Random Topologies on Finite Sets and Their Threshold Functions

Catriona Maclean*, Neil O'Connell
Basic Research Institute in the Mathematical Sciences
HP Laboratories Bristol
HPL-BRIMS-97-18
October, 1997

random topologies,
random functions,
power sets

For each integer n, there is a natural family of probability distributions on the set of topologies on a set of n elements, parameterised by an integer variable, m. We will describe how these are constructed and analysed, and find threshold functions (for m in terms of n) for various topological properties.

*Department of Mathematics, University of Cambridge, Cambridge, United Kingdom

Random Topologies on Finite Sets and their Threshold functions

C.F. Maclean N. O'Connell

20 September, 1997

Abstract

For each integer n, there is a natural family of probability distributions on the set of topologies on a set of n elements, parameterised by an integer variable, m. We will describe how these are constructed and analysed, and find threshold functions (for m in terms of n) for various topological properties.

Let us suppose we have some set, S, and we wish to pick a topology on S at random. What do we mean when we say "at random"? In the case where S is a finite set, we may say that all topologies on S are equally likely; however, if S happens to be an infinite set, this makes no sense. We would like to be able to talk about a random topology on any set, and so we are led to what is essentially a problem in measure theory: given S, how do we define a probability measure on the set {possible topologies on S} in a reasonably natural way?

1 The Probability Measure

Now, while we do not know how to define a sigma-algebra, let alone a probability measure, on the set {possible topologies on S}, we do know how to pick collections of random subsets of S. Given any cardinal number, m, which may be either infinite or finite, let us consider the set of functions, $\{f||f : \{m\} \mapsto \wp(S)\}$; ie the set of {collections of subsets of S indexed by m} (by $\{m\}$, I mean the set $\{1, 2 \ldots m\}$). We can think of such a function as assigning to every member of the product set $\{m\} \times S$,(i,t) either a 1 or a 0, depending on whether or not the element, t of S is an element of $f(i)$.

This means that functions, $\{f \| f : \{m\} \mapsto \wp(S)\}$ are in natural one-to-one correspondence with elements of the set $\{0,1\}^{\{m\}\times S}$. Now, we can put a measure on $\{0,1\}$, given by $\mu(0) = 1/2$ and $\mu(1) = 1/2$, called Bernoulli measure. And this means that in principle we can put a measure on $\{0,1\}^{\{m\}\times S}$, namely product measure with respect to Bernoulli measure.

So, we have a natural measure, μ_m, on the set of {collections of m random subsets of S}. But, we also have a natural function,f, from {collections of random subsets of S} to {topologies on S}, namely the function which takes a collection of subsets to the topology generated by them. Now, this means we can define a probability measure, p_m, on the set, $T =$ {topologies on S}, given by $p_m(T') = \mu_m(f^{-1}(T'))$, whenever the right hand side is defined.

So, for every set, S, we have constructed a natural family of measures on the set of topologies on S, indexed by the cardinal numbers. We will consider what happens when S is a finite set; without loss of generality, S is the set $\{1, 2 \ldots, n\}$, for some n. We will denote this set by $\{n\}$. For every natural number, m, there is a measure μ_m on the set of topologies on $\{n\}$, and we will denote a random topology on $\{n\}$, chosen with respect to the measure, μ_m, by $T_m(n)$.

2 Preliminaries

Before starting on the real work of determining threshold functions for topological properties in this model, we will need some technical results, which will form the basis for our later work.

We will think of a choice of m random subsets of the set $\{n\}$ as being a random function, $\{f | f : \{m\} \mapsto \wp(\{n\})\}$, where all functions are equally likely.

We observe first that there is a one-to-one correspondence, C, between functions $f : \{m\} \mapsto \wp(\{n\})$ and functions $g : \{n\} \mapsto \wp(\{m\})$ given by

$$i \in f(j) \Leftrightarrow j \in g(i).$$

We note that the law of g is uniform on the set of all mappings.

Throughout what follows, f will denote a random function from $\{m\}$ to $\wp(\{n\})$ and g will denote its image under C. This function, g will be useful to us, because it will give us a more concrete description of the topology, $T_m(n)$

than " the topology generated by this random collection of open sets". We observe secondly that, associated with a topology on a finite set, there is

1. an equivalence relation on the set and
2. a partial order on the equivalence classes formed by the relation.

The relation is given by $a \sim b \Leftrightarrow \forall S$ open $a \in S \Leftrightarrow b \in S$.

The partial order is given by $[a] \leq [b] \leftrightarrow \forall S$ open $a \in S \Rightarrow b \in S$.

Given such a relation, R, and a poset, P, they define a topology,T, given by

S open $\Leftrightarrow$ S is a collection of equivalence classes, closed under ascending P.

We claim that, if S is a finite set, then this is the only possible topology giving rise to R and P

Lemma 1 *For S finite, P and R completely determine the topology, T.*

Proof: Suppose that 2 topologies, T_1 and T_2 define R and P. Now, consider an open set in T_1, S. On taking intersections, we get that, $\forall x \in \{n\}$

$$\{y|[y] \geq [x]\}$$

is an open set in both T_1 and T_2. Call it $S(x)$. Equally, on taking unions, we get:

$$S = \cup\{S(x)|x \in S\}$$

$\forall$ open S. Thus S is necessarily open in T_2. Thus $T_1 = T_2$.

Now, we want to know what this partial order and equivalence relation are, for a topology $T_m(n)$. Recall that, when we chose a random topology $T_m(n)$, we also chose a random function, $g : \{n\} \mapsto \wp(\{m\}$. $g(i)$ is interpreted as being the labels of the random basis elements (in the random basis which generates $T_m(n)$) which contain i. We would like to have a concrete description of the equivalence relation and partial order which define $T_m(n)$.

When is it true that, in the topology $T_m(n)$, an open set contains i iff it contains j? This happens precisely when every element of our random basis either contains both i and j or neither i nor j. That is, exactly when $g(i) = g(j)$. When is it true that an open set in $T_m(n)$ which contains a necessarily contains b? This is when there is no element of our random basis which contains a but not b: ie whenever $g(a)$ is a subset of $g(b)$. So the equivalence relation and partial order associated with our topology , $T_m(n)$ are precisely the equivalence relation and partial order induced by the random function, g.

That is, a random topology $T_m(n)$ is induced by a random mapping from n to $\wp(\{m\})$. There is a large literature on such mappings; see for example Kolchin; "Random Mappings". However, the properties of this random mapping, g, which will prove to be important in the study of random topologies, are somewhat unusual and have not, so far as we are aware, been previously studied .

We will need the following lemma:

Lemma 2 *If P is a poset on n elements, then,*

$$p(P \subset P(T_m(n))) = (\frac{a}{2^n})^m$$

where $a = |\{$ antichains in the poset, $P\}|$

Proof:

The probability we seek is given by $p = \frac{|\{f:\{n\} \mapsto \wp(\{m\}) | f induces P\}|}{2^{mn}}$. It will be enough to show that:

$\#\{f \in F(m,n) | f \text{ induces } P\} = a^m$.

Now, f induces P iff ,for every element of m, $f(m)$ is closed under taking supersets with respect to P. In other words, such functions are precisely functions from $\{m\}$ to the set

$S' = \{S \subset \{n\} | S \text{ is closed under superset with respect to } P\}$.

It will be enough to show that S' is in one-to-one correspondence with $\{$antichains in P$\}$. But the function, $S \mapsto$ minimal elements of S provides such a correspondence, and the result follows.

3 Threshold functions

We will now turn our attention to the problem of threshold functions for topological properties. A threshold function for some property of a topology (for example, connectedness), call the property Q, is a function, m(n), which satisfies the following 2 properties:

$$\frac{m^*(n)}{m(n)} \to \infty \Rightarrow P(Q|T_{m^*(n)}(n)) \to 1$$

$$\frac{m^*(n)}{m(n)} \to 0 \Rightarrow P(Q|T_{m^*(n)}(n)) \to 0$$

or conversely.

In fact, I will instead look for functions, $m(n)$, for which I can prove a theorem of the following form:

$$\frac{m^*(n)}{m(n)} \to \lambda \Rightarrow P(Q|T_{m^*(n)}(n)) \to f(\lambda)$$

where f is some function of λ.

For more information on threshold functions, see Bollobas, Combinatorics.

We will present proofs of threshold functions for the following 4 topological properties:

1. Discreteness of the topology $T_m(n)$.
2. Surjectivity of the function, g.
3. Distinguishability of all points in the topology $T_m(n)$.
4. Connectedness of the topology $T_m(n)$.

Theorem 1 *If* $n^2 = \lambda(\frac{4}{3})^m$ *then* $lim_{n\to\infty}P(T_m(n)$ *is the discrete topology*)$=$ $e^{-\lambda}$.

Proof:

Suppose that $n^2 = \lambda(\frac{4}{3})^m$.

($T_m(n)$ discrete)$\Leftrightarrow$(none of the posets $i \leq j$ are induced by f).

Now, there are n^2 events of this form, each occuring with probability $(\frac{3}{4})^m$. If they were independent, we could conclude the argument here. However, they are not— but events of the form $x_i \leq y_i$ are independent, provided that all the x_i's and y_i's are distinct. There is an inclusion-exclusion proof that, given N independent events, each happening with probability λ/N, the probability that none of them happen is $e^{-\lambda}$. Our aim will be to modify this proof so that it applies to these almost-independent events. To do this, we will need to show that the probability of 2 non-independent events happening is small. Let us consider this probability— ie, let us consider $q = P(E)$, where E is the event that at least one of the following holds:

1. $\exists i, j, k$ such that $i \leq j$ and $k \leq j$

2. $\exists i, j, k$ such that $i \geq j$ and $k \geq j$

3. $\exists i, j, k$ such that $i \leq j$ and $j \leq k$

4. $\exists i, j$ such that $f(i) = f(j)$

Adding together probabilities:

$$q \leq \frac{n^3}{2}(\frac{5}{8})^m + \frac{n^3}{2}(\frac{5}{8}) + \frac{n^3}{2}(\frac{1}{2})^m + n^2(\frac{1}{2})^m$$

Substituting in $n = \sqrt{\lambda(\frac{4}{3})^m}$, we see that $q \to 0$. Let us consider

$P(T$ is not the discrete topology and E does not happen$) = p$

It will be enough to show that $p \to e^{-\lambda}$. By inclusion-exclusion principle we have that, $\forall k$:

$$\Sigma_{i=1}^{2k}(-1)^{i+1}(|\{\{< x_j, y_j >\}_{j=1}^{j=i}\}|P_i) \leq p \leq \Sigma_{i=1}^{2k+1}(-1)^{i+1}(|\{\{< x_j, y_j >\}_{j=1}^{j=i}\}|P_i)$$

where

$$P_i = P(x_j \leq y_j \forall j \& not E)$$

Now,

$$|\{\{< x_j, y_j >\}_{j=1}^{j=i}\}| = n^{2i/i!}$$

And,

$$\frac{n^{2i/i!}}{\frac{4}{3}^{mi}} \to \frac{\lambda^i}{i!}$$

as $n, m \to \infty$

So, our equation is eventually very close to:

$$\Sigma_{i=1}^{2k}(-1)^{i+1}\frac{\lambda^i}{i!}(P_i((\frac{4}{3})^{mi})) \leq p \leq \Sigma_{i=1}^{2k+1}(-1)^{i+1}\frac{\lambda^i}{i!}(P_i((\frac{4}{3})^{mi}))$$

We see from this that it will be enough to show that

$$\lim_{n \to \infty} \frac{P_i}{(\frac{3}{4})^{mi}} = 1.$$

Note that $P_i = (\frac{3}{4})^{mi} - r$, where

$$r = P(x_j \leq y_j \forall j \& E.$$

It remains to show only that $\lim_{n \to \infty} r(\frac{4}{3})^{mi} = 0$.

We note that $r = P(F)$, where F is the event, one of the following happens (k is assumed to be not equal to any of the x_j's or y_j's throughout):

1. $x_j \leq y_j \forall j$ and a poset of the form (1)-(4) ,which is independent of the x_j's and y_j's occurs.
2. $x_j \leq y_j \forall j$ and $\exists k | x_j \leq k, k \neq y_j$
3. $x_j \leq y_j \forall j$ and $\exists k | y_j \geq k, k \neq x_j$
4. there is a proper extension of the poset $x_j \leq y_j (= P)$ on the set x_j, y_j

Now, P(1 holds)$\leq q(\frac{3}{4})^{mi}$.

P(2 holds)= P(3 holds) $\leq ni(\frac{3}{4})^{mi}(\frac{5}{6})^m$.

(This holds because there are $< n$ possible choices for k, i possible choices for the pair x_j, y_j, and $P(k \leq y_j | x_j \leq y_j) = (\frac{5}{6})^m$, by Lemma 2.)

Now, $(\frac{5}{6} < \sqrt{\frac{3}{4}}) \Rightarrow \lim_{n \to \infty}$ P(2 holds)$(\frac{4}{3})^{mi}) = 0$

There are only a finite number of proper poset extensions of P, each of which occurs with a probability of the form $(\mu)^m$, where $\mu < (\frac{3}{4})^i$.

So, $\lim_{n \to \infty} r(\frac{4}{3})^{mi} = 0$. We are done.

The next property we will look at will be surjectivity of the function, g. This interests us because, in the case where g is surjective, our little studied problem of random topologies reduces to 2 much-studied problems, namely the problem of random partitions of a set, and the partial order on a power set.

Lemma 3 *If* $n = \lambda 2^m \log 2^m$ *then, as* $n \to \infty$, *if* $\lambda < 1$ $P(f$ *is surjective*$) \to 0$ *and if* $\lambda > 1 P(f$ *is surjective*$) \to 1$.

Proof:

Throughout the proof of this lemma, we will treat g as a random function from $\{n\}$ to $\{2^m\}$.

The case $\lambda > 1$ follows on computing the expected number on elements of $\wp(\{m\})$ not in the image of g. Given some element, i of $\{2^m\}$, then

$$P(\nexists k | g(k) = i) = (1 - 2^{-m})^n.$$

Further,

$$(1 - 2^{-m})^n = ((1 - 2^m)^{2^m})^{\log 2^m \lambda}.$$

Recall that

$$(1 - 2^m)^{2^m} \to e^{-1};$$

thus, eventually

$$(1 - 2^m)^{2^m \lambda} < e - \epsilon$$

for some positive ϵ. We deduce from this that

$$\lim_{n \to \infty} 2^m (1 - 2^{-m})^n = 0$$

and thus that

$$\lim_{n \to \infty} E(\{i | \nexists k, s.t. g(k) = i\}|) = 0$$

The result follows.

For the case $\lambda < 1$; consider

$$p_j = P(\nexists k, s.t. g(k) = j | \forall i < j \exists k, s.t. g(k) = i).$$

We claim that

$$p_j \geq P(\nexists k, s.t. g(k) = j) = p(*)$$

Now, Bayes formula tells us that

$$P(A|B) \geq P(A) \Leftrightarrow P(A + B) \geq P(A)P(B) \Leftrightarrow P(B|A) \geq P(B)$$

for any events A, B.

In this case, to prove (*), it will be enough to show that

$$P(\forall i < j \exists k, s.t. g(k) = i | \nexists k, s.t. g(k) = j) \geq P(\forall i < j \exists k, s.t. g(k) = i)$$

In other words we have to show that a random function from a set, $\{n\}$ to a set of size $2^m - 1$ is more likely to hit all the first $j - 1$ elements than a random function from $\{n\}$ to a set of size 2^m.

Now, let us consider a random function from a set, $\{n\}$, to a k-element set. We can regard this as being a choice of a random subset of $\{n\}$, S (to be interpreted as the elements which go to k under our random function), and a random function from the set $\{n\} \setminus S$ to a $k-1$ element set. So it will be enough to show that the probability of a random function from $\{n\}$ to a k element set hitting all the first $j-1$ elements is increasing in n— which it is.

So,

$$P(\forall i \leq j \exists k, s.t. g(k) = i) \leq (1-p)^j.$$

We know $p = (1-2^{-m})^n$ For $n = 2^m \log 2^m \lambda$,where $\lambda < 1$,

$$(1-x\epsilon)^{\frac{1}{\epsilon}} \rightarrow_{\epsilon \rightarrow 0} e^{-1} \Rightarrow \lim_{n\rightarrow\infty} 2^m(1-2^{-m})^n = \infty$$

.

Thus, given any real number, x, eventually

$$P(\forall i \exists j | g(j) = i) \leq (1-p)^{2^m} \leq (1-x2^{-m})^{2^m} \rightarrow e^{-x}.$$

The theorem follows.

We now turn to the problem of distinguishability: how many sets in a randomly chosen basis do we need to be almost sure that in $T_m(n)$, any 2 points can be separated in $T_m(n)$?

Lemma 4 *If $n^2 = 2\lambda 2^m$, then as $n \rightarrow \infty$, $P(T$ distinguishes the points of $n) \rightarrow e^{-\lambda}$*

Proof:

$p = P(T_m(n)$ distinguishes the points$) = P(g$ injective$)=\prod_{i=0}^{n-1}(1-\frac{i}{2^m})$.

Let us take logs of this equation:

$$\log p = -\Sigma_{i=0}^{i=n-1} \frac{i}{2^m} + \Sigma_{i=0}^{i=n-1} \epsilon(i)$$

where $\epsilon(i)$ is an error term. Rearranging:

$$\log p = -\lambda + \Sigma_{i=0}^{i=n-1} \epsilon(i)$$

Now Taylor's theorem tells us that

$$\epsilon(i) = O(\frac{i}{2^m})^2$$

Therefore, $\log p \rightarrow -\lambda$;and the result follows .

Our next result will be a threshold function for topological connectedness.

Theorem 2 *(1)If* $n = \mu \log n(2^{-\lceil \frac{m}{2} \rceil} + 2^{-\lfloor \frac{m}{2} \rfloor})^{-1}$*,and* μ *is bounded below by* $\lambda > 1$*, then* $P(T$ *is connected*$) \to 1$ *as* $n \to \infty$*.(*)*

(2)If $n = \mu \log n(2^{-\lceil \frac{m}{2} \rceil} + 2^{-\lfloor \frac{m}{2} \rfloor})^{-1}$*, and* μ *is bounded above by* $\lambda < 1$*, then* $P(T$ *is connected*$) \to 0$ *as* $n \to \infty$*.(**)*

Proof of (1):

We note first that a topology, T, is connected iff the associated poset,P is connected. Indeed, since open sets in T are precisely the sets which are closed under ascending P, the components of T are precisely the components of the P.

Recall further that the partial order associated with $T_m(n)$ was the same as the partial order induced by a randomly chosen function $g : \{n\} \mapsto \wp(\{m\})$. So, our theorem is that, for these values of m and n, such a function eventually almost surely induces a connected partial order.

Now, let us consider $g(i)$, and let us consider the set {subsets of $\{m\}$ which are comparable with $g(i)$}. Call this set S. Now, if $\#(g(i)) = l$, then $\#(S) = 2^l + 2^{m-l}$. In particular, the size of S grows exponentially with l, for $l > \lceil m/2 \rceil$, and decreases exponentially with l for $l < \lfloor m/2 \rfloor$. So, if we can find i such that $g(i)$ is either unusually large or unusually small, then we expect that i will be comparable with an unusually large number of points in the poset induced by g , and our strategy in this proof will be to show that there does, indeed, exist such an i , and then use it as a "hub", out of which will we "grow" a component, and evantaully show that this component covers the whole of the set $\{n\}$.

This motivates the following lemma:

Lemma 5 $\exists \eta$ *such that,* $\eta > \frac{3}{4}$*, and, for n,m such that (*) holds,* $P(\exists i \| g(i) | \geq \eta m) \to 1$,

Proof:

Stirling's formula tells us that $C^m_{\eta m} \to q^m/p(\sqrt{m})$, where $q = (\frac{1}{\eta})^\eta (\frac{1}{1-\eta})^{1-\eta}$ and p is a polynomial. Note that q is a continuous function of η.

$q(3/4) > \sqrt{2} \Rightarrow \exists \eta$ such that $\eta > 3/4$ and $q(\eta) > \sqrt{2}$.(a)

Let us choose η such that (a) holds. Consider $P(|g(i)| \geq \eta m) = p(\eta)$;

$$p(\eta) \geq \frac{C^m_{\eta m}}{2^m} \approx (\frac{q(\eta)}{2})^m / (p(\sqrt{m}))$$

Let $\frac{C^m_{\eta m}}{2^m} = t$. Now,

$$(q/2) > \sqrt{\frac{1}{2}} \Rightarrow \frac{t}{(\sqrt{2})^{-m}} \to \infty \Rightarrow tn \to \infty.$$

So, for any real number, x, eventually

$$P(\nexists i||g(i)| \geq \eta m) \leq (1 - \frac{x}{n})^n (\approx e^{-x})$$

The result follows.

Note that, by symmetry arguments,

$$P(\nexists i||g(i)| \leq (1-\eta)m) \to 0.$$

Now, we are almost sure to have an element of $\{n\}$, i, such that $|g(i)| \geq \eta m$. We will now start "growing" the component out from i. We will start with the following lemma:

Lemma 6 $\exists \gamma < 3/4$ *such that,* $P(\exists i, j, s.t. |g(i)| \geq \eta m, |g(j)| \geq \gamma m$, *and* i, j *in different components of* $T_m(n)) \to 0$.

Proof:

Our strategy here will be to show that there almost certainly exists some k, such that $g(k)$ is contained in $g(i) \cap g(j)$, for every such pair i, j. We will then be done, since k will then be comparable with both i and j in the poset induced by g. Consider γ such that

$$\gamma < 3/4$$

$$gamma + \eta > 3/2$$

Suppose that $|g(i)| \geq \eta m$ and $|g(j)| \geq \gamma m$. Consider $g(i) \cap g(j)$. We have

$$|g(i) \cap g(j)| > (\gamma + \eta - 1)m :$$

we set $(\gamma + \eta - 1) = \nu (> 1/2)$. Consider

$$P(g(k) \in g(i) \cap g(j)) = p \geq 2^{(\nu - 1)m}$$

$$(*) \Rightarrow \exists \epsilon > 0 | 2^{(\nu-1)m} \geq n^{\epsilon-1}$$

for sufficiently large n.

Thus $P(g(k) \in g(i) \cap g(j)) \geq n^{\epsilon-1}$, for sufficiently large n.

Consider

$$P(\nexists k | g(k) \subset g(i) \cap g(j)) \leq (1 - n^{\epsilon-1})^{n-2}$$

Recall that $(1-\delta)^{\frac{1}{\delta}} \to e^{-1}$ as $\delta \to 0$. Thus, for n sufficiently large:

$$P(\nexists k | g(k) \subset g(i) \cap g(j)) \leq e^{-n^{\epsilon/2}}$$

$$P(\exists i, j || g(i)| \geq \eta m, |g(j)| \geq \gamma m \& \ \nexists k, s.t. g(k) \subset g(i) \cap g(j)) \leq n^{\underline{2}} e^{-n^{\epsilon/2}},$$

since there are at most $n^{\underline{2}}$ possible choices of i, j,

Now, $\lim_{n \to \infty} n^{\underline{2}} e^{-n^{\epsilon/2}} = 0 \Rightarrow$ the result follows.

So we are now almost sure that all elements of $\{n\}$, whose image under g has size at least γm, are in the same component of $T_m(n)$. Of course, by symmetry we are also almost sure that any two points whose images under g have size at most $(1-\gamma)m$ are in the same component of the topology.

So, we have 2 components of our topology, one containing any point whose image under g is small, one containing any point whose image under g is large.

Our next lemma will show that these 2 components between them cover our set.

Lemma 7 *For m, n satisfying (*), $P(E) \to 1$, where E is the event:*
$\forall i, \exists k \neq i$, s.t. either
(1) $g(k) \subset g(i)$ and $|g(k)| \leq (1-\gamma)m$
or
(2) $g(k) \supset g(i)$ and $|g(k)| \geq \gamma m$

Proof:

Consider $g(i)$. Consider $S' = \{s \in \wp(\{m\}) | (1) \text{or} (2) holds\}$. Our aim is to show that

$$nP(\nexists i | g(i) \in S') \to 0.$$

We will start by finding a lower bound for the size of S'. Suppose that $|g(i)| = l$. Then,

$$h(l) = |S'| = \Sigma_{(1-\gamma)m \geq i \geq 0}(C_i^l + C^{m-l}i).$$

We claim that $min_l|S'| = h(\lceil m/2 \rceil)$.
For, consider

$$(C_i^l + C^{m-l}i) = f_i(l)$$

Combinatorial identities tell us that

$$f_i(l) - f_i(l-1) = C_{i-1}^{l-1} - C_{i-1}^{m-l}.$$

$$\Rightarrow (i < m/2 \Rightarrow (f_i(l) \geq f_i(l-1) - 1 \geq m - l).)$$

$$f_i(l) = f_i(m-l) \Rightarrow min_l(f_i(l)) = f_i(\lceil (m/2) \rceil).$$

$$h(l) = \Sigma_{(1-\gamma)m \geq i \geq 0} f_i(l) \Rightarrow min_l h(l) = h(\lceil (m/2) \rceil).$$

Consider $h(\lceil (m/2) \rceil)$. Suppose $|g(i)| = \lceil (m/2) \rceil$; define

$$S_1 = \{s : s \in \wp(\{m\}) | s \supset g(i)\},$$

$$S_1' = \{s \in S_1 | |s| \geq \gamma m\} = S_1 \cap S'$$

We claim that $\frac{|S_1'|}{|S_1|} \rightarrow 1$.

For, let us consider the 1-1 function,

$$f : S_1 \mapsto \wp(\{m\} \setminus g(i)), f(A) = A \setminus g(i).$$

Note that $|A| = \lceil (m/2) \rceil + |f(A)|$, and that, by the weak law of large numbers

$$P(|s| \geq (\gamma - \frac{1}{2})m | s \in \wp(\{m\} \setminus g(i))) \rightarrow 1.$$

(Incidentally, it is here that we have used our assumption that $\gamma < 3/4$.)
We are done.

Similarly, if

$$S_2 = \{s : s \in \wp(\{m\}), s \subset g(i)\}, S_2' = \{s : s \in S_2 | |s| \leq (1-\gamma)m\}$$

then $\lim_{n\to\infty} \frac{|S_2'|}{|S_2|} = 1$.

We have established the following equation:

$$\frac{min|S'|}{2^{\lceil (m/2) \rceil} + 2^{\lfloor (m/2) \rfloor}} \to 1(b)$$

Now, let us consider

$$P(\nexists i | g(i) \in S') = (1 - \frac{|S'|}{2^m})^{n-1}.$$

We aim to show that $n(1 - \frac{|S'|}{2^m})^{n-1} \to 0$
We have that, from (*) and (b);

$$\frac{min|S'|}{2^m} \geq \frac{\log n}{n} \lambda c(n,m)$$

where $c(n,m) \to 1$ as $n \to \infty$.

Recall that $(1 - x\epsilon)^{1/\epsilon} \to e^{-x}$ as ϵ goes to $0(\dagger)$. We have,

$$(1 - \frac{|S'|}{2^m})^{(n-1)} \leq (1 - \frac{\log n}{n} \lambda c(n,m))^{n-1} = ((1 - \delta(n,m))^{\frac{1}{\delta(n,m)} C(n,m)})^{\log n}.$$

where $\delta(n,m) \to 0$ as $n \to \infty$, and $C(n,m) \to \lambda$ as $n \to \infty$.

Now, by $(\dagger), \exists \epsilon > 0$, s.t., for sufficiently large n

$$(1 - \delta(n,m))^{\frac{1}{\delta(n,m)} C(n,m)} \leq e - \epsilon$$

Therefore,

$$((1 - \delta(n,m))^{\frac{1}{\delta(n,m)} C(n,m)})^{\log n} \leq (e - \epsilon)^{\log n}$$

Note that

$$n(e - \epsilon)^{\log n} \to 0$$

Thus, $nP(i$ is such that neither 1 nor 2 holds for any $k) \to 0$.

$\Rightarrow E(\#(i|$ neither 1 nor 2 holds for any $k)) \rightarrow 0$.
The lemma holds.

So, we have shown that, almost certainly, our topology has at most 2 components, one of which contains all the points whose image under g is small, and one containing all the points whose image under g is large. All that it remains to show is that these 2 components are in fact the same component. In other words, to prove our theorem, it will be enough to establish the following lemma:

Lemma 8 *For n, m satisfying (*), $P(E) \rightarrow 1$, where E is the event:*
$\exists i, j$ such that

1. $|g(i)| \geq \eta m$
2. $|g(j)| \leq (1-\gamma)m$
3. $g(j) \subset g(i)$

Proof:

Consider i, such that $|g(i)| \geq \eta m$.

Consider $P(g(j)$ satisfies 2 and 3$) = p$ — we aim to find a lower bound for p .Define:

$$h(l) = |\{s : |s| \leq (1-\gamma)m \& s \subset g(i)\}||(|g(i)| = l).$$

$h(l)$ is increasing in l; therefore, we may assume that $|g(i)| = \frac{3}{4}m$.

$$P(g(j) \subset g(i)) = 2^{-\frac{1}{4}m}.$$

Consider $P(|g(j)| \leq (1-\gamma)m | g(j) \subset g(i)) = q$.

We know $(1-\gamma) > \frac{1}{4}$, therefore,

$$q > \frac{C_{\frac{1}{4}m}^{\frac{3}{4}m}}{2^{\frac{3}{4}m}}$$

$$\approx \left(\frac{3^{\frac{1}{3}}(\frac{3}{2})^{\frac{2}{3}}}{2}\right)^{\frac{3}{4}m} \frac{1}{r(m)}$$

where r is some polynomial function of $\sqrt{m}$.
(Stirling's formula)
So,$\forall\epsilon > 0$, eventually

$$p > (1-\epsilon)\frac{(3^{\frac{1}{3}}(\frac{3}{2})^{\frac{2}{3}})^{\frac{3}{4}m}}{2^m r(m)}$$

$$= (1-\epsilon)\frac{t^m}{r(m)},$$

where $t > \sqrt{\frac{1}{2}}$. (computational check).

In particular, $|g(i)| \geq \eta m \Rightarrow nP(j$ satisfies (2) and (3)$) \rightarrow \infty$. Thus, given i

$$P(\exists j, s.t. 1-3 satisfied \| (|g(i)| \geq \eta m)) \rightarrow 1.$$

Together with Lemma 4, this proves the lemma.

Proof of (2):

Consider
$p = P(\exists k, s.t. \| |g(k)| = \lceil m/2 \rceil \& k$ a singleton in the poset, $P(T_m(n)))$.

We will show that, for n, m satisfying (**) $\lim_{n\to\infty} p = 1$.

We first define a subset, S of $\{n\}$, as follows:

$i \in S \Leftrightarrow |g(i)| = \lceil m/2 \rceil \& \forall j < i, j \notin S$ or $|g(j) \triangle g(i)| \geq \beta m$,

where β is some suitably chosen constant (to be chosen later).

We aim to show that, almost surely, there is an element of S which is a singleton in the induced poset.

We will need first get a bound on the size of S. β will be chosen so that S is almost surely sufficiently large.

We note, firstly, that, given j,

$$|\{s : |s \triangle g(j)|) < \beta m\}| \approx q^m/(p(\sqrt{m})),$$

p some polynomial, q is a function of β,s.t.

$$\forall\epsilon > 0, \exists\beta s.t. q(\beta) < 1+\epsilon).$$

(This result comes from Stirling's formula).

We choose β such that, for sufficiently large n, m satisfying (**).

$$n(\sum_{i=0}^{i=\beta m} C_i^m) < 1/2.C^m_{\lceil m/2 \rceil}(= O(\frac{2^m}{\sqrt{m}}))$$

,

Thus, $\forall \{g(1), g(2) \ldots g(i-1)\}$,

$$P(i \in S) \Leftrightarrow g(i) \in S' \subset \wp(\{m\}), |S'| \geq 1/2.C^m_{\lceil m/2 \rceil}$$

.

(S' depends on $\{g(1), g(2) \ldots g(i-1)\}$)
Therefore, $\forall \{s_j\}$

$$p(i \in S | g(j) = s_j, j < i) \geq \frac{1/2.C^m_{\lceil m/2 \rceil}}{2^m}$$

Stirling's formula tells us

$$\frac{1/2.C^m_{\lceil m/2 \rceil}}{2^m} = O(\frac{1}{\sqrt{m}})$$

Therefore,$P(|S| \geq n/m) \rightarrow 1$.
Consider
$P(\exists i \in S | i$ is a singleton in the induced poset $P(T_m(n)) = p'$.
We will rephrase this problem as follows:

We have a collection $C_1(=S)$ (not necessarily randomly chosen) of $\lceil (m/2) \rceil$-sets in $\wp(\{m\})$, such that:

$$\forall s, t \in C_1, |s \bigtriangleup t| \geq \beta m, |C_1| \geq \frac{n}{m}$$

We also have a collection, C_2, of randomly chosen elements of $\wp(\{m\})/\{m\}^{\lceil (m/2) \rceil}$, such that $|C_2| \leq n$.

We wish to find a lower bound for

$$P(\exists i \in C_1 | \nexists j \in C_2, s.t. i \supset j \& \nexists j, s.t. j \supset i) = p'.$$

As C_1 is not randomly chosen, we require this bound to be valid for any possible C_1: that is, we seek a lower bound for $\min_{C_1}(p')$.

As $\min_{C_1}(p)$ is decreasing in $|C_2|$, we may assume $|C_2| = N$, for any $N \geq n$. Similarly, we may assume that $|C_1| = \lceil n/m \rceil$.

Our strategy here will be to show that, for any positive γ, there is a choice of N, such that, eventually $N \geq n$, and $\min_{C_1}(p')|(|C_2| = N)) = 1 - e^{-\gamma}$

We have a collection of $\lceil n/m \rceil$ non-independent events, $\{E_1, E_2 \ldots E_{\lceil (m/2) \rceil}\}$, ($E_i$=(i is a singleton)) and we want to know $P(E_i$ happens, some i).

Consider $P(E_1, E_2 ... E_k$ all happen)$= p_k(N)$.

$$p_k(N) = (1 - \frac{|Q|}{2^m - C^m_{\lceil (m/2) \rceil}})^N$$

,

where $Q = \{S : S \in \wp\{m\}/\{m\}^{\lceil (m/2) \rceil}, \exists j \leq k | S \subset S_j \text{ or } S \supset S_j\}$.

By inclusion-exclusion:

$$k(2^{\lceil \frac{m}{2} \rceil} + 2^{\lfloor \frac{m}{2} \rfloor} - 2) \geq |Q| \geq k(2^{\lceil \frac{m}{2} \rceil} + 2^{\lfloor \frac{m}{2} \rfloor} - 2) - C^k_2(2^{\lceil \frac{m}{2} \rceil - \beta \frac{m}{2}} + 2^{\lfloor \frac{m}{2} \rfloor - \beta \frac{m}{2}})$$

Therefore:

$$(1 - \frac{k(2^{\lceil \frac{m}{2} \rceil} + 2^{\lfloor \frac{m}{2} \rfloor} - 2)}{2^m})^N \leq p_k(N) \leq (1 - \frac{k(2^{\lceil \frac{m}{2} \rceil} + 2^{\lfloor \frac{m}{2} \rfloor} - 2)}{2^m} + C^k_2(\frac{2^{\lceil \frac{m}{2} \rceil - \beta \frac{m}{2}} + 2^{\lfloor \frac{m}{2} \rfloor - \beta \frac{m}{2}}}{2^m}))^N . (\dagger$$

Our next aim will be to show that, for suitable choices of N, both sides of the equation ($\dagger$) $\sim (\frac{m\gamma}{n})^k$.

Proof:

We choose N such that $N = \log \frac{n}{m\gamma'} (2^{-\lceil \frac{m}{2} \rceil} + 2^{-\lfloor \frac{m}{2} \rfloor} - 2^{1-m})^{-1}$, where $\gamma' \to \gamma$.

By (**), eventually, $N > n$. Now, let us consider $p_k(N)$. In particular, let us consider the eqn. ($\dagger$).

Consider the LHS of this eqn:

$$(1 - \frac{k(2^{\lceil \frac{m}{2} \rceil} + 2^{\lfloor \frac{m}{2} \rfloor} - 2)}{2^m})^N = (1 - k \frac{\log \frac{n}{m\gamma'}}{N})^N$$

In particular, consider:

$$\log (1 - k \frac{\log \frac{n}{m\gamma'}}{N})^N = -k \log \frac{n}{m\gamma'} + \epsilon$$

where ϵ is an error term. Taylor's theorem, applied to log(1-x), tells us that, for sufficiently large n, $|\epsilon| < k\frac{(\log \frac{n}{m\gamma})^2}{N} \rightarrow_{n\rightarrow\infty} 0$. Thus, we see that the LHS of (†) $\sim (\frac{m\gamma}{n})^k$ as $n \rightarrow \infty$. RTS only that the RHS does, as well.

But let us consider

$$\frac{RHS(\dagger)}{LHS(\dagger)} = (1 + x(C_2^k(\frac{2^{\lceil \frac{m}{2} \rceil - \beta\frac{m}{2}} + 2^{\lfloor \frac{m}{2} \rfloor - \beta\frac{m}{2}}}{2^m})))^N = y$$

where $x \rightarrow 1$ as $n, m \rightarrow \infty$.

Observe that

$$\lim_{n\rightarrow\infty} Nx(C_2^k(\frac{2^{\lceil \frac{m}{2} \rceil - \beta\frac{m}{2}} + 2^{\lfloor \frac{m}{2} \rfloor - \beta\frac{m}{2}}}{2^m}) = 0$$

Therefore, $\lim_{n\rightarrow\infty} y \rightarrow 1$.

Thus, for this choice of N, $\forall\epsilon > 0$, for n sufficiently large

$$\forall P_1 \forall\{E_{i_1}, \ldots E_{i_k}\}, (1-\epsilon)(\frac{m\gamma}{n})^k \leq P(\bigcap\{E_{i_1}, \ldots E_{i_k}\}) \leq (1+\epsilon)(\frac{m\gamma}{n})^k$$

Now, if we apply the inclusion-exclusion principle to the events E_i, we see that

$P(E_i$ happens, some $i) \rightarrow 1 - e^{-\gamma}$

We are done.

4 The Largest Component Problem

If we have a random topology, $T_m(n)$, then we have an associated random variable, $X_m(n)$, given by $X_m(n) = (|\text{largest component of } T_m(n)|)/n$.

Our aim in the current section is to examine the behaviour of this random variable as $n, m \rightarrow \infty$.

We will need the following technical lemma:

Lemma 9 *$\exists p < \sqrt{2}$and $\gamma < 3/4$ such that if $n = p^m$, then with probability eventually close to 1, all elements of $\{n\}$ whose images under g have size $< (1-\gamma)m$ or $> \gamma m$ lie in the same component of the induced topology.*

Proof:

This will be a modification of the proof of this same result for $n = \log n 2^{\frac{m}{2}}\lambda$

Let us consider $C^{m}_{\eta m} \approx q^m/\mathrm{f}(\sqrt{m})$.(f some polynomial)(Stirling's formula).

$$qp > 2 \Rightarrow nP(|g(i)| \geq \eta m) \to \infty \Rightarrow P(\exists i||g(i)| \geq \eta m) \to 1.$$

So, choose $\eta > \frac{3}{4}, s.t. q(\eta) > \sqrt{2}$.

$$p > \frac{2}{q(\eta)} \Rightarrow P(\exists i||g(i)| \geq \eta m) \to 1.$$

Let us also choose γ, s.t.

$$\gamma < \frac{3}{4}, \gamma + \eta > \frac{3}{2}.$$

Consider

$$P(\exists i, j : |g(i)| \geq \eta m, |g(j)| \geq \gamma m \not\exists k|g(k) \in g(i) \cap g(j)) \to 0$$

$$\Leftarrow 2^{(\gamma+\eta-2)m}n \to \infty \Leftrightarrow 2^{(\gamma+\eta-2)m}p^m \to \infty \Leftarrow p > 2^{2-(\gamma+\eta)}.$$

We also wish to establish that, almost surely, $\exists i, j$ such that

1. $\#(g(i)) \geq \eta m$
2. $\#(g(j)) \leq (1-\gamma)m$
3. $g(j) \subset g(i)$

Now,set
$p' = P(j \text{ s.t.} j \text{ satisfies (2) and (3)}||g(i)| \geq \eta m) \approx q'^m/(\mathrm{f}(\sqrt{m}))$,
q' is some real number,$q' > \sqrt{\frac{1}{2}}$.

$p > \frac{1}{q'} \Rightarrow np' \to \infty \Rightarrow P(\exists j, s.t. i, j,$ satisfy 1-3 $||g(i)| \geq \eta m) \to 1$

But, $P(\exists i||g(i)| \geq \eta m) \to 1$
Therefore, provided p satisfies

1. $p > \frac{1}{q'}$
2. $p > 2^{2-(\gamma+\eta)}$
3. $qp > 2$

p also satisfies the conditions of the lemma. The result follows.

We will also need the following lemma:

Lemma 10 *For $n > p^m$, p as in the statement of Lemma 9, $\forall i, P(E_i) \to 1$, where E_i is the following event:*

Either i is a singleton in the induced poset, P, or i lies in the component of P which contains all sets of size $\geq \gamma m$.

Proof:

We note that, for any positive ϵ,

$$2(1-\gamma-\epsilon)m > |g(i)| > (2\gamma - 1 + \epsilon)m \Rightarrow$$

$P(|(|s| - \frac{1}{2})| \geq (\gamma - \frac{1}{2})m)|s$ comparable with $g(i)) \to 1$.

So, let us consider i, such that i is not a singleton in the poset. Let j be minimal such that $g(j)$ is comparable with $g(i)$. g(j) is then a randomly chosen element of

$\{S : S \subset \{m\}, S$ comparable with $g(i)\}$.

Now, for some sufficiently small ϵ

$P(E') = P((2(1-\gamma) - \epsilon)m > |g(i)| > (2\gamma - 1 + \epsilon)m) \to 1$, and

$P(|g(j)| \geq \gamma m$ or $|g(j)| \leq (1-\gamma)m|E') \to 1$.

The result follows.

Theorem 3 *If $n = 2^{m/2 - 2\lambda\sqrt{m}}$, then the size of the largest component of the induced topology, $X_m(n)$, converges to $\int_{|x|\geq\lambda} e^{-x^2}/\sqrt{2\pi x}$ in probability.*

Proof:

We have already seen that

$$E(X_m(n) + X'_m(n)) \to 1,$$

where $X'_m(n) = |\{i | i \text{ a singleton}\}|/n$ Since $X_m(n) + X'_m(n) \leq 1 + \frac{1}{n}$

$$X_m(n) + X'_m(n) \Rightarrow_{\mathcal{P}} 1$$

Choose $\mu < 2\lambda$ and $\nu > 2\lambda$. We will show that $P(E) \to 0$, where E is the following event:

Either $abs(|g(i)| - m/2) \leq \mu\sqrt{m}$ and i is not a singleton.†

Or $abs(|g(i)| - m/2) \geq \nu\sqrt{m}$ and i is a singleton.‡

For, let us consider the probability that † happens, $p(\dagger)$

$$p(\dagger) = \Sigma_l c_l (1 - (1 - 2^{-l} - 2^{l-m} + 2^{-m})^{n-1}),$$

where $l = |g(i)|$, the sum is taken over $abs(l - m/2) \leq \mu\sqrt{m}$ and $c_l = P(|g(i)| = l)$. So , given that

$$1 - (1 - 2^{-l} - 2^{l-m} + 2^{-m})^{n-1} \leq 1 - (1 - 2^{1-m/2+\mu\sqrt{m}})^{n-1}$$

we have

$$p(\dagger) \leq 1 - (1 - 2^{1-m/2+\mu\sqrt{m}})^{n-1}$$

Now,

$$n2^{1-m/2+\mu\sqrt{m}} = 2^{1+(\mu-2\lambda)\sqrt{m}} \to 0$$

and, thus

$$(1 - 2^{1-m/2+\mu\sqrt{m}})^{n-1} \to 1$$

as $n \to \infty$. Therefore, $p(\dagger) \to 0$ RTS only that $p(\ddagger) \to 0$. The same argument as above shows that

$$P(\ddagger) \leq (1 - 2^{-m/2+\nu\sqrt{m}})^{n-1}$$

$$n2^{-m/2+\nu\sqrt{m}} = 2^{(\nu-2\lambda)\sqrt{m}} \to \infty$$

and, thus

$$(1 - 2^{-m/2+\nu\sqrt{m}})^{n-1} \to 0$$

We are done.

So, if C is our largest component, and $S = \{i|abs(|g(i)| - m/2) \leq \mu\sqrt{m}$, then $E(|S \bigtriangleup C|) \rightarrow 0$: since $|S \bigtriangleup C| \geq 0$ necessarily,

$$|S \bigtriangleup C| \rightarrow_{\mathcal{P}} 0.$$

The Central limit theorem tells us that

$$P(g(i) \in S) \rightarrow \int_{|x| \leq \lambda} \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

So, now, the weak law of large numbers tells us that

$$|S| \rightarrow_{\mathcal{P}} \int_{|x| \leq \lambda} \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

We are done.

5 Acknowledgements

The authors wish to thank: Eleni Katirtzoglou, who, along with Neil O'Connell, first thought of looking at random topologies;

Jeremy Gunawardena, director of Hewlett Packard's Basic Research Institute in the Mathematical Sciences, for his support;

All those at Hewlett Packard Labs in Bristol who have helped us over the last 3 months;

Hewlett Packard, for their support for this research.

6 References

1. Kolchin, Random Mappings.

2. Bollobas, Combinatorics (C.U.P, 1986)