

Modelling the Behaviour of Retweets

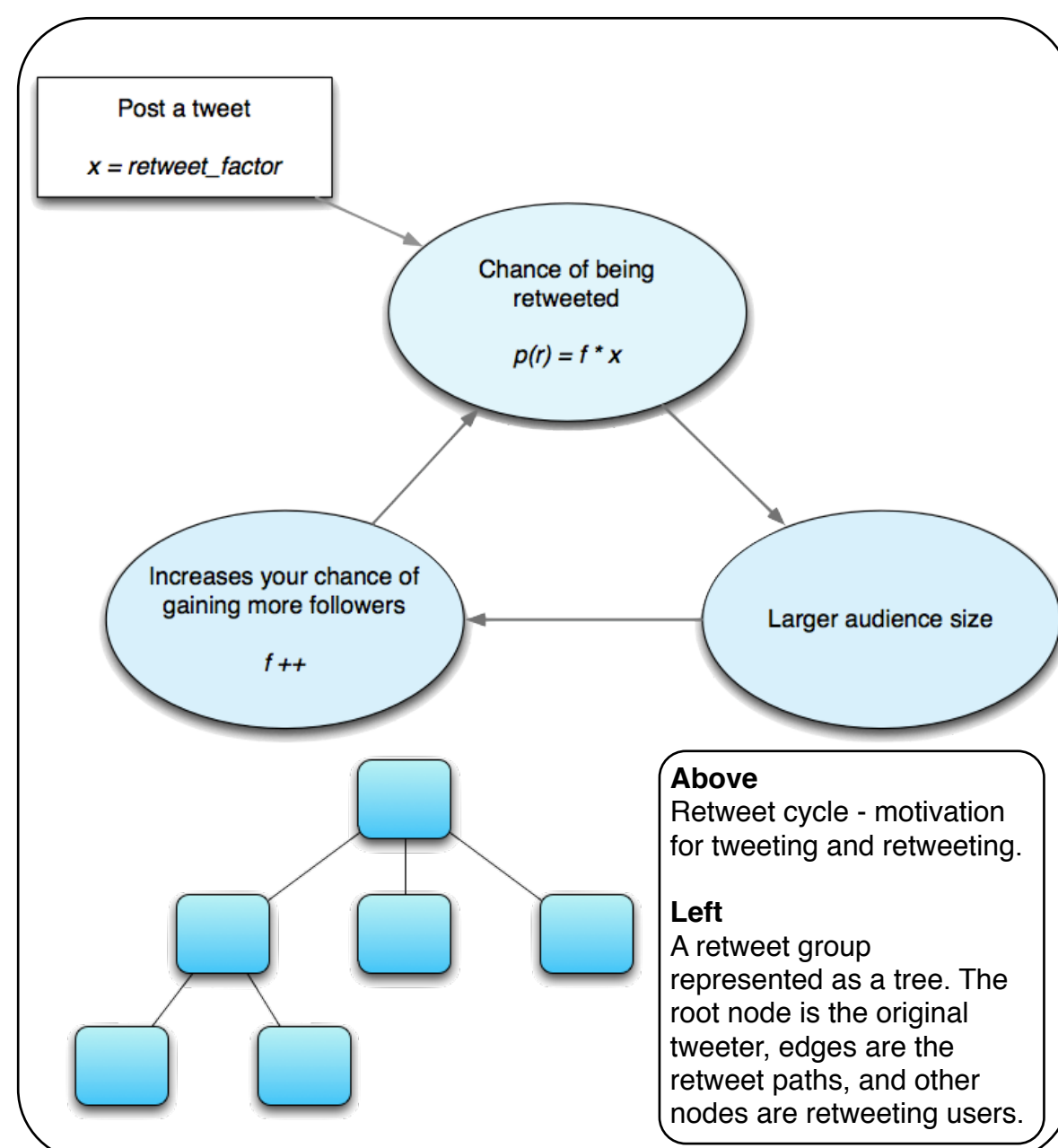
Introduction

Retweeting is the method by which users can **propagate** information through Twitter.

Tweets can therefore spread in different fashions. The *distance* travelled by a tweet (number of hops) is known as the **path-length**. A tweet along with all of its associated retweets is known as a **retweet group**. Groups have various sizes depending on the tweet's popularity.

More widespread retweet groups have the opportunity to propagate the tweet across many different **communities**, thus enabling more users to be exposed to the tweet.

The interesting part of studying retweets comes from the **decentralised** nature of message-forwarding, similar to that of **opportunistic** gossiping networks.



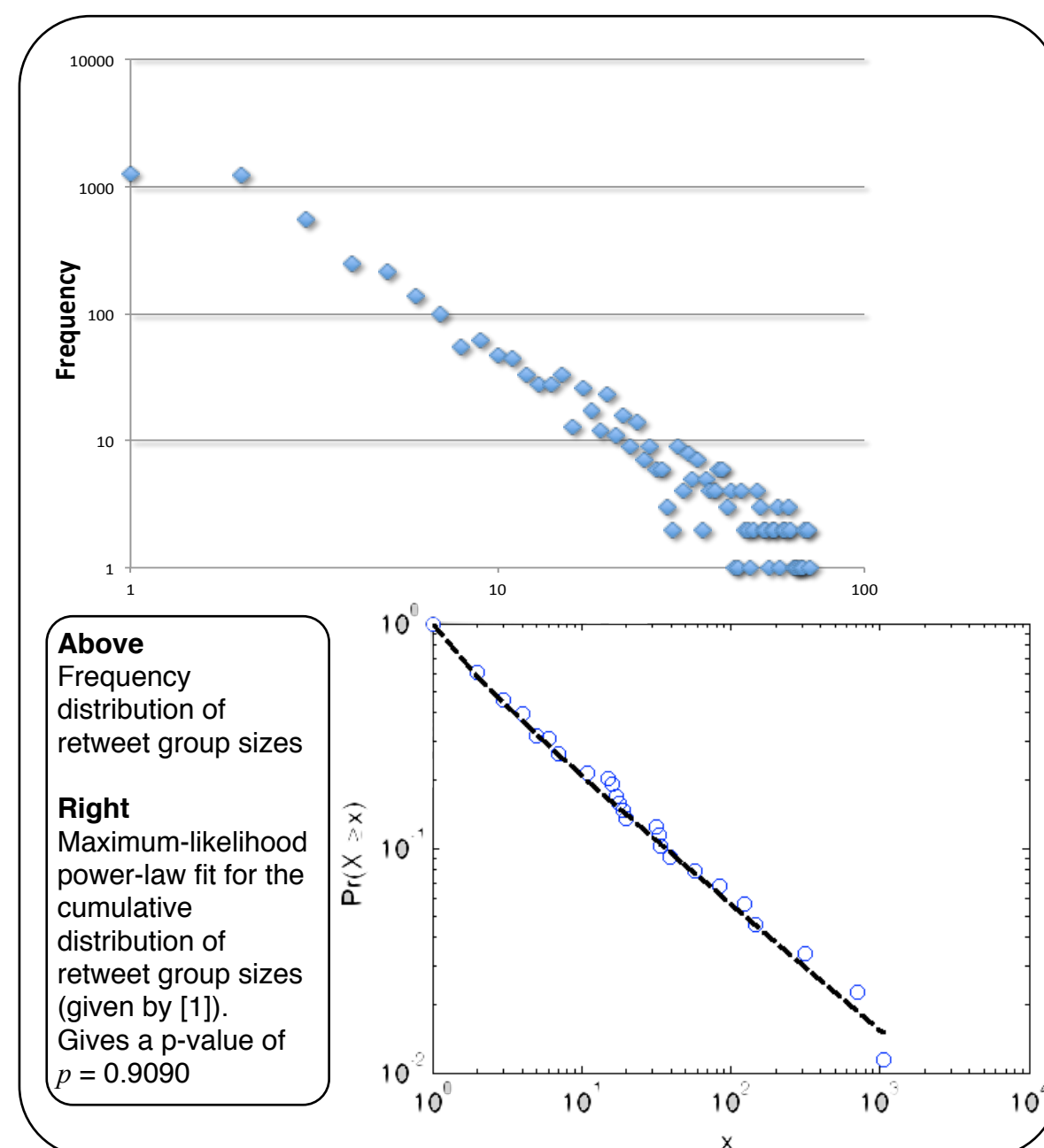
Twitter Data

[1] Clauset A, Shalizi C R, Newman M E J. *Power-Law Distributions in Empirical Data*. ArXiv e-prints, June 2007.

In the hope of understanding the retweet phenomenon more closely, initial research was carried out into the behaviour of retweets in terms of their **spread** and the size of the **audience** of users reached.

More specifically, the **distance** travelled by tweets in terms of the **path-length** as well as the width of propagation were of interest. The frequency distribution of **maximum** path-lengths was discovered, and that the distribution of retweet group **sizes** follows a power-law trend. The latter result is reinforced by applying the statistical models introduced by Clauset et al ^[1].

From this, data has been extracted regarding the number of **unique** viewers of retweet groups in order to gain an understanding of how different types of retweet groups affect the **audience size**.

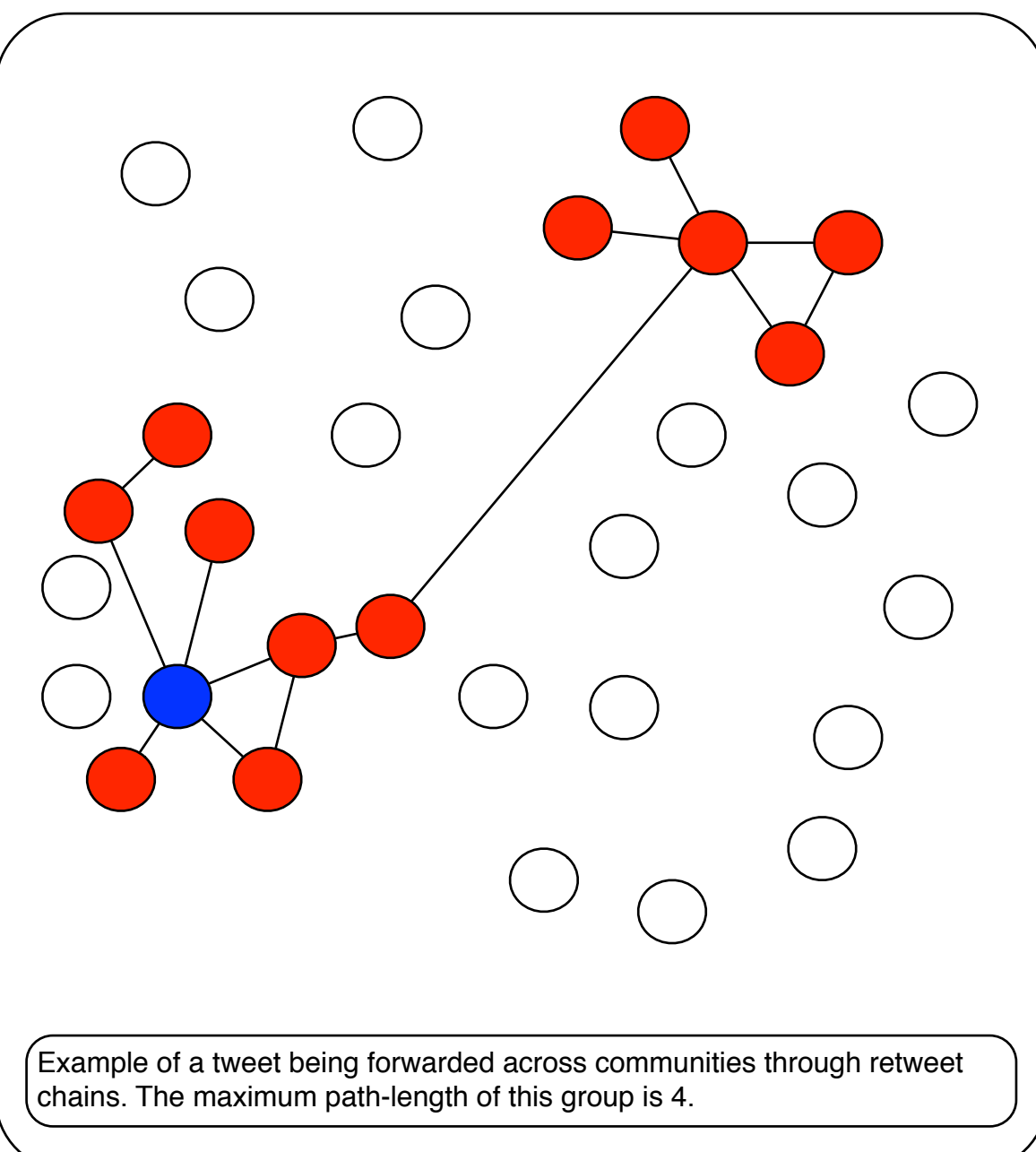


Modelling Retweet Behaviour

In order to predict how information is disseminated around Twitter, development has started on a retweet **model**. It is hoped that the model can be used to realistically simulate tweet propagation in different types of networks.

NetworkX methods are currently being used to produce **scale-free** networks of users using **preferential attachment** algorithms, generated with different sizes and densities. Networks of other types (namely random and linear) have been produced in order to highlight the differences between results produced by different networks.

However, more data is being collected from Twitter in the hope that the social ties between users represented within this data can be used instead to generate more realistic networks.



Model Results

To check the **validity** of the model, comparisons need to be made with the results from the model against the **real-life** results from the data collected from Twitter.

The same experiments that were carried out on Twitter were also done on the model, and sets of data have been produced and analysed in the same way.

Some of the data has, thus far, been relatively similar in terms of the trends and patterns emerging, but further work needs to be done to verify this **statistically**.

Research into relevant statistical methods, including the Kolmogorov-Smirnov statistic, has been made in the hope that this will go some way to validating the model results.

