A RANDOM GRAPH

The construction. A random (undirected) graph with n vertices is constructed in the following way: pairs of vertices are randomly 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? Yes. For example, if the vertex pair $\{x, y\}$ were to be selected subsequently, there would now be more than one edge connecting x and y: a *multiple edge* or *cycle 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 $n = O(m)$.

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 $p^{(n)}$

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

$$p = \begin{cases} 0 & \text{if } 0 < c \leq 2 \\ e^{\frac{1}{c}}\sqrt{1 - \frac{2}{c}} & \text{if } c > 2\,. \end{cases}$$

"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

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

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

$$\lim_{n \to \infty} p_k^{(n)} = e^{-\lambda_k},$$

where

$$\lambda_k = \left(\frac{2}{c}\right)^k \frac{1}{2k}.$$

It follows that

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

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

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

$$\lim_{n\to\infty} p^{(n)} = e^{-\lambda} \,,$$

where

$$\lambda = \sum_{k=2}^{\infty} \lambda_k \,. \tag{1}$$

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 (1) is divergent. Thus, we may deduce immediately that if $c \leq 2$, $\lim_{n\to\infty} p^{(n)}$ exists and equals 0.

The interesting case $c > 2$ is also easily dealt with. When $c > 2$, we have that $0 < \lambda_k < 1$ and that

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

From Markov's inequality we have

$$\Pr(X_k^{(n)} \geq 1) \leq \mathrm{E} X_k^{(n)}$$

and so

$$p_k^{(n)} = \mathrm{Pr}(X_k^{(n)} = 0) \geq 1 - \mathrm{E}X_k^{(n)} .$$

But, by Lemma 2 of [HM], we know have that, for each fixed $k \geq 2$, $\mathrm{E}X_k^{(n)} \uparrow \lambda_k$, as $n \to \infty$. In particular, for each fixed $k \geq 2$, the sequence $\{\mathrm{E}X_k^{(n)}\}$ is bounded above by λ_k. It follows that $\{q_k^{(n)}\}$ is bounded above by $c_k := -\log(1 - \lambda_k)$. Further, since $\lambda_k < 1$, we have

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

Thus, by the Lebesgue Dominated Convergence Theorem, we have that

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

and, hence, that $\lim_{n\to\infty} p^{(n)}$ exists and is equal to $e^{-\lambda}$.