# Political Stance Prediction Based on Online Information Diffusion

Pilli Yogesh Kumar

**Laboratory Hubert Curien - University of Sydney**

*Advisors:- Christine Largeron Andrei-Marian Rizoiu*

7th July 2020

# Overview

1. State of Art
2. Dataset
3. Social Network Analysis
   - Illustration of Comment Tree
   - Notion of Edge Homogeneity
   - Notion of Source Node and Target Node on smallest thread
4. Political Stance Predictor
   - Long Short Term Memory -Neural Network
5. Results
6. Conclusion and Future Work
7. QA

# State of Art

**Myers, Seth A., Chenguang Zhu, and Jure Leskovec. (2012) "Information diffusion and external influence in networks."**

**Information Diffusion**: Spreading of information from node to node along the edges of a social network

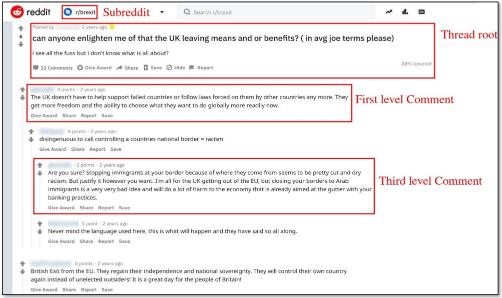
**Shuai, X., Ding, Y., Busemeyer, J., Chen, S., Sun, Y., Tang, J. (IJSWIS , 2012) - "Modeling indirect influence on twitter**

**Social Influence**: Process in which an individual's thoughts, feelings or actions are affected by other people

**Benoit, K. and Matsuo, A. (2018) - "Network analysis of Brexit discussion on social media."**

**Politial Stance Classifier**: Labels users Stance as AgainsBrexit, ProBrexit or Neutral.

# Data




- **Time Frame:** Nov'2015 - Apr'2019
- **Users:** 14,362 users
- **Threads:** 21,725 threads
- **Comments:** 207,894 comments

# Social Network Analysis

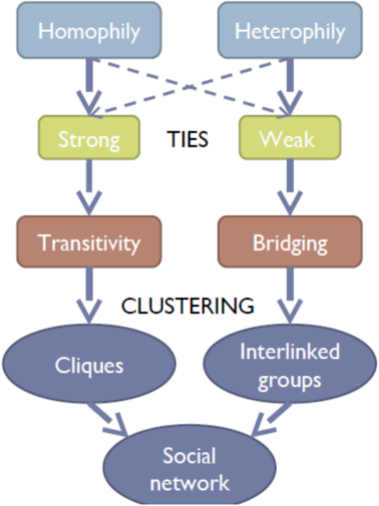
**Homophily**

It is tendency of people to form friendships with others with similar characteristics

**Social Influence**

It is tendency of people may modify their behaviors to bring them more closely into alignment with the behaviors of their friends.

# Social Network Analysis



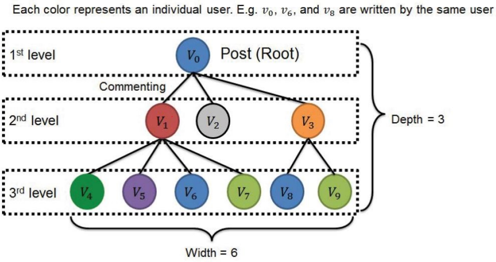

- **Transitivity** in SNA is a property of ties: if there is a tie between A and B and one between B and C, then in a transitive network A and C will also be connected
- **Bridges** are nodes and edges that connect across groups

# Social Network Analysis

- **Edge** can represent to connections, streams of data or social relations
- **Specifically for social relations, an intermediary's or the quality of a tie can be**
  - the recurrence of association (communication) or the measure of information exchange
  - different qualities of the nodes or ties (e.g., connections)
  - The structure of the nodes' neighborhood (e.g.many shared 'companions')

# Social Network Analysis

## Illustration of Comment Tree




- We define Comment Tree as an undirected Tree, T = (V,E), where V is the set of all messages, which includes the original post(root) and follow-up comments in the thread, and E is the set of edges, each of which connects two messages that are linked by comment

# Social Network Analysis

**Notion of Edge Homogeneity**

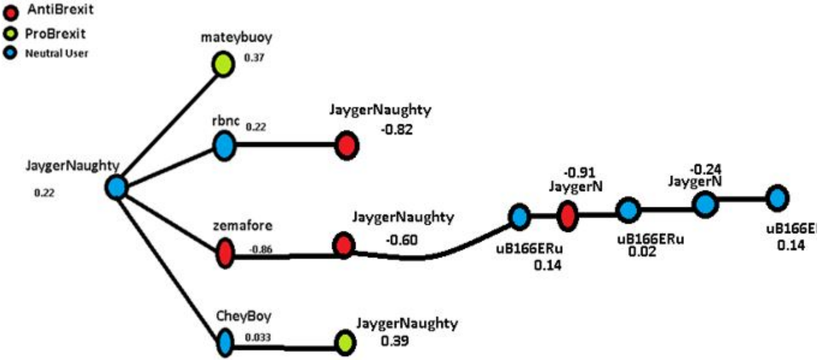
We define user polarization $\sigma = 2p - 1$ , where $p$ is probability, which lies in [0,1] and hence $\sigma$ lies in between [-1,1]

From user polarization we define edge homogeneity , for any edge $e_{ij}$ between the nodes $i$ and $j$

$$\sigma_{ij} = \sigma_i \sigma_j \tag{1}$$

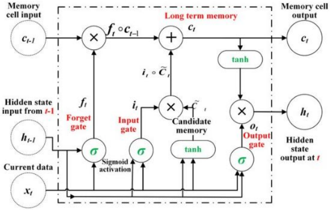
# Social Network Analysis

## Notion of Source Node and Target Node on smallest thread



# Political Stance Predictor

- LSTM is Variant of Recurrent Neural Network which introduces number of special and internal gates
- Internal gates help with the problem of learning relationships between long and short sequences in data
- **PRO:** Introduces many more internal parameters which must be learned - **Flexible**
- **CON:** Due to introducing many more internal parameters, It take time to learn - **Time Consuming**



# Political Stance Predictor

- Input gate $i$ :
  - Takes previous output $h_{t1}$ and current input $x_t$

$$i_t = \sigma(W_i \cdot [h_{t1}, x_t] + b_i)$$

- Forget gate $f$:
  - Takes previous output $h_{t1}$ and current input $x_t$

$$f_t = \sigma(W_f \cdot [h_{t1}, x_t] + b_f)$$

  - if $f_t = 0$: **Forget** previous state, otherwise pass through prev. state.
- Read gate $q$:
  - Takes previous output $h_{t1}$ and current input $x_t$

$$q_t = \tanh(W_q \cdot [h_{t1}, x_t] + b_q)$$

# Political Stance Predictor

- Memory Cell gate $c$:
  - New values depends on $f_t$, its previous state $c_{t-1}$ and read gate $q_t$

$$c_t = ftc_{t1} + i_t q_t)$$

- Output gate $o_t$ :
  - Will be fed as input into next block

$$o_t = \sigma(W_o \cdot [h_{t1}, x_t] + b_o)$$

$$h_t = o_t tanh(c_t)$$

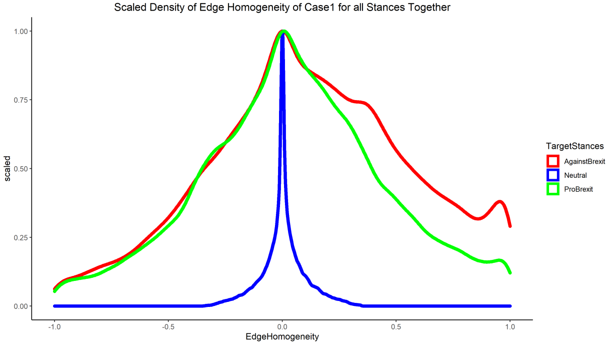
- **Intuition:**
  - We learn when to retain a state, or when to forget it.
  - Parameters are constantly updated as new data arrive

# Political Stance Predictor

- **LSTM without Textual Features**
  - We take current user stance label as LSTM input
  - We create X and Y pairs, Y sequence is created by one next shift in the sequence
  - We reshape the data in LSTM friendly 3D format as (samples, timesteps, features) use categorical crossentropy as loss function and softmax as activation function
- **LSTM with textual features**
  - We take textual features as input
  - We label each sequence as one user stance
  - We reshape the data in LSTM friendly 3D format as (samples, timesteps, features) use categorical crossentropy as loss function and softmax as activation function
- **Performance metrics is measures with F1 Score, Accuracy, Recall and Precision**

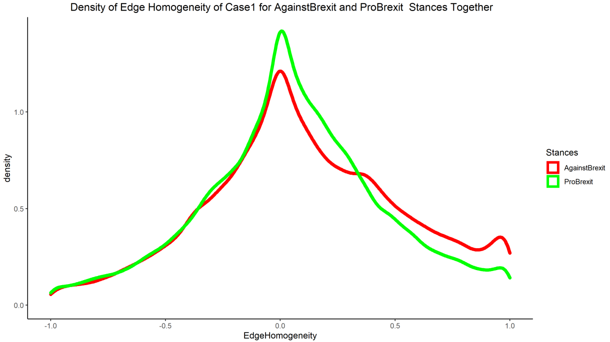
# Results

## Probability Density Function of Edge Homogeneity for Case1 (Source Node is Fixed and Target Node is Combined of All Stance)-All the Scenarios



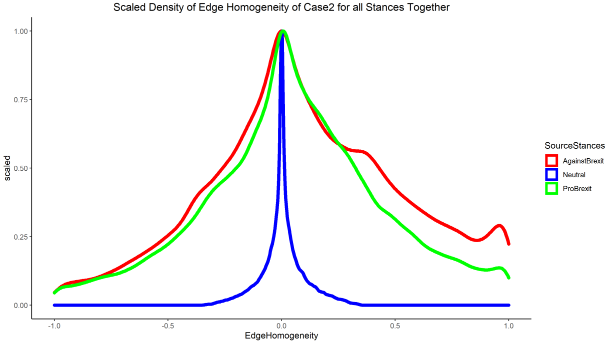
# Results

## Probability Density Function of Edge Homogeneity for Case1 (Source Node is Fixed and Target Node is Combined of All Stance)-with Against and Pro Brexit Stances



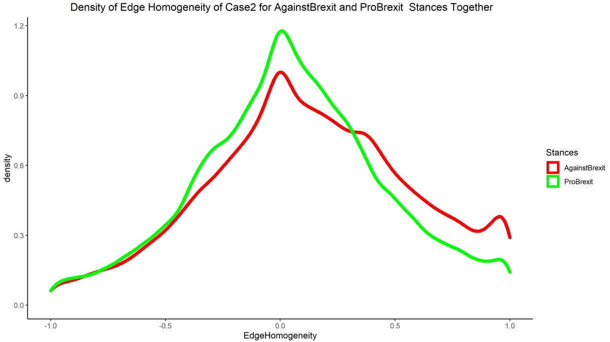
# Results

## Probability Density Function of Edge Homogeneity for Case2 (Target Node is Fixed and Source Node is Combined of All Stance)- combined with other scenarios



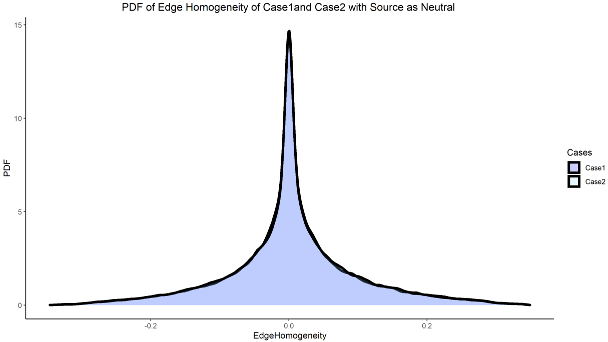
# Results

## Probability Density Function of Edge Homogeneity for Case2 (Target Node is Fixed and Source Node is Combined of All Stance)-with Against and Pro Brexit Stances



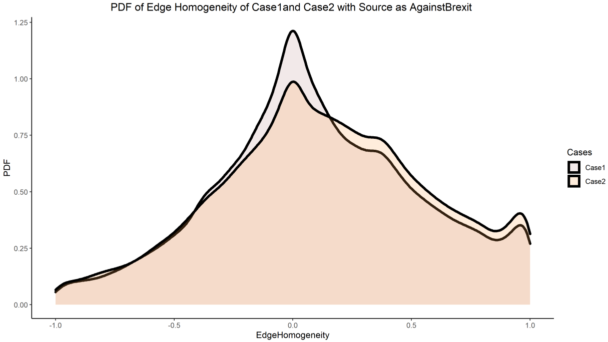
# Results

## Probability Density Function of Edge Homogeneity for Case1 and Case2 For Neutral Stance



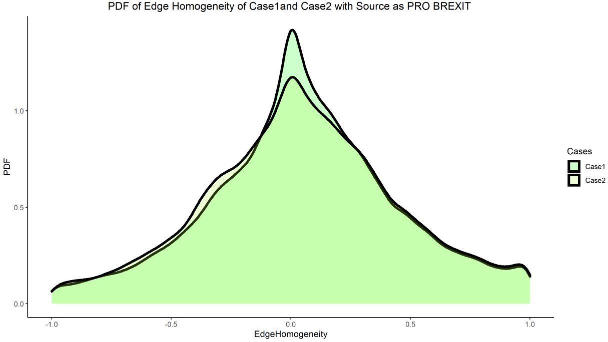
# Results

## Probability Density Function of Edge Homogeneity for Case1 and Case2 For Against Stance



# Results

## Probability Density Function of Edge Homogeneity for Case1 and Case2 For ProBrexit Stance



# Results

## Performance Evaluation of the Political Stance Predictor

Table: Confusion Matrix of Actual Stance and Predicted Stance from LSTM model

| *Actual —Predicted* | **Against** | **Neutral** | **Brexit** |
|---|---|---|---|
| **Against** | 3790 | 5544 | 3354 |
| **Neutral** | 5817 | 8174 | 4818 |
| **Brexit** | 3857 | 5238 | 3210 |

# Results

Table: Performance Metrics for LSTM Model without Textual Features used for predicting user's stance based on their submitted posts.

| **Set** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| **Test** | 0.350 | 0.3496 | 0.3496 | 0.3496 |

Table: Performance Metrics for LSTM Model with Textual Features used for predicting user's stance based on their submitted posts.

| **Set** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| **Test** | 0.430 | 0.4236 | 0.4216 | 0.4236 |

# Conclusion and Future Work

- Edge Homogeneity on Source Node and Target Node around Brexit on reddit dataset
- Prediction of future political stance based on different features defined using the structure online diffusion
- Apply the Model on Dividing the data on different time periods
- Improve the model with adding additional hyper parameters

# Thank You