Exploring the Open Source Scala Community Network on Twitter

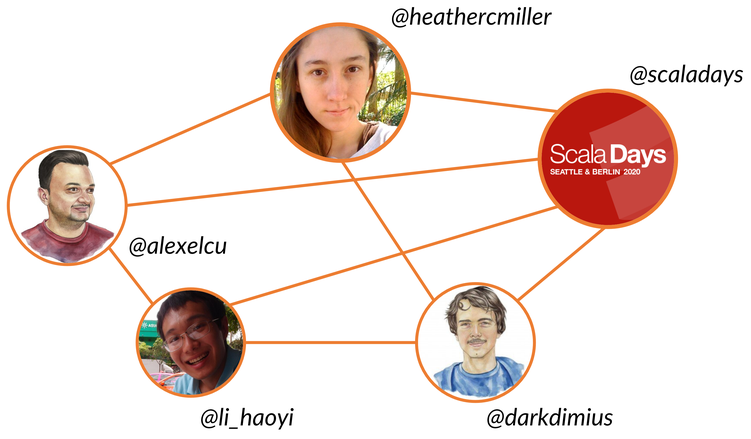
Mengchen (Sam) Yong, Hemank Lamba, Prof. Heather Miller, Prof. Bogdan Vasilescu

Open-source developers often share and receive ideas about new projects and technologies on social media platforms. Their interactions on the platform form a complex network of relationships.
Our case study of the Scala community on Twitter explored the research question:

How do technical ideas spread through a social network?

Part I – Network Construction

We created a follower relationship graph of
- 112 known Scala-related "seed" users and
- around 20,000 "connection" users who follow at least 5 "seed" users.



We then used Louvain method for community detection to partition the network into 5 clusters. Domain expert Prof. Miller verified that most top connected users are indeed well-known members of the Scala community.

Part II – Scala Ideas in Tweets

We collected the most recent around 3200 tweets from all "seed" and "connections" users. We then used python nltk library to find most common ngram collocations among all tweets.

To identify ideas relevant to Scala, we filtered the ngrams to only include those that contain at least one word from the glossary (114 terms) of a recommended textbook on the Scala official documentation website.

*"When I teach the concept of **referential transparency** to my students, I use memoization as a litmus. Probably thanks to @BartoszMilewski, though I can't recall where I got the idea." – tweeted by @kenpu*

*"Everything you ever wanted to know about using **sealed traits** in Scala http://t.co/qPZGpjeDzf" – tweeted by @underscoreio*

Bigram (▼ Frequency)	Specific to…
referential transparency	Functional
sealed trait	Scala
mutually recursive	Functional
pattern matching	Functional
visitor pattern	Object-oriented
companion object	Scala
partially applied	Functional

What's Next

Refine the vocabulary

The glossary we used lacks some important Scala terms (e.g., "implicit") and contains many general functional programming or object-oriented programming terms. We plan to

1. Collect tags that often co-occur with tag "scala" on Stack Overflow

scala | types | scala-cats | category-theory | type-constructor

2. Intersect the tags with indexes from a few Scala textbooks

Identify more relevant ngrams

We will not only find ngrams that contain at least one term from the vocabulary but prioritize ngrams that contain two- or three-word phrases.

Follow-up research questions

- How have the ngrams found in Part II spread on Twitter over time?
- How have the ngrams found in Part II spread within sub-communities we discovered in Part I?
- Can we characterize communities instead by common ideas (ngrams)?