Predicting Interesting Tweets in your Network

AUTHORS

Will Webberley, Dr Stuart M Allen, Prof Roger M Whitaker

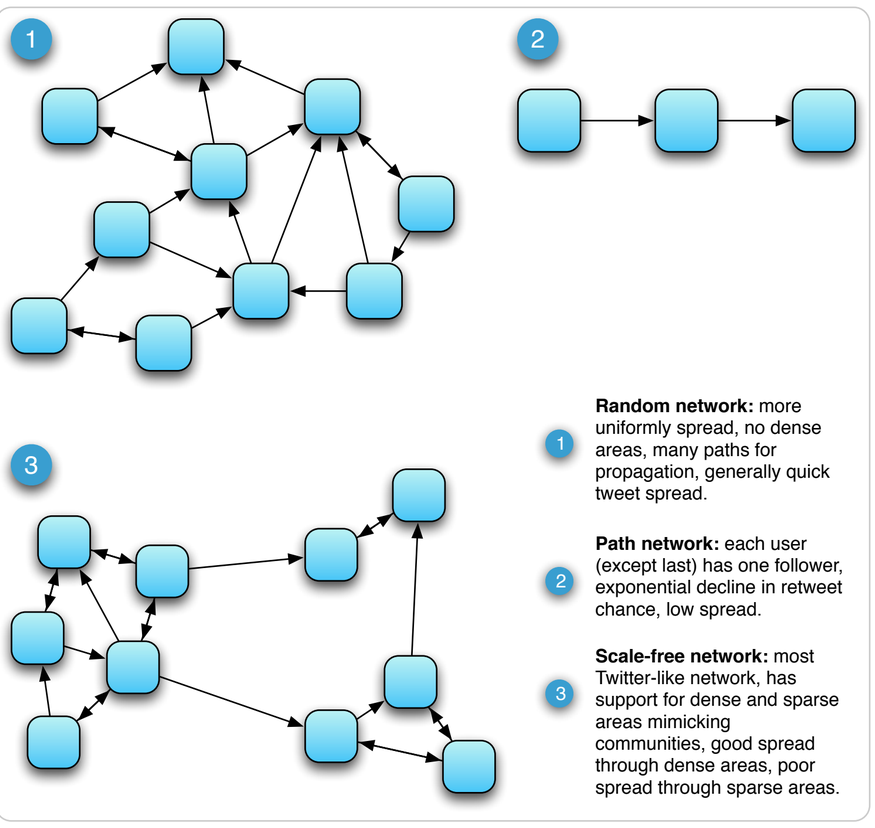
EVENT

Poster Day, Cardiff School of Computer Science & Informatics

WHEN

January 2013

Tweets



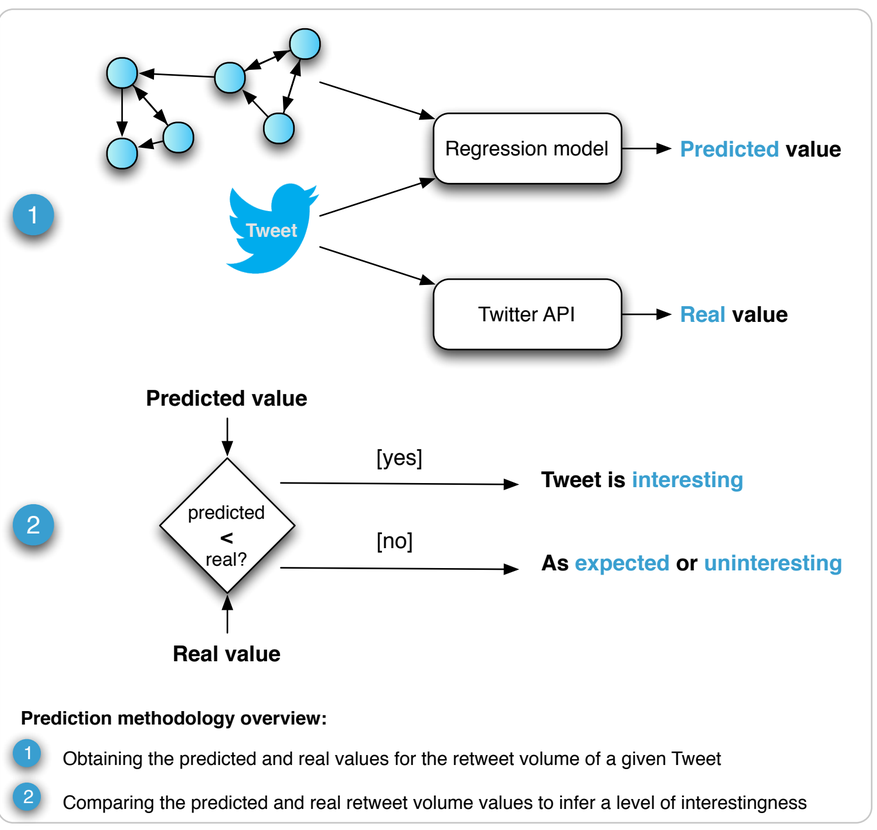
Simulating networks using a regression model — Base research

Retweets are made by users, who must first judge a Tweet to be **interesting** and then actually make the decision on whether or not to retweet it. This means that user interests and the **relevance** of information play a large role in how far a Tweet may propagate.

It is believed that Twitter's **network structure** itself also plays a large role in how far Tweets can 'travel'. In Twitter, users elect to follow others, forging a **directional link** between them. As more edges are created in the graph, more avenues are made down which propagation can occur.

Here, three network types are compared; **path**, **random** and **scale-free** graphs. The patterns in retweet behaviour can be shown to vary considerably in the different network types.

In addition, it is demonstrable that the generated scale-free networks have a very similar retweet behaviour pattern to that of Twitter itself, indicating that Twitter's own social structure **has scale-free properties**.



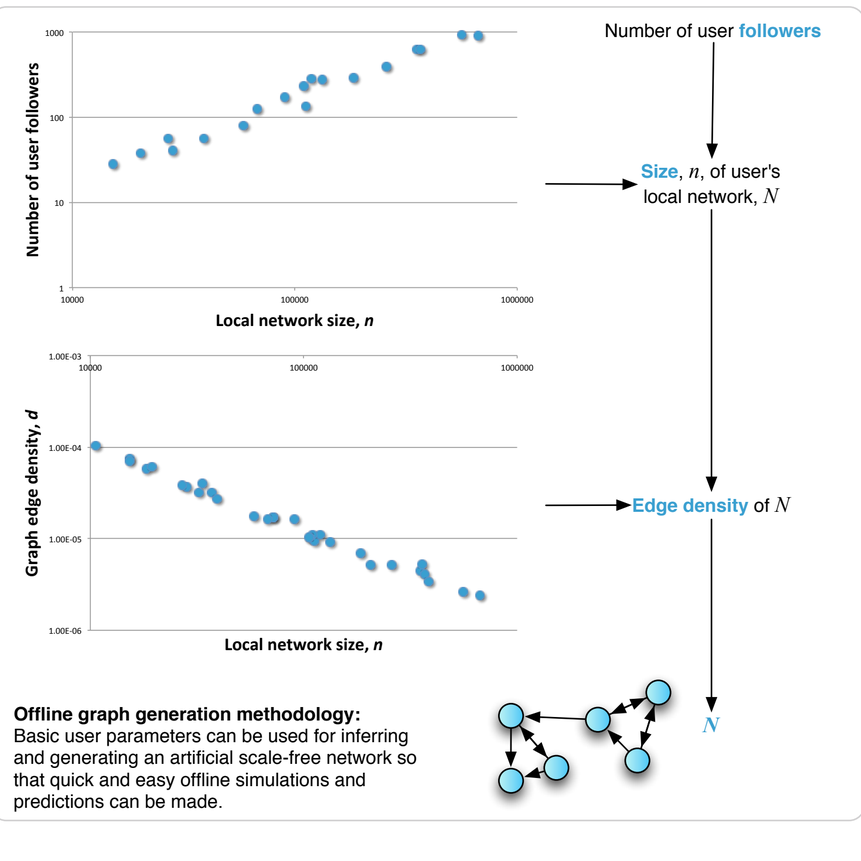
Predicting tweet interestingness — Online predictions using the model

Building on the base research, a potential methodology emerged for **predicting** whether a Tweet is **interesting** (or uninteresting). This uses part of the work above, and, specifically, the retweet volume for a given Tweet in a particular network.

If, instead, a particular Twitter user's Tweet was simulated through that user's actual Twitter **local network** (within two hops), an estimated prediction value for the number of times that Tweet is retweeted can be obtained.

From this, a level of interestingness can be inferred. If a Tweet was predicted to be retweeted twice, but was actually retweeted five times on Twitter, then we can say that this Tweet is **more interesting**, on average, than other tweets with similar features.

Preliminary results from this ongoing work suggest that this method can relatively **accurately predict** the retweet outcome for a given Tweet (~90% accuracy). Amazon's Mechanical Turk was used to help verify the Tweet interestingness.



Inferring network parameters for offline predictions — Future work

The research above relies on the collection of a user's entire local network to be collected from Twitter in order to model their tweets' behaviours. This is **impractical** due to speed and Twitter's REST API's **restrictions**.
The usefulness of this would be improved if a network could be artificially, and quickly, generated in which to model Tweets to predict their interestingness.

Ongoing work suggests how various basic user parameters can be used to generate an artificial scale-free network using a base edge-density as a parameter.
Whilst the edge density of a user's local network cannot be obtained directly from Twitter, it can be **estimated** from other, more accessible information.

It can be shown that the **size** of a user's local network can be inferred from the **follower count** of the user. From this, an estimation can be made on the **edge-density** of the network, from which an **artificial scale-free** network can be generated.

It is hoped that this can be used for making **quick**, **offline** predictions on Tweet interestingness.