ON RANDOM $k$-OUT GRAPHS WITH PREFERENTIAL ATTACHMENT

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By

Nicholas Richard Peterson, MS
Graduate Program in Mathematics

The Ohio State University
2013

Dissertation Committee:
Dr. Boris Pittel, Advisor
Dr. Neil Falkner
Dr. Jeffery McNeal
ABSTRACT

In a series of papers, Hansen and Jaworski explored a very general model for choosing random mappings with exchangeable in-degrees. The special case in which the in-degrees are obtained by conditioning independent random variables with a specially chosen negative binomial distribution on their sum corresponds in distribution to a process for choosing random mappings which exhibits preferential attachment - images are chosen one at a time, and vertices chosen as images earlier in the process are more likely to be chosen again. They consider the functional digraph as a random object, and study its properties.

We generalize Hansen and Jaworski’s preferential attachment model to a new setting: our model $M_{n,k}^\alpha$ produces digraphs with labeled arcs and uniform out-degree $k$; other than a few technical wrinkles, the graph $G_{n,k}^\alpha$ obtained by ignoring arc directions, loops, and multiple arcs is essentially a preferential attachment version of the $k$-out graph model first studied by Fenner and Frieze. This dissertation splits nicely in to three parts: first, we build a collection of analytical tools for studying $M_{n,k}^\alpha$; second, we study structural properties of the graph $G_{n,k}^\alpha$ induced by $M_{n,k}^\alpha$, including minimum vertex degree, vertex connectivity, and the $k$-core; finally, we present a very strong measurement of the differences and similarities between $M_{n,k}^\alpha$ and the limiting case $M_{n,k}^\infty$, which is uniformly random.
To the four people without whom this would have been impossible: my advisor, Boris Pittel; my fiancée, Lizabeth Goldstein; and my parents, Donald and Sheri Peterson.
ACKNOWLEDGMENTS

First and foremost, I must thank my dissertation advisor, Boris Pittel. The jump from combinatorics to combinatorial probability was a scary one for me; but, I could not have asked for a guide more experienced, knowledgeable, patient, or enthusiastic than Dr. Pittel. I am lucky to be able to call Dr. Pittel my advisor, mentor, and friend, and I can only hope that he is pleased with this product of our many hours of work together.

I also owe a debt of gratitude to Jeffery McNeal. Dr. McNeal signed on as my academic advisor for my doctorate as a result of Dr. Pittel’s retirement, and I am very grateful for his assistance. Helping me is nothing new for Dr. McNeal; in addition to having been my analysis instructor, Dr. McNeal both sat on my candidacy exam committee and was my advisor for my masters degree.

Thanks also to my third committee member, Neil Falkner. In addition to sitting on my defense committee, Dr. Falkner sat on my candidacy exam committee and was my instructor in probability theory; his thoroughness, attention to detail, and willingness to help have been invaluable to me.

Huseyin Acan, Dan Poole, and Christopher Ross - my fellow members of Dr. Pittel’s last cohort of students - have been a great help to me. Their advice and suggestions have helped me through many a tough spot, and their thoughtful questions during our research seminars have helped me to gain a better understanding of the intricacies of, and intuition behind, our work. Thank you all very much!
Last, but certainly not least: thank you to three of the most important people in my life: my fiancée Lizabeth Goldstein and my parents Donald and Sheri Peterson. They may not have been able to help with my research... but, their love and support may be the only reason that I survived it!
VITA

1985 ..............................  Born in Edina, MN

2007 ..............................  BS in Mathematical Sciences
                                (Discrete Concentration)
                                Michigan Technological University

2010 ..............................  MS in Mathematics
                                The Ohio State University

2007-Present .......................... Graduate Teaching Associate
                                 The Ohio State University

FIELDS OF STUDY

Major Field: Mathematics

Specialization: Combinatorial Probability
# TABLE OF CONTENTS

Abstract ................................................................. ii  
Dedication ................................................................. iii  
Acknowledgments ......................................................... iv  
Vita ................................................................. vi  
List of Figures ............................................................ ix  

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction ................................................................. 1</td>
</tr>
<tr>
<td>1.1</td>
<td>Background: $k$-Out Graphs ................................................................. 1</td>
</tr>
<tr>
<td>1.2</td>
<td>Background: Preferential Attachment ................................................................. 3</td>
</tr>
<tr>
<td>1.3</td>
<td>The Probability Space ................................................................. 4</td>
</tr>
<tr>
<td>1.4</td>
<td>Description of the Model ................................................................. 6</td>
</tr>
<tr>
<td>1.5</td>
<td>The Uniform Model ................................................................. 9</td>
</tr>
<tr>
<td>1.6</td>
<td>Outline and Discussion of Results ................................................................. 10</td>
</tr>
<tr>
<td>1.7</td>
<td>Notes on Notation ................................................................. 11</td>
</tr>
<tr>
<td>2</td>
<td>Methods for Analyzing $M_{n,k}^\alpha$ ................................................................. 13</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction ................................................................. 13</td>
</tr>
<tr>
<td>2.2</td>
<td>Probabilities and the $\alpha$-Weight ................................................................. 13</td>
</tr>
<tr>
<td>2.3</td>
<td>Alternative Formulation ................................................................. 16</td>
</tr>
<tr>
<td>2.4</td>
<td>Symmetry ................................................................. 21</td>
</tr>
<tr>
<td>2.5</td>
<td>Coupling and Monotone Properties ................................................................. 24</td>
</tr>
<tr>
<td>3</td>
<td>Structure of the Induced Graph $G_{n,k}^\alpha$ ................................................................. 26</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction ................................................................. 26</td>
</tr>
<tr>
<td>3.2</td>
<td>Minimum Vertex Degree ................................................................. 27</td>
</tr>
<tr>
<td>3.3</td>
<td>Vertex Connectivity of $G_{n,k}^\alpha$ ................................................................. 31</td>
</tr>
<tr>
<td>3.4</td>
<td>The $k$-Core of $G_{n,k}^\alpha$ ................................................................. 42</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>4</td>
<td>Relationship of $M_{n,k}^\alpha$ to the Uniform Map $M_{n,k}^\infty$</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction and Theorem Statement</td>
</tr>
<tr>
<td>4.2</td>
<td>Preliminaries</td>
</tr>
<tr>
<td>4.3</td>
<td>Proof for the case $\alpha/\sqrt{n} \to 0$</td>
</tr>
<tr>
<td>4.4</td>
<td>Proof for the case $\alpha/\sqrt{n} \to \infty$</td>
</tr>
<tr>
<td>4.5</td>
<td>Proof for the case $\alpha = \beta \sqrt{n}$</td>
</tr>
<tr>
<td>5</td>
<td>Open Problems and Related Models</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>5.2</td>
<td>The $k$-Core of $G_{n,k}^\alpha$ and the Model $G_{k-out}^\alpha$</td>
</tr>
<tr>
<td>5.3</td>
<td>The $h$-Cores of $G_{n,k}^\alpha$ for $h &gt; k$</td>
</tr>
<tr>
<td>5.4</td>
<td>Hamiltonicity of $M_{n,k}^\alpha$ and $G_{n,k}^\alpha$</td>
</tr>
<tr>
<td>5.5</td>
<td>Epidemic Processes</td>
</tr>
<tr>
<td>5.6</td>
<td>Variations on $M_{n,k}^\alpha$</td>
</tr>
<tr>
<td>5.7</td>
<td>Generalization of Hansen and Jaworski’s Exchangeable In-Degree Model</td>
</tr>
<tr>
<td>5.8</td>
<td>The Buckley-Osthus Model</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
</tr>
</tbody>
</table>

**APPENDICES**

A  Glossary of Notation                     74
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>8</td>
</tr>
<tr>
<td>3.1</td>
<td>34</td>
</tr>
<tr>
<td>5.1</td>
<td>65</td>
</tr>
</tbody>
</table>

1.1 Functional digraph for $M: 1 \mapsto (1, 2), 2 \mapsto (1, 3), 3 \mapsto (2, 2)$

1.2 Example conditional distribution of a step in generating $M_{3,2}^{\alpha}$

1.3 Functional digraph from Figure 1.1 and its induced graph.

3.1 Example of a (3,1)-separation: $T = \{1, 2, 3\}$, $P = \{4\}$, $R = \{5, 6, 7, 8\}$

5.1 Maximum $h$ (in simulation) such that $G_{n,k}^{\alpha}$ contains an $h$-core.
CHAPTER 1
INTRODUCTION

1.1 Background: $k$-Out Graphs

In the study of random graphs, two (closely related) models dominate: the models $G(n, M)$ and $G(n, p)$. The Erdős-Rényi random graph $G(n, M)$ is chosen uniformly at random from the set of all graphs with vertex set $[n]$ and precisely $M$ edges; it was introduced by Erdős and Rényi in the paper [15] that initiated the study of random graphs. The model $G(n, p)$, often called the Bernoulli random graph, is the graph on vertex set $[n]$ in which a given edge exists with probability $p$ and does not exist with probability $1 - p$, independent of all other edges.

Usually, $G(n, M)$ and $G(n, p)$ are not studied for fixed $p$ or $M$; rather, $p$ and $M$ are allowed to vary with the number $n$ of vertices, and the probability that the graph satisfies various properties is studied asymptotically as $n \rightarrow \infty$. Graph characteristics often behave completely differently for $p$ and $M$ on one side or the other of a specific function of $n$ - a phenomenon called a threshold for the characteristic.

A common theme in the study of these random graphs is that necessary but insufficient conditions for complicated graph properties - existence of a Hamilton cycle, for instance - are often almost sufficient. For instance: Erdős and Rényi [15] proved that the thresholds for $G(n, M)$ to have minimum degree 1 and to be connected coincide. They continued on to show in [17] that this generalizes: the thresholds for
$G(n, M)$ to have minimum degree $k$ and to be $k$-connected agree. Bollobás and Thomason [8] proved the stronger result that if you view $G(n, M)$ as a process, where one edge is added at a time, then with high probability the edge that makes the minimum degree $k$ also makes the graph $k$-connected. Komlós and Szemerédi [29], building on work by Pósa [32], proved that the thresholds for minimum degree 2 and Hamiltonicity coincide. The pattern continues.

Noticing that many of these complicated properties appear as soon as basic minimum degree conditions are satisfied, Fenner and Frieze [19] came up with a new model for random graphs. They called them $m$-orientable graphs; they are now more commonly referred to as $k$-out graphs. In their model, a digraph is chosen uniformly at random from the set of directed graphs in which each vertex has $k$ out-arcs to distinct targets; they study the graph obtained by ignoring arc directions.

This model builds a graph which has minimum degree $k$; their hope was to show that many of the usual minimum degree results from $G(n, M)$ and $G(n, p)$ translated to this new setting. Fenner and Frieze [19] verified that $G_{k-\text{out}}$ is $k$-connected with high probability for $k \geq 2$; Hamilton cycles in $G_{k-\text{out}}$ were studied by Fenner and Frieze [20], Frieze [22], Frieze and Luczak [21], and Bohman and Frieze [4], with the result of Bohman and Frieze finally establishing that $G_{3-\text{out}}$ is Hamiltonian with high probability. Interestingly, $G_{1-\text{out}}$ is w.h.p. not connected even though it has minimum degree 1, and $G_{2-\text{out}}$ is w.h.p. not Hamiltonian even though it has minimum degree 2.

The object of study in this paper will be a $k$-out graph model, with three wrinkles; two minor differences are that our model will permit loops and multiple edges in the digraph (which might allow the minimum degree to fall below $k$), and that our digraph will have uniquely labeled arcs. The major difference, however, is that neighbors will be chosen through a process that exhibits preferential attachment.
1.2 Background: Preferential Attachment

*Preferential attachment* in random graphs generally refers to any process for building a graph edge-by-edge, in which vertices of higher degree are more likely to be incident to newly formed edges. These models exist in many different flavors. A few examples:

- Barabási and Albert [2] studied a model (now referred to as the Barabási-Albert model) in which new vertices are added one at a time, each with edges to some fixed number of previously existing vertices; these neighbors are chosen with probability proportional to current vertex degree.

- Pittel [31] studied a model with a fixed vertex set $n$, where edges are inserted one at a time; if the degree sequence of the current graph is $d = (d_1, \ldots, d_n)$, then the non-adjacent vertices $i$ and $j$ are connected by the next edge with probability proportional to $(d_i + \alpha)(d_j + \alpha)$, where $\alpha$ is a fixed constant.

- Bollobás, Borgs, Chayes, and Riordan [6] studied a directed model; at each step, either a new vertex is added with an out arc, a new vertex is added with an in-arc, or an arc is added between two existing vertices. Whenever the target for an arc must be chosen, it is chosen with probability proportional to $\delta_{in} + d_{in}(v)$, where $\delta_{in} > 0$ is a fixed constant and $d_{in}(v)$ is the current in-degree of $v$; whenever the origin for an arc must be chosen, it is chosen with probability proportional to $\delta_{out} + d_{out}(v)$, where $\delta_{out} > 0$ is a fixed constant and $d_{out}(v)$ is the current out-degree of $v$.

For an overview of similar models written by Bollobás and Riordan, see [7].

The particular model of preferential attachment that we seek to generalize is actually a special case of a very general model for random mappings $[n] \to [n]$ studied by Hansen and Jaworski [25, 26]. Essentially: let $D_1, \ldots, D_n$ be $n$ exchangeable
random variables such that $D_1 + \cdots + D_n = n$; conditioned on their values, choose a mapping uniformly at random from the set of mappings with in-degree sequence $(D_1, \ldots, D_n)$. They proceed to show that, given no information about the distribution of $D_1, \ldots, D_n$ other than exchangeability and summing to $n$, the probabilities of many events can be written as expected values of explicit functions of $D_1, \ldots, D_n$.

As a special case, Hansen and Jaworski [25–27] considered the case where the in-degree sequence $(D_1, \ldots, D_n)$ is distributed as a vector of independent random variables with a specially chosen negative binomial distribution, conditioned on summing to $n$. They showed that the corresponding distribution on the set of mappings could also be generated by a process in which the vertices choose their images one at a time, and are more likely to choose vertices that have been chosen earlier in the process - a process that exhibits preferential attachment.

Our objective in this work is to generalize Hansen and Jaworski’s preferential attachment model from the realm of mappings (which are usually studied as 1-out graphs) to the realm of $k$-out graphs.

1.3 The Probability Space

For the remainder of this work, all probability measures will be defined on the measurable space $\mathcal{M}_{n,k}$, defined as follows:

**Definition.** For $n, k \in \mathbb{N}$, let $\mathcal{M}_{n,k}$ denote the collection of functions $M : [n] \to [n]^k$, where $[n]^k$ denotes the set of $k$-long vectors with elements in $[n]$. Consider $\mathcal{M}_{n,k}$ as a measurable space imbued with the discrete $\sigma$-algebra.

Note that $\mathcal{M}_{n,k}$ stands in one-to-one correspondence with the following two sets:

1) The collection of mappings $[n] \times [k] \to [n]$. The mapping $M \in \mathcal{M}_{n,k}$ corresponds
to the mapping $M' : [n] \times [k] \to [n]$ by letting $M'(i, j)$ be the element in the $j$th coordinate of the vector $M(i)$.

2) The set of digraphs with labeled arcs on vertex set $[n]$ such that each vertex $v \in [n]$ has exactly $k$ outbound arcs, labeled $1, \ldots, k$. The mapping $M \in \mathcal{M}_{n,k}$ corresponds to its *functional digraph*: if $M(i) = (j_1, \ldots, j_k)$, then the $\ell$th arc out of vertex $i$ targets vertex $j_{\ell}$. For instance, if $M \in \mathcal{M}_{3,2}$ is given by

$$M(1) = (1, 2) \quad M(2) = (1, 3) \quad M(3) = (2, 2),$$

then the functional digraph for $M$ is shown in Figure 1.1.

![Figure 1.1: Functional digraph for $M$](image)

With these correspondences in hand, we will identify these three sets. That is, we will simultaneously consider $M \in \mathcal{M}_{n,k}$ as a mapping $[n] \to [n]^k$, a mapping $[n] \times [k] \to [n]$, and as its functional digraph, interchangeably. We will refer to $M \in \mathcal{M}_{n,k}$ as a *k-out map* or *k-out digraph*. The identification with a mapping $[n] \times [k] \to [n]$ gives us the convenient notation $M(i, j)$ to reference the $j$th coordinate of the vector $M(i)$, while the identification with a digraph will give us an interesting object to study going forward.
Identifying $M \in \mathcal{M}_{n,k}$ with its functional digraph gives us a notion of the in-degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$ of $M$. Here, $d_i$ is the number of arcs which point to $i$; equivalently, it is the number of times that $i$ occurs as an image in $M$, including multiplicity. How can we characterize the in-degree sequences of mappings in $\mathcal{M}_{n,k}$?

**Lemma 1.3.1.** An $n$-tuple $\mathbf{d} = (d_1, \ldots, d_n)$ of non-negative integers is the in-degree sequence of a mapping in $\mathcal{M}_{n,k}$ if and only if $d_1 + \cdots + d_n = kn$. The number of mappings in $\mathcal{M}_{n,k}$ with in-degree sequence $\mathbf{d}$ is

$$\binom{kn}{d_1, \ldots, d_n}.$$

**Proof.** That $d_1 + \cdots + d_n = kn$ is necessary is immediate: there are $k$ out-arcs from each vertex in $[n]$, for a total of $kn$ out-arcs; hence the total in-degree must also be $kn$.

To show that $d_1 + \cdots + d_n = kn$ is sufficient, and to count the mappings with in-degree sequence $\mathbf{d}$, note that any choice of targets for the $kn$ (distinguishable) arcs such that $d_1$ choose vertex 1, $d_2$ choose vertex 2, etc. is a valid mapping with in-degree sequence $\mathbf{d}$. □

We will occasionally wish to refer only to those mappings $M \in \mathcal{M}_{n,k}$ which admit a given in-degree sequence:

**Definition.** For $\mathbf{d} = (d_1, \ldots, d_n)$ such that $d_1 + \cdots + d_n = kn$, let $\mathcal{M}_{n,k}(\mathbf{d})$ denote the set of all $M \in \mathcal{M}_{n,k}$ that have in-degree sequence $\mathbf{d}$.

### 1.4 Description of the Model

Note that any $M \in \mathcal{M}_{n,k}$ is determined by $kn$ choices - the choices of targets for each of the $k$ labeled arcs belonging to each of $n$ vertices. We build a mapping $M_{n,k}^\alpha \in \mathcal{M}_{n,k}$, one choice at a time:
Definition. For \( k, n \in \mathbb{N} \) and \( \alpha \in (0, \infty) \), let \( M^\alpha_{n,k} \) be a random element of \( \mathcal{M}_{n,k} \), chosen as follows: the \( kn \) targets are chosen one at a time, ordered first by vertex number and then by arc label. (So that the first choice is the target for vertex 1 arc 1, the second is for vertex 1 arc 2, and so on.) Say that each vertex \( v \in [n] \) has initial weight \( \omega_0(v) = \alpha \). The \( j \)th choice (out of \( kn \)) is made using the conditional distribution

\[
P(j \text{th choice is } v \mid \text{first } j - 1 \text{ choices}) = \frac{\omega_{j-1}(v)}{\sum_{w=1}^{n} \omega_{j-1}(w)}.
\]

Increase the weight of the chosen vertex by 1; that is, define

\[
\omega_j(v) = \begin{cases} 
\omega_{j-1}(v) + 1 & \text{if the } j \text{th choice is } v \\
\omega_{j-1}(v) & \text{else}
\end{cases}
\]

Continue this process until all \( kn \) choices have been made.

By way of example, consider the following scenario in the development of \( M^\alpha_{3,2} \): suppose the first three choices made, in order, are 1, 2, and 1. Then so far, vertex 1 has chosen 1 for its first arc and 2 for its second arc, and vertex 2 has chosen 1 for its first arc. Given these three choices, the current weights are \( \omega(1) = \alpha + 2 \), \( \omega(2) = \alpha + 1 \), and \( \omega(3) = \alpha \); Figure 1.2 illustrates the conditional distribution of the fourth choice, with the candidate for the fourth edge dashed.

Note that what we have really done here is generated a \( kn \)-long vector with elements in \([n]\), representing the choices made in the process, and then used these choices and a specified ordering to choose a mapping \( M^\alpha_{n,k} \in \mathcal{M}_{n,k} \). This view of the process will be helpful when we discuss symmetry in Section 2.4.

There is a natural mapping from the set of digraphs on \([n]\) to the set of graphs on \([n]\): given a digraph \( M \), the associated graph is obtained by ignoring arc directions, loops, multiple edges, and (in our case) arc labels. Figure 1.3 shows the graph obtained from the mapping \( M \in \mathcal{M}_{3,2} \) from Figure 1.1.
In Chapter 3, we will study properties of the graph obtained from $M_{n,k}^{\alpha}$ in this manner:

**Definition.** Let $G_{n,k}^{\alpha}$ be the graph obtained from $M_{n,k}^{\alpha}$ by ignoring arc directions, loops, multiple edges, and arc labels.
1.5 The Uniform Model

In the process for choosing $M_{n,k}^\alpha$ described in Section 1.4, the initial weight $\alpha$ of each of the vertices in $[n]$ can be thought of as a measure of the independent-mindedness of the decision-makers. To see why, let us refer back to the step in generating $M_{3,2}^\alpha$ shown in Figure 1.2. Here, vertex 1 has chosen $(1, 2)$ as its image, and vertex 2 has chosen 1 as its first image. The conditional distribution of the next step, with $\alpha$ unspecified, is given in Figure 1.2 (page 8).

Note that allowing $\alpha$ to tend to 0, future choices are almost surely determined by past ones - in other words, the effect of the first decision on later decisions is very strong, or the decisions are made with almost no independence. Conversely, allowing $\alpha$ to tend to infinity, the probability of any given decision tends to $\frac{1}{3}$, regardless of previous choices - in other words, the first decision has little impact on later decisions, or decisions are made almost independently.

In this sense, we can think of the uniform distribution on $\mathcal{M}_{n,k}$ as the limiting case of $M_{n,k}^\alpha$ where $\alpha = \infty$:

**Definition.** Let $M_{n,k}^\infty$ be a uniformly random element of $\mathcal{M}_{n,k}$, and let $G_{n,k}^\infty$ denote the graph obtained from $M_{n,k}^\infty$ by ignoring arc directions, loops, multiple edges, and arc labels.

In this uniform process, decisions are made in the same order as usual, but each decision is independent of the decisions made before it. The induced graph $G_{n,k}^\infty$ is closely related to the $k$-out model studied by Fenner and Frieze [19]; as discussed in Section 1.1, it only differs in that we allow loops/multiple edges in $M_{n,k}^\infty$, and that the arcs of $M_{n,k}^\infty$ are labeled.

We will see that $M_{n,k}^\alpha$ and $G_{n,k}^\alpha$ exhibit a sort of “continuity” at $\alpha = \infty$: taking essentially any result for fixed $\alpha$ or $\alpha = \alpha(n)$ bounded away from 0 and formally
letting $\alpha \to \infty$ independently of $n$ yields the corresponding result for the uniform case. Further, the same methods of proof used in the $\alpha < \infty$ case will work in the uniform case; so, separate proofs are not really necessary.

This is analogous to a phenomenon in the preferential attachment process of Pittel [31], denoted $\{G_\alpha(n, M)\}_{M=0}^N$. Pittel’s model starts with the empty graph on $n$ vertices, and at each step a new edge is added; the probability that the new edge is $\{i, j\}$ is proportional to $(d_i + \alpha)(d_j + \alpha)$, where $d_i$ and $d_j$ are the current degrees of the vertices and $\alpha > 0$ is a fixed parameter. Letting $\alpha \to \infty$ recovers the Erdős-Rényi process $\{G(n, M)\}_{M=0}^N$. Pittel’s main result in [31] is that w.h.p. $G_\alpha(n, M)$ develops a giant component when the average vertex degree $c := 2M/n$ exceeds $\frac{\alpha}{\alpha+1}$, and the giant component has size asymptotic to

$$n \left[ 1 - \left( \frac{\alpha + c^*}{\alpha + c} \right)^\alpha \right], \quad c^* < \frac{\alpha}{\alpha + 1} \quad \text{and} \quad \frac{c}{(\alpha + c)^{2+\alpha}} = \frac{c^*}{(\alpha + c^*)^{2+\alpha}}.$$ 

Letting $\alpha \to \infty$, we recover the well-known result of Erdős and Rényi [16] that $G(n, M)$ develops a giant component when $c := 2M/n$ exceeds 1 and that the giant component has size $n(1 - \frac{c^*}{c})$, where $c^* < 1$ satisfies $c^*e^{-c^*} = ce^{-c}$.

We will precisely quantify the interesting relationship between $M_{n,k}^\alpha$ and $M_{n,k}^\infty$ in Chapter 4.

1.6 Outline and Discussion of Results

We hope, here, to accomplish four goals, essentially devoting one chapter to each:

- In Chapter 2, we establish several analytical tools that can be leveraged in the study of $M_{n,k}^\alpha$ and $G_{n,k}^\alpha$. Most of these tools (though not all) will prove essential in later chapters, and the rest may prove helpful in future research.

- In Chapter 3, we study the structural properties of the induced graph $G_{n,k}^\alpha$, using the result for $G(n, p)$, $G(n, M)$, and $G_{k-out}$ (see Section 1.1) as our guides. In
particular, we will study minimum vertex degree, vertex connectivity, and the
$k$-core.

• In Chapter 4, we will formalize and quantify the relationship between $M_{n,k}^\alpha$ and
the uniform model $M_{n,k}^\infty$. In particular, we will use probabilistic results on the
in-degree sequence and asymptotics related to rising factorials to establish the
*total variation distance* between the two distributions.

• Finally, in Chapter 5, we will examine several possible courses for future research
in $M_{n,k}^\alpha$ as well as a couple of related models.

### 1.7 Notes on Notation

Throughout this text, several symbols will be used time and time again, always
with the same definition. As a rule, the definitions of such symbols will be typeset
specially (with a bold “Definition.”), while more transitory symbols will be defined
inline as needed. Further, a glossary of important, document-wide symbols is given
in Appendix A.

This section contains some other notational/terminological conventions used throughout this work. First, a definition:

**Definition.** An event for $M_{n,k}^\alpha$ is said to occur with high probability, usually abbre-
viated “w.h.p.”, if the probability of the event tends to 1 as $n \to \infty$.

Second: while studying $M_{n,k}^\alpha$, we will need to make extensive use of both rising
factorials and falling factorials. Each is typically represented by the Pochhammer
symbol $(x)_n$ in the absence of the other; however, this obviously creates some confusion
when both are necessary. We will use the notation of Knuth [28] for rising factorials,
and use the Pochhammer symbol for falling factorials:
**Definition.** Let $a^\bar{b}$ denote the rising factorial

$$a^\bar{b} = a(a + 1) \cdots (a + b - 1),$$

and let $(a)_b$ denote the falling factorial

$$(a)_b = a(a - 1) \cdots (a - (b - 1)).$$

Finally: as mentioned in Section 1.5, $M^\alpha_{n,k}$ tends to exhibit a sort of “continuity” at $\alpha = \infty$, in the sense that properties for the uniform mapping $M^\infty_{n,k}$ are usually precisely the limits of the same properties for $M^\alpha_{n,k}$ as we let $\alpha \to \infty$ and keep $n$ fixed. This relationship is essentially formal - we do not claim that proving the $\alpha$ result and then taking a limit proves the corresponding result for $\alpha = \infty$ - however, the formulas are generally consistent, and so we will not list separate formulas for $\alpha < \infty$ and $\alpha = \infty$ when it can be helped. Further, our method of proof for the $\alpha < \infty$ case can usually be repeated step-by-step in the $\alpha = \infty$ case, with each step matching up perfectly. So, unless there is a difference in method needed, we will not carry out separate proofs for the two cases.
CHAPTER 2
METHODS FOR ANALYZING $M_{n,k}^\alpha$

2.1 Introduction

Working with $M_{n,k}^\alpha$ has several complications not present in more classical models of
random graphs, digraphs, or mappings. Some of these include:

1. The decisions made in generating $M_{n,k}^\alpha$ are not made independently.

2. The process to build $M_{n,k}^\alpha$ lacks symmetry, as decisions about arcs are made in
a specified, fixed order.

3. The definition of our probability measure naturally leads to rising factorials in
formulas, complicating asymptotic analysis.

In this chapter, we aim to ease these difficulties by exploring basic properties of $M_{n,k}^\alpha$
and developing general methods for solving more complex problems.

2.2 Probabilities and the $\alpha$-Weight

Of course, our first order of business must be getting a handle on the distribution of
$M_{n,k}^\alpha$. To do so, we make the following definition:

**Definition.** For $M \in \mathcal{M}_{n,k}$, define the $\alpha$-weight of $M$, $w_\alpha(M)$, by

$$w_\alpha(M) = \prod_{i=1}^{n} \alpha^{x_i},$$

13
where \( \mathbf{d} = (d_1, \ldots, d_n) \) is the in-degree sequence of \( M \).

Interestingly, although the \( \alpha \)-weight of a mapping depends only on its in-degree sequence, it completely determines the distribution of \( M_{n,k}^\alpha \).

**Theorem 2.2.1.** For \( M \in M_{n,k} \),

\[
P(M_{n,k}^\alpha = M) = \frac{w_\alpha(M)}{\sum_{\mu \in M_{n,k}} w_\alpha(\mu)} = \frac{w_\alpha(M)}{(\alpha n)^{kn}}.
\]

**Proof.** Let \( \mathbf{c} = (c_1, \ldots, c_{kn}) \) denote the vertices which must be chosen as images, in order of choice, to make \( M_{n,k}^\alpha = M \) - that is,

\[
c_{k(i-1)+j} = M(i, j), \quad \text{for } 1 \leq i \leq n, \ 1 \leq j \leq k.
\]

Let \( \mathcal{F}_i \) denote the event that the first \( i \) choices made in generating \( M_{n,k}^\alpha \) are precisely \((c_1, \ldots, c_i)\). Then for any \( 1 \leq i \leq kn \),

\[
P(\mathcal{F}_i) = \frac{\prod_{v=1}^{i} \alpha^{d_v(i)}}{(\alpha n)^{i}}, \quad d_v(i) := \#\{j \leq i : c_j = v\}. \tag{2.2.1}
\]

We prove this by induction. For \( i = 1 \): the first choice is made uniformly at random among \( n \) possible choices. Note that \( d_{c_1}(1) = 1 \), while \( d_v(1) = 0 \) for all \( v \neq c_1 \). So,

\[
P(\mathcal{F}_1) = \frac{1}{n} = \frac{\alpha}{\alpha n} = \frac{\prod_{v=1}^{i} \alpha^{d_v(i)}}{(\alpha n)^{i}}.
\]

Suppose that (2.2.1) holds for the first \( i - 1 \) choices; consider choice \( i \). Write

\[
P(\mathcal{F}_i) = P(\text{ith choice is } c_i \mid \mathcal{F}_{i-1}) \cdot P(\mathcal{F}_{i-1})
\]

\[
= P(\text{ith choice is } c_i \mid \mathcal{F}_{i-1}) \cdot \frac{\prod_{v=1}^{i} \alpha^{d_v(i-1)}}{(\alpha n)^{i-1}}.
\]

Now, vertex \( c_i \) was chosen precisely \( d_{c_i}(i - 1) \) times in the first \( i - 1 \) steps of the process; so, it now has weight \( \alpha + d_{c_i}(i - 1) \). Since \( i - 1 \) choices have been made, the total weight of all vertices is \( \alpha n + i - 1 \). So,

\[
P(\text{ith choice is } c_i \mid \mathcal{F}_{i-1}) = \frac{\alpha + d_{c_i}(i - 1)}{\alpha n + i - 1}.
\]
Noting that \( d_c(i) = d_c(i - 1) + 1 \) and \( d_v(i) = d_v(i - 1) \) for all \( v \neq c \), we find that
\[
P(F_i) = \frac{(\alpha + d_c(i - 1))\alpha^{d_c(i - 1)}\prod_{v \neq c} \alpha^{d_v(i - 1)}}{(\alpha n + i - 1)(\alpha n)^{i-1}} = \frac{\prod_{i=1}^n \alpha^{d_v(i)}}{(\alpha n)^i}.
\]
So, inductively, (2.2.1) holds for \( 1 \leq i \leq kn \). Taking \( i = kn \) and noting that \((d_1(kn), \ldots, d_n(kn))\) is precisely the in-degree sequence of \( M \), we find
\[
P(M_{n,k}^\alpha = M) = P(F_{kn}) = \frac{w_\alpha(M)}{(\alpha n)^{kn}}.
\]
That the total weight of all \( \mu \in \mathcal{M}_{n,k} \) is \((\alpha n)^{kn}\) follows immediately from
\[
\sum_{M \in \mathcal{M}_{n,k}} P(M_{n,k}^\alpha = M) = 1.
\]

This result implies that the in-degree sequence will be extremely important going forward; let us give it a name:

**Definition.** Let \( D_n = (D_{n,1}, \ldots, D_{n,n}) \) denote the in-degree sequence of \( M_{n,k}^\alpha \), where \( D_{n,j} \) is the in-degree of vertex \( j \).

Theorem 2.2.1 allows us to shift gears: rather than having to consider \( M_{n,k}^\alpha \) as a growing process and compute events by a sequence of conditioning arguments, we can simply consider the total \( \alpha \)-weight of all mappings satisfying a desired property.

This transition to counting arguments can be extremely useful! One simple example: let us compute the distribution of the in-degree sequence.

**Lemma 2.2.2.** Let \( d = (d_1, \ldots, d_n) \) be a valid in-degree sequence for a mapping in \( \mathcal{M}_{n,k} \); that is, \( d_1, \ldots, d_n \geq 0 \) and \( d_1 + \cdots + d_n = kn \). Then
\[
P(D_n = d) = \binom{kn}{d_1, \ldots, d_n} \frac{\prod_{i=1}^n \alpha^{d_i}}{(\alpha n)^{kn}}.
\]

**Proof.** Every mapping with the in-degree sequence \( d \) has the same \( \alpha \)-weight; so,
\[
P(D_n = d) = \sum_{M \in \mathcal{M}_{n,k}(d)} \frac{w_\alpha(M)}{(\alpha n)^{kn}} = |\mathcal{M}_{n,k}(d)| \frac{\prod_{i=1}^n \alpha^{d_i}}{(\alpha n)^{kn}} = \binom{kn}{d_1, \ldots, d_n} \frac{\prod_{i=1}^n \alpha^{d_i}}{(\alpha n)^{kn}},
\]
proving the result. \( \Box \)
2.3 Alternative Formulation

As discussed in Section 1.2, Hansen and Jaworski arrived at the mapping model \((M_{n,1}^\alpha, \text{in our notation})\) along a different route, as a special case of their exchangeable in-degree model. In essence: take \(n\) independent, identically distributed random variables \(Z_{n,1}, \ldots, Z_{n,n}\) with the generalized negative binomial distribution

\[
P(Z_{n,j} = d) = \frac{\alpha^d}{d!} \left( \frac{\alpha}{\alpha + 1} \right)^\alpha \left( \frac{k}{\alpha + k} \right)^d.
\]  

Choose the in-degree sequence \(\tilde{D}_n\) of a random mapping by conditioning these negative binomials on their sum being \(n\). Conditioned on this degree sequence, choose a mapping uniformly at random from \(M_{n,k}(\tilde{D}_n)\). Hansen and Jaworski showed that this random mapping has the same distribution as the random map process we refer to as \(M_{n,1}^\alpha\).

They also showed that the uniformly random map can be realized in the same way, by taking \(Z_{n,1}, \ldots, Z_{n,n}\) to be independent and Poisson-distributed with parameter 1. It is interesting to note that this is in keeping with our theme of viewing the uniform mapping as the limiting case \(\alpha = \infty\): letting \(\alpha \to \infty\) but keeping \(n\) fixed in (2.3.1) yields precisely \(e^{-1}/d!\), the probability that a Poisson-distributed random variable with mean 1 takes the value \(d\).

This idea generalizes to \(M_{n,k}^\alpha\), and this alternative characterization will prove helpful in establishing methods for computing the probabilities of various structural events in \(M_{n,k}^\alpha\) and \(G_{n,k}^\alpha\). Here is our generalization:

**Definition.** Let \(Z_{n,1}, \ldots, Z_{n,n}\) denote independent random variables with the negative binomial distribution

\[
P(Z_{n,j} = d) = \frac{\alpha^d}{d!} \left( \frac{\alpha}{\alpha + k} \right)^\alpha \left( \frac{k}{\alpha + k} \right)^d.
\]

and let \(Z_n = (Z_{n,1}, \ldots, Z_{n,n})\). In the case \(\alpha = \infty\), let \(Z_{n,1}, \ldots, Z_{n,n}\) be distributed
as the formal limit of (2.3.2) suggests: as Poisson-distributed random variables with mean \( k \).

Before we prove that this is the correct generalization of (2.3.1), let us establish some helpful properties of the variables \( Z_{n,1}, \ldots, Z_{n,n} \) in (2.3.2):

**Lemma 2.3.1.** Let \( Z_{n,1}, \ldots, Z_{n,n} \) be defined as in (2.3.2).

i) The probability generating function for \( Z_{n,j} \) is

\[
E[x^{Z_{n,j}}] = \left( \frac{\alpha}{\alpha + k} \right)^\alpha \left( 1 - \frac{kx}{\alpha + k} \right)^{-\alpha}, \quad |x| < \frac{\alpha + k}{k},
\]

ii) The falling factorial moments for \( Z_{n,j} \) are given by

\[
E[(Z_{n,j})^\ell] = \frac{\alpha^\ell k^\ell}{\alpha^\ell},
\]

iii) The moments for \( Z_{n,j} \) are given by

\[
E[Z_{n,j}^s] = \sum_{\ell=1}^{s} \binom{s}{\ell} \frac{\alpha^\ell k^\ell}{\alpha^\ell},
\]

where \( \{ \binom{s}{\ell} \} \) is the Stirling partition number: the number of ways to partition a set of \( s \) distinct elements into \( \ell \) non-empty parts.

iv) The probability that \( Z_n \) is the in-degree sequence of a \( k \)-out map is

\[
P(Z_{n,1} + \cdots + Z_{n,n} = kn) = \frac{(\alpha n)^{kn}}{(kn)!} \left( \frac{\alpha}{\alpha + k} \right)^{an} \left( \frac{k}{\alpha + k} \right)^{kn}.
\]

In particular, if \( \alpha = \alpha(n) \) is bounded away from 0 then

\[
P(Z_{n,1} + \cdots + Z_{n,n} = kn) \sim \sqrt{\frac{\alpha}{2\pi k(\alpha + k)n}}.
\]

**Note 1:** We use the notation \( Z_{n,j} \) rather than just \( Z_j \) because the distribution of \( Z_{n,j} \) depends on \( \alpha \), which we will sometimes allow to vary with \( n \).
Note 2: In fact, letting the $Z_{n,j}$ have any generalized negative binomial distribution with shape parameter $\alpha$ will lead to the correct distribution when we condition on their sum; we take this particular distribution because it makes $E[Z_{n,j}] = k$, maximizing $P(Z_{n,1} + \cdots + Z_{n,n} = kn)$.

Proof. For (i), note that if $\left| \frac{kx}{\alpha + k} \right| < 1$, then
\[
E[x^{Z_{n,j}}] = \left( \frac{\alpha}{\alpha + k} \right)^{\alpha} \sum_{d=0}^{\infty} \frac{\alpha^d}{d!} \left( \frac{kx}{\alpha + k} \right)^d = \left( \frac{\alpha}{\alpha + k} \right)^{\alpha} \left( 1 - \frac{kx}{\alpha + k} \right)^{-\alpha}.
\]

For (ii): for $|x| < \frac{\alpha + k}{k}$, we have
\[
E[(Z_{n,j})_\ell x^{Z_{n,j}-\ell}] = \frac{d^\ell}{dx^\ell} \left[ E[x^{Z_{n,j}}] \right] = \alpha^\ell \left( \frac{\alpha}{\alpha + k} \right)^{\alpha} \left( \frac{k}{\alpha + k} \right)^\ell \left( 1 - \frac{kx}{\alpha + k} \right)^{-(\alpha + \ell)}.
\]

Taking $x = 1$ and simplifying yields (ii). Result (iii) follows immediately from (ii) and the identity
\[
z^s = \sum_{\ell=1}^{s} \binom{s}{\ell} (z)_\ell.
\]

Finally, for (iv): as in Lemma 1.3.1, $Z_n$ is the in-degree sequence of a $k$-out map if and only if $Z_{n,1} + \cdots + Z_{n,n} = kn$. Since $Z_{n,1}, \ldots, Z_{n,n}$ are independent, we have
\[
P(Z_{n,1} + \cdots + Z_{n,k} = kn) = [x^{kn}] E[x^{Z_{n,1} + \cdots + Z_{n,n}}]
\]
\[
= [x^{kn}] \left( \frac{\alpha}{\alpha + k} \right)^{\alpha n} \left( 1 - \frac{kx}{\alpha + k} \right)^{-\alpha n}
\]
\[
= \frac{(\alpha n)^{kn}}{(kn)!} \left( \frac{\alpha}{\alpha + k} \right)^{\alpha n} \left( \frac{k}{\alpha + k} \right)^{kn},
\]
as claimed. The asymptotic estimate follows by noting $a^b = \Gamma(b)/\Gamma(a)$ and applying Stirling’s formula. \qed

We are now ready to prove our alternative formulation for $M_{n,k}^\alpha$:

**Theorem 2.3.2.** Let $Z_n = (Z_{n,1}, \ldots, Z_{n,n})$ be as in (2.3.2).
i) Given \( d = (d_1, \ldots, d_n) \) such that \( d_1 + \cdots + d_n = kn \),

\[
P(D_n = d) = P(Z_n = d \mid Z_{n,1} + \cdots + Z_{n,n} = kn).
\]

ii) For \( M \in \mathcal{M}_{n,k}(d) \),

\[
P(M^\alpha_{n,k} = M \mid D_n = d) = \frac{1}{|\mathcal{M}_{n,k}(d)|} = \left( \frac{kn}{d_1, \ldots, d_n} \right)^{-1}.
\]

**Proof.** In Lemma 2.2.2, we computed

\[
P(D_n = d) = \left( \frac{kn}{d_1, \ldots, d_n} \right) \prod_{i=1}^{n} \frac{\alpha d_i}{(\alpha n)^{kn}}.
\]

Further, by Lemma 2.3.1(iv),

\[
P(Z_{n,1} + \cdots + Z_{n,n} = kn) = \left( \frac{\alpha n}{kn} \right)^{kn} \left( \frac{\alpha}{\alpha + k} \right)^{\alpha n} \left( \frac{k}{\alpha + k} \right)^{kn}.
\]

Independence of \( Z_{n,1}, \ldots, Z_{n,j} \), and the assumption \( d_1 + \cdots + d_n = kn \), then imply

\[
P(Z_n = d, Z_{n,1} + \cdots + Z_{n,n} = kn) = \left( \frac{\alpha}{\alpha + k} \right)^{\alpha n} \left( \frac{k}{\alpha + k} \right)^{kn} \prod_{i=1}^{n} \frac{\alpha d_i}{d_i!}.
\]

Combining these observations yields

\[
P(Z_n = d \mid Z_{n,1} + \cdots + Z_{n,n} = kn) = \left( \frac{kn}{d_1, \ldots, d_n} \right) \prod_{i=1}^{n} \frac{\alpha d_i}{(\alpha n)^{kn}} = P(D_n = d),
\]

proving (i). For (ii): if \( M \in \mathcal{M}_{n,k}(d) \), then

\[
P(M^\alpha_{n,k} = M \mid D_n = d) = \frac{P(M^\alpha_{n,k} = M)}{P(D_n = d)} = \left( \frac{kn}{d_1, \ldots, d_n} \right)^{-1},
\]

as claimed.

This alternative characterization of \( M^\alpha_{n,k} \) is of immense value: independence of \( Z_{n,1}, \ldots, Z_{n,n} \) allows us to bring generating function and characteristic function methods to the table. Further, for any collection \( C_n \) of in-degree sequences, we have the conditioning device

\[
P(D_n \in C_n) = \frac{P(Z_n \in C_n, Z_{n,1} + \cdots + Z_{n,n} = kn)}{P(Z_{n,1} + \cdots + Z_{n,n} = kn)} \leq \frac{P(Z_n \in C_n)}{P(Z_{n,1} + \cdots + Z_{n,n} = kn)}.
\]
if $\alpha = \alpha(n)$ is bounded away from 0, then by Lemma 2.3.1 this implies

$$P(D_n \in C_n) = O \left( \sqrt{n} P(Z_n \in C_n) \right). \quad (2.3.3)$$

We will see an interesting use of the $Z_{n,j}$ to prove a central limit theorem for $D_{n,1}^2 + \cdots + D_{n,n}^2$ in Section 4.5. For now, we can use it to compute the moments for the $D_{n,j}$:

**Lemma 2.3.3.** Let $D_n = (D_{n,1}, \ldots, D_{n,n})$ and $Z_n = (Z_{n,1}, \ldots, Z_{n,n})$ be as above.

i) The falling factorial moments for $D_{n,j}$ are given by

$$E[(D_{n,j})^\ell] = \frac{\alpha^\ell(kn)_\ell}{(\alpha n)^\ell}. \quad (2.3.4)$$

ii) The moments of $D_{n,j}$ are given by

$$\mu_{s,\alpha} := E[D_{n,j}^s] = \sum_{\ell=1}^{s} \left\{ \frac{s}{\ell} \right\} \frac{\alpha^\ell(kn)_\ell}{(\alpha n)^\ell}. \quad (2.3.5)$$

iii) The mixed factorial moments of $D_{n,1}, \ldots, D_{n,n}$ are given by

$$E[(D_{n,i})^\ell(D_{n,j})^m] = \frac{\alpha^\ell \alpha^m(kn)^{\ell+m}}{(\alpha n)^{\ell+m}}. \quad (2.3.6)$$

iv) If $\alpha = \alpha(n)$ is bounded away from 0, then

$$E[(D_{n,i})^\ell(D_{n,j})^m] = E[(D_{n,i})^\ell] E[(D_{n,j})^m] + O \left( \frac{1}{n} \right). \quad (2.3.7)$$

**Proof.** Using the probability generating function for $Z_{n,j}$ computed in Lemma 2.3.1, we find

$$E[(Z_{n,1})^\ell x^{Z_{n,1}}] = x^{\ell} \frac{d^\ell}{dx^\ell} E[x^{Z_{n,1}}] = x^{\ell} \alpha^\ell \left( \frac{\alpha}{\alpha + k} \right)^\ell \left( \frac{k}{\alpha + k} \right)^\ell \left( 1 - \frac{kx}{\alpha + k} \right)^{-(\alpha + \ell)}. \quad (2.3.8)$$

Noting that $D_n$ is distributed as $Z_n$ conditioned on $Z_{n,1} + \cdots + Z_{n,n} = kn$, and leveraging independence of the $Z_{n,j}$, we find

$$P(D_{n,1} = d) = \frac{P(Z_{n,1} = d) \cdot P(Z_{n,2} + \cdots + Z_{n,n} = kn - d)}{P(Z_{n,1} + \cdots + Z_{n,n} = kn)}. \quad (2.3.9)$$
As such,
\[
\mathbb{E}[(D_{n,1})^\ell] = \sum_{d=0}^{\infty} \frac{(d_1^\ell P(Z_{n,1} = d) \cdot P(Z_{n,2} + \cdots + Z_{n,n} = kn - d)}{P(Z_{n,1} + \cdots + Z_{n,n} = kn)}
\]
\[
= \left[ x^{kn} \right] \mathbb{E}[(Z_{n,1})^\ell x^{Z_{n,1}}] \cdot \mathbb{E}[x^{Z_{n,2} + \cdots + Z_{n,n}}]
\]
\[
= \alpha^\ell \left( \frac{k}{\alpha + k} \right)^\ell \frac{[x^{kn-\ell}](1 - \frac{kx}{\alpha + k})^{-(\alpha n + \ell)}}{[x^{kn}](1 - \frac{kx}{\alpha + k})^{-\alpha n}}
\]
\[
= \alpha^\ell \frac{(kn)^\ell}{(\alpha n)^{\ell+1}}.
\]
As the in-degrees are exchangeable, this proves (i). Claim (ii) follows immediately from claim (i) and the identity
\[
z^s = \sum_{\ell=1}^{s} \left\{ \frac{s}{\ell} \right\} (z)^\ell,
\]
as in the proof of Lemma 2.3.1. For (iii), for \(\ell, m \in \mathbb{N}\) we compute
\[
\mathbb{E}[(D_{n,1})^\ell (D_{n,2})_m] = \sum_{d_1,d_2=0}^{\infty} (d_1)^\ell (d_2)_m P(D_{n,1} = d_1, D_{n,2} = d_2)
\]
\[
= \left[ x^{kn} \right] \mathbb{E}[(Z_{n,1})^\ell x^{Z_{n,1}}] \cdot \mathbb{E}[(Z_{n,2})_m x^{Z_{n,2}}] \cdot \mathbb{E}[x^{Z_{n,3} + \cdots + Z_{n,n}}]
\]
\[
= \alpha^\ell \frac{\alpha^m}{(\alpha n)^{\ell+1}} \left( \frac{k}{\alpha + k} \right)^{\ell+m} \frac{[x^{kn-\ell-m}](1 - \frac{kx}{\alpha + k})^{-(\alpha n + \ell+m)}}{[x^{kn}](1 - \frac{kx}{\alpha + k})^{-\alpha n}}
\]
\[
= \alpha^\ell \frac{\alpha^m (kn)^{\ell+m}}{(\alpha n)^{\ell+1}},
\]
and again note that the more general result again holds by exchangeability of the in-degrees. Claim (iv) follows by noting that if \(\alpha\) is bounded away from 0, then
\[
\frac{(kn)^{\ell}}{(\alpha n)^t} = \left( \frac{k}{\alpha} \right)^t + O \left( \frac{1}{n} \right).
\]

### 2.4 Symmetry

One disturbing aspect of the process that produces \(M_{n,k}^\alpha\) is that it seems to lack symmetry - the first choice made by vertex 1 is made uniformly at random, while the
kth choice made by vertex \( n \) depends (heavily) on the \( kn - 1 \) choices made earlier in the process. This lack of symmetry would seem to make the goal of computing probabilities of structural events unlikely - different sets of vertices would have different probabilities of having the same structure!

However, in Theorem 2.3.2, we saw that the in-degrees \( D_n,1, \ldots, D_n,n \) of \( M_{n,k}^\alpha \) are actually exchangeable random variables; this is a hopeful sign that there may be more symmetry to \( M_{n,k}^\alpha \) than meets the eye.

In fact, there is a great deal of symmetry in \( M_{n,k}^\alpha \): rather than having vertex 1 make its choices in order, then vertex 2, and so on, we can pick any arbitrary order for the choices to be made. The end result will have the same distribution as our mapping \( M_{n,k}^\alpha \)!

**Theorem 2.4.1.** Let \( \sigma : [n] \times [k] \to [kn] \) be injective. Let \( \tilde{M}_{n,k}^\alpha \) be chosen from \( M_{n,k}^\alpha \) in the same manner as \( M_{n,k}^\alpha \), except with choices made in the order given by \( \sigma \) - that is, \( \sigma(i,j) = m \) means that the \( m \)th decision made is \( \tilde{M}_{n,k}^\alpha(i,j) \). Then \( \tilde{M}_{n,k}^\alpha \equiv M_{n,k}^\alpha \).

**Proof.** Let \( \tau : [n] \times [k] \to [kn] \) denote the usual order in which choices are made - that is,

\[
\tau(i,j) = k(i-1) + j, \quad 1 \leq i \leq n, \ 1 \leq j \leq k.
\]

We couple the processes that generate \( M_{n,k}^\alpha \) and \( \tilde{M}_{n,k}^\alpha \). Let \( C_{n,k}^\alpha = (C_1, \ldots, C_{kn}) \), \( C_i \in [n] \), be chosen randomly as follows: say that each \( v \in [n] \) has initial weight \( \omega_0(v) = \alpha \). Choose \( C_1, C_2, \ldots, C_{kn} \) in order, with conditional distribution

\[
P(C_i = v \mid C_1, \ldots, C_{i-1}) = \frac{\omega_{j-1}(v)}{\sum_{w=1}^n \omega_{j-1}(w)}.
\]

Increase the weight of the chosen vertex by 1; that is, if \( v \) is chosen, then

\[
\omega_j(v) = \omega_{j-1}(v) + 1, \quad \text{and} \quad \omega_j(w) = \omega_{j-1}(w) \text{ for } w \neq v.
\]

Think of \( C_{n,k}^\alpha \) as the choices that are made in generating either \( M_{n,k}^\alpha \) or \( \tilde{M}_{n,k}^\alpha \), in the order they are made - but without any reference to which decision is being made.
We can then generate a mapping \( M_\sigma \in \mathcal{M}_{n,k} \) by letting \( M_\sigma(i, j) = C_\sigma(i, j) \), and \( M_\tau \in \mathcal{M}_{n,k} \) by letting \( M_\tau(i, j) = C_\tau(i, j) \) - in other words, use the orderings \( \sigma \) and \( \tau \) of the choices to be made, as well as the list \( C^\alpha_{n,k} \) of the decisions made in order, to determine the mappings.

Then \( M_\sigma \overset{D}{=} M^\alpha_{n,k} \) and \( M_\tau \overset{D}{=} M^\alpha_{n,k} \). Further, \( M_\sigma \) and \( M_\tau \) are both determined by \( C^\alpha_{n,k} \), leading to the relationship

\[
M_\tau(i, j) = C_\tau(i, j) = M_\sigma(\sigma^{-1}\tau(i, j)), \quad 1 \leq i \leq n, \ 1 \leq j \leq k.
\]

So, for any \( M \in \mathcal{M}_{n,k} \),

\[
P(M^\alpha_{n,k} = M) = P(M_\sigma = M) = P(M_\tau = M') = P(M^\alpha_{n,k} = M'),
\]

where \( M' \) is defined by \( M'(i, j) = M(\sigma^{-1}\tau(i, j)) \) for \((i, j) \in [n] \times [k]\). However, note that \( M' \) and \( M \) have the same in-degree sequence: the elements of \( M \) are permuted to get \( M' \), so each element occurs the same number of times in \( M \) and \( M' \). Thus \( w_\alpha(M) = w_\alpha(M') \), and

\[
P(M^\alpha_{n,k} = M) = P(M^\alpha_{n,k} = M') = \frac{w_\alpha(M')}{(\alpha n)^k} = \frac{w_\alpha(M)}{(\alpha n)^k} = P(M^\alpha_{n,k} = M),
\]

proving the result.

Theorem 2.4.1 will be absolutely invaluable going forward, for two reasons:

First: by changing the order in which choices are made, we can relate the probability that some set \( V \subseteq [n] \) has a structure to the probability that \( \{1, 2, \ldots, |V|\} \) has that same structure. So, when computing expected numbers of arrangements in \( M^\alpha_{n,k} \), we can use symmetry to greatly reduce the set of possibilities.

Second: when computing the probability that \( \{1, 2, \ldots, |V|\} \) has a given structure, we can rearrange the choices so that at each step in generating \( M^\alpha_{n,k} \), we know the distribution of the weights of the vertices. This allows us to quickly and easily compute closed-form formulas for probabilities, rather than having to start from multiple summations and simplify.
2.5 Coupling and Monotone Properties

Think of a graph property as a set $Q$ of graphs. A graph property $Q$ is called monotone increasing if it is maintained by the addition of edges - that is, if $G$ and $H$ are graphs on the same vertex set and $H \subseteq G$, then $H \in Q$ implies $G \in Q$. For example, connectivity is a monotone increasing graph property.

Similarly, $Q$ is called monotone decreasing if it is maintained through the deletion of edges. For example, the property of being bipartite is monotone decreasing.

In the study of classical random graphs, for instance the Erdős-Rényi random graph $G(n, M)$, there are nice theorems which relate the probabilities of monotone graph properties over different values of parameters: for instance, if $Q$ is monotone increasing and $0 \leq M_1 \leq M_2 \leq \binom{n}{2}$, then

$$P(G(n, M_1) \in Q) \leq P(G(n, M_2) \in Q).$$

(See, for instance, [5, Theorem 2.1].)

The relationship between $\alpha$ and $M_{n,k}^\alpha$ is too complex to nicely encapsulate in a theorem of this sort; however, the symmetry results of Section 2.4 do allow us to relate probabilities of monotone properties for various values of $k$:

**Theorem 2.5.1.** Suppose that $Q$ is a monotone increasing digraph property and $R$ is a monotone decreasing digraph property. Let $\alpha > 0$ be given. Then for $1 \leq k_1 \leq k_2$,

$$P(M_{n,k_1}^\alpha \in Q) \leq P(M_{n,k_2}^\alpha \in Q) \quad \text{and} \quad P(M_{n,k_1}^\alpha \in R) \geq P(M_{n,k_2}^\alpha \in R).$$

**Proof.** Relying on Theorem 2.4.1, let us build $M_{n,k_2}^\alpha$ using the following decision order: first the vertices (in order of increasing vertex number) each make their first decision; then they each make their second decision; and so on. Let $M_{n,k_1}^\alpha$ be the snapshot of this process after all $n$ vertices have made their first $k_1$ decisions. Then by another
appeal to Theorem 2.4.1, $\tilde{M}_{n,k_1}^\alpha$ has the same distribution as $M_{n,k_1}^\alpha$; further, since the remainder of the process only involves adding edges, $\tilde{M}_{n,k_1}^\alpha$ is a subgraph of $M_{n,k_2}^\alpha$.

Since $Q$ is monotone increasing, $M_{n,k_2}^\alpha \in Q$ whenever $\tilde{M}_{n,k_1}^\alpha \in Q$; so,

$$P(M_{n,k_1}^\alpha \in Q) = P(\tilde{M}_{n,k_1}^\alpha \in Q) \leq P(M_{n,k_2}^\alpha \in Q).$$

Since $R$ is monotone decreasing, $\tilde{M}_{n,k_1}^\alpha \in R$ whenever $M_{n,k_2}^\alpha \in R$; so,

$$P(M_{n,k_1}^\alpha \in R) = P(\tilde{M}_{n,k_1}^\alpha \in R) \geq P(M_{n,k_2}^\alpha \in R).$$

The primary benefit of this result is that it allows us to leverage information from Hansen and Jaworski’s papers (for instance [25], [26], and [27]) that deal with the $k = 1$ case. For an easy example:

**Corollary 2.5.2.** Suppose $\alpha = \alpha(n)$ is such that $\alpha n \to 0$ as $n \to \infty$. Let $k \geq 1$.

Then w.h.p., $G_{n,k}^\alpha$ is connected.

**Proof.** Hansen and Jaworski [27, Theorem 2(iii)] proved that if $\alpha n \to 0$, then

$$\lim_{n \to \infty} P(G_{n,1}^\alpha \text{ is connected}) = 1.$$

Let $Q$ be the digraph property “the graph induced by the digraph $D$ is connected.” Then $Q$ is monotone increasing: adding arcs to a digraph $D$ can only possibly add edges to the induced graph, and connectivity is a monotone increasing graph property.

So, for $k \geq 1$, Theorem 2.5.1 implies

$$P(G_{n,k}^\alpha \text{ is connected}) = P(M_{n,k}^\alpha \in Q) \geq P(M_{n,1}^\alpha \in Q) = P(G_{n,1}^\alpha \text{ is connected});$$

so, since $G_{n,1}^\alpha$ is connected w.h.p., it is also true that $G_{n,k}^\alpha$ is connected w.h.p..
CHAPTER 3

STRUCTURE OF THE INDUCED GRAPH $G_{n,k}^\alpha$

3.1 Introduction

Let us now restrict our attention to $G_{n,k}^\alpha$, the graph obtained from $M_{n,k}^\alpha$ by ignoring arc directions, loops, repeated arcs, and arc labels. In this chapter, we will address several common structural questions about $G_{n,k}^\alpha$: namely its minimum vertex degree, its vertex connectivity, and the existence, size, and properties of its $k$-core.

At the core of all of these properties, as might be expected, is the distribution of the vertex degrees. We will see that vertices of degree $k - 2$ or less are very unlikely; the number of vertices of degree $k - 1$ will be asymptotically Poisson-distributed, with mean the following function of $\alpha$ and $k$:

**Definition.** Let $\lambda : (0, \infty] \times \mathbb{N} \to \mathbb{R}$ be defined by

$$
\lambda(\alpha, k) := \begin{cases} 
\left( \frac{\alpha}{\alpha + k} \right)^\alpha \left( \frac{k\alpha}{\alpha + k} + \frac{k}{2} \frac{\alpha + 1}{\alpha} \right) & \text{if } \alpha \in (0, \infty) \\
\frac{k(k + 1)}{2e^k} & \text{if } \alpha = \infty
\end{cases}.
$$

(3.1.1)

Note that $\lambda(\alpha, k) \to \lambda(\infty, k)$ as $\alpha \to \infty$; so, as discussed in Section 1.7, we will usually refer to the first definition and allow $\alpha \in (0, \infty]$.
3.2 Minimum Vertex Degree

First, we will address the minimum vertex degree of $G_{n,k}^\alpha$. As discussed in Section 1.1, $G_{n,k}^\alpha$ lives on a slightly different probability space than does the $k$-out model studied by Fenner and Frieze [19], in that we allow loops and repeated arcs in $M_{n,k}^\alpha$. This means that $G_{n,k}^\alpha$ does not necessarily have minimum degree $k$. However, we shall see that even when $G_{n,k}^\alpha$ falls short of minimum degree $k$, it does not do so by much.

Let us name the set of vertices with specified degree in $G_{n,k}^\alpha$:

**Definition.** Let $V_{n,d}$ denote the set of vertices of degree exactly $d$ in $G_{n,k}^\alpha$.

We are now prepared for the main result on minimum degree:

**Theorem 3.2.1.** For a graph $G$, let $\delta(G)$ denote the minimum vertex degree in $G$. Suppose $\alpha = \alpha(n)$ is bounded away from 0. Let $\lambda(\alpha, k)$ be as in (3.1.1). Then:

i) For any $h \in \mathbb{N}$,

$$P(\delta(G_{n,k}^\alpha) \geq h) = \begin{cases} 1 + o(1) & \text{if } h \leq k - 1 \\ e^{-\lambda(\alpha,k)} + o(1) & \text{if } h = k \\ o(1) & \text{if } h \geq k + 1 \end{cases}$$

as $n \to \infty$.

ii) In the limit, $|V_{n,k-1}|$ is Poisson-distributed with parameter $\lambda(\alpha, k)$, in the sense that for all $m \in \mathbb{N}$,

$$\lim_{n \to \infty} \left| P(|V_{n,k-1}| = m) - \frac{e^{-\lambda(\alpha,k)}(\lambda(\alpha,k))^m}{m!} \right| = 0.$$ 

Conditioned on $|V_{n,k-1}|$, $V_{n,k-1}$ is chosen uniformly at random from all subsets of $[n]$ of size $|V_{n,k-1}|$. 

27
Proof. We will prove this result as follows: first, we show that w.h.p. $\delta(G^\alpha_{n,k}) \geq k - 1$; second, that w.h.p. $\delta(G^\alpha_{n,k}) \leq k$; and third, that the number of vertices of degree $k - 1$ is asymptotically Poisson.

For the first: if there is a vertex $v$ of $G^\alpha_{n,k}$ which has degree at most $k - 2$, then necessarily there is a set $L$ of size $\ell \in [0, k - 2]$ so that all out-arcs of $v$ in $M^\alpha_{n,k}$ end in $L \cup \{v\}$. Using Markov’s inequality to bound by the expected number of such arrangements, we find that

$$P(\delta(G^\infty_{n,k}) \leq k - 2) \leq \sum_{\ell=0}^{k-2} n \binom{n-1}{\ell} \frac{(\alpha(\ell + 1))^k}{(\alpha n)^k} = O\left(\frac{1}{n}\right)$$

as $n \to \infty$ because $\alpha = \alpha(n)$ is bounded away from 0.

Next, to show that w.h.p. some vertex has degree $k$ or less: note that any vertex with in-degree 0 in $M^\alpha_{n,k}$ necessarily has degree at most $k$, as all degrees come from in-degrees. By our conditioning device (see (2.3.3)),

$$P(\forall j, D_{n,j} > 0) \leq \sqrt{n} P(\forall j, Z_{n,j} > 0) = \sqrt{n} \left(1 - \left(\frac{\alpha}{\alpha + k}\right)^{\alpha\alpha}\right)^n \leq \sqrt{n}(1-e^{-k})^n \to 0$$

as $n \to \infty$, where “$\leq_b$” indicates that the left side is bounded by a constant times the right. So, w.h.p. there is a vertex of in-degree 0 and hence of degree at most $k$, as claimed.

Finally, we consider the distribution of the number of vertices of degree $k - 1$. To get there, let us consider binomial moments.

For fixed $t \in \mathbb{N}$, $\mathbb{E}\left[\binom{|V_n| - k}{t - 1}\right]$ is the expected number of sets $T$ of vertices, $|T| = t$, such that every vertex of $T$ has degree $k - 1$ in $G^\alpha_{n,k}$. Any such $T$ must be of one of the following types:

Type 1: All neighbors in $G^\alpha_{n,k}$ of the vertices in $T$ are outside of $T$, and are distinct; $M^\alpha_{n,k}$ contains no arcs directed from $[n] \setminus T$ to $T$.

Type 2: $M^\alpha_{n,k}$ contains no arcs directed from $[n] \setminus T$ to $T$, but there are two vertices in $T$ which either are neighbors or have a common neighbor in $[n] \setminus T$. 28
Type 3: $M_{n,k}^\alpha$ contains at least one arc directed from $[n] \setminus T$ to $T$.

Let $E_{n,i}$ denote the contribution to $\mathbb{E}[\binom{\left| V_{n,k-1} \right|}{t}]$ by sets $T$ of type $i$, so that

$$
\mathbb{E}\left[\binom{\left| V_{n,k-1} \right|}{t}\right] = E_{n,1} + E_{n,2} + E_{n,3}.
$$

Consider $E_{n,1}$. If $T$ is of type 1, what must $M_{n,k}^\alpha$ look like? Each $v \in T$ must choose exactly $k-1$ images in $[n] \setminus T$; either $v$ has a loop and each of the $k-1$ chosen images gets exactly one arc, or $v$ has no loop, one of the $k-1$ images gets 2 arcs, and the rest each get 1. Further, all arcs from $[n] \setminus T$ target $[n] \setminus T$.

If $E_{n,1}^{(\ell)}$ is the contribution to $E_{n,1}$ by $T$ where exactly $\ell$ vertices of $T$ have a loop, then

$$
E_{n,1}^{(\ell)} = \binom{n}{t} \frac{(n-t)!}{(n-k\ell)!} \left(\binom{k\alpha}{\ell} \frac{(k^{\alpha})^\ell}{(\alpha n)^{k\ell}} \right) \left(\frac{(k\alpha)^{k-1}(\alpha+1)^{t-\ell}}{(\alpha n + k\ell)^{k(t-\ell)}}\right) \frac{(\alpha(n-t) + kt - \ell)^{k(n-t)}}{(\alpha n + k\ell)^{k(n-t)}}
$$

$$
= \frac{1}{t!} \binom{t}{\ell} \frac{(\alpha_k + k)^{\ell}}{(\alpha_k + \alpha)^{\ell}} \left(\binom{k\alpha}{\ell} \frac{(k\alpha)^{k-1}(\alpha+1)^{t-\ell}}{(\alpha n + k\ell)^{k(t-\ell)}}\right). + O\left(\frac{1}{n}\right).
$$

Why? We choose our set of $t$ vertices; choose $t(k-1)$ distinct targets for them in $[n] \setminus T$; and choose which $\ell$ vertices of $t$ have loops. The first bracketed expression is the probability that the “loop” vertices do as they should; the second that the remaining vertices of $t$ behave, conditioned on the loop vertices’ choices; and the last that $[n] \setminus T$ chooses all images in $[n] \setminus T$, conditioned on the choices of $T$. The asymptotics follow by simplifying, some basic bounds, and applying Stirling’s formulas to rising factorials (similar to (4.5.1)). It is important to note that this last expression is bounded away from 0 for $\alpha$ bounded away from 0.

Summing over $\ell \in \{0, \ldots, t\}$ and applying the Binomial Theorem yields

$$
E_{n,1} = \frac{1}{t!} (\lambda(\alpha, k))^t + O\left(\frac{1}{n}\right).
$$

(3.2.1)

Next, consider $E_{n,2}$. If $T$ is of type 2, then there can be at most $(k-1)t - 1$ vertices in $[n] \setminus T$ which are adjacent to vertices in $T$; all other edges incident to $T$
are contained in $T$. Bounding $E_{n, 2}$ above by the expected number of $T$ satisfying this property, we find

$$E_{n, 2} \leq \binom{n}{t} \frac{(n-t)}{(k-1)t-1} \frac{\alpha(kt-1)^{kt}}{(\alpha n)^{kt}}$$

so that

$$E_{n, 2} \leq \frac{n^{kt-1}}{t!((k-1)t-1)!} \cdot \frac{((\alpha + 1)kt)^{kt}}{(\alpha n)^{kt}} = O\left(\frac{1}{n}\right) \quad (3.2.2)$$

as $n \to \infty$. The first inequality follows from the fact that there are a total of $kt - 1$ possible targets for the $kt$ arcs from $T$; the second from the observation $a^b \leq a^b \leq (a + b)^b$.

Finally, consider $E_{n, 3}$. If $T$ is of type 3, then necessarily there exists some vertex $v \in [n] \setminus T$ which chooses an image in $T$, and there is a set $B$ of size $0 \leq b \leq t(k-1) - 1$ such that all out-arcs from $T$ land in $T \cup \{v\} \cup B$. Bounding $E_{n, 3}$ by the expected number of such configurations, and over-counting, yields

$$E_{n, 3} \leq \binom{n}{t} (n-t) \left(1 - \frac{(\alpha(n-t))^r}{(\alpha n)^k}\right)^{t(k-1)-1} \sum_{b=0}^{n-t-1} \binom{n-t-1}{b} \frac{(\alpha(t+1+b)+k)^{kt}}{(\alpha n+k)^{kt}}.$$  

Why? We choose $T$ and a neighbor $v$ for $T$; require that $v$ have an image in $T$; then choose possible additional images for $T$, noting that only $k$ choices have been made so far. Bounding the above yields

$$E_{n, 3} = O\left(n^{t+1} \cdot \frac{1}{t!} \cdot \sum_{b=0}^{n-t-1} \frac{n^b}{n^{kt}}\right) = O\left(\frac{1}{n}\right). \quad (3.2.3)$$

Combining equations (3.2.1), (3.2.2), and (3.2.3), we see that

$$\mathbb{E}\left[\binom{|V_{n,k-1}|}{t}\right] = \frac{(\lambda(\alpha, k))^t}{t!} + O\left(\frac{1}{n}\right). \quad (3.2.4)$$
Now, by Bonferroni’s inequalities (see, for instance, the textbook [14] by Durrett), for any fixed \( \ell \in \mathbb{N} \) we have

\[
\sum_{t=j}^{j+2\ell+1} (-1)^{t-j} \binom{t}{j} \mathbb{E} \left[ \left( \frac{|V_{n,k-1}|}{t} \right) \right] \leq P(|V_{n,k-1}| = j) \leq \sum_{t=j}^{j+2\ell} (-1)^{t-j} \binom{t}{j} \mathbb{E} \left[ \left( \frac{|V_{n,k-1}|}{t} \right) \right].
\]

Subtracting \( e^{-\lambda(\alpha,k)} \frac{(\lambda(\alpha,k))^j}{j!} \) from each side and taking limits supremum and infimum yields

\[
- \limsup_{n \to \infty} \left| \sum_{t=j+2\ell+2}^{\infty} \frac{(-\lambda(\alpha,k))^{t-j}}{(t-j)!} \right| \leq \liminf_{n \to \infty} P(|V_{n,k-1}| = 0) - e^{-\lambda(\alpha,k)} \leq \limsup_{n \to \infty} P(|V_{n,k-1}| = 0) - e^{-\lambda(\alpha,k)} \leq \limsup_{n \to \infty} \left| \sum_{t=j+2\ell+1}^{\infty} \frac{(-\lambda(\alpha,k))^{t-j}}{(t-j)!} \right|.
\]

Noting that \( \lambda(\alpha,k) \) is bounded uniformly in \( n \), letting \( \ell \to \infty \) shows that

\[
P(|V_{n,k-1}| = j) - e^{-\lambda(\alpha,k)} \frac{(\lambda(\alpha,k))^j}{j!} \to 0 \text{ as } n \to \infty,
\]
as claimed. That \( V_{n,k-1} \) is (conditioned on its size) chosen uniformly at random is immediate by symmetry.

This answers the question of the minimum vertex degree: although it may possibly fall short of \( k \), the minimum degree will w.h.p. be at least \( k - 1 \), and the number of vertices of degree \( k - 1 \) very small.

### 3.3 Vertex Connectivity of \( G^n_{\alpha,n,k} \)

Given any random graph model, one very common question to address is the vertex connectivity of the resulting graph. We say that a graph \( G \) is \( h \)-vertex connected (or,
just $h$-connected) if removing any set of fewer than $h$ vertices from $G$, along with any edges incident to them, yields a connected graph.

Note that if $G$ is $h$-connected for some $h \geq 2$, then $G$ is also $(h - 1)$-connected. On the other hand, a minimum requirement for $h$-connectivity is that every vertex has degree at least $h$; otherwise, deleting the (at most $h - 1$) neighbors of such a vertex disconnects the graph. So, the set of $h$ such that $G$ is $h$-connected is bounded above by minimum degree, which is finite. Combining these, there must be some finite maximum $h$ such that $G$ is $h$-connected; this $h$ is called the vertex connectivity of $G$, denoted $C_v(G)$. If you think of the vertices of a graph as communication hubs in a network, vertex connectivity is a measure of their redundancy - how many of them can be destroyed without harming the function of the network.

As discussed in Section 1.1, a common theme in random graphs is that graphs with minimum degree $h$ are most likely also $h$-connected. This was, in fact, part of the motive for Fenner and Frieze [19] to study their $k$-out model - to see if these minimum degree conditions were still “almost sufficient” in this new setting.

Their results do not translate directly to our model, as their model does not permit multiple arcs in the directed version, whereas our does. We shall see, however, that the heuristic relationship between minimum degree and vertex connectivity still holds; further, we have the result that $G_{n,1}^\alpha$ (for $\alpha$ bounded away from 0) is w.h.p. not connected, as a corollary to a result by Hansen and Jaworski [27, Theorem 2(i)].

**Theorem 3.3.1.** For a graph $G$, let $\delta(G)$ denote the minimum vertex degree and $C_v(G)$ denote the vertex connectivity of $G$. Let $k \geq 2$ be fixed. Suppose that $\alpha = \alpha(n)$ is bounded away from 0, so that $\epsilon := \inf\{\alpha(n) : n \in \mathbb{N}\} > 0$. Then

$$P(C_v(G_{n,k}^\alpha) = \delta(G_{n,k}^\alpha)) = 1 - O\left(\frac{1}{n^{k-1}}\right) \text{ as } n \to \infty.$$ 

**Note:** The proof of this result is somewhat involved; so, for readability, let us outline the argument here, then fill in the details after with a series of lemmas.
Outline of Proof. Given a graph $G$, a set $P$ of vertices of $G$ is called a separator set of $G$ if the graph obtained from $G$ by removing the vertices of $P$ and all associated edges is not connected. Separators are intimately tied to vertex connectivity; in fact, $G$ is $h$-connected if and only if $G$ contains no separator set of size $h - 1$. (see [5, Section 7.2]).

If $v$ is a vertex of degree $d$, then the neighbors $N(v)$ of $v$ form a separator set of size $d$; Bollobás refers to separators of this type as trivial separators in [5]. All trivial separators have size at least $\delta(G)$, and at least one has size exactly $\delta(G)$; so, the event that $C_v(G) = \delta(G)$ is precisely the event that $G$ has no non-trivial separator set of size $\delta(G) - 1$ or less.

Since we showed in Theorem 3.2.1 that w.h.p. $\delta(G^\alpha_{n,k}) \in \{k - 1, k\}$, it suffices to show that with sufficiently high probability, $G^\alpha_{n,k}$ contains no non-trivial separator sets of size $k - 1$ or less; that is, any separator $P$ of size $p \leq k - 1$ such that the smallest connected component in the graph $G \setminus P$ has size at least 2. To do this, we introduce the following object:

**Definition.** Let $G$ be a graph with vertex set $[n]$. A $(t,p)$-separation of $G$ is a triple $(T,P,R)$ such that:

- $|T| = t$ and $|P| = p$.
- $T \cap P \cap R = \emptyset$.
- $T \cup P \cup R = [n]$.
- There are no edges in $G$ between $T$ and $R$.

In other words, $P$ is a separator of $G$ of size $p$, which separates $T$ from $R$. Note that the $(1,p)$-separations correspond to vertices of degree at most $p$. 

33
1
2
3
4
5
6
7
8

T

P

R

Figure 3.1: Example of a (3, 1)-separation: $T = \{1, 2, 3\}$, $P = \{4\}$, $R = \{5, 6, 7, 8\}$.

We will proceed by showing that with probability $1 - O(n^{-(k-1)})$, there are no $(t,p)$-separations of $G_{n,k}$ where $p \leq k - 1$ and $t \geq 2$. Note that $(T, P, R)$ is a $(t,p)$-separation if and only if $(R, P, T)$ is a $(n - t - p, p)$-separation; hence we need only consider $t \leq \frac{n-p}{2}$. We do this in three ranges:

- "Small" separations $2 \leq t \leq n^{2/5}$: that $G_{n,k}$ admits no small separations with probability $1 - O(n^{-(k-1)})$ is established in Lemma 3.3.2.

- "Medium" separations $n^{2/5} < t \leq \delta n$: we establish a specific constant $\delta = \delta(\epsilon, k) > 0$ in Lemma 3.3.3, and establish in Lemma 3.3.4 that with probability $1 - O(n^{-(k-1)})$ there are no medium separations of $G_{n,k}$.

- "Large" separations $\delta n < t \leq \frac{n-p}{2}$: we split into two cases. In Lemma 3.3.5 and Corollary 3.3.6, we establish a constant $C > 0$ and show that, with probability $1 - O(n^{-(k-1)})$, $G_{n,k}$ admits no large separations $(T, P, R)$ such that more than $Cp \log n$ arcs go from $T$ to $P$ in $M_{n,k}$. In Lemma 3.3.7, we establish that the probability that there exists a large separation $(T, P, R)$ such that at most $Cp \log n$ arcs go from $T$ to $P$ in $M_{n,k}$ is exponentially small.

This completes the proof. $\square$
For the rest of the section, let $S_{n,t}$ denote the number of $(t,p)$-separations of $G_{n,k}^\alpha$ (where we consider $p$ as fixed, $p \leq k - 1$). Let us begin by dismissing the existence of $(t,p)$-separations where $t$ is "small"; this will allow easier asymptotic analysis later.

**Lemma 3.3.2.** Let $k \geq 2$. Suppose that $\alpha = \alpha(n)$ is bounded away from 0 and $p \leq k - 1$ is fixed. Then with probability $1 - O(n^{-(k-1)})$, there are no $(t,p)$-separations of $G_{n,k}^\alpha$ with $2 \leq t \leq n^{2/5}$.

**Proof.** If $(T,P,R)$ is a bad $(t,p)$-separation of $G_{n,k}^\alpha$, then necessarily all out-arcs of $T$ in $M_{n,k}^\alpha$ land in $T \cup P$. Bounding $\mathbb{E}[S_{n,t}]$ above by the expected number of such arrangements yields

$$\mathbb{E}[S_{n,t}] \leq \binom{n}{p} \binom{n-p}{t} \frac{(\alpha(t+p))^{kt}}{(\alpha n)^{kt}} =: b_t.$$

Taking ratios of successive terms for $2 \leq t \leq n^{2/5}$, we find that uniformly over such $t$,

$$\frac{b_{t+1}}{b_t} = \frac{n-p-t}{t+1} \cdot \frac{(\alpha(t+p+1))^{kt}}{(\alpha(t+p))^{kt}} \cdot \frac{(\alpha(t+p+1) + kt)^k}{(\alpha n + kt)^k} \leq n \cdot \left(\frac{t+p+1}{t+p}\right)^{kt} \cdot \left(\frac{(\alpha + k)(t + 1) + \alpha p}{\alpha n + kt}\right)^k$$

$$= O\left(\frac{n}{n^{3k/5}}\right) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

so that by comparison with a geometric series

$$\sum_{t=2}^{n^{2/5}} \mathbb{E}[S_{n,t}] = O(b_2) = O\left(\frac{1}{n^{2(k-1)-p}}\right) = O\left(\frac{1}{n^{k-1}}\right),$$

proving the result.

Next, we handle $n^{2/5} < t \leq \delta n$. We will squeeze this result out of the same bound used in Lemma 3.3.2; unfortunately, that bound is too crude to get all the way up to $\frac{n-p}{2}$ directly. However, being able to restrict our attention to $t$ linear in $n$ will allow us to use heavy asymptotic machinery to great effect later on.
Of course, we must start by defining $\delta$. To eliminate any potential ugly behavior caused by variations in $\alpha$, we will define $\delta$ as depending only on $k$ and $\epsilon := \inf\{\alpha(n) : n \in \mathbb{N}\}$.

**Lemma 3.3.3.** Let $\epsilon > 0$ and $k \in \mathbb{N}$ be fixed. Then there exists $\delta = \delta(\epsilon, k) \in (0, \frac{1}{2}]$ such that

i) $\delta$, $\epsilon$, and $k$ satisfy the inequality

$$\frac{\delta \epsilon^{1-\delta} (k + \epsilon)^{\delta(k+\epsilon)}}{\delta^2 (1 - \delta)^{1-\delta} (\delta k + \epsilon)^{\delta k+\epsilon}} < 1.$$  

(3.3.1)

ii) For $n$ sufficiently large, the function $g_n(t)$ is concave up for $t \in [n^{2/5}, \delta n]$, where $g_n(t)$ is defined for $t \in \mathbb{N}$ by

$$g_n(t) := \left(\begin{array}{c} n \\ p \end{array}\right) \left(\begin{array}{c} n - p \\ t \end{array}\right) \frac{(\epsilon(t + p))^{t+1}}{(\epsilon n)^{(k+1)t}}$$

and extended to real $t$ by way of the $\Gamma$-function.

**Proof.** First, for (i): letting $\delta \to 0$, the left side of (3.3.1) is

$$\exp [(k - 1)\delta \log \delta + O(\delta)] < 1$$

for $\delta > 0$ but sufficiently small. So, there is $M = M(\epsilon, k) > 0$ such that (3.3.1) is satisfied for $\delta \in (0, M)$.

It remains to show that there is $\delta \in (0, M)$ such that, for $n$ sufficiently large, $g_n''(t) > 0$ for $t \in [n^{2/5}, \delta n]$; to that end, it suffices to show that $\log g_n(t)$ is convex. So, we compute

$$\frac{d^2}{dt^2} \left[ \log g_n(t) \right] = -\psi'(t + 1) - \psi'(n - t + 1)$$

$$+ (\epsilon + k)^2 \psi'((\epsilon + k)t + \epsilon p) - \epsilon^2 \psi'(\epsilon(t + p)) - k^2 \psi'(\epsilon n + kt),$$

where $\psi$ is the logarithmic derivative of $\Gamma$. Applying Alzer’s asymptotics for $\psi'$ [1], we find that that:
i) Uniformly over \( n^{2/5} \leq t \leq n^{2/3} \),
\[
\frac{d^2}{dt^2} \left[ \log g_n(t) \right] = \frac{k-1}{t} + O \left( \frac{1}{n^{4/5}} \right) \geq 0
\]
for \( n \) sufficiently large.

ii) If \( \delta > 0 \) is constant and satisfies
\[
\delta < \frac{\epsilon(k-1)}{(1+\epsilon)k}
\]
then uniformly over \( n^{2/3} \leq t \leq \delta n \),
\[
\frac{d^2}{dt^2} \left[ \log g_n(t) \right] = -\frac{1}{t} - \frac{1}{n-t} + \frac{\epsilon+k}{t} - \frac{\epsilon}{t} - \frac{k^2}{\epsilon n + kt} + O \left( \frac{1}{n^{4/3}} \right)
\]
\[
= \frac{k(\epsilon+1)n^2}{t(n-t)(\epsilon n + kt)} \left( \frac{\epsilon(k-1)}{k(\epsilon+1) - \frac{t}{n}} + \frac{t}{n} \right) + O \left( \frac{1}{n^{4/3}} \right)
\]
\[
\geq 0
\]
for \( n \) sufficiently large, since the last expression is \( \Theta \left( \frac{1}{n} \right) \) and positive. Choose \( \delta \) smaller than both \( M \) and \( \frac{\epsilon(k-1)}{(1+\epsilon)k} \).

\( \square \)

**Lemma 3.3.4.** Let \( k \geq 2 \) be fixed. Suppose \( \epsilon := \inf \{ \alpha(n) : n \in \mathbb{N} \} > 0 \). Let \( \delta = \delta(k, \epsilon) \) be as in Lemma 3.3.3. Then the probability that there exists a \( (t,p) \)-separation of \( G_{n,k}^{\alpha} \) with \( n^{2/5} \leq t \leq \delta n \) tends to 0 exponentially fast.

**Proof.** As in the proof of Lemma 3.3.2, we have
\[
\mathbb{E}[S_{n,t}] \leq \binom{n}{p} \binom{n-p}{t} \frac{(\alpha(t+p))^{kt}}{(\alpha n)^{kt}}.
\]
Note that
\[
\frac{(\alpha(t+p))^{kt}}{(\alpha n)^{kt}} = \prod_{j=0}^{kt-1} \frac{\alpha(t+p) + j}{\alpha n + j},
\]
each term of this product is decreasing as a function of \( \alpha \) as long as \( t + p < n \). So, because \( \alpha = \alpha(n) \geq \epsilon \) for all \( n \),
\[
\mathbb{E}[S_{n,t}] \leq \binom{n}{p} \binom{n-p}{t} \frac{(\epsilon(t+p))^{kt}}{\epsilon n^{kt}} =: g_n(t).
\]
By Lemma 3.3.3, \( g_n(t) \) is concave up on \([n^{2/5}, \delta n]\); so, for all \( t \) in this range,

\[
g_n(t) \leq \max\{g_n(n^{2/5}), g_n(\delta n)\} \leq g_n(n^{2/5}) + g_n(\delta n)
\]

and

\[
\sum_{t=n^{2/5}}^{\delta n} \mathbb{E}[S_{n,t}] = O\left(n g_n(n^{2/5}) + n g_n(\delta n)\right).
\]

Applying Stirling’s formula yields

\[
n g_n(n^{2/5}) \sim M \left(\frac{\epsilon + k}{\epsilon}\right)^{(\epsilon + k)n^{2/5}} \frac{1}{e^{(k-1)n^{2/5}} n^{\frac{5}{2}(k-1)n^{2/5} - \frac{p}{5}}},
\]

where \( M = M(\epsilon, k) \) is constant; this tends to 0 exponentially fast because \( k \geq 2 \). An additional application of Stirling’s formula yields

\[
n g_n(\delta n) \sim M \frac{n^{p+1}}{\sqrt{n}} \left[ \frac{\delta^k \epsilon^{(1-\delta)k}(k + \epsilon)^{\delta(k+\epsilon)}}{\delta^k (1 - \delta)^{1-\delta} (\delta k + \epsilon)^{\delta k + \epsilon}} \right] \to 0 \text{ as } n \to \infty
\]

exponentially fast, where \( M = M(\epsilon, k) \) is a constant (different from the previous \( M \), but still dependent only on \( \epsilon \) and \( k \)) and we use the fact that \( \delta \) makes the bracketed term positive but smaller than 1.

Next, we try to eliminate \( \delta n < t \leq \frac{n-p}{2} \). Trying to simply count the number of bad \((t, p)\)-separations of \( G_{n,k}^\alpha \) where \( t \) is large can get us in to trouble if we allow the set \( T \) to choose a large number of images in \( P \); in this case, many subsets of \( T \) have no occupancy, and so could be moved to \( R \) to form another separation. This means that we over-count drastically in this situation!

The remedy is to show that the total occupancy of \( P \) is, with sufficiently high probability, fairly small; we do this by bounding the maximum in-degree \( \Delta_{in}(M_{n,k}^\alpha) \):

**Lemma 3.3.5.** For a digraph \( M \), let \( \Delta_{in}(M) \) denote the maximum vertex in-degree in \( M \). Suppose that \( \epsilon := \inf\{\alpha(n) : n \in \mathbb{N}\} > 0 \). Let \( C = C(\epsilon, k) \) be defined by

\[
C = k \left(\log \left[\frac{\epsilon + k}{k}\right]\right)^{-1}.
\]
Then

$$P(\Delta_{in}(M_{n,k}^\alpha) < C \log n) = 1 - O\left(\frac{1}{n^{k-1}}\right).$$

Proof. For $\ell \in \mathbb{N}$, let $X_n(\ell)$ denote the number of vertices of $M_{n,k}^\alpha$ which have in-degree exactly $\ell$. Then, by Markov’s inequality,

$$P(\Delta_{in}(M_{n,k}^\alpha) \geq C \log n) \leq \sum_{\ell = C \log n}^{kn} \mathbb{E}[X_n(\ell)] = \sum_{\ell = C \log n}^{kn} n\left(\frac{k}{\ell}\right) \frac{\alpha^\ell(\alpha(n-1))^{kn-\ell}}{(\alpha n)^{kn}}.$$

Note that

$$\frac{\alpha^\ell(\alpha(n-1))^{kn-\ell}}{(\alpha n)^{kn}} = \left(\prod_{j=0}^{\ell-1} \frac{\alpha + j}{\alpha n + j}\right) \left(\prod_{j=0}^{kn-\ell-1} \frac{\alpha(n-1) + j}{\alpha n + \ell + j}\right);$$

each of these fractions decreases as a function of $\alpha$, so that

$$n\left(\frac{k}{\ell}\right) \frac{\alpha^\ell(\alpha(n-1))^{kn-\ell}}{(\alpha n)^{kn}} \leq n\left(\frac{k}{\ell}\right) \frac{\bar{\epsilon}^\ell(\bar{\epsilon}(n-1))^{kn-\ell}}{(\bar{\epsilon} n)^{kn}} =: h_n(\ell).$$

Extend $h_n$ to a function of continuous $\ell$ by rewriting in terms of the $\Gamma$-function. Note that for $n$ sufficiently large, uniformly over $\ell \geq C \log n$ we have

$$h'_n(\ell) = \psi(kn - \ell + 1) - \psi(\ell + 1) + \psi(\ell + \ell) - \psi((\epsilon + k)n - \epsilon - \ell)$$

$$= -\log\left(\frac{\epsilon + k}{k}\right) + O\left(\frac{\log n}{n}\right)$$

$$\leq -\frac{1}{2} \log\left(\frac{\epsilon + k}{k}\right)$$

where $\psi$ is the logarithmic derivative of the $\Gamma$-function, by Alzer’s sharp asymptotics for $\psi$ [1]. This last expression shows that $h'_n(\ell)$ is bounded below 0, for $n$ sufficiently large; thus, for such $n$,

$$\mathbb{E}[X_n(\ell)] = \exp[h_n(\ell)] \leq \exp\left[h_n(C \log n) - \frac{1}{2} \log\left(\frac{\epsilon + k}{k}\right) (\ell - C \log n)\right],$$

so that

$$\sum_{\ell = C \log n}^{kn} \mathbb{E}[X_n(\ell)] \leq \frac{\exp[h_n(C \log n)]}{1 - \sqrt{\frac{k}{\epsilon + k}}}.$$
Applying Stirling’s formula yields
\[
\exp[h_n(C \log n)] \leq b \exp \left[ \log n + (\epsilon - 1) \log(C \log n) - C \log n \log \left( \frac{\epsilon + k}{k} \right) \right]
\leq \exp \left[ - \left( C \log \left( \frac{\epsilon + k}{k} \right) - 1 \right) \log n \right]
= O \left( \frac{1}{n^{k-1}} \right),
\]
proving the claim.

A direct corollary of this is precisely what we want:

**Corollary 3.3.6.** Suppose \( \alpha = \alpha(n) \geq \epsilon > 0 \) for all \( n \). Let \( C = C(\epsilon, k) \) as in Lemma 3.3.5, and \( \delta = \delta(\epsilon, k) \) as in Lemma 3.3.3. Then with probability \( 1 - O(n^{-(k-1)}) \), there are no \((t, p)\)-separations \((T, P, R)\) of \( G_{n,k}^{\alpha} \), with \( \delta n \leq t \leq \frac{n-p}{2} \), such that \( T \) chooses \( Cp \log n \) or more images in \( P \) in \( M_{n,k}^{\alpha} \).

**Proof.** If there were such a separation, then \( T \) would have in-degree at least \( Cp \log n \) in \( M_{n,k}^{\alpha} \), requiring that at least one element of \( T \) has in-degree of \( Cp \log n \) or more. \( \square \)

Finally, we are at the end of the argument; here, we must be more careful than before. For \( \ell \in [kt] \), let \( S_{n,t,\ell} \) denote the number of \((t, p)\)-separations of \( G_{n,k}^{\alpha} \) in which there are precisely \( \ell \) arcs from \( T \) to \( P \) in \( M_{n,k}^{\alpha} \). Then
\[
\mathbb{E}[S_{n,t,\ell}] := \binom{n}{p} \binom{n-p}{t} \binom{kt}{\ell} \frac{(\alpha p)^t(\alpha t)^{kt-\ell}(\alpha(n-t) + \ell)^{k(n-t-p)}}{(\alpha n)^{k(n-p)}}. \tag{3.3.2}
\]
Why? Given \( T \) and \( P \), and which set of \( \ell \) out-arcs from \( T \) land in \( P \), the total weight of valid targets for the \( \ell \) balls is \( \alpha p \); conditioned on this, the total weight of valid targets (in \( T \)) for the remaining \( kt - \ell \) arcs out of \( T \) is \( \alpha t \); and conditioned on this, the total weight of valid targets for \( R = [n] \setminus (T \cup P) \) is \( \alpha(n-t) + \ell \).

In Corollary 3.3.6, we dismissed the existence of \((t, p)\)-separations in which \( \ell > Cp \log n \). So, all that remains to consider is the existence of \((t, p)\)-separations with
δn ≤ t ≤ \frac{n-p}{2} in which \ell < Cp \log n arcs go from T to P in \( M_{α,n,k}^α \). Fortunately, we now have a great deal of control over the relationships between our parameters, allowing us to use Alzer’s asymptotic approximations to the polygamma functions [1] to great effect:

**Lemma 3.3.7.** Suppose that \( α = α(n) ≥ ε > 0 \) for all \( n \). Let \( C = C(ε, k) \) be as in Lemma 3.3.5 and \( δ = δ(ε, k) \) as in Lemma 3.3.4. Then with probability \( 1 - O(n^{-(k-1)}) \), \( G_{α,n,k}^α \) admits no \((t,p)\)-separations such that \( δn ≤ t ≤ \frac{n-p}{2} \) and \( M_{α,n,k}^α \) contains fewer than \( Cp \log n \) arcs from \( T \) to \( P \).

**Proof.** Let \( f_n(t, \ell) := \log \mathbb{E}[S_{n,t,\ell}] \). Rewrite \( f_n(t, \ell) \) in terms of the \( Γ \)-function, so that it can be considered as a function of continuous \( t \). Uniformly over \( t ∈ [δn, \frac{n-p}{2}] \) and \( \ell ≤ Cp \log n \), Alzer’s asymptotics for \( \log Γ \) [1] yield

\[
\frac{∂}{∂t}[f_n(t, \ell)] = -k \log k - (k - 1) \log \left( \frac{n-t}{t} \right) + O \left( \frac{\ln n}{n} \right),
\]

so that for \( n \) sufficiently large, all such \( t \) and \( \ell \) have

\[
\frac{∂}{∂t}[f_n(t, \ell)] ≤ -\frac{k}{2} \log k.
\]

It follows that for \( n \) sufficiently large, uniformly over \( \ell ≤ Cp \log n \),

\[
\sum_{t=δn}^{\frac{n-p}{2}} \mathbb{E}[S_{n,t,\ell}] ≤ \sum_{t=δn}^{\frac{n-p}{2}} \exp \left[ f_n(δn, \ell) - \frac{k}{2} \log k(t-δn) \right] ≤ \frac{\exp[f_n(δn, \ell)]}{1 - \frac{1}{\frac{gy}{2}}}.
\]

But, for \( g_n \) defined as in Lemma 3.3.3,

\[
\sum_{\ell=0}^{Cp \log n} \exp[f_n(δn, \ell)] ≤ \sum_{\ell=0}^{kn} \frac{n}{p} \binom{n-p}{\delta n} \binom{kδn}{\ell} \frac{(αp)^{kδn-\ell}}{(αn)^{kδn}}
\]

\[
= \binom{n}{p} \binom{n-p}{\delta n} \frac{(α(t+p))^{kδn}}{(αn)^{kδn}}
\]

\[
≤ g_n(δn) → 0 \text{ as } n → ∞
\]

exponentially quickly, as in the proof of Lemma 3.3.4. So, the expected number of \((t,p)\)-separations with \( δn ≤ t ≤ \frac{n-p}{2} \) such that \( T \) sends no more than \( Cp \log n \) arcs to \( P \) tends to 0, so that Markov’s inequality completes the proof. ∎
3.4 The $k$-Core of $G_{n,k}^\alpha$

For a graph $G$ and $h \in \mathbb{N}$, the $h$-core of $G$ is defined as the maximal induced subgraph of $G$ that has minimum degree $h$. It is the graph obtained from $G$ as follows: delete all vertices of $G$ which have degree $h-1$ or less. If any vertices lose enough edges in this way to have degree $h-1$ or less in the resulting graph, delete them too. Continue with this process until the resulting subgraph is either empty or has minimum degree $h$. What is left, if anything, is the $h$-core; if nothing is left, then we say that $G$ does not have an $h$-core.

Interestingly, it is possible for a graph which has only one vertex of degree $h-1$ to have absolutely no $h$-core; the removal of that vertex could cause all of its neighbors to have degree at most $h-1$, and so on. So, the $h$-core can be thought of as not only a measure of the extent to which the graph fails to have minimum degree $h$, but also of the fragility of the minimum degree condition.

A consequence of Theorem 3.2.1 is that w.h.p. $G_{n,k}^\alpha$ has minimum degree either $k-1$ or $k$; so, in particular, $G_{n,k}^\alpha$ is w.h.p. its own $(k-1)$-core. Given the fact that we wish to compare our model with the $k$-out models studied by Frieze et. al, in which the minimum degree is at least $k$, it is crucial that we be able to measure how far off of $G_{n,k}^\alpha$ is from having minimum degree $k$.

We have the following explicit description of the $k$-core of $G_{n,k}^\alpha$:

**Theorem 3.4.1.** Suppose that $\epsilon := \sup\{\alpha(n) : n \in \mathbb{N}\} > 0$. Let $V_{n,k-1}$ denote the set of vertices of degree exactly $k-1$ in $G_{n,k}^\alpha$, and let $K_n$ denote the $k$-core of $G_{n,k}^\alpha$. Then w.h.p., the $k$-core of $G_{n,k}^\alpha$ is precisely its induced subgraph on $[n] \setminus V_{n,k-1}$.

**Proof.** Let $K_n$ be the induced subgraph of $G_{n,k}^\alpha$ on the vertex set $[n] \setminus V_{n,k-1}$. Then the $k$-core must be a subgraph of $K_n$, as the vertices of $V_{n,k-1}$ are all deleted in the
first round of the algorithm for determining the $k$-core. So, we seek to show that w.h.p., every vertex of $K_n$ has degree at least $k$.

By Theorem 3.2.1, w.h.p. $G_{n,k}^\alpha$ contains no vertex of degree $k-2$ or less. So, w.h.p. any vertex in $K_n$ with degree smaller than $k$ must have had degree at least $k$ in $G_{n,k}^\alpha$, and lost edges when the vertices of $V_{n,k-1}$ were removed.

How much can the degree of a vertex in $[n] \setminus V_{n,k-1}$ suffer from the deletion of $V_{n,k-1}$? As it turns out, the degree can decrease by at most one, as w.h.p. no two vertices of degree $k-1$ have a common neighbor. Why? Let $B_{n,1}$ denote the event that there are vertices of $V_{n,k-1}$ which have a common neighbor. If $v$ and $w$ are two such vertices in $V_{n,k-1}$, then all out-arcs from $v$ and $w$ target $v$, $w$, or one of their neighbors. There are at most $2k-3$ such neighbors to choose from, allowing the possibility that $v$ and $w$ have multiple neighbors in common.

Let us bound the probability that two vertices of degree $k-1$ have a common neighbor by the probability that there exist two adjacent vertices whose out-arcs land in at most $2k-3$ vertices other than themselves. Then by Markov’s inequality,

$$P(B_{n,1}) \leq \binom{n}{2} \sum_{a=1}^{2k-3} \binom{n-2}{a} \frac{\alpha(a+2))^{2k}}{(\alpha n)^{2k}} = O\left(\frac{1}{n}\right),$$

and therefore w.h.p. any vertex in $[n] \setminus V_{n,k-1}$ loses at most one degree when we delete $V_{n,k-1}$.

So, w.h.p., any vertex of degree $k-1$ or less in $K_n$ had degree exactly $k$ in $G_{n,k}^\alpha$, and was adjacent to exactly one vertex in $V_{n,k-1}$. Let $B_{n,2}$ denote the event that such a vertex exists. If $v$ has degree $k$ and $w$ has degree $k-1$, then necessarily the out-arcs from $v$ and $w$ in $M_{n,k}^\alpha$ all target a set of size at most $2k-3$: $v$ and $w$, at most $k-1$ additional neighbors of $v$, and at most $k-2$ additional neighbors of $w$. Bounding $P(B_{n,2})$ above by the probability of this less restrictive event, and again
applying Markov’s inequality, yields

\[ P(B_{n,2}) \leq (n)_2 \sum_{a=0}^{2k-3} \binom{n-2}{a} \frac{(\alpha(a+2))^{2k}}{(\alpha n)^{2k}} = O\left(\frac{1}{n}\right). \]

Putting this all together, we have now shown that \( K_n \), the induced subgraph of \( G_{n,k}^\alpha \) on vertex set \([n] \setminus V_{n,k-1}\), is w.h.p. the \( k \)-core of \( G_{n,k}^\alpha \). \( \square \)

We have now seen that removing the most obvious obstacle to having minimum degree \( k \) - namely the vertices of degree \( k - 1 \) - is sufficient to get a graph whose minimum degree is \( k \). Further, the number of vertices that we remove in this way is very small - it is Poisson-distributed with finite expectation \( \lambda(\alpha, k) \) as in Section 3.2.

So, after removing \( V_{n,k-1} \), we are left with very nearly all of the original vertices of \( G_{n,k}^\alpha \) but now have minimum degree \( k \). Further, we saw in the proof of Theorem 3.2.1 that w.h.p., there are no element of \( V_{n,k-1} \) has any edges caused by in-bound arcs. So, in the mapping \( M_{n,k}^\alpha \), restricting to \([n] \setminus V_{n,k-1}\) actually yields a \( k \)-out map on \( n - |V_{n,k-1}| \) vertices. Intuitively, it seems likely that this map would be structurally very similar to \( M_{n,k}^\alpha \) - and so, because the induced graph does have minimum degree \( k \), it seems reasonable that it should be \( k \)-connected. We will return to this in Section 5.2.
CHAPTER 4

RELATIONSHIP OF $M^\alpha_{n,k}$ TO THE UNIFORM MAP $M^\infty_{n,k}$

4.1 Introduction and Theorem Statement

In what follows, we will need to simultaneously refer to two different mappings, with different values for $\alpha$. We will also need to refer directly to the measures induced by $M^\alpha_{n,k}$ and $M^\infty_{n,k}$. So, we make the following definitions:

**Definition.** Where confusion is possible, let $D^\alpha_n = (D^\alpha_{n,1}, \ldots, D^\alpha_{n,n})$ denote the indegree sequence for $M^\alpha_{n,k}$.

*Note:* Where we deal with both $M^\alpha_{n,k}$ for some $\alpha < \infty$ and $M^\infty_{n,k}$, the sequence $Z_n = (Z_{n,1}, \ldots, Z_{n,n})$ will always be used in conjunction with $M^\alpha_{n,k}$.

**Definition.** Let $P^\alpha_{n,k}$ be the measure on $M_{n,k}$ induced by $M^\alpha_{n,k}$.

Throughout the previous chapters, we have discussed the uniform map $M^\infty_{n,k}$ as the limiting case of $M^\alpha_{n,k}$ where we let $\alpha \to \infty$. Further, in every result we have proved as yet, we have not needed to do anything special to handle the case $\alpha = \infty$; in some sense, all of the properties of $M^\alpha_{n,k}$ are “continuous” at $\alpha = \infty$, so that simply taking limits as $\alpha \to \infty$ of all equations, distributions, etc. yield the corresponding results for the uniform map.

Given this, it is only natural to ask the question: how large must $\alpha$ be in order to make the differences between $M^\alpha_{n,k}$ and $M^\infty_{n,k}$ insignificant? Is there a threshold
function for this behavior? Before we can answer any of these questions, we must decide how to measure the difference between $P_{\alpha n,k}$ and $P_{\infty n,k}$. We will quantify this in terms of the total variation distance:

**Definition.** Let $P_1$ and $P_2$ be two probability measures on the same discrete probability space $\Omega$. The total variation distance between $P_1$ and $P_2$, denoted $\|P_1 - P_2\|_{TV}$, is defined by

$$\|P_1 - P_2\|_{TV} := \sup_{A \subseteq \Omega} |P_1(A) - P_2(A)| = \frac{1}{2} \sum_{x \in \Omega} |P_1(x) - P_2(x)|.$$ 

**Note:** The first equality is the definition; the second is a simplification of the formula in the special case where $\Omega$ is discrete. To see that they agree, partition $\Omega$ in to three sets based on equality or inequality of $P_1(x)$ and $P_2(x)$.

The total variation distance is a very strong measure of the difference in the distributions $P_1$ and $P_2$ - given any measurable event $A \subseteq \Omega$, the two probability measures differ on $A$ by at most $\|P_1 - P_2\|_{TV}$. So, if $\|P_{\alpha n,k} - P_{\infty n,k}\|_{TV} \to 0$ as $n \to \infty$, then ALL properties of the preferential attachment map agree (in the limit) with those of the uniform map.

We will see that $\alpha = \Theta(\sqrt{n})$ marks a transition in the behavior of $M_{\alpha n,k}$:

**Theorem 4.1.1.** Let $k \in \mathbb{N}$ be fixed, and suppose $\alpha = \alpha(n) < \infty$ for all $n$.

i) If $\alpha/\sqrt{n} \to 0$ as $n \to \infty$, then

$$\lim_{n \to \infty} \|P_{\alpha n,k} - P_{\infty n,k}\|_{TV} = 1.$$ 

ii) If $\alpha/\sqrt{n} \to \infty$ as $n \to \infty$, then

$$\lim_{n \to \infty} \|P_{\alpha n,k} - P_{\infty n,k}\|_{TV} = 0.$$ 

46
iii) If \( \alpha = \beta \sqrt{n} \), where \( \beta > 0 \) is fixed, then

\[
\lim_{n \to \infty} \| P_{n,k}^{\alpha} - P_{n,k}^{\infty} \|_{TV} = \frac{1}{2} \mathbb{E} \left| 1 - \exp \left[ -\mathcal{N} \left( \frac{k^2}{4\beta^2}, \frac{k^2}{\beta^2} \right) \right] \right|,
\]

where \( \mathcal{N}(\mu, \sigma^2) \) denotes a normally-distributed random variable with mean \( \mu \) and variance \( \sigma^2 \).

We proceed to prove these three cases separately, after some preliminary results.

### 4.2 Preliminaries

First, let us establish some helpful estimates for rising factorials:

**Lemma 4.2.1.** Suppose \( a \in (0, \infty) \) and \( b \in \mathbb{Z} \cap [0, a) \). Then the rising factorial \( a^b \) satisfies the bounds

\[
\exp \left[ -\frac{b^3}{6a^2} \right] \leq \frac{a^b}{a^b \exp \left( \frac{b(b-1)}{2a} \right)} \leq 1
\]

and

\[
1 \leq \frac{a^b}{a^b \exp \left( \frac{b(b-1)2b-1}{12a^2} \right)} \leq \exp \left[ \frac{b^4}{12a^3} \right].
\]

**Proof.** Write

\[
a^b = a^b \exp \left[ \sum_{j=0}^{b-1} \log \left( 1 + \frac{j}{a} \right) \right].
\]

For \( x \in (0, 1) \), the power series for \( \log(1 + x) \) is alternating and its terms decrease in absolute value, so that \( \log(1 + x) \) is bounded between any two successive partial sums. From this, we get the bounds

\[
\frac{j}{a} - \frac{j^2}{2a^2} \leq \log \left( 1 + \frac{j}{a} \right) \leq \frac{j}{a}
\]

and

\[
\frac{j}{a} - \frac{j^2}{2a^2} \leq \log \left( 1 + \frac{j}{a} \right) \leq \frac{j}{a} - \frac{j^2}{2a^2} + \frac{j^3}{3a^3}
\]

for \( 0 \leq j \leq b - 1 < a \). The results follow by applying these bounds and the formulas for sums of consecutive integers, squares, and cubes. \( \square \)
The following corollary to Lemmas 2.3.1 and 2.3.3 will be helpful in simplifying asymptotic estimates:

**Corollary 4.2.2.** Suppose that $\alpha = \alpha(n)$ is bounded away from 0; further, suppose $\omega = \omega(n) \to \infty$ as $n \to \infty$. Fix $s \in \mathbb{N}$. Then, \textit{w.h.p.}, the in-degree sequence $D_{n}^{\alpha} = (D_{1}^{\alpha}, \ldots, D_{n,n}^{\alpha})$ of $M_{n,k}^{\alpha}$ satisfies

$$
| (D_{n,1}^{\alpha})^{s} + \cdots + (D_{n,n}^{\alpha})^{s} - \mu_{s,\alpha} n | < \omega \sqrt{n},
$$

where

$$
\mu_{s,\alpha} := \mathbb{E}[(D_{n,1}^{\alpha})^{s}] = \sum_{\ell=1}^{s} \left\{ \frac{s}{\ell} \frac{\alpha^{\ell}(kn)^{\ell}}{(\alpha n)^{\ell}} \right\}. 
$$

**Proof.** Note that

$$
\mathbb{E}[(D_{n,1}^{\alpha})^{s} + \cdots + (D_{n,n}^{\alpha})^{s}] = n \mathbb{E}[(D_{n,1}^{\alpha})^{s}] = \mu_{s,\alpha} n.
$$

Further,

$$
\mathbb{E}[(D_{n,1}^{\alpha})^{s} + \cdots + (D_{n,n}^{\alpha})^{s} + (D_{n,1}^{\alpha})^{s}] = n \mathbb{E}[(D_{n,1}^{\alpha})^{s}] = n(n - 1) \mathbb{E}[(D_{n,1}^{\alpha})^{s}] = 0(n),
$$

so that

$$
\mathbb{E} \left[ \left( \sum_{j=1}^{n} (D_{n,j}^{\alpha})^{s} \right)^{2} \right] = n^{2} \mathbb{E} \left[ \left( \sum_{m=1}^{s} \binom{s}{m} (D_{n,1}^{\alpha})^{m} \right) \left( \sum_{t=1}^{s} \binom{s}{t} (D_{n,2}^{\alpha})^{t} \right) \right] + O(n)
$$

$$
= n^{2} \sum_{m=1}^{s} \sum_{t=1}^{s} \binom{s}{m} \binom{s}{t} \mathbb{E}[(D_{n,1}^{\alpha})^{m}(D_{n,2}^{\alpha})^{t}] + O(n)
$$

$$
= n^{2} \sum_{m=1}^{s} \sum_{t=1}^{s} \binom{s}{m} \binom{s}{t} \mathbb{E}[(D_{n,1}^{\alpha})^{m}] \mathbb{E}[(D_{n,2}^{\alpha})^{t}] + O(n)
$$

$$
= (\mu_{s,\alpha} n)^{2} + O(n)
$$

by Lemma 2.3.3. Combining these yields

$$
\text{Var}[(D_{n,1}^{\alpha})^{s} + \cdots + (D_{n,n}^{\alpha})^{s}] = O(n),
$$

48
and so by Chebyshev’s inequality

\[ P\left(\left|\left(D_{n,1}\right)^{\alpha} + \cdots + \left(D_{n,n}\right)^{\alpha} - \mu_{s,\alpha} n\right| \geq \omega \sqrt{n}\right) = O\left(\frac{1}{\omega^2}\right) \rightarrow 0 \text{ as } n \rightarrow \infty. \]

4.3 Proof for the case \( \alpha/\sqrt{n} \rightarrow 0 \)

Let us write \( \alpha = \sqrt{n}/\omega; \) note that \( \omega \rightarrow \infty \) as \( n \rightarrow \infty. \)

For \( n \in \mathbb{N}, \) define

\[ G_n := \{ \mathbf{d} : d_1 + \cdots + d_n = kn \text{ and } \left|d_1^2 + \cdots + d_n^2 - \mu_{2,\alpha} n\right| < \sqrt{\omega n}\}, \]

where

\[ \mu_{2,\alpha} := \mathbb{E}\left[(D_{n,1}^{\alpha})^2\right] = \frac{\alpha^2 (kn)^2}{(\alpha n)^2} + k. \]

By Corollary 4.2.2, \( P(D_n^{\alpha} \in G_n) \rightarrow 1 \) as \( n \rightarrow \infty, \) and the uniform map satisfies

\[ P\left(\left|\sum_{j=1}^{n} (D_{n,j}^{\alpha})^2 - \mu_{2,\alpha} n\right| \geq \sqrt{\omega n}\right) \rightarrow 0, \quad \mu_{2,\alpha} := \mathbb{E}\left[(D_{n,1}^{\alpha})^2\right] = \frac{(kn)^2}{n^2} + k. \]

But, for \( n \) sufficiently large,

\[ \left|\mu_{2,\alpha} n - \mu_{2,\alpha} n\right| = \frac{k(n-1)(kn-1)}{\alpha n + 1} \geq \frac{k^2 n^2}{2\alpha n} = \frac{1}{2} k^2 \omega \sqrt{n} \geq \sqrt{\omega n}, \]

so that \( P(D_n^{\alpha} \in G_n) \rightarrow 0 \) as \( n \rightarrow \infty. \) So, we have

\[ 1 \geq \left\|P_{n,k}^{\alpha} - P_{n,k}^{\infty}\right\|_{TV} \geq \left|P(D_n^{\alpha} \in G_n) - P(D_n^{\infty} \in G_n)\right| \rightarrow 1 \text{ as } n \rightarrow \infty, \]

proving the claim.

4.4 Proof for the case \( \alpha/\sqrt{n} \rightarrow \infty \)

Again, let \( D_n^{\alpha} \) and \( D_n^{\infty} \) denote the in-degree sequences of \( M_{n,k}^{\alpha} \) and \( M_{n,k}^{\alpha}, \) respectively.

Let us write \( \alpha = \omega \sqrt{n}; \) note that \( \omega \rightarrow \infty \) as \( n \rightarrow \infty. \)
First, note that the total variation distance between $P_{n,k}^\alpha$ and $P_{n,k}^\infty$ is precisely the total variation distance between their in-degree sequences: conditioned on having in-degree sequence $d$, each mapping is uniformly distributed on $M_{n,k}(d)$ (see Theorem 2.3.2), so that

$$\left\| P_{n,k}^\alpha - P_{n,k}^\infty \right\|_{TV} = \frac{1}{2} \sum_{M \in M_{n,k}} \left| P_{n,k}^\alpha(M) - P_{n,k}^\infty(M) \right|$$

$$= \frac{1}{2} \sum_d \sum_{M \in M_{n,k}(d)} \frac{1}{|M_{n,k}(d)|} \left| P(D_n^\alpha = d) - P(D_n^\infty = d) \right|$$

$$= \frac{1}{2} \sum_d \left| P(D_n^\alpha = d) - P(D_n^\infty = d) \right|,$$

where these sums have $d$ ranging over all valid in-degree sequences $d = (d_1, \ldots, d_n)$: that is, $d_1, \ldots, d_n$ are non-negative integers, and $d_1 + \cdots + d_n = kn$. So, in particular,

$$\left\| P_{n,k}^\alpha - P_{n,k}^\infty \right\|_{TV} = \frac{1}{2} \sum_d \left( \frac{kn}{d_1, \ldots, d_n} \right) \left\| \frac{\prod_{j=1}^{kn} \alpha_d^j}{(\alpha n)^{kn}} - \frac{1}{n^{kn}} \right\|$$

$$= \frac{1}{2n^{kn}} \sum_d \left( \frac{kn}{d_1, \ldots, d_n} \right) \left| 1 - \frac{n^{kn} \prod_{j=1}^{kn} \alpha_d^j}{(\alpha n)^{kn}} \right|.$$
as $n \to \infty$. It only remains to show that the sum over $d \in B_n$ tends to 0 as well.

Applying the first bound in Lemma 4.2.1 yields that

$$\left(\alpha n \right)^{kn} = (\alpha n)^{kn} \exp \left[ \frac{k^2 \sqrt{n}}{2\omega} + O\left(\frac{1}{\omega \sqrt{n}}\right) + O\left(\frac{1}{\omega^2}\right) \right],$$

while

$$\prod_{j=1}^{n} \alpha^{d_j} = \alpha^{kn} \exp \left[ \sum_{j=1}^{n} \frac{d_j(d_j - 1)}{2\alpha} + O\left(\frac{\sum_{j=1}^{n} d_j^3}{\alpha^2}\right) \right]
= \alpha^{kn} \exp \left[ \frac{k^2 \sqrt{n}}{2\omega} + O\left(\frac{1}{\sqrt{\omega}}\right) + O\left(\frac{1}{\omega^2}\right) \right],$$

uniformly over $d \in B_n$. So, uniformly over $d \in B_n$,

$$\frac{n^{kn} \prod_{j=1}^{n} \alpha^{d_j}}{(\alpha n)^{kn}} = \exp \left[ O\left(\frac{1}{\sqrt{\omega}}\right) \right] = 1 + O\left(\frac{1}{\sqrt{\omega}}\right),$$

so that

$$\frac{1}{2n^{kn}} \sum_{d \in B_n} \left(\frac{kn}{d_1, \ldots, d_n}\right) \left| 1 - \frac{n^{kn} \prod_{j=1}^{n} \alpha^{d_j}}{(\alpha n)^{kn}} \right| = O\left(\frac{1}{n^{kn}} \sum_{d} \left(\frac{kn}{d_1, \ldots, d_n}\right) \frac{1}{\sqrt{\omega}}\right) = o(1),$$

proving the claim. \qed

4.5 Proof for the case $\alpha = \beta \sqrt{n}$

This case requires a significantly more delicate touch than was necessary in either of the previous cases. We will proceed by reducing the total variation distance to a sum over terms with meaningful contributions; in this case, “meaningful” terms are very nearly determined by the sum of the squared in-degrees!

The issue, of course, is that the in-degrees are not independent: they always sum to $kn$. However, this sort of dependence is not terribly strong, and each $D_{a,j}$ makes a minor contribution to the whole; so, we might expect the sum of squares, appropriately centered and scaled, to converge in distribution to a normal distribution.
Let us proceed by formalizing this idea. To do so, we will need to go somewhat out of the way, as the sum of squares of the in-degree sequence is not particularly amenable to attack by characteristic functions. However, we know that the in-degree sequence is distributed as a sequence of independent generalized negative binomial variables \((Z_{1,n}, \ldots, Z_{n,n})\), conditioned on their sum being \(kn\); this gives us our opening.

Let us start by proving a local limit theorem for the two-dimensional random variable \((\sum_{j=1}^{n} Z_{n,j}, \sum_{j=1}^{n} Z_{2,n,j})\), then use the local limit result to achieve the desired central limit theorem for \(\sum_{j=1}^{n}(D_{n,j}^\alpha)^2\).

Why do this, rather than going directly for a central limit theorem? Our device for going from the two-dimensional variable back to the sum of squares is the conditional probability formula - and we are conditioning on the event \(Z_{1,n} + \cdots + Z_{n,n} = kn\), which is very thin (it has probability on the order of \(n^{-1/2}\)). So, we need some sort of uniform convergence if we are to stand a chance of getting the desired central limit theorem for the sum of squares in this manner.

Note that in what follows, all pairs and vectors are treated as column matrices (so that \((a,b)\) acts as the matrix \([a\ b]\)).

**Lemma 4.5.1.** Suppose \(\alpha = \alpha(n) < \infty\) for all \(n\), but that \(\alpha(n) \to \infty\) as \(n \to \infty\). Let \(Z_n = (Z_{1,n}, \ldots, Z_{n,n})\) be defined as in Section 2.3, and let \(S_n\) denote the normalized vector

\[
S_n = \left(\frac{\sum_{j=1}^{n} Z_{n,j} - kn}{\sqrt{n}}, \frac{\sum_{j=1}^{n} Z_{2,n,j}^2 - \mu'_{2,\alpha} n}{\sqrt{n}}\right), \quad \mu'_{2,\alpha} := \mathbb{E}[Z_{n,1}^2] = k^2 + \frac{k^2}{\alpha} + k.
\]

Then

\[
\lim_{n \to \infty} \sup_{x \in \text{Supp}(S_n)} \left| \frac{n}{2} P(S_n = x) - \eta(x) \right| = 0,
\]

where

\[
\eta(x) := \frac{\exp \left( -\frac{1}{2} x^T \Sigma^{-1} x \right)}{2\pi \sqrt{|\det \Sigma|}}
\]

52
is the density function of a Gaussian random vector on $\mathbb{R}^2$ with mean $0$ and covariance matrix
\[
\Sigma := \begin{bmatrix} k & 2k^2 + k \\ 2k^2 + k & 4k^3 + 6k^2 + k \end{bmatrix}.
\]

Note: Our proof follows the pattern of the proof of a one-variable local limit theorem in the textbook by Durrett [14, Section 2.5, Theorem 5.2], allowing for two wrinkles: that we deal with random vectors, and that the distributions of $Z_{n,1}, \ldots, Z_{n,n}$ depend on $n$ (by way of the dependence of $\alpha$ on $n$). Most of the facts used in Durrett’s proof generalize easily to the multivariable case; my reference for these generalizations is the textbook [3] by Bhattacharya and Rao.

Proof. For ease of notation, define
\[
V_{n,j} := (Z_{n,j} - \mathbb{E}[Z_{n,j}], Z_{n,j}^2 - \mathbb{E}[Z_{n,j}^2]) = \left( Z_{n,j} - k, Z_{n,j}^2 - \left( k^2 + k + \frac{k^2}{\alpha} \right) \right).
\]
Note that $V_{n,1}, \ldots, V_{n,n}$ are independent, and $S_n = (V_{n,1} + \cdots + V_{n,n})/\sqrt{n}$. Let $\phi_n$, $\Phi_n$, and $\psi$ denote the characteristic functions for $V_{n,j}$, $S_n$, and $\mathcal{N}(0, \Sigma)$, respectively.

The minimal additive subgroup $\mathcal{L} \subseteq \mathbb{R}^2$ such that there exists $x_0 \in \mathbb{R}^2$ with $P(V_{n,j} \in x_0 + \mathcal{L}) = 1$ is generated by $(1,1)$ and $(1,3)$: these vectors are both necessary to accommodate the fact that $(0,0), (1,1), (2,4) \in \text{Supp}(Z_{n,j}, Z_{n,j}^2)$, and we cannot leave their span because $m^2 \equiv m \ (\text{mod} \ 2)$ for all non-negative integers $m$.

Since $\mathcal{L}$ is the minimal subgroup for $V_{n,j}$ and $V_{n,1}, \ldots, V_{n,n}$ are independent, $\mathcal{L}$ is also the minimum subgroup for $\sum_{j=1}^n V_{n,j}$; hence the minimum subgroup for $S_n$ is $\mathcal{L}/\sqrt{n}$, the lattice generated by $(1,1)/\sqrt{n}$ and $(1,-1)/\sqrt{n}$.

So, by the inversion formula for lattice-distributed variables [3, p.230, eq. 21.28], for any $x \in \text{Supp}(S_n)$ we have
\[
P(S_n = x) = \frac{1}{4\pi^2} \cdot \frac{2}{n} \cdot \int_{\mathbb{R}^2} e^{-i(t,x)} \phi_n(t) \, dt,
\]
53
where
\[ \mathcal{F}_n := \{ t : |t_1 + t_2| < \pi \sqrt{n} \text{ and } |t_1 - t_2| < \pi \sqrt{n} \}. \]

By the inversion formula for continuous variables [3, Theorem 4.1(iv)],
\[
\eta(x) = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} e^{-i\langle t, x \rangle} \psi(t) \, dt = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} e^{-i\langle t, x \rangle} e^{-\frac{1}{2} t^T \Sigma t} \, dt.
\]

So, for any \( x \in \text{Supp}(S_n) \),
\[
\left| \frac{n}{2} P(S_n = x) - \eta(x) \right| \leq \frac{1}{4\pi^2} \int_{\mathcal{F}_n} \left| \Phi_n(t) - e^{-\frac{1}{2} t^T \Sigma t} \right| \, dt + \int_{\mathbb{R}^2 \setminus \mathcal{F}_n} e^{-\frac{1}{2} t^T \Sigma t} \, dt \quad (4.5.1)
\]

This last expression has no dependence on \( x \); so, uniformity follows by proving that the quantity on the right side of (4.5.1) tends to 0 with \( n \). That the second integral tends to 0 is immediate, as \( e^{-\frac{1}{2} t^T \Sigma t} \) is integrable and \( \mathcal{F}_n \) increases to \( \mathbb{R} \).

For the first, we begin by noting that since \( S_n = (V_{n,1} + \cdots + V_{n,n})/\sqrt{n} \), independence of the \( V_{n,j} \) implies
\[
\Phi_n(t) = \phi_n \left( \frac{t}{\sqrt{n}} \right)^n.
\]

The vectors \( V_{n,j} \) have mean \( 0 \); denote their covariance by \( \Sigma_n \). The third moment of \( V_{n,j} \) is bounded uniformly in \( n \), and so for \( y \to 0 \),
\[
\phi_n(y) = 1 - \frac{y^T \Sigma_n y}{2} + O(|y|^3);
\]
computing \( \Sigma_n \) using Lemma 2.3.1, we find that \( \Sigma_n \) differs from \( \Sigma \) (in, say, the Euclidean norm) by at most \( \sqrt{n} \). hence
\[
\Phi_n(t) = \phi_n \left( \frac{t}{\sqrt{n}} \right)^n = \left( 1 - \frac{t^T \Sigma t}{2n} + o(n) \right)^n \to e^{-\frac{1}{2} t^T \Sigma t} \quad \text{as } n \to \infty
\]
pointwise, for all \( t \).

The third moments of \( V_{n,j} \) are bounded uniformly in \( n \). So, as \( y \to 0 \), uniformly over \( n \) we have
\[
|\phi_n(y)| = 1 - \frac{1}{2} y^T \Sigma_n y + O \left( |y|^3 \right).
\]
Because $\det \Sigma_n$ is also uniformly bounded in $n$, there exists a constant $\delta' > 0$ such that for all $n$,

$$|\phi_n(y)| \leq 1 - \frac{1}{4} y^T \Sigma_n y$$

whenever $|y| < \delta'$.

Now, for $n$ sufficiently large, there is a constant $\gamma > 0$ such that $y^T \Sigma_n y \geq \gamma |y|^2$ for all $y \in \mathbb{R}^2$: note that $\Sigma$ and $\Sigma_n$ are positive definite. So, if this is not true, then there are a subsequence $(\Sigma_n)_{n \in M}$ and associated non-zero vectors $(y_n)_{n \in M}$ such that $(y_n^T \Sigma_n y_n / |y_n|^2)_{n \in M}$ tends to 0. Then the vectors $v_n := y_n / |y_n|$ all have length 1, and $(v_n^T \Sigma_n v_n)_{n \in M}$ also tends to 0. The sequence $(v_n)_{n \in M}$ is bounded, and so has a subsequence which converges to some $v \in \mathbb{R}^2$; but, this $v$ then has $|v| = 1$ and $v^T \Sigma v = 0$, contradicting positive-definiteness of $\Sigma$.

So, for $n$ sufficiently large and $|y| < \delta'$,

$$|\phi_n(y)| \leq 1 - \frac{1}{4} \gamma |y|^2 \leq e^{-\frac{1}{4} \gamma |y|^2}.$$

Choose $\delta > 0$ such that $t \in \mathcal{F}_n^{(1)}$ implies $|t| < \delta' \sqrt{n}$, where

$$\mathcal{F}_n^{(1)} := \{ t : |t_1 + t_2| < \delta \sqrt{n} \text{ and } |t_1 - t_2| < \delta \sqrt{n} \}.$$

Then for $t \in \mathcal{F}_n^{(1)}$,

$$|\Phi_n(t)| = \left| \phi_n \left( \frac{t}{\sqrt{n}} \right) \right|^n \leq \left( e^{-\frac{1}{4} \gamma |t|^2 / n} \right)^n = e^{-\frac{1}{4} \gamma |t|^2}.$$

By the triangle inequality, for $t \in \mathcal{F}_n^{(1)}$, for all $n$,

$$|\Phi_n(t) - e^{-\frac{1}{2} t^T \Sigma t}| \leq e^{-\frac{1}{4} \gamma |t|^2} + e^{-\frac{1}{2} t^T \Sigma t},$$

which is integrable on $\mathbb{R}^2$. So, because this integrand tends to 0 pointwise,

$$\int_{\mathcal{F}_n^{(1)}} \left| \Phi_n(t) - e^{-\frac{1}{2} t^T \Sigma t} \right| dt \to 0 \text{ as } n \to \infty,$$

by the Dominated Convergence Theorem.
We finish by considering $F_n^{(2)} := F_n \setminus F_n^{(1)}$. Note that as $n \to \infty$, $V_{n,j}$ converges in distribution to $(Y - k, Y^2 - (k^2 + k))$, where $Y$ is Poisson-distributed with parameter $k$; so, by a slight generalization of [18, Section XV, Theorem 2], $\phi_n$ converges to the characteristic function $\phi^*$ of $(Y, Y^2)$ uniformly on compact sets. By [3, Lemma 21.6], $|\phi^*(y)| = 1$ (and $|\phi_n(y)| = 1$) if and only if $|y_1 + y_2|, |y_1 - y_2| \in 2\pi \mathbb{Z}$. In particular, for $t \in F_n^{(2)}$, $t/\sqrt{n}$ is bounded away from all such points, so that by uniform continuity there exists $\epsilon \in (0, 1)$ so that $|\phi^*(t/\sqrt{n})| < \epsilon/2$ for all $t \in F_n^{(2)}$. But $F_n^{(2)}/\sqrt{n}$ is contained in a compact set, so that $\phi_n$ converges uniformly to $\phi^*$ on $F_n^{(2)}/\sqrt{n}$, implying that $|\phi_n(t/\sqrt{n})| < \epsilon$ for $t \in F_n^{(2)}$ and $n$ sufficiently large. But then such $n$ and $t$ have $|\Phi_n(t)| < \epsilon^n$, and therefore

$$\int_{F_n^{(2)}} |\Phi_n(t) - e^{-\frac{1}{2}t^T \Sigma t}| \, dt \leq \epsilon^n \text{Vol}(F_n^{(2)}) + \int_{F_n^{(2)}} e^{-\frac{1}{2}t^T \Sigma t} \, dt \to 0 \text{ as } n \to \infty,$$

completing the proof.

With this in hand, we can prove the desired central limit theorem:

**Lemma 4.5.2.** Suppose $\alpha = \alpha(n) \to \infty$ as $n \to \infty$. Let $D_n^\alpha = (D_{n,1}^\alpha, \ldots, D_{n,n}^\alpha)$ denote the in-degree sequence of $M_{n,k}^\alpha$. Then

$$\frac{(D_{n,1}^\alpha)^2 + \cdots + (D_{n,n}^\alpha)^2 - \mu_{2,\alpha}^n}{\sqrt{n}} \Rightarrow \mathcal{N}(0, 2k^2), \quad \mu_{2,\alpha} := \mathbb{E}[Z_{n,1}^2] = k^2 + \frac{k^2}{\alpha} + k.$$

**Proof.** Let $Z_{n,1}, \ldots, Z_{n,n}$ and $S_n$ be as in Lemma 4.5.1, and let

$$S_n := \frac{(D_{n,1}^\alpha)^2 + \cdots + (D_{n,n}^\alpha)^2 - \mu_{2,\alpha}^n}{\sqrt{n}}.$$

By Theorem 2.3.2, $(0, S_n)$ is distributed as $S_n$ conditioned on $Z_{n,1} + \cdots + Z_{n,n} = kn$. For $a < b$,

$$P(a \leq S_n \leq b) = \sum_{a \leq x \leq b} \frac{\eta(0, x)}{P(Z_{n,1} + \cdots + Z_{n,n} = kn)} \cdot \frac{2}{n} + R_n(a, b),$$

56
where \( \eta(x) \) is as in Lemma 4.5.1 and

\[
R_n(a, b) := \sum_{\substack{a \leq x \leq b \\ x \in \text{Supp}(S_n)}} \frac{n}{2} P(S_n = x) - \eta(x) \cdot \frac{2}{n}.
\]

Using Lemma 2.3.1(iv), the identity \( a^b = \Gamma(a + b)/\Gamma(a) \), and Stirling’s formula for the \( \Gamma \)-function yields the asymptotic estimate

\[
P(Z_{n,1} + \cdots + Z_{n,n} = kn) = \frac{1}{\sqrt{2\pi kn}} \left( 1 + O \left( \frac{1}{\sqrt{n}} \right) \right).
\]

Since \( \text{Supp}(S_n) \) is a lattice with span \( \frac{2}{\sqrt{n}} \), it follows by Lemma 4.5.1 that

\[
|R_n(a, b)| \leq \frac{b - a}{\sqrt{n} P(Z_{n,1} + \cdots + Z_{n,n} = kn)} \sup_{x \in \text{Supp}(S_n)} \left| \frac{n}{2} P(S_n = x) - \eta(x) \right| \to 0
\]

as \( n \to \infty \). This, along with the fact that

\[
\eta(0, x) = \frac{e^{-x^2/(2k^2)}}{2\pi \sqrt{2k^3}},
\]

implies

\[
P(a \leq S_n \leq b) = \sum_{\substack{a \leq x \leq b \\ x \in \text{Supp}(S_n)}} \frac{e^{-x^2/(2k^2)}}{2k\sqrt{\pi}} \cdot \frac{2}{\sqrt{n}} + o(1) \text{ as } n \to \infty.
\]

Since \( \text{Supp}(S_n) \) is a lattice with span \( \frac{2}{\sqrt{n}} \), this last expression is a Riemann sum! Thus

\[
P(a \leq S_n \leq b) \to \int_a^b \frac{e^{-x^2/(2k^2)}}{2k\sqrt{\pi}} \, dx = P(a \leq N(0, 2k^2) \leq b) \text{ as } n \to \infty,
\]

and therefore \( S_n \Rightarrow N(0, 2k^2) \) as \( n \to \infty \) as claimed.

Finally, we are ready for the last part of the theorem:

**Proof of Theorem 4.1.1(iii).** As usual, let \( D_n^\alpha \) and \( D_n^\infty \) denote the in-degree sequences of \( M_{n,k}^\alpha \) and \( M_{n,k}^\infty \), respectively. For \( s \in \mathbb{N} \), define

\[
\mu_s,\infty := \mathbb{E}[(D_n^\infty)_s^s] = \sum_{\ell=1}^s \left\{ \begin{array}{c} s \\ \ell \end{array} \right\} k^\ell \quad \text{and} \quad \mu_s,\alpha := \mathbb{E}[(D_n^\alpha)_s^s] = \sum_{\ell=1}^s \left\{ \begin{array}{c} s \\ \ell \end{array} \right\} \alpha^\ell (kn)_\ell / (\alpha n)^\ell.
\]
Fix a constant $A > \frac{2k^2}{\beta}$ and $\omega = \omega(n)$ such that $\omega \to \infty$ as $n \to \infty$ but $\omega/\sqrt{n} \to 0$.

Let

$$A_n = A_n(A) := \left\{ d : \left| \sum_{j=1}^{n} d_j^2 - \mu_{2,\alpha} n \right| < A\sqrt{n} \right\}$$

and

$$A'_n = A'_n(A) := \left\{ d \in A_n : \left| \sum_{j=1}^{n} d_j^s - \mu_{s,\alpha} n \right| < \omega\sqrt{n} \text{ for } s = 3, 4 \right\}.$$

Define

$$f(x) := \left| 1 - \exp \left[ -\frac{k^2}{2\beta^2} - \frac{x}{2\beta} \right] \right| \quad \text{and} \quad f_A(x) := \begin{cases} f(x) & \text{if } |x| \leq A \\ f(A) & \text{if } x > A \\ f(-A) & \text{if } x < -A \end{cases}.$$

Finally, as in Lemma 4.5.2, let

$$S_n := \sum_{j=1}^{n} (D_{n,j}^\alpha)^2 - \mu_{2,\alpha}^\prime n, \quad \mu_{2,\alpha}^\prime := \mathbb{E}[Z_{n,1}^2] = k^2 + k + \frac{k^2}{\alpha}.$$

We proceed to estimate the total variation distance. As before, we have

$$\|P_{n,k}^\alpha - P_{n,k}^\infty\|_{TV} = \sum_d \left| \frac{k^n}{(\alpha n)^{kn}} \prod_{j=1}^{n} \alpha d_j - \frac{1}{n^{kn}} \right|.$$

The error committed in restricting this sum to $d \in A'_n$ can be made small by choosing $A$ large: in fact, there is a constant $C > 0$ such that

$$0 \leq \limsup_{n \to \infty} \sum_{d \notin A'_n} \left( \frac{k^n}{(\alpha n)^{kn}} \prod_{j=1}^{n} \alpha d_j - \frac{1}{n^{kn}} \right) \leq \frac{C}{(A - \frac{k^2}{\beta})^2}. \quad (4.5.2)$$

That 0 is a lower bound is immediate; for the upper bound, the triangle inequality implies

$$\sum_{d \notin A'_n} \left( \frac{k^n}{(\alpha n)^{kn}} \prod_{j=1}^{n} \alpha d_j - \frac{1}{n^{kn}} \right) \leq P(D^n_{\alpha} \notin A'_n) + P(D^n_{\infty} \notin A'_n).$$

58
Now, we bound
\[ P(D_n^\infty \notin A'_n) \leq P \left( \left| \sum_{j=1}^n (D_{n,j}^\infty)^2 - \mu_{2,\alpha} n \right| \geq A\sqrt{n} \right) \]
\[ + \sum_{s=3}^4 P \left( \left| \sum_{j=1}^n (D_{n,j}^\infty)^s - \mu_{s,\alpha} n \right| \geq \omega \sqrt{n} \right). \]

Note the following, where the inequalities hold for \( n \) sufficiently large:
\[
|\mu_{2,\alpha} - \mu_{2,\infty}| = \frac{k^2}{\beta} \sqrt{n} + O(1) \leq \frac{2k^2}{\beta} \sqrt{n}.
\]
\[
|\mu_{3,\alpha} - \mu_{3,\infty}| = \frac{3k^3 + 3k^2}{\beta} \sqrt{n} + O(1) \leq \frac{6k^3 + 6k^2}{\beta} \sqrt{n} =: c_3 \sqrt{n}.
\]
\[
|\mu_{4,\alpha} - \mu_{4,\infty}| = \frac{6k^4 + 18k^3 + 7k^2}{\beta} \sqrt{n} + O(1) \leq \frac{12k^4 + 36k^3 + 7k^2}{\beta} \sqrt{n} =: c_4 \sqrt{n}.
\]

So, for \( s = 3 \) and \( s = 4 \),
\[
P \left( \left| \sum_{j=1}^n (D_{n,j}^\infty)^3 - \mu_{s,\alpha} n \right| \geq \omega \sqrt{n} \right) \leq P \left( \left| \sum_{j=1}^n (D_{n,j}^\infty)^s - \mu_{s,\infty} n \right| \geq (\omega - c_s) \sqrt{n} \right)
\]
\[ \leq \frac{\text{Var} \left[ \sum_{j=1}^n (D_{n,j}^\infty)^s \right]}{(\omega - c_s)^2 n} \]

by Chebyshev’s inequality; similarly,
\[
P \left( \left| \sum_{j=1}^n (D_{n,j}^\infty)^2 - \mu_{2,\alpha} n \right| \geq A\sqrt{n} \right) \leq P \left( \left| \sum_{j=1}^n (D_{n,j}^\infty)^2 - \mu_{2,\infty} n \right| \geq \left( A - \frac{2k^2}{\beta} \right) \sqrt{n} \right)
\]
\[ \leq \frac{\text{Var} \left[ \sum_{j=1}^n (D_{n,j}^\infty)^2 \right]}{(A - \frac{2k^2}{\beta})^2 n} \]

Using the moments in Lemma 2.3.3, these variances are all \( O(n) \). More direct application of Chebyshev’s inequality yields similar bounds for \( (D_{n,j}^\alpha)^s, s = 2, 3, 4 \), with variances being \( O(n) \) by Lemma 2.3.3. So, taking limits supremum yields (4.5.2).

Next, we claim
\[
\frac{1}{2} \sum_{d \in A'_n} \left( \frac{kn}{d_1, \ldots, d_n} \right) \left| \prod_{j=1}^n \alpha^d_j \right| \frac{1}{n^{kn}} = \frac{1}{2} \mathbb{E} \left[ f(S_n) 1_{(D_n^\alpha \in A'_n)} \right] + O \left( \frac{\omega}{\sqrt{n}} \right). \tag{4.5.3}
\]
Why? Start by factoring \((\prod_{j=1}^{n} \alpha^{\overline{d_j}}) / (\alpha n)^{kn}\) out of the absolute value on the left:

\[
\left(\frac{kn}{d_1, \ldots, d_n} \right) \left| \frac{\prod_{j=1}^{n} \alpha^{\overline{d_j}} - 1}{(\alpha n)^{kn}} \right| = \left(\frac{kn}{d_1, \ldots, d_n} \right) \left| \frac{\prod_{j=1}^{n} \alpha^{\overline{d_j}}}{(\alpha n)^{kn}} \right| 1 - \frac{(\alpha n)^{kn}}{n^{kn} \prod_{j=1}^{n} \alpha^{\overline{d_j}}}.\]

By the second bound in Lemma 4.2.1,

\[
(\alpha n)^{kn} = (\alpha n)^{kn} \exp \left[ \frac{k^2 n}{2\beta \sqrt{n}} - \frac{k^3}{6\beta^2} + O \left( \frac{1}{\sqrt{n}} \right) \right],
\]

while uniformly over all \(d\),

\[
\prod_{j=1}^{n} \alpha^{\overline{d_j}} = \alpha^{kn} \exp \left[ \sum_{j=1}^{n} \left( \frac{d_j^2 - d_j}{2\beta \sqrt{n}} - \frac{2d_j^3 - 3d_j^2 + d_j}{12\beta^2 n} \right) \right] + O \left( \sum_{j=1}^{n} d_j^4 \right)\]

Restricting to \(d \in A'_n\), this implies

\[
\prod_{j=1}^{n} \alpha^{\overline{d_j}} = \alpha^{kn} \exp \left[ \sum_{j=1}^{n} d_j^2 \frac{k}{2\beta \sqrt{n}} - 2\mu_3, \alpha - 3\mu_2, \alpha + k \frac{k^2}{12\beta^2} + O \left( \frac{\omega}{\sqrt{n}} \right) \right]
= \alpha^{kn} \exp \left[ \sum_{j=1}^{n} d_j^2 \frac{k}{2\beta \sqrt{n}} - \frac{2\mu(n, j)}{2\beta \sqrt{n}} \right] + O \left( \frac{\omega}{\sqrt{n}} \right)
= \alpha^{kn} \exp \left[ \sum_{j=1}^{n} d_j^2 \frac{k}{2\beta \sqrt{n}} - \frac{k^2}{6\beta^2} - \frac{k^3}{4\beta^2} + O \left( \frac{\omega}{\sqrt{n}} \right) \right]
\]

by applying the asymptotic approximations of the moments in Lemma 2.3.3, where

\[
\mu'_{2, \alpha} := \mathbb{E}[Z^2_{n,j}] = k^2 + \frac{k^2}{\alpha} + k = k^2 + \frac{k^2}{\beta \sqrt{n}} + k.
\]

Combining these results yields

\[
\frac{(\alpha n)^{kn}}{n^{kn} \prod_{j=1}^{n} \alpha^{\overline{d_j}}} = \exp \left[ - \frac{k^2}{4\beta^2} - \frac{\sum_{j=1}^{n} d_j^2 - \mu'_{2, \alpha n}}{2\beta \sqrt{n}} + O \left( \frac{\omega}{\sqrt{n}} \right) \right],
\]

so that

\[
\left| 1 - \frac{(\alpha n)^{kn}}{n^{kn} \prod_{j=1}^{n} \alpha^{\overline{d_j}}} \right| = f \left( \frac{\sum_{j=1}^{n} d_j^2 - \mu'_{2, \alpha n}}{\sqrt{n}} \right) + O \left( \frac{\omega}{\sqrt{n}} \right)
\]

uniformly over \(d \in A'_n\) by boundedness of \(f\) over such \(d\); hence

\[
\frac{1}{2} \sum_{d \in A'_n} \left( \frac{kn}{d_1, \ldots, d_n} \right) \frac{\prod_{j=1}^{n} \alpha^{\overline{d_j}}}{(\alpha n)^{kn}} \left| 1 - \frac{(\alpha n)^{kn}}{n^{kn} \prod_{j=1}^{n} \alpha^{\overline{d_j}}} \right| = \frac{1}{2} \mathbb{E} \left[ f(S_n) \mathbf{1}_{\{D_n \in A'_n\}} \right] + O \left( \sum_{d} \left( \frac{kn}{d_1, \ldots, d_n} \frac{\omega}{\sqrt{n}} \right) \right)
= \frac{1}{2} \mathbb{E} \left[ f(S_n) \mathbf{1}_{\{D_n \in A'_n\}} \right] + O \left( \frac{\omega}{\sqrt{n}} \right),
\]

60
proving (4.5.3).

Next, I claim that there is a constant $C$ (independent of $A$) such that

$$\limsup_{n \to \infty} \left| \frac{1}{2} \mathbb{E} \left[ f(S_n) \mathbf{1}_{D_n^k \in \mathcal{A}_n} \right] - \frac{1}{2} \mathbb{E} [f_A(S_n)] \right| \leq C \left( 1 + e^{A/2} e^{-A^2/4k^2} \right).$$ (4.5.4)

Consider these expectations as sums over in-degree sequences. The two expressions are identical for sequences in $\mathcal{A}_n'$. The first expression is 0 in all other cases. On the other hand,

$$|f_A(x)| \leq 1 + e^{A/2},$$

and so

$$0 \leq \frac{1}{2} \mathbb{E}[f_A(S_n) \mathbf{1}_{D_n^k \in \mathcal{A}_n \setminus \mathcal{A}_n'}] \leq (1 + e^{A/2}) P(D_n^a \notin \mathcal{A}_n') = O \left( \frac{\omega}{\sqrt{n}} \right).$$

For in-degree sequences not in $\mathcal{A}_n$, we bound

$$\left| \mathbb{E}[f_A(S_n) \mathbf{1}_{D_n^k \notin \mathcal{A}_n}] \right| \leq (1 + e^{A/2}) P(|S_n| \geq A) \to (1 + e^{A/2}) P(|\mathcal{N}(0, 2k^2)| \geq A)$$

as $n \to \infty$ because $A$ is fixed and $S_n \Rightarrow \mathcal{N}(0, 2k^2)$ by Lemma 4.5.2. By the usual tail inequality for normal variables (see, for example, [14, Theorem 1.1.4]),

$$P(|\mathcal{N}(0, 2k^2)| \geq A) \leq \frac{2k}{\sqrt{\pi}} \cdot \frac{e^{-A^2/4k^2}}{A}.$$

This establishes (4.5.4).

Next: notice that $f_A$ is bounded and continuous; so, because $S_n \Rightarrow \mathcal{N}(0, 2k^2)$ (by Lemma 4.5.2),

$$\lim_{n \to \infty} \frac{1}{2} \mathbb{E}[f_A(S_n)] = \frac{1}{2} \mathbb{E}[f_A(\mathcal{N}(0, 2k^2))].$$ (4.5.5)

Finally, we are ready to start putting the pieces together. Combining (4.5.2), (4.5.3), (4.5.4), and (4.5.5) yields

$$\limsup_{n \to \infty} \left| \left\| P_{n,k}^\infty - P_{n,k}^\infty \right\|_{TV} - \frac{1}{2} \mathbb{E}[f_A(\mathcal{N}(0, 2k^2))] \right| \leq b \frac{1}{(A - \frac{2k^2}{\beta})^2} + \frac{(1 + e^{A/2}) e^{-A^2/4k^2}}{A},$$ (4.5.6)
where “≤b” denotes that the left side is bounded by a constant (in this case, independent of A) multiple of the right.

Now, (4.5.6) holds for any fixed A. Letting A increase, the right side of (4.5.6) tends to 0; further, calculations similar to the proof of (4.5.4) show that

$$\frac{1}{2} \mathbb{E}[f_A(\mathcal{N}(0, 2k^2))] \to \frac{1}{2} \mathbb{E}[f(\mathcal{N}(0, 2k^2))] = \frac{1}{2} \mathbb{E} \left| 1 - \exp \left[ -\mathcal{N} \left( \frac{k^2}{4\beta^2}, \frac{k^2}{2\beta^2} \right) \right] \right|$$

as $A \to \infty$, proving the result. $\Box$
CHAPTER 5
OPEN PROBLEMS AND RELATED MODELS

5.1 Introduction

The following sections contain questions for future study. The first several sections outline problems for further study in $M^\alpha_{n,k}$ and $G^\alpha_{n,k}$; the list is not by any means exhaustive, but is rather a list of the most immediate follow-up questions to this work. The remaining sections will then detail a few different models, related to $M^\alpha_{n,k}$, which may warrant further investigation.

5.2 The $k$-Core of $G^\alpha_{n,k}$ and the Model $G_{k-out}$

In Section 3.4, we established that w.h.p. the $k$-core of $G^\alpha_{n,k}$ is precisely the induced subgraph on the set of vertices of degree $k$ or more. However, one of our primary hopes in studying the $k$-core is as yet unanswered.

Even the uniform model $G^{\infty}_{n,k}$ is qualitatively different from $k$-out graph studied by Fenner and Frieze [19], in that the possibility of loops and repeated arcs in $M^\alpha_{n,k}$ means that $G^\alpha_{n,k}$ need not have minimum degree $k$. However, Theorem 3.2.1 tells us that any possible shortfall in the minimum degree is slight - at worst, a Poisson-distributed number of vertices have degree $k - 1$, and w.h.p. none have degree $k - 2$ or lower. The natural question to ask, then: how similar is the behavior of the $k$-core to the graph $G^\alpha_{n,k}$ conditioned on minimum degree at least $k$? If these distributions
are essentially similar, it will provide a very strong basis for comparison between \( G_{n,k}^\alpha \) and \( G_{k-\text{out}} \). The following conjecture forms the most basic confirmation of this relationship:

**Conjecture 5.2.1.** The \( k \)-core of \( G_{n,k}^\alpha \) is w.h.p. \( k \)-connected.

### 5.3 The \( h \)-Cores of \( G_{n,k}^\alpha \) for \( h > k \)

We showed in Theorem 3.4.1 that w.h.p. there is a giant \( k \)-core in \( G_{n,k}^\alpha \) - in fact, the number of vertices not included in the \( k \)-core is bounded in probability. What about \( h \)-cores for \( h > k \)? A vertex of \( G_{n,k}^\alpha \) can have at most \( k \) neighbors of its own choosing, but may have many more neighbors that choose it in \( M_{n,k}^\alpha \). In fact, it is not difficult to show that the number of vertices of degree \( k + 1 \) or more is likely linear in \( n \). How large can \( h \) be so that we expect a giant \( h \)-core? For that matter, for how large of \( h \) is there likely an \( h \)-core at all?

The largest that such \( h \) could possibly be is \( 2k \). Why? The edges in the induced subgraph of \( G_{n,k}^\alpha \) on some set \( T \subseteq [n] \) are precisely the edges due to arcs in \( M_{n,k}^\alpha \) that both start and end in \( T \); there can be at most \( k |T| \) such arcs, and so there are at most \( k |T| \) edges in this induced graph, and the total degree is at most \( 2k |T| \). So, the highest the minimum degree could possibly be is \( 2k \), and it is likely lower.

Based on computer simulation, it seems that there is a slowly growing function \( f \) such that w.h.p. \( G_{n,k}^\alpha \) contains giant \( h \)-cores for \( h \leq k + f(k) \), and no \( h \)-cores for \( h > k + f(k) \). The (experimental) values for this \( h \) are summarized in the following chart:
Figure 5.1: Maximum $h$ (in simulation) such that $G_{n,k}^\alpha$ contains an $h$-core.

5.4 Hamiltonicity of $M_{n,k}^\alpha$ and $G_{n,k}^\alpha$

One of the primary areas of study regarding the $k$-out model of Fenner and Frieze [19] has been Hamiltonicity. A Hamilton cycle in a graph $G$ is a cycle which passes through all of the vertices of $G$, entering and leaving each vertex exactly once; a graph which contains a Hamilton cycle is called Hamiltonian. A directed Hamilton
cycle is a Hamilton cycle of arcs in a directed graph, where the arcs have a common orientation.

Fenner and Frieze [20] proved that $G_{k-out}$ is Hamiltonian for $k \geq 23$. Frieze [22] improved this to $k \geq 10$. This was further improved to $k \geq 5$ by Frieze and Łuczak [21], before finally being reduced to $k \geq 3$ by Bohman and Frieze [4]. As $G_{2-out}$ is demonstrably not Hamiltonian, this completed the study of Hamiltonicity of $G_{k-out}$.

The graph $G_{n,k}^a$ does not necessarily have minimum degree 3 like $G_{k-out}$; however, we have seen that the shortfall is minimal. Despite this, the expected number of collections of $n$ arcs in $M_{n,k}^a$ that induce Hamilton cycles in $G_{n,k}^a$ seems to tend to 0 for $a$ sufficiently close to 0 and fixed, so that w.h.p. $G_{n,k}^a$ does not admit a Hamilton cycle. This is surprising; in all other properties, $M_{n,k}^a$ and $M_{n,k}^\infty$ have been qualitatively the same (although quantitatively different) as long as $a$ is bounded away from 0.

This begs the question: supposing it is the case that $G_{n,k}^a$ is w.h.p. not Hamiltonian for $a$ sufficiently small, is it then the case that there is a threshold $a^*$ for Hamiltonicity? If so, how strong of a threshold? This is likely to be a very difficult question to answer, but would be a very nice result. The same questions can be asked about (directed) Hamiltonicity of $M_{n,k}^a$; presumably, if there is a threshold $a^*$ for directed Hamiltonicity, then $a^* > a^*$.

### 5.5 Epidemic Processes

The study of “infection spread” processes on random mappings was initiated by Gertsbakh [23]. Essentially: suppose that we are given a mapping $f : [n] \to [n]$, and some subset $K \subseteq [n]$ of initially infected vertices. Say that infection spreads in one of the following three ways: in the direct process, infection spreads from $j$ to $f(j)$; in the inverse process, infection spreads from $f(j)$ to $j$; in the two-sided process, infection spreads both from $i$ to $f(i)$ and from $f(j)$ to $j$. 66
Considering uniformly random mappings \([n] \to [n]\) which have no fixed points, Gertsbakh found that in the direct and inverse processes, no significant fraction of the vertices is infected if \(o(\sqrt{n})\) vertices are infected initially; on the other hand, a significant fraction of the vertices is infected in the two-sided process if \(m = m(n) \to \infty\) vertices start infected, no matter how slowly \(m \to \infty\).

Burtin [10] studied the inverse process with \(m\) initially infected vertices, finding explicit limiting distributions for the number of eventually infected vertices for several different (asymptotic) relationships between \(m\) and \(n\). Pittel [30] undertook a similar study of the direct and two-sided processes.

These processes could make for an interesting study in \(M_{n,k}^\alpha\); in fact, they could make an interesting study in Hansen and Jaworski’s exchangeable in-degree model. Many of the ingredients necessary to study these processes in the exchangeable in-degree model are already in place; for instance, Hansen and Jaworski [25] have already established results regarding the size of the successor and predecessor sets.

An interesting generalization of these processes, appropriate for study in \(M_{n,k}^\alpha\), is bootstrap percolation. This process, initially introduced by Chalupa, Leith, and Reich [11], works as the epidemic processes, except that a vertex must receive the infection from at least \(r\) sources before it becomes infected. The cases \(r = k - 1\) and \(r = k\) are very natural questions for study on \(M_{n,k}^\alpha\); in particular, there is hope to get results in the direct process, as we have a great deal of information about the in-degrees of vertices.

5.6 Variations on \(M_{n,k}^\alpha\)

In Section 2.4, we saw that the order in which decisions are made in generating \(M_{n,k}^\alpha\) can be changed without changing the distribution of the result. Let us consider the
following order: starting with vertex 1, each vertex makes their first choice; then each
makes their second; etc.

This can be thought of as representing a voting scheme, for instance in a committee
electing its chairman, as follows: there are \( n \) members of the committee; there are \( k \)
rounds of voting; within each round, votes are taken one at a time and spoken out
loud. Thus when a particular vote is cast, the voter has full information regarding all
previous votes and current standings - and is in some way swayed by that information,
to an extent determined by the size of \( \alpha \).

This voting scheme is very natural, and a fitting interpretation of \( M_{n,k}^\alpha \); but, it
is of course not the only natural voting scheme possible for this type of situation.
Following are a couple of examples of other schemes, and descriptions of how they
translate in to \( k \)-out digraph models.

Multiple rounds of written ballots. In this scheme, there are \( k \) rounds of voting
in which each member casts one vote, but votes are cast simultaneously - that is,
voters have full information about previously completed rounds of voting, but no
information about the other votes made during the current round. Formally:

**Definition.** Choose a sequence of \( k \) mappings \( M_1, M_2, \ldots, M_k : [n] \rightarrow [n] \) as follows:

choose \( M_1 \) uniformly at random among all mappings \( [n] \rightarrow [n] \). Conditioned on
\( M_1, \ldots, M_j \), choose \( M_{j+1} : [n] \rightarrow [n] \) by letting each \( v \in [n] \) independently choose an
image, with

\[
P(M_{j+1}(v) = w \mid M_1, \ldots, M_j) = \frac{D_j(w)}{\sum_{i=1}^{n} D_j(i)},
\]

where \( D_j(w) \) is the number of times \( w \) was chosen as an image in \( M_1, \ldots, M_j \).

Let \( M_{n,k}^\alpha \in \mathcal{M}_{n,k} \) be the \( k \)-out map obtained by combining these \( k \) mappings:
that is, the \( j \)th arc from vertex \( v \) targets \( M_j(v) \).

One round of spoken votes, multiple choices. The idea here is that each voter
places \( k \) votes, which are spoken out loud - so that people who vote later have the
information in hand. There are several decisions to be made here, however - for instance, how are the $k$ choices made? Are they independent of each other? Are repeats allowed? There are several avenues of exploration here.

### 5.7 Generalization of Hansen and Jaworski’s Exchangeable In-Degree Model

Part of the beauty of Hansen and Jaworski’s original model for random mappings with exchangeable in-degrees lies in the sheer number of exact results that can be proved with very little knowledge of the distribution in play. Attempting to generalize this to the setting of $k$-out maps could prove very interesting.

However, it is a study fraught with difficulty. Mappings are amenable to analysis by generating function methods: a mapping is essentially a forest of rooted trees, together with a permutation on the roots. The collection of $k$-out mappings, however, has no such nice structure known; they are much more complicated and chaotic. This makes finding closed-form expressions for probabilities unlikely; however, it is still possible that asymptotic results may be obtained by making minimal assumptions on the in-degrees.

### 5.8 The Buckley-Osthus Model

The *web graph* is an abstraction of the world wide web, where vertices represent web pages and edges represent links between documents. Analysis of the web, along with several other large-scale networks, has revealed a structure that is markedly different from the structure of the Erdős-Rényi random graph; most remarkably, the proportion of vertices of degree $d$ is proportional to $d^{-\gamma}$ for some fixed constant $\gamma$. In
In this case, we say that the in-degree sequence satisfies a *scale-free power law*. See, for instance, the survey by Bollobás and Riordan [7].

In [9], Buckley and Osthus formalized a preferential attachment model that was previously discussed non-rigorously by Dorogovtsev, Mendes, and Samukhin [12] and Drinea, Enachescu, and Mitzenmacher [13]. The model is described in two stages. For a fixed parameter $\alpha > 0$, build $\{G_n^1\}_{n=1}^{\infty}$ as follows: $G_1^1$ is a graph with vertex set $\{1\}$ and one loop $1 \to 1$. For $n > 1$, $G_n^1$ is obtained from $G_{n-1}^1$ by adding a new vertex, $n$, and one new arc $n \mapsto \nu_n$ for some $\nu_n \in [n]$, with

$$P(\nu_n = j \mid G_{n-1}^1) = \frac{d_{n-1}(j) + \alpha}{\alpha n + n - 1},$$

where $d_{n-1}(j)$ is the in-degree of vertex $j$ in $G_{n-1}^1$ (and $d_{n-1}(n) = 0$). Now, for $k \in \mathbb{N}$, we define a $k$-out version of $G_n^1$, say $G_n^k$, by computing $G_{kn}^1$ and identifying vertices $1, \ldots, k$, vertices $k+1, \ldots, 2k$, and so on.

Their main result in [9]: if $\alpha \in \mathbb{N}$, then the proportion of vertices with in-degree $d$ is asymptotically proportional to $d^{-2-\alpha}$, uniformly over $0 \leq d \leq n^{1/100(\alpha+1)}$. Grechnikov [24] then extended this result to the case where $\alpha \notin \mathbb{N}$, as well as removing the upper bound on $d$.

The parallels between $G_n^k$ and our mapping $M_{n,k}^\alpha$ are clear; but, the development of $G_n^k$ as a growing process fundamentally changes the structure involved. It seems that very little study has been carried out on this model beyond establishing the desired result that the in-degrees satisfy a scale-free power law; however, every extra structural property of such a model gives more of a basis for comparison with the web graph, and could therefore prove useful. This is perhaps an area in which I can be of help.
BIBLIOGRAPHY


### Appendix A

#### GLOSSARY OF NOTATION

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a^\bar{b}$</td>
<td>The rising factorial $a(a + 1) \cdots (a + b - 1)$. See page 12.</td>
</tr>
<tr>
<td>$(a)_b$</td>
<td>The falling factorial $a(a - 1) \cdots (a - (b - 1))$. See page 12.</td>
</tr>
<tr>
<td>${a}_{b}$</td>
<td>The Stirling partition number: the number of ways to partition $a$ elements into $b$ nonempty sets.</td>
</tr>
<tr>
<td>$C_v(G)$</td>
<td>The <em>vertex connectivity</em> of the graph $G$; that is, the maximum $h$ such that $G$ is $h$-connected. See Section 3.3.</td>
</tr>
<tr>
<td>$D_n, D^\alpha_n$</td>
<td>The in-degree sequence for $M^\alpha_{n,k}$. The $\alpha$ is left off when doing so will not create confusion. See page 15.</td>
</tr>
<tr>
<td>$D_{n,j}, D^\alpha_{n,j}$</td>
<td>The in-degree of vertex $j$ in $M^\alpha_{n,k}$. The $\alpha$ is left off when doing so will not create confusion. See page 15.</td>
</tr>
<tr>
<td>$\delta(G)$</td>
<td>The minimum vertex degree of the graph $G$.</td>
</tr>
<tr>
<td>$\Delta_m(M)$</td>
<td>The maximum vertex in-degree of the digraph/mapping $M$.</td>
</tr>
<tr>
<td>$G^\alpha_{n,k}$</td>
<td>The graph obtained from $M^\alpha_{n,k}$ by ignoring loops and multiple edges. See page 8.</td>
</tr>
<tr>
<td>$\lambda(\alpha, k)$</td>
<td>The (limiting) expected number of vertices of degree $k - 1$ in $G^\alpha_{n,k}$. See page 26.</td>
</tr>
</tbody>
</table>
\( \mathcal{M}_{n,k} \) The set of \( k \)-out mappings on \([n]\) - that is, functions \([n] \to [n]^k\) - given the discrete \( \sigma \)-algebra. See page 4.

\( \mathcal{M}_{n,k}(d) \) The set of \( M \in \mathcal{M}_{n,k} \) which have in-degree sequence \( d \). See page 6.

\( M_{n,k}^\alpha \) The random \( k \)-out mapping with preferential attachment. See page 7.

\( \mu_{s,\alpha} \) The \( s \)th moment of \( D_{n,j} \). Computed in Lemma 2.3.3(ii), page 20.

\( \mu_{s,\alpha}' \) The \( s \)th moment of \( Z_{n,j} \). Computed in Lemma 2.3.1(iii), page 17.

\( \mathcal{N}(\mu,\sigma^2) \) A normally-distributed random variable with mean \( \mu \) and variance \( \sigma^2 \).

\([n]\) For \( n \in \mathbb{N} \), the set \( \{1, \ldots, n\} \).

\([n]^k\) The set of \( k \)-long vectors with elements in \([n]\).

\( P_{n,k}^\alpha \) The probability measure on \( \mathcal{M}_{n,k} \) associated with \( M_{n,k}^\alpha \). See page 45.

\( \|P_1 - P_2\|_{TV} \) The total variation distance between measures \( P_1 \) and \( P_2 \). See page 46.

\( V_{n,d} \) The set of vertices in \( G_{n,k}^\alpha \) which have degree exactly \( d \). See page 27.

\( w_\alpha(M) \) The \( \alpha \)-weight of the mapping \( M \in \mathcal{M}_{n,k} \). See page 13.

w.h.p. Abbreviation of “with high probability”. Indicates that an event for \( M_{n,k}^\alpha \) occurs with probability approaching 1 as \( n \to \infty \). See page 11.

\( Z_n \) The vector \((Z_{n,1}, \ldots, Z_{n,n})\) used in the alternative formulation of \( M_{n,k}^\alpha \).

See page 16.

\( Z_{n,j} \) A specially chosen random variable with a generalized negative binomial distribution, used in the alternative formulation of \( M_{n,k}^\alpha \). Defined in (2.3.1), page 16.