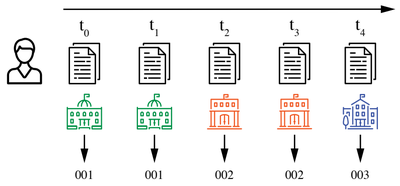
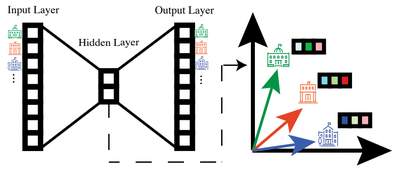
# How far apart are two places?

*A suprisingly difficult question to answer in a world that has rapidly globalized, and where the importance of geography to many kinds of mobility has grown more tenuous.* ***Can we learn it from actual scientific mobility data?***




**Mobility Sentence:** "001 — 001 — 002 — 002 — 003"

$t_0$ $t_1$ $t_2$ $t_3$ $t_4$




**Data** on scientific mobility from the **Web of Science** was used to construct **mobility trajectories** of organizations for authors using papers published between **2008** and **2019**

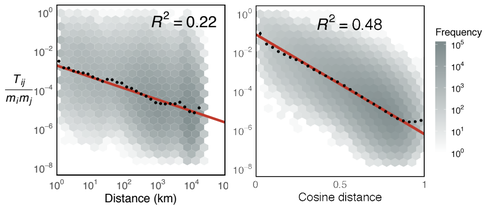
These trajectories were fed into **word2vec**, which learns a **vector representation of organizations**

How well does cosine distance between organization vectors explain actual flows?

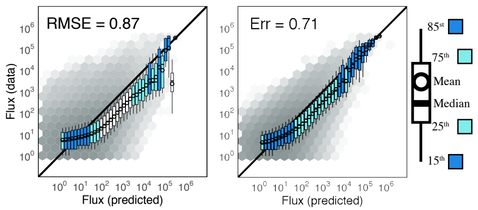
We assess using the **Gravity Model** of Mobility**.** Under this model, expected flow between organizations are proportional to their size and a function of distance

$$\hat{T}_{ij} = C m_i m_j f(r_{ij})$$

## Embedding distance outperforms geographic distance




Distance in the embedding space explains more of the expected flow than geographic distance (left)

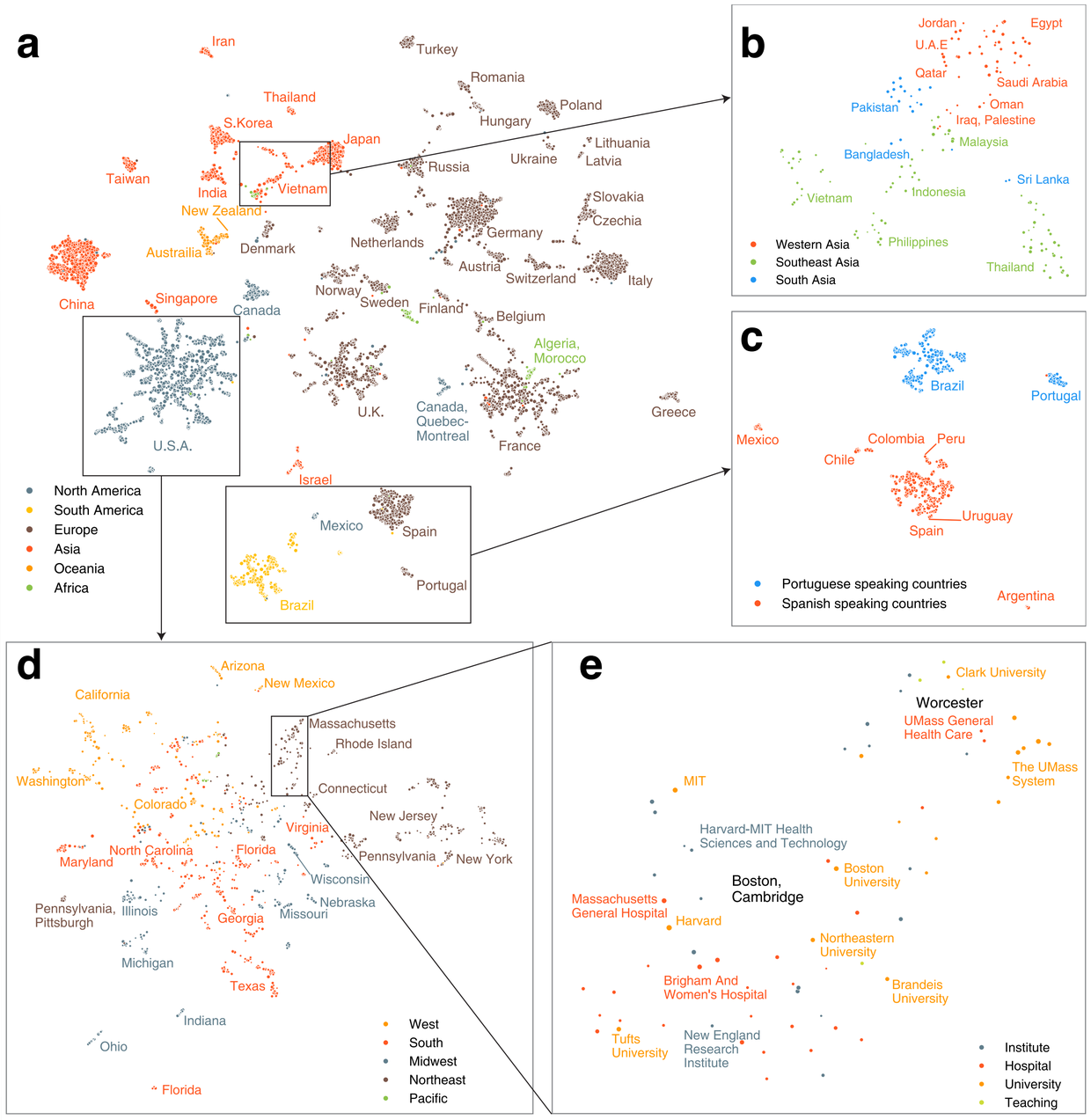



Predictions made as a function of the embedding distance out-perform predictions using geographic distance

# Embedding co-affiliation trajectories captures structure of global scientific mobility

[1]Dakota Murray, [12]Jisung Yoon, [2]Woo-Sung Jung, [1]Stasa Milojevic, [3]Rodrigo Costas [1]Yong-Yeol Ahn

*[1]Indiana Univeirsty Bloomington, [2]Ponahg Univeristy of Science and Technology, [3]Leiden University*




**a** **UMAP projection** of global univeristies reveals geographic clustering

**b** Re-projection of subsets of countires reveals the geographic and cultural gradient between Central Asian and Southeast Asian countries

**c** Spanish-speaking countires in South and Central America neighbor Spain in the embedding space. Similarly, portouguese-speaking Brazil is neaghbors Portugal

**d** Re-projection of organizations within the U.S. reveals the multi-level structure that is captured. Organizations are largely clustered by state, and states by regions

**e** Within Massachuessets, organizations are largely clustered by urban center. However, note the separtion of the teaching-oriented UMass system from other prestigious research univeresities.

Similar patterns are observed for California, New York, Texas, and Pennsylvania

**In conclusion,** this data-driven approach that encodes geography, language, and prestige at multiple scales into a high-resolution, dense, coherent, and meaningful vector space representation

d***akmurra@iu.edu***

***@dakotasmurray***

IC²S²