

Article

A General Computational Approach for Counting Labeled Graphs

Ravi Goyal ^{1,*}  and Victor De Gruttola ²

¹ Division of Infectious Diseases and Global Public Health, University of California San Diego, La Jolla, CA 92093, USA

² Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

* Correspondence: r1goyal@health.ucsd.edu

Abstract: This paper presents a general recursive formula to estimate the number of labeled graphs as well as details to evaluate the formula for the following graph properties: number of edges (graph density), degree sequence, degree distribution, classification mixing, and degree mixing, i.e., the formula estimates the number of labeled graphs that have given values for graph properties. The proposed approach can be extended to additional graph properties (e.g., number of triangles) as well as properties of bipartite graphs. For special settings in which formulas exist from previous research, simulation studies demonstrate the validity of the proposed approach. In addition, we demonstrate how our approach can be used to quantify the level of variability in values of a graph property in the subset of graphs that hold a specified value of a different graph property (or properties) constant.

Keywords: labeled graphs; graph enumeration; degree sequence; computational approach

1. Introduction

Graph enumeration is a well-established area of combinatorics for counting graphs with particular features. Examples of such enumeration include determining how many graphs exist with a given number of vertices or edges, or a given degree sequence. Approaches for counting graphs fall into two categories based on whether their vertices are labeled or unlabeled. In the former case, vertices of a graph are labeled in a way that makes them distinguishable from one another. In the latter, all permutations of vertices are considered to form the same graph [1]. In social network analysis—our area of interest—vertices are most often distinguishable from each other; hence, we focus on labeled graph enumeration.

Recently, Iniquez et al. discussed bridging the gap between graph theory and social network analysis [2]. As they note, connecting these two disciplines may clarify the role of randomness in modeling dynamical systems, such as those that govern spread of infectious diseases within a population. Simulation studies that investigate such spread and the impact of interventions on it often make use of agent-based models (ABMs) [3,4]. An important component of ABMs is the formulation of interactions among the agents in the model; these interactions can be represented as a graph. We refer to the collection of such interactions as contact networks (in keeping with infectious disease transmission literature). Often ABMs generate the contact network from a stochastic process. Graph enumeration can aid in interpreting results of simulation studies of processes that operate on graphs (e.g., spread of infection) by permitting assessment of the contribution to variation in the results (e.g., total number infected in unit time) that arises from variation in the generated contact networks. Higher levels of the latter might be expected to lead to higher levels of the former. We provide an application of our methods to demonstrate how graph enumeration can help in quantifying variation in graphs.

Current solutions to graph enumeration problems are individually tailored to particular properties (such as degree sequence) [5]. These solutions are either closed-form mathematical expressions or asymptotic formulas. Equations to calculate the number of



Citation: Goyal, R.; De Gruttola, V. A General Computational Approach for Counting Labeled Graphs. *Algorithms* **2023**, *16*, 16. <https://doi.org/10.3390/a16010016>

Academic Editor: Mircea-Bogdan Radac

Received: 23 September 2022

Revised: 15 December 2022

Accepted: 16 December 2022

Published: 27 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

labeled graphs with various characteristics have been reported; these include rooted graphs, connected graphs, and directed graphs [1]. Considerable research has been devoted to estimating the number of labeled graphs with a given degree sequence—a property important in social network analysis [6–9]. However, there has been little research focusing on other important properties in social network analysis, such as degree mixing and number of triangles.

Below, we propose a general approach for counting labeled graphs that applies to several graph properties, including degree sequence. Furthermore, our approach deviates from the standard one of developing a closed-form or asymptotic formula. By contrast, we propose an algorithmic method to the graph enumeration problem. The next section provides terminology used in the paper. Section 3 presents a general recursive formula to estimate the number of labeled graphs as well as details to evaluate the formula for the following graph properties: number of edges (graph density), degree sequence, degree distribution, classification mixing, and degree mixing. For settings in which formulas exist from previous research, Section 4 presents simulation studies demonstrating the degree of similarity between our proposed methods and those that are currently available. In Section 5, we apply the proposed approach to estimate the number of labeled graphs associated with different values of degree distribution and degree mixing that arise from the Barabási–Albert model to investigate the variation across graphs generated with this model [10]. The paper concludes with a discussion and further research.

2. Terminology

We represent a graph, $G = (V, E)$, as an adjacency matrix with dimensions equal to the size of set V . Therefore, G has dimensions $|V| \times |V|$, where $|V|$ denotes the size of set V . Let n and m represent the number of vertices in G , i.e., $n = |V|$ and number of edges, i.e., $m = |E|$, respectively. Let $\{v_1, \dots, v_n\}$ denote the vertices in set V , which are labeled (arbitrarily) but enables them to be distinguishable from one another. Let (v_i, v_j) denote an edge between v_i and v_j . Let $G[i, j] = 1$ indicate that there is an edge between v_i and v_j , where $v_i, v_j \in \{v_1, \dots, v_n\}$, while $G[i, j] = 0$ indicates that there is no edge. Denote the neighbors of v_i as $\eta(v_i)$, i.e., $\eta(v_i) = \{v_j : G[i, j] = 1\}$. Let \mathcal{G}_n be the set of all simple labeled graphs with n vertices.

Let ϕ_1 denote an algebraic map from a graph G to its number of edges, i.e., $\phi_1(G) = m$. The degree of vertex v_i , denoted as $d_i(G)$, is the number of edges the vertex has with other vertices in V ; therefore $d_i(G) = \sum_j G[i, j]$. Let $d(G) = (d_1(G), \dots, d_n(G))$ represent the vector of degrees for nodes in set V , commonly referred to as a degree sequence. The degree distribution, denoted as $D(G)$, is a vector representing the number of these degrees over all vertices in set V ; the k th entry represents the number of vertices with degree k , i.e., $D_k(G) = \sum_{i=1}^n I_{\{d_i(G)=k\}}$. Let ϕ_2 and ϕ_{2a} denote the mapping from a graph to its degree distribution, i.e., $\phi_2(G) = D(G)$, and degree sequence, i.e., $\phi_{2a}(G) = d(G)$, respectively.

Let $m_i(G)$ represent a discrete classification for vertex v_i in graph G ; we denote the number of distinct classifications as q . Let $m(G) = (m_1(G), \dots, m_n(G))$ be a vector containing the characteristics of all vertices. The classification distribution, denoted as $M(G)$, is a vector representing the number of individuals with these classification over all vertices; the k th entry represents the number of vertices with classification k , i.e., $M_k(G) = \sum_{i=1}^n I_{\{m_i(G)=k\}}$. Let $MM(G)$ be a $q \times q$ symmetric matrix representing the mixing by classification of graph G ; we refer to $MM(G)$ as a classification mixing matrix. The entry $MM_{k,l}(G)$ is the total number of edges between a vertex with classification k and vertex with classification l . Let ϕ_3 denote the mapping from a graph to its classification mixing matrix. Let $DMM(G)$ be a particular mixing matrix where the classification represents vertex degrees. Therefore, the entry $DMM_{k,l}(G)$ is the total number of edges between vertices of degree k and l . Let ϕ_4 denote the mapping from a graph to its degree mixing matrix.

Denote the inverse images associated with a map ϕ_i as $c_{\phi_i}(x) = \{G : \phi_i(G) = x, G \in \mathcal{G}_n\}$. These inverse images of singleton sets have been referred to as fibers in algebraic statistics literature [11]. The graph enumeration problem is calculating the size of a fiber,

denoted as $|c_{\phi_i}(x)|$, which represents the number of graphs where the graph property associated with ϕ_i equals x ; this quantity has been referred to as a volume factor [12]. We refer to x as a graphical value associated with ϕ_i if $|c_{\phi_i}(x)| \geq 1$.

3. Methods

This section first presents a general recursive formula to estimate the number of labeled graphs; details for specific graph properties, e.g., degree distribution, follow afterwards. Equation (1) provides a recursive formula to estimate the number of graphs, $|c_{\phi_i}(x_k)|$, with specific value(s), x_k , for particular graph properties associated with ϕ_i :

$$|c_{\phi_i}(x_k)| = r_{\phi_i}(x_k, x_{k-1}) * |c_{\phi_i}(x_{k-1})|, \tag{1}$$

where $r_{\phi_i}(x_k, x_{k-1})$ is the ratio between the sizes of fibers $c_{\phi_i}(x_k)$ and $c_{\phi_i}(x_{k-1})$, i.e.,

$$r_{\phi_i}(x_k, x_{k-1}) = \frac{|c_{\phi_i}(x_k)|}{|c_{\phi_i}(x_{k-1})|}. \tag{2}$$

Goyal et al. [13] provides equations to calculate $r_{\phi}(x_i, x_{i-1})$ for a range of graph properties including number of edges, classification mixing, degree distribution, degree mixing, and number of triangles (controlling for degree mixing) when x_i and x_{i+1} are specified such that there exists graphs G_i and G_{i-1} where:

- s1. G_i and G_{i-1} differ by the presence or absence of a single edge;
- s2. $\phi_i(G_i) = x_i$; and
- s3. $\phi_i(G_{i-1}) = x_{i-1}$.

To make use of the recursive formula and previous work by Goyal et al. [13], it is necessary to specify a sequence of values $\{x_0, \dots, x_k\}$ such that there exists graphs $\{G_0, \dots, G_k\}$ where each consecutive pair satisfies s1 – s3. In addition, we need to be able to calculate $|c_{\phi_i}(x_0)|$. Although there is no constraint on x_0 , it is often useful to set x_0 equal to the specific value of the graph properties associated with the empty graph; hence, typically, $|c_{\phi}(x_0)| = 1$. Throughout this paper, we follow this approach. In the sections below, we provide details for calculating $|c_{\phi_i}(x)|$ for when ϕ_i and x are associated with a given number of edges, degree distribution, degree sequence, classification mixing matrix, and degree mixing matrix.

3.1. Graph Enumeration Problem: Calculate the Number of Labeled Graphs of Size n with m Edges

In this section, we calculate $|c_{\phi_1}(m)|$. To address this graph enumeration problem, we specify the following sequence of number of edges for the recursive procedure: $x_0, \dots, x_k = m$ where $x_i = i$. Theorem 1 proves that there exists a collection of graphs that is consistent with this specification of number of edges, i.e., there exists a collection of graphs G_0, \dots, G_k where each consecutive pair satisfies s1 – s3 for ϕ_1 .

Theorem 1. For a sequence of number of edges: x_0, \dots, x_k where $x_i = i$ and $k \leq \binom{n}{2}$, there exists a collection of graphs G_0, \dots, G_k where each consecutive pair satisfies s1 – s3 for ϕ_1 and $G_i \in \mathcal{G}_n$ for all $i \in \{0, \dots, k\}$.

Proof. Let $E = \{e_1, \dots, e_k\}$ be a set of distinct edges among vertices $\{1, \dots, n\}$; this is possible because $k \leq \binom{n}{2}$. Let G_i denote the graph formed with the first i edges from E , i.e., G_i contains edges $\{e_1, \dots, e_i\}$. Based on the definition of G_i , $\phi_1(G_i) = x_i = i$, $\phi_1(G_{i-1}) = x_{i-1} = i - 1$, and G_i and G_{i-1} differ by a single edge. \square

Since x_0, \dots, x_k satisfies s1 – s3, we can use results from Goyal et al. [13] to calculate $r_{\phi_1}(x_i, x_{i-1})$ as shown below:

$$r_{\phi_1}(x_i, x_{i-1}) = \frac{\binom{n}{2} - x_{i-1}}{x_i}. \tag{3}$$

Using Equation (3) along with the specification of $x_0, \dots, x_k = m$ as $x_i = i$ and $|c_{\phi_1}(x_0)| = 1$, it is possible to calculate $|c_{\phi_1}(m)|$. Section 4.1.1 provides a numerical example for calculating $|c_{\phi_1}(m)|$ when $m = 10$, while Section 4.1.2 provides a comparison between the recursive formula and a previously established formula.

3.2. Graph Enumeration Problem: Calculate the Number of Labeled Graphs of Size n with Degree Distribution D

To calculate $|c_{\phi_2}(D)|$, the number of labeled graphs with degree distribution D , using the recursive formula, we need to specify a sequence of degree distributions, $x_0, \dots, x_k = D$. We specify such a sequence by leveraging the Havel–Hakimi algorithm [14,15]. Let d be any degree sequence that is consistent with degree distribution D . The Havel–Hakimi algorithm permits identification of a set of edges, denoted as E , that can be used to construct a graph with degree sequence d . Algorithm 1 provides a procedure to identify E .

Algorithm 1: Degree distribution

Input: Degree distribution x ;
 Part 1: Generate degree sequence;
 $S \leftarrow \emptyset$;
for $j \leftarrow 0$ **to** n **do**
 $S \leftarrow \text{Union}(S, \text{Rep}(j, x_j))$;
 Part 2: Generate edges;
 $E \leftarrow \emptyset$;
for $i \leftarrow 1$ **to** n **do**
 $v \leftarrow$ Index of maximum elt. in S ;
 $l \leftarrow S[v]$;
 $S[v] \leftarrow 0$;
 $v_1, \dots, v_l \leftarrow$ Indices of the l maximum elts. in S ;
 for $j \leftarrow 1$ **to** l **do**
 $S[v_j] \leftarrow S[v_j] - 1$;
 $E \leftarrow \text{Union}(E, (v, v_j))$;
return E ;

Let G_i denote the graph formed with the first i edges from E , i.e., contains edges $\{e_1, \dots, e_i\}$. Let x_i denote the degree distribution associated with graph G_i , i.e., $x_i = D(G_i)$. Theorem 2 states that $x_0, \dots, x_k = D$ satisfies s1 – s3.

Theorem 2. Let E denote the collection of edges outputted from Algorithm 1 with a graphical degree distribution D as input. Let G_i denote the graph formed with the first i edges from E , i.e., contains edges $\{e_1, \dots, e_i\}$. Let $x_i = \phi_2(G_i)$. Each consecutive pair in the collection of graphs G_0, \dots, G_k satisfies s1 – s3 for ϕ_2 .

Proof. The edges in E are distinct. Therefore, by design, the conditions are satisfied. □

Let (v_i, v_j) be the single edge that differs between G_i and G_{i-1} . Based on results from Goyal et al. [13]:

$$r_{\phi}(x_i, x_{i-1}) = \frac{\beta_1(G_{i-1}) - \alpha_1(G_{i-1})}{\alpha_1(G_i)}, \tag{4}$$

where

$$\alpha_1(G) = E(\text{DMM}_{d_i(G), d_j(G)} | D(G)); \tag{5}$$

$$\beta_1(G) = \begin{cases} D_{d_i(G)}(G)D_{d_j(G)}(G) & \text{if } d_i(G) \neq d_j(G) \\ \binom{D_{d_i(G)}(G)}{2} & \text{else,} \end{cases} \tag{6}$$

and based on Newman [16],

$$E(DMM_{x,y}|D) \approx \frac{x D_x \times y D_y}{.5(\sum_z z D_z)} \times \left(\frac{1}{2}\right)^{I\{x=y\}} \tag{7}$$

Using Equation (4) along with the specification of $x_0, \dots, x_k = D$ as defined above and $|c_{\phi_2}(x_0)| = 1$ (as only the empty graph has degree distribution x_0), it is possible to calculate $|c_{\phi_2}(D)|$.

3.3. Graph Enumeration Problem: Calculate the Number of Labeled Graphs of Size n with Degree Sequence d

The number of graphs with degree sequence d , $|c_{\phi_{2a}}(d)|$, can be computed by dividing the number of labeled graphs with the degree distribution consistent with d , denoted as $D(d)$, by the number of permutations of assigning vertices to degrees. Specifically,

$$|c_{\phi_{2a}}(x)| = \frac{|c_{\phi_2}(D(d))|}{\prod_{j=0}^n \binom{n - \sum_{k=0}^{j-1} D_k(d)}{D_j(d)}} \tag{8}$$

Section 4.2.1 provides a numerical example for calculating $|c_{\phi_{2a}}(d)|$, while Section 4.2.2 provides a comparison between the presented recursive formula and a formula by Liebenau et al. [9].

3.4. Graph Enumeration Problem: Calculate the Number of Labeled Graphs of Size n with Classification Mixing Matrix MM

To calculate the number of labeled graphs with mixing matrix MM , we assume that the classification of all vertices, M , is known and MM is graphical. We specify $x_0, \dots, x_k = MM$ as the following for $l \leq m$ (x_i is symmetric):

$$x_{i,l,m} = \begin{cases} 0 & \text{if } i \leq \sum_{a=1}^{l-1} \sum_{b=a}^q x_{k_{a,b}} + \sum_{b=l}^{m-1} x_{k_{l,b}} \\ x_{k_{l,m}} & \text{else if } i \geq \sum_{a=1}^{l-1} \sum_{b=a}^q x_{k_{a,b}} + \sum_{b=l}^m x_{k_{l,b}} \\ i - \sum_{a=1}^{l-1} \sum_{b=a}^q x_{k_{a,b}} + \sum_{b=l}^{m-1} x_{k_{l,b}} & \text{else,} \end{cases} \tag{9}$$

where q is the number of distinct classifications. Theorem 3 proves that there are graphs consistent with $\{x_0, \dots, x_k\}$ that satisfy $s1 - s3$.

Theorem 3. For a sequence of mixing matrices $\{x_0, \dots, x_k\}$ defined by Equation (9), there exists a collection of graphs G_0, \dots, G_k where each consecutive pair satisfies $s1 - s3$ for ϕ_3 .

Proof. To show this, let $E = \{e_1, \dots, e_k\}$ be a set of distinct edges where the first $x_{k_{1,1}}$ are between vertices with classification 1, the next $x_{k_{1,2}}$ are between vertices with classification 1 and 2, and so on. Let G_i denote the graph formed with the first i edges from E , i.e., G_i contains edges $\{e_1, \dots, e_i\}$. Based on the definition of G_i , $\phi_3(G_i) = x_i$, $\phi_3(G_{i-1}) = x_{i-1}$, and G_i and G_{i-1} differ by a single edge. \square

Let (v_l, v_j) be the single edge that differs between G_i and G_{i-1} . Given Theorem 3, $r_{\phi_3}(x_i, x_{i-1})$ is the following:

If $m_l \neq m_j$

$$r_{\phi_3}(x_i, x_{i-1}) = \frac{M_{m_l} \times M_{m_j} - x_{i-1,l,j}}{x_{i,l,j}} \tag{10}$$

else,

$$r_{\phi_3}(x_i, x_{i-1}) = \frac{\binom{M_{m_i}}{2} - x_{i-1, i, j}}{x_{i, i, j}}, \quad (11)$$

Using Equations (10) and (11) along with the specification of $x_0, \dots, x_k = MM$ as defined above and $|c_{\phi_3}(x_0)| = 1$, it is possible to calculate $|c_{\phi_3}(MM)|$.

3.5. Graph Enumeration Problem: Calculate the Number of Graphs of Size n with Degree Mixing Matrix DMM

To calculate the number of labeled graphs with degree mixing DMM , we follow a similar approach as that for degree distribution. Specifically, we use a constructive proof for assessing whether a degree mixing matrix is graphical to specify a set of edges, E , that can be used to construct a graph with degree mixing DMM [13]; Algorithm 2 provides a procedure to construct E .

Algorithm 2: Degree Mixing

Part 1: Generate degree distribution;

for $j \leftarrow 0$ **to** $n - 1$ **do**

$D_j \leftarrow (\sum_{i=0} DMM_{i,j} + DMM_{j,j}) / j$;

Part 2: Generate degree sequences;

$S_c \leftarrow \text{Rep}(0, n)$;

$S_f \leftarrow \emptyset$;

for $j \leftarrow 0$ **to** n **do**

$S_f \leftarrow \text{Union}(S_f, \text{Rep}(j, D_j))$;

Part 3: Edges between vertices with same final degree;

$E \leftarrow \emptyset$;

for $j \leftarrow 0$ **to** n **do**

$P \leftarrow 0, \dots, 2 * DMM_{j,j} - 1$ modulo D_j ;

$T \leftarrow \text{Table}(P)$;

$S_j \leftarrow \emptyset$;

for $i \leftarrow 0$ **to** n **do**

if $i = j$ **then**

$S_j \leftarrow \text{Union}(S_j, T)$;

else

$S_j \leftarrow \text{Union}(S_j, \text{Rep}(0, D_i))$;

$S_c \leftarrow S_c + S_j$;

$E_j \leftarrow \text{Use degree distribution algorithm with } S_j$;

$E \leftarrow \text{Union}(E, E_j)$;

Part 4: Edges between vertices with different final degrees;

for $j \leftarrow 1$ **to** $n - 1$ **do**

for $i \leftarrow j + 1$ **to** n **do**

$P \leftarrow 0, \dots, DMM_{i,j} - 1$ modulo D_j ;

$T \leftarrow \text{Table}(P)$;

for $k \leftarrow 1$ **to** D_j **do**

$v \leftarrow \text{Index of min. elt. in } S_c \text{ s.t. } S_f = j$;

$w_1, \dots, w_{T_k} \leftarrow \text{Indices of } T_k \text{ min. elts. in } S_c \text{ s.t. } S_f = i$;

for $m \leftarrow 1$ **to** T_k **do**

$S_c[v] \leftarrow S_c[v] + 1$;

$S_c[w_m] \leftarrow S_c[w_m] + 1$;

$E \leftarrow \text{Union}(E, (v, w_m))$;

return E ;

Theorem 4. Let E denote the collection of edges output from Algorithm 2 with a graphical degree mixing matrix DMM as input. Let G_i denote the graph formed with the first i edges from E , i.e., contains edges $\{e_1, \dots, e_i\}$. Let $x_i = \phi_4(G_i)$. Each consecutive pair in the collection of graphs G_0, \dots, G_x satisfies $s1 - s3$ for ϕ_4 .

Proof. The edges in E are distinct. Therefore, by design, the conditions are satisfied. \square

Based on the definition of x_0 in Theorem 4, $|c_{\phi_4}(x_0)| = 1$. Let (v_l, v_j) be the single edge that differs between G_i and G_{i-1} . Based on results from Goyal et al. [13]:

$$r_{\phi_4}(x_i, x_{i-1}) = \frac{[\gamma_1(G_{i-1}) - \alpha_2(G_{i-1})] \times \beta_{(l,j)}^0(G_{i-1})}{DMM_{d_l(G_i), d_j(G_i)}(G_i) \times \beta_{(l,j)}^1(G_i)}, \tag{12}$$

where

$$\alpha_2(G) = DMM_{d_l(G), d_j(G)}(G); \tag{13}$$

$$\gamma_1(G) = \begin{cases} D_{d_i(G)}(G)D_{d_j(G)}(G) & \text{if } d_i(G) \neq d_j(G) \\ \binom{D_{d_i(G)}(G)}{2} & \text{else;} \end{cases} \tag{14}$$

and based on concepts from Newman [16], if $d_i(z) \neq d_j(z)$,

$$\beta_{(l,j)}^s(G) \approx \frac{\prod_z \binom{DMM'_{d_l(G),z}(G) - I_{\{d_j(G)=z\}} \cdot s}{n_l^z - I_{\{d_j(G)=z\}} \cdot s}}{\binom{\sum_z DMM'_{d_l(G),z}(G) - I_{\{d_j(G)=z\}} \cdot s}{d_l(G) - s}} \times \frac{\prod_z \binom{DMM'_{z,d_j(G)}(G) - I_{\{d_l(G)=z\}} \cdot s}{n_j^z - I_{\{d_l(G)=z\}} \cdot s}}{\binom{\sum_z DMM'_{z,d_j(G)}(G) - I_{\{d_l(G)=z\}} \cdot s}{d_j(G) - s}} \tag{15}$$

else,

$$\beta_{(l,j)}^s(G) \approx \frac{\prod_z \binom{DMM'_{d_l(G),z}(G) - I_{\{d_j(G)=z\}} \cdot s}{n_l^z + n_j^z - 2I_{\{d_j(G)=z\}} \cdot s}}{\binom{\sum_z DMM'_{d_l(G),z}(G) - I_{\{d_j(G)=z\}} \cdot s}{d_l(G) + d_j(G) - 2s}}. \tag{16}$$

where $DMM'_{a,b}(G) = DMM_{a,b}(G)$ if $a \neq b$ and $DMM'_{a,b}(G) = 2 * DMM_{a,b}(G)$ if $a = b$ and n_l^z and n_j^z denote the number of vertices that are neighbors of l and j and equal to z .

Using Equation (12) along with the specification of $x_0, \dots, x_k = DMM$ as defined above and $|c_{\phi_4}(x_0)| = 1$, it is possible to calculate $|c_{\phi_4}(DMM)|$.

3.6. Additional Graph Properties and Bipartite Graphs

The recursive formula and associated framework we propose can be used to calculate the number of labeled graphs for many additional graph properties. In particular, Goyal et al. [13] provide equations for $r_{\phi}(x_k, x_{k-1})$ for number of triangles (controlling for degree mixing) as well as jointly specifying classification mixing matrix and degree distribution. In addition, Goyal et al. [17] enables extending the calculation of $r_{\phi}(x_k, x_{k-1})$ to the setting of bipartite graphs.

4. Results

In this section, we present numerical examples—including validation results for estimating the number of labeled graphs—for two graph properties: number of edges (Section 4.1) and degree sequence (Section 4.2). To the author’s knowledge, our paper is the first to provide formulas for estimating the number of labeled graphs consistent with values for several graph properties (e.g., degree mixing matrix) described in Section 3; hence, for these properties, we are not able to compare our approach to any existing approach. In

the application section (Section 5), we present results for the number of labeled graphs associated with particular degree distributions and degree mixing matrices using our presented approach.

4.1. Number of Edges

Although there is a closed-form expression for calculating the number of labeled graphs of size n with m number of edges, we use Equation (1) for calculation of this graph property to illustrate our recursive approach for graph enumeration.

4.1.1. Example

To illustrate the use of the recursive formula, we provide a numerical example where we estimate the number of graphs of size $n = 1000$ with exactly $m = 10$ edges; that is, we calculate $|c_{\phi_1}(m = 10)|$. As discussed in Section 3.1, we set $x_0 = 0, x_1 = 1, \dots, x_k = 10$. Therefore, based on Equation (1):

$$|c_{\phi_1}(x_k = 10)| = r_{\phi_1}(x_k = 10, x_{k-1} = 9) \cdot |c_{\phi_1}(x_{k-1} = 9)|. \tag{17}$$

Based on Equation (3),

$$r_{\phi_1}(x_k = 10, x_{k-1} = 9) = \frac{\binom{1000}{2} - 9}{10} = 49949.1. \tag{18}$$

Therefore,

$$|c_{\phi_1}(x_k = 10)| = 49949.1 \cdot |c_{\phi_1}(x_{k-1} = 9)|. \tag{19}$$

The next step in the procedure is the calculation of $|c_{\phi_1}(x_{k-1} = 9)|$ using the following:

$$|c_{\phi_1}(x_{k-1} = 9)| = r_{\phi_1}(x_{k-1} = 9, x_{k-2} = 8) \cdot |c_{\phi_1}(x_{k-2} = 8)|. \tag{20}$$

The procedure ends at $x_0 = 0$. Therefore,

$$\log(|c_{\phi_1}(x_k = 10)|) = \sum_{i=1}^{i=10} \log(r_{\phi_1}(x_i, x_{i-1})) + \log(|c_{\phi_1}(x_0 = 0)|). \tag{21}$$

Table 1 provides the log values for $r(x_i, x_{i-1})$ for $i = 1$ to 10. Based on these values and noting that $|c_{\phi_1}(x_0 = 0)| = 1$, we calculate $\log(|c_{\phi_1}(x_k = 10)|) = 116.11$.

Table 1. Comparison of methods to calculate number of edges.

x	$\log(r(x, x - 1))$	$\log(c_{\phi_1}(x))$ [Equation (1)]	$\log(c_{\phi_1}(x))$ [Equation (22)]
0	–	0	0
1	13.12	13.12	13.12
2	12.42	25.55	25.55
3	12.02	37.57	37.57
4	11.74	49.31	49.31
5	11.51	60.82	60.82
6	11.33	72.15	72.15
7	11.18	83.32	83.32
8	11.04	94.37	94.37
9	10.92	105.29	105.29
10	10.82	116.11	116.11

4.1.2. Comparison

To illustrate the validity of the recursive formula in this setting, we compare the estimates of the number of graphs of size n with m edges, $|c_{\phi_1}(m = x)|$, based on the proposed recursive formula to those based on the following existing formula [1]:

$$|c_{\phi_1}(x)| = \binom{\binom{n}{2}}{x}. \tag{22}$$

Our comparison considers values of x ranging from $\{1, \dots, 10\}$ for graphs of size $n = 1000$.

For each value of $x \in \{1, \dots, 10\}$, Table 1 provides the log values for $r(x, x - 1)$ and $|c_{\phi_1}(x)|$ based on Equations (1) and (22). The estimates obtained from the proposed recursive formula and the known formula are identical—an expected finding given that a closed-form equation for $r_{\phi_1}(x, x - 1)$ exists.

Regarding complexity, computing the recursive formula for $|c_{\phi_1}(x)|$ requires computing $\binom{n}{2}$ as well as x number of subtractions and divisions. Therefore, the complexity of the recursive formula is $O(x)$ where subtractions and divisions are $O(1)$.

4.2. Degree Sequence

This section illustrates the use of Equation (1) to estimate the number of graphs with a fixed degree sequence.

4.2.1. Example

To illustrate the use of the recursive formula, we provide a numerical example wherein we estimate the number of graphs with $n = 1000$ and fixed degree sequence. In particular, we estimate the number of 2-regular graphs of size $n = 1000$; a δ -regular graph is one in which each vertex has exactly degree δ . Therefore, the degree sequence for a δ -regular graph is $d = \{\delta, \dots, \delta\}$.

As the number of distinct degree sequences for a degree distribution associated with a δ -regular graph is one, Equation (8) simplifies to:

$$|c_{\phi_{2a}}(d)| = |c_{\phi_2}(D(d))|. \tag{23}$$

Using Algorithm 1, we generate a sequence of edges E , where E specifies a 2-regular graph of size n . The edge list E contains 1000 edges. A partial list of the edges in E , denoted as a pair consisting of Vertex 1 and Vertex 2 $\in \{1, \dots, 1000\}$, are shown in Table 2. Using the edge list E , we generate a series of graphs, G_0, \dots, G_{1000} , where G_i is the graph that contains the first i edges in E . Next, we generate a series of degree distributions $\{x_0, \dots, x_{1000}\}$, where $x_i = \phi_2(G_i)$. Table 3 shows a partial list of the degree distribution sequences, i.e., x_i . For each consecutive entry in the degree distribution sequences, we can use Equations (4)–(7) to estimate $r_{\phi}(x_i, x_{i-1})$. As x_0 is only consistent with the empty graph, $|c_{\phi_2}(x_0)| = 1$. We estimate the log number of 2-regular graphs as 5907.8999; as we see in the next section an existing formula by Liebenau et al. [9] estimates this as 5907.352.

Table 2. A partial list of the edges in E , denoted as a pair consisting of Vertex 1 and Vertex 2.

Vertex 1	Vertex 2
v_1	v_2
v_1	v_3
v_4	v_5
v_4	v_6
...	...
v_{14}	v_{15}
v_{11}	v_{12}
v_8	v_9
v_5	v_6
v_2	v_3

Table 3. A partial list of the degree distribution sequences.

Degree 0	Degree 1	Degree 2	Degree 3	
x_0	1000	0	0	0
x_1	998	2	0	0
x_2	997	2	1	0
x_3	995	4	1	0
x_4	994	4	2	0
...
x_{996}	0	8	992	0
x_{997}	0	6	994	0
x_{998}	0	4	996	0
x_{999}	0	2	998	0
x_{1000}	0	0	1000	0

4.2.2. Comparison

As in the previous section, we illustrate the validity of the recursive formula in this setting by comparing the estimates of $|c_{\phi_{2a}}(d)|$ based on the proposed recursive formula to those that result from an existing formula. In particular, we compare results of the proposed recursive approach to those resulting from the available formula for estimating the number of δ -regular graphs of size $n = 1000$.

Liebenau et al. [9] proved the validity of a general asymptotic formula—conjectured in 1990—for the number of graphs with given degree sequence. They also provide a formula that converges to the number of δ -regular graphs as $n \rightarrow \infty$. This formula allows comparison of $|c_{\phi_{2a}}(d)|$, where d is the degree sequence $d = \{\delta, \dots, \delta\}$, obtained from this asymptotic formula to those from the proposed recursive formula.

Figure 1 shows log estimates for the number of δ -regular graphs for δ from 1 to 10 for the two approaches. The red bars depict estimates based on the recursive formula introduced in this paper; the blue bars are estimates based on Liebenau et al. [9]. Each plot in Figure 1 shows log estimates for graphs of size 1000, 5000, and 10,000. The log estimates differ by less than 0.01%. To calculate the number of graphs with a given degree sequence based on the recursive formula, we first make use of the Havel–Hakimi algorithm, which has a complexity of $O(n^2)$ [18]. Next, the ratio presented in Equation (4) must be computed $(\sum_{i=1}^n d)/2$ times, where each calculation of the ratio is $O(1)$. Finally, we calculate the number of degree sequences that are associated with a degree distribution, which has complexity $O(n)$. Therefore, the complexity of the recursive formula for calculating the number of graphs of size n with a degree sequence d is $O(n^2) + O((\sum_{i=1}^n d)/2)$.

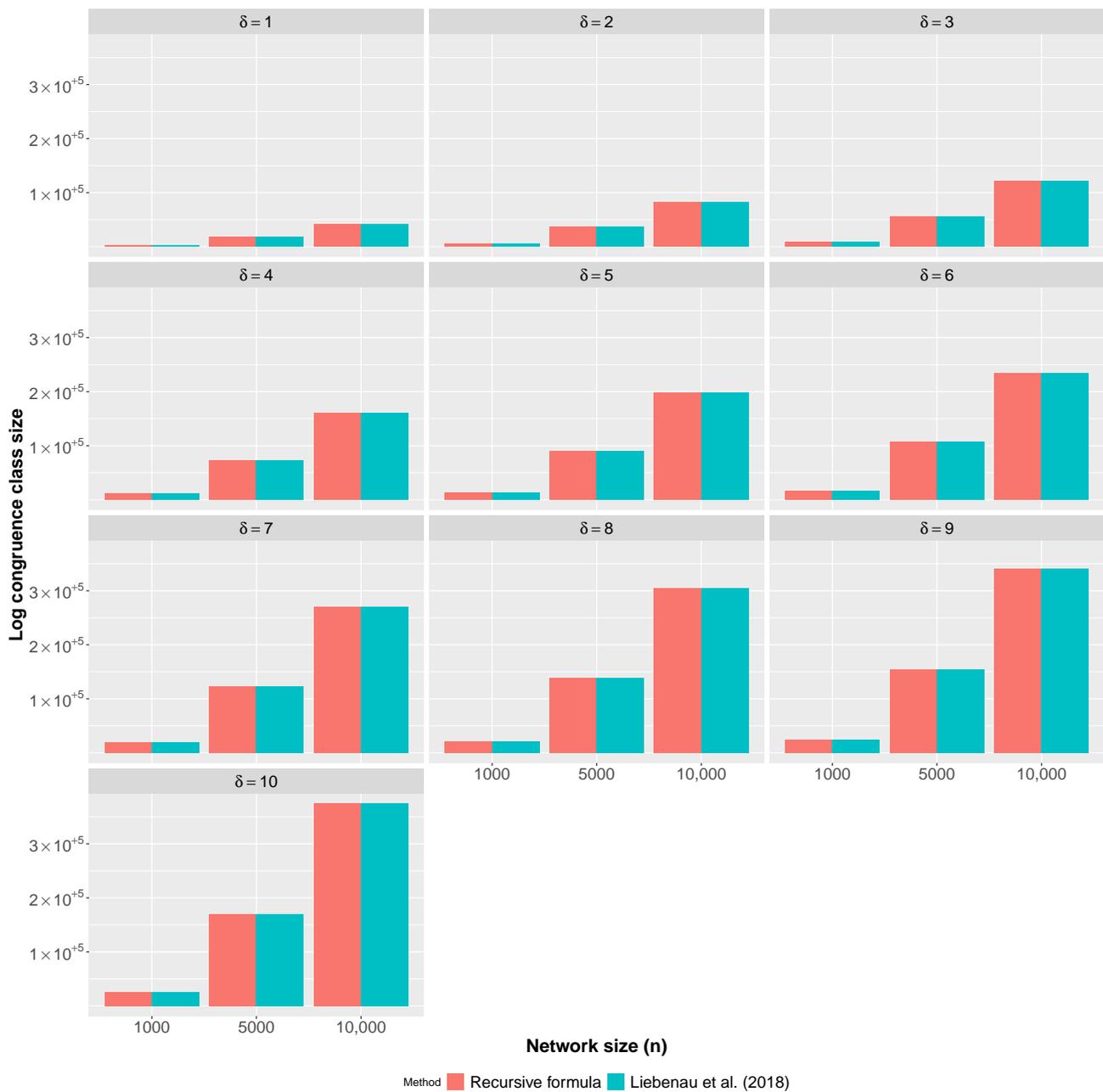


Figure 1. Comparison: Log estimates the number of δ -regular graphs for degrees from 1 to 10. The red bars depict estimates based on the recursive formula introduced in this paper; the blue bars are estimates based on Liebenau et al. [9]. Each plot shows log estimates for graphs of size 1000, 5000, and 10,000.

5. Application

In this section, we estimate the variation in degree mixing matrices consistent with degree distributions formed by the Barabási–Albert (BA) model [10]. The BA model can be initiated with a small seed graph that grows by the addition of new vertices one at a time. Each new vertex forms a new edge with an existing vertex based on preferential attachment rules. Vertices and edges, once introduced, are never deleted. The BA model fixes the number of (undirected) edges connected to each new vertex. The BA model provides a mechanism to generate graphs with a fat-tailed degree distribution—specifically a power-law degree distribution—wherein the probability, $P(k)$, that a vertex in the graph has degree k , decays as a power-law $P(k) \sim k^{-\gamma}$.

To calculate the variation in degree mixing matrices consistent with degree distributions formed by the BA model, we first estimate the number graphs consistent with a degree distribution associated with the BA. Second, we estimate the number of graphs consistent with a degree mixing matrix associated with a degree distribution from the BA model. Third, we estimate the number of distinct degree mixing matrices associated with a degree distribution generated from the BA model, which provides a metric for the variation of graphs generated by the BA model.

For the first step, we generate 100 graphs using the BA model ($n = 5000$), denoted as $\{G_1^{BA}, \dots, G_{100}^{BA}\}$. Figure 2 shows density plots for the log estimates for $|c_{\phi_2}(\phi_2(G_i^{BA}))|$ in the first panel. The average number of labeled graphs associated with a degree distribution generated from the BA model was estimated as $1.26e16988$ (exponential of the mean of the first panel in Figure 2). Second, for each graph, we estimate $|c_{\phi_4}(\phi_4(G_i^{BA}))|$. Figure 2 (second panel) shows density plots for the log estimates for $|c_{\phi_4}(\phi_4(G_i^{BA}))|$. Figure 2 (third panel) shows a density plot for the log estimates for the number of distinct degree mixing matrices associated with a degree distribution generated from the BA model. The exponential of the mean gives an estimate of $4.16e634$ distinct degree mixing matrices associated with a degree distribution generated from the BA model.

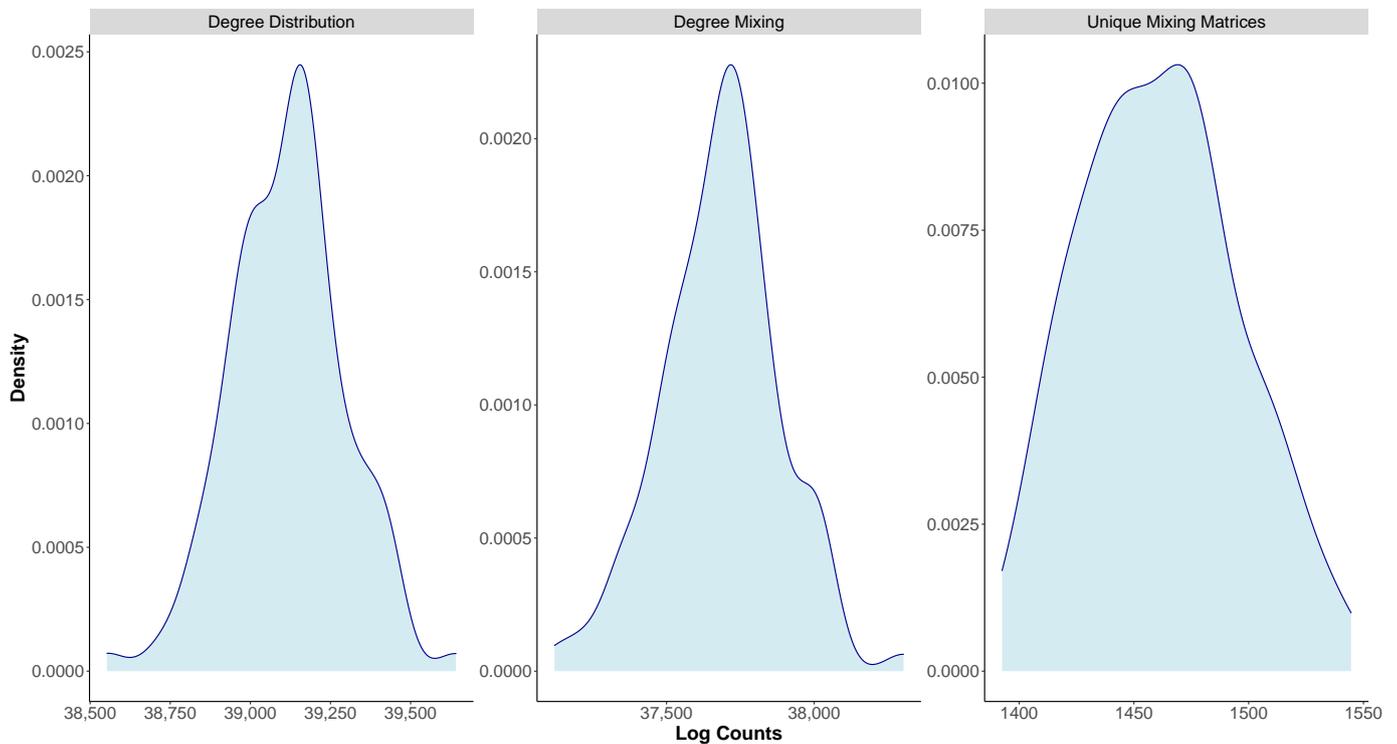


Figure 2. Barabási–Albert (BA) model: Density plots for the log estimates for $|c_{\phi_2}(\phi_2(G_i^{BA}))|$ (first panel) [degree distribution], $|c_{\phi_4}(\phi_4(G_i^{BA}))|$ (second panel) [degree mixing], and number of distinct degree mixing matrices associated with a degree distribution generated from the BA model.

6. Discussion

This paper presents a general recursive formula to estimate the number of labeled graphs with specific values for graph properties of interest. We consider those with particular relevance for social network analysis: number of edges (graph density), degree sequence, degree distribution, classification mixing, and degree mixing. The proposed method can easily be extended to additional graph properties, including number of triangles (controlling for degree mixing), as well as to bipartite graphs; the formulas for Equation (2) are currently available. The proposed recursive formula differs from other available approaches for graph enumeration both in its overall approach and in the breadth

of graph properties that can be considered; it may be profitable to investigate the theoretical connections between the proposed method and other approaches. Furthermore, graph enumeration has the potential to play an important role in statistical network analysis, because formulating the likelihood of observing a real-world graph with particular properties is necessary for making principled inferences.

One current area of research addresses the question of how to make use of results obtained from a study in one population setting to predict what results of a similar study would have been in a different setting. Causal methodologists refer to such research as the study of transportability. This notion is related to the idea of generalizability of results to populations different from the one under study, but true generalizability requires that two populations be similar in all factors that impact study results in important ways. For example, if characteristics such as age or sex of recipients of interventions impacted their efficacy, then generalizability would require that the two populations be similar in these characteristics. Transportability analyses attempt to adjust for differences in populations in prediction of quantities such as intervention effects in new populations. In the settings we consider, adjustment would be required not only for individual characteristics, but also potentially for graph features that impact intervention effects. For example, if degree assortativity—a summary measure of the degree mixing matrix—impacts the spread of disease or the effectiveness of interventions, then this factor would need to be taken into account when predicting spread or effectiveness in a population different from the one that was studied. The methods we describe would aid in investigation of transportability in such settings by facilitating development of ABM-based simulation studies in which graph properties can be chosen to reflect knowledge about those properties (including their uncertainty) in the setting of interest. Of course, in many settings detailed information about potentially important properties may be unavailable. This issue can be addressed using the methods described above to assess the extent to which the unknown properties might impact intervention effectiveness (or other quantities of interest) in the new population [19]. Ideally it would be safe to exclude these properties from the graph model; but if not, an investigator could use the proposed methods to consider plausible ranges of these properties. Doing so would appropriately increase the uncertainty of the prediction of intervention effectiveness in the new population.

Author Contributions: Conceptualization, R.G. and V.D.G.; methodology, R.G. and V.D.G.; software, R.G.; validation, R.G.; writing—original draft preparation, R.G.; writing—review and editing, V.D.G.; visualization, R.G.; funding acquisition, R.G. and V.D.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institute of Health grants R37 AI-51164 and R01 AI-147441.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the paper; or in the decision to publish the results.

References

1. Harary, F.; Palmer, E.M. *Graphical Enumeration*; Elsevier: Amsterdam, The Netherlands, 2014.
2. Iñiguez, G.; Battiston, F.; Karsai, M. Bridging the gap between graphs and networks. *Commun. Phys.* **2020**, *3*, 1–5. [[CrossRef](#)]
3. Goyal, R.; Hotchkiss, J.; Schooley, R.T.; De Gruttola, V.; Martin, N.K. Evaluation of SARS-CoV-2 transmission mitigation strategies on a university campus using an agent-based network model. *Clin. Infect. Dis.* **2021**, *73*, 1735–1741. [[CrossRef](#)]
4. Hambridge, H.L.; Kahn, R.; Onnela, J.P. Examining sars-cov-2 interventions in residential colleges using an empirical network. *Int. J. Infect. Dis.* **2021**, *113*, 325–330. [[CrossRef](#)]
5. Lucatero, C.R. Combinatorial Enumeration of Graphs. In *Probability, Combinatorics and Control*; IntechOpen: London, UK, 2019.

6. Read, R. The enumeration of locally restricted graphs (I). *J. Lond. Math. Soc.* **1959**, *1*, 417–436. [[CrossRef](#)]
7. Bender, E.A.; Canfield, E.R. The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory Ser. A* **1978**, *24*, 296–307. [[CrossRef](#)]
8. Bollobás, B. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.* **1980**, *1*, 311–316. [[CrossRef](#)]
9. Liebenau, A.; Wormald, N. Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph. *arXiv* **2017**, arXiv:1702.08373.
10. Barabasi, A.L.; Albert, R. Emergence of Scaling in Random Networks. *Science* **1999**, *286*, 509–512. [[CrossRef](#)]
11. Petrovic, S. A survey of discrete methods in (algebraic) statistics for networks. *Algebr. Geom. Methods Discret. Math.* **2017**, *685*, 260–281.
12. Shalizi, C.R.; Rinaldo, A. Consistency under sampling of exponential random graph models. *Ann. Stat.* **2013**, *41*, 508–535. [[CrossRef](#)]
13. Goyal, R.; Blitzstein, J.; De Gruttola, V. Sampling networks from their posterior predictive distribution. *Netw. Sci.* **2014**, *2*, 107–131. [[CrossRef](#)] [[PubMed](#)]
14. Hakimi, S.L. On realizability of a set of integers as degrees of the vertices of a linear graph. *J. SIAM* **1962**, *10*, 496–506. [[CrossRef](#)]
15. Havel, V. A remark on the existence of Finite graphs. *Časopis Pest. Mat.* **1955**, *80*, 477–480. [[CrossRef](#)]
16. Newman, M. Assortative mixing in networks. *Phys. Rev. Lett.* **2002**, *89*, 208701. [[CrossRef](#)] [[PubMed](#)]
17. Goyal, R.; De Gruttola, V. Inference on network statistics by restricting to the network space: applications to sexual history data. *Stat. Med.* **2018**, *37*, 218–235. [[CrossRef](#)] [[PubMed](#)]
18. Sensarma, D.; Sen Sarma, S. Role of graphic integer sequence in the determination of graph integrity. *Mathematics* **2019**, *7*, 261. [[CrossRef](#)]
19. DeGruttola, V.; Goyal, R.; Martin, N.K.; Wang, R. Network methods and design of randomized trials: Application to investigation of COVID-19 vaccination boosters. *Clin. Trials* **2022**, *19*, 363–374. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.