Deliberation and Resolution on Wikipedia: A Case Study of Requests for Comment

Jane Im*, University of Michigan
Amy Zhang, MIT CSAIL
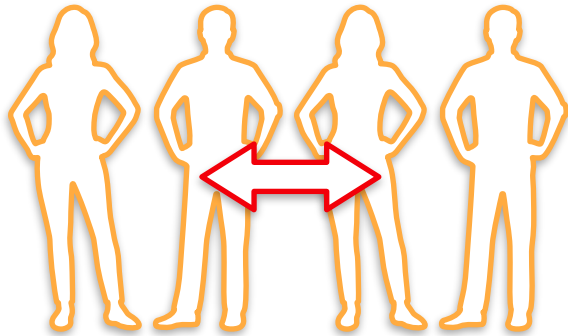Christopher Schilling, Wikimedia Foundation
David Karger, MIT CSAIL
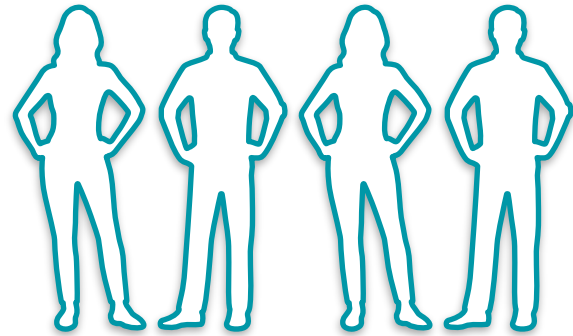Jonathan Morgan, Wikimedia Foundation
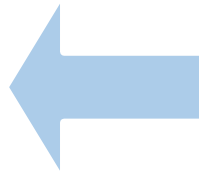
1

What are **Requests for Comment (RfC)** ?

A major and common process used by Wikipedia editors for requesting input from uninvolved editors concerning content related disputes

Unresolved content dispute

Uninvolved editors' new input

What are **Requests for Comment (RfC)** ?

A major and common process used by Wikipedia editors for requesting input from uninvolved editors concerning content related disputes
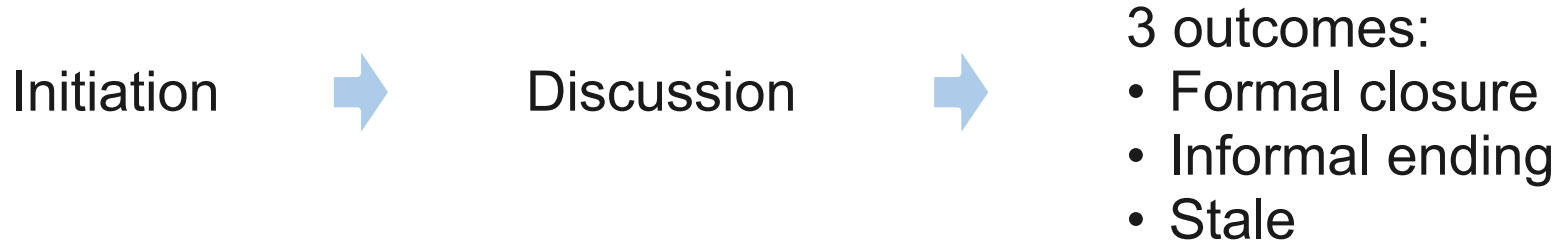
Initiation ➡ Discussion ➡ 3 outcomes:
- Formal closure
- Informal ending
- Stale

Request for comments (RFC) about whether the lead paragraph should say he's the current president

An editor has **requested comments from other editors for this discussion.** Within 24 hours, this page will be added to the following lists:

- **Biographies**
- **Politics, government, and law**

When discussion has ended, remove this tag and it will be removed from the list. If this page is on additional lists, they will be noted below.

Please consider joining the feedback request service.

Assuming that the hatnote is edited to remove that he is the current ("incumbent") president, should the lead paragraph be edited to say that he is the current president?15:58, 27 March 2017 (UTC)

Survey about saying he's the current president

- **Support** as proposer. This is a sensitive subject, in view of the Not my president and Not My Presidents Day movements. If the lead paragraph merely says that he is the 45th president, that does not indicate he is the current president. After all, reliable sources say that Jimmy Carter is still the 39th president, and Bill Clinton is still the 42nd president. According to the Washington Post⧉, and CBS News⧉, "Jimmy Carter is the 39th president of the United States". According to the Washington Post⧉ and Variety⧉, Bill Clinton "is the 42nd President of the United States". It's true that our talk page includes a list of current consensuses⧉, but that list is merely for exempting people from 1RR, not for making it more difficult to change any consensus. The list of consensuses includes the lead sentence, but removal of incumbency from the hatnote changes the context of the lead sentence, to such an extent that I think a slight modification of the lead sentence is very much justified. This RFC is not about how exactly it should be phrased ("current" vs. "incumbent" vs. "serving since", etc.).Anythingyouwant (talk) 15:58, 27 March 2017 (UTC)
- **Support**. Defining characteristic. Current, incumbent, serving all good... or put brackets round (45th). Siuenti (talk) 16:06, 27 March 2017 (UTC)
- **Oppose** as unnecessary and superfluous. "Is" is a present tense verb, regardless of how some sources choose to use it (I don't think WP:V applies to grammar). The notion that this is needed because of a "not my

Request for comments (RFC) about whether the lead paragraph should say he's the current president

*The following discussion is closed. **Please do not modify it.** Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.*

Consensus reached to **include "current"**, exact language agreed elsewhere, see Talk:Donald Trump#Current consensus, item #17. — JFG talk 20:00, 2 April 2017 (UTC)

Assuming that the hatnote is edited to remove that he is the current ("incumbent") president, should the lead paragraph be edited to say that he is the current president?15:58, 27 March 2017 (UTC)

An uninvolved editor (closer) came and formally closed the RfC.

Consensus reached to **include "current"**, exact language agreed elsewhere, see Talk:Donald Trump#Current consensus, item #17. — JFG talk 20:00, 2 April 2017 (UTC)

Closer evaluated that the consensus is to include "current".

This is the ideal case but it doesn't always happen to all RfCs.

	Formally closed	**Informally ended**	**Stale**
Dispute is resolved	O	O	X
Ended by	Uninvolved editor (closer)	Participant, initiator, or uninvolved editor	None
RfC tag is removed by	Uninvolved editor (closer)	Participant, initiator, or uninvolved editor	Legobot

Comparison of three possible outcomes of RfCs

	Formally closed	**Informally ended**	**Stale**
Dispute is resolved	O	O	X
Ended by	Uninvolved editor (closer)	Participant, initiator, or uninvolved editor	None
RfC tag is removed by	Uninvolved editor (closer)	Participant, initiator, or uninvolved editor	Legobot

Usually for relatively more contentious discussions.

Request: Enlarge section on policies?

As with most of these types of articles, the emphasis is mostly biographical, but given that Sanders' politics may diverge significantly from other US senators (D & R), could some information be included from manifestos or third-party sources? In other words, a bit more on how Sanders' socialism is manifest. --94.194.57.116 (talk) 13:55, 12 April 2009 (UTC)

I deleted the link to "Sanders Turns Blind Eye To Victims, Sponsors Deadly MOTHERS Act" because it was a dead link. —Preceding unsigned comment added by 207.47.218.9 (talk) 12:51, 27 August 2009 (UTC)

Sanders stand on the Patriot Acts

Why does the article say nothing about wether or not Bernie Sanders voted for the Patriot Acts or not? —Preceding unsigned comment added by 84.215.44.195 (talk) 09:23, 13 October 2009 (UTC)

	Formally closed	Informally ended	Stale
Dispute is resolved	O	O	X
Ended by	Uninvolved editor (closer)	Participant, initiator, or uninvolved editor	None
RfC tag is removed by	Uninvolved editor (closer)	Participant, initiator, or uninvolved editor	Legobot

Usually due to overwhelming agreement by participants or withdrawal of the RfC by the initiator.

	Formally closed	**Informally ended**	**Stale**
Dispute is resolved	O	O	X
Ended by	Uninvolved editor (closer)	Participant, initiator, or uninvolved editor	None
RfC tag is removed by	Uninvolved editor (closer)	Participant, initiator, or uninvolved editor	Legobot

Problematic case where the dispute is unresolved.

- **Comment** If multiple human rights organizations and publications are cited which label certa[in]
  original research? If this material is deleted, we have no objective basis whatsoever for the c[...]
  thrown around without any precise meaning. I'm not saying there is only one, objective yards[...]
  17 April 2014 (UTC)

- **Oppose**. I think I have made my case against Ghostofnemos repeated and inherent OR ten[...]
  Saddhiyama (talk) 22:10, 17 April 2014 (UTC)

- **Oppose**. They are either fork of authoritarianism or WP:OR with WP:POV. "The Economist,
  are not in this case. Both represent specific political positions. Sietecolores (talk) 22:50, 18 A[...]

Another, sad Wikipedia fail. Ghostofnemo (talk) 01:00, 27 April 2014 (UTC)

No, this was a success ... material that did not refer to police states was correctly rejected as[...]
"police state" and "authoritarian regime" or "unfree state" as synonymous (they clearly are n[...]
your say-so would have been a failure. -- 184.189.217.91 (talk) 04:40, 23 April 2015 (UTC)
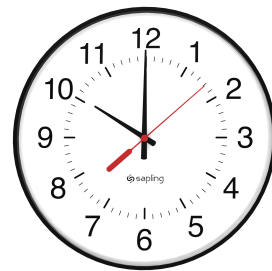
14

We found that

- 1/3 of RfCs are **stale without any closure**

We found that

- 1/3 of RfCs are **stale without any closure**

- Many do not get formally closed in a timely fashion
  - average: 45.56 days ($\sigma = 81.14$)

1.5 times more than the time an RfC is allowed by default

So it appears many Requests for Comments are stale.

Why are they a problem?

Stale RfCs…

- Can be **discouraging to editors** if an RfC never gets closed when they put effort into it

- Can be a **problem for productivity** as editors involved in RfCs may wait on the outcome before further editing

Why do many RfCs remain stale?

OUTLINE

1. Collection of data

2. Description of RfCs over time

3. Reasons behind stale RfCs

4. Model for predicting RfC's outcome

OUTLINE

1. **Collection of data**

2. Description of RfCs over time

3. Reasons behind stale RfCs

4. Model for predicting RfC's outcome

Quantitative Study

- Gathered and analyzed 7,316 RfCs

  from the English Wikipedia 2011~2017
- Built models predicting RfCs' outcomes

Quantitative Study

- Gathered and analyzed 7,316 RfCs from the English Wikipedia 2011~2017
- Built models predicting RfCs' outcomes

Qualitative Study

- Interviewed frequent 10 closers
- Inspected 40 randomly chosen stale RfCs from the dataset

Quantitative Study

- Gathered and analyzed 7,316 RfCs

  from the English Wikipedia 2011~2017
- Built models predicting RfCs' outcomes

Qualitative Study

- Interviewed frequent 10 closers
- Inspected 40 randomly chosen

  stale RfCs from the dataset

- Consulted with 2 members of the Wikimedia Foundation
- Discussed the study on Wikimedia's research mailing list

RfC Discussion and Closing Data

1. Use revision history of talk pages provided by MediaWiki API and Legobot's edits

```
| 601484529 | 2014-03-27 10:00:41.000000 | Adding RFC ID.
```

```
| 605875895 | 2014-04-26 10:00:23.000000 | Removing expired RFC template.
```

RfC Discussion and Closing Data

1. Use revision history of talk pages provided by MediaWiki API and Legobot's edits

```
| 601484529 | 2014-03-27 10:00:41.000000 | Adding RFC ID.
```

```
| 605875895 | 2014-04-26 10:00:23.000000 | Removing expired RFC template.
```

2. Gather url of the original page where the RfC was started at

RfC Discussion and Closing Data

1. Use revision history of talk pages provided by MediaWiki API and Legobot's edits

```
| 601484529 | 2014-03-27 10:00:41.000000 | Adding RFC ID.
```

```
| 605875895 | 2014-04-26 10:00:23.000000 | Removing expired RFC template.
```

2. Gather url of the original page where the RfC was started at

3. Iterate through archives to find the current url of RfC

RfC Discussion and Closing Data

1. Use revision history of talk pages provided by MediaWiki API and Legobot's edits

```
| 601484529 | 2014-03-27 10:00:41.000000 | Adding RFC ID.
```

```
| 605875895 | 2014-04-26 10:00:23.000000 | Removing expired RFC template.
```

2. Gather url of the original page where the RfC was started at

3. Iterate through archives to find the current url of RfC

4. Using the url, retrieve/parse RfC content using libraries (MediaWiki, WikiChatter)

RfC Discussion and Closing Data

- 7,316 RfCs on English Wikipedia (2011 ~2017)

- Extracted initiator, participant, closer information, comments, initiating and closing statements

- Kept reply structure intact using Python libraries

OUTLINE

1. Collection of data

2. **Description of RfCs over time**

3. Reasons behind stale RfCs

4. Model for predicting RfC's outcome
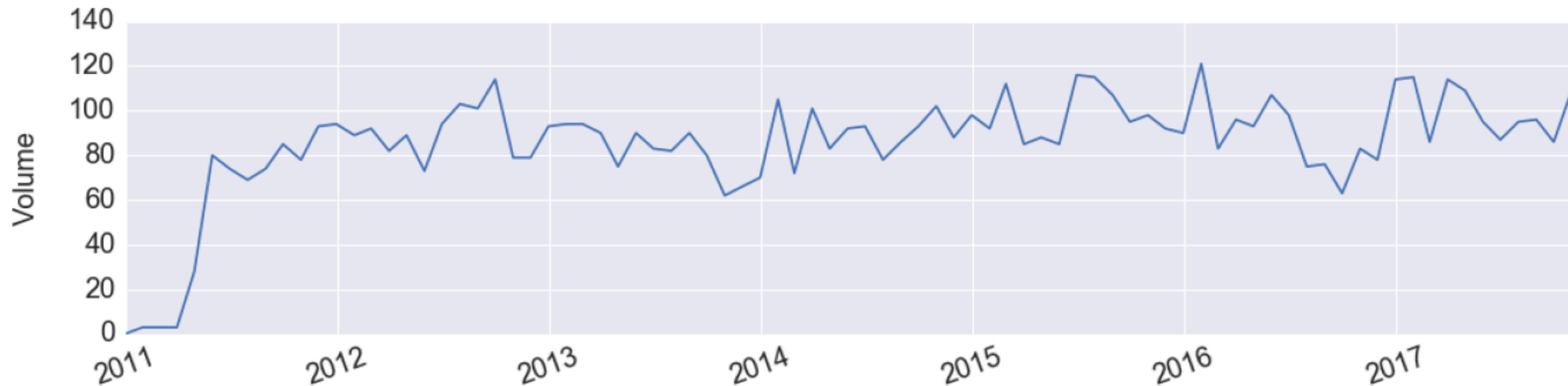
The number of RfCs initiated each month in our dataset from 2011 to end of 2017

Overall a **steady volume** of RfCs are initiated: 60~120 per month

Closers are:

1. **More experienced** than initiators/participants
   - higher average edit counts

	closers	participants	initiators
number of edits	39,759	14,055	23,432

2. **Smaller number** of closers compared to initiators/participants
   - May mean not everyone closes

	closers	participants	initiators
num of editors	759	14,815	3,346

So there's a steady volume of usage of RfCs with experienced editors as closers…
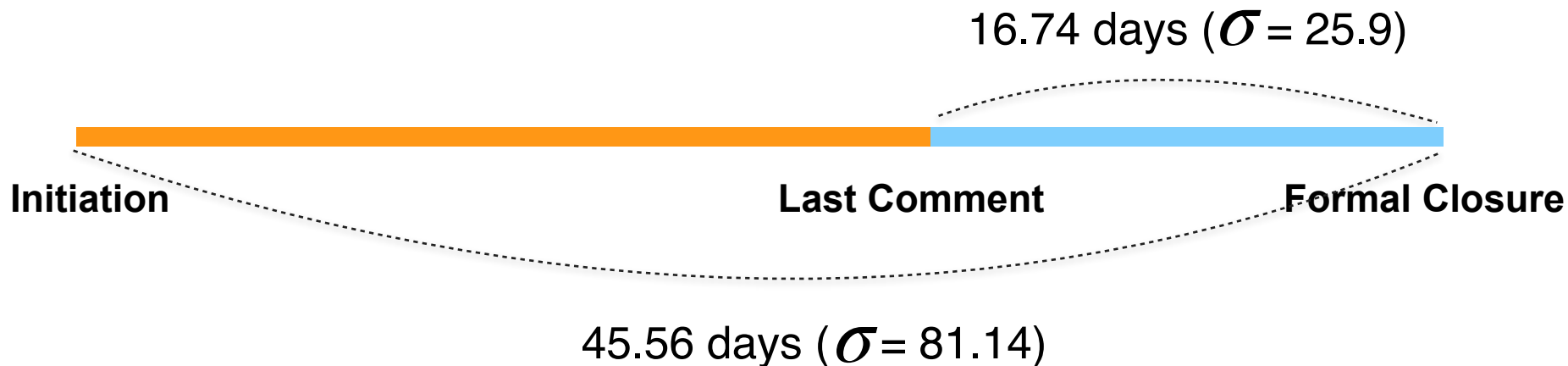
Back to the main question:
Major issues we found in the RfC process

1. Many Requests for Comment are stale.

	Formally closed	Informally ended	Stale
Dispute is resolved	O	O	X
Ended by	Uninvolved editor (closer)	Participant, initiator, or uninvolved editor	None
Number of RfCs	4,086 (58%)	672 (9%)	2,329 (33%)

2. Many RfCs are not formally closed with a timely fashion.

16.74 days ($\sigma = 25.9$)

Initiation

Last Comment

Formal Closure

45.56 days ($\sigma = 81.14$)

Crazy outliers exist:
- RfCs that took a long time to close after last comment
- RfCs that just dragged on, not being formally closed

Time between opening and last comment   Time between last comment and closing
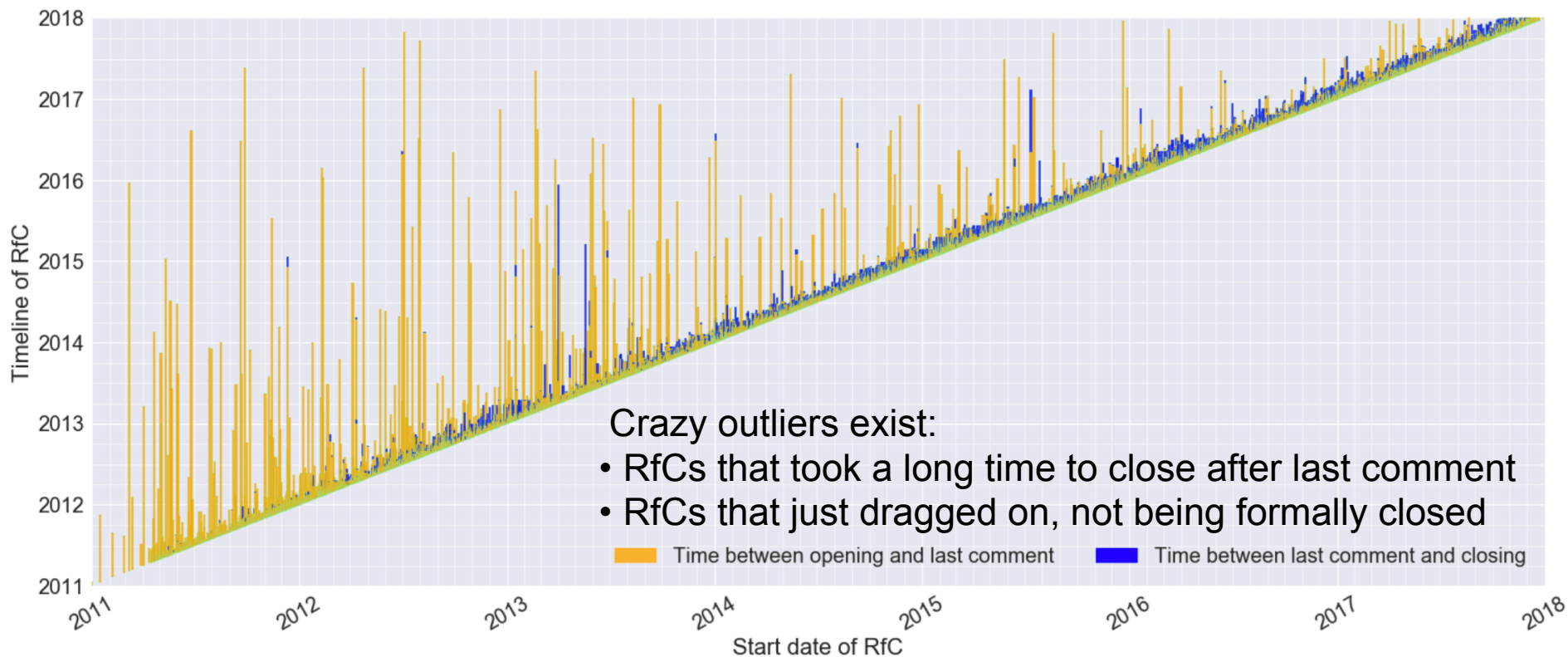
Why are many Requests for Comment stale?

OUTLINE

1. Collection of data

2. Description of RfCs over time

3. **Reasons behind stale RfCs**

4. Model for predicting RfC's outcome

Recap of Methods

 A. Interviews with 10 frequent closers on English Wikipedia

 B. Qualitative analysis of randomly selected 40 stale RfCs

1. Problems with initiators and initial proposals

> I was randomly selected by RFCbot to comment here. This request is too vague to serve as the basis for any consensus. Please state the request explicitly (and neutrally). What is it you are asking for input about?
> Jojalozzo 20:22, 7 March 2013 (UTC)

"**An RfC not well-formed -** … **results are unclear because of the structure of the RfC.** For example, the RfC might **have no clear question**…"

1. Problems with initiators and initial proposals

I was randomly selected by RFCbot to comment here. This request is too vague to serve as the basis for any consensus. Please state the request explicitly (and neutrally). What is it you are asking for input about?
Jojalozzo 20:22, 7 March 2013 (UTC)

"**An RfC not well-formed -** … **results are unclear because of the structure of the RfC.** For example, the RfC might **have no clear question**…"

precisely due to such nonsensical POV-pushing that no one has come to your aid, despite all your canvassing, in case you're wondering. Athenean (talk) 20:39, 2 March 2013 (UTC)

2. Behavior of participants: bickering and sock-puppeting

This is a meta-comment, about the dispute rather than the substance: Both Aprock and Mirade are spending too much time bickering over this. Both of them need to slow down and let other editors comment. Both of them would do well to stop responding to the other person's comments within minutes. If you've opened this RFC to get comments rather than to get another place to argue with each other, then you need to make this forum more accessible to other people by *not* posting.

2. Behavior of participants: bickering and sock-puppeting

This is a meta-comment, about the dispute rather than the substance: Both Aprock and Mirade are spending too much time bickering over this. Both of them need to slow down and let other editors comment. Both of them would do well to stop responding to the other person's comments within minutes. If you've opened this RFC to get comments rather than to get another place to argue with each other, then you need to make this forum more accessible to other people by *not* posting.

"If I would have a **suspicion that there was socking going on**, I probably **wouldn't be closing it.**"

43

3. Lack of interest or expertise from uninvolved editors

Should this split proceed?

I would just do it, but lack of objection is not quite the same thing as a show of support, especially on a page with few watchlisters. I'm including a `|style` parameter in the RfC tag since followers of Wikipedia-internal style discussions are usually also interested in the progress of our reader-facing articles on such subjects (or should be!).

"When no one cares enough because even if you get it wrong, you've affected one small part of one article that might get 15 views a day…
passed on an RfC because I thought '… My time is better used elsewhere.'"

3. Lack of interest or expertise from uninvolved editors

"…in some cases a **certain amount of background** may also be a requirement. This is especially relevant for more technical subjects… You may be able to remedy this by studying, or it may be **better to leave the discussion for someone else to close**."

4. RfC is too complicated/contentious

"…a **few that I avoid** just because… **anything with like 300 plus comments or where feelings are running very high.** Eventually I think '**That needs one of Wikipedia's big names to close**…'"

"And I tried to read it, I looked it over and I **realized I couldn't make heads or tails of it.**'"

5. Interpersonal issues and "wikipolitics"

"… Now suppose **people with whom I do not share a particularly good relationship has initiated the RfC**, **I don't generally close it**."

"…if you have **people who don't like something you did**, … , if it's **not popular amongst enough people**, they can … sway a discussion**."**

OUTLINE

1. Collection of data

2. Description of RfCs over time

3. Reasons behind stale RfCs

4. **Model for predicting RfC's outcome**

Goal

1. Understand the features that can predict whether an RfC will go stale.

Goal

1. Understand the features that can predict whether an RfC will go stale

2. Help initiator/participants take action to prevent stale RfCs

Goal

1.Understand the features that can predict whether an RfC will go stale

2.Help initiator/participants take action to prevent stale RfCs

Need to also build timely models to prevent in advance

Building a predictor

1. Select features

   ↑

   5 reasons found through qualitative study

2. Train classifiers using those features

1.Initiator's experience ⬅ **Problems with initiators/initial proposals**

- initiator's edit count

- age of the initiator account

- number of revisions to the talk page of the RfC by the initiator

- whether the initiator is an admin

2.Participants' interest ⬅ **Lack of interest from editors**

- number of participants

- ratio of new participants to the talk page where the RfC is at

3.Participants' experience ⬅ **Behavior of participants**

- age of the account of participants(average, standard deviation, sum, maximum)

- participants' edit count (average, sum)

4.Size and shape of discussion   ⬅   **RfC being complicated/long**

- number of comments

- average depth of replies per comment

- average number of replies to each comment

5.Contentiousness   **RfC being contentious**

- number of supports/opposes

- ratio of supports over total votes

- number of replies that support/oppose comments receive

- weighted reciprocity

6.Tone of participant discourse ⬅ **Behavior of participants**

- hostility, swear words, anger, positive, negative, affect

- cognition, percept, insight

- first-person singular word, inclusive, exclusive

- certainty, tentativeness

7.Initial proposal tone & length ⬅ **Problems with initial proposals**

- number of words and characters in proposal

- same tone related features as the participants'

8.Popularity of RfC and topic ⬅ **Lack of interest from editors**

- number of words and characters in the RfC

- number of revisions made on talk page of RfC

- number of revisions made 1/2/3 week(s), 1/2 month(s)

  prior to initiation

Algorithm	Precision	Recall	F1	AUC	Accuracy
LG	0.762	0.868	0.812	0.657	**0.73**
ADT	0.788	0.864	0.825	0.695	**0.753**
RF	0.75	0.909	0.822	0.645	**0.736**
SVM	0.71	0.955	0.815	0.58	**0.709**
Baseline (predicting close)	0.672	1	0.803	0.5	**0.672**

Adaptive Boosted Decision Trees perform the best overall except the recall score with **75.3%** accuracy.

➡ **8.1%** increase over the baseline performance of 67.2%.

Category	Precision	Recall	F1	AUC	Accuracy
Size and Shape of Discussion	0.75	0.903	0.819	0.644	**0.733**
Participant Experience	0.757	0.86	0.805	0.647	**0.72**
Participant Interest	0.722	0.897	0.8	0.595	**0.699**
Contentiousness	0.674	0.98	0.799	0.506	**0.669**
Popularity of RfC and Topic	0.687	0.947	0.797	0.533	**0.675**
Tone of Discourse	0.691	0.925	0.791	0.54	**0.673**
Initiator Experience	0.675	0.984	0.801	0.508	**0.672**
Initial Proposal Tone and Length	0.673	0.978	0.798	0.504	**0.667**

Performance of ADT classifiers using features from each category

Category	Precision	Recall	F1	AUC	Accuracy
Size and Shape of Discussion	0.75	0.903	0.819	0.644	**0.733**
Participant Experience	0.757	0.86	0.805	0.647	**0.72**
Participant Interest	0.722	0.897	0.8	0.595	**0.699**
Contentiousness	0.674	0.98	0.799	0.506	0.669
Popularity of RfC and Topic	0.687	0.947	0.797	0.533	0.675
Tone of Discourse	0.691	0.925	0.791	0.54	0.673
Initiator Experience	0.675	0.984	0.801	0.508	0.672
Initial Proposal Tone and Length	0.673	0.978	0.798	0.504	0.667

Performance of ADT classifiers using features from each category

Features	Importance	ρ	p
Number of comments	0.08	-0.053	< 0.0001
Maximum Wikipedia age of participants	0.06	0.12	< 0.0001
Cognitive tone of Rfc	0.06	-0.049	< 0.0001
Average Wikipedia age of participants	0.06	0.03	< 1
σ of Wikipedia age of participants	0.04	0.215	< 0.0001
Sum of edit counts of participants	0.04	0.147	< 0.0001
Average edit counts of participants	0.04	0.146	< 0.0001
Number of participants	0.04	0.13	< 0.0001
Average reply depth of comments	0.04	-0.13	< 0.0001
Average number of replies	0.04	0.061	< 0.0001
Affective tone of RfC	0.04	-0.054	< 0.0001
Wikipedia age of RfC initiator	0.04	0.028	< 0.005
Hostile tone of initial proposal	0.04	0.013	< 0.5
First person singular word usage of Rfc	0.04	0.015	< 0.5

Top 14 features in the ADT model incorporating all data, including correlation to closure

Features	Importance	ρ	p
Number of comments	0.08	-0.053	< 0.0001
Average reply depth of comments	0.04	-0.13	< 0.0001
Average number of replies	0.04	0.061	< 0.0001

Features related to size and shape of discussion within top 14 features

Features	Importance	ρ	p
Number of participants	0.04	0.13	< 0.0001

Feature related to participant interest within top 14 features

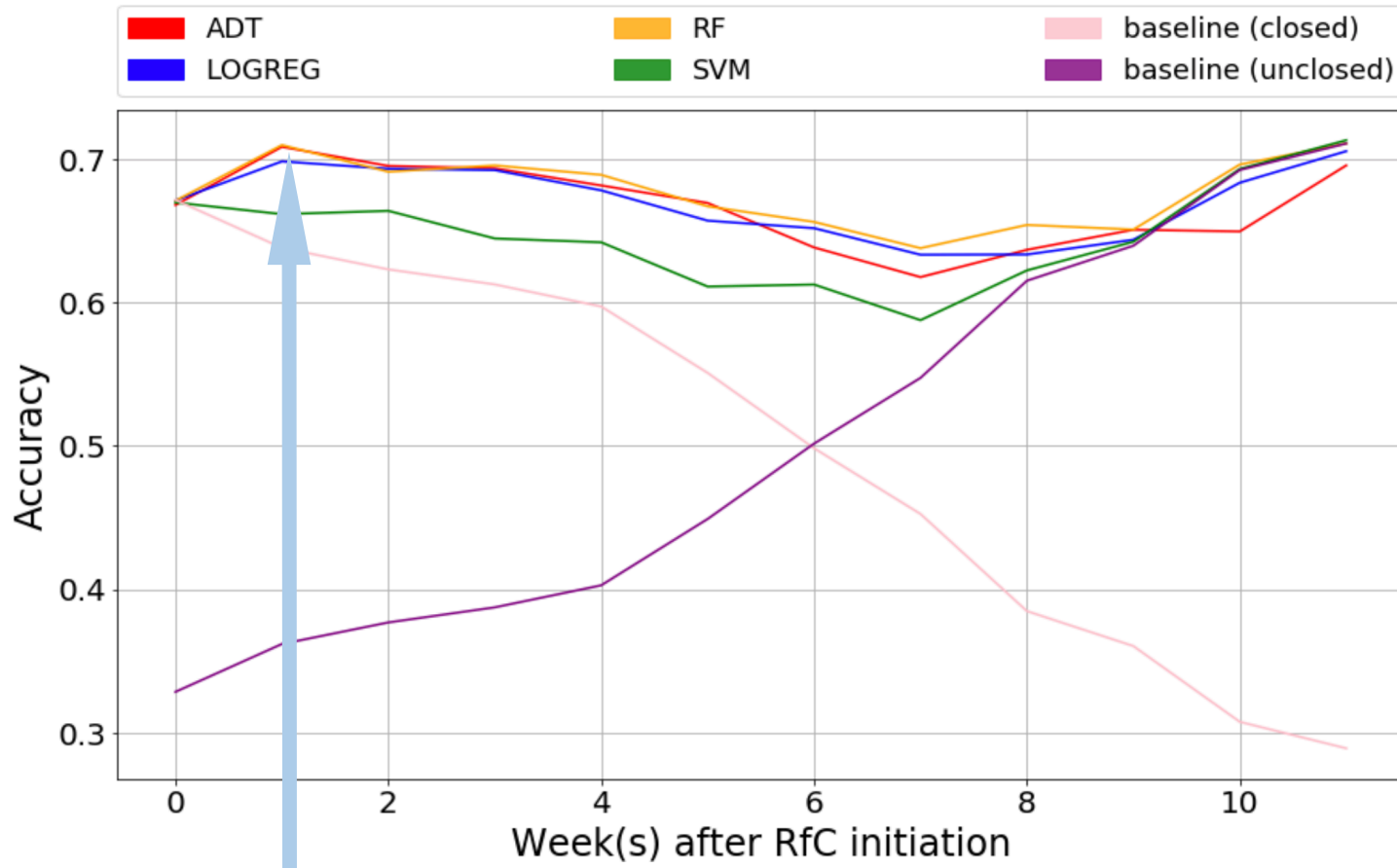| Features	Importance	ρ	p
Maximum Wikipedia age of participants	0.06	0.12	< 0.0001
Average Wikipedia age of participants	0.06	0.03	< 1
σ of Wikipedia age of participants	0.04	0.215	< 0.0001
Sum of edit counts of participants	0.04	0.147	< 0.0001
Average edit counts of participants	0.04	0.146	< 0.0001

Features related to participant experience within top 14 features

**How soon** after an RfC is initiated can we predict the likelihood of closure with reasonable accuracy?

How soon after an RfC is initiated can we predict the likelihood of closure with reasonable accuracy?

- Immediately after initiation, 1 week, 2 weeks, ... ,11 weeks

- All 61 features + number of days since the last comment up to the current time
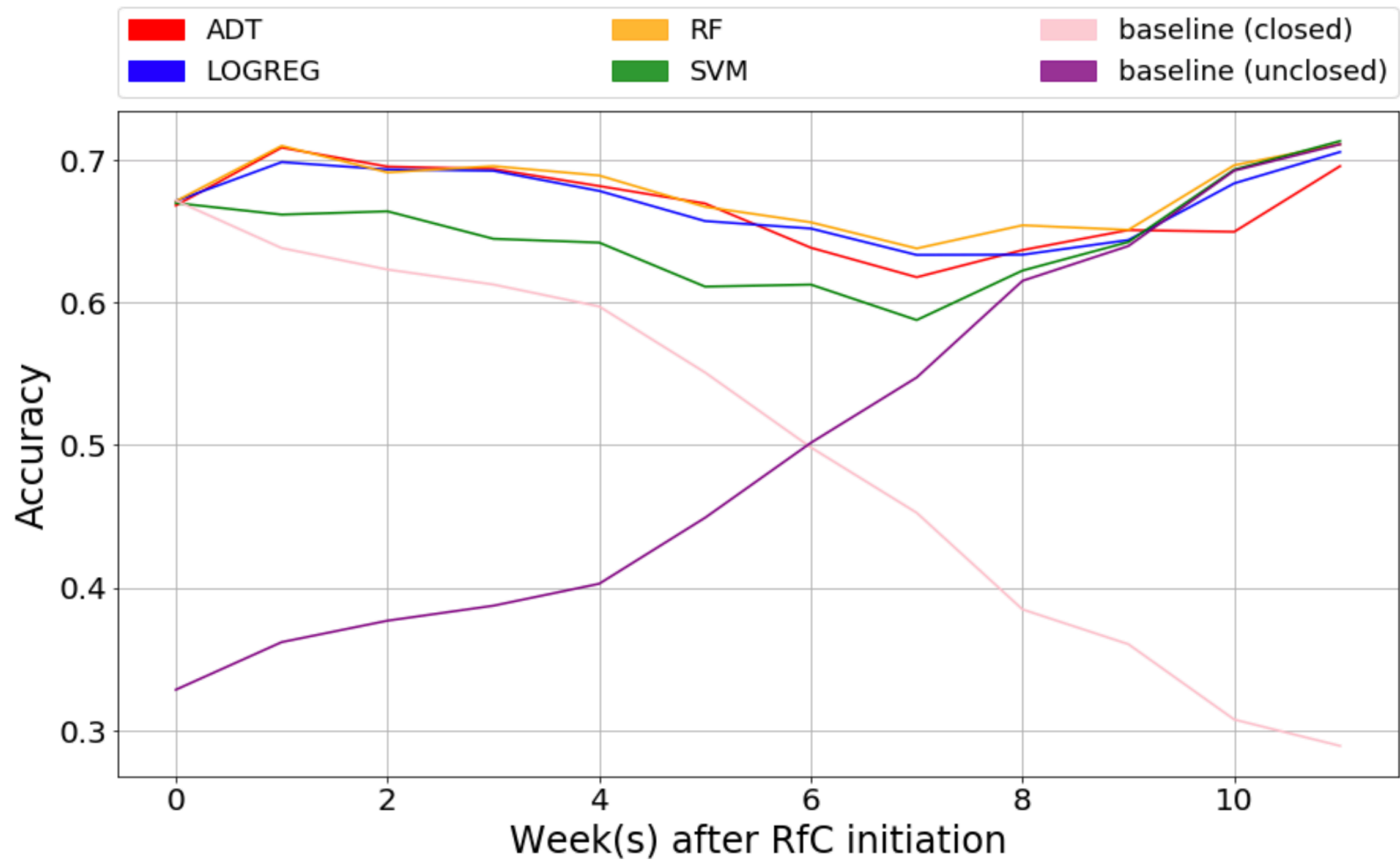
Can predict **above 70%** accuracy as early as **one week after initiation**
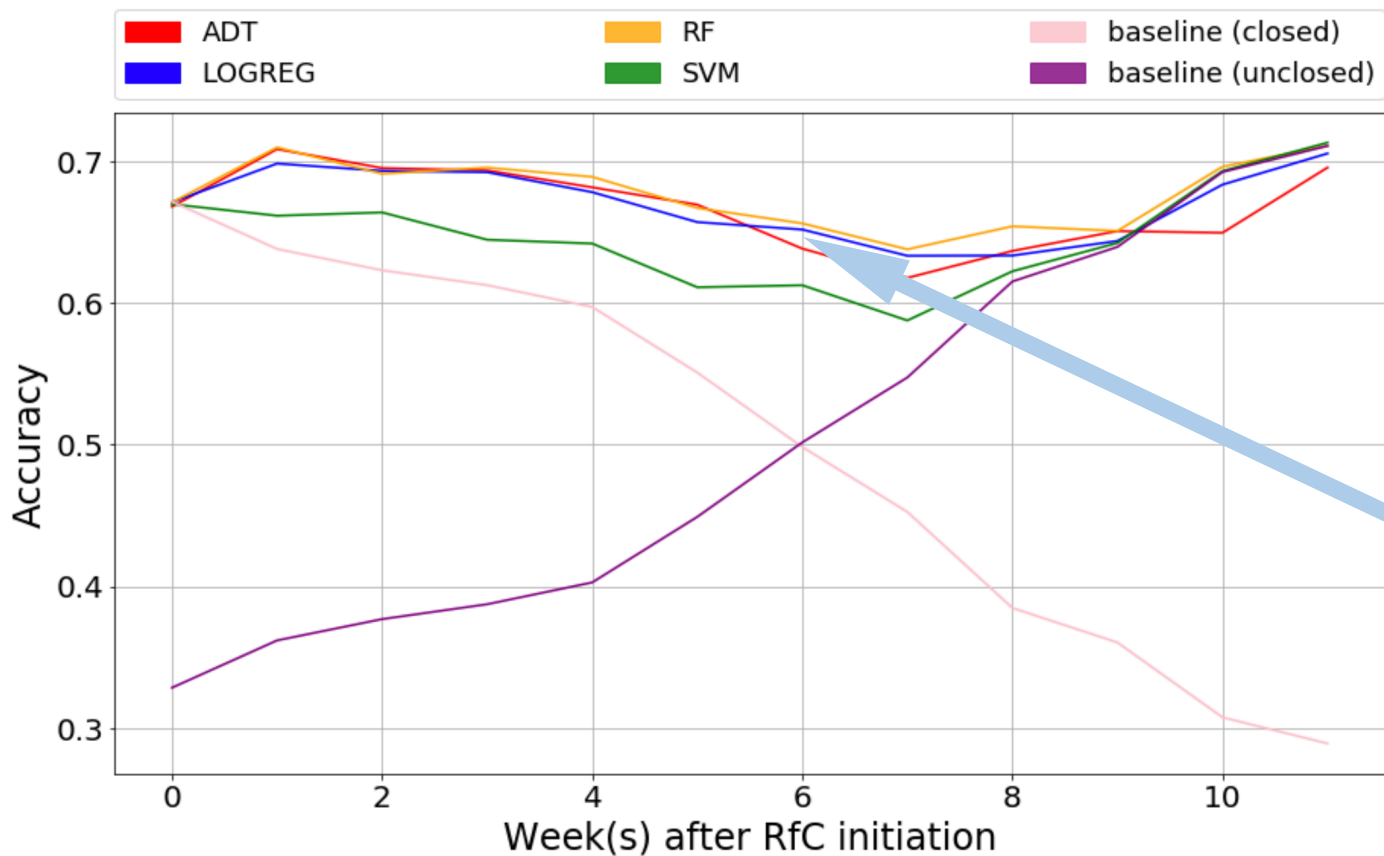
When it's a 50/50 chance at **6 weeks,** best models improve over the baseline by **over 15%**

Implication of timely models

Show **which features are important at a time point** along with prediction of outcome

Initiator/participants can **take actions using the top features**

Implication of timely models

After 2 weeks the models show participants' expertise level is crucial.

Participants invite experienced editors to the discussion.

Summary of contributions

1) New comprehensive **dataset of RfCs**
   - https://figshare.com/articles/rfc_sql/7038575

2) 1/3 of RfCs **do not get closed at all** and many do not get closed within time

3) Qualitative study showing insight from the closers as to **why** this is the case

4) New **models** to help predict which RfCs are likely to go stale

- Paper: trusttri.github.io/papers/wiki_deliberation.pdf

Jane Im, Amy X. Zhang, Christopher J. Schilling, and David Karger. 2018. Deliberation and Resolution on Wikipedia: A Case Study of Requests for Comments. In Proceedings of the ACM on Human-Computer Interaction, Vol. 2, CSCW, Article 74 (November 2018). ACM, New York, NY

- Contact: imjane@umich.edu, axz@mit.edu

Appendix

RfC Category	Num RfCs initiated	RfC Category	Num RfCs initiated
Politics, government & law	2650	Religion & philosophy	949
History & geography	2573	Wikipedia style & naming	749
Biographies	2123	Wikipedia proposals	634
Wikipedia policies & guidelines	1767	Economy, trade, & companies	585
Uncategorized	1732	Wikipedia technical issues & templates	381
Society, sports & culture	1634	Language & linguistics	372
Art, architecture, literature, & media	1601	WikiProjects & collaborations	259
Maths, science, & technology	1165		

Number of RfCs issued from 2004 to 2017 by categories.
One RfC may have multiple categories, for example, { { rfc | econ | bio } }.