

Chapter 26. User-Generated Content: Lines in the Sand

Content Upload Success

(Content upload success also applies to two-sided marketplaces)

If there's an action on your site that you want users to take because it's key to success, it has a funnel you can track and optimize. On Facebook, for example, sharing photos is one of the most common things users do. In 2010, Facebook's Adam Mosseri revealed some data on how Facebook's photo upload funnel worked:^[126]

- 57% of users successfully find and select their photo files.
- 52% of users find the upload button.
- 42% successfully upload a picture.

Success can be a complicated thing to define. For example, 85% of users chose only one picture for an album, which wasn't good for the way Facebook organized pictures. So the developers added another step that allowed users to select more than one picture more easily. After the change, the number of single-picture albums dropped to 40%.

Bottom Line

There's no clear number, but if a content generation function (such as uploading photos) is core to the use of your application, optimize it until all your users can do it, and track error conditions carefully to find out what's causing the problem.

Time on Site Per Day

(Time on site per day also applies to media sites)

There's a surprisingly consistent rule of thumb for social networks and UGC websites. Across many companies we polled, the average time on site per day seemed to be 17 minutes. This number was mentioned several times by companies participating in the TechStars accelerator program at a recent demo day; it's also what reddit sees for an average user. One study showed that Pinterest users spend 14 minutes on the site each day, Tumblr users spend 21 minutes a day, and Facebook users spend an hour a day on the site. ^[127]

Bottom Line

You'll have a very good indicator of stickiness when site visitors are spending 17 minutes a day on your site.

Reddit Part 1—From Links to a Community

From humble beginnings as a startup in the first cohort of Paul Graham's Y Combinator accelerator, reddit has grown to be one of the highest-traffic destinations on the Web.

Reddit began as a simple link-sharing site, but over the years it's changed significantly. "A lot of features were just us sitting down and thinking, 'what would be cool to have?'" says Jeremy Edberg, who was reddit's first employee and ran infrastructure operations. "When the site first launched, it was just for sharing and voting on links. The idea to add comments was pretty much because [reddit co-founder] Steve Huffman decided he wanted to comment on some links."

Even after commenting was enabled, there was no way to start a discussion within reddit itself. So users found ways to do this themselves. The comment threads became discussions in their own right. Seeing this, the team added a feature, called *self-posts*, that let someone start a conversation without linking elsewhere on the Web. "When we first did [self-posts], it was pretty much just a response to things users were already doing using hacks, so we decided to make it easier," says Jeremy. This is a great example of what Marc Andreessen says: "In a great market—a market with lots of real potential customers—the market *pulls* product out of the startup."^[128] Self-posts have since become a cornerstone of the site, creating a community of users who interact with one another. "Today, more submissions are self-posts than not."

Reddit has an engaged, passionate community, and it's perfectly designed to collect feedback. "The entire site is set up for giving feedback, which makes it very easy for the users to give direct feedback and for the company to know which feedback is important," says Jeremy. But he cautions that it's not enough to listen to users—you have to watch what they do. "Direct feedback, even on reddit, is usually not an accurate depiction of how users actually feel. The phrase 'actions speak louder than words' applies just as much to business as anything else. Your users' *actions* should drive your business."

Summary

- Reddit pivoted from simple link sharing to commenting to a platform for

moderated, on-site discussions by watching how users were using what it had built.

- Despite copious feedback from vocal users, the real test was what users were actually doing.

Analytics Lessons Learned

While it's important not to overbuild beyond your initial feature set or core function—in reddit's case, link sharing—a thriving community will pull features out of you if you know how to listen. Reddit included only basic functionality, but made it easy for users to extend the site, then learned from what was working best and incorporated it into the platform.

Engagement Funnel Changes

Leading web usability consultant Jakob Nielsen once observed that in an online population, 90% of people lurk, 9% contribute intermittently, and 1% are heavy contributors.^[129] His numbers suggest that there are power laws at work in engagement funnels. These patterns predate the Web—they occurred in online forums like CompuServe, AOL, and Usenet. **Table 26-1** shows some of his estimates.

Table 26-1. Jakob Nielsen's engagement estimates

Platform	Lurkers	Occasional	Frequent
Usenet	?	580,000	19,000
Blogs	95%	5%	0.1%
Wikipedia	99.8%	0.2%	0.003%
Amazon reviews	99%	1%	Tiny
Facebook donation app	99.3%	0.7%	?

Nielsen has a number of approaches for moving lurkers toward participation, including making it easier to participate and making participation an automatic side-effect of usage. For example, if you have a link-sharing site, you might time

side-effect of usage. For example, if you have a link-sharing site, you might time how long it takes a user to return from viewing a link and use that as a measurement of the link's quality—the user wouldn't have to rate the link. Any attempt to optimize contribution and engagement would then become a hypothesis for testing.

Nielsen's ratio is changing as web use becomes part of our daily lives. A 2012 BBC study of online engagement showed that 77% of the UK's online population is participating online, partly due to the ubiquity of the Internet as a social platform and how easy it is to participate lazily, by uploading a picture or updating a status.^[130]

The Altimeter Group's Charlene Li has done a lot of research into engagement. Her engagement pyramid details several kinds of user engagement. In her book *Open Leadership* (Jossey-Bass), she cites the 2010 Global Web Index Source, which surveyed web users from various countries about the kinds of activities in which they engaged online.^[131] Roughly 80% of respondents consumed content passively, 62% shared content, 43% commented, and 36% produced content. (See **Table 26-2**.)

Table 26-2. Engagement by country

	China	France	Japan	UK	USA
Watchers: Watch video, listen to a podcast, read a blog, visit a consumer review site or forum.	86.0%	75.4%	70.4%	78.9%	78.1%
Sharers: Share videos or photos, update social network or blog.	74.2%	48.9%	29.2%	61.8%	63.0%
Commenters: Comment on a news story, blog, or retail site.	62.1%	35.6%	21.7%	31.9%	34.4%
Producers: Write a blog or news story, upload a video.	59.1%	20.2%	28.0%	21.1%	26.1%

The difference between countries is notable—more than half of Chinese web users produced their own content, but only 20% of French and English respondents did. Clearly, “normal” engagement is dependent on user culture.

Participation, then, is tied to cultural expectations and the purpose of the platform. Facebook has a high engagement rate from its users because their

interactions are highly personal, and users upload to Flickr because, well, that's where their pictures live. But highly directed participation (like writing a Wikipedia entry, or posting a product review) that isn't the central reason for the platform to exist remains elusive for many startups.

The BBC's model breaks users down into four groups:

- 23% of Internet users are passive, choosing only to consume
- 16% of users will react to something (voting, commenting, or flagging it)
- 44% will initiate something (posting content, starting a thread, etc.)
- 17% of users are contributing intensely, doing something even when it's difficult or not core to the platform, such as reviewing a book on an e-commerce site

A thread on reddit that discussed user engagement on the site had some interesting numbers.^[132] One user posted that he'd submitted a picture that received 75,000 views in 24 hours on Imgur. The topic itself had 1,347 up-votes, 640 down-votes, and 108 comments. That suggests a 2.5% "easy" engagement and a 0.14% "difficult" engagement.

Jeremy Edberg says that in 2009 reddit's user contribution followed the 80/20 rule seen on many UGC sites; that is, 20% of users were logged in and voting, and 20% of *those* were commenting. While the site's behavior has shifted significantly as it has become more social and community-oriented, the percentage of visitors who comment is still small.

Even lurking, disengaged visitors may be doing something. A 2011 study from MIT's Sloane School of Management suggests that many of them share passively, via channels you don't see, such as email or conversations elsewhere.^[133] Yammer says that over 60% of its users subscribe to a regular digest of activity, which means the company has permission to reach them.^[134]

Bottom Line

By our estimates, expect 25% of your visitors to lurk, 60–70% of your visitors to do things that are easy and central to the purpose of your product or service, and 5–15% of your users to engage and create content for you. Among those

engaged users, expect 80% of your content to come from a small, hyperactive group of users, and expect 2.5% of users to interact casually with content and less than 1% to put some effort into interaction.

Reddit Part 2—There's Gold in Those Users

Once reddit had pivoted from link sharing to a community, it had engaged users, but it still wasn't making money, sometimes struggling to pay for enough infrastructure to handle its growing traffic load. While advertising was a possible source of revenue, it came at the expense of user satisfaction. Enough of reddit's users employed ad-blocking software on their browsers that reddit even ran the occasional ad thanking people for not using it.

Then the company found an alternate source of revenue: donations. "Users would constantly joke that such-and-such a feature is only available via reddit gold," says Jeremy Edberg. "At some point, our parent company came to us and asked us to think of ways to increase our revenue (which, to their credit, was something that took three years for them to ask). We thought, 'Hey, let's make this reddit gold thing real.'"

The team added the ability to buy "gold," which didn't really have any effect beyond bragging rights. "When it launched, the only benefit you got was access to a secret forum and an (electronic) trophy. We didn't even have a price—we asked people to pay what they thought it was worth. One person paid \$1,000 for a month of reddit gold, some paid a penny," says Jeremy. "But the average was right around \$4, which is how we set the price."

Over time, reddit gold users got early access to new features. As dedicated users, they were more likely to provide useful feedback—and the limited number of people using the new feature shielded servers from heavy load.

Eventually, reddit added the ability to gift gold to others, and reward good posts with a donation of gold. While the company hasn't disclosed the revenue it makes from gold, it's a significant part of its income, and it's taken steps to build it into the site. "We also realized people were buying gold for others as a way of 'tipping' for great content, so we made that easier to do," says Jeremy.

Summary

- Despite healthy user growth, reddit wasn't paying its bills and was constantly skimping on new infrastructure.
- Building on considerable goodwill and user feedback, the team tried a

donation model that fit the tone and culture of the community.

- They analyzed the results of a “pay what you will” campaign to set pricing.
- Once they saw some success, they found ways to make donation easier and expand how it was used.

Analytics Lessons Learned

Remember the business model flipbook: just because you’re a UGC business doesn’t mean your revenue must come from ads. Wikipedia and reddit both generate revenue from their community, and it helps them stay true to their culture and retain their users.

Spam and Bad Content

UGC sites thrive because they have good content. For many of the UGC companies we spoke with—such as Community Connect and reddit—fraudulent content is a very real problem that requires constant analysis and a significant engineering investment. In addition to algorithms and machine heuristics, companies like Google and Facebook pay people full-time to screen content for criminal or objectionable material, which can be a grueling job.^[135] Jeremy Edberg estimates that 50% of reddit’s development time focused on stopping spam and vote cheating—although for the first 18 months of the site’s life, user voting was enough to block all spam, and there was no spam protection in place.

Spammers often create one-time accounts, which are easy to spot. Hijacked accounts are harder to pinpoint, but most UGC sites allow users to flag spammy content, which makes it easier to review. But despite the promise of a self-policing community, users aren’t a good way to find bad content. Many of the posts flagged on reddit were actually spammers flagging everyone *else* in the hopes of boosting their own content. At reddit, “we had to build a system to analyze the quality of the reports per user (how many reports ultimately turned into verified spam),” says Jeremy.

At reddit, automated filters, along with moderators, catch most of the spam—which, in 2011, represented about half of all submitted content. “That 50% comes from far less than 50% of the users,” says Jeremy. “Pretty much the way

all the anti-cheating was developed was by finding a case of a cheater who was successful, analyzing why they were successful, finding other examples in the corpus, and then developing a model to find that type of cheating.”

Ultimately, spam suggested the site’s advertising revenue model, too. “We figured spammers were trying to get their links seen though cheating; why not just let them pay and then make it obvious they paid?” recalls Jeremy. “If you look at the sponsored link today, you’ll see that the styling and execution is almost identical to how Google highlighted sponsored links around 2008.”

Bottom Line

Expect to spend a significant amount of time and money fighting spam as you become more popular. Start measuring what’s good and bad, and which users are good at flagging bad content, early on—the key to effective algorithms is a body of data to train them. Content quality is a leading indicator of user satisfaction, so watch for a decline in quality and deal with it before it alienates your community.

[126] <http://blog.kissmetrics.com/analytics-that-matter-to-facebook/>

[127] <http://tellemgrodypr.com/2012/04/04/how-popular-is-pinterest/>

[128] <http://www.stanford.edu/class/ee204/ProductMarketFit.html>

[129] See http://www.useit.com/alertbox/participation_inequality.html, which has a number of excellent tips for improving participation inequality.

[130] http://www.bbc.co.uk/blogs/bbcinternet/2012/05/bbc_online_briefing_spring_201_1.html

[131] Global Web Index Wave 2 (January 2010), trendstream.net.

[132] http://www.reddit.com/r/AskReddit/comments/bg7b8/what_percentage_of_redditors_are_lurkers/

[133] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1041261

[134] <http://blog.yammer.com/blog/2011/07/your-community-hidden-treasure-lurking.html>

[135] <http://www.buzzfeed.com/reghan/tech-confessional-the-googler-who-looks-at-the-wo>