Minimum Number of Input Clues in Robust Information Retrieval

Ville Junnila and **Tero Laihonen**
Department of Mathematics and Statistics
University of Turku, FI-20014 Turku, Finland
viljun@utu.fi and terolai@utu.fi

Abstract

Information retrieval in associative memories was considered recently by Yaakobi and Bruck in [15]. In their model, a stored information unit is retrieved using input clues. In this paper, we study the problem where at most s ($s \geq 0$) of the received input clues can be false and we still want to determine the sought information unit uniquely. We use a coding theoretical approach to estimate the maximum number of stored information units with respect to a given s. Moreover, optimal results for the problem are given, for example, in the infinite king grid. We also discuss the problem in the class of line graphs where a characterization and a connection to k-factors is given.

Keywords: Information retrieval, associative memory, robustness, Johnson bound, k-factor

1 Introduction

Let $G = (V, E)$ be a simple, undirected and connected graph. The graphic distance $d(x, y)$ between vertices $x \in V$ and $y \in V$ is the number of edges in any shortest path between them. If vertices x and y are adjacent, we use the notation $x \sim y$. The degree of a vertex x, denoted by $\deg(x)$, is the number of vertices adjacent to x. The minimum degree of G is $\delta = \delta(G)$.

Information retrieval in an associative memory is modeled as follows. Let $G = (V, E)$ be a graph where the vertices correspond to the memory entries containing the stored information units. The edges between the vertices represent associations between information units and two vertices $x, y \in V$ are said to be *t-associated* if $d(x, y) \leq t$. The set of t-associated vertices to x constitutes a ball of radius t centered at x, that is,

$$B_t(x) = \{y \in V \mid d(x, y) \leq t\}.$$

If $t = 1$, we omit t and write the closed neighbourhood $B_1(x) = B(x)$.

The information retrieval in the memory is performed based on associations as follows. We have a *reference set* $C \subseteq V$. We retrieve the sought information unit $x \in V$ by receiving *input clues* from C one after another (until we can determine x uniquely). Moreover, if the input clues are t-associated to x then we say that they are *correct* input clues and, if they are not t-associated to x then they are called *false* input clues. We assume that there are at most s false input clues for some fixed $s \geq 0$. The set

$$I_t(x) = I_t(C; x) = B_t(x) \cap C$$

is the set of all possible correct input clues for $x \in V$. Again, if $t = 1$, we write $I_1(x) = I(x)$. Next we consider when it is possible to find a sought information unit uniquely assuming there are at most s false input clues.

If $|I_t(x) \setminus I_t(y)| \leq s$ for some $y \in V$, $x \neq y$, then we cannot retrieve x without ambiguity. Namely, even receiving all the possible correct input clues in $I_t(x)$, we cannot determine x uniquely. Indeed, in spite of receiving $I_t(x)$, y could also be the sought information unit, because we may have received $I_t(x) \cap I_t(y)$ correct input clues for y and the rest $I_t(x) \setminus I_t(y)$ (at most s of them!) are false clues (regarding y). Hence, based on the input clues in $I_t(x)$, the sought information unit could be either x or y (and we cannot expect to receive any more clues for x).

If $|I_t(x) \setminus I_t(y)| \geq s + 1$ for every $y \in V$, $y \neq x$, then we can determine x uniquely. Moreover, in order to determine x, it is enough to receive any

$$m_t(C; x, s) = \max_{x \neq y} |I_t(x) \cap I_t(y)| + s + 1$$

correct clues from $I_t(x)$. Namely, suppose we have received (at least) such amount of correct clues together with at most s false clues. Denote the set of received input clues by U. So $|U \cap I_t(x)| \geq m_t(C; x, s)$ and $|U \setminus I_t(x)| \leq s$. If now some $y \in V$ would be the sought information unit instead of x, then, by the definition of $m_t(C; x, s)$, there are at least $s + 1$ input clues in U which are not in $I_t(y)$. Hence y cannot be the sought information unit, because there would be at least $s + 1$ false input clues regarding y and we allowed at most s.

In the sequel, we assume that we can retrieve all information units in V without ambiguity, so we require from the reference set C that for *all* distinct vertices $x, y \in V$ we have

$$|I_t(C; x) \setminus I_t(C; y)| \geq s + 1. \tag{1}$$

Recall that we may receive up to s false clues. Therefore, in the worst case, we need to listen to at most $m_t(C; x, s) + s$ input clues in order to determine x uniquely. It is natural to require that there is some fixed upper bound m on number of input clues needed to determine any information unit. This leads to the following definition.

Definition 1. Let $G = (V, E)$ and $C \subseteq V$. We say that a pair (G, C) is a *sequential* (t, m)*-associative memory robust against at most* s *false input clues* with the reference set C if (1) holds for all distinct vertices $x, y \in V$ and

$$m_t(C; x, s) + s \leq m \quad \text{for any } x \in V.$$

In short, the sequential (t, m)-associative memory robust against at most s false input clues is called s*-robust* $\mathcal{SAM}_G(t, m)$. We also say that C *gives* an s-robust $\mathcal{SAM}_G(t, m)$ if it is a reference set of the memory. In the sequel, we call the reference set a *code* and its elements *codewords*.

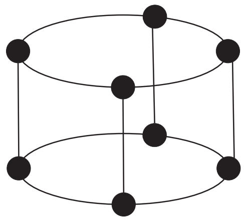


Figure 1: The graph $G_8 = (V, E)$. The set $C = V$ gives a 1-robust $\mathcal{SAM}_{G_8}(1, 5)$.

Naturally, we would like to find the sought information unit with as small number of input clues as possible, so we prefer a code (a reference set) $C \subseteq V$ giving an s-robust $\mathcal{SAM}_G(t, m)$ with as small value of

$$m = \max_{x \neq y}\{|I_t(C; x) \cap I_t(C; y)|\} + 2s + 1. \quad (2)$$

as possible.

Example 2. Consider the graph G_8 in Figure 1. The whole vertex set $C = V$ of G_8 is a code giving a 1-robust $\mathcal{SAM}_{G_8}(1, 5)$ as shown next. The graph is 3-regular and for $C = V$ we have $I(x) = I(V; x) = B(x) \cap V = B(x)$. Hence $|I(x)| = 4$ for all $x \in V$ and it is easy to check that $|I(x) \cap I(y)| \leq 2$ for any $x, y \in V$, $x \neq y$. These observations yield $|I(x) \setminus I(y)| \geq 2$ for distinct vertices $x, y \in V$. Hence the condition (1) is satisfied when $s = 1$. Using (2) we get $m = 5$, as claimed.

Let a code C give an s-robust $\mathcal{SAM}_G(t, m)$. As discussed, we listen to clues (sequentially) one after another in order to find the sought information unit $x \in V$. Suppose we have received clues $U \subseteq C$. We discard any $y \in V$ such that $|U \setminus I_t(y)| \geq s + 1$, since it cannot be the sought information unit (too many false clues regarding y). It is guaranteed by C, that after receiving at most m clues, we are left with one single non-discarded vertex in V. This is the sought information unit x.

The model of information retrieval in an associative memory discussed above was introduced by Yaakobi and Bruck [15] without the notion of false clues (i.e., they considered the case $s = 0$). In this paper, we allow the possibility of at most s false clues. Besides [15], the information retrieval has also been studied in [16], [6], [7], [9] and [8] — again when $s = 0$. Related codes concerning (1) can also be found in [5, Theorem 3]. Apart from the information retrieval in an associative memory, the concept of Definition 1 has applications to Levenshtein's sequences reconstruction problem (see [10, 15, 8]) and to the following problem of RF-based localization.

Remark 3. Consider a *sensor network monitoring* with RF-based localization, which is discussed in [4, 14] for indoor environments. Sensors in a building are mapped to vertices of a graph $G = (V, E)$ and a pair of vertices is connected by an edge if the two corresponding sensors are within each other's communication range. A small portion of all sensors $C \subseteq V$ are kept active while the others can be put in energy-saving mode. The system periodically broadcasts

ID packets from the active sensors. Suppose C gives an s-robust $\mathcal{SAM}_G(t,m)$. An observer can determine her location $x \in V$ from the set of ID packets U that she receives even if there were s false ID packets in U caused by changes in a harsh environment [14] (like door openings permitting ID package from an active sensor, which is not normally at the communication range from x). For more details, see [4, 14].

The paper is organized as follows. In Section 2, we provide optimal results for codes giving an s-robust $\mathcal{SAM}_{\mathcal{K}}(t,m)$ in the king grid $\mathcal{K}$, which is a widely studied graph (see, for example, [3] and numerous other papers in [11]). The Section 3 discusses a lower bound on the number of input clues m when $t = 1$. We use a method from coding theory in Section 4 to estimate the number of possible information units when s is given. It is also shown that this estimate is attained for a family of strongly regular graphs $\mathcal{A}^r$, $r \geq 1$. In the last section, we show that in the class of *line graphs* we can characterize the structure of s-robust codes. We also give optimal results with the aid of suitable k-factors.

2 The infinite king grid

In this section, we consider as the underlying graph the infinite two-dimensional square lattice with diagonals (illustrated in Figure 2). Its vertex set is $V = \mathbb{Z} \times \mathbb{Z} = \mathbb{Z}^2$ and two vertices $u = (u_1, u_2)$ and $v = (v_1, v_2)$ are adjacent if their Euclidean distance is at most $\sqrt{2}$, that is, $|u_1 - v_1| \leq 1$ and $|u_2 - v_2| \leq 1$. We call the resulting infinite graph the *king grid* and denote it by $\mathcal{K}$.

Observe that the ball of radius t centered at $u = (u_1, u_2)$ in $\mathcal{K}$ is a $(2t+1) \times (2t+1)$-square (see Figure 2(a))

$$B_t(u) = \{(x, y) \in \mathbb{Z}^2 \mid |u_1 - x| \leq t, |u_2 - y| \leq t\}.$$

The next theorem provides the optimal results in the king grid $\mathcal{K}$ for any $t \geq 1$ and (all possible values of) s.

Theorem 4. *For each $t \geq 1$ and $0 \leq s \leq 2t$, there exists a code C giving an s-robust $\mathcal{SAM}_{\mathcal{K}}(t,m)$ with $m = 2t(s+1)+2s+1$. Moreover, this is the smallest possible m in the sense that for any code giving an s-robust $\mathcal{SAM}_{\mathcal{K}}(t,m)$ we have $m \geq 2t(s+1) + 2s + 1$.*

Proof. Denote $S_s = \{0, 1, 2, \ldots, s\}$. Let

$$C_{s,t} = \{(a, b) \in \mathbb{Z}^2 \mid a - b \equiv i \pmod{2t+1} \text{ for } i \in S_s\}.$$

We have illustrated $C_{1,2}$ in Figure 2(b). We will verify using (1) and (2) that $C_{s,t}$ gives an s-robust $\mathcal{SAM}_{\mathcal{K}}(t,m)$ with $m = 2t(s+1) + 2s + 1$.

First we consider (1). Let $x = (x_1, x_2) \in \mathbb{Z}^2$ and $y = (y_1, y_2) \in \mathbb{Z}^2$, $x \neq y$.

Assume first that $x_1 < y_1$. Then $B_t(x) \setminus B_t(y)$ contains the set of $2t + 1$ vertically consecutive vertices $\{x + (-t, -t + j) \mid j = 0, 1, \ldots, 2t\}$. Clearly, $s + 1 = |S_s|$ of these consecutive points satisfies the congruence which defines $C_{s,t}$. Hence, $|I(x) \setminus I(y)| \geq s + 1$ as required in (1).

The case $x_1 > y_1$ is analogous to the previous one, so it suffices to assume that $x_1 = y_1$. Then either $x_2 < y_2$ or $x_2 > y_2$. Suppose $x_2 < y_2$ (the case $x_2 > y_2$ is similar). Now $B_t(x) \setminus B_t(y)$ contains the set of the following $2t + 1$

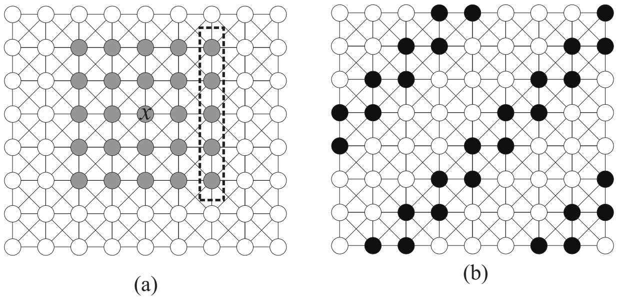


Figure 2: (a) The gray vertices constitute the ball $B_2(x)$. The dashed box illustrates $P_0(x)$. (b) The black vertices constitute the code $C_{1,2}$.

horizontally consecutive points $\{x + (-t + j, -t) \mid j = 0, 1, \ldots, 2t\}$ and $s + 1$ of these belong to $C_{s,t}$ giving again $|I(x) \setminus I(y)| \geq s + 1$.

Next we determine m using (2). If $x_1 < y_1$, then $B_t(x) \cap B_t(y)$ is a subset of $B_t(x) \cap B_t(x + (1, 0))$. The set $B_t(x) \cap B_t(x + (1, 0))$ consists of $2t$ disjoint subsets $P_i(x)$, $i = 0, 1, \ldots, 2t - 1$ where (see Figure 2(a))

$$P_i(x) = \{x + (t - i, -t + j) \mid j = 0, 1, 2, \ldots, 2t\}.$$

Each of these subsets $P_i(x)$ consists of $2t + 1$ vertically consecutive points, so exactly $s + 1$ of them belongs to $C_{s,t}$. Consequently, $|I_t(x) \cap I_t(y)| \leq 2t(s + 1)$. The case $x_1 > y_1$ is analogous. If $x_1 = y_1$ we use horizontal disjoint subsets instead of $P_i(x)$ and the same argument works. Therefore, by (2), we have $m = 2t(s + 1) + 2s + 1$.

Finally, we estimate m for any code D giving an s-robust $\mathcal{SAM}_{\mathcal{K}}(t, m)$. Now the vertical subsets $P_i(x)$ each contain *at least* $s+1$ codewords of D due to (1), which must be satisfied by D. Thus $|I_t(D; x) \cap I_t(D; x + (1, 0))| \geq 2t(s + 1)$. This yield the estimate on $m \geq 2t(s + 1) + 2s + 1$ for D. □

3 Lower bound on the number of input clues

The next result discusses the relation between the number of input clues m and the robustness $s \geq 0$ when the radius $t = 1$.

Theorem 5. *Let G be a d-regular graph and $\Omega = \min_{x \sim y} |B(x) \cap B(y)|$. If C gives an s-robust $\mathcal{SAM}_G(1, m)$, then*

$$m \geq \frac{(s + 2)\Omega}{d} + 2s + 1. \tag{3}$$

Proof. Let C give an s-robust $\mathcal{SAM}_G(1, m)$ in $G = (V, E)$. First we will show that

$$|C| \geq \frac{|V|(s + 2)}{d}. \tag{4}$$

- For each non-codeword $x \in V \setminus C$, we have $|I(x)| \geq s + 2$. Indeed, by (1), we have $|I(x)| \geq s + 1$. Choose $c \in I(x)$. The set $I(x)$ contains the set

$I(x) \setminus I(c)$ and in addition at least the codeword c. By (1), $|I(x) \setminus I(c)| \geq s+1$ and we are done. For a codeword $x \in C$ we have $|I(x)| \geq s+3$. Indeed, let $c' \in I(x)$, $x \neq c'$. Now $I(x) \setminus I(c')$ contains neither x nor c' (and $I(x)$ contains both of them), which gives the claim since $|I(x) \setminus I(c')| \geq s+1$.

Counting in two ways the number N of pairs (x, c) where $x \in V$ and $c \in C$ with $d(x, c) \leq 1$ we get

$$|C|(d+1) = N \geq |V \setminus C|(s+2) + |C|(s+3).$$

This yields the bound (4).

Now we will prove the claimed bound (3) on m. Since each intersection $I(x) \cap I(y)$, where $x \sim y$, contains by (2) at most $m - 2s - 1$ codewords of C, we obtain (since G is d-regular) that

$$(m - 2s - 1)|V|d \geq \sum_{\substack{x,y \in V \\ x \sim y}} |I(x) \cap I(y)|.$$

Next we will verify the estimate

$$\sum_{\substack{x,y \in V \\ x \sim y}} |I(x) \cap I(y)| \geq |C|d\Omega.$$

Indeed, for a fixed $c \in C$ let us count the number of pairs x and y such that $x \sim y$ and $c \in I(x) \cap I(y)$. If $x = c$, then there exist d such pairs (x, y). If $x \neq c$, there are at least $d(\Omega - 1)$ such pairs (notice that $x \neq y$). Consequently, the number of pairs it at least $d\Omega$ for each $c \in C$.

The two inequalities above yield $(m - 2s - 1)|V| \geq |C|\Omega$. Combining this with (4) gives

$$(m - 2s - 1)|V| \geq \frac{|V|(s+2)\Omega}{d}$$

from which the assertion follows. □

The lower bound (3) is attained, for example in the case of Example 2. There we observed that $C = V$ gives a 1-robust $\mathcal{SAM}_{G_8}(1, 5)$. In G_8 we have $\Omega = 2$ and $d = 3$. Consequently, the bound $m \geq 5$ provided by (3) is achieved by the code C.

4 A family of strongly regular graphs

We start the section by discussing in Theorem 6 the relation between the number of information units $|V|$ and the robustness s using an approach from coding theory. Then we show that the bound (5) is attained for a family of strongly regular graphs $\mathcal{A}^r$, $r \geq 1$. Finally, we construct codes (see Example 8 and Theorem 9) giving an s-robust $\mathcal{SAM}(1, m)$ in the graph $\mathcal{A}^r$.

Let n and w be positive integers where $0 \leq w \leq n$. Denote the binary field by $\mathbb{F} = \{0, 1\}$ and the Cartesian product $\mathbb{F}^n = \mathbb{F} \times \mathbb{F} \times \cdots \times \mathbb{F}$ (n times). The elements of $\mathbb{F}^n$ are called *words* and a nonempty subset C of $\mathbb{F}^n$ is called a *code*. For $x = x_1 x_2 \ldots x_n \in \mathbb{F}^n$, we define the *(Hamming) weight* $w(x)$ as the number of coordinates with 1 in x. Let *binary Johnson space* $\mathbb{F}^{n,w}$ consists of all the words in $\mathbb{F}^n$ with weight exactly w. We endow the space with the *Hamming distance* $d(x, y)$ which is the number of bits where words x and y differ.

A subset C of the Johnson space $\mathbb{F}^{n,w}$ is called a *constant weight code* and its elements are called *codewords*. We define further $A(n, d, w)$ to be the greatest cardinality of a constant weight code in which the distinct codewords have Hamming distance at least d apart.

Theorem 6. *Let $G = (V, E)$ be a k-regular graph. If C gives an s-robust $\mathcal{SAM}_G(1, m)$, then*

$$|V| \leq A(|V|, 2(s + 1), k + 1).$$

In particular,

$$|V| \leq \left\lfloor \frac{(s+1)|V|}{(k+1)^2 - (k+1)|V| + (s+1)|V|} \right\rfloor \tag{5}$$

provided that the denominator is positive.

Proof. Let $C = \{c_1, c_2, \ldots, c_{|C|}\}$ give an s-robust $\mathcal{SAM}_G(1, m)$ in a k-regular graph $G = (V, E)$. Let us write the sets $I(C; x)$, $x \in V$, using binary words $b(x) = b_1^x b_2^x \ldots b_{|C|}^x \in \mathbb{F}^{|C|}$ in the following way:

$$b_i^x = \begin{cases} 1 & \text{if } c_i \in B(x), \\ 0 & \text{otherwise.} \end{cases}$$

Let us denote by $\mathcal{B}(C)$ the binary code obtained in this way. Obviously, the distance between any two words of $\mathcal{B}(C)$ is at least $2(s + 1)$, due to (1).

Because C gives an s-robust $\mathcal{SAM}_G(1, m)$, then also V gives an s-robust $\mathcal{SAM}_G(1, m')$ for some $m' \geq m$. Now, since G is k-regular, the code $\mathcal{B}(V)$ is a constant weight code with weight $k + 1$ and the the distance between any two codewords of $\mathcal{B}(V)$ is still at least $2(s + 1)$. Clearly, the number of codewords in $\mathcal{B}(V)$ equals $|V|$. Consequently,

$$|V| \leq A(|V|, 2(s + 1), k + 1).$$

For the second claim (5) we use the Johnson bound [12, p. 525]

$$A(n, 2\delta, w) \leq \left\lfloor \frac{\delta n}{w^2 - wn + \delta n} \right\rfloor$$

if the denominator is positive. □

The following theorem 7 shows that the bound of the previous theorem can be attained.

Let us first introduce some definitions and notations. A graph $G = (V, E)$ is called *strongly regular* with parameters (n, k, λ, μ) if $|V| = n$, G is k-regular and any two adjacent vertices have exactly λ common neighbours and any two nonadjacent vertices have exactly μ common neighbours. See [1] for more information.

Let $\mathbb{A} = \{1, 2, 3, 4\}$ and $r \geq 1$. We will focus on a graph $\mathcal{A}^r$ where the vertex set is the Cartesian product $\mathbb{A}^r = \mathbb{A} \times \cdots \times \mathbb{A}$ (r times) and two different vertices $i_1 \ldots i_r \in \mathbb{A}^r$ and $j_1 \ldots j_r \in \mathbb{A}^r$ are adjacent if and only if the number of coordinates ρ such that $i_\rho + j_\rho = 5$ is even. For any $r \geq 1$, it is known [2] that $\mathcal{A}^r$ is a strongly regular graph with parameters

$$(2^{2r}, 2^{2r-1} + 2^{r-1} - 1, 2^{2r-2} + 2^{r-1} - 2, 2^{2r-2} + 2^{r-1}).$$

Theorem 7. *Let $r \geq 1$. The code $C = \mathbb{A}^r$ gives a $(2^{2r-2}-1)$-robust $\mathcal{SAM}_{\mathcal{A}^r}(1, 3 \cdot 2^{2r-2} + 2^{r-1} - 1)$, which attains the bound (5).*

Proof. Consider the graph $\mathcal{A}^r$ and choose as the code C the whole set of vertices. Since in the graph $\mathcal{A}^r$ the cardinality of the ball $B(x)$ equals $2^{2r-1} + 2^{r-1}$ for all x and the intersection $|B(x) \cap B(y)| = \lambda + 2 = \mu = 2^{2r-2} + 2^{r-1}$ for any x and y with $1 \leq d(x, y) \leq 2$, the set $B(x) \setminus B(y)$ contains 2^{2r-2} elements. Hence it follows that C is a $(2^{2r-2} - 1)$-robust $\mathcal{SAM}_{\mathcal{A}^r}(1, m)$ with

$$m = 3 \cdot 2^{2r-2} + 2^{r-1} - 1.$$

Now clearly, $|\mathbb{A}^r| = 4^r$ and it is easy to check that the right-hand side of (5) equals 4^r also. □

Theorem 7 gives the maximum value for s of an s-robust $\mathcal{SAM}(1, m)$ in $\mathcal{A}^r$. Next we consider smaller values of s in the following example and in Theorem 9.

Example 8. In this example, we provide codes giving an s-robust $\mathcal{SAM}_{\mathcal{A}^r}(1, m)$ for $r = 2$. (For $r = 1$, $\mathcal{A}^1$ is a 4-cycle, and there are clearly no other values of s except the $s = 0$ of the previous theorem). Let

$$C_0 = \{14, 22, 23, 24, 31, 33, 34, 41, 42, 44\},$$

$$C_1 = \{13, 14, 23, 24, 31, 32, 33, 34, 41, 42, 43, 44\}$$

and $C_2 = \mathbb{A}^2 \setminus \{11\}$. The previous codes C_s are such s-robust $\mathcal{SAM}_{\mathcal{A}^2}(1, m_s)$ codes that (as computer searches show) they have the smallest values of (2), namely, $m_0 = 6$, $m_1 = 9$ and $m_2 = 11$. In addition to these, we have the code $C = \mathbb{A}^2$ of the previous theorem giving $m = 13$.

Next we provide an approach to give codes when r is general. Let C be a code giving an s-robust $\mathcal{SAM}(1, m)$ in $\mathcal{A}^r$, $r \geq 1$. In the following theorem, we provide a method to obtain using C a robust $\mathcal{SAM}(1, m')$ for the larger graph $\mathcal{A}^{r+1}$. Let $F \subseteq \mathbb{A}^r$. Denote by

$$\text{App}(F, i) = \{x_1 \dots x_r i \mid x_1 \dots x_r \in F\} \subseteq \mathbb{A}^{r+1}$$

the set where we have appended the last fixed coordinate $i \in \mathbb{A}$.

Theorem 9. *Let $r \geq 1$ and C give an s-robust $\mathcal{SAM}(1, m)$ in $\mathcal{A}^r$. Denote $i_{max} = \max\{|I(x)| \mid x \in \mathbb{A}^r\}$, $i_{min} = \min\{|I(x)| \mid x \in \mathbb{A}^r\}$ and $l = \min\{|I(x) \cup I(y)| \mid x, y \in \mathbb{A}^r, x \neq y\}$. Then the code*

$$D = App(C, 1) \cup App(C, 2) \cup App(C, 3)$$

gives an S-robust $\mathcal{SAM}(1, m')$ in $\mathcal{A}^{r+1}$ with parameters

$$S = \min\{2(s + 1), i_{min}, |C| - i_{max}\} - 1$$

and

$$m' = \max\{3(m - 2s - 1), 2i_{max}, 2(m - 2s - 1) + |C| - l\} + 2S + 1.$$

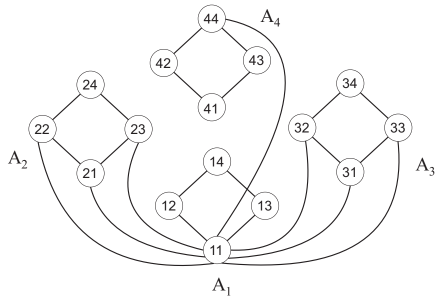


Figure 3: The graph $\mathcal{A}^1$ is a 4-cycle. In the figure there are the vertices of the graph $\mathcal{A}^2$ given as the four "copies" of $\mathcal{A}^1$. The vertices between the different layers A_i are not given except those that are incident with 11 indicating $B(11)$.

Proof. Let us first consider the structure of $\mathcal{A}^{r+1}$ with the aid of $\mathcal{A}^r$. We divide the set of vertices $\mathbb{A}^{r+1}$ into four layers depending on the last coordinate, namely,

$$\mathbb{A}^{r+1} = A_1 \cup A_2 \cup A_3 \cup A_4$$

where

$$A_i = \{x_1x_2 \dots x_r i \mid x_1x_2 \dots x_r \in \mathbb{A}^r\}$$

for any $i = 1, 2, 3, 4$. We also divide a vertex $x \in \mathbb{A}^{r+1}$ into two parts, namely, $x = x'i$ where $x' \in \mathbb{A}^r$ and $i \in \mathbb{A}$.

Next we consider a ball $B(x)$ in $\mathbb{A}^{r+1}$ and examine its parts $B(x) \cap A_j$ in the layers $j = 1, 2, 3, 4$:

- Denote by $\tilde{B}(x')$ be the ball centered at x' in the smaller graph $\mathcal{A}^r$ for any $x = x'i \in \mathbb{A}^{r+1}$. The ball $B(x) = B(x'i)$ consists of the four parts in the different layers (see Figure 3) where

$$B(x'i) \cap A_j = \begin{cases} \mathrm{App}(\tilde{B}(x'), j) & \text{if } i + j \neq 5 \\ \mathrm{App}(\mathbb{A}^r \setminus \tilde{B}(x'), j) & \text{if } i + j = 5. \end{cases} \tag{6}$$

Let us consider the code D where $\mathrm{App}(C, j)$ is a "copy" of the code C in the layer A_j for $j = 1, 2, 3$. Notice that the layer A_4 contains no codewords of D.

Let again $x = x'i$ where $x' \in \mathbb{A}^r$ and $i \in \mathbb{A}$. In what follows, we write in short $I(x') = I(C; x')$ as usual, but in $I(D; x)$ we always keep the D to distinguish between the two codes C and D. Due to (6) we have in the different layers $j = 1, 2, 3, 4:$

$$I(D; x'i) \cap A_j = \begin{cases} \emptyset & \text{if } j = 4, \\ \mathrm{App}(I(x'), j) & \text{if } j \neq 4 \text{ and } i + j \neq 5, \\ \mathrm{App}(C \setminus I(x'), j) & \text{if } j \neq 4 \text{ and } i + j = 5. \end{cases} \tag{7}$$

We use (1) and (2) to verify our claim that D gives an S-robust $\mathcal{SAM}_{\mathcal{A}^{r+1}}(1, m')$ with the given parameters S and m'. For that we also write $y \in \mathbb{A}^{r+1}$, $y \neq x$, as $y = y'h$ where $y' \in \mathbb{A}^r$ and $h \in \mathbb{A}$.

First of all, we show (1), that is, $|I(D; x) \setminus I(D; y)| \geq S + 1$ for any distinct x and y. Denote $T = I(D; x) \setminus I(D; y)$. We calculate $|T|$ examining three cases depending on $x = x'i$ and $y = y'h$. First the case where $i = h$, secondly, when $x' = y'$, and finally the case that $x' \neq y'$ and $i \neq h$.

Case 1: Assume first that $i = h$, so both x and y are in the same layer A_i.

(i) Let first $i = 1$. Notice that the layer A_4 contains no codewords of D. In the other three layers A_j, $j \neq 4$, we have $i + j \neq 5$ and thus in each of these layers we have the set $\mathrm{App}(I(x'), j) \setminus \mathrm{App}(I(y'), j)$ which contributes to T. This set contains the same codewords as $I(x') \setminus I(y')$ appended with j. Hence the cardinality of it equals $|I(x') \setminus I(y')|$. Consequently, we get $|T| = 3|I(x') \setminus I(y')|$.

(ii) Assume then that $i = 2$ (the cases $i = 3, 4$ go similarly). Now the two layers A_j, $j = 1, 2$, both contribute $\mathrm{App}(I(x'), j) \setminus \mathrm{App}(I(y'), j)$ to T. The layer A_3 (since $i + j = 2 + 3 = 5$) contributes $\mathrm{App}(C \setminus I(x'), 3) \setminus \mathrm{App}(C \setminus I(y'), 3)$ to T. This equals the set $I(y') \setminus I(x')$ appended with 3. Hence $|T| = 2|I(x') \setminus I(y')| + |I(y') \setminus I(x')|$.

Case 2: Assume then that $x' = y'$ (and thus $i \neq h$).

(i) Let first $i = 1$. Assume that $h = 2$ (the cases $h = 3, 4$ go analogously and h cannot be 1). For the layers $j = 1, 2$ we have $h + j \neq 5$ and $i + j \neq 5$, so these layers contribute to T the sets $\mathrm{App}(I(x'), j) \setminus \mathrm{App}(I(y'), j)$. However, since $x' = y'$ this set is the empty set. So neither layer contributes to T. In the layer A_3, since $i + 3 \neq 5$ and $h + 3 = 5$, the layer 3 contributes $\mathrm{App}(I(x'), 3) \setminus \mathrm{App}(C \setminus I(y'), 3)$ to T. The fact that $y' = x'$ implies that this set equals $I(x')$ appended with 3. Therefore, $|T| = |I(x')|$.

(ii) Assume next that $i = 2$ (again $i = 3, 4$ are similar). Let first $h = 1$. The layer A_3 contributes $\mathrm{App}(C \setminus I(x'), 3) \setminus \mathrm{App}(I(y'), 3)$ which equals $\mathrm{App}(C \setminus I(x'), 3)$ since $x' = y'$. Other layers contribute nothing, so $|T| = |C \setminus I(x')|$. Assume then that $h = 3$ (the case $h = 4$ is similar and h cannot be 2). Now the layer A_2 (resp. A_3) contributes $\mathrm{App}(I(x'), 2) \setminus \mathrm{App}(C \setminus I(y'), 2)$ (resp. $\mathrm{App}(C \setminus I(x'), 3) \setminus \mathrm{App}(I(y'), 3)$). The layer A_1 contributes nothing, so $|T| = |I(x')| + |C \setminus I(x')| = |C|$.

Case 3: Assume next that both $x' \neq y'$ and $i \neq h$.

(i) Let first $i = 1$. Suppose first that $h = 2$ (again the other cases $h = 3, 4$ are analogous). The layer A_1 (resp. A_2) contributes $I(x') \setminus I(y')$ appended with 1 (resp. 2) to T. The layer A_3 contributes $\mathrm{App}(I(x'), 3) \setminus \mathrm{App}(C \setminus I(y'), 3)$. This set equals $I(x') \cap I(y')$ appended with 3. Consequently, $|T| = 2|I(x') \setminus I(y')| + |I(x') \cap I(y')|$.

(ii) Let then $i = 2$ (the cases $i = 3, 4$ go the same way). Assume first that $h = 1$. Again the layer A_1 (resp. A_2) contributes $I(x') \setminus I(y')$ appended with 1 (resp. 2) to T. The layer A_3 contributes $\mathrm{App}(C \setminus I(x'), 3) \setminus \mathrm{App}(I(y'), 3)$. This set equals $C \setminus (I(x') \cup I(y'))$ appended with 3. Therefore, $|T| = 2|I(x') \setminus I(y')| + |C| - |I(x') \cup I(y')|$.

Suppose then that $h = 3$ ($h = 4$ is similar). Now the layer A_1 contributes $I(x') \setminus I(y')$ appended with 1. The layer A_2 contributes $\mathrm{App}(I(x'), 2) \setminus \mathrm{App}(C \setminus I(y'), 2)$ which corresponds to $I(x') \cap I(y')$. The layer A_3 contributes $\mathrm{App}(C \setminus I(x'), 3) \setminus \mathrm{App}(I(y'), 3)$ which corresponds to $C \setminus (I(x') \cup$

$I(y'))$. This implies that $|T| = |I(x') \setminus I(y')| + |I(x') \cap I(y')| + |C| - |I(x') \cup I(y')| = |I(x')| + |C| - |I(x') \cup I(y')|$.

It is now straightforward to check that each of these values of $|T|$ is bounded from below by $\min\{2(s+1), i_{min}, |C| - i_{max}\}$. The claim for S follows from (1).

We can get the parameter m' using (2) and the previous calculations. For (2) we need (besides S above) the cardinality of the intersections $I(D; x) \cap I(D; y)$, $x \neq y$. Since for any sets X and Y, we have $X \cap Y = X \setminus (X \setminus Y)$, we obtain the value $|I(D; x) \cap I(D; y)|$ using T (which corresponds to $X \setminus Y$) and the set $I(D; x)$ (corresponding to X), which is given in (7).

If $x = x'1$ and $y = y'1$, then the three layers A_j, $j = 1, 2, 3$ contribute $\mathrm{App}(I(x'), j)$ to $I(D; x)$. Consequently, $|I(D; x)| = 3|I(x)|$ and the intersection $|I(D; x) \cap I(D, y)| = |I(D; x)| - |T| = 3|I(x')| - 3|I(x') \setminus I(y')| = 3|I(x') \cap I(y')|$ where $|T|$ comes from Case 1(i) above. The other cases are similar and we only give the results:

The size of the intersection $|(D; x) \cap I(D; y)|$ belongs to the set $\{3|I(x') \cap I(y')|, 2|I(x') \cap I(y')| + |C| - |I(x') \cup I(y')|, 2|I(x')|, |I(x')|, |I(x') \cap I(y')| + |I(x')|, |I(x') \cap I(y')| + |I(y')|, |I(x') \cup I(y')|\}$. It is easy to check that all of these values are bounded from above by $\max\{3(m - 2s - 1), 2i_{max}, 2(m - 2s - 1) + |C| - l\}$. This yields the claim for m'.

□

5 Line graphs

In this section, we consider the problem of robust associative memories in the line graphs. Assuming $G = (V, E)$ is a graph, the line graph $L(G)$ of G is defined as follows. The vertex set of $L(G)$ consists of the edges of G, and two vertices in $L(G)$ are adjacent if the corresponding edges are adjacent in G, i.e., they share a common vertex. Due to the definition of line graphs, studying associative memories in $L(G)$ is equivalent to considering analogous problems in the original graph G for the edges. Hence, we propose the following definitions for the edges of G. For an edge $e \in E$, the notation $B(e)$ is used for the set of edges associated to e, i.e., adjacent to e (also including e itself). Assuming $C \subseteq E$ is code, we define $I(e) = B(e) \cap C$. Notice that in the case of edges we only consider t-association with $t = 1$ implying the similar restriction for the line graphs. Therefore, for the rest of the section, let $t = 1$.

Let s be a nonnegative integer. Recall that we had the earlier condition (1) for distinct vertices $x, y \in V$. In the case of *edges*, assuming $C \subseteq E$, this is equivalent to stating that

$$|I(C; e_1) \setminus I(C; e_2)| \geq s + 1 \tag{8}$$

for any distinct edges $e_1, e_2 \in E$. In the following theorem, we give a characterization of the graphs admitting a code meeting the previous condition.

Theorem 10. *Let G be a simple connected graph on at least three vertices. The graph G admits a code satisfying the condition (8) if and only if $\delta(G) \geq s + 2$ and for any 3-cycle $\mathcal{C}_3$ in G each vertex u belonging to $\mathcal{C}_3$ has $\deg(u) \geq s + 3$.*

Proof. Let $G = (V, E)$ be a graph. Notice first that if a code $C \subseteq E$ satisfies the condition (8), then any C' such that $C \subseteq C'$ also meets the condition.

Therefore, the graph G admits such a code if and only if the code $C = E$ satisfies the condition. Hence, for the graph G, it suffices to verify that

$$|B(e_1) \setminus B(e_2)| \geq s + 1$$

for any distinct $e_1, e_2 \in E$.

Assume first that G admits a code satisfying the condition (8). Assume to the contrary that $\delta(G) \leq s + 1$, or there exists a 3-cycle $\mathcal{C}_3$ and a vertex $u \in \mathcal{C}_3$ such that $\deg(u) \leq s + 2$. Clearly, $\delta(G) \geq 1$. If there exists a vertex $u \in V$ such that $\deg(u) \leq s + 1$, then consider an edge $e_1 = uv$ incident with u. Let us first observe that the vertex v cannot have $\deg(v) = 1$. Indeed, in that case there would be an edge $e' = uz$, $e_1 \neq e'$, due to the fact that G is connected on at least three vertices. This would yield $|B(e_1) \setminus B(e')| = 0$, contradicting (8). Hence $\deg(v) \geq 2$, so we can choose an edge $e_2 = vw$ distinct from uv. Then, if $\deg(u) \leq s+1$, we obtain a contradiction as $|B(e_1) \setminus B(e_2)| \leq s$. Assume then that there exist a 3-cycle $\mathcal{C}_3$ and a vertex u in it such that $\deg(u) \leq s + 2$. Let then $e_1 = uv$ and $e_2 = uw$ be distinct edges of $\mathcal{C}_3$. This again implies a contradiction since $e_3 = vw$ belongs to E as part of the 3-cycle and $|B(e_1) \setminus B(e_3)| \leq s$.

Assume then that $\delta(G) \geq s + 2$ and for any 3-cycle $\mathcal{C}_3$ in G each vertex u belonging to $\mathcal{C}_3$ has $\deg(u) \geq s+3$. Let $e_1 = uv$ and e_2 be distinct edges in G. If e_1 and e_2 are not adjacent to each other, then we obtain (using the assumption) that

$$|B(e_1) \setminus B(e_2)| \geq 2(s + 1) + 1 - 2 = 2s + 1$$

and we are done. Hence, we may assume that e_2 is adjacent to e_1 and denote $e_2 = vw$. Assume first that there is no edge between u and w. This implies that $|B(e_1) \setminus B(e_2)| \geq s + 2 - 1 = s + 1$ and we are done. Therefore, we may assume that $e_3 = uw \in E$. Thus e_1, e_2 and e_3 form a 3-cycle. By the assumption, we now have $\deg(u) \geq s + 3$. Thus, we obtain that (the graph G is simple)

$$|B(e_1) \setminus B(e_2)| \geq s + 3 - 2 = s + 1$$

and the claim follows. □

In the following theorem, we give a characterization for the codes satisfying the condition (8).

Theorem 11. *Let $G = (V, E)$ be a simple connected graph on at least three vertices admitting a code satisfying the condition* (8)*. Then a code $C \subseteq E$ satisfies the condition* (8) *if and only if*

(a) each vertex of G is incident with at least $s + 2$ edges of C and

(b) for any 3-cycle $\mathcal{C}_3$ in G each vertex of $\mathcal{C}_3$ is incident with at least $s + 1$ edges of C not belonging to $\mathcal{C}_3$

Proof. Assume first that $C \subseteq E$ is a code satisfying the condition (8). Assume to the contrary that the condition (a) is not satisfied, i.e., there exists a vertex $u \in V$ such that u is incident with at most $s+1$ edges of C. Let $e_1 = uv$ be any edge of C incident with u. Then, for any distinct edge $e_2 = vw$ (not necessarily in C), we have

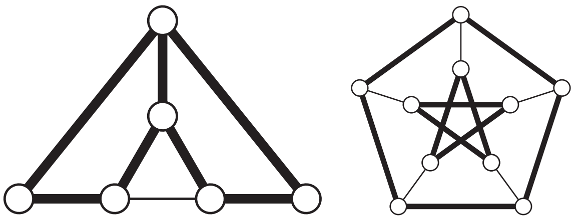
$$|I(C; e_1) \setminus I(C; e_2)| \leq s$$

and a contradiction follows. Assume then to the contrary that the condition (b) is not satisfied, i.e., there exists a 3-cycle $\mathcal{C}_3$ and a vertex $u \in \mathcal{C}_3$ such that u is incident with at most s edges of C not belonging $\mathcal{C}_3$. As the condition (a) is satisfied, there exists distinct edges $e_1 = uv \in C$ and $e_2 = uw \in C$ such that e_1 and e_2 are part of the 3-cycle $\mathcal{C}_3$. Hence, considering the edge $e_3 = vw \in E$, we obtain a contradiction as

$$|I(C; e_1) \setminus I(C; e_3)| \leq s.$$

Assume then that the conditions (a) and (b) hold. Now reasoning similarly as in the proof of Theorem 10, we obtain that (8) is satisfied.

□



(a) Graph with a 3-cycle (b) The Petersen graph

Figure 4: The code consists of the bold edges.

In Figure 4(a), there is an example of a graph with 3-cycle. Let the code C consist of the bold edges in the figure. Clearly, C satisfies the conditions (a) and (b) of Theorem 11 for $s = 0$.

Recall that a reference set $C \subseteq E$ satisfying the condition (8) or the conditions of the previous theorem gives an s-robust $\mathcal{SAM}_G(1, m)$ (regarding the edge set of G) with

$$m = \max_{e_1 \neq e_2} \{|I(C; e_1) \cap I(C; e_2)|\} + 2s + 1.$$

In order to construct such reference sets, we introduce the concept of k-factors, where k is a positive integers. We say that a subgraph H of G is *k-factor* if H contains all the vertices of G and is k-regular. For an extensive coverage on k-factors, we refer the interested reader to the survey [13]. Now we are ready to present the following corollary of Theorem 11.

Corollary 12. *If $G = (V, E)$ is a graph without 3-cycles, then the edges of any $(s+2)$-factor form a reference set $C \subseteq E$ giving an s-robust $\mathcal{SAM}_G(1, m)$ (regarding the edge set of G) with $m = 3s + 3$ and no reference set implying smaller m exists.*

An example of this corollary for $s = 0$ is the 2-factor (denoted by the bold edges) in the Petersen graph of Figure 4(b).

References

[1] A. E. Brouwer, A. M. Cohen, and A. Neumaier. *Distance-regular graphs*, volume 18 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1989.

[2] A. E. Brouwer and J. H. van Lint. Strongy regular graphs and partial geometries. In: *Enumeration and design (Waterloo, Ont, 1982)*, pages 85–122, Academic Press, Toronto, 1984

[3] I. Charon, I. Honkala, O. Hudry and A. Lobstein The minimum density of an identifying code in the king lattice. Discrete Math. 276: 95–109, 2004

[4] N. Fazlollahi, D. Starobinski and A. Trachtenberg. Connected identifying codes. *IEEE Trans. Inform. Theory*, 58(7): 4814–4824, 2012.

[5] I. Honkala and T. Laihonen. On a new class of identifying codes in graphs. *Inform. Process. Lett.*, 102(2-3):92–98, 2007.

[6] V. Junnila and T. Laihonen. Codes for information retrieval with small uncertainty. *IEEE Trans. Inform. Theory*, 60:976–985, 2014.

[7] V. Junnila and T. Laihonen. Information retrieval with unambiguous output. *Information and Computation*, 242: 354–368, 2015.

[8] V. Junnila and T. Laihonen. Information retrieval with varying number of input clues. *IEEE Trans. Inform. Theory*, 62:1–14, 2016.

[9] T. Laihonen. Information retrieval and the average number of input clues. Submitted for publication.

[10] V. Levenshtein. Efficient reconstruction of sequences. *IEEE Trans. Inform. Theory*, 47(1): 2—22, 2001.

[11] A. Lobstein. Watching systems, identifying, locating-dominating and discrminating codes in graphs, a bibliography. Published electronically at `http://perso.telecom-paristech.fr/~lobstein/debutBIBidetlocdom.pdf`.

[12] F. J. MacWilliams and N. J. A. Sloane. The theory of error-correcting codes. North-Holland, Amsterdam, 1977.

[13] M. D. Plummer. Graph factors and factorization: 1985–2003: a survey. *Discrete Math*, 307(7–8): 791—821, 2007.

[14] S. Ray, D. Starobinski, A. Trachtenberg and R. Ungrangsi. Robust location detection with sensor networks. *IEEE Journal on Selected Areas in Communications*, 22(6): 1016-1025, 2004.

[15] E. Yaakobi and J. Bruck. On the uncertainty of information retrieval in associative memories. In *Proceedings of 2012 IEEE International Symposium on Information Theory*, pages 106–110, 2012.

[16] E. Yaakobi, M. Schwartz, M. Langberg, and J. Bruck. Sequence reconstruction for grassmann graphs and permutations. In *Proceedings of 2013 IEEE International Symposium on Information Theory*, pages 874–878, 2013.