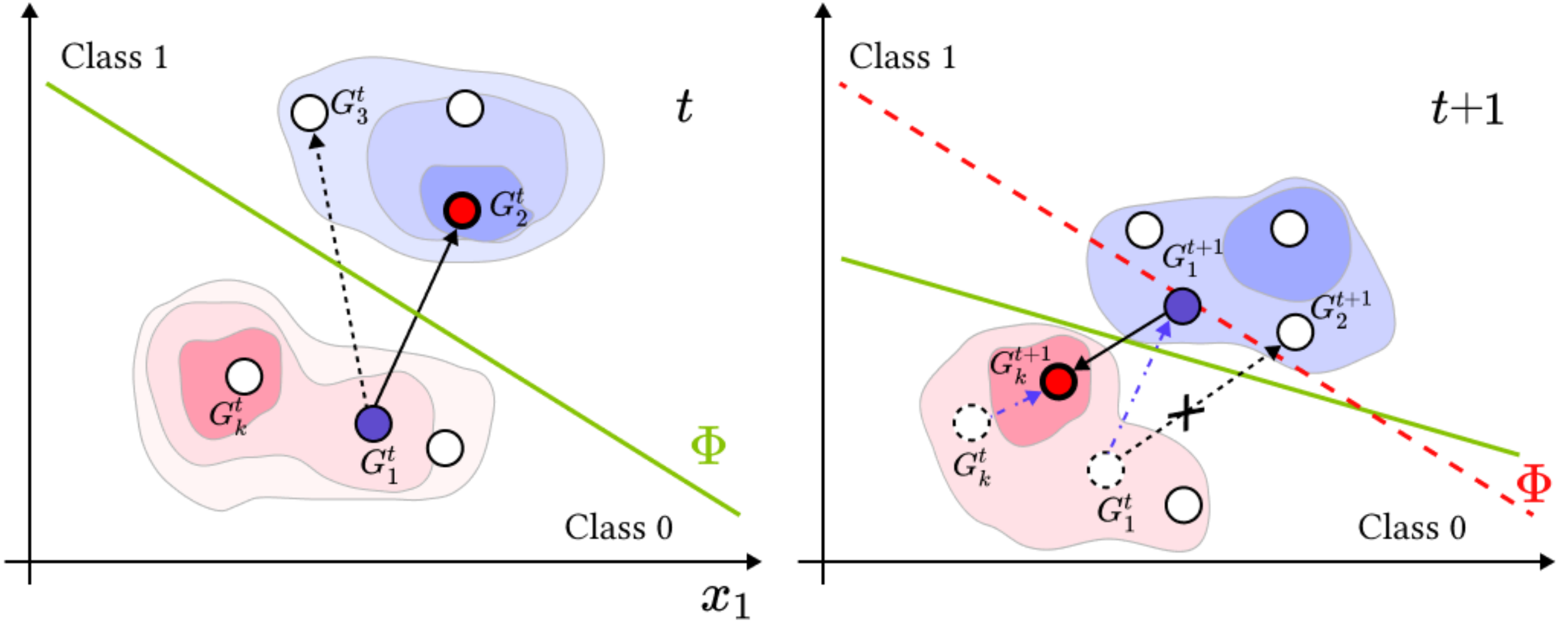
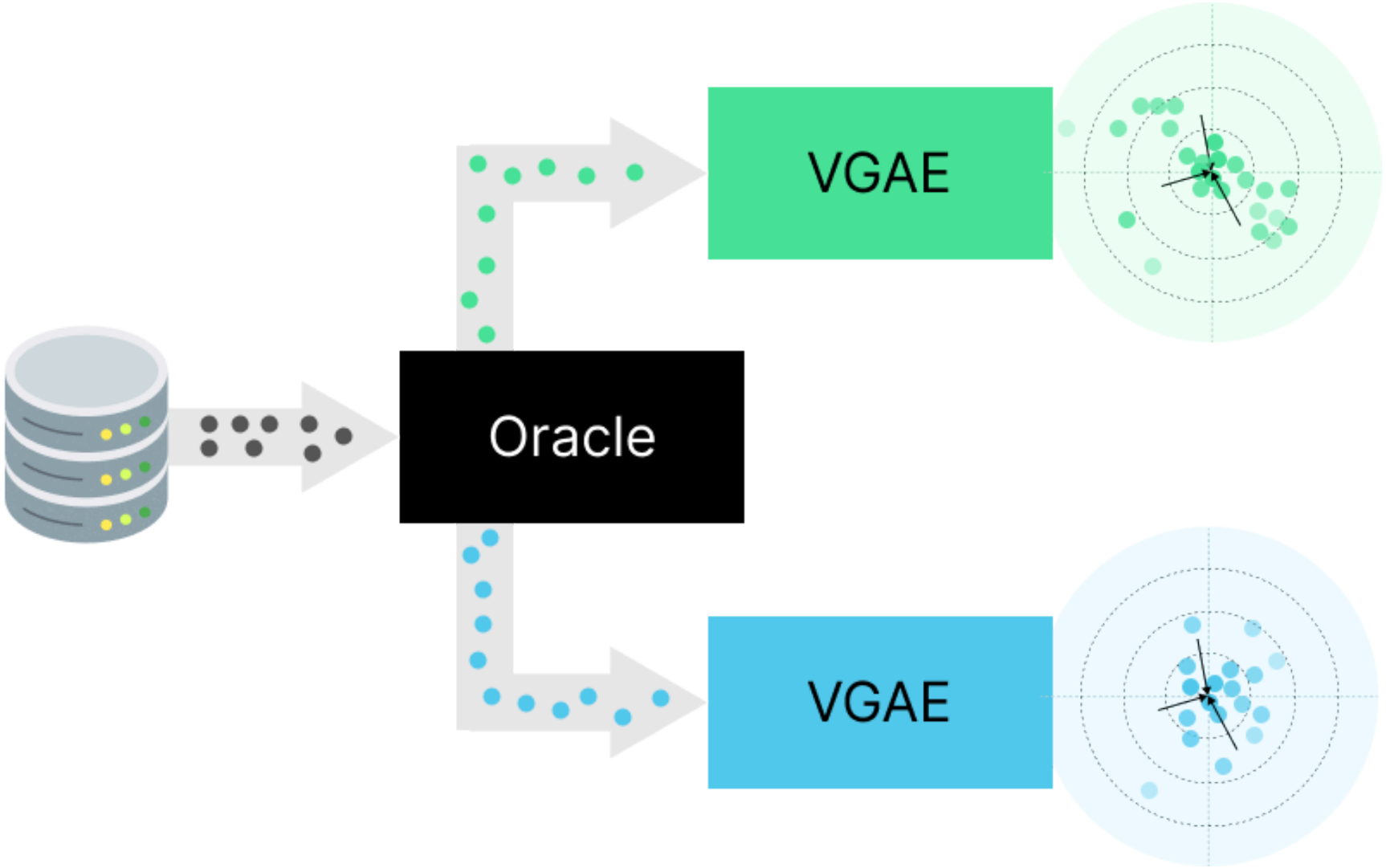
Unifying Evolution, Explanation, and Discernment: A Generative Approach for Dynamic Graph Counterfactuals

Bardh Prenkaj
Post-doc in AI
bardh.prenkaj@tum.de



Correct oracle decision boundary
Old oracle decision boundary
Potential counterfactual association
Correct counterfactual association
Old counterfactual association
Graph movement from t to $t+1$
Graph at current time
Graph at previous time

Class Representation Experts



Variational Autoencoders

We consider the following generative model where the graphs G are generated from factored latent representation
and the true class label y

$$p\left(G|y\right) = \int_{\mathbf{z}\in\mathcal{Z}} p\left(G|\mathbf{z},y\right) p\left(\mathbf{z}|y\right) dz$$

Variational Autoencoders (Decoder)

To represent $p(G|y)$, we use a single VGAE for each class $y \in \mathsf{Y}$, which is dependent on the class where each node has a latent vector and then define

$$p_{\theta_y}(G|\mathbf{z}, y) = p_{\theta_y}(\mathbf{A}, \mathbf{X}|\mathbf{z}, y)$$
$$= p_{\theta_y}(\mathbf{X}|\mathbf{A}, \mathbf{z}, y)\, p_{\theta_y}(\mathbf{A}|\mathbf{z}, y)$$

Variational Autoencoders (Encoder)

$$q_{\varphi_y}\left(\mathbf{z}|G,y\right)=\prod_{v_i} q_{\varphi_y}\left(\mathbf{z}_{v_i}|G,y\right)$$

$$q\left(z_{v_i}|G,y\right)=\mathcal{N}\left(z_{v_i}|\mu_{v_i},\gamma^2\mathbf{I}\right),$$

> 0 and fixed hyperparameter

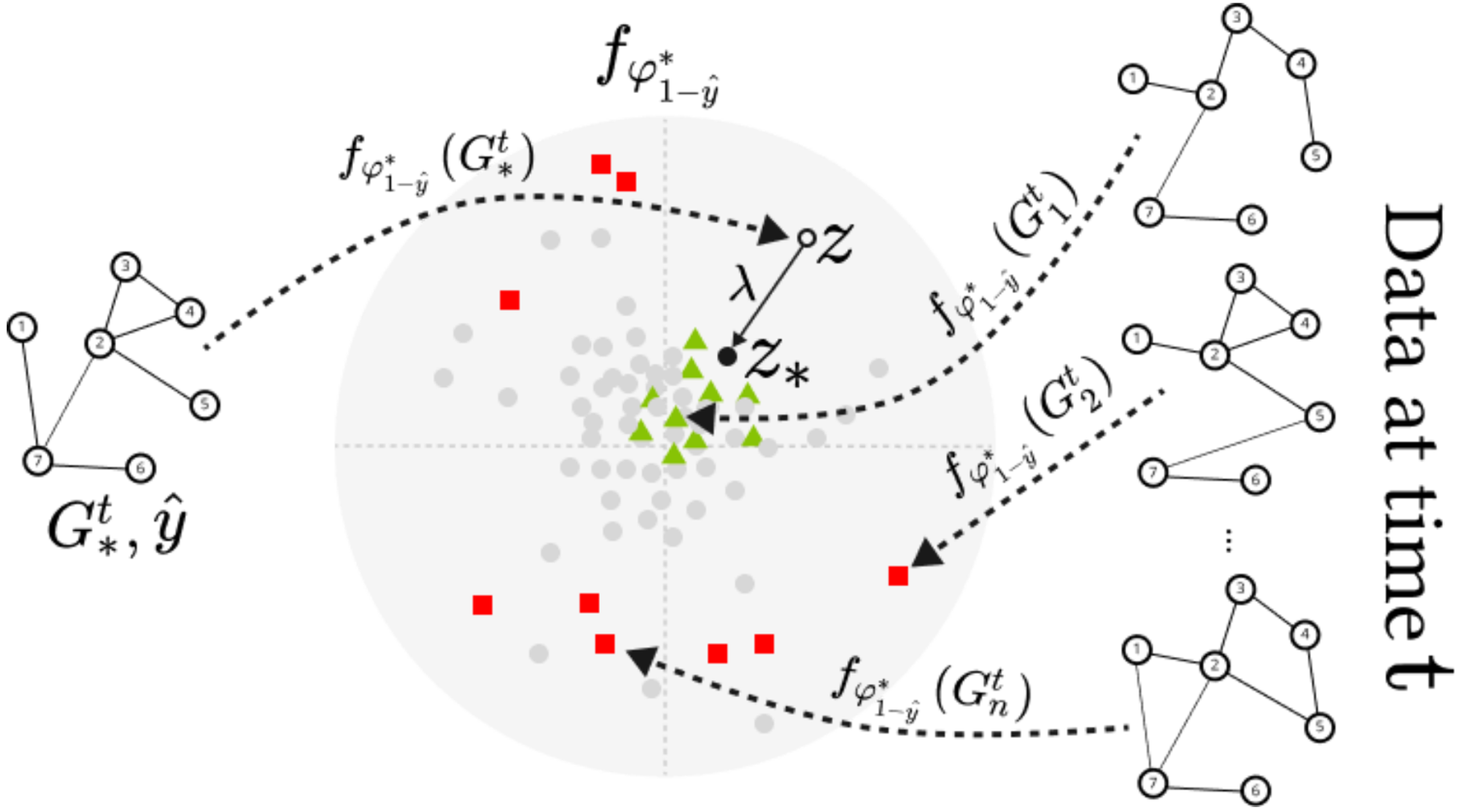
$$\mu=\left[\mu_{v_1},\ldots,\mu_{v_n}\right]=\mathrm{GCN}_{\varphi_y}\left(G\right)$$

Training GRACIE

$$-\mathrm{ELBO}_y\left(\theta_y, \varphi_y\right) = \mathcal{L}_{rec} + \mathcal{L}_{dist}$$

$$= \frac{1}{2}\left(\mathop{\mathbb{E}}_{q_{\varphi_y}(\mathbf{z}|G)}\left[\frac{||g_{\theta_y}(\mathbf{z}) - G||_2^2}{\sigma^2}\right] + ||f_{\varphi_y}(G)||_2^2\right)$$

Inference & Update

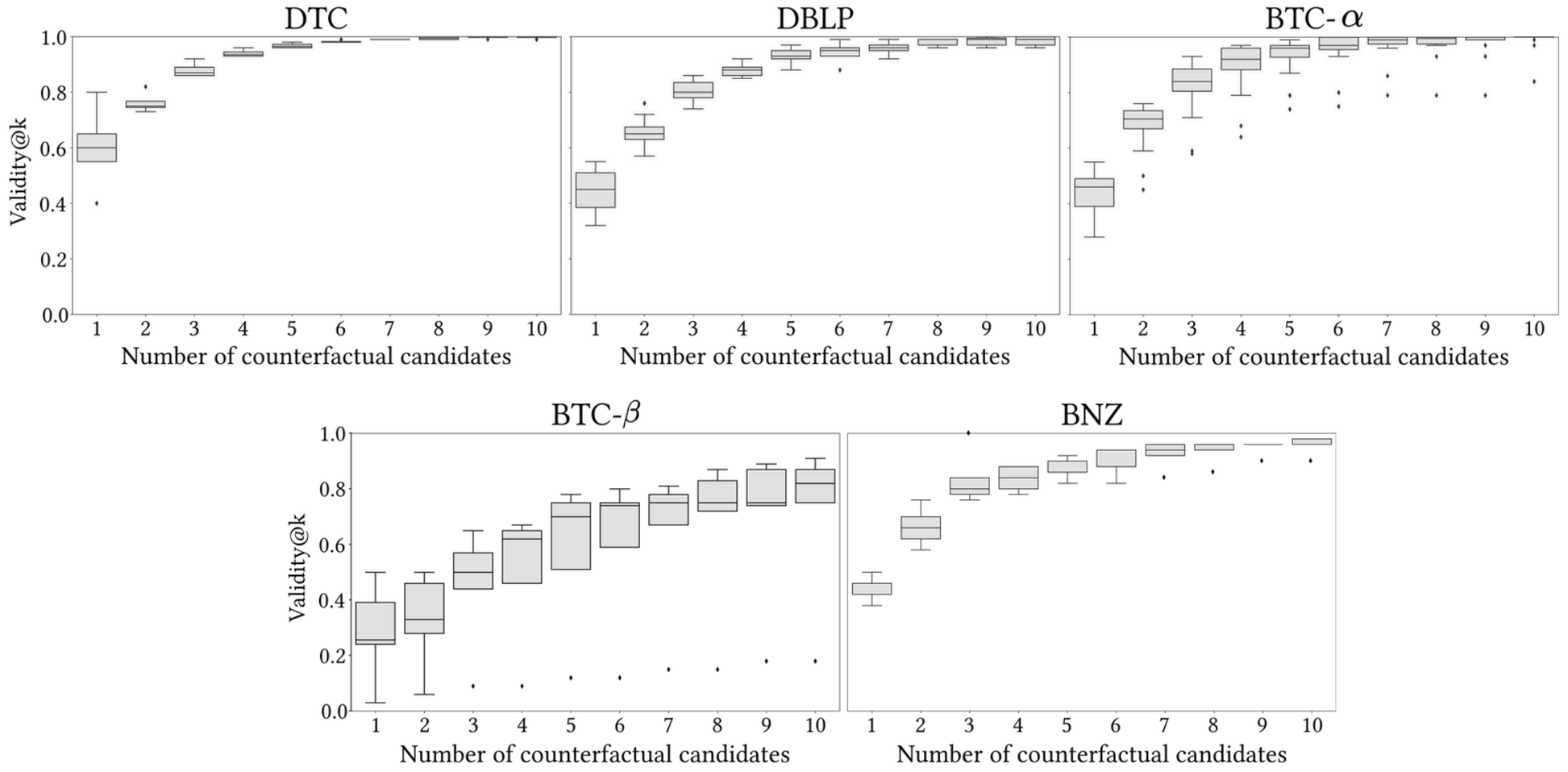


GRACIE is SoTA

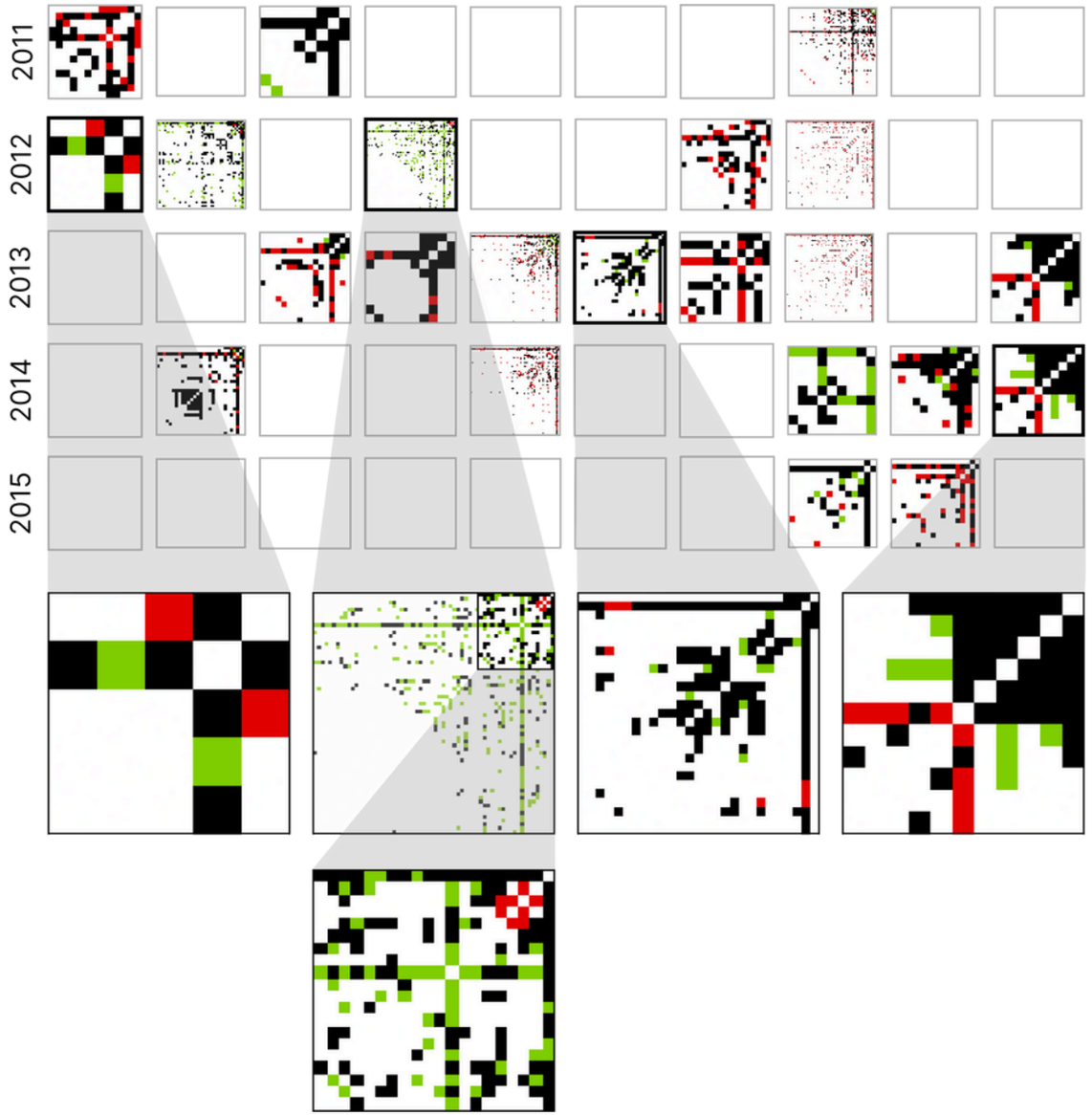
	DTC	DBLP	BTC-α	BTC-β	BNZ
BDDS	0.465	<u>0.381</u>	<u>0.360</u>†	0.235	0.136
MEG	0.250	0.209	×	<u>0.260</u>	0.120†
CLEAR	0.458	0.024	0.214	0.125	0.000
G-CounteRGAN	0.507	0.256	0.236	×	<u>0.404</u>
DyGRACE	<u>0.525</u>	0.307	0.232	0.000†	0.232
GRACIE	**0.600**	**0.442**	**0.440**	**0.284**	**0.441**

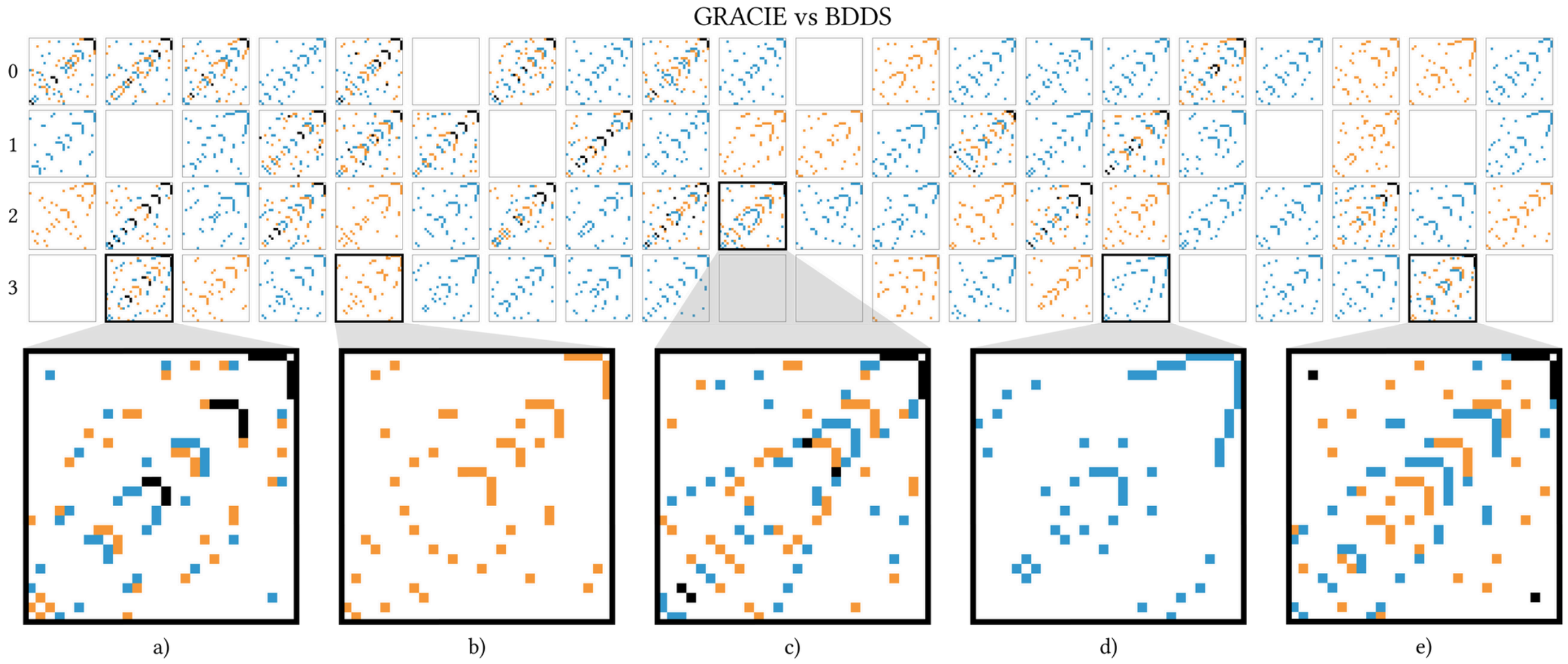
[a]The criterion of non-convergence is to fail to produce at least one counterfactual within 14 days of execution on an HPC SGE Cluster of 6 nodes with 360 cumulative cores, 1.2Tb of RAM, and two GPUs (i.e., one Nvidia A30 and one A100).

More sampling = more validity



Qualitative on BTC-β




GRACIE vs BDDS
0
1
2
3
a)
b)
c)
d)
e)

Thank you!

Come to my poster (#10 today at 13:00)