seq2net: a Python library for converting behavior data to social networks

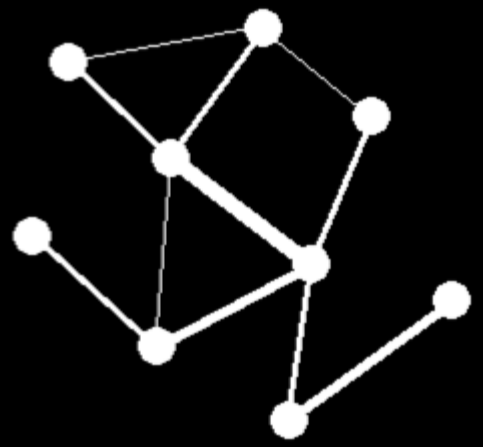
Alexander M. Procton
Programming and Data Science for Biologists
Columbia University
April 23, 2018

The problem

- Social network data uses unique formats
- Behavior data is recorded sequentially
- No package to convert sequential data to network format

subj	beeps	activity	partner
ID1	4	gmd/ssitco	ID14, ID20
ID1	3	sitco	ID20
ID2	13	R	NA
ID2	3	gmd	ID8
ID2	4	gm	ID8

	ID1	ID2	ID3	ID4	…
ID1	0	2	5	12	…
ID2	0	0	10	0	…
ID3	4	7	0	5	…
ID4	9	3	1	0	…
…	…	…	…	…	…



Package goals

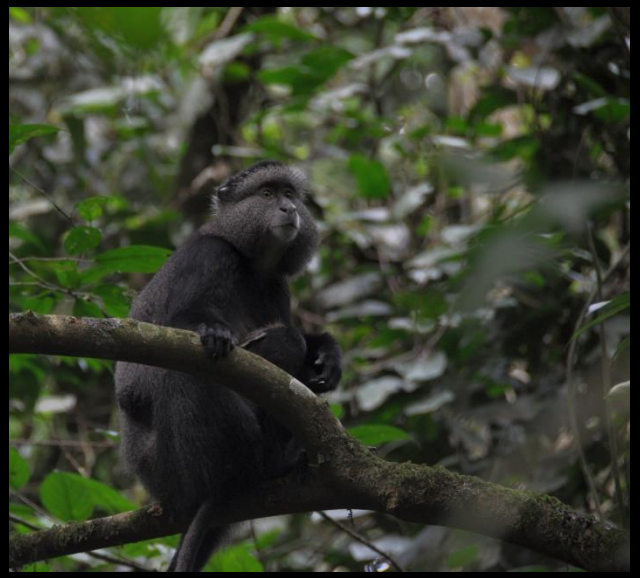
- Use SequentialData object to create weighted networks for easy analysis
- Can input date range and target behavior
- Create directed or undirected networks
- Provide statistics about behavior
- Combine multiple behaviors as new variables
- Network visualizations (time permitting)
- Merging datasets (time permitting)

Tools

- ipython/jupyter
- numpy
- pandas
- igraph
- toyplot

Data

- Behavior data collected on blue monkeys (*Cercopithecus mitis*) as part of Marina Cords's long-term study
  - Kakamega Forest, Western Kenya
  - Daily focal animal sampling for 35+ years
- One focal group's interactions in 2015



Data

- ***subj*** – anonymous ID of the subject (ID#)
- ***date*** – the date of observation
- ***beeps*** – the number of consecutive beeps (1 minute intervals) the subject was observed doing the behavior
- ***activity*** – what behavior was observed (e.g. feeding, moving, resting, social contact)
- ***gm, gmd_by, sitco*** – social partners for giving and receiving grooming, contact-sitting
- ***partner*** – all social partners (previous columns combined)

Progress

- Data subsetted and anonymized
- Implementing and improving R code in Python
- So far: can input data and subset DataFrame to relevant columns
- Next steps
  - Dictionary containing unique subjects
  - Loop through dictionary twice to get all records of interactions

Problems

- Troubleshooting igraph on Windows
- For loops as alternative to R dcast?
- Package will probably run in $O(n^2)$ time
  - Limiting factor is number of individuals
- How to make igraph a dependency?