n,p)$ is equal to the number of isolated vertices in $\mathcal{G}_{n,1-p}$. From Theorem 3 (i), there exists almost surely at least two universal vertices in $\mathcal{G}(n,p)$ for such a p. A graph having two universal vertices has no ID code. □

The results of Theorem 5 can be represented as in Figure 1, where we tried to sketch the evolution of the limit of $\Pr(G_{n,p}$ has an ID code) as a function of $p(n)$. Note the (quite unusual) fact that there are two intervals of existence of an ID code with high probability.

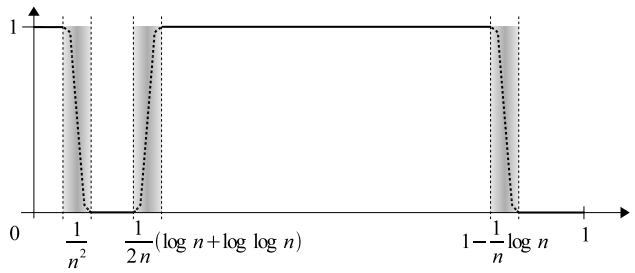


Figure 1: Graphical representation of the thresholds for the property of having an ID code. On the vertical axis is the asymptotic value of $\Pr(G_{n,p}$ has an ID code), while on the horizontal axis is $p(n)$.

Due to the precision of the results in Theorems 3 and 4, we are actually able to describe rather accurately what happens *at* the thresholds, *i.e.* when p is in one of the three shaded areas of Figure 1.

Theorem 6 *For any constant $z > 0$, if $n^2 p \to z$ then the probability that a graph in $\mathcal{G}(n,p)$ has an ID code tends to $e^{-\frac{z}{2}}$ as n tends to infinity.*

Proof : We know that $G_{n,p}$ has no ID code if and only if there exists a pair of distinct vertices $u \neq v$ such that $N[u] = N[v]$, but we can restrict ourselves to vertices $u \neq v$ such that $N[u] = N[v] = \{u, v\}$, that is to isolated edges. Indeed, the presence of $u \neq v$ such that $N[u] = N[v]$ with $|N[u]| \geq 3$ implies the presence of a triangle in $G_{n,p}$, and for such a p the probability that $G_{n,p}$ contains a triangle is bounded by $\binom{n}{3}p^3$, which tends to 0 as n tends to infinity. Hence, for large n,

$$\Pr\left(G_{n,p} \text{ has no ID code}\right) \sim \Pr\left(G_{n,p} \text{ has an isolated edge}\right)$$

Thanks to Theorem 4 (ii), we know that the number of isolated edges tends to the Poisson distribution with mean $\frac{z}{2}$. □

Theorem 7 *For any constant $x \in \mathbb{R}$, if $2np - (\log n + \log\log n)$ tends to x as n tends to infinity, then the probability that a graph of $\mathcal{G}(n,p)$ has an ID code tends to $e^{-e^{-\frac{x}{4}}}$ as n tends to infinity.*

Proof : As in the previous theorem, it is enough to look for isolated edges. Indeed, the presence of two vertices $u \neq v$ such that $N[u] = N[v]$ with $|N[u]| \geq 4$ implies the presence of a subgraph isomorphic to H_4 in $G_{n,p}$, where H_4 denotes the graph with five edges on four vertices. The expected number of subgraphs isomorphic to H_4 contained in $G_{n,p}$ is equal to $6\binom{n}{4}p^5$, which tends to 0 when n tends to infinity for such a p. Then, the probability that $G_{n,p}$ contains two vertices $u \neq v$ such that $N[u] = N[v]$ with $|N[u]| \geq 4$ tends to 0 as n tends to infinity. Now, let us look at the probability that $G_{n,p}$ contains two vertices $u \neq v$ such that $N[u] = N[v]$ and $|N[u]| = 3$. The expected number of such pairs of vertices is $3\binom{n}{3}p^3(1-p)^{2(n-3)}$, which tends to 0 when n tends to infinity. Hence $G_{n,p}$ almost surely does not contain two vertices $u \neq v$ such that $N[u] = N[v]$ and $|N[u]| = 3$. Hence, for large n,

$$\Pr\left(G_{n,p} \text{ has no ID code}\right) \sim \Pr\left(G_{n,p} \text{ has an isolated edge}\right)$$

We conclude using Theorem 4 (iv). □

Theorem 8 *For any constant $x \in \mathbb{R}$, if $n(1-p) - \log n$ tends to x as n tends to infinity, then the probability that a graph of $\mathcal{G}(n,p)$ has an ID code tends to $e^{-e^{-x}}(1+e^{-x})$ as n tends to infinity.*

Proof : We know that $G_{n,p}$ has no ID code if and only if there exists a pair of vertices $u \neq v$ such that $N[u] = N[v]$, but we can actually restrict ourselves to universal vertices, *i.e.* to the case where $|N[u]| = |N[v]| = n$. This is the case because the expected number of pairs of vertices $u \neq v$ such that $N[u] = N[v]$ and $|N[u]| \leq n-1$ is $\binom{n}{2}p\left((p^2+(1-p)^2)^{n-2} - p^{2(n-2)}\right)$, which tends to 0 as n tends to infinity for such a p. Hence, for large n,

$$\begin{aligned} &\Pr\left(G_{n,p} \text{ has no ID code}\right) \\ &\quad\sim\quad \Pr\left(G_{n,p} \text{ has at least two universal vertices}\right) \end{aligned}$$

We conclude using Theorem 3 (ii), using the fact that the number of universal vertices of $G_{n,p}$ is equal to the number of isolated vertices of $G_{n,1-p}$. □

3 ℓ-ID codes in random graphs

We call a pair of subsets (X,Y) of the vertex set with $1 \leq |X|, |Y| \leq \ell$ maximal, if

- either $|X| = \ell - 1$, $|Y| = \ell$, and $X \subseteq Y$
- or $|Y| = \ell - 1$, $|X| = \ell$, and $Y \subseteq X$,

- or $|X| = \ell$ and $|Y| = \ell$.

The following lemma shows that in order to get an ℓ-ID code we can restrict ourselves to separate maximal pairs of subsets (X, Y):

Lemma 3 *C is an ℓ-ID code of G if and only if $I(X, C) \neq \emptyset$ for all $X \subseteq V$ such that $|X| \leq \ell$, and the condition*

$$I(X, C) \neq I(Y, C)$$

is satisfied for all maximal pairs (X, Y).

Proof : If (X, Y) is maximal then we are done. If (X, Y) is not maximal, then let us assume that $|X| \leq |Y|$. Two cases follow.

(a) If $X \subseteq Y$, then let $Z \subseteq V \setminus Y$ be of cardinality $\ell - |Y|$ and let $y_0 \in Y \setminus X$. Now set $X' := Y \cup Z \setminus \{y_0\}$ and $Y' := Y \cup Z$. It is easy to see that if C doesn't separate X and Y, then C doesn't separate X' and Y' either. This would contradict the fact that (X', Y') is maximal.

(b) If $X \not\subseteq Y$, then let $Z \subseteq V \setminus Y$ of cardinality $\ell - |Y|$ and let $T \subseteq Y \setminus X$ such that $|X| + |T| + |Z| = \ell$. Now set $X' := X \cup T \cup Z$ and $Y' := Y \cup Z$. It is easy to see that if C doesn't separate X and Y, then C doesn't separate X' and Y' either. This would contradict the fact that (X', Y') is maximal.

□

3.1 Minimum cardinality of an ℓ-ID code

The following lemma is analogous to Proposition 1.

Lemma 4 *Let $C \neq V$ be a subset of vertices of a random graph $G_{n,p}$. Then the probability that C is not an ℓ-ID code is bounded by:*

$$\Pr(C \textit{ is not a code}) \leq n^{2\ell} \left(1 - \min\{p, 2p(1-p)\}(1-p)^{\ell-1}\right)^{|C|-2\ell}$$

In the case where $C = V$, we have:

$$\Pr(V \textit{ is not a code}) \leq n^{2\ell} \left(1 - (1-p)^{\ell}\right) \left(1 - \min\{p, 2p(1-p)\}(1-p)^{\ell-1}\right)^{n-2\ell}$$

Proof : Fix C a subset of vertices of $G_{n,p}$. Let $\mathcal{S}$ denote the set $\{(X, Y) \mid X \subseteq V, Y \subseteq V, X \neq Y, |X| \leq \ell, |Y| \leq \ell\}$, and let $\mathcal{S}'$ denote the set of maximal pairs of $\mathcal{S}$. For any maximal pair $(X, Y) \in \mathcal{S}'$, let us denote $A_{X,Y}$ the event $\{I(X) = I(Y)\}$, where $I(X)$ denotes $I(X, C)$. Then we have

$$\Pr(C \text{ is not a code}) \leq \sum_{(X,Y) \in \mathcal{S}'} \Pr(A_{X,Y})$$

Since $|\mathcal{S}'| = \Theta(n^{2\ell})$, we have now to compute an upper bound of $\Pr(A_{X,Y}) = \Pr(\cap_{z\in C} A_{X,Y}(z))$, where $A_{X,Y}(z)$ denotes the event $\{z \in I(X) \cap I(Y)\} \cup \{z \notin (I(X) \cup I(Y))\}$. We can decompose this quantity as a product as follows:

$$\Pr(A_{X,Y}) \leq \left\{ \prod_{z\in C\setminus(X\cup Y)} \Pr(A_{X,Y}(z)) \right\} \times \Pr\left(\bigcap_{z\in C\cap(X\cup Y)} A_{X,Y}(z) \right)$$

Indeed, for $z \in C \setminus (X \cup Y)$ the events $A_{X,Y}(z)$ are independent from each other, and are independent from the intersection $\bigcap_{z\in C\cap(X\cup Y)} A_{X,Y}(z)$. Now let us find bounds on each term of this product:

(a) Bound on $\prod_{z\in C\setminus(X\cup Y)} \Pr(A_{X,Y}(z))$:
We decompose the event $A_{X,Y}(z)$ as follows:

$$\begin{aligned} A_{X,Y}(z) \quad = \quad & \{z \in I(X \cap Y)\} \\ & \cup \{z \in I(X \setminus Y) \texttt{ AND } z \in I(Y \setminus X) \texttt{ AND } z \notin I(X \cap Y)\} \\ & \cup \{z \notin I(X \cup Y)\} \end{aligned}$$

This leads to the bound:

$$\begin{aligned} \Pr(A_{X,Y}(z)) \leq f(X,Y) \quad := \quad & 1 - (1-p)^{|X\cap Y|} \\ & + \left(1 - (1-p)^{|X\setminus Y|}\right)\left(1 - (1-p)^{|Y\setminus X|}\right) \\ & \quad (1-p)^{|X\cap Y|} \\ & + (1-p)^{|X\cup Y|} \end{aligned}$$

Without loss of generality, we may assume $|X| \leq |Y|$. Two cases follow:

(a.1) $X \subseteq Y$:
Since (X,Y) is maximal, then we have $|X| = \ell - 1$ and $|Y| = \ell$, hence:

$$f(X,Y) = 1 - (1-p)^{\ell-1} + (1-p)^{\ell} = 1 - p(1-p)^{\ell-1}$$

(a.2) $X \not\subseteq Y$:
Since (X,Y) is maximal, then we have $|X| = |Y| = \ell$. If $|X \cap Y| < \ell - 1$, then let $x_0 \in X \setminus Y$ and $y_0 \in Y \setminus X$. Now set $X' := X \setminus \{x_0\} \cup \{y_0\}$ and $Y' = Y$. As (X,Y) is maximal, then (X',Y') is also maximal, and it is easy to check that $f(X',Y') > f(X,Y)$. Iterating this, we obtain that the maximum of f is attained in the case $|X \cap Y| = \ell - 1$, where we have:

$$\begin{aligned} \Pr(A_{X,Y}(z)) \quad \leq \quad & 1 - (1-p)^{\ell-1} + (1-(1-p))^2(1-p)^{\ell-1} \\ & + (1-p)^{\ell+1} \\ = \quad & 1 - 2p(1-p)^{\ell} \end{aligned}$$

Since $|C \setminus (X \cup Y)| \geq |C| - 2\ell$, this leads to the following bound:

$$\prod_{z \in C \setminus (X \cup Y)} \Pr(A_{X,Y}(z)) \leq \left(1 - \min\{p, 2p(1-p)\}(1-p)^{\ell-1}\right)^{|C|-2\ell}$$

(b) Bound on $\Pr\left(\bigcap_{z \in C \cap (X \cup Y)} A_{X,Y}(z)\right)$:
In the case where $|C| < n$, we simply bound this quantity by 1 and we get the desired result. If $C = V$, then for each pair $(X, Y) \in \mathcal{S}'$ there exists a vertex $z_0 \in X \Delta Y$, the symmetric difference of sets X and Y. Without loss of generality, let us assume that $z_0 \in Y \setminus X$. For such a vertex z_0, we have simply $A_{X,Y}(z_0) = \{z_0 \in N(X)\}$, which has probability $1 - (1-p)^{|X|}$. Since $|X| \leq \ell$, we obtain:

$$\begin{aligned} \Pr\left(\bigcap_{z \in C \cap (X \cup Y)} A_{X,Y}(z)\right) &\leq \Pr(A_{X,Y}(z_0)) \\ &\leq 1 - (1-p)^{\ell} \end{aligned}$$

which leads to the desired bound.

□

Theorem 9 *Let ϵ be such that $n^{\epsilon} \to +\infty$, and p constant, $p \neq 0, 1$. Then almost every graph in $\mathcal{G}(n,p)$ has an ℓ-ID code C of cardinality*

$$|C| \leq \frac{2(\ell + \epsilon)\log n}{\log(1/q_\ell)}$$

where $q_\ell = 1 - \min\{p, 2p(1-p)\}(1-p)^{\ell-1}$.

Proof : With the above settings, we know by Lemma 4 that

$$\Pr(C \text{ is not a code}) \leq n^{2\ell} q_\ell^{|C|-2\ell}$$

It then suffices to check that $n^{2\ell} q_\ell^{|C|-2\ell} \to 0$ if $|C| = \frac{2(\ell+\epsilon)\log n}{\log(1/q_\ell)}$, which is straightforward. □

Notice that this doesn't mean that almost surely the minimum cardinality of an ℓ-ID code in $G_{n,p}$ is $O(\ell \log n)$, because $\frac{1}{\log(1/q_\ell)}$ is $O(2^\ell)$. We rather have that the minimum cardinality of an ℓ-ID code in $G_{n,p}$ is almost surely $O(\ell 2^\ell \log n)$.

The theorem above is analogous to the upper bound in Theorem 1. We were not able to find tight lower bound for the ℓ-ID case, $\ell \neq 1$ and we pose this as an open problem.
As usual, we can derive an existence result from Lemma 4.

Proposition 6 *Let ϵ be such that $n^{\epsilon} \to +\infty$. Then for all $n \in \mathbb{N}$ there exists a graph G^n having an ℓ-ID code C^n of cardinality*

$$|C^n| \leq \sqrt{2}(\ell^2 + \epsilon)\log n$$

Proof : Let $p = \frac{1}{\ell}$. With the above settings, we know by Lemma 4 that

$$\Pr(C \text{ is not a code}) \leq n^{2\ell} \left(1 - 2\frac{1}{\ell}\left(1 - \frac{1}{\ell}\right)^{\ell}\right)^{|C|-2\ell}.$$

For some constant K_ℓ, we have

$$\begin{aligned}
\Pr(C \text{ is not a code}) &\leq K_\ell n^{2\ell} \left(1 - 2\frac{1}{\ell}\left(1 - \frac{1}{\ell}\right)^{\ell}\right)^{|C|} \\
&\leq K_\ell \exp\left(2\ell \log n - 2\frac{|C|}{\ell}\left(1 - \frac{1}{\ell}\right)^{\ell}\right) \\
&\leq K_\ell \exp\left(2\ell \log n - \sqrt{2}\frac{|C|}{\ell}\right),
\end{aligned}$$

since $\left(1 - \frac{1}{\ell}\right)^{\ell} \geq \frac{1}{\sqrt{2}}$ for $\ell > 1$. Concluding the computations,

$$\begin{aligned}
\Pr(C \text{ is not a code}) &\leq K_\ell \exp\left(\left(2\ell - 2\ell - 2\frac{\epsilon}{\ell}\right)\log n\right) \\
&\leq K_\ell \exp\left(-2\frac{\epsilon}{\ell}\log n)\right) \\
&\leq K_\ell n^{-2\epsilon/\ell}
\end{aligned}$$

Since $n^\epsilon \to +\infty$, we have that $\Pr(C \text{ is not a code}) \to 0$. Hence for such a p almost every graph of $\mathcal{G}(n,p)$ has an ℓ-ID code of cardinality $\sqrt{2}(\ell^2 + \epsilon)\log n$, in particular there exists a graph on n vertices G^n having an ℓ-ID code C^n of cardinality $|C^n| \leq \sqrt{2}(\ell^2 + \epsilon)\log n$. □

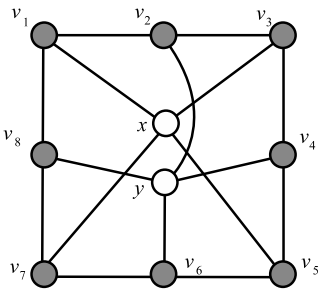
3.2 ℓ-ID and superimposed codes

Given a graph G together with an ℓ-ID code C, we can construct a binary matrix $M = M(G, C)$ as follows: The rows of the matrix correspond to the vertices of the code, and the columns of M correspond to the vertices of G, with M_{ij} equal to 1 if and only if the vertex i is a neighbor of the vertex j or $i = j$. Alternatively, we obtain M as the concatenation of the characteristic vectors of the sets $I(x)$, $x \in V$. (See the example in Figure 2.) Note that if A is the adjacency matrix of G and I is the identity matrix, M is a submatrix of $A + I$.

As C is an ℓ-ID code of G, the columns of M satisfy the following property:

> The bitwise `OR` of any set of at most ℓ columns of M is distinct from the bitwise `OR` of any other set of at most ℓ columns of M. (1)

Indeed, the bitwise `OR` of a set of at most ℓ columns of M is nothing else than the characteristic vector of an identifying set $I(X)$ for a certain X having



	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	x	y
v_1	1	1	0	0	0	0	0	1	1	0
v_2	1	1	1	0	0	0	0	0	0	1
v_3	0	1	1	1	0	0	0	0	1	0
v_4	0	0	1	1	1	0	0	0	0	1
v_5	0	0	0	1	1	1	0	0	1	0
v_6	0	0	0	0	1	1	1	0	0	1
v_7	0	0	0	0	0	1	1	1	1	0
v_8	1	0	0	0	0	0	1	1	0	1

Figure 2: A graph together with a code identifying sets of at most 2 vertices, and its corresponding matrix.

no more than ℓ vertices. A set of 0-1 vectors satisfying (1) is called a UD_ℓ-*code* or ℓ-*superimposed code*. This notion was introduced in [11] by Kautz and Singleton. A connection between identifying codes and superimposed codes was introduced in [10]. To clarify terminology: The *dimension of the space* in which ℓ-superimposed code lies is the number of rows in the matrix M and the *cardinality* of the code is the number of columns.

There are known bounds on the cardinality of ℓ-superimposed codes:

Theorem 10 *[3, 7, 14] There exists an absolute constant c such that, in a space of dimension N, the maximum cardinality of an ℓ-superimposed code is less or equal to*

$$\exp\left(cN\frac{\log \ell}{\ell^2}\right).$$

As a result, we have a bound on ℓ-ID codes.

Corollary 2 *Let C be an ℓ-ID code of a graph G on n vertices. Then there exists a constant c such that*

$$|C| \geq c\frac{\ell^2}{\log \ell}\log n.$$

Proof : An ℓ-ID code C in a graph on n vertices gives us an ℓ-superimposed code consisting of n vectors in a space of dimension $|C|$. Thus we have $n \leq \exp\left(c|C|\frac{\log \ell}{\ell^2}\right)$. This leads to the desired lower bound in Theorem 11. □

We would like to get bounds for graphs which have ℓ-ID codes that are as small as possible. Let

$$m(n) = \min_{|V(G)|=n}\{\min|C| : C \text{ is an } \ell\text{-ID code of } G\}.$$

Karpovsky, et al. [10] established that $m(n) = \omega(\ell \log n)$. If we put Proposition 6 and Corollary 2 together, then we get sharper bounds on $m(n)$.

Theorem 11 *For appropriate constants c_1 and c_2, $m(n)$ satisfies these inequalities*

$$c_1 \frac{\ell^2}{\log \ell} \log n \leq m(n) \leq c_2 \ell^2 \log n.$$

We have seen that, given an ℓ-ID code, we can easily construct an ℓ-superimposed code. A natural question is the converse: Can we construct ℓ-ID codes from ℓ-superimposed codes? Clearly, things don't work so easily in this direction. Though, if we restrict ourselves to *optimal* ℓ-superimposed codes (that is, ℓ-superimposed codes of maximum cardinality), then we get a correspondence with ℓ-ID codes in *oriented* graphs.

The notion of an ℓ-ID code in an oriented graph $G = (V, A)$ is obtained by replacing, in the definition of ℓ-ID codes, the expression $N[x] = N(x) \cup \{x\}$ by $\Gamma^+[x] := \Gamma^+(x) \cup \{x\}$. Here, $\Gamma^+(x)$ is the set of vertices v which are contained on arcs (x, v). We require the sets

$$I^+(X, C) := \bigcup_{x \in X} \Gamma^+[x] \cap C$$

to be nonempty and distinct for all $X \subseteq V$ such that $|X| \leq \ell$.

Theorem 12 *An optimal ℓ-superimposed code of cardinality t in $\{0,1\}^N$ can be realized as an oriented graph on t vertices together with an ℓ-ID code of cardinality N.*

Proof : Let $\{v_1, \ldots, v_t\}$ be an optimal ℓ-superimposed code in $\{0,1\}^N$, and let M be the $N \times t$ matrix whose columns are the vectors $v_1, \ldots, v_t$. If we get an $N \times N$ submatrix M' having only 1's on its diagonal then we can easily construct an oriented graph on t vertices together with an ℓ-ID code of cardinality N: The vertices corresponding to M' are codewords and the other are non-codewords.

Let $\{A, B\}$ be the 'vectors-coordinates' bipartite graph associated to M: $A = \{1, \ldots, N\}$ and $B = \{v_1, \ldots, v_t\}$, and there is an edge between i and v_j if and only if the i-th coordinate of v_j is 1. We claim that there exist a matching of $\{A, B\}$ which covers A. Indeed, using Hall's theorem, if not, then there exists $X \subseteq A$ such that $|N(X)| < |X|$. Now, if we replace these $|N(X)|$ vectors by the $|X|$ unit vectors of coordinate set X, then we get an ℓ-superimposed code of cardinality greater than the original code, a contradiction. Hence there exists a matching which covers A, which corresponds to an $N \times N$ submatrix M' having only 1's on its diagonal. □

Note that also from an ℓ-ID code of an *oriented* graph we can get an ℓ-superimposed code as in the non-oriented case. Hence we have a complete correspondence between maximum ℓ-superimposed codes and minimum ℓ-ID codes in oriented graphs.

3.3 Threshold probabilities admitting ℓ-ID codes

In the general case $\ell > 1$ we only have partial results about threshold functions for the property of admitting an ℓ-ID code. Clearly, some results about ID codes

still apply:

Proposition 7 *If $p = o(n^{-2})$ then almost every graph in $\mathcal{G}(n,p)$ has an ℓ-ID code.*

Indeed for such a p almost surely $G_{n,p}$ has no edge, hence has almost surely the unique ℓ-ID code $C = V$.

Proposition 8 *Let $\epsilon > 0$. If $p \geq 1 - \frac{1}{n}(\log n - \epsilon \log\log n)$ then almost no graph in $\mathcal{G}(n,p)$ has an ℓ-ID code.*

This follows from the fact that for such a p the graph $G_{n,p}$ contains almost surely two universal vertices.

Using the second part of Lemma 4, we can prove the following:

Proposition 9 *Let $\epsilon > 0$. If*

$$\frac{\ell 2^{\ell-1}}{n}(\log n + \epsilon \log\log n) \leq p \leq 1 - \left(\frac{1}{n}(\log n + \epsilon \log\log n)\right)^{1/\ell}$$

then almost every graph in $\mathcal{G}(n,p)$ has an ℓ-ID code.

Proof : Using Lemma 4, we have that:

$$\begin{aligned}
&\Pr(V \text{ is not a code}) \\
&\leq \quad n^{2\ell}\left(1-(1-p)^{\ell}\right)\left(1-\min\{p, 2p(1-p)\}(1-p)^{\ell-1}\right)^{n-2\ell} \\
&\leq \quad \begin{cases} K_\ell n^{2\ell}\left(1-(1-p)^{\ell}\right)\exp\left(-pn(1-p)^{\ell-1}\right), & \text{if } p \leq 1/2; \\ K'_\ell n^{2\ell}\left(1-(1-p)^{\ell}\right)\exp\left(-2pn(1-p)^{\ell}\right), & \text{if } p \geq 1/2. \end{cases}
\end{aligned}$$

where K_ℓ and K'_ℓ are two constants, each depending only on ℓ. It is easy to see that these quantities tend to 0 for a p satisfying the above inequalities. □

For the next proposition we need a result of Bollobás about the degree sequence of a random graph, that we cite here as in [2, Theorem 3.1 (ii)].

Theorem 13 *Let $\epsilon > 0$ be fixed and p such that $\epsilon n^{-\frac{3}{2}} \leq p \leq 1 - \epsilon n^{-\frac{3}{2}}$. Let k be a natural number and let X_k denote the random variable equal to the number of vertices of degree k in $G_{n,p}$. Set*

$$\lambda_k := \lambda_k(n) = n\binom{n-1}{k}p^k(1-p)^{n-k}$$

Then we have:

$$\lim \lambda_k(n) = +\infty \Longrightarrow \lim \Pr(X_k \geq t) = 1,$$

for every fixed $t \geq 0$.

This result is of use because a graph admitting a vertex v with degree $1 \leq d(v) \leq \ell - 1$ has no ℓ-ID code. Indeed, we can't separate the set of neighboring vertices (without v) from the set of closed neighborhood of v.

Proposition 10 *For any $\epsilon > 0$, if $pn^2 \to \infty$ and $p \leq \frac{1}{n}(\log n + (\ell - 1 - \epsilon)\log\log n)$ then almost no graph in $\mathcal{G}(n,p)$ has an ℓ-ID code.*

Proof : We use Theorem 13 with $k = \ell - 1$ and $t = 1$. It is easy to see that if $pn^2 \to \infty$ and $p \leq \frac{1}{n}(\log n + (\ell - 1 - \epsilon)\log\log n)$ then we have $\lambda_k(n) = n\binom{n-1}{k}p^k(1-p)^{n-k} \to +\infty$. Consequently, the graph $G_{n,p}$ contains almost surely a vertex x_0 of degree $\ell - 1$. Now consider the subsets of vertices $X := N(x_0)$ and $Y := N(x_0) \cup \{x_0\}$: X and Y are both of cardinality less or equal to ℓ, and satisfy $N(X) = N(Y)$, hence $G_{n,p}$ has no ℓ-ID code. □

4 Remarks, open problems

Some results of this paper are partial, for instance the threshold function for the property of admitting an ℓ-ID code is unknown. We do not know what happens for

$$\frac{\log n}{n} \leq p \leq \ell 2^{\ell-1}\frac{\log n}{n}$$

and

$$1 - \left(\frac{\log n}{n}\right)^{1/\ell} \leq p \leq 1 - \frac{\log n}{n}.$$

It would be also interesting to diminish the gap between the upper and lower bounds in Theorem 11 and to *explicitly* construct graphs $\{G_n\}$, for all n, such that G_n has a ℓ-ID code C_n of cardinality $|C_n| = \Theta(\ell^2 \log n)$.

We've established in 3.2 a complete correspondence between maximum ℓ-super-imposed codes and minimum ℓ-ID codes in oriented graphs. Our question is: In non-oriented graphs, what can we do to obtain ℓ-ID codes from ℓ-superimposed codes? Up to know, our attempts to establish such a connection have failed. However, we are pretty convinced that this connection exists.

References

[1] N. Alon, J. H. Spencer, *The probabilistic method,* Wiley-Interscience [John Wiley & Sons] (2000).

[2] B. Bollobás, *Random Graphs,* Cambridge University Press (2001).

[3] A. G. D'yachkov, V. V. Rykov, *Bounds on the length of disjunctive codes* Problems of Information Transmission **18** (1983), 166–171.

[4] P. Erdős, L. Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions,* in Infinite and Finite Sets (to Paul Erdős on his 60th birthday), 609–627, North-Holland, Amsterdam (1975).

[5] P. Erdős, A. Rényi, *On the evolution of random graphs,* Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5** (1960), 17–61.

[6] P. Erdős, A. Rényi, *On the evolution of random graphs,* Bulletin de l'Institut I nternational de Statistique **38(4)** (1961), 343–347.

[7] Z. Füredi, *On r-cover-free families,* Journal of Combinatorial Theory, Series A **73(1)** (1996), 172–173.

[8] S. Janson, *New versions of Suen's correlation inequality,* Random Structures Algorithms **13(3-4)** (1998), 467–483.

[9] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, John Wiley and Sons, (2000).

[10] M. G. Karpovsky, K. Chakrabarty, L. B. Levitin, *On a new class of codes for identifying vertices in graphs,* IEEE Transactions on Information Theory **44(2)** (1998), 599–611.

[11] W. H. Kautz, R. R. Singleton, *Nonrandom binary superimposed codes,* IEEE Transformations on Information Theory **10(4)** (1964), 363–377.

[12] T. Laihonen, S. Ranto, *Families of optimal codes for strong identification,* Discrete Applied Mathematics, **121** (2002), 203–213.

[13] S. Ranto, I. Honkala, T. Laihonen, *Two families of optimal identifying codes in binary Hamming spaces,* IEEE Trans. Inf. Theory **48** (2002), 1200–1202.

[14] M. Ruszinkó, *On the upper bound of the size of the r-cover-free families,* Journal of Combinatorial Theory, Series A **66(2)** (1994), 302–310.

[15] W.-C. S. Suen, *A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph,* Random Structures Algorithms **1(2)** (1990), 231–242.