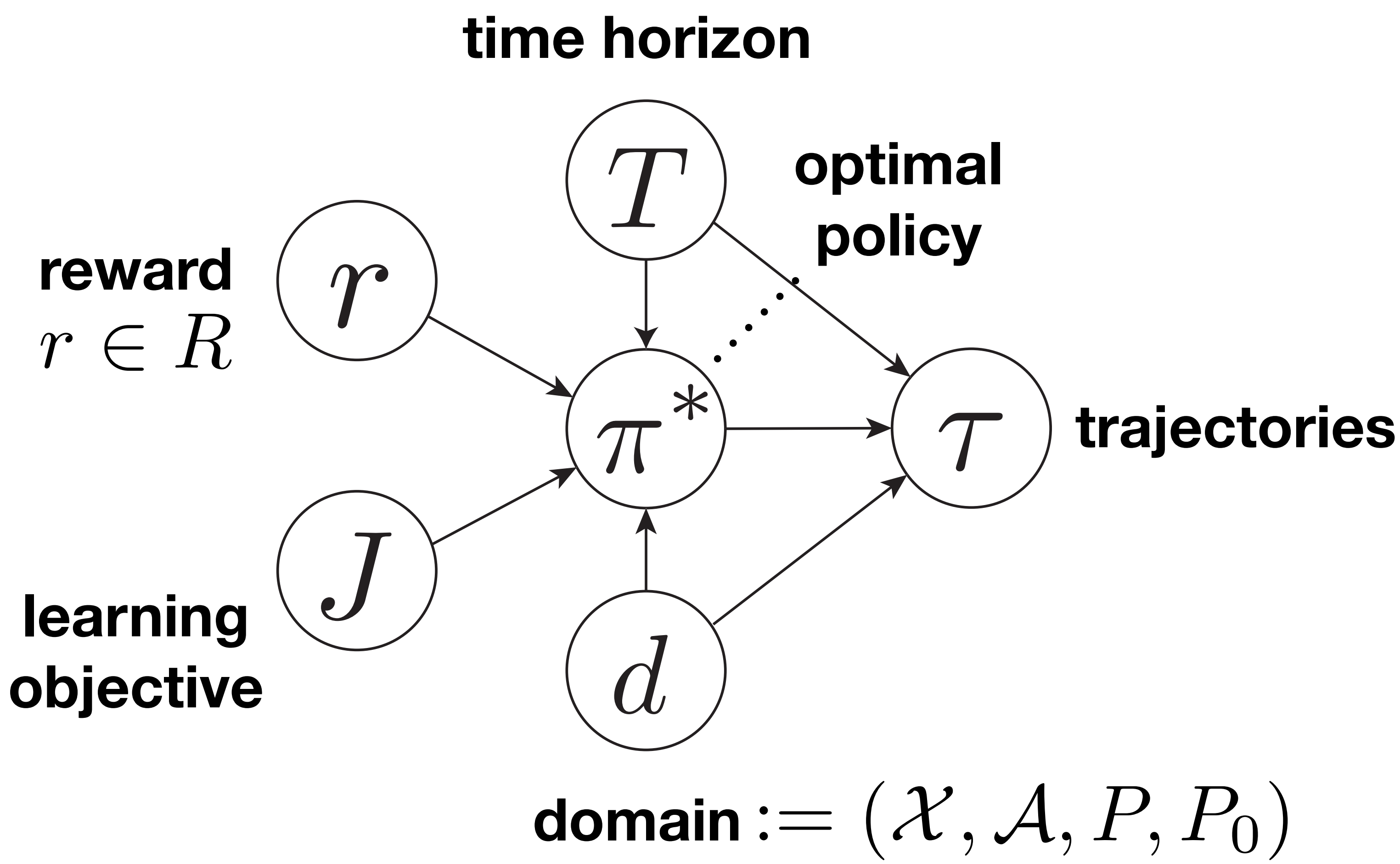
ICML 2021

# Reward Identification in IRL

Kuno Kim, Kirankumar Shiragur, Shivam Garg, Stefano Ermon

# Definition: MDP Models




**MDP Model**

$$\mathcal{P}_{MDP}[R; d, T, J] := \{p_r(\tau; \pi^*, d, T) : r \in R\}$$

**trajectory distribution induced by optimal policy for reward $r$ in domain $d$**

# Inverse Reinforcement Learning (IRL)

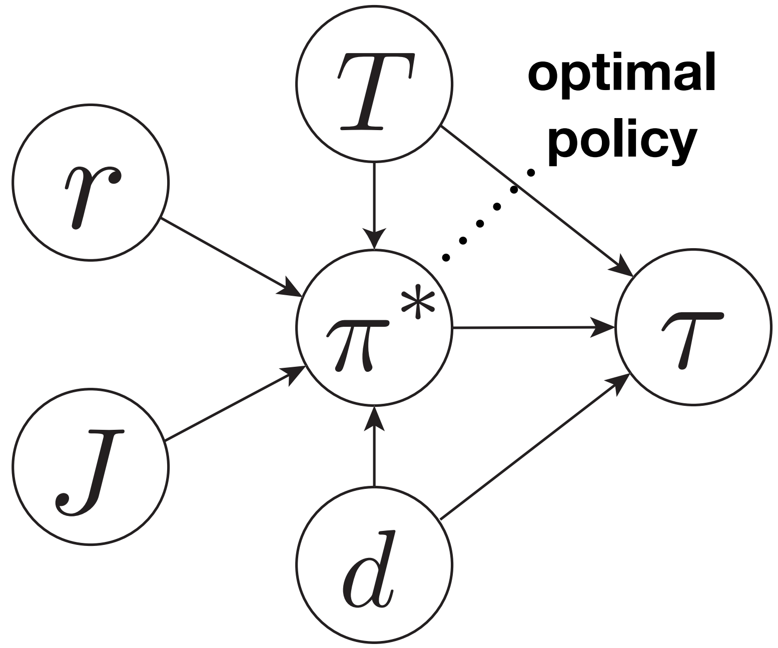
**RL**: Learn the optimal behavior for a given reward

$$r \rightarrow p_r(\tau; \pi^*, d, T)$$

**IRL**: Infer the underlying reward given optimal behavior

$$p_r(\tau; \pi^*, d, T) \rightarrow r$$

# The Reward Identification Problem




**MDP Model**

$$\mathcal{P}_{MDP}[R; d, T, J] := \{p_r(\tau; \pi^*, d, T) : r \in R\}$$

**IRL**: Infer the underlying reward given optimal behavior

$$p_r(\tau; \pi^*, d, T) \to r$$

**When is it possible to identify a reasonable equivalence class of rewards given knowledge of $(p_r, d, T, J)$?**

# Definition: Identifiabililty

**Definition 1.** (Identifiability) *An MDP model* $\mathcal{P}_{\text{MDP}}[R; d, T, J] = \{p_r(\tau; d, T, J) \mid r \in R\}$ *is **identifiable** up to an equivalence relation* $\cong$ *if for all* $r, \hat{r} \in R$,

$$r \cong \hat{r} \iff p_r = p_{\hat{r}}$$

# Definition: Weak Identifiability

**Definition 3.** (Weak Identifiability) *An MDP model* $\mathcal{P}_{\mathrm{MDP}}[R; d, T, J]$ *is* ***weakly identifiable*** *if it is identifiable up to* $\cong_\tau$*, i.e trajectory equivalence.*

**Trajectory Equivalence**

$$r \cong_\tau \hat{r} \iff \forall x \in \mathcal{X}^0, \tau', \tau'' \in \Omega[x, d, T],$$
$$\hat{r}(\tau') - r(\tau') = \hat{r}(\tau'') - r(\tau'')$$

# Definition: Strong Identifiability

**Definition 4.** (Strong Identifiability) *An MDP model is **strongly identifiable** if it is identifiable up to rewards shifted by a constant, i.e* $\cong_{x,a}$.

State-action Equivalence

$$r \cong_{x,a} \hat{r} \iff \forall (x', a'), (x'', a'') \in \mathcal{X} \times \mathcal{A},$$
$$\hat{r}(x', a') - r(x', a') = \hat{r}(x'', a'') - r(x'', a'')$$

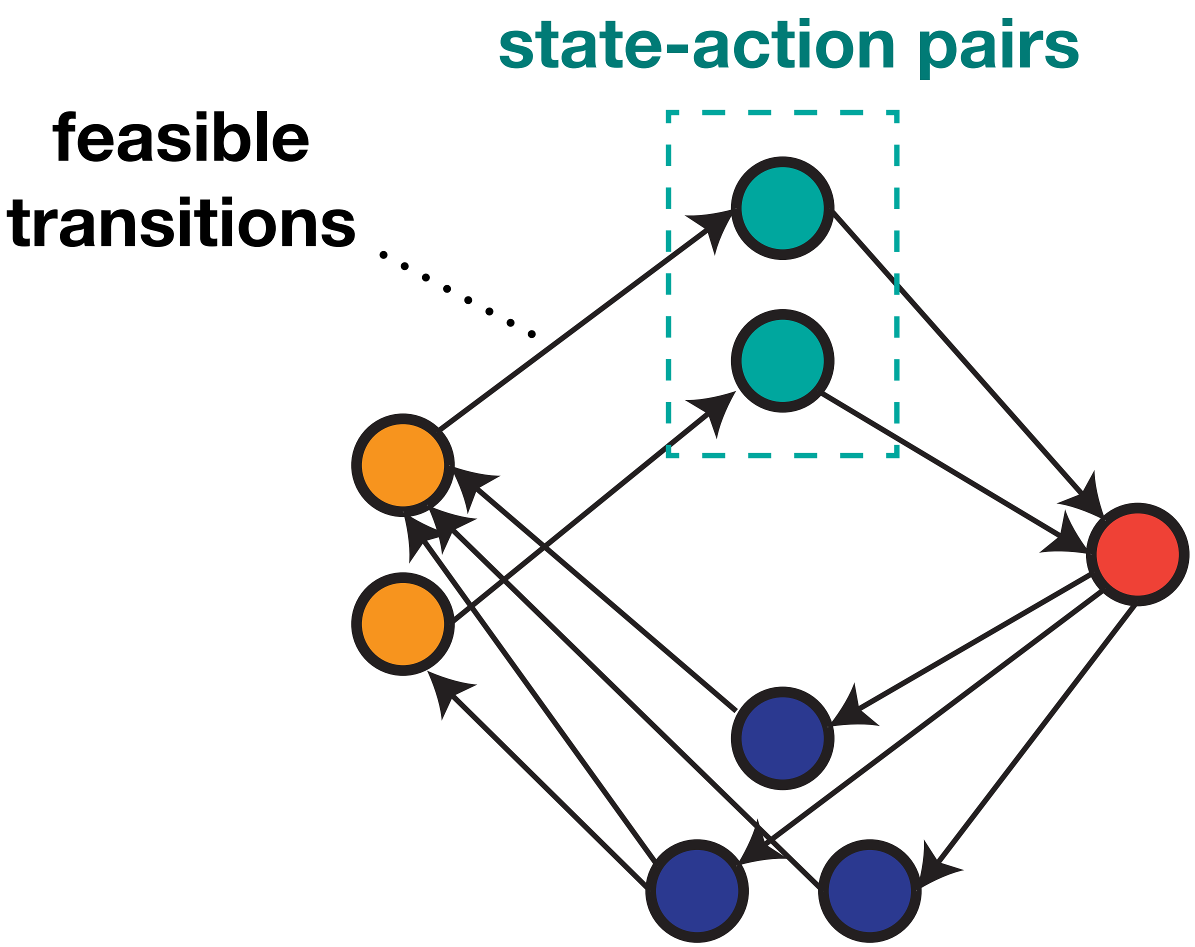
**Proposition 1.** *A proper MDP model is strongly identifiable only if it is weakly identifiable*

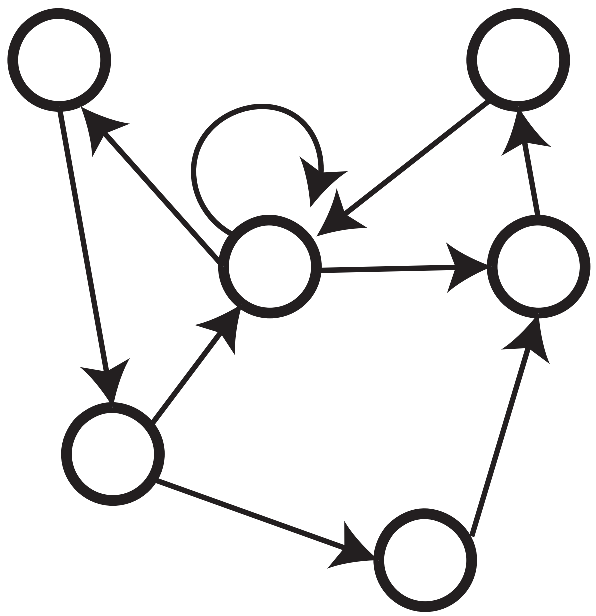
# Theorem: Weak Identification

**Theorem 1.** *Let* $\mathcal{P}_{\text{MDP}}[R; d, T, J_{\text{MaxEnt}}]$ *be a MaxEnt MDP model and* $R \subseteq \{r \mid r : \mathcal{X} \times \mathcal{A} \to \mathbb{R}\}$ *be any set of rewards. Then, for all domains* $d := (\mathcal{X}, \mathcal{A}, P, P_0, \gamma)$ *consisting of deterministic transition dynamics, i.e* $\forall(x, a), |\text{supp}(P(\cdot|x, a))| = 1$*, a deterministic initial state, i.e* $|\text{supp}(P_0)| = 1$*, and* $T \geq 0$*,* $\mathcal{P}_{\text{MDP}}[R; d, T, J_{\text{MaxEnt}}]$ *is weakly identifiable.*

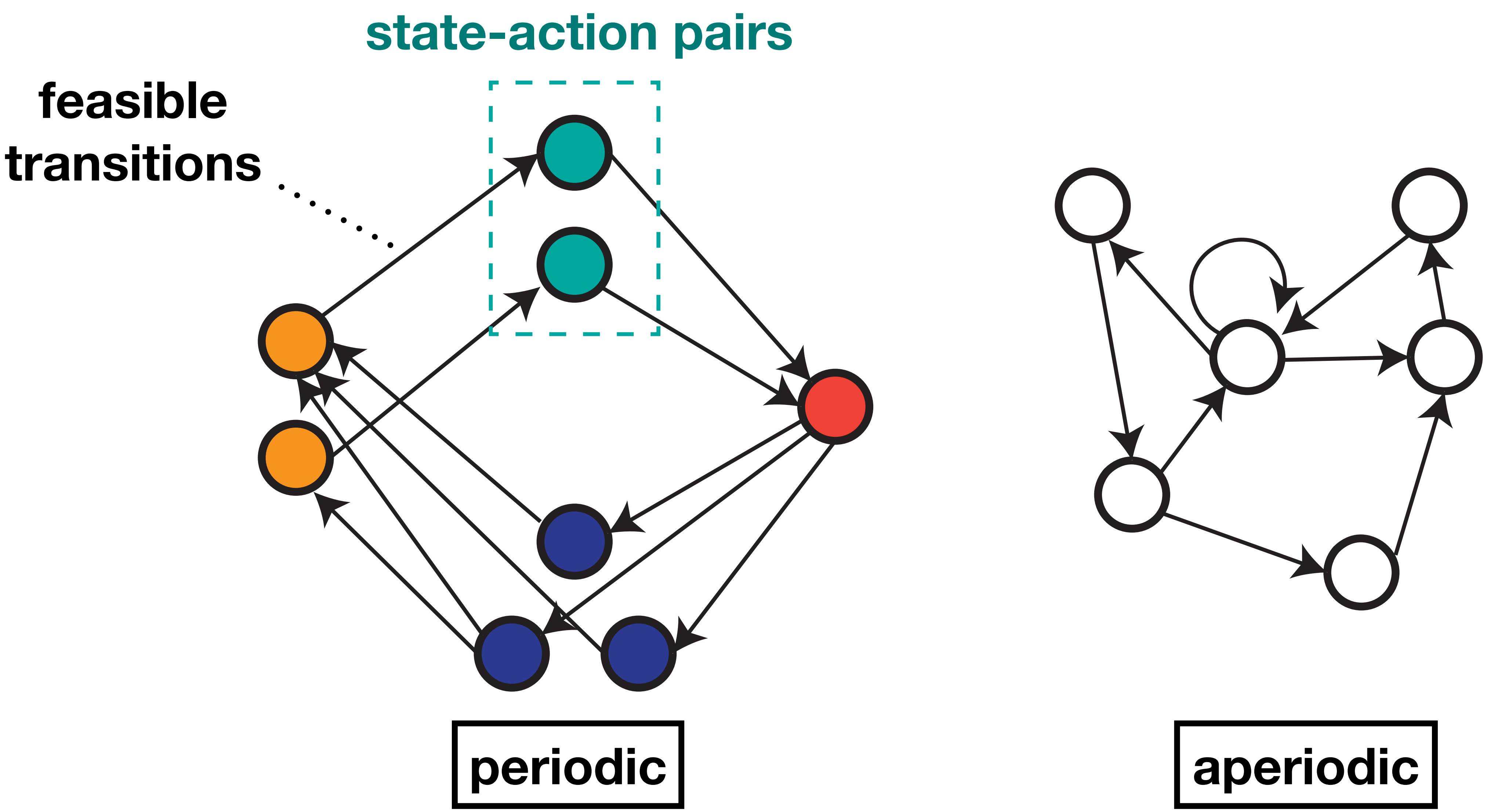
**TLDR: Deterministic, MaxEnt MDP models are weakly identifiable regardless of the domain properties**

# Domain Graphs



# Domain Graphs




A domain graph is aperiodic if the GCD of the periods of all cycles in the graph is 1, and periodic otherwise

# Theorem: Strong Identification

**Corollary 2.** (Strong Identification Condition) *For all* $(d, r, T, J)$ *such that* $\mathcal{P}_{\mathrm{MDP}}[R; d, T, J]$ *is a proper MDP model and* $G_{\mathrm{d}}$ *is strongly connected,*

- *(Sufficiency)* $\mathcal{P}_{\mathrm{MDP}}[R; d, T, J]$ *is weakly identifiable,* $G_d$ *aperiodic* $\Rightarrow \exists T_0 \geq 0$ *such that* $\forall T \geq T_0, \mathcal{P}_{\mathrm{MDP}}[R; d, T, J]$ *is strongly identifiable*
- *(Necessity)* $\mathcal{P}_{\mathrm{MDP}}[R; d, T, J]$ *is strongly identifiable* $\Rightarrow$ $\mathcal{P}_{\mathrm{MDP}}[R; d, T, J]$ *is weakly identifiable,* $G_{\mathrm{d}}$ *is aperiodic.*

**TLDR: MDP Models with Aperiodic Domain Graphs are Strongly Identifiable**

# Algorithms for Identification Testing

**Algorithm 1** Strong Identifiability Test for MDP models with Strongly Connected Domain Graphs

**Procedure** `MDPIdTest` ($\mathcal{P}_{\text{MDP}}[R; d, T, J]$)

Construct a domain graph $G_d = (V_d, E_d, V_d^0)$ from $d$.

Set $gcd =$ `Period Finder`$(V_d, E_d)$ (Denardo, 1977)

**return** $gcd == 1$

# Acknowledgements


SONY

NSF

Office of Naval Research
ONR
Science & Technology

Air Force Office of Scientific Research
United States Air Force

TOYOTA
RESEARCH INSTITUTE