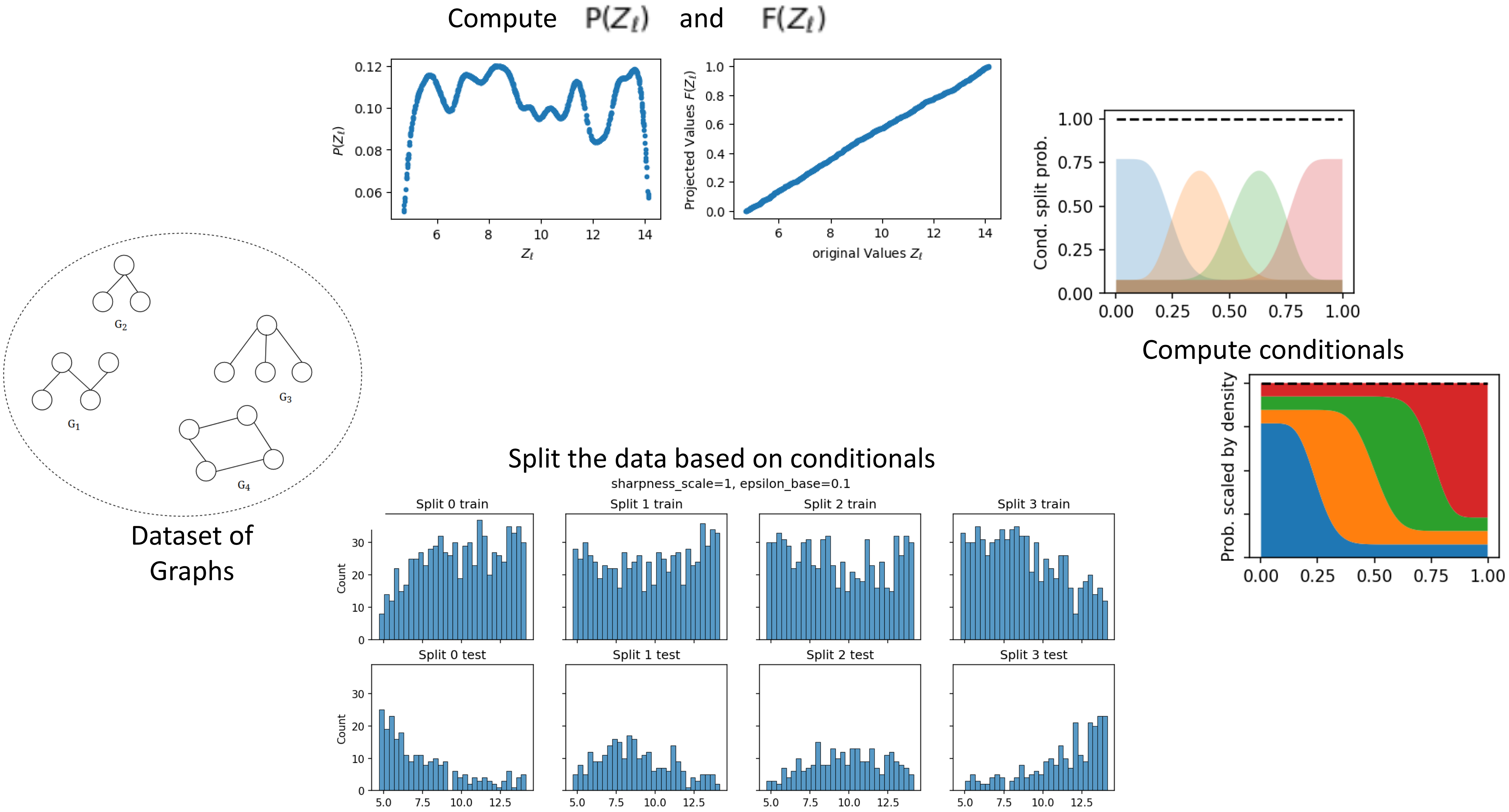
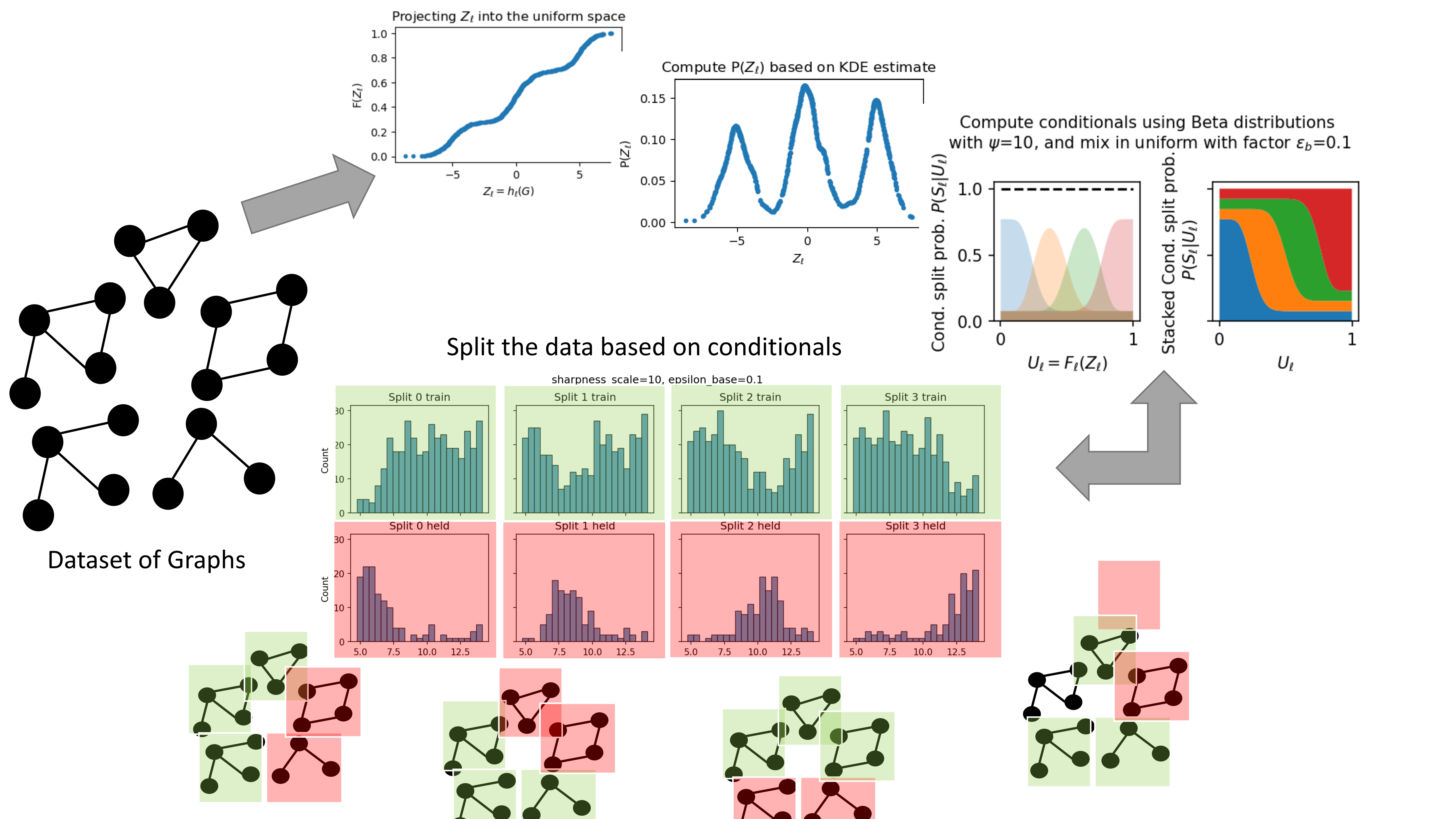
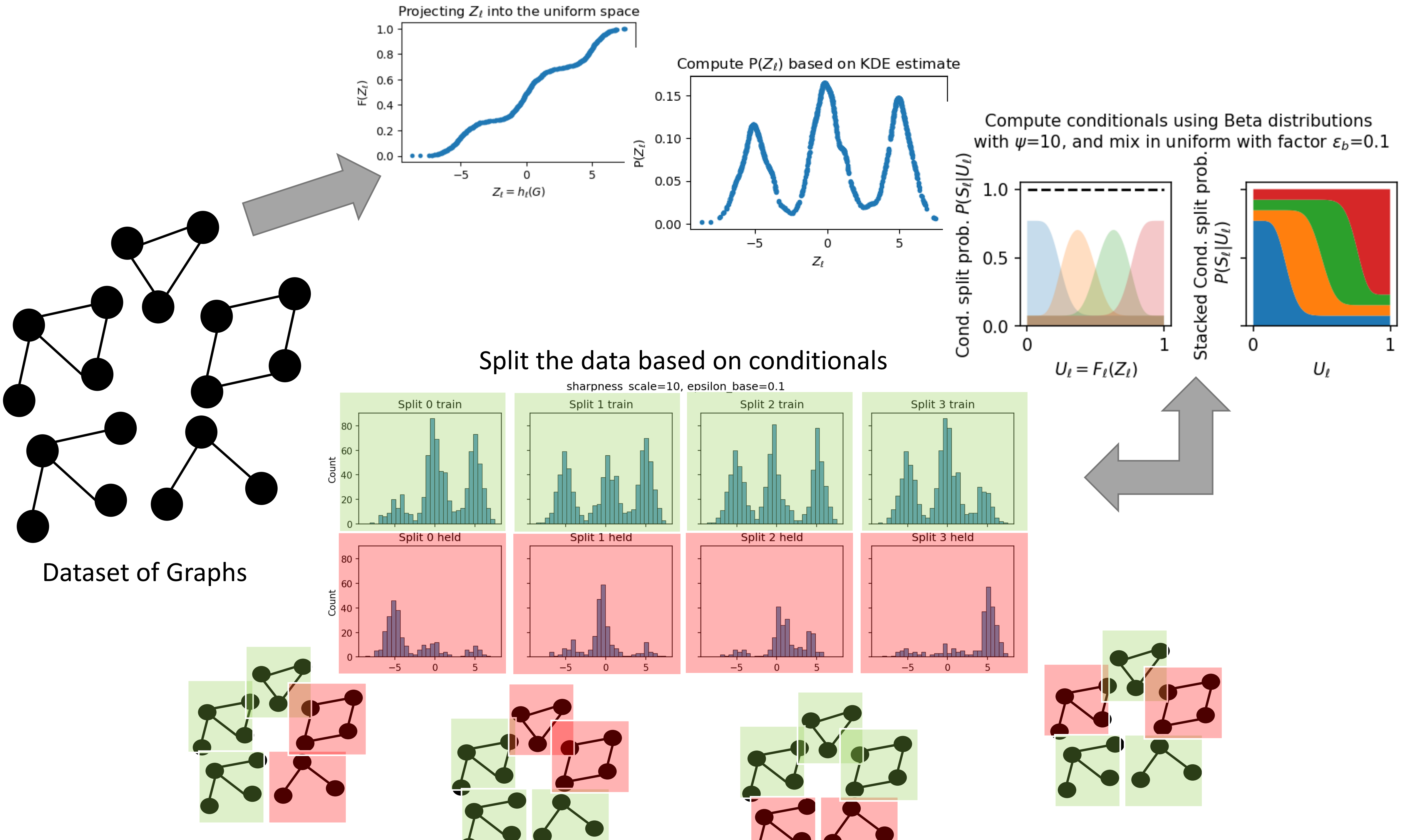
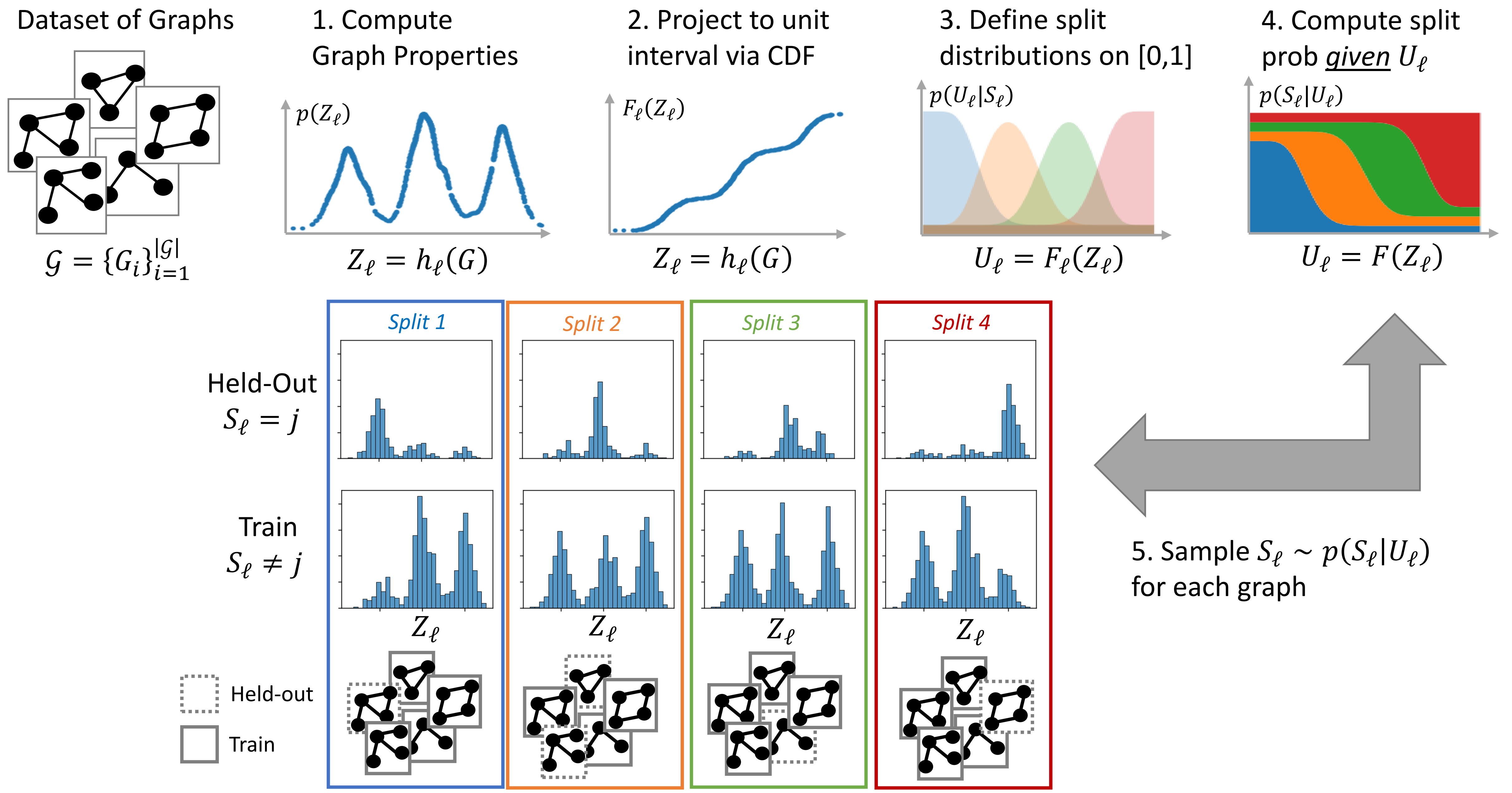


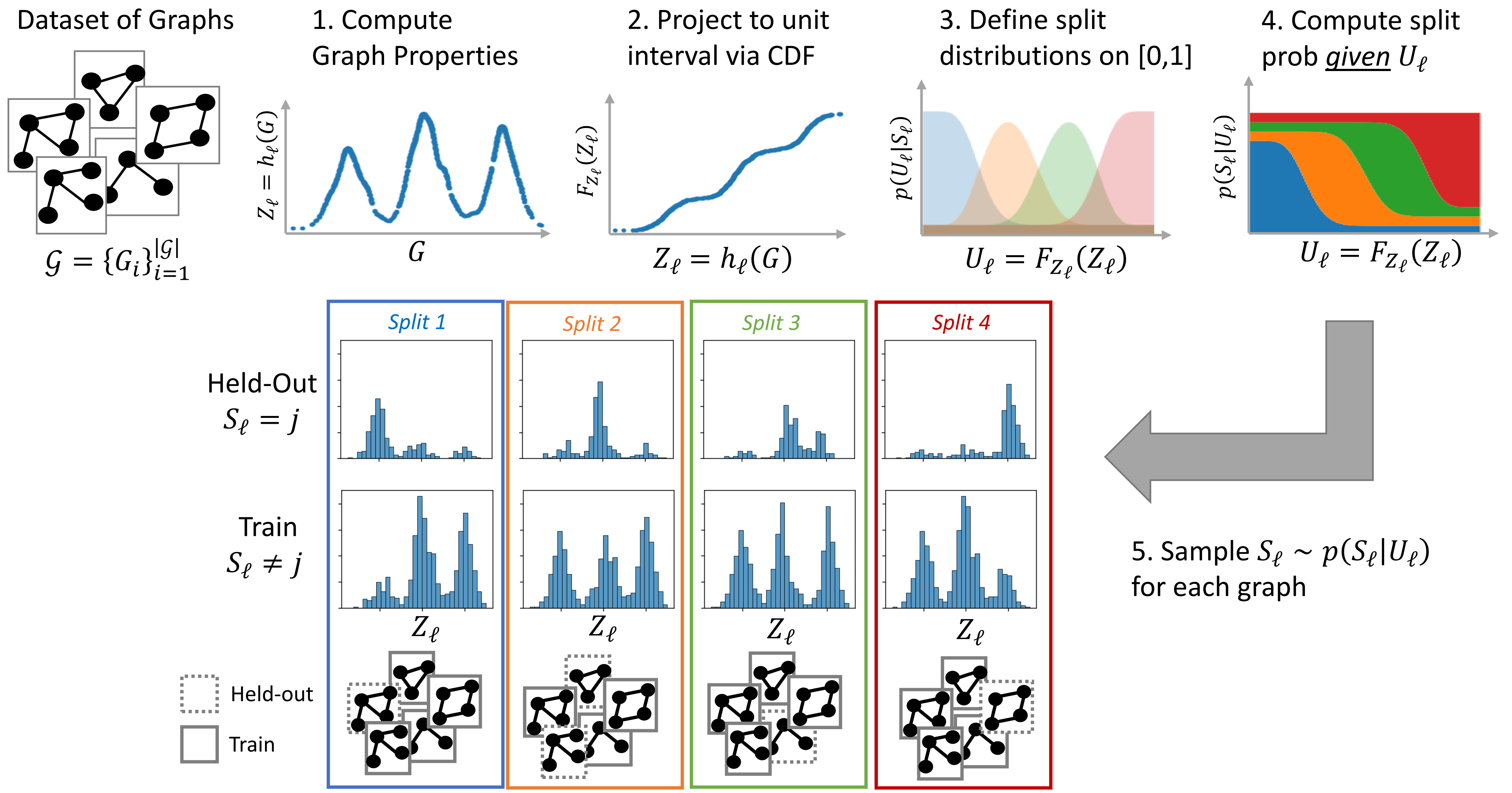
Compute $P(Z_\ell)$ and $F(Z_\ell)$
$P(Z_\ell)$
Z_ℓ
Projected Values $F(Z_\ell)$
original Values Z_ℓ
Cond. split prob.
Compute conditionals
Prob. scaled by density
G_1
G_2
G_3
G_4
Dataset of Graphs
Split the data based on conditionals
sharpness_scale=1, epsilon_base=0.1
Split 0 train
Split 1 train
Split 2 train
Split 3 train
Split 0 test
Split 1 test
Split 2 test
Split 3 test
Count


Projecting Z_ℓ into the uniform space
$F(Z_\ell)$
$Z_\ell = h_\ell(G)$
Compute $P(Z_\ell)$ based on KDE estimate
$P(Z_\ell)$
Z_ℓ
Compute conditionals using Beta distributions with ψ=10, and mix in uniform with factor ϵ_b=0.1
Cond. split prob. $P(S_\ell|U_\ell)$
$U_\ell = F_\ell(Z_\ell)$
Stacked Cond. split prob. $P(S_\ell|U_\ell)$
U_ℓ
Split the data based on conditionals
sharpness_scale=10, epsilon_base=0.1
Split 0 train
Split 1 train
Split 2 train
Split 3 train
Split 0 held
Split 1 held
Split 2 held
Split 3 held
Count
Dataset of Graphs

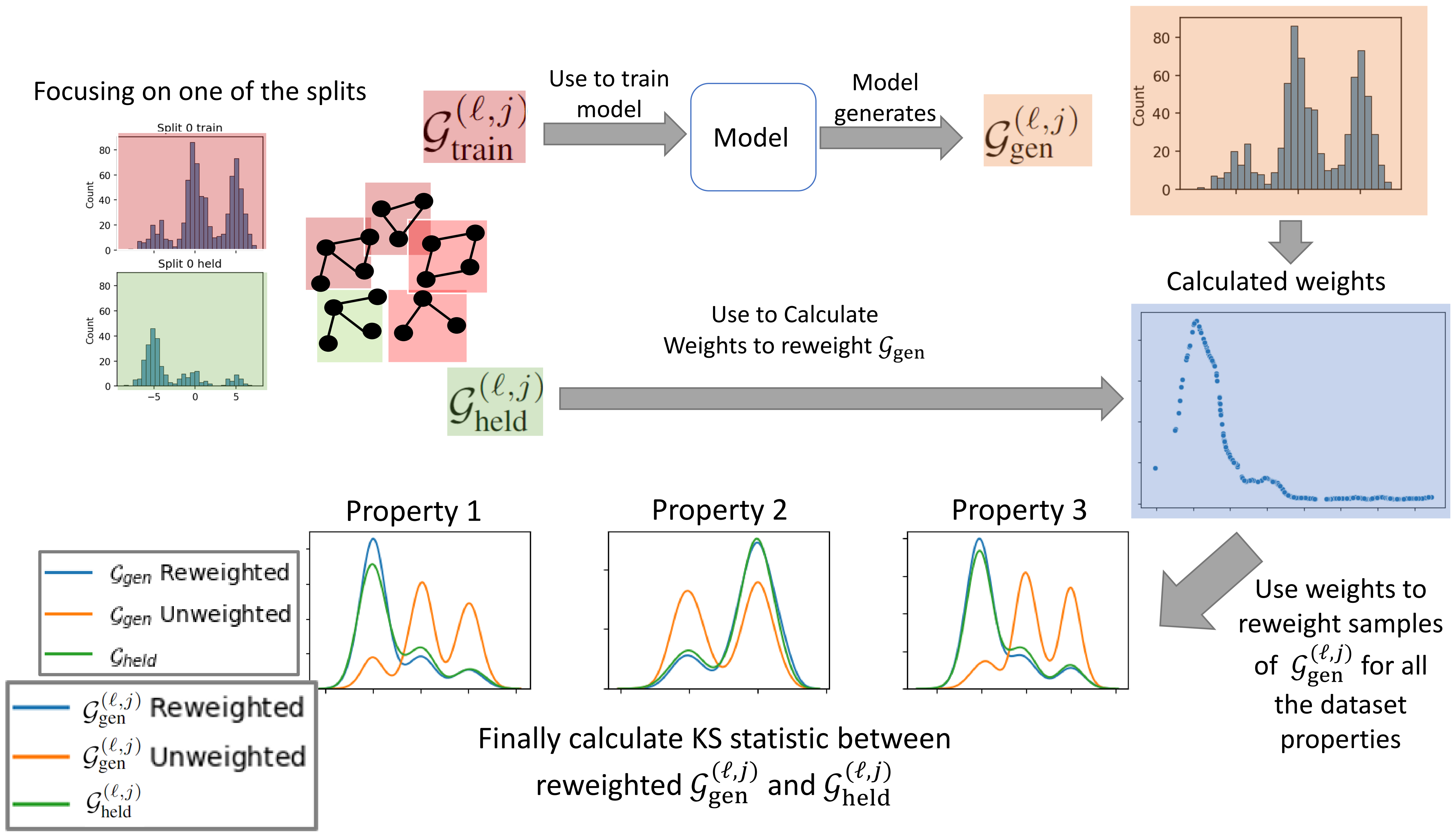


Projecting Z_ℓ into the uniform space
$F(Z_\ell)$
$Z_\ell = h_\ell(G)$
Compute $P(Z_\ell)$ based on KDE estimate
$P(Z_\ell)$
Z_ℓ
Compute conditionals using Beta distributions with ψ=10, and mix in uniform with factor ϵ_b=0.1
Cond. split prob. $P(S_\ell|U_\ell)$
$U_\ell = F_\ell(Z_\ell)$
Stacked Cond. split prob. $P(S_\ell|U_\ell)$
U_ℓ
Split the data based on conditionals
sharpness_scale=10, epsilon_base=0.1
Split 0 train
Split 1 train
Split 2 train
Split 3 train
Split 0 held
Split 1 held
Split 2 held
Split 3 held
Count
Dataset of Graphs

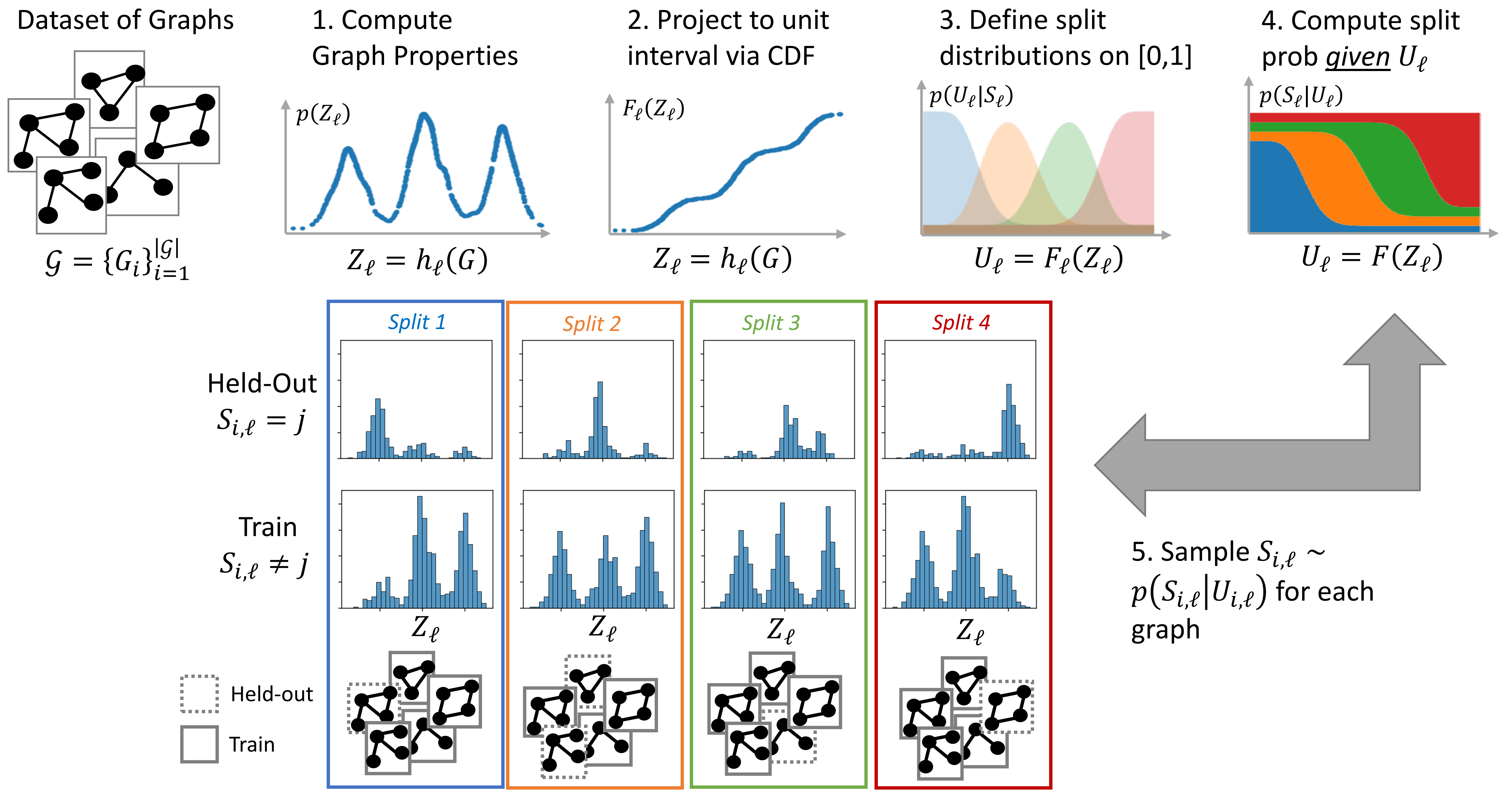

Dataset of Graphs
$\mathcal{G} = \{G_i\}_{i=1}^{|\mathcal{G}|}$
1. Compute Graph Properties
$p(Z_\ell)$
$Z_\ell = h_\ell(G)$
2. Project to unit interval via CDF
$F_\ell(Z_\ell)$
$Z_\ell = h_\ell(G)$
3. Define split distributions on [0,1]
$p(U_\ell|S_\ell)$
$U_\ell = F_\ell(Z_\ell)$
4. Compute split prob *given* U_ℓ
$p(S_\ell|U_\ell)$
$U_\ell = F(Z_\ell)$
Split 1
Split 2
Split 3
Split 4
Held-Out $S_\ell = j$
Train $S_\ell \neq j$
Z_ℓ
Z_ℓ
Z_ℓ
Z_ℓ
Held-out
Train
5. Sample $S_\ell \sim p(S_\ell|U_\ell)$ for each graph



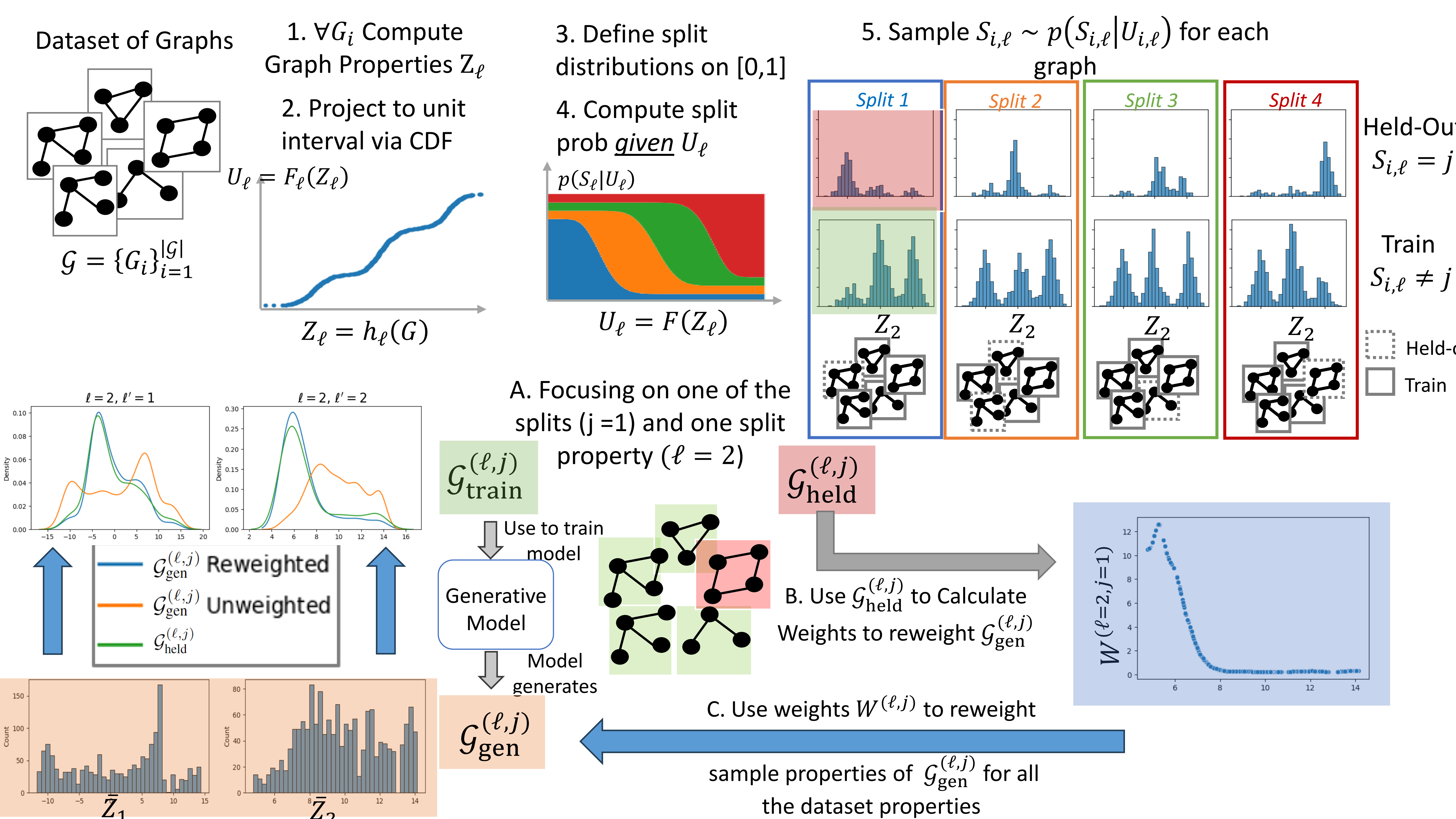
Dataset of Graphs
$\mathcal{G} = \{G_i\}_{i=1}^{|\mathcal{G}|}$
1. Compute Graph Properties
$Z_\ell = h_\ell(G)$
G
2. Project to unit interval via CDF
$F_{Z_\ell}(Z_\ell)$
$Z_\ell = h_\ell(G)$
3. Define split distributions on [0,1]
$p(U_\ell|S_\ell)$
$U_\ell = F_{Z_\ell}(Z_\ell)$
4. Compute split prob *given* U_ℓ
$p(S_\ell|U_\ell)$
$U_\ell = F_{Z_\ell}(Z_\ell)$
5. Sample $S_\ell \sim p(S_\ell|U_\ell)$ for each graph
Split 1
Split 2
Split 3
Split 4
Held-Out $S_\ell = j$
Train $S_\ell \neq j$
Z_ℓ
Held-out
Train



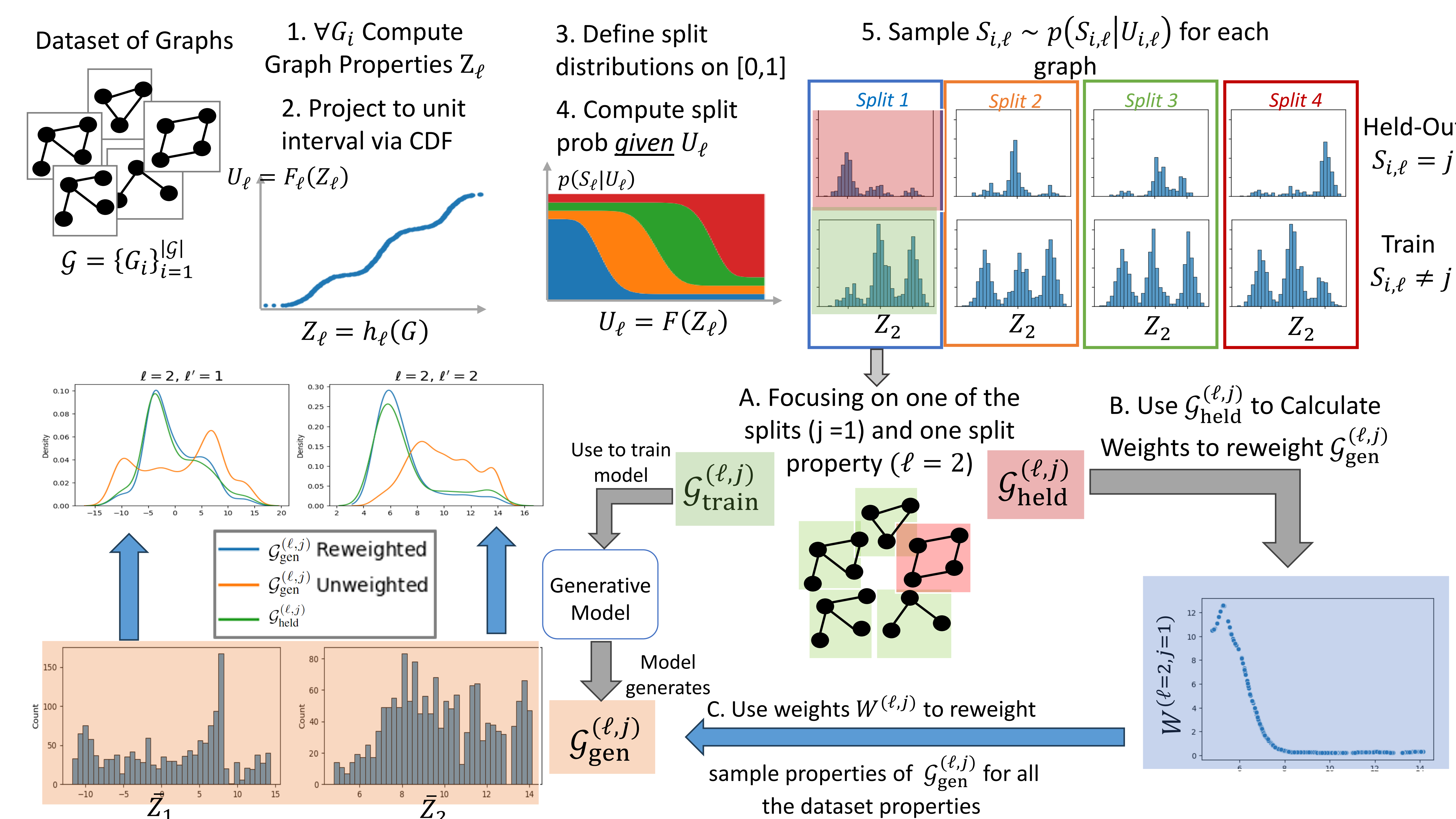
Focusing on one of the splits
Split 0 train
Split 0 held
Count
$\mathcal{G}_{\text{train}}^{(\ell,j)}$
Use to train model
Model
Model generates
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$
Calculated weights
Use to Calculate Weights to reweight $\mathcal{G}_{\text{gen}}$
$\mathcal{G}_{\text{held}}^{(\ell,j)}$
Property 1
Property 2
Property 3
$\mathcal{G}_{gen}$ Reweighted
$\mathcal{G}_{gen}$ Unweighted
$\mathcal{G}_{held}$
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$ Reweighted
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$ Unweighted
$\mathcal{G}_{\text{held}}^{(\ell,j)}$
Use weights to reweight samples of $\mathcal{G}_{\text{gen}}^{(\ell,j)}$ for all the dataset properties
Finally calculate KS statistic between reweighted $\mathcal{G}_{\text{gen}}^{(\ell,j)}$ and $\mathcal{G}_{\text{held}}^{(\ell,j)}$



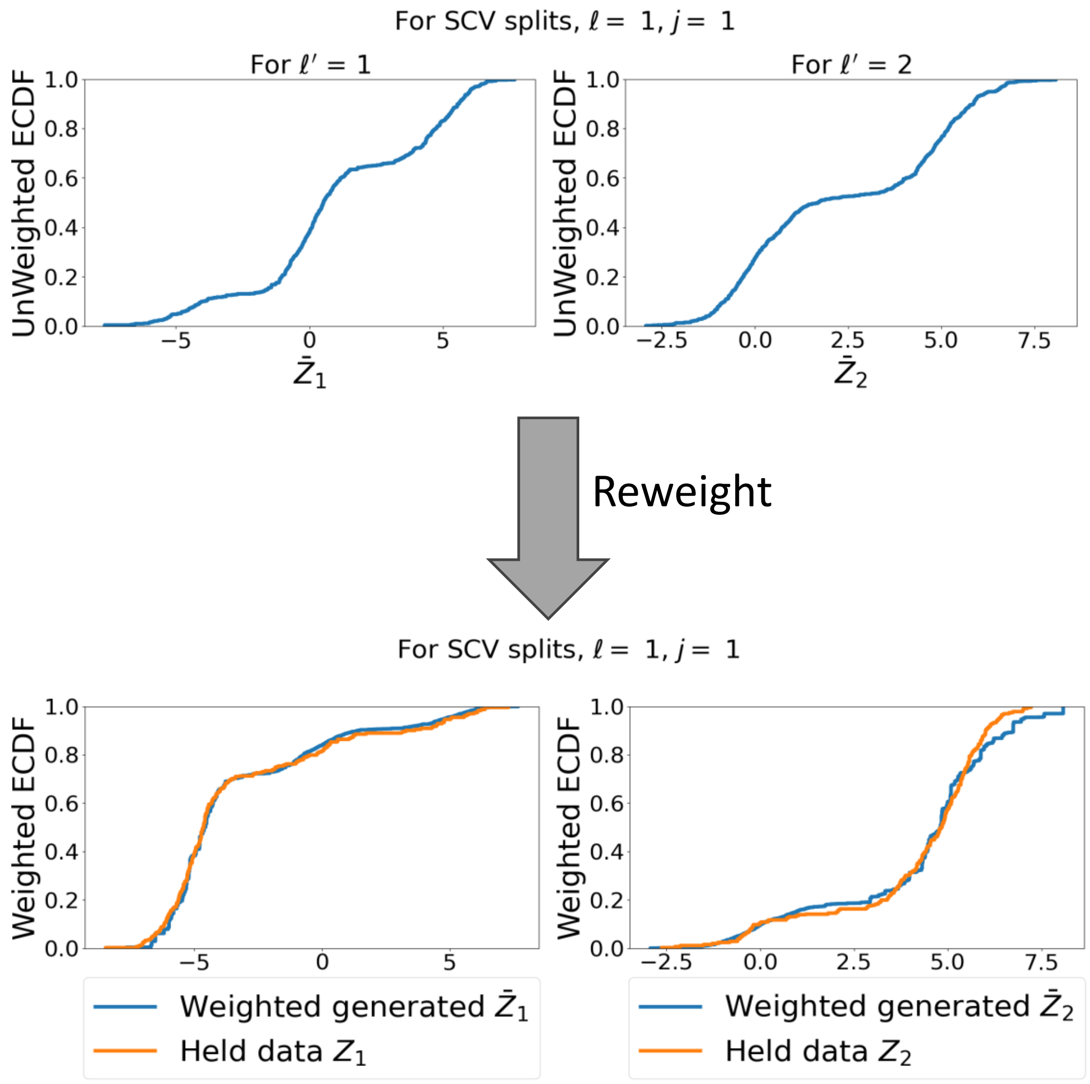
Dataset of Graphs
$\mathcal{G} = \{G_i\}_{i=1}^{|\mathcal{G}|}$
1. Compute Graph Properties
$p(Z_\ell)$
$Z_\ell = h_\ell(G)$
2. Project to unit interval via CDF
$F_\ell(Z_\ell)$
$Z_\ell = h_\ell(G)$
3. Define split distributions on [0,1]
$p(U_\ell|S_\ell)$
$U_\ell = F_\ell(Z_\ell)$
4. Compute split prob *given* U_ℓ
$p(S_\ell|U_\ell)$
$U_\ell = F(Z_\ell)$
Split 1
Split 2
Split 3
Split 4
Held-Out $S_{i,\ell} = j$
Train $S_{i,\ell} \neq j$
Z_ℓ
Held-out
Train
5. Sample $S_{i,\ell} \sim p(S_{i,\ell}|U_{i,\ell})$ for each graph



Dataset of Graphs
$\mathcal{G} = \{G_i\}_{i=1}^{|\mathcal{G}|}$
1. $\forall G_i$ Compute Graph Properties Z_ℓ
2. Project to unit interval via CDF
$U_\ell = F_\ell(Z_\ell)$
$Z_\ell = h_\ell(G)$
3. Define split distributions on [0,1]
4. Compute split prob *given* U_ℓ
$p(S_\ell|U_\ell)$
$U_\ell = F(Z_\ell)$
5. Sample $S_{i,\ell} \sim p(S_{i,\ell}|U_{i,\ell})$ for each graph
Split 1
Split 2
Split 3
Split 4
Z_2
Held-Out $S_{i,\ell} = j$
Train $S_{i,\ell} \neq j$
Held-out
Train
A. Focusing on one of the splits (j =1) and one split property ($\ell = 2$)
$\mathcal{G}_{\text{train}}^{(\ell,j)}$
$\mathcal{G}_{\text{held}}^{(\ell,j)}$
Use to train model
Generative Model
Model generates
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$
B. Use $\mathcal{G}_{\text{held}}^{(\ell,j)}$ to Calculate Weights to reweight $\mathcal{G}_{\text{gen}}^{(\ell,j)}$
$W^{(\ell=2,j=1)}$
C. Use weights $W^{(\ell,j)}$ to reweight sample properties of $\mathcal{G}_{\text{gen}}^{(\ell,j)}$ for all the dataset properties
$\ell = 2,\ \ell' = 1$
$\ell = 2,\ \ell' = 2$
Density
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$ Reweighted
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$ Unweighted
$\mathcal{G}_{\text{held}}^{(\ell,j)}$
Count
$\bar{Z}_1$
$\bar{Z}_2$

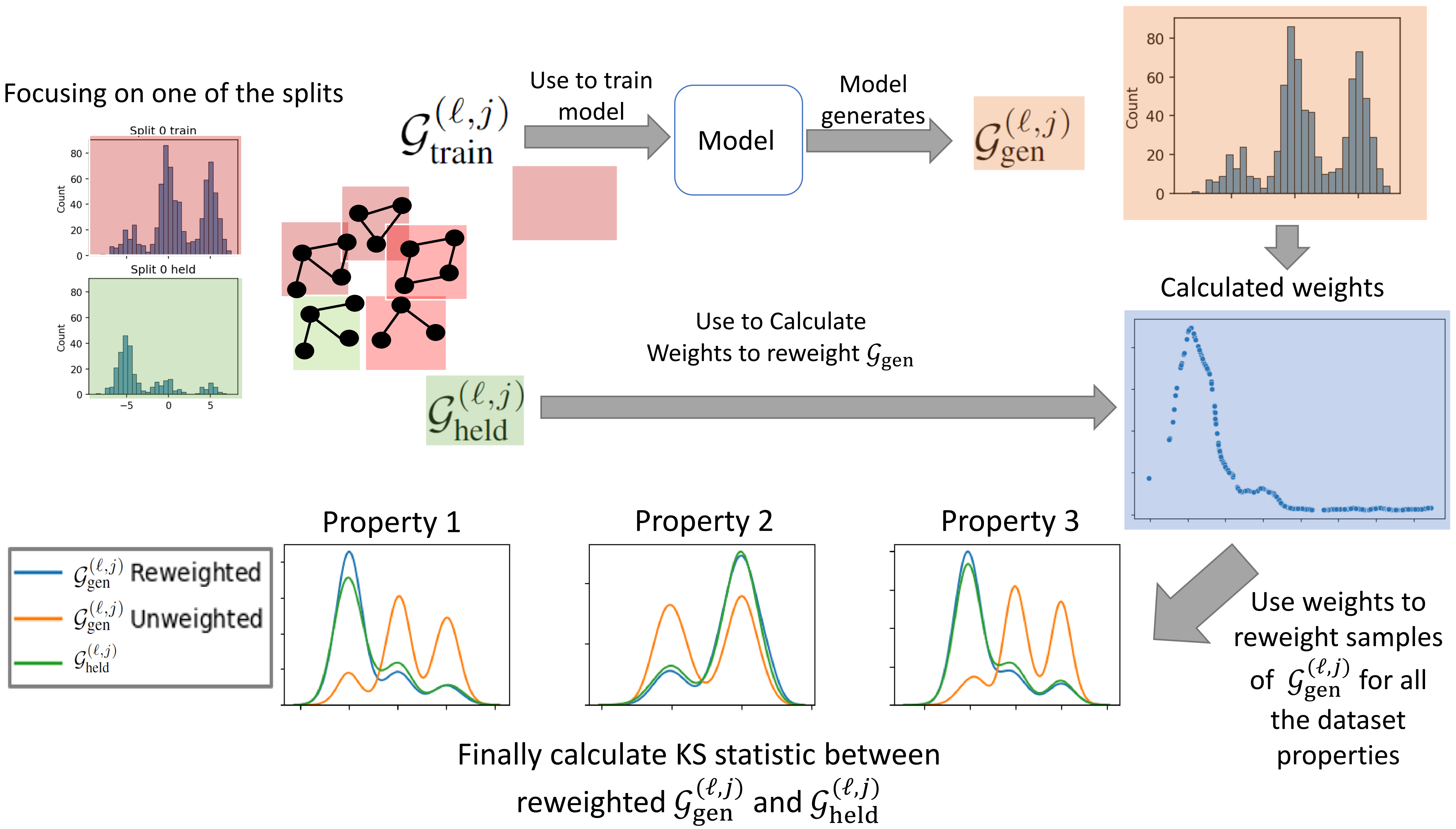


Dataset of Graphs
$\mathcal{G} = \{G_i\}_{i=1}^{|\mathcal{G}|}$
1. $\forall G_i$ Compute Graph Properties Z_ℓ
2. Project to unit interval via CDF
$U_\ell = F_\ell(Z_\ell)$
$Z_\ell = h_\ell(G)$
3. Define split distributions on [0,1]
4. Compute split prob *given* U_ℓ
$p(S_\ell|U_\ell)$
$U_\ell = F(Z_\ell)$
5. Sample $S_{i,\ell} \sim p(S_{i,\ell}|U_{i,\ell})$ for each graph
Split 1
Split 2
Split 3
Split 4
Held-Out $S_{i,\ell} = j$
Train $S_{i,\ell} \neq j$
Z_2
A. Focusing on one of the splits (j =1) and one split property ($\ell = 2$)
B. Use $\mathcal{G}_{\text{held}}^{(\ell,j)}$ to Calculate Weights to reweight $\mathcal{G}_{\text{gen}}^{(\ell,j)}$
Use to train model
$\mathcal{G}_{\text{train}}^{(\ell,j)}$
$\mathcal{G}_{\text{held}}^{(\ell,j)}$
Generative Model
Model generates
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$
C. Use weights $W^{(\ell,j)}$ to reweight sample properties of $\mathcal{G}_{\text{gen}}^{(\ell,j)}$ for all the dataset properties
$W^{(\ell=2,j=1)}$
$\ell = 2, \ell' = 1$
$\ell = 2, \ell' = 2$
Density
Count
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$ Reweighted
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$ Unweighted
$\mathcal{G}_{\text{held}}^{(\ell,j)}$
$\bar{Z}_1$
$\bar{Z}_2$

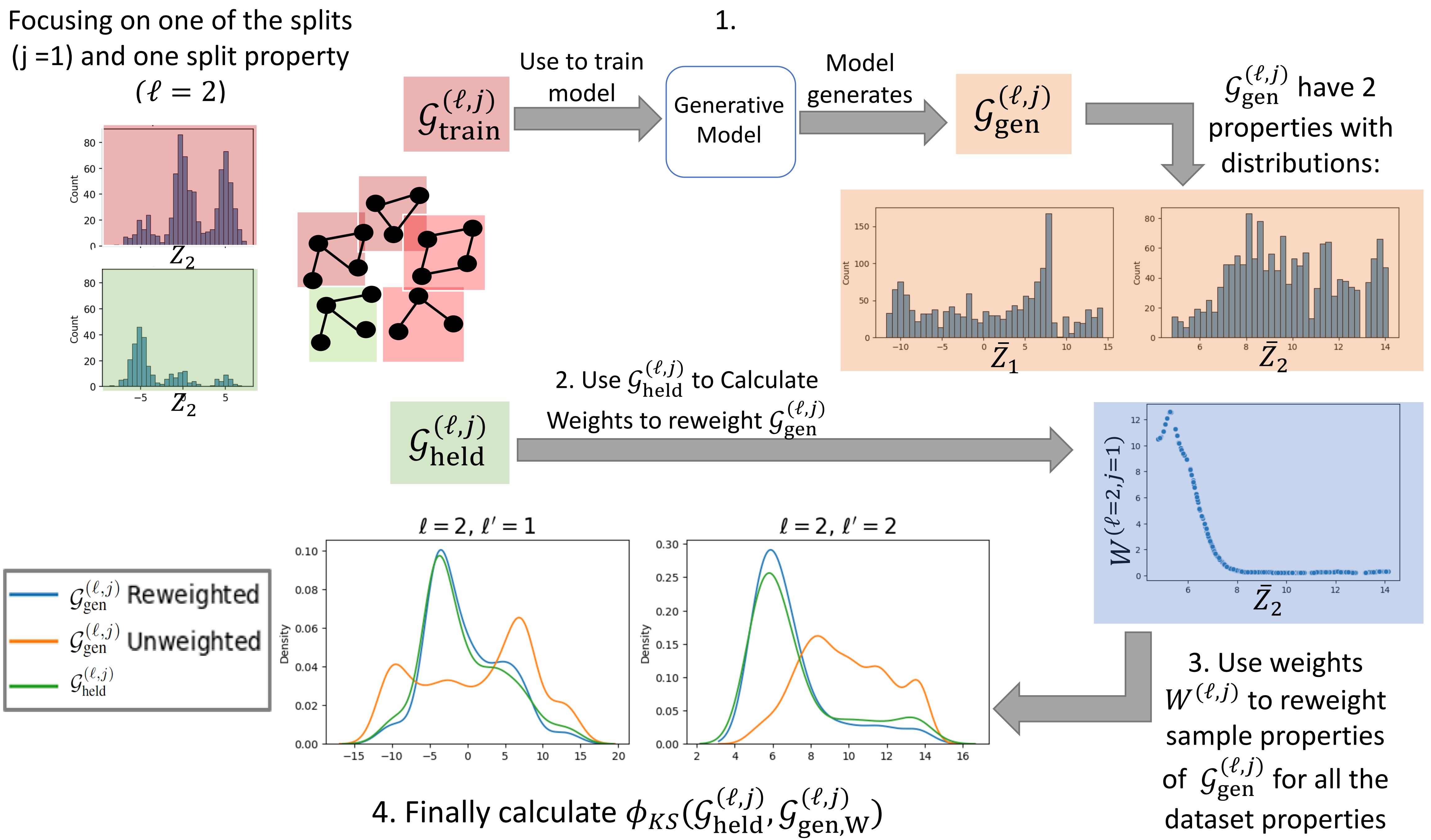


1) Reweight to bring the distribution of held data and generated data for split property ℓ to become closer

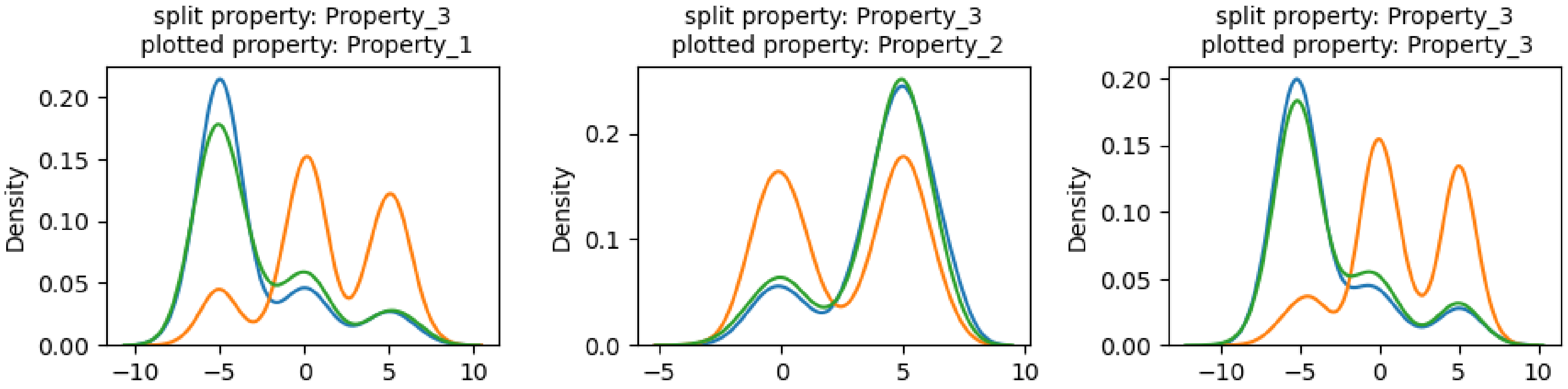
2) Calculate ϕ_{ks} on all remaining properties $\ell' \neq \ell$

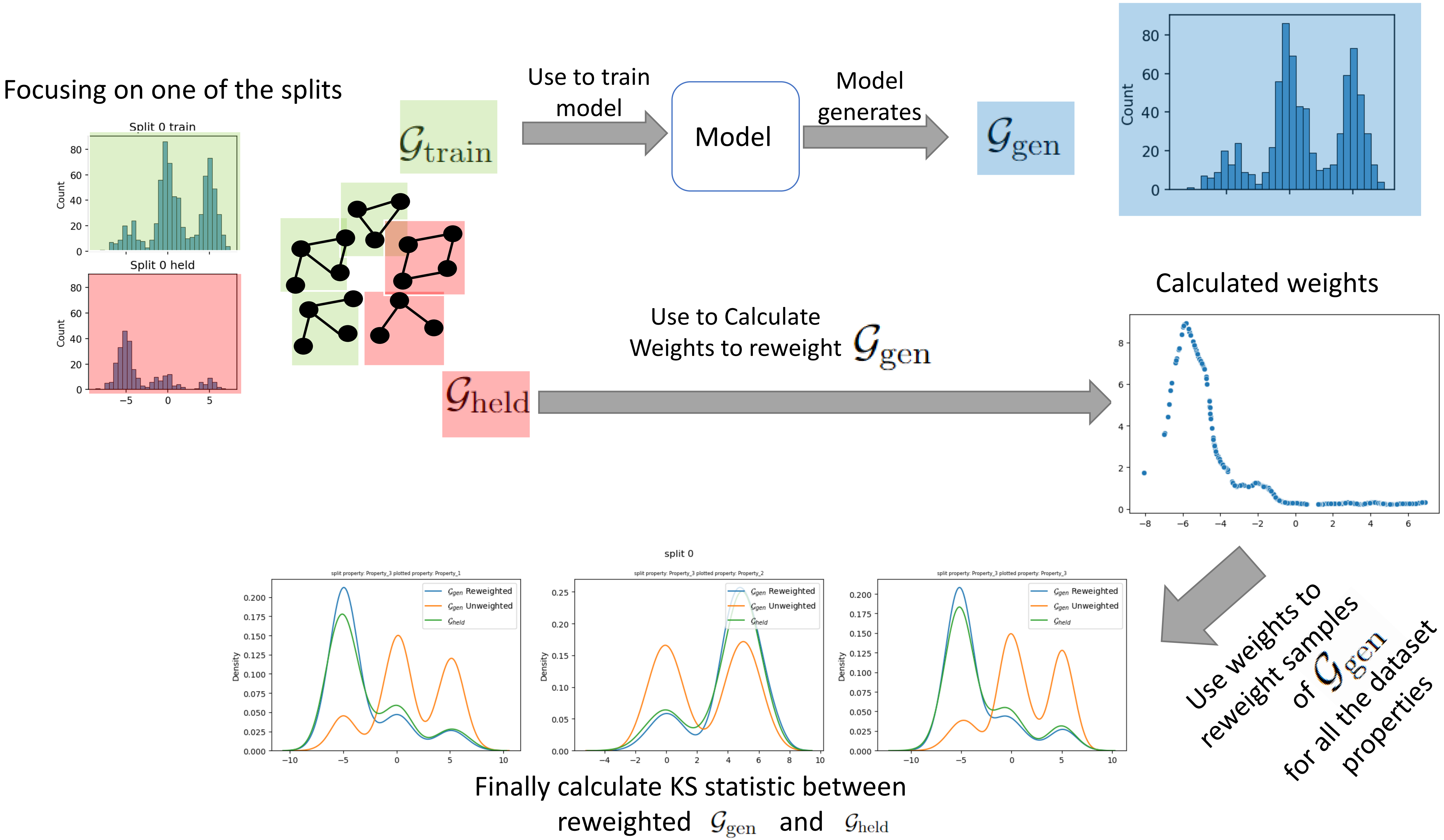


Focusing on one of the splits
Split 0 train
Count
Split 0 held
Count
$\mathcal{G}_{\text{train}}^{(\ell,j)}$
Use to train model
Model
Model generates
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$
Count
Calculated weights
Use to Calculate Weights to reweight $\mathcal{G}_{\text{gen}}$
$\mathcal{G}_{\text{held}}^{(\ell,j)}$
Property 1
Property 2
Property 3
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$ Reweighted
$\mathcal{G}_{\text{gen}}^{(\ell,j)}$ Unweighted
$\mathcal{G}_{\text{held}}^{(\ell,j)}$
Use weights to reweight samples of $\mathcal{G}_{\text{gen}}^{(\ell,j)}$ for all the dataset properties
Finally calculate KS statistic between reweighted $\mathcal{G}_{\text{gen}}^{(\ell,j)}$ and $\mathcal{G}_{\text{held}}^{(\ell,j)}$

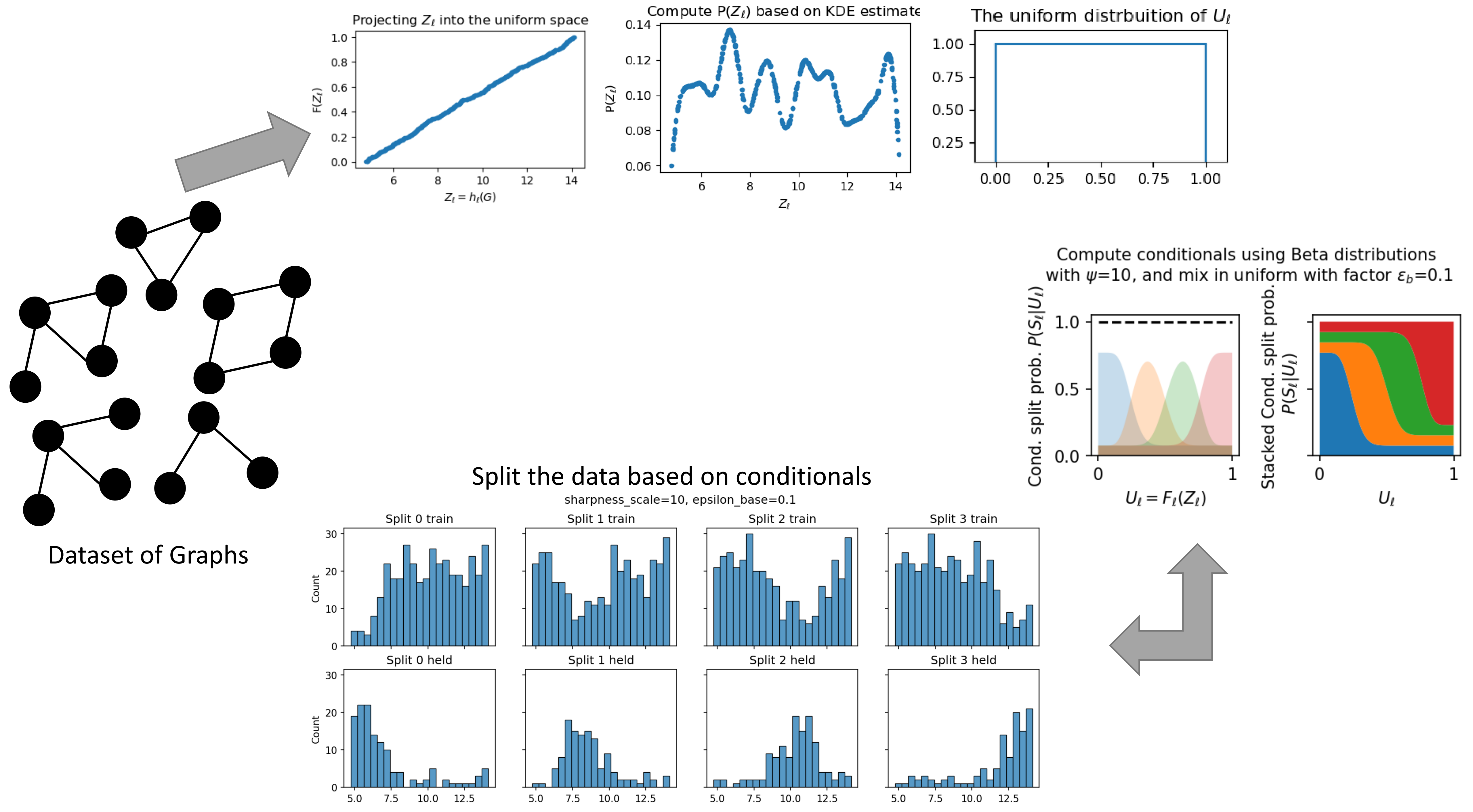


Focusing on one of the splits (j =1) and one split property ($\ell = 2$)
1.
$\mathcal{G}^{(\ell,j)}_{\text{train}}$
Use to train model
Generative Model
Model generates
$\mathcal{G}^{(\ell,j)}_{\text{gen}}$
$\mathcal{G}^{(\ell,j)}_{\text{gen}}$ have 2 properties with distributions:
$\bar{Z}_1$
$\bar{Z}_2$
Z_2
Count
2. Use $\mathcal{G}^{(\ell,j)}_{\text{held}}$ to Calculate Weights to reweight $\mathcal{G}^{(\ell,j)}_{\text{gen}}$
$\mathcal{G}^{(\ell,j)}_{\text{held}}$
$W^{(\ell=2,j=1)}$
3. Use weights $W^{(\ell,j)}$ to reweight sample properties of $\mathcal{G}^{(\ell,j)}_{\text{gen}}$ for all the dataset properties
$\ell = 2, \ell' = 1$
$\ell = 2, \ell' = 2$
Density
$\mathcal{G}^{(\ell,j)}_{\text{gen}}$ Reweighted
$\mathcal{G}^{(\ell,j)}_{\text{gen}}$ Unweighted
$\mathcal{G}^{(\ell,j)}_{\text{held}}$
4. Finally calculate $\phi_{KS}(\mathcal{G}^{(\ell,j)}_{\text{held}}, \mathcal{G}^{(\ell,j)}_{\text{gen,W}})$


split property: Property_3
plotted property: Property_1
Density
0.00
0.05
0.10
0.15
0.20
−10
−5
0
5
10
split property: Property_3
plotted property: Property_2
Density
0.0
0.1
0.2
−5
0
5
10
split property: Property_3
plotted property: Property_3
Density
0.00
0.05
0.10
0.15
0.20
−10
−5
0
5
10



Focusing on one of the splits
Split 0 train
Count
Split 0 held
Count
$\mathcal{G}_{\text{train}}$
Use to train model
Model
Model generates
$\mathcal{G}_{\text{gen}}$
Count
$\mathcal{G}_{\text{held}}$
Use to Calculate Weights to reweight $\mathcal{G}_{\text{gen}}$
Calculated weights
Use weights to reweight samples of $\mathcal{G}_{\text{gen}}$ for all the dataset properties
split 0
split property: Property_3 plotted property: Property_1
split property: Property_3 plotted property: Property_2
split property: Property_3 plotted property: Property_3
$\mathcal{G}_{gen}$ Reweighted
$\mathcal{G}_{gen}$ Unweighted
$\mathcal{G}_{held}$
Density
Finally calculate KS statistic between reweighted $\mathcal{G}_{\text{gen}}$ and $\mathcal{G}_{\text{held}}$



Dataset of Graphs
Projecting Z_ℓ into the uniform space
$F(Z_\ell)$
$Z_\ell = h_\ell(G)$
Compute $P(Z_\ell)$ based on KDE estimate
$P(Z_\ell)$
Z_ℓ
The uniform distrbuition of U_ℓ
Compute conditionals using Beta distributions
with ψ=10, and mix in uniform with factor ε_b=0.1
Cond. split prob. $P(S_\ell|U_\ell)$
$U_\ell = F_\ell(Z_\ell)$
Stacked Cond. split prob.
$P(S_\ell|U_\ell)$
U_ℓ
Split the data based on conditionals
sharpness_scale=10, epsilon_base=0.1
Split 0 train
Split 1 train
Split 2 train
Split 3 train
Split 0 held
Split 1 held
Split 2 held
Split 3 held
Count