On visibility graphs generated by sequences of primes

Brian Lee Mayer[1], Luiz H. A. Monteiro[1]

Abstract

Prime numbers is a well studied yet mysterious subject of mathematics, by generalizing the meaning of "prime" we can perceive some new patterns and relations. Graphs are simply a representation of connected entities, providing a set of properties like, degree, centrality, distance &c; which creates a framework to study series of numbers. In this paper we use graphs to investigate some properties of series arising from differences between numbers with the same number of divisors.

Keywords: divisors, graph, prime number, time series

1. Introduction

Prime number theory is a fascinating subject, many works where made since their discovery, and there are still open problems. In order to perceive new properties and relations the sequence of the differences of consecutive primes is considered, then this sequences if represented as a graph using two algorithms from [1, 2]. The approach here is to generalize the concept of prime number by considering other quantities of divisors, those cases are collected and their first differences is taken, generating a sequence. This sequence is then represented as two different graphs, by using visibility algorithms. Finally these graphs are studied, and their properties like mean degree and degree distribution are extracted.

By using this new approach some relations between numbers with a fixed quantity of divisors are observed, also by only inspecting the divisors, some cases show a simple pattern of generation. These discoveries can be used to improve the existing techniques and many others can still be found, the full discussion is found in 5.

2. Preliminary definitions

In the following paragraphs all definitions used in this study are exposed, followed by our methodology, which will explain the techniques and methods of

Email addresses: `bleemayer@gmail.com` (Brian Lee Mayer), `luizm@usp.br` (Luiz H. A. Monteiro)

the research employed for achieving the result. Lastly, all results are shown in 4 and closed by conclusions in 5.

Firstly, a prime number is a positive integer that is greater than 1 and is only divisible by 1 and itself [3].

Hence the set $\{2, 3, 5, 7, 11\}$ contains only prime numbers. It is true that there are infinitely many prime numbers, and many proofs have been already given. Although there are no known formulas that can generate a prime number p_n explicitly in function of n [3].

The difference operator is used to study the relations between consecutive numbers in a sequence, it is, sometimes, considered the discrete version of the derivative operator.

So let f be a function and h a constant for which $x + h$ is in the domain of f for any x, then the first difference of f, noted Δf, is:

$$\Delta f(x) = f(x + h) - f(x)$$

the constant h is named the difference interval [5]. As this paper is interested in finding differences between consecutive elements h is set to 1.

For example, consider the sequence $(a_n)_{n\in\mathbb{N}} = (a_0, a_1, a_2, \ldots)$, then the sequence of the first differences is the sequence: $(a_{n+1} - a_n)_{n\in\mathbb{N}} = (a_1 - a_0, a_2 - a_1, a_3 - a_2, \ldots)$.

In order to understand the relations among different terms of the sequence a graph is generated by means of an algorithm. A graph is a double (V, S), where V is the set of vertices, or nodes, and S is the set of arcs of the system, commonly written as (i, j) to indicate that i is connected to j [11].

Some graphs are said *directed* to indicate that $(i, j) \not\Longrightarrow (j, i)$, which means there is a path from i to j but not from j to i, in this article all graphs are **not** directed.

One of the algorithms used to generate the graph is the visibility algorithm, defined below.

Two points (t_a, y_a) and (t_b, y_b) with $t_a < t_b$ of a time series are said to meet the visibility criterion, if

$$y_c < y_b + (y_a - y_b)\frac{t_b - t_c}{t_b - t_a}, \quad \forall t_c : t_a < t_c < t_b.$$

Then (t_a, y_a) and (t_b, y_b) are connected nodes in the resulting graph [2].

The other algorithm used is the horizontal visibility, and is defined as follows.

Two points (t_a, y_a) and (t_b, y_b) with $t_a < t_b$ of a time series are said to meet the horizontal visibility criterion, if

$$t_a, t_b > t_c, \quad \forall t_c : t_a < t_c < t_b.$$

Then (t_a, y_a) and (t_b, y_b) are connected nodes in the resulting graph [1].

Studies [1, 2, 6, 7] show that algorithms that maps a time series into a graph conserves various aspects of the time series. One example is that fractal series maps to large-scale graphs.

3. Methodology

For this research a new definition of prime number was employed: here the meaning of being prime means a number has only two divisors, and, as a consequence, these divisors are 1 and itself, which is equivalent with the former definition of prime number. In this new definition there are many classes of primes: one class per number of divisors, and one writes $p_d(n)$ to denote the n-th number that has d divisors. This means, the classical primes are the set $\{p_2(x) : x \in \mathbb{N}^*\}$, so $p_2(1) = 2$ is the first number with 2 divisors, $p_2(2) = 3$, $p_2(3) = 5$ and so fourth. In the same sense, $p_3(1) = 4$, $p_3(2) = 9$, $p_3(3) = 25$, $p_3(4) = 49$ are a few 3-primes, which are spoken like "three-primes". In words, a n-prime is a number that has exactly n divisors. This rule in mathematical terms is given as follows:

A n-prime p_n is natural number such that the set of its divisors $|\{d : d|p_n\}| = n$.

For now, how this set of divisors is generated for each number is not the main concern. Some examples are show in table 1.

After generating primes up to 5 million, their differences are calculated, for each n, using the difference operator defined previously, the resulting in a new time series. This series is named *difference series* and is seen on table 2. This method of considering differences between primes has been used in similar research [8, 9].

Each term of this sequence represents the difference between two n-primes, which are used to study how the original series is changing. Also, the study of small and big differences between prime number is of great interest for many researchers. On this research only 500 terms of each difference series is considered for the graphs generation, due to computational limitations, and on each series both algorithms are run, their images can be seen on the next section.

The next step is to represent each series of differences into two graphs:

- The visibility graph
- The horizontal visibility graph

The visibility graph is generated by applying the visibility algorithm on the difference series. The graphs in 7 are the results for some values on n.

Likewise the horizontal visibility graph is generated by applying the horizontal visibility algorithm. The resulting graphs are shown in 7. For each graph the distribution of degrees was calculated and the graph of k by $P(k)$, where $P(k)$ is the relative frequency of k, was plotted.

Some outlier points had to be removed before modeling with a curve, these points are either the first transition, or the last one: the first because they can only see the next, or because the series was truncated at the 500th term.

- 2-primes: Visibility algorithm: Transitions $1 \to 2$ and $499 \to 500$, which are the only nodes with degree 1. The same is true for the horizontal visibility algorithm.

- 3-primes: Visibility algorithm: The last transition, $499 \to 500$, due to truncation this node degree 1. For the horizontal visibility algorithm transitions $1 \to 2$ and $499 \to 500$, where removed for the same reasons.

- 4-primes: Visibility algorithm: Transitions $1 \to 2$, $1 \to 3$, $488 \to 500$ and $499 \to 500$, are the only nodes with degree 2, there are no nodes with degree 1. Horizontal visibility algorithm: Transitions $1 \to 2$ and $499 \to 500$ are the only nodes with degree 1.

- 5-primes: Visibility algorithm: The last transition, $499 \to 500$ is the only node with degree 1. Transitions $1 \to 2$ and $499 \to 500$, which are the only nodes with degree 1 where removed for the horizontal visibility algorithm.

Using visual inspection one could infer the distribution is a power law, which is compatible with those of large-scale networks [10]. From that fact the function that is know to be the best fit is $y(x) = Cx^{-\gamma}$. So the goal is to find γ and C for the fitted function such that give the best fit. These two numbers are important because they characterize the graph.

Also, the mean degree $\langle k \rangle$ is of interest as it indicates the receiving node probability to a new connection. The results of this analysis are found in the next section.

4. Results

This section shows results obtained by applying the techniques considered in 3, it also shows a relation found during the generation of the n-primes discussed.

The coefficients' values for the fitted curves are seen in 3. The small values of the standard deviation indicates that the curves had a good fit, confirming that the graphs that generated those distributions of degrees are scale-free.

Figure 3 is the assemble of graphic results. The left column has the results for the visibility algorithm in [2], and the right column for [1]. Each row has a different value of n, the number of positive divisors.

During the generation of the n-prime time series for n odd, a pattern emerged, as can be seen in 4.

By visual inspection it is clear that on each line a prime, and its square are seen as divisors, and that on the next line the same happens with the next prime, and so on. This pattern can be expressed in a generic form as $p_3(x) = p_2(x)^2$, in fact this holds for $x < 5 \times 10^6$, the first 5 million primes, limited by the computational power only.

Also note that 5 shows a similar pattern: $p_5(x) = p_2(x)^4$, which holds for the same interval, indicating that the same pattern holds for different values of n. Considering that this pattern holds for $n = 7, 9, 11 \ldots$ the formula can be rewritten as $p_n(x) = p_2(x)^{n-1}$.

These observations led to the formula

$$p_n(x) = p_2(x)^{n-1} \quad \forall n \text{ odd}, \forall x \in \mathbb{N}^* \tag{1}$$

which was already noted in [12]. Interestingly no pattern for n even was observed.

5. Conclusion

This article shows that the visibility of graphs generated from first differences of numbers with 2, 3, 4 and 5 divisors are scale-free graphs, which by [2, 1], one can say that the generating series are of fractal type. This research opens a new frontier on prime research as relations between primes and numbers with odd number of divisors where observed and conjectured, also, many new studies can be employed for n with different values or visibility algorithms.

Since all values of γ and C are close; and $\langle k \rangle$ increased linearly with n there is a strong indication that a relation amongst them exists, and that it holds for bigger values of n. This result let alone opens space for new research.

Finally, these discoveries can be useful for cryptography and other fields as the mean degree and their distribution can indicate trends between prime numbers, which could help the discovery of flaws or points of exploration in current cryptographic algorithms.

References

[1] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, Horizontal visibility graphs: Exact results for random time series, Physical Review E, October 2009.

[2] Lucas Lacasa, Bartolo Luque, Fernando Ballesteros, Jordi Luque, and Juan Carlos Nuño, From time series to complex networks: The visibility graph, PNAS, April 1, 2008, vol. 105, no. 13.

[3] Serge Tabachnikov, Kvant Selecta: Algebra and analysis II, American Mathematical Society, 1999.

[4] Underbook Dudley, Elementary number theory, Dover Publications, 2 ed. 2008.

[5] Samuel Goldberg, Introduction to difference equations, Dover Publications, New York, 1958.

[6] Bruna Amin Gonçalves, Laura Carpi, Osvaldo A. Rosso, Martín G. Ravetti, Time series characterization via horizontal visibility graph and Information Theory, Physica A 464 (2016) 93–102.

[7] I.V. Bezsudnov, A.A. Snarskii, From the time series to the complex networks: The parametric natural visibility graph, Physica A 414 (2014) 53–60.

[8] George G. Szpiro, The gaps between the gaps: some patterns in the prime number sequence, Physica A 341 (2004) 607–617.

[9] Anjan Kumar Chandra, Subinay Dasgupta, A small world network of prime numbers, Physica A 357 (2005) 436–446.

[10] Albert-Laszlo Barabasi, Reka Albert, Hawoong Jeong, Mean-field theory for scale-free random networks, Elsevier, 2018.

[11] Jørgen Bang-Jensen, Gregory Gutin, Digraphs Theory. Algorithms and Applications,Springer-Verlag, 2017.

[12] Omar E. Pol, The On-Line Encyclopedia of Integer Sequences, (2008). http://oeis.org/A001248 (accessed 15 October 2019).

6. Table and table captions

n	n-primes
2	$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, \ldots$
3	$4, 9, 25, 49, 121, 169, 289, 361, 529, 841, \ldots$
4	$6, 8, 10, 14, 15, 21, 22, 26, 27, 33, 34, 35, \ldots$
5	$16, 81, 625, 2401, 14641, 28561, 83521, \ldots$

Table 1: Some n-primes for $n = 2, 3, 4, 5$

n	Δp_n
2	$1, 2, 2, 4, 2, 4, 2, 4, 6, 2, 6, 4, \ldots$
3	$5, 16, 24, 72, 48, 120, 72, 168, 312, 120, 408, 312, \ldots$
4	$2, 2, 4, 1, 6, 1, 4, 1, 6, 1, 1, 3, 1, 7, 5, 4, 2, 1, \ldots$
5	$65, 544, 1776, 12240, 13920, 54960, 46800, 149520, \ldots$

Table 2: First differences of n-primes for $n = 2, 3, 4, 5$

Algorithm	n	$\langle$k$\rangle$	γ	C
visibility	2	3.456	1.12 ± 0.2	0.49 ± 0.1
horizontal	2	3.456	1.65 ± 0.4	1.25 ± 0.5
visibility	3	6.776	0.98 ± 0.2	0.37 ± 0.1
horizontal	3	3.844	1.54 ± 0.2	1.05 ± 0.3
visibility	4	6.128	1.50 ± 0.2	1.10 ± 0.4
horizontal	4	3.416	1.86 ± 0.2	1.59 ± 0.3
visibility	5	16.092	0.53 ± 0.1	0.09 ± 0.03
horizontal	5	3.756	1.54 ± 0.2	1.06 ± 0.3

Table 3: Results of each fitted model

$p_3(x)$	divisors
4	$\{1, 2, 4\}$
9	$\{1, 3, 9\}$
25	$\{1, 5, 25\}$
49	$\{1, 7, 49\}$
121	$\{1, 11, 121\}$
169	$\{1, 13, 169\}$
289	$\{1, 17, 289\}$
361	$\{1, 19, 361\}$
529	$\{1, 23, 529\}$
841	$\{1, 29, 841\}$

Table 4: Pattern on divisors n-primes, for $n = 3$

$p_5(x)$	divisors
16	$\{1, 2, 4, 8, 16\}$
81	$\{1, 3, 9, 27, 81\}$
625	$\{1, 5, 25, 125, 625\}$
2401	$\{1, 7, 49, 2401\}$
14641	$\{1, 11, 121, 1331, 14641\}$
28561	$\{1, 13, 169, 2197, 28561\}$
83521	$\{1, 17, 289, 4913, 83521\}$
130321	$\{1, 19, 361, 6859, 130321\}$
279841	$\{1, 23, 529, 12167, 279841\}$
707281	$\{1, 29, 841, 24389, 707281\}$

Table 5: Pattern on divisors n-primes, for $n = 3$

7. Figure and figure captions

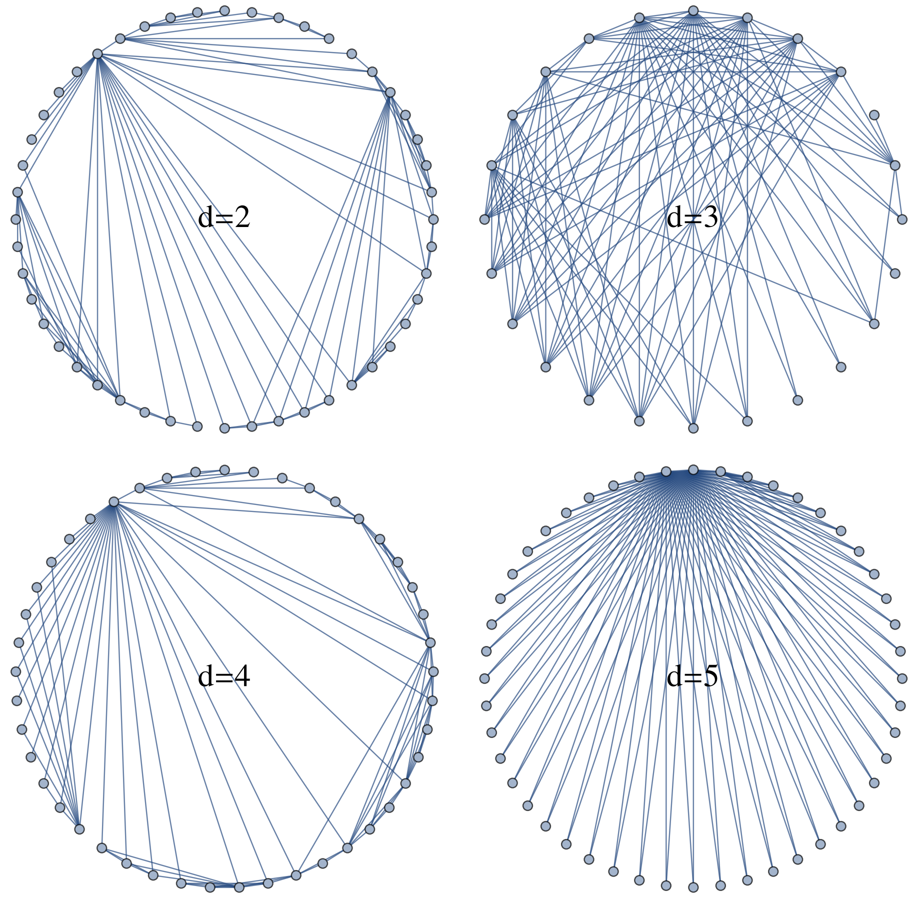

Figure 1: Visibility graphs for $d = 2, 3, 4, 5$ in writing order generated using only the first 100 differences

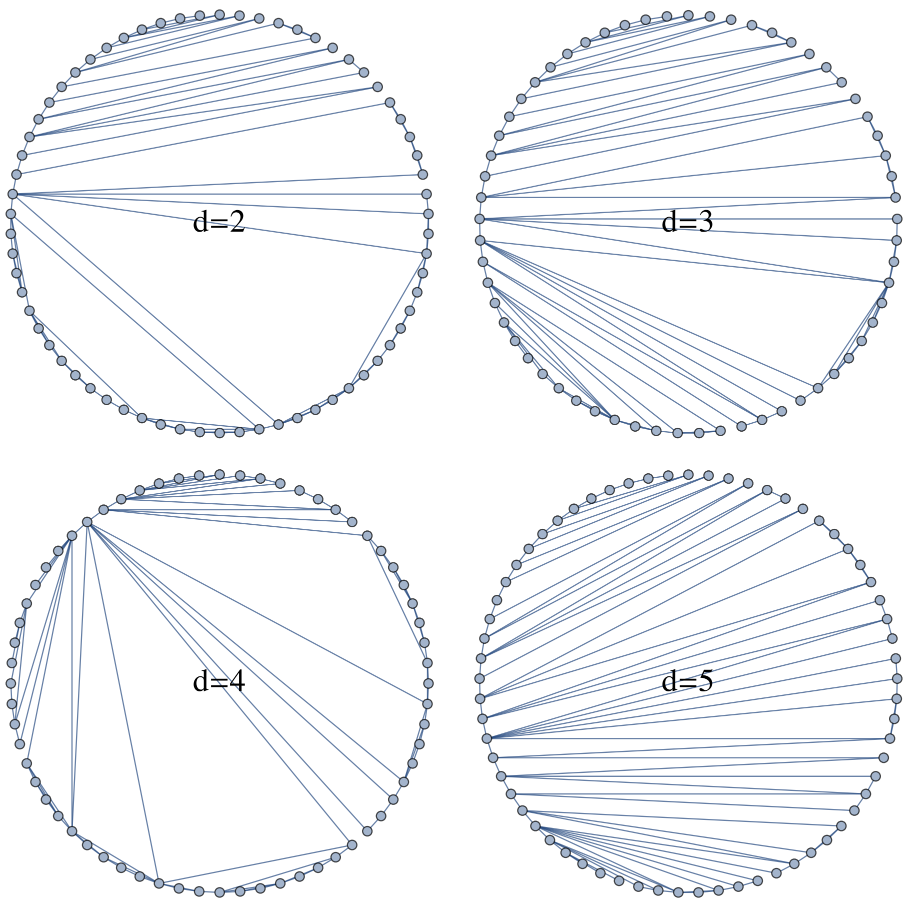

Figure 2: Horizontal visibility graphs for $d = 2, 3, 4, 5$ in writing order generated using only the first 100 differences

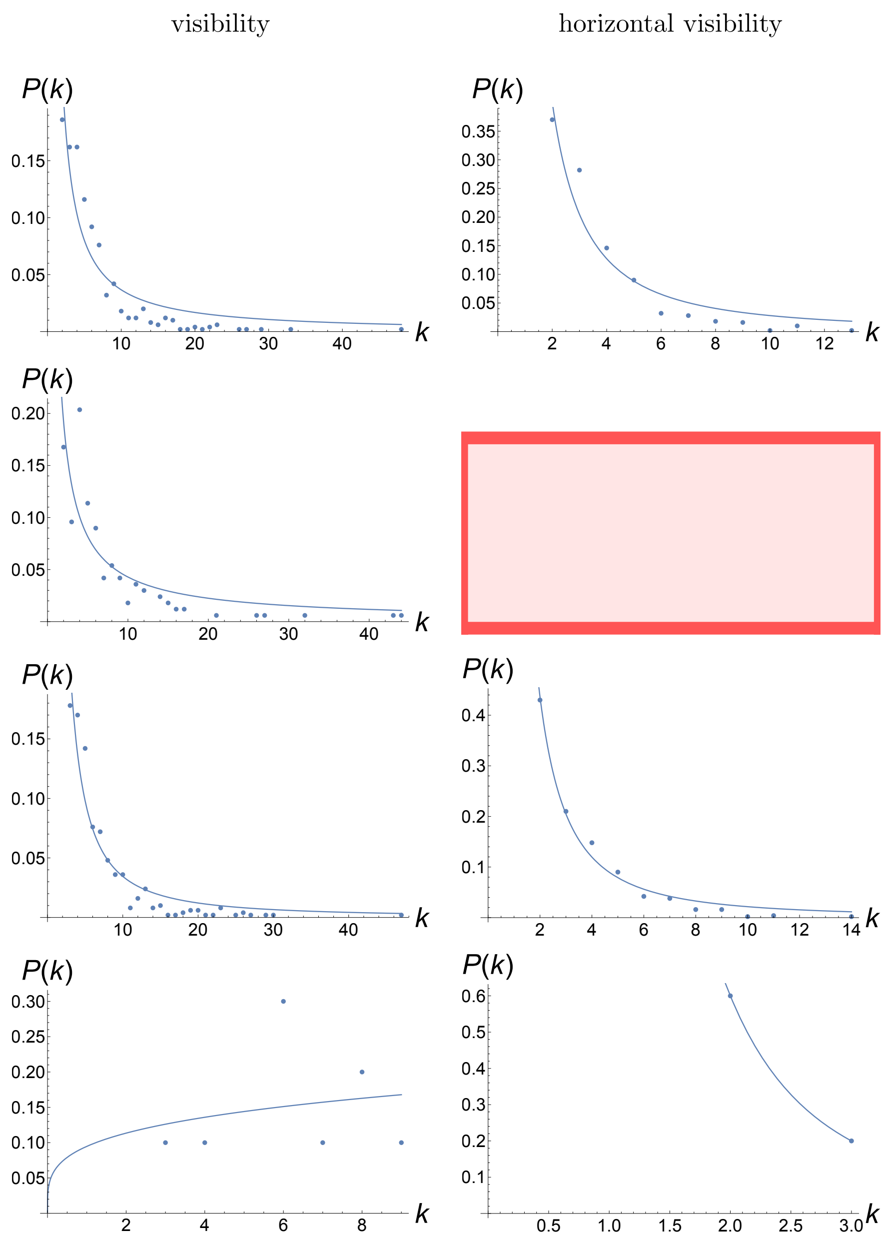

Figure 3: Fitted models for 2, 3, 4, 5 divisors