Precise and Concise Graphical Representation of the Natural Numbers

David W. Matula
Computer Science Department
Southern Methodist University
Dallas, USA
Email: matula@smu.edu

Zizhen Chen
Computer Science Department
Southern Methodist University
Dallas, USA
Email: zizhenc@smu.edu

***Abstract*—A graphical number representation system based on the formal logic foundation of a recursively defined function is presented. The function exposes a one-to-one correspondence between the natural numbers and the set of rooted trees. Secondly, a relation between pairs of natural numbers is shown to partition the integers into disjoint finite sets each visualized as a tree in the forest of all finite trees. This relation encapsulates in visual form the natural structure and distribution of primes in a manner not previously investigated.**

These representations allow arithmeticians to visually experience natural numbers in a 2D graphical layout. This form is free of cultural choices such as the number of symbols in an alphabet and/or one-dimensional symbol strings. The fundamental nature of these mappings is argued to be accessible to a creative mind independent of formal schooling. This short note includes examples illustrating the simplicity of the constructions.

***Keywords*-graphical representation; prime factorization; prime number function; recursive function; number fonts**

I. Introduction and Summary

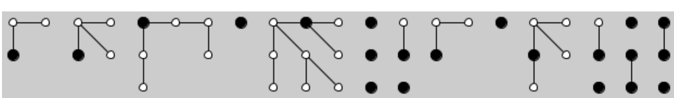


Figure 1: *A graphic is worth a thousand digits.*

The main purpose of this paper is to initiate an investigation of graphic fonts for the natural numbers that are easy to decipher from fundamental principles of arithmetic. For this purpose we introduce a graphic number construction paradigm founded on properties of prime numbers and the prime number function. Our number symbol fonts are given by trees that can grow in two dimensions. They provide a countable alphabet of graphic symbols, with size and form of each symbol reflecting both the size and the multiplicative structure of the number represented.

The current visual confinement of numeric data to uninspiring chains of culturally determined digit choices has strangled the life out of numbers and arithmetic. Until current times this situation was at the behest of mechanical and early electronic devices. Current architectural advances in inputting, processing, and presenting graphically represented data along with the emerging power of quantum computing has released these digital representation constraints.

Can insight into "big data" be better gleaned from a mural or 3D printed sculptural representation? Will this extend to arithmeticians the inspirational capability to realize operations in science equivalent to those experiences captured and preserved by artists in carvings and paintings throughout cultures over millennia.

Primes are immediately observed as those linear sequences of marks that cannot be reduced to a 2D rectangular layout. They have been discovered and themselves counted across cultures by those without schooling. This observation motivates our major result in Section II. This is to introduce a graphic representation of natural numbers (counts 1, 2, ...) utilizing a prime number inspired one-to-one mapping between natural numbers and rooted trees. We emphasize that this culture-free system for displaying counts is founded on three *fundamental principles of arithmetic*, specifically a theorem (prime factorization), an operation (counting), and the computational process of recursion. The intent of a graphic system of counts is to be concise, for both presentation and storage with precise completeness of range.

Sequences of these self-delimiting graphic numbers provide the basis of a second major result in Section II. They allow a finite graphic representation of each ratio of counts, complete over the rationals. Furthermore they also can represent well-known accurate approximations to fundamental geometric and scientific constants, such as the graphic for a 21 partial quotient approximation of π initiating this Section.

The main result of Section III is the introduction of the concept of adjacency of integers based simply on the prime number function. The concept is the basis for an equivalence relation partitioning the integers into finite sets, each corresponding to a non isomorphic tree of the forest of all finite trees. Each tree is shown to encapsulate the integers of the rooted trees isomorphic to it absent the specification of a root location. The potential for application of these graphics and useful properties now unseen are hopefully suggested by a now well-known but previously unseen property of primes in the foundation of RSA cryptography.

Section IV provides conclusions and some challenges for further study.

II. Graphic Number Representation Systems

We utilize the notation p_k for the kth prime, with k the index of the prime. Let T^* denote the set of all finite rooted trees and Z^+ the natural numbers. A one-to-one mapping between T^* and Z^+ was introduced in 1968 [1] and rediscovered twice in the 1980's [2], [3]. The tree numbers were termed "Matula numbers" in [4] with applications there and in later papers to chemical nomenclature. Properties of the mapping have been studied in [2], [3], [5] and in websites. A proof is included here for completeness.

Theorem 1: There exists a one-to-one mapping $\mu : T^* \to Z^+$ with the following properties. For each rooted tree there is a prime label on each edge directed towards the root where:

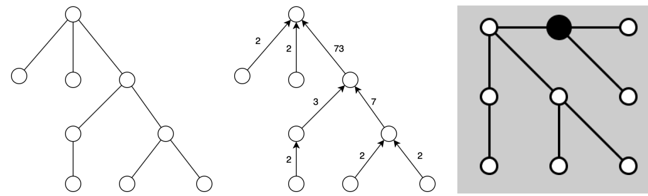
(i) the directed edge out of each leaf is labeled with the prime $p_1 = 2$,
(ii) an edge out of a node directed towards the root is labeled with the prime p_k where the index k is the product of the primes labeling the edges directed into the node,
(iii) the root of the single node rooted tree is labeled one,
(iv) the root of any rooted tree having at least two nodes is labeled with the integer n, where n is the product of all primes on the edges directed into the root. The root label value n then defines the mapping $\mu(t) = n$ for $t \in T^*$.

Furthermore, given an integer $n \in Z^+$, the labeling can be determined from root to leaves yielding the inverse mapping $\mu^{-1}(n) = t \in T^*$ with $\mu(\mu^{-1}(n)) = n$.

Proof: The directed edge and root labeling may be viewed constructively with reference to Figure 2. Assuming the rooted tree has depth $j \geqslant 1$, all edges from depth j are labeled with $p_1 = 2$ as required for step (i). Then the labels on all edges from depth j to depth $j-1$ may be assigned satisfying steps (i) and (ii). When all edges directed into any node at depth one have been labeled, then by steps (i) and (ii) all edges into the root may be labeled with primes and their product by step (iv) determining the root label and the value of $\mu(t) \in Z^+$ for any $t \in T^*$.

For the inverse, determining $\mu^{-1}(n) \in T^*$ for any natural number n, the tree $\mu^{-1}(n)$ has edges from the root labeled with the prime factors of n as determined by unique prime factorization. For each prime factor p_k, the edge from the root leads to the subtree $\mu^{-1}(k)$. The procedure is continued recursively with $\mu^{-1}(1)$ denoting a single node or a leaf of the tree. ■

Figure 2a shows a rooted tree $t \in T^*$, and 2b illustrates the corresponding directed edge prime number labels and determination of the root value $\mu(t) = 292$. For the inverse starting with $n = 292$, then $\mu^{-1}(292) = \mu^{-1}(2^2 \times 72) = \mu^{-1}(p_1^2 \times p_{21})$. So the root node of $\mu^{-1}(292)$ leads to three subtrees, two being leaves, and the third the subtree $\mu^{-1}(21)$. Recursion leads to all the leaves, as also seen in Figure 2b.



(a) traditional root at top node (b) labeled edges directed towards root (c) 3×3 grid format with root highlighted

Figure 2: Three formats of a rooted tree

Figure 2c introduces a square grid font for $n = 292$ for comparison with existing decimal digit fonts. The root is highlighted in the font.

This numbering scheme for rooted trees is arguably fundamental as it comprises three fundamental principles of arithmetic,

(i) a fundamental theorem: unique prime factorization is known as the Fundamental Theorem of Arithmetic,
(ii) a fundamental operation: counting is a fundamental operation of arithmetic,
(iii) a fundamental process: recursion is a fundamental computational procedure, with recursive functions a building block of logic.

For later reference it is useful to note certain derivative portions of the result of Theorem 1.

Definition 1: A prime branch of a rooted tree is a maximal subgraph containing a single child of the root.

A prime branch is by itself a rooted tree where the root is then also a leaf. By ordering the prime branches of the root, we obtain:

Observation 1: Given $n_1 = \mu(t_1)$, $n_2 = \mu(t_2)$, the integer product $n_1 n_2 = \mu(t_3)$ has t_3 determined by merging all prime branches of t_1 and t_2.

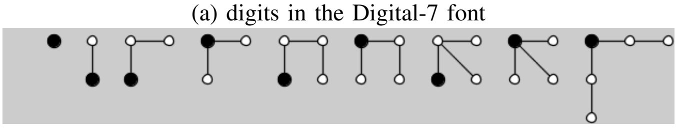
Observation 2: The determination of whether $\mu(t_1)$ exactly divides $\mu(t_2)$, and if so the resulting quotient $\mu(t_3)$, is readily determined from the prime branches of t_1 and t_2 yielding the prime branches of t_3.

Observation 3: For a set $\{t_1, t_2, ..., t_i\} \subset T^*\}, i \geqslant 2$, the GCD and LCM of the set $\{\mu(t_1), \mu(t_2), ..., \mu(t_i)\}$ is readily determined.

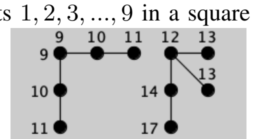
While these multiplicative properties of the mapping μ are potentially useful for application, that is not our main purpose here. That purpose is to utilize the mapping μ to facilitate the construction of a "natural" visual format for the "natural" numbers.

The rooted tree for $\mu^{-1}(n)$ can be drawn in many styles of font. We utilize here a square grid format such as shown in Figure 2c in order to compare size and style with that of the well-known Digital-7 integer font which is shown in Figure 3a. The rooted trees for natural numbers $1, 2, ..., 9$

(a) digits in the Digital-7 font



(b) counts $1, 2, 3, ..., 9$ in a square grid font



(c) compressed square grid font for selecting counts 9-14 and 17

Figure 3: Three fonts for small integers in a grid

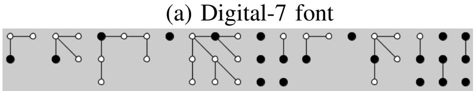
are shown in Figure 3b in a similarly styled 2×2 square grid for numbers $2, 3, 4, 5, 6, 7, 8$ and in a 3×3 square grid for numbers 9.

Note that the edge choices for 3 and 4 are the same, as are the edge choices for the pair 5, 6 and the pair 7, 8 with the only distinction being the highlighted nodes designating the root. This allows for compressing to just six the number of distinct edge formats for the nine distinct values using root node highlighting. The compression continues for the 3×3 square grid for the triple 9, 10, 11 and the consecutive triple 12, 13, 14, as well as 17 for Figure 3c. Note that Figure 2c had utilized this 3×3 square grid with the root node highlighted for the natural number 292.

This representation is concise with the size of the square grid suggesting the size of the number. The corresponding prime branches are available for potential applications.

Importantly, the visual forms for the counting numbers can be extended to a unique and concise numeric representation of the rationals. Recall that partial quotients of a continued fraction are all natural numbers. Since each partial quotient can effectively have its own form as a rooted tree, there is no need for delimiters as is needed when each number is represented by a digit sequence. Figure 4 shows the 21

(a) Digital-7 font



(b) square grid font

Figure 4: Fonts for a 21 partial quotient approximation of π

partial quotient "best rational approximation" of π utilizing the Digital-7 format for the total of 25 decimal digits in the format. The 21 partial quotients require 20 delimiter marks, yielding a total of about 60 columns to represent this rational. For comparison, the same representation of this 21 partial quotient representation takes only 21 columns in the square grid partial quotient sequence given in Figure 4. That 21 column format employs some transparent compression that has saved 8 columns, but even spread over 29 columns the compression is about 50%.

The main aspects we emphasize in this representation are,

(i) there is no need for delimiters,
(ii) the size of the characters in two dimensions gives a strong visual feel to the implicit series of better approximations of π,
(iii) stopping before a large character such as that for 292 provides a very good approximation, measuring the size of the character as the effort needed to only slightly improve the accuracy of the approximation.

The large frequency of $1's$ and $2's$ in the continued fraction's initial partial quotients of π is replicated in many expansions. For a random number picked uniformly over the interval $[0, 1)$, the frequency of small digits has been studied [6]. In particular $1's$ occur about 41.5% of the time and $2's$ about 17.0%, both of whose representations fit in a single column. The frequency of partial quotients in the range one to eight is over 80% in total, and all of these integers are represented requiring no more than the 2×2 square grid. The occurrence of the particular partial quotient 292 occurring in $\pi's$ list has in the random over $[0, 1)$ model a frequency of less than 0.002%. The partial quotient 84 which is the next $(22nd)$ in the continued fraction for π after those shown in Figure 4 has a frequency of only 0.02%. Utilizing these frequencies note that a single column in our square grid representation is sufficient for over 58% of the partial quotients, and another 22% fit in the 2×2 square grid.

This suggests only about 1.6 columns per partial quotient in the square grid number representation system. In contrast the Digital-7 format along with the delimiters takes at least 3 columns for over 58% of the partial quotients, and at least 4 columns for about 14% of the total number of partial quotients. This suggests over 2.6 columns per partial quotient for the Digital-7 decimal digit string representation which is over 50% greater.

III. Rooted Tree Sets and Integer Adjacency

In Section II we utilized alternative placement of the root node in a tree to reduce the number of tree formats for visualizing the natural numbers. In this manner a subset of Z^+ can be associated with each tree. Consider the graph theoretic concept of the forest of all non-isomorphic trees. This set of trees is denoted by T. Classified by size there are only single members of T for trees of sizes 1, 2 and 3, two members for trees of size 4, and three of size 5. Their numbers grow rapidly with six of size 6, eleven of size 7, twenty three of size 8, forty seven of size 9 and over a hundred of size 10.

Many nodes of these trees yield the same rooted tree number due to automorphisms of each tree. The sets of rooted tree numbers associated with the trees form a disjoint partition of the natural numbers with tree labeling of edges and nodes indicating the structure as follower.

Theorem 2: Every tree $t \in T$ can have its nodes and edges (directed in both directions) labeled with the labels having the following properties:

(i) every directed edge from node u to node v is labeled with the prime corresponding to the prime branch from u into root node v,

(ii) every node is labeled with the product of the primes on the edges directed into the node.

Proof: The result follows by considering all prime branches determined for each root location in tree t. ■

The integer sets corresponding to the trees of T are said to yield a *tree set* partition of Z^+. Each tree may be uniquely identified with the minimum root node label for the tree, allowing $\alpha : T \rightarrow Z^+$ to be defined with $\alpha(t) = 20$ for the tree of Figure 5. Given the rooted tree labeling of

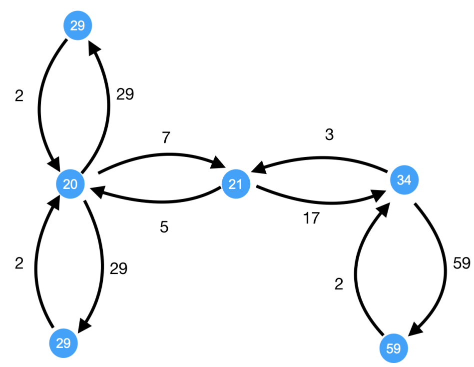


Figure 5: The tree for tree set $\{20, 21, 29, 34, 59\}$

a designated root node in the tree as determined by the procedure in the proof of Theorem 1, it is evident that each edge directed out of that root node may be labeled by a prime p_k with k a product readily determined from the appropriate labeled edges directed into that root node. Recursively the rest of the labeling can be completed with an n node tree requiring no more than $2n - 2$ calls to an oracle for prime number function evaluation.

There is an interesting alternative approach to obtaining the same partition of Z^+. The following definition of *integer adjacency* is based solely on the prime number function.

Definition 2: For every pair $i, j \in Z^+$, the integer $i \times p_j$ is adjacent to $j \times p_i$.

It is not difficult to show that i adjacent to j implies j is adjacent to i. This adjacency relation can readily be extended to an equivalence relation termed *integer connectivity*. The resulting equivalence relation partition of Z^+ under this integer connectivity relation is readily shown to be identical to the tree set natural number partition.

Proofs of these results and further properties are beyond the scope of this paper. They present attractive avenues for further development of special arithmetic functions.

IV. Conclusions and Future Directions

We have introduced a graphical number representation system based on several fundamental arithmetic and computational properties of primes. The system foundations avoid the existing cultural and machine dependent choices of base and digit set, as well as the limitation to one dimensional strings of characters from a finite alphabet. The graphic number system has a larger two dimensional footprint as needed for large natural numbers which are presented in a visually insightful manner. Evidence was included of the conciseness of format compared to traditional digit fonts. As ”big data” bit strings and their access indices grow ever larger, 2D graphic font representations provide a potential alternative index structure to reduce storage size maintaining unique precise values.

Selected figures laying out our graphic number representations are given both to teach and exhibit properties of the system. The facility of unique sequences of these graphic symbols to represent any rational and to support “best rational approximation” of reals provides universality to storing the results of standard arithmetic operations. Comparisons of distinct rational approximation of reals is facilitated by unique partial quotient initial sequences being identical until two partial quotients differ. A large partial quotient indicates the sequence up to and not including that quotient is relatively very accurate. This provides both concise and precise approximation of reals.

The structure of RSA cryptography was gleaned from a separability property of primes. We speculate that intriguing properties of primes and their distribution seamlessly embedded in the graphic number representations will provide fertile ground to nurture new special arithmetic functions. There are many directions here for further research.

Acknowledgment

The authors would like to thank Brittni Watkins for contributing to the grid font designs in the Figures.

References

[1] Matula, D. W. ”A natural rooted tree enumeration by prime factorization.” SIAM Review. Vol. 10: 273, 1968.

[2] Göbel, F. ”On a 1-1-correspondence between rooted trees and natural numbers.” Journal of Combinatorial Theory, Series B 29.1 (1980): 141-143.

[3] Cappello, P. ”A new bijection between natural numbers and rooted trees.” 4th SIAM Conf. on Discrete Mathematics. 1988.

[4] Gutman, Ivan, Aleksandar Ivic, and S. B. Elk. ”Matula numbers for coding chemical structures and some of their properties.” Journal Serbian Chemical Society 58 (1993): 193-193.

[5] Gutman, Ivan, and Aleksandar Ivić. ”On Matula numbers.” Discrete Mathematics 150.1-3 (1996): 131-142.

[6] Knuth, Donald. ”The art of computer programming, 2 (Seminumerical Algorithms).” Vol. 2. 2nd Edition, 1981, p.352.