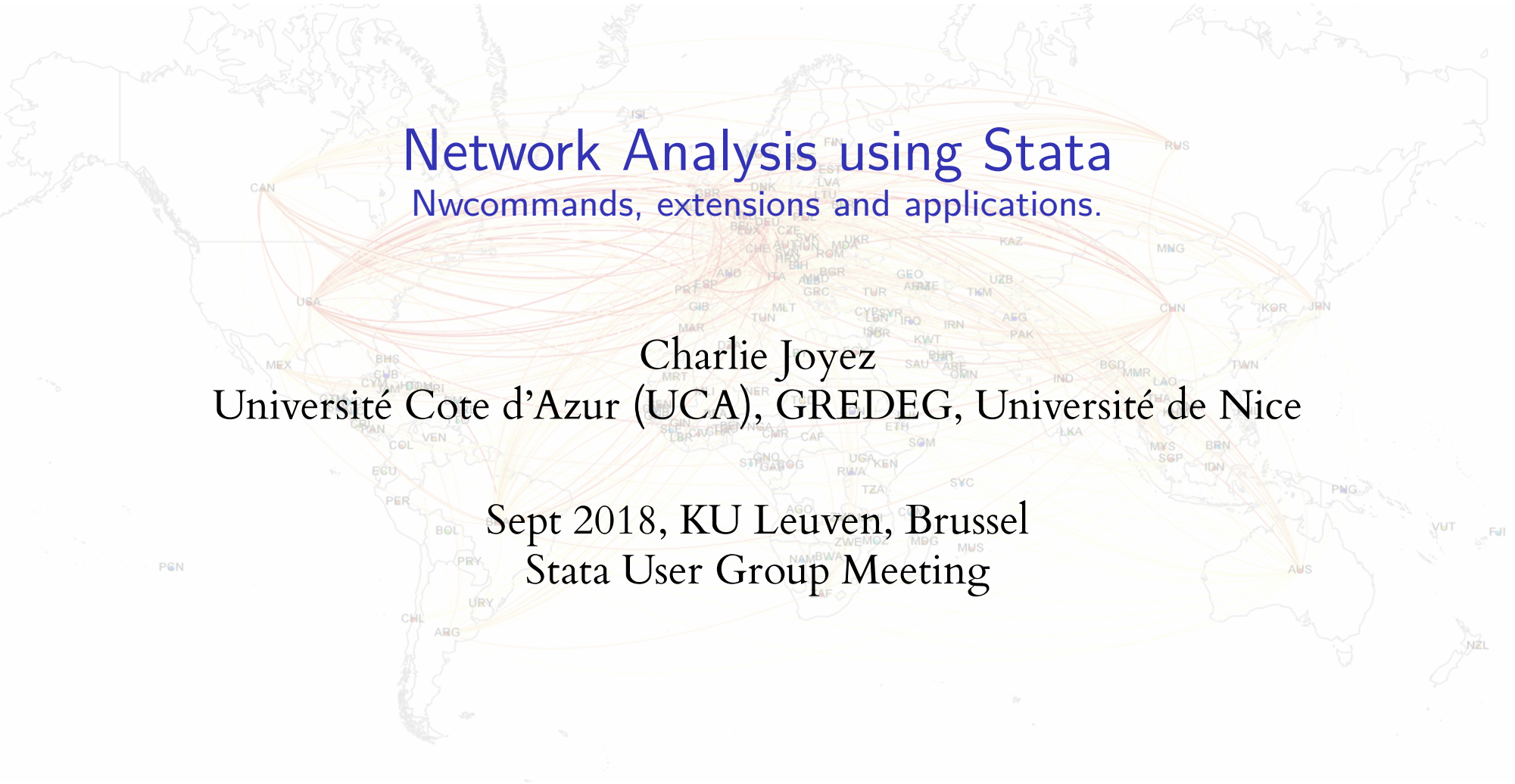


Network Analysis using Stata

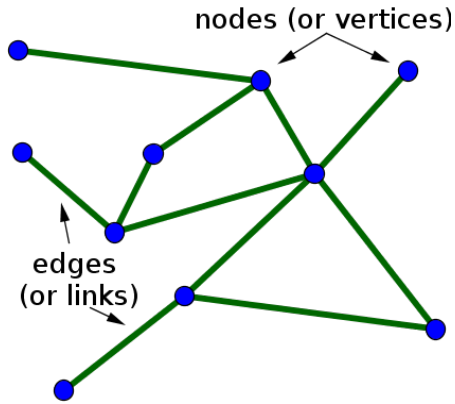
Nwcommands, extensions and applications.

Charlie Joyez
Université Cote d'Azur (UCA), GREDEG, Université de Nice

Sept 2018, KU Leuven, Brussel
Stata User Group Meeting

Motivation

Networks are everywhere. flexible mathematical object



- Complex systems, interactions, interdependence's.
- Two type of use in (Social) Sciences
  - Theoretical modeling with complex micro-foundations
  - Empirical analysis of existing networks.
    - Booming in several fields with data availability and computing capabilities.
    - Increasing interest (See Stata news january 2018 (33-1))

Objective

How to easily proceed to network analysis using Stata?

- Node level and network wide analysis

Outline

nwcommands - Presentation

- Developed (maintained) by Thomas Grund - Univ. College Dublin
  - http://nwcommands.org
  - `install nwcommands-ado, from(http://www.nwcommands.org)`
- Entire suite of commands, close to Stata commands (nw prefix)
  - declare, use, save network data
  - Manipulate (keep, drop, permute, etc.) nodes or entire networks
  - Compute network metrics
    - At the node level (centrality, etc)
    - At the entire network level (density, overall clustering coeff).

Declare Data

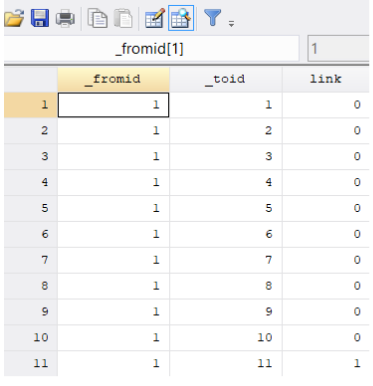
- From a Mata Matrix (Adjacency matrix)
  - `mata A=(0,10,1 \5,0,0 \0,2,0)`

    `mata A`

```
. mata A
        1    2    3
    +-------------------+
  1 |   0   10    1     |
  2 |   5    0    0     |
  3 |   0    2    0     |
    +-------------------+
```

    `nwset, mat(A) name(netA)`

- From an edge list



_fromid[1] 1

	_fromid	_toid	link
1	1	1	0
2	1	2	0
3	1	3	0
4	1	4	0
5	1	5	0
6	1	6	0
7	1	7	0
8	1	8	0
9	1	9	0
10	1	10	0
11	1	11	1

  - `nwfromedge _fromid _toid link, name(Net1) undirected`

Node-level metrics

- **`nwdegree`**
  - `_degree`: Number of direct neighbors
  - $d_i = \sum_j m_{i,j}$, $M = A$: /A Unweighted adjacency matrix
  - returns Freeman (1979) index

$$C_x = \frac{\sum_{i=1}^{N} C_x(p_*) - C_x(p_i)}{\max \sum_{i=1}^{N} C_x(p_*) - C_x(p_i)}$$

- **`nwdegree, valued`**
  - `_strength`: Sum of edges weights
  - $s_i = \sum_j a_{ij}$

- Other node centrality metrics : Betweeness & closeness, Katz, Eigenvector.

Network-wide information

- **`tnwsummarize`**

```
. nwsummarize
```

```
   Network name:  test
   Network id:  1
   Directed: true
   Nodes: 6
   Arcs: 14
   Minimum value:  0
   Maximum value:  1
   Density:  .4666666666666667
```

- **`nwgeodesic`**
  - Longest past, diameter, avg shortest path (unweighted)
- **`nwclustering`** Overall clustering coefficient (nb triads / nb possible triads)

Outline

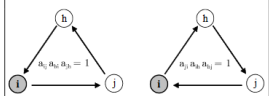
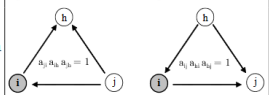
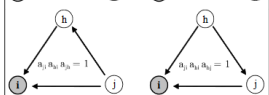
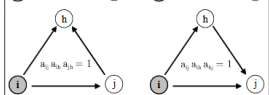
Node-level metrics 1/2

- Average Nearest Neighbors Degree (Strength)
  - `nwannd` : Average nearest neighbor degree.
    -
```
mata
neighbor = mymat:>0
Z=st_data(.,"_degree")
mata:  totdegreemat = neighbor*Z
mata:  ANNDmat=totdegreemat:/Z
end mata
```
- nwdisparity (Barthélemy et al., 2005) : distribution of edge's weight (concentration)
  $disparity_i = \sum_j (w_{ij}/s_i)^2$
- nw_harmonic centrality (suited for disconnected graphs)
  - $H(x) = \sum_{y \neq x} \frac{1}{d(y,x)}$

Node-level metrics 2/2

Weighted / directed extension of existing commands

- **`nwcluster`** : directions and/or weighted generalization (Onnela et al., 2005)
- **`nw_wcc`** : Weighted Clustering Coefficients (Fagiolo, 2006)

Patterns	Graphs	t_i^*	T_i^*	CCs for BDNs	CCs for WDNs
Cycle		$(A)^3_{ii}$	$d_i^{in}d_i^{out} - d_i^{\leftrightarrow}$	$C_i^{cyc} = \frac{(A)^3_{ii}}{d_i^{in}d_i^{out} - d_i^{\leftrightarrow}}$	$\tilde{C}_i^{cyc} = \frac{(\hat{W})^3_{ii}}{d_i^{in}d_i^{out} - d_i^{\leftrightarrow}}$
Middleman		$(AA^TA)_{ii}$	$d_i^{in}d_i^{out} - d_i^{\leftrightarrow}$	$C_i^{mid} = \frac{(AA^TA)_{ii}}{d_i^{in}d_i^{out} - d_i^{\leftrightarrow}}$	$\tilde{C}_i^{mid} = \frac{(\hat{W}\hat{W}^T\hat{W})_{ii}}{d_i^{in}d_i^{out} - d_i^{\leftrightarrow}}$
In		$(A^TA^2)_{ii}$	$d_i^{in}(d_i^{in} - 1)$	$C_i^{in} = \frac{(A^TA^2)_{ii}}{d_i^{in}(d_i^{in}-1)}$	$\tilde{C}_i^{in} = \frac{(\hat{W}^T\hat{W}^2)_{ii}}{d_i^{in}(d_i^{in}-1)}$
Out		$(A^2A^T)_{ii}$	$d_i^{out}(d_i^{out} - 1)$	$C_i^{out} = \frac{(A^2A^T)_{ii}}{d_i^{out}(d_i^{out}-1)}$	$\tilde{C}_i^{out} = \frac{(\hat{W}^2\hat{W}^T)_{ii}}{d_i^{out}(d_i^{out}-1)}$
All (D)	All 8 graphs above	$\frac{(A+A^T)^3_{ii}}{2}$	$d_i^{tot}(d_i^{tot} - 1) - 2d_i^{\leftrightarrow}$	$C_i^D = \frac{(A+A^T)^3_{ii}}{2T_i^D}$	$\tilde{C}_i^D = \frac{(\hat{W}+\hat{W}^T)^3_{ii}}{2T_i^D}$

- **`nw_geodesic`** (weights as distance)

Network level

- **nwreciprocity** (Barrat et al., 2004)
  - ```
    mata
    s=sum(W)
    Z = W :* (W :< W') + W' :* (W' :< W) /*min of symmetrics
    elements = reciprocated ties*/
    E=sum(Z)
    r=E/s
    end mata
    ```
  - Compares reciprocity with N random draws (same size, density).
- **nwstrengthcent** : (Freeman, 1979) index based on Strength.
    ```

Declaration

- From neigbor lists (only existing ties). A variable may indicate the sequence.
  - **nw_fromneighbor** `nw_fromlist test,node(NODE) id(ID) direction(year)`

Initial data


Data Editor (Browse) - [Untitled]
File Edit View Data Tools
var5[4]

	ID	NODE	year
1	1	AUT	1
2	1	USA	2
3	2	USA	1
4	2	GBR	1
5	3	GBR	2
6	3	FRA	2
7	3	ESP	1
8	4	ESP	1
9	4	ITA	1
10	4	USA	2
11	4	USA	3
12	.	USA	3
13	4		3

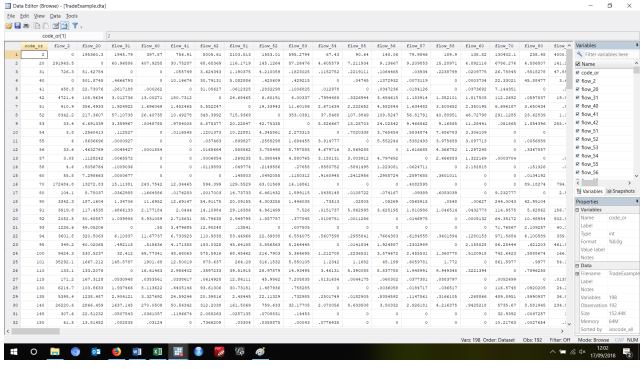
Final data

	NODE	_AUT	_ESP	_FRA	_GBR	_ITA	_USA
1	AUT	0	0	0	0	0	1
2	ESP	0	0	1	1	1	1
3	FRA	0	1	0	0	0	0
4	GBR	0	1	1	0	0	1
5	ITA	0	1	0	0	0	1
6	USA	0	1	0	1	1	0

Outline

Application to international economics

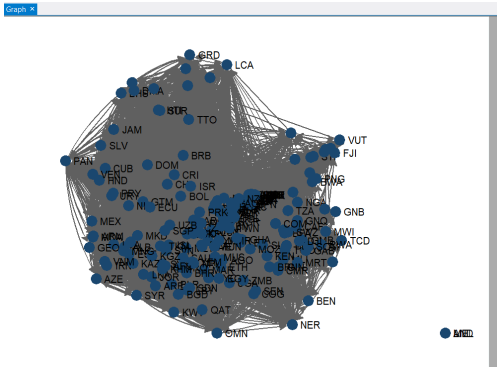
- World Trade Web on 2016



- Declare to be weighted directed network : 192*192 table

```
mkmat flow_*, matrix(M)
mata A=st_matrix("M")
nwset , mat(A) name(TradeNet‘v’)
```

Application - 2



- Most central nodes?
  - Eigenvector centrality

Network name: **TradeNet_3**

Eigenvector centrality

Variable	Obs	Mean	Std. Dev.	Min	Max
_evcent	183	.0716463	.0182512	.0114232	.0927111

```
. rename _evcent evcent_unw

. nw_evcent
```

Network name: **TradeNet_3**

Eigenvector centrality

Variable	Obs	Mean	Std. Dev.	Min	Max
_evcent	183	.0246924	.0698673	5.23e-06	.5755326

```
. rename _evcent evcent_w

.
```

TZA	.039391	.0000414
UGA	.0506185	.0001651
UKR	.0907891	.0191053
URY	.0808169	.0013135
USA	.0924482	.5755326
UZB	.0913563	.1499816

Degree centralization 0.364

Strength Centralization 0.176

Application - 3

Econometrics of networks

- Use of network metrics (e.g. centrality indexes of nodes) into traditional analysis. (Hidalgo et al., 2007)
- Regress network structure (dyadic data)
  - Individuals in networks not iid
    - OLS biased unless FE or clustering
    - QAP : unit = dyadic value + random permutations of rows and columns.
  - `nwqap MNEnet_2011 GVCnet_2010 , mode(dist) type(reg) permutation(500)`

```
Permutation: 500 out of 500


Multiple Regression Quadratic Assignment Procedure

  Estimation           =  QAP
  Regression           =  reg
  Permutations         =  500
  Number of vertices   =  54
  Number of edges      =  1430
```

MNEnet_forGVC2011	Coef.	P-value
GVCnet_forMNE2010	.001906	0
_cons	37.52447	

Outline

Conclusion

- Network analysis made easy through Stata
  - easy to learn and contribute
  - suited to a wide range of issues

Next steps

  - generalize metrics to weighted, directed, unconnected graphs.
    - Fit to complex networks.
  - improve network graphs & plots vizualization
  - Incorporate nwcommands into Stata 16?
  - Promote network analysis to colleague/students already familiar with Stata.

Thank you

charlie.joyez@unice.fr

Many thanks to **Thomas Grund** for its **nwcommands: Network Analysis with Stata**
Additional Stata commands used for this paper are available on my RePEc Ideas page or directy from SSC (e.g. `ssc install nwannd`)

References

Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, 2004.

Barthélemy, M., Barrat, A., Pastor-Satorras, R., and Vespignani, A. Characterization and modeling of weighted networks. *Physica a: Statistical mechanics and its applications*, 346(1): 34–43, 2005.

Fagiolo, G. Directed or undirected? a new index to check for directionality of relations in socio-economic networks. *Economics Bulletin*, 3(34):1–12, 2006.

Freeman, L. C. Centrality in social networks conceptual clarification. *Social networks*, 1(3): 215–239, 1979.

Hidalgo, C. A., Klinger, B., Barabasi, A.-L., and Hausmann, R. The product space conditions the development of nations. Paper 0708.2090, arXiv.org, 2007. 00468.

Joyez, C. On the topological structure of multinationals network. *Physica A: Statistical Mechanics and its Applications*, 473:578 – 588, 2017.

Onnela, J.-P., Saramäki, J., Kertész, J., and Kaski, K. Intensity and coherence of motifs in weighted complex networks. *Physical Review E*, 71(6):065103, 2005.