Comment on "Families and clustering in a natural numbers network"

Jeffrey D. Achter*
Colorado State University, Fort Collins, Colorado 80523, USA
(Received 23 April 2004; published 29 November 2004)

Corso [Phys. Rev. E **69** 036106 (2004)] constructs a family of graphs from subsets of the natural numbers, and numerically estimates diameter, degree and clustering. We give exact asymptotic formulas for these quantities, and thereby argue that number theory is a more appropriate tool than simulation.

DOI: 10.1103/PhysRevE.70.058103 PACS number(s): 89.75.Fb, 02.10.De

In [1], the author examines an infinite class of finite graphs constructed from the natural numbers. Small-world networks—that is, those which are simultaneously of low connectivity, small distance, and high clustering coefficient [2]—arise in stunningly diverse contexts. Given this, one would like to estimate degree distributions and related data for natural families of graphs.

For a given natural number X, [1] constructs a graph $M = M(X)$. The vertices are the natural numbers $2,\ldots,X$, and vertices (corresponding to) m and n are connected exactly if they share a nontrivial divisor. Using a combination of numerical experiments and heuristic arguments, Ref. [1] advances claims about interesting quantities associated to M such as its number of connected components, average degree, maximal degree, and average intervertex distance. While the combinatorial arguments and simulation results are sound, the behavior of these quantities as X grows is quite difficult to estimate numerically. Just as it is hard to verify with direct computation that the harmonic series diverges, the experimental results in [1] misrepresent the asymptotic behavior of the graphs $M(X)$. In this note, we use fundamental results in analytic number theory [3] to retell the story. Specifically, we will obtain precise values for quantities estimated in Table 1 of [1], and show that the conclusions deduced from Fig. 3 are incorrect. Broadly, our message is that by availing ourselves of centuries of research in number theory, we can gain insight into the matters at hand.

A classical result ([3], I.3.4) says that the chance two numbers are relatively prime—that is, the associated vertices in M *do not* share an edge—is $\sim 6/\pi^2$. [Here, and in the rest of the paper, we use notation $f(X) \sim g(X)$ if $f(X) = g(X) + o(X)$, where the $o(X)$ term is a known, suppressed error.) Therefore, the number of edges in M is $\sim (1-6/\pi^2)\binom{X-1}{2}$. Corso estimates ([1], Table 1) through simulation that the average degree of a vertex in $M(X)$ is 0.45. We immediately see that as X gets ever larger, the average vertex degree approaches $1-6/\pi^2 \approx 0.39$. Already, classical analytic number theory allows us to compute precisely a quantity difficult to apprehend through simulation.

We similarly investigate the structure of the connected components of $M(X)$. On one hand, we will see that the set of singletons in $M(X)$, while growing with X, accounts for a vanishingly small proportion of the vertices of $M(X)$. On the other hand, we will construct a subgraph $P(X)$ of $M(X)$ which contains almost all vertices of $M(X)$, in the sense that $\lim_{X\to\infty}|V(P(X))|/|V(M(X))| = 1$, where $V(G)$ denotes the vertex set of the graph G. Therefore, we can compute average properties of $M(X)$ by restricting to $P(X)$.

Let $S(X)$ be the set of singletons in $M(X)$, that is, vertices with no neighbors in $M(X)$. On one hand, $|S(X)|$ is expected to be unbounded as X goes to infinity; on the other hand, $\#S(X)/X$ vanishes as X gets larger. Indeed, following standard notation let $\pi(X)$ denote the number of primes less than X. The celebrated prime number theorem states that $\pi(X) \sim X/\ln X$, again with explicit control over the error term. Now, as [1] observes (although with an unfortunate mistake in the inequality as written), a number m is a singleton in $M(X)$ if and only if m is a prime such that $2m > X > m$. Therefore, the number of singletons is $\sim \pi(X) - \pi(X/2)$. We see that $\lim_{X\to\infty}\#S(X) = \infty$, while $\lim_{X\to\infty}\#S(X)/X = 0$. It is at best difficult to see this from numerical examples with small values of X, and this is typical of the difficulties in [1]. Average quantities, such as average vertex degree, are asymptotically insensitive to the presence of singletons.

We can also obtain and exploit a lower bound on the maximal degree of a vertex in $M(X)$. A theorem of Hanson ([3], I.1.2) states that

$$\prod_{p<X} p < 3^X.$$

Armed with this, let $Y = Y(X) = \log_3(X)$, and set $n = n(X) = \Pi_{p<Y} p$, so that n is smaller than X. The number of vertices in $M(X)$ which are not connected to n is ([3], III.6.2)

$$\sim \alpha(X) \overset{\text{def}}{=} \frac{Xe^{-\gamma} - Y}{\ln Y},$$

where γ is Euler's constant. Since $\lim_{X\to\infty}\alpha(X)/X = 0$, the degree of $n(X)$ approaches the number X as X gets large.

Bearing this in mind, we consider Fig. 3 of [1] and the commentary in the last paragraph of ([1], IIA). In Fig. 3, vertices are ranked by degree so that $\deg(v_i) \geqslant \deg(v_{i+1})$, and $\deg(v_i)/X$ is plotted as a function f_X of i/X. The text of the paper seems to assert that $f_X(0)$ is bounded above by 0.8, and that $f_X(x)$ is zero for $x > 0.9$. This is *false*. In fact, since $\lim_{X\to\infty}\deg n(X) = X$, $\lim_{X\to\infty} f_X(0) = 1$. Moreover, since

*Electronic address: j.achter@colostate.edu
URL: http://lamar.colostate.edu/~jachter

singletons form a vanishingly small proportion of vertices in $M(X)$, the measure of the interval on which f_X is zero approaches zero. The only other facts we can readily establish are that f_X is nonincreasing (by construction), and that the area under f_X on [0, 1] approaches $1-6/\pi^2$.

Let $P(X)$ be the subgraph of $M(X)$ obtained by considering only those vertices connected to $n(X)$; $P(X)$ contains all but a vanishing small proportion of the vertices (and, thus, edges) of $M(X)$. By construction, the diameter of $P(X)$ is 2, as any pair of vertices may be connected via $n(X)$.

Moreover, we can obtain an upper bound for the average distance $d(P(X))$, defined as the average distance between all pairs of vertices in $P(X)$. [The paper [1] claims to simulate the distance of $M(X)$, but it is unclear how to define the average intervertex distance of a disconnected graph. Therefore, we restrict ourselves to the largest connected component.] As noted above, the chance that any two vertices share an edge is $1-6/\pi^2$. Given that $u,v\in P(X)$, the chance that they share an edge is *at least* $1-6/\pi^2$. Since the maximum distance between vertices in $P(X)$ is 2, the average distance between vertices of $P(X)$ is at most $(1-6/\pi^2)\times 1+(6/\pi^2)\times 2=1+6/\pi^2$. We can thus invoke the logic of ([1], III)—bearing in mind that in a random, nonsparse graph the distance is 2—and conclude that $M(X)$ does not behave like a random graph.

We close this discussion with a heuristic for clustering in $M(X)$. The clustering coefficient of a vertex in a graph is the proportion of pairs of neighbors of that vertex which are themselves directly connected. For want of a direct estimate of the average clustering coefficient—although the argument in the previous paragraph also indicates that $M(X)$ is highly clustered—we will show that, for a fixed vertex v, as $X\to\infty$ the clustering coefficient of v approaches unity.

Recall that the radical of a natural number is the product of all primes which divide that number. If two numbers m and m' share the same radical, then they are indistinguishable in M in the following sense. For all $X>\max(m,m')$ and all $n\in M(X)$, n is connected to m if and only if n is connected to m'. Every class of mutually indistinguishable vertices has the same radical. Since radicals are square-free, and since the number of square-free numbers less than X is $\sim(6/\pi^2)X$, we see that such classes tend to be small relative to X. We now restrict our attention to square-free numbers, hoping that in doing so we capture the asymptotic behavior of all of $M(X)$.

Thus, let n be any square-free number. We have already seen that any two vertices, and in particular any two vertices connected to n, have a positive chance of being joined by an edge. It turns out that their shared connection to n makes them even more likely to share an edge. If n is a prime number, then any two neighbors of n are connected, and thus the clustering coefficient is one. For square-free n which is not prime, let m and m' be two vertices connected to n. Then the chance that $\gcd(\gcd(m,m'),n)$ is nontrivial, given that both $\gcd(m,n)$ and $\gcd(m',n)$ are nontrivial, approaches

$$\beta(n)\overset{\text{def}}{=}1-\frac{\varphi(n)}{n(n-\varphi(n))}\frac{\sigma(n)-\tau(n)}{n}$$

for $X\gg n$. Here, φ is Euler's totient function, which counts the number of natural numbers less than n which are relatively prime to n; σ is the sum of all divisors of n; and τ is the number of divisors of n. We interpret this as the probability that, solely by virtue of being connected to n, two neighbors of n share an edge. Let $\beta^*(n)$ be 1 if n is a power of a prime, and $\beta(n)$ otherwise. As X gets large, the average value of β^* approaches 1. To the extent that statistics of $M(X)$ track those of the square-free vertices, we expect that the average clustering coefficient of $M(X)$ approaches unity.

Corso also considers a variant $M_\ell(X)$, in which m and n are connected if and only if some prime larger than ℓ divides both m and n. We expect that the same sort of sieve results should apply, and thus that $M_\ell(X)$ and $M(X)$ have the same asymptotic behavior. This will be difficult to verify by computer, however, since a large ℓ will significantly retard the approach to stationarity in X. Also, the constants which appear will be different; in particular, the average connectivity $1-6/\pi^2$ is surely a different (nonzero) value in $M_\ell(X)$.

Theoretical analysis shows more directly than numerical simulation that $M(X)$ is a dense, clustered graph whose largest component has small diameter; we leave it to the reader to decide for herself to what extent it behaves like a small-world network. Corso [1] identifies $M(X)$ as "a promising laboratory in the study of degree distribution and cluster families." Hopefully, the asymptotic calculations here can guide those studies.

[1] G. Corso, Phys. Rev. E **69**, 036106 (2004).
[2] D. Watts and S. Strogatz, Nature (London) **393**, 440 (1998).
[3] G. Tennenbaum, *Introduction to Analytica and Probabilistic Number Theory* (Cambridge University Press, Cambridge, England, 1995).