

Methods

Pablo E. Gutiérrez-Fonseca

2022-05-08

Glossary

Edges: number of ties.

Nodes: can e.g. be people in a social network, genes in a co-expression network, etc. Nodes are connected via ties/edges.

Function setup -> `graph_from_data_frame()`

We used ‘directed = False’ (this is Undirected network) since Source (main author) and Target (co-authors) are not meaningful.

It is an undirected network, a graph with bidirectional edges in contrast with a directed graph in which the direction of an edge from one vertex to another is considered

<https://www.jessesadler.com/post/network-analysis-with-r/>

Directed and Undirected definition: <https://ona-book.org/working.html>

Many papers on coauthorship analysis use Undirected network:

1- Azondekon, R., Harper, Z. J., Agossa, F. R., Welzig, C. M., & McRoy, S. (2018). Scientific authorship and collaboration network analysis on malaria research in Benin: papers indexed in the web of science (1996–2016). *Global Health Research and Policy*, 3(1), 1-11.

2- Han, S. H., Chae, C., & Passmore, D. L. (2019). Social network analysis and social capital in human resource development research: A practical introduction to R use. *Human Resource Development Quarterly*, 30(2), 219-243.

3- Ho, T. M., Nguyen, H. V., Vuong, T. T., Dam, Q. M., Pham, H. H., & Vuong, Q. H. (2017). Exploring Vietnamese co-authorship patterns in social sciences with basic network measures of 2008-2017 Scopus data. *F1000Research*, 6.

4- Medina, A. M. (2018). Why do ecologists search for co-authorships? Patterns of co-authorship networks in ecology (1977–2016). *Scientometrics*, 116(3), 1853-1865.

Find duplicate in a column

`duplicated(df$column)`

Centrality

Central Authors likely wield a disproportionate amount of influence on the network. Therefore, high centralization provides fewer Authors with more power and control.

High centrality networks have few nodes with many connections, low centrality networks have many nodes with similar numbers of edges.

The first variable **\$res** provides node-level centrality scores.

The **\$centralization** variable provides our actual centrality score, which is quite small suggesting a very decentralized network.

The **\$theoretical_max** provides the maximum theoretical graph level centralization score for our graph based on the number of vertices.

Example:

<https://rpubs.com/smbrook3/unit-2-case-study>

Density

Density refers to the number of ties in the network reported as a fraction of the total possible number of ties... The closer this number is to 1.0, the denser the network.

The igraph packages has two different function for calculating the density of a network, `edge_density()` and `graph.density()`, but both yield the exact same result. The `edge_density()` function is the more recent function and should be used.

Example:

<https://rpubs.com/smbrook3/unit-2-case-study>

Homophily (or assortativity)

Values less than 1 indicate stronger subgroup cohesion than would be expected in a random network (i.e., homophily), and values higher than 1 represent higher than expected rates of collaboration (i.e., heterophily).

We used:

```
assortativity_nominal(network2004, as.factor(V(network2004)$affiliation))
```

1- “as.factor” Convert Affiliation in numbers 2- Since continent affiliation is a categorical variable, we should use the corresponding definition of the assortativity coefficient. The assortativity coefficient for categorical properties can be calculated by

```
assortativity_nominal(network2004, as.factor(V(network2004)$affiliation))
```

Example: <https://methods.sagepub.com/dataset/howtoguide/assortativity-in-ukfaculty-2008>

https://kateto.net/wp-content/uploads/2016/01/NetSciX_2016_Workshop.pdf

References

<https://www.r-bloggers.com/2017/05/network-analysis-of-game-of-thrones-family-ties/>

https://www.jstor.org/stable/pdf/26447831.pdf?refreqid=excelsior%3A5e716af69759e46e51fc7561026d4596&ab_segments=&origin=

<https://rpubs.com/MarkusLoew/226759>

<https://rpubs.com/nr2462/845876>

<https://igraph.org/r/html/latest/>