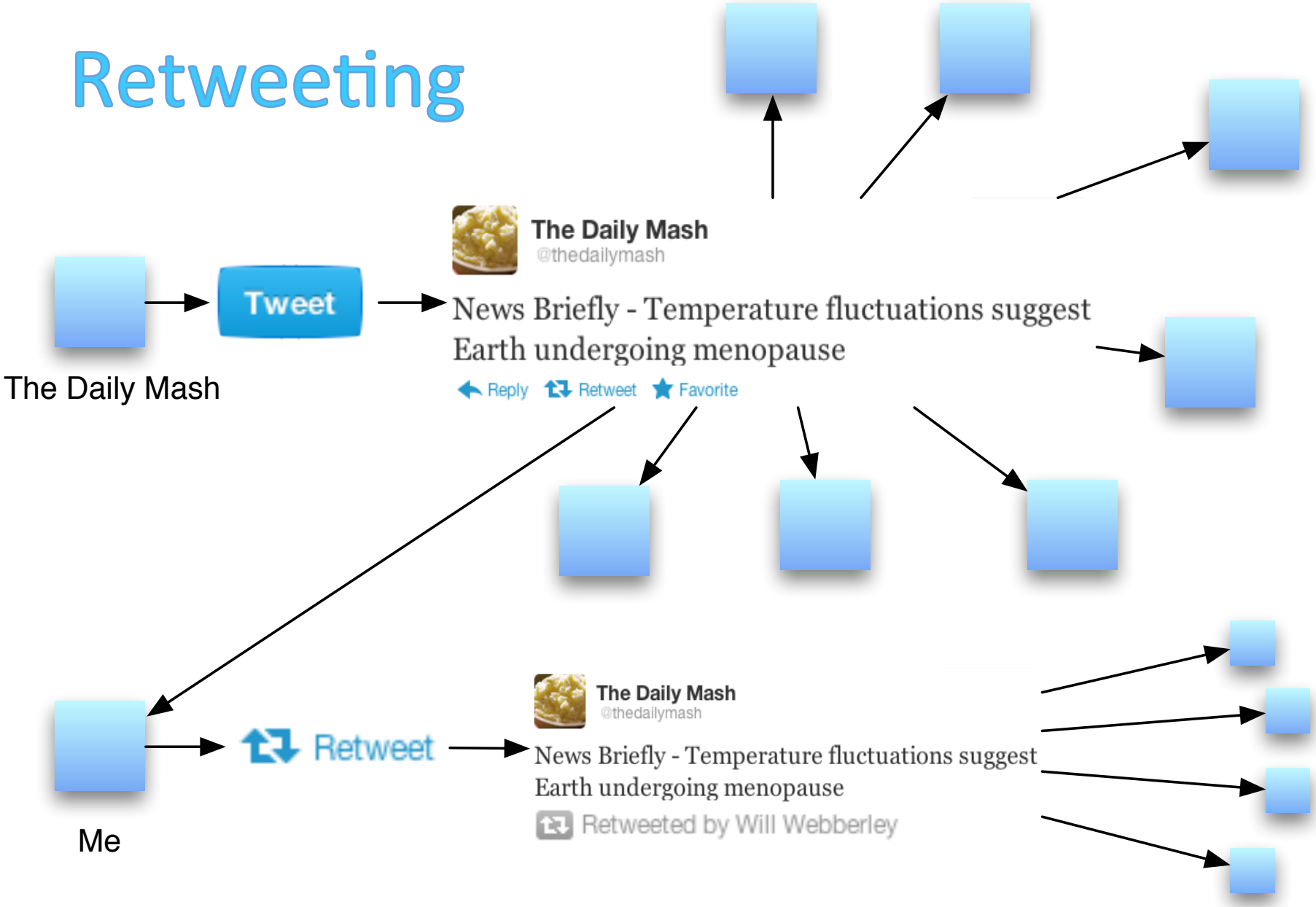
Modelling the Behaviour of Retweets

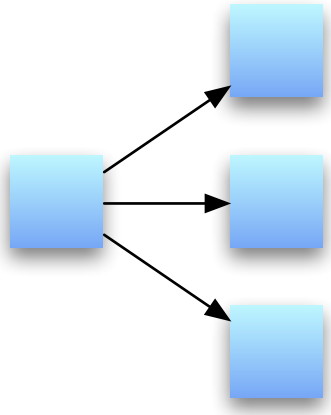
Will Webberley

Retweeting

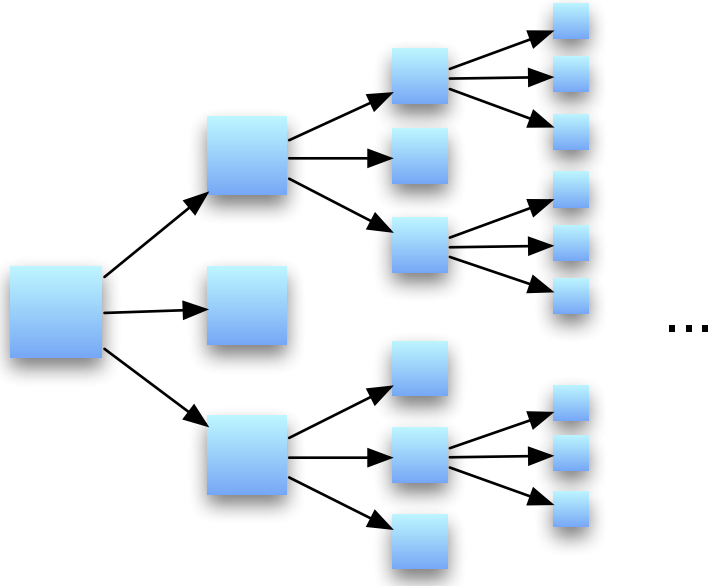

Tweet
The Daily Mash
The Daily Mash
@thedailymash
News Briefly - Temperature fluctuations suggest Earth undergoing menopause
Reply
Retweet
Favorite
Retweet
Me
The Daily Mash
@thedailymash
News Briefly - Temperature fluctuations suggest Earth undergoing menopause
Retweeted by Will Webberley

Retweeting

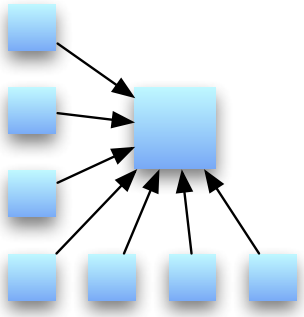
Without retweeting



With retweeting



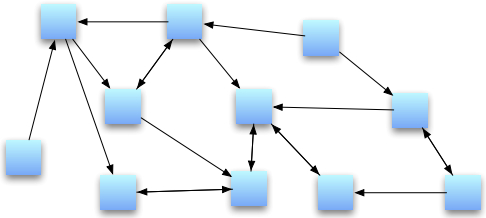
Motivation



Retweet chance

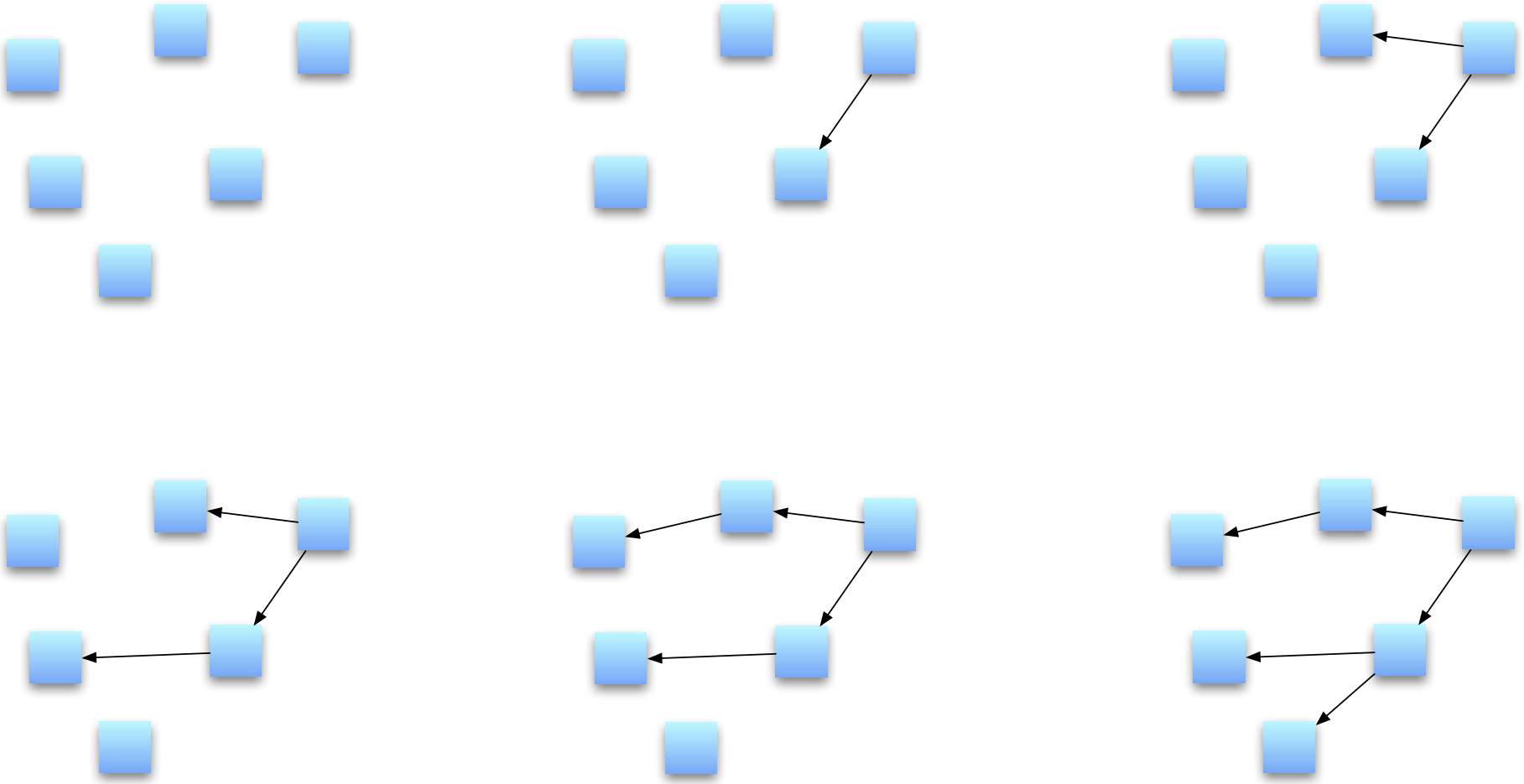
http:// …

Retweet chance

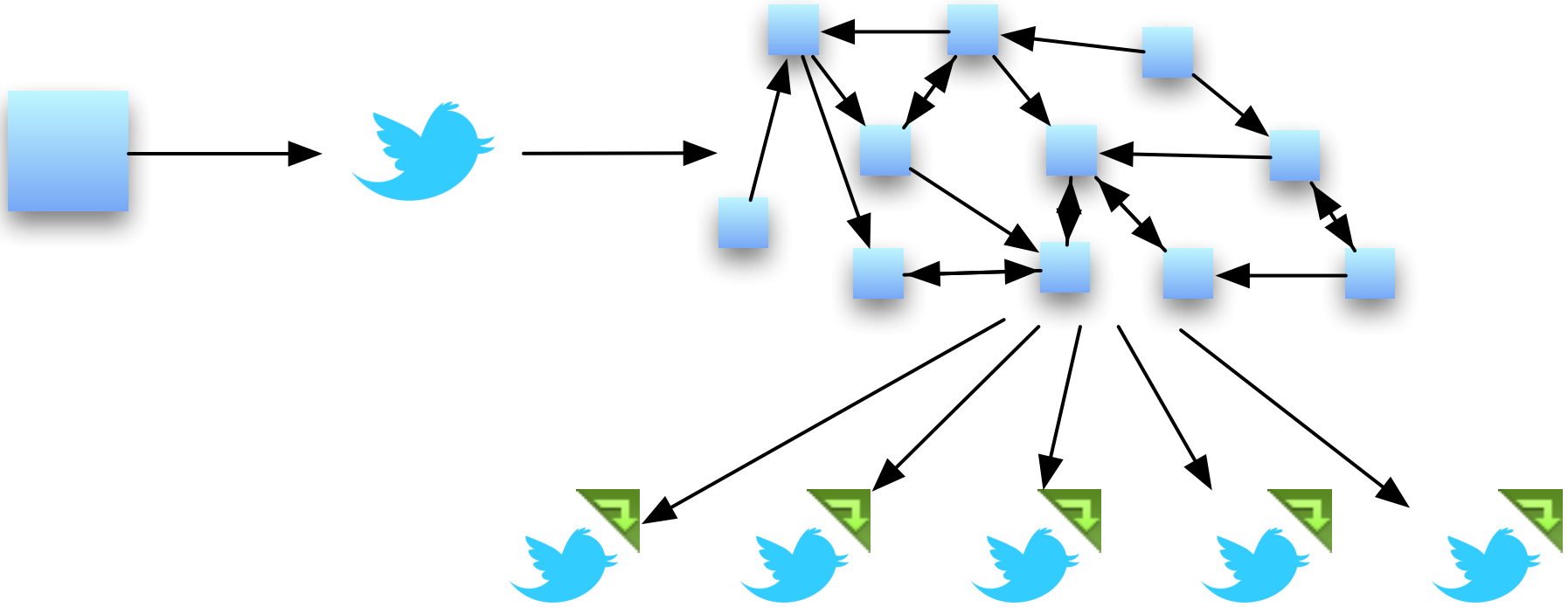


Retweet chance

Twitter Social Structure



Modelling Retweeting



Given a tweet, how many retweets are produced by the network?

Simulation Algorithm

```
users = {User 1}
RT    = {}
```

For each user in users

Calculate retweet probability

Remove user from `users`

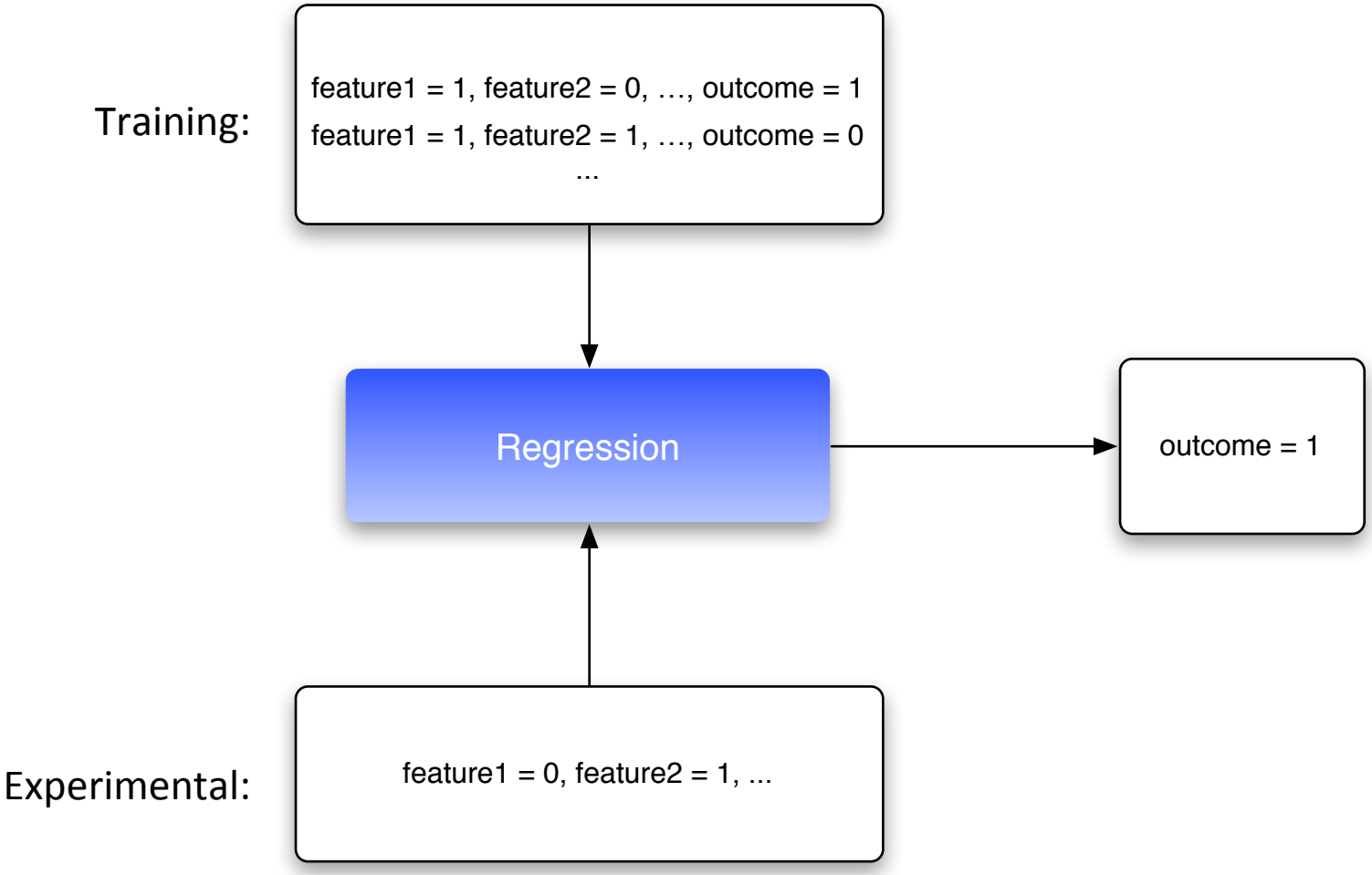
Add followers of user to `users`

Add user to `RT`

Repeat

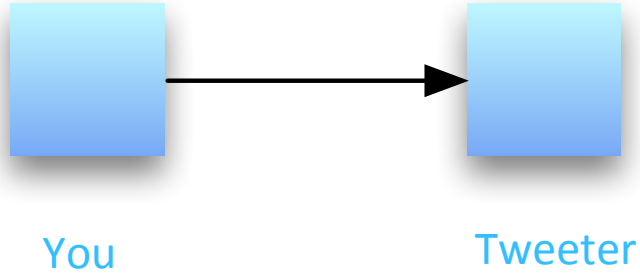
Repeat

Training the Logistic Regression

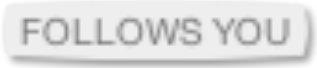


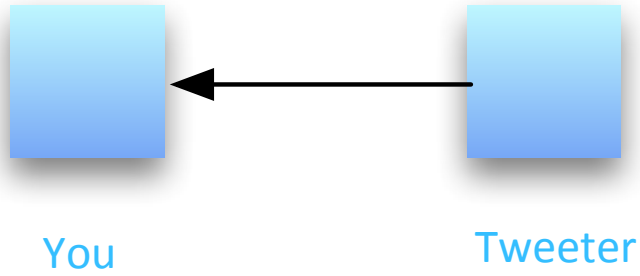
Features

following



followed





Features

`URL`

http:// ...

Will Webberley
@flyingSparx

Where would you end up if you dug straight through the Earth? antipodemap.com

---

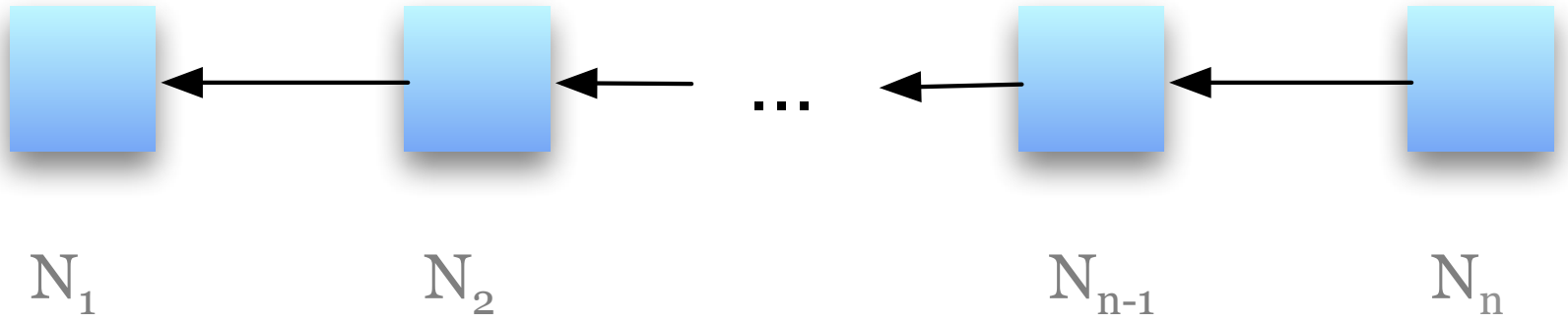
`mentioned`

@ ...

Michael
@billynomates

@flyingSparx Everyone knows it's China

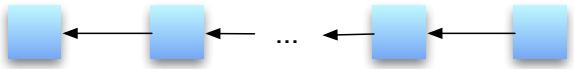
Linear Networks

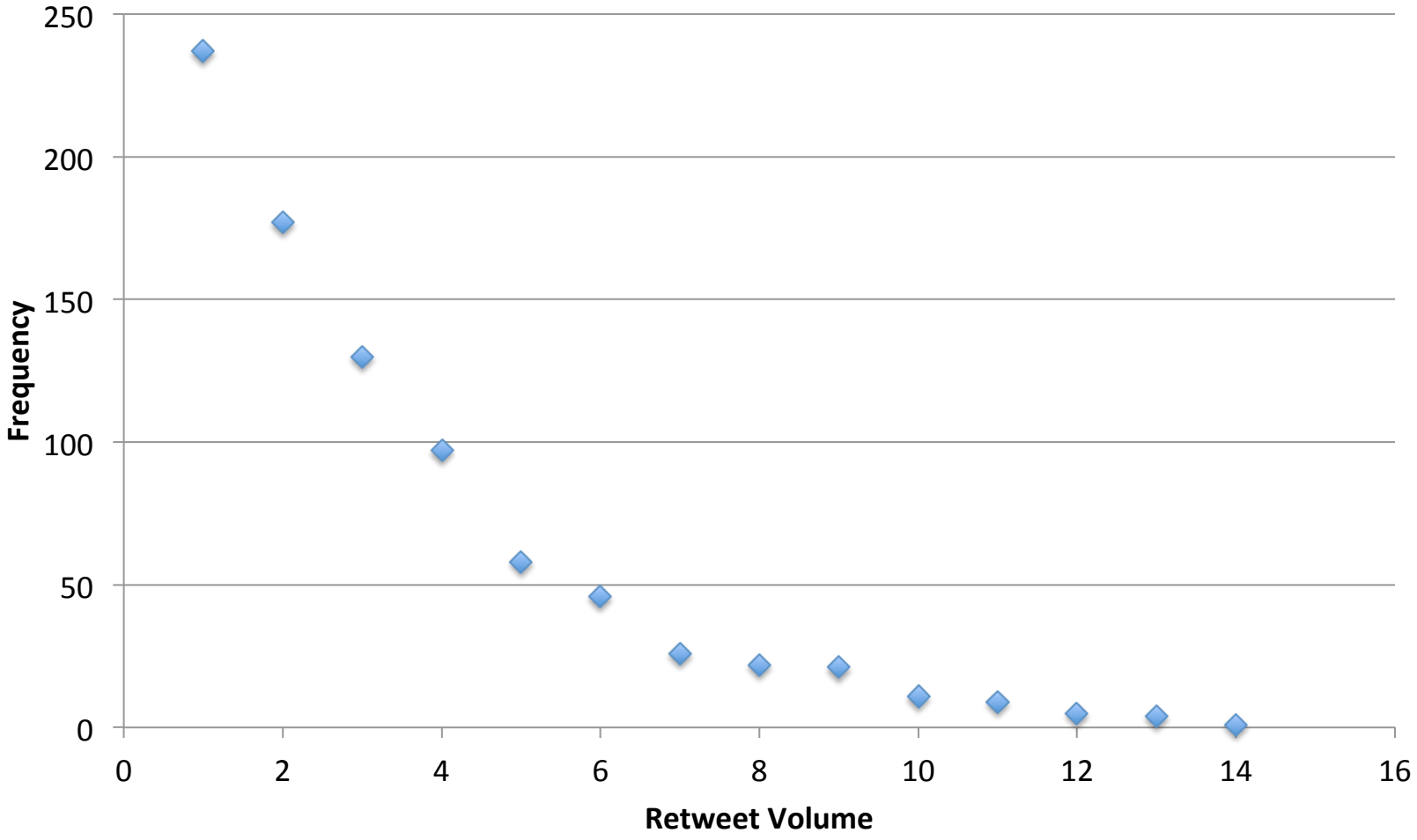


Each user (except last) has one follower

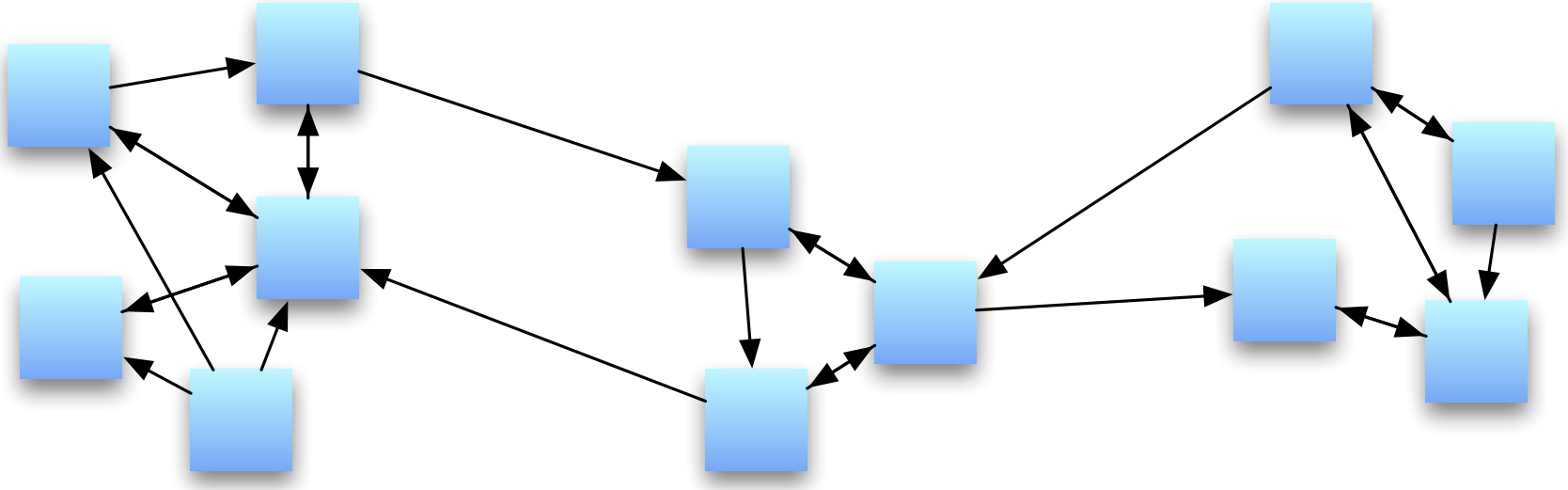
Chance of a user retweeting relies on all preceding users also retweeting

Linear Networks





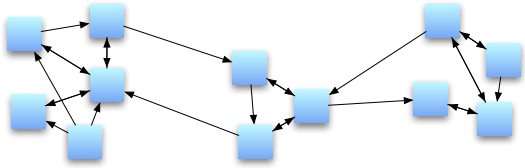
Scale-Free Networks

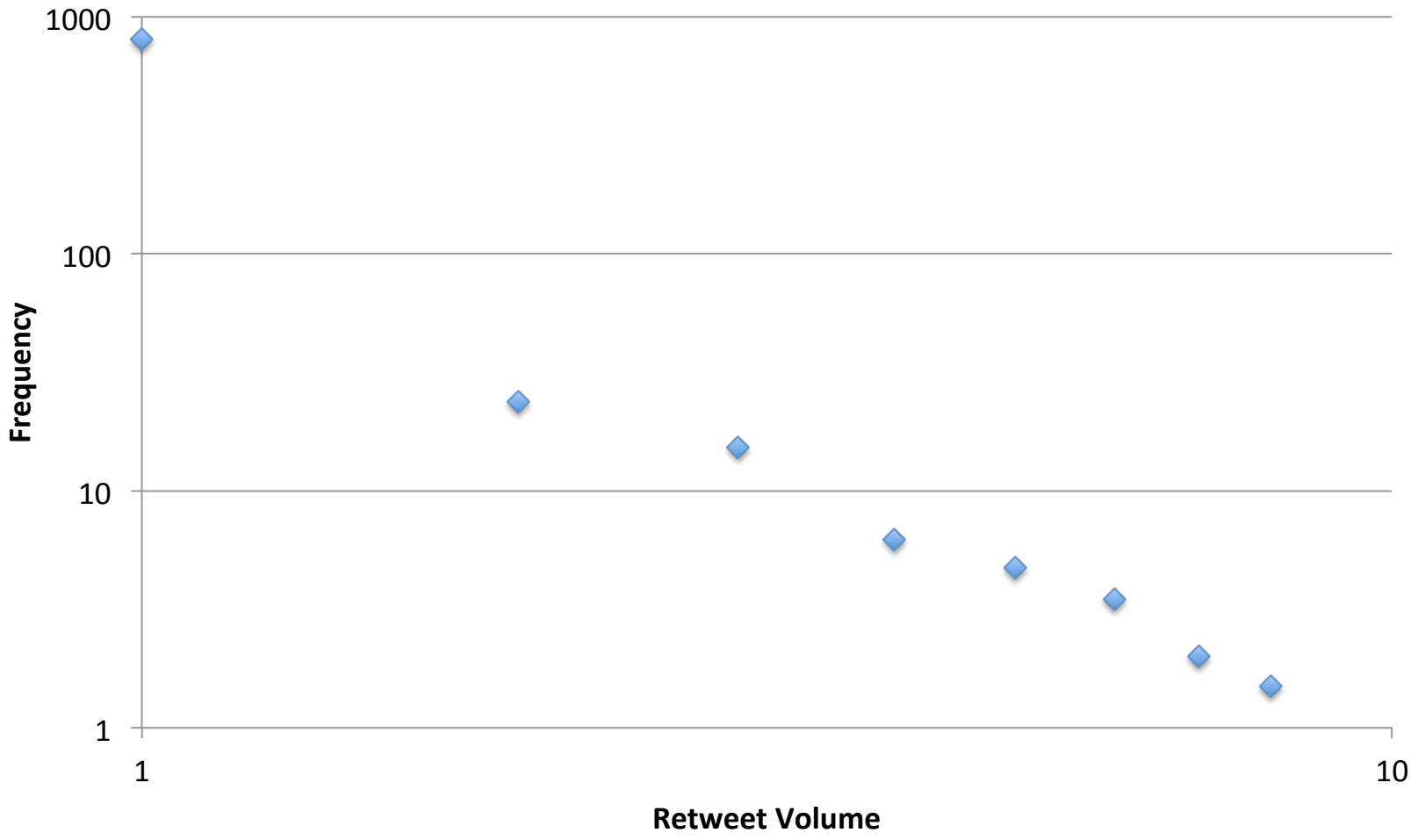


More realistic social network structure

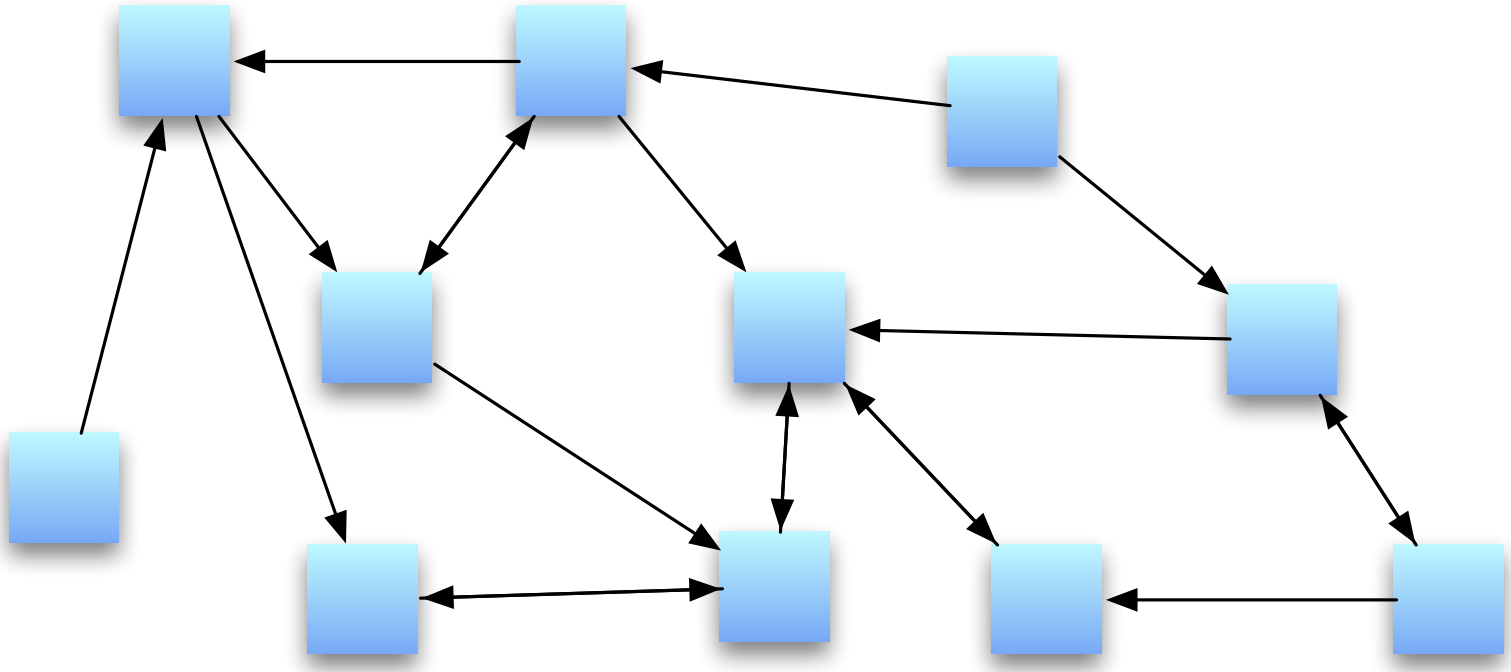
Allows for clustering and areas of dense edges

Scale-Free Networks





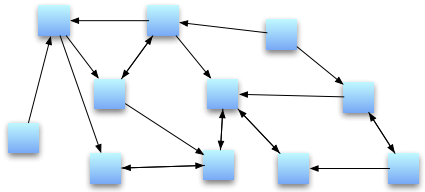
Random Networks

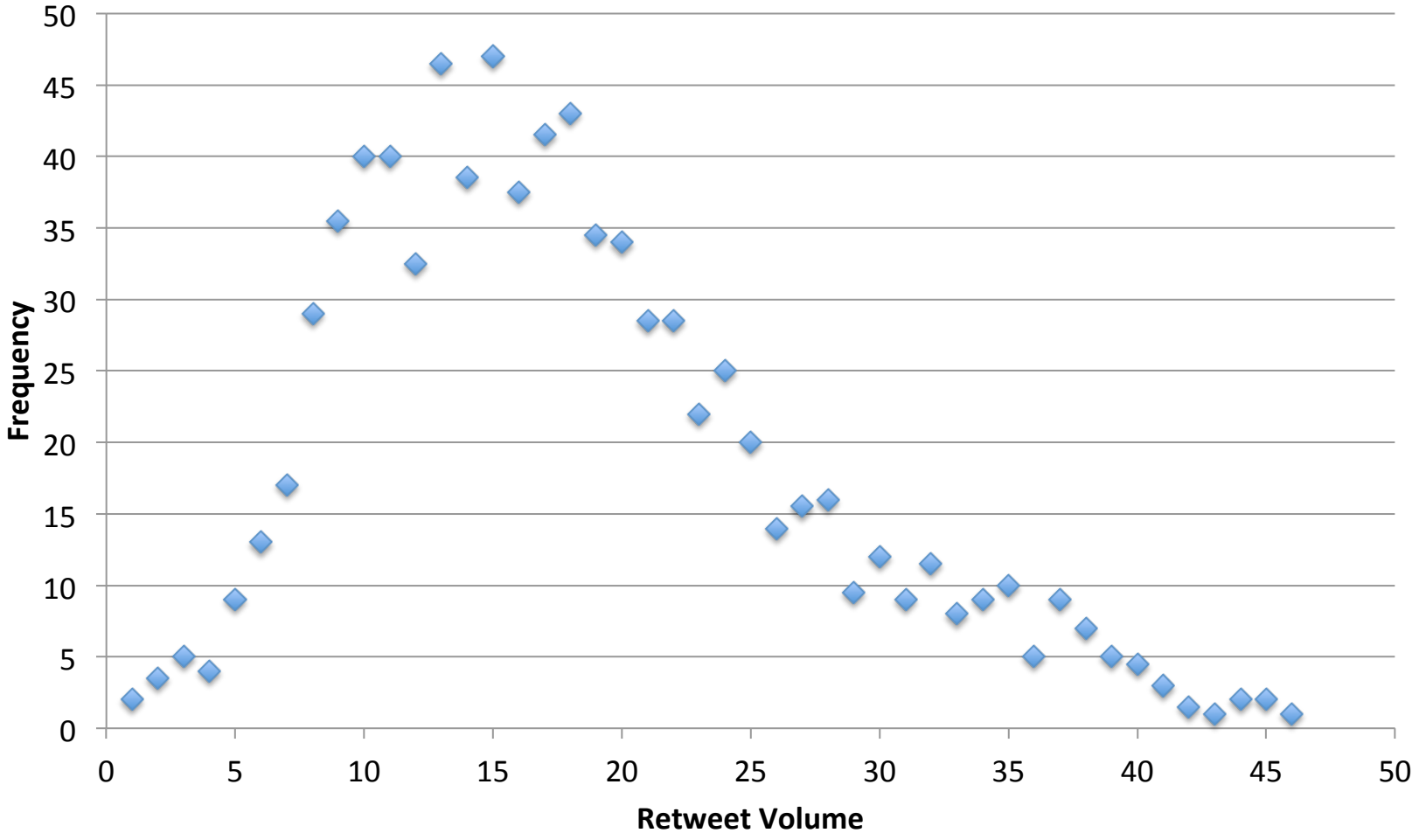


More uniform network structure

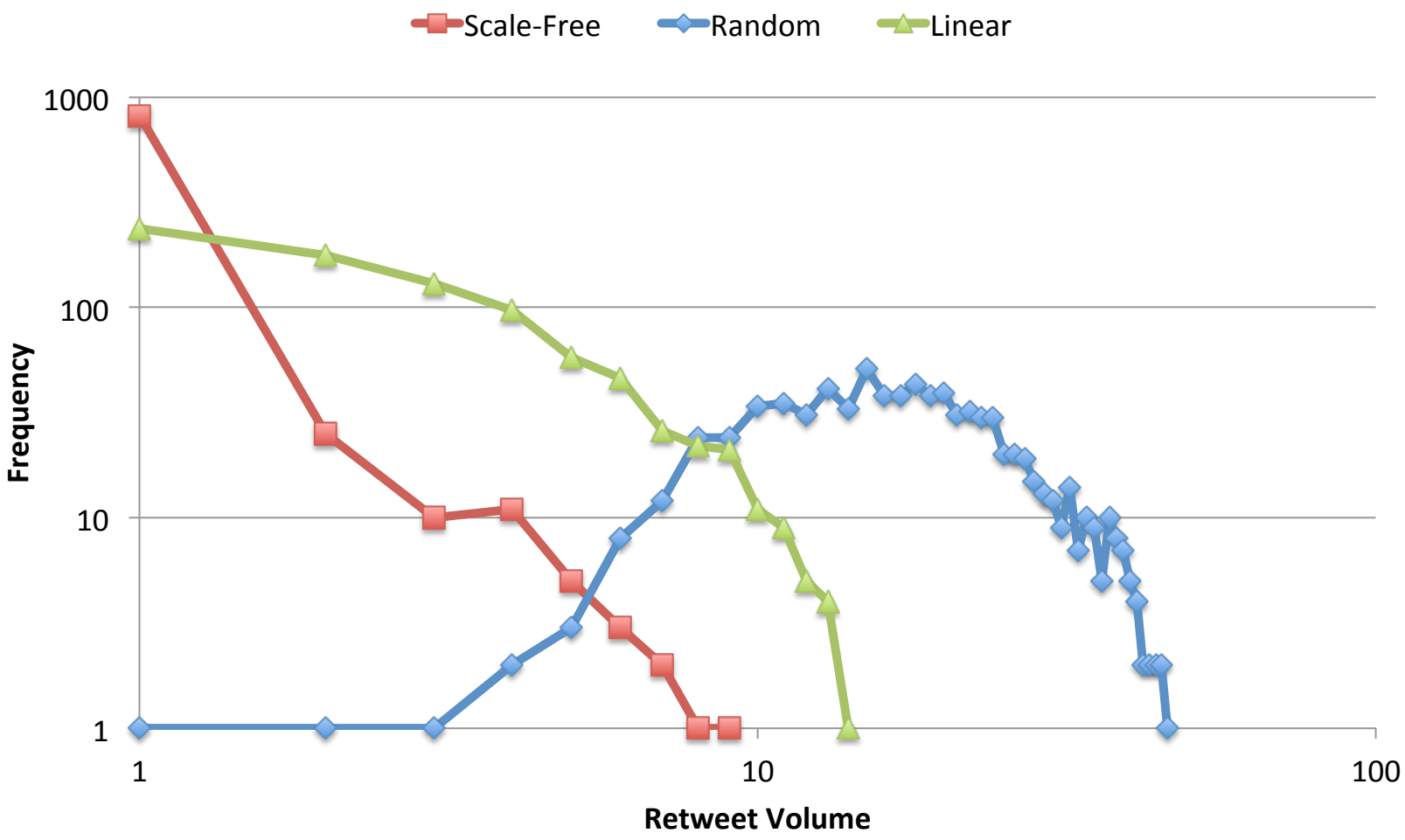
Many communication avenues

Random Networks

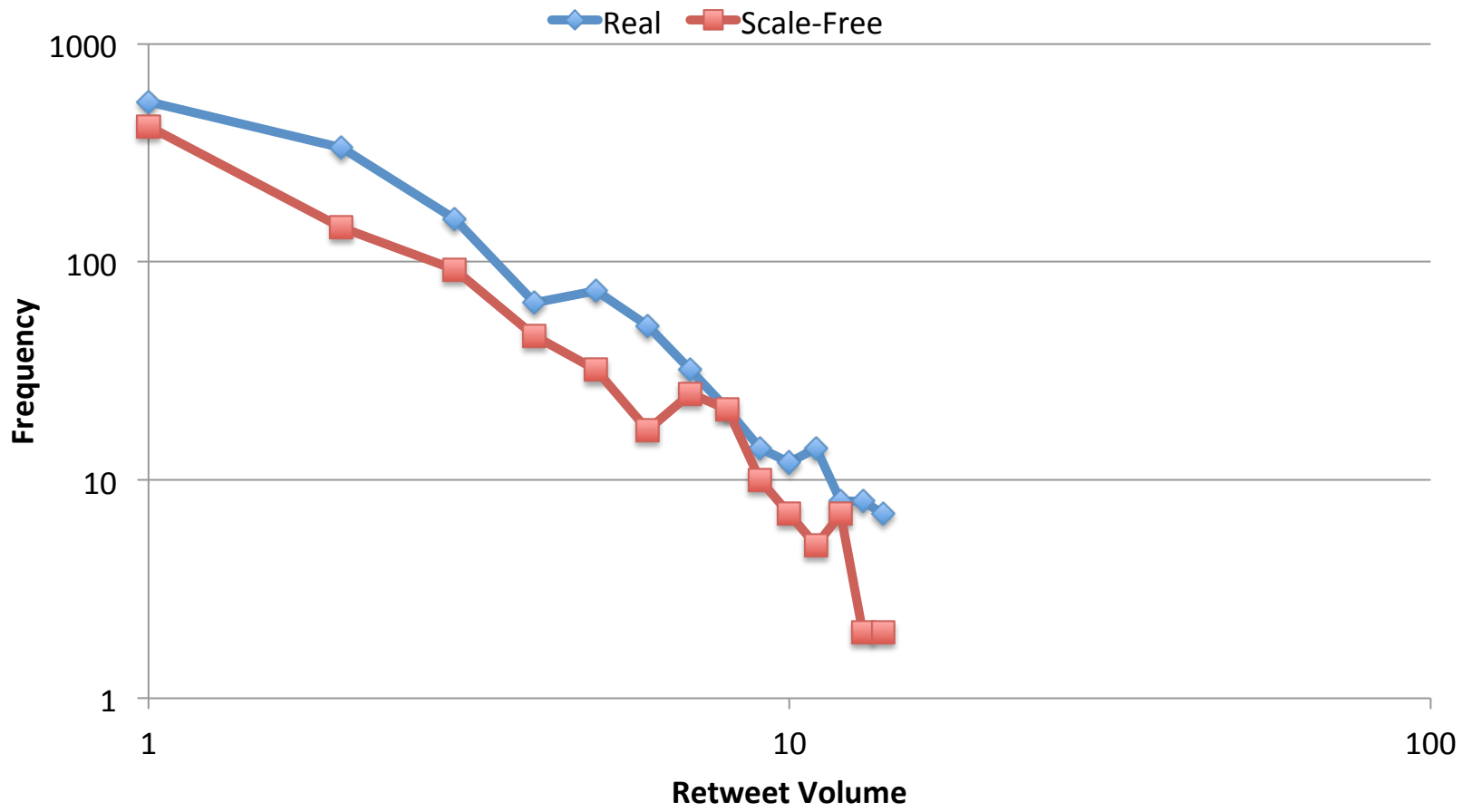




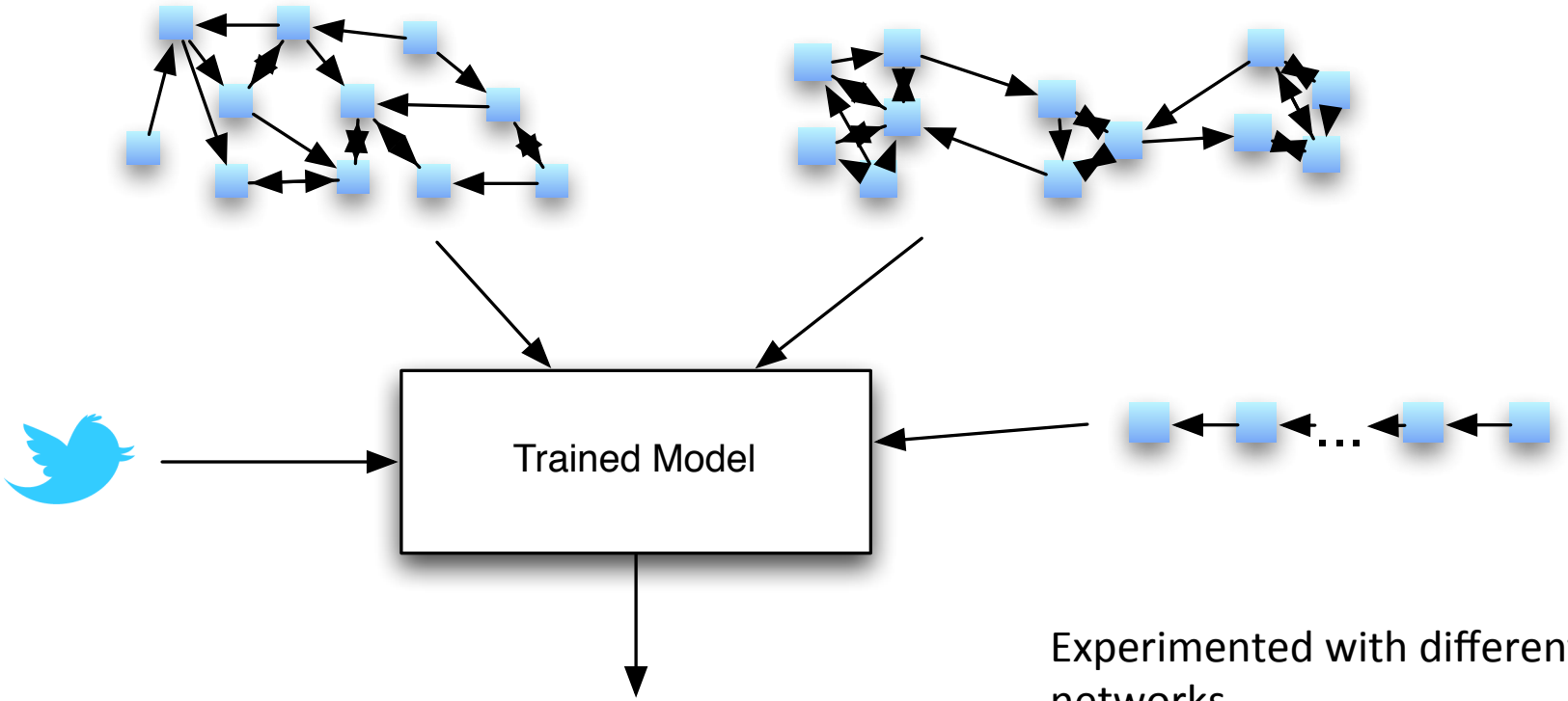
Comparisons

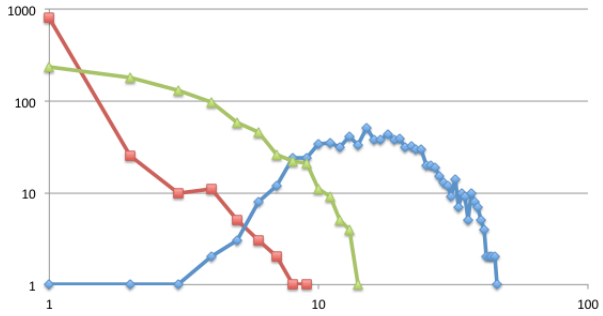


Real Twitter Network Data



Summary





Experimented with different networks

Significant retweet behaviour differences

Closing & What's Next?

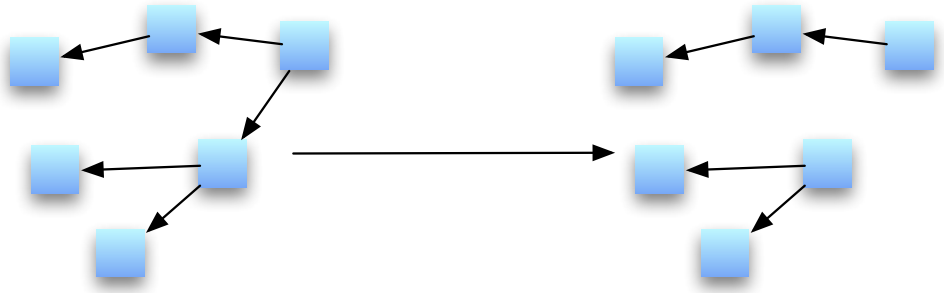
- High traffic in random networks

Propagation

Recall

Precision

- Experiment with adding / removing links



- Can we get relevant tweets to users *without* these links?