Reply in Zoom chat:
Describe a community you used to enjoy before it got popular. What happened to it?
We act differently in different spaces.

Norms — informal rules that govern behavior — play a massive role in determining how you act in a give space, giving the character to that socio-technical system.

Descriptive and injunctive norms operate differently, but people notice them remarkably quickly, and they are most influential when they are made salient.

Design defaults can influence norms; seeding the community can likewise set expectations.
Wikipedia’s growth

Wikipedia emerged as the leading collaboratively edited encyclopedia and experienced rapid growth.

From just a few editors to about 150,000 monthly active editors in just five years.

https://stats.wikimedia.org/v2/#/en.wikipedia.org/contributing/active-editors/normal|line|All|~total
Wikipedia’s growth and decline

…but then something changed.

https://stats.wikimedia.org/v2/#/en.wikipedia.org/contributing/active-editors/normal|line|All|~total
Wikipedia’s growth and decline

...and has continued to change.

What happened? [2min]

https://stats.wikimedia.org/v2/#/en.wikipedia.org/contributing/active-editors/normal|line|All|~total
Non-English Wikipedias: same pattern.

They’re all different sizes, so it’s not that they ran out of articles.

The peak hit at different dates, so it’s not exogenous.
So if it’s not because they ran out of content, and it’s not because they ran out of people...

What happened?
Less and less of the editing is on the pages themselves; more and more in the discussion pages. [Kittur et al. 2007]

On CNN.com, the community is becoming more and more downvote-oriented over time [Cheng et al. 2017]
Do communities get worse as they grow?

Is this decline inevitable?
Today: the challenge of growth

What changes about the dynamics of social computing systems as they grow?

What do you need to change, as a designer or community organizer, to keep a social computing system vibrant as it grows?

Topics today:

Why is growth hard?

Moderation (pt 1)

Ranking
What changes about a socio-technical system as it grows?
What happened?

Harvard undergraduates
What happened?

Anyone with a college email address
What happened?

International
What happened?

What started out narrow, necessarily broadened. New members mean new norms, culture and contestation.

Russia’s IRA
Broader participation exposes cultural rifts

Why Are Trans People Being Banned from Tinder?

The dating app won't really explain why it keeps happening.

Cis straight men reporting female-identifying trans women: trans members get auto-banned
Newcomers challenge norms

New members of the system are typically more energetic than existing members and also interested in a broader range of discussion than the community’s current focus [Jeffries 2006]

Newcomers have not been enculturated: they don’t know the norms of the system, so they are more likely to breach them [Kraut, Burke, and Riedl 2012]

…and, there are a lot of newcomers, with more constantly joining, exhausting the resources of the existing members.
Eternal September: the permanent destruction of a community’s norms due to an influx of newcomers.

Usenet, the internet’s original discussion forum, would see an influx of norm-breaking newcomers each September as college freshmen arrived on campus and got their first access to the internet.

In September 1993, America Online gave its users access to Usenet, flooding it with so many newcomers that it never recovered. It was the September that never ended: the Eternal September.

Have you ever read: “This was so much better when it was smaller”?
Surviving an Eternal September

What allows a community to stay vibrant following a massive surge in user growth?

Classic case: small subreddits getting defaulted — added to the default set for new Reddit users

Cases that survived: [Kiene, Monroy-Hernandez, Hill 2016; Lin et al. 2017]

1) Required strong moderation
2) A small % of posts now get attention

Let’s unpack these each in turn
Moderation
Scale does not come free.

To survive massive growth, moderators must step up their efforts to shepherd behavior toward the community’s desired norms.

- Removing off-content and rule-breaking content
- Banning persistent rule breakers
- Updating rules and handling angry flare-ups
Moderation

“Three imperfect solutions”

h/t Gillespie [2018]
Paid moderation

Rough estimates:

~15,000 contractors on Facebook [Statt 2018, theverge.com],

~10,000 contractors on YouTube [Popper 2017, theverge.com]

Moderators at Facebook are trained on over 100 manuals, spreadsheets and flowcharts to make judgments about flagged content.
Paid moderation

“Think like that there is a sewer channel and all of the mess/dirt/waste/shit of the world flow towards you and you have to clean it.”

- Paid Facebook moderator

Paid moderation

Strengths

A third party reviews any claims, which helps avoid brigading and supports more calibrated and neutral evaluation.

Weaknesses

Major emotional trauma and PTSD for moderators.

Evaluators may have only seconds to make a snap judgment.
Community moderation

Members of the community, or moderators who run the community, handle reports and proactively remove comments.

Examples: Reddit, Twitch, Steam

It’s best practice for the moderator team to publish their rules, rather than let each moderate act unilaterally.

Moderator Guidelines for Healthy Communities

Effective April 17, 2017.

1. Engage in Good Faith
   Healthy communities are those where participants engage in good faith, and with an assumption of good faith for their co-collaborators. It’s not appropriate to attack your own users. Communities are active, in relation to their size and purpose, and where they are not, they are open to ideas and leadership that may make them more active.

2. Management of your own Community
   Moderators are important to the Reddit ecosystem. In order to have some consistency:

   Community Descriptions:
   Please describe what your community is, so that all users can find what they are looking for on the site.
Community moderation

“I really enjoy being a gardener and cleaning out the bad weeds and bugs in subreddits that I’m passionate about. Getting rid of trolls and spam is a joy for me. When I’m finished for the day I can stand back and admire the clean and functioning subreddit, something a lot of people take for granted. I consider moderating a glorified janitor’s job, and there is a unique pride that janitors have.”

- /u/noeatnosleep, moderator on 60 subreddits including /r/politics, /r/history, /r/futurology, and /r/listentothis

[https://thebetterwebmovement.com/interview-with-reddit-moderator-unoeatnosleep/]
Contribution pyramid redux

Imagine a 10x dropoff between levels

- Lurkers
- Likers
- Commenters
- Contributors
- Mods

This is why most communities only have a few mods
Community moderation design

Community feedback: up/downvotes, flagging

Discourse

Reddit
Community moderation design: hellbanning

When people know that they’re banned, they create new accounts and try to game the system.

Instead, ban them into one of the “circles of hell”, where their comments are only able to be seen by other people in the same circle of hell.

The trolls feed the trolls.
Community moderation

Strengths:

- Leverages intrinsic motivation
- Local experts are more likely to have context to make hard calls

Weaknesses:

- Mods don’t feel they get the recognition they deserve
- Resentment that the platform makes money off free labor
- Not necessarily consistent, fair, or just
Algorithmic moderation

Train an algorithm to automatically flag or take down content that violates rules (e.g., nudity). Example via YouTube:
## Algorithmic moderation: just-in-time norm reminders

<table>
<thead>
<tr>
<th>Question</th>
<th>Votes</th>
<th>Views</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can I log in with Google and forum log in too?</td>
<td>3</td>
<td>223</td>
<td>Sep '15</td>
</tr>
<tr>
<td>What’s the coolest thing you have worked on?</td>
<td>3</td>
<td>227</td>
<td>Sep '15</td>
</tr>
<tr>
<td>Can I find stuff while writing replies?</td>
<td>2</td>
<td>231</td>
<td>Sep '15</td>
</tr>
<tr>
<td>How can I make a post Wiki?</td>
<td>1</td>
<td>109</td>
<td>Sep '15</td>
</tr>
<tr>
<td>What is a onebox?</td>
<td>10</td>
<td>355</td>
<td>Jul '15</td>
</tr>
<tr>
<td>Favorite TV show scene?</td>
<td>5</td>
<td>312</td>
<td>Jun '15</td>
</tr>
<tr>
<td>Is the Second Amendment still relevant today?</td>
<td>21</td>
<td>145</td>
<td>May '15</td>
</tr>
</tbody>
</table>
Algorithmic moderation

Examples of errors via Ali Alkhatib [2019, al2.in/street]

Google's Anti-Bullying AI Mistakes Civility for Decency

by Jillian York, motherboard.vice.com
August 18, 2017 12:00 PM

Something is wrong on the internet

by James Bridle, medium.com
November 6, 2017 10:09 AM

YouTube is still restricting and demonetizing LGBT videos — and adding anti-LGBT ads to some

by Megan Farokhmanesh, theverge.com
June 4, 2018 01:46 PM
Algorithmic moderation

Strengths:
Can act quickly, before people are hurt by the content.

Weaknesses:
These systems make embarrassing errors, often ones that the creators didn’t intend. Errors are often interpreted as intentional platform policy.

Even if a perfectly fair, transparent and accountable (FAT*) algorithm were possible, culture would evolve and training data would become out of date [Alkhatib 2019].
So...what do we do?

Many social computing systems use multiple tiers:

**Tier I: Algorithmic moderation** for the most common and easy-to-catch problems. Tune the algorithmic filter conservatively to avoid false positives, and route uncertain judgments to human moderators.

**Tier II: Human moderation**, paid or community depending on the platform. Moderators monitor flagged content, review an algorithmically curated queue, or monitor all new content, depending on platform.
Tools help facilitate moderator decisions by automatically flagging problematic posts, and providing relevant information.
Appeals

Most modern platforms allow users to appeal unfair decisions. If the second moderator disagrees with the first moderator, the post goes back up.

Instagram, 2019
More on moderation later.

Today:

- Moderation design
- Moderation algorithms

Later in the course:

- What effect does moderation actually have on a community?
- Moderation as invisible labor
- Moderation as classification
- Safe Harbor regulation
Ranking
“In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients.”  - Herb Simon, 1971
In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients.
Information overload causes attention underprovision

As Usenet groups grow in size, members (1) respond to simpler messages, (2) generate simpler responses, and (3) are more likely to leave. [Jones, Ravid, and Rafaeli 2004]

As a subreddit gets larger, its users cluster their comments around a smaller and smaller proportion of posts [Lin et al. 2017]

Fewer than half of Reddit’s most popular links get noticed and upvoted the first time they were submitted to the site [Gilbert 2013]
Designing for info overload

**Ranking**

Unintuitive mental model, but when right, a front page is helpful

Facebook
Twitter (top)
Pinterest

**Chronological**

Simple mental model but spammy accounts can dominate

Twitter
Email
Slack

Facebook
Instagram
Reddit
Spotify

iMessage
WhatsApp
Twitch
Feed ranking algorithm

1) Featurize
2) Predict
3) Rank
Featurize

- Tie strength w/ MSB: 6
- Content type: mobile phone photo
- Vision algorithm: stuffed animal
- Text features (e.g., BERT embeddings)
- Interactions so far: 101
- Platform: iPhone XoXo
- % haha reactions: 15%
- Day of year
- Age of content
- Internet: 10 mbps
Predict

Relevance score
From engagement signals: like, click, comment, share, hide, report, etc.

$$score = \sum_{s \in \text{signals}} weight_s \cdot s$$
Predict

How do we train this deep learning algorithm?

Use prior behavior on the platform

The algorithm’s loss is minimized by (“the machine learning’s goal is”) accurately predicting users’ engagement signals on past posts

The algorithm may be further fine-tuned on the user’s specific behavior, or on a learned embedding of users (e.g., socialites, jokesters, political junkies)
Rank the items in the feed by their predicted relevance score
Feed ranking algorithm

- Featurize
- Predict
- Rank

1.4
1.58
1.59
1.6
1.8
However.

The ranking algorithm will only optimize what it’s trained to optimize.

Facebook later had to add weights for improving well-being.

How should platforms optimize for long-term community health rather than short term engagement signals?
Back to the beginning
Wikipedia’s growth and decline

Returning to the original question:

What happened?

https://stats.wikimedia.org/v2/#/en.wikipedia.org/contributing/active-editors/normal|line|All|~total
Growing pains [Halfaker et al. 2012]

1. Wikipedia starts small, with little moderation needed and strongly motivated contributors

2. The formula works — Wikipedia grows

3. As Wikipedia grows, the percentage and volume of low-quality contributions rises, creating strain on the reputation of Wikipedia and invisible labor for the Wikipedia editors
Growing pains [Halfaker et al. 2012]

4. To manage the strain, Wikipedia admins stem the tide: they reject more contributions and create bots and tools to help them quickly revert bad work. [Suh et al. 2009]
Growing pains [Halfaker et al. 2012]

5. The increased rejections lead to newcomers less likely to stay.
Growing pains [Halfaker et al. 2012]

1. Start small, little moderation
2. Get popular and grow
3. Strain under newcomer contributions
4. Institute policies to reduce junk
5. Lose newcomers w/ new policies
Not just Wikipedia [TeBlunthuis et al. 2018]

Replicated across hundreds of Wikia wikis
e.g., runescape, yugioh, harrypotter, ewrestling, onepiece, clubpenguin

1. Start small, little moderation
2. Get popular and grow
3. Strain under newcomer contributions
4. Institute policies to reduce junk
5. Lose newcomers w/ new policies
So what do we do?

The temptation as a community grows is to implement new features to keep the community excited. However, I suggest that you instead focus on tools that support the community and its ability to stay upright through potentially massive norm shifts.

This means tools for recognizing and supporting the moderators and enculturating newcomers.

This means tools for empowering users to manage overwhelming amounts of content.
Summary

Growth is a double-edged sword: the rules of the playground weren’t set up for so many people.

Moderation allows social computing systems to maintain their norms in the face of massive growth.

Design can help empower this by making it easier for moderators to identify problematic behaviors and act on them.

Proportionally less content can get attention as the system grows. Feed ranking algorithms prioritize content by training on past engagement behaviors.
Assignment I
meme voting

Every social system is designed...including this evaluation.
TrueSkill comparisons

[Herbrich, Minka, and Graepel 2007]

TrueSkill is a skill rating system that was used at Xbox Live to identify skill levels for players.

It's a Bayesian generalization of the Elo chess ranking system. In it, we play “games” between pairs of options, and record which one won.

Intuitively, if you win repeatedly against another player, you should have a higher skill score.

The voting portal is open tonight. Go vote 25 times before it closes.
Assignment 2: Revive a Community

Pick a community that you were a member of that has become a ghost town

Diagnose: why did it become a ghost town? What changes to the norms, motivation, and other compositional factors would be needed to revive the community for long-term sustainability?

One week: redesign

One week: launch and participate