EPFL

DATA SCIENCE IN PRACTICE MGT-415

Christopher Bruffaerts

Seminar #3

Assistant: Omar Ballester

Networking !

Your department at the Telecom. company is *on fire*! Since they hired you, they have boosted their productivity, and are now capable of taking data-driven decision much more quickly. Given the development of the new techniques, your boss has decided he wants to patent the data processing methods that you use, in order to add some value to the company. While this is great news... it means more work for you! It is the first time the company looks into this kind of patents, so they know very little. You have been asked to take all the patents from 2016 in the relevant class, and study your competitors. You are specially interested in knowing : who is the central player in this technology area ? Who is the leader ? Who is an irrelevant actor ?

The lawyers will subsequently use this information in order to study their patent portfolios and make the best of our resources !

For this assignment, use the data from *Google Patents Datasets*. You have to register to the free-tier of Google BigQuery, and use the patents-public-data/publications table. Identify and extract all patents from 2016 from that have a patent classification (cpc) starting with G06Q30/06... Then, extract all the patents from 2016-2020 that cite these patents, and build a directed network of patent citations. Try to answer the questions stated above.