Team +id +members +skills +date +location +generate sparse matrix() +get one hot() +remove outliers() +get_stats() +plot stats() +read_data()

Publication

members = authors skills = fields of study date = year location = venue

+title +references

read_data()

Patent

members = inventors skills = subgroups date = year location = country

+title +withdrawn

read_data()

Movie

members = cast'ncrew skills = genres date = year location = none

+title +runtime

read_data()

Repository

members = developers/contributors skills = programming Languages date = created at location = none

+name +nforks

read data()