# SCOT: Soft Correspondence via Optimal Transport for Multi-Source Cross-City Transfer


**Yuyao Wang[1] Min Yang[2] Meng Chen[2] Weiming Huang[3] Yilong Yin[2] Yongshun Gong[2*]**

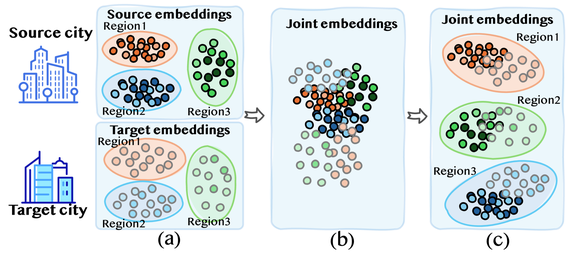
[1]Boston University [2]Shandong University [3]University of Leeds



**NeurIPS 2026**


## 1 Motivation

Labels (GDP, population, $CO_2$) exist only in a few well-instrumented cities. **How can we transfer them to label-scarce cities?**




### Why is it hard?

| No correspondence | $n_s \neq n_t$ | Partial transfer |
|---|---|---|
| Partitions differ across cities | Unequal, rectangular matching | Corridors ✓ Tourist spots ✗ |

**Key insight:** the bottleneck is the **alignment**, not the encoder.

**< 3%** encoder / readout swap *vs* **25–40%** **alignment design**

| Anchor / NN | Distribution | Soft OT (SCOT) |
|---|---|---|
| *RP · HBP · HSA* | *MMD · DANN* | *Sinkhorn + contrastive* |
| ✗ **Hubness** many-to-one collapse | ✗ **Over-mixing** regions blur together | ✓ **Selective** mass-controlled, many-to-many |

**Missing ingredient:** explicit, mass-controlled **soft correspondence** between unequal region sets → **Optimal Transport**

## Problem Setup & Data

| City | #Regions | #Trips |
|---|---|---|
| **Xi'an** | 1,306 | 559,729 |
| **Chengdu** | 1,056 | 384,618 |
| **Beijing** | 1,311 | 78,945 |
| New York | *cross-country evaluation* | |

**Mobility** $M_{ij} = \dfrac{\text{count}(i \to j)}{\sum_k \text{count}(i \to k)}$

**Goal** $\mathbf{z}_s \in \mathbb{R}^{n_s \times d}, \quad \mathbf{z}_t \in \mathbb{R}^{n_t \times d}, \quad n_s \neq n_t$

**Protocol** $h = \text{Ridge}(\mathbf{z}_s, \mathbf{y}_s), \quad \hat{\mathbf{y}}_t = h(\mathbf{z}_t)$

**Tasks:** region GDP · population · $CO_2$ (MAE / MAPE). *No region correspondence is ever used.*

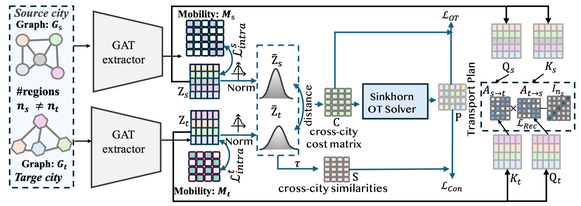
## Contributions

i. **OT soft correspondence:** Sinkhorn OT + OT-weighted contrastive + one-sided cycle.

ii. **Shared prototype hub:** target-induced prior + balanced OT; no source selection.

iii. **Consistent gains:** 5–50% lower MAE/MAPE across cities, tasks & countries.

## 2 Method: SCOT




**1 Intra-city encoding** *→ preserve mobility*

$$\mathbf{z}_c = \text{GAT}(\mathbf{H}_c^{(0)}; \mathbf{A}_c), \qquad \hat{P}_{ij}^{(c)} = \frac{\exp(\mathbf{z}_{c,i}^\top \mathbf{z}_{c,j})}{\sum_k \exp(\mathbf{z}_{c,i}^\top \mathbf{z}_{c,k})}$$

$$\mathcal{L}_{\text{intra}} = -\sum_{c \in \{s,t\}} \sum_{i,j} (\mathbf{M}_c)_{ij} \log \hat{P}_{ij}^{(c)}$$

**2 Sinkhorn soft correspondence** *→ no hubness*

$$C_{ij} = \left\| \tilde{\mathbf{z}}_i^s - \tilde{\mathbf{z}}_j^t \right\|_2, \qquad \mathbf{K} = \exp(-\mathbf{C}/\varepsilon)$$

$$\mathbf{u} \leftarrow \mathbf{a} \oslash (\mathbf{K}\mathbf{v}), \quad \mathbf{v} \leftarrow \mathbf{b} \oslash (\mathbf{K}^\top \mathbf{u}) \qquad (T \text{ iterations})$$

$$\mathbf{P} = \text{diag}(\mathbf{u})\, \mathbf{K}\, \text{diag}(\mathbf{v}), \qquad \mathcal{L}_{\text{OT}} = \frac{\langle \mathbf{P}, \mathbf{C} \rangle}{\min(n_s, n_t)}$$

**3 OT-weighted contrastive** *→ no over-mixing*

$$S_{ij} = \tilde{\mathbf{z}}_i^{s\top} \tilde{\mathbf{z}}_j^t / \tau, \qquad \mathcal{L}_{\text{Con}} = -\frac{1}{n_s} \sum_{i=1}^{n_s} \log \frac{\sum_j P_{ij}\, e^{S_{ij}}}{\sum_j e^{S_{ij}}}$$

**4 One-sided cycle regularizer** *→ stability*

$$\mathbf{A}_{s \to t} = \text{softmax}\left( \frac{\mathbf{Q}_s \mathbf{K}_t^\top}{\sqrt{d}} \right), \qquad \mathcal{L}_{\text{Rec}} = \left\| \mathbf{A}_{s \to t} \mathbf{A}_{t \to s} - \mathbf{I}_{n_s} \right\|_F^2 + \beta\, \mathcal{R}_{\text{ent}}$$

$$\mathcal{L} = \mathcal{L}_{\text{intra}}^{(s)} + \mathcal{L}_{\text{intra}}^{(t)} + \lambda_{\text{align}} \left( \mathcal{L}_{\text{OT}} + \eta\, \mathcal{L}_{\text{Con}} \right) + \lambda_{\text{rec}}\, \mathcal{L}_{\text{Rec}}$$

$$\varepsilon = 0.15, \quad \tau = 0.1, \quad \eta = 0.5, \quad \beta = 0.05, \quad \lambda_{\text{align}} = 1, \quad \lambda_{\text{rec}} = 0.5$$

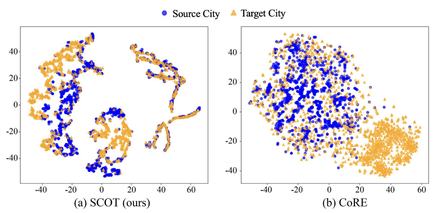
**Proposition:** **lower contrastive loss ⇒ tighter transfer bound** (empirical r = 0.95)

$$\mathcal{R}_t^b(h) \;\leq\; \mathcal{R}_s^a(h) + (L_h + L_g)\sqrt{2 - 2\,\underline{m}}$$

$$\underline{m} = \max\left\{ -1,\; \tau \log n_t + \tau H(\mathbf{a}) - \tau\, \mathcal{L}_{\text{Con}}(\mathbf{P}) - 1 - \frac{1}{2\tau} \right\}$$

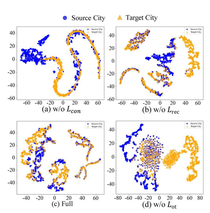
### What SCOT learns: Aligned, not collapsed

t-SNE, XA→BJ



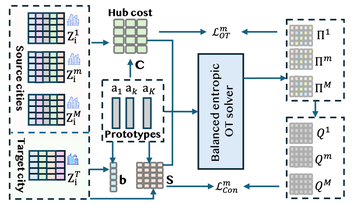

**SCOT vs. CoRE:** cities interleave; CoRE leaves target clusters unmatched. **1.49×** semantic lift.

### Every loss matters




**Ablation:** w/o $\mathcal{L}_{\text{Con}}$ (a), $\mathcal{L}_{\text{Rec}}$ (b), $\mathcal{L}_{\text{OT}}$ (d); full (c) aligns best.

## 3 Multi-Source Hub




- K = 32 shared prototypes as common anchors
- Target-induced prior steers the hub
- Balanced OT → no source domination
- **~30× cheaper than pairwise OT**

**Target prior** $\bar{s}_k = \dfrac{1}{n_t} \sum_{j=1}^{n_t} \tilde{\mathbf{z}}_j^{t\top} \tilde{\mathbf{a}}_k, \qquad b_k \propto \max\left\{ \exp(\bar{s}_k/\tau_b),\, \epsilon_b \right\}$

**City→hub OT** $\boldsymbol{\Pi}^m = \underset{\mathbf{P}\mathbf{1}=\mathbf{a}^m,\ \mathbf{P}^\top\mathbf{1}=\mathbf{b}}{\arg\min} \langle \mathbf{P}, \mathbf{C}^m \rangle - \varepsilon H(\mathbf{P}), \qquad C_{ik}^m = \left\| \tilde{\mathbf{z}}_i^m - \tilde{\mathbf{a}}_k \right\|_2$

**Objective** $\mathcal{L} = \sum_{m \in \mathcal{S} \cup \{t\}} \mathcal{L}_{\text{intra}}^m + \dfrac{1}{|\mathcal{S}|+1} \sum_m \left( \lambda_{\text{align}} \mathcal{L}_{\text{Align}}^m + \lambda_{\text{rec}} \mathcal{L}_{\text{Rec}}^m \right)$

## 4 Results

| 18/18 | −54.5% | +21–32% | +3.9% |
|---|---|---|---|
| best MAE & MAPE (single-source) | max MAPE ↓ vs. best baseline (multi-source) | MAE gain with NYC as target | 3-source hub vs. best single (→NYC) |

**Baselines (8 methods)** grouped by alignment paradigm

| | |
|---|---|
| Correspondence-based | RP · HBP · HSA — surrogate anchors (rank / hierarchical prototypes) + Procrustes / affine map |
| Correspondence-free | MMD · Adv (DANN) · CrossTReS (region re-weighting) · CoRE (global + region-level latent alignment) |
| Reference | Non-Alignment — intra-city objectives only, no cross-city signal |

**MAE ↓ vs. best baseline (6 directions)**

| Direction | GDP | Pop. | $CO_2$ |
|---|---|---|---|
| XA→BJ | **−26.9%** | −10.2% | −9.8% |
| BJ→XA | −4.5% | −9.5% | −9.1% |
| XA→CD | −4.8% | −6.6% | −8.2% |
| CD→XA | −8.9% | −10.4% | −5.8% |
| BJ→CD | −8.2% | −8.6% | −4.3% |
| CD→BJ | −9.1% | −8.2% | −2.0% |

**Single-source MAE, XA ↔ BJ** red: best · blue: runner-up

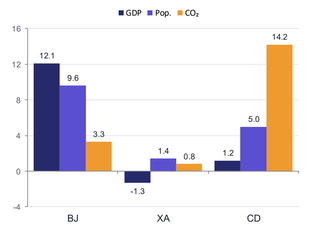
| Method | XA → BJ | | | BJ → XA | | |
|---|---|---|---|---|---|---|
| | GDP | Pop. | $CO_2$ | GDP | Pop. | $CO_2$ |
| Non-Align | 264.30 | 981.07 | 288.41 | 252.91 | 946.93 | 270.42 |
| RP | 189.84 | 684.50 | 196.05 | 181.08 | 670.44 | 191.85 |
| HBP | 177.69 | 665.46 | 188.72 | 196.20 | 627.46 | 176.35 |
| HSA | 201.27 | 619.33 | 176.40 | 188.40 | 636.33 | 182.91 |
| MMD | 183.32 | **588.34** | **165.70** | 180.73 | **499.94** | **141.57** |
| Adv | 192.59 | 702.19 | 199.23 | 199.21 | 805.01 | 203.27 |
| CrossTReS | 207.43 | 633.25 | 179.75 | 170.27 | 639.44 | 182.72 |
| CoRE | **157.83** | 611.18 | 166.28 | **162.19** | 547.74 | 153.63 |
| SCOT (ours) | **115.33** | **528.50** | **149.42** | **154.92** | **452.67** | **128.74** |

**Multi-source MAE (two sources → target)** shared hub vs. baselines with adaptive source weighting

| Method | Target: Beijing | | | Target: Xi'an | | | Target: Chengdu | | |
|---|---|---|---|---|---|---|---|---|---|
| | GDP | Pop. | $CO_2$ | GDP | Pop. | $CO_2$ | GDP | Pop. | $CO_2$ |
| RP | 172.76 | 679.86 | 166.80 | 195.91 | 642.01 | 181.16 | 187.13 | 630.44 | 167.76 |
| HBP | 164.25 | 662.60 | 165.53 | 200.04 | 670.41 | 150.16 | 176.35 | 631.26 | 125.97 |
| HSA | 156.40 | 644.89 | 160.05 | 183.55 | 648.67 | 155.20 | 180.71 | 651.06 | 159.86 |
| MMD | **127.45** | **605.81** | 160.52 | **163.78** | **506.22** | 144.61 | 163.82 | 639.32 | 145.74 |
| Adv | 196.76 | 717.96 | 189.41 | 221.31 | 731.92 | 184.73 | 183.26 | 687.72 | 156.37 |
| CrossTReS | 151.17 | 666.74 | 187.59 | 179.01 | 625.63 | 151.22 | 167.53 | 668.00 | 139.22 |
| CoRE | 152.88 | 620.34 | **152.24** | 173.72 | 549.37 | **134.19** | **156.10** | **621.11** | **121.33** |
| SCOT (ours) | **104.16** | **525.10** | **143.53** | **156.94** | **446.13** | **127.66** | **133.94** | **546.82** | **98.43** |

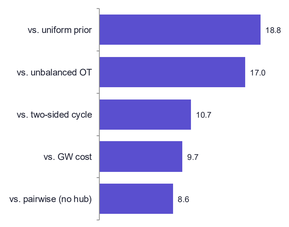
**Multi- vs. best single-source**

MAE reduction (%), per target



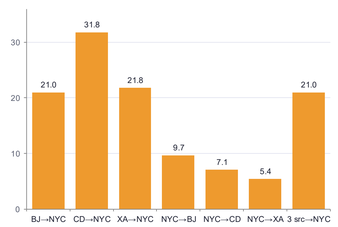

**Design ablations**

Pop. MAE reduction vs. alternative (%)




**Cross-country: New York**

Pop. MAE gain vs. strongest baseline (CoRE, %)




**Diagnostics & robustness**

| 35 / 1311 | ≈ 4 | 63% | 1.49× | < 3% |
|---|---|---|---|---|
| targets hold 50% of each row's OT mass | active prototypes per region (hub) | OT mass between same-function blocks | GT correspondence lift (synthetic; OT-only ≈ 1.0) | MAE change across encoders & readouts |

**Takeaways**

- Invest in alignment, not encoders: it drives cross-city transfer.
- OT + contrastive: selective many-to-many matches, no hubness or over-mixing.
- A shared hub lets multiple sources help each other.