FRANK VEGA
Computational Complextiy

Sparse complete sets for coNP: Solution of the P versus NP problem

Frank Vega

March 4, 2019

Abstract: P versus NP is considered as one of the most important open problems in computer science. This consists in knowing the answer of the following question: Is P equal to NP? A precise statement of the P versus NP problem was introduced independently by Stephen Cook and Leonid Levin. Since that date, all efforts to find a proof for this problem have failed. Another major complexity class is coNP. Whether NP = coNP is another fundamental question that it is as important as it is unresolved. In 1979, Fortune showed that if any sparse language is coNP-complete, then P = NP. We prove there is a possible sparse language in coNP-complete. In this way, we demonstrate the complexity class P is equal to NP.

1 Introduction

The *P* versus *NP* problem is a major unsolved problem in computer science [6]. This is considered by many to be the most important open problem in the field [6]. It is one of the seven Millennium Prize Problems selected by the Clay Mathematics Institute to carry a US$1,000,000 prize for the first correct solution [6]. It was essentially mentioned in 1955 from a letter written by John Nash to the United States National Security Agency [1]. However, the precise statement of the $P = NP$ problem was introduced in 1971 by Stephen Cook in a seminal paper [6].

ACM Classification: F.1.3.3, F.1.3.2

AMS Classification: 68Q15, 68Q17

Key words and phrases: complexity classes, complement language, sparse, completeness, polynomial time

In 1936, Turing developed his theoretical computational model [19]. The deterministic and nondeterministic Turing machines have become in two of the most important definitions related to this theoretical model for computation [19]. A deterministic Turing machine has only one next action for each step defined in its program or transition function [19]. A nondeterministic Turing machine could contain more than one action defined for each step of its program, where this one is no longer a function, but a relation [19].

Another relevant advance in the last century has been the definition of a complexity class. A language over an alphabet is any set of strings made up of symbols from that alphabet [7]. A complexity class is a set of problems, which are represented as a language, grouped by measures such as the running time, memory, etc [7].

In the computational complexity theory, the class P contains those languages that can be decided in polynomial time by a deterministic Turing machine [13]. The class NP consists in those languages that can be decided in polynomial time by a nondeterministic Turing machine [13]. The biggest open question in theoretical computer science concerns the relationship between these classes: Is P equal to NP? In 2012, a poll of 151 researchers showed that 126 (83%) believed the answer to be no, 12 (9%) believed the answer is yes, 5 (3%) believed the question may be independent of the currently accepted axioms and therefore impossible to prove or disprove, 8 (5%) said either do not know or do not care or don't want the answer to be yes nor the problem to be resolved [12].

It is fully expected that $P \neq NP$ [18]. Indeed, if $P = NP$ then there are stunning practical consequences [18]. For that reason, $P = NP$ is considered as a very unlikely event [18]. Certainly, P versus NP is one of the greatest open problems in science and a correct solution for this incognita will have a great impact not only for computer science, but for many other fields as well [1]. Whether $P = NP$ or not is still a controversial and unsolved problem [1]. In this work, we proved the complexity class P is equal to NP. Hence, we solved one of the most important open problems in computer science.

2 Basic Definitions

Let Σ be a finite alphabet with at least two elements, and let Σ^* be the set of finite strings over Σ [2]. A Turing machine M has an associated input alphabet Σ [2]. For each string w in Σ^* there is a computation associated with M on input w [2]. We say that M accepts w if this computation terminates in the accepting state, that is $M(w) =$ "*yes*" [2]. Note that M fails to accept w either if this computation ends in the rejecting state, that is $M(w) =$ "*no*", or if the computation fails to terminate [2].

The language accepted by a Turing machine M, denoted $L(M)$, has an associated alphabet Σ and is defined by

$$L(M) = \{w \in \Sigma^* : M(w) = \text{"}yes\text{"}\}.$$

We denote by $t_M(w)$ the number of steps in the computation of M on input w [2]. For $n \in \mathbb{N}$ we denote by $T_M(n)$ the worst case run time of M; that is

$$T_M(n) = max\{t_M(w) : w \in \Sigma^n\}$$

where Σ^n is the set of all strings over Σ of length n [2]. We say that M runs in polynomial time if there is a constant k such that for all n, $T_M(n) \leq n^k + k$ [2]. In other words, this means the language $L(M)$ can be

accepted by the Turing machine M in polynomial time. Therefore, P is the complexity class of languages that can be accepted in polynomial time by deterministic Turing machines [7]. A verifier for a language L is a deterministic Turing machine M, where

$$L = \{w : M(w,c) = \text{“}yes\text{”} \textit{ for some string } c\}.$$

We measure the time of a verifier only in terms of the length of w, so a polynomial time verifier runs in polynomial time in the length of w [2]. A verifier uses additional information, represented by the symbol c, to verify that a string w is a member of L. This information is called certificate. *NP* is also the complexity class of languages defined by polynomial time verifiers [18]. If *NP* is the class of problems that have succinct certificates, then the complexity class *coNP* must contain those problems that have succinct disqualifications [18]. That is, a “*no*” instance of a problem in *coNP* possesses a short proof of its being a “*no*” instance [18].

A function $f : \Sigma^* \to \Sigma^*$ is a polynomial time computable function if some deterministic Turing machine M, on every input w, halts in polynomial time with just $f(w)$ on its tape [19]. Let $\{0,1\}^*$ be the infinite set of binary strings, we say that a language $L_1 \subseteq \{0,1\}^*$ is polynomial time reducible to a language $L_2 \subseteq \{0,1\}^*$, written $L_1 \leq_p L_2$, if there is a polynomial time computable function $f : \{0,1\}^* \to \{0,1\}^*$ such that for all $x \in \{0,1\}^*$,

$$x \in L_1 \textit{ if and only if } f(x) \in L_2.$$

An important complexity class is *NP–complete* [13]. A language $L \subseteq \{0,1\}^*$ is *NP–complete* if

- $L \in NP$, and
- $L' \leq_p L$ for every $L' \in NP$.

If L is a language such that $L' \leq_p L$ for some $L' \in$ *NP–complete*, then L is *NP–hard* [13]. Moreover, if $L \in NP$, then $L \in$ *NP–complete* [13]. A principal *NP–complete* problem is *HAM–CYCLE* [7].

A simple graph is an undirected graph without multiple edges or loops [7]. An instance of the language *HAM–CYCLE* is a simple graph $G = (V,E)$ where V is the set of vertices and E is the set of edges, each edge being an unordered pair of vertices [7]. We say $(u,v) \in E$ is an edge in a simple graph $G = (V,E)$ where u and v are vertices. For a simple graph $G = (V,E)$ a simple cycle in G is a sequence of distinct vertices $\langle v_0, v_1, v_2, ..., v_k \rangle$ such that $(v_k, v_0) \in E$ and $(v_{i-1}, v_i) \in E$ for $i = 1, 2, ..., k$ [7]. A Hamiltonian cycle is a simple cycle of the simple graph which contains all the vertices of the graph. A simple graph that contains a hamiltonian cycle is said to be hamiltonian; otherwise, it is nonhamiltonian [7]. The problem *HAM–CYCLE* asks whether a simple graph is hamiltonian [7].

3 Summary

In computational complexity theory, a sparse language is a formal language (a set of strings) such that the complexity function, counting the number of strings of length n in the language, is bounded by a polynomial function of n. The complexity class of all sparse languages is called *SPARSE*. *SPARSE* contains *TALLY*, the class of unary languages, since these have at most one string of any one length.

Fortune showed in 1979 that if any sparse language is *coNP–complete*, then $P = NP$ (this is Fortune's theorem) [10]. Mahaney used this to show in 1982 that if any sparse language is *NP–complete*, then $P = NP$ [15]. A simpler proof of this based on left-sets was given by Ogihara and Watanabe in 1991 [17]. Mahaney's argument does not actually require the sparse language to be in *NP*, so there is a sparse *NP–hard* set if and only if $P = NP$ [15].

We create a class with the opposite definition, that is a class of languages that are dense instead of sparse. We show there is a sequence of languages that are in *NP–complete*, but their density grows as much as we go forward into the iteration of the sequence. The first element of the sequence is a variation of the *NP–complete* problem known as *HAM–CYCLE* [18]. The next element in the sequence is constructed from this new version of *HAM–CYCLE*. Indeed, each language is created from its previous one in the sequence.

Since the density grows according we move forward into the sequence, then there must be a language so much dense such that its complement is sparse. Fortunately, we find this property from a language created with the elements of these languages on the sequence when the bit length n of the binary strings tends to infinity. However, this incredible dense language is still *NP–complete*. Thus, the complement of this language remains in *coNP–complete*, because the complement of every *NP–complete* language is complete for *coNP* [18]. As a consequence of Fortune's theorem, we demonstrate that P is equal to NP. To sum up, we proved there is a sparse complete set for *coNP* and therefore, we just solved the P versus NP problem.

4 Results

Definition 4.1. A dense language on m is a formal language (a set of **binary** strings) such that for a positive integer n_0, the counting of the number of strings of length $n \geq n_0$ in the language is greater than or equal to 2^{n-m} where m is a real number and $0 \leq m \leq 1$. The complexity class of all dense languages on m is called $DENSE(m)$.

In this work, we are going to represent the simple graphs with an adjacency-matrix [7]. For the adjacency-matrix representation of a simple graph $G = (V, E)$, we assume that the vertices are numbered $1, 2, \ldots, |V|$ in some arbitrary manner. The adjacency-matrix representation of a simple graph G consists of a $|V| \times |V|$ matrix $A = (a_{i,j})$ such that $a_{i,j} = 1$ when $(i, j) \in E$ and $a_{i,j} = 0$ otherwise [7]. In this way, every simple graph of k vertices is represented by k^2 bits.

Observe the symmetry along the main diagonal of the adjacency matrix in this kind of graph that is called simple. We define the transpose of a matrix $A = (a_{i,j})$ to be the matrix $A^T = (a^T_{i,j})$ given by $a^T_{i,j} = a_{j,i}$. Hence the adjacency matrix A of a simple graph is its own transpose $A = A^T$.

Definition 4.2. The language *NON–SIMPLE* contains all the graph that are represented by an adjacency-matrix A such that $A \neq A^T$ or there is some $a_{i,j} = 1$ where $i = j$.

Lemma 4.3. *NON–SIMPLE* $\in P$.

Proof. Given a binary string x, we can check whether x is an adjacency-matrix which is not equal to its own transpose in time $O(|x|^2)$ just iterating each bit $a_{i,j}$ in x and checking whether $a_{i,j} \neq a_{j,i}$ or $a_{i,j} = 1$ when $i = j$ where $|\ldots|$ represents the bit-length function [7]. □

Definition 4.4. The language *HAM–CYCLE'* contains all the binary strings z such that $z = xy$, the bit-length of x is equal to $(\lfloor\sqrt{|z|}\rfloor)^2$ and $x \in$ *HAM–CYCLE* or $x \in$ *NON–SIMPLE* where y could be the empty string when $|\ldots|$ and $\lfloor\ldots\rfloor$ represent the bit-length function and the floor function respectively.

Lemma 4.5. *HAM–CYCLE' ∈ NP–complete.*

Proof. Given a binary string z such that $z = xy$ and the bit-length of x is equal to $(\lfloor\sqrt{|z|}\rfloor)^2$, we can decide in polynomial time whether $x \notin$ *NON–SIMPLE* just verifying when $x = x^T$ and $a_{i,i} = 0$ for all vertex i. In this way, we can reduce in polynomial time a simple graph $G = (V,E)$ of k vertices encoded as the binary string x such that when x has k^2 bits and $x \notin$ *NON–SIMPLE* then

$$x \in \textit{HAM–CYCLE if and only if } xy \in \textit{HAM–CYCLE'}$$

where y could be the empty string. In this way, we can reduce in polynomial time each element of *HAM–CYCLE* to some element of *HAM–CYCLE'*. Therefore, *HAM–CYCLE'* is in *NP–hard*. Moreover, we can check in polynomial time over a binary string z such that $z = xy$ and the bit-length of x is equal to $(\lfloor\sqrt{|z|}\rfloor)^2$ whether $x \in$ *HAM–CYCLE* or $x \in$ *NON–SIMPLE* since *HAM–CYCLE* $\in$ *NP* and *NON–SIMPLE* $\in$ *NP* because of $P \subseteq NP$ [18]. Consequently, *HAM–CYCLE'* is in *NP*. Hence, *HAM–CYCLE' ∈ NP–complete*. □

Lemma 4.6. *HAM–CYCLE' ∈ DENSE*(1).

Proof. OEIS A000088 gives some number of graphs on n unlabeled points [20]. For 8 points there are 12346 so just over half the graphs on 8 points are Hamiltonian [20]. For 12 points, there are 152522187830 Hamiltonian graphs out of 165091172592 which would claim that over 92% of the 12 point graphs are Hamiltonian [20]. For $n = 2$ there are two graphs, neither of which is Hamiltonian [20]. For $n < 8$ over half the graphs are not Hamiltonian [20]. It does not seem surprising that once n gets large most graphs are Hamiltonian [20].

Choosing a graph on n vertices at random is the same as including each edge in the graph with probability $\frac{1}{2}$, independently of the other edges [4]. You get a more general model of random graphs if you choose each edge with probability p [4]. This model is known as $G_{n,p}$ [4]. It turns out that for any constant $p > 0$, the probability that G contains a Hamiltonian cycle tends to 1 when n tends to infinity [4]. In fact, this is true whenever $p > \frac{c \times \log n}{n}$ for some constant c. In particular this is true for $p = \frac{1}{2}$, which is our case [4].

For all the binary strings z such that $z = xy$ and the bit-length of x is equal to $(\lfloor\sqrt{|z|}\rfloor)^2$, the amount of elements of size $|z|$ in *HAM–CYCLE'* is equal to the number of binary strings $x \in$ *HAM–CYCLE* or $x \in$ *NON–SIMPLE* of size $(\lfloor\sqrt{|z|}\rfloor)^2$ multiplied by $2^{|z|-(\lfloor\sqrt{|z|}\rfloor)^2}$. Since the number of Hamiltonian graphs increases as much as we go further on n, it does not seem surprising either that once n gets large most binary strings belong to *HAM–CYCLE'*. Moreover, the amount of binary strings which have some bit-length n^2 and belongs to *NON–SIMPLE* is considerably superior to the amount of strings with the same bit-length which are valid simple graphs. Actually, we can affirm for a sufficiently large positive integer n_0', all the binary strings of length $n \geq n_0'$ which belong to *HAM–CYCLE'* are indeed more than or equal to 2^{n-1} elements. In this way, we show *HAM–CYCLE' ∈ DENSE*(1). □

Definition 4.7. We will define a sequence of languages *HAM–CYCLE'*$_k$ for every possible integer $1 \leq k$. We state *HAM–CYCLE'*$_1$ as the language *HAM–CYCLE'*. Recursively, from a language *HAM–CYCLE'*$_k$, we define *HAM–CYCLE'*$_{k+1}$ as follows: A binary string xy complies with $xy \in$ *HAM–CYCLE'*$_{k+1}$ if and only if x and y are binary strings, $x \in$ *HAM–CYCLE'*$_k$ or $y \in$ *HAM–CYCLE'*$_k$ such that $|x| = \lfloor \log |xy| \rfloor$ where $|\ldots|$ represents the bit-length function and $\lfloor \ldots \rfloor$ the floor function.

Lemma 4.8. *For every integer* $1 \leq k$, *HAM–CYCLE'*$_k \in$ *NP*.

Proof. This is true for $k = 1$ as we see in Lemma 4.5. Every string xy which belongs to *HAM–CYCLE'*$_2$ complies with $x \in$ *HAM–CYCLE'*$_1$ or $y \in$ *HAM–CYCLE'*$_1$ such that $|x| = \lfloor \log |xy| \rfloor$. Moreover, every string $xyvw$ which belongs to the language *HAM–CYCLE'*$_3$ complies with at least one of these memberships $x \in$ *HAM–CYCLE'*$_1$ or $y \in$ *HAM–CYCLE'*$_1$ or $v \in$ *HAM–CYCLE'*$_1$ or finally $w \in$ *HAM–CYCLE'*$_1$ such that $|xy| = \lfloor \log |xyvw| \rfloor$, $|x| = \lfloor \log |xy| \rfloor$ and $|v| = \lfloor \log |vw| \rfloor$. Furthermore, we can extend this property for every positive integer $k > 3$ in *HAM–CYCLE'*$_k$. Indeed, *HAM–CYCLE'*$_k$ is in *NP* for every integer $1 \leq k$, because the verification of whether the whole string or substrings are indeed elements of *HAM–CYCLE'*$_1$ can be done in polynomial time with the appropriated certificates. □

Theorem 4.9. *For every integer* $1 \leq k$, *HAM–CYCLE'*$_k \in$ *NP–complete*.

Proof. This is true for $k = 1$ by Lemma 4.5. Let's assume is valid for some positive integer $1 \leq k'$. Let's prove this for $k' + 1$. We already know the adjacency-matrix of n^2 zeros represents a simple graph of n vertices which does not contain any edge. This kind of a simple graph does not belong to *HAM–CYCLE'*$_1$. As a consequence, this string will not belong to any *HAM–CYCLE'*$_{k'}$, because its substrings of a quadratic length are also adjacency-matrix of only zeros. Suppose, we have an instance y of *HAM–CYCLE'*$_{k'}$. We can reduce y in *HAM–CYCLE'*$_{k'}$ to zy in *HAM–CYCLE'*$_{k'+1}$ such that

$$y \in \textit{HAM–CYCLE'}_{k'} \textit{ if and only if } zy \in \textit{HAM–CYCLE'}_{k'+1}$$

where the binary string z is exactly a sequence of $\lfloor \log |zy| \rfloor$ zeros. We can do this since we already know $z \notin$ *HAM–CYCLE'*$_{k'}$. Certainly, if the membership $zy \in$ *HAM–CYCLE'*$_{k'+1}$ is true, $z \notin$ *HAM–CYCLE'*$_{k'}$ and $|z| = \lfloor \log |zy| \rfloor$, then $y \in$ *HAM–CYCLE'*$_{k'}$ is also holds according to the Definition 4.7. Due to this reduction remains in polynomial time for every positive integer $1 \leq k'$, then we show *HAM–CYCLE'*$_{k'+1}$ is in *NP–hard*. Moreover, *HAM–CYCLE'*$_{k'+1}$ is also in *NP–complete*, because of Lemma 4.8. □

Theorem 4.10. *For every integer* $1 \leq k$, *if the language HAM–CYCLE'*$_k$ *is in* $DENSE(k')$ *for every instance of bit-length* $n' \geq n_0$, *then HAM–CYCLE'*$_{k+1}$ *is in* $DENSE(\frac{k'}{2})$ *for every instance of bit-length* $n' \geq n_0 + \lfloor \log n_0 \rfloor$.

Proof. If the language *HAM–CYCLE'*$_k$ is in $DENSE(k')$ for every instance of bit-length $n' \geq n_0$, then for every integer $n \geq n_0$ the amount of elements of size $n+i$ in *HAM–CYCLE'*$_{k+1}$ (where $i = \lfloor \log n \rfloor$) is greater than or equal to

$$2^{i-k'} \times 2^n + 2^{n-k'} \times (2^i - 2^{i-k'}).$$

This is because there must be more than or equal to $2^{i-k'}$ elements of size i in *HAM–CYCLE'*$_k$ which are prefixes of the binary strings of size $n+i$ in the language *HAM–CYCLE'*$_{k+1}$. We multiply that amount by 2^n since this is the number of different combinations of suffixes with length n in the binary strings

of size $n+i$. Moreover, there must be more than or equal to $2^{n-k'}$ elements of size n in *HAM–CYCLE'*$_k$ which are suffixes of the binary strings of size $n+i$ in *HAM–CYCLE'*$_{k+1}$. We multiply that amount by $(2^i - 2^{i-k'})$ since this is the number of different combinations of prefixes with length i in the binary strings of size $n+i$ just avoiding to count the previous prefixes twice. If we join both properties, we obtain the sum described by the formula above.

Indeed, this formula can be simplified to

$$2^{n+i-k'} + 2^{n+i-k'} \times (2^0 - 2^{-k'})$$

and extracting a common factor we obtain

$$2^{n+i-k'} \times (1 + (1 - 2^{-k'})$$

which is equal to

$$2^{n+i-k'} \times (2 - \frac{1}{2^{k'}}).$$

Nevertheless, for every real number $0 \leq k' \leq 1$ we have

$$(2 - \frac{1}{2^{k'}}) \geq 2^{\frac{k'}{2}}.$$

Certainly, if we multiply both member of the inequality by $2^{k'}$, we obtain

$$(2^{k'+1} - 1) \geq 2^{k'+\frac{k'}{2}}$$

which is equivalent to

$$2^{k'} \times (2 - 2^{\frac{k'}{2}}) \geq 1$$

that it is true for every real number $0 \leq k' \leq 1$. Thus

$$2^{n+i-k'} \times (2 - \frac{1}{2^{k'}}) \geq 2^{n+i-k'} \times 2^{\frac{k'}{2}}$$

where

$$2^{n+i-k'} \times 2^{\frac{k'}{2}} = 2^{n+i-(k'-\frac{k'}{2})} = 2^{n+i-\frac{k'}{2}}.$$

Since every binary string of size $n' > 2$ has also the bit-length $n+i$ for some natural number n (where $i = \lfloor \log n \rfloor$), then there are more than or equal to $2^{n'-(\frac{k'}{2})}$ elements of the language *HAM–CYCLE'*$_{k+1}$ with length $n' \geq n_0 + \lfloor \log n_0 \rfloor$. In this way, we show *HAM–CYCLE'*$_{k+1}$ is in $DENSE(\frac{k'}{2})$ for every instance of bit-length $n' \geq n_0 + \lfloor \log n_0 \rfloor$. □

Lemma 4.11. *HAM–CYCLE'*$_k \in DENSE(\frac{1}{2^{k-1}})$ *for every instance of bit-length* $n \geq n'_0 + k \times (\log n'_0 + \log k)$ *where the constant* n'_0 *is the positive integer used in the Definition 4.1 and Lemma 4.6 for HAM–CYCLE'.*

Proof. According to Lemma 4.6, *HAM–CYCLE'*$_1$ is in $DENSE(1)$ for every instance of bit-length $n \geq n'_0 \leq n'_0 + \log n'_0$. Consequently, due to Theorem 4.10, *HAM–CYCLE'*$_2$ is in $DENSE(\frac{1}{2})$ for every instance of bit-length $n \geq n'_0 + \lfloor \log n'_0 \rfloor \leq n'_0 + 2 \times (\log n'_0 + 1)$. Moreover, *HAM–CYCLE'*$_3$ is in $DENSE(\frac{1}{4})$ for every instance of bit-length $n \geq n'_0 + \lfloor \log n'_0 \rfloor + \lfloor \log(n'_0 + \lfloor \log n'_0 \rfloor) \rfloor \leq n'_0 + 3 \times (\log n'_0 + \log 3)$ and so forth ... and thus, for every language *HAM–CYCLE'*$_k$, we have *HAM–CYCLE'*$_k \in DENSE(\frac{1}{2^{k-1}})$ for every instance of bit-length $n \geq n'_0 + k \times (\log n'_0 + \log k)$. □

Definition 4.12. We will define a language *HAM–CYCLE'*$_\infty$ as follows: A binary string x complies with $x \in$ *HAM–CYCLE'*$_\infty$ if and only if we obtain that $x \in$ *HAM–CYCLE'*$_k$ and $k^2 \leq |x| < (k+1)^2$ where $|\ldots|$ represents the bit-length function.

Lemma 4.13. *HAM–CYCLE'*$_\infty \in NP$.

Proof. We can calculate the possible value of k from some binary string x using the value $\sqrt{|x|}$. In this way, we should know if $x \in$ *HAM–CYCLE'*$_\infty$, then $x \in$ *HAM–CYCLE'*$_k$. However, for every positive integer k, we can check in polynomial time whether $x \in$ *HAM–CYCLE'*$_k$ just splitting the binary string x into the following substrings $x = x_1x_2x_3 \ldots x_m$ and verifying later whether $x_1 \in$ *HAM–CYCLE'*$_1$ or $x_2 \in$ *HAM–CYCLE'*$_1$ or $x_3 \in$ *HAM–CYCLE'*$_1$ and so forth ... until we finally check whether $x_m \in$ *HAM–CYCLE'*$_1$ where m is polynomially bounded by k and therefore, for the bit-length string $|x|$ as well. Indeed, the language *HAM–CYCLE'*$_\infty$ is in NP, because the verification of whether the whole string or a polynomially amount of substrings are indeed elements of *HAM–CYCLE'*$_1$ can be done in polynomial time with the appropriated certificates. □

Theorem 4.14. *HAM–CYCLE'*$_\infty \in$ *NP–complete.*

Proof. We already know the adjacency-matrix of n^2 zeros represents a simple graph of n vertices which does not contain any edge. This kind of a simple graph does not belong to *HAM–CYCLE'*$_1$. Suppose, we have an instance y of *HAM–CYCLE'*$_1$. We can reduce y in *HAM–CYCLE'*$_1$ to zy in *HAM–CYCLE'*$_\infty$ such that

$$y \in \textit{HAM–CYCLE'}_1 \textit{ if and only if } zy \in \textit{HAM–CYCLE'}_\infty$$

where z is a binary string of a polynomially sequence zeros such that $k^2 \leq |zy| < (k+1)^2$ and the membership in $zy \in$ *HAM–CYCLE'*$_k$ implies that $y \in$ *HAM–CYCLE'*$_1$. Certainly, the argument is based on if $y \in$ *HAM–CYCLE'*$_1$ then $z'y \in$ *HAM–CYCLE'*$_2$ where z' is a sequence of zeros such that $|z'| = \lfloor \log |z'y| \rfloor$. Moreover, $z''z'y \in$ *HAM–CYCLE'*$_3$ where z'' is a sequence of zeros such that $|z''| = \lfloor \log |z''z'y| \rfloor$. Furthermore, $z'''z''z'y \in$ *HAM–CYCLE'*$_4$ where z''' is a sequence of zeros such that $|z'''| = \lfloor \log |z'''z''z'y| \rfloor$. Therefore, the amount of substrings $z^{'\cdots'}$ in z is polynomially bounded by $|y|^2$. In addition, the size of each substring $z^{'\cdots'}$ in z is bounded by $|y| + \log|y| + \log(|y| + \log|y|) + \log(|y| + \log|y| + \log(|y| + \log|y|)) + \ldots$ that is bounded by $|y| + \log|y| + \log(2 \times |y|) + \log(3 \times |y|) + \ldots \leq |y| + \log|y| + \log 2 + \log|y| + \log 3 + \log|y| + \ldots$. Since the amount of substrings $z^{'\cdots'}$ in z is bounded $|y|^2$, then the longest substring $z^{'\cdots'}$ in z is bounded by $|y| + |y|^2 \times \log|y| + \log((|y|^2)!) < |y| + |y|^2 \times \log|y| + |y|^2 \times \log|y|^2$ which is polynomially bounded by $|y|$. In this way, we show *HAM–CYCLE'*$_\infty$ is in *NP–hard*. Moreover, we demonstrate *HAM–CYCLE'*$_\infty$ is also in *NP–complete*, because of Lemma 4.13. □

Lemma 4.15. *HAM–CYCLE'*$_\infty \in DENSE(0)$ *when the bit length n of the binary strings tends to infinity.*

Proof. When k tends to infinity, then $\frac{1}{2^{k-1}}$ tends to 0. In this way, when k tends to infinity, then *HAM–CYCLE'*$_k \in DENSE(0)$ as a consequence of Lemma 4.11. *HAM–CYCLE'*$_\infty$ contains the elements of the languages *HAM–CYCLE'*$_k$ into the interval of binary strings between the bit-length $k^2 \leq n < (k+1)^2$. Those elements will have a bit-length greater than $n'_0 + k \times (\log n'_0 + \log k)$ for a sufficiently large k and by the Lemma 4.11 the density in the interval will be $DENSE(\frac{1}{2^{k-1}})$ which is also in $DENSE(\frac{1}{2^{\sqrt{n}-2}})$ in the same interval of binary strings between the bit-length $k^2 \leq n < (k+1)^2$. In this way, the density totally grows for the language *HAM–CYCLE'*$_\infty$ until we reach the ultimate $DENSE(0)$ into the infinite intervals when the bit length n of the binary strings tends to infinity. □

Theorem 4.16. *There is a sparse language in coNP–complete.*

Proof. In Lemma 4.15, the complement of *HAM–CYCLE'*$_\infty$ is sparse when the bit length n of the binary strings tends to infinity. Thus, the complexity of counting the number of strings with length n in the complement of this language is bounded by a polynomial function of n. Indeed, a language is sparse if and only if its complement is in $DENSE(0)$ when the bit length n of the binary strings tends to infinity [15]. Indeed, the sparse languages are called sparse because there are a total of 2^n strings of length n, and if a language only contains polynomially many of these, then the proportion of strings of length n that it contains rapidly goes to zero as n grows (which means its complement should be in $DENSE(0)$ when n tends to infinity) [15]. Furthermore, if the language is sparse from some interval of instances of bit-length greater than some positive integer n_0, then this will remain sparse for the instances of bit-length lesser than or equal to n_0 since they are bounded by the polynomial function of n^{n_0+1} for $n \geq 2$. However, according to Theorem 4.14, the complement of this language *HAM–CYCLE'*$_\infty$ must be in *coNP-complete*, because the complements of the *NP-complete* problems are complete for *coNP* [18]. □

Lemma 4.17. $P = NP$.

Proof. By the Fortune's theorem, if any sparse language is *coNP–complete*, then $P = NP$ [10]. As result of Theorem 4.16, there is a sparse language in *coNP–complete*. Finally, we demonstrate that P is equal to NP. □

5 Discussion

A logarithmic space Turing machine has a read-only input tape, a write-only output tape, and a read/write work tape [19]. The work tape may contain $O(\log n)$ symbols [19]. In computational complexity theory, *LOGSPACE* is the complexity class containing those decision problems that can be decided by a logarithmic space Turing machine which is deterministic [18]. Whether $LOGSPACE = P$ is another fundamental question that it is as important as it is unresolved [18].

A logarithmic space Turing machine M may compute a function $f : \Sigma^* \to \Sigma^*$, where $f(w)$ is the string remaining on the output tape after M halts when it is started with w on its input tape [19]. We call f a logarithmic space computable function [19]. We say that a language $L_1 \subseteq \{0,1\}^*$ is logarithmic space reducible to a language $L_2 \subseteq \{0,1\}^*$, written $L_1 \leq_l L_2$, if there exists a logarithmic space computable function $f : \{0,1\}^* \to \{0,1\}^*$ such that for all $x \in \{0,1\}^*$,

$$x \in L_1 \text{ if and only if } f(x) \in L_2.$$

The logarithmic space reduction is frequently used for the class *P-complete* [18].

In 1999, Jin-Yi Cai and D. Sivakumar, building on work by Ogihara, showed that if there exists a sparse *P-complete* problem, then $LOGSPACE = P$ [5]. We might extend the proof of this paper to demonstrate that $LOGSPACE = P$. Certainly, we might only need to find some *P-complete* which belongs to $DENSE(1)$ because the *P-completeness* is closed under complement [18]. Indeed, the other steps of that possible proof might be similar to the arguments that we follow in this paper. Consequently, this work would help us not only to solve *P* versus *NP*, but also *LOGSPACE* versus *P*.

6 Conclusions

No one has been able to find a polynomial time algorithm for any of more than 300 important known *NP–complete* problems [11]. A proof of $P = NP$ will have stunning practical consequences, because it leads to efficient methods for solving some of the important problems in *NP* [6]. The consequences, both positive and negative, arise since various *NP–complete* problems are fundamental in many fields [6]. This result explicitly concludes supporting the existence of a practical solution for the *NP–complete* problems because of $P = NP$.

Cryptography, for example, relies on certain problems being difficult. A constructive and efficient solution to an *NP–complete* problem such as $3SAT$ will break most existing cryptosystems including: Public-key cryptography [14], symmetric ciphers [16] and one-way functions used in cryptographic hashing [8]. These would need to be modified or replaced by information-theoretically secure solutions not inherently based on *P–NP* equivalence.

Learning becomes easy by using the principle of Occam's razor-we simply find the smallest program consistent with the data [9]. Near perfect vision recognition, language comprehension and translation and all other learning tasks become trivial [9]. We will also have much better predictions of weather and earthquakes and other natural phenomenon [9].

There are enormous positive consequences that will follow from rendering tractable many currently mathematically intractable problems. For instance, many problems in operations research are *NP–complete*, such as some types of integer programming and the traveling salesman problem [11]. Efficient solutions to these problems have enormous implications for logistics [6]. Many other important problems, such as some problems in protein structure prediction, are also *NP–complete*, so this will spur considerable advances in biology [3].

But such changes may pale in significance compared to the revolution an efficient method for solving *NP–complete* problems will cause in mathematics itself. Stephen Cook says: "...it would transform mathematics by allowing a computer to find a formal proof of any theorem which has a proof of a reasonable length, since formal proofs can easily be recognized in polynomial time." [6].

Research mathematicians spend their careers trying to prove theorems, and some proofs have taken decades or even centuries to find after problems have been stated. For instance, Fermat's Last Theorem took over three centuries to prove. A method that is guaranteed to find proofs to theorems, should one exist of a "reasonable" size, would essentially end this struggle.

Indeed, with a polynomial algorithm for an *NP–complete* problem, we could solve not merely one Millennium Problem but all seven of them [1]. This observation is based on once we fix a formal system such as the first-order logic plus the axioms of *ZF* set theory, then we can find a demonstration in time

polynomial in n when a given statement has a proof with at most n symbols long in that system [1]. This is assuming that the other six Clay conjectures have ZF proofs that are not too large such as it was the Perelman's case [1].

Besides, a $P = NP$ proof reveals the existence of an interesting relationship between humans and machines [1]. For example, suppose we want to program a computer to create new Mozart-quality symphonies and Shakespeare-quality plays. When $P = NP$, this could be reduced to the easier problem of writing a computer program to recognize great works of art [1].

References

[1] SCOTT AARONSON: P $\stackrel{?}{=}$ NP. *Electronic Colloquium on Computational Complexity, Report No. 4*, 2017. 1, 2, 10, 11

[2] SANJEEV ARORA AND BOAZ BARAK: *Computational complexity: a modern approach*. Cambridge University Press, 2009. 2, 3

[3] BONNIE BERGER AND TOM LEIGHTON: Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete. *Journal of Computational Biology*, 5(1):27–40, 1998. [doi:10.1145/279069.279080] 10

[4] BÉLA BOLLOBÁS: *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001. [doi:10.1017/CBO9780511814068] 5

[5] JIN-YI CAI AND D. SIVAKUMAR: Sparse hard sets for P: resolution of a conjecture of Hartmanis. *Journal of Computer and System Sciences*, 58(2):280–296, 1999. [doi:10.1006/jcss.1998.1615] 10

[6] STEPHEN A COOK: The P versus NP Problem, April 2000. at `http://www.claymath.org/sites/default/files/pvsnp.pdf`. 1, 10

[7] THOMAS H CORMEN, CHARLES E LEISERSON, RONALD L RIVEST, AND CLIFFORD STEIN: *Introduction to Algorithms*. The MIT Press, 3rd edition, 2009. 2, 3, 4

[8] DEBAPRATIM DE, ABISHEK KUMARASUBRAMANIAN, AND RAMARATHNAM VENKATESAN: Inversion attacks on secure hash functions using SAT solvers. In *International Conference on Theory and Applications of Satisfiability Testing*, pp. 377–382. Springer, 2007. 10

[9] LANCE FORTNOW: The Status of the P Versus NP Problem. *Commun. ACM*, 52(9):78–86, September 2009. [doi:10.1145/1562164.1562186] 10

[10] S. FORTUNE: A note on sparse complete sets. *SIAM Journal on Computing*, 8(3):431–433, 1979. [doi:10.1137/0208034] 4, 9

[11] MICHAEL R GAREY AND DAVID S JOHNSON: *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco: W. H. Freeman and Company, 1 edition, 1979. 10

[12] WILLIAM I GASARCH: Guest column: The second P $\stackrel{?}{=}$ NP poll. *ACM SIGACT News*, 43(2):53–77, 2012. 2

[13] ODED GOLDREICH: *P, NP, and NP-Completeness: The basics of computational complexity*. Cambridge University Press, 2010. 2, 3

[14] SATOSHI HORIE AND OSAMU WATANABE: Hard instance generation for SAT. *Algorithms and Computation*, pp. 22–31, 1997. [doi:10.1007/3-540-63890-3_4] 10

[15] S. R. MAHANEY: Sparse complete sets for NP: Solution of a conjecture by Berman and Hartmanis. *Journal of Computer and System Sciences*, 25:130–143, 1982. [doi:10.1016/0022-0000(82)90002-2] 4, 9

[16] FABIO MASSACCI AND LAURA MARRARO: Logical cryptanalysis as a SAT problem. *Journal of Automated Reasoning*, 24(1):165–203, 2000. [doi:10.1023/A:1006326723002] 10

[17] M. OGIWARA AND O. WATANABE: On polynomial time bounded truth-table reducibility of NP sets to sparse sets. *SIAM Journal on Computing*, 20:471–483, 1991. [doi:10.1137/0220030] 4

[18] CHRISTOS H PAPADIMITRIOU: *Computational complexity*. Addison-Wesley, 1994. 2, 3, 4, 5, 9, 10

[19] MICHAEL SIPSER: *Introduction to the Theory of Computation*. Volume 2. Thomson Course Technology Boston, 2006. 2, 3, 9

[20] THE ON-LINE ENCYCLOPEDIA OF INTEGER SEQUENCES: Number of graphs on n unlabeled nodes, August 2018. at http://oeis.org/A000088. 5

AUTHOR

Frank Vega
Computational Researcher
Joysonic
Belgrade, Serbia
vega.frank@gmail.com
https://uh-cu.academia.edu/FrankVega

ABOUT THE AUTHOR

FRANK VEGA is essentially a back-end programmer graduated in Computer Science since 2007. In August 2017, he was invited as a guest reviewer for a peer-review of a manuscript about Theory of Computation in the flagship journal of IEEE Computer Society. In October 2017, he contributed as co-author with a presentation in the 7^{th} International Scientific Conference on economic development and standard of living ("EDASOL 2017 - Economic development and Standard of living"). In February 2017, his book "Protesta" (a book of poetry and short stories in Spanish) was published by the Alexandria Library Publishing House. He was also Director of two IT Companies (Joysonic and Chavanasoft) created in Serbia.