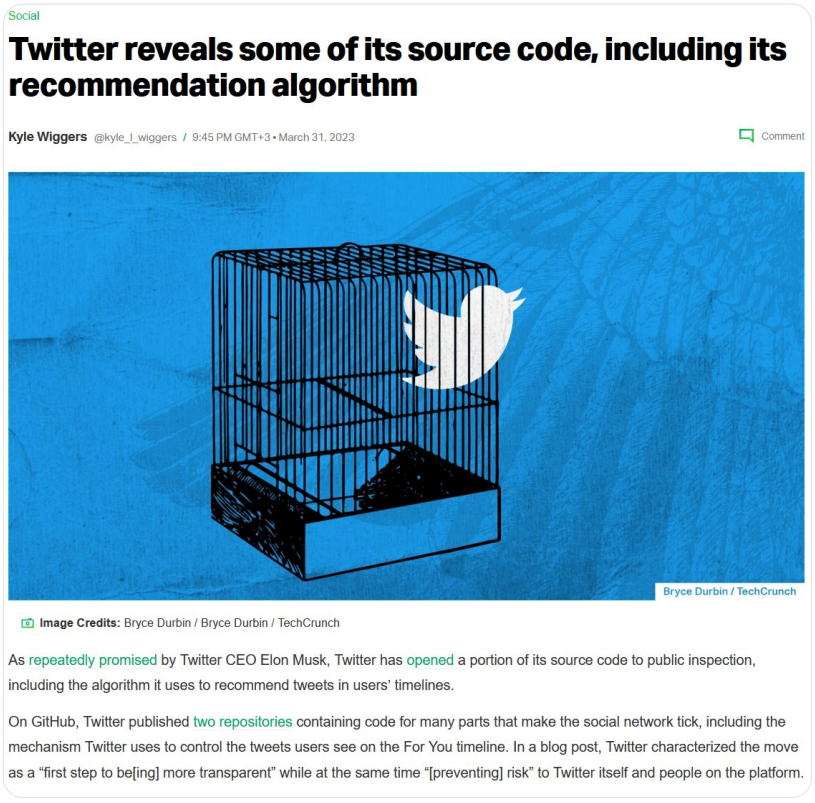
**Pekka Kallioniemi** @P_Kallioniemi

Apr 2 · 23 tweets · P_Kallioniemi/status/1642425184513662977

In today's #vatniksoup, I'll talk about Twitter. Elon Musk recently released the Twitter's algorithm as open-source for all to see (kudos!), and people have already found some interesting things from the code.

Here's my analysis on the most interesting findings.

1/21



Social

**Twitter reveals some of its source code, including its recommendation algorithm**

**Kyle Wiggers** @kyle_l_wiggers / 9:45 PM GMT+3 • March 31, 2023 Comment



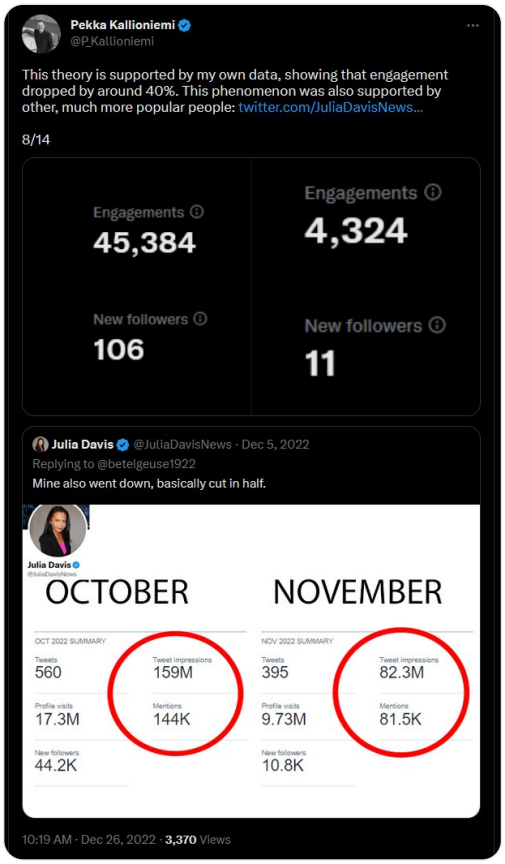
**Image Credits:** Bryce Durbin / Bryce Durbin / TechCrunch

As repeatedly promised by Twitter CEO Elon Musk, Twitter has opened a portion of its source code to public inspection, including the algorithm it uses to recommend tweets in users' timelines.

On GitHub, Twitter published two repositories containing code for many parts that make the social network tick, including the mechanism Twitter uses to control the tweets users see on the For You timeline. In a blog post, Twitter characterized the move as a "first step to be[ing] more transparent" while at the same time "[preventing] risk" to Twitter itself and people on the platform.

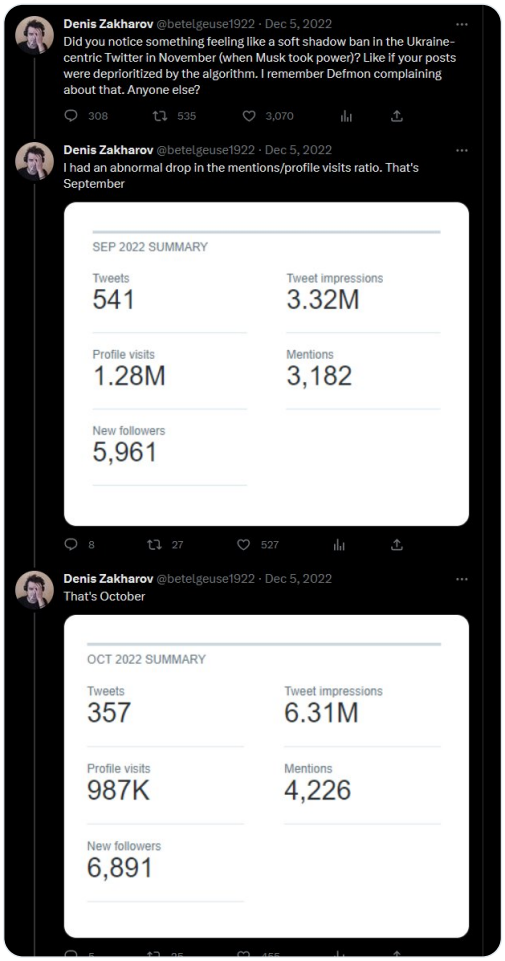
First of all, let's start with the "I told you so" moment: On 26 Dec 2022, I theorized that pre-Elon there was a adjustment in the algorithm that prioritized Ukraine-related content, making it more visible, thus making it gaining more likes and retweets.

2/21

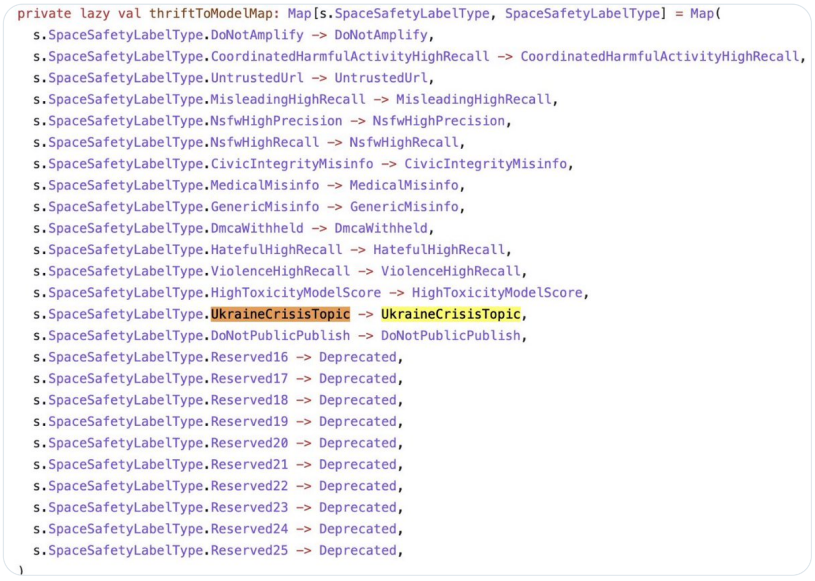
After Elon took over, my engagement in Ukraine-related content went down by 40%. This same drop was experienced and reported by Denis Zakharov (@betelgeuse1922) and Julia Davis (@JuliaDavisNews). Apparently, right after taking over Twitter, Elon Musk put in place an...

3/21

...algorithm change that PENALIZED all Ukraine-related content, making it less visible. For bigger accounts like Davis', 40-50% drop in engagement is HUGE. Now, this could've been due to there being a Ukraine-boosting prioritization before, which he then reversed...

4/21



```
private lazy val thriftToModelMap: Map[s.SpaceSafetyLabelType, SpaceSafetyLabelType] = Map(
  s.SpaceSafetyLabelType.DoNotAmplify -> DoNotAmplify,
  s.SpaceSafetyLabelType.CoordinatedHarmfulActivityHighRecall -> CoordinatedHarmfulActivityHighRecall,
  s.SpaceSafetyLabelType.UntrustedUrl -> UntrustedUrl,
  s.SpaceSafetyLabelType.MisleadingHighRecall -> MisleadingHighRecall,
  s.SpaceSafetyLabelType.NsfwHighPrecision -> NsfwHighPrecision,
  s.SpaceSafetyLabelType.NsfwHighRecall -> NsfwHighRecall,
  s.SpaceSafetyLabelType.CivicIntegrityMisinfo -> CivicIntegrityMisinfo,
  s.SpaceSafetyLabelType.MedicalMisinfo -> MedicalMisinfo,
  s.SpaceSafetyLabelType.GenericMisinfo -> GenericMisinfo,
  s.SpaceSafetyLabelType.DmcaWithheld -> DmcaWithheld,
  s.SpaceSafetyLabelType.HatefulHighRecall -> HatefulHighRecall,
  s.SpaceSafetyLabelType.ViolenceHighRecall -> ViolenceHighRecall,
  s.SpaceSafetyLabelType.HighToxicityModelScore -> HighToxicityModelScore,
  s.SpaceSafetyLabelType.UkraineCrisisTopic -> UkraineCrisisTopic,
  s.SpaceSafetyLabelType.DoNotPublicPublish -> DoNotPublicPublish,
  s.SpaceSafetyLabelType.Reserved16 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved17 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved18 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved19 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved20 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved21 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved22 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved23 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved24 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved25 -> Deprecated,
)
```

...- I guess we'll never know. If the penalization list is in chronological order, the addition to punish Ukraine-related content was the latest addition.

Now, the rest of the algorithm works pretty much like all social media algorithms work: promote thought-provoking...

5/21


News, Tech Roundup

## How social media algorithms harm your mental health

Cora Quigley | December 14, 2022 0

4 min read

Social media algorithms have long been criticized for various reasons, and mounting evidence strongly suggests the detriment they can have on users' mental health. A team of psychology researchers has urged social media companies to increase transparency around how algorithms work in a recent article published in the journal Body Image. While this article focuses on the impact of social media on teens and body image, other studies have found a link between social media use and depression.

Many point to algorithms as a key cause for these adverse mental health outcomes.

...content and kill any conversation around it. Each like gains 30x boost, re-tweets get a 20x bonus, and replies have no effect at all. So appease to your audience and refuse to talk with the dissidents, and you'll be rewarded by the algorithm.

6/21

```
private def getLinearRankingParams: ThriftRankingParams = {
  ThriftRankingParams(
    `type` = Some(ThriftScoringFunctionType.Linear),
    minScore = -1.0e100,
    retweetCountParams = Some(ThriftLinearFeatureRankingParams(weight = 20.0)),
    replyCountParams = Some(ThriftLinearFeatureRankingParams(weight = 1.0)),
    reputationParams = Some(ThriftLinearFeatureRankingParams(weight = 0.2)),
    luceneScoreParams = Some(ThriftLinearFeatureRankingParams(weight = 2.0)),
    textScoreParams = Some(ThriftLinearFeatureRankingParams(weight = 0.18)),
    urlParams = Some(ThriftLinearFeatureRankingParams(weight = 2.0)),
    isReplyParams = Some(ThriftLinearFeatureRankingParams(weight = 1.0)),
    favCountParams = Some(ThriftLinearFeatureRankingParams(weight = 30.0)),
    langEnglishUIBoost = 0.5,
    langEnglishTweetBoost = 0.2,
    langDefaultBoost = 0.02,
    unknownLanguageBoost = 0.05,
    offensiveBoost = 0.1,
    inTrustedCircleBoost = 3.0,
    multipleHashtagsOrTrendsBoost = 0.6,
    inDirectFollowBoost = 4.0,
    tweetHasTrendBoost = 1.1,
    selfTweetBoost = 2.0,
    tweetHasImageUrlBoost = 2.0,
    tweetHasVideoUrlBoost = 2.0,
    useUserLanguageInfo = true,
    ageDecayParams = Some(ThriftAgeDecayRankingParams(slope = 0.005, base = 1.0))
  )
}
```

Adding media helps, and adding both videos and images gives you a 2.0x boost. External links, on the other hand, may hurt your engagement especially if the account is fresh. I assume this was done to counter the spam bots that spread harmful and spammy links.

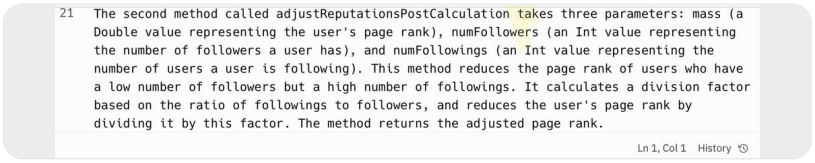
7/21

```
selfTweetBoost = 2.0,
tweetHasImageUrlBoost = 2.0,
tweetHasVideoUrlBoost = 2.0,
useUserLanguageInfo = true,
```

```
@Override
protected float score(float luceneQueryScore) throws IOException {
  if (documentFeatures.isFlagSet(EarlybirdFieldConstant.FROM_VERIFIED_ACCOUNT_FLAG)) {
    return NOT_SPAM_SCORE;
  }

  int tweepCredThreshold = 0;
  if (documentFeatures.isFlagSet(EarlybirdFieldConstant.HAS_LINK_FLAG)
      && !documentFeatures.isFlagSet(EarlybirdFieldConstant.HAS_IMAGE_URL_FLAG)
      && !documentFeatures.isFlagSet(EarlybirdFieldConstant.HAS_VIDEO_URL_FLAG)
      && !documentFeatures.isFlagSet(EarlybirdFieldConstant.HAS_NEWS_URL_FLAG)) {
    // Contains a non-media non-news link, definite spam vector.
    tweepCredThreshold = MIN_TWEEPCRED_WITH_LINK;
  }

  int tweepcred = (int) documentFeatures.getFeatureValue(EarlybirdFieldConstant.USER_REPUTATION);

  // For new user, tweepcred is set to a sentinel value of -128, specified at
  // src/thrift/com/twitter/search/common/indexing/status.thrift
  if (tweepcred >= tweepCredThreshold
      || tweepcred == (int) RelevanceSignalConstants.UNSET_REPUTATION_SENTINEL) {
    return NOT_SPAM_SCORE;
  }

  double retweetCount =
      documentFeatures.getUnnormalizedFeatureValue(EarlybirdFieldConstant.RETWEET_COUNT);
  double replyCount =
      documentFeatures.getUnnormalizedFeatureValue(EarlybirdFieldConstant.REPLY_COUNT);
  double favoriteCount =
      documentFeatures.getUnnormalizedFeatureValue(EarlybirdFieldConstant.FAVORITE_COUNT);

  // If the tweet has enough engagements, do not mark it as spam.
  if (retweetCount + replyCount + favoriteCount >= ENGAGEMENTS_NO_FILTER) {
    return NOT_SPAM_SCORE;
  }

  return SPAM_SCORE;
}
```

Your following-to-follower ratio also matters, and there's an algorithm that reduces the engagement of users who follow a lot of accounts but have only few followers. This could be used to counter the "follow me and I follow back" accounts who try to...

8/21




...garner a large following in a short time.

If a lot people have muted or blocked you, your engagement goes down. Same happens, if a lot of people have recently unfollowed you or reported your account for spam and/or abuse.

9/21

```
// group all features by (src, dest)
val allEdgeFeatures: SCollection[Edge] =
  getEdgeFeature(SCollection.unionAll(Seq(blocks, mutes, abuseReports, spamReports, unfollows)))

val negativeFeatures: SCollection[KeyVal[Long, UserSession]] =
  allEdgeFeatures
    .keyBy(_.sourceId)
    .topByKey(maxDestinationIds)(Ordering.by(_.features.size))
    .map {
      case (srcId, pqEdges) =>
        val topKNeg =
          pqEdges.toSeq.flatMap(toRealGraphEdgeFeatures(hasNegativeFeatures))
        KeyVal(
          srcId,
          UserSession(
            userId = Some(srcId),
            realGraphFeaturesTest =
              Some(RealGraphFeaturesTest.V1(RealGraphFeaturesV1(topKNeg)))))
    }
```

Having Twitter Blue gives you a HUGE boost in engagement, ranging from 0 to 100 with the default of 4x.

Besides Ukraine-related content, there are penalties for posting disinformation, medical misinformation (most probably related to COVID-19), calls for...

10/21

```
object BlueVerifiedAuthorInNetworkMultiplierParam
    extends FSBoundedParam[Double](
      name = "home_mixer_blue_verified_author_in_network_multiplier",
      default = 4.0,
      min = 0.0,
      max = 100.0
    )

object BlueVerifiedAuthorOutOfNetworkMultiplierParam
    extends FSBoundedParam[Double](
      name = "home_mixer_blue_verified_author_out_of_network_multiplier",
      default = 2.0,
      min = 0.0,
      max = 100.0
    )
```

harmful acts (probably installed after Jan 6th), "not safe for work" content (usually porn), content with withheld DMCA strikes, and hateful and/or violent content.

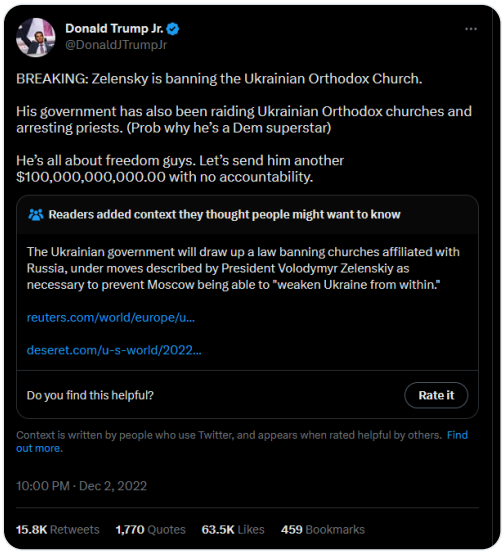
But if the account is big enough, these penalties don't even matter. Take for example Donald Trump Jr.'s...

11/21

```
private lazy val thriftToModelMap: Map[s.SpaceSafetyLabelType, SpaceSafetyLabelType] = Map(
  s.SpaceSafetyLabelType.DoNotAmplify -> DoNotAmplify,
  s.SpaceSafetyLabelType.CoordinatedHarmfulActivityHighRecall -> CoordinatedHarmfulActivityHighRecall,
  s.SpaceSafetyLabelType.UntrustedUrl -> UntrustedUrl,
  s.SpaceSafetyLabelType.MisleadingHighRecall -> MisleadingHighRecall,
  s.SpaceSafetyLabelType.NsfwHighPrecision -> NsfwHighPrecision,
  s.SpaceSafetyLabelType.NsfwHighRecall -> NsfwHighRecall,
  s.SpaceSafetyLabelType.CivicIntegrityMisinfo -> CivicIntegrityMisinfo,
  s.SpaceSafetyLabelType.MedicalMisinfo -> MedicalMisinfo,
  s.SpaceSafetyLabelType.GenericMisinfo -> GenericMisinfo,
  s.SpaceSafetyLabelType.DmcaWithheld -> DmcaWithheld,
  s.SpaceSafetyLabelType.HatefulHighRecall -> HatefulHighRecall,
  s.SpaceSafetyLabelType.ViolenceHighRecall -> ViolenceHighRecall,
  s.SpaceSafetyLabelType.HighToxicityModelScore -> HighToxicityModelScore,
  s.SpaceSafetyLabelType.UkraineCrisisTopic -> UkraineCrisisTopic,
  s.SpaceSafetyLabelType.DoNotPublicPublish -> DoNotPublicPublish,
  s.SpaceSafetyLabelType.Reserved16 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved17 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved18 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved19 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved20 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved21 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved22 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved23 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved24 -> Deprecated,
  s.SpaceSafetyLabelType.Reserved25 -> Deprecated,
)
```

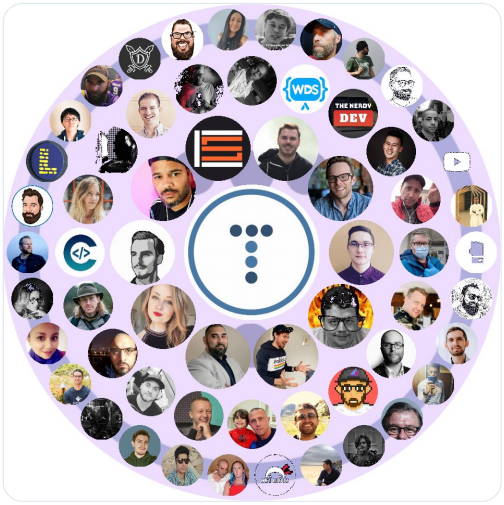
..disinformative tweet about Zelenskyy banning the Ukrainian Orthodox Church: this was clearly a lie and was labelled as such, yet the tweet itself gained 16 000 retweets and 64 000 likes. Once you've got big enough audience on Twitter, none of the penalties apply.

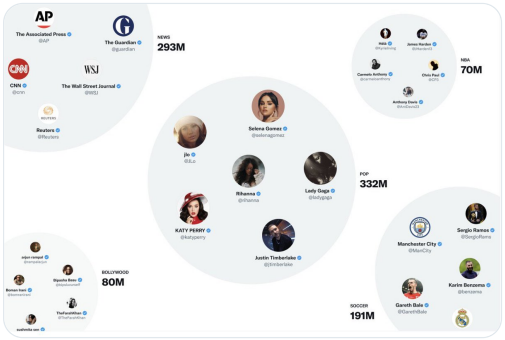
12/21





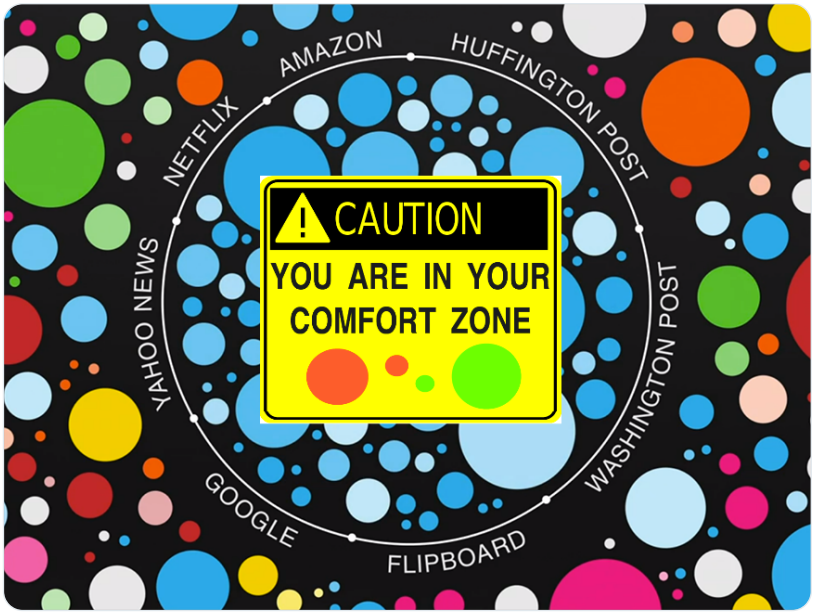
Remember those Twitter circles which shows who you interact with most? Twitter does those,too.Most of the users who tweet regularly are put inside a circle & posting on topics outside of your circle actually hurts your engagement. Now,this penalty is 10x which is again huge.13/21






```
// subtractive penalty applied after boosts for out-of-network replies.
120: optional double outOfNetworkReplyPenalty = 10.0
```

So just keep on posting stuff that appeases your audience and you won't get punished. Any new perspectives and challenging of people is strictly forbidden, at least if you want to stay relevant. Twitter also tracks the time people spend on your tweets, probably to increase..14/21




..the engagement of threads. If people spend over 2 minutes reading your tweet or if people check your profile through one of your tweets, you also get bonuses.

Twitter also detects for "unknown language", so misspelling, typos or using of words that don't exist penalizes..15/21

..your engagement a LOT (0.01x penalty). So even if you have a good message, it can be penalized if it's badly written. This again seems like a measure against auto-translation bots that spew out bad English en masse.

Interestingly, Twitter also tracks closely the...

16/21

```
// Boost (demotion) if the tweet language is not one of user's understandable languages,
// nor interface language.
43: optional double unknownLanguageBoost = 0.01
```

...engagement of US political system, by tracking the impressions on both Republican and Democrat users. It also tracks how so-called Power Users and Elon himself are doing in terms of impressions.

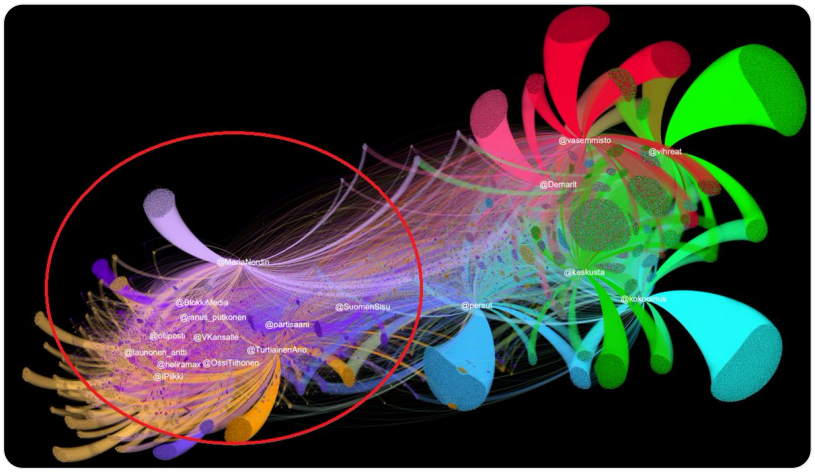
To conclude: Twitter's algorithm is working exactly like other social...

17/21

```
224       (
225         "author_is_elon",
226         candidate =>
227           candidate
228             .getOrElse(AuthorIdFeature, None).contains(candidate.getOrElse(DDGStatsElonFeature, 0L))),
229       (
230         "author_is_power_user",
231         candidate =>
232           candidate
233             .getOrElse(AuthorIdFeature, None)
234             .exists(candidate.getOrElse(DDGStatsVitsFeature, Set.empty[Long]).contains)),
235       (
236         "author_is_democrat",
237         candidate =>
238           candidate
239             .getOrElse(AuthorIdFeature, None)
240             .exists(candidate.getOrElse(DDGStatsDemocratsFeature, Set.empty[Long]).contains)),
241       (
242         "author_is_republican",
243         candidate =>
244           candidate
245             .getOrElse(AuthorIdFeature, None)
246             .exists(candidate.getOrElse(DDGStatsRepublicansFeature, Set.empty[Long]).contains)),
247     )
```

...media platforms, meaning that it promotes the idea of "information bubbles" and echo chambers. It seems that Elon Musk also lied when he said that Twitter now promotes all viewpoints equally - as there's a engagement penalty for all Ukraine-related content.

18/21



I actually appreciate Musk for releasing the algorithm as open-source, but there's also one big problem: it allows social media manipulation on Twitter to be taken to a whole new level. Let me give you an example: if troll farms want to decrease the engagement of a ...

19/21

...dissident, all they have to do is create hundreds of accounts that then mute/block that dissident's account. They can also increase engagement for propagandists and grifters by retweeting and liking their content. Actually, anyone can do it - buying 1000 likes ...

20/21

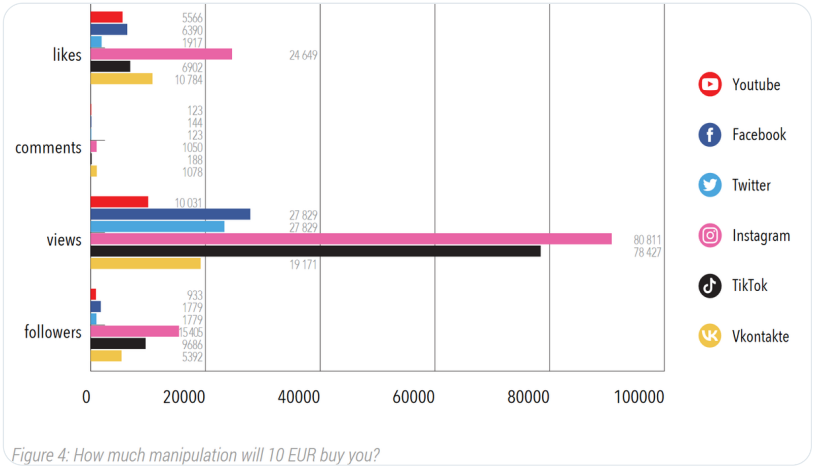



*Figure 4: How much manipulation will 10 EUR buy you?*

...costs you today around 30 USD, and 500 retweets can be bought for 15 USD. Or, if you're the owner of the platform, you can just change the algorithm so that EVERYONE sees your tweets whether they care about your message or not.

21/21


Big shoutout to all analysts, including @amasad @0xCygaar @xerocooleth @steventey @aakashg0

Support my work:

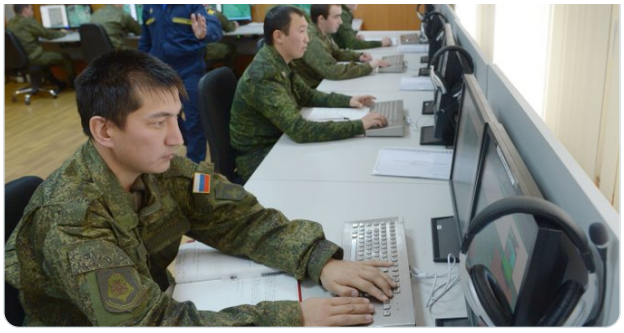
and production of vatnik soup or in the creation of web portal (vatniksoup.com)...

https://buymeacoffee.com/PKallioniemi

Past soups: [vatniksoup.com](vatniksoup.com)

Related soups:
Social media manipulation:


Troll farms:



7:22 AM · Dec 9, 2022

Read the full conversation on Twitter

1K Reply Copy link

Read 24 replies

APPENDIX: Some of the claims here have been disputed, see for example this thread:

**Crimson**
@CrimsonShadowMK · Follow

I'd like to address this misleading analysis of #Twitter code, which seems to be cobbled together with a minimum of investigation. I'll preface this by saying that I fully support Ukraine 🇺🇦, and that I'm just as unhappy as many with the direction this platform has taken. 📜

> **Aakash Gupta 🚀 Product Growth Guy** @aakashg0
> Twitter revealed its algorithm to the world.
>
> But what does it mean for you?
>
> I spent the evening analyzing it.
>
> Here's what you need to know:

12:12 AM · Apr 2, 2023

Read the full conversation on Twitter

633 Reply Copy link

Read 21 replies

• • •