

Modularity of the ABCD Random Graph Model with Community Structure

Bogumił Kamiński* Bartosz Pankratz† Paweł Prałat‡ François Théberge§

October 27, 2022

Abstract

The **Artificial Benchmark for Community Detection (ABCD)** graph is a random graph model with community structure and power-law distribution for both degrees and community sizes. The model generates graphs with similar properties as the well-known **LFR** one, and its main parameter ξ can be tuned to mimic its counterpart in the **LFR** model, the mixing parameter μ .

In this paper, we investigate various theoretical asymptotic properties of the **ABCD** model. In particular, we analyze the modularity function, arguably, the most important graph property of networks in the context of community detection. Indeed, the modularity function is often used to measure the presence of community structure in networks. It is also used as a quality function in many community detection algorithms, including the widely used *Louvain* algorithm.

1 Introduction

One of the most important features of real-world networks is their community structure, as it reveals the internal organization of nodes [12]. In social networks communities may represent groups by interest, in citation networks they correspond to related papers, in the Web communities are formed by pages on related topics, etc. Being able to identify communities in a network could help us to exploit this network more effectively.

Unfortunately, there are very few datasets with ground-truth identified and labelled. As a result, there is need for synthetic random graph models with community structure that resemble real-world networks in order to benchmark and tune clustering algorithms that are unsupervised by nature. The **LFR** (Lancichinetti, Fortunato, Radicchi) model [32, 30] generates networks with communities and at the same time it allows for the heterogeneity in the distributions of both node degrees and of community sizes. It became a standard and extensively used method for generating artificial networks.

*Decision Analysis and Support Unit, SGH Warsaw School of Economics, Warsaw, Poland; e-mail: bogumil.kaminski@sgh.waw.pl

†Department of Mathematics, Toronto Metropolitan University, Toronto, ON, Canada; e-mail: bartosz.pankratz@ryerson.ca

‡Department of Mathematics, Toronto Metropolitan University, Toronto, ON, Canada; e-mail: pralat@ryerson.ca

§Tutte Institute for Mathematics and Computing, Ottawa, ON, Canada; email: theberge@ieee.org

In this paper, we analyze the Artificial Benchmark for Community Detection (**ABCD** graph) [24] that was recently introduced and implemented*, including a fast implementation that uses multiple threads (**ABCDe**)†. The undirected variant of **LFR** and **ABCD** produce graphs with comparable properties but **ABCD/ABCDe** is faster than **LFR** [21] and can be easily tuned to allow the user to make a smooth transition between the two extremes: pure (disjoint) communities and random graph with no community structure. Moreover, the model is quite flexible which allows us to adjust it to include, for example, outliers [27] and hypergraphs [26]. More importantly from the perspective of this paper, it is easier to analyze theoretically. See Subsection 2.3 for a definition of the model.

The key ingredient for many clustering algorithms is *modularity*, which is at the same time a global criterion to define communities, a quality function of community detection algorithms, and a way to measure the presence of community structure in a network. The definition of modularity for graphs was first introduced by Newman and Girvan in [38]. We present the definition in Subsection 2.4.

Despite some known issues with this function such as the “resolution limit” reported in [13], many popular algorithms for partitioning nodes of large graphs use it [11, 37, 31] and perform very well. The list includes one of the mostly used unsupervised algorithms for detecting communities in graphs, the *Louvain* (hierarchical) algorithm [7]. For more details we direct the reader to any book on complex networks, including the following recent additions [25, 29].

1.1 Summary of Results

In this paper, we investigate the modularity function for the **ABCD** model \mathcal{A} . The paper is structured as follows. The **ABCD** model is introduced in Subsection 2.3 and the modularity function is defined in Subsection 2.4. Results for other random graph models in the context of the modularity function are summarized in Section 3. Some useful general observations are made in Section 4: concentration results that are used in almost all of our proofs are outlined in Subsection 4.1, useful expansion properties of **random d -regular graphs** are presented in Subsection 4.2.

We start analyzing the **ABCD** model by investigating some basic properties—see Section 5. These properties will be needed to establish results for the modularity function but they are important on their own. Lemma 5.2 shows that the degree distribution is well concentrated around the corresponding expectations. Corollary 5.5 shows a concentration for the number of communities and well as the distribution of their sizes. The same generating process is applied in **LFR** so the two results hold for that model as well. The **ABCD** model assigns nodes to communities randomly. Clearly, there is no hope to predict the volumes of small communities of constant size but sufficiently large communities have their volumes as well as the number of internal edges well concentrated around the corresponding expectations—see Lemma 5.6.

Then we move to the results for the modularity function. By design of the **ABCD** model, a $1 - \xi$ fraction of the edges should become community edges and so should end up in some part of the ground truth partition \mathbf{C} . (ξ is the main parameter of the model responsible for the level of noise.) It is indeed the case but it turns out that a negligible fraction of the background graph join them there. As a result, the modularity function of the ground-truth partition \mathbf{C} is asymptotic to $1 - \xi$, as proved in Theorem 6.1.

Analyzing the maximum modularity is much more complex. We have two types of results. The first result (Theorem 6.2) shows that when the level of noise is sufficiently large (ξ close to

*<https://github.com/bkamins/ABCDGraphGenerator.jl/>

†<https://github.com/tolcz/ABCDeGraphGenerator.jl/>

one), then the maximum modularity $q^*(\mathcal{A})$ is asymptotically larger than $q(\mathbf{C})$, the modularity of the ground-truth. In this regime, the number of edges within community graphs G_i is relatively small so a partition of the background graph into small connected pieces yields a better modularity function. To show this result, we need to investigate the degree distribution of the background graph (Lemma 6.3) which might be of independent interest.

The second set of results is concerned with graphs with low level of noise (ξ close to zero). For these graphs, the situation is rather the opposite. It turns out that the ground truth partition is asymptotically the best possible, that is, the maximum modularity $q^*(\mathcal{A})$ is only $o(1)$ away from $q(\mathbf{C})$, the modularity of the ground truth partition \mathbf{C} ; both of them are asymptotic to $1 - \xi$ (see Theorem 6.7). For some technical reason, it is assumed that δ , the minimum degree of \mathcal{A} , is sufficiently large: the lower bound of 100 easily works but it may be improved with more detailed treatment. Moreover, it seems that one needs a different approach to uncover the real bottleneck. On the other hand, the above property is not true if $\delta = 1$ (see Theorem 6.8): if $\delta = 1$, then $q^*(\mathcal{A})$ is substantially larger than $q(\mathbf{C})$, regardless of how close to zero ξ is.

Finally, let us mention that the approach used to prove Theorem 6.7 uses a coupling of random graph $\mathcal{P}(\mathbf{w})$ on n' nodes and a given degree sequence with a random b -regular graph $\mathcal{P}_{n'',b}$ on n'' nodes. We use this coupling to show that good expansion properties of $\mathcal{P}_{n'',b}$ imply good expansion properties for $\mathcal{P}(\mathbf{w})$. This coupling seems powerful and might be potentially useful for some other applications.

1.2 Simulations

This paper focuses on asymptotic theoretical results of the **ABCD** model. Moreover, we performed a number of simulations and compared asymptotic predictions with graphs generated by a computer. These simulations show that the behaviour of small random instances is similar to what is predicted by the theory. This is a good news for practitioners as it shows that, despite the fact that the generative algorithm is randomized, the model has good stability. The code is accessible on GitHub repository[‡].

1.3 Open Problems

Theoretical results and simulations suggest that if δ , the minimum degree of \mathcal{A} , satisfies $\delta \geq \delta_0$ for some $\delta_0 \geq 2$, then there exists a constant $\xi_0 = \xi_0(\delta)$ (that possibly depends also on other parameters of the **ABCD** model \mathcal{A}) such that the following holds *w.h.p.* (that is, with probability tending to one as $n \rightarrow \infty$):

- if $0 < \xi < \xi_0$, then $q^*(\mathcal{A}) \sim q(\mathbf{C})$, where \mathbf{C} is the ground truth partition of the set of nodes of \mathcal{A} ,
- if $\xi > \xi_0$, then $q^*(\mathcal{A})$ is separated by a constant from $q(\mathbf{C})$.

Our results make the first step towards this conjecture by showing upper and lower bounds for such threshold constant ξ_0 , when $\delta_0 = 100$. The bounds for ξ_0 are not close to each other. The next step would be to narrow the gap down or perhaps to determine the threshold value exactly, provided that δ_0 is sufficiently large. Another natural direction would be to decrease the lower bound for δ , that is, to decrease the value of δ_0 . We showed that $\delta = 1$ does not have the desired property but maybe

[‡]<https://github.com/tolcz/ABCDeGraphGenerator.jl/>

$\delta_0 = 2$? Or maybe one can always construct a better partition than \mathbf{C} when $\delta = 2$, regardless how small parameter ξ is? These questions are left as open questions for future investigation.

2 Definitions (of ABCD Model and Modularity)

2.1 Asymptotic Notation

Our results are asymptotic in nature, that is, we will assume that the number of nodes $n \rightarrow \infty$. Formally, we consider a sequence of graphs $G_n = (V_n, E_n)$ and we are interested in events that hold *with high probability (w.h.p.)*, that is, events that hold with probability tending to 1 as $n \rightarrow \infty$. It will also be convenient to consider events that hold *with extreme probability (w.e.p.)*, that is, events that hold with probability at least $1 - \exp(-\Omega((\log n)^2))$. An easy but convenient property is that if a polynomial number of events hold *w.e.p.*, then *w.e.p.* all of them hold simultaneously.

Given two functions $f = f(n)$ and $g = g(n)$, we will write $f(n) = \mathcal{O}(g(n))$ if there exists an absolute constant $c \in \mathbb{R}_+$ such that $|f(n)| \leq c|g(n)|$ for all n , $f(n) = \Omega(g(n))$ if $g(n) = \mathcal{O}(f(n))$, $f(n) = \Theta(g(n))$ if $f(n) = \mathcal{O}(g(n))$ and $f(n) = \Omega(g(n))$, and we write $f(n) = o(g(n))$ or $f(n) \ll g(n)$ if $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$. In addition, we write $f(n) \gg g(n)$ if $g(n) = o(f(n))$ and we write $f(n) \sim g(n)$ if $f(n) = (1 + o(1))g(n)$, that is, $\lim_{n \rightarrow \infty} f(n)/g(n) = 1$.

2.2 Other Notation

We will use $\log n$ to denote the natural logarithm of n . For a given $n \in \mathbb{N} := \{1, 2, \dots\}$, we will use $[n]$ to denote the set consisting of the first n natural numbers, that is, $[n] := \{1, 2, \dots, n\}$. Finally, as typical in the field of random graphs, for expressions that clearly have to be an integer, we round up or down but do not specify which: the choice of which does not affect the argument.

2.3 ABCD Model

Table 1: Parameters of the **ABCD** model

parameter	range	description
n	\mathbb{N}	number of nodes
γ	$(2, 3)$	power-law exponent of degree distribution
δ	\mathbb{N}	minimum degree at least δ
ζ	$(0, \frac{1}{\gamma-1}]$	maximum degree at most n^ζ
β	$(1, 2)$	power-law exponent of distribution of community sizes
s	$\mathbb{N} \setminus [\delta]$	community sizes at least s
τ	$(\zeta, 1)$	community sizes at most n^τ
ξ	$(0, 1)$	level of noise

The **ABCD** model is governed by 8 parameters summarized in Table 1. For a fixed set of parameters, we generate the **ABCD** graph $\mathcal{A} = \mathcal{A}(n, \gamma, \delta, \zeta, \beta, s, \tau, \xi)$ following the steps outlined below. Each time we refer to graph \mathcal{A} in this paper, we implicitly (or explicitly, but it happens rather rarely) fix all of these parameters.

2.3.1 Degree Distribution

Let $\gamma \in (2, 3)$, $\delta \in \mathbb{N}$, and $\zeta \in (0, 1)$. Degrees of nodes of **ABCD** graph \mathcal{A} are generated randomly following the (truncated) *power-law distribution* $\mathcal{P}(\gamma, \delta, \zeta)$ with exponent γ , minimum value δ , and maximum value $D = n^\zeta$. In order to make sure the sum of the degrees is even, if needed, we decrease by one the degree of one node of the largest degree.

It is easy to show (see Lemma 5.1) that for any $\omega = \omega(n)$ tending to infinity as $n \rightarrow \infty$ *w.h.p.* the maximum degree of \mathcal{A} is at most $n^{1/(\gamma-1)}\omega$ (of course, by definition, it is deterministically at most n^ζ). As a result, for any two values of $\zeta_1, \zeta_2 \in (\frac{1}{\gamma-1}, 1)$ one may couple the two corresponding **ABCD** graphs \mathcal{A} so that *w.h.p.* they produce exactly the same graph. Hence, for convenience but without loss of generality, we will later on assume that $\zeta \in (0, \frac{1}{\gamma-1}]$.

2.3.2 Distribution of the Community Sizes

Let $\beta \in (1, 2)$, $s \in \mathbb{N} \setminus [\delta]$, and $\tau \in (\zeta, 1)$. Community sizes of **ABCD** graph \mathcal{A} are generated randomly following the (truncated) *power-law distribution* $\mathcal{P}(\beta, s, \tau)$ with exponent β , minimum value s , and maximum value $S = n^\tau$. Communities are generated with this distribution as long as the sum of their sizes is less than n , the desired number of nodes. Suppose that the last community has size z and after adding it to the remaining ones, the sum of their sizes will exceed n by $k \in \mathbb{N} \cup \{0\}$. If $k = 0$, then there is nothing else to do. If $z - k \geq s$, then the size of the last community is reduced to $z - k$ so that the total number of nodes is exactly n . Otherwise, we select $z - k < s$ old communities at random, increase their sizes by one, and remove the last community so that the desired property holds.

The assumption that $\tau > \zeta$ is introduced to make sure large degree nodes have large enough communities to be assigned to. Similarly, the assumption that $s \geq \delta + 1$ is required to guarantee that small communities are not too small and so that they can accommodate small degree nodes.

2.3.3 Assigning Nodes into Communities

At this point, the degree distribution ($w_1 \geq w_2 \geq \dots \geq w_n$) and the distribution of community sizes ($c_1 \geq c_2 \geq \dots \geq c_\ell$) are already fixed. The final **ABCD** graph \mathcal{A} will be formed as the union of $\ell + 1$ independent graphs: ℓ community graphs $G_i = (C_i, E_i)$, $i \in [\ell]$, and a single background graph $G_0 = (V, E_0)$, where $V = \bigcup_{i \in [\ell]} C_i$. Roughly ξw_i edges incident to node i will, by definition, belong to its own community but a few additional edges from the background graph might end up in that community. In order to create enough room for these edges, a node of degree w_i will be allowed to be assigned to a community of size c_j if the following inequality is satisfied:

$$\lceil (1 - \xi\phi)w_i \rceil \leq c_j - 1, \quad \text{where } \phi = 1 - \sum_{k \in [\ell]} (c_k/n)^2.$$

Note that this condition is equivalent to the following one:

$$w_i \leq \frac{c_j - 1}{1 - \xi\phi}. \quad (1)$$

An assignment of nodes into communities will be called *admissible* if the above inequality is satisfied for all nodes. We will show in Subsection 5.3 that there are many admissible assignments. In particular, there are linearly many nodes of degree δ but, fortunately, *w.h.p.* communities of size

more than n^ζ (more than the maximum degree) have space for almost all nodes. We select one admissible assignment uniformly at random. Sampling uniformly one of such assignments turns out to be relatively easy from both theoretical and practical points of view. We will discuss it in detail in Subsection 5.3.

2.3.4 Distribution of Weights

Parameter $\xi \in (0, 1)$ reflects the amount of noise in the network. It controls the fraction of edges that are between communities. Indeed, asymptotically (but not exactly) a $1 - \xi$ fraction of the edges are going to end up within one of the communities. Each node will have its degree w_i split into two parts: *community degree* y_i and *background degree* z_i ($w_i = y_i + z_i$). Our goal is to get $y_i \approx (1 - \xi)w_i$ and $z_i \approx \xi w_i$. However, both y_i and z_i have to be non-negative integers and for each community $C \subseteq V$, $\sum_{i \in C} y_i$ has to be even. Note that since $\sum_{i \in V} w_i$ is even, it will imply that

$$\sum_{i \in V} z_i = \sum_{i \in V} (w_i - y_i) = \sum_{i \in V} w_i - \sum_C \sum_{i \in C} y_i$$

is even too.

For each community $C \subseteq V$ we identify the *leader*, a node of largest degree w_i associated with community C . (If many nodes in C have the largest degree, then we arbitrarily select one of them to be the leader.) For non-leaders we split the weights as follows:

$$y_i = \lfloor (1 - \xi)w_i \rfloor \quad \text{and} \quad z_i = w_i - y_i,$$

where for a given integer $a \in \mathbb{Z}$ and real number $b \in [0, 1)$ the random variable $\lfloor a + b \rfloor$ is defined as

$$\lfloor a + b \rfloor = \begin{cases} a & \text{with probability } 1 - b \\ a + 1 & \text{with probability } b. \end{cases} \quad (2)$$

(Note that $\mathbb{E}[\lfloor a + b \rfloor] = a(1 - b) + (a + 1)b = a + b$.) For the leader of community C we round $(1 - \xi)w_i$ up or down so that the sum of weights in each cluster is even. If $(1 - \xi)w_i \in \mathbb{N}$ and the sum of weights y_i in C is odd, then we randomly make a decision whether to subtract or add one to make the sum to be even.

2.3.5 Creating Graphs

As already mentioned, the final **ABCD** graph $\mathcal{A} = (V, E)$ will be formed as the union of $\ell + 1$ independent graphs: ℓ community graphs $G_i = (C_i, E_i)$, $i \in [\ell]$, and a single background graph $G_0 = (V, E_0)$, where $V = \bigcup_{i \in [\ell]} C_i$, that is, $E = \bigcup_{i \in [\ell] \cup \{0\}} E_i$. Each of these $\ell + 1$ graphs will be created independently. The partition $\mathbf{C} = \{C_1, C_2, \dots, C_\ell\}$ will be called a *ground-truth* partition.

Suppose then that our goal is to create a graph on n nodes with a given degree distribution $\mathbf{w} := (w_1, w_2, \dots, w_n)$, where \mathbf{w} is any vector of non-negative integers such that $w := \sum_{i \in [n]} w_i$ is even. We define a random multi-graph $\mathcal{P}(\mathbf{w})$ with a given degree sequence known as the **configuration model** (sometimes called the **pairing model**), which was first introduced by Bollobás [8]. (See [6, 41, 42] for related models and results.) We start with w *points* that are partitioned into n *buckets* labelled with labels v_1, v_2, \dots, v_n ; bucket v_i consists of w_i points. It is easy to see that there are $\frac{w!}{(w/2)!2^w}$ pairings of points. We select one of such pairings uniformly at

random, and construct a multi-graph $\mathcal{P}(\mathbf{w})$, with loops and parallel edges allowed, as follows: nodes are the buckets v_1, v_2, \dots, v_n , and a pair of points xy corresponds to an edge $v_i v_j$ in $\mathcal{P}(\mathbf{w})$ if x and y are contained in the buckets v_i and v_j , respectively.

2.3.6 Simulation Corner

Note that the **ABCD** model \mathcal{A} allows loops and multiple edges. Indeed, they can occur both in any of the generated graphs G_i ($i \in [\ell] \cup \{0\}$) or after taking a union of their edge sets. In general, however, there will not be very many of them. To keep the theoretical model simple, in this paper we allow \mathcal{A} be a multi-graph but, alternatively, one may condition on \mathcal{A} to be a simple graph. In practice, the algorithm performs some kind of edges “switching” that is known to generate a random graph that is very close to the uniform distribution [18]. We will use simple graphs for our experiments to show that, indeed, the difference is not detectable. Moreover, to get closed formulas in some theoretical results proved in this paper we use the continuous variant of the power law distribution. For example, if $X \in \mathcal{P}(\gamma, \delta, \zeta)$, then for any $k \in \{\delta, \delta + 1, \dots, D\}$ we assume that

$$q_k = \Pr(X = k) = \frac{\int_k^{k+1} x^{-\gamma} dx}{\int_\delta^{D+1} x^{-\gamma} dx}.$$

Alternatively, one may use the discrete counterpart, namely, assume that

$$r_k = \Pr(X = k) = \frac{k^{-\gamma}}{\sum_{x=\delta}^D x^{-\gamma}}. \quad (3)$$

All results proved in this paper hold for both variants. We state them for the continuous one but one may replace q_k with r_k to get the discrete counterparts. Since the default implementation of the **ABCD** model uses the discrete distribution, we use it for our simulations. For more details we direct the reader to the original paper on the **ABCD** model [24].

2.4 Modularity Function

The modularity function favours partitions of the set of nodes of a graph G in which a large proportion of the edges falls entirely within the parts but benchmarks it against the expected number of edges one would see in those parts in the corresponding **Chung-Lu** random graph model [10] which generates graphs with the expected degree sequence following exactly the degree sequence in G .

Formally, for a graph $G = (V, E)$ and a given partition $\mathbf{A} = \{A_1, A_2, \dots, A_\ell\}$ of V , the *modularity function* is defined as follows:

$$q(\mathbf{A}) = \sum_{A_i \in \mathbf{A}} \frac{e(A_i)}{|E|} - \sum_{A_i \in \mathbf{A}} \left(\frac{\text{vol}(A_i)}{\text{vol}(V)} \right)^2, \quad (4)$$

where for any $A \subseteq V$, $e(A) = |\{uv \in E : u, v \in A\}|$ is the number of edges in the subgraph of G induced by set A , and $\text{vol}(A) = \sum_{v \in A} \deg(v)$ is the *volume* of set A . In particular, $\text{vol}(V) = 2|E|$. The first term in (4), $\sum_{A_i \in \mathbf{A}} e(A_i)/|E|$, is called the *edge contribution* and it computes the fraction of edges that fall within one of the parts. The second one, $\sum_{A_i \in \mathbf{A}} (\text{vol}(A_i)/\text{vol}(V))^2$, is called the *degree tax* and it computes the expected fraction of edges that do the same in the corresponding random graph (the null model). The modularity measures the deviation between the two.

It is easy to see that for any partition \mathbf{A} , $q(\mathbf{A}) \leq 1$. On the other hand, it can be shown that $q(\mathbf{A}) \geq -1/2$. Also, if $\mathbf{A} = \{V\}$, then $q(\mathbf{A}) = 0$, and if $\mathbf{A} = \{\{v_1\}, \{v_2\}, \dots, \{v_n\}\}$, then $q(\mathbf{A}) = -\sum (\deg(v)/\text{vol}(V))^2 < 0$. The maximum *modularity* $q^*(G)$ is defined as the maximum of $q(\mathbf{A})$ over all possible partitions \mathbf{A} of V ; that is, $q^*(G) = \max_{\mathbf{A}} q(\mathbf{A})$. In order to maximize $q(\mathbf{A})$ one wants to find a partition with large edge contribution subject to small degree tax. If $q^*(G)$ approaches 1 (which is the trivial upper bound), we observe a strong community structure; conversely, if $q^*(G)$ is close to zero (which is the trivial lower bound), there is no community structure. The definition in (4) can be generalized to weighted edges by replacing edge counts with sums of edge weights. It can also be generalized to hypergraphs [22, 23].

3 Related Results for Random Graphs

Analyzing the maximum modularity $q^*(G)$ for sparse random graphs is a challenging task. The most attention was paid to **random d -regular graphs** $\mathcal{G}_{n,d}$ but even for this family of graphs we only know upper and lower bounds for $q^*(\mathcal{G}_{n,d})$ that are quite apart from each other. For example, for the random 3-regular graph $\mathcal{G}_{n,3}$ we only know that *w.h.p.*

$$0.667026 \leq q^*(\mathcal{G}_{n,3}) \leq 0.789998.$$

These bounds were recently proved in [33] but the main goal of that paper was to confirm the conjecture from [35] that *w.h.p.* $q^*(\mathcal{G}_{n,3}) \geq 2/3 + \varepsilon$ for some $\varepsilon > 0$. We refer the reader to [35, 40] for numerical bounds on $q^*(\mathcal{G}_{n,d})$ for other values of $d \geq 3$ and for some explicit but weaker bounds. It is also known that *w.h.p.* $q^*(\mathcal{G}_{n,2}) \sim 1$ [35].

The **binomial random graphs** $\mathcal{G}(n, p)$ were studied in [36] where it was shown that *w.h.p.* $q^*(\mathcal{G}(n, p)) \sim 1$, provided that $pn \leq 1$ whereas *w.h.p.* $q^*(\mathcal{G}(n, p)) = \Theta(1/\sqrt{pn})$, provided that $pn \geq 1$ and $p < 1 - \varepsilon$ for some $\varepsilon > 0$. The modularity of the well-known **Preferential Attachment (PA) model** [5] and the **Spatial Preferential Attachment (SPA) model** [1] was studied in [40]. Finally, the modularity of a model of random geometric graphs on the hyperbolic plane [28], known as the **KPKBV model** after its inventors, was recently studied in [9].

4 Preliminaries

4.1 Chernoff Bounds and Their Generalization

Let us first state a specific instance of Chernoff's bound that we will find often useful. Let $X \in \text{Bin}(n, p)$ be a random variable with the binomial distribution with parameters n and p . Then, a consequence of *Chernoff's bound* (see e.g. [19, Corollary 2.3]) is that

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon \mathbb{E}[X]) \leq 2 \exp\left(-\frac{\varepsilon^2 \mathbb{E}[X]}{3}\right) \quad (5)$$

for $0 < \varepsilon < 3/2$. However, at some point we will need need a stochastic upper bound for X when $\mathbb{E}[X]$ is small. In such situations the following bound can be applied instead of (5) (see e.g. [19, Theorem 2.1]):

$$\mathbb{P}(X \geq \mathbb{E}[X] + u) \leq \exp\left(-\frac{u^2}{2(\mathbb{E}[X] + u/3)}\right). \quad (6)$$

Let us mention that the above bounds hold for the general case in which $X = \sum_{i=1}^n X_i$ and $X_i \in \text{Bernoulli}(p_i)$ with (possibly) different p_i (e.g. see [19, Theorem 2.8]). Moreover, they also hold for a hypergeometric distribution with parameters n , z , and t , where $\max(z, t) \leq n$. Let T be a subset of $[n]$ of size t selected uniformly at random. Then a random variable with a hypergeometric distribution is defined as follows: $X = |T \cap [z]|$. Then the above inequalities hold with $\mathbb{E}[X] = zt/n$ (again, e.g. see [19, Theorem 2.10]).

We will also need the following result which can be viewed as a generalization of Chernoff bounds. In particular, the two inequalities above ((5 and (6)) are special cases when $c = 1$.

Lemma 4.1. *Let (c_1, c_2, \dots, c_r) be a sequence of natural numbers with $c = \max_i c_i$. Let $S_j = \sum_{i=1}^j c_i Z_i$, where $Z_i, i \in [r]$ are independent Bernoulli(p) random variables. Let $\mu_j = \mathbb{E}[S_j] = p \sum_{i=1}^j c_i$, and let $\mu = \mu_r = \mathbb{E}[S_r]$. Then for $u \geq 0$ we have that*

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq j \leq r} (S_j - \mu_j) \geq u\right) &\leq \exp\left(-\frac{u^2}{2c(\mu + u/3)}\right) \text{ and} \\ \mathbb{P}\left(\max_{1 \leq j \leq r} (\mu_j - S_j) \geq u\right) &\leq \exp\left(-\frac{u^2}{2c\mu}\right). \end{aligned}$$

In particular, for $\varepsilon \leq 3/2$ we have that

$$\mathbb{P}\left(\max_{1 \leq j \leq r} |S_j - \mu_j| \geq \varepsilon \mu\right) \leq 2 \exp\left(-\frac{\varepsilon^2 \mu}{3c}\right).$$

To prove this lemma, one can easily adjust the proof of the classic Chernoff bound. Alternatively, the same bounds come from [34]. In that paper, the counterpart of Lemma 4.1 is stated for $S_r - \mu$ (see Theorem 2.3); however, the author comments that $S_r - \mu$ can be replaced with $\max_{1 \leq j \leq r} (S_j - \mu_j)$ (which is a slightly stronger version than we need here) as follows. A standard martingale bound shows that eg. for any $h > 0$:

$$\mathbb{P}\left(\max_{1 \leq j \leq r} (S_j - \mu_j) \geq u\right) \leq e^{-hu} \mathbb{E}\left[e^{h(S_r - \mu)}\right].$$

Then plugging this into the appropriate place in the proof of Theorem 2.3 yields the desired bounds.

4.2 Expansion Properties of Random d -regular graphs

Let $\mathcal{G}_{n,d}$ be the probability space of **random d -regular simple graphs** with uniform probability distribution ($d \geq 2$ is fixed and n is even if d is odd). It is an easy fact that the probability of a random pairing $\mathcal{P}_{n,d} := \mathcal{P}(\mathbf{w})$ with $\mathbf{w} := (d, d, \dots, d)$ corresponding to a given simple d -regular graph G is independent of the graph. As a result, the restriction of the probability space $\mathcal{P}_{n,d}$ to simple graphs is precisely $\mathcal{G}_{n,d}$. Moreover, it is well known that a random pairing generates a simple graph with probability asymptotic to $e^{(1-d^2)/4}$ depending on d , so that any event holding *w.h.p.* over the probability space of random pairings also holds *w.h.p.* over the corresponding space $\mathcal{G}_{n,d}$. For this reason, asymptotic results over random pairings can be immediately transferred to $\mathcal{G}_{n,d}$. For more information on this model, see [42].

We will use well-known expansion properties of random d -regular graphs that follow from their eigenvalues. These expansion properties are known to hold *w.h.p.* for $\mathcal{G}_{n,d}$ but, fortunately, they actually hold *w.h.p.* for $\mathcal{P}_{n,d}$ (and so, by the argument mentioned above, they immediately hold *w.h.p.* for $\mathcal{G}_{n,d}$ which is what is typically used). In one of our proofs (see the proof of Theorem 6.7), we will couple the **ABCD** model \mathcal{A} with a random pairing $\mathcal{P}_{n,d}$. This coupling will allow us to deduce some useful expansion properties of \mathcal{A} from the corresponding properties of $\mathcal{P}_{n,d}$.

The adjacency matrix $A = A(G)$ of a given a d -regular (multi)graph G with n nodes, is an $n \times n$ real and symmetric matrix. Thus, the matrix A has n real eigenvalues which we denote by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. It is known that certain properties of a d -regular graph are reflected in its spectrum but, since we focus on expansion properties, we are particularly interested in the following quantity: $\lambda = \lambda(G) = \max(|\lambda_2|, |\lambda_n|)$. In words, λ is the largest absolute value of an eigenvalue other than $\lambda_1 = d$. For more details, see the general survey [17] about expanders, or [4, Chapter 9].

The value of λ for random d -regular graphs has been studied extensively. A major result due to Friedman [14] is the following:

Lemma 4.2 ([14]). *For every fixed $\varepsilon > 0$ and for $G \in \mathcal{P}_{n,d}$ (and so also for $G \in \mathcal{G}_{n,d}$), w.h.p.*

$$\lambda(G) \leq 2\sqrt{d-1} + \varepsilon.$$

The number of edges $|E(S, T)|$ between sets S and T is expected to be close to the expected number of edges between S and T in a random graph of edge density d/n , namely, $d|S||T|/n$. A small λ (or large spectral gap) implies that this deviation is small. The following useful bound is essentially proved in [2] (see also [4]):

Lemma 4.3 (Expander Mixing Lemma). *Let $G = (V, E)$ be a d -regular (multi)graph with n nodes and set $\lambda = \lambda(G)$. Then for all $S, T \subseteq V$*

$$\left| |E(S, T)| - \frac{d|S||T|}{n} \right| \leq \lambda \sqrt{|S||T|}.$$

(Note that $S \cap T$ does not have to be empty; in general, $|E(S, T)|$ is defined to be the number of edges between $S \setminus T$ to T plus twice the number of edges that contain only nodes of $S \cap T$.)

The Expander Mixing Lemma is very useful but for our purpose it is better to apply a slightly stronger lower estimate for $|E(S, V \setminus S)|$, namely,

$$|E(S, V \setminus S)| \geq \frac{(d - \lambda)|S||V \setminus S|}{n} \tag{7}$$

for all $S \subseteq V$. This is proved in [3], see also [4]. Combining this inequality with a simple averaging argument gives immediately the following useful bound for the maximum modularity that was observed in [40].

Lemma 4.4 ([40]). *Let $G = (V, E)$ be a d -regular (multi)graph with n nodes and set $\lambda = \lambda(G)$. Then,*

$$q^*(G) \leq \frac{\lambda}{d}.$$

5 Some Properties of ABCD

5.1 Degree Distribution

Let $\gamma \in (2, 3)$, $\delta \in \mathbb{N}$, and $\zeta \in (0, 1)$. We will show soon that we may assume that $\zeta \leq \frac{1}{\gamma-1}$ but for now we allow ζ to be any value from $(0, 1)$. Recall that the degrees of nodes of the **ABCD** model are generated randomly following the (truncated) *power-law distribution* $\mathcal{P}(\gamma, \delta, \zeta)$ with exponent γ , minimum value δ , and maximum value $D = n^\zeta$. More precisely, if $X \in \mathcal{P}(\gamma, \delta, \zeta)$, then for any $k \in \{\delta, \delta + 1, \dots, D\}$,

$$\begin{aligned}
q_k &= \Pr(X = k) = \frac{\int_k^{k+1} x^{-\gamma} dx}{\int_\delta^{D+1} x^{-\gamma} dx} = \frac{k^{1-\gamma} - (k+1)^{1-\gamma}}{\delta^{1-\gamma} - (D+1)^{1-\gamma}} \\
&= (1 + \mathcal{O}(n^{-\zeta(\gamma-1)})) \left(k^{-(\gamma-1)} - (k+1)^{-(\gamma-1)} \right) \delta^{\gamma-1} \\
&= (1 + \mathcal{O}(n^{-\zeta(\gamma-1)})) k^{-(\gamma-1)} \left(1 - (1 + 1/k)^{-(\gamma-1)} \right) \delta^{\gamma-1} \\
&= (1 + \mathcal{O}(n^{-\zeta(\gamma-1)})) k^{-(\gamma-1)} \left(1 - (1 - (\gamma-1)/k + \mathcal{O}(1/k^2)) \right) \delta^{\gamma-1} \\
&= (1 + \mathcal{O}(n^{-\zeta(\gamma-1)}) + \mathcal{O}(k^{-1})) k^{-(\gamma-1)} \frac{\gamma-1}{k} \delta^{\gamma-1} \\
&= (1 + \mathcal{O}(n^{-\zeta(\gamma-1)}) + \mathcal{O}(k^{-1})) k^{-\gamma} (\gamma-1) \delta^{\gamma-1}.
\end{aligned} \tag{8}$$

As promised earlier, we start with a proof of an upper bound for the maximum degree, which justifies our future assumption that $\zeta \in (0, 1/(\gamma-1)]$.

Lemma 5.1. *Let $\gamma \in (2, 3)$, $\delta \in \mathbb{N}$, and $\zeta \in (0, 1)$. Let $\omega = \omega(n)$ be any function tending to infinity as $n \rightarrow \infty$. Then, w.h.p. the maximum degree of \mathcal{A} is at most $\min(n^\zeta, n^{1/(\gamma-1)}\omega)$.*

Proof. Trivially, by definition, the maximum degree is at most $D = n^\zeta$. Hence, the desired property holds (deterministically) if $\zeta \leq 1/(\gamma-1)$. Let us then assume that $\zeta > 1/(\gamma-1)$. Our goal is to show that w.h.p. the maximum degree of \mathcal{A} is at most $K := n^{1/(\gamma-1)}\omega$. Note that the expected number of nodes with degree at least K is equal to

$$n \sum_{k=K}^D q_k = n \frac{\int_K^{D+1} x^{-\gamma} dx}{\int_\delta^{D+1} x^{-\gamma} dx} = n \frac{K^{1-\gamma} - (D+1)^{1-\gamma}}{\delta^{1-\gamma} - (D+1)^{1-\gamma}} = \mathcal{O}(n K^{-(\gamma-1)}) = \mathcal{O}(\omega^{-(\gamma-1)}) = o(1).$$

Hence, by the first moment method, w.h.p. there is no node with degree at least K and the proof of the lemma is finished. \square

Now, we are ready to investigate the degree distribution.

Lemma 5.2. *Let $\gamma \in (2, 3)$, $\delta \in \mathbb{N}$, and $\zeta \in (0, \frac{1}{\gamma-1}]$. For $k \in \mathbb{N}$, let Y_k be the random variable counting the number of nodes in \mathcal{A} that are of degree k . For $k \in \mathbb{N}$ and $\eta = \eta(n)$, let Y_k^η be the random variable counting the number of nodes in \mathcal{A} that are of degree at least k but at most $(1+\eta)k$.*

The following properties hold w.e.p.:

(a) If $\zeta \in (0, \frac{1}{\gamma})$, then for any $k \in \mathbb{N}$ such that $\delta \leq k \leq n^\zeta$ we have

$$\begin{aligned} Y_k &= (1 + \mathcal{O}((\log n)^{-1})) nq_k \\ &= (1 + \mathcal{O}((\log n)^{-1})) n \left(k^{-(\gamma-1)} - (k+1)^{-(\gamma-1)} \right) \delta^{\gamma-1} \\ &= (1 + \mathcal{O}(k^{-1}) + \mathcal{O}((\log n)^{-1})) nk^{-\gamma}(\gamma-1)\delta^{\gamma-1}. \end{aligned} \quad (10)$$

(b) If $\zeta \in [\frac{1}{\gamma}, \frac{1}{\gamma-1})$, then for any $k \in \mathbb{N}$ such that $\delta \leq k \leq n^{1/\gamma}(\log n)^{-4/\gamma} \ll n^\zeta$ the random variable Y_k satisfies (10), and for any $k \in \mathbb{N}$ such that $n^{1/\gamma}(\log n)^{-4/\gamma} \leq k \leq n^\zeta/(1+\eta)$ we have

$$\begin{aligned} Y_k^\eta &= (1 + \mathcal{O}((\log n)^{-1})) n\eta kq_k \\ &= (1 + \mathcal{O}((\log n)^{-1})) (\gamma-1)\delta^{\gamma-1}(\log n)^4, \end{aligned} \quad (11)$$

where

$$\eta = \eta(k) = n^{-1}(\log n)^4 k^{\gamma-1} = \mathcal{O}((\log n)^{-1}) = o(1).$$

(c) If $\zeta = \frac{1}{\gamma-1}$, then for any $k \in \mathbb{N}$ such that $\delta \leq k \leq n^{1/\gamma}(\log n)^{-4/\gamma} \ll n^\zeta$ the random variable Y_k satisfies (10), and for any $k \in \mathbb{N}$ such that $n^{1/\gamma}(\log n)^{-4/\gamma} \leq k \leq n^\zeta(\log n)^{-5/(\gamma-1)}$ the random variable Y_k^η satisfies (11). The number of nodes of degree at least $n^\zeta(\log n)^{-5/(\gamma-1)}$ is equal to $\Theta((\log n)^5)$.

Before we move to the proof of this lemma, let us mention the following straightforward corollary. Recall that for a given set of nodes $A \subseteq V$, the *volume* of A is defined as follows:

$$\text{vol}(A) = \sum_{v \in A} \text{deg}(v).$$

In particular, $\text{vol}(V) = 2|E|$.

Corollary 5.3. *The volume of all nodes in \mathcal{A} is w.e.p. equal to*

$$\text{vol}(V) = \sum_{k=\delta}^D kY_k = (1 + \mathcal{O}((\log n)^{-1})) dn, \quad \text{where } d := \sum_{k=\delta}^D kq_k.$$

Unfortunately, there is no closed formula for the constant d but one can easily approximate it numerically and obtain some theoretical bounds. For example, note that

$$\begin{aligned} d &= \sum_{k=\delta}^D k \frac{\int_k^{k+1} x^{-\gamma} dx}{\int_\delta^{D+1} x^{-\gamma} dx} \leq \sum_{k=\delta}^D \frac{\int_k^{k+1} x^{1-\gamma} dx}{\int_\delta^{D+1} x^{-\gamma} dx} = \frac{\int_\delta^{D+1} x^{1-\gamma} dx}{\int_\delta^{D+1} x^{-\gamma} dx} = \frac{(\delta^{2-\gamma} - (D+1)^{2-\gamma})/(\gamma-2)}{(\delta^{1-\gamma} - (D+1)^{1-\gamma})/(\gamma-1)} \\ &= (1 + \mathcal{O}(n^{-\zeta(\gamma-2)})) \delta \frac{\gamma-1}{\gamma-2}. \end{aligned}$$

On the other hand,

$$\begin{aligned} d &= \sum_{k=\delta}^D k \frac{\int_k^{k+1} x^{-\gamma} dx}{\int_\delta^{D+1} x^{-\gamma} dx} \geq \sum_{k=\delta}^D \frac{k}{k+1} \cdot \frac{\int_k^{k+1} x^{1-\gamma} dx}{\int_\delta^{D+1} x^{-\gamma} dx} \geq \frac{\delta}{\delta+1} \cdot \frac{\int_\delta^{D+1} x^{1-\gamma} dx}{\int_\delta^{D+1} x^{-\gamma} dx} \\ &= (1 + \mathcal{O}(n^{-\zeta(\gamma-2)})) \frac{\delta^2}{\delta+1} \cdot \frac{\gamma-1}{\gamma-2}. \end{aligned}$$

Proof of Lemma 5.2. For now, suppose that the degree sequence is simply a sequence of n independent random variables, each of them following power-law distribution $\mathcal{P}(\gamma, \delta, \zeta)$. In the end, the degree of a node of the largest degree might possibly be decreased by one to make sure the sum of the degrees is even. Clearly, this small adjustment will not affect out asymptotic bounds.

We will call a node *light* if its degree is at most $K := n^{1/\gamma}(\log n)^{-4/\gamma}$; otherwise, it will be called *heavy*. Note that if $\zeta < 1/\gamma$, then trivially all nodes are light. On the other hand, if $\zeta \geq 1/\gamma$, then (as will be shown soon) *w.h.p.* the maximum degree is equal to $(1 + o(1))n^\zeta$ and so we will have a mixture of light and heavy nodes. We need to make a distinction because for a given value of k , $\delta \leq k \leq K$, there are plenty of light nodes with their degrees are equal to k , that is, we will show that *w.e.p.* $Y_k > 0$. Unfortunately, for a given value of k , $K < k \leq n^\zeta$, it might happen (in fact, it is quite often the case) that there will be no heavy node that has its degree equal to k , that is, $Y_k = 0$. In order to solve this problem, we will need to group heavy nodes together and consider them in batches, that is, consider Y_k^η instead of Y_k .

We start with light nodes. Let us fix $k \in \mathbb{N}$ such that $\delta \leq k \leq K$. Using (9), we get that

$$\mathbb{E}[Y_k] = nq_k = \Theta(nk^{-\gamma}) = \Omega(nK^{-\gamma}) = \Omega((\log n)^4).$$

After applying Chernoff's bound (5) with $\varepsilon = (\log n)^{-1}$ we get that *w.e.p.*

$$Y_k = (1 + \mathcal{O}((\log n)^{-1}))\mathbb{E}[Y_k] = (1 + \mathcal{O}((\log n)^{-1}))nq_k.$$

The other two equalities in (10) follow immediately from (8) and (9). This finishes part (a) and the first half of part (b).

We move now to heavy nodes. Assume that $\zeta \geq 1/\gamma$ and let us fix $k \in \mathbb{N}$ such that $K \leq k \leq n^\zeta/(1 + \eta)$. If $\zeta = 1/(\gamma - 1)$ (part (c)), then we additionally assume that $k \leq n^\zeta(\log n)^{-5/(\gamma-1)}$. (We will independently deal with nodes of degrees that are very close to the maximum degree.) Recall that

$$\eta = \eta(k) = n^{-1}(\log n)^4 k^{\gamma-1}.$$

Recall also that the random variable Y_k^η counts a *batch* of nodes with degrees at least k but at most $(1 + \eta)k$. Note that $\eta(k)$ is an increasing function of k . Moreover, $\eta(K) = 1/K$ so only a few values are considered for $k = K$ (note that $\eta(K)K = 1$). On the other extreme, $\eta(n^\zeta) = n^{-1+\zeta(\gamma-1)}(\log n)^4$ which is $\mathcal{O}((\log n)^{-1}) = o(1)$, provided that $\zeta < 1/(\gamma - 1)$. If $\zeta = 1/(\gamma - 1)$ (part (c)), then the additional upper bound for k also guarantees that at the extreme case $\eta(n^\zeta(\log n)^{-5/(\gamma-1)}) = (\log n)^{-1} = o(1)$. In any case, if $\eta = \mathcal{O}((\log n)^{-1})$, then all nodes in the batch have degrees equal to $(1 + \mathcal{O}(\eta))k = (1 + \mathcal{O}((\log n)^{-1}))k$. Using (9) as before, we get that

$$\mathbb{E}[Y_k^\eta] = n \sum_{i=k}^{(1+\eta)k} q_i = \Theta(n(\eta k)k^{-\gamma}) = \Theta((\log n)^4).$$

After applying Chernoff's bound (5) with $\varepsilon = (\log n)^{-1}$ we get that *w.e.p.*

$$Y_k^\eta = (1 + \mathcal{O}((\log n)^{-1}))\mathbb{E}[Y_k^\eta] = (1 + \mathcal{O}((\log n)^{-1})) n\eta k q_k.$$

The second equality in (11) follows immediately from (9). This finishes the second half of part (b).

To finish part (c), it remains to concentrate on the case $\zeta = 1/(\gamma - 1)$ and deal with nodes of degrees that are very close to the maximum degree. Since the expected number of nodes of degree at least $n^\zeta(\log n)^{-5/(\gamma-1)}$ is of order

$$n \left(n^\zeta (\log n)^{-5/(\gamma-1)} \right)^{1-\gamma} = (\log n)^5,$$

we may apply Chernoff's bound (5) with $\varepsilon = (\log n)^{-1}$ for the last time to get the desired concentration. The proof of the lemma is finished. \square

Simulation Corner

In order to see whether asymptotic results can be used to predict the behaviour for relatively small values of n , we generated the degree distributions for two **ABCD** graphs \mathcal{A} on $n = 1,000$ and, respectively, $n = 1,000,000$ nodes. In both cases, we used parameters $\gamma = 2.5$, $\delta = 5$, and $\zeta = 1/2 < 2/3 = 1/(\gamma - 1)$ (that is, $D = \sqrt{n}$). In Figure 1 we plot the complement of the cumulative degree distribution (that is, the fraction of nodes of degree at least K) and compare it with asymptotic, theoretical predictions, namely, the function

$$\sum_{k=K}^D q_k = \frac{\int_K^{D+1} x^{-\gamma} dx}{\int_{\delta}^{D+1} x^{-\gamma} dx} = \frac{K^{1-\gamma} - (D+1)^{1-\gamma}}{\delta^{1-\gamma} - (D+1)^{1-\gamma}}.$$

We observe almost perfect agreement, especially for a larger graph. As mentioned earlier, the theoretical distribution uses the continuous model whereas when generating **ABCD** we used the discrete distribution. As a consequence, we see a slight deviation between the curves indicating that the theoretical continuous model generates slightly larger node degrees.

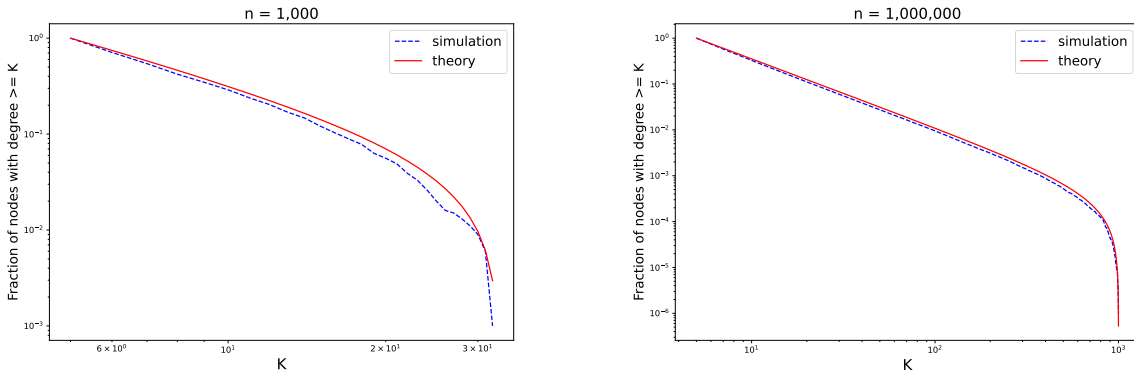


Figure 1: Complement of cumulative degree distribution for small ($n = 1,000$; left plot) and large ($n = 1,000,000$; right plot) graphs using the following parameters: $\gamma = 2.5$, $\delta = 5$, and $\zeta = 1/2$.

Our next experiment investigates how well Corollary 5.3 predicts the volume of \mathcal{A} in practice. For each value of $n = 1000 \cdot 2^i$, $i \in \{0, 1, \dots, 15\}$, we independently generated 100 graphs with the same parameters as in the first experiment. In Figure 2 we present the average value and the standard deviation of $\text{vol}(V)$. We use two different scalings: dn , the theoretical asymptotic prediction using continuous power-law distribution, and $\hat{d}n$ where $\hat{d} = \sum_{k=\delta}^D kr_k$ (see (3) for a definition of r_k), the discrete counterpart of d . Since the discrete distribution is used by default by the **ABCD** generator, in line what we observed in Figure 1, the continuous prediction is a little bit off but the discrete one works well, especially for large graphs.

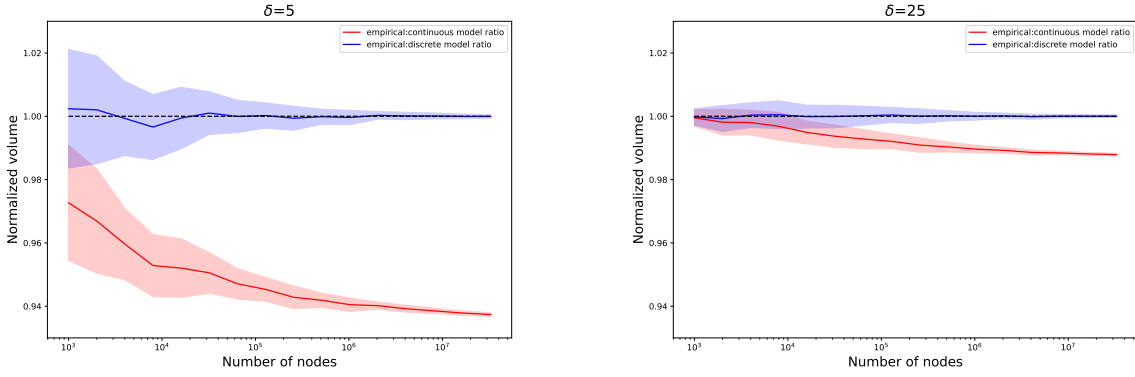


Figure 2: The average volume of 100 independently generated graphs; the shaded area represents the standard deviation. Both quantities were normalized by the theoretical prediction for the continuous model (red curve) and the discrete one (blue curve). The dashed line at 1 corresponds to a perfect prediction. Parameters used: $\gamma = 2.5$, $\delta = 5$, and $\zeta = 1/2$ (left plot), and $\gamma = 2.5$, $\delta = 25$, and $\zeta = 1/2$ (right plot).

5.2 Distribution of the Community Sizes

Let $\beta \in (1, 2)$, $s \in \mathbb{N}$, and $\tau \in (\zeta, 1)$. Recall that community sizes of the **ABCD** model are generated randomly following the (truncated) *power-law distribution* $\mathcal{P}(\beta, s, \tau)$ with exponent β , minimum value s , and maximum value $S = n^\tau$. More precisely, if $X \in \mathcal{P}(\beta, s, \tau)$, then after following exactly the same computation as in (8) and (9) we get that for any $k \in \{s, s+1, \dots, S\}$,

$$\begin{aligned}
 p_k &= \Pr(X = k) = \frac{\int_k^{k+1} x^{-\beta} dx}{\int_s^{S+1} x^{-\beta} dx} = \frac{k^{1-\beta} - (k+1)^{1-\beta}}{s^{1-\beta} - (S+1)^{1-\beta}} \\
 &= (1 + \mathcal{O}(n^{-\tau(\beta-1)})) \left(k^{-(\beta-1)} - (k+1)^{-(\beta-1)} \right) s^{\beta-1} \quad (12)
 \end{aligned}$$

$$= (1 + \mathcal{O}(n^{-\tau(\beta-1)}) + \mathcal{O}(k^{-1})) k^{-\beta} (\beta - 1) s^{\beta-1}. \quad (13)$$

Community sizes of **ABCD** are generated randomly and independently, each of them following power-law distribution $\mathcal{P}(\beta, s, \tau)$. Our first task is to investigate how many communities there are in a graph on n nodes, and how their sizes are distributed. It is slightly easier to fix the number of communities, generate their sizes independently, and check how large graph they span. Once we establish this, we will simply inverse the process, fix the number of nodes of the graph to be n and compute the number of communities ℓ that need to be generated to reach the desired number of nodes.

The assumption that $\tau > \zeta$ is introduced to make sure large degree nodes have large enough communities to be assigned to. We do not need this assumption in the first lemma as it is concerned with a sequence of random variables, not the **ABCD** model.

Lemma 5.4. *Let $\beta \in (1, 2)$, $s \in \mathbb{N}$, and $\tau \in (0, 1)$. Let*

$$\ell = \ell(n) = cn^{1-\tau(2-\beta)}$$

for some function $c = c(n) \rightarrow \hat{c}$, where $\hat{c} \in \mathbb{R}_+$. Let $(X_1, X_2, \dots, X_\ell)$ be a sequence of independent random variables, $X_i \in \mathcal{P}(\beta, s, \tau)$ for any $i \in [\ell]$. For $k \in \mathbb{N}$, let Y_k be the random variable counting the number of variables X_i that are equal to k . For $k \in \mathbb{N}$ and $\eta = \eta(n)$, let Y_k^η be the random variable counting the number of variables X_i that are at least k but at most $(1 + \eta)k$.

The following properties hold w.e.p.:

(a) If $\tau \in (0, 1/2)$, then for any $k \in \mathbb{N}$ such that $s \leq k \leq n^\tau$ we have

$$\begin{aligned} Y_k &= (1 + \mathcal{O}((\log n)^{-1})) \ell p_k \\ &= (1 + \mathcal{O}((\log n)^{-1})) cn^{1-\tau(2-\beta)} \left(k^{-(\beta-1)} - (k+1)^{-(\beta-1)} \right) s^{\beta-1} \\ &= (1 + \mathcal{O}(k^{-1}) + \mathcal{O}((\log n)^{-1})) cn^{1-\tau(2-\beta)} k^{-\beta} (\beta-1) s^{\beta-1}. \end{aligned} \quad (14)$$

(b) If $\tau \in [1/2, 1)$, then for any $k \in \mathbb{N}$ such that $s \leq k \leq n^{\tau-(2\tau-1)/\beta} (\log n)^{-4/\beta} \ll n^\tau$ the random variable Y_k satisfies (14), and for any $k \in \mathbb{N}$ such that $n^{\tau-(2\tau-1)/\beta} (\log n)^{-4/\beta} \leq k \leq n^\tau / (1 + \delta)$ we have

$$\begin{aligned} Y_k^\eta &= (1 + \mathcal{O}((\log n)^{-1})) \ell \eta k p_k \\ &= (1 + \mathcal{O}((\log n)^{-1})) c(\beta-1) s^{\beta-1} (\log n)^4, \end{aligned}$$

where

$$\eta = \eta(k) = n^{\tau(2-\beta)-1} (\log n)^4 k^{\beta-1} \leq n^{-(1-\tau)} (\log n)^4 = \mathcal{O}((\log n)^{-1}) = o(1).$$

In any case,

$$\sum_{i=1}^{\ell} X_i = (1 + \mathcal{O}((\log n)^{-1})) cns^{\beta-1} \frac{\beta-1}{2-\beta} = (1 + o(1))n, \quad (15)$$

provided

$$\hat{c} = \frac{2-\beta}{(\beta-1)s^{\beta-1}}.$$

Before we prove the lemma, let us state the following straightforward implication for the **ABCD** model.

Corollary 5.5. *Let $\beta \in (1, 2)$, $s \in \mathbb{N}$, and $\tau \in (\zeta, 1)$. The following properties hold w.e.p. for \mathcal{A} .*

(a) *The number of communities is equal to*

$$\ell = \ell(n) = (1 + \mathcal{O}((\log n)^{-1})) \hat{c} n^{1-\tau(2-\beta)},$$

where

$$\hat{c} = \frac{2-\beta}{(\beta-1)s^{\beta-1}}.$$

(b) *For $k \in \mathbb{N}$, let Y_k be the number of communities of size k . For $k \in \mathbb{N}$ and $\eta = \eta(n)$, let Y_k^η be the number of communities of size at least k but at most $(1 + \eta)k$. The random variables Y_k and Y_k^η satisfy properties (a) and (b) in Lemma 5.4.*

Proof. Recall that community sizes are generated independently following power-law distribution $\mathcal{P}(\beta, s, \tau)$ until the sum of their sizes reaches n . By Lemma 5.4 (see (15)), *w.e.p.* after generating

$$\ell_\ell = \ell_\ell(n) = (1 - a(\log n)^{-1}) \hat{c} n^{1-\tau(2-\beta)}$$

communities, the sum is still below n provided that a is a large enough constant. On the other hand, after generating

$$\ell_u = \ell_u(n) = (1 + a(\log n)^{-1}) \hat{c} n^{1-\tau(2-\beta)}$$

communities, *w.e.p.* the sum exceeds n , again, provided that a is large enough. It follows that *w.e.p.* $\ell_\ell \leq \ell \leq \ell_u$. This finishes part (a) of the corollary.

Recall that once the desired number of nodes is reached, either the last community is “trimmed” or at most $s - 1 = O(1)$ communities increase their sizes. To see that part (b) holds, note that the bounds for Y_k and Y_k^η implied by Lemma 5.4 are asymptotically the same, regardless whether the lemma is applied for ℓ_ℓ or for ℓ_u . Moreover, these bounds are of order at least $(\log n)^4$ and so affecting $O(1)$ communities at the end of the process will not affect the final bounds. The proof of the corollary is finished. \square

Now we are ready to get back to Lemma 5.4. The proof of the lemma is a straightforward adaptation of the one of Lemma 5.2.

Proof of Lemma 5.4. We will call a random variable X_i *light* if $X_i \leq K$, where

$$K := \min \left(n^{\tau-(2\tau-1)/\beta} (\log n)^{-4/\beta}, n^\tau \right);$$

otherwise, X_i will be called *heavy*. Note that if $\tau < 1/2$, then $K = n^\tau$ so all variables are light. On the other hand, if $\tau \geq 1/2$, then $K = n^{\tau-(2\tau-1)/\beta} (\log n)^{-4/\beta} \ll n^\tau$ and so we will have a mixture of light and heavy variables.

We start with light variables. Let us fix $k \in \mathbb{N}$ such that $s \leq k \leq K$. Using (13), we get that

$$\mathbb{E}[Y_k] = \ell p_k = \Theta(n^{1-\tau(2-\beta)} k^{-\beta}) = \Omega(n^{1-\tau(2-\beta)} K^{-\beta}) = \Omega((\log n)^4).$$

After applying Chernoff’s bound (5) with $\varepsilon = (\log n)^{-1}$ we get that *w.e.p.*

$$Y_k = (1 + \mathcal{O}((\log n)^{-1})) \mathbb{E}[Y_k] = (1 + \mathcal{O}((\log n)^{-1})) \ell p_k.$$

The other two equalities in (14) follow immediately from (12) and (13). This finishes part (a) and the first half of part (b).

We move now to heavy variables. Assume that $\tau \geq 1/2$ and let us fix $k \in \mathbb{N}$ such that $K \leq k \leq n^\tau / (1 + \eta)$. Let us also fix

$$\eta = \eta(k) = n^{\tau(2-\beta)-1} (\log n)^4 k^{\beta-1}.$$

Recall that the random variable Y_k^η counts a *batch* of variables X_i that are at least k but at most $(1 + \eta)k$. Note that $\eta(k)$ is an increasing function of k . Moreover, $\eta(K) = 1/K$ so only a few values are considered for $k = K$ (note that $\eta(K)K = 1$). On the other extreme, $\eta(n^\tau) = n^{-(1-\tau)} (\log n)^4$ which is still $\mathcal{O}((\log n)^{-1}) = o(1)$. It implies that all variables in the batch are equal to $(1 + \mathcal{O}(\eta))k = (1 + \mathcal{O}((\log n)^{-1}))k$. As before, using (13) we get that

$$\mathbb{E}[Y_k^\eta] = \ell \sum_{i=k}^{(1+\eta)k} p_k = \Theta \left(n^{1-\tau(2-\beta)} (\eta k) k^{-\beta} \right) = \Omega((\log n)^4).$$

After applying Chernoff's bound (5) with $\varepsilon = (\log n)^{-1}$ we get that *w.e.p.*

$$Y_k^\eta = (1 + \mathcal{O}((\log n)^{-1}))\mathbb{E}[Y_k^\eta] = (1 + \mathcal{O}((\log n)^{-1})) \ell \eta k p_k.$$

This finishes the second half of part (b).

It remains to show (15). Since we aim for a statement that holds *w.e.p.*, we may assume that parts (a) and (b) hold. First, let us note that

$$\sum_{i=1}^{\ell} X_i = \sum_{k=s}^{n^\tau} k Y_k.$$

Light variables can be dealt with immediately: for any $k \in \mathbb{N}$ such that $s \leq k \leq K$ we have

$$k Y_k = (1 + \mathcal{O}((\log n)^{-1})) k \ell p_k.$$

For heavy variables we again use the fact that all variables in one batch have asymptotically the same value. For any $k \in \mathbb{N}$ such that $K \leq k \leq n^\tau / (1 + \eta)$ we have

$$\begin{aligned} \sum_{i=k}^{(1+\eta)k} i Y_i &= (1 + \mathcal{O}(\eta)) k \sum_{i=k}^{(1+\eta)k} Y_i = (1 + \mathcal{O}((\log n)^{-1})) k Y_k^\eta \\ &= (1 + \mathcal{O}((\log n)^{-1})) k \ell \eta k p_k = (1 + \mathcal{O}((\log n)^{-1})) \sum_{i=k}^{(1+\eta)k} i \ell p_i. \end{aligned}$$

It follows that

$$\sum_{i=1}^{\ell} X_i = \sum_{k=s}^{n^\tau} k Y_k = (1 + \mathcal{O}((\log n)^{-1})) \sum_{k=s}^{n^\tau} k \ell p_k.$$

Now, note that the contribution from the first $\log n$ terms is negligible:

$$\begin{aligned} \sum_{k=s}^{\log n} k \ell p_k &= \mathcal{O} \left(n^{1-\tau(2-\beta)} \sum_{k=s}^{\log n} k^{1-\beta} \right) = \mathcal{O} \left(n^{1-\tau(2-\beta)} \int_s^{\log n} x^{1-\beta} dx \right) \\ &= \mathcal{O} \left(n^{1-\tau(2-\beta)} (\log n)^{2-\beta} \right) = \mathcal{O}(n / \log n). \end{aligned}$$

On the other hand,

$$\begin{aligned} \sum_{k=\log n}^{n^\tau} k \ell p_k &= (1 + \mathcal{O}((\log n)^{-1})) c n^{1-\tau(2-\beta)} (\beta - 1) s^{\beta-1} \sum_{k=\log n}^{n^\tau} k^{1-\beta} \\ &= (1 + \mathcal{O}((\log n)^{-1})) c n^{1-\tau(2-\beta)} (\beta - 1) s^{\beta-1} \int_{\log n}^{n^\tau} x^{1-\beta} dx + \mathcal{O}(1) \\ &= (1 + \mathcal{O}((\log n)^{-1})) c n^{1-\tau(2-\beta)} (\beta - 1) s^{\beta-1} \frac{n^{\tau(2-\beta)} - (\log n)^{2-\beta}}{2 - \beta} \\ &= (1 + \mathcal{O}((\log n)^{-1})) c n s^{\beta-1} \frac{\beta - 1}{2 - \beta}. \end{aligned}$$

This finishes the proof of (15), and so the theorem holds. \square

Simulation Corner

We generated community sizes for two **ABCD** graphs \mathcal{A} on $n = 1,000$ and, respectively, $n = 1,000,000$ nodes. In both cases, we used parameters $\beta = 1.5$, $s = 50$, and $\tau = 3/4$ (that is, $S = n^{3/4}$). In Figure 3 we plot the complement of the cumulative distribution (that is, the fraction of communities that consist of at least K nodes) and compare it with asymptotic, theoretical predictions, namely, the function

$$\sum_{k=K}^S p_k = \frac{\int_K^{S+1} x^{-\beta} dx}{\int_s^{S+1} x^{-\beta} dx} = \frac{K^{1-\beta} - (S+1)^{1-\beta}}{s^{1-\beta} - (S+1)^{1-\beta}}.$$

Simulation results are averaged over 30 independent runs in view of the small number of communities compared to n . Note that since $s = 50$ is much larger than $\delta = 5$, in Figure 1 we do not observe a large difference between theoretical and simulated curves (the largest discrepancy between the continuous (q_k) and the discrete (r_k) distributions is observed for small values of k).

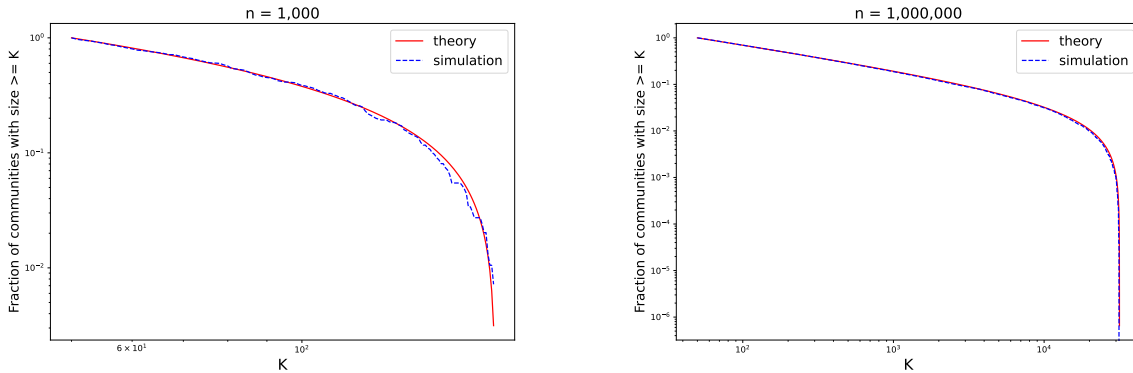


Figure 3: Complement of cumulative community size distribution for small ($n = 1,000$; left plot) and large ($n = 1,000,000$; right plot) graphs. Parameter used: $\beta = 1.5$, $s = 50$, and $\tau = 3/4$.

To see how well Corollary 5.5 predicts the number of communities in practice, for each value of $n = 1000 \cdot 2^i$, $i \in \{0, 1, \dots, 15\}$, we independently generated 100 graphs with the same parameters as in the above experiment and three other sets of parameters. In Figure 4 we present the average number of communities and its standard deviation. In our theoretical results, we do not pay attention to signs of error terms as all of them tend to zero as $n \rightarrow \infty$. However, based on the simulation, it seems that the error term is negative for $\beta < 1.5$ and positive for $\beta > 1.5$ but, as expected, it tends to zero asymptotically. Additionally, note that the standard deviation of the normalized number of communities is significant even for large graphs, as opposed to the standard deviation of the normalized volume presented in Figure 2.

5.3 Assigning Nodes into Communities and Distribution of Weights

Recall that at this point of the process, the degree distribution ($w_1 \geq w_2 \geq \dots \geq w_n$) and the distribution of community sizes ($c_1 \geq c_2 \geq \dots \geq c_\ell$) are already fixed. In order to assign nodes to communities we will use the following easy and natural algorithm. We consider nodes, one by one, starting from w_1 (high degree node) and finishing with w_n (low degree node). Recall that node i of

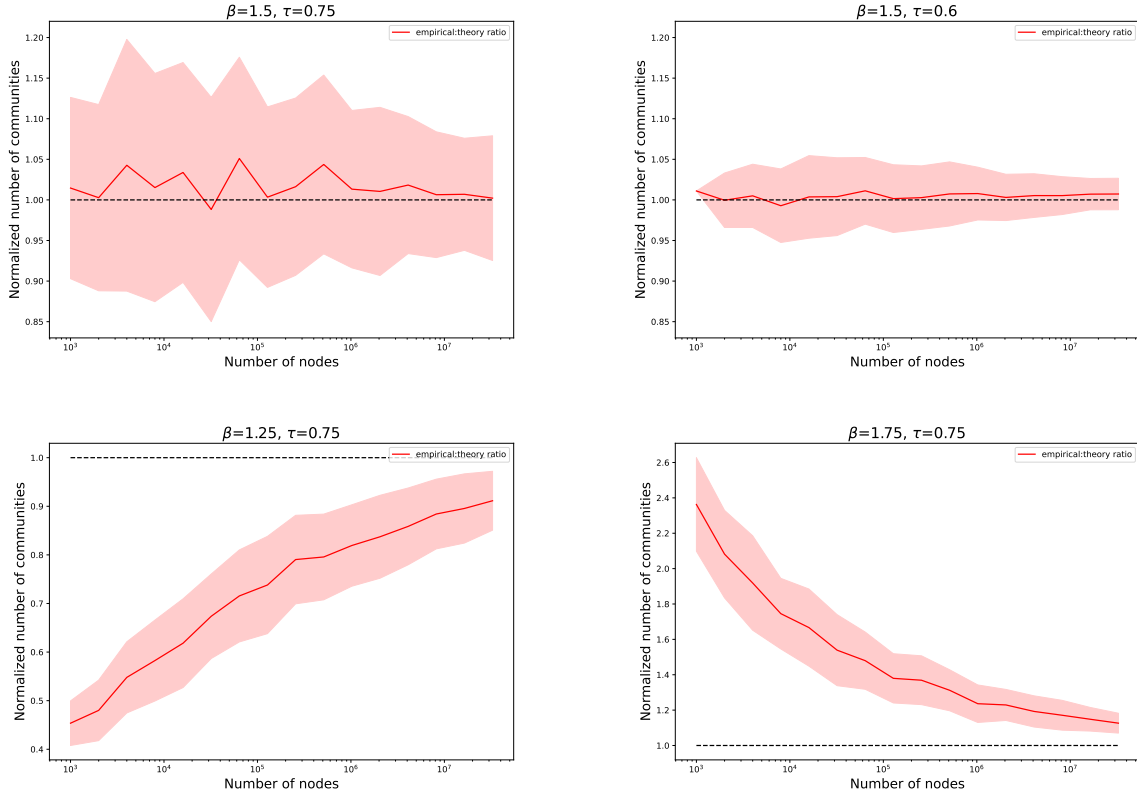


Figure 4: The average number of communities for 100 independently generated graphs; the shaded area represents the standard deviation. Both quantities were normalized by the theoretical prediction. The dashed line at 1 corresponds to a perfect prediction. All plots are with $s = 50$. The top two plots are shown with the same range on the y axis for comparison, i.e. decreasing τ lowers the variability. In the bottom plots, we see that convergence is slower for values of β away from 1.5.

degree w_i has to be assigned to a community of size c_j so that inequality (1) holds. We assign node w_i randomly to one of the communities that have size larger than $\lceil (1 - \xi\phi)w_i \rceil$ and still have some “available spots”. We do it with probability proportional to the number of available spots left.

Note that it follows immediately from Corollary 5.5 that *w.e.p.* $\phi = 1 - \sum_{k \in [\ell]} (c_k/n)^2 = 1 - o(1)$, since

$$\begin{aligned} \sum_{k \in [\ell]} (c_k/n)^2 &= \sum_{k=s}^{n^\tau} Y_k (k/n)^2 = \sum_{k=s}^{n^\tau} \mathcal{O} \left(n^{1-\tau(2-\beta)} k^{-\beta} (k/n)^2 \right) \\ &= \mathcal{O} \left(n^{-1-\tau(2-\beta)} \sum_{k=s}^{n^\tau} k^{2-\beta} \right) = \mathcal{O} \left(n^{-1-\tau(2-\beta)} n^{\tau(3-\beta)} \right) = \mathcal{O} \left(n^{-(1-\tau)} \right) = o(1). \end{aligned}$$

In order to see that the above simple algorithm generates one of the admissible assignments uniformly at random, let t_i be the number of available spots (at the beginning of the process) for

node i , that is, the total number of nodes that belong to communities satisfying inequality (1):

$$t_i = \sum_{j \in I_i} c_j, \quad \text{where } I_i = \left\{ j \in [\ell] : w_i \leq \frac{c_j - 1}{1 - \xi\phi} \right\} = \left\{ j \in [\ell] : c_j \geq (1 - \xi\phi)w_i + 1 \right\}.$$

Since nodes are considered in non-increasing order of their degrees, exactly $i - 1$ of these t_i available spots are taken by other nodes when it is time for node i to be assigned to some community.

To see that *w.e.p.* $t_i \geq i$, consider any node of degree k such that $\delta \leq k \leq n^\zeta (\log n)^{-5/(\gamma-1)}$. It follows from Lemma 5.2 that when this node is considered by the algorithm, *w.e.p.* its index i satisfies the following property:

$$i = \Theta \left(\sum_{i=\delta}^k nq_i \right) = \Theta \left(n \sum_{i=\delta}^k i^{-\gamma} \right) = \Theta \left(nk^{-(\gamma-1)} \right).$$

On the other hand, by Corollary 5.5, *w.e.p.*

$$t_i = n - \Theta \left(\sum_{i=s}^{(1-\xi\phi)k} lip_i \right) = n - \Theta \left(n^{1-\tau(2-\beta)} \sum_{i=s}^{(1-\xi\phi)k} i^{1-\beta} \right) = n - \Theta \left(n \left(\frac{k}{n^\tau} \right)^{2-\beta} \right).$$

Both i and t_i are decreasing functions of k but clearly $t_i \geq i$. In fact, note that *w.e.p.* $t_1 = n(1 - o(1))$ so almost all nodes are available right from the very beginning when a node of maximum degree is considered. Since *w.e.p.* there are linearly many nodes of degree δ , we process all nodes of degrees at least $\delta + 1$ before there is potentially a problem.

It follows that each admissible assignment is used with probability equal to

$$\prod_{i=1}^n \frac{1}{t_i - (i - 1)},$$

which is a fixed number that depends on the sequences $(w_1 \geq w_2 \geq \dots \geq w_n)$ and $(c_1 \geq c_2 \geq \dots \geq c_\ell)$ but not on the choice of admissible assignment. Since the algorithm cannot produce any non-admissible assignment, this shows uniformity.

We say that the community is *large* if its size is at least $(\log n)^8$ and *very large* if it is of size at least $n^\zeta (\log n)^4$; otherwise, it will be called *small*. The volumes of small communities are not well concentrated around their means. For example, a community $C \subseteq V$ of constant size ($|C| = O(1)$) has to have average degree satisfying

$$\delta \leq \frac{\text{vol}(C)}{|C|} \leq \left\lfloor \frac{|C| - 1}{1 - \xi\phi} \right\rfloor,$$

but it could be equal to any of the two extreme values with probability bounded away from zero. Hence, in this situation there is no hope for determining the asymptotic value for its volume that holds *w.e.p.* On the other hand, the volumes of very large communities are well concentrated around their means, as the next lemma shows.

Recall that each node has its degree w_i randomly split into community degree y_i and background degree z_i . For non-leaders, we have precisely $\mathbb{E}[y_i] = (1 - \xi)w_i$ but the leaders might require adjustment

implying that $\mathbb{E}[y_i] = (1 - \xi)w_i + O(1)$ (which is, of course, negligible from our asymptotic results point of view). For any community $C \subseteq V$, let

$$\text{vol}_c(C) := \sum_{i \in C} y_i$$

be the community volume. We expect a $(1 - \xi)$ fraction of the total volume of each community to be assigned to community degrees, that is, $\mathbb{E}[\text{vol}_c(C)] = (1 - \xi)\text{vol}(C) + O(1)$. Similarly to the total volumes, for very large communities we will prove that $\text{vol}_c(C)$ is well concentrated around their expectations.

Now we are ready to state the next lemma.

Lemma 5.6. *Let $C \subseteq V$ be any large community in \mathcal{A} , that is, a community of size $|C| \geq (\log n)^8$. Then,*

$$\begin{aligned} \mathbb{E}[\text{vol}(C)] &= (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) d|C|, \text{ and} \\ \mathbb{E}[\text{vol}_c(C)] &= (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1 - \xi)d|C|, \end{aligned} \quad \text{where } d := \sum_{k=\delta}^D kq_k.$$

Moreover, the following properties hold w.e.p. If C is very large (that is, $|C| \geq n^\zeta (\log n)^4$), then

$$\begin{aligned} \text{vol}(C) &= (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) d|C|, \text{ and} \\ \text{vol}_c(C) &= (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1 - \xi)d|C|. \end{aligned}$$

If C is large (that is, $(\log n)^8 \leq |C| < n^\zeta (\log n)^4$), then $\text{vol}(C) = \mathcal{O}(n^\zeta (\log n)^4)$.

If C is small (that is, $|C| < (\log n)^8$), then trivially $\text{vol}(C) \leq D|C| < n^\zeta (\log n)^8$.

Finally, $\text{vol}(F) = \mathcal{O}(n(\log n)^{-2})$, where F is the union of small and large communities.

Let us note that the constant d used in the above lemma is the same as in Corollary 5.3. Some useful and explicit bounds for d are provided right after the statement of this corollary.

Proof. Let us fix any community $C \subseteq V$ of size $z = |C| \geq z_0 := (\log n)^8$. If, for example, $z > n^\zeta$, then community C has enough room even for a node of the largest degree. However, if z is small then only nodes of degree at most $\left\lfloor \frac{z-1}{1-\xi\phi} \right\rfloor$ are allowed to be assigned to C . Fortunately, w.e.p. the number of nodes of degree larger than $K_0 := \left\lfloor \frac{z-1}{1-\xi\phi} \right\rfloor \geq k_0 := \left\lfloor \frac{z_0-1}{1-\xi\phi} \right\rfloor = \Omega((\log n)^8)$ is, by Lemma 5.2, equal to

$$\begin{aligned} n' &= \sum_{k > K_0} Y_k \leq \sum_{k > k_0} Y_k = (1 + \mathcal{O}((\log n)^{-1})) \sum_{k > k_0} nq_k = \mathcal{O} \left(\sum_{k > k_0} nk^{-\gamma} \right) \\ &= \mathcal{O} \left(nk_0^{1-\gamma} \right) = \mathcal{O} \left(n (\log n)^{-8(\gamma-1)} \right) = \mathcal{O} \left(n (\log n)^{-2} \right) = o(n), \end{aligned}$$

and so is negligible.

In order to track which nodes end up in community C , we need to start paying attention to the algorithm generating admissible assignments from time $n' + 1$ when considered nodes have a chance to end up in community C (because their degrees are at most K_0). Based on the observation above,

right before this happens there are only $n' = \mathcal{O}(n(\log n)^{-2})$ spots taken from communities of size at least $z + 1$. On the other hand, all the remaining nodes may be assigned to any community of size more than n^ζ , the trivial upper bound for the maximum degree. By Corollary 5.5, *w.e.p.* the total number of spots available in communities of size at most n^ζ is equal to

$$\begin{aligned} n'' &= (1 + \mathcal{O}((\log n)^{-1})) \sum_{k \leq n^\zeta} \ell k p_k = \mathcal{O} \left(\sum_{k \leq n^\zeta} n^{1-\tau(2-\beta)} k^{1-\beta} \right) \\ &= \mathcal{O} \left(n^{1-\tau(2-\beta)} n^{\zeta(2-\beta)} \right) = \mathcal{O} \left(n^{1-(\tau-\zeta)(2-\beta)} \right) = \mathcal{O} \left(n(\log n)^{-2} \right) = o(n). \end{aligned} \quad (16)$$

It follows that once $n' + t$ nodes are assigned to communities, the number of nodes assigned to community C can be stochastically upper bounded by U_t , the hypergeometric random variable with parameters $n - n' - n'' = n(1 - \mathcal{O}((\log n)^{-2}))$, z , t and lower bounded by L_t , the hypergeometric random variable with parameters $n - n' = n(1 - \mathcal{O}((\log n)^{-2}))$, z , t .

Suppose first that $t \leq n(\log n)^6/z = \mathcal{O}(n(\log n)^{-2})$. Note that $\mathbb{E}[U_t] = zt/(n - n' - n'') = \mathcal{O}((\log n)^6)$. Chernoff bound (6) applied with $u = (\log n)^6$ gives us that *w.e.p.* $U_t = \mathcal{O}((\log n)^6)$. Hence, during this phase of the algorithm, a node considered at time $t + 1$ is assigned to community C with probability

$$\begin{aligned} \exp(-\Omega((\log n)^2)) + \frac{z - \mathcal{O}(U_t)}{n(1 - \mathcal{O}((\log n)^{-2})) - t} &= \frac{z - \mathcal{O}((\log n)^6)}{n(1 - \mathcal{O}((\log n)^{-2})) - \mathcal{O}(n(\log n)^{-2})} \\ &= \frac{z}{n} (1 + \mathcal{O}((\log n)^{-2})). \end{aligned}$$

Suppose now that $n(\log n)^6/z \leq t \leq n - n(\log n)^{-1}$ so that $\mathbb{E}[U_t] \geq \mathbb{E}[L_t] = zt/(n - n') \geq (\log n)^6$. Chernoff bound (5) applied with $\varepsilon = (\log n)^{-2}$ gives us that *w.e.p.*

$$U_t = \mathbb{E}[U_t] (1 + \mathcal{O}((\log n)^{-2})) = \frac{zt}{n} (1 + \mathcal{O}((\log n)^{-2})).$$

Similarly, *w.e.p.* $L_t = (zt/n)(1 + \mathcal{O}((\log n)^{-2}))$. During this phase of the algorithm, a node considered at time $t + 1$ is assigned to community C with probability

$$\begin{aligned} \exp(-\Omega((\log n)^2)) + \frac{z - (zt/n)(1 + \mathcal{O}((\log n)^{-2}))}{n(1 - \mathcal{O}((\log n)^{-2})) - t} &= \frac{z(1 - t/n + \mathcal{O}((t/n)(\log n)^{-2}))}{(n - t)(1 - \mathcal{O}((\log n)^{-1}))} \\ &= \frac{z(1 - t/n)(1 + \mathcal{O}((\log n)^{-1}))}{(n - t)(1 - \mathcal{O}((\log n)^{-1}))} \\ &= \frac{z}{n} (1 + \mathcal{O}((\log n)^{-1})), \end{aligned}$$

since $t \leq n - n(\log n)^{-1}$. The end of the algorithm is unpredictable; we claim no bound for the probability of assigning a node to community C when $t > n - n(\log n)^{-1}$. However, fortunately, the contribution from these nodes will turn out to be negligible.

Let us now “rewind” the process and “play” it from the very beginning, this time paying attention what kind of nodes are assigned to community C . Our goal is to estimate $\text{vol}(C)$, the volume of community C of size z and $\text{vol}_c(C)$, the community volume of C . Since the adjustment needed for the leader of C changes the value of $\text{vol}_c(C)$ by at most 1 (and so is negligible), we may assume

that each node of degree w_i assigned to community C is non-leader, that is, its community degree is equal to $y_i = \lfloor (1 - \xi)w_i \rfloor$.

Let A be the set of nodes of degree larger than $K_0 = \lfloor \frac{z-1}{1-\xi\phi} \rfloor \geq k_0 = \lfloor \frac{z_0-1}{1-\xi\phi} \rfloor = \Omega((\log n)^8)$ that cannot be assigned to community C . By Lemma 5.2, *w.e.p.* the volume of A is equal to

$$\begin{aligned} \text{vol}(A) &= \sum_{k>K_0} kY_k \leq \sum_{k>k_0} kY_k = (1 + \mathcal{O}((\log n)^{-1})) \sum_{k>k_0} nkq_k = \mathcal{O} \left(\sum_{k>k_0} nk^{1-\gamma} \right) \\ &= \mathcal{O} \left(nk_0^{2-\gamma} \right) = \mathcal{O} \left(n (\log n)^{-8(\gamma-2)} \right) = o(n), \end{aligned}$$

and so is negligible comparing to the total volume that is *w.e.p.* linear (see Corollary 5.3). Let B be the set of the last $n (\log n)^{-1}$ nodes that we do not control, that is, we cannot predict the probability that a given node joins C but we will be able to say how many of them do it. Since *w.e.p.* these nodes have degree δ and so $\text{vol}(B) = \mathcal{O}(n (\log n)^{-1})$. More importantly, the expected number of spots already taken from community C when the nodes from B are about to be considered is equal to

$$\frac{z}{n} (1 + \mathcal{O}((\log n)^{-1})) \cdot (|V| - |A| - |B|) = z (1 + \mathcal{O}((\log n)^{-1})).$$

Hence, the expected number of spots left is $\mathcal{O}(z(\log n)^{-1})$ and so *w.e.p.* at most $\mathcal{O}(z(\log n)^{-1})$ are indeed left by Chernoff bound (6) applied with $u = z(\log n)^{-1} \geq (\log n)^7$. It follows that

$$\begin{aligned} \mathbb{E}[\text{vol}(C)] &= \frac{z}{n} (1 + \mathcal{O}((\log n)^{-1})) \cdot (\text{vol}(V) - \text{vol}(A) - \text{vol}(B)) + \mathcal{O}(z(\log n)^{-1}) \\ &= \text{vol}(V) \cdot \frac{z}{n} \left(1 + \mathcal{O}((\log n)^{-\min(1, 8(\gamma-2))}) \right) \\ &= dz \left(1 + \mathcal{O}((\log n)^{-(\gamma-2)}) \right). \end{aligned}$$

From this we immediately get that

$$\mathbb{E}[\text{vol}_c(C)] = (1 - \xi)\mathbb{E}[\text{vol}(C)] = (1 - \xi)dz \left(1 + \mathcal{O}((\log n)^{-(\gamma-2)}) \right).$$

The concentration for very large communities follows from Lemma 4.1. Note that if C is very large, then both $\mathbb{E}[\text{vol}(C)]$ and $\mathbb{E}[\text{vol}_c(C)]$ are of order $z = \Omega(n^\zeta (\log n)^4)$. We get the desired bound for the failure probability by applying the lemma with $\varepsilon = (\log n)^{-1}$ and $c = D = n^\zeta$. If C is large, then $\mathbb{E}[\text{vol}(C)] = \mathcal{O}(n^\zeta (\log n)^4)$. The lemma applied with $u = n^\zeta (\log n)^4$ and $c = D = n^\zeta$ gives us that *w.e.p.* $\text{vol}(C) = \mathcal{O}(n^\zeta (\log n)^4)$. Finally, by performing the same computation as in (16) we get that *w.e.p.* $|F| = \mathcal{O}(n (\log n)^{-2}) = o(n)$. It follows that $\mathbb{E}[\text{vol}(F)] = \mathcal{O}(n (\log n)^{-2})$ and so *w.e.p.* $\text{vol}(F) = \mathcal{O}(n (\log n)^{-2})$, which finishes the proof of the lemma. \square

Simulation Corner

In order to see whether asymptotic predictions of volumes of large communities can be used to predict the behaviour for all communities and for relatively small values of n , we generated two **ABCD** graphs \mathcal{A} on $n = 1,000$ and, respectively, $n = 1,000,000$ nodes. In both cases, we used parameters $\gamma = 2.5$, $\delta = 5$, $\zeta = 1/2 < 2/3 = 1/(\gamma - 1)$ (that is, $D = \sqrt{n}$), $\beta = 1.5$, $s = 50$, and $\tau = 3/4$ (that is, $S = n^{3/4}$). In Figure 5 for each generated graph \mathcal{A} we plot ℓ points (x_i, y_i) , one for each community C_i , $i \in [\ell]$: $x_i = |C_i|$ and $y_i = \text{vol}(C_i)/(\hat{d}|C_i|)$, where $\hat{d} = \sum_{k=\delta}^D kr_k$ is the discrete

counterpart of d (see (3) for a definition of r_k). As expected, larger communities in a larger graph on $n = 1,000,000$ nodes show good concentration but small graphs on $n = 1,000$ nodes are too small and deviate from the expectation even for the largest communities (that are still too small).

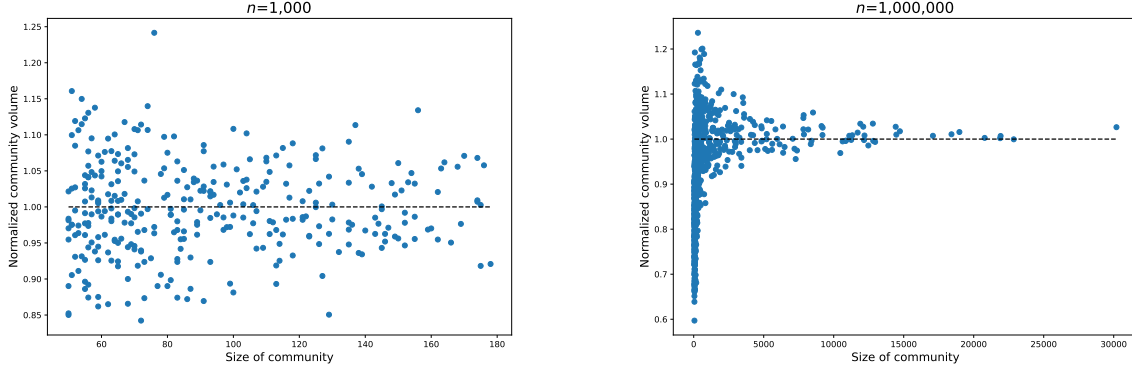


Figure 5: Volumes of communities C_i scaled by $\hat{d}|C_i|$. The dashed line at 1 corresponds to a perfect prediction. 30 independent small graphs were generated ($n = 1,000$; left plot) but only one large ($n = 1,000,000$; right plot).

6 Modularity

6.1 Modularity of the Ground-truth Partition: $q(\mathbf{C})$

Let us start by investigating the modularity of the ground-truth partition of \mathcal{A} .

Theorem 6.1. *Let $\mathbf{C} = \{C_1, C_2, \dots, C_\ell\}$ be the ground-truth partition of the set of nodes of \mathcal{A} . Then, w.e.p.*

$$q^*(\mathcal{A}) \geq q(\mathbf{C}) = (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1 - \xi).$$

Proof. Let us first estimate the degree tax. By Corollary 5.3, w.e.p. $\text{vol}(V) = (1 + \mathcal{O}((\log n)^{-1})) dn$, where $d = \sum_{k=\delta}^D kq_k$. By Lemma 5.6, w.e.p. for each community C_i we have $\text{vol}(C_i) = \mathcal{O}(n^\zeta (\log n)^8)$. It follows that w.e.p.

$$\sum_{C_i \in \mathbf{C}} \left(\frac{\text{vol}(C_i)}{\text{vol}(V)} \right)^2 = \mathcal{O}(n^{-(1-\zeta)} (\log n)^8) \sum_{C_i \in \mathbf{C}} \frac{\text{vol}(C_i)}{\text{vol}(V)} = \mathcal{O}(n^{-(1-\zeta)} (\log n)^8) = \mathcal{O}((\log n)^{-2}) = o(1),$$

and so it is negligible.

Let us now move to the edge contribution that is more challenging to estimate. We will use the terminology introduced in Lemma 5.6. In particular, we will call community C_i very large if $|C_i| \geq n^\zeta (\log n)^4$, and F is the union of communities that are not very large. By Lemma 5.6, w.e.p. $\text{vol}(F) = \mathcal{O}(n (\log n)^{-2})$ and so the contribution (to the edge contribution) from communities that are *not* very large is w.e.p. equal to

$$\sum_{C_i \in \mathbf{C}, |C_i| < n^\zeta (\log n)^4} \frac{e(C_i)}{|E|} \leq \sum_{C_i \in \mathbf{C}, |C_i| < n^\zeta (\log n)^4} \frac{\text{vol}(C_i)/2}{\text{vol}(V)/2} = \frac{\text{vol}(F)}{\text{vol}(V)} = \mathcal{O}((\log n)^{-2}) = o(1),$$

and so it is negligible. It remains to concentrate on very large communities.

Let C_i be any very large community. By definition, trivially, all edges of the community graph $G_i = (C_i, E_i)$ appear in C_i . By Lemma 5.6, *w.e.p.* the number of edges in G_i is equal to

$$\begin{aligned} |E_i| = \text{vol}_c(C_i)/2 &= (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1 - \xi)d|C_i|/2 \\ &= (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1 - \xi)\text{vol}(C_i)/2. \end{aligned}$$

To estimate the number of edges in the background graph G_0 that appear within C_i we can use the following useful property of the pairing model. One does not need to select one pairing uniformly at random from the set of all pairings but, instead, pairs of points may be chosen sequentially. Moreover, the first point may be selected using any rule whatsoever as long as the second point is chosen uniformly at random from the set of the remaining unchosen points. By Lemma 5.6, *w.e.p.* the number of points in the background graph that are associated with nodes in C_i is equal to

$$W = \text{vol}(C_i) - \text{vol}_c(C_i) = (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) \xi d|C_i|.$$

In our application, we will always select the first point of the j th pair from the set of unchosen points associated with nodes in C_i (arbitrarily). The probability that the second point is also in C_i is equal to

$$p_j = \frac{W - (j + e_j)}{\text{vol}(V) - (2j - 1)} \leq \frac{W}{\text{vol}(V)} =: p,$$

where e_j is the number of pairs of points that already appeared within C_i . (Indeed, at this point $2(j - 1) + 1 = 2j - 1$ points are already chosen, $j + e_j$ of them are associated with nodes in C_i .) Hence, the number of edges from the background graph that end up within community C_i can be stochastically upper bounded by the binomial random variable $X \in \text{Bin}(w, p)$ with $\mathbb{E}[X] = Wp = W^2/\text{vol}(V)$. If $|C_i| \geq \sqrt{n}(\log n)$, then $\mathbb{E}[X] = \Omega((\log n)^2)$ and we get from Chernoff bound (applied with $\varepsilon = 1$) that *w.e.p.*

$$X = \mathcal{O}(W^2/n) = \mathcal{O}(Wn^{-(1-\tau)}) = \mathcal{O}(W(\log n)^{-2}) = \mathcal{O}(\text{vol}(C_i)(\log n)^{-2}),$$

since, by definition, $W = \Theta(|C_i|) = \mathcal{O}(n^\tau)$. On the other hand, if $|C_i| < \sqrt{n}(\log n)$, then $\mathbb{E}[X] = \mathcal{O}((\log n)^2)$ and we get from Chernoff bound (applied with $u = (\log n)^2$) that *w.e.p.* $X = \mathcal{O}((\log n)^2)$.

Since there are clearly at most $n/(n^\zeta(\log n)^4) = n^{1-\zeta}(\log n)^{-4}$ very large communities, *w.e.p.* the total number of edges in the background graph G_0 that appear within some very large community C_i is at most

$$\begin{aligned} &\sum_{C_i \in \mathbf{C}, |C_i| \geq n^\zeta(\log n)^4} \mathcal{O}(\text{vol}(C_i)(\log n)^{-2}) + n^{1-\zeta}(\log n)^{-4} \cdot \mathcal{O}((\log n)^2) \\ &= \mathcal{O}(\text{vol}(V)(\log n)^{-2}) + \mathcal{O}(n^{1-\zeta}(\log n)^{-2}) = \mathcal{O}(n(\log n)^{-2}). \end{aligned}$$

The contribution (to the edge contribution) from very large communities is then *w.e.p.* equal to

$$\begin{aligned}
\sum_{C_i \in \mathbf{C}, |C_i| \geq n^\zeta (\log n)^4} \frac{e(C_i)}{|E|} &= \frac{\mathcal{O}(n(\log n)^{-2})}{|E|} + \sum_{C_i \in \mathbf{C}, |C_i| \geq n^\zeta (\log n)^4} \frac{|E_i|}{|E|} \\
&= \mathcal{O}((\log n)^{-2}) + (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) \sum_{C_i \in \mathbf{C}, |C_i| \geq n^\zeta (\log n)^4} \frac{(1-\xi)\text{vol}(C_i)/2}{\text{vol}(V)/2} \\
&= \mathcal{O}((\log n)^{-2}) + (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1-\xi) \frac{\text{vol}(V) - \text{vol}(F)}{\text{vol}(V)} \\
&= \mathcal{O}((\log n)^{-2}) + (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1-\xi) (1 + \mathcal{O}((\log n)^{-2})) \\
&= (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1-\xi),
\end{aligned}$$

which is the only non-negligible contribution to the modularity function. This finishes the proof of the theorem. \square

Simulation Corner

In order to see how well Theorem 6.1 predicts the modularity function $q(\mathbf{C})$ in practice, for each value of $n = 1000 \cdot 2^i$, $i \in \{0, 1, \dots, 15\}$, we independently generated 30 graphs with the same parameters as in the previous experiment: $\gamma = 2.5$, $\delta = 5$, $\zeta = 1/2 < 2/3 = 1/(\gamma - 1)$ (that is, $D = \sqrt{n}$), $\beta = 1.5$, $s = 50$, and $\tau = 3/4$ (that is, $S = n^{3/4}$). In Figure 6 we present the average value and the standard deviation of $q(\mathbf{C})$, the modularity of the ground-truth partition, for two values of ξ : $\xi = 0.2$ (low level of noise) and $\xi = 0.7$ (high level of noise). We also present the edge contribution part of $q(\mathbf{C})$, again, its average value and the standard deviation. It seems that the edge contribution for small graphs is slightly larger than the corresponding asymptotic prediction but it converges quite fast. As expected, the degree tax for small graphs is non-negligible but it converges to zero quickly. As a result, both the value of $q(\mathbf{C})$ and the edge contribution tend to $1 - \xi$ as n grows.

Additionally, in order to investigate whether there is any difference for various levels of noise, for each value of $\xi = (0.1)i$, $i \in [9]$, we independently generated 30 graphs on $n = 1,000$ nodes and $n = 1,000,000$ nodes. The results are presented in Figure 7. As observed earlier, the edge contribution for small graphs is slightly larger than $1 - \xi$, the asymptotic prediction. As expected, the difference is more visible for larger values of ξ as background graph edges in noisy graphs contribute more. On the other hand, the modularity function is closer to its asymptotic prediction for noisier graphs. Large graphs show almost perfect agreement with the asymptotic prediction.

6.2 Maximum Modularity: $q^*(G)$

As mentioned in Section 3, analyzing the maximum modularity $q^*(G)$ for sparse random graphs is a challenging task and typically only bounds for $q^*(G)$ are known that are far apart from each other. Since the **ABCD** model \mathcal{A} is more complex than other sparse random graphs, especially random d -regular graphs, there is no hope for tight bounds for the maximum modularity function but we will make some interesting observations below.

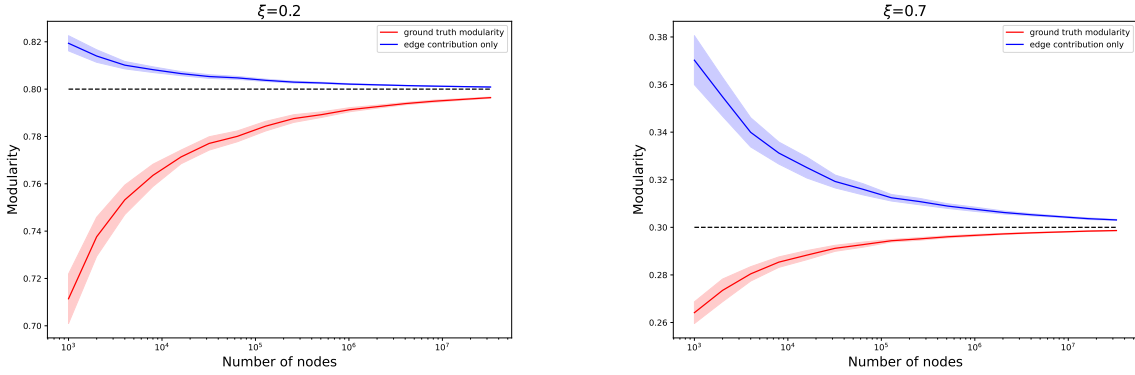


Figure 6: The modularity $q(\mathbf{C})$ of the ground-truth partition (red) and the corresponding edge contribution (blue) for 30 independently generated graphs; the shaded areas represent the standard deviation. The dashed line at $1 - \xi$ corresponds to a perfect prediction. Parameters used: $\gamma = 2.5$, $\delta = 5$, $\zeta = 1/2$, $\beta = 1.5$, $s = 50$, and $\tau = 3/4$. Two different levels of noise are investigated.

Large Level of Noise

Let us start with investigating graphs with a large level of noise, that is, with ξ close to one. For such graphs, one should focus on the background graph G_0 which involves all but a small fraction of edges. It turns out that G_0 is connected *w.h.p.*, provided that its minimum degree is at least 3, or otherwise *w.h.p.* it has a giant component. By restricting ourselves to a spanning tree of the giant component of G_0 , we may partition the set of nodes into small parts such that each part induces a connected graph. This is not much, but for noisy graphs it yields a lower bound for q^* that is strictly larger than the modularity of the ground-truth partition.

Theorem 6.2. *Let $\gamma \in (2, 3)$, $\delta \in \mathbb{N}$, $\zeta \in (0, \frac{1}{\gamma-1}]$, and $\xi \in (0, 1)$. There exists a universal constant α bounded away from 0, that can be chosen to be 1 if $\xi\delta \geq 3$ holds, and there exists a partition \mathbf{C} of the set of nodes V of \mathcal{A} such that the following properties hold w.h.p.*

$$\begin{aligned}
 q^*(\mathcal{A}) \geq q(\mathbf{C}) &\geq (1 + \mathcal{O}(n^{-(1-\zeta)/2})) \frac{2\alpha n}{\text{vol}(V)} \\
 &= (1 + \mathcal{O}((\log n)^{-1})) \frac{2\alpha}{d}, \quad \text{where } d = \sum_{k=\delta}^D kq_k.
 \end{aligned}$$

(Note that q_i is defined in (8).)

Recall that the modularity function of the ground-truth partition is *w.e.p.* asymptotic to $1 - \xi$. The above theorem implies that if $\delta \geq 4$ and the graph has a large level of noise, namely, $\xi \geq 3/\delta$ and $\xi > 1 - 2/d$, then *w.h.p.* the modularity function obtained from dissecting the spanning tree of G_0 is larger! The same conclusion can be derived when $\delta \leq 3$ by considering ξ sufficiently close to one.

In order to prove the above theorem, we first investigate the degree distribution of the background graph G_0 .

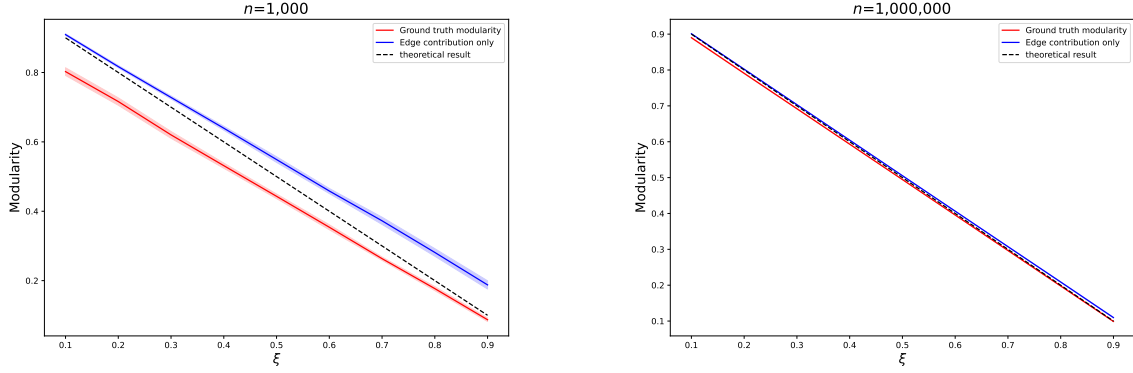


Figure 7: The modularity $q(\mathbf{C})$ of the ground-truth partition (red) and the corresponding edge contribution (blue) for 30 independently generated graphs; the shaded areas represent the standard deviation. The dashed line at $1 - \xi$ corresponds to a perfect prediction. Parameters used: $\gamma = 2.5$, $\delta = 5$, $\zeta = 1/2$, $\beta = 1.5$, $s = 50$, and $\tau = 3/4$. Two different graph sizes are investigated.

Lemma 6.3. *Let $\gamma \in (2, 3)$, $\delta \in \mathbb{N}$, $\zeta \in (0, \frac{1}{\gamma-1}]$, and $\xi \in (0, 1)$. For $k \in \mathbb{N} \cup \{0\}$, let \hat{Y}_k be the random variable counting the number of nodes in the background graph G_0 of \mathcal{A} that are of degree k . For $k \in \mathbb{N}$ and $\eta = \eta(n)$, let \hat{Y}_k^η be the random variable counting the number of nodes in the background graph G_0 of \mathcal{A} that are of degree at least k but at most $(1 + \eta)k$. Finally, for $k \in \mathbb{N} \cup \{0\}$, let*

$$u_k := \sum_{\substack{i \in \mathbb{N} \\ k-1 < \xi i < k+1 \\ \delta \leq i \leq \Delta}} (1 - |\xi i - k|) q_i,$$

where q_i is defined in (8).

The following properties hold w.e.p.:

(a) *If $\zeta \in (0, \frac{1}{\gamma})$, then for any $k \in \mathbb{N}$ such that $\lfloor \xi \delta \rfloor \leq k \leq \xi \lfloor n^\zeta \rfloor$ we have*

$$\hat{Y}_k = (1 + \mathcal{O}((\log n)^{-1})) nu_k = \Theta(nq_k) = \Theta(nk^{-\gamma}). \quad (17)$$

(b) *If $\zeta \in [\frac{1}{\gamma}, \frac{1}{\gamma-1})$, then for any $k \in \mathbb{N}$ such that $\lfloor \xi \delta \rfloor \leq k \leq n^{1/\gamma} (\log n)^{-4/\gamma} \ll n^\zeta$ the random variable \hat{Y}_k satisfies (17), and for any $k \in \mathbb{N}$ such that $n^{1/\gamma} (\log n)^{-4/\gamma} \leq k \leq \xi \lfloor n^\zeta \rfloor / (1 + \eta)$ we have*

$$\hat{Y}_k^\eta = \Theta(n\eta k q_k), \quad (18)$$

where

$$\eta = \eta(k) = n^{-1} (\log n)^4 k^{\gamma-1} = \mathcal{O}((\log n)^{-1}) = o(1).$$

(c) *If $\zeta = \frac{1}{\gamma-1}$, then for any $k \in \mathbb{N}$ such that $\lfloor \xi \delta \rfloor \leq k \leq n^{1/\gamma} (\log n)^{-4/\gamma} \ll n^\zeta$ the random variable \hat{Y}_k satisfies (17), and for any $k \in \mathbb{N}$ such that $n^{1/\gamma} (\log n)^{-4/\gamma} \leq k \leq n^\zeta (\log n)^{-5/(\gamma-1)} \ll n^\zeta$ the random variable \hat{Y}_k^η satisfies (18). The number of nodes in G_0 of degree at least $n^\zeta (\log n)^{-5/(\gamma-1)}$ is equal to $\Theta((\log n)^5)$.*

(d) The minimum and the maximum degrees of G_0 are respectively $\delta_0 \geq \lfloor \xi \delta \rfloor$ and $\Delta_0 \leq \lceil \xi n^\zeta \rceil \leq n^\zeta$.

(e) The volume of G_0 satisfies

$$\text{vol}_{G_0}(V) = \sum_{k=\delta}^D k \hat{Y}_k = (1 + \mathcal{O}((\log n)^{-1})) \xi dn, \quad \text{where } d := \sum_{k=\delta}^D k q_k.$$

Proof. As in the statement of Lemma 5.2, let Y_k and Y_k^η be the counterparts of \hat{Y}_k and \hat{Y}_k^η but defined for the whole graph \mathcal{A} instead of its subgraph, the background graph G_0 . By Lemma 5.2, *w.e.p.* the degree distribution of \mathcal{A} is well concentrated around its expectation.

Recall that a node of degree i in \mathcal{A} has degree $\lfloor \xi i \rfloor$ in G_0 , where $\lfloor \xi i \rfloor$ is a random variable equal to $\lfloor \xi i \rfloor$ or $\lceil \xi i \rceil$, and the probabilities are tuned such that the expectation is equal to ξi —see definition (2). We immediately get part (d) of the lemma (in fact, it holds deterministically, not only *w.e.p.*). Moreover, it implies that $\mathbb{E}[\text{vol}_{G_0}(V)] = \xi \text{vol}_{\mathcal{A}}(V)$ and so, by Corollary 5.3 and Chernoff's bound, we get part (e). More importantly, in order for a node of degree i in \mathcal{A} to have a chance to be of degree k in G_0 , we must have $k - 1 < \xi i < k + 1$. If $k - 1 < \xi i \leq k$, then its degree in G_0 is k with probability $\xi i - (k - 1) = 1 - |\xi i - k|$. On the other hand, if $k < \xi i < k + 1$, then the probability is equal to $(k + 1) - \xi i = 1 - |\xi i - k|$. Hence, we expect $1 - |\xi i - k|$ fraction of nodes of degree i in \mathcal{A} to be of degree k in G_0 and so

$$\mathbb{E}[\hat{Y}_k] = \sum_{\substack{i \in \mathbb{N} \\ k-1 < \xi i < k+1 \\ \delta \leq i \leq \Delta}} (1 - |\xi i - k|) \mathbb{E}[Y_i].$$

Note that for any value of k , $\mathbb{E}[\hat{Y}_k]$ is the linear combination of $\mathbb{E}[Y_i]$ with values of $i \in \mathbb{N}$ in the interval of length $2/\xi$. So there are at least $\lfloor 2/\xi \rfloor \geq 2$ terms in the above sum but no more than $2/\xi + 1$. The coefficient $1 - |\xi i - k|$ of at least one of them is bounded away from zero so we get that for the range considered in the statement of the lemma, we have $\mathbb{E}[\hat{Y}_k] = \Theta(\mathbb{E}[Y_k])$. As the result, all expectations tend to infinity fast enough, that is, are of order at least $(\log n)^4$. Hence, Lemma 5.2 combined with the Chernoff's bound, implies the concentration for \hat{Y}_k and \hat{Y}_k^η . The proof of the lemma is finished. \square

We will also need the following three results from [15], [20], and [40]. The first two provide sufficient conditions for a graph with a given degree sequence to be connected and, respectively, to have a giant component, that is, a component of linear order. (Both papers provide also necessary conditions but we do not need them so we only concentrate on the sufficient ones.)

Recall that a degree sequence is graphical if there exists at least one simple graph with such degree sequence. For any graphical degree sequence $\mathbf{w} := (w_1, w_2, \dots, w_n)$, let $\mathcal{G}(\mathbf{w})$ denote a random graph on the set of nodes $[n]$ selected uniformly at random from the family of simple graphs where node i has degree w_i for every $i \in [n]$. Our background graph is a multi-graph ($\mathcal{P}(\mathbf{w})$ instead of $\mathcal{G}(\mathbf{w})$) but the result below applies to both families of graphs, and the condition of \mathbf{w} being graphical is not needed for $\mathcal{P}(\mathbf{w})$. (In fact, as discussed in Subsection 4.2, in order to prove results that hold *w.h.p.* for $\mathcal{G}(\mathbf{w})$ one typically proves them for $\mathcal{P}(\mathbf{w})$ and then transfers them to $\mathcal{G}(\mathbf{w})$.)

The first lemma will be useful when the background graph G_0 has minimum degree at least 3.

Lemma 6.4 ([15]). Let $\mathbf{w} := (w_1, w_2, \dots, w_n)$ be any graphical sequence such that $w_1 \geq w_2 \geq \dots \geq w_n \geq 1$. Let

$$\Delta = w_1 \quad (\text{maximum degree}), \quad M = \sum_{i=1}^n w_i \quad (\text{volume}), \quad J = \sum_{i=1}^{\Delta} w_i,$$

and for each k let

$$n_k = \sum_{i=1}^n \mathbf{1}_{w_i=k} \quad (\text{the number of nodes of degree } k).$$

If $n_1 = o(\sqrt{M})$, $n_2 = o(M)$, and $J = o(M)$, then w.h.p. $\mathcal{G}(\mathbf{w})$ is connected.

The next lemma will be used when the background graph has minimum degree at most 2.

Lemma 6.5 ([20]). Let $\mathbf{w} := (w_1, w_2, \dots, w_n)$ be any graphical sequence such that $w_1 \leq w_2 \leq \dots \leq w_n$. Let

$$\begin{aligned} J &= \min \left(\left\{ j : j \in [n] \text{ and } \sum_{i=1}^j w_i(w_i - 2) > 0 \right\} \cup \{n\} \right), \\ R &= \sum_{i=J}^n d_i, \\ M &= \sum_{i \in [n]: d_i \neq 2} d_i. \end{aligned}$$

If $R \geq \varepsilon M$ for some $\varepsilon > 0$, then there exists $\alpha = \alpha(\varepsilon)$ such that w.h.p. $\mathcal{G}(\mathbf{w})$ has a component of order at least αn .

The final observation that we will need, regardless of the minimum degree of G_0 , is that the spanning tree of the background graph (and so also a spanning tree of \mathcal{A}) can be decomposed into small subtrees. This can be relatively easily done by analyzing some greedy algorithm. We direct the reader to [40] for more details.

Lemma 6.6 ([40]). Suppose that $G = (V, E)$ is a connected graph on n nodes. Then there exists a partition \mathbf{C} of V such that

$$q^*(G) \geq q(\mathbf{C}) \geq \frac{2n}{\text{vol}(V)} - 3\sqrt{\frac{\Delta}{\text{vol}(V)}} - \frac{\Delta}{\text{vol}(V)},$$

where Δ is the maximum degree of G and $\text{vol}(V)$ is the volume of G . In fact, the edge contribution is at least $2n/\text{vol}(V) - 2\sqrt{\Delta/\text{vol}(V)}$ and the degree tax is at most $\sqrt{\Delta/\text{vol}(V)} + \Delta/\text{vol}(V)$.

Now, we are ready to combine all of these observations and prove Theorem 6.2.

Proof of Theorem 6.2. Suppose first that $\xi\delta \geq 3$. We will use Lemma 6.4 to show that the background graph G_0 is w.h.p. connected. By Lemma 6.3(d) we know that the minimum degree of G_0 is at least $\lfloor \xi\delta \rfloor \geq 3$. Let $\mathbf{w} := (w_1, w_2, \dots, w_n)$ be the degree sequence of G_0 such that $w_1 \geq w_2 \geq \dots \geq w_n \geq 3$. The only property we need to check before we may apply Lemma 6.4 is to verify that $J = o(M)$.

Let $\omega = \omega(n)$ be any function tending to infinity as $n \rightarrow \infty$. By Lemma 6.3(a-c), *w.e.p.* the number of nodes of degree at least ω is $\sum_{k \geq \omega} \Theta(nk^{-\gamma}) = \Theta(n\omega^{-(\gamma-1)}) = n^{1-o(1)}$, much more than the maximum degree of G_0 ; trivially, $\Delta \leq n^\zeta$. Hence, *w.e.p.*

$$J = \sum_{i=1}^{\Delta} w_i \leq \sum_{i, w_i \geq \omega} w_i = \sum_{k \geq \omega} \Theta(k \cdot nk^{-\gamma}) = \Theta(n\omega^{-(\gamma-2)}) = o(n) = o(M),$$

since $M = \text{vol}_{G_0}(V) = \Theta(n)$ by Lemma 6.3(e). Lemma 6.4 implies that *w.h.p.* G_0 is connected and so, trivially, \mathcal{A} is connected *w.h.p.* too.

Since $\Delta \leq n^\zeta$, we get immediately from Lemma 6.6 that there exists a partition \mathbf{C} of V such that

$$q(\mathbf{C}) \geq (1 + \mathcal{O}(n^{-(1-\zeta)/2})) \frac{2n}{\text{vol}_{\mathcal{A}}(V)}.$$

The conclusion follows from Corollary 5.3 which implies that *w.e.p.* $\text{vol}_{\mathcal{A}}(V) = (1 + \mathcal{O}((\log n)^{-1})) dn$. This finishes part (a) of the theorem.

Suppose now that $\xi\delta < 3$. Lemma 6.3(a-c) tells us that *w.e.p.* there are linearly many nodes of degree 2 so, unfortunately, Lemma 6.4 cannot be applied. We will use Lemma 6.5 instead. This time we need to label nodes of G_0 such that $w_1 \leq w_2 \leq \dots \leq w_n$. Note that

$$\sum_{i=1}^j w_i(w_i - 2) = -n_1 + \sum_{i=n_1+n_2+1}^j w_i(w_i - 2) = -n_1 + \sum_{i=n_1+n_2+1}^j w_i^2 - 2 \sum_{i=n_1+n_2+1}^j w_i,$$

where n_1 and n_2 denote the number of nodes of degree one and two, respectively. By Lemma 6.3(a-c), *w.e.p.* $n_1 = (1 + \mathcal{O}((\log n)^{-1})) nu_k$. On the other hand, for any constant $K \geq 3$ we have

$$\begin{aligned} \sum_{i \in [n]: 3 \leq w_i \leq K} w_i^2 &= \sum_{k=3}^K \Theta(k^2 \cdot nk^{-\gamma}) = \Theta(nK^{3-\gamma}) \quad \text{and} \\ \sum_{i \in [n]: 3 \leq w_i \leq K} w_i &= \sum_{k=3}^K \Theta(k \cdot nk^{-\gamma}) = \Theta(n), \end{aligned}$$

where the constants hidden in the $\Theta(\cdot)$ notation depend on parameters of the model but do not depend on n nor K . Since the first sum grows with K and the second one does not, $\sum_{i=1}^j w_i(w_i - 2)$ becomes positive when the sum is taken over nodes of degree at most K for some sufficiently large constant K . (Let us stress it again that K is a universal constant, possibly large, but *not* a function of n .) We get that $J \leq (1 - \lambda)n$ for some $\lambda > 0$ and so both R and M are of order n . Since $R \geq \varepsilon M$ for some $\varepsilon > 0$, Lemma 6.5 can be applied and we get that *w.h.p.* G_0 has a component of order at least αn for some constant α which only depends on ε . (One may try to estimate R and M better, which would give some bound for ε but Lemma 6.5 does not provide an explicit function $\alpha = \alpha(\varepsilon)$ anyway so there is no point to do it.)

As before, it is enough to apply Lemma 6.6 but this time to the giant component of $G_0 = (V, E)$, graph $G'_0 = (V', E')$, instead of G_0 . By Corollary 5.3, *w.e.p.* $\text{vol}_{G_0}(V') = \Theta(n) = \Theta(\text{vol}_{\mathcal{A}}(V))$ (note that since G_0 is connected, $\text{vol}_{G_0}(V') \geq |V'| \geq \alpha n$). By Lemma 6.6, there exists a partition \mathbf{C}' of V' such that the corresponding edge contribution for the giant component in G_0 is at least

$(1 + \mathcal{O}(n^{-(1-\zeta)/2}))2\alpha n/\text{vol}_{G_0}(V')$, that is, the number of edges within one of the parts is at least $(1 + \mathcal{O}(n^{-(1-\zeta)/2}))\alpha n$. On the other hand, the degree tax is $\mathcal{O}(n^{-(1-\zeta)/2})$.

Let \mathbf{C} be a partition of V obtained by extending the partition \mathbf{C}' of $V' \subseteq V$ by adding a trivial partition of $V \setminus V'$ consisting of parts of size one (singletons). Since the partition \mathbf{C} captures exactly the same edges within some of its parts as partition \mathbf{C}' , the corresponding edge contribution for \mathcal{A} is $(1 + \mathcal{O}(n^{-(1-\zeta)/2}))2\alpha n/\text{vol}_{\mathcal{A}}(V)$. On the other hand, since *w.e.p.* $\text{vol}_{G_0}(V') = \Theta(\text{vol}_{\mathcal{A}}(V))$, the contribution to the degree tax for \mathcal{A} from parts in partition \mathbf{C}' is still $\mathcal{O}(n^{-(1-\zeta)/2})$. The contribution to the degree tax for \mathcal{A} from singletons is clearly

$$\begin{aligned} \sum_{i \in V \setminus V'} \left(\frac{w_i}{\text{vol}_{\mathcal{A}}(V)} \right)^2 &\leq \frac{\Delta}{\text{vol}_{\mathcal{A}}(V)^2} \sum_{i \in V \setminus V'} w_i \leq \frac{\Delta}{\text{vol}_{\mathcal{A}}(V)^2} \sum_{i \in V} w_i = \frac{\Delta}{\text{vol}_{\mathcal{A}}(V)} \\ &= \mathcal{O}(n^{-(1-\zeta)}) = \mathcal{O}(n^{-(1-\zeta)/2}). \end{aligned}$$

We get that *w.h.p.*

$$q(\mathbf{C}) \geq (1 + \mathcal{O}(n^{-(1-\zeta)/2})) \frac{2\alpha n}{\text{vol}(V)},$$

which finishes the proof of the theorem. \square

Low Level of Noise

This time we will investigate graphs with a low level of noise, that is, with ξ close to zero. Let us fix a value of $\delta \in \mathbb{N}$ such that $\delta \geq 100$. For any $a \in \mathbb{N}$ and $b \in \mathbb{N} \setminus \{1, 2\}$ such that $ab < \delta$, let

$$c(a, b) := \frac{b - 2\sqrt{b-1}}{2b} \frac{ab}{ab + b - 1} - \frac{b-1}{ab + b - 1} - 0.011. \quad (19)$$

Let

$$\xi_0(\delta) := \max_{a \in \mathbb{N}, b \in \mathbb{N} \setminus \{1, 2\}, ab < \delta} \min \left(1 - \frac{ab}{\delta}, \frac{c(a, b)}{4}, \frac{1}{20} \right). \quad (20)$$

It is clear that $\xi_0(\delta)$ is a non-decreasing function of δ . Moreover, $\xi_0(100) \approx 0.0217$ (the maximum is achieved for $a = 8$ and $b = 12$), and $\xi_0(\delta) = 1/20$ for $\delta \geq 340$.

Our first result says that for the **ABCD** graph \mathcal{A} with minimum degree $\delta \geq 100$ and $\xi \in (0, \xi_0(\delta))$, the maximum modularity $q^*(\mathcal{A})$ is *w.h.p.* asymptotically equal to the modularity function on the ground-truth.

Theorem 6.7. *Let $\delta \in \mathbb{N}$ such that $\delta \geq 100$ and $0 < \xi < \xi_0(\delta)$, where $\xi_0(\delta)$ is defined in (20). Let $\mathbf{C} = \{C_1, C_2, \dots, C_\ell\}$ be the ground-truth partition of the set of nodes of \mathcal{A} . Then, *w.h.p.**

$$q^*(\mathcal{A}) \sim q(\mathbf{C}) \sim 1 - \xi.$$

The lower bound of 100 for δ as well as the constants $\xi_0(\delta)$ are not tuned for the strongest result. Since the proof technique we use will not allow us to close the gap anyway, we aimed for a simple argument that works for large enough δ and relatively simple constants. Having said that, the above property is not true for $\delta = 1$; that is, if \mathcal{A} has minimum degree $\delta = 1$, then one may find a partition of the nodes of \mathcal{A} that yields larger modularity than the one associated with the ground-truth.

Theorem 6.8. Fix $\delta = 1$ and let $0 < \xi < 1$. Let $\mathbf{C} = \{C_1, C_2, \dots, C_\ell\}$ be the ground-truth partition of the set of nodes of \mathcal{A} . Then, w.e.p.

$$\begin{aligned} q^*(\mathcal{A}) &\geq (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) \left((1 - \xi) + \frac{\xi q_1}{d} \left(2 - \frac{q_1}{d} \right) \right) \\ &> (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1 - \xi) = q(\mathbf{C}), \end{aligned}$$

where q_k is defined in (8) and $d = \sum_{k=\delta}^D k q_k$.

The rest of this subsection is devoted to proving the above theorems. Recall that our ground-truth communities as well as the background graph are random multi-graphs $\mathcal{P}(\mathbf{w})$ on n' nodes with a given degree sequence $\mathbf{w} := (w_1, w_2, \dots, w_{n'})$ generated by the *configuration model*—see Subsection 2.3.5 for more details. In order to derive some useful expansion properties of $\mathcal{P}(\mathbf{w})$, we need to couple $\mathcal{P}(\mathbf{w})$ with $\mathcal{P}_{n'',b}$, the random b -regular multi-graph on n'' nodes, for some integer $b \geq 3$. Note that not only $\mathcal{P}(\mathbf{w})$ but also $\mathcal{P}_{n'',b}$ is generated by the configuration model. In particular, both of them are defined on the set of points that are eventually contracted to form nodes. In what follows, we establish a relationship between points used to generate both of them.

Coupling Between $\mathcal{P}(\mathbf{w})$ and $\mathcal{P}_{n'',b}$: Let us fix $a \in \mathbb{N}$ and $b \in \mathbb{N} \setminus \{1, 2\}$ such that $ab \leq (1 - \xi)\delta$, where $\xi \in (0, 1)$ and $\delta \geq 4$ are the parameters of the **ABCD** model \mathcal{A} . Note that the minimum degree of each community graph G_i is $\lfloor (1 - \xi)\delta \rfloor$ so ab is at most the minimum degree of each G_i .

Consider any random multi-graphs $\mathcal{P}(\mathbf{w})$ on n' nodes with a given degree sequence $\mathbf{w} := (w_1, w_2, \dots, w_{n'})$ and the minimum degree at most $(1 - \xi)\delta$. Let $W = \sum_{i=1}^{n'} w_i$ be the volume of $\mathcal{P}(\mathbf{w})$ which is equal to the number of points in the corresponding configuration model, as it was defined in Subsection 2.3.5. Recall that the model guarantees that W is even. Our goal is to couple $\mathcal{P}(\mathbf{w})$ and $\mathcal{P}_{n'',b}$ such that they use, if possible, the same set of points. If W is divisible by b , then we have the right number of points to create a random b -regular graph on $n'' = W/b$ nodes. Suppose then that W is congruent to $j \pmod{b}$ for some $j \in [b - 1]$. If $b - j$ is even, then we add $b - j$ *additional* points; otherwise, the number of additional points is equal to $2b - j$. Note that if b is even, then j has to be even as W is even. So if $b - j$ is odd, then b must be odd and so $2b - j$ is even. Hence, regardless whether $b - j$ is even or not, we add an even number of additional points which will be important for the argument below to hold. This time, a random b -regular graph will have a few more points than $\mathcal{P}(\mathbf{w})$ but, as before, W points will be shared by both models.

Recall that in the configuration model, points are partitioned into buckets that are eventually contracted to form nodes. In $\mathcal{P}(\mathbf{w})$, there are n' buckets, the i th bucket (corresponding to the i th node) consists of w_i points for a total of W points. In $\mathcal{P}_{n'',b}$, there are n'' buckets (that we will call *auxiliary buckets*), each consisting of b points for a total of $bn'' \geq W$ points (a potential discrepancy is taken care of additional points). Before we expose any edges, we need to assign each of the regular W points to buckets in $\mathcal{P}(\mathbf{w})$ and to auxiliary buckets in $\mathcal{P}_{n'',b}$; additional points will only be assigned to auxiliary buckets in $\mathcal{P}_{n'',b}$.

First, we arbitrarily place n' buckets associated with $\mathcal{P}(\mathbf{w})$ on W regular points. Then, for each $i \in [n']$, the i th bucket in $\mathcal{P}(\mathbf{w})$ consisting of w_i points is arbitrarily partitioned into sets of points of size b and, possibly, one set of at most $b - 1$ points. By construction, there are at least a sets of points of size b that become buckets in $\mathcal{P}_{n'',b}$ that we will call *internal auxiliary buckets*. The remaining points (including additional points if W is not divisible by b) are arbitrarily partitioned into b -element *external auxiliary buckets* in $\mathcal{P}_{n'',b}$. Note that this is possible as, by construction, W plus the number of additional points is divisible by b . As mentioned earlier, the original buckets will

be used to generate $\mathcal{P}(\mathbf{w})$ and the auxiliary buckets will correspond to the random b -regular graph generated by $\mathcal{P}_{n'',b}$.

It is time to start exposing edges, that is, start randomly matching points! Recall that one of the advantages of using the pairing model is that the pairs may be chosen sequentially, at each step choosing a point using any rule (possibly randomized) that depends only on the pairs so far chosen and pairing it with a point chosen uniformly at random over the remaining (unchosen) points. We start with exposing additional points (if we have them), one by one. Note that, during this initial phase, it might happen that two additional points are paired together. (This happens with probability $\mathcal{O}(1/n) = o(1)$ as there are $\mathcal{O}(1)$ additional points, so we could announce that the coupling fails and deal with such rare situations in the proof of the theorem differently. However, the coupling can be established even if this rare situation happens.) Suppose that r additional points are matched with not additional ones: for each $i \in [r]$, the additional point p_i is matched with point q_i . Recall that there is an even number of additional points. So, even if an even number of them are matched with other additional points, the number of pairs $p_i q_i$ is even (that is, r is even). Pairs $p_i q_i$ ($i \in [r]$) will be present in $\mathcal{P}_{n'',b}$ but $\mathcal{P}(\mathbf{w})$ will consist of pairs $q_{2i-1} q_{2i}$ instead ($i \in [r/2]$). Note that the sequence of points q_i ($i \in [r]$) is a sequence of r points selected uniformly at random from the set of non-repeating sequences of points from $\mathcal{P}(\mathbf{w})$ of that length. Hence, the two corresponding pairings are valid partial random pairings of $\mathcal{P}_{n'',b}$ and, respectively, $\mathcal{P}(\mathbf{w})$. We continue pairing the remaining points, keeping the obtained pairs in both models. This finishes the coupling.

Remark: Let us summarize the important observations. The coupling between $\mathcal{P}(\mathbf{w})$ and $\mathcal{P}_{n'',b}$, despite the fact that the number of points associated with these models might be different, has the following properties: a) $\mathcal{P}(\mathbf{w})$ and $\mathcal{P}_{n'',b}$ are perfect configuration models with their respected fixed degree distributions, and b) almost all pairs of points are coupled; the only discrepancy occurs with pairs associated with nodes in $\mathcal{P}_{n'',b}$ containing additional points and their neighbours ($\mathcal{O}(1)$ nodes in $\mathcal{P}_{n'',b}$ and so also $\mathcal{O}(1)$ nodes in $\mathcal{P}(\mathbf{w})$).

Now, we are ready to prove the first theorem, Theorem 6.7.

Proof of Theorem 6.7. The behaviour of some nodes might be challenging to predict. For example, nodes that belong to a community of constant size may (with positive probability) be partitioned into two parts with no edges between the parts. Such small communities might end up in different parts of the optimal partition yielding the maximum modularity $q^*(\mathcal{A})$ of \mathcal{A} . Fortunately, *w.h.p.* the number of problematic nodes will be small and so their contribution to the modularity function will be negligible.

We will use the terminology introduced in Lemma 5.6. In particular, we will call community C_i very large if $|C_i| \geq n^\zeta (\log n)^4$, and F is the union of communities that are not very large. Each community that is not very large will be called *problematic* and, as a result, all nodes of F will be called *problematic*. By Lemma 5.6, *w.e.p.* $\text{vol}(F) = \mathcal{O}(n (\log n)^{-2}) = o(n)$. On the other hand, by Corollary 5.3, *w.e.p.* $\text{vol}(V) = (1 + \mathcal{O}((\log n)^{-1})) dn$, where $d = \sum_{k=\delta}^D k q_k$. Hence, *w.e.p.* $\text{vol}(F) = o(\text{vol}(V))$. Since we aim for a statement that holds *w.h.p.*, we may assume that this property holds.

Let $\delta \geq 100$ be the minimum degree of \mathcal{A} . Let $a \in \mathbb{N}, b \in \mathbb{N} \setminus \{1, 2\}$ be the constants that yield $\xi_0(\delta) > 0$ in (20), and $c = c(a, b) > 0$ is defined as in (19). We couple the community graphs G_i of very large communities C_i that are generated according to the model $\mathcal{P}(\mathbf{w})$ with the model $\mathcal{P}_{n'',b}$. Note that, since $\xi < \xi_0(\delta) \leq 1 - ab/\delta$, we have $(1 - \xi)\delta \geq ab$ and so the coupling may be applied. Let λ_i be the largest absolute value of an eigenvalue other than $\lambda = b$ of the adjacency matrix

of the random b -regular graph $\mathcal{P}_{n'',b}$ associated with G_i . (See Section 4.2 for more on that.) If $\lambda_i > 2\sqrt{b-1} + 10^{-5}$, then we call the associated community C_i and all of its nodes *problematic*. By Lemma 4.2, *w.h.p.* each very large community is not problematic so the expected volume of nodes that become problematic is $o(n)$. By the first moment method, we conclude that *w.h.p.* the number of problematic nodes is $o(n)$. Very large communities G_i that are *not* problematic are coupled with b -regular graphs that are good expanders ($\lambda_i \leq 2\sqrt{b-1} + 10^{-5}$). However, as remarked above, there could be some constant number of nodes in G_i that are incident with edges in $\mathcal{P}(\mathbf{w})$ that are not present in $\mathcal{P}_{n'',b}$. We will also call these nodes *problematic*. Since the maximum degree is at most n^ζ and each very large community has at least $n^\zeta(\log n)^4$ nodes, the volume associated with new problematic nodes is $\mathcal{O}(n/(\log n)^4) = o(n)$.

We call a partition \mathbf{P} of the set of nodes of \mathcal{A} *nice* if the following two properties hold: a) all problematic communities (not very large or bad expanders) form a separate part in \mathbf{P} , and b) problematic nodes from non-problematic communities belong to a largest part of the partition of its own community that is induced by \mathbf{P} (if there are at least two communities that are largest, then all problematic nodes belong to one of them). Let \mathbf{P}^* be a partition that yields $q^*(\mathcal{A})$. One can move the problematic nodes around, if needed, to transform \mathbf{P}^* into a nice partition \mathbf{P} . Since the volume of the problematic nodes is $o(\text{vol}(V))$, both the edge contribution and the degree tax change by $o(1)$, and so we get that $q(\mathbf{P}) = q(\mathbf{P}^*) + o(1)$. Hence, it is enough to show that the maximum modularity over the family of nice partitions is $1 - \xi + o(1)$.

Let \mathbf{P} be a nice partition that yields the largest modularity over this family of partitions. We will show that each non-problematic community is contained in one part of \mathbf{P} . For a contradiction, suppose that partition \mathbf{P} partitions a non-problematic community of volume W into $j \geq 2$ parts U_i , $i \in [j]$. Let $W_i = \text{vol}(U_i)$. Without loss of generality, we may assume that $W_1 \geq W_2 \geq \dots \geq W_j$ and that all problematic nodes, if present, belong to U_1 . Consider any part U_i ($i \geq 2$). Since U_1 has the largest volume, $W_i \leq W/2$. Our goal is to estimate $E(U_i, V \setminus U_i)$, the number of edges going from U_i to its complement. To that end, we will use the coupling with b -regular graphs. Let U'_i be a subset of nodes of the coupled b -regular graph $\mathcal{P}_{n'',b}$ that are associated with internal auxiliary buckets, and let W'_i be its volume. (In order to make it easier for the reader to distinguish sets of nodes in the original graph $\mathcal{P}(\mathbf{w})$ from those in the coupled graph $\mathcal{P}_{n'',b}$, we will use U_i and, respectively, U'_i as well as W_i and W'_i .) Trivially, $W'_i \leq W_i$. Moreover, since at least a internal auxiliary buckets of b points (nodes in $\mathcal{P}_{n'',b}$) are contained in each node in $\mathcal{P}(\mathbf{w})$,

$$W'_i \geq \frac{ab}{ab + b - 1} W_i.$$

(The extreme case corresponds to a node in $\mathcal{P}(\mathbf{w})$ of degree $ab + b - 1$ which is associated with ab internal buckets.) Since $\mathcal{P}_{n'',b}$ is a good expander, it follows from (7) that the number of edges from U'_i to its complement satisfies the following inequality:

$$\begin{aligned} E(U'_i, V \setminus U'_i) &\geq (b - \lambda) \frac{|U'_i| |V \setminus U'_i|}{n''} \\ &= (b - \lambda + o(1)) \frac{(W'_i/b)((W - W'_i)/b)}{W/b} \\ &\geq \frac{b - \lambda + o(1)}{b} \frac{ab}{ab + b - 1} W_i \frac{W - W'_i}{W} \\ &\geq \frac{b - \lambda + o(1)}{2b} \frac{ab}{ab + b - 1} W_i. \end{aligned}$$

Note that not all of these edges go from U_i to its complement (in $\mathcal{P}(\mathbf{w})$) but clearly at most $W_i - W'_i$ of them can stay within U_i as there are $W_i - W'_i$ points in U_i that are assigned to external auxiliary buckets. Since $\lambda \leq 2\sqrt{b-1} + 10^{-5}$, we get that

$$\begin{aligned}
E(U_i, V \setminus U_i) &\geq E(U'_i, V \setminus U'_i) - (W_i - W'_i) \\
&\geq \frac{b - \lambda + o(1)}{2b} \frac{ab}{ab + b - 1} W_i - \frac{b - 1}{ab + b - 1} W_i \\
&\geq \left(\frac{b - 2\sqrt{b-1} - 10^{-5} + o(1)}{2b} \frac{ab}{ab + b - 1} - \frac{b - 1}{ab + b - 1} \right) W_i \\
&\geq (c - 10^{-5} + 0.011) W_i,
\end{aligned}$$

where $c = c(a, b)$ is defined in (19).

Now, we will modify the partition \mathbf{P} slightly, keeping the modified partition nice, and show that it improves its modularity which will give us the desired contradiction. We need to independently consider two cases. Suppose first that U_1 belongs to a part P_1 of \mathbf{P} of volume $\text{vol}(P_1) \leq \frac{c-1.1\xi}{2} \text{vol}(V)$. We move all nodes in parts U_i , $i \geq 2$, to P_1 . That operation puts together all edges of the non-problematic community we consider within one part but some edges of the background graph might get lost. To estimate the number of background edges that might get removed from some part of partition \mathbf{P} , note that the ratio between the background degree and the community degree of a node of degree $w \geq \delta \geq 100$ is at most

$$\frac{\lceil \xi w \rceil}{\lfloor (1 - \xi)w \rfloor} \leq \frac{\xi w + 1}{(1 - \xi)w - 1} \leq \frac{\xi \delta + 1}{(1 - \xi)\delta - 1} \leq \frac{100\xi + 1}{94} \leq 1.07\xi + 0.0107,$$

since $\xi \leq 1/20$. Hence, the edge contribution increases by at least

$$(c - 10^{-5} + 0.011) \sum_{i=2}^j \frac{W_i}{\text{vol}(V)} - (1.07\xi + 0.0107) \sum_{i=2}^j \frac{W_i}{\text{vol}(V)} \geq (c - 1.07\xi) \sum_{i=2}^j \frac{W_i}{\text{vol}(V)}.$$

On the other hand, the degree tax increases by at most

$$\begin{aligned}
\frac{(\text{vol}(P_1) + \sum_{i=2}^j W_i)^2}{\text{vol}(V)^2} - \frac{\text{vol}(P_1)^2}{\text{vol}(V)^2} &= \frac{2 \text{vol}(P_1) \sum_{i=2}^j W_i}{\text{vol}(V)^2} + \frac{\left(\sum_{i=2}^j W_i\right)^2}{\text{vol}(V)^2} \\
&= \frac{(2 + o(1)) \text{vol}(P_1) \sum_{i=2}^j W_i}{\text{vol}(V)^2} \\
&\leq (c - 1.1\xi + o(1)) \sum_{i=2}^j \frac{W_i}{\text{vol}(V)}.
\end{aligned}$$

Since

$$(c - 1.07\xi) - (c - 1.1\xi + o(1)) = 0.03\xi + o(1) > 0,$$

the modification of \mathbf{P} increases its modularity and we get a contradiction that \mathbf{P} maximizes the modularity over the family of nice partitions.

Suppose now that U_1 belongs to a part P_1 of \mathbf{P} of volume $\text{vol}(P_1) > \frac{c-1.1\xi}{2} \text{vol}(V)$. This time we move all nodes in parts U_i , $i \geq 1$, and form an independent part. Since we also disconnect U_1 from its part (P_1), the edge contribution increases by at least

$$\left(c - 1.07\xi\right) \sum_{i=2}^j \frac{W_i}{\text{vol}(V)} - \left(1.07\xi + 0.0107\right) \frac{W_1}{\text{vol}(V)}.$$

The second term might potentially dominate the change so the edge contribution might actually decrease. Trivially, it may decrease at most by the absolute value of the second term above. Fortunately, disconnecting U_1 from a large part decreases the degree tax substantially. Indeed, the degree tax decreases by at least

$$\begin{aligned} \frac{\text{vol}(P_1)^2}{\text{vol}(V)^2} - \frac{(\text{vol}(P_1) - W_1)^2}{\text{vol}(V)^2} - \frac{(\sum_{i=1}^j W_i)^2}{\text{vol}(V)^2} &= \frac{(2 + o(1)) \text{vol}(P_1)W_1}{\text{vol}(V)^2} \\ &\geq \left(c - 1.1\xi + o(1)\right) \frac{W_1}{\text{vol}(V)}. \end{aligned}$$

Note that $c \geq 4\xi_0(\delta) \geq 4\xi_0(100) > 0.08$. Since $\xi < \xi_0(\delta) \leq c/4$, we get that

$$\begin{aligned} \left(c - 1.1\xi + o(1)\right) - \left(1.07\xi + 0.0107\right) &= c - 2.17\xi - 0.0107 + o(1) \\ &\geq 0.4575c - 0.0107 + o(1) \\ &= 0.0366 - 0.0107 + o(1) > 0, \end{aligned}$$

and so the modification of \mathbf{P} increases its modularity and we get a contradiction too.

It follows that the nice partition \mathbf{P} that yields the largest modularity over this family of partitions has each non-problematic (very large and good expander) community contained in one part of \mathbf{P} . It remains to show that one cannot improve the modularity function by combining some non-problematic communities together. We will use $q_{G_0}(\mathbf{P})$, $e_{G_0}(A_i)$, and $\text{vol}_{G_0}(A_i)$ for counterparts of $q(\mathbf{P})$, $e(A_i)$, and $\text{vol}(A_i)$ that are applied for the background graph G_0 instead of the entire graph \mathcal{A} . We get that

$$\begin{aligned} q(\mathbf{P}) &= \sum_{A_i \in \mathbf{P}} \frac{e(A_i)}{|E|} - \sum_{A_i \in \mathbf{P}} \left(\frac{\text{vol}(A_i)}{\text{vol}(V)}\right)^2 \\ &= \left(1 - \xi + o(1)\right) + \sum_{A_i \in \mathbf{P}} \frac{e_{G_0}(A_i)}{|E|} - \sum_{A_i \in \mathbf{P}} \left(\frac{\text{vol}(A_i)}{\text{vol}(V)}\right)^2. \end{aligned}$$

By Lemma 5.6, *w.e.p.* the volume of each very large community C satisfies $\text{vol}_{G_0}(C) = \text{vol}(C) - \text{vol}_c(C) \sim \xi \text{vol}(C)$. By Lemma 6.3(e), *w.e.p.* $\text{vol}_{G_0}(V) \sim \xi \text{vol}(V)$. We get that *w.e.p.*

$$\begin{aligned} q(\mathbf{P}) &\leq \left(1 - \xi + o(1)\right) + \left(\xi + o(1)\right) \sum_{A_i \in \mathbf{P}} \frac{e_{G_0}(A_i)}{|E(G_0)|} - (1 + o(1)) \sum_{A_i \in \mathbf{P}} \left(\frac{\text{vol}_{G_0}(A_i)}{\text{vol}_{G_0}(V)}\right)^2 \\ &\leq \left(1 - \xi + o(1)\right) + (1 + o(1)) \left(\sum_{A_i \in \mathbf{P}} \frac{e_{G_0}(A_i)}{|E(G_0)|} - \sum_{A_i \in \mathbf{P}} \left(\frac{\text{vol}_{G_0}(A_i)}{\text{vol}_{G_0}(V)}\right)^2\right) \\ &\leq \left(1 - \xi + o(1)\right) + (1 + o(1)) q_{G_0}(\mathbf{P}). \end{aligned}$$

It remains to show that $q_{G_0}(\mathbf{P}) = o(1)$. We may contract each non-problematic community into a single node, since they must belong to one part. Similarly, we contract all problematic communities into a single node. Now, we may couple the entire background graph that is generated as the pairing model $\mathcal{P}(\mathbf{w})$ with $\mathcal{P}_{n'',d}$ with $a = 1$ and $b = d$ for an arbitrarily large integer d (note that after contraction the minimum degree in $\mathcal{P}(\mathbf{w})$ tends to infinity as $n \rightarrow \infty$ so $ab = d$ is certainly less than the minimum degree of $\mathcal{P}(\mathbf{w})$). By Lemma 4.4 and Lemma 4.2 we get that *w.h.p.* $q^*(\mathcal{P}_{n'',d}) = \mathcal{O}(1/\sqrt{d})$ and so *w.h.p.* $q^*(\mathcal{P}(\mathbf{w})) = \mathcal{O}(1/\sqrt{d})$. Since d can be made arbitrarily large, we conclude that $q_{G_0}(\mathbf{P}) = o(1)$ and the proof of the theorem is finished. \square

Now, let us move to the proof of the second theorem. Since the proof is rather straightforward and the reader is already warmed-up, we only sketch it.

Proof of Theorem 6.8. Suppose that $\delta = 1$. By Lemma 5.2, *w.e.p.* there are $(1 + \mathcal{O}((\log n)^{-1})) q_1 n$ nodes of degree 1 in \mathcal{A} , where q_k is defined in (8). It follows easily from Chernoff's bound that *w.e.p.* $(1 + \mathcal{O}((\log n)^{-1})) \xi q_1 n$ of them have degree 1 in the background graph (and so degree 0 in their own community graph)—see also Lemma 6.3. We will call such nodes *lucky*.

Consider the ground-truth partition $\mathbf{C} = \{C_1, C_2, \dots, C_\ell\}$ of the set of nodes of \mathcal{A} . It follows from Theorem 6.1 that *w.e.p.* $q(\mathbf{C}) = (1 + \mathcal{O}((\log n)^{-(\gamma-2)})) (1 - \xi)$. We will modify it to improve slightly the modularity function. All lucky nodes will be moved to the community of their neighbours. Note that two lucky nodes could be neighbours of each other, that is, they may form an isolated edge in \mathcal{A} . Such lucky nodes will be called *super-lucky* and we arbitrarily assign them to a community of one of them. In fact, edges formed by super-lucky nodes should form independent parts to increase the modularity function but the improvement would be negligible so there is no point to do it. By Lemma 6.3(e), the volume of the background graph is well concentrated around its mean and so the probability that a lucky node is super-lucky is equal to $(1 + \mathcal{O}((\log n)^{-1})) (\xi q_1 n)/(\xi d n)$, where $d = \sum_{k=\delta}^D k q_k$. Hence, the expected number of super-lucky nodes is equal to

$$(1 + \mathcal{O}((\log n)^{-1})) \xi q_1 n \frac{\xi q_1 n}{\xi d n} = (1 + \mathcal{O}((\log n)^{-1})) \frac{\xi q_1^2}{d} n,$$

and so the expected number of the associated isolated edges is half of it. On the other hand, the expected number of lucky nodes that are not super-lucky is $(1 + \mathcal{O}((\log n)^{-1})) \xi q_1 (1 - q_1/d) n$. The concentration follows easily from Chernoff's bound. Since all edges from the community graphs still remain in some part, *w.e.p.* this modification increases the edge contribution by

$$\begin{aligned} (1 + \mathcal{O}((\log n)^{-1})) \left(\frac{\xi q_1^2}{2d} + \xi q_1 \left(1 - \frac{q_1}{d}\right) \right) \frac{n}{|E|} &= (1 + \mathcal{O}((\log n)^{-1})) \left(\xi q_1 \left(1 - \frac{q_1}{2d}\right) \right) \frac{n}{dn/2} \\ &= (1 + \mathcal{O}((\log n)^{-1})) \frac{\xi q_1}{d} \left(2 - \frac{q_1}{d}\right), \end{aligned}$$

as $|E| = \text{vol}(V)/2 = (1 + \mathcal{O}((\log n)^{-1})) dn/2$ by Corollary 5.3.

Since each part may at most double its volume (deterministically), the degree tax after the modification is of the same order as before the modification, that is, it is $\mathcal{O}((\log n)^{-2})$. This finishes the proof of the theorem. \square

Simulation Corner

Theorem 6.2 shows that the maximum modularity is larger than the modularity of the ground-truth partition, provided that \mathcal{A} is very noisy. On the other hand, Theorem 6.7 shows the opposite: the

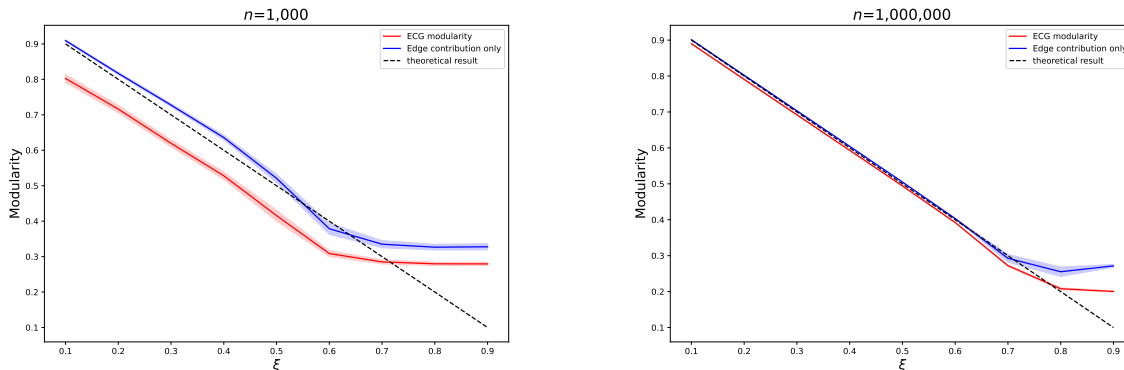


Figure 8: The modularity $q(\mathbf{C})$ obtained with ECG (red) and the corresponding edge contribution (blue) for 30 independently generated graphs; the shaded areas represent the standard deviation. The dashed line at $1 - \xi$ corresponds to a perfect prediction for the ground-truth. Parameters used: $\gamma = 2.5$, $\delta = 5$, $\zeta = 1/2$, $\beta = 1.5$, $s = 50$, and $\tau = 3/4$. Two different graph sizes are investigated.

maximum modularity is asymptotic to the modularity of the ground-truth partition, provided \mathcal{A} has low level of noise. For some technical reason, we assumed in that theorem that δ , the minimum degree of \mathcal{A} , is large enough but the same property should hold for much smaller values of δ .

In order to investigate this, for each value of $\xi = (0.1)i$, $i \in [9]$, we independently generated 30 graphs on $n = 1,000$ nodes and $n = 1,000,000$ nodes and with the same parameters as in the previous experiment: $\gamma = 2.5$, $\delta = 5$, $\zeta = 1/2 < 2/3 = 1/(\gamma - 1)$ (that is, $D = \sqrt{n}$), $\beta = 1.5$, $s = 50$, and $\tau = 3/4$ (that is, $S = n^{3/4}$). In order to approximate the maximum modularity, we used the ensemble clustering algorithm for unweighted graphs (ECG) which is based on the Louvain algorithm [7] and the concept of consensus clustering [39], and is shown to have good stability.

The experiments coincide with theoretical predictions, despite the fact that $\delta = 5$ in the experimental graphs is much smaller than the lower bound of 100 assumed in the corresponding theorem. For small values of ξ , the modularity found by ECG is very close to $1 - \xi$, an asymptotic prediction for the modularity of the ground-truth. On the other hand, if ξ is large, then the partition found by the algorithm is of better quality than the ground-truth partition—see Figure 8.

In Table 2, we compare of the modularity of the ground-truth partition and a partition found using ECG. Additionally, we provide AMI (the adjusted mutual information) and ARI (the adjusted Rand index) measures of similarities between the two partitions [16]. If ξ is close to 0, then ground-truth and ECG partitions are similar. Based on our theoretical results, we also expect that they are close to the maximum graph modularity, q^* . The reason is that there is a low level of noise coming from the background graph and so ECG has no problem finding a good partition that is close to the ground-truth one. On the other hand, if ξ is close to 1, then the two partitions are dissimilar. Also, note that then the modularity of the ECG partition is much higher than the one of the ground-truth, as expected based on the theoretical results. For intermediate values of ξ we observe two effects. The first observation is that AMI and ARI of the ECG partition and ground-truth partition drops sharply when ξ becomes greater than 0.5. The reason for this is that for such values of ξ the volume of the background graph, which is independent from the community graphs, becomes dominant and ECG tries to recover its structure. The second observation, which is a consequence of the first one,

ξ	size (n)	$q(\text{ground truth})$	$q(\text{ECG})$	AMI	ARI
0.1	10^3	0.802594	0.802594	1.000000	1.000000
	10^6	0.890027	0.890038	0.999367	0.999968
0.2	10^3	0.716289	0.716190	0.999721	0.9996570
	10^6	0.790705	0.790722	0.998952	0.999943
0.3	10^3	0.620174	0.619678	0.997716	0.997692
	10^6	0.692048	0.692070	0.998584	0.999903
0.4	10^3	0.531436	0.528090	0.974488	0.973746
	10^6	0.593213	0.593239	0.997954	0.999763
0.5	10^3	0.443102	0.415710	0.688916	0.632881
	10^6	0.494549	0.494295	0.986875	0.987448
0.6	10^3	0.353922	0.308721	0.199603	0.109946
	10^6	0.395613	0.392237	0.889011	0.856030
0.7	10^3	0.263305	0.285068	0.071929	0.022793
	10^6	0.296665	0.272223	0.366945	0.164060
0.8	10^3	0.176988	0.279705	0.032120	0.008722
	10^6	0.197707	0.207766	0.027720	0.001900
0.9	10^3	0.087200	0.279385	0.013933	0.003501
	10^6	0.098857	0.200353	0.001541	0.000025

Table 2: Comparison of the modularity of the ground-truth communities and the modularity obtained using ECG along with AMI and ARI measures between the two partitions. Data is presented for graphs with 10^3 and 10^6 nodes. The results are averages over 30 randomly sampled graphs. Parameters used: $\gamma = 2.5$, $\delta = 5$, $\zeta = 1/2$, $\beta = 1.5$, $s = 50$, and $\tau = 3/4$.

is that for intermediate values of ξ the modularity of the ground-truth is noticeably larger than for the ECG partition (for example, for $\xi = 0.6$ with $n = 10^3$ and for $\xi = 0.7$ with $n = 10^6$). The reason is that ECG already starts to get a lot of signal from the background graph while it still would be more efficient to stick to the partition closer to the ground-truth. However, because of the large level of noise already present in the graph coming from the background graph, the ECG algorithm is unable to recover it.

References

- [1] William Aiello, Anthony Bonato, Colin Cooper, Jeanette Janssen, and Paweł Prałat. A spatial web graph model with local influence regions. *Internet Mathematics*, 5(1-2):175–196, 2008.
- [2] Noga Alon and Fan RK Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72(1-3):15–19, 1988.
- [3] Noga Alon and Vitali D Milman. λ_1 , isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.
- [4] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.
- [5] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

- [6] Edward A Bender and E Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307, 1978.
- [7] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
- [8] 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.
- [9] Jordan Chellig, Nikolaos Fountoulakis, and Fiona Skerman. The modularity of random graphs on the hyperbolic plane. *Journal of Complex Networks*, 10(1):cnab051, 2022.
- [10] Fan Chung Graham and Linyuan Lu. *Complex graphs and networks*. Number 107. American Mathematical Soc., 2006.
- [11] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.
- [12] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.
- [13] Santo Fortunato and Marc Barthelemy. Resolution limit in community detection. *Proceedings of the national academy of sciences*, 104(1):36–41, 2007.
- [14] Joel Friedman. *A proof of Alon’s second eigenvalue conjecture and related problems*. American Mathematical Soc., 2008.
- [15] Pu Gao and Yuval Ophapkin. Subgraph probability of random graphs with specified degrees and applications to chromatic number and connectivity. *arXiv preprint arXiv:2007.02216*, 2020.
- [16] Alexander J. Gates and Yong-Yeol Ahn. The impact of random models on clustering similarity. *J. Mach. Learn. Res.*, 18(1):3049–3076, 2017.
- [17] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.
- [18] Svante Janson. Random graphs with given vertex degrees and switchings. *Random Structures & Algorithms*, 57(1):3–31, 2020.
- [19] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random graphs*, volume 45. John Wiley & Sons, 2011.
- [20] Felix Joos, Guillem Perarnau, Dieter Rautenbach, and Bruce Reed. How to determine if a random graph with a fixed degree sequence has a giant component. *Probability Theory and Related Fields*, 170(1):263–310, 2018.
- [21] Bogumił Kamiński, Tomasz Olczak, Bartosz Pankratz, Paweł Prałat, and François Théberge. Properties and performance of the abcde random graph model with community structure. *Big Data Research*, (30):100348, 2022.
- [22] Bogumił Kamiński, Valérie Poulin, Paweł Prałat, Przemysław Szufel, and François Théberge. Clustering via hypergraph modularity. *PloS one*, 14(11):e0224307, 2019.

- [23] Bogumił Kamiński, Paweł Prałat, and François Théberge. Community detection algorithm using hypergraph modularity. In *International Conference on Complex Networks and Their Applications*, pages 152–163. Springer, 2020.
- [24] Bogumił Kamiński, Paweł Prałat, and François Théberge. Artificial benchmark for community detection (abcd)—fast random graph model with community structure. *Network Science*, pages 1–26, 2021.
- [25] Bogumił Kamiński, Paweł Prałat, and François Théberge. Mining complex networks. 2021.
- [26] Bogumił Kamiński, Paweł Prałat, and François Théberge. Hypergraph artificial benchmark for community detection (h-abcd). *arXiv preprint*, 2022.
- [27] Bogumił Kamiński, Paweł Prałat, and François Théberge. Outliers in the abcd random graph model with community structure (abcd+o). In *11th International Conference on Complex Networks and Their Applications*. Springer, 2022 (in press).
- [28] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010.
- [29] Renaud Lambiotte and M Schaub. Modularity and dynamics on complex networks. 2021.
- [30] Andrea Lancichinetti and Santo Fortunato. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E*, 80(1):016118, 2009.
- [31] Andrea Lancichinetti and Santo Fortunato. Limits of modularity maximization in community detection. *Physical review E*, 84(6):066122, 2011.
- [32] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical review E*, 78(4):046110, 2008.
- [33] Lyuben Lichev and Dieter Mitsche. On the modularity of 3-regular random graphs and random graphs with given degree sequences. *Random Structures & Algorithms*, 2022.
- [34] Colin McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, pages 195–248. Springer, 1998.
- [35] Colin McDiarmid and Fiona Skerman. Modularity of regular and treelike graphs. *Journal of Complex Networks*, 6(4):596–619, 2018.
- [36] Colin McDiarmid and Fiona Skerman. Modularity of Erdős-Rényi random graphs. *Random Structures & Algorithms*, 57(1):211–243, 2020.
- [37] Mark EJ Newman. Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133, 2004.
- [38] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
- [39] Valérie Poulin and François Théberge. Ensemble clustering for graphs. In *International Conference on Complex Networks and their Applications*, pages 231–243. Springer, 2018.

- [40] Liudmila Ostroumova Prokhorenkova, Paweł Prałat, and Andrei Raigorodskii. Modularity of complex networks models. *Internet Mathematics*, 2017.
- [41] Nicholas C Wormald. Generating random regular graphs. *Journal of algorithms*, 5(2):247–280, 1984.
- [42] Nicholas C Wormald et al. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.