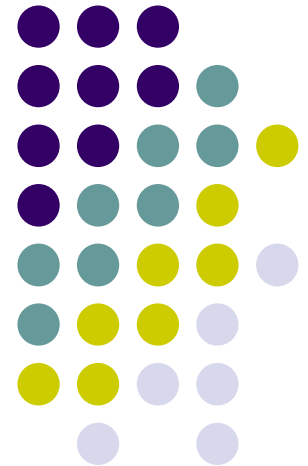


# Big Data Analytics

---

## Fifth Assignment

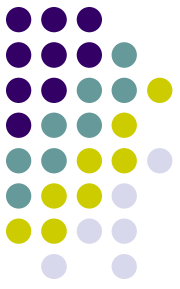


# Deadline: 15 February 2021



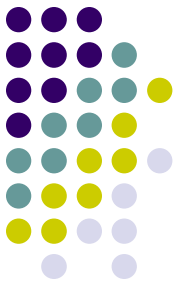
1. Retrieve the last 1,000 tweets from the official account of Joe Biden “@JoeBiden” via `rtweet`. Do you remember how to do it? `get_timeline(c("JoeBiden"), n = 1000, include_rts=TRUE)`

# Deadline: 15 February 2021



2. Run a STM on the set of tweets using as covariate for assessing the **topic prevalence** in your Structural Topic Model the day in which a tweet has been posted that you treat as a “continuous” variable
  - REMEMBER! When you assess a “continuous” topic prevalence model, your covariate should be a number, not a character or a date! Use the option: *as.numeric(the name of your time variable)* to convert your time-variable into a number. Save the new variable you just created in the data frame you got via your `rtweet` query. Then from this data frame create first your corpus, then your dfm, then convert your dfm into a stm object!

# Deadline: 15 February 2021



3. Briefly comment your results (for example, the content of the topics, the number of them, the results you got for topic prevalence, etc.)