# Express More Than a Thousand Words

## problem

- What are **relations** between emojis ?
- Can we identify emoji synomins?
- Can we define the meaning of emoji?
- Can we **assign** emoji to text?

## METHOD

### 1 Data acquisition

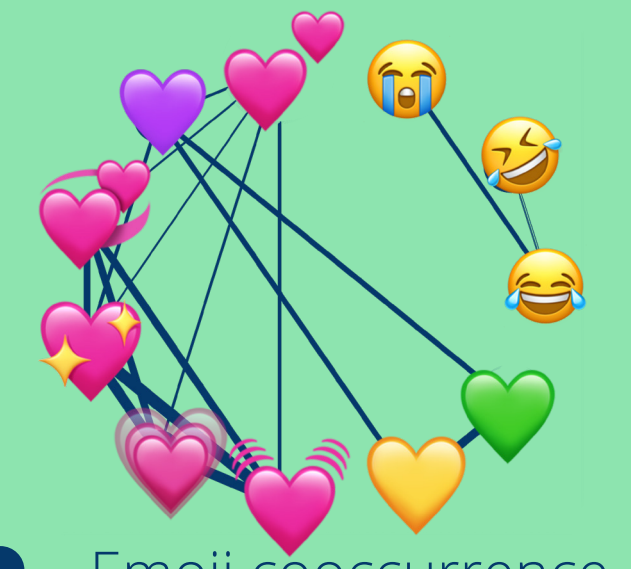TWINT - a Twitter scrapper
**+**
C E L E R Y
**+**
mongoDB.

- Twitter scraping using twint
- 45 distinct emojis
- 0,5 + M tweets
- 8 months:  4-11/2019
- selected hours

### 2 Preprocessing

- tokenization
- cleaning: links, emails, mentions, dates, phones, ...
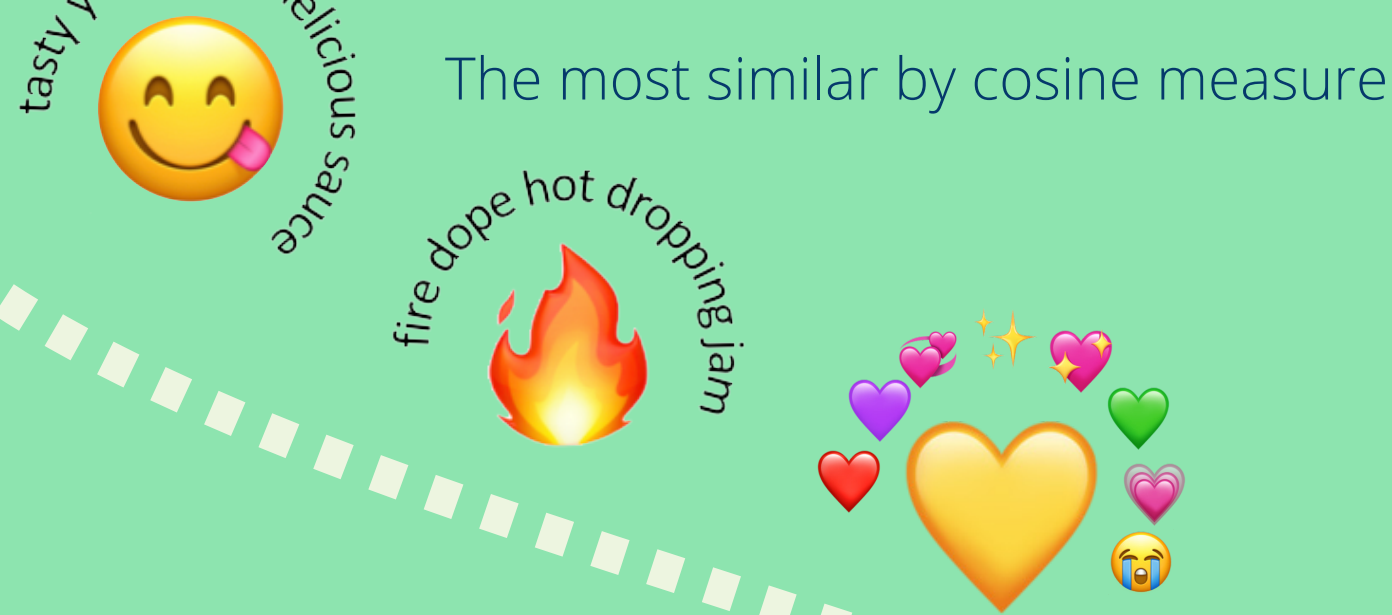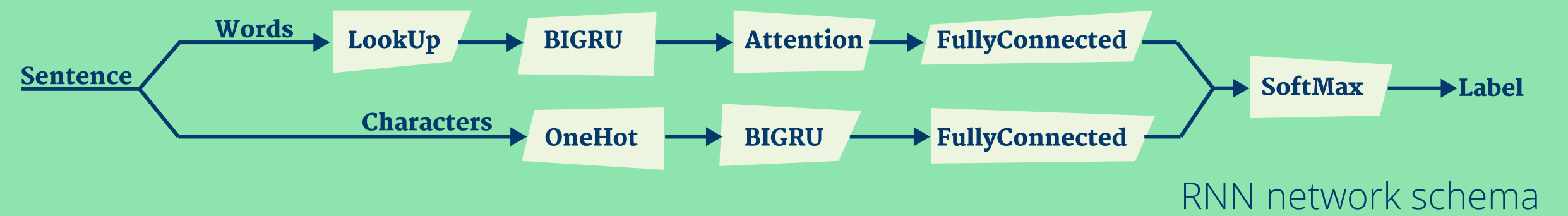- separation: text + emoji

### 3 Analysis

happy birthday

- word cloud for 🎉

Emoji cooccurrence with  threshold=600

### 4 Embeddings

fastText 🙂 → [124, 43, 53, ....]
Hello → [561, 52, 22, ... ]
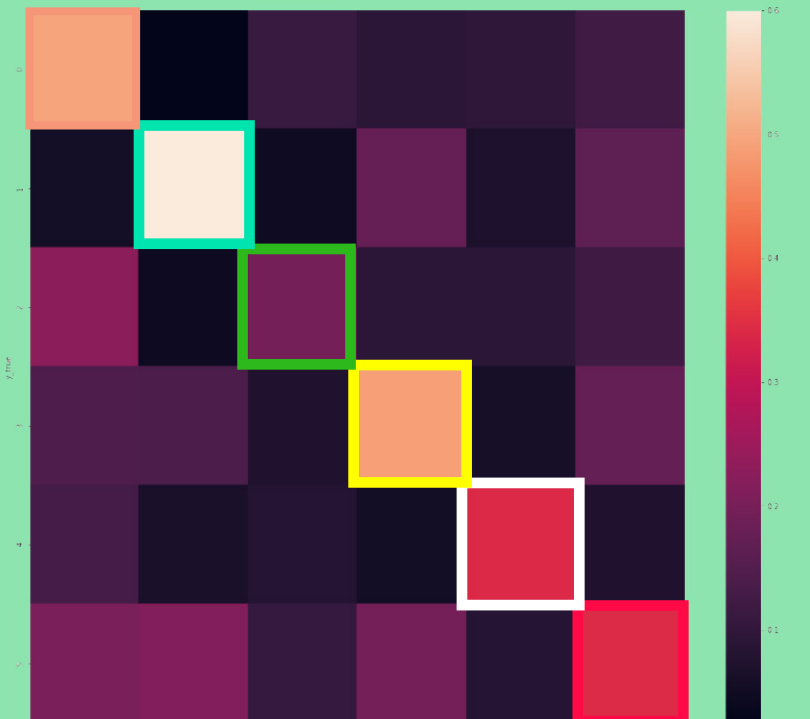
The most similar by cosine measure

tasty yum salad delicious sauce 😋

fire dope hot dropping jam 🔥

💖💕 💚💞 ❤️💛💖😭

### 5 Clustering

2 component PCA

- **KMeans**: PCA for 6 clusters
  - 😂 together with 🤣
  - 😘 together with 🙂 ?

- **DBScan**:
  - only few groups
  - a lot of outliers
  - the strongest group is 💖💗💕💞💕

### 6 Classification

Sentence → Words → LookUp → BIGRU → Attention → FullyConnected
Sentence → Characters → OneHot → BIGRU → FullyConnected
→ SoftMax → Label

RNN network schema

**Classification:**

Results:
- acc:uracy: 15%
- f-score: 13%
- min. f-score: 2% 🤣💞
- max f-score 33% 🤣💞

Discovery:
- There are emojis that could be seen as synonymous like ❤️ and 💖
- Some emojis are rarely assigned to sentences by model
- "Sounds perfect another very talented very fine great choice by" (true: 🤣 preicted: 👍)

The same groups as on 5

**Classification of emoji clusters:**

Results:
- accuracy: 44%
- f-score: 42%
- significantly better

Discovery:
- More clusters generate more unbalanced dataset: ex. 29% vs 1.5% of data
- Some clusters are too general, diversified

The same groups as on 5

**Attention discovery:**

Correct:
i know it is and i am about to burst into tears just watching it again 😢
i hope i win your giveaway wish me luck 🙏
missing our little bunny so much my heart hurts 💔

Incorrect:
where did yall come from never had this much likes lmao real talk tho i had an amazing day
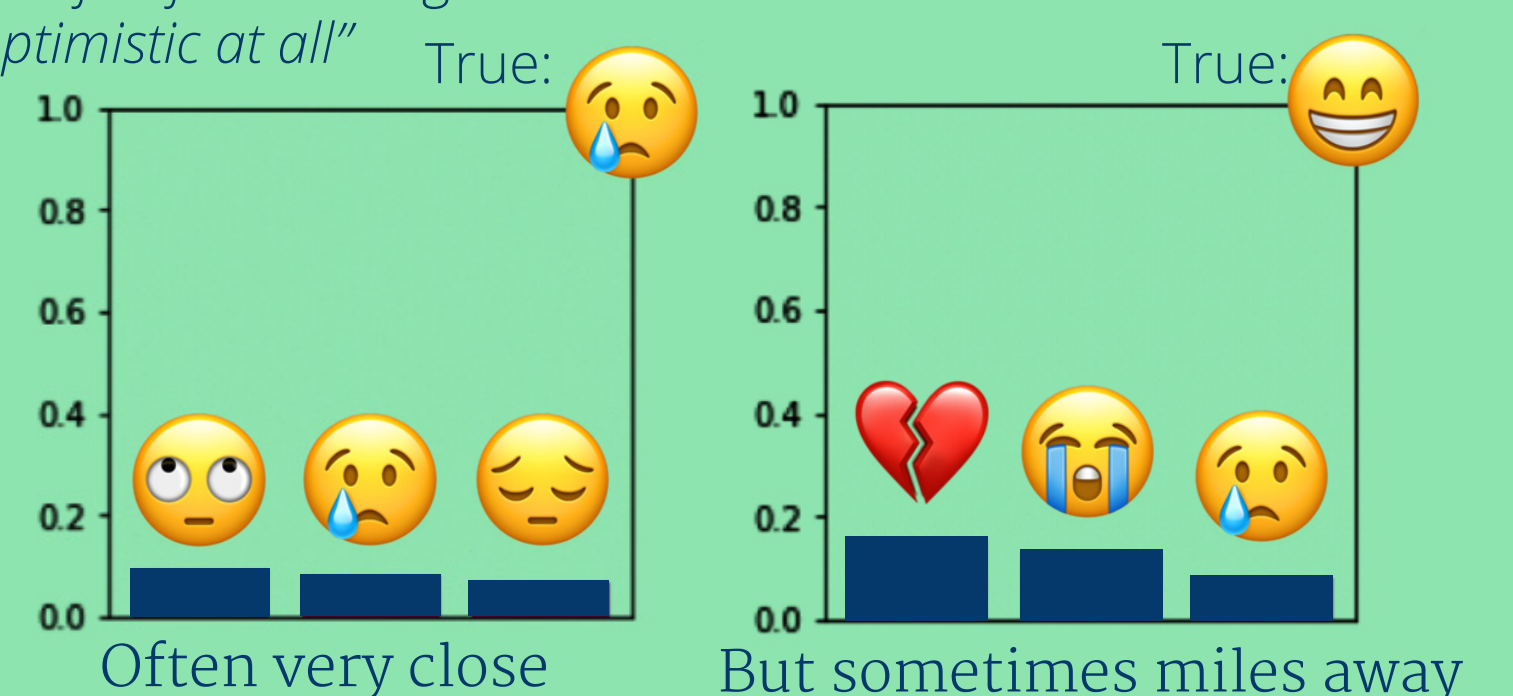true: ❤️ predicted: 😂

Controversial:
they shudve ask u to give them a lil gym workout
true: 🙂 predicted: 💪
be strong enough to walk away from what s hurting you and be patient enough to wait for the blessings you deserve
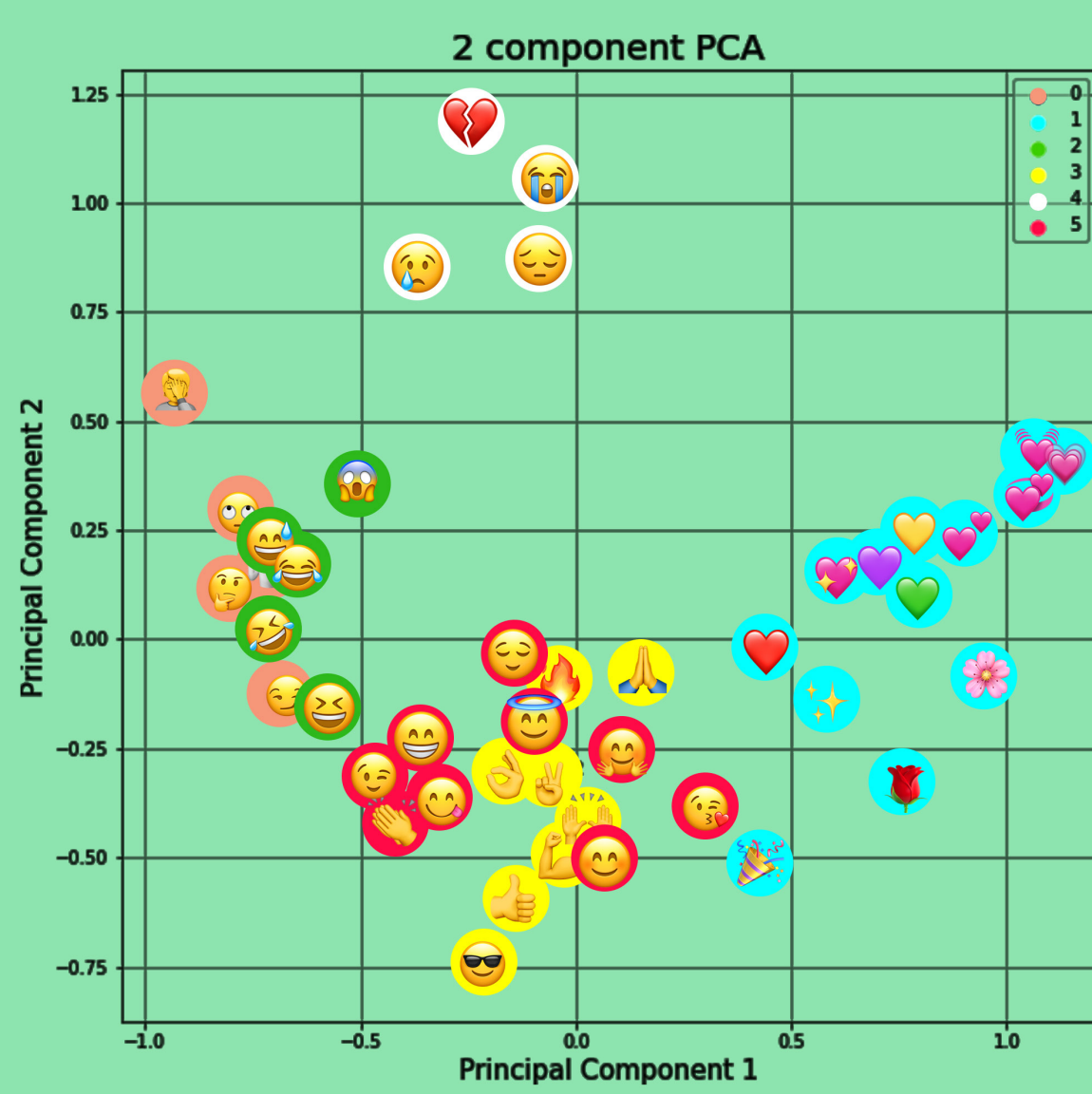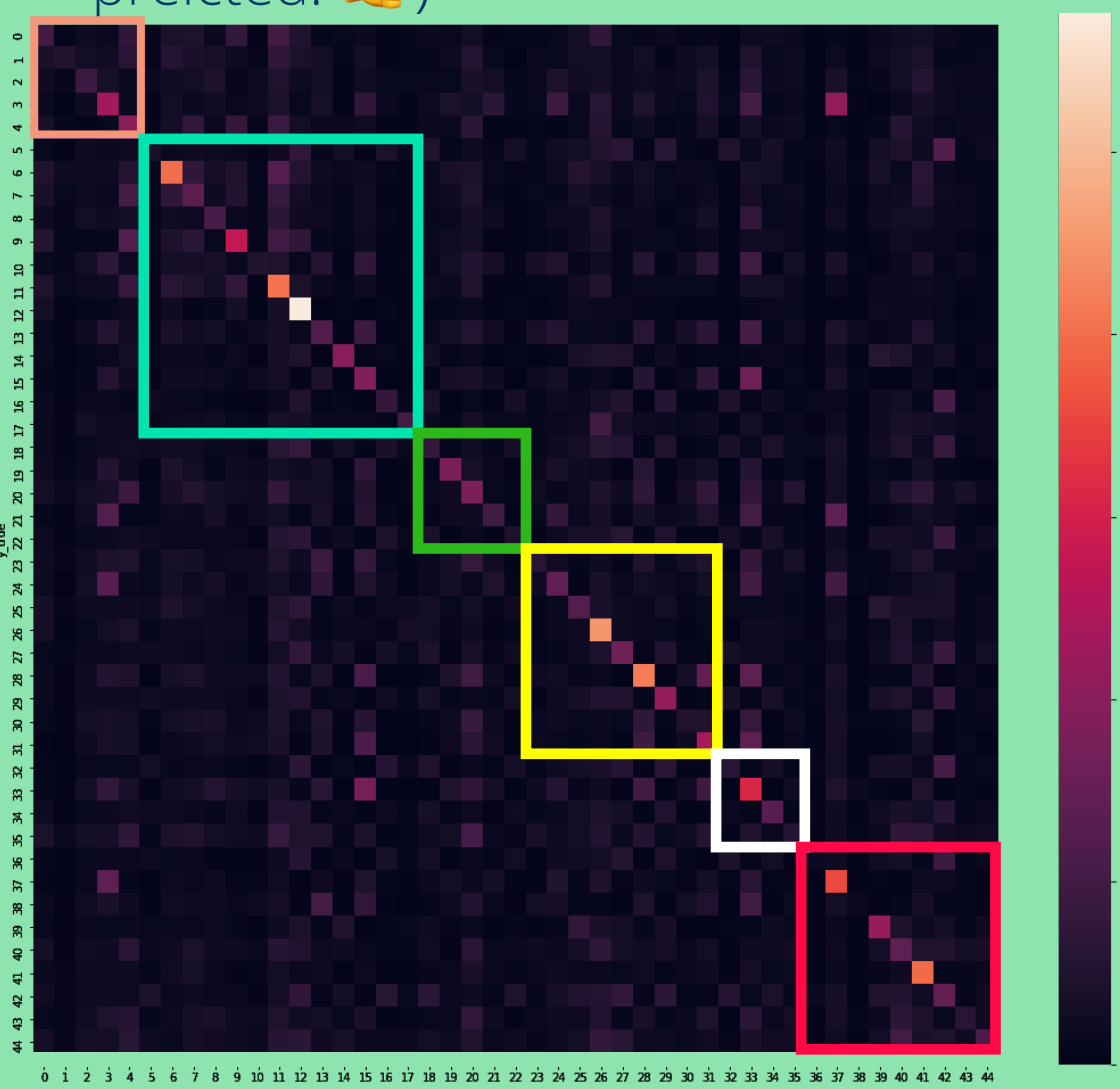true: 🙂 predicted: 🙏

**Soft evaluation**
*Is the right answer close or far away?*

"homo sapiens is too aggressive and too fearful to change i am not optimistic at all" True: 😢

Often very close

"he did nothing to me so i can not hate" True: 😁

But sometimes miles away

## Takeaway message!

- Some emojis are used in very **wide** and some in **narrow** context

- Predicting emoji to text is difficult task because emojis are used in a subjective way. There is no objective way

**Authors**
😂 Miriam Jańczak
🤣 Anna Makuś
😅 Gabriela Relikowska
😆 Adrianna Zielińska