Poisson cloning model for random graphs

Jeong Han Kim

Abstract. In the random graph $G(n, p)$ with pn bounded, the degrees of the vertices are almost i.i.d. Poisson random variables with mean $\lambda := p(n-1)$. Motivated by this fact, we introduce the Poisson cloning model $G_{\mathrm{PC}}(n, p)$ for random graphs in which the degrees are i.i.d. Poisson random variables with mean λ.

We first establish a theorem that shows that the new model is equivalent to the classical model $G(n, p)$ in an asymptotic sense. Next, we introduce a useful algorithm to generate the random graph $G_{\mathrm{PC}}(n, p)$, called the cut-off line algorithm. Then $G_{\mathrm{PC}}(n, p)$ equipped with the cut-off line algorithm enables us to very precisely analyze the sizes of the largest component and the t-core of $G(n, p)$. This new approach for the problems yields not only elegant proofs but also improved bounds that are essentially best possible.

We also consider the Poisson cloning model for random hypergraphs and the t-core problem for random hypergraphs.

Mathematics Subject Classification (2000). Primary 05C80; Secondary 05D40.

Keywords. Random graph, giant component, core, Poisson distribution.

1. Introduction

The notion of a random graph was first introduced in 1947 by Erdős [14] to show the existence of a graph with a certain Ramsey property. A decade later, the theory of random graphs began with the paper entitled *On Random Graphs I* by Erdős and Rényi [15], and they further developed the theory in a series of papers [16], [17], [18], [19], [20]. Since then, the subject has become one of the most active research areas. Many researchers have devoted themselves to studying various properties of random graphs, such as the emergence of the giant component [16], [5], [32], the connectivity [15], [17], [11], the existence of perfect matching [18], [19], [20], [11], the existence of Hamiltonian cycle(s) [31], [6], [10], the k-core problem [6], [34], [38], and the graph invariants like the independence number [9], [36] and the chromatic number [39], [8], [33]. (The list of references here is far from being exhaustive.)

There are two canonical models for random graphs, both of which were originated in the simple model introduced in [14]. In the binomial model $G(n, p)$ on a set V of n vertices, each of $\binom{n}{2}$ possible edges is in the graph with probability p, independently of other edges. Thus, the probability of $G(n, p)$ being a fixed graph G with m edges is $p^m(1-p)^{\binom{n}{2}-m}$. The uniform model $G(n, m)$ on V is a graph chosen uniformly at random from the set of all graphs on V with m edges. Hence $G(n, m)$ becomes a fixed

Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006

graph G with probability $\binom{\binom{n}{2}}{m}^{-1}$, provided G has m edges. Most of the asymptotic behavior of the two models is almost identical if their expected numbers of edges are the same. (See Proposition 1.13 in [27].) The random graph process, in which random edges are added one by one, is also extensively studied. For more about models and/or basics of random graphs we recommend two books with the same title: *Random Graphs* by Bollobás [7] and by Janson, Łuczak and Ruciński [27].

The phase transition phenomenon is one of the interesting topics in random graphs. Specifically, the phase transition phenomena regarding the emergences of the giant (connected) component and the t-core have attracted much attention. In the monumental paper *On the Evolution of Random Graphs* [16], Erdős and Rényi proved that, for the size $\ell_1(n, p)$ of the largest component of $G(n, p)$,

$$\ell_1(n, p) = \begin{cases} O(\log n), & \text{if } \limsup_{n\to\infty} p(n-1) < 1, \\ (1 + o(1))\theta_\lambda n, & \text{if } \lim_{n\to\infty} pn = \lambda > 1, \end{cases}$$

where θ_λ is the positive solution of the equation $1 - \theta - e^{-\lambda\theta} = 0$ and all other components are of size $O(\log n)$.

Why does the size of the largest component change so dramatically around $\lambda = 1$? It was Karp [28] who nicely explained the reason. To find a component $C(v)$ of a fixed vertex v in $G(n, p)$, one may first expose the vertices that are adjacent to v and keep on repeating the same procedure by taking each of those adjacent vertices. Initially, v is active and all other vertices are neutral. At each step we take an active vertex w and expose all neutral vertices adjacent to w. This can be done by checking $\{w, w'\} \in G(n, p)$ or not for all neutral vertices w'. Then activate all neutral vertices that are adjacent to w. The vertex w is no longer active, and only non-activated neutral vertices remain neutral. The process terminates when there is no more active vertex left. Clearly, the process will stop after finding all the vertices in the component containing v. If the number of neutral vertices does not decrease so fast, the number of newly activated vertices is close to the binomial random variable $\mathrm{Bin}(n - 1, p)$, where

$$\Pr[\mathrm{Bin}(n-1, p) = \ell] = \binom{n-1}{\ell} p^\ell (1-p)^{n-1-\ell}.$$

Particularly, the mean number is close to pn. If $pn \leq 1 - \delta$ for a fixed $\delta > 0$, then the process is expected to die out quickly almost every time. Thus, all $C(v)$ are expected to be small. If $pn \geq 1 + \delta$, then the process may survive forever with positive probability. Hence, $C(v)$ can be large with positive probability. As there are n trials, at least one of the $C(v)$'s is expected to be large. Applying this approach to the random directed graph, Karp was able to prove a phase transition phenomenon for the size of the largest strong component.

Notice that, when $pn = \Theta(1)$, the distribution of $\mathrm{Bin}(n - 1, p)$ is very close the Poisson distribution with parameter $\lambda := p(n-1)$. Hence we may further expect that the process described above could be approximated by the Galton–Watson branching process defined by a Poisson random variable $\mathrm{Poi}(\lambda)$ with mean λ, where

$$\Pr[\mathrm{Poi}(\lambda) = \ell] = e^{-\lambda}\frac{\lambda^\ell}{\ell!}.$$

Generally, the Galton–Watson branching process defined by a random variable X starts with a single unisexual organism. The organism will give birth to X_1 children, where X_1 is a random variable with the same distribution as X. The same but independent birth process continues from each of the children, the grandchildren and so on, until no more descendant exists. (For more information regarding Galton–Watson branching processes, one may refer [4].) For simplicity we call the Galton–Watson branching process defined by $\mathrm{Poi}(\lambda)$ the *Poisson*(λ) *branching process*.

The Poisson cloning model. To convert the above observation to a rigorous proof, it is needed to overcome or bypass two main obstacles. Firstly, the degrees of vertices of $G(n, p)$ are not exactly i.i.d. Poisson random variables. Though they have the same distribution as $\mathrm{Bin}(n-1, p)$, they are not mutually independent. For example, the sum of all degrees must be even as it is twice the number of edges, which cannot be guaranteed if the degrees are independent. Secondly, the number of neutral vertices keeps decreasing. Even if both obstacles do not cause substantial differences in many cases, one needs at least to keep tracking small differences for rigorous proofs. Since this kind of small differences occurs almost everywhere in the analysis, it sometimes makes rigorous analysis significantly difficult, if not impossible.

As an approach to bypass the first obstacle, we introduce the Poisson cloning model $G_{\mathrm{PC}}(n, p)$ for random graphs in which the degrees are i.i.d. Poisson random variables with mean $\lambda = p(n-1)$. Moreover, the new model is equivalent to the classical model $G(n, p)$ in an asymptotic sense. Actually, defining the model is not extremely difficult: First take i.i.d. Poisson λ random variables $d(v)'s$ indexed by all vertices in the vertex set V. Then take $d(v)$ copies, or clones, of each vertex v. If the sum of $d(v)$'s is even, then we generate a uniform random perfect matching on the set of all clones. An edge $\{v, w\}$ is in the random graph $G_{\mathrm{PC}}(n, p)$ if a clone of v is matched to a clone of w in the random perfect matching. If the sum is odd, one may just take a graph with a self loop. Hence the graph is not simple if the sum is odd.

It is also possible to extend the model to uniform hypergraphs, where a k-uniform hypergraph on the vertex set V is a collection of subsets of V with size k. A graph is then a 2-uniform hypergraph. In the binomial model $H(n, p\,; k)$ for random k-uniform hypergraphs each of $\binom{n}{k}$ edges is in the hypergraph with probability p, independently of other edges. The Poisson cloning model for random k-uniform hypergraphs is denoted by $H_{\mathrm{PC}}(n, p\,; k)$. In the next section the Poisson cloning model is defined in detail.

The following theorem shows that the Poisson cloning model is essentially equivalent to the binomial model.

Theorem 1.1. *Suppose $k \geq 2$ and $p = \Theta(n^{1-k})$. Then for any collection $\mathcal{H}$ of k-uniform simple hypergraphs,*

$$c_1 \Pr[H_{\mathrm{PC}}(n, p\,; k) \in \mathcal{H}] \leq \Pr[H(n, p\,; k) \in \mathcal{H}] \leq c_2\big(\Pr[H_{\mathrm{PC}}(n, p\,; k) \in \mathcal{H}]^{\frac{1}{k}} + e^{-n}\big),$$

where

$$c_1 = k^{1/2} e^{\frac{p}{n}\binom{k}{2}\binom{n}{k} + \frac{p^2}{2}\binom{n}{k}} + O(n^{-1/2}), \quad c_2 = \Big(\frac{k}{k-1}\Big)\big(c_1(k-1)\big)^{1/k} + o(1),$$

and $o(1)$ goes to 0 as n goes to infinity.

To overcome the second obstacle we present an algorithm, called the cut-off line algorithm, that enables us to generate the Poisson cloning model and analyze problems simultaneously. As a consequence the size of the largest component of $G_{\mathrm{PC}}(n, p)$ can be described very precisely. It is also possible to analyze the size of t-core of the random hypergraph $H_{\mathrm{PC}}(n, p\,; k)$, where the t-core of a hypergraph H is the largest subhypergraph of H with minimum degree at least t.

The emergence of the giant component. After the phase transition result of Erdős and Rényi it remained to determine the size of the largest component when $pn \to 1$. Though Erdős and Rényi suggested that the size $\ell_1(n, p)$ of the largest component could be one of $O(\log n)$, $\Theta(n^{2/3})$, and $\Theta(n)$, Bollobás [5] showed that $\ell_1(n, p)$ increases rather continuously by estimating it quite accurately for $pn - 1 \geq n^{-1/3}\sqrt{\log n}/2$. Later Łuczak [32] was able to estimate $\ell_1(n, p)$ for $pn - 1 \gg n^{-1/3}$.

Before stating the result of Łuczak a convention is introduced: when the expression $x \gg y$ is used as part of the hypotheses, it means 'there exists a (large) constant $K > 0$ so that, if $x \geq Ky \ldots$'. We also denote $\lambda(n, p) = p(n-1)$.

Theorem 1.2 (Supercritical phase). *Suppose $\lambda = \lambda(p, n) = 1 + \varepsilon$ with $\varepsilon \gg n^{-1/3}$. Then for large enough n, with probability at least $1 - 7(\varepsilon^3 n/8)^{-1/9}$*

$$|\ell_1(n, p) - \theta_\lambda n| \leq \frac{n^{2/3}}{5},$$

and all other components are smaller than $n^{2/3}$.

Using estimations for the numbers of connected graphs with certain number of vertices and edges and the first and second moment methods, one may also obtain the following result for the subcritical phase.

Theorem 1.3 (Subcritical phase). *Let $\lambda(n, p) = 1 - \varepsilon$ with $n^{-1/3} \ll \varepsilon \ll 1$. Then for any positive constant $\delta \leq 1/3$ and large enough n, with probability at least $1 - \big(\frac{8}{\varepsilon^3 n}\big)^{\delta/4}$*

$$\left|\ell_1(n, p) - \frac{2\log(\varepsilon^3 n)}{\varepsilon^2}\right| \leq \frac{\delta \log(\varepsilon^3 n)}{\varepsilon^2}.$$

For results regarding the structure of the largest component readers are referred to [27], [32], [25], [35] and references therein.

For Poisson branching processes a duality principle has been known. A pair (μ, λ) with $\mu < 1 < \lambda$ is called a *conjugate pair* if $\mu e^{-\mu} = \lambda e^{-\lambda}$. It is easy to see that $\mu = (1-\theta_\lambda)\lambda$ for a conjugate pair (μ, λ). For a conjugate pair (μ, λ) the distribution of the Poisson(λ) branching process conditioned that the process dies out is exactly the same as that of the Poisson(μ) branching process. (See e.g. [2], p. 164.) A similar duality was observed for the random graph $G(n, p)$ and $G(n^*, p)$ with $\lambda = \lambda(n, p) > 1$ and $n^* = (1-\theta_\lambda)n$. (Recall that $\lambda(n, p) = p(n-1)$.) Notice that $1-\theta_\lambda$ is the extinction probability for the Poisson(λ) branching process. It is known that the component sizes of $G(n^*, p)$ and those of $G(n, p)$ excluding the largest component are the same in an asymptotic sense (see [2]).

The Poisson cloning model $G_{\text{PC}}(n, p)$ equipped with the cut-off line algorithm enables us to not only estimate $\ell_1(n, p)$ more accurately but also to establish a discrete duality principle: in the supercritical phase $\lambda := \lambda(n, p) = 1+\varepsilon$ with $n^{-1/3} \ll \varepsilon \ll 1$; $G_{\text{PC}}(n, p)$ can be decomposed by three vertex disjoint graphs C, S and G whp (with high probability), where C is a connected graph of size about $\theta_\lambda n$, S is a smaller graph of size about $\varepsilon^{-2} \ll \theta_\lambda n$, and G has the same distribution as $G_{\text{PC}}(n^*, p^*)$ with $n^* \approx (1-\theta_\lambda)n$ and $p^* \approx p$, which yields $\lambda(n^*, p^*) \approx \mu := (1-\theta_\lambda)\lambda$. In the subcritical phase $\lambda = 1-\varepsilon$ with $n^{-1/3} \ll \varepsilon \ll 1$ the largest component is of size

$$\frac{\log(\varepsilon^3 n) - 2.5 \log\log(\varepsilon^3 n) + O(1)}{-(\varepsilon + \log(1-\varepsilon))}$$

whp. The precise statements are as follows. We concentrate on the cases $\varepsilon \ll 1$ for which more careful analysis is required. It is believed that the proofs are easily modified for the cases of positive constants ε.

Theorem 1.4. *Supercritical phase: Let* $\lambda := \lambda(n, p) = 1+\varepsilon$ *with* $n^{-1/3} \ll \varepsilon \ll 1$, $\mu := (1-\theta_\lambda)\lambda$ *and* $1 \ll \alpha \ll (\varepsilon^3 n)^{1/2}$. *Then with probability* $1 - e^{-\Omega(\alpha^2)}$ $G_{\text{PC}}(n, p)$ *may be decomposed by three vertex disjoint graphs C, S and G, where C is connected and*

$$\theta_\lambda n - \alpha(n/\varepsilon)^{1/2} \le |C| \le \theta_\lambda n + \alpha(n/\varepsilon)^{1/2},$$

$|S| \le \frac{\alpha^2}{\varepsilon^2}$, *G has the same distribution as* $G_{\text{PC}}(n^*, p^*)$ *for some* n^* *and* p^* *satisfying*

$$(1-\theta_\lambda)n - \alpha(n/\varepsilon)^{1/2} \le n^* \le (1-\theta_\lambda)n + \alpha(n/\varepsilon)^{1/2},$$

and

$$\mu - \alpha(\varepsilon n)^{-1/2} \le \lambda(n^*, p^*) \le \mu + \alpha(\varepsilon n)^{-1/2}.$$

Subcritical phase: Suppose $\lambda := \lambda(n, p) = 1-\varepsilon$ *with* $n^{-1/3} \ll \varepsilon \ll 1$. *Then the size* $\ell_1^{\text{PC}}(n, p)$ *of the largest component of* $G_{\text{PC}}(n, p)$ *satisfies*

$$\Pr\left[\ell_1^{\text{PC}}(n, p) \ge \frac{\log(\varepsilon^3 n) - 2.5\log\log(\varepsilon^3 n) + c}{-(\varepsilon + \log(1-\varepsilon))}\right] \le 2e^{-\Omega(c)},$$

and

$$\Pr\left[\ell_1^{\mathrm{PC}}(n,p) \le \frac{\log(\varepsilon^3 n) - 2.5\log\log(\varepsilon^3 n) - c}{-(\varepsilon + \log(1-\varepsilon))}\right] \le 2e^{-e^{\Omega(c)}}$$

for any constant $c > 0$.

Inside window: Suppose $\lambda := \lambda(n,p) = 1+\varepsilon$ *with* $|\varepsilon| = O(n^{1/3})$. *Then whp*

$$\ell_1^{\mathrm{PC}}(n,p) = \Theta(n^{2/3}).$$

(*All constants in the* $\Omega(\,\cdot\,)$*'s do not depend on any of* ε, α *and* c.)

By Theorem 1.1 a corollary regarding $G(n,p)$ follows.

Corollary 1.5. *Supercritical region: Suppose* $\lambda = \lambda(n,p) = 1+\varepsilon$ *with* $n^{-1/3} \ll \varepsilon \ll 1$, *and* $1 \ll \alpha \ll (\varepsilon^3 n)^{1/2}$. *Then, in* $G(n,p)$,

$$\Pr[\,|\ell_1(n,p) - \theta_\lambda n| \ge \alpha(n/\varepsilon)^{1/2}] \le 2e^{-\Omega(\alpha^2)}.$$

Moreover, for the size $\ell_2(n,p)$ *of the second largest component and* $\varepsilon^* = 1-(1-\theta_\lambda)\lambda$,

$$\Pr\left[\ell_2(n,p) \ge \frac{\log((\varepsilon^*)^3 n) - 2.5\log\log((\varepsilon^*)^3 n) + c}{-(\varepsilon^* + \log(1-\varepsilon^*))}\right] \le 2e^{-\Omega(c)}$$

and

$$\Pr\left[\ell_2(n,p) \le \frac{\log((\varepsilon^*)^3 n) - 2.5\log\log((\varepsilon^*)^3 n) - c}{-(\varepsilon^* + \log(1-\varepsilon^*))}\right] \le 2e^{-e^{\Omega(c)}},$$

for any constant $c > 0$.

Subcritical region: Suppose $\lambda = 1-\varepsilon$ *with* $n^{-1/3} \ll \varepsilon \ll 1$. *Then for any constant* $c > 0$,

$$\Pr\left[\ell_1(n,p) \ge \frac{\log(\varepsilon^3 n) - 2.5\log\log(\varepsilon^3 n) + c}{-(\varepsilon + \log(1-\varepsilon))}\right] \le 2e^{-\Omega(c)}$$

and

$$\Pr\left[\ell_1(n,p) \le \frac{\log(\varepsilon^3 n) - 2.5\log\log(\varepsilon^3 n) - c}{-(\varepsilon + \log(1-\varepsilon))}\right] \le 2e^{-e^{\Omega(c)}}.$$

Inside window: Suppose $\lambda := \lambda(n,p) = 1+\varepsilon$ *with* $|\varepsilon| = O(n^{1/3})$. *Then whp*

$$\ell_1(n,p) = \Theta(n^{2/3}).$$

The emergence of the *t*-core. There are at least two possible directions to extend the problem of (connected) component. Observing that the minimum degree in a component must be larger than or equal to 1, one may consider subgraphs with minimum degree at least $t \ge 2$. For a graph G, the t-core is the largest subgraph with minimum degree at least t. Since the minimum degree of the union of two subgraphs is larger than or equal to the smaller minimum degree of the two, the t-core of a graph is unique. It is also easy to see that the t-core must be an induced subgraph. For this reason the t-core of G sometimes refers to its vertex set. Denote by $V_t(G)$ (the vertex set of) the t-core of G. As the 1-core $V_1(G)$ is the set of all non-isolated vertices, we consider the cases $t \geq 2$ throughout this paper. If there is no subgraph with minimum degree t, the t-core is defined to be empty.

Another direction is to consider the t-connectivity, where a graph is t-connected if the graph remains connected after any $t-1$ vertices are removed. Higher orders of connectivity have been used to understand various structures of graphs. Clearly, if a non-empty subgraph is t-connected, then its minimum degree must be t or larger.

In 1984, Bollobás [6] initiated the study of t-core, $t \geq 2$, and observed that, provided $t \geq 3$ and pn is larger than a fixed constant, the t-core of $G(n, p)$ is non-empty and t-connected whp. Łuczak [34] proved that for $t \geq 3$ there is an absolute constant c such that the t-core of $G(n, p)$ is either empty, or larger than cn and t-connected, whp. In particular, as far as the random graph $G(n.p)$ is concerned, the t-core problem is the same as the t-connectivity problem. Moreover, if $\lambda(n, p)$ is less than 1, then the t-core of $G(n, p)$ is empty whp since the size of the largest component is $O(n^{2/3})$ whp. As p increases while n is fixed, the probability of the t-core of $G(n, p)$ being non-empty keeps increasing. Let $p_t(n, \delta)$ be the infimum of all p that makes the probability larger than or equal to a constant δ with $0 < \delta < 1$. Then Bollobás's result implies that $np_t(n, \delta)$ is bounded from above by a constant. Though $np_t(n, \delta)$ may still have no limit value as n goes to infinity, it seems to be more natural to expect that the limit exists. Furthermore, as it happens often in phase transition phenomena, the limit, if it exists, is also expected to be independent of δ. In other words, the phase transition is expected to be sharp.

For $t = 2$, the 2-core of a graph G is non-empty if and only if G contains a cycle. It is easy to see by the first moment method that $G(n, p)$ with $p = o(1/n)$ does not contain a cycle whp. For a constant c with $0 < c < 1$, $G(n, p)$ may or may not have a cycle with positive probability. In particular, the phase transition for the existence of a non-empty 2-core is not sharp. In the graph process $(G(n, m))_{m=0,1,\ldots}$ in which a random edge is added one by one without repetition, Janson [24] found the limiting distribution for the length of the first cycle, especially he showed that the length is bounded whp. However, the expectation of the length is known to be $\Theta(n^{1/6})$ due to Flajolet et al. [23]. The two facts do not contradict each other, since there are random variables X that are bounded whp, but $E[X]$ is not. For example, $\Pr[X = 1] = 1 - 1/n$ and $\Pr[X = n^2] = 1/n$.

Bollobás [6] proved that, if $t \geq 5$ and $\lambda(n, p) := p(n-1) \geq \max\{67, 2t + 6\}$, then $G(n, p)$ has a non-empty t-core. Chvátal [12] introduced the notion of critical λ_t, without proving existence, satisfying the following. As n goes to infinity,

$$\Pr[G(n, p) \text{ has a non-empty } t\text{-core }] \longrightarrow \begin{cases} 0 & \text{if } \lambda(n, p) < \lambda_t - \delta, \\ 1 & \text{if } \lambda(n, p) > \lambda_t + \delta \end{cases}$$

for any constant $\delta > 0$. He also proved that $\lambda_3 \geq 2.88$ if it exists, and claimed that

$\lambda_4 \geq 4.52$ and $\lambda_5 \geq 6.06$ etc. could be proven by the same method. Pittel, Spencer and Wormald [38] proved a general theorem which implies that λ_t exists for fixed $t \geq 3$ and identified their values. We present a slightly weaker version of this result.

For a Poisson random variable $\mathrm{Poi}(\rho)$ with mean ρ let $P(\rho, i) = \Pr[\mathrm{Poi}(\rho) = i]$ and $Q(\rho, i) = \Pr[\mathrm{Poi}(\rho) \geq i]$, i.e.,

$$P(\rho, i) = e^{-\rho}\frac{\rho^i}{i!} \quad \text{and} \quad Q(\rho, i) := \sum_{j=i}^{\infty} P(\rho, j) = e^{-\rho}\sum_{j=i}^{\infty}\frac{\rho^j}{j!},$$

and let

$$\lambda_t = \min_{\rho>0}\frac{\rho}{Q(\rho, t-1)}.$$

Theorem 1.6. *Let $t \geq 3$, $\lambda(n, p) = p(n-1)$. Then*

$$\Pr[G(n,p) \textit{ has a non-empty } t\textit{-core}] \longrightarrow \begin{cases} 0 & \textit{if } \lambda(n,p) < \lambda_t - n^{-\delta}, \\ 1 & \textit{if } \lambda(n,p) > \lambda_t + n^{-\delta} \end{cases}$$

for any $\delta \in (0, 1/2)$, and the t-core when $\lambda(n, p) > \lambda_t + n^{-\delta}$ has $(1+o(1))Q(\theta_\lambda \lambda, t)n$ vertices, whp, where θ_λ is the largest solution for the equation

$$\theta - Q(\theta\lambda, t-1) = 0.$$

There are many studies about the t-cores of various types of random graphs and random hypergraphs. Fernholz and Ramachandran [21], [22] studied random graph conditions on given degree sequences. Cooper [13] found the critical values for t-cores of a uniform multihypergraph with given degree sequences that includes the random k-uniform hypergraph $H(n, p; k)$. Molloy [37] considered cores for random hypergraphs and random satisfiability problems for Boolean formulas. Recently, S. Janson and M. J. Luczak [26] also gave seemingly simpler proofs for t-core problems that contain the result of Pittel, Spencer and Wormald. For more information and techniques used in the above mentioned papers readers are referred to [26].

Using the Poisson cloning model for random hypergraphs together with the cut-off line algorithm we are able to completely analyze the t-core problem for the random uniform hypergraph. We also believe that the cut-off line algorithm can be used to analyze the t-core problem for random hypergraphs conditioned on certain degree sequences as in [13], [21], [22], [26].

As the 2-core of $G(n, p)$ behaves quite differently from the other t-cores of $H(n, p; k)$, we exclude the case $k = t = 2$, which will be studied in a subsequent paper. The critical value for the problem turns out to be the minimum λ such that there is a positive solution for the equation

$$\theta - Q(\theta^{k-1}\lambda, t-1) = 0. \tag{1.1}$$

It is not difficult to check that the minimum is

$$\lambda_{\mathrm{crt}}(k, t) := \min_{\rho>0}\frac{\rho}{Q(\rho, t-1)^{k-1}}. \tag{1.2}$$

For $\lambda > \lambda_{\text{crt}}(k,t)$, let θ_λ be the largest solution of the equation $\theta^{\frac{1}{k-1}} - Q(\theta\lambda, t-1) = 0$.

Theorem 1.7. *Let $k, t \geq 2$, excluding $k = t = 2$, and $\sigma \gg n^{-1/2}$.*

Subcritical phase: If $\lambda(n, p\,; k) := p\binom{n-1}{k-1} = \lambda_{\text{crt}} - \sigma$ is uniformly bounded from below by 0 and $i_0(k,t)$ is the minimum i such that $\binom{i}{k} \geq ti/k$, then

$$\Pr[V_t(H(n, p\,; k)) \neq \emptyset] = e^{-\Omega(\sigma^2 n)} + O\big(n^{-(t-1-t/k)i_0(k,t)}\big),$$

and for any $\delta > 0$,

$$\Pr[|V_t(H(n, p\,; k))| \geq \delta n] = e^{-\Omega(\sigma^2 n)} + e^{-\Omega(\delta^{2k/(k-1)} n)}. \tag{1.3}$$

Supercritical phase: If $\lambda = \lambda(n, p\,; k) = \lambda_{\text{crt}} + \sigma$ is uniformly bounded from above, then for all α in the range $1 \ll \alpha \ll \sigma n^{1/2}$,

$$\Pr\Big[\big||V_t(n, p\,; k)| - Q(\theta_\lambda\lambda, t)n\big| \geq \alpha(n/\sigma)^{1/2}\Big] = e^{-\Omega(\alpha^2)}, \tag{1.4}$$

and, for any $i \geq t$ and the sets $V_t(i)$ (resp. $W_t(i)$) of vertices of degree i (resp. larger than or equal to i) in the t-core,

$$\Pr\Big[\big||V_t(i)| - P(\theta_\lambda\lambda, i)n\big| \geq \delta n\Big] \leq 2e^{-\Omega(\min\{\delta^2\sigma n, \sigma^2 n\})},$$

and

$$\Pr\Big[\big||W_t(i)| - Q(\theta_\lambda\lambda, i)n\big| \geq \delta n\Big] \leq 2e^{-\Omega(\min\{\delta^2\sigma n, \sigma^2 n\})}.$$

In particular, for $\lambda = \lambda_{\text{crt}} + \sigma$ and $\rho_{\text{crt}} := \theta_{\lambda_{\text{crt}}(k,t)}\lambda_{\text{crt}}(k,t)$,

$$|V_t(i)| = \big(1 + O(\sigma^{1/2})\big)^i P(\rho_{\text{crt}}, i)n + O\big((n/\sigma)^{1/2}\log n\big),$$

with probability $1 - 2e^{-\Omega(\min\{\log^2 n, \sigma^2 n\})}$.

As one might guess, we will prove a stronger theorem (Theorem 6.2) for the Poisson cloning model $H_{\text{PC}}(n, p\,; k)$, from which Theorem 1.7 easily follows.

In the next section the Poisson cloning model is defined in detail. The cut-off line algorithm and the cut-off line lemma are presented in Section 3. In Section 4 we study Chernoff type large derivation inequalities that will be used in most of our proofs. In Section 5, a generalized core is defined and the main lemma is presented. Section 6 is devoted to the proof of Theorem 1.7. As the proof of Theorem 1.4 is more sophisticated, we only give the proof ideas in Section 7. We conclude this paper with final remarks in Section 8. Due to the space limitation, many proofs are omitted. They can be found on the author's web site.

2. The Poisson cloning model

To construct the Poisson cloning model $G_{\mathrm{PC}}(n, p)$ for random graphs, let V be a set of n vertices. We take i.i.d. Poisson $\lambda = p(n-1)$ random variables $d(v)$, $v \in V$, and then take $d(v)$ copies of each vertex $v \in V$. The copies of v are called *clones of* v, or simply v*-clones*. Since the sum of Poisson random variables is also a Poisson random variable, the total number $N_\lambda := \sum_{v \in V} d(v)$ of clones is a Poisson λn random variable. It is sometimes convenient to take a reverse, but equivalent, construction. We first take a Poisson $\lambda n = 2p\binom{n}{2}$ random variables N_λ and then take N_λ unlabelled clones. Each clone is independently labelled as v-clone uniformly at random, in the sense that v is chosen uniformly at random from V. It is well known that the numbers $d(v)$ of v-clones are i.i.d. Poisson random variables with mean λ.

If N_λ is even, the multigraph $G_{\mathrm{PC}}(n, p)$ is defined by generating a (uniform) random perfect matching of those N_λ clones, and contracting clones of the same vertex. That is, if a v-clone and a w-clone are matched, then the edge $\{v, w\}$ is in $G_{\mathrm{PC}}(n, p)$ with multiplicity. In the case that $v = w$, it produces a loop that contributes 2 to the degree of v. If N_λ is odd, we may define $G_{\mathrm{PC}}(n, p)$ to be any graph with a special loop that, unlike other loops, contributes only 1 to the degree of the corresponding vertex. In particular, if N_λ is odd, $G_{\mathrm{PC}}(n, p)$ is not a simple graph.

Strictly speaking, $G_{\mathrm{PC}}(n, p)$ varies depending on how to define it when N_λ is odd. However, if only simple graphs are concerned, the case of N_λ being odd would not matter. For example, the probability that $G_{\mathrm{PC}}(n, p)$ is a simple graph with a component larger than $0.1n$ does not depend on how $G_{\mathrm{PC}}(n, p)$ is defined when N_λ is odd, as it is not a simple graph anyway. Generally, for any collection $\mathcal{G}$ of simple graphs, the probability that $G_{\mathrm{PC}}(n, p)$ is in $\mathcal{G}$ is totally independent of how $G_{\mathrm{PC}}(n, p)$ is defined when N_λ is odd. Notice that properties of simple graphs are actually mean collections of simple graphs. Therefore, when properties of simple graphs are concerned, it is not necessary to describe $G_{\mathrm{PC}}(n, p)$ for odd N_λ.

Here are two specific ways to generate the uniform random matching.

Example 2.1. One may keep matching two clones chosen uniformly at random among all unmatched clones.

Example 2.2. One may keep choosing his or her favorite unmatched clone, and matching it to a clone selected uniformly at random from all other unmatched clones.

If N_λ is even both examples would yield uniform random perfect matchings. If N_λ is odd, then each of them would yield a matching and an unmatched clone. We may create the special loop consisting of the vertex for which the unmatched clone is labelled. More specific ways to choose random clones will be described in the next section.

Generally for $k \geq 3$, the Poisson cloning model $H_{\mathrm{PC}}(n, p\,; k)$ for k-uniform hypergraphs may be defined in the same way: We take i.i.d. Poisson $\lambda = p\binom{n-1}{k-1}$ random variables $d(v)$, $v \in V$, and then take $d(v)$ clones of each v. If $N_\lambda := \sum_{v \in V} d(v)$ is divisible by k, the multihypergraph $H_{\text{PC}}(n, p; k)$ is defined by generating a uniform random perfect matching consisting of k-tuples of those N_λ clones, and contracting clones of the same vertex. That is, if v_1-clone, v_2-clone, …, v_k-clone are matched in the perfect matching, then the edge $\{v_1, v_2, \ldots, v_k\}$ is in $H_{\text{PC}}(n, p\,; k)$ with multiplicity. If N_λ is not divisible by k, $H_{\text{PC}}(n, p\,; k)$ may be any hypergraph with a special edge consisting of $N_\lambda - k\lfloor N_\lambda/k\rfloor$ vertices. In particular, $H_{\text{PC}}(n, p\,; k)$ is not k-uniform when N_λ is not divisible by k. Therefore, as long as properties of k-uniform hypergraphs are concerned, we do not have to describe $H_{\text{PC}}(n, p\,; k)$ when N_λ is not divisible by k.

We show that the Poisson cloning model $H_{\text{PC}}(n, p\,; k)$, $k \geq 2$, is contiguous to the classical model $H(n, p\,; k)$ when the expected average degree is a constant.

Theorem 1.1 (restated). *Suppose $k \geq 2$ and $p = \Theta(n^{1-k})$. Then for any collection $\mathcal{H}$ of k-uniform simple hypergraphs,*

$$c_1 \Pr[H_{\text{PC}}(n, p\,; k) \in \mathcal{H}] \leq \Pr[H(n, p\,; k) \in \mathcal{H}] \leq c_2\big(\Pr[H_{\text{PC}}(n, p\,; k) \in \mathcal{H}]^{\frac{1}{k}} + e^{-n}\big),$$

where

$$c_1 = k^{1/2} e^{\frac{p}{n}\binom{k}{2}\binom{n}{k} + \frac{p^2}{2}\binom{n}{k}} + O(n^{-1/2}), \quad c_2 = \Big(\frac{k}{k-1}\Big)\Big(c_1(k-1)\Big)^{1/k} + o(1),$$

and $o(1)$ goes to 0 as n goes to infinity.

Proof. See [30]. □

3. The λ-cell and the cut-off line algorithm

To generate a uniform random perfect matching of N_λ clones, we may keep matching k unmatched clones uniformly at random (cf. Example 2.1). Another way is to choose the first clone as we like and match it to $k-1$ clones selected uniformly at random among all other unmatched clones (cf. Example 2.2). As there are many ways to choose the first clone, we may take a way that makes the given problem easier to analyze. Formally, a sequence $\mathcal{S} = (S_i)$ of choice functions determines how to choose the first clone at each step, where S_i tells us which unmatched clone is to be the first clone for the i^{th} edge in the random perfect matching. A choice function may be deterministic or random. If less than k clones remain unmatched, the edge consisting of those clones will be added. The clone chosen by S_i is called the i^{th} chosen clone, or simply a chosen clone.

We also present a more specific way to select the $k-1$ random clones to be matched to the chosen clone. The way introduced here will be useful to solve problems mentioned in the introduction. First, independently assign to each clone a uniform random real number between 0 and $\lambda = p\binom{n-1}{k-1}$. For the sake of convenience, a clone is called the largest, the smallest, etc. if so is the number assigned to it. In addition, a clone is called $\theta\lambda$-*large* (resp. $\theta\lambda$-small) if its assigned number is larger than or equal to (resp. smaller than) $\theta\lambda$. To visualize the labelled clones with assigned numbers, one may consider n horizontal line segments from $(0, j)$ to (λ, j), $j = 0, \ldots, n-1$ in the two-dimensional plane $\mathbb{R}^2$. The v_j-clone with assigned number x can be regarded as the point (x, j) in the corresponding line segment. Then each line segment with the points corresponding to clones with assigned numbers is an independent Poisson arrival process with density 1, up to time λ. The set of these Poisson arrival processes is called a *Poisson* (λ, n)-*cell* or simply a λ-*cell*.

We will consider sequences of choice functions that choose an unmatched clone without changing the joint distribution of the numbers assigned to all other unmatched clones. Such a choice function is called oblivious. A sequence of oblivious choice functions is also called oblivious. The choice function that chooses the largest unmatched clone is not oblivious, as the numbers assigned to the other clones must be smaller than the largest assigned number. As an example of an oblivious choice function one may consider the choice function that chooses a v-clone for a vertex v with fewer than 3 unmatched clones. For a more general example, let a vertex v and its clones be called t-*light* if there are fewer than t unmatched v-clones.

Example 3.1. Suppose there is an order of all clones which is independent of the assigned numbers. The sequence of the choice functions that choose the first t-light clone is oblivious.

A cut-off line algorithm is determined by an oblivious sequence of choice functions. Once a clone is obliviously chosen, the largest $k-1$ clones among all unmatched clones are to be matched to the chosen clone. This may be further implemented by moving the cut-off line to the left until $k-1$ vertices are found: Initially, the cut-off line of the λ-cell is the vertical line in $\mathbb{R}^2$ containing the point $(\lambda, 0)$. The initial cut-off value, or cut-off number, is λ. At the first step, once the chosen clone is given, move the cut-off line to the left until exactly $k-1$ unmatched clones, excluding the chosen clone, are on or in the right side of the line. The new cut-off value, which is denoted by Λ_1, is to be the number assigned to the $(k-1)^{\text{th}}$ largest clone. The new cut-off line is, of course, the vertical line containing $(\Lambda_1, 0)$. Repeating this procedure, one may obtain the i^{th} cut-off value Λ_i and the corresponding cut-off line.

Notice that, after the i^{th} step ends with the cut-off value Λ_i, all numbers assigned to unmatched clones are i.i.d. uniform random numbers between 0 to Λ_i, as the choice functions are oblivious. Let N_i be the number of unmatched clones after step i. That is, $N_i = N_\lambda - ik$. Since the $(i+1)^{\text{th}}$ choice function tells how to choose the first clone to form the $(i+1)^{\text{th}}$ edge without changing the distribution of the assigned numbers, the distribution of Λ_{i+1} is the distribution of the $(k-1)^{\text{th}}$ largest number among $N_i - 1$ independent uniform random numbers between 0 and Λ_i. Let $1 - T_j$ be the random variable representing the largest number among j independent uniform random numbers between 0 and 1. Or equivalently, T_j is the random variable representing the smallest number among the random numbers. Then the largest number among the $N_i - 1$ random numbers has the same distribution as $\Lambda_i(1 - T_{N_i-1})$. Repeating this $k-1$ times, we have

$$\Lambda_{i+1} = \Lambda_i(1 - T_{N_i-1})(1 - T_{N_i-2}) \dots (1 - T_{N_i-k+1}),$$

and hence

$$\begin{aligned}\Lambda_{i+1} &= \Lambda_i(1 - T_{N_i-1}) \dots (1 - T_{N_i-k+1}) \\ &= \Lambda_{i-1}(1 - T_{N_{i-1}-1}) \dots (1 - T_{N_{i-1}-k+1}) \cdot (1 - T_{N_i-1}) \dots (1 - T_{N_i-k+1}) \\ &= \lambda \prod_{\substack{j=N_\lambda-1 \\ k \nmid N_\lambda - j}}^{N_\lambda-(i+1)k+1} (1 - T_j).\end{aligned}$$

It is crucial to observe that, once N_λ is given, all T_i are mutually independent random variables. This makes the random variable Λ_i highly concentrated near its mean, which enables us to develop theories as if Λ_i were a constant. The cut-off value Λ_i will provide enough information to resolve some otherwise difficult problems.

For θ in the range $0 \le \theta \le 1$, let $\Lambda(\theta)$ be the cut-off value when $\big(1 - \theta^{\frac{k}{k-1}}\big)\lambda n$ or more clones are matched for the first time. Conversely, let $N(\theta)$ be the number of matched clones until the cut-off line reaches $\theta\lambda$.

Lemma 3.2 (Cut-off line lemma). *Let $k \ge 2$ and $\lambda > 0$ be fixed. Then for $\theta_1 < 1$ uniformly bounded below from 0 and $0 < \Delta \le n$,*

$$\Pr\Big[\max_{\theta:\theta_1 \le \theta \le 1} |\Lambda(\theta) - \theta\lambda| \ge \tfrac{\Delta}{n}\Big] \le 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{(1-\theta_1)n}\})}$$

and

$$\Pr\Big[\max_{\theta:\theta_1 \le \theta \le 1} |N(\theta) - (1 - \theta^{\frac{k}{k-1}})\lambda n| \ge \Delta\Big] \le 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{(1-\theta_1)n}\})}.$$

Proof. See [30]. □

4. Large deviation inequalities

In this section a generalized Chernoff bound and an inequality for random process is given. Let $X_1, \dots, X_m$ be a sequence of random variables such that the distribution of X_i is determined if all the values of $X_1, \dots, X_{i-1}$ are known. For example, $X_i = \Lambda(\theta_i)$ with $1 \ge \theta_1 \ge \dots \ge \theta_m \ge 0$ in a Poisson λ-cell. If the upper and/or lower bounds are known for the conditional means $E[X_i|X_1, \dots, X_{i-1}]$ and for the conditional second and third monuments, then Chernoff type large deviation inequalities may be obtained not only for $\sum_{j=1}^m X_j$ but for $\min_{1\le i \le m} \sum_{j=1}^i X_j$ and/or

$\max_{1\le i\le m}\sum_{j=1}^{i} X_j$. Large deviation inequalities for such minima or maxima are especially useful in various situations. Lemma 3.2 can be shown using such inequalities too.

Lemma 4.1. *Let* $X_1,\ldots,X_m$ *be a sequence of random variables. Suppose that*

$$E[X_i|X_1,\ldots,X_{i-1}]\le \mu_i, \tag{4.1}$$

and that there are positive constants a_i *and* b_i *such that*

$$E[(X_i-\mu_i)^2|X_1,\ldots,X_{i-1}]\le a_i, \tag{4.2}$$

and

$$E[(X_i-\mu_i)^3 e^{\xi(X_i-\mu_i)}|X_1,\ldots,X_{i-1}]\le b_i \quad \textit{for all } 0\le\xi\le\xi_0. \tag{4.3}$$

Then for any α *with* $0<\alpha\le\xi_0\big(\sum_{i=1}^m a_i\big)^{1/2}$,

$$\Pr\Big[\sum_{i=1}^m X_i \ge \sum_{i=1}^m \mu_i + \alpha\Big(\sum_{i=1}^m a_i\Big)^{1/2}\Big] \le \exp\bigg(-\frac{\alpha^2}{2}\bigg(1+\frac{\alpha\sum_{i=1}^m b_i}{3\big(\sum_{i=1}^m a_i\big)^{3/2}}\bigg)\bigg).$$

Similarly,

$$E[X_i|X_1,\ldots,X_{i-1}]\ge \mu_i \tag{4.4}$$

together with (4.2) *and*

$$E[(X_i-\mu_i)^3 e^{\xi(X_i-\mu_i)}|X_1,\ldots,X_{i-1}]\ge b_i \quad \textit{for all } \xi_0\le\xi<0 \tag{4.5}$$

implies that

$$\Pr\Big[\sum_{i=1}^m X_i \le \sum_{i=1}^m \mu_i - \alpha\Big(\sum_{i=1}^m a_i\Big)^{1/2}\Big] \le \exp\bigg(-\frac{\alpha^2}{2}\bigg(1-\frac{\alpha\sum_{i=1}^m b_i}{3\big(\sum_{i=1}^m a_i\big)^{3/2}}\bigg)\bigg).$$

Proof. See [30]. □

As it is sometimes tedious to point out the value of α and to check the required bounds for it, the following forms of inequalities are often more convenient.

Corollary 4.2 (Generalized Chernoff bound). *If* $\delta\xi_0\sum b_i\le\sum a_i$ *for some* $0<\delta\le 1$, *then* (4.1)–(4.3) *imply*

$$\Pr\Big[\sum_{i=1}^m X_i \ge \sum_{i=1}^m \mu_i + R\Big] \le e^{-\frac{1}{3}\min\{\delta\xi_0 R,\ R^2/\sum_{i=1}^m a_i\}}$$

for all $R>0$. *Similarly, if* $-\delta\xi_0\sum b_i\le\sum a_i$ *for some* $0<\delta\le 1$, *then* (4.2), (4.4) *and* (4.5) *yield*

$$\Pr\Big[\sum_{i=1}^m X_i \le \sum_{i=1}^m \mu_i - R\Big] \le e^{-\frac{1}{3}\min\{\delta\xi_0 R,\ R^2/\sum_{i=1}^m a_i\}}$$

for all $R>0$.

Let $X_\theta, \theta \geq 0$, be random variables which are possibly set-valued. Here θ may be integers as well as real numbers. Suppose that $\Gamma(\theta)$ is a random variable depending on $\{X_{\theta'}\}_{\theta' \leq \theta}$ and θ, and

$$\psi = \psi(\{X_{\theta'}\}_{\theta' \leq \theta_1}; \theta_0, \theta_1) \quad \text{and} \quad \psi_\theta = \psi_\theta(\{X_{\theta'}\}_{\theta' \leq \theta_1}; \theta_0, \theta, \theta_1).$$

The random variables ψ and ψ_θ are used to bound $\Gamma(\theta)$.

Example 4.3. Let $X_1, X_2, \ldots$ be i.i.d. Bernoulli random variables with mean p and $S_i = \sum_{j=1}^{i} X_j$. Set $\Gamma(i) = |S_i - ip|$ and

$$\psi = \Gamma(n) \quad \text{and} \quad \psi_i = |S_n - S_i - (n-i)p|.$$

Then, since

$$S_i - ip = S_n - np - (S_n - S_i - (n-i)p)$$

we have

$$\Gamma(i) \leq \psi + \psi_i.$$

Example 4.4. Consider the (λ, n)-cell defined in the previous section. Let v_θ be the vertex that has its largest clone at $(1-\theta)\lambda$. If such a vertex does not exist, v_θ is defined to be $\aleph$, assuming $\aleph \notin V$. As there is no possibility that two distinct clones are assigned the same number, v_θ is well-defined. Let $X_\theta = v_\theta$ and $V(\theta)$ be the set of vertices that contain no clone larger than or equal to $(1-\theta)\lambda$. That is, $V(\theta) = V \setminus \{v_{\theta'} : 0 \leq \theta' \leq \theta\}$. Clearly, $E[|V(\theta)|] = e^{-\theta\lambda} n$. Observing that for $\theta_0 \leq \theta \leq \theta_1$ one has

$$e^{-(\theta_1-\theta)\lambda} \big| |V(\theta)| - e^{-\theta\lambda} n \big| \leq \big| |V(\theta_1)| - e^{-\theta_1\lambda} n \big| + \big| |V(\theta_1)| - e^{-(\theta_1-\theta)\lambda} |V(\theta)| \big|,$$

we may set $\Gamma(\theta) = ||V(\theta)| - e^{-\theta\lambda} n|$,

$$\psi = e^{(\theta_1-\theta_0)} \Gamma(\theta_1) \quad \text{and} \quad \psi_\theta = e^{(\theta_1-\theta_0)} \big| |V(\theta_1)| - e^{-(\theta_1-\theta)\lambda} |V(\theta)| \big|.$$

We bound the probabilities $\max_{\theta_0 \leq \theta \leq \theta_1} \Gamma(\theta) \geq R$ and $\min_{\theta_0 \leq \theta \leq \theta_1} \Gamma(\theta) \leq R$ under some conditions.

Lemma 4.5. *Let* $0 \leq \theta_0 < \theta_1$, $R = R_1 + R_2$, $R_1, R_2 > 0$ *and* Φ_θ *be events depending on* $\{X_{\theta'}\}_{\theta' \leq \theta}$. *If*

$$\Gamma(\theta) \leq \psi + \psi_\theta \quad \textit{for all } \theta_0 \leq \theta \leq \theta_1,$$

then

$$\Pr\Big[\max_{\theta_0 \leq \theta \leq \theta_1} \Gamma(\theta) \geq R\Big] \leq \Pr\big[\psi \geq R_1\big] + \Pr\Big[\bigcup_{\theta : \theta_0 \leq \theta \leq \theta_1} \overline{\Phi_\theta}\Big] + \max_{\theta : \theta_0 \leq \theta \leq \theta_1} \max_{\{X_{\theta'}\}_{\theta' \leq \theta}} 1(\Phi_\theta) \Pr\big[\psi_\theta \geq R_2 \mid \{X_{\theta'}\}_{\theta' \leq \theta}\big].$$

Similarly, if

$$\Gamma(\theta) \geq \psi + \psi_\theta \quad \textit{for all } \theta_0 \leq \theta \leq \theta_1,$$

then

$$\Pr\Big[\min_{\theta_0 \leq \theta \leq \theta_1} \Gamma(\theta) \leq -R\Big] \leq \Pr\big[\psi \leq -R_1\big] + \Pr\Big[\bigcup_{\theta:\theta_0 \leq \theta \leq \theta_1} \overline{\Phi}_\theta\Big]$$
$$+ \max_{\theta:\theta_0 \leq \theta \leq \theta_1} \max_{\{X_{\theta'}\}_{\theta' \leq \theta}} 1(\Phi_\theta) \Pr\big[\psi_\theta \leq -R_2 \mid \{X_{\theta'}\}_{\theta' \leq \theta}\big].$$

Proof. See [30]. □

Example 4.3 (continued). As

$$\Pr[\psi \geq R_1] \leq e^{-\Omega(\min\{R_1, \frac{R_1^2}{p(1-p)n}\})}$$

and

$$\Pr[\psi_i \geq R_2 | X_1, \ldots, X_i] = \Pr[\psi_i \geq R_2] \leq e^{-\Omega(\min\{R_2, \frac{R_2^2}{p(1-p)(n-i)}\})},$$

Lemma 4.5 for $R_1 = R_2 = R/2$ and $\Phi_\theta = \emptyset$ gives

$$\Pr[\max_{i:0 \leq i \leq n} |S_i - pi| \geq R] \leq e^{-\Omega(\min\{R, \frac{R^2}{p(1-p)n}\})}.$$

Example 4.4 (continued). Since

$$|V(\theta)| = \sum_{v \in V} 1(v \text{ has no } (1-\theta)\lambda\text{-large clone})$$

is a sum of i.i.d. Bernoulli random variables with mean $e^{-\theta\lambda}$,

$$\Pr\Big[\big||V(\theta)| - e^{-\theta\lambda} n\big| \geq R\Big] \leq e^{-\Omega(\min\{R, \frac{R^2}{\theta n}\})},$$

especially

$$\Pr\big[\psi \geq R/2\big] \leq e^{-\Omega(\min\{R, \frac{R^2}{\theta_1 n}\})}.$$

Once $\{X_{\theta'}\}_{\theta' \leq \theta}$ is given, $V(\theta)$ is determined and

$$V(\theta_1) = \sum_{v \in V(\theta)} 1(v \text{ has no } (1-\theta_1)\lambda\text{-large clone})$$

is a sum of i.i.d. Bernoulli random variables with mean $e^{-(\theta_1 - \theta)\lambda}$. Thus

$$\Pr\big[\psi_\theta \geq R/2 \mid \{X_{\theta'}\}_{\theta' \leq \theta}\big] \leq 2e^{-\Omega(\min\{R, \frac{R^2}{(\theta_1-\theta)|V(\theta)|}\})} \leq 2e^{-\Omega(\min\{R, \frac{R^2}{\theta n}\})},$$

and Lemma 4.5 for $\theta_0 = 0$ and $\Phi_\theta = \emptyset$ yields

$$\Pr\Big[\max_{\theta:0 \leq \theta \leq \theta_1} |V(\theta) - e^{-\theta\lambda} n| \geq R\Big] \leq 2e^{-\Omega(\min\{R, \frac{R^2}{\theta n}\})}.$$

5. Generalized cores and the main lemma

In this section we introduce generalized cores and the main lemma. The main lemma will play a crucial roles in the proofs of the theorems mentioned in the introduction.

We start with some terminology. A *generalized degree* is an ordered pair (d_1, d_2) of non-negative integers. The inequality between two generalized degrees is determined by the inequality between the first coordinates and the reverse inequality between the second coordinates. That is, $(d_1, d_2) \geq (d_1', d_2')$ if and only if $d_1 \geq d_1'$ and $d_2 \leq d_2'$. A *property* for generalized degrees is simply a set of generalized degrees. A property P is *increasing* if generalized degrees larger than an element in P are also in P. When a property P depends only on the first coordinate of generalized degrees, it is simply a property of degrees. For the t-core problem, we will use $P_{t\text{-core}} = \{(d_1, d_2) : d_1 \geq t\}$. To estimate the size of the largest component, we will set $P_{\text{comp}} = \{(d_1, d_2) : d_2 = 0\}$.

Given the Poisson λ-cell on the set V of n vertices and θ with $0 \leq \theta \leq 1$, let $d_v(\theta)$ be the number of v-clones smaller than $\theta\lambda$. Similarly, $\bar{d}_v(\theta)$ is the number of v-clones larger than or equal to $\theta\lambda$. Then $D_v(\theta) := (d_v(\theta), \bar{d}_v(\theta))$ are i.i.d. random variables. In particular, for any property P the events $D_v(\theta) \in P$ are independent and occur with the same probability, say $p(\theta, \lambda; P)$, or simply $p(\theta)$.

For an increasing property P, the P-process is defined as follows. Construct the Poisson λ-cell as described in Section 3, where $\lambda = p\binom{n-1}{k-1}$. The vertex set $V = \{v_0, \ldots, v_{n-1}\}$ will be regarded as an ordered set so that the i^{th} vertex is v_{i-1}. The P-process is a generalization of Example 2.2 for which choice functions choose t-light clones.

The *P*-process. Initially, the cut-off value $\Lambda = \lambda$. Activate all vertices v with $D_v(1) \notin P$. All clones of the activated vertices are activated too. Put those clones in a stack in an arbitrary order. However, this does not mean that the clones are removed from the λ-cell.

(a) If the stack is empty, go to (b). If the stack is nonempty, choose the first clone in the stack and move the cut-off line to the left until the largest $k-1$ unmatched clones, excluding the chosen clone, are found. (So, the cut-off value Λ keeps decreasing.) Then match the $k-1$ clones to the chosen clone. Remove all matched clones from the stack and repeat the process. A vertex that has not been activated is to be activated as soon as $D_v(\Lambda/\lambda) \notin P$. This can be done even before all $k-1$ clones are found. Its unmatched clones are to be activated too and put into the stack immediately. Clones found while moving the cut-off line are also in the stack until they are matched.

(b) Activate the first vertex in V which has not been activated. Its clones are activated too. Put those clones into the stack. Then go to (a).

Clones in the stack are called active. The steps carried out by the instruction described in (b) are called forced steps as it is necessary to artificially activate a vertex.

When the cut-off line is at $\theta\lambda$, all $\theta\lambda$-large clones are matched or will be matched at the end of the step and all vertices v with $D_v(\theta) \notin P$ have been activated. All other vertices can have been activated only by forced steps. Let $V(\theta) = V_P(\theta)$ be the set of vertices v with $D_v(\theta) \in P$, and let $M(\theta) = M_P(\theta)$ be the number of $\theta\lambda$-large clones plus the number of $\theta\lambda$-small clones of vertices v not in $V(\theta)$. That is,

$$M(\theta) = \sum_{v \in V} \bar{d}_v(\theta) + d_v(\theta) 1(v \notin V(\theta)) = \sum_{v \in V} \bar{d}_v(\theta) + d_v(\theta) 1(D_v(\theta) \notin P).$$

Recalling that $N(\theta)$ is the number of matched clones until the cut-off line reaches $\theta\lambda$, the number $A(\theta)$ of active clones (when the cut-off value Λ is) at $\theta\lambda$ is at least as large as $M(\theta) - N(\theta)$. On the other hand, the difference $A(\theta) - (M(\theta) - N(\theta))$ is at most the number $F(\theta)$ of clones activated in forced steps until $\theta\lambda$, i.e.,

$$M(\theta) - N(\theta) \leq A(\theta) \leq M(\theta) - N(\theta) + F(\theta). \tag{5.1}$$

As the cut-off lemma gives a concentration inequality for $N(\theta)$,

$$\Pr\Big[\max_{\theta:\theta_1 \leq \theta \leq 1} |N(\theta) - (1 - \theta^{\frac{k}{k-1}})\lambda n| \geq \Delta \Big] \leq 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{(1-\theta_1)n}\}},$$

a concentration inequality for $M(\theta)$ will be enough to obtain a similar inequality for $B(\theta) := M(\theta) - N(\theta)$. More precisely, we will show that under appropriate hypotheses

$$\Pr\Big[\max_{\theta:\theta_1 \leq \theta \leq 1} |M(\theta) - (\lambda - q(\theta))n| \leq \Delta \Big] \leq 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{(1-\theta_1)n}\})},$$

where

$$q(\theta) = q(\theta, \lambda; P) = E\big[d_v(\theta) 1(D_v(\theta) \in P)\big].$$

As the $d_v(\theta)$'s and $D_v(\theta)$'s are identically distributed, $q(\theta)$ does not depend on v. Also, recall that $p(\theta) = \Pr[D_v(\theta) \in P]$.

As we will see later, $B(\theta)$ is very close to $A(\theta)$. Hence a concentration inequality for $B(\theta)$ plays a very important roles in all of our proofs.

Lemma 5.1 (Main lemma). *In the P-process, if $\theta_1 < 1$ uniformly bounded from below by 0, $1 - p(\theta_1) = O(1 - \theta_1)$ and $p(\theta_1) = \Omega(1)$, then for all Δ in the range $0 < \Delta \leq n$ we have*

$$\Pr\Big[\max_{\theta:\theta_1 \leq \theta \leq 1} \big| |V(\theta)| - p(\theta)n \big| \leq \Delta \Big] \leq 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{(1-\theta_1)n}\})}$$

and

$$\Pr\Big[\max_{\theta:\theta_1 \leq \theta \leq 1} \big| B(\theta) - (\lambda\theta^{\frac{k}{k-1}} - q(\theta))n \big| \leq \Delta \Big] \leq 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{(1-\theta_1)n}\})}.$$

Proof. See [30]. □

6. Cores of random hypergraphs

In this section we prove Theorem 1.7. Let $\lambda > 0$ and $H(\lambda) = H_{\rm PC}(n, p)$, where $\lambda = p\binom{n-1}{r-1}$. Let the property $P = \{(d_1, d_2) : d_1 \ge t\}$. Then

$$p(\theta) = Q(\theta\lambda, t) \quad \text{and} \quad q(\theta) = \theta\lambda Q(\theta\lambda, t-1).$$

The main lemma gives

Corollary 6.1. *For $\theta_1 \le 1$ uniformly bounded from below by 0 and Δ in the range $0 < \Delta \le n$,*

$$\Pr\Big[\max_{\theta:\theta_1\le\theta\le 1}\big||V(\theta)| - Q(\theta\lambda, t)n\big| \ge \Delta\Big] \le 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{n}\})}$$

and

$$\Pr\Big[\max_{\theta:\theta_1\le\theta\le 1}\big|B(\theta) - (\theta^{\frac{1}{k-1}} - Q(\theta\lambda, t-1))\theta\lambda n\big| \ge \Delta\Big] \le 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{n}\})}.$$

Subcritical Region: For $\lambda = \lambda_{\rm crt} - \sigma$, $\sigma > 0$ and $\theta_1 = \delta/\lambda_{\rm crt}$ with $\delta = 0.1$ it is easy to see that there is a constant $c > 0$ such that

$$(\theta^{\frac{1}{k-1}} - Q(\theta\lambda, t-1))\theta\lambda n \ge c\sigma n \quad \text{for all } \theta \text{ with } \theta_1 \le \theta \le 1.$$

Let τ be the first time the number $A(\theta)$ of active clones at $\theta\lambda$ becomes 0. Then the second part of Corollary 6.1 gives

$$\begin{aligned}\Pr[\tau \ge \theta_1] &\le \Pr[B(\theta) = 0 \text{ for some } \theta \text{ with } \theta_1 \le \theta \le 1]\\ &\le \Pr\Big[\max_{\theta:\theta_1\le\theta\le 1}\big|B(\theta) - (\theta^{\frac{1}{k-1}} - Q(\theta\lambda, t-1))\theta\lambda n\big| \ge c\sigma n\Big]\\ &\le 2e^{-\Omega(\sigma^2 n)}.\end{aligned}$$

As $\theta_1\lambda \le \theta_1\lambda_{\rm crt} = \delta$, and hence $Q(\theta_1\lambda, t) \le \delta/2$ for $t \ge 2$, the first part of Corollary 6.1 yields

$$\Pr[|V_t(H_{\rm PC}(n, p\,; k))| \ge \delta n] \le \Pr[\tau \ge \theta_1] + \Pr[|V(\theta_1)| \ge \delta n] \le 2e^{-\Omega(\sigma^2 n)}.$$

Therefore Theorem 1.1 implies that

$$\Pr[|V_t(H(n, p\,; k))| \ge \delta n] \le 2e^{-\Omega(\sigma^2 n)}.$$

To complete the proof, we observe that the t-core of size i has at least ti/k edges. Let Z_i be the number of subgraphs on i vertices with at least ti/k edges, $i = i_0, \ldots, \delta n$, where $i_0 = i_0(k, t)$ is the least i such that $\binom{i}{k} \ge ti/k$. Then

$$E[Z_i] \le \binom{n}{i}\binom{\binom{i}{k}}{ti/k}p^{ti/k} \le \frac{n^i}{i!}\frac{i^{ti}}{(ti/k)!}p^{ti/k} =: L_i, \tag{6.1}$$

where ti/k actually means $\lceil ti/k \rceil$. Hence

$$\frac{L_{i+k}}{L_i} = O\Big(\frac{n^k}{i^k}\frac{i^{kt}}{i^t}n^{-(k-1)t}\Big) = O\Big(\Big(\frac{i}{n}\Big)^{(k-1)t-k}\Big) = O(\delta^{(k-1)(t-1)-1}).$$

That is, L_{i+k}/L_i exponentially decreases. For $i = i_0, \ldots, i_0 + k - 1$,

$$L_i = O(n^i n^{-i(k-1)t/k}) = O(n^{-i(t-1-t/k)})$$

implies that

$$\Pr[V_t(H(n, p\,; k)) \neq \emptyset] \leq 2e^{-\Omega(\sigma^2 n)} + O(n^{-i_0(t-1-t/k)}),$$

as desired. □

Supercritical region: We will prove the following theorem.

Theorem 6.2. *Suppose that $p\binom{n-1}{k-1} \geq \lambda_{\rm crt} + \sigma$ and $0 < \delta \leq 1$. Then, with probability $1 - 2e^{-\Omega(\min\{\delta^2\sigma n, \sigma^2 n\})}$, $V_t = V_t(H_{\rm PC}(n, p\,; k))$ satisfies*

$$Q(\theta_\lambda\lambda, t)n - \delta n \leq |V_t| \leq Q(\theta_\lambda\lambda, t)n + \delta n, \tag{6.2}$$

and the degrees of vertices of the t-core are i.i.d. t-truncated Poisson random variables with parameter $\Lambda_t := \theta_\lambda\lambda + \beta$ for some β with $|\beta| \leq \delta$. Moreover, the distribution of the t-core is the same as that of the t-truncated Poisson cloning model with parameters $|V_t|$ and Λ_t.

Proof. Let $\lambda = \lambda_{\rm crt} + \sigma$, $\sigma > 0$ and θ_λ be the largest solution for the equation

$$\theta^{\frac{1}{k-1}} - Q(\theta\lambda, t-1) = 0.$$

Then it is not hard to check that there are constants $c_1, c_2 > 0$ such that for θ in the range $\theta_\lambda \leq \theta \leq 1$,

$$\theta^{\frac{1}{k-1}} - Q(\theta\lambda, t-1) \geq c_1\sigma^{1/2}(\theta - \theta_\lambda),$$

and for θ in the range $\theta_\lambda - c_2\sigma^{1/2} \leq \theta \leq \theta_\lambda$,

$$\theta^{\frac{1}{k-1}} - Q(\theta\lambda, t-1) \leq -c_1\sigma^{1/2}(\theta_\lambda - \theta).$$

Let τ be the largest θ with $A(\theta) = 0$. Then $V(\tau)$ is the t-core of $H_{\rm PC}(n, p; k)$. For $\theta_1 = \theta_\lambda + \delta$ and $\theta_2 = \theta_\lambda - \min\{\delta, c_2\sigma^{1/2}\}$ with $0 < \delta \leq 1$, Corollary 6.1 gives

$$\begin{aligned}\Pr[\tau \geq \theta_1] &\leq \Pr[B(\theta) = 0 \text{ for some } \theta \text{ with } \theta_1 \leq \theta \leq 1] \\ &\leq \Pr\Big[\max_{\theta:\theta_1\leq\theta\leq 1} |B(\theta) - (\theta^{\frac{1}{k-1}} - Q(\theta\lambda, t-1))\theta\lambda n| \geq c_1\sigma^{1/2}\delta n\Big] \\ &\leq 2e^{-\Omega(\delta^2\sigma n)}\end{aligned}$$

and

$$\begin{aligned}\Pr[\tau < \theta_2] &\le \Pr[B(\theta_2) > 0]\\ &\le \Pr\left[|B(\theta_2) - (\theta_2^{\frac{1}{k-1}} - Q(\theta_2\lambda, t-1))\theta_2\lambda n| \ge c_1\sigma^{1/2}\min\{\delta, c_2\sigma^{1/2}\}n\right]\\ &\le 2e^{-\Omega(\min\{\delta^2\sigma n, \sigma^2 n\})}.\end{aligned}$$

Since $\frac{d}{d\theta}Q(\theta\lambda, t) = \lambda P(\theta\lambda, t-1) \le \lambda$, we have

$$Q(\theta_1\lambda, t) \le Q(\theta_\lambda\lambda, t) + \lambda\delta, \quad \text{and} \quad Q(\theta_2\lambda, t) \ge Q(\theta_\lambda\lambda, t) - \lambda\delta,$$

and Corollary 6.1 implies that

$$\Pr[V(\theta_1) - Q(\theta_\lambda\lambda, t)n \ge 2\lambda\delta n] \le 2e^{-\Omega(\delta^2 n)}$$

and

$$\Pr[V(\theta_2) - Q(\theta_\lambda\lambda, t)n \le -2\lambda\delta n] \le 2e^{-\Omega(\delta^2 n)}.$$

Therefore

$$\Pr[|\tau - \theta_\lambda| > \delta] \le \Pr[\tau \ge \theta_1] + \Pr[\tau \le \theta_2] \le 2e^{-\Omega(\min\{\delta^2\sigma n, \sigma^2 n\})}$$

and, replacing δ by $\frac{\delta}{2\lambda}$,

$$\begin{aligned}\Pr[|V(\tau) - Q(\theta_\lambda\lambda, t)n| \ge \delta n] &\le \Pr[\tau \ge \theta_1] + \Pr[\tau \le \theta_2] + 2e^{-\Omega(\delta^2 n)}\\ &\le 2e^{-\Omega(\min\{\delta^2\sigma n, \sigma^2 n\})}.\end{aligned}$$

Clearly, once $V(\tau)$ and $\Lambda_t := \tau\lambda$ are given, the residual degrees $d_v(\tau)$, $v \in V(\tau)$, are i.i.d. t-truncated Poisson random variables with parameter Λ_t. □

Once V_t and Λ_t are given, $|V_t(i)|$, $i \ge t$, is the sum of i.i.d. Bernoulli random variables with mean $p_i(\Lambda_t) := \frac{P(\Lambda_t, i)}{Q(\Lambda_t, t)}$. Similarly, the size of $W_t(i) = \bigcup_{j \ge i} V_t(j)$ is the sum of i.i.d. Bernoulli random variables with mean $q_i(\Lambda_t) := \frac{Q(\Lambda_t, i)}{Q(\Lambda_t, t)}$. Applying the generalized Chernoff bound (Lemma 4.2), we have

$$\Pr\left[\big||V_t(i)| - p_i(\Lambda_t)|V_t|\big| \ge \delta|V_t| \,|\, V_t, \Lambda_t\right] \le 2e^{-\Omega(\delta^2|V_t|)}$$

and

$$\Pr\left[\big||W_t(i)| - q_i(\Lambda_t)|V_t|\big| \ge \delta|V_t| \,|\, V_t, \Lambda_t\right] \le 2e^{-\Omega(\delta^2|V_t|)}.$$

Combining this with Lemma 6.2 and using

$$|P(\rho, i) - P(\rho', i)| \le |\rho - \rho'|, \quad \text{and} \quad |Q(\rho, i) - Q(\rho', i)| \le |\rho - \rho'|,$$

we obtain, for any i,

$$\Pr\left[\big||V_t(i)| - P(\theta_\lambda\lambda, i)n\big| \ge \delta n\right] \le 2e^{-\Omega(\min\{\delta^2\sigma n, \sigma^2 n\})},$$

and

$$\Pr\Big[\big||W_t(i)| - Q(\theta_\lambda\lambda, i)n\big| \geq \delta n\Big] \leq 2e^{-\Omega(\min\{\delta^2\sigma n, \sigma^2 n\})}.$$

In particular, as $\theta_\lambda = \theta_{\mathrm{crt}} + \Theta(\sigma^{1/2})$ for uniformly bounded σ it follows that for $\lambda = \lambda_{\mathrm{crt}} + \sigma$,

$$|V_t(i)| = (1 + O(\sigma^{1/2}))^i P(\theta_{\mathrm{crt}}\lambda_{\mathrm{crt}}, i)n + O\big((n/\sigma)^{1/2}\log n\big),$$

with probability $1 - 2e^{-\Omega(\min\{\log^2 n, \sigma^2 n\})}$.

7. The emergence of the giant component

In this section we just give ideas for the proof of Theorem 1.4. Let the property P be $\{(d_1, d_2) : d_2 = 0\}$. Then $p(\theta) = e^{-(1-\theta)\lambda}$ and $q(\theta) = \theta\lambda e^{-(1-\theta)\lambda}$, and the main lemma gives

Corollary 7.1. *For $\theta_1 \geq 1$ uniformly bounded from above by 1 and Δ in the range $0 < \Delta \leq n$,*

$$\Pr\Big[\max_{\theta:0\leq\theta\leq\theta_1}\big||V(\theta)| - e^{-(1-\theta)\lambda}n\big| \geq \Delta\Big] \leq 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{(1-\theta_1)n}\})}$$

and

$$\Pr\Big[\max_{\theta:0\leq\theta\leq\theta_1}\big|B(\theta) - (\theta - e^{-(1-\theta)\lambda})\theta\lambda n\big| \geq \Delta\Big] \leq 2e^{-\Omega(\min\{\Delta, \frac{\Delta^2}{(1-\theta_1)n}\})}.$$

To estimate $A(\theta)$ it is now enough for us to estimate $F(\theta)$ by (5.1). Once good estimations for $F(\theta)$ are established, we may take similar (but slightly more complicated) approaches used in the previous section.

It is convenient to consider an (imaginary) secondary stack with parameter ρ, or simply ρ-secondary stack. Initially, the secondary stack with parameter ρ consists of the first ρn vertices $v_0, \ldots, v_{\rho n-1}$ of V. The set of those ρn vertices is denoted by V_ρ. Whenever the primary stack is empty, the first vertex in the secondary stack that has not been activated must be activated. Its clones are activated too and put into the primary stack. The activated vertex as well as vertices activated by other means are no longer in the secondary stack. If the secondary stack is empty, go back to the regular procedure. This does not change the P-process at all, but will be used just for the analysis. Let τ_ρ be the largest τ such that, at $\tau\lambda$, the primary stack becomes empty after the secondary stack is empty. Thus, once the cut-off line reaches $\tau_\rho\lambda$, no active clones are provided from the secondary stack. Denote by $C(\rho)$ the union of the components containing any vertex in V_ρ.

The following lemma is useful to predict how large τ_ρ is.

Lemma 7.2. *Suppose* $0 < \delta,\ \rho < 1$ *and* $\theta_1, \theta_2 \leq 1$ *are uniformly bounded from below by* 0*. Then*

$$\Pr[\tau_\rho \geq \theta_1] \leq \Pr\Big[\min_{\theta:\theta_1\leq\theta\leq 1} B(\theta) \leq -(1-\delta)\theta_1\lambda e^{-(1-\theta_1)\lambda}\rho n\Big] + 2e^{-\Omega(\delta^2\rho n)},$$

and conversely,

$$\Pr[\tau_\rho \leq \theta_2] \leq \Pr[B(\theta_2) \geq -(1+\delta)\theta_2\lambda e^{-(1-\theta_2)\lambda}\rho n] + 2e^{-\Omega(\delta^2\rho n)}.$$

Proof. See [30]. □

Once the value of τ_ρ is known quite precisely, a good estimation of $F(\theta)$ is possible. Using similar (but slightly more complicated) arguments used in the previous section, estimation of $A(\theta)$ is also possible. Due to space limitation, the proof of Theorem 1.4 is omitted.

8. Closing remarks

The Poisson λ-cell is introduced to analyze those properties of $G_{\mathrm{PC}}(n, p)$, for which the degrees are i.i.d. Poisson random variables with mean $\lambda = p(n-1)$. Then various nice properties of Poisson random variables are used to analyze sizes of the largest component and the t-core of $G_{\mathrm{PC}}(n, p)$. We believe that the approaches presented in this paper are useful to analyze problems with similar flavors, especially problems related to branching processes. For example, we can easily modify the proofs of Theorem 1.7 to analyze the pure literal rule for the random k-SAT problems, $k \geq 3$. Another example may be the Karp–Sipser Algorithm to find a large matching of the random graph. (See [29], [3].) In a subsequent paper, we will analyze the structure of the 2-core of $G(n, p)$ and the largest strong component of the random directed graph as well as the pure literal rule for the random 2-SAT problem.

For the random (hyper)graph with a given sequence (d_i), we may also introduce the (d_i)-cell, in which the vertex v_i has d_i clones and each clone is assigned a uniform random real number between 0 and the average degree $\frac{1}{n}\sum_{i=0}^{n-1} d_i$. Though it is not possible to use all of the nice properties of Poisson random variables any more, we believe that the (d_i)-cell equipped the cut-off line algorithm can be used to prove stronger results for the t-core problems considered in various papers including [13], [21], [22], [26], [37].

Recall that the degrees in $G(n, p)$ has the binomial distribution with parameters $n-1$ and p. By introducing the Poisson cloning model, we somehow first take the limit of the binomial distribution, which is the Poisson distribution. In general, many limiting distributions like Poisson and Gaussian ones have nice properties. In our opinion this is because various small differences are eliminated by taking the limits, and limiting distributions have some symmetric and/or invariant properties. Thus one may wonder whether there is an infinite graph that shares most properties of the random graphs $G(n, p)$ with large enough n. So, in a sense, the infinite graph, if it exists, can be regarded as the limit of $G(n, p)$. An infinite graph which Aldous [1] considered to solve the linear assignment problem may or may not be a (primitive) version of such an infinity graph. Though it may be impossible to construct such a graph, the approaches taken in this paper might be useful to find one, if any.

Acknowledgement. The author thanks C. Borgs, J. Chayes, B. Bollobás and Y. Peres for helpful discussions.

References

[1] Aldous, D., The zeta(2) limit in the random assignment problem. *Random Structures Algorithms* **18** (2001), 381–418.

[2] Alon, N., Spencer, J., *The Probabilistic Method.* 2nd ed., Wiley-Interscience, New York, NY, 2000.

[3] Aronson, J., Frieze, A., Pittel, B., Maximum matchings in sparse random graphs: Karp-Sipser revisited. *Random Structures Algorithms* **12** (1998), 111–178.

[4] Athreya, K. B., Ney, P. E., *Branching processes*. Grundlehren Math. Wiss. 196, Springer-Verlag, Berlin 1972.

[5] Bollobás, B., The evolution of random graphs. *Trans. Amer. Mat. Soc.* **286** (1984), 257–274.

[6] Bollobás, B., The evolution of sparse graphs. In *Graph Theory and Combinatorics* (ed. by B. Bollobás), Academic Press, London 1984, 35–57.

[7] Bollobás, B., *Random graphs*. Academic Press, London 1985.

[8] Bollobás, B., The chromatic number of random graphs. *Combinatorica* **8** (1988), 49–56.

[9] Bollobás, B., Erdős, P., Cliques in random graphs. *Math. Proc. Cambridge Philos. Soc.* **80** (1976), 419–427.

[10] Bollobás, B., Frieze A. M., On matchings and Hamiltonian cycles. In *Random Graphs '83* (ed. by M. Karoński and A. Ruciński), North-Holland, Amsterdam, New York 1985, 23–46.

[11] Bollobás, B., Thomason A., Random graphs of small order. In *Random Graphs '83* (ed. by M. Karoński and A. Ruciński), North-Holland, Amsterdam, New York 1985, 47–97.

[12] Chvátal, V., Almost all graphs with 1.44 edges are 3-colorable. *Random Structures Algorithms* **2** (1991), 11–28.

[13] Cooper, C., The cores of random hypergraphs with a given degree sequence. *Random Structures Algorithms* **25** (2004), 353–375.

[14] Erdős, P., Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.* **53** (1947), 292–294.

[15] Erdős, P., Rényi, A., On random graphs I. *Publ. Math. Debrecen* **6** (1959), 290–297.

[16] Erdős, P., Rényi, A., On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960), 17–61.

[17] Erdős, P., Rényi, A., On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hung.* **12** (1961), 261–267.

[18] Erdős, P., Rényi, A., On random matrices. *Publ. Math. Inst. Hung. Acad. Sci.* **8** (1964), 455–461.

[19] Erdős, P., Rényi, A., On the existence of a factor of degree one of a connected random graph. *Acta Math. Acad. Sci. Hung.* **17** (1966), 359–368.

[20] Erdős, P., Rényi, A., On random matrices II. *Stud. Sci. Math. Hung.* **3** (1968), 459–464.

[21] Fernholz D., Ramachandran V., The giant k-core of a random graph with a specified degree sequence. Manuscript, 2003.

[22] Fernholz D., Ramachandran V., Cores and connectivity in sparse random graphs. Tech. Report TR-04-13, University of Texas at Austin, Department of Computer Science, Austin, 2004.

[23] Flajolet, D., Knuth D. E., Pittel, B., The first cycle in an evolving graph. *Discrete Math.* **75** (1989), 167–215.

[24] Janson, S., Poisson convergence and Poisson processes with applications to random graphs. *Stochastic Process. Appl.* **26** (1988), 1–30.

[25] Janson, S., Knuth D. E., Łuczak, T., Pittel, B., The birth of the giant component. *Random Structures Algorithms* **3** (1993), 233–358.

[26] Janson, S., Łuczak, M. J., A simple solution to the k-core problem. Manuscript, 2005.

[27] Janson, S., Łuczak, T., Ruciński, A., *Random graphs*. Wiley-Interscience, New York, NY, 2000.

[28] Karp, R. M., The transitive closure of a random digraph. *Random Structures Algorithms* **1** (1990), 73–93.

[29] Karp, R. M., Sipser, M., Maximum matchings in sparse random graphs. In *Proceedings of the 22nd IEEE Symposium on Foundations of Computer Science*, IEEE Comput. Soc. Press, Los Alamitos, CA, 1981, 364–375.

[30] Kim., J.-H., Poisson cloning model for random graph. Manuscript, 2005.

[31] Komlś, J., Szemerédi E., Limit distributions for the existence of Hamilton cycles in a random graph. *Discrete Math.* **43** (1983), 55–63.

[32] Łuczak, T., Component behavior near the critical point of the random graph process. *Random Structures Algorithms* **1** (1990), 287–310.

[33] Łuczak, T., The chromatic number of random graphs. *Combinatorica* **11** (1991), 45–54.

[34] Łuczak, T., Size and connectivity of the k-core of a random graph. *Discrete Math.* **91** (1991), 61–68.

[35] Łuczak, T., Pittel, B., Wierman, J., The birth of the giant component. *Trans. Amer. Mat. Soc.* **341** (1994), 721–748.

[36] Matula, D., The largest clique size in a random graph. *Tech. Rep., Dept. Comp. Sci.*, Southern Methodist University, Dallas, 1976.

[37] Molloy, M., Cores in random random hypergraphs and Boolean formulas. *Random Structures Algorithms* **27** (2005), 124–135.

[38] Pittel, B., Spencer J., Wormald, N., Sudden emergence of a giant k-core in a random graph. *J. Combin. Theory Ser. B* **67** (1996), 111–151.

[39] Shamir, E., Spencer, J., Sharp concentration of the chromatic number on random graph $G_{n,p}$. *Combinatorica* **7** (1987), 124–129.

Microsoft Research, One Microsoft Way, Redmond, WA 98052, U.S.A
E-mail: jehkim@microsoft.com