

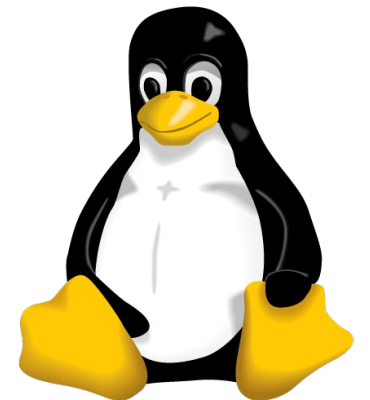
Let's Talk Open-Source – An Analysis of Conference Talks and Community Dynamics

Kimberly Truong
Oregon State University



Motivations

- There's a disconnect between practitioners and researchers
- What do practitioners experience? What do they want to talk about? What is **relevant**?
- Community interactions in a project are an indication of the project's success. So what do practitioners choose to share?
- How can we further sustain and support open-source software?



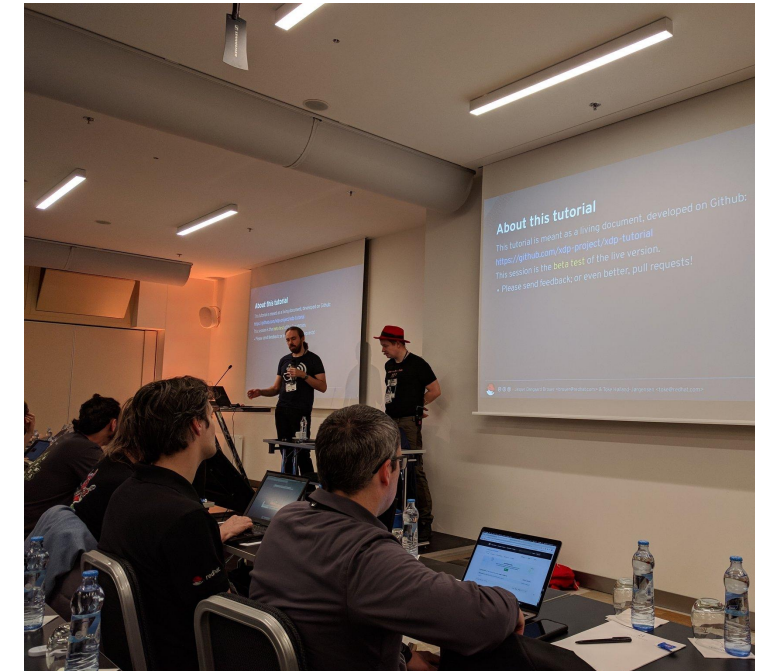
1 year, 3 conferences – What's new?



PyCon Africa 2019,
<https://www.ict4d.at/tag/pycon/>



Open Source Summit 2019,
https://live.staticflickr.com/65535/51647514916_acc4a7e90b_c.jpg



NetDev 2019,
<https://bootlin.com/blog/tag/netdev/>

Data Curation

doi.org/10.5281/zenodo.6395342

Dataset

- Large and diverse
 - Disciplines
 - Size
 - Experiences/roles
 - Countries
 - Times
- Grey literature analysis
 - First-hand accounts
 - Captures what happened in time

Tool

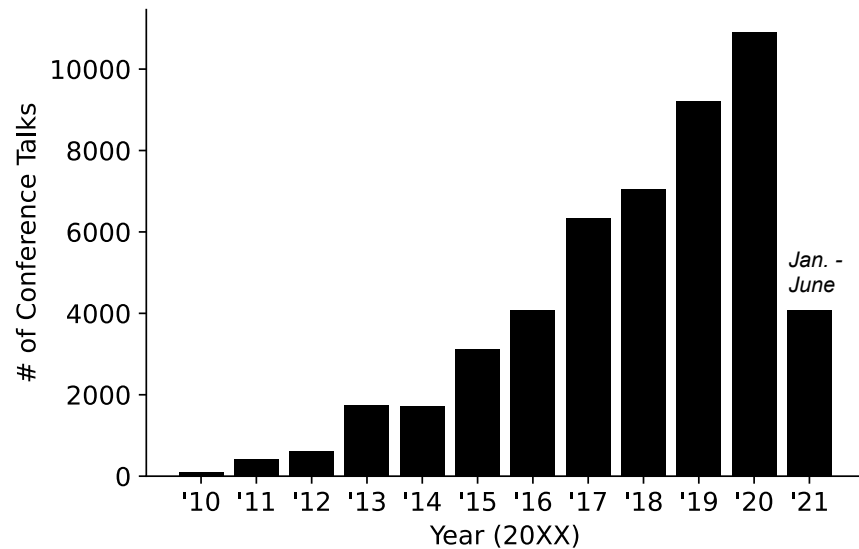
- Extract YouTube video data
 - Video transcripts
 - Full videos
 - Expand beyond open-source software



Data Curation

doi.org/10.5281/zenodo.6395342

- 87 Conferences
- 24,669 Talks



Metadata

- Focus/theme
- Size
 - # of talks
 - # of speakers
 - # of attendees
- Affiliated conferences/organizations,
- Sponsorship information
- Main (or most recent) website
- Wikipedia page.

New Findings

- Topic model analysis
 - What are the most common topics in open-source software?
 - How has the frequency of these topics changed over ten years?
- Qualitative analysis of contributor disengagement in open-source software
 - What do contributors cite as their reasons for leaving open-source?
 - What changes do they recommend?
 - What is the pattern after they disengage?

Methodology

1. Establish a **conference list**



- Google search for top 30 links
 - “open source conference”; “open source conference call for proposals”

2. Collect metadata and **filter** conferences:



- Two documented editions
- > 50 attendees or > 10 speakers/talks

3. **Compile** data



- Generate separate text files for each video.

Resulting Data

- Name of video
- Publication date
- Playlist (often the conference edition)
- Description
- Transcript*
- YouTube URL

* Not shared with the Dataset due to the YouTube License

```
1 Title: Russell Keith-Magee - Keynote - PyCon 2019
2 Publication date: 2019-05-05
3 Playlist: PyCon 2019 - Keynotes
4 Description:
5     "Speaker: [REDACTED]"
6
7 Keynote
8
9 Slides can be found at: https://speakerdeck.com/pycon2019 and https://
10 Captions:
11     00:00:41,270 --> 00:00:51,399
12 good morning
13
14     00:00:42,940 --> 00:00:53,800
15 [Applause]
16
17     00:00:51,399 --> 00:01:00,820
18 I couldn't be more excited to welcome
19
20     00:00:53,800 --> 00:01:03,699
21 you to the 2019 Pike on pike on 2019
22
23     00:01:00,820 --> 00:01:05,770
24 here in Cleveland Ohio on behalf of the
25
26     00:01:03,699 --> 00:01:08,380
27 pike on 2019 staff I want to start by
28
29     00:01:05,770 --> 00:01:09,610
30 saying thank you thank you to all the
31
32     00:01:08,380 --> 00:01:11,770
33 volunteers that make this conference
34
35     00:01:09,610 --> 00:01:13,509
36 possible to the Python Software
37
38     00:01:11,770 --> 00:01:15,820
39 Foundation for taking on the fiscal
```


Disengagement Diaries

<https://disengagement-diaries.github.io>

Method

1. Generate a sample from keywords
 - 'leave', 'abandon', 'pause', 'hostility', etc.
2. Thematically code top 1500 transcripts.
3. Record data.
 - Reasons for disengagement
 - Recommended interventions
 - GitHub Activity

Not Enough Time (Internal) No Longer Enjoyable

Who?

Github: Removed for privacy

Project: [Django](#)



Why?

Anna cited three reasons for leaving their open source communities at DjangoCon US in 2018. These reasons arose because they had to change their priorities as new events unfolded in their life. The first issue was the lack of a work life balance as newer, more important responsibilities arose.

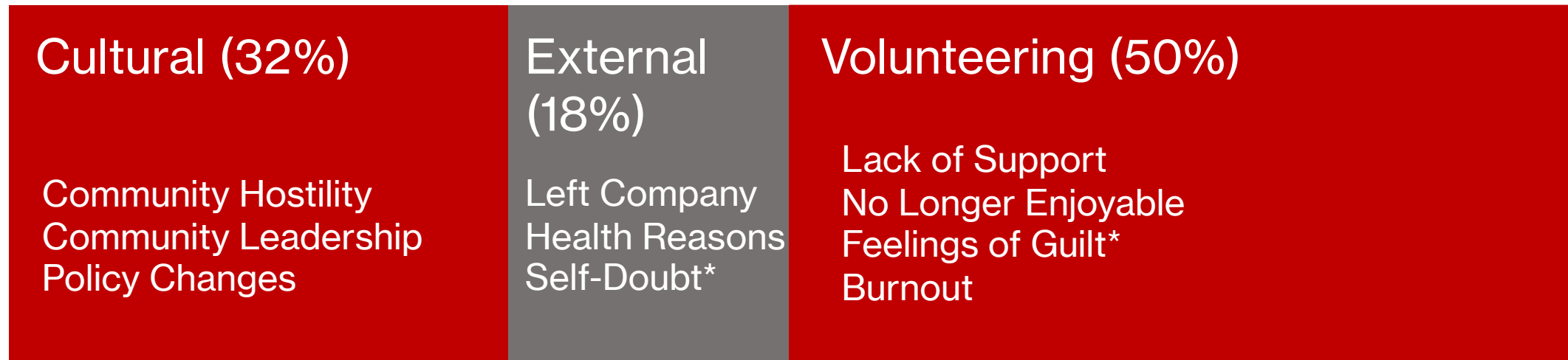
"People's living situations change or your priorities change and so people may say after a while I can't do it anymore for whatever reason." (15:44)

"Tech kind of takes over our whole life. I'm definitely guilty of that so I'm kind of like actively saying no." (30:00)

This led to a lack of time internally due to their shift in priorities. Although they still appreciated the Django community, Anna found contributing to open source to take away from more significant priorities, like a new job and changing living situations.

"I'm actually moving cities so I also kind of want to prioritize my personal life a little more I've served the community a lot and now it's kind of like time to look after myself a little more I'm out I also took on this

Results



- Acknowledge contributors “have a life”.
- Encourage understanding and inclusive communities.
- Encourage a work like balance within communities.

Topic Modelling

Method

1. Preprocess text
2. Trial and error with 7 to 40 topics.
3. Consolidate topics

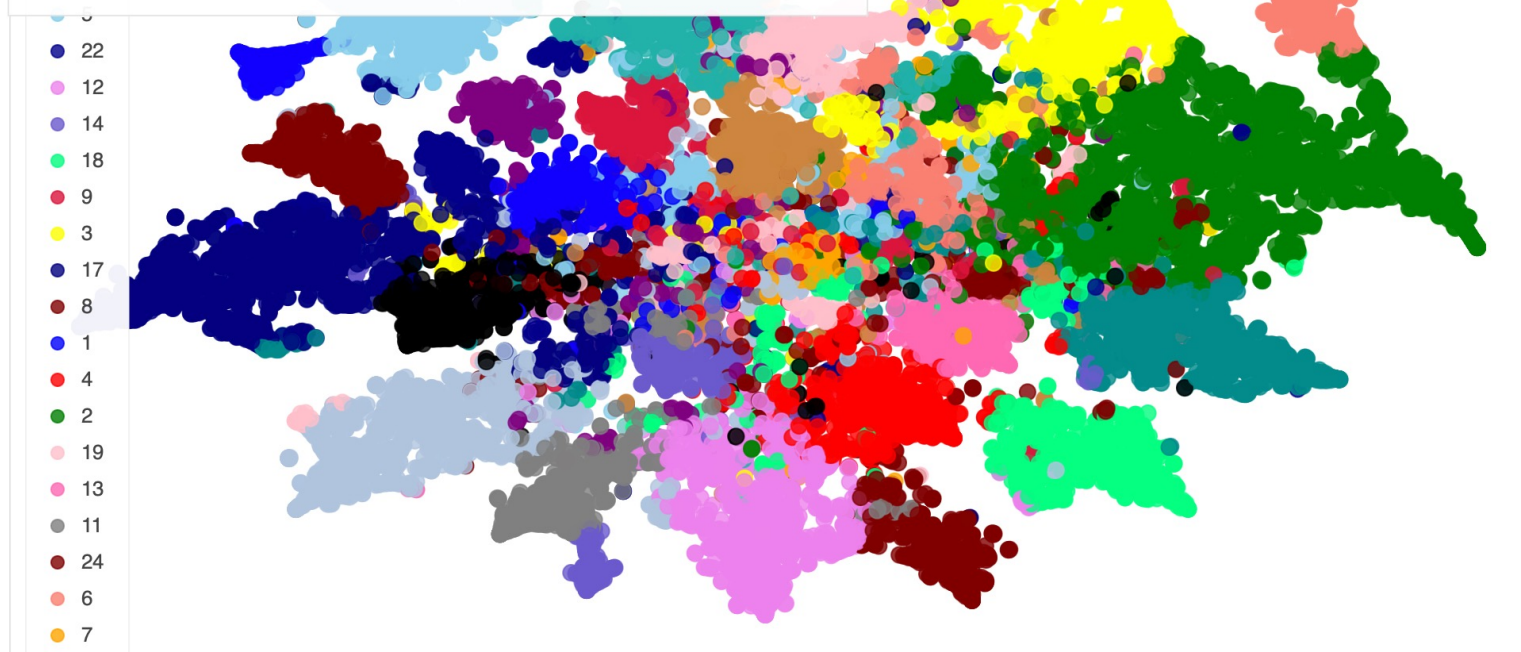
T-SNE_with_dates

content: Title: LPC2019 , Conference: LPC2019 , Date: 2019, KeyWords: wanna slot encryption password - Topic: 20

content: Title: Applications of TPM 2, Conference: DevConfCZ 2019, Date: 2019, KeyWords: measurement encryption hash trusted - Topic: 20

content: Title: Better Data Security with Commodity TPM Chips , Conference: Open Source Summit , Date: 2020, KeyWords: chip encryption boot password - Topic: 20

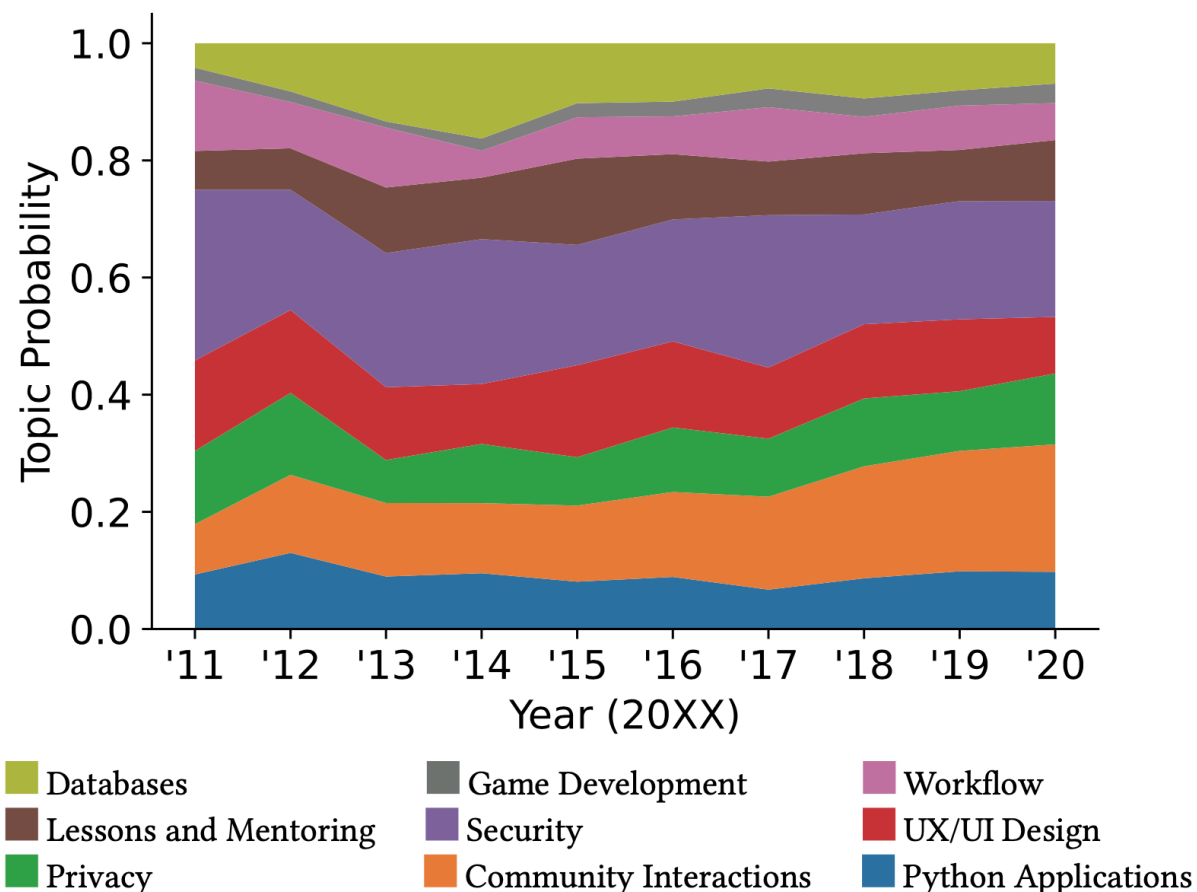
content: Title: Applications of TPM 2, Conference: DevConfCZ 2019, Date: 2019, KeyWords: measurement encryption hash trusted - Topic: 20



Topic Modelling

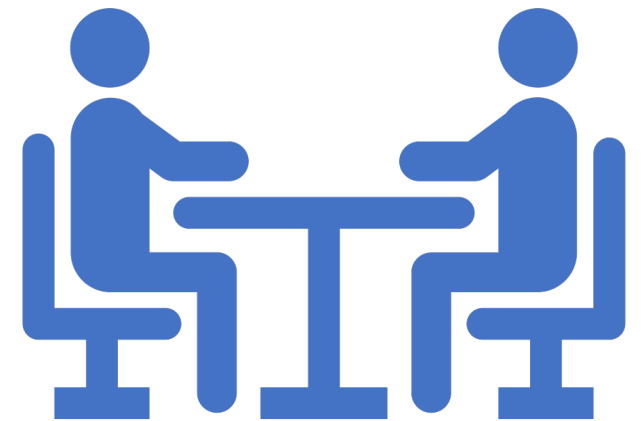
Observations

- Community interactions is a more common topic
- Most disengagement talks were found in more recent years of Community Interactions
- Talks about Databases have decreased



Summary

- A dataset of 24,669 open-source software conference talks
- Contributed to a database of common reasons for open-source disengagement and possible interventions.
- Present nine common topics and their frequency from 2011-2021



Possible Future Work

- Analyze the popularity and diffusion of tools and practices
- Understand common challenges discussed by practitioners
- Identify major influences on certain communities
- Filter to create samples by time, focus, size, etc.

Dataset: doi.org/10.5281/zenodo.6395342

Disengagement Diaries: <https://disengagement-diaries.github.io>

e. truonkim@oregonstate.edu

GitHub: /KimberlyTruong

LinkedIn: /in/kimberly-le-truong-cs

