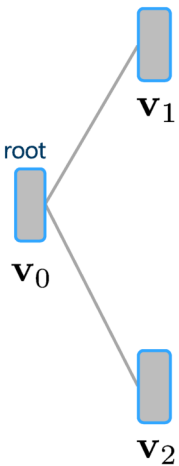
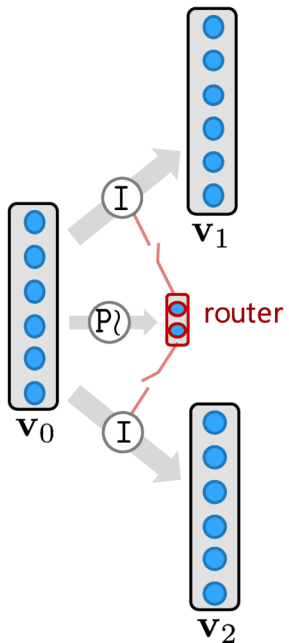


Old notation



New notation



Algebraically

Data transformation

$$\mathbf{v}_i = r_i \mathbf{I} \mathbf{v}_0 \quad i \in \{1, 2\}$$

Routing weights (with ReLU non-linearity)

$$\mathbf{r} = \sigma(\mathbf{P} \mathbf{v}_0) \quad r_i \in [0, 1]$$

The router can produce $N \geq 2$ outputs.