COMPLEX NETWORKS

Analyzing Information Cascades in Online Communication

When does a small signal become a large cascade?

Aarti Srivastava | Faculty of Business and Information Technology, Ontario Tech University | aarti.srivastava@ontariotechu.net

1. The question:

How does the probability of sharing change the final reach of information in a connected network?

This project uses computational experiments to study diffusion. Users become active and can activate others through network connections.

2. Data and experiment

Higgs Twitter dataset [1]

Public, anonymized network data surrounding the July 2012 Higgs boson announcement.

456,626 users

14.86 million social connections

Five starting users

The simulation starts with the five users who have the most incoming and outgoing connections combined.

Five sharing probabilities

0.1%, 0.5%, 1%, 5% and 10%. The same five users start every run.

100 runs at each setting

Repeated trials give the mean, minimum and maximum number of users activated.

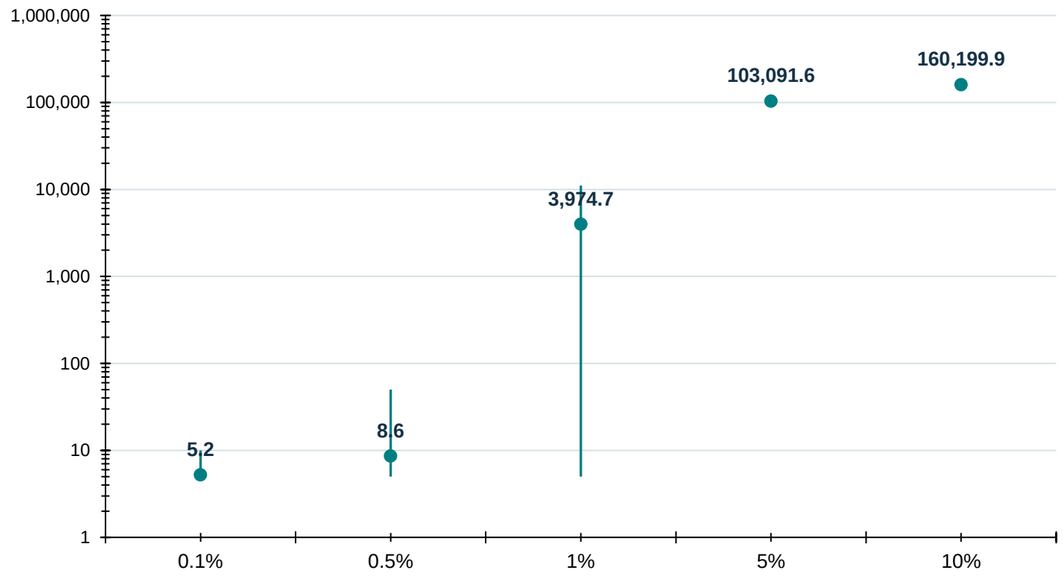
3. Sharing probability and cascade reach

Higher sharing probability produced much larger simulated reach.

Independent Cascade model | Five fixed seeds | 100 runs per setting

Final active users, including seeds

Each step on the vertical axis represents ten times as many users.



Sharing probability (%) | Dots show the mean. Whiskers span the smallest and largest runs.

At 1% sharing probability

The mean was 3,974.7 users. Individual runs ranged from 5 to 11,077 users.

Final reach was sensitive to the assumed sharing probability. The five tested settings leave the precise location of any critical threshold unresolved.

4. How the model works

Independent Cascade [2]

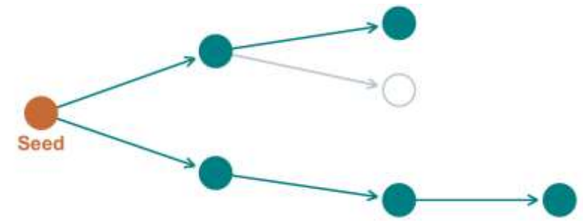


Illustration of activation attempts

Each newly active user gets one chance to activate each inactive neighbour with probability p. Spread stops when no new users activate.

The project also implements Linear Threshold. A user activates when the fraction of active incoming neighbours reaches that user's threshold.

5. Safety and security

Potential application

A future study could test how emergency guidance spreads or how changes in sharing affect misinformation reach. This dataset contains no misinformation labels or emergency outcomes.

Limits that matter

The simulation follows links from followers to the accounts they follow. Using it to predict who sees a message requires validating that direction. Observed retweets are grouped by account, which prevents tracing individual posts.

Sharing assumptions can change reach dramatically, even with the same five starting users.

[1] De Domenico et al. (2013). The anatomy of a scientific rumor. Scientific Reports 3, 2980. Dataset available at snap.stanford.edu/data/higgs-twitter.html.

[2] Kempe, Kleinberg & Tardos (2003). Maximizing the spread of influence through a social network. KDD. doi.org/10.1145/956750.956769

Simulation results from tipping_points.csv in the project archive. Tools used include Python, NetworkX and pandas.