# Documentation README (last update: 22.02.2023)

Main paths:

**Scripts:** /zi-flstorage/data/Luise/NoSeMaze2023/scripts

**Output:** /zi-flstorage/data/Luise/NoSeMaze2023/DLC\_output\_windows

# Parameter and centrality measure definitions

|  |  |
| --- | --- |
| Term | Definition |
| Closeness | Closeness finds the shortest path between nodes. In closeness takes only incoming (to a vertex), while out only takes outgoing (from a vertex), all treats the network as undirected and takes shortest overall |
| Betweenness | Which nodes are important for the flow of the network - if you remove the link how affected is the network. The extent to which a certain vertex lies on the shortest paths between other vertices |
| Eigenvector | Which nodes are powerful in the network, not only strength towards others, but also how well are your friends connected. Strong connection to a ‘loner’ isn’t as good as one to a well connected individual- can be used on both directed and undirected, but not very accurate for directed. Look at KC or pagerank instead |
| PageRank | an individual can have a high PageRank if there are many (or strong) links that point to it, or if there are some individuals that link to it and have a high PageRank. Can be used on both directed and undirected, but for undirected Eigenvector is better, and directed this is better. PR weights alot on indegree (incoming) edges so better for directed |
| KC Authority | Kleinberg Centrality for with individuals are ‘sources’, ie lots of strong incoming connections- very similar to eigenvector but for directed networks (has same value as eigenvector when used on an undirected network) |
| KC Hub | Kleinberg Centrality for which individuals link to others, ie lots of strong outgoing connections - very similar to eigenvector but for directed networks (has same value as eigenvector when used on an undirected network) |
| time\_summed | total time focal and partner spent together on specific day (undirected) |
| interactions | total number of interactions dyad had (undirected) |
| mean\_dist | mean distance between mice over all interactions that day (undirected) |
| traj\_int | Number of interactions from which a trajectory could be calculated (undirected) |
| approaches | number of approaches focal made (directed) |
| arena\_time | total time in the arena of focal |
| total\_socialt | total time spent in proximity of at least one other mouse (if in proximity of more than 1 mouse, the time doesn’t get counted double, so this may be less time than if social time is calculated by adding all time\_summed for a focal) |
| paired\_social\_prop | Of the focals total social time, how much of it was spent with that partner (time\_summed/total\_socialt)(directed) |
| paired\_arena\_prop | Of the focals total arena time, how much was spent with that partner (time\_summed/arena\_time)(directed) |
| social\_prop | Proportion of time focal was social (total\_socialt/arena\_time) |
| HWI\_time | Half weight ratio of time\_summed) (undirected) |
| HWIG\_time | HWI with correction for gregarioussness (undirected) |
| approach\_prop | Out of the interactions with trajectories, how often was the focal the approacher (approaches/traj\_int) (directed) |

[https://bookdown.org/markhoff/social\_network\_analysis/centrality.html](https://bookdown.org/markhoff/social_network_analysis/centrality.html#betweenness-centrality) ← Good source for SNA

<https://www.sci.unich.it/~francesc/teaching/network/kleinberg.html> ← KC centrality measures

<http://infolab.stanford.edu/~backrub/google.html> ← PageRank equation and logic

<https://www.sci.unich.it/~francesc/teaching/network/> ← good source on general network measures

# ReadMe (scripts)

|  |  |  |  |
| --- | --- | --- | --- |
| Order to Run: | Input | Output | Time |
| Social\_groups | Raw DLC | filtered DLC, social groups | 3-5hrs |
| Extract\_dyads | Social Group File | dyadic interations | 20min |
| Trajectory\_script | Dyadic Interactions, filtered DLC | trajectory distances before interaction (approaches) | 10min |
| extracting\_master\_matrix\_variables | Social group file, filtered DLC | Total time in arean, ROI, social time | 10min |
| SNA\_parameters | Dyadic Interations, trajectories | dyadic social parameters | 1min |
| Centrality\_measures (use centralitymeasures\_and\_filecombination instead) | SNA Parameters | centrality measures for different time resolutions, combined SNA files for resolutions | 5min |
| QualityTests | All above files | Makes file with all errors for group | 1min |
| Correlation\_file.R | combined SNA files | Correlation and boxplots | 1min |
| MasterFile\_Creator.R | SNA, ROI and centrality files | Combined files by resolution D1/D3/D7, master file (long format) | 1min |

## Social\_groups.R

* INPUT: Raw DLC
* OUTPUT: p0.8\_GX\_DX\_windowed.csv, MOUSEID\_socialgroup.csv
* USAGE:
  + variables to change: ***group***
    - ensure that all groups are in the video\_framerates.xlsx file, at position group+1
    - If in Oxytocin, uncomment OXYTOCIN section - should be able to change to any origin folder by altering parent\_path (which should lead to saved location of raw DLC)

## extract\_dyads.R

* INPUT: MOUSEID\_socialgroup.csv
* OUTPUT: dyadic\_interactions\_DX.csv
* USAGE:
  + variables to change: ***group***
    - Takes social groups for each mouse and turns it into dyadic data with; **focal, partner, socialgroup, day, dist, end, start, totaltime** (in frames)
    - Filters: frames must be consecutive, or with a gap of less than 5frames to be same social event, distance must be less than 100px, dyad must exist for at least 30 frames

## trajectory\_script.R

* INPUT: dyadic\_interactions\_DX.csv, p0.8\_GX\_DX\_windowed.csv
* OUTPUT: GX\_DX\_trajectories\_pre120frames.csv
* USAGE:
  + variables to change: **group, flstorage**
    - Potentially change ***traj\_fps*** & **min\_frames** if the fps for the videos is below 30fps.
      * Current setting (Feb 2022) : **traj\_fps**=120 (4sec), **min\_frames**=21
    - Requires at least **min\_frames** to calculated trajectory, must be an uneven number!
    - Code for raw and linear distances commented out, can graph trajectories (bottom of script)

## extracting\_master\_matrix\_variables.R

* INPUT: XMouseID.csv, p0.8\_GX\_DX\_windowed.csv
* OUTPUT: GXsocial\_time\_groups.csv, GX\_ROI\_times\_seconds.csv, GX\_ROI\_wide.csv
* USAGE:
  + variables to change: **group, flstorage**
    - ensure all videos crop dimensions are in Social\_analysis\_summary.xlsx at position group+1
    - Gives time in corners, borders, middle, total time in arena, also total social time, in long and wide format
    - total social time= total time where at least 1 other mouse was within 100px (this means that when adding time with all partners, it can be more than total social time, time isn’t counted twice when within range of more than one mouse)

## SNA\_parameters.R

* INPUT: GXsocial\_time\_groups.csv, GX\_ROI\_times\_seconds.csv, dyadic\_interactions.csv, GX\_DX\_trajectories\_pre120frames.csv,
* OUTPUT: GX\_dyadic\_SNA\_parameters.csv
* USAGE:
  + variables to change: **group, flstorage**
    - Puts all parameters for SNA into a single file, calculates HWI and HWIG for total time

## centralitymeasures\_and\_filecombination.R

* INPUT: GX\_dyadic\_SNA\_parameters
* OUTPUT: Combined\_dyadic\_SNA\_parameters\_D1.csv, Combined\_dyadic\_SNA\_parameters\_D3.csv, Combined\_dyadic\_SNA\_parameters\_D7.csv, GX\_centrality\_linear\_resDX.csv, combined\_centrality\_linear\_resDX.csv
* USAGE:
  + variables to change: ***resolution, flstorage***
    - First combines all available SNA\_parameter files at different time resolutions (D1, D3, D7), then calculates the centrality measures for resolution

## Correlation\_file.R

* INPUT: combined\_dyadic\_SNA\_parameters\_resDX.csv
* OUTPUT: correlation matrices jpegs, box plots of group by group variation, individuals\_repetitions.csv
* USAGE:
  + variables to change: ***resolution, setwd(), parameter\_type***
    - Finds where animals had repeat rounds in NoSeMaze and finds mean or sum of variable (whichever appropriate) for that round (ie all days of group combined)
    - Creates boxplots to visualise the variation of variables and centrality measures in each group
    - Reads combined files and creates correlation graphs for each variable
      * dyadic files each interaction pair gets 2 points (one for each animal as focal)

## Saving\_traj\_graphs.R

* INPUT: dyadic\_interactions, p0.8\_GX\_windowed
* OUTPUT: plots with distance travelled in 4sec before socila event starts
* USAGE:
  + variables to change: ***group***
    - Plots **ALL** cases where a trajectory can be calculated with raw and smoothed line of focal and partner

## Pattern\_reliability.R

* INPUT: Table\_minTime30\_groupX\_dayX.xlsx (from Carlas scripts)
* OUTPUT: TTpatterns.csv, Pattern\_Validity.jpeg
* USAGE:
  + Plots detection time by pattern over all groups

## locomotion\_script.R

* INPUT: p0.8\_GX\_windowed
* OUTPUT: locomotion\_GX.csv
* USAGE:
  + variables to change: **group**
    - Calculates total distance travelled for each mouse over the whole day

## MasterFile\_Creator.R

* INPUT: dyadic\_SNA\_parameters, centrality\_linear, ROI\_wide
* OUTPUT: MasterFile.xlsx
* USAGE:
  + - Wide format table for all variables, all individuals, all groups
      * Dyad\_master naming: variable\_partnerpatternID\_DX
      * Center\_master naming: centralitymeasure\_variable\_DX
      * ROI\_master naming: area\_day
    - In sheet naming, the D refers to resultion (D1 = day by day, D3= groups of 3 days, D7= week by week). D3 and D7 have res3\_X in naming. (e.g. res3\_4 indicates the 4th group of 3 days, so days 10,11,12)

## QualityTest.R

* INPUT: dyadic\_interactions, \_trajectories\_pre120, \_ROI\_times\_seconds.csv, social\_time\_groups.csv
* OUTPUT: GX \_ErrorFile.csv.csv
* USAGE:
  + Variable to change: ***group***
    - Tests for errors in files. Will only create an error file if errors present
    - Error tested: Dyad/trajectory file missing, is there a mismatch in dyads or trajectories (each should appear twice, once for each mouse as focal), do the corners+middle+border add to total time in arena, is the total social time under total time, is there missing time input, is social time missing

## NoSeMaze\_functions.R

***IMPORTANT FILE FOR ALL OTHER SCRIPTS***

|  |  |
| --- | --- |
| get\_folder\_paths | Gets iterable list of folders from a parent folder |
| total\_social\_time | Total time at least 1 other mouse was within 100px of the focal mouse |
| paste\_noNA | Pasting together columns, ignoring NA cells |
| ROI\_time | Evenly splits arena into nx x ny rectangle and calculates time in corners, borders, middle, total |
| count\_nonNaN | Counts nonNaN cells - can then be converted to time by dividing by fps |
| df\_to\_matrix | converts long format edge list (focal, partner, value) to igraph adjacency matrix |
| centrality\_measures | takes adj matrix and calculates; betweenness, closeness, in\_closeness, out\_closeness, eigenvector, pagerank, KC\_authority, KC\_hub |
| HWIG | takes half weight ratio column already calculated and corrects for gregarrioussness depending on grouped variable (day, week, 3 day etc) |
| panel.cor | in pairs() function, colors significant correlations red (color can be changed in function) |
| panel.smooth | in pairs() function, allows editing of colors of dotplot |
| panel.hist | in pairs() function, can change color of histograms |
| correlation\_plot\_center | correlations centrality measures by variable and plots |
| correlation\_plot\_dyads | correlates variables and plots |

# OLD - better alternatives available

## centrality\_measures.R

* INPUT: GX\_dyadic\_SNA\_parameters.csv
* OUTPUT: Centrality\_measures\_GX.xlsx, GX\_centrality\_linear.csv
* USAGE:
  + variables to change: **group, flstorage**
    - Gets centrality measures (betweenness, closeness, eigenvector, pagerank) for each of the parameters in SNA\_parameters

## Calculating\_NoNNaN\_values.R

* INPUT: p0.8\_GX\_windowed
* OUTPUT: GX\_p0.8\_nonNaNcount.csv
* USAGE:
  + variables to change: ***group***
    - Calculates total time in arena on day

## ROI\_wide\_master.R

* INPUT: GX\_ROI\_times\_seconds.csv
* OUTPUT: GX\_MASTER\_ROI\_wide.csv
* USAGE:
  + Converts long format ROI to wide format