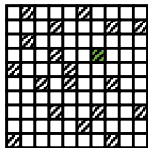
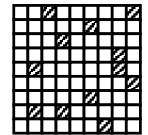


(A) Inputs**Partially observed GIs**

X - Target species GIs



- Missing
 - Observed



Y - Source species GIs

Cross-species Information $sim(\dots)$

Real valued gene-gene
similarity measure
(e.g. BLAST)

**Side Information
(Optional)**

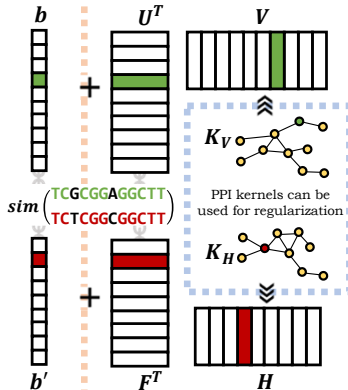
Side information (e.g.
PPI networks) in target
and source can be
incorporated via
kernelization

**(B) Kernelized cross-species
matrix factorization with bias (K-XSMF-b)**

Per-gene biases

Latent factors

Similarity measure
between source
and target genes is
used to *link* latent
factors and biases
and transfer
knowledge across
species

**(C) Parameter
learning**Learn latent factors
and biases

Loss is minimized with
gradient descent to learn
latent factors

$$\theta = \{U, V, F, H, b, b'\}$$

$$\underset{\theta}{\operatorname{argmin}} \mathcal{L}(\theta; X, Y, \lambda)$$

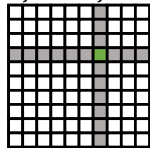
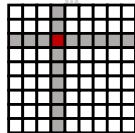


Automatic hyperparameter
optimization maximizes
imputation performance

(D) Outputs

Imputed GI scores

$$\hat{x}_{ij} = u_i^T v_j + b_i$$

**Target****Source**

$$\hat{y}_{ij} = f_i^T h_j + b'_i$$