Random Graphs G22.3033-007
Assignment 9. SOLUTIONS

Note: NO CLASS Monday, April 24. LAST CLASS Monday, May 1

1. Let $D$ denote the unit disk in the plane with center $\vec{0} = (0, 0)$. Let $\vec{P}$ be at distance $1 - s$ from the $\vec{0}$, $s \in (0, 1)$. Let $L$ be the line through $\vec{P}$ perpendicular to the line from $\vec{0}$ to $\vec{P}$. Then $L$ splits $D$ into two parts. Let $f(s)$ denote the area of the smaller part.

(a) Find $f(s)$ precisely.
   
   Solution: We can assume $\vec{P} = (1 - s, 0)$ so that
   
   $$f(s) = \int_{1-s}^{1} 2\sqrt{1 - x^2} dx = \cos^{-1}(1 - s) - (1 - s)\sqrt{1 - (1-s)^2}$$

(b) Give an asymptotic formula $f(s) \sim As^B$ as $s \to 0^+$.
   
   Solution: Setting $y = 1 - x$, $f(s) = \int_0^s 2\sqrt{2y - y^2} dy$. But as $s \to 0^+$ we have $y = o(1)$ throughout the range so $2\sqrt{2y - y^2} \sim 2\sqrt{2y^{1/2}}$ and $f(s) \sim \int_0^s 2\sqrt{2y^{1/2}} dy \sim As^B$ with $A = 4\sqrt{2}/3$, $B = 3/2$.

(c) Let $\vec{P}_1, \ldots, \vec{P}_n$ be chosen uniformly and independently from $D$. Call $\vec{P}_i$ *extremal* if, splitting $D$ by a line through $\vec{P}_i$ as defined above, all the $\vec{P}_j, j \neq i$, lie in the larger part. Find an exact formula (as an integral) for the expected number of extremal points.
   
   Solution: For a uniformly chosen $\vec{P}$ let $r$ be the distance from $\vec{P}$ to the origin so that $\Pr[r \leq D] = D^2$. Setting $s = 1 - r$, $\Pr[s \geq D] = 1 - (1 - D)^2 = 2D - D^2$. Hence a uniformly chosen point has $s$ with density function $2 - 2s$. The probability that $P_1$ is extremal given this value $s$ is $(1 - f(s)/\pi)^{n-1}$ as the other $n - 1$ points must not be in the area. Thus
   
   $$\Pr[\vec{P}_1 \text{ is extremal}] = \int_0^1 (2 - 2s)(1 - f(s)/\pi)^{n-1} ds$$

   and the expected number $E$ of extremal points (by Linearity of Expectation) is $n$ times this:

   $$E = n \int_0^1 (2 - 2s)(1 - f(s)/\pi)^{n-1} ds$$
(d) Find an asymptotic formula (as \( n \to \infty \)) for the above formula. (Note: The main term will be when \( s \) is “small” but finding the right parametrization for \( s \) in terms of \( n \) is the key to the problem.)

Solution: As \( f(s) \sim As^{3/2} \) we parametrize \( s = xn^{-2/3} \) so that \( ds = n^{-2/3}dx \). In the critical range (leaving out the technical justification) we have \( 2 - 2s \sim 2 \) and \( f(s)/\pi \sim (A/\pi)x^{3/2}n^{-1} \) so that \( (1 - f(s)/\pi)^{n-1} \sim \exp[-(A/\pi)x^{3/2} \) and

\[
E \sim n^{1/3} \int_0^\infty 2 \exp[-(A/\pi)x^{3/2}]dx
\]

so \( E \sim A^*n^{1/3} \) with \( A^* \) a definite constant.

(Remark: The convex hull of \( n \) randomly chosen points has been the object of much study. Extremal points are necessarily on the convex hull, though the converse is not true.)

2. By a dumbbell we mean two cycles joined by a path. Let \( r, s \) denote the cycle lengths and \( t \) the number of interior points of the path so that the dumbbell has \( k = r + s + t \) points.

(a) Find the number of dumbbells with parameters \( r, s, t \) on \( n \) vertices.

Solution: Setting \( k = r + s + t \) there are \((n)_k\) labellings and four automorphisms as you can flip each cycle so that answer is \((n)_k/4 unless \( r = s \) in which flip the whole structure so it would be \((n)_k/8\).

(b) Find an exact expression \( A = A(n, p) \) for the expected number of dumbbells in \( G(n, p) \). The expression should be a sum over \( k \) as above.

Solution: There are \((k-1)(k-2)/2\) choices for \( r, s, t \geq 1 \) of which \([(k-1)/2]\) have \( r = s \). Hence there are

\[
D(k) := (n)_k\frac{(k-1)(k-2)}{8} - \frac{[(k-1)/2]}{8}
\]

dumbbells on \( k \) points. Each has probability \( p^{k+1} \) of being in \( G \) so

\[
A = A(n, p) = \sum_{k=7}^{n} D(k)p^{k+1}
\]
(c) Let $c < 1$ be fixed and let $G \sim G(n,p)$ with $p = \frac{c}{n}$. Prove (by showing $A(n,p) \to 0$) that almost surely $G$ does not contain a dumbbell.

Solution: The rough bound $D(k) \leq k^2n^k$ already gives

$$A \leq p \sum_{k=7}^{n} k^2c^k = o(1)$$

as the sum converges.

(d) (*) Let $\rho = \frac{1}{n} + \lambda n^{-4/3}$ where $\lambda = \lambda(n) \to -\infty$. (This is known as the subcritical range.) Prove (by showing $A(n,p) \to 0$) that almost surely $G$ does not contain a dumbbell.

Solution: Again

$$A \leq p \sum_{k=7}^{n} k^2(\rho n)^k \leq n^{-1} \sum_{k=7}^{n} k^2(\rho n)^k$$

One could work the sum out precisely ($\sum_{k} k^2 \alpha^k$ has an exact expression) but here is the rough idea of why it works: Set $\rho n = 1 - \epsilon$ so that $(\rho n)^k \leq e^{-\epsilon k}$. For $k$ up to $O(\epsilon^{-1})$ this term is $\Theta(1)$, after it drops off exponentially and has negligible contribution to the sum. So for $k = 7$ to $\Theta(\epsilon^{-1})$ we are adding $k^2$ which gives a sum $\Theta(\epsilon^{-3})$. We are assuming $\epsilon \gg n^{-1/3}$ so that the sum is $o(n)$, when we multiply by $n^{-1}$ we get $A = o(1)$.

[Note: Connected components of $G$ which are neither trees nor unicyclic are called complex. Complex components either contain dumbbells or something similar. Extending this analysis one can show that complex components do not appear in the subcritical range.]

3. Consider a branching process beginning with root Eve in which each node independently has number of children given by a Poisson distribution with mean one. List all possible outcomes (trees) for which the family (including Eve) has precisely five nodes. For each give the probability of obtaining that outcome. What is the total probability of getting a family of size precisely five?

Solution: All cases not named below are childless and hence have factor of $e^{-1}$.

(a) Eve has four children:

$$(e^{-1}/4!)(e^{-1})^4 = e^{-5}/24$$
(b) Eve has three children, one has one:
\[
(e^{-1}/3!)(e^{-1})^3 = e^{-5}/2. \text{ (Note: The 3 is the choice of the child with a child)}
\]

(c) Eve has two children, both with one:
\[
(e^{-1}/2!)(e^{-1})^2 = e^{-5}/2
\]

(d) Eve has two children, one with two:
\[
(e^{-1}/2!)(e^{-1}/2!)(e^{-1})^2 = e^{-5}/2
\]

(e) Eve has two children, one with one who has one:
\[
(e^{-1}/2!)(e^{-1})e^{-1} = e^{-5}
\]

(f) Eve has Fran who has three children:
\[
e^{-1}(e^{-1}/3!)(e^{-1})^3 = e^{-5}/6
\]

(g) Eve has Fran who has two children, one with one child:
\[
e^{-1}2(e^{-1}/2!)(e^{-1})e^{-1} = e^{-5}
\]

(h) Eve has Fran who has Glenda who has two children:
\[
e^{-1}(e^{-1}/2!)(e^{-1})e^{-1} = e^{-5}/2
\]

(i) Eve has Fran who has Glenda who has Harriet who has Isidora:
\[
e^{-5}
\]

Total: \[
e^{-5}\frac{125}{24}
\]

In general, for trees of size \(u\) the answer will be \(e^{-\alpha}u^{\alpha-2}/(\alpha-1)!\) as was shown in class.

4. Consider a branching process beginning with root Eve in which each node independently has number of children given by a Poisson distribution with mean \(c\).

(a) What is the probability Eve has precisely two children?
\[
\text{Solution: } e^{-c}c^2/2!
\]

(b) What is the probability Eve has no children with precisely two children?
\[
\text{Solution: As Eve has } Po(c) \text{ children she has } Po(ce^{-c}c^2/2!) \text{ children with precisely two children (this is part of the amazing property of Poisson that if you have } Po(\alpha) \text{ blips and each blip is a blop with probability } \beta \text{ then you have } Po(\alpha\beta) \text{ blops) and so the probability she has no such is } \exp[-c^3e^{-c}/2].
\]

(c) Draw the family tree (including males) with root your maternal grandmother. [If you’d rather not, just make one up or use your officemate’s grandmother.]
Solution: Root $A$ had $B, C, D, E$, $E$ had $F, G$, $G$ had $H, I$, $H$ had $J$.

(d) True or false: she had no children that had no children that had no children.

Solution: $B$ had no children with no children as he had no children at all so the statement is false: $A$ has a child with no children that had no children.

(e) With the branching process defined for Eve what is the probability that she has no children that have no children that have no children?

Solution: An $X$ has probability $e^{-c}$ of having no children. $Y$ has $Po(c)$ children and hence $Po(ce^{-c})$ children with no children and hence probability $\exp[ce^{-c}]$ of having no children with no children. Eve has $Po(c)$ children and hence $Po(c)\exp[ce^{-c}]$ children with no children with no children and hence probability $\exp[-c\exp[ce^{-c}]]$ of having no children with no children with no children.

5. For $1 \leq i < j \leq n$ let $X_{ij}$ be independent and uniform in $[0, 1]$. Let $T$ be the size of the minimal spanning tree (MST) on $K_n$ with $X_{ij}$ being the length of the edge $ij$. Here we derive (minus some technical details) a remarkable formula for the asymptotic expectation of $T$.

(a) Argue $E[T] = \binom{n}{2} \int_0^1 t f^-(t) dt$ where $f^-(t)$ is the probability that edge $\{1, 2\}$ is in the MST conditional on $X_{12} = t$.

Solution: $T = \sum X_{ij} I_{ij}$ where $I_{ij}$ is the indicator of $ij$ being in the MST. By Linearity of Expectation $E[X] = \sum E[X_{ij} I_{ij}]$ and by symmetry all these are the same so $E[X] = \binom{n}{2} E[X_{12} I_{12}]$ If these were discrete the latter would be $\sum \Pr[X_{12} = t] E[X_{12} I_{12} | X_{12} = t]$ and the latter expectation is $t \Pr[I_{12} | X_{12} = t]$ which is just $f^-(t)$. As we have continuous probability the sum becomes an integral.

(b) Assume $X_{12} = t$. Argue that $\{1, 2\}$ is in the MST if and only if there is no path from 1 to 2 consisting of edges all of weight less than $t$. (Hint: Consider Kruskal’s algorithm.)

Solution: From Kruskal’s Algorithm an edge 1, 2 lies on the MST if and only if 1, 2 lie in different components in the graph of all edges $i, j$ with weight less than that of 1, 2. We’re conditioning on $X_{12} = t$ so that graph is $G(n, t)$ except that 1, 2 are not adjacent.
(c) Deduce $f^-(t)$ is the probability that 1, 2 don’t lie in the same component of $G(n, t)$, conditional on 1, 2 not being adjacent.

Solution: Right, that’s what I just said.

(d) Now set $t = \frac{c}{n}$ with $c > 0$ arbitrary and asymptotics as $n \to \infty$. Let $f(t)$ be the probability that 1, 2 don’t lie in the same component of $G(n, t)$ (with no conditioning). Argue that $f^-(t) = f(t) + o(1)$. (This is part of a very general principle that conditioning on an almost sure event changes a probability by $o(1)$.)

Solution: $f(t) = t + (1 - t)f^-(t)$ as we split $f(t)$ into when 1, 2 are adjacent in $G$ and when they are not. Thus $|f(t) - f^-(t)| = |t - tf^-(t)| = o(1)$ as $t = o(1)$ and $f^-(t) \leq 1$.

(e) (The meat of the problem.) Find the limiting value of $f(t)$ as a function $g(c)$. (There will be two cases and $g(c)$ may be given implicitly. The key is to look at the almost sure picture of $G(t)$ given in class.)

Solution: When $c < 1$ $G(t)$ almost surely has largest component $O(\ln n)$ which is $o(n)$ so that $f(t) = o(1)$, as the second random vertex has probability $o(1)$ of being in the same vertex as any given first vertex. So $g(c) = 0$. For $t > 1$ let $y = y(c)$ be such that the giant component has size $\sim yn$, so that $y$ is given implicitly by $1 - y = e^{-ty}$ and $y > 0$. Then $f(t) = y^2(c) + o(1)$, as the two random vertices must both be in the giant component. So $g(c) = y^2(c)$

(f) For $K > 0$ let $E_K[T]$ denote the sum of the lengths of the minimal spanning tree, counting only lengths that are at most $K/n$. Express $\lim_{n \to \infty} E_K[T]$ as an integral.

(g) It can be shown that $\lim_{n \to \infty} E[T] = \lim_{K \to \infty} \lim_{n \to \infty} E_K[T]$ and lets assume that. (But I hope you see that this is not obvious. If, for example, most of the weight from the minimal spanning tree came from edges of weight between $n^{-1/2}$ and $2n^{-1/2}$ then it wouldn’t be true.) Given this assumption find $\lim_{n \to \infty} E[T]$ as an integral. Evaluate the integral numerically.

Solution: Setting $t = c/n$ we have $dt = n^{-1}dc$ so

\[ \left( \frac{n}{2} \right) \int_0^{K/n} tf(t)dt \rightarrow \frac{1}{2} \int_0^K cg(c)dc \]

(h) (*) Express $\lim_{n \to \infty} E[T]$ in a nice form which overlaps the name of a movie star.
Solution: One gets \( \frac{1}{2} \int_0^\infty cg(c)dc \) which after considerable work becomes \( \sum_{i=1}^\infty i^{-3} \) and the movie star is the only one I know with a Greek letter in her name: Catherine Zeta-Jones.

I think that it is a relatively good approximation to truth - which is much too complicated to allow anything but approximations - that mathematical ideas originate in empirics, although the genealogy is sometime long and obscure. But, once they are so conceived, the subject begins to live a particular life of its own and it is better compared to a creative one, governed by almost entirely aesthetical motivations, than to anything else and, in particular, to an empirical science.
– John von Neumann