On the Propagation of Low-Rate Measurement Error to Subgraph Counts in Large, Sparse Networks

Prakash Balachandran, Eric D. Kolaczyk, and Weston Viles

Department of Mathematics and Statistics, Boston University

Abstract: Our work in this paper is motivated by an elementary but also fundamental and highly practical observation – that uncertainty in constructing a network graph \( \hat{G} \), as an approximation (or estimate) of some true graph \( G \), manifests as errors in the status of (non)edges that must necessarily propagate to any summaries \( \eta(G) \) we seek. Mimicking the common practice of using plug-in estimates \( \eta(\hat{G}) \) as proxies for \( \eta(G) \), our goal is to characterize the distribution of the discrepancy \( D = \eta(\hat{G}) - \eta(G) \), in the specific case where \( \eta(\cdot) \) is a subgraph count. In the empirically relevant setting of large, sparse graphs with low-rate measurement errors, we demonstrate under an independent and unbiased error model and for the specific case of counting edges that a Poisson-like regime maintains. Specifically, we show that the appropriate limiting distribution is a Skellam distribution, rather than a normal distribution. Next, because dependent errors typically can be expected when counting subgraphs in practice, either at the level of the edges themselves or due to overlap among subgraphs, we develop a parallel formalism for using the Skellam distribution in such cases. In particular, using Stein’s method, we present a series of results leading to the quantification of the accuracy with which the difference of two sums of dependent Bernoulli random variables may be approximated by a Skellam. This formulation is general and likely of some independent interest. We then illustrate the use of these results in our original context of subgraph counts, where we examine (i) the case of counting edges, under a simple dependent error model, and (ii) the case of counting chains of length 2 under an independent error model. We finish with a discussion of various open problems raised by our work.

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35.

Keywords and phrases: sample, \L A T E X \v .

1. Introduction

The analysis of network data is widespread across the scientific disciplines. Technological and infrastructure, social, biological, and information networks are a few of the major network classes in which such analyses have been employed. However, despite the already substantial body of work in network analysis generally (e.g., see [9, 13, 15]), with contributions from a variety of different fields of study, much work still remains to more fully develop the theory and methods of statistical analysis of network data, particularly for certain classes of problems of a fairly fundamental nature. Here in this paper we pose and address a version of one such fundamental problem, that regarding the propagation of error through the process of network construction and summary.

*This work was supported in part by AFOSR award 12RSL042 and NSF grant CNS-0905565. Correspondence may be addressed to Eric Kolaczyk, 111 Cummington Mall, Boston, MA 02215 USA (email: kolaczyk@bu.edu).
In applied network analysis, a common modus operandi is to (i) gather basic measurements relevant to the interactions among elements in a system of interest, (ii) construct a network-based representation of that system, with nodes serving as elements and links indicating interactions between pairs of elements, and (iii) summarize the structure of the resulting network graph using a variety of numerical and visual tools. Key here is the point that the process of network analysis usually rests upon some collection of measurements of a more basic nature. For example, online social networks (e.g., Facebook) are based on the extraction and merging of lists of ‘friends’ from millions of individual accounts. Similarly, biological networks (e.g., of gene regulatory relationships) are often based on notions of association (e.g., correlation, partial correlation, etc.) among experimental measurements of gene activity levels. Finally, maps of the logical Internet traditionally have been synthesized from the results of surveys in which paths along which information flows are learned through a large set of packet probes (e.g., traceroute).

Importantly, while it is widely recognized that there is measurement error associated with these and other common types of network constructions, most network analyses in practice effectively proceed as if there were in fact no error. There are at least two possible reasons for this current state of affairs: (1) there is comparatively little in the way of formal probabilistic analyses characterizing the propagation of such error and of statistical methods accounting for such propagation, and (2) in many settings (arguably due in large part to (1)), much attention is given at the stages of measurement and network construction to trying to keep the rate of error ‘low’ in declaring the presence and absence of links between nodes.

Here we offer what is, to the best of our knowledge, the first formal and general treatment of the problem of propagation of error, in which we provide a framework in which to characterize the manner in which errors made in assigning links between nodes accumulate in the reporting of functions of the network as a whole. We provide a probabilistic treatment, wherein our goal is to understand the nature of the distribution induced on the graph functions by that of the errors in the graph construction.

More formally, we consider a setting wherein an underlying (undirected) network-graph $G$ possesses a network characteristic $\eta(G)$ of interest. While there are many types of functions $\eta(\cdot)$ used in practice to characterize networks (e.g., centralities, path-based metrics, output from methods of community detection, etc.) we restrict our attention here to the canonical problem of subgraph counting. That is, we are interested in the class of functions $\eta$ of the form

$$
\eta_H(G) = \frac{1}{|\text{Iso}(H)|} \sum_{H' \subseteq K_{n_v}, H' \simeq H, H \subseteq G} 1\{H' \subseteq G\}, \quad (1.1)
$$

where $n_v = |V(G)|$ is the number of vertices in $G$, $K_{n_v}$ is the complete graph on $n_v$ vertices, $H$ is a graph of interest (i.e., copies of which we desire to count), and $H \subseteq G$ indicates that $H$ is a subgraph of $G$ (i.e., $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$). The value $|\text{Iso}(H)|$ is a normalization factor for the number of isomorphisms of $H$.

If $\hat{G}$ is a network-graph resulting from an attempt to construct $G$ from some collection of basic measurements, then the common practice of reporting the analogous characteristics of $\hat{G}$ is equivalent to the use of $\eta(\hat{G})$, i.e. effectively a plug in estimator.
Let the discrepancy between these two quantities be defined as \( D = \eta(\hat{G}) - \eta(G) \).

Our goal is to make precise probabilistic statements about the behavior of \( D \) under certain conditions. In the case where \( \eta \) is defined as a subgraph count, as in (1.1), \( D \) may be expressed as the difference of (i) the number of times the subgraph \( H \) arises somewhere in \( \hat{G} \) but does not in fact exist in same manner in \( G \), and (ii) vice versa. Hence, \( D \) may be understood in this context to be the difference in total number of Type I and Type II errors, respectively.

Intuitively, in the cases where a relatively low rate of error occurs on a large graph \( G \), each of these two sums should have a Poisson-like behavior. This observation suggests that the propagation of low-rate measurement error to subgraph counts should behave, under appropriate conditions, like the difference of two Poisson random variables, i.e., a so-called Skellam distribution [17]. Our contribution in this paper is to provide both finite sample and asymptotic results on the accuracy with which the Skellam distribution may be used in approximating the distribution of \( D \), under the setting where the graph \( G \) is both large and sparse (reflecting two typical and basic characteristics of networks in the field complex network analysis) and the rate of error is low. Our approach is through the use of Stein’s method (e.g., [5]). Specifically, we present a Stein operator for the Skellam probability distribution and, in a manner consistent with the Stein methodology, we derive several bounds on the discrepancy between the distribution of the difference of two arbitrary sums of binary random variables to an appropriately parameterized Skellam distribution. The latter in turn may be used to establish the rate of weak convergence, which we demonstrate through a handful of examples of choice of subgraph \( H \), on the one hand, and distribution of the underlying measurement errors, on the other hand.

As remarked above, there appears to be little in the way of a formal and general treatment of the error propagation problem we consider here. However, there are, of course, several areas in which the probabilistic or statistical treatment of uncertainty enters prominently in network analysis. The closest area to what we present here is the extensive literature on distributions of subgraph counts in random graphs. See [10], for example, for comprehensive coverage. Importantly, there the graph \( G \) is assumed to emerge according to a (classical) random graph and uncertainty typically is large enough that normal limit theorems are the norm (although Poisson limit theorems also have been established). In contrast, in our setting we assume that \( G \) is a fixed, true underlying graph, and then study the implications of observing a ‘noisy’ version \( \hat{G} \) of that graph, under various assumptions on the nature of the noise, which involves two specific types of error (i.e., Type I and II errors), the contributions of which are informed in part by the topology of \( G \) itself. An area related to this work in random graphs is the work in statistical models for networks, such as exponential random graph models (ERGMs). See [14] for a recent treatment. Here, while these models are inherently statistical in nature, the randomness due to generation of the graph \( G \) and due to observation of \( G \) – resulting in what we call \( \hat{G} \) – usually are combined into a single realization from the underlying distribution. And while subgraph counts do play a key role in traditional ERGMs, they typically enter as covariates in these (auto)regressive models. In a somewhat different direction, uncertainty in network construction due to sampling has also been studied in some depth. See, for example, [13, Ch 5] or [2] for
surveys of this area. However, in this setting, the uncertainty arises only from sampling – the subset of vertices and edges obtained through sampling are typically assumed to be observed without error. Finally, we note that there just recently have started to emerge in the statistics literature formal treatments of the same type of graph observation error model that we propose here. There the emphasis is on producing estimators (e.g., [3]) or classifiers (e.g., [16]), for example, rather than on the type of probabilistic treatment we present here.

The organization of this paper is as follows. We provide an initial treatment of our problem in Section 2, wherein we establish necessary notation and assumptions and then focus on the setting with independent edge noise, for the specific case where the subgraph count is the total number of edges in the graph. There we establish formally under a precise sense of ‘large’, ‘sparse’, and ‘low-rate’ noise that the distribution of the subgraph count is closer to a Skellam than a normal distribution. In Section 3 we then provide a general set of results useful for a variety of other versions of this problem. Specifically, we establish a bound for the Kolmogorov-Smirnov distance between the distribution of the difference of two arbitrary sums of binary random variables from a certain Skellam. This work is based on the application of Stein’s method to the Skellam distribution, a first of its kind to the best of our knowledge, and the results therefore are of some independent interest as well. In Section 4 we then illustrate the way in which these general results may be used to understand the propagation of error in networks for subgraph counts. Finally, there is further discussion in Section 5. Proofs of our key results may be found in the appendices.

2. Propagation of Error to Subgraph Counts: an Initial Treatment

2.1. Notation and Assumptions

By $G = (V, E)$ we will mean an undirected graph, with vertex set $V$ of cardinality $|V|$ and edge set $E$ of cardinality $|E|$. Much of the results that follow will be stated as a function of the number of vertices which, for notational convenience, we denote $n_v = |V|$. We assume the vertex set $V$ is known but that the edge set $E$ is unknown. However, we assume there is information by which to construct an approximation to $E$ or, more formally, by which to infer $E$, as a set $\hat{E}$, yielding an inferred network graph $\hat{G} = (V, \hat{E})$.

While there are many ways in practice by which the set $\hat{E}$ is obtained, one principled way of viewing the task is as one of performing \( \binom{n_v}{2} \) hypothesis tests, using the data underlying the graph construction process as input, one for each of the vertex pairs \( \{i, j\} \) in the network graph $G$. In some contexts, $G$ is literally obtained through hypothesis testing procedures; for instance, in constructing some gene regulatory networks from microarray expression data. See [13, Ch 7], for example. Formally, in such cases we can think of $\hat{G}$ resulting from a collection of testing problems

$$H_0 : \{i, j\} \notin E \text{ versus } H_1 : \{i, j\} \in E,$$

for $\{i, j\} \in V^{(2)}$, where

$$V^{(2)} = \{\{i, j\} : i, j \in V; i < j\}.$$
These tests amount to a collection of \( \binom{n_v}{2} \) binary random variables \( \{ Y_{ij} : \{i,j\} \in V^{(2)} \} \), where

\[
Y_{ij} = \begin{cases} 
1 & \text{if } H_0 \text{ is rejected} \\
0 & \text{if } H_0 \text{ is not rejected}.
\end{cases}
\]

Note that the random variables \( Y_{ij} \) need not be independent and, in fact, in many contexts will most likely be dependent. Gene regulatory networks inferred by correlating gene expression values at each vertex \( i \) with that of all other vertices \( j \in V \setminus \{i\} \) and maps of the logical Internet obtained through merging paths learned by sending traffic probes between many sources and destinations are just two examples where dependency can be expected.

Whether obtained informally or formally, we can define the inferred edge set \( \hat{E} \) as

\[
\hat{E} = \left\{ \{i,j\} \in V^{(2)} : Y_{ij} = 1 \right\}.
\]

It is useful to think of the collection of random variables \( \{ Y_{ij} : \{i,j\} \in V^{(2)} \} \) as characterizing two types of errors. That is, we express the marginal distributions of the \( Y_{ij} \) in the form

\[
Y_{ij} \sim \begin{cases} 
\text{Bernoulli}(\alpha_{ij}), & \text{if } \{i,j\} \in E^c, \\
\text{Bernoulli}(1 - \beta_{ij}), & \text{if } \{i,j\} \in E,
\end{cases}
\]

where \( E^c = V^{(2)} \setminus E \). Again pursuing the example of network construction based on hypothesis testing, \( \alpha_{ij} \) can be interpreted as the probability of Type-I error for the test on vertex pair \( \{i,j\} \in E^c \), while \( \beta_{ij} \) is interpreted as the probability of Type-II error for the test on vertex pair \( \{i,j\} \in E \).

Our interest in this paper is in characterizing the manner in which the uncertainty in the \( Y_{ij} \) propagates to subgraph counts on \( \hat{G} \). More specifically, we seek to characterize the distribution of the difference

\[
D = \frac{1}{|Iso(H)|} \sum_{H' \subseteq K_{n_v}, \text{Iso}(H') = H} \left[ 1\{H' \subseteq \hat{G}\} - 1\{H' \subseteq G\} \right],
\]

for a given choice of subgraph \( H \). Naturally, this distribution will depend in no small part on context. Here we focus on a general formulation of the problem in which we make the following four assumptions.

(A1) **Large Graphs:** \( n_v \to \infty \).

(A2) **Sparse Graphs:** \( |E| = \Theta(n_v \log n_v) \).

(A3) **Edge Unbiasedness:** \( \sum_{\{i,j\} \in E^c} \alpha_{ij} = \sum_{\{i,j\} \in E} \beta_{ij} (\equiv \lambda) \).

(A4) **Low Error Rate:** \( \lambda = \Theta(n_v^\gamma \log^\kappa n_v), \) for \( \gamma \in [0,1) \) and \( \kappa \geq 0 \).

Assumption (A1) reflects both the fact that the study of large graphs is a hallmark of modern applied work in complex networks and, accordingly, our desire to make statements that are asymptotic in \( n_v \).

Similarly, one of the most basic characteristics of such large networks in practice is that they tend to be sparse, in that the number of edges scales slower than quadratic
in \( n_v \). Assumption (A2), therefore, is that this number instead scales just slightly more quickly than linear.

Our use of assumption (A3) reflects the understanding that a ‘good’ approximation \( \hat{G} \) to the graph \( G \) should at the very least have roughly the right number of edges. The difference of the two sums defined in (A3) is in fact the expectation of the statistic \( D \) in (2.2) for the case where the subgraph being counted is just a single edge, i.e., it is the expected discrepancy between the number of observed edges \( |\hat{E}| \) and the actual number of edges \( |E| \). So (A3) states that this particular choice of \( D \) have expectation zero. Alternatively, (A3) may be interpreted as saying that the total numbers of Type I and Type II errors should be equal to a common value \( \lambda \).

Finally, in (A4) we encode the notion of there being a ‘low’ rate of error in \( \hat{G} \). Since the network is sparse, \( |E| \) is small compared to \( n_v^2 \). And since the total Type I and Type II error rates are equal, the requirement that \( \lambda \) be small compared to \( |E| \) means it should scale sublinear in \( n_v \). Our parameterization of \( \lambda \) in (A4) includes the cases of constant error rate (i.e., \( \gamma = \kappa = 0 \)), logarithmic error (i.e., \( \gamma = 0, \kappa = 1 \)), and error with root powers (e.g., \( \gamma = 1/3, \kappa = 0 \)).

Lastly, we recall the definition of the Skellam distribution. A random variable \( W \) defined on the integers is said to have a Skellam distribution, with parameters \( \lambda_1, \lambda_2 > 0 \), i.e., \( W \sim \text{Skellam}(\lambda_1, \lambda_2) \), if

\[
P(W = k) = e^{-(\lambda_1 + \lambda_2)} \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^k I_k \left( 2\sqrt{\frac{\lambda_1 \lambda_2}{}} \right) \quad \text{for } k \in \mathbb{Z},
\]  

(2.3)

where \( I_k \left( 2\sqrt{\frac{\lambda_1 \lambda_2}{}} \right) \) is the modified Bessel function of the first kind with index \( k \) and argument \( 2\sqrt{\lambda_1 \lambda_2} \). The Skellam distribution may be constructed by defining \( W \) through the difference of two independent Poisson random variables, with means \( \lambda_1 \) and \( \lambda_2 \), respectively. The mean and variance of this distribution are given by \( \mathbb{E}[W] = \lambda_1 - \lambda_2 \) and \( \text{Var}(W) = \lambda_1 + \lambda_2 \). The distribution of \( W \) is in general nonsymmetric, with symmetry holding if and only if \( \lambda_1 = \lambda_2 \).

The main results we provide in this paper are in the form of bounds on the extent to which the distribution of the discrepancy \( D \) in (2.2) may be well-approximated by an appropriate Skellam distribution. For this purpose, we adopt the Kolmogorov-Smirnov distance to quantify the distance between distributions of two random variables, say, \( X_1 \) and \( X_2 \), i.e.,

\[
ds_{KS}(X_1, X_2) \equiv \sup_{x \in \mathbb{R}} \left| P(X_1 \leq x) - P(X_2 \leq x) \right|.
\]  

(2.4)

2.2. Behavior of Edge Counts Under Independent Edge Noise

We now turn to the main contribution in this section, which is to provide an initial proof-of-concept for the occurrence of the Skellam distribution in characterizing the behavior of \( D \). We let our choice of subgraph \( H \) be simply a single edge, in which case the statistic \( \eta(G) \) in (1.1) is just the total number of edges in \( G \), i.e., \( \eta(G) = |E| \). In addition, we also assume for the moment that the variables \( Y_{ij} \) are independent. This allows for the use of existing tools and results in the literature in obtaining Theorem 1.
below, and is particularly important to establishing the lower bound in that theorem. Finally, for convenience, we add to the core assumptions (A1)-(A4) a fifth assumption, upon which we will call periodically throughout the paper when desiring to simplify some of our expressions.

\[ (A5) \quad \text{Homogeneity: } \alpha_{ij} \equiv \alpha \text{ and } \beta_{ij} \equiv \beta, \text{ for } \alpha, \beta \in (0, 1). \]

In other words, we assume that the probability of making a Type I or II error (as the case may be) does not depend upon the specific (non)edge in question.

The discrepancy \( D \) we study in this section then is the value

\[
D_E = |\hat{E}| - |E| = \sum_{\{i,j\} \in E^c} Y_{ij} - \sum_{\{i,j\} \in E} (1 - Y_{ij}).
\]

This random variable has expectation \( E[|D_E|] = \alpha|E^c| - \beta|E| = \lambda - \lambda = 0 \) and variance \( \sigma^2 = \alpha(1 - \alpha)|E^c| + \beta(1 - \beta)|E| \). We have the following result characterizing the behavior of \( D_E \).

**Theorem 1.** Under assumptions (A1)-(A5) and independence among errors in declaring (non)edge status (i.e., among the \( Y_{ij} \)),

\[
d_{KS}(D_E, \text{Skellam}(\lambda, \lambda)) \leq O\left((n_v^1 - \gamma \log^1 n_v)^{-1}\right).
\]

At the same time,

\[
d_{KS}(D_E/\sigma, \mathcal{N}(0, 1)) \leq O\left((n_v^\gamma/2 \log^{\kappa/2} n_v)^{-1}\right)
\]

and, for sufficiently large \( n_v \),

\[
d_{KS}(D_E/\sigma, \mathcal{N}(0, 1)) \geq \Omega\left((n_v \log n_v)^{-1}\right)
\]

where \( \mathcal{N}(0, 1) \) refers to a standard normal random variable.

Proof of this theorem may be found in the appendix, in Section 6.1. To put the statement of the theorem in context, suppose first that the error rate \( \lambda \) is constant in \( n_v \), i.e., \( \gamma = \kappa = 0 \). Then the theorem indicates that in large, sparse networks, with independent and homogeneous low-rate errors, the distribution of the discrepancy \( D_E \) tends to that of a Skellam distribution symmetric and centered on zero, with variance \( 2\lambda \), at a rate slightly better than linear in the number of vertices \( n_v \). At the same time, an approximation by a normal distribution will be no better than the same rate. We summarize in the following corollary.

**Corollary 1.** When \( \gamma = \kappa = 0 \),

\[
d_{KS}(D_E, \text{Skellam}(\lambda, \lambda)) \leq O\left((n_v \log n_v)^{-1}\right),
\]

while

\[
d_{KS}(D_E/\sigma, \mathcal{N}(0, 1)) \geq \Omega\left((n_v \log n_v)^{-1}\right).
\]
More generally, for $\gamma \in [0, 1)$, the upper bound in (2.6) will decay at a faster rate than that in (2.7). We suspect, in fact, that stronger statements can be made and that the Skellam approximation dominates the standard normal approximation for a nontrivial range of moderately growing low-rate error levels. In Figure 1 we show the results of a small simulation study, comparing the two approximations as $n_v$ ranges from 100 to 1000 to 10,000, for error rates $\lambda$ defined to be constant, logarithmic, square root, or linear functions of $n_v$. It is apparent that the behavior of the case $\lambda = \log n_v$ is nearly the same as that of the constant case (where we chose $\lambda = \log 100$), for this range of $n_v$, with the Skellam clearly dominating the standard normal. At the same time, in the case of $\lambda = n_v^{1/2}$, where the standard normal shows improvement over its performance in the constant or logarithmic cases, the Skellam still shows clear superiority (by over an order of magnitude). Only by the time $\lambda = n_v$ do we see the standard normal become a better approximation than the Skellam for these choices of $n_v$. Note that by this stage, $\beta = O(1)$, and so essentially there is no ‘signal’ standing out from the ‘noise’. The Skellam distribution dominates as an approximation when there can be expected to be a clear graph ‘signal’ standing out against the ‘noise’ induced by underlying low-rate measurement errors, which is precisely the context we have attempted to capture in assumptions (A1)-(A4).

![Figure 1](image_url)
3. General Results on Approximation by Skellam

Recall the general form of our statistic of interest \( D \) in (2.2), as the difference of two sums of binary random variables. While in the previous section, for the special case of counting edges under an independent error model, these binary random variables are independent, this need not be the case in general. Several examples were given earlier of contexts in practice where the underlying measurement process informing construction of the graph \( \hat{G} \) can be expected to yield dependent errors. Furthermore, even if the errors for individual (non)edges are independent, for subgraphs \( H \) of larger order than 2 the overlap of candidate subgraphs \( H' \) in the sum defining \( \eta(G) \) in (1.1) can be expected to induce some dependency among the indicator variables in (2.2). We will illustrate with the example of chains of length 2 later, in Section 4. Finally, for such larger subgraphs \( H \) it similarly cannot be expected that necessarily the expectation of the two sums in (2.2) be the same, and hence accurate approximation of the distribution of \( D \) may involve nonsymmetric distributions.

Nevertheless, under appropriate conditions it seems reasonable to expect that an approximation by Skellam be useful. In the previous section, proof of our results quantifying the accuracy of such approximations under the stated conditions were derived, due to the assumption of independence, from manipulation of existing results for approximating sums by Poisson distributions. Without independence, however, it is necessary to approach the problem directly, by explicitly using the Skellam distribution. In this section, we therefore provide the results of such an approach. This is a completely general treatment – devoid of the motivating context of counting subgraphs – and therefore of independent interest. In Section 4 we return to the problem of counting subgraphs under low-rate error and illustrate the use of the results presented here in this section through certain applications.

Our approach is through Stein’s method. This choice is reminiscent, naturally, of the Chen-Stein treatment for Poisson approximations. However, the task is technically more involved at several points, as it involves handling a Stein function that is defined through a second-order difference, rather than the first-order difference encountered in the Poisson problem. Moreover, the kernel of the Skellam distribution includes a modified Bessel function of the first kind, which emerges in ways necessitating a somewhat delicate treatment.

3.1. A Stein Bound for the Skellam Distribution

Let \( U \) be a random variable defined as

\[
U = \sum_{k=1}^{n} L_k - \sum_{k=1}^{m} M_k ,
\]

(3.1)

where \( \{L_k\}_{k=1}^{n} , \{M_k\}_{k=1}^{m} \) is a collection of two sets of indicator random variables with \( \mathbb{E}[L_k] = p_k \) for \( k = 1, \ldots, n \) and \( \mathbb{E}[M_k] = q_k \) for \( k = 1, \ldots, m \). In the case of our subgraph counting problem, \( U = D \), where \( D \) is defined in (2.2), although for the remainder of this current section \( U \) is defined generally.
Recall the definition of a Skellam random variable $W$ in (2.3). We desire a bound

$$d_{KS}(U,W) := \sup_{x \in \mathbb{R}} |\mathbb{P}(U \leq x) - \mathbb{P}(W \leq x)|,$$  \hspace{1cm} (3.2)

quantifying how close the distribution of $U$ is to that of $W$. In pursuing the standard paradigm for Stein’s method, we first determine an operator $\mathcal{A}[f(k)]$ such that

$$\mathbb{E}\mathcal{A}[f(W)] = 0 \quad \text{if and only if} \quad W \sim \text{Skellam}\ (\lambda_1, \lambda_2)$$

for any bounded function $f : \mathbb{Z} \rightarrow \mathbb{R}$. This operator need not be unique, but the theory only requires one. This is accomplished through the following result, the proof of which uses several properties of the modified Bessel function of the first kind, as detailed in the appendix, in Section 6.2.

**Theorem 2.** A Stein operator $\mathcal{A}$ for the Skellam $(\lambda_1, \lambda_2)$ distribution is

$$\mathcal{A}[f(k)] = \lambda_1 f(k+1) - kf(k) - \lambda_2 f(k-1).$$

With this operator in hand, and again following the usual paradigm under Stein’s method, we set

$$\mathcal{A}[f(k)] = g(k)$$

for a class of test functions $g(k)$, and allow that to implicitly define the function $f$. The choice of the test functions $g$ is guided by the choice of the metric used to measure the distance between $U$ and $W$. Since the metric we choose to measure the distance between $U$ and $W$ is given by $d_{KS}(U,W)$ in (3.2), we choose the test function $g := g_x$ given by

$$g_x(k) = 1 \{k \leq x\} - \mathbb{P}(W \leq x)$$ \hspace{1cm} (3.3)

for any $x \in \mathbb{R}$.

At this point it is common to exhibit a solution $f$ defined by our choice of $g$. Instead, we forestall that step until later in this section, choosing rather to state a general result that will allow us to more quickly gain insight into the nature of the bounds we are able to obtain. Our result employs a minor variant of the notion of coupling that is common to the literature on Chen-Stein approximations.

**Theorem 3.** Let $U$ be as in (3.1) and let $\mathcal{L}(U)$ denote the law of $U$. Let

$$\mathcal{L}\left(U_{k}^{(L)} + 1\right) = \mathcal{L}(U|L_k = 1) \quad \text{for} \quad k = 1, \ldots, n$$

and

$$\mathcal{L}\left(U_{k}^{(M)} - 1\right) = \mathcal{L}(U|M_k = 1), \quad \text{for} \quad k = 1, \ldots, m$$

be a collection of random variables all defined on a common probability space. If $\lambda_1 = \sum_{k=1}^{n} p_k$ and $\lambda_2 = \sum_{k=1}^{m} q_k$, and $W \sim \text{Skellam}(\lambda_1, \lambda_2)$, then

$$d_{KS}(U,W) \leq ||\Delta f|| \left( \sum_{k=1}^{n} p_k \mathbb{E}\left|U - U_{k}^{(L)}\right| + \sum_{k=1}^{m} q_k \mathbb{E}\left|U - U_{k}^{(M)}\right| \right),$$ \hspace{1cm} (3.4)
where

\[ ||\Delta f|| = \sup_{x \in \mathbb{R}} \sup_{j \in \mathbb{Z}} |f_x(j + 1) - f_x(j)| \]

and \( f_x \) is a solution to \( A[f_x(k)] = g_x(k) \) for \( k \in \mathbb{Z} \).

The proof of this result relies on elementary considerations of the equation \( A[f_x(k)] = g_x(k) \) and may be found in the appendix in Section 6.3. The extent to which it allows one to obtain error estimates of practical interest in a particular setting will depend on the extent to which both the main expression within brackets in (3.4) and the quantity \( ||\Delta f|| \) can be further controlled. We address these each in turn.

3.2. Error Estimates Resulting from the Stein Bound

The terms \( \mathbb{E} \left| U - U^{(L)}_k \right| \) and \( \mathbb{E} \left| U - U^{(M)}_k \right| \) in (3.4) measure the dependence of \( U \) on the events \( L_k = 1 \) and \( M_k = 1 \), respectively. As such, the term in brackets in (3.4) plays a central role in controlling \( d_{KS}(U, W) \), the details of which will be dictated by the context available in the application at hand. In Section 4 we will show examples drawing on context imposed by the application to subgraph counting. The following intermediate results will be useful in that regard.

First, note that if \( U^{(L)}_k \) and \( U^{(M)}_k \) are independent of their respective events, then we obtain

\[ d_{KS}(U, W) \leq ||\Delta f|| \left\{ \sum_{k=1}^{n} p_k^2 + \sum_{k=1}^{m} q_k^2 \right\}, \]  

(3.5)

which is analogous to the classical form of the bound for individual sums of independent indicator random variables (e.g., [5]). This bound could also be produced through the use of the same technique we used in proving the bound (2.6) in Theorem 1.

In the absence of independence, it is possible to bound the right-hand side of (3.4) by quantities involving the first two moments of the random variables \( L_k \) and \( M_k \). Specifically, using manipulations standard to coupling approaches in the Poisson problem (e.g., [5, p. 75]), we may argue that

\[ p_k \mathbb{E} \left| U - U^{(L)}_k \right| \leq \sum_{j \neq k} \mathbb{E}[L_j L_k] + p_j p_k + \sum_{\ell} \mathbb{E}[M_{\ell} L_k] + q_{\ell} p_k, \]

and similarly

\[ q_k \mathbb{E} \left| U - U^{(M)}_k \right| \leq \sum_{j} \mathbb{E}[L_j M_k] + p_j q_k + \sum_{\ell \neq k} \mathbb{E}[M_{\ell} M_k] + q_{\ell} q_k. \]

And from these it follows that

\[ d_{KS}(U, W) \leq ||\Delta f|| \left\{ \sum_{k=1}^{n} p_k^2 + \sum_{k=1}^{m} q_k^2 \right\} + 2 \left[ \sum_{j < k} \text{Cov}(L_j, L_k) + \sum_{\ell < k} \text{Cov}(M_{\ell}, M_k) + \sum_{j, \ell} \text{Cov}(L_j, M_{\ell}) \right]. \]
Finally, still further refinements are possible with specific knowledge of the behavior of the covariances in (3.6) or, even stronger, of the nature of the dependency among these random variables. For example, as just one illustration, we recall (e.g., [5, p. 78]) that a binary random vector \( (X_1, \ldots, X_n) \) is said to be negatively associated if, whenever \( \phi_1 \) and \( \phi_2 \) are bounded, increasing functions and \( I_1 \) and \( I_2 \) are disjoint subsets of \( \{1, \ldots, n\} \), we have

\[
E [\phi_1 (X_i; i \in I_1) \phi_2 (X_i; i \in I_2)] \leq E [\phi_1 (X_i; i \in I_1)] E [\phi_2 (X_i; i \in I_2)] .
\]

Suppose that the entire collection of variables \( L_k \) and \( M_k \) are negatively associated. Then from (3.6), and using arguments directly analogous to those in the classical Poisson setting (e.g., [5, pp. 75-78]), it follows that

\[
d_{KS}(U, W) \leq ||\Delta f|| [E(T_1 + T_2) - \text{Var}(T_1 + T_2)] ,
\]

where \( T_1 = \sum_{k=1}^{n} L_k \) and \( T_2 = \sum_{k=1}^{m} M_k \). We will make use of this result in Section 4.1.

### 3.3. Controlling the term \( ||\Delta f|| \).

Now consider the quantity \( ||\Delta f|| \) in (3.4). To control this quantity first requires understanding the solution \( f_x(k) \). We provide a family of closed-form expressions for this solution in the following.

**Theorem 4.** Let \( g_k \) be defined as in equation (3.3). If \( f_x \) is a bounded solution to the difference equation

\[
\lambda_1 f_x (k + 1) - k f_x (k) - \lambda_2 f_x (k - 1) = g_x (k)
\]

for \( k \in \mathbb{Z} \), then \( f_x \) is given by

\[
f_x (m) = \begin{cases} 
(-1)^m \left( \frac{\lambda_1 + \lambda_2}{2} \right)^m I_m \left( -1 \right) \left( \frac{\sqrt{\lambda_1 + \lambda_2}}{2} \right) \frac{1}{\lambda} f (c) \\
\frac{\lambda_1 + \lambda_2}{\sqrt{\lambda_1 + \lambda_2}} \sum_{n=c}^{m} \left( -1 \right)^n \left( \frac{\sqrt{\lambda_1 + \lambda_2}}{2} \right) \frac{1}{\lambda} f (c) \\
\frac{\lambda_1 + \lambda_2}{\sqrt{\lambda_1 + \lambda_2}} \sum_{n=m+1}^{\infty} \left( -1 \right)^n \left( \frac{\sqrt{\lambda_1 + \lambda_2}}{2} \right) \frac{1}{\lambda} f (c) \\
\end{cases}
\]

for any initial condition \( (c, f_x (c)) \) with \( c \in \mathbb{Z} \) and \( f_x (c) \in \mathbb{R} \).

The proof of this theorem is similar to that of solving a second order linear differential equation. An integrating factor is found, integration is performed with a boundary condition at \( -\infty \), and then a second integration is performed with the initial condition. Details are provided in the appendix in Section 6.4.
Leveraging our insight into $f_x$ to control $||\Delta f||$ means producing a bound on the absolute differences $|\Delta f_x(j)| = |f_x(j + 1) - f_x(j)|$ independent of $x \in \mathbb{R}$ and $j \in \mathbb{Z}$. For arbitrary $\lambda_1, \lambda_2 > 0$, we are unable to produce a general bound. However, for the special case where $\lambda_1$ and $\lambda_2$ are equal, we are able to offer the following result.

**Theorem 5.** Suppose that $\lambda_1 = \lambda_2 = \lambda$, so that $\mathbb{E}[U] = 0$ and $W$ is a Skellam($\lambda, \lambda$) distribution, symmetric about zero. Then

$$||\Delta f|| \leq \frac{156}{2\lambda}.$$  

The proof of this theorem is highly technical in nature, and relies on a concentration inequality for the Skellam($\lambda, \lambda$) distribution [4] with several other technical arguments. A sketch of the proof may be found in the appendix in Section 6.5, while a detailed presentation is available in the Supplementary Materials.

The above result is of immediate relevance to the subgraph counting problem examined in Section 2.2, given assumption (A3), where the subgraph $H$ of interest is a single edge and the count of interest is the total number of edges $|E|$ in the graph $G$. We provide details in the application featured in Section 4.1 below.

Note that the bound in Theorem 5 is essentially the analogue of the classical result for Poisson approximation, in which, for sufficiently large $\lambda$, the term $1/\lambda$ is the standard constant. In both cases, therefore, the corresponding term $||\Delta f||$ is bounded by the inverse of the expected total number of counts, where here that is $\mathbb{E}[T_1 + T_2] = 2\lambda$. Comparing this quantity to, for instance, the expression in brackets on the right-hand side of (3.7), we see in the case of negatively associated indicator variables the accuracy of the Skellam approximation is determined by the relative magnitude of the discrepancy $\mathbb{E}[T_1 + T_2] - \text{Var}(T_1 + T_2)$ to $\mathbb{E}[T_1 + T_2]$. In the case where the random variables $T_1$ and $T_2$ are precisely Poisson distributed and independent, the discrepancy will, of course, be zero. Under dependency, it can be expected to deviate from zero.

For the general case of $\lambda_1 \neq \lambda_2$, we conjecture that $||\Delta f||$ can be bounded by a term of the form $C/(\lambda_1 + \lambda_2)$, for some constant $C$ independent of $\lambda_1, \lambda_2$, at least for $\lambda_1$ and $\lambda_2$ sufficiently close. Some preliminary numerical calculations seem to support this conjecture but we have not succeeded in showing it formally.

### 4. Application of General Results to Network Subgraph Counts

We now return to the primary concern of this paper, that of characterizing the propagation of low-rate measurement error to subgraph counts in large, sparse graphs. We have already provided one such detailed characterization in Section 2, for counting edges under the assumption of an independent error model. Here we provide two additional characterizations, both of which fall under the more general context that was the focus in Section 3. In the first, we return to the problem of counting edges, but now under a simple dependent error model, while in the second we examine the task of counting chains of length two under an independent error model.
4.1. Edge Counts Under Dependent Edge Noise

Recall the context of the application in Section 2.2, wherein our subgraph count of interest was \( \eta(G) = |E| \) and our goal was to characterize the accuracy with which the distribution of the difference \( D_E = |\hat{E}| - |E| \) in (2.2) is approximated by a Skellam(\( \lambda, \lambda \)) distribution.

While it often can be expected in practice that the error associated with construction of the empirical graph \( \hat{G} \) will involve dependency across (non)edges, a precise characterization of such dependency is typically problem-specific and, more often than not, nontrivial in nature. Here, for the purposes of illustration, we instead describe a comparatively simpler model of dependency, one which allows further analysis and interpretation.

Let \( T = T_1 + T_2 \) be the total number of errors (of either type) made in constructing \( \hat{G} \), where
\[
T_1 = \sum_{\{i,j\} \in E^c} Y_{ij} \quad \text{and} \quad T_2 = \sum_{\{i,j\} \in E} (1 - Y_{ij})
\]
are the total number of Type I errors and Type II errors, respectively. We will model \( T_1 \) and \( T_2 \) as independent, but with the \( Y_{ij} \) defining each of these totals as dependent among themselves. To accomplish this, we equip \( T_\ell \) with a Conway-Maxwell binomial (COMB) distribution, with parameters \((\pi_\ell, \nu_\ell)\), for \( \ell = 1, 2 \), and assign to the corresponding \( Y_{ij} \) the unique exchangeable distribution. As a result, the homogeneity assumption (A5) is satisfied.

The COMB distribution has been proposed recently by Kadane [12] as a simple extension of the binomial distribution that allows for dependency among the corresponding Bernoulli random variables. We say \( S \) follows a COMB distribution, with parameters \( \pi \in [0, 1] \) and \( \nu \in \mathbb{R} \), if
\[
P(S = k) = \frac{\pi^k (1 - \pi)^{n-k} \binom{n}{k}^\nu}{S(\pi, \nu)},
\]
for \( k = 0, \ldots, n \), where
\[
S(\pi, \nu) = \sum_{k=0}^{n} \pi^k (1 - \pi)^{n-k} \binom{n}{k}^\nu
\]
is a normalizing constant. The case of \( \nu = 1 \) corresponds to the standard binomial distribution. To create a collection \( X_1, \ldots, X_n \) of \( n \) dependent Bernoulli random variables, we assign probability \( P(S = k)/\binom{n}{k}^\nu \) to any sequence for which \( \sum_{i=1}^{n} X_i = k \). By construction, the \( X_i \)'s are exchangeable.

As \( \nu \to -\infty \), the distribution \( P(S = k) \) puts more and more of its weight towards the extremes of 0 and \( n \) [12]. In fact, we can show that, for \( \nu \leq 1 \), the Bernoulli random variables corresponding to the COMB distribution are negatively associated. To do so, it is sufficient to show that the distribution defined by the probabilities \( P(S = k)/\binom{n}{k}^\nu \) is log-supermodular [8]. That is, letting \( \mu(x) \) denote this distribution, evaluated at the \( n \)-length binary vector \( x \), we wish to show that \( \mu(x) \mu(x') \leq \mu(x \land x') \mu(x \lor x') \).
for all $\mathbf{x}, \mathbf{x}' \in \{0, 1\}^n$, where $\wedge$ and $\vee$ denote elementwise min and max operations, respectively. Straightforward calculations yield that this inequality holds if and only if

$$1 \leq \left[ \left( \frac{\sum_i^n x_i}{n} \right) \left( \frac{\sum_i^n x'_i}{n} \right) \right]^{1-\nu}.$$ 

Interestingly, the parameter $\pi$ does not play a role here. Since the terms in the numerator are greater than or equal to those in the denominator, the quantity in brackets is greater than or equal to 1, and thus the inequality will hold if and only if $1 - \nu \geq 0$, i.e., for $\nu \leq 1$.

Therefore, we define the subcollections $\{Y_{ij}, \{i, j\} \in E\}$ and $\{Y_{ij}, \{i, j\} \in E\}$ through independent COMB distributions, with parameters $\nu_1 \leq 0$, $\ell = 1, 2$. Then each of these subcollections of random variables are negatively associated. This captures the simple notion that while the making of Type I and Type II errors at the level of (non)edges may depend on unrelated mechanisms, nevertheless the tendency to, for example, make a Type II (I) error on a given edge (nonedge) in $G$ will increase the chances of similarly making an error on other edges (nonedges). Furthermore, since the union of independent sets of negatively associated random variables is also negatively associated, the entire collection of errors $Y_{ij}$ are negatively associated (e.g., [11, Prop 7]).

As a result, we can make use of the bound in (3.7). Under assumption (A3), $\lambda_1 = \lambda_2$, and so, utilizing the result in Theorem 5, we have that

$$d_{KS}(D_E, \text{Skellam}(\lambda, \lambda)) \leq O \left( \frac{2\lambda - \text{Var}(T_1 + T_2)}{2\lambda} \right).$$

If, for example, $\lambda = \log n_v$ (i.e., $\gamma = 0$ and $\kappa = 1$), then in order to have the bound above in (4.1) achieve the same rate as that in (2.6), i.e., a rate of $O(n_v^{-1})$, we find that we would need $\text{Var}(T_1 + T_2) = \Theta (2\log n_v (1 - C/n_v)).$

In principle, from here one can explore the implications of this result on the parameters $(\pi_\ell, \nu_\ell), \ell = 1, 2$. Since $T_1$ and $T_2$ are independent in our model, $\text{Var}(T_1 + T_2) = \text{Var}(T_1) + \text{Var}(T_2)$. Additionally, under the COMB distribution, the moment generating function has the form [12]

$$\mathbb{E} \left[ e^{tS} \right] = \Psi(e^t \pi/(1 - \pi), \nu)/(\Psi(\pi/(1 - \pi), \nu),$$

where $\Psi(x, \nu) = \sum_{k=0}^n x^k (\binom{n}{k})^\nu$. From this, expressions for the mean and variance may be obtained for $T_1$ and $T_2$, and numerical methods may be used to explore the space of values $(\pi, \nu)$ that satisfy the constraints on mean and variance imposed under our problem, for a given choice of $n_v$.

4.2. Counts of Chains of Length 2 Under Independent Edge Noise

Now consider the problem wherein the goal is to count chains of length 2 in a graph $G$, but the errors in declaring (non)edges in $\hat{G}$ are again independent. Even in this simple variation, there is sufficient richness to illustrate the additional complexity faced
in moving away from the problem of counting edges in $G$. Our goal in this subsection is modest in simply providing a sketch of key elements of the structure of the problem faced here, with an eye primarily towards highlighting the nature of some of the challenges that remain to be addressed.

Let $V^{(3)}$ be the set of triples $\{i,j,k\}$, for $i \neq j \neq k$, and let $C^{(3)}_2$ be the collection of all such triples that possess exactly two edges among them in the true graph $G$, i.e., they form a chain of length 2. Our interest is in the discrepancy $D_{2C} = |\hat{C}_2^{(3)}| - |C^{(3)}_2|$, where $C^{(3)}_2$ is the number of length-2 chains in the observed graph $G$.

In order to better understand the nature of the task we now face, we need to impose additional structure on $G$. We assume, for convenience, that $G$ is a realization of an Erdos-Renyi binomial random graph, i.e., where each edge in $E$ is determined independently according to a coin flip with probability $p$. Motivated by the assumption of sparseness in (A2), we let $p$ scale like $(\log n_v)/n_v$. Most immediately relevant to whether or not $D_{2C}$ has an approximate Skellam distribution or not is whether or not length-2 chains occur sparsely in $G$. Since $|C^{(3)}_2|$ scales like $n_v \log^2 n_v$ in expectation under the Erdos-Renyi model, comparison to $|V^{(3)}|$ is $O(n_v^3)$ suggests that the distribution of $D_{2C}$ is indeed more likely to be Poisson-like rather than normal.

Focusing then on the use of a Skellam approximation, there is the question of what values the parameters $\lambda_1, \lambda_2$ should take. Using notation analogous to that in equation (3.1), in defining the general difference $U$, we have

$$D_{2C} = \sum_{(i,j,k) \in C^{(3)}_2} L_{ijk} - \sum_{(i,j,k) \in C^{(3)}_2} M_{ijk},$$

where $L_{ijk}$ indicates the relevant Type I error in declaring a non-chain to be a chain, and $M_{ijk}$, the relevant Type II error, in declaring a chain to be a non-chain. Under the assumptions of independence and homogeneity among errors on (non)edges, $q_{ijk} \equiv \mathbb{E}[L_{ijk}]$ is equal to $1 - (1 - \beta)^2$, while $p_{ijk} \equiv \mathbb{E}[M_{ijk}]$ is either $\alpha^2$ or $(1 - \beta)\alpha$, depending on whether $\{i,j,k\}$ possess no edges or one edge among them, respectively. Letting $C^{(3)}_0$ be the number of triples with no edges among them, and $C^{(3)}_1$, the number with exactly one edge among them, we therefore have

$$\lambda_1 = |C^{(3)}_0| \alpha^2 + |C^{(3)}_1| (1 - \beta)\alpha \quad \text{and} \quad \lambda_2 = |C^{(3)}_2| [1 - (1 - \beta)^2].$$

The precise values of these quantities will vary with the structure of $G$. Suppose again that we assume an Erdos-Renyi structure for the true graph $G$, with parameter $p$ scaling like $(\log n_v)/n_v$. Also, recall that under the homogeneity assumption (A5), we have $\alpha = \lambda/|E^c|$ and $\beta = \lambda/|E|$, where $\lambda$ is the total expected number of errors in counting (non)edges. Then substituting the expected values for the cardinalities of the sets $C^{(3)}_i$, $\ell = 0, 1, 2$, under the Erdos-Renyi regime, tedious but straightforward calculations yield that to first order $\lambda_1$ and $\lambda_2$ both scale like $(\lambda/3) \log n_v$. However, they differ in their second leading terms, with that of $\lambda_2$ scaling like $\lambda^2/(6n_v)$, and that of $\lambda_1$, like $(\lambda/(3n_v)) \log^2 n_v$.

These calculations suggest, therefore, that to first order an assumption of $\lambda_1 = \lambda_2 = (\lambda/3) \log n_v$ is reasonable in this particular setting. However, in next seeking to bound the accuracy with which this Skellam approximation holds, note that despite the fact...
that our errors in observing (non)edges are assumed independent, some of the Type I and Type II errors we make in counting chains of length 2 can be expected to be dependent. So a bound like that in (3.5) cannot be used. On the other hand, unlike in Section 4.1 above, a property like negative association seems much too strong to expect to hold here. Therefore, rather than making use of a relatively simple bound like that in (3.7), we need to resort instead to something like (3.6).

The first two terms in this latter bound are the sums of the squared Type I and II errors, respectively, in counting length-2 chains. That is, sums of the terms \( p^2_{ijk} \) and \( q^2_{ijk} \). Using calculations analogous to those above, again assuming an Erdos-Renyi structure for \( G \), the first sum can be shown to scale in expectation like \( \lambda^2 \log n_v / n_v \), and the second, like \( \lambda^2 / n_v \). Hence, the second dominates and, under assumption (A4), we might hope to have a bound of order \( O \left( (n_v^2 \log 2k n_v) / n_v \right) \). If the Type I and II errors we make in counting chains of length 2 were independent, this would indeed be the case, as (3.6) would reduce to the simpler bound in (3.5). Given the absence of such independence, it is necessary to consider the next set of terms in (3.6), which will involve the covariances \( \text{Cov} (L_{ijk}, L_{i'j'k'}) \), \( \text{Cov} (M_{ijk}, M_{i'j'k'}) \), and \( \text{Cov} (L_{ijk}, M_{i'j'k'}) \).

First, note that under our independent error model these covariances will be non-zero only when the sets \( \{i, j, k\} \) and \( \{i', j', k'\} \) consist of just four unique vertices among them, i.e., say \( \{i, j, k\} \) and \( \{j, k, \ell\} \). In principle, we would then proceed by evaluating each of the covariances above in turn. Note that the expression for those covariances involving Type I errors (i.e., variables \( L_{ijk} \)) will vary depending on whether the corresponding triple of vertices contains no edges or only one edge. However, rather than enumerate all of these possibilities, it is sufficient for our purposes simply to consider \( \text{Cov} (M_{ijk}, M_{jkt}) \). These two variables pertain to a length-3 chain on a set of four vertices \( \{i, j, k, \ell\} \), resulting in two length-2 chains that overlap in a single edge. Since \( E \left[ M_{ijk} \right] M_{jkt} \right] = 1 - (1 - \beta)^3 \), and \( E \left[ M_{ijk}\right] = 1 - (1 - \beta)^3 \), we find that \( \text{Cov} (M_{ijk}, M_{jkt}) = \beta (3 - \beta)(1 - \beta)^2 \), which scales like \( 3\beta + O(\beta^2) \). Under the same Erdos-Renyi model used above, the expected number of length-3 chains scales like \( (n_v/24) \log^3 n_v \). So the total contribution to the bound in (3.6) by terms of the form \( \text{Cov} (M_{ijk}, M_{i'j'k'}) \) is expected to scale like \( (\lambda/8) \log^2 n_v \).

Therefore, this term will dominate the contributions of the first two sums defining the bound in (3.6), which we argued above would scale like \( \lambda^2 / n_v \). There is the term \( ||\Delta f|| \) in (3.6) still to be taken into account. Formally, this term cannot be controlled without an extension of our Theorem 5 to the case of \( \lambda_1 \neq \lambda_2 \). However, having argued above that \( \lambda_1 \approx \lambda_2 \approx (\lambda/3) \log n_v \), we might conjecture that \( ||\Delta f|| \leq O \left( (\lambda \log n_v)^{-1} \right) \). But even if true this will not be enough to compensate for the \( O \left( \lambda \log^2 n_v \right) \) contribution described just above.

Hence, a more refined bound than that in (3.6) is needed to quantify the accuracy of Skellam distribution in this problem and thus, by extension, in problems involving larger and more complicated subgraphs \( H \).

5. Discussion

Our work in this paper is, to the best of our knowledge, the first to present and treat a general formulation of the problem of propagating errors from a modern complex
network to descriptors of that network. Our formulation is reminiscent of the type of ‘signal plus noise’ model commonly used in nonparametric function estimation and digital signal processing. In particular, in our formulation the true underlying graph \( G \) is fixed. This necessitates a different treatment than, say, traditional analysis of subgraph counts in classical Erdos-Renyi random graphs. In the special case where an Erdos-Renyi model is assumed (as we have in Section 4.2, for the purposes of illustration), as well as assuming independence among the measurement errors, and the analysis is done without conditioning on \( G \), then the problem may be viewed as involving a classical random graph with two values for the probability of an edge arising in \( \hat{G} \) (i.e., depending on whether or not there was an edge between a given pair of vertices in \( G \)). In general, however, either when \( G \) is fixed or from some other class of random graph models (e.g., various models with heterogeneous degree distributions), or when the measurement errors are dependent, the problem is more involved. Our formulation allows us to focus our analysis firstly on a high-level notion of Type I and II errors among (non)edges, and then secondly on the manner in which the structure of the underlying graph \( G \) may interact with those errors.

We view our work as laying a foundation on an important new problem area, and thus – not surprisingly – our work also suggests several challenging directions for future research. First, there is need to control the term \( ||\Delta f|| \) in our general Stein bound in (3.4) for the case where \( \lambda_1 \neq \lambda_2 \). Our proof of Theorem 5, in Section 3.3, bounding this quantity for the case of \( \lambda_1 = \lambda_2 \), required the control of alternating sums of differences of the ratios of modified Bessel functions of the first kind. As such, the treatment is necessarily delicate. Furthermore, the literature on quantities of this sort is lacking and, hence, we were required to develop several novel analysis results. These results, which are of independent interest, are available in a separate manuscript [4]. Second, even in the case where \( \lambda_1, \lambda_2 \) are equal, or approximately equal, as in the example of Section 4.2, more refined bounds will often be needed than the Stein bounds we provide here (e.g., in (3.6) and (3.7)) on the Kolmogorov-Smirnov distance \( d_{KS}(D, W) \) between the distribution of the discrepancy \( D = \eta(\hat{G}) - \eta(G) \) and its approximating Skellam \( W \). Alternately, there may be specific models of measurement error (perhaps context-specific) that strike a useful balance between empirical relevance, analytical tractability, and sufficient control of quantities like the covariances in (3.6). Finally, our work suggests the need for additional work characterizing subgraph counts in large, sparse random graphs of a more general nature than classical Erdos-Renyi random graphs. Our example in Section 4.2 illustrates how it can be expected that the ‘signal plus noise’ formulation we have posed here leads to a need for understanding the large-graph limit (i.e., \( n_v \to \infty \)) of subgraph counts in general settings. In this direction, presumably the recent work on the theory of graph limits for sparse graphs will be useful (e.g., [6, 7]).

6. Appendix

6.1. Proof of Theorem 1.

The proof of this theorem is presented in two parts, first treating the approximation by the Skellam distribution (i.e., as summarized in (2.6)), and then treating the approxi-
mation by the standard normal distribution (i.e., as summarized in (2.7) and (2.8)).

6.1.1. Approximation by Skellam: Proof of (2.6).

To establish the bound in (2.6), we exploit the independence of the random variables $Y_{ij}$ to combine existing bounds for Poisson approximations to each of the individual sums in (2.5). For this we require the following proposition.

**Proposition 1.** Let $X, Y, Z,$ and $W$ be real-valued random variables taking at most countably infinite many values on the same support $S$. Suppose that

$$d(X, Z) := \sup_{t \in \mathbb{R}} |\mathbb{P}[X \leq t] - \mathbb{P}[Z \leq t]| < \frac{\epsilon}{2}$$

and

$$d(Y, W) = \sup_{t \in \mathbb{R}} |\mathbb{P}[Y \leq t] - \mathbb{P}[W \leq t]| < \frac{\epsilon}{2}.$$  

and that $(X, Y)$ and $(Z, W)$ are independent. Then,

$$d(X - Y, Z - W) < \epsilon.$$  

**Proof of Proposition 1:** While we suspect that this proposition is not novel, we include its proof here for completeness. Let $\{s_n\}_{n=-\infty}^{\infty}$ be an ordering of the support $S$. Since $(X, Y)$ are independent:

$$\mathbb{P}[X - Y \leq t] = \sum_{n=-\infty}^{\infty} \mathbb{P}[X \leq t + s_n] \mathbb{P}[Y = s_n].$$

and similarly

$$\mathbb{P}[Z - W \leq t] = \sum_{n=-\infty}^{\infty} \mathbb{P}[Z \leq t + s_n] \mathbb{P}[W = s_n].$$

So,

$$|\mathbb{P}[X - Y \leq t] - \mathbb{P}[Z - W \leq t]|$$

$$= \left| \sum_{n=-\infty}^{\infty} \mathbb{P}[X \leq t + s_n] \mathbb{P}[Y = s_n] - \sum_{n=-\infty}^{\infty} \mathbb{P}[Z \leq t + s_n] \mathbb{P}[W = s_n] \right|$$

$$\leq \left| \sum_{n=-\infty}^{\infty} (\mathbb{P}[X \leq t + s_n] - \mathbb{P}[Z \leq t + s_n]) \mathbb{P}[Y = s_n] \right|$$

$$+ \left| \sum_{n=-\infty}^{\infty} \mathbb{P}[Z \leq t + s_n] (\mathbb{P}[Y = s_n] - \mathbb{P}[W = s_n]) \right|$$

$$\leq \sup_n |\mathbb{P}[X \leq t + s_n] - \mathbb{P}[Z \leq t + s_n]|$$

$$+ \left| \sum_{n=-\infty}^{\infty} \mathbb{P}[Z \leq t + s_n] (\mathbb{P}[Y = s_n] - \mathbb{P}[W = s_n]) \right|$$
Next, use the fact $P[Y = s_n] = P[Y \leq s_n] - P[Y \leq s_{n-1}]$, and similarly for $W$, and define $A(n) = P[Y \leq s_n] - P[W \leq s_n]$, so that

$$\left| \sum_{n=-\infty}^{\infty} P[Z \leq t + s_n] (P[Y = s_n] - P[W = s_n]) \right|$$

$$= \left| \sum_{n=-\infty}^{\infty} P[Z \leq t + s_n] (P[Y \leq s_n] - P[Y \leq s_{n-1}] - (P[W \leq s_n] - P[W \leq s_{n-1}])) \right|$$

$$= \left| \sum_{n=-\infty}^{\infty} P[Z \leq t + s_n] (A(n) - A(n-1)) \right|$$

$$= \sum_{n=-\infty}^{\infty} |A(n)| \left| P[Z \leq t + s_n] - P[Z \leq t + s_{n+1}] \right|$$

$$\leq \sum_{n=-\infty}^{\infty} |A(n)| \left| P[Z \leq t + s_n] - P[Z \leq t + s_{n+1}] \right|$$

$$\leq \sup_n |A(n)|$$

Thus,

$$|P[X - Y \leq t] - P[Z - W \leq t]|$$

$$\leq \sup_n |P[X \leq t + s_n] - P[Z \leq t + s_n]| + \sup_n |P[Y \leq s_n] - P[W \leq s_n]|$$

Taking supremums over $t$ now yields the result.

$\square$

With $\lambda = |E^c|\alpha = |E|\beta$, and applying Proposition 1 to the standard bounds obtained for each of the sums in $D_E$ by Stein’s method for the Poisson distribution [5],
we obtain
\[ d_{KS} \left( D_E, \text{Skellam}(\lambda, \lambda) \right) \leq \frac{1}{\lambda} \left( \sum_{(i,j) \in E^c} \alpha^2 + \sum_{(i,j) \in E} \beta^2 \right) = \frac{|E^c| \alpha^2 + |E| \beta^2}{|E^c| \lambda} \]
\[ = \alpha + \frac{|E^c| \alpha}{|E|} = \alpha + \frac{\binom{n}{2} - |E|}{|E|} \alpha \]
\[ = \frac{\binom{n}{2}}{|E|} \alpha . \]

Noting \( \alpha = \lambda/|E^c| \), the last quantity above is seen to behave like \( \lambda/|E| \) which, under assumptions (A2) and (A4) behaves like \( O \left( (n^{1-\gamma} \log^{1-\alpha} n_{\text{v}})^{-1} \right) \), and so the bound in (2.6) is established.

### 6.1.2. Approximation by standard normal: Proof of (2.7) and (2.8).

Next, to establish the bounds in (2.7) and (2.8), we use the following result from Stein’s method for the normal distribution (e.g., [5]).

**Theorem 6.** Let \( \xi_1, \ldots, \xi_n \) be independent random variables which have zero means and finite variances \( \mathbb{E} \left[ \xi_i^2 \right] = \sigma_i^2, \ 1 \leq i \leq n \), and satisfy \( \sum_{i=1}^n \sigma_i^2 = 1 \). If \( F_n(x) \) is the cdf of \( \sum_{i=1}^n \xi_i \), then, for every \( \epsilon > 0 \),

\[ 1 - e^{-\frac{\pi}{4}} \sum_{i=1}^n \mathbb{E} \left[ \xi_i^2 I_{(\xi_i > \epsilon)} \right] - \sum_{i=1}^n \sigma_i^4 \leq \sup_{x \in \mathbb{R}} \left| F_n(x) - \Phi(x) \right| \leq 7 \sum_{i=1}^n \mathbb{E} \left[ |\xi_i|^3 \right] . \]

We apply this theorem, with \( \xi_i = X_i/\sigma \) where \( X_i \) is a term in one of the sums of \( D_E \), to establish each of our upper and lower bounds in turn.

**Upper Bound in (2.7):** First, note that since

\[ \sum_{i=1}^n \mathbb{E} \left[ |\xi_i|^3 \right] = \frac{\sum_{i=1}^n \mathbb{E} \left[ |X_i|^3 \right]}{\sigma^3} , \]

and

\[ \mathbb{E} \left[ |X_i|^3 \right] = \alpha (1 - \alpha) \left[ (1 - \alpha)^2 + \alpha^2 \right] \quad \text{or} \quad \beta (1 - \beta) \left[ (1 - \beta)^2 + \beta^2 \right] , \]
with $n$ understood to be either $|E^c|$ or $|E|$, it follows that
\[
\sum_{i=1}^{n} E \left[ |\xi_i|^3 \right] = \frac{\alpha(1-\alpha) \left[ (1-\alpha)^2 + \alpha^2 \right] |E^c| + \beta(1-\beta) \left[ (1-\beta)^2 + \beta^2 \right] |E|}{(\alpha(1-\alpha)|E^c| + \beta(1-\beta)|E|)^{\frac{3}{2}}} \\
\leq \max \{ (1-\alpha)^2 + \alpha^2, (1-\beta)^2 + \beta^2 \} \frac{\alpha(1-\alpha)|E^c| + \beta(1-\beta)|E|}{(\alpha(1-\alpha)|E^c| + \beta(1-\beta)|E|)^{\frac{3}{2}}} \\
= \max \{ (1-\alpha)^2 + \alpha^2, (1-\beta)^2 + \beta^2 \} \frac{1}{\sqrt{2-(\alpha+\beta)}} \\
= \max \{ (1-\alpha)^2 + \alpha^2, (1-\beta)^2 + \beta^2 \} \frac{1}{\sqrt{2-(\alpha+\beta)}},
\]
where in the last equality we have used the fact that $\beta = (|E^c|/|E|) \alpha$ follows from (A3). Finally, note that
\[(1-\alpha)^2 + \alpha^2 = 1 - 2\alpha + 2\alpha^2 = 1 - 2\alpha(1-\alpha) \leq 1\]
and the same holds for $(1-\beta)^2 + \beta^2$, since $0 \leq \alpha, \beta \leq 1$, so that
\[
\sum_{i=1}^{n} E \left[ |\xi_i|^3 \right] \leq \frac{1}{\sqrt{2-(\alpha+\beta)}} \cdot \frac{1}{\sqrt{\alpha|E^c|}}.
\]
This immediately implies, after another application of $\beta = (|E^c|/|E|) \alpha$,
\[
\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{7}{\sqrt{2-(\alpha + |E^c|/|E|) \alpha}} \cdot \frac{1}{\sqrt{\alpha|E^c|}}.
\]
Again using $\alpha = \lambda/|E^c|$ and invoking assumptions (A2) and (A4), the expression in (2.7) follows.

**Lower bound in (2.8):** First, note that since $\xi_i = X_i/\sigma$, $\sigma_i^2 = \alpha(1-\alpha)/\sigma^2$ or $\sigma_i^2 = \beta(1-\beta)/\sigma^2$. Thus,
\[
\sum_{i=1}^{n} \sigma_i^4 = \frac{(\alpha(1-\alpha))^2|E^c| + (\beta(1-\beta))^2|E|}{(\alpha(1-\alpha)|E^c| + \beta(1-\beta)|E|)^2} \\
= \frac{(\alpha(1-\alpha))^2|E^c| + \left(\frac{|E^c|}{|E|} \alpha(1-\beta)\right)^2|E|}{(2-(\alpha+\beta))^2(\alpha|E^c|)^2} \\
= \frac{1}{|E^c|} \cdot \frac{1}{(2-(\alpha+\beta))^2} \cdot \left[ (1-\alpha)^2 + \frac{|E^c|}{|E|} (1-\beta)^2 \right] \\
= \frac{1}{(2-(\alpha+\beta))^2} \left[ (1-\alpha)^2 + \frac{|E^c|}{|E|} (1-\beta)^2 \right],
\]
where in the second equality, we have used $\beta = (|E^c|/|E|) \alpha$. 


Next, choose $\epsilon = 1/(2\sigma)$. Note that this is the midpoint of the intervals

$$\left(\frac{\alpha}{\sigma}, 1 - \frac{\alpha}{\sigma}\right), \quad \text{and} \quad \left(\frac{\beta}{\sigma}, 1 - \frac{\beta}{\sigma}\right)$$

if $\alpha, \beta < 1/2$ and of the intervals

$$\left(\frac{1 - \alpha}{\sigma}, \frac{\alpha}{\sigma}\right), \quad \text{and} \quad \left(\frac{1 - \beta}{\sigma}, \frac{\beta}{\sigma}\right).$$

if $\alpha, \beta \geq 1/2$. In either case, these are the endpoints of the interval formed by the values of $|\xi| = |X_\epsilon|/\sigma$.

Due to the symmetry in these intervals about $\frac{1}{2}$, we may, without loss of generality, assume $\alpha, \beta < 1/2$. In doing so, and using $\beta = (|E^c|/|E|)\alpha$,

$$\frac{1 - e^{-\frac{\epsilon^2}{4}}}{40} \sum_{i=1}^n \mathbb{E} \left[ \xi_i^2 I_{(|\xi_i| > \epsilon)} \right]$$

$$= \frac{1 - e^{-\frac{\epsilon^2}{4}}}{40} \cdot \frac{(1 - \alpha)^2 \alpha |E^c| + (1 - \beta)^2 \beta |E|}{\alpha(1 - \alpha)|E^c| + \beta(1 - \beta)|E|}$$

$$= \frac{1 - e^{-\frac{\epsilon^2}{4}}}{40} \cdot \frac{2 - (\alpha + \beta)}{(1 - \alpha)^2 + (1 - \beta)^2}$$

$$\geq e^{-\frac{\epsilon^2}{4}} \cdot \frac{1}{160} \cdot \frac{1}{2 - (\alpha + \beta)}$$

$$= e^{-\frac{\epsilon^2}{4}} \cdot \frac{1}{640} \cdot \frac{1}{\alpha |E^c|} \cdot \frac{(1 - \alpha)^2 + (1 - \beta)^2}{(2 - (\alpha + \beta))^2}.$$ 

Combining the two sets of expressions above, the lower bound becomes

$$\frac{1 - e^{-\frac{\epsilon^2}{4}}}{40} \sum_{i=1}^n \mathbb{E} \left[ \xi_i^2 I_{(|\xi_i| > \epsilon)} \right] - \sum_{i=1}^n \sigma_i^4$$

$$\geq e^{-\frac{\epsilon^2}{4}} \cdot \frac{1}{640} \cdot \frac{1}{\alpha |E^c|} \cdot \frac{(1 - \alpha)^2 + (1 - \beta)^2}{(2 - (\alpha + \beta))^2} - \frac{1}{(2 - (\alpha + \beta))^2} \left[ \frac{(1 - \alpha)^2}{|E^c|} + \frac{(1 - \beta)^2}{|E|} \right]$$

$$= \frac{1}{(2 - (\alpha + \beta))^2} \left[ (1 - \alpha)^2 \left( \exp \left( \frac{-1}{16} \cdot \frac{1}{\alpha |E^c|} \cdot \frac{1}{2 - (\alpha + \beta)} \right) \cdot \frac{1}{\alpha |E^c|} - \frac{1}{|E^c|} \right) \right.$$

$$\left. + (1 - \beta)^2 \left( \exp \left( \frac{-1}{16} \cdot \frac{1}{\alpha |E^c|} \cdot \frac{1}{2 - (\alpha + \beta)} \right) \cdot \frac{1}{\alpha |E^c|} - \frac{1}{|E^c|} \right) \right] (6.1)$$
Note that each of the two quantities in the brackets is positive when
\[ \alpha, \beta < \exp \left( -\frac{1}{16} \cdot \frac{1}{\alpha |E|^\gamma} \cdot \frac{1}{2^{-(\alpha + \beta)}} \right) \]
This expression may be rewritten as
\[ \min(\log |E|^\gamma, \log |E|) - \log \lambda > \log 640 + [16\lambda(2 - \alpha/\delta)]^{-1} \]
where \( \delta = |E|/(n_v^2) \) is the density of edges in \( G \). Recalling that \( |E| \) will be smaller than \( |E^c| \) and ignoring the modest contribution of the second term on the right above, we see that (6.2) is essentially equivalent to requiring that \( |E|/\lambda > 640 \). Given the constraints imposed on each of \( |E| \) and \( \lambda \) in assumptions (A2) and (A4), respectively, this requirement will be met for \( n_v \) sufficiently large.

6.2. Proof of Theorem 2.

We begin with the operator,
\[ \mathcal{A}[f(k)] = \lambda_1 f(k+1) - kf(k) - \lambda_2 f(k-1) \]
with the intent of showing that the random variable \( W \sim \text{Skellam}(\lambda_1, \lambda_2) \) if and only if \( \mathbb{E}\mathcal{A}[f(W)] = 0 \) for any bounded function \( f: \mathbb{Z} \rightarrow \mathbb{R} \).

We begin with the necessity direction and the computation of
\[ \mathbb{E}\mathcal{A}[f(W)] = \mathbb{E} \left[ \lambda_1 f(W+1) - Wf(W) - \lambda_2 f(W-1) \right] \]
\[ \propto \sum_{k=-\infty}^{\infty} \left[ \lambda_1 f(k+1) - kf(k) - \lambda_2 f(k-1) \right] \left( \sqrt{\lambda_1/\lambda_2} \right)^k I_k \]
where \( \propto \) is to be read as “proportional to,” and as shorthand, we write \( I_k \) for \( I_k(2\sqrt{\lambda_1\lambda_2}) \).

By standard properties of \( I_k \) (e.g., [11]) we have that
\[ I_{k-1} - I_{k+1} = \frac{k}{\sqrt{\lambda_1\lambda_2}} I_k \]
or, in other words,
\[ \sqrt{\lambda_1\lambda_2} \frac{I_{k-1}}{I_k} - \sqrt{\lambda_1\lambda_2} \frac{I_{k+1}}{I_k} = k. \] (6.3)

This means that
\[ \mathbb{E}\mathcal{A}[f(W) 1\{W \leq n\}] \propto \sum_{k=-\infty}^{n} \left[ \sqrt{\frac{\lambda_1}{\lambda_2}} f(k+1) - \frac{I_{k-1}}{I_k} f(k) + \right. \]
\[ \frac{I_{k+1}}{I_k} f(k) - \sqrt{\frac{\lambda_2}{\lambda_1}} f(k-1) \left( \frac{\sqrt{\lambda_1}}{\lambda_2} \right)^k I_k = \sum_{k=-\infty}^{n} \left[ \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^{k+1} I_k f(k+1) - \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^k I_{k-1} f(k) \right] \\
+ \sum_{k=-\infty}^{n} \left[ \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^k I_{k+1} f(k) - \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^{k-1} I_k f(k-1) \right] \\
= \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^{n+1} I_n f(n+1) + \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^n I_{n+1} f(n). \]

Now, since \( f \) is bounded,
\[
\lim_{n \to \infty} \left\{ \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^{n+1} I_n f(n+1) + \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^n I_{n+1} f(n) \right\} = 0
\]
so that by monotone convergence,
\[
\mathbb{E} [A[f(W)]] = \lim_{n \to \infty} \mathbb{E} A[f(W) 1 \{W \leq n\}] \\
= \lim_{n \to \infty} \left\{ \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^{n+1} I_n f(n+1) + \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^n I_{n+1} f(n) \right\} \\
= 0
\]
which proves the claim.

To prove sufficiency, we begin with \( \mathbb{E} A[f(W)] = 0 \) and suppose that \( f_k(j) = 1 \{j = k\} \) for some \( j \in \mathbb{Z} \) in which case
\[
\lambda_1 p(k-1) - kp(k) - \lambda_2 p(k+1) = 0
\]
where \( p(k) = \mathbb{P}(W = k) \). An ansatz of
\[
S(k) = \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^k I_k(2\sqrt{\lambda_1 \lambda_2}) \quad \text{and} \quad T(k) = \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^k K_k(2\sqrt{\lambda_1 \lambda_2})
\]
shows that \( S \) and \( T \) form two linearly independent solutions to this second order linear difference equation, where \( I_k(x) \) and \( K_k(x) \) are the modified Bessel functions of the first and second kinds. Thus, we know that the general solution is given by,
\[
p(k) = C_1 S(k) + C_2 T(k)
\]
for some constants \( C_1, C_2 \in \mathbb{R} \).
Now, to determine the constants $C_1$ and $C_2$ we appeal to the fact that $\sum_{k=-\infty}^{\infty} p(k) = 1$. Since $I_k, K_k > 0$ for all $k \in \mathbb{Z}$ and $\sum_{k=-\infty}^{\infty} K_k = \infty$ it must be that $C_2 = 0$. Now, consider the generating function

$$e^{z(t+1/t)} = \sum_{k=-\infty}^{\infty} t^k I_k(z)$$

which means that

$$C_1 = \frac{1}{\sum_{k=-\infty}^{\infty} \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^k I_k \left( 2\sqrt{\lambda_1\lambda_2} \right)} = \frac{1}{e^{\lambda_1\lambda_2 \left( \frac{1}{\sqrt{\lambda_1}} + \frac{1}{\sqrt{\lambda_2}} \right)}} = e^{-(\lambda_1+\lambda_2)}$$

so that

$$p(k) = e^{-(\lambda_1+\lambda_2)} \left( \frac{\lambda_1}{\lambda_2} \right)^k I_k \left( 2\sqrt{\lambda_1\lambda_2} \right)$$

so that $W \sim \text{Skellam}(\lambda_1, \lambda_2)$.

6.3. Proof of Theorem 3.

Given that $f_x$ is a solution to $A[f_x(k)] = g_x(k)$, we have

$$\lambda_1 f_x(k+1) - kf_x(k) - \lambda_2 f_x(k-1) = 1(k \leq x) - \mathbb{P}[W \leq x].$$

Substituting $k = U$ and taking expected values, we obtain,

$$|\mathbb{P}[U \leq x] - \mathbb{P}[W \leq x]| = |\mathbb{E}[\lambda_1 f_x(U+1) - U f_x(U) - \lambda_2 f_x(U-1)]|, \quad (6.4)$$

Next, recall from (3.1) that $U = \sum_{k=1}^{n} L_k - \sum_{k=1}^{m} M_k$. Since $\lambda_1 = \sum_{k=1}^{n} p_k$ and $\lambda_2 = \sum_{k=1}^{m} q_k$, we have after conditioning on $L_k$ and $M_k$,
|E [λ₁fₓ(U + 1) − Ufₓ(U) − λ₂fₓ(U − 1)]|
\[= \sum_{k=1}^{n} |\mathbb{E}[p_k f_x(U + 1) - L_k f_x(U)] + \sum_{k=1}^{m} |\mathbb{E}[M_k f_x(U) - q_k f_x(U - 1)]|
\[= \sum_{k=1}^{n} p_k (\mathbb{E}[f_x(U + 1) - f_x(U)|L_k = 1]) + \sum_{k=1}^{m} q_k (\mathbb{E}[f_x(U)|M_k = 1] - \mathbb{E}[f_x(U - 1)])|
\[\leq \sum_{k=1}^{n} p_k ||\Delta f||\mathbb{E}|U - U_{k}^{(L)}| + \sum_{k=1}^{m} q_k ||\Delta f||\mathbb{E}|U - U_{k}^{(M)}|
\[= ||\Delta f|| \left[\sum_{k=1}^{n} p_k \mathbb{E}|U - U_{k}^{(L)}| + \sum_{k=1}^{m} q_k \mathbb{E}|U - U_{k}^{(M)}|\right].
\]

Combining this with (6.4) yields the result.

\[\square\]


First, consider the solution to
\[λ₁f(k + 1) - kf(k) - λ₂f(k - 1) = g(k), \quad (6.5)\]
for some bounded function \(g : \mathbb{Z} \mapsto \mathbb{R}\), with the boundary condition
\[\lim_{k \to -\infty} \left(\frac{λ₁}{λ₂}\right)^k I_k f(k) = 0. \quad (6.6)\]

We use (6.3) to substitute for \(k\) in (6.5). Then, multiplying both sides of (6.5) by \(\left(\frac{λ₁}{λ₂}\right)^k I_k\), we obtain,
\[λ₁ \left(\frac{λ₁}{λ₂}\right)^k I_k f(k + 1) - \lambda₁ \left(\frac{λ₁}{λ₂}\right)^{k-1} I_{k-1} f(k)
+ \lambda₂ \left(\frac{λ₁}{λ₂}\right)^{k+1} I_{k+1} f(k) - \lambda₂ \left(\frac{λ₁}{λ₂}\right)^k I_k f(k - 1)
= \left(\frac{λ₁}{λ₂}\right)^k I_k g(k)\]

which is the same as,
\[
\left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^{k+1} I_k f(k+1) - \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^k I_{k-1} f(k) \\
+ \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^k I_{k+1} f(k) - \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^{k-1} I_k f(k-1) \\
= \frac{1}{\sqrt{\lambda_1 \lambda_2}} \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^k I_k g(k).
\]

Notice that we have grouped terms together so that summing over \( k \) yields a telescoping sum. So, summing over \( k \in \{-\infty, \ldots, n\} \) and using the boundary condition (6.6),

\[
\left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^{n+1} I_n f(n+1) + \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^n I_{n+1} f(n) = \frac{1}{\sqrt{\lambda_1 \lambda_2}} \sum_{k=-\infty}^{n} \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^k I_k g(k).
\]

Now, multiplying both sides by \((-1)^{n+1} / (I_n I_{n+1})\) and summing over \( n \in \{c, c+1, \ldots, m\} \) for \( m > c \) and over \( n \in \{m, m+1, \ldots, c-1\} \), for some initial condition \( c \in \mathbb{Z} \) and \( f(c) \in \mathbb{R} \), we obtain

\[
f(m) = \begin{cases} 
(-1)^m \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^m I_m \left[ (-1)^c \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^c \frac{1}{c} f(c) 
\right. \\
+ \frac{1}{\sqrt{\lambda_1 \lambda_2}} \sum_{n=c}^{m-1} (-1)^{n+1} \sum_{k=-\infty}^{n} \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^k I_k g(k) & \text{if } m > c \\
(-1)^m \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^m I_m \left[ (-1)^c \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^c \frac{1}{c} f(c) 
\right. \\
- \frac{1}{\sqrt{\lambda_1 \lambda_2}} \sum_{n=m}^{c-1} (-1)^{n+1} \sum_{k=-\infty}^{n} \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^k I_k g(k) & \text{if } m < c.
\end{cases}
\]

Note that if

\[
g(k) = g_x(k) = 1 \{k \leq x\} - \mathbb{P}(W \leq x)
\]

then

\[
\sum_{k=-\infty}^{n} \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^k I_k g(k) = \begin{cases} 
e^{-\lambda_1-\lambda_2} \mathbb{P}(W \leq n) \mathbb{P}(W > x) & \text{if } n \leq x \\
e^{-\lambda_1-\lambda_2} \mathbb{P}(W \leq x) \mathbb{P}(W > n) & \text{if } n \geq x
\end{cases}
\]

since, for example if \( n \leq x \):

\[
\sum_{k=-\infty}^{n} \left( \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}} \right)^k I_k g(k) = e^{-\lambda_1-\lambda_2} \sum_{k=-\infty}^{n} \mathbb{P}(W = k) g(k) \\
= e^{-\lambda_1-\lambda_2} \sum_{k=-\infty}^{n} \mathbb{P}(W = k) [1 \{k \leq x\} - \mathbb{P}(W \leq x)] \\
= e^{-\lambda_1-\lambda_2} \mathbb{P}(W \leq \min\{n, x\}) - \mathbb{P}(W \leq x) \mathbb{P}(W \leq n)
\]
\[ f_x(m) = \begin{cases} 
(-1)^m \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^m I_m \left[ (-1)^c \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^c \frac{1}{\lambda} f(c) \right. \\
+ \frac{e^{\lambda_1 + \lambda_2}}{\sqrt{\lambda_1 \lambda_2}} \sum_{n=0}^{m-1} (-1)^{n+1} I_n I_{n+1} \mathbb{P} (W \leq \min \{n, x\}) \mathbb{P} (W > \max \{n, x\}) \right] & \text{if } m > c \\
(-1)^m \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^m I_m \left[ (-1)^c \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^c \frac{1}{\lambda} f(c) \right. \\
- \frac{e^{\lambda_1 + \lambda_2}}{\sqrt{\lambda_1 \lambda_2}} \sum_{n=m}^{c-1} (-1)^{n+1} I_n I_{n+1} \mathbb{P} (W \leq \min \{n, x\}) \mathbb{P} (W > \max \{n, x\}) \right] & \text{if } m < c.
\end{cases} \]

The case that \( n \geq x \) is similar. This means that

\[ f_x(m) = \begin{cases} 
(-1)^m \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^m I_m \left[ (-1)^c \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^c \frac{1}{\lambda} f(c) \right. \\
+ \frac{e^{\lambda_1 + \lambda_2}}{\sqrt{\lambda_1 \lambda_2}} \sum_{n=0}^{m-1} (-1)^{n+1} I_n I_{n+1} \mathbb{P} (W \leq \min \{n, x\}) \mathbb{P} (W > \max \{n, x\}) \right] & \text{if } m > c \\
(-1)^m \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^m I_m \left[ (-1)^c \left( \sqrt{\frac{\lambda_1}{\lambda_2}} \right)^c \frac{1}{\lambda} f(c) \right. \\
- \frac{e^{\lambda_1 + \lambda_2}}{\sqrt{\lambda_1 \lambda_2}} \sum_{n=m}^{c-1} (-1)^{n+1} I_n I_{n+1} \mathbb{P} (W \leq \min \{n, x\}) \mathbb{P} (W > \max \{n, x\}) \right] & \text{if } m < c.
\end{cases} \]

6.5. Proof of Theorem 5.

Our proof of Theorem 5 is highly involved, from an analysis perspective, but the overall program can be stated in a relatively succinct manner. Accordingly, we sketch here the overall program behind our proof and refer the interested reader to the Supplementary Materials for a detailed account.

Recall that we are trying to obtain a bound on \(|\Delta f_x(j)| = |f_x(j + 1) - f_x(j)|\) independent of \(x \in \mathbb{R}\) and \(j \in \mathbb{Z}\). From Theorem 4, we have the solution to the Stein equation, however to use it to bound \(|\Delta f_x(j)|\), we need to simplify it further. For ease of notation, we simply refer to \(f\) instead of \(f_x\) and \(g\) instead of \(g_x\).

First, note that we have the freedom to choose the initial condition \((c, f(c))\). Making the choice that \(c = \lambda_2 - \lambda_1\), and hence that \(c = 0\) under the assumption that \(\lambda_1 = \lambda_2\), we are able to simplify our expression for \(f\) in Theorem 4 to read, in the case that \(m > 0\), as

\[ f(m + 1) = (-1)^{m+1} I_{m+1} \left[ \frac{1}{I_0} f(0) \right. \\
- \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0 I_1} \mathbb{P} (W \leq \min \{0, x\}) \mathbb{P} (W > \max \{0, x\}) \\
- \frac{e^{2\lambda}}{\lambda} \sum_{n=0}^{m-1} \frac{1}{I_{n+1} I_{n+2}} \mathbb{P} (W \leq \min \{n + 1, x\}) \mathbb{P} (W > \max \{n + 1, x\}) \right] \]

and, in the case that if \(m < 0\), as

\[ f(m - 1) = (-1)^{m-1} I_{m-1} \left[ \frac{1}{I_0} f(0) \right. \\
- \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0 I_1} \mathbb{P} (W \leq \min \{-1, x\}) \mathbb{P} (W > \max \{-1, x\}) \\
+ \frac{e^{2\lambda}}{\lambda} \sum_{n=m}^{-1} \frac{1}{I_{n+1} I_{n-1}} \mathbb{P} (W \leq \min \{n - 1, x\}) \mathbb{P} (W > \max \{n - 1, x\}) \right] .\]
Finally, for the case \( m = 0 \), we have

\[
f(0) = \frac{e^{2\lambda}}{2\lambda I_0 + I_1} \left[ \mathbb{P}(W \leq \min\{0, x\}) \mathbb{P}(W > \max\{0, x\}) \right. \\
+ \mathbb{P}(W \leq \min\{-1, x\}) \mathbb{P}(W > \max\{-1, x\}) \right].
\]

(6.8)

Next, through manipulation of the arguments in the sums defining the above expressions for \( f \), exploiting properties of the modified Bessel functions \( I_k \), and applying the triangle inequality, we are able to produce bounds on the differences \( |f(m+1) - f(m)| \) of the form

\[
|f(m+1) - f(m)| \leq \frac{\mathbb{P}(W \leq x)}{\lambda} \left\{ \sum_{n=1,3,\ldots}^{m-1} \frac{I_{m+1}I_n}{I_{n+2}} - \frac{I_m}{I_{n-1}} \right\} \\
+ \sum_{n=1,3,\ldots}^{m-1} H(n) \left| \frac{I_{m+1}I_n}{I_{n+1}I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_n} + \frac{I_m}{I_{n-1}} \right| \\
+ \left| I_m \frac{1}{I_0} f(0) + I_{m+1} \left\{ \frac{1}{I_0} f(0) \\
- \frac{e^{2\lambda}}{\lambda I_0 I_1} \mathbb{P}(W \leq \min\{0, x\}) \mathbb{P}(W > \max\{0, x\}) \right\} \right|,
\]

if \( m \) is even, and

\[
|f(m+1) - f(m)| \leq \frac{\mathbb{P}(W \leq x)}{\lambda} \left\{ \sum_{n=1,3,\ldots}^{m-1} \frac{I_{m+1}I_n}{I_{n+2}} - \frac{I_m}{I_{n-1}} \right\} \\
+ \sum_{n=1,3,\ldots}^{m-1} H(n) \left| \frac{I_{m+1}I_n}{I_{n+1}I_{n+2}} - \frac{I_m}{I_{n+1}} - \frac{I_{m+1}}{I_n} + \frac{I_m}{I_{n-1}} \right| \\
+ \left| I_m \frac{1}{I_0} f(0) + I_{m+1} \left\{ \frac{1}{I_0} f(0) \\
- \frac{e^{2\lambda}}{\lambda I_0 I_1} \mathbb{P}(W \leq \min\{0, x\}) \mathbb{P}(W > \max\{0, x\}) \right\} \right| \\
+ |H(m+1) - H(m)|,
\]

if \( m \) is odd. Here \( H(n) = \mathbb{P}(W > n) / \mathbb{P}(W = n) \) is the inverse of the hazard function of the Skellam distribution (and is not to be confused with our use of \( H \) in the main body of the paper as a subgraph of the graph \( G \)).

Note that both (6.9) and (6.10) consist of \( \mathbb{P}(W \leq x) / \lambda \) times a sum of distinct terms. Furthermore, note that there are three such terms in (6.9), while (6.10) has the same three augmented by the addition of a fourth, i.e., \( |H(m+1) - H(m)| \). Since the multiplicative factor \( \mathbb{P}(W \leq x) / \lambda \) is trivially bounded by \( 1/\lambda \), it remains for us to
show that each of the terms it multiplies are $O(1)$ and to characterize the constant in each.

Through a series of arguments (the result for each of which is presented as a separate proposition in the Supplementary Materials), we are able to show the following. First, that

$$\sup_{m \in \mathbb{N}^+} \left| \sum_{n=1,3,...}^{m-1} \frac{I_{m+1}}{I_{n+2}} - \frac{I_m}{I_{n+1}} \right| \leq 3 \quad (6.11)$$

Next that,

$$\sum_{n=1,3,...}^{m-1} H(n) \left| \frac{I_{m+1}I_n}{I_{n+1}I_{n+2}} - \frac{I_m}{I_{n+1}} + \frac{I_{m+1}}{I_{n-1}} \right| \leq 73 \quad (6.12)$$

for $\lambda \geq 1$. And furthermore, that

$$\left| \frac{1}{I_0} f(0) + I_{m+1} \left\{ \frac{1}{I_0} f(0) - \frac{e^{2\lambda}}{\lambda} \frac{1}{I_0I_1} \mathbb{P}(W \leq \min\{0,x\}) \mathbb{P}(W > \max\{0,x\}) \right\} \right| \in O\left(\frac{1}{\lambda}\right),$$

indicating that this term is negligible compared to the others, for sufficiently large $\lambda$. Finally, it is clear that

$$H(m) - H(m+1) = \frac{\mathbb{P}(W > m)}{\mathbb{P}(W = m)} - \frac{\mathbb{P}(W > m + 1)}{\mathbb{P}(W = m + 1)}$$

$$= \frac{1}{\mathbb{P}(W = m)} [\mathbb{P}(W > m) - \mathbb{P}(W > m + 1)]$$

$$= \frac{\mathbb{P}(W = m + 1)}{\mathbb{P}(W = m)}$$

$$\leq 1$$

and, at the same time $H(m) - H(m+1) \geq 0$ so, we have that we may bound the magnitude of the fourth and final term by 1.

As a result of all of the above, we may conclude that

$$|f(m+1) - f(m)| \in O\left(\frac{78}{\lambda}\right) \quad (6.13)$$

for $m > 0$. Or, equivalently, we may express the right-hand side above as $O(156/2\lambda)$.

The argument for the case of $m < 0$ involves similar reasoning, as described in the Supplementary Materials.

\[\square\]

**Acknowledgements**

We thank Kostas Spiliopoulos and Tasso Kaper for helpful discussions at certain stages of this work.
References


