MASTER

Ultra-small world phenomenon in the directed configuration model

Colenbrander, D.G.J.

Award date:
2022

Technische Universiteit Eindhoven
University of Technology
Department of Mathematics and Computer Science
Architecture of Information Systems Research Group

Ultra-small world phenomenon in the directed configuration model

Master Thesis

Deen Colenbrander

Supervisors:
Remco van der Hofstad
Pim van der Hoorn
Oliver Tse

Final version

Eindhoven, June 2022

Abstract

In this thesis we investigate the structure under which the typical distances in directed configuration model scale double-logarithmically with the size of the graph. We study the necessary conditions and show how they lead to this phenomena. We then study how these typical distances fluctuate under changes to the graph size. Moreover, we add a possible truncation the the in- and out-degrees, and study its effect on the structure of the model.

Contents

Chapter 1

Introduction

Anyone who is interested in random graphs, most likely heard of the game 'six degrees of Kevin Bacon'. The game consists of a network of actors, in which the actors are connected if they have worked on the same movie. To play it, you start with an arbitrary actor, and try to find the shortest path to the actor Kevin Bacon, using the connection between the actors. The 'six degrees' is a reference to the theorem that people in the world are typically only six degrees of separation apart. This phenomenon is known as a *small world*. An important property of the connections between people in this network, is that they are undirected. This means that if person 'A' knows person 'B', that it must be that person 'B" also knows person 'A'. A similar game where this is not the case is the Wiki Game. This game consists of a network of Wikipedia articles, connected through their hyperlinks. Here, you start at an arbitrary chosen article, and try to reach another given article in the least amount of steps, using the hyperlinks in each article. It can be the case that article 'A' has an hyperlink to article 'B', but 'B' has no hyperlink to article 'A'. Thus, the connections in this network are *directed*.

1.1 Motivation and summary of main results

The study of small worlds in networks attracted attention after S. Milgram published an article [10], describing the results of a social experiment he had conducted. For the experiment a randomly selected starting and target person were taken from the American population, where the goal was to send a note from the starting person to the target person. The catch is that the note could only be passed to someone known by first name basis, such that each person needs to pass the note to someone who is likely to be closer to the target. Against intuition, the experiment showed that on average only five to six intermediate acquaintances were needed to reach the target. This result sparked the curiosity of many researchers how this phenomenon arises and to find what other networks were small worlds. One such example is the Kevin Bacon network [18] that was mentioned before. A very different example is the network formed by the neurons and synapses in our brain [9], which presumably results from natural selection under the pressure of a cost-efficiency balance.

For many of these small world networks, it has been shown that they follow a *scale free* paradigm [17]. When network sciences were still young, this was defined to mean that the number of nodes with k neighbors for $k \geq k_{\min}$ is proportional to $ck^{-\gamma}$, where c represents the normalization constant. In reality, measurement are prone to different types of magnitudes of noise and fluctuations, such that it is quite rare to find such clean patterns in real-world networks [4]. Hence, we take vertex degree i.e., the number of neighbors of each node, to have a *regularly varying distribution*, which contains distributions with probability density functions of the form

$$p(k) = l(k)k^{-\tau}, \tag{1.1}$$

where $l(k)$ is a slowly varying function. The precise definition of a scale-free network that we shall take, is that the tail distribution of the degrees is of the form

$$\bar{F}(k) = \sum_{k'>k} p(k) = l'(k)k^{-(\tau-1)}, \tag{1.2}$$

for which $l'(k)$ is also a slowly varying function. Distributions with a tail distribution of the form (1.2) are defined to be *power law*. Note that in the directed case it can be that *both* the proportion for the connections towards and from nodes, i.e. the in- and out-degree distribution, are power law but it more commonly holds for just for the in- *or* out-degree distribution. Examples can be found in the article [17], e.g. the degree of the number of people that follow a twitter user and the number the user follows are both power law, and the interactions between proteins in humans has only a power-law distributed out-degree.

A consequence of the scale-free pattern, is that there are large numbers of high degree vertices, which we call *hubs*. As these hubs are connected to a large portion of the nodes, they act as a short-cut in the network, drastically decreasing the typical distances between the nodes. It has been shown for undirected small-world networks, that these distances further rely on the variability of the degrees [5]. The typical length of the distances of the scale free network grow *logarithmically* with the size of the network, but for networks with estimated power law exponent $\hat{\tau} \in (2, 3)$ as defined in (1.2), the distances are much smaller. These networks are defined as *ultra-small worlds*. The knowledge gained from studying these small world networks can be applied in a broad range of sciences. For instance, it can used to model the spread of epidemic diseases by epidemiologist to prevent outbreaks, or engineers can use the network of the brain as an inspiration to reconstruct its efficiency in devices.

Choice of mathematical models: When constructing mathematical models to study these networks, the main thing we want to recreate are the characteristics of the degrees of the nodes. An easy way to do this, is by starting with a fixed sequence of degrees, which can be assigned to the nodes, and randomly pairing the nodes based on their degree. Such a model was introduced by Bollobas [3], called the *configuration model*. For the construction of this model, each node is assigned a fixed number of *half-edges*. Connections between the nodes are then randomly constructed, by pairing the half-edges uniformly at random. For the construction of the directed network a similar approach is used, but now every node is assigned a fix number of *inbound* and *outbound* half-edges. The directed connections are constructed, by pairing each outbound half-edge with an uniformly at random chosen inbound half-edge that has not been paired yet. The resulting random graph is called the *directed configuration model*. A more precise construction of the models will be introduced further along the thesis. The versatility of these models makes them useful for studying the properties of real world networks [12][8]. The property we are interested in is the typical distance in the graph, i.e., the least amount of edges that need to be traversed to travel between two randomly chosen vertices. More specifically, we want to know under what circumstances these typical distances in the directed network become ultra small.

Description of main results: For the configuration model of size n, it has been shown that the typical distances between vertices, conditioned on them being in the same component, grows logarithmic with n if the empirical degree distribution follows a power law degree distribution with finite variance. When the distribution has infinite variance, which is the case when the empirical degree distribution follows a power law with exponent $\tau \in (2, 3)$, it is shown ([14], Section 7) that the typical distances are even $2 \log \log n / |\log(\tau - 2)|$, confirming the phenomenon seen in real networks.

In the directed configuration model, the covariance between the in- and out-degree distribution also needs to be taken in consideration. It has been shown that under the assumption that the covariance is finite, the typical distances grow logarithmically, even if the variance of the in- and/or out-degree is infinite [16]. This bring us to the research topic of this thesis: under what conditions does the directed configuration model experience the ultra-small world phenomenon? To study this, we take a close look at the reason behind the conditions that are used in the proof from ([14], Section 7) are required to obtain the ultra-small distances in the configuration model. Next, we discuss why these exact conditions are not enough to replicate the phenomenon in the directed configuration model. Then we reconstruct the conditions to replicate the phenomena that lead to ultra-small typical distances in the directed configuration model. The following describes the first result obtained in this thesis. For now the details of the assumptions are omitted, but will be described precisely in Theorem 1.5.4.

Main result 1: *Consider the directed configuration model on n vertices for which the in- and out-degree satisfy some appropriate regularity conditions, and the empirical distribution of the in- and out-degree of neighboring nodes follows a power law with respective exponent* $\tau^{(in)} \in (2,3)$ *and* $\tau^{(out)} \in (2,3)$. *Let* $\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1, o_2)$ *denote the (directed) distance between two randomly chosen vertices. Then, on the condition that the vertices are in the same component it follows that*

$$\frac{\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1, o_2)}{\log\log n} \xrightarrow{\mathbb{P}} \frac{1}{|\log(\tau^{(\mathrm{in})}-2)|} + \frac{1}{|\log(\tau^{(\mathrm{out})}-2)|}. \tag{1.3}$$

The next result adds a possible truncation to the power law distribution of the neighboring in- and out-degrees, and describes how the typical distances fluctuate with n. The proof is derived from the article [15]. Here it is shown for the configuration model, that if the power law is truncated at some value n^{β_n}, where $\beta_n(\log n)^\eta \to \infty$ for some $\eta \in (0,1)$, that the typical distances are centered around $2\log\log(n^{\beta_n})/|\log(\tau-2)| + 1/(\beta_n(3-\tau))$. By taking the dependence of the truncation values of in- and out-degree into consideration, we obtain the following result:

Main result 2: *Consider the directed configuration model on n vertices and* l_n *out- and inbound half-edges, for which the in- and out-degree satisfy some appropriate regularity conditions. Moreover, the empirical distribution of the in- and out-degree of neighboring nodes follows a power law with respective exponent* $\tau^{(in)} \in (2,3)$ *and* $\tau^{(out)} \in (2,3)$ *truncated at degrees* $l_n^{\beta_n^{(in)}}$ *and* $l_n^{\beta_n^{(out)}}$, *where both* $\beta_n^{(q)}(\log n)^\eta \to \infty$ *for* $q \in \{out, in\}$ *and some* $\eta \in (0,1)$. *Then, given that*

$$\beta_n^{(\mathrm{in})} + \beta_n^{(\mathrm{out})}(\tau^{(\mathrm{out})}-2) = \beta_n^{(\mathrm{out})} + \beta_n^{(\mathrm{in})}(\tau^{(\mathrm{in})}-2) \in (0,1), \tag{1.4}$$

it follows that

$$\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1, o_2) - \frac{\log\log l_n^{\beta_n^{(\mathrm{out})}}}{|\log(\tau^{(\mathrm{out})}-2)|} - \frac{\log\log l_n^{\beta_n^{(\mathrm{in})}}}{|\log(\tau^{(\mathrm{in})}-2)|} - \frac{2}{\beta_n^{(\mathrm{out})}(3-\tau^{(\mathrm{out})}) + \beta_n^{(\mathrm{in})}(3-\tau^{(\mathrm{in})})} \tag{1.5}$$

is a tight sequence of random variables.

For the intuitive explanation behind the result, we look at a common trait of the directed and the undirected configuration models. Namely, that a small enough neighborhood of a typical vertex is unlikely to contain any cycles or multi-edges. This so called local-tree like structure, allows these neighborhoods to be approximated by a properly defined branching process, which experience a double exponential generational growth rate [7] under the conditions of the results. Each generation k of these branching processes represents either, the number of vertices that can be reached in k steps from a typical vertex, or the number of vertices that can reach a typical vertex in k steps. As these are the same in the configuration model, the double exponential growth results in the typical distances $2\log\log n/|\log(\tau-2)|$, representing the similar distance from the starting vertex towards the hubs and from the hub to the target vertex. As the in- and out-degree have their own distributions in directed configuration model, these branching processes also have different distributions. Hence, we obtain $\log\log l_n/|\log(\tau^{(\mathrm{out})}-2)|$ to represent the distance to the highest order out-degree vertices called the outbound-hub, and $\log\log l_n/|\log(\tau^{(\mathrm{in})}-2)|$ for the distance *from* the highest order in-degree vertices, called the inbound-hub. For the non-truncated model, the vertices in the outbound-hub are very likely to be connected to all the vertices in the inbound-hub, such that the typical distance contains just these two terms. For the truncated model, the highest order out- and in-degrees and the typical distances to and from them, depends on the truncation values. And such, it no longer has to be the case they are likely connected. Hence the third term represents the value around which the remaining distance from the outbound to the inbound hubs is centered.

The remainder of this section introduces and discusses the results of the thesis in detail. First, in Section 1.2 we describe how the directed configuration model is constructed and introduce some notation and definitions. Next, in Section 1.3 we study the local tree-like structures and show how its distribution relates to a branching process. In Section 1.4 the conditions are shown under which uniformly chosen vertices are almost surely in the same unique strongly connected component. Finally, in Section 1.5 the necessary power law conditions are discussed and the asymptotic ultra-small world result is introduced. Then, the results describing the fluctuations of the typical distances are shown in Section 1.6 with a possibly truncated model.

Notation: We write $[n]$ to indicate the set of integers $\{1, 2, ..., n\}$. For some graph G_n with $n \in \mathbb{N}$ vertices we use $V(G_n) = [n]$ and $E(G_n)$ to indicate the respective set of vertices and edges of the graph . We will use $\xrightarrow{d}$, $\xrightarrow{\text{a.s.}}$ and $\xrightarrow{\mathbb{P}}$ to denote convergence in distribution, almost surely and convergence in probability.

1.2 Construction of the directed configuration model

Let $\boldsymbol{d}^{(\text{out})} = (d_u^{(\text{out})})_{u\in[n]}$ be a sequence of out-degrees, where $d_u^{(\text{out})}$ denotes the number outbound half-edges assigned to vertex $v \in [n]$. Similarly, let $\boldsymbol{d}^{(\text{in})} = (d_u^{(\text{in})})_{u\in[n]}$ be a sequence of in-degrees, with the condition that the equality

$$l_n = \sum_{u\in[n]} d_u^{(\text{out})} = \sum_{u\in[n]} d_u^{(\text{in})} \tag{1.6}$$

is satisfied. By assigning the in- and out-degree pair $(d_u^{(\text{in})}, d_u^{(\text{out})})$ to each vertex $u \in [n]$ we obtain the degree sequence $\boldsymbol{d} = (\boldsymbol{d}^{(\text{in})}, \boldsymbol{d}^{(\text{out})})$. To construct the path we pick an arbitrary in- or outbound half-edge and a randomly chosen respective out- or inbound half-edge, without replacement. The chosen in- and outbound half-edges are then paired to form an edge. So at step $k+1$ an arbitrary in- or outbound half-edge is taken from the $l_n - k$ that remain, and is randomly paired with a respective out- or inbound half-edge from the $l_n - k$ that remain from this type of half-edge. Say that the outbound half-edge attached to vertex u is paired with an inbound half-edge attached to vertex v. Then this edge is denoted as (u, v), and can only be traversed from u towards v. After all the half-edges have been paired, the directed configuration model with fixed degree sequence $\boldsymbol{d}$ is constructed, denoted as $\text{DCM}_n(\boldsymbol{d})$.

The directed configuration model is a multi-graph, which means that there can be multiple edges between two vertices. The model also contains self-loops, which arise when an inbound half-edge of a vertex is paired with one of its own outbound half-edges. Let $D_n = (D_n^{(\text{in})}, D_n^{(\text{out})})$ denote the in- and out-degree of a randomly chosen vertex, which we will refer to as a *typical vertex* of the graph. The proportional distribution of the degrees is denoted as

$$f_n(k,l) = \mathbb{P}(D_n = (k,l)) = \frac{1}{n} \sum_{u\in[n]} \mathbb{1}\{(d_u^{(\text{in})}, d_u^{(\text{out})}) = (k,l)\}, \tag{1.7}$$

for which the cumulative distribution function is denoted by F_n. The graph distance between two vertices $u, v \in [n]$ is k if this is the least amount of directed edges that need to be traversed to get from vertex u to v, denoted as the event $\{\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(u, v) = k\}$. Ofcourse, there needs to be a directed path for it to even be possible to travel from u to v. If this is not the case, such that no path exists, we say that $\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(u, v) = \infty$.

1.3 Directed branching process approximation

This section we will study the directed configuration model from the perspective of a typical vertex. It is possible that an event that is likely to occur on the whole graph, is uncommon to take place on a small enough subgraph. This may allow us to dismiss this event when only studying the properties of this particular subgraph. We will show that on the locally finite directed configuration model, this is the case for cycles and multi-edges.

To understand this phenomenon, we take another look at the construction process of the model. As explained before, to construct an edge we can take *any* arbitrary in- or outbound half-edge, and pair it with a respective uniformly chosen out- or inbound half-edge. It follows that we can also construct the model in a breadth-first order, without affecting the randomness of the model. That is, instead of taking any arbitrary half-edge, we start with an uniformly chosen vertex, and uniformly pair all its outbound and inbound half-edges. Next, we repeat this process one-by-one, with all the vertices attached to the uniformly chosen half-edges, until again all their respective half-edges have been paired. We call this process the *forward-backward exploration* of the graph. The same process can also be performed by only pairing the outbound *or* inbound half-edges of each vertex, which we respectively call the *forward* and *backward* exploration processes. By repeating these processes with each of the added vertices a sufficient number of times, it is possible to construct the following neighborhoods.

Definition 1.3.1. (Neighborhoods) *The forward r-neighborhood of vertex o, denoted as $B_r^{(G;out)}(o)$, is the subgraph containing all vertices at most distance r from vertex o, such that*

$$\begin{aligned} V(B_r^{(G;\text{out})}(o)) &= \{v \in V(G) : \text{dist}_G(o,v) \le r\} \\ E(B_r^{(G;\text{out})}(o)) &= \{(u,v) \in E(G) : \text{dist}_G(o,u) \le r, \text{dist}_G(o,v) \le r\}. \end{aligned} \tag{1.8}$$

The backward r-neighborhood of vertex o, denoted as $B_r^{(G;in)}(o)$, is the subgraph containing all vertices with distances at most r towards vertex o, such that

$$\begin{aligned} V(B_r^{(G;\text{in})}(o)) &= \{v \in V(G) : \text{dist}_G(v,o) \le r\} \\ E(B_r^{(G;\text{in})}(o)) &= \{(u,v) \in E(G) : \text{dist}_G(u,o) \le r, \text{dist}_G(v,o) \le r\}. \end{aligned} \tag{1.9}$$

The forward-backward r-neighborhood of vertex o, denoted as $B_r^{(G)}(o)$, is the union of both neighborhoods $B_r^{(G;out)}(o) \cup B_r^{(G;in)}(o)$.

To ensure that the size of these neighborhoods is bounded for fixed values of r as $n \to \infty$, we introduce the condition that the typical degree converges weakly to a random variable with finite mean.

Assumption 1.3.2. (Regularity conditions) *We assume that the vertex in- and out-degree satisfy the following regularity conditions.*

(a) **Weak convergence of vertex in- and out-degrees.** *There exists a distribution function F, such that*

$$D_n = (D_n^{(\text{in})}, D_n^{(\text{out})}) \xrightarrow{d} (D^{(\text{in})}, D^{(\text{out})}) = D, \tag{1.10}$$

where D has cumulative distribution F.

(b) **Convergence of average vertex in- and out-degrees**.

$$\lim_{n\to\infty} \mathbb{E}[D_n^{(\text{in})}] = \lim_{n\to\infty} \mathbb{E}[D_n^{(\text{out})}] = \mathbb{E}[D^{(\text{in})}] = \mathbb{E}[D^{(\text{out})}] > 0. \tag{1.11}$$

Under Assumption 1.3.2 the total number of out- and inbound half-edge is very large, as $l_n/n = \mathbb{E}[D_n^{(\text{out})}] \to \mathbb{E}[D^{(\text{out})}] > 0$. Therefore, many exploration steps can be taken until the neighborhood contains any significant proportion of the total l_n outbound or inbound half-edges. Consequently, for each of these pairing steps it is unlikely that the in- or outbound half-edge is paired with a half-edge attached to a vertex that already belongs to the explored neighborhood. By bounding the number of exploration steps compared to the graph size, it is unlikely to add any multi-edges or cycles to the neighborhood.

In the following sections we will show that it is possible to approximate these neighborhoods with a properly defined branching processes. In Section 1.3.1 we explain how each of the exploration processes can be coupled with branching processes. In Section 1.3.2 we introduce the marked graph and its applications in the study of local weak convergence. We construct a one-to-one mapping from the directed graph to these marked graphs, which will enable us to apply these applications to study the local properties of the directed graph. The results of this study are introduced in Section 1.3.3.

1.3.1 Coupling explorations with branching processes

Let us take a look at the distribution of the explored neighborhood. The exploration process starts at a typical vertex with degree distribution D_n. If the first exploration step starts from an outbound half-edge, the probability that the randomly chosen inbound half-edge is attached to a vertex with degree (k, l) depends on the number of outbound half-edges that attached to vertices with degree (k, l). Hence, the probability equals $\frac{k}{\mathbb{E}[D_n^{(\text{in})}]}\mathbb{P}(D_n = (k, l))$, for which $\mathbb{E}[D_n^{(\text{in})}]$ is the normalization constant. Similarly, if the first step starts from an inbound half-edge, the degree distribution of the following explored vertex equals $\frac{l}{\mathbb{E}[D_n^{(\text{out})}]}\mathbb{P}(D_n = (k, l))$. For each of the following exploration steps, this distribution changes slightly as the total number of available in- and outbound half-edges depletes. Though, as the total number of free out- and inbound half-edges l_n at the start is so large under Assumption 1.3.2, a significant number of exploration steps can be performed before the degree distribution of the following explored vertex has a noticeable difference. By bounding the number of exploration steps, the degrees of the explored vertices are close to independently distributed. Combined with the unlikely occurrence of cycles and multi-edges, this intuitively shows the breadth-first exploration has a very similar distribution as to a branching process.

It remains to define these branching processes, to represent each type of exploration. As we mentioned before, for the forward-backward exploration the degree distribution of the explored vertices depends on whether they were chosen by their out- or inbound half-edge. Hence, we define the (delayed) two-type branching process $\mathcal{Z}_n$ with root distribution D_n and offspring distributions

$$\begin{aligned} f_n^{*(\text{in})}(k,l) &= \frac{(k+1)}{\mathbb{E}[D_n^{(\text{in})}]} f_n(k+1,l), \\ f_n^{*(\text{out})}(k,l) &= \frac{(l+1)}{\mathbb{E}[D_n^{(\text{out})}]} f_n(k,l+1), \end{aligned} \tag{1.12}$$

for the following generations. Whether an individual has offspring distribution $f_n^{*(\text{in})}$ or $f_n^{*(\text{out})}$ depends on the type of half-edge it is connected with to the previous generation. We say the individuals with offspring distribution $(f_n^{*(\text{in})}(k,l))_{k,l\geq 0}$, connected to the previous generation by an inbound half-edge is an "in-type", and individuals with offspring distribution $(f_n^{*(\text{out})}(k,l))_{k,l\geq 0}$, connected to the previous generation by an outbound half-edge is an "out-type".

The forward and backward exploration only expands by their respective out- or inbound half edges. Hence, the forward exploration is represented by the branching process $\mathcal{Z}_n^{(\text{out})}$ with root distribution $D_n^{(\text{out})}$ and for the following generations $f_n^{(\text{out})}(l) = \sum_k \frac{k}{\mathbb{E}[D_n^{(\text{out})}]} f_n(k,l)$, denoting the total number of inbound half-edges attached to vertices with out-degree l. For the same reasoning, the backward exploration is represented by the branching process $\mathcal{Z}_n^{(\text{in})}$ with root distribution $D_n^{(\text{in})}$ and for the following generations the offspring distribution $f_n^{(\text{in})}(k) = \sum_l \frac{l}{\mathbb{E}[D_n^{(\text{in})}]} f_n(k,l)$. We will write

$$\begin{aligned} F_n^{(\text{out})}(x) &= \sum_{l \le x} f_n^{(\text{out})}(l) = \frac{1}{l_n} \sum_{u \in [n]} d_u^{(\text{in})} \mathbb{1}\{d_u^{(\text{out})} \le x\} = \mathbb{P}(D_n^{*(\text{out})} \le x) \\ F_n^{(\text{in})}(x) &= \sum_{k \le x} f_n^{(\text{in})}(k) = \frac{1}{l_n} \sum_{u \in [n]} d_u^{(\text{out})} \mathbb{1}\{d_u^{(\text{in})} \le x\} = \mathbb{P}(D_n^{*(\text{in})} \le x), \end{aligned} \tag{1.13}$$

with

$$\begin{aligned} D_n^{*(\text{out})} &= (\text{size biased version of } D_n^{(\text{out})} \text{ w.r.t. the in-degree}), \\ D_n^{*(\text{in})} &= (\text{size biased version of } D_n^{(\text{in})} \text{ w.r.t. the out-degree}). \end{aligned} \tag{1.14}$$

Denoting the respective generation sizes of $\mathcal{Z}_n$, $\mathcal{Z}_n^{(\text{out})}$ and $\mathcal{Z}_n^{(\text{in})}$ as $(Z_k^{(n)})_{k\ge 0}$, $(Z_k^{(n;\text{out})})_{k\ge 0}$ and $(Z_k^{(n;\text{in})})_{k\ge 0}$, the precise result of the coupling is the following:

Lemma 1.3.3. *Consider the directed configuration model for which regularity Assumptions 1.3.2 hold. Define $d_{max} = \max_{u\in[n]}\{d_u^{(in)}, d_u^{(out)}\}$ and take r such that for uniformly chosen vertex o*

$$|B_r^{(G_n)}(o)| \le (n/d_{max})^{\frac{1}{2}-\delta}, \tag{1.15}$$

for some $\delta > 0$. Then $B_r^{(G_n)}(o)$ can be coupled to the branching process with generation sizes $(Z_k^{(n)})_{k\ge 0}$. Moreover, take r such that for two uniformly and independently chosen vertices o_1 and o_2

$$|B_r^{(G_n;\text{out})}(o_1) \cup B_r^{(G_n;\text{in})}(o_2)| \le (n/d_{max})^{\frac{1}{2}-\delta}, \tag{1.16}$$

for some $\delta > 0$. Then, $(B_r^{(G_n;out)}(o_1), B_r^{(G_n;in)}(o_2))$ can be coupled to the two independent branching processes with generation sizes $(Z_k^{(n;out)}, Z_k^{(n;in)})_{k\ge 0}$.

Next, we study the local behaviour of the graph with sizes in the limiting regime. Before we do this, we introduce marked (undirected) graphs.

1.3.2 Marked graphs

A *marked graph* is the undirected graph $\tilde{G}_n = (V(\tilde{G}_n), E(\tilde{G}_n))$ together with a set of marks $M(\tilde{G}_n)$, containing the mappings from $V(\tilde{G}_n)$ and $E(\tilde{G}_n)$ to the complete and separable metric space Ξ. As shown in ([14], Section 2), there are convenient tools for the marked graph that describe its local convergent properties. Conveniently, we will show that it is possible to describe the directed graph as a marked graph. This will enable us to define an one-to-one mapping of the directed graph to the marked graph, allowing us to make use of these tools to study the directed graph. As the marked graph is undirected, the information about the direction of the edges are stored in their respective marks, i.e.,

$$\Xi = \{\{(u,k),(v,l)\}, \{u,v\} \in E_n, k,l \in \{\text{out}, \text{in}\}, k \ne l\}\}. \tag{1.17}$$

Although the edges are not directed, we will say an edge u, v with mark $\{(u,\text{out}),(v,\text{in})\}$ is an out-edge of u and an in-edge of v. Take $G_n = \text{DCM}_n(\boldsymbol{d})$ and denote the mapping to the space marked graph as $\psi : G_n \mapsto (\tilde{G}_n, M(\tilde{G}_n))$, where the mapping is defined as

$$(u,v) \in E_n \Rightarrow \{u,v\} \in \tilde{E}_n, \text{ with } m(\{u,v\}) = \{(u,\text{out}),(v,\text{in})\} \in M(\tilde{G}_n). \tag{1.18}$$

As no information about the directed edge is lost, we also have the inverse mapping ψ^{-1} : $(\tilde{G}_n, M(\tilde{G}_n)) \mapsto G_n$, defined as

$$\{u,v\} \in \tilde{E}_n \text{ with } m(\{u,v\}) = \{(u,\text{out}),(v,\text{in})\} \in M(\tilde{G}_n) \Rightarrow (u,v) \in E_n. \tag{1.19}$$

As we study the neighborhood of a vertex a *root* vertex is also included, which will act as the centre of the neighborhood. Including this root, we obtain the *marked rooted graph*, denoted as $(\tilde{G}, o, M(\tilde{G}))$. Next, we define the neighborhood of the root.

Definition 1.3.4. (Marked neighborhood) *Let* dist_G *denote the graph distance of graph* $G = (V(G), E(G))$. *For a marked rooted graph* $(G, o, M(G))$, *the subgraph containing all vertices at most distance* r *from* o *is denoted as* $B_r^{(G,M(G))}(o)$, *such that*

$$\begin{aligned} V(B_r^{(G,M(G))}(o)) &= \{v \in V(G) : \mathrm{dist}_G(o, v) \leq r\} \\ E(B_r^{(G,M(G))}(o)) &= \{\{u, v\} \in E(G) : \mathrm{dist}_G(o, u) \leq r, \mathrm{dist}_G(o, v) \leq r\}. \end{aligned} \tag{1.20}$$

This is defined as the r*-neighborhood of* o.

Note that $\psi(B_r^{(G_n)}(o)) = B_r^{(\tilde{G}_n, M(\tilde{G}_n))}(o)$. For ease of notation, we will denote $B_r^{(\tilde{G}_n)}(o) = B_r^{(\tilde{G}_n, M(\tilde{G}_n))}(o)$ and simply mention whether we are using a marked graph. In order to study the local weak convergence, we first need to know the topology which we are working with. Let

$$x_m^{(i)} := \#\{e \in E(\tilde{G}_i) : m(e) = m\} \quad m \in \Theta, k, l \in \Xi, \tag{1.21}$$

denote the number of edges in $E(G_i)$ with label $m \in \Xi$. The following definition shows when marked graphs are considered to be equal.

Definition 1.3.5. (Marked rooted graph isomorphism) *Two marked rooted graph* $(\tilde{G}_1, o_1, M_1(\tilde{G}_1))$ *and* $(\tilde{G}_2, o_2, M_2(\tilde{G}_2))$ *are called isomorphic, notated as* $(\tilde{G}_1, o_1, M_1(\tilde{G}_1)) \simeq (\tilde{G}_2, o_2, M_2(\tilde{G}_2))$, *when there exists a bijection* $\phi : V(G_1) \mapsto V(G_2)$, *such that*

1. $\phi(o_1) = o_2$,
2. $x_m^{(1)} = x_{\phi(m)}^{(2)}$ *for all* $m \in \Xi$,

where for $m(\{u, v\}) = \{(u, k), (v, l)\}$ we have $\phi(m(\{u, v\})) = \{(\phi(u), k), (\phi(v), l)\}$.

The bijective function ϕ ensures that the given indexes of the vertices is not taken into consideration when comparing marked rooted graphs. We let $\mathcal{G}_*$ be the set of marked rooted graphs modulo isomorphisms. Next, we introduce the metric which we use to compare marked graphs.

Definition 1.3.6. (Marked rooted graphs metric) *Let* $(\tilde{G}_1, o_1, M_1(\tilde{G}_1))$ *and* $(\tilde{G}_2, o_2, M_2(\tilde{G}_2))$ *denote two marked rooted graphs. Let*

$$R_* = \sup\{r : B_r^{\tilde{G}_1}(o_1) \simeq B_r^{G_2}(o_2)\}, \tag{1.22}$$

and define

$$d_{\mathcal{G}_*}((\tilde{G}_1, o_1, M_1(\tilde{G}_1)), (\tilde{G}_2, o_2, M_2(\tilde{G}_2))) = 1/(R_* + 1) \tag{1.23}$$

The value R_* is the largest value of r such that $B_r^{\tilde{G}_1}(o_1)$ is isomorphic to $B_r^{\tilde{G}_2}(o_2)$. It can be shown that the space $\mathcal{G}_*$ with the metric $d_{\mathcal{G}_*}$ is a separable metric space.

Finally, we define the marked local weak convergence of a marked rooted graph.

Definition 1.3.7. (Marked local weak convergence) *Let* $(\tilde{G}_n, o_n, M(\tilde{G}_n))$ *denote a marked random graph. Then,*

(a) *we say that* $(\tilde{G}_n, M(\tilde{G}_n))$ *converges marked locally weakly to* $(G, o, M(G))$ *having law* μ, *when*

$$\mathbb{E}[h(\tilde{G}_n, o_n, M(\tilde{G}_n))] \xrightarrow{n\to\infty} \mathbb{E}_\mu[h(\tilde{G}, o, M(\tilde{G}))], \tag{1.24}$$

for every bounded and continuous function $h : \mathcal{G}_* \mapsto \mathbb{R}$, *where the expectation* $\mathbb{E}$ *is w.r.t. the random vertex* o_n, *and the random graph* $\tilde{G}_n$. *This is equivalent to* $(\tilde{G}_n, o_n, M(\tilde{G}_n)) \xrightarrow{d} (\tilde{G}, o.M(\tilde{G}))$,

(b) *we say that* $(\tilde{G}_n, M(\tilde{G}_n))$ *converges locally in probability to* $(\tilde{G}, o, M(\tilde{G}))$ *having law* μ *when*

$$\mathbb{E}[h(\tilde{G}_n, o_n, M(\tilde{G}_n))|\tilde{G}_n] \xrightarrow{\mathbb{P}} \mathbb{E}_\mu[h(\tilde{G}, o, M(\tilde{G}))] \tag{1.25}$$

for every bounded and continuous function $h : \mathcal{G}_* \mapsto \mathbb{R}$.

1.3.3 Marked locally weak convergence

To describe the local asymptotic behaviour of the directed configuration model, we introduce the asymptotic versions of the branching processes used in the coupling. As we use first show the properties on the marked graphs, the edges need to be labeled as described before. First, we have the (edge) marked branching process $\mathcal{Z}$, with root distribution $(f(k,l))_{k,l\geq 0}$ and the two types of offspring distribution $(f^{*(\text{out})}(k,l))_{k,l\geq 0}$ or $(f^{*(\text{in})}(k,l))_{k,l\geq 0}$, with

$$\begin{aligned} f(k,l) &= \lim_{n\to\infty} f_n(k,l) \\ f^{*(\text{out})}(k,l) &= \lim_{n\to\infty} f_n^{*(\text{out})}(k,l), \quad f^{*(\text{in})}(k,l) = \lim_{n\to\infty} f_n^{*(\text{in})}(k,l), \end{aligned} \tag{1.26}$$

which are properly defined under Assumption 1.3.2. The marked forward branching process $\mathcal{Z}^{(\text{out})}$ has root distribution $D^{(\text{out})}$ and $f^{(\text{out})}(l) = \lim_{n\to\infty} f_n^{(\text{out})}(l)$ for the following generations. Note that each generation is connected by their respective (undirected) out-edges to the following generation. The marked backward branching process $\mathcal{Z}^{(\text{in})}$ is defined similarly, except the role of the out- and in- is reversed.

Theorem 1.3.8. (Locally tree-like nature directed configuration model) *Consider the* $\text{DCM}_n(\boldsymbol{d})$, *where the in- and out-degrees* $\boldsymbol{d} = (\boldsymbol{d}^{(in)}, \boldsymbol{d}^{(out)})$ *satisfy Assumptions 1.3.2. Then, the directed configuration model converges marked locally weakly in probability to the marked branching process. Consequently, the respective forward and backward neighborhoods converges marked weakly in probability to the marked forward and backwards branching process.*

It can be convenient to view the neighborhoods that are approximated by the branching process as a single vertex, for which the total degree of the final generation indicates the degree of this vertex. To do this, we take the two uniformly chosen vertices $o_1, o_2 \in [n]$. For convenience, we take $H_r^{(n;\text{out})} = H_0^{(n;\text{in})} = 1$, and let $H_r^{(n;\text{out})}$ denote the number of unpaired outbound half-edges attached to the vertices at graph distance $r-1$ from vertex o_1, after forward exploring $B_r^{(G_n;\text{out})}(o_1)$ in a breadth-first manner. And let $H_r^{(n;\text{in})}$ denote the number of unpaired inbound half-edges attached to the vertices at graph distance $r-1$ *towards* vertex o_2, after backward exploring $B_r^{(G_n;\text{in})}(o_1)$ in a breadth-first manner. Thus, $H_1^{(n;\text{out})} = d_{o_1}^{(\text{out})}$ and $H_1^{(n;\text{in})} = d_{o_2}^{(\text{in})}$. Note that you might think that $H_r^{(n;q)} = |\partial B_r^{(G_n;q)}|$, which is only the case when $B_r^{(G_n;q)}$ is a tree. As this is whp indeed the case for for fixed r according to Theorem 1.3.8, we obtain the following corollary which shows that $(H_k^{(n;\text{out})}, H_k^{(n;\text{in})})_{k=0}^m$ are close to the independent forward and backward branching processes.

Corollary 1.3.9. *Consider the directed configuration model* $G_n = DCM_n(\boldsymbol{d})$*, where the out- and in-degrees satisfy Assumption 1.3.2. Let* $Z_k^{(out)}$ *denote the number of vertices in the k'th generation of the marked forward branching process, and* $Z_k^{(in)}$ *similarly for the marked backward branching process. Then, for* $m > 0$ *arbitrary small*

$$(H_k^{(n;\text{out})}, H_k^{(n;\text{in})})_{k=0}^m \xrightarrow{d} (Z_k^{(\text{out})}, Z_k^{(\text{in})})_{k=0}^m. \tag{1.27}$$

1.4 Strongly connected component

Let us look at the *strongly connected component*, which is the set of vertices $u, v \in [n]$ such that a directed path exists from u to v and from v to u. Our goal is to construct a directed graph for which nearly all vertices are contained in the same strongly connected component, such that two typical vertices are whp connected. To find sufficient conditions, we will make use of a theorem that describes the dependence of the strongly connected components size on the survival probability of a forward and backward branching process [6] [13].

To start with, let denote $\theta^{(\text{out})}$ and $\theta^{(\text{in})}$ denote the survival probability of the branching processes with respective distribution $(f^{(\text{out})}(k))_{k\geq 0}$ and $(f^{(\text{in})}(k))_{k\geq 0}$. Note that these branching processes are the undelayed versions of the marked forward and backward branching processes, described in Section 1.3.3. Next, take

$$\zeta^{(\text{in})} = 1 - \sum_{k,l\geq 0} f(k,l)(1-\theta^{(\text{in})})^k, \quad \zeta^{(\text{out})} = 1 - \sum_{k,l\geq 0} f(k,l)(1-\theta^{(\text{out})})^l, \tag{1.28}$$

where $(f(k,l))_{k,l\geq 0}$ denotes the asymptotic degree distribution. Define a *backward cluster* of vertex v as all the vertices u with a path to v, and a *forward cluster* of v to be all vertices u for which there is a path from v to u. Then, we can interpret $\zeta^{(\text{out})}$ as the asymptotic probability that a typical vertex has a large forward cluster, and $\zeta^{(\text{in})}$ as the asymptotic probability a typical vertex has a large backward cluster. Further, we take

$$\gamma = \sum_{k,l\geq 0} f(k,l)(1-\theta^{(\text{in})})^k(1-\theta^{(\text{out})})^l, \tag{1.29}$$

which then represents the asymptotic probability that a typical vertex has both a large forward and backward cluster. It follows that $1-\gamma$ can be interpreted as the asymptotic probability an uniform vertex has either a large forward *or* large backward cluster. Hence,

$$\zeta = \zeta^{(\text{in})} + \zeta^{(\text{out})} - (1-\gamma), \tag{1.30}$$

represents the asymptotic probability a typical vertex has both a large forward *and* large backward cluster. For

$$\nu = \frac{\mathbb{E}[D^{(\text{in})}D^{(\text{out})}]}{\mathbb{E}[D^{(\text{out})}]}, \tag{1.31}$$

the following theorem describes when the directed configuration model is dominated by a unique strongly connected component [13].

Theorem 1.4.1. (Phase transition in $\text{DCM}_n(\boldsymbol{d})$) *Suppose that the in- and out- degrees of the directed configuration model satisfy Assumption 1.3.2. Let $\mathcal{C}_{max}$ and $\mathcal{C}_{(2)}$ denote the vertices in the respective largest and second largest strongly connected component.*

(a) *When $\nu > 1$, ζ in (1.30) satisfies $\zeta \in (0,1]$, and*

$$|\mathcal{C}_{\max}|/n \xrightarrow{\mathbb{P}} \zeta, \tag{1.32}$$

while $|\mathcal{C}_{(2)}|/n \xrightarrow{\mathbb{P}} 0$.

(b) *When $\nu < 1$, ζ in (1.30) satisfies $\zeta = 0$ and $|\mathcal{C}_{max}|/n \xrightarrow{\mathbb{P}} 0$.*

The theorem tells us that for $\nu < 1$ it is very unlikely a path exists between two uniformly chosen vertices. We are interested in the conditions for which $\nu > 1$, which brings us the following assumption.

Assumption 1.4.2. (Minimal out- and in-degree)

$$\min_{u\in[n]} d_u^{(\text{out})} \geq 1 \text{ and } \min_{u\in[n]} d_u^{(\text{in})} \geq 1. \tag{1.33}$$

Notice that if $\mathbb{P}(D^{(\text{in})} > 1, D^{(\text{out})} > 1) > 0$ that under Assumption 1.4.2 we clearly have $\nu > 1$. Moreover, we have that $\theta^{(\text{out})} = \theta^{(\text{in})} = 1$, as each individual in the branching processes have at least one offspring. It follows that $\zeta = 1$, such that Theorem 1.4.1 states that under Assumption 1.3.2 the directed configuration model contains a unique strongly connected component of size $n - o_{\mathbb{P}}(1)$.

1.5 Ultra-small phenomenon in the directed configuration model

Recall that in small world directed networks, the in- or out-degree follows a power law, which is formally defined as follows:

Definition 1.5.1. (Slowly varying functions and power-law distributions) *A function $x \mapsto L(x)$ is slowly varying at infinity when, for every $t > 0$,*

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1. \tag{1.34}$$

We say that X has a power-law distribution with exponential τ when there exists a function $x \mapsto L(x)$ that is slowly varying at infinity, such that

$$[1 - F_X](x) = \mathbb{P}(X > x) = L(x)x^{-(\tau-1)}. \tag{1.35}$$

If the function $L(x)$ is constant for all x, we say X has a pure power-law distribution.

To study the behaviour of the large directed graphs in the limiting regime, the specific slowly varying function of the degree distribution is not important. Thus, we shall use Potter's Theorem ([2], Theorem 1.5.6) to bound the slowly varying functions in the degree distribution to simplify calculations.

Theorem 1.5.2. (Potter's Theorem) *Let $x \mapsto L(x)$ be slowly varying at infinity. For every δ, there exists a constant $c = c(\delta)$, such that, for all $x \geq 1$*

$$x^{-\delta}/c \leq L(x) \leq cx^{\delta}. \tag{1.36}$$

Potter's Theorem implies that the tail of any power-law distribution can be bounded from both sides by a pure power-law distribution, with a slightly adapted power-law exponent. Remember that $F_n^{(\text{in})}(x)$ (resp., $F_n^{(\text{out})}(x)$) denotes the respective proportion of outbound (resp., inbound) half-edges attached to vertices with in-degree (resp., out-degree) at most x, with

$$F_n^{(\text{in})}(x) = \frac{1}{l_n}\sum_{u\in[n]} d_u^{(\text{out})}\mathbb{1}\{d_u^{(\text{in})} \leq x\}, \quad F_n^{(\text{out})}(x) = \frac{1}{l_n}\sum_{u\in[n]} d_u^{(\text{in})}\mathbb{1}\{d_u^{(\text{out})} \leq x\}. \tag{1.37}$$

For our directed configuration model, we impose these distribution functions satisfy the conditions of the folliwng assumption:

Assumption 1.5.3. (Power-law bounds) *For both $q \in \{out, in\}$ there exists $\tau^{(q)} \in (2,3)$, and for all $\delta^{(q)} > 0$, there exists $c_1^{(q)} = c_1^{(q)}(\delta^{(q)})$, $c_2^{(q)} = c_2^{(q)}(\delta^{(q)})$ such that, uniformly in n,*

$$c_1^{(q)} x^{-(\tau^{(q)}-2+\delta^{(q)})} \leq [1 - F_n^{(q)}](x) \leq c_2^{(q)} x^{-(\tau^{(q)}-2-\delta^{(q)})} \tag{1.38}$$

where the upper bound holds for all $x \geq 1$, while the lower bound is only required to hold for $1 \leq x \leq l_n^{\beta}$, for some $\beta \in (1/2, 1)$.

Note that the assumption states that $[1 - F_n^{(\text{in})}](x)$ and $[1 - F_n^{(\text{out})}](x)$ obey the power-law bounds shown in Potter's Theorem 1.5.2. This is a more general assumption than simply taking them to be power-law. Under Assumption 1.3.2 and 1.5.3 it can be shown that

$$\sum_{k\geq 0}\sum_{l\geq 0} klf_n(k,l) \geq l_n \sum_{x\geq 1} c_1^{(q)} x^{-(\tau^{(q)}-2+\delta^{(q)})} \xrightarrow{n\to\infty} \infty. \tag{1.39}$$

As $\mathbb{E}[D_n^{(\text{out})}] \to \mathbb{E}[D^{(\text{out})}]$ and $\mathbb{E}[D_n^{(\text{in})}] \to \mathbb{E}[D^{(\text{in})}]$, it follows that the covariance of the in- and out-degree of a typical vertex is infinite. The following is the first main result of this thesis, describing the ultra-small typical distances of the directed configuration model under the conditions that were described along this section so far.

Theorem 1.5.4. (Typical distance for infinite covariance directed configuration model) *Consider the directed configuration model* $\mathrm{DCM}_n(\boldsymbol{d})$ *with empirical in- and out-degree distribution satisfying Assumption 1.3.2, 1.4.2 and the size-biased in- and out-degrees satisfying Assumption 1.5.3, for some* $\tau^{(in)}, \tau^{(out)} \in (2.3)$*. Then, for the uniformly and independently chosen vertices* o_1 *and* o_2*, it follows that*

$$\frac{\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1, o_2)}{\log\log n} \xrightarrow{\mathbb{P}} \frac{1}{|\log(\tau^{(\mathrm{out})}-2)|} + \frac{1}{|\log(\tau^{(\mathrm{in})}-2)|}. \tag{1.40}$$

To intuitively understand this result, note that the vertices o_1 and o_2 are taken randomly and independent from each other. Hence, the distance between these vertices is smaller than some value k, if either: we pick the starting vertex o_1 first and then by chance pick o_2 from the set of vertices that can be reached in k or less steps from o_1, denoted by $B_k^{(G_n:\mathrm{out})}(o_1)$, or we pick the target vertex o_2 first and then by chance o_1 is picked from the set of vertices that can reach o_2 in k or less steps, denoted by $B_k^{(G_n:\mathrm{in})}(o_2)$. It follows that

$$\mathbb{P}(\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1, o_2) \le k) = \frac{\mathbb{E}[|B_k^{(G_n:\mathrm{out})}(o_1)|]}{n} = \frac{\mathbb{E}[|B_k^{(G_n:\mathrm{in})}(o_2)|]}{n}, \tag{1.41}$$

which seems to indicate that the k is related to the size of the forward and backward neighborhoods. Consequently, the equality (1.41) show that for the probability to be non-zero for large n, it is required that both $\mathbb{E}[|B_k^{(G_n:\mathrm{out})}(o_1)|]$ and $\mathbb{E}[|B_k^{(G_n:\mathrm{in})}(o_2)|]$ are both of order $\Theta(n)$. In Section 1.3.3 we have shown the branching process approximations $|B_k^{(G_n:\mathrm{out})}(o_1)| \approx Z_k^{(\mathrm{out})}$ and $|B_k^{(G_n:\mathrm{in})}(o_2)| \approx Z_k^{(\mathrm{in})}$ under the regularity conditions from Assumption 1.3.2, given that k is not too large. Under the power-law Assumptions 1.5.3 the offspring distribution for both of these branching processes have an infinite mean. For such branching processes it is known that they experience super-exponential generational growth [7], i.e. for some constant $C_q > 0$ and $q \in \{\mathrm{out}, \mathrm{in}\}$

$$Z_k^{(q)} \approx C_q^{\frac{1}{(\tau^{(q)}-2)^k}}. \tag{1.42}$$

The equality (1.41) suggest that we need to take $k_n^{(q)}$ such that $Z_{k_n^{(q)}}^{(q)} = \Theta(n)$, which is the case for $k_n^{(q)} \approx \log\log n / |\log(\tau^{(q)} - 2)|$. To keep the approximation accurate we use

$$\mathbb{P}(\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1, o_2) \le k_n^{(\mathrm{out})} + k_n^{(\mathrm{in})}) = \mathbb{P}(B_{k_n^{(\mathrm{out})}}^{(G_n:\mathrm{out})}(o_1) \cap B_{k_n^{(\mathrm{in})}}^{(G_n:\mathrm{in})}(o_2) \neq \emptyset), \tag{1.43}$$

and approximations both $|B_k^{(G_n:\mathrm{out})}(o_1)|$ and $|B_k^{(G_n:\mathrm{in})}(o_2)|$ simultaneously. This shows that under the additional Assumption 1.4.2 on the minimal in- and out-degree, such that typical vertices are likely in the same component, it follows that

$$\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1, o_2) \approx \frac{\log\log n}{|\log(\tau^{(\mathrm{out})}-2)|} + \frac{\log\log n}{|\log(\tau^{(\mathrm{in})}-2)|}. \tag{1.44}$$

Remember that the typical paths make use of the highest order out- and in-degree vertices. Here, $\frac{\log\log n}{|\log(\tau^{(\mathrm{out})}-2)|}$ is the contribution of the part of the path with rapidly increasing out-degree vertices towards the outbound hub and $\frac{\log\log n}{|\log(\tau^{(\mathrm{in})}-2)|}$ the contribution of the part with rapidly decreasing in-degree vertices from the inbound hub to the target vertex.

1.6 Fluctuations of the typical distances

We now study how the typical distances fluctuate with the size of the directed configuration model and add a possible truncation to the power-law distributions. The choice of these assumptions are taken to recreate the conditions of [15], in which these properties are studied in the configuration model. To be able to do this we will need to know how the in- and out-degree distribution behaves with changes to the graph size n. Hence, for the power-law assumption we include the slowly varying function in the defined pure power-law.

Assumption 1.6.1. *For both* $q \in \{out, in\}$ *define* $\gamma^{(q)}(x) = C^{(q)}(\log x)^{\gamma^{(q)}-1}$ *for some* $\gamma^{(q)} \in [0,1)$ *and* $C^{(q)} > 0$ *and take* $\tau^{(q)} \in (2,3)$. *Then there exists* $\beta_n^{(q)}$ *such that for all* $\varepsilon > 0$, $F_n^{(q)} = 1$ *for* $x \geq l_n^{\beta_n^{(q)}(1+\varepsilon)}$, *and*

$$x^{-(\tau^{(q)}-2)-\gamma^{(q)}(x)} \leq [1 - F_n^{(q)}](x) \leq x^{-(\tau^{(q)}-2)+\gamma^{(q)}(x)}, \quad \forall x \leq l_n^{\beta_n^{(q)}(1-\varepsilon)}, \tag{1.45}$$

given that

$$\beta_n^{(\text{in})} + \beta_n^{(\text{out})}(\tau^{(\text{out})} - 2) = \beta_n^{(\text{out})} + \beta_n^{(\text{in})}(\tau^{(\text{in})} - 2) \in (0,1). \tag{1.46}$$

Note that the equality (1.46) is the same as stating that $\beta_n^{(\text{in})}(3 - \tau^{(\text{in})}) = \beta_n^{(\text{out})}(3 - \tau^{(\text{out})})$. This equality condition arises naturally as each outbound half-edge is paired with an inbound half-edge. This ensures that the heavier tailed distribution is truncated earlier, such that the total number of out- and inbound half-edges remains in proportion. Next, we want to be able to relate the behaviour of $F_n, F_n^{(\text{out})}$ and $F_n^{(\text{in})}$ for different values of n to their limiting distributions. Let

$$d_{TV}(F,G) = \frac{1}{2}\sum_{x \in \mathbb{N}} |F(x+1) - F(x) - (G(x+1) - G(x))|, \tag{1.47}$$

denote the total variation distance between two discrete probability measures. The weakest form of assumption we can state to show this relation is of the following form.

Assumption 1.6.2. *Assume there exist distribution functions* $F(x), F^{(out)}$ *and* $F^{(in)}$ *such that* $F_n \to F$, $F_n^{(out)} \to F^{(out)}$ *and* $F_n^{(in)} \to F^{(in)}$ *in all continuity points of* $F(x), F^{(out)}$ *and* $F^{(in)}$. *Moreover, there exists some* $\kappa > 0$, *such that for* $\beta_n = \max\{\beta_n^{(out)}, \beta_n^{(in)}\}$

$$\max\{d_{TV}(F_n^{(\text{out})}, F^{(\text{out})}), d_{TV}(F_n^{(\text{in})}, F^{(\text{in})}), d_{TV}(F_n, F)\} \leq n^{-\beta_n \kappa}. \tag{1.48}$$

As the total variance convergence equals weak convergence for discrete random variables, $D_n \to D$, $D_n^{*(\text{out})} \to D^{*(\text{out})}$ and $D_n^{*(\text{in})} \to D^{*(\text{in})}$ under Assumption 1.6.2. Note that the limiting random variables $F, F^{(\text{out})}$ and $F^{(\text{in})}$ are not truncated. It follows that under Assumption 1.6.1 the bound (1.48) is the best possible we can take, as for both $q \in \{\text{out}, \text{in}\}$

$$d_{TV}(F_n^{(q)}, F^{(q)}) \geq \mathbb{P}(D^{*(q)} > l_n^{\beta_n^{(q)}}) \geq n^{-\beta_n(\tau^{(q)}-2-\delta)}. \tag{1.49}$$

The second main result uses a random variable, that describes the super exponential growth rates of both the branching processes $(Z_k^{(\text{out})})_{k\geq 0}$ and $(Z_k^{(\text{in})})_{k\geq 0}$, which are coupled with the respective forward and backward explorations. So before we can state the result, we define the following:

Definition 1.6.3. *Let* $Z_k^{(out)}$ *and* $Z_k^{(in)}$ *denote the size of the k'th generation of a marked forward branching process and marked backward branching process, described in Section 1.3.3. Then, for* $\tau = \max\{\tau^{(out)}, \tau^{(in)}\}$ *and some*

$$\delta' < (\tau - 2)\min\{n^{-\beta_n^{(\text{in})}(\kappa-\delta)}, n^{-\beta_n^{(\text{out})}(\kappa-\delta)}, n^{(1-\beta_n^{(\text{out})}(1+\varepsilon)-\delta)/2}, n^{(1-\beta_n^{(\text{in})}(1+\varepsilon)-\delta)/2}\}, \tag{1.50}$$

define

$$Y_n^{(\text{out})} = (\tau^{(\text{out})} - 2)^{t(n^{\delta'})} Z_{t(n^{\delta'})}^{(\text{out})} \quad Y_n^{(\text{in})} = (\tau^{(\text{in})} - 2)^{t(n^{\delta'})} Z_{t(n^{\delta'})}^{(\text{in})}, \tag{1.51}$$

where $t(n^{\delta'}) = \inf_k\{\max\{Z_k^{(out)}, Z_k^{(in)}\} \geq n^{\delta'}\}$. *Further define*

$$Y^{(\text{out})} = \lim_{k\to\infty} (\tau^{(\text{out})} - 2)^k Z_k^{(\text{out})}, \quad Y^{(\text{in})} = \lim_{k\to\infty} (\tau^{(\text{in})} - 2)^k Z_k^{(\text{in})}. \tag{1.52}$$

Note that the pair $(Y_n^{(\text{out})}, Y_n^{(\text{in})})$ is a subsequence of the convergent sequence $((\tau^{(\text{out})} - 2)^k Z_k^{(\text{out})}, (\tau^{(\text{in})} - 2)^k Z_k^{(\text{in})})$. As for any $\delta' > 0$ $t(n^{\delta'}) \to \infty$ with $n \to \infty$, it follows that under Assumption 1.6.2 $(Y_n^{(\text{out})}, Y_n^{(\text{in})}) \xrightarrow{d} (Y^{(\text{out})}, Y^{(\text{in})})$. For ease of notation, we make numerous use of the following definition:

Definition 1.6.4. ($\sim$ notation) *We use the short-hand notation $X_n \sim a_n$ to indicate the property*

$$X_n \sim a_n \Leftrightarrow \mathbb{P}\left(X_n \in [a_n e^{-\log(a_n)^\theta}, a_n e^{\log(a_n)^\theta}]\right), \tag{1.53}$$

for some $\theta \in (0,1)$.

Note that $X_n \sim n^a$ is a stronger statement than $X_n = n^{a(1+o_{\mathbb{P}}(1))}$. We will call vertices with out-degree $\sim l_n^{\beta_n^{(\text{out})}(\tau^{(\text{out})}-2)}$ outbound hubs and vertices with in-degree $\sim l_n^{\beta_n^{(\text{in})}(\tau^{(\text{in})}-2)}$ inbound hubs. To state the main result shortly, define

$$\begin{aligned} T_{(\text{out})} = T_{(\text{out})}(\beta_n^{(\text{out})}) &= -1 + \left\lfloor \frac{\log\log(l_n^{\beta_n^{(\text{out})}}) - \log(Y_n^{(\text{out})})}{|\log(\tau^{(\text{out})}-2)|} \right\rfloor \\ T_{(\text{in})} = T_{(\text{in})}(\beta_n^{(\text{in})}) &= -1 + \left\lfloor \frac{\log\log(l_n^{\beta_n^{(\text{in})}}) - \log(Y_n^{(\text{in})})}{|\log(\tau^{(\text{in})}-2)|} \right\rfloor \end{aligned} \tag{1.54}$$

and

$$b_n^{(\text{out})} = \left\{ \frac{\log\log(l_n^{\beta_n^{(\text{out})}}) - \log(Y_n^{(\text{out})})}{|\log(\tau^{(\text{out})}-2)|} \right\}, b_n^{(\text{in})} = \left\{ \frac{\log\log(l_n^{\beta_n^{(\text{in})}}) - \log(Y_n^{(\text{in})})}{|\log(\tau^{(\text{in})}-2)|} \right\}, \tag{1.55}$$

where $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part of x. Thing brings us to the main result, describing how the typical distances fluctuate with the size of the graph and the truncation values.

Theorem 1.6.5. (Distance in truncated power-law directed configuration models) *Consider the directed configuration model $\mathrm{DCM}_n(\boldsymbol{d})$ with empirical degree distribution satisfying Assumption 1.4.2,1.6.1 and 1.6.2 and for both $q \in \{out, in\}$, $\beta_n^{(q)}(\log n)^\eta \to \infty$ for some $\eta \in (0,1)$. Then, for two uniformly chosen vertices $o_1, o_2 \in [n]$*

$$\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1,o_2) = T_{(\text{out})} + T_{(\text{in})} + \left\lceil \frac{1 - \beta_n^{(\text{out})}(\tau^{(\text{out})}-2)^{b_n^{(\text{out})}} - \beta_n^{(\text{in})}(\tau^{(\text{in})}-2)^{b_n^{(\text{in})}}}{\frac{1}{2}(\beta_n^{(\text{out})}(3-\tau^{(\text{out})}) + \beta_n^{(\text{in})}(3-\tau^{(\text{in})}))} \right\rceil + 1. \tag{1.56}$$

Note that the condition that for both $q \in \{\text{out}, \text{in}\}$ $\beta_n^{(q)}(\log n)^\eta \to \infty$ is slightly stronger than the condition that empirical second moment of D_n and thus, the first moment of the forward out- and in-degrees $D_n^{*(\text{out})}$ and $D_n^{*(\text{in})}$ are infinite. This is can be seen in (3.106), where it is shown that $v_n^{(q)} = \mathbb{E}[D_n^{*(q)}] \sim l_n^{\beta_n^{(q)}(3-\tau^{(q)}}$, as in definition (1.53). This value tends to infinity if $\beta_n^{(q)} \log n \to \infty$. Note that

$$\begin{aligned} \sum_{k\geq 0}\sum_{l\geq 0} kl f_n(k,l) &= \sum_{k\geq 0} k \sum_{l\geq 0} l \frac{1}{n} \sum_{u\in[n]} \mathbb{1}\{d_u = (k,l)\} \\ &= \sum_{k\geq 0} k \frac{1}{n} \sum_{u\in[n]} d_u^{(\text{out})} \mathbb{1}\{d_u^{(\text{in})} = k\} \\ &= \frac{l_n}{n} \sum_{k\geq 0} k f_n^{(\text{in})}(k) = \mathbb{E}[D_n^{(\text{in})}]\mathbb{E}[D_n^{*(\text{in})}]. \end{aligned} \tag{1.57}$$

So also the covariance of the in- and out-degree of a typical vertex, which equals

$$\mathbb{E}[D_n^{(q)}]\mathbb{E}[D_n^{*(q)}] - \mathbb{E}[D_n^{(q)}]^2, \tag{1.58}$$

tends to infinity under this condition. As the expectation of the empirical first moment of the degrees is finite, this shows the same dependence on the value of $v_n^{(q)}$.

Let us now compare this result with the typical distances in the non-truncated model. For the truncated model, the highest out-degree vertices with degree $l_n^{\beta_n^{(\text{out})}}$ in the outbound-hub and the highest in-degree vertices with degree $l_n^{\beta_n^{(\text{in})}}$ in the inbound-hub. The approximation from (1.42) suggests that a typical vertex has a path of length $\log\log(l_n^{\beta_n^{(\text{out})}})/(|\log(\tau^{(\text{out})}-2)|)+$ tight number of steps to the outbound hub, and can be reached from the inbound hub by a path of length $\log\log(l_n^{\beta_n^{(\text{in})}})/(|\log(\tau^{(\text{in})}-2)|)+$ tight number of steps. In contrary to the non-truncated model, the vertices in the outbound hub are not necessarily likely connected to the vertices in the inbound hub. For the distance from the out- to the inbound hub, we use the approximation $Z_k^{(n;q)} \approx \mathbb{E}[Z_k^{(n;q)}] = \mathbb{E}[D_n^{(q)}](\nu_n^{(q)})^{k-1}$. Note that as the main contribution of $\nu_n^{(q)}$ comes from the degrees of the vertices in the hubs, this approximation is only remotely accurate for the neighborhood of these particular vertices. As $v_n^{(q)} \sim l_n^{\beta_n^{(q)}(3-\tau^{(q)}}$, this shows that an additional $\log n/\log(\nu_n^{(q)}) = \frac{1}{\beta_n^{(q)}(3-\tau^{(q)})}+$ tight number of steps are needed to get from the outbound hub to the inbound hub, where under the condition (1.46) from Assumption 1.6.1

$$\frac{1}{\beta_n^{(q)}(3-\tau^{(q)})} = \frac{2}{\beta_n^{(\text{out})}(3-\tau^{(\text{out})})+\beta_n^{(\text{in})}(3-\tau^{(\text{in})})}. \tag{1.59}$$

This explains how these typical distances arise.

Chapter 2

Ultra-small world phenomenon on the directed configuration model

This section we will prove the ultra-small world phenomenon on the strongly connected directed configuration model with infinite covariance between the in- and out-degrees. The proof uses many arguments from ([14], Section 7), which shows the conditions under which the configuration model experiences the ultra-small world phenomenon. As the construction process of the directed and the undirected configuration model are very similar, it is natural to assume that both models can be approached similarly when studying their properties. For this reason, we start Section 2.1 with a summary of the conditions used in ([14], Section 7), which lead to the ultra-small phenomenon in of the configuration model (Remco vd Hofstad, [14]). The precise proofs are omitted, but we will discuss the importance of each of the conditions, by explaining their effect on the behaviour of the graph. In addition, we introduce the methods how these effects are used to prove the properties. In Section 2.2 the local tree-like structure of the directed configuration model is proven, where we first show the coupling of each type of breadth-first exploration process with a properly defined branching process in Section 2.2.1, which we shall use to prove the local convergence results in Section 2.2.2. In Section 2.3 we show an upper bound on the typical distances shown in Theorem 1.5.4 by proving the typical vertices are connected by a path passing the out- and inbound hub. Finally, we apply path-counting techniques in Section 2.4 to bound the path probabilities to obtain a matching lower bound for Theorem 1.5.4.

2.1 Inspiration from the configuration model

The properties of the configuration model have been extensively studied ever since it was introduced by B. Bollobas [3]. This includes the conditions under which the model experiences double logarithmic typical distances. It feels natural to study the effects of these conditions as a starting point and see where their desired results fail in the directed setting, which will give a good insight how the conditions need to be adapted. This section we will study the conditions under which the ultra-small phenomenon on the configuration model is proven in ([14], Section 7). The precise proofs of most results is omitted, but how the given conditions lead to the result is discussed in detail.

Important remark: Section 2.1 is a stand alone part of the thesis. Instead of using new notation to indicate each object of the graph, a lot of the notation used for the directed configuration model is reused. This decision is made in order to restrict the number of different symbols used along this thesis, which could otherwise confuse the reader. Hence, the notation used this section, should not be confused with their meaning in the remainder of the thesis.

2.1.1 Model construction and branching process approximation

Start by fixing a sequence $\boldsymbol{d} = (d_1, ..., d_n)$ of non-negative integers, for which d_u denotes the number of half-edges assigned to vertex $u \in [n]$, under the condition that

$$l_n = \sum_{u \in [n]} d_u \tag{2.1}$$

is an even number. The edges of the graph are constructed from pairs of half-edges, which are randomly taken without replacement. This process is repeated such that at step $k+1$, an arbitrary half-edge is chosen from the $l_n - 2k$ remaining, and randomly paired to one of the other $l_n - 2k - 1$ remaining free half-edges. After all the half-edges have been paired, the configuration model with degrees $\boldsymbol{d}$ is constructed, denoted as $\mathrm{CM}_n(\boldsymbol{d})$. A typical vertex in the configuration model has degree distribution D_n, for which

$$\mathbb{P}(D_n = k) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}\{d_u = k\}, \tag{2.2}$$

denotes the proportion of vertices with degree k. The cumulative degree distribution is written as

$$F_n(x) = \frac{1}{n} \sum_{u \in [n]} \mathbb{1}\{d_u \leq x\}. \tag{2.3}$$

The graph distance between vertices u and v equals k, denoted by the event $\{\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(u, v) = k\}$, when this is the least amount of edges that need to be traversed to get from u to v. Note that as the edges are not directed it follows that $\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(u, v) = \mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(v, u)$. The typical distance of the graph is the distance between two uniformly and independently chosen vertices o_1 and o_2.

Exploration process: The similar construction of the undirected configuration model, allows it to be constructed in a breadth-first manner, similarly to the exploration process described in Section 1.3. Though note that as $\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(u, v) = \mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(v, u)$, there is only one type of exploration and thus, one type of neighborhood to be explored. Taking $G_n = \mathrm{CM}_n(\boldsymbol{d})$, the r-neighborhood of a typical vertex o is denoted by $B_r^{G_n}(o)$, such that

$$\begin{aligned} V(B_r^{G_n}(o)) &= \{u \in V(G_n) : \mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(o, u) \leq r\}, \\ E(B_r^{G_n}(o)) &= \{\{u, v\} \in E(G_n) : \mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(o, u) \leq r, \mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(o, v) \leq r\}. \end{aligned} \tag{2.4}$$

For fixed values of r, the size of the this neighborhood stays bounded as $n \to \infty$ under the following regularity conditions:

Assumption 2.1.1. (Regularity conditions)

(a) **Weak convergence of vertex degrees.** There exists a distribution function F, such that

$$D_n \xrightarrow{d} D, \tag{2.5}$$

where D_n and D have cumulative distribution F_n and F, respectively.

(b) **Convergence of average vertex degrees**. There exists a distribution function F, such that

$$\lim_{n \to \infty} \mathbb{E}[D_n] = \mathbb{E}[D], \tag{2.6}$$

where D_n and D have cumulative distribution functions, F_n and F, respectively.

The topology on which the local convergence results are shown is described in detail in ([14], Section 2). It is very similar to the topology described for the marked graphs in Section 1.3.2.

The main difference is that there is no need to mark the edges as there is only one type of half-edge. To understand the local convergence results that will be introduced shortly, note that the asymptotic degree has distribution D, where $\mathbb{P}(D = k) = \lim_{n\to\infty} \mathbb{P}(D_n = k)$. Under Assumption 2.1.1 it follows that $l_n/n = \mathbb{E}[D_n] \to \mathbb{E}[D] > 0$, such that the total number of half-edges is very large for large values of n. So after a bounded number of exploration steps the depletion of the number of available half-edges is barely noticeable. Ignoring the fact that some half-edges are attached to the root, the probability a half-edge of the root pairs with a half-edge attached to a vertex with degree k equals $\frac{k}{\mathbb{E}[D_n]} p_n(k)$. As one half-edge has been used to connect with the root, the vertex needs to have degree $(k + 1)$ to have k available half-edges from which the exploration process can expand. So the offspring distribution D_n^*, also called the forward degree, equals

$$p_n^*(k) = \mathbb{P}(D_n^* = k) = \frac{(k+1)}{\mathbb{E}[D_n]} \mathbb{P}(D_n = k + 1), \tag{2.7}$$

for which the distribution function can be written as

$$F_n^* = \mathbb{P}(D_n^* \leq k) = \frac{1}{l_n} \sum_{u\in[n]} d_u \mathbb{1}\{d_u \leq k + 1\}. \tag{2.8}$$

Under Assumption 2.1.1 it follows that

$$\lim_{n\to\infty} \mathbb{P}(D_n^* = k) = \mathbb{P}(D^* = k) = \frac{(k+1)}{\mathbb{E}[D]} \mathbb{P}(D = k + 1). \tag{2.9}$$

The precise result of the local convergence in the configuration model that is proven in ([14], Theorem 4.1) is the following:

Theorem 2.1.2. (Local tree-like nature configuration model) *Consider the configuration model* $\mathrm{CM}_n(\boldsymbol{d})$ *for which Assumption 2.1.1 is satisfied. Then* $\mathrm{CM}_n(\boldsymbol{d})$ *converges locally in probability to the branching process with root distribution* D *and offspring distribution* D^* *for the following generations.*

It can be convenient to view the breadth-first explored neighborhood as a singe super vertex, for which the free half-edges denote the degree of this super vertex. Take the uniformly chosen vertices o_1 and o_2, $r \geq 1$ and $i \in \{1, 2\}$. Let $H_r^{(n;i)}$ denote the number of unpaired half-edges incident to vertices that are at distance $r - 1$ from uniformly chosen vertex o_i, where we take $H_0^{(i)} = 1$. So by definition we have $H_1^{(n;i)} = d_{o_i}$, and $H_2^{(n;i)}$ is obtained by pairing the d_{o_i} half-edges and counting the number of unpaired half-edges at distance 1 from o_i. The following Corollary shows that the process $(H_l^{(n;1)}, H_l^{(n;2)})_{l=0}^r$ is close to two independent branching processes with root distribution D, and offspring distribution D^* for the following generations

Corollary 2.1.3. (*Approximating neighborhoods with independent branching processes*) Let the degrees $\boldsymbol{d}$ satisfy Assumption 2.1.1, and let $(Z_l^{(1)}, Z_l^{(2)})_{l=0}^r$ denote two delayed branching processes with root distribution D and offspring distribution D^* for the following generations. Then for every $r \geq 1$,

$$(H_l^{(n;1)}, H_l^{(n;2)})_{l=0}^r \xrightarrow{d} (Z_l^{(1)}, Z_l^{(2)})_{l=0}^r. \tag{2.10}$$

2.1.2 Ultra-small phenomenon on the configuration model

To ensure the degree distribution has an infinite variance, we introduce the following assumption:

Assumption 2.1.4. *There exists a* $\tau \in (2, 3)$ *and for all* $\delta > 0$*, there exist* $c_1 = c_1(\delta)$ *and* $c_2 = c_2(\delta)$*, such that, uniformly in* n*,*

$$c_1 x^{-(\tau-1+\delta)} \leq [1 - F_n](x) \leq c_2 x^{-(\tau-1-\delta)}, \tag{2.11}$$

where the upper bound holds for every $x \geq 1$ *and the lower bound is only required to hold for* $1 \leq x \leq n^\beta$*, for some* $\beta \geq 1/2$*.*

Note that $[1 - F_n](x) > 0$ implies that $[1 - F_n](x) \geq \frac{1}{n}$, which means at least one vertex has a degree of x or larger. Hence, if $x >> n^{1/(\tau-1)}$ the lower and upper bound will contradict each other. For this reason the lower bound can only hold for $x = O(n^{1/(\tau-1)})$. As for $\tau \in (2,3)$ we have $1/(\tau - 1) \in (1/2, 1)$, it is only required that the lower bound holds for $x \leq n^{\beta}$, for *some* $\beta \in (1/2, 1)$. Finally, we introduce the following condition on the minimal degrees.

Assumption 2.1.5. (Minimal degree)

$$\min_{u \in [n]} d_u \geq 2. \tag{2.12}$$

Under Assumption 2.1.4 and Assumption 2.1.5 the configuration model almost surely has a unique connected component of size $n(1 - o_{\mathbb{P}}(1))$ [11]. The main result describing the typical distances in the configuration under the introduced conditions is shown in the following theorem ([14], Theorem 7.2):

Theorem 2.1.6. *Consider the configuration model for which the empirical degree distribution satisfies Assumption 2.1.1, 2.1.4 and 2.1.5. Then, for* $\tau \in (2,3)$ *it holds that*

$$\frac{\text{dist}_{\text{CM}_n(\boldsymbol{d})}(o_1, o_2)}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau - 2)|}. \tag{2.13}$$

In the following section we will describe how these conditions lead to these double logarithmic typical distances. In Section 2.1.3 how the ultra-small typical distances results from the condition, and describe the characteristics of these typical paths. Next, we introduce path-counting techniques in Section 2.1.4, and show they they can be applied to show it is unlikely an even shorter path exists between typical vertices.

2.1.3 Upper bound typical distances configuration model

Here we will summarise the results of ([14], Section 7.3.4), providing an upper bound on the typical distances. First, we introduce the following lemma which plays an important role is describing the connectivity between vertices in the configuration model:

Lemma 2.1.7. (Connectivity set in $CM_n(\boldsymbol{d})$) *For any two sets of vertices* $A, B \subseteq [n]$,

$$\mathbb{P}(A \text{ not directly connected to } B) \leq e^{-d_A d_B/(2l_n)}, \tag{2.14}$$

where, for any $A \subseteq [n]$,

$$d_A = \sum_{u \in A} d_u \tag{2.15}$$

denotes the total degree of vertices in A.

Proof. There are a total of d_A half-edges incident to vertices of set A, which will be paired to d_A different half-edges. Assume that the first k half-edges are paired with half-edges not incident to vertices of set B. The probability that the next half-edge is also not paired with any half-edges incident to set B equals

$$1 - \frac{1}{l_n - 2k + 1} \leq 1 - \frac{1}{l_n}. \tag{2.16}$$

Note that half-edges incident to vertices in set A can be paired to other half-edges incident to vertices in set A, such that two half-edges are used during the pairing. This means that we have to do at least $d_A/2$ pairings, such that we obtain the upper bound

$$\mathbb{P}(A \text{ not directly connected to } B) \leq \left(1 - \frac{1}{l_n}\right)^{d_A/2} \leq e^{-d_A d_B/(2l_n)}, \tag{2.17}$$

where we use that $1 - x \leq e^{-x}$. ∎

Lemma 2.1.3 shows that the probability a vertex with degree d is not connected to a vertex with degree greater than y is bounded by

$$e^{-d\sum_{u\in[n]} d_u \mathbb{1}\{d_u>y\}/(2l_n)} = e^{-d[1-F_n^*](y)/2}, \tag{2.18}$$

for $F_n^*(x)$ defined in (2.8). Under power-law conditions from Assumption 2.1.4 it follows that $[1-F_n^*](y)$ is close to $y^{-(\tau-2)}$. So the probability that a vertex with degree d is not connected to a vertex with degree greater than y is at most $e^{-dy^{-(\tau-2)}/2}$. For large values of d, this probability becomes very small when $y < d^{\frac{1}{(\tau-2+\varepsilon)}}$ for arbitrary small $\varepsilon > 0$. This shows that a vertex with a large degree d is whp connected to a vertex with approximate degree $d^{\frac{1}{(\tau-2)}}$. Moreover, the highest order degree vertices in the hub have degree n^{β}, where $\beta \in (\frac{1}{2}, 1)$. So the probability they are not connected is at most $e^{n^{2\beta}/2l_n}$, where $n^{2\beta}/l_n \to \infty$ as $n \to \infty$. This shows that whp the vertices in the hub form a clique.

So far we have discussed the connectivity of high-degree vertices, though typical vertices are not likely to have such a large degree. This is where the neighborhood approximation from Theorem 2.1.3 with the pair of branching processes with generation sizes $(Z_l^{(1)}, Z_l^{(2)})_{l\geq 0}$ is applied. It can be shown that under Assumption 2.1.4 there exists $c_1^* = c_1^*(\delta)$ and $c_2^* = c_2^*(\delta)$, such that

$$c_1^* x^{-(\tau-2+\delta)} \leq [1-F_n^*](x) \leq c_2^* x^{-(\tau-2-\delta)}, \tag{2.19}$$

where F_n^* is the offspring distribution function of the branching process, defined in (2.8). This shows the offspring distribution has an infinite mean, for which we have mentioned before that they experience super-exponential generational growth[7]. With the approximation from Theorem 2.1.3 we view the respective neighborhoods $B_k^{G_n}(o_1)$ and $B_k^{G_n}(o_2)$ as two super vertices with degree $Z_k^{(1)}$ and $Z_k^{(2)}$, which both grow super-exponential with k. By using these two super vertices as a start and end of the path, Lemma 2.1.7 shows the likely existence of a path connecting o_1 and o_2 for which the degrees grow larger with every vertex we move away from o_1 or o_2. In other words, starting at o_1 the degrees grows super-exponential every step we move forward till we reach the hub containing the highest order degree vertices, after which the degrees will shrink in the same order till we reach the target vertex o_2. This explains how the double logarithmic distances from Theorem 2.1.6 arise.

2.1.4 Lower Bound typical distances configuration model

Here we will summarise the proof given in ([14], Theorem 7.8), of the double logarithmic lower bound of the typical path length. The proof makes use of path-counting techniques to obtain an upper bound on the expected number of a certain type of path in the graph. By applying Markov's inequality, this is used to bound the probability that such a path exists.

First, we need a precise definition of a path. Let $\mathcal{P}_k(u, v)$ denote the set of paths of length k from vertex u to vertex v. The elements $\pi \in \mathcal{P}_k(u, v)$ of the set are the paths

$$\pi = ((\pi_0, t_0), (\pi_1, s_1, t_1), ..., (\pi_{k-1}, s_{k-1}, t_{k-1}), (\pi_k, s_k)), \tag{2.20}$$

where π_i denotes the index of the i'th vertex in the path, and t_i is the half-edge attached to vertex π_i, which is paired to the half-edge s_{i+1} which is attached to the following vertex in the path π_{i+1}, for all $i = 0, ..., k-1$ (note $\pi_0 = u$ and $\pi_k = v$ in this case). When two half-edges are paired during the construction process, we say that the edge they formed is *occupied*. Such that a path is occupied if all of its edges are occupied. For any $\pi \in \mathcal{P}_k(u, v)$, the probability this is the case equals

$$\mathbb{P}(\pi \text{ occupied}) = \prod_{i=1}^{k} \frac{1}{l_n - 2i + 1}. \tag{2.21}$$

For the fixed sequence of vertices $\pi_0, \pi_1, ..., \pi_k$, there are d_{π_0} ways to leave the first vertex π_0, d_{π_k} ways to enter the last vertex π_k, and $d_{\pi_i}(d_{\pi_i} - 1)$ ways to travel though vertex $i \in [1, k-1]$ along the path. Hence, the number of ways to create paths with this sequence of fixed vertices equals

$$d_{\pi_0}\left(\prod_{i=1}^{k-1} d_{\pi_i}(d_{\pi_i}-1)\right) d_{\pi_k}. \tag{2.22}$$

Now let us see how we can use these expression to bound the path probabilities. First note that

$$\begin{aligned}\mathbb{P}(\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(o_1,o_2)\le k_n) &= \frac{1}{n^2}\sum_{u,v\in[n]}\mathbb{P}(\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(u,v)\le k_n)\\ &= \frac{1}{n}+\frac{1}{n^2}\sum_{\substack{u,v\in[n]\\ u\neq v}}\sum_{k=1}^{k_n}\mathbb{P}(\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(u,v)=k),\end{aligned} \tag{2.23}$$

where the second equality follows as for all $u\in[n]$ it holds that $\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(u,u)=1$. Note that $\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(u,v)\le k_n$ if there exists an occupied path $\pi\in P_k(u,v)$ with $k\le k_n$. Hence, by taking $\nu_n=\sum_{u\in[n]}\frac{d_u(d_u-1)}{l_n}$ we obtain the bounds

$$\begin{aligned}\mathbb{P}(\mathrm{dist}_{\mathrm{CM}_n(\boldsymbol{d})}(u,v)=k) &\le \mathbb{P}(\exists\pi\in\mathcal{P}_k(u,v):\pi\text{ occupied})\\ &\le \sum_{\pi\in\mathcal{P}_k(u,v)}\mathbb{P}(\pi\text{ occupied})\\ &\le \frac{d_ud_v}{l_n-2k+1}\sum_{\pi_1,\dots,\pi_{k-1}}^{*}\prod_{i=1}^{k-1}\frac{d_{\pi_i}(d_{\pi_i}-1)}{l_n-2i+1}\\ &\le \frac{d_ud_v}{l_n}\frac{(l_n-2k-1)!!}{(l_n-1)!!}\nu_n^{k-1},\end{aligned} \tag{2.24}$$

where $\sum^*$ indicates the sum over distinct vertices. Here, the last inequality follows by including paths without the restriction that it uses unique vertices. The problem with this upper bound is that under the power-law Assumption 2.1.4, the value of ν_n is dominated by the high degree vertices, and thus too large to provide any useful bound. The reason this upper bound is so large, is because it does not consider the fact that it is unlikely that a typical vertex is connected with a vertex with a high degree. This can be seen, as the probability that a vertex with degree d is connected with a vertex with degree at least y is at most

$$\frac{1}{l_n}\sum_{u\in[n]} d\cdot d_u\mathbb{1}\{d_u>y)\} = d[1-F_n^*](y), \tag{2.25}$$

which is small for large values of y.

This issue is resolved by adding a suitable truncation argument on the degrees of the vertices along the paths. This will split the set of paths in so called "good paths" containing vertices with relatively small degree and "bad paths" containing high degree vertices. As we expect to pass vertices with increasingly large degrees the further we move away from a vertex, the truncation values along the path are indicated by an increasing sequence of positive numbers $(b_l)_{l=0}^{\infty}$. We say a path $\pi\in\mathcal{P}_k(u,v)$ is *good* when $d_{\pi_l}\le b_l\wedge b_{k-l}$ for $l=0,...,k$, and a *bad* path otherwise. Denote the set of these good paths by

$$\mathcal{GP}_k(u,v)=\{\pi\in\mathcal{P}_k(u,v): d_{\pi_l}\le b_l\wedge b_{k-l},\quad l=0,...,k\}, \tag{2.26}$$

where

$$\mathcal{E}_k(u,v)=\{\exists\pi\in\mathcal{GP}_k(u,v):\pi\text{ occupied}\}, \tag{2.27}$$

denotes the event that such a good path exists. Let $\mathcal{P}_k(u)=\bigcup_{v\in[n]}\mathcal{P}_k(u,v)$ be the set of all paths of length k starting from vertex u, and denote the subset of bad paths as

$$\mathcal{BP}_k(u)=\{\pi\in\mathcal{P}_k(u): d_{\pi_k}>b_k, d_{\pi_s}\le b_s,\quad \forall s<k\}. \tag{2.28}$$

Furthermore, denote the event that a bad path $\pi \in \mathcal{P}_k(u)$ is occupied as

$$\mathcal{F}_k(u) = \{\exists \pi \in \mathcal{BP}_k(u) : \pi \text{ occupied}\}. \tag{2.29}$$

Note that if $\text{dist}_{\text{CM}_n(\boldsymbol{d})}(u, v) \leq k_n$, a path of length k from vertex u to vertex v is occupied with $k \leq k_n$. This path can either be a good path $\pi \in \mathcal{GP}_k(u, v)$, or there exists a $l \leq \lceil k/2 \rceil$ such that $d_{\pi_s} \leq b_s$ for all $s < l$ and $d_{\pi_l} > b_l$, or it is the case that $d_{\pi_{k-s}} \leq b_{k-s}$ for all $s < l$ and $d_{\pi_{k-l}} > b_{k-l}$. This yields the bound

$$\{\text{dist}_{\text{CM}_n(\boldsymbol{d})}(u, v) \leq k_n\} \subseteq \bigcup_{k \leq k_n} \left(\mathcal{F}_k(u) \cup \mathcal{F}_k(v) \cup \mathcal{E}_k(u, v)\right). \tag{2.30}$$

From (2.23) and Boole's inequality it follows that

$$\mathbb{P}(\text{dist}_{\text{CM}_n(\boldsymbol{d})}(u, v) \leq k_n) \leq \frac{1}{n} + \sum_{k=1}^{k_n} \left[\frac{2}{n} \sum_{u \in [n]} \mathbb{P}(\mathcal{F}_k(u)) + \frac{1}{n^2} \sum_{\substack{u, v \in [n] \\ u \neq v}} \mathbb{P}(\mathcal{E}_k(u, v)) \right]. \tag{2.31}$$

Denoting

$$\nu_n(b) = \frac{1}{l_n} \sum_{u \in [n]} d_u(d_u - 1)\mathbb{1}\{d_u \leq b\}, \tag{2.32}$$

the following Lemma shows bound on the good and bad path probabilities.

Lemma 2.1.8. (Truncated path probabilities) *For every $k \geq 1$, $(b_l)_{l \geq 0}$ with $b_l \geq 0$ and $l \mapsto b_l$ non-decreasing,*

$$\mathbb{P}(\mathcal{F}_k(u)) \leq d_u \frac{l_n^k (l_n - 2k - 1)!!}{(l_n - 1)!!} [1 - F_n^*](b_k) \prod_{l=1}^{k-1} \nu_n(b_l), \tag{2.33}$$

and

$$\mathbb{P}(\mathcal{E}_k(u, v) \leq \frac{d_u d_v}{l_n} \frac{l_n^k (l_n - 2k - 1)!!}{(l_n - 1)!!} \prod_{l=1}^{k-1} \nu_n(b_l \wedge b_{k-l}). \tag{2.34}$$

Moreover,

$$\nu_n(b) \leq c_\nu b^{3-\tau}. \tag{2.35}$$

These bounds are obtained by applying the same path counting techniques that were used to obtain (2.24). For some constant $C > 0$ and arbitrary small $\delta \in (0, \tau - 2)$ define $b_l = C^{(\tau-2-\delta)^{-l}}$, such that $b_{l+1} = b_l^{\frac{1}{(\tau-2-\delta)}}$. It is shown that under this truncation sequence for $k_n = 2 \log\log n / |\log(\tau - 2)|$, that the bound (2.31) becomes arbitrary small. Intuitively, this bound on the bad paths tells us that it is unlikely a typical vertices are connected by a path through the hub, which is shorter than the path from the upper bound. And the bound on the good paths shows that it is unlikely there exists a shorter path that does not make use of the vertices with highest order degrees.

2.2 Local tree-like structure

This section we will prove the local tree-like structure of the locally finite directed configuration model. First, we will show how each exploration process can be coupled to a properly defined branching process in Section 2.2.1. Then in Section 2.2.2, we make use of the marked graphs and its available tools described in Section 1.3.2, to prove the marked local convergence in probability.

2.2.1 Coupling of the exploration process with a branching process

Take $\mathrm{DCM}_n(\boldsymbol{d}) = G_n$ for ease of notation and let $(G_n(s))_{(s)\in\mathbb{N}_0}$ denote the forward-backward exploration of G_n. Here, $G_n(s)$ is the graph after pairing a combined total of s out- and inbound half-edges, where $G_n(0)$ only contains the root and $G_n(1) = d_o$. Let $\mathcal{Z}_n$ denote the branching process with root distribution $(g_n(k,l))_{k,l\geq 0}$ and offspring distribution $(g_n^{*(\mathrm{out})}(k,l))_{k,l\geq 0}$ and $(g_n^{*(\mathrm{in})}(k,l))_{k,l\geq 0}$ for the following generations. Note the branching process is not constructed by pairing half-edges. So we define the *tree-exploration*, where the number of children of each vertex is inspected in a bread-first manner. Let $\mathcal{Z}_n(s)$ denotes the branching process after s tree-exploration steps. Similarly, $\mathcal{Z}_n(0)$ only contains the root o, and $\mathcal{Z}_n(1)$ the root o in addition with its degree d_0. The following Lemma shows that we can couple the forward-backward exploration and tree-exploration in such a way, that $(G_n(s))_{s=0}^{m_n}$ equals $(\mathcal{Z}_n(s)_{s=0}^{m_n}$ for $m_n \to \infty$ arbitrary slowly.

Lemma 2.2.1. (Coupling graph exploration to a branching process) *Take $d_{max} = \max_{u\in[n]}\{d_u^{(out)}, d_u^{(in)}\}$, denoting the maximal value of all the in- and out-degrees. Under Assumption 1.3.2 it holds that $d_{max} = o(n)$, and there exists a coupling $\left(\hat{G}_n(s), (\hat{\mathcal{Z}}_n(s)\right)_{(s)\in\mathbb{N}_0}$ of $(G_n(s))_{s\in\mathbb{N}_0}$ and $(\mathcal{Z}_n(s))_{s\in\mathbb{N}_0}$, such that*

$$\mathbb{P}\left((\hat{G}_n(s))_{s=0}^{m_n} \neq (\hat{\mathcal{Z}}_n(s))_{s=0}^{m_n}\right) \leq \frac{m_n^2(2+d_{\max})}{l_n}. \tag{2.36}$$

Proof. Start by taking a uniformly chosen root $o \in [n]$ with degrees $d_o = (d_o^{(\mathrm{in})}, d_o^{(\mathrm{out})})$, such that

$$\hat{G}_n(0) = \hat{\mathcal{Z}}(0) \text{ and } \hat{G}_n(1) = \hat{\mathcal{Z}}(1). \tag{2.37}$$

Given that for the first $m_n - 1$ coupling steps it holds that

$$(\hat{G}_n(s))_{s=0}^{m_n-1} = (\hat{\mathcal{Z}}_n(s))_{s=0}^{m_n-1}, \tag{2.38}$$

we construct $\hat{G}_n(m_n)$ depending or the first free half-edge of $\hat{G}_n(m_n-1)$ is an out- or inbound half-edge. If it is an outbound half-edge, the construction will go as follows. We let x_{m_n} denote the first unpaired outbound half-edge of $\hat{G}_n(m_n-1)$. Next, we draw an inbound half-edge y_{m_n} uniformly at random from *all* inbound half-edges. Let vertex U_n be the vertex to which inbound half-edge y_m is attached. Then let the m_n'th individual $\hat{\mathcal{Z}}_n(m_n)$ have $d_{U_n}^{(\mathrm{in})} - 1$ "in" marked offspring and $d_{U_n}^{(\mathrm{out})}$ "out" marked offspring. If y_{m_n} has not been paired yet in the process $(\hat{G}_n(s))_{s=0}^{m_n-1}$, we pair the half-edges and construct $(\hat{G}_n(t))_{s=0}^{m_n}$. However, if it has already been paired, we draw a new inbound half-edge $\hat{y}_{m_n}$ from the set of unpaired inbound half-edges and then, pair this with x_{m_n}.

Note that differences between the explorations can occur, when a new inbound or outbound half-edge has to be selected. Another possibility is that an outbound or inbound is paired with an respective inbound or outbound half-edge attached to a vertex that has already been used, resulting in a cycle. We will look at the probability that one of these errors occurs, and show that that the bound in holds.

Half-edge reuse: After $m_n - 1$ pairing steps, there are $m_n - 1$ inbound and outbound half-edges that will result in reuse. Hence, the probability that in- or outbound half edge y_m is already used is equal to $\frac{m_n-1}{l_n}$. It follows that the expected number of in- or outbound reuses before step m_n is

$$\sum_{m=1}^{m_n}(m-1)/l_n \leq \frac{(m_n)^2}{l_n}, \tag{2.39}$$

which we can use as an upper bound on the half-edge reuse probability by applying Markov's inequality.

Vertex reuse: Denoting $d_u^{\max} = \max\{d_u^{(\text{in})}, d_u^{(\text{out})}\}$, the probability that an inbound or outbound half-edge incident to vertex $u \in [n]$ is chosen is bounded by $d_u^{\max}/l_n$, such that the probability that an inbound or outbound half-edge incident to vertex $u \in [n]$ is chosen at least twice before step m_n is at most

$$\frac{m_n(m_n-1)}{2}\left(\frac{d_u^{\max}}{l_n}\right)^2. \tag{2.40}$$

From (2.40) it follows that the expected number of vertex reuses before step m_n is at most

$$\sum_{u\in[n]} \frac{m_n(m_n-1)}{2}\left(\frac{d_u^{\max}}{l_n}\right)^2 \leq \frac{1}{2}\left(\frac{m_n}{l_n}\right)^2 \sum_{u\in[n]} (d_u^{\max})^2. \tag{2.41}$$

Note that by Assumption (1.3.2) , we have that

$$d_{\max} = \max_{u\in[n]}\{d_u^{\max}\} = o(n). \tag{2.42}$$

As under Assumption (1.3.2) l_n is of order n, we can further bound (2.41) by

$$\frac{d_{\max} m_n^2}{2l_n}. \tag{2.43}$$

We can again apply Markov's inequality to bound the probability of vertex reuse before step m_n by (2.43). Summarizing both the bounds (2.39) and (2.43) results in the bound of the claim (2.36). □

Note that the same bound (2.36) holds when we couple the respective forward and backward exploration to branching processes $\mathcal{Z}_n^{(\text{out})}$ and $\mathcal{Z}_n^{(\text{in})}$ with generation sizes $(Z_k^{(n;\text{out})})_{k\geq 0}$ and $(Z_k^{(n;\text{in})})_{k\geq 0}$, described in Section 1.3.1. This will only add one vertex, which slightly increases the probability of vertex reuse, but the bound (2.41) will still hold, and with it the bound (2.36) in Lemma 2.2.1. Denoting the respective forward and backward explorations as $\left(G_n^{(\text{out})}(s)\right)_{s\in\mathbb{N}}$ $\left(G_n^{(\text{in})}(s)\right)_{s\in\mathbb{N}}$, this result is shown in the following corollary:

Corollary 2.2.2. *Consider the directed configuration model for which Assumption 1.3.2 is satisfied. Then there exists a coupling*

$$\left((\hat{G}_n^{(\text{out})}(s), \hat{G}_n^{(\text{in})}(t)), (\hat{\mathcal{Z}}_n^{(\text{out})}(s), \hat{\mathcal{Z}}_n^{(\text{in})}(t))\right)_{s+t=0}^{m_n} \tag{2.44}$$

of $(G_n^{(out)}(s), G_n^{(in)}(t))_{s+t\in\mathbb{N}_0}$ *and the couple of independent branching processes* $(\mathcal{Z}_n^{(out)}(s), \mathcal{Z}_n^{(in)}(t))_{s+t\in\mathbb{N}_0}$*, such that*

$$\mathbb{P}\left((\hat{G}_n^{(\text{out})}(s), \hat{G}_n^{(\text{in})}(t))_{s+t=0}^{m_n} \neq (\hat{\mathcal{Z}}_n^{(\text{out})}(s), \hat{\mathcal{Z}}_n^{(\text{in})}(t))_{s+t=0}^{m_n}\right) \leq \frac{m_n^2(2+d_{max})}{l_n}. \tag{2.45}$$

2.2.2 Marked Local weak convergence

This section we will prove asymptotic properties for the neighborhood of a typical vertex. To do this, we will apply the mapping of the directed graph to the marked graph described in Section 1.3.2. This will allow us to make use of the tools for the marked graph, that are introduced in the same section. We will reuse some notation and denote the marked graph by $(G_n, M(G_n)$. To start, we prove that the branching processes that we have used in the coupling in Lemma 2.2.1 converge to the marked branching processes $\mathcal{Z}$ defined in 1.3.3.

Lemma 2.2.3. *Let Assumption 1.3.2 be satisfied. Let $\mathcal{Z}(s)$ denote the tree exploration process on the marked branching process after exploring s vertices. Then, for all trees $\boldsymbol{t} \in \mathcal{G}^*$, we have*

$$\mathbb{P}(\mathcal{Z}_n(s) \simeq \boldsymbol{t}(s)) \xrightarrow{n\to\infty} \mathbb{P}(\mathcal{Z}(s) \simeq \boldsymbol{t}(s)) \tag{2.46}$$

Proof. Excluding the root, let $t_{(\text{out})}$ and $t_{(\text{in})}$ denote the respective number of "out" and "in" vertices in the first s explored vertices of $\boldsymbol{t}$. Starting from the root with index 0, we index the "in" vertices as $1, \dots, t_{(\text{in})}$, and the "out" vertices as $t_{(\text{in})}+1, \dots, t_{(\text{out})}$, such that $s = t_{(\text{in})} + t_{(\text{out})} + 1$. Under Assumption 1.3.2, it follows that

$$\begin{aligned} f_n(k,l) &\xrightarrow{n\to\infty} f(k,l), \\ f_n^{*(\text{out})}(k,l) &\xrightarrow{n\to\infty} g^{*(\text{out})}(k,l), \\ f_n^{*(\text{in})}(k,l) &\xrightarrow{n\to\infty} f^{*(\text{in})}(k,l). \end{aligned} \tag{2.47}$$

So given that $\boldsymbol{t}$ has the degree-sequence $(\hat{d}^{(\text{in})}(i), \hat{d}^{(\text{out})}(i))_{i\in[s]}$, we have

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(\mathcal{Z}_n(s) \simeq \boldsymbol{t}(s)) = &\lim_{n\to\infty} f_n(\hat{d}^{(\text{in})}(0), \hat{d}^{(\text{out})}(0)) \prod_{i=1}^{t_{(\text{out})}} f_n^{*(\text{out})}(\hat{d}^{(\text{in})}(i), \hat{d}^{(\text{out})}(i)) \\ &\times \prod_{j=t_{(\text{out})}+1}^{t_{(\text{in})}+t_{(\text{out})}} f_n^{*(\text{in})}(\hat{d}^{(\text{in})}(j), \hat{d}^{(\text{out})}(j)) \\ =&f(\hat{d}^{(\text{in})}(0), \hat{d}^{(\text{out})}(0)) \prod_{i=1}^{t_{(\text{out})}} f^{*(\text{out})}(\hat{d}^{(\text{in})}(i), \hat{d}^{(\text{out})}(i)) \\ &\times \prod_{j=t_{(\text{out})}+1}^{t_{(\text{in})}+t_{(\text{out})}} f^{*(\text{in})}(\hat{d}^{(\text{in})}(j), \hat{d}^{(\text{out})}(j)) \\ =& \mathbb{P}(\mathcal{Z}(s) \simeq \boldsymbol{t}(s)). \end{aligned} \tag{2.48}$$

□

To prove the marked local convergence results, we will make use of the following convenient theorem from ([14], Theorem 2.15):

Theorem 2.2.4. *Let $(G_n, (M(G_n))_{n\geq 1}$ be a sequence of marked graphs. Then*

1. $(G_n, o_n, M(G_n)) \xrightarrow{d} (G, o, M(G))$ *having probability law μ when, for every marked rooted graph $\boldsymbol{t} \in \mathcal{G}_*$ and all integers $r \geq 0$,*

$$\mathbb{E}[\mathbb{1}\{B_r^{(G_n)}(o) \simeq \boldsymbol{t}\}] = \frac{1}{n}\sum_{u\in[n]} \mathbb{P}(B_r^{(G_n)}(v) \simeq \boldsymbol{t}) \xrightarrow{n\to\infty} \mu(B_r^{(G)}(o) \simeq \boldsymbol{t}), \tag{2.49}$$

2. $(G_n, M(G_n))$ *converges marked locally weakly in probability to $(G, o, M(G))$ having probability law μ when, for every marked rooted graph $\boldsymbol{t} \in \mathcal{G}_*$ and all integers $r \geq 0$,*

$$\frac{1}{n}\sum_{u\in[n]} \mathbb{1}\{B_r^{(G_n)}(v) \simeq \boldsymbol{t}\} \xrightarrow{\mathbb{P}} \mu(B_r^{(G)}(o) \simeq \boldsymbol{t}). \tag{2.50}$$

Next, with the following claim we show that any bounded mapping on a bounded neighborhood is bounded and continuous, such that Theorem 2.2.4 applies.

Claim 2.2.5. (Mappings on neighborhoods are continuous) *Take the marked rooted graph $(G, o, M(G))$. Then for all $r \geq 1$ mappings on the neighborhood $B_r^G(o)$ are continuous on $\mathcal{G}_*$.*

Proof. For every marked rooted graph $(G_1, o_1, M(G_1))$ and $r \geq 1$ we can take a marked rooted graph $(G_2, o_2, M(G_2)) \in \mathcal{G}_*$ such that

$$d_{\mathcal{G}_*}((G_1, o_1, M_1(G_1)), (G_2, o_2, M_2(G_2))) < 1/(r+1). \tag{2.51}$$

It follows by Definition 1.3.5 that $B_r^{G_1}(o_1) \sim B_r^{G_2}(o_2)$, such that they are mapped to the same element. So the claim follows. □

For all $\boldsymbol{t} \in \mathcal{G}^*$, let

$$N_{n,r}(\boldsymbol{t}) = \sum_{u \in [n]} \mathbb{1}\{B_r^{(G_n)}(u) \simeq \boldsymbol{t}\}, \quad \boldsymbol{t} \in \mathcal{G}_* \tag{2.52}$$

denoting the number vertices $u \in [n]$ such that the neighborhood $B_r^{(G_n)}(u)$ is identical to $\boldsymbol{t}$ modulo isomorphism.

Theorem 2.2.6. *Denote the marked branching process restricted to the first r generations as $\mathcal{Z}_{\leq r}$. Given that regularity Assumption 1.3.2 is satisfied, it follows that for all marked rooted trees $\boldsymbol{t} \in \mathcal{G}_*$*

$$\frac{1}{n} N_{n,r}(\boldsymbol{t}) \xrightarrow{\mathbb{P}} \mu(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t}). \tag{2.53}$$

Consequently, $(G_n, M(G_n))$ converges marked locally weakly in probability to the marked branching process.

Proof. For the proof we will make use of the second moment method, i.e., we will show that for all $\boldsymbol{t} \in \mathcal{G}_*$, the first moment

$$\mathbb{E}[N_{n,r}(\boldsymbol{t})]/n \xrightarrow{n\to\infty} \mu(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t}), \tag{2.54}$$

and for the second moment it holds that $\mathrm{Var}(N_{n,r}(\boldsymbol{t})]/n) = o(n^2)$. Then, we will apply Chebyshev's inequality to prove the claim.

First moment Let t_r denote the number of individuals of the first $r-1$ generations in $\boldsymbol{t} \in \mathcal{G}_*$, and $(\boldsymbol{t}(s))_{s\in[t_r]}$ denote the breadth-first tree exploration of the first t_r individuals. Then,

$$\begin{aligned}\mathbb{E}[N_{n,r}(\boldsymbol{t})]/n =& \mathbb{P}((G_n(s))_{s\in[t_r]} = (\boldsymbol{t}(s))_{s\in[t_r]}) \\ =& \mu((\mathcal{Z}_n(s))_{s\in[t_r]} = (\boldsymbol{t}(s))_{s\in[t_r]}) + o(1) \\ =& \mu((\mathcal{Z}(s))_{s\in[t_r]} = (\boldsymbol{t}(s))_{s\in[t_r]}) + o(1) \\ =& \mu\left(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t}\right) + o(1),\end{aligned} \tag{2.55}$$

where the second equality follows from the coupling of Lemma 2.2.1, and the third equality from the convergence of the branching process from Lemma 2.2.3.

Second moment For the second moment, we first note that for uniformly chosen vertices o_1 and o_2 we have the equality

$$\mathbb{E}\left[\left(\frac{N_{n,r}(\boldsymbol{t})}{n}\right)^2\right] = \mathbb{P}\left(B_r^{(G_n)}(o_1), B_r^{(G_n)}(o_2) \simeq \boldsymbol{t}\right). \tag{2.56}$$

Note that with Theorem 2.2.4 we can conclude from (2.55) that $(G_n, o_n, M(G_n)) \xrightarrow{d} \mathcal{Z}$. Moreover, by Claim 2.2.5 the mapping $(G_n, o_n, M(G_n)) \mapsto \mathbb{1}\{|B_r^{G_n}(o)| = k\}$ is bounded and continuous on $\mathcal{G}_*$, such that again by Theorem 2.2.4 $|B_r^{G_n}(o_n)| \xrightarrow{d} |\mathcal{Z}_{\leq r}|$, which is a tight random variable. It follows that $|B_r^{(G_n)}(o_1)|/n = o(1)$, such that

$$\mathbb{E}\left[\left(\frac{N_{n,r}(\boldsymbol{t})}{n}\right)^2\right] = \mathbb{P}\left(B_r^{(G_n)}(o_1), B_r^{(G_n)}(o_2) \simeq \boldsymbol{t}, o_2 \notin B_{2r}^{(G_n)}(o_1)\right) + o(1). \tag{2.57}$$

Furthermore, by conditioning on $B_r^{(G_n)}(o_2) \simeq \boldsymbol{t}$, we can take

$$\begin{aligned}
&\mathbb{P}(B_r^{(G_n)}(o_1), B_r^{(G_n)}(o_2) \simeq \boldsymbol{t}, o_2 \notin B_{2r}^{(G_n)}(o_1)) \\
&= \mathbb{P}\left(B_r^{(G_n)}(o_2) \simeq \boldsymbol{t} | B_r^{(G_n)}(o_1) \simeq \boldsymbol{t}, o_2 \notin B_{2r}^{(G_n)}(o_1)\right) \\
&\times \mathbb{P}(B_r^{(G_n)}(o_1) \simeq \boldsymbol{t}, o_2 \notin B_{2r}^{(G_n)}(o_1)).
\end{aligned} \tag{2.58}$$

For the first moment, we have already shown that $\mathbb{P}(B_r^{(G_n)}(o_1) \simeq \boldsymbol{t}) \xrightarrow{n\to\infty} \mu\left(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t}\right)$, such that

$$\mathbb{P}(B_r^{(G_n)}(o_1) \simeq \boldsymbol{t}, o_2 \notin B_{2r}^{(G_n)}(o_1)) \xrightarrow{n\to\infty} \mu\left(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t}\right). \tag{2.59}$$

As $|B_{2r}^{(G_n)}(o_1)|$ is a tiht random variable, given that $o_2 \notin B_{2r}^{(G_n)}(o_1)$ and $B_r^{(G_n)}(o_1) \simeq \boldsymbol{t}$, the probability that $B_r^{(G_n)}(o_2) \simeq \boldsymbol{t}$ is the same as the probability this event happens in $\mathrm{DCM}_n(\boldsymbol{d}')$, which is obtained by removing all the vertices in $B_{2r}^{(G_n)}(o_1)$ from $\mathrm{DCM}_n(\boldsymbol{d})$. The important result is that the regularity assumptions 1.3.2 still hold for $\mathrm{DCM}_n(\boldsymbol{d}')$, such that

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(B_r^{(G_n)}(o_2) \simeq \boldsymbol{t} | B_r^{(G_n)}(o_1) \simeq \boldsymbol{t}, o_2 \notin B_{2r}^{(G_n)}(o_1)) \\
&\quad = \mu\left(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t}\right).
\end{aligned} \tag{2.60}$$

Applying (2.59) and (2.60) on (2.58). we obtain that

$$\mathbb{E}\left[\left(\frac{N_{n,r}(\boldsymbol{t})}{n}\right)^2\right] \xrightarrow{n\to\infty} \mu\left(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t}\right)^2. \tag{2.61}$$

Chebyshev's inequality To finish the proof of the first claim, we use Chebyshev's inequality, and the results of the first and second moment to conclude

$$\begin{aligned}
&\mathbb{P}(|\frac{N_{n,r}(\boldsymbol{t})}{n} - \mu(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t})| \geq \varepsilon) \\
&\leq \mathbb{P}(|\frac{N_{n,r}(\boldsymbol{t})}{n} - \mathbb{E}[\frac{N_{n,r}(\boldsymbol{t})}{n}]| \geq \varepsilon - |\mathbb{E}[\frac{N_{n,r}(\boldsymbol{t})}{n}] - \mu(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t})|) \\
&\leq \frac{\mathrm{Var}(\frac{N_{n,r}(\boldsymbol{t})}{n})}{\varepsilon - |\mathbb{E}[\frac{N_{n,r}(\boldsymbol{t})}{n}] - \mu(\mathcal{Z}_{\leq r} \simeq \boldsymbol{t})|} \xrightarrow{n\to\infty} 0.
\end{aligned} \tag{2.62}$$

As $B_r^{(G_n;\mathrm{out})}(o_n)$ and $B_r^{(G_n;\mathrm{in})}(o_n)$ are subgraphs of $B_r^{(G_n)}(o_n)$, the marked forward and marked backward locally weak convergence follows immediately from Theorem 2.2.4 and Claim 2.2.5. □

Next, we will prove that the respective number of vertices that are k steps away from a typical vertex, and the number of vertices that can reach an independently chosen typical vertex in k steps, converges locally weakly in distribution to $Z_k^{(\mathrm{out})}$ and $Z_k^{(\mathrm{in})}$.

Theorem 2.2.7. *Assume that regularity Assumption 1.3.2 is satisfied. Let $Z_k^{(out)}$ denote the number of vertices in the k'th generation of the marked forward branching process $\mathcal{Z}^{(out)}$, and $Z_k^{(in)}$ similarly for the marked backward branching process $\mathcal{Z}^{(in)}$. Then, for $m > 0$ arbitrary small*

$$(H_r^{(n;\mathrm{out})}, H_r^{(n;\mathrm{in})})_{r=0}^m \xrightarrow{d} (Z_r^{(\mathrm{out})}, Z_r^{(\mathrm{in})})_{r=0}^m. \tag{2.63}$$

Proof. Note that by Claim 2.2.5 both the functions $q \in \{\mathrm{out}, \mathrm{in}\}$

$$\mathbb{1}\{|\partial B_r^{(G_n;q)}(o)| = l_r, \forall r \leq m\}, \tag{2.64}$$

are bounded and continuous functions on $\mathcal{G}_*$. It follows from Theorem 2.2.4 and Theorem 2.2.6 that both $\left(|\partial B_r^{G_n;q)}(o)|\right)_{r=0}^m \xrightarrow{\mathbb{P}} \left(Z_r^{(q)}\right)_{r=0}^m$. Now note if $B_r^{G_n;q)}(o)$ is a tree, which happens whp for fixed r according to Theorem 1.3.8, that $|\partial B_r^{G_n;q)}(o)| = H_r^{(n;q)}$. Thus, we also have that $\left(H_r^{(n;q)}\right)_{r=0}^m \xrightarrow{\mathbb{P}} \left(Z_r^{(q)}\right)_{r=0}^m$. For ease of notation, denote the events

$$\begin{aligned} H^{(q)} &= \{H_r^{(n;q)} = l_r, \forall r \leq m\} \\ T^{(q)} &= \{Z_r^{(q)} = l_r, \forall r \leq m\}. \end{aligned} \tag{2.65}$$

As o_1 and o_2 are independently and uniformly chosen, and the two branching processes are independent, it follows that

$$\mathbb{1}\{H^{(\text{out})}, H^{(\text{in})}\} = \mathbb{1}\{H^{(\text{out})}\}\mathbb{1}\{H^{(\text{in})}\}, \quad \mathbb{1}\{T^{(\text{out})}, T^{(\text{in})}\} = \mathbb{1}\{T^{(\text{out})}\}\mathbb{1}\{T^{(\text{in})}\}. \tag{2.66}$$

Take

$$|\mathbb{E}[\mathbb{1}\{H^{(\text{out})}\}\mathbb{1}\{H^{(\text{in})}\}] - \mu(T^{(\text{out})})\mu(T^{(\text{in})})| \leq \mathbb{E}[|\mathbb{1}\{H^{(\text{out})}\}\mathbb{1}\{H^{(\text{in})}\} - \mu(T^{(\text{out})})\mu(T^{(\text{in})})|]. \tag{2.67}$$

By adding and subtracting the term $\mathbb{1}\{H^{(\text{out})}\}\mu(T^{(\text{in})})$ inside the expectation, we can apply the triangle inequality to bound (2.67) by

$$\begin{aligned} &\mathbb{E}[|\mu(T^{(\text{in})}|||\mathbb{1}\{H^{(\text{out})}\} - \mu(T^{(\text{out})})|] + \mathbb{E}[|\mathbb{1}\{H^{(\text{out})}\}||\mathbb{1}\{B^{(\text{in})}\} - \mu(T^{(\text{in})})|] \\ &\leq \mathbb{E}[||\mathbb{1}\{H^{(\text{out})}\} - \mu(T^{(\text{out})})|] + \mathbb{E}[|\mathbb{1}\{H^{(\text{in})}\} - \mu(T^{(\text{in})})|]. \end{aligned} \tag{2.68}$$

For ease for notation let $\mathbb{E}_{<\varepsilon}[X] = \mathbb{E}[X\mathbb{1}\{X < \varepsilon\}]$ and $\mathbb{E}_{\geq\varepsilon}[X] = \mathbb{E}[X\mathbb{1}\{X \geq \varepsilon\}]$ for some $\varepsilon > 0$. We can split both terms on the RHS by

$$\mathbb{E}_{<\varepsilon}[||\mathbb{1}\{H^{(q)}\} - \mu(T^{(q)})|] + \mathbb{E}_{\geq\varepsilon}[||\mathbb{1}\{H^{(q)}\} - \mu(T^{(q)})|], \tag{2.69}$$

for $q \in \{\text{out}, \text{in}\}$. The expectation on the left clearly is smaller than ε. Moreover, from the marked weak local convergence in probability,

$$\begin{aligned} \mathbb{E}_{\geq\varepsilon}[|\mathbb{1}\{H^{(q)}\} - \mu(T^{(q)})|] \leq &\mathbb{E}[\mathbb{1}\{|\mathbb{1}\{H^{(q)}\} - \mu(T^{(q)})| \geq \varepsilon\}] \\ = &\mathbb{P}(|\mathbb{1}\{H^{(q)}\} - \mu(T^{(q)})| \geq \varepsilon) = o(1). \end{aligned} \tag{2.70}$$

Hence, the limit as $n \to \infty$ of the expression in (2.67) becomes smaller than any arbitrary $\varepsilon > 0$. This proves the marked local weak convergence in distribution. □

2.3 Upper bound typical distances directed configuration model

This section we will prove a double logarithmic upper bound for the typical distance in the directed configuration model under the conditions that are given in Theorem 1.5.4. To do this, we adapt the approach used for the configuration Model shown in Section 2.1.3 to the directed setting. We first prove that under the assumptions of Theorem 1.5.4, a vertex with a sufficiently large out-degree is likely to have a neighbor with an even exponentially larger out-degree. And similarly, a vertex with a high in-degree is likely to be the neighbor of a vertex with an exponentially larger in-degree. Consequently, we shall see that the number of steps needed to get from any of these high out-degree vertices to any high in-degree vertex, grows at most double logarithmic with the size of the graph. To finish the proof, we will apply the approximation of the forward and backward explored neighborhoods with their respective branching processes, as shown in Theorem 1.3.9. This will show the length of a path from typical vertex to a high-degree vertex, and the path from a high in-degree vertex to a typical vertex, are both at most of order $\log\log(n)$. We will consistently apply the lower bound given by Potter's Theorem 1.5.2. Hence, for ease of notation we take

$$\gamma^{(q)} = \gamma^{(q)}(\delta^{(q)}) = (\tau^{(q)} + \delta^{(q)}), \quad q \in \{\text{out}, \text{in}\}, \tag{2.71}$$

and show at the end of this section that each proof holds for all $\delta^{(q)} > 0$ arbitrary small.

2.3.1 Distances from the outbound to the inbound core

To analyze the connectivity between the high in- and high out-degree vertices, we start with the following adaption of Lemma 2.3.2 which provides a bound on the probability that two sets are *not* connected.

Lemma 2.3.1. (Connectivity between sets in $\mathrm{DCM}_n(\boldsymbol{d})$) *For any two sets of vertices $A, B \subseteq [n]$,*

$$\mathbb{P}(A \text{ not directly connected to } B) \leq e^{-d_A^{(\text{out})} d_B^{(\text{in})}/l_n}, \tag{2.72}$$

where

$$d_A^{(\text{out})} = \sum_{u \in A} d_u^{(\text{out})}, \quad d_B^{(\text{in})} = \sum_{u \in B} d_u^{(\text{in})} \tag{2.73}$$

denotes the respective total out-degree of A and total in-degree of B.

Proof. From the construction of the directed configuration model, we are allowed to pick the outbound half-edges attached to the vertices in the set A one by one, and pair them with an uniformly chosen unpaired inbound half-edge. If the first k outbound half-edges are not paired with any of the $d_B^{(\text{in})}$ inbound half-edges attached to vertices of set B, then the probability that this also holds for the following outbound half-edge is

$$1 - \frac{d_B^{(\text{in})}}{l_n - k} \leq 1 - \frac{d_B^{(\text{in})}}{l_n}. \tag{2.74}$$

We apply (2.74) to bound the probability that this happens for all of the $d_A^{(\text{out})}$ outbound half-edges by

$$\mathbb{P}(A \text{ not directly connected to } B) \leq \left(1 - \frac{d_B^{(\text{in})}}{l_n}\right)^{d_A^{(\text{out})}} \leq e^{-d_A^{(\text{out})} d_B^{(\text{in})}/l_n}, \tag{2.75}$$

where we have used that $1 - x \leq e^{-x}$ for all real valued x. □

The bound (2.72) appears sharper than the bound for the undirected configuration model from Lemma 2.3.2, which contains an extra 1/2 in the exponent. This difference arises as the configuration model contains only one type of half-edge. Therefore, a half-edge of set A can pair with another half-edge of set A, using two half-edges in one step. This reduces the number of chances the set A has to connect with a half-edge attached to set B. As for the $\mathrm{DCM}_n(\boldsymbol{d})$ the outbound half-edges can only pair with an inbound half-edge, there is no possibility of a single pairing step using two outbound half-edges. But before comparing the sharpness of the bounds, we need to take some more factors into consideration. For one, while in the configuration model the set A has a chance to connect with B through *all* of set A and B's available half-edges, the directed configuration needs to connect A's outbound half-edges to B's inbound half-edges. This reduces the number of chances for the sets to connect for the two-type half-edges model, compared to the one-type model. However, the l_n in the configuration model denotes the number of *all* half-edges, and for the directed configuration model it only denotes the number of one type of half-edges, which is half of the total number. Considering all these factors, we can not make any comparisons between the two bounds without further knowledge about the sets A and B.

Define the vertex sets

$$\begin{aligned} \mathcal{N}^{(\text{out})}(v) &:= \{u \in [n] : \mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(v, u) = 1\} \\ \mathcal{N}^{(\text{in})}(v) &:= \{u \in [n] : \mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(u, v) = 1\}, \end{aligned} \tag{2.76}$$

such that $\mathcal{N}^{(\text{out})}(v)$ denotes the set of neighboring vertices of v and $\mathcal{N}^{(\text{in})}(v)$ denotes the set of vertices that have v as a neighbor. Lemma 2.3.1 can be used to provide a bound on the probability that some vertex u is connected to a vertex v, depending on the out-degree of u and the in-degree of v. For the configuration model we have shown in Section 2.3 that given that the degree distribution follows a power law, each high degree vertex is likely to have a neighbor which has an even exponentially larger degree, creating a $\log\log n$ paths toward the highest order vertices. To create a similar phenomenon in the directed configuration model, it will not be enough to directly assume the out- and the in-degree distributions to follow a power law. As Lemma 2.3.1 provides a bound depending on the out-degree of u and in-degree of v, this will not ensure any properties for the out-degree of v. So a vertex might have a large number "entrances" (inbound half-edges), it does not have to contain a large number of "exits" (outbound half-edges) out-degree. So it might would not be likely to find such a path of rapidly increasing out-degree vertices toward the high order out-degree vertices. To resolve this issue, we adapt the power law assumption to hold for the forward out- and in-degree distributions $F_n^{(\text{out})}$ and $F_n^{(\text{in})}$, as defined in (1.37). Lemma 2.3.1 shows that for a vertex with degree $d_v = (d_v^{(\text{in})}, d_v^{(\text{out})})$, the probability that it does not have a neighbor with an out-degree larger than y_1, and the probability that it is not the neighbor of a vertex with an in-degree larger than y_2, can be bounded by

$$\begin{aligned}\mathbb{P}\left(\forall u \in \mathcal{N}^{(\text{out})}(v), d_u^{(\text{out})} \leq y_1\right) &\leq e^{-\frac{d_v^{(\text{out})}}{l_n}\sum_u d_u^{(\text{in})}\mathbb{1}\{d_u^{(\text{out})}>y_1\}} = e^{-d_v^{(\text{out})}[1-F_n^{(\text{out})}](y_1)},\\ \mathbb{P}\left(\forall u \in \mathcal{N}^{(\text{in})}(v), d_u^{(\text{in})} \leq y_2\right) &\leq e^{-\frac{d_v^{(\text{in})}}{l_n}\sum_u d_u^{(\text{out})}\mathbb{1}\{d_u^{(\text{in})}>y_2\}} = e^{-d_v^{(\text{in})}[1-F_n^{(\text{in})}](y_2)}.\end{aligned} \tag{2.77}$$

With the power-law Assumption 1.5.3, the tail distribution $[1 - F_n^{(q)}](y)$ is close to $y^{-(\tau^{(q)}-2)}$ for both $q \in \{\text{out}, \text{in}\}$. So the bounds (2.77) become extremely small for $d_v^{(q)}$ large, when $y \ll d_v^{1/(\tau^{(q)}-2)}$. This shows that a vertex with a large out-degree $d_v^{(\text{out})}$ is likely to have at least one neighbor with approximate out-degree $d^{1/(\tau^{(\text{out})}-2)}$. Similarly, a vertex with large in-degree $d_v^{(\text{in})}$ is likely to be the neighbor of a vertex with approximate in-degree $d_v^{1/(\tau^{(\text{in})}-2)}$. To further study this phenomenon, take

$$\sigma^{(\text{out})} > 1/(3-\gamma^{(\text{out})}), \quad \sigma^{(\text{in})} > 1/(3-\gamma^{(\text{in})}), \tag{2.78}$$

and define the set of high in- or out-degree vertices, called the *inbound core* and the *outbound core*, as

$$\text{Core}_n^{(\text{in})} = \{u : d_u^{(\text{in})} \geq (\log n)^{\sigma^{(\text{in})}}\}, \quad \text{Core}_n^{(\text{out})} = \{u : d_u^{(\text{out})} \geq (\log n)^{\sigma^{(\text{out})}}\}. \tag{2.79}$$

We will give a precise proof of the previous arguments, and apply this phenomenon to show that each vertex in $\text{Core}_n^{(\text{out})}$ has a path to the highest order out-degree vertices $\{u : d_u^{(\text{out})} \geq l_n^\beta\}$, and each of the highest in-degree vertices $\{u : d_u^{(\text{in})} \geq l_n^\beta\}$ has a path to any vertex in $\text{Core}_n^{(\text{in})}$, for which the length of both is at most order $\log\log(n)$.

To construct these paths we divide both $\text{Core}_n^{(\text{out})}$ and $\text{Core}_n^{(\text{in})}$ into segments. Then we show that each vertex in a segment of $\text{Core}_n^{(\text{out})}$ has a neighbor in the subsequent segment, and each vertex a segment of $\text{Core}_n^{(\text{in})}$ has a neighbor in the previous segment. For both $q \in \{\text{out}, \text{in}\}$, take $u_1^{(q)} = l_n^\beta$, and define the first segment as

$$\Gamma_1^{(q)} = \{u : d_u^{(q)} \geq u_1^{(q)}\}, \tag{2.80}$$

containing the highest order in- or out-degree vertices. To construct the following segments, take some constant $C > 0$ and for $k \geq 2$ define

$$u_k^{(q)} = C\log n (u_{k-1}^{(q)})^{\gamma^{(q)}-2}, \tag{2.81}$$

with segments

$$\Gamma_k^{(q)} = \{u : d_u^{(q)} \geq u_k^{(q)}\}. \tag{2.82}$$

As $u_k^{(q)} \leq u_{k-1}^{(q)}$ for all $k \geq 1$, the segments are ordered as $\Gamma_{k-1}^{(q)} \subset \Gamma_k^{(q)}$. Recursively solving (2.81) for every $k \geq 1$, leads to the expression

$$u_k^{(q)} = (C\log(n))^{a_k^{(q)}} l_n^{b_k^{(q)}}, \tag{2.83}$$

where

$$a_k^{(q)} = \sum_{i=1}^{k-1} (\gamma^{(q)} - 2)^{i-1} = \frac{1}{3-\gamma^{(q)}}[1 - (\gamma^{(q)} - 2)^{k-1}], \quad b_k^{(q)} = \beta(\gamma^{(q)} - 2)^{k-1}. \tag{2.84}$$

This can be verified, by simply substituting these results in the expression (2.81). The following lemma is an adaption of ([14], Prop 7.14), which shows the connectivity of the segments as previously described:

Lemma 2.3.2. (Connectivity between $\Gamma_k^{(q)}$ and $\Gamma_{k-1}^{(q)}$) *Let Assumptions 1.3.2 and 1.5.3 be satisfied. Fix $k \geq 2$ and for $c = \max\{c_2^{(out)}, c_2^{(in)}\}$ defined in 1.5.3, take $C > 1/c$. Then the probability that there exists an $u \in \Gamma_k^{(out)}$ that does not have a neighbor in the set $\Gamma_{k-1}^{(out)}$ is at most*

$$\mathbb{P}\left(\exists u \in \Gamma_k^{(\text{out})} : \mathcal{N}^{(\text{out})}(u) \cap \Gamma_{k-1}^{(\text{out})} = \emptyset\right) \leq n^{1-c_2^{(\text{out})}C} = o(1). \tag{2.85}$$

Moreover, the probability that there exists an $u \in \Gamma_k^{(in)}$ that is not the neighbor of some vertex in the set $\Gamma_{k-1}^{(in)}$ is at most

$$\mathbb{P}\left(\exists u \in \Gamma_k^{(\text{in})} : \mathcal{N}^{(\text{in})}(u) \cap \Gamma_{k-1}^{(\text{in})}) = \emptyset\right) = n^{1-c_2^{(\text{in})}C} = o(1). \tag{2.86}$$

Furthermore, every vertex $v \in \Gamma_1^{(in)}$ is whp a neighbor of every $u \in \Gamma_1^{(out)}$, i.e.

$$\mathbb{P}\left(\Gamma_1^{(\text{in})} \subset \mathcal{N}^{(\text{out})}(\Gamma_1^{(\text{out})})\right) = 1 - o(1). \tag{2.87}$$

Proof. For both $q \in \{\text{out}, \text{in}\}$ the mapping $k \mapsto u_k^{(q)}$ is decreasing, such that from Assumption 1.5.3 we get

$$\begin{aligned} \sum_{v \in \Gamma_{k-1}^{(\text{out})}} d_v^{(\text{in})} &= l_n[1 - F_n^{(\text{out})}](u_{k-1}^{(\text{out})}) \geq l_n c_2^{(\text{out})} (u_{k-1}^{(\text{out})})^{2-\gamma^{(\text{out})}}, \\ \sum_{v \in \Gamma_{k-1}^{(\text{in})}} d_v^{(\text{out})} &= l_n[1 - F_n^{(\text{in})}](u_{k-1}^{(\text{in})}) \geq l_n c_2^{(\text{in})} (u_{k-1}^{(\text{in})})^{2-\gamma^{(\text{in})}}. \end{aligned} \tag{2.88}$$

Applying (2.88) in the bounds from Lemma 2.3.1, we use the fact that $|\Gamma_k^{(q)}| \leq n$ and Boole's inequality to bound the probability that some $v \in \Gamma_k^{(\text{out})}$ does not have a neighbor in $\Gamma_{k-1}^{(\text{out})}$ by

$$\begin{aligned} \mathbb{P}\left(\exists u \in \Gamma_k^{(\text{out})} : \mathcal{N}^{(\text{out})}(u) \cap \Gamma_{k-1}^{(\text{out})} = \emptyset\right) &\leq n e^{-u_k^{(\text{out})} c_2^{(\text{out})} (u_{k-1}^{(\text{out})})^{2-\gamma^{(\text{out})}}} \\ &= n e^{-c_2^{(\text{out})} C \log(n)} = n^{1-c_2^{(\text{out})}C}, \end{aligned} \tag{2.89}$$

and similarly

$$\mathbb{P}\left(\exists u \in \Gamma_k^{(\text{in})} : \mathcal{N}^{(\text{in})}(u) \cap \Gamma_{k-1}^{(\text{in})}) = \emptyset\right) \leq n^{1-c_2^{(\text{in})}C}. \tag{2.90}$$

Finally, as $\beta \in (\frac{1}{2}, 1)$ the lemma 2.3.1 shows that

$$\mathbb{P}\left(\forall u \in \Gamma_1^{(\text{out})}, \Gamma_1^{(\text{in})} \not\subset \mathcal{N}^{(\text{out})}(u)\right) \leq e^{-l_n^{2\beta}/l_n} = o(1), \tag{2.91}$$

which proves the claim (2.87). ☐

We will apply this lemma to provide an upper bound of the distance from $\text{Core}_n^{(\text{out})}$ to $\text{Core}_n^{(\text{in})}$ in the following theorem.

Theorem 2.3.3. (Path between the Cores) *Consider the directed configuration model, where the Assumptions 1.3.2 and 1.5.3 are satisfied. For any* $\sigma^{(in)} > 1/(3-\gamma^{(in)})$ *and* $\sigma^{(out)} > 1/(3-\gamma^{(out)})$*, the number of (directed) steps needed to get from any* $u \in \mathrm{Core}_n^{(out)}$ *to any* $v \in \mathrm{Core}_n^{(in)}$ *is whp bounded from above by*

$$\frac{\log\log n}{|\log(\gamma^{(\mathrm{in})}-2)|} + \frac{\log\log n}{|\log(\gamma^{(\mathrm{out})}-2)|} + 1. \tag{2.92}$$

Proof. Take

$$k_{(\mathrm{out})} = \frac{\log\log n}{|\log(\gamma^{(\mathrm{out})}-2)|}, \quad k_{(\mathrm{in})} = \frac{\log\log n}{|\log(\gamma^{(\mathrm{in})}-2)|}. \tag{2.93}$$

From Lemma 2.3.2 it follows that we can whp bound the length of the path from any vertex $u \in \Gamma^{(\mathrm{out})}_{k_{(\mathrm{out})}}$ to any a vertex in $\Gamma^{(\mathrm{out})}_1$ by $k_{(\mathrm{out})}$, and any vertex $u \in \Gamma^{(\mathrm{in})}_1$ to a vertex in $\Gamma^{(\mathrm{in})}_{k_{(\mathrm{in})}}$ by $k_{(\mathrm{in})}$. As Theorem 2.3.2 shows that whp $\Gamma^{(\mathrm{in})}_1 \subset \mathcal{N}(\Gamma^{(\mathrm{out})}_1)$, we can conclude that the number of steps from any vertex in $\Gamma^{(\mathrm{out})}_{k_{(\mathrm{out})}}$ to any vertex in $\Gamma^{(\mathrm{in})}_{k_{(\mathrm{in})}}$ can be bounded from above by $k_{(\mathrm{out})} + k_{(\mathrm{in})} + 1$. To complete the claim of the theorem, it remains to show that

$$\mathrm{Core}_n^{(\mathrm{out})} \subseteq \Gamma^{(\mathrm{out})}_{k_{(\mathrm{out})}} \text{ and } \mathrm{Core}_n^{(\mathrm{in})} \subseteq \Gamma^{(\mathrm{in})}_{k_{(\mathrm{in})}}. \tag{2.94}$$

This is the same as showing that $u^{(q)}_{k_{(q)}} \leq (\log n)^{\sigma^{(q)}}$, where

$$u^{(q)}_{k_{(q)}} = (\log(n))^{a^{(q)}_{k_{(q)}}} l_n^{b_{k_{(q)}}}. \tag{2.95}$$

From the expression of in $b_{k_{(q)}}$ in(2.84), we find that $l_n^{b^{(q)}_{k_{(q)}}} = e^{\frac{\beta}{\gamma^{(q)}-2}(\gamma^{(q)}-2)^{k_{(q)}}}$. For $\gamma^{(q)} \in (2,3)$, the equality

$$x(\gamma^{(q)}-2)^{\frac{\log x}{|\log(\gamma^{(q)}-2)|}} = x \cdot x^{-1} = 1 \tag{2.96}$$

holds. By substituting $x = \log n$, we find that $n^{b^{(q)}_{k_{(q)}}} = e^{\frac{\beta}{\gamma^{(q)}-2}}$. Moreover, $a^{(q)}_k \to 1/(\gamma^{(q)}-3)$ from below as $k \to \infty$, such that

$$(C\log n)^{a^{(q)}_{k_{(q)}}} = (C\log n)^{1/(3-\gamma^{(q)})-o(1)}. \tag{2.97}$$

Substituting all these equations into (2.95), it follows that

$$u^{(q)}_{k_{(q)}} = e^{\frac{\beta}{\gamma^{(q)}-2}}(C\log n)^{1/(3-\gamma^{(q)})-o(1)}, \tag{2.98}$$

such that, for sufficiently large n, we can make $u^{(q)}_{k_{(q)}} \leq (\log n)^{\sigma^{(q)}}$. This proves that the number of (directed) steps needed to get from any $u \in \mathrm{Core}_n^{(\mathrm{out})}$ to any $v \in \mathrm{Core}_n^{(\mathrm{in})}$ is whp bounded above by

$$\frac{\log\log n}{|\log(\gamma^{(\mathrm{in})}-2)|} + \frac{\log\log n}{|\log(\gamma^{(\mathrm{out})}-2)|} + 1. \tag{2.99}$$

□

2.3.2 Typical distance to the outbound core and from the inbound core

The bound of the length of a path from vertices in the outbound core to vertices in the inbound core in Theorem 2.3.3 already closely resembles the typical distances in the directed configuration model from Theorem 1.5.4. Though Theorem 2.3.3 only provides a bound for the distance between high out-degree vertices to high in-degree vertices. To complete the upper bound for Theorem 1.5.4, we will show that both the distance of a typical vertex to $\mathrm{Core}_n^{(\mathrm{out})}$ and the distance from $\mathrm{Core}_n^{(\mathrm{in})}$ to a typical vertex can both be bounded by $c\log\log(n)$ for any arbitrary small $c > 0$. To do this, we shall make use of the approximation of the forward explored neighborhood and backward explored neighborhood to the branching processes described in Section 1.3.3. Under Assumption 1.5.3, both branching processes have an infinite mean which experience double exponential generational growth [7]. By viewing these neighborhoods as a single vertex with a large in- or out-degree, we will be able to apply Lemma 2.3.1 to prove the likely existence of double logarithmic paths to and from these cores.

Theorem 2.3.4. *Consider the directed configuration model for which Assumptions 1.3.2, 1.4.2 and 1.5.3 are satisfied. For all $\varepsilon > 0$, the length of the path of a typical vertex to $\mathrm{Core}_n^{(out)}$ can whp be bounded by*

$$\varepsilon \frac{\log\log n}{|\log(\tau^{(\mathrm{out})} - 2)|}, \tag{2.100}$$

and the path from $\mathrm{Core}_n^{(in)}$ to a typical vertex whp by

$$\varepsilon \frac{\log\log n}{|\log(\tau^{(\mathrm{in})} - 2)|}. \tag{2.101}$$

Consequently, the length of a typical the path is bounded from above by

$$(1+\varepsilon)\left(\frac{\log\log n}{|\log(\tau^{(\mathrm{out})} - 2)|} + \frac{\log\log n}{|\log(\tau^{(\mathrm{in})} - 2)|}\right). \tag{2.102}$$

Proof. For ease of notation, denote $G_n = \mathrm{DCM}_n(\boldsymbol{d})$ and fix $r \geq 1$. Remember that $H_r^{(n;\mathrm{out})}$ denotes the number of unpaired outbound half edges attached to vertices at distance $r-1$ from a forward explored typical vertex, and $H_r^{(n;\mathrm{in})}$ is similarly defined for the inbound half-edges of vertices with distance $r-1$ to a backward explored typical vertex. We have shown that under Assumption 1.3.2 that $(H_r^{(n;\mathrm{out})}, H_r^{(n;\mathrm{in})}) \xrightarrow{d} (Z_r^{(\mathrm{out})}, Z_r^{(\mathrm{in})})$ for all $r \geq 1$ fixed. Moreover, under the minimal degree Assumption 1.4.2, both $H_r^{(n;\mathrm{out})}$ and $H_r^{(n;\mathrm{in})}$ are whp very large for large values of r. To start with, we define for some small valued $\varepsilon_0 \in (0, 3-\gamma^{(q)})$ and $q \in \{\mathrm{out}, \mathrm{in}\}$

$$u_k^{(q)} = (H_r^{(n;q)})^{1/(\gamma^{(q)}-2+\varepsilon_0)^k}, \tag{2.103}$$

and define the layers

$$\tilde{\Gamma}_k^{(q)} := \{u : d_u^{(q)} \geq u_k^{(q)}\}. \tag{2.104}$$

Define the events

$$\begin{aligned} \mathcal{E}_k^{(\mathrm{out})} &= \{\exists u \in \tilde{\Gamma}_{k-1}^{(\mathrm{out})} : \mathcal{N}^{(\mathrm{out})}(u) \cap \tilde{\Gamma}_k^{(\mathrm{out})} = \emptyset\} \\ \mathcal{E}_k^{(\mathrm{in})} &= \{\exists u \in \tilde{\Gamma}_{k-1}^{(\mathrm{in})} : \mathcal{N}^{(\mathrm{in})}(u) \cap \tilde{\Gamma}_k^{(\mathrm{in})} = \emptyset\}. \end{aligned} \tag{2.105}$$

Similarly as we have shown in (2.89), the probability that any vertex with q-degree at least $u_{k-1}^{(q)}$ does not have a neighbor or is the neighbor of (depending on q) a vertex with q-degree at least $u_k^{(q)}$ is at most

$$e^{-c_2^{(q)}(u_{k-1}^{(q)})^{\varepsilon_0/(\gamma^{(q)}-2+\varepsilon_0)}}. \tag{2.106}$$

We iteratively construct the sequences $u_k^{(q)}$ process till we reach vertices that are in the outbound core for $q = \{\mathrm{out}\}$ and inbound core for $q = \{\mathrm{in}\}$, i.e., till

$$k_{*(q)} = \inf\{k : u_k^{(q)} \geq \log(n)^{1/(3-\gamma^{(q)})}\}. \tag{2.107}$$

Some rearranging using the expression of $u_k^{(q)}$ in (2.103) shows that

$$k_{*(q)} = \left\lceil \frac{\log\log\log n - \log\log(H_r^{(n;q)}) + \log(\frac{1}{3-\gamma^{(q)}})}{|\log(\gamma^{(q)} - 2 + \varepsilon_0)|} \right\rceil. \tag{2.108}$$

Conditioning on the values of $Z_r^{(n;\text{out})}$ and $Z_r^{(n;\text{in})}$, we apply the union bound to find

$$\begin{aligned}\mathbb{P}\left(\bigcup_{k=1}^{k_{*(out)}} \mathcal{E}_k^{(\text{out})} | H_r^{(n;\text{out})}\right) &\leq \sum_{k=1}^{k_{*(out)}} e^{-c_2^{(out)}(u_{k-1}^{(out)})^{\varepsilon_0/(\gamma^{(out)}-2+\varepsilon_0)}} \\ \mathbb{P}\left(\bigcup_{k=1}^{k_{*(in)}} \mathcal{E}_k^{(\text{in})} | H_r^{(n;\text{in})}\right) &\leq \sum_{k=1}^{k_{*(in)}} e^{-c_2^{(in)}(u_{k-1}^{(in)})^{\varepsilon_0/(\gamma^{(in)}-2+\varepsilon_0)}}\end{aligned} \tag{2.109}$$

As $(\gamma^{(q)} - 2 + \varepsilon_0) \in (0,1)$, under Assumption 1.4.2 it follows from Corollary 1.3.9, that for r large enough that the limit of this bound is

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=1}^{k_{*(q)}} \mathcal{E}_k^{(q)} | H_r^{(n;q)}\right) \leq \sum_{k=1}^{\infty} e^{-c_2^{(q)}(Z_r^{(q)})^{\varepsilon_0/(\gamma^{(q)}-2+\varepsilon_0)^k}}. \tag{2.110}$$

Under Assumptions 1.4.2 + 1.5.3 the branching processes $(Z_r^{(q)})_{r\geq 1}$ are ensured to survive, such that the random variables $Z_r^{(q)} \xrightarrow{\mathbb{P}} \infty$ as $r \to \infty$. Consequently, the bound on the RHS of (2.110) converges whp to zero, and with it the probability of the events $\mathcal{E}_{k_1}^{(out)}$ and $\mathcal{E}_{k_2}^{(in)}$ happening any for $k_1 \in [k_{*(out)}]$ and $k_2 \in [k_{*(in)}]$. Given that both events do not hold, for any fixed $r \geq 1$ and $\varepsilon > 0$ we can take n large enough such that whp

$$k_{*(q)} + r \leq \varepsilon \frac{\log\log n}{|\log(\gamma^{(q)} - 2)|}. \tag{2.111}$$

To obtain a bound on the typical distances we add the bound for the distance between the out- an inbound cores from Theorem 2.3.3 to obtain that whp

$$\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(o_1, o_2) \leq (1+\varepsilon)\left(\frac{\log\log n}{|\log(\gamma^{(\text{in})} - 2)|} + \frac{\log\log n}{|\log(\gamma^{(\text{out})} - 2)|}\right). \tag{2.112}$$

Note that (2.112) holds any arbitrary small value of $\delta^{(q)} > 0$ in $\gamma^{(q)} = (\tau^{(q)} + \delta^{(q)})$, for $q \in \{\text{out}, \text{in}\}$. As the mapping $x \mapsto 1/|\log(x)|$ is strictly increasing for $x > 0$, we can take $\delta^{(\text{out})}, \delta^{(\text{in})} > 0$ arbitrary small, such that the smallest upper bound for the typical distances becomes

$$\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(o_1, o_2) \leq (1+\varepsilon)\left(\frac{\log\log n}{|\log(\tau^{(\text{in})} - 2)|} + \frac{\log\log n}{|\log(\tau^{(\text{out})} - 2)|}\right), \tag{2.113}$$

completing the proof of the claim. □

2.4 Lower bound typical distances directed configuration model

This section we will provide the lower bound for the typical distances in Theorem 1.5.4. To do this, we follow the steps of the proof from ([14], Theorem 7.8), summarized in Section 2.1.4, which shows a lower bound of the typical distances in the configuration model. The structure of the proof is as follows. By applying path counting techniques to obtain bounds on the expected number of paths of a given length, we are able to apply Markov's inequality to bound the probability of the paths existence. Thought to obtain meaningful bounds, the paths need to be divided, by truncating the increasing out- or decreasing in-degrees along the paths. For one, this will create the set of paths which experience an even larger exponential difference between the out- or in-degree for some subsequent vertices along the path, than the path constructed for the upper bound in Theorem 2.3.4. We will show that even under the assumption that the size biased out- and in-degree distributions follow a power law, such differences are unlikely to occur on a path shorter then Theorem 1.5.4. The remaining paths contain relatively average out-degree vertices near the starting vertex and in-degree vertices near the target vertex. For these paths we will show it is likely to reach another typical vertex in fewer steps than than 1.5.4. The main difference between the undirected setting that we need to consider is the asymmetry between the number of paths from the starting vertex, and towards the target vertex, caused by the different distributions of the in- and out-degree.

2.4.1 Path counting

To start the study of the paths, we start by providing a formal definition.

Definition 2.4.1. (Directed paths) *The set of directed paths from vertex u to v of length k is denoted as $\mathcal{P}_k(u,v)$, containing sequences*

$$\vec{\pi} = \{(\pi_0, t_0), (\pi_1, s_1, t_1), \ldots, (\pi_{k-1}, s_{k-1}, t_{k-1}), (\pi_k, s_k)\}, \tag{2.114}$$

for which π_i denotes the i'th vertex along the path, and t_i the label of the outbound half-edge attached the vertex π_i, that is paired with the inbound half-edge with label s_{i+1}, attached to vertex π_{i+1}. These paths are self-avoiding, meaning that $\pi_i \neq \pi_j$, $\forall i \neq j$.

When half-edges are paired during the construction process, we say that the directed edge they formed is *occupied*. Hence, a path $\vec{\pi} \in \mathcal{P}_k(u,v)$ is occupied if this is the case for all k directed edges, for which the probability equals

$$\mathbb{P}(\vec{\pi} \text{ occupied}) = \prod_{i=1}^{k} \frac{1}{l_n - i + 1}. \tag{2.115}$$

To find the number of ways to construct a path using fixed distinct vertices $\pi_0, \pi_1, \ldots, \pi_k$, note that we can exit the first vertex from $d_{\pi_0}^{(\text{out})}$ outbound half-edges, enter the final vertex from $d_{\pi_k}^{(\text{in})}$ inbound half-edges, and traverse vertex $i \in [1, k-1]$ along the path in $d_{\pi_i}^{(\text{out})} d_{\pi_i}^{(\text{in})}$ ways, such that the number equals

$$d_{\pi_0}^{(\text{out})} \prod_{i=1}^{k-1} \left(d_{\pi_i}^{(\text{out})} d_{\pi_i}^{(\text{in})} \right) d_{\pi_k}^{(\text{in})}. \tag{2.116}$$

Now let us look at the probability distribution of the typical distances of the directed configuration model. First, we take

$$\begin{aligned} \mathbb{P}\left(\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(o_1, o_2) \leq k_n\right) &= \frac{1}{n^2} \sum_{u,v \in [n]} \mathbb{P}\left(\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(u, v) \leq k_n\right) \\ &= \frac{1}{n} + \frac{1}{n^2} \sum_{\substack{u,v \in [n] \\ u \neq v}} \sum_{k=1}^{k_n} \mathbb{P}\left(\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(u, v) = k\right), \end{aligned} \tag{2.117}$$

where the second equality follows as $\mathbb{P}\left(\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(u, v) = 0\right) = 1$ only for all $u = v \in [n]$ and has probability zero otherwise. Defining $\nu_n = \sum_{u \in [n]} \frac{d_u^{(\text{in})} d_u^{(\text{out})}}{l_n}$, use (2.115), (2.116) and the union bound to find the bound for all $k \geq 1$

$$\begin{aligned} \mathbb{P}\left(\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(u, v) = k\right) &\leq \mathbb{P}(\vec{\pi} \in \mathcal{P}_k(u,v) : \vec{\pi} \text{ occupied}) \\ &\leq \sum_{\pi \in \mathcal{P}_k(u,v)} \mathbb{P}(\vec{\pi} \text{ occupied}) \\ &\leq \frac{d_u^{(\text{out})} d_v^{(\text{in})}}{l_n - k + 1} \sum_{\pi_1, \ldots, \pi_{k-1}}^{*} \prod_{i=1}^{k-1} \frac{d_{\pi_i}^{(\text{out})} d_{\pi_i}^{(\text{in})}}{l_n - i + 1} \\ &\leq \frac{d_u^{(\text{out})} d_v^{(\text{in})}}{l_n} \frac{l_n^k (l_n - k)!}{l_n!} \nu_n^{k-1}, \end{aligned} \tag{2.118}$$

where the sum after the third inequality denotes the sum over all combinations of distinct vertices $\pi_1, ..., \pi_{k-1}$. The final inequality follows as we also include the paths that are not self-avoiding, increasing the number of paths that are available. Under the power law Assumption 1.5.3, the value of ν_n grows too rapidly to provide a meaningful bound to the path probabilities. This is caused by the large amount of vertices with a high in- or out-degree. For the upper bound we used this phenomenon to create a path to and from the in- and outbound hubs, but these paths are very rare. The bound (2.118) ignores the fact that it is unlikely that a typical vertex is connected with such a significantly higher out- or lower in- degree vertex. This issue is resolved by dividing the paths in two different routes as explained at the start of this section. Next, we introduce the truncation argument of the out- or in-degrees to create this division and provide bounds for both path probabilities.

2.4.2 Good and bad paths

As shown in Section 2.3, the odds of finding of finding higher out-degree vertices increases, the further we walk from the typical starting vertex. Similarly, for the in-degree, the further we look at vertices along the path towards the typical target vertex. Hence, to truncate the respective in- or out-degrees of paths we fix two *increasing* sequences of numbers $(b_l^{(\text{in})})_{l=0}^{\infty}$ and $(b_l^{(\text{out})})_{l=0}^{\infty}$. As the distribution of the in- and out degrees are not identical, the number of vertices for which we truncate the out-degree is possibly not the same as the number of vertices for which we truncate the in-degree, along the same path. So let s denote the fraction of vertices along the path for which we truncate the out-degree, and therefore $(1-s)$ the fraction of vertices for which the in-degree is truncated.

We start by defining the paths that do not contain the large differences of in- or out-degrees between subsequent vertices. We say $\vec{\pi} \in \mathcal{P}_k(u,v)$ is a *good path* if $d_{\pi_l}^{(\text{out})} \leq b_l^{(\text{out})}$ for all $1 \leq l \leq \lfloor k \cdot s \rfloor$ and $d_{\pi_l}^{(\text{in})} \leq b_{k-l}^{(\text{in})}$ all $\lceil k \cdot s \rceil \leq l \leq k$. Note the truncation value for the in-degree increases the further we move from vertex π_k. We denote the set of good paths of length k as

$$\mathcal{GP}_k(u,v) := \{\vec{\pi} \in \mathcal{P}_k(u,v) : d_{\pi_i}^{(\text{out})} \leq b_i^{(\text{out})}, d_{\pi_j}^{(\text{in})} \leq b_{k-j}^{(\text{in})}, 1 \leq i \leq \lfloor k \cdot s \rfloor < j \leq k\}. \tag{2.119}$$

The remaining paths $\vec{\pi} \in \mathcal{P}_k(u,v) \backslash \mathcal{GP}_k(u,v)$ containing exceptional large differences in the in- or out-degree between subsequent vertices, are defined as *bad paths*. From the definition it follows that a bad path contains a vertex with either $d_{\pi_l}^{(\text{out})} > b_l^{(\text{out})}$ for some $1 \leq l \leq \lfloor ks \rfloor$, or $d_{\pi_l}^{(\text{in})} > b_{k-l}^{(\text{in})}$, for some $\lceil ks \rceil \leq l \leq k$. Denote the set of all paths of length k starting from vertex u as $\mathcal{P}_k^{(\text{out})}(u) = \bigcup_{v \in [n]} \mathcal{P}_k(u,v)$ and the set of all paths ending at vertex v as $\mathcal{P}_k^{(\text{in})}(v) = \bigcup_{u \in [n]} \mathcal{P}_k(u,v)$. We denote their respective subset of bad paths as

$$\begin{aligned} \mathcal{BP}_k^{(\text{out})}(u) &= \{\vec{\pi} \in \mathcal{P}_k^{(\text{out})}(u) : d_{\pi_k}^{(\text{out})} > b_k^{(\text{out})}, d_{\pi_s}^{(\text{out})} \leq b_s^{(\text{out})}, \forall s < k\} \\ \mathcal{BP}_k^{(\text{in})}(v) &= \{\vec{\pi} \in \mathcal{P}_k^{(\text{in})}(v) : d_{\pi_0}^{(\text{in})} > b_k^{(\text{in})}, d_{\pi_{k-s}}^{(\text{in})} \leq b_s^{(\text{in})}, \forall s < k\}. \end{aligned} \tag{2.120}$$

Moreover, denote the event that such a path is occupied as

$$\begin{aligned} \mathcal{F}_k^{(\text{out})}(u) &= \{\exists \vec{\pi} \in \mathcal{BP}_k^{(\text{out})}(u) : \vec{\pi} \text{ occupied}\} \\ \mathcal{F}_k^{(\text{in})}(u) &= \{\exists \vec{\pi} \in \mathcal{BP}_k^{(\text{in})}(u) : \vec{\pi} \text{ occupied}\}. \end{aligned} \tag{2.121}$$

Note that if $\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(u,v) \leq k_n$ it follows that there exists either a good or a bad path from vertex u to v with length equal or less than k_n. The event that a bad path exists is a subset of the event either some $\vec{\pi} \in \mathcal{BP}_k^{(\text{out})}(u)$ *or* $\vec{\pi} \in \mathcal{BP}_l^{(\text{in})}(v)$ for $k \leq \lfloor k_n \cdot s \rfloor$ and $l \leq \lceil k_n \cdot (1-s) \rceil$. This is why we can bound the event

$$\{\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(u,v) \leq k_n\} \subseteq \bigcup_{k \leq k_n} \left(\mathcal{E}_k(u,v) \cup \mathcal{F}_{\lfloor k \cdot s \rfloor}^{(\text{out})}(u) \cup \mathcal{F}_{\lceil k \cdot (1-s) \rceil}^{(\text{in})}(v)\right). \tag{2.122}$$

Applying Boole's inequality results in the bound

$$\mathbb{P}\left(\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(u,v) \le k_n\right) \le \sum_{k_0=0}^{k_n} \mathbb{P}(\mathcal{E}_{k_0}(u,v)) + \sum_{k_1=0}^{\lfloor k_n s\rfloor} \mathbb{P}(\mathcal{F}^{(\mathrm{out})}_{k_1}(u)) + \sum_{k_2=0}^{\lceil k_n(1-s)\rceil} \mathbb{P}(\mathcal{F}^{(\mathrm{in})}_{k_2}(v)). \quad (2.123)$$

Moreover, applying (2.117) on (2.123) further gives

$$\begin{aligned}\mathbb{P}\left(\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1,o_2) \le k_n\right) \le & \frac{1}{n} + \frac{1}{n}\sum_{u\in[n]}\sum_{k_1=1}^{\lfloor k_n s\rfloor} \mathbb{P}(\mathcal{F}^{(\mathrm{out})}_{k_1}(u)) + \sum_{k_2=1}^{\lceil k_n(1-s)\rceil} \mathbb{P}(\mathcal{F}^{(\mathrm{in})}_{k_2}(u)) \\ & + \frac{1}{n^2}\sum_{\substack{u,v\in[n]\\ u\neq v}}\sum_{k_0=1}^{k_n} \mathbb{P}(\mathcal{E}_{k_0}(u,v)).\end{aligned} \quad (2.124)$$

Now let us look at each of these events separately. For ease of notation, denote the restriction of ν_n with respect to the out- or in-degree, for $b>0$ as

$$\begin{aligned}\nu_n^{(\mathrm{out})}(b) &= \frac{1}{l_n}\sum_{u\in[n]} d_u^{(\mathrm{in})} d_u^{(\mathrm{out})} \mathbb{1}\{d_u^{(\mathrm{out})} \le b\} \\ \nu_n^{(\mathrm{in})}(b) &= \frac{1}{l_n}\sum_{u\in[n]} d_u^{(\mathrm{in})} d_u^{(\mathrm{out})} \mathbb{1}\{d_u^{(\mathrm{in})} \le b\}.\end{aligned} \quad (2.125)$$

The following lemma provides a bound on the good and bad path probabilities, that depends on the forward and backward tail distribution, and these restrictions of ν_n.

Lemma 2.4.2. (Truncated directed path probabilities) *For every $k \ge 1$, $(b_l^{(out)})_{l\ge 0}, (b_l^{(in)})_{l\ge 0}$ with $b_l^{(out)}, b_l^{(in)}, \ge 0$ and $l \mapsto b_l^{(out)}, b_l^{(in)}$ non-decreasing,*

$$\begin{aligned}\mathbb{P}(\mathcal{F}_k^{(\mathrm{out})}(u)) &\le d_u^{(\mathrm{out})} e^{k^2/l_n}[1-F_n^{(\mathrm{out})}](b_k^{(\mathrm{out})}) \times \prod_{l=1}^{k-1} \nu_n^{(\mathrm{out})}(b_l^{(\mathrm{out})}) \\ \mathbb{P}(\mathcal{F}_k^{(\mathrm{in})}(v)) &\le d_v^{(\mathrm{in})} e^{k^2/l_n}[1-F_n^{(\mathrm{in})}](b_k^{(\mathrm{in})}) \times \prod_{l=1}^{k-1} \nu_n^{(\mathrm{in})}(b_l^{(\mathrm{in})}).\end{aligned} \quad (2.126)$$

and

$$\mathbb{P}(\mathcal{E}_k(u,v)) \le \frac{d_u^{(\mathrm{out})} d_v^{(\mathrm{in})}}{l_n} e^{k^2/l_n} \prod_{i=1}^{ks} \nu_n^{(\mathrm{out})}(b_i^{(\mathrm{out})}) \prod_{j=1}^{k(1-s)} \nu_n^{(\mathrm{in})}(b_j^{(\mathrm{in})}). \quad (2.127)$$

Proof. We start by proving the upper bound for the bad path probability in (2.126). Take $q,p \in \{\mathrm{out},\mathrm{in}\}$, with $q \neq p$. Using a similar approach as that in (2.118) restricted to the bad paths, we obtain

$$\begin{aligned}\mathbb{P}(\mathcal{F}_k^{(q)}(u)) &\le d_u^{(q)} \sum_{\pi_k: d_{\pi_k}^{(q)} > b_k^{(q)}} \frac{d_{\pi_k}^{(p)}}{l_n - k + 1} \times \prod_{i=1}^{k-1} \sum_{\pi_i: d_{\pi_i}^{(q)} \le b_i^{(q)}} \frac{d_{\pi_i}^{(q)} d_{\pi_i}^{(p)}}{l_n - i + 1} \\ &= d_u^{(q)} \frac{l_n^k (l_n-k)!}{l_n!} \sum_{\pi_k: d_{\pi_k}^{(q)} > b_k^{(q)}} \frac{d_{\pi_k}^{(p)}}{l_n} \times \prod_{i=1}^{k-1} \sum_{\pi_i: d_{\pi_i}^{(q)} \le b_i^{(q)}} \frac{d_{\pi_i}^{(q)} d_{\pi_i}^{(p)}}{l_n} \\ &\le d_u^{(q)} e^{k^2/l_n}[1-F_n^{(q)}](b_k^{(q)}) \prod_{i=1}^{k-1} \nu_n^{(q)}(b_i^{(q)}),\end{aligned} \quad (2.128)$$

where e^{k^2/l_n} is a bound on the factor $\frac{l_n^k(l_n-k)!}{l_n!}$. Following the same approach, we can find the upper bound (2.127) for the probability that there is a good path from vertex u to v.

$$\begin{aligned}\mathbb{P}(\mathcal{E}_k(u,v)) \leq & \sum_{\pi\in\mathcal{GP}_k(u,v)} \frac{d_u^{(\text{out})}d_v^{(\text{in})}}{l_n-k+1}\prod_{i=1}^{k-1}\frac{d_{\pi_i}^{(\text{out})}d_{\pi_i}^{(\text{in})}}{l_n-i+1} \\ \leq & \frac{d_u^{(\text{out})}d_v^{(\text{in})}}{l_n}\frac{l_n^k(l_n-k)!}{l_n!}\prod_{i=1}^{\lceil ks\rceil}\sum_{\pi_i:d_{\pi_i}^{(\text{out})}\leq b_i^{(\text{out})}}\frac{d_{\pi_i}^{(\text{out})}d_{\pi_i}^{(\text{in})}}{l_n} \\ & \times\prod_{j=\lceil ks\rceil+1}^{k-1}\sum_{\pi_j:d_{\pi_j}^{(\text{in})}\leq b_{k-j}^{(\text{in})}}\frac{d_{\pi_j}^{(\text{out})}d_{\pi_j}^{(\text{in})}}{l_n} \\ \leq & \frac{d_u^{(\text{out})}d_v^{(\text{in})}}{l_n}e^{k^2/l_n}\prod_{i=1}^{\lfloor ks\rfloor}\nu_n^{(\text{out})}(b_i^{(\text{out})})\prod_{j=1}^{\lceil k(1-s)\rceil}\nu_n^{(\text{in})}(b_j^{(\text{in})}).\end{aligned} \tag{2.129}$$

□

2.4.3 Proof of lower bound of typical distances

For the remainder of this section we will make numerous use of the upper bound given by Potter's Theorem 1.5.2, hence we take

$$\begin{aligned}\gamma^{(\text{out})} &= \gamma^{(\text{out})}(\delta^{(\text{out})}) = \tau^{(\text{out})} - \delta^{(\text{out})} \\ \gamma^{(\text{in})} &= \gamma^{(\text{in})}(\delta^{(\text{in})}) = \tau^{(\text{in})} - \delta^{(\text{in})},\end{aligned} \tag{2.130}$$

for ease of notation. Note that all the bounds in Lemma 2.4.2 depend on $\nu_n^{(\text{out})}$ and $\nu_n^{(\text{in})}$. With the following lemma we further bound these objects:

Lemma 2.4.3. *Given that Assumption 1.5.3 is satisfied, there are some* $c_2^{(out)}, c_2^{(in)}$ *such that*

$$\begin{aligned}\nu_n^{(\text{out})}(b) &\leq c_2^{(\text{out})}b^{3-\gamma^{(\text{out})}} \\ \nu_n^{(\text{in})}(b) &\leq c_2^{(\text{in})}b^{3-\gamma^{(\text{in})}}.\end{aligned} \tag{2.131}$$

Proof. Let $p, q \in \{\text{out}, \text{in}\}$ with $p \neq q$, then

$$\begin{aligned}\nu_n^{(q)}(b) &= \frac{1}{l_n}\sum_{u\in[n]} d_u^{(p)}d_u^{(q)}\mathbb{1}\{d_u^{(q)}\leq b\} \\ &= \frac{1}{l_n}\sum_{x=1}^{\infty} x\sum_{u\in[n]} d_u^{(p)}\mathbb{1}\{d_u^{(q)}=x\}\mathbb{1}\{d_u^{(q)}\leq b\} \\ &= \frac{1}{l_n}\sum_{x=1}^{b}\sum_{k=1}^{x}\sum_{u\in[n]} d_u^{(p)}\mathbb{1}\{d_u^{(q)}=x\} \\ &= \frac{1}{l_n}\sum_{k=1}^{b}\sum_{x=k}^{b}\sum_{u\in[n]} d_u^{(p)}\mathbb{1}\{d_u^{(q)}=x\} \\ &= \sum_{k=1}^{b}[F_n^{(q)}(b) - F_n^{(q)}(k)] \leq \sum_{k=1}^{b}[1-F_n^{(q)}](k) \\ &\leq \sum_{k=1}^{b} c_2^{(q)}k^{-(\gamma^{(q)}-2)} \leq c_2^{(q)}b^{3-\gamma^{(q)}}.\end{aligned} \tag{2.132}$$

□

It remains to define the values of the truncation sequences and the ratio of the path for which the out-degree of the vertices are truncated. First, for the ratio we take

$$\begin{aligned} s &= \frac{1/|\log(\gamma^{(\mathrm{out})}-2)|}{1/|\log(\gamma^{(\mathrm{out})}-2)|+1/|\log(\gamma^{(\mathrm{in})}-2)|} \\ \Rightarrow (1-s) &= \frac{1/|\log(\gamma^{(\mathrm{in})}-2)|}{1/|\log(\gamma^{(\mathrm{out})}-2)|+1/|\log(\gamma^{(\mathrm{in})}-2)|}. \end{aligned} \tag{2.133}$$

For the truncation values, we refer to the step sizes of the path (2.103) from the upper bound. Here, some arbitrary small $\varepsilon_0 > 0$, the out-degree increased by powers of $1/(\tau^{(\mathrm{out})}-2+\varepsilon_0)^k$ and decreased the in-degree by the power $1/(\tau^{(\mathrm{in})}-2+\varepsilon_0)^k$ each step, as we moved away from the starting and target vertex. For our bad paths, we increase the step size by taking some arbitrary small value $\delta^{(q)} \in (\gamma^{(q)}-2)$, and define

$$a_{(q)} = 1/(\gamma^{(q)}-2-\delta^{(q)}). \tag{2.134}$$

Take $b_0^{(q)} = e^{A^{(q)}}$ for some constant $A^{(q)} > 0$ sufficiently large, and we recursively define both $(b_l^{(q)})_{l\geq 0}$ by

$$b_l^{(q)} = (b_{l-1}^{(q)})^{a_{(q)}} = (b_0^{(q)})^{a_{(q)}^l} = e^{A^{(q)}(\gamma^{(q)}-2-\delta^{(q)})^{-l}}. \tag{2.135}$$

With the following theorem we will complete the proof of the lower bound of Theorem 1.5.4, by applying the bounds we shown this section and show these become asymptotically small as n grows large.

Theorem 2.4.4. *Suppose that the in- and out-degrees of* $\mathrm{DCM}_n(\boldsymbol{d})$ *satisfy Assumption 1.3.2, and the upper bound of Assumption 1.5.3 holds for the tail forward out-degree and backward in-degree distribution. Then for every* $\varepsilon > 0$,

$$\mathbb{P}\left(\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1,o_2) \leq (1-\varepsilon)(\frac{\log\log n}{|\log(\tau^{(\mathrm{out})}-2)|}+\frac{\log\log n}{|\log(\tau^{(\mathrm{in})}-2)|})\right) = o(1). \tag{2.136}$$

Proof. First, we take

$$k_n = (1-\varepsilon)\left(\frac{\log\log n}{|\log(\gamma^{(\mathrm{out})}-2)|}+\frac{\log\log n}{|\log(\gamma^{(\mathrm{in})}-2)|}\right), \tag{2.137}$$

and remember that we showed from (2.138) we have the bound

$$\begin{aligned} \mathbb{P}\left(\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(o_1,o_2) \leq k_n\right) \leq & \frac{1}{n} + \frac{1}{n}\sum_{u\in[n]}\sum_{k_1=1}^{\lfloor k_n s\rfloor}\mathbb{P}(\mathcal{F}_{k_1}^{(\mathrm{out})}(u)) + \sum_{k_2=1}^{\lceil k_n(1-s)\rceil}\mathbb{P}(\mathcal{F}_{k_2}^{(\mathrm{in})}(u)) \\ & + \frac{1}{n^2}\sum_{\substack{u,v\in[n] \\ u\neq v}}\sum_{k_0=1}^{k_n}\mathbb{P}(\mathcal{E}_{k_0}(u,v)). \end{aligned} \tag{2.138}$$

As the first step, we note the bound of each of the event probabilities in (2.138) that we have shown in (2.4.2) contains the factor consisting of the product of $v_n^{(q)}(b_l)$, for $q \in \{\mathrm{out},\mathrm{in}\}$. Applying the bound from Lemma 2.4.3, it follows that

$$\begin{aligned} \prod_{i=1}^{k-1}\nu_n^{(q)}(b_l^{(q)}) &\leq \prod_{i=1}^{k-1}c_2^{(q)}(b_l^{(q)})^{3-\gamma^{(q)}} = (c_2^{(q)})^{k-1}e^{A^{(q)}(3-\gamma^{(q)})\sum_{l=1}^{k-1}a_{(q)}^l} \\ &\leq (c_2^{(q)})^{k-1}e^{A^{(q)}(3-\gamma^{(q)})a_{(q)}^k/(a_{(q)}-1)} = (c_2^{(q)})^{k-1}(b_k^{(q)})^{(3-\gamma^{(q)})/(a_{(q)}-1)}. \end{aligned} \tag{2.139}$$

Bad paths We will first apply this bound to show we can bound the probability that a bad path exists by an arbitrary small value. Applying the event probability bound from Lemma 2.4.2 in addition with (2.139) and the upper bound from Assumption 1.5.3, we obtain

$$\begin{aligned}\mathbb{P}(\mathcal{F}_k^{(q)}(u)) &\leq d_u^{(q)} e^{k/l_n} [1 - F_n^{(q)}](b_k^{(q)}) \times \prod_{l=1}^{k-1} \nu_n^{(q)}(b_l^{(q)}) \\ &\leq d_u^{(q)} e^{k/l_n} (c_2^{(q)})^k (b_k^{(q)})^{-(\gamma^{(q)}-2)+(3-\gamma^{(q)})/(a_{(q)}-1)}.\end{aligned} \tag{2.140}$$

With some rewriting it can be shown that

$$(\gamma^{(q)} - 2) - (3 - \gamma^{(q)})/(a_{(q)} - 1) = \delta^{(q)}/(3 - \gamma^{(q)} + \delta^{(q)}) > \delta^{(q)} > 0. \tag{2.141}$$

Applying this bound on the exponent of $b_k^{(q)}$ in (2.140) gives

$$\mathbb{P}(\mathcal{F}_k^{(q)}(u)) \leq d_u^{(q)} (e^{1/l_n} c_2^{(q)})^k (b_k^{(q)})^{-\delta^{(q)}}. \tag{2.142}$$

Taking $k_n^{(\text{out})} = \lfloor k_n s \rfloor$ and $k_n^{(\text{in})} = \lceil k_n(1-s) \rceil$, we use this bound to find that for all arbitrary small $\delta^{(q)} > 0$,

$$\begin{aligned}\frac{1}{n} \sum_{u \in [n]} \sum_{k=1}^{k_n^{(q)}} \mathbb{P}(\mathcal{F}_k^{(q)}(u)) &\leq \frac{1}{n} \sum_{u \in [n]} \sum_{k=1}^{k_n^{(q)}} d_u^{(q)} (e^{1/l_n} c_2^{(q)})^k (b_k^{(q)})^{-\delta^{(q)}} \\ &= \frac{l_n}{n} \sum_{k=1}^{k_n^{(q)}} (e^{1/l_n} c_2^{(q)})^k (b_k^{(q)})^{-\delta^{(q)}}.\end{aligned} \tag{2.143}$$

To finish the proof for the bad paths, we note that from the definition in (2.135) that

$$(b_k^{(q)})^{-\delta^{(q)}} = e^{-A^{(q)}\delta^{(q)}(\gamma^{(q)}-2-\delta^{(q)})^{-k}}. \tag{2.144}$$

As $(\gamma^{(q)} - 2 - \delta^{(q)}) > 1$, and $l_n/n = O(1)$, it is possible for all $\varepsilon_0 > 0$ to take $A^{(q)}(\delta^{(q)}, \varepsilon_0)$ large enough, such that

$$\frac{1}{n} \sum_{u \in [n]} \sum_{k=1}^{k_n^{(q)}} \mathbb{P}(\mathcal{F}_k^{(q)}(u)) \leq \frac{l_n}{n} \sum_{k=1}^{\infty} (e^{1/l_n} c_2^{(q)})^k (b_k^{(q)})^{-\delta^{(q)}} < \varepsilon_0. \tag{2.145}$$

This proves that

$$\frac{1}{n} \sum_{u \in [n]} \sum_{k_1=1}^{\lfloor k_n s \rfloor} \mathbb{P}(\mathcal{F}_{k_1}^{(\text{out})}(u)) + \sum_{k_2=1}^{\lceil k_n(1-s) \rceil} \mathbb{P}(\mathcal{F}_{k_2}^{(\text{in})}(u)) \xrightarrow{n \to \infty} 0. \tag{2.146}$$

As this itself shows that the probability that a bad path exists between two typical vertices of length k_n, given in (2.137), converges to zero as $n \to \infty$.

Good paths: It now remains to show that the probability of the existence of a good path with length at most k_n can also be bounded by an arbitrary small value in the limit of n. We start with the bound from Lemma 2.4.2, which showed that

$$\mathbb{P}(\mathcal{E}_k(u, v) \leq \frac{d_u^{(\text{out})} d_v^{(\text{in})}}{l_n} e^{k/l_n} \prod_{i=1}^{\lfloor ks \rfloor} \nu_n^{(\text{out})}(b_i^{(\text{out})}) \prod_{j=1}^{\lceil k(1-s) \rceil} \nu_n^{(\text{in})}(b_j^{(\text{in})}). \tag{2.147}$$

Applying the bound (2.139) on the product factors of (2.147) further gives

$$\mathbb{P}(\mathcal{E}_k(u,v)) \leq \frac{d_u^{(\text{out})}d_v^{(\text{in})}}{l_n} e^{k/l_n}(c_2^{(\text{out})})^{\lfloor k\cdot s\rfloor}(b^{(\text{out})}_{\lfloor k\cdot s\rfloor})^{(3-\gamma^{(\text{out})})/(a_{(\text{out})}-1)} \times (c_2^{(\text{in})})^{\lceil k(1-s)\rceil}(b^{(\text{in})}_{\lceil k(1-s)\rceil})^{(3-\gamma^{(\text{in})})/(a_{(\text{in})}-1)}. \tag{2.148}$$

Applying (2.148) on the summation of good path event probabilities from (2.138) gives

$$\begin{aligned}\frac{1}{n^2}\sum_{u,v\in[n]}\sum_{k=1}^{k_n}\mathbb{P}(\mathcal{E}_k(u,v)) \leq& \frac{1}{n^2}\sum_{u,v\in[n]}\frac{d_u^{(\text{out})}d_v^{(\text{in})}}{l_n}\sum_{k=1}^{k_n} e^{k/l_n}(c_2^{(\text{out})})^{\lfloor k\cdot s\rfloor}(b^{(\text{out})}_{\lfloor k\cdot s\rfloor})^{(3-\gamma^{(\text{out})})/(a^{(\text{out})}-1)}\\ &\times (c_2^{(\text{in})})^{\lceil k(1-s)\rceil}(b^{(\text{in})}_{\lceil k(1-s)\rceil})^{(3-\gamma^{(\text{in})})/(a^{(\text{in})}-1)}\\ \leq& \frac{l_n}{n^2}k_n e^{k_n/l_n}(c_2^{(\text{out})})^{\lfloor k_n\cdot s\rfloor}(b^{(\text{out})}_{\lfloor k_n\cdot s\rfloor})^{(3-\gamma^{(\text{out})})/(a^{(\text{out})}-1)}\\ &\times (c_2^{(\text{in})})^{\lceil k_n(1-s)\rceil}(b^{(\text{in})}_{\lceil k_n(1-s)\rceil})^{(3-\gamma^{(\text{in})})/(a^{(\text{in})}-1)}.\end{aligned} \tag{2.149}$$

From the definitions of the ratio value s in (2.133) and the number of steps k_n in (2.137) we get

$$\begin{aligned}k_n s &= (1-\varepsilon)\frac{\log\log n}{|\log(\gamma^{(\text{out})}-2)|}\\ k_n(1-s) &= (1-\varepsilon)\frac{\log\log n}{|\log(\gamma^{(\text{in})}-2)|}.\end{aligned} \tag{2.150}$$

Hence, we can take $\delta^{(\text{out})}, \delta^{(\text{in})}$ so small that both

$$\begin{aligned}(\gamma^{(\text{out})}-2-\delta_1^{(\text{out})})^{-k_n * s} &\leq (\log n)^{1-\varepsilon/4}\\ (\gamma^{(\text{in})}-2-\delta_1^{(\text{in})})^{-(k_n*(1-s)+1)} &\leq (\log n)^{1-\varepsilon/4}.\end{aligned} \tag{2.151}$$

From the expression of $(b_l^{(q)})_{l\geq 0}$ in (2.135) and the bound (2.151) we get for both $q \in \{\text{out}, \text{in}\}$

$$\begin{aligned}(b^{(q)}_{\lceil k_n^{(q)}\rceil})^{(3-\gamma^{(q)})/(a_{(q)}-1)} &\leq \exp\left\{A^{(q)}\frac{3-\gamma^{(q)}}{a_{(q)}-1}(\log n)^{1-\varepsilon/4}\right\}\\ &\leq \exp\left\{A^{(q)}(3-\gamma^{(q)})(\log n)^{1-\varepsilon/4}\right\}.\end{aligned} \tag{2.152}$$

Take $c = \max\{c_2^{(\text{out})}, c_2^{(\text{in})}\}$, such that $(c_2^{(\text{in})})^{\lceil k_n(1-s)\rceil}\cdot(c_2^{(\text{out})})^{\lfloor k_n\cdot s\rfloor} \leq c^{k_n+1}$. Since we take $k_n = O(\log\log n)$, we have $l_n/n^2 = \Theta(1/n)$ and $e^{k_n/l_n} = 1+o(1)$. we can conclude that

$$\begin{aligned}\frac{1}{n^2}\sum_{u,v\in[n]}\sum_{k=1}^{k_n}\mathbb{P}(\mathcal{E}_k(u,v)) &\leq \frac{l}{n^2}k_n c^{k_n+1}\exp\left\{(A^{(\text{out})}(3-\gamma^{(\text{out})})+(A^{(\text{in})}(3-\gamma^{(\text{in})})(\log n)^{1-\varepsilon/4}\right\}\\ &= o(1).\end{aligned} \tag{2.153}$$

This shows that the probability a good path exists between two typical vertex with length less than k_n, converges to zero as $n\to\infty$. In addition with the bound on the bad paths this shows that

$$\mathbb{P}\left(\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(o_1,o_2) \leq (1-\varepsilon)(\frac{\log\log n}{|\log(\gamma^{(\text{out})}-2)|}+\frac{\log\log n}{|\log(\gamma^{(\text{in})}-2)|})\right) = o(1). \tag{2.154}$$

Remember that $\gamma^{(q)} = \tau^{(q)} - \delta^{(q)}$, and we can take $\delta^{(q)} > 0$ arbitrary small. It follows that the largest lower bound is

$$(1-\varepsilon)(\frac{\log\log n}{|\log(\gamma^{(\text{out})}-2)|}+\frac{\log\log n}{|\log(\gamma^{(\text{in})}-2)|}), \tag{2.155}$$

which completes the proof of the claim. □

As the $\varepsilon > 0$ from the upper bound in Theorem 2.3.4 and the lower bound in Theorem 2.4.4 can be taken arbitrary small, this shows that the typical distances in the directed configuration model are approximately

$$\left(\frac{\log\log n}{|\log(\gamma^{(\mathrm{out})} - 2)|} + \frac{\log\log n}{|\log(\gamma^{(\mathrm{in})} - 2)|}\right), \tag{2.156}$$

which proves the claim of Theorem 1.5.4.

Chapter 3

Fluctuation typical distances directed configuration model

This section we will provide a proof for Theorem 1.6.5, which describes how the typical distances in the truncated directed configuration model fluctuate with the size of the graph. We start Section 3.1 by coupling the the respective forward and backward breadth-first exploration of the starting and target vertex with two independent branching processes. This will allow us to apply properties of the branching processes to describe the growth rate of the neighborhoods. Then, in Section 3.2 we show the typical number of steps needed to go from the starting vertex to the outbound hub, and from the inbound hub to the target vertex. In Section 3.3 we show that the combined number of steps is likely not enough to reach the target vertex through any path. Then in Section 3.4 we finish the proof by showing the typical distance from the outbound hub to the inbound hub.

3.1 Growth rate of the branching process

In Section 2 it is shown how the forward and backward breadth-first exploration can be coupled with two independent branching processes, as long as the explored neighborhoods are disjoint. The branching process coupled with the forward exploration has generation sizes $(Z_k^{(n;\text{out})})_{k\geq 0}$, with root distribution $D_n^{(\text{out})}$ and for the following generations offspring distribution $D_n^{*(\text{out})}$, defined in (1.14). The backward exploration is coupled with the branching process with generation sizes $(Z_k^{(n;\text{in})})_{k\geq 0}$, with root distribution $D_n^{(\text{in})}$ and for the following generations offspring distribution $D_n^{*(\text{in})}$, defined in (1.14). For two uniformly and independently chosen (typical) vertices $v_{(\text{out})}$ and $v_{(\text{in})}$, let the simultaneous forward and backward exploration at time t contain all the vertices at most distance t from $\mathcal{C}_0^{(\text{out})} = v_{(\text{out})}$, and all the vertices with distance at most t towards $\mathcal{C}_0^{(\text{in})} = v_{(\text{in})}$. We will denote these respective clusters by $\mathcal{C}_t^{(\text{out})}$ and $\mathcal{C}_0^{(\text{in})}$. Hence, the coupling from Corollary 2.2.2 shows that the simultaneous forward and backward neighborhood exploration can be coupled to the independent pair $((Z_k^{(n;\text{out})})_{k\geq 0}, (Z_k^{(n;\text{in})})_{k\geq 0})$. We will expand this coupling result by showing it is possible to add the independent asymptotic forward and backward branching processes with generation sizes $(Z_k^{(\text{out})})_{k\geq 0}$ and $Z_k^{(\text{in})})_{k\geq 0}$. The offspring distribution are simply the asymptotic variant, as under Assumption 1.6.2 for both $q \in \{\text{out}, \text{in}\}$ $D_n^{(q)} \to D^{(q)}$ and $D_n^{*(q)} \to D^{*(q)}$.

Corollary 3.1.1. *Consider the directed configuration model for which Assumptions 1.6.1 and 1.6.2 hold, and take r such that*

$$|\mathcal{C}_r^{(\text{out})} \cup \mathcal{C}_r^{(\text{in})}| \leq \min\{n^{-\beta_n^{(\text{in})}(\kappa-\delta)}, n^{-\beta_n^{(\text{out})}(\kappa-\delta)}, n^{(1-\beta_n^{(\text{out})}(1+\varepsilon)-\delta)/2}, n^{(1-\beta_n^{(\text{in})}(1+\varepsilon)-\delta)/2}\}, \quad (3.1)$$

for some $\delta > 0$. *Then* $(\mathcal{C}_r^{(out)}, \mathcal{C}_r^{(in)})$ *can whp be coupled to two independent branching processes* $(Z_k^{(out)}, Z_k^{(in)})_{k\leq r}$.

Proof. Note that Corollary 2.2.2 holds as long as $d_{max} = \max_{u\in[n]}\{d_u^{(\text{in})}, d_u^{(\text{out})}\} = o(n)$, which is the case under the truncation from Assumption 1.6.1. So apply Corollary 2.2.2 to couple $(G_n^{(\text{out})}(s), G_n^{(\text{in})}(t))_{s+t\in\mathbb{N}_0}$ and $(\mathcal{Z}_n^{(\text{out})}(s), \mathcal{Z}_n^{(\text{in})}(t))_{s+t\in\mathbb{N}_0}$. Now let $(\mathcal{Z}^{(\text{out})}(s), \mathcal{Z}^{(\text{in})}(t))_{s+t\in\mathbb{N}_0}$ denote the simultaneous tree-exploration of the forward and marked backward branching processes. Here, $(\mathcal{Z}^{(\text{out})}(s), \mathcal{Z}^{(\text{in})}(t))$ denotes the graph after s tree-exploration steps on the forward branching process, and t steps on the backward branching process. Our goal is to couple $(\mathcal{Z}^{(\text{out})}(s), \mathcal{Z}^{(\text{in})}(t))_{s+t\in\mathbb{N}_0}$ with $(\mathcal{Z}_n^{(\text{out})}(s), \mathcal{Z}_n^{(\text{in})}(t))_{s+t\in\mathbb{N}_0}$, and thus by extension with $(G_n^{(\text{out})}(s), G_n^{(\text{in})}(t))_{s+t\in\mathbb{N}_0}$. On the event that

$$(\hat{\mathcal{Z}}_n^{(\text{out})}(s), \hat{\mathcal{Z}}_n^{(\text{in})}(t))_{s+t=0}^{m_n-1} = (\hat{\mathcal{Z}}^{(\text{out})}(s), \hat{\mathcal{Z}}^{(\text{in})}(t))_{s+t=0}^{m_n-1}, \tag{3.2}$$

we construct the coupling step between the forward branching processes in the following way. Given that the latest tree-explored vertex has offspring $D_n^{*(\text{out})} = l$, we apply the optimal coupling that realizes the total variation defined in (1.47), between $D_n^{*(\text{out})}$ and $D^{*(\text{out})}$. Namely

$$\mathbb{P}\left(D^{*(\text{out})} = l | D_n^{*(\text{out})} = l\right) = \frac{\min\{\mathbb{P}(D_n^{*(\text{out})} = l), \mathbb{P}(D^{*(\text{out})} = l)\}}{\mathbb{P}(D_n^{*(\text{out})} = l)}. \tag{3.3}$$

So the coupling error of this step equals

$$\mathbb{P}(D^{*(\text{out})} \neq D_n^{*(\text{out})}) = 1 - \sum_{l\geq 1} \min\{\mathbb{P}(D_n^{*(\text{out})} = l), \mathbb{P}(D^{*(\text{out})} = l)\} = d_{TV}(F_n^{(\text{out})}, F^{(\text{out})}). \tag{3.4}$$

With the Union bound we can bound the probability an error occurs between the coupling for any of the m_n steps by

$$m_n \max\{d_{TV}(F_n^{(\text{out})}, F^{(\text{out})}), d_{TV}(F_n^{(\text{in})}, F^{(\text{in})})\} \leq m_n n^{-\beta_n \kappa}. \tag{3.5}$$

Similarly we can couple the out-degree and in-degree $d_{v_{(\text{out})}}^{(\text{out})}$ and $d_{v_{(\text{in})}}^{(\text{in})}$ of the roots to two independent $D^{(\text{out})}$ and $D^{(\text{in})}$ with coupling error at most $2d_{TV}(F_n, F)$. Hence the coupling error between $(\mathcal{Z}^{(\text{out})}(s), \mathcal{Z}^{(\text{in})}(t))_{s+t\in\mathbb{N}_0}$ and $(\mathcal{Z}_n^{(\text{out})}(s), \mathcal{Z}_n^{(\text{in})}(t))_{s+t\in\mathbb{N}_0}$ converges to zero for $m_n = o(n^{\beta_n\kappa})$. Now note that under Assumption 1.6.1 the maximal out- and in-degree are less than $n^{\beta_n^{(\text{out})}(1+\varepsilon)}$ and $n^{\beta_n^{(\text{in})}(1+\varepsilon)}$ for $\varepsilon > 0$. So the coupling error of Corollary 2.2.2 converges to zero for $m_n = o(n^{(1-\beta_n(1+\varepsilon))/2}$. This completes the proof of the claim. ☐

To study the growth rate of the clusters $\mathcal{C}_r^{(\text{out})}$ and $\mathcal{C}_r^{(\text{in})}$ with r, we can now make use of some known results about the growth rate of branching processes. We start by introducing the theorem by Davies [7], describing the growth rate of an infinite-mean non-delayed branching process.

Theorem 3.1.2. (Growth rate of an infinite mean BP) *Let* $(\tilde{Z}_k)_{k\geq 0}$ *be a non-delayed branching process with offspring distribution* $Z_1 = D^*$ *having distribution function* F^*. *Suppose that there exists a non-negative, non-decreasing function* $x \mapsto \gamma(x)$ *such that*

$$x^{-(\tau-2)-\gamma(x)} \leq [1 - F^*](x) \leq x^{-(\tau-2)+\gamma(x)}, \tag{3.6}$$

where $x \mapsto \gamma(x)$ *satisfies*

1. $x \mapsto x^{\gamma(x)}$ *is non-decreasing,*
2. $\int_0^\infty \gamma(e^{e^x})dx < \infty$.

Then $(\tau - 2)^k \log(\tilde{Z}_k \vee 1) \xrightarrow{a.s.} \tilde{Y}$ *as* $k \to \infty$, *with* $\mathbb{P}(\tilde{Y} = 0)$ *equal to the extinction probability of* $(\tilde{Z}_k)_{k\geq 0}$.

Note that Davies' theorem describes the growth of a non-delayed branching process, while the coupled branching processes $(Z_k^{(\mathrm{out})})_{k\geq 0}$ and $(Z_k^{(\mathrm{in})})_{k\geq 0}$ are both delayed. Hence, the following lemma expands the result of Davies' theorem to describe the growth rate of the delayed branching process.

Lemma 3.1.3. (Growth rate infinite mean delayed BP) *Let $(Z_k)_{k\geq 0}$ denote the branching process with root distribution D and offspring distribution D^* for the following generations having distribution function F^*. Let Y be the limiting variable of $(\tau-2)^k \log(Z_k)$, for some $\tau \in (2,3)$. Then Y satisfies the distributional identity*

$$Y = (\tau - 2) \max_{1\leq i\leq D} \tilde{Y}_i, \tag{3.7}$$

where $\tilde{Y}_i$ are i.i.d. copies of the limiting random variable of the undelayed branching processes with offspring distribution F_n.

Proof. Branching processes have the property that all the subtrees connected to the root are themselves independently distributed undelayed branching processes, which we denote $\hat{Z}_{k-1}^{(i)}$ for $i \in [1, D]$. Hence, for every $k \geq 1$

$$Z_k \overset{\mathrm{d}}{=} \sum_{i=1}^{D} \hat{Z}_{k-1}^{(i)}. \tag{3.8}$$

Substituting this in the expression of our limiting variable gives

$$Y = \lim_{k\to\infty} (\tau-2)^k \log(Z_k) = \lim_{k\to\infty} (\tau-2)^k \log(\sum_{i=1}^{D} \hat{Z}_{k-1}^{(i)}). \tag{3.9}$$

We bound the right hand side of the equality from both sides by

$$\begin{aligned}(\tau-2)^k \log(\max_{i=1,\dots,D} \hat{Z}_{k-1}^{(i)}) &\leq (\tau-2)^k \log(\sum_{i=1}^{D} \hat{Z}_{k-1}^{(i)}) \\ &\leq (\tau-2)^k \log(D \max_{i=1,\dots,D} \hat{Z}_{k-1}^{(i)}).\end{aligned} \tag{3.10}$$

As $(\tau-2) \in (0,1)$ it follows that $(\tau-2)^k \log D \overset{\mathbb{P}}{\to} 0$, which leaves a similar bound on both sides. Moreover, from the monotonicity of the logarithm, the order of taking the max and the logarithm can be exchanged. So by applying Davis' Theorem 3.1.2 to find the limiting random variable for $(\tau-2)^k \log(\hat{Z}_{k-1}^{(i)})$ it follows from (3.9) and (3.10) that

$$Y = \lim_{k\to\infty} \max_{i=1,\dots,D} (\tau-2)\hat{Y}_{k-1}^{(i)}. \tag{3.11}$$

By exchanging the order of taking the limit and the max, the proof of the claim follows. □

Remember that the respective offspring distribution functions of the branching processes $(Z_k^{(n;\mathrm{out})})_{k\geq 0}$ and $(Z_k^{(n;\mathrm{in})})_{k\geq 0}$ are $F_n^{(\mathrm{out})}$ and $F_n^{(\mathrm{in})}$, with

$$F_n^{(\mathrm{out})}(x) = \frac{1}{l_n} \sum_{u\in[n]} d_u^{(\mathrm{in})} \mathbb{1}\{d_u^{(\mathrm{out})} \leq x\}, \quad F_n^{(\mathrm{in})}(x) = \frac{1}{l_n} \sum_{u\in[n]} d_u^{(\mathrm{out})} \mathbb{1}\{d_u^{(\mathrm{in})} \leq x\}. \tag{3.12}$$

Before we can apply Lemma 3.1.3, we show that the power-law Assumption 1.6.1 satisfies the conditions of Davies' Theorem 3.1.2.

Lemma 3.1.4. *Let $\gamma(x) = c(\log x)^{\gamma-1}$, for some $\gamma \in (0,1)$ and $c > 0$. Then,*

1. $x \mapsto x^{\gamma(x)}$ *is non-decreasing for all $x \geq 1$,*

2. $\int_0^\infty \gamma(e^{e^x})dx < \infty$.

Proof. Note that $\gamma(1) = 0$ and

$$\frac{d}{dx} c\log(x)^\gamma = \frac{c\gamma}{x}(\log x)^{\gamma-1} > 0, \quad x > 1, \tag{3.13}$$

which shows that $\gamma(x)$ is non-decreasing. Moreover, substituting $u = \log(x)$ with $du = \frac{1}{x}dx$,

$$\int_e^\infty \frac{c\log(x)^{\gamma-2}}{y}dy = c\int_e^\infty u^{\gamma-2} = [\frac{u^{\gamma-1}}{b-1}]_e^\infty = \frac{c}{\gamma-1} < \infty. \tag{3.14}$$

□

This shows that both $\gamma^{(\text{out})}$ and $\gamma^{(\text{in})}$ from Assumption 1.6.1 satisfy the conditions for Davies' Theorem, such that the extended result can describes the growth rate of our branching processes.

Coupling neighborhood with branching process Following the approach of ([15], Section 2) we want to apply the coupling to an as large as possible neighborhood for both the uniformly chosen starting and target vertex $v_{(\text{out})}$ and $v_{(\text{in})}$. As the growth rate of the branching processes $(Z_k^{(\text{out})})_{k\geq 0}$ and $(Z_k^{(\text{in})})_{k\geq 0}$ can vastly differ, we will simply apply the coupling till one of the neighborhood sizes approaches the limit from Corollary 3.1.1. Without loss of generality we assume that $\tau^{(\text{out})} > \tau^{(\text{in})}$ as we can simply exchange the labels if this is not the case. We will apply the coupling till the first branching process reaches the size limit $n^{\delta'}$ for $\delta' = (\tau^{(\text{out})} - 2)\delta$, where the

$$\delta = \min\{n^{-\beta_n^{(\text{in})}\kappa}, n^{-\beta_n^{(\text{out})}\kappa}, n^{(1-\beta_n^{(\text{out})}(1+\varepsilon))/2}, n^{(1-\beta_n^{(\text{in})}(1+\varepsilon))/2}\} \tag{3.15}$$

is from Lemma 3.1.1. Let $t(n^{\delta'}) = \inf\{k : Z_k^{(\text{out})} \geq n^{\delta'}\}$ denote the smallest number of steps, such that the forward neighborhood of $v_{(\text{out})}$ exceeds $n^{\delta'}$. Define the random variables

$$Y_n^{(\text{out})} = (\tau^{(\text{out})} - 2)^{t(n^{\delta'})}\log(Z_{t(n^{\delta'})}^{(\text{out})}), \quad Y_n^{(\text{in})} = (\tau^{(\text{in})} - 2)^{t(n^{\delta'})}\log(Z_{t(n^{\delta'})}^{(\text{in})}). \tag{3.16}$$

By rewriting the expression for $Y_n^{(\text{out})}$, we obtain

$$t(n^{\delta'}) = \frac{\log(\delta'/Y_n^{(\text{out})}) + \log\log(n)}{|\log(\tau^{(\text{out})} - 2)|} - a_n^{(\text{out})} + 1, \tag{3.17}$$

where

$$a_n^{(\text{out})} = \left\{\frac{\log(\delta'/Y_n^{(\text{out})}) + \log\log(n)}{|\log(\tau^{(\text{out})} - 2)|}\right\}, \tag{3.18}$$

for which $\{x\} = x - \lfloor x \rfloor$. By substituting the expression (3.17) in (3.16) and some rearrangements we find that the size of the last generation of the branching processes coupled with the forward neighborhood of $v_{(\text{out})}$ and backward neighborhood of $v_{(\text{in})}$ are

$$Z_{t(n^{\delta'})}^{(\text{out})} = n^{\delta'(\tau^{(\text{out})}-2)^{a_n^{(\text{out})}-1}} = m_{(\text{out})}, \tag{3.19}$$

and

$$Z_{t(n^{\delta'})}^{(\text{in})} = \exp\left\{Y_n^{(\text{in})}\left(\frac{\delta'}{Y_n^{(\text{out})}}\log n\right)^{\frac{|\log(\tau^{(\text{in})}-2)|}{|\log(\tau^{(\text{out})}-2)|}}(\tau^{(\text{in})} - 2)^{a_n^{(\text{out})}-1}\right\} = m_{(\text{in})}. \tag{3.20}$$

With the assumption that $m_n^{(\text{in})} \leq m_n^{(\text{out})}$, the choice of δ' ensures that the total size of the two branching processes does not exceed n^δ and the coupling stays accurate as we take larger sized graphs. Denote the coupled forward and backward explored neighborhoods as $\mathcal{C}^{(\text{out})}_{t(n^{\delta'})}$ and $\mathcal{C}^{(\text{in})}_{t(n^{\delta'})}$. As we shall study the typical distances based on these neighborhoods, we introduce the σ-algebra generated by these induced subgraph as

$$\mathcal{T}_{\delta'} = \sigma\left(\mathcal{C}^{(\text{out})}_{t(n^{\delta'})} \cup \mathcal{C}^{(\text{in})}_{t(n^{\delta'})}\right). \tag{3.21}$$

We denote the probability conditioned on this sigma-algebra as

$$\mathbb{P}_{\mathcal{T}}(\cdot) = \mathbb{P}(\cdot|\mathcal{T}_{\delta'}), \quad \mathbb{E}_{\mathcal{T}}[\cdot] = \mathbb{E}[\cdot|\mathcal{T}_{\delta'}], \tag{3.22}$$

and say an event A_n holds $\mathbb{P}_{\mathcal{T}}$-whp if $\mathbb{P}(A_n|\mathcal{T}_{\delta'}) \to 1$ as $n \to \infty$.

3.2 Number of steps towards the outbound and from the inbound hubs

This section we will study the typical distance towards the highest order out-degree vertices contained in the outbound hub, and the typical distance from the highest order in-degree vertices contained in the inbound hub. More precisely, define the vertex sets

$$\begin{aligned} \text{hub}^{(\text{out})} &:= \{u \in [n] : d^{(\text{out})}_u \geq l_n^{\beta_n^{(\text{out})}(\tau^{(\text{out})}-2)}\} \\ \text{hub}^{(\text{in})} &:= \{u \in [n] : d^{(\text{in})}_u \geq l_n^{\beta_n^{(\text{in})}(\tau^{(\text{in})}-2)}\}. \end{aligned} \tag{3.23}$$

For $b_n^{(\text{out})}$, $b_n^{(\text{out})}$, $T_{(\text{out})}$ and $T_{(\text{in})}$ as defined in (1.55) and (1.54), we shall prove that $\mathbb{P}_{\mathcal{T}}$-whp both

$$\begin{aligned} \text{dist}_{\text{DCM}_n(\boldsymbol{d})}(v_{(\text{out})}, \text{hub}^{(\text{out})}) = T_{(\text{out})} &= -1 + \frac{\log\log(l_n^{\beta_n^{(\text{out})}}) - \log(Y_n^{(\text{out})})}{|\log(\tau^{(\text{out})}-2)|} - b_n^{(\text{out})} \\ \text{dist}_{\text{DCM}_n(\boldsymbol{d})}(\text{hub}^{(\text{in})}, v_{(\text{in})}) = T_{(\text{in})} &= -1 + \frac{\log\log(l_n^{\beta_n^{(\text{in})}}) - \log(Y_n^{(\text{in})})}{|\log(\tau^{(\text{in})}-2)|} - b_n^{(\text{in})}. \end{aligned} \tag{3.24}$$

To do this this we shall follow the same steps as the proof of ([15], Proposition 2.1), which shows a similar result for the undirected configuration model. We start from the neighborhoods $\mathcal{C}^{(\text{out})}_{t(n^{\delta'})}$ and $\mathcal{C}^{(\text{in})}_{t(n^{\delta'})}$ that are coupled with the branching processes. Then, in Section 3.2.1 provide an upper bound, for which we segment the vertices for which the out- or in-degree is polynomial in n. In section 3.2.2 we provide a matching lower bound by using path-counting techniques. This proof will provide an upper bound on the out-degree of vertices that can be reached from $\mathcal{C}^{(\text{out})}_{t(n^{\delta'})}$ in a certain number of steps, and on the in-degree of vertices that can reach $\mathcal{C}^{(\text{in})}_{t(n^{\delta'})}$ in a certain number of steps.

To define the segments for the upper bound and the bounds on the in- and out-degree for the lower bound, we shall make use of the functions

$$h^{(q)}(x) := x^{\frac{2C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}}}(\log x)^{\eta^{(q)}-1}} = \exp\left\{\frac{2C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}}}(\log x)^{\eta^{(q)}}\right\}, \tag{3.25}$$

with $q \in \{\text{out}, \text{in}\}$, for which the constant $C^{(q)} > 0$ is from the power law Assumption 1.6.1 and $\eta^{(q)} \in (\gamma^{(q)}, 1)$. This function is chosen such that

$$\begin{aligned} &\left(x^{\frac{1}{\tau^{(q)}-2}}\right)^{-\gamma^{(q)}\left(x^{\frac{1}{\tau^{(q)}-2}}\right)} h^{(q)}(x) \to \infty, \\ &\left(x^{\frac{1}{\tau^{(q)}-2}}\right)^{\gamma^{(q)}\left(x^{\frac{1}{\tau^{(q)}-2}}\right)} / h^{(q)}(x) \to 0, \end{aligned} \tag{3.26}$$

as $x \to \infty$. As we shall see in the following sections, this will allow us to obtain matching bound on the distances, for which the difference becomes asymptotically small for large values of n.

3.2.1 Upper bound distance to the outbound and from the inbound hub

To construct the path towards the outbound hub and from the inbound hub, we shall use a similar approach as Section 5. Namely, we will prove the phenomenon discussed below Lemma 2.3.1, that each high out-degree vertex is likely to have a neighbor with exponentially larger out-degree and each high in-degree vertex is likely to be the neighbor of a vertex with exponentially larger in-degree. We will show that the path created by these particular neighbors, creates a path to and from the hubs for which we length can be bounded by $T_n^{(q)}$, for $q \in \{\text{out}, \text{in}\}$. Conditioning on the neighborhoods $C^{(\text{out})}_{t(n^{\delta'})}$ and $C^{(\text{in})}_{t(n^{\delta'})}$, we study paths starting from $C^{(\text{out})}_{t(n^{\delta'})}$ and paths ending in $C^{(\text{in})}_{t(n^{\delta'})}$. To construct the paths we segment the high in- and out-degree vertices into the layers

$$\Gamma_k^{(q)} := \{v \in [n] : d_v^{(q)} \geq u_k^{(q)}\}, \tag{3.27}$$

and choose $u_k^{(q)}$ such that $\Gamma_0^{(q)}$ intersects $\mathcal{C}^{(q)}_{t(n^{\delta'})}$. Using the function $h^{(q)}$ defined in (3.25), recursively define the minimal q-degrees of each layer $\Gamma_{k+1}^{(q)}$ by

$$u_{k+1}^{(q)} = \left(\frac{u_k^{(q)}}{h^{(q)}(u_k^{(q)})}\right)^{\frac{1}{(\tau^{(q)}-2)}}, \text{ with } u_0^{(q)} = \left(\frac{m_q}{h(m_n^{(q)})}\right)^{\frac{1}{(\tau^{(q)}-2)}}, \tag{3.28}$$

and $u_{-1}^{(q)} = m_n^{(q)}$. Solving this recursion results in the expression

$$u_k^{(q)} = m_q^{(\tau^{(q)}-2)^{-(k+1)}} / \prod_{i=1}^{k+1} h(u_{k-i}^{(q)})^{(\tau^{(q)}-2)^{-i}}. \tag{3.29}$$

As these values will serve as a lower bound on the highest in- or out-degree vertices that can be reached in $t(n^{\delta'}) + k$ steps, let us take a closer look at these values. The first thing to notice is that $u_k^{(q)} \leq m_{(q)}^{(\tau^{(q)}-2)^{-(k+1)}}$. We can apply this same inequality to obtain an upper bound for each term of the product in the denominator of (3.29) and thus a lower bound on the whole expression. First, we apply the inequality on $h^{(q)}(u_l^{(q)})$ to obtain the bound

$$\begin{aligned} h^{(q)}(u_l^{(q)}) &= \exp\left\{\frac{2C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}}}(\log u_l^{(q)})^{\eta^{(q)}}\right\} \\ &\leq \exp\left\{\frac{2C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}}}(\log m_q^{(\tau^{(q)}-2)^{-(l+1)}})^{\eta^{(q)}}\right\} \\ &= \exp\left\{\frac{2C^{(q)}(\tau^{(q)}-2)^{-\eta^{(q)}(l+1)}}{(\tau^{(q)}-2)^{\eta^{(q)}}}(\log m_q)^{\eta^{(q)}}\right\}. \end{aligned} \tag{3.30}$$

Applying this bound on each term of the product $\left(\prod_{i=1}^{k+1} h(u_{k-i}^{(q)})^{(\tau^{(q)}-2)^{-i}}\right)^{-1}$, we can bound it from below by

$$\begin{aligned} &\prod_{i=1}^{k+1} \exp\{-\frac{2C(\tau^{(q)}-2)^{-\eta^{(q)}(k-i+1)}}{(\tau^{(q)}-2)^{\eta^{(q)}}}(\log m_q)^{\eta^{(q)}}(\tau^{(q)}-2)^{-i}\} \\ \geq &\exp\left\{-\frac{2C^{(q)}\left((\tau^{(q)}-2)^{-(\eta^{(q)}k+1)} - (\tau^{(q)}-2)^{\eta^{(q)}-k-2}\right)}{K_\eta(\tau^{(q)}-2)^{\eta^{(q)}}}(\log m_q)^{\eta^{(q)}}\right\} \\ \geq &\exp\left\{-\frac{2C^{(q)}(\tau^{(q)}-2)^{-(k+1)}}{K_\eta(\tau^{(q)}-2)^{\eta^{(q)}}}(\log m_q)^{\eta^{(q)}}\right\}, \end{aligned} \tag{3.31}$$

where $K_\eta = 1 - (\tau^{(q)} - 2)^{-(1-\eta^{(q)})}$. Hence, we obtain the lower bound

$$u_k^{(q)} \geq (m_n^{(q)})^{(\tau^{(q)}-2)^{-(k+1)}} \exp\left\{-\frac{2C^{(q)}(\tau^{(q)}-2)^{-(k+1)}}{K_\eta(\tau^{(q)}-2)^{\eta^{(q)}}}(\log m_q)^{\eta^{(q)}}\right\}. \tag{3.32}$$

Note as $(\tau^{(q)}-2)^{-1} > 1$ the exponent of the second factor is positive, but as $m_{(q)}$ tends to infinity and $\eta^{(q)} \in (0,1)$ we can see that for large n the second factor is of a smaller order than the first. This shows that $u_k^{(q)} \sim m_q^{(\tau^{(q)}-2)^{-(k+1)}}$, as defined in (1.53).

Now let us study the connectivity between the layers defined in (3.27). For the first step, we show that $\mathbb{P}_{\mathcal{T}}$-whp the layer $\Gamma_0^{(q)}$ intersects the neighborhood $\mathcal{C}^{(q)}_{t(n^{\delta'})}$. To do this, we use the following lemma:

Lemma 3.2.1. *Let X_i, $i = 1, ...m$ be i.i.d. random variables from distribution $F_n^{(q)}$ or $F^{(q)}$, $q \in \{in, out\}$. Then, given that Assumption 1.6.1 is satisfied,*

$$\mathbb{P}\left(\max_{i=1,\dots,m} X_i < \left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right) \leq \exp\left\{-\exp\left\{\frac{C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}}}\log(m)^{\eta^{(q)}}\right\}\right\}. \tag{3.33}$$

Proof. First, as each of the random variables are i.i.d., it follows that

$$\mathbb{P}\left(\max_{i=1,\dots,m} X_i < \left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right) \leq F_n^{(q)}\left(\left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right)^m. \tag{3.34}$$

Using the inequality $1 + x \leq e^x$, leads to

$$\begin{aligned} F_n^{(q)}\left(\left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right)^m &= \left(1 + \left(F_n^{(q)}\left(\left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right) - 1\right)\right)^m \\ &\leq \exp\left\{-m\left(1 - F_n^{(q)}\left(\left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right)\right)\right\}. \end{aligned} \tag{3.35}$$

Using the lower bound for $[1 - F_n^{(q)}](x)$ from Assumption 1.6.1 and the definition of $h^{(q)}(x)$ in (3.25) gives

$$\begin{aligned} [1 - F_n^{(q)}]\left(\left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right) &\geq \left(\left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right)^{-(\tau^{(q)}-2)-\gamma^{(q)}\left(\left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right)} \\ &= \frac{h^{(q)}(m)}{m}\exp\left\{-C^{(q)}\log\left(\left(\frac{m}{h^{(q)}(m)}\right)^{\frac{1}{\tau^{(q)}-2}}\right)^{\gamma^{(q)}}\right\} \\ &= \frac{h^{(q)}(m)}{m}\exp\left\{-\frac{C^{(q)}}{(\tau^{(q)}-2)^{\gamma^{(q)}}}\log\left(\frac{m}{h^{(q)}(m)}\right)^{\gamma^{(q)}}\right\} \\ &\geq \frac{1}{m}\exp\left\{\frac{C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}}}\log(m)^{\eta^{(q)}}\right\}, \end{aligned} \tag{3.36}$$

where the last inequality follows as $\eta^{(q)} > \gamma^{(q)}$, such that $(\tau^{(q)}-2)^{-\eta^{(q)}} > (\tau^{(q)}-2)^{-\gamma^{(q)}}$. Using this bound on the exponent of the RHS of (3.35) proves the bound (3.33) □

The offspring distribution of each individual of the final generation of the coupled branching processes are independent. Therefore, we can apply Lemma 3.2.1 with $m = m_q$. As $m_q \to \infty$ with $n \to \infty$, the bound (3.33) converges to zero. From the definition of $u_0^{(q)}$ in (3.28) we obtain that the event

$$\{\exists v \in \mathcal{C}^{(q)}_{t(n^{\delta'})} : d_v^{(q)} \geq u_0^{(q)}\} = \{\Gamma_0^{(q)} \cap \mathcal{C}^{(q)}_{t(n^{\delta'})} \neq \emptyset\}, \tag{3.37}$$

holds $\mathbb{P}_{\mathcal{T}}$-whp. To show the segments construct a path to and from the hubs, it remains to prove that all subsequent segments $\Gamma^{(q)}_{k-1}$ and $\Gamma^{(q)}_{k}$ are connected, for all k up to

$$k^*_{(q)} = \sup\{k : u^{(q)}_k \leq l_n^{\beta^{(q)}}\}. \tag{3.38}$$

We will do this in the following lemma, where we prove that $\mathbb{P}_{\mathcal{T}}$-whp, each $v \in \Gamma^{(\text{out})}_{k-1}$ has a neighbor in $\Gamma^{(\text{out})}_{k}$, and each vertex $v \in \Gamma^{(\text{in})}_{k}$ is the neighbor of a vertex in $\Gamma^{(\text{in})}_{k-1}$. Remember that $\mathcal{N}^{(\text{out})}(A)$ denotes the set of neighboring vertices of the vertex set A, and $\mathcal{N}^{(\text{in})}(A)$ denotes the set of vertices and have a neighbor in set A.

Lemma 3.2.2. *Let Assumption 1.6.2, 1.4.2 and 1.6.1 be satisfied. Then, for $k^*_{(out)}$ and $k^*_{(in)}$ as defined in (3.38), the events*

$$\left\{\forall k \in [1, k^*_{(\text{out})}], v \in \Gamma^{(\text{out})}_{k-1} : \mathcal{N}^{(\text{out})}(v) \cap \Gamma^{(\text{out})}_{k} \neq \emptyset\right\}, \tag{3.39}$$

and

$$\left\{\forall k \in [1, k^*_{(\text{in})}], v \in \Gamma^{(\text{in})}_{k-1} : \mathcal{N}^{(\text{in})}(v) \cap \Gamma^{(\text{in})}_{k} \neq \emptyset\right\}, \tag{3.40}$$

both hold $\mathbb{P}_{\mathcal{T}}$-whp.

Proof. We start by applying Lemma 2.3.1, which shows the probability that a vertex with out-degree at least $u^{(\text{out})}_{k-1}$ does not have a neighboring vertex in $\Gamma^{(\text{out})}_{k}$, is bounded by

$$\exp\left\{-u^{(\text{out})}_{k-1}[1-F^{(\text{out})}_n](u^{(\text{out})}_k)\right\}, \tag{3.41}$$

and the probability that a vertex with in-degree at least $u^{(\text{in})}_{k-1}$ is not the neighbor of any vertex in $\Gamma^{(\text{in})}_{k}$ is bounded by

$$\exp\left\{-u^{(\text{in})}_{k-1}[1-F^{(\text{in})}_n](u^{(\text{in})}_k)\right\}. \tag{3.42}$$

From the lower bound on $[1-F^{(q)}_n](x)$ from Assumption 1.6.1 and the definition of $u^{(q)}_k$ in (3.28), we can bound the exponents of (3.41) and (3.42) by

$$\begin{aligned} u^{(q)}_{k-1}[1-F^{(q)}_n](u^{(q)}_k) &\geq u^{(q)}_{k-1}(u^{(q)}_k)^{-(\tau^{(q)}-2)}\exp\{-C^{(q)}\log(u^{(q)}_k)^{\gamma^{(q)}}\} \\ &= h^{(q)}(u^{(q)}_{k-1})\exp\{-C^{(q)}\log(u^{(q)}_k)^{\gamma^{(q)}}\}. \end{aligned} \tag{3.43}$$

Note $u^{(q)}_k \leq (u^{(q)}_{k-1})^{\frac{1}{\tau^{(q)}-2}}$, such that from the definition of $h^{(q)}$ it follows that

$$\begin{aligned} u^{(q)}_{k-1}[1-F^{(q)}_n](u^{(q)}_k) &\geq \exp\left\{\frac{2C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}}}\log(u^{(q)}_{k-1})^{\eta^{(q)}} - C^{(q)}\log(u^{(q)}_k)^{\gamma^{(q)}}\right\} \\ &\geq \exp\left\{\frac{C^{(q)}}{(\tau^{(q)}-2)}\log(u^{(q)}_{k-1})^{\eta^{(q)}}\right\} \\ &\geq \exp\left\{\frac{\tilde{C}^{(q)}}{(\tau^{(q)}-2)^k}\log(m^{(q)})^{\eta^{(q)}}\right\}, \end{aligned} \tag{3.44}$$

for some $\tilde{C}^{(q)} > 0$. By applying the bound (3.44) on the exponent of both (3.41) and (3.42), we find that for all $v^{(\text{out})} \in \Gamma^{(\text{out})}_{k-1}$ and $v^{(\text{in})} \in \Gamma^{(\text{out})}_{k}$ that

$$\begin{aligned} \mathbb{P}_{\mathcal{T}}(\mathcal{N}^{(\text{out})}(v^{(\text{out})}) \cap \Gamma^{(\text{out})}_k = \emptyset) &\leq \exp\left\{-\exp\{\tilde{C}^{(\text{out})}(\tau^{(\text{out})}-2)^{-k}(\log(m^{(\text{out})}_n)^{\eta^{(\text{out})}}\}\right\}, \\ \mathbb{P}_{\mathcal{T}}(\mathcal{N}^{(\text{in})}(v^{(\text{in})}) \cap \Gamma^{(\text{in})}_k = \emptyset) &\leq \exp\left\{-\exp\{\tilde{C}^{(\text{in})}(\tau^{(\text{in})}-2)^{-k}(\log(m^{(\text{in})}_n)^{\eta^{(\text{in})}}\}\right\}. \end{aligned} \tag{3.45}$$

To complete the proof of the claim, note that even if we sum the RHS of (3.45) over all $k \geq 1$, the bound converges to zero as $m_q \to \infty$ with $n \to \infty$. ☐

Lemma 3.2.1 and Lemma 3.2.2 show that $\mathbb{P}_{\mathcal{T}}$-whp a path of length $k^*_{(\text{out})}$ from the neighborhood $\mathcal{C}^{(\text{out})}_{t(n^\delta)}$ to layer $\Gamma^{(\text{out})}_{k^*_{(\text{out})}}$, and a path of length $k^*_{(\text{in})}$ from layer $\Gamma^{(\text{in})}_{k^*_{(\text{in})}}$ to $\mathcal{C}^{(\text{in})}_{t(n^\delta)}$ exists. It remains to find the value for both $k^*_{(q)}$, $q \in \{\text{out}, \text{in}\}$. From the definition of $u^{(q)}_k$ in (3.28), the number of steps towards the final layer $\Gamma^{(\text{in})}_{k^*_{(\text{out})}}$ is equivalent to

$$k^*_{(\text{out})} = -1 + \frac{\log\left(\beta^{(\text{out})}_n \log(l_n)/(\delta'(\tau^{(\text{out})}-2)^{a_n - 1}\log(n))\right)}{|\log(\tau^{(\text{out})}-2)|} - b^{(\text{out})}_n, \tag{3.46}$$

where $b^{(\text{out})}_n$ is the fractional of its previous term. Using the value of $a^{(\text{out})}_n$ and that $\{x-1+\{y\}\} = \{x+y\}$, yields

$$b^{(\text{out})}_n = \left\{ \frac{\log(\beta^{(\text{out})}_n / Y^{(\text{out})}_n) + \log\log(l_n)}{|\log(\tau^{(\text{out})}-2)|} \right\}. \tag{3.47}$$

Taking $\nu = \frac{|\log(\tau^{(\text{in})}-2)|}{|\log(\tau^{(\text{out})}-2)|}$, we similarly find that the number of steps from $\Gamma^{(\text{in})}_{k^*_{(\text{in})}}$ equals

$$k^*_{(\text{in})} = -1 + \frac{\log\left(\beta^{(\text{in})}_n \log(l_n)/\left(Y^{(\text{in})}_n \left(\frac{\delta'}{Y^{(\text{out})}_n}\log n\right)^{\nu} (\tau^{(\text{in})}-2)^{a^{(\text{out})}_n - 1}\right)\right)}{|\log(\tau^{(\text{in})}-2)|} - b^{(\text{in})}_n, \tag{3.48}$$

with

$$b^{(\text{in})}_n = \left\{ \frac{\log\log(l_n^{\beta^{(\text{in})}_n}) - \log(Y^{(\text{in})}_n)}{|\log(\tau^{(\text{in})}-2)|} \right\}. \tag{3.49}$$

From the definition of $u^{(q)}_k$ it follows that the segment $\Gamma^{(q)}_{k^*_{(q)}}$ contains vertices with $q-$degree at least $u^{(q)}_{k^*_{(q)}} \sim l_n^{(\tau^{(q)}-2)^{b^{(q)}_n}}$. Denoting the length of the path of vertex $o_{(\text{out})}$ to the outbound hub as $T^{(\text{out})}_n$ and the path from the inbound hub to $o_{(\text{in})}$ as $T^{(\text{in})}_n$ we find that

$$T^{(q)}_n = t(n^{\delta'}) + k^*_{(q)} = -1 + \frac{\log\log(l_n^{\beta^{(q)}_n}) - \log(Y^{(q)}_n)}{|\log(\tau^{(q)}-2)|} - b^{(q)}_n, \tag{3.50}$$

for $q \in \{\text{out}, \text{in}\}$. This shows that

$$\begin{aligned} \text{dist}_{\text{DCM}_n(\boldsymbol{d})}(v_{(\text{out})}, \text{hub}^{(\text{out})}) &\leq T_{(\text{out})} \\ \text{dist}_{\text{DCM}_n(\boldsymbol{d})}(\text{hub}^{(\text{in})}, v_{(\text{in})}) &\leq T_{(\text{in})}. \end{aligned} \tag{3.51}$$

3.2.2 Lower bound on distance towards the outbound and from the inbound hub

This section we will provide a matching lower bound on the typical distances to or from the out- and inbound hub, by providing an upper bound on the out- and in-degrees of vertices at a certain distance from a typical vertex. To do this, we again use the function $h^{(q)}$ defined in (3.25), to recursively define

$$\hat{u}^{(q)}_{k+1} = (\hat{u}^{(q)}_k h^{(q)}(\hat{u}^{(q)}_k))^{\frac{1}{\tau^{(q)}-2}}, \quad \hat{u}^{(q)}_0 = (m_q h^{(q)}(m_{(q)}))^{\frac{1}{\tau^{(q)}-2}}. \tag{3.52}$$

Note that as we multiply with $h^{(q)}$ instead of dividing, the values of $\hat{u}^{(q)}_k$ grow faster than $u^{(q)}_k$, defined in (3.28). We will use the same path counting techniques that were applied in Section 6 to show that it is unlikely there is a vertex with q-degree $\hat{u}^{(q)}_k$ that is k steps from (q = out) or k steps towards (q = in) the neighborhood $\mathcal{C}^{(q)}_{t(n^{\delta'})}$. Remember that $\mathcal{BP}^{(\text{out})}_k(u)$ defined in (2.120), denotes the set of "bad" paths $((\pi_0,t_0),(\pi_1,s_1,t_1),\dots,(\pi_{k-1},s_{k-1},t_{k-1}),(\pi_k,s_k))$ such that $d^{(\text{out})}_{\pi_s} < \hat{u}^{(\text{out})}_s$ for all $s<k$ and $d^{(\text{out})}_{\pi_k} \geq \hat{u}^{(\text{out})}_k$, and that $\mathcal{BP}^{(\text{in})}_k(u)$ is the set of paths that is truncated on the in-degree in the opposite direction, such that $d^{(\text{in})}_{\pi_{k-s}} < \hat{u}^{(\text{in})}_s$ for all $s<k$ and $d^{(\text{in})}_{\pi_0} \geq \hat{u}^{(\text{in})}_k$. Define

$$\mathcal{BP}^{(\text{out})}_k = \bigcup_{v\in\mathcal{C}^{(\text{out})}_{t(n^{\delta'})}} \mathcal{BP}^{(\text{out})}_k(v), \quad \mathcal{BP}^{(\text{in})}_k = \bigcup_{v\in\mathcal{C}^{(\text{in})}_{t(n^{\delta'})}} \mathcal{BP}^{(\text{in})}_k(v), \tag{3.53}$$

denoting the respective set of bad paths starting in a free outbound half-edge of $\mathcal{C}^{(\text{out})}_{t(n^{\delta'})}$ or ending in a free inbound half-edge of $\mathcal{C}^{(\text{in})}_{t(n^{\delta'})}$. With the following lemma we shall show it is unlikely such a bad path exist.

Lemma 3.2.3. *Let Assumptions 1.4.2, 1.6.1, and 1.6.2 be satisfied. Then, for both $q\in\{out,in\}$*

$$\mathbb{P}_{\mathcal{T}}\left(\exists k\in[0,k^{(q)}_*]:\mathcal{BP}^{(q)}_k\neq\emptyset\right) \leq \exp\{-C\log(m_q)^{\eta^{(q)}}\}, \tag{3.54}$$

for some constant $C>0$.

Proof. Before we start with counting the number of paths, we need to consider the $O(n^{\delta(\tau^{(\text{out})}-2)})$ in- and outbound half-edges that have already been paired for the construction of both $\mathcal{C}^{(\text{out})}_{t(n^{\delta'})}$ and $\mathcal{C}^{(\text{in})}_{t(n^{\delta'})}$. Denoting the remaining number of out- or inbound half-edges by l^*_n, it is easy to see from the neighborhood sizes (3.19) and (3.20) that $l^*_n = l_n(1-o_{\mathbb{P}}(1))$. Taking the number of used half-edges into consideration, we can apply the same path counting methods as used in Lemma 2.4.2 to obtain for $p,q\in\{\text{out},\text{in}\}$ with $p\neq q$ the bounds

$$\mathbb{E}_{\mathcal{T}}\left[|\mathcal{BP}^{(q)}_k|\right] \leq \sum_{\pi_0\in\mathcal{C}^{(q)}_{t(n^{\delta'})}} d^{(q)}_{\pi_0} e^{\frac{k^2}{l^*_n}} \sum_{\pi_k:d^{(q)}_{\pi_k}>\hat{u}^{(q)}_k} \frac{d^{(p)}_{\pi_k}}{l^*_n} \times \prod_{i=1}^{k-1} \sum_{\pi_i:d^{(q)}_{\pi_i}\leq\hat{u}^{(q)}_i} \frac{d^{(\text{out})}_{\pi_i} d^{(\text{in})}_{\pi_i}}{l^*_n}. \tag{3.55}$$

Replacing the l^*_n in the denominators of (3.55) with l_n causes an error of at most

$$1+O(n^{\delta'(\tau^{(\text{out})}-2)-1})^k \leq \exp\{kn^{\delta'(\tau^{(\text{out})}-2)-1}\}, \tag{3.56}$$

which tends to 1 as $n\to\infty$ for any $k=O(\log\log n)$ as $\delta'(\tau^{(\text{out})}-2)<1$. After this replacement the expression closely resembles (2.128), such that we can apply Lemma 2.4.2 to obtain

$$\mathbb{E}_{\mathcal{T}}\left[|\mathcal{BP}^{(q)}_k|\right] \leq \sum_{\pi_0\in\mathcal{C}^{(q)}_{t(n^{\delta'})}} d^{(q)}_{\pi_0} e^{k^2/l^*_n}[1-F^{(q)}_n](\hat{u}^{(q)}_k) \times \prod_{l=1}^{k-1} \sum_{\pi_i:d^{(q)}_{\pi_i}\leq\hat{u}^{(q)}_i} \frac{d^{(\text{out})}_{\pi_i} d^{(\text{in})}_{\pi_i}}{l_n}(1+o(1)). \tag{3.57}$$

We can bound $[1-F^{(q)}_n](\hat{u}^{(q)}_k)$ using our Assumption 1.6.1, for the remaining factors we will provide bounds with the following lemma:

Lemma 3.2.4. *Let $\left(D^{*(q)}_{n,i}\right)_{i=1}^m$ be i.i.d. random variables from distribution $F^{(q)}_n$ or $F^{(q)}$. Then, under Assumption 1.6.1*

$$\mathbb{P}\left(\sum_{i=1}^m D^{*(q)}_{n,i} \geq (mh^{(q)}(m))^{\frac{1}{(\tau^{(q)}-2)}}\right) \leq (\frac{2}{3-\tau^{(q)}}+1)\exp\left\{-\frac{C}{(\tau^{(q)}-2)\gamma^{(q)}}\log(m)^{\gamma^{(q)}}\right\}. \tag{3.58}$$

Moreover, for all sequences $y_n\to\infty$ and n large enough

$$\sum_{\pi:d^{(q)}_\pi\leq y_n} \frac{d^{(\text{out})}_\pi d^{(\text{in})}_\pi}{l_n} \leq \frac{2}{3-\tau^{(q)}} y_n^{3-\tau^{(q)}+C^{(q)}(\log y_n)^{\gamma^{(q)}-1}}. \tag{3.59}$$

Proof. Denoting $M_m^{(q)} := (mh^{(q)}(m))^{\frac{1}{\tau^{(q)}-2}}$ it follows from the union bound that

$$\mathbb{P}(\sum_{i=1}^{m} D_{n,i}^{*(q)} \geq M_m^{(q)}) \leq \mathbb{P}(\exists i \leq m : D_{n,i}^{*(q)} \geq M_m^{(q)}) + \mathbb{P}(\sum_{i=1}^{m} D_{n,i}^{*(q)} \mathbb{1}\{D_{n,i}^{*(q)} \leq M_m^{(q)}\} \geq M_m^{(q)}). \tag{3.60}$$

For the first probability on the RHS it follows from the upper bound of Assumption 1.6.1 and the union bound that

$$\begin{aligned}\mathbb{P}(\exists i \leq m : D_{n,i}^{*(q)} \geq M_m^{(q)}) &\leq m[1 - F_n^{(q)}](M_m^{(q)}) \leq m(M_m^{(q)})^{-(\tau^{(q)}-2)+\gamma^{(q)}(M_m^{(q)})} \\ &= \frac{1}{h^{(q)}(m)} \exp\left\{\frac{C^{(q)}}{(\tau^{(q)}-2)^{\gamma^{(q)}}} \log(mh^{(q)}(m))^{\gamma^{(q)}}\right\}.\end{aligned} \tag{3.61}$$

For the second probability on the RHS of (3.60) we first apply Markov's inequality

$$\mathbb{P}(\sum_{i=1}^{m} D_{n,i}^{*(q)} \mathbb{1}\{D_{n,i}^{*(q)} \leq M_m^{(q)}\} \geq M_m^{(q)}) \leq \frac{m\mathbb{E}[D_{n,i}^{*(q)} \mathbb{1}\{D_{n,i}^{*(q)} \leq M_m^{(q)}\}]}{M_m^{(q)}}, \tag{3.62}$$

where the summation becomes a multiplication by the number of terms due to the independence of the random variables. Next, we use the standard method to relate the expectation of a random variable to its tail distribution, namely

$$\begin{aligned}\mathbb{E}[D_{n,i}^{*(q)} \mathbb{1}\{D_{n,i}^{*(q)} \leq M_m^{(q)}\}] &= \sum_{j=1}^{M_m^{(q)}} j\mathbb{P}(D_{n,i}^{*(q)} = j) = \sum_{j=1}^{M_m^{(q)}} \sum_{s=1}^{j} \mathbb{P}(D_{n,i}^{*(q)} = j) \\ &= \sum_{s=1}^{M_m^{(q)}} \sum_{j=s}^{M_m^{(q)}} \mathbb{P}(D_{n,i}^{*(q)} = j) = \sum_{s=1}^{M_m^{(q)}} F_n^{(q)}(M_m^{(q)}) - F_n^{(q)}(s-1) \\ &\leq \sum_{s=1}^{M_m^{(q)}} [1 - F_n^{(q)}](s-1) \leq \sum_{s=0}^{M_m^{(q)}} [1 - F_n^{(q)}](s) \\ &\leq \sum_{s=0}^{M_m^{(q)}} s^{-(\tau^{(q)}-2)+\gamma^{(q)}(s)},\end{aligned} \tag{3.63}$$

where the last inequality follows from Assumption 1.6.1. As $s^{\gamma^{(q)}(s)}$ is slowly varying with s and $M_m^{(q)} \to \infty$ for $n \to \infty$, we are able to apply the direct half of Karamata's Theorem ([2], Prop 1.5.9) to bound the RHS of (3.63) by

$$\mathbb{E}[D_{n,i}^{*(q)} \mathbb{1}\{D_{n,i}^{*(q)} \leq M_m^{(q)}\}] \leq \frac{2}{3-\tau^{(q)}} (M_m^{(q)})^{3-\tau^{(q)}+\gamma^{(q)}(M_m^{(q)})}, \tag{3.64}$$

which proves the bound (3.59). Applying this bound to the RHS of (3.62), and adding (3.61), we obtain

$$\begin{aligned}\mathbb{P}(\sum_{i=1}^{m} D_{n,i}^{*(q)} \geq M_m^{(q)}) &\leq (\frac{2m}{3-\tau^{(q)}} + m)(M_m^{(q)})^{2-\tau^{(q)}+\gamma^{(q)}(M_m^{(q)})} \\ &= (\frac{2}{3-\tau^{(q)}} + 1) \frac{1}{h^{(q)}(m)} \exp\left\{\frac{C^{(q)}}{(\tau^{(q)}-2)^{\gamma^{(q)}}} \log(mh^{(q)}(m))^{\gamma^{(q)}}\right\} \\ &\leq (\frac{2}{3-\tau^{(q)}} + 1) \exp\left\{-\frac{C^{(q)}}{(\tau^{(q)}-2)^{\gamma^{(q)}}} \log(m)^{\gamma^{(q)}}\right\},\end{aligned} \tag{3.65}$$

which is the claim of the Lemma. ☐

As $m_q \to \infty$ with $n \to \infty$, Lemma 3.2.4 shows that the event $\mathcal{E}_n := \left\{\sum_{i=1}^{m_q} D_i^{*(q)} \leq \hat{u}_0^{(q)}\right\}$ holds $\mathbb{P}_{\mathcal{T}}$-whp. Conditioning on this event and applying the bounds (3.59) and the upper bound form Assumption 1.6.1 on (3.57) yields

$$\mathbb{E}_{\mathcal{T}}\left[|\mathcal{BP}_k^{(q)}||\mathcal{E}_n\right] \leq \hat{u}_0^{(q)}(\hat{u}_k^{(q)})^{-(\tau^{(q)}-2)+\gamma^{(q)}(\hat{u}_k^{(q)})} \prod_{l=1}^{k-1} \frac{2(\hat{u}_l^{(q)})^{3-\tau^{(q)}+\gamma^{(q)}(\hat{u}_l^{(q)})}}{3-\tau^{(q)}}. \tag{3.66}$$

Using the definition of $\hat{u}_k^{(q)}$ in (3.52) it can be shown that $\hat{u}_l^{(q)}(\hat{u}_{l+1}^{(q)})^{3-\tau^{(q)}}h^{(q)}(\hat{u}_l^{(q)}) = \hat{u}_{l+1}^{(q)}$, and that

$$\hat{u}_0^{(q)}(\hat{u}_k^{(q)})^{2-\tau^{(q)}}h^{(q)}(\hat{u}_{k-1}^{(q)}) \prod_{l=1}^{k-1}(\hat{u}_l^{(q)})^{3-\tau^{(q)}}h^{(q)}(\hat{u}_{l-1}^{(q)}) = 1. \tag{3.67}$$

Applying this equality for the bound (3.66) we obtain

$$\begin{aligned}
\mathbb{E}_{\mathcal{T}}\left[|\mathcal{BP}_k^{(q)}||\mathcal{E}_n\right] &\leq \prod_{l=0}^{k-1} \frac{2}{3-\tau^{(q)}} \frac{(\hat{u}_{l+1}^{(q)})^{\gamma^{(q)}(\hat{u}_{l+1}^{(q)})}}{h^{(q)}(\hat{u}_l^{(q)})} \\
&\leq \frac{2}{3-\tau^{(q)}} \prod_{l=0}^{k-1} \exp\left\{-\frac{C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}}} \log(\hat{u}_l^{(q)})^{\eta^{(q)}}\right\} \\
&\leq \frac{2}{3-\tau^{(q)}} \exp\left\{-C^{(q)} \sum_{l=0}^{k-1} (\tau^{(q)}-2)^{-\eta^{(q)}(l+2)} \log(m_q)^{\eta^{(q)}}\right\},
\end{aligned} \tag{3.68}$$

where the last inequality follows from $\hat{u}_k^{(q)} \geq m_q^{(\tau^{(q)}-2)^{-(k+1)}}$ as seen in definition (3.52). As $(\tau^{(q)}-2)^{-\eta^{(q)}} > 1$ the summation in the exponent is of the order $(\tau^{(q)}-2)^{-k\eta^{(q)}}(\log m_{(q)})^{\eta^{(q)}}$. So by Markov's inequality there exists a constant $\tilde{C}^{(q)} > 0$ such that

$$\mathbb{P}_{\mathcal{T}}\left(\exists k > 0 : \mathcal{BP}_k^{(q)} \neq \emptyset\right) \leq \sum_{k=1}^{\infty} \mathbb{E}_{\mathcal{T}}\left[|\mathcal{BP}_k^{(q)}|\right] \leq \exp\left\{-\tilde{C}^{(q)}(\log m_q)^{\eta^{(q)}}\right\} \to 0, \tag{3.69}$$

as $n \to \infty$ as $m_q \to \infty$. Note that for $k = 0$, the probability that there is a bad path is simply the event that there exists a vertex in the last generation of the branching process coupled with $\mathcal{C}_{t(n^{\delta'})}^{(q)}$ with degree at least $\hat{u}_0^{(q)}$, for which we have shown a bound in Lemma 3.2.4. Merging this error term with the RHS of (3.69) finishes the proof of the Lemma 3.2.3. □

Conditioned on the event $\left\{\forall k \in [0, k_q^*] : \mathcal{BP}_k^{(q)} = \emptyset, q \in \{\text{out}, \text{in}\}\right\}$, which Lemma 3.2.3 shows it holds $\mathbb{P}_{\mathcal{T}}$-whp, the value of $\hat{u}_i^{(\text{out})}$ is an upper bound on the out-degree of vertices $t(n^{\delta'}) + i$ steps from $v_{(\text{out})}$, and the value $\hat{u}_i^{(\text{in})}$ an upper bound on the in-degree of vertices $t(n^{\delta'}) + i$ steps towards $v_{(\text{in})}$. We now have all have shown all the bounds that we need to proof the typical distance to the outbound and from the inbound hub, described in the following Lemma:

Lemma 3.2.5. (Distance to the outbound and from the inbound hubs) *Consider the directed configuration model on n vertices that satisfies Assumptions 1.3.2, 1.4.2 and 1.6.1. Then, for a uniformly chosen vertices $v_{(out)}$ and $v_{(in)}$, both*

$$\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(v_{(\text{out})}, \text{hub}^{(\text{out})}) = T_{(\text{out})} = -1 + \frac{\log\log(l_n^{\beta_n^{(\text{out})}}) - \log(Y_n^{(\text{out})})}{|\log(\tau^{(\text{out})}-2)|} - b_n^{(\text{out})}, \tag{3.70}$$

and

$$\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(\text{hub}^{(\text{in})}, v_{(\text{in})}) = T_{(\text{in})} = -1 + \frac{\log\log(l_n^{\beta_n^{(\text{in})}}) - \log(Y_n^{(\text{in})})}{|\log(\tau^{(\text{in})}-2)|} - b_n^{(\text{in})}, \tag{3.71}$$

hold $\mathbb{P}_{\mathcal{T}}$-whp. Consequently, $\mathbb{P}$-whp there is a vertex $v^*_{(out)} \in \mathrm{hub}^{(out)}$ at distance $T_{(out)}$ from $v_{(out)}$ and a vertex $v^*_{(in)} \in \mathrm{hub}^{(in)}$ with distance $T_{(in)}$ towards $v_{(in)}$, such that

$$d^{(\mathrm{out})}_{v^*_{(\mathrm{out})}} \sim l_n^{\beta_n^{(\mathrm{out})}(\tau^{(\mathrm{out})}-2)^{b_n^{(\mathrm{out})}}}, \quad d^{(\mathrm{in})}_{v^*_{(\mathrm{in})}} \sim l_n^{\beta_n^{(\mathrm{in})}(\tau^{(\mathrm{in})}-2)^{b_n^{(\mathrm{in})}}}, \tag{3.72}$$

while this is not case for vertices at other distances.

Proof. Define

$$\hat{k}_*^{(q)} := \inf\{k : \hat{u}_k^{(q)} \geq l_n^{\beta_n^{(q)}(\tau^{(q)}-2)}\}, \tag{3.73}$$

which is similar to the definition of $k_*^{(q)}$ in (3.38), by the super exponential growth of $\hat{u}_k^{(q)}$ by powers of $1/(\tau^{(q)}-2)$. We have already shown the upper bound for the distances in (3.51). To show the matching lower bound, it remains to show that $k_*^{(q)} = \hat{k}_*^{(q)}$ holds $\mathbb{P}_{\mathcal{T}}$-whp. We do this by showing that similarly to $u_k^{(q)}$, it also holds that $\hat{u}_k^{(q)} \sim m_q^{(\tau^{(q)}-2)^{-(k+1)}}$. Solving the recursion of $\hat{u}_k^{(q)}$ in (3.52) leads to the expression

$$\hat{u}_k^{(q)} = m_q^{(\tau^{(q)}-2)^{-(k+1)}} \prod_{i=1}^{k+1} h^{(q)}(\hat{u}_{k-i}^{(q)})^{(\tau^{(q)}-2)^{-i}}. \tag{3.74}$$

To obtain a bound for the product, note that $h^{(q)}(\hat{u}_{k-i}^{(q)})^{(\tau^{(q)}-2)^{-i}}$ equals

$$\exp\left\{\frac{2C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}+i}}\left(\log((\hat{u}_{k-i-1}^{(q)})^{\frac{1}{\tau^{(q)}-2}}) + \frac{2C^{(q)}}{(\tau^{(q)}-2)}\log((\hat{u}_{k-i-1}^{(q)})^{\frac{1}{\tau^{(q)}-2}})^{\eta^{(q)}}\right)^{\eta^{(q)}}\right\}. \tag{3.75}$$

After some rewriting we can express this as

$$\exp\left\{\frac{2C^{(q)}}{(\tau^{(q)}-2)^{\eta^{(q)}+i}}\left(\log((\hat{u}_{k-i-1}^{(q)})^{\frac{1}{\tau^{(q)}-2}}) + \frac{2C^{(q)}}{(\tau^{(q)}-2)}\log((\hat{u}_{k-i-1}^{(q)})^{\frac{1}{\tau^{(q)}-2}})^{\eta^{(q)}}\right)^{\eta^{(q)}}\right\}. \tag{3.76}$$

Repeat this same rewriting process $(i-k-1)$ times results in the expression

$$\exp\left\{\frac{2C^{(q)}\log((\hat{u}_0^{(q)})^{\frac{1}{(\tau^{(q)}-2)^{k-i}}})^{\eta^{(q)}}}{(\tau^{(q)}-2)^{\eta^{(q)}+i}}\prod_{l=0}^{k-i-1}\left(1+\frac{2C^{(q)}}{(\tau^{(q)}-2)}\log((\hat{u}_l^{(q)})^{\frac{1}{\tau^{(q)}-2}})^{\eta^{(q)}-1}\right)^{\eta^{(q)}}\right\}. \tag{3.77}$$

Taking $\hat{u}_{-1}^{(q)} = m_q$ and using the definition of $\hat{u}_0^{(q)}$ show that $h^{(q)}(\hat{u}_{k-i}^{(q)})^{(\tau^{(q)}-2)^{-i}}$ equals

$$\exp\left\{\frac{2C^{(q)}\log\left(m_q^{\frac{1}{(\tau^{(q)}-2)}}\right)^{\eta^{(q)}}}{(\tau^{(q)}-2)^{\eta^{(q)}(k-i+1)+i}}\prod_{l=-1}^{k-i-1}\left(1+\frac{2C^{(q)}}{(\tau^{(q)}-2)}\log((\hat{u}_l^{(q)})^{\frac{1}{\tau^{(q)}-2}})^{\eta^{(q)}-1}\right)^{\eta^{(q)}}\right\}. \tag{3.78}$$

As $\eta^{(q)}-1<0$ and $\hat{u}_l^{(q)}$ tends to infinity as $n\to\infty$ it follows that term of the product can approximate 1 arbitrary close from above, by taking n large enough. So for some constant $C>0$ we can bound $\prod_{i=1}^{k+1} h^{(q)}(\hat{u}_{k-i}^{(q)})^{(\tau^{(q)}-2)^{-i}}$ by

$$\exp\left\{\frac{C\log\left(m_q^{\frac{1}{(\tau^{(q)}-2)}}\right)^{\eta^{(q)}}}{(\tau^{(q)}-2)^{k+1}}\sum_{i=0}^{k}\frac{1}{(\tau^{(q)}-2)^{i(\eta^{(q)}-1)}}\right\}, \tag{3.79}$$

which again can be bounded for some constant $\tilde{C} > 0$ by

$$\exp\left\{ \frac{\tilde{C}}{(\tau^{(q)}-2)^{k+1}} \log\left(m_q^{\frac{1}{(\tau^{(q)}-2)}} \right)^{\eta^{(q)}} \right\}. \tag{3.80}$$

The bound (3.80) shows that the product in (3.74) is of a much smaller order than $m_q^{(\tau^{(q)}-2)^{-(k+1)}}$, where $m_q \to \infty$ as $n \to \infty$. It thus follows that $\hat{u}_k^{(q)} \sim m_q^{(\tau^{(q)}-2)^{-(k+1)}}$, from which we can conclude that $k_*^{(q)} = \hat{k}_*^{(q)}$ $\mathbb{P}_{\mathcal{T}}$-whp. This concludes the proof of Lemma 3.2.5. □

3.3 Early meeting is unlikely

For the lower bound of the typical distance in Theorem 1.6.5, we show that the neighborhoods $\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}}$ and $\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}}$ are $\mathbb{P}_{\mathcal{T}}$-whp disjoint.

Lemma 3.3.1. *Consider the directed configuration model with n vertices that satisfies Assumption 1.3.2, 1.4.2 and 1.6.1. Then, the event*

$$\left\{ \mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}} \cap \mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}} = \emptyset \right\} \tag{3.81}$$

holds $\mathbb{P}_{\mathcal{T}}$-whp. Moreover, let $H^{(out)}(\mathcal{C}^{(out)}_{T_n^{(out)}})$ denote the number of outbound half-edges attached to vertices of $\mathcal{C}^{(out)}_{T_n^{(out)}}$ and $H^{(in)}(\mathcal{C}^{(in)}_{T_n^{(in)}})$ the number of inbound half-edges attached to vertices in $\mathcal{C}^{(in)}_{T_n^{(in)}}$. Then, $\mathbb{P}_{\mathcal{T}}$-whp

$$H^{(\text{out})}(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}}) \sim l_n^{\beta_n^{(\text{out})}(\tau^{(\text{out})}-2)^{b_n^{(\text{out})}}}, \quad H^{(\text{in})}(\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}}) \sim l_n^{\beta_n^{(\text{in})}(\tau^{(\text{in})}-2)^{b_n^{(\text{in})}}}. \tag{3.82}$$

Proof. We will prove the claim under the condition that the event

$$\text{NoBad} = \left\{ \forall k \in [0, k_q^*] : \mathcal{BP}_k^{(q)} = \emptyset, q \in \{\text{out}, \text{in}\} \right\} \tag{3.83}$$

holds for both $q \in \{\text{out}, \text{in}\}$, which Lemma 3.2.3 shows it will happen $\mathbb{P}_{\mathcal{T}}$-whp. As for any event A it holds that $\mathbb{P}_{\mathcal{T}}(\cdot) \geq \mathbb{P}_{\mathcal{T}}(\cdot|\text{NoBad})\mathbb{P}_{\mathcal{T}}(\text{NoBad})$, it is enough to show that the event (3.81) holds $\mathbb{P}_{\mathcal{T}}$-whp, conditioned on the event NoBad. To obtain a bound on the probability the neighborhoods intersect, we start by calculating the number of free outbound half-edges in $\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}-l_1}$, denoted by $\mathcal{H}^{(\text{out})}(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}-l_1})$, and the number of free inbound half-edges in $\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}-l_2}$, denoted by $\mathcal{H}^{(\text{in})}(\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}-l_2})$, for each $l_1 \in [k^*_{(\text{out})}]$ and $l_2 \in [k^*_{(\text{in})}]$. For this we will make use of the following paths:

Definition 3.3.2. (Open paths) *An open-ended path of length k is defined as the sequence*

$$((\pi_0, s_0), (\pi_1, t_1, s_1), \ldots, (\pi_k, s_k, t_k)), \tag{3.84}$$

for which π_i denotes the i'th vertex along the path, and t_i the label of the outbound half-edge attached the vertex π_i, that is paired with the inbound half-edge with label s_{i+1}, attached to vertex π_{i+1}. Similarly, an open-starting path is defined as the sequence

$$((\pi_0, t_0, s_0), (\pi_1, t_1, s_1), \ldots, (\pi_k, s_k)). \tag{3.85}$$

These paths are self-avoiding, meaning that $\pi_i \neq \pi_j$ for all $i \neq j$.

These open-ended and open-starting paths look similar to the directed paths defined in Definition 2.4.1. The difference is that for the k-length open-ended paths the outbound half-edge attached to the last vertex in the path remains unpaired, and similarly the first inbound half-edge for the k-length open-starting paths. As it is likely that there are multiple paths of length $T_n^{(\text{out})} - l$ that can be taken from vertex $v_{(\text{out})}$ to the vertex attached to a particular free outbound half-edge, and similarly paths of length $T_n^{(\text{in})} - l$ from some vertex attached to a particular free inbound half-edge to the vertex $v_{(\text{in})}$, the number of these of open paths serve as an upper bound for $\mathcal{H}^{(\text{out})}(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}-l_1})$ and $\mathcal{H}^{(\text{in})}(\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}-l_2})$.

By using these open paths as upper bounds on the available out- or inbound half-edges, we will be able to approach the proof with our familiar path-counting techniques. Let $\mathcal{R}_k^{(\text{out})}(A)$ denote the set, and $R_k^{(\text{out})}(A)$ the number of k-length open ended paths starting from a vertex in set A. Similarly, let $\mathcal{R}_k^{(\text{in})}(A)$ denote the set and $R_k^{(\text{in})}(A)$ the number of k-length open-starting paths ending at a vertex in set A. As $T_n^{(q)} = t(n^\delta) + k_*^{(q)}$ we obtain for any $k \leq k_*^{(q)}$ the bound

$$H^{(q)}\left(\mathcal{C}^{(q)}_{T_n^{(q)}-l}\right) \leq R^{(q)}_{k_*^{(q)}-l}\left(\mathcal{C}^{(q)}_{t(n^\delta)}\right). \tag{3.86}$$

Conditioned on the event NoBad from (3.83), the value of $\hat{u}_k^{(\text{out})}$ in (3.52) bounds the out-degree of any vertex at most distance $t(n^{\delta'}) + k$ from vertex $v_{(\text{out})}$ and similarly the $\hat{u}_k^{(\text{in})}$ bounds the in-degree of any vertex at most distance $t(n^{\delta'}) + k$ towards vertex $v_{(\text{in})}$. When counting the number of open-ended or open-starting paths we can use similar path counting arguments as we used to obtain the bound for the bad paths in (3.55). The difference is that the out-degree of the final vertex in the open-ended paths and the in-degree of the first vertex for the open-starting paths need to be taken into consideration. This is resolved quickly by replacing the last term in (3.55) with an additional term $d_{\pi_k}^{(\text{in})} d_{\pi_k}^{(\text{out})}$ in the second factor, i.e. we obtain

$$\mathbb{E}[R^{(q)}_{k_*^{(q)}-l}\left(\mathcal{C}^{(q)}_{t(n^\delta)}\right)|\text{NoBad}] \leq e^{k_*^{(q)}/l_n} \sum_{v \in \mathcal{C}^{(q)}_{t(n^{\delta'})}} d_v^{(q)} \prod_{i=1}^{k_*^{(q)}-l} \left(\sum_{\pi_i : d_{\pi_i}^{(q)} \leq \hat{u}_i^{(q)}} \frac{d_{\pi_i}^{(\text{in})} d_{\pi_i}^{(\text{out})}}{l_n} \right). \tag{3.87}$$

Note that each $l_n^* = l_n(1 - o_{\mathbb{P}}(1))$ in the denominator is replaced by l_n, which leads to an error which tends to one as $n \to \infty$. Remember that we paired both $\mathcal{C}^{(q)}_{t(n^\delta)}$ with a branching process, for which the number of offspring for each individual of the last generation is i.i.d.. As all the free outbound or inbound half-edges of these neighborhoods belong to the individuals of the final generation, we can apply Lemma 3.2.4 to $\mathbb{P}_\mathcal{T}$-whp bound the first summation factor of (3.87) by $\hat{u}_0^{(q)}$. To bound each term of the product, we apply the bound from (3.59) to obtain

$$\mathbb{E}_\mathcal{T}[R^{(q)}_{k_*^{(q)}-l}\left(\mathcal{C}^{(q)}_{t(n^\delta)}\right)|\text{NoBad}] \leq e^{k_*^{(q)}/l_n} \hat{u}_0^{(q)} \prod_{i=1}^{k_*^{(q)}-l} \frac{2(\hat{u}_i^{(q)})^{-(\tau^{(q)}-3)+\gamma^{(q)}(\hat{u}_i^{(q)})}}{3-\tau^{(q)}}. \tag{3.88}$$

Using the equality $\hat{u}_l^{(q)}(\hat{u}_{l+1}^{(q)})^{3-\tau^{(q)}} h^{(q)}(\hat{u}_l^{(q)}) = \hat{u}_{l+1}^{(q)}$, this bound equals

$$\mathbb{E}_\mathcal{T}[R^{(q)}_{k_*^{(q)}-l}\left(\mathcal{C}^{(q)}_{t(n^\delta)}\right)|\text{NoBad}] \leq e^{k_*^{(q)}/l_n} \frac{2\hat{u}^{(q)}_{k_*^{(q)}-l}}{3-\tau^{(q)}} \prod_{i=0}^{k_*^{(q)}-l-1} \frac{(\hat{u}_{i+1}^{(q)})^{\gamma^{(q)}(\hat{u}_{i+1}^{(q)})}}{h^{(q)}(\hat{u}_i^{(q)})}. \tag{3.89}$$

Applying Markov's inequality in addition with the union bound we obtain

$$\mathbb{P}_\mathcal{T}\left(\exists l \in [0, k_*^{(q)}] : H^{(q)}(\mathcal{C}^{(q)}_{T_n^{(q)}-l}) \geq \hat{u}^{(q)}_{k_*^{(q)}-l}|\text{NoBad}\right) \leq \frac{2e^{k_*^{(q)}/l_n}}{3-\tau^{(q)}} \sum_{l=0}^{k_*^{(q)}} \prod_{i=0}^{k_*^{(q)}-l-1} \frac{(\hat{u}_{i+1}^{(q)})^{\gamma^{(q)}(\hat{u}_{i+1}^{(q)})}}{h^{(q)}(\hat{u}_i^{(q)})}. \tag{3.90}$$

We have shown in (3.46) and (3.48) that $k_*^{(q)}$ is a tight random variable, and as $l_n = n\mathbb{E}[D_n^{(q)}]$ is of order n the first term is of order $(1 + o_{\mathbb{P}}(1))$. For the summation in (3.90) the bound shown in (3.69) is applied to obtain

$$\mathbb{P}_{\mathcal{T}}\left(\exists l \in [0, k_*^{(q)}] : H^{(q)}(\mathcal{C}^{(q)}_{T_n^{(q)}-l}) \geq \hat{u}^{(q)}_{k_*^{(q)}-l} | \text{NoBad}\right) \leq \tilde{C}^{(q)}\exp\left\{-\tilde{C}^{(q)}(\log m^{(q)})^{\eta^{(q)}}\right\}. \tag{3.91}$$

We will use this bound to show that $\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}}$ and $\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}}$ are $\mathbb{P}_{\mathcal{T}}$-whp disjoint in the following way. As $H^{(\text{out})}(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}-l})$ is largest for $l = 0$, we will expand the forward exploration of $v_{(\text{out})}$ till this point. Next, we look at the number of free inbound half-edges $H^{(\text{in})}(\mathcal{C}^{(\text{in})}_{t(n^{\delta'}+1)}), H^{(\text{in})}(\mathcal{C}^{(\text{in})}_{t(n^{\delta'}+2)}), \ldots, H^{(\text{in})}(\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}})$ of the backward exploration neighborhood of $v_{(\text{in})}$ step by step, and at each of these steps we will look or any of the free inbound half-edges are paired with any of the $H^{(\text{out})}(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}})$ outbound half-edges. As the probability that there is a connection before or at time $T_n^{(\text{in})} - l$ has the same order of magnitude as the probability there is a connection at time $T_n^{(\text{in})} - l$, it is enough to calculate if the latter is the case for any $l \in [0, k_*^{((\text{in}))}]$. Given $H^{(\text{out})}(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}})$ and $H^{(\text{in})}(\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}-l})$, we use the union bound to find the probability is at most

$$\mathbb{P}_{\mathcal{T}}\left(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}} \leftrightarrow \mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}-l} | H(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}}), H(\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}-l})\right) \leq \frac{H(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}})H(\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}-l})}{l_n(1 - o_{\mathbb{P}}(1))}. \tag{3.92}$$

Denote the event

$$\mathcal{D}_n := \left\{H^{(q)}(\mathcal{C}^{(q)}_{T_n^{(q)}-l}) \leq \hat{u}^{(\text{in})}_{k_*^{(q)}-l}, \forall l \in [0, k_*^{(q)}\right\}, \tag{3.93}$$

for which we know from (3.91) that it holds $\mathbb{P}_{\mathcal{T}}$-whp. Using the definition of $\hat{u}_k^{(q)}$ in (3.52) and the expression we found for $k_*^{(q)}$ in (3.46) and (3.48), we obtain

$$\begin{aligned}\mathbb{P}\left(\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}} \cap \mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}} \neq \emptyset | \mathcal{D}_n\right) &\leq \frac{\hat{u}^{(\text{out})}_{k_*^{(\text{out})}}}{l_n(1 - o_{\mathbb{P}}(1))} \sum_{l=1}^{k_*^{(\text{in})}} \hat{u}^{(\text{in})}_{k_*^{(\text{in})}-l} \\ &\leq l_n^{\beta_n^{(\text{out})}(\tau^{(\text{out})}-2)^{b_n^{(\text{out})}}-1} \sum_{l=1}^{k_*^{(\text{in})}} l_n^{\beta_n^{(\text{in})}(\tau^{(\text{in})}-2)^{b_n^{(\text{in})}}+l}.\end{aligned} \tag{3.94}$$

The bound on the RHS of (3.94) converges to zero for

$$\beta_n^{(\text{out})}(\tau^{(\text{out})} - 2)^{b_n^{(\text{out})}} + \beta_n^{(\text{in})}(\tau^{(\text{in})} - 2)^{b_n^{(\text{in})}+1} < 1. \tag{3.95}$$

Note that the expression on the RHS is at most $\beta_n^{(\text{out})} + \beta_n^{(\text{in})}(\tau^{(\text{in})} - 2)$, which we assumed to have a smaller value than 1. This proves that the neighborhoods $\mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}}$ and $\mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}}$ are $\mathbb{P}_{\mathcal{T}}$-whp disjoint, confirming the claim of the lemma. □

3.4 Typical graph distances

Before we continue with the proof of typical distances in the directed configuration model shown in Theorem 1.6.5, let us recap what we have shown so far. From Lemma 3.2.5 we know that $\mathbb{P}_{\mathcal{T}}$-whp (defined below (3.22)) there is a vertex $v^*_{(\text{out})}$ at distance

$$T_{(\text{out})} = -1 + \left\lfloor \frac{\log\log(l_n^{\beta_n^{(\text{out})}}) - \log(Y_n^{(\text{out})})}{|\log(\tau^{(\text{out})} - 2)|} \right\rfloor \tag{3.96}$$

from the typical vertex $v_{(\text{out})}$ with out-degree $d^{(\text{out})}_{v^*_{(\text{out})}} \sim l_n^{\beta_n^{(\text{out})}(\tau^{(\text{out})}-2)^{b_n^{(\text{out})}}}$, and a vertex $v^*_{(\text{in})}$ with distance

$$T_{(\text{in})} = -1 + \left\lfloor \frac{\log\log(l_n^{\beta_n^{(\text{in})}}) - \log(Y_n^{(\text{in})})}{|\log(\tau^{(\text{in})}-2)|} \right\rfloor \tag{3.97}$$

towards the typical vertex $v_{(\text{in})}$ with in-degree $d^{(\text{in})}_{v^*_{(\text{in})}} \sim l_n^{\beta_n^{(\text{in})}(\tau^{(\text{in})}-2)^{b_n^{(\text{in})}}}$. In Lemma 3.3.1 we have shown that the neighborhoods $\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}$ and $\mathcal{C}^{(\text{in})}_{T_{(\text{in})}}$ are $\mathbb{P}_{\mathcal{T}}$-whp disjoint. To complete the proof for the typical distances from Theorem 1.6.5, it thus remains to show the distance from $\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}$ towards $\mathcal{C}^{(\text{in})}_{T_{(\text{in})}}$. As usual, the proof follows by providing an upper bound by showing the existence of a certain path, and a matching lower bound which is found with our familiar path-counting techniques. As the lower bound is easier in this case, we will start with this part.

Before we start counting paths, we show that even considering the out- and inbound half-edges used to construct both $\mathcal{C}^{(q)}_{T_{(q)}}$, that $l_n^* = l_n(1-o_{\mathbb{P}}(1))$ still applies for the number of remaining in- or outbound half-edges. This follows from the event $\mathcal{D}_n$ from (3.93), for which we have shown it holds $\mathbb{P}_{\mathcal{T}}$-whp. This means we can bound the number of unavailable out- or inbound half-edges by the sum of $\sim \hat{u}_k^{(q)}$ over all $k \leq k_*^{(q)}$ and $q \in \{\text{out}, \text{in}\}$. Note that the highest order terms in these sums are $\hat{u}_{k_*^{(q)}} \sim l_n^{\beta_n^{(q)}(\tau^{(q)}-2)^{b_n^{(q)}}}$. As $\beta_n^{(q)} < 1$ and $(\tau^{(q)}-2)^{b_n^{(q)}} \leq 1$, this shows that the property still applies. Hence we can easily apply our familiar path counting techniques for the number of paths starting at any of the $\mathcal{H}^{(\text{out})}(\mathcal{C}^{(\text{out})}_{T_{(\text{out})}})$ free outbound half-edges in $\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}$ or ending at any of the $\mathcal{H}^{(\text{in})}(\mathcal{C}^{(\text{in})}_{T_{(\text{in})}})$ free inbound half-edges in $\mathcal{C}^{(\text{in})}_{T_{(\text{in})}}$. Let $N_z(\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}, \mathcal{C}^{(\text{in})}_{T_{(\text{in})}})$ denote the number of paths of length $z+1$, thus passing z vertices, from $\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}$ towards $\mathcal{C}^{(\text{in})}_{T_{(\text{in})}}$. The expression obtained from the path counting is similar to (3.55), but the in- and out-degree of the vertices along the path have no restrictions. Thus, we obtain

$$\begin{aligned} &\mathbb{E}_{\mathcal{T}}[N_z\left(\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}, \mathcal{C}^{(\text{in})}_{T_{(\text{in})}}\right)|\text{NoBad}] \\ &\qquad \leq e^{z^2/l_n} \sum_{\pi_0 \in \mathcal{C}^{(\text{out})}_{T_n^{(\text{out})}}} d^{(\text{out})}_{\pi_0} \prod_{i=1}^{z} \left(\sum_{\pi_i \in [n]} \frac{d^{(\text{in})}_{\pi_i} d^{(\text{out})}_{\pi_i}}{l_n^*} \right) \sum_{\pi_{z+1} \in \mathcal{C}^{(\text{in})}_{T_n^{(\text{in})}}} d^{(\text{in})}_{\pi_3}. \end{aligned} \tag{3.98}$$

The first and last summation already have bounds, shown in Lemma 3.3.1. To obtain a bound for the remaining summation we will use the restrictions on the in- and out-degree from Assumption 1.6.1, from which the highest order out- and in-degrees are $l_n^{\beta_n^{(\text{out})}}$ and $l_n^{\beta_n^{(\text{in})}}$. Depending whether we restrict the in- out out-degrees, we can apply (3.59) from Lemma 3.2.4 to get

$$\sum_{\pi_i \in [n]} \frac{d^{(\text{in})}_{\pi_i} d^{(\text{out})}_{\pi_i}}{l_n} = \sum_{\substack{\pi_i \in [n] \\ d^{(q)}_{\pi_i} \leq l_n^{\beta_n^{(q)}}}} \frac{d^{(\text{in})}_{\pi_i} d^{(\text{out})}_{\pi_i}}{l_n} \leq l_n^{\beta_n^{(q)}(3-\tau^{(q)}) + \gamma^{(q)}(l_n^{\beta_n^{(q)}})} \sim l_n^{\beta_n^{(q)}(3-\tau^{(q)})}. \tag{3.99}$$

Assumption 1.6.1 has the condition that $\beta_n^{(\text{out})}(3-\tau^{(\text{out})}) = \beta_n^{(\text{in})}(3-\tau^{(\text{in})})$, such that it does not make much of a difference which type of half-edge we restrict. For ease of notation, we define "$\lesssim x$" to mean "less than x times a factor at most $\exp\{\pm \log(x)^\theta\}$ for some $\theta \in (0,1)$". Applying the bound from Lemma 3.3.1 and (3.99) shows that

$$\mathbb{E}_{\mathcal{T}}[N_z\left(\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}, \mathcal{C}^{(\text{in})}_{T_{(\text{in})}}\right)|\text{NoBad}] \lesssim l_n^{-1} l_n^{\beta_n^{(\text{out})}(\tau^{(\text{out})}-2)^{b_n^{(\text{out})}}} l_n^{z\beta_n^{(q)}(3-\tau^{(q)})} l_n^{\beta_n^{(\text{in})}(\tau^{(\text{in})}-2)^{b_n^{(\text{in})}}}. \tag{3.100}$$

With Markov's inequality, we can bound the probability of the event that at least one path of length $z+1$ or less exists from $\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}$ to $\mathcal{C}^{(\text{in})}_{T_{(\text{in})}}$ by

$$\mathbb{P}_{\mathcal{T}}\left(\text{dist}\left(\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}, \mathcal{C}^{(\text{in})}_{T_{(\text{in})}}\right) \leq z+1|\text{NoBad}\right) \lesssim l_n^{-1+\beta_n^{(\text{out})}(\tau^{(\text{out})}-2)^{b_n^{(\text{out})}}+z\beta_n^{(q)}(3-\tau^{(q)})+\beta_n^{(\text{in})}(\tau^{(\text{in})}-2)^{b_n^{(\text{in})}}}. \tag{3.101}$$

To obtain a lower bound for the distance between the neighborhoods, we want to find the smallest value for z such that the bound (3.101) does not go to zero, i.e.,

$$\begin{aligned} z_n^* &= \inf\{z \in \mathbb{N} : \beta_n^{(\text{out})}(\tau^{(\text{out})}-2)^{b_n^{(\text{out})}} + z\beta_n^{(q)}(3-\tau^{(q)}) + \beta_n^{(\text{in})}(\tau^{(\text{in})}-2)^{b_n^{(\text{in})}} > 1\} \\ &= \left\lceil \frac{1-\beta_n^{(\text{out})}(\tau^{(\text{out})}-2)^{b_n^{(\text{out})}} - \beta_n^{(\text{in})}(\tau^{(\text{in})}-2)^{b_n^{(\text{in})}}}{\frac{1}{2}(\beta_n^{(\text{out})}(3-\tau^{(\text{out})}) + \beta_n^{(\text{in})}(3-\tau^{(\text{in})}))} \right\rceil, \end{aligned} \tag{3.102}$$

where we have used that $\beta_n^{(q)}(3-\tau^{(q)}) = \frac{1}{2}(\beta_n^{(\text{out})}(3-\tau^{(\text{out})})+\beta_n^{(\text{in})}(3-\tau^{(\text{in})}))$ for both $q \in \{\text{out}, \text{in}\}$. As the bound (3.101) tends to zero with $n \to \infty$ for all $z < z_n^*$, it follows that $\mathbb{P}_{\mathcal{T}}$-whp there is no path of length $T_n^{(\text{out})} + T_n^{(\text{in})} + z_n^*$ or less, from vertex o_1 to o_2. This shows that $\mathbb{P}_{\mathcal{T}}$-whp

$$\text{dist}_{\text{DCM}_n(\boldsymbol{d})}(o_1, o_2) \geq T_n^{(\text{out})} + T_n^{(\text{in})} + z_n^* + 1. \tag{3.103}$$

This concludes the lower bound on the typical distances in the configuration model.

Upper Bound For the upper bound we want to prove the likely existence of a path from $\mathcal{C}^{(\text{out})}_{T_{(\text{out})}}$ towards $\mathcal{C}^{(\text{in})}_{T_{(\text{in})}}$, for which the length matches the lower bound (3.103). First, note that from Lemma 3.2.5 we know there is a vertex $v^*_{(\text{out})} \in \mathcal{C}^{(\text{out})}_{T_{(\text{out})}}$ with $d^{(\text{out})}_{v^*_{(\text{out})}} \sim l_n^{\beta_n^{(\text{out})}(\tau^{(\text{out})}-2)^{b_n^{(\text{out})}}}$ and a vertex $v^*_{(\text{in})} \in \mathcal{C}^{(\text{in})}_{T_{(\text{in})}}$ with $d^{(\text{in})}_{v^*_{(\text{in})}} \sim l_n^{\beta_n^{(\text{in})}(\tau^{(\text{in})}-2)^{b_n^{(\text{in})}}}$. Instead of looking at all possible paths, we study paths starting at $v^*_{(\text{out})}$ and ending in $v^*_{(\text{in})}$. Moreover, the ith vertex along the path is taken from from a different disjoint set of vertices for all $i \geq 0$. For the configuration model it is shown in ([1], figure 7) that the number of ways two such paths can overlap in the undirected graph, is quite limited. As the possible shapes that two paths form does not change by taking directed paths, we can apply this property to simplify the calculation of the second moment of the number of such paths. This will allow us to apply Chebyshev's inequality to bound the probability such path of length $z+1$ does not exist. To define the number of these disjoint sets of vertices, note that both $b_n^{(\text{out})}, b_n^{(\text{in})} \in [0,1)$, such that

$$z_n^* + 2 \leq \left\lceil \frac{1-\beta_n^{(\text{out})}(\tau^{(\text{out})}-2) - \beta_n^{(\text{in})}(\tau^{(\text{in})}-2)}{\frac{1}{2}(\beta_n^{(\text{out})}(3-\tau^{(\text{out})}) + \beta_n^{(\text{in})}(3-\tau^{(\text{in})}))} \right\rceil + 2 := M_\beta. \tag{3.104}$$

We will divide all the vertices in M_β disjoint sets, for which we denote set $i \in [M_\beta]$ by Δ_i. The vertices will be divided between the M_β sets, such that for some $0 < c_1 < c_1 < \infty$,

$$\begin{aligned} v_i &= \sum_{v \in \Delta_i} \frac{d_v^{(\text{out})} d_v^{(\text{in})}}{l_n} \in \left[\frac{c_1}{M_\beta}, \frac{c_2}{M_\beta}\right] \cdot \sum_{v \in [n]} \frac{d_v^{(\text{out})} d_v^{(\text{in})}}{l_n} \\ \kappa_i^{(\text{in})} &= \sum_{v \in \Delta_i} \frac{d_v^{(\text{out})} d_v^{(\text{in})}(d_v^{(\text{in})}-1)}{l_n} \in \left[\frac{c_1}{M_\beta}, \frac{c_2}{M_\beta}\right] \cdot \sum_{v \in [n]} \frac{d_v^{(\text{out})} d_v^{(\text{in})}(d_v^{(\text{in})}-1)}{l_n} \\ \kappa_i^{(\text{out})} &= \sum_{v \in \Delta_i} \frac{d_v^{(\text{out})} d_v^{(\text{in})}(d_v^{(\text{out})}-1)}{l_n} \in \left[\frac{c_1}{M_\beta}, \frac{c_2}{M_\beta}\right] \cdot \sum_{v \in [n]} \frac{d_v^{(\text{out})} d_v^{(\text{in})}(d_v^{(\text{out})}-1)}{l_n}. \end{aligned} \tag{3.105}$$

Let $v_*^{(\text{out})} \in \Delta_0$ and $v_*^{(\text{in})} \in \Delta_{M_\beta}$. Let $N_z^o\left(v_*^{(\text{out})}, v_*^{(\text{in})}\right)$ denote the number of paths on vertices $(v_*^{(\text{out})} = v_0, v_1, ..., v_{z-1}, v_*^{(\text{in})} = v_z)$, for which $v_j \in \Delta_j$ when $j \leq z/2$ and $v_j \in \Delta_{M_\beta+1-j}$ when

$j > z/2$. Let us start by obtaining a lower and upper bound on the expected number of paths $N_z^o(v_*^{(\mathrm{out})}, v_*^{(\mathrm{in})})$. From Assumption 1.6.1 it follows that $F_n^{(q)}(l_n^{\beta_n^{(q)}}) = 1$. Following similar steps as the proof of (3.59) in Lemma 3.2.4 shows that

$$\begin{aligned}
\mathbb{E}[D_n^{*(q)}] = \nu^{(q)}(l_n^{\beta_n^{(q)}}) &= \sum_{\substack{\pi_i \in [n] \\ d_{\pi_i}^{(q)} \le l_n^{\beta_n^{(q)}}}} \frac{d_{\pi_i}^{(\mathrm{in})} d_{\pi_i}^{(\mathrm{out})}}{l_n} \\
&= \sum_{k=1}^{l_n^{\beta_n^{(q)}}} [F_n^{(q)}(l_n^{\beta_n^{(q)}}) - F_n^{(q)}(k-1)] = \sum_{k=1}^{l_n^{\beta_n^{(q)}}} [1 - F_n^{(q)}](k-1) \\
&= \sum_{s=0}^{l_n^{\beta_n^{(q)}}} [1 - F_n^{(q)}](s) \ge \sum_{s=0}^{l_n^{\beta_n^{(q)}}} s^{2-\tau^{(q)}} e^{-C^{(q)} \log(s)^{\gamma^{(q)}}} \\
&\ge \frac{1}{2(3-\tau^{(q)})} l_n^{\beta_n^{(q)}(3-\tau^{(q)})} e^{-C^{(q)} \log(l_n^{\beta_n^{(q)}})^{\gamma^{(q)}}},
\end{aligned} \tag{3.106}$$

where we applied the lower bound from Assumption 1.6.1 for the first inequality, and Karamata's Theorem for the second inequality. Combined with the bound shown in (3.99) this shows that $\nu^{(q)}(l_n^{\beta_n^{(q)}}) \sim l_n^{\beta_n^{(q)}(3-\tau^{(q)})}$. The expected number of restricted paths can be obtained similarly as to (3.98). The difference is that the path starts at $v_*^{(\mathrm{out})}$ and ends at $v_*^{(\mathrm{in})}$, and we need to apply the restriction on the middle summation that $\pi_j \in \Delta_j$ for $j \le z/2$ and $\pi_j \in \Delta_{M_\beta+1-j}$ when $j > z/2$. Applying these changes yields

$$\mathbb{E}_{\mathcal{T}}[N_z^o(v_{(\mathrm{out})}^*, v_{(\mathrm{in})}^*)] \mathrel{\underset{\sim}{\in}} l_n^{-1} l_n^{\beta_n^{(\mathrm{out})}(\tau^{(\mathrm{out})}-2)^{b_n^{(\mathrm{out})}}} l_n^{z\beta_n^{(q)}(3-\tau^{(q)})} l_n^{\beta_n^{(\mathrm{in})}(\tau^{(\mathrm{in})}-2)^{b_n^{(\mathrm{in})}}} [(c_1/M_\beta)^z, (c_2/M_\beta)^z], \tag{3.107}$$

where $\underset{\sim}{\in}$ means the value is contained in an interval, where an addition factor of at most $\exp\{\pm(\log l_n^{\beta_n^{(q)}})^\theta\}$ for some $\theta < 1$ could be multiplied with the prefactor of the interval. From the definitions of z_n^* and M_β in (3.102) and (3.104), it follows from the bound $\beta_n^{(q)} \ge (\log n)^{-\eta}$ that for all $z \le z_n^*$ and both $i \in \{1, 2\}$

$$(c_i/M_\beta)^z \ge \exp\{\log(c_i) - \eta \log\log(n)(\log n)^\eta\} \ge \exp\{-(\log n)^\theta\}, \tag{3.108}$$

for $\theta \in (\eta, 1)$ and n sufficiently large. Applying this bound on both sides of the interval from (3.107), we obtain

$$\mathbb{E}_{\mathcal{T}}[N_z^o(v_{(\mathrm{out})}^*, v_{(\mathrm{in})}^*)] \sim l_n^{-1} l_n^{\beta_n^{(\mathrm{out})}(\tau^{(\mathrm{out})}-2)^{b_n^{(\mathrm{out})}}} l_n^{z\beta_n^{(q)}(3-\tau^{(q)})} l_n^{\beta_n^{(\mathrm{in})}(\tau^{(\mathrm{in})}-2)^{b_n^{(\mathrm{in})}}}. \tag{3.109}$$

The smallest value for z such that this value does not converge to zero is z_n^* as defined in (3.102).

We now use the restricted paths and Chebyshev's inequality to bound the probability there is no path of length $z + 1$ from $v_{(\mathrm{out})}^*$ to $v_{(\mathrm{in})}^*$:

$$\mathbb{P}_{\mathcal{T}}\left(N_z(v_{(\mathrm{out})}^*, v_{(\mathrm{in})}^*) = 0\right) \le \mathbb{P}_{\mathcal{T}}\left(N_z^o(v_{(\mathrm{out})}^*, v_{(\mathrm{in})}^*) = 0\right) \le \frac{\mathrm{Var}_{\mathcal{T}}\left(N_z^o(v_{(\mathrm{out})}^*, v_{(\mathrm{in})}^*)\right)}{\mathbb{E}_{\mathcal{T}}[N_z^o(v_{(\mathrm{out})}^*, v_{(\mathrm{in})}^*)]^2}. \tag{3.110}$$

To obtain a bound on the variance of these restricted paths, we can closely follow the proof of ([1], Lemma 7.1 (7.5)) with similar adjustments made in ([15], Section 4), which bound the expected number and variance of the number of similarly restricted paths for the undirected configuration model. The main difference which needs to be considered is the asymmetry that results from the directed edges. In the undirected proof, κ_i is used to indicate the number of ways two paths can disperse or meet at the ith vertex, which they call the start or the end of an excursion of one of the paths. In the directed graph we need to distinguish between the start and the end of these excursions. The reason is that for $i \notin \{0, z\}$, the vertex at the start of an excursion is entered from 1 inbound half-edge and left from 2 outbound half-edges, and the vertex at which the excursion ends is entered from 2 inbound half-edges and entered from 1 outbound half-edge. For this reason we use $\kappa_i^{(\text{out})}$ to indicate the number of ways the paths can disperse at the ith vertex, and $\kappa_i^{(\text{in})}$ for the number of ways they can meet.

We omit the proof of the bound on the variance, and simply direct the reader to ([1], Lemma 7.1 (7.5)). For the directed setting the structure can be followed step by step. The most notable difference is in ([1], equation (A.25)) which shows a bound on the number of combinatorial factors to pick half-edges for each shape formed by the two paths. From the earlier mentioned distinction between the start and end of a excursion, we obtain the expression

$$\begin{aligned} & d^{(\text{out})}_{v^*_{(\text{out})}}(d^{(\text{out})}_{v^*_{(\text{out})}}-1)^{\delta_{(\text{out})}} d^{(\text{in})}_{v^*_{(\text{in})}}(d^{(\text{in})}_{v^*_{(\text{in})}}-1)^{\delta_{(\text{in})}} \prod_{\substack{s \\ d_\sigma(\pi_s)=(1,2)}} d^{(\text{out})}_{v_s} d^{(\text{in})}_{v_s}(d^{(\text{in})}_{v_s}-1) \\ & \times \prod_{\substack{t \\ d_\rho(\pi_t)=(2,1)}} d^{(\text{in})}_{v_t} d^{(\text{out})}_{v_t}(d^{(\text{out})}_{v_t}-1) \prod_{\substack{i \\ d_\sigma(\pi_i)=(1,1)}} d^{(\text{in})}_{\pi_i} d^{(\text{out})}_{\pi_i} \\ & \times \prod_{\substack{d_\sigma(\rho_u)=(1,1) \\ \rho_u\cap\pi=\emptyset}} d^{(\text{in})}_{v_u} d^{(\text{out})}_{v_u}, \end{aligned} \tag{3.111}$$

where $\delta_{\text{out}} = \mathbb{1}\{\text{an excursion starts at } v^*_{(\text{out})}\}$ and $\delta_{\text{in}} = \mathbb{1}\{\text{an excursion ends at } v^*_{(\text{in})}\}$ and $d_\sigma(v)$ denotes the number of in- and outbound half-edges vertex v uses in the shape formed by the two paths. So by keeping in mind where to use $\kappa_i^{(\text{in})}$ and $\kappa_i^{(\text{out})}$ it can be shown that

$$\begin{aligned} \operatorname{Var}\left(N^o_{z^*_n}(v^*_{(\text{out})}, v^*_{(\text{in})})\right) &\le \mathbb{E}_{\mathcal{T}}[N^o_{z^*_n}(v^*_{(\text{out})}, v^*_{(\text{in})})] \\ & + \overline{\mathbb{E}_{\mathcal{T}}[N^o_{z^*_n}(v^*_{(\text{out})}, v^*_{(\text{in})})]}^2 \left(\frac{\nu_1}{\nu_1-\tilde{C}} \frac{C}{\nu_1^2}\left(\frac{\kappa_1^{(\text{out})}}{u^{(\text{out})}_{k^*_{(\text{out})}}} + \frac{\kappa_1^{(\text{in})}}{u^{(\text{in})}_{k^*_{(\text{in})}}}\right) + \frac{\nu_1^2}{(\nu_1-\tilde{C})^2} \frac{C^2\kappa_1^{(\text{in})}\kappa_1^{(\text{out})}}{\nu_1^4} \frac{1}{u^{(\text{out})}_{k^*_{(\text{out})}} u^{(\text{in})}_{k^*_{(\text{in})}}} \right. \\ & \left. + \frac{8(z^*_n)^2}{l_n} + \left(1 + \frac{C\kappa_1^{(\text{out})}\nu_1}{\nu_1^2 u^{(\text{out})}_{k^*_{(\text{out})}}}\right)\left(1 + \frac{C\kappa_1^{(\text{in})}\nu_1}{\nu_1^2 u^{(\text{in})}_{k^*_{(\text{in})}}}\right) \frac{z^*_n}{\nu_1-\tilde{C}} \left(2\frac{z^*_n\nu_1}{l_n}\frac{C^2\kappa_1^{(\text{out})}\kappa_1^{(\text{in})}}{\nu_1^4}\right)\right), \end{aligned} \tag{3.112}$$

where the $\overline{\mathbb{E}_{\mathcal{T}}[N^o_{z^*_n}(v^*_{(\text{out})}, v^*_{(\text{in})})]}$ indicates the upper bound given in (3.109). Applying similar methods as we have used to show that $\nu^{(q)}(l_n^{\beta_n^{(q)}}) \sim l_n^{\beta_n^{(q)}(3-\tau^{(q)})}$, it can be shown for $p, q \in \{\text{out}, \text{in}\}$ with $p \neq q$, that

$$\sum_{\substack{v\in[n] \\ d^{(q)}_u \le l_n^{\beta_n^{(q)}}}} \frac{d^{(p)}_u d^{(q)}_u (d^{(q)}_u - 1)}{l_n} \sim l_n^{\beta_n^{(q)}(4-\tau^{(q)})}. \tag{3.113}$$

From the definitions given in (3.105), this implies that $\kappa_1^{(q)} \sim l_n^{\beta_n^{(q)}(4-\tau^{(q)})}$ and $\nu_1 \sim l_n^{\beta_n^{(q)}(3-\tau^{(q)})}$. As $u^{(q)}_{k^*_{(q)}} \sim l_n^{\beta_n^{(q)}(\tau^{(q)}-2)^{b_n^{(q)}}}$, it follows that for both $q \in \{\text{out}, \text{in}\}$

$$\frac{\kappa_1^{(q)}}{\nu_1^2}\frac{1}{u^{(q)}_{k^*_{(q)}}} \sim l_n^{\beta_n^{(q)}((\tau^{(q)}-2)-(\tau^{(q)}-2)^{b_n^{(q)}})}, \tag{3.114}$$

where the error hidden in the $\sim$ symbol is at most $\exp\{\pm\log(n^{\beta_n^{(q)}})^\theta\}$ for some $\theta < 1$. Since $b_n^{(q)} < 1$ it follows that $(\tau^{(q)}-2)-(\tau^{(q)}-2)^{b_n^{(q)}} < 0$, such that (3.114) converges to zero as $n \to \infty$.

This shows that both terms in the first line of (3.112) that is multiplied with $\overline{\mathbb{E}_{\mathcal{T}}[N_{z_n^*}(v^*_{(\mathrm{out})}, v^*_{(\mathrm{in})})]}^2$ converges to zero as $n \to \infty$. Similarly, for the main contribution in the second line of (3.114) we find

$$\frac{\kappa_1^{(q)}}{\nu_1 u^{(q)}_{k^*_{(q)}}} \frac{\kappa_1^{(\mathrm{out})}\kappa_1^{(\mathrm{in})}}{\nu_1^4} \frac{1}{l_n} \sim l_n^{\beta_n^{(q)}(1-(\tau^{(q)}-2)^{b_n^{(q)}}} l_n^{\beta_n^{(\mathrm{out})}(\tau^{(\mathrm{out})}-2)+\beta_n^{(\mathrm{in})}(\tau^{(\mathrm{in})}-2)} l_n^{-1}. \tag{3.115}$$

Given that $\beta_n^{(q)} + \beta_n^{(p)}(\tau^{(p)} - 2) < 1$ from Assumption 1.6.1, it can be shown that the exponent of (3.115) is negative. Hence, this term also converges to zero as $n \to \infty$. Combining both estimates (3.114) and (3.115) it can be shown that the variance of $N^o_{z_n^*}(v^*_{(\mathrm{out})}, v^*_{(\mathrm{in})})$ is of smaller order than its expectation squared, such that the bound on the RHS of (3.112) converges to zero. This proves that $\mathbb{P}_{\mathcal{T}}$-whp there exists a path of length $z_n^* + 1$ from $\mathcal{C}^{(\mathrm{out})}_{T_{(\mathrm{out})}}$ to $\mathcal{C}^{(\mathrm{in})}_{T_{(\mathrm{in})}}$. Combined with the lower bound on the typical distances from (3.103), this shows that

$$\begin{aligned}\mathrm{dist}_{\mathrm{DCM}_n(\boldsymbol{d})}(v_{(\mathrm{out})}, v_{(\mathrm{out})}) &= T_n^{(\mathrm{out})} + T_n^{(\mathrm{in})} + z_n^* + 1 \\ &= T_n^{(\mathrm{out})} + T_n^{(\mathrm{in})} + \left\lceil \frac{1 - \beta_n^{(\mathrm{out})}(\tau^{(\mathrm{out})} - 2)^{b_n^{(\mathrm{out})}} - \beta_n^{(\mathrm{in})}(\tau^{(\mathrm{in})} - 2)^{b_n^{(\mathrm{in})}}}{\frac{1}{2}(\beta_n^{(\mathrm{out})}(3 - \tau^{(\mathrm{out})}) + \beta_n^{(\mathrm{in})}(3 - \tau^{(\mathrm{in})}))} \right\rceil + 1.\end{aligned} \tag{3.116}$$

this completes the proof of Theorem 1.6.5.

Chapter 4

Conclusions

In this thesis we have studied the conditions under which a directed network experiences the ultra-small phenomenon, and how the distances fluctuate with the size of the network. For the mathematical model we have chosen to use the directed configuration model, as its simple construction makes it easy to recreate the desired characteristics of the network.

The first main result of the thesis in Theorem 1.5.4 shows the conditions under which the typical distances in the directed configuration model are ultra-small. The first one being that the degree distribution converges in distribution to a random variable with finite mean. This ensures that the size of the typical neighborhoods, consisting of either the vertices that can reach the typical vertex, or those that can be reached from the typical vertex in a certain number of steps, stays bounded. Under this condition, both of these neighborhoods have a tree-like structure and can be approximated by properly defined branching processes.

Next, we add two additional conditions. Namely, that each vertex has at least one in- and out-degree, and that the forward out- and in-degrees follow a power-law, with power-law exponents $\tau^{(\text{out})}, \tau^{(\text{in})} \in (2, 3)$. These conditions have a number of consequences, which lead to the desired phenomenon. For one, the graph contains a unique strongly connected component, meaning that a path is likely to exist between typical vertices. Moreover, each of the highest order out-degree vertices in the outbound hub, are likely connected with each of the highest order in-degree vertices from the inbound hub. And finally, the covariance between the in- and out-degree of a typical vertex is infinite. From this positive correlation between the in- and out- degree, it follows that the offspring distribution of both branching processes have an infinite mean, which experience super-exponential generational growth. As these generation sizes approximate the sizes of the previous mentioned neighborhoods, this describes the size of these neighborhoods as well. This further implies that the path between typical vertices passes vertices with respective larger and large out- or in-degree, the further we move away from the vertices at the start and end of the path. More precise, a path for which the out-degree increases doubly-exponential with each vertex we move forward till we reach the outbound hub, which is connected to the inbound hub. Then from the inbound hub, the in-degree decreases super-exponentially with each vertex till we reach the target vertex. This type of path describes the typical paths under the mentioned conditions, and explains how the double logarithmic distances arise.

The second main result in Theorem 1.6.5 describes how these typical distances fluctuate with changes to the graph size, after adding a possible truncation to the in- and out-degrees. Naturally, the values at which the in- and out-degrees are truncated depends on how heavily tailed their distributions are. As we want the same amount of inbound as outbound half-edges, it follows that the more heavily tailed distributed type needs to be truncated at a lower value. Moreover, under the truncation the covariance between the in- and out-degree must remain infinite, to preserve the ultra-small distances. Depending on these truncation values, the outbound hub might not be directly connected with the inbound hub anymore. Hence, the expression of the typical distances consists of a third factor, representing the remaining distance between the hubs.

The subject of this thesis was inspired by previous research on both the directed and the undirected configuration model. For the undirected configuration model it was shown that, conditioned on the vertices being in the same component, the typical distances scale doubly logarithmic if the degree distribution has infinite variance. For the directed configuration model it is shown under a similar setting, for which the in- and out-degree distribution have infinite variance, this is not sufficient to result in ultra-small typical distances.

To understand the difference, when the variance of the degrees in the undirected configuration model increases, presumably the number of vertices with extremely large degrees grows along with it. These vertices then act as short-cuts between the typical vertices in the graph, resulting in the small distances. While increasing the variance of the in- and out-degree in the directed configuration model also presumably increases the number of vertices with extremely large out- or in-degree, these two qualities are not likely on the same vertex. It follows that without some dependence between the in- and out-degrees, this property fails to result in the appearance of short-cuts. The resulting typical distances after adding the infinite covariance condition is what we expected to find at the start of the study. The typical distances in the undirected configuration model consist of two identical factors, which can be seen to represent the distance for the starting vertex to the highest order degree vertices called the hub, and the distance from the hub to the target vertex. As we had expected, the expression for the directed configuration model consists of two differing factors, which follows from the asymmetry between the in- and out-degree distributions. First, we have the factor representing the distance to the outbound hub, which depends on the power-exponent of the out-degree distribution. And the second factor representing the distance from the inbound hub to the target vertex, depending on the power-exponent of the in-degree distribution.

While many real world networks are directed, for research they are often modeled as undirected networks. This is presumably due to the complexity of the directed setting, which makes it difficult to find a suitable model to represent the characteristics of the model. With the results of this thesis we take a step forward in understanding how the ultra-small phenomenon arises in real-world directed networks, and thus improving our insight in how to model such networks.

Open questions:

- To obtain the ultra-small distances in the directed configuration model, we assumed that both the forward out- and in-degree follow a power law with power-exponent $\tau^{(\text{out})}, \tau^{(\text{in})} \in (2, 3)$. It would be interesting to see how the typical distances behave when this property only holds for one type of half-edge. For the ultra-small distances, it will still be required that the in- and out-degree have infinite covariance, but this is possible even if the distribution of the other type of half-edge has finite variance.

- For both truncation values $\beta_n^{(q)}$ with $q \in \{\text{out}, \text{in}\}$, we set the condition that $\beta_n^{(q)}(\log n)^\eta \to \infty$ for some $\eta \in (0, 1)$. It is possible that the same results can be shown under the weaker condition $\beta_n^{(q)} \log n \to \infty$, which is the weakest condition under which the second moment and the covariance of the in- and out-degrees are infinite. This could provide a perfect interpolation between the logarithmic and doubly logarithmic distances.

- For the undirected configuration model of size n it has been shown that for power-law exponent $\tau = 3$, the typical distance are of order $\log n / \log \log n$. It would be interesting to see if the directed configuration model experiences similar typical distances when $\tau^{(\text{out})} = \tau^{(\text{in})} = 3$. Moreover, what would happen for each case when $\tau^{(\text{out})}, \tau^{(\text{in})} \in \{2, 3\}$.

Bibliography

[1] Enrico Baroni, Remco van der Hofstad, and Júlia Komjáthy. Fixed speed competition on the configuration model with infinite variance degrees: Unequal speeds. *Electronic Journal of Probability*, 20, 11 2015. 62, 63, 64

[2] N H Bingham, C M Goldie, and J L Teugels. *Regular Variation.* Cambridge University Press, Cambridge, 1987. 11, 55

[3] Bela Bollobas. A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs. Technical report, 1980. 2, 17

[4] Anna D. Broido and Aaron Clauset. Scale-free networks are rare. *Nature Communications*, 10(1), 12 2019. 1

[5] Reuven Cohen and Shlomo Havlin. Scale-Free Networks Are Ultrasmall. *Physical Review Letters*, 90(5):4, 2003. 2

[6] Colin Cooper and Alan Frieze. The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combinatorics Probability and Computing*, 13(3):319–337, 5 2004. 9

[7] P L Davies. The Simple Branching Process: A Note on Convergence When the Mean Is Infinite. Technical Report 3, 1978. 3, 12, 21, 34, 46

[8] K. I. Goh, E. Oh, B. Kahng, and D. Kim. Betweenness centrality correlation in social networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 67(1):4, 2003. 2

[9] Xuhong Liao, Athanasios V. Vasilakos, and Yong He. Small-world human brain networks: Perspectives and challenges, 6 2017. 1

[10] S. Milgram. The Small World Problem. *Psychology Today.*, 1967. 1

[11] Michael Molloy and Bruce Reed. The Size of the Giant Component of a Random Graph with a Given Degree Sequence. *Combinatorics Probability and Computing*, 7(3):295–305, 1998. 20

[12] M. E.J. Newman. Spread of epidemic disease on networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 66(1), 7 2002. 2

[13] Mathew D. Penrose. The strong giant in a random digraph. *Journal of Applied Probability*, 53(1):57–70, 3 2016. 9, 10

[14] Remco Van der Hofstad. Random graphs and complex networks. Technical report, 2022. 2, 7, 17, 18, 19, 20, 21, 26, 32, 35

[15] Remco van der Hofstad and Júlia Komjáthy. When is a Scale-Free Graph Ultra-Small? *Journal of Statistical Physics*, 169(2):223–264, 10 2017. 3, 12, 48, 49, 63

[16] Pim van der Hoorn and Mariana Olvera-Cravioto. Typical distances in the directed configuration model, 6 2018. 2

[17] Ivan Voitalov, Pim Van Der Hoorn, Remco Van Der Hofstad, and Dmitri Krioukov. Scale-free networks well done. *Physical Review Research*, 1(3), 10 2019. 1, 2

[18] Duncan J Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness.* Princeton University Press, 2018. 1