Generalizing prime numbers using the number of divisors

B.L. Mayer[a,1] and L.H.A. Monteiro[a,b,2]

(a) Universidade Presbiteriana Mackenzie,
PPGEEC, São Paulo, SP, Brazil

(b) Universidade de São Paulo,
Escola Politécnica, São Paulo, SP, Brazil

(1) bleemayer@gmail.com
(2) luizm@mackenzie.br, luizm@usp.br

Corresponding author:
Luiz Henrique Alves Monteiro
Universidade Presbiteriana Mackenzie
Escola de Engenharia
Rua da Consolação, n.896
01302-907, São Paulo, SP, Brazil
E-mail addresses: luizm@mackenzie.br, luizm@usp.br
Telephone number: (55)(11)2114-8711
Fax number: (55)(11)2114-8600

ORCID ID: 0000-0002-2309-1254

Abstract

First let a d-prime denote a positive integer number with exactly d divisors. As a consequence, the usual prime numbers correspond to the particular case $d = 2$. This work consists of numerically investigating series and properties of those numbers. In one scenario, two graphs are built by applying rules to these numbers, and these graphs are analyzed. In the second scenario we study the decimal digits of irrational numbers using informational entropy, and by employing two visibility algorithms these sequences are mapped into graphs. Computer simulations shows that these numbers are not distinguished.

Keywords: d-prime, informational entropy, power law, time series, visibility graph.

1 Introduction

Since 300 BC, prime numbers fascinate mathematicians and math lovers [2, 3]. Attracting the attention of most of the great minds, such as Eratosthenes, Euclid, Dirichlet, Erdös, Euler, Fermat, Legendre, and Riemann [2,3]. Primes are known to be the foundation of analytical studies of many areas, such as number theory [2, 3]. In applied Mathematics, primes have been used in cryptographic keys [4], and can be found in the life cycles of cicadas [5], in Physics, primes characterize the energy spectrum of chaotic quantum systems [6]. Their irregular gaps is a field of study on its own and has been extensively investigated for many centuries [2,32,34]. A prominent research is using graphs to study relations between numbers, for instance, if primes are the nodes of a graph, the edges can represent that these numbers sum to an even number [11]. In this work graphs are employed to study structural properties of sequences of prime numbers.

In this work a positive integer with d divisors is noted as

$$p_d(x), \qquad x \in \mathbb{N}$$

Therefore, the usual prime numbers correspond to the case $d = 2$. Also, this notation is convenient to write sequences of numbers with equal number of divisors, see Table 1. One series investigated is the series of number of divisors, that is, $(|\{d : d|x\}|)$ for $x \in \mathbb{N}$. To evaluate their variability, the informational entropy [13] of these series is computed. In addition, these series are transformed into undirected graphs by applying two visibility algorithms [14, 15], the degree distribution and the average degree of these graphs is investigated.

The aim of this study based on d-primes is to understand how the gap sequences depend on d. This study can provide an alternative way of investigating the occurrence of the usual primes, which are the factors of any d-prime with $d > 2$. The proposed approach can be used to analyze other sequences of numbers found in nature [16], such as the energy levels of atomic nuclei and the quantum space-time structure [17].

2 Methodology

Let a d-prime be defined as a positive integer greater than 1 that is divisible by exactly d positive integers. A d-prime with $d > 2$ is usually called a composite number [2, 3]. Also note that 2-primes are just the usual primes. Thus, d-prime is a naive generalization of the definition of prime number. In mathematical terms: x is a d-prime if $|\{k : k|x\}| = d$. From this definition a sequence of d-primes is noted as:

$$(p_d(x))_{x\in\mathbb{N}}$$

From this definition series of numbers will be generated for different values of d, the following table gives a sample.

Those series are catalogued on OEIS (On-Line Encyclopedia of Integer Sequences) [1].

Table 1: Sequences of d-primes, denoted by $p_d(n)$ with $n \in \mathbb{N}$, for $d \in \{2, 3, ..., 11\}$.

d	$p_d(n)$
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	$16, 81, 625, 2401, 14641, 28561, 83521, \ldots$
6	$12, 18, 20, 28, 32, 44, 45, 50, 52, 63, 68, 75 \ldots$
7	$64, 729, 15625, 117649, 1771561, 4826809, \ldots$
8	$24, 30, 40, 42, 54, 56, 66, 70, 78, 88, 102, \ldots$
9	$36, 100, 196, 225, 256, 441, 484, 676, 1089, \ldots$
10	$48, 80, 112, 162, 176, 208, 272, 304, 368, \ldots$
11	$1024, 59049, 9765625, 282475249, \ldots$

2.1 Visibility algorithms

The following two algorithms are used to convert time series into undirected graphs [14, 15]. These algorithms have been employed, for instance, in the analysis of stock indices [20] and electroencephalography recordings [21]. Here, in the visibility graphs, each node represents a distinct value of x_d.

Consider that $n_a < n_i < n_b$. In the natural visibility (NV) graph [14], the nodes corresponding to $x_d(n_a)$ and $x_d(n_b)$ are connected if any intermediate point $(n_i, x_d(n_i))$ in the time series satisfies the inequality:

$$x_d(n_i) < x_d(n_a) + (x_d(n_b) - x_d(n_a)) \left(\frac{n_i - n_a}{n_b - n_a} \right) \tag{1}$$

Thus, these nodes are connected if there is a straight line joining $(n_a, x_d(n_a))$ and $(n_b, x_d(n_b))$ in the plot $x_d(n) \times n$, provided that any intermediate point $(n_i, x_d(n_i))$ is below such a line.

In the horizontal visibility (HV) graph [15], the nodes associated with $x_d(n_a)$ and $x_d(n_b)$ are connected if:

$$\{x_d(n_a), x_d(n_b)\} > x_d(n_i) \tag{2}$$

with $n_a < n_i < n_b$. Thus, these nodes are connected if any intermediate point $(n_i, x_d(n_i))$ in the plot $x_d(n) \times n$ is below the horizontal line joining $(n_a, x_d(n_a))$ and $(n_b, x_d(n_b))$.

Then, the corresponding degree distributions $P(k)$ and the average degree $\langle k \rangle$ were determined. Recall that the degree of a node is the number of edges connected to this node [22]. Recall also that the degree distribution expresses how the percentage $P(k)$ of nodes with degree k varies with k [22]. Usually, $P(k)$ is interpreted as the probability of randomly picking a node with degree k.

2.2 Relation between primes and numbers with a prime number of divisors

Any natural number can be factored into primes, called prime factors, for prime numbers, the factors are only 1 and itself. If n is prime:

$$n = p_1 \times 1$$

And

$$n^a = (p_1 \times 1)^a = p_1^a \times 1$$

That means n^a has $a+1$ divisors, or, in mathematical notation: $p_d(x) = p(x)^{d-1}$, for d 2-prime. Which is a result that can be found on [2]. This criterion is used to validate results.

2.3 Entropy

The entropy $H = -\sum_{i=1}^{q} p_i \ln p_i$ and its maximum value $H_{max} = \ln q$ (obtained in the case of $p_i = 1/q$ for $i = 1, ..., q$ [13]) are calculated by taking p_i as the relative frequency of occurrence of a distinct x_d. In these expressions, q is the number of distinct values of x_d.

2.4 n-prime gaps

Prime gaps is another subject related to prime numbers that had received great attention in the past [33], and still receives [32]. One recent result [34], was made and improved the formula for the maximum gap. In this section we explore prime gaps for n-primes.

Let $p_d(n)$ be the n-th d-prime, with $n \in \mathbb{N}^*$. For instance, $p_3(4)$ is the fourth 3-prime, which is equal to 49. Table 1 presents a list of the first d-primes for $d \in \{2, 3, ..., 11\}$. Let $x_d(n) = p_d(n+1) - p_d(n)$ be the gap between consecutive d-primes. For instance, $x_5(1) = 65$, because $p_5(2) = 81$ and $p_5(1) = 16$. Note that the sequence $x_d(n)$ can be taken as a time series, in which n corresponds to the time variable. Table 2 shows the first numbers of the series $x_d(n)$ for $d \in \{2, 3, ..., 11\}$.

Per the formula for n-primes, with n prime, we can calculate the gaps for these sequences:

$$x_3(n) = p_3(n+1) - p_3(n) = p(n+1)^2 - p(n)^2 = [p(n+1) + p(n)][p(n+1) - p(n)]$$

But we know that two primes differ at least by 2, the case of twin primes.

$$p(n+1) - p(n) \geq 2$$

Using this on the previous equation:

$$x_3(n) \geq 2(p(n+1) + p(n))$$

In order to evaluate the variability of the series x_d for $d \in \{2, 3, ..., 11\}$, the informational entropy H [13] was computed. This entropy has been calculated, for instance, in investigations on the dynamics of biological [18] and social systems [19]. Its normalized value, denoted by Δ, is given by:

$$\Delta = \frac{H}{H_{max}} \tag{3}$$

2.5 Analysis of irrational numbers' digits

This exploration was made using the 10,000 first decimal digits of the numbers $\sqrt{2}$, e, ϕ and π, the table bellow shows some digits for each number. Subscript notation is used to indicate the numbers composed with more than one digit.

number	first digits
$\sqrt{2}$	(4, 1, 4, 2, 1, 3, 5, 6, 2, 3, 7, 3, 0, 9, 5, 0, 4, 8, ...)
e	(7, 1, 8, 2, 8, 1, 8, 2, 8, 4, 5, 9, 0, 4, 5, 2, 3, 5, 3, ...)
ϕ	(6, 1, 8, 0, 3, 3, 9, 8, 8, 7, 4, 9, 8, 9, 5, ...)
π	(1, 4, 1, 5, 9, 2, 6, 5, 3, 5, 8, 9, 7, 9, 3, 2, 3, 8, ...)

We also considered the series of pairs and triples of digits, i.e. by taking digits two by two and forming a number: 4 and 1 gives 41, 0 and 5 gives 5, and so on, resulting in our second dataset of digits for the same analysis. The table below has a sample.

number	first digits
$\sqrt{2}_2$	(41, 42, 13, 56, 23, 73, 9, 50, 48, 80, 16, 88, 72, 42, ...)
e_2	(71, 82, 81, 82, 84, 59, 4, 52, 35, 36, 2, 87, 47, 13, ...)
ϕ_2	(61, 80, 33, 98, 87, 49, 89, 48, 48, 20, 45, 86, 83, 43, 65, ...)
π_2	(14, 15, 92, 65, 35, 89, 79, 32, 38, 46, 26, 43, 38, ...)

And triples:

number	first digits
$\sqrt{2}_3$	(414, 213, 562, 373, 95, 48, 801, 688, 724, 209, ...)
e_3	(718, 281, 828, 459, 45, 235, 360, 287, 471, ...)
ϕ_3	(618, 33, 988, 749, 894, 848, 204, 586, 834, 365, ...)
π_3	(141, 592, 653, 589, 7932, 384, 626, 433, 832, 795, 28, ...)

Numerical experiments were performed by applying the two algorithms defined on previous sections to convert the series of digits into a graph. Finally this graph is studied and its properties are shown.

Additionally the digit distribution was investigated, first considering single digits, secondly taking pairs of digits and forming a number with the formula $10 \times d_1 + d_2$, with d_1 and d_2 the two digits from the pair. Lastly, the same for triples of digits: $100 \times d_1 + 10 \times d_2 + d_3$. For this investigation one million of digits where used, that is, one million single digits, pairs or triples.

2.6 Series of quantity of divisors

The first is generated by the points $(x, |\{d : d|x\}|)$ for $n \in N$ that is: each number to its quantity of divisors. That is:

$$\{(2,2),(3,2),(4,3),(5,2),\ldots,(10000,25)\}$$

The second applies the divisor function iteratively until the result reaches 2, that is, is a prime number, this series will be called as the orbit series through this paper. For example, take the number 60:

- 60 has 12 divisors, first time
- 12 has 6 divisors, second time
- 6 has 4 divisors, third time
- 4 has 3 divisors, fourth time
- and finally, 3 has 2 divisors, fifth time

Is this case we applied the function 5 times, hence the orbit of 60 has length 5, and the point is $(60, 5)$. Generating the following sequence:

$$\{(2,1),(3,1),(4,2),(5,1),\ldots,(10000,3)\}$$

The intention is to discover for each case how the natural visibility and horizontal visibility measures relate. And in consequence, how the two series are related.

Also, for the two cases the mean shortest path was calculated at each new vertex.

3 Results

This section describes the procedure to build and analyse each series. The analysis consists of calculating the natural and horizontal visibilities for the series, then fitting the model $Cx^{-\gamma}$ on the degree distribution of the generated graph and analyse the coefficients.

In the studies the numbers started at 2 to avoid the only case of 1 divisor. On the second study specifically, the orbit of 2 was set to 1 instead of 0 to avoid this special case and, as consequence, dividing the graph in two disconnected subgraphs. All links are considered undirected.

The following table summarizes the parameter values of the two studies.

series	visibility	γ	C
divisors	natural	1.21489 ± 0.18265	0.559692 ± 0.132711
divisors	horizontal	0.637136 ± 0.658635	0.190621 ± 0.174726
orbits	natural	0.556285 ± 0.414888	0.149885 ± 0.0999156
orbits	horizontal	0.353234 ± 1.56196	0.198959 ± 0.379138

(4)

In the following paragraphs all results for the two studies are displayed, for each set we show first the graph, for illustration purposes, then the plot of the mean shortest path.

The figure (1) shows a graph that is generated using the divisors series, by adding links $(x, y) \mapsto x \to y$

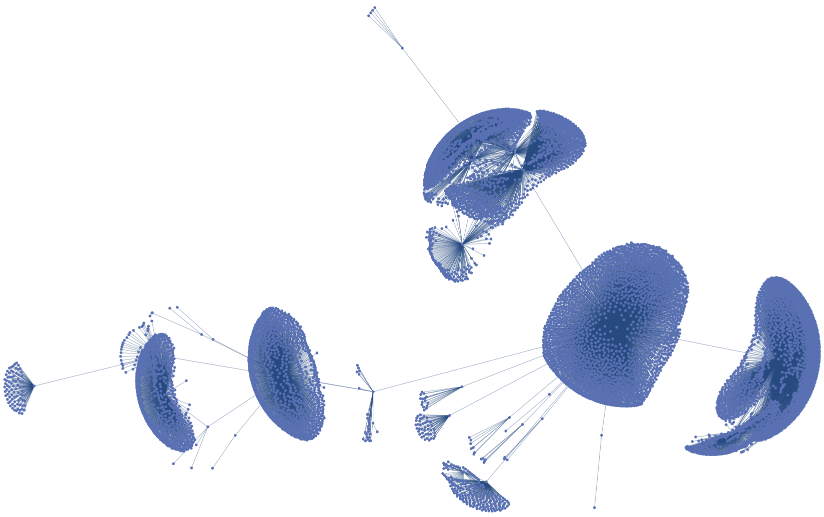

Figure 1: Graph illustrating the divisors series

The figure (2) shows the corresponding plot of mean shortest path in relation to the number of nodes for the divisors series.

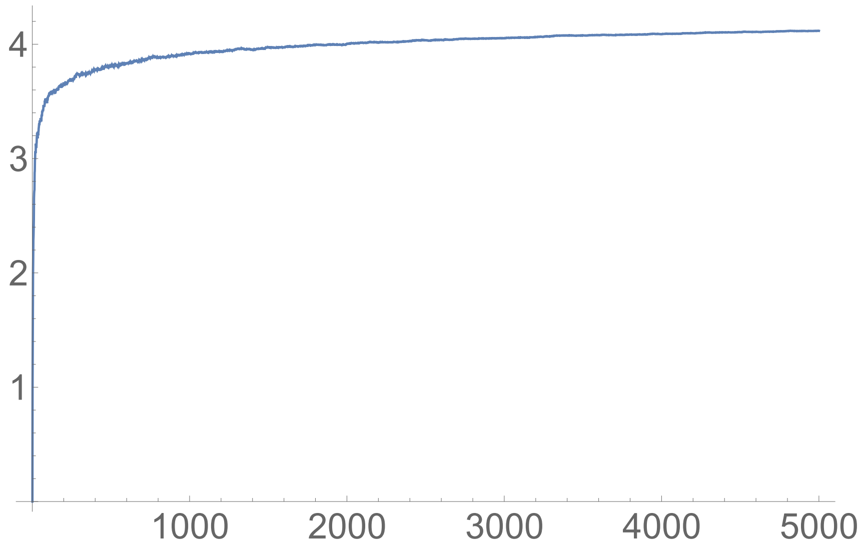

Figure 2: Graph of the mean shortest path for each new vertex up to 5000 nodes

The figure (3) shows the graph generated by the orbits series.
And figure (4), the corresponding plot of mean shortest path.

3.1 Digits analysis

The table below shows numerical results for the two visibility graphs, for each number.

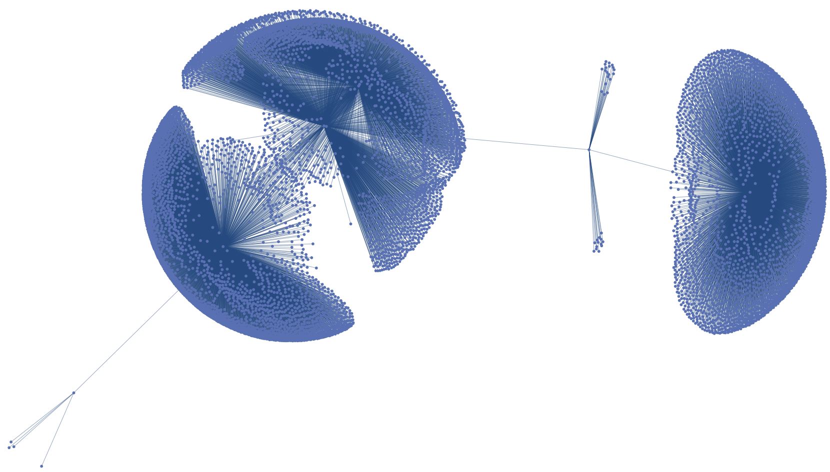

Figure 3: Graph illustrating the orbits series

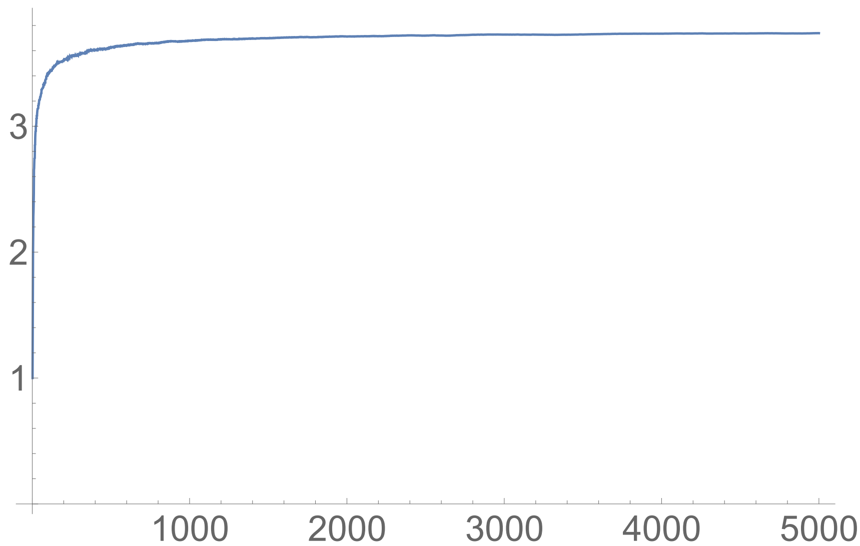

Figure 4: Graph of the mean shortest path for each new vertex up to 5000 nodes

			Natural Visibility			Horizontal Visibility		
	SDL	H/H_{max}	$\langle k\rangle$	C	γ	$\langle k\rangle$	C	γ
$\sqrt{2}$	0.194	0.948	5.19	0.696±0.134	1.338±0.163	3.63	1.334±0.256	1.734±0.201
$\sqrt{2}_2$	0.135	0.964	5.40	0.653±0.121	1.303±0.154	3.95	1.215±0.171	1.687±0.143
$\sqrt{2}_3$	0.097	0.975	5.39	0.658±0.121	1.308±0.153	3.99	1.210±0.161	1.688±0.135
ϕ	0.192	0.949	5.21	0.687±0.135	1.329±0.166	3.63	1.323±0.265	1.726±0.209
ϕ_2	0.136	0.964	5.37	0.664±0.123	1.314±0.154	3.95	1.219±0.175	1.689±0.146
ϕ_3	0.096	0.975	5.37	0.659±0.126	1.308±0.159	3.99	1.202±0.168	1.680±0.142
e	0.195	0.948	5.21	0.689±0.127	1.331±0.156	3.64	1.328±0.243	1.733±0.191
e_2	0.149	0.961	5.35	0.653±0.130	1.300±0.165	3.94	1.219±0.174	1.690±0.145
e_3	0.096	0.975	5.35	0.672±0.122	1.321±0.152	3.99	1.221±0.154	1.695±0.129
π	0.192	0.949	5.19	0.699±0.134	1.339±0.162	3.64	1.333±0.248	1.733±0.195
π_2	0.145	0.962	5.38	0.648±0.125	1.299±0.159	3.95	1.211±0.169	1.684±0.142
π_3	0.100	0.974	5.37	0.665±0.119	1.315±0.149	3.99	1.201±0.173	1.679±0.147

The histogram for each series is depicted below.

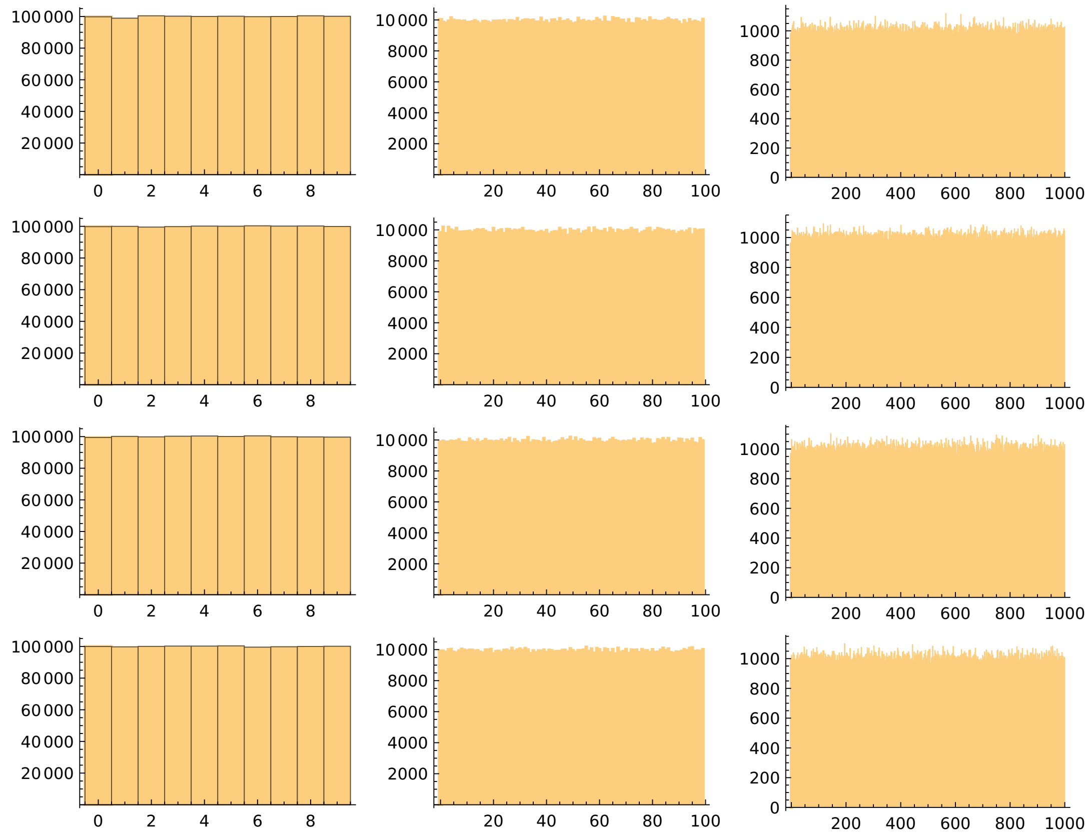

Figure 5: Histograms of digit frequency: From top to bottom: $\sqrt{2}, \phi, e, \pi$, from left to right: single, pairs and triples.

4 Discussion and Conclusion

By visual inspection the graphs are alike, they show a fractal structure and the associated γ coefficient for the orbits series is approximately half of the divisors series, but has shown a big variance, which indicates that the model may not be the best choice.

In the two studies the mean minimum path seemed to converge, the divisors series converges to 4.11735 and the orbits series to 3.73921. Using more nodes can increase the precision, in the divisor series the value seems to take longer to converge.

In both cases the number of nodes can be increased to give more precise results, also the orbit of the orbit can be studied.

A possible further study is to determine the likelihood of a node forming a link with the new node during the construction of a graph. This can lead to the values of convergence of the mean minimum path for both graphs.

4.1 Digits analysis

As all values for single digits are all very close we conclude that the two visibility algorithms do not distinguish irrational numbers from transcendental numbers.

By taking more digits to form a number the quantity of possible values increases, this reflected in slight increases of entropy and mean degree, but the SDL complexity decreased.

References

[1] OEIS Foundation Inc. (2022), The On-Line Encyclopedia of Integer Sequences, Published electronically at http://oeis.org

[2] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford, UK, 2008.

[3] H. G. Diamond, "Elementary methods in the study of the distribution of prime numbers," *Bulletin of the American Mathematical Society*, vol. 7, no. 3, pp. 553-589, 1982.

[4] U. M. Maurer, "Fast generation of prime numbers and secure public-key cryptographic parameters," *Journal of Cryptology*, vol. 8, no. 3, pp. 123-155, 1995.

[5] J. Yoshimura, T. Hayashi, Y. Tanaka, K. Tainaka, and C. Simon, "Selection for prime-number intervals in a numerical model of periodical cicada evolution," *Evolution*, vol. 63, no. 1, pp. 288-294, 2009.

[6] M. V. Berry and J. P. Keating, "The Riemann zeros and eigenvalue asymptotics," *SIAM Review*, vol. 41, no. 2, pp. 236-266, 1999.

[7] E. Bombieri and H. Davenport, "Small differences between prime numbers," *Proceedings of the Royal Society of London Series A*, vol. 293, no. 1432, pp. 1-18, 1966.

[8] S. Ares and M. Castro, "Hidden structure in the randomness of the prime number sequence?," *Physica A*, vol. 360, no. 2, pp. 285-296, 2006.

[9] G. G. Szpiro, "The gaps between the gaps: some patterns in the prime number sequence," *Physica A*, vol. 341, pp. 607-617, 2004.

[10] G. G. Szpiro, "Peaks and gaps: spectral analysis of the intervals between prime numbers," *Physica A*, vol. 384, no. 2, pp. 291-296, 2007.

[11] A. K. Chandra and S. Dasgupta, "A small world network of prime numbers," *Physica A*, vol. 357, no. 3-4, pp. 436-446, 2005.

[12] G. Corso, "Families and clustering in a natural numbers network," *Physical Review E*, vol. 69, no. 3, article ID 036106, 2004.

[13] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Chicago, USA, 1949.

[14] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, "From time series to complex networks: the visibility graph," *Proceedings of the National Academy of Sciences of the USA*, vol. 105, no. 13, pp. 4972ï¿½4975, 2008.

[15] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, "Horizontal visibility graphs: exact results for random time series," *Physical Review E*, vol. 80, no. 4, article ID 046103, 2009.

[16] K. Soundararajan, “Small gaps between prime numbers: the work of Goldston-Pintz- Yildirim,” *Bulletin of the American Mathematical Society*, vol. 44, no. 1, pp. 1-18, 2007.

[17] N. G. Sanchez, “New quantum structure of space-time,” *Gravitation and Cosmology*, vol. 25, no. 2, pp. 91ï¿½102, 2019.

[18] J. R. C. Piqueira, F. A. Serboncini, and L. H. A. Monteiro, “Biological models: measuring variability with classical and quantum information,” *Journal of Theoretical Biology*, vol. 242, no. 2, pp. 309-313, 2006.

[19] V. M. Netto, J. Meirelles, and F. L. Ribeiro, “Social interaction and the city: the effect of space on the reduction of entropy,” *Complexity*, vol. 2017, article ID 6182503, 2017.

[20] D. R. Chen, C. Liu, Y. C. Zhang, and Z. K. Zhang, “Predicting financial extremes based on weighted visual graph of major stock indices,” *Complexity*, vol. 2019, article ID 5320686, 2019.

[21] L. Wang, X. Long, J. B. A. M. Arends, and R. M. Aarts, “EEG analysis of seizure patterns using visibility graphs for detection of generalized seizures,” *Journal of Neuroscience Methods*, vol. 290, pp. 85-94, 2017.

[22] C. Gros, *Complex and Adaptive Dynamical Systems: A Primer*, Springer, New York, USA, 2008.

[23] R. Matsushita and S. da Silva, “A Power Law Governing Prime Gaps,” *Open Access Library Journal*, vol. 3, article ID e2989, 2016.

[24] J. R. C. Piqueira, L. H. A. Monteiro, T. M. C. de Magalhães, R. T. Ramos, R. B. Sassi, and E. G. Cruz, “Zipf’s law organizes a psychiatric ward,” *Journal of Theoretical Biology*, vol. 198, no. 3, pp. 439-443, 1999.

[25] B. C. Dai, F. Zhang, D. Tarzia, and K. Ahn, “Forecasting financial crashes: revisit to log-periodic power law,” *Complexity*, vol. 2018, article ID 4237471, 2018.

[26] A. D. Broido and A. Clauset, “Scale-free networks are rare,” *Nature Communications*, vol. 10, Article ID 1017, 2019.

[27] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Upper Saddle River, USA, 1998.

[28] E. F. Keller, “Revisiting “scale-free” networks,” *BioEssays*, vol. 27, no. 10, pp. 1060-1068, 2005.

[29] M. E. J. Newman, “Power laws, Pareto distributions and Zipf’s law,” *Contemporary Physics*, vol. 46, no. 5, pp. 323-351, 2005.

[30] C. M. Song, S. Havlin, and H. A. Makse, “Self-similarity of complex networks,” *Nature*, vol. 433, no. 7024, pp. 392-395, 2005.

[31] A. Nuñez, L. Lacasa, E. Valero, J. P. Gomez, and B. Luque, "Detecting series periodicity with horizontal visibility graphs," *International Journal of Bifurcation and Chaos*, vol. 22, no. 7, article ID 1250160, 2010.

[32] R. C. Baker, G. Harman, and J. Pintz, "The difference between consecutive primes. II", Proc. London Math. Soc. 83 (2001)

[33] P. ERDOS, "On The Difference Of Consecutive Primes", The Quarterly Journal of Mathematics, Volume os-6, Issue 1, 1935

[34] K. Ford, B. Green, S. Konyagin and T. Tao, "Large gaps between consecutive prime numbers", *Annals of Mathematics* vol. 195, no. 1, 2022