Canonical Labeling of Sparse Random Graphs

Oleg Verbitsky
Institut für Informatik, Humboldt-Universität zu Berlin, Germany

Maksim Zhukovskii
School of Computer Science, University of Sheffield, UK

Abstract

We show that if $p = O(1/n)$, then the Erdős-Rényi random graph $G(n, p)$ with high probability admits a canonical labeling computable in time $O(n \log n)$. Combined with the previous results on the canonization of random graphs, this implies that $G(n, p)$ with high probability admits a polynomial-time canonical labeling whatever the edge probability function p. Our algorithm combines the standard color refinement routine with simple post-processing based on the classical linear-time tree canonization. Noteworthy, our analysis of how well color refinement performs in this setting allows us to complete the description of the automorphism group of the 2-core of $G(n, p)$.

2012 ACM Subject Classification Mathematics of computing → Random graphs; Mathematics of computing → Graph algorithms

Keywords and phrases Graph isomorphism, random graphs, canonical labeling, color refinement

Digital Object Identifier 10.4230/LIPIcs.STACS.2025.75

Related Version *Full Version*: `https://arxiv.org/abs/2409.18109` [33]

Funding *Oleg Verbitsky*: Supported by DFG grant KO 1053/8–2. On leave from the IAPMM, Lviv, Ukraine.

Acknowledgements The authors would like to thank Michael Krivelevich for helpful discussions.

1 Introduction

On an n-vertex input graph G, a *canonical labeling algorithm* computes a bijection $\lambda_G : V(G) \to \{1, \ldots, n\}$ such that if another graph G' is isomorphic to G, then the isomorphic images of G and G' under respective permutations λ_G and $\lambda_{G'}$ are equal. Given the labelings λ_G and $\lambda_{G'}$, it takes linear time to check whether G and G' are isomorphic. The existence of polynomial-time algorithms for testing isomorphism of two given graphs and, in particular, for producing a canonical labeling remain open. Babai's breakthrough quasi-polynomial algorithm for testing graph isomorphism [7] was subsequently extended to a canonical labeling algorithm of the same time complexity [8]. In the present paper, we address the canonical labeling problem for the Erdős-Rényi (or binomial) random graph $G(n, p)$. Recall that the vertex set of $G(n, p)$ is $\{1, \ldots, n\}$, and each pair of vertices is adjacent with probability $p = p(n)$, independently of the other pairs.

Babai, Erdős, and Selkow [5] proved that the simple algorithmic routine known as *color refinement* (*CR* for brevity) with high probability produces a *discrete* coloring of the vertices of $G(n, 1/2)$, that is, a coloring where the vertex colors are pairwise different. Since the vertex colors are isomorphism-invariant, this yields a canonical labeling of $G(n, 1/2)$ by numbering the color names in the lexicographic order. Here and throughout, we say that an event happens for $G(n, p)$ *with high probability* (*whp* for brevity) if the probability of this event tends to 1 as $n \to \infty$. The result of [5] has a fundamental meaning: almost all graphs admit an easily computable canonical labeling and, hence, the graph isomorphism problem has low average-case complexity.

42nd International Symposium on Theoretical Aspects of Computer Science (STACS 2025).
Editors: Olaf Beyersdorff, Michał Pilipczuk, Elaine Pimentel, and Nguyễn Kim Thắng;
Article No. 75; pp. 75:1–75:20
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The argument of [5] can be extended to show [14, Theorem 3.17] that the CR coloring of $G(n,p)$ is whp discrete for all $n^{-1/5}\ln n \ll p \le 1/2$. Note that it is enough to consider the case of $p \le 1/2$ since $G(n,1-p)$ has the same distribution as the complement of $G(n,p)$. Remarkably, the algorithm of Babai, Erdős, and Selkow performs only a bounded number of color refinement steps and, due to this, works in linear time.

A different algorithm suggested by Bollobás [12] works in polynomial time and whp produces a canonical labeling of $G(n,p)$ in a much sparser regime, namely when $\frac{(1+\delta)\ln n}{n} \le p \le 2n^{-11/12}$ for any positive constant δ. The next improvement was obtained by Czajka and Pandurangan [16] who proved that, in a bounded number of rounds, CR yields a discrete coloring of $G(n,p)$ whp when $\frac{\ln^4 n}{n \ln\ln n} \ll p \le \frac{1}{2}$. Finally, Linial and Mosheiff [27] designed a polynomial time algorithm for canonical labeling of $G(n,p)$ when $\frac{1}{n} \ll p \le \frac{1}{2}$. As shown by Gaudio, Rácz, and Sridhar [21], in the subdiapason $p \ge \frac{(1+\delta)\ln n}{n}$ for any fixed $\delta > 0$, a canonical labeling can still be provided by CR in a bounded number of rounds.

The decades-long line of research summarized above leaves open the question whether a random graph $G(n,p)$ admits efficient canonization in the regime $p = O(1/n)$. Note that the case of $p = o(1/n)$ is easy. Indeed, as long as $pn = 1 - \omega(n^{-1/3})$, whp $G(n,p)$ is a vertex-disjoint union of trees and *unicyclic* graphs (i.e., connected graphs containing exactly one cycle). Canonization of such graphs is tractable due to the classical linear-time canonical labeling algorithms for trees [1] and even planar graphs (see [6] for a survey of the early work on graph isomorphism covering these graph classes). Thus, efficient canonization remains unknown for all $p = p(n)$ such that, for some $C > 0$ and all n, $1 - Cn^{-1/3} \le pn \le C$ (even though $G(n,p)$ stays planar with a non-negligible probability as long as $pn = 1 + O(n^{-1/3})$; see [32]). Our first result closes this gap.

▶ **Theorem 1.** *If $p = O(1/n)$, then $G(n,p)$ whp admits a canonical labeling computable in time $O(n \log n)$.*

The development of canonical labeling algorithms for $G(n,p)$ is summarized in Table 1. The sources marked by * show that canonical labeling in the corresponding range can be obtained by CR in a constant number of refinement rounds. An inspection of the other algorithms reveals that all of them can be implemented as a combination of the 2-WL

Table 1 Canonical labeling algorithms for random graphs in the full-scale Erdős-Rényi evolutional model $G(n,p)$.

Edge probability	Algorithm
$p = \frac{1}{2}$	Babai, Erdős, and Selkow [5]*
$n^{-1/5}\ln n \ll p \le 1/2$	Bollobás [14]*
$\frac{(1+\delta)\ln n}{n} \le p \le 2n^{-11/12}$	Bollobás [12]
$\frac{\ln^4 n}{n\ln\ln n} \ll p \le \frac{1}{2}$	Czajka and Pandurangan [16]*
$\frac{(1+\delta)\ln n}{n} \le p \ll n^{-5/6}$	Gaudio, Rácz, and Sridhar [21]*
$\frac{1}{n} \ll p \le \frac{1}{2}$	Linial and Mosheiff [27]
$p = O(\frac{1}{n})$	this paper

(2-dimensional Weisfeiler-Leman) algorithm [34] with tree canonization. The 2-WL is an extension of CR which computes a canonical coloring of *pairs* of vertices. Thus, in these cases, canonical labeling can be obtained in time $O(n^3 \log n)$, matching the running time bound for 2-WL (see [23]). If $p = O(1/n)$, the running time is actually $O(n \log n)$ because our algorithm, as discussed below, uses CR along with simple pre- and post-processing.

A simple argument shows that Theorem 1, combined with the previous results, implies that the Erdős-Rényi random graph $G(n, p)$ whp admits an efficiently computable canonical labeling whatever the edge probability function $p(n)$.

▶ **Corollary 2.** *There exists a polynomial time algorithm that, for any function $p = p(n)$ with values in $[0, 1]$, whp produces a canonical labeling of $G(n, p)$.*

We now recall some highlights of the evolution of the random graph. Erdős and Rényi [19] proved their spectacular result that when p passes a certain threshold around $1/n$, then the size of the largest connected component in $G(n, p)$ rapidly grows from $\Theta(\log n)$ to $\Theta(n)$. A systematic study of the structure of connected components in the random graph when p is around the critical value $1/n$ was initiated in the influential paper of Bollobás [13]. For more details about the phase transition, see, e.g., [24, Chapter 5].

A connected graph is called *complex*, if it has more than one cycle. The union of all complex components of a graph G will be called the *complex part* of G, and the union of the other components will be referred to as the *simple part*. As we already mentioned, if $pn = 1 - \omega(n^{-1/3})$ then the complex part of $G(n, p)$ is whp empty. This is the so-called *subcritical phase*. In the *critical phase*, when $pn = 1 \pm O(n^{-1/3})$, the complex part of $G(n, p)$ whp has size $O_P(n^{2/3})$ and its structure is thoroughly described in [29, 30]. Here and below, for a sequence of random variables ξ_n and a sequence of reals a_n we write $\xi_n = O_P(a_n)$ if the sequence ξ_n/a_n is stochastically bounded[1]. Finally, in the *supercritical phase*, when $pn = 1 + \omega(n^{-1/3})$, whp $G(n, p)$ contains a unique complex connected component and this component has size $\Theta(n^2(p - 1/n))$. In particular, when $p = O(1/n)$ and $p > (1 + \delta)/n$ for a constant $\delta > 0$ (we refer to this case as *strictly supercritical regime*), whp this component has linear size $\Omega(n)$. It is called the *giant component* as all the other connected components have size $O(\log n)$.

In general, the simple part of a graph G is easily canonizable by the known techniques, which reduces our problem to finding a canonical labeling for the complex part of G. Furthermore, recall that the 2-core of a graph H, which we will for brevity call just *core* and denote by $\mathrm{core}(H)$, is the maximal subgraph of H that does not have vertices of degree 1. Equivalently, $\mathrm{core}(H)$ can be defined as the subgraph of H obtained by iteratively pruning all vertices in H that have degree at most 1 until there are no more such vertices. Thus, if H is the (non-empty) complex part of G, then H consists of $\mathrm{core}(H)$ and some (possibly empty) rooted trees planted at the vertices of the core. It follows that if we manage to canonically label the vertices of $\mathrm{core}(H)$, then this labeling easily extends to a canonical labeling of the entire graph G.

Suppose that CR is run on H. In the most favorable case, it would output a vertex coloring discrete on $\mathrm{core}(H)$. It turns out that, though not exactly true, this is indeed the case to a very large extent.

[1] I.e. for every $\varepsilon > 0$, there exists $C > 0$ and n_0 such that $\mathbb{P}(|\xi_n/a_n| > C) < \varepsilon$ for all $n \geq n_0$.

▶ **Theorem 3.** *Let $G_n = G(n,p)$ and assume that $p = O(1/n)$. Let H_n denote the complex part of G_n and $\mathsf{C}_n = \mathrm{core}(\mathsf{H}_n)$. When CR is run on H_n, then*

1. *CR assigns individual colors to all but $O_P(1)$ vertices in C_n;*
2. *the other color classes whp have size 2;*
3. *whp, every such color class is an orbit of the automorphism group $\mathrm{Aut}(\mathsf{H}_n)$ consisting of two vertices with degree 2 in C_n.*

▶ Remark 4. From our proofs it is easy to derive that, when $np = 1 + o(1)$, then CR distinguishes between all vertices of C_n whp.

Theorem 3 allows us to obtain an efficient canonical labeling algorithm for $G(n,p)$, as stated in Theorem 1, by combining CR with simple post-processing whose most essential part is invoking the linear-time tree canonization. Another consequence of Theorem 3 is that CR alone is powerful enough to solve the standard version of the graph isomorphism problem for the complex part of $G(n,p)$. Specifically, we say that a graph H is *identifiable* by CR if CR distinguishes H from any non-isomorphic graph H' (in the sense that CR outputs different multisets of vertex colors on inputs H and H'). It is not hard to see that H is identifiable by CR whenever the CR coloring of H is discrete. Fortunately, the properties of the CR coloring ensured by Theorem 3 are still sufficient for CR-identifiability.

▶ **Corollary 5.** *Under the assumption of Theorem 3,*

1. *H_n is whp identifiable by CR and, consequently,*
2. *whp, G_n is identifiable by CR exactly when the simple part of G_n is identifiable.*

The CR-identifiability of the simple part of a graph admits an explicit, efficiently verifiable characterization, which we give in Theorem 14. This characterization can be used to show that the random graph G_n is identifiable by CR with probability asymptotically bounded away from 0 and 1.

Our techniques for proving Theorems 1 and 3 can also be used for deriving a structural information about the automorphisms of a random graph. As proved by Erdős and Rényi [20] and by Wright [35], $G(n,p)$ for $p \leq 1/2$ is asymmetric, i.e., has no non-identity automorphism, if $np - \ln n \to \infty$ as $n \to \infty$. This result is best possible because if, $np - \ln n \leq \gamma$ for some constant $\gamma > 0$, then the random graph has at least 2 isolated vertices with non-vanishing probability. It is noteworthy that the asymmetry of $G(n,p)$ in the regime $p \geq \frac{(1+\delta)\ln n}{n}$ can be certified by the fact that CR coloring of $G(n,p)$ is discrete due to the aforementioned result of Gaudio, Rácz, and Sridhar [21]. In the diapason of p forcing $G(n,p)$ to be disconnected, the action of the automorphism group can be easily understood on the simple part and on the tree-like pieces of the complex part, and full attention should actually be given to the core of the complex part. Theorem 3 provides a pretty much precise information about the action of $\mathrm{Aut}(G_n)$ on C_n. More subtle questions arise if, instead of considering the action of $\mathrm{Aut}(G_n)$ on C_n, we want to understand the automorphisms of C_n itself. It is quite remarkable that the CR algorithm, applied to C_n rather than to H_n, can serve as a sharp instrument for tackling this problem (and, in fact, the proof of Theorem 3 is based on an analysis of the output of CR on C_n).

We begin with describing simple types of potential automorphisms of C_n (with the intention of showing that, whp, all automorphism of C_n are actually of this kind). If a vertex x has degree 2 in C_n, then it belongs to a (unique) path from a vertex s of degree at least 3 to a vertex t of degree at least 3 with all intermediate vertices having degree 2. We call such a path in C_n *pendant*. It is possible that $s = t$, and in this case we speak of a *pendant cycle*. Clearly, the reflection of a pendant cycle fixing its unique vertex of degree more than 2 is an automorphism of C_n. Furthermore, call two pendant paths *transposable* if they have the same length and share the endvertices. Note that C_n has an automorphism transposing such paths (and fixing their endvertices). Let A_1 denote the set of the automorphisms provided by pendant cycles, and let A_2 be the set of the automorphisms provided by transposable pairs of pendant paths. Moreover, C_n can have a connected component consisting of two vertices of degree 3 and three pendant paths of pairwise different lengths between these vertices. Such a component has a single non-trivial automorphism, which contributes in $\mathrm{Aut}(\mathsf{C}_n)$. The set of such automorphisms of C_n will be denoted by A_3.

Recall that an elementary abelian 2-group is a group in which all non-identity elements have order 2 or, equivalently, a group isomorphic to the group $(\mathbb{Z}_2)^k$ for some k.

▶ **Theorem 6.** *Let $G_n = G(n,p)$ and assume that $p = O(1/n)$. Let C_n be the core of the complex part of G_n.*

1. *The order of $\mathrm{Aut}(\mathsf{C}_n)$ is stochastically bounded, i.e., $|\mathrm{Aut}(\mathsf{C}_n)| = O_P(1)$.*
2. *Whp, $\mathrm{Aut}(\mathsf{C}_n)$ is an elementary abelian 2-group. Moreover, $A_1 \cup A_2 \cup A_3$ is a minimum generating set of $\mathrm{Aut}(\mathsf{C}_n)$.*[2]
3. *In addition,*
   - **(a)** *if $pn \geq 1 + \delta$ for a constant $\delta > 0$, then both A_1 and A_2 are non-empty with non-negligible probability, while $A_3 = \varnothing$ whp.*
   - **(b)** *If $pn = 1 + o(1)$ and $pn = 1 + \omega(n^{-1/3})$, then $A_1 \neq \varnothing$ with non-negligible probability, while $A_2 = A_3 = \varnothing$ whp.*
   - **(c)** *If $pn = 1 \pm O(n^{-1/3})$, then both A_1 and A_3 are non-empty with non-negligible probability, while $A_2 = \varnothing$ whp.*

This theorem makes a final step in the study of the automorphisms group of a random graph. Recall that H_n is whp empty when $np = 1 - \omega(n^{-1/3})$ and that $G(n,p)$ is connected and asymmetric when $np = \ln n + \omega(1)$. We, therefore, focus on the intermediate diapason. If $np \to \infty$ as $n \to \infty$, then the core of the giant component of $G(n,p)$ is whp still asymmetric, as proved independently by Łuczak [28] and Linial and Mosheiff [27]. Moreover, Łuczak described the automorphisms group of the core of the giant component of $G(n,p)$ when $np > \gamma$ for a large enough constant γ, and obtained an analogue of Theorem 6 for this case; see [28, Theorem 4]. Our Theorem 6 not only extends [28, Theorem 4] to the full range of $p = O(1/n)$ but also refines this result even for $np > \gamma$ by showing that $\mathrm{Aut}(\mathsf{C}_n)$ is actually an elementary 2-group. Another interesting observation is that some automorphisms of the core do not extend to automorphisms of the entire $G(n,p)$. Indeed, if $np = 1 + o(1)$, then whp $\mathrm{Aut}(G(n,p))$ acts trivially on the core; see Remark 4.

Related work. As we already mentioned, Theorem 1 combined with the previous results on canonical labeling of $G(n,p)$ for $1/n \ll p \leq 1/2$ implies the existence of a polynomial-time canonical labeling algorithm succeeding on $G(n,p)$ whp for an *arbitrary* edge probability function $p = p(n)$. In this form, this result has been independently obtained by Michael Anastos, Matthew Kwan, and Benjamin Moore [2]. Another result in their paper describes the action of $\mathrm{Aut}(G(n,p))$ on the core of $G(n,p)$, which follows also from our Theorem 3 and the results of Łuczak [28] and Linial and Mosheiff [27]. The techniques used in [2] and in our paper are completely different, though both proofs rely on color refinement. The two approaches have their own advantages. The method developed in [2] is used there also

[2] Consequently, whp C_n contains neither a triple of pairwise transposable paths, nor two isomorphic components with an automorphism in A_3, nor a cyclic component with a single chord between diametrically opposite vertices. Moreover, whp no two pendant cycles in C_n share a vertex.

to show that color refinement is helpful for canonical labeling of the random graph when $p \gg 1/n$ and to study the smoothed complexity of graph isomorphism. Our method allows obtaining precise results on the automorphism group of the core (Theorem 6).

Immerman and Lander [23] showed a tight connection between CR-identifiability and definability of a graph in first-order logic with counting quantifiers. Corollary 5 can, therefore, be recast in logical terms as follows. If $p = O(1/n)$, then H_n is whp definable in this logic with using only two first-order variables (where the definability of a graph H means the existence of a formula which is true on H and false on any graph non-isomorphic to H). Definability of the giant component of $G(n, p)$ in the standard first-order logic (without counting quantifiers) was studied by Bohman et al. [11].

The rest of the paper and proof strategy. Section 2 begins with formal description of the color refinement algorithm in Subsection 2.1 and then, in Subsection 2.2, presents a useful criterion of CR-distinguishability in terms of universal covers. The concept of a universal cover appeared in algebraic and topological graph theory [10, 15, 31], and its tight connection to CR was observed in [3]. Subsection 2.3 pays special attention to the CR-identifiability of unicyclic graphs, which in Subsection 2.4 allows us to obtain an explicit criterion of CR-identifiability for general graphs in terms of the complex and the simple part of a graph. Finally, in Subsection 2.5 we use the relationship between CR and universal covers to prove useful properties of the CR-coloring of the core of an arbitrary graph.

Theorem 1 and Corollaries 2 and 5 are proved in Section 3. The proofs of Theorem 1 and Corollary 5 are based on Theorem 3. A crucial ingredient of the proof of Theorem 3 is our Main Lemma (Lemma 20). This lemma says that CR is unable to distinguishe between two vertices in the core only if they lie either on pendant paths (with the same endvertices) transposable by an automorphism of the graph or on a pendant cycle admitting a reflection by an automorphism. Note that this statement alone, which is a part of the information provided by Theorem 3, is enough to derive Theorem 1 and Corollary 5.

The proof of Main Lemma is outlined in Section 4. It heavily relies on the notion of a kernel. The *kernel* $K(G)$ of a graph G is a multigraph obtained from core(G) by contracting all pendant paths. That is, $K(G)$ is obtained via the following iterative procedure: at every step if there exists a vertex u with only two neighbors v_1, v_2, we remove u with both incident edges and add the edge $\{v_1, v_2\}$ instead. Note that this transformation can lead to appearance of multiple edges and loops.

To prove that CR colors vertices of the core in the described manner, we use the contiguous models due to Ding, Lubetzky, and Peres [18] in strictly supercritical regime and due to Ding, Kim, Lubetzky, and Peres in critical regime [17]. They allow to transfer properties of random multigraphs with given degree sequences to the kernel of the giant component in the random graph. Another important ingredient in our proofs is the fact that in the kernel of the supercritical random graph there are whp no small complex subgraphs. We consider separately two types of vertices: first, we prove that CR colors differently all vertices such that their neighborhoods induce trees. This is done in Sections 4.1 and 4.2 for $p = 1 + \omega(n^{-1/3})$ and $p = 1 + O(n^{-1/3})$ respectively. Then, in Section 4.3, we prove that these vertices are helpful to distinguish between all the remaining vertices.

A complete proof of Theorem 3 and the proof of Theorem 6 are omitted due to the space constraints and can be found in the full version of the paper [33].

2 Color refinement: From identifiability to canonical labeling

2.1 Description of the CR algorithm

We now give a formal description of the *color refinement* algorithm (*CR* for short). CR operates on vertex-colored graphs but applies also to uncolored graphs by assuming that their vertices are colored uniformly. An input to the algorithm consists either of a single graph or a pair of graphs. Consider the former case first. For an input graph G with initial coloring C_0, CR iteratively computes new colorings

$$C_i(x) = \Big(C_{i-1}(x), \{\!\{C_{i-1}(y)\}\!\}_{y\in N(x)}\Big), \tag{1}$$

where $\{\!\{\}\!\}$ denotes a multiset and $N(x)$ is the neighborhood of a vertex x. Denote the partition of $V(G)$ into the color classes of C_i by $\mathcal{P}_i$. Note that each subsequent partition $\mathcal{P}_{i+1}$ is either finer than or equal to $\mathcal{P}_i$. If $\mathcal{P}_{i+1} = \mathcal{P}_i$, then $\mathcal{P}_j = \mathcal{P}_i$ for all $j \geq i$. Suppose that the color partition stabilizes in the t-th round, that is, t is the minimum number such that $\mathcal{P}_t = \mathcal{P}_{t-1}$. CR terminates at this point and outputs the coloring $C = C_t$. Note that if the colors are computed exactly as defined by (1), they will require exponentially long color names. To prevent this, the algorithm renames the colors after each refinement step, using the same set of no more than n color names.

If an input consists of two graphs G and H, then it is convenient to assume that their vertex sets $V(G)$ and $V(H)$ are disjoint. The vertex colorings of G and H define an initial coloring C_0 of the union $V(G) \cup V(H)$, which is iteratively refined according to (1). The color partition $\mathcal{P}_i$ is defined exactly as above but now on the whole set $V(G) \cup V(H)$. As soon as the color partition of $V(G) \cup V(H)$ stabilizes, CR terminates and outputs the current coloring $C = C_t$ of $V(G) \cup V(H)$. The color names are renamed for both graphs synchronously.

We say that CR *distinguishes* G and H if $\{\!\{C(x)\}\!\}_{x\in V(G)} \neq \{\!\{C(x)\}\!\}_{x\in V(H)}$. If CR fails to distinguish G and H, then we call these graphs *CR-equivalent* and write $G \equiv_{\mathrm{CR}} H$. A graph G is called *CR-identifiable* if $G \equiv_{\mathrm{CR}} H$ always implies $G \cong H$.

2.2 Covering maps and universal covers

A surjective homomorphism from a graph K onto a graph G is a *covering map* if its restriction to the neighborhood of each vertex in K is bijective. We suppose that G is a finite graph, while K can be an infinite graph. If there is a covering map from K to G (in other terms, K *covers* G), then K is called a *covering graph* of G. Let G be connected. We say that a graph U is a *universal cover* of a graph G if U covers every connected covering graph of G. A universal cover $U = U^G$ of G is unique up to isomorphism. Alternatively, U^G can be defined as a tree that covers G. If G is itself a tree, then $U^G \cong G$; otherwise the tree U^G is infinite.

A straightforward inductive argument shows that a covering map α preserves the coloring produced by CR, that is, $C_i(u) = C_i(\alpha(u))$ for all i, where C_i is defined by (1). It follows that, if two connected graphs G and H have a common universal cover, i.e., $U^G \cong U^H$, then $\{\,C(u) : u \in V(G)\} = \{\,C(v) : v \in V(H)\}$. The converse implication is also true, as a consequence of the following lemma.

▶ **Lemma 7** (cf. Lemmas 2.3 and 2.4 in [26]). *Let U^G and U^H be universal covers of connected graphs G and H respectively. Furthermore, let α be a covering map from U^G to G and β be a covering map from U^H to H. For a vertex x of U^G and a vertex y of U^H, let U^G_x and U^H_y be the rooted versions of U^G and U^H with roots at x and y respectively. Then $U^G_x \cong U^H_y$ (isomorphism of rooted trees) if and only if $C(\alpha(x)) = C(\beta(y))$.*

The union of CR-identifiable graphs does not need be CR-identifiable. However, the concept of a universal cover allows us to state the following criterion, which is an extension of [4, Thm. 5.4] (see [4, p. 649] for details).

▶ **Lemma 8.** *Let $G_1, \ldots, G_k$ be connected CR-identifiable graphs and G be their vertex-disjoint union. Then G is CR-identifiable if and only if, for every pair of distinct i and j such that neither G_i nor G_j is a tree, the universal covers of G_i and G_j are non-isomorphic.*

2.3 Unicyclic graphs

2.3.1 Universal covers of unicyclic graphs

For a unicyclic graph G, its core(G) is the set of vertices lying on the unique cycle of G. We use the notation $c(G) = |\operatorname{core}(G)|$ for the length of this cycle. For a vertex x in core(G), let G_x denote the subgraph of G induced by the vertices reachable from x along a path avoiding the other vertices in core(G). This is obviously a tree. Moreover, we define G_x as a rooted tree with root at x. Let $t(x)$ denote the isomorphism class of the rooted tree G_x. We treat t as a coloring of core(G) and write $R(G)$ to denote the cycle of G endowed with this coloring. Thus, $R(G)$ is defined as a vertex-colored cycle graph. It will also be useful to see $R(G)$ as a *circular word* over the alphabet $\{t(x) : x \in \operatorname{core}(G)\}$; see, e.g., [22] and the references therein for more details on this concept in combinatorics on words. In fact, $R(G)$ is associated with two circular words, depending on one of the two directions in which we go along $R(G)$. However, the choice of one of the two words is immaterial in what follows.

Speaking about a *word*, we mean a standard, non-circular word. Two words are conjugated if they are obtainable from one another by cyclic shifts. A circular word is formally defined as the conjugacy class of a word. A word u is a period of a word v if $v = u^k$ for some $k \geq 1$. A word u is a period of a circular word w if u is a period of some representative in the conjugacy class of w. Note that if u is a period of a word v, then any conjugate of u is a period of some conjugate of v. This allows us to consider periods of circular words themselves being circular words. We define the *periodicity* $p(w)$ of a circular word w to be the minimum length of a period of w. It may be useful to keep in mind that a period of length $p(w)$ is also a period of every period of w; cf. [22, Proposition 1] (note that our terminology is different from [22]).

For a unicyclic graph G, we define its periodicity by $p(G) = p(R(G))$, where $R(G)$ is seen as a circular word as explained above. Note that $p(G)$ is a divisor of $c(G)$ and that $1 \leq |\{t(x)\}_{x \in \operatorname{core}(G)}| \leq p(G) \leq c(G)$.

Like trees, unicyclic graphs are also characterizable in terms of universal covers.

▷ Claim 9. A connected graph G is unicyclic if and only if U^G has a unique infinite path.

The unique infinite path subgraph of U^G will be denoted by $P(U_G)$. The structure of U^G is clear: The cycle of G is unfolded into the infinite path $P(U^G)$. Moreover, let α be a covering map from U^G to G. Then U^G is obtained by planting a copy of the rooted tree $G_{\alpha(x)}$ at each vertex x on $P(U^G)$. The path $P(U^G)$ will be considered being a vertex colored graph, with each vertex x colored by $t(\alpha(x))$.

The following observation is quite useful in what follows. Let α be a covering map from U^G to G. The restriction of α to $P(U^G)$ is a covering map from the vertex-colored path $P(U^G)$ to the vertex-colored cycle $R(G)$. Note that a covering map must preserve vertex colors.

▶ **Lemma 10.** *Let G and H be connected unicyclic graphs. Then $U^G \cong U^H$ if and only if the circular words $R(G)$ and $R(H)$ have a common period. Moreover, if $U^G \cong U^H$, then $p(G) = p(H)$.*

Proof. In one direction the statement is clear: if $R(G)$ and $R(H)$ have a common period, then $U^G \cong U^H$ by the definition. Let $U \cong U^G \cong U^H$ be a common universal cover of G and H. We can naturally see $P(U)$ as an infinite word. An arbitrary subword of length $p(G)$ of $P(U)$ is a period of $P(U)$, and the same is true for an arbitrary subword of length $p(H)$ of $P(U)$. It follows that $P(U)$ has a period $u = u_1 \dots u_q$ of length $q = \gcd(p(G), p(H))$. Indeed, it is sufficient to note that there exist integers β_1, β_2 such that $q = \beta_1 p(G) - \beta_2 p(H)$. Thus, for any u_0, u_q at distance q in $P(U)$, we get that $u_0 = u_{\beta_1 p(G)} = u_{q+\beta_2 p(H)} = u_q$.

If α and β are covering maps from U to G and H respectively, then $\alpha(u_1) \dots \alpha(u_q)$ is a period of $R(G)$ and $\beta(u_1) \dots \beta(u_q)$ is a period of $R(H)$. Since α and β preserve the vertex colors, we have the equality $\alpha(u_1) \dots \alpha(u_q) = \beta(u_1) \dots \beta(u_q)$. This also implies that, in fact, $q = p(G) = p(H)$. ◀

2.3.2 CR-identifiability of unicyclic graphs

▷ Claim 11. Let G be a connected unicyclic graph. Suppose that $G \equiv_{\mathrm{CR}} H$ and H consists of connected components $H_1, \dots, H_m$. Then

1. $U^{H_i} \cong U^G$ for all i,
2. every H_i is unicyclic, and
3. $c(G) = c(H_1) + \dots + c(H_m)$.

Proof. 1. Fix $i \in [m]$. Let U^G and $W = U^{H_i}$ and α^U, α^W be the respective covering maps. Let y be a vertex in U^{H_i}. Since G and H are CR-equivalent, U^G must contain a vertex x such that $C(\alpha^U(x)) = C(\alpha^W(y))$. By Lemma 7, $U_x \cong W_y$. It follows that $U \cong W$.

2. Immediately by Part 1 and Claim 9.

3. Fix a period of $R(G)$ and set $T = \sum_x |V(G_x)|$ where the summation goes over all x in this period (this definition obviously does not depend on the choice of the period). Let $p = p(G)$. Note that $|V(G)| = \frac{c(G)}{p} T$. By Part 1 and Lemma 10, we similarly have $|V(H_i)| = \frac{c(H_i)}{p} T$. The required equality now follows from the trivial equality $|V(G)| = |V(H_1)| + \dots + |V(H_m)|$. ◁

▶ **Lemma 12.** *A connected unicyclic graph G is CR-identifiable if and only if one of the following conditions is true:*

- $p(G) = 1$ *and* $c(G) \in \{3, 4, 5\}$,
- $p(G) = 2$ *and* $c(G) \in \{4, 6\}$,
- $p(G) = c(G)$.

Proof. ($\Leftarrow$) Suppose that a connected unicyclic graph G satisfies one of the three conditions and show that it is CR-identifiable. Assuming that H is CR-equivalent to G, we have to check that G and H are actually isomorphic.

Assume first that H is connected. If H is not unicyclic, then G and H have equal number of vertices but different number of edges. This implies that G and H have different degree sequences, contradicting the assumptions that $G \equiv_{\mathrm{CR}} H$. Therefore, H must be unicyclic. By Part 3 of Claim 11, we have $c(G) = c(H)$. Along with Lemma 10, which is applicable because $U^G \cong U^H$ whenever $G \equiv_{\mathrm{CR}} H$, this implies that $R(G) \cong R(H)$. The last relation, in its turn, implies that $G \cong H$.

Assume now that H is disconnected. Let $H_1, \ldots, H_m$ be the connected components of H. Combining Claim 11 and Lemma 10, we see that $p(G) = p(H_1) \leq c(H_1) < c(G)$. It follows that G satisfies one of the first two conditions. The restrictions on $c(G)$, however, rule out the equality in Part 3 of Claim 11. In particular, if $c(G) = 6$, then the only possible case is $m = 2$ and $c(H_1) = c(H_2) = 3$. However, it contradicts the equality $p(H_1) = p(H_2) = p(G) = 2$. Thus, the case of disconnected H is actually impossible, that is, all such H are distinguishable from G by CR.

($\Rightarrow$) Suppose that all three conditions are false. That is, either $p(G) = 1$ and $c(G) \geq 6$, or $p(G) = 2$ and $c(G) \geq 8$ (note that $c(G)$ is even in this case), or $3 \leq p(G) < c(G)$ (in the last case, $p(G)$ is a proper divisor of $c(G)$). In each case, $R(G)$ is CR-equivalent to a disjoint union of two shorter vertex-colored cycles R_1 and R_2, both sharing the same period of length $p(G)$ with $R(G)$. Taking the connected unicyclic graphs H_1 and H_2 such that $R(H_1) \cong R_1$ and $R(H_2) \cong R_2$, we see that G is CR-equivalent to the disjoint union of H_1 and H_2 and is, therefore, not CR-identifiable. ◀

▶ **Lemma 13.** *Let G and H be connected unicyclic graphs with $c(H) \leq c(G)$. Assume that both G and H are CR-identifiable. Then $U^G \cong U^H$ if and only if $\{t(x) : x \in \text{core}(G)\} = \{t(x) : x \in \text{core}(H)\}$ and one of the following conditions is true:*

- *$G \cong H$,*
- *$p(G) = p(H) = 1$ and $3 \leq c(H) < c(G) \leq 5$,*
- *$p(G) = p(H) = 2$ and $c(H) = 4$ while $c(G) = 6$.*

Proof. ($\Leftarrow$) By Lemma 10.

($\Rightarrow$) Let G and H be CR-identifiable connected unicyclic graphs with $c(H) \leq c(G)$. Assume that $U^G \cong U^H$. The equality $\{t(x) : x \in \text{core}(G)\} = \{t(x) : x \in \text{core}(H)\}$ immediately follows from Lemma 10. By the same corollary, $p(G) = p(H) = p$. If $p \geq 3$, then Lemma 12 yields the equality $c(G) = p(G) = p(H) = c(H)$, which readily implies $G \cong H$ by using Lemma 10 once again. If $p \leq 2$, then either $c(G) = c(H)$ and $G \cong H$ or $c(H) < c(G)$ and then, by Lemma 12, $c(H)$ and $c(G)$ are as claimed. ◀

2.4 A general criterion of CR-identifiability

Deciding whether a given graph is CR-identifiable is an efficiently solvable problem [4, 25]. For our purposes, it is beneficial to have a more explicit description of CR-identifiable graphs in terms of the complex and the simple part of a graph. We now derive such a description from the facts obtained for unicyclic graphs in the preceding subsection.

▶ **Theorem 14.**

1. *A graph G is CR-identifiable if and only if both the complex and the simple parts of G are CR-identifiable.*
2. *The simple part of G is CR-identifiable if and only if both of the following two conditions are true:*
   (a) *every unicyclic component of G is CR-identifiable, i.e., is as described in Lemma 12;*
   (b) *every two unicyclic components of G have non-isomorphic universal covers, i.e., there is no pair of connected components as described in Lemma 13.*

Proof.

1. If G is CR-identifiable, then its complex and simple parts are both CR-identifiable as a consequence a more general fact: The vertex-disjoint union of any set of connected components of G is CR-identifiable. This fact is easy to see directly, and it also immediately follows from Lemma 8.

In the other direction, assuming that the complex and the simple parts of G are CR-identifiable, we have to conclude that G is CR-identifiable. Lemma 8 reduces our task to verification that if H is a complex connected component of G and S is a simple connected component of G (a tree or a unicyclic graph), then the universal covers of H and S are non-isomorphic. The last condition follows from the fact that the universal cover of a tree is the tree itself and from Claim 9.

2. The second part of the theorem follows immediately from Lemma 8 due to the well-known fact [23] that every tree is CR-identifiable. ◀

2.5 Coloring the cores of general graphs

We conclude this section by collecting useful general facts about the CR-colors of vertices in the core of a graph. Let G be an arbitrary graph. If x is a vertex in $\mathrm{core}(G)$, then in G we have a tree growing from the root x that shares with $\mathrm{core}(G)$ only the vertex x. We denote this rooted tree by T_x.

▷ Claim 15. Let G and H be graphs. Let x be a vertex in $\mathrm{core}(G)$ and y be a vertex in $\mathrm{core}(H)$. If $T_x \not\cong T_y$, then $C(x) \neq C(y)$.

Proof. Clearly, it suffices to prove this for connected G and H. The condition $T_x \not\cong T_y$ readily implies that $U_x^G \not\cong U_y^H$, and the claim follows from Lemma 7. ◁

▷ Claim 16. Let G and H be two graphs (it is not excluded that $G = H$). For vertices $u \in V(G)$ and $v \in V(H)$ assume that $C(u) = C(v)$. Then $u \in \mathrm{core}(G)$ if and only if $v \in \mathrm{core}(H)$.

Proof. Assume that G and H are connected (the general case will easily follow). Let α_G be a covering map from U^G to G, and α_H be a covering map from U^H to H. Consider $x \in V(U^G)$ and $y \in V(U^H)$ such that $\alpha_G(x) = u$ and $\alpha_H(y) = v$. Note that $u \in \mathrm{core}(G)$ if and only if there is a cycle in U^G containing x, and the same is true about v any y. This proves the claim because $U_x^G \cong U_y^H$ by Lemma 7. ◁

In our proofs, we will deal with cores that locally have a tree structure, that is, the balls of sufficiently large radii around most of its vertices induce trees. In this case, CR distinguishes vertices that have non-isomorphic neighborhoods.

▷ Claim 17. Let $B_r(v)$ denote the set of vertices at distance at most r from a vertex v. Let $v_1, v_2 \in V(G)$. If, for some r, the r-neighborhoods $B_r(v_1)$ and $B_r(v_1)$ induce non-isomorphic trees rooted in v_1 and v_2 respectively, then $C(v_1) \neq C(v_2)$.

Proof. This is a direct consequence of Lemma 7. ◁

Throughout the paper, we identify the vertex set of the kernel of G with the set of vertices of $\mathrm{core}(G)$ having degrees at least 3 in the core. We now state another consequence of Lemma 7.

▷ Claim 18. Let G be a graph with minimum degree at least 2 and let K be its kernel. Let r be a positive integer. For $v \in V(\mathrm{K})$, let $\mathcal{B}_r^{\mathrm{K}}(v)$ be the subgraph of K induced by the set of vertices at distance at most r from v in K. Let $\mathcal{B}_r(v) \subset G$ be the subdivided version of $\mathcal{B}_r^{\mathrm{K}}(v)$. Let v_1, v_2 be vertices of K such that, for some r, graphs $\mathcal{B}_r(v_1), \mathcal{B}_r(v_2) \subset G$ are non-isomorphic trees rooted in v_1, v_2. Then $C^G(v_1) \neq C^G(v_2)$.

Finally, we need the fact that the partition produced by CR on a graph refines the partition produced by CR on its core.

▷ Claim 19. Let u and v be vertices in core(G). Let C and C' be the colorings produced by CR run on G and core(G) respectively. If $C'(u) \neq C'(v)$, then also $C(u) \neq C(v)$.

Proof. Clearly, it is enough to prove the claim for a connected graph G. Let us assume towards a contradiction that $C(u) = C(v)$. Let α be a covering map from U^G to G. Let $x, y \in V(U^G)$ be such that $\alpha(x) = u$ and $\alpha(y) = v$. Due to Lemma 7, $U^G_x \cong U^G_y$. Therefore, $U_x^{\mathrm{core}(G)} = \mathrm{core}(U^G_x) \cong \mathrm{core}(U^H_y) = U_y^{\mathrm{core}(H)}$. But then, again by Lemma 7, $C'(u) = C'(v)$, a contradiction. ◁

3 Proofs of main results

3.1 Derivation of Corollary 5 from Theorem 3

Part 1. Any isomorphism of graphs obviously respects their cores; cf. Claim 16. Note that the CR-color of any vertex x in the core C_n contains a complete information about the isomorphism type of the rooted tree T_x "growing" from this vertex (cf. Claim 15). This has the following consequence. Let C'_n denote the colored version of C_n where each vertex x is colored by the isomorphism type of T_x. Then H_n is CR-identifiable if and only if C'_n is CR-identifiable. In order to show that C'_n is CR-identifiable it suffices to show that C'_n is reconstructible up to isomorphism from the multiset of the vertex colors produced by CR on input C'_n. The CR-color partition of C'_n is equal to the restriction of the CR-color partition of H_n to C_n (recall Claim 16). Theorem 3, therefore, provides us with the following information (whp):[3]

(a) every CR-color class of C'_n has size either 1 or 2,
(b) every two equally colored vertices have degree 2,
(c) every two equally colored vertices are transposable by an automorphism of C'_n.

Moreover, our Main Lemma (Lemma 20) ensures that H_n whp has no involutory automorphism of type A_3 described in Section 1. Along with this fact, the above conditions readily imply that the color classes of size 2 occur either "along" a pair of transposable pendant paths between two vertices of degree at least 3 or correspond to the reflectional symmetry of a pendant cycle. Here we use the notions introduced in Section 1 in the context of Aut(C_n), which should now be refined by taking into account the coloring of C'_n.

If $\{u\}$ and $\{v\}$ are two color classes of size 1, then the colors $C(u)$ and $C(v)$ yield the information on whether the vertices u and v are adjacent or not. For color classes $\{u\}$ and $\{v, v'\}$, note that u and v are adjacent if and only if u and v' are adjacent. This adjacency pattern is as well reconstructible from the colors $C(u)$ and $C(v) = C(v')$. If $\{u, u'\}$ and $\{v, v'\}$ are two color classes of size 2, then they span either a complete or empty bipartite graph or a matching (for example, u is adjacent to v, u' is adjacent to v', and there is no other edges between these color classes). Each of these three possible adjacency patterns is reconstructible from the colors $C(u) = C(u')$ and $C(v) = C(v')$. A crucial observation, completing the proof, is that all ways to put a matching between $\{u, u'\}$ and $\{v, v'\}$ lead to isomorphic graphs.

Part 2. This follows from part 1 by part 1 of Theorem 14.

[3] Note that Conditions (b) and (c) are provided by Main Lemma. Condition (a) is not essential for the argument in Subsections 3.1 and 3.2, which easily extends to the case of more than two mutually transposable pendant paths.

3.2 Derivation of Theorem 1 from Theorem 3

Before proceeding to the proof, we remark that when we say that a canonical labeling algorithm succeeds on a random graph G_n, we mean that the algorithm works correctly on a certain efficiently recognizable (closed under isomorphisms) class of graphs $\mathcal{C}$ such that G_n belongs to $\mathcal{C}$ whp. Though not explicitly stated in the argument below, it will be clear that, in our case, $\mathcal{C}$ is the class of all graphs satisfying the conditions of Theorem 3. Note that these conditions are easy to check after running CR on a graph.

First of all, we distinguish the complex and the simple parts of G_n and compute a canonical labeling of the simple part separately. This is doable in linear time. It remains to handle the complex part H_n.

It is enough to compute a suitable injective coloring of H_n and subsequently to rename the colors in their lexicographic order by using the labels that were not used for the simple part. To this end, we run CR on H_n. This takes time $O(n \log n)$ as CR can be implemented [9] in time $O((n+m)\log n)$, where m denotes the number of edges (which is linear for the sparse random graph under consideration). Then we begin with coloring the vertices of the core C_n. Theorem 3 along with Claim 16 ensures that the vertices of degree at least 3 already received individual colors. The duplex colors occur along transposable pendant paths and pendant cycle (like in Section 3.1, these notions are understood with respect to H_n rather than to C_n alone). To make such vertex colors unique, we keep the original colors along one of two transposable paths and concatenate their counterparts in the other path with a special symbol. We proceed similarly with symmetric pendant cycles. In this way, every vertex x in the core C_n receives an individual color $\ell(x)$. In the last phase, we compute a canonical labeling for each tree part T_x of H_n, regarding T_x as a tree rooted at x. This coloring is not injective yet because some T_x and T_y can be isomorphic. This is rectified by concatenating all vertex colors in T_x with $\ell(x)$.

3.3 Proof of Corollary 2

As well known, if $1/n \ll p \leq 1/2$ then the core of the giant component of $G(n,p)$ coincides with the core of the entire graph. Due to the classical linear-time algorithms for canonical labeling of trees, this observation reduces canonical labeling of $G(n,p)$ with $1/n \ll p \leq 1/2$ to canonical labeling of its core.

Linial and Mosheiff [27] suggested an algorithm A_1 that, for any p with $\frac{1}{n} \ll p(n) < n^{-2/3}$, whp labels canonically $G(n,p)$ in time $O(n^4)$ by distinguishing between all vertices of the core. In [16], it was proved that, if $\frac{\ln^4 n}{n} \leq p \leq \frac{1}{2}$, then CR whp distinguishes between all vertices of the entire $G(n,p)$. Finally, our Theorem 1 provides an algorithm A_2 that, for any $p = O(1/n)$, whp labels canonically $G(n,p)$ in time $O(n \ln n)$. Now, consider the following algorithm A:

1. Run CR. If it colors differently all vertices, then halt and output the canonical labeling produced by CR.
2. If the algorithm does not halt in Step 1, then run A_1. If it succeeds (i.e., colors differently all vertices in the core of the input graph), then halt and output the labeling produced by A_1.
3. If the algorithm does not halt in Steps 1 and 2, then run A_2 and output the labeling it produces (or give up if A_2 fails).

Let us show that the algorithm A succeeds whp for any p with $p(n) \leq 1/2$. Assume, to the contrary, that there exist a constant $\varepsilon > 0$ and a sequence $(n_k)_{k\in\mathbb{N}}$ such that

$$\mathbb{P}(\mathsf{A} \text{ fails on } G(n_k, p(n_k))) > \varepsilon$$

for all k. If there is a subsequence $(n_{k_i})_{i\in\mathbb{N}}$ and a constant $C>0$ such that $p(n_{k_i})<C/n_{k_i}$ for all i, then we get a contradiction with the performance of the algorithm A_2. Therefore, $p(n_k)\gg\frac{1}{n_k}$. If there is a subsequence $(n_{k_i})_{i\in\mathbb{N}}$ such that $p(n_{k_i})<n_{k_i}^{-2/3}$ for all i, then we get a contradiction with the performance of the algorithm A_1. It follows that $p(n_k)\geq n_k^{-2/3}$ for all k. This, however, contradicts the result of [16] that CR in this regime produces a discrete coloring of $G(n,p)$ whp.

In order to obtain canonical labeling, whp, for all p with $p(n)\in[0,1]$, we run the algorithm A on input G and if it fails, then we run A once again on the complement of G.

4 CR-coloring of the random graph

In this section, we state and prove our Main Lemma that describes the output of CR on the random graph. Given a graph G, we call vertices u and v in $\mathrm{core}(G)$ *interchangeable*, if

- they both have degree 2 in $\mathrm{core}(G)$,
- u and v belong to a cycle $F\subset\mathrm{core}(G)$ with the following property: there exists a vertex w on the cycle such that w has degree at least 3 in $\mathrm{core}(G)$, $d_F(u,w)=d_F(v,w)$, and all the other vertices on the cycle, but the vertex opposite to w when $|V(F)|$ is even, have degree 2 in $\mathrm{core}(G)$. In other words, u and v either belong to a pendant cycle or to two transposable pendant paths, and the respective transposition replaces u and v.

▶ **Lemma 20** (Main Lemma). *Let $\gamma>1$ be a constant, $pn\leq\gamma$, and $G_n=G(n,p)$. Let H_n be the union of complex components in G_n, and C_n be its core. If CR is run on H_n, then whp any pair of vertices in C_n receiving the same color is interchangeable. Under the condition $pn=1+\omega(n^{-1/3})$, this is true also if CR is run on C_n.*

The proof of Main Lemma is given in Sections 4.1–4.3. We consider separately large p (supercritical phase) and small p (critical phase). In both cases, we specify good sets of vertices and show that all vertices from good sets are distinguished by CR. This is done in Section 4.1 for large p and Section 4.2 for small p. Finally, in Section 4.3 we complete the proof: we show that distinguishing between good vertices in the core is sufficient to distinguish between all pairs in the core that are not interchangeable.

For a graph G, $d_G(u,v)$ is the shortest-path distance between u and v in G. Sometimes, when the graph is clear from the context, we omit the subscript G. For a vertex v and a real number r, we denote by $\mathcal{B}_r^G(v)$ the ball of radius r around v in G, i.e., the graph induced on the set of all vertices at distance at most r from v in G. For a non-negative integer r, we denote by $\mathcal{S}_r^G(v)\subset\mathcal{B}_r^G(v)$ the sphere of radius r around v in G, i.e., the graph induced on the set of all vertices at distance exactly r from v in G.

For a connected graph G, its *excess* is the difference between the number of edges and the number of vertices. In particular, a tree has excess -1. We call *ℓ-complex* a connected graph with excess ℓ. The *total excess* of a graph without unicyclic components is the sum of excesses of all its components.

4.1 Distinguishing good vertices in the core in the supercritical and strictly supercritical phases

In this subsection, we let $p=p(n)$ be such that $\gamma\geq np=1+\omega(n^{-1/3})$ for some constant $\gamma>1$. Denote $\delta_n:=np-1$. We denote the kernel and the core of the giant component of $G_n\sim G(n,p)$ by K_n and C_n. Let C^{core} be the coloring produced by CR on C_n.

We assign to every edge e of C_n the weight $1/\ell$, where $\ell-1$ is the number of vertices that subdivide the edge of the kernel e belongs to. The weight of a path is the sum of weights of its edges. For $u,v\in V(\mathrm{C}_n)$, let $d^f(u,v)$ be the *fractional distance* between u and v, i.e. the minimum weight of a path between u and v. We denote the respective metric space by $\mathcal{M}_n$.

Fix a positive real s. Let D_s be the set of all $v\in V(\mathrm{C}_n)$ such that the ball around v in $\mathcal{M}_n$ of radius s induces an acyclic graph. For every vertex $v\in D_s$ and integer $r<s$, let $\mathcal{P}_r(v)$ be the multiset of lengths of edge-disjoint paths from C_n that are produced by subdividing edges $\{x,y\}\in E(\mathrm{K}_n)$, where $d^f(x,v)\le r$ while $d^f(y,v)>r$. We will need the following facts.

▷ Claim 21. Let $s\ge 0.6(\ln(\delta_n^3 n))^{2/3}$. Whp for any two different $u,v\in D_s\cap V(\mathrm{K}_n)$, there exists an integer $r\le 0.5(\ln(\delta_n^3 n))^{2/3}$ such that the multisets $\mathcal{P}_r(u)$ and $\mathcal{P}_r(v)$ are different.

Proof. We fix $s\ge 0.6(\ln(\delta_n^3 n))^{2/3}$ and let $D:=D_s$. We prove this claim in the contiguous models $\tilde{G}_n$, defined in [17, Thm. 2] and [18, Thm. 1] and then use these theorems to conclude that it also holds in G_n. So, in what follows, $\tilde{\mathrm{K}}_n=\mathrm{K}(\tilde{G}_n)$, $\tilde{\mathrm{C}}_n=\mathrm{C}(\tilde{G}_n)$, and $\tilde{D}=D(\tilde{K}_n)$.

Let us expose $\tilde{\mathrm{K}}_n$ and let $u,v\in\tilde{D}\cap V(\tilde{\mathrm{K}}_n)$. As proved in [17, 18], whp $N=|V(\tilde{K}_n)|=\Theta(\delta_n^3 n)$. Assume first that the distance between u and v is at most $0.4(\ln(\delta_n^3 n))^{2/3}$ in $\tilde{\mathrm{K}}_n$. Let P be the shortest path between u and v – it is unique due to the definition of $\tilde{D}$. Let v' be a neighbor of v in $\tilde{\mathrm{K}}_n$ that does not belong to P. Then, by the definition of $\tilde{D}$, $\left|\mathcal{S}_r^{\tilde{\mathrm{K}}_n}(v')\setminus\mathcal{B}_r^{\tilde{\mathrm{K}}_n}(v)\right|\ge 2^r$ for all $r\in\left[0.4(\ln(\delta_n^3 n))^{2/3},0.5(\ln(\delta_n^3 n))^{2/3}\right]$. Since $u,v\in\tilde{D}$, we have that $\mathcal{B}_s^{\tilde{\mathrm{K}}_n}(u)$ and $\mathcal{B}_s^{\tilde{\mathrm{K}}_n}(v)$ are trees. It immediately implies, that for every such r, $\left|\mathcal{S}_{r+1}^{\tilde{\mathrm{K}}_n}(v)\setminus\mathcal{B}_{r+1}^{\tilde{\mathrm{K}}_n}(u)\right|\ge 2^r$.

We then generate subdivisions of the edges of the kernel from the definition of $\tilde{G}_n$ in the following order: for every $r=\left\lceil 0.4(\ln(\delta_n^3 n))^{2/3}\right\rceil,\ldots,\left\lfloor 0.5(\ln(\delta_n^3 n))^{2/3}\right\rfloor$, we, first, subdivide all edges growing from $\mathcal{B}_{r+1}^{\tilde{\mathrm{K}}_n}(u)$ outside of the ball, and then all edges growing from $\mathcal{S}_{r+1}^{\tilde{\mathrm{K}}_n}(v)$ outside of $\mathcal{B}_{r+1}^{\tilde{\mathrm{K}}_n}(v)$. Notice that all sets $\mathcal{S}_{r+1}^{\tilde{\mathrm{K}}_n}(v)$ are disjoint for different r. For every r, as soon as the edges that correspond to the vertex u are subdivided, the event that $\mathcal{P}_{r+1}(u)=\mathcal{P}_{r+1}(v)$ immediately implies that the *random* multiset of lengths of paths from $\tilde{\mathrm{C}}_n$, that are produced by subdividing edges from $\tilde{\mathrm{K}}_n$ that grow from $\mathcal{B}_{r+1}^{\tilde{\mathrm{K}}_n}(v)$ outside, should be equal to a predefined value. This multiset has size at least 2^r. Since the geometric random variables considered in [17, 18] do not have atoms with probability measure $1-o(1)$, the latter event has probability at most $2^{-\Theta(r)}$ due to the de Moivre–Laplace local limit theorem. Eventually,

$$\mathbb{P}\left(\mathcal{P}_{r+1}(u)=\mathcal{P}_{r+1}(v)\text{ for all }r\in\left[0.4(\ln(\delta_n^3 n))^{2/3},0.5(\ln(\delta_n^3 n))^{2/3}\right]\right)\le$$
$$\le\exp\left(-\Theta((\log(\delta_n^3 n))^{4/3})\right).$$

Assume now that the distance between u and v is bigger than $0.4(\ln(\delta_n^3 n))^{2/3}$ in $\tilde{\mathrm{K}}_n$. Then, by the definition of $\tilde{D}$, sets $\mathcal{B}_{0.2(\ln(\delta_n^3 n))^{2/3}}^{\tilde{\mathrm{K}}_n}(v)$ and $\mathcal{B}_{0.2(\ln(\delta_n^3 n))^{2/3}}^{\tilde{\mathrm{K}}_n}$ are disjoint and sets $\mathcal{S}_r^{\tilde{\mathrm{K}}_n}(v)$ have size at least 2^r for all $r\in[0.15(\ln(\delta_n^3 n))^{2/3},0.2(\ln(\delta_n^3 n))^{2/3}-1]$. As above, we get that $\mathcal{P}_r(u)=\mathcal{P}_r(v)$ for all $r\in\left[0.15(\ln(\delta_n^3 n))^{2/3},0.2(\ln(\delta_n^3 n))^{2/3}-1\right]$ with probability at most $\exp\left(-\Theta((\log(\delta_n^3 n))^{4/3})\right)$.

The union bound over all pairs $u,v\in\tilde{D}$ along with [17, Thm. 2] and [18, Thm. 1] completes the proof. ◁

▷ Claim 22. Let $s^*:=\lfloor(\ln(\delta_n^3 n))^{2/3}\rfloor$ and $D=D_{s^*}$. Whp, $C^{\mathrm{core}}(u)\ne C^{\mathrm{core}}(v)$ for any distinct $u,v\in D$.

Proof. Assume that the assertion of Claim 21 holds for $s = s^*$ and $s = s^* - 1$ deterministically. Let $u, v \in D \cap V(\mathrm{K}_n)$. Let B_u and B_v be the subdivided versions of $\mathcal{B}^{\mathrm{K}_n}_{s^*}(u)$ and $\mathcal{B}^{\mathrm{K}_n}_{s^*}(v)$ in C_n. Since $B_u \not\cong B_v$ due to the conclusion of Claim 21, we get that $C^{\mathrm{core}}(u) \neq C^{\mathrm{core}}(v)$ due to Claim 18. It remains to consider the case $v \in D \setminus V(\mathrm{K}_n)$ and $v \neq u \in D$. Assume towards contradiction that $C^{\mathrm{core}}(u) = C^{\mathrm{core}}(v)$. Then, both u and v have degree 2 in C_n. In particular, $u \notin V(\mathrm{K}_n)$. Consider the edges e_u, e_v of K_n that u and v subdivide. Let P_u, P_v be the subdivided versions of e_u, e_v. Due to the assertion of Claim 21 applied to $s = s^* - 1$, we get that all vertices of K_n from $e_u \cup e_v$ have different colors. On the other hand, by the definition of CR, the neighbors of u should have exactly the same color as the neighbors of v. Thus, by induction, we get that the entire paths P_u, P_v are colored identically. It may only happen if the endpoints of P_u coincide with the endpoints of P_v. By the definition of D, it means that $P_u = P_v =: P$. Since the endpoints of P are colored in different colors, it can be easily shown by induction that all vertices in P are also colored in different colors. Thus, $u = v$, yielding a contradiction. ◁

4.2 Distinguishing good vertices in the core in the critical regime

Let A be a large positive number. Let $1 - n^{-1/3} \ln n \leq pn = 1 + o(1)$. Whp any complex component in $G_n \sim G(n, p)$ has size at least $100A \ln n$ due to the following well-known fact.

▷ Claim 23. Let $\gamma > 1$, $np \leq \gamma$, and $G_n \sim G(n, p)$. There exists $\varepsilon = \varepsilon(\gamma)$ such that $\varepsilon \to \infty$ as $\gamma \to 1$ and whp any connected subgraph of G_n of size at most $\varepsilon \ln n$ is not complex.

Let us say that a path $u_1 \dots u_k$ *extends* the path $v_1 \dots v_k$ if, for some $i \in \{2, \dots, k\}$ the sets $\{v_1, \dots, v_{i-1}\}$ and $\{u_{k-i+2}, \dots, u_k\}$ are disjoint and $u_1 = v_i, \dots, u_{k-i+1} = v_k$. For convenience, we assume that this notion is closed under rotations of paths, i.e. if $u_1 \dots u_k$ extends $v_1 \dots v_k$, then it also extends $v_k \dots v_1$ and we also say that $u_k \dots u_1$ extends both $v_1 \dots v_k$ and $v_k \dots v_1$ in this case. We call two paths $v_1 \dots v_k$ and $u_1 \dots u_k$ *weakly disjoint*, if they are either vertex-disjoint or one paths extends the other one.

▷ Claim 24. Whp in G_n there are no two weakly disjoint paths $v_1 \dots v_k$ and $u_1 \dots u_k$ of length $k = \lfloor A \ln n \rfloor$ such that, for every $i \in \{2, \dots, k-1\}$, v_i has degree 2 if and only if u_i has degree 2.

Proof. Due to Claim 23, whp in G_n there are no complex subgraphs with at most $2k$ vertices.

For a path $P = v_1 \dots v_k$ in G_n, let us consider a binary word $w(P) = (w_2, \dots, w_{k-1})$ defined as follows: $w_i = 1$ if and only if v_i has degree 2 in G_n. Notice that, if a path $u_1 \dots u_k$ extends the path $v_1 \dots v_k$ so that $u_1 = v_i, \dots, u_{k-i+1} = v_k$ and $w(v_1 \dots v_k) = w(u_1 \dots u_k)$, then $w(u_1 \dots u_k)$ is periodic and defined by $w(v_1 \dots v_{i+1}) = (w_2, \dots, w_i)$.

Let X be the number of pairs of paths as in the statement of the claim and such that there are at most 2 edges between the paths (we are allowed to assume this since there are no complex subgraphs of size at most $2k$). Fix two weakly disjoint path $\mathbf{v} = v_1 \dots v_k$ and $\mathbf{u} = u_1 \dots u_k$ and assume without loss of generality that either $\mathbf{u}$ extends $\mathbf{v}$, or they are disjoint. Let i be such that $u_1 = v_i$. If there is no such i, i.e. the paths are disjoint, set $i = k + 1$. Then, expose edges from all v_j, $j \leq i$, and assume that they send at most 2 edges to $u_2, \dots, u_{k-1}$, other than the edge $\{u_1, u_2\}$. Then, probability that for every $j \in \{2, \dots, k-1\}$, v_j has degree 2 if and only if u_j has degree 2, is at most

$$\max\left\{(1-p)^{n-2k}, (1-(1-p)^n)\right\}^{k-4} \leq \left(1 - e^{-(1+o(1))}\right)^{k-4} = o\left(\left(\frac{2}{3}\right)^k\right).$$

We then get

$$\mathbb{E}X \le \sum_{i=2}^{k+1} n^{k+i-1}p^{k+i-3}\left(\frac{2}{3}\right)^k \le kn^2(1+o(1))^{2k}\left(\frac{2}{3}\right)^k = o(1),$$

for an appropriate choice of A. Due to Markov's inequality, $\mathbb{P}(X \ge 1) \le \mathbb{E}X = o(1)$, completing the proof. ◁

Let D be the set of all vertices v in $\mathrm{C}_n = \mathrm{core}(G_n)$ that belong to a complex component of G_n and such that $B_v := \mathcal{B}^{G_n}_{3A\ln n}(v)$ is a tree. Let C be the coloring produced by CR on G_n.

▷ Claim 25. Whp, $C(u) \neq C(v)$ for any $u, v \in D$.

Proof. Assume that the statement of Claim 24 holds deterministically in G_n. Fix two different $u, v \in D$. Let us show towards contradiction that trees B_v and B_u are not isomorphic.

Take an arbitrary path $v_1 \dots v_k$ of length $k := \lfloor 1.9A\ln n \rfloor$, where $v_1 = v$. Since, by assumption, $B_v \cong B_u$, there exists a path $\mathbf{u} = u_1 \dots u_k$ such that $u_1 = u$ and, for every $i \in \{2, \dots, k-1\}$, v_i has degree 2 if and only if u_i has degree 2. In the same way, since $v \in \mathrm{core}(G_n)$ and B_v is a tree, we may consider a path $v'_1 \dots v'_k$ that shares only the vertex $v'_1 = v = v_1$ with $v_1 \dots v_k$. Since $B_u \cong B_v$, there should be a path $\mathbf{u}' = u'_1 \dots u'_k$ that shares with $u_1 \dots u_k$ the only vertex $u'_1 = u = u_1$ and such that, for every $i \in \{2, \dots, k-1\}$, v'_i has degree 2 if and only if u'_i has degree 2. Since B_v is acyclic and since pairs of paths $v_1 \dots v_k$, $u_1 \dots u_k$ and $v'_1 \dots v'_k$, $u'_1 \dots u'_k$ cannot be disjoint due to Claim 24, u must lie on the path $P := v_k \dots v_1 \dots v'_k$. Moreover, since B_u is acyclic, once $\mathbf{u}$ or $\mathbf{u}'$ leave P, they never meet with P again. Thus, the path $u_k \dots u_1 \dots u'_k$ is divided by P in at most 3 parts: the first part does not have common vertices with P, the second part is a subpath of P, and the third part does not have common vertices with P again. Let Q be the longest part of the three. Then Q has length $\ell \ge \frac{1}{3}(2k-1) > A\ln n$. Moreover, since $\deg_{G_n} u = \deg_{G_n} v$ by assumption and $u \neq v$, there should be a subpath $P' \subset P$ such that P' and Q are weakly disjoint, and the degrees of internal vertices in P' and Q are aligned in the sense that the i-th inner vertex of P' have degree 2 if and only if the i-th vertex of Q has degree 2. This is impossible due to Claim 24. Thus, $B_v \ncong B_u$ implying $C(u) \neq C(v)$ due to Claim 17. ◁

4.3 Completing the proof of Main Lemma (Lemma 20)

Due to Claims 22, 25, and 19, it remains to prove the following:

1. If $1 + \omega(n^{-1/3}) = pn \le \gamma$, then
   - $C^{\mathrm{core}}(u) \neq C^{\mathrm{core}}(v)$ for every $v \in V(\mathrm{C}_n) \setminus D$ and $u \in D$;
   - $C^{\mathrm{core}}(u) \neq C^{\mathrm{core}}(v)$ for any non-interchangeable pair $u, v \in V(\mathrm{C}_n) \setminus D$;
2. If $1 - n^{-1/3}\ln n \le pn = 1 + o(1)$, then
   - $C(u) \neq C(v)$ for every $v \in V(\mathrm{C}_n) \setminus D$ and $u \in D$;
   - $C(u) \neq C(v)$ for any non-interchangeable pair $u, v \in V(\mathrm{C}_n) \setminus D$.

We will use the following technical fact, which follows from [17, Thm. 2] and [18, Thm. 1].

▷ Claim 26. Let $\delta > 0$ be a constant, $n^{-1/3} \ll \delta_n := pn - 1 \le \delta$, and $G_n \sim G(n, p)$. Then whp in $K(G_n)$ there are no complex subgraphs of size at most $(\ln(n\delta_n^3))^{3/4}$.

For the sake of brevity, below we prove both statements in two different regimes simultaneously. Thus, with some abuse of notation, in the supercritical phase (i.e., $1 + \omega(n^{-1/3}) = pn \le \gamma$), we let $G_n := \mathrm{C}_n$ and $C := C^{\mathrm{core}}$ as we only consider CR on C_n. We also assume that when $1 + \omega(n^{-1/3}) = pn \le \gamma$, the core is equipped with the fractional distance d^f, constituting the metric space $\mathcal{M}_n$. If $1-n^{-1/3}\ln n \leq pn \leq 1+o(1)$, then G_n is equipped with the usual shortest-path distance, that we denote by d^f as well. We also use the following notation: $d=\lfloor(\ln(\delta_n^3 n))^{2/3}\rfloor$ when we prove the assertion for $1+\omega(n^{-1/3})=pn\leq\gamma$ and $d=\lfloor 3A\ln n\rfloor$ when we prove it for $1-n^{-1/3}\ln n\leq pn=1+o(1)$. In what follows, we assume that the assertions of Claims 22, 23, 25, and 26 hold deterministically in G_n.

1. Assume that some $v\notin D$ and $u\in D$ have $C(u)=C(v)$. We know that v is d-close to a cycle F of length at most $2d$. If $v\in V(F)$, then let v' be the closest to v vertex on F that has degree more than 2 in the core. Otherwise, let $v'=v$. Let P be the shortest path from v' to F. Let us extend this path by a path P' of length $10d$ beyond v'. Due to Claim 23 and Claim 26, it has a subpath $w\ldots w'$ of length $5d$ consisting of vertices from D only such that $d^f(w,v')\leq d$. We know that all elements of the vector $\mathbf{c}:=(C(w),\ldots,C(w'))$ are different. Then, due to our assumption, u must have a vertex z at distance at most $d^f(w,v)$ such that z is the first vertex of a path $z\ldots z'$ with $C(w)=C(z)$, $\ldots$, $C(w')=C(z')$.
   We now consider separately two cases: $w=z$ and $w\neq z$. In the first case, we have that the distance from u to the closest cycle (which is F, the same as for v) is at most
   $$d^f(w,u)+d^f(w,v')+d^f(v',F)\leq \text{length of } F+2(d^f(w,v')+d^f(v',F))\leq 6d.$$
   Let P be the shortest path between u and v. Due to Claim 23 and Claim 26, there exists a path $u\tilde{w}\ldots\tilde{w}'$ of length $5d+1$ that does not meet P and consists of vertices from D only. Due to Claim 22 and Claim 25 all elements of the vector $\tilde{\mathbf{c}}:=(C(\tilde{w}),\ldots,C(\tilde{w}'))$ are different and no element of $\tilde{\mathbf{c}}$ equals to any element of $\mathbf{c}$. Moreover, by construction, $d^f(v,\tilde{w})>d^f(u,\tilde{w})$. Then, due to our assumption, v must have a neighbor $\tilde{z}\neq\tilde{w}$ such that $\tilde{z}$ is the first vertex of a path $\tilde{z}\ldots\tilde{z}'$ with $C(\tilde{w})=C(\tilde{z})$, $\ldots$, $C(\tilde{w}')=C(\tilde{z}')$. Note that $\tilde{w}\neq\tilde{z},\ldots,\tilde{w}'\neq\tilde{z}'$ due to Claim 23 and Claim 26. Since all vertices in D are distinguished by $C(\cdot)$, we conclude that all vertices $\tilde{z},\ldots,\tilde{z}'$ must be outside D. Due to Claim 23, Claim 26, and the definition of D, they constitute a (self-avoiding) path and are d-close to a cycle of length at most $2d$. Since the path has length $5d$, we get a contradiction with Claim 23 or Claim 26.
   We then assume $w\neq z$. It may only happen when $z\notin D$. Moreover, all $z,\ldots,z'$ are not in D. Indeed, otherwise, different paths $w\ldots w'$ and $z\ldots z'$ have common vertices. Then the path from z to F that goes through w has length greater than d. However, due to Claim 23 and Claim 26, there are no two different paths from z to F, both of length at most $12d$ and, also, there is no other cycle F' of length at most $2d$ such that a path from z to F' has at most d vertices. This contradicts the fact that $z\notin D$. Thus, we again get a (self-avoiding) path consisting of vertices $z,\ldots,z'$ that are d-close to a cycle of length at most $2d$. This contradicts Claim 23 or Claim 26 again, since the path has length $5d$.
   We conclude that every vertex $u\in D$ has $C(u)$ that does not equal to the color of any other vertex in the core.
2. It remains to prove that, for any two distinct $u,v\notin D$ that are not interchangeable, $C(u)\neq C(v)$. Fix two such vertices u and v. We may assume that $T_u\cong T_v$ since otherwise $C(u)\neq C(v)$ due to Claim 15. Let F_u and F_v be two cycles of length at most $2d$ that are closest to u and v respectively (both are at distance at most d from the respective vertices). If $F_u\neq F_v$, then set $F:=F_u$. In this case, we let $u'=u$ when $u\notin V(F)$ and let u' be the closest vertex of degree 3 in F to u otherwise. If $F_u=F_v=:F$, then, without loss of generality we assume that either u is not in F or both u,v are in F. Let $u'=u$ when $u\notin V(F)$ and let u' be a vertex of F that has degree at least 3 and such that $d^f(u,u')\neq d^f(v,u')$ otherwise. Note that such a vertex exists due to the definition

of an interchangeable pair. Consider a path P of length $10d$ that starts at u' and does not meet F. Due to Claim 23 and Claim 26, this path has a vertex w in D such that $d(w, u') \leq d$. If $F_u \neq F_v$, then

$$d^f(v, w) \geq d^f(F_u, F_v) - d^f(v, F_v) - d^f(w, F_u) > d^f(u, w)$$

due to Claim 23 and Claim 26. Finally, let $F_u = F_v$. Assume, in addition, $u \notin V(F)$. Then the only possibility for $C(u)$ to be equal to $C(v)$ is to have a path P' between v and w of length $d^f(u, w)$. Let us extend P' by a path of length $10d$ beyond v. Due to Claim 23 and Claim 26 this path has a vertex w' from D such that $d^f(w', v) \leq d$. But then $d^f(u, w') > d^f(v, w')$, implying $C(u) \neq C(v)$. If $u, v \in V(F)$, then

$$d^f(v, w) = d^f(v, u') + d^f(u', w) \neq d^f(u, u') + d^f(u', w) = d^f(u, w).$$

In either case, we get $d^f(v, w) \neq d^f(u, w)$ or $C(u) \neq C(v)$. Recalling that w has a unique color, we readily conclude that $C(u) \neq C(v)$, completing the proof.

References

1 Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullmann. *The Design and Analysis of Computer Algorithms.* Addison-Wesley Longman Publishing Co., 1974.

2 Michael Anastos, Matthew Kwan, and Benjamin Moore. Smoothed analysis for graph isomorphism, 2024. arXiv:2410.06095.

3 D. Angluin. Local and global properties in networks of processors. In *The 12th Annual ACM Symposium on Theory of Computing*, pages 82–93, 1980.

4 Vikraman Arvind, Johannes Köbler, Gaurav Rattan, and Oleg Verbitsky. Graph isomorphism, color refinement, and compactness. *Comput. Complex.*, 26(3):627–685, 2017. doi:10.1007/s00037-016-0147-6.

5 L. Babai, P. Erdős, and S. M. Selkow. Random graph isomorphism. *SIAM Journal on Computing*, 9(3):628–635, 1980. doi:10.1137/0209047.

6 László Babai. Moderately exponential bound for Graph Isomorphism. In *Proc. of the 3rd Int. Conf. on Fundamentals of Computation Theory (FCT'81)*, volume 117 of *Lecture Notes in Computer Science*, pages 34–50. Springer, 1981. doi:10.1007/3-540-10854-8_4.

7 László Babai. Graph isomorphism in quasipolynomial time. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing (STOC'16)*, pages 684–697, 2016. doi:10.1145/2897518.2897542.

8 László Babai. Canonical form for graphs in quasipolynomial time: preliminary report. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC'19)*, pages 1237–1246, 2019. doi:10.1145/3313276.3316356.

9 Cristoph Berkholz, Paul Bonsma, and Martin Grohe. Tight lower and upper bounds for the complexity of canonical colour refinement. *Theory of Computing Systems*, 60:581–614, 2017. doi:10.1007/S00224-016-9686-0.

10 N. Biggs. *Algebraic graph theory.* Cambridge University Press, 2nd edition, 1994.

11 Tom Bohman, Alan M. Frieze, Tomasz Luczak, Oleg Pikhurko, Clifford D. Smyth, Joel Spencer, and Oleg Verbitsky. First-order definability of trees and sparse random graphs. *Comb. Probab. Comput.*, 16(3):375–400, 2007. doi:10.1017/S0963548306008376.

12 Bela Bollobás. Distinguishing vertices of random graphs. *Ann. Discrete Math.*, 13:33–50, 1982.

13 Bela Bollobás. The evolution of random graphs. *Transactions of the American Mathematical Society*, 286(1):257–274, 1984.

14 Bela Bollobás. *Random graphs.* Cambridge University Press, 2001.

15 D. M. Cvetković, M. Doob, and H. Sachs. *Spectra of graphs. Theory and applications.* Leipzig: J. A. Barth Verlag, 3rd edition, 1995.

16 Tomek Czajka and Gopal Pandurangan. Improved random graph isomorphism. *Journal of Discrete Algorithms*, 6:85–92, 2008. doi:10.1016/J.JDA.2007.01.002.

17 Jian Ding, Jeong Han Kim, Eyal Lubetzky, and Yuval Peres. Anatomy of young giant component in the random graph. *Random Structures & Algorithms*, 39(2):139–178, 2011. doi:10.1002/RSA.20342.

18 Jian Ding, Eyal Lubetzky, and Yuval Peres. Anatomy of the giant component: The strictly supercritical regime. *European Journal of Combinatorics*, 35:155–168, 2014. doi:10.1016/J.EJC.2013.06.004.

19 Paul Erdős and Alfred Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:17–61, 1960.

20 Paul Erdős and Alfred Rényi. Asymmetric graphs. *Acta Math Acad Sci Hung*, 14:295–315, 1963.

21 Julia Gaudio, Miklós Z. Rácz, and Anirudh Sridhar. Average-case and smoothed analysis of graph isomorphism, 2023. arXiv:2211.16454.

22 László Hegedüs and Benedek Nagy. On periodic properties of circular words. *Discrete Mathematics*, 339(3):1189–1197, 2016. doi:10.1016/j.disc.2015.10.043.

23 Neil Immerman and Eric Lander. *Describing Graphs: A First-Order Approach to Graph Canonization*, pages 59–81. Springer New York, 1990.

24 Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs.* John Wiley & Sons, 2000.

25 Sandra Kiefer, Pascal Schweitzer, and Erkal Selman. Graphs identified by logics with counting. *ACM Trans. Comput. Log.*, 23(1):1:1–1:31, 2022. doi:10.1145/3417515.

26 Andreas Krebs and Oleg Verbitsky. Universal covers, color refinement, and two-variable counting logic: Lower bounds for the depth. In *30th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS'15)*, pages 689–700. IEEE Computer Society, 2015. doi:10.1109/LICS.2015.69.

27 N. Linial and J. Mosheiff. On the rigidity of sparse random graphs. *Journal of Graph Theory*, 85(2):466–480, 2017. doi:10.1002/JGT.22073.

28 T. Łuczak. The automorphism group of random graphs with a given number of edges. *Math Proc Camb Phil Soc*, 104:441–449, 1988.

29 Tomasz Łuczak. The phase transition in a random graph. In D. Miklós, V.T. Sós, and T. Szőnyi, editors, *Combinatorics, Paul Erdős is Eighty*, volume 2, pages 399–422. Bolyai Soc. Math. Stud. 2, J. Bolyai Math. Soc., Budapest, 1996.

30 Tomasz Łuczak, Boris Pittel, and John C. Wierman. The structure of a random graph at the point of the phase transition. *Transactions of the American Mathematical Society*, 341(2):721–748, 1994.

31 W. S. Massey. *Algebraic topology: An introduction*, volume 56 of *Graduate Texts in Mathematics.* Springer, 5th edition, 1981.

32 Marc Noy, Vlady Ravelomanana, and Juanjo Rué. On the probability of planarity of a random graph near the critical point. *Proc. Am. Math. Soc.*, 143(3):925–936, 2015. doi:10.1090/S0002-9939-2014-12141-1.

33 Oleg Verbitsky and Maksim Zhukovskii. Canonical labeling of sparse random graphs, 2024. arXiv:2409.18109.

34 B.Yu. Weisfeiler and A.A. Leman. The reduction of a graph to canonical form and the algebra which appears therein. *NTI, Ser. 2*, 9:12–16, 1968. In Russian. English translation is available at https://www.iti.zcu.cz/wl2018/pdf/wl_paper_translation.pdf.

35 E. M. Wright. Asymmetric and symmetric graphs. *Glasgow Math J*, 15:69–73, 1974.