Growing Pains
CS 278 | Stanford University | Michael Bernstein
Group and team collaboration requires interdependence, which leads to a distinct set of design constraints and affordances.

Aiming just to replicate the experience of being there is quixotic; better to aim for beyond being there by looking for affordances unique to the digital realm.

Social translucence is a general principle for designing these systems with awareness and accountability.

If incentives are misaligned, these systems will get abandoned.
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/#/en.wikipedia.org/contributing/editors/normal|line|all|~total|monthly
Wikipedia’s growth and decline

…but then something changed.

https://stats.wikimedia.org/#/en.wikipedia.org/contributing/editors/normal|line|all|~total|monthly
Wikipedia’s growth and decline

...and has continued to change.

What happened? [2min]

https://stats.wikimedia.org/#/en.wikipedia.org/contributing/editors/normal|line|all|~total|monthly
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?

Content ranking

Designing for a global community
What changes about a socio-technical system as it grows?
Tempting POV: designs that scale

Use moderators, upvoting, report links, and algorithms to build a design that scales from 1,000 people to 1B people.

In other words, your design should not need changes as your system becomes more popular.
Why that POV fails

It’s not just the design that needs to scale, it’s also the norms.

Recall that these are socio-technical systems, so the design itself is not enough to guarantee the same experience at 1K, 1M, and 1B.

The community leaders that established and enforced norms will no longer be visible to the vast majority of users.
What happened?

Harvard undergraduates
What happened?

Anyone with a college email address
What happened?
What happened?

What started out narrow, necessarily broadened. New members mean new norms, culture and contestation.
Broader participation exposes cultural rifts

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.

Example: CS 278 as it grew
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

Successful cases: [Kiene, Monroy-Hernandez, Hill 2016; Lin et al. 2017]

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

A future lecture
Today
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

<table>
<thead>
<tr>
<th>Social Media</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>Intuitive mental model, but when right, a front page is helpful</td>
</tr>
<tr>
<td>Twitter (top)</td>
<td>Simple mental model but spammy accounts can dominate</td>
</tr>
<tr>
<td>Pinterest</td>
<td></td>
</tr>
<tr>
<td>Instagram</td>
<td></td>
</tr>
<tr>
<td>Reddit</td>
<td></td>
</tr>
<tr>
<td>TikTok</td>
<td></td>
</tr>
</tbody>
</table>

### Chronological

<table>
<thead>
<tr>
<th>Social Media</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter</td>
<td>Simple mental model but spammy accounts can dominate</td>
</tr>
<tr>
<td>Email</td>
<td></td>
</tr>
<tr>
<td>Slack</td>
<td></td>
</tr>
<tr>
<td>iMessage</td>
<td></td>
</tr>
<tr>
<td>WhatsApp</td>
<td></td>
</tr>
<tr>
<td>Twitch</td>
<td></td>
</tr>
</tbody>
</table>
Designing for info overload

**Ranking**

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

**Chronological**

Simple mental model but spammy accounts can dominate

How do you think a system should be directing attention in an overloaded community? [2min]
Global ranking

e.g., traditional Reddit “hot” ranking 🧘‍♂️

First shot: how many upvotes does it have?

e.g., 100 upvotes ranked above 10 upvotes

…but this ignores if lots of people saw it but a large % disliked it

Second shot: normalize upvotes by impressions (how many people saw it)

e.g., 10 upvotes out of 100 impressions=10%, ranked above 100 upvotes out of 10000 impressions=1%
Global ranking

Reddit also has downvote data!

Third shot: upvotes – downvotes

e.g., 10 upvotes + 1 downvote ranked above 100 upvotes + 100 downvotes

…but this ranking would never stay fresh! The most popular items of all time would never change.
Global ranking

Final shot: normalize

\[ \log \text{score} = \log(\max(\text{upvotes} - \text{downvotes}, 1)) \]

Now, decay the log score linearly over time

(Why log?)

(Because the most popular posts have orders of magnitude more upvotes than others: without the log transform, the top posts would never decay fast enough, relative to the other posts)

While this approach is relatively simple, it only works for global rankings and can’t personalize. So, Reddit has shifted toward…
Personalized feed: machine learning

1) Featurize
2) Predict
3) Rank
Featurize

- Tie strength w/ MSB: 6
- Content type: mobile phone photo
- Vision algorithm: stuffed animal, bear
- 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

\[ \text{score} = \sum_{s \in \text{signals}} \text{weight}_s \cdot s \]

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

“The app is able to map a user's preferences in relation to similar users and group them into "clusters." Simultaneously, it also groups videos into "clusters" based on similar themes, like "basketball" or "bunnies."

Using machine learning, the algorithm serves videos to users based on their proximity to other clusters of users and content that they like. TikTok's logic aims to avoid redundancies that could bore the user, like seeing multiple videos with the same music or from the same creator.”

However.

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

And what it’s trained to optimize, e.g., likes, prioritizes short-term dopamine hits over long-term wellbeing or community health.

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?
Another side effect: filter bubbles

Filter bubbles occur when everyone is shown only content that they like. This happens as a natural outcome of optimizing for likes.

Example: YouTube recommendation radicalization: channels that are slightly less mainstream become recommendation gateways to more and more radical channels [Ribeiro et al. 2020]
Techniques for designing for a global community
How do you build empathy?

How can you build empathy with a huge number of communities? How do you prevent yourself from designing for your own prototypical user?

One approach successfully used in product teams: show user videos to engineers. Bring in stakeholder groups for participatory design. Don’t assume you can. Instead, create local governance (e.g., subreddits) and be responsive to it.
How do you test new ideas?

How do you A/B test new ideas, when there’s no easy way to bucket people into group A or B? Everyone's connected…

The most common answer is country comparisons, where versions are launched to different countries that have similar properties.

   e.g., launch one version in New Zealand and another in Australia

Want an advanced answer? Go chat with Johan Ugander in MS&E:

Graph Cluster Randomization: Network Exposure to Multiple Universes

Johan Ugander  Brian Karrer  Lars Backstrom  Jon Kleinberg
But more than those...

Michael suggests that first, rather than building new features, focus on tools that support the community and its ability to stay upright. This means tools for managing counter-norm behavior and enculturating newcomers.
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 the Wikipedia editors
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]
5. The increased rejections leads newcomers to be 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
Assignment 3: Feed Me

Programming assignment, in pairs

Two weeks to complete: 2x the size

Try your hand at feed ranking, using vote data from the Going Viral memes

Global feed

Personalized ranked feed

Personalized ranked feed where groups will leave the platform if too much content distasteful to them is ranked highly for others
Summary

Growth is a double-edged sword. It's great that lots of people want to play in the same playground, but the rules of the playground weren't set up for so many people.

Proportionally less content gets attention as the system grows. Feed ranking algorithms draw on machine learning to find content that each member is interested in.

Smash that like button!
Social Computing
CS 278 | Stanford University | Michael Bernstein

Creative Commons images thanks to Kamau Akabueze, Eric Parker, Chris Goldberg, Dick Vos, Wikimedia, MaxPixel.net, Mescon, and Andrew Taylor.

Slide content shareable under a Creative Commons Attribution-NonCommercial 4.0 International License.