Network analysis; an introduction (with igraph in R)

Ali (Aliakbar Akbaritabar)

Email: Akbaritabar@gmail.com

21/11/2019 - Humboldt University of Berlin

Outline of brief introduction to Network Analysis

1. What is relational view and network analysis?

Outline of brief introduction to Network Analysis

1. What is relational view and network analysis?
2. Ethnography of network ties! Context of interactions

Outline of brief introduction to Network Analysis

1. What is relational view and network analysis?
2. Ethnography of network ties! Context of interactions
3. How to gather and use network data?

Outline of brief introduction to Network Analysis

1. What is relational view and network analysis?
2. Ethnography of network ties! Context of interactions
3. How to gather and use network data?
4. Possible questions to ask!

Outline of brief introduction to Network Analysis

1. What is relational view and network analysis?
2. Ethnography of network ties! Context of interactions
3. How to gather and use network data?
4. Possible questions to ask!
5. A real life example from science studies!

Outline of brief introduction to Network Analysis

1. What is relational view and network analysis?
2. Ethnography of network ties! Context of interactions
3. How to gather and use network data?
4. Possible questions to ask!
5. A real life example from science studies!
6. Where to next?!

What is this?!

Now what?!

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

A poll/survey results?

Variables/attributes

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

A variable by observation table

Variables/attributes

Respondents/
Observations

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

What if respondents know each other?!

Variables/attributes

Respondents/
Observations

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

Different contexts of familiarity

- Family, college, gym, . . .

Variables/attributes

Respondents/
Observations

ID	Name	Age	Political view	Education
1	Tom	24	left	NA
2	Sara	22	right	BA
3	Bill	30	neutral	MA
4	Margaret	31	NA	PhD

Stories behind ties!

- Independence of observations? (in many cases it is violated!)

Respondents/ Observations (columns); Respondents/ Observations (rows)

	Tom	Sara	Bill	Margaret
Tom	-	Instagram celebrity	Is not sure remembers	NA
Sara	NA	-	Brother	Brother's crush
Bill	(Former) same gym member	Sister	-	MA classmate / friend / smartest batchmate
Margaret	NA	NA	MA classmate	-

Adjacency (familiarity) matrix

Respondents/
Observations

	Tom	Sara	Bill	Margaret
Tom	-	1	0	0
Sara	0	-	1	1
Bill	1	1	-	1
Margaret	0	0	1	-

Respondents/
Observations

Read Edge List as CSV (to construct a network)

```
edge_list2_use <- read_csv("../1_data/humans_ties.csv")
kable(edge_list2_use)
```

source	target	weight	label
Tom	Sara	0.5	Acquaintance
Sara	Bill	1.0	Sibling
Sara	Margaret	0.5	Acquaintance
Bill	Tom	0.5	Acquaintance
Bill	Sara	1.0	Sibling
Bill	Margaret	1.0	Friend
Margaret	Bill	0.5	Acquaintance

Convert it to a (network) graph object[1]

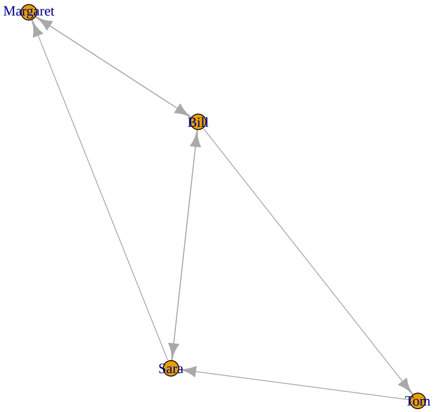
```
gg = graph_from_data_frame(d = edge_list2_use, directed = TRUE)

print(gg)
```

```
## IGRAPH 1464730 DNW- 4 7 --
## + attr: name (v/c), weight (e/n), label (e/c)
## + edges from 1464730 (vertex names):
## [1] Tom     ->Sara     Sara    ->Bill     Sara    ->Margaret Bill    ->Tom
## [5] Bill    ->Sara     Bill    ->Margaret Margaret->Bill
```

[1] Python users, check script “09_example_network_igraph_python.py” in code directory

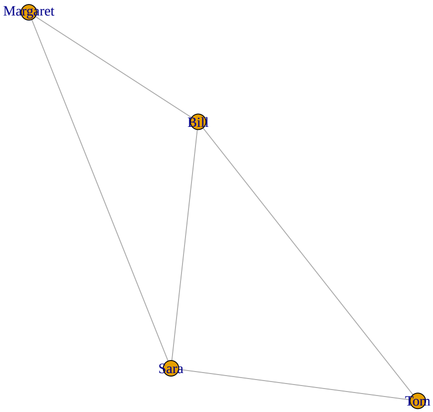
Plot the graph with a layout (directed)

```
set.seed(2225235)
gg_layout = layout.fruchterman.reingold(graph = gg)
plot(gg, layout = gg_layout, edge.label = NA, vertex.size=8)
```



Plot the graph with a layout (un-directed)

```
gg_undirected = graph_from_data_frame(d = edge_list2_use, directed = F)
gg_undirected = simplify(graph = gg_undirected, remove.multiple = T)
plot(gg_undirected, layout = gg_layout, edge.label = NA, vertex.size=8)
```



Add a new attribute to nodes?

```
print(V(gg))
```

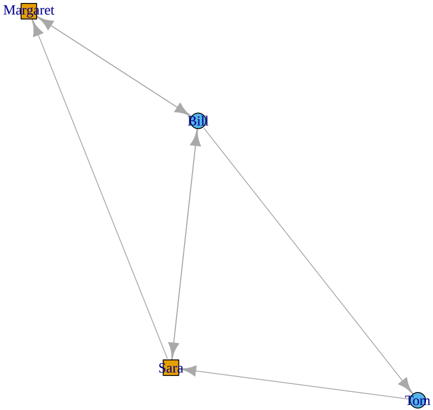
```
## + 4/4 vertices, named, from 1464730:
## [1] Tom      Sara     Bill     Margaret
```

```
V(gg)$gender <- c('male', 'female', 'male', 'female')
V(gg)$shape <- c('circle', 'square', 'circle', 'square')
print(gg)
```

```
## IGRAPH 1464730 DNW- 4 7 --
## + attr: name (v/c), gender (v/c), shape (v/c), weight (e/n), label
## | (e/c)
## + edges from 1464730 (vertex names):
## [1] Tom     ->Sara     Sara    ->Bill     Sara    ->Margaret Bill    ->Tom
## [5] Bill    ->Sara     Bill    ->Margaret Margaret->Bill
```

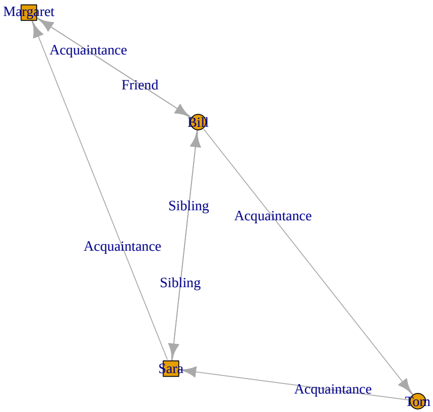
Color and shape of nodes based on gender

```
plot(gg, edge.label = NA, vertex.color = factor(V(gg)$gender),
     vertex.shape = V(gg)$shape, layout = gg_layout, vertex.size=8)
```



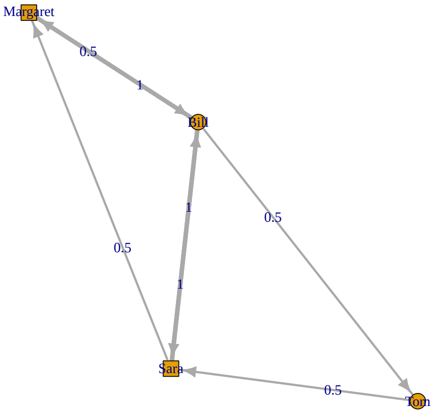
Name ties based on types

```
plot(gg, edge.label = E(gg)$label, layout = gg_layout, vertex.size=8)
```



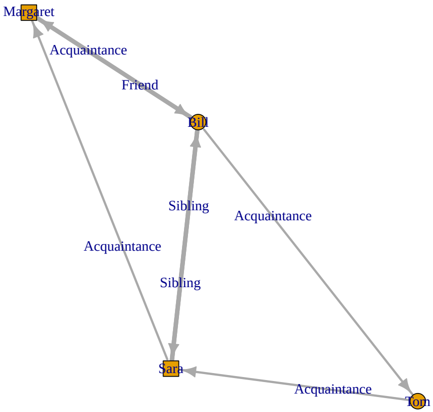
Weight ties based on importance

```
plot(gg, edge.width = E(gg)$weight*5, edge.label = E(gg)$weight, layout = gg_layout, vertex.size=8)
```



Mixture of weight/label

```
plot(gg, edge.label = E(gg)$label, edge.width = E(gg)$weight*5, layout = gg_layout, vertex.size=8)
```

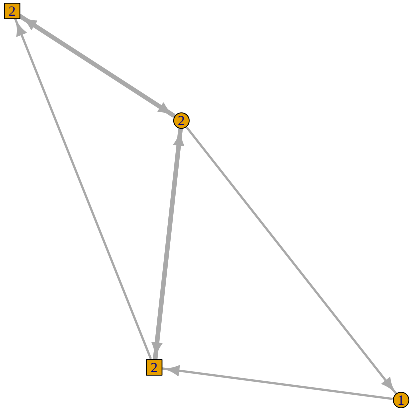


A glimpse to more serious analysis

- After simple visualization (if possible), a five number summary!
  1. **Size**: V, E (N of vertices/nodes and ties/edges, respectively)
  2. **Density** (ratio of ties to possible ties, $1 =$ fully connected)
  3. **Components** & (dis)connectivity (more connection inside groups, less among them)
  4. **Diameter** (how compact the network is?)
  5. **Clustering Coefficient** (transitivity and triangles)
- **Centrality** in network (different measures of importance in structure)
  - Degree, Closeness, Betweenness, Eigenvector, ...

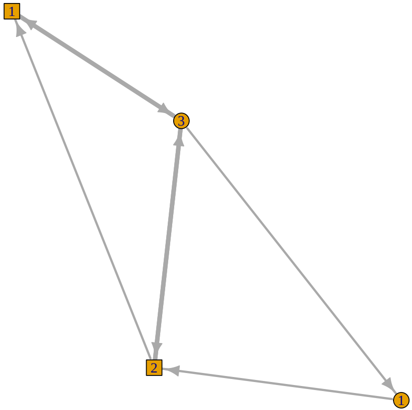
In-degree of a node (incoming ties)

```
plot(gg, edge.label = NA, edge.width = E(gg)$weight*5,
     vertex.label = degree(gg, mode = 'in'), layout = gg_layout, vertex.size=8)
```



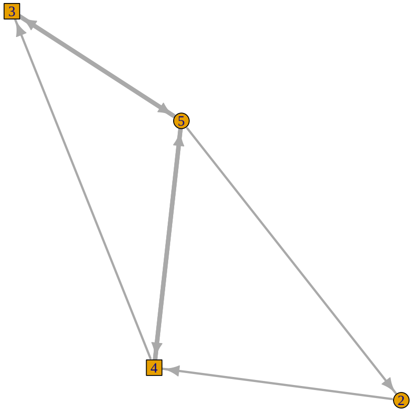
Out-degree of a node (outgoing ties)

```
plot(gg, edge.label = NA, edge.width = E(gg)$weight*5,
     vertex.label = degree(gg, mode = 'out'), layout = gg_layout, vertex.size=8)
```



Degree of a node (both incoming/outgoing ties)

```
plot(gg, edge.label = NA, edge.width = E(gg)$weight*5,
     vertex.label = degree(gg, mode = 'all'), layout = gg_layout, vertex.size=8)
```

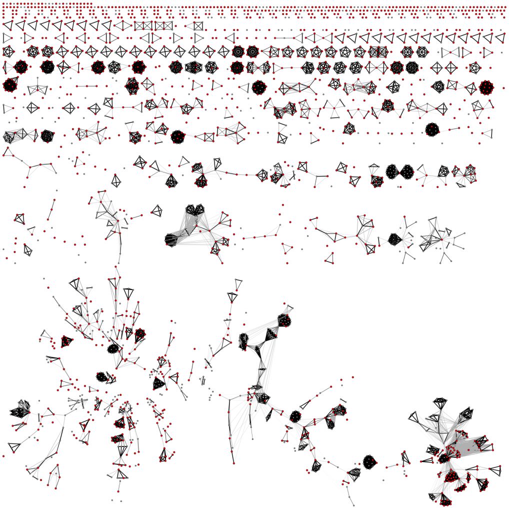


Sociological theories (& SNA conceptualization)[2]

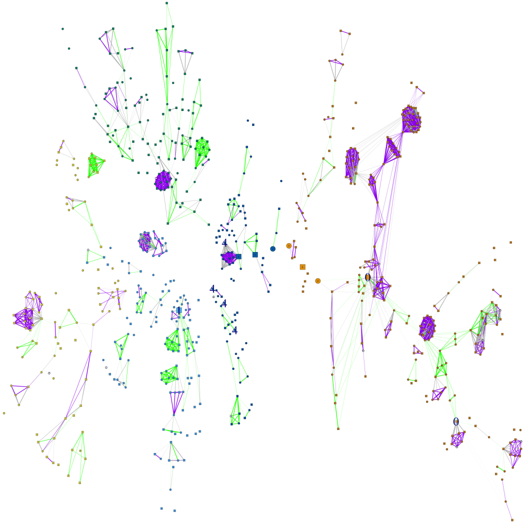
- **Matteo effect**, winner takes all?
  - Highly prolific scientists attract higher collaborations from other scientists?
  - Attaching preferably to a few **star scientists**/leaders?
- **Fragmentation** of ideas, sociology as a interstitial science?
  - Methodologists bridging the islands?
- [Sociological] **small world** of disconnected islands?
- **Core** of leaders and **periphery** of followers?

[2] Akbaritabar, A., Traag, V. A., Caimo, A., & Squazzoni, F. (2020). Italian Sociologists: A Community of Disconnected Groups. Scientometrics. https://doi.org/10.1007/s11192-020-03555-w

Coauthorship of Italian sociologists



Communities in the giant component



What can we learn from these communities? (1/2)[3]

Table 2: Gender composition and internationality of members of the communities detected from the giant component (Percentages are calculated by rows separately for gender and country)

		Gender			Country			
Community	# members	Female	Male	Missing Gender	Europe	Italy	Other	Missing Country
0	254	43%	54%	3%	54%	29%	11%	5%
1	142	50%	49%	1%	36%	55%	6%	3%
2	122	38%	61%	1%	37%	56%	3%	4%
3	103	45%	54%	1%	41%	44%	5%	11%
4	91	47%	49%	3%	32%	57%	9%	2%

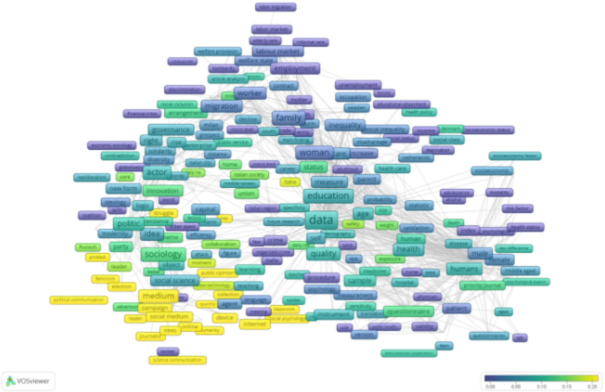
Table 3: Sectors composition of members of the communities detected from the giant component (Percentages are calculated by rows)

		Scientific Disciplinary Sectors						
Community	# members	postdoc	SPS/07	SPS/08	SPS/09	SPS/10	SPS/11	Missing Sector
0	254	2%	1%	5%	0	0%	0%	91%
1	142	2%	6%	3%	8%	1%	1%	78%
2	122	5%	10%	1%	7%	0	1%	76%
3	103	2%	4%	2%	12%	1%	0	80%
4	91	1%	7%	7%	0	1%	2%	82%

[3]Akbaritabar, A., Traag, V. A., Caimo, A., & Squazzoni, F. (2020). Italian Sociologists: A Community of Disconnected Groups. Scientometrics. https://doi.org/10.1007/s11192-020-03555-w

What can we learn from these communities? (2/2)[4]

- 65% foreigners
- Medium, science communication, social medium, internet, political communication & public opinion



[4] Akbaritabar, A., Traag, V. A., Caimo, A., & Squazzoni, F. (2020). Italian Sociologists: A Community of Disconnected Groups. Scientometrics. https://doi.org/10.1007/s11192-020-03555-w

Where to next?!

- **Awesome network analysis** list: https://github.com/briatte/awesome-network-analysis