Supplemental information

Inferring global-scale temporal latent topics

from news reports to predict public

health interventions for COVID-19

Zhi Wen, Guido Powell, Imane Chafi, David L. Buckeridge, and Yue Li

S1   Experiment Details

The number of topics 25 is chosen based on the best topic quality, as shown in **Fig.** S1.
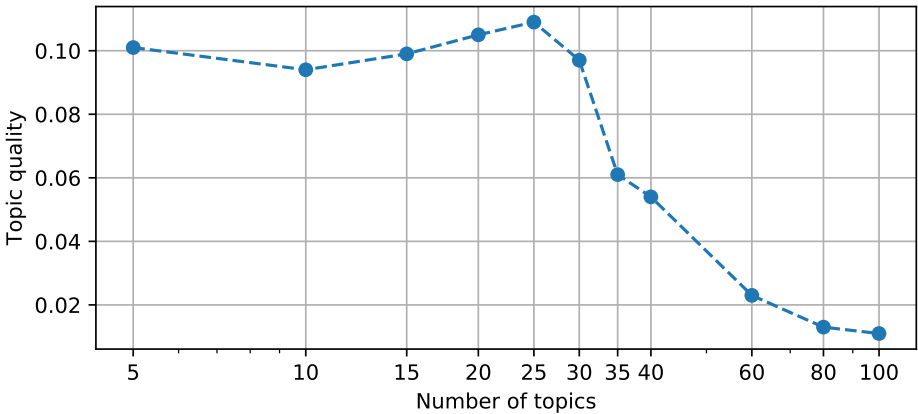

Figure S1: Topic quality versus the number of topics. Using 25 topics leads to the highest topic quality, and that number is used throughout this work. The x-axis is log-scaled.

S2   Precision-Recall Curves of All NPIs

The precision-recall curves for the all NPIs are in **Fig.** S2.
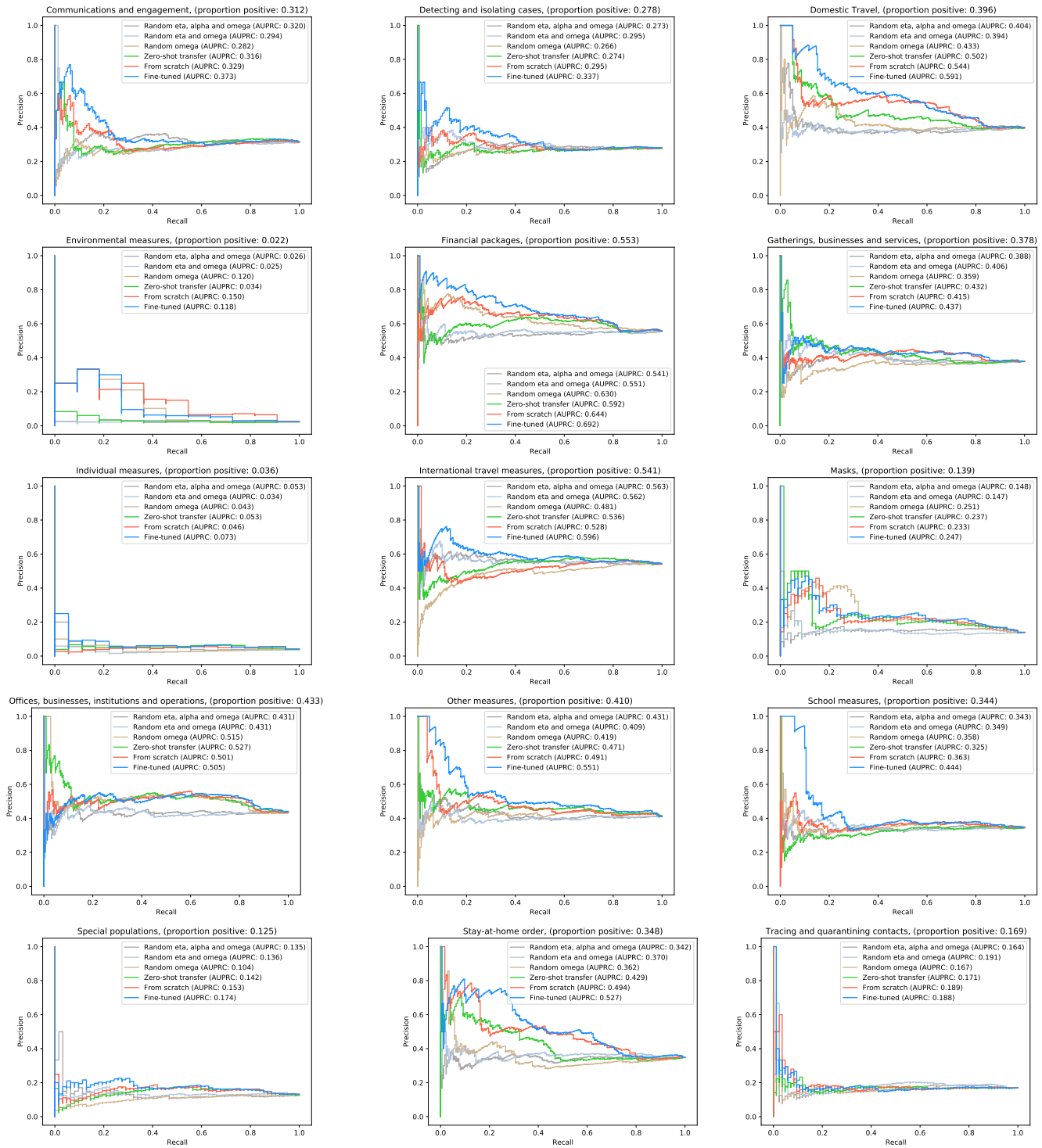
Figure S2: Country-level precision-recall curves of all 15 NPIs over 42 countries. Similar to **Fig.** 6b in the main text, each subfigure corresponds to one NPI, and together they correspond to all the NPIs in **Fig.** 6a. The predictions are made by a linear classifier on the country-specific topic mixture, and different methods in comparison learn the classifier weight differently as described in **Section** 4.5 in the main text. In particular, Random, zero-shot transfer, from-scratch, and fine-tuned are methods that predict NPI using random features, pre-trained linear coefficients from Stage 2 on document-level NPI predictions, training the linear coefficients from random initialization, and fine-tuning the pre-trained linear coefficients from Stage 2, respectively. The numbers in the brackets indicate the proportion of the positive labels for that NPI, which are positively correlated with the corresponding AUPRC.

S3   Dataset Details

The number of positive labels at the document-level in the WHO dataset and at the country-level are shown in **Fig.** S4.
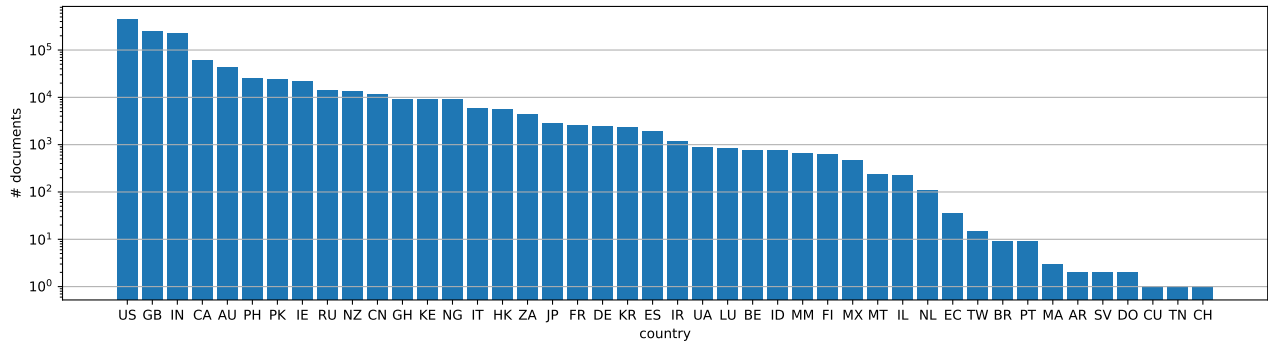

Figure S3: Number of documents of each country in the AYLIEN dataset. The y-axis is log-scaled.

(a) WHO documents

(b) Country-time pairs

Figure S4: The number of positive labels at the document and country levels. The x-axis is log-scaled.
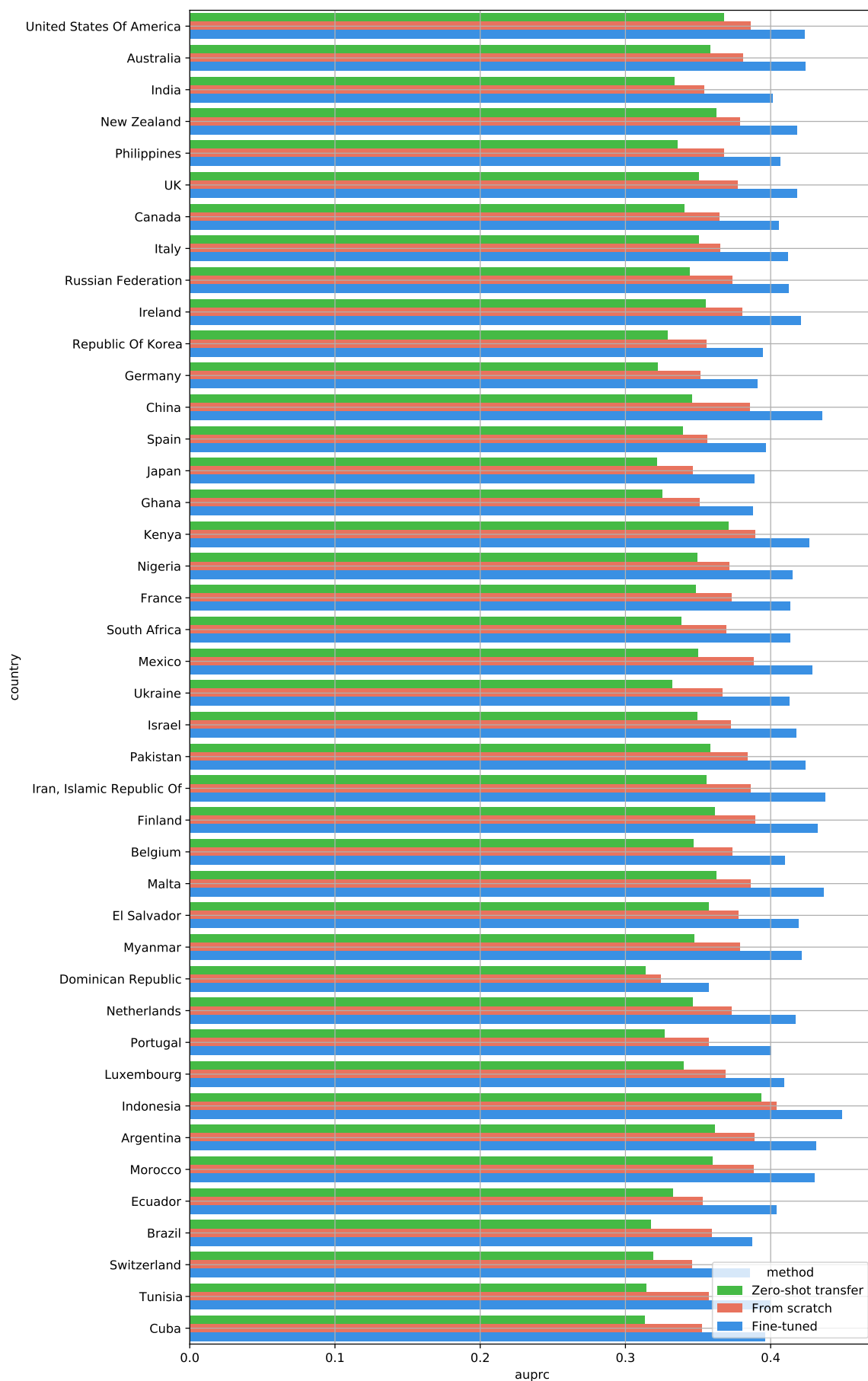
Figure S5: Performance of NPI prediction for each country in terms of macro AUPRC.

S4   Supplementary Tables

Table S1: Grouping of WHO-NPI categories into 15 groups, which are treated as labels in our application of NPI predictions. Excel file named TableS1_npi_labels.xlsx is provided for this table.