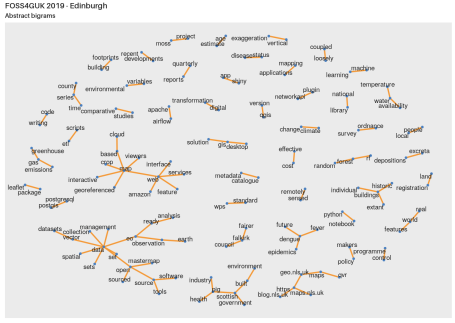
[](https://scottishsnow.files.wordpress.com/2019/09/abstract_bigram.png)

FOSS4GUK 2019 abstract bigrams

A bi-gram is a pair of words. In the last sentence “is a”, “a pair” and “pair of” are all bi-grams, they are pairs of words which are adjacent. As before each abstract is stored in a separate file, so I’ve read each of those in and then turned them into a tidy bi-gram table:

library(tidyverse)

library(tidytext)

library(tidygraph)

library(ggraph)

library(extrafont)

# ----------------------------

data("stop\_words")

f = list.files("~/Cloud/Michael/FOSS4G/talks/abstracts\_clean/")

abstracts = lapply(f, function(i){

read\_table(paste0("~/Cloud/Michael/FOSS4G/talks/abstracts\_clean/", i),

col\_names = F) %>%

gather(key, word) %>%

select(-key) %>%

add\_column(author = str\_remove(i, ".txt")) %>%

unnest\_tokens(bigram, word, token = "ngrams", n = 2)

})

abstracts = do.call("rbind.data.frame", abstracts)

bigrams = abstracts %>%

separate(bigram, c("word1", "word2"), sep = " ") %>%

filter(!word1 %in% stop\_words$word[stop\_words$word != "open"]) %>%

filter(!word2 %in% stop\_words$word[stop\_words$word != "open"]) %>%

filter(!str\_detect(word1, "[0-9]")) %>%

filter(!str\_detect(word2, "[0-9]")) %>%

filter(!str\_detect(word1, "NA")) %>%

filter(!str\_detect(word2, "NA"))

bigram\_counts = bigrams %>%

count(word1, word2, sort = TRUE)

Then I write out a graph to a png. There’s some nifty stuff on the repel line which keeps labels on the plot and I’ve event put the text into the conference font:

png("~/Cloud/Michael/FOSS4G/talks/abstract\_bigram.png",

width=1200, height=850, res=110)

bigram\_counts %>%

filter(n > 1) %>%

as\_tbl\_graph() %>%

ggraph(layout = "fr") +

geom\_edge\_link(width = 1.1, colour = "#f49835") +

geom\_node\_point(colour = "#497fbf") +

geom\_node\_text(aes(label = name),

colour = "grey10",

vjust = 1, hjust = 1,

repel = T, force = 0.1, box.padding = 0)+

labs(title = "FOSS4GUK 2019 - Edinburgh",

subtitle = "Abstract bigrams") +

theme(text = element\_text(family = "Aileron SemiBold",

colour = "grey10"))

dev.off()