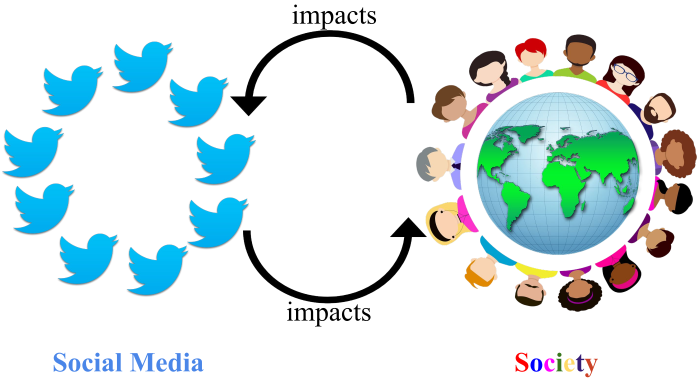
# Semi-supervised Geolocation via Graph Convolutional Networks

Afshin Rahimi, Trevor Cohn and Tim Baldwin

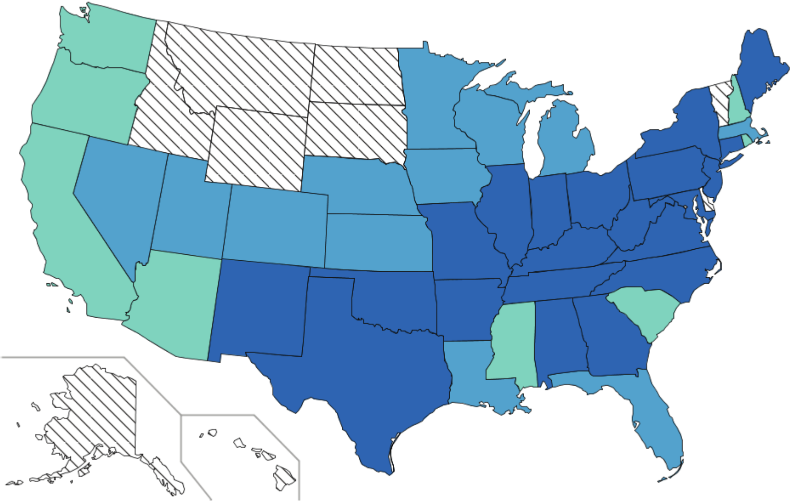
THE UNIVERSITY OF MELBOURNE

July 16, 2018

# Location Lost in Translation

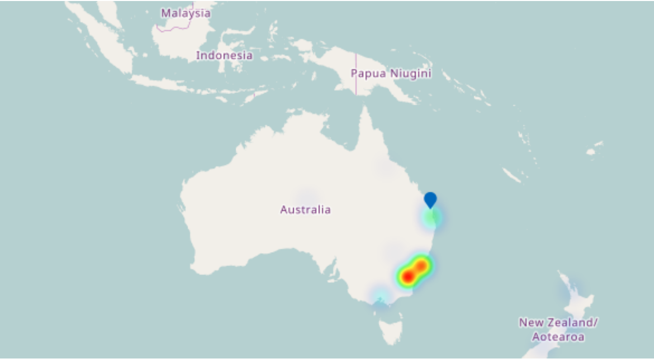


## Applications: Public Health Monitoring

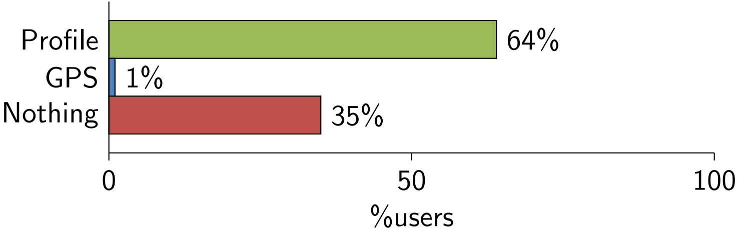


**Allergy Rates** (Paul and Dredze, 2011)

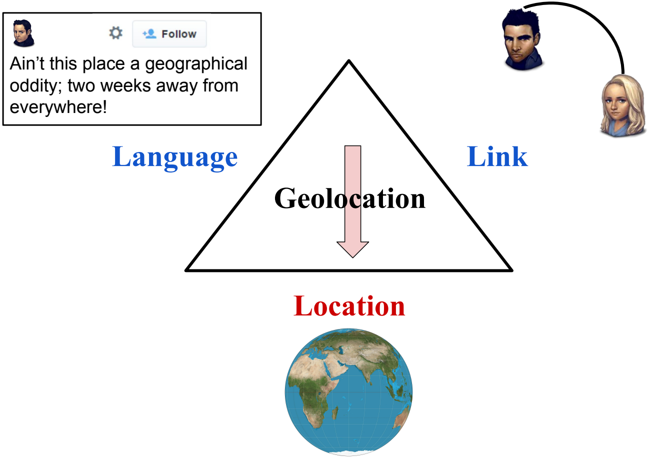
# Applications: Emergency Situation Awareness: Bushfires, Floods and Earthquakes




**Fight bushfire with #fire:** **Alert hospital before anybody calls**
(Cameron et al., 2012)




Profile field is noisy (Hecht et. al, 2011), GPS data is scarce (Hecht and Stephens, 2014), and biased toward younger urban users (Pavalanathan and Eisenstein, 2015)

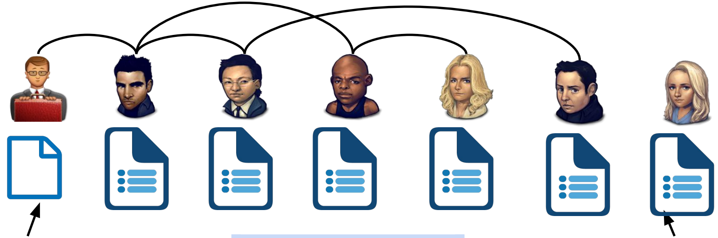



User geolocation is the task of identifying the "home" location of a social media user using contextual information such as geographical variation in **language use** and in **social interactions**.

- Huge amounts of unlabelled data, little labelled data
- Multiple views of Data: Text, Network

| Text-based Supervised Classification |
|---|
| Backstrom et al. (2008) |
| Cheng et al. (2010) |
| Wing and Baldridge (2011, 2014) |

| Network-based Semi-supervised Regression |
|---|
| Backstrom et al. (2010) |
| Davis Jr et al. (2011) |
| Jurgens (2013) |



No Text

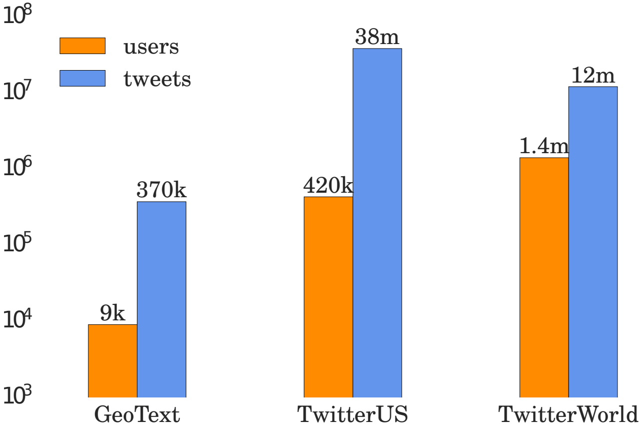
No Network

| Joint/Hybrid Text+Network |
|---|
| Rahimi et al. (2015) |
| Do et al. (2017) |
| Miura et al. (2017) |

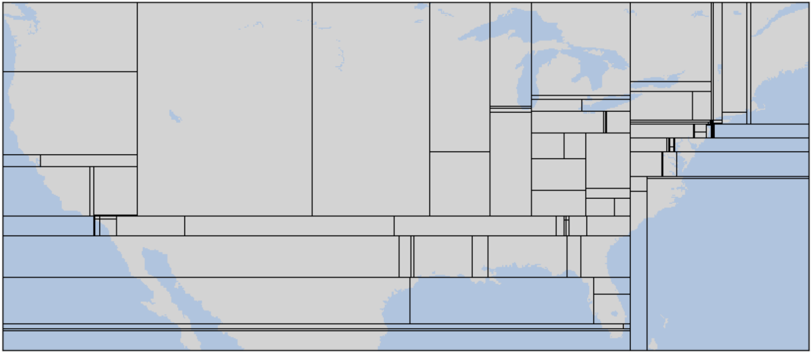
Don’t utilise unlabelled text data

**Our work:** Text+Network Semi-supervised Geolocation

# Twitter Geolocation Datasets



# Discretisation of Labels

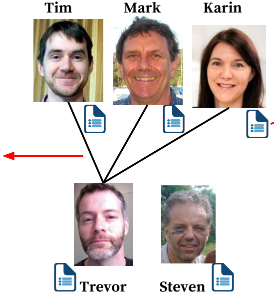


- Cluster continuous lat/lon: cluster ids are labels.
- Use the median training point of the predicted region as the final continuous prediction.
- **Evaluate** using Mean and Median errors between the known and the predicted coordinates.

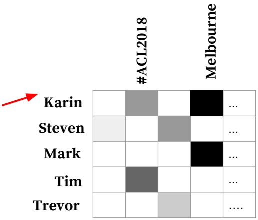
# Text and Network Views of Data

| | Karin | Mark | Steven | Tim | Trevor |
|---|---|---|---|---|---|
| **Karin** | 0.5 | 0 | 0 | 0 | 0.5 |
| **Mark** | 0 | 0.5 | 0 | 0 | 0.5 |
| **Steven** | 0 | 0 | 1 | 0 | 0 |
| **Tim** | 0 | 0 | 0 | 0.5 | 0.5 |
| **Trevor** | 0.25 | 0.25 | 0 | 0.25 | 0.25 |

**Normalised Adj. Matrix: A**




**@-mention Graph**



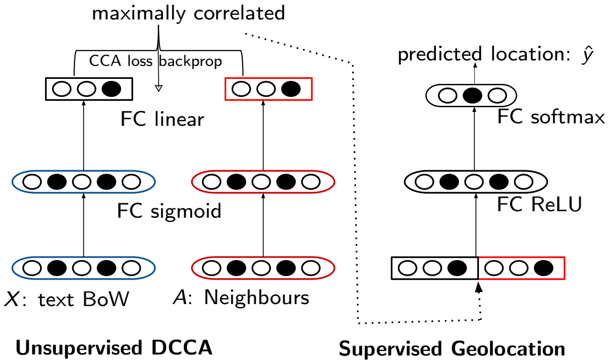

**Text BoW: X**

Two users are connected if they have a common @-mention.

## Baseline 1: FeatConcat

- Concatenate $A$ and $X$, and feed them to a DNN: $Y = f([X, A])$
- The dimensions of $A$, and consequently the number of parameters grow with the number of samples.

# Baseline 2: DCCA

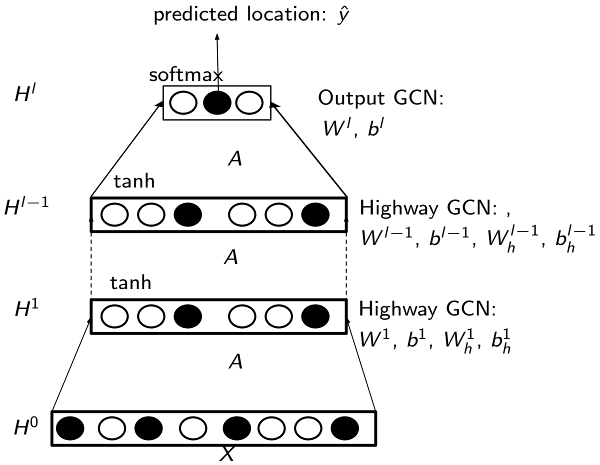



Learn a shared representation using Deep Canonical Correlation Analysis (Andrew et al., 2013):

$$\rho = \mathrm{corr}(f_1(X), f_2(A)) = \frac{\mathrm{cov}(f_1(X), f_2(A))}{\sqrt{\mathrm{var}(f_1(X)).\mathrm{var}(f_2(A))}}$$

$$Y = f\left([f_1(X), f2(A)]\right)$$

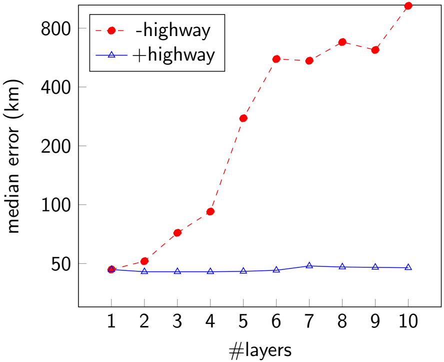
# Proposed Model: GCN




GCN Layer: $H^{(l+1)} = \mathrm{ReLU}\left(AH^{(l)}W^{(l)} + b\right)$

Adding more layers results in expanded neighbourhood smoothing: control with highway gates $W_h^l$, $b_h^l$

# Highway GCN: Control Neighbourhood Smoothing



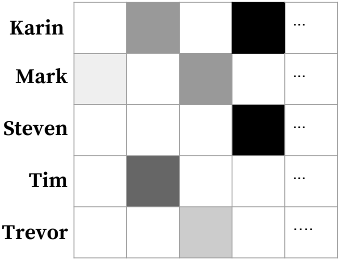

$$\text{layer gates:} \quad T(\vec{h}^l) = \sigma\left(W_h^l \vec{h}^l + b_h^l\right)$$

$$\text{layer output:} \quad \vec{h}^{l+1} = \underbrace{\vec{h}^{l+1} \circ T(\vec{h}^l) + \vec{h}^l \circ (1 - T(\vec{h}^l))}_{\text{weighted sum of layer input and output}}$$

| | Karin | Mark | Steven | Tim | Trevor |
|---|---|---|---|---|---|
| **Karin** | 0.5 | 0 | 0 | 0 | 0.5 |
| **Mark** | 0 | 0.5 | 0 | 0 | 0.5 |
| **Steven** | 0 | 0 | 1 | 0 | 0 |
| **Tim** | 0 | 0 | 0 | 0.5 | 0.5 |
| **Trevor** | 0.25 | 0.25 | 0 | 0.25 | 0.25 |

**Normalised Adj. Matrix: A**

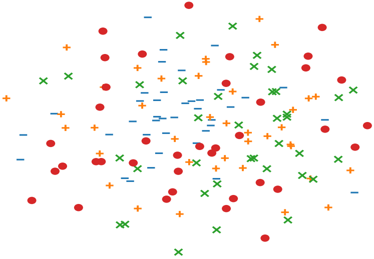
$\times$



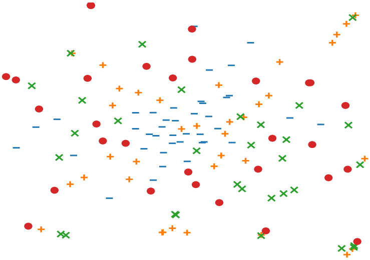

**Text BoW: X**

Smoothing immediate neighbourhood: $A \cdot X$
smoothing expanded neighbourhood: $A \cdot A \cdot X$

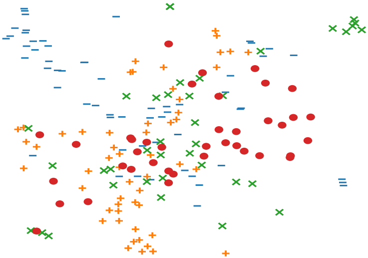
# Sample Representation using t-SNE



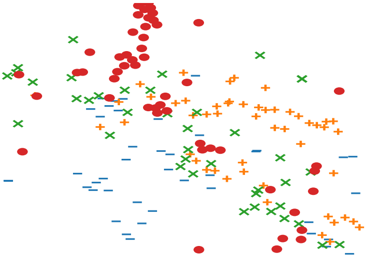
FeatConcat $[X, A]$



DCCA

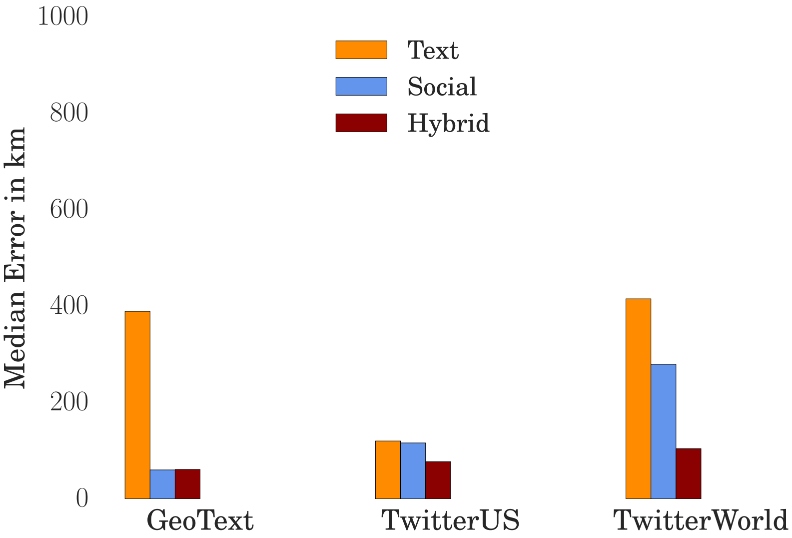


1 GCN $A \cdot X$

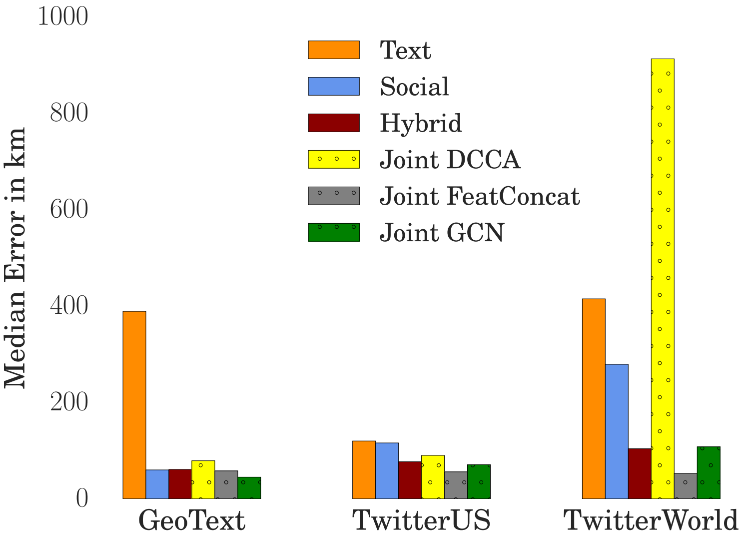


2 GCN $A \cdot A \cdot X$

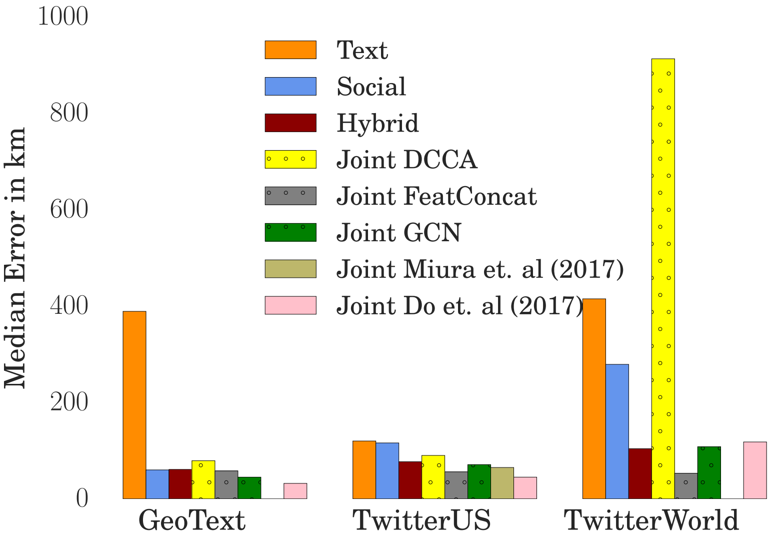
# Test Results: Median Error



# Test Results: Median Error



## Test Results: Median Error

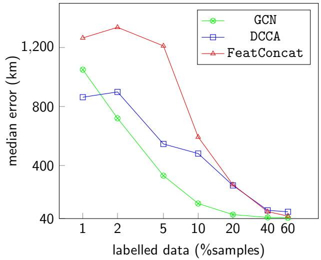


## Top Features Learnt from Unlabelled Data (1% Supervision)

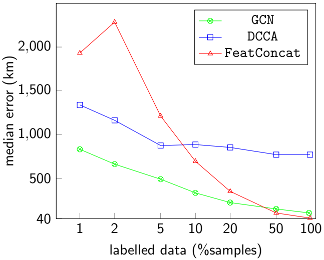
| Seattle, WA | Austin, TX | Jacksonville, FL | Columbus, OH |
|---|---|---|---|
| #goseahawks | stubb | unf | laffayette |
| smock | gsd | ribault | #weareohio |
| traffuck | #meatsweats | wahoowa | #arcgis |
| ferran | lanterna | wjct | #slammin |
| promissory | pupper | fscj | #ouhc |
| chowdown | effaced | floridian | #cow |
| ckrib | #austin | #jacksonville | mommyhood |
| #uwhuskies | lmfbo | #mer | beering |

Top terms for a few regions detected by `GCN` using only 1% of TWITTER-US for supervision. The terms that existed in labelled data are removed.

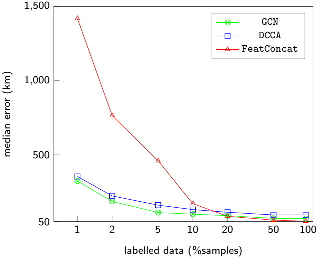
# Dev. Results: How much labelled data do we really have?



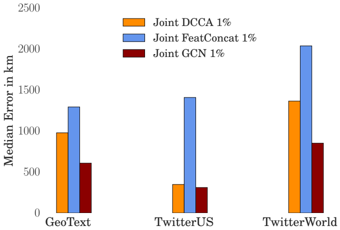

GeoText




Twitter-US

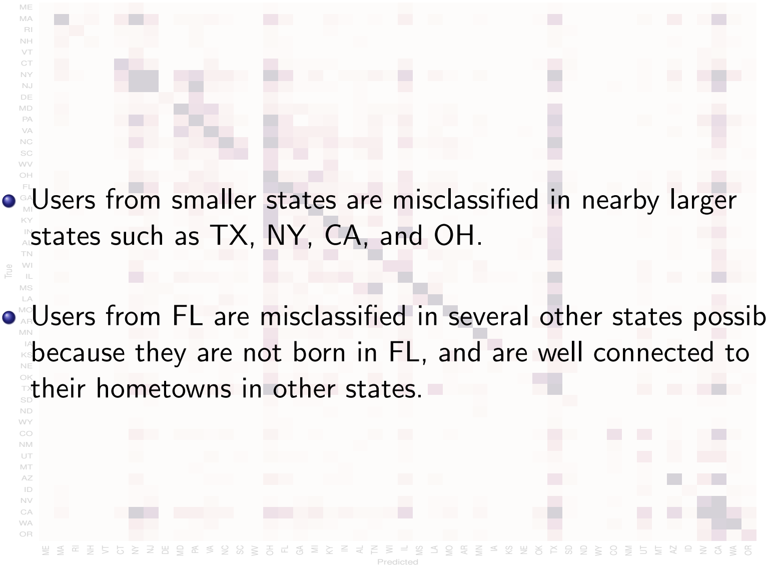



Twitter-World




Test results with 1% labelled data

# Confusion Matrix Between True Location and Predicted Location




- Users from smaller states are misclassified in nearby larger states such as TX, NY, CA, and OH.
- Users from FL are misclassified in several other states possibly because they are not born in FL, and are well connected to their hometowns in other states.

# Conclusion

- Simple concatenation in `FeatConcat` is a strong baseline with large amounts of labelled data.
- `GCN` performs well with both large and small amounts of labelled data by effectively using unlabelled data.
- Gating mechanisms (e.g. highway gates) are essential for controlling neighbourhood smoothing in `GCN` with multiple layers.
- The models proposed here are applicable to other demographic inference tasks.

*Thank you!*

Code available at:
https://github.com/afshinrahimi/geographconv