

Blogosphere

Communities

Bonus

# Homeplay for week 02's lab: code

## Blogosphere

Let's look at the blogosphere data we have now seen a few times. You haven't actually touched the data, so we are going to do that now.

1. Load the `blogosphere` network from the `SNA4DSData` package.
2. Clean it up a bit: there are some blogs that refer to themselves ("loops"). That is not what we want, so remove these loops. How do I do that, you ask? Well, search through the documentation for `igraph`. Alternatively, see if you can find the solution in the course cheatsheet. From here on, continue with this cleaned graph.

```
blog1 <- igraph::simplify(blogosphere)
```

```
## This graph was created by an old(er) igraph version.  
## i Call `igraph::upgrade_graph()` on it to use with the current igraph version.  
## For now we convert it on the fly...
```

1. Calculate summary statistics for this network. At least determine:
  - mean distance
  - diameter
  - dyad census
  - reciprocity
  - transitivity
  - density

Interpret these numbers, do you understand what they mean? Are they what you would expect?

```
# mean distance  
snafun::g_mean_distance(blog1)
```

```
## Registered S3 methods overwritten by 'snafun':  
##   method      from  
##   plot.igraph igraph  
##   print.igraph igraph
```

```
## [1] 3.390184
```

```
# diameter
snafun::g_diameter(blog1)
```

```
## [1] 9
```

```
# dyad census
snafun::count_dyads(blog1)
```

```
## Mutual Asymmetric Null
##      2307      14408 1092590
```

```
# reciprocity
snafun::g_reciprocity(blog1)
```

```
## [1] 0.2425612
```

```
# transitivity
snafun::g_transitivity(blog1)
```

```
## [1] 0.2395381
```

```
# density
snafun::g_density(blog1)
```

```
## [1] 0.008573837
```

## Communities

I already told you that the network is quite polarized: republican blogs see to refer mainly to other republican blogs, and the same holds for democratic blogs. So, you would expect some clear subgroups in this network.

1. Use the walktrap algorithm to determine communities in the blogosphere network. How many communities do you get? Does the partitioning make sense? What do you notice about the result? (hint: use the cheatsheet to see how you can combine `snafun` and `igraph` here)

```
walk <- snafun::extract_comm_walktrap(blog1)
igraph::membership(walk) |> table()
```

```

##
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
## 560  4 640  2  2  4  2  2  2  2  2  2  1  1  1  1  1  1  1  1
## 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1

```

```
igraph::sizes(walk)
```

```
## Community sizes
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
## 560 4 640 2 2 4 2 2 2 2 2 2 1 1 1 1 1 1 1 1
## 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
igraph::sizes(walk) |> length()
```

```
## [1] 278
```

```
igraph::modularity(walk)
```

```
## [1] 0.4302334
```

1. One thing that might affect the result is that a bunch of the blogs do not point to another blog and no other blog points to them. In SNA we call these nodes *isolates*.
2. What do you think would happen to the results if you would remove those isolates first? First, think about this. Then actually remove the isolates from the graph and run the walktrap again. Check if you were correct.

```
blog0 <- snafun::remove_isolates(blog1)
walk0 <- snafun::extract_comm_walktrap(blog0)
igraph::membership(walk0) |> table()
```

```
##
##   1   2   3   4   5   6   7   8   9  10  11  12
## 560   4 640   2   2   4   2   2   2   2   2   2
```

```
igraph::modularity(walk0)
```

```
## [1] 0.4302334
```

```
igraph::sizes(walk0) |> length()
```

```
## [1] 12
```

## Bonus

Continue with this smaller network. You should have 1224 vertices left and 19022 edges between them.

1. Add the group membership (which you get with `igraph::membership`) to the network as a vertex attribute. Call this new vertex attribute `community`.
2. Run the following function: `snafun::make_mixingmatrix(graph_without_isolates, "community")` and check the help documentation of `snafun::make_mixingmatrix` to interpret the result. Does it make sense? What does this tell you about the boundaries around the communities? (obviously, replace `graph_without_isolates` with the actual object name you are using).
3. Extract all of the vertex attributes using `snafun::extract_all_vertex_attributes` and explore whether it is indeed the case that like-minded blogs tend cluster together inside the communities you found. (hint: the simplest approach is to use a table)

```
blog0 <- snafun::add_vertex_attributes(blog0, "community", value = igraph::membership(walk0))
# alternatively
# igraph::V(blog0)$community <- igraph::membership(walk0)

# mixing matrix
snafun::make_mixingmatrix(blog0, "community")
```

```
##      to
## from    1    2    3    4    5    6    7    8    9   10   11   12
## 1    8339    1   633    0    1    0    0    0    0    2    1    0
## 2      2    3    3    0    0    0    0    0    0    0    0    0
## 3    675    0  9338    0    0    1    0    1    1    0    0    0
## 4      1    0    1    1    0    0    0    0    0    0    0    0
## 5      1    0    0    0    1    0    0    0    0    0    0    0
## 6      0    0    0    0    0    5    0    0    0    0    0    0
## 7      1    0    0    0    0    0    1    0    0    0    0    0
## 8      0    0    0    0    0    0    0    1    0    0    0    0
## 9      0    0    0    0    0    0    0    0    1    0    0    0
## 10     2    0    0    0    0    0    0    0    0    2    0    0
## 11     1    0    0    0    0    0    0    0    0    0    1    0
## 12     0    0    0    0    0    0    0    0    0    0    0    1
## Sum 9022    4  9975    1    2    6    1    2    2    4    2    1
##      to
## from    Sum
## 1    8977
## 2      8
## 3   10016
## 4      3
## 5      2
## 6      5
## 7      2
## 8      1
## 9      1
## 10     4
## 11     2
## 12     1
## Sum 19022
```

```
# alternative, outside of the graph object
all_attr <- snafun::extract_all_vertex_attributes(blog0)
table(all_attr$community, all_attr$party)
```

```
##
##      conservative liberal
## 1          23      537
## 2           1        3
## 3         604      36
## 4           2        0
## 5           0        2
## 6           4        0
## 7           0        2
## 8           0        2
## 9           2        0
## 10          0        2
## 11          0        2
## 12          0        2
```

Overall, you shouldn't be surprised about the strong political divide in the USA at the moment. This was already apparent in the 2004 elections... You just showed that yourself!

