Homework 3

Distributed Computing

By Su-Young Hong

1.1: See homework3.py

1.2: See homework3.py

1.3: See homework3.py

1.4: See homework3.py

1.5: 130160392 and

1.6:

a) AUTHS: [Row(page=u"United\_States'\_Country\_Reports\_on\_Human\_Rights\_Practices", auth=0.37843687670043263), Row(page=u'Geographic\_coordinates', auth=0.29735638171906054), Row(page=u'2008,\_Madrid\_plane\_crash', auth=0.2890848187118112), Row(page=u'2007,\_Canada\_budget', auth=0.26910488097800983), Row(page=u'Biography.com', auth=0.15610089669910465), Row(page=u'2006-04-01', auth=0.1477031658607502), Row(page=u"United\_Kingdom's", auth=0.14062666454972816), Row(page=u'Music\_genres', auth=0.13116609590214443), Row(page=u'Record\_label\_Hypnos', auth=0.1249404704440951), Row(page=u"France's\_National\_Archives", auth=0.12470731229221291), Row(page=u'2005,\_S\_1313', auth=0.12121296804192645), Row(page=u"England's", auth=0.11927402081350351), Row(page=u'Population\_dispersal', auth=0.10084140536037495), Row(page=u"Canada's", auth=0.1005820042996229), Row(page=u"Germany's\_Basic\_Law", auth=0.09625697104934057), Row(page=u'Internet\_Movie\_Databse', auth=0.09532433844181178), Row(page=u'Biological\_clock', auth=0.09110629551230656), Row(page=u'Time\_zone\_(disambiguation)', auth=0.090571343908421), Row(page=u'Area(Travel\_Agency)', auth=0.08834473402507846), Row(page=u'2004-05\_AHL\_season', auth=0.0831133110761372)]

>>> hubs\_results = sqlContext.createDataFrame(final\_hubs, ['page','hubs'])

b)

[(u'77935', 0.009848901978055179), (u'2496661', 0.00974483484765591), (u'79583', 0.009727889113675819), (u'84707', 0.00958871973804183), (u'3211540', 0.009510428985060758), (u'3000061', 0.009497314290398944), (u'81615', 0.009336508448418787), (u'76573', 0.009293679261479205), (u'65954', 0.009288383140475878), (u'4529923', 0.009158187518859206), (u'1331646', 0.008989923453599892), (u'3211271', 0.008975654608151772), (u'2638932', 0.008939934619409683), (u'75323', 0.008896146900461128), (u'1715261', 0.008893359847670326), (u'47849', 0.00888938260905526), (u'2496868', 0.008856839881086173), (u'65064', 0.008838156228705273), (u'3283893', 0.008823480702114821), (u'1715441', 0.008813295141250531)]

1.7:

a)

[(u'4504608', 0.006158049357462614), (u'1273793', 0.006103541463316293), (u'5296341', 0.00608430646587461), (u'2278916', 0.006069132899760532), (u'3102538', 0.0060357289399881), (u'1601947', 0.006009303157071975), (u'2528778', 0.00598825511525346), (u'3337035', 0.005965529910613182), (u'77935', 0.0059642558217318144), (u'490113', 0.005962484236429215), (u'3862152', 0.005957370763910136), (u'1473256', 0.005952949411546176), (u'5181938', 0.005944243991134256), (u'2496661', 0.005927220178201412), (u'4346832', 0.005909425298294766), (u'1794072', 0.005906170268028396), (u'3242067', 0.005904174569753349), (u'4830148', 0.005901618951379806), (u'612607', 0.0059007505077467125), (u'3211540', 0.0058944423915780475)]

b)

[(u'5302153', 0.2980290830322092), (u'88822', 0.2745145816496703), (u'1921890', 0.2482397862804333), (u'84707', 0.21017421901657926), (u'4015997', 0.20046321716804083), (u'434174', 0.1778273799413172), (u'5115901', 0.17084667300068518), (u'1569176', 0.16854485839201244), (u'3013310', 0.1641218377082484), (u'2496936', 0.15324252436345157), (u'964099', 0.1495315508259344), (u'1201051', 0.14881617379651352), (u'4720054', 0.14742109886988553), (u'5303198', 0.14550040535717362), (u'79583', 0.14243862936745513), (u'1319777', 0.1399791267967069), (u'3240494', 0.1345671038058234), (u'3906702', 0.13415373410173997), (u'4031749', 0.1329618722798994), (u'4002307', 0.12350243429169601)]

2.1: See homework3.py

2.2: See homework3.py

2.3: See homework3.py

2.4: 130160392 and

2.5

a) AUTHS: [Row(page=u"United\_States'\_Country\_Reports\_on\_Human\_Rights\_Practices", auth=0.37843687670043263), Row(page=u'Geographic\_coordinates', auth=0.29735638171906054), Row(page=u'2008,\_Madrid\_plane\_crash', auth=0.2890848187118112), Row(page=u'2007,\_Canada\_budget', auth=0.26910488097800983), Row(page=u'Biography.com', auth=0.15610089669910465), Row(page=u'2006-04-01', auth=0.1477031658607502), Row(page=u"United\_Kingdom's", auth=0.14062666454972816), Row(page=u'Music\_genres', auth=0.13116609590214443), Row(page=u'Record\_label\_Hypnos', auth=0.1249404704440951), Row(page=u"France's\_National\_Archives", auth=0.12470731229221291), Row(page=u'2005,\_S\_1313', auth=0.12121296804192645), Row(page=u"England's", auth=0.11927402081350351), Row(page=u'Population\_dispersal', auth=0.10084140536037495), Row(page=u"Canada's", auth=0.1005820042996229), Row(page=u"Germany's\_Basic\_Law", auth=0.09625697104934057), Row(page=u'Internet\_Movie\_Databse', auth=0.09532433844181178), Row(page=u'Biological\_clock', auth=0.09110629551230656), Row(page=u'Time\_zone\_(disambiguation)', auth=0.090571343908421), Row(page=u'Area(Travel\_Agency)', auth=0.08834473402507846), Row(page=u'2004-05\_AHL\_season', auth=0.0831133110761372)]

>>> hubs\_results = sqlContext.createDataFrame(final\_hubs, ['page','hubs'])

b)

[(u'77935', 0.009848901978055179), (u'2496661', 0.00974483484765591), (u'79583', 0.009727889113675819), (u'84707', 0.00958871973804183), (u'3211540', 0.009510428985060758), (u'3000061', 0.009497314290398944), (u'81615', 0.009336508448418787), (u'76573', 0.009293679261479205), (u'65954', 0.009288383140475878), (u'4529923', 0.009158187518859206), (u'1331646', 0.008989923453599892), (u'3211271', 0.008975654608151772), (u'2638932', 0.008939934619409683), (u'75323', 0.008896146900461128), (u'1715261', 0.008893359847670326), (u'47849', 0.00888938260905526), (u'2496868', 0.008856839881086173), (u'65064', 0.008838156228705273), (u'3283893', 0.008823480702114821), (u'1715441', 0.008813295141250531)]

2.6:

a)

[(u'4504608', 0.006158049357462614), (u'1273793', 0.006103541463316293), (u'5296341', 0.00608430646587461), (u'2278916', 0.006069132899760532), (u'3102538', 0.0060357289399881), (u'1601947', 0.006009303157071975), (u'2528778', 0.00598825511525346), (u'3337035', 0.005965529910613182), (u'77935', 0.0059642558217318144), (u'490113', 0.005962484236429215), (u'3862152', 0.005957370763910136), (u'1473256', 0.005952949411546176), (u'5181938', 0.005944243991134256), (u'2496661', 0.005927220178201412), (u'4346832', 0.005909425298294766), (u'1794072', 0.005906170268028396), (u'3242067', 0.005904174569753349), (u'4830148', 0.005901618951379806), (u'612607', 0.0059007505077467125), (u'3211540', 0.0058944423915780475)]

b)

[(u'5302153', 0.2980290830322092), (u'88822', 0.2745145816496703), (u'1921890', 0.2482397862804333), (u'84707', 0.21017421901657926), (u'4015997', 0.20046321716804083), (u'434174', 0.1778273799413172), (u'5115901', 0.17084667300068518), (u'1569176', 0.16854485839201244), (u'3013310', 0.1641218377082484), (u'2496936', 0.15324252436345157), (u'964099', 0.1495315508259344), (u'1201051', 0.14881617379651352), (u'4720054', 0.14742109886988553), (u'5303198', 0.14550040535717362), (u'79583', 0.14243862936745513), (u'1319777', 0.1399791267967069), (u'3240494', 0.1345671038058234), (u'3906702', 0.13415373410173997), (u'4031749', 0.1329618722798994), (u'4002307', 0.12350243429169601)]