Contents

Introduction

Notation

Chapter 1

Preliminaries

Chapter 2

Probabilities of Sentences about Very Sparse Random Graphs

In this chapter we will review the results obtained in the paper with the same name by James F. Lynch [1]. In there, limit probabilities of sentences in the first order language of graphs $\mathcal{L}$ are discussed for the binomial model $G(n, p)$ in the cases $p = \beta/n$ and $p = \beta n^{-\alpha}$ with $\alpha = (l+1)/l$.

More precisely, it is proven that in those cases the probability of every sentence converges and it is shown that for any of those sentences, its limit probability is among the values taken by some analytic formulas with parameter β.

We are interested in the case $p = \frac{\beta}{n}$, which is the one discussed more extensively in [1]. According to the author, the relevant theorems for the other case can be proven analogously. From now on we will only refer as random graphs to the ones in $G(n, \beta/n)$

From now on we will denote by Poi_λ the probability function of the Poisson distribution with mean λ. That is, the one given by $Poi_\lambda(n) = e^{-\lambda}\lambda^n/n!$ for any $n \in \mathbb{N}$. Also, we define $Poi_\lambda(\leq n)$ and $Poi_\lambda(> n)$ as $\sum_{i=0}^{n} Poi_\lambda(n)$ and $1 - Poi_\lambda(\leq n)$ respectively. Notice that for a fixed n, both $Poi_\lambda(\leq n)$ and $Poi_\lambda(> n)$ can be considered real functions of parameter λ.

We define the following sets of functions. Let Λ be the smallest set of expressions with parameter β such that:

- $1 \in \Lambda$,
- For any $\lambda \in \Lambda$ and $i \in \mathbb{N}$, both $Poi_{\beta\lambda}(n)$ and $Poi_{\beta\lambda}(> n)$ are in Λ.
- For any $\lambda_1, \lambda_2 \in \Lambda$, $\lambda_1\lambda_2$ belongs to Λ as well.

And let Θ be the smallest set of functions with parameter β such that:

- For any $\lambda \in \Lambda$ and $n, a, i \in \mathbb{N}$, with $i \geq 3$, both $Poi_{\beta^i\lambda/a}(\leq n)$ and $Poi_{\beta^i\lambda/a}(> n)$ are in Θ.

The main result is the following:

Theorem 2.1 (Lynch, 1992)**.** *Let ϕ be a sentence in the first order theory of graphs. Then the limit $\lim\limits_{n\to\infty} P(G(n, \beta/n) \models \phi)$ exists for all positive real numbers β, and it is a finite sum of expressions in Θ.*

We show now an outline of the proof.

We show that for any quantifier rank k there are some classes of graphs $C_1^k, \ldots, C_{n_k}^k$ such that

(1) a.a.s the rank k type of any two graphs in the same class coincide,
(2) a.a.s. any random graph belongs to some of them, and
(3) the limit probability of random graph belonging to any of them is an expression in Θ.

After this is archived the theorem follows easily. Indeed, let ϕ be a sentence in the first order language $\mathcal{L}$ of graphs whose quantifier rank is k, and denote by G a random graph in $G(n, \beta/n)$. We define the events $E_1, \ldots, E_{n_k}$ as

$$E_i := (G \models \phi) \wedge (G \in C_i),$$

and the event F as

$$F := (G \models \phi) \bigwedge_{i=1}^{n_k} (G \notin C_i).$$

Then, for any $n \in \mathbb{N}$

$$P(G \models \phi) = \sum_{i=1}^{n_k} P(E_i) + P(F), \tag{2.1}$$

as the events E_i together with F form a partition of all the cases where G satisfies ϕ.

Fix and index $i \in \{1, \ldots, n_k\}$. From the property (1) of the classes $C_1, \ldots, C_{n_k}$ it follows that if G and H are random graphs, then

$$\lim_{n\to\infty} P((G \models \phi) \wedge \neg(H \models \phi) \,|\, G \in C_i \wedge H \in C_i \,) = 0.$$

This is because G and H share a.a.s the same rank k type if they both belong to C_i. In consequence the limit

$$\lim_{n\to\infty} P(G \models \phi \,|\, G \in C_i)$$

is either zero or one, and

$$\lim_{n\to\infty} P(E_i) = \lim_{n\to\infty} P(G \in C_i) \cdot P(G \models \phi \,|\, G \in C_i) = \text{ either } 0 \text{ or } \lim_{n\to\infty} P(G \in C_i).$$
$$\tag{2.2}$$

Also, as a consequence of property (2) we obtain

$$\lim_{n\to\infty} P(\bigwedge_{i=1}^{n_k} G \notin C_i) = 0,$$

6

so

$$\lim_{n\to\infty} P(F) = \lim_{n\to\infty} P(\bigwedge_{i=1}^{n_k} G \notin C_i) \cdot P(G \models \phi \mid \bigwedge_{i=1}^{n_k} G \notin C_i) = 0. \qquad (2.3)$$

Taking limits in equation 2.1 and using equations 2.2 and 2.3 we get

$$\lim_{n\to\infty} P(G \models \phi) = \sum_{C_i \in \mathcal{C}} \lim_{n\to\infty} P(G \in C_i),$$

where $\mathcal{C}$ is a (possibly empty) subset of $\{C_1, \ldots, C_{n_k}\}$. Finally, because of property (3) for each i the limit $\lim_{n\to\infty} P(G \in C_i)$ is an expression in Θ. Thus $\lim_{n\to\infty} P(G \models \phi)$ is a finite sum of expressions in Θ and the theorem follows.

The objective of next sections will be to define the classes $C_1, \ldots, C_{n_k}$ and to show that they satisfy properties (1), (2) and (3). Later we will prove a stronger result, so we will allow ourselves to just sketch some of the proofs during this chapter.

2.1 Agreability Classes

It is known that $n^{-v/e}$ is the t hreshold probability for the appearance of a balanced graph of density v/e. In our case $v/e = 1$, so in consequence any connected graph H with $e(H) < v(H)$ a.a.s will not appear as a subgraphs of $G(n, \beta/n)$. It can be easily shown that such graphs H are exactly the ones containing more than one cycle.

If H is a connected graph with $v = e$, then H is an uni-cyclic graph. In this case, the number X_H of copies of H in $G(n, \beta/n)$ will asymptotically have non-zero bounded expectancy m. It does not take much work to prove, using Brun's sieve, that X_H converges in distribution to a Poisson random variable with mean m as n goes to infinity.

Finally, if H is a connected graph with $v > e$ then it must be a tree. Here the expected number of copies of H in $G(n, \beta/n)$ diverges asymptotically. Informally, trees of any kind will occur arbitrarily often.

This all means, in a sense, that a.a.s the only difference between large graphs in $G(n, \beta/n)$ lies in their uni-cyclic subgraphs. More precisely, because of the "locality" of first order logic of quantifier rank k we will only be interested in the "small" neighborhoods of the "short" cycles. Thus, our goal will be to classify uni-cyclic graphs in a way that respects equivalence under first order logic of quantifier rank k.

To make our classification suitable for proofs involving E.F. games we need to work graphs to which we "attach" labels. We define the set of symbols $Const = \{c_i\}_{i\in\mathbb{N}}$ as the set of constants. Also, we will denote by $Const_n$ the set $\{c_1, \ldots, c_n\}$.

Definition 2.1. A **co-labeling** of a graph $G = (V, E)$ is a map $\sigma : D \to V$, where $D \subset C$ is a finite set of constant symbols. Given $c_i \in D$, we will say that the vertex $\sigma(c_i)$ is labeled c_i. Equivalently, we can denote a labeling σ as a tuple $(c_{i_1}[x_1], \ldots, c_{i_m}[x_m])$ where each c_{i_j} is a constant symbol, and x_j is the vertex in V labeled c_{i_j}.

Definition 2.2. A **graph with constants**[1] $G(c_{i_1}[x_1], \ldots, c_{i_m}[x_m])$ is a graph G together with a co-labeling $(c_{i_1}[x_1], \ldots, c_{i_m}[x_m])$.

To keep our notation compact we will often drop he x_i's and say $G(c_{i_1}, \ldots, c_{i_m})$.

Definition 2.3. Let G be a graph with constants. A subgraph H of G is a graph with constants such that $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$ and all vertices in $V(H)$ have the same labels in H and G.

An important abuse of notation we are going to make will be to identify the constants c_i with their labeled vertices $\sigma(c_i)$. This way things like $c_i \sim c_j$ will make sense. In this context, notice that the expression $c_i = c_j$ is ambiguous because the vertices labeled c_i and c_j may be the same for some $i \neq j$, but the constant symbols c_i and c_j will be equal only if $i = j$. We will make sure to leave no room for ambiguity in this situations.

Proposition / Definition 2.1. Let $G = (V, E, c_{i_1}, \cdots, c_{i_m})$ be a connected graph with constants. Then it has a unique minimal connected subgraph H containing all its constants and cycles. We will call the **center** of G to such subgraph and denote it by $Center(G)$. If $\bar{G}$ is an arbitrary graph with constants, then its center $Center\bar{G}$ will be the union of the centers of its connected components.

Proof. TO DO ▢

For an arbitrary graph with constants we define the metric $d(\cdot, \cdot)$ on $V(G)$ as the one such that $d(x, y)$ is the minimum length of a path connecting x and y in G or ∞ if such path does not exist. For any vertex $x \in V(G)$ and $r \in \mathbb{N}$ we define the co-labeled subgraph $N(x; r)$ as the ball of radius r centered at v. That is, the induced subgraph with vertex set

$$V(N(x; r)) = \{\, y \in V \mid d(x, y) \leq r \,\}.$$

In a similar vein, given $X \subseteq V(G)$ we define its neighborhood of radius r as the induced co-labeled subgraph $N(X; r)$ whose vertex set is

$$V(N(X; r)) = \{\, y \in V \mid \forall x \in X : \ d(x, y) \leq r \,\}.$$

Let $G = (V, E)$, and $V' \subseteq V$. Another important abuse of notation we will make is writing $H = (V', E)$ for a subgraph H to mean that the edge set of $E(H)$ is the one induced by $E(G)$ on V'.

---

[1] Compare with [1], where they are called "rooted graphs".

Definition 2.4. A **rooted tree** $T = (V, E, x)$ is a tree (V, E) with distinguished vertex $x \in V$ with we will call **root** of the tree.

Proposition / Definition 2.2. Let $G = (V, E, c_{i_1}, \cdots, c_{i_m})$ be a connected graph and $x \in V$. Then define $Tree(x, G)$ as the rooted tree

$$Tree(x, G) = (V_x, E, x),$$

where

$$V_x = \{ y \in V \mid d(Center(G), y) = d(Center(G), x) + d(x, y) \}.$$

Proof. TO DO $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

The radius $r(T)$ of a rooted tree $T = (V, E, x)$ is the maximum distance between its root x and any other of its vertices. The branches of T are the rooted trees of the form $Tree(y, T)$, where $y \sim x$. We will denote by $Br(T)$ the set of branches of T.

We begin by defining an equivalence relation between rooted trees for each quantifier rank k.

Definition 2.5. Let $k \in \mathbb{N}$ with $k \geq 1$. The **k-morphism** equivalence relation $\overset{k}{\simeq}$ between graph with constantss is the one inductively defined as follows:

- If T_1, T_2 are rooted trees of radius 0 -i.e., they consist only of their roots- they are k-morphic.

- Let T_1, T_2 be rooted trees of radius r whose rots have the same label. Then $T_1 \overset{k}{\simeq} T_2$ if for any k-morphism class C of trees with radius less than r and root either

  "T_1 and T_2 have the same number of branches of type C"
  $$|Br(T_1) \cap C| = |Br(T_2) \cap C|,$$

  or

  "T_1 and T_2 have both more than k branches of type C"
  $$|Br(T_i) \cap C| \geq k + 1 \text{ for } i = 1, 2.$$

It follows from the definition that k-morphic trees have the same radius. It is also easy to check that the k-morphism relation is indeed an equivalence one.

Proposition 2.1. *For all $k, r \in \mathbb{N}$ and with $k \geq 1$, the set of classes of k-morphic trees with radius lesser or equal than r is finite.*

Proof. TO DO $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

We define now the k-morphism relation for arbitrary graph with constantss.

Definition 2.6. Let $G^1 = (V^1, E^1, c_{i_1}[x_1^1], \ldots, c_{i_m}[x_m^1])$ and $G^2 = (V^2, E^2, c_{i_1}[x_2^1],$ $\ldots, c_{i_m}[x_m^2])$ be graph with constantss with the same constant symbols. We will say that they are k-morphic (denoted by $G^1 \overset{k}{\simeq} G^2$) if there is a bijection $f :$ $V(Center(G^1)) \to V(Center(G^2))$ such that

- "f preserves edges"

$$\forall x, y \in V(Center(G^1)): \quad x \sim y \iff f(x) \sim f(y).$$

- "f preserves labels"

$$\forall j \in \{1, \ldots, m\}: \quad f(x_j^1) = x_j^2.$$

- "f preserves k-morphism classes of trees"

$$\forall x \in V(Center(G^1)): \quad Tree(x, G^1) \overset{k}{\simeq} Tree(f(x), G^2).$$

In this case we will say that $G^1 \overset{k}{\simeq} G^2$ via f.

We are going to show that the rank k type of a random graph a.a.s only depends on the neighborhoods of its small cycles. In consequence the following definition is motivated:

Definition 2.7. Let $G = (V, E, c_{i_1}, \cdots, c_{i_m})$ be a graph with constants. Then its core of radius r, $Core(G, r)$ is the co-labeled subgraph $N(X; r)$, where X is the union of the (vertex sets of the) cycles in G with size at most $2r + 1$ and all of the labeled vertices in G.

Chapter 3

First Order Logic of Sparse Random General Hypergraphs

3.1 General Hypergraphs

$\forall n \in N : \quad [n] := \{1, \ldots, n\}$

S_n : symmetric group on n elements.

Let X be a set, then:

- $\Delta \subset X^2 : \Delta$ is the diagonal subset. $\Delta : \{(a, a) \in X^2\}$.

- S_n acts on X^n in the obvious way.

- When the subgroup ϕ is understood or (is not relevant) $\langle x_1, \ldots, x_n \rangle$ denotes the equivalence class of $(x_1, \ldots, x_n)$ in X^n/ϕ.

Definition 3.1. Let V be a finite set. Given a natural number $a \in \mathbb{N}$, a group $\phi \leq S_a$ and a subset $R \subseteq [a]^2 \setminus \delta$, the total (hiper-)edge set of size a, symmetry group ϕ and restrictions R is the set:

$$\mathcal{H}_{V,a}^{\phi,R} = V^n/\phi \; \setminus \{ \langle x_1, \ldots, x_n \rangle \in V^n/\phi \mid x_i = x_j \text{ for some } (i, j) \in R\}$$

Definition 3.2. An hypergraph $(V, H_1, \ldots, H_l)$ with edge sizes $a_1, \ldots, a_l$, symmetry groups $\phi_1, dots, \phi_1$ and restrictions $R_1, \ldots, R_l$ consists of

- A finite vertex set V.

- For $i = 1, \ldots, l$, an edge set $H_i \subseteq \mathcal{H}_{V,a_i}^{\phi_i, R_i}$.

Definition 3.3. Random model $G(n, p_1, \ldots, p_l)$ with edge sizes $a_1, \ldots, a_l$, symmetry groups $\phi_1, dots, \phi_1$ and restrictions $R_1, \ldots, R_l$:

- We obtain a hypergraph $(V, H_1, \ldots, H_l)$ with $V = [n]$

- The probability of each edge $e \in \mathcal{H}_{V,a_i}^{\phi_i,R_i}$ is p_i independently.

From now on we fix the edge sizes $a_1, \ldots, a_l$, symmetry groups $\phi_1, dots, \phi_1$ and restrictions $R_1, \ldots, R_l$.

In a graph $G(V, E_1, \ldots, E_l)$, the type of an edge e is the index i such that $e \in E_i$.

The language is the first order relational language with relations $R_1, ..., R_l$ of arities $a_1, \ldots, a_l$ respectively. The interpretation on $(V, H_1, \ldots, H_l)$ is the evident one.

Definition 3.4. The characteristic $ch(G)$ of an hypergraph $G = (V, E_1, \ldots, E_l)$ is

$$\left(\sum_{i=1}^{l} |E_i|(a_i - 1) \right) - |V|.$$

An hypergraph is c-balanced if it contains no subgraph of greater characteristic than itself.

Proposition 3.1. *A connected graph cannot have characteristic lesser than* -1.

Definition 3.5. A unicycle is a connected c-balanced graph with characteristic 0.

Definition 3.6. A cluster is a minimal c-balanced graph.

Remark: Note that clusters have to be connected.

Definition 3.7. A cycle is a minimal unicycle, or equivalently a characteristic 0 cluster.

Proposition 3.2. *Any unicycle contains exactly one cycle.*

Definition 3.8. A (hiper)forest is a graph with no clusters. A (hiper)tree is a connected forest.

Definition 3.9. A path between x and y is a connected graph containing both x and y that is minimal among the ones with those properties.

Proposition 3.3. *There is a unique path between any two vertices of a tree.*

We define the constant set $Const := \{c_i\}_{i \in \mathbb{N}}$.

Definition 3.10. As before, a (hiper)graph with constants $G(c_{i_1}, \cdots, c_{i_m})$ is a hypergraph G together with a co-labeling.

Definition 3.11. The center of a connected graph with constants $G(c_{i_1}, \cdots, c_{i_m})$ is its minimal connected subgraph containing all the constants and clusters. If $G(c_{i_1}, \cdots, c_{i_m})$ is not connected then its center is the union of the centers of its connected components.

We define a distance over $G(c_{i_1}, \cdots, c_{i_m})$.

$$d(x,y) = (\text{minimum size of a connected graph containing } x \text{ and } y) - 1$$

If such graph does not exist we define $d(x,y) = \infty$.

Given $x \in V(G)$ and $X \subseteq V(G)$ we define $N(x;r)$ and $N(X;r)$ as

$$N(x;r) = G[S], \quad \text{where } S = \{\, y \in V(G) \,|\, dist(x,y) \le r \,\}$$

$$N(X;r) = G[S], \quad \text{where } S = \{\, y \in V(G) \,|\, dist(X,y) \le r \,\}.$$

Definition 3.12. A rooted (hiper)tree is a tree with a distinguished vertex called root. The radius of the tree is the maximum distance between its root and any of its vertices.

In a rooted tree (T,x) all the (hiper)edges can be rooted in a canonical way. The root of an edge e is the vertex $y \in e$ such that $dist(x,y) = dist(x,e)$.

Definition 3.13. Let $G(c_{i_1}, \cdots, c_{i_m})$ be a graph with constants and $x \in V(G)$. Then $Tree(x, G) = (G[V_x], x)$, where

$$V_x = \{\, y \in V \,|\, dist(Center(G), y) = dist(Center(G), x) + dist(x,y) \,\}.$$

Definition 3.14. The k-morphism equivalence relation $\overset{k}{\simeq}$ between rooted trees, and the $\overset{k}{\sim}$ relation for rooted (hyper)edges are defined inductively as follows:

- If T_1 and T_2 are radius 0 trees then they consist only of their roots and $T_1 \overset{k}{\simeq} T_2$.

- If e_1 and e_2 are edges in rooted trees T_1, T_2 of radii at most r, then $e_1 \overset{k}{\sim} e_2$ if e_1 and e_2 are edges of the same type and $e_1 = [x_1, \ldots, x_m]$, $e_2 = [y_1, \ldots, y_m]$ in a way that

  - x_i is the root of e_1 if and only if y_i is the root of e_2.
  - If x_j is not the root of e_1, then $Tree(x_j, T_1) \overset{k}{\simeq} Tree(y_j, T_2)$.

- If T_1 and T_2 are trees of radii at most r and roots x and y, then $T_1 \overset{k}{\simeq} T_2$ means that for any $\overset{k}{\sim}$ class of edges C either

  - the number of edges in T_1 of type C containing x is the same as the number of edges in T_2 of tipe C containing y, or
  - both T_1 and T_2 contain no less than $k+1$ edges of type C containing x and y respectively.

Definition 3.15. Let $G_1(c_{i_1}, \cdots, c_{i_m})$ and $G_2(c_{i_1}, \cdots, c_{i_m})$ be graphs with the same constants. Then we say $G_1(c_{i_1}, \cdots, c_{i_m}) \overset{k}{\simeq} G_2(c_{i_1}, \cdots, c_{i_m})$ if there is a bijection $f : V(Center(G_1)) \to V(Center(G_2))$ such that

- $f(c_i^1) = c_i^2$ for all constants, and

- $Tree(x, G_1) \overset{k}{\simeq} Tree(f(x), G_2)$ for all $x \in V(Center(G_1))$.

The diameter of a (hiper)graph is the maximum distance between any two of its vertices.

Definition 3.16. Let $G(c_{i_1}, \cdots, c_{i_m})$ be a graph with constants. We define $Core(G, r)$ to be $N(X, r)$ where X is the union of constants and clusters of diameter at most r.

Definition 3.17. A (hyper)graph G is r-simple if $Core(G, r)$ is a union of unicycles.

Definition 3.18. A (hyper)graph G is i, j, r-rich for some $i, j, r \in \mathbb{N}$, if for any rooted tree T of radius at most r there are $x_1, \ldots, x_i \in V(G)$ such that

- The $N(x_l; r)$'s are disjoint.

- The $N(x_l; r)$'s do not intersect $Core(G, r)$

- $N(x_l; r)(x_l) \overset{j}{\simeq} T$.

Theorem 3.1. If $G_1(c_1, \ldots, c_m) \overset{k}{\simeq} G_2(c_1, \ldots, c_m)$, then for any $x \in V(G_1)$ such that $d(x, Center(G_1)) \neq \infty$ there is a $y \in V(G_2)$ such that $G_1(c_1, \ldots, c_m, c_{m+1}[x]) \overset{k-1}{\simeq} G_2(c_1, \ldots, c_m, c_{m+1}[y])$.

Theorem 3.2. If $Core(G_1, r) \overset{k}{\simeq} Core(G_2, r)$ then for any $s < r$, $Core(G_1, s) \overset{k}{\simeq} Core(G_2, s)$.

Definition 3.19. Two graphs with constants G_1 and G_2 are k-agreeable if for each k-morphism equivalence class C they either have the same number of connected components of type C or they both have no less than $k + 1$ components of type C.

Theorem 3.3. *Let G_1, G_2 be graphs with constants satisfying:*

- *They are both $(k, k, 3^k)$-rich.*

- *$Core(G_1, 3^k)$ is k-agreeable with $Core(G_2, 3^k)$.*

Then the G_1 and G_2 share the same rank k type.

From now on fix $\beta_1, \ldots, \beta_l$ positive real numbers.

Theorem 3.4. *For any r,*

$$\lim_{n \to \infty} Pr(G(n, \beta_1/n^{a_1-1}, \ldots, \beta_l/n^{a_l-1}) \text{ is } r\text{-simple}) = 1.$$

Define the families of expressions with arguments $\beta_1, \ldots, \beta_l$: Λ, $\mathcal{E}$ and Θ.

Theorem 3.5. *Let $\phi(x_1, \ldots, x_l)$ be an open formula, and let $k, r, s \in \mathbb{N}$ $s \leq l$. Then*

$$\lim_{n \to \infty} Pr(\bigwedge_{i=i}^{s} Tree(x_i, G) \overset{k}{\simeq})$$

Bibliography

[1] James F Lynch. Probabilities of sentences about very sparse random graphs. *Random Structures & Algorithms*, 3(1):33–53, 1992.