These text files contain information regarding the spread of individual tweets. Each file contains information for the root tweet and all retweets of this root tweet. With this data once can reconstruct the temporal evolution of the "retweet graph" for this root tweet. The data files are named with the following format:

**Source_followers_tweet_[number of retweets+1]_[tweet source screen_name].txt**

Each line of the data file contains tab delimited information for a tweet. The first line is for the root tweet, and the subsequent lines are for each retweet of this root tweet. The lines in the file have the following format:

**RetweetCount UserId ScreenName FollowerCount DistanceFromRoot Time ParentScreenName Text**

These fields are explained below.

**RetweetCount – the number of the retweet (root has RetweetCount 0).**

**UserId – Twitter Id of the author of the tweet**

**ScreenName – Twitter screen name of the author of the tweet**

**FollowerCount – Number of followers of the author of the tweet.**

**DistanceFromRoot – the distance (in hops) of the author of the tweet from the root tweet in the retweet graph for this particular root tweet. For the root, DistanceFromRoot =0.**

**Time – the time when the tweet was created**

**ParentScreenName – the screen name of the author of the source tweet for the retweet. This can also can be thought of as the parent of the author of the retweet in the retweet graph.**

**Text – the text of the retweet**

The Excel files contain the file names of the training and prediction tweets for each data partition used to test different predictive models. The analysis in the paper used the partition found in file **Partition_1.xlsx**.