R-bloggers

R news and tutorials contributed by (563) R bloggers

- [Home](#)
- [About](#)
- [RSS](#)
- [add your blog!](#)
- [R jobs���](#)
- [Contact us](#)

Welcome!

Here you will find daily **news and tutorials about R**, contributed by over 555 bloggers. There are many ways to **follow us -**

By e-mail:

| Your e-mail here |
| Subscribe |

15553 readers
BY FEEDBURNER

On Twitter:

Follow @rbloggers 14.5K

On Facebook:

R bloggers
Like

21,052 people like R bloggers.

If you are an R blogger yourself you are invited to add your own R content feed to this site (**Non-English** R bloggers should add themselves- here)

Jobs for R-users

- Improve UI for our general online data analysis and visualization tools (R)
- Looking for a partner to code an algorithm which will trade base on twitter and other social data feed
- R Integration with D3 charts
- Data Analyst at uSwitch

- Director, Pew Research Center Labs

Search & Hit Enter

Popular Searches

- heatmap
- web scraping
- hadoop
- twitter
- PCA
- shiny
- finance
- boxplot
- maps
- ggplot
- ggplot2
- animation
- LaTeX
- trading
- time series
- excel
- quantmod
- how to import image file to R
- knitr
- googlevis
- tutorial
- rstudio
- eclipse
- market research
- rattle
- gis
- rcmdr
- ecdf
- coplot
- sweave

Recent Posts

- Microsoft to acquire Revolution Analytics: Now this gets interesting…
- RSS feeds for statistics and related journals
- R Open Source and Commercial Support
- Assertive R programming in dplyr/magrittr pipelines
- Revolution Analytics joins Microsoft
- Shiny 0.11, themes, and dashboard
- R Package to Download Fitbit Data
- A first look at Spark
- Data Wrangling Webinar and IFAR Chapter

Other sites

Information Transmission in a Social Network: Dissecting the Spread of a Quora Post

September 7, 2011
By Edwin Chen

Like   Share   2   Tweet   4

(This article was first published on **Edwin Chen's Blog » r**, and kindly contributed to R-bloggers)

tl;dr See this movie visualization for a case study on how a post propagates through Quora.

How does information spread through a network? Much of Quora's appeal, after all, lies in its social graph — and when you've got a network of users, all broadcasting their activities to their neighbors, information can cascade in multiple ways. How do these social designs affect which users see what?

Think, for example, of what happens when your kid learns a new slang word at school. He doesn't confine his use of the word to McKinley Elementary's particular boundaries, between the times of 9-3pm — he introduces it to his friends from other schools at soccer practice as well. A couple months later, he even says it at home for the first time; you like the word so much, you then start using it at work. Eventually, Justin Bieber uses the word in a song, at which point the word's popularity really starts to explode.

So how does information propagate through a social network? What types of people does an answer on Quora reach, and how does it reach them? (Do users discover new answers individually, or are hubs of connectors more key?) How does the activity of a post on Quora rise and fall? (Submissions on other sites have limited lifetimes, fading into obscurity soon after an initial spike; how does that change when users are connected and every upvote can revive a post for someone else's eyes?)

To start answering some of these questions, I dug into one of my more popular posts, on a layman's introduction to random forests.

Users, Topics

Before looking deeper into the voting dynamics of my post, let's first get some background on what kinds of users my answer reached.

Here's a graph of the topics that question upvoters follow. (Each node is a topic, and every time upvoter X follows both topics A and B, I add an edge between A and B.)

We can see from the graph that upvoters tend to be interested in three kinds of topics:

- **Machine learning and other technical matters** (the green cluster): Classification, Data Mining, Big Data, Information Retrieval, Analytics, Probability, Support Vector Machines, R, Data Science, …
- **Startups/Silicon Valley** (the red cluster): Facebook, Lean Startups, Investing, Seed Funding, Angel Investing, Technology Trends, Product Managment, Silicon Valley Mergers and Acquisitions, Asana, Social Games, Quora, Mark Zuckerberg, User Experience, Founders and Entrepreneurs, …
- **General Intellectual Topics** (the purple cluster): TED, Science, Book Recommendations, Philosophy, Politics, Self-Improvement, Travel, Life Hacks, …

Also, here's the network of the upvoters themselves (there's an edge
between users A and B if A follows B):
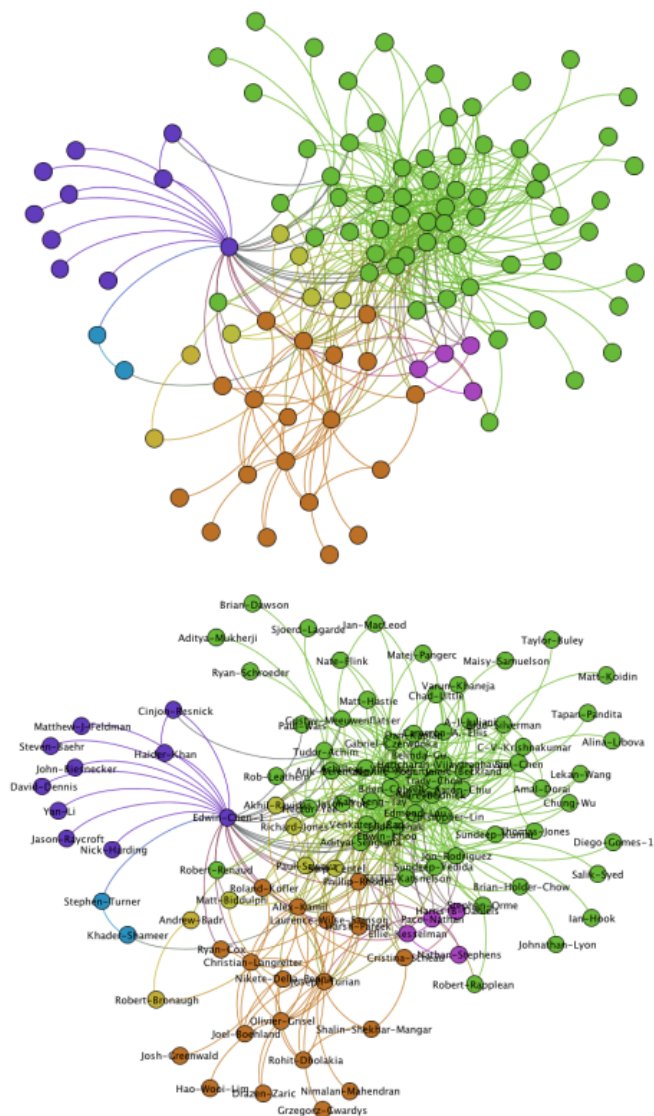
We can see three main clusters of users:

- A large group in **green** centered around a lot of power users and
  Quora employees.
- A group of folks in **orange** centered around people like Oliver
  Grisel, Christian Langreiter, and Joseph Turian.
- A group of people following me, in **purple**.
- Plus some smaller clusters in blue and yellow. (There were also
  a bunch of isolated users, connected to no one, that I filtered out
  of the picture.)

Digging into how these topic and user graphs are related:

- The orange cluster of users is more heavily into machine
  learning: 79% of users in that cluster follow more green topics
  (machine learning and technical topics) than red and purple
  topics (startups and general intellectual matters).
- The green cluster of users is reversed: 77% of users follow more
  of the red and purple clusters of topics (on startups and general
  intellectual matters) than machine learning and technical topics.

More interestingly, though, we can ask: how do the connections
between upvoters relate to the way the post spread?

Social Voting Dynamics

So let's take a look. Here's a visualization I made of upvotes on my

answer across time (click here for a larger view).
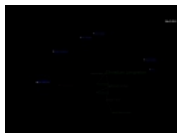
To represent the social dynamics of these upvotes, I drew an edge from user A to user B if user A transmitted the post to user B through an upvote. (Specifically, I drew an edge from Alice to Bob if Bob follows Alice and Bob's upvote appeared within five days of Alice's upvote; this is meant to simulate the idea that Alice was the key intermediary between my post and Bob.) Also,

- Green nodes are users with at least one upvote edge.
- Blue nodes are users who follow at least one of the topics the post is categorized under (i.e., users who probably discovered the answer by themselves).
- Red nodes are users with no connections and who do not follow any of the post's topics (i.e, users whose path to the post remain mysterious).
- Users increase in size when they produce more connections.

Here's a play-by-play of the video:

- On Feb 14 (the day I wrote the answer), there's a flurry of activity.
- A couple of days later, Tracy Chou gives an upvote, leading to another spike in activity.
- Then all's quiet until… bam! Alex Kamil leads to a surge of upvotes, and his upvote finds Ludi Rehak, who starts a small surge of her own. They're quickly followed by Christian Langreiter, who starts a small revolution among a bunch of machine learning folks a couple days later.
- Then all is pretty calm again, until a couple months later when… bam! Aditya Sengupta brings in a smashing of his own followers, and his upvote makes its way to Marc Bodnick, who sets off a veritable storm of activity.

(Already we can see some relationships between the graph of user connections and the way the post propagated. Many of the users from the orange cluster, for example, come from Alex Kamil and Christian Langreiter's upvotes, and many of the users from the green cluster come from Aditya Sengupta and Marc Bodnick's upvotes. What's interesting, though, is, why didn't the cluster of green users appear all at once, like the orange cluster did? People like Kah Seng Tay, Tracy Chou, Venkatesh Rao, and Chad Little upvoted the answer pretty early on, but it wasn't until Aditya Sengupta's upvote a couple months later that people like Marc Bodnick, Edmond Lau, and many of the other green users (who do indeed follow that first set of folks) discovered the answer. Did the post simply get lost in users' feeds the first time around? Was the post perhaps ignored until it received enough upvotes to be considered worth reading? Are some users' upvotes just trusted more than others'?)

For another view of the upvote dynamics, here's a static version of the movie, where we can again easily see the clusters of activity:
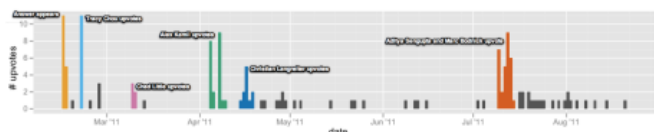
Fin

To summarize a bit what we've found, here are some statistics on the role the social graph played in spreading the answer:

- There are 5 clusters of activity after the initial post, sparked both by power users and less-connected folks. In an interesting cascade of information, some of these sparks led to further spikes in activity as well (as when Aditya Sengupta's upvote found its way to Marc Bodnick, who set off even more activity).

- 35% of users made their way to my answer because of someone else's upvote.
- Through these connections, the post reached a fair variety of users: 32% of upvoters don't even follow any of the post's topics.
- 77% of upvotes came from users over two weeks *after* my answer appeared.
- If we look only at the upvoters who follow at least one of the post's topics, 33% didn't see my answer until someone else showed it to them. In other words, a full one-third of people who presumably would have been interested in my post anyways only found it because of their social network.

So it looks like the social graph played quite a large part in the post's propagation, and I'll end with a big shoutout to Stormy Shippy, who provided an awesome set of scripts I used to collect a lot of this data.

Comments: 6   Add to Facebook!   Tweet it!

2	4
Like	Tweet

Share

To **leave a comment** for the author, please follow the link and comment on his blog: **Edwin Chen's Blog » r**.

R-bloggers.com offers **daily e-mail updates** about R news and tutorials on topics such as: visualization (ggplot2, Boxplots, maps, animation), programming (RStudio, Sweave, LaTeX, SQL, Eclipse, git, hadoop, Web Scraping) statistics (regression, PCA, time series, trading) and more...

If you got this far, why not **subscribe for updates** from the site? Choose your flavor: e-mail, twitter, RSS, or facebook...

Like   Share 2   Tweet 4

Tags: Gephi, information propagation, processing, R, social graphs, social network analysis, Uncategorized

Comments are closed.

Top 3 Posts from the past 2 days

- Revolution Analytics joins Microsoft
- A first look at Spark
- Shiny 0.11, themes, and dashboard

Search & Hit Enter

Top 9 articles of the week

1. Learn Statistics and R online from Harvard
2. Installing R packages
3. In-depth introduction to machine learning in 15 hours of expert videos
4. Scatterplots
5. Using apply, sapply, lapply in R
6. Basics of Histograms
7. Visualization series: Insight from Cleveland and Tufte on plotting numeric data by groups
8. R in Business Intelligence
9. K-means clustering is not a free lunch

Sponsors
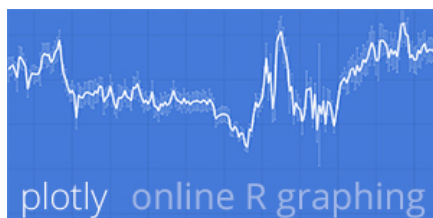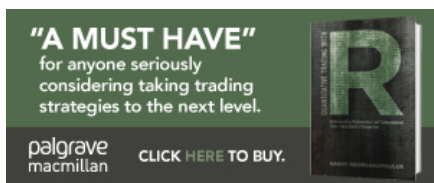
Full list of contributing R-bloggers

R-bloggers was founded by Tal Galili, with gratitude to the R community.

Is powered by WordPress using a bavotasan.com design.

Copyright © 2015 **R-bloggers**. All Rights Reserved. Terms and Conditions for this website