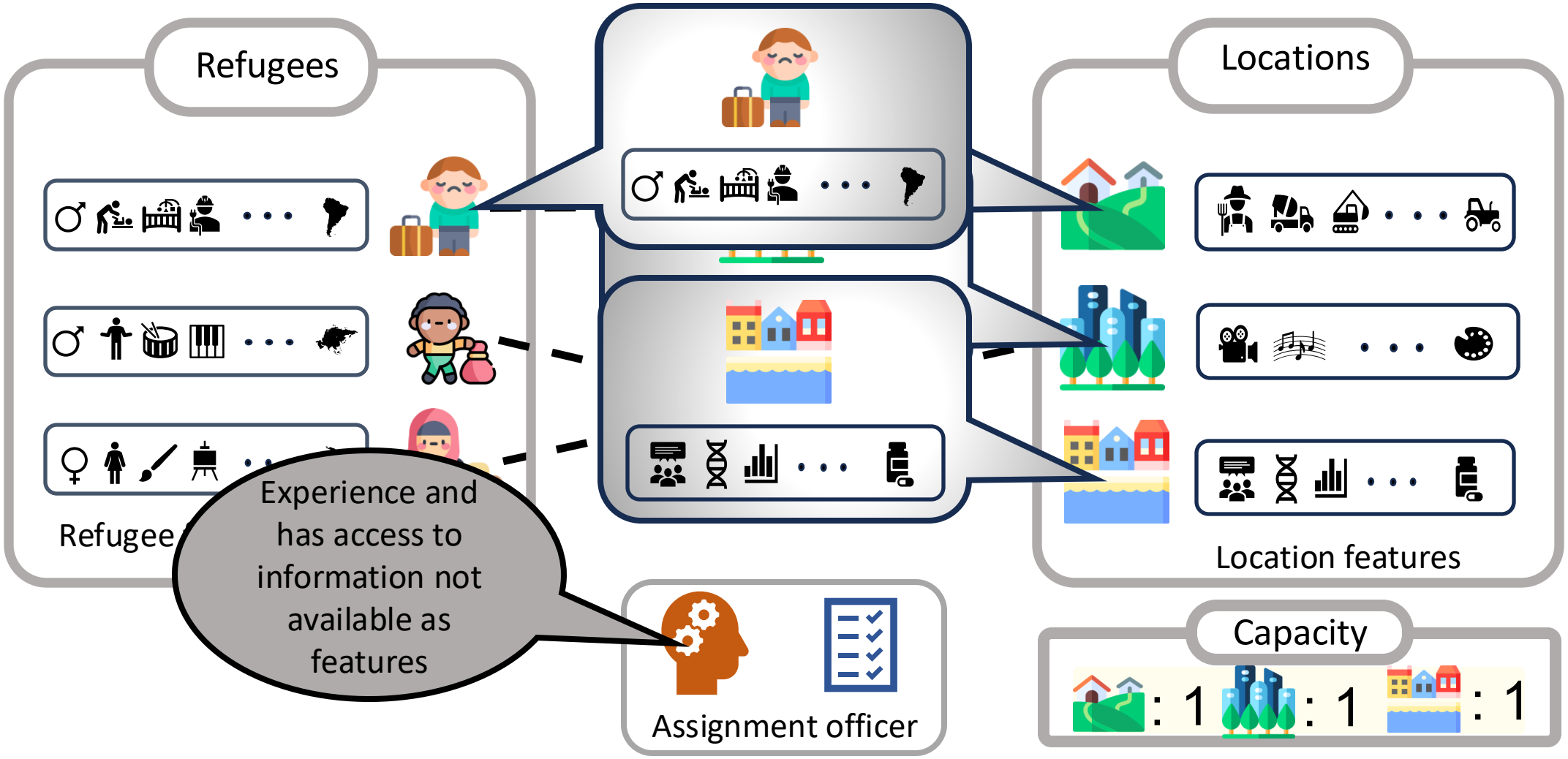
Matchings, Predictions and Counterfactual Harm in Refugee Resettlement Processes

Seungeon Lee [2], Nina Corvelo Benz [1,3], **Suhas Thejaswi** [1], Manuel Gomez-Rodriguez [1]

[1] Max Planck Institute for Software Systems
[2] Korea Advanced Institute of Science and Technology
[3] ETH Zurich

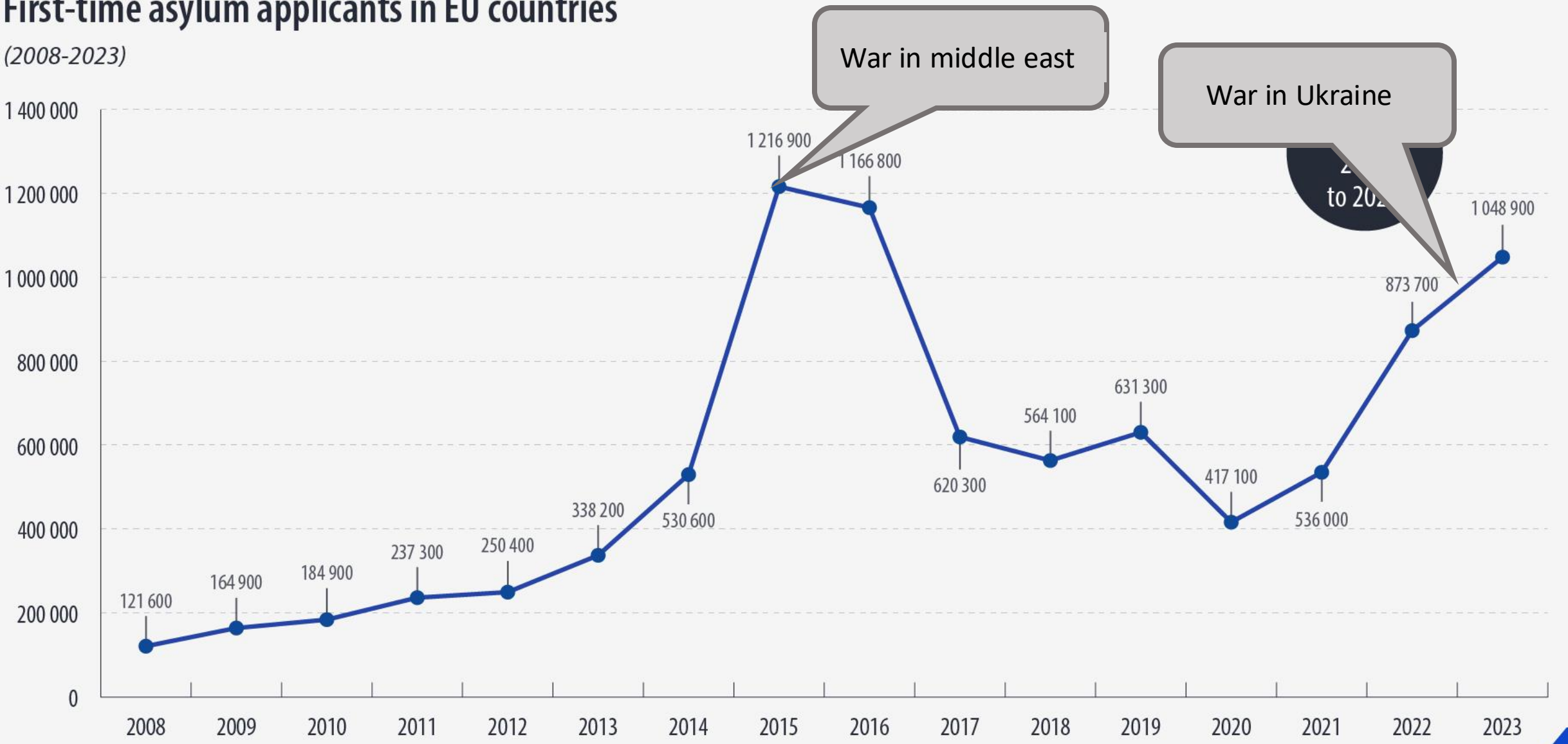
Historical refugee relocation process



This we will refer as **the default policy**

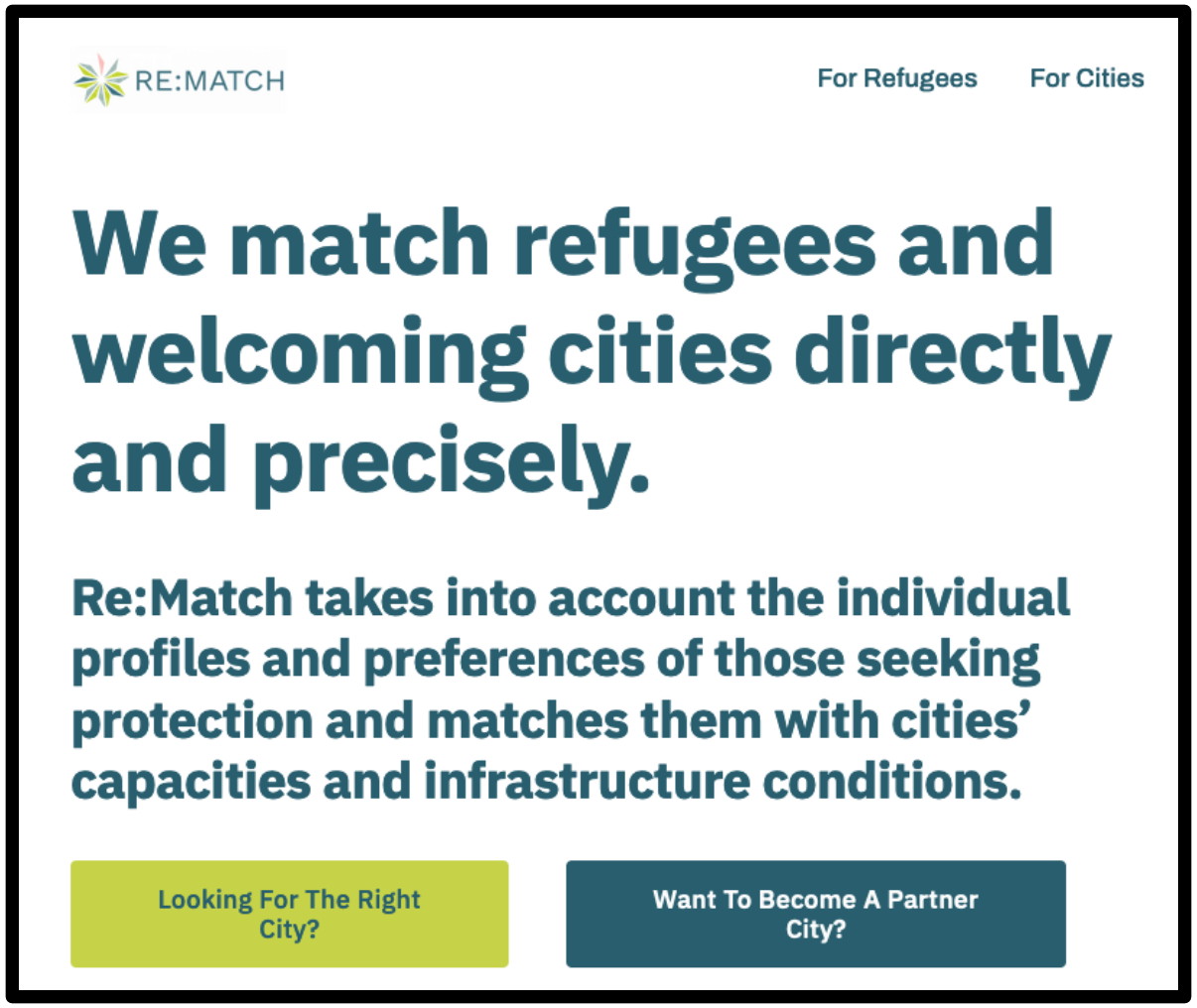
First-time asylum applicants in EU countries

(2008-2023)



Notes: 2008-2013 sum of available data. Numbers are rounded to the nearest 100.

Data-driven refugee relocation process



We are using tools from machine learning, integer optimization, and matching theory to find the best matches between refugees and local communities.

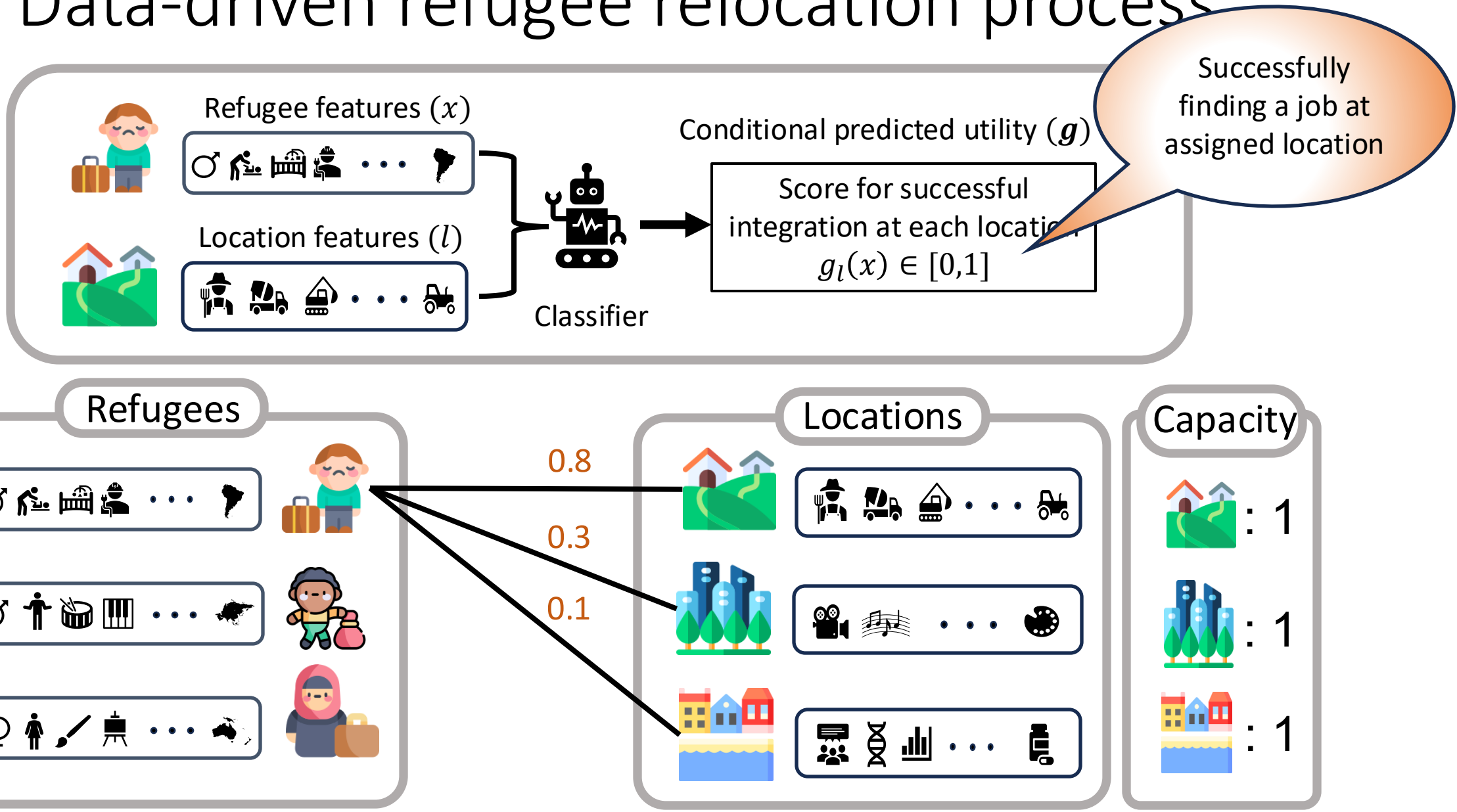
An official report by the Swedish Government and the Independent Chief Inspector of Borders and Immigration in the UK have recommended improving the geographic matching of refugees.

Introducing *Annie*™ MOORE

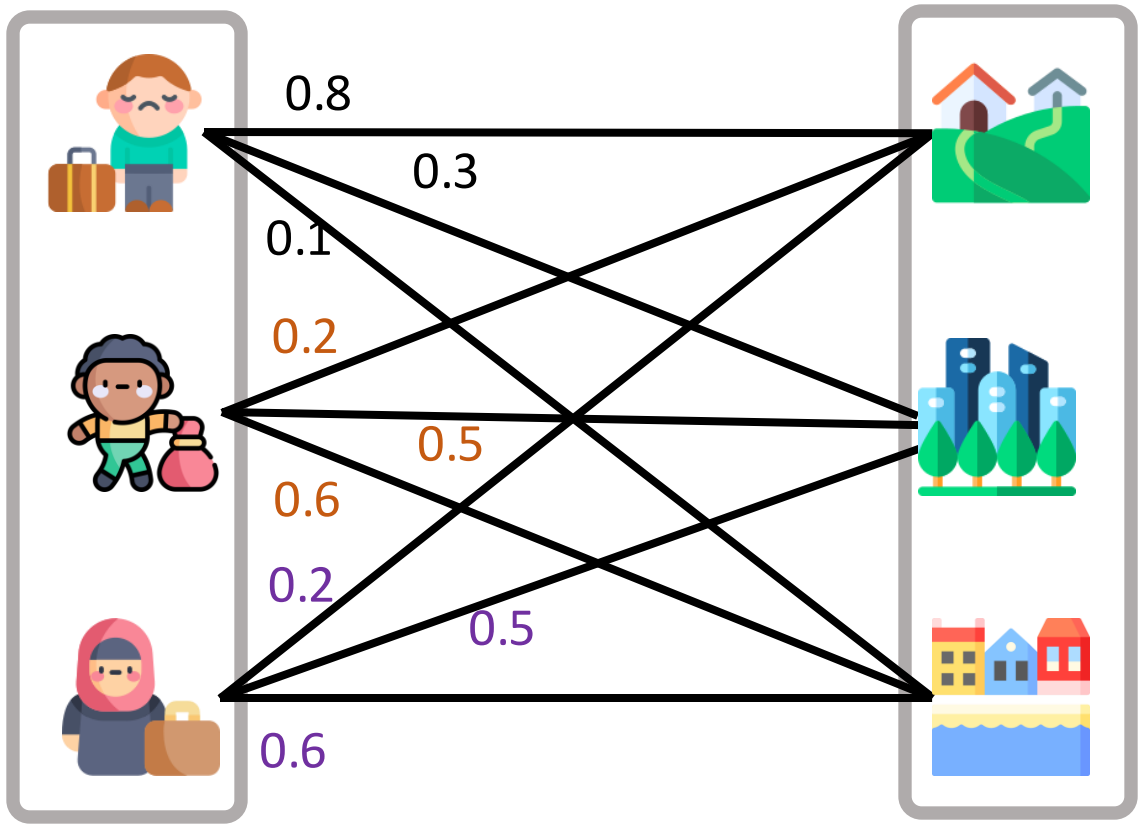
Annie™ *MOORE* (*M*atching and *O*utcome *O*ptimization for *R*efugee *E*mpowerment) is the world's first software that helps resettlement agencies optimize their initial placement of refugees within host countries. It is named in honor of Annie Moore, the first person to be processed at Ellis Island in 1892.

Read about how *Annie*™ has been helping a US resettlement agency since May 2018 here.

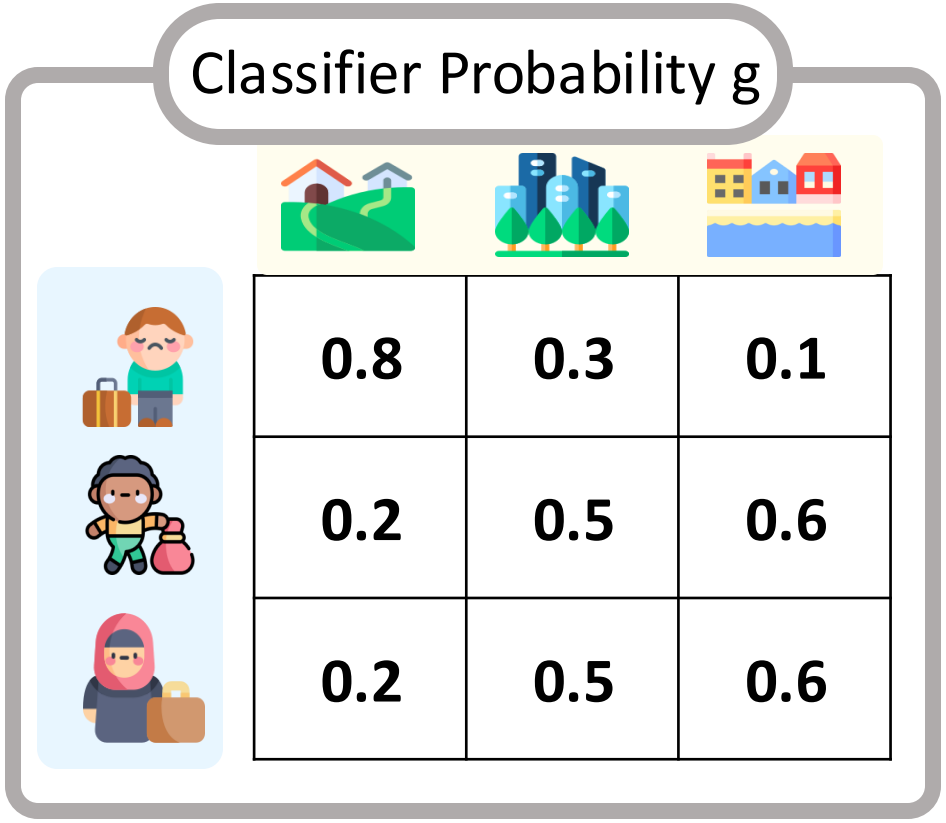
Data-driven refugee relocation process



Data-driven refugee relocation process

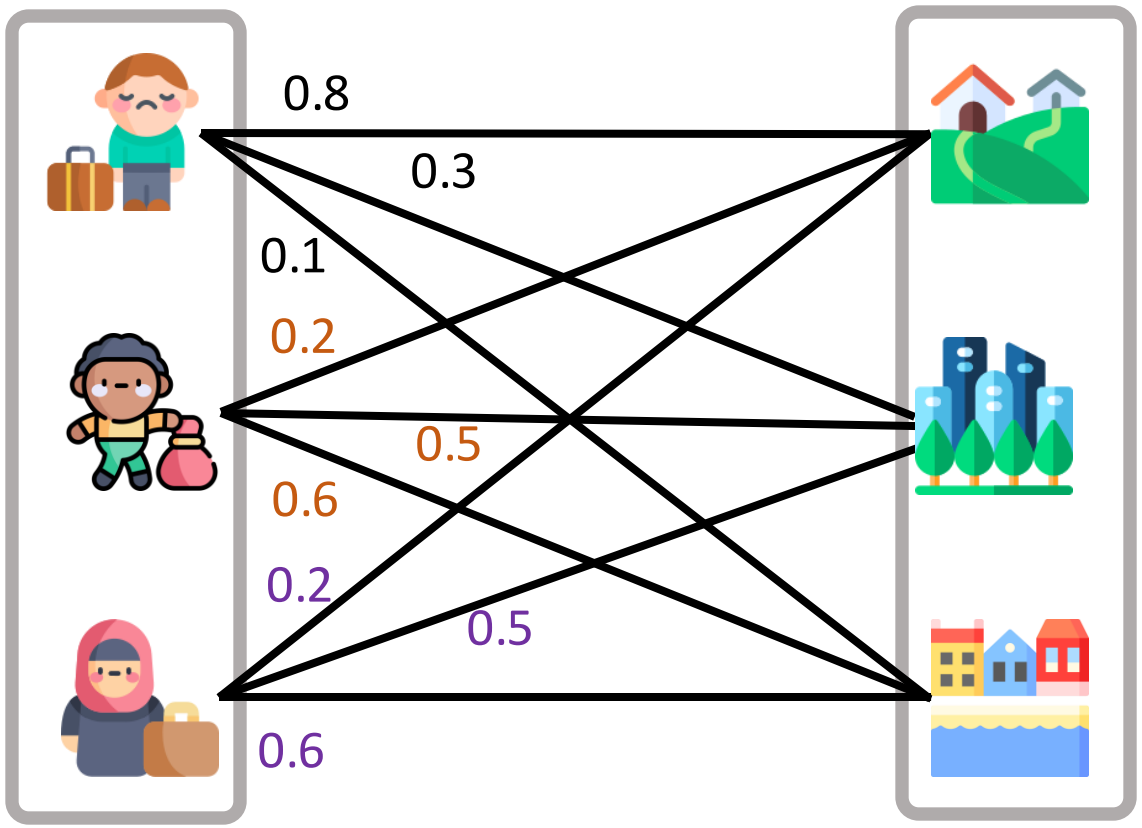


A bipartite graph

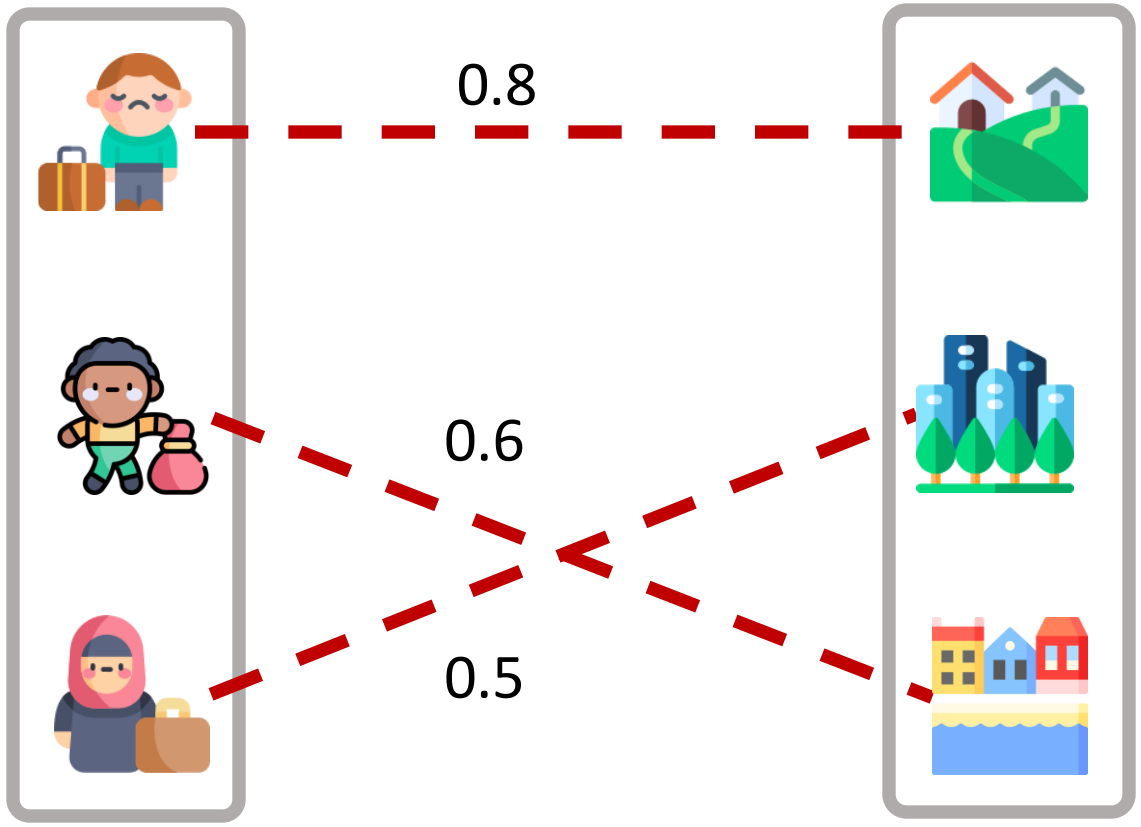


A matrix representation

Data-driven refugee relocation process



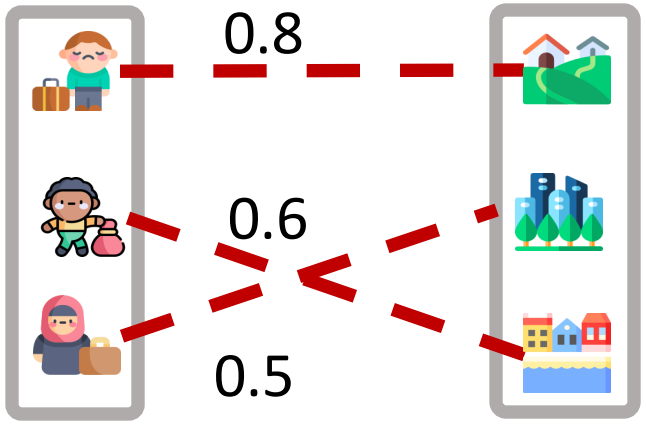
Maximum weight bipartite matching problem



Utility of $\pi(g)$ = 0.8 + 0.6 + 0.5

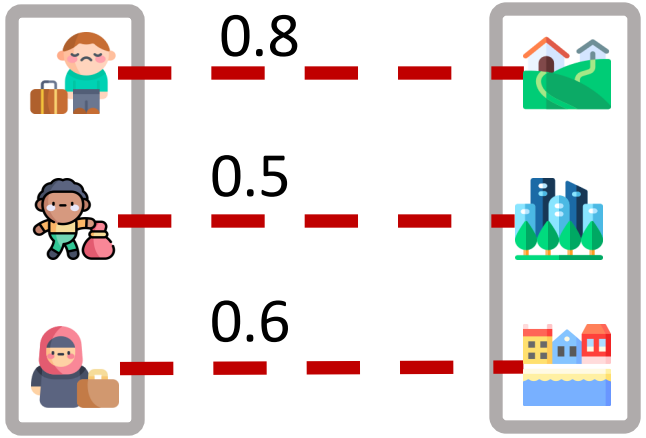
Find a matching $\pi(g)$ that maximizes the predicted utility

Data-driven refugee relocation process



Classifier probability g

0.8	0.3	0.1
0.2	0.5	**0.6**
0.2	**0.5**	0.6

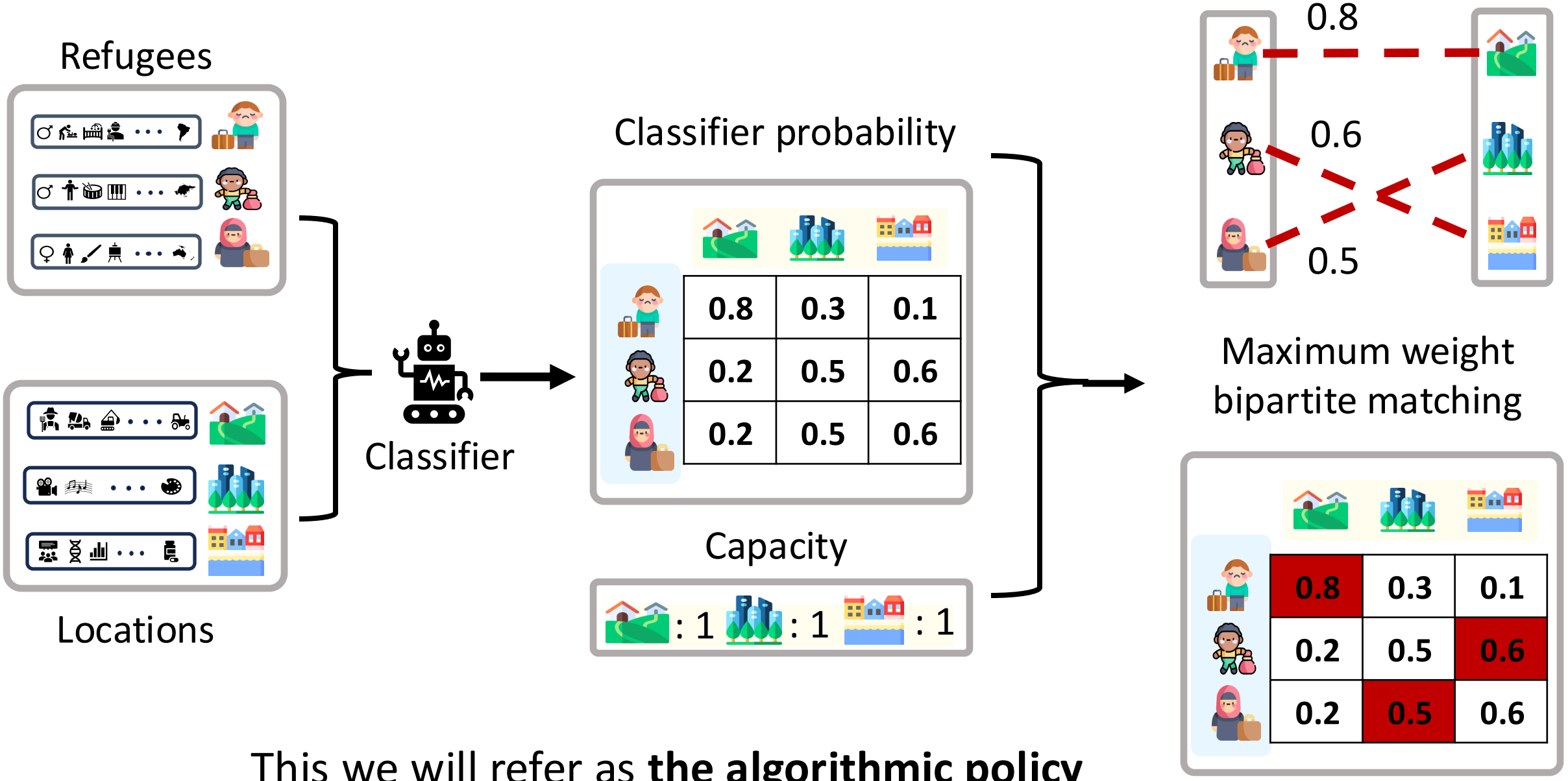


Classifier probability

0.8	0.3	0.1
0.2	**0.5**	0.6
0.2	0.5	**0.6**

Utility of $\pi(g)$ = 0.8 + 0.6 + 0.5

Data-driven refugee relocation process



This we will refer as **the algorithmic policy**

Predicted vs expected utility

- a matching that maximizes **the (conditional) predicted utility**

$$\hat{\pi}(\mathbf{g}) \in \hat{\Pi}(\mathbf{g}) = \operatorname{argmax}_{\pi(\mathbf{g}) \in \Pi(\mathbf{g})} \sum_{i \in \mathcal{I}} g_{\pi_i(\mathbf{g})}(x_i)$$

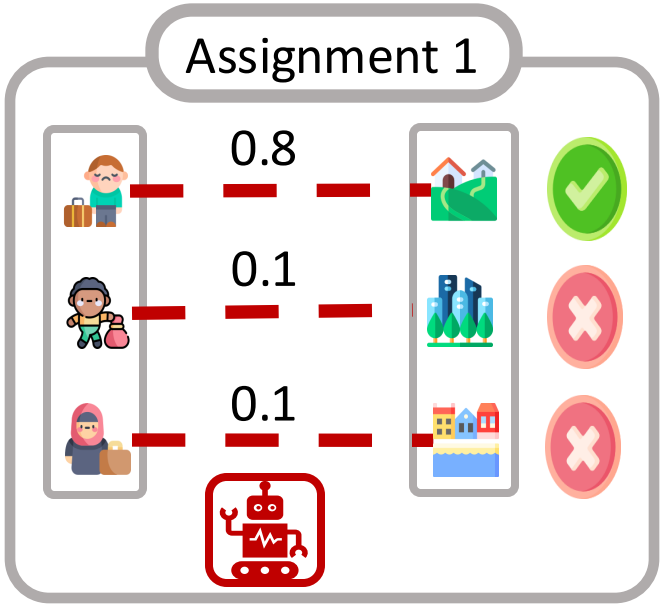
- a matching that maximizes **the (conditional) expected utility**

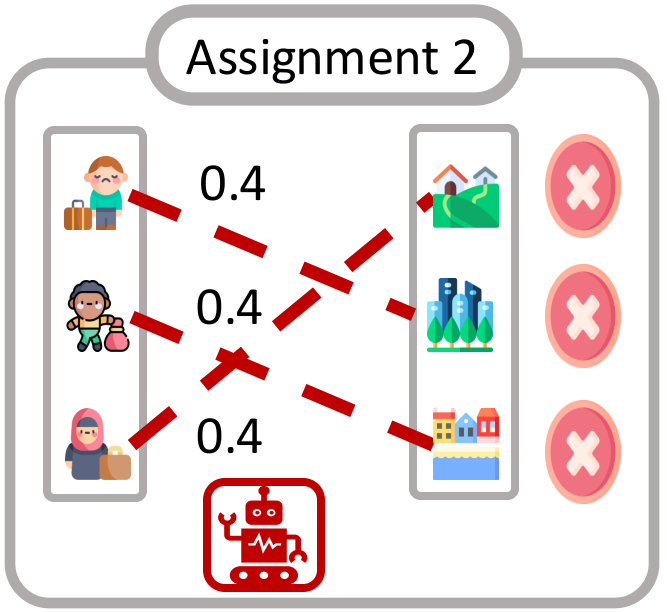
$$\pi^*(\mathbf{g}) \in \Pi^*(\mathbf{g}) = \operatorname{argmax}_{\pi(\mathbf{g}) \in \Pi(\mathbf{g})} \mathbb{E}_{\mathbf{Y} \sim P^{\mathcal{M}\,;\,\mathrm{do}(\mathbf{L}=\pi(\mathbf{g}))}} [\mathbf{1}^T \mathbf{Y} \mid \mathbf{X} = \mathbf{x}]$$

If the classifier is perfectly calibrated a matching that maximizes the predicted utility will also maximize the expected utility

This is not possible to achieve

What do we mean by calibration?



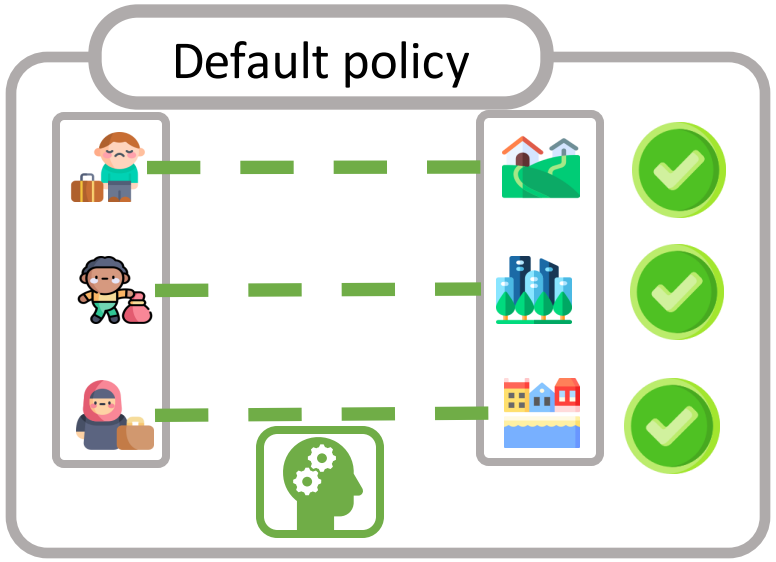


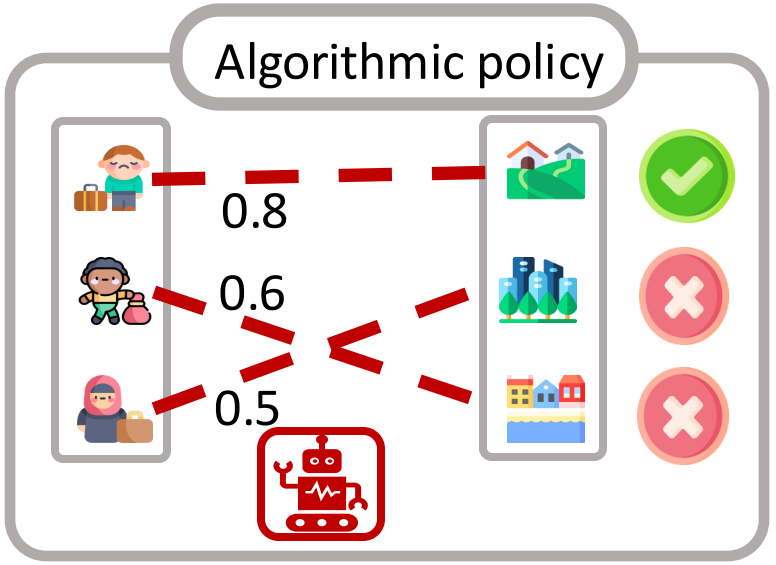
Utility = 0.8 + 0.1 + 0.1 = 1.0 $<$ Utility = 0.4 + 0.4 + 0.4 = **1.2**

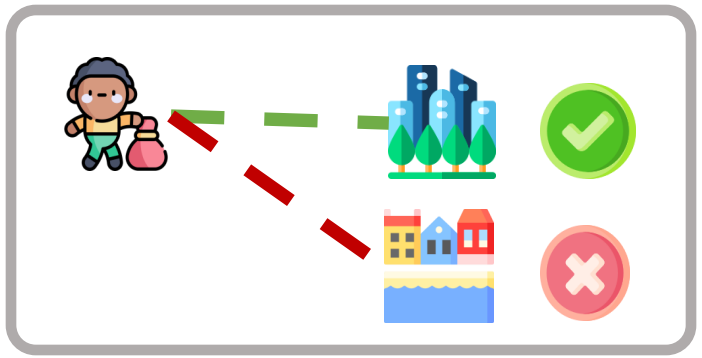
# of refugees finding jobs $\not\leq$ # of refugees finding jobs

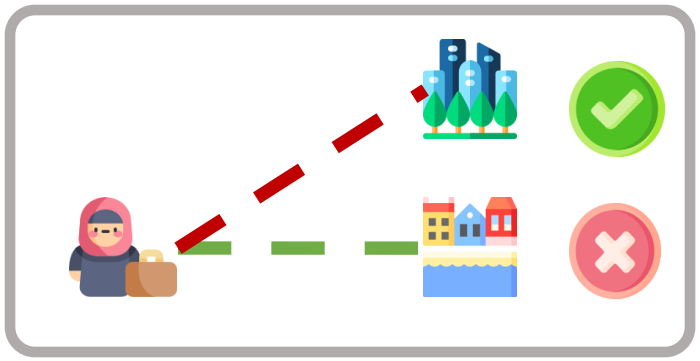
With increase in the (total) predicted utility, the number of refugees finding jobs should not decrease

Counterfactual harm



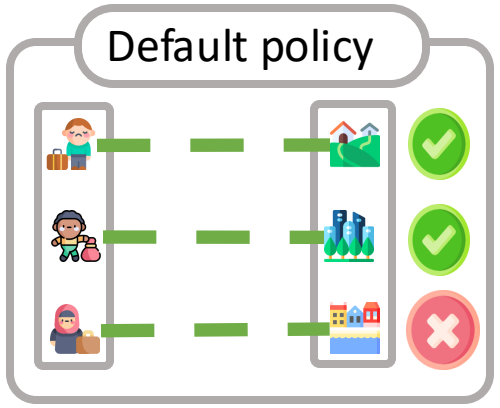


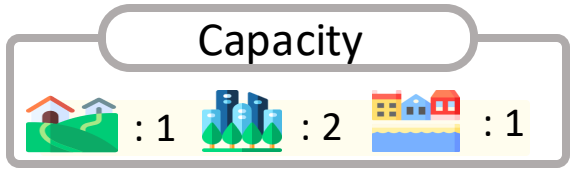


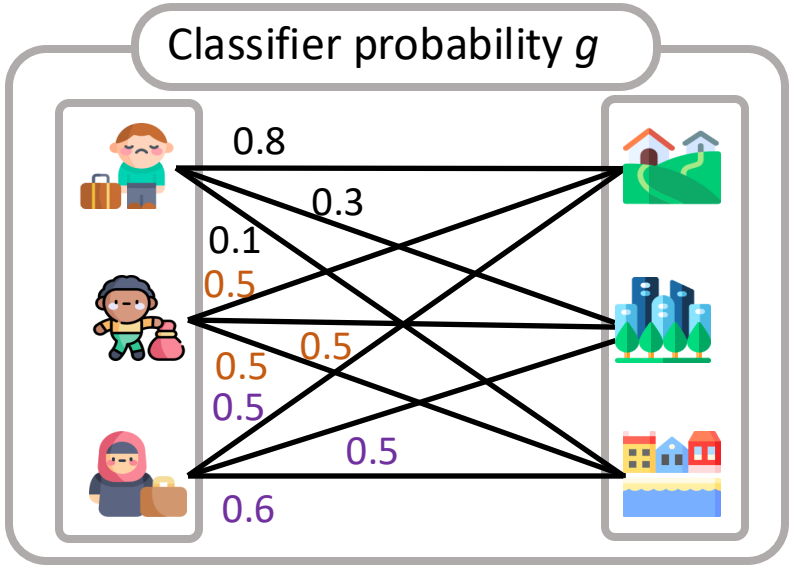


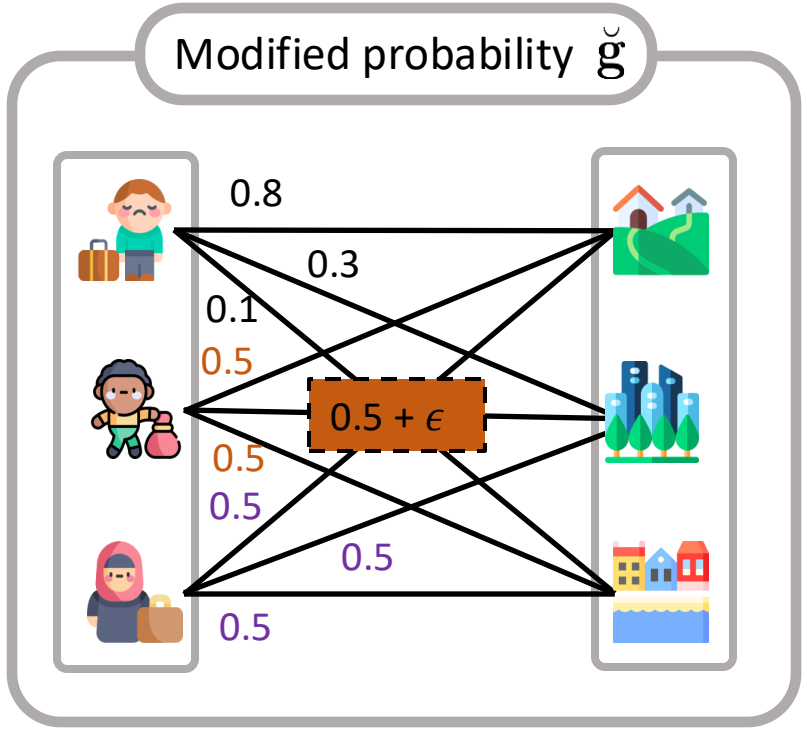
A refugee that would have succeded in finding a job if the default policy had been implemented failed because of the algorithmic policy

Partial inverse bipartite matching problem

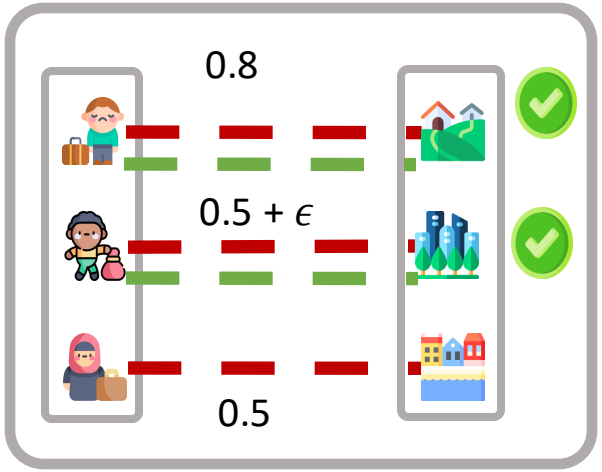




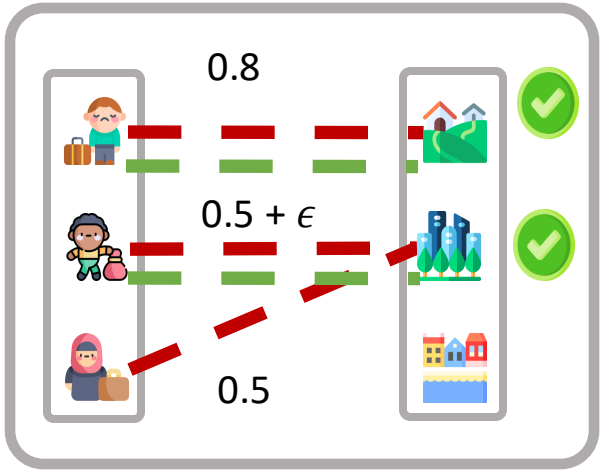




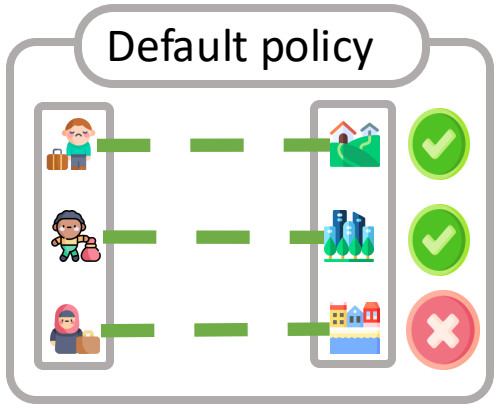
modify the edge weights such that the positive decisions in the default policy become part of every matching that maximizes the predicted utility

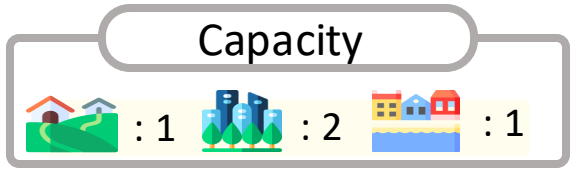


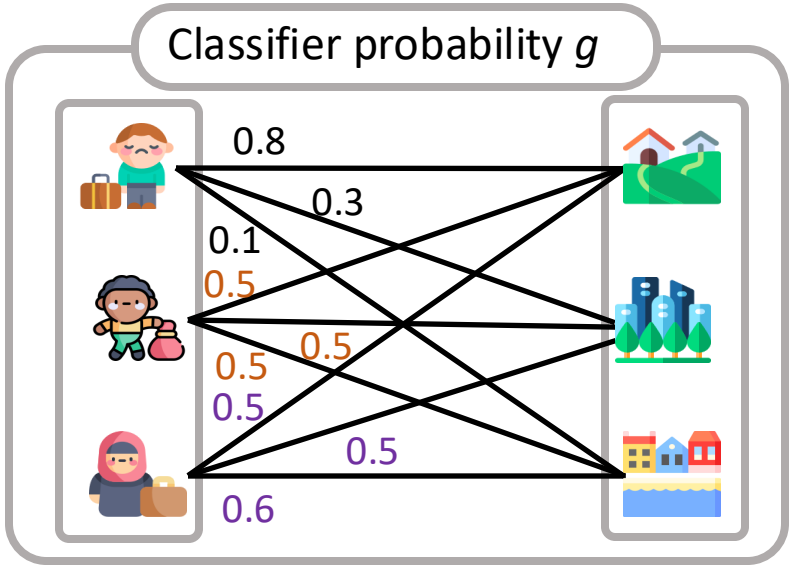
Only positive decisions needs to be preserved

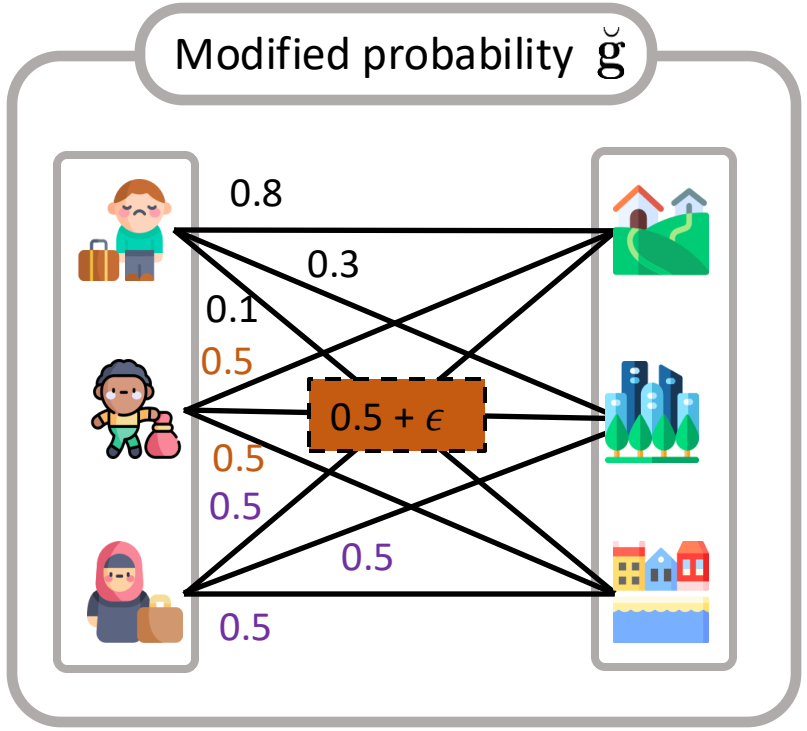


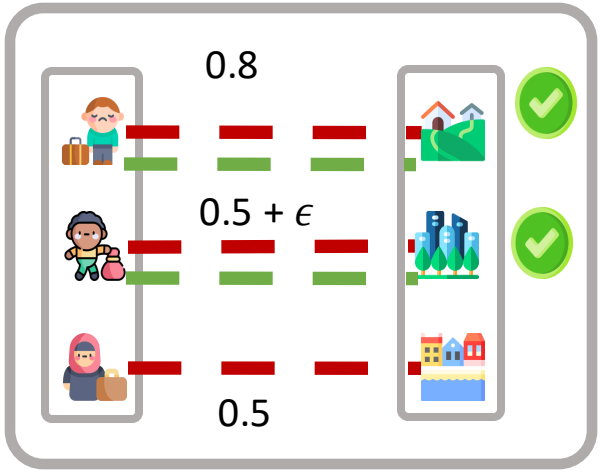
Partial inverse bipartite matching problem



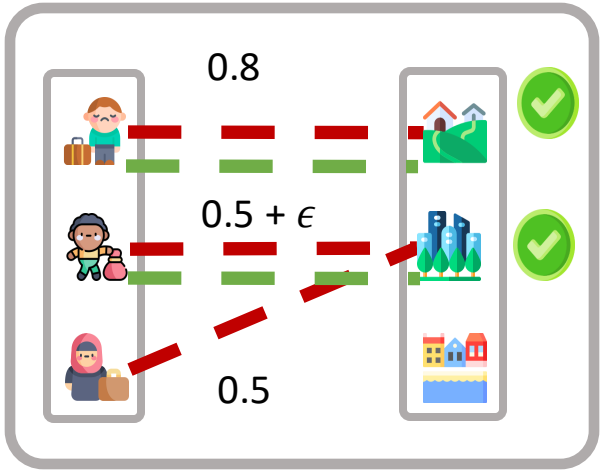






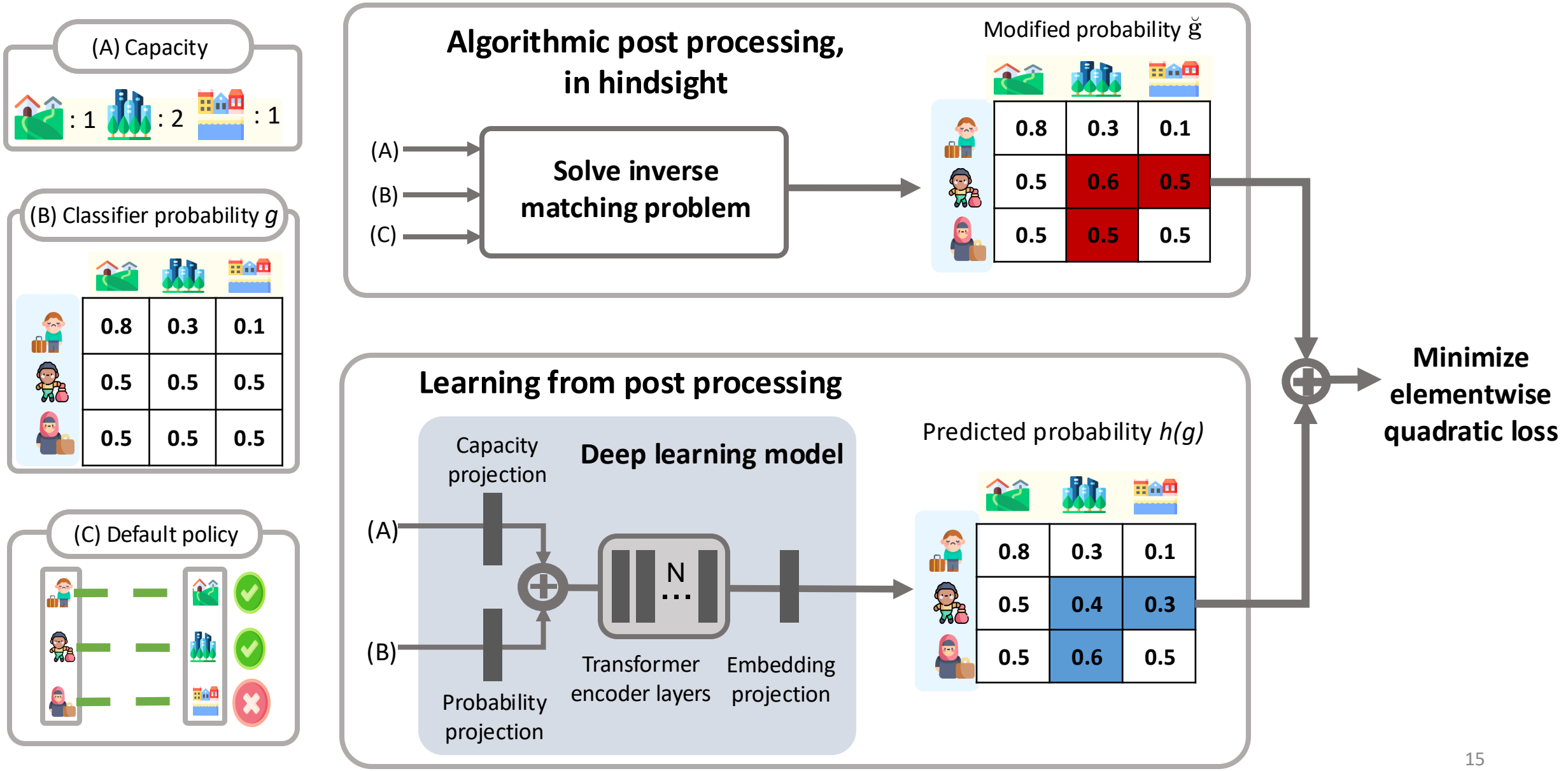


Only positive decisions needs to be preserved



We can only solve this problem for historical data

Post-processing framework



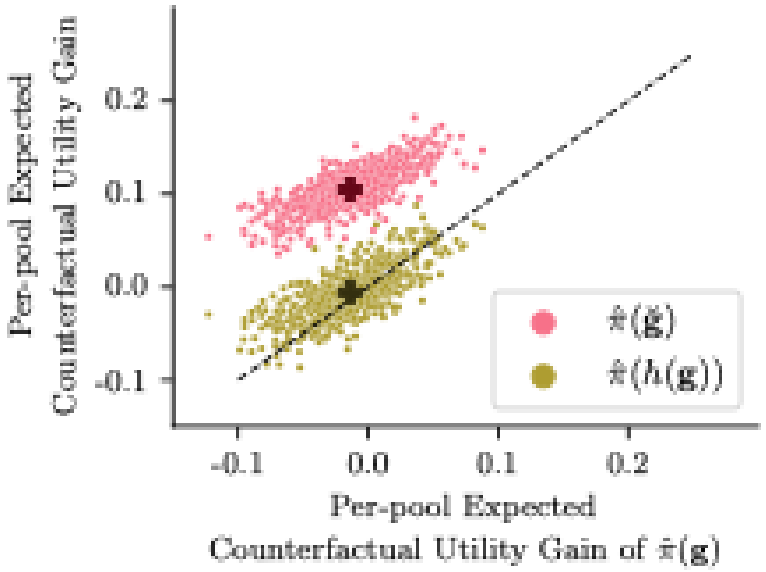
Experiments on synthetic data

- Synthetic data
  - statistics from UNHCR, Migration policy institute, US department of labor, US census
  - generate data that matches the distribution of these statistics
  - possible to compute the true expected utility achieved by any algorithmic matching
- Experimental setting
  - 5000 synthetic pools of refugees to be settled across 10 locations
  - each pool has 100 refugees and features are age, sex, country of origin, education
  - each location corresponds to a US state and features are labour market, demography of labor force

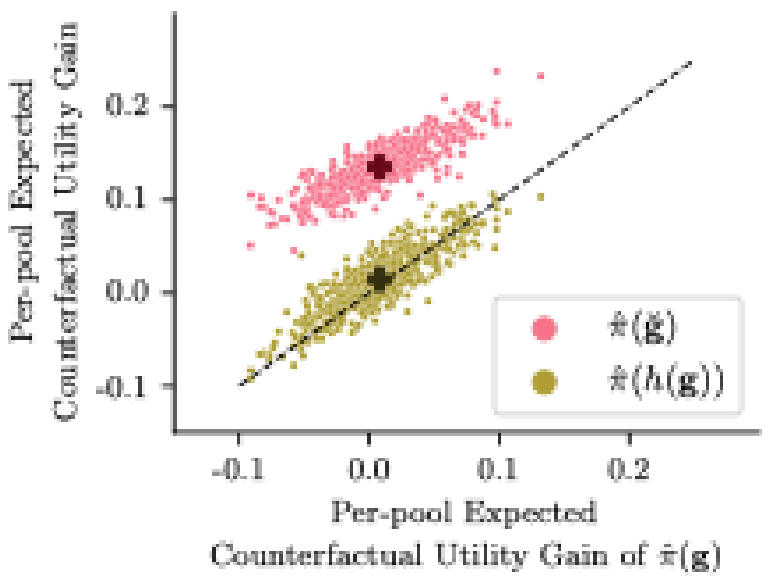
The classifier *g* overestimates the true employment probability for half of the locations, picked at random, and underestimates its value for the remaining half

Experiments on synthetic data

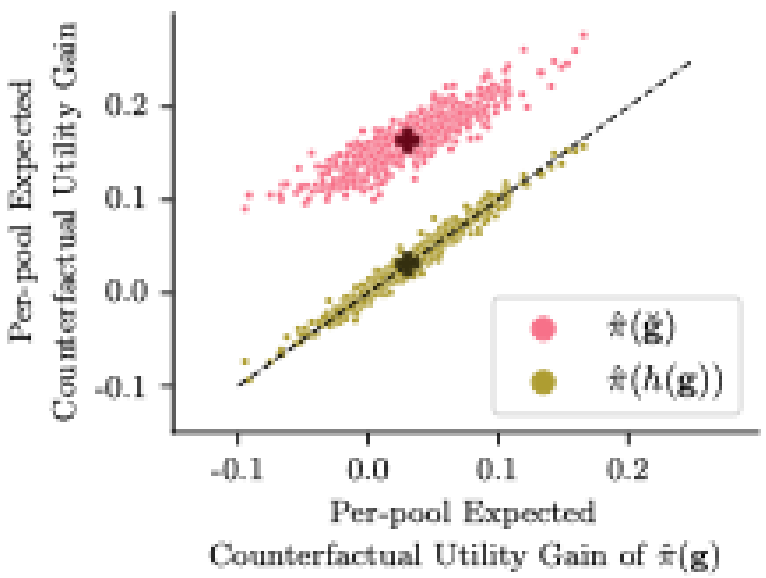
Low noise level (0.25)



Medium noise level (0.5)



High noise level (0.875)



Maximize the conditional predicted utility under the following policies

- $\hat{\pi}(\mathbf{g})$: employment probabilities of classifier
- $\hat{\pi}(\breve{\mathbf{g}})$: modified employment probabilities
- $\hat{\pi}(h(\mathbf{g}))$: predicted employment probabilities of the transformer model

As the level of noise increases in the default policy, it is difficult for our framework to learn to avoid harm from past placement decisions made by the default policy

Summary

- We show that data-driven algorithmic matching may cause (counterfactual) harm to some refugees
- We initiate the discussion on designing of systems that cause less (counterfactual) harm in practice
- Our framework may also have limitations and these limitations must be addressed before deploying

Paper

Code

Anyone could find themselves in the position of being a refugee some day. If that happens, would you be satisfied with the same system that has been used for others?

Paper

Code