The Poisson Cloning Model for Random Graphs

Jeong Han Kim

Microsoft Research

jehkim@microsoft.com

Motivation

Models of Random graphs

Poisson Cloning

***k*-core problem**

Random k-SAT

• Motivation

In $G(n,p)$ on the vertex set

$$V = \{1, ..., n\},$$

all degrees $d(v)$ are identically distributed and

$$d(v) \in \mathrm{Bin}(n-1, p)$$

(for any $v \in V$). If

$$p = \frac{\lambda}{n-1}$$

for a constant $\lambda > 0$, then

$$d(v) \text{ is almost } \mathrm{Poi}(\lambda).$$

Any model for random graphs so that all degrees are i.i.d $\mathrm{Poi}(\lambda)$ **??**

Models of random graphs

- Binomial Model $G(n, p)$ (Usual Random Graph)
- Uniform Model $G_{UN}(n, p)$

  (cf. Random graph $G(n, m)$ without repetition)
- Independent Edge Model $G_{IE}(n, p)$

  (cf. Random Graph $G(n, m)$ with repetition)
- Independent Vertex Model $G_{IV}(n, p)$
- Poisson Cloning Model $G_{PC}(n, p)$

  (independent degree model)

- Binomial Model $G(n, p)$

each of $\binom{n}{2}$ edges is independently
in $G(n, p)$ with probability p

$p = 1$: complete graph

$p = 0$: empty graph

Let M_p be the number of edges. Then

$$M_p \in \mathrm{Bin}\Big(\binom{n}{2}, p\Big)$$

For fixed $G = (V, E)$ with m edges,

$$\Pr[G(n, p) = G] = p^m (1 - p)^{\binom{n}{2} - m}.$$

- **Uniform model** $G_{UN}(n, p)$

$G(n, m)$: a graph chosen uniformly at random

among all graphs with m edges.

Or equivalently,

Choose a collection of m edges uniformly at random among all possible $\binom{\binom{n}{2}}{m}$ collections.

For G with m edges

$$\Pr[G(n, m) = G] = \binom{\binom{n}{2}}{m}^{-1}.$$

Take a binomial random variable

$$M_p \in B\left(\binom{n}{2}, p\right)$$

and define

$$G_{UN}(n, p) = G(n, M_p).$$

Is $G_{UN}(n,p) = G(n,p)$*??*

Recall for a graph G with m edges,

$$\Pr[G(n,p) = G] = p^m (1-p)^{\binom{n}{2}-m}.$$

On the other hand,

$$\begin{aligned}
\Pr[G_{UN}(n,p) = G] &= \Pr[M_p = m] \Pr[G_{UN}(n,m) = G] \\
&= \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m} \times \binom{\binom{n}{2}}{m}^{-1} \\
&= p^m (1-p)^{\binom{n}{2}-m}.
\end{aligned}$$

Note that

when we choose M_p random edges for $G_{UN}(n, p)$,

$$2M_p \approx 2p\binom{n}{2} \text{ vertices are chosen.}$$

with multiplicity,

• **Poisson cloning model** $G_{PC}(n, p)$

Take a random variables

$$N_P \in \mathrm{Poi}\Big(2p\binom{n}{2}\Big),$$

and N_p i.i.d. (uniform) random vertices.

The random vertices chosen are called **clones**.

If the total number N_p of clones is even,

generate a random perfect matching
on the set of all clones and then
contract clones of the same vertex.

In general, $G_{PC}(n, p)$ is a pseudograph. ...

Equivalently, for each $v \in V$, independently take

$$d(v) \in \mathrm{Poi}(p(n-1))$$

clones of v. It is well-known that

$$N_p := \sum_{v \in V} d(v) \ \in \ \mathrm{Poi}(pn(n-1)).$$

Then generate a (uniform) random perfect matching on N_p clones.

(Partially motivated by the pairing model of random regular graphs due to Bollobás ('79) similar models used by Bender and Canfield ('78), Békéssy, Békéssy and Komlós ('74), Wormald ('78))

$G(n,p)$ **vs.** $G_{PC}(n,p)$

Lemma. If $pn^2 = O(n)$, then there are positive constants c_1 and c_2 so that for any collection $\mathcal{G}$ of SIMPLE graphs

$$c_1 \Pr[G_{PC}(n,p) \in \mathcal{G}] - e^{-\quad (pn^2)} \cdot \quad \Pr[G(n,p) \in \mathcal{G}]$$

and

$$\Pr[G(n,p) \in \mathcal{G}] \cdot \quad c_2 \Big(\Pr[G_{PC}(n,p) \in \mathcal{G}] \Big)^{1/2} + e^{-\quad (pn^2)}.$$

- **Generating Random Perfect Matching**

For each clone w, assign a uniform random (real) number from 0 to $\lambda := p(n-1)$.

We say that a clone is larger than another clone if the assigned numbers are so.

Take two largest clones and match them
and repeat it for the remaining, or unmatched, clones.

Or,

choose the first unmatched clone
according to a certain selection rule (SR)
without looking assigned numbers,
then match it to the largest unmatched clone.

(Any selection rule without changing the independency and the uniform distribution of the assigned numbers is fine.)

If N_p is even, the matching is
the uniform perfect matching.

If N_p is odd, the matching obtained this way leaves only one clone uncovered. The distribution of the matching varies depending upon section rules, which also instruct what to do for the clone.

E.g. Add a loop to the corresponding vertex of the uncovered clone.

- The k-core Problem

A k-core of a graph is a largest subgraph
with minimum degree at least k.

(due to Bollobás)

Let

$$\lambda_k = \min_{\rho > 0} \frac{\rho}{P(\rho, k-1)},$$

where

$$P(\rho, k-1) := \Pr\Big(\mathrm{Poi}(\rho) \geq k-1\Big) = e^{-\rho} \sum_{l=k-1}^{\infty} \frac{\rho^l}{l!}.$$

(Pittel, Spencer & Wormald '96)

For $k \geq 3$,

$$\Pr\Big[G(n, \lambda/(n-1)) \text{ has a } k\text{-core}\Big] \to \begin{cases} 0 & \text{if } \lambda < \lambda_{k} - n^{-\delta} \\ 1 & \text{if } \lambda > \lambda_{k} + n^{-\delta}, \end{cases}$$

for any $\delta \in (0, 1/2)$, and

$$\Pr\Big[\text{either } \exists \text{ no } k\text{-core or } \exists \ k\text{-core of size} \geq (1 - n^{-\delta})\lambda_{k}^{*} n\Big]$$

$$\to 1.$$

Recall,

$$\lambda_k = \min_{\rho > 0} \frac{\rho}{P(\rho, k-1)}.$$

(improving Łuczak's result).

For $k \geq 3$,

$$\Pr\left[G_{PC}(n, \lambda/(n-1)) \text{ has a } k\text{-core }\right] \to$$

$$\begin{cases} 0 & \text{if } \lambda < \lambda_k - n^{-1/2} \log n \\ 1 & \text{if } \lambda > \lambda_k + \omega(n) n^{-1/2}, \end{cases}$$

for any $\omega(n) \to \infty$.

$$\Pr\Big[\text{either } \exists \text{ no } k\text{-core or}$$

$$\exists\ k\text{-core of size} \geq (1 - n^{-1/2} \log n)\lambda_k^* n\Big] = o(1).$$

Recall,

$$\lambda_k = \min_{\rho > 0} \frac{\rho}{P(\rho, k-1)}.$$

In $G_{PC}(n, p)$ and $\lambda = p(n-1)$,

At step 0,

A vertex v is light if $d(v) < k$,
or the number of v-clones is less than k.

It is heavy, otherwise.

Take a light clone w and
then choose the largest clone except w.

In general, at step i,

A vertex v is light if
the number of unmatched v-clones is less than k.

It is heavy, otherwise.

Take a unmatched light clone w and
then choose the largest unmatched clone except w.

.

- **Parameters**

$$N_i = \text{the number of unmatched clones},$$

$$N_0 \approx \lambda n.$$

$$\lambda_i = \text{the length of the interval},$$

$$\lambda_0 = \lambda.$$

$$H_i = \text{the number of unmatched heavy clones},$$

$$H_0 = \sum_{v \in V} X_v 1(X_v \geq k),$$

$$E[H_0] = \lambda P(\lambda, k-1),$$

where X_v's are i.i.d. Poi(λ).

Generally, if λ_i is given,

$$H_i = \sum_{v \in V} X_i 1(X_i \geq k)$$

and

$$E[H_i] = \lambda_i P(\lambda_i, k-1),$$

where X_v's are i.i.d. Poi(λ_i).

Note that

$$N_i - H_i = \text{the total number of light clones.}$$

Thus

$$\exists \text{ no } k\text{-core} \quad \text{iff} \quad N_i - H_i > 0 \text{ for all } i.$$

Trivially,

$$N_i = N_0 - 2i.$$

In expectation,

$$\lambda_1 = \Big(1 - \frac{1}{N_0 - 1}\Big)\lambda$$

since we took the largest number among $N_0 - 1$ i.i.d uniform random numbers from 0 to λ. Similarly, in expectation,

$$\lambda_i = \lambda \prod_{j=0}^{i-1} \Big(1 - \frac{1}{N_i - 1}\Big).$$

Actually,

$$\lambda_i = \lambda \prod_{j=0}^{i-1} \Big(1 - T_i\big),$$

where T_i are mutually independent with

$$\Pr[T_i \geq t] \approx e^{-(N_{i-1}-1)t}.$$

Thus, for $\theta_i^2 := N_i/N_0$

$$\lambda_i \approx \lambda \exp\Big(-\sum_{j=0}^{i-1} \frac{1}{N_p - 2j - 1}\Big) \approx \theta_i \lambda.$$

In terms of θ_i,

$$N_i = \theta_i^2 N \approx \theta_i^2 \lambda n,$$

and

$$H_i \approx \lambda_i P(\lambda_i, k-1)n \approx \theta_i \lambda P(\theta_i \lambda, k-1).$$

Since $\exists$ no k-core $\quad$ iff $\quad N_i - H_i > 0$ for all i,

$$\exists \text{ no } k\text{-core} \quad \text{iff} \quad \theta - P(\theta\lambda, k-1) > 0 \quad \forall \theta \in (0,1),$$

or

$$\exists\ k\text{-core} \quad \text{iff} \quad \theta - P(\theta\lambda, k-1) = 0 \quad \text{for some } \theta \in (0,1).$$

EASY: If

$$\lambda < \lambda_k = \min_{\rho > 0} \frac{\rho}{P(\rho, k-1)},$$

then

$$\forall\ \theta \in (0,1), \quad \theta - P(\theta\lambda, k-1) > 0,$$

and if

$$\lambda > \lambda_k,$$

then

$$\exists \theta \in (0,1), \quad \theta - P(\theta\lambda, k-1) = 0.$$

.

Hypergraph

In terms of θ_i,

$$N_i = \theta_i^h N \approx \theta_i^h \lambda n,$$

and

$$H_i \approx \lambda_i P(\lambda_i, k-1)n \approx \theta_i^{h-1}\lambda P(\theta_i^{h-1}\lambda, k-1).$$

For

$$f_h(\theta) = \theta - P(\theta^{h-1}\lambda, k-1)$$

$$\exists \text{ no } k\text{-core} \quad \text{iff} \quad f_h(\theta) > 0 \qquad \forall \theta \in (0,1),$$

or

$$\exists\ k\text{-core} \quad \text{iff} \quad f_h(\theta) = 0 \quad \text{for some } \theta \in (0,1).$$

For h-uniform hypergraph,

$$\Pr\left[H_h(n, \lambda/(n-1)) \text{ has a } k\text{-core}\right] \to$$

$$\begin{cases} 0 & \text{if } \lambda < \lambda_k^{(h)} - n^{-1/2}\log n \\ 1 & \text{if } \lambda > \lambda_k + \omega(n)n^{-1/2}, \end{cases}$$

where

$$\lambda_k^{(h)} = \min_{\rho>0} \frac{\rho}{(P(\rho, k-1))^{h-1}}.$$

For 2-core for 3-uniform hypergrah,

$$\lambda_k^{(h)} \approx 2.45542\cdots, \quad \text{or} \quad m_k^{(h)} = 1.63694\cdots.$$

(Note that $2 \times 0.818 = 1.636$.)

Random k-SAT

Random Digraph

Random 2-SAT

Boolean Variables: $x_1, ..., x_n \in \{0, 1\}$

Negation of x: $\bar{x} = 1 - x$

$2n$ literals: $x_1, \bar{x}_1, ..., x_n, \bar{x}_n$

x and y are *strictly distinct (s.d.)*
if $x \neq y$ and $x \neq \bar{y}$

k-clause:

$$C = v_1 \vee \cdots \vee v_k$$

where $v_1, ..., v_k$ are s.d. literals

How many k-clauses??

Take k Boolean variables out of n.

Then $\exists$ two choices (negation or not) for each variable.

$$2^k \binom{n}{k}$$

k-SAT Formula:

$$F = F(x_1, ..., x_n) = C_1 \wedge \cdots \wedge C_m$$

where $C_1, ..., C_m$ are k-clauses.

F is *satisfiable* if

$$F(x_1, ..., x_n) = 1$$

for some $x_1, ..., x_n \in \{0, 1\}$

k-SAT problem: NP-Complete if $k \geq 3$

(P if $k = 2$)

- Random 2-SAT $F(n,p)$:

$$\text{Each 2-clause appears in } F$$
$$\text{with probability } p$$

Expected # of clauses

$$m = 2^2 p \binom{n}{2}$$

(Goerdt '92, Chvátal & Reed '92, F. de la Vega '92) For $k = 2$,

$$\Pr[F_2 \text{ is SAT }] \to \begin{cases} 1 & \text{if } \; m/n \to c < 1 \\ 0 & \text{if } \; m/n \to c > 1 \end{cases}$$

- **Poisson Cloning Model** $F_{PC}(n, p)$

For each literal v, independently
take $d(v) \in \text{Poi}(p(2n-1))$ clones and
match all clones uniformly at random.

E.g.

For each clone w, assign a uniform random (real) number from 0 to $\lambda := 2p(n-1)$.

We say that a clone is larger than another clone if the assigned numbers are so.

Take two largest clones and match them
and repeat it for the remaining, or unmatched, clones.

Or,

choose the first unmatched clone
according to a certain selection rule (SR)
without looking assigned numbers,
then match it to the largest unmatched clone.

$F(n,p)$ vs. $F_{PC}(n,p)$

Lemma. If $pn^2 = O(n)$, then there are positive constants c_1 and c_2 so that for any collection $\mathcal{F}$ of (SIMPLE) formulae

$$c_1 \Pr[F_{PC}(n,p) \in \mathcal{F}] - e^{- \quad (pn^2)} \cdot \quad \Pr[F(n,p) \in \mathcal{F}]$$

and

$$\Pr[F(n,p) \in \mathcal{F}] \cdot \quad c_2 \Big(\Pr[F_{PC}(n,p) \in \mathcal{F}] \Big)^{1/2} + e^{- \quad (pn^2)}.$$

- Pure Literal of a formula F

  a literal x is pure iff v appears in F but not $\bar{v}$

- Light Clone

  clones of pure literals are called light.

- Pure Literal Rule for $F_{PC}(n, p)$

  take a light clone and match it to the largest (unmatched) clone

- **Parameters**

$$N_i = \text{the number of unmatched clones},$$

$$N_0 \approx 2\lambda n.$$

$$\lambda_i = \text{the length of the interval},$$

$$\lambda_0 = \lambda.$$

$$H_i = \text{the number of unmatched heavy clones},$$

$$H_0 = \sum_{j=1}^{n} (X_j + Y_j) 1(X_j Y_j \geq 1),$$

$$E[H_0] = 2\lambda(1 - e^{-\lambda}),$$

where X_i, Y_i's are i.i.d. Poi(λ).

Generally, if λ_i is given,

$$H_i = \sum_{j=1}^{n} (X_j^{(i)} + Y_j^{(i)}) 1(X_j^{(i)} Y_j^{(i)} \geq 1),$$

and

$$E[H_i] = 2\lambda_i (1 - e^{-\lambda_i}),$$

where $X_j^{(i)}, Y_j^{(i)}$'s are i.i.d. Poi(λ_i).

Note that

$$N_i - H_i = \text{the total number of light clones.}$$

Thus

$$\exists \text{ no } (1,1)\text{-core} \quad \text{iff} \quad N_i - H_i > 0 \text{ for all } i.$$

Trivially,

$$N_i = N_0 - 2i.$$

In expectation,

$$\lambda_1 = \Big(1 - \frac{1}{N_0 - 1}\Big)\lambda$$

since we took the largest number among $N_0 - 1$ i.i.d uniform random numbers from 0 to λ. Similarly, in expectation,

$$\lambda_i = \lambda \prod_{j=0}^{i-1} \Big(1 - \frac{1}{N_i - 1}\Big).$$

Actually,

$$\lambda_i = \lambda \prod_{j=0}^{i-1} \Big(1 - T_i\Big),$$

where T_i are mutually independent with

$$\Pr[T_i \geq t] \approx e^{-(N_{i-1}-1)t}.$$

Thus, for $\theta_i^2 := N_i/N_0$

$$\lambda_i \approx \lambda \exp\Big(- \sum_{j=0}^{i-1} \frac{1}{N_p - 2j - 1}\Big) \approx \theta_i \lambda.$$

In terms of θ_i,

$$N_i = \theta_i^2 N \approx 2\theta_i^2 \lambda n,$$

and

$$H_i \approx 2\lambda_i(1 - e^{-\lambda_i})n \approx 2\theta_i\lambda(1 - e^{-\theta_i\lambda}).$$

Since $\exists$ no $(1,1)$-core iff $N_i - H_i > 0$ for all i,

for $f(\theta) := \theta - (1 - e^{-\theta\lambda})$,

$$\exists \text{ no } (1,1)\text{-core iff } f(\theta) > 0 \quad \forall\theta \in (0,1),$$

or

$$\exists\ (1,1)\text{-core} \quad \text{iff} \quad f(\theta) = 0 \quad \text{for some } \theta \in (0,1).$$

Since $f(0) = 0$ and

$$f'(\theta) = 1 - \lambda e^{-\theta\lambda} = \begin{cases} > 0 \ \text{ for all } \theta \in (0,1) & \text{if } \lambda \cdot \ \ 1 \\ < 0 \ \text{ for all } \theta \in (0,\delta_\lambda) & \text{if } \lambda > 1 \end{cases}$$

for some $\delta_\lambda > 0$,

$$f(\theta) = 0 \text{ has a solution in } \theta \in (0,1) \text{ iff } \lambda > 1 \ .$$

Pure Literal Alg. for k-SAT

In terms of θ_i,

$$N_i = \theta_i^k N \approx 2\theta_i^k \lambda n,$$

and

$$H_i \approx 2\lambda_i(1 - e^{-\lambda_i})n \approx 2\theta_i^{k-1}\lambda(1 - e^{-\theta_i^{k-1}\lambda}).$$

For

$$f_k(\theta) := \theta - (1 - e^{-\theta^{k-1}\lambda})) = \theta - P(\theta^{k-1}\lambda, 1),$$

$$\exists \text{ no } (1,1)\text{-core iff } f(\theta) > 0 \quad \forall \theta \in (0,1),$$

or

$$\exists\ (1,1)\text{-core} \quad \text{iff} \quad f(\theta) = 0 \quad \text{for some } \theta \in (0,1).$$

Thus

$$\lambda_k = \min_{\rho>0} \frac{\rho}{(1-e^{-\rho})^{k-1}}, \quad \text{or} \quad m_k = \min_{\rho>0} \frac{2\rho}{k(1-e^{-\rho})^{k-1}}$$

E.g.,

$$m_3 = 1.63694\cdots, \quad m_4 = 1.54456\cdots, \quad m_5 = 1.40356\cdots.$$

(Mitzenmacher('97), Molloy & Wormald(?))