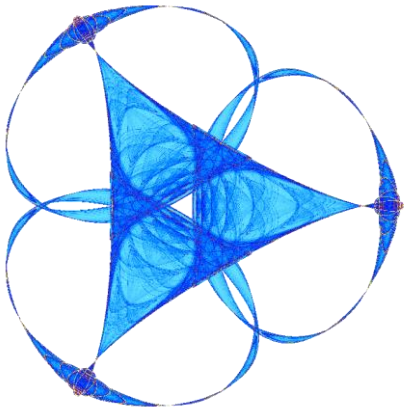
DISTINGUISHING VERTICES OF INHOMOGENEOUS RANDOM GRAPHS

By

Paolo Codenotti

IMA Preprint Series #2419

(July 2013)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS
UNIVERSITY OF MINNESOTA
400 Lind Hall
207 Church Street S.E.
Minneapolis, Minnesota 55455-0436
Phone: 612-624-6066 Fax: 612-626-7370
URL: http://www.ima.umn.edu

Distinguishing Vertices of Inhomogeneous Random Graphs

Paolo Codenotti*
Institute for Mathematics and its Applications
University of Minnesota
207 Church Street SE, 306 Lind Hall, Minneapolis, MN 55455
paolo@ima.umn.edu

July 25, 2013

Abstract

We explore under what conditions simple combinatorial attributes and algorithms such as the distance sequence and degree-based partitioning and refinement can be used to distinguish vertices of inhomogeneous random graphs. In the classical setting of Erdős-Renyi graphs and random regular graphs it has been proven that vertices can be distinguished in a constant number of rounds of degree-based refinement or the distance sequence at a logarithmic distance. This yields a high-probability canonical labeling algorithm, and hence an efficient high-probability isomorphism test.

In this paper we analyze the same attributes in the context of random graphs that come from distributions that more closely model real-world networks. We first prove a technical result about the effects of one refinement step in the general setting were edges are chosen independently at random. This allows us to prove that an algorithm based on distinguishing vertices yields a canonical labeling for graphs with scale-free degree distribution where edges are added independently, and for Stochastic Kronecker Product Graphs with certain settings of the parameters. Along the way we prove results on the degree distribution, connectedness and diameter of Stochastic Kronecker Graphs with generating matrices of arbitrary size.

1 Introduction

Several random graph models have been proposed over the last two decades to model real-world networks. For a more complete bibliograpy we refer the reader to the following books [19, 33], and review articles [11, 36].

These models have applications to a variety of fields. In computer science they are used for modeling the Internet, the World Wide Web, and social networks [2, 3, 15, 25, 24, 27]. In biology they are used as models for protein-protein, gene-gene, and gene-protein interaction networks [38]. Random network models have also been extensively used in statistics and social sciences (see e. g. [17]).

The new random graph models try to explain properties such as small-world, clustering coefficients, and shrinking diameters, which are not captured by the classical models of Erdős-Renyi

*This research was supported by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation.

graphs [20] and random regular graphs [10, 9]. One important difference between the new models and the classical models is that the vertices are not homogeneous.

We explore under what conditions simple combinatorial attributes which can be computed efficiently such as the distance sequence and a degree-based partitioning and refinement procedure can be used to distinguish vertices of inhomogeneous random graphs. In the classical setting of Erdős-Renyi graphs and random regular graphs it has been proven that vertices can be distinguished in a constant number of rounds of degree-based refinement and by the distance sequence at a logarithmic distance [4, 5, 9, 10]. This yields a high-probability canonical labeling algorithm, and hence an efficient high-probability isomorphism test.

In this paper we analyze the same attributes in the context of random graphs that come from distributions that more closely model real-world networks.

Our main results are for settings where edges are added independently at random. In general, let P be an $n \times n$ matrix with entries $0 \leq P[i,j] \leq 1$. We consider a probability distribution $\mathcal{G}(n,P)$ on the set of all graphs on n vertices. A graph $G \in \mathcal{G}(n,P)$ is selected by independently adding each edge (i,j) with probability $P[i,j]$. In other words

$$\mathbb{P}(G) = \prod_{(i,j)\in E(G)} P[i,j] \prod_{(i,j)\notin E(G)} (1 - P[i,j]).$$

The main application of our technique in this paper is the following (see subsection 2.2 for the precise definition of partition-refinement and the distance sequence):

Theorem 1.1 *Let $P[i,j] = \Theta(1/\sqrt{ij})$, and pick the graph $G \sim \mathcal{G}(n,P)$. Note that with high probability G will have scale-free degree distribution. With probability $1 - o(1/n)$, we have:*

- $\deg(1) > \deg(2) > \cdots > \deg(k)$, *for all $k = o(n^{1/5}(\ln n)^{-2/5})$. In other words the degree of every vertex $i = o(n^{1/5}(\ln n)^{-2/5})$ is unique in G.*
- *Every vertex $i = O\left(\frac{n}{\ln^8 n}\right)$ is uniquely determined by the degrees of its neighbors.*
- *We call two vertices* equivalent *if they have the same set of neighbors. Every vertex that is connected to the giant connected component will be distinguished from every other vertex that is not equivalent to it by partition-refinement initialized by the distance sequence.*

This theorem implies that our algorithm is a canonical labeling algorithm for almost every graph in the distribution $\mathcal{G}(n,P)$.

In subsection 3.1 we prove the more general technical results that we need for the theorem above. These results can be used in various other applications, for example they also underly our results for Stochastic Kronecker Product Graphs [27]. We defer the definition of this model to section 4. Our main result in this setting is the following: in a stochastic Kronecker graph, every vertex that has expected degree at least n^ϵ, for any constant ϵ will be distinguished after one round of naive partition-refinement. In fact, this gives us a canonical labeling algorithm for the setting where the Stochastic Kronecker Graphs are connected with high probability. The precise statement is more involved, and requires to first prove results about the degree distribution and connectedness of Stochastic Kronecker Graphs with generating matrices of arbitrary size. These had previously only been analyzed for 2×2 generating matrices [30].

We study the above random graph models for two reasons: they are mathematically clean, and several other models build upon them (see e.g. [25, 22]). Simulations show that these attributes

can be used to distinguish vertices in very few rounds even for preferential attachment graphs [3], where the edges are not independent.

Our algorithms can be easily implemented in a distributed setting. Whenever the algorithms require few rounds, the corresponding distributed implementations become very efficient, with few communication steps, and low running time.

Finally we note that procedures to distinguish vertices can be useful even in cases where they do not distinguish all the vertices in the graph (and hence do not obtain a canonical labeling). For example, they can be used for leader selection in distributed settings, for anti-aliasing, or to reduce the search space for node correspondence in maximum-likelihood parameter estimation. Our main result shows that in graphs that exhibit a core-periphery structure, vertices in the core will be easier to distinguish. This is good news, since these are the more "interesting" vertices.

The rest of the paper is organized as follows. In section 2 we discuss related work and precisely describe the algorithms. In section 3 we present our main technical tools and apply them to scale-free random graphs with independent edges. Then in section 4 we discuss our results for stochastic Kronecker graphs. In section 5 we present our experimental results for preferential attachment graphs. We conclude and discuss further work in section 6.

2 Background and Preliminaries

2.1 Graph Isomorphism

The Graph Isomorphism problem asks whether two input graphs are isomorphic or not. Graph Isomorphism has a very interesting computational complexity status: it is in co-AM, and hence is not NP-complete unless the polynomial-time hierarchy collapses to the second level [8]. However, we do not know any polynomial-time algorithms for this problem. Graph Isomorphism is the only important natural problem with this complexity status.

Graph Isomorphism has applications to several areas, including chemistry and SAT-solvers, and there is considerable interest in practical algorithms to solve the problem. There are software packages [31, 21, 18] that perform well on most instances that arise in practice, but take exponential time on some instances [32].

The fastest worst-case algorithm for Graph Isomorphism runs in time $\exp(O(\sqrt{n \log n}))$ [7]. This algorithm was obtained by combining a combinatorial trick due to Zemlyachenko [39] and Luks' polynomial-time isomorphism test for bounded degree graphs [29]. Luks' bounded degree result is the crowning achievement of the group theoretic methods for graph isomorphism, first introduced by Babai in 1979 [6].

2.2 Algorithms

We now describe two simple algorithms to distinguish vertices of graphs: degree-based partition and refinement, and distance-sequence computation.

Degree sequence and refinement. Partitioning and refinement is an effective heuristic for graph isomorphism (cf. Read and Corneil [35]). This heuristic is at the heart of most practical graph isomorphism algorithms (see e.g. [31, 21]). For almost all graphs this technique provably works in polynomial time, giving some theoretical explanation for its success [4, 5].

We consider colored graphs. The refinement procedure will work for any coloring, but one good example to keep in mind for this discussion is to initially color vertices by their degree. The goal of the refinement technique is to get closer to a coloring of the vertices such that each vertex gets a unique color. Given a graph $X = (V, E)$ and a coloring $c : V \to \mathcal{C}$, where $\mathcal{C}$ is a set of colors, one "refinement step" computes a new coloring c' that is a refinement of c. The new color of a node is defined by the multiset of colors of the neighbors, more precisely,

$$c'(v) := (c(v), \{c(u) \mid (u, v) \in E\}) \qquad \forall v \in V.$$

We call the coloring c an *invariant* if it is preserved by all automorphisms. Note that any isomorphism that preserves c must preserve c'. Hence if c was invariant, then so is c'.

The set of potential colors of c' may be large, but at most n colors actually appear, so we can number those colors that do appear, and assume $c' : V \to \{1, \ldots, n\}$. This trick allows us to recursively apply the refinement step until we reach a 'stable coloring,' that is, any further refinement does not change it. The number of refinement steps that can occur is at most n, since if $c \neq c'$, at least one color class must be partitioned. Therefore the overall running time is bounded by $O(n(n + m))$.

If the number of refinement steps, r, is small, then the algorithm can be implemented as an r-round distributed distributed algorithm, with run time Δ^r, where Δ is the maximum degree.

This is a general procedure that works for any initial vertex-coloring of a graph. Often it is used as a subroutine in isomorphism algorithms. The Naive Refinement algorithm starts by coloring each vertex according to its degree, and recursively refines the coloring until a stable coloring is reached.

Babai-Erdős-Selkow [4] who showed that apart from a $n^{-1/7}$ fraction of graphs, the Naive Refinement procedure finds a canonical labeling in one refinement step. The heart of their proof is an estimate on the top degrees of vertices in a random graph: they show that for every constant c, for almost all graphs, the highest $c \log n$ degrees are distinct.

Using a more complicated argument, Naive Refinement was shown to canonically label all but an exponentially small fraction of graphs in just three refinement steps [5].

Distance sequence. For random graphs the naive refinement procedure based on degrees described above does not even start. To distinguish vertices of random regular graphs, it is useful to look at another combinatorial attribute of vertices, the distance sequence [9, 10].

For a vertex v, let $d_i(v)$ be the number of vertices at distance i from v. The *distance sequence* of vertex v is the sequence of $d_i(v)$. Bollobas proved that for (Erdős-Renyi) Random Graphs, and random regular graphs, with probability 1, the distance sequence up to length $\ell_0 \leq O(\log n)$ uniquely determines every vertex in the graphs [9, 10]. In particular this gives a canonical labeling procedure for almost every regular graph, and hence an isomorphism test for these graphs and any other graph.

Our algorithm Our goal is to see how far these combinatorial techniques go for other distributions on the set of all graphs.

The most general version of our algorithm is to first color every vertex by its distance sequence, then run partition-refinement. In fact, this will be overkill for many vertices. We show that high degree vertices will be distinguished by few rounds of naive partition-refinement (precise bounds for this statement will depend on the model, see Theorem 1.1 for an example). Low degree vertices require computing the distance sequence and possibly more rounds of refinement.

3 Independent edges

In this section we prove our main technical tools for distinguishing vertices. We consider the following random graph model. Given an integer n, and an $n \times n$ matrix P of probabilities, we construct a random graph $G(P, n)$ with n vertices, and edges added at random with $\mathbb{P}((i,j) \in E) = P[i,j]$.

3.1 Coin tosses with non-uniform probabilities

Here we consider the setting where we toss t independent coins, each with a (possibly different) probability p_i to come up heads. First we prove a specific concentration bound.

Lemma 3.1 *We toss t coins, each has probability p_i to come up heads. Let $\lambda = p_1 + \cdots + p_t = \mathbb{E}[\#$ of heads $]$. Let s denote the number of heads. $\forall \epsilon > 0, \forall y > 0$, let $z = \max\{\lambda, \frac{2y \ln t}{\epsilon^2}\}$, then $\mathbb{P}[|s - \lambda| > \epsilon z] \leq t^{-y}$.*

Proof. The random variable s is the sum of t independent Bernoulli random variables. By Bennett's inequality, we have:

$$\begin{aligned}
\mathbb{P}[|s - \lambda| > \epsilon z] &\leq \exp\left[-\frac{\epsilon z}{2} \ln\left(1 + \frac{\epsilon z}{\lambda}\right)\right] \\
&\leq \exp\left[-\frac{z}{2}\, \epsilon \ln(1+\epsilon)\right] \\
&\leq \exp\left[-\frac{y \ln t}{\epsilon^2}\, \epsilon \ln(1+\epsilon)\right] \\
&\leq \exp\left[-y \ln t\right] = t^{-y}
\end{aligned}$$

The first inequality is Bennett's inequality. The second follows since $z \geq \lambda$, hence $z/\lambda \geq 1$. The third inequality since $z \geq \frac{2y \ln t}{\epsilon^2}$, and the fourth by the fact that $\ln(1+\epsilon) \leq \epsilon$. □

REMARK. In particular this applies for a vertex u, where $\lambda = \mathbb{E}[\deg(u)]$, and $t = n$.

Next we bound the probability of attaining any particular value.

Lemma 3.2 *We toss t coins, each has probability p_i to come up heads. Let $\lambda = p_1 + \cdots + p_t = \mathbb{E}[\#$ of heads $]$, and $\gamma = \max_i p_i$. For every x,*

$$\mathbb{P}(\textit{exactly } x \textit{ heads}) \leq \frac{e^{\lambda\gamma}}{\sqrt{2\pi}} \frac{1}{\sqrt{\lambda}}$$

Proof. Let $f(x) = \mathbb{P}(\text{exactly } x \text{ heads})$. The probability distribution $f(x)$ will not be too far from a Poisson distribution:

$$f(x) \leq \frac{\lambda^x e^{-\lambda + x\gamma}}{x!}, \tag{1}$$

While the proof of Equation (1) is elementary, and the setting is a common setting in statistics, this was only proved by Wang in 1993 [37, Equation (31)].

Next we use the following form of Stirling's approximation(c.f. [1, 6.1.38]):

$$\sqrt{2\pi}\, n^{n+1/2} e^{-n} \leq n! \leq e^{1/12}\sqrt{2\pi}\, n^{n+1/2} e^{-n} \tag{2}$$

Combining Equations (1) and (2), we obtain:

$$f(x) \leq \frac{\lambda^x e^{-\lambda+x\gamma}}{\sqrt{2\pi} x^{x+1/2} e^{-x}} \leq \frac{1}{\sqrt{2\pi}} \frac{\cancel{\lambda^x}}{\cancel{\lambda^x}} \frac{\cancel{e^{-\lambda}}+\lambda\gamma}{\cancel{e^{-\lambda}}} \frac{1}{\sqrt{\lambda}} = \frac{e^{\lambda\gamma}}{\sqrt{2\pi\,\lambda}},$$

where the second inequality follows since the function is maximized for $x = \lambda$. □

Consider the random graph model $G(P, n)$, and fix a vertex v. Now suppose that there are k color classes in the graph. Let $\lambda_c = \lambda_c(v)$ be the expected number of neighbors of v that have color c, and let $\gamma_c = \gamma_c(v)$ be the maximum probability of any edge from v to a vertex of color c: $\gamma_c = \max_{u|c(u)=c} P[v, u]$.

Lemma 3.3 *Fix a vertex v, and a set of colors $S \subseteq \mathcal{C}$. Using the notation from above, let $\lambda_{\min} = \min_{c\in S} \lambda_c$, and $K = \max_{c\in S} \gamma_c\lambda_c$. Then the probability that v is uniquely determined by the colors of its neighbors (i. e. that v has a unique color after one step of refinement) is $1 - n^{-y}$, where $y = \frac{|S|(\ln(\sqrt{2\pi\lambda_{\min}})-K)}{\ln n}$.*

Proof. For a vertex v, let $\mathcal{C}(v)$ be the multi-set of colors of the neighbors of v. Then for two vertices u, v, we have

$$\begin{aligned}\mathbb{P}(\mathcal{C}(u) = \mathcal{C}(v)) &\leq \max_S \mathbb{P}(\mathcal{C}(v) = S) \\ &\leq \max_{x_1,\ldots,x_k} \prod_{c\leq k} \mathbb{P}(v \text{ has exactly } x_c \text{ neighbors of color } c).\end{aligned}$$

By Lemma 3.2, we have that:

$$\begin{aligned}\mathbb{P}(\mathcal{C}(u) = \mathcal{C}(v)) &\leq \prod_{c\in \mathrm{S}} \frac{e^{\gamma_c\lambda_c}}{\sqrt{2\pi\lambda_c}} \leq \prod_{c\in \mathrm{S}} \frac{e^K}{\sqrt{2\pi\lambda_{\min}}} \\ &= \exp\left(|S|(K - \ln(2\sqrt{\pi}))\right) = \exp(-y\ln n) = n^{-y}.\end{aligned}$$

□

Corollary 3.4 *If there is a vertex v and set of colors S with $|S| \geq 8\ln n$, and such that for all $c \in S$, $\lambda_c \geq 2$ and $\gamma_c\lambda_c < 1$, then the probability of v being uniquely determined by the colors of its neighbors is $1 - n^{-2}$.*

3.2 Application to random scale-free graphs

In this subsection we prove Theorem 1.1. If we choose $P[i, j] = \theta(1/\sqrt{ij})$, then the expected number of vertices of degree d will be $\theta(n/d^3)$, and with high probability, the number of vertices of degree d will be concentrated around its expected value (cf. [12])

Lemma 3.5 *For any constant y, with probability $1 - o(n^{-y})$, every vertex $i = o(n^{1/5}(\ln n)^{-2/5})$ will have a unique degree.*

Proof. First we show that vertices i for $i = o(n^{1/3}/\ln n)$ will all have distinct degrees with high probability. Let

$$\lambda_i = \mathbb{E}[\deg(i)] = \theta\left(\sqrt{\frac{n}{i}}\right), \quad \text{and} \quad \delta_i = (\lambda_i - \lambda_{i+1})/2 = \Theta\left(\frac{\sqrt{n}}{\sqrt{i}} - \frac{\sqrt{n}}{\sqrt{i+1}}\right) = \Theta(\sqrt{n}\, i^{-3/2}).$$

We have

$$\mathbb{P}[\deg(i) > deg(i+1)] \geq (1 - \mathbb{P}[deg(i) < \lambda_i - \delta_i])(1 - \mathbb{P}[deg(i+1) > \lambda_{i+1} + \delta_i])$$

Since $\delta_i > \delta_{i+1}$, it is enough to show that

$$\mathbb{P}[|\deg(i) - \lambda_i| > \delta_{i+1}] < n^{-y'}, \qquad y' = y + 1/3$$

.

We apply Lemma 3.1, with $t = n$, , and set $\epsilon = \frac{C\, y \ln n\, i^{3/2}}{\sqrt{n}}$ for some (small) constant C depending on the Θs above. Note that since $i = o(n^{1/5}(\ln n)^{-2/5})$, we have that $\epsilon < 1/i$. By Lemma 3.1, with probability $1 - n^{-y'}$, we have

$$|\deg(i) - \lambda_i| < \epsilon z < \max\left(\epsilon\lambda_i, \frac{2y \ln n}{\epsilon}\right) < \frac{C'\sqrt{n}}{i^{3/2}} < \delta_{i+1}.$$

(C' is another constant we can make as small as we want by setting C). □

Lemma 3.6 *With probability $1 - o(n^{-1})$ every vertex $i \leq n/\log^8 n$ is uniquely determined by the degrees of its neighbors.*

Proof. Now consider the neighbors of vertex i. We want to apply Corollary 3.4 in this case, so we are interested in which values of d and i are such that $\mathbb{E}[|N_c(i)|] \geq 2$, where $N_c(i)$ stands for the neighbors of i of degree (color) c. Suppose $i \leq \frac{n}{\log^8 n}$.

$$\begin{aligned}\mathbb{E}[|N_c(i)|] &= \sum_{j\,:\,d(j)=c} P[i,j] = \sum_{j\,:\,d(j)=c} \frac{1}{\sqrt{ij}} \\ &= \Omega\left(\frac{n}{d^3\sqrt{i\,n}}\right) = \Omega\left(\frac{\sqrt{n}}{d^3\sqrt{i}}\right).\end{aligned}$$

So if $i \leq \frac{n}{\log^8 n}$, then

$$\mathbb{E}[|N_c(i)|] \geq \Omega\left(\frac{\sqrt{n}}{c^3\sqrt{\frac{n}{\log^8 n}}}\right) = \Omega\left(\frac{\log^4 n}{c^3}\right)$$

which will be at least 2 for $c \leq T \log n$, for any constant T. Notice further that all vertices $j \leq \sqrt{n}$ have expected degree $\Omega\left(n^{1/4}\right)$, and hence, by Lemma 3.1 have degree greater than $T \log n$. Therefore for the colors c we are considering, we have $\gamma_c \leq 1/\sqrt{n}$, and $\lambda_c\gamma_c \leq 1$. So by Corollary 3.4, with probability $o(n^{-1})$, all vertices $i \leq \frac{n}{\log^8 n}$ will be uniquely determined by the degrees of their neighbors. □

We now turn our attention to other vertices. Notice that there may be some isolated vertices, or vertices that are part of very small connected components, and we cannot hope to distinguish these. Notice further that a constant fraction of vertices will have a constant number of neighbors, so we cannot hope to distinguish them just by the degrees of the neighbors. Moreover some of the vertices of low degree will share all their neighbors. We call such sets of vertices "equivalent." However we can still say something about our general algorithm. Recall this algorithm first colors vertices by their degree sequence, and then runs partition-refinement starting from this coloring. We can show that with high probability all vertices that are connected to the giant connected component of the graph will be distinguished by this algorithm from all other vertices they are not equivalent to.

Lemma 3.7 *If v is a vertex that is connected to the giant connected component, then with probability $1 - o(n^{-1})$, the main algorithm will distinguish v from all other vertices exception those equivalent to v.*

Sketch of the proof. This proof is more technical than the previous ones. First of all, we only need to consider vertices $i > \frac{n}{\log^8 n}$. The idea is to look at two such vertices u, and v, and grow two BFS starting at u and v. If any BFS layer for v has a different number of nodes than u then we distinguished u and v. If at any point the BFS layer is large enough, then with high probability it will contain some vertices $i \leq \frac{n}{\log^8 n}$, and with high probability these will be different for u and v, and since these are already uniquely labelled we are done. The important detail that we swept under the rug is that we need to maintain the property that the BFS levels for u and v remain disjoint enough unless they are equal after one step (in which case u and v are equivalent). □

4 Stochastic Kronecker Graphs

4.1 Motivation

Kronecker prodcut graphs and stochastic Kronecker product graphs were introduced to model real world networks [26, 27]. From the theoretical side, Mahdian and Xu [30] studied the case of 2×2 initiator matrices, finding thresholds for connectivity, the emergence of a giant connected component, and proving that the effective diameter is shrinking for certain ranges of the parameters. While the case of larger initiator matrices has been used in many contexts(see [27]), we are not aware of theoretical results in this case.

Further motivation to study initiator matrices of arbitrary size arises from the relationship to a different random graph model. It has been shown that, in the limit to an infinite number of nodes, the adjacency matrix of a dense graph is well approximated by a continuous function on the unit square [28, 16] (see also [11, 14] for similar results in the sparse graph case). This fact has been used to generate "multifractal random networks" by a continuous analog to the Kronecker product [34]. An initiator matrix of arbitrary size can be used as an approximation to a continuous function on the unit square, and hence stochastic Kronecker product graphs with large initiator matrices can model multifractal random networks. One important difference between the two models is that multifractal random networks require row sums of 1, and hence in the infinite limit, the graphs are not interesting: $n - o(n)$ vertices will be isolated [34, Supporting Information]. The stochastic Kronecker graph model is more general in the sense that row sums can be arbitrary. As we will see, the graphs will be connected exactly when the row sums are greater than 1.

4.2 Definitions

We will denote by $A[i,j]$ or A_{ij} the entry in row i, column j of a matrix A.

The Kronecker product of an $m \times n$ matrix A with entries $a_{i,j} = A[i,j]$, and a $p \times q$ matrix B with entries $B[k,l]$ is the $mp \times nq$ matrix $C = A \otimes B$. We will index the rows of C by pairs $(ik) \in [m] \times [p]$, and the columns by pairs $(jl) \in [n] \times [q]$. The entries of C are defined by: $C[(ik),(jl)] = A[i,j]B[k,l]$. In other words,

$$A \otimes B \;=\; \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

Given an $r \times r$ initiator matrix θ, with entries in $[0,1]$, and an integer k, we construct a random graph $G(r^k, \theta)$ with $n = r^k$ vertices, each vertex labelled by a unique string of length k over the alphabet $[r]$. We add each edge independently at random. Given two vertices u and v with labels $u_1u_2\ldots u_k$ and $v_1v_2\ldots v_k$ resp., the probability with which we chose edge (u,v) is $P[u,v] = \prod_{i=1}^{k} \theta[u_i, v_i]$. In other words, the matrix $P = \theta^{\otimes k} = \theta \otimes \theta \otimes \cdots \otimes \theta$ is the Kronecker product of θ with itself k times.

For a vertex u with label $u_1u_2\ldots u_k$, and $j \in [r]$, let w_j be the number of coordinates i such that $u_i = j$. We say that u has weight vector $(w_1, \ldots, w_r)$.

4.3 Expected Degrees and Concentration Results

First we prove that the expected in- and out-degrees of the vertices are distributed like multinomials.

Lemma 4.1 *The expected in- and out-degree of a vertex u with weight vector $(w_1, \ldots, w_r)$ are:*

$$\mathbb{E}[d_u^+] = (\theta_{11} + \cdots + \theta_{1r})^{w_1}(\theta_{21} + \cdots + \theta_{2r})^{w_2} \ldots (\theta_{r1} + \cdots + \theta_{rr})^{w_r};$$

and

$$\mathbb{E}[d_u^-] = (\theta_{11} + \cdots + \theta_{r1})^{w_1}(\theta_{12} + \cdots + \theta_{r2})^{w_2} \ldots (\theta_{1r} + \cdots + \theta_{rr})^{w_r}.$$

Proof. We prove the result for the out-degree, d_u^+, the result for in-degree is proven analogously. For notational convenience, let $\theta_i = \theta_{r1} + \cdots + \theta_{rr}$ be the row-sums of the initiator matrix θ.

For every vertex v, let $I_{(u,v)}$ be the indicator function of the event that (u,v) is an edge. Then $\mathbb{E}[d_u^+] = \mathbb{E}[\sum_v I_{(u,v)}]$. By the definition of the Kronecker product, $\mathbb{E}[I_{(u,v)}] = \mathbb{P}[(u,v)] = \prod_{i=1}^{k} \theta[u_i, v_i]$. Therefore we have $\mathbb{E}[d_u^+] = \sum_v \prod_{i=1}^{k} \theta[u_i, v_i]$. For $\ell, m \in [r]$, let $n_{\ell,m}$ be the number of coordinates i such that $u_i = \ell$, and $v_i = m$. Note that $\mathbb{P}[(u,v)] = \prod_{\ell,m \in [r]} \theta[\ell, m]^{n_{\ell,m}}$; therefore we can re-write the above sum as:

$$
\begin{aligned}
\mathbb{E}[d_u^+] &= \sum_{n_{1,1}+\cdots+n_{1,r}=w_1} \binom{w_1}{n_{1,1},\ldots n_{1,r}} \\
&\left(\cdots\left(\sum_{n_{r,1}+\cdots+n_{r,r}=w_r} \binom{w_r}{n_{r,1},\ldots n_{r,r}} \prod_{\ell=1}^{r}\prod_{m=1}^{r} \theta[\ell,m]^{n_{\ell,m}}\right)\cdots\right) \\
&= \sum_{n_{1,1}+\cdots+n_{1,r}=w_1} \binom{w_1}{n_{1,1},\ldots n_{1,r}} \prod_{m=1}^{r} \theta[1,m]^{n_{1,m}} \\
&\left(\cdots\left(\sum_{n_{r,1}+\cdots+n_{r,r}=w_r} \binom{w_r}{n_{r,1},\ldots n_{r,r}} \prod_{m=1}^{r} \theta[r,m]^{n_{r,m}}\right)\cdots\right) \\
&= \sum_{n_{1,1}+\cdots+n_{1,r}=w_1} \binom{w_1}{n_{1,1},\ldots n_{1,r}} \prod_{m=1}^{r} \theta[1,m]^{n_{1,m}} \\
&\left(\sum_{n_{2,1}+\cdots+n_{2,r}=w_2} \cdots\right)(\theta[r,1]+\cdots+\theta[r,r])^{w_r} \\
&= (\theta_1)^{w_1}(\theta_2)^{w_2}\ldots(\theta_r)^{w_r}.
\end{aligned}
$$

□

Lemma 3.1 gives us concentration results for the degrees of the vertices.

4.4 Connectedness

We now turn our attention to determining for what initiator matrices the stochastic Kronecker graphs will be connected.

As above, we have $P = \theta^{\otimes k}$. For subsets $S, T \subseteq V$, let $P[S,T] = \sum_{s\in S, t\in T} P[s,t]$.

We use the following result:

Theorem 4.2 ([30, Theorem 1]) *There exists an (absolute) constant c such that if $\forall S \subset V,\ S \neq \emptyset,\ P(S, V\setminus S) \geq c\ln n$, then with high probability $G(n,P)$ is connected.*

For u, v in V, let $e(u,v)$ be the edit distance between u and v, that is, the number of coordinates i such that $u_i \neq v_i$.

First we consider the setting where the initiator matrix has no zero entries. In this case, we define $\theta_{\min} = \min_{i,j}\theta[i,j]$, $\theta_{\max} = \max_{i,j}\theta[i,j]$, and $\theta_R = \theta_{\max}/\theta_{\min}$. The following lemma plays the role of [30, Lemma 2].

Lemma 4.3 *Suppose all entries of the initiator matrix are non-zero. Then for ever $u, v \in V$, and every $S \subseteq V$,*

$$\frac{P[v,S]}{\theta_R} \leq P[u,S] \leq \theta_R P[v,S].$$

Proof. Without loss of generality, u and v differ in the last coordinate, r.

$$\begin{aligned} P[u,S] &= \sum_{w\in S} P[u,w] = \sum_{w\in S}\prod_{i=1}^{r}\theta_{u_i\,w_i} \geq \theta_{\min}\sum_{w\in S}\prod_{i=1}^{r-1}\theta_{u_i\,w_i} \\ &= \theta_{\min}\sum_{w\in S}\prod_{i=1}^{r-1}\theta_{v_i,w_i} \geq \theta_{\min}\sum_{w\in S}\prod_{i=1}^{r}\theta_{v_i\,w_i}/\theta_{\max} \\ &= \sum_{w\in S} P[v,w]/\theta_R. \end{aligned}$$

□

We are now able to describe sufficient and necessary conditions for connectivity.

Theorem 4.4 *Let θ be symmetric, with no zero entries. Then $G(n,P)$ is connected with high probability if and only if the row sums satisfy $\sum_j \theta_{ij} > 1$ for every i.*

Proof. We first prove the "only if." If $\exists i$ such that $\theta_i < 1$, then with constant probability, the vertex $\overrightarrow{i} = (i,i,\ldots,i)$ has expected degree $\theta_i^k = o(1)$, hence it will be isolated with probability 1. If $\theta_i = 1$, then either $\overrightarrow{i}$ is disconnected with probability at least e^{-2}, or row i is deterministic (exactly one 1 and every other position 0). But the second case cannot happen since we assumed θ doesn't have any zero entries. The proof follows along the lines as the proof that $\overrightarrow{0}$ is isolated in [30, Theorem 4]:

$$\mathbb{P}[\overrightarrow{i} \text{ isolated}] = \prod_v (1 - P[\overrightarrow{i}, v])$$

$$= \prod_{w_1+\cdots+w_r=k} (1-\theta[i,1]^{w_1}\theta[i,2]^{w_2}\ldots\theta[i,r]^{w_r})^{\binom{k}{w_1,\ldots,w_r}}$$

$$\geq \prod_{w_1+\cdots+w_r=k} \exp\left(-2\binom{k}{w_1,\ldots,w_r}\theta[i,1]^{w_1}\ldots\theta[i,r]^{w_r}\right)$$

$$= \exp\left(-2\sum_{w_1+\cdots+w_r=k}\binom{k}{w_1,\ldots w_k}\theta[i,1]^{w_1}\ldots\theta[i,r]^{w_r}\right)$$

$$= \exp\left(-2(\theta[i,1]+\cdots+\theta[i,r])^k\right) = \exp\left(-2(\theta_i)^k\right).$$

So, as claimed, if $\theta_i < 1$, the probability that vertex $\overrightarrow{i}$ is isolated is $1-o(1)$, if $\theta_i = 1$, then vertex $\overrightarrow{i}$ is isolated with probability at least e^{-2}.

Now for the other direction. Suppose $\theta_i > 1$ for every i. Our goal is to show that P satisfies the conditions of Theorem 4.2. Let $S \subset V$, $S \neq \emptyset$. There exist $u \in S$ and $v \in V \setminus S$ that differ in only 1 coordinate. We have

$$[u, V\setminus\{u\}] = \mathbb{E}[\deg(u)] = \prod_{i=1}^{r}(\theta_i)^{w_i(u)} \geq (\theta_{\min})^k.$$

Since $k = \log n$ (in some base), for k large enough, we have $(\theta_{\min})^k \geq (\theta_R + 1)c \ln n$. Now either $P[u, V \setminus S] \geq c \ln n$ or $P[u, S] \geq c\,\theta_R \ln n \implies P[v, S] \geq c \ln n$. The last implication follows by Lemma 4.3. And therefore $P[S, V \setminus S] \geq c \ln n$. □

In Appendix A we discuss possible extensions of the connectedness result to the setting where the initiator matrix has zero entries.

4.5 Distinguishing vertices

Theorem 4.5 *In a stochastic Kronecker graph, with high probability every vertex that has expected degree at least n^ϵ, for any constant ϵ will be distinguished after one round of naive partition-refinement.*

Sketch of the proof. By Lemma 3.1, and Lemma 4.1, with high probability all vertices of degree at least n^ϵ will have degree at least $n^{\epsilon/2}$. The result follows along the lines of the proofs of Lemmas 3.5 and 3.6. □

Remark. In the setting where the initiator matrix has no zero entries and the graph is connected with high probability, then by Lemma 4.1 the expected degree of every vertex is at least n^ϵ for some epsilon, so the theorem above allows us to distinguish all vertices. In particular we obtain a canonical labeling algorithm.

We expect that it is possible to extend the above theorem to other vertices, using a more careful analysis.

5 Experimental results for preferential-attachment graphs

We now consider preferential attachment graphs. This model was first suggested by Albert-Barabasi [3], and analyzed more carefully by Bollobas et al. [15, 12, 13]. We consider the model of Bollobas et al. [15]. There is one parameter, m, which is half the average degree. A preferential attachment (multi)graph with n vertices is constructed as follows. Start with two vertices and m edges between them. Now add the remaining vertices one at a time. When adding a vertex v, add m edges to it, where each edge (v, u) has probability proportional to $\deg(u)$. The idea of this model is that "popular" vertices are more likely to be linked in the future. For $m = 1$, the graph will be a tree, and it is known that trees (non-random) are canonically labelled by the partition-refinement technique (see [35]). For $m = 2$, with constant probability there will be at least one pair of vertices of degree 2 with exactly the same neighbors; hence these vertices cannot be distinguished. This is easy to compute directly (see [23]). For $m \geq 3$ with high probability there are no such pairs (again see [23]).

For both $m = 2$ and $m = 3$ we generated 10^5 graphs with 10^5 vertices by preferential attachment. We then ran the naive partition-refinement algorithm and computed distance sequences (recall the definitions from subsection 2.2). Both algorithms distinguished all vertices of the graphs, except (in the case $m = 2$) those pairs of degree 2 vertices that share the same neighbors. Indeed naive partition refinement converged in at most three rounds on all graphs.

In Appendix B we discuss some preliminary thoughts on proving that partition refinement and the distance sequence distinguish all vertices for almost all preferential attachment graphs, as the experiments suggest.

6 Conclusions and future work

The gap between mathematically nice random models and real-world networks is an interesting and promising area of research. We show that we can distinguish vertices and then develop efficient algorithms for several important families of real-world networks. On the other hand, there are examples of real-world networks for which canonical labeling is difficult for these simple algorithms.

It is interesting to study the question of distinguishing vertices in other random graph models. We believe that the results for stochastic Kronecker graphs apply in other settings of the parameters; extending our results requires understanding the behavior of the low-degree vertices in this setting.

Another direction is to prove theoretical results for models where edges are not added independently, such as preferential attachment. We suspect that for most models vertices will be distinguished in a few rounds of the combinatorial algorithms described in this paper.

Finding "hard" examples for graph isomorphism is a notoriously difficult task. If there were models where distinguishing vertices is difficult, this would provide a class of such graphs. So both positive and negative results in this direction are interesting.

References

[1] M. Abramowitz and I.A. Stegun, editors. *Handbook of Mathematical Functions*. Number 55 in Applied Mathematics Series. National Bureau of Standards, tenth edition, 1972.

[2] William Aiello, Fan Chung, and Linyuan Lu. A random graph model for massive graphs. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, STOC '00, pages 171–180, New York, NY, USA, 2000. ACM.

[3] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[4] L. Babai, P. Erdős, and S.M. Selkow. Random graph isomorphism. *SIAM Journal on Computing*, 9(3):628–635, 1980.

[5] L. Babai and L. Kucera. Canonical labelling of graphs in linear average time. In *Foundations of Computer Science, 1979., 20th Annual Symposium on*, pages 39–46. IEEE, 1979.

[6] László Babai. Monte carlo algorithms in graph isomorphism testing. In *Université de Montréal Technical Report*, volume 70 of *DMS*, page 42, 1979. `http://people.cs.uchicago.edu/~laci/lasvegas79.pdf`.

[7] László Babai and Eugene M. Luks. Canonical labeling of graphs. In *Proc. 15th STOC*, pages 171–183. ACM Press, 1983.

[8] László Babai and Shlomo Moran. Arthur-Merlin games: A randomized proof system, and a hierarchy of complexity classes. *Journal of Computer and System Sciences*, 36(2):254 – 276, 1988.

[9] B. Bollobás. Distinguishing vertices of random graphs. *North-Holland Mathematics Studies*, 62:33–49, 1982.

[10] B. Bollobás. *Random graphs*, volume 73. Cambridge University Press, 2001.

[11] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

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

[13] B. Bollobás and O. Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 1(1):1–35, 2004.

[14] B. Bollobás and O. Riordan. Sparse graphs: Metrics and random models. *Random Structures & Algorithms*, 39(1):1–38, 2011.

[15] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Struct. Algorithms*, 18(3):279–290, May 2001.

[16] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851, 2008.

[17] Nicholas A. Christakis and James H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379, 2007.

[18] P. Cordella, L. P. Foggia, C. Sansone, and M. Vento. Evaluating performance of the vf graph matching algorithm. In *Proc. of the 10th International Conference on Image Analysis and Processing*, pages 1172–1177. IEEE Computer Society Press, 1999.

[19] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

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

[21] H. Katebi, K. Sakallah, and I. Markov. Conflict anticipation in the search for graph automorphisms. In *Logic for Programming, Artificial Intelligence, and Reasoning*, pages 243–257. Springer, 2012.

[22] Myunghwan Kim and Jure Leskovec. Multiplicative attribute graph model of real-world networks. *Internet Mathematics*, 8(1-2):113–160, 2012.

[23] R.D. Kleinberg and J.M. Kleinberg. Isomorphism and embedding problems for infinite limits of scale-free graphs. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 277–286. Society for Industrial and Applied Mathematics, 2005.

[24] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 57–65. IEEE, 2000.

[25] S. Lattanzi and D. Sivakumar. Affiliation networks. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 427–434. ACM, 2009.

[26] J. Leskovec, D. Chakrabarti, J. Kleinberg, and C. Faloutsos. Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication. *Knowledge Discovery in Databases: PKDD 2005*, pages 133–145, 2005.

[27] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, and Z. Ghahramani. Kronecker graphs: An approach to modeling networks. *The Journal of Machine Learning Research*, 11:985–1042, 2010.

[28] L. Lovász and B. Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006.

[29] Eugene M. Luks. Isomorphism of graphs of bounded valence can be tested in polynomial time. *J. Comp. Sys. Sci.*, 25:42–65, 1982.

[30] Mohammad Mahdian and Ying Xu. Stochastic kronecker graphs. In *Proc. 5th international conference on algorithms and models for the web-graph*, WAW'07, pages 179–186, 2007.

[31] Brendan D. McKay. Practical graph isomorphism. In *Congressus Numerantium, 30*, pages 45–87, 1981.

[32] T. Miyazaki. The complexity of McKay's canonical labeling algorithm. In L. Finkelstein and W.M. Kantor, editors, *Groups and Computation, II*, pages 239–256. Amer. Math. Soc., 1997.

[33] M. Newman. *Networks: An Introduction*. Oxford University Press, Inc., 2010.

[34] G. Palla, L. Lovász, and T. Vicsek. Multifractal network generator. *Proceedings of the National Academy of Sciences*, 107(17):7640–7645, 2010.

[35] R. C. Read and D. G. Corneil. The graph isomorphism disease. *Journal of Graph Theory*, pages 339–363, 1977.

[36] M. Salter-Townshend, A. White, I. Gollini, and T. B. Murphy. Review of statistical network analysis: models, algorithms, and software. *Statistical Analysis and Data Mining*, 5(4):243–264, 2012.

[37] H Wang, Y.˙ On the number of successes in independent trials. *Statistica Sinica*, 3:295–312, 1993.

[38] Gezhi Weng, Upinder S Bhalla, and Ravi Iyengar. Complexity in biological signaling systems. *Science*, 284:92–96, 1999.

[39] V. Zemlyachenko, N. Kornienko, and R. Tyshkevich. Graph isomorphism problem. *The Theory of Computation I*, 118, 1982.

A About extensions of the connectedness result for stochastic Kronecker graphs

If θ has zero entries, the situation is more complicated.

It still holds that if $\theta_i < 1$ for some i, then $\overrightarrow{i}$ is disconnected with probability 1 and hence G is not connected. It is also still true that if $\theta_i = 1$, then the ith row must be deterministic. Moreover, it is not hard to see that if θ is periodic, then P will be disconnected.

Therefore the following are necessary (but not sufficient) conditions for G to be connected with high probability:

1. $\theta_i \geq 1 \ \forall i$ and if $\theta_i = 1$, then the ith row is deterministic.

2. θ is aperiodic.

In view of Theorem 4.2 [30, Theorem 1], a way to prove connectedness is to bound the size of any cut. One way to do this is to look at the second eigenvalue of the Laplacian. Unfortunately, while eigenvalues behave nicely under Kronecker product, eigenvalues of the Laplacian do not. A more promising approach is to prove the conjecture below:

Conjecture A.1 *There exist constants c_1 and c_2 so that if θ is aperiodic and for every vertex $v \in [r]^k$, the cut that separates v from the rest of the graph has weight at least $c_1 \log n$, then every cut of P has weight at least $c_2 \log n$.*

If the conjecture holds, it would be sufficient to prove that with high probability there are no isolated vertices.

B Towards canonical labeling of preferential attachment graphs

The general idea for the algorithm is as follows. The first step of the proof is to show that all vertices that have degree at least n^ϵ, for some constant ϵ, will be distinguished after one round of refinement. The top vertices will have degree at least n^ϵ. Indeed Bollobas and Riordan [13] note that with probability $1 - o(n^{-\delta})$ (for a fixed constant δ), every vertex i with expected degree $\mathbb{E}[d(i)] \geq 2n^\epsilon$ has $d(i) \geq n^\epsilon$. Bollobas and Riordan do not make any attempt to optimize δ. However, based on the proof of [12, Lemma 7] (specifically property E_1) the statement should be true for $\delta = 5$ or 6. In particular, we will have many vertices that will be distinguished after one round of refinement.

The second step is to take care of vertices of small degree. There are two lines of attack, using ideas from [12, 13], and it is likely that using a combination will be useful. The first idea is to use the fact that with probability $1 - o(n^{-\delta})$, all vertices are connected by a path of length at most $8 \log \log n$ to early vertices [12, Lemma 8] (in this case δ is small: Bollobas-Riordan prove their results for $\delta = 1/1000$). Then we use the fact that early vertices are already distinguished, and ideas about neighborhood growth from [13] to show that vertices will be distinguished either by their distance sequence up to some small value ($O(\log \log n)$), or by the closest early vertices.

B.1 Degree distribution of vertices from t on

Bollobas, Riordan, Spencer and Tusnady [15] studied the degree distribution of graphs generated by preferential attachment. We are interested in the degree distribution of the neighbors of a vertex v. The out-degree distribution (vertices that were added before v) is given by the construction. The in-degree distribution requires lookin at vertices added after v. As a first step in that direction we characterize the degree distribution of vertices added after time t.

Extending the notation of [15], let $\#_m^{n,t}(d)$ be the number of vertices in G_m^n that have degree $m+d$ and are added at or after time t. Let $r = t/n$.

Lemma B.1

$$\mathbb{E}[\#_m^{n,t}(d)] \sim \frac{2m(m+1)n}{(d+m)(d+m+1)(d+m+2)} \, \mathrm{I}_{1-\sqrt{k}}(d+1, m+2)$$

$$= \mathbb{E}[\#_m^{n}(d)] \;\; \mathrm{I}_{1-\sqrt{k}}(d+1, m+2)$$

where $I_s(a,b) = B_s(a,b)/B(a,b)$ is the regularized incomplete beta function.

We should be able to get high confidence bounds like the ones in [15], proving that the error is at most an additive factor of $\sqrt{n \log n}$ with probability 1.

Since

$$\mathrm{I}_s(a,b) = \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} s^j (1-s)^{a+b-1-j},$$

we have:

$$\mathbb{E}[\#_m^{n,t}(d)] \sim \frac{2m(m+1)n}{(d+m)(d+m+1)(d+m+2)}$$
$$\sum_{j=d+1}^{d+m+2} \binom{d+m+2}{j} (1-\sqrt{r})^j \sqrt{k}^{d+m+2-j}.$$

Remark. Our main use of this lemma will be for relatively small values of r ($r \le n^{1-\epsilon}$). It is likely that the estimates we need for these values of r can be obtained more immediately, but we keep the result as general as possible.

Proof.

The proof follows along the lines of Bollobas et al. Here is a sketch. They obtain

$$\mathbb{P}[d'_{k+1} = d+m]$$
$$= o(n^{-1}) + (1+o(1)) \binom{d+m-1}{m-1} k^{m/2} (1-\sqrt{k})^d,$$

and compute this by computing the integral

$$\int_0^1 k^{m/2} (1-\sqrt{k})^d dk.$$

We want to compute the same integral, but stop early, namely start at r instead of 0. As in Bollobas et al., we substitute $k = (1-u)^2$.

$$\int_r^1 k^{m/2} (1-\sqrt{k})^d dk = 2 \int_0^{1-\sqrt{r}} (1-u)^{m+1} u^d du$$
$$= 2B_{1-\sqrt{r}}(d+1, m+2)$$

Where $B_s(a,b)$ is the incomplete Beta function. Therefore the estimate is the same as the estimate for $\mathbb{E}[\#_m^n(d)]$ in [15], except the beta function is replaced by the incomplete beta function. The result follows. □