Random Abstractions of Great Circle Graphs

by

Christina Knöbel

A master's project
presented to the University of Waterloo
in fulfillment of the
project requirement for the degree of
Master of Mathematics
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 2010

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The aim of this project paper is to analyze models that abstract from great circle graphs. After a brief overview of random d-regular graphs and great circle graphs, we abstract these graphs to two models, G_n^* and $G_n^{\mathcal{A}}$. We will show that the short cycles distribution in G_n^* is asymptotically Poisson; We will present a conjecture on contiguity to $G_{n,4}$ and then show that asymptotically almost surely all graphs in G_n^* are 4-connected. We will end by presenting some open problems.

Acknowledgements

First and foremost, I would like to thank my advisor, Professor Nick Wormald, for pointing me towards open problems and having a good feeling for what was feasible. Thank you for your patience.

I would like to thank the Austrian government and the BMWF: The post-graduate scholarship of the BMWF for non-German speaking countries enabled me to do my master's and research here. Thank you.

I would also like to thank Māris Ozols for proof-reading my paper, helping me with Figure 1.1 and cooking me dinner. I really appreciate your support.

Last but not least, I would like to thank my parents for always believing in me.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

Random regular graphs are a field of research that has been studied extensively, and many beautiful models exist.

In this project paper, we will start by presenting the history of random graphs and explain the pairing model that is used to describe the uniform d-regular graph model (Chapter 1). We will then present great circle graphs, and then, in Chapter 2, abstract them to two random models, G_n^* and $G_n^{\mathcal{A}}$. Next, in Chapter 3, for G_n^*, we will show that the short cycles distribution is asymptotically Poisson; We will then present a conjecture on contiguity to $G_{n,4}$ in Chapter 4, and finally, in chapter 5, we will show that asymptotically almost surely all graphs in G_n^* are 4-connected. We will end by presenting some open problems.

1.2 Preliminaries and notation

To start out, we need some definitions and notations, from graph theory as well as from probability theory.

The basic concepts are assumed to be known to the reader. For a given graph $G = (V, E)$, we will let $V(G)$ denote the vertices of G, and $E(G)$ denote the edges of G. When we talk of a graph G, we will assume it is simple. Similarly, we will write $|G|$ for the number of vertices of G, and $||G||$ for the number of edges.

For $n \in \mathbb{N}$, we let $[n]$ denote the set $\{1, 2, \dots, n\}$. Let $\mathbb{P}(\mathcal{A})$ denote the probability of event $\mathcal{A}$ occurring.

We will also be using the "big-O" notation, see e.g. [16]: For two functions, $f(n)$ and $g(n)$,

- $f(n) = O(g(n))$ if there exist constants C and n_0 such that $f(n) \leq Cg(n)$ for all $n \geq n_0$,
- $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$,
- $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$,
- $f(n) \sim g(n)$ if $f(n)/g(n) \to 1$ as $n \to \infty$ and finally
- $f(n) = o(g(n))$ as $n \to \infty$ if $f(n)/g(n) \to 0$ as $n \to \infty$.

As for notations from probability theory, let $\mathrm{Po}(\lambda)$ denote the Poisson distribution with parameter λ. A probabilistic event $\mathcal{A}_n$ is said to hold *asymptotically almost surely (a.a.s.)* if $\mathbb{P}(\mathcal{A}_n) \to 1$ for $n \to \infty$. For any given random variable Z, we say that a sequence $X_1, X_2, \dots$ converges towards Z in distribution as $n \to \infty$ if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(Z \leq x)$ for every real x such that the distribution of Z is continuous at x [16]. We write $X \xrightarrow{d} Z$. At some point, we will also need Chebyshev's inequality:

Theorem 1.1 (Chebyshev's inequality)**.** *For a random variable X with expected value μ and variance $\sigma^2 < \infty$, for any $k > 0$, it holds that*

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}. \tag{1.1}$$

1.3 Brief history of random graphs

While Erdős used random graphs already in [7] to prove the existence of graphs with a specific Ramsey property, one might say that the active study of random graphs goes back to Erdős and Rényi's groundbreaking papers from 1959 and 1960 [8, 9]. The two models they introduced, $G_{n,p}$ and $G_{n,M}$, have been extremely well analyzed; for an extensive overview of what has been done, see, for example, [2, 16]. Here, I will briefly explain how these models work, and what properties of these models are known.

Let $N := \binom{n}{2}$. The model $G_{n,p}$ is parametrized by a number $0 \leq p \leq 1$. The probability space $(\Omega^{(n)}, \mathcal{F}, \mathbb{P})$ consists of the set of all graphs on n vertices, $\Omega^{(n)}$, where each of the graphs $G \in \Omega^{(n)}$ is realized with probability $\mathbb{P}(G) = p^{e(G)}(1-p)^{N-e(G)}$. This means that for each edge $e \in E(G)$, a coin is tossed: with probability p, e is included in the graph G, while with probability $1-p$, e is not chosen.

For the other approach to classical random graphs, $G_{n,M}$, the probability space $G_{n,M}$ consists of $(\Omega_M^{(n)}, \mathcal{F}_M, \mathbb{P}_M)$, where $\Omega_M^{(n)}$ is the space of all graphs on n vertices that have exactly M edges. A graph G is chosen from $\Omega_M^{(n)}$ uniformly at random, so $\mathbb{P}(G) = \frac{1}{\left|\Omega_M^{(n)}\right|}$, where $\left|\Omega_M^{(n)}\right| = \binom{N}{M}$.

A nice way to view a graph in $G_{n,M}$ is as the result of the mth step of a graph process: Starting with a graph on n vertices and no edges, at each time step t an edge that is not yet in the graph is chosen at random and added to the graph.

It can be shown (see, i.e. [16, p. 14]) that these two models are asymptotically equivalent. This means that, if $M \sim \binom{n}{2}p$, a majority of properties of graphs in $G_{n,p}$ are very similar to properties that hold for graphs in $G_{n,M}$.

Many properties about $G_{n,p}$ are known, such as:

- For any integer l, let $X_{c,l}(G)$ denote the number of l-cycles in a given graph G. If $np \to c > 0$, then for the cycle count X_{c_l}, of $G \in G_{n,p}$, we have $X_{c_l}(G) \xrightarrow{d} \mathrm{Po}(\lambda)$, the Poisson distribution with expectation $\lambda = \frac{c^l}{l}$ [16].
- For p large enough, we can speak of the diameter of a random graph in a non-trivial way [2]: If c is a positive constant, $d = d(n) \geq 2$ a natural number, and for $p = p(n, c, d)$, $0 < p < 1$,
$$p^d n^{d-1} = \log(n^2/c),$$
such that $pn/(\log n)^3 \to \infty$, then in $G_{n,p}$ we have
$$\lim_{n\to\infty} \mathbb{P}(\text{diam G} = d) = e^{-c/2} \quad \text{and} \quad \lim_{n\to\infty} \mathbb{P}(\text{diam G} = d+1) = 1 - e^{-c/2}.$$
- Many other questions such as the emergence of a giant component [15] and the chromatic number [20], have also been studied intensely.

The study of random graphs has generated many beautiful results. In fact, the first use of the probabilistic method is believed to be by Erdős in conjecture with a random graph [7]. However, as significant as "classical" random graphs are, for many cases they are just too general. For example, if we want a graph to model the links on the Internet, it may seem that these links occur in a random manner, but clearly not uniformly at random. Popular sites such as *Google* or *Facebook* tend to become ever more popular, the more hyper references there already are linking to them. Other models had to be found to analyze this type of behaviour [4, 5].

It is also easy to find examples where one wishes to model a random graph where each vertex has the same degree, d, i.e. a d-regular graph. Consider for example a peer-to-peer network such that each user is represented by a vertex of the graph, vertices leave and enter the system at random, and at each point t in time, the graph should be connected and every vertex should connect to exactly d other vertices. Properties of models for a type of system like this have been analyzed [11, 12], though of course, mathematicians started with the most general case of random d-regular graphs first.

1.4 Uniform random regular graphs and the pairing model

Let $G_{n,d}$ be the space of d-regular random graphs on n vertices, such that each graph appears with equal probability. This is also called the *uniform model*. Clearly, dn must be even.

What seems to have started in [1] as the *configuration model* and is described in [25] as the *pairing model* is often used for analyzing $G_{n,d}$.

1.4.1 The pairing model

Consider a set of dn balls and n bins, such that each bin contains d balls. Uniformly at random, pick a random matching between these dn balls. This creates $dn/2$ pairs of balls, and is called a *pairing*. Form a graph as follows: All balls in a given bin are identified as one vertex representing the bin. For every matched pair of balls, an edge is placed between the corresponding bins. Note that this gives a random, d-regular multigraph, as it is possible that two balls in the same bin are matched (a loop), and also that several balls in some bin i are matched with some other balls in bin j, so multi-edges are also allowed. This model is denoted by $\mathcal{P}_{n,d}$.

Since graphs in $G_{n,d}$ have no loops or multiple edges, each G in $G_{n,d}$ corresponds to $(d!)^n$ pairings. Randomly choosing a graph from $\mathcal{P}_{n,d}$ under the condition that the graph is simple gives a uniform distribution over all d-regular graphs. It is thus possible to analyze $G_{n,d}$ by analyzing $\mathcal{P}_{n,d}$ and conditioning on the fact that the multigraph chosen from $\mathcal{P}_{n,d}$ is simple.

However, note that randomly choosing a multigraph from the non-simple outcomes of the pairing model is not uniform: For each loop or multiple edge, the number of corresponding pairings decreases.

Lemma 1.1 ([25]). *Let A be an event in $G_{n,d}$ and let A' be the set of pairings that corre-*

spond to A. Then

$$\mathbb{P}_{G_{n,d}}(A) = \frac{\mathbb{P}_{\mathcal{P}_{n,d}}(A')}{\mathbb{P}(G \textit{ simple})}.$$

Proof.

$$\begin{aligned}\mathbb{P}_{G_{n,d}}(A) &= \mathbb{P}_{\mathcal{P}_{n,d}}(A' \mid G \text{ simple }) \\ &= \frac{\mathbb{P}_{\mathcal{P}_{n,d}}(A')}{\mathbb{P}(G \text{ simple})}.\end{aligned}$$

Here, the first equality follows because each graph $G \in G_{n,d}$ corresponds to the same number of pairings, and as both models are uniform, going from the probability space $G_{n,d}$ to $\mathcal{P}_{n,d}$ does not change probabilities as long as we are only considering simple pairings. Finally, the last equality follows by the definition of conditional probabilities. □

With help of the pairing model, it is possible to give an approximation for $|G_{n,d}|$. Let r be an even number. How many ways are there of matching r numbers? Consider the following algorithm (Algorithm 1) that runs on the set $A := [r]$.

Algorithm 1 An algorithm that produces a random matching on $[r]$

Procedure RandomPairing (r: even integer)
$A := [r]$;
for $i = 1$ to r **do**
  if $i \in A$ **then**
    $A \leftarrow A \setminus i$
    Choose an element j in A uniformly at random;
    Pair i with j;
    $A \leftarrow A \setminus j$
  end if
end for
return ($r/2$ pairs)

For $i = 1, \ldots, r$, if $i \in A$, it generates uniformly at random a number $j \neq i$, $j \in A$, and matches i with j. Then, i and j are removed from A. Each different outcome of this algorithm is equally likely. At the first step, there are $r-1$ choices to choose what to pair with 1, at the next step, there are $r-3$ choices as to what to pair with the next smallest element left in A, and so on. Finally, we have

$$(r-1)(r-3)\cdot\cdots\cdot 2 = (r-1)!! = \frac{r!}{(r/2)!2^{r/2}}$$

different matchings of $[r]$.

In particular, consider $r = dn$ in the pairing model. Each of the d balls in each bin is equivalent, so there are

$$|\mathcal{P}_{n,d}| = \frac{(dn)!}{(dn/2)!2^{dn/2}(d!)^n} \tag{1.2}$$

outcomes of the pairing model. With this, it follows that

$$|G_{n,d}| = \frac{(dn)!}{(dn/2)!2^{dn/2}(d!)^n}\mathbb{P}(\text{simple}).$$

The only thing necessary now is to estimate $\mathbb{P}_{\mathcal{P}_{n,d}}(\text{simple})$. In [18], McKay and Wormald showed as a corollary that, for $d = o(\sqrt{n})$, asymptotically

$$\mathbb{P}(\text{simple}) = \exp\left(\frac{1-d^2}{4} - \frac{d^3}{12n} + O\left(\frac{d^2}{n}\right)\right).$$

For example, for $d = 4$, as n gets large, there are asymptotically

$$\frac{(4n)!}{(2n)!2^{2n}24^n}\exp\left(\frac{-15}{4} - \frac{64}{12n} + O\left(\frac{1}{n}\right)\right) \tag{1.3}$$

graphs in $G_{n,4}$.

Corollary 1.1. *If an event $\mathcal{A}$ holds a.a.s. in $\mathcal{P}_{n,d}$, it also holds a.a.s. in $G_{n,d}$.*

Proof.

$$\begin{aligned}
\mathbb{P}_{\mathcal{P}_{n,d}}(\mathcal{A}) &= \mathbb{P}_{\mathcal{P}_{n,d}}(\mathcal{A} \cap G \text{ simple}) + \mathbb{P}_{\mathcal{P}_{n,d}}(\mathcal{A} \cap G \text{ not simple}) \\
&= \mathbb{P}_{\mathcal{P}_{n,d}}(\mathcal{A} \mid G \text{ simple})\mathbb{P}_{\mathcal{P}_{n,d}}(G \text{ simple}) + \mathbb{P}_{\mathcal{P}_{n,d}}(\mathcal{A} \mid G \text{ not simple})\mathbb{P}_{\mathcal{P}_{n,d}}(G \text{ not simple}) \\
&= \mathbb{P}_{G_{n,d}}(\mathcal{A})\mathbb{P}_{\mathcal{P}_{n,d}}(G \text{ simple}) + \mathbb{P}_{\mathcal{P}_{n,d}}(\mathcal{A} \mid G \text{ not simple})\mathbb{P}_{\mathcal{P}_{n,d}}(G \text{ not simple}) \\
&= \mathbb{P}_{G_{n,d}}(\mathcal{A})\mathbb{P}_{\mathcal{P}_{n,d}}(G \text{ simple}) + \mathbb{P}_{\mathcal{P}_{n,d}}(G \text{ not simple})
\end{aligned}$$

The last equality follows from $\mathbb{P}_{\mathcal{P}_{n,d}}(\mathcal{A}) \to 1$, and as $\mathbb{P}_{\mathcal{P}_{n,d}}(G \text{ simple})$ is bounded away from 0. With the same reasoning, as $\mathbb{P}_{\mathcal{P}_{n,d}}(\mathcal{A}) = 1$ *a.a.s.*, we have $\mathbb{P}_{G_{n,d}}(\mathcal{A}) = 1$ *a.a.s.* □

1.4.2 Results on graphs in $G_{n,d}$

The field of random regular graphs has been intensely studied. For a very thorough overview, see [25].

Many of these results were in fact obtained using the pairing model, such as

Theorem 1.2. *[1, 23] The distribution of the number of i-cycles in $G \in G_{n,d}$ is asymptotically Poisson, with mean $\lambda_i = \frac{(d-1)^i}{2i}$, for all fixed i.*

See also Section 3.3.

Theorem 1.3. *[24] For $d \geq 3$, $G_{n,d}$ is d-connected a.a.s.*

Theorem 1.4. *[3] The diameter of a graph $G \in G_{n,d}$ is $O(\log(n))$ a.a.s.*

For this project paper, another important result will be the following:

Theorem 1.5. *[21] A graph $G \in G_{n,4}$ has chromatic number 3 a.a.s.*

1.5 Great circle graphs

Consider the following geometrical construction: We are given a 3-dimensional sphere, without loss of generality the sphere of radius 1 centred at the origin. We are also given n planes that pass through the origin. Consider the intersection of the planes with the sphere: Each intersection forms a great circle on the sphere. (See Figure 1.1) For any two distinct great circles i and j, it holds that they meet in precisely two points. We construct a graph as follows:

- Each intersection point of two or more circles becomes a vertex embedded in the plane.
- Two vertices v_i and v_j are adjacent if the arc between the points v_i and v_j is not intersected by any other arc.

This construction is then a *great circle graph*. If for every point it holds that at most two great circles intersect in it, we call the corresponding graph simple. In this case, it is a 4-regular graph. Except when otherwise stated, we will always assume a great circle graph to be simple.

In fact, this construction can be generalized a bit [10]. Consider the sphere S and a family $\{c_1, \ldots, c_n\}$ of simple closed curves on it such that

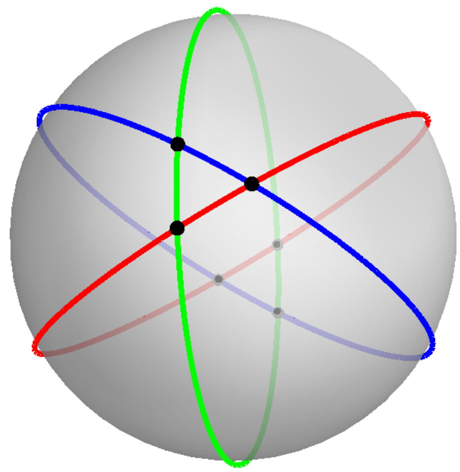


Figure 1.1: A great circle graph – the intersection of the great circles form the vertices, while the arcs between the vertices form edges.

- every two curves have exactly two points in common at which they cross and
- for every three different indices $i, j, k \in [n]$, curve c_k separates the two intersections of c_i and c_j.

We call the set of graphs G_n^a that this construction gives *arrangement graphs*. Note that simple great circle graphs form a subset of arrangement graphs.

For any arrangement graph G, several properties hold. The first three follow easily.

- First and foremost, G is *planar*. This means that, by the four-colour theorem [19], G is four colourable.
- G is 4-regular.
- G has $n(n-1)$ vertices and $2n(n-1)$ edges.
- For $n \geq 3$, G is 4-connected.

In [10], the last item is proved using Menger's theorem. (See, for example, [6]).

Wiring diagrams corresponding to graphs in $G_n^{\mathcal{A}}$ are a set of n lines in the plane each corresponding to a curve on S. Their y coordinate is fixed except when two curves, say c_i and c_j, cross. At this point, the lines corresponding to c_i and c_j "swap" y-coordinates. This construction was introduced in [13], and was used in [10] to prove following theorem:

Theorem 1.6 ([10]). *Every pseudo-cycle arrangement can be decomposed into two edge-disjoint Hamilton cycles, and the decomposition can be found efficiently [10].*

Note that the following theorem, also [10], follows as a direct corollary:

Theorem 1.7 ([10]). *Circle arrangement graphs are four edge colourable.*

This follows from the fact that, after decomposing a graph on an even number of vertices into two Hamilton cycles, one of the cycles can be coloured in colours 1 and 2, while the other cycle can be coloured in colours 3 and 4.

Conjecture 1.1 ([10]). *Circle arrangement graphs are 3-vertex colourable.*

This has been proved for arrangements of up to 11 circles by Aichholzer[1]. For cases where there are more than 11 circles, this seems to be stated as an open problem[2],[3], though Cahit claims[4] to have proved it by counting chains of triangles.

All these constructions of graphs lead to the idea of the models analyzed in this paper.

[1] Personal communication

[2] http://www.tex.ac.uk/cgi-bin/texfaq2html?label=citeURL, 2 Nov 2010

[3] http://garden.irmacs.sfu.ca/?q=category/arrangement_graph, 2 Nov 2010

[4] http://arxiv.org/abs/math/0408363

Chapter 2

The models

From the different models presented in the previous chapter — $G_{n,d}$ and arrangement graphs — we came up with two random generalizations of arrangement graphs. We are aiming for a random graph model of an abstraction of great circle graphs that does not include the planarity of the graphs— not including planarity makes the model much easier to analyse.

Instead of circles, we will be considering cycles. If two cycles c_i and c_j in a graph share a vertex v , we call v an *intersection* of c_i and c_j. We say a vertex $v \neq w_1, w_2$ separates vertices w_1 and w_2 in a path P going from w_1 to w_2 if $v \in P$.

In this sense, arrangement graphs translate to graphs with n cycles on $2n-2$ vertices such that for each three cycles c_i, c_j, c_k we have that in both c_i and c_j the intersections of c_i and c_j are separated by a vertex in c_k.

We can abstract from this even further by taking n cycles on $2n-2$ vertices such that each pair of cycles c_i and c_j must intersect in exactly two vertices. We will start by describing this abstraction first.

2.1 Random cycle arrangement graphs

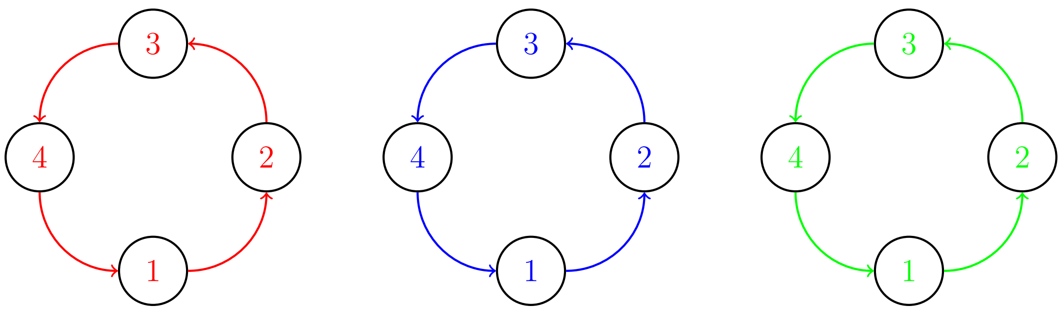
The idea for this model stems from an abstraction of the great circle graph presented in Section 1.5. Formally, consider a graph whose edges are partitioned into n cycles each of length $2n-2$, such that each two cycles intersect in precisely two vertices and each vertex has degree 4. We will call these n cycles *initial* cycles. This setup is similar to that of the great cycle graphs, except that we have taken away the planarity. Also, the outcome might not be a simple graph as this model permits double edges.

Create such a graph by starting with n disjoint cycles and identifying two vertices for each pair of cycles, in a random manner such that each valid way of doing the n identifications is equally likely. We will call a graph generated like this a *random cycle arrangement graph*.

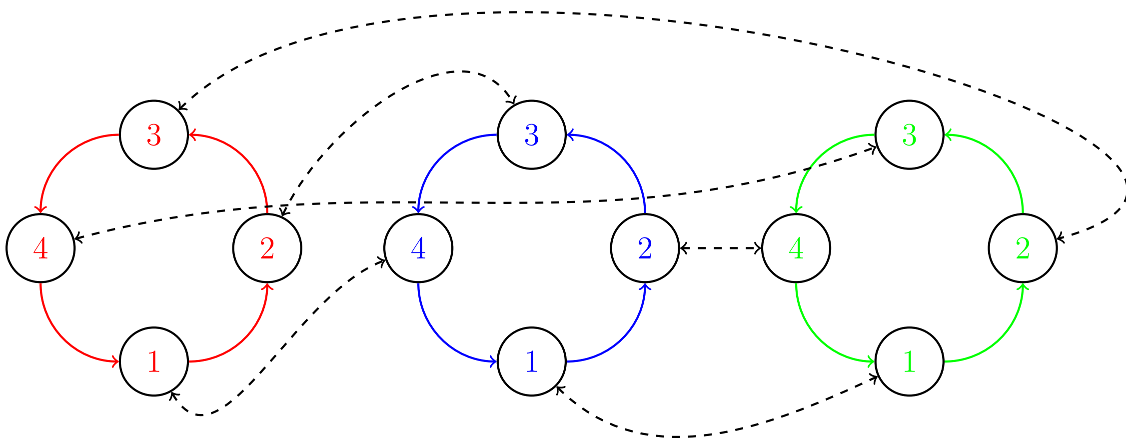
2.1.1 An easy example in pictures

Here is an example, in pictures, of how this would work for the case $n = 3$. Following this very easy example will help understand the notation and calculations later.

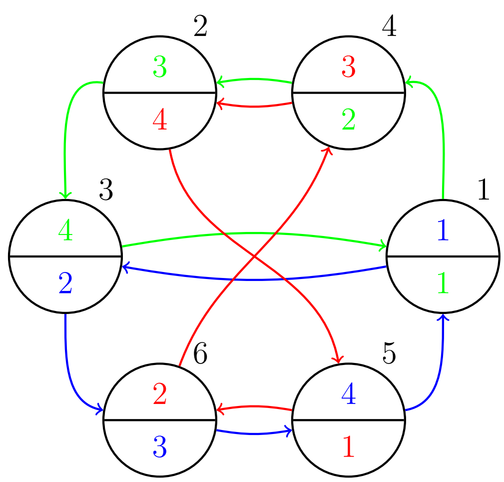
We start with considering three directed cycles, each of length four. To distinguish between them, let one of them be red, one blue and one green.



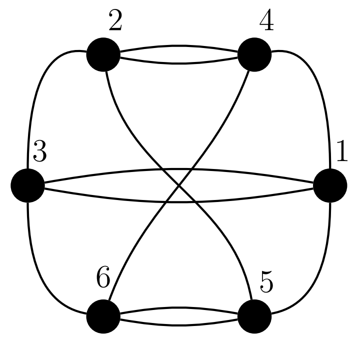
As the next step, match two vertices of the red cycle with two vertices of the blue cycle. The other two vertices of the red cycle are matched with two vertices of the green cycle in the same manner, while two vertices of the blue cycle are matched with two vertices of the green cycle. This is done in a random manner that makes each such identification equally likely. (How this random manner works will be described in Section 2.1.2). The result might look something like this:



Next, all vertices that have been matched are identified with each other, while the initial edges remain in the graph. Thus, we go from $n(2n-2) = 12$ vertices to $n(n-1) = 6$ vertices. At this point, a labelling of the vertices will also occur by some random algorithm.



Finally, to obtain the graph we want, we take away the colours, the two initial labels each vertex has, and the orientation of the edges. We end up with a graph that looks like this:



Note that different labellings of the initial matchings of vertices correspond to the same graph, and also that the outcome of these vertex identifications has double edges.

We will permit these double edges in the examination of this model. In further research, conditioning on having no double edges would be an interesting step to take.

Let us denote this model by G_n^*. If a given graph can be decomposed as in this model and is thus in G_n^*, we call such a decomposition a *lock* decomposition.

What properties does this random 4-regular graph have? In particular, the question one may ask oneself is if *almost all* 4-regular graphs have such a lock decomposition. The right way to pose this question is to ask if every graph $G_n^* \in G_{n,4}$ has a lock decomposition *a.a.s.* We can formalize this by asking if the two models, G_n^* and $G_{n,4}$, are *contiguous*. (See chapter 4). This would imply that properties that hold *a.a.s.* for $G_{n,4}$ would also hold *a.a.s.* in G_n^*. In particular, by Theorems 1.4 and 1.5 we would know that

- $G \in G_n^*$ would be 3-colourable *a.a.s.*
- $G \in G_n^*$ would have a diameter of size $O(\log(n))$ *a.a.s.*

2.1.2 Generating graphs in G_n^*

Before we can examine what a graph in G_n^* typically looks like, we need to figure out how to find graphs in G_n^*. We use the following model to generate graphs in G_n^*.

Consider n labelled, directed cycles, each coloured with a colour in $\{1, 2, \dots, n\}$. These are our *initial* cycles. For every vertex v in an initial cycle, say cycle c_i, it is desired that v must be matched to another vertex w in a different initial cycle $j \neq i$, such that v and w are identified in the final graph. How do we do this?

Define $(v_i)_{c_j}$ to be vertex v_i in cycle c_j. When not otherwise stated, we will assume that the vertices in each initial cycle c_i are labelled from 1 to $2n-2$, where there is a directed edge from i to $i+1$ for all $i < 2n-2$ and an edge from $2n-2$ to 1.

Let the identification of two vertices, say v_1 in cycle c_i and v_2 in cycle c_j, $j \neq i$, be described by

$$(v_1)_{c_i} \longleftrightarrow (v_2)_{c_j}.$$

For $i \in \{1, 2, \ldots, n\}$, consider a random bijective function ϕ_{c_i} that assigns $([n] \setminus \{i\}) \times \{u, d\}$ uniformly at random to the vertices in c_i.

$$\phi_{c_i} : [2n-2] \to ([n] \setminus \{i\}) \times \{u, d\}$$

Let $\phi_{c_i}[1]$ denote the first component of ϕ_{c_i}, while $\phi_{c_i}[2]$ denotes the second. Let us call $\phi_{c_2}[2](v)$ the label of vertex v. To get some intuition, assume that u stands for *up* and d stands for *down*. After identifying two vertices with each other, we can imagine that one former vertex will inhabit the "upper" floor of the new vertex, while the other one lives "downstairs" — see the example in Subsection 2.1.1— and for each new vertex both upper and lower floor must have exactly one initial vertex.

So $(v)_{c_i} \to (j, k)$. The values (j, k) which ϕ_{c_i} assigns to each vertex v can be explained as follows: c_j is the other initial cycle v will be in, and k is v's label. In c_j, there is some vertex w such that $\phi_{c_j}(w) = (i, k')$, $k \neq k'$, i.e. there is some other vertex w that is assigned to the initial cycle c_i and has the opposite label from v. After the identifications, v and w will be one vertex. This means that

$$(v)_{c_i} \longleftrightarrow (v')_{c_j} \iff \begin{cases} \phi_{c_i}[1](v) & = j, \\ \phi_{c_j}[1](v') & = i, \\ \phi_{c_i}[2](v) & \neq \phi_{c_j}[2](v'). \end{cases} .$$

2.1.3 How many different graphs are there in G_n^*?

To find the number of graphs in G_n^*, $|G_n^*|$, consider that there are $2n-2$ outcomes for each mapping ϕ_c, and the bijections ϕ consider labelled, directed and coloured graphs. This model gives

$$\frac{((2n-2)!)^n}{n!(2n-2)^n 2^n} 2^{-n(n-1)/2},$$

different graphs, where the terms in the fraction come from the different possible bijections, accounting for the over-counting due to considering labelled (the $(2n-2)^n$ term), directed (the 2^n term) graphs such that each cycle is given a colourings (the $n!$ term). The $2^{-n(n-1)/2}$ term occurs because for every identification of vertices in two given cycles c_i and c_j, the labellings ($\phi_{c_i}[2]$ and $\phi_{c_j}[2]$) can be exchanged.

Considering all possible relabellings of the vertices, and using Stirling's formula $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$, this then gives

$$
\begin{aligned}
|G_n^*| &= \frac{(n(n-1))!((2n-2)!)^n}{n!(n-1)^n} 2^{-n(n+3)/2} \\
&\sim \frac{\sqrt{2\pi n(n-1)}(\frac{n(n-1)}{e})^{n(n-1)}(\sqrt{2\pi 2(n-1)}(\frac{2(n-1)}{e})^{2(n-1)})^n}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n (n-1)^n} 2^{-n(n+3)/2} \\
&= \frac{\sqrt{n(n-1)}(n(n-1))^{n(n-1)}(\sqrt{\pi(n-1)})^n (n-1)^{2(n-1)n} 2^{2(n-1)n+n}}{e^{3n(n-1)-n}\sqrt{n} n^n (n-1)^n} 2^{-n(n+3)/2} \\
&\sim \frac{n \cdot n^{2n(n-1)} \cdot \sqrt{\pi}^n \cdot n^{n/2} \cdot n^{2(n-1)n} \cdot 2^{2(n-1)n+n}}{e^{3n(n-1)-n} \cdot n^{1/2} \cdot n^n \cdot n^n} 2^{-n(n+3)/2} \\
&= 2^{(3n^2-5n)/2}\sqrt{\pi}^n e^{-3n^2+4n} n^{4n^2-\frac{11}{2}n+1/2}
\end{aligned}
$$

possible decompositions of labelled 4-regular graphs in total. Let $\mathcal{Z}$ denote the number of decompositions of a given multigraph $G \in P_{n,4}$ into such cycles. By equation (1.2) and Stirling's formula, we know that

$$
\begin{aligned}
|\mathcal{P}_{n,4}|^{-1} &= \frac{(4n)!}{(2n)! 2^{2n} (4!)^n} \\
&\sim \frac{\sqrt{2\pi 4n}\left(\frac{4n}{e}\right)^{4n}}{\sqrt{2\pi 2n}\left(\frac{2n}{e}\right)^{2n} 2^{2n}(24)^n} \\
&= \Theta\left(\frac{\left(\frac{4n}{e}\right)^{4n}}{\left(\frac{2n}{e}\right)^{2n} 2^{2n}(24)^n}\right) \\
&= \Theta(n^{2n} 2^n e^{-2n} 3^{-n}).
\end{aligned} \tag{2.1}
$$

With these calculations, as $|G_n^*|$ grows as a rate $\Omega(n^{n^4})$, while $\mathcal{P}_{n,4}$ grows at a rate $o(n^{n^2})$, it is thus easy to see that

$$\mathbb{E}(\mathcal{Z}) = \frac{|G_n^*|}{|\mathcal{P}_{n,4}|} \to \infty.$$

This is good news, as it is an indication that contiguity (see Chapter) may hold. However, there is not much to be deduced from this for sure: Even if $\mathbb{E}(\mathcal{Z}) \to \infty$, it is still possible that *not* almost all graphs in $G_{n,4}$ can be decomposed initial cycles.

2.2 Great cycles random graphs

As described before, arrangement graphs translate to a graph with n cycles $c_1, c_2, \ldots, c_n$ on $2n-2$ vertices, labelled by $1, 2, \ldots, 2n-2$. We will continue to call $c_1, c_2, \ldots, c_n$ initial cycles. The model works as follows:

- For each pair of cycles c_i and c_j, $i \neq j$, it holds that c_i and c_j intersect in precisely two vertices.
- Let v_{c_i,c_j} and v'_{c_i,c_j} denote the intersections of c_i and c_j, and let $P^i_{c_i,c_j}$ and $P^{i'}_{c_i,c_j}$ denote the the two paths between v_{c_i,c_j} and v'_{c_i,c_j} in c_i, and $P^j_{c_i,c_j}$ and $P^{j'}_{c_i,c_j}$ denote the the two paths between v_{c_i,c_j} and v'_{c_i,c_j} in c_j. For every $k \neq i, j$, it must hold that the intersections of c_k with c_i and c_j separate v_{c_i,c_j} and v'_{c_i,c_j} on $P^i_{c_i,c_j}$ and $P^{i'}_{c_i,c_j}$, and on $P^j_{c_i,c_j}$ and $P^{j'}_{c_i,c_j}$, respectively.

Consider only c_i. Note that for any given c_j, it holds that *every* c_k, $k \neq i, j$, separates both intersections of c_i and c_j, on both $P^i_{c_i,c_j}$ and $P^{i'}_{c_i,c_j}$. This means that for both paths along c_i from v_{c_i,c_j} to v'_{c_i,c_j}, we have n vertices. In fact, as each initial cycle has $2n-2$ vertices, with v_{c_i,c_j} and v'_{c_i,c_j}, these are all the vertices there are, and so $P^i_{c_i,c_j}$ and $P^{i'}_{c_i,c_j}$ will always have the same length, for all i, j. This is important, as it tells us that in a certain sense, each cycle in this model is symmetrical. As every pair v_{c_i,c_j}, v'_{c_i,c_j} will be separated by two paths of length $n-1$ on c_i, it suffices to know where v_{c_i,c_j} is located to know the location of v'_{c_i,c_j}. See example in Figure 2.1.

For an initial cycle c_i, denote the cycle vertex j is paired with by $c(j)$. With the above reasoning, it must hold that $c(j) = c(j+n-1)$ for $j \leq n-1$.

Take the space of all such graphs of cycle arrangements on n cycles, and pick one of them uniformly at random. We call this model the *great cycle random graph* model, and denote graphs in this model with $G^{\mathcal{A}}_n$.

2.2.1 Generating graphs in $G^{\mathcal{A}}_n$

Similarly to Section 2.1.2, we want to find a function such that each graph in $G^{\mathcal{A}}_n$ is generated uniformly at random. To do this, consider n directed cycles, $c_1, c_2, \ldots, c_n$, on $2n-2$ vertices, labelled $1, 2, \ldots, 2n-2$, such that there is an arc between i and $i+1$, for all $i < 2n-2$ and there is an arc between $2n-2$ and 1. In any cycle c_i, as $c(s) = c(s+n-1)$ for $s \leq n-1$, it is sufficient to generate $c(s)$ for the first $n-1$ vertices in the cycle.

What we do not know is if vertex s is identified with a vertex in $[n-1]$ of $c(s)$, or if it is identified with a vertex in $[2n-2] \setminus [n-1]$. As in Section 2.1.2, the function generating these random graphs will need to assign a labelling to each vertex. For every c_i, consider an initial random function $\tilde{\phi}^i_{c_i}$ that assigns $([n] \setminus \{i\}) \times \{-1, 1\}$ uniformly at random to the first $n-1$ vertices in c_i.

$$\tilde{\phi}^i_{c_i} : [n-1] \to ([n] \setminus \{i\}) \times \{-1, 1\}.$$

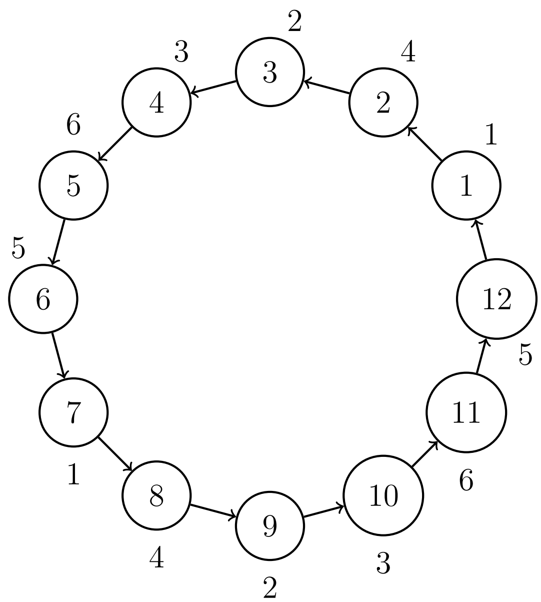


Figure 2.1: An example of a possible alignment when $n = 7$ and a given cycle on 12 vertices. The labelling beside vertex i gives $c(i)$. Note that $c(i) = c(i + 6)$, for $i \leq 6$.

such that each $\tilde{\phi}^i_{c_i}$ maps to each element in $([n] \setminus \{i\})$ exactly once. As in section 2.1.2, for a given vertex s, let $\phi^f_{c_i}[1](k)$ denote the first component of the mapping at s, while $\phi^f_{c_i}[2](k)$ denotes the second.

Then, to finalize the generation of this random graph, we define a second, bijective function $\tilde{\phi}^f_{c_i}$ that assigns some value from $([n] \setminus \{i\}) \times \{-1, 1\}$ to all vertices in c_i.

$$\tilde{\phi}^f_{c_i} : [2n - 2] \to ([n] \setminus \{i\}) \times \{-1, 1\},$$

such that

$$\tilde{\phi}^f_{c_i}(s) = \begin{cases} \tilde{\phi}^i_{c_i}(s) & \text{if } s \leq n - 1 \\ (\phi^i_{c_i}[1](s - (n - 1)), -\phi^i_{c_i}[2](s)) & \text{otherwise.} \end{cases}$$

So if $(v)_{c_i} \to (j, k)$, the values (j, k) which $\phi^f_{c_i}$ assigns to each vertex v have the same explanation as in section 2.1.2, except that the label is a bit different: c_j is the other initial cycle v will be in, and k is v's label. In c_j, there is some vertex w with $\phi_{c_j}(w) = (i, k')$, such that $k \cdot k' = 1$, i.e. there is some other vertex w that is signed to the initial cycle c_i and has the same label as v. After the identifications, v and w will be one vertex. This means that

$$(v)_{c_i} \longleftrightarrow (v')_{c_j} \quad \Longleftrightarrow \quad \begin{cases} \phi^f_{c_i}[1](v) & = j, \\ \phi^f_{c_j}[1](v') & = i, \\ \phi^t_{c_i}[2](v) & = \phi_{c_j}[2](v'). \end{cases}.$$

2.2.2 How many different graphs are there in $G_n^{\mathcal{A}}$?

Similarly to Section 2.1.2, we can count the number of graphs in $G_n^{\mathcal{A}}$. There are $n-1$ outcomes for each mapping ϕ_c^i, and the bijections ϕ^f consider labelled, directed and coloured graphs. This model gives

$$\frac{((n-1)!2^{n-1})^n}{n!(2n-2)^n 2^n} 2^{-n(n-1)/2}$$

different graphs, where the terms in the fraction come from the different possible bijections of ϕ_c^f, accounting for the over-counting due to considering labelled (the $(2n-2)^n$ term), directed (the 2^n term) graphs with given colours (the $n!$ term). The $2^{-n(n-1)/2}$ term occurs because for every identification of vertices in two given cycles c_i and c_j, the labellings ($\phi^f_{c_i}[2]$ and $\phi^f_{c_j}[2]$) can be exchanged.

Considering all possible relabellings of the vertices and using Stirling's formula, this then gives

$$\begin{aligned}
|G_n^{\mathcal{A}}| &= \frac{(n(n-1))!((n-1)!)^n}{n!(n-1)^n} 2^{\frac{n(n-1)}{2}-2n} \\
&\sim \frac{\sqrt{2\pi n(n-1)}(\frac{n(n-1)}{e})^{n(n-1)}(\sqrt{2\pi(n-1)}(\frac{(n-1)}{e})^{(n-1)})^n}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n (n-1)^n} 2^{\frac{n(n-1)}{2}-2n} \\
&\sim \frac{\sqrt{n(n-1)}(\frac{n(n-1)}{e})^{n(n-1)}(\sqrt{2\pi(n-1)}(\frac{(n-1)}{e})^{(n-1)})^n}{\sqrt{n}\left(\frac{n}{e}\right)^n (n-1)^n} 2^{\frac{n(n-1)}{2}-2n} \\
&\sim e^{-2n^2+3n}\pi^{n/2}2^{1/2n^2-2n}n^{3n^2-\frac{9}{2}n+1/2}
\end{aligned}$$

possible decompositions of labelled 4-regular graphs in total. This implies that if $\mathcal{Z}^{\mathcal{C}}$ denotes the number of decompositions of a given graph $G \in P_{n,4}$ into great cycles, from equation (2.1)

$$\mathbb{E}(\mathcal{Z}^{\mathcal{C}}) = \frac{|G_n^{\mathcal{A}}|}{P_{n,4}} \to \infty.$$

Chapter 3

Short cycles

When given a random graph model, one of the first things that is often done is to count the number of short cycles, or rather find out the asymptotic distribution of short cycles, as in [1, 22]. The easiest way to do so is using the *method of moments*. In this chapter, I will briefly describe the method of moments, and then give proofs that short cycles in both the pairing model as well as in the great cycles random graph model are asymptotically distributed according to the Poisson law. As these calculations are quite tedious, I will only state the expected number of i-cycles in the random cycle arrangement model and explain why it should also be asymptotically distributed according to the Poisson law.

To simplify reading and writing of these proofs, by abuse of notation we will assume n to be fixed and very large, so that instead of writing $Y_i(n)$ for a function, we will write Y_i.

3.1 The method of moments

Here, we will briefly describe the method of moments, see, e.g. [16]. What it says is that, given some circumstances, if the moments of a random variable are known, it is possible to find its distribution. Let us first define moments.

Definition 3.1. The *moments* of a random variable X are the numbers

$$\mathbb{E}(X^k),$$

$k \geq 1$. The *factorial moments* $\mathbb{E}(X)_k$ of a random variable X are the numbers defined by

$$\mathbb{E}(X)_k := \mathbb{E}X(X-1)(X-2)\cdots(X-k+1),$$

for $k \geq 1$.

Definition 3.2. We say the distribution of a random variable X is *determined by its moments* if

- X has finite moments and
- every random variable that has the same moments as X has the same distribution.

In particular, if X has normal or Poisson distribution, X is determined by its moments. Note that if the moments of a random variable are finite, then so are its factorial moments — switching from moments to factorial moments can be seen as a mere change of basis. We now have enough background to state the theorem behind the method of moments:

Theorem 3.1 (Method of moments, [16])**.** *Let X be a random variable with a distribution that is determined by its moments. If $X_1, X_2, \ldots$ are random variables with finite moments such that $\mathbb{E}(X_n)_k \to \mathbb{E}(X)_k$ as $n \to \infty$ for every integer $k \geq 1$, then $X_n \xrightarrow{d} X$.*

In particular, for X_n integer valued random variables, this theorem can also be stated in the less general version that we will be using:

Theorem 3.2 ([25])**.** *Let $\lambda = \lambda(n)$ be non-negative and bounded. For all $n \geq 1$, let X_n be a non-negative integer, bounded random variable such that for all $k \geq 0$,*

$$\mathbb{E}(X_n)_k = \lambda(n)^k + o(1).$$

Then for all $i \geq 0$,

$$\mathbb{P}(X_n = i) = e^{-\lambda(n)} \frac{\lambda(n)^i}{i!} + o(1).$$

For λ fixed this implies that $X \xrightarrow{d} Po(\lambda)$.

3.2 The expected number of i-cycles in G_n^*

To make things easier to read, we start by giving the expected number of i-cycles in G_n^* before giving the higher moments.

Let Y_i denote the number of i-cycles in a graph $G \in G_n^*$. Let $(u, v)_c$ denote the edge from u to v in an initial cycle c. Note that, in G_n^*, a cycle is determined by a sequence of vertices, edges and vertex identifications denoting a change of initial cycles, i.e.

$$(v_1)_c, (v_1 v_2)_c, (v_2)_c, \ldots, (v_{i-1} v_i)_c, (v_i)_c, (v_i)_c \longleftrightarrow (u_1)_d, (u_1)_d, \ldots.$$

Define a random variable as follows

$$X_{c,v}^{i,j} = \begin{cases} 1 & (v, v+1)_c \text{ is an edge in an } i\text{-cycle consisting of } j \\ & \text{edge-disjoint paths in initial cycles and } (v-1, v) \text{ is not an edge in this } i\text{-cycle} \\ 0 & \text{otherwise.} \end{cases}$$

3.2.1 2-cycles in G_n^*

Recall that a graph in G_n^* is not necessarily simple. We will start by counting the expected number of double edges, or 2-cycles, $\mathbb{E}Y_2$. This calculation will be very detailed. The calculations of larger cycles will be in less depth.

Using the definition of $X_{c,v}^{i,j}$, we get

$$\begin{aligned}\mathbb{E}(Y_2) &= \frac{1}{2}\mathbb{E}\sum_{v,c} X_{c,v}^{2,2} \\ &= \frac{1}{2}n(2n-2)\mathbb{E}X_{1,1}^{2,2},\end{aligned}$$

because there are n initial cycles in the model, each having $2n-2$ vertices, and because all vertices are equivalent. Also, note that we count each 2-cycle twice. As $X_{c,v}^{i,j}$ is an indicator variable,

$$\begin{aligned}\mathbb{E}(Y_2) &= n(n-1)\mathbb{P}(X_{1,1}^{2,2} = 1) \\ &= n(n-1)\sum_{c,j}\mathbb{P}(X_{1,1}^{2,2} = 1 \wedge \phi_1(1) = (c,j)).\end{aligned}$$

Here, we are just using the law of total probability and summing over all possible values of $\phi_1(1)$. Because ϕ_i assigns each value uniformly at random, for all i, it follows that

$$\mathbb{E}(Y_2) = n(n-1)(2n-2)\mathbb{P}(X_{1,1}^{2,2} = 1 \wedge \phi_1(1) = (2,1)).$$

Having set the beginning vertex and its neighbouring cycle, we can thus proceed to define the vertex in the neighbouring cycle.

$$\begin{aligned}\mathbb{E}(Y_2) &= n(n-1)(2n-2)\sum_{i=1}^{2n-2}\mathbb{P}(X_{1,1}^{2,2} = 1 \wedge \phi_1(1) = (2,1) \wedge \phi_2(i) = (1,1)) \\ &= n(n-1)(2n-2)(2n-2)\mathbb{P}(X_{1,1}^{2,2} = 1 \wedge \phi_1(1) = (2,1) \wedge \phi_2(1) = (1,1)),\end{aligned}$$

which again follows because all vertices are equivalent.

Next, after deciding that vertex 1 from cycle 1 will be identified with vertex 1 from cycle 2, and because we are only considering the directed edges away from vertex 1, to consider the double edge we must distinguish between two cases: Either it will form a directed cycle, $(12)_1, 2_1 \longleftrightarrow (2n-2)_2, (2n-2,1)_2, (1)_2 \longleftrightarrow (1)_1$, or it will not, i.e. $(2)_1 \longleftrightarrow (2)_2$. So

$$\begin{aligned}\mathbb{E}(Y_2) &= 2^2 n(n-1)^3 \mathbb{P}(\phi_1(1)=(2,1) \wedge \phi_2(1)=(1,1) \wedge \phi_1(2)=(2,2) \wedge (\phi_2(2)=(1,2) \vee \phi_2(2n-2)=(1,2))) \\ &= 2^3 n(n-1)^3 \mathbb{P}(\phi_1(1)=(2,1) \wedge \phi_2(1)=(1,1) \wedge \phi_1(2)=(2,2) \wedge \phi_2(2)=(1,2)) \\ &= 2^3 n(n-1)^3 (\mathbb{P}(\phi_1(1)=(2,1) \wedge \phi_1(2)=(2,2)))^2\end{aligned}$$

where the last equality follows because all cycles are independent and equivalent. The rest is then basic probability theory:

$$\begin{aligned}\mathbb{E}(Y_2) &= 2^3 n(n-1)^3 (\mathbb{P}(\phi_1(1)=(2,1) \mid \phi_1(2)=(2,2)) \mathbb{P}(\phi_1(2)=(2,2)))^2 \\ &= 2^3 n(n-1)^3 \left(\frac{1}{2n-3}\frac{1}{2n-2}\right)^2 \\ &\sim \frac{1}{2}\end{aligned}$$

Note here that it also follows that $\mathbb{E}(X_{c,v}^{2,2}) \sim \frac{1}{2}n^{-2}$.

3.2.2 i-cycles in G_n^*

Using the same procedures as for computing $\mathbb{E}(Y_2)$, we can easily compute $\mathbb{E}(Y_i)$. Because the assignment of vertices to each other is chosen uniformly at random, for each vertex, $\mathbb{P}(X_{c,v}^{i,j}=1) = \mathbb{P}(X_{c,v}^{i',j}=1)$, for $i, i' \geq j$, and thus

$$\mathbb{P}(X_{c,v}^{i,j}=1) = \binom{i-1}{j-1}\mathbb{P}(X_{c,v}^{j,j})$$

where $\binom{i-1}{j-1}$ is due to the choices there are of placing different length cycles.

Any i-cycle can contain edges of between 2 and i initial cycles.

For Y_i, the number of i-cycles in the graph, it holds that

$$\mathbb{E}(Y_i) = \mathbb{E}\sum_{j=2}^{i}\frac{1}{j}\sum_{c,v} X_{c,v}^{j,j} = n(2n-2)\sum_{j=2}^{i}\frac{1}{j}\binom{i-1}{j-1}\mathbb{P}(X_{1,1}^{j,j}=1). \tag{3.1}$$

What is the probability that a given edge $(v, v+1)$ will be part of a path in an i-cycle on i-paths, i.e. $\mathbb{P}(X_{c,v}^{i,i}=1)$? We must distinguish between two cases here: the i paths come from i different initial cycles, or at least one initial cycle is used more than once.

$$\begin{aligned}&\mathbb{P}(X_{1,1}^{i,i} = 1 \wedge \text{all paths from different initial cycle}) \\ &\qquad = (n-1)_{i-1}(2n-2)^{i-1}2^{i-1}2^i(2n-2)^{-i}(2n-3)^{-i} \\ &\qquad \sim 2^{i-2}n^{-2},\end{aligned}$$

where the first two terms of the product come from the different ways the cycles, vertices and directions can be chosen, next come the possible labellings the different ϕs may give, while the last two terms come from the probabilities that the different ϕs will choose the correct cycle.

Next let us consider the case where there exist paths that use the same initial cycle. Define the event A_j as the existence of j paths that use some initial cycle that is already used by a previous path. Here, previous means closer to $(1)_1$ in the direction of the cycle going through arc $(1,2)_1$.

$$\begin{aligned}&\mathbb{P}(X_{1,1}^{i,i} = 1 \wedge A_j)\\ &\quad \le (n-1)_{i-j-1}(2n-2)^{i-1}2^{i-1}(2n-2)^{-i+j}(2n-3)^{-i+j}(2n-2-j-1)^{-2j}2^i c(i,j)\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \sim n^{-2-j}2^{i-2}c(i,j) = o(n^{-2})\end{aligned}$$

where $c(i,j)$ is a constant.

Thus, from equation (3.1) and some algebra, it follows that

$$\begin{aligned}\mathbb{E}(Y_i) &= 2n(n-1)\left(\sum_{j=3}^{i}\frac{1}{j}\binom{i-1}{j-1}2^{j-2}n^{-2} + \frac{1}{2}\binom{i-1}{1}\frac{1}{2}n^{-2}\right)(1+o(1))\\ &= \frac{3^i - i^2 - i - 1}{2i}(1+o(1)).\end{aligned}$$

3.3 The distribution of short cycles in the pairing model

Before things get somewhat complicated and we start working out higher moments of short cycles in G_n^*, let us consider the distribution of short cycles in the pairing model.

Let $\tilde{Y}_i$ be the number of i-cycles in the pairing model, $\mathcal{P}_{n,4}$. The expected value of $\tilde{Y}_i$ can be calculated as follows:

$$\mathbb{E}(\tilde{Y}_i) \sim (n)_i \frac{1}{2i}\left(\frac{3}{n}\right)^i \sim \frac{3^i}{2i},$$

where we are again considering all $(n)_i$ possible directed cycles, we take away the direction and then calculate possibilities of the corresponding edges being present in the pairing model.

What about higher moments? Let $\tilde{Z}^i_j$ denote the jth possible i-cycle in a graph in $\mathcal{P}_{n,4}$. Let

$$\begin{aligned}\mathcal{J}_l := \{(j_1,\dots,j_k) \mid\ & \tilde{Z}^i_1,\dots,\tilde{Z}^i_k \text{ all distinct and exactly } l \text{ vertices of the} \\ & \tilde{Z}^i_{j_r}\text{s are used already by a cycle of lower index}\}\end{aligned}$$

for $l = 0,\dots,ik$. Clearly,

$$\mathbb{E}[\tilde{Y}]_i = \sum_{j_1,\dots j_k \in \bigcup \mathcal{J}_l} \tilde{Z}^i_1 \cdots \tilde{Z}^i_k.$$

First let us consider $\mathcal{J}_0$. As these cycles do not share any edge, they are independent of each other, so we have

$$\sum_{j_1,\dots j_k \in \mathcal{J}_0} \tilde{Z}^i_1 \cdots \tilde{Z}^i_k \sim (n)_{ik} \left(\frac{1}{2i}\right)^k \left(\frac{3}{n}\right)^{ik} \sim \left(\frac{3^i}{2i}\right)^k.$$

Next, we will consider $\mathcal{J}_j$, for some $1 \le j \le ik$, such that $\mathcal{J}_j$ is non-empty. Note here that, if a cycle shares several vertices and some edges with another cycle, it will always share at least one more vertex than edges. Also note that if all vertices and edges of a cycle are present in already existing cycles (i.e. cycles of lower index) then these cycles must share more vertices than edges. This means that, if we let m denote the number of edges shared, $m < l$. It thus follows that

$$\begin{aligned}\sum_{j_1,\cdots j_k \in \mathcal{J}_l} \tilde{Z}^i_1 \cdots \tilde{Z}^i_k &\le \sum_{m=0}^{l-1} (n)_{ik-l} \left(\frac{3}{n}\right)^{ik-m} \left(\frac{1}{2i}\right)^k c(l,m,k) \\ &\le (n)_{ik-l} \left(\frac{3}{n}\right)^{ik-l+1} \left(\frac{1}{2i}\right)^k c'(l,m,k) \\ &= o(1),\end{aligned}$$

where $c(l,m,k)$ and $c'(l,m,k)$ are constants counting the different possible constellations of cycles. Because there are less than ik $\mathcal{J}_r$s to be counted, it follows that $\mathbb{E}[\tilde{Y}_i]_k = (\mathbb{E}\tilde{Y}_i)^k + o(1)$, so with the method of moments, it follows that in $\mathcal{P}_{n,4}$, the number of i-cycles is asymptotically Poisson distributed with mean $\frac{3^i}{2i}$.

3.4 The distribution of short cycles in G^*_n

To find out how short cycles are distributed in G^*_n, we will use the method of moments again, (see Section 3.1.) so we will need to calculate higher moments of the variables Y_i. We will start with the higher moments of 2-cycles and then move to general i-cycles.

3.4.1 The distribution of double edges in G_n^*

We start with $\mathbb{E}[Y_2]_k$, the kth factorial moment of double edges. Let Z_j, $j = 1, 2, \ldots$ denote the jth 2-cycle in G. Let

$$\mathcal{J} := \{(i_1, \ldots, i_k) \mid i_1, \ldots, i_k \text{ all distinct and } Z_{i_1}, \ldots, Z_{i_k} \text{ do not share an initial cycle}\}$$

and

$$\mathcal{J}' := \{(i_1, \ldots, i_k) \mid i_1, \ldots, i_k \text{ all distinct and at least two of the } Z_{i_1}, \ldots, Z_{i_k} \text{ share an initial cycle}\}.$$

Let

$$\begin{aligned}
\mathbb{E}[Y_2]_k &= \mathbb{E} \sum_{j_1,\ldots,j_k \in \mathcal{J} \cup \mathcal{J}'} Z_{j_1} Z_{j_2} \cdots Z_{j_k} \\
&= \mathbb{E} \sum_{j_1,\ldots,j_k \in \mathcal{J}} Z_{j_1} Z_{j_2} \cdots Z_{j_k} + \mathbb{E} \sum_{j_1,\ldots,j_k \in \mathcal{J}'} Z_{j_1} Z_{j_2} \cdots Z_{j_k} \qquad (3.2)
\end{aligned}$$

The first part of the sum in (3.2) can be calculated as follows:

$$\begin{aligned}
\sum_{j_1,\ldots,j_k \in \mathcal{J}} \mathbb{E} Z_{j_1} Z_{j_2} \cdots Z_{j_k} &= (n)_{2k} 2^{-k} (2n-2)^{2k} 2^k (2n-2)^{-2k} (2n-3)^{-2k} 2^k \qquad (3.3) \\
&= 2^{-k} + o(1).
\end{aligned}$$

Here, (3.3) is explained as follows: there are $(n)_k 2^{-k}$ ways to choose k pairs of initial cycles that form the Z_js, the factor 2^{-k} accounts for the fact that these pairs are unordered. Each of these $2k$ cycles has $2n-2$ edges to choose from, and for each pair of edges there are two possible orientations of the edges towards each other to be found (i.e. if the edges form a directed cycle or not). Finally, the probability that the chosen vertices will pair up is $(2n-2)^{-2k}(2n-3)^{-2k}2^k$, where the factor 2^k comes from the two choices of labellings ($\phi_c[2]$) for each pair of cycles.

Consider the second part of (3.2). Define

$$\begin{aligned}
\mathcal{J}_l' :=& \{(i_1, \ldots, i_k) \mid i_1, \ldots, i_k \text{ all distinct and } l \text{ edges of the } Z_{i_1}, \ldots, Z_{i_k} \\
& \text{are in an initial cycle that is used by other cycle with lower index}\}.
\end{aligned}$$

Clearly,

$$\mathbb{E} \sum_{j_1,\ldots,j_k \in \mathcal{J}'} Z_{j_1} Z_{j_2} \cdots Z_{j_k} = \mathbb{E} \sum_{l=2}^{k} \sum_{j_1,\ldots,j_k \in \mathcal{J}_l'} Z_{j_1} Z_{j_2} \cdots Z_{j_k}$$

Consider each $\mathcal{J}_l'$ separately:

$$\mathbb{E} \sum_{j_1,\ldots,j_k \in \mathcal{J}_l'} Z_{j_1} Z_{j_2} \cdots Z_{j_k} \leq (n)_{2k-l} 2^{-k+l} (2n-2)^{2k} 2^k (2n-3-l)^{-4k} 2^k \qquad (3.4)$$
$$\sim 2^{-k+l} n^{-l} = o(1).$$

Here, the $(n)_{2k-l} 2^{-k+l}$ in (3.4) comes from the different possible ways of choosing the initial cycles involved, while for each pair of cycles there are less than $2(2n-2)^2$ ways of choosing the edges involved times their orientation towards each other. The probability of all these vertex identifications taking place in this way is less than $(2n-3-l)^{-4k} 2^k$.

Thus, with the method of moments, it follows that asymptotically the number of double edges in G_n^* have a Poisson distribution with parameter $\frac{1}{2}$.

3.4.2 Higher moments and distribution of larger cycles in G_n^*

Let us next consider the higher moments of larger cycles, i.e. $[Y_i]_k$, $i > 2$. Let Z_j^i denote the indicator function of jth i-cycle in the graph. As before, set

$$\mathcal{J} := \{(j_1, \ldots, j_k) \mid j_1, \ldots, j_k \text{ all distinct and } Z_{j_1}^i, \ldots, Z_{j_k}^i \text{ do not share an initial cycle}\}$$

and

$$\begin{aligned}\mathcal{J}' := \{&(j_1, \ldots, j_k) \mid j_1, \ldots, j_k \text{ all distinct and at least two of the} \\ &Z_{j_1}^i, \ldots, Z_{j_k}^i \text{ share an initial cycle}\}.\end{aligned}$$

Again, we have

$$\mathbb{E}[Y_i]_k = \sum_{j_1,\ldots,j_k \in \mathcal{J} \cup \mathcal{J}'} \mathbb{E} Z_{j_1}^i Z_{j_2}^i \cdots Z_{j_k}^i =$$
$$\mathbb{E} \sum_{j_1,\ldots,j_k \in \mathcal{J}} Z_{j_1}^i Z_{j_2}^i \cdots Z_{j_k}^i + \mathbb{E} \sum_{j_1,\ldots,j_k \in \mathcal{J}'} Z_{j_1}^i Z_{j_2}^i \cdots Z_{j_k}^i$$

Now, note first that

$$\begin{aligned}\mathbb{E} \sum_{j_1,\ldots,j_k \in \mathcal{J}} Z_{j_1}^i Z_{j_2}^i \cdots Z_{j_k}^i &= \sum_{j_1,\ldots,j_k \in \mathcal{J}} \mathbb{P}(Z_{j_1}^i = 1 \wedge Z_{j_2}^i = 1 \wedge \ldots Z_{j_k}^i = 1) \\ &= \sum_{j_1,\ldots,j_k \in \mathcal{J}} \mathbb{P}(Z_{j_1}^i = 1) \mathbb{P}(Z_{j_2}^i = 1) \cdots \mathbb{P}(Z_{j_k}^i = 1)\end{aligned}$$

where the last equality follows because the i-cycles, in this case, are completely independent. Recall that for i-cycles, $i \geq 3$, it holds that

$$\begin{aligned}\mathbb{E}(Y_i) &= 2n(n-1)\left(\sum_{j=3}^{i}\frac{1}{j}\binom{i-1}{j-1}2^{j-2}n^{-2} + \frac{1}{2}\binom{i-1}{1}\frac{1}{2}n^{-2}\right)(1+o(1)) \\ &= \frac{3^i - i^2 - i - 1}{2i}(1+o(1)).\end{aligned}$$

Because $(n)_i \sim (n-r)_i$, $r \leq ik$, it follows that

$$\mathbb{E}\sum_{j_1,\dots,j_k\in\mathcal{J}} Z^i_{j_1}Z^i_{j_2}\cdots Z^i_{j_k} = \left(\frac{3^i - i^2 - i - 1}{2i}\right)^k (1+o(1)).$$

Next, let us consider what happens with $\mathcal{J}'$. Define $\mathcal{J}'_l$ as for 2-cycles. Let $R_k := \{2,3,\dots i\}^k$. For a vector $\mathbf{r_k} \in R_k$, let $\mathbf{r_k}(j)$ denote the jth entry in $\mathbf{r_k}$. Let $B(\mathbf{r_k})$ denote the event that, for given k cycles $1,2,\dots,k$, cycle j uses exactly $\mathbf{r_k}(j)$ initial cycles. It will be clear from context which k cycles are meant. Let $|\mathbf{r_k}| := \sum_{j=1}^{k}\mathbf{r_k}(j)$. We are now set to consider the second part of the expected value of $[Y_i]_k$.

$$\mathbb{E}\sum_{j_1,\dots,j_k\in\mathcal{J}'} Z^i_{j_1}Z^i_{j_2}\cdots Z^i_{j_k} = \sum_{l=1}^{ik}\mathbb{E}\sum_{j_1,\dots,j_k\in\mathcal{J}'_l} Z^i_{j_1}Z^i_{j_2}\cdots Z^i_{j_k}$$

Consider each $\mathcal{J}'_l \neq \varnothing$ separately:

$$\begin{aligned}\mathbb{E}\sum_{j_1,\dots,j_k\in\mathcal{J}'_l} Z^i_{j_1}Z^i_{j_2}\cdots Z^i_{j_k} &= \sum_{j_1,\dots,j_k\in\mathcal{J}'_l}\mathbb{P}(Z^i_{j_1}Z^i_{j_2}\cdots Z^i_{j_k}=1) \\ &= \sum_{\mathbf{r_k}\in R_k}\sum_{j_1,\dots,j_k\in\mathcal{J}'_l}\mathbb{P}(Z^i_{j_1}Z^i_{j_2}\cdots Z^i_{j_k}=1\cap B(\mathbf{r_k})) \\ &\leq \sum_{\mathbf{r_k}\in R_k}(n)_{|\mathbf{r_k}|-l}(2n-2)^{|\mathbf{r_k}|}2^{2|\mathbf{r_k}|}c(\mathbf{r_k})(2n-2-l)^{-2|\mathbf{r_k}|+l-1} \qquad (3.5) \\ &= o(1),\end{aligned}$$

where (3.5) can be explained as follows: In sum, $|\mathbf{r_k}| - l$ cycles are chosen, and for each cycle counted there will be at most $2n-2$ vertices to be chosen from as "beginning vertices" in the initial cycles. There will also be 2 labellings possible each, and 2 directions the cycle may take from these vertices. $c(\mathbf{r_k})$ is a constant that counts the different possible path lengths there are of these cycles intersecting initial cycles, i.e.

$$c(\mathbf{r_k}) = \frac{1}{|\mathbf{r_k}(j)|}\prod_{j=1}^{k}\binom{i-1}{\mathbf{r_k}(j)-1}.$$

The last term is a crude estimate of the probabilities involved: Even if some of the paths of some cycles are identical on some initial cycles, the overlap must stop at some point. As R_k is finite, it follows that this expression is $o(1)$.

We can thus conclude, with the method of moments, that the distribution of i-cycles in G_n^* tends towards a Poisson law with mean $\frac{3^i-i^2-i-1}{2i}$.

3.5 Expected number of i-cycles in $G_n^{\mathcal{A}}$

The calculations for the short cycle count in $G_n^{\mathcal{A}}$ follow the calculations for G_n^* closely. The main difference here is that $G_n^{\mathcal{A}}$ cannot have any double edge, and for $i < n-1$, it can also not have an i-cycle on only two initial cycles. Thus, we start at $i = 3$. Let us again call the number of i-cycles Y_i, as well as define $X_{c,v}^{i,j}$ as in Section 3.2.

Again with the same reasoning, we have that

$$\begin{aligned}\mathbb{E}(Y_i) &= \mathbb{E}\sum_{j=3}^{i}\frac{1}{j}\sum_{c,v}X_{c,v}^{i,j}\\ &= n(2n-2)\sum_{j=3}^{i}\frac{1}{j}\binom{i-1}{j-1}\mathbb{P}(X_{1,1}^{j,j}=1).\end{aligned} \tag{3.6}$$

We can now proceed to calculate $\mathbb{P}(X_{1,1}^{j,j}=1)$. Again, consider two cases, either all j edges are in different initial cycles, or they are not. Let A_l denote the event that l edges use an initial cycle that was used before, where before means closer to vertex $(1)_1$ in the path. First, we have that

$$\begin{aligned}&\mathbb{P}_{G_n^{\mathcal{A}}}(X_{1,1}^{j,j}=1\wedge \text{all paths are from different initial cycles})\\ &\qquad = (n-1)_{j-1}(2n-2)^{j-1}2^{j-1}2^j(n-1)^{-j}2^{-j}(n-2)^{-j}2^{-j}\\ &\qquad \sim 2^{j-2}n^{-2},\end{aligned} \tag{3.7}$$

where the first term counts the number of possible initial cycles, the next term counts the number of vertices to choose from, then come the number of directions to go and the possible initial label of each vertex. Finally, we divide by the probabilities of ϕ^f assigning these vertices to each other.

On the other hand, for a constant $c(j,l)$, we have

$$\begin{aligned}\mathbb{P}(X_{1,1}^{j,j}=1\wedge A_l) \leq&(n-1)_{j-l-1}(2n-2)^{j-1}2^{j-1}2^j\\ &(n-1)^{-j+l}2^{-j+l}(n-2)^{-j+l}2^{-j+l}(n-1-j-)^{-2l}2^{-2l}2^ic(j,l)\\ =&o(n^{-2}).\end{aligned}$$

Note that the value of (3.7) dominates this expression, and also that $\mathbb{P}_{G_n^{\mathcal{A}}}(X_{1,1}^{i,i} = \mathbb{P}_{G_n^*}(X_{1,1}^{i,i})$.

Thus, from equation (3.6), we get that

$$\begin{aligned}\mathbb{E}(Y_i) &= n(2n-2)\sum_{j=3}^{i}\frac{1}{j}\binom{i-1}{j-1}(2^{j-2}n^{-2}+o(n^{-2})) \\ &= \frac{3^i-2i^2-1}{2i}(1+o(1))\end{aligned}$$

As the calculations for the expected number of i-cycles in $G_n^{\mathcal{A}}$ were extremely similar to the expected number of i-cycles in G_n^*, the calculations for higher moments of i-cycles in $G_n^{\mathcal{A}}$ should also be similar, as should be the result. In future research, if necessary, it would probably be easy to prove that the number of i-cycles in $G_n^{\mathcal{A}}$ also asymptotically follows a Poisson distribution, while there seems to be little need to do so at present.

Chapter 4

Contiguity

When given different random graph models on the same sequence of spaces of graphs (the space depends on n), a question of great interest is that of contiguity. Informally, contiguity means that if almost any graph in one space has a given property, almost every graph in the other space will also have this property.

In this section, we will briefly explain what contiguity is and give some examples of contiguity. We will state the following conjecture, that G_n^* and $G_{n,4}$ might be contiguous. This would be a wonderful result, as it would imply that almost all graphs in G_n^* are 3-colourable! (See Theorem 1.5.) Finally, we will explain why we think this conjecture might hold, citing a theorem with which contiguity can be proven.

4.1 Definition and examples of contiguity

Let us begin by stating formally what contiguity means, see e.g. [16, 25]:

Definition 4.1. Given two sequences of probability spaces $(P_n, \Omega_n, \mathcal{F}_n)$ and $(Q_n, \Omega_n, \mathcal{F}_n)$, $n \geq 1$ that are defined on the same measurable space $(\Omega_n, \mathcal{F}_n)$, we call two sequences *contiguous* if for every event $\mathcal{A}$

$$\mathbb{P}_{P_n}(\mathcal{A}) = 1 \qquad a.a.s. \qquad \Longleftrightarrow \qquad \mathbb{P}_{Q_n}(\mathcal{A}) = 1 \qquad a.a.s.$$

We then write $(P_n, \Omega_n, \mathcal{F}_n) \approx (Q_n, \Omega_n, \mathcal{F}_n)$.

Note that contiguity defines an equivalence relation.

To clarify this concept, let us give some examples [25] and explain what they mean. For two probability spaces of random graphs on the same vertex set, $\mathcal{G}$ and $\mathcal{G}'$, define the *sum* $\mathcal{G}+\mathcal{G}' := \mathcal{G} \cup \mathcal{G}'$. As this gives a multigraph, define the *graph-restricted sum* $\mathcal{G} \oplus \mathcal{G}' := \mathcal{G} \cup \mathcal{G}'$

conditioned on the resulting graph being simple. Note that $\mathcal{G} \oplus \mathcal{G}'$ is only defined if there exists at least one such simple graph. Let $\mathcal{H}_n$ denote a uniformly random Hamilton cycle. The following theorem holds:

Theorem 4.1 ([25]). *It holds that*

$$G_{n,2} \oplus \mathcal{H}_n \approx G_{n,4}$$

and

$$\mathcal{H}_n \oplus \mathcal{H}_n \approx G_{n,4}.$$

What does this mean? This means that, *a.a.s.*, any graph in $G_{n,4}$ decomposes into a random 2-regular graph and a Hamiltonian cycle, but also that *a.a.s.*, any graph in $G_{n,4}$ decomposes into two Hamiltonian cycles. This immediately implies that for n even, any graph in $G_{n,4}$ is *a.a.s.* 4-edge colourable.

In this line of thought, a big goal would be to prove that $G_n^* \approx G_{n(n-1),4}$. This would mean that anything that applies *a.a.s.* to $G_{n,4}$ also applies to G_n^*. In particular, G_n^* would be *a.a.s.* 3-colourable. (See section 1.4.2.) As I will explain in the next section, there is reason to believe that this might hold. Thus, we have the following

Conjecture 4.1. $G_{n(n-1),4} \approx G_n^*$.

4.2 Implications of Contiguity

Showing that G_n^* and$G_{n(n-1),4}$ are contiguous would facilitate the proof of many open problems mentioned here; in fact, make them redundant. If G_n^* and $G_{n(n-1),4}$ were proven to be contiguous, results about properties that hold *a.a.s.* for $G_{n,4}$ would hold also for G_n^*. In particular, we would have

- an easier proof that G_n^* is *a.a.s.* 4-connected, (See Theorem 1.3 and also the next chapter, Chapter 5.)
- the result that the diameter of a graph $G \in G_n^*$ is also *a.a.s.* $O(\log(n))$, (See Theorem 1.4.), and, most importantly
- we would know that *a.a.s.* a graph $G \in G_n^*$ is *a.a.s.* 3-colourable. (See Theorem 1.5.)

The last point would be of great interest, especially considering Conjecture 1.1.

4.3 The small subgraph conditioning method and a theorem for contiguity

Clearly, it would be wonderful to prove contiguity. There exist methods of doing this, though they are not necessarily straightforward. One of them uses the so-called *small subgraph conditioning method*.

This section follows [25]. In some cases, for a random graph process $\mathcal{G} = \mathcal{G}(n)$, we wish to count the expected value of some random variable $Y = Y(n)$. For example, Y could be the number of perfect matchings on $G \in \mathcal{G}(n)$, or the number of lock decompositions. Let σ_Y denote the variance of Y. We want to show that $Y > 0$ *a.a.s.*. The problem occurs when $\mathbb{E}(Y) = \Theta(\sigma_Y)$, so we cannot use Chebyshev's inequality (Theorem 1.1) to show this.

The method to circumvent this problem is, in some cases, to realize that Y depends strongly on some small and not too common subgraphs that do occur — short cycles, mainly — and that conditioning on these small subgraphs affects $\mathbb{E}(Y)$ and also significantly reduces the variance of Y. This method can then be used to show that two models are contiguous.

We need some more definitions before we can state the theorem. Staying in the space $\mathcal{G}$, assume that $\mathbb{E}(Y) \geq 0$. Let $\mathbb{P}_{\mathcal{G}}(\mathcal{A})$ denote the probability of the event $\mathcal{A}$ in this space. Let us define a new space, $\mathcal{G}^{(Y)}$, such that the probability of a graph $G \in \mathcal{G}^{(Y)}$ occurring is

$$\mathbb{P}_{\mathcal{G}^{(Y)}}(G) = \mathbb{P}_{\mathcal{G}}(G)\frac{Y(G)}{\mathbb{E}(Y)}.$$

Note that for an event $\mathcal{A}$ it then holds that

$$\begin{aligned}
\mathbb{P}_{\mathcal{G}^{(Y)}}(\mathcal{A}) &= \sum_{\substack{G\in\mathcal{G}\\ \mathcal{A}\text{ occurs in }G}} \frac{\mathbb{P}_{\mathcal{G}}(G)Y(G)}{\mathbb{E}(Y)} \\
&= \sum_{\substack{G\in\mathcal{G}\\ Y(G)>0}} \frac{\mathbb{P}_{\mathcal{G}}(G)Y(G)I_{\mathcal{A}\in G}}{\mathbb{E}(Y)} \\
&= \frac{\mathbb{E}_{\mathcal{G}}(I_{\mathcal{A}} \wedge Y)}{\mathbb{E}(Y)},
\end{aligned}$$

where $I_{\mathcal{A}}$ denotes the indicator function of the event $\mathcal{A}$.

To see a connection to G_n^*, note that G_n^* is a model where every lock decomposition appears with equal probability. For a given graph G, we let Y denote the number of lock decompositions. It holds that $\mathbb{P}_{G_n^*}(G) = \frac{Y(G)}{|G_n^*|}$. It also holds that

$$\begin{aligned} \mathbb{P}_{G^{(Y)}_{n(n-1),4}}(G) &= \mathbb{P}_{G_n^*}(G)\frac{Y(G)}{\mathbb{E}(Y)} \\ &= \frac{1}{|G_{n(n-1),4}|}\frac{Y(G)}{\frac{\sum_G Y(G)}{|G_{n(n-1),4}|}} \\ &= \frac{Y(G)}{|G_n^*|}, \end{aligned}$$

so G_n^* is the same as the space $G^{(Y)}_{n(n-1),4}$.

We have the following theorem:

Theorem 4.2 (Ad verbatim from [25])**.** *Let $\lambda_i > 0$ and $\delta_i \geq -1$, $i = 1, 2, \ldots,$ be real numbers and suppose that for each n there are random variables $X_i = X_i(n)$, $i = 1, 2, \ldots,$ and $Y = Y(n)$ defined on the same probability space $\mathcal{G} = \mathcal{G}(n)$ such that X_i is non-negative integer valued, Y is non-negative and $\mathbb{E}Y > 0$ (for n sufficiently large). Suppose furthermore that*

1. *For each $k \geq 1$ X_i, $i = 1, 2, \ldots, k$ are asymptotically independent Poisson random variables with $\mathbb{E}X_i \to \lambda_i$;*

2. *For every finite sequence $j_1, \ldots, j_k$ of non-negative integers:*

$$\frac{\mathbb{E}(Y(X_1)_{j_1}\ldots(X_k)_{j_k})}{\mathbb{E}Y} \to \prod_{i=1}^{k}(\lambda_i(1+\delta_i))^{j_i}$$

3. *$\sum_i \lambda_i \delta_i^2 < \infty$;*

4. *$\frac{\mathbb{E}Y_n^2}{(\mathbb{E}Y_n)^2} \leq \exp\left(\sum_i \lambda_i\delta_i^2\right) + o(1)$ as $n \to \infty$.*

Then

$$\mathbb{P}(Y_n > 0) = \exp\left(-\sum_{\delta_i=-1}\lambda_i\right) + o(1)$$

and, provided $\sum_{\delta_i=-1}\lambda_i < \infty$,

$$\bar{\mathcal{G}} \approx \bar{\mathcal{G}}^{(Y)}$$

where $\bar{\mathcal{G}}$ is the probability space obtained from $\mathcal{G}$ by conditioning on then event $\wedge_{\delta_i=-1}(X_i = 0)$.

With this theorem, we can work towards proving Conjecture 4.1. Let Y again be the number of lock decompositions, and X_i be the number of i-cycles in a graph $G \in G_{n(n-1),4}$. Let Z_i be the number of i-cycles in $G \in G_n^*$. Following the notation from Theorem 4.2, note that

1. The distribution of i-cycles in $G_{n(n-1),4}$ asymptotically follows a Poisson law, $\sim \mathrm{Po}(\lambda_i)$ (See Section 3.3), where $\lambda_i = \frac{3^i}{2i}$.

2. Let $W_{i_1}, W_{i_2}, \ldots$ denote the different subgraphs of $K_{n(n-1)}$ that are i-cycles. We get that

$$\begin{aligned}
\frac{\mathbb{E}Y(X_i)}{\mathbb{E}Y} &= \sum_{G \in G_{n,4}} \sum_{i_1, i_2, \ldots, i_j \text{ distinct}} \frac{I_{(W_{i_1} \in G)} I_{(W_{i_2} \in G)} \cdots I_{(W_{i_j} \in G)} Y(G) \mathbb{P}_{G_{n,4}}(G)}{\mathbb{E}(Y)} \\
&= \sum_{G \in G_{n,4}} \sum_{i_1, i_2, \ldots, i_j \text{ distinct}} I_{W_{i_1} \in G} I_{W_{i_2} \in G} \cdots I_{W_{i_j} \in G} \mathbb{P}_{G_{n,4}^{(Y)}}(G) \\
&= \mathbb{E}_{G_{n,4}^{(Y)}}(X_i)_j \\
&= \mathbb{E}_{G_n^*}(X_i)_j \\
&= \mathbb{E}(Z_i) \\
&= \frac{3^i - i^2 - i - 1}{2i}(1 + o(1)) \\
&\to \mu_i,
\end{aligned}$$

   where $I_{(\mathcal{H})}$ denotes the indicator function for a subgraph $\mathcal{H}$ being in a graph. The last three equations follow from Section 3.4. The same thing should not be too hard to show for combinations of different sizes of i-cycles.

3. Let $\delta_i = -1 + \mu_i/\lambda_i = \frac{-i^2-i-1}{3^i}$. The sum

$$\sum_{i \geq 3} \lambda_i \delta_i^2 = \sum_{i \geq 3} \frac{3^i}{2i} \left(\frac{i^2 + i + 1}{3^i} \right)^2 = \frac{335}{144} + 1/2(\log(3/2)) < \infty$$

   converges

4. We have calculated $\mathbb{E}Y$. $\mathbb{E}Y^2$ may or may not be fairly difficult to obtain and is a problem worthy of future research.

As three out of four conditions already hold, there is a good chance that Conjecture 4.1 is true.

Chapter 5

Connectivity of G_n^*

As shown in [2, 24], for $d \geq 3$, it holds that *a.a.s.* $G \in G_{n,d}$ is d-connected. In particular, this holds for 4-regular graphs. Proving that graphs in G_n^* are *not* *a.a.s.* 4-connected would show that G_n^* and $G_{n(n-1),4}$ are not contiguous, while proving that they are *a.a.s.* 4-connected gives another indication that Conjecture 4.1 holds. As we will see, the latter is the case.

Theorem 5.1. *Conditioning on having no double edges, a random graph in G_n^* is 4-connected a.a.s.*

Proof. Let $n \geq 3$. We will start by proving that deterministically, every graph in G_n^* is 2-connected.

Consider two vertices v_1 and v_2 in a graph $G \in G_n^*$, $v_1 \neq v_2$. Let v_1 be in initial cycles c_1 and d_1 and let v_2 be in initial cycles c_2 and d_2. We consider several cases:

1. $\{c_1, d_1\} \cap \{c_2, d_2\} \neq \varnothing$: Without loss of generality, let $c_1 = c_2$. In this case, v_1 and v_2 are in cycle c_1, thus there exist two vertex disjoint paths that connect v_1 and v_2.
2. $\{c_1, d_1\} \cap \{c_2, d_2\} = \varnothing$: There exist two vertices u_1 and u_2 that are the intersections of initial cycles c_1 and c_2. Thus, there exist vertex disjoint paths from v_1 to u_1 (in c_1), to v_2 (in c_2) and from v_1 to u_2 (in c_1) to v_2 (in c_2). (See Figure 5.1.)

Thus, by Menger's theorem (see, e.g. [6]), we have that for any $G \in G_n^*$ it holds that G is deterministically 2-connected.

Next, let us assume there is a graph $G \in G_n^*$ that is not 3-connected. It suffices to find a graph that has a 2-cut. Denote the vertices in the cut by v_1 and v_2. $G \setminus \{v_1, v_2\}$ has at least 2 components. Denote the largest of these components by G', and all the other

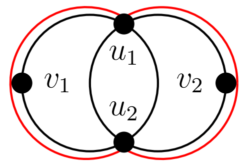


Figure 5.1: 2-connectivity. The red lines indicate the path chosen.

components together by G''. Thus $|G'| > |G''|$, i.e. $|G'| \geq n(n-1)/8 - 2 = \Theta(n^2)$. This means that there exist $\Theta(n)$ initial cycles at least partially in G'.

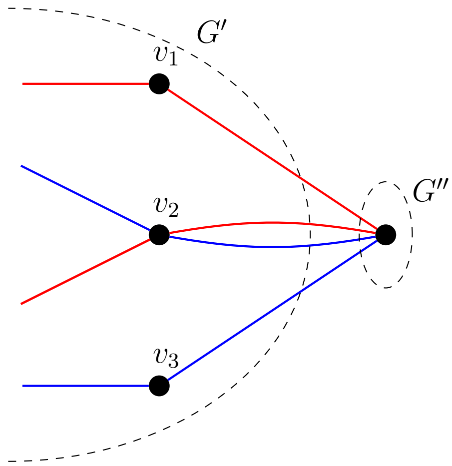
Now consider the initial cycles in G'': Assume there exists an initial cycle c that is completely in G''. Clearly, for every cycle c' at least partially in G' it holds that c and c' must have two common vertices. As c is completely in G'', it must hold that c' uses the cut to access c. However, then there are $\Theta(n)$ cycles in G' that must pass through c, while the cut only has size 2, and each vertex has degree 4. Thus, only two cycles can pass through the cut, a contradiction.

It follows that there can be no cycle c completely in G''. How many vertices can be in G''? We know that at most two cycles pass through the cut, so without loss of generality, let cycle 1 and cycle 2 be partially in G''. As these cycles already share v_1 and v_2 as common vertices, they can not intersect in G''. However, this means that any vertex in G'' would only be in one cycle, thus have degree 2, contradiction. The same argument holds if there is only one cycle partially in G''. Thus, G'' must be empty, so there can not exist a cut of size 2. We have shown that $G \in G_n^*$ is deterministically 3-connected.

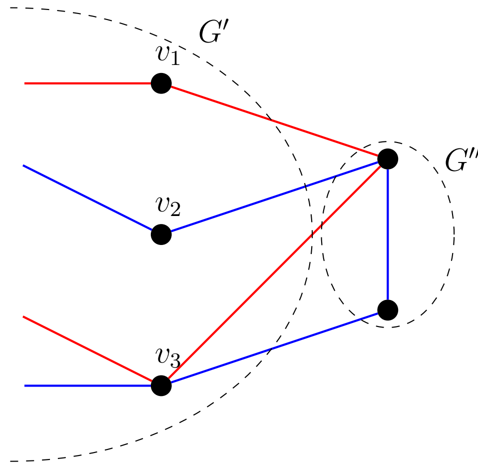
Assume now there exists a cut of size 3, on vertices v_1, v_2, v_3, that separates a connected component G' from a subgraph G'' such that G' is the largest component in the cut. By the same line of argument as for 2-cuts, there can only be at most three cycles entering G''. How big can G'' become here?

Note first that there cannot be only one cycle entering G'': any vertex in G'' would have degree 2. Next, note that there cannot be only two cycles entering G'': These cycles entering must share at least one vertex in $\{v_1, v_2, v_3\}$. If there is only one vertex in G'', both would contain it, yielding a double edge. If there is more than one vertex in G'', there exists a vertex that only one cycle contains, yielding a degree 2 vertex, which is impossible. See Figures 5.2 for these configurations.

Thus, if such a cut exists, there must be three cycles passing through it. For each v_i, there must be two different cycles passing through it. Note that if there were an edge between two vertices in $\{v_1, v_2, v_3\}$, only two cycles could enter G'', and so we would be in the previous case (see Figure 5.2). Thus, without loss of generality, cycles 1, 2 and 3 are partially in G''. We can distinguish, and possibly rule out, three different cases:



(a) If G'' has only one vertex, the graph is forced to have a double edge — impossible.



(b) If G'' has two vertices, the graph is forced to have a vertex of degree two — impossible.

Figure 5.2: Configurations of 3-cuts with two cycles—impossible

1. $|G''| = 1$: This would yield double edges, and so we can rule out this case.
2. $|G''| = 2$: Let $V(G'') = \{u_1, u_2\}$. There exists a configuration that is feasible in the model, namely that, without loss of generality, cycle 1 uses edges v_1u_1 and u_1v_2, cycle 2 uses edges v_1u_2 and u_2v_3 and cycle three uses edges v_2u_2, u_1u_2 and u_1v_3. See Figure 5.3. This is the only configuration possible without double edges such that u_1 and u_2 both have degree 4. Note at this point that edge u_1u_2 is in three triangles, with all three vertices. From the proof of the short cycle count, we know that this is highly unlikely. Let us call this configuration case 1. We will consider the calculations for this case this case below.
3. $|G''| = 3$: Denote the vertices in G'' by u_1, u_2, u_3. Note first that we cannot have more vertices in G''. Each pair of cycles has an intersection in the cut, and there can be only three more vertices in G'' where the cycles might intersect — if we would have more vertices, we would have a vertex of degree < 4.

   Next, note that each vertex in the cut contributes two half-edges that enter G'', so six edges in total. However, every vertex in G'' must have degree 4, so we need twelve half edges in total. Because we would be otherwise "missing" six half edges, all edges between vertices in $\{u_1, u_2, u_3\}$ must be present. Also, each v_i connects to two different vertices u_i, and each pair v_i, v_j shares exactly one common neighbour. We consider two sub-cases

   - For all $i \neq j$, each edge u_iu_j is in a different cycle. Without loss of generality,

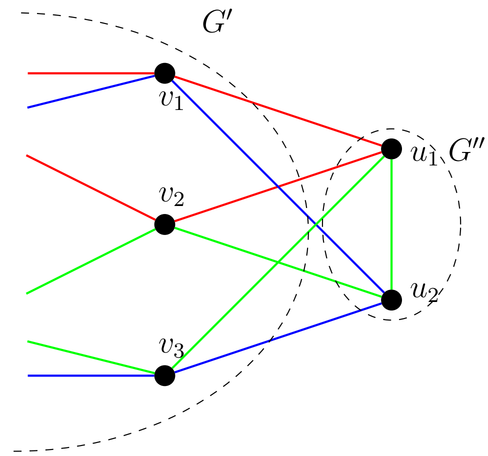
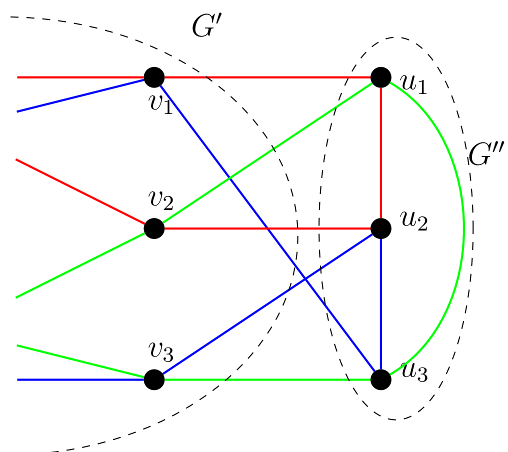
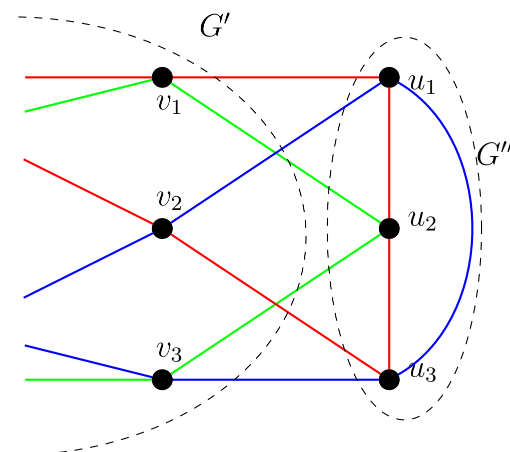


Figure 5.3: The only possible configuration with $|G''| = 2$ and three cycles

let v_1u_1, u_1u_2, v_2u_2 be in cycle 1, v_2u_1, u_1u_3, v_3u_3 in cycle 2 and v_1u_3, u_2u_3, u_2v_3 in cycle 3. (See Figure 5.4a)



(a) The first possible configuration with $|G''| = 3$ and three cycles



(b) The second possible configuration with $|G''| = 3$ and three cycles

Figure 5.4: The cases where three cycles pass through G''.

Note here that each edge is in a triangle, so there are two triangles per initial cycle. Recall from the proof of the short cycle count that this is unlikely. Denote this as case 2.

- The last case is that two of the edges are completely in G'', u_1u_2 and u_2u_3, say, are in the same initial cycle. Then, u_1u_3 is in another initial cycle, and there is one initial cycle that does not use any of the edges in G''. Without loss of generality, let the sequence of edges $v_1u_1, u_1u_2, u_2u_3, u_3v_2$ be in cycle 1, v_2u_1, u_1u_3, u_3v_3 be in cycle 2, and v_1u_2, u_2v_3 be in cycle 3. See Figure 5.4b. Denote this as case 3.

Just how unlikely is one of these cases to occur? Using the union bound,

$$\mathbb{P}(\text{Case 1 or case 2 or case 3 occurs }) \leq \sum_{i=1}^{3} \mathbb{P}(\text{Case } i \text{ occurs}).$$

Consider first case 1: G'' has two vertices. Let Z_1 denote the number of occurrences of case 1 in a given graph. There are $O(n^3(2n)^3)$ ways of choosing initial cycles to be present, and their vertices involved. The probability that the vertices identify as in case 1 is $O(n^{-10})$, (a factor of n^{-2} for each vertex identification). Thus, $\mathbb{E}(Z_1) = O(n^{-4})$. As for any non-negative integer random variable X, it holds by the first moment principle that

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} i\mathbb{P}(X = i) \geq \mathbb{P}(X > 0),$$

we have that

$$\mathbb{P}(\text{case 1 occurs}) = O(n^{-4}).$$

The same line of argument works for case 2 and 3: Now, there are $O(n^6)$ ways of choosing the vertices and the cycles, and the probability for the six vertices to pair up in a manner that would lead to a 3–cut is $O(n^{-12})$. Thus, $\mathbb{P}(\text{case 2 or 3 occurs}) = O(n^{-6})$.

In sum, it can be said that the probability of a three cut occurring is $O(n^{-4})$, which tends to 0 as $n \to \infty$. It follows that, asymptotically almost surely, any graph in G_n^* is 4-connected. $\square$

Chapter 6

Conclusions and future work

So far, in this project paper, we have two major results: The distribution of short cycles in G_n^* asymptotically follows a Poisson law (See Chapter 3), and graphs in G_n^* are *a.a.s* 4-connected (Chapter 5). We have also touched upon some open problems in the paper (Conjecture 4.1). This one seems to be the most important one:

6.1 Contiguity

Showing that G_n^* and $G_{n,4}$ are contiguous would facilitate the proof of almost all the other open problems mentioned here; in fact, make them redundant. If G_n^* and $G_{n,4}$ were proven to be contiguous, everything that holds for $G_{n(n-1),4}$ would also hold for G_n^* — see Subsection 1.4.2. The main hurdle to proove this, however, is to calculate the variance of lock compositions for a graph $G \in G_n^*$. The approach used by Kim and Wormald [17] does not work with the model formulation we have, as this proof is strongly dependent on the use of the pairing model. Another approach will have to be found.

6.2 Colourability

We know by Brook's theorem (see, e.g. [6]) that any graph $G \in G_n^*$ is 4-colourable. Considering Conjecture 1.1, it would be of special interest to show that (almost) all graphs $G \in G_n^*$ are 3-colourable. Of course, this would follow immediately after having proved contiguity. (See Theorem 1.5.) After having proved 3-colourability, it might be interesting to restrict the model to planar graphs, and consider colourability on them.

6.3 Diameter

As for the diameter of G_n^*, it is fairly obvious that for any graph $G \in G_n^*$, it holds that

$$\text{diam } G \leq 2n,$$

as for any two vertices v_1 and v_2 it holds that either they are on the same initial cycle — in which case there exists a path from v_1 to v_2 that is shorter than $\frac{2n-2}{2}$, or they are on two different initial cycles, c_1 and c_2, say, such that there are paths of length $\leq n-1$ connecting v_1 and v_2 with an intersection of c_1 and c_2.

We expect that the diameter should be much smaller than this, maybe around $\log(n)$. If we can show that contiguity holds, we immediately get this result (see Theorem 1.4). Otherwise, it might be possible to show this result using branching processes. (See, e.g. [14])

6.4 Conclusion

To summarize, in this project paper we started with a brief history of random graphs and explained the pairing model to describe the uniform d-regular graph model. We motivated our model with some notes on great circle graphs, and then explained two models, G_n^* and $G_n^{\mathcal{A}}$. For G_n^*, we showed that the short cycles distribution is asymptotically Poisson; we presented a conjecture on contiguity to $G_{n,4}$, and finally, we showed that *a.a.s.* all graphs in G_n^* are 4-connected. In this final section, we presented some open problems that give room for further research.

Bibliography

[1] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980.

[2] Béla Bollobás. *Random graphs*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985.

[3] Béla Bollobás and Wenceslas Fernandez de la Vega. The diameter of random regular graphs. *Combinatorica*, 2(2):125–134, 1982.

[4] Béla Bollobás and Oliver Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

[5] Colin Cooper and Alan Frieze. A general model of web graphs. *Random Structures Algorithms*, 22(3):311–335, 2003.

[6] Reinhard Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 2005.

[7] Paul Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, 53:292–294, 1947.

[8] Paul Erdős and Alfréd Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[9] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[10] Stefan Felsner, Ferran Hurtado, Marc Noy, and Ileana Streinu. Hamiltonicity and colorings of arrangement graphs. In *In Proc. 11th Symp. Discrete Algorithms*, pages 155–164, 2000.

[11] Pu Gao and Nicholas C. Wormald. Rate of convergence of the short cycle distribution in random regular graphs generated by pegging. *Electron. J. Combin.*, 16(1):Research Paper 44, 19, 2009.

[12] Pu Gao and Nicholas C. Wormald. Short cycle distribution in random regular graphs recursively generated by pegging. *Random Structures Algorithms*, 34(1):54–86, 2009.

[13] Jacob E. Goodman. Proof of a conjecture of Burr, Grünbaum, and Sloane. *Discrete Mathematics*, 32(1):27–35, 1980.

[14] Theodore E. Harris. *The theory of branching processes.* Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, 1963.

[15] Svante Janson, Donald E. Knuth, Tomasz Łuczak, and Boris Pittel. The birth of the giant component. *Random Structures Algorithms*, 4(3):231–358, 1993.

[16] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[17] Jeong Han Kim and Nicholas C. Wormald. Random matchings which induce Hamilton cycles, and Hamiltonian decompositions of random regular graphs. *J. Combin. Theory Ser. B*, 81:20–44, 2001.

[18] Brendan D. McKay and Nicholas C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11(4):369–382, 1991.

[19] Neil Robertson, Daniel Sanders, Paul Seymour, and Robin Thomas. The four-colour theorem. *J. Combin. Theory Ser. B*, 70(1):2–44, 1997.

[20] Eli Shamir and Joel Spencer. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, 7(1):121–129, 1987.

[21] Lingsheng Shi and Nicholas C. Wormald. Colouring random 4-regular graphs. *Combin. Probab. Comput.*, 16(2):309–344, 2007.

[22] Nicholas C. Wormald. The asymptotic connectivity of labelled regular graphs. *Journal of Combinatorial Theory, Series B*, (31):156–167, 1981.

[23] Nicholas C. Wormald. The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B*, 31(2):168–182, 1981.

[24] Nicholas C. Wormald. The asymptotic connectivity of labelled regular graphs. *Journal of Combinatorial Theory, Series B*, 31(2):156–167, October 1981.

[25] Nicholas C. Wormald. Models of random regular graphs. In *Surveys in combinatorics, 1999 (Canterbury)*, volume 267 of *London Math. Soc. Lecture Note Ser.*, pages 239–298. Cambridge Univ. Press, Cambridge, 1999.