A RANDOM GRAPH

The construction. A random (undirected) graph with n vertices is constructed in the following way: pairs of vertices are selected one at a time in such a way that each pair has the same probability of being selected on any given occasion, and, each selection is made independently of previous selections. If the vertex pair $\{x, y\}$ is selected, then an edge is constructed which connects x and y.

Are multiple edges possible? In my model, *yes!* For example, if the vertex pair $\{x, y\}$ were to be selected k times, there would be k edges connecting x and y: a *multiple edge* contributing $\binom{k}{2}$ *cycles of length 2*.

ASYMPTOTIC BEHAVIOUR

Suppose that m edges have been selected. We shall be concerned with the behaviour of the graph in the limit as n and m become large, but in such a way that $m = O(n)$.

The problem. Our problem is to determine the limiting probability that the graph is acyclic.

Motivation. Havas and Majewski* present an algorithm for *minimal perfect hashing* (used for memory-efficient storage and fast retrieval of items from static sets) based on this random graph. Their algorithm is optimal when the graph is acyclic.

*[HM] Havas, G. and Majewski, B.S. (1992), *Optimal algorithms for minimal perfect hashing*, Technical Report No. 234, Key Centre for Software Technology, Department of Computer Science, The University of Queensland.

WHY ACYCLIC?

Consider a set W of m words (or keys). Every *bijection* $h : W \to I$, where $I = \{0, \ldots, m-1\}$, is called a *minimal perfect hash function.* HM find hash functions of the form

$$h(w) = (g(f_1(w)) + g(f_2(w))) \bmod m\,;$$

f_1, f_2 map keys to integers (they identify the pair of vertices of the graph corresponding to the edge w) and g maps integers to I.

Given f_1 and f_2, can g be chosen so that h is a bijection?

If the graph is acyclic then, yes, it is easy to construct g from h. Traverse the graph: if vertex w is reached from vertex u then set

$$g(w) = (h(e) - g(u)) \bmod m\,,$$

where $e = (u, w)$.

EFFICIENCY

HM's algorithm generates f_1 and f_2 at random until an acyclic graph is found:

$$f_k(w) = \left(\sum_{i=1}^{|w|} T_k(i) w[i] \right) \bmod m \,,$$

where T_1 and T_2 are tables of random integers and $w[i]$ denotes the i-th character (an integer) of key i.

The efficiency of the algorithm is determined by the probability $p^{(n)}$ that the graph is acyclic: the expected number of iterations needed to find an acyclic graph will be $1/p^{(n)}$ (typically between 2 and 3).

EVALUATING $p^{(n)}$

Theorem. If n and m tend to ∞ in such a way that $m \sim cn$, where c is a positive constant, the limiting probability p that the graph is acyclic is given by

$$p = \begin{cases} e^c\sqrt{1-2c} & \text{if } 0 < c < 1/2 \\ \\ 0 & \text{if } c \geq 1/2 \end{cases}$$

Proof. On request. It uses results from [HM] and Erdös and Renyi*.

*[ER] Erdös, P. and Renyi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–61.

SKETCH PROOF

Let $X_k^{(n)}$ be the number of cycles of length k and let $p_k^{(n)} = \Pr(X_k^{(n)} = 0)$. Following [HM] write

$$p^{(n)} = \prod_{k=2}^{\infty} p_k^{(n)}, \qquad n = 2, 3, \ldots .$$

Now let $q_k^{(n)} = -\log p_k^{(n)}$, so that $0 \leq q_k^{(n)} < \infty$ and

$$p^{(n)} = \exp\left(-\sum_{k=2}^{\infty} q_k^{(n)}\right), \quad n = 2, 3, \ldots .$$

ER show that the distribution of $X_k^{(n)}$ is asymptotically Poisson: in particular,

$$\lim_{n\to\infty} p_k^{(n)} = e^{-\lambda_k}, \text{ where } \lambda_k = (2c)^k/2k.$$

It follows that

$$\lim_{n\to\infty} q_k^{(n)} = -\log\left(\lim_{n\to\infty} p_k^{(n)}\right) = \lambda_k\,.$$

So, formally,

$$\lim_{n\to\infty} \sum_{k=2}^{\infty} q_k^{(n)} = \sum_{k=2}^{\infty} \lim_{n\to\infty} q_k^{(n)} = \sum_{k=2}^{\infty} \lambda_k\,,$$

and hence

$$\lim_{n\to\infty} p^{(n)} = e^{-\lambda},\ \text{where } \lambda = \sum_{k=2}^{\infty} \lambda_k\,. \qquad (2)$$

By Fatou's Lemma, we always have

$$\liminf_{n\to\infty} \sum_{k=2}^{\infty} q_k^{(n)} \geq \sum_{k=2}^{\infty} \liminf_{n\to\infty} q_k^{(n)} = \sum_{k=2}^{\infty} \lambda_k\,,$$

from which it follows immediately that

$$\limsup_{n\to\infty} p^{(n)} \leq e^{-\lambda}\,;$$

this argument is valid even if the sum in (2) is divergent. We deduce immediately that if $c \geq 1/2$, $p^{(n)} \to 0$.

When $c < 1/2$, we have $0 < \lambda_k < 1$ and

$$\lambda = \sum_{k=2}^{\infty} \lambda_k = -c + \frac{1}{2} \ln \left(\frac{1}{1-2c} \right) .$$

From Markov's inequality we have $\Pr(X_k^{(n)} \geq 1) \leq \mathsf{E} X_k^{(n)}$ and so $p_k^{(n)} = \Pr(X_k^{(n)} = 0) \geq 1 - \mathsf{E} X_k^{(n)}$. By Lemma 2 of [HM], we have, for each fixed $k \geq 2$, that $\mathsf{E} X_k^{(n)} \uparrow \lambda_k$ as $n \to \infty$. In particular, for each $k \geq 2$, the sequence $\{\mathsf{E} X_k^{(n)}\}$ is bounded above by λ_k. It follows that $\{q_k^{(n)}\}$ is bounded above by $d_k := -\log(1 - \lambda_k)$. Further, since $\lambda_k < 1$,

$$\sum_{k=2}^{\infty} d_k = -\log \left(\prod_{k=2}^{\infty} (1 - \lambda_k) \right) < \infty .$$

Thus, by Dominated Convergence, we have

$$\lim_{n \to \infty} \sum_{k=2}^{\infty} q_k^{(n)} = \sum_{k=2}^{\infty} \lim_{n \to \infty} q_k^{(n)} = \lambda ,$$

and, hence, $p^{(n)} \to e^{-\lambda}$.

THE FIVE STAGES OF EVOLUTION

PRIMORDIAL STEW: $m(n) = o(n)$

If $m(n)/n \to 0$, then (with limiting probability 1) *all components are trees*.

Trees of order k appear when m reaches order $n^{(k-2)/(k-1)}$. In particular, T_k, the number of trees of order k, has a (limiting) Poisson distribution with mean $\lambda_k = (2\rho)^{k-1} k^{k-2}/k!$, where

$$\rho = \lim_{n\to\infty} m(n) n^{(k-1)/(k-2)}.$$

Finally, if $m(n)n^{(k-1)/(k-2)} \to \infty$, the number of trees of order k is asymptotically normally distributed with mean and variance equal to

$$\mu_n = n\frac{k^{k-2}}{k!}\left(\frac{2m(n)}{n}\right)^{k-1} e^{-2km(n)/n}.$$

To be precise, $(T_k - \mu_n)/\sqrt{\mu_n} \Rightarrow N(0,1)$. This result holds in the next two stages of evolution; we only require $\mu_n \to \infty$.

SPOOKY: $m(n) \sim cn$, where $0 < c < 1/2$

Cycles of all orders start to appear: C_k, the number of cycles of order k, has a (limiting) Poisson distribution with mean $\lambda_k = (2c)^k/(2k)$.

Furthermore, with limiting probability 1, all components are either trees or consist of exactly one cycle (k vertices and k edges), the latter having a Poisson distribution with mean

$$\lambda_k = \frac{(2ce^{-2c})^k}{k!} \sum_{i=0}^{k-3} \frac{k^i}{i!},$$

where k is the order of the cycle.

The largest component is a tree; it has

$$\frac{1}{2c - 1 - \log 2c} \left(\log n - \frac{5}{2} \log \log n \right)$$

vertices (with probability tending to 1).

A MONSTER APPEARS:

$m(n) \sim cn$, where $c \geq 1/2$

When $m(n) \sim n$ ($c = 1/2$), the largest component has (with probability tending to 1) $n^{2/3}$ vertices. When $m(n) \sim cn$ with $c > 1/2$, *a giant component appears:* the largest component in the graph has $G(c)n$ vertices, where $G(c) = 1 - X(c)/2c$ and

$$X(c) = \sum_{i=1}^{\infty} \frac{i^{i-1}}{i!}(2ce^{-2c})^i .$$

Note that $G(1/2) = 0$ and $G(c) \to 1$ as $c \to \infty$.

Almost all the other vertices belong to trees: the total number of vertices belonging to trees is almost surely $n(1 - G(c)) + o(n)$.

For $c > 1/2$, the expected number of components in the graph is asymptotically

$$\frac{n}{2c}\left(X(c) - \frac{1}{2}X^2(c)\right) .$$

CONNECTEDNESS:

$m(n) \sim cn \log n$, **where** $0 < c \leq 1/2$

The graph is becoming connected: if

$$m(n) = \frac{n}{2k} \log n + \frac{k-1}{2k} n \log\log n + \alpha n + o(n),$$

then (with probability tending to 1) there are only trees of order $\leq k$ outside the giant component, the limiting distribution of the number of trees of order l being Poisson with mean $e^{-2\alpha l}/(l.l!)$. For example ($k = 1$), if

$$m(n) = \frac{n}{2} \log n + \alpha n + o(n),$$

there are (almost surely) only isolated vertices outside the giant component, the number of these having a limiting Poisson distribution with mean $e^{-2\alpha}$. And, the chance that the graph is indeed connected tends to $\exp(-e^{-2\alpha})$ (which itself tends to 1 as α grows).

ASYMPTOTIC REGULARITY:

$m(n) \sim \omega(n) n \log n$, **where** $\omega(n) \to \infty$

The whole graph becomes regular: with probability tending to 1, the graph becomes connected and the orders of all vertices are equal.