Some proof sketches concerning the closure of limiting probabilities in sparse Erdős-Rényi random graphs

Tobias

September 29, 2019

When $c \geq 1$ we are dense in $[0,1]$

Let N_k denote the number of cycles of length exactly k. Then

$$N_k \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{Po}\left(\frac{c^k}{2k}\right),$$

and moreover for any fixed k, the random variables $N_3, \dots, N_k$ are asymptotically independent. (We need a reference for this – which surely exists.) Hence also

$$N_{\leq k} := N_3 + \cdots + N_k \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{Po}\left(\sum_{i=3}^{k} \frac{c^k}{2k}\right),$$

for any fixed k. Since $c \geq 1$ we can pick a k such that this mean is as large as we like. Note that for any k and $a \in \mathbb{R}$ the property that $N_{\leq k} \leq a$ can be expressed in FO. By the central limit theorem we have, for any fixed $x \in \mathbb{R}$:

$$\mathbb{P}(\mathrm{Po}(\mu) \leq \mu + x\sqrt{\mu}) \xrightarrow[\mu\to\infty]{} \Phi(x),$$

where $\Phi(.)$ denotes the cdf of the standard normal.

For $0 < p < 1$ and $\varepsilon > 0$ we can find x such that $\Phi(x) = p$, a μ_0 such that $\mathbb{P}(\mathrm{Po}(\mu) \leq \mu + x\sqrt{\mu}) \in (p-\varepsilon, p+\varepsilon)$ for all $\mu \geq \mu_0$, and then finally a k such that $\mu_{\leq k} := \sum_{i=3}^{k} \frac{c^k}{2k} \geq \mu_0$.

Conclusion : there exists an FO-property φ with limiting probability within ε of p.

When $c < 1$ then we get a finite union of intervals

The "fragment" F_n will be the union of all unicyclic components in $G(n.c/n)$. Rephrasing the findings of Marc and Alberto:

$$\mathbb{P}(F_n \cong H) \xrightarrow[n\to\infty]{} \frac{(ce^{-c})^{v(H)}}{\mathrm{aut}(H)} \cdot e^{-\sum_{k\geq 3} \frac{c^k}{2k}}, \tag{1}$$

for H any graph that consists of unicyclic components. Let us denote by F a random (unlabelled) graph with probability distribution given by the RHS of (1).

Side remark : this is the "Boltzmann-Poisson distribution"

$$\mathbb{P}(F = H) = \frac{1}{G(s)} \cdot \frac{s^{v(H)}}{\mathrm{aut}(H)} = e^{-C(s)} \cdot \frac{s^{v(H)}}{\mathrm{aut}(H)},$$

with $C(.)$ the generating function for connected, unicyclic graphs and $G(.)$ the generating function for graphs all of whose components are unicyclic, and $s := ce^{-c}$. Note that $s = ce^{-c} < e^{-1}$, the radius of convergence of C and G. Also note it is not hard to show that when $c < 1$ the expected number of unicyclic components equals $1 + o(1)$ times the expected number of cycles. I believe (1) alternatively can be derived from the fact that if $H_1, H_2, \dots$ is an enumeration of the unicyclic, connected graphs then the corresponding numbers of components $N(H_1), N(H_2), \dots$ are (in the limit) independent Poisson random variables – again a standard fact with probably a convenient reference – together with standard observations on the Boltzmann-Poisson distribution. In fact Theorem 1.3 from McDiarmid "Random Graphs from a Minor-Closed Class" does the trick for the Botlzmann-Poisson part of the argument I think. The Boltzmann-Poisson distribution has the property that the number of components isomorphic to $H_1, H_2, \dots$ are independent Poissons with the correct means and this completely specifies the distribution of the fragment.

Let $\mathcal{U}$ denote the unlabelled graphs whose components are unicyclic, and let $H_1, H_2, \dots$ be an enumeration of $\mathcal{U}$ such that $p_1 \geq p_2 \geq \dots$, where of course $p_i = \mathbb{P}(F = H_i)$ is as provided by the formula above. By an argument analogous to the one from the Heinig-Müller-Noy-Taraz paper

$$\mathrm{cl}(L) := \mathrm{cl}(\{\lim_{n\to\infty} \mathbb{P}(G(n, c/n) \models \varphi) : \varphi \in \mathrm{FO}\}) = \left\{ \sum_{i \in I} p_i : I \subseteq \mathbb{N} \right\}.$$

By Kakeya's criterion, if it holds that $p_i \leq \sum_{j>i} p_j$ for all sufficiently large i, then $\mathrm{cl}(L)$ is a finite union of intervals.

Let $i \in \mathbb{N}$ be arbitrary and define $k = k(i)$ via

$$\frac{s^{k-1}}{G(s)} \geq p_i > \frac{s^k}{G(s)}.$$

Now we consider the set $\mathcal{U}_k$ of all (unlabelled) graphs in $\mathcal{U}$ with exactly k verices. Let $T_{x,y,z}$ denote the graph consisting of a triangle with paths of length x, resp. y, resp. z, attached to its vertices. Note that $\mathrm{aut}(T_{x,y,z}) = 1$ if x, y, z are distinct, $\mathrm{aut}(T_{x,y,z}) = 6$ if $x = y = z$ and $\mathrm{aut}(T_{x,y,z}) = 2$ otherwise.

We have that $T_{0,0,k-3}, T_{0,1,k-4}, T_{0,2,k-5}, \dots, T_{0,\lfloor (k-3)/2 \rfloor, \lceil (k-3)/2 \rceil}$ are non-isomorphic members of $\mathcal{U}_k$. If k is odd then $T_{0,0,k-3}$ and $T_{0,(k-3)/2,(k-3)/2}$ both have two automorphisms, and the remaining $(k-5)/2$ members of the sequence each have only one automorphism. If k is even then $T_{0,0,k-3}$ is the only graph in the sequence with two automorphisms and the remaining $(k-4)/2$ members of the sequence have exactly one automorphism. This gives

$$\sum_{H \in \mathcal{U}_k} \frac{1}{\mathrm{aut}(H)} \geq \frac{k-3}{2}.$$

Hence, if $(k-3)/2 > 1/s$ (i.e. $k > 2/s + 3$) then

$$\sum_{j>i} p_j \geq \sum_{H \in \mathcal{U}_k} \frac{s^k}{\mathrm{aut}(H) G(s)} \geq \left(\frac{k-3}{2} \right) \cdot \frac{s^k}{G(s)} > \frac{s^{k-1}}{G(s)} \geq p_i,$$

Note that $k(i) > 2/s + 3$ whenever $p_i \leq s^{2/s+2}/G(s)$. This of course is true for all sufficiently large i as the p_i sum to one and in particular $p_i \xrightarrow[i\to\infty]{} 0$.

We have seen that, for any $0 < c < 1$, it is indeed the case that $p_i < \sum_{j>i} p_j$ for all sufficiently large i. This implies that $\mathrm{cl}(L)$ is a finite union of intervals.

(And unfortunately we do not get a Cantor set ...)

When $0.9368317138 \leq c < 1$

As argued convincingly by Marc and Alberto, when $c < 0.9368317138$ (the solution of $e^{c/2+c^2/4}\sqrt{1-c} = 1/2$) then $\mathbb{P}(\text{acyclic}) > \frac{1}{2}$ and there is at least one gap, containing $1/2$.

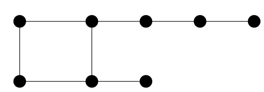
Now we assume $0.9368317138 \leq c < 1$. We will again make use of Kakeya's criterion, this time arguing that, with such a choice of c, in fact $p_i \leq \sum_{j>i} p_j$ **for all** i.

Note that if $0.9368317138 \leq c < 1$ then $s = ce^{-c}$ satisfies

$$\frac{1}{3} < s < \frac{1}{e}.$$

So in particular, since $\frac{k-3}{2} \geq 3 > 1/s$ for all $k \geq 9$, if $p_i \leq s^8/G(s)$ then $p_i < \sum_{j>i} p_j$. (Reusing the considerations from the previous section.) In particular, if $v(H_i) \geq 8$ then $p_i < \sum_{j>i} p_j$.

Now we consider $k = 8$. Observe that the graph obtained taking a four cycle, taking two vertices that are consecutive on the cycle and adding paths of length one, resp. 3, is an element of $\mathcal{U}_8$ that has exactly one automorphism.



exactly one automorphism

Hence

$$\sum_{H\in\mathcal{U}_8} \frac{1}{\mathrm{aut}(H)} \geq 1 + \frac{8-3}{2} > 3,$$

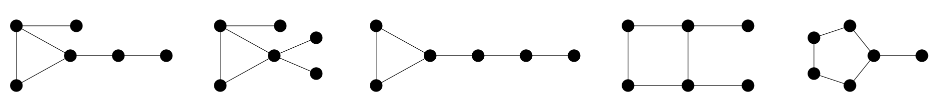
and we conclude that $p_i < \sum_{j>i} p_j$ whenever $v(H_i) \geq 7$.

The case when $k = 7$ is essentially the same. The four cycle with a paths of lenght one and two attached to two consecutive vertices on the four cycle has exactly one automorphism, so

$$\sum_{H\in\mathcal{U}_7} \frac{1}{\mathrm{aut}(H)} \geq 1 + \frac{7-3}{2} = 3,$$

giving that that $p_i < \sum_{j>i} p_j$ whenever $v(H_i) \geq 6$.

The following figure gives a (non-exhaustive) list of non-isomorphic graphs in $\mathcal{U}_6$, stating also their number of automorphisms.



aut = 1 aut = 2 aut = 2 aut = 2 aut = 2

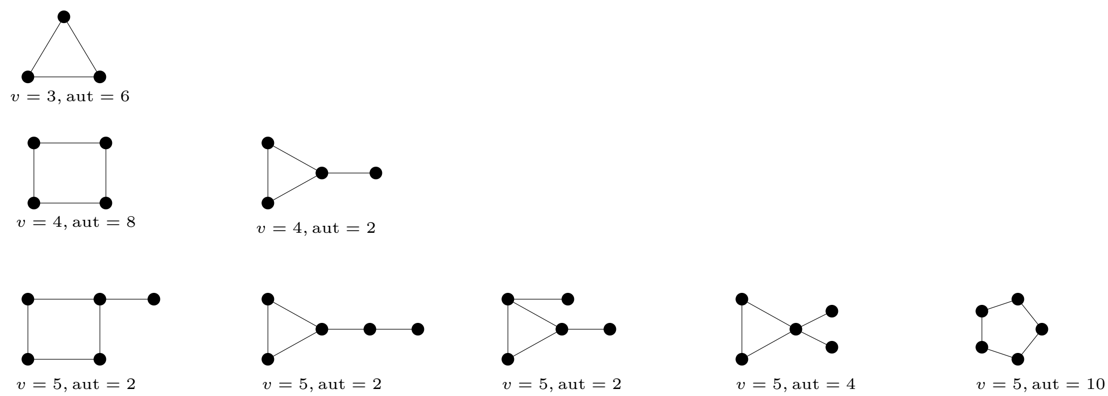
We see that

$$\sum_{H \in \mathcal{U}_6} \frac{1}{\mathrm{aut}(H)} \geq 3,$$

and thus $p_i < \sum_{j>i} p_j$ whenever $v(H_i) \geq 5$.

Of course the first graph in our ordering of $\mathcal{U}$ is $H_1 = \emptyset$, the empty graph, with probability $p_1 = \frac{1}{G(s)}$. As explained by Marc and Alberto, when $c \geq 0.9368317138$ then $p_1 \leq 1 - p_1 = \sum_{j>1} p_j$.

The following figure lists all non-empty graphs in $\mathcal{U}$ on at most five vertices, together with their numbers of automorphisms.



As argued above, if $v(H_i) \geq 5$ then $p_i < \sum_{j>i} p_j$. In fact this holds when $p_i \leq s^5/G(s)$. Note that

$$\mathbb{P}(F = C_4) = \frac{s^4}{8G(s)} \leq \frac{s^5}{G(s)},$$

so if $H_i = C_4$ the also $p_i < \sum_{j>i} p_j$.

We are only left with the triangle and the triangle with a path of length one attached. Their probabilities are $s^3/(6G(s))$, respectively $s^4/(2G(s))$. The numbers satisfy

$$\frac{s^4}{2G(s)} > \frac{s^3}{6G(s)}.$$

On the other hand

$$\frac{s^3}{6G(s)} > \frac{s^5}{2G(s)} = \max_{H \in \mathcal{U}_5} P(F = H).$$

So the triangle and the triangle with a path of lenght one attached occur in our ordering before anuy graph of order five. Now note

$$\sum_{H \in \mathcal{U}_5} \mathbb{P}(F = H) = 1.85 \cdot \frac{s^5}{G(s)} > \frac{s^4}{2G(s)}.$$

It follows that $p_i < \sum_{j>i} p_j$ also if H_i happens to be one of the two graphs left over, the triangle or the triangle with a path of length one attached.

We've seen that $p_i \leq \sum_{j>i} p_j$ for every i, as claimed.

Where to go from here?

Assuming the above is correct (please do check carefully!), we of course need to flesh out the argument and fill in the missing details.

Some follow up questions could be : When do we have exactly one, exactly two, etc., gaps? Or, maybe less ambitiously, can we show that the number of intervals tends to infinity as $c \searrow 0$? Maybe even provide the rate?