# Graphon Mixtures

**Sevvandi Kandanaarachchi**
CSIRO's Data61, Melbourne, Australia

**Cheng Soon Ong**
CSIRO's Data61, Canberra, Australia

## Abstract

Social networks have a small number of large hubs, and a large number of small dense communities. We propose a generative model that captures both hub and dense structures. Based on recent results about graphons on line graphs, our model is a graphon mixture, enabling us to generate sequences of graphs where each graph is a combination of sparse and dense graphs. We propose a new condition on sparse graphs (the max-degree), which enables us to identify hubs. We show theoretically that we can estimate the normalized degree of the hubs, as well as estimate the graphon corresponding to sparse components of graph mixtures. We illustrate our approach on synthetic data and real-world networks, showing the benefits of explicitly modeling sparse graphs.

## 1 INTRODUCTION

Edge density is a graph attribute that helps to characterize graph sequences into dense or sparse graphs. A graph sequence is called dense if the edges grow quadratically with the nodes, and it is called sparse if the edges grow sub-quadratically with the nodes. A characteristic of dense graphs is that nodes are more connected to each other. A characteristic of sparse graphs is that they are less connected. Simple examples of sparse graphs include stars, paths and rings.

Observed sparse graphs such as social networks exhibit two contrasting types of behavior: a small number of high degree nodes called *hubs* inducing sparsity and a large number of small, dense communities (Zang et al., 2018; Zhao et al., 2021). Such hubs and communities are also observed in complex multi-functional networks such as neurological networks (Van den Heuvel and Sporns, 2013; Schwarz et al., 2008). A hub, viewed in isolation can be thought of as a star graph and a community by itself can be thought of as a dense graph. The prevalence of these two components in both biological and social networks is evidence that they are fundamental structures in evolving graphs.

Motivated by these considerations we consider graph mixtures – a combination of two graphs, one sparse and one dense generated from two different types of graph generators: graphons. A graphon is a symmetric and measurable function $W : [0, 1]^2 \to [0, 1]$. Graphons are used to learn the underlying structure or representation of graphs (Xu et al., 2021) and recently graphons are used in neural operators and graph neural networks (Levie, 2023; Tieu et al., 2024; Cheng and Peng, 2023). A graphon can be thought of as a graph blueprint and can be used to generate graphs of an arbitrarily large size. As a consequence of the Aldous-Hoover theorem (Aldous, 1981), graphs generated from a graphon $W$ are dense. Thus, while it straightforward to generate dense graphs from a graphon $W$, generating sparse graphs from a graphon is not quite that simple.

There have been several extensions to graphons to model sparse graphs. Caron et al. (2023) model graphs as exchangeable point processes and extend the classical graphon framework to the sparse regime. Their graphons are defined on $\mathbb{R}_+^2$ instead of on the unit square $[0, 1]^2$. Research led by Borgs and Chayes tackle sparsity in different ways. For example, Borgs et al. (2018) consider 'stretched' and 'rescaled' graphons that can represent sparse graphs. Veitch and Roy (2015) introduce graphexes, a triple describing isolated edges, infinite stars and a graphon defined on $\mathbb{R}_+^2$. Kandanaarachchi and Ong (2024) model sparse graphs by considering graphons of line graphs. We use results from Kandanaarachchi and Ong (2024) and Janson (2016) to define graphon mixtures, which can generate sparse and dense graphs depending on the mixture properties.

Our approach for sparse graph generation considers line graphs. Line graphs, also known as non-backtracking graphs (Krzakala et al., 2013) are obtained by mapping

---

Proceedings of the 29[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. 

edges to vertices. A line graph of a graph $G_n$ denoted by $L(G_n)$ maps edges of $G_n$ to vertices of $L(G_n)$ and connects vertices in $L(G_n)$ if the corresponding edges in $G_n$ share a vertex (Figure 1a). Kandanaarachchi and Ong (2024) showed that for a subset of sparse graphs, which they called *square-degree*, the line graphs are dense. Hence, for square-degree graphs, by taking line graphs, we are back in the space of dense graphs which we know how to model using the standard graphon. However, this hardly tells us anything about the structure of the line graph graphon. Janson (2016) showed that line graph limits are disjoint clique graphs and further showed that disjoint clique graphs can be modelled by a sequence of positive numbers, which he called a *mass-partition*. This result combined with the inverse line graph operation enable us to define graphon mixtures.

The contributions of this paper are: (1) we propose graphon mixtures – a novel approach that explicitly models sparsity and thereby generates graph mixtures; (2) given a graph mixture, we estimate the degree of hubs of unseen graphs and the graphon corresponding to the sparse components, and we provide theoretical guarantees on our estimates that have exponential or polynomial convergence; (3) we show empirically that our approach is useful on synthetic and real data. All proofs are presented in the Appendix.

# 2 LINE GRAPH LIMITS AND SPARSITY

We review the setting of graphons, in particular with respect to the recent results on graphons on line graphs (Kandanaarachchi and Ong, 2024).

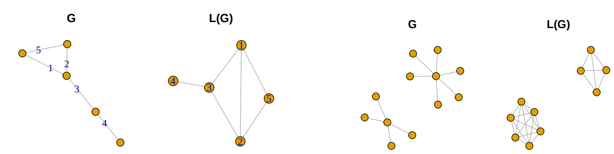
**Definition 2.1 (Dense and sparse graph sequences).** *A sequence of graphs $\{G_n\}_n$ is dense if the number of edges $m$ grow quadratically with the number of nodes $n$, i.e., $\liminf_{n\to\infty} \frac{m}{n^2} = c > 0$. A sequence of graphs $\{G_n\}_n$ is sparse if the number of edges $m$ grow sub-quadratically with the number of nodes $n$, i.e., $\lim_{n\to\infty} \frac{m}{n^2} = 0$.*

A graphon is a symmetric and measurable function $W : [0,1]^2 \to [0,1]$. Graphs and graphons are linked by the adjacency matrix. The *empirical graphon* (Definition A.2) of a graph is obtained by scaling the adjacency matrix to the unit square and coloring smaller squares of size $1/n \times 1/n$ black if the corresponding entry in the adjacency matrix is 1 and coloring it white if the entry is 0. Given a graphon $W$, a graph with $n$ nodes for any $n \in \mathbb{N}$ can be generated as follows:

**Definition 2.2.** *Uniformly pick $x_1, x_2, \ldots x_n$ from $[0,1]$. A **W-random graph** $\mathbb{G}(n, W)$ has the vertex set $1, 2, \ldots n$ and vertices $i$ and $j$ are connected with probability $W(x_i, x_j)$.*

Given a sequence of graphs $\{G_n\}_n$, convergence is defined using the *cut norm* (Definition A.3) and the *cut metric* (Definition A.4). Borgs et al. (2008) showed that every convergent graph sequence converges to a graphon $W$. Therefore, graphons are graph limits. The problem is that all sparse graph sequences converge to $W = 0$. Hence, the standard graphon is not representative of sparse graphs, in the sense that graphs generated from $W = 0$ are isolated nodes without any edges.

## 2.1 Sparse graphs are stars in the limit




(a) (b)

Figure 1: (a) A graph $G$ and its line graph $L(G)$. Vertices in $L(G)$ are edges of $G$. Vertices of $L(G)$ are connected if the corresponding edges in $G$ share a vertex. (b) Line graphs of disjoint stars are disjoint cliques. Thus, the inverse line graph of disjoint cliques are disjoint stars.

To address this problem, more mathematically intricate constructions and convergences have been defined (Borgs et al., 2018; Caron et al., 2023). Kandanaarachchi and Ong (2024) showed that if sparse graphs satisfy a certain condition, then their line graphs (Definition A.1, Figure 1a) are dense. They called this condition the *square-degree property* (Definition A.9). The set of graph sequences satisfying the square-degree property is denoted by $S_q$. They showed that if $\{G_n\}_n \in S_q$ and the line graphs $\{L(G_n)\}_n$ converge, then they converge to a graphon $U \neq 0$. Here, we elucidate the structure of $U$, based on an earlier result by Janson (2016).

**Theorem 2.3. (Janson (2016) Thm 8.3)** *A graph limit is a line graph limit if and only if it is a disjoint clique graph limit.*

A clique is a subset of nodes in the graph that are connected to each other. This result is important because as shown in Figure 1b, the inverse line graph of a disjoint clique graph is a disjoint star graph, and stars are the most sparse of the graphs that exhibit power-law degrees. Therefore, the line graph limit explains sparsity in a fundamental way via the inverse line graph operation. It is worth noting that if a graph is a line graph, then its inverse exists and is unique (Whitney,

1932) with the only exception when the line graph is a triangle. Janson (2016) represented disjoint clique graph limits by a sequence of proportions called the *mass-partition* (Definition A.5) where the $j$th element of the mass-partition gives the proportion of nodes in the $j$th disjoint clique. A mass-partition $\boldsymbol{p} = \{p_i\}_{i=1}^{\infty}$ uniquely defines a disjoint clique graphon $W_{\boldsymbol{p}}^{\mathcal{M}}$ (Definition A.6) resembling a block-diagonal matrix (see $U$ in Figure 2). We combine disjoint clique graph limits of Janson (2016) with converging sparse square-degree graphs $S_q$ (Definition A.9) of Kandanaarachchi and Ong (2024) in the following lemma. Furthermore, we provide a new definition (so called max-degree) which is equivalent to square-degree property.

**Lemma 2.4.** *Let $\{G_n\}_n \in S_q$ (Definition A.9) and let $H_m = L(G_n)$ where $L$ denotes the line graph operation. If $\{H_m\}_m$ converges to $U$ then $U$ is a disjoint clique graphon and there exists a unique mass-partition $\boldsymbol{p}$ that describes $U$. That is, $U = W_{\boldsymbol{p}}^{\mathcal{M}}$ (see Definition A.6) for some mass-partition $\boldsymbol{p}$.*

*Proof.* Immediate from Theorem 2.3 and A.8 (Janson, 2016). □

**Definition 2.5. (Max-degree condition)** *Let $\{G_n\}_n$ be a sequence of graphs where $G_n$ has $n$ nodes and $m$ edges. Let the maximum degree of $G_n$ be denoted by $d_{\max,n}$. We say the $\{G_n\}_n$ satisfies the max-degree condition if there exists $c > 0$*

$$\liminf_{n,m\to\infty} \frac{d_{\max,n}}{m} = c > 0 \, .$$

*Let $S_x$ denote the set of graph sequences satisfying the max-degree condition.*

We show that for graph sequences with converging line graphs, the square-degree property is equivalent to the max-degree condition

**Lemma 2.6.** *Let $\{G_n\}_n$ be a graph sequence with $H_m = L(G_n)$. Suppose $\{H_m\}_m$ converges to $U$. Then*

$$\{G_n\}_n \in S_q \equiv \{G_n\}_n \in S_x \, .$$

### 2.2 Related work

Sparse graphs including scale-free and power-law networks have been widely studied. In the Barabási–Albert model (Barabási and Albert, 1999), the degree distribution satisfies $P(d) \propto d^{-\gamma}$, where $\gamma \approx 3$, and Bollobás et al. (2001) showed the maximum degree grows as $\Theta(\sqrt{n})$, where $n$ is the number of nodes. Kronecker graphs (Leskovec et al., 2010) provide another generative mechanism, producing heavy-tailed degree distributions through deterministic or stochastic Kronecker products.

Classical graphons were used to describe dense graphs. Over time the original definitions, constructions and convergences were extended to incorporate sparse graphs (Borgs et al., 2021). Caron and Fox (2017) and Caron et al. (2023) modeled graphs as exchangeable point processes where the graphon $W$ is defined on $\mathbb{R}_+^2$. These were also known as graphex processes. They used Kallenberg exchangeability in their theoretical framework and sparsity was achieved by extending the support of the graphon from a finite subset to the whole positive quadrant. In machine learning, recent work used graphon interpolation in a dense setting aiding graph augmentation and classification (Han et al., 2022; Azizpour et al., 2025). Their graphons represent dense graphs and do not generalize to sparse graphs.

Our approach is different to the above as we use inverse line graphs of disjoint clique graphs to generate the sparse component. As both graphons in our mixture are defined on the unit square, norms are well behaved making it easier to estimate these graphons in practice, especially the sparse component. By varying mixture properties we can generate both dense and sparse graphs.

## 3 MIXTURES OF SPARSE AND DENSE GRAPHS

We generate a graph mixture by picking a graphon $W$ and a disjoint clique graphon $U$ (see Figure 2). First we generate two graph sequences using $U$ and $W$, noting that both graph sequences are dense, but the graph sequence generated by $U$ lives in the line graph space. The sparse graph sequence is obtained by taking the inverse line graphs of the graphs generated by $U$. Then we join the $i$th graph generated by $W$ with the $i$th inverse line graph generated by $U$ according to a set of simple joining rules and obtain the graph mixture. The results in Sections 4, 5 and 6 do not have any terms pertaining to the join. They hold if we consider the disjoint union of the two graph sequences. However, the real graphs in Table 2 are not disjoint unions. As such, we consider a simple way to join nodes randomly without adding a lot of extra edges. We ensure the main structural parts are contributed by $U$ and $W$, but not the join.

A key component in the mixing process is the ratio of the number of nodes in the sparse graph to that of the dense graph. If this ratio goes to infinity along the sequence, then the mixture sequence is sparse. If the ratio is bounded, then the mixture sequence is dense. Thus, for a given $W$ and $U$ by changing the mixture ratio function, we can generate graph sequences ranging from dense to sparse.

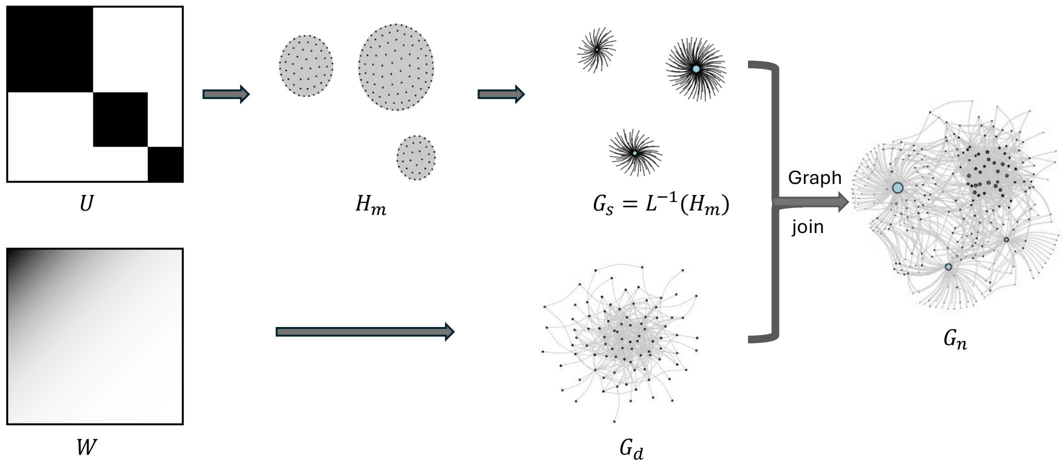



Figure 2: The overview of $(U, W)$-mixture graphs. A disjoint clique graph $H_m$ is sampled from graphon $U$. The inverse line graph $G_s = L^{-1}(H_m)$ is a disjoint set of stars. A graph $G_d$ is sampled from graphon $W$. Then, $G_s$ and $G_d$ are joined according to graph joining rules resulting in the mixture graph $G_n$.

### 3.1 The graph mixture model

A disjoint clique graphon $U$ is defined by a sequence of proportions (the mass-partition) $\boldsymbol{p} = \{p_i\}_{i=1}^{\infty}$. Janson (2016) considered $\sum_i p_i \leq 1$. In our work we consider a rescaled version of the original mass-partition and let the elements of the mass-partition add up to 1. We explore the inverse line graphs of graphs generated by $U$ when $\sum_i p_i < 1$ in Appendix B.6.

**Definition 3.1.** ***(Mass-partition)*** *We define a mass-partition to be a sequence $\boldsymbol{p} = \{p_i\}_{i=1}^{\infty}$ of non-negative real numbers such that*

$$p_1 \geq p_2 \geq \cdots \geq p_i \geq \cdots \quad \text{and} \quad \sum_{i=1}^{\infty} p_i = 1\,.$$

*We refer to the number of non-zero elements in $\boldsymbol{p}$ as partitions. Thus $U$ corresponding to $\boldsymbol{p}$ can have finite or infinite partitions.*

**Definition 3.2.** *Given a graphon $W$, a disjoint clique graphon $U$, and positive, integer valued sequences $\{n_{d_i}\}_i$ and $\{m_{s_i}\}_i$ increasing with $i$, a **$(U, W)$-mixture graph sequence** $G_{n_i} \sim \mathbb{G}(U, W, n_{d_i}, m_{s_i})$ is constructed as follows:*

1. *Let $G_{d_i} \sim \mathbb{G}(n_{d_i}, W)$ and $H_{s_i} \sim \mathbb{G}(m_{s_i}, U)$, i.e., $G_{d_i}$ is a $W$-random graph (Definition 2.2) and $H_{s_i}$ is a $U$-random graph. Sequences $\{G_{d_i}\}_i$ and $\{H_{s_i}\}_i$ consist of dense graphs.*
2. *As $U$ is a disjoint clique graphon, the inverse line graph of $H_{s_i}$ exists and is a disjoint union of stars. Suppose $G_{s_i} = L^{-1}(H_{s_i})$ has $n_{s_i}$ nodes where $L^{-1}$ denotes the inverse line graph operation. Thus, $\{G_{s_i}\}_i$ is a sparse graph sequence.*
3. *Join $\{G_{d_i}\}_i$ and $\{G_{s_i}\}_i$ satisfying graph joining rules in Definition 3.3. Let $\{G_{n_i}\}_i$ denote the resulting graph sequence.*

*We say $G_{n_i} \sim \mathbb{G}(U, W, n_{d_i}, m_{s_i})$ is a $(U, W)$-mixture graph. We call $G_{d_i}$ the **dense part** and $G_{s_i}$ the **sparse part** of $G_{n_i}$.*

Given two graph sequences – one dense and one sparse – we consider a simple way to randomly join a dense graph and a sparse graph with a given number of edges, making sure the main structural parts are contributed by the two graphs but not the join. We formally define the joining rules below.

**Definition 3.3.** ***(Graph Joining Rules)*** *Let $\{G_{s_i}\}_i$ and $\{G_{d_i}\}_i$ be two graph sequences. We are interested in joining $G_{s_i}$ and $G_{d_i}$ for every $i$. Let $G_{n_i}$ denote the joined graph. We consider graph joins that satisfy the following conditions.*

1. ***No new nodes:*** *no new nodes are added as part of the joining process.*
2. ***No deletion:*** *No nodes or edges are deleted as part of the joining process.*
3. ***Random edges:*** *if edges are added, they are added randomly in the sense that nodes within a graph are equally likely to be selected to form edges.*
4. ***New edges:*** *if edges are added, the number of new edges added $m_{new_i}$ satisfies $m_{new_i} = cm_{d_i}$, for small $c \in \mathbb{R}$, where $m_{d_i}$ denotes the number of edges in $G_{d_i}$.*

In defining these rules our motivation is to keep the joining process as simple as possible. As such, we have employed a random join. These rules can be further explored and modified to suit different needs that result in more targeted mixtures.

### 3.2 Expectations and edge density

In graphon mixtures, randomness arise in two ways: (1) when the graphs are generated from $U$ and $W$ and (2) when the graphs are joined according to graph joining rules. We compute expectations of hub nodes and nodes in the dense part with respect to both sources of randomness in Lemmas B.1 and B.2. Depending on the mixing ratio $\frac{n_{s_i}}{n_{d_i}}$ the $(U, W)$ mixture graphs can be sparse or dense as shown in Lemma 3.4.

**Lemma 3.4.** *Let $\{G_{n_i}\}_i$ be a sequence of $(U, W)$-mixture graphs (Definition 3.2) with $G_{n_i} \sim \mathbb{G}(U, W, n_{d_i}, m_{s_i})$. Let $G_{d_i}$ be the dense part of $G_{n_i}$ and let $G_{s_i}$ be the sparse part. Let $n_{d_i}$ and $n_{s_i}$ be the number of nodes in $G_{d_i}$ and $G_{s_i}$ respectively. Then*

1. *If there exists $c \in \mathbb{R}^+$ such that $\limsup_{i\to\infty} \frac{n_{s_i}}{n_{d_i}} = c$ then $\{G_{n_i}\}_i$ is dense.*
2. *If $\lim_{i\to\infty} \frac{n_{s_i}}{n_{d_i}} = \infty$, then $\{G_{n_i}\}_i$ is sparse.*

Lemma 3.4 shows that $(U, W)$-mixture graphs are a general construction that allows us to generate sparse or dense graphs. While our focus is on sparse graphs, dense graphs that cannot be fully explained by a single graphon $W$ can be generated by a $(U, W)$ mixture (see Appendix B.2). Thus, $(U, W)$-mixture graphs can model a richer set of dense graphs compared to a single graphon $W$. We give examples of $(U, W)$-mixture graphs in Appendix B and show that by changing mixture properties we get graph sequences ranging from dense to sparse using the same $(U, W)$ combination.

## 4 TOP-$k$ DEGREES OF SPARSE $(U, W)$ MIXTURE GRAPHS

In a network the highest-degree nodes (the hubs) often represent the most influential or central entities of the network. In social networks they maybe key individuals, in transportation networks they can represent critical infrastructure and in a neurological context hub overload and failure can explain neurological disorders (Stam, 2014). Thus modeling high-degree nodes is important. By observing a mixture graph without knowing $U$, $W$ or the mixture ratio function, we predict the high-degrees of an unseen graph, by knowing only the number of nodes in that graph.

For a sparse $(U, W)$ mixture sequence, we show that the high-degree nodes are contributed by $U$ (Lemma C.3)). Recall that graphon $U$ is in the line graph space, and by Theorem 2.3 is formed by disjoint cliques. Hence the inverse line graphs are disjoint stars. As $U$ is described by a mass-partition $\boldsymbol{p} = (p_1, p_2, \ldots)$, if $p_j > p_k$ the degree of the star corresponding to $p_j$ is larger than that of $p_k$ even after joining the sparse and dense parts (Lemma C.1). Furthermore, the joining process does not change the order of the highest degrees as stated below.

**Proposition 4.1.** ***(Order Preserving Property)*** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) associated with $U$ which has at least $k$ partitions. Let $\tilde{q}_{j,i}$ be the degree of the star in $G_{s_i}$ corresponding to $p_j \neq 0$. Let $q_{j,i}$ denote the degree of the corresponding vertex in $G_{n_i}$. Let $deg_{\,G_{n_i}} v_{(r)}$ denote the $r$th highest degree in $G_{n_i}$. Then*

$$P\left(\bigcap_{j=1}^{k}\left(q_{j,i} = deg_{\,G_{n_i}} v_{(j)}\right)\right) \geq \left(1 - \frac{c_1}{m_{s_i}}\right)^k \times \left(1 - \exp\left(-c_2 \frac{m_{s_i}^2}{n_{d_i}^2}\right) - \exp\left(-c_3 m_{s_i}\right)\right)$$

*That is, with high probability the order of the stars in the sparse part are preserved by joining.*

The Order Preserving Property helps us to estimate the top-$k$ degrees of graphs when mass-partitions have at least $k$ non-zero entries.

**Lemma 4.2.** *Suppose $G_{n_i}$ and $G_{n_j}$ are two graphs from a sparse $(U, W)$-mixture graph sequence (Definition 3.2). We treat $G_{n_i}$ as the training graph and $G_{n_j}$ as the test graph. Suppose $G_{n_i}$ and $G_{n_j}$ have $n_i$ and $n_j$ nodes respectively. Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) associated with $U$, which has at least $k$ partitions. Let $deg_{\,G_{n_i}} v_{(\ell)}$ denote the $\ell$th largest degree in $G_{n_i}$ where $\ell \leq k$. Then we estimate the $\ell$th largest degree in $G_{n_j}$ as*

$$deg_{\,G_{n_j}} \hat{v}_{(\ell)} = deg_{\,G_{n_i}} v_{(\ell)} \times \frac{n_j}{n_i}\,, \tag{1}$$

*which satisfies*

$$\frac{\left|deg_{\,G_{n_j}} \hat{v}_{(\ell)} - \mathbb{E}(q_{\ell,j})\right|}{m_{s_j}} \leq c p_\ell \left|1 - \frac{n_j m_{s_i}}{n_i m_{s_j}}\right| \tag{2}$$

*with high probability, where $q_{\ell,j}$ denotes the degree of the hub vertex corresponding to $p_\ell \neq 0$ in $G_{n_j}$ and $m_{s_i}$ and $m_{s_j}$ denote the number of edges in the sparse parts $G_{s_i}$ and $G_{s_j}$.*

We use equation (1) to predict degrees and thus model highest-degree nodes. Equation (2) gives an error bound.

# 5 ESTIMATING FINITE PARTITION $U$

Graphon $U$ describes the relative strength (or vulnerability) of the hubs in a mixture. In a social media network, hubs act as superspreaders of information and reach; $U$ shows the relative power of the big players in the network. In an energy/telecommunications network, it can show the vulnerability/resilience in the event of an attack. Consider two energy networks $A$ and $B$ with $U_A = (0.5, 0.3, 0.2)$ and $U_B = (0.1, \ldots, 0.1)$. Then mixture $A$ is far more vulnerable from a resilience viewpoint compared to mixture $B$ if attacked. By attacking one node (the hub corresponding to 0.5), approximately 50% of the network can be crippled in an energy/telecommunications network scenario.

Graphon $U$ can have finite or infinite partitions. By observing mixture graphs, we estimate the mass-partition $\boldsymbol{p}$ (Definition 3.1), both when $\boldsymbol{p}$ has finite and infinite partitions. In this section we consider finite partitions. For both cases, we consider a class of well-behaved $W$ in our mixture model. We do not need $W$ to be well-behaved to estimate the topmost degrees of $G_{n_i}$. However, we need some conditions on $W$ so that we can estimate $U$.

**Definition 5.1.** *(**Degree Function**) As in Delmas et al. (2021) we define the degree function*

$$D(x) = \int_0^1 W(x, y)\, dy \,,$$

*where $D(x)$ is defined on $x \in [0, 1]$.*

**Assumption 5.2.** *We assume $W$ to have a continuous degree function $D(x)$ (Definition 5.1).*

When $W$ satisfies Assumption 5.2, the degrees in the dense part $G_{d_i}$ are closely packed, i.e., given a degree value, there is another degree value quite close in $G_{d_i}$ (see Lemma D.4). Using Assumption 5.2 we estimate $k$, the number of partitions.

**Proposition 5.3.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ having nodes $n_{d_i}$ and $n_{s_i}$ respectively. Suppose $W$ satisfies Assumption 5.2 and $W \neq 1$. Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) associated with $U$ which has only $k$ partitions. Let $deg_{\,G_{n_i}} v_{(\ell)}$ denote the $\ell$th largest degree in $G_{n_i}$. Then there exists $I_0$ such that for $i > I_0$ we have*

$$k = \max_{\ell} \left( \log deg_{\,G_{n_i}} v_{(\ell)} - \log deg_{\,G_{n_i}} v_{(\ell+1)} \right) \,, \quad (3)$$

*where we exclude small degrees. That is, in the log scale the difference between successive top $k$ degrees is largest at $k$.*

**Remark 5.4.** *Equation (3) can be used to estimate $k$ when $k$ is finite and $i$ is large. Procedure 6.1 extends $k$ estimation for infinite $k$, which also works when $k$ is finite and on small graphs.*

**Remark 5.5.** *If Assumption 5.2 is not met, then we can still estimate the number of partitions $k$ using Proposition 5.3. The difference is that if the degree function $D(x)$ is discontinuous, the ratio of successive degrees in the dense part converge to a constant $c > 1$, whereas when it is continuous it converges to $c = 1$. In contrast, the ratios of the top $k$ successive degrees in the sparse part keeps increasing.*

Once we have found the partitions $k$, we can estimate the mass-partition $\boldsymbol{p} = (p_1, \ldots, p_k, 0, 0, \ldots)$.

**Proposition 5.6.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) associated with $U$ with only $k$ partitions. Let $q_{j,i}$ denote the degree of the hub vertex corresponding to $p_j \neq 0$ in $G_{n_i}$. Then using the second order approximation of the Taylor expansion for the expectation and the first order approximation for the variance, with $Q_{k,i} = \sum_{\ell=1}^{k} q_{\ell,i}$ and for some constants $c_1$ and $c_2$ we have*

$$\left| \mathbb{E}\left( \frac{q_{j,i}}{Q_{k,i}} \right) - p_j \right| \leq \frac{c_1}{m_{s_i}} \,, \; Var\left( \frac{q_{j,i}}{Q_{k,i}} \right) \leq \frac{c_2}{m_{s_i}} \,.$$

Therefore, when $U$ has finite partitions, first we estimate $k$ using equation (3). Then we estimate the mass-partition $\boldsymbol{p} = (p_1, p_2, \ldots, p_k, 0, \ldots)$ by

$$\hat{p}_j = \frac{\deg_{\,G_{n_i}} v_{(j)}}{\sum_{\ell=1}^{\hat{k}} \deg_{\,G_{n_i}} v_{(\ell)}} \,, \quad (4)$$

where $\deg_{\,G_{n_i}} v_{(\ell)}$ denotes the observed $\ell$th highest degree in $G_{n_i}$. From the Order Preserving Property (Proposition 4.1) we know that for large $i$, $\deg_{\,G_{n_i}} v_{(\ell)} = q_{\ell,i}$, and thus the bounds in Proposition 5.6 are satisfied.

# 6 ESTIMATING INFINITE PARTITION $U$

When $U$ has infinite partitions, Proposition 5.3 does not hold. In the finite partition case, there is a widening gap between the hub degrees generated by $U$ and the degrees generated by $W$. But in the infinite partition case, there are always smaller values $p_i$ that fill this gap. However, we show that the degrees generated by $U$ are different to those generated by $W$. Let $k_i$ denote the number of hub nodes in $G_{n_i}$ generated by $U$ with degrees higher than those generated by $W$. Our

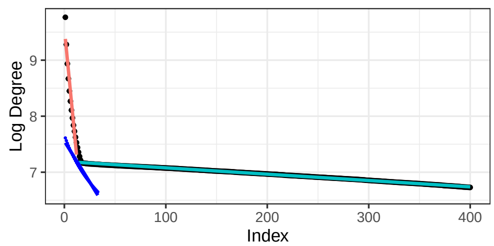



Figure 3: Unique log degree values of a $(U, W)$ mixture graph $G_{n_i}$. The line fitted to points $\{(j, \log(\deg_{G_{n_i}} v_{(j)}))\}_{j=1}^{k_i}$ is shown in red and the line fitted to degrees generated from $W$ is shown in green. The line fitted to points $\{(j, \log \frac{m_{s_i}}{j})\}_{j=1}^{N}$ is shown in blue.

goal is to estimate $k_i$. Thus, we focus on the larger degrees of $G_{n_i}$ and consider the unique degree values of $G_{n_i}$ greater than some percentile $C$. We consider the ordered, unique log degree values $\log(\deg_{G_{n_i}} v_{(j)})$, where $\deg_{G_{n_i}} v_{(j)}$ is the $j$th highest degree in $G_{n_i}$. To the points $(j, \log(\deg_{G_{n_i}} v_{(j)}))$, we fit two lines using the OLS estimator. We show that the line fitted to the first $k_i$ points $\{(j, \log(\deg_{G_{n_i}} v_{(j)}))\}_{j=1}^{k_i}$ has a steeper slope compared to the line fitted to the rest of the points, which are generated by $W$. In Figure 3 these two lines are shown in red and green respectively. We do this by comparing with the line fitted to points on the curve $y = \log(m_{s_i}/x)$, which is shown in blue in Figure 3.

We do this in two parts. First we show that the red line has a steeper slope than the blue line. The key observation is that the series $\sum_j \frac{1}{j}$ diverges to infinity, and as such cannot represent a mass-partition, implying that a mass-partition $\{p_i\}_{i=1}^{\infty}$ has to converge to zero more steeply (or faster) than $\{1/j\}_{j=1}^{\infty}$. This is shown in Lemma E.1. Next we show that the green line is much less steep than the blue line. As $W$ has a continuous degree function $D(x)$ (Assumption 5.2) the sorted, unique degree values are close to each other, and for large $i$, they are mostly consecutive integers. Using this we show that the green line has a less steep slope than the blue line. This is shown in Lemma E.2.

**Procedure 6.1.** *Let $G_{n_i}$ be a graph from a sparse $(U, W)$-mixture graph sequence with $U$ having infinite partitions. Let $k_i$ denote the number of hub nodes in $G_{n_i}$ generated by $U$ with degrees higher than those generated by $W$. We consider the points $(j, \log(deg_{G_{n_i}} v_{(j)})) \in \mathbb{R}^2$ for unique degree values greater than some percentile $C$. Suppose there are $N$ such points. For different cutoff points $r$ we fit two lines – one for $\{(j, \log(deg_{G_{n_i}} v_{(j)}))\}_{j=1}^{r}$ and another for $\{(j, \log(deg_{G_{n_i}} v_{(j)}))\}_{j=r+1}^{N}$ using OLS regression. Suppose the $L_2$ loss functions for the first and second fitted lines are given by $\mathcal{L}_1(r)$ and $\mathcal{L}_2(N-r)$ respectively. Then we estimate $\hat{k}_i$ as*

$$\hat{k}_i = \arg\min_r \left(\mathcal{L}_1(r) + \mathcal{L}_2(N-r)\right) . \tag{5}$$

*That is, $\hat{k}_i$ is the number of points on the first fitted line segment when the sum of the loss functions is minimized.*

Lemma E.3 shows that the sequence $\{\hat{k}_i\}_i$ tends to infinity with high probability. This allows us to show that given $\hat{k}_i$ the ratio

$$\hat{p}_{j,i} = \frac{\deg_{G_{n_i}} v_{(j)}}{\sum_{\ell=1}^{\hat{k}_i} \deg_{G_{n_i}} v_{(j)}} \tag{6}$$

converges to $p_j$ in expectation. The proof is similar to that of Proposition 5.6.

**Proposition 6.2.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Suppose $U$ has infinite partitions and let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the associated mass-partition (Definition 3.1). Let $q_{j,i}$ denote the degree of the hub vertex corresponding to $p_j$ in $G_{n_i}$ and $\hat{k}_i$ be estimated using Procedure 6.1. Then using the second order approximation of the Taylor expansion for the expectation and the first order approximation for the variance, with $Q_{\hat{k},i} = \sum_{\ell=1}^{\hat{k}_i} q_{\ell,i}$ and for some constants $c_1$ and $c_2$ we have*

$$\left| \mathbb{E}\left(\frac{q_{j,i}}{Q_{\hat{k},i}}\right) - p_j \right| \leq \frac{c_1}{m_{s_i}} , \; Var\left(\frac{q_{j,i}}{Q_{\hat{k},i}}\right) \leq \frac{c_2}{m_{s_i}}$$

We use equation (5) to compute $\hat{k}_i$ and equation (6) to compute $\hat{p}_j$ for $j \in \{1, \ldots, \hat{k}_i\}$, which estimates the first $\hat{k}_i$ elements of the mass-partition $\boldsymbol{p}$ describing $U$. Additionally, if we know the rate at which the elements in the mass-partition $\{p_i\}_i$ go to zero, then we can estimate $\sum_{j=\hat{k}_i}^{\infty} p_j$, the sum of the unestimated elements in the partition. This is shown in Lemma E.4.

**Time complexity of estimating $\hat{\boldsymbol{p}}$**

For a graph $G_n$ with $n$ nodes, the degree computation using the adjacency matrix representation is $O(n^2)$ as an $n \times n$ matrix needs to be summed. However, current programs compute the degree distributions much faster. Procedure 6.1 takes the logarithm of unique degrees and fits two lines to the top half of them (say $r$ points). This can be completed in $O(r)$ time. When line fitting is iterated over $r$ points, the time complexity increases to $O(r^2)$. But $r << n$ as the number of nodes with small degrees is large. This process estimates the number of partitions $k$. Once $\hat{k}$ is computed, estimating the mass-partition involves sorting the degrees and dividing by their sum as shown in equation (6), which can be achieved in $O(n)$ time. Thus, the worst case time complexity is $O(n^2)$. Figure 4 shows the time taken to estimate $\hat{\boldsymbol{p}}$ for different values of $n$ illustrating the algorithm generally takes $O(n)$ time.

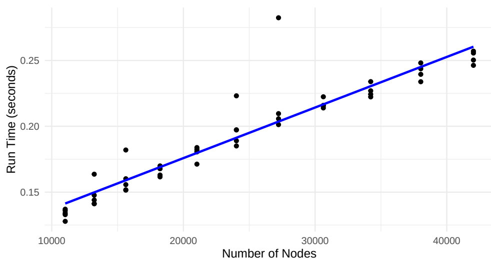



Figure 4: Time taken to estimate $\hat{\boldsymbol{p}}$.

# 7 EXPERIMENTS

We conduct experiments with synthetic and real data on a standard laptop. As the evaluation metric we use the Mean Absolute Percentage Error (MAPE) given by $100 \times \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{y_i}$. For degree prediction, $y_i$ denotes the top-$k$ degree values and for estimating $U$, $y_i$ denotes the mass-partition elements $p_i$. More details of the experiments are given in Appendix F.

## 7.1 Illustration with synthetic data

With synthetic data experiments we focus on 3 tasks: (1) predict the top-$k$ degrees, (2) estimate $U$ when $\boldsymbol{p}$ has finite partitions and (3) estimate $U$ when $\boldsymbol{p}$ has infinite partitions. To predict the top-$k$ degrees we use training graphs with $n_i$ nodes and test graphs with $n_j$ nodes. Using equation (1) we predict the top-$k$ degrees for graphs in the test set. To estimate $U$ we only use training graphs. When $U$ has finite partitions we use equation (4) and for infinite partitions we use equation (6). For each task we conduct 4 experiments. For degree prediction we consider $W_1 = \exp(-(x+y))$, $W_2 = 0.1$ and $U_1$ with mass-partition $\boldsymbol{p}_1 \propto \{1/j^{1.2}\}_{j=2}^{50}$ and $U_2$ with mass-partition $\boldsymbol{p}_2 \propto \{1/1.2^j\}_{j=2}^{50}$. The mass-partitions are normalized so that they add up to 1. A baseline comparison is conducted using scale-free graph properties (Bollobás et al., 2001). Table 1 gives the results of the experiments.

## 7.2 Sensitivity to method of joining

The graph joining rules (Definition 3.3) randomly joins nodes in the dense and sparse parts. However, in social networks nodes might connect to "more famous" nodes. To test this, we join nodes with a probability proportional to the degree of nodes in the sparse part. This makes nodes in the dense part connect with highly connected or famous nodes. We call this the *Pro-join*.

In addition, we test the inverse scenario where nodes in the dense part connect with nodes in the sparse part with a probability inversely proportional to their degree. To purposely skew the $\hat{\boldsymbol{p}}$ estimation, we exclude degree-1 nodes from the joining process, making nodes prefer "less famous" hubs more. We note that this join is unintuitive. It is intended as a "bad join" that can break the estimation of $\hat{\boldsymbol{p}}$. We call this the *Inv-join*.

Table 2 gives the results with Pro-join and Inv-join on estimating $\hat{\boldsymbol{p}}$ . We conducted the same experiments as in Table 1 for $\hat{\boldsymbol{p}}$ estimation. The only difference was that either Pro-join or Inv-join was used to connect the dense and the sparse parts instead of the random join given in the joining rules. We see that Pro-join results are similar or better than the random-join, especially for the infinite $U$ task. The Inv-join produces worse results as expected.

## 7.3 Real-world networks

For observed real world graphs, we do not have a ground truth model for $U$, hence we only compute MAPE for predicted degrees. We use Facebook (FB) links (Viswanath et al., 2009), Hep-PH Physics citations (Leskovec et al., 2005; Gehrke et al., 2003), MOOC interactions (Kumar et al., 2019), SMS (Wu et al., 2010), UCI Messages (Panzarasa et al., 2009) and Yahoo messages (Rossi and Ahmed, 2015) datasets. Each dataset is given as an edgelist with timestamped edges. We use daily graphs for Facebook links, MOOC interactions, SMS, and UCI messages datasets. For Yahoo messages we use 2-hourly graphs and for Hep-PH citations monthly graphs. Each dataset consists of a sequence of growing graphs $\{G_{n_i}\}_i$ and we consider $G_{n_i}$ for $i \in \{20, \ldots, 24\}$ as training graphs. We select the test graphs $G_{n_j}$ such that $j = \min_k\{|G_{n_k}| : |G_{n_k}| - |G_{n_i}| \geq 500\}$, i.e., $G_{n_j}$ is the first graph that has more than 500 nodes compared to $G_{n_i}$. We predict the top-10 degrees of $G_{n_j}$ using Bollobás et al. (2001), Caron and Fox (2017) and 3 variations of Kronecker graphs (Leskovec et al., 2010). Furthermore, we test for statistical significance between the top 2 methods. Table 3 gives the results with blanks denoting timed out instances. The $(U, W)$-mixture estimate is significantly better for 5 out of 6 datasets.

## 7.4 A downstream task

We present a toy example illustrating how $(U, W)$-mixture properties can support downstream tasks in instances where sparsity is important. The estimated sparse component $\hat{\boldsymbol{p}}$, which corresponds to the graphon $U$ can be used as a feature for downstream tasks such

Table 1: Proposed method is an order of magnitude better than baseline on synthetic data. MAPE (smaller is better) reported with standard deviation in parenthesis.

| Task | Description | Experiment 1 | Experiment 2 | Experiment 3 | Experiment 4 |
|---|---|---|---|---|---|
| Top-$k$ deg. | Proposed | 0.386 (0.337) | 0.225 (0.178) | 0.394 (0.288) | 0.341 (0.391) |
| | Baseline | 9.078 (0.436) | 9.098 (0.238) | 9.185 (0.336) | 9.074 (0.488) |
| Finite $U$ | Proposed | 1.800 (1.571) | 0.602 (0.428) | 0.845 (0.644) | 1.621 (1.448) |
| | Baseline | 98.709 (0.030) | 98.702 (0.012) | 97.952 (0.024) | 97.766 (0.025) |
| Infinite $U$ | Proposed | 10.917 (0.098) | 1.559 (0.069) | 6.351 (0.051) | 0.339 (0.171) |
| | Baseline | 97.238 (0.002) | 95.233 (0.003 ) | 95.070 (0.003) | 98.245 (0.003) |

Table 2: Sensitivity analysis using Pro-join and Inv-join in the joining process. MAPE (smaller is better) reported with standard deviation in parenthesis.

| Task | Join | Experiment 1 | Experiment 2 | Experiment 3 | Experiment 4 |
|---|---|---|---|---|---|
| Finite $U$ | Pro-join | 0.738 (0.428) | 0.894 (0.364) | 1.333 (0.358) | 1.388 (0.225 ) |
| | Inv-join | 13.635 (0.456) | 0.734 (0.329) | 0.973 (0.253) | 0.964 (0.250) |
| Infinite $U$ | Pro-join | 0.376 (0.037) | 0.805 (0.093) | 0.441 (0.040) | 0.231 (0.106) |
| | Inv-join | 74.233 (0.049) | 191.228 (9.381) | 167.116 (0.073) | 4887.076 (16.587) |

Table 3: Comparison of proposed method with Bollobás et al. (2001), Caron and Fox (2017) and 3 versions of Kronecker graphs (Leskovec et al., 2010) on 6 datasets. Average MAPE reported with standard deviation in parenthesis. Best results in bold. Sig. denotes statistical significance with $\alpha = 0.1$.

| Dataset | Bollobas | Caron-Fox | Kronfit1 | Kronfit2 | Kronfit3 | UW-Mixture | Sig. |
|---|---|---|---|---|---|---|---|
| FB | 0.130 (0.041) | 0.175 (0.103) | 0.289 (0.068) | 111.0 (14.9) | 0.518 (0.047) | **0.101** (0.044) | ✓ |
| HEP-PH | 0.177 (0.105) | 0.435 (0.066) | 0.351 (0.165) | 20.20 (4.01) | 0.294 (0.115) | **0.061** (0.046) | ✓ |
| MOOC | 0.015 (0.012) | 0.854 (0.014) | 0.772 (0.068) | 0.797 (0.022) | 0.835 (0.031) | **0.005** (0.005) | ✓ |
| SMS | 0.051 (0.033) | 0.426 (0.051) | 0.702 (0.032) | – | 0.747 (0.024) | **0.029** (0.020) | ✓ |
| UCI | 0.201 (0.106) | 0.757 (0.044) | 0.404 (0.098) | 2.19 (0.409) | 0.908 (0.011) | **0.185** (0.090) | ✗ |
| Yahoo | 0.030 (0.037) | 0.446 (0.046) | 0.356 (0.080) | – | 0.834 (0.014) | **0.021** (0.011) | ✓ |

as graph classification as shown below.

Using $(U, W)$-mixtures we generate graphs of 2 classes. We let $U$ for graphs belonging to class A have a mass partition $\boldsymbol{p} = (0.5, 0.3, 0.2)$ and for graphs belonging to class B have a mass partition $\boldsymbol{p} = (0.34, 0.33, 0.33)$. For both mixtures we let $W(x, y) = \exp(-(x+y))$. We generate 100 graphs of each class. Then, we extract the mass-partition for each graph as detailed in Procedure 6.1 and equation (6). This gives us a feature matrix $P$ where each row gives the estimated $\hat{\boldsymbol{p}}$ for each graph. Using 10-fold cross validation we train and test a logistic regression classifier on $P$. We obtain an average cross-validated accuracy of 98.5% with a standard deviation of 2.5% showing that $\hat{\boldsymbol{p}}$ is a discerning feature. A clustering experiment can be done similarly.

## 8 CONCLUSIONS

We present graphon mixtures, an approach that explicitly models sparse and dense components, and generates graph mixtures using two graphons. The graphon $W$ is used to generate dense graphs and the disjoint clique graphon $U$ is used to generate sparse graphs via the inverse line graph operation. Then the sparse and the dense sequences are joined resulting in a mixture sequence.

Our focus has been on $U$, the new piece in the puzzle. We have modeled the highest degrees of sparse $(U, W)$ mixtures and estimated $U$. We can estimate $U$ with high accuracy when $U$ has finite disjoint cliques. When $U$ has infinitely many disjoint cliques, the accuracy depends on the rate at which the clique proportions go to zero. We have focused on sparse $(U, W)$ mixtures. While we briefly touch on dense $(U, W)$ mixtures in Appendix B, this topic can be explored further. Exploring sparsity resulting from other types of graphs such as paths and rings is an avenue for future work.

## Acknowledgments

We thank the anonymous reviewers for their careful reading of the manuscript and their insightful comments and suggestions, which improved the quality of this work.


## References


Aldous, D. J. (1981), 'Representations for partially exchangeable arrays of random variables', *Journal of Multivariate Analysis* **11**(4), 581–598.

Azizpour, A., Ramezanpour, R., Sabharwal, A. and Segarra, S. (2025), 'From moments to models: Graphon mixture-aware mixup and contrastive learning', *arXiv preprint arXiv:2510.03690* .

Barabási, A.-L. and Albert, R. (1999), 'Emergence of scaling in random networks', *Science* **286**(5439), 509–512.

Beineke, L. W. and Bagga, J. S. (2021), *Line graphs and line digraphs*, Springer.

Bollobás, B. e., Riordan, O., Spencer, J. and Tusnády, G. (2001), 'The degree sequence of a scale-free random graph process', *Random Structures & Algorithms* **18**(3), 279–290.

Borgs, C., Chayes, J. T., Cohn, H. and Holden, N. (2018), 'Sparse exchangeable graphs and their limits via graphon processes', *Journal of Machine Learning Research* **18**(210), 1–71.

Borgs, C., Chayes, J. T., Dhara, S. and Sen, S. (2021), 'Limits of sparse configuration models and beyond: Graphexes and Multigraphexes', *Annals of Probability* **49**(6), 2830–2873.

Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T. and Vesztergombi, K. (2008), 'Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing', *Advances in Mathematics* **219**(6), 1801–1851.

Caron, F. and Fox, E. B. (2017), 'Sparse graphs using exchangeable random measures', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **79**(5), 1295–1366.

Caron, F., Panero, F. and Rousseau, J. (2023), 'On sparsity, power-law, and clustering properties of graphex processes', *Advances in Applied Probability* **55**(4), 1211–1253.

Cheng, C. and Peng, J. (2023), 'Equivariant neural operator learning with graphon convolution', *Advances in Neural Information Processing Systems* **36**, 61960–61984.

Delmas, J.-F., Dhersin, J.-S. and Sciauveau, M. (2021), 'Asymptotic for the cumulative distribution function of the degrees and homomorphism densities for random graphs sampled from a graphon', *Random Structures & Algorithms* **58**(1), 94–149.

Frieze, A. and Kannan, R. (1999), 'Quick approximation to matrices and applications', *Combinatorica* **19**(2), 175–220.

Gehrke, J., Ginsparg, P. and Kleinberg, J. (2003), 'Overview of the 2003 KDD Cup', *ACM SIGKDD Explorations Newsletter* **5**(2), 149–151.

Han, X., Jiang, Z., Liu, N. and Hu, X. (2022), G-mixup: Graph data augmentation for graph classification, *in* 'International conference on machine learning', PMLR, pp. 8230–8248.

Janson, S. (2016), 'Graph limits and hereditary properties', *European Journal of Combinatorics* **52**, 321–337. Special Issue: Recent Advances in Graphs and Analysis.

Kandanaarachchi, S. and Ong, C. S. (2024), 'Graphons of line graphs', *arXiv preprint arXiv:2409.01656* .

Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L. and Zhang, P. (2013), 'Spectral redemption in clustering sparse networks', *Proceedings of the National Academy of Sciences* **110**(52), 20935–20940.

Kumar, S., Zhang, X. and Leskovec, J. (2019), Predicting dynamic embedding trajectory in temporal interaction networks, *in* 'Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining', ACM, pp. 1269–1278.

Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C. and Ghahramani, Z. (2010), 'Kronecker graphs: an approach to modeling networks', *Journal of Machine Learning Research* **11**(2).

Leskovec, J., Kleinberg, J. and Faloutsos, C. (2005), Graphs over time: Densification laws, shrinking diameters and possible explanations, *in* 'Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining', pp. 177–187.

Levie, R. (2023), 'A graphon-signal analysis of graph neural networks', *Advances in Neural Information Processing Systems* **36**, 64482–64525.

Panzarasa, P., Opsahl, T. and Carley, K. M. (2009), 'Patterns and dynamics of users' behavior and interaction: Network analysis of an online community', *Journal of the American Society for Information Science and Technology* **60**(5), 911–932.

Rossi, R. A. and Ahmed, N. K. (2015), The Network Data Repository with Interactive Graph Analytics and Visualization, *in* 'AAAI'.
**URL:** *https://networkrepository.com*

Schwarz, A. J., Gozzi, A. and Bifone, A. (2008), 'Community structure and modularity in networks of correlated brain activity', *Magnetic resonance imaging* **26**(7), 914–920.

Stam, C. J. (2014), 'Modern network science of neurological disorders', *Nature Reviews Neuroscience* **15**(10), 683–695.

Stuart, A. and Ord, K. (2010), *Kendall's advanced theory of statistics, distribution theory*, John Wiley & Sons.

Tieu, K., Fu, D., Zhu, Y., Hamann, H. and He, J. (2024), 'Temporal graph neural tangent kernel with graphon-guaranteed', *Advances in Neural Information Processing Systems* **37**, 94568–94606.

Van den Heuvel, M. P. and Sporns, O. (2013), 'Network hubs in the human brain', *Trends in cognitive sciences* **17**(12), 683–696.

Veitch, V. and Roy, D. M. (2015), 'The class of random graphs arising from exchangeable random measures', *arXiv preprint arXiv:1512.03099* .

Viswanath, B., Mislove, A., Cha, M. and Gummadi, K. P. (2009), On the evolution of user interaction in Facebook, *in* 'Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks (WOSN'09)'.

Whitney, H. (1932), 'Congruent graphs and the connectivity of graphs', *American Journal of Mathematics* **1**, 150–168.

Wu, Y., Zhou, C., Xiao, J., Kurths, J. and Schellnhuber, H. J. (2010), 'Evidence for a bimodal distribution in human communication', *Proceedings of the national academy of sciences* **107**(44), 18803–18808.

Xia, X., Mishne, G. and Wang, Y. (2023), Implicit graphon neural representation, *in* 'International Conference on Artificial Intelligence and Statistics', PMLR, pp. 10619–10634.

Xu, H., Luo, D., Carin, L. and Zha, H. (2021), Learning graphons via structured gromov-wasserstein barycenters, *in* 'Proceedings of the AAAI Conference on Artificial Intelligence', Vol. 35, pp. 10505–10513.

Zang, C., Cui, P., Faloutsos, C. and Zhu, W. (2018), 'On power law growth of social networks', *IEEE Transactions on Knowledge and Data Engineering* **30**(9), 1727–1740.

Zhao, X., Liang, J. and Wang, J. (2021), 'A community detection algorithm based on graph compression for large-scale social networks', *Information Sciences* **551**, 358–372.

## Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Not Applicable]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Yes]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

# Graphon Mixtures: Supplementary Materials

## A Background: Limits of sparse graphs and line graphs

First we define line graphs.

**Definition A.1.** *(**Line graph**) Let $G$ denote a graph with at least one edge. Then its **line graph** denoted by $L(G)$ is the graph whose vertices are the edges of $G$, with two vertices being adjacent if the corresponding edges are adjacent in $G$ (Beineke and Bagga, 2021).*

The **empirical graphon** maps a graph to a function defined on a unit square.

**Definition A.2.** *(**empirical graphon**) Given a graph $G$ with $n$ vertices labeled $\{1, \dots, n\}$, we define its **empirical graphon** $W_G : [0,1]^2 \to [0,1]$ as follows: We split the interval $[0,1]$ into $n$ equal intervals $\{J_1, J_2, \dots, J_n\}$ (first one closed, all others half open) and for $x \in J_i, y \in J_j$ define*

$$W_G(x,y) = \begin{cases} 1 & \textit{if} \quad ij \in E(G) \\ 0 & \textit{otherwise}, \end{cases}$$

*where $E(G)$ denotes the edges of $G$. The empirical graphon replaces the the adjacency matrix with a unit square and the $(i,j)$th entry of the adjacency matrix is replaced with a square of size $(1/n) \times (1/n)$.*

Graph convergence is defined using the **cut-norm** and the **cut-metric**.

**Definition A.3.** *The **cut norm** of graphon $W$ (Frieze and Kannan, 1999; Borgs et al., 2008) is defined as*

$$\|W\|_\square = \sup_{A,B} \left| \int_{A \times B} W(x,y)\, dxdy \right| ,$$

*where the supremum is taken over all measurable sets $A$ and $B$ of $[0,1]$.*

**Definition A.4.** *(**cut metric**) Given two graphons $W_1$ and $W_2$ the **cut metric** (Borgs et al., 2008) is defined as*

$$\delta_\square(W_1, W_2) = \inf_\varphi \|W_1 - W_2^\varphi\|_\square ,$$

*where the infimum is taken over all measure preserving bijections $\varphi : [0,1] \to [0,1]$.*

The cut metric is a pseudo-metric because $\delta_\square(W_1, W_2) = 0$ does not imply $W_1 = W_2$, i.e., $\delta_\square(W_1, W_2) \geq 0$ for $W_1 \neq W_2$.

Next we give some definitions from Janson (2016).

**Definition A.5.** *(**Janson's mass-partition**) We define a mass-partition to be a sequence $\boldsymbol{p} = \{p_i\}_{i=1}^\infty$ of non-negative real numbers such that*

$$p_1 \geq p_2 \geq \cdots \geq 0 \quad \textit{and} \quad \sum_{i=1}^\infty p_i \leq 1 .$$

*Let $\mathcal{M}$ be the set of all mass-partitions.*

For every mass-partition we can define a graphon as follows:

**Definition A.6.** *Given a mass-partition $\boldsymbol{p}$ (Definition A.5) we define a graphon $W_{\boldsymbol{p}}^{\mathcal{M}}$ by taking disjoint subsets $\{A_i\}_{i=1}^\infty$ of a probability space $(S, \mu)$ such that $\mu(A_i) = p_i$ and defining $W_{\boldsymbol{p}}^{\mathcal{M}} = \sum_{i=1}^\infty \mathbf{1}_{A_i \times A_i}$.*

We can define a mass-partition for any graph. To do that we use component sizes.

**Definition A.7.** *For any graph $G$, we denote the component sizes by $C_1(G) \geq C_2(G) \geq \cdots$, ordered such that when $k$ is larger than the number of components $C_k(G) = 0$.*

Each graph $G$ defines a mass-partition $\{C_i(G)/|G|\}_{i=1}^{\infty}$ where $|G|$ denotes the number of nodes in $|G|$.

The next theorem links mass partitions, graphons and disjoint clique graphs.

**Theorem A.8. (Janson (2016) Thm 7.5)** *If $\{G_n\}_n$ is a sequence of disjoint clique graphs with $|G_n| \to \infty$ and $W$ is a graphon, then $G_n \to W$ if an only if $W = W_{\boldsymbol{p}}^{\mathcal{M}}$ for some $\boldsymbol{p} \in \mathcal{M}$ and $\{(C_i(G_n)/|G_n|)_i\} \to \boldsymbol{p}$.*

Theorem A.8 tells us that a mass-partition $\boldsymbol{p}$ can be used to describe a disjoint clique graphon, i.e., there is a one-to-one correspondence between disjoint clique graphons and mass-partitions.

Theorem 2.3 and Theorem A.8 (Janson, 2016) tells us that graphons of line graphs are disjoint clique graphons and as such they can be described by mass-partitions.

### A.1 Square-degree property and max-degree condition

**Definition A.9 (Square-degree property).** *Let $\{G_n\}_n$ denote a sequence of graphs. Then $\{G_n\}_n$ exhibits the square-degree property if there exists some $c_1 > 0$ and $N_0 \in \mathbb{N}$ such that for all $n \geq N_0$*

$$\sum deg\ v_{i,n}^2 \geq c_1 \left(\sum deg\ v_{i,n}\right)^2 .$$

*The set of graph sequences satisfying the square-degree property is denoted by $S_q$.*

**Lemma 2.6.** *Let $\{G_n\}_n$ be a graph sequence with $H_m = L(G_n)$. Suppose $\{H_m\}_m$ converges to $U$. Then*

$$\{G_n\}_n \in S_q \equiv \{G_n\}_n \in S_x .$$

*Proof.* See Lemmas A.10 and A.11. □

**Lemma A.10.** *Let $\{G_n\}_n$ be a graph sequence with $H_m = L(G_n)$. Suppose $\{H_m\}_m$ converges to $U$. Then*

$$\{G_n\}_n \in S_x \implies \{G_n\}_n \in S_q ,$$

*where $S_x$ denotes the set of graph sequences satisfying the max-degree condition (Definition 2.5) and $S_q$ denotes the set of graph sequences satisfying the square-degree property (Definition A.9).*

*Proof.* We compute the ratio of the sum of degree squares to the number of edges squared as this quantity determines the square-degree property (Definition A.9). For $\{G_n\}_n \in S_x$

$$\frac{\sum \deg\ v_{i,n}^2}{m^2} \geq \frac{d_{\max,n}^2}{m^2} \geq c^2 > 0 ,$$

we obtain the result. □

**Lemma A.11.** *Let $\{G_n\}_n$ be a graph sequence with $H_m = L(G_n)$. Suppose $\{H_m\}_m$ converges to $U$. Then*

$$\{G_n\}_n \in S_q \implies \{G_n\}_n \in S_x ,$$

*where $S_x$ denotes the set of graph sequences satisfying the max-degree condition (Definition 2.5) and $S_q$ denotes the set of graph sequences satisfying the square-degree property (Definition A.9).*

*Proof.* For $\{G_n\}_n \in S_q$ as $H_m$ converges to $U$

$$\liminf_{n\to\infty} \frac{\sum \deg\ v_{i,n}^2}{m^2} = \lim_{n\to\infty} \frac{\sum \deg\ v_{i,n}^2}{m^2} = c > 0$$

Let us denote the ordered degrees by $d_{(1)}, d_{(2)}$, and so on where $d_{max,n} = d_{(1)}$. Then for any $n$

$$\frac{d_{(1)}}{m} \geq \frac{d_{(2)}}{m} \geq \cdots \geq \frac{d_{(n)}}{m} ,$$

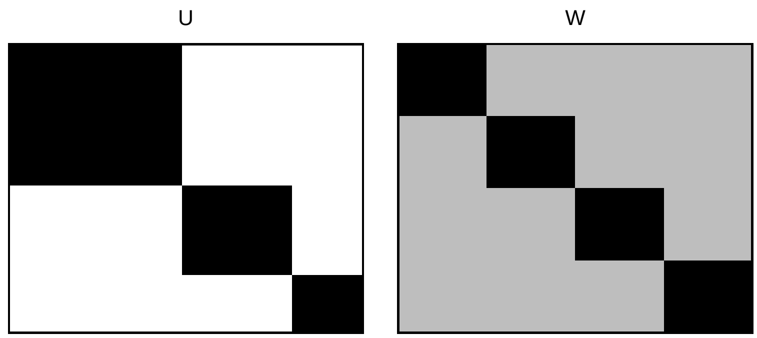



Figure 5: Graphons $U$ and $W$ used in Example 1.

and

$$\sum_i \deg\, v_{i,n}^2 = d_{(1)}^2 + d_{(2)}^2 + \cdots + d_{(n)}^2 \,.$$

Noting the series $\frac{\sum \deg\, v_{i,n}^2}{m^2}$ is absolutely convergent and the individual terms are positive $\left(\frac{d_{(i)}^2}{m^2} > 0\right)$ we have

$$\lim_{n\to\infty} \frac{\sum \deg\, v_{i,n}^2}{m^2} = \lim_{n\to\infty} \frac{d_{(1)}^2}{m^2} + \lim_{n\to\infty} \frac{d_{(2)}^2}{m^2} + \cdots + \lim_{n\to\infty} \frac{d_{(n)}^2}{m^2} = c\,.$$

If each individual limit on the right hand side of the above equation were zero, then we would get a contradiction. As such, at least some limits need to be positive. As $d_{(1)}/m$ is the largest, there exists a constant $c_2 > 0$ such that

$$\frac{d_{(1)}^2}{m^2} = \frac{d_{\max,n}^2}{m^2} = c_2\,,$$

which completes the proof. □

Lemmas A.10 and A.11 show that if we consider sparse graph sequences that have converging line graphs, then the square-degree property is equivalent to the max-degree property.

# B Mixtures of sparse and dense graphs

## B.1 $(U, W)$ mixture examples

We give two examples of $(U, W)$-mixture graphs in this section.

### B.1.1 Example 1

We consider $U$ and $W$ shown in Figure 5 where $U$ is given by the mass-partition $\boldsymbol{p} = (0.5, 0.3, 0.2)$ and $W$ is given by a stochastic block model graphon. Graphon $W$ has 4 equally sized communities where the edge probability of two nodes $i$ and $j$ within the same community is given by $p_{ij} = 0.6$ and the edge probability of two nodes $i$ and $k$ in two different communities is given by $p_{ik} = 0.025$.

Figures 6 and 7 show examples of mixture graphs from this $U$ and $W$. In Figure 6 the dense part $G_{d_i}$ has 100 nodes and the sparse part $G_{s_i}$ has 600 nodes. As $G_{s_i}$ has a larger number of nodes, the resulting mixture graph $G_{n_i}$ has large hubs, resembling a graph from a sparse sequence. In contrast, in Figure 7 the dense part $G_{d_i}$ has 100 nodes and the sparse part $G_{s_i}$ has 50 nodes making the mixture graph $G_{n_i}$ resemble a graph from a dense sequence.

From this $U$ and $W$ we construct a sparse $(U, W)$-mixture sequence with $n_{s_i} = \lceil n_{d_i} \times \sqrt{5i} \rceil$ for $i \in \{1, \ldots 20\}$. For each mixture graph $G_{n_i}$, we compute the global clustering coefficient, the mean distance of the graph and its spectral radius. Table 4 gives the results. We see that the clustering coefficient does not go to zero in this example.

Table 4: The global clustering coefficient, mean distance and spectral radius of the sparse $(U, W)$ mixture graphs sequence discussed in Example 1

| $n_{d_i}$ | $n_{s_i}$ | Clustering Coefficient | Mean Distance | Spectral Radius |
|---|---|---|---|---|
| 100 | 224 | 0.32 | 4.06 | 17.08 |
| 200 | 633 | 0.31 | 3.83 | 33.60 |
| 300 | 1162 | 0.33 | 3.67 | 50.50 |
| 400 | 1789 | 0.33 | 3.75 | 67.46 |
| 500 | 2500 | 0.33 | 3.66 | 84.18 |
| 600 | 3287 | 0.33 | 3.57 | 101.17 |
| 700 | 4142 | 0.33 | 3.54 | 117.99 |
| 800 | 5060 | 0.33 | 3.42 | 134.99 |
| 900 | 6038 | 0.33 | 3.48 | 151.80 |
| 1000 | 7072 | 0.33 | 3.48 | 168.85 |
| 1100 | 8158 | 0.33 | 3.47 | 186.21 |
| 1200 | 9296 | 0.33 | 3.41 | 203.01 |
| 1300 | 10481 | 0.33 | 3.46 | 219.61 |
| 1400 | 11714 | 0.33 | 3.37 | 236.22 |
| 1500 | 12991 | 0.33 | 3.32 | 252.36 |
| 1600 | 14311 | 0.33 | 3.28 | 270.54 |
| 1700 | 15674 | 0.33 | 3.27 | 287.00 |
| 1800 | 17077 | 0.33 | 3.25 | 303.76 |
| 1900 | 18519 | 0.33 | 3.29 | 320.70 |
| 2000 | 20000 | 0.33 | 3.26 | 337.55 |

#### B.1.2 Example 2

Next we consider $U$ having the mass-partition $\boldsymbol{p} \propto \left(\frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{10}\right)$ such that $\sum_i p_i = 1$. Let $W$ be given by

$$W = 0.9 \exp\left(\frac{(-y^2 - (x-1)^2)}{0.05^2}\right) + 0.9 \exp\left(\frac{(-(y-1)^2 - x^2)}{0.05^2}\right) + 0.9 \exp\left(-\frac{(\sin(3\pi/4)\,x + \cos(3\pi/4)\,x)}{0.05^2}\right),$$

as in Xia et al. (2023). Graphon $W$ produces dense ring-type graphs. The two graphons are shown in Figure 8. Figure 9 shows the dense part $G_{d_i}$ sampled from $W$ with $n_{d_i} = 100$, the sparse part $G_{s_i}$ with $n_{s_i} = 100$ and the mixture graph $G_{n_i}$. Figure 10 shows the dense part $G_{d_i}$ with $n_{d_i} = 100$, the sparse part $G_{s_i}$ with $n_{s_i} = 600$ and the mixture graph $G_{n_i}$. We see that the hub structure is stronger in Figure 10 as $n_{s_i} > n_{d_i}$.

Again we construct a sparse mixture sequence from this $U$ and $W$. Similar to Example 1, we let $n_{s_i} = \lceil n_{d_i} \times \sqrt{5i} \rceil$ for $i \in \{1, \ldots 20\}$. For each mixture graph $G_{n_i}$, we compute the global clustering coefficient, the mean distance of the graph and its spectral radius. Table 5 gives the results. Similar to Example 1, we see a robust clustering coefficient as $n_{s_i}$ increases.

We see that the mixture graphs $G_{n_i}$ shown in Figures 6 and 7 are quite different from those in Figures 9 and 10. The main reason is that graphon $W$ stipulates the dense part of the mixture and as such, different $W$ produces different mixture graphs.

### B.2 An example where the cut metric cannot explain the anomalous node

Consider the following example: Let $\{G_n\}_n$ be a sequence of graphs such that originally $G_n \sim G(n, p)$ where $G(n, p)$ denotes a graph generated from the Erdős–Rényi model where $G_n$ has $n$ nodes and the edge probability is $p$. Then we modify each graph in the sequence such that a randomly selected node in $G_n$ is connected to all other nodes in $G_n$. Figure 11 shows the empirical graphons for $n \in \{50, 100, 200, 400\}$ with $p = 0.1$ for this construction.

The empirical graphons in this example converge to $W = 0.1$ in the cut metric (Definition A.4). However if we inspect the degree distribution, we see a persistent anomalous node with high degree as shown in Figure 12.

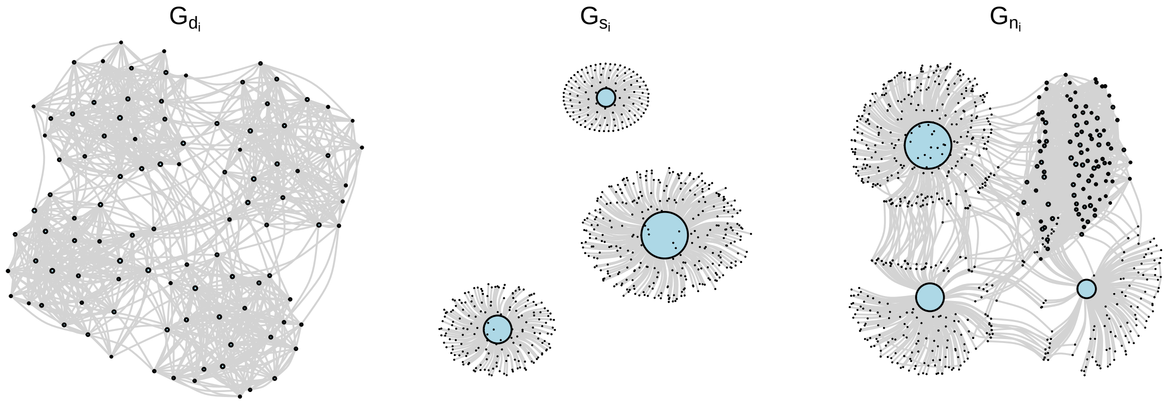



Figure 6: A $(U, W)$-mixture graph with $n_{d_i} = 100$ and $n_{s_i} = 600$ where $U$ and $W$ are shown in Figure 5. The mixture graph $G_{n_i}$ resembles a graph from a sparse sequence.

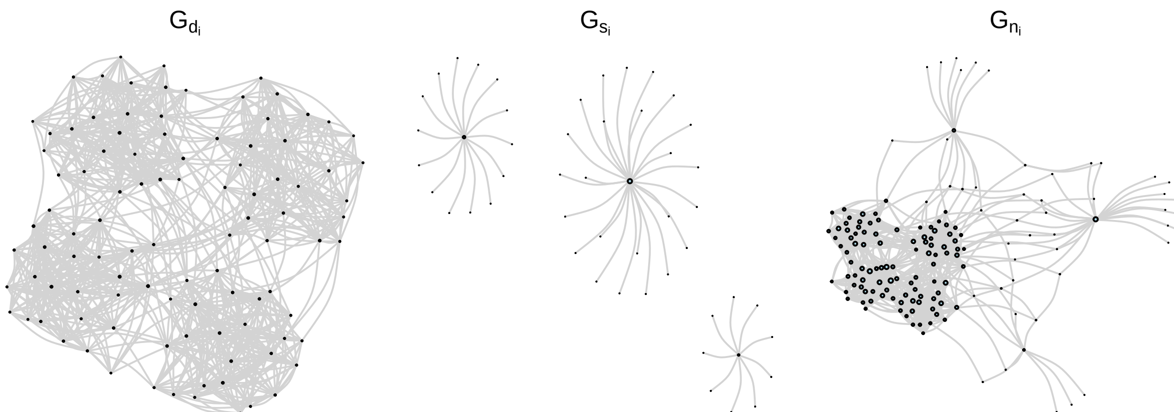



Figure 7: A $(U, W)$-mixture graph with $n_{d_i} = 100$ and $n_{s_i} = 50$ where $U$ and $W$ are shown in Figure 5. The mixture graph $G_{n_i}$ resembles a graph from a dense sequence.

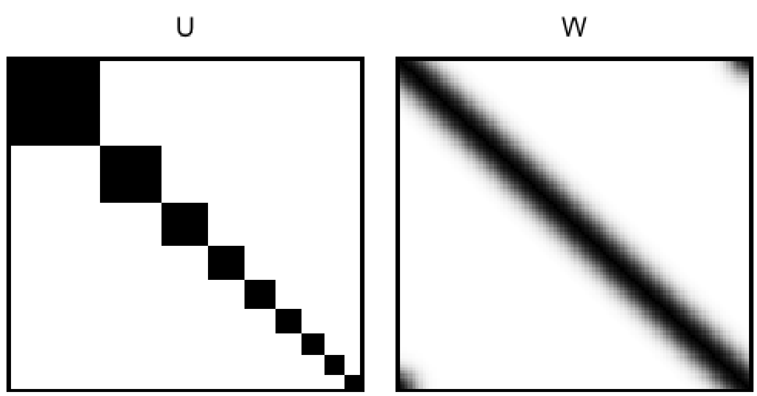



Figure 8: Graphons $U$ and $W$ used in Example 2.

The empirical graphons (Definition A.2) converge to $W = 0.1$ in the cut-metric (Definition A.4). While the graph sequence is dense, it has a sparse component that a single graphon $W$ does not fully explain. That is, graphs

Table 5: The global clustering coefficient, mean distance and spectral radius of the sparse $(U, W)$ mixture graphs sequence discussed in Example 2

| $n_{d_i}$ | $n_{s_i}$ | Clustering Coefficient | Mean Distance | Spectral Radius |
|---|---|---|---|---|
| 100 | 224 | 0.27 | 6.06 | 10.62 |
| 200 | 633 | 0.29 | 4.51 | 21.84 |
| 300 | 1162 | 0.31 | 4.28 | 33.22 |
| 400 | 1789 | 0.31 | 4.23 | 44.12 |
| 500 | 2500 | 0.32 | 4.06 | 55.80 |
| 600 | 3287 | 0.32 | 4.14 | 66.76 |
| 700 | 4142 | 0.31 | 4.08 | 77.61 |
| 800 | 5060 | 0.32 | 4.09 | 89.42 |
| 900 | 6038 | 0.32 | 3.98 | 100.76 |
| 1000 | 7072 | 0.32 | 4.03 | 111.96 |
| 1100 | 8158 | 0.32 | 4.04 | 123.18 |
| 1200 | 9296 | 0.32 | 4.03 | 133.78 |
| 1300 | 10481 | 0.32 | 4.02 | 145.67 |
| 1400 | 11714 | 0.32 | 3.88 | 156.22 |
| 1500 | 12991 | 0.32 | 3.97 | 167.71 |
| 1600 | 14311 | 0.32 | 3.96 | 178.63 |
| 1700 | 15674 | 0.32 | 3.96 | 190.49 |
| 1800 | 17077 | 0.32 | 3.92 | 201.24 |
| 1900 | 18519 | 0.32 | 3.86 | 212.29 |
| 2000 | 20000 | 0.32 | 3.91 | 223.61 |

generated from $W = 0.1$, would not have the persistent anomalous node shown in Figure 12. This shows that $W = 0.1$ does not fully capture the observed graphs. Modeling these graphs as a mixture gives the flexibility to account for the persisting high degree node.

### B.3 Generating dense and sparse graph sequences with a $(U, W)$ mixture

In this section we consider $(U, W)$-mixture graph sequences $\{G_{n_i}\}_i$ generated with $W = \exp(-(x + y))$ and $U$ with mass-partition $\left(\frac{2}{3}, \frac{1}{3}\right)$. The graphons $U$ and $W$ are shown in Figure 13. We consider graph generation for different rates of evolution of $\frac{n_{s_i}}{n_{d_i}}$.

**Dense $\{G_{n_i}\}_i$, when $\frac{n_{s_i}}{n_{d_i}} \to c \in \mathbb{R}$:** When $\frac{n_{s_i}}{n_{d_i}} \to c \in \mathbb{R}$ the graph sequence $\{G_{n_i}\}_i$ is dense (Lemma 3.4), i.e., the edge density of $G_{n_i}$ does not go to zero. Figure 14 shows an example of the dense part $G_{d_i}$, the sparse part $G_{s_i}$ and the mixture graph $G_{n_i}$ for $n_{d_i} = n_{s_i} = 100$. Figure 15 shows the degree distributions for different values of $n_{d_i}$ and $n_{s_i}$ when the ratio $\frac{n_{s_i}}{n_{d_i}} = 1$. When the ratio $\frac{n_{s_i}}{n_{d_i}}$ converges to $c \leq 1$, depending on $W$ and $U$, the largest degree may be contributed by the dense part. In Figure 15 the largest degree is actually produced by $U$, however, it is so close to the degrees generated by $W$ that we cannot distinguish the effect of $U$ in this example.

Figure 16 shows the dense part $G_{d_i}$ and the sparse part $G_{s_i}$ when $n_{d_i} = 100$ and $n_{s_i} = 300$. Figure 17 shows the degree distribution when $\frac{n_{s_i}}{n_{d_i}} = 3$. When the ratio $\frac{n_{s_i}}{n_{d_i}}$ converges to $c > 1$, for large enough $i$ the sparse part contributes to highest-degree nodes, which are anomalous. In Figure 17 two anomalous nodes are contributed by $U$ for each pair of $n_{s_i}$ and $n_{d_i}$.

**Sparse $\{G_{n_i}\}_i$, when $\frac{n_{s_i}}{n_{d_i}} \to \infty$:** If $\{G_{n_i}\}_i$ is sparse, then $n_{s_i}$ grows faster than $n_{d_i}$ making $\frac{n_{s_i}}{n_{d_i}}$ go to infinity (Lemma 3.4). We used $n_{s_i} = \lceil \sqrt{5i} \times n_{d_i} \rceil$ for this experiment and Figure 19 shows the degree distributions. Two anomalous nodes contributed by $U$ can be seen in 3 of the 4 subplots. Figure 18 shows the dense part $G_{d_i}$, the sparse part $G_{s_i}$ and the mixture graph $G_{n_i}$ when $n_{d_i} = 300$ and $n_{s_i} = 1162$. We see that the effect of the stars is higher in this mixture.

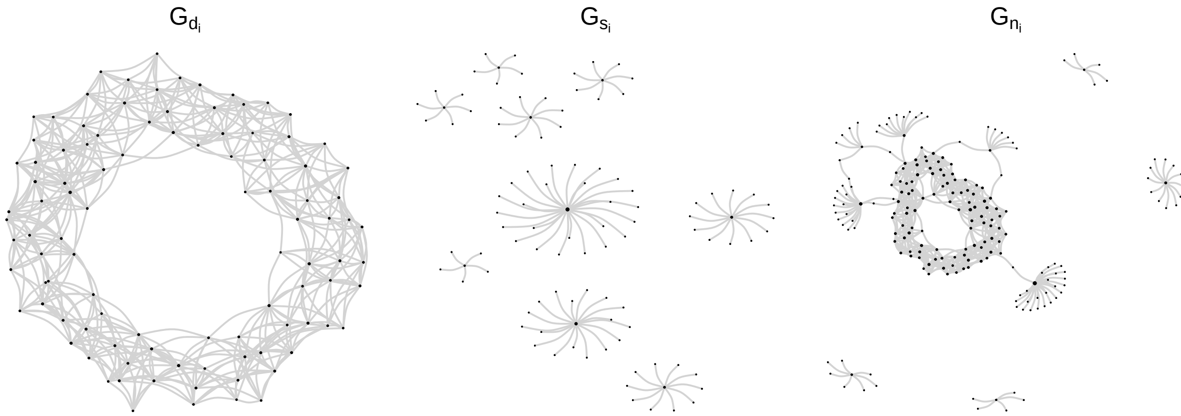



Figure 9: A $(U, W)$-mixture graph with $n_{d_i} = 100$ and $n_{s_i} = 100$ where $U$ and $W$ are shown in Figure 8.

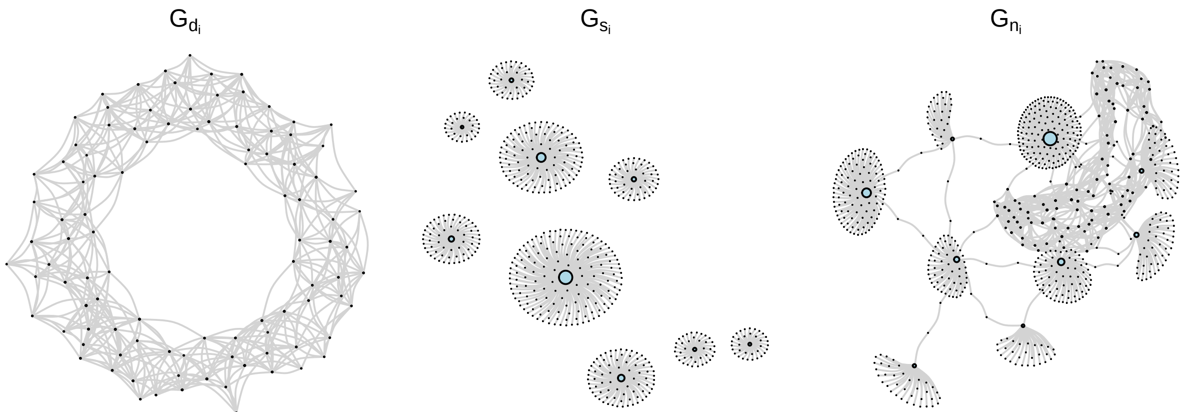



Figure 10: A $(U, W)$-mixture graph with $n_{d_i} = 100$ and $n_{s_i} = 600$ where $U$ and $W$ are shown in Figure 8

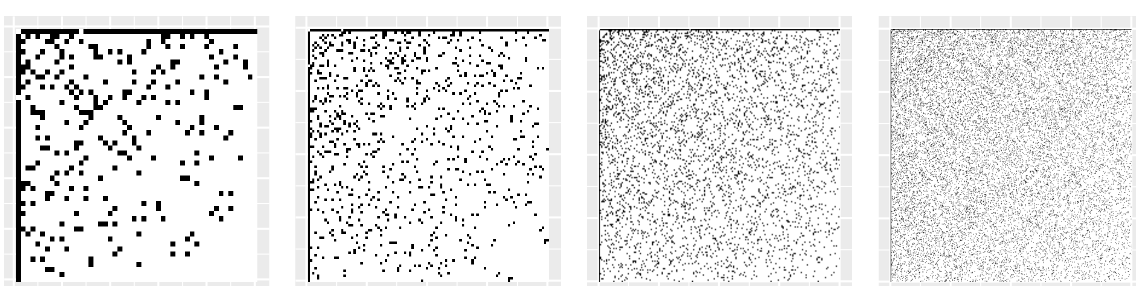


Figure 11: Empirical graphons of $G_n \sim G(n, p)$ with one node connected to all other nodes for $n \in \{50, 100, 200, 400\}$.

**Different $n_{s_i}/n_{d_i}$ ratios:** Figure 20 shows the edge density along the graph sequence for six functions of $n_{s_i}/n_{d_i}$. The first experiment (Exp 1 in Figure 20) considers $n_{s_i}/n_{d_i} = 1$ and experiment 2 considers $n_{s_i}/n_{d_i} = 2$. For experiment 3 we consider $n_{s_i}/n_{d_i} \approx 1/\sqrt{i}$. As $n_{s_i}/n_{d_i}$ are bounded from above for the first 3 experiments we get dense graphs. For experiment 4, we considered $n_{s_i}/n_{d_i} \approx \sqrt{i}$ making the mixture sequence sparse. For experiments 5 and 6 we considered $n_{s_i}/n_{d_i} = i$ and $n_{s_i}/n_{d_i} = i^2$ respectively. Experiments 4, 5 and 6 produced sparse sequences. For sparse sequences $\{n_{s_i}/n_{d_i}\}_i$ determines the rate at which the mixture becomes sparse.

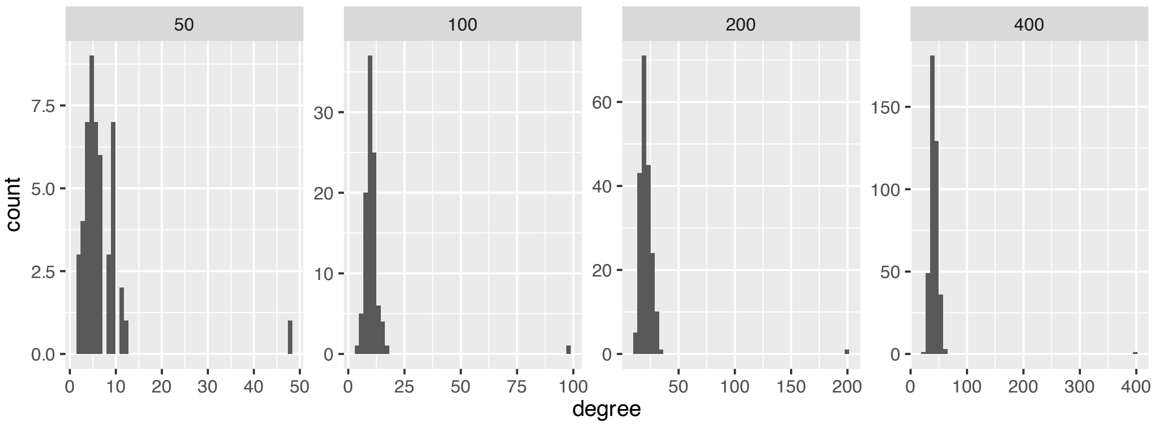



Figure 12: Degree histograms of the graphs corresponding to empirical graphons in Figure 11.

### B.4 Expectations

**Lemma B.1.** *Let* $\{G_{n_i}\}_i$ *be a sequence of* $(U, W)$*-mixture graphs (Definition 3.2) with* $G_{n_i} \sim \mathbb{G}(U, W, n_{d_i}, m_{s_i})$ *with dense and sparse parts* $G_{d_i}$ *and* $G_{s_i}$ *respectively. Let* $n_{d_i}$ *and* $n_{s_i}$ *be the number of nodes in* $G_{d_i}$ *and* $G_{s_i}$ *respectively. Suppose* $m_{new_i}$ *edges are added as part of the joining process. Suppose* $v_\ell$ *is originally a vertex in the dense part* $G_{d_i}$*. We denote its degree in* $G_{d_i}$ *by deg* $_{G_{d_i}} v_\ell$ *and its degree in* $G_{n_i}$ *by deg* $_{G_{n_i}} v_\ell$*. Then taking expectations over the joining process we have*

$$\mathbb{E}\left(deg_{\;G_{n_i}} v_\ell\right) = deg_{\;G_{d_i}} v_\ell + c_1 \frac{m_{new_i}}{n_{d_i}}$$
$$Var(deg_{\;G_{n_i}} v_\ell) = m_{new_i} \frac{c_1}{n_{d_i}} \left(1 - \frac{c_1}{n_{d_i}}\right) ,$$

*where* $c_1 > 0$ *depend on the joining process. Similarly, suppose* $v_j$ *is originally a vertex in the sparse part* $G_{s_i}$*. We denote its degree in* $G_{s_i}$ *by deg* $_{G_{s_i}} v_j$ *and its degree in* $G_{n_i}$ *by deg* $_{G_{n_i}} v_j$*. Then, taking expectations over the joining process,*

$$\mathbb{E}\left(deg_{\;G_{n_i}} v_j\right) = deg_{\;G_{s_i}} v_j + c_2 \frac{m_{new_i}}{n_{s_i}} ,$$
$$Var(deg_{\;G_{n_i}} v_j) = m_{new_i} \frac{c_2}{n_{s_i}} \left(1 - \frac{c_2}{s_{d_i}}\right) ,$$

*where* $c_2 > 0$*.*

*Proof.* The *Random edges condition* (Definition 3.3 Condition 3) stipulates that nodes within a graph are equally likely to be selected. That is, the probability of selecting a node is inversely proportional to the number of nodes in the graph. For a new edge, the probability of selecting a node in $G_{d_i}$ is $\frac{c_1}{n_{d_i}}$ and the probability of selecting a node in $G_{s_i}$ is $\frac{c_2}{n_{s_i}}$. This gives rise to a binomial distribution. Thus, the expected number of new edges for node $v_\ell$ in $G_{d_i}$ is $c_1 \frac{m_{new_i}}{n_{d_i}}$ and its variance is $m_{new_i} \frac{c_1}{n_{d_i}} \left(1 - \frac{c_1}{n_{d_i}}\right)$. For node $v_j$ in $G_{s_i}$ the expected number of new edges is $c_2 \frac{m_{new_i}}{n_{s_i}}$ and its variance is $m_{new_i} \frac{c_2}{n_{s_i}} \left(1 - \frac{c_2}{s_{d_i}}\right)$.

□

**Lemma B.2.** *Let* $\{G_{n_i}\}_i$ *be a sequence of* $(U, W)$*-mixture graphs (Definition 3.2) with* $G_{n_i} \sim \mathbb{G}(U, W, n_{d_i}, m_{s_i})$ *with sparse part* $G_s$*. Let* $\boldsymbol{p} = (p_1, p_2, \ldots)$ *be the mass-partition (Definition 3.1) corresponding to* $U$*. Let* $\tilde{q}_{j,i}$ *be the degree of the star in* $G_{s_i}$ *corresponding to* $p_j \neq 0$ *and let* $q_{j,i}$ *denote the degree of the corresponding node in* $G_{n_i}$*.*

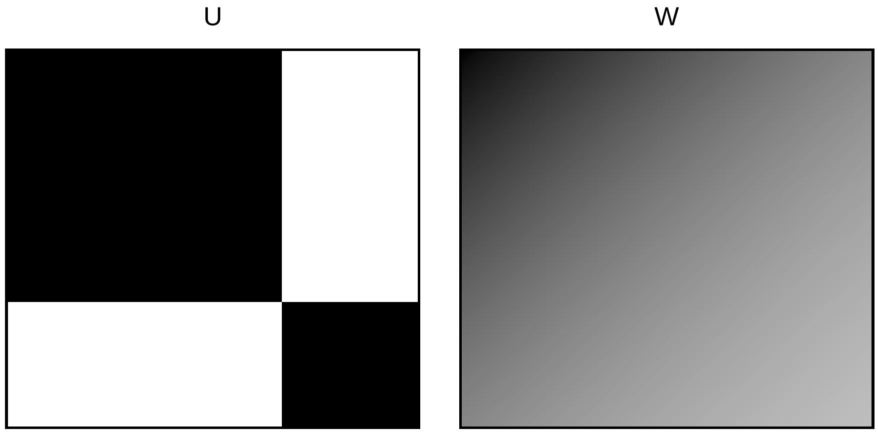



Figure 13: Graphons $U$ and $W$ in the $(U, W)$ mixture with $W = \exp(-(x + y))$ and $U$ having the mass-partition $(2/3, 1/3)$. Graphon $U$ lives in the line graph space and is used to generate sparse graphs via the inverse line graph operation. Graphon $W$ is the limit of the dense part.

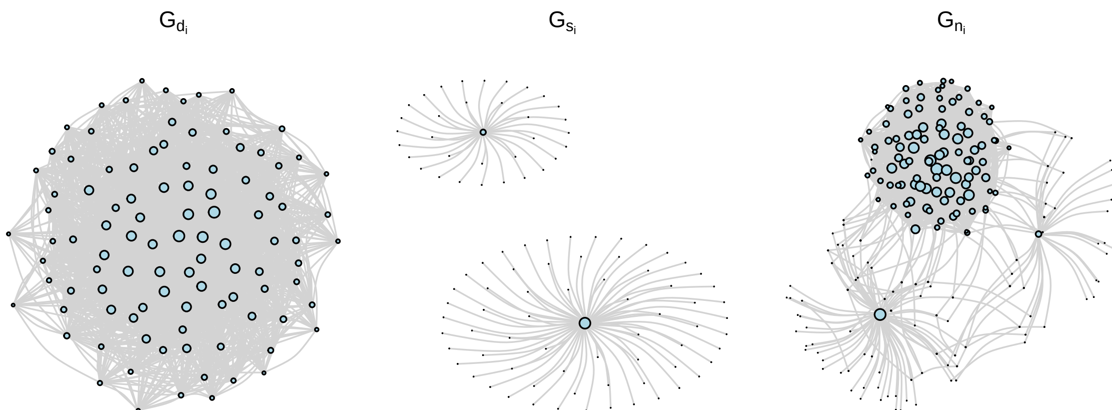



Figure 14: The dense part $G_{d_i}$, sparse part $G_{s_i}$ and the mixture graph $G_{n_i}$ Both dense part $G_{d_i}$ and sparse part $G_{s_i}$ have 100 nodes each.

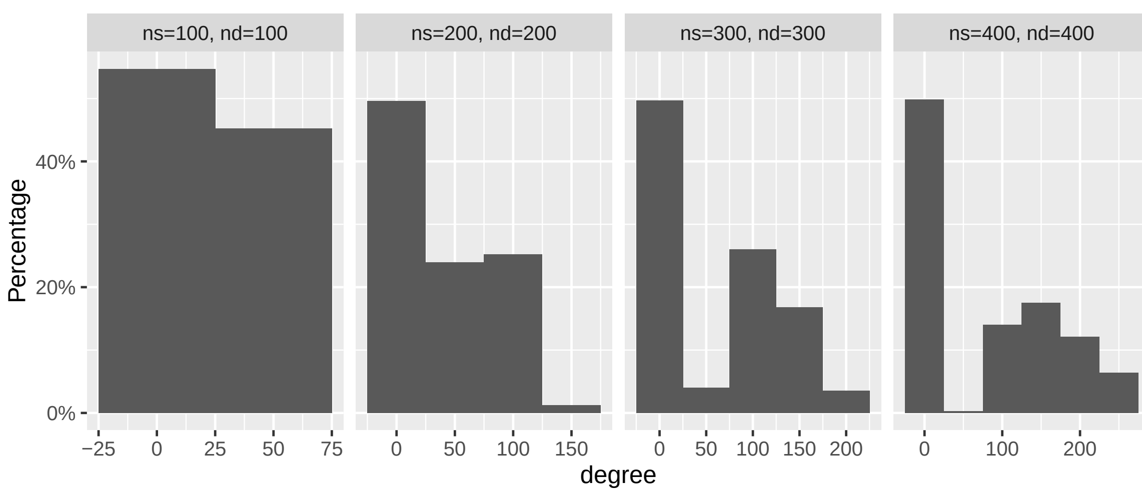



Figure 15: Degree distribution when $\frac{n_{s_i}}{n_{d_i}} = 1$

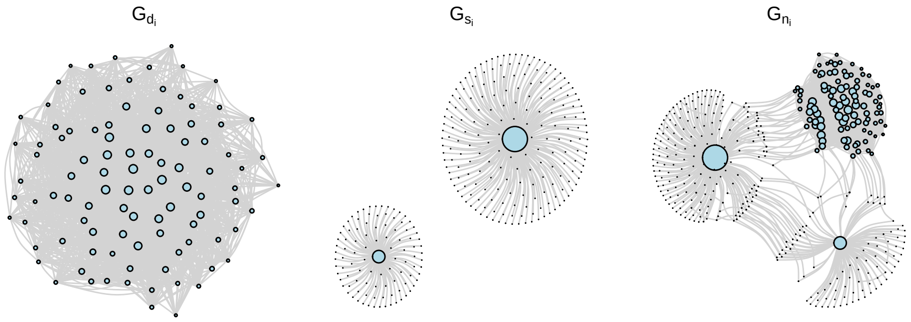



Figure 16: The dense part $G_{d_i}$, sparse part $G_{s_i}$ and the mixture graph $G_{n_i}$ with $n_{d_i} = 100$ and $n_{s_i} = 300$.

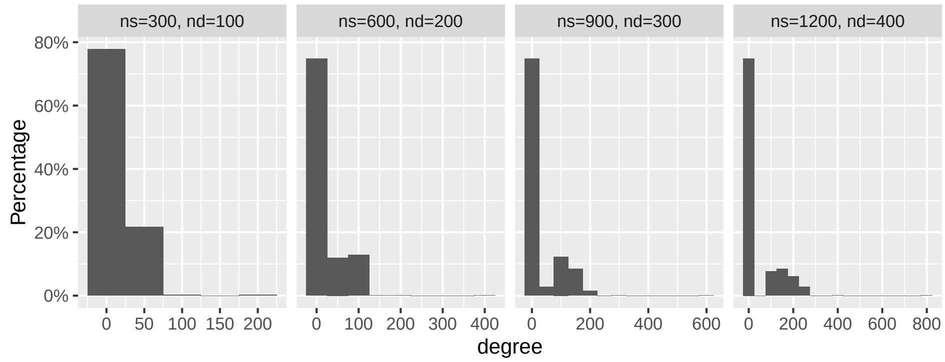



Figure 17: When $\frac{n_{s_i}}{n_{d_i}} = 3$

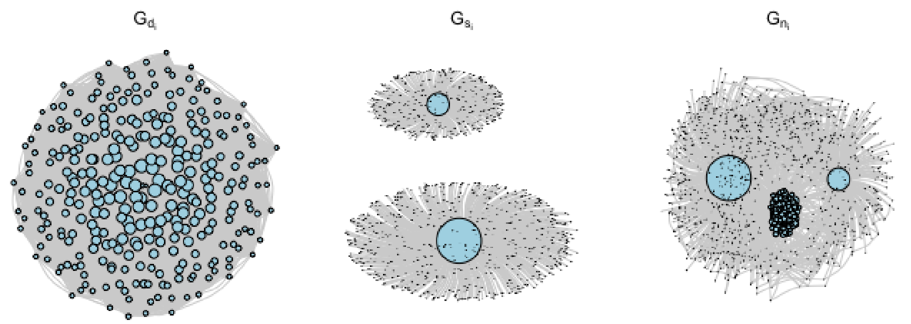



Figure 18: The dense part $G_{d_i}$, sparse part $G_{s_i}$ and the mixture graph $G_{n_i}$ with $n_{d_i} = 300$ and $n_{s_i} = 1162$.

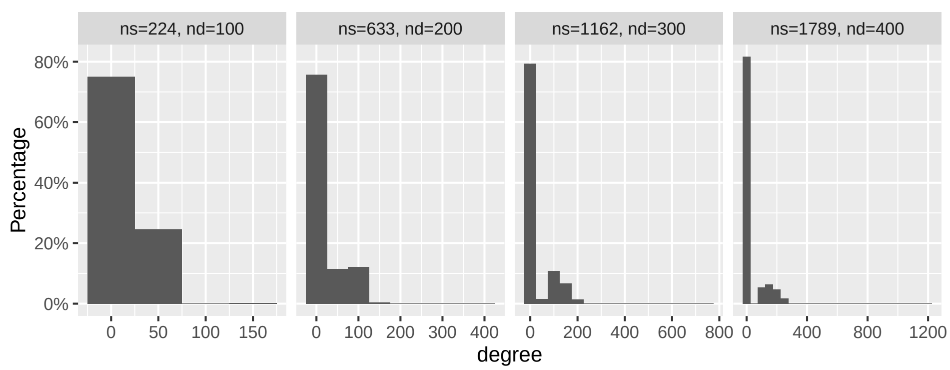



Figure 19: When $\frac{n_{s_i}}{n_{d_i}} \to \infty$

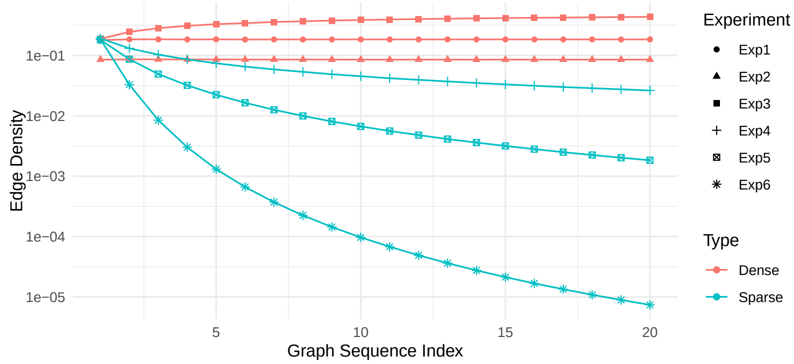



Figure 20: Sparse and dense $(U, W)$ sequences for different functions of $n_{s_i}/n_{d_i}$.

*Then taking expectations over graph generation and the joining process*

$$\mathbb{E}(q_{j,i}) = m_{s_i} p_j + \frac{c m_{new_i}}{n_{s_i}} ,$$
$$Var(q_{j,i}) = m_{s_i} p_j (1 - p_j) + m_{new_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right) .$$

*Proof.* The *Random edges condition* in Definition 3.3 specifies that edges are added randomly. From Lemma B.4 we know that

$$\mathbb{E}(\tilde{q}_{j,i}) = m_{s_i} p_j \quad \text{and} \quad \text{Var}(\tilde{q}_{j,i}) = m_{s_i} p_j (1 - p_j) .$$

Combining with Lemma B.1 we obtain

$$\mathbb{E}(q_{j,i}) = \mathbb{E}(\tilde{q}_{j,i}) + \frac{c m_{new_i}}{n_{s_i}}$$
$$\mathbb{E}(q_{j,i}) = m_{s_i} p_j + \frac{c m_{new_i}}{n_{s_i}} .$$

Similarly the variance satisfies

$$\text{Var}(q_{j,i}) = \text{Var}(\tilde{q}_{j,i}) + \text{Var}(\text{new edges for this vertex}) ,$$
$$\text{Var}(q_{j,i}) = m_{s_i} p_j (1 - p_j) + m_{new_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right) .$$

□

### B.5 Edge density of $(U, W)$-mixture graphs

**Lemma 3.4.** *Let $\{G_{n_i}\}_i$ be a sequence of $(U, W)$-mixture graphs (Definition 3.2) with $G_{n_i} \sim \mathbb{G}(U, W, n_{d_i}, m_{s_i})$. Let $G_{d_i}$ be the dense part of $G_{n_i}$ and let $G_{s_i}$ be the sparse part. Let $n_{d_i}$ and $n_{s_i}$ be the number of nodes in $G_{d_i}$ and $G_{s_i}$ respectively. Then*

1. *If there exists $c \in \mathbb{R}^+$ such that $\limsup_{i\to\infty} \frac{n_{s_i}}{n_{d_i}} = c$ then $\{G_{n_i}\}_i$ is dense.*

2. *If $\lim_{i\to\infty} \frac{n_{s_i}}{n_{d_i}} = \infty$, then $\{G_{n_i}\}_i$ is sparse.*

*Proof.* Suppose $G_{d_i}$, $G_{s_i}$ and $G_{n_i}$ have $n_{d_i}$, $n_{s_i}$ and $n_i$ nodes and $m_{d_i}$, $m_{s_i}$ and $m_i$ edges respectively. Then $m_{d_i} + m_{s_i} \leq m_i$ and $n_i = n_{d_i} + n_{s_i}$.

1. Then the edge density of $G_{n_i}$ satisfies

$$\begin{aligned}
\text{density}(G_{n_i}) &\geq \frac{2(m_{d_i}+m_{s_i})}{(n_{d_i}+n_{s_i})^2}\,, \\
&= \frac{2(m_{d_i}+m_{s_i})}{(n_{d_i}^2+2n_{d_i}n_{s_i}+n_{s_i}^2)}\,, \\
&= 2\frac{\frac{m_{d_i}}{n_{d_i}^2}+\frac{m_{s_i}}{n_{d_i}^2}}{1+2\frac{n_{s_i}}{n_{d_i}}+\frac{n_{s_i}^2}{n_{d_i}^2}}\,, \\
&= 2\frac{\frac{m_{d_i}}{n_{d_i}^2}+\frac{m_{s_i}}{n_{s_i}^2}\frac{n_{s_i}}{n_{d_i}^2}}{1+2\frac{n_{s_i}}{n_{d_i}}+\frac{n_{s_i}^2}{n_{d_i}^2}}\,, \\
&\geq \frac{\rho_{d_i}}{1+2c+c^2}\,,
\end{aligned}$$

where $\rho_{d_i}$ is the edge density of the dense part $G_{d_i}$. As $G_{d_i}$ is sampled from $W$, the edge density $\rho_{d_i}$ is bounded away from zero as $i$ goes to infinity. This makes the edge density of $G_{n_i}$ strictly positive as $i$ goes to infinity, making $\{G_{n_i}\}_i$ dense.

2. Suppose $\lim_{i\to\infty}\frac{n_{s_i}}{n_{d_i}}=\infty$. Equivalently, $\lim_{i\to\infty}\frac{n_{d_i}}{n_{s_i}}=0$. As $m_i=m_{d_i}+m_{s_i}+c'm_{d_i}$ where the new edges $m_{new_i}=c'm_{d_i}$ depend on the joining process, the edge density of $G_{n_i}$ is given by

$$\begin{aligned}
\text{density}(G_{n_i}) &= \frac{2(m_{d_i}+m_{s_i}+c'm_{d_i})}{(n_{d_i}+n_{s_i})^2}\,, \\
&= \frac{2((1+c')m_{d_i}+m_{s_i})}{n_{d_i}^2+2n_{d_i}n_{s_i}+n_{s_i}^2}\,, \\
&= 2\frac{(1+c')\frac{m_{d_i}}{n_{s_i}^2}+\frac{m_{s_i}}{n_{s_i}^2}}{\frac{n_{d_i}^2}{n_{s_i}^2}+2\frac{n_{d_i}}{n_{s_i}}+1} \\
&= 2\frac{(1+c')\frac{m_{d_i}}{n_{d_i}^2}\frac{n_{d_i}^2}{n_{s_i}^2}+\frac{m_{s_i}}{n_{s_i}^2}}{\frac{n_{d_i}^2}{n_{s_i}^2}+2\frac{n_{d_i}}{n_{s_i}}+1} \\
&= 2\frac{(1+c')\rho_{d_i}\frac{n_{d_i}^2}{n_{s_i}^2}+\frac{m_{s_i}}{n_{s_i}^2}}{\frac{n_{d_i}^2}{n_{s_i}^2}+2\frac{n_{d_i}}{n_{s_i}}+1} \\
\lim_{i\to\infty}\text{density}(G_{n_i}) &= 0\,,
\end{aligned}$$

as $\{G_{s_i}\}_i$ is sparse and as $\lim_{i\to\infty}\frac{n_{d_i}}{n_{s_i}}=0$.

□

**Lemma B.3.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U,W)$-mixture graphs (Definition 3.2) with dense and sparse parts of $G_{n_i}$ being $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p}=(p_1,p_2,\ldots)$ be the mass-partition (Definition 3.1) associated with $U$ and suppose $\boldsymbol{p}$ has $k$ non-zero elements. Recall $G_{s_i}$ is a union of disjoint stars and isolated edges. Let $\tilde{q}_{j,i}$ be the degree of the star in $G_{s_i}$ corresponding to $p_j\neq 0$. Then, if we consider $G_{n_i}$ to be the disjoint union of $G_{d_i}$ and $G_{s_i}$ the top $k$ degrees denoted by deg $v_{(1\ldots k)}$ satisfies*

$$\lim_{i\to\infty}P\left(\bigcup_{j=1}^{k}\tilde{q}_{j,i}\subseteq deg\ v_{(1\ldots k)}\right)=1\,.$$

*Proof.* From Lemma 3.4 we know if $\{G_{n_i}\}_i$ is sparse, then $n_{s_i}/n_{d_i} \to \infty$ as $i$ tends to infinity. Furthermore, $n_i = n_{d_i} + n_{s_i}$. In the dense part $G_{d_i}$, vertices have degree less than or equal to $n_{d_i}$.

From Lemma B.4 the degree of the star in $G_{s_i}$ corresponding to $p_j$ satisfies $\mathbb{E}(\tilde{q}_j) = m_{s_i} p_j$. Then using Chernoff bounds

$$P\left(\tilde{q}_j \leq (1-\epsilon) m_{s_i} p_j\right) \leq \exp\left(-\frac{m_{s_i} p_j \epsilon^2}{2}\right).$$

The sparse part $G_{s_i}$ has $n_{s_i}$ nodes and $m_{s_i}$ edges and is the inverse line graph of $H_{s_i} \sim \mathbb{G}(U, m_{s_i})$. As the inverse line graph of a clique is a star and the inverse line graph of an isolated node is an isolated edge, we have $m_{s_i} \in \Theta(n_{s_i})$. This gives

$$m_{s_i} p_j \geq c_1 p_j n_{s_i} > n_{d_i}.$$

for large enough $i$ as $n_{s_i}/n_{d_i} \to \infty$. Thus, by letting $n_{d_i} \leq (1-\epsilon) m_{s_i} p_j$ we get

$$P\left(\tilde{q}_j \leq n_{d_i}\right) \leq \exp\left(-\frac{m_{s_i} p_j}{2}\left(1 - \frac{n_{d_i}}{m_{s_i} p_j}\right)^2\right),$$

which goes to zero fast as $n_{d_i}/(m_{s_i} p_j)$ goes to zero and as $m_{s_i}$ goes to infinity. For $j \in \{1, \ldots, k\}$ if all $\tilde{q}_{j,i} > n_{d_i}$ then $\bigcup_{j=1}^{k} \tilde{q}_{j,i} \subseteq \deg v_{(1\ldots k)}$ for large $i$. Then

$$\begin{aligned}
P\left(\bigcup_{j=1}^{k} \tilde{q}_{j,i} \subseteq \deg v_{(1\ldots k)}\right) &= P\left((\tilde{q}_{1,i} > n_{d_i}) \cap (\tilde{q}_{2,i} > n_{d_i}) \cdots (\tilde{q}_{k,i} > n_{d_i})\right), \\
&\geq 1 - \sum_{j=1}^{k} P\left(\tilde{q}_{j,i} \leq n_{d_i}\right), \\
&\geq 1 - \sum_{j=1}^{k} \exp\left(-\frac{m_{s_i} p_j}{2}\left(1 - \frac{n_{d_i}}{m_{s_i} p_j}\right)^2\right),
\end{aligned}$$

giving the result.

□

### B.6 When $\sum_i p_i < 1$

If we focus on $U$ in the $(U, W)$-mixture graphs, we see that $U \to H_m \to G_s = L^{-1}(H_m)$. The graphon $U$ is used to generate a disjoint clique graph $H_m$, of which the inverse line graph is computed to obtain $G_s$, the sparse part. Suppose $\boldsymbol{p}$ is the mass-partition corresponding to $U$. If $\sum_j p_j = 1$, then $H_m$ consists of disjoint cliques and $G_s$ consists of disjoint stars. If $\sum_j p_j < 1$, then $H_m$ consists of disjoint cliques including isolated vertices, which are technically cliques of size 1. The inverse line graph of isolated vertices are isolated edges, which are technically stars of 2 vertices, $K_{1,1}$. This is illustrated in Figure 21.

**Lemma B.4.** *Let $\{G_{n_i}\}_i$ be a sequence of $(U, W)$-mixture graphs (Definition 3.2) with $G_{n_i} \sim \mathbb{G}(U, W, n_{d_i}, m_{s_i})$ with sparse part $G_s$. Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) corresponding to $U$. Then every non-zero $p_j$ results in a star $K_{1,\tilde{q}_j}$ in $G_s$ with degree of the hub-node $\tilde{q}_{j,i}$ satisfying*

$$\mathbb{E}(\tilde{q}_{j,i}) = m_{s_i} p_j \quad \textit{and} \quad Var(\tilde{q}_{j,i}) = m_{s_i} p_j (1 - p_j).$$

*If $\sum_j p_j < 1$ then $G_s$ has isolated edges in addition to disjoint stars. Let $\tilde{I}_i$ denote the number of isolated edges in $G_s$. Then*

$$\mathbb{E}(\tilde{I}_i) = m_{s_i}\left(1 - \sum_j p_j\right) \quad \textit{and} \quad Var(\tilde{I}_i) = m_{s_i} \sum_j p_j \left(1 - \sum_j p_j\right).$$

*Proof.* For a $(U, W)$-mixture graph $G_{n_i}$, the sparse part is given by $G_{s_i}$. For the sparse part, we draw $m_s$ points $x_1, x_2, \ldots, x_{m_s}$ uniformly from $[0, 1]$ (see Definition 3.2). The vertices $k$ and $\ell$ are connected with probability $U(x_k, x_\ell)$. This graph is called $H_s$. As $U$ is a disjoint clique graphon it can be represented by a mass-partition

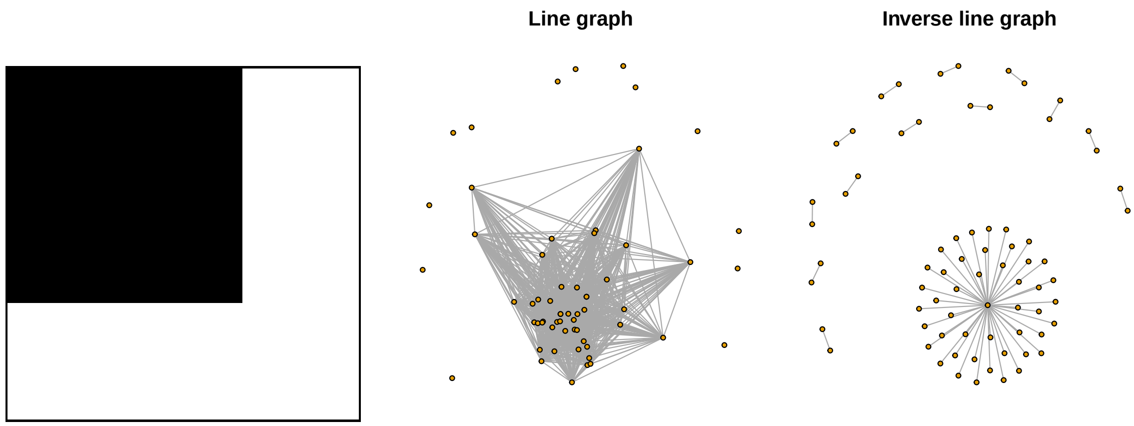



Figure 21: A disjoint clique graphon $U$ on the left, graph $H_m \sim \mathbb{G}(U, m)$ in the middle and $G_n = L^{-1}(H_m)$ on the right. Graph $H_m$ has a large clique and some isolated nodes. Graph $G_n$ is the inverse line graph of $H_m$ and has a star and isolated edges. The star is the inverse line graph of the large clique and isolated edges are the inverse of isolated vertices.

$\boldsymbol{p} = (p_1, p_2, \ldots)$. Thus, $H_s$ is a collection of disjoint cliques. As in Definition A.7 let $C_j$ denote the component size or the number of nodes in the $j$th largest component. As $x_1, x_2, \ldots, x_{m_s}$ are uniformly drawn, for non-zero $p_j$, the component size $C_j(H_s)$ has a binomial distribution with parameters $m_s$ and $p_j$. Thus for non-zero $p_j$

$$\mathbb{E}(C_j(H_s)) = m_{s_i} p_j \, ,$$

and

$$\mathrm{Var}(C_j(H_s)) = m_{s_i} p_j (1 - p_j)$$

where expectation is computed with respect to different randomizations. When we compute the inverse line graph of $L^{-1}(H_s)$, each clique corresponding to non-zero $p_j$ is converted to a star $K_{1,\tilde{q}_j}$ such that the degree of the hub node is equal to the number of nodes in the clique, i.e., $\tilde{q}_{j,i} = C_j(H_s)$, giving us

$$\mathbb{E}(\tilde{q}_{j,i}) = m_{s_i} p_j \quad \text{and} \quad \mathrm{Var}(\tilde{q}_{j,i}) = m_{s_i} p_j (1 - p_j) \, .$$

When $\sum_j p_j < 1$, this results in isolated nodes in $H_s$. The number of isolated nodes $H_s$ is equal to $\tilde{I}_i$, the number of isolated edges in $G_s$ as $G_s$ is the inverse line graph of $H_s$. As $x_1, x_2, \ldots, x_{m_s}$ are uniformly drawn from $[0, 1]$, $\tilde{I}_i$ has a binomial distribution with parameters $m_{s_i}$ and $p_0 = 1 - \sum_j p_j$ making

$$\mathbb{E}(\tilde{I}_i) = m_{s_i} p_0 = m_{s_i} \left( 1 - \sum_j p_j \right) \quad \text{and} \quad \mathrm{Var}(\tilde{I}_i) = m_{s_i} p_0 (1 - p_0) = m_{s_i} \sum_j p_j \left( 1 - \sum_j p_j \right) .$$

□

## C Predicting top-$k$ degrees

**Lemma C.1.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) associated with $U$ and suppose $p_j > p_k \neq 0$. Let $\tilde{q}_{j,i}$ and $\tilde{q}_{k,i}$ denote the degree of star vertices in $G_{s_i}$ corresponding to $p_j$ and $p_k$ respectively. Let $q_{j,i}$ and $q_{k,i}$ denote their degrees in $G_{n_i}$. Then*

$$P\left(q_{k,i} > q_{j,i}\right) \leq \frac{C}{m_{s_i}}$$

*That is, with high probability the graph mixture does not change the order of the degree of star vertices.*

*Proof.* Let $G_{d_i}, G_{s_i}$ and $G_{n_i}$ have $n_i$, $n_{d_i}$ and $n_{s_i}$ number of nodes and $m_i$, $m_{d_i}$ and $m_{s_i}$ number of edges respectively. Let $m_{new_i}$ denote the number of edges added as part of the joining process. From Lemma B.2 we know that

$$\begin{aligned}\mathbb{E}(q_{j,i}) &= m_{s_i} p_j + \frac{c m_{new_i}}{n_{s_i}}\,, \\ \mathrm{Var}(q_{j,i}) &= m_{s_i} p_j (1 - p_j) + m_{new_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right).\end{aligned}$$

We define $X_i = q_{k,i} - q_{j,i}$, where $q_{k,i}$ and $q_{j,i}$ are the degrees of the hubs in $G_{n_i}$ corresponding to $p_k$ and $p_j$ respectively. Then

$$\begin{aligned}\mathbb{E}(X_i) &= m_{s_i}(p_k - p_j)\,, \\ \mathrm{Var}(X_i) &= m_{s_i}\left(p_j(1-p_j) + p_k(1-p_k)\right) + 2 m_{new_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right), \\ &= m_{s_i}\left(p_j(1-p_j) + p_k(1-p_k)\right) + 2c' m_{d_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right).\end{aligned}$$

As $m_{s_i} \in \Theta(n_{s_i})$, $m_{d_i} \in \Theta(n_{d_i}^2)$ and $n_{d_i}/n_{s_i} \to 0$ for sparse graphs we obtain

$$\frac{\mathrm{Var} X_i}{\mathbb{E}(X_i)^2} \leq \frac{C}{m_{s_i}}\,, \tag{7}$$

$$\text{making} \quad \lim_{i \to \infty} \frac{\mathrm{Var} X_i}{\mathbb{E}(X_i)^2} = 0\,,$$

As $p_j > p_k$, we have $\mathbb{E}(X_i) < 0$. Furthermore, the probability

$$\begin{aligned}P(X_i > 0) &= P(X_i - \mathbb{E}(X_i) \geq -\mathbb{E}(X_i))\,, \\ &= P(X_i - \mathbb{E}(X_i) \geq |\mathbb{E}(X_i)|) \quad \text{as } -\mathbb{E}(X_i) = |\mathbb{E}(X_i)|\,, \\ &\leq P(|X_i - \mathbb{E}(X_i)| \geq |\mathbb{E}(X_i)|)\,, \\ &\leq \frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2}\end{aligned}$$

where we have used Chebyshev's inequality in the last step and the fact that $P(|X| \geq a) \geq P(X \geq a)$ as the set $X \geq a$ is a subset of the set $|X| \geq a$. Combining with equation (7) we get

$$P(q_{k,i} > q_{j,i}) \leq \frac{C}{m_{s_i}}\,,$$

making

$$\lim_{i \to \infty} P(q_{k,i} > q_{j,i}) = 0\,.$$

□

**Lemma C.2.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) associated with $U$ and suppose $p_k \neq 0$. Let $v_\ell$ be a vertex originally from the dense part $G_{d_i}$ and let us denote the degree of $v_\ell$ in $G_{d_i}$ by $deg_{\,G_{d_i}} v_\ell$ and its degree in $G_{n_i}$ by $deg_{\,G_{n_i}} v_\ell$. Then*

$$P\left(deg_{\,G_{n_i}} v_\ell \geq (1-\epsilon) m_{s_i} p_k\right) \leq \exp\left(-C \frac{m_{s_i}^2}{n_{d_i}^2}\right)$$

*where $\epsilon > 0$ and $m_{s_i}$ denotes the number of edges in the sparse part. That is, the probability of a node in the dense part having a larger degree than the expected degree of a star node $G_{n_i}$ goes to zero.*

*Proof.* Recall the joining process (Definition 3.3) does not delete or add vertices. Hence each vertex in $G_{n_i}$ comes either from the dense part or from the sparse part. Suppose vertex $v_\ell$ comes from the dense part $G_{d_i}$. Let us denote the degree of $v_\ell$ in $G_{d_i}$ by $\deg_{G_{d_i}} v_\ell$ and denote its degree in $G_{n_i}$ by $\deg_{G_{n_i}} v_\ell$. Then from Lemma B.1 we know that

$$\mathbb{E}(\deg_{G_{n_i}} v_\ell) = \deg_{G_{d_i}} v_\ell + c\frac{m_{new_i}}{n_{d_i}} \leq \deg_{G_{d_i}} v_\ell + c'\frac{m_{d_i}}{n_{d_i}} \, .$$

Rewriting $\mu_{\ell,i} = \mathbb{E}(\deg_{G_{n_i}} v_\ell)$ for $\delta > 0$ we have the following Chernoff bound

$$P\left(\deg_{G_{n_i}} v_\ell \geq (1+\delta)\mu_{\ell,i}\right) \leq \exp\left(-\frac{\mu_{\ell,i}\delta^2}{3}\right) \, .$$

From Lemma B.4 we know $\mathbb{E}(\tilde{q}_{k,i}) = m_{s_i} p_k$ where $\tilde{q}_{k,i}$ denotes the degree of the hub vertex corresponding to $p_k$ in $G_{s_i}$. Then as $n_{s_i}/n_{d_i}$ goes to infinity for sparse graphs, for a fixed $\epsilon > 0$ we have $(1-\epsilon)m_{s_i}p_k >> (1+\delta)\mu_{\ell,i}$ for increasing $i$. Hence, the probability of $\deg_{G_{n_i}} v_\ell \geq (1-\epsilon)m_{s_i}p_k$ decreases for a fixed $\epsilon > 0$. Thus, for $\delta \leq (1-\epsilon)m_{s_i}p_k/\mu_{\ell,i} - 1$ we obtain

$$P\left(\deg_{G_{n_i}} v_\ell \geq (1-\epsilon)m_{s_i}p_k\right) \leq \exp\left(-\frac{\mu_{\ell,i}}{3}\left(\frac{(1-\epsilon)m_{s_i}p_k}{\mu_{\ell,i}} - 1\right)^2\right)$$

As $\mu_{\ell,i} \in \mathcal{O}(n_{d_i})$, $m_{s_i} \in \times(n_{s_i})$ and $n_{s_i}/n_{d_i}$ goes to infinity for sparse graphs the expression $\left(\frac{(1-\epsilon)m_{s_i}p_k}{\mu_{\ell,i}} - 1\right)^2$ goes to infinity. Therefore, the probability $P\left(\deg_{G_{n_i}} v_\ell \geq (1-\epsilon)m_{s_i}p_k\right)$ goes to zero satisfying

$$P\left(\deg_{G_{n_i}} v_\ell \geq (1-\epsilon)m_{s_i}p_k\right) \leq \exp\left(-C\frac{m_{s_i}^2}{n_{d_i}^2}\right) \, .$$

□

**Lemma C.3.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U,W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p} = (p_1, p_2, \dots)$ be the mass-partition (Definition 3.1) associated with $U$. Let $\tilde{q}_{j,i}$ be the degree of the star in $G_{s_i}$ corresponding to $p_j \neq 0$. Let $q_{j,i}$ denote the degree of the corresponding vertex in $G_{n_i}$. Let $v_\ell$ be a vertex originally from the dense part $G_{d_i}$ and let us denote the degree of $v_\ell$ in $G_{d_i}$ by deg $_{G_{d_i}} v_\ell$ and its degree in $G_{n_i}$ by deg $_{G_{n_i}} v_\ell$. Then*

$$P\left(\deg_{G_{n_i}} v_\ell \geq q_{j,i}\right) \leq \exp\left(-c\frac{m_{s_i}^2}{n_{d_i}^2}\right) + \exp(-c' m_{s_i})$$

*That is, the probability of a node originally in the dense part having a larger degree than that of a star node in $G_{n_i}$ goes to zero.*

*Proof.* Let $G_{d_i}, G_{s_i}$ and $G_{n_i}$ have $n_i$, $n_{d_i}$ and $n_{s_i}$ number of nodes and $m_i$, $m_{d_i}$ and $m_{s_i}$ number of edges respectively. Then for $\epsilon > 0$

$$\begin{aligned}
P\left(\deg_{G_{n_i}} v_\ell \geq q_{j,i}\right) &= P\left(\deg_{G_{n_i}} v_\ell \geq q_{j,i} | q_{j,i} \geq (1-\epsilon)m_{s_i}p_j\right) P\left(q_{j,i} \geq (1-\epsilon)m_{s_i}p_j\right) \\
&\quad + P\left(\deg_{G_{n_i}} v_\ell \geq q_{j,i} | q_{j,i} \leq (1-\epsilon)m_{s_i}p_j\right) P\left(q_{j,i} \leq (1-\epsilon)m_{s_i}p_j\right) \, , \\
&\leq P\left(\deg_{G_{n_i}} v_\ell \geq (1-\epsilon)m_{s_i}p_j\right) P\left(q_{j,i} \geq (1-\epsilon)m_{s_i}p_j\right) \\
&\quad + P\left(\deg_{G_{n_i}} v_\ell \geq q_{j,i} | q_{j,i} \leq (1-\epsilon)m_{s_i}p_j\right) P\left(q_{j,i} \leq (1-\epsilon)m_{s_i}p_j\right) \, , \\
&\leq P\left(\deg_{G_{n_i}} v_\ell \geq (1-\epsilon)m_{s_i}p_j\right) + P\left(q_{j,i} \leq (1-\epsilon)m_{s_i}p_j\right) \qquad (8)
\end{aligned}$$

From Lemma C.2 we know that

$$P\left(\deg_{G_{n_i}} v_\ell \geq (1-\epsilon)m_{s_i}p_j\right) \leq \exp\left(-C\frac{m_{s_i}^2}{n_{d_i}^2}\right) \, .$$

From Lemma B.2 we know

$$\mathbb{E}(q_{j,i}) = m_{s_i} p_j + \frac{c m_{new_i}}{n_{s_i}} \, .$$

Using Chernoff bounds we get

$$P\left(q_{j,i} \leq (1-\epsilon) m_{s_i} p_j\right) \leq \exp\left( - \frac{m_{s_i} p_j \epsilon^2}{2} \right) \, .$$

Substituting in equation (8) we get

$$P\left(\deg_{G_{n_i}} v_\ell \geq q_{j,i}\right) \leq \exp\left( -C \frac{m_{s_i}^2}{n_{d_i}^2} \right) + \exp\left(-c' m_{s_i}\right) \, .$$

□

**Proposition 4.1.** *(**Order Preserving Property**) Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) associated with $U$ which has at least $k$ partitions. Let $\tilde{q}_{j,i}$ be the degree of the star in $G_{s_i}$ corresponding to $p_j \neq 0$. Let $q_{j,i}$ denote the degree of the corresponding vertex in $G_{n_i}$. Let $deg_{G_{n_i}} v_{(r)}$ denote the $r$th highest degree in $G_{n_i}$. Then*

$$P\left( \bigcap_{j=1}^{k} \left( q_{j,i} = deg_{G_{n_i}} v_{(j)} \right) \right) \geq \left(1 - \frac{c_1}{m_{s_i}}\right)^k \left( 1 - \exp\left( -c_2 \frac{m_{s_i}^2}{n_{d_i}^2} \right) - \exp\left(-c_3 m_{s_i}\right) \right)$$

*That is, with high probability the order of the stars in the sparse part are preserved by joining.*

*Proof.* Let us explore the condition $q_{j,i} = \deg_{G_{n_i}} v_{(j)}$ for $j \in [k]$. This can only happen when $q_{1,i} > q_{2,i} > \cdots > q_{k,i}$ and when all other nodes have degrees lower than $q_{k,i}$. Let us denote the maximum degree of the other nodes by $\deg_{G_{n_i}} v_u$. The probability

$$\begin{aligned} P\left( \bigcap_{j=1}^{k} q_{j,i} = \deg_{G_{n_i}} v_{(j)} \right) &= P\left(q_{1,i} > q_{2,i} > \cdots > q_{k,i}\right) \times P\left( \deg_{G_{n_i}} v_u < q_{k,i} \right) \, , \\ &= \prod_{j=1}^{k-1} P\left(q_{j,i} > q_{j+1,i}\right) \times P\left( \deg_{G_{n_i}} v_u < q_{k,i} \right) \, . \end{aligned} \tag{9}$$

From Lemma C.1 we know that

$$P\left(q_{j+1,i} > q_{j,i}\right) \leq \frac{C}{m_{s_i}} \, ,$$

making

$$P\left(q_{j,i} > q_{j+1,i}\right) \geq 1 - \frac{C}{m_{s_i}} \, . \tag{10}$$

Let $v_\ell$ be a vertex originally from the dense part $G_{d_i}$ and let us denote the degree of $v_\ell$ in $G_{d_i}$ by $\deg_{G_{d_i}} v_\ell$ and its degree in $G_{n_i}$ by $\deg_{G_{n_i}} v_\ell$. From Lemma C.3 we know that for vertices in the dense part

$$P\left(\deg_{G_{n_i}} v_\ell \geq q_{j,i}\right) \leq \exp\left( -C \frac{m_{s_i}^2}{n_{d_i}^2} \right) + \exp\left(-c' m_{s_i}\right) \, .$$

making

$$P\left(\deg_{G_{n_i}} v_\ell < q_{j,i}\right) \geq 1 - \exp\left( -C \frac{m_{s_i}^2}{n_{d_i}^2} \right) - \exp\left(-c' m_{s_i}\right) \, . \tag{11}$$

Suppose $U$ has $k$ partitions. Then the maximum degree of the "other" nodes is obtained by a vertex in the dense part. In addition to the vertices in the dense part, there are degree-1 vertices in the sparse part. These are the leaf or non-hub nodes in every star. Their degrees are much smaller than the rest. Thus, equation (11) holds for all "other" vertices including the vertex which has the maximum degree of these vertices. Combining equations (10) and (11) with equation (9) we obtain

$$
\begin{aligned}
P\left(\bigcap_{j=1}^{k}\left(q_{j,i} = \deg_{G_{n_i}} v_{(j)}\right)\right) &\geq \left(1 - \frac{c_1}{m_{s_i}}\right)^{k-1}\left(1 - \exp\left(-c_2\frac{m_{s_i}^2}{n_{d_i}^2}\right) - \exp\left(-c_3 m_{s_i}\right)\right), \\
&> \left(1 - \frac{c_1}{m_{s_i}}\right)^{k}\left(1 - \exp\left(-c_2\frac{m_{s_i}^2}{n_{d_i}^2}\right) - \exp\left(-c_3 m_{s_i}\right)\right),
\end{aligned}
$$

where $c_1$ denotes the largest $C$ in equation (10) for different $j \leq k-1$.

Suppose $U$ has either infinite partitions or $\ell$ finite partitions with $\ell > k$. Then the probability

$$
P\left(\bigcap_{j=1}^{k} q_{j,i} = \deg_{G_{n_i}} v_{(j)}\right) = P\left(q_{1,i} > q_{2,i} > \cdots > q_{k,i}\right) \times P\left(q_{k,i} > \max_{j\in\{k+1,\dots\}} q_{j,i}\right) \times P\left(\deg_{G_{n_i}} v_u < q_{k,i}\right),
$$

where $\max_{j\in\{k+1,\dots\}} q_{j,i}$ denotes the maximum degree of other hub nodes contributed by $U$. Similar to equation (10) the probability

$$
P\left(q_{k,i} > \max_{j\in\{k+1,\dots\}} q_{j,i}\right) \geq 1 - \frac{C}{m_{s_i}}
$$

making

$$
P\left(\bigcap_{j=1}^{k}\left(q_{j,i} = \deg_{G_{n_i}} v_{(j)}\right)\right) \geq \left(1 - \frac{c_1}{m_{s_i}}\right)^{k}\left(1 - \exp\left(-c_2\frac{m_{s_i}^2}{n_{d_i}^2}\right) - \exp\left(-c_3 m_{s_i}\right)\right).
$$

□

**Lemma 4.2.** *Suppose $G_{n_i}$ and $G_{n_j}$ are two graphs from a sparse $(U, W)$-mixture graph sequence (Definition 3.2). We treat $G_{n_i}$ as the training graph and $G_{n_j}$ as the test graph. Suppose $G_{n_i}$ and $G_{n_j}$ have $n_i$ and $n_j$ nodes respectively. Let $\boldsymbol{p} = (p_1, p_2, \dots)$ be the mass-partition (Definition 3.1) associated with $U$, which has at least $k$ partitions. Let $deg_{G_{n_i}} v_{(\ell)}$ denote the $\ell$th largest degree in $G_{n_i}$ where $\ell \leq k$. Then we estimate the $\ell$th largest degree in $G_{n_j}$ as*

$$
deg_{G_{n_j}} \hat{v}_{(\ell)} = deg_{G_{n_i}} v_{(\ell)} \times \frac{n_j}{n_i}, \tag{1}
$$

*which satisfies*

$$
\frac{\left|deg_{G_{n_j}} \hat{v}_{(\ell)} - \mathbb{E}(q_{\ell,j})\right|}{m_{s_j}} \leq c p_\ell \left|1 - \frac{n_j m_{s_i}}{n_i m_{s_j}}\right| \tag{2}
$$

*with high probability, where $q_{\ell,j}$ denotes the degree of the hub vertex corresponding to $p_\ell \neq 0$ in $G_{n_j}$ and $m_{s_i}$ and $m_{s_j}$ denote the number of edges in the sparse parts $G_{s_i}$ and $G_{s_j}$.*

*Proof.* For large $i$ and $j$ with high probability the expression

$$
\begin{aligned}
\frac{\left|\deg_{G_{n_j}} \hat{v}_{(\ell)} - \mathbb{E}(q_{\ell,j})\right|}{m_{s_j}} &= \frac{1}{m_{s_j}}\left|\deg_{G_{n_i}} v_{(\ell)}\frac{n_j}{n_i} - \left(m_{s_j} p_\ell + \frac{c m_{new_j}}{n_{s_j}}\right)\right|, \\
&= \frac{1}{m_{s_j}}\left|q_{\ell,i}\frac{n_j}{n_i} - \left(m_{s_j} p_\ell + \frac{c m_{new_j}}{n_{s_j}}\right)\right|,
\end{aligned}
$$

where we have used equation (1) and Lemma B.2 for $\mathbb{E}(q_{\ell,j})$ in the first line and the Order Preserving Property to say $\deg_{G_{n_i}} v_{(\ell)} = q_{\ell,i}$ in the second line for large $i$. From Lemma B.2 we can see that

$$
\lim_{i\to\infty} \frac{\mathrm{Var}(q_{\ell,i})}{\mathbb{E}(q_{\ell,i})^2} = 0,
$$

implying that the observed values of $q_{\ell,i}$ lie close to its expected value for increasing $i$. Therefore,

$$\begin{aligned}
\frac{\left|\deg_{G_{n_j}} \hat{v}_{(\ell)} - \mathbb{E}(q_{\ell,j})\right|}{m_{s_j}} &\approx \frac{1}{m_{s_j}} \left| \left( m_{s_i} p_\ell + \frac{c m_{new_i}}{n_{s_i}} \right) \frac{n_j}{n_i} - \left( m_{s_j} p_\ell + \frac{c m_{new_j}}{n_{s_j}} \right) \right| , \\
&= \frac{n_j}{m_{s_j}} \left| \left( \frac{m_{s_i} p_\ell}{n_i} + \frac{c m_{new_i}}{n_{s_i} n_i} \right) - \left( \frac{m_{s_j} p_\ell}{n_j} + \frac{c m_{new_j}}{n_{s_j} n_j} \right) \right| , \\
&\leq c_1 \frac{n_j}{m_{s_j}} \left| \frac{m_{s_i} p_\ell}{n_i} - \frac{m_{s_j} p_\ell}{n_j} \right| , \\
&= c_1 \frac{n_j p_\ell}{m_{s_j}} \left| \frac{m_{s_i}}{n_i} - \frac{m_{s_j}}{n_j} \right| , \\
&= c_1 p_\ell \left| 1 - \frac{n_j m_{s_i}}{n_i m_{s_j}} \right| .
\end{aligned} \tag{12}$$

where we have used $\frac{c m_{new_i}}{n_{s_i} n_i}$ goes to zero. This is because $m_{new_i} \in \Theta(n_{d_i}^2)$ as a result of the joining process (Definition 3.3) and $n_{s_i}/n_i$ goes to 1 as $n_{d_i}/n_{s_i}$ goes to zero for sparse graphs.

The number of nodes and edges in the sparse part are linked by

$$m_{s_i} + k_i = n_{s_i}$$

where $k_i$ denotes the number of stars in the sparse part $G_{s_i}$ and $n_{s_i}$ denote the number of nodes. If $U$ has finite partitions, then $k_i$ is bounded for all large $i$. However, if $U$ has infinite partitions then the number of stars observed $k_i$ will increase with increasing $i$. Nevertheless, $k_i/n_{s_i}$ goes to zero. This is because the number of nodes in a collection of stars grows much faster than the number of stars when we uniformly sample from the disjoint clique graphon $U$. Thus we have

$$\lim_{i \to \infty} \frac{m_{s_i}}{n_{s_i}} = 1 . \tag{13}$$

Let $n_{d_i}$ denote the number of nodes in the dense part. Then

$$\begin{aligned}
n_{s_i} + n_{d_i} &= n_i , \\
\frac{n_{s_i}}{n_i} + \frac{n_{d_i}}{n_i} &= 1 .
\end{aligned}$$

As $n_{s_i}/n_{d_i}$ goes to infinity for sparse graphs we have $n_{d_i}/n_i$ going to zero making $n_{s_i}/n_i$ tending to one. Combining with (13) we get

$$\lim_{i \to \infty} \frac{m_{s_i}}{n_i} = 1 .$$

Substituting this in equation (12) , we get

$$\lim_{i,j \to \infty} \frac{\left|\deg_{G_{n_j}} \hat{v}_{(\ell)} - \mathbb{E}(q_{\ell,j})\right|}{m_{s_j}} = 0 .$$

□

## D Estimating $U$ for finite partitions

Using the Taylor series expansion, we derive the second order approximation of the expectation of a ratio of random variables and the first order approximation for the variance.

**Lemma D.1.** *(Stuart and Ord (2010)) Let $X$ and $Y$ be two random variables where $Y$ has no mass at zero. Then the second order Taylor series approximation of $\mathbb{E}(X/Y)$ is given by*

$$\mathbb{E}\left(\frac{X}{Y}\right) = \frac{\mathbb{E}(X)}{\mathbb{E}(Y)} - \frac{Cov(X,Y)}{\mathbb{E}(Y)^2} + \frac{\mathbb{E}(X)\, Var(Y)}{\mathbb{E}(Y)^3} ,$$

*and the first order Taylor approximation of Var(X/Y) is given by*

$$Var\left(\frac{X}{Y}\right) = \frac{\mathbb{E}(X)^2}{\mathbb{E}(Y)^2}\left(\frac{Var(X)}{\mathbb{E}(X)^2} - 2\frac{Cov(X,Y)}{\mathbb{E}(X)\mathbb{E}(Y)} + \frac{Var(Y)}{\mathbb{E}(Y)^2}\right).$$

*Proof.* For any $f(x, y)$, the second order Taylor expansion about $\boldsymbol{a} = (a_x, a_y)$ is given by

$$\begin{aligned} f(x,y) &= f(\boldsymbol{a}) + f'_x(\boldsymbol{a})(x-a_x) + f'_y(\boldsymbol{a})(y-a_y) \\ &+ \frac{1}{2}f''_{xx}(\boldsymbol{a})(x-a_x)^2 + f''_{xy}(\boldsymbol{a})(x-a_x)(y-a_y) + \frac{1}{2}f''_{yy}(\boldsymbol{a})(y-a_y)^2 + R, \end{aligned}$$

Let $\mathbb{E}(X) = \mu_x,\ \mathbb{E}(Y) = \mu_y,\ \ f(X,Y) = X/Y$ and $\boldsymbol{a} = (\mu_x, \mu_y)$. Then the expectation of the second order approximation is given by

$$\begin{aligned} \mathbb{E}(f(X,Y)) &= \mathbb{E}\left(f(\boldsymbol{a}) + f'_x(\boldsymbol{a})(X-a_x) + f'_y(\boldsymbol{a})(Y-a_y)\right), \\ &+ \mathbb{E}\left(\frac{1}{2}f''_{xx}(\boldsymbol{a})(X-a_x)^2 + f''_{xy}(\boldsymbol{a})(X-a_x)(Y-a_y) + \frac{1}{2}f''_{yy}(\boldsymbol{a})(Y-a_y)^2\right), \\ &= f(\boldsymbol{a}) + f'_x(\boldsymbol{a})\mathbb{E}(X-\mu_x) + f'_y(\boldsymbol{a})\mathbb{E}(Y-\mu_y) \\ &+ \frac{1}{2}f''_{xx}\mathrm{Var}(X) + f''_{xy}(\boldsymbol{a})\mathrm{Cov}(X,Y) + \frac{1}{2}f''_{yy}\mathrm{Var}(Y), \\ &= f(\boldsymbol{a}) + \frac{1}{2}f''_{xx}(\boldsymbol{a})\mathrm{Var}(X) + f''_{xy}(\boldsymbol{a})\mathrm{Cov}(X,Y) + \frac{1}{2}f''_{yy}(\boldsymbol{a})\mathrm{Var}(Y) \end{aligned} \tag{14}$$

as $\mathbb{E}(X-\mu_x) = \mathbb{E}(X-\mu_y) = 0$. As $f''_{xx} = 0$, $f''_{xy} = -\frac{1}{y^2}$ and $f''_{yy} = \frac{2x}{y^3}$ we obtain

$$\mathbb{E}\left(\frac{X}{Y}\right) = \frac{\mathbb{E}(X)}{\mathbb{E}(Y)} - \frac{\mathrm{Cov}(X,Y)}{\mathbb{E}(Y)^2} + \frac{\mathbb{E}(X)\mathrm{Var}(Y)}{\mathbb{E}(Y)^3}.$$

The variance is defined as

$$\mathrm{Var}(f(X,Y)) = \mathbb{E}\left([f(X,Y) - \mathbb{E}(f(X,Y))]^2\right)$$

From equation (14) we let

$$\mathbb{E}(f(X,Y)) \approx f(\boldsymbol{a})$$

and get

$$\mathrm{Var}(f(X,Y)) \approx \mathbb{E}\left([f(X,Y) - f(\boldsymbol{a})]^2\right).$$

Then we use the first order Taylor expansion for $f(X,Y)$ around $\boldsymbol{a}$ inside the expectation term

$$\begin{aligned} \mathrm{Var}(f(X,Y)) &\approx \mathbb{E}\left(\left[f(\boldsymbol{a}) + f'_x(\boldsymbol{a})(x-a_x) + f'_y(\boldsymbol{a})(y-a_y) - f(\boldsymbol{a})\right]^2\right), \\ &= \mathbb{E}\left(\left[f'_x(\boldsymbol{a})(X-a_x) + f'_y(\boldsymbol{a})(Y-a_y)\right]^2\right), \\ &= \mathbb{E}\left(f'^2_x(\boldsymbol{a})(X-a_x)^2 + 2f'_x(\boldsymbol{a})(X-a_x)f'_y(\boldsymbol{a})(Y-a_y) + f'^2_y(\boldsymbol{a})(Y-a_y)^2\right), \\ &= f'^2_x(\boldsymbol{a})\mathrm{Var}(X) + 2f'_x(\boldsymbol{a})f'_y(\boldsymbol{a})\mathrm{Cov}(X,Y) + f'^2_y(\boldsymbol{a})\mathrm{Var}(Y). \end{aligned}$$

For $f(X,Y) = X/Y$ we have $f'(X) = 1/Y$, $f'(Y) = -\frac{X}{Y^2}$, $f'^2_x(\boldsymbol{a}) = 1/\mu_y^2$, $f'^2_y(\boldsymbol{a}) = \frac{\mu_x^2}{\mu_y^4}$ and $f'_x(\boldsymbol{a})f'_y(\boldsymbol{a}) = -\frac{\mu_x}{\mu_y^3}$. Substituting these we obtain

$$\begin{aligned} \mathrm{Var}(f(X,Y)) &\approx \frac{1}{\mu_y^2}\mathrm{Var}(X) - 2\frac{\mu_x}{\mu_y^3}\mathrm{Cov}(X,Y) + \frac{\mu_x^2}{\mu_y^4}\mathrm{Var}(Y), \\ &= \frac{\mu_x^2}{\mu_y^2}\left(\frac{\mathrm{Var}(X)}{\mu_x^2} - 2\frac{\mathrm{Cov}(X,Y)}{\mu_x\mu_y} + \frac{\mathrm{Var}(Y)}{\mu_y^2}\right), \\ &= \frac{\mathbb{E}(X)^2}{\mathbb{E}(Y)^2}\left(\frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2} - 2\frac{\mathrm{Cov}(X,Y)}{\mathbb{E}(X)\mathbb{E}(Y)} + \frac{\mathrm{Var}(Y)}{\mathbb{E}(Y)^2}\right). \end{aligned}$$

□

### D.1 Estimating $p_j$

**Proposition 5.6.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p} = (p_1, p_2, \dots)$ be the mass-partition (Definition 3.1) associated with $U$ with only $k$ partitions. Let $q_{j,i}$ denote the degree of the hub vertex corresponding to $p_j \neq 0$ in $G_{n_i}$. Then using the second order approximation of the Taylor expansion for the expectation and the first order approximation for the variance, with $Q_{k,i} = \sum_{\ell=1}^{k} q_{\ell,i}$ and for some constants $c_1$ and $c_2$ we have*

$$\left| \mathbb{E}\left( \frac{q_{j,i}}{Q_{k,i}} \right) - p_j \right| \leq \frac{c_1}{m_{s_i}} \quad \textit{and} \quad \textit{Var}\left( \frac{q_{j,i}}{Q_{k,i}} \right) \leq \frac{c_2}{m_{s_i}} .$$

*Proof.* As $\tilde{q}_{j,i}$ is the degree of the star in $G_{s_i}$ corresponding to $p_j \neq 0$ and as $q_{j,i}$ denotes the degree of the corresponding vertex in $G_{n_i}$ we have

$$q_{j,i} = \tilde{q}_{j,i} + m_{j,new_i}$$

where $m_{j,new_i}$ denotes the number of edges added to that vertex as part of the joining process. Then

$$\sum_{\ell=1}^{k} q_{\ell,i} = \sum_{\ell=1}^{k} \left( \tilde{q}_{\ell,i} + m_{\ell,new_i} \right) = m_{s_i} + \sum_{\ell=1}^{k} m_{\ell,new_i} ,$$

as the sum of the edges in the $k$ stars equal $m_{s_i}$. This is because of two reasons: First we consider a mass-partition $\boldsymbol{p}$ with $k$ non-zero entries with $\sum_{\ell=1}^{k} p_\ell = 1$. That is $G_{s_i}$ has only $k$ stars. The second reason comes from the construction of $(U, W)$-mixture graphs (Definition 3.2). Recall we sample $m_{s_i}$ nodes from $U$ and construct a graph, which is a disjoint clique graph. Then we find its inverse line graph, which we call $G_{s_i}$. Consequently $G_{s_i}$ has $m_{s_i}$ edges, which are spread across $k$ stars.

The number of edges $m_{s_i}$, $m_{new_i}$ and the number of nodes $n_{s_i}$ are not random variables. They are parameters of the $(U, W)$-mixture graph process. The degree of each node is a random variable as sampling from the graphons are involved in determining the degree. Let $X_i = q_{j,i}$ and $Y_i = \sum_{\ell=1}^{k} q_{\ell,i}$.

Then from Lemma B.2

$$\begin{aligned}
\mathbb{E}(X_i) &= m_{s_i} p_j + \frac{c m_{new_i}}{n_{s_i}} , \\
\mathbb{E}(Y_i) &= m_{s_i} + \frac{k c m_{new_i}}{n_{s_i}} , \\
\mathrm{Var}(X_i) &= m_{s_i} p_j (1 - p_j) + m_{new_i} \frac{c}{n_{s_i}} \left( 1 - \frac{c}{n_{s_i}} \right) , \\
\mathrm{Var}(Y_i) &= \mathrm{Var}\left( \sum_{\ell=1}^{k} m_{j,new_i} \right) , \\
&= \sum_{\ell=1}^{k} \mathrm{Var}\left( m_{j,new_i} \right) , \\
&= k m_{new_i} \frac{c}{n_{s_i}} \left( 1 - \frac{c}{n_{s_i}} \right) .
\end{aligned}$$

The covariance

$$\begin{aligned}
\mathrm{Cov}(X_i, Y_i) &= \mathrm{Cov}\left( q_{j,i}, \sum_{\ell=1}^{k} q_{\ell,i} \right) , \\
&= \mathrm{Cov}\left( q_{j,i}, q_{j,i} + \sum_{\substack{\ell=1 \\ \ell \neq j}}^{k} q_{\ell,i} \right) , \\
&= \mathrm{Var}(q_{j,i}) , \\
&= m_{s_i} p_j (1 - p_j) + m_{new_i} \frac{c}{n_{s_i}} \left( 1 - \frac{c}{n_{s_i}} \right) .
\end{aligned}$$

From Lemma D.1 using the second order Taylor series approximation we have

$$\mathbb{E}\left(\frac{X_i}{Y_i}\right) = \frac{\mathbb{E}(X_i)}{\mathbb{E}(Y_i)} - \frac{\mathrm{Cov}(X_i, Y_i)}{\mathbb{E}(Y_i)^2} + \frac{\mathbb{E}(X_i)\mathrm{Var}(Y_i)}{\mathbb{E}(Y_i)^3} \,. \tag{15}$$

We compute each term separately and let $i$ go to infinity. The first term

$$\begin{aligned}
\frac{\mathbb{E}(X_i)}{\mathbb{E}(Y_i)} &= \frac{m_{s_i} p_j + \frac{c m_{new_i}}{n_{s_i}}}{m_{s_i} + \frac{k c m_{new_i}}{n_{s_i}}} \,, \\
&= \frac{p_j + \frac{c m_{new_i}}{n_{s_i} m_{s_i}}}{1 + \frac{k c m_{new_i}}{n_{s_i} m_{s_i}}} \,, \\
\lim_{i\to\infty} \frac{\mathbb{E}(X_i)}{\mathbb{E}(Y_i)} &= p_j \,,
\end{aligned} \tag{16}$$

as $m_{new_i} \in \mathcal{O}(n_{d_i}^2)$, $m_{s_i} \in \Theta(n_{s_i})$ and $n_{d_i} \in o(n_{s_i})$ the ratio $m_{new_i}/(n_{s_i} m_{s_i}) \to 0$. The second term

$$\begin{aligned}
\frac{\mathrm{Cov}(X_i, Y_i)}{\mathbb{E}(Y_i)^2} &= \frac{m_{s_i} p_j (1 - p_j) + m_{new_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right)}{\left(m_{s_i} + \frac{k c m_{new_i}}{n_{s_i}}\right)^2} \,, \\
&= \frac{m_{s_i} p_j (1 - p_j) + m_{new_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right)}{m_{s_i}^2 + 2 m_{s_i} \frac{k c m_{new_i}}{n_{s_i}} + \left(\frac{k c m_{new_i}}{n_{s_i}}\right)^2} \,, \\
&= \frac{\frac{1}{m_{s_i}} p_j (1 - p_j) + \frac{m_{new_i}}{m_{s_i}^2} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right)}{1 + 2 \frac{k c m_{new_i}}{m_{s_i} n_{s_i}} + \left(\frac{k c m_{new_i}}{n_{s_i} m_{s_i}}\right)^2} \,, \\
\lim_{i\to\infty} \frac{\mathrm{Cov}(X_i, Y_i)}{\mathbb{E}(Y_i)^2} &= 0 \,,
\end{aligned} \tag{17}$$

as both $m_{new_i}/(n_{s_i} m_{s_i}) \to 0$ and $m_{new_i}/m_{s_i}^2 \to 0$.

The third term

$$\begin{aligned}
\frac{\mathbb{E}(X_i)\mathrm{Var}(Y_i)}{\mathbb{E}(Y_i)^3} &= \frac{\left(m_{s_i} p_j + \frac{c m_{new_i}}{n_{s_i}}\right)\left(k m_{new_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right)\right)}{\left(m_{s_i} + \frac{k c m_{new_i}}{n_{s_i}}\right)^3} \,, \\
&= \frac{\left(\frac{p_j}{m_{s_i}} + \frac{c m_{new_i}}{n_{s_i} m_{s_i}}\right)\left(k \frac{m_{new_i}}{m_{s_i}^2} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right)\right)}{\left(1 + \frac{k c m_{new_i}}{n_{s_i} m_{s_i}}\right)^3} \,, \\
\lim_{i\to\infty} \frac{\mathbb{E}(X_i)\mathrm{Var}(Y_i)}{\mathbb{E}(Y_i)^3} &= 0 \,,
\end{aligned} \tag{18}$$

as $p_j \leq 1$, $p_j/m_{s_i}$ goes to zero in addition to the terms $m_{new_i}/(n_{s_i} m_{s_i})$ and $m_{new_i}/m_{s_i}^2$. Therefore, using the second order Taylor series approximation in equation (25) and equations (16), (17) and (18) we obtain

$$\lim_{i\to\infty} \mathbb{E}\left(\frac{X}{Y}\right) = \lim_{i\to\infty} \mathbb{E}\left(\frac{q_{j,i}}{\sum_{\ell=1}^{k} q_{\ell,i}}\right) = p_j \,.$$

Furthermore from the above computations we have

$$\frac{\mathrm{Cov}(X_i, Y_i)}{\mathbb{E}(Y_i)^2} \leq \frac{c}{m_{s_i}} \,,$$

and

$$\frac{\mathbb{E}(X_i)\mathrm{Var}(Y_i)}{\mathbb{E}(Y_i)^3} \leq \frac{c}{m_{s_i}^2}\,,$$

where each $c$ denotes a different constant. Thus we have

$$\left| \mathbb{E}\left( \frac{q_{j,i}}{\sum_{\ell=1}^{k} q_{\ell,i}} \right) - p_j \right| \leq \frac{c}{m_{s_i}}\,.$$

From Lemma D.1 we know

$$\mathrm{Var}\left(\frac{X}{Y}\right) = \frac{\mathbb{E}(X)^2}{\mathbb{E}(Y)^2}\left( \frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2} - 2\frac{\mathrm{Cov}(X,Y)}{\mathbb{E}(X)\mathbb{E}(Y)} + \frac{\mathrm{Var}(Y)}{\mathbb{E}(Y)^2} \right)\,.$$

We consider the three terms inside the paranthesis separately. The first term

$$\begin{aligned}
\frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2} &= \frac{m_{s_i} p_j(1-p_j) + m_{new_i}\frac{c}{n_{s_i}}\left(1-\frac{c}{n_{s_i}}\right)}{\left(m_{s_i}p_j + \frac{c m_{new_i}}{n_{s_i}}\right)^2}\,, \\
&= \frac{\frac{1}{m_{s_i}} p_j(1-p_j) + \frac{m_{new_i}}{m_{s_i}^2}\frac{c}{n_{s_i}}\left(1-\frac{c}{n_{s_i}}\right)}{\left(p_j + \frac{c m_{new_i}}{n_{s_i} m_{s_i}}\right)^2}\,, \\
\lim_{i\to\infty} \frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2} &= \lim_{i\to\infty} \frac{\frac{1}{m_{s_i}} p_j(1-p_j) + \frac{m_{new_i}}{m_{s_i}^2}\frac{c}{n_{s_i}}\left(1-\frac{c}{n_{s_i}}\right)}{\left(p_j + \frac{c m_{new_i}}{n_{s_i} m_{s_i}}\right)^2}\,, \\
\lim_{i\to\infty} \frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2} &= 0\,.
\end{aligned}$$

The second term

$$\begin{aligned}
\frac{\mathrm{Cov}(X_i,Y_i)}{\mathbb{E}(X_i)\mathbb{E}(Y_i)} &= \frac{m_{s_i} p_j(1-p_j) + m_{new_i}\frac{c}{n_{s_i}}\left(1-\frac{c}{n_{s_i}}\right)}{\left(m_{s_i}p_j + \frac{c m_{new_i}}{n_{s_i}}\right)\left(m_{s_i} + \frac{kc m_{new_i}}{n_{s_i}}\right)}\,, \\
&= \frac{\frac{1}{m_{s_i}} p_j(1-p_j) + \frac{m_{new_i}}{m_{s_i}^2}\frac{c}{n_{s_i}}\left(1-\frac{c}{n_{s_i}}\right)}{\left(p_j + \frac{c m_{new_i}}{n_{s_i} m_{s_i}}\right)\left(1 + \frac{kc m_{new_i}}{n_{s_i} m_{s_i}}\right)}\,, \\
\lim_{i\to\infty} \frac{\mathrm{Cov}(X_i,Y_i)}{\mathbb{E}(X_i)\mathbb{E}(Y_i)} &= 0\,.
\end{aligned}$$

The term term is similar to the first term. Combining these three terms we get using the first order approximation

$$\lim_{i\to\infty} \mathrm{Var}\left(\frac{X}{Y}\right) = 0\,.$$

Similar to the computation of the expectation we get

$$\mathrm{Var}\left(\frac{X}{Y}\right) \leq \frac{c}{m_{s_i}}\,.$$

□

### D.2 Estimating $k$

**Proposition D.2.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U,W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) associated with $U$*

*with $k$ finite partitions.Let $\tilde{q}_{j,i}$ be the degree of the star in $G_{s_i}$ corresponding to $p_j \neq 0$. Let $q_{j,i}$ denote the degree of the corresponding vertex in $G_{n_i}$. Let deg $v_{(\ell)}$ denote the $\ell$th largest degree in $G_{n_i}$. Suppose deg $v_{(\ell)} = q_{\ell,i}$ for $\ell \in \{1, \ldots, k\}$ and let $\alpha \geq 0$ be a constant. Then for any $\ell \in \{1, \ldots, (k-1)\}$*

$$P\left((1+\alpha)q_{\ell,i}n_{d_i} > q_{k,i}q_{\ell+1,i}\right) \leq \frac{c}{m_{s_i}}$$

*where $n_{d_i}$ denotes the number of nodes in the dense part. Consequently, with high probability*

$$\log q_{k,i} - \log(1+\alpha)n_{d_i} \geq \log q_{\ell,i} - \log q_{\ell+1,i}\,.$$

*That is, in the log scale the difference between successive top $k$ degrees is largest at $k$, where we have considered the highest degree in the dense part to be $n_{d_i}$.*

*Proof.* Let $X_i = (1+\alpha)q_{\ell,i}n_{d_i} - q_{k,i}q_{\ell+1,i}$. We show that $P(X_i > 0)$ goes to zero. From Lemma B.2 we know

$$\begin{aligned}
\mathbb{E}(q_{\ell,i}) &= m_{s_i}p_\ell + \frac{cm_{new_i}}{n_{s_i}} = m_{s_i}p_\ell + c_{e_i}\,, \\
\mathrm{Var}(q_{\ell,i}) &= m_{s_i}p_\ell(1-p_\ell) + m_{new_i}\frac{c}{n_{s_i}}\left(1 - \frac{c}{n_{s_i}}\right) \\
&= m_{s_i}p_\ell(1-p_\ell) + c_{v_i}\,,
\end{aligned}$$

where $c_{e_i} = \frac{cm_{new_i}}{n_{s_i}}$ and $c_{v_i} = m_{new_i}\frac{c}{n_{s_i}}\left(1 - \frac{c}{n_{s_i}}\right)$. Note that $\frac{c_{e_i}}{m_{s_i}}$ and $\frac{c_{v_i}}{m_{s_i}}$ goes to zero as $m_{new_i} \in \mathcal{O}(n_{d_i}^2)$ and as $n_{d_i}/n_{s_i}$ goes to zero. Given $m_{s_i}$, the random variables $q_{k,i}$ and $q_{\ell,i}$ are independent for $k \neq \ell$. Hence

$$\begin{aligned}
\mathbb{E}(X_i) &= \mathbb{E}\left((1+\alpha)q_{\ell,i}n_{d_i} - q_{k,i}q_{\ell+1,i}\right)\,, \\
&= (1+\alpha)n_{d_i}\mathbb{E}\left(q_{\ell,i}\right) - \mathbb{E}\left(q_{k,i}\right)\mathbb{E}\left(q_{\ell+1,i}\right)\,, \\
&= (1+\alpha)n_{d_i}\left(m_{s_i}p_\ell + c_{e_i}\right) - \left(m_{s_i}p_k + c_{e_i}\right)\left(m_{s_i}p_{\ell+1} + c_{e_i}\right)\,, \\
\frac{\mathbb{E}(X_i)}{m_{s_i}^2} &= \frac{(1+\alpha)n_{d_i}}{m_{s_i}}\left(p_\ell + \frac{c_{e_i}}{m_{s_i}}\right) - \left(p_k + \frac{c_{e_i}}{m_{s_i}}\right)\left(p_{\ell+1} + \frac{c_{e_i}}{m_{s_i}}\right)\,, \\
\lim_{i\to\infty}\frac{\mathbb{E}(X_i)}{m_{s_i}^2} &= -p_kp_{\ell+1}\,,
\end{aligned} \tag{19}$$

as $n_{d_i}/m_{s_i}$ goes to zero because $n_{d_i}/n_{s_i}$ goes to zero and $m_{s_i} \in \Theta(m_{s_i})$. The variance of the product of two independent random variables $A$ and $B$ is given by

$$\mathrm{Var}(AB) = \mathrm{Var}(A)\mathrm{Var}(B) + \mathbb{E}(A)^2\mathrm{Var}(B) + \mathbb{E}(B)^2\mathrm{Var}(A)\,.$$

Using the product formula we get

$$\begin{aligned}
\mathrm{Var}(q_{k,i}q_{\ell+1,i})) &= \mathrm{Var}(q_{k,i})\mathrm{Var}(q_{\ell+1,i})) + \mathbb{E}(q_{k,i})^2\mathrm{Var}(q_{\ell+1,i})) + \mathbb{E}(q_{\ell+1,i}))^2\mathrm{Var}(q_{k,i})\,, \\
&= \left(m_{s_i}p_k(1-p_k) + c_{v_i}\right)\left(m_{s_i}p_{\ell+1}(1-p_{\ell+1}) + c_{v_i}\right)\,, \\
&\quad + \left(m_{s_i}p_k + c_{e_i}\right)^2\left(m_{s_i}p_{\ell+1}(1-p_{\ell+1}) + c_{v_i}\right)\,, \\
&\quad + \left(m_{s_i}p_{\ell+1} + c_{e_i}\right)^2\left(m_{s_i}p_k(1-p_k) + c_{v_i}\right)\,, \\
\frac{\mathrm{Var}(q_{k,i}q_{\ell+1,i}))}{m_{s_i}^3} &= \frac{1}{m_{s_i}}\left(p_k(1-p_k) + \frac{c_{v_i}}{m_{s_i}}\right)\left(p_{\ell+1}(1-p_{\ell+1}) + \frac{c_{v_i}}{m_{s_i}}\right)\,, \\
&\quad + \left(p_k + \frac{c_{e_i}}{m_{s_i}}\right)^2\left(p_{\ell+1}(1-p_{\ell+1}) + \frac{c_{v_i}}{m_{s_i}}\right)\,, \\
&\quad + \left(p_{\ell+1} + \frac{c_{e_i}}{m_{s_i}}\right)^2\left(p_k(1-p_k) + \frac{c_{v_i}}{m_{s_i}}\right)\,, \\
\lim_{i\to\infty}\frac{\mathrm{Var}(q_{k,i}q_{\ell+1,i}))}{m_{s_i}^3} &= p_k^2p_{\ell+1}(1-p_{\ell+1}) + p_{\ell+1}^2p_k(1-p_k)
\end{aligned} \tag{20}$$

As $n_{d_i}$ is not a random variable we get

$$\begin{aligned}\mathrm{Var}((1+\alpha)n_{d_i}q_{\ell,i}) &= (1+\alpha)^2 n_{d_i}^2 \mathrm{Var}(q_{\ell,i})\,, \\ &= (1+\alpha)^2 n_{d_i}^2 \left(m_{s_i}p_\ell(1-p_\ell)+c_{v_i}\right)\,, \\ \frac{\mathrm{Var}(n_{d_i}q_{\ell,i})}{m_{s_i}^3} &= \frac{(1+\alpha)^2 n_{d_i}^2}{m_{s_i}^2}\left(p_\ell(1-p_\ell)+\frac{c_{v_i}}{m_{s_i}}\right)\,, \\ \lim_{i\to\infty}\mathrm{Var}(n_{d_i}q_{\ell,i}) &= 0\,,\end{aligned} \tag{21}$$

as $n_{d_i}/m_{s_i}$ goes to zero. Combining terms for $X_i = (1+\alpha)q_{\ell,i}n_{d_i} - q_{k,i}q_{\ell+1,i}$ in equations (19), (20) and (21) we get

$$\begin{aligned}\mathrm{Var}(X_i) &= \mathrm{Var}\left((1+\alpha)q_{\ell,i}n_{d_i}\right) + \mathrm{Var}\left(q_{k,i}q_{\ell+1,i}\right)\,, \\ \lim_{i\to\infty}\frac{\mathrm{Var}(X_i)}{m_{s_i}^3} &= p_k^2 p_{\ell+1}(1-p_{\ell+1}) + p_{\ell+1}^2 p_k(1-p_k)\,, \\ \lim_{i\to\infty}\frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2} &= \lim_{i\to\infty}\frac{\mathrm{Var}(X_i)/m_{s_i}^4}{\mathbb{E}(X_i)^2/m_{s_i}^4}\,, \\ &= \lim_{i\to\infty}\frac{\frac{1}{m_{s_i}}\left(p_k^2 p_{\ell+1}(1-p_{\ell+1}) + p_{\ell+1}^2 p_k(1-p_k)\right)}{(p_k p_{\ell+1})^2} = 0\,,\end{aligned} \tag{22}$$

as $\mathrm{Var}(X_i) \in \Theta(m_{s_i}^3)$ and $\mathbb{E}(X_i) \in \Theta(m_{s_i}^2)$. Furthermore

$$\frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2} \leq \frac{c}{m_{s_i}}\,,$$

for some $c$. As shown in equation 19 for large $i$, $\mathbb{E}(X_i) \leq 0$. We know that

$$\begin{aligned}P(X_i > 0) &= P(X_i - \mathbb{E}(X_i) \geq -\mathbb{E}(X_i))\,, \\ &= P(X_i - \mathbb{E}(X_i) \geq |\mathbb{E}(X_i)|) \quad \text{as } -\mathbb{E}(X_i) = |\mathbb{E}(X_i)|\,, \\ &\leq P(|X_i - \mathbb{E}(X_i)| \geq |\mathbb{E}(X_i)|)\,, \\ &\leq \frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2}\end{aligned} \tag{23}$$

where we have used Chebyshev's inequality. Using equations (23) and (22) we obtain

$$\begin{aligned}P\left((1+\alpha)q_{\ell,i}n_{d_i} > q_{k,i}q_{\ell+1,i}\right) &= P(X_i > 0)\,, \\ &\leq \frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2} = 0\,, \\ &\leq \frac{c}{m_{s_i}}\end{aligned}$$

giving us

$$\lim_{i\to\infty} P\left((1+\alpha)q_{\ell,i}n_{d_i} > q_{k,i}q_{\ell+1,i}\right) = 0\,.$$

Consequently

$$\lim_{i\to\infty} P\left((1+\alpha)q_{\ell,i}n_{d_i} \leq q_{k,i}q_{\ell+1,i}\right) = 1\,.$$

Therefore, with high probability we have

$$\begin{aligned}q_{k,i}q_{\ell+1,i} &\geq (1+\alpha)q_{\ell,i}n_{d_i}\,, \\ \log\left(\frac{q_{k,i}q_{\ell+1,i}}{(1+\alpha)q_{\ell,i}n_{d_i}}\right) &\geq 0\,, \\ \log q_{k,i} - \log(1+\alpha)n_{d_i} &\geq \log q_{\ell,i} - \log q_{\ell+1,i}\,.\end{aligned}$$

That is, if we take successive differences of the top-$k$ degrees in the log scale, with high probability the difference between the $k$th largest degree and $(1+\alpha)n_{d_i}$ is larger than the other successive degree differences for $\ell < k$ where we have used the $(1+\alpha)n_{d_i}$ in place of the highest degree in the dense part $G_{d_i}$. □

**Lemma D.3.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $max\_deg_{\widetilde{G}_{d_i}}$ denote the maximum degree in $G_{n_i}$ restricted to nodes from $G_{d_i}$. That is, $max\_deg_{\widetilde{G}_{d_i}}$ belongs to a node from the dense part $G_{d_i}$. Then*

$$\mathbb{E}\left(max\_deg_{\widetilde{G}_{d_i}}\right) \leq (1+\alpha)n_{d_i}\,,$$

*where $\alpha$ depends on $W$ and the joining mechanism.*

*Proof.* Suppose $v_\ell$ is originally a vertex in the dense part $G_{d_i}$ Then, from Lemma B.1

$$\begin{aligned}\mathbb{E}\left(\deg_{\ G_{n_i}} v_\ell\right) &= \deg_{\ G_{d_i}} v_\ell + c_1\frac{m_{new_i}}{n_{d_i}}\,,\\ &\leq \deg_{\ G_{d_i}} v_\ell + c'\frac{m_{d_i}}{n_{d_i}}\,,\\ &= \deg_{\ G_{d_i}} v_\ell + c' n_{d_i}\frac{m_{d_i}}{n_{d_i}^2}\,,\\ &\leq n_{d_i} + c' n_{d_i}\rho_i\,,\\ &= (1+c'\rho_i)n_{d_i}\end{aligned}$$

where number of nodes in the dense part $n_{d_i}$ is a large upper bound for the largest degree in $G_{d_i}$. The constant $c_1$ and $m_{new_i}$ depend on the joining mechanism (Definition 3.3) and $m_{new_i} \leq c m_{d_i}$. The density of $G_{d_i}$ is denoted by $\rho_i$, which is a bounded quantity and depends on $W$. □

**Lemma D.4.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Suppose $W$ satisfies Assumption 5.2. We consider the degrees in $G_{d_i}$. Let $r_{x_j,i}$ and $r_{x_j+\epsilon_i,i}$ denote the degree of two nodes corresponding to $D(x_j)$ and $D(x_j+\epsilon_i)$ for $D(x_j), D(x_j+\epsilon_i) > \xi > 0$ for some $\xi > 0$. Suppose $\lim_{i\to\infty}\epsilon_i = 0$. Then using the second order approximation of the Taylor expansion, we have*

$$\left|\mathbb{E}\left(\frac{r_{x_j,i}}{r_{x_j+\epsilon_i,i}}\right) - 1\right| \leq \frac{c}{n_{d_i}}$$

*That is, if $W$ has a continuous degree function $D(x)$, then as $i$ goes to infinity, excluding small degrees the ratio of consecutive degrees in $G_{d_i}$ tend to one.*

*Proof.* The degree function acts like the edge probability. Thus, the expectation and variance of $r_{x_j,i}$ are given by

$$\begin{aligned}\mathbb{E}(r_{x_j,i}) &= (n_{d_i}-1)D(x_j)\,,\\ \mathrm{Var}(r_{x_j,i}) &= (n_{d_i}-1)D(x_j)(1-D(x_j))\,.\end{aligned}$$

Using Lemma D.1 we have

$$\mathbb{E}\left(\frac{X}{Y}\right) = \frac{\mathbb{E}(X)}{\mathbb{E}(Y)} - \frac{\mathrm{Cov}(X,Y)}{\mathbb{E}(Y)^2} + \frac{\mathbb{E}(X)\mathrm{Var}(Y)}{\mathbb{E}(Y)^3}\,.$$

for two random variables $X$ and $Y$. Given $D(x_j)$ the degrees of the two nodes are independent because each edge is independently sampled from a Bernoulli($p$) for $p = D(x_j), D(x_j+\epsilon_i)$. This gives us

$$\begin{aligned}\mathbb{E}\left(\frac{r_{x_j,i}}{r_{x_j+\epsilon_i,i}}\right) &= \frac{(n_{d_i}-1)D(x_j)}{(n_{d_i}-1)D(x_j+\epsilon_i)} + \frac{(n_{d_i}-1)D(x_j)(n_{d_i}-1)D(x_j+\epsilon_i)(1-D(x_j+\epsilon_i))}{(n_{d_i}-1)^3D(x_j+\epsilon_i)^3}\,,\\ &= \frac{D(x_j)}{D(x_j+\epsilon_i)} + \frac{D(x_j)(1-D(x_j+\epsilon_i))}{(n_{d_i}-1)D(x_j+\epsilon_i)^2}\,,\end{aligned}$$

As $n_{d_i}$ goes to infinity, the second term goes to zero and as $\epsilon_i$ goes to 0, we have the result. Continuity of $D(x)$ is used to say that $D(x_j+\epsilon_i)$ goes to $D(x_j)$. A discontinuous function with a gap at $x_j$ does not satisfy this condition.

Excluding small degrees such as 0, 1, and 2, we can think of consecutive degrees as realizations of nodes sampled with probabilities $D(x_j)$ and $D(x_j+\epsilon_i)$ for some $x_j$ where $D(x_j) > \xi > 0$. The condition $D(x_j) > \xi > 0$ guarantees that we stay away from small degrees. As $i$ goes gets larger, more and more nodes are sampled from the graphon $W$. As such $\epsilon_i$ goes to zero as $i$ tends to infinity. Therefore, the ratio of consecutive degrees go to 1. □

**Proposition 5.3.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ having nodes $n_{d_i}$ and $n_{s_i}$ respectively. Suppose $W$ satisfies Assumption 5.2 and $W \neq 1$. Let $\boldsymbol{p} = (p_1, p_2, \dots)$ be the mass-partition (Definition 3.1) associated with $U$ which has only $k$ partitions. Let deg $_{G_{n_i}} v_{(\ell)}$ denote the $\ell$th largest degree in $G_{n_i}$. Then there exists $I_0$ such that for $i > I_0$ we have*

$$k = \max_{\ell} \left( \log deg \, _{G_{n_i}} v_{(\ell)} - \log deg \, _{G_{n_i}} v_{(\ell+1)} \right) , \tag{3}$$

*where we exclude small degrees. That is, in the log scale the difference between successive top $k$ degrees is largest at $k$.*

*Proof.* Let $\tilde{q}_{j,i}$ be the degree of the star in $G_{s_i}$ corresponding to $p_j \neq 0$. Let $q_{j,i}$ denote the degree of the corresponding vertex in $G_{n_i}$. Then from the Order Preserving Property (Proposition 4.1) we know

$$\lim_{i \to \infty} P \left( \bigcap_{j=1}^{k} \left( q_{j,i} = \deg \, _{G_{n_i}} v_{(j)} \right) \right) = 1 \, .$$

Combining with Proposition D.2 with high probability we have

$$\log \deg \, _{G_{n_i}} v_{(k)} - \log(1+\alpha) n_{d_i} \geq \log \deg \, _{G_{n_i}} v_{(\ell)} - \log \deg \, _{G_{n_i}} v_{(\ell+1)} \, , \tag{24}$$

for $\ell \in \{1, \dots, k-1\}$. Let $\text{max\_deg}_{\widetilde{G}_{d_i}}$ denote the maximum degree in $G_{n_i}$ restricted to nodes from $G_{d_i}$. That is, $\text{max\_deg}_{\widetilde{G}_{d_i}}$ belongs to a node from the dense part $G_{d_i}$. For $W = 1$ we have $\text{max\_deg}_{\widetilde{G}_{d_i}} = n_{d_i} - 1$ and

$$\mathbb{E} \left( \text{max\_deg}_{\widetilde{G}_{d_i}} \right) \leq (1 + c') n_{d_i} \, .$$

But for all other $W \neq 1$, $\text{max\_deg}_{\widetilde{G}_{d_i}} << n_{d_i}$. For large enough $i$, $(k+1)$st highest degree

$$\deg \, _{G_{n_i}} v_{(k+1)} = \text{max\_deg}_{\widetilde{G}_{d_i}} \, .$$

Using Chernoff bounds, the probability

$$P \left( \deg \, _{G_{n_i}} v_{(k+1)} > (1+\epsilon)\mu \right) \leq \exp \left( -\frac{\mu \epsilon^2}{3} \right) \, ,$$

making

$$P \left( \deg \, _{G_{n_i}} v_{(k+1)} > (1+\alpha) n_{d_i} \right) \to 0 \, ,$$

for appropriate $\alpha$ and $W \neq 1$. With high probability we have

$$\begin{aligned}
(1+\alpha) n_{d_i} &\geq \deg \, _{G_{n_i}} v_{(k+1)} \, , \\
\log(1+\alpha) n_{d_i} &\geq \log \deg \, _{G_{n_i}} v_{(k+1)} \, , \\
\log \deg \, _{G_{n_i}} v_{(k)} - \log \deg \, _{G_{n_i}} v_{(k+1)} &\geq \log \deg \, _{G_{n_i}} v_{(k)} - \log(1+\alpha) n_{d_i} \, , \\
&\geq \log \deg \, _{G_{n_i}} v_{(\ell)} - \log \deg \, _{G_{n_i}} v_{(\ell+1)} \, ,
\end{aligned}$$

where we have used equation (24). In the logscale the difference between $k$th highest vertex and $(k+1)$st highest vertex is larger than the $\ell$th highest vertex and the $(\ell+1)$st highest vertex for $\ell < k$. That is, in the log scale, successive difference of the top $k$ degrees is highest at $k$. From Lemma D.4 we know that for degrees in the dense part $G_{d_i}$, the ratios of successive degrees converge to 1, when we exclude small degrees. As we are taking ordered values $\deg \, _{G_{n_i}} v_{(\ell)} / \deg \, _{G_{n_i}} v_{(\ell+1)}) \approx p_\ell / p_{\ell+1} > 1$. The ratio $\deg \, _{G_{n_i}} v_{(k)} / \deg \, _{G_{n_i}} v_{(k+1)} \approx m_{s_i} p_k / n_{d_i} D_{\max}$ where $D_{\max} = \max D(x)$. Thus, the log difference $\log \deg \, _{G_{n_i}} v_{(k)} - \log \deg \, _{G_{n_i}} v_{(k+1)}$ is much greater than 1 for $i > I_0$. Therefore we have

$$k = \max_{\ell} \left( \log \deg \, _{G_{n_i}} v_{(\ell)} - \log \deg \, _{G_{n_i}} v_{(\ell+1)} \right) \, ,$$

with high probability. □

**Lemma D.5.** ***(The 3 Degree Groups)*** *Suppose $\{G_{n_i}\}_i$ is a sparse $(U, W)$-mixture graph sequence (Definition 3.2) with $W$ having a continuous degree function (Definition 5.1). Let $\boldsymbol{p} = (p_1, p_2, \ldots)$ be the mass-partition (Definition 3.1) associated with $U$. Then there are 3 groups of nodes satisfying different degree expectations and variances. They are (1)* ***Group 1****: the large degrees generated from $U$ (2)* ***Group 2****: the degrees of nodes generated by $W$ and (3)* ***Group 3****: the very small degrees generated from $U$.*

*Proof.* The sparse part $G_{s_i}$ is a collection of stars with large degree hub nodes and degree-1 leaf nodes. Let $q_{j,i}$ be the degree of the hub node in $G_{n_i}$ corresponding to $p_j$. These are the Group 1 nodes. Then from Lemma B.2

$$\mathbb{E}(q_{j,i}) = m_{s_i} p_j + \frac{c m_{new_i}}{n_{s_i}} ,$$
$$\mathrm{Var}(q_{j,i}) = m_{s_i} p_j (1 - p_j) + m_{new_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right) .$$

In addition the the large degree hub nodes, the sparse part $G_{s_i}$ has a large number of nodes with degree 1. Let us denote the degree of these nodes in $G_{n_i}$ as $e_i$. These are Group 3 nodes. Their degree can change as a result of the joining process (Definition 3.3). Then

$$\mathbb{E}(e_i) = 1 + \frac{c m_{new_i}}{n_{s_i}} ,$$
$$\mathrm{Var}(e_i) = m_{new_i} \frac{c}{n_{s_i}} \left(1 - \frac{c}{n_{s_i}}\right) .$$

Next we consider nodes generated from $W$, the Group 2 nodes. Combining Lemma C.2 and Lemma D.4 we know that nodes generated by $W$ have degree

$$\mathbb{E}(r_{x_j,i}) = (n_{d_i} - 1) D(x_j) + \frac{c m_{new_i}}{n_{d_i}} ,$$
$$\mathrm{Var}(r_{x_j,i}) = (n_{d_i} - 1) D(x_j)(1 - D(x_j)) + m_{new_i} \frac{c}{n_{d_i}} \left(1 - \frac{c}{n_{d_i}}\right) .$$

where $r_{x_j,i}$ denotes the degree of a node in $G_{n_i}$ corresponding to $D(x_j)$. Thus, the degree expectations and variances of these 3 groups of nodes are different. Furthermore as $n_{s_i}/n_{d_i}$ goes to infinity for sparse graphs, we have

$$\mathbb{E}(q_{j,i}) > \mathbb{E}(r_{x_j,i}) > \mathbb{E}(e_i) ,$$
$$\mathrm{Var}(q_{j,i}) > \mathrm{Var}(r_{x_j,i}) > \mathrm{Var}(e_i) ,$$

for large $i$.

□

# E Estimating $U$ for infinite partitions

**Lemma E.1.** *Let $\{G_{n_i}\}_i$ be a sparse sequence of $(U, W)$-mixture graphs. Let $\boldsymbol{p}$ be the mass-partition corresponding to $U$ and suppose $\boldsymbol{p}$ has infinite non-zero elements. Let $k_i$ be the number of nodes in $G_{n_i}$ that are generated by $U$ with degrees higher than those generated by $W$. Let $\deg_{G_{n_i}} v_{(j)}$ denote the $j$th highest degree in $G_{n_i}$. Then the line fitted to the points $(j, \log \deg_{G_{n_i}} v_{(j)})_{j=1}^{k_i}$ using OLS regression is steeper than the line fitted to $\left\{\left(j, \log \frac{m_{s_i}}{j}\right)\right\}_{j=1}^{k_i}$ for large enough $i$, i.e., if $\beta_{actual,i}$ and $\beta_{1,i}$ are the slopes of the lines fitted to points $\left\{(j, \log \deg_{G_{n_i}} v_{(j)})\right\}_{j=1}^{k_i}$ and $\left\{\left(j, \log \frac{m_{s_i}}{j}\right)\right\}_{j=1}^{k_i}$ respectively, then there exits $i > I_0$ such that*

$$|\beta_{actual,i}| > |\beta_{1,i}| .$$

*Proof.* First we show for any $\alpha > 0$ and for some $N \in \mathbb{N}$ the OLS line fitted to the points $\left\{\left(j, \log \frac{m_{s_i}}{j^{1+\alpha}}\right)\right\}_{j=1}^{N}$ is steeper than the OLS line fitted to the points $\left\{\left(j, \log \frac{m_{s_i}}{j}\right)\right\}_{j=1}^{N}$. For a set of points $\{(x_j, y_j)\}_{j=1}^{N}$ the slope of the OLS line is given by

$$\beta = \frac{\sum_j x_j y_j - \bar{x} y_j}{\sum_j (x_j - \bar{x})^2} \, .$$

As the denominator of the slope $\beta$ is the same for both sets of points $\left\{\left(j, \log \frac{m_{s_i}}{j^{1+\alpha}}\right)\right\}_{j=1}^{N}$ and $\left\{\left(j, \log \frac{m_{s_i}}{j}\right)\right\}_{j=1}^{N}$ we focus on the numerator. Let $\beta_{num,\alpha} = \sum_j x_j y_j - \bar{x} y_j$ for the points $\left\{\left(j, \log \frac{m_{s_i}}{j^{1+\alpha}}\right)\right\}_{j=1}^{N}$ and $\beta_{num,1} = \sum_j x_j y_j - \bar{x} y_j$ for the points $\left\{\left(j, \log \frac{m_{s_i}}{j}\right)\right\}_{j=1}^{N}$. Then we show that

$$\beta_{num,\alpha} < \beta_{num,1} \, .$$

Noting that $\bar{x}$ denotes the mean of $x_i$ values, which is the same for both lines we compute

$$\begin{aligned}
\beta_{num,1} - \beta_{num,\alpha} &= \sum_j \left( j \log \frac{m_{s_i}}{j} - \bar{x} \log \frac{m_{s_i}}{j} \right) - \sum_j \left( j \log \frac{m_{s_i}}{j^{1+\alpha}} - \bar{x} \log \frac{m_{s_i}}{j^{1+\alpha}} \right) , \\
&= \sum_j j \left( \log \frac{m_{s_i}}{j} - \log \frac{m_{s_i}}{j^{1+\alpha}} \right) - \bar{x} \sum_j \left( \log \frac{m_{s_i}}{j} - \log \frac{m_{s_i}}{j^{1+\alpha}} \right) , \\
&= \sum_j \left( j\alpha \log j - \bar{x} \alpha \log j \right) , \\
&= \alpha \sum_j \left( j \log j - \bar{x} \log j \right) , \\
&= \alpha \beta_{num,-1} \, ,
\end{aligned}$$

where $\beta_{num,-1} = \sum_j x_j y_j - \bar{x} y_j$ for the points $(i, \log j)_{j=1}^{N}$. These points lie on the curve $y = \log x$ and as such its slope is positive. This gives us

$$\beta_{num,1} - \beta_{num,\alpha} > 0 \, .$$

Dividing by the same denominator we get

$$\beta_{1,i} > \beta_{\alpha,i} \, ,$$

where $\beta_{\alpha,i}$ is the slope of the line fitted to the points $\left\{\left(j, \log \frac{m_{s_i}}{j^{1+\alpha}}\right)\right\}_{j=1}^{N}$. But these slopes are negative because the $y$ values are decreasing. Therefore

$$|\beta_{\alpha,i}| > |\beta_{1,i}| \, .$$

A mass-partition $\boldsymbol{p}$ with infinite non-zero entries cannot decrease similar to $\{c/j\}_j$ because the series $\sum_j c/j$ diverges to infinity. As $\sum_j p_j = 1$, the sequence $\{p_j\}_j$ has to decrease faster than $\{c/j\}_j$. As shown above for a set of points decreasing faster then $\{c/j\}_j$ the slope of the OLS line fitted to the log values is steeper than the slope of the line fitted to $\left\{\left(j, \log \frac{m_{s_i}}{j}\right)\right\}_{j=1}^{N}$, making

$$|\beta_{actual,i}| > |\beta_{1,i}| \, .$$

□

For the next lemma we consider nodes with relatively large degrees generated by $W$. As such we consider the unique degree values and take degree values greater than a given percentile $C$. In practice $C$ can be the median unique degree value.

**Lemma E.2.** *Let $\{G_{n_i}\}_i$ be a sparse sequence of $(U, W)$-mixture graphs with $W$ satisfying Assumption 5.2. We consider nodes generated by $W$ with degrees greater than some percentile $C$ when considering unique degree values. Suppose there are $N_i$ such points in $G_{n_i}$. We consider $\{(j, \log(deg_{\,G_{n_i}} v_{(\ell)}))\}_{j=1}^{N_i}$ where the unique degree values* *are sorted in decreasing order. Then the line fitted to the points* $\{(j, \log(deg_{\ G_{n_i}} v_{(\ell)}))\}_{j=1}^{N_i}$ *is less steeper than the line fitted to the points* $\{(j, \log n_{d_i}/j)\}_{j=1^{N_i}}$. *If* $\beta_{dense,i}$ *and* $\beta_{1,i}$ *are the slopes of the lines fitted to points* $\{(j, \log deg_{\ G_{n_i}} v_{(j)})\}_{j=1}^{N_i}$ *and* $\{(j, \log n_{d_i}/j)\}_{j=1}^{N_i}$ *respectively, then*

$$|\beta_{dense,i}| < |\beta_{1,i}| \, .$$

*Proof.* Assumption 5.2 is on $W$ having a continuous degree function $D(x)$ (Definition 5.1). When $W$ has a continuous degree function, for large enough $i$, the sorted unique degree values are mostly consecutive integers. Thus, the points $\{(j, \log \deg_{\ G_{n_i}} v_{(j)})\}_{j=1}^{N_i}$ have a large proportion of points in the form of $(j, \log(N_0 - j))$ for some $N_0$. The slope of the OLS line fitted to these points is much less steeper than the slope of the line fitted to $\{(j, \log n_{d_i}/j)\}_{j=1}^{N_i}$. That is,

$$|\beta_{dense,i}| < |\beta_{1,i}| \, .$$

□

**Lemma E.3.** *Suppose* $\{G_{n_i}\}_i$ *is a sparse* $(U, W)$*-mixture graph sequence (Definition 3.2) with* $W$ *having a continuous degree function (Definition 5.1). Let* $\boldsymbol{p} = (p_1, p_2, \ldots)$ *be the mass-partition (Definition 3.1) associated with* $U$. *Suppose* $U$ *has infinite partitions. Let* $G_{n_\ell}$ *be a graph in the sequence* $\{G_{n_i}\}_i$. *We conduct Procedure 6.1 for* $G_{n_\ell}$ *and estimate* $\hat{k}_\ell$ *using equation* (5). *Similarly we estimate* $\hat{k}_i$ *for each* $G_{n_i}$. *Then as* $i$ *goes to infinity*

$$\lim_{i \to \infty} P(\hat{k}_i > \hat{k}_\ell) = 1 \, .$$

*Furthermore, with high probability* $\lim_{i\to\infty} \hat{k}_i = \infty$. *That is, for any* $M \in \mathbb{N}$

$$\lim_{i \to \infty} P(\hat{k}_i < M) = 0 \, .$$

*Proof.* Let $G_{s_\ell}$ denote the sparse part of $G_{n_\ell}$. Suppose the number of stars in the sparse part $G_{s_\ell}$ is given by $st_\ell$. We know that $\hat{k}_\ell < st_\ell$ because some stars have degrees that are indistinguishable from the dense part. Pick $r \in \mathbb{N}$ such that $\hat{k}_\ell < r \leq st_\ell$. The star corresponding to $r$ has expected degree

$$\mathbb{E}(q_{r,\ell}) = m_{s_\ell} p_r + \frac{c m_{new_\ell}}{n_{s_\ell}}$$

from Lemma B.2. Furthermore as $\mathrm{Var}(q_{r,i})/\mathbb{E}(q_{r,i})^2$ goes to zero with $i$ the observed values of $q_{r,i}$ get more centered at its expected value. For sparse graphs $n_{s_i}/n_{d_i}$ goes to infinity. This results in $m_{s_i}/n_{d_i}$ going to infinity and

$$\mathbb{E}(q_{r,i}) = m_{s_i} p_r + \frac{c m_{new_i}}{n_{s_i}} > C n_{d_i}$$

for some $i$ where $C > 1$. Furthermore from Chernoff bounds

$$P(q_{r,i} < C n_{d_i}) << P(q_{r,i} < (1 - \epsilon)\mu)) = \exp\left(-\frac{\mu \epsilon^2}{2}\right)$$

where $\mu = \mathbb{E}(q_{r,i})$. As $m_{s_i}$ goes to infinity, this probability goes to zero making

$$\lim_{i \to \infty} P(q_{r,i} > C n_{d_i}) = 1 \, .$$

The nodes generated by $W$ have degrees less than $C n_{d_i}$ for some $C$, and as $i$ goes to infinity, $q_{r,i}$ gets much larger than these nodes. As such, when fitting 2 lines as described in Procedure 6.1, $q_{r,i}$ gets fitted with the first line for larger $i$, making $\hat{k}_i$ at least as big as $r$, which is greater than $\hat{k}_\ell$. Thus,

$$\lim_{i \to \infty} P(\hat{k}_i \geq r > \hat{k}_\ell) = 1 \, .$$

Next consider the graph sequence $\{G_{n_\ell}\}_\ell$. For a given $m_{s_\ell}$ the number of stars in the sparse part $G_{s_\ell}$ is finite, but as $\ell$ increases, $m_{s_\ell}$ increases and the number of stars $st_\ell$ increases making the sequence $\{st_\ell\}_\ell$ tend to infinity because $\boldsymbol{p}$ has infinite non-zero elements. Thus, we consider an increasing sequence $\{r_\ell\}_\ell$ such that $\hat{k}_\ell < r_\ell \leq st_\ell$ and $\lim_{\ell\to\infty} r_\ell = \infty$. For every $r_\ell$ there exists $I_\ell$ such that for $i > I_\ell$ we have

$$\hat{k}_i \geq r_\ell > \hat{k}_\ell \, ,$$

with high probability. Therefore, the sequence

$$\lim_{i\to\infty} \hat{k}_i = \infty$$

with high probability. That is, for any $M \in \mathbb{N}$

$$\lim_{i\to\infty} P(\hat{k}_i < M) = 0\,.$$

□

**Proposition 6.2.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U,W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Suppose $U$ has infinite partitions and let $\boldsymbol{p} = (p_1, p_2, \dots)$ be the associated mass-partition (Definition 3.1). Let $q_{j,i}$ denote the degree of the hub vertex corresponding to $p_j$ in $G_{n_i}$ and $\hat{k}_i$ be estimated using Procedure 6.1. Then using the second order approximation of the Taylor expansion for the expectation and the first order approximation for the variance, with $Q_{\hat{k},i} = \sum_{\ell=1}^{\hat{k}_i} q_{\ell,i}$ and for some constants $c_1$ and $c_2$ we have*

$$\left|\mathbb{E}\left(\frac{q_{j,i}}{Q_{\hat{k},i}}\right) - p_j\right| \le \frac{c_1}{m_{s_i}} \quad \textit{and} \quad \textit{Var}\left(\frac{q_{j,i}}{Q_{\hat{k},i}}\right) \le \frac{c_2}{m_{s_i}}$$

*Proof.* The proof is similar to that of Proposition 5.6. Let $X_i = q_{j,i}$ and $Y_i = \sum_{\ell=1}^{\hat{k}_i} q_{\ell,i}$. The difference is that $k$ is replaced with $\hat{k}_i$. Some of the expectations, variances and covariance are changed to

$$\begin{aligned}
\mathbb{E}(X_i) &= m_{s_i} p_j + \frac{c m_{new_i}}{n_{s_i}}\,, \\
\mathbb{E}(Y_i) &= m_{s_i} \sum_{\ell=1}^{\hat{k}_i} p_\ell + \frac{\hat{k}_i c m_{new_i}}{n_{s_i}}\,, \\
\mathrm{Var}(X_i) &= m_{s_i} p_j (1-p_j) + m_{new_i} \frac{c}{n_{s_i}}\left(1 - \frac{c}{n_{s_i}}\right)\,, \\
\mathrm{Var}(Y_i) &= m_{s_i} \sum_{\ell=1}^{\hat{k}_i} p_\ell \left(1 - \sum_{\ell=1}^{\hat{k}_i} p_\ell\right) + \hat{k}_i m_{new_i} \frac{c}{n_{s_i}}\left(1 - \frac{c}{n_{s_i}}\right)\,, \\
\mathrm{Cov}(X_i, Y_i) &= m_{s_i} p_j (1-p_j) + m_{new_i} \frac{c}{n_{s_i}}\left(1 - \frac{c}{n_{s_i}}\right)\,.
\end{aligned}$$

We turn out attention to the term $\hat{k}_i$ . We note that $\hat{k}_i \in O(m_{s_i}/n_{d_i})$. The sequence $\{\hat{k}_i\}_i$ increases at a slower rate compared to $m_{s_i}/n_{d_i}$ depending on the rate $\{p_j\}_j$ decreases. For $\sum_j p_j$ to add to 1, $p_j$ cannot decrease like $c/j$ for come constant $c$, because $\sum_j c/j$ diverges to infinity. Hence $p_j \approx c/j^{1+\alpha}$ is an option. The estimate $\hat{k}_i$ satisfies

$$m_{s_i} p_{\hat{k}_i} \ge C n_{d_i}\,,$$

for some $C \in \mathbb{R}$ because the hub vertex corresponding to $p_{\hat{k}_i}$ needs have a larger degree than the degrees generated by $W$. Approximating $p_j = c/j^{1+\alpha}$ and letting $j \le \hat{k}_i$ we have

$$\begin{aligned}
p_j &\ge C \frac{n_{d_i}}{m_{s_i}}\,, \\
\frac{c}{j^{1+\alpha}} &\ge C \frac{n_{d_i}}{m_{s_i}}\,, \\
j^{1+\alpha} &\le C' \frac{m_{s_i}}{n_{d_i}}\,, \\
j &\le \left(C' \frac{m_{s_i}}{n_{d_i}}\right)^{\frac{1}{1+\alpha}}\,,
\end{aligned}$$

making

$$\hat{k}_i = \max\left\{ j : j \le \left(C' \frac{m_{s_i}}{n_{d_i}}\right)^{\frac{1}{1+\alpha}} \right\}\,.$$

As this is true for any $\alpha > 0$ we have

$$\hat{k}_i \in O(m_{s_i}/n_{d_i})\,.$$

This makes

$$\begin{aligned}
\mathbb{E}(Y_i) &= m_{s_i} \sum_{\ell=1}^{\hat{k}_i} p_\ell + \frac{\hat{k}_i c m_{new_i}}{n_{s_i}}\,, \\
\lim_{i\to\infty} \frac{1}{m_{s_i}} \mathbb{E}(Y_i) &= \lim_{i\to\infty} \sum_{\ell=1}^{\hat{k}_i} p_\ell + \frac{\hat{k}_i c m_{new_i}}{m_{s_i} n_{s_i}}\,, \\
&= \sum_{\ell=1}^{\hat{k}_i} p_\ell\,,
\end{aligned}$$

following the same reasoning as in Proposition 5.6 and because $\hat{k}_i \in O(m_{s_i}/n_{d_i})$. That is, $\hat{k}_i$ is bounded by the rate $m_{s_i}/n_{d_i}$, and it is offset by $m_{new_i}/(m_{s_i} n_{s_i})$, which behaves like $n_{d_i}^2/m_{s_i}^2$. Similarly,

$$\lim_{i\to\infty} \frac{1}{m_{s_i}} \mathrm{Var}(Y_i) = \sum_{\ell=1}^{\hat{k}_i} p_\ell \left(1 - \sum_{\ell=1}^{\hat{k}_i} p_\ell\right)\,.$$

Therefore the term $\hat{k}_i$ does not affect the expectation or the variance estimates in a substantial way. The rest is the same as in Proposition 5.6.

As in Proposition 5.6 using the second order Taylor series approximation of the expectation of a ratio of two random variables

$$\mathbb{E}\left(\frac{X_i}{Y_i}\right) = \frac{\mathbb{E}(X_i}{\mathbb{E}(Y_i)} - \frac{\mathrm{Cov}(X_i, Y_i)}{\mathbb{E}(Y_i)^2} + \frac{\mathbb{E}(X_i)\mathrm{Var}(Y_i)}{\mathbb{E}(Y_i)^3}\,. \tag{25}$$

we obtain

$$\begin{aligned}
\lim_{i\to\infty} \mathbb{E}\left(\frac{X_i}{Y_i}\right) &= \lim_{i\to\infty} \frac{m_{s_i} p_j}{m_{s_i} \sum_{\ell=1}^{\hat{k}_i} p_\ell}\,, \\
&= \lim_{i\to\infty} \frac{p_j}{\sum_{\ell=1}^{\hat{k}_i} p_\ell}\,, \\
&= p_j\,,
\end{aligned}$$

and

$$\left| \mathbb{E}\left(\frac{X_i}{Y_i}\right) - p_j \right| \leq \frac{c}{m_{s_i}}\,.$$

Similarly, the first order Taylor approximation of the variance is given by

$$\mathrm{Var}\left(\frac{X_i}{Y_i}\right) = \frac{\mathbb{E}(X_i)^2}{\mathbb{E}(Y_i)^2} \left(\frac{\mathrm{Var}(X_i)}{\mathbb{E}(X_i)^2} - 2\frac{\mathrm{Cov}(X_i, Y_i)}{\mathbb{E}(X_i)\mathbb{E}(Y_i)} + \frac{\mathrm{Var}(Y_i)}{\mathbb{E}(Y_i)^2}\right)\,.$$

The terms inside the parenthesis go to zero as $i$ goes to infinity making

$$\mathrm{Var}\left(\frac{X_i}{Y_i}\right) = 0\,,$$

and

$$\mathrm{Var}\left(\frac{X_i}{Y_i}\right) \leq \frac{c}{m_{s_i}}\,.$$

□

**Lemma E.4.** *Let $\{G_{n_i}\}_i$ be a sequence of sparse $(U, W)$-mixture graphs (Definition 3.2) with dense and sparse parts $G_{d_i}$ and $G_{s_i}$ respectively. Let $\boldsymbol{p} = (p_1, p_2, \dots)$ be the mass-partition (Definition 3.1) associated with $U$ which has infinite partitions. For a given graph $G_{n_i}$ we conduct Procedure 6.1 and estimate $\hat{k}_i$. Furthermore we estimate the mass partition $\hat{\boldsymbol{p}} = (\hat{p}_i, \hat{p}_2, \dots)$ using equation (6). Suppose $p_i < 1/(i+1)^{1+\alpha}$, i.e., $\alpha$ gives an upper bound for the rate at which $p_i$ goes to zero. Then*

$$\sum_{j=\hat{k}_i}^{\infty} p_j \leq \frac{1}{\alpha \hat{k}_i^{\alpha}} \,.$$

*Proof.* We consider $f(x) = \frac{1}{x^{1+\alpha}}$ and have

$$\frac{1}{(j+1)^{1+\alpha}} < \int_j^{j+1} \frac{1}{x^{1+\alpha}}\, dx \,,$$

$$\sum_{j=\hat{k}_i}^{\infty} p_j < \sum_{j=\hat{k}_i}^{\infty} \frac{1}{(j+1)^{1+\alpha}} \leq \int_{\hat{k}_i}^{\infty} \frac{1}{x^{1+\alpha}}\, dx \,,$$

$$= \frac{1}{\alpha \hat{k}_i^{\alpha}} \,.$$

□

# F Experiments

## F.1 Illustration with synthetic data

### F.1.1 Degree prediction

To predict the top-$k$ degrees using synthetic data we use $(U, W)$-mixture graphs with 4 combinations. We consider $W_1 = \exp(-(x+y))$, $W_2 = 0.1$ and $U_1$ with mass-partitions $\boldsymbol{p}_1 \propto \{1/j^{1.2}\}_{j=2}^{50}$ and $U_2$ with mass-partition $\boldsymbol{p}_2 \propto \{1/1.2^j\}_{j=2}^{50}$. The mass-partitions are normalized so that they add up to 1. For each $(U, W)$ combination we have training graphs with $n_i = 11{,}000$ nodes and test graphs with $n_j = 13{,}200$ nodes. We then predict the top-$k$ degrees in graph $G_{n_j}$ using equation (1) for $k = \hat{k}_i$ in each instance. For the four experiments the estimated $\hat{k}_i$ values were $8, 13, 14, 22$ respectively.

As a baseline comparison method we used a scale-free graph property discussed in Bollobás et al. (2001). For a scale-free graph with $n$ nodes the maximum degree behaves like $\Theta(\sqrt{n})$. Suppose $\deg_{G_{n_i}} v_{(\ell)}$ denotes the $\ell$th highest degree in $G_{n_i}$. Then, for an unseen graph $G_{n_j}$ with $n_j$ nodes we estimate the $\ell$th highest degree using

$$\deg_{G_{n_j}} \hat{v}_{(\ell)} = \deg_{G_{n_i}} v_{(\ell)} \times \sqrt{\frac{n_j}{n_i}} \,, \tag{26}$$

as a means of comparison. Figures 22 and 23 show the results of the four experiments using equations 1 and (26) respectively.

### F.1.2 Estimating $U$ with finite partitions

We conduct 4 experiments with $W(x, y) = \exp(-(x+y))$ and different mass-partitions $\boldsymbol{p}$ corresponding to $U$. We consider $\boldsymbol{p}_1 = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)$ for experiment 1, $\boldsymbol{p}_2 = (0.27, 0.26, 0.24, 0.23)$ for experiment 2, $\boldsymbol{p}_3 = (0.1, 0.1, \cdots, 0.1)$, i.e., 0.1 repeated 10 times for experiment 3 and $\boldsymbol{p}_4 \propto \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}\right)$ in experiment 4 where the mass-partition is obtained by dividing by $\sum_{i=2}^{6} 1/i$. For each experiment we use a set of graphs with a given number of nodes and we estimate $\hat{k}_i$ using equation (5), which gives the same result as equation (3). Using $\hat{k}_i$ and equation (4) we estimate the mass-partition.

As a comparison method we use

$$\hat{p}_j = \frac{\deg_{G_{n_i}} v_{(j)}}{\sum_{\ell=1}^{n_i} \deg_{G_{n_i}} v_{\ell)}} \,, \tag{27}$$

i.e., the ratio of the $j$th highest degree to the sum of all degrees. Figures 24 and 25 show the results of the 4 experiments using equations (4) and (27) respectively.

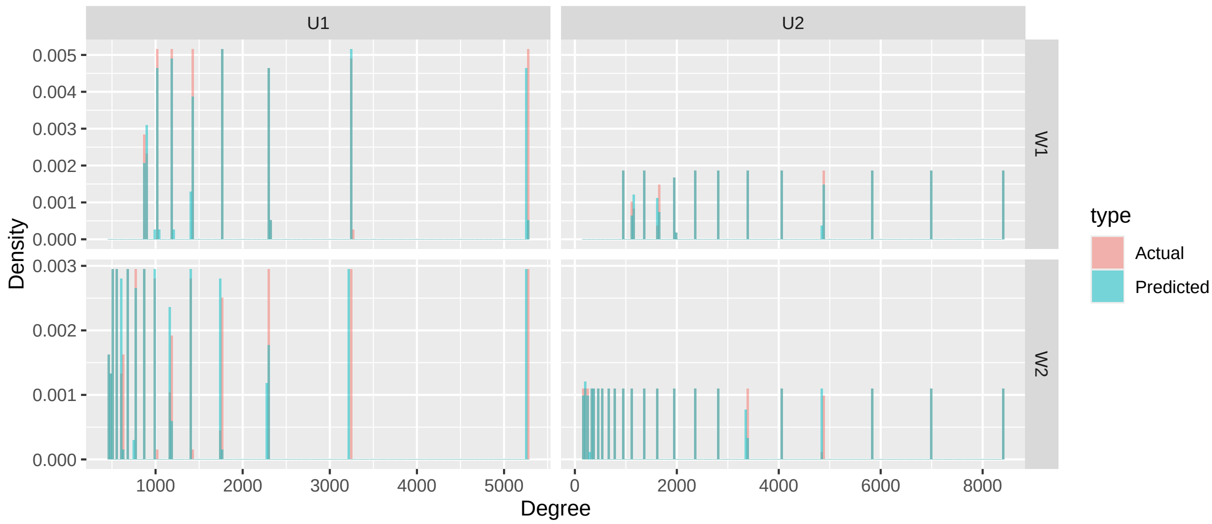



Figure 22: Densities of predicted degree using equation (1) for experiments with $(W_1, U_1)$, $(W_1, U_2)$, $(W_2, U_1)$ and $(W_2, U_2)$-mixture graphs.

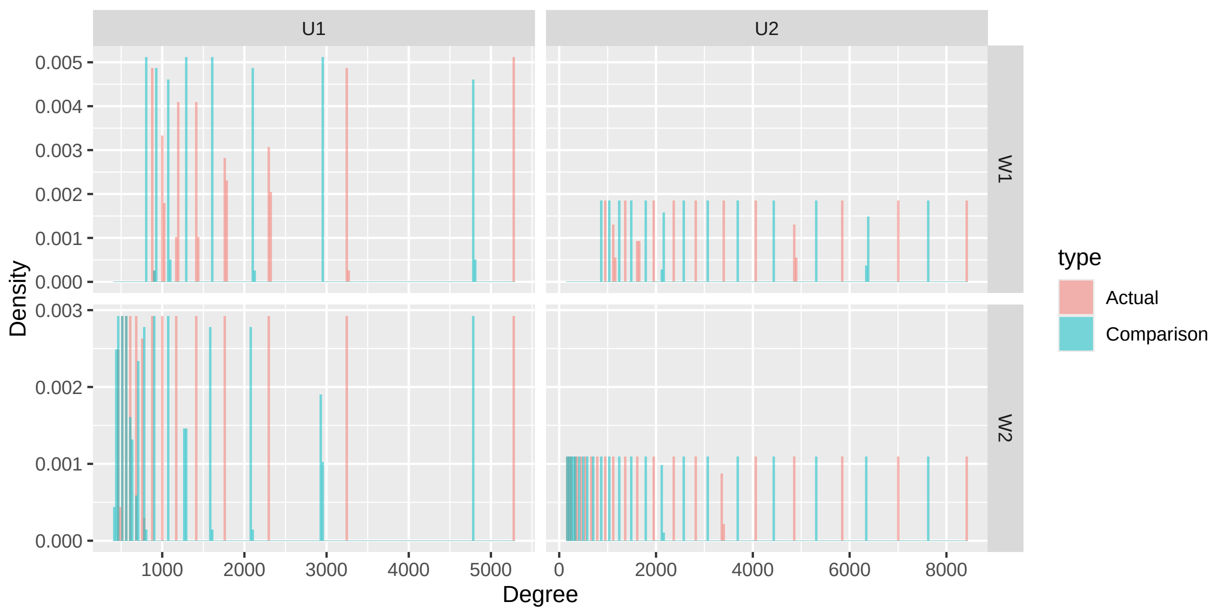



Figure 23: Densities of predicted degrees using equation (26) for experiments with $(W_1, U_1)$, $(W_1, U_2)$, $(W_2, U_1)$ and $(W_2, U_2)$-mixture graphs.

### F.1.3 Estimating $U$ with infinite partitions

We use $W(x, y) = \exp(-(x+y))$ for this set of experiments as well. In practice we cannot generate a mass-partition with infinite non-zero elements. As such we consider a finite number of partitions from an infinite sequence. For the 4 experiments we consider these sequences: for experiment 1 we want the $j$th element of the mass-partition $p_j \propto \frac{1}{j+1^{1.2}}$. Noting that we cannot have a mass partition proportional to $\frac{1}{j}$ we consider a relatively low exponent of 1.2. For experiment 2 we consider $p_j \propto \frac{1}{1.2^{j+1}}$. This is a geometric series. For experiment 3 we consider $p_j \propto \frac{1}{(j+1)\log(j+1)}$. For experiment 4 we consider $p_j \propto \frac{1}{(j+1)!}$. For all experiments we consider $\boldsymbol{p}$ to have 49 elements and rescale the mass-partition to add up to 1.

For each experiment we use a set of graphs and using equation (5) we estimate $\hat{k}_i$, which are 30, 23, 30 and 4 for the four experiments. Figures 26 and 27 show the estimated and actual $p_j$ values using equations (4) and (27)

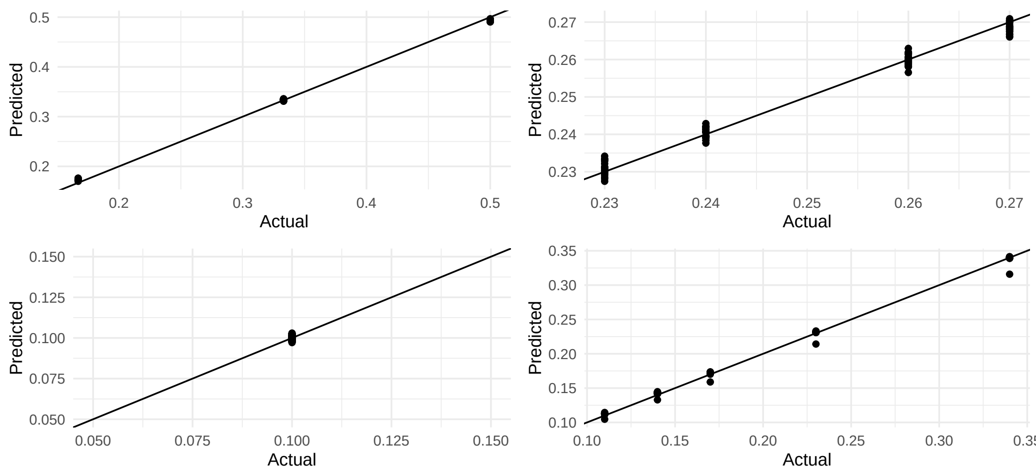



Figure 24: Experiments for $U$ with finite partitions. The actual $p_j$ in mass-partition in $\boldsymbol{p} = (p_1, \ldots, p_n, 0, \ldots)$ is given on the $x$-axis and the predicted $\hat{p}_j$ using equation (4) on the $y$-axis.

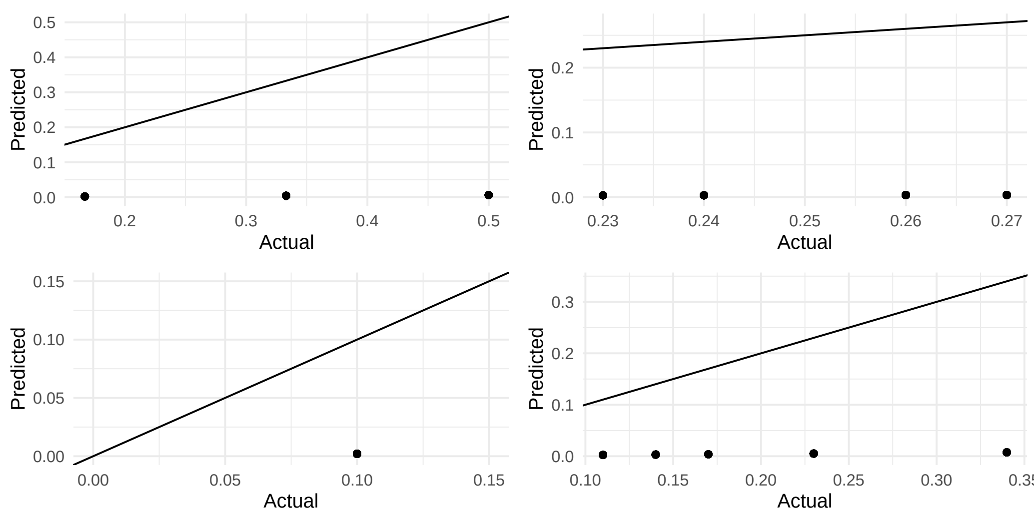



Figure 25: Experiments for $U$ with finite partitions. The actual $p_j$ in mass-partition in $\boldsymbol{p} = (p_1, \ldots, p_n, 0, \ldots)$ is given on the $x$-axis and the predicted $\hat{p}_j$ using equation (27) on the $y$-axis.

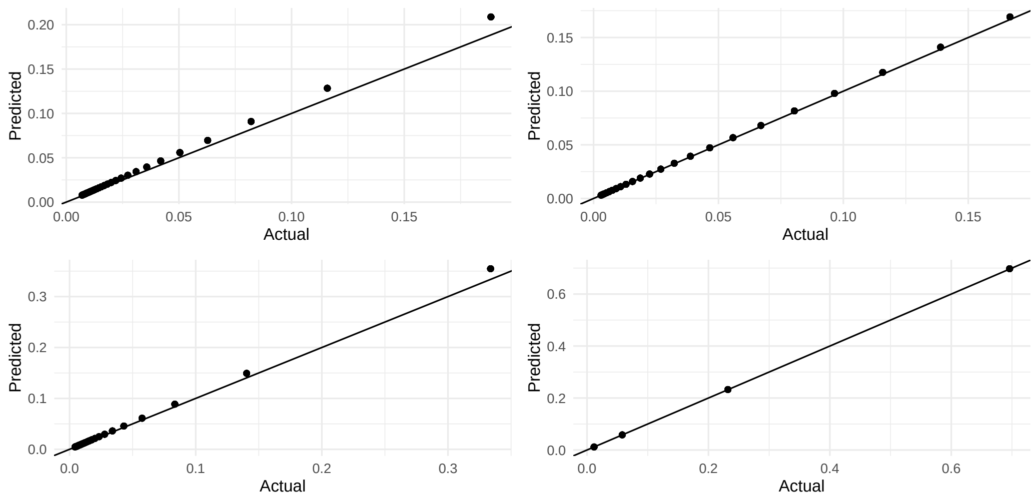



Figure 26: Experiments for infinite mass partition $\boldsymbol{p}$ with experiment 1 (top left), 2 (top right), 3 (bottom left) and 4 (bottom right). Equation (4) is used to estimate $p_i$.

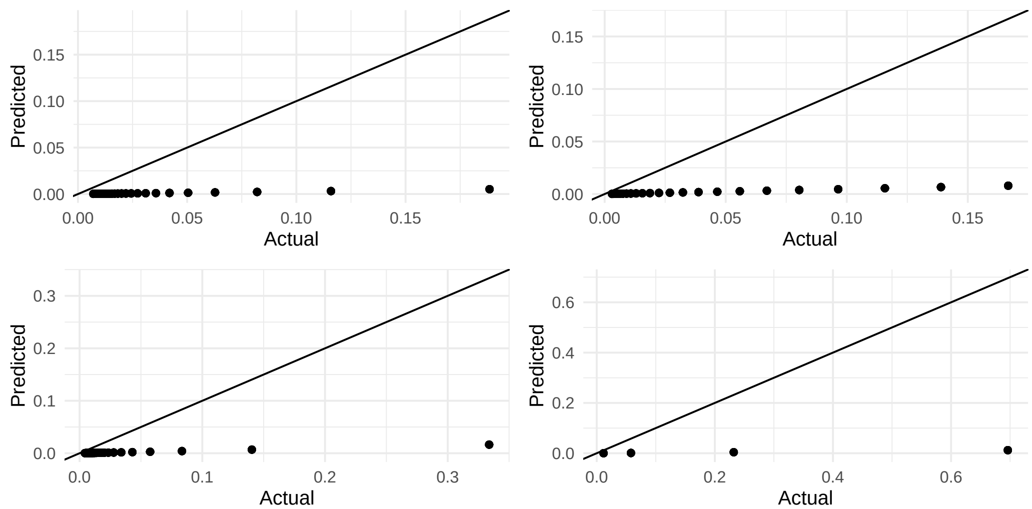



Figure 27: Experiments for infinite mass partition $\boldsymbol{p}$ with experiment 1 (top left), 2 (top right), 3 (bottom left) and 4 (bottom right). Equation (27) is used to estimate $p_i$.

respectively.

For fast decreasing sequences such as $\frac{1}{(j+1)!}$ a smaller number of elements contribute to a larger sum $\sum_{j=1}^{\hat{k}_i} p_j$ and as such both $k_i$ (the number of hubs with degrees greater than those generated by $W$) and the estimate $\hat{k}_i$ are small. Notwithstanding this, the estimate $\hat{p}_j$ is quite accurate because the sum $\sum_{j=1}^{\hat{k}_i} p_j$ is quite large. In contrast, for slowly decreasing sequences such as $\frac{1}{j+1^{1.2}}$, as the proportions decrease slowly, the estimate $\hat{k}_i$ is large, but the value $\sum_{j=1}^{\hat{k}_i} p_j$ is not that large. As such $\hat{p}_j$ deviates from $p_j$ quite a bit for small $j$. This is seen in the topleft subplot of Figure 26, which corresponds to experiment 1. The estimate $\hat{k}_i$ and the sum $\sum_{j=1}^{\hat{k}_i} p_j$ for these 4 experiments are given in Table 6.

### F.2 Real datasets

We use Facebook (FB) links (Viswanath et al., 2009), Hep-PH Physics citations (Leskovec et al., 2005; Gehrke et al., 2003), MOOC interactions (Kumar et al., 2019), SMS (Wu et al., 2010), UCI Messages (Panzarasa et al., 2009) and Yahoo messages (Rossi and Ahmed, 2015) datasets. Each dataset is given as an edgelist with

Table 6: The estimates $\hat{k}_i$ and $\sum_{j=1}^{\hat{k}_i} p_j$ for infinite $U$ experiments.

| Task | Description | Experiment 1 | Experiment 2 | Experiment 3 | Experiment 4 |
|---|---|---|---|---|---|
| Infinite $U$ | $\hat{k}_i$ | 30 | 23 | 30 | 4 |
| | $\sum_{j=1}^{\hat{k}_i} p_j$ | 0.902 | 0.985 | 0.941 | 0.998 |

timestamped edges. The edges accumulate at different speeds in each dataset. In Yahoo messages dataset the edges accumulate quite fast and as such we use a 2-hourly time window to construct graphs. In contrast, edges accumulate slowly in Hep-PH citation dataset and we use monthly graphs. For the other datasets we use daily graphs. In each example we consider growing networks, i.e., new edges and nodes are added to the existing graph as time passes.

Each dataset comprises a sequence of growing graphs $\{G_{n_i}\}_i$ and we consider $G_{n_i}$ for $i \in \{20, \ldots, 24\}$ as training graphs. For each $G_{n_i}$ we select the test graph $G_{n_j}$ such that $j = \min_k\{|G_{n_k}| : |G_{n_k}| - |G_{n_i}| \geq 500\}$, i.e., $G_{n_j}$ is the first graph that has more than 500 nodes compared to $G_{n_i}$. This is to ensure that the test graph is different from the training graph. Using the training graphs we predict the top-10 degrees of the test graphs.

For Bollobás et al. (2001) we used equation (26) to predict degrees. Caron and Fox (2017) have made their code available at `https://www.stats.ox.ac.uk/~caron/code/bnpgraph/`. We used their code to train their graph generation model on the training graph $G_{n_i}$ and using the trained model generated a graph having $n_k$ nodes, which we compared with $G_{n_k}$. Kronecker graphs have 2 settings: deterministic and stochastic, and their code is available at `https://github.com/snap-stanford/snap/tree/master/examples/krongen`. We generated Kronecker graphs with (1) the default setting (stochastic $2 \times 2$ initiator matrix), (2) stochastic setting with a $3 \times 3$ initiator matrix and (3) the deterministic setting with a $3 \times 3$ initiator matrix. For the stochastic settings, we first fitted a Kronecker model on the training graph $G_{n_i}$ and then using the fitted parameters generated the graph with $n_k$ nodes, which we compared with the test graph $G_{n_k}$. Leskovec et al. (2010) explored the initiator matrix resulting from a 3-star ($K_{1,3}$) with self loops added at each node. The adjacency matrix of this graph is given by

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} . \tag{28}$$

We explored the deterministic case with this initiator matrix. Using the predicted top-10 degrees from each method, we computed the average MAPE and the standard deviation, which is given in Table 3. We compared the performance of the top 2 methods on each dataset using the Student's t-test with $\alpha = 0.1$. For all datasets except UCI Messages we found that $(U, W)$-mixture was significantly better than Bollobás et al. (2001), which was the ranked the second. For UCI, while $(U, W)$-mixtures performed better on average, it was not significantly better. Figures 28 to 33 show the actual and predicted degrees of the test graphs with Actual = Predicted line drawn. In some instances, the $y = x$ line is out of range.

#### F.2.1 HEP-TH dataset comparison as in Leskovec et al. (2005)

Leskovec et al. (2010) have used the initiator matrix $A$ given in equation (28) to generate graphs similar to the HEP-TH dataset using the deterministic setting. For the stochastic setting, they have replaced 1s with $\alpha = 0.41$ and 0s with $\beta = 0.11$. With these two initiators we ran their algorithm and verified we could produce the degree distribution in their paper. In their paper they say that Kronecker graph results qualitatively match the original results. We found this to be true in our simulation as well. Table 7 gives the actual and predicted top-10 degrees using $(U, W)$-mixture graphs, Kronecker deterministic and stochastic versions and the absolute percentage error $|\hat{y} - y| \times 100/y$.

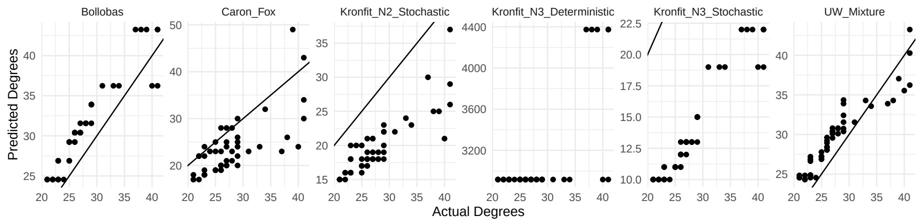



Figure 28: Facebook links dataset

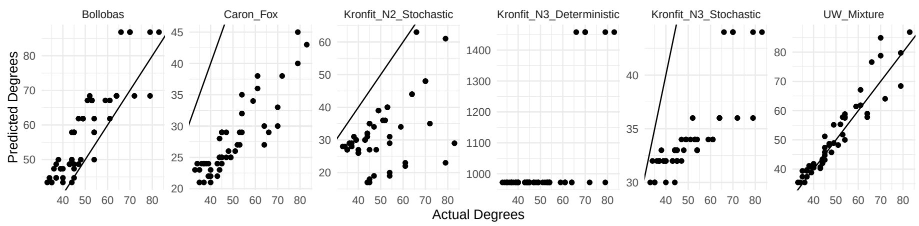



Figure 29: HEP-PH dataset

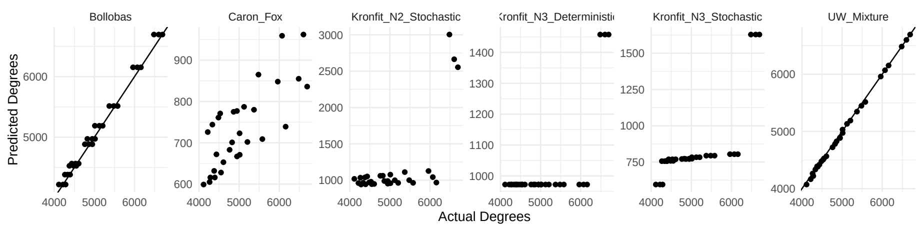



Figure 30: MOOC dataset

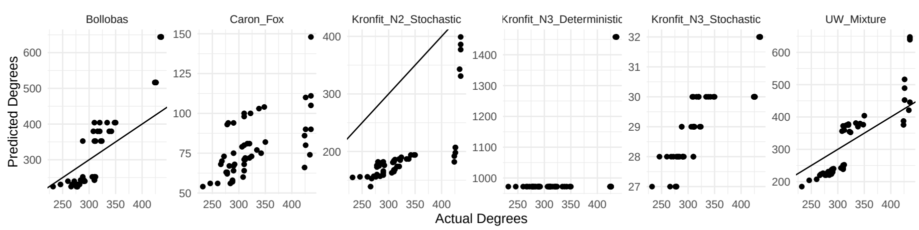



Figure 31: UCI Messages dataset

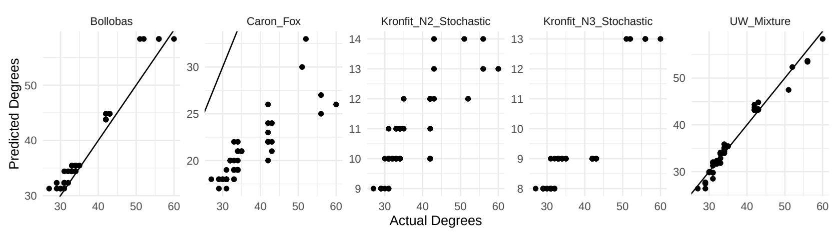



Figure 32: SMS dataset

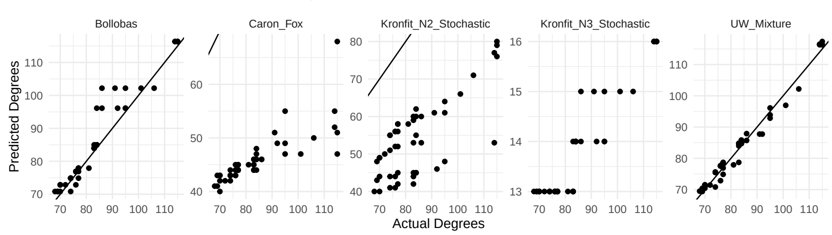



Figure 33: Yahoo dataset

Table 7: The top-10 degree comparison results on HEP-TH dataset where APE denotes the absolute percentage error $|\hat{y} - y| \times 100/y$.

| Actual | Predicted $(U, W)$ degree | Predicted Kronecker Deterministic. | Predicted Kronecker Stochastic | APE $(U, W)$ | APE Kronecker Det. | APE Kronecker Stoch. |
|---|---|---|---|---|---|---|
| 2468 | 2249 | 32768 | 74 | 8.89 | 1227.71 | 97.00 |
| 1797 | 1633 | 16384 | 51 | 9.14 | 811.74 | 97.16 |
| 1653 | 1503 | 16384 | 47 | 9.09 | 891.17 | 97.16 |
| 1369 | 1281 | 16384 | 45 | 6.44 | 1096.79 | 96.71 |
| 1308 | 1253 | 16384 | 44 | 4.19 | 1152.60 | 96.64 |
| 1219 | 1147 | 16384 | 44 | 5.92 | 1244.05 | 96.39 |
| 1218 | 1124 | 16384 | 43 | 7.69 | 1245.16 | 96.47 |
| 1165 | 1101 | 16384 | 42 | 5.51 | 1306.35 | 96.39 |
| 1124 | 1019 | 16384 | 42 | 9.35 | 1357.65 | 96.26 |
| 1038 | 996 | 16384 | 42 | 4.01 | 1478.42 | 95.95 |