On the distribution of the largest connected component size in random graphs with fixed edges set size

Robin Petit

March-April 2017

Contents

1 Introduction

1.1 Definitions and preliminary results

Let's consider $V = \{v_1, \dots, v_{|V|}\}$ a set of vertices. We denote by $|V|$ the cardinality of the set V. Let's define the function:

$$X : \mathbb{N} \to \mathbb{N} : n \mapsto \frac{n(n-1)}{2}.$$

Definition 1.1. An undirected graph Γ is denoted $\Gamma(V, E)$ for V its vertices set, and E its edges set, with $E = \{e_1, \dots, e_{|E|}\}$ and $\forall i \in [\![1, |E|]\!] : e_i = \{v_{i1}, v_{i2}\}$ for $1 \leq i_1, i_2 \leq |V|$.

Remark. $|E|$ is usually denoted as m, and $|V|$ is sometimes denoted as n. Both these numbers are (non-strictly) positive integers.

Definition 1.2. The set of all the existing graphs having given vertices set V is denoted by $\Gamma(V, \cdot)$. We denote $\Gamma_m(V, \cdot)$ the subset of $\Gamma(V, \cdot)$ such that $|E| = m$.

Remark. We observe that:

$$\Gamma(V, \cdot) = \bigsqcup_{m \in \mathbb{N}} \Gamma_m(V, \cdot).$$

Definition 1.3. For every $n \in \mathbb{N}$, we define $\mathcal{K}_n$ as the *complete graph* of size n.

Lemma 1.4. *For a graph $\Gamma(V, E)$, we have $|E| \leq X(|V|)$.*

Proof. We know that $\Gamma(V, E) \leq \mathcal{K}_{|V|}$, and $\mathcal{K}_{|V|}$ has exactly $X(|V|)$ edges (vertex v_i is connected to vertices v_{i+1} to $v_{|V|}$, so the number of edges is equal to $\sum_{i=1}^{|V|}(|V| - i) = \sum_{i=0}^{|V|-1} i = X(|V|)$). $\square$

Lemma 1.5. *For given vertices set V and fixed number of edges $m \in \mathbb{N}$, we have:*

$$\left|\Gamma_m(V, \cdot)\right| = \begin{cases} \binom{X(|V|)}{m} & \text{if } m \leq X(|V|) \\ 0 & \text{else} \end{cases}.$$

Corollary 1.6. *For given vertices set V, we have* $\left|\Gamma(V,\cdot)\right| = 2^{X(\!|V|\!)}$.

Proof. Since $\Gamma(V,\cdot)$ is given by a disjoint union over m, its cardinality is equal to the sum of the individual cardinalities:

$$\left|\Gamma(V,\cdot)\right| = \sum_{m\in\mathbb{N}} \left|\Gamma_m(V,\cdot)\right| = \sum_{k=0}^{X(\!|V|\!)} \left|\Gamma_m(V,\cdot)\right| = \sum_{k=0}^{X(\!|V|\!)} \binom{X(\!|V|\!)}{m} = 2^{X(\!|V|\!)}.$$

□

Definition 1.7. A graph $\Gamma(V,E)$ is said to be connected if for each $v,w \in V$, there exists a path between v and w. We denote by $\chi(V,\cdot)$ the set of all connected graphs having vertices set V. Again, for $m \in \mathbb{N}$, we denote by $\chi_m(V,\cdot) \subset \chi(V,\cdot)$ the set of connected graphs having m edges.
Remark. $\chi(V,\cdot) \subset \Gamma(V,\cdot)$, and:

$$\chi(V,\cdot) = \bigsqcup_{m\in\mathbb{N}} \chi_m(V,\cdot).$$

Lemma 1.8. *For $m < |V|$ or $m > X(\!|V|\!)$, we have* $\left|\chi_m(V,\cdot)\right| = 0$.
Definition 1.9. For every $W \in \mathcal{P}(V)$, we define $\Delta_W : \Gamma(V,\cdot) \to \Gamma(W,\cdot) : \Gamma(V,E) \mapsto \Gamma'(W,E')$ such that:

$$E' = \left\{\{v_i,v_j\} \in E \text{ s.t.} v_i, v_j \in W\right\}.$$

Definition 1.10. We define the *connected component of vertex $v_i \in V$ in graph* $\Gamma(V,E)$ by the biggest subset (in the sense of inclusion) W of V such that $\Delta_W(\Gamma(V,E)) \in \chi(W,\cdot)$.

We then define the *largest connected component of the graph* $\Gamma(V,E)$ as:

$$\mathrm{LCC}(\Gamma(V,E)) \coloneqq \underset{\substack{W\in\mathcal{P}(W)\\ \Delta_W(\Gamma(V,E)\in\chi(W,\cdot)}}{\arg\max} |W| = \underset{W\in\mathcal{P}(V)}{\arg\max} |W|\, \mathbb{I}_{\left[\Delta_W(\Gamma(V,E)\in\chi(W,\cdot)\right]}.$$

The set $\Lambda_k^m(V,\cdot)$ is then the set of all graphs $\Gamma(V,E) \in \Gamma(V,\cdot)$, such that $|E| = m$ and $\left|\mathrm{LCC}(\Gamma(V,E))\right| = k$.
Remark. Since $\Lambda_k(V,\cdot) = \bigsqcup_{m=0}^{X(\!|V|\!)} \Lambda_k^m(V,\cdot)$ and:

$$\Gamma(V,\cdot) = \bigsqcup_{k=1}^{|V|} \bigsqcup_{m=0}^{X(\!|V|\!)} \Lambda_k^m(V,\cdot),$$

we want to know what is $\left|\Lambda_k^m(V,\cdot)\right|$ equal to.
Definition 1.11. Let's declare a new random variable $\mathscr{G}(V)$, a graph uniformly distributed in $\Gamma(V,\cdot)$, thus such that:

$$\forall\Gamma(V,E) \in \Gamma(V,\cdot) : \mathbb{P}[\mathscr{G}(V) = \Gamma(V,E)] = \frac{1}{\left|\Gamma(V,\cdot)\right|} = 2^{-X(\!|V|\!)}.$$

1.2 Objectives

The objective now is to find an expression for $\left|\Lambda_k(V,\cdot)\right|$ since we are looking for:

$$\mathbb{P}[\mathrm{LCC}(\mathscr{G}(V)) = k] = \frac{\left|\Lambda_k(V,\cdot)\right|}{\left|\Gamma(V,\cdot)\right|} = \frac{1}{\left|\Gamma(V,E)\right|} \sum_{m=0}^{X(\!|V|\!)} \left|\Lambda_k^m(V,\cdot)\right|.$$

Let's denote this value $p_k \coloneqq \mathbb{P}\left[\left|\mathrm{LCC}(\mathscr{G}(V)\right| = k\right]$.

2 Results

The general idea in order to determine $\left|\Lambda_k^m(V,\cdot)\right|$ is to insert a connected component of size k on vertices set V, and then to tally the configurations placing $m-k$ vertices without making a bigger connected component than the first one.

2.1 Examples

It is trivial to tell $\left|\Lambda_1^m(V,\cdot)\right| = \delta_0^m$, i.e. equals one if $m = 0$ and equals zero if $m > 0$: a graph having at least one edge, cannot have a largest connected component of size 1.

Lemma 2.1 (Upper boundary of edges amount for $k = 2$)**.** *For $m > \frac{|V|}{2}$, we have $\Lambda_2^m(V,\cdot) = \emptyset$.*

Proof. To have a largest connected component of size 2, each vertex must have degree 0 or 1. Take $m \in \mathbb{N}$ such that $m > \frac{V}{2}$. Take $\Gamma(V,E)$ such that $|E| = m$, and take $\mathcal{V}_1 \coloneqq \{v \in V \text{ s.t.} \deg(v) \leq 1\} \subset V$. The restriction $\Gamma'(\mathcal{V}_1, E') = \Delta_{\mathcal{V}_1}(\Gamma(V,E))$.

Since in a graph, the sum of the degree of each vertex is equal to twice the amount of edges, when applied on Γ', it follows that:

$$2\left|E'\right| = \sum_{v\in\mathcal{V}_1} \deg(v) \leq \sum_{v\in\mathcal{V}_1} 1 = |\mathcal{V}_1|\,.$$

We then deduce that $\left|E'\right| \leq \frac{|\mathcal{V}_1|}{2} \leq \frac{|V|}{2}$. Thus $\mathcal{V}_1$ must be *strictly* included in V, and then there must exist $v \in V$ such that $\deg(v) \geq 2$. Thus:

$$\forall m > \frac{|V|}{2} : \forall \Gamma(V,E) \in \Gamma_m(V,\cdot) : \Gamma(V,E) \notin \Lambda_2^m(V,\cdot).$$

□

Lemma 2.2. *For $\Gamma(V,E) \in \Gamma(V,\cdot)$ a graph and $k \in [\![1,|V|]\!]$, if there exists a vertex $v \in V$ such that $\deg(v) = k$, then $\left|\mathrm{LCC}(\Gamma(V,E))\right| \geq k+1$.*

Proof. Take $v \in V$ such that $\deg(v) = k$. There exist $\{v_{i_1},\ldots,v_{i_k}\} \subset V$ such that:

$$\forall j \in [\![1,k]\!] : \{v, v_{i_j}\} \in E.$$

Thus $\{v, v_{i_1}, v_{i_k}\}$ is a connected component of size $k+1$. Thus the largest connected component must have size at least that big. □

Proposition 2.3 (Upper boundary of edges amount generalized)**.** *For $k \in [\![1,|V|]\!]$, and $m > \frac{|V|(k-1)}{2}$, we have $\Lambda_k^m(V,\cdot) = \emptyset$.*

Proof. Take $m > \frac{(k-1)|V|}{2}$, and $\Gamma(V,E) \in \Gamma_m(V,\cdot)$. Take $\mathcal{V}_k \coloneqq \{v \in V \text{ s.t.} \deg(v) \leq k-1\}$. Let $\Gamma'(\mathcal{V}_k, E')$ be defined by $\Delta_{\mathcal{V}_k}(\Gamma(V,E))$. We know that:

$$2\left|E'\right| = \sum_{v\in\mathcal{V}_k} \deg(v) \leq (k-1)|\mathcal{V}_k| \leq (k-1)|V|\,.$$

We deduce that $\left|E'\right| \leq \frac{(k-1)|V|}{2} < m = |E|$. Thus $|E| \gneqq \left|E'\right|$, and this implies that there exists $v \in V$ such that $\deg(v) \geq k$. By previous lemma, largest connected component size must be at least $k+1$. □