

Eternal September: Online Information-Sharing Community Analysis

Ehsan Gholami, John Karasinski, Jenette Sellin, Dmitry Shemetov
May 26, 2016



StackExchange 

Outline

- Goals and motivation
- Previous works
- 3 datasets
 - Reddit
 - HackerNews
 - StackExchange
- Discussion and conclusions

Research Questions

1. How does the amount of users' activity change over time?
2. How do users' centrality in the network change over time, both overall and by cohort?
3. How does comment length change in the community over time, both overall and by cohort?
4. How do users' activities change in the community over time, both overall and by cohort?

Research Questions (for today)

1. How does the amount of users' activity change over time?
 - a. Average number of active users in each cohort
 - b. Average number of posts/comments from each cohort
2. How do users' centrality in the network change over time, both overall and by cohort?
 - a. Average centrality of each cohort over time

Networks change with time

To investigate changes in users' behavior over time, we divide the users into cohorts (communities).

We then investigate how cohorts behave differently from each other, and how cohorts change over time.

Fire, M. et. al “Analyzing Complex Network User Arrival Patterns and Their Effect on Network Topologies”

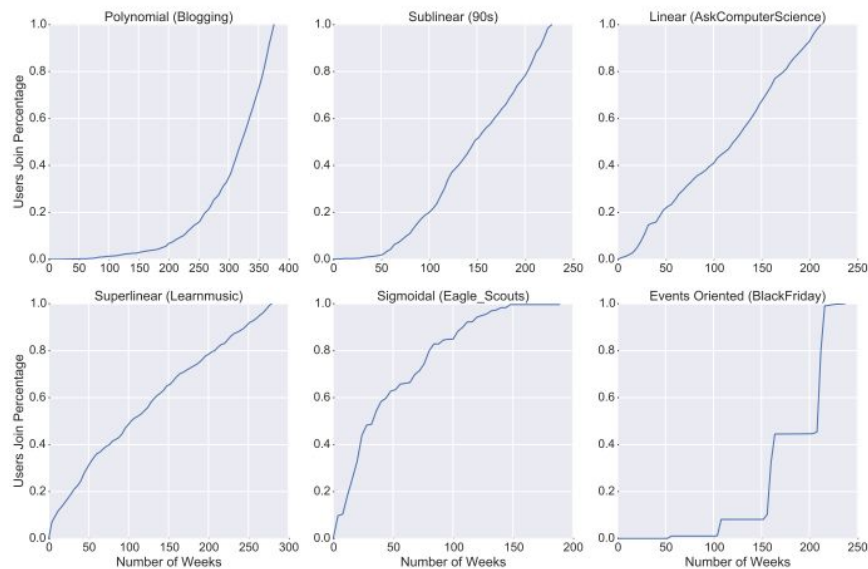
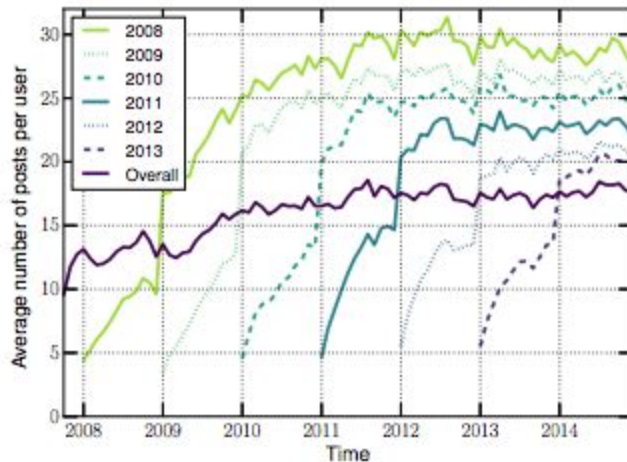


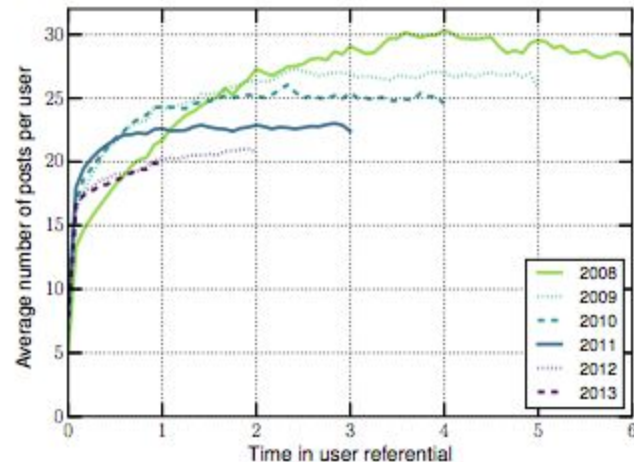
Figure 1: Common UACs of the following subreddits: Bloggng, 90s, AskComputerScience, Learnmusic, Eagle-Scouts, and BlackFriday.

Previous work

Barbosa, S. et. al. “Averaging Gone Wrong: Using Time-Aware Analyses to Better Understand Behavior”



(a)



(b)

Datasets

	Reddit (/r/math)	Hacker News	StackExchange (math)
Time	2008 - 2014	2006 - 2016	2011 - 2015
Users	38,832	251,715	66,702
Comments	326,355	9,082,312	1,358,756

Reddit

Reddit is a social website for sharing links, for discussion, and cat photos.

Founded in 2005.

Reddit

Messenger x Inbox - dmtryshmtv x Inbox - dshemetov x MAE253 Present... x Network Theory/ x Reddit data analy x centrality_cohort x reddit screenshot x top scoring links x Dmitry

https://www.reddit.com/top/

Apps tools cmxd Current Projects News Later spt 23 cal blg Trello PSX Longplay [019] To JupyterHome Evernote Web MAE253 Presentation

MY SUBREDDITS FRONT ALL - RANDOM ASKREDDIT FUNNY - PICS - TODAYILEARNED - VIDEOS - NEWS - GIFS - WORLDNEWS - GAMING - MOVIES - AWW - SHOWERTHOUGHTS - TELEVISION - MILDLYINTERESTING - JOKES - EXPLAINLIKEIMFIVE - TIFU - SCIENCE - NOTTHEONION - FUTUROLOGY - MUSIC - IAM MORE »

reddit hot new rising controversial top gilded wiki promoted

Want to join? Log in or sign up in seconds. | English

links from: past 24 hours ▼

Save \$\$. (www.ebay.com)
promoted by eBayOfficial

sponsored link what's this?

Work in IT, got this in the mail today. (imgur.com)
submitted 22 hours ago by Rick91981 to /r/funny
4553 comments share

Greek news sites are reporting that Aristotle's tomb was found. (ekathimerini.com)
submitted 6 hours ago by sotricious to /r/worldnews
1337 comments share

A baby fox was brought to our local vet. Looks almost like CGI on this pic. (i.imgur.com)
submitted 22 hours ago by teo_sk to /r/aww
512 comments share

The bottles really make it (i.imgur.com)
submitted 5 hours ago by xMeta4x to /r/pics
963 comments share

A 66-year-old hiker who was found dead two years after she went missing on the Appalachian Trail kept a journal of her ordeal after getting lost in the wilderness. (nzherald.co.nz)
submitted 16 hours ago by tehgerbil to /r/news
1974 comments share

.30 mm drill bit. (imgur.com)
submitted 16 hours ago by GrimJeeper81 to /r/mildlyinteresting
1220 comments share

Bouncers (i.imgur.com)
submitted 14 hours ago by cantbelieveitsbacon to /r/funny
712 comments share

Forgot how great Pearls Before Swine was (imgur.com)
submitted 16 hours ago by Tempo_44 to /r/funny
146 comments share

Picked my dog up from surgery today. She's a little high still. (i.redd.it/uploads.com)

discuss this ad on reddit

Submit a new link

Submit a new text post

discuss this ad on reddit

daily reddit gold goal 51%

help support reddit

Network Theory SUNN O))) - The Gri... top scoring links : red... Spotify "Inbox - dmtryshmtv...

11:39 AM

Reddit

so so good

permalink embed parent

[-] **Hellectika** 277 points 22 hours ago

That's fantastic, mister

permalink embed

[-] **eelaws** 29 points 19 hours ago

I logged in just to give you an upvote.

permalink embed parent

[-] **bullet4mv92** 69 points 17 hours ago

I bought a computer just to give you an upvote.

permalink embed parent

[-] **Nerochi** 57 points 17 hours ago

I was born just to give you an upvote.

permalink embed parent

[+] **Superlizard** 36 points 16 hours ago (3 children)

load more comments (5 replies)

[-] **kubrick23** 1 point 16 hours ago

Mister F...

permalink parent

[-] **anklefat** 0 points 15 hours ago

Ah yes. Arrested development!

permalink parent

[-] **Capt_Tommy_Bags** 114 points 21 hours ago

He is just getting cleared for duty before he gets into his starship!

permalink embed

[-] **raptor102888** 36 points 16 hours ago

Do a barrel roll!

permalink embed parent

[-] **white_android** 4 points 16 hours ago

I just fucking did a barrel roll. I'm not doing another.

permalink embed parent

[-] **nasian_** 11 points 16 hours ago

20xx begins

permalink embed parent

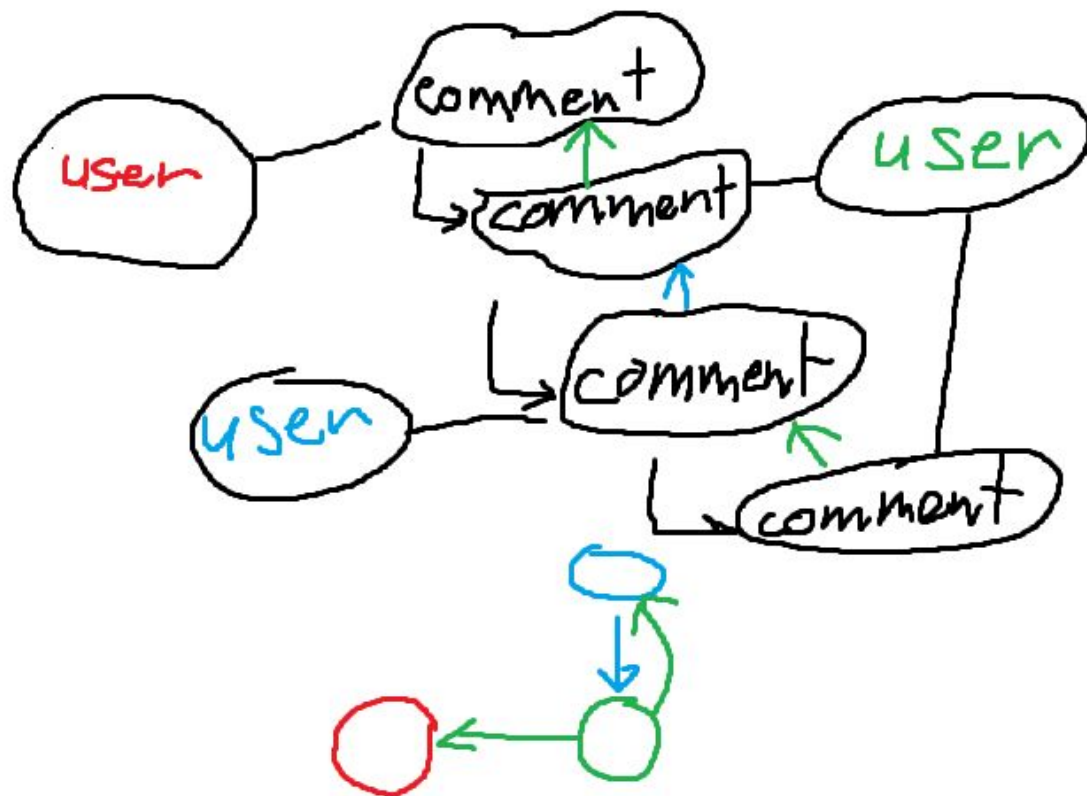
[-] **knecr99** 10 points 16 hours ago

discussions in /r/aww

205 points • 62 comments

Puppy tub

Creating a network out of comments

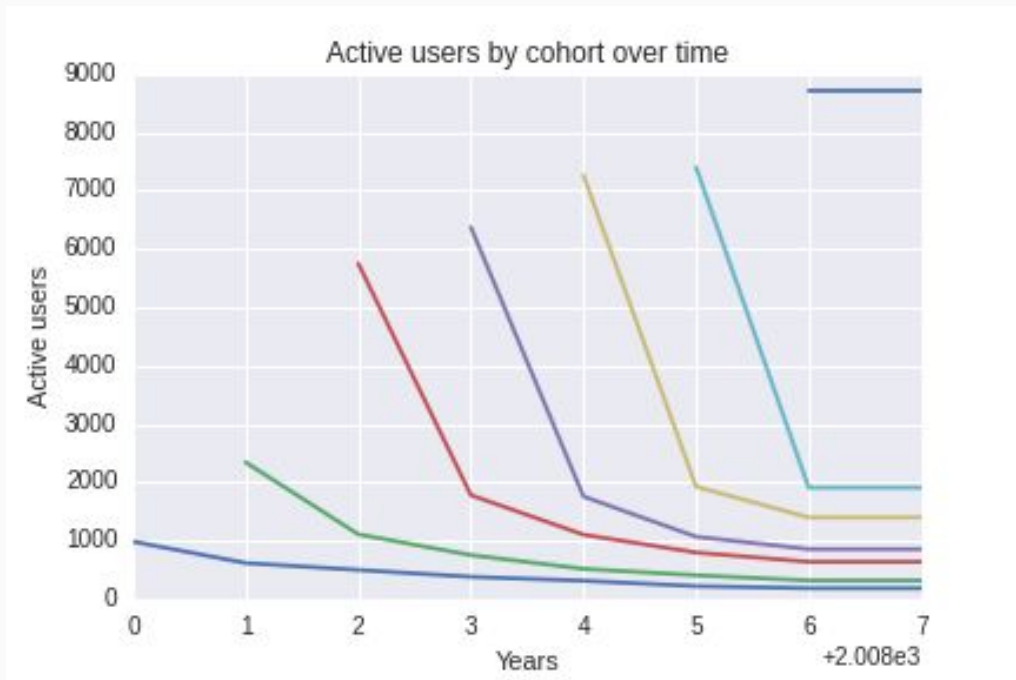


Analysis

Split the users into cohorts (communities) based on year of first comment. Create a network for every year and check the behavior statistics of each cohort in that year. Behavior statistics:

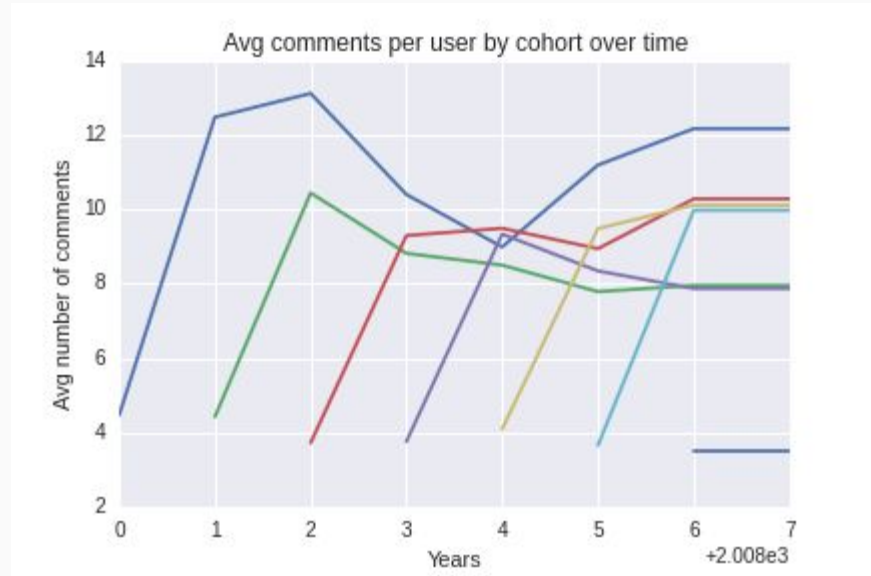
- Comments per user
- Closeness centrality

In each cohort, count the number of unique active users per year



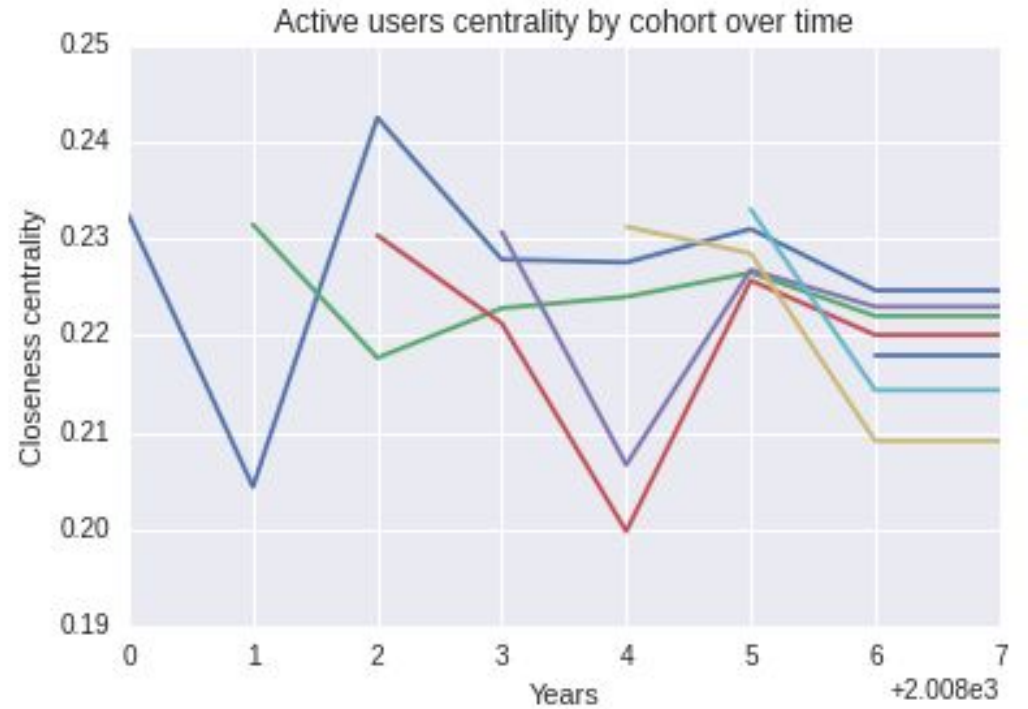
Comments per user over time

Average number of posts per user over time.



Closeness centrality

Closeness centrality: the reciprocal sum of the shortest distances from every node to a given node.



Hacker News

Hacker News is a social news website focusing on computer science and entrepreneurship. In general, content that can be submitted is defined as

“Anything that good hackers would find interesting. That includes more than hacking and startups. If you had to reduce it to a sentence, the answer might be: anything that gratifies one's intellectual curiosity.”

-[Hacker News Guidelines](#)

Netflix Will Be the Exclusive

John

← → ↺ https://news.ycombinator.com/item?id=11756381 ☆ ≡

Y Hacker News

new | threads | comments | show | ask | jobs | submit

karasinski (18) | logout

▲ Netflix Will Be the Exclusive US TV Home of Disney, Marvel, Lucasfilm and Pixar (netflix.com)

246 points by mgr86 3 hours ago | past | web | 166 comments

add comment

▲ Animats 2 hours ago

Netflix is now behaving like a channel, rather than a library. Originally, they had a huge library from which you could choose. Now, they announce "Here are some of the great and varied selection of films coming to our U.S. service in the next few months". Some of those films are decades old. Gradually, their back catalog has shrunk.

[reply](#)

▲ sseveran 2 hours ago

That is largely driven by the content creators and they don't have nearly as much control over it. They signed the Starz deal before streaming was big and when it came up for renewal Starz wanted more for catalog access than netflix was willing (maybe able) to pay. I would pay more for a bigger back catalog, but i guess I am in the minority.

[reply](#)

▲ magicalist 1 hour ago

Not to mention the content creators desperately want to avoid another Amazon/iTunes situation, with one distributor being able to dictate what the market does. They go out of their way to strategically deny licensing things completely to ensure Netflix doesn't have everything at once.

Hulu used to be their great big hope, since it's owned by several media companies, but they seem to have given up on that (even as Hulu has finally been worth using as they finally gave an option to pay and not still see ads).

[reply](#)

▲ vonklaas 44 minutes ago

Hulu is a great example of an company working super hard to un-innovate innovation. They had the distribution, rights, timing, capital and credibility to at least compete with, if not overtake netflix, but they couldn't Clayton Christiansen it. Who they fuck would want to pay, to watch selective programming...with commercials? Not only that but HBO? are they kidding. I am an HBO customer and I pirate HBO shows because they load faster and are in at least as good quality.

I used to do this to Adobe. I commend them for lowering their product price and really getting competitive/offering a tn of value. I pay \$20 for creative cloud which is totally worth it for me vs. time/aggravation of pirating.

I would've paid for hulu if I could watch shows as they aired, and anytime after, without commercials. Obviously, I can never pay for/trust them and never will, now. Netflix *really* needs a competitor...

[reply](#)

▲ ChuckMcM 1 hour ago

To some extent this is inevitable and I think ultimately it will be to their detriment. One of the things that struck me during my recent "cut the cable" exercise was that while the offerings on Amazon Prime video were significantly less to my liking than those on NetFlix, if it wasn't on prime I could "buy" it or "rent" it from the same interface almost regardless of what it was. On NetFlix things vanish and then you no longer have access, period. So its fine "in the now" but want to watch an old episode of some show that used to be there and now its off the list.

I suspect that one of the things that Netflix cannot afford using AWS is to have content available that not enough people watch. They end up paying the storage charges with no revenue.

[reply](#)

▲ loeg 1 hour ago

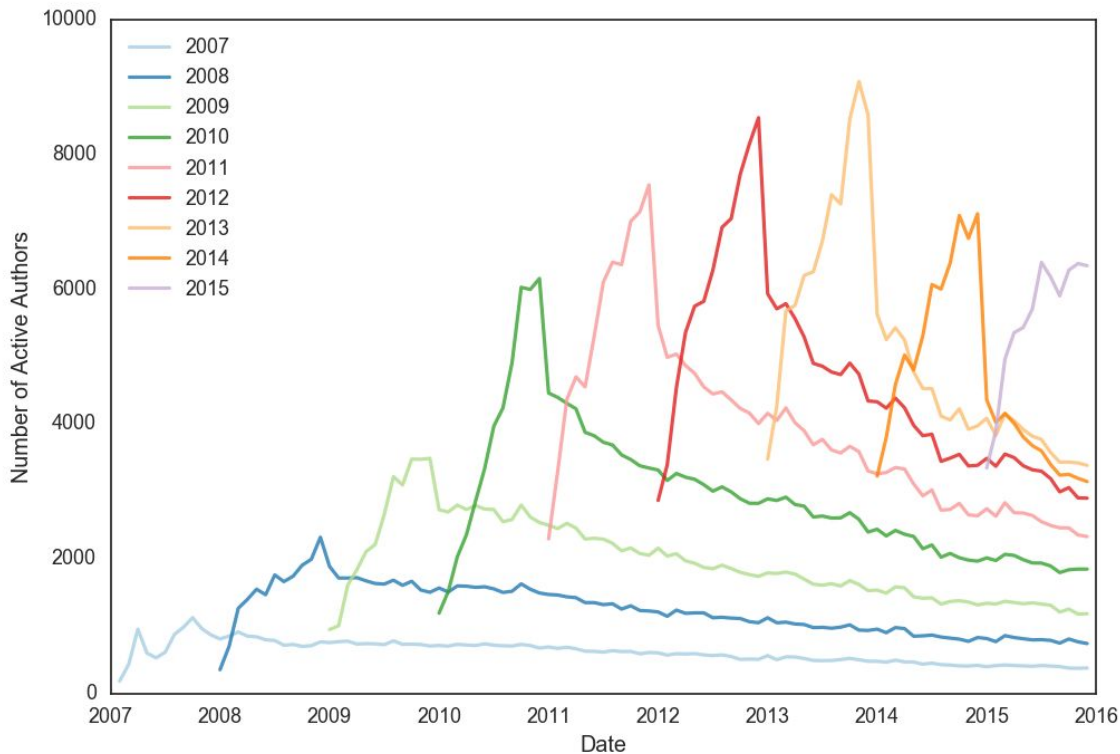
I think your first paragraph is spot on.

Active users

- Active authors by month
- An author is active if they posted at least once in a month
- Cohorts are defined by the year that an author first commented

Many users that join do not stick around for long.

Number of new users by year appears to have peaked in 2013.

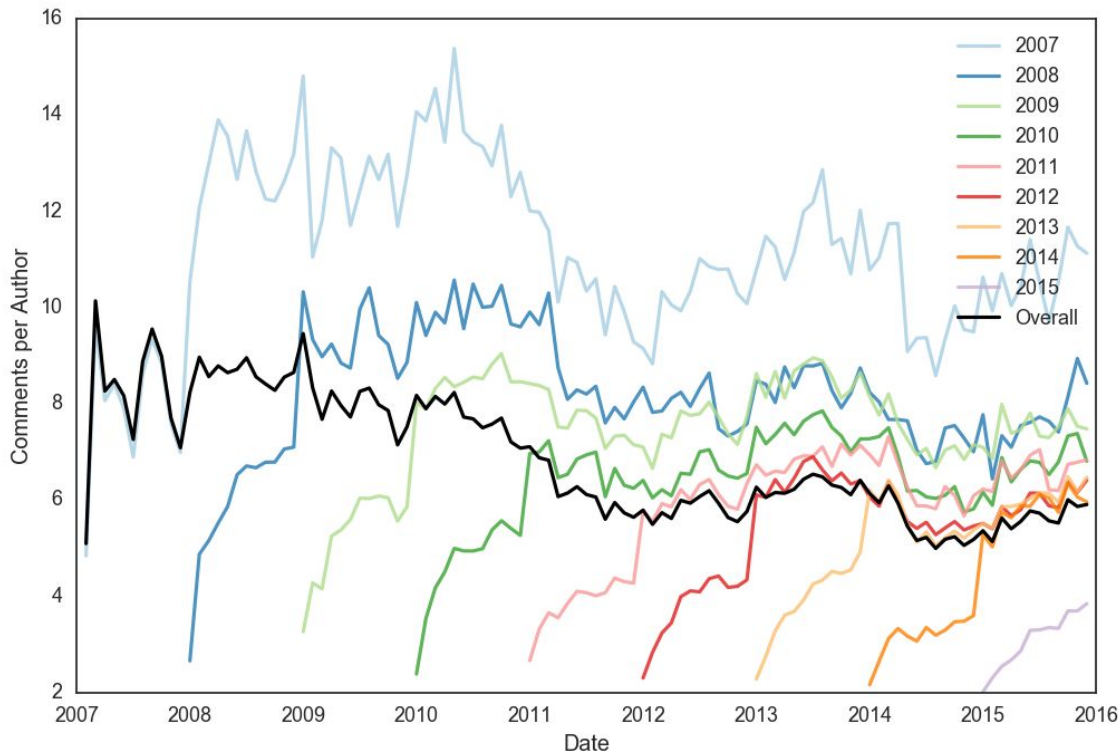


Comments per month per active user

- Overall number of comments per month per user seems to be decreasing slightly over time

Earlier cohorts post more.

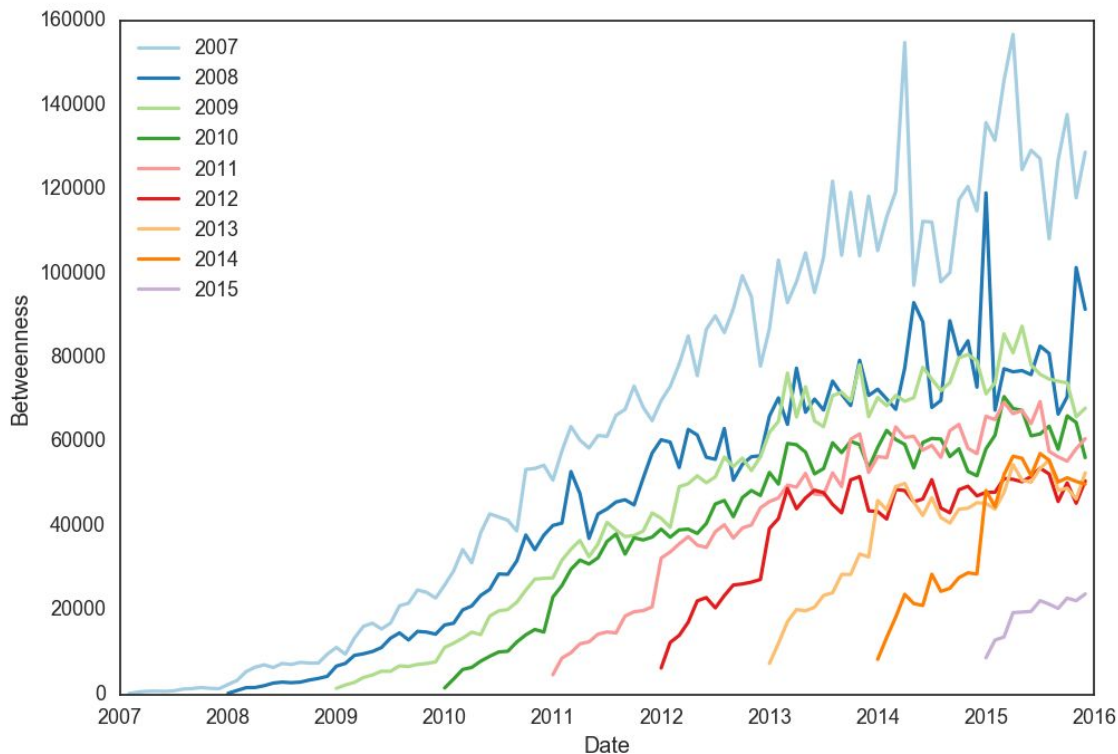
Newer cohorts take time to post more often.



Identifying Central Users

- Calculate the betweenness of the network during each month.
- Find the average betweenness for users in each cohort.

Users in older cohorts have a larger number of shortest paths from all vertices to all others that pass through their nodes.



StackExchange

StackExchange is a series of question and answer sites, each relating to a specific topic. In particular, we look at Math StackExchange, which is a forum for anyone studying math to ask and answer questions in mathematics. Users come from all levels of math.

Content ranges from introductory homework problems to open ended theoretical discussion.

Proving $1^3 + 2^3 + \dots + n^3 = \left(\frac{n(n+1)}{2}\right)^2$ using induction



How can I prove that

32

$$1^3 + 2^3 + \dots + n^3 = \left(\frac{n(n+1)}{2}\right)^2$$



8

for all $n \in \mathbb{N}$? I am looking for a proof using mathematical induction.

Thanks

(algebra-precalculus) (induction) (summation) (faq)

share cite improve this question

edited Nov 30 '12 at 8:59

community wiki
13 revs, 6 users 38%
t.b.



2

In order for you to understand the linked questions, it is maybe worth mentioning that $\sum_{k=1}^n k = \frac{n(n+1)}{2}$. I

hope you know that. – t.b. Sep 6 '11 at 2:45

2

@Steve: See [this answer](#) for general comments on induction, and [this one](#) for specific advice on doing proofs by induction. The example there may be enough for you to figure out how to prove *this* statement by induction. – Arturo Magidin Sep 6 '11 at 3:34

1

Since this question is asked quite frequently, it has been added to the list of [Generalizations of Common questions](#). It has been kept separate from the [version which does not use induction](#). – Eric Naslund Aug 30 '12 at 0:23

add a comment

7 Answers



8

You are trying to prove something of the form,

$$A = B.$$

asked 4 years ago

viewed 8851 times

active 6 months ago



Love this site?

Get the **weekly newsletter**! In it, you'll get:

- The week's top questions and answers
- Important community announcements
- Questions that need answers

Sign up for the newsletter

see an [example newsletter](#)

Linked

7

Prove that

$$1^3 + 2^3 + \dots + n^3 = (1 + 2 + \dots + n)^2$$

4

Sum of cubes proof

1

I'm having trouble with induction. Prove

$$1 + 2^3 + 3^3 + \dots + n^3 = \frac{(n^2)(n+1)^2}{4}$$

3

Prove by induction $\sum_{i=1}^n i^3 = \frac{n^2(n+1)^2}{4}$ for $n \geq 1$

3

Proof by induction help. I seem to be stuck and my algebra is a little rusty

3

Proving that two summations are equivalent:
 $\sum_{i=1}^n i^3 = (\sum_{i=1}^n i)^2$

2

Prove

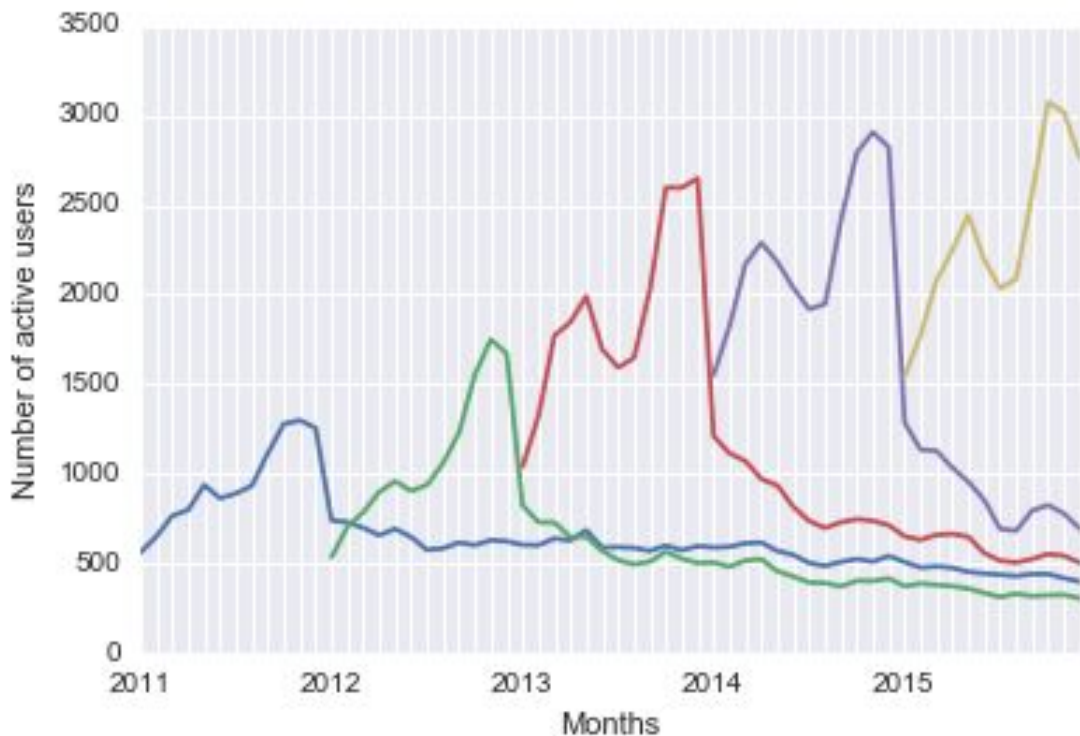
$$1^3 + 2^3 + \dots + n^3 = (1 + 2 + \dots + n)^2$$

Number of active users per month

Cohorts are determined by the year of a user's first post (when they joined).

Active users are those who have posted in a given month

- Most new users are active for about a year, and then become inactive
- Users from all cohorts tend to the same level of activity over time
- That level of activity is slowly decreasing

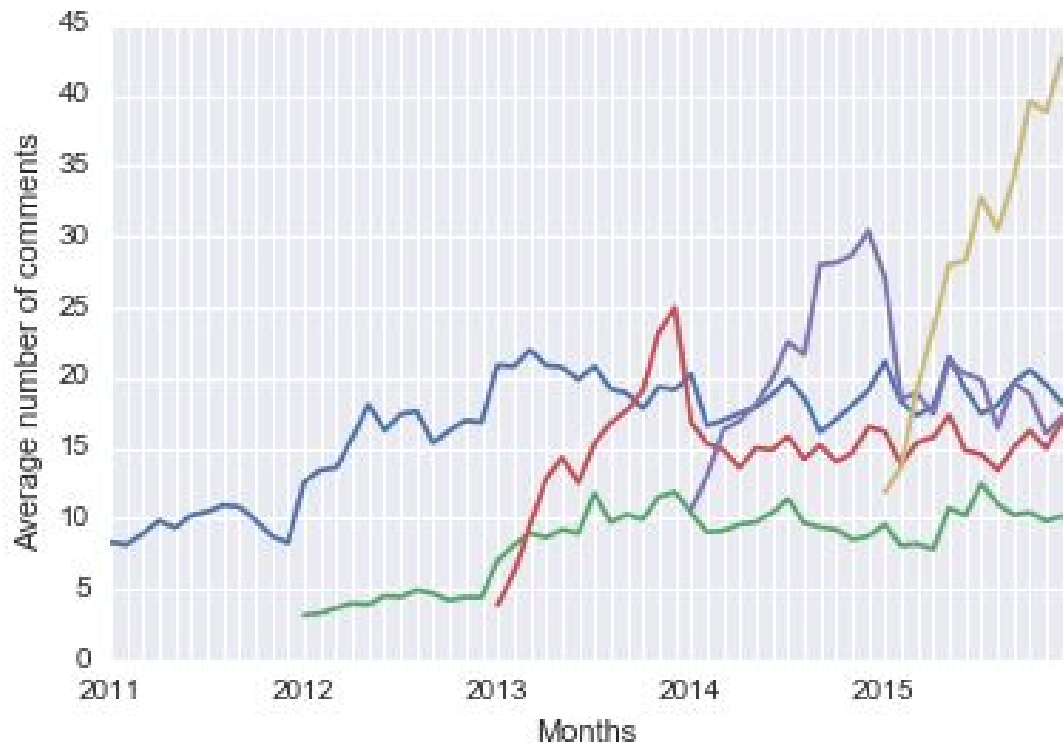


Average number of comments per month

- The average number of comments per user has decreased over time, with the exception of the most recent year.

New cohorts have a spike in activity during their first year, and then die back down gradually.

This reflects the fluctuation of active users over their first year.



Conclusions

- Old users are more active even though new users outnumber them.
- Old users have higher centrality compared to new ones.
- Old users are loyal to the system.

Future work

- Measure inter-cohort interaction
- Detecting community migration from subreddit to subreddit: TrueReddit -> TrueTrueReddit.
- It is unclear which metrics can accurately measure the “quality” of the information flowing on a network. Best bet might be to detect community splintering: find points where a piece of the community creates a separate discussion space.

Thank you!

Questions? Suggestions?

Finding Key Actors

One method to identify key actors is to plot actors' scores for Eigenvector centrality versus Betweenness. Theoretically, these metrics should be approximately linear; therefore, any non-linear outliers will be of note.

- An actor with very high betweenness but low EC may be a critical gatekeeper to a central actor
- Likewise, an actor with low betweenness but high EC may have unique access to central actors

