Should We Reduce the MSI Weight of Sticker Comments?

tl;dr

- Sticker comments receive **66%** fewer comment replies than short text comments (<=5 tokens)
- Sticker comments are only about **3%** less meaningful on average than short text comments for both the giver and receiver
- Receiving a sticker comments results in about **6%** fewer sessions in the following 4 hours vs. short text comment
- We should consider reducing the weight of sticker comments from the current value (15) to something between reactions (5) and short text comments (15) in the range of **10-12**

Introduction

To date, stickers have been treated the same in MSI as short text comments (<=5 tokens), despite clear qualitative differences. Stickers often do not contain any text and have the appearance of over-sized emojis. They also take up quite a bit more screen real estate when rendered in a comment thread. Stickers may be less likely to be part of a conversation and instead tend to serve a role similar to reactions (supporting research).

Presently, stickers comprise **5.7%** of all nmMSI and **9%** of all comment MSI. Stickers account for **9.5%** of all comments left on FB, but are skewed particularly towards India, where **20%** of all comments are stickers vs **3%** in the US. In countries where stickers are very popular, such as India, we see much less reaction usage (reactions are **8%** of all likes+reactions+comments) relative to countries, such as US, where sticker usage is minimal (reactions are **30%** of all likes+reactions+comments), suggesting that stickers and reactions serve similar needs and a given market tends to choose one vs. the other.

Due to Comment Guide, usage of stickers has increased by 80% world-wide and more than tripled in the US and India. Guide has also increased total comments by 3.3% (9% in India). However, visitation metrics have remained flat to slightly neutral on both comment giver and receiver side Comment Guide holdouts, which is inconsistent with our expectation that more comments=more sessions, and suggests that we should re-consider the MSI weight of sticker comments to incentivize teams to focus more on other types of interactions, such as text comments.

Analysis

Analysis

METHODOLOGY

In all cases below, we are comparing user-to-user (U2U) sticker comments to U2U short text comments (<=5 tokens) w/o mentions.

Propensity weighting was done by first training a GBDT classifier model. The class probabilities were then used to re-weight the data based on the likelihood of receiving the "treatment" (e.g. leaving or receiving a sticker comment). More details can be found in this note.
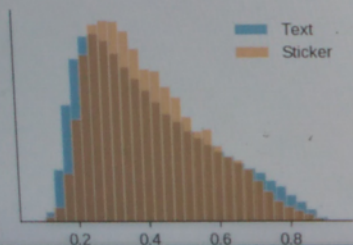
COMMENT REPLY RATE

One way to measure the value to the FB ecosystem of different comment types is reply rate. Comments with a higher reply rate will more strongly feed into the feedback flywheel and therefore have a larger net benefit to the FB engagement ecosystem.

After propensity weighting to correct for cofounders, we find that sticker comments receive **0.12** replies vs **0.37** for short text comments, on average, or **66%** less.
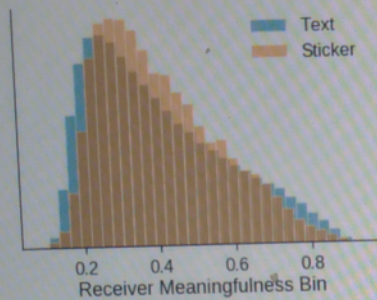
COMMENT MEANINGFULNESS

Another approach is to consider how meaningful a given comment is to both the giver and receiver of that comment. The Feed Relevance team has invested significantly in modeling the meaningfulness of comments, trained on survey data.

Interestingly, comment meaningfulness does not appear to differ much between sticker and short text comments, with average meaningfulness for comment givers and receivers being **0.45** for sticker comments and **0.47** for short text comments (after propensity weighting), a difference of only about **3%**. For comparison, the un-weighted average meaningfulness score is **0.42** for both stickers and text on the receiver-side and **0.47** for stickers vs. **0.45** for short text on the giver-side.

Looking at the histogram (left) of per-comment receiver meaningfulness values, we see significant overlap between short text (blue) and sticker (orange) comments in the distribution of comment meaningfulness, although there is greater variance for short text comments, with more high and low values than sticker comments.

Given that the propensity-weighted delta in meaningfulness is smaller than expected for sticker vs. text, I trained a GBDT regression model which was able

Looking at the histogram (left) of per-comment receiver meaningfulness values, we see significant overlap between short text (blue) and sticker (orange) comments in the distribution of comment meaningfulness, although there is greater variance for short text comments, with more high and low values than sticker comments.

Given that the propensity-weighted delta in meaningfulness is smaller than expected for sticker vs. text, I trained a GBDT regression model which was able to explain ~**33%** of the observed variance in comment meaningfulness.

Looking at the coefficient weights (see table below), after the post being a fundraiser, the closeness multiplier seems to be very strongly associated with comment meaningfulness, along with other details of the post such as it containing photos or being an MLE (major life event). Comment stickers in fact tend to be slightly positively associated with meaningfulness, perhaps because they are often left as a "reaction" on MLE posts.

These results suggest that meaningfulness as measured by the comment meaningfulness model is more related to the relationship between the two people interacting and the content being interacted with, as opposed to the details of the interaction itself. For example, if grandma leaves you a sticker comments vs. a short text comment on a picture of your recent wedding, you are likely to find both nearly equally meaningful.

	Coefficient
content_type_fundraiser_donation	0.201602
interactor_content_creator_msi_multiplier	0.095081
post_content_type_multimedia post	0.072564
content_type_life_event	0.062583
has_sticker	0.042215
content_type_post	0.037533
content_type_photo	0.034885
post_content_type_photo post	0.034481
post_content_type_other	0.032077
has_gif	0.030191
has_photo	0.029659
content_type_photo_album	0.020473
post_content_type_video post	0.018221
significance_significant	0.009490
content_type_video	0.007270
interaction_mod_newsfeed	0.005999
interactor_poster_relationship_self	0.005212
interaction_mod_story_view_from_newsfeed	-0.009235
interaction_mod_story_view	-0.009292
post_content_type_text post	-0.017000
interaction_interface_other_interface	-0.021262
interaction_interface_iphone_native	-0.021306
post_content_type_reshare.other	-0.023595
post_content_type_question	-0.025956
interactor_poster_relationship_group_members	-0.026142
interaction_interface_fblite_android	-0.036725
has_attachment	-0.045540
content_type_og_post	-0.048087
post_business_segment_other	-0.057132
post_business_segment_Wall Posts	-0.060901
interactor_poster_relationship_none	-0.071533
commenter_country_US	-0.073113
interactor_poster_relationship_page_follower	-0.073507
post_business_segment_Broadcast (reshares)	-0.084954
post_business_segment_Page Posts	-0.101463
commenter_country_MX	-0.106376

	Coefficient
content_type_fundraiser_donation	0.201602
interactor_content_creator_msi_multiplier	0.095081
post_content_type_multimedia post	0.072564
content_type_life_event	0.062583
has_sticker	0.042215
content_type_post	0.037533
content_type_photo	0.034885
post_content_type_photo post	0.034481
post_content_type_other	0.032077
has_gif	0.030191
has_photo	0.029659
content_type_photo_album	0.020473
post_content_type_video post	0.018221
significance_significant	0.009490
content_type_video	0.007270
interaction_mod_newsfeed	0.005999
interactor_poster_relationship_self	0.005212
interactor_poster_relationship_friend	0.002421
comment_type_threaded_comment	0.001384
has_text	0.000139

interaction_mod_story_view_from_newsfeed	-0.009235
interaction_mod_story_view	-0.009292
post_content_type_text post	-0.017000
interaction_interface_other_interface	-0.021262
interaction_interface_iphone_native	-0.021306
post_content_type_reshare.other	-0.023595
post_content_type_question	-0.025956
interactor_poster_relationship_group_members	-0.026142
interaction_interface_fblite_android	-0.036725
has_attachment	-0.045540
content_type_og_post	-0.048087
post_business_segment_other	-0.057132
post_business_segment_Wall Posts	-0.060901
interactor_poster_relationship_none	-0.071533
commenter_country_US	-0.073113
interactor_poster_relationship_page_follower	-0.073507
post_business_segment_Broadcast (reshares)	-0.084954
post_business_segment_Page Posts	-0.101463
commenter_country_MX	-0.106376
post_business_segment_Broadcast (original)	-0.108344
commenter_country_other_country	-0.110043
commenter_country_TH	-0.111815

OWNSTREAM SESSIONS

DOWNSTREAM SESSIONS

Finally, let's take a look at how many downstream sessions people have after receiving a sticker vs short text comment. We attribute any session in the 4 hours following receiving either a short text or sticker comment to that comment, and again use propensity weighting to compare total average attributed sessions for each comment type.

We find that the weighted average difference in downstream sessions for sticker vs short text comments to be -6% (vs +16% before weighting). Again, we find only a relatively small difference between the comment types, but it may be enough to explain some or all of the slight sessions regression we see in the Comment Guide long-term holdout. Given that notifications are delivered deterministically in jewel, it perhaps is not surprising that the the difference in sessions is relatively slight. However, if we do decide to start treating notifications from sticker comments differently than text comments, a larger difference in sessions may become apparent. For example, jewel notifications from Guided sticker comments have a CTR of 58% vs 64% for non-sticker Guided comments.
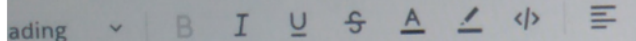
Discussion

The analysis presented above does not definitively point to the correct weight for sticker comments. It is clear that stickers provide value to both those people giving and receiving them, and data suggests that in some countries (such as India) stickers may play the role that reactions do in countries such as the US. Additionally, metrics such as reply rate and sessions suggest that a short text comment does indeed have more overall ecosystem value than a sticker comment.

So what should be the weight for stickers? In prior discussions, people have suggested a base weight equivalent to reactions (5 vs. the current 15). However, sticker comments do have some clear benefits vs. reactions, primarily the ability for people to reply and react to them. Additionally, people can now combine stickers with text in a single comment, in which case we should likely treat these sticker+text comments the same as text comments. Finally, should we also reduce the weight for stickers that are essentially a stylized short text comment (e.g. Happy Birthday)?

Given that there is some obvious daylight between reactions and sticker comments, it seems more appropriate to assign sticker comments a weight in-between that of reactions and short text comments (e.g. 10-12). Doing so will still incentivize teams (particularly Comment Guide) to skew more towards preserving and growing text comments, but will still acknowledge that there is a difference between stickers and reactions.

Going forward, I do think we should invest in understanding how people in sticker-heavy markets (e.g. India) use and perceive sticker comments vs. other types of interactions. Do they see these as a substitute for reactions? If yes, should we consider alternative UX treatments for stickers that reduce their prominence in the app relative to text comments? Or do they seem the more like short text comments? Generally, do people make a distinction between stickers containing text vs. those that do not? Regardless of where we land on the MSI weight for stickers, there are still many questions to be answered, all with meaningful product implications.

attribute any session in the 4 hours following receiving either a short text or sticker comment to that comment, and again use propensity weighting to compare total average attributed sessions for each comment type.

We find that the weighted average difference in downstream sessions for sticker vs short text comments to be **-6%** (vs +16% before weighting). Again, we find only a relatively small difference between the comment types, but it may be enough to explain some or all of the slight sessions regression we see in the Comment Guide long-term holdout. Given that notifications are delivered deterministically in jewel, it perhaps is not surprising that the the difference in sessions is relatively slight. However, if we do decide to start treating notifications from sticker comments differently than text comments, a larger difference in sessions may become apparent. For example, jewel notifications from Guided sticker comments have a CTR of 58% vs 64% for non-sticker Guided comments.

Discussion

The analysis presented above does not definitively point to the correct weight for sticker comments. It is clear that stickers provide value to both those people giving and receiving them, and data suggests that in some countries (such as India) stickers may play the role that reactions do in countries such as the US. Additionally, metrics such as reply rate and sessions suggest that a short text comment does indeed have more overall ecosystem value than a sticker comment.

So what should be the weight for stickers? In prior discussions, people have suggested a base weight equivalent to reactions (5 vs. the current 15). However, sticker comments do have some clear benefits vs. reactions, primarily the ability for people to reply and react to them. Additionally, people can now combine stickers with text in a single comment, in which case we should likely treat these sticker+text comments the same as text comments. Finally, should we also reduce the weight for stickers that are essentially a stylized short text comment (e.g. Happy Birthday)?

Given that there is some obvious daylight between reactions and sticker comments, it seems more appropriate to assign sticker comments a weight in-between that of reactions and short text comments (e.g. **10-12**). Doing so will still incentivize teams (particularly Comment Guide) to skew more towards preserving and growing text comments, but will still acknowledge that there is a difference between stickers and reactions.

Going forward, I do think we should invest in understanding how people in sticker-heavy markets (e.g. India) use and perceive sticker comments vs. other types of interactions. Do they see these as a substitute for reactions? If yes, should we consider alternative UX treatments for stickers that reduce their prominence in the app relative to text comments? Or do they seem the more like short text comments? Generally, do people make a distinction between stickers containing text vs. those that do not? Regardless of where we land on the MSI weight for stickers, there are still many questions to be answered, all with meaningful product implications.

n the 4 hours following receiving either a short text or sticker comment to that comment, and again
ting to compare total average attributed sessions for each comment type.

hted average difference in downstream sessions for sticker vs short text comments to be **-6%** (vs
ing). Again, we find only a relatively small difference between the comment types, but it may be
me or all of the slight sessions regression we see in the Comment Guide long-term holdout. Given that
ered deterministically in jewel, it perhaps is not surprising that the the difference in sessions is
ever, if we do decide to start treating notifications from sticker comments differently than text
ifference in sessions may become apparent. For example, jewel notifications from Guided sticker
R of **58%** vs **64%** for non-sticker Guided comments.

ed above does not definitively point to the correct weight for sticker comments. It is clear that stickers
n those people giving and receiving them, and data suggests that in some countries (such as India)
e role that reactions do in countries such as the US. Additionally, metrics such as reply rate and sessions
text comment does indeed have more overall ecosystem value than a sticker comment.

he weight for stickers? In prior discussions, people have suggested a base weight equivalent to
current 15). However, sticker comments do have some clear benefits vs. reactions, primarily the ability
nd react to them. Additionally, people can now combine stickers with text in a single comment, in
ld likely treat these sticker+text comments the same as text comments. Finally, should we also reduce
ers that are essentially a stylized short text comment (e.g. Happy Birthday)?

ome obvious daylight between reactions and sticker comments, it seems more appropriate to assign
weight in-between that of reactions and short text comments (e.g. **10-12**). Doing so will still incentivize
Comment Guide) to skew more towards preserving and growing text comments, but will still
here is a difference between stickers and reactions.

think we should invest in understanding how people in sticker-heavy markets (e.g. India) use and
nments vs. other types of interactions. Do they see these as a substitute for reactions? If yes, should we
e UX treatments for stickers that reduce their prominence in the app relative to text comments? Or do
e like short text comments? Generally, do people make a distinction between stickers containing text vs.
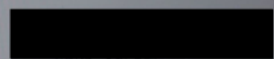

Conversation

You're new to this document

edits · Jul 2, 2019

differ much between sticker and
short text comments, with ~~both~~
average meaningfulness for
comment givers and receivers

View Changes

a comment thread. Stickers
~~generally can not contribute~~ may be
less likely to be part of a
conversation ~~but~~ and instead ~~may~~

View Changes

2 .. Stickers generally can not
contribute to a conversation, but...

View comments (1 new)

BF   Jul 12, 2019
There was a bit of research done
that supported this idea, which I
added a link to, but I also edited t...

↰ Reply