An Axiomatic and an Average-Case Analysis of Algorithms and Heuristics for Metric Properties of Graphs

Michele Borassi*   Pierluigi Crescenzi†   Luca Trevisan ‡

Abstract

In recent years, researchers proposed several algorithms that compute metric quantities of real-world complex networks, and that are very efficient in practice, although there is no worst-case guarantee.

In this work, we propose an axiomatic framework to analyze the performances of these algorithms, by proving that they are efficient on the class of graphs satisfying certain properties. Furthermore, we prove that these properties are verified asymptotically almost surely by several probabilistic models that generate power law random graphs, such as the Configuration Model, the Chung-Lu model, and the Norros-Reittu model. Thus, our results imply average-case analyses in these models.

For example, in our framework, existing algorithms can compute the diameter and the radius of a graph in subquadratic time, and sometimes even in time $n^{1+o(1)}$. Moreover, in some regimes, it is possible to compute the k most central vertices according to closeness centrality in subquadratic time, and to design a distance oracle with sublinear query time and subquadratic space occupancy.

In the worst case, it is impossible to obtain comparable results for any of these problems, unless widely-believed conjectures are false.

1   Introduction.

We study problems motivated by network analysis, such as computing the diameter of a graph, the radius, the closeness centrality, and so on. All these problems admit polynomial-time algorithms, based on computing the distance between all pairs of vertices. These algorithms, however, do not terminate in reasonable time if the input is a real-world graph with millions of nodes and edges. Such worst-case inefficiency is probably due to complexity-theoretic bottlenecks: indeed, a faster algorithm for any of these problems would falsify widely believed conjectures [42, 36, 1, 7, 12, 2].

In practice, these problems are solved via heuristics and algorithms that do not offer any performance guarantee, apart from empirical evidence. These algorithms are widely deployed, and they are implemented in major graph libraries, like Sagemath [41], Webgraph [8], NetworKit [40], and SNAP [30].

In this work, we develop a theoretical framework in which these algorithms can be evaluated and compared. Our framework is axiomatic in the sense that we define some properties, we experimentally show that these properties hold in most real-world graphs, and we perform a worst-case analysis on the class of graphs satisfying these properties. The purpose of this analysis is threefold: we validate the efficiency of the algorithms considered, we highlight the properties of the input graphs that are exploited, and we perform a comparison that does not depend on the specific dataset used for the evaluation. A further confirmation of the validity of this approach comes from the results obtained, that are very similar to existing empirical results.

Furthermore, we show that the properties are verified on some models of random graphs, asymptotically almost surely (a.a.s.), that is, with probability that tends to 1 as the number of nodes n goes to infinity: as a consequence, all results can be turned into average-case analyses on these models, with no modification. This modular approach to average-case complexity analysis has two advantages: since our properties are verified by different models, we can prove results in all these models with a single worst-case analysis. Furthermore, we clearly highlight which properties of random graphs we are using: this way, we can experimentally validate the choice of the probabilistic model, by showing that these properties are reflected by real-world graphs.

In the past, most average-case analyses were performed on the Erdös-Renyi model, which is defined by fixing the number n of nodes, and connecting each pair of nodes with probability p [26, 37, 46, 32]. However many algorithms that work well in practice have poor average-case running time on this model.[1] Indeed, these

---

*IMT School for Advanced Studies Lucca

†University of Florence

‡EECS Department and Simons Institute, U.C. Berkeley

[1]The poor performances of some of these algorithms in the Erdös-Renyi model were empirically shown in [22], and they can be proved with a simple adaptation of the analysis in this paper.

algorithms are efficient if there are some nodes with very high degree, and such nodes are not present in Erdös-Renyi graphs. Conversely, most real-world graphs contain such nodes, because their degree distribution is power law [6], that is, the number of vertices with degree d is close to $\frac{n}{d^\beta}$ for some $\beta > 1$. For this reason, we only consider models that generate power law random graphs. Our framework encompasses almost all values of β, and many of these models: the Configuration Model [9], and Rank-1 Inhomogeneous Random Graph models ([45], Chapter 3), such as the Chung-Lu [20] and the Norros-Reittu model [35].

Our approach is based on four properties: one simply says that the degree distribution is power law, and the other three study the behavior of $\boldsymbol{\tau}_s\left(n^x\right)$, which is defined as the smallest integer ℓ such that the number of vertices at distance ℓ from s is at least n^x. The first of these properties describes the typical and extremal behavior of $\boldsymbol{\tau}_s\left(n^x\right)$, where s ranges over all vertices in the graph. The next two properties link the distance between two vertices s and t with $\boldsymbol{\tau}_s\left(n^x\right) + \boldsymbol{\tau}_t\left(n^y\right)$: informally, $\operatorname{dist}(s,t)$ is close to $\boldsymbol{\tau}_s\left(n^x\right) + \boldsymbol{\tau}_t\left(n^{1-x}\right)$. We prove that these properties are verified in the aforementioned graph models.

The definition of these properties is one of the main technical contributions of this work: they do not only validate our approach, but they also provide a very simple way of proving other metric properties of random graphs, and their use naturally extends to other applications. Indeed, the proof of our probabilistic analysis is very simple, when one assumes these properties. On the other hand, the proof of the properties is very technical, and it uses different techniques in the regimes $\beta > 2$, and $1 < \beta < 2$. In the regime $\beta > 2$, the main technical tool used is branching processes: it is well-known that the size of neighborhoods of a given vertex in a random graph resembles a branching process [35, 25, 10, 45, 44], but this tool was almost always used either as an intuition [25, 45, 44] (and different techniques were used in the actual proof), or it was applied only for specific models, such as the Norros-Reittu model [35]. Conversely, in this work, we provide a quantitative result, from which we deduce the proof of the properties. In the regime $\beta < 2$, the branching process approximation does not hold anymore (indeed, the distribution of the branching process is not even defined). For this reason, in the past, very few results were obtained in this case [43]. In this work, we overcome this difficulty through a different technique: we prove that the graph contains a very dense "core" made by the nodes with highest degree, and the distance between two nodes s, t is almost always the length of a shortest path from s to the core, and from the core to t. This technique lets us compute

the exact value of the diameter, and it lets us prove that the asymptotics found in [43] for the Configuration Model also hold in other models.

Assuming the four properties, we can easily prove consequences on the main metric properties of the graphs $G = (V, E)$ under consideration: we start by estimating the eccentricity of a given vertex s, which is defined as $\operatorname{ecc}(s) = \max_{t \in V} \operatorname{dist}(s,t)$. From this result, we can estimate the diameter $D = \max_{s \in V} \operatorname{ecc}(s)$. Similarly, we can estimate the *farness* $f(s)$ of s, that is, $\sum_{t \in V} \operatorname{dist}(s,t)$, the *closeness centrality* of s, which is defined as $\frac{1}{f(s)}$, and the average distance between two nodes. By specializing these results to the random graph models considered, we retrieve known asymptotics for these quantities, and we prove some new asymptotics in the regime $1 < \beta < 2$.

After proving these results, we turn our attention to the analysis of many heuristics and algorithms, by proving all the results in Table 1 (a plot of the results is available in Figure 1).[2] For approximation algorithms, we usually know the running time and we analyze the error; conversely, for exact algorithms, we bound the running time. All algorithms analyzed are exactly the algorithms published in the original papers, apart from the SUMSH and the SUMS, where we need a small variation to make the analysis work.

In many regimes, our results improve corresponding worst-case bounds: indeed, under reasonable complexity assumptions, for any $\varepsilon > 0$, there is no algorithm that computes a $\frac{3}{2} - \varepsilon$-approximation of the diameter or the radius in $\mathcal{O}\left(n^{2-\varepsilon}\right)$ [36, 12, 2], the complexity of computing the most closeness central vertex is $\Omega\left(n^{2-\varepsilon}\right)$ [2], and there are hardness results on the possible trade-offs between space needed and query time in distance oracles [42, 39]. The difference is very significant, both from a theoretical and from a practical point of view: for instance, we can compute the diameter and the radius of a graph in linear time, in many regimes. This means that, on standard real-world graphs with millions of nodes, the heuristic are millions of times faster than the standard quadratic algorithms. It is also worth mentioning that our results strongly depend on the exponent β: in particular, there are two phase transitions corresponding to $\beta = 2$ and $\beta = 3$. This is due to the fact that, if $1 < \beta < 2$, the average degree is unbounded, if $2 < \beta < 3$, the average degree is finite, but the vari-

---

[2]Some of the results contain a value $o(1)$: this value comes from the four properties, which depend on a parameter ε. In random graphs, this notation is formally correct: indeed, we can let ε tend to 0, since the properties are satisfied a.a.s. for each ε. In real-world graphs, we experimentally show that these properties are verified for small values of ε, and with abuse of notation we write $o(1)$ to denote a function bounded by $c\varepsilon$, for some constant c.

Table 1: a summary of the results of our probabilistic analyses. The value of the constant C is $\frac{2\operatorname{dist}_{avg}(n)}{D-\operatorname{dist}_{avg}(n)}$, where D is the diameter of the graph, $\operatorname{dist}_{avg}(n)$ is the average distance. The values marked with $(*)$ are proved using further characteristics of the probabilistic models.

Parameter	Algorithm	Running time		
		$\beta > 3$	$2 < \beta < 3$	$1 < \beta < 2,$
Diameter (lower bound)	BFS from n^γ random nodes	$\Theta(n^{1+\gamma})$ $\varepsilon_{rel} = \frac{1-\gamma+o(1)}{2+C}$	$\Theta(n^{1+\gamma})$ $\varepsilon_{rel} = \frac{1-\gamma+o(1)}{2}$	$\Theta(mn^\gamma)$ $\varepsilon_{abs} = \left\lvert \frac{2(\beta-1)}{2-\beta}\right\rvert - \left\lvert\frac{(\gamma+1)(\beta-1)}{2-\beta}\right\rvert$
Diameter (lower bound)	2-Sweep [31]	$\Theta(n)$ $\varepsilon_{rel} = o(1)$	$\Theta(n)$ $\varepsilon_{rel} = o(1)$	$\Theta(m)$ $\varepsilon_{abs} \le \begin{cases} 1 & D \text{ even} \\ 2 & D \text{ odd}\end{cases}$
Diameter (lower bound)	RW [36]	$\Theta(n^{\frac{3}{2}}\log n)$ $\varepsilon_{rel} = o(1)$	$\Theta(n^{\frac{3}{2}}\log n)$ $\varepsilon_{rel} = o(1)$	$\Theta(m\sqrt{n}\log n)$ $\varepsilon_{abs} \le \begin{cases} 1 & D \text{ even} \\ 2 & D \text{ odd}\end{cases}$
All eccentricities (lower bound)	SumSH [13]	$n^{1+o(1)}$ $\varepsilon_{abs} = 0$	$n^{1+o(1)}$ $\varepsilon_{abs} = 0$	$\le mn^{1-\frac{2-\beta}{\beta-1}\left(\left\lfloor\frac{\beta-1}{2-\beta}-\frac{3}{2}\right\rfloor-\frac{1}{2}\right)}$ $\varepsilon_{abs} = 0$
Diameter	iFub [22]	$\le n^{1+\left(\frac{1}{2}-\frac{1}{\beta-1}\right)C+o(1)}$	$n^{1+o(1)}$	$\le mn^{1-\frac{2-\beta}{\beta-1}\left\lfloor\frac{\beta-1}{2-\beta}-\frac{1}{2}\right\rfloor+o(1)}$
Diameter	SumS [13, 14]	$\le n^{1+\frac{C}{C+\frac{\beta-1}{\beta-3}}}$ (*)	$n^{1+o(1)}$	$\le mn^{1-\frac{2-\beta}{\beta-1}\left(\left\lfloor\frac{\beta-1}{2-\beta}-\frac{3}{2}\right\rfloor-\frac{1}{2}\right)}$
Radius	SumS [13, 14]	$n^{1+o(1)}$	$n^{1+o(1)}$	$\le mn^{1-\frac{2-\beta}{\beta-1}\left(\left\lfloor\frac{\beta-1}{2-\beta}-\frac{3}{2}\right\rfloor-\frac{1}{2}\right)}$
Top-k closeness	BCM [7]	$n^{2-\frac{1}{\beta-1}}$ (*)	$n^{2-o(1)}$	$m^{1+o(1)}$
Distance oracle (query time) (space needed)	AIY [4]	$n^{1-o(1)}$ $n^{2-o(1)}$	$\le n^{f(\beta)}$ (*) (no closed form) $\le n^{1+f(\beta)}$ (*)	$\le n^{\frac{1}{2}+o(1)}$ $\le n^{\frac{3}{2}+o(1)}$

ance is unbounded, while if $\beta > 3$ also the variance is finite. Furthermore, all the results with $\beta > 3$ can be easily generalized to any degree distribution with finite variance, but the results become more cumbersome and dependent on specific characteristics of the distribution, such as the maximum degree of a vertex (for this reason, we focus on the power law case). Conversely, in the case $\beta < 3$, our results strongly depend on the degree distribution to be power law, because random graphs generated with different degree distributions can have very different behaviors. The only open cases are $\beta = 2$ and $\beta = 3$, which are left for future work (note that, if $\beta \le 1$, the degree distribution is not well defined).

Approximating the diameter. We confirm the empirical results in [31], proving that the 2-Sweep heuristic is significantly better than the basic sampling algorithm, which returns the maximum eccentricity of a random set of vertices. Furthermore, we show that the SumSH is even better than the 2-Sweep, confirming the experimental results in [13, 14]. Finally, we analyze the well-known RW algorithm, which provides a guaranteed $\frac{3}{2}$-approximation of the diameter in time $\Theta(m\sqrt{n})$. In our framework, it does not improve the 2-Sweep algorithm (which is much faster): this might theoretically explain why many graph libraries implement (variations of) the 2-Sweep heuristic, but not the RW algorithm (for instance, Sagemath [41], Webgraph [8], NetworKit [40]).

Computing the diameter. The aforementioned heuristics can be turned into exact algorithms, that always provide the correct result, but that can be inefficient in the worst case. We analyze two of these algorithms, proving that, for small values of β, both the iFub and the SumS algorithm are very efficient; for big values of β, the SumS algorithm is usually better, because it is always subquadratic. These results explain the surprisingly small running time on most graphs, and the reason why the SumS algorithm is usually faster on "hard" instances, as observed in [13, 14]. It is interesting to note that all the running times for $\beta > 3$ depend on the same constant C, which we prove to be close to $\frac{2\operatorname{dist}_{avg}(n)}{D-\operatorname{dist}_{avg}(n)}$, where D is the diameter and $\operatorname{dist}_{avg}(n)$ is the average distance of two nodes in the input graph. Intuitively, if this ratio is small, it means that there are "few far vertices", and the algorithms are quite efficient because they only need to analyze these vertices (the only exception is the sampling algorithm, which is not able to find these vertices, and hence achieves better performances when C is large). For $2 < \beta < 3$, a very similar argument applies, but in this case $C = 0$, because $D = \mathcal{O}(\log n)$ and $\operatorname{dist}_{avg}(n) = \mathcal{O}(\log\log n)$.

Other algorithms. Our framework lets us also analyze algorithms for computing other quantities. For example, the SumS algorithm is also able to compute
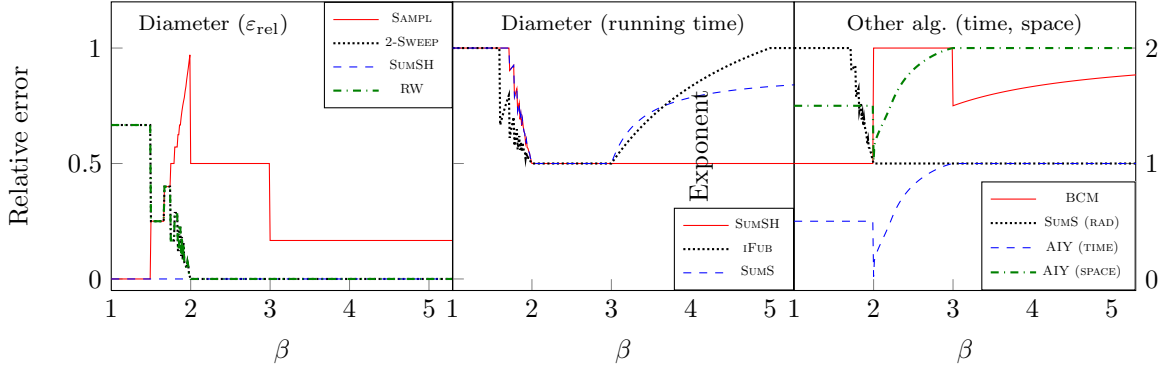
Figure 1: plot of the running time and relative errors of the heuristics and algorithms considered. The constant C was set to 3, and the $o(1)$ were ignored.

the radius: in this case, in all regimes, the running time is almost linear, confirming the results in [14], where it is shown that the algorithm needed at most 10 BFSes on all inputs but one, and in the last input it needed 18 BFSes. The other two algorithms analyzed are the BCM algorithm, to compute the k most central vertices according to closeness centrality [15], and the distance oracle AIY in [4]. In the first case, we show significant improvements with respect to the worst-case in the regime $1 < \beta < 2$ and $\beta > 3$, and we show that the algorithm is not efficient if $2 < \beta < 3$: this is the *only* result in this paper which is not reflected in practice. The problem is that our analysis relies on the fact that $\mathrm{dist}_{\mathrm{avg}}(n) = \Theta(\log \log n)$ tends to infinity, but the experiments were performed on graphs where $n < 10\,000\,000$, and consequently $\log \log(n) < 4$. The last probabilistic analysis confirms the efficiency of AIY: we show that, if $\beta < 3$, the expected time needed to compute the distance between two random nodes is sublinear, and the space occupied is subquadratic.

Finally, as a side result of our analysis, we compute for the first time the diameter of random graph in the regime $1 < \beta < 2$.

1.1 Related Work. This work combines results in several research fields: the analysis of random graphs, axiomatic approaches in the study of social networks, the design of heuristics and algorithms that are efficient on real-world graphs, the average and worst-case analysis of algorithms. Since it is impossible to provide a comprehensive account of the state-of-the-art in all these areas, here we just point the reader to the most recent and comprehensive surveys.

There are several works that study metric properties of random graphs: most of these results are summarized in [45, 44]. In the regime $\beta > 2$, we take inspiration from the proofs in [25] for the Configuration Model, and in [35, 10] for Inhomogeneous Random Graphs. In this

setting, we give a formal statements that links neighborhood sizes with branching processes (Theorem B.2): although it was used very often as a heuristic argument [45, 44], or used in specific settings [35], as far as we know, it was never formalized in this general setting. Furthermore, in the regime $1 < \beta < 2$, we use new techniques to prove the four properties, and as a result we obtain new asymptotics for the diameter. As far as we know, the only work that addresses the latter case is [43], which only computes the typical distance between two nodes.

Furthermore, our work relies on several works that outline the main properties of complex networks, and that develop models that satisfy such properties: for example, the choice of the power law degree distribution is validated by extensive empirical work (see [34] for a survey).

Despite this large amount of research on models of real-world graphs, few works have tried to address the problem of evaluating heuristics and algorithms on realistic models. For example, several works have addressed the efficiency of computing shortest paths in road networks [27, 38, 24, 23]. In [3], the authors provide an explanation of their efficiency, based on the concept of *highway dimension*. Another example is the algorithm in [8], which is used to compress a web graph: in [19], the authors prove that in most existing models no algorithm can achieve good compression ratio, and they provide a new model where the algorithm in [8] works well. Also in [29], the authors develop an axiomatic framework, but they study triangle density, and not distances (that is, they assume that the input graph contains many triangles, a characteristic that is shared by most real-world graphs). That paper sets forth the research agenda of defining worst-case conditions on graphs generalizing all the popular generative models: it discusses the main advantages and disadvantages of the approach, and it leaves as an open problem to find algorithms that are

more efficient on the class of triangle-dense graphs. Another related work is [17], where the authors develop an axiomatic approach that is similar to ours: assuming only that the degree distribution is power law, they manage to analyze some algorithms, and to prove that their analysis improves the worst-case analysis. Our work is orthogonal to their work: indeed, they only assume the degree distribution to be power law, using a variation of our Property 2.4. Their properties are weaker than ours, since they only assume a variation of our Property 2.4: for this reason, they manage to analyze algorithms that compute local properties, such as patterns in subgraphs, but not the global metric properties considered in this paper (indeed, graphs with the same degree distribution can have very different metric properties). An approach that is more similar to ours is provided in [16]: among other results, it is proved that it is possible to compute a shortest path between two nodes s, t in sublinear time, in the models considered in this paper. The running-time is $\mathcal{O}(n^{\frac{1}{2}+\varepsilon})$ if $\beta > 3$, and $\mathcal{O}(n^{\frac{4-\beta}{2}})$ if $2 < \beta < 3$ (while, in the worst-case, this task can be performed in $\mathcal{O}(n)$).

Finally, some works have tried to explain and motivate the efficiency of some heuristics and algorithms for diameter and radius computation. A first attempt uses the *Gromov hyperbolicity* of the input graph [28]: for example, the 2-Sweep heuristic provides good approximations of the diameter of hyperbolic graphs [18]. However, this approach cannot be applied to some algorithms, like the iFub, and when it can be applied, the theoretical guarantees are still far from the empirical results, because real-world graphs are usually not hyperbolic according to Gromov's definition [11].

1.2 Structure of the Paper. In Section 2, we state the four properties considered; in Section 3, we define the models considered and we sketch the proof that they satisfy the four properties. In Section 4 we experimentally show that they are satisfied by real-world graphs. Then, in Section 5, we prove some consequences of the axioms, that are extensively used throughout the paper, such as asymptotics for the diameter, the average distance, etc. In Sections 6 to 11, we perform the probabilistic analysis for diameter and radius algorithms (for the other two algorithms, BCM and AIY, we refer to Appendices C and D). Section 12 concludes the paper.

2 The Four Properties.

In this section, we define the four properties used in our framework. Let us start with some definitions.

DEFINITION 2.1. *Given a graph $G = (V, E)$, if $s \in V$,*

let $\mathbf{\Gamma}^\ell(s)$ be the set of vertices at distance exactly ℓ from s, let $\boldsymbol{\gamma}^\ell(s) = |\mathbf{\Gamma}^\ell(s)|$, let $\mathbf{N}^\ell(s)$ be the set of vertices at distance at most ℓ from s, and let $\boldsymbol{n}^\ell(s) = |\mathbf{N}^\ell(s)|$. We define $\boldsymbol{\tau}_s(k) = \min\{\ell \in \mathbb{N} : \boldsymbol{\gamma}^\ell(s) > k\}$, and $T(d \to k)$ as the average number of steps for a node of degree d to obtain a neighborhood of k nodes. More formally, $T(d \to k)$ is the average $\boldsymbol{\tau}_s(k)$ over all vertices s of degree d (note that, since the diameter is $\mathcal{O}(\log n)$, $T(d \to n^x)$ is defined for each $x < 1$).

Our properties depend on a parameter ε: for instance, the first property bounds the number of vertices such that $\boldsymbol{\tau}_s(n^x) \geq (1 + \varepsilon)T(d \to n^x)$. Intuitively, one can think of ε as a constant which is smaller than any other constant appearing in the proofs, but bigger than $\frac{1}{n}$, or any other infinitesimal function of n. Indeed, in random graphs, we prove that if we fix $\varepsilon, \delta > 0$, we can find $n_{\varepsilon,\delta}$ such that the properties hold for each $n > n_{\varepsilon,\delta}$, with probability at least $1 - \delta$. In real-world graphs, we experimentally show that the four properties are verified with good approximation for $\varepsilon = 0.2$. In our analyses, the time bounds are of the form $n^{c+\mathcal{O}(\varepsilon)}$, and the constants in the $\mathcal{O}$ are quite small. Since, in our dataset, $n^{0.2}$ is between 6 and 19, we can safely consider $n^{c+\mathcal{O}(\varepsilon)}$ close to n^c.

The first property analyzes the typical and extremal values of $\boldsymbol{\tau}_s(n^x)$, where s is any vertex.

PROPERTY 2.1. *There exists a constant c such that:*

- *for each vertex s with degree $d > n^\varepsilon$, $\boldsymbol{\tau}_s(n^x) \leq (1 + \varepsilon)(T(d \to n^x) + 1)$;*

- *the number of vertices verifying $\boldsymbol{\tau}_s(n^x) \geq (1 + \varepsilon)(T(d \to n^x) + \alpha)$ is $\mathcal{O}(nc^{\alpha-x})$;*

- *the number of vertices verifying $\boldsymbol{\tau}_s(n^x) \geq (1 - \varepsilon)(T(1 \to n^x) + \alpha)$ is $\Omega(nc^{\alpha-x})$.*

In random graphs, the values of $T(d \to n^x)$ depend on the exponent β (see Table 2). In many of our analyses, we do not use the actual values of $T(d \to n^x)$, but we use the following properties:

- $T(d \to n^{x+\varepsilon}) \leq T(d \to n^x)(1 + \mathcal{O}(\varepsilon))$;

- $\sum_{d=1}^\infty |\{v \in V : \deg(v) = d\}|T(d \to n^x) = (1 + o(1))nT(1 \to n^x)$;

- $T(1 \to n^x) + T(1 \to n^{1-x}) - 1 = (1 + o(1)) \text{dist}_{\text{avg}}(n)$, where $\text{dist}_{\text{avg}}(n)$ is a function not depending on x (this function is very close to the average distance, as we prove in Section 5).

The next two properties relate the distance between two vertices s, t with the values of $\boldsymbol{\tau}_s(n^x)$, $\boldsymbol{\tau}_t(n^y)$,

Table 2: the values of $T(d \to n^x)$, $\text{dist}_{\text{avg}}(n)$ and c, depending on the value of β.

Regime	$T(d \to n^x)$	$\text{dist}_{\text{avg}}(n)$	c
$1 < \beta < 2$	1 if $d \geq n^x$, 2 otherwise	3	$n^{-\frac{2-\beta}{\beta-1}(1+o(1))}$
$2 < \beta < 3$	$(1+o(1))\log_{\frac{1}{\beta-2}} \frac{\log n^x}{\log d}$ if $n^x < n^{\frac{1}{\beta-1}}$ $(1+o(1))\log_{\frac{1}{\beta-2}} \frac{\log n^x}{\log d} + \mathcal{O}(1)$ if $n^x > n^{\frac{1}{\beta-1}}$	$(2+o(1))\log_{\frac{1}{\beta-2}} \log n$	$\eta(1)+o(1)$
$\beta > 3$	$(1+o(1))\log_{M_1(\mu)} \frac{n^x}{d}$	$(1+o(1))\log_{M_1(\mu)} n$	$\eta(1)+o(1)$

where x, y are two reals between 0 and 1. The idea behind these two properties is to apply the "birthday paradox", assuming that $\mathbf{\Gamma}^{\boldsymbol{\tau}_s(n^x)}(s)$ and $\mathbf{\Gamma}^{\boldsymbol{\tau}_t(n^y)}(t)$ are random sets of n^x and n^y vertices. In this idealized setting, if $x + y > 1$, there is a vertex that is common to both, and $\text{dist}(s,t) \leq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y)$; conversely, if $x + y < 1$, $\text{dist}(s,t)$ is likely to be bigger than $\boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y)$. Let us start with the simplest property, which deals with the case $x + y > 1$.

PROPERTY 2.2. *Let us fix two real numbers $0 < x, y < 1$ such that $x + y > 1 + \varepsilon$. For each pair of vertices s, t, $\text{dist}(s,t) < \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y)$.*

The next property is a sort of converse: the main idea is that, if the product of the size of two neighborhoods is smaller than n, then the two neighborhoods are usually not connected. The simplest way to formalize this is to state that, for each pair of vertices s, t, $\text{dist}(s,t) \geq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y)$. However, there are two problems with this statement: first, in random graphs, if we fix s and t, $\text{dist}(s,t) \geq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y)$ a.a.s., not w.h.p., and hence there might be $o(n)$ vertices t such that $\text{dist}(s,t) < \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y)$ (for example, if s and t are neighbors, they do not verify $\text{dist}(s,t) \geq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y)$). To solve this, our theorem bounds the number of vertices t verifying $\text{dist}(s,t) \geq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y)$. The second problem is more subtle: for example, if s has degree 1, and its only neighbor has degree $n^{\frac{1}{2}}$, $\boldsymbol{\tau}_s\left(n^{\frac{1}{4}}\right) = \boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) = 2$, and the previous statement cannot hold for $x = \frac{1}{4}$. However, this problem does not occur if $x \geq y$: the intuitive idea is that we can "ignore" vertices with degree bigger than n^x. Indeed, if a shortest path from s to t passes through a vertex v with degree bigger than n^x, then $\boldsymbol{\tau}_s(n^x) \leq \text{dist}(s,v)+1$, $\boldsymbol{\tau}_t(n^y) \leq \text{dist}(t,v)+1$, and hence $\text{dist}(s,t) = \text{dist}(s,v) + \text{dist}(v,t) \geq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y) - 2$.

PROPERTY 2.3. *Let s be any vertex, let $0 < z \leq y < x < 1$, let $x + y \geq 1 + \varepsilon$, and let α, ω be integers. If $T_{\alpha,\omega,z}$ is the set of vertices t such that $\boldsymbol{\tau}_t(n^z)$ is between α and ω, there are at most $|T_{\alpha,\omega,z}| \frac{n^{x+y+\varepsilon}}{n}$ vertices $t \in T$ such that $\text{dist}(s,t) < \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y) - 2$.*

Finally, in some analyses, we also need to use the fact that the degree distribution is power law. To this purpose, we add a further property (in random graphs, this result is well-known [45, 44]).

PROPERTY 2.4. *The number of vertices with degree bigger than d is $\Theta\left(\frac{n}{d^{\max(1,\beta-1)}}\right)$.*

Although the definition of the four properties is quite complicated, the intuition is natural. Indeed, Properties 2.2 and 2.3 simply say in a formal way that $\text{dist}(s,t) \approx \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^{1-x})$, and this is the property which is used in all the probabilistic analysis. As far as we know, in the context of the analysis of real-world graphs, this property was never stated or formalized: we believe that it can give further insight in the field of the analysis of real-world graphs. A further confirmation of the importance of this property is that the algorithms considered are not very efficient on graphs where this property is not satisfied, such as road networks [13, 14, 7].

Conversely, Properties 2.1 and 2.4 are more specific, and they are specifically suited to the analysis of the real-world networks and the random graphs under consideration. They were chosen because they are satisfied by the graphs under consideration, but one might be interested in using variations of these properties on different kinds of networks, since the proofs usually do not depend on the specific values of the parameters considered.

3 Validity of the Properties in Random Graphs: Overview.

In order to transform the axiomatic worst-case analyses into average-case analyses on random graphs, we use the following theorem.

THEOREM 3.1. *For each fixed $\varepsilon > 0$, Properties 2.1 to 2.4 are verified in the random graphs defined in all the models considered, a.a.s..*

In other words, for each $\varepsilon, \delta > 0$, there exists $n_{\varepsilon,\delta}$ such that the probability that a random graph with $n > n_{\varepsilon,\delta}$ nodes does not verify the four properties is at most $1 - \delta$.

In this section, we sketch the proof of this theorem, while we provide the complete proof in Appendix B.

3.1   The Models The models considered are the Configuration Model (CM) and Rank-1 Inhomogeneous Random Graphs (IRG), such as the Norros-Reittu model and the Chung-Lu model. All these models fix a set V of n vertices, and they assign a weight ρ_v to each vertex $v \in V$ (we choose the weights ρ_v according to a power law distribution with exponent β). Then, we create edges in a way that the degree of v is close to ρ_v: in the CM, this is done by associating to v ρ_v half-edges, and pairing these half-edges at random, while in IRG, an edge between vertices v and w exists with probability close to $\frac{\rho_v \rho_w}{M}$, where $M = \sum_{v \in V} \rho_v$.

Furthermore, we need to consider only the *giant component* of the graph considered, and, differently from other works, we do not assume the graph generated through the CM to be simple (anyway, multiple edges and self-loops have no effect on distances). For more details of the models considered, and for some additional technical assumptions used to avoid pathological cases, we refer to Appendix A.

3.2   Properties 2.2 to 2.4. It is quite easy to prove that Property 2.4 holds: indeed, it is enough to show that the degree of a vertex v is close to its weight ρ_v, and this can be done through a Chernoff-type probability bound.

Then, we need to prove that Properties 2.2 and 2.3 hold: these two properties bound $\text{dist}(s,t)$ with $\boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y)$. Let us assume that $\boldsymbol{\gamma}^\ell(s) = n^x$, and $\boldsymbol{\gamma}^{\ell'}(t) = n^y$: if all vertices are in $\boldsymbol{\Gamma}^\ell(s)$ with the same probability, $\boldsymbol{\Gamma}^\ell(s)$ will be a random subset of the set of vertices, and the probability that a vertex in $\boldsymbol{\Gamma}^\ell(s)$ is also in $\boldsymbol{\Gamma}^{\ell'}(t)$ is close to $\frac{\boldsymbol{\gamma}^{\ell'}(t)}{n} = \frac{1}{n^{1-y}}$. Hence, the probability that $\text{dist}(s,t) \geq \ell + \ell'$ is related to the probability that $\boldsymbol{\Gamma}^\ell(s)$ does not intersect $\boldsymbol{\Gamma}^{\ell'}(t)$, which is close to $\left(1 - \frac{1}{n^{1-y}}\right)^{n^x} \approx e^{-n^{x+y-1}}$. For $x + y > 1$, this means that $\text{dist}(s,t) \leq \ell + \ell'$ w.h.p., and this is very close to the statement of Property 2.2. For $x+y < 1$, $\boldsymbol{\Gamma}^\ell(s)$ does not intersect $\boldsymbol{\Gamma}^{\ell'}(t)$ with probability $e^{-n^{x+y-1}} \approx 1 - n^{x+y-1}$, and hence $\text{dist}(s,t) \leq \ell + \ell'$ with probability close to n^{x+y-1}. The proof that Property 2.3 holds is then concluded by applying concentration inequalities, exploiting the fact that T is "enough random".

3.3   Property 2.1, $\beta > 2$. The proof that Property 2.1 holds is much more complicated: in the proof, we have to distinguish between the case $\beta < 2$ and $\beta > 2$. In the case $\beta > 2$, we use two different techniques.

1. When $\boldsymbol{\gamma}^\ell(s) = |\boldsymbol{\Gamma}^\ell(s)|$ is small (say, smaller than n^ε), we show that the behavior of $\boldsymbol{\gamma}^\ell(s)$ is well approximated by a μ-distributed branching process, where μ is the residual distribution of λ (the definition of residual distribution depends on the model, and it is provided in Definition A.1). Furthermore, if s and t are two different vertices, and if $\boldsymbol{\gamma}^\ell(s)$ and $\boldsymbol{\gamma}^{\ell'}(t)$ are small, the behavior of $\boldsymbol{\Gamma}^\ell(s)$ and the behavior of $\boldsymbol{\Gamma}^{\ell'}(t)$ are "almost" independent.

2. When $\boldsymbol{\gamma}^\ell(s)$ is large, the branching process approximation and the independence do not hold anymore. We need a different technique: since $\boldsymbol{\gamma}^\ell(s) > n^\varepsilon$, a Chernoff-type probability bound gives guarantees of the form e^{-n^ε}, which is bigger than any polynomial in n. This way, we can prove very precise bounds on the size of $\boldsymbol{\gamma}^{\ell+1}(s)$ given the size of $\boldsymbol{\gamma}^\ell(s)$, and through a union bound we can show that these bounds hold for any vertex s.

The second technique was already used in some works [20, 35, 25]; however, the formalization of the connection between neighborhood expansion and branching processes is original (Theorem B.2), it formalizes existing intuitive explanations [45, 44], and it generalizes proofs that were performed in restricted classes of models [35, 10]. Let us provide some more details: we define a branching process $\boldsymbol{\delta}^\ell(s)$ coupled with $\boldsymbol{\gamma}^\ell(s)$ (that is, $\boldsymbol{\gamma}^\ell(s)$ and $\boldsymbol{\delta}^\ell(s)$ are defined on the same probability space, and the probability that they are equal is high). Then, we analyze the size of $\boldsymbol{\delta}^\ell(s)$: if the first moment $M_1(\mu)$ of the distribution μ of the branching process is finite (or, equivalently, if $M_2(\lambda)$ is finite), it is well known [5] that the expected size of $\boldsymbol{\delta}^\ell(s)$ is $\boldsymbol{\delta}^1(s) M_1(\mu)^{\ell-1} = \deg(s) M_1(\mu)^{\ell-1}$; if λ is a power law distribution with $2 < \beta < 3$, the typical size of $\boldsymbol{\delta}^\ell(s)$ is close to $\boldsymbol{\delta}^1(s)^{\left(\frac{1}{\beta-2}\right)^{\ell-1}} = \deg(s)^{\left(\frac{1}{\beta-2}\right)^{\ell-1}}$. Hence, heuristically, we can estimate $\boldsymbol{\tau}_s(n^x)$, by setting $\deg(s) M_1(\mu)^{\ell-1} = n^x$ if $M_1(\mu)$ is finite and strictly bigger than 1, and $\deg(s)^{\left(\frac{1}{\beta-2}\right)^{\ell-1}} = n^x$ if μ is power law with exponent $1 < \beta < 2$. Solving with respect to ℓ, we obtain the values in Table 2.

Through a more refined analysis, we can use the branching process approximation to estimate the deviations from these value: first, we remove from the branching process all branches that have a finite number of descendants, since they have little impact on the total size of the branching process (if the whole branching process is finite, it means that the starting vertex is not in the *giant component*, and we can ignore it). It is proved in [5, 1.D.12] that we obtain another branching process, with distribution η that depends only on μ, and such that $\eta(0) = 0$, so that all branches are infinite. Then,

we prove that the "worst" that can happen is that $\boldsymbol{\delta}^\ell(s)$ is 1 for a long time, and then it grows normally: this means that $\mathbb{P}\left(\boldsymbol{\tau}_s(n^x) > T\left(\deg(s) \to n^x\right) + k\right) \approx \eta(1)^k$, and, since the growths of different vertices are almost independent, we obtain that the number of vertices verifying $\boldsymbol{\tau}_s(n^x) > T\left(\deg(s) \to n^x\right) + k$ is approximately $n\eta(1)^k$.

Summarizing, we sketched the proof that the values appearing in Table 2 are correct, and that Property 2.1 holds, at least when x is small. For big values of x, the branching process approximation does not hold anymore: however, as soon as $\boldsymbol{\gamma}^\ell(s)$ is large enough, we can prove directly that $\boldsymbol{\gamma}^{\ell+1}(s) \approx \boldsymbol{\gamma}^\ell(s) M_1(\mu)$ if $M_1(\mu)$ is finite, and $\boldsymbol{\gamma}^{\ell+1}(s) \approx \boldsymbol{\gamma}^\ell(s)^{\frac{1}{\beta-2}}$ if λ is power law with exponent $2 < \beta < 3$, w.h.p.. This way, we can prove results on $\boldsymbol{\tau}_s(n^x)$ by proving the same results for $\boldsymbol{\tau}_s(n^y)$ for some small y, and extending the result to $\boldsymbol{\tau}_s(n^x)$ using this argument. This concludes the proof that the values appearing in Table 2 are correct, and that Property 2.1 holds.

3.4 Property 2.1, $\beta < 2$. In this case, the branching process approximation does not hold: indeed, the residual distribution μ cannot be even defined! We use a completely different technique. First, we consider the N vertices with highest weight, where N is a big constant: using order statistics, we can prove that each of these vertices has weight $\Theta(M)$, where $M = \sum_{v \in V} \rho_v$. From this, we can prove that each vertex with degree at least n^ε is connected to each of these N vertices, and these N vertices have degree $\Theta(n)$. This is enough to characterize the size of neighbors of any vertex v with degree bigger than n^ε: there are $\deg(v)$ vertices at distance 1 and $\Theta(n)$ vertices at distance 2.

Let us now consider the neighborhood growth of other vertices: given a vertex v, the probability that it is not connected to any vertex w with weight smaller than n^ε is approximately $\prod_{\rho_w < n^\varepsilon}\left(1 - \frac{\rho_v \rho_w}{M}\right) \approx 1 - \frac{\rho_v}{M}\sum_{\rho_w < n^\varepsilon}\rho_w \approx 1 - \frac{n}{n^{\frac{1}{\beta-1}}}$ (it is possible to prove that $M \approx n^{\frac{1}{\beta-1}}$). As a consequence, the probability that a vertex v is connected to another vertex with weight $w < n^\varepsilon$ is quite small, being approximately $n^{-\frac{2-\beta}{\beta-1}} = c$. Let us consider three cases separately.

1. If v is connected to a vertex w with degree at least n^ε, we deduce results on neighbors of v from results on neighbors of w.

2. If v is not connected to a vertex w with degree at least n^ε, the following cases might occur:

   (a) if v is not connected to a vertex with weight smaller than n^ε, we can ignore it, because it is not in the giant component;

   (b) the last case is that v is connected to another vertex w with weight smaller than n^ε, which occurs with probability c; in this case, we iterate our argument with w, until we hit a vertex with degree at least n^ε.

In particular, the probability that Item 2b occurs ℓ times before hitting a vertex with degree at least n^ε is approximately c^ℓ: this means that the number of vertices whose neighbors reach size n^x after ℓ steps is at most $nc^{\ell + \mathcal{O}(1)}$. Through a more thorough analysis of the constant $\mathcal{O}(1)$, we obtain the results in Table 2, proving upper bounds for Property 2.1. For lower bounds, surprisingly, we only have to consider vertices with degree 1 and 2: in particular, the probability that a vertex with degree 1 is linked to another vertex of degree 2 turns out to be approximately c. For this reason, there are at least nc^ℓ vertices of degree 1 that are starting points of a path of length ℓ, which terminates in a vertex with larger degree. This concludes the proof that Property 2.1 holds.

4 Validity of the Properties in Real-World Graphs.

In this section, we experimentally show that the first three properties hold in real-world graphs, with good approximation (we do not perform experiments on the fourth property, because it is well known that the degree distribution of many real-world graphs is power law [6, 33]). To this purpose, we consider a dataset made by 18 real-world networks of different kinds (social networks, citation networks, technological networks, and so on), taken from the well-known datasets SNAP (`snap.stanford.edu/`) and KONECT (`http://konect.uni-koblenz.de/networks/`). Then, for each of the properties, we compute the quantities considered, on all graphs in the dataset, and we show that the actual behavior reflects the predictions.

We start with Property 2.1: to verify the first claim, we consider all vertices with degree at least $n^{0.2}$, which is between 6 and 19 in our inputs. For each of these vertices, we compute $\boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) - T\left(\deg(s) \to n^{\frac{1}{2}}\right)$ (in this paper, we show the results for $x = \frac{1}{2}$, but very similar results hold for all values of x). The results obtained are represented in Table 3.

The table shows that in all the graphs considered, the first statement of Property 2.1 is verified with good approximation: almost all vertices with degree at least $n^{0.2}$ verify $\boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) - \left\lceil T\left(\deg(s) \to n^{\frac{1}{2}}\right)\right\rceil \leq 2$; the percentage of vertices verifying $\boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) - \left\lceil T\left(\deg(s) \to n^{\frac{1}{2}}\right)\right\rceil = 2$ is always below 0.5%,

Network	$n^{0.2}$	Vert.	$k=-2$	$k=-1$	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
p2p-Gnutella09	6.1	2811	0.00%	61.37%	38.63%	0.00%	0.00%	0.00%	0.00%
oregon1-010526	6.5	640	0.00%	58.75%	41.25%	0.00%	0.00%	0.00%	0.00%
ego-gplus	7.5	348	0.00%	2.87%	97.13%	0.00%	0.00%	0.00%	0.00%
oregon2-010526	6.5	1113	0.00%	55.17%	44.83%	0.00%	0.00%	0.00%	0.00%
ca-HepTh	6.1	1987	2.21%	48.97%	43.48%	4.98%	0.25%	0.00%	0.10%
ca-CondMat	7.3	6519	0.00%	45.25%	51.20%	3.27%	0.23%	0.05%	0.00%
ca-HepPh	6.5	4644	0.00%	46.32%	50.39%	2.84%	0.45%	0.00%	0.00%
email-Enron	8.0	6354	0.00%	69.00%	30.33%	0.66%	0.02%	0.00%	0.00%
loc-brightkite	8.9	9929	0.00%	69.45%	29.94%	0.42%	0.18%	0.00%	0.00%
email-EuAll	11.8	2654	0.00%	59.08%	40.66%	0.23%	0.00%	0.00%	0.04%
ca-AstroPh	7.1	9812	0.00%	58.55%	41.10%	0.18%	0.16%	0.00%	0.00%
gowalla-edges	11.5	33263	0.00%	65.69%	34.07%	0.23%	0.01%	0.00%	0.00%
munmun-twitter	13.6	6670	0.00%	70.57%	29.43%	0.00%	0.00%	0.00%	0.00%
com-dblp	12.6	33363	1.65%	63.03%	32.41%	2.57%	0.32%	0.01%	0.00%
com-lj.all.cmty	12.5	5258	0.51%	65.96%	32.98%	0.53%	0.02%	0.00%	0.00%
enron	9.7	7792	0.00%	77.71%	21.79%	0.37%	0.13%	0.00%	0.00%
com-youtube	16.3	46471	0.00%	79.01%	20.32%	0.45%	0.15%	0.04%	0.02%
wiki-Talk	18.9	27536	0.00%	62.63%	37.37%	0.00%	0.00%	0.00%	0.00%

Table 3: the percentage of vertices with degree at least $n^{0.2}$ that verify $\boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) - \left\lceil T\left(d \to n^{\frac{1}{2}}\right)\right\rceil = k$ (the other values of k are 0, for each graph in the dataset).
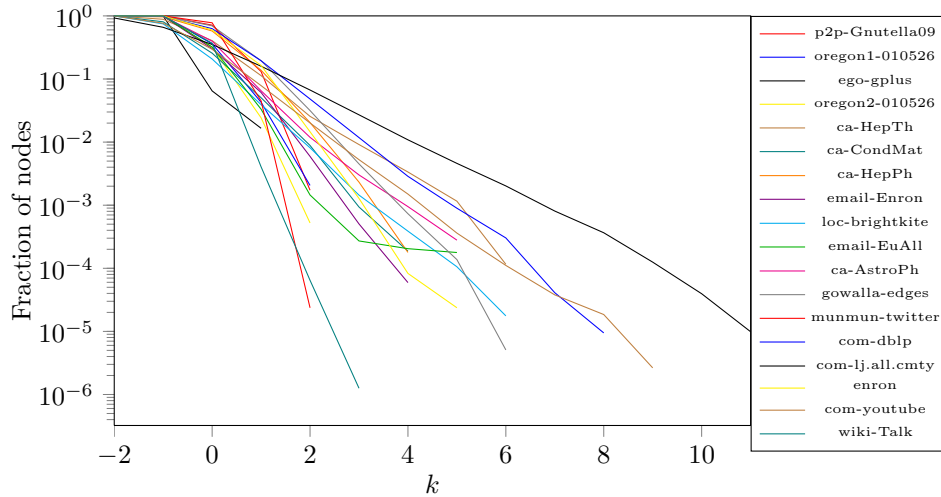

Figure 2: the percentage of vertices verifying $\boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) - T\left(\deg(s) \to n^{\frac{1}{2}}\right) \geq k$, in all the graphs in our dataset.

and the percentage of vertices verifying $\boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) - \left\lceil T\left(\deg(s) \to n^{\frac{1}{2}}\right)\right\rceil = 1$ is always below 5%.

For the other two points of Property 2.1, for each vertex s, we have computed $\boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) - T\left(\deg(s) \to n^{\frac{1}{2}}\right)$. We want to prove that the number of vertices that verify $\boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) - T\left(\deg(s) \to n^{\frac{1}{2}}\right) \geq k$ is close to nc^k, for some constant c smaller than 1. For

this reason, we have plotted the fraction of vertices verifying this inequality in logarithmic scale, in Figure 2.

This plot confirms the last two points of Property 2.1: indeed, in logarithmic scale, the number of vertices satisfying $\boldsymbol{\tau}_s\left(n^{\frac{1}{2}}\right) - T\left(\deg(s) \to n^{\frac{1}{2}}\right) \geq k$ decreases almost linearly with k, when $k > 0$.

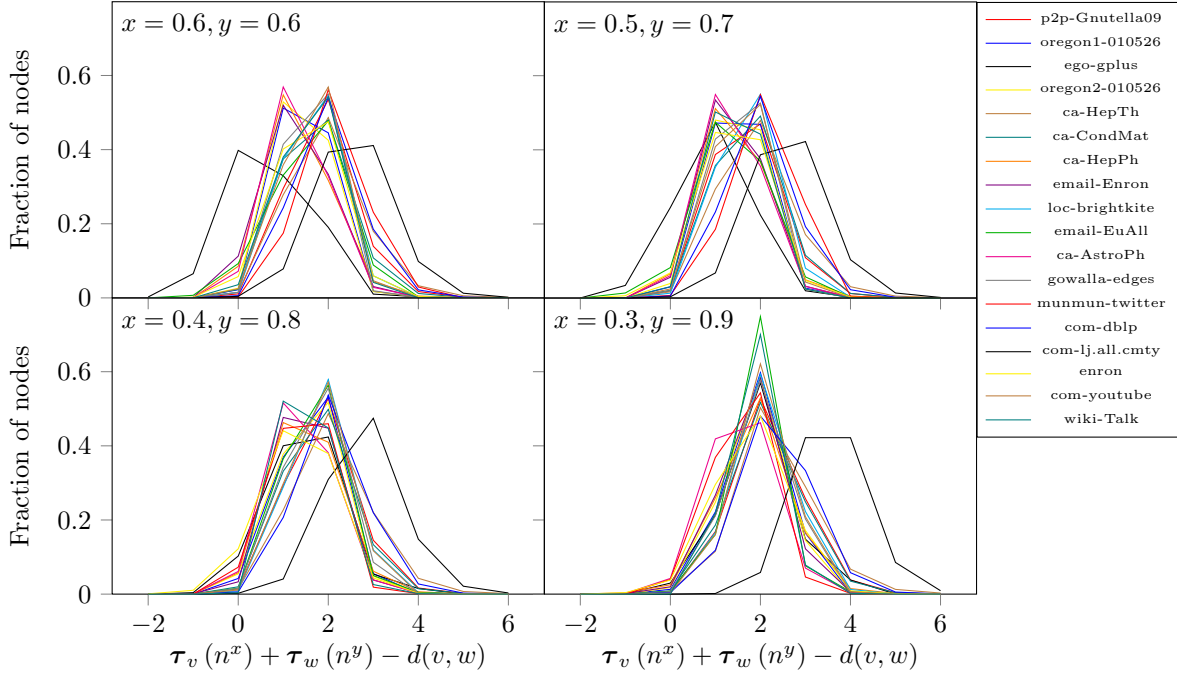Then, let us validate Property 2.2, which says that, whenever $x + y > 1 + \varepsilon$, for each pair of vertices s, t, $\text{dist}(s, t) < \boldsymbol{\tau}_s\left(n^x\right) + \boldsymbol{\tau}_t\left(n^y\right)$:

Figure 3: the values of $\boldsymbol{\tau}_s\left(n^x\right) + \boldsymbol{\tau}_t\left(n^y\right) - \text{dist}(s,t)$ for $10\,000$ pairs of vertices in each graph.

we have tested this condition with $(x,y) = (0.3,0.9),(0.4,0.8),(0.5,0.7),(0.6,0.6)$. For each graph $G=(V,E)$ in the dataset, and for each of the aforementioned pairs (x,y), we have chosen a set $T \subseteq V$ made by $10\,000$ random vertices (or the whole V if $|V| < 10\,000$), and for each i we have plotted the percentages of pairs $(s,t) \in T^2$ such that $\boldsymbol{\tau}_s\left(n^x\right) + \boldsymbol{\tau}_t\left(n^y\right) - \text{dist}(s,t) = i$. The plots are shown in Figure 3.

From the figure, it is clear that $\boldsymbol{\tau}_s\left(n^x\right) + \boldsymbol{\tau}_t\left(n^y\right)$ is almost always at least $\text{dist}(s,t)$, as predicted by Property 2.2. However, in some cases, $\text{dist}(s,t) = \boldsymbol{\tau}_s\left(n^x\right) + \boldsymbol{\tau}_t\left(n^y\right)$: we think that this is due to the fact that, in our random graph models, the guarantee is $\mathcal{O}\left(e^{-n^{\varepsilon}}\right)$, and for $\varepsilon = 0.2$, this value is not very small (for instance, if $n = 10\,000$, $e^{-n^{\varepsilon}} = 0.012$). However, this value tends to 0 when n tends to infinity, and this is reflected in practice: indeed, the fit is better when the number of nodes is larger. Overall, we conclude that Property 2.2 is valid with good approximation on the networks in the dataset, and we conjecture that the correspondance is even stronger for bigger values of n.

Finally, we need to validate Property 2.3, which says that, given a vertex s, for "many" sets of vertices T, $|\{t \in T : \boldsymbol{\tau}_s\left(n^x\right) + \boldsymbol{\tau}_t\left(n^y\right) < \text{dist}(s,t) + 2\}| \leq |T|n^{1-x-y+\varepsilon}$. Hence, we have chosen a random vertex s and a random set T made by $10\,000$ vertices, and for each $t \in T$, we have computed $z_t = \min\{x + y : x > y, \boldsymbol{\tau}_s\left(n^x\right) + \boldsymbol{\tau}_t\left(n^y\right) < \text{dist}(s,t) + 2\}$. If the number N_z

of vertices t such that $z_t < z$ is at most $|W|n^{-1+z+\varepsilon}$, then we can guarantee that the theorem holds for each x and y. Solving with respect to z, we want that $N_z \leq |T|n^{-1+z+\varepsilon}$, that is, $\log \frac{N_z}{|T|} \leq (-1 + z + \varepsilon)\log n$, that is, $z \geq 1 - \varepsilon + \frac{\log \frac{N_z}{|T|}}{\log n}$. Hence, Figure 4 shows the values of the function $1 + \frac{\log \frac{N_z}{|T|}}{\log n}$, for each graph in our dataset. Furthermore, since Property 2.3 also deals with sets T defined depending on $\boldsymbol{\tau}_t\left(n^x\right)$, we have also repeated the experiment on sets T containing only vertices t verifying $0 \leq \boldsymbol{\tau}_t\left(n^{\frac{1}{2}}\right) < \frac{D}{6}$, $\frac{D}{6} \leq \boldsymbol{\tau}_t\left(n^{\frac{1}{2}}\right) < \frac{D}{3}$, $\boldsymbol{\tau}_t\left(n^{\frac{1}{2}}\right) > \frac{D}{3}$, where D is the diameter of the graph.

From the plot, it is clear the claim is verified even with $\varepsilon = 0$, by all but one case. Also the latter case is verified with a very small value of ε.

For the validation of Property 2.4, we rely on extensive studies that show that the degree distribution of many real-world graphs is power law (for more information, we refer to [33] and the references therein).

5 Technical Preliminaries on Diameter, Eccentricity, Closeness Centrality, and Average Distance.

In this section, we prove some basic metric properties in the graphs satisfying our four properties. By specializing these results to random graphs, we obtain a new proof of known asymptotics, and we prove new asymp-
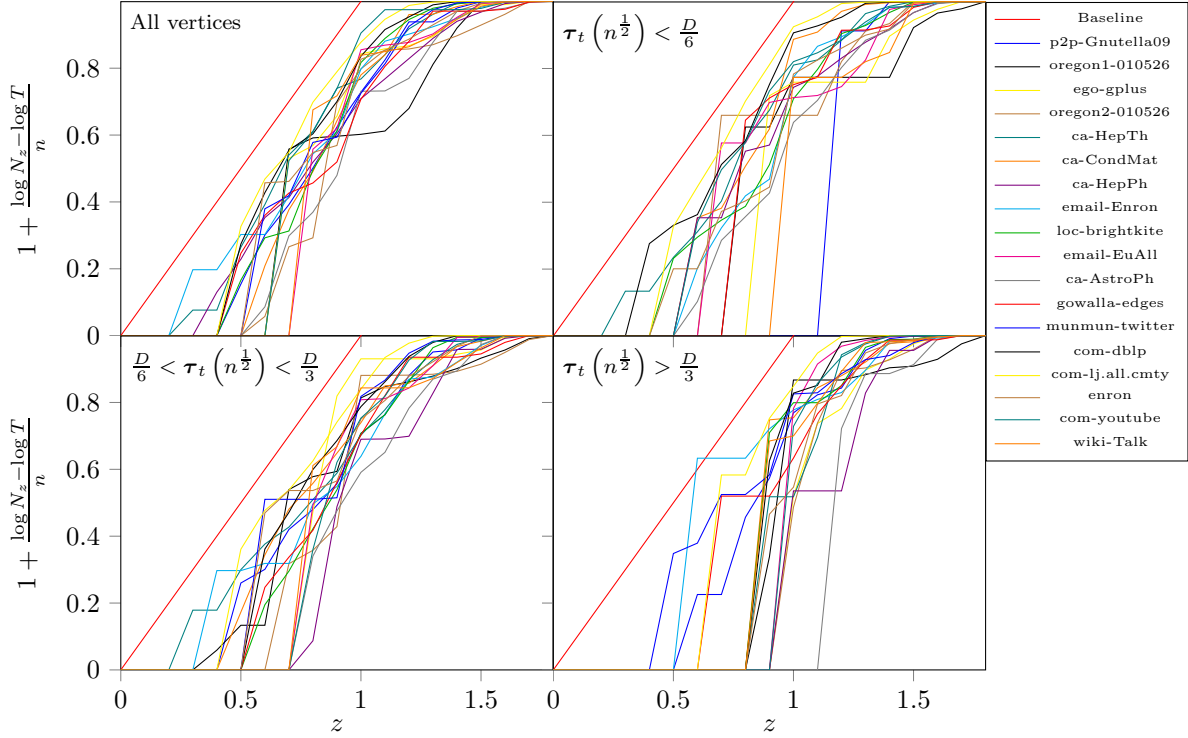
Figure 4: the values of $1 + \frac{\log \frac{N_z}{|T|}}{\log n}$, as a function of z.

totics in the case $1 < \beta < 2$. In all the following lemmas, with abuse of notation, we write $\mathcal{O}(\varepsilon)$ even if ε is a constant, in order to indicate a function bounded by $c\varepsilon$ for some constant c.

LEMMA 5.1. *All vertices s with degree d verify* $\boldsymbol{\tau}_s(n^x) \leq \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(T(d \to n^x) + \frac{\log n}{-\log c} + x \right) \right\rfloor$. *Moreover, for each $\delta > 0$, there are $\Omega(n^\delta)$ vertices s with degree 1 verifying $\boldsymbol{\tau}_s(n^x) \geq \left\lceil (1 - \varepsilon - \delta) \left(T(1 \to n^x) + \frac{\log n}{-\log c} - 1 + x \right) \right\rceil$.*

Proof. By Property 2.1 applied with $\alpha = (1 + \varepsilon)\frac{\log n}{-\log c} + x$, there are $\mathcal{O}(nc^{\alpha - x}) = \mathcal{O}\left(nc^{(1+\varepsilon)\frac{\log n}{-\log c}} \right) \leq \mathcal{O}(n^{-\varepsilon}) < 1$ vertices s such that $\boldsymbol{\tau}_s(n^x) \geq (1 + \varepsilon)(T(d \to n^x) + \alpha) = (1 + \varepsilon)\left(T(d \to n^x) + (1 + \varepsilon)\frac{\log n}{-\log c} + x \right)$. By observing that $\boldsymbol{\tau}_s(n^x)$ is an integer, we obtain the first claim.

For the other inequality, let us apply Property 2.1 with $\alpha = (1 - \delta)\frac{\log n}{-\log c} - 1 + x$: there are $\Omega(nc^{\alpha + 1 - x}) = \Omega\left(nc^{(1-\delta)\frac{\log n}{-\log c}} \right) = \Omega(n^\delta)$ vertices s such that $\boldsymbol{\tau}_s(n^x) \geq (1 - \varepsilon)(T(1 \to n^x) + \alpha) = (1 - \varepsilon)\left(T(1 \to n^x) + (1 - \delta)\frac{\log n}{-\log c} - 1 + x \right) \geq (1 - \varepsilon - \delta)\left(T(1 \to n^x) + \frac{\log n}{-\log c} - 1 + x \right)$. By observing that

$\boldsymbol{\tau}_s(n^x)$ is an integer, the second claim is proved.

By combining the previous lemma with Properties 2.2 and 2.3, we can estimate the eccentricity of each vertex.

THEOREM 5.1. *For each vertex s and for each x between 0 and 1,*

$$\mathrm{ecc}(s) \leq \boldsymbol{\tau}_s(n^x) + \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(T(1 \to n^{1-x}) + \frac{\log n}{-\log c} - x \right) \right\rfloor.$$

Furthermore, for each s and for each $x \geq \frac{1}{2}$:

$$\mathrm{ecc}(s) \geq \boldsymbol{\tau}_s(n^x) + \left\lceil (1 + \mathcal{O}(\varepsilon)) \left(T(1 \to n^{1-x}) + \frac{\log n}{-\log c} - x \right) \right\rceil - 2.$$

Proof. By Property 2.2, for each vertex t, $\mathrm{dist}(s,t) \leq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^{1-x+\varepsilon}) - 1$. By Lemma 5.1, for each t,

$$\boldsymbol{\tau}_t(n^{1-x+\varepsilon}) \leq \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(T(\deg(t) \to n^{1-x+\varepsilon}) + \frac{\log n}{-\log c} + 1 - x + \varepsilon \right) \right\rfloor,$$

and consequently $\mathrm{ecc}(s) = \max_{t \in V} \mathrm{dist}(s,t) \leq \boldsymbol{\tau}_s (n^x) + \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^{1-x+\varepsilon} \right) + \frac{\log n}{-\log c} + 1 - x + \varepsilon \right) \right\rfloor - 1 = \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^{1-x} \right) + \frac{\log n}{-\log c} - x \right) \right\rfloor$.

For the other inequality, if $x \geq \frac{1}{2}$, let $y = 1 - x - \varepsilon < x$, and let T be the set of vertices t such that $\boldsymbol{\tau}_t (n^y) \geq \left\lceil (1 - 3\varepsilon) \left(T \left(1 \rightarrow n^y \right) + \frac{\log n}{-\log c} - 1 + y \right) \right\rceil$ (by Lemma 5.1, $|T| \geq n^{2\varepsilon}$). By Property 2.3, there is at least a vertex $t \in T$ verifying $\mathrm{dist}(s,t) \geq \boldsymbol{\tau}_s (n^x) + \boldsymbol{\tau}_t (n^y) - 2 \geq \boldsymbol{\tau}_s (n^x) + \left\lceil (1 - 3\varepsilon) \left(T \left(1 \rightarrow n^y \right) + \frac{\log n}{-\log c} - 1 + y \right) - 2 \right\rceil$. The second claim follows.

Thanks to this lemma, we can compute the diameter of a graph as the maximum eccentricity.

THEOREM 5.2. *For each x, the diameter of our graph is $D = \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(\mathrm{dist}_{\mathrm{avg}} (n) + \frac{2 \log n}{-\log c} \right) \right\rfloor$.*

Proof. By combining the upper bounds in Theorem 5.1 and Lemma 5.1, we can prove that $D \leq \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^x \right) + \frac{\log n}{-\log c} + x \right) \right\rfloor + \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^{1-x} \right) + \frac{\log n}{-\log c} - x \right) \right\rfloor$. If we choose x such that $(1 + \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^x \right) + \frac{\log n}{-\log c} + x \right) = i - \varepsilon$, we obtain that $D \leq i - 1 + \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^{1-x} \right) + \frac{\log n}{-\log c} - x \right) \right\rfloor \leq \left\lfloor i + (1 + \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^{1-x} \right) + \frac{\log n}{-\log c} - x - 1 \right) \right\rfloor \leq \left\lfloor (1 + \mathcal{O}(\varepsilon)) \left(\mathrm{dist}_{\mathrm{avg}} (n) + \frac{2 \log n}{-\log c} \right) \right\rfloor$.

Let us combine the lower bounds in Theorem 5.1 and Lemma 5.1: we obtain that $D \geq \left\lceil (1 - \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^x \right) + \frac{\log n}{-\log c} + x - 1 \right) \right\rceil + \left\lceil (1 - \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^{1-x} \right) + \frac{\log n}{-\log c} - x - 1 \right) \right\rceil - 1$. For all but a constant number of values of x, this value is equal to $\left\lfloor (1 - \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^x \right) + \frac{\log n}{-\log c} + x \right) \right\rfloor + \left\lfloor (1 - \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^{1-x} \right) + \frac{\log n}{-\log c} - x \right) \right\rfloor - 1$. Furthermore, if $(1 - \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^x \right) + \frac{\log n}{-\log c} + x \right) = i + \varepsilon$ for some integer i, this value is $\left\lfloor i + (1 - \mathcal{O}(\varepsilon)) \left(T \left(1 \rightarrow n^{1-x} \right) + \frac{\log n}{-\log c} - x \right) \right\rfloor - 1 \geq \left\lfloor (1 - \mathcal{O}(\varepsilon)) \left(\mathrm{dist}_{\mathrm{avg}} (n) + \frac{2 \log n}{-\log c} \right) \right\rfloor$. A similar argument can be applied if the second term is $i + \varepsilon$: hence, it only remains to prove that we can find a value of x between $\frac{1}{2}$ and 1 such that one of the two parts is close to an integer. This is true because $T \left(1 \rightarrow n^x \right) + x$ is continuous and increasing with respect to x, and $T \left(1 \rightarrow n^{1-x} \right) - x$ is continuous and decreasing. Since

the increase and the decrease are at least $\frac{1}{2}$, the sum of the two is at least 1.

Given Theorem 3.1, and given the values in Table 2, Theorem 5.2 gives diameter bounds for power law graphs generated through the models considered (since the four properties hold for each ε, we can safely let ε tend to 0, and transform $\mathcal{O}(\varepsilon)$ into $o(1)$). As far as we know, the bound for $1 < \beta < 2$ is new, while the other bounds are already known [25, 10].

COROLLARY 5.1. *If λ is a power law degree distribution with exponent β, the diameter of a random graph with degree distribution λ is:*

- *if $1 < \beta < 2$, $D = \left\lfloor 3 + \frac{\beta - 2}{\beta - 1} \right\rfloor$;*

- *if $2 < \beta < 3$, $D = (1 + o(1)) \left(\frac{2}{-\log \eta(1)} \right) \log n$;*

- *if $\beta > 3$,*

$$D = (1 + o(1)) \left(\frac{2}{-\log \eta(1)} + \frac{1}{\log M_1(\mu)} \right) \log n.$$

All the previous results deal with "extremal" properties of the distance distribution. Instead, the next results deal with properties that hold on average. Let us start by estimating the farness of a node s, that is, $\sum_{t \in V} \mathrm{dist}(s,t)$.

THEOREM 5.3. *For each vertex s and for each $0 < x < 1$, the farness $f(s)$ of s verifies*

$$f(s) \leq n(1 + \mathcal{O}(\varepsilon))(\boldsymbol{\tau}_s (n^x) - T \left(1 \rightarrow n^x \right) + \mathrm{dist}_{\mathrm{avg}} (n)) - \deg(s).$$

Proof. By Property 2.2, for each vertex t, $\mathrm{dist}(s,t) \leq \boldsymbol{\tau}_s (n^x) + \boldsymbol{\tau}_t \left(n^{1-x+\varepsilon} \right) - 1$, and hence

$$\sum_{t \in V} \mathrm{dist}(s,t)$$

$$\leq n \left(\boldsymbol{\tau}_s (n^x) - 1 \right) + \sum_{d=1}^{+\infty} \sum_{\deg(t) = d} \boldsymbol{\tau}_t \left(n^{1-x+\varepsilon} \right)$$

$$= n \left(\boldsymbol{\tau}_s (n^x) - 1 \right) + \sum_{d=1}^{+\infty} |\{t \in V : \deg(t) = d\}| \cdot$$

$$T \left(d \rightarrow n^{1-x+\varepsilon} \right) - n$$

$$\leq n \left(\boldsymbol{\tau}_s (n^x) - 1 \right) + n(1 + o(1)) T \left(1 \rightarrow n^{1-x+\varepsilon} \right)$$

$$+ nT \left(1 \rightarrow n^{x-\varepsilon} \right) - nT \left(1 \rightarrow n^{x-\varepsilon} \right)$$

$$\leq n(1 + \mathcal{O}(\varepsilon)) \left(\boldsymbol{\tau}_s (n^x) - T \left(1 \rightarrow n^x \right) + \mathrm{dist}_{\mathrm{avg}} (n) \right).$$

We need to subtract $\deg(s)$ from this result. To this purpose, we observe that in the first estimate, all

neighbors of s with degree at most n^{1-x} were given a distance $\boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^{1-x+\varepsilon}) - 1 = \boldsymbol{\tau}_s(n^x) + 2 - 1 \geq 2$, and consequently the other estimates remain correct if we subtract the number of neighbors of s with degree at most n^{1-x}, or equivalently if we subtract $\deg(s)$ and we sum the number of neighbors of s with degree at least n^{1-x}. Since $|E| \leq n^{1+\varepsilon}$ by Property 2.4, the number of vertices with degree at least n^{1-x} is at most $n^{x+\varepsilon}$, and the latter contribution is negligible.

THEOREM 5.4. *For each vertex s and for each $\frac{1}{2} \leq x < 1$,*

$$f(s) \geq n(1 - \mathcal{O}(\varepsilon)) \cdot$$
$$(\boldsymbol{\tau}_s(n^x) - T(1 \to n^x) + \text{dist}_{\text{avg}}(n) - 1).$$

Proof. Let s be any vertex, and let us apply Property 2.3 with $T = V$: there are at most $\mathcal{O}(n^{1-\varepsilon})$ vertices $t \in V$ such that $\text{dist}(s,t) < \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^{1-x-2\varepsilon}) - 2$. Let $T' := \{t \in V : \text{dist}(s,t) \geq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^{1-x-2\varepsilon}) - 2\}$.

$$f(s) = \sum_{t \in V} \text{dist}(s,t)$$
$$\geq \sum_{t \in V'} \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^{1-x-2\varepsilon}) - 2$$
$$= n(1 - o(1))(\boldsymbol{\tau}_s(n^x) - 2)$$
$$+ \sum_{t \in V} \boldsymbol{\tau}_t(n^{1-x-2\varepsilon}) - \boldsymbol{\tau}_t(n^{1-x-2\varepsilon})$$
$$= n(1 - \mathcal{O}(\varepsilon))(\boldsymbol{\tau}_s(n^x) - T(1 \to n^x) + \text{dist}_{\text{avg}}(n)$$
$$- 1) - \sum_{i=1}^{+\infty} |V - V'|\mathcal{O}(\log n)$$
$$= n(1 - \mathcal{O}(\varepsilon))(\boldsymbol{\tau}_s(n^x) - T(1 \to n^x)$$
$$+ \text{dist}_{\text{avg}}(n) - 1).$$

By computing the inverse of the farness, we can compute the closeness centrality of a vertex.

COROLLARY 5.2. *For each x such that $\frac{1}{2} \leq x < 1$, the closeness centrality of a vertex s verifies* $\frac{1 - \mathcal{O}(\varepsilon)}{n(\boldsymbol{\tau}_s(n^x) - T(1 \to n^x) + \text{dist}_{\text{avg}}(n)) - \deg(s)} \leq c(s) \leq \frac{1 + \mathcal{O}(\varepsilon)}{n(\boldsymbol{\tau}_s(n^x) - T(1 \to n^x) + \text{dist}_{\text{avg}}(n) - 1)}.$

COROLLARY 5.3. *The average distance between two vertices is between $(1 - \mathcal{O}(\varepsilon)) \text{dist}_{\text{avg}}(n) - 1$ and $(1 + \mathcal{O}(\varepsilon)) \text{dist}_{\text{avg}}(n)$.*

Proof. The average distance is the sum of the farness of all vertices, divided by $n(n-1)$. By the two previous theorems, for each $x \geq \frac{1}{2}$, $n(1 + \mathcal{O}(\varepsilon))(\boldsymbol{\tau}_s(n^x) - T(1 \to n^x) + \text{dist}_{\text{avg}}(n) - 1) \leq f(s) \leq n(1 + \mathcal{O}(\varepsilon))(\boldsymbol{\tau}_s(n^x) - T(1 \to n^x) + \text{dist}_{\text{avg}}(n))$.

Let us compute $\sum_{s \in V} \boldsymbol{\tau}_s(n^x) = \sum_{d=1}^{+\infty} \sum_{\deg(s)=d} \boldsymbol{\tau}_s(n^x) = \sum_{d=1}^{\infty} |\{s : \deg(s) = d\}| T(d \to n^x) = (1 + o(1))T(1 \to n^x)$. Combining this estimate with the previous equation, we obtain:

$$(1 - \mathcal{O}(\varepsilon)) \text{dist}_{\text{avg}}(n) - 1 \leq \frac{1}{n(n-1)} \sum_{s \in V} f(s)$$
$$\leq (1 + \mathcal{O}(\varepsilon)) \text{dist}_{\text{avg}}(n).$$

Again, assuming Theorem 3.1, and given the values in Table 2, we have proved the following asymptotics for the average distance in random graphs.

COROLLARY 5.4. *If λ is a power law degree distribution with exponent β, the average distance in a random graph with degree distribution λ is:*

- *if $1 < \beta < 2$, $2 - o(1) \leq \text{dist}_{\text{avg}} \leq 3 + o(1)$;*

- *if $2 < \beta < 3$, $\text{dist}_{\text{avg}} = (2 + o(1))\left(\log_{\frac{1}{\beta-1}} \log n\right)$;*

- *if $\beta > 3$, $\text{dist}_{\text{avg}} = (1 + o(1))\frac{\log n}{\log M_1(\mu)}$.*

6 Bounding the Diameter Through Sampling.

The first algorithm we analyze is very simple: it lower bounds the diameter of a graph by performing k BFSes from random nodes $s_1, \ldots, s_k$, and returning $\max_{i=1,\ldots,k} \text{ecc}(s_i)$. Clearly, the running time is $\mathcal{O}(mk)$: we want to analyze the error of this method on graphs that satisfy our assumptions. The main idea behind this analysis is that $\text{ecc}(s)$ is strongly correlated with $\boldsymbol{\tau}_s(n^x)$, and the number of vertices verifying $\boldsymbol{\tau}_s(n^x) > \alpha$ decreases exponentially with respect to α. This means that the number of vertices with high eccentricity is very small, and it is difficult to find them by sampling: this means that the error should be quite big.

More formally, by Theorem 5.1, the eccentricity of a vertex s verifies $\text{ecc}(s) \leq \boldsymbol{\tau}_s(n^x) + \left\lfloor (1 + \mathcal{O}(\varepsilon))\left(T(1 \to n^{1-x}) + \frac{\log n}{-\log c} - x\right)\right\rfloor$, and consequently the output is at most $\max_{i=1,\ldots,k} \boldsymbol{\tau}_{s_i}(n^x) + \left\lfloor (1 + \mathcal{O}(\varepsilon))\left(T(1 \to n^{1-x}) + \frac{\log n}{-\log c} - x\right)\right\rfloor$. We want to estimate $\max_{i=1,\ldots,k} \boldsymbol{\tau}_{s_i}(n^x)$ through Property 2.1: the number of vertices s verifying $\boldsymbol{\tau}_s(n^x) \geq (1 + \mathcal{O}(\varepsilon))(T(1 \to n^x) + \alpha)$ is at most $nc^{\alpha-x}$, and consequently a random set of k vertices does not contain any such vertex, a.a.s., if $k \leq \frac{n^{1-\varepsilon}}{nc^{\alpha-x}} \ll \frac{n}{nc^{\alpha-x}}$. Solving the first inequality with respect to α, we obtain $\alpha \geq x + \frac{\varepsilon \log n + \log k}{-\log c}$.

We conclude that, a.a.s., if $\alpha = x + \frac{\varepsilon \log n + \log k}{-\log c}$, we do not perform any BFS from a vertex s such that $\boldsymbol{\tau}_s(n^x) \geq (1 + \varepsilon)(T(1 \to n^x) + \alpha) = (1 + $

$\mathcal{O}(\varepsilon))\left(T\left(1 \to n^x\right) + \frac{\gamma \log n}{-\log c} + x\right)$. This means that, for a suitable choice of x, the output is smaller than:

$$
\max_{s \in X} \boldsymbol{\tau}_s\left(n^x\right)
$$
$$
+ \left\lfloor (1 + \mathcal{O}(\varepsilon))\left(T\left(1 \to n^{1-x}\right) + \frac{\log n}{-\log c} - x\right)\right\rfloor
$$
$$
\leq \left\lfloor (1 + \mathcal{O}(\varepsilon))\left(T\left(1 \to n^x\right) + \frac{\gamma \log n}{-\log c} + x\right)\right\rfloor
$$
$$
+ \left\lfloor (1 + \mathcal{O}(\varepsilon))\left(T\left(1 \to n^{1-x}\right) + \frac{\log n}{-\log c} - x\right)\right\rfloor
$$
$$
\leq \left\lfloor (1 + \mathcal{O}(\varepsilon))\left(\mathrm{dist}_{\mathrm{avg}}(n) + \frac{(1+\gamma)\log n}{-\log c} - 1\right)\right\rfloor.
$$

By replacing the values in Table 2, we obtain the desired results. In order to obtain a lower bound on the error, it is enough to perform similar computations after replacing ε with $-\varepsilon$.

7 The 2-Sweep Heuristic.

The 2-Sweep heuristic [31] finds a lower bound on the diameter, by performing a BFS from a vertex s, finding a vertex t that maximizes the distance from s, and returning the eccentricity of t (since only 2 BFSes are performed, the running time is linear in the input size). Following the intuitive proof, let us show that $\boldsymbol{\tau}_t\left(n^x\right)$ is high, and consequently the eccentricity of t, which depends on $\boldsymbol{\tau}_t\left(n^x\right)$, is high as well.

Lemma 7.1. *For each vertex s, let t be a vertex maximizing the distance from s. Then, for each $x \geq \frac{1}{2}$, $\boldsymbol{\tau}_t\left(n^{1-x}\right) \geq \left\lceil (1 - \mathcal{O}(\varepsilon))\left(T\left(1 \to n^{1-x}\right) + \frac{\log n}{-\log c} - 1 - x\right)\right\rceil$.*

Proof. By Property 2.2 and Theorem 5.1, $\boldsymbol{\tau}_s\left(n^{x+\varepsilon}\right) + \left\lceil (1 - \mathcal{O}(\varepsilon))\left(T\left(1 \to n^{1-x-2\varepsilon}\right) + \frac{\log n}{-\log c} - 2 - x\right)\right\rceil \leq \mathrm{ecc}(s) = \mathrm{dist}(s,t) \leq \boldsymbol{\tau}_s\left(n^{x+\varepsilon}\right) + \boldsymbol{\tau}_t\left(n^{1-x}\right) - 1$. From this inequality, we obtain that $\boldsymbol{\tau}_t\left(n^{1-x}\right) \geq \left\lceil (1 - \mathcal{O}(\varepsilon))\left(T\left(1 \to n^{1-x}\right) + \frac{\log n}{-\log c} - 1 - x\right)\right\rceil$.

By Theorems 5.1 and 5.2, if t is the vertex maximizing the distance from s:

$$
\mathrm{ecc}(t) \geq \boldsymbol{\tau}_t\left(n^{\frac{1}{2}}\right)
$$
$$
+ \left[(1 - \mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{1}{2}\right)\right] - 2
$$
$$
\geq \left[(1 - \mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{3}{2}\right)\right]
$$
$$
+ \left[(1 - \mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{5}{2}\right)\right]
$$
$$
\geq 2\left\lfloor (1 - \mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{1}{2}\right)\right\rfloor - 1
$$
$$
\geq 2\left\lfloor (1 - \mathcal{O}(\varepsilon))\frac{D}{2}\right\rfloor - 1
$$

(in this analysis, we used that $T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c}$ is not an integer, and that ε is small enough).

We conclude that the output of the 2-Sweep heuristic is $2\left\lfloor (1 - \mathcal{O}(\varepsilon))\frac{D}{2}\right\rfloor - 1$, proving the results in Table 1.

8 The RW Algorithm.

The RW algorithm [36] is a randomized algorithm that computes a $\frac{3}{2}$-approximation of the diameter of a graph, in time $\Theta(m\sqrt{n}\log n)$. The algorithm works as follows: we choose $k = \Theta(\sqrt{n}\log n)$ vertices $s_1, \ldots, s_k$, and we perform a BFS from each of these vertices. Then, we compute the vertex t maximizing $\min_{i=1,\ldots,k} \mathrm{dist}(s_i, t)$, and we let $t_1, \ldots, t_k$ be the k vertices closest to t. Then, if there exist i, j such that $s_i = t_j$, we return the maximum eccentricity among the s_is and the t_js, otherwise the algorithm fails, and we can decide to run it again (anyway, the probability that it fails are small).

For the worst-case analysis of this algorithm, we refer to [36]: in this work, we analyze its performances in our framework. The running time is still $\Theta(m\sqrt{n}\log n)$, since the algorithm requires $\Theta(\sqrt{n}\log n)$ BFSes from the vertices s_i, t_j, but the approximation factor can be better than the worst-case. Intuitively, this algorithm is quite similar to the 2-Sweep heuristic (if k was 1, the algorithm would be the 2-Sweep heuristic), and we conjecture that also its behavior should be similar.

For a formal proof, let v be any vertex: since the vertices s_i are random, $\mathbb{P}\left(\forall i, s_i \in \boldsymbol{N}^{\boldsymbol{\tau}_v(n^x)}(v)\right) \geq \left(1 - \frac{n^x}{n}\right)^{\sqrt{n}\log n} = e^{n^{x - \frac{1}{2} + o(1)}}$, and similarly $\mathbb{P}\left(\forall i, s_i \notin \boldsymbol{N}^{\boldsymbol{\tau}_v(n^x)-1}(v)\right) \leq \left(1 - \frac{\mathcal{O}(n^x \log n)}{n}\right)^{\sqrt{n}\log n} = e^{n^{x - \frac{1}{2} + o(1)}}$ (because there are at most $n^x \boldsymbol{\tau}_v\left(n^x\right) = \mathcal{O}(n^x \log n)$ vertices in $\boldsymbol{N}^{\boldsymbol{\tau}_v(n^x)-1}(v)$). This means that $\min_{i=1,\ldots,k} \mathrm{dist}(s_i, v) \leq \boldsymbol{\tau}_v\left(n^{\frac{1+\varepsilon}{2}}\right)$ w.h.p.,

and $\min_{i=1,\dots,k} \operatorname{dist}(s_i,v) \geq \boldsymbol{\tau}_v\left(n^{\frac{1-\varepsilon}{2}}\right) - 1$ a.a.s.. Hence, if v is one of the vertices maximizing $\boldsymbol{\tau}_v\left(n^{\frac{1}{2}}\right)$, $\min_{i=1,\dots,k} \operatorname{dist}(s_i,v) \geq \boldsymbol{\tau}_v\left(n^{\frac{1-\varepsilon}{2}}\right) - 1 \geq \left\lceil (1-\mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{1}{2}\right)\right\rceil - 1$ a.a.s., by Lemma 5.1.

This means that the vertex t maximizing $\min_{i=1,\dots,k} \operatorname{dist}(s_i,t)$ verifies $\boldsymbol{\tau}_t\left(n^{\frac{1+\varepsilon}{2}}\right) \geq \min_{i=1,\dots,k} \operatorname{dist}(s_i,t) \geq \min_{i=1,\dots,k} \operatorname{dist}(s_i,v) \geq \left\lceil (1-\mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{1}{2}\right)\right\rceil$.

This means that $\operatorname{ecc}(t) \geq \boldsymbol{\tau}_t\left(n^{\frac{1}{2}}\right) + \left\lceil (1-\mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{1}{2}\right)\right\rceil - 2 \geq \left\lceil (1-\mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{1}{2}\right)\right\rceil - 1 + \left\lceil (1-\mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}-\varepsilon}\right) + \frac{\log n}{-\log c} - \frac{1}{2}\right)\right\rceil - 2 = 2\left\lceil (1-\mathcal{O}(\varepsilon))\left(T\left(1 \to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{1}{2}\right)\right\rceil - 3$, and this value is exactly the same value we obtained for the 2-Sweep heuristic.

Note that this analysis only uses the vertex t, and a more refined analysis could (in principle) obtain better bounds.

9 The SumSweep Heuristic.

The SumS heuristic [13, 14] provides a lower bound on the eccentricity of all vertices, by performing some BFSes from vertices $t_1,\dots,t_k$, and defining $L(v) = \max_{i=1,\dots,k} \operatorname{dist}(v,t_i) \leq \operatorname{ecc}(v)$ for each vertex v. The vertices t_i are chosen as follows: we start from a random vertex s_1, then we choose t_1 as the vertex maximizing $\operatorname{dist}(s_1,t_1)$. Then, we choose again s_2 as a random vertex, and we choose t_2 as the vertex in $V-\{t_1\}$ maximizing $\operatorname{dist}(s_1,t_2)+\operatorname{dist}(s_2,t_2)$. In general, after $2i$ BFSes are performed, we choose a random vertex s_{i+1}, we perform a BFS from s_{i+1}, and we choose t_{i+1} as the vertex in $V-\{t_1,\dots,t_i\}$ maximizing $\sum_{j=1}^{i+1} \operatorname{dist}(s_j,t_{i+1})$.[3]

The idea behind the analysis of the 2-Sweep heuristic and the RW algorithm is to exploit the existence of few vertices t with big values of $\boldsymbol{\tau}_t(n^x)$: both algorithms find a single vertex t such that $\boldsymbol{\tau}_t(n^x)$ is high, and they lower bound the diameter with the eccentricity of this vertex, which is peripheral. Instead, the SumS heuristic highlights all vertices t_i having big values of $\boldsymbol{\tau}_t(n^x)$, and it performs a BFS from each of these vertices. Then, for each vertex v, if t is the vertex farthest from v, $\boldsymbol{\tau}_t(n^x)$ is

---

[3]Actually, in the original SumS heuristic, there is no distinction between the vertices s_i and t_i: we simply choose t_{i+1} as the vertex maximizing $\sum_{j=1}^{i} \operatorname{dist}(t_j,t_{i+1})$. However, for our analyses, we need to use this variation.

big, and this means that $t = t_i$ for some small i. Consequently, if we lower bound $\operatorname{ecc}(s) \geq \max_{i=1,\dots,k} \operatorname{ecc}(t_i)$, the lower bounds obtained are tight after few steps. Let us formalize this intuition: first, we need to prove that the vertices t with high value of $\boldsymbol{\tau}_t\left(n^{1-x}\right)$ are chosen soon by this procedure.

LEMMA 9.1. *Let S be a random set of vertices, let $\boldsymbol{\tau}_S(n^y) = \frac{1}{|S|}\sum_{s \in S} \boldsymbol{\tau}_s(n^y)$, and let t be any vertex in the graph. Then, $\frac{\sum_{s \in S} \operatorname{dist}(s,t)}{|S|} \leq \boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_S\left(n^{1-x+\varepsilon}\right) - 1$. Furthermore, if $|S| > n^{3\varepsilon}$, $x \geq \frac{1}{2}$, $\frac{\sum_{s \in S} \operatorname{dist}(s,t)}{|S|} \geq (1-o(1))\left(\boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_S\left(n^{1-x-2\varepsilon}\right) - 1\right)$, w.h.p..*

Proof. For the upper bound, by Property 2.2, $\sum_{s \in S} \operatorname{dist}(s,t) \leq \sum_{s \in S} \boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_s\left(n^{1-x+\varepsilon}\right) - 1 = |S|\left(\boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_S\left(n^{1-x+\varepsilon}\right) - 1\right)$.

For the lower bound, by Property 2.3, for each vertex t, the number of vertices $s \in V$ verifying $\operatorname{dist}(s,t) < \boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_s\left(n^{1-x-2\varepsilon}\right) - 2$ is at most $n^{1-\varepsilon}$. Let $S' \subseteq S$ be the set of vertices verifying $\operatorname{dist}(s,t) \geq \boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_s\left(n^{1-x-2\varepsilon}\right) - 2$: since S is random, the probability that a vertex $s \in S$ does not belong to S' is at least $n^{-\varepsilon}$. From this bound, we want to prove that $|S'| \geq (1-\mathcal{O}(n^{-\varepsilon}))|S|$, using Hoeffding's inequality. For each $s \in S$, let $\boldsymbol{X}_s = 1$ if $\operatorname{dist}(s,t) \geq \boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_s\left(n^{1-x-2\varepsilon}\right) - 2$, 0 otherwise: clearly, $|S'| = \sum_{s \in S} \boldsymbol{X}_s$, the variables $\boldsymbol{X}_s$ are independent, and $\mathbb{P}(\boldsymbol{X}_s = 1) \geq 1 - n^{-\varepsilon}$. By Hoeffding's inequality, $\mathbb{P}\left(\sum_{s \in S} \boldsymbol{X}_s < \mathbb{E}\left[\sum_{s \in S} \boldsymbol{X}_s\right] - \lambda\right) \leq e^{-\frac{\lambda^2}{|S|}}$. Since $\mathbb{E}\left[\sum_{s \in S} \boldsymbol{X}_s\right] \geq |S|(1-n^{-\varepsilon})$, if we choose $\lambda = |S|n^{-\varepsilon}$, we obtain that $\mathbb{P}\left(|S'| < (1-2n^{-\varepsilon})|S|\right) \leq e^{-|S|n^{-2\varepsilon}}$. We proved that, w.h.p., $|S'| \geq |S|(1-\mathcal{O}(n^{-\varepsilon}))$. As a consequence:

$$\sum_{s \in S} \operatorname{dist}(s,t)$$
$$\geq \sum_{s \in S'} \operatorname{dist}(s,t)$$
$$\geq \sum_{s \in S'} \boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_s\left(n^{1-x-2\varepsilon}\right) - 2$$
$$\geq |S'|\left(\boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_S\left(n^{1-x-2\varepsilon}\right) - 2\right) - \sum_{s \in S-S'} \mathcal{O}(\log n)$$
$$\geq |S'|\left(\boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_S\left(n^{1-x-2\varepsilon}\right) - 2\right) - \mathcal{O}\left(n^{-\varepsilon}|S|\log n\right)$$
$$\geq (1-o(1))|S|\left(\boldsymbol{\tau}_t(n^x) + \boldsymbol{\tau}_S\left(n^{1-x-2\varepsilon}\right) - 2\right).$$

By Lemma 7.1, if t maximizes $\operatorname{dist}(u,t)$ for some $u \in V$, then for each $y \geq \frac{1}{2}$, $\boldsymbol{\tau}_t\left(n^{1-y-2\varepsilon}\right) \geq \left\lceil (1-\mathcal{O}(\varepsilon))\left(T\left(1 \to n^{1-y}\right) + \frac{\log n}{-\log c} - 1 - y\right)\right\rceil$. If we choose $x = y = \frac{1}{2}$, the previous

lemma proves that, if a vertex v is chosen before t in this procedure, then $\boldsymbol{\tau}_v\left(n^{\frac{1}{2}+3\varepsilon}\right) + \boldsymbol{\tau}_S\left(n^{\frac{1}{2}-2\varepsilon}\right) - 1 \geq \frac{\sum_{s\in S}\operatorname{dist}(s,v)}{|S|} \geq \frac{\sum_{s\in S}\operatorname{dist}(s,t)}{|S|} \geq (1-o(1))\left(\boldsymbol{\tau}_t\left(n^{\frac{1}{2}}\right) + \boldsymbol{\tau}_S\left(n^{\frac{1}{2}-2\varepsilon}\right) - 2\right)$.
Rearranging this inequality, we obtain $\boldsymbol{\tau}_v\left(n^{\frac{1}{2}+3\varepsilon}\right) \geq (1-o(1))\left(\boldsymbol{\tau}_t\left(n^{\frac{1}{2}}\right) - 1\right) \geq \left\lceil(1-\mathcal{O}(\varepsilon))\left(T\left(1\to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{5}{2}\right)\right\rceil = \left\lfloor(1-\mathcal{O}(\varepsilon))\left(T\left(1\to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{3}{2}\right)\right\rfloor$.

If we apply Property 2.1 with the value of α verifying $(1+\varepsilon)\left\lfloor T\left(1\to n^{\frac{1}{2}+3\varepsilon}\right) + \alpha\right\rfloor = \left\lfloor(1-\mathcal{O}(\varepsilon))\left(T\left(1\to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{3}{2}\right)\right\rfloor$, we obtain that the number of vertices v satisfying the latter equation is $\mathcal{O}(nc^{\alpha-x}) = nc^{\left\lfloor(1-\mathcal{O}(\varepsilon))\left(T\left(1\to n^{\frac{1}{2}}\right) + \frac{\log n}{-\log c} - \frac{3}{2}\right)\right\rfloor - T\left(1\to n^{\frac{1}{2}+3\varepsilon}\right) - \frac{1}{2}}$.
If $\beta > 2$, this value is simply $\mathcal{O}(\varepsilon)$, while if $1 < \beta < 2$, this value is $n^{1 - \frac{2-\beta}{\beta-1}\left(\left\lfloor\frac{\beta-1}{2-\beta} - \frac{3}{2}\right\rfloor - \frac{1}{2}\right)}$, if ε is small enough.

We conclude that, after $n^{3\varepsilon} + \mathcal{O}(nc^{\alpha-x})$ BFSes, we have performed a BFS from all vertices t that maximize $\operatorname{dist}(u,t)$ for some $u \in V$: this means that the lower bounds on all eccentricities are tight.

10   The iFub Algorithm.

The IFUB algorithm is an exact algorithm to compute the diameter of a graph. Its worst-case running time is $\mathcal{O}(mn)$, but it performs much better in practice [22]. It works as follows: it performs a BFS from a vertex v, and it uses the fact that, if $D = \operatorname{dist}(s,t)$, either $\operatorname{dist}(s,v) \geq \frac{D}{2}$, or $\operatorname{dist}(v,t) \geq \frac{D}{2}$. Hence, after the first BFS from v, the IFUB algorithm computes the eccentricity of all the other vertices, in decreasing order of distance from v. During this process, it keeps track of the maximum eccentricity found D_L, which is a lower bound on the diameter. As soon as we are processing a vertex s such that $\operatorname{dist}(v,s) \leq \frac{D_L}{2}$, we know that, for each pair (s,t) of unprocessed vertices, $\operatorname{dist}(s,t) \leq 2\frac{D_L}{2} = D_L$: this means that we have processed at least one of the vertices in a diametral pair, and $D_L = D$. The running time is $\mathcal{O}\left(mN_{\frac{D}{2}}(v)\right)$, where $N_{\frac{D}{2}}(v)$ is the number of vertices at distance at least $\frac{D}{2}$ from v (indeed, the algorithm performs a BFS from each of these vertices).

For our analysis, we only need to estimate $N_{\frac{D}{2}}(v)$. Intuitively, the diameter is the sum of two contributions: one is $\operatorname{dist}_{\text{avg}}(n)$, which is close to the average distance between two nodes, and the other is twice the maximum deviation from this value, that is, $2\frac{\log n}{-\log c}$. Hence, $N_{\frac{D}{2}}(v)$ is the number of vertices at distance $\frac{\operatorname{dist}_{\text{avg}}(n)}{2} + \frac{\log n}{-\log c}$ from v: if the second term is dominant (for

instance, if $2 < \beta < 3$), we are considering only vertices with very big deviations, and the time is much smaller than n. Conversely, if the deviation is smaller than $\frac{\operatorname{dist}_{\text{avg}}(n)}{2}$, we expect this number to be $\mathcal{O}(n)$ (for instance, if $\beta > 3$ and $\eta(1)$ is small).

Let us formalize this intuition. By Theorem 5.2, $\frac{D}{2} \geq \frac{1}{2}\left\lfloor(1+\mathcal{O}(\varepsilon))\left(\operatorname{dist}_{\text{avg}}(n) + \frac{2\log n}{-\log c}\right)\right\rfloor$. In order to estimate $N_{\frac{D}{2}}(v)$, we use the fact that if $x + y \geq 1 + \varepsilon$, $\operatorname{dist}(v,w) \leq \boldsymbol{\tau}_v(n^x) + \boldsymbol{\tau}_w(n^y) - 1$, and consequently, if $\operatorname{dist}(v,w) \geq \frac{D}{2}$, then $\boldsymbol{\tau}_w(n^y) \geq \left\lceil\frac{D}{2} + 1 - \boldsymbol{\tau}_v(n^x)\right\rceil \geq \left\lfloor\frac{1}{2}(1+\mathcal{O}(\varepsilon))\left(\operatorname{dist}_{\text{avg}}(n) + \frac{2\log n}{-\log c}\right)\right\rfloor + 1 - \boldsymbol{\tau}_v(n^x)$.

Let us apply Property 2.1 with α such that $(1+\varepsilon)(T(d\to n^y) + \alpha) = \left\lfloor\frac{1}{2}(1+\mathcal{O}(\varepsilon))\left(\operatorname{dist}_{\text{avg}}(n) + \frac{2\log n}{-\log c}\right)\right\rfloor + 1 - \boldsymbol{\tau}_v(n^x)$: we obtain that $N_{\frac{D}{2}}(v) \leq nc^{\alpha-y} = n^{1-\frac{\log c}{\log n}(\alpha-y)}$. Let us estimate:

$$\alpha - y$$
$$\leq (1+\mathcal{O}(\varepsilon))\left(\left\lfloor\frac{1}{2}(1+\mathcal{O}(\varepsilon))\left(\operatorname{dist}_{\text{avg}}(n) + \frac{2\log n}{-\log c}\right)\right\rfloor \right.$$
$$\left. + 1 - \boldsymbol{\tau}_v(n^x) - T(d\to n^y)\right)$$
$$\leq (1+\mathcal{O}(\varepsilon))\left(\left\lfloor\frac{1}{2}\left(\operatorname{dist}_{\text{avg}}(n) + \frac{2\log n}{-\log c}\right)\right\rfloor \right.$$
$$\left. - \operatorname{dist}_{\text{avg}}(n) + T\left(1\to n^{\max\left(1-\varepsilon,\frac{1}{\beta-1}\right)}\right)\right).$$

For $\beta > 3$, the number of BFSs is $n^{(1+\mathcal{O}(\varepsilon))\frac{-\log c}{\log n}\left(\frac{1}{2}\operatorname{dist}_{\text{avg}}(n) - T\left(1\to n^{\frac{1}{\beta-1}}\right)\right)} = n^{(1+\mathcal{O}(\varepsilon))\frac{-\log \eta(1)}{\log n}\left(\frac{1}{2}-\frac{1}{\beta-1}\right)\frac{\log n}{\log M_1(\mu)}} = n^{\left(\frac{1}{2}-\frac{1}{\beta-1}+\mathcal{O}(\varepsilon)\right)\frac{-\log \eta(1)}{\log M_1(\mu)}}$. Hence, the running time is $\mathcal{O}\left(mN_{\frac{D}{2}}(v)\right) = n^{1+\left(\frac{1}{2}-\frac{1}{\beta-1}+\mathcal{O}(\varepsilon)\right)\frac{-\log \eta(1)}{\log M_1(\mu)}}$.

For $2 < \beta < 3$, the computation is similar, but $M_1(\mu)$ is infinite: the running time is $n^{1+\mathcal{O}(\varepsilon)}$.

Finally, for $1 < \beta < 2$, if v is the maximum degree vertex, this value is at most $n^{1-(1+\mathcal{O}(\varepsilon))\frac{2-\beta}{\beta-1}\left\lfloor\frac{3}{2}+\frac{\beta-1}{2-\beta}-2\right\rfloor} = n^{1+\mathcal{O}(\varepsilon)-\frac{2-\beta}{\beta-1}\left\lfloor\frac{\beta-1}{2-\beta}-\frac{1}{2}\right\rfloor}$. The running-time is $mn^{1-\frac{2-\beta}{\beta-1}\left\lfloor\frac{\beta-1}{2-\beta}-\frac{1}{2}\right\rfloor+\mathcal{O}(\varepsilon)} = n^{2-\frac{2-\beta}{\beta-1}\left\lfloor\frac{\beta-1}{2-\beta}-\frac{1}{2}\right\rfloor+\mathcal{O}(\varepsilon)}$.

11   The Exact SumSweep Algorithm.

The SUMS algorithm [13, 14] is based on keeping lower bounds $L(v)$ and upper bounds $U(v)$ on the eccentricity of each vertex v. In particular, assume that we have performed BFSes from vertices $s_1, \ldots, s_k$: we can set an upper (resp., lower) bound $U(v) = \min_{i=1,\ldots,k}(\operatorname{ecc}(s_i) + \operatorname{dist}(s_i,v))$ (resp., $L(v) = \max_{i=1,\ldots,k}(\operatorname{dist}(v_i,s))$) on the eccentricity of v. Furthermore the algorithm keeps a lower bound D_L (resp., an upper bound R_U) on the

diameter (resp., radius), defined as the maximum (resp., minimum) eccentricity of a processed vertex s_i. As soon as $D_L \geq \min_{v \in V} U(v)$, we can safely output D_L as the diameter; similarly, as soon as $R_U \leq \min_{v \in V} L(v)$, we know that R_U is the exact radius. It remains to define how the vertices $s_1, \ldots, s_k$ are chosen: we start by performing a SumSH, and after that we alternatively maximize L and minimize U (obviously, we never choose the same vertex twice). Actually, in order to perform the analysis in the case $\beta > 3$, we also need to perform a BFS from a vertex maximizing the degree every k steps, for some constant k (differently from the original SumS algorithm).

The analysis for the radius computation is very simple: after the initial SumSH, all lower bounds are tight w.h.p., and consequently it is enough to perform a further BFS from a vertex minimizing L to obtain the final value of R_U. Then, the running time is the same as the running time of the SumSH. For the diameter, the analysis is more complicated, because we have to check when all upper bounds are below the diameter, and the upper bounds are not tight, in general.

Intuitively, if $\beta < 3$, the radius is very close to half the diameter, and the first BFS is performed from a radial vertex s: consequently, after the first BFS, the upper bound of a vertex v becomes $\mathrm{ecc}(s) + \mathrm{dist}(s, v) \leq D$ if $\mathrm{dist}(s, v) \leq D - \mathrm{ecc}(s) = D - R \approx R = \mathrm{ecc}(s)$. This means that, after this BFS, we have to perform a BFS from each vertex whose distance from s is approximately the eccentricity of s, and there are not many such vertices, as shown by Lemma 7.1. Hence, we obtain that, in this regime, the running time of the EXACTSumS algorithm is the same as the running time of the initial SumSH. Conversely, if $\beta > 3$, a BFS from a vertex s sets upper bounds smaller than D to all vertices closer to s than $D - \mathrm{ecc}(s)$, and the number of such vertices is close to $M_1(\mu)^{D - \mathrm{ecc}(s)}$. Since $D - \mathrm{ecc}(s)$ is usually $\mathcal{O}(\log n)$, a BFS sets correct bounds to $M_1(\mu)^{\mathcal{O}(\log n)} = n^{\mathcal{O}(1)}$ vertices: hence, we expect the number of BFSes needed to be subquadratic.

11.1 The Case $1 < \beta < 3$. As we said before, the first BFS is performed from a radial vertex s: by Theorem 5.1, if $\overline{s}$ is a vertex maximizing $\boldsymbol{\tau}_s(n^x)$, $\mathrm{ecc}(s) \leq \mathrm{ecc}(\overline{s}) \leq \boldsymbol{\tau}_{\overline{s}}(n^x) + \left\lfloor (1 + \varepsilon)\left(T\left(1 \to n^{1-x}\right) + \frac{\log n}{-\log c} - x\right)\right\rfloor$. Let $x := 1 - \varepsilon$, if $1 < \beta < 2$, $x := \frac{1}{\beta - 1} - \varepsilon$ if $2 < \beta < 3$: this value is at most $\left\lfloor (1 + 2\varepsilon)\left(2 + \frac{\log n}{-\log c}\right)\right\rfloor$. As a consequence, after the first BFS, the algorithm sets upper bounds smaller

than D to any vertex closer to s than

$$D - \mathrm{ecc}(s) \geq \left\lfloor (1 - \mathcal{O}(\varepsilon))\left(\mathrm{dist}_{\mathrm{avg}}(n) + \frac{2\log n}{-\log c}\right)\right\rfloor$$
$$- \left\lfloor (1 + 2\varepsilon)\left(2 + \frac{\log n}{-\log c}\right)\right\rfloor$$
$$\geq (1 - \mathcal{O}(\varepsilon))\left(\left\lfloor \frac{2\log n}{-\log c}\right\rfloor - \left\lfloor \frac{\log n}{-\log c}\right\rfloor + 1\right).$$

This means that we only have to analyze vertices v such that $D - \mathrm{ecc}(s) \leq \mathrm{dist}(s, v) \leq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_v\left(n^{1-x+\varepsilon}\right) - 1$, that is:

$$\boldsymbol{\tau}_v\left(n^{1-x+\varepsilon}\right)$$
$$\geq D - \mathrm{ecc}(s) - \boldsymbol{\tau}_s(n^x) + 1$$
$$\geq (1 - \mathcal{O}(\varepsilon))\left(\left\lfloor \frac{2\log n}{-\log c}\right\rfloor - \left\lfloor \frac{\log n}{-\log c}\right\rfloor + 1\right).$$

Let us apply Property 2.1 with α such that $(1 + \varepsilon)\left(T\left(1 \to n^{1-x+\varepsilon}\right) + \alpha\right) = (1 - \mathcal{O}(\varepsilon))\left(\left\lfloor \frac{2\log n}{-\log c}\right\rfloor - \left\lfloor \frac{\log n}{-\log c}\right\rfloor + 1\right)$: we obtain that the number of vertices v that do not receive bounds smaller than D is at most

$$nc^{\alpha - 1 + x - \varepsilon}$$
$$= nc^{(1 - \mathcal{O}(\varepsilon))\left(\left\lfloor \frac{2\log n}{-\log c}\right\rfloor - \left\lfloor \frac{\log n}{-\log c}\right\rfloor + 1\right) - T\left(1 \to n^{1-x+\varepsilon}\right) - 1 + x - \varepsilon}$$
$$= nc^{\alpha - 1 + x - \varepsilon}$$
$$= nc^{(1 - \mathcal{O}(\varepsilon))\left(\left\lfloor \frac{2\log n}{-\log c}\right\rfloor - \left\lfloor \frac{\log n}{-\log c}\right\rfloor\right) - T\left(1 \to n^{1-x+\varepsilon}\right) - 1},$$

which is smaller than the number of iteration of the SumSH. Hence, the total running time is bounded by the time needed to perform the initial SumSH.

11.2 The Case $\beta > 3$. In the case $\beta > 3$, the previous argument does not work, because $\mathrm{dist}_{\mathrm{avg}}(n)$ can be small. We need a different technique: we prove that, for each vertex v and for some x, either $\boldsymbol{\tau}_v(n^x)$ is quite large, or there is a vertex s with high degree that is "not far" from s. After $\mathcal{O}(k)$ steps, we have performed a BFS from the k vertices with highest degree, and consequently all vertices which are quite close to one of these vertices have bounds smaller than D: this means that there are few vertices with upper bound bigger then D. Then, since every $\mathcal{O}(1)$ steps, the number of vertices with upper bound bigger than D decreases by 1, after few more BFSes, all upper bounds are smaller than or equal to D.

More formally, let $s_1, \ldots, s_k$ be all the vertices with degree bigger than n^x: by Property 2.4, $k = \frac{n^{1 \pm \mathcal{O}(\varepsilon)}}{n^{x(\beta - 1)}}$, and after at most $\frac{n^{1 + \mathcal{O}(\varepsilon)}}{n^{x(\beta - 1)}}$ BFSes (apart from the initial SumSH), we have performed a BFS from each of these vertices.

We start by estimating $\mathrm{ecc}(s_i)$, because, after the BFS from s_i, for each vertex v, $U(v) \leq \mathrm{dist}(v, s_i) + \mathrm{ecc}(s_i)$. By Theorem 5.1, $\mathrm{ecc}(s_i) \leq$ $\boldsymbol{\tau}_{s_i}(n^x) + \left\lfloor (1+\varepsilon)\left(T\left(1 \to n^{1-x}\right) + \frac{\log n}{-\log c} - x \right) \right\rfloor \leq$ $1 + (1+\varepsilon)\left((1-x)\frac{\log n}{\log M_1(\mu)} + \frac{\log n}{-\log c} \right) \leq$ $(1+2\varepsilon)\left((1-x)\frac{\log n}{\log M_1(\mu)} + \frac{\log n}{-\log c} \right)$. Hence, after the BFS from vertex s_i, the upper bound of any vertex v is smaller than $\mathrm{dist}(v, s_i) + (1 + 2\varepsilon)\left((1-x)\frac{\log n}{\log M_1(\mu)} + \frac{\log n}{-\log c} \right)$, which is smaller than D if $\mathrm{dist}(v, s_i) \leq D - (1+2\varepsilon)\left((1-x)\frac{\log n}{\log M_1(\mu)} + \frac{\log n}{-\log c} \right) =$ $(1+4\varepsilon)\left(\frac{x \log n}{\log M_1(\mu)} + \frac{\log n}{-\log c} \right)$ by Theorem 5.2.

Now, we want to compute the number of vertices that are at distance at least $(1+4\varepsilon)\left(\frac{x \log n}{\log M_1(\mu)} + \frac{\log n}{-\log c} \right)$ from each s_i. To estimate this quantity, we use the following lemma, which does not follow directly from the four properties (for a proof, see Appendix B.7).

LEMMA 11.1. *Assume that* $\beta > 2$, *and let* T *be the set of vertices with degree at least* n^x. *Then,* $\mathrm{dist}(s, T) := \min_{t \in T} \mathrm{dist}(s, t) \leq \boldsymbol{\tau}_s\left(n^{x(\beta-2)+\varepsilon}\right) + 1$ *w.h.p..*

In other words, this lemma says that, for each vertex v, $\min_{i=1,\dots,k} \mathrm{dist}(v, s_i) \leq \boldsymbol{\tau}_v\left(n^{x(\beta-2)+\varepsilon}\right)$: hence, after a BFS from each vertex s_i has been performed, the upper bound of v is at most D if $\boldsymbol{\tau}_v\left(n^{x(\beta-2)+\varepsilon}\right) \leq (1+4\varepsilon)\left(\frac{x \log n}{\log M_1(\mu)} + \frac{\log n}{-\log c} \right)$. We conclude that, after $n^{1+\varepsilon-x(\beta-1)}$ BFSes, only vertices verifying $\boldsymbol{\tau}_v\left(n^{x(\beta-2)+\varepsilon}\right) > (1 + 4\varepsilon)\left(\frac{x \log n}{\log M_1(\mu)} + \frac{\log n}{-\log c} \right)$ have upper bounds bigger than D.

By Property 2.1, the number of vertices that verify the latter inequality is at most $\mathcal{O}\left(nc^{(1-\mathcal{O}(\varepsilon))\left(\frac{x \log n}{\log M_1(\mu)} + \frac{\log n}{-\log c} - \frac{x(\beta-2)\log n}{\log M_1(\mu)} \right)} \right) =$ $n^{1 - \frac{-\log c}{\log n}\left(\frac{\log n}{-\log c} - \frac{x(\beta-3)\log n}{\log M_1(\mu)} \right) + \mathcal{O}(\varepsilon)} = n^{\frac{x(\beta-3)(-\log c)}{\log M_1(\mu)} + \mathcal{O}(\varepsilon)}$. Hence, by performing $\mathcal{O}\left(n^{\frac{x(\beta-3)(-\log c)}{\log M_1(\mu)} + \mathcal{O}(\varepsilon)} \right)$ more BFSes, the algorithm terminates.

We conclude that the total number of BFSes is at most $\max\left(n^{\frac{x(\beta-3)(-\log c)}{\log M_1(\mu)} + \mathcal{O}(\varepsilon)}, n^{1-x(\beta-1)} \right)$: if we substitute $x = \frac{1}{\beta - 1 + (\beta-3)\frac{-\log c}{\log M_1(\mu)}}$, we obtain $n^{\frac{1}{1 + \frac{\beta-1}{\beta-3}\frac{\log M_1(\mu)}{\log c}} + \mathcal{O}(\varepsilon)}$. Then, the running time is at most $n^{1 + \frac{1}{1 + \frac{\beta-1}{\beta-3}\frac{\log M_1(\mu)}{\log c}} + \mathcal{O}(\varepsilon)}$.

12 Conclusions and Open Problems.

In this paper, we have developed an axiomatic framework to evaluate heuristics and algorithms that compute metric properties of real-world graphs. The analyses performed in this framework motivate the empirical results obtained by previous works, they let us compare the different algorithms, and they provide more insight into their behavior. Furthermore, these results can be turned into average-case analyses in realistic models of random graphs.

This work leaves several open problems. First of all, it would be useful to improve the analysis with respect to the parameter ε, by computing the exact constants instead of simply writing $\mathcal{O}(\varepsilon)$.

Furthermore, in some cases, if we ignore ε, we have exactly computed the constants appearing in the exponent. However, in other cases, we just proved upper bounds: it would be interesting to understand if these bounds are tight. We conjecture that the bounds for the algorithms to compute the diameter are tight, or almost tight, but the bounds for other algorithms might be improved (for example, to perform the analysis of the AIY algorithm, we used estimates that are probably not optimal).

Finally, it could be interesting to generalize these results to other models: for instance, a possible generalization is to include all Inhomogeneous Random Graphs [10] (while in this paper we only considered only Rank-1 Inhomogeneous Random Graphs). We conjecture that also these graphs satisfy the four properties, because the known asymptotics for diameter and average distance are very similar to the asymptotics obtained in this paper.

Another possible generalization is to consider directed graphs: most of the algorithms we analyze in this paper can be generalized to the directed case, and the four properties can be generalized, as well. However, in the literature, there are no well-established models of power law random directed graphs: for this reason, it would be interesting to develop natural generalizations of the models considered, and prove that these generalizations satisfy the (generalized) properties.

References

[1] Amir Abboud, Fabrizio Grandoni, and Virginia V. Williams. Subcubic equivalences between graph centrality problems, APSP and diameter. In *Proceedings of the 26th ACM-SIAM Symposium on Discrete Algorithms*, pages 1681–1697, 2015.

[2] Amir Abboud, Virginia V. Williams, and Joshua Wang. Approximation and fixed parameter subquadratic algorithms for radius and diameter. In *Proceedings of the 26th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 377–391, may 2016.

[3] Ittai Abraham and Amos Fiat. Highway dimension, shortest paths, and provably efficient algorithms. In

Proceedings of the 21st ACM-SIAM Symposium on Discrete Algorithms (SODA), pages 782–793, 2010.

[4] Takuya Akiba, Yoichi Iwata, and Yuichi Yoshida. Fast exact shortest-path distance queries on large networks by pruned landmark labeling. In *Proceedings of the 2013 ACM-SIGMOD International Conference on Management of Data*, pages 349–360, 2013.

[5] Krishna B. Athreya and Peter Ney. *Branching processes*. Dover Books on Mathematics Series. Springer-Verlag Berlin Heidelberg New York, 1972.

[6] Albert L. Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[7] Elisabetta Bergamini, Michele Borassi, Pierluigi Crescenzi, Andrea Marino, and Henning Meyerhenke. Computing top-k closeness centrality faster in unweighted graphs. In *Proceedings of the Meeting on Algorithm Engineering and Experiments (ALENEX)*, pages 68–80, 2016.

[8] Paolo Boldi and Sebastiano Vigna. The webgraph framework I: compression techniques. In *Proceedings of the 13th international conference on World Wide Web (WWW)*, pages 595–602, 2004.

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

[10] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, 31(1):3–122, 2007.

[11] Michele Borassi, Alessandro Chessa, and Guido Caldarelli. Hyperbolicity Measures Democracy in Real-World Networks. *Physical Review E*, 92(3):032812, 2015.

[12] Michele Borassi, Pierluigi Crescenzi, and Michel Habib. Into the square - On the complexity of some quadratic-time solvable problems. *Electronic Notes in Computer Science*, pages 1–17, 2016.

[13] Michele Borassi, Pierluigi Crescenzi, Michel Habib, Walter Kosters, Andrea Marino, and Frank Takes. On the solvability of the Six Degrees of Kevin Bacon game - A faster graph diameter and radius computation method. In *Proceedings of the 7th International Conference on Fun with Algorithms (FUN)*, pages 57–68, 2014.

[14] Michele Borassi, Pierluigi Crescenzi, Michel Habib, Walter A. Kosters, and Frank W. Takes. Fast diameter and radius BFS-based computation in (weakly connected) real-world graphs - With an application to the Six Degrees of Separation games. *Theoretical Computer Science*, 586:59–80, 2014.

[15] Michele Borassi, Pierluigi Crescenzi, and Andrea Marino. Fast and simple computation of top-k closeness centralities. *arXiv preprint 1507.01490*, 2015.

[16] Michele Borassi and Emanuele Natale. KADABRA is an adaptive algorithm for betweenness via random approximation. In *Proceedings of the 24th European Symposium on Algorithms (in publication)*, 2016.

[17] Pawel Brach, Marek Cygan, Jakub Lacki, and Piotr Sankowski. Algorithmic complexity of power law networks. In *Proceedings of the 26th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1306–1325, 2016.

[18] Victor Chepoi, Fedor F. Dragan, Bertrand Estellon, Michel Habib, and Yann Vaxès. Diameters, centers, and approximating trees of δ-hyperbolic geodesic spaces and graphs. In *Proceedings of the 24th Annual Symposium on Computational Geometry*, pages 59–68, 2008.

[19] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, Alessandro Panconesi, and Prabhakar Raghavan. Models for the compressible web. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 331–340, 2009.

[20] Fan Chung and Linyuan Lu. *Complex graphs and networks*. American Mathematical Society, Boston, MA, USA, 2006.

[21] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms (3rd edition)*. MIT Press, 2009.

[22] Pierluigi Crescenzi, Roberto Grossi, Michel Habib, Leonardo Lanzi, and Andrea Marino. On computing the diameter of real-world undirected graphs. *Theoretical Computer Science*, 514:84–95, 2013.

[23] Daniel Delling, Andrew V. Goldberg, and Renato F. Werneck. Hub label compression. In *Proceedings of the 12th International Symposium on Experimental Algorithms (SEA)*, pages 18–29, 2013.

[24] Daniel Delling and Renato F. Werneck. Faster customization of road networks. In *Proceedings of the 12th International Symposium on Experimental Algorithms (SEA)*, pages 30–42, 2013.

[25] Daniel Fernholz and Vijaya Ramachandran. The diameter of sparse random graphs. *Random Structures and Algorithms*, 31(4):482–516, 2007.

[26] Alan Frieze and Colin McDiarmid. Algorithmic theory of random graphs. *Random Structures and Algorithms*, 10(1-2):5–42, 1997.

[27] Andrew V. Goldberg and Chris Harrelson. Computing the shortest path: A* search meets graph theory. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 156–165, 2005.

[28] Mikhael Gromov. Hyperbolic groups. *Essays in Group Theory*, 8:75–265, 1987.

[29] Rishi Gupta, Tim Roughgarden, and Comandur Seshadhri. Decompositions of Triangle-Dense Graphs. *SIAM Journal on Computing*, 45(2):197–215, 2016.

[30] Jure Leskovec and Rok Sosič. SNAP: A general purpose network analysis and graph mining library in C++. http://snap.stanford.edu/snap, jun 2014.

[31] Clémence Magnien, Matthieu Latapy, and Michel Habib. Fast computation of empirically tight bounds for the diameter of massive graphs. *Journal of Experimental Algorithmics (JEA)*, 13:1.10:1–1.10:9, 2009.

[32] Zoltan A. Mann and Anikò Szajkó. Average-case complexity of backtrack search for coloring sparse random graphs. *Journal of Computer and System*

Sciences, 79(8):1287–1301, 2013.

[33] Mark E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[34] Mark E. J. Newman. *Networks: an introduction*. OUP Oxford, 2010.

[35] Ilkka Norros and Hannu Reittu. On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38(1):59–75, 2006.

[36] Liam Roditty and Virginia V. Williams. Fast approximation algorithms for the diameter and radius of sparse graphs. In *Proceedings of the 45th annual ACM Symposium on Theory of Computing (STOC)*, pages 515–524, New York, New York, USA, 2013. ACM Press.

[37] Benjamin Rossman. *Average-case complexity of detecting cliques*. PhD thesis, Massachussets Institute of Technology, 2010.

[38] Peter Sanders and Dominik Schultes. Highway hierarchies hasten exact shortest path queries. In *Proceedings of the 13th European Symposium on Algorithms (ESA)*, pages 568–579, 2005.

[39] Christian Sommer, Elad Verbin, and Wei Yu. Distance oracles for sparse graphs. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 703–712, 2009.

[40] Christian L. Staudt, Aleksejs Sazonovs, and Henning Meyerhenke. Networkit: an interactive tool suite for high-performance network analysis. *arXiv preprint 1403.3005*, pages 1–25, 2014.

[41] William Stein and David Joyner. Sage: system for algebra and geometry experimentation. *SIGSAM Bulletin*, 39(2):61–64, 2005.

[42] Mikkel Thorup and Uri Zwick. Approximate distance oracles. *Journal of the ACM*, 52(1):1–24, 2005.

[43] Henri Van Den Esker, Remco Van Der Hofstad, Gerard Hooghiemstra, and Dmitri Znamenski. Distances in random graphs with infinite mean degrees. *Extremes*, 8(3):111–141, 2005.

[44] Remco van der Hofstad. Random graphs and complex networks. Vol. I, 2014.

[45] Remco van der Hofstad. Random graphs and complex networks. Vol. II, 2014.

[46] Aravindan Vijayaraghavan. *Beyond worst-case analysis in approximation algorithms*. PhD thesis, Princeton, 2012.

A The Model.

We assume the reader to be familiar with the basic notions of graph theory (see, for example, [21]), and we assume that all the graphs we consider are undirected and unweighted. Furthermore, we say that an event E holds asymptotically almost surely or a.a.s. if, when n tends to infinity, $\mathbb{P}(E) = 1 - o(1)$; it holds with high probability or w.h.p. if $\mathbb{P}(E) = 1 - o\left(n^{-k}\right)$ for each $k \in \mathbb{N}$.

In this paper, we consider different models of random graphs: the Configuration Model (CM, [9]), and Rank-1 Inhomogeneous Random Graph models (IRG, [45], Chapter 3), such as the Chung-Lu model [20], and the Norros-Reittu model [35]. All these models are defined by fixing in advance the number n of nodes, and n weights ρ_v, one for each vertex in the graph. Then, edges are created at random, trying to give ρ_v outgoing edges to each vertex v. We assume that the weights ρ_v are chosen according to a power law distribution λ, which is the degree distribution of many real-world graphs [33]: more specifically, we assume that, for each d, the number of vertices with weight bigger than d is $\Theta\left(\frac{n}{d^{\beta-1}}\right)$, for some constant β.[4]

After defining the weights, we have to define how we generate the edges:

- in the CM, we give ρ_v half-edges, or stubs to a vertex v; edges are created by pairing these $M = \sum_{v \in V} \rho_v$ stubs at random (we assume the number of stubs to be even, by adding a stub to a random vertex if necessary).

- in IRG, an edge between a vertex v and vertex w is created independently with probability $f(\frac{\rho_v \rho_w}{M})$, where $M = \sum_{v \in V} \rho_v$, and

    - in general, we assume the following:
        * f is derivable at least twice in 0;
        * f is increasing;
        * $f'(0) = 1$;
        * $f(x) = 1 - o(x^k)$ for each k, when x tends to infinity.
    - in the Chung-Lu model, $f(x) = \min(x, 1)$;
    - in the Norros-Reittu model, $f(x) = 1 - e^{-x}$.

REMARK A.1. *The first two assumptions in IRG are needed to exclude pathological cases. The third assumption is just needed to simplify notation, but it can be easily lifted by modifying the weights ρ_v: for instance, if $f'(0) = c$, we may multiply all ρ_vs by $\sqrt{c}$, and redefine $f_1(x) = f\left(\frac{x}{c}\right)$, obtaining the same graph with a function verifying $f'_1(0) = 1$. The fourth assumption is less natural, and there are models where it is not satisfied, like the Generalized Random Graph model ([44], Chapter 6). However, if the average degree is finite (that is, $\beta > 2$), the proofs do not need this assumption (in this work, we have chosen to use this assumption in order to simplify the statements).*

In order to prove our results, we further need some technical assumptions, to avoid pathological cases. In particular, we exclude from our analysis the values of β corresponding to the phase transitions: $\beta = 2$, and $\beta = 3$. Furthermore, in the regime $1 < \beta < 2$, we have other phase transitions related to the diameter of the graph, which is $\left\lfloor 3 + \frac{\beta-1}{2-\beta} \right\rfloor$: we assume that $\frac{\beta-1}{2-\beta}$ is not an integer, and, with abuse of notation, we write $\left\lfloor \frac{\beta-1}{2-\beta} - \varepsilon \right\rfloor = \left\lfloor \frac{\beta-1}{2-\beta} \right\rfloor = \left\lceil \frac{\beta-1}{2-\beta} \right\rceil - 1 = \left\lceil \frac{\beta-1}{2-\beta} + \varepsilon \right\rceil - 1$.

Finally, we need a last assumption on the degree distribution λ: all our metric quantities make sense only if the graph is connected. Hence, we need to assume that λ does not contain "too many vertices" of small degree, so that a.a.s. there is a unique connected component of size $\Theta(n)$, named giant component. All our results hold in the giant component of the graph considered.

In the remainder of this section, we define precisely this assumption, and we further define some more constants that appear in the main theorems. A reader who is not interested in these technicalities might just skip this part, assuming that the graph is connected, and that the main theorems hold (our probabilistic analyses do not depend on the definition of these constants).

---

[4]In some cases, a stronger definition of power law is used, that is, it is assumed that there are $\Theta\left(\frac{n}{d^{\beta}}\right)$ vertices with degree d, for each d. However, our proofs still work with the weaker definition.

The first definition is the *residual* distribution μ [25, 45, 44]: intuitively, if we choose a random node v, and we choose a random neighbor w of v, the degree of w is μ-distributed (see Theorem B.2, in the case $\ell = 1$). This distribution is defined as follows.

DEFINITION A.1. *Given a distribution λ, its first moment $M_1(\lambda)$ is the expected value of a λ-distributed random variable. The residual distribution μ of the distribution λ is:*

- *in the CM, $\mu(i) = \frac{(i+1)\lambda(i+1)}{M_1(\lambda)}$;*

- *in IRG, let $\mu'(i) = \frac{i\lambda(i)}{M_1(\lambda)}$: $\mu(i)$ is a Poisson distribution with random parameter μ'.*

In Appendix B.3, we show that the number of vertices at distance ℓ from a given vertex v is very close to a μ-distributed branching process $\boldsymbol{Z}^\ell$ (for more background on branching processes, we refer to [5]). If $M_1(\mu) < 1$, this branching process dies a.a.s.: in terms of graphs, it means that the biggest component has size $\mathcal{O}(\log n)$, and there is no giant component. Conversely, if $M_1(\mu)$ is bigger than 1, then the branching process has an infinite number of descendants with positive probability p: in terms of graphs, it means that there is a connected component of size close to pn (see [44] for a proof). Hence, we assume $M_1(\mu) > 1$ and we ignore all vertices that are not in the giant component.

Finally, given a μ-distributed branching process, we may consider only the branches that have an infinite number of descendant (see [5], I.D.12): we obtain another branching process with offspring distribution η depending on μ. In particular, our results depend on $\eta(1)$, that is, the probability that an η-distributed random variable has value 1. For more information on the value $\eta(1)$, we refer to [25]. In the following, we also assume that $\eta(1) > 0$: this is true if and only if $\mu(0) \neq 0$ or $\mu(1) \neq 0$. In IRG, this is automatically implied by the definition of μ, while in the CM this is an additional technical assumption.

B Validity of the Properties in Random Graphs: Formal Proof.

In this section, we prove Theorem 3.1, that states that the four properties are a.a.s. verified if a graph is generated with the Configuration Model, or with Rank-1 Inhomogeneous Random Graph models. We follow the sketch in Section 3. In Appendix B.1 we state some basic lemmas that are used throughout this section, while in Appendix B.2 we analyze the size of $\boldsymbol{\gamma}^\ell(s)$ when $\boldsymbol{\gamma}^\ell(s)$ is "big" (at least n^ε). Then, Appendix B.3 completes Appendix B.2 by analyzing the size of $\boldsymbol{\gamma}^\ell(s)$ when $\boldsymbol{\gamma}^\ell(s)$ is small, using branching process approximation. Then, in Appendix B.4 we analyze separately the case $1 < \beta < 2$, which has a different behavior. Appendix B.5 develop tools to convert probabilistic results into results on the number of vertices satisfying a certain property. Finally, Appendix B.6 proves Theorem 3.1, relying on the results of all previous sections, and Appendix B.7 proves other results that were used in some analyses (these analyses are marked with $(*)$ in Table 1).

B.1 Probabilistic Preliminaries.
In this section, we state some basic probabilistic theorems that are used in the proof of our main results. For a more thorough discussion and for their proof, we refer to [20].

LEMMA B.1. (MULTIPLICATIVE FORM OF CHERNOFF BOUND) *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ be independent Bernoulli random variables, and let $\boldsymbol{S} = \sum_{i=1}^{k} \boldsymbol{X}_i$. Then,*

$$\mathbb{P}\left(\boldsymbol{S} < (1-\varepsilon)\mathbb{E}[\boldsymbol{S}]\right) \leq \left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{\mathbb{E}[\boldsymbol{S}]} \qquad \mathbb{P}\left(\boldsymbol{S} > (1+\varepsilon)\mathbb{E}[\boldsymbol{S}]\right) \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mathbb{E}[\boldsymbol{S}]}$$

LEMMA B.2. (HOEFFDING'S INEQUALITY) *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ be independent random variables such that $a_i < \boldsymbol{X}_i < b_i$ almost surely, and let $\boldsymbol{S} = \sum_{i=1}^{k} \boldsymbol{X}_i$. Then,*

$$\mathbb{P}\left(|\boldsymbol{S} - \mathbb{E}[\boldsymbol{S}]| > \lambda\right) \leq 2e^{-\frac{2\lambda^2}{\sum_{i=1}^{k} |b_i - a_i|^2}}$$

The next lemmas deal with supermartingales and submartingales, which are defined as follows.

DEFINITION B.1. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ be a sequence of random variables, let $\mathcal{F}_1, \ldots, \mathcal{F}_k$ be a sequence of σ-fields such that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_i$ are $\mathcal{F}_i$-measurable. The sequence is a martingale if the conditional expectation $\mathbb{E}[\boldsymbol{X}_{i+1}|\mathcal{F}_i]$ is equal to $\boldsymbol{X}_i$, it is a supermartingale if $\mathbb{E}[\boldsymbol{X}_{i+1}|\mathcal{F}_i] \leq \boldsymbol{X}_i$, and it is a submartingale if $\mathbb{E}[\boldsymbol{X}_{i+1}|\mathcal{F}_i] \geq \boldsymbol{X}_i$.*

The terms "submartingale" and "supermartingale" have not been used consistently in the literature, since in some works a supermartingale verifies $\mathbb{E}[\boldsymbol{X}_{i+1}|\mathcal{F}_i] \geq \boldsymbol{X}_i$, and a submartingale verifies $\mathbb{E}[\boldsymbol{X}_{i+1}|\mathcal{F}_i] \leq \boldsymbol{X}_i$ [20]. In this work, we use the most common definition.

LEMMA B.3. (AZUMA INEQUALITY FOR SUPERMARTINGALES) *Let $\boldsymbol{X}_k$ be a supermartingale, and let us assume that $|\boldsymbol{X}_k - \boldsymbol{X}_{k+1}| < M$ almost surely. Then, $\mathbb{P}(\boldsymbol{X}_n - \boldsymbol{X}_0 \geq t) \leq e^{-\frac{t^2}{2nM^2}}$.*

LEMMA B.4. (AZUMA INEQUALITY FOR SUBMARTINGALES) *Let $\boldsymbol{X}_k$ be a submartingale, and let us assume that $|\boldsymbol{X}_k - \boldsymbol{X}_{k+1}| < M$ almost surely. Then, $\mathbb{P}(\boldsymbol{X}_n - \boldsymbol{X}_0 \leq -t) \leq e^{-\frac{t^2}{2nM^2}}$.*

LEMMA B.5. (STRENGTHENED VERSION OF AZUMA INEQUALITY) *Let $\boldsymbol{X}_k$ be a supermartingale associated with a filter $\mathcal{F}$, and assume that $\mathrm{Var}(\boldsymbol{X}_k|\mathcal{F}_{k-1}) \leq \sigma^2$, and $\boldsymbol{X}_k - \mathbb{E}(\boldsymbol{X}_k|\mathcal{F}_{k-1}) \leq M$. Then,*

$$\mathbb{P}\left(\boldsymbol{X}_k \geq \boldsymbol{X}_0 + \lambda\right) \leq e^{\frac{-\lambda^2}{2k\sigma^2 + \frac{M\lambda}{3}}}.$$

Finally, we need a technical lemma on the sum of power law random variables.

LEMMA B.6. *Let $\boldsymbol{X} = \sum_{i=1}^{k} \boldsymbol{X}_i$, where k tends to infinity and the $\boldsymbol{X}_i$s are power law random variables with exponent $1 < \beta < 2$. Then, for each $c > 0$, $\mathbb{P}\left(\boldsymbol{X} > k^{\frac{1+c}{\beta-1}}\right) = \mathcal{O}(k^{-c})$.*

Proof. For each i, $\mathbb{P}\left(\boldsymbol{X}_i > k^{\frac{1+c}{\beta-1}}\right) = \mathcal{O}\left(\left(k^{-\frac{1+c}{\beta-1}}\right)^{\beta-1}\right) = \mathcal{O}\left(\frac{1}{k^{1+c}}\right)$, and consequently the probability that there exists i such that $X_i > k^{\frac{1+c}{\beta-1}}$ is $\mathcal{O}\left(k^{-c}\right)$.

Conditioned on $\boldsymbol{X}_i \leq k^{\frac{1+c}{\beta-1}}$ for each i,

$$\mathbb{E}\left[\boldsymbol{X}\right] = \mathbb{E}\left[\sum_{\ell=1}^{\infty} |\{i : \boldsymbol{X}_i > \ell\}|\right] = \mathbb{E}\left[\sum_{\ell=1}^{k^{\frac{1+c}{\beta-1}}} |\{i : \boldsymbol{X}_i > \ell\}|\right] = \sum_{\ell=1}^{k^{\frac{1+c}{\beta-1}}} \mathbb{E}\left[|\{i : \boldsymbol{X}_i > \ell\}|\right]$$

$$= \sum_{\ell=1}^{k^{\frac{1+c}{\beta-1}}} \mathcal{O}(k\ell^{-\beta+1}) = \mathcal{O}\left(k^{1+\frac{(1+c)(2-\beta)}{\beta-1}}\right) = k^{\frac{1+c(2-\beta)}{\beta-1}}.$$

We conclude that $\mathbb{P}\left(\boldsymbol{X} > k^{\frac{1+c}{\beta-1}}\right) = \mathbb{P}\left(\boldsymbol{X} > k^{\frac{1+c}{\beta-1}} \Big| \exists i, \boldsymbol{X}_i > k^{\frac{1+c}{\beta-1}}\right) \mathbb{P}\left(\exists i, \boldsymbol{X}_i > k^{\frac{1+c}{\beta-1}}\right) + \mathbb{P}\left(\boldsymbol{X} > k^{\frac{1+c}{\beta-1}} \Big| \forall i, \boldsymbol{X}_i < k^{\frac{1+c}{\beta-1}}\right) \mathbb{P}\left(\forall i, \boldsymbol{X}_i < k^{\frac{1+c}{\beta-1}}\right) \leq \mathcal{O}\left(k^{-c}\right) + \frac{1}{k^{\frac{1+c}{\beta-1}}} \mathbb{E}\left(\boldsymbol{X} > k^{\frac{1+c}{\beta-1}} \Big| \forall i, \boldsymbol{X}_i < k^{\frac{1+c}{\beta-1}}\right) = \mathcal{O}\left(k^{-c} + k^{\frac{1+c(2-\beta)}{\beta-1} - \frac{1+c}{\beta-1}}\right) = \mathcal{O}\left(k^{-c} + k^{-c}\right) = \mathcal{O}\left(k^{-c}\right)$ by Markov inequality.

B.2 Big Neighborhoods. First of all, let us define precisely the typical time needed by a vertex of degree d to reach size n^x. In Section 2, we defined $T\left(d \rightarrow n^x\right)$ as the smallest ℓ such that $\boldsymbol{\gamma}^\ell(s) > n^x$, and then we stated which are the typical values of $T\left(d \rightarrow n^x\right)$ in different regimes. In this section, we do the converse: we define $F\left(d \rightarrow S\right)$ as a function of the degree distributions, and we show that there is a high chance that $\boldsymbol{\gamma}^{(1-\varepsilon)F(d\rightarrow S)}(s) < S < \boldsymbol{\gamma}^{(1+\varepsilon)F(d\rightarrow S)}(s)$, if s is a vertex of degree d in the giant component.

DEFINITION B.2. *In the following for any $0 < d < S$, we denote by*

$$F\left(d \rightarrow S\right) = \begin{cases} \log_{M_1(\mu)}\left(\frac{S}{d}\right) & \text{if } M_1(\mu) \text{ is finite.} \\ \log_{\frac{1}{\beta-2}}\left(\frac{\log S}{\log d}\right) & \text{if } \lambda \text{ is a power law distribution with } 2 < \beta < 3. \end{cases}$$

Following the intuitive proof, in this section we fix x, y bigger than ε, and we bound $\boldsymbol{\tau}_s\left(n^y\right) - \boldsymbol{\tau}_s\left(n^x\right)$. The main technique used is to prove that, w.h.p., each neighbor which is big enough satisfies some constraints, and these constraints imply bounds on $\boldsymbol{\tau}_s\left(n^y\right) - \boldsymbol{\tau}_s\left(n^x\right)$. More formally, we prove the following theorem.

THEOREM B.1. *For each $0 < x < y < 1$, $\boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \geq (1-\varepsilon)F(n^x \to n^y)$ a.a.s., and $\boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \leq (1+\varepsilon)F(n^x \to n^y)$ w.h.p..*

The proof of this theorem is different for the CM and for IRG. In particular, the main tool used to prove this theorem is an estimate on $\boldsymbol{\gamma}^{\ell+1}(s)$ knowing $\boldsymbol{\gamma}^\ell(s)$: intuitively, in the CM, for each vertex in $\boldsymbol{\Gamma}^\ell(s)$ we count how many neighbors it has in $\boldsymbol{\Gamma}^{\ell+1}(s)$, while in IRG we count how many vertices outside $\boldsymbol{N}^\ell(s)$ have a neighbor in $\boldsymbol{\Gamma}^\ell(s)$.

B.2.1 Configuration Model. Let us assume that we know the structure of $\boldsymbol{N}^\ell(s)$ (that is, we consider all possible events $\boldsymbol{E}_i$ that describe the structure of $\boldsymbol{N}^\ell(s)$, and we prove bounds conditioned on $\boldsymbol{E}_i$; finally, though a union bound, we remove the conditioning). Let us define a random variable $\boldsymbol{\Delta}^\ell(s)$, which measures "how big a neighbor is".

DEFINITION B.3. *Given a graph $G = (V, E)$ generated through the CM, we denote by $\boldsymbol{\Delta}^\ell(s)$ the set of stubs of vertices in $\boldsymbol{\Gamma}^\ell(s)$, not paired with stubs of vertices in $\boldsymbol{\Gamma}^{\ell-1}(s)$. We denote $\boldsymbol{\delta}^\ell(s) = |\boldsymbol{\Delta}^\ell(s)|$.*

In order to make this analysis work, we need to assume that $\rho_{\boldsymbol{N}^\ell(s)} < n^{1-\varepsilon}$ and $\boldsymbol{\Delta}^\ell(s) > n^\varepsilon$.

Let us consider the following process: we sort all the stubs in $\boldsymbol{\Delta}^\ell(s)$, obtaining $a_1, \ldots, a_{\boldsymbol{\delta}^\ell(s)}$, and, starting from a_1, we choose uniformly at random the "companion" of a_i among all free stubs (if a_i is already paired with a stub a_j for some $j < i$, we do not do anything). The companion of a_i can be one of the following:

1. a stub of a vertex v that already belongs to $\boldsymbol{\Gamma}^{\ell+1}(s)$ (because another stub of v was already paired with a stub in $\boldsymbol{\gamma}^\ell(s)$);

2. a stub of a "new" vertex;

3. another unpaired stub in $\boldsymbol{\Delta}^\ell(s)$.

Let us prove that the number of stubs in $\boldsymbol{\Delta}^\ell(s)$ that are paired with other stubs in $\boldsymbol{\Delta}^\ell(s)$ is small (Item 3): at each step, the probability that we choose one of these stubs is the ratio between the number of unpaired stubs in $\boldsymbol{\Delta}^\ell(s)$ with respect to the total number of unpaired stubs. Since $\rho_{\boldsymbol{N}^\ell(s)} < n^{1-\varepsilon}$, the number of unpaired stubs in $\boldsymbol{\Delta}^\ell(s)$ is at most $n^{1-\varepsilon}$, and the total number of unpaired stubs is at least $M - n^{1-\varepsilon} = M(1 - o(1))$. Hence, the probability that we choose one of these stubs is at most $\frac{n^{1-\varepsilon}}{M} < n^{-\varepsilon}$. Let $\boldsymbol{X}_a$ be a Bernoulli random variable which is 1 if we pair a with another stub inside $\boldsymbol{\Delta}^\ell(s)$, 0 if a is already paired when we process it, or if it is paired outside $\boldsymbol{\Delta}^\ell(s)$ (observe that the number of vertices paired inside $\boldsymbol{\Delta}^\ell(s)$ is $2\sum_{a \in \boldsymbol{\Delta}^\ell(s)} \boldsymbol{X}_a$). We want to apply Azuma's inequality: first, we sort the stubs in $\boldsymbol{\Delta}^\ell(s)$, obtaining $a_1, \ldots, a_{\boldsymbol{\delta}^\ell(s)}$. By the previous argument, $\boldsymbol{S}_k = \sum_{i=1}^k \boldsymbol{X}_{a_i} - kn^{-\varepsilon}$ is a supermartingale, and hence $\mathbb{P}(\boldsymbol{X}_k > \varepsilon k) \leq e^{-\frac{\varepsilon^2 k^2}{2k}}$: for $k = \boldsymbol{\Delta}^\ell(s)$, this probability is at most $e^{-\varepsilon^3 n^\varepsilon}$. In conclusion, w.h.p., at most $2\varepsilon \boldsymbol{\Delta}^\ell(s)$ stubs in $\boldsymbol{\Delta}^\ell(s)$ are paired to other stubs in $\boldsymbol{\Delta}^\ell(s)$.

Let us consider a stub a paired outside $\boldsymbol{\Delta}^\ell(s)$ with a random stub a': if the number of stubs that are already in $\boldsymbol{\Delta}^{\ell+1}(s)$ is at most $n^{1-\varepsilon^2}$, then the probability that a' is already in $\boldsymbol{\Delta}^{\ell+1}(s)$ is at most $n^{-\varepsilon^2}$. In order to solve the case where $\rho_{\boldsymbol{\Gamma}^{\ell+1}(s)} > n^{1-\varepsilon^2}$, let us assume that $\boldsymbol{\Delta}^\ell(s) < n^{1-\varepsilon}$: in this case, since the number of elements in $\boldsymbol{\Delta}^{\ell+1}(s)$ decreases at most by 1 at each step, $\boldsymbol{\Delta}^{\ell+1}(s) \geq n^{1-\varepsilon^2} - n^{1-\varepsilon} \geq n^{1-\varepsilon}$.

Hence, the case that "almost always" occurs is that the new stub a' belongs to a "new" vertex. Relying on this, we can lower bound $\boldsymbol{\gamma}^{\ell+1}(s)$: by definition, $\boldsymbol{\gamma}^{\ell+1}(s) \leq \boldsymbol{\delta}^\ell(s)$, and we want to prove that $\boldsymbol{\gamma}^{\ell+1}(s) \geq (1-\varepsilon)\boldsymbol{\delta}^\ell(s)$. Since the number of stubs in $\boldsymbol{\Delta}^\ell(s)$ paired with other stubs in $\boldsymbol{\Delta}^\ell(s)$ is negligible w.h.p., we can write $\boldsymbol{\gamma}^{\ell+1}(s) = \sum_{i=1}^{(1-\varepsilon)\boldsymbol{\delta}^\ell(s)} \boldsymbol{X}_i$, where $\boldsymbol{X}_i = 1$ with probability at least $1 - n^{-\varepsilon^2}$, 0 otherwise (note that the $\boldsymbol{X}_i$s are not independent, but if $\boldsymbol{\Delta}^{\ell+1}(s) < n^{1-\varepsilon}$, then $\mathbb{P}(\boldsymbol{X}_i = 1) \geq 1 - n^{-\varepsilon^2}$, as before). We want to apply Azuma's inequality: $\boldsymbol{S}_k = \sum_{i=1}^k \boldsymbol{X}_i - k(1 - n^{-\varepsilon^2})$ is a submartingale, and hence $\mathbb{P}(\boldsymbol{S}_k < -\varepsilon k) \leq e^{-\frac{\varepsilon^2 k^2}{2k}}$: for $k = \boldsymbol{\delta}^\ell(s)$, this probability is at most $e^{-\varepsilon^3 n^\varepsilon}$. Hence, w.h.p., $\boldsymbol{X}_i \geq k(1 - n^{-\varepsilon^2}) - \varepsilon k \geq (1 - 2\varepsilon)k$, and for $k = \boldsymbol{\delta}^\ell(s)$ we have proved the following lemma.

LEMMA B.7. *Given a random graph $G = (V, E)$ generated through the CM and a vertex $s \in V$, if $\boldsymbol{\delta}^\ell(s) > n^\varepsilon$ and $\rho_{\boldsymbol{N}^\ell(s)} < n^{1-\varepsilon}$, then $(1 - 2\varepsilon)\boldsymbol{\delta}^\ell(s) \leq \boldsymbol{\gamma}^{\ell+1}(s) \leq \boldsymbol{\delta}^\ell(s)$ w.h.p..*

COROLLARY B.1. *For each vertex s, let $\boldsymbol{\tau}'_s(S)$ be the smallest integer such that $\boldsymbol{\delta}^\ell(s) > S$. Then, for each $0 < x < 1$, $\boldsymbol{\tau}'_s(n^x) + 1 \leq \boldsymbol{\tau}_s(n^x) \leq \boldsymbol{\tau}'_s\left(\frac{n^x}{1-\varepsilon}\right) + 1$ w.h.p..*

Proof. For the first inequality, if $\ell = \boldsymbol{\tau}_s(n^x)$, $\boldsymbol{\delta}^{\ell-1}(x) \geq \boldsymbol{\gamma}^\ell(s) \geq n^x$. For the second inequality, for each $i < \ell - 1$, $n^x > \boldsymbol{\gamma}^{i+1}(s) \geq (1-\varepsilon)\boldsymbol{\delta}^i(s)$ by the previous lemma. Hence, $\boldsymbol{\tau}'_s\left(\frac{n^x}{1-\varepsilon}\right)$ cannot be smaller than $\ell - 1$.

Hence, in order to understand $\boldsymbol{\tau}_s(n^x) - \boldsymbol{\tau}_s(n^y)$, we may as well understand $\boldsymbol{\tau}'_s(n^x) - \boldsymbol{\tau}'_s(n^y)$, and we do it by estimating $\boldsymbol{\delta}^{\ell+1}(s)$ from $\boldsymbol{\delta}^\ell(s)$. As before, $\boldsymbol{\delta}^{\ell+1}(s) = \sum_{a\in\boldsymbol{\Delta}^\ell(s)} \boldsymbol{Y}_a$, where $\boldsymbol{Y}_a$ is 0 if the stub $\boldsymbol{a}$ paired with a is in $\boldsymbol{\Delta}^\ell(s)$, -1 if $\boldsymbol{a}$ is in $\boldsymbol{\Gamma}^{\ell+1}(s)$, otherwise it the number of stubs of the vertex of $\boldsymbol{a}$, minus one (because $\boldsymbol{a}$ is not in $\boldsymbol{\Delta}^{\ell+1}(s)$). By definition, the distribution of $\boldsymbol{Y}_a$ is very close to μ (more specifically, $\sum_{k=0}^\infty |\mu(k) - \mathbb{P}(\boldsymbol{Y}_a = k)| < \frac{1}{n^\varepsilon}$).

It remains to estimate this sum: we need to do it differently for upper and lower bounds, and for different regimes of β.

Lower bound, $2 < \beta < 3$. The probability that at least one of the $\boldsymbol{Y}_a$ is at least $\boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}$ is close to $\mathbb{P}\left(\mu > \boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}\right)$, because, w.h.p., no visited vertex can have weight bigger than $\boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}$ (otherwise, there would be a $\ell' < \ell$ such that $\boldsymbol{\delta}^{\ell'}(s) \geq \boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}$). Hence, the probability that one of the $\boldsymbol{Y}_a$s is at least $\boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}$ is $\Theta\left(\frac{1}{\boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}(\beta-2)}}\right) = \Theta\left(\boldsymbol{\delta}^\ell(s)^{-1+\varepsilon}\right)$. We want to apply Azuma's inequality to prove that at least one of $\boldsymbol{Y}_a$s is bigger than $\boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}$. Let us number the stubs in $\boldsymbol{\Delta}^\ell(s)$, obtaining $a_1, \ldots, a_{\boldsymbol{\delta}^\ell(s)}$, and let $\boldsymbol{S}_k = \sum_{i=0}^k \boldsymbol{Y}'_{a_i} - ck\boldsymbol{\delta}^\ell(s)^{-1+\varepsilon}$, where $\boldsymbol{Y}'_{a_i} = 1$ if $\boldsymbol{Y}_{a_i} > \boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}$, 0 otherwise, and c is a small enough constant, so that $\boldsymbol{S}_k$ is a submartingale. Furthermore, $\text{Var}(\boldsymbol{Y}'_{a_i}) \leq \mathbb{E}[(\boldsymbol{Y}'_{a_i})^2] = \mathbb{E}[\boldsymbol{Y}'_{a_i}]) = \mathcal{O}\left(\boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}\right)$. Then, by the strengthened version of Azuma's inequality (Lemma B.5), if $k = \boldsymbol{\delta}^\ell(s)$, $\mathbb{P}\left(\boldsymbol{S}_k \leq \frac{c}{2}k\boldsymbol{\delta}^\ell(s)^{-1+\varepsilon}\right) \leq e^{-\Omega\left(\frac{k^2\boldsymbol{\delta}^\ell(s)^2}{2k\boldsymbol{\delta}^\ell(s)+k\boldsymbol{\delta}^\ell(s)}\right)} \leq e^{-\Omega(\boldsymbol{\delta}^\ell(s)^\varepsilon)} \leq e^{-n^{\varepsilon^3}}$. Hence, w.h.p., $\boldsymbol{S}_{\boldsymbol{\delta}^\ell(s)} \geq \frac{c}{2}k\boldsymbol{\delta}^\ell(s)^{-1+\varepsilon} > 0$, and consequently there is i such that $\boldsymbol{Y}'_{a_i} \neq 0$. This means that, for each i, $\boldsymbol{\delta}^{\ell+i}(s) \geq \boldsymbol{\delta}^\ell(s)^{\left(\frac{1-\varepsilon}{\beta-2}\right)^i}$.

Upper bound, $2 < \beta < 3$. By Lemma B.6, since μ is a power law with exponent $\beta - 1$, and $2 < \beta < 3$, the probability that $\sum_{a\in\boldsymbol{\Delta}^\ell(s)} \boldsymbol{Y}_a$ is bigger than $k^{\frac{1+\varepsilon}{\beta-2}}$ is at most $\mathcal{O}(k^{-\varepsilon}) = \mathcal{O}\left(n^{-\varepsilon^2}\right)$. Consequently, by a union bound, $\boldsymbol{\delta}^{\ell+i}(s) \leq \boldsymbol{\delta}^\ell(s)^{\left(\frac{1+\varepsilon}{\beta-2}\right)^i}$ for each $i < n^{\varepsilon^3}$, with probability $1 - o(1)$.

Lower bound, $\beta > 3$. We cannot apply directly Azuma's inequality to say that $\boldsymbol{\delta}^{\ell+1}(s)$ is close to $\mathbb{E}[\boldsymbol{\delta}^{\ell+1}(s)] = (1 + o(1))M_1(\mu)\boldsymbol{\delta}^\ell(s)$, because $\boldsymbol{Y}_a$ can assume very large values. However, we can "cut the distribution", by defining $\boldsymbol{Y}'_a = \boldsymbol{Y}_a$ if $\boldsymbol{Y}_a < N$, 0 otherwise. If N is big enough, $\mathbb{E}[\boldsymbol{Y}'_a] > M_1(\mu) - \varepsilon$. By a straightforward application of Azuma's inequality (Lemma B.4), $\boldsymbol{\delta}^{\ell+1}(s) \geq \sum_{a\in\boldsymbol{\Delta}^\ell(s)} \boldsymbol{Y}_a \geq (1-\varepsilon)(M_1(\mu)-\varepsilon)\boldsymbol{\delta}^\ell(s)$ w.h.p.. Consequently, $\boldsymbol{\delta}^{\ell+i}(s) \geq (M_1(\mu) - \mathcal{O}(\varepsilon))^i\boldsymbol{\delta}^\ell(s)$, w.h.p..

Upper bound, $\beta > 3$. The expected value of $\boldsymbol{\delta}^{\ell+i}(s)$ is at most $(M_1(\mu) + \varepsilon)^i\boldsymbol{\delta}^\ell(s)$. A straightforward application of Markov inequality lets us conclude that $\mathbb{P}\left(\boldsymbol{\delta}^{\ell+i}(s) > (M_1(\mu) + \varepsilon)^i\boldsymbol{\delta}^\ell(s)n^\varepsilon\right) \leq n^{-\varepsilon}$.

Proof. [Proof of Theorem B.1, CM] By Corollary B.1, $\boldsymbol{\tau}'_s(n^x) + 1 \leq \boldsymbol{\tau}_s(n^x) \leq \boldsymbol{\tau}'_s((1+\varepsilon)n^x) + 1$. Hence, $\boldsymbol{\tau}'_s(n^y) - \boldsymbol{\tau}'_s((1+\varepsilon)n^x) \leq \boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \leq \boldsymbol{\tau}'_s(n^y(1+\varepsilon)) - \boldsymbol{\tau}'_s(n^x)$.

If we apply the lower bounds with $i = F\left(Z^0 \to S\right)(1+\varepsilon')$, $\ell = \boldsymbol{\tau}'_s(n^x)$, we obtain the following.

- If $2 < \beta < 3$, either $\boldsymbol{n}^{\ell+j}(s) > n^{1-\varepsilon}$ for some $j < i$, or, w.h.p., $\boldsymbol{\delta}^{\ell+i}(s) \geq \boldsymbol{\delta}^\ell(s)^{\left(\frac{1-\varepsilon}{\beta-2}\right)^i} \geq n^{x\left(\frac{1-\varepsilon}{\beta-2}\right)^{(1+\varepsilon')\log\frac{1}{\beta-2}\frac{y}{x}}} = $

  $n^{xe^{\log\left(\frac{y}{x}\right)(1+\varepsilon')\frac{\log\frac{1-\varepsilon}{\beta-2}}{\log\frac{1}{\beta-2}}}} \geq (1+\varepsilon)n^y$ if ε is small enough with respect to ε'. In both cases, $\boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \leq \boldsymbol{\tau}'_s((1+\varepsilon)n^y) - \boldsymbol{\tau}'_x(n^x) \leq F\left(Z^0 \to S\right)(1+\varepsilon')$. With a very similar computation, one can conclude that $\boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \geq F\left(Z^0 \to S\right)(1-\varepsilon')$ a.a.s. (the only difference is how to handle the case where $\boldsymbol{n}^{\ell+j}(s) > n^{1-\varepsilon}$: to this purpose, it is enough to observe that if $n^x < n^y < n^{1-\varepsilon}$, for the whole process $\boldsymbol{n}^{\ell+j}(s) < n^y < n^{1-\varepsilon}$).

- If $\beta > 3$, as before, either $\boldsymbol{n}^{\ell+j}(s) > n^{1-\varepsilon}$ for some $j < i$, or, w.h.p., $\boldsymbol{\delta}^{\ell+i}(s) \geq \boldsymbol{\delta}^{\ell}(s)(M_1(\mu) - \varepsilon)^i \geq \boldsymbol{\delta}^{\ell}(s)(M_1(\mu) - \varepsilon)^{(1+\varepsilon)\log_{M_1(\mu)} n^{y-x}} = n^x e^{\log(n^{y-x})(1+\varepsilon')\frac{M_1(\mu)-\varepsilon}{M_1(\mu)}} \geq n^y(1 + \varepsilon)$ if ε is small enough with respect to ε'. We conclude that $\boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \leq \boldsymbol{\tau}_s'((1+\varepsilon)n^y) - \boldsymbol{\tau}_x'(n^x) \leq F\left(Z^0 \to S\right)(1+\varepsilon')$ w.h.p.. A similar computation yields $\boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \geq \boldsymbol{\tau}_s'(n^y) - \boldsymbol{\tau}_x'((1+\varepsilon)n^x) \leq F\left(Z^0 \to S\right)(1-\varepsilon')$ a.a.s..

To conclude this section, we prove a stronger upper bound in the case $\beta > 3$, which is used in two of our probabilistic analyses.

LEMMA B.8. *Assume that $\boldsymbol{\delta}^{\ell}(s) > d_{\max} n^\varepsilon$, where $d_{\max}$ is the maximum degree in the graph, and that $M_1(\mu)$ is finite. Then, w.h.p., $\boldsymbol{\delta}^{\ell+1}(s) \leq \boldsymbol{\delta}^{\ell}(s)(M_1(\mu) + \varepsilon)$.*

Proof. We want to apply Azuma's inequality as in the lower bound. More precisely, $\boldsymbol{\delta}^{\ell+1}(s) \leq \sum_{i=1}^{\boldsymbol{\delta}^{\ell}(s)} \boldsymbol{Y}_{a_i}$, where $\mathbb{E}[\boldsymbol{Y}_{a_i}]$ is at most $M_1(\mu) + \varepsilon$, conditioned on the values of $\boldsymbol{Y}_{a_j}$ for each $j < i$. Hence, $\sum_{i=1}^{k} \boldsymbol{Y}_{a_i} - k(M_1(\mu) + \varepsilon)$ is a supermartingale, and $\boldsymbol{Y}_{a_i} < n^{\frac{1}{\beta-1}} < n^{\frac{1}{2}-\varepsilon}$ for ε small enough. By Azuma's inequality (Lemma B.3), $\mathbb{P}\left(\boldsymbol{Y}_k \leq \varepsilon k\right) \leq e^{-\frac{\varepsilon^2 k^2}{2kd_{\max}}} \leq e^{-\varepsilon^3 n^\varepsilon}$ for $k = \boldsymbol{\delta}^{\ell}(s) > d_{\max} n^\varepsilon$. We proved that, w.h.p., $\boldsymbol{\delta}^{\ell+1}(s) - \boldsymbol{\delta}^{\ell}(s)(M_1(\mu) + \varepsilon) = \sum_{i=1}^{\boldsymbol{\delta}^{\ell}(s)} \boldsymbol{Y}_{a_i} - \boldsymbol{\delta}^{\ell}(s)(M_1(\mu) + \varepsilon) \leq \varepsilon\boldsymbol{\delta}^{\ell}(s)$, and consequently $\boldsymbol{\delta}^{\ell+1}(s) \leq \boldsymbol{\delta}^{\ell}(s)(M_1(\mu) + 2\varepsilon)$. $\quad\blacksquare$

Combining this lemma with Lemma B.7, we obtain the following corollary.

COROLLARY B.2. *Assume that $d_{\max} n^\varepsilon < \boldsymbol{\gamma}^{\ell}(s) < n^{1-\varepsilon}$, where $d_{\max}$ is the maximum degree in the graph, and that $M_1(\mu)$ is finite. Then, w.h.p., $\boldsymbol{\gamma}^{\ell+1}(s) \leq \boldsymbol{\gamma}^{\ell}(s)(M_1(\mu) + \varepsilon)$.*

COROLLARY B.3. *For each vertex v, and for each $0 < x < y < 1$ such that $d_{\max} < n^{x-\varepsilon}$, $\boldsymbol{\tau}_v(n^y) - \boldsymbol{\tau}_v(n^x) \geq (1-\varepsilon)\log_{M_1(\mu)} n^{y-x}$, w.h.p..*

B.2.2 Inhomogeneous Random Graphs. Let us assume that we know the structure of $\boldsymbol{N}^{\ell}(s)$. Following the proof for the CM, we define the auxiliary quantity $\boldsymbol{\delta}^{\ell}(s) = \rho_{\boldsymbol{\gamma}^{\ell}(s)}$. Again, we need to assume that $\rho_{\boldsymbol{N}^{\ell}(s)} < n^{1-\varepsilon}$ and that $\boldsymbol{\delta}^{\ell}(s) > n^\varepsilon$.

Let w be a vertex with weight at most $\frac{n^{1-\varepsilon}}{\boldsymbol{\delta}^{\ell}(s)}$, outside $\boldsymbol{N}^{\ell}(s)$: $\mathbb{P}\left(w \notin \boldsymbol{\gamma}^{\ell+1}(s)\right) = \prod_{v \in \boldsymbol{\gamma}^{\ell}(s)}\left(1 - f\left(\frac{\rho_v \rho_w}{M}\right)\right) = \prod_{v \in \boldsymbol{\gamma}^{\ell}(s)}\left(1 - (1+o(1))\left(\frac{\rho_v \rho_w}{M}\right)\right) = e^{-\sum_{v \in \boldsymbol{\gamma}^{\ell}(s)}(1+o(1))\left(\frac{\rho_v \rho_w}{M}\right)} = e^{-(1+o(1))\left(\frac{\boldsymbol{\delta}^{\ell}(s)\rho_w}{M}\right)} = 1 - (1+o(1))\left(\frac{\boldsymbol{\delta}^{\ell}(s)\rho_w}{M}\right)$. Hence, $\mathbb{P}\left(w \in \boldsymbol{\gamma}^{\ell+1}(s)\right) = (1+o(1))\left(\frac{\boldsymbol{\delta}^{\ell}(s)\rho_w}{M}\right)$, and $\boldsymbol{\gamma}^{\ell+1}(s) = \sum_{w \notin \boldsymbol{N}^{\ell}(s)} \boldsymbol{X}_w$, where the $\boldsymbol{X}_w$s are independent Bernoulli random variables with success probability $(1+o(1))\left(\frac{\boldsymbol{\delta}^{\ell}(s)\rho_w}{M}\right)$ if this quantity is much smaller than 1, otherwise $\mathcal{O}(1)$. We want to compute the number of vertices in $\boldsymbol{\Gamma}^{\ell+1}(s)$, knowing $\boldsymbol{\delta}^{\ell}(s)$: first, we observe that the number of vertices with weight at least $\frac{n^{1-\varepsilon}}{\boldsymbol{\delta}^{\ell}(s)}$ is $\mathcal{O}\left(n\left(\frac{\boldsymbol{\delta}^{\ell}(s)}{n^{1-\varepsilon}}\right)^{\beta-1}\right) = \mathcal{O}\left(\boldsymbol{\delta}^{\ell}(s)\frac{n^{\varepsilon(\beta-1)}\boldsymbol{\delta}^{\ell}(s)^{\beta-2}}{n^{\beta-2}}\right) = \mathcal{O}\left(\boldsymbol{\delta}^{\ell}(s)n^{\varepsilon(\beta-1)-\sqrt{\varepsilon}(\beta-2)}\right) = o(\boldsymbol{\delta}^{\ell}(s))$, assuming $\boldsymbol{\delta}^{\ell}(s) < n^{1-\sqrt{\varepsilon}}$, and we can safely ignore these vertices. By the multiplicative form of Chernoff bound (Lemma B.1), if $\boldsymbol{S} = \sum_{w \notin \boldsymbol{N}^{\ell}(s), \rho_w < \frac{n^{1-\varepsilon}}{\boldsymbol{\delta}^{\ell}(s)}} \boldsymbol{X}_w$:

$$\mathbb{P}\left(\boldsymbol{S} < (1-\varepsilon)\mathbb{E}[\boldsymbol{S}]\right) \leq \left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{\mathbb{E}[\boldsymbol{S}]} \leq e^{(-\varepsilon-(1-\varepsilon)\log(1-\varepsilon))n^\varepsilon} \leq e^{-\varepsilon^3 n^\varepsilon}$$

$$\mathbb{P}\left(\boldsymbol{S} > (1+\varepsilon)\mathbb{E}[\boldsymbol{S}]\right) \leq \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mathbb{E}[\boldsymbol{S}]} \leq e^{(\varepsilon+(1+\varepsilon)\log(1+\varepsilon))n^\varepsilon} \leq e^{-\varepsilon^3 n^\varepsilon}$$

if ε is small enough. By changing the value of ε with $\sqrt{\varepsilon}$, we have proved the following lemma.

LEMMA B.9. *Assume that $\rho_{\boldsymbol{N}^{\ell}(s)} < n^{1-\varepsilon}$ and that $\boldsymbol{\delta}^{\ell}(s) > n^\varepsilon$. Then, $(1-\varepsilon)\boldsymbol{\delta}^{\ell}(s) \leq \boldsymbol{\gamma}^{\ell+1}(s) \leq (1+\varepsilon)\boldsymbol{\delta}^{\ell}(s)$ w.h.p..*

COROLLARY B.4. *For each vertex s, let $\boldsymbol{\tau}'_s(S)$ be the smallest integer such that $\boldsymbol{\delta}^\ell(s) > S$. Then, for each $0 < x < 1$, $\boldsymbol{\tau}'_s((1-\varepsilon)n^x) + 1 \le \boldsymbol{\tau}_s(n^x) \le \boldsymbol{\tau}'_s((1+\varepsilon)n^x) + 1$.*

As in the CM, we need to estimate $\boldsymbol{\tau}'_s(n^x) - \boldsymbol{\tau}'_s(n^y)$. To this purpose, we compute $\rho_{\boldsymbol{\Gamma}^{\ell+1}(s)}$ knowing $\boldsymbol{\delta}^\ell(s)$. Using the previous notations, $\rho_{\boldsymbol{\Gamma}^{\ell+1}(s)} = \sum_{w \notin \boldsymbol{N}^\ell(s)} \rho_w \boldsymbol{X}_w$, and we estimate this sum by considering separately upper and lower bounds, and the different possible values of β.

Lower bound, $2 < \beta < 3$. Let us consider all vertices with weight at least $\boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}$. The probability that one of these vertices is connected to a vertex in $\boldsymbol{\Gamma}^\ell(s)$ is $\Theta\left(\boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}+1}n^{-1}\right) = \Theta\left(\boldsymbol{\delta}^\ell(s)^{\frac{\beta-1-\varepsilon}{\beta-2}}n^{-1}\right)$. Through a straigthforward application of the multiplicative for of Chernoff bound (Lemma B.1), since there are $\Theta\left(n\boldsymbol{\delta}^\ell(s)^{-\frac{1-\varepsilon}{\beta-2}}\right)$ such vertices, we can prove that there is at least one node with weight at least $\boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}$ which is connected to $\boldsymbol{\Gamma}^\ell(s)$, w.h.p.. This means that $\rho_{\boldsymbol{\gamma}^{\ell+1}(s)} \ge \boldsymbol{\delta}^\ell(s)^{\frac{1-\varepsilon}{\beta-2}}$, and consequently, by a union bound, $\rho_{\boldsymbol{\Gamma}^{\ell+i}(s)} \ge \rho_{\boldsymbol{\Gamma}^\ell(s)}^{\left(\frac{1-\varepsilon}{\beta-2}\right)^i}$ for each i such that $\rho_{\boldsymbol{\Gamma}^{\ell+i}(s)} < n^{1-\varepsilon}$.

Upper bound, $2 < \beta < 3$. Let us consider all vertices with weight at least $\boldsymbol{\delta}^\ell(s)^{\frac{1+\varepsilon}{\beta-2}}$. The probability that one of these vertices is connected to a vertex in $\boldsymbol{\Gamma}^\ell(s)$ is $\Theta\left(\boldsymbol{\delta}^\ell(s)^{\frac{1+\varepsilon}{\beta-2}+1}n^{-1}\right) = \Theta\left(\boldsymbol{\delta}^\ell(s)^{\frac{\beta-1+\varepsilon}{\beta-2}}n^{-1}\right)$. Since there are $\Theta\left(n\boldsymbol{\delta}^\ell(s)^{-\frac{(1+\varepsilon)(\beta-1)}{\beta-2}}\right)$ such vertices, by a union bound, the probability that none of these vertices is connected to a vertex in $\boldsymbol{\Gamma}^\ell(s)$ is $1 - \Theta\left(\boldsymbol{\delta}^\ell(s)^{-\frac{\beta-1+\varepsilon-(1+\varepsilon)(\beta-1)}{\beta-2}}\right) = 1 - \Theta\left(n^{-\varepsilon^2}\right)$. Conditioned on this event, the expected value of $\rho_{\boldsymbol{\Gamma}^{\ell+1}(s)}$ is at most $\boldsymbol{\delta}^\ell(s)\sum_{\rho_v < \boldsymbol{\delta}^\ell(s)^{\frac{1+\varepsilon}{\beta-2}}}\frac{\rho_v^2}{n} = \Theta\left(\boldsymbol{\delta}^\ell(s)^{1+\frac{(1+\varepsilon)(3-\beta)}{\beta-2}}\right) = \Theta\left(\boldsymbol{\delta}^\ell(s)^{\frac{1+\varepsilon(3-\beta)}{\beta-2}}\right)$. By Markov inequality, with probability at least $1 - n^{-\varepsilon^3}$, $\rho_{\boldsymbol{\Gamma}^{\ell+1}(s)} \le \boldsymbol{\delta}^\ell(s)^{\frac{1+\varepsilon(3-\beta)}{\beta-2}+\varepsilon} = \boldsymbol{\delta}^\ell(s)^{\frac{1}{\beta-2}+\varepsilon}$. As a consequence, $\rho_{\boldsymbol{\Gamma}^{\ell+i}(s)} \le \boldsymbol{\delta}^\ell(s)^{\left(\frac{1}{\beta-2}\right)^i}$ for each $i < n^{\varepsilon^4}$, a.a.s..

Lower bound, $\beta > 3$. We want to apply Hoeffding's inequality to prove that, if $\boldsymbol{\delta}^\ell(s) > n^\varepsilon$, $\boldsymbol{\delta}^{\ell+1}(s) \ge (1-\varepsilon)\mathbb{E}[\boldsymbol{\delta}^{\ell+1}(s)] \ge (1-2\varepsilon)M_1(\mu)\boldsymbol{\delta}^\ell(s)$. To this purpose, let N be a big constant (to be chosen later): $\boldsymbol{\delta}^{\ell+1}(s) = (1+o(1))\sum_{w \in V} \rho_w \boldsymbol{X}_w \ge (1+o(1))\sum_{\rho_w < N} \rho_w \boldsymbol{X}_w$. By Hoeffding's inequality (Lemma B.2), w.h.p., $\sum_{\rho_w < N} \rho_w \boldsymbol{X}_w$ is at least $(1-\varepsilon)\mathbb{E}\left[\sum_{\rho_w < N} \rho_w \boldsymbol{X}_w\right]$: if we choose N big enough, the latter value is at least $(1-2\varepsilon)M_1(\mu)$. This means that $\boldsymbol{\delta}^{\ell+1}(s) \ge (1-2\varepsilon)M_1(\mu)\boldsymbol{\delta}^\ell(s)$ w.h.p., and by a union bound $\boldsymbol{\delta}^{\ell+i}(s) \ge (1-2\varepsilon)^i M_1(\mu)^i \boldsymbol{\delta}^\ell(s)$.

Upper bound, $\beta > 3$. Conditioned on $\boldsymbol{\delta}^\ell(s)$, the expected value of $\boldsymbol{\delta}^{\ell+1}(s)$ is at most $\sum_{w \notin \boldsymbol{N}^\ell(s)} \rho_w \mathbb{E}[\boldsymbol{X}_w] = (1+o(1))\sum_{w \in V}\rho_w \frac{\boldsymbol{\delta}^\ell(s)\rho_w}{M} = (1+o(1))\boldsymbol{\delta}^\ell(s)\frac{M_2(\lambda)}{M_1(\lambda)} = (1+o(1))\boldsymbol{\delta}^\ell(s)M_1(\mu) \le (M_1(\mu)+\varepsilon)\boldsymbol{\delta}^\ell(s)$. By Markov inequality, $\mathbb{P}\left(\boldsymbol{\delta}^{\ell+i}(s) > (M_1(\mu)+\varepsilon)^i\boldsymbol{\delta}^\ell(s)n^\varepsilon\right) \le n^{-\varepsilon}$.

Proof. [Proof of Theorem B.1, IRG] By Corollary B.4, $\boldsymbol{\tau}'_s((1-\varepsilon)n^x) + 1 \le \boldsymbol{\tau}_s(n^x) \le \boldsymbol{\tau}'_s((1+\varepsilon)n^x) + 1$. Hence, $\boldsymbol{\tau}'_s((1-\varepsilon)n^y) - \boldsymbol{\tau}'_s((1+\varepsilon)n^x) \le \boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \le \boldsymbol{\tau}'_s(n^y(1+\varepsilon)) - \boldsymbol{\tau}'_s(n^x(1-\varepsilon))$. By the aforementioned lower bounds on $\boldsymbol{\delta}^{d+i}(s)$, the conclusion follows.

As before, we conclude this section by proving a stronger upper bound in the case $\beta > 3$, which is used in two of our probabilistic analyses.

LEMMA B.10. *Assume that $\boldsymbol{\delta}^\ell(s) > d_{\max}n^\varepsilon$, where $d_{\max}$ is the maximum degree in the graph, and assume that $\beta > 3$. Then, w.h.p., $\boldsymbol{\delta}^{\ell+1}(s) \le \boldsymbol{\delta}^\ell(s)(M_1(\mu)+\varepsilon)$.*

Proof. As in the lower bound, we write $\boldsymbol{\delta}^{\ell+1}(s) = \rho_{\boldsymbol{\Gamma}^{\ell+1}(s)} \le \sum_{w \in V} \rho_w \boldsymbol{X}_w$, where $\boldsymbol{X}_w$ is a Bernoulli random variable with success probability $1 - \prod_{v \in \boldsymbol{\Gamma}^\ell(s)}\left(1 - f\left(\frac{\rho_v \rho_w}{n}\right)\right) = 1 - \prod_{v \in \boldsymbol{\Gamma}^\ell(s)} e^{-(1+o(1))\frac{\rho_v \rho_w}{n}} = 1 - e^{-(1+o(1))\frac{\rho_{\boldsymbol{\Gamma}^\ell(s)}\rho_w}{n}} = (1+o(1))\left(\frac{\rho_{\boldsymbol{\Gamma}^\ell(s)}\rho_w}{n}\right)$. Hence, $\mathbb{E}\left[\sum_{w \in V}\rho_w \boldsymbol{X}_w\right] = (1+o(1))\left(\sum_{w \in V}\frac{\rho_w^2 \rho_{\boldsymbol{\Gamma}^\ell(s)}}{n}\right) \le (M_1(\mu)+\varepsilon)\boldsymbol{\delta}^\ell(s)$. A simple application of Hoeffding's inequality (Lemma B.2) lets us conclude, since $\rho_w < d_{\max}$ for each w.

Combining this lemma with Lemma B.7, we obtain the following corollary.

COROLLARY B.5. *Assume that $d_{\max}n^\varepsilon < \Gamma^\ell(s) < n^{1-\varepsilon}$, where $d_{\max}$ is the maximum degree in the graph, and that $M_1(\mu)$ is finite. Then, w.h.p., $\boldsymbol{\gamma}^{\ell+1}(s) \leq \boldsymbol{\gamma}^\ell(s)(M_1(\mu) + \varepsilon)$.*

COROLLARY B.6. *For each vertex v, and for each $0 < x < y < 1$ such that $d_{\max} < n^{x-\varepsilon}$, $\boldsymbol{\tau}_v(n^y) - \boldsymbol{\tau}_v(n^x) \geq (1-\varepsilon)\log_{M_1(\mu)} n^{y-x}$, w.h.p..*

B.3 Small Neighborhoods. Using Theorem B.1, we reduced ourselves to prove Property 2.1 and the bounds in Table 2 for some small values of x. We start the proof by formalizing Item 1 in Section 3: the main tool is the relationship between the size $\boldsymbol{\gamma}^\ell(s)$ of a neighbor of a vertex s and a μ-distributed branching process $\boldsymbol{\delta}^\ell(s)$.

THEOREM B.2. *Let $\boldsymbol{G} = (V, \boldsymbol{E})$ be a random graph with degree distribution λ, let μ be the corresponding residual distribution, and let $s \in V$. There are multisets $\boldsymbol{\Delta}^\ell(s)$ of vertices such that:*

1. *the cardinality $\boldsymbol{\delta}^\ell(s)$ of $\boldsymbol{\Delta}^\ell(s)$ is a μ-distributed branching process;*

2. *if $\boldsymbol{\Theta}^\ell(s) = \cup_{i=0}^\ell \boldsymbol{\Delta}^i(s)$, $\mathbb{P}\left(\boldsymbol{\Gamma}^{\ell+1}(s) = \boldsymbol{\Delta}^{\ell+1}(s)\Big|\boldsymbol{\Theta}^\ell(s) = \boldsymbol{N}^\ell(s)\right) = \mathcal{O}\left(\rho^2_{\boldsymbol{\Theta}^\ell(s)}\frac{M_2(\lambda)}{n}\right)$.*

B.3.1 Proof for the Configuration Model. For each vertex v, let us fix a set of stubs $a_{v,1}, a_{v,\rho_v}$ attached to v (let A the set of all stubs).

We define a procedure that generates a random pairing of stubs (and, hence, a graph), by fixing a vertex s and pairing stubs "in increasing order of distance from s", obtaining something similar to a breadth-first search (BFS). This way, in order to understand the structure of $\boldsymbol{n}^\ell(s)$, we only have to consider the first ℓ levels of this BFS, and we may ignore how all other stubs are paired.

The procedure keeps the following information:

- a partial function $\boldsymbol{\alpha} : A \to A$, that represent a partial pairing of stubs;

- for each ℓ, a set $\boldsymbol{I}^\ell$ of all stubs at distance ℓ from s;

- for each ℓ, a set $\boldsymbol{\Delta}^\ell(s)$ of all vertices at distance ℓ from s.

The random part of our procedure is given by a set of random variables $\{\boldsymbol{b}_{a,i}\}_{a\in A, i\in\mathbb{N}}$, whose range is the set of stubs. Informally, stub a "wants to be paired" with stub $\boldsymbol{b}_{a,0}$, if available, otherwise $\boldsymbol{b}_{a,1}$, and so on. We assume that, for each a, $A = \{b \in A : \boldsymbol{b}_{a,i} = b$ for some $i\}$ is infinite (this event occurs with probability 1).

DEFINITION B.4. *The procedure* P1 *starts with $\boldsymbol{\alpha}_1$ as the empty function, $\boldsymbol{\Delta}_1^0(s) = \{s\}$, $\boldsymbol{\Delta}_1^\ell(s) = \emptyset$, $\boldsymbol{I}_1^0 = \{a_{s,1}, \ldots, a_{s,\rho_s}\}$, $\boldsymbol{I}_1^\ell = \emptyset$ for each $\ell > 0$. Then, for increasing values of ℓ, for each stub a in $\boldsymbol{I}_1^\ell$ (any order is fine):*

1. *it sets $\boldsymbol{b}$ as the first $\boldsymbol{b}_{a,i}$ such that $\boldsymbol{\alpha}_1(\boldsymbol{b}_{a,i})$ is undefined;*

2. *it defines $\boldsymbol{\alpha}_1(a) = \boldsymbol{b}$, $\boldsymbol{\alpha}_1(\boldsymbol{b}) = a$;*

3. *if $\boldsymbol{b}$ is not in $\boldsymbol{I}_1^\ell$ for any ℓ:*

   (a) *it adds to $\boldsymbol{I}_1^{\ell+1}$ all stubs of $V(\boldsymbol{b})$ except $\boldsymbol{b}$;*
   (b) *it adds $V(\boldsymbol{b})$ to $\boldsymbol{\Delta}_1^{\ell+1}(s)$, and it sets $V(a)$ as the father of $V(\boldsymbol{b})$;*

4. *else:*

   (a) *it removes $\boldsymbol{b}$ from $\boldsymbol{I}_1^\ell$.*

The procedure ends when $\boldsymbol{I}_1^\ell$ is empty, and all remaining stubs are paired uniformly at random (so that $\boldsymbol{\alpha}_1$ becomes a total function).

At the end, the pairing $\boldsymbol{\alpha}_1$ is uniformly distributed, because, at each step, we choose the "companion" of a stub uniformly among all unpaired stubs. Furthermore, if we consider the graph obtained with the pairing $\boldsymbol{\alpha}_1$, the set $\boldsymbol{\Delta}_1^\ell(s)$ is the set of vertices at distance ℓ from s.

Now, we define another similar, simplified procedure. This time, we let $\boldsymbol{b}$ be any stub, and we do not test if $\boldsymbol{b}$ is not in $\boldsymbol{I}_2^\ell$ for some ℓ, and we add it anyway to $\boldsymbol{I}_2^{\ell+1}$. This way, $\boldsymbol{I}_2^\ell$ and $\boldsymbol{\Delta}^\ell(s)$ become multisets (that is, repetitions are allowed).

DEFINITION B.5. *The procedure* P2 *starts with* $\boldsymbol{\Delta}^0(s) = \{s\}$, $\boldsymbol{\Delta}^\ell(s) = \emptyset$ *for each* $d > 0$, $\boldsymbol{I}_2^0 = \{a_{s,1}, \ldots, a_{s,\rho_s}\}$, $\boldsymbol{I}_2^\ell = \emptyset$ *for each* $d > 0$. *Then, for increasing values of* ℓ, *for each stub* a *in* $\boldsymbol{I}_2^\ell$ *(any order is fine):*

1. *it sets* $\boldsymbol{b} = \boldsymbol{b}_{a,0}$, *and it shifts the* $\boldsymbol{b}_{a,i}s$ *by one;*

2. *it defines* $\boldsymbol{\alpha}_2(a) = \boldsymbol{b}$, $\boldsymbol{\alpha}_2(\boldsymbol{b}) = a$ *(in case, it replaces its value);*

3. *in any case:*

   (a) *it adds to* $\boldsymbol{I}_2^{\ell+1}$ *all stubs of* $V(\boldsymbol{b})$ *except* $\boldsymbol{b}$;
   (b) *it adds* $V(\boldsymbol{b})$ *to* $\boldsymbol{\Delta}^{\ell+1}(s)$, *and it sets* $V(a)$ *as the father of* $V(\boldsymbol{b})$;

The procedure ends when $\boldsymbol{I}_2^\ell$ *is empty, or continues indefinitely.*

Thanks to these simplifications, we are able to prove that, in procedure P2, the cardinality $\boldsymbol{\delta}^\ell(s)$ of $\boldsymbol{\Delta}^\ell(s)$ is a branching process, starting from $\boldsymbol{\delta}^1(s) = \rho_s$.

LEMMA B.11. *In procedure* P2, *the stochastic process* $\boldsymbol{\delta}^\ell(s)$ *is a* μ-*distributed branching process, starting from* $\boldsymbol{\delta}^1(s) = \rho_s$.

Proof. First of all, we observe that $\boldsymbol{\delta}^{\ell+1}(s) = |\boldsymbol{I}_2^\ell|$, so it is enough to prove that $\boldsymbol{I}_2^\ell$ is a branching process. It is clear that $\boldsymbol{I}_2^0 s = \rho_s$. Moreover, $\boldsymbol{I}_2^{\ell+1} = \sum_{a \in \boldsymbol{I}_2^\ell} \rho_{V(\boldsymbol{b}_{a,0})} - 1$. Let $\boldsymbol{X}_i : \rho_{V(\boldsymbol{b}_{a,0})} - 1$: since the $\boldsymbol{b}_{a,0}s$ are independent, also the $\boldsymbol{X}_i s$ are independent, and $\mathbb{P}(\boldsymbol{X}_i = k) = \mathbb{P}(\rho_{V(\boldsymbol{b}_i)}) = k + 1 = \frac{(k+1)n\lambda(k+1)}{\sum_{j=0}^\infty jn\lambda(j)} = \mu(k)$.

With this choice of $\boldsymbol{\delta}^\ell(s)$, we have proved the first part of Theorem B.2. Now, we have to bound the probability that $\boldsymbol{\Delta}^{\ell+1}(s) \neq \boldsymbol{\Gamma}^{\ell+1}(s)$, assuming $\boldsymbol{\Theta}^\ell(s) = \boldsymbol{N}^\ell(s)$: the next lemma gives a bound which is stronger than the bound in Theorem B.2, and it concludes the proof.

LEMMA B.12. *Let* $\boldsymbol{N}^\ell(s) = \bigsqcup_{i=0}^\ell \boldsymbol{\Gamma}^\ell(s)$, $\boldsymbol{\Theta}^\ell(s) = \bigsqcup_{i=0}^\ell \boldsymbol{\Delta}^\ell(s)$, *where* $\bigsqcup$ *denotes the disjoint union, and let* $\boldsymbol{n}^\ell(s) = |\boldsymbol{N}^\ell(s)|$, $\boldsymbol{\theta}^\ell(s) = |\boldsymbol{\Theta}^\ell(s)|$. *Assuming* $\boldsymbol{\Theta}^\ell(s) = \boldsymbol{N}^\ell(s)$, *the probability that* $\boldsymbol{\Delta}^{\ell+1}(s) \neq \boldsymbol{\Gamma}^{\ell+1}(s)$ *is at most* $\frac{1}{n}2\rho_{\boldsymbol{\Theta}^\ell(s)}^2$.

Proof. Let us pair the stubs in $\boldsymbol{\Delta}^\ell(s)$ one by one, and try to bound the probability that a stub is paired "differently". More formally, for each stub a paired by the procedures P1, P2, we bound the probability that a is the first stub that was paired differently (so that we can assume that the pairing of all other stubs was the same). In particular, both procedures choose the companion $\boldsymbol{b}$ of a as $\boldsymbol{b}_{a,0}$, if $\boldsymbol{b}_{a,0}$ is not already in $\boldsymbol{\Theta}^\ell(s)$, and it is not already paired with a stub in $\boldsymbol{\Theta}^\ell(s)$. Hence, the probability that the companion of a is the same in the two procedures is at most the probability that $\boldsymbol{b}_{a,0}$ is not in $\boldsymbol{\Theta}^\ell(s)$, and it is not already paired with a stub in $\boldsymbol{\Theta}^\ell(s)$. This probability is at most $\frac{2\rho_{\boldsymbol{\Theta}^\ell(s)}}{M}$. By a union bound, we can estimate that the probability that at least a stub is paired differently in the two procedures is at most $\rho_{\boldsymbol{\Delta}^\ell(s)} \frac{2\rho_{\boldsymbol{\Theta}^\ell(s)}}{M} \leq \frac{2\rho_{\boldsymbol{\Theta}^\ell(s)}^2}{M} \leq \frac{2\rho_{\boldsymbol{\Theta}^\ell(s)}^2}{n}$.

B.3.2 Proof for Rank-1 Inhomogeneous Random Graphs. In this section, we prove Theorem B.2 in IRG. Let us fix a set V of vertices, let us fix the expected degree ρ_v of each vertex $v \in V$, and let us fix $M = \sum_{v \in V} \rho_v$.

Let s be any vertex, and let us define a procedure that considers edges "in increasing order of distance from s", obtaining something similar to a BFS. This way, in order to understand the structure of $\boldsymbol{n}^\ell(s)$, we only have to consider the first d levels of this BFS, and we may ignore all other edges.

We denote by $\{\boldsymbol{X}_{v,w}\}$ a random variable that has value 1 if the edge (v, w) exists, 0 otherwise. Note that the $\boldsymbol{X}_{v,w}s$ are independent Bernoulli random variables with success probability $f\left(\frac{\rho_v \rho_w}{M}\right)$.

DEFINITION B.6. *The procedure* P1 *starts with* $\boldsymbol{\Delta}_1^0(s) = \{s\}$, $\boldsymbol{\Delta}_1^\ell(s) = \emptyset$. *Then, for increasing values of* d, *for each vertex* $v \in \boldsymbol{\Delta}_1^\ell(s)$:

1. *for each vertex* w *such that* $\boldsymbol{X}_{v,w} = 1$:

(a) if w is not in $\bigcup_{i=0}^{\infty} \boldsymbol{\Delta}_1^i(s)$:

    i. add w to $\boldsymbol{\Delta}_1^{\ell+1}(s)$.

The procedure ends when $\boldsymbol{\Delta}_1^\ell(s)$ is empty.

There are two reasons why this procedure is not a branching process. The first and simplest problem, that occurred also in the CM, is that we need to check that w is "a new vertex", and hence there is dependance between the number of children of different vertices. However, there is also a more subtle problem: if we assume that there is no dependency, we can informally write $\boldsymbol{\delta}_1^{\ell+1}(s) = \sum_{v \in \boldsymbol{\Delta}_1^\ell(s)} \sum_{w \in V} \boldsymbol{X}_{v,w}$. To turn this into a branching process, we have to link $\boldsymbol{\delta}_1^\ell(s)$ with $\boldsymbol{\delta}_1^{\ell+1}(s)$, but the previous formula also depends on which vertices are in $\boldsymbol{\Delta}_1^\ell(s)$. If we condition on which vertices we find in $\boldsymbol{\Delta}_1^\ell(s)$, then the random variables $\sum_{w \in V} \boldsymbol{X}_{v,w}$ are not identically distributed. So, we have to fix $\boldsymbol{\delta}_1^\ell(s)$, write $\boldsymbol{\Delta}_1^\ell(s) = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$, where $\boldsymbol{v}_i$ is a random variable taking values in V, and then set $\boldsymbol{\delta}_1^{\ell+1}(s) = \sum_{i=1}^{\boldsymbol{\Delta}_1^\ell(s)} \sum_{w \in V} \boldsymbol{X}_{\boldsymbol{v}_i, w}$. Now, the random variables $\sum_{w \in V} \boldsymbol{X}_{\boldsymbol{v}_i, w}$ are i.i.d., but the distribution of the weight of $\boldsymbol{v}_i$ (and hence the distribution of the sum) depends on $\boldsymbol{\gamma}^\ell(s)$ in general, so we do not obtain a branching process. Summarizing, the second problem is that we need somehow to choose $\boldsymbol{\delta}^\ell(s)$ before choosing which vertices are in $\boldsymbol{\Delta}^\ell(s)$, then choose $\boldsymbol{v}_i$ in a way that is independent from $\boldsymbol{\delta}^\ell(s)$. It turns out that, if the random variables $\boldsymbol{X}_{v,w}$ are Poisson-distributed, actually the two choices can be made independent. So, in the second procedure we define, we do not only ignore already visited vertices, but we also define new Poisson random variables $\boldsymbol{Y}_{v,w}$ such that $\boldsymbol{Y}_{v,w} = \boldsymbol{X}_{v,w}$ with probability $1 - \mathcal{O}\left(\left(\frac{\rho_v \rho_w}{M}\right)^2\right)$, and we work with $\boldsymbol{Y}_{v,w}$.

LEMMA B.13. *Given a Bernoulli random variable $\boldsymbol{X}$ with success probability $f(p)$, it is possible to define a random variable $\boldsymbol{Y} = \mathrm{Poisson}(p)$ such that $\boldsymbol{X} = \boldsymbol{Y}$ with probability $1 - \mathcal{O}\left(p^2\right)$.*

Proof. Let $\boldsymbol{E}_0, \boldsymbol{E}_1$ be the events $\boldsymbol{X} = 0$, $\boldsymbol{X} = 1$. Let $\boldsymbol{E}_0'$ be an event such that $\boldsymbol{E}_0' \subseteq \boldsymbol{E}_0$ or $\boldsymbol{E}_0' \supseteq \boldsymbol{E}_0$, and $\mathbb{P}(\boldsymbol{E}_0') = \mathbb{P}(\mathrm{Poisson}(p) = 0)$: we define $\boldsymbol{Y} = 0$ in $\boldsymbol{E}_0'$. Similarly, let $\boldsymbol{E}_1'$ be an event such that $\boldsymbol{E}_1' \subseteq \boldsymbol{E}_1$ or $\boldsymbol{E}_1' \supseteq \boldsymbol{E}_1$, $\boldsymbol{E}_1' \cap \boldsymbol{E}_0' = \emptyset$, and $\mathbb{P}(\boldsymbol{E}_1') = \mathbb{P}(\mathrm{Poisson}(p) = 1)$. We define $\boldsymbol{Y} = 1$ in $\boldsymbol{E}_1'$. Then, we cover the rest of the space as we wish.

We know that $\boldsymbol{Y} = \boldsymbol{X}$ on $\boldsymbol{E}_0 \cap \boldsymbol{E}_0'$ and on $\boldsymbol{E}_1 \cap \boldsymbol{E}_1'$: let us prove that the probability of these events is $1 - \mathcal{O}\left(p^2\right)$. Indeed, the probability of $\boldsymbol{E}_0$ is $1 - f(p) = 1 - p + \mathcal{O}(p^2)$, the probability of $\boldsymbol{E}_1$ is $p + \mathcal{O}(p^2)$, the probability of $\boldsymbol{E}_0'$ is $e^{-p} = 1 - p + \mathcal{O}(p^2)$, and the probability of $\boldsymbol{E}_1'$ is $pe^{-p} = p + \mathcal{O}(p^2)$. In any case, $\mathbb{P}((\boldsymbol{E}_0 \cap \boldsymbol{E}_0') \cup (\boldsymbol{E}_1 \cap \boldsymbol{E}_1')) = \min(\mathbb{P}(\boldsymbol{E}_0), \mathbb{P}(\boldsymbol{E}_0')) + \min(\mathbb{P}(\boldsymbol{E}_1), \mathbb{P}(\boldsymbol{E}_1')) = 1 - p + \mathcal{O}(p^2) + p + \mathcal{O}(p^2) = 1 + \mathcal{O}(p^2)$.

DEFINITION B.7. *The procedure* P2 *starts with $\boldsymbol{\Delta}^0(s) = \{s\}$, $\boldsymbol{\Delta}^\ell(s) = \emptyset$. Then, for increasing values of ℓ, for each vertex $v \in \boldsymbol{\Delta}^\ell(s)$:*

  1. *for each vertex w:*

    (a) *in any case:*

        i. *add $\boldsymbol{Y}_{v,w}$ times w to $\boldsymbol{\Delta}^{\ell+1}(s)$;*

        ii. *replace $\boldsymbol{Y}_{v,w}$ with another $\mathrm{Poisson}\left(\frac{\rho_v \rho_w}{M}\right)$ random variable, independent from all previous events.*

The procedure ends when $\boldsymbol{\Delta}^\ell(s)$ is empty, or it continues forever.

THEOREM B.3. *The cardinality $\boldsymbol{\delta}^\ell(s)$ of $\boldsymbol{\Delta}^\ell(s)$ is a μ-distributed branching process, starting from $\boldsymbol{\Delta}^1(s) = \deg(s)$.*

Proof. In this procedure, we got rid of the dependencies between different zones of the branching tree. Hence, if $\boldsymbol{\Delta}^\ell(s) = \{v_1, \dots, v_{\boldsymbol{\delta}^\ell(s)}\}$, we formalize the previous computation by saying that $\boldsymbol{\delta}^{\ell+1}(s) = \sum_{i=1}^{\boldsymbol{\delta}^\ell(s)} \sum_{w \in V} \boldsymbol{Y}_{\boldsymbol{v}_i, w} = \sum_{i=1}^{\boldsymbol{\delta}^\ell(s)} \sum_{w \in V} \mathrm{Poisson}\left(\frac{\rho_{\boldsymbol{v}_i} \rho_w}{M}\right) = \sum_{i=1}^{\boldsymbol{\delta}^\ell(s)} \mathrm{Poisson}\left(\rho_{\boldsymbol{v}_i}\right)$.

It only remains to prove that the probability that $\mathbb{P}\left(\rho_{\boldsymbol{v}_i} = k\right) = \frac{k\lambda(k)}{M_1(\lambda)}$, independently from $\boldsymbol{\Delta}^\ell(s)$. We need the following facts:

- $\boldsymbol{\delta}^\ell(s) = \sum_{v \in \boldsymbol{\delta}^{\ell-1}(s)} \sum_{w \in V} \mathrm{Poisson}\left(\frac{\rho_v \rho_w}{M}\right) = \mathrm{Poisson}\left(\rho_{\boldsymbol{\delta}^{\ell-1}(s)}\right) = \mathrm{Poisson}(\eta)$ if $\eta = \rho_{\boldsymbol{\delta}^{\ell-1}(s)}$;

- if $\boldsymbol{T}_u$ is the number of times that vertex u appears in $\boldsymbol{\Delta}^\ell(s)$, $\boldsymbol{T}_u = \sum_{v\in\boldsymbol{\delta}^{\ell-1}(s)} \mathrm{Poisson}\left(\frac{\rho_u\rho_v}{M}\right) = \mathrm{Poisson}\left(\frac{\rho_u\rho_{\boldsymbol{\Delta}^{\ell-1}(s)}}{M}\right) = \mathrm{Poisson}(\theta)$ if $\theta = \frac{\rho_u\rho_{\boldsymbol{\Delta}^{\ell-1}(s)}}{M}$;

- conditioned on $\boldsymbol{T}_u = k$, $\boldsymbol{\delta}^\ell(s) - k = \sum_{v\in\boldsymbol{\Delta}^{\ell-1}(s)}\sum_{w\in V-\{u\}} \mathrm{Poisson}\left(\frac{\rho_v\rho_w}{M}\right) = \mathrm{Poisson}\left(\rho_{\boldsymbol{\Delta}^{\ell-1}(s)}\frac{M-\rho_u}{M}\right) = \mathrm{Poisson}(\eta - \theta)$.

Using these three results, we can prove that:

$$
\begin{aligned}
\mathbb{P}\left(\boldsymbol{T}_u = k | \boldsymbol{\delta}^\ell(s) = h\right) &= \frac{\mathbb{P}(\boldsymbol{\delta}^\ell(s) = h | \boldsymbol{T}_u = k)\mathbb{P}(\boldsymbol{T}_u = k)}{\mathbb{P}(\boldsymbol{\delta}^\ell(s) = h)} \\
&= \frac{\mathbb{P}(\mathrm{Poisson}(\eta - \theta) = h - k)\mathbb{P}(\mathrm{Poisson}(\theta) = k)}{\mathrm{Poisson}(\eta) = h} \\
&= \frac{e^{-(\eta-\theta)}\frac{(\eta-\theta)^{h-k}}{(h-k)!}e^{-\theta}\frac{\theta^k}{k!}}{e^{-\eta}\frac{\eta^h}{h!}} \\
&= \frac{h!}{k!(h-k)!}\left(\frac{\theta}{\eta}\right)^k\left(1-\frac{\theta}{\eta}\right)^{h-k} = \binom{h}{k}\left(\frac{\rho_u}{M}\right)^k\left(1-\frac{\rho_u}{M}\right)^{h-k}
\end{aligned}
$$

Hence, the probability that u appears k times in our process is exactly the probability that u appears k times if we select $\boldsymbol{\delta}^\ell(s)$ vertices, by picking u with probability $\frac{\rho_u}{M}$. Summing over all vertices u with weight k, $\mathbb{P}(\rho_{\boldsymbol{v}_i} = k) = n\lambda(k)\frac{k}{M} = \frac{k\lambda(k)}{M_1(\lambda)}$. This concludes the proof: indeed, $\boldsymbol{\delta}^{\ell+1}(s) = \sum_{i=1}^{\boldsymbol{\delta}^\ell(s)}\mathrm{Poisson}(\rho_{\boldsymbol{v}_i})$, and $\mathbb{P}(\rho_{\boldsymbol{v}_i} = k) = \frac{k\lambda(k)}{M_1(\lambda)}$: hence, $\mathrm{Poisson}(\rho_{\boldsymbol{v}_i})$ is μ-distributed.

With this choice of $\boldsymbol{\Delta}^\ell(s)$, we have proved the first part of Theorem B.2. Now, we have to bound the probability that $\boldsymbol{\Delta}^{\ell+1}(s) \neq \boldsymbol{\Gamma}^{\ell+1}(s)$, assuming $\boldsymbol{\Delta}^\ell(s) = \boldsymbol{\Gamma}^\ell(s)$: the next lemma gives a bound which is stronger than the bound in Theorem B.2, and it concludes the proof.

LEMMA B.14. *Let* $\boldsymbol{\Theta}^\ell(s) = \bigsqcup_{i=0}^\ell \boldsymbol{\Delta}^i(s)$, *where* $\bigsqcup$ *denotes the disjoint union, and let* $\theta^\ell(s) = |\boldsymbol{\Theta}^\ell(s)|$. *Assuming* $\boldsymbol{\Theta}^\ell(s) = \boldsymbol{N}^\ell(s)$, $\mathbb{P}\left(\boldsymbol{\Delta}^{\ell+1}(s) \neq \boldsymbol{\Gamma}^{\ell+1}(s)\right) \leq \frac{1}{n}2\rho_{\boldsymbol{\Theta}^\ell(s)}^2$.

Proof. The procedures P1 and P2 behave differently only if one of the following holds:

1. $\boldsymbol{Y}_{v,w} > 0$ for some $v \in \boldsymbol{\Delta}^\ell(s), w \in \boldsymbol{\Theta}^\ell(s)$;

2. $\boldsymbol{Y}_{v,w} \neq \boldsymbol{X}_{v,w}$ for some $v \in \boldsymbol{\Delta}^\ell(s), w \in V$

The probability that the first case occurs is $\sum_{v\in\boldsymbol{\Delta}^\ell(s)}\sum_{w\in\boldsymbol{\Theta}^\ell(s)}\frac{\rho_v\rho_w}{M} \leq \frac{\rho_{\boldsymbol{\Theta}^\ell(s)}^2}{M} \leq \frac{\rho_{\boldsymbol{\Theta}^\ell(s)}^2}{n}$. The probability that the second case occurs is $\sum_{v\in\boldsymbol{\Delta}^\ell(s)}\sum_{w\in V}\mathcal{O}\left(\frac{\rho_v^2\rho_w^2}{M^2}\right) = \mathcal{O}\left(\rho_{\boldsymbol{\Delta}^\ell(s)}^2\frac{nM_2(\lambda)}{n^2M_1(\lambda)^2}\right) = \mathcal{O}\left(\rho_{\boldsymbol{\Delta}^\ell(s)}^2\frac{M_2(\lambda)}{n}\right)$.

B.3.3 Bounds for Branching Processes. In order to analyze the neighborhood sizes, we need to better understand the behavior of branching processes. For this reason, we need the following lemmas.

LEMMA B.15. *Let* $\boldsymbol{Z}$ *be a* μ-*distributed branching process, let* ℓ, S *be integers such that* $S \leq \log^2\ell$. *Then, for* ℓ *tending to infinity,* $\mathbb{P}\left(0 < \boldsymbol{Z}^\ell < S\right) \leq (\eta(1) + o(1))^\ell$.

Proof. We divide the proof in two different cases: in the first case, we condition on the fact that $\boldsymbol{Z}^\ell$ eventually dies (for more background on branching processes conditioned on death/survival, we refer to [5]). Conditioned on death, the expected number of descendants after ℓ steps is $Z^0\eta(1)^\ell \leq e^{-\ell(-\log\eta(1))}$: by Markov inequality, the probability that $\boldsymbol{Z}^\ell \geq 1$ is at most $\mathbb{E}[\boldsymbol{Z}^\ell] = e^{-\ell(-\log\eta(1))}$.

In the second case, let $\tilde{\boldsymbol{Z}}^\ell$ be the process $\boldsymbol{Z}^\ell$ conditioned on survival: since $\boldsymbol{Z}^\ell \geq \tilde{\boldsymbol{Z}}^\ell$, it is enough to prove the claim for $\tilde{\boldsymbol{Z}}$. We name "bad" a step of this process in which $\tilde{\boldsymbol{Z}}^{\ell+1} = \tilde{\boldsymbol{Z}}^\ell$ (note that $\tilde{\boldsymbol{Z}}^{\ell+1} \geq \tilde{\boldsymbol{Z}}^\ell$): let us perform

$h = \ell - S$ steps, trying to find S good steps. A step is bad with probability $\eta(1)$, and the probability that at least $\ell - S$ steps are bad is

$$\sum_{i=\ell-S}^{\ell} \binom{\ell}{i} \eta(1)^i (1-\eta(1))^{\ell-i} \leq S\ell^S \eta(1)^{\ell-S} = e^{\log S + S\log \ell - (-\log(\eta(1)))(\ell-S)} = e^{-(1+o(1))\ell(-\log \eta(1))}.$$

If i is a good step, $\tilde{\boldsymbol{Z}}^i \geq \tilde{\boldsymbol{Z}}^{i-1} + 1$, otherwise $\tilde{\boldsymbol{Z}}^i \geq \boldsymbol{Z}^{i-1}$: hence, if there are at least S good steps, $\tilde{\boldsymbol{Z}}^\ell \geq S$.

LEMMA B.16. *Let $\boldsymbol{Z}$ be a μ-distributed branching process with $Z^0 = \omega(1)$, and let $S > Z^0$. Then, for each $\ell > (1+\varepsilon)F\left(Z^0 \to S\right)$, $\mathbb{P}\left(\boldsymbol{Z}^\ell < S\right) \leq e^{-\Omega(Z^0)} + o(1)^{\ell - F\left(Z^0 \to S\right)}$. If $\mu(0) = 0$, $\mathbb{P}\left(\boldsymbol{Z}^\ell < S\right) \leq o(1)^{\ell - F\left(Z^0 \to S\right)}$.*

Proof. We can view the branching process as the sum of Z^0 different branching processes. A standard theorem in the theory of branching processes [5, 1.A.5, Theorem 1] says that the probability that one of this branching processes dies is z_μ, where z_μ is the only integer between 0 and 1 such that $z_\mu = \sum_{i \in \mathbb{N}} \mu(i) z_\mu^i$. Since the different processes are independent, by Chernoff bound, the probability that at least $\frac{Z^0 \mu}{2}$ processes survive is at least $e^{-\frac{Z^0 z_\mu}{8}} = e^{-\Omega(Z^0)}$. Hence, if $\tilde{\boldsymbol{Z}}$ is the process $\boldsymbol{Z}$ conditioned on survival, $\tilde{\boldsymbol{Z}}^0 = \Omega(Z^0)$ with probability $e^{-\Omega(Z^0)}$. Furthermore, if $\mu(0) = 0$, $\tilde{\boldsymbol{Z}}^0 = Z^0$ by definition.

Then, let us perform ℓ steps, and let us estimate $\tilde{\boldsymbol{Z}}^\ell$: a step is "bad" if $\tilde{\boldsymbol{Z}}^{i+1} \leq \tilde{\boldsymbol{Z}}^i M_1(\mu)^{1-\varepsilon}$, if $M_1(\mu)$ is finite, or $\tilde{\boldsymbol{Z}}^{i+1} \leq \left(\tilde{\boldsymbol{Z}}^i\right)^{\frac{1-\varepsilon}{\beta-1}}$ if μ is a power law with exponent β: it is simple to prove that a step is bad with probability at most $o(1)$, if $\tilde{\boldsymbol{Z}}^\ell \geq \tilde{\boldsymbol{Z}}^0 = \Omega(Z^0)$ tends to infinity. If the number of good steps is at least $(1+3\varepsilon)F\left(Z^0 \to S\right)$:

- if $M_1(\mu)$ is finite, $\tilde{\boldsymbol{Z}}^\ell \geq \tilde{\boldsymbol{Z}}^0 M_1(\mu)^{(1-\varepsilon)(1+3\varepsilon)F\left(Z^0 \to S\right)} \geq \tilde{\boldsymbol{Z}}^0 M_1(\mu)^{(1+\varepsilon)F\left(Z^0 \to S\right)} \geq \tilde{\boldsymbol{Z}}^0 \frac{S}{Z^0} \omega(1) \geq S$;

- if μ is power law with $1 < \beta < 2$, $\tilde{\boldsymbol{Z}}^\ell \geq \left(\tilde{\boldsymbol{Z}}^0\right)^{\left(\frac{1-\varepsilon}{\beta-1}\right)^{(1+3\varepsilon)F\left(Z^0 \to S\right)}} \geq \left(\tilde{\boldsymbol{Z}}^0\right)^{\left(\frac{1}{\beta-1}\right)^{(1+\varepsilon)F\left(Z^0 \to S\right)}} \geq$ $e^{\log\left(\tilde{\boldsymbol{Z}}^0\right)\frac{\log(S)}{\log(Z^0)}\Omega(1)} \geq S$.

Let $\ell' = \ell - (1+3\varepsilon)F\left(Z^0 \to S\right)$ be the maximum number of bad steps: by changing the value of ε in the statement, we can assume that $\ell(1-\varepsilon) \geq (1+3\varepsilon)F\left(Z^0 \to S\right)$, and hence $\ell' = \ell - (1+3\varepsilon)F\left(Z^0 \to S\right) \geq \varepsilon\ell$. We need to bind the probability that at least ℓ' steps are bad: this is equal to the probability that the sum of ℓ Bernoulli variables with success probability $o(1)$ is at least ℓ'. This probability is

$$\sum_{i=\ell'}^{\ell} \binom{\ell}{i} (o(1))^i (1-o(1))^{n-i} \leq \ell 2^\ell (o(1))^{\ell'} \leq 2^{\mathcal{O}(\ell')}(o(1))^{\ell'} = o(1)^{\ell'}.$$

COROLLARY B.7. *Let $\boldsymbol{Z}$ be a μ-distributed branching process, and let S be an integer. If $\ell = \omega(1)$, and $\ell > (1+\varepsilon)F\left(Z^0 \to S\right)$, then $\mathbb{P}\left(0 < \boldsymbol{Z}^{(1+\varepsilon)\ell} < S\right) \leq \eta(1)^{\ell - F\left(Z^0 \to S\right)}$.*

Proof. If the process dies, by Lemma B.15 it dies before performing ℓ steps with probability smaller than $\eta(1)^\ell$. Otherwise, by Lemma B.15, $\tilde{\boldsymbol{Z}}^{(1+\varepsilon)\ell - F\left(Z^0 \to S\right)} \geq \log \ell = \omega(1)$ with probability $1 - \frac{\eta(1)^{\ell - F\left(Z^0 \to S\right)}}{2}$, if ℓ is big enough. Conditioned on this event, by Lemma B.16, $\tilde{\boldsymbol{Z}}^{(1+\varepsilon)\ell - F\left(Z^0 \to S\right) + (1+2\varepsilon)F\left(Z^0 \to S\right) + \varepsilon\ell} \geq S$ with probability $1 - o(1)^{\varepsilon\ell + 2\varepsilon F(\log \ell \to S))} \geq 1 - o(1)^{\varepsilon\ell} \geq 1 - \frac{\eta(1)^\ell}{2}$. Summing the two probabilities, $\mathbb{P}\left(0 < \boldsymbol{Z}^{(1+2\varepsilon)\ell} < S\right) \leq \eta(1)^{\ell - F\left(Z^0 \to S\right)}$.

Now, let us prove upper bounds on neighborhood sizes, that correspond to lower bounds on $\boldsymbol{\tau}_s(n^x)$.

LEMMA B.17. *Let us fix $\varepsilon > 0$, and a μ-distributed branching process $\boldsymbol{Z}^\ell$. Given a value $S = \omega(1)$, $\mathbb{P}\left(\forall \ell < (1-\varepsilon)F\left(Z^0 \to S\right), \boldsymbol{Z}^\ell < S\right) \geq 1 - o(1)$.*

Proof. Assume that $M_1(\mu)$ is finite: $\mathbb{E}\left[\boldsymbol{Z}^\ell\right] \leq \mathbb{E}\left[\boldsymbol{Z}^{(1-\varepsilon)F\left(Z^0 \to S\right)}\right] = Z^0 M_1(\mu)^{(1-\varepsilon)F\left(Z^0 \to S\right)} = Z^0 M_1(\mu)^{(1-\varepsilon)\log_{M_1(\mu)}\frac{S}{Z^0}} = Z^0 \left(\frac{S}{Z^0}\right)^{1-\varepsilon} = S^{1-\varepsilon} Z_0^\varepsilon = S\left(\frac{Z^0}{S}\right)^\varepsilon$. We conclude by Markov inequality.

Let us consider the case where μ is a power law distribution with $1 < \beta < 2$. By Lemma B.6 applied with $k = \boldsymbol{Z}^i$, with probability at least $\left(\frac{1}{\boldsymbol{Z}^i}\right)^\varepsilon$, $\boldsymbol{Z}^{i+1} < \left(\boldsymbol{Z}^i\right)^{\frac{1+\varepsilon}{\beta-1}}$. Let us assume $\boldsymbol{Z}^i > \log S$ for each i (increasing the number of elements, we can only increase the number of descendants). Consequently, the probability that $\boldsymbol{Z}^\ell$ is bigger than $\left(Z^0\right)^{\left(\frac{1+\varepsilon}{\beta-1}\right)^\ell}$ is at most $\sum_{i=1}^\ell \left(\frac{1}{\boldsymbol{Z}^i}\right)^\varepsilon \geq \ell\left(\frac{1}{\log S}\right)^\varepsilon = o(1)$ if $\ell = (1-\varepsilon)F\left(Z^0 \to S\right)$. With probability $1 - o(1)$, $\boldsymbol{Z}^k < \left(Z^0\right)^{\left(\frac{1+\varepsilon}{\beta-1}\right)^\ell}$: since $\ell < (1-\varepsilon')F\left(Z^0 \to S\right)$, the claim follows.

Now, we need to prove a corresponding bound for tail probabilities.

LEMMA B.18. *Let us fix $\varepsilon > 0$, and a μ-distributed branching process $\boldsymbol{Z}^\ell$ such that $\boldsymbol{Z}^0 = 1$. Given integers $\ell = \omega(1)$, S such that $F\left(1 \to \ell^2\right) \leq \varepsilon F\left(\ell^2 \to S\right)$, $\mathbb{P}\left(\forall i < (1-\varepsilon)(F\left(Z^0 \to S\right) + \ell), 0 < \boldsymbol{Z}^i < S\right) \geq \eta(1)^\ell$.*

Proof. First of all, let $\tilde{\boldsymbol{Z}}^\ell$ be the corresponding branching process conditioned on survival ($\tilde{\boldsymbol{Z}}^0 = 1$ with probability $\Omega(1)$). Assuming $\tilde{\boldsymbol{Z}}^0 = 1$, the probability that $\tilde{\boldsymbol{Z}}^\ell$ starts with a path of length ℓ is $\eta(1)^\ell$. Now, let us estimate $\mathbb{P}\left(\boldsymbol{Z}^\ell = k \wedge \tilde{\boldsymbol{Z}}^\ell = 1\right) \leq \mathbb{P}\left(\tilde{\boldsymbol{Z}}^\ell = 1 \Big| \boldsymbol{Z}^\ell = k\right) \leq k z_\mu^k$, where z_μ is the probability that a μ-distributed branching process has an infinite number of descendants. Hence, $\mathbb{P}(\tilde{\boldsymbol{Z}}^\ell = 1 \wedge \boldsymbol{Z}^\ell < k) \geq \eta(1)^\ell - \sum_{i=k}^\infty i z_\mu^{i-1} = \eta(1)^\ell - \frac{k z_\mu^{k-1}(1-z_\mu) + z_\mu^k}{(1-z_\mu)^2} = \eta(1)^\ell - \mathcal{O}\left(k z_\mu^{k-1}\right)$.

For $k = \ell^2$, $\mathbb{P}(\boldsymbol{Z}^\ell < \ell^2) \geq \mathbb{P}\left(\boldsymbol{Z}^\ell < \ell^2 \wedge \tilde{\boldsymbol{Z}}^\ell = 1\right) = \eta(1)^\ell - \mathcal{O}\left(\ell^2 z_\mu^{\ell^2-1}\right) = \eta(1)^\ell(1 - o(1))$. Then, let us consider the process $\boldsymbol{Z}_1$ defined by $\boldsymbol{Z}_1^k = \boldsymbol{Z}^{\ell+k}$: since $\boldsymbol{Z}_1^0 < \ell^2$, we know by Lemma B.17 that $\mathbb{P}\left(\forall k < (1-\varepsilon)F\left(\ell^2 \to S\right) + \ell, \boldsymbol{Z}_1^k < S\right) \geq 1 - o(1)$. Since the behavior of $\boldsymbol{Z}_1$ is independent from the behavior of $\boldsymbol{Z}$, and since $F\left(\ell^2 \to S\right) = F\left(1 \to S\right) - F\left(1 \to \ell^2\right) \geq (1-\varepsilon)F\left(1 \to S\right)$,

$$\mathbb{P}\left(\forall k < (1-2\varepsilon)F\left(1 \to S\right) + \ell, \boldsymbol{Z}^k < S\right) = \Omega\left(1 \cdot \eta(1)^\ell(1 - o(1)) \cdot 1\right) = \Omega\left(\eta(1)^\ell\right).$$

We conclude by replacing ℓ with $(1-\varepsilon)\ell$.

B.3.4 Bounds on Neighborhood Sizes. Now, we need to translate the results in the previous section from the realm of branching processes to the realm of random graphs, using Theorem B.2. First of all, the following corollary of Theorem B.2 gives us a simpler bound to decide when the branching process approximation works.

COROLLARY B.8. *Let G be a random graph, $s \in G$. There exists a constant c_λ only depending on λ such that, for each $\ell < n^{c_\lambda}$ $\mathbb{P}\left(\boldsymbol{\Theta}^\ell(v) \neq \boldsymbol{N}^\ell(v)\right) = \mathcal{O}\left(n^{-c_\lambda}\right)$, assuming $\rho_{\boldsymbol{\Delta}^i(s)} < n^{c_\lambda}$ for each $i < \ell$. The same is true if we condition on the size of $\boldsymbol{\Delta}^\ell(v)$.*

Proof. By Theorem B.2,

$$\mathbb{P}\left(\boldsymbol{\Theta}^\ell(v) \neq \boldsymbol{N}^\ell(v)\right) = \sum_{i=1}^\ell \mathbb{P}\left(\boldsymbol{\Theta}^i(v) \neq \boldsymbol{N}^i(v) \wedge \boldsymbol{\Theta}^{i-1}(v) = \boldsymbol{N}^{i-1}(v)\right)$$

$$= \sum_{i=1}^\ell \mathbb{P}\left(\boldsymbol{\Delta}^i(v) \neq \boldsymbol{\Gamma}^i(v) \wedge \boldsymbol{\Theta}^{i-1}(v) = \boldsymbol{N}^{i-1}(v)\right)$$

$$\leq \sum_{i=1}^\ell \mathbb{P}\left(\boldsymbol{\Delta}^i(v) \neq \boldsymbol{\Gamma}^i(v)\big|\boldsymbol{\Theta}^{i-1}(v) = \boldsymbol{N}^{i-1}(v)\right)$$

$$= \mathcal{O}\left(\sum_{i=1}^\ell \rho^2_{\boldsymbol{\Theta}^i(s)} \frac{M_2(\lambda)}{n}\right)$$

$$= \mathcal{O}\left(\ell^3 n^{2c_\lambda} \frac{M_2(\lambda)}{n}\right)$$

$$= \mathcal{O}\left(\frac{M_2(\lambda)}{n^{1-5c_\lambda}}\right).$$

We conclude because $M_2(\lambda) = \mathcal{O}(1)$ if λ has finite variance, $M_2(\lambda) = n^{3-\beta}$ if λ is power law with exponent $2 < \beta < 3$: this means that it is enough to choose c_λ such that $n^{\max(0,3-\beta)-1+5c_\lambda} < n^{-c_\lambda}$, that is, $c_\lambda < \frac{1-\max(0,3-\beta)}{6}$.

From this corollary, it is easy to translate Lemmas B.17 and B.18 in terms of random graphs, at least for values of x smaller than c_λ.

COROLLARY B.9. *Let $G = (V, E)$ be a random graph, let $s \in V$ be in the giant component, and let $x < c_\lambda$ be a fixed, small enough constant. Then, $\mathbb{P}\left(\boldsymbol{\tau}_s(n^x) > (1-\varepsilon)F(\deg(s) \to n^x)\right) \geq 1 - o(1)$. Furthermore, if s has degree 1, $\mathbb{P}\left(\boldsymbol{\tau}_s(n^x) > (1-\varepsilon)(F(1 \to n^x) + \alpha)\right) \geq \eta(1)^\alpha$ for $\alpha = \omega(1)$.*

Proof. By Corollary B.8, assuming $x < c_\lambda$, $\boldsymbol{\Delta}^\ell(s) = \boldsymbol{\gamma}^\ell(s)$ with probability $1-o(1)$; furthermore, by Lemma B.17, $\boldsymbol{\Delta}^\ell(s) < n^x$ for each $\ell < (1-\varepsilon)F(\deg(s) \to n^x)$ with probability $1 - o(1)$.

Similarly, by Lemma B.18, $\mathbb{P}(\forall i < (1-\varepsilon)(F(\deg(s) \to n^x)+\alpha), \boldsymbol{\delta}^i(s) < n^x) \geq \eta(1)^\alpha$, and since $\boldsymbol{\gamma}^i(s) = \boldsymbol{\delta}^i(s)$ with probability $1 - o(1)$ for each $i < \boldsymbol{\tau}_s(n^x)$, we conclude that $\mathbb{P}\left(\boldsymbol{\tau}_s(n^x) > (1-\varepsilon)(F(1 \to n^x) + \alpha)\right) \geq (1 - o(1))\eta(1)^\alpha$.

The translation of the lower bounds is more complicated: the main problem is that, when $\boldsymbol{\Theta}^\ell(s) \neq \boldsymbol{N}^\ell(s)$, we know very little on the size of $\boldsymbol{\Theta}^\ell(s)$. In order to deal also with this case, as soon as $\boldsymbol{\Theta}^\ell(s) \neq \boldsymbol{N}^\ell(s)$, we remove the whole $\boldsymbol{N}^\ell(s)$ from the graph, and we consider the neighborhood growth of a new vertex s' which was in $\boldsymbol{\Gamma}^{\ell+1}(s)$ in the previous graph. We prove that the behavior of the neighbors of s' in the new graph is "very similar" to the behavior of the neighbors of s in the old graph: basically, the only difference is that we re-start from size 1 instead of $\boldsymbol{\delta}^\ell(s)$. However, this difference is compensated by the fact that the probability that $\boldsymbol{\Theta}^\ell(s) \neq \boldsymbol{N}^\ell(s)$ is small. In other words, it is more likely that $\boldsymbol{\delta}^\ell(s)$ remains 1 for ℓ steps, rather than that $\boldsymbol{\Theta}^\ell(s) \neq \boldsymbol{N}^\ell(s)$.

Let us formalize this intuitive proof. First of all, we need to understand what happens when we remove a neighbor from the graph.

LEMMA B.19. *Let $G = (V, E)$ be a random graph, let $s \in V$, let $\ell \in \mathbb{N}$, and let us assume that $\rho_{\boldsymbol{N}^\ell(s)} < n^{1-\varepsilon}$. Then, conditioned on the structure of $\boldsymbol{N}^\ell(s)$, the subgraph induced by $V - \boldsymbol{N}^\ell(s)$ is again a random graph, and the values of $\eta(1)$, $M_1(\mu)$ change by $\mathcal{O}\left(\frac{1}{n^\varepsilon}\right)$.*

Proof. [Proof for the CM] Let us consider the graph obtained from G by removing all the stubs in $\boldsymbol{N}^\ell(s)$, and all the stubs paired with stubs in $\boldsymbol{N}^\ell(s)$. The pairing on the remaining stubs is clearly a random pairing, and the number of stubs removed is at most $n^{1-\varepsilon}$, and if λ' is the degree distribution of $G - \boldsymbol{N}^\ell(s)$, $\sum_{i \in \mathbb{N}} |\lambda(i) - \lambda'(i)| < \frac{1}{n^\varepsilon}$. From this condition, it is easy to prove that $\eta(1)$ and $M_1(\mu)$ cannot change by more than $\mathcal{O}(n^\varepsilon)$, if n is big enough.

Proof. [Proof for IRG] In this case, let us remove $\boldsymbol{N}^\ell(s)$, and let us consider the probability that two vertices outside $\boldsymbol{N}^\ell(s)$ are connected: $\mathbb{P}(E(v,w)) = f\left(\frac{\rho_v \rho_w}{M}\right) = f\left(\frac{\rho_v \rho_w}{M - \rho_{\boldsymbol{N}^\ell(s)}} \frac{M - \rho_{\boldsymbol{N}^\ell(s)}}{M}\right)$. Let $\rho'_v = \rho_v \sqrt{\frac{M - \rho_{\boldsymbol{N}^\ell(s)}}{M}}$: clearly, $\rho'_v = \rho_v(1 + o(1))$, and $G - \boldsymbol{N}^\ell(s)$ is a random graph with weights ρ'_v. Furthermore, if λ' is the degree distribution of $G - \boldsymbol{N}^\ell(s)$, it is clear that the required conditions are satisfied, because the dependency between λ, μ, and η is continuous.

Using this lemma, we may translate Corollary B.7 to the context of random graphs.

LEMMA B.20. *Let G be a graph with a power law degree distribution λ with exponent β, let μ, η be as before. There exists a positive constant c_λ only depending on λ such that, for each ℓ, S such that $\ell = \mathcal{O}(\log n)$, $n^\varepsilon < S < n^{c_\lambda}$,*
$$\mathbb{P}\left(\forall \ell' < \ell(1 + \varepsilon), 0 < \boldsymbol{\gamma}^{\ell'}(s) < S\right) = \mathcal{O}\left(\eta(1)^{\ell - F\left(Z^0 \to S\right)}\right).$$

Proof. First of all, we may assume that $\ell - F\left(Z^0 \to S\right) = \omega(1)$, otherwise the probabilistic bound is trivial. By Corollary B.8, the three following cases are possible:

- $\boldsymbol{N}^\ell(s) = \boldsymbol{\Theta}^\ell(s)$;

- $\rho_{\boldsymbol{\Gamma}^i(s)} \geq 4Sn^\varepsilon$ for some $i < \ell$;

- none of the two cases above applies.

In the first case, the result follows directly by Corollary B.7. In the second case, let i be the smallest integer such that $\rho_{\boldsymbol{\Gamma}^i(s)} \geq 4Sn^\varepsilon$: in IRG, by Lemma B.9, $\boldsymbol{\Gamma}^{i+1}(s) \geq (1 - \varepsilon)\rho_{\boldsymbol{\Gamma}^i(s)}$ w.h.p., and $\boldsymbol{\tau}_s\left(n^x\right) < i + 1 < \ell$. In the CM, $\boldsymbol{\delta}^{i-1}(s) + \boldsymbol{\delta}^i(s) \geq 4Sn^\varepsilon$, and as a consequence either $\boldsymbol{\delta}^{i-1}(s) \geq 2Sn^\varepsilon$ or $\boldsymbol{\delta}^i(s) \geq 2Sn^\varepsilon$: by Lemma B.7, $\boldsymbol{\tau}_s\left(n^x\right) < i + 1 < \ell$.

It only remains to solve the third case. The probability that this case occurs is $\mathcal{O}\left(n^{-c_\lambda}\right)$ by Corollary B.8. However, n^{-c_λ} is not sufficient for our purposes, because $\eta(1)^{\ell - F\left(Z^0 \to S\right)}$ can be much smaller. Let us consider the following process: we explore neighbors of v of increasing size, until we hit a neighbor i verifying $\boldsymbol{\Delta}^i(s) \neq \boldsymbol{\Gamma}^i(s)$. If $\boldsymbol{\Delta}^i(s) \neq \boldsymbol{\Gamma}^i(s)$, either all vertices in $\boldsymbol{\Gamma}^i(s)$ have all edges directed inside $\boldsymbol{\Gamma}^i(s)$, and $\boldsymbol{\Gamma}^{i+1}(s)$ is empty, or there is at least a vertex v with an edge directed outside $\boldsymbol{\Gamma}^i(s)$. In the former case, we know that $\boldsymbol{\gamma}^{i+1}(s) = 0$, and the conclusion follows. In the latter case, we remove $\boldsymbol{\Gamma}^i(s)$ from the graph: the size of $\boldsymbol{\Gamma}^{i+j}(s)$ is at least the size of $\boldsymbol{\Gamma}^j(v')$, where v' is the neighbor of v outside $\boldsymbol{\Gamma}^i(s)$. Furthermore, by Lemma B.19, $G - \boldsymbol{\Gamma}^i(s)$ is a random graph, with degree distribution very similar to the degree distribution of G: indeed, $i < \ell = \mathcal{O}(\log n)$, and the volume of vertices removed is at most $S \log n \leq n^{1-\varepsilon}$. Moreover, the size of the neighbors of v' is independent from all previous events, because all we knew about v' has been removed from the graph. Then, we can restart the exploration from v', in the new graph: if $\boldsymbol{\Delta}^j(v') \neq \boldsymbol{\Gamma}^j(v')$, we proceed again as before.

More formally, let us fix ℓ, and let $P(\ell, h)$ be the probability that $\boldsymbol{\Gamma}^\ell(s) < S$, and that $\boldsymbol{\Delta}^j(s) \neq \boldsymbol{\Gamma}^j(s)$ happened h times in the aforementioned process. We prove by induction on h that $P(\ell, h) \leq e^{-(1+\varepsilon)(\ell - F\left(Z^0 \to S\right))(-\log \eta(1))}$. The base case follows by our initial argument. For inductive step, let ℓ' be the smallest integer such that $\boldsymbol{\Gamma}^{\ell'}(s) \neq \boldsymbol{\Delta}^{\ell'}(s)$: note that $\mathbb{P}\left(\ell' = i\right) \leq \mathbb{P}\left(\ell' < \ell\right) \leq \frac{n^{-k}}{\ell} \leq n^{-k+\varepsilon}$, and that, by inductive hypothesis, $P(\ell - \ell', h) \leq e^{-(-\log \eta(1) + \varepsilon)(\ell - \ell')}$ if $\ell - \ell' \geq \log S$, and consequently $P(\ell - \ell', h) \leq e^{-(-\log \eta(1) + \varepsilon)(\ell - \ell' - \log S)}$.

$$\mathbb{P}(\ell, h+1) \leq \sum_{i=0}^\ell \mathbb{P}(\ell' = i) P(i, 0) P(\ell - i, S, h)$$

$$\leq \sum_{i=0}^\ell n^{-k+\varepsilon} e^{-(-\log \eta(1) + \varepsilon)(i - F\left(Z^0 \to S\right))} e^{-(-\log \eta(1) + \varepsilon)(\ell - i - \log S - F\left(Z^0 \to S\right))}$$

$$\leq n^{-k+\varepsilon} e^{-(-\log \eta(1) + \varepsilon)(\ell - 2F\left(Z^0 \to S\right) - \log S)}$$

$$\leq e^{-(-\log \eta(1) + \varepsilon)(\ell - F\left(Z^0 \to S\right))} e^{-(k-\varepsilon)\log n + F\left(Z^0 \to S\right) + \log S}.$$

The inductive step is proved, if $e^{-(k-\varepsilon)\log n + F\left(Z^0 \to S\right) + \log S} < 1$, that is, $(\min(1, \beta - 2) - 2\log_n S - 2\varepsilon)\log n > F\left(Z^0 \to S\right) + \log S$, which is implied by $(\min(1, \beta - 2) - 2\varepsilon)\log n > \log_{M_1(\mu)} S + 3\log S$, that is, $S\left(3 + \frac{1}{\log M_1(\mu)}\right) < n^{\min(1, \beta - 2) - 2\varepsilon}$. The lemma follows by choosing the right value of c_λ.

By combining this lemma with Theorem B.1, we have proved the following theorem.

THEOREM B.4. *Let $G = (V, E)$ be a random graph, let λ be the degree distribution of G, let μ, η be as before, and let $0 < x < 1$. Then, if $s \in V$, $\deg(v) = d$, the following hold:*

- $\boldsymbol{\tau}_s(n^x) \geq (1 - \varepsilon)F(d \to n^x)$ *a.a.s.;*

- $\mathbb{P}(\boldsymbol{\tau}_s(n^x) \geq (1 + \varepsilon)(\alpha + F(d \to n^x))) = \mathcal{O}(\eta(1)^\alpha);$

- $\mathbb{P}(\boldsymbol{\tau}_s(n^x) \geq (1 - \varepsilon)(\alpha + F(d \to n^x))) = \Omega(\eta(1)^\alpha).$

B.4 The Case $1 < \beta < 2$. In the case $1 < \beta < 2$, the branching process approximation does not work: indeed, the distribution μ is not even defined, because $M_1(\lambda)$ is infinite. For this reason, we need a different analysis, which looks similar to the "big neighbors" analysis in the case $\beta > 2$. We prove that the graph which is generated from this distribution has a very dense core, which is made by all vertices whose degree is big enough: almost all the other vertices are either connected to the core, or isolated, so that the average distance between two nodes is 2 or 3. There are also some paths of length $\mathcal{O}(1)$ leaving the core, whose length depends on the value β of the distribution.

In our analysis, in order to avoid pathological cases, we have to assume that $\rho_v < (1 - \varepsilon)M$ for each v in the Chung-Lu model (otherwise, all vertices with weight at least $1 + \varepsilon$ would be connected to the maximum degree vertex). Note that this event holds with probability $\mathcal{O}(1)$.

Before entering the details of our analysis, we need some probabilistic lemmas that describe the relationship between the weight and the degree of a vertex.

LEMMA B.21. ([43], EQUATION A.1.7) *For each $\varepsilon > 0$, there exists N_ε and C_ε not depending on n such that the following hold a.a.s.:*

- $M = (1 + \varepsilon) \sum_{i=1}^{N_\varepsilon} \rho_i;$

- $\rho_1 \leq C_\varepsilon \rho_{N_\varepsilon}.$

COROLLARY B.10. *The vertex with maximum weight has weight $\Theta\left(n^{\frac{1}{\beta-1}}\right)$ a.a.s., and $M = \Theta\left(n^{\frac{1}{\beta-1}}\right)$.*

In this regime, we still need the definitions of $\boldsymbol{\Delta}^\ell(s)$ as in the case $\beta > 2$, but in this case we will not prove that $\boldsymbol{\gamma}^{\ell+1}(s)$ is close to $\boldsymbol{\delta}^\ell(s)$ w.h.p.: for example, let s be a vertex with weight $M = \Theta\left(n^{\frac{1}{\beta-1}}\right)$: clearly, $\boldsymbol{\gamma}^{\ell+1}(s)$ cannot be M, which is bigger than n. Indeed, we prove that $\boldsymbol{\gamma}^{\ell+1}(s)$ is close to $\boldsymbol{\Delta}^\ell(s)^{\beta-1}$: this way, the number of neighbors of a vertex s with weight $\Theta(M)$ is close to n, which makes sense. In order to prove this result, we need a technical lemma on the volume of some vertices.

LEMMA B.22. *Given a random graph with a degree distribution λ which is power law with exponent $1 < \beta < 2$, $\sum_{\rho_w \leq d} \rho_w = \mathcal{O}\left(nd^{2-\beta}\right)$.*

Proof. This result is a simple application of Abel's trick to estimate a sum: $\sum_{\rho_w \leq d} \rho_w = \sum_{i=1}^d i(|\{w : \rho_w \geq i\}| - |\{w : \rho_w \geq i + 1\}|) = \sum_{i=1}^d i|\{w : \rho_w \geq i\}| - \sum_{i=2}^{d+1} |(i-1)\{w : \rho_w \geq i\}|) \leq \sum_{i=1}^d |\{w : \rho_w \geq i\}| = \sum_{i=1}^d \mathcal{O}\left(\frac{n}{i^{\beta-1}}\right) = \mathcal{O}\left(n \int_{\frac{1}{2}}^d x^{1-\beta} dx\right) = \mathcal{O}\left(nd^{2-\beta}\right).$

Using this lemma, we can formally prove the relation between $\boldsymbol{\delta}^\ell(s)$ and $\boldsymbol{\gamma}^{\ell+1}(s)$.

LEMMA B.23. *For each $\varepsilon > 0$, and for each ℓ such that $\boldsymbol{n}^\ell(s) < \boldsymbol{\delta}^\ell(s)^{\beta-1}n^{-\varepsilon}$, and $\boldsymbol{\delta}^\ell(s) > n^\varepsilon$, $\boldsymbol{\gamma}^{\ell+1}(s) = \Theta\left(\boldsymbol{\delta}^\ell(s)^{\beta-1}\right)$.*

Proof. Let us fix $\varepsilon > 0$, and let us prove that $\boldsymbol{\delta}^{\ell+1}(v) \geq \boldsymbol{\gamma}^\ell(v)^{\beta-1}$. Let us consider the set W made by all vertices with weight at least $\frac{M}{\boldsymbol{\delta}^\ell(s)}$, not in $\boldsymbol{N}^\ell(s)$: there are $\Theta\left(n\left(\frac{\boldsymbol{\delta}^\ell(s)}{M}\right)^{\beta-1}\right) = \Theta\left(\boldsymbol{\delta}^\ell(s)^{\beta-1}\right)$ such vertices,

because $\boldsymbol{n}^\ell(s) < \boldsymbol{\delta}^\ell(s)^{\beta-1}$. We want to apply concentration inequalities to prove that there are $\mathcal{O}(|W|)$ vertices in W that are in $\boldsymbol{\Delta}^{\ell+1}(s)$. First of all, let us assume without loss of generality that $\boldsymbol{\delta}^\ell(s) > n^\varepsilon$, otherwise this inequality is empty. In the Configuration Model, let us sort the vertices in W, obtaining $w_1, \ldots, w_k$, and let us consider a procedure where we pair stubs of w_i until we find a connection to $\boldsymbol{\Delta}^\ell(s)$. Since, at each step, the number of stubs in $\boldsymbol{\Delta}^\ell(s)$ that are not paired with a vertex in W is $\mathcal{O}(\boldsymbol{\Delta}^\ell(s))$, and w_i has $\frac{Mn^\varepsilon}{\boldsymbol{\Delta}^\ell(s)}$ stubs, at each step there is probability $\mathcal{O}(1)$ that w_i is connected to a vertex in $\boldsymbol{\Gamma}^\ell(s)$. A simple application of Azuma's inequality lets us conclude. In IRG, the probability that a vertex $w \in W$ is linked to a vertex in $\boldsymbol{\Gamma}^{\ell+1}(s)$ is at least $\sum_{v \in \boldsymbol{\Gamma}^\ell(s)} f\left(\frac{\rho_v}{\boldsymbol{\delta}^\ell(s)}\right) = \mathcal{O}(1)$: a simple application of the multiplicative form of Chernoff bound (Lemma B.1) lets us conclude.

For an upper bound, we can divide the vertices in $\boldsymbol{\Gamma}^{\ell+1}(s)$ in two sets W, W', where W is the set of vertices with weight at most $\frac{M}{\rho_v}$, $W' = W^C$. For W', the number of vertices with weight at least $\frac{M}{\rho_v}$ is $\mathcal{O}\left(n\left(\frac{\rho_v}{M}\right)^{\beta-1}\right) = \mathcal{O}\left(\rho_v^{\beta-1}\right)$, and hence the number of neighbors of v in W' is at most $\rho_v^{\beta-1}$. For the set W, we have to consider separately IRG and the CM. In the first case, let $\boldsymbol{X}_w = 1$ if $w \in \boldsymbol{\Gamma}^{\ell+1}(s)$, 0 otherwise: we want to estimate $\sum_{w \in W} \boldsymbol{X}_w$. Through the previous lemma, the expected value of this sum is:

$$
\begin{aligned}
\mathbb{E}\left[\sum_{w \in W} \boldsymbol{X}_w\right] &= \sum_{w \in W} \sum_{v \in \boldsymbol{\Gamma}^\ell(s)} f\left(\frac{\rho_v \rho_w}{M}\right) \\
&= \sum_{w \in W} \sum_{v \in \boldsymbol{\Gamma}^\ell(s)} (1 + o(1)) \frac{\rho_v \rho_w}{M} \\
&= (1 + o(1)) \frac{\boldsymbol{\delta}^\ell(s)}{M} \sum_{w \in W} \rho_w \\
&= \mathcal{O}\left(\frac{\boldsymbol{\delta}^\ell(s)}{M} n \left(\frac{M}{\boldsymbol{\delta}^\ell(s)}\right)^{2-\beta}\right) \\
&\leq \boldsymbol{\delta}^\ell(s)^{\beta-1}.
\end{aligned}
$$

Since these random variables are independent, we can apply Chernoff bound to prove that $\sum_{w \in W} \boldsymbol{X}_w \leq \mathbb{E}\left[\sum_{w \in W} \boldsymbol{X}_w\right]$.

In the CM, let $a_1, \ldots, a_{\boldsymbol{\delta}^\ell(s)}$ be the stubs in $\boldsymbol{\Delta}^\ell(s)$. By the previous lemma, the number of stubs in W is $\sum_{w \in W} \rho_w = \mathcal{O}\left(\frac{nM^{2-\beta}}{\boldsymbol{\delta}^\ell(s)^{2-\beta}}\right) = \mathcal{O}\left(\frac{M}{\boldsymbol{\delta}^\ell(s)^{2-\beta}}\right)$. The number of vertices in $W \cap \boldsymbol{\gamma}^{\ell+1}(s)$ is at most the number of stubs in $\boldsymbol{\delta}^\ell(s)$ which are paired with stubs in W: let us pair stubs in $\boldsymbol{\delta}^\ell(s)$ in order: at each step, the probability that we hit a stub in W is $\mathcal{O}\left(\frac{1}{M}\frac{M}{\boldsymbol{\delta}^\ell(s)^{2-\beta}}\right)$, because there are still $\mathcal{O}(M)$ stubs outside W. A simple application of Azuma's inequality proves that $\boldsymbol{\gamma}^{\ell+1}(s) \leq \mathcal{O}\left(\frac{\boldsymbol{\delta}^\ell(s)}{\boldsymbol{\delta}^\ell(s)^{2-\beta}}\right) = \mathcal{O}\left(\boldsymbol{\delta}^\ell(s)^{\beta-1}\right)$.

COROLLARY B.11. *For each vertex s with degree at least n^ε, the number of neighbors of s is $\Theta\left(\rho_s^{\beta-1}\right)$.*

Proof. Apply the previous lemma with $\ell = 0$.

Using the last two results, we can transform statements dealing with the number of vertices to statements dealing with weights. For this reason, we can analyze the weights, which are much simpler.

LEMMA B.24. *The probability that a vertex v with weight ρ_v is connected to a vertex with weight at most ρ is $\mathcal{O}\left(\frac{n\rho_v \rho^{2-\beta}}{M}\right)$.*

Proof. First, we can assume that $\rho_v \ll \frac{M}{n\rho^{2-\beta}}$, otherwise the thesis of the lemma is trivially true.

In the CM, let us pair all the stubs of v in order. At each step, the probability that we hit a stub whose vertex has weight at most ρ is $\mathcal{O}\left(\frac{1}{M}\sum_{w:\rho_w<\rho} \rho_w\right)$, because we have paired at most $\rho_v \ll M$ vertices. Summing over all stubs of v, we obtain $\frac{\rho_v}{M}\sum_{w:\rho_w<\rho} \rho_w = \mathcal{O}\left(\frac{n\rho_v \rho^{2-\beta}}{M}\right)$ by Lemma B.22.

In IRG, this probability is $\sum_{w:\rho_w<\rho} f\left(\frac{\rho_v\rho_w}{M}\right) = (1+o(1))\frac{\rho_v}{M}\sum_{w:\rho_w<\rho}\rho_w = \mathcal{O}\left(\frac{n\rho_v\rho^{2-\beta}}{M}\right)$ by Lemma B.22.

LEMMA B.25. *A vertex v with degree at least $\log^2 n$ is w.h.p. connected to all vertices with weight at least εM.*

Proof. In IRG, this lemma follows from our assumptions on f. In the CM, let v be a vertex with degree at least $\log^2 n$, let w be a vertex with degree at least εM, and let $a_1,\ldots,a_k$ be the stubs of v. Let us pair the stubs a_i in order: at each step, the probability that a_i is connected to a stub in w is at least ε. Hence, by Azuma's inequality (Lemma B.4), at least one of the stubs a_i is connected to a stub in W.

LEMMA B.26. *For each vertex s with degree at most $n^{1-\varepsilon}$, $\mathbb{P}\left(\boldsymbol{\tau}_s\left(n^{1-\varepsilon}\right)=2\right)=1-\frac{1}{n^{\mathcal{O}(\varepsilon)}}$.*

Proof. Since $\deg(s)<n^{1-\varepsilon}$, $\boldsymbol{\tau}_s\left(n^{1-\varepsilon}\right)\geq 2$. For the lower bound, if s is connected to a vertex with weight $M^{1-\frac{\varepsilon}{2}}$, then $\boldsymbol{\tau}_s\left(n^{1-\varepsilon}\right)\leq 2$ by Corollary B.11. By Lemma B.25, this happens w.h.p. if $\deg(v)>\log^2 n$: for this reason, the only remaining case is when $\deg(v)<\log^2 n$, and v is not connected to any vertex with weight $n^{1-\frac{\varepsilon}{2}}$. In this case, we prove that v is likely to be isolated: indeed, let us bind the probability that v is connected to a vertex w with degree at most $n^{1-\frac{\varepsilon}{2}}$ (hence, with weight at most $M^{1-\varepsilon'}$). By Lemma B.24, this probability is $\mathcal{O}\left(\frac{n\rho_v M^{(2-\beta)(1-\varepsilon')}}{M}\right)=\mathcal{O}\left(n\log n M^{1-\beta}M^{-\varepsilon'(2-\beta)}\right)=\mathcal{O}\left(n^{-\varepsilon''}\right)$. This means that, by Markov inequality, the number of vertices that are not isolated and not connected to a vertex with degree $n^{1-\varepsilon}$ is at most $n^{1-\varepsilon''}$, a.a.s.. We conclude that $T\left(d\to n^{1-\varepsilon}\right)\geq 2+\mathcal{O}\left(n^{-\varepsilon''}\right)$.

Let us now estimate the deviations from this probability.

LEMMA B.27. *For each vertex s, $\mathbb{P}\left(\boldsymbol{\tau}_s\left(n^x\right)=\ell\right)\leq n^{1-\frac{2-\beta}{\beta-1}(\ell-2-x)+o(1)}$.*

Proof. If $\deg(s)>\log^2 n$, $\boldsymbol{\tau}_s\left(n^x\right)\leq 2$ w.h.p.. Otherwise, since all vertices with degree at least $\log^2 n$ are connected to the vertex with maximum degree, $\boldsymbol{\tau}_s\left(n^x\right)=\ell$ implies that all vertices at distance at most $\ell-3$ from s have degree at most $\log^2 n$. Hence, $\boldsymbol{\gamma}^i(v)\leq\log^{2\ell} n=n^{o(1)}$ for each $i\leq\ell-3$. This means that, for each $i\leq\ell-4$, there is a vertex in $\boldsymbol{\gamma}^i(v)$ with weight $n^{o(1)}$ connected to another vertex with weight $n^{o(1)}$. The probability that this happens is at most $n^{-\frac{2-\beta}{\beta-1}+o(1)}$, because there are $n^{o(1)}$ such vertices, and we may apply Lemma B.24 to each of them.

Since these events are independent, if we multiply the probabilities for each i between 0 and $\ell-4$, the probability becomes $n^{-(\ell-3)\frac{2-\beta}{\beta-1}+o(1)}$. Finally, all vertices in $\boldsymbol{\gamma}^{\ell-3}(s)$ should be connected to vertices with degree at most n^x, and hence to vertices with weight at most $\mathcal{O}\left(n^{\frac{x}{\beta-1}}\right)$. Again by Lemma B.24, the probability that this event happens is $n^{o(1)}\frac{nn^{o(1)}n^{\frac{x(2-\beta)}{\beta-1}}}{M}=n^{1-\frac{1}{\beta-1}+\frac{x(2-\beta)}{\beta-1}+o(1)}=n^{-\frac{2-\beta}{\beta-1}(1-x)+o(1)}$. Overall, the probability that $\boldsymbol{\tau}_s\left(n^x\right)=\ell$ is at most $n^{-\frac{2-\beta}{\beta-1}(\ell-2-x)+o(1)}$.

LEMMA B.28. *For each vertex s with degree 1, $\mathbb{P}\left(\boldsymbol{\tau}_s\left(n^x\right)=\ell\right)\geq n^{1-\frac{2-\beta}{\beta-1}(\ell-2-x)+o(1)}$.*

Proof. Let s be a vertex of weight 1: we want to estimate the probability that s is connected to a vertex of weight 2 in the CM, with weight 1 in IRG. This probability is $\frac{2n\lambda(2)}{M}=n^{-\frac{2-\beta}{\beta-1}+o(1)}$ in the CM, $1-\left(1-f\left(\frac{1}{M}\right)\right)^n=1-\left(1-\frac{1+o(1)}{M}\right)^n=1-e^{\frac{-n(1+o(1))}{M}}=\frac{n(1+o(1))}{M}=n^{-\frac{2-\beta}{\beta-1}+o(1)}$. Assuming this event holds, the probability that s is not connected to any other vertex is 1 in the CM, and it is $\mathcal{O}(1)$ in IRG, assuming the maximum weight is smaller than $(1-\varepsilon)M$. This means that, with probability $\mathcal{O}(1)n^{-\frac{2-\beta}{\beta-1}+o(1)}=n^{-\frac{2-\beta}{\beta-1}+o(1)}$, s is connected to a single vertex s_1 with weight 1 in IRG, 2 in the CM. We may re-iterate the process with s_1, finding a new vertex s_2, and so on, for $\ell-3$ steps. The probability that we find a path of length $\ell-3$ is $n^{-(\ell-3)\frac{2-\beta}{\beta-1}+o(1)}$. Then, let us estimate the probability that $s_{\ell-3}$ is connected only to a vertex with degree at most n^x. In the CM, the number of stubs of vertices with degree at most n^x is $\sum_{\rho_w<n^{\frac{x}{\beta-1}}}\rho_w=\mathcal{O}\left(n^{1+x\frac{2-\beta}{\beta-1}}\right)$, and hence the probability that we hit a stub of a vertex with degree at most n^x is $\mathcal{O}\left(\frac{n^{1+x\frac{2-\beta}{\beta-1}}}{M}\right)=n^{-(1-x)\frac{2-\beta}{\beta-1}+o(1)}$. In IRG, the

probability is $1 - \prod_{\rho_w < n^{\frac{x+o(1)}{\beta-1}}} \left(1 - f\left(\frac{\rho_w}{M}\right)\right) = 1 - \prod_{\rho_w < n^{\frac{x+o(1)}{\beta-1}}} \left(1 - \frac{\rho_w(1+o(1))}{M}\right) = 1 - \prod_{\rho_w < n^{\frac{x+o(1)}{\beta-1}}} e^{-\frac{\rho_w(1+o(1))}{M}} =$

$1 - e^{-\frac{n^{\frac{1+x\frac{2-\beta}{\beta-1}+o(1)}}{M}}} = 1 - e^{-n^{-(1-x)\frac{2-\beta}{\beta-1}+o(1)}} = n^{-(1-x)\frac{2-\beta}{\beta-1}+o(1)}.$

In both cases, we proved that the probability of having a path of length $\ell - 2$ followed by a vertex with degree at most n^x is at most $n^{-(\ell-2-x)\frac{2-\beta}{\beta-1}+o(1)}$. It is clear that in this case $\boldsymbol{\tau}_s(n^x) \geq \ell$.

Summarizing the results obtained in this section, we have proved the following theorem.

THEOREM B.5. *Let* $G = (V, E)$ *be a random graph with degree distribution* λ*, which is power law with* $1 < \beta < 2$*. Then, if* $s \in V$*,* $\deg(v) = d$*, for each* x *between* 0 *and* 1*, the following hold:*

- $\boldsymbol{\tau}_s(n^x) \leq 2$ *a.a.s.;*

- $\mathbb{P}(\boldsymbol{\tau}_s(n^x) \geq \alpha + 2) \leq nc^{\alpha-x+o(1)}$;

- $\mathbb{P}(\boldsymbol{\tau}_s(n^x) \geq \alpha + 2) \geq nc^{\alpha+1-x+o(1)}$.

B.5 Applying the Probabilistic Bounds. Until now, we have proved bounds on the probability that $\boldsymbol{\tau}_s(n^x)$ has certain values. In this section, we turn these probabilistic bounds into bounds on the number of vertices that satisfy a given constraint, concluding the proof of the main theorems, and of the values in Table 2. The main tool used in the following lemma.

LEMMA B.29. *For each vertex* t*, let* $\boldsymbol{E}_\ell(t)$ *be an event that only depends on the structure of* $\boldsymbol{N}^\ell(t)$*. Then, for each set* $\boldsymbol{T} \subseteq V$*,* $0 < x < 1$*, if* $\boldsymbol{E}(t)$ *is the event* $\forall \ell < \boldsymbol{\tau}_t(n^x) - 1, \boldsymbol{E}_\ell(t)$

$$|\{t \in \boldsymbol{T} : \boldsymbol{E}(t)\}| = (1 \pm o(1)) \sum_{t \in \boldsymbol{T}} \mathbb{P}(\boldsymbol{E}(t)|t \in \boldsymbol{T}) \pm |\boldsymbol{T}| \frac{M^{2x}}{M^{1-o(1)}}.$$

If we condition on the structure of a neighbor with volume at most n^y*, a very similar result holds:*

$$|\{t \in T : \boldsymbol{E}(t)\}| = (1 \pm o(1)) \sum_{t \in T} \mathbb{P}(\boldsymbol{E}(t)|t \in T) \pm |T| \left(\frac{M^{x+y} + M^{2x}}{M^{1-o(1)}}\right).$$

Proof. First of all, we assume without loss of generality that $|T| < n^{2\varepsilon}$, by dividing T in several sets if this is not the case. Let us sort the vertices in T, obtaining $t_1, \ldots, t_k$, let $\boldsymbol{X}_i$ be 1 if $\boldsymbol{E}(t_i)$ holds, 0 otherwise, and let us assume that we know the structure of $\boldsymbol{N}^{\boldsymbol{\tau}_{t_j}(n^x)-2}(t_j)$ for each $j < i$ (in other words, let $\mathfrak{A}_i$ be the σ-field generated by all possible structures of $\boldsymbol{N}^{\boldsymbol{\tau}_{t_j}(n^x)-2}(t_j)$ for each $j < i$, and of the neighbor with volume at most n^y). Then, the probability that $\boldsymbol{N}^{\boldsymbol{\tau}_{t_i}(n^x)-2}(t_i)$ touches $\boldsymbol{N}^{\boldsymbol{\tau}_{t_j}(n^x)-2}(t_j)$ is at most $\sum_{\ell,\ell' < \mathcal{O}(\log n)} \mathbb{P}\left(\ell \leq \boldsymbol{\tau}_{t_j}(n^x) - 2 \wedge \ell' \leq \boldsymbol{\tau}_{t_i}(n^x) - 2 \wedge \boldsymbol{\Gamma}^\ell(t_i) \cap \boldsymbol{\Gamma}^{\ell'}(t_j) \neq \emptyset\right) \leq$

$\sum_{\ell,\ell' < \mathcal{O}(\log n)} \mathbb{P}\left(\boldsymbol{\Gamma}^\ell(t_i) \cap \boldsymbol{\Gamma}^{\ell'}(t_j) \neq \emptyset \middle| \ell \leq \boldsymbol{\tau}_{t_j}(n^x) - 2 \wedge \ell' \leq \boldsymbol{\tau}_{t_i}(n^x) - 2\right) \leq \mathcal{O}(\log^2 n) \frac{M^{2x}+M^{x+y}}{M^{1-\varepsilon}}$, because $\boldsymbol{\gamma}^{\boldsymbol{\tau}_{t_i}(n^x)-1}(t_i) < n^x$ for each i, and consequently $\boldsymbol{\delta}^{\boldsymbol{\tau}_{t_i}(n^x)-2}(t_i) < M^{x+\varepsilon}$ w.h.p., by Lemmas B.7, B.9 and B.23. As a consequence $p_i = \mathbb{P}(\boldsymbol{E}(t_i)) - \frac{M^{2x+\varepsilon}}{M} \leq \mathbb{P}(\boldsymbol{E}(t_i)|\mathfrak{A}_i) \leq \mathbb{P}(\boldsymbol{E}(t_i)) + \frac{M^{2x+\varepsilon}}{M} = q_i$.

We have proved that $\boldsymbol{S}_k = \sum_{i=1}^k \boldsymbol{X}_i - p_i, \boldsymbol{S'}_k = \sum_{i=1}^k q_i - \boldsymbol{X}_i$ are submartingales. If $p = \sum_{i=1}^k p_i$, by the strengthened version of Azuma's inequality (Lemma B.5), $\mathbb{P}(\boldsymbol{S}_k > \varepsilon kp) \leq e^{-\mathcal{O}\left(\frac{\varepsilon^2 k^2 p^2}{kp + \varepsilon kp}\right)} \leq e^{-\varepsilon^3 kp} \leq e^{-\varepsilon^3 n^\varepsilon}$. This proves that $|\{t \in T : \boldsymbol{E}(t)\}| \geq (1-\varepsilon) \sum_{t \in V} \mathbb{P}(\boldsymbol{E}_\ell(t)|\ell < \boldsymbol{\tau}_t(n^x) - 1, t \in T) + |T|\frac{M^{2x}+M^{x+y}}{M^{1-\varepsilon}}$, w.h.p.. The other inequality follows from a very similar argument applied to $\boldsymbol{S'}_k$.

COROLLARY B.12. *Let* $p = \mathbb{P}(\boldsymbol{\tau}_t(n^x)) \leq \ell | \deg(t) = d)$*, and let us assume that* $p > M^{2x+\varepsilon-1}$*. Then,* $(1-\varepsilon)p|T| \leq |\{t \in T : \boldsymbol{\tau}_s(n^x)) \leq \ell\} \leq (1+\varepsilon)p|T|$*.*

Proof. We apply Lemma B.29 with T as the set of vertices of degree d, $\boldsymbol{E}_\ell(t)$ as the event that $\ell \leq 2 + (1-\varepsilon)F(k \to n^x)$. We obtain that $|\{t \in T : \boldsymbol{\tau}_s(n^y) \leq (1-\varepsilon)F(d \to n^x)\}| = |\{t \in T : \forall \ell < \boldsymbol{\tau}_s(n^y) - 1, \ell \leq (1-\varepsilon)F(d \to n^x) - 2\}| = (1 \pm o(1))p|T| \pm |T|\frac{M^{2x+o(1)}}{M} = (1 \pm o(1))p|T|$.

B.6 Proof of Theorem 3.1.

Proof. [Proof that Property 2.1 holds] For the first statement, if $\deg(s) = n^\alpha$ with $\alpha > \varepsilon$, in the case $\beta > 2$, we know by Theorem B.1 that $\boldsymbol{\tau}_s(n^y) \leq \boldsymbol{\tau}_s(n^\alpha) + (1+\varepsilon)F(n^\alpha \to n^y) \leq 1 + (1+\varepsilon)T(n^\alpha \to n^y) \leq (1+2\varepsilon)T(n^\alpha \to n^y)$. In the case $1 < \beta < 2$, we know by B.25 that s is connected to the maximum degree vertex, which has degree $\Theta(n)$: hence, $\boldsymbol{\tau}_s(n^x) \leq 2 = T(n^\alpha \to n^x)$.

For the other statements, if x is small enough, this result follows by Corollary B.12 and Theorems B.4 and B.5. For bigger values of x, we can extend it with Theorem B.1.

Proof. [Proof that Property 2.2 holds, CM] Let us recall the definition of $\boldsymbol{\Delta}^\ell(s)$ as the set of stubs of vertices at distance ℓ from s, not paired with stubs at distance $\ell-1$. We know that $\boldsymbol{\Delta}^\ell(s) \geq \boldsymbol{\Gamma}^{\ell+1}(s)^{\max\left(\frac{1}{\beta-1}\right)}$ by Lemmas B.7 and B.23. For $\ell_s = \boldsymbol{\tau}_s(n^x) - 1$, $\ell_t = \boldsymbol{\tau}_t(n^y) - 1$, $\boldsymbol{\delta}^{\ell_s}(s) \geq \boldsymbol{\gamma}^{\ell_s+1}(s)^{\max\left(1,\frac{1}{\beta-1}\right)}n^{-\varepsilon} \geq n^{x\max\left(1,\frac{1}{\beta-1}\right)-\varepsilon} \geq M^{x-\varepsilon}$, and similarly $\boldsymbol{\delta}^{\ell_t}(t) \geq M^{y-\varepsilon}$. Consequently, $\boldsymbol{\Delta}^{\ell_s}(s)\boldsymbol{\Delta}^{\ell_t}(t) \geq M^{x+y-2\varepsilon} \geq M^{1+\varepsilon'}$. We claim that, w.h.p., a stub in $\boldsymbol{\Delta}^{\ell_s}(s)$ is paired with a stub in $\boldsymbol{\Delta}^{\ell_t}(t)$, and consequently $\mathrm{dist}(s,t) \leq \ell_s + \ell_t + 1 = \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y) - 1$, proving the theorem. To prove this claim, let us first observe that if $\boldsymbol{N}^{\ell_s}(s)$ and $\boldsymbol{N}^{\ell_t}(t)$ touch each other, then $\mathrm{dist}(s,t) \leq \ell_s + \ell_t < \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y) - 1$, and the result follows. Otherwise, let us assume without loss of generality that $x < y$ (if $x > y$, we swap the roles of s and t, if $x = y$, we can decrease x by a small amount, and we change the value of ε). Let us consider the $M^{x-\varepsilon}$ unpaired stubs $a_1, \dots, a_{M^{x-\varepsilon}}$ in $\boldsymbol{\Delta}^{\ell_t}(s)$, and let us pair these stubs one by one, by defining $\boldsymbol{X}_i = 1$ if the stub is paired to a stub in $\boldsymbol{\Delta}^{\ell_t}(t)$, 0 otherwise. Note that, conditioned on all possible pairings of a_j with $j < i$, $\mathbb{E}[\boldsymbol{X}_i] \geq \frac{M^{y-\varepsilon}-M^{x-\varepsilon}}{M} \geq \frac{M^{y-2\varepsilon}}{M}$. Hence, $\boldsymbol{S}_k = \frac{kM^{y-2\varepsilon}}{M} - \sum_{i=1}^k \boldsymbol{X}_i$ is a supermartingale, and $\mathrm{Var}\left[\boldsymbol{X}_i\right] \leq \mathbb{E}\left[\boldsymbol{X}_i^2\right] \leq \mathbb{E}\left[\boldsymbol{X}_i\right] \leq \frac{M^{y-2\varepsilon}}{M}$. By a strengthened version of Azuma's inequality (Lemma B.5), $\mathbb{P}\left(\sum_{i=1}^k \boldsymbol{X}_i = 0\right) \leq \mathbb{P}\left(k\frac{M^{y-2\varepsilon}}{M} - \sum_{i=1}^k \boldsymbol{X}_i < \varepsilon i \frac{M^{y-2\varepsilon}}{M}\right) \leq e^{\frac{-\varepsilon^2 k^2 M^{2(y-2\varepsilon)}}{\Omega\left(kM^{y-2\varepsilon}M\right)}} = e^{-\Omega\left(\frac{\varepsilon^2 kM^{y-2\varepsilon}}{M}\right)}$. For $k = M^{x-\varepsilon}$, we have proved that, w.h.p., the number of stubs in $\boldsymbol{\Delta}^{\ell_s}(s)$ that are paired with stubs in $\boldsymbol{\Delta}^{\ell_t}(t)$ is at least $(1-\varepsilon)\frac{M^{x+y-3\varepsilon}}{M} \geq 1$, and consequently $\mathrm{dist}(s,t) \leq \ell_s + \ell_t < \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y) - 1$.

Proof. [Proof that Property 2.2 holds, IRG] As in Appendix B.2, let $\boldsymbol{\Delta}^\ell(s)$ be the volume of $\boldsymbol{\Gamma}^\ell(s)$, and let $\ell_s = \boldsymbol{\tau}_s(n^x) - 1$, $\ell_t = \boldsymbol{\tau}_t(n^y) - 1$. If $\beta > 2$, by Lemma B.9, $\boldsymbol{\delta}^{\ell_s}(s) > (1-\varepsilon)M^x$, and $\boldsymbol{\delta}^{\ell_t}(t) > (1-\varepsilon)M^y$. The probability that a vertex $v \in \boldsymbol{\Gamma}^{\ell_s}(s)$ is not connected to any vertex $w \in \boldsymbol{\Gamma}^{\ell_t}(t)$ is $\prod_{w \in \boldsymbol{\Gamma}^{\ell_t}(t)} 1 - f\left(\frac{\rho_v \rho_w}{M}\right)$. We have to consider different cases separately.

- If $\sum_{v \in \boldsymbol{\Gamma}^{\ell_s}(s), \rho_v < \frac{M}{M^y}} \rho_v > \frac{Mn^\varepsilon}{M^y}$, by removing some vertices we can assume that all vertices in $\boldsymbol{\Gamma}^{\ell_s}(s)$ have weight at most $\frac{M}{M^y}$. In this case, the number of vertices $v \in \boldsymbol{\Gamma}^{\ell_s}(s)$ having a connection to $\boldsymbol{\Gamma}^{\ell_t}(t)$ is $\sum_{v \in \boldsymbol{\Gamma}^{\ell_s}(s)} \boldsymbol{X}_v$, where the $\boldsymbol{X}_v$s are independent random variables with success probability $1 - \prod_{w \in \boldsymbol{\Gamma}^{\ell_t}(t)} 1 - f\left(\frac{\rho_v \rho_w}{M}\right) = 1 - \prod_{w \in \boldsymbol{\Gamma}^{\ell_t}(t)} e^{-\Omega\left(\frac{\rho_v \rho_w}{M}\right)} = 1 - e^{-\Omega\left(\frac{\rho_v M^y}{M}\right)} = \Omega\left(\frac{\rho_v M^y}{M}\right)$. We conclude by a straightforward application of the multiplicative form of Chernoff bound (Lemma B.1).

- If we do not fall into the previous case, $\sum_{v \in \boldsymbol{\Gamma}^{\ell_s}(s), \rho_v < \frac{M}{M^y}} \rho_v < \frac{Mn^\varepsilon}{M^y}$, and by slightly decreasing x we can assume without loss of generality that all vertices in $\boldsymbol{\Gamma}^{\ell_s}(s)$ have weight at least $\frac{M}{M^y}$. By changing the roles of s and t, we can also assume that all vertices in $\boldsymbol{\Gamma}^{\ell_t}(t)$ have weight at least $\frac{M}{M^x}$. Assuming this, we still have to divide the analysis in two possible cases.

    - if $\boldsymbol{\gamma}^{\ell_s}(s)\boldsymbol{\gamma}^{\ell_t}(t) > n^\varepsilon$, the number of connections between $\boldsymbol{\Gamma}^{\ell_s}(s)$ and $\boldsymbol{\Gamma}^{\ell_t}(t)$ is at least $\sum_{v \in \boldsymbol{\Gamma}^{\ell_s}(s), w \in \boldsymbol{\Gamma}^{\ell_t}(t)} \boldsymbol{X}_{v,w}$, where the $\boldsymbol{X}_{v,w}$s are independent random variables with success probability $f\left(\frac{\rho_v \rho_w}{M}\right) = \Theta(1)$. Since the sum is made by at least n^ε terms, we can conclude by a straightforward application of the multiplicative form of Chernoff bound (Lemma B.1).

    - If $\boldsymbol{\gamma}^{\ell_s}(s)\boldsymbol{\gamma}^{\ell_t}(t) < n^\varepsilon$, there is at least a vertex $v \in \boldsymbol{\Gamma}^{\ell_s}(s)$ with weight $n^{x-\varepsilon}$, and a vertex $w \in \boldsymbol{\Gamma}^{\ell_t}(t)$ with weight $n^{y-\varepsilon}$. Then, $\mathbb{P}(E(v,w)) = f\left(\frac{\rho_v \rho_w}{M}\right) = f\left(\frac{M^{x+y-2\varepsilon}}{M}\right) \geq f(M^\varepsilon) \geq 1 - o(M^{\varepsilon k})$ for each k (we recall that, in our assumptions, $f(x) = 1 - o(x^k)$ for each k, if x tends to infinity). We conclude because this means that v is connected to w w.h.p..

Proof. [Proof that Property 2.3 holds] Let us fix $x \geq \frac{1}{2}$, let s be any vertex, and let us fix an integer ℓ_s such that $\boldsymbol{\delta}^{\ell_s}(s) < M^{x+\varepsilon}$. Let us consider a vertex $t \in W$, and let ℓ_t be an integer such that $\boldsymbol{\delta}^{\ell_t}(t) < M^{y+\varepsilon}$: if $\boldsymbol{E}\left(\boldsymbol{\delta}^{\ell_s}(s), \boldsymbol{\delta}^{\ell_t}(t)\right)$ is the event that there is an edge between $\boldsymbol{\Delta}^{\ell_s}(s)$ and $\boldsymbol{\Delta}^{\ell_t}(t)$, $\mathbb{P}\left(\boldsymbol{E}\left(\boldsymbol{\delta}^{\ell_s}(s), \boldsymbol{\delta}^{\ell_t}(t)\right)\Big|\boldsymbol{\delta}^{\ell_s}(s) < M^{x+\varepsilon}, \boldsymbol{\delta}^{\ell}(t) < M^{y+\varepsilon}\right) < \frac{M^{x+y+3\varepsilon}}{M}$. Hence,

$$P\left(\exists \ell_s, \ell_t : \boldsymbol{\delta}^{\ell_s}(s) < M^{x+\varepsilon} \wedge \boldsymbol{\delta}^{\ell_t}(t) < M^{y+\varepsilon} \wedge \boldsymbol{E}\left(\boldsymbol{\delta}^{\ell_s}(s), \boldsymbol{\delta}^{\ell_t}(t)\right)\right)$$

$$\leq \sum_{\ell_s, \ell_t = 0}^{\mathcal{O}(\log n)} \mathbb{P}\left(\boldsymbol{\delta}^{\ell_s}(s) < M^{x+\varepsilon} \wedge \boldsymbol{\delta}^{\ell_t}(t) < M^{y+\varepsilon} \wedge \boldsymbol{E}\left(\boldsymbol{\delta}^{\ell_s}(s), \boldsymbol{\delta}^{\ell_t}(t)\right)\right)$$

$$\leq \sum_{\ell_s, \ell_t = 0}^{\mathcal{O}(\log n)} \mathbb{P}\left(\boldsymbol{E}\left(\boldsymbol{\delta}^{\ell_s}(s), \boldsymbol{\delta}^{\ell_t}(t)\right)\Big|\boldsymbol{\delta}^{\ell_s}(s) < M^{x+\varepsilon}, \boldsymbol{\delta}^{\ell_t}(t) < M^{y+\varepsilon}\right)$$

$$\leq \frac{M^{x+y+4\varepsilon}}{M}.$$

This means that, with probability $1 - \frac{M^{x+y+4\varepsilon}}{M}$, $\text{dist}(s,t) \geq \ell_s + \ell_t + 2$, where ℓ_s (resp., ℓ_t) is the maximum integer such that $\boldsymbol{\delta}^{\ell_s}(s) < M^{x+\varepsilon}$ (resp., $\boldsymbol{\delta}^{\ell_t}(t) < M^{y+\varepsilon}$). By definition of ℓ_s, ℓ_t, $\boldsymbol{\delta}^{\ell_s+1}(s) > n^{x+\varepsilon}$, and by Lemmas B.7, B.9 and B.23, $\boldsymbol{\gamma}^{\ell_s+2}(s) > n^x$, meaning that $\boldsymbol{\tau}_s(n^x) \leq \ell_s + 2$, w.h.p.. Since the same holds for t, $\text{dist}(s,t) \geq \ell_s + \ell_t + 2 \geq \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y) - 2$, with probability $1 - \frac{M^{x+y+4\varepsilon}}{M}$.

We have to translate this probabilistic result into a result on the number of vertices t such that $\text{dist}(s,t) < \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y) - 2$. To this purpose, we apply Lemma B.29, by fixing s, conditioning on $\boldsymbol{N}^{\boldsymbol{\tau}_s(n^x)-2}(s)$ (which has volume at most n^x), and defining $\boldsymbol{E}(t)$ as $\text{dist}(s,t) < \boldsymbol{\tau}_s(n^x) + \boldsymbol{\tau}_t(n^y) - 2$. Since $y < x$, and $x + y < 1$, $|\{t \in T : \text{dist}(s,t) < \boldsymbol{\tau}_s(n^x)\}| \leq (1 + o(1))|T|\frac{M^{x+y+4\varepsilon-1}}{M} \pm |T|\frac{M^{x+y}+M^{2x}}{M^{1-o(1)}} \leq |T|M^{x+y+5\varepsilon-1}$.

Proof. [Proof that Property 2.4 holds] For values of d bigger than n^ε, by Lemmas B.7, B.9 and B.23, a vertex with weight d has degree $\Theta\left(d^{\max(1, n^{\frac{1}{\beta-1}})}\right)$. Hence, since the number of vertices with weight at least d is $\Theta\left(\frac{n}{d^{\beta-1}}\right)$, the conclusion follows.

For smaller values of d, a vertex with weight d has degree bigger than $\frac{1}{2}d^{\max(1, \frac{1}{\beta-1})}$ with probability $p = \mathcal{O}(1)$: through simple concentration inequalities it is possible to prove that the degree number of vertices with degree at least $\frac{1}{2}d^{\max(1, \frac{1}{\beta-1})}$ is $\mathcal{O}(|\{v \in V : \rho_v \geq d\}|) = \mathcal{O}(\frac{n}{d^{\beta-1}})$. By defining $d' = d^{\max(1, \frac{1}{\beta-1})}$, we conclude.

B.7 Other Results. Before concluding, we need to prove some lemmas that are used in some probabilistic analyses, even if they do not follow from the main theorems.

LEMMA B.30. *Assume that $\beta > 2$, and let T be the set of vertices with degree at least n^x. Then, $\text{dist}(s,T) := \min_{t \in T} \text{dist}(s,t) \leq \boldsymbol{\tau}_s\left(n^{x(\beta-2)+\varepsilon}\right) + 1$ w.h.p..*

Proof. By removing some vertices from T, we can redefine T as the set of vertices with weight at least $n^{x+\varepsilon}$ (because each vertex with weight at least $n^{x+\varepsilon}$ has degree at least n^x by Lemmas B.7, B.9 and B.23). After this modification, the number of vertices in T is $\Theta\left(\frac{n}{n^{(x+\varepsilon)(\beta-1)}}\right) = \Theta\left(n^{1-(x+\varepsilon)(\beta-1)}\right)$, and the volume of T is $\Omega\left(n^{1-(x+\varepsilon)(\beta-1)+x+\varepsilon}\right) = \Omega\left(n^{1-(x+\varepsilon)(\beta-2)}\right)$. We recall the definition of $\boldsymbol{\delta}^\ell(s)$: in the CM, it is the number of stubs at distance ℓ from s, not paired with stubs at distance $\ell-1$, while in IRG it is the volume of the set of vertices at distance ℓ from s. By Lemmas B.7, B.9 and B.23, if $\ell = \boldsymbol{\tau}_s\left(n^{(x+3\varepsilon)(\beta-2)}\right) - 1$, $\boldsymbol{\delta}^\ell(s) \geq n^{(x+2\varepsilon)(\beta-2)}$. In the CM, since the pairing of stubs is random, there is w.h.p. a stub in $\boldsymbol{\Delta}^\ell(s)$ which is paired with a stub of a vertex in T. In IRG, the probability that a vertex in $\boldsymbol{\Gamma}^\ell(s)$ is paired with a vertex in T is at least $\sum_{v \in \boldsymbol{\Gamma}^\ell(s)} \sum_{t \in T} \boldsymbol{X}_{vt}$, where the $\boldsymbol{X}_{vt}$s are Bernoulli random variables with success probability $f\left(\frac{\rho_v \rho_t}{M}\right)$. We conclude by a straightforward application of Chernoff bound (Lemma B.1).

LEMMA B.31. *Given a vertex v and an integer ℓ, assume that $n^\varepsilon < \boldsymbol{\gamma}^\ell(v) < n^{1-\varepsilon}$, and let $S = \{s \in V : n^\alpha < \deg(s) < n^{\alpha+\varepsilon}\}$, for some $\alpha > 0$. Then, $|S \cap \boldsymbol{\Gamma}^\ell(v)| \leq \boldsymbol{\gamma}^\ell(v)|S|n^{-1+\alpha+\varepsilon}$ w.h.p..*

Proof. [Proof for the CM] By Lemma B.7, we can assume that $n^\varepsilon \leq \boldsymbol{\delta}^{\ell-1}(v) \leq n^{1-\varepsilon}$. Let us sort the stubs in $\boldsymbol{\Delta}^{\ell-1}(v)$, and let $\boldsymbol{X}_i$ be 1 if the i-th stub is paired with a stub of a vertex in S, 0 otherwise. Clearly, $|S \cap \boldsymbol{\Gamma}^\ell(v)| \leq \sum_{i=1}^{\boldsymbol{\delta}^{\ell-1}(v)} \boldsymbol{X}_i$. Since $\boldsymbol{\delta}^{\ell-1}(v) < n^{1-\varepsilon}$, conditioned on the outcome of the previous variables $\boldsymbol{X}_j$, $\mathbb{P}(\boldsymbol{X}_i = 1) = \mathcal{O}(\frac{1}{n}\sum_{v \in S}\rho_v) \leq |S|n^{-1+\alpha+\varepsilon}$ (because we have already paired at most $o(n)$ stubs). Hence, $\boldsymbol{S}_k = k|S|n^{-1+\alpha+\varepsilon} - \sum_{i=0}^{k}\boldsymbol{X}_i$ is a submartingale, and if $k = \boldsymbol{\delta}^{\ell-1}(v)$, by the strenghtened version of Azuma's inequality (Lemma B.5), w.h.p., $\boldsymbol{S}_k \geq -k|S|n^{-1+\alpha+\varepsilon}$, that is, $k|S|n^{-1+\alpha+\varepsilon} - \sum_{i=0}^{k}\boldsymbol{X}_i \geq -k|S|n^{-1+\alpha+\varepsilon}$, and $|S \cap \boldsymbol{\Gamma}^\ell(v)| \leq \sum_{i=0}^{k}\boldsymbol{X}_i \leq 2k|S|n^{-1+\alpha+\varepsilon}$. The result follows.

Proof. [Proof for IRG] By Lemma B.9, we can assume that $n^\varepsilon \leq \boldsymbol{\delta}^{\ell-1}(v) \leq n^{1-\varepsilon}$. The probability that a vertex $s \in S$ is not linked to any vertex in $\boldsymbol{\Gamma}^{\ell-1}(v)$ is $\prod_{w \in \boldsymbol{\Gamma}^{\ell-1}(v)}\left(1 - f\left(\frac{\rho_w \rho_s}{n}\right)\right) = \prod_{w \in \boldsymbol{\Gamma}^{\ell-1}(v)}\left(1 - \mathcal{O}\left(\frac{\rho_w \rho_s}{n}\right)\right) \leq \prod_{w \in \boldsymbol{\Gamma}^{\ell-1}(v)} e^{-\mathcal{O}\left(\frac{\rho_w \rho_s}{n}\right)} = e^{-\mathcal{O}\left(\frac{\boldsymbol{\delta}^{\ell-1}(v)\rho_s}{n}\right)} = e^{-\mathcal{O}\left(\frac{\boldsymbol{\delta}^{\ell-1}(v)n^{\alpha+\varepsilon}}{n}\right)}$. If $\boldsymbol{\delta}^{\ell-1}(v) > n^{1-\alpha-2\varepsilon}$, the result of the lemma is trivial, if we change the value of ε. If $\boldsymbol{\delta}^{\ell-1}(v) < n^{1-\alpha-2\varepsilon}$, the probability that a vertex in S is not linked to any vertex in $\boldsymbol{\delta}^{\ell-1}(s)$ is $e^{-\mathcal{O}\left(\frac{\boldsymbol{\delta}^{\ell-1}(s)n^{\alpha+\varepsilon}}{n}\right)} = 1 - \mathcal{O}\left(\frac{\boldsymbol{\delta}^{\ell-1}(s)n^{\alpha+\varepsilon}}{n}\right)$, and hence the probability that it is connected to a vertex in $\boldsymbol{\delta}^{\ell-1}(s)$ is $\mathcal{O}\left(\frac{\boldsymbol{\delta}^{\ell-1}(s)n^{\alpha+\varepsilon}}{n}\right)$. By a straightforward application of Chernoff bound, the number of vertices in S that belong to $\boldsymbol{\Gamma}^\ell(s)$ is $\mathcal{O}\left(\frac{|S|\boldsymbol{\delta}^{\ell-1}(s)n^{\alpha+\varepsilon}}{n}\right)$, w.h.p..

LEMMA B.32. *Assume that $\beta > 3$, and let v a vertex with degree $\omega(1)$. Let S be the set of vertices with degree between n^α and $n^{\alpha+\varepsilon}$. Then, the number of pairs of vertices $s, t \in S$ such that $\mathrm{dist}(s,v) + \mathrm{dist}(v,t) \leq c\log_{M_1(\mu)} n$, and $\mathrm{dist}(s,w) + \mathrm{dist}(w,t) > c\log_{M_1(\mu)} n$ for each w such that $\deg(w) > \deg(v)$ is at most $\deg(v)^2|S|^2 n^{-2+c+2\alpha+\varepsilon}$.*

Proof. First of all, we want to assume without loss of generality that v is the vertex with maximum degree. To this purpose, we remove from the graph all vertices with degree bigger than $\deg(v)$: by Lemma B.19, we obtain a new random graph G', and the value of $M_1(\mu)$ changes by $o(1)$. Furthermore, v is the vertex with maximum degree in the new graph, and the shortest paths not passing from vertices with degree bigger than $\deg(v)$ are conserved.

Then, let us assume that v is the vertex with maximum degree. By Corollaries B.2 and B.5, $\boldsymbol{\gamma}^i(v) \leq n^\varepsilon \deg(v)(M_1(\mu) + \varepsilon)^i$, and by Lemma B.31, $|S \cap \boldsymbol{\Gamma}^i(v)| \leq \boldsymbol{\gamma}^i(v)|S|n^{-1+\alpha+\varepsilon} \leq \deg(v)|S|(M_1(\mu) + \varepsilon)^i n^{-1+\alpha+2\varepsilon}$. We conclude that the number of pairs $(s,t) \in S^2$ such that $\mathrm{dist}(s,v) + \mathrm{dist}(v,t) \leq c\log_{M_1(\mu)} n$ is at most:

$$\sum_{i+j=c\log_{M_1(\mu)} n} |\{s : \mathrm{dist}(s,v) \leq i\}||\{t : \mathrm{dist}(t,v) \leq c\log_{M_1(\mu)} n - i\}|$$

$$\leq \sum_{i+j=c\log_{M_1(\mu)} n} i\deg(v)|S|(M_1(\mu) + \varepsilon)^i n^{-1+\alpha+2\varepsilon} \cdot j\deg(v)|S|(M_1(\mu) + \varepsilon)^j n^{-1+\alpha+2\varepsilon}$$

$$\leq \sum_{i+j=c\log_{M_1(\mu)} n} \deg(v)^2|S|^2(M_1(\mu) + \varepsilon)^{i+j} n^{-2+2\alpha+5\varepsilon}$$

$$\leq \deg(v)^2|S|^2 n^{-2+c+2\alpha+\varepsilon'}.$$

C   The BCM Algorithm.

The BCM algorithm [15, 7] exactly computes the k most central vertices according to closeness centrality. We recall that the farness $f(s)$ of a vertex s is defined as $\sum_{t \in V}\mathrm{dist}(s,t)$; the closeness centrality of s is $c(s) = \frac{1}{f(s)} = \frac{1}{\sum_{t \in V}\mathrm{dist}(s,t)}$. Intuitively, a vertex with high closeness centrality needs a few step to "talk" to all other vertices, and consequently it is considered central.

The textbook algorithm that computes the k most central vertices simply computes the farness $f(s)$ of each vertex s through n BFSes, and it returns the k vertices with smallest $f(s)$ values. The running time is $\mathcal{O}(mn)$. The improvement proposed by the BCM algorithm stops the BFS from a vertex s as soon as we can guarantee that s is not in the top-k. To this purpose, assume that the k-th smallest farness found until now is f_k, and we

have visited all vertices up to distance ℓ. Then, we can lower bound the farness of s by setting distance $\ell + 1$ to all unvisited vertices, and if this lower bound is bigger than f_k, then we can safely interrupt the BFS.

A further speed-up can be obtained by computing a better bound: we set distance $\ell + 1$ to a number of vertices equal to $\sum_{t \in \mathbf{\Gamma}^\ell(s)} \deg(t) - 1$ (which is the number of edges exiting from level ℓ of the BFS tree), and distance $\ell + 2$ to all other vertices. In formula, we stop a visit from a vertex s after ℓ steps if:

$$\tilde{f}_\ell(s) := \sum_{t \in \mathbf{N}^\ell(s)} \operatorname{dist}(s,t) + (\ell+1)\gamma_U^\ell(s) + (\ell+2)\left(n - |\mathbf{N}^\ell(s)| - \gamma_U^\ell(s)\right) \geq f_k$$

where $\mathbf{N}^\ell(s)$ is the set of vertices at distance at most ℓ from s, $\mathbf{\Gamma}^\ell(s)$ is the set of vertices at distance exactly ℓ from s, $\gamma_U^\ell(s) = \sum_{v \in \mathbf{\Gamma}^\ell(s)} \deg(v) - 1$.

It remains to choose the order in which vertices are processed. In order to speed-up the computation as much as possible, the algorithm processes vertices in decreasing order of degree, so that we obtain quite soon high values of f_k.

In order to analyze this algorithm, we first provide a deterministic bound on the running time. The idea behind these bounds is that a BFS from s visits all vertices at distance at most $f_k - 2$, then it might find a lower bound bigger than f_k. In particular, if the number of vertices visited at distance at most $f_k - 2$ is much smaller than n, the bound is likely to be sufficient to stop the visit. Otherwise, the BFS from s has already visited $\Theta(n)$ vertices: in both cases, the number of visited vertices is close to the number of vertices at distance at most $f_k - 2$ from s.

LEMMA C.1. *Let f_k be the k-th smallest farness among the k vertices with highest degree, and let $\ell \geq (1+\alpha)\frac{f_k}{n} - 2$ be an integer. Then, the running time of the algorithm is at most $\mathcal{O}\left(m + \frac{1}{\alpha}\sum_{s \in V}\sum_{v \in \mathbf{N}^\ell(s)} \deg(v)\right)$.*

Proof. The first k BFSes need time $\mathcal{O}(m)$, because they cannot be cut. For all subsequent BFSes, the k-th smallest farness found is at least f_k. Let us consider a BFS from s, and let us assume that we have visited all vertices up to distance ℓ: our lower bound on the farness of s is $\sum_{v \in \mathbf{N}^\ell(s)} \operatorname{dist}(s,v) + (\ell+1)\gamma_U^\ell(s) + (\ell+2)\left(n - |\mathbf{N}^\ell(s)| - \gamma_U^\ell(s)\right) \geq (\ell+2)\left(n - |\mathbf{N}^\ell(s)| - \gamma_U^\ell(s)\right) \geq (\ell+2)\left(n - \sum_{v \in \mathbf{N}^\ell(s)} \deg(v)\right) \geq (1+\alpha)\frac{f_k}{n}\left(n - \sum_{v \in \mathbf{N}^\ell(s)} \deg(v)\right)$. We claim that the BFS from s visits at most $\frac{m}{n\alpha}\sum_{v \in \mathbf{N}^\ell(s)} \deg(v)$ edges: this is trivially true if $\sum_{v \in \mathbf{N}^\ell(s)} \deg(v) > \frac{\alpha}{2}n$, because the value becomes $\mathcal{O}(m)$, while if $\sum_{v \in \mathbf{N}^\ell(s)} \deg(v) < \frac{\alpha}{2}n$, the lower bound is at least $(1+\alpha)\frac{f_k}{n}\left(1 - \frac{\alpha}{2}\right)n \geq f_k$, and the BFS is stopped after ℓ steps.

LEMMA C.2. *Let f_k be the k-th smallest farness, and let $\ell \leq \frac{f_k}{n} - 2$. Then, the running time of the algorithm is $\Omega\left(\sum_{s \in V} \mathbf{n}^\ell(s)\right)$.*

Proof. Clearly, at any step, the k-th minimum farness found is at least f_k. We want to prove that the BFS from s reaches level ℓ: indeed, the lower bound on the farness of s is $\sum_{v \in \mathbf{N}^\ell(s)} \operatorname{dist}(s,v) + (\ell+1)\gamma_U^\ell(v) + (\ell+2)\left(n - |\mathbf{N}^\ell(s)| - \gamma_U^\ell(s)\right) \leq n(\ell+2) \leq f_k$. Hence, the BFS is not cut until level ℓ, and at least $\Omega\left(\sum_{s \in V} \mathbf{n}^\ell(s)\right)$ vertices are visited.

We now need to compute these values in graphs in our framework. We analyze separately the running time in the case $1 < \beta < 2$, in the case $2 < \beta < 3$, and in the case $\beta > 3$.

C.1   The Case $1 < \beta < 2$. By Property 2.4, if s is one of the k vertices with maximum degree, then $\deg(s) > 2cn$ for some c only depending on k, and by Theorem 5.3, the farness of s is at most $(2 + o(1))n - \deg(s) \leq (2 - c)n$ if n is big enough. By Lemma C.1 applied with $\ell = 0$ and $\alpha = c$, the running time is at most $\mathcal{O}\left(m + \frac{1}{c}\sum_{s \in V}\sum_{t \in \mathbf{N}^0(s)} \deg(t)\right) = \mathcal{O}\left(m + \sum_{s \in V} \deg(s)\right) = \mathcal{O}(m)$.

C.2   The Case $2 < \beta < 3$. In this case, we know by Theorem 5.4 applied with $x = \frac{1}{2}$ that the minimum farness is at least $n(1+o(1))\left(\boldsymbol{\tau}_s(n^x) - T(1 \to n^x) + \operatorname{dist}_{\text{avg}}(n) - 1\right) = n\left(\frac{1}{2} + o(1)\right)\operatorname{dist}_{\text{avg}}(n) = (1+o(1))\log_{\frac{1}{\beta-2}}\log n = \Theta(\log\log n)$.

We claim that all vertices with degree at least n^x are at distance $\mathcal{O}(1)$ from each other: indeed, if $\deg(s), \deg(t) > n^x$, by Property 2.1, $\boldsymbol{\tau}_s\left(n^{\frac{2}{3}}\right), \boldsymbol{\tau}_t\left(n^{\frac{2}{3}}\right) = \mathcal{O}(1)$, and by Property 2.2, $\mathrm{dist}(s,t) < \boldsymbol{\tau}_s\left(n^{\frac{2}{3}}\right) + \boldsymbol{\tau}_t\left(n^{\frac{2}{3}}\right) = \mathcal{O}(1)$.

By Lemma C.2, if $\ell + 2$ is smaller than f_k, the running time is $\Omega\left(\sum_{s \in V} \boldsymbol{n}^\ell(v)\right) = \Omega\left(\sum_{\deg(s) > n^x} \boldsymbol{n}^{\Theta(\log \log n)}(s)\right) = \Omega\left(\sum_{\deg(s) > n^x} |\{t : \deg(t) > n^x\}|\right) \geq \Omega\left(\left(n^{1-x(\beta-1)}\right)^2\right) = \Omega\left(n^{2-2x(\beta-1)}\right)$. If we choose $x = \varepsilon$, the running time is $\Omega\left(n^{2-\mathcal{O}(\varepsilon)}\right)$.

C.3 The Case $\beta > 3$. Let us estimate the farness of the k vertices with highest degree. By Property 2.4, the k maximum degrees are $\Theta\left(n^{\frac{1}{\beta-1}}\right)$, and their farness is $n(1 + \mathcal{O}(\varepsilon))\left(\boldsymbol{\tau}_s\left(n^{\frac{1}{\beta-1}}\right) - T\left(1 \to n^{\frac{1}{\beta-1}}\right) + \mathrm{dist}_{\mathrm{avg}}(n)\right) \leq n(1 + \mathcal{O}(\varepsilon)) \log_{M_1(\mu)}\left(n^{1-\frac{1}{\beta-1}}\right)$. Hence, by Lemma C.2 applied with $\ell = (1 + \mathcal{O}(\varepsilon)) \log_{M_1(\mu)}\left(n^{1-\frac{1}{\beta-1}}\right)$, we obtain that the running time is $\mathcal{O}\left(\sum_{s \in V} \sum_{v \in \boldsymbol{N}^\ell(s)} \deg(v)\right) = \mathcal{O}\left(\sum_{s \in V} \sum_{v \in \boldsymbol{N}^\ell(s)} \rho_v\right) = \mathcal{O}\left(n^\varepsilon \sum_{s \in V} \boldsymbol{n}^{\ell+1}(v)\right)$, because $\deg(v) \leq n^\varepsilon \rho_v$, and $\boldsymbol{\gamma}^{\ell+1}(v) > \rho_{\boldsymbol{\Gamma}^\ell(v)}$ in random graphs, as shown in Lemmas B.7 and B.9. For the lower bound, we use two partial results that we obtain in the proof of the main theorems: Corollaries B.3 and B.6. These corollaries say that $\boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \geq (1 - \varepsilon)(y - x) \log_{M_1(\mu)} n$, for each $\frac{1}{\beta-1} < x < y < 1$. Hence, for each vertex s, $f(s) \geq \left(n - n^{1-\varepsilon}\right) \boldsymbol{\tau}_s\left(n^{1-\varepsilon}\right) \geq \left(1 - \frac{1}{\beta-1} - \mathcal{O}(\varepsilon)\right) \log_{M_1(\mu)} n \geq (1 - \mathcal{O}(\varepsilon)) \log_{M_1(\mu)}\left(n^{1-\frac{1}{\beta-1}}\right)$.

We conclude that, by Lemmas C.1 and C.2, the running time of the algorithm is $\sum_{s \in V} \boldsymbol{n}^\ell(s)$, where $\ell = (1 \pm \mathcal{O}(\varepsilon)) \log_{M_1(\mu)}\left(n^{1-\frac{1}{\beta-1}}\right)$. To estimate this value, we use the following result, that we prove in Appendix B (Theorem B.1).

THEOREM C.1. *For each $0 < x < y < 1$, $\boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \geq (1 - \varepsilon) \log_{M_1(\mu)} n^{y-x}$ a.a.s.. Moreover, $\boldsymbol{\tau}_s(n^y) - \boldsymbol{\tau}_s(n^x) \leq (1 + \varepsilon) \log_{M_1(\mu)} n^{y-x}$ w.h.p..*

Thanks to this result, for each vertex s, $\boldsymbol{\tau}_s(n^y) \leq \boldsymbol{\tau}_s(n^\varepsilon \deg(s)) + (1 + \varepsilon) \log_{M_1(\mu)} n^{y-x} \leq (1 + \mathcal{O}(\varepsilon)) \log_{M_1(\mu)} \frac{n^y}{\deg(s)} \leq \ell$ if $y = 1 - \frac{1}{\beta-1} + \frac{\log \deg(s)}{\log n} + \mathcal{O}(\varepsilon)$, and consequently $\boldsymbol{\gamma}^i(s) < n^y$ for each $i < \ell$. Hence, the running time is smaller than $n^{\mathcal{O}(\varepsilon)} \sum_{s \in V} \boldsymbol{n}^\ell(s) \leq n^{\mathcal{O}(\varepsilon)} \sum_{s \in V} \ell n^y \leq n^{1-\frac{1}{\beta-1}+\mathcal{O}(\varepsilon)} \sum_{s \in V} \deg(s) \leq n^{2-\frac{1}{\beta-1}+\mathcal{O}(\varepsilon)}$. An analogous argument proves that the running time is at least $n^{2-\frac{1}{\beta-1}-\mathcal{O}(\varepsilon)}$.

D Distance Oracle.

In this section, we analyze the performances of the distance oracle in [4]. Basically, this distance oracle assigns a label $L(s)$ to each vertex s: each label is a set of pairs $(v, \mathrm{dist}(s, v))$, where v is a vertex in the graph, and $\mathrm{dist}(s, v)$ is the distance between s and v. The construction of these labels enforces the so-called 2-hop cover property: for each pair of vertices s, t, there is a vertex v in a shortest path between s and t that belongs both to $L(s)$ and to $L(t)$. Using the 2-hop cover property, it is possible to compute $\mathrm{dist}(s, t) = \min_{v \in L(s) \cap L(t)} \mathrm{dist}(s, v) + \mathrm{dist}(v, t)$, in time $\mathcal{O}(|L(s)| + |L(t)|)$. The space required is $\Theta\left(\sum_{s \in V} |L(s)|\right)$.

In order to compute a set of labels that satisfies the 2-hop cover property, we sort all vertices s obtaining $s_1, \ldots, s_n$ (any order is fine), and we perform a BFS from each vertex, following this order. During the BFS from s_i, as soon as we visit a vertex t, we add $(s_i, \mathrm{dist}(s_i, t))$ to the label of t. Furthermore, we prune each BFS at each vertex t such that $\mathrm{dist}(s, t) = \min_{v \in L(s) \cap L(t)} \mathrm{dist}(s, v) + \mathrm{dist}(v, t)$, where $L(s)$ and $L(t)$ are the current labels. It is proved in [4] that the labels generated by the algorithm satisfy the 2-hop cover property, and that s_i is in the label of t if and only if there is no vertex s_j for some $j < i$ that belongs to an (s, t)-shortest path. It remains to define how vertices are sorted: in [4], it is suggested to sort them in order of degree (tie-breaks are solved arbitrarily).

First, as we did in the previous analyses, we compute a deterministic bound on the expected time of a distance query between two random vertices.

LEMMA D.1. *For each s_i, let N_{s_i} be the number of vertices $t \in V$ such that no (s, t)-shortest path passes from a vertex s_j with $j < i$. Then, the average query time is $\mathcal{O}\left(\frac{1}{n} \sum_{s \in V} N(s)\right)$, and the space used is $\Theta\left(\sum_{s \in V} N(s)\right)$.*

Proof. If the labels are sorted, in order to intersect $L(s)$ and $L(t)$, we need time $\mathcal{O}\left(|L(s)| + |L(t)|\right)$. Hence, the expected time of a distance query between two random vertices is $\frac{1}{n^2}\sum_{s,t\in V}\mathcal{O}\left(|L(s)| + |L(t)|\right) = \mathcal{O}\left(\frac{1}{n}\sum_{t\in V}|L(t)|\right) = \mathcal{O}\left(\frac{1}{n}\sum_{s,t\in V}X_{st}\right) = \mathcal{O}\left(\frac{1}{n}\sum_{s\in V}N(s)\right)$, where $X_{st} = 1$ if $s \in L(t)$, 0 otherwise. Similarly, the space used is $\Theta\left(\frac{1}{n}\sum_{t\in V}|L(t)|\right) = \Theta\left(\sum_{s\in V}N(s)\right)$.

D.1 The Case $1 < \beta < 2$. Let us fix $\varepsilon > 0$, and let us consider vertices s with small degree (at most, $n^{2\varepsilon}$): the number of vertices reachable from s at distance k passing only through vertices of degree smaller than $\deg(s)$ is at most $\deg(s)^k$: hence, $N(s) \leq D\deg(s)^D \leq Dn^{2\varepsilon D} = n^{\mathcal{O}(\varepsilon)}$ (because the diameter D is constant).

Let us consider vertices s such that $\deg(s) > n^{2\varepsilon}$: by Property 2.2, all these vertices are connected to each vertex with degree at least $n^{1-\varepsilon}$, and no vertex t with degree bigger than n^ε can contain s in their label, unless t is a neighbor of s. Consequently, the vertices that contain s in their label are at most $\deg(s)$ vertices at distance 1 from s, $N_2(s)$ vertices at distance 2 from s, and at most $\left(n^{2\varepsilon}\right)^D N_2(s) = n^{\mathcal{O}(\varepsilon)}N_2(s)$ vertices at a bigger distance. Summing these values, we obtain that $N(s) \leq \deg(s) + N_2(s)n^{\mathcal{O}(\varepsilon)}$: summing over all vertices s, the average query time is $\frac{1}{n}\mathcal{O}\left(n^{1+\varepsilon} + \sum_{s\in V, \deg(s)>n^{2\varepsilon}}\deg(s) + N_2(s)n^{2\varepsilon}\right) = n^{\mathcal{O}(\varepsilon)}\left(1 + \frac{1}{n}\sum_{s\in V}N_2(s)\right)$. Since $t \in N_2(s)$ implies that $\deg(t) < n^{2\varepsilon}$, and since the number of vertices with degree bigger than n^x is $n^{1-x+o(1)}$ by Property 2.4,

$$\sum_{s\in V}N_2(s) = \sum_{s\in V}\sum_{v\in\mathbf{\Gamma}^1(s),\deg(v)<\deg(s)}\sum_{t\in\mathbf{\Gamma}^1(v),\deg(t)<n^{2\varepsilon}}1$$

$$= \sum_{v\in V}\left|\{t \in \mathbf{\Gamma}^1(v) : \deg(t) < n^{2\varepsilon}\}\right|\left|\{s \in \mathbf{\Gamma}^1(v) : \deg(s) > \deg(v)\}\right|$$

$$\leq \sum_{v\in V}\deg(v)\max\left(\deg(v), \frac{n^{1+\varepsilon}}{\deg(v)}\right)$$

$$= \sum_{d=1}^{n}|\{v : \deg(v) = d\}|\,d\max\left(d, \frac{n^{1+\varepsilon}}{d}\right)$$

$$\leq \sum_{d=1}^{n^{\frac{1}{2}}}|\{v : \deg(v) = d\}|\,d^2 + \sum_{d=n^{\frac{1}{2}}}^{n}|\{v : \deg(v) = d\}|\,n^{1+\varepsilon}$$

$$\leq S_1 + n^{\frac{3}{2}+\varepsilon}.$$

Let us estimate S_1 using Abel's summation technique:

$$S_1 = \sum_{d=1}^{n^{\frac{1}{2}}}|\{v : \deg(v) = d\}|\,d^2$$

$$= \sum_{d=1}^{n^{\frac{1}{2}}}|\{v : \deg(v) \geq d\}|\,d^2 - \sum_{d=1}^{n^{\frac{1}{2}}}|\{v : \deg(v) \geq d+1\}|\,d^2$$

$$\leq n + \sum_{d=1}^{n^{\frac{1}{2}}}|\{v : \deg(v) \geq d\}|\left(d^2 - (d-1)^2\right)$$

$$\leq n + \sum_{d=1}^{n^{\frac{1}{2}}}\frac{n}{d}2d$$

$$= \mathcal{O}\left(n^{\frac{3}{2}}\right).$$

Then, the average query time is $n^{\mathcal{O}(\varepsilon)}\left(1 + \frac{1}{n}\sum_{s\in V}N_2(s)\right) = n^{\frac{1}{2}+\mathcal{O}(\varepsilon)}$. The space occupied is n multiplied by the average query time, that is, $n^{\frac{3}{2}+\mathcal{O}(\varepsilon)}$, by Lemma D.1.

D.2 The Case $2 < \beta < 3$. Let us consider the number $N_\ell(s)$ of vertices t such that $s \in L(t)$ and $\text{dist}(s,t) = \ell$. For small values of ℓ, we estimate $N_\ell(s) \leq N_{\ell-1}(s)\deg(s)$, while for bigger values of ℓ, we prove that all vertices with high degree are at distance at most $\ell - 2$ from s, obtaining that $N_\ell(s) \leq f(\ell, \deg(s))N_{k-1}(s)$, for some function f.

More formally, let s, t be two vertices with degree at least $\log^2 n$: the distance between s and t is at most $\boldsymbol{\tau}_s\left(n^{\frac{1+\varepsilon}{2}}\right) + \boldsymbol{\tau}_t\left(n^{\frac{1+\varepsilon}{2}}\right) \leq (1+\varepsilon)\left(\log_{\frac{1}{\beta-2}}\left(\frac{\log^2 n^{\frac{1+\varepsilon}{2}}}{\log\deg(s)\log\deg(t)}\right)\right)$ (our properties say that this approximation holds for $\deg(s), \deg(t) > n^\varepsilon$, but it is easy to extend the results in Appendix B.2 to the case $\deg(s) > \log^2 n$). Consequently there is no vertex at distance k from s with degree bigger than $\max\left(\log^2 n, e^{(\beta-2)^{\frac{k}{1+\varepsilon}}\frac{\log^2 n^{\frac{1+\varepsilon}{2}}}{\log(\deg(s))}}\right)$.

As we said before, for $k < k_0 = (1+\varepsilon)\left(\log_{\frac{1}{\beta-2}}\left(\frac{\log^2 n^{\frac{1+\varepsilon}{2}}}{\log^2 \deg(s)}\right)\right)$, we estimate $N_{k+1}(s) \leq N_k(s)\deg(s)$, and consequently $N_k(s) \leq \deg(s)^k \leq \deg(s)^{(1+\varepsilon)\left(\log_{\frac{1}{\beta-2}}\left(\frac{\log^2 n^{\frac{1+\varepsilon}{2}}}{\log^2 \deg(s)}\right)\right)}$. For bigger values of k, since $\boldsymbol{\Gamma}^{k+1}(v)$ does not contain any vertex with degree bigger than $\max\left(\log^2 n, e^{(\beta-2)^{\frac{k}{1+\varepsilon}}\frac{\log^2 n^{\frac{1+\varepsilon}{2}}}{\log(\deg(s))}} + \log^2 n\right)$, $N_{k+1}(S) \leq N_k(S)\max\left(\log^2 n, e^{(\beta-2)^{\frac{k-1}{1+\varepsilon}}\frac{\log^2 n^{\frac{1+\varepsilon}{2}}}{\log(\deg(s))}}\right)$. As a consequence, we can prove by induction that, if $k = \mathcal{O}(\log\log n)$:

$$
\begin{aligned}
N_k(s) &\leq N_{k_0}(s)\prod_{i=k_0}^{k} e^{(\beta-2)^{\frac{i-1}{1+\varepsilon}}\frac{\log^2 n^{\frac{1+\varepsilon}{2}}}{\log^2 \deg(s)}}\log^{2k} n \\
&\leq N_{k_0}(s)n^\varepsilon \prod_{i=k_0}^{k} e^{(\beta-2)^{\frac{i-k_0}{1+\varepsilon}}(\beta-2)^{\frac{k_0-1}{1+\varepsilon}}\frac{\log^2 n^{\frac{1+\varepsilon}{2}}}{\log(\deg(s))}} \\
&\leq N_{k_0}(s)n^\varepsilon e^{\sum_{i=k_0}^{k}(\beta-2)^{\frac{i-k_0}{1+\varepsilon}}\log\deg(s)} \\
&\leq N_{k_0}(s)n^\varepsilon \deg(s)^{\frac{1}{1-(\beta-2)^{\frac{1}{1+\varepsilon}}}}. \\
&\leq n^{2\varepsilon}\deg(s)^{(2+2\varepsilon)\left(\log_{\frac{1}{\beta-2}}\left(\frac{\log n^{\frac{1+\varepsilon}{2}}}{\log\deg(s)}\right)\right)+\frac{1}{3-\beta}} = f(\deg(s)).
\end{aligned}
$$

For bigger values of k, there are very few vertices at distance k from s, and their contribution is negligible. Hence, we know that $N(s) \leq \min(n, n^\varepsilon f(\deg(s))) = g(\deg(s))$: we want to compute $\sum_{s \in V} g(\deg(s))$.

LEMMA D.2. *If G is a random graph with power law degree distribution,*

$$
\sum_{s \in V, \deg(s) > \log^2 n} g(d) = \mathcal{O}\left(n\int_{\log^2 n}^{n^{\frac{1}{\beta-1}+\varepsilon}} \frac{g(x)}{x^\beta}dx + n^{1+\varepsilon}\right).
$$

Proof. We use Abel's summation technique twice:

$$\sum_{s\in V,\deg(s)>\log^2 n} g(d) = \sum_{d=\log^2 n}^{n^{\frac{1+\varepsilon}{\beta-1}}} |\{s\in V:\deg(s)=d\}|\,g(d)$$

$$= \sum_{d=\log^2 n}^{n^{\frac{1+\varepsilon}{\beta-1}}} \left(|\{s\in V:\deg(s)\geq d\}| - |\{s\in V:\deg(s)\geq d+1\}|\right)g(d)$$

$$\leq \sum_{d=\log^2 n}^{n^{\frac{1+\varepsilon}{\beta-1}}} \left(|\{s\in V:\deg(s)\geq d\}| - |\{s\in V:\deg(s)\geq d+1\}|\right)g(d)$$

$$\leq \sum_{d=\log^2 n}^{n^{\frac{1+\varepsilon}{\beta-1}}} \frac{n}{d^{\beta-1}}(g(d)-g(d-1)) + \mathcal{O}(ng(\log^2 n))$$

$$\leq \mathcal{O}(ng(\log^2 n)) + \sum_{d=\log^2 n}^{n^{\frac{1+\varepsilon}{\beta-1}}} ng(d)\left(\frac{1}{d^{\beta-1}} - \frac{1}{(d+1)^{\beta-1}}\right)$$

$$\leq \mathcal{O}(ng(\log^2 n)) + \sum_{d=\log^2 n}^{n^{\frac{1+\varepsilon}{\beta-1}}} ng(d)\left(\frac{1}{d^{\beta-1}} - \frac{1}{(d+1)^{\beta-1}}\right)$$

$$= \mathcal{O}\left(n^{1+\varepsilon} + \sum_{d=\log^2 n}^{n^{\frac{1+\varepsilon}{\beta-1}}} \frac{ng(d)}{d^{\beta}}\right).$$

We have to transform this sum into an integral: to this purpose, we observe that $g(d+\varepsilon) = \mathcal{O}(g(d))$ for each $d > \log^2 n - 1$, and for each $\varepsilon \leq 1$. Hence, $\sum_{d=\log^2 n}^{n^{\frac{1+\varepsilon}{\beta-1}}} \frac{g(d)}{d^\beta} = \int_{\log^2 n-1}^{n^{\frac{1}{\beta-1}+\varepsilon}} \frac{g(\lceil x\rceil)}{\lceil x\rceil^\beta}dx = \mathcal{O}\left(\int_{\log^2 n-1}^{n^{\frac{1}{\beta-1}+\varepsilon}} \frac{g(x)}{x^\beta}dx\right).$

It remains to estimate this integral.

$$n\int_{\log^2 n}^{n^{\frac{1}{\beta-1}}} \frac{1}{x^\beta}\min\left(n, x^{(2+2\varepsilon)\log_{\frac{1}{\beta-2}}\left(\frac{\log n^{\frac{1+\varepsilon}{2}}}{\log x}\right)+\frac{1}{3-\beta}}\right)dx = \left(\frac{\log x}{\log n}=t\right)$$

$$\leq n^{1+\mathcal{O}(\varepsilon)}\int_0^{\frac{1}{\beta-1}} n^{-\beta t}\min\left(n, n^{t\left(2\log_{\frac{1}{\beta-2}}\left(\frac{1}{2t}\right)+\frac{1}{3-\beta}\right)}\right)n^t\log n\,dt$$

$$= n^{1+\mathcal{O}(\varepsilon)}\int_0^{\frac{1}{\beta-1}} n^{t\left(\min\left(1,2\log_{\frac{1}{\beta-2}}\left(\frac{1}{2t}\right)+\frac{1}{3-\beta}\right)-\beta+1\right)}dt$$

$$= n^{1+\mathcal{O}(\varepsilon)+\max t\in\left[0,\frac{1}{\beta-1}\right]t\left(\min\left(1,2\log_{\frac{1}{\beta-2}}\left(\frac{1}{2t}\right)+\frac{1}{3-\beta}\right)-\beta+1\right)}.$$

Then, the average query time is $n^{\max_{t\in\left[0,\frac{1}{\beta-1}\right]}t\left(\min\left(1,2\log_{\frac{1}{\beta-2}}\left(\frac{1}{2t}\right)+\frac{1}{3-\beta}\right)-\beta+1\right)+\mathcal{O}(\varepsilon)}$. We are not able to find an analytic form for this function, but the result is plotted in Figure 5. The exponent of the total space occupancy is this function, plus one.

D.3   The Case $\beta > 3$. In this case, we prove that the algorithm does not provide a significant improvement: indeed, the time needed for a distance query is $n^{1-\mathcal{O}(\varepsilon)}$. To prove this result, it is enough to show that, for a set of graphs that satisfy the four properties, the algorithm is not efficient on this set of graph. The set of graphs we
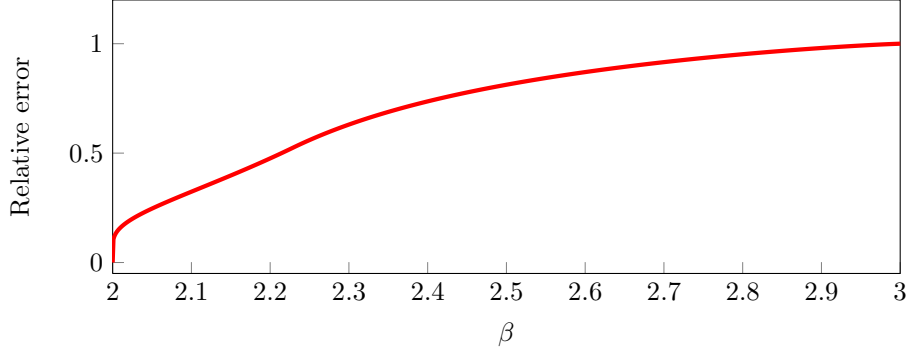
Figure 5: an upper bound on the exponent in the average distance query time of the distance oracle for $2 < \beta < 3$.

choose is the set of random graphs generated through the CM, or through IRG: for this reason, we are allowed to use theorems that are specific of these models.

In this proof, we show that, if S is the set of vertices with degree between n^α and $n^{\alpha+\varepsilon^2}$, then the number of pairs $(s,t) \in S^2$ such that $s \in L(t)$ is $\Omega\left(|S|^2\right)$, and hence the average label size is big, because, by Property 2.4, $|S| = n^{1-\alpha(\beta-1)+o(1)} = n^{1-o(1)}$ if α tends to 0.

More formally, by Properties 2.1 and 2.2, for each pair of vertices $s,t \in S$, $\mathrm{dist}(s,t) \leq \boldsymbol{\tau}_s\left(n^{\frac{1}{2}+\varepsilon}\right) + \boldsymbol{\tau}_t\left(n^{\frac{1}{2}+\varepsilon}\right) \leq (2+2\varepsilon)T\left(n^\alpha \to n^{\frac{1}{2}+\varepsilon}\right) \leq (2+2\varepsilon)\frac{\log n^{\frac{1}{2}+\varepsilon-\alpha}}{\log M_1(\mu)} \leq (1-2\alpha+4\varepsilon)\frac{\log n}{\log M_1(\mu)} =: D_S$. We want to prove that, a.a.s., if $s,t \in S$ and $\deg(t) < \deg(s)$, there is a high chance that $s \in L(t)$. To this purpose, we consider all vertices v with degree bigger than $\deg(s)$, and we count the number of pairs $s,t \in S$ such that $\mathrm{dist}(s,v) + \mathrm{dist}(v,t) \leq D_S$. Then, we sum this contribution over all vertices v: if this sum is $o\left(|S|\right)$, it means that $\Omega(|S|)$ vertices in S have s in their label.

More formally, we start by estimating, for each vertex v, the number of vertices $s,t \in S$ such that $\mathrm{dist}(s,v) + \mathrm{dist}(v,t) \leq D_S$.

LEMMA D.3. (FOR A PROOF, SEE LEMMA B.32) *Let v be a vertex with degree $\omega(1)$. Then, the number of pairs of vertices $s,t \in S$ such that $\mathrm{dist}(s,v) + \mathrm{dist}(v,t) \leq D_S$, and $\mathrm{dist}(s,w) + \mathrm{dist}(w,t) > D_S$ for each w such that $\deg(w) > \deg(v)$, is at most $\deg(v)^2|S|^2n^{-1+\mathcal{O}(\varepsilon)}$.*

Let us consider the ordering of all vertices $s_1, \ldots, s_n$, and let us estimate:

$$
\begin{aligned}
&|\{(s_i,s_j) \in S^2 : i < j, s_i \notin L(s_j)\}| \\
&\leq |\{(s_i,s_j) \in S^2 : i < j, \exists k < i, \mathrm{dist}(s_i,s_j) = \mathrm{dist}(s_i,s_k) + \mathrm{dist}(s_k,s_j)\}| \\
&\leq |\{(s_i,s_j) \in S^2 : i < j, \exists k < i, \mathrm{dist}(s_i,s_k) + \mathrm{dist}(s_k,s_j) \leq D_S\}| \\
&\leq \sum_{k<i} \deg(s_k)^2|S|^2n^{-1+\mathcal{O}(\varepsilon)} \\
&\leq n^{1-\alpha(\beta-3)+\varepsilon}|S|^2n^{-1+\mathcal{O}(\varepsilon)} \\
&= o(|S|^2).
\end{aligned}
$$

We used the fact that $\sum_{k<i} \deg(s_k)^2 \leq n^{1-\alpha(\beta-3)+\varepsilon}$: let us prove it formally, using Abel's summation technique and Property 2.4.

$$\sum_{k<i} \deg(s_k)^2 = \sum_{d=n^\alpha}^{+\infty} d^2 |\{v : \deg(v) = n^\alpha\}|$$

$$= \sum_{d=n^\alpha}^{+\infty} d^2 |\{v : \deg(v) \geq d\}| - \sum_{d=n^\alpha+1}^{+\infty} (d-1)^2 |\{v : \deg(v) \geq d\}|$$

$$\leq n^{2\alpha} |\{v : \deg(v) \geq n^\alpha\}| + \sum_{d=n^\alpha}^{+\infty} 2d |\{v : \deg(v) \geq d\}|$$

$$\leq n^{2\alpha} \frac{n}{n^{\alpha(\beta-1)}} + \sum_{d=n^\alpha}^{+\infty} 2d \frac{n}{d^{\beta-1}} = \mathcal{O}\left(n^{1-\alpha(\beta-3)}\right).$$

We have proved that $|\{(s_i, s_j) \in S^2 : i < j, s_i \in L(s_j)\}| = \Omega\left(|S|^2\right)$, and consequently the total label size is at least $\Omega\left(|S|^2\right) \geq n^{2-3\alpha(\beta-1)}$. We claim that this means that there are many labels with size bigger than $n^{1-4\alpha(\beta-1)}$, because no label has size bigger than n. Indeed, if ℓ_i is the size of label i, $n^{2-3\alpha(\beta-1)} \leq \sum_{i=1}^{n} \ell_i \leq n^{1-4\alpha(\beta-1)} |\{i : \ell_i \leq n^{1-4\alpha(\beta-1)}\}| + n |\{i : \ell_i > n^{1-4\alpha(\beta-1)}\}| \leq n^{2-4\alpha(\beta-1)} + n |\{i : \ell_i > n^{1-4\alpha(\beta-1)}\}|$, and hence $|\{i : \ell_i > n^{1-4\alpha(\beta-1)}\}| \geq n^{1-3\alpha(\beta-1)} - n^{1-4\alpha(\beta-1)} \geq n^{1-4\alpha(\beta-1)}$.

We have proved that, for each $\alpha' = 4\alpha(\beta - 1)$, there are at least $n^{1-\alpha'}$ labels of size $n^{1-\alpha'}$: consequently, the expected time to perform a distance query is at least $\frac{n^{2-2\alpha'}}{n^2} n^{1-\alpha'}$, because the probability that we hit two vertices s, t whose labels are bigger than $n^{1-\alpha'}$ is at least $\frac{n^{2-2\alpha'}}{n^2}$. If we let $\alpha' = \mathcal{O}(\varepsilon)$, the average time for a distance query becomes at least $n^{1-\mathcal{O}(\varepsilon)}$. Similarly, the space occupied is $n^{2-\mathcal{O}(\varepsilon)}$.