CSC 555 Social Computing

P1 Analytics Report

Part1: Hypothesis

Hypothesis:

Majority of the people belong to three social groups. These groups include school, university, and work.

Analysis:

The social circle of a user would be all the connections the user has. The network formed in such a manner would be including connections from the usual places where the user may be present at. Logically, any human spends time at school, university and work for most of their lives. So, the hypothesis therefore makes some sense initially.

After looking at the data we can see that universities haven't been specifically mentioned in the data sets. So, for analyzing here, I just took values of school and work ids, to see if the hypothesis still holds true. This seemed logical since even universities are considered as schools. So, the work was done by checking if the social circle of one user has most of its members from the same school or the same workplace.

For a specific egonet, we check if from every connection, you have maximum number of connections having a school id and a work id is present. This will make sure that the majority of the user social circle will be made of people having school and work mentioned in their profile. These users should constitute a lion's share of the user social network. We then get the data of the people sharing their school and their workplace from there.

I worked on numerous egonets to see if the hypothesis holds true, and it did. The output looked like this:

```
**** ------ Hypothesis 1 observation ------- ****
Total number of nodes in egonet 0: 239
Total number of nodes with same school_id or workId as same as egonet 0 : 168
Percentage of nodes in same social network: 70.29288702928869


Total number of nodes in egonet 850: 249
Total number of nodes with same school_id or workId as same as egonet 850 : 176
Percentage of nodes in same social network: 70.68273092369478


Total number of nodes in egonet 2255: 110
Total number of nodes with same school_id or workId as same as egonet 2255 : 79
Percentage of nodes in same social network: 71.81818181818181
```

As we can see, egonet 0 has about 70% nodes with same school or workplace as the user in question, egonet 850 has about 71% nodes with same school or workplace as the user in question, and egonet 2255 has about 72% nodes with same school or workplace as the user in question. The hypothesis holds true over the network.

Part2: For an egonet, show specific values for it, which include:

Egonet used is 0

Graph statistics

- Number of Nodes
  Number of nodes for egonet 0 are 239

- Number of Edges
  Number of edges for egonet 0 are 4448

Sociometric network properties

- Betweenness Centrality (of ego nodes)

  Betweenness Centrality: {0: 0.6585235979076569, 1: 0.0007683021532231889, 146: 0.004112139518594241, 189: 0.004483061587532435, 229: 0.00013711113932016942, 201: 0.0005186733769216206, 204: 0.00030086008310240675, 60: 0.0003756554498805189, 215: 0.0013971217264378572, 35: 0.00036056226814815004, 91: 0.0008374147014976589, 238: 2.0361430705443873e-06, 88: 0.0007505271581390334, 28: 0.00037837234037174406, 166: 0.0001866414507160146, 218: 0.0005501933439077681, 156: 0.0005744553674709442, 6:

0.0006020801048628486, 2: 0.00039276171805999795, 8: 0.005446978382353888, 231: 0.0012647793943072892, 165: 0.0005580048132291283, 184: 0.0006474621479658505, 20: 0.0004710421440527905, 221: 0.000957470074687156, 128: 0.0007288876332837805, 39: 0.0005181484032937139, 48: 0.0008111555072275328, 109: 0.0002982633601413987, 84: 0.0010518369930759698, 15: 0.001774838141605593, 139: 9.14600874445399e-05, 21: 0.0004027836162349535 6, 108: 0.00020379233440384742, 207: 0.0005571857509333241, 161: 0.00121369667829862, 147: 0.0002796585320066206, 155: 0.0007514477074551579, 140: 0.0033558543231437346, 41: 9.710863786484747e-05, 12: 0.0017713950660038712, 160: 0.0006313070587128418, 127: 0.0006894057406197789, 227: 0.0011593231049714832, 47: 0.000566118442177089, 65: 0.0007695507788235215, 105: 0.0004347186918483436, 62: 0.00032719194232255067, 182: 0.0003478557037950357, 79: 0.00027466281428990694, 90: 0.000527491084922899, 100: 0.00038523262583550186, 148: 0.0006012534862466249, 137: 0.007208716429851828, 170: 0.0003692196480081889, 158: 0.0004192420999777948, 58: 0.02372843243398391, 134: 0.0004087517823154403, 179: 0.00012009506281738887, 123: 0.0002403974832582978 5, 167: 0.0016313817376146513, 45: 0.002074856016587298, 51: 0.0008188681849897085, 85: 0.0076197405869284, 54: 9.782403994755761e-05, 75: 0.0002863213029707258, 191: 0.00032295775508590273, 95: 0.0006467454283852879, 141: 0.0003966938760425534, 235: 0.00010244105239103284, 46: 0.00018965439651700696, 210: 0.0001430351210282844 4, 126: 0.00038896857177602736, 129: 0.0001567906413739389, 205: 0.004192826230060243, 76: 0.0006880953392226528, 97: 0.0005115115928778933, 3: 0.00787560840080064, 185: 0.0005279442046543641, 80: 0.0004570148967696635, 61: 0.001097047096959957, 188: 0.0002480244214902454, 22: 0.0005179063555568404, 222: 0.0020118947539577853, 118: 0.0005947808051307558, 4: 0.0007240883417893049, 223: 0.0004108988793409527 4, 138: 0.0002755514277650863, 52: 0.0010066997293744675, 212: 0.0019655388296526614, 213: 0.00036302995446890084, 226: 0.0005931477918163065, 117: 0.0001213913377710801 1, 103: 6.163452073461646e-05, 18: 8.889109475150649e-05, 154: 0.0001081531288613580 5, 43: 4.085718797000871e-05, 176: 5.765683687095226e-05, 181: 0.0002708053033161599 4, 33: 0.0020672458002294713, 59: 0.0008668992836647834, 17: 0.0011264969439368133, 44: 0.0012077377752240997, 89: 0.001193812807261505, 36: 0.0005650938281332092, 232: 0.0008773975639526387, 224: 0.000303097834073234, 114: 0.0014146167589212783, 145: 0.0002767426651103541, 228: 0.0001974565476463745 5, 66: 0.0015512063345830235, 199: 0.0011389071157796033, 34: 0.000878053666571414, 209: 0.0041023624854874 05, 101: 0.0004139163756256937, 78: 0.0004629729909630518, 225: 0.0005321628370203 6, 190: 0.001453521502764407, 112: 0.0011007454015660358, 169: 0.0019438266898629737, 113: 0.0008641084417597554, 11: 0.0006514092063327243, 157: 0.0003361314933621364, 120: 0.00020854655301782193, 72: 3.351339037754112e-05, 187: 0.000754245047437874 9, 163: 0.0014518934638548503, 177: 0.0005184473370544953, 106: 0.0001689488361413867, 16: 0.0003143873583188549, 92: 3.1785297391574516e-05, 42: 0.0091694191381436 79, 174: 0.0002181605915105628 7, 5: 2.9547684052523962e-05, 111: 4.84582018461393e-05, 71: 0.0035803212777502113, 132: 0.0003353140010110375, 31: 0.0004778614893324103 6, 81: 0.000305171178427277, 230: 0.0001720018572934634 8, 162: 0.0022406196227445 54, 37: 0.0004871692643220724, 7: 0.000765654156094472 4, 216: 0.0001980853564227047 4, 142: 0.0005119769426758044, 104: 8.391542270916806e-05, 194: 0.0005027641899814447, 233: 0.0007550001054446765, 217: 0.0, 203:

0.0035843075092654303, 219: 0.00020588826247798704, 197: 0.00046547583847171087, 195: 0.00030154446158375646, 67: 0.00024642768499804984, 23: 0.0005134013562403533, 121: 0.0008011643475955784, 180: 0.00037515912260356805, 25: 0.0003532167025579547, 13: 0.0009073680839057325, 122: 0.00039788644342045863, 40: 0.0009181554759786414, 234: 9.573449633017764e-05, 74: 8.096065430391565e-05, 116: 3.3684359819877315e-05, 9: 6.020671992068806e-06, 193: 4.963565218858307e-05, 198: 0.0001474990394092814, 119: 0.0018427004194640282, 200: 3.51037622518304e-05, 10: 5.112413100924961e-05, 192: 4.136675767353354e-05, 236: 0.00012334559191276562, 151: 0.0016481778451048405, 133: 0.00014189231827503089, 168: 0.0002960978056114447, 115: 0.00011366750894778849, 153: 2.1021066654509905e-05, 38: 0.00020278576340797104, 164: 0.00011430384520893898, 30: 0.0, 87: 0.00012528218038270158, 86: 0.00038382353029730963, 110: 0.0008192958136693126, 206: 0.0015041538864651947, 98: 7.597975899220448e-05, 125: 0.0004205479660218517, 14: 0.00012182088025083449, 211: 0.00018235142158129069, 96: 0.000352841558564354, 77: 0.0002163312582416933, 175: 9.680512020688595e-05, 135: 0.0, 69: 0.0013201905234667707, 183: 0.00017032598318277145, 171: 0.00040421231783852783, 50: 0.0, 186: 5.682246933177685e-06, 152: 1.8608546956595035e-05, 237: 0.00024024336300251876, 159: 0.0005514897334491104, 220: 1.5107659684477754e-05, 150: 1.8661695191067764e-06, 99: 4.407335593008497e-05, 19: 0.00011050833835643962, 144: 0.0, 214: 0.00020055302824364626, 178: 0.00021983476935077828, 68: 0.0001252821803827016, 24: 8.459079834465432e-05, 149: 1.7728610431514378e-05, 26: 0.0, 27: 3.841198926828115e-05, 93: 0.00015647022057226222, 29: 0.0, 32: 0.0, 124: 0.00047402927415692015, 57: 0.0008505441833354212, 143: 0.00014687990245684026, 55: 2.0251388528359228e-05, 49: 2.068337883676677e-05, 53: 0.0, 56: 0.00042075902090794116, 64: 3.3684359819877315e-05, 63: 5.082201657034122e-05, 82: 0.00010046212577858147, 70: 0.0, 94: 6.205013651030033e-05, 131: 3.545722086302875e-05, 83: 0.0, 172: 6.205013651030033e-05, 73: 1.7728610431514378e-05, 136: 0.0, 202: 0.0, 102: 0.0, 107: 0.0, 173: 0.0, 130: 0.0, 208: 0.0, 196: 0.0}

- Clustering coefficient

  Clustering value: 0.7391599143006116

  Output window

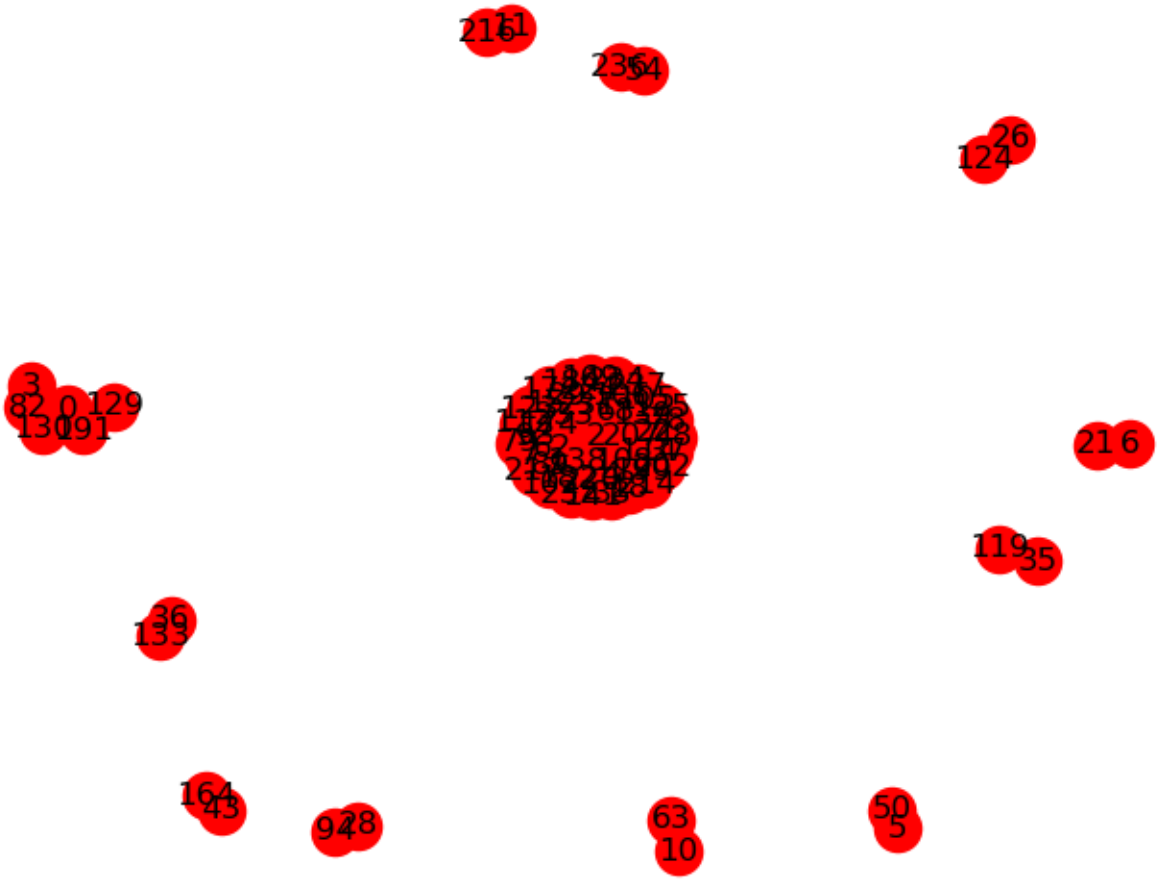
```
Number of nodes: 239
Number of edges: 4448
Betweenness Centrality: {0: 0.6585235979076569, 1: 0.0007683021532231889, 146: 0.004112139518594241, 189: 0.004483061587532435, 229: 0.00013711113932016942, 201:
0.0005186733769216206, 204: 0.00030086008310240675, 60: 0.0003756554498805189, 215: 0.0013971217264378572, 35: 0.00036056226814815004, 91: 0.0008374147014976589, 238:
2.0361430705443873e-06, 88: 0.0007505271581390334, 28: 0.00037837234037174406, 166: 0.0001866414507160146, 218: 0.0005501933439077681, 156: 0.0005744553674709442, 6:
0.0006020801048628486, 2: 0.00039276171805999795, 8: 0.005446978382353888, 231: 0.0012647793943072892, 165: 0.0005580048132291283, 184: 0.0006474621479658505, 20:
0.0004710421440527905, 221: 0.000957470074687156, 128: 0.0007288876332837805, 39: 0.0005181484032937139, 48: 0.0008111555072275328, 109: 0.0002982633601413987, 84:
0.0010518369930759698, 15: 0.001774838141605593, 139: 9.14600874445399e-05, 21: 0.00040278361623495356, 108: 0.00020379233440384742, 207: 0.0005571857509333241, 161:
0.00121369667829862, 147: 0.0002796585320066206, 155: 0.0007514477074551579, 140: 0.0033558543231437346, 41: 9.710863786484747e-05, 12: 0.0017713950660038712, 160:
0.0006313070587128418, 127: 0.0006894057406197789, 227: 0.0011593231049714832, 47: 0.000566118442177089, 65: 0.0007695507788235215, 105: 0.0004347186918483436, 62:
0.0003271919423225506 7, 182: 0.0003478557037950357, 79: 0.00027466281428990694, 90: 0.000527491084922899, 100: 0.0003852326258355018 6, 148: 0.0006012534862466249, 137:
0.007208716429851828, 170: 0.0003692196480081889, 158: 0.0004192420999777948, 58: 0.02372843243398391, 134: 0.0004087517823154403, 179: 0.0001200950628173888 7, 123:
0.0002403974832582978 5, 167: 0.0016313817376146513, 45: 0.002074856016587298, 51: 0.0008188681849897085, 85: 0.0076197405869284, 54: 9.782403994755761e-05, 75: 0.0002863213029707258,
191: 0.0003229577550859027 3, 95: 0.0006467454283852879, 141: 0.0003966938760425534, 235: 0.0001024410523910328 4, 46: 0.0001896543965170069 6, 210: 0.0001430351210282844 4, 126:
0.0003889685717760273 6, 129: 0.0001567906413739389, 205: 0.004192826230060243, 76: 0.0006880953392226528, 97: 0.0005115115928778933, 3: 0.007875608400800 64, 185:
0.0005279442046543641, 80: 0.0004570148967696635, 61: 0.001097047096959957, 188: 0.0002480244214902454, 22: 0.0005179063555568404, 222: 0.0020118947539577853, 118:
0.0005947808051307558, 4: 0.0007240883417893049, 223: 0.0004108988793409527 4, 138: 0.0002755514277650863, 52: 0.0010066997293744675, 212: 0.0019655388296526614, 213:
0.0003630299544689008 4, 226: 0.0005931477918163065, 117: 0.0001213913377710801 1, 103: 6.163452073461646e-05, 18: 8.889109475150649e-05, 154: 0.0001081531288613580 5, 43:
4.085718797000871e-05, 176: 5.765683687095226e-05, 181: 0.0002708053033161599 4, 33: 0.0020672458002294713, 59: 0.0008668992836647834, 17: 0.0011264969439368133, 44:
0.0012077377752240997, 89: 0.001193812807261505, 36: 0.0005650938281332092, 232: 0.0008773975639526387, 224: 0.0003030978340732 34, 114: 0.0014146167589212783, 145:
0.0002767426651103541, 228: 0.0001974565476463745 5, 66: 0.0015512063345830235, 199: 0.0011389071157796033, 34: 0.000878053666571414, 209: 0.0041023624854874 05, 101:
0.0004139163756256937, 78: 0.0004629729909630518, 225: 0.0005321628370203 6, 190: 0.0014535215027644 07, 112: 0.0011007454015660358, 169: 0.0019438266898629737, 113:
```

0.0008641084417597554, 11: 0.0006514092063327243, 157: 0.0003361314933621364, 120: 0.0002085465530178219 3, 72: 3.351339037754112e-05, 187: 0.0007542450474378749, 163: 0.0014518934638548503, 177: 0.0005184473370544953, 106: 0.0001689488361413867, 16: 0.0003143873583188549, 92: 3.1785297391574516e-05, 42: 0.009169419138143679, 174: 0.0002181605915105628 7, 5: 2.9547684052523962e-05, 111: 4.84582018461393e-05, 71: 0.0035803212777502113, 132: 0.0003353140010110375, 31: 0.0004778614893324103 6, 81: 0.0003051711784272 77, 230: 0.0001720018572934 6348, 162: 0.0022406196227445 54, 37: 0.0004871692643220724, 7: 0.0007656541560944724, 216: 0.0001980853564227047 4, 142: 0.0005119769426758044, 104: 8.39154227091680 6e-05, 194: 0.0005027641899814447, 233: 0.0007550001054446765, 217: 0.0, 203: 0.0035843075092654303, 219: 0.0002058882624779870 4, 197: 0.0004654758384717108 7, 195: 0.0003015444615837564 6, 67: 0.0002464276849980498 4, 23: 0.0005134013562403533, 121: 0.0008011643475955784, 180: 0.0003715912260356805, 25: 0.0003532167025579547, 13: 0.0009073680839057325, 122: 0.0003978864434204586 3, 40: 0.0009181554759786414, 234: 9.573449633017764e-05, 74: 8.0960654303915 65e-05, 116: 3.3684359819877315e-05, 9: 6.020671992068806e-06, 193: 4.963565218858307e-05, 198: 0.0001474990394092814, 119: 0.0018427004194640282, 200: 3.5103762251830 4e-05, 10: 5.112413100924961e-05, 192: 4.136675767353354e-05, 236: 0.0001233455919127656 2, 151: 0.0016481778451048405, 133: 0.0001418923182750308 9, 168: 0.0002960978056114447, 115: 0.0001136675089477884 9, 153: 2.1021066654509905e-05, 38: 0.0002027857634079710 4, 164: 0.0001143038452089389 8, 30: 0.0, 87: 0.0001252821803827015 8, 86: 0.0003838235302973096 3, 110: 0.0008192958136693126, 206: 0.0015041538864651947, 98: 7.597975899220448e-05, 125: 0.0004205479660218517, 14: 0.0001218208802508344 9, 211: 0.0001823514215812906 9, 96: 0.000352841558564354, 77: 0.0002163312582416933, 175: 9.680512020688595e-05, 135: 0.0, 69: 0.0013201905234667707, 183: 0.0001703259831827714 5, 171: 0.0004042123178385278 3, 50: 0.0, 186: 5.682246933177685e-06, 152: 1.8608546956595035e-05, 237: 0.0002402433630025187 6, 159: 0.0005514897334491104, 220: 1.5107659684477754e-05, 150: 1.8661695191067764e-06, 99: 4.407335593008497e-05, 19: 0.0001105083383564396 2, 144: 0.0, 214: 0.0002005302824364626, 178: 0.0002198347693507782 8, 68: 0.0001252821803827016, 24: 8.459079834465432e-05, 149: 1.7728610431514378e-05, 26: 0.0, 27: 3.841198926828115e-05, 93: 0.0001564702205722622 2, 29: 0.0, 32: 0.0, 124: 0.0004740292741569201 5, 57: 0.000850544183335421 2, 143: 0.0001468799024568402 6, 55: 2.0251388528359228e-05, 49: 2.068337883676677e-05, 53: 0.0, 56: 0.0004207590209079411 6, 64: 3.3684359819877315e-05, 63: 5.082201657034122e-05, 82: 0.0001004621257785814 7, 70: 0.0, 94: 6.205013651030033e-05, 131: 3.545722086302875e-05, 83: 0.0, 172: 6.205013651030033e-05, 73: 1.7728610431514378e-05, 136: 0.0, 202: 0.0, 102: 0.0, 107: 0.0, 173: 0.0, 130: 0.0, 208: 0.0, 196: 0.0}
Clustering value: 0.7391599143006116

Part 3: Social Circles

The approach for creating a social circle required a parameter to be chosen for a proper graph to be shown. The parameter I chose was languages. I got a social circle from the data over egonet 0, to see the trend the data shows.

While there was a common language amongst most people which resulted in the biggest cluster in the graph, there were a few other clusters with multiple other languages too. The graph looked a bit like this:



While we do have outliers, the languages spoken are so many that we get a number of clusters from the whole data.