Learning Optimal and Interpretable Summaries of Galaxy Catalogs with SBI
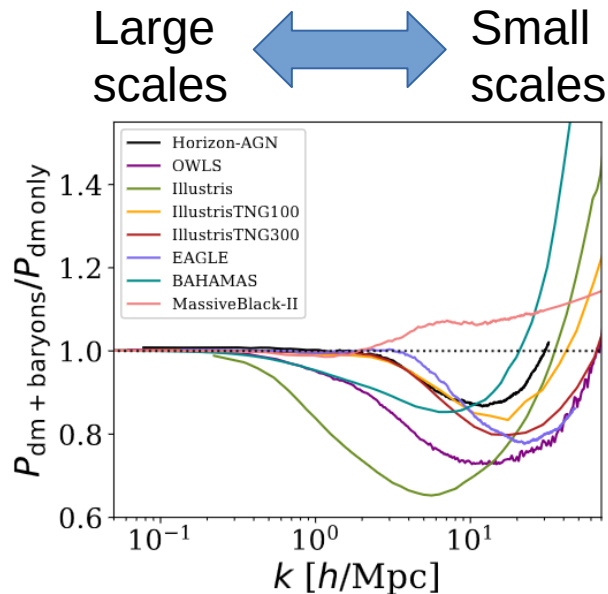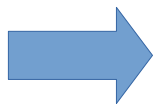
with Sven Krippendorf, Jochen Weller, and Klaus Dolag

- Numerical simulations for non-linear structures in the Universe

- Uncertainties in small-scale Physics yield significant disagreement in known summaries

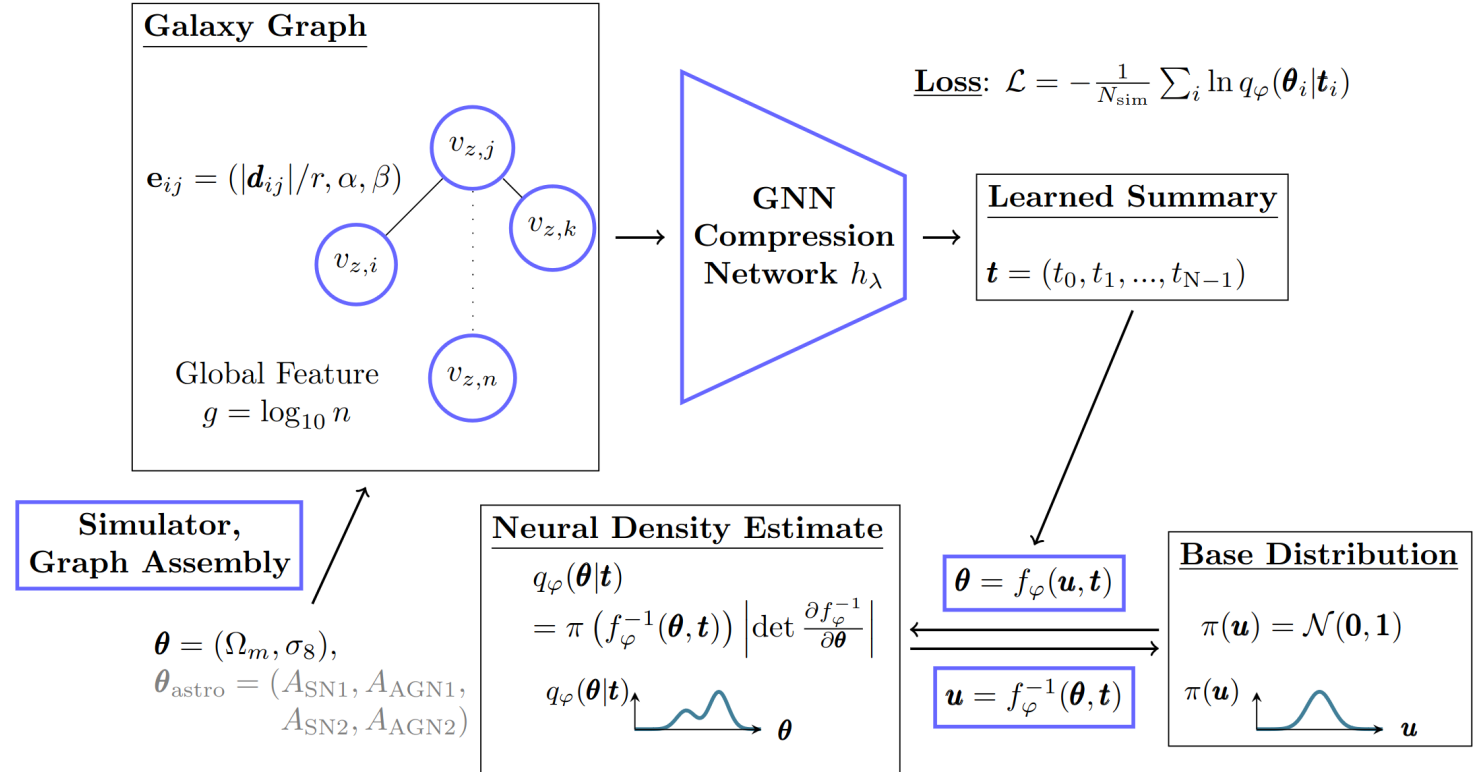- How is this reflected beyond known summary statistics?

➡ Find and analyze optimized summary statistics

Large scales ⟷ Small scales

Chisari+ 19

Summary Extraction with Simulation Based Inference

- Galaxy catalogs are embedded as graphs
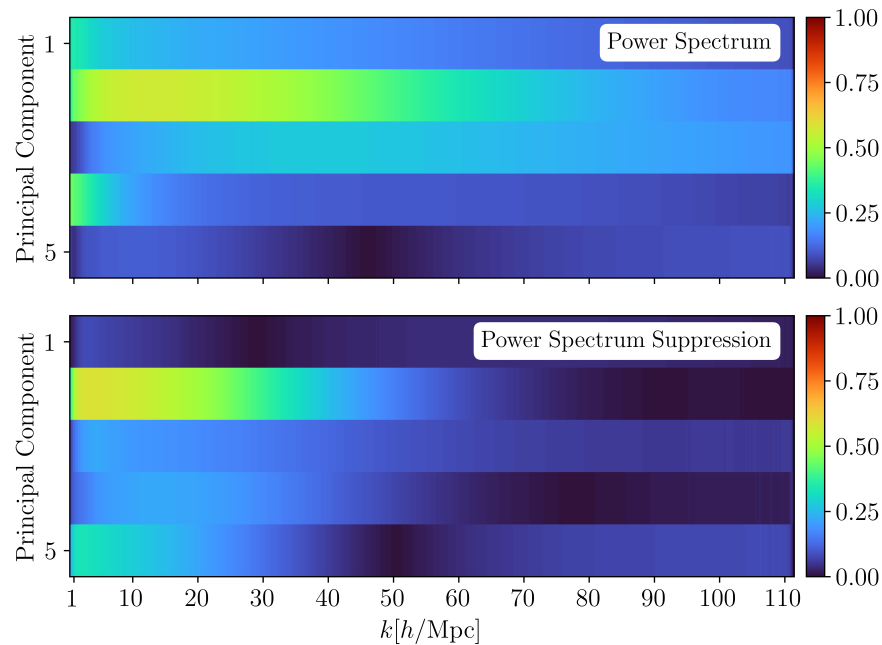
- SBI machine learns optimized summaries of these graphs

Galaxy Graph

$\mathbf{e}_{ij} = (|\boldsymbol{d}_{ij}|/r, \alpha, \beta)$

$v_{z,j}$

$v_{z,k}$

$v_{z,i}$

$v_{z,n}$

Global Feature
$g = \log_{10} n$

GNN Compression Network h_λ

Loss: $\mathcal{L} = -\frac{1}{N_{\text{sim}}} \sum_i \ln q_\varphi(\boldsymbol{\theta}_i | \boldsymbol{t}_i)$

Learned Summary

$\boldsymbol{t} = (t_0, t_1, ..., t_{N-1})$

Simulator, Graph Assembly

$\boldsymbol{\theta} = (\Omega_m, \sigma_8),$
$\boldsymbol{\theta}_{\text{astro}} = (A_{\text{SN1}}, A_{\text{AGN1}}, A_{\text{SN2}}, A_{\text{AGN2}})$

Neural Density Estimate

$q_\varphi(\boldsymbol{\theta}|\boldsymbol{t})$
$= \pi\left(f_\varphi^{-1}(\boldsymbol{\theta}, \boldsymbol{t})\right) \left|\det \frac{\partial f_\varphi^{-1}}{\partial \boldsymbol{\theta}}\right|$

$q_\varphi(\boldsymbol{\theta}|\boldsymbol{t})$ $\boldsymbol{\theta}$

$\boldsymbol{\theta} = f_\varphi(\boldsymbol{u}, \boldsymbol{t})$

$\boldsymbol{u} = f_\varphi^{-1}(\boldsymbol{\theta}, \boldsymbol{t})$

Base Distribution

$\pi(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{1})$

$\pi(\boldsymbol{u})$ $\boldsymbol{u}$

What Can We Learn from Optimized Summaries?
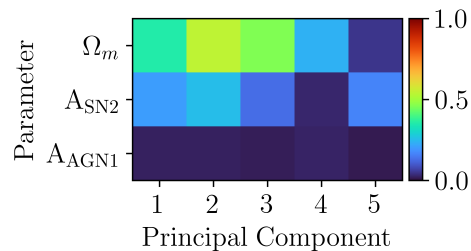
Important Parameters
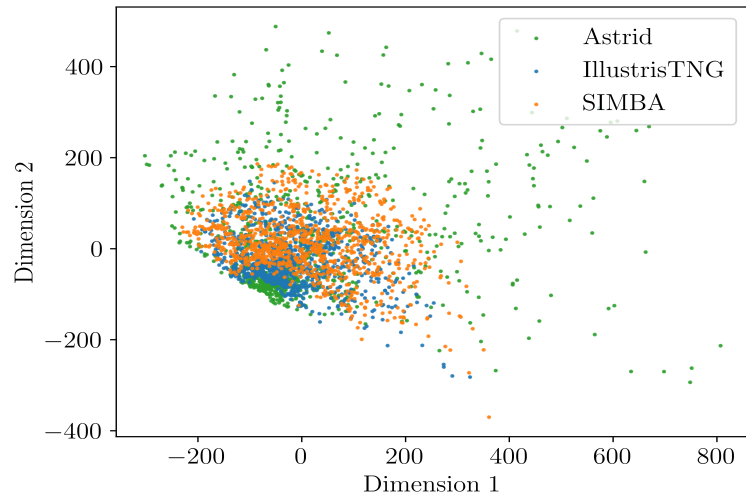
Comparison of Simulation Models

Scales