Statistic	Value
# of documents in corpus	6.9 million
# of unique authors	8.3 million
# of unique keyphrases	823,677
# of unique venues	23,672
avg. # of incoming citations	7.4 ($\pm$ 38.1)
avg. # of outgoing citations	8.4 ($\pm$ 14.4)
size of training set [years 1991 to 2014]	5.5 million
size of dev set [years 2014 to 2015]	689,000
size of test set [years 2015 to 2016]	20,000