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 $1 \leqslant i_1, i_2 \leqslant |V|$ with $i_1 \neq i_2$ (i.e. loops are not tolerated).

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* $|E| \leqslant X(|V|)$.

Proof. We know that $\Gamma(V,E) \leqslant \mathcal{K}_{|V|}$, 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 $\sum_{i=1}^{|V|}(|V|-i) = \sum_{i=0}^{|V|-1} i = X(|V|)$). □

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

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

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 $v_i \in W$ and $\Delta_W(\Gamma(V,E)) \in \chi(W,\cdot)$.

For graph $\Gamma(V,E) \in \Gamma(V,\cdot)$, we define $\left|LCC(\Gamma(V,E))\right|$ by:

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

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

$$LCC(\Gamma(V,E)) \coloneqq \underset{\substack{W\in\mathcal{P}(V) \\ |W|=\left|LCC(\Gamma(V,E))\right|}}{\arg\min} \min_{i\in[\![1,|V|]\!]} i \times \mathbb{I}_{[v_i\in W]}.$$

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 $\left|LCC(\Gamma(V,E))\right| = k$.
Remark. The notations here are consistent since $\left|LCC(\Gamma(V,E))\right|$ corresponds indeed to the cardinality of $LCC(\Gamma(V,E))$.

Furthermore, this definition of largest connected component allows to define uniquely the LCC, even though a graph $\Gamma(V,E)$ has several connected component of same size. For example, following graph has two connected component of size 2, i.e. $\{1,2\}$ (in red) and $\{3,4\}$ (in blue).

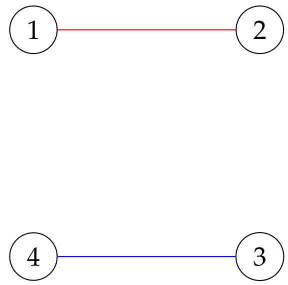

Figure 1: Graph $\Gamma\left(\{1,2,3,4\},\left\{\{1,2\},\{3,4\}\right\}\right)$

Nevertheless, the LCC operator yields $\{1,2\}$ since it minimizes the lowest id of element in connected component (1 for this graph).
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:

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

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

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

2.1.1 $\left|\Lambda_{k=1}(V,\cdot)\right|$

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.

2.1.2 Upper boundary of m for $\left|\Lambda_k^m(V,\cdot)\right|$

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 $\mathcal{V}_1 \coloneqq \{v \in V \text{ s.t.} \deg(v) \leqslant 1\} \subset V$. Take 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:

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

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

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

$\square$

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 $\left|\mathrm{LCC}(\Gamma(V,E))\right| \geqslant k+1$.*

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

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

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

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 $\mathcal{V}_k \coloneqq \{v \in V \text{ s.t.} \deg(v) \leqslant k-1\}$. Let $\Gamma'(\mathcal{V}_k, E')$ be defined by $\Delta_{\mathcal{V}_k}(\Gamma(V,E))$. We know that:

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

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

Remark. We can understand this upper boundary as $m > \frac{|V|(k-1)}{2} = \frac{|V|}{k}\frac{k(k-1)}{2} = \frac{|V|}{k} \cdot X(k)$. So in order to have a LCC of size k, edges can be distributed to make $\left\lfloor \frac{|V|}{k} \right\rfloor$ complete graphs having each $X(k)$ edges. The maximum amount of edges is then given by $\frac{|V|(k-1)}{2}$.

2.1.3 $\left|\Lambda_{k=2}(V,\cdot)\right|$

Example of size 2 is a bit more complicated:

$$\forall m \in \left[\!\left[1, \left\lfloor \frac{|V|}{2} \right\rfloor \right]\!\right] : \left|\Lambda_2^m(V,\cdot)\right| = \begin{cases} \frac{1}{m!} \prod_{k=0}^{m-1} \binom{|V|-2k}{2} & \text{if } m \leqslant \frac{|V|}{2} \\ 0 & \text{else} \end{cases}.$$

Proof. For $m > \frac{|V|}{2}$, result is shown in Lemma 2.1. The part $\prod_{k=0}^{m-1} \binom{|V|-2k}{2}$ corresponds to the choice of m edges without making a connected component of size $\geqslant 3$.

$\binom{|V|-2\cdot 0}{2}$ is the choice of the first edge (two vertices) among $|V|$ vertices, $\binom{|V|-2}{2}$ is the choice of the second edge (two vertices) among the $|V|-2$ vertices left, etc. At step ℓ, only $|V|-2(\ell-1)$ vertices are available because two are selected per step, and a selected vertex cannot be used again, otherwise its degree would be $\geqslant 2$, and then the largest component size would be $\geqslant 3$.

The $\frac{1}{m!}$ comes from the fact that the order the edges are selected doesn't matter (so for each choice of m edges, there are $m!$ permutations of these). ☐

Remark. This can also be expressed as:

$$\left|\Lambda_2^m(V,\cdot)\right| = \frac{1}{m!} \frac{|V|!}{2^m \left(|V|-2m\right)!},$$

by simplification of the product.

3 Processing on examples

$$\begin{aligned}
\left|\Lambda_3^0(V,\cdot)\right| = \left|\Lambda_3^1(V,\cdot)\right| &= 0 \\
\left|\Lambda_3^2(V,\cdot)\right| &= \binom{|V|}{3}\binom{3}{2} \\
\left|\Lambda_3^3(V,\cdot)\right| &= \binom{|V|}{3}\binom{3}{3} + \binom{|V|}{3}\binom{3}{2}\binom{|V|-3}{2} \\
\left|\Lambda_3^4(V,\cdot)\right| &= \binom{|V|}{3}\binom{3}{3}\binom{|V|-3}{2} + \binom{|V|}{3}\binom{3}{2}\binom{|V|-3}{3}\binom{3}{2} + \binom{|V|}{3}\binom{3}{2}\binom{|V|-3}{2}\binom{|V|-5}{2}.
\end{aligned}$$

Definition 3.1. Let's denote equally $\left|\Lambda_k^m(n)\right| = \Lambda_k^m(n) \equiv \left|\Lambda_k^m(V,\cdot)\right|$ for V such that $|V| = n$.

This notation allows to lighten the expressions.

Conjecture 3.2.

$$\left|\Lambda_k^m(|V|)\right| = \binom{|V|}{k} \sum_{\ell=k-1}^{\min\left(m, X(k)\right)} \left|\Lambda_k^\ell(k)\right| \sum_{p=1}^{k} \left|\Lambda_p^{m-\ell}(|V|-k)\right| \beta_{p\,\ell}(m,k,|V|),$$

with $\beta_{p\,\ell}(m,k,|V|)$, a coefficient.

Remark. The idea behind this formula is explained in introduction of Section 2: to find the amount of graphs having n vertices, m edges and a largest connected component of size k, let's place a connected component of size k somewhere in the graph (so choose k in $|V|$ vertices), and then multiply this by the amount of possible graphs of largest connected component of size $p \in \{1,\ldots,k\}$ (so lower or equal to k).

Idea of proof of conjecture. In order to prove the equality of the cardinalities, let's find a bijective function Ω between $\Lambda_k^m(V,\cdot)$ and a set like:

$$\mathfrak{Q}_k^m(V) \coloneqq \bigsqcup_{\substack{W\in\mathcal{P}(V)\\ |W|=k}} \bigsqcup_{\ell=k-1}^{\min\left(m,X(k)\right)} \Lambda_k^\ell(W,\cdot) \times \left(\bigsqcup_{p=1}^{k} \Lambda_p^{m-\ell}(V\setminus W,\cdot)\right).$$

□

Lemma 3.3. *The sets $\chi_\ell(V,\cdot)$ and $\Lambda_{|V|}^\ell(V,\cdot)$ are equal.*

Proof. A graph $\Gamma(V,E)$ is connected if and only if its largest connected component contains all its vertices, i.e. $\mathrm{LCC}(\Gamma(V,E)) = V$.

This is equivalent to say that $\left|\mathrm{LCC}(\Gamma(V,E))\right| = |V|$ since $\forall W \in \mathcal{P}(V) : |W| = |V| \Rightarrow V = W$:

$$\forall W \in \mathcal{P}(V) : \left|\left\{\widetilde{W} \in \mathcal{P}(V) \text{ s.t.} |W| = \left|\widetilde{W}\right|\right\}\right| = \binom{|V|}{|W|},$$

and $\binom{|V|}{|V|} = 1$, thus $\left\{W \in \mathcal{P}(V) \text{ s.t.} |W| = |V|\right\} = \{V\}$. □

3.1 Decomposing set $\Lambda_k(V)$

Definition 3.4. For $k \in \mathbb{N}$, and $\alpha \in \mathbb{N}$, we define:

$$\Lambda_{k,\alpha}(V,\cdot) \coloneqq \left\{\Gamma(V,E) \in \Lambda_k(V,\cdot) \text{ s.t. } \left\{W \in \mathcal{P}(V) \text{ s.t.} \Delta_W(\Gamma(V,E)) \in \chi(W) \text{ and } |W| = \left|\mathrm{LCC}(\Gamma(V,E))\right|\right\} = \alpha\right\},$$

the class of all graphs in $\Lambda_k(V,\cdot)$ having exactly α connected components of maximum size.
Remark. Even though several connected components of maximum size do exist in a graph, *the one* LCC is still defined unambiguously!

Lemma 3.5.

1. *For $k > |V|$ or $k = 0$, we have: $\forall \alpha \in \mathbb{N} : \Lambda_{k,\alpha}(V,\cdot) = \emptyset$.*
2. *For $k \in [\![1, |V|]\!]$ and $\alpha > \left\lfloor \frac{|V|}{k} \right\rfloor$, we have $\Lambda_{k,\alpha}(V,\cdot) = \emptyset$.*

Proof.

1. For $k > |V|$ or $k = 0$, it is obvious that: $\Lambda_k(V,\cdot) = \emptyset$ (and then $\Lambda_{k,\alpha}(V,\cdot)$).
2. Take such k and α. Assume (*ad absurdum*) that there exists $\Gamma(V,E) \in \Lambda_{k,\alpha}(V,\cdot)$. We have then $L_1,\dots,L_\alpha \in \mathcal{P}(V)$ such that $\forall i \in [\![1,\ \alpha]\!] : |L_i| = k$. Also, since the L_i's are connected component, they are disjoint, i.e. $\forall (i,j) \in [\![1,\ \alpha]\!]^2 : i \neq j \Rightarrow L_i \cap L_j = \emptyset$.

   Thus $\bigcup_{i=1}^{\alpha} L_i \subset V$, and $\sum_{i=1}^{\alpha} |L_i| \leqslant |V|$. But:

   $$\sum_{i=1}^{\alpha} |L_i| = \alpha k > \left\lfloor \frac{|V|}{k} \right\rfloor k > \frac{|V|}{k} k = |V|,$$

   which leads a contradiction: $|V| > |V|$. We deduce that $\Lambda_{k,\alpha}(V,\cdot) = \emptyset$.

□

Remark. Again, for $m \in [\![1, X(\!|V|\!)]\!]$, we define $\Lambda_{k,\alpha}^m(V,\cdot)$ by $\Lambda_k^m(V,\cdot) \cap \Lambda_{k,\alpha}(V,\cdot)$.

Corollary 3.6.

$$\forall k < |V| : \Lambda_k(V,\cdot) = \bigsqcup_{m=k-1}^{X(\!|V|\!)} \bigsqcup_{\alpha=1}^{\left\lfloor \frac{|V|}{k} \right\rfloor} \Lambda_{k,\alpha}^m(V,\cdot).$$

Proof. Unions are trivially disjointed.

Now show the equality. The right-hand side is trivially included in $\Lambda_k(V,\cdot)$ (by definition of $\Lambda_{k,\alpha}^m(V,\cdot)$).

Now take $\Gamma(V,E) \in \Lambda_k(V,\cdot)$. We know that $\Lambda_k^{|E|}(V,\cdot)$ with $|E| \leqslant X(\!|V|\!)$. As well, we know that the amount of connected components of size $\left|\mathrm{LCC}(\Gamma(V,E))\right| = k$ is at least 1 (because $\Gamma(V,E) \in \Lambda_k(V,\cdot)$), and lower or equal to $\left\lfloor \frac{|V|}{k} \right\rfloor$ by previous Lemma. □

From now on, let's write:

$$\mathfrak{Q}_{k,1}^m(V) \coloneqq \bigsqcup_{\substack{W \in \mathcal{P}(V) \\ |W|=k}} \bigsqcup_{\ell=k-1}^{\min\left(m, X(k)\right)} \chi_\ell(W,\cdot) \times \left(\bigsqcup_{p=1}^{k-1} \Lambda_p^{m-\ell}(V,\cdot) \right).$$

Proposition 3.7. *For* $k \in [\![1, |V|]\!]$ *and* $m \in [\![1, X(\!|V|\!)]\!]$, *we have:*

$$\Lambda_{k,1}^m(V,\cdot) \cong \mathfrak{Q}_{k,1}^m(V),$$

i.e. there exists a bijection between these two sets.

Proof. Consider the function defined by:

$$\Omega : \Lambda_{k,1}^m(V,\cdot) \to \mathfrak{Q}_{k,1}^m(V) : \Gamma(V,E) \mapsto \left(\Delta_{\mathrm{LCC}(\Gamma(V,E))}(\Gamma(V,E)), \Delta_{V \setminus \mathrm{LCC}(\Gamma(V,E))}(\Gamma(V,E)) \right).$$

We know that $\Omega(\Gamma(V,E))$ gives a partition of $\Gamma(V,E)$ because there doesn't exist $e = \{v_{i_1}, v_{i_2}\} \in E$ such that $v_{i_1} \in \mathrm{LCC}(\Gamma(V,E))$ and $v_{i_2} \in V \setminus \mathrm{LCC}(\Gamma(V,E))$ by definition of connected component. Thus if:

$$\left(\Gamma_1(\mathrm{LCC}(\Gamma(V,E)), E_{\mathrm{LCC}}), \Gamma_2(V \setminus \mathrm{LCC}(\Gamma(V,E)), E_{V \setminus \mathrm{LCC}}) \right) = \Omega(\Gamma(V,E)),$$

we have $E_{\mathrm{LCC}} \sqcup E_{V \setminus \mathrm{LCC}} = E$.

Let's prove that Ω is injective. Take $\Gamma_1(V,E_1)$ and $\Gamma_2(V,E_2)$ in $\Lambda_k^m(V,\cdot)$. Assume that $\Omega(\Gamma_1(V,E_1)) = \Omega(\Gamma_2(V,E_2))$. This implies that $\mathrm{LCC}(\Gamma_1(V,E_1)) = \mathrm{LCC}(\Gamma_2(V,E_2))$ (and thus that their complementary in V are equal as well) and that the restriction of $\Gamma_1(V,E_1)$ and $\Gamma_2(V,E_2)$ to their common LCC are equal. Their restriction to the complementary of the common LCC in V are equal as well.

Thus $\Gamma_1(V,E_1)$ and $\Gamma_2(V,E_2)$ have the same partition and are then equal (i.e. $E_1 = E_2$).

Now, let's prove that Ω is surjective. For $W \in \mathcal{P}(V)$, take:

$$\left(\Gamma_1(W,E_1), \Gamma_2(V \setminus W, E_2) \right) \in \mathfrak{Q}_{k,1}^m(V).$$

We deduce that if $\ell \coloneqq |E_1|$, we have $|E_2| = m - \ell$. Also, $\left|\mathrm{LCC}(\Gamma_2(V \setminus W, E_2))\right| \lneqq k$ and $\Gamma_1(W,E_1) \in \chi_\ell(W)$.

For $\Gamma(V,E) \coloneqq \Gamma(V \setminus W \sqcup W, E_1 \sqcup E_2)$, we have indeed that $\Gamma(V,E) \in \Lambda_k^m(V,\cdot)$ and $\Omega(\Gamma(V,E)) = (\Gamma_1,\Gamma_2)$ because $\left|\mathrm{LCC}(\Gamma(V,E))\right| = |W| = k$ since W is the only connected component of size k. $\square$