Simulation and mining of social networks in a pandemic context

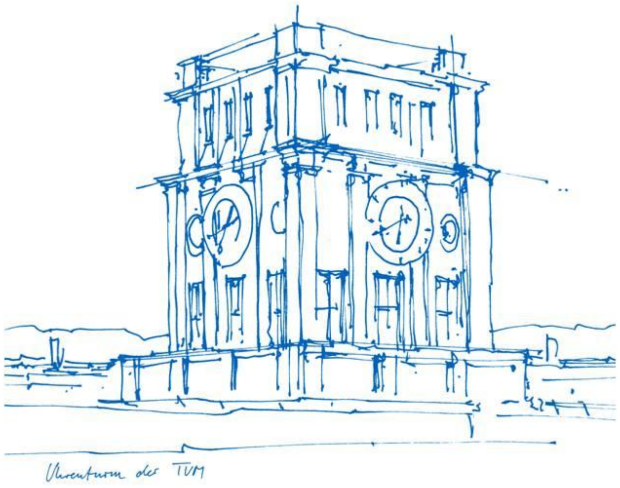
Aldi Topalli, Miriam Anschütz

Technische Universität München

Faculty of Informatics

Chair of Social Computing

Garching, 19.04.2021



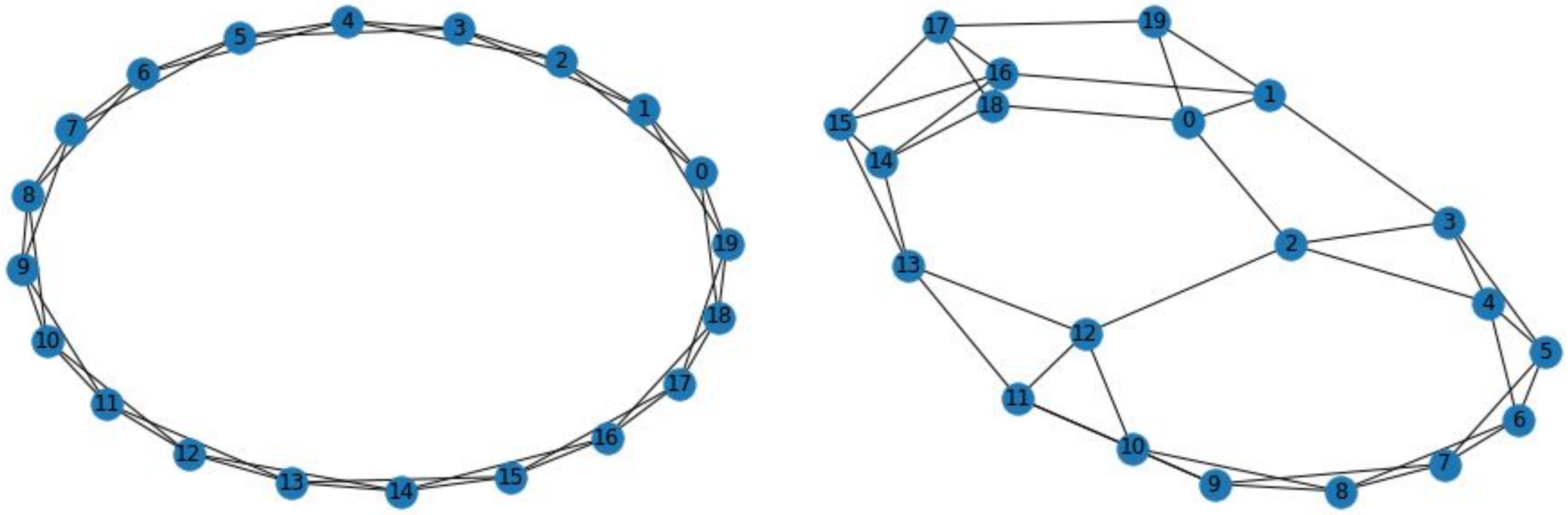
Outline

- Generation of long-term social network
- Individual risk prediction
- Action-based risk prediction

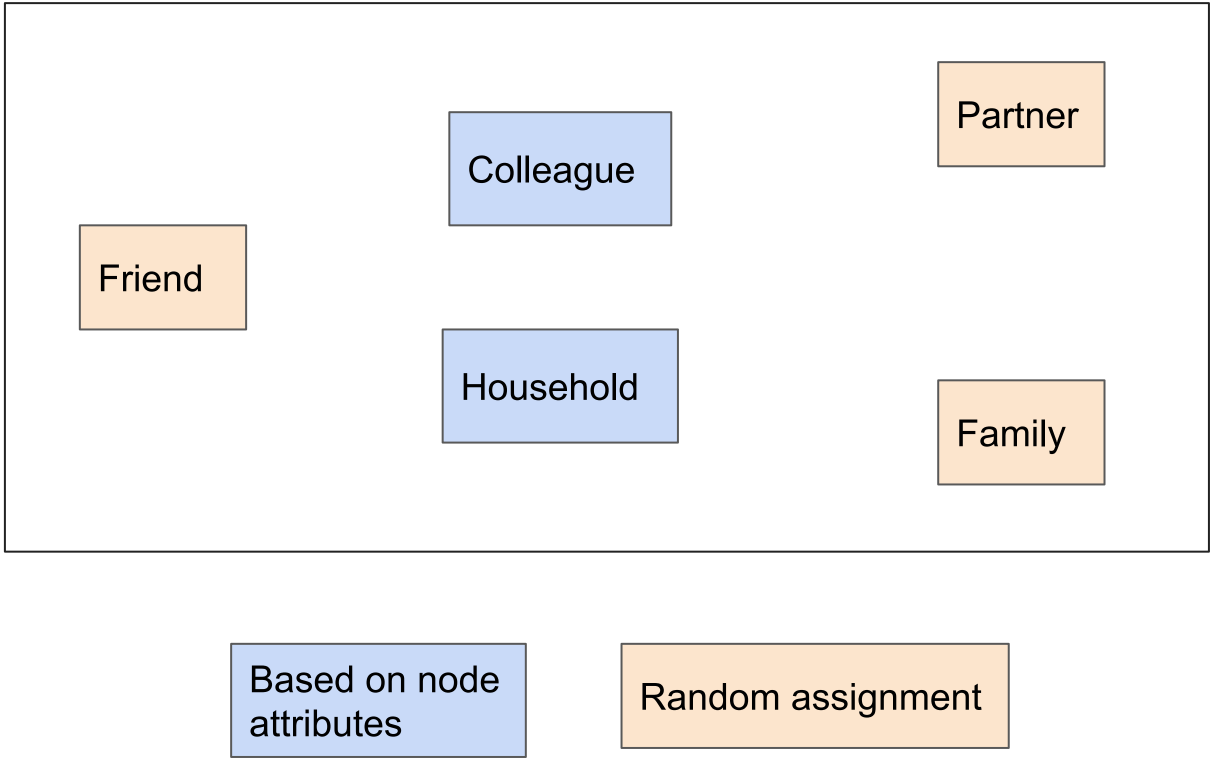
Objective

- Generate a real-world graph
- Model long-term relationships
- Store work and home facility

Watts-Strogatz model

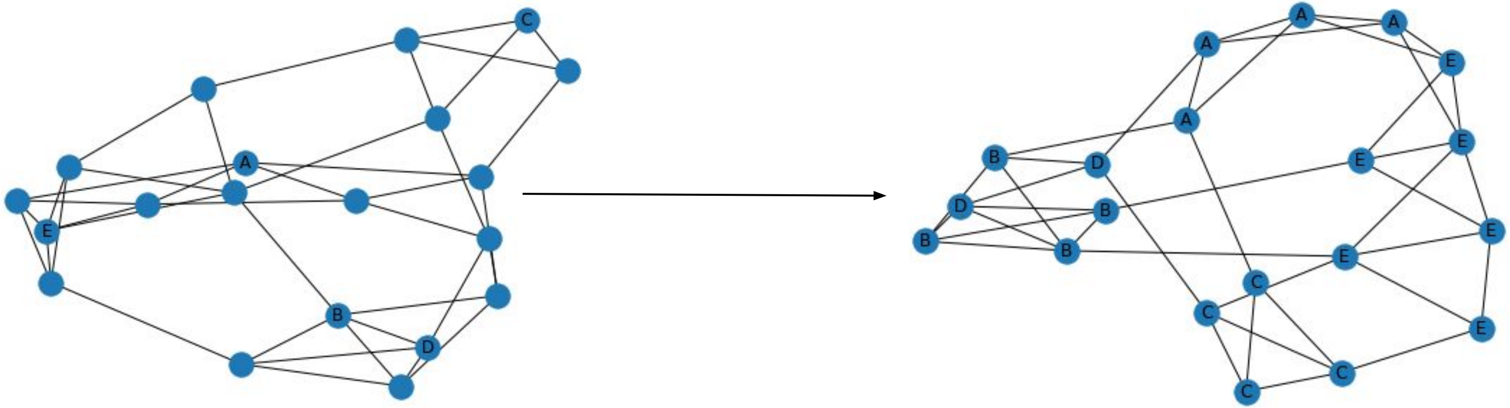


Edge types


Partner
Colleague
Friend
Household
Family
Based on node attributes
Random assignment

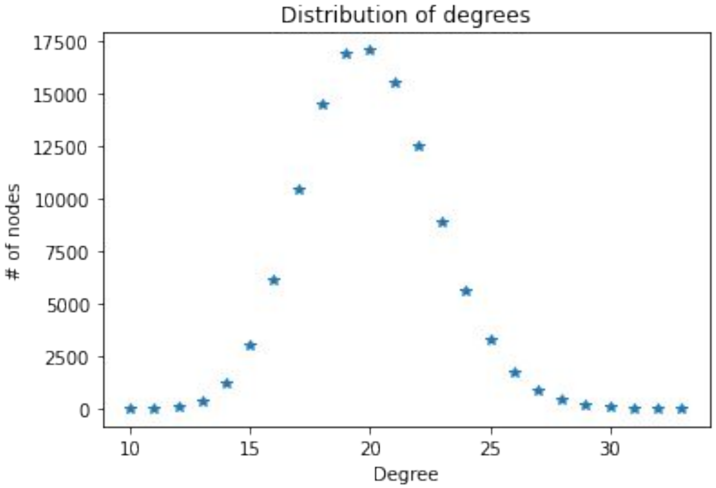
Node classification [1]

- Assign given facility to a random node
- Propagate attribute: Close nodes should have similar attributes

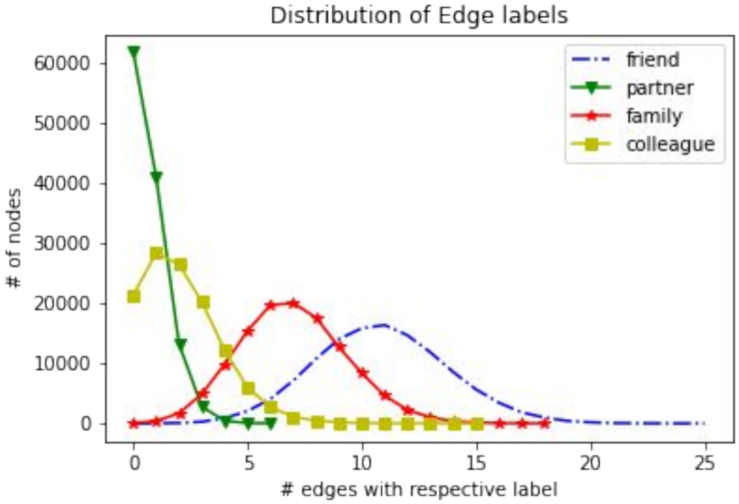


[1] Zhou, D., Bousquet, O., Lal, T., Weston, J. and Schölkopf, B., 2003. Learning with local and global consistency. *Advances in neural information processing systems*, *16*, pp.321-328.

Results

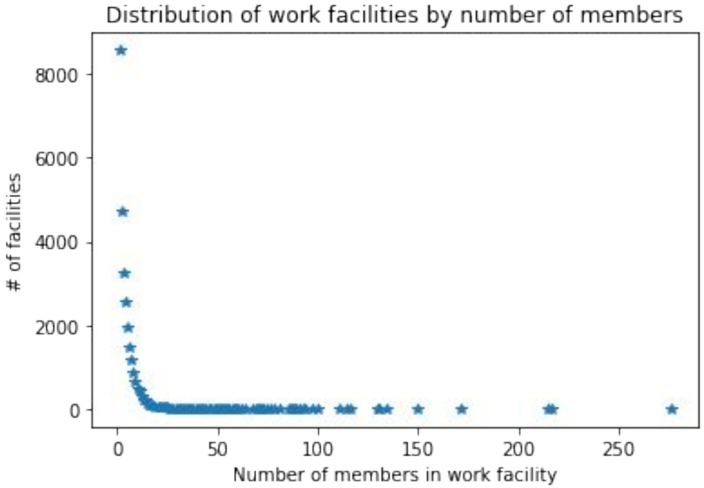


Distribution of degrees
17500
15000
12500
10000
7500
5000
2500
0
# of nodes
10
15
20
25
30
Degree

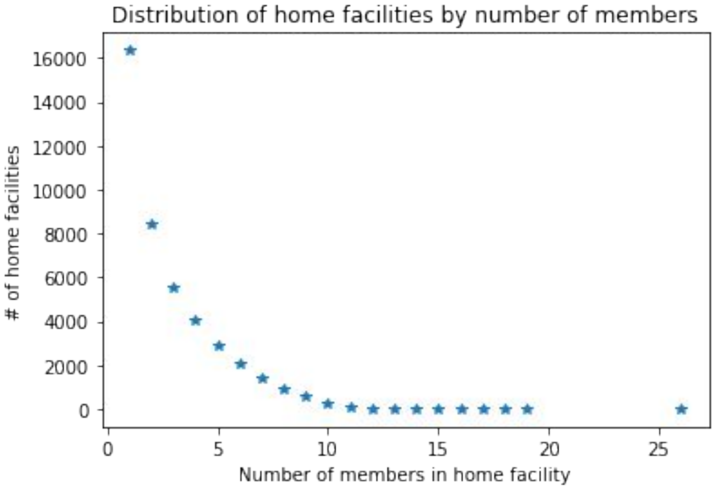


Distribution of Edge labels
friend
partner
family
colleague
60000
50000
40000
30000
20000
10000
0
# of nodes
0
5
10
15
20
25
edges with respective label

Results



Distribution of work facilities by number of members
# of facilities
Number of members in work facility
0
2000
4000
6000
8000
0
50
100
150
200
250



Distribution of home facilities by number of members
# of home facilities
Number of members in home facility
0
2000
4000
6000
8000
10000
12000
14000
16000
0
5
10
15
20
25

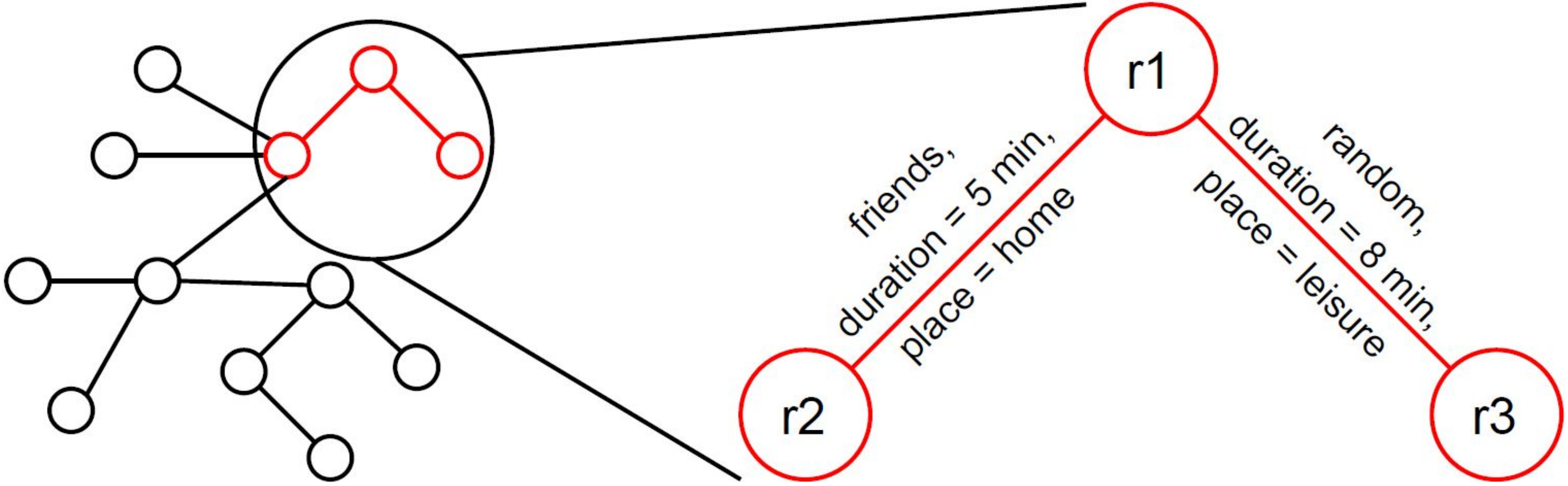
Outline

- Generation of long-term social network
- Individual risk prediction
- Action-based risk prediction

Objective: Extend the German Corona Warn app

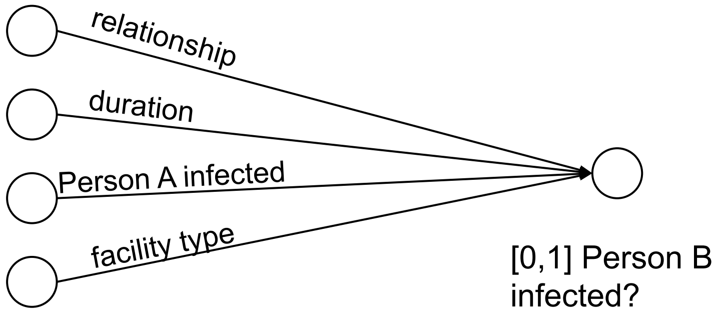
- Encounter-based live prediction
- Continuous instead of binary risk

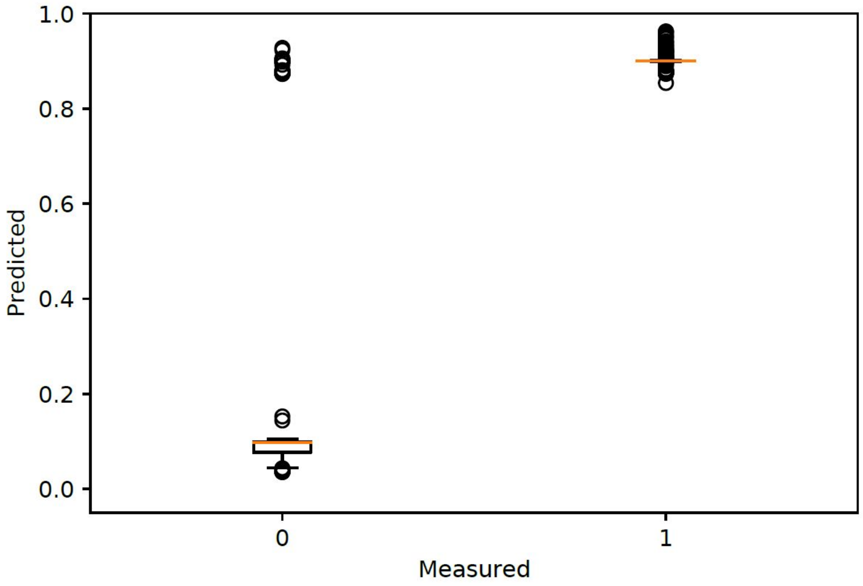
Encounter network



Encounter risk with regression

Person A meets Person B:

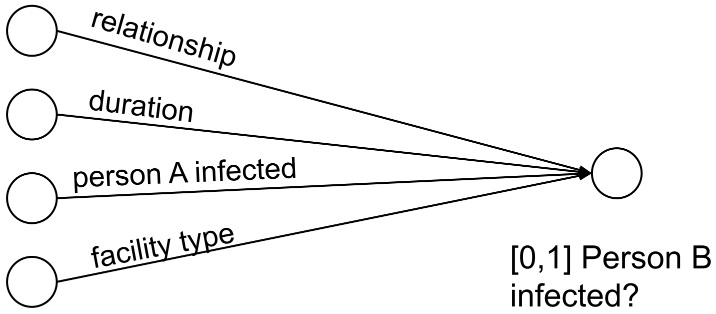




MAE: 0.12

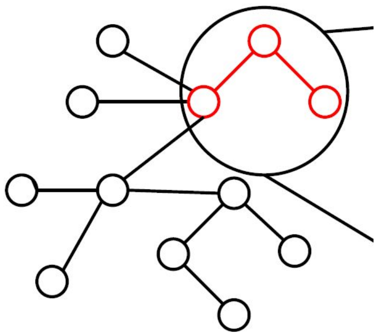
Discussion: Important features

Person A meets Person B:



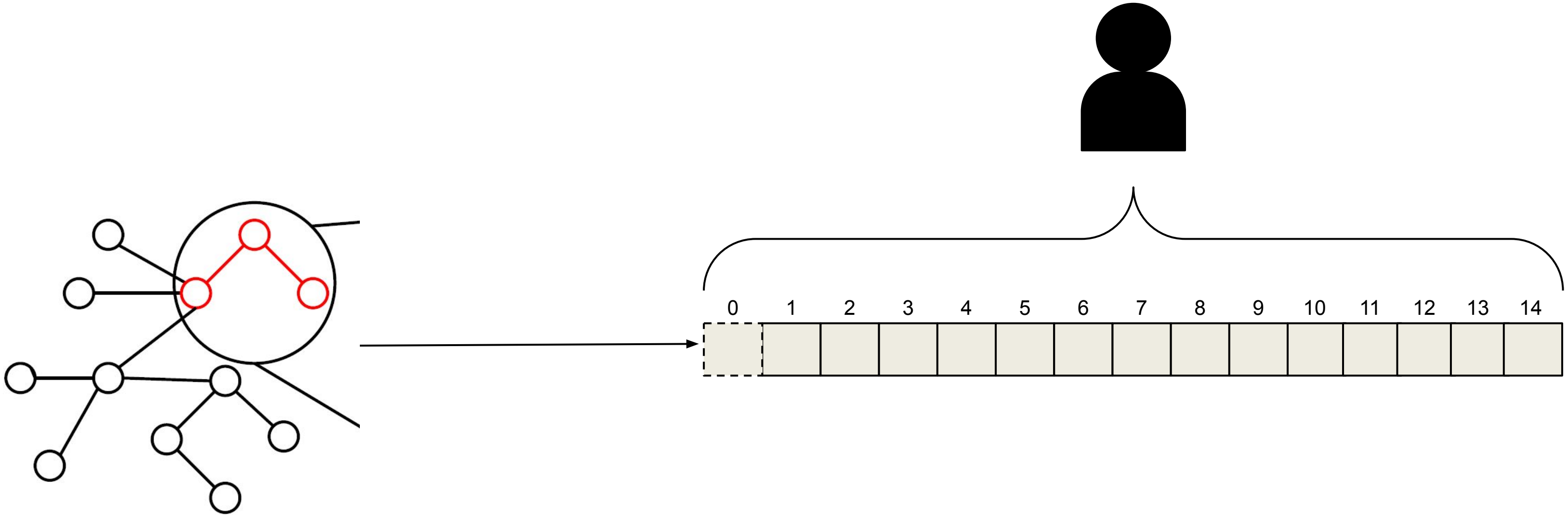
Feature left out	Mean absolute error
None	0.12
relationship	0.12
encountered person's infect risk	0.35
facility type	0.12
duration	0.12
all except person's infect risk	0.13

Predicting the daily risk

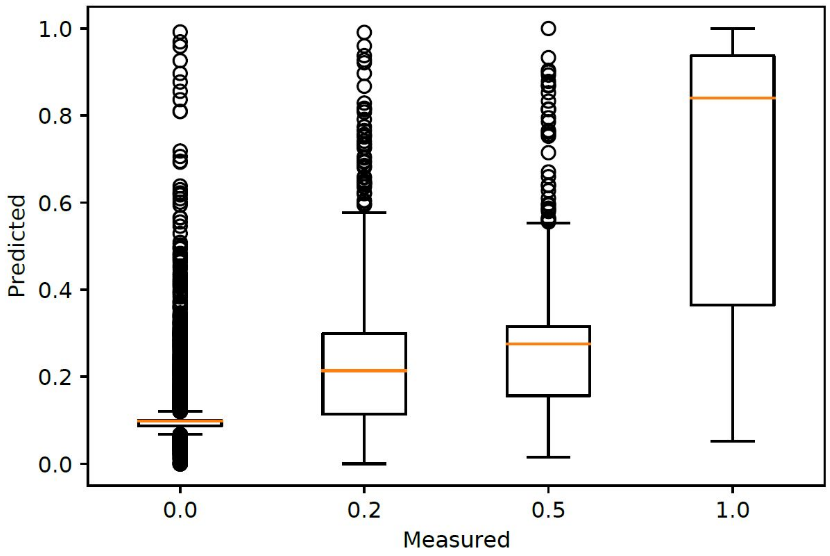


$$r_p = max\left(1, \sum_{i=1}^{n} w_i\right)$$

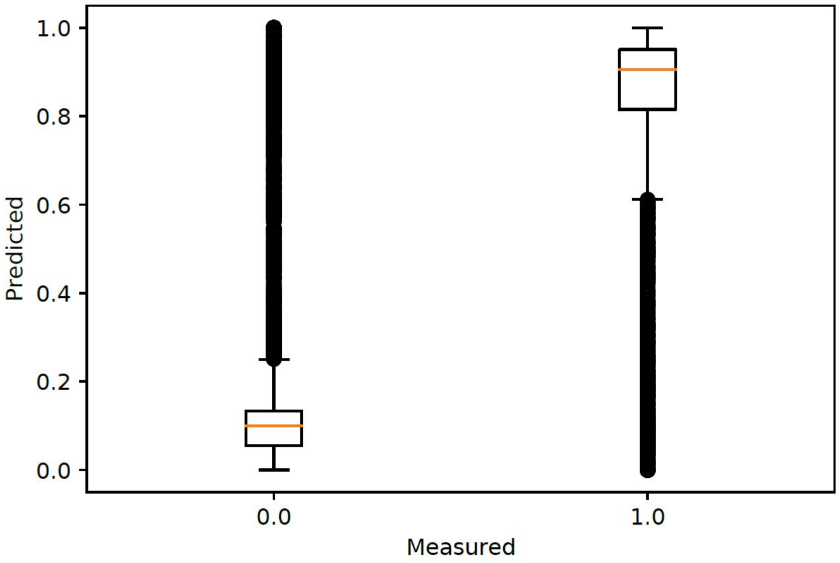
14 day risk aggregation



Results

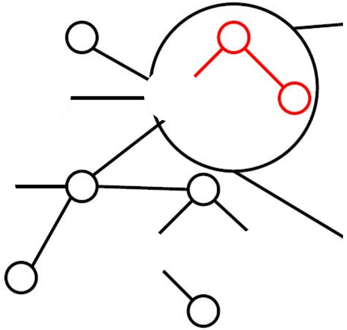


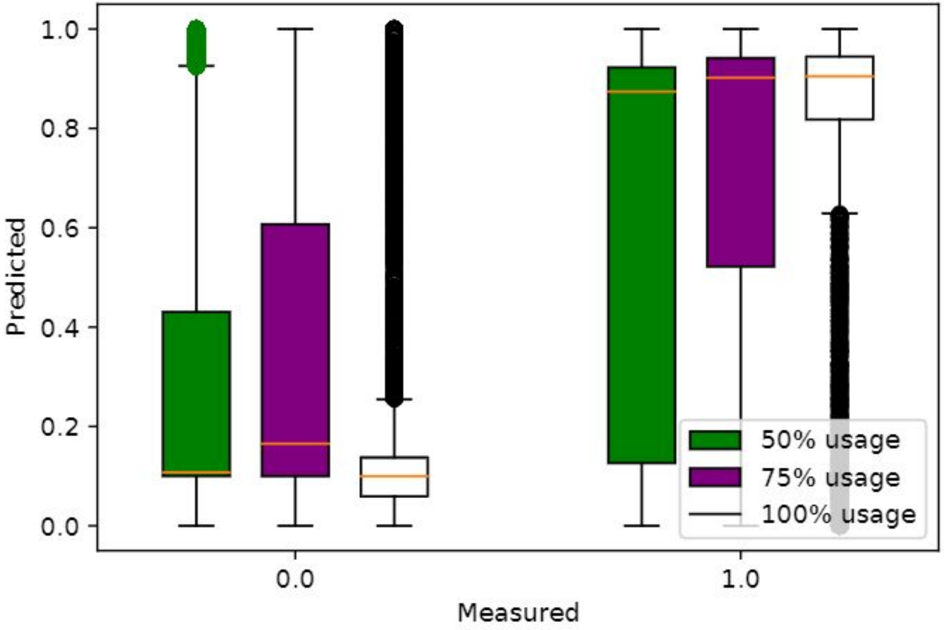
1.0
0.8
0.6
0.4
0.2
0.0
Predicted
0.0
0.2
0.5
1.0
Measured



1.0
0.8
0.6
0.4
0.2
0.0
Predicted
0.0
1.0
Measured

Discussion: App usage in population





Outline

- Generation of long-term social network
- Individual risk prediction
- Action-based risk prediction

Daily “sentence”

Home → Bus → Work → Supermarket → Bus → Home

Daily “sentence”

Home → Bus → Work → Supermarket → Bus → Home

↓

Daily Risk of infection

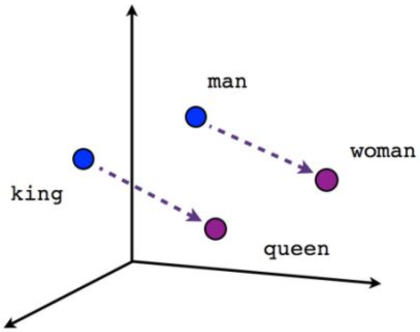
Actions

Each action has

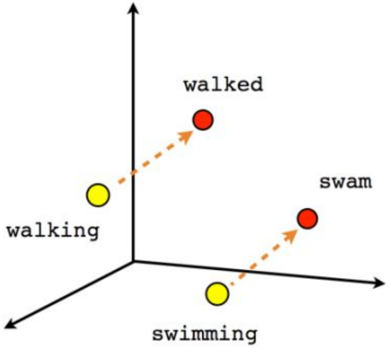
- A start time
- A group size
- A duration
- A facility (location)

Enter high dimensional space

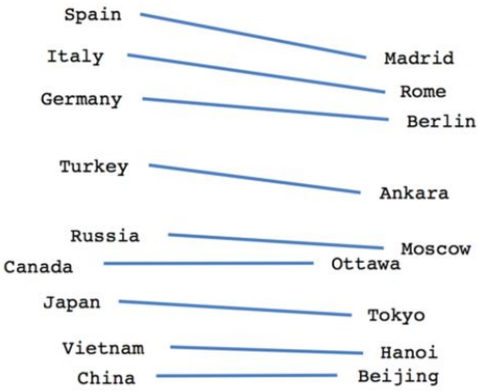
Map actions to an embedding space



Male-Female



Verb tense



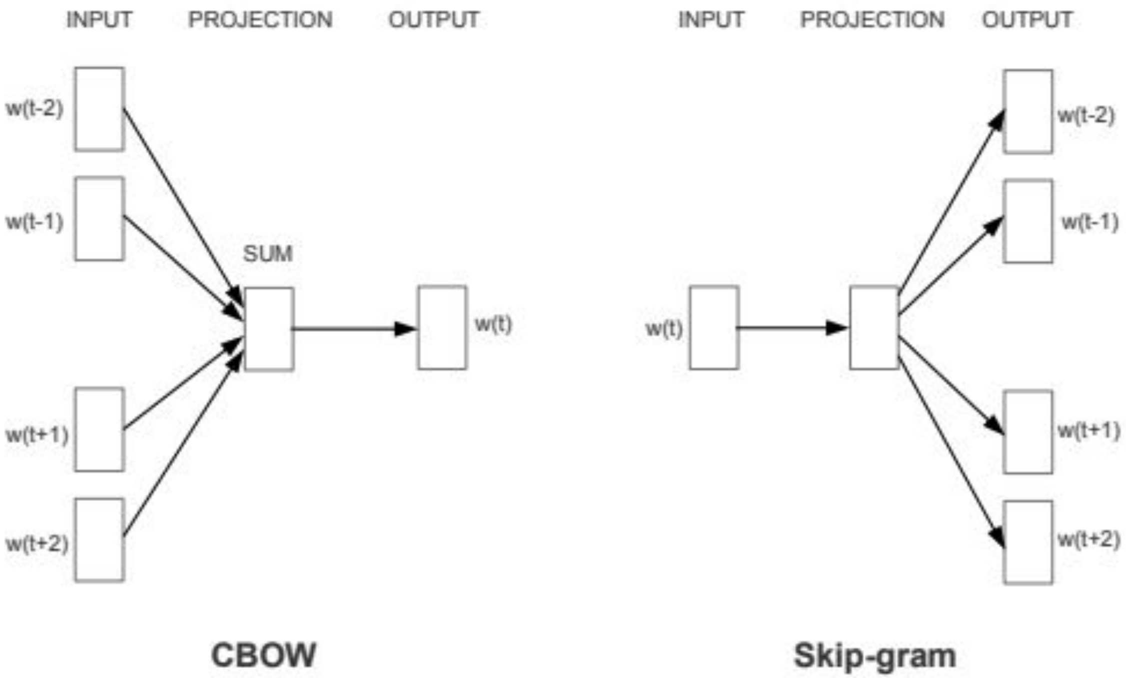
Country-Capital

Enter high dimensional space

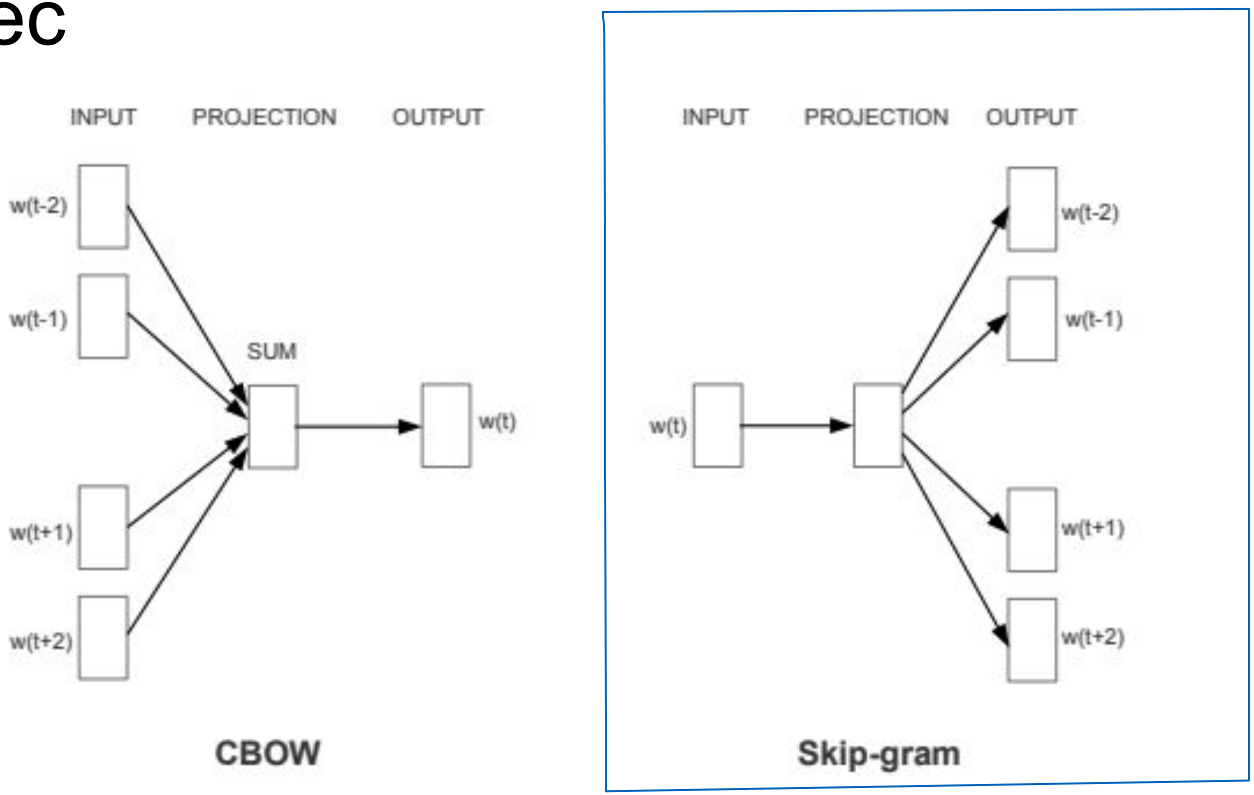
Map actions to an embedding space

- Pull similar actions closer together
- Based on the surrounding actions (context)
- Weighted Word2Vec

Word2Vec


INPUT
PROJECTION
OUTPUT
w(t-2)
w(t-1)
SUM
w(t)
w(t+1)
w(t+2)
CBOW
INPUT
PROJECTION
OUTPUT
w(t)
w(t-2)
w(t-1)
w(t+1)
w(t+2)
Skip-gram

Word2Vec


INPUT
PROJECTION
OUTPUT
w(t-2)
w(t-1)
SUM
w(t)
w(t+1)
w(t+2)
CBOW
INPUT
PROJECTION
OUTPUT
w(t)
w(t-2)
w(t-1)
w(t+1)
w(t+2)
Skip-gram

Training objective

Word2Vec

$$C(\theta) = \frac{1}{T} \sum_{t=1}^{T} \sum_{-k \leq j \leq k, j \neq 0} L(w_t, w_{t+j}; \theta)$$

Time2Vec

$$C(\theta) = \frac{1}{T} \sum_{t=1}^{T} \sum_{c_j \in S_t} \alpha(c_t, c_j) L(c_t, c_j)$$

Training objective

Word2Vec

$$C(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{-k\leq j\leq k, j\neq 0} L(w_t, w_{t+j}; \theta)$$

L is the loss when we predict word w_{t+j} is in the context of w_t

Time2Vec

$$C(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{c_j\in S_t} \alpha(c_t, c_j) L(c_t, c_j)$$

Training objective

Word2Vec

$$C(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{-k\leq j\leq k, j\neq 0} L(w_t, w_{t+j}; \theta)$$

Time2Vec

$$C(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{c_j\in S_t} \alpha(c_t, c_j) L(c_t, c_j)$$

Actions similarity

- Start time difference
- Duration difference
- Group Size difference
- Facility name used as concept name

Actions similarity

- Start time difference
- Duration difference
- Group Size difference
- Facility name used as concept name

Steps

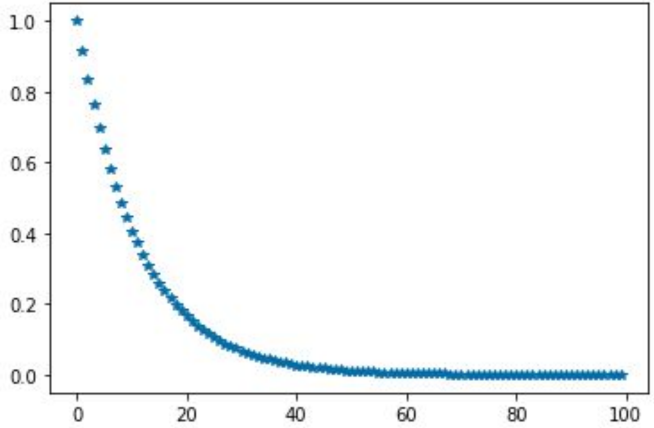
1. Build a vocabulary of concepts (map to indices)
2. Create pairs
3. Compute weights
4. Create the train file
5. Train embedding
6. Clustering and risk prediction

Inter-Action Weights

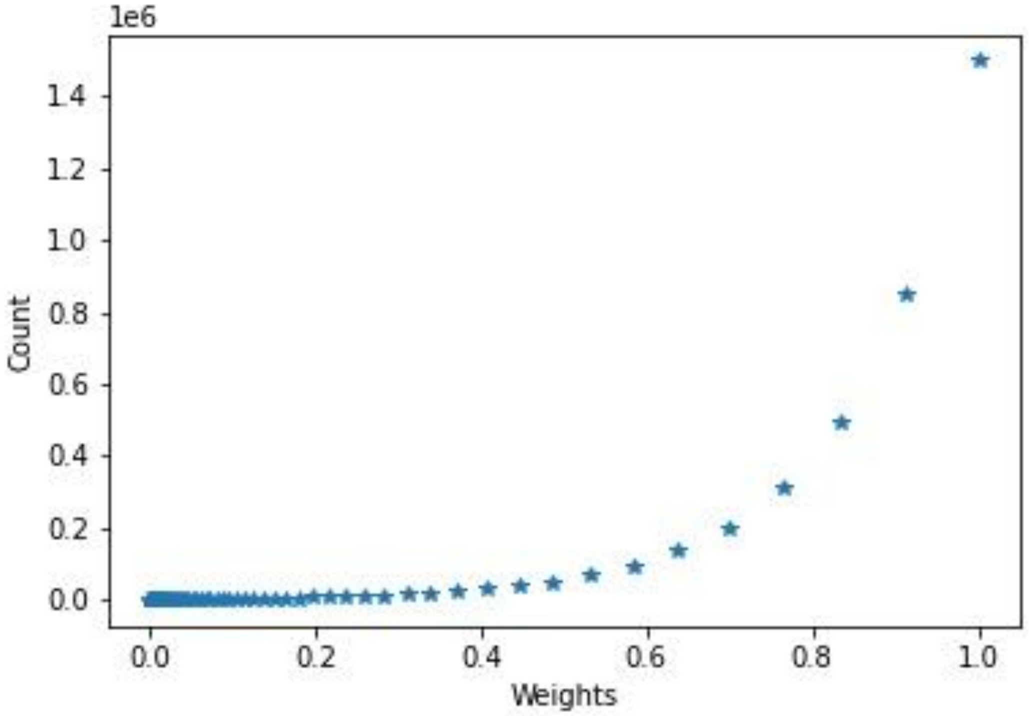
	input	target	time_difference	duration_difference	group_size_difference
0	1234	5746	30558.0	25018.0	29.0
1	5746	1234	30558.0	25018.0	29.0
2	1234	5747	35456.0	29066.0	30.0
3	5747	1234	35456.0	29066.0	30.0
4	5746	5747	4898.0	4048.0	1.0

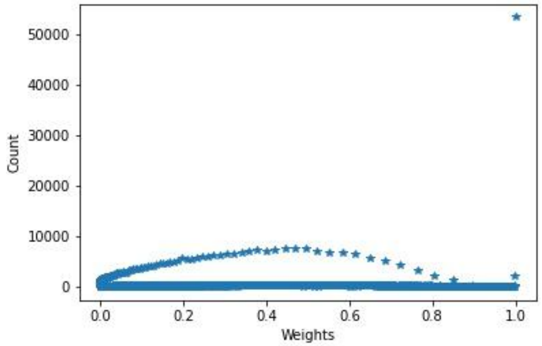
Differences

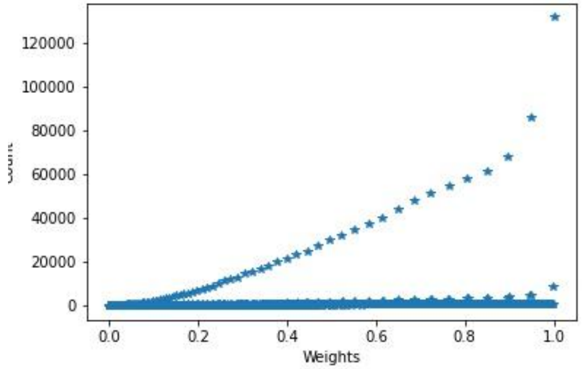
Exponential decay



Apply exponential decay



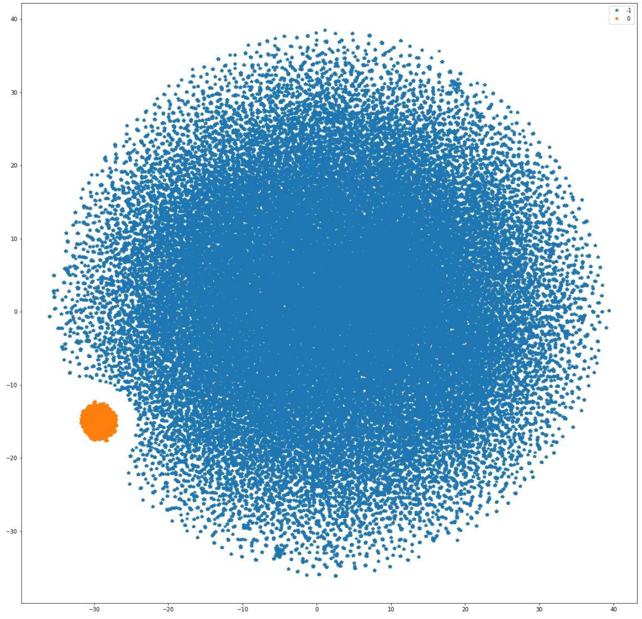


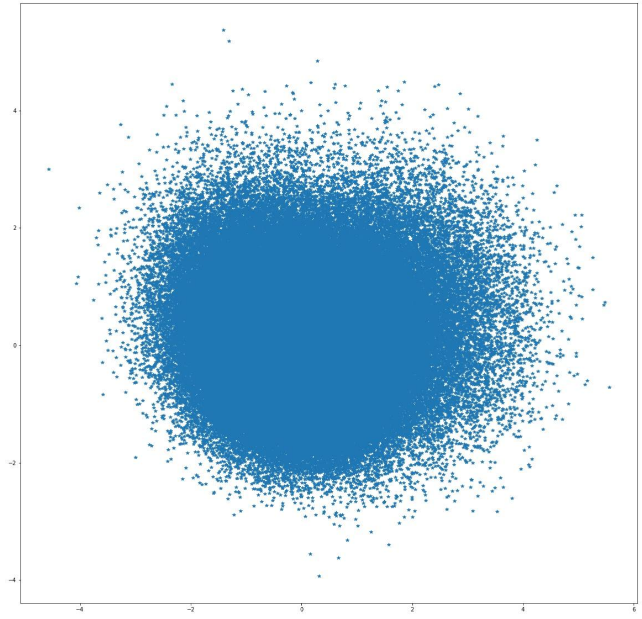


Train embeddings

- ~211.000 concepts (facilities)
- Training time: 4 days (on CPU)
- Embedding dimensions: 50

Clustering (T-SNE and PCA)

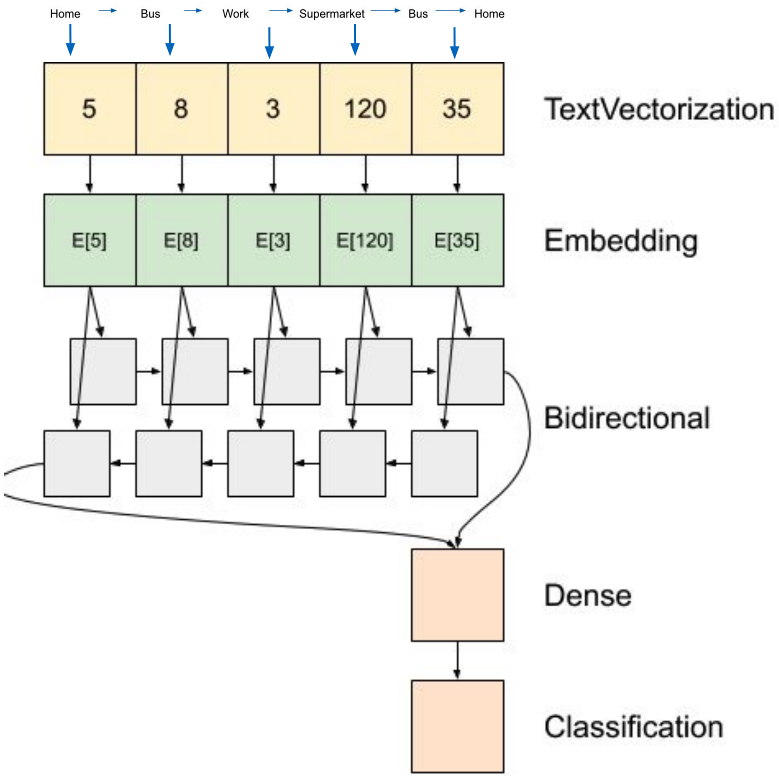




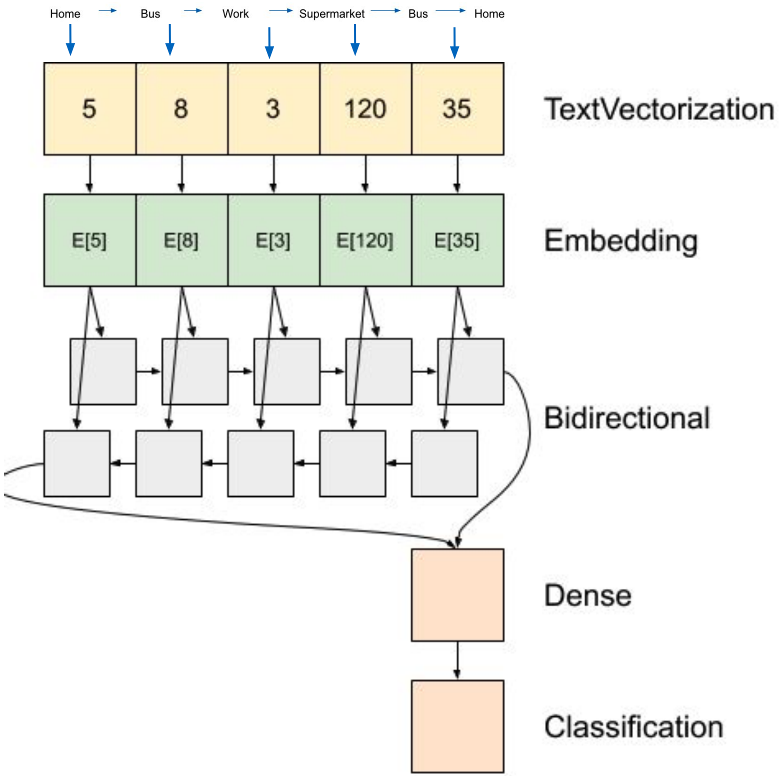
Risk prediction

- Structure actions of a person as sentence Home → Bus → Work → Supermarket → Bus → Home
- Train a model to predict risk of infection given a sentence
  - Binary classification
  - Similar to sentiment analysis (Prediction of +/- sentiment)

Models tried



Models tried



Models

- Trained with 1 BiLSTM (10 epochs)
- Trained with 2 BiLSTM (10 epochs)

Results

- Training remained at random (~50%)
  - No clear patterns in the data
  - Both models performed same

Conclusion

- The whole process needs to be redone → to improve performance

End