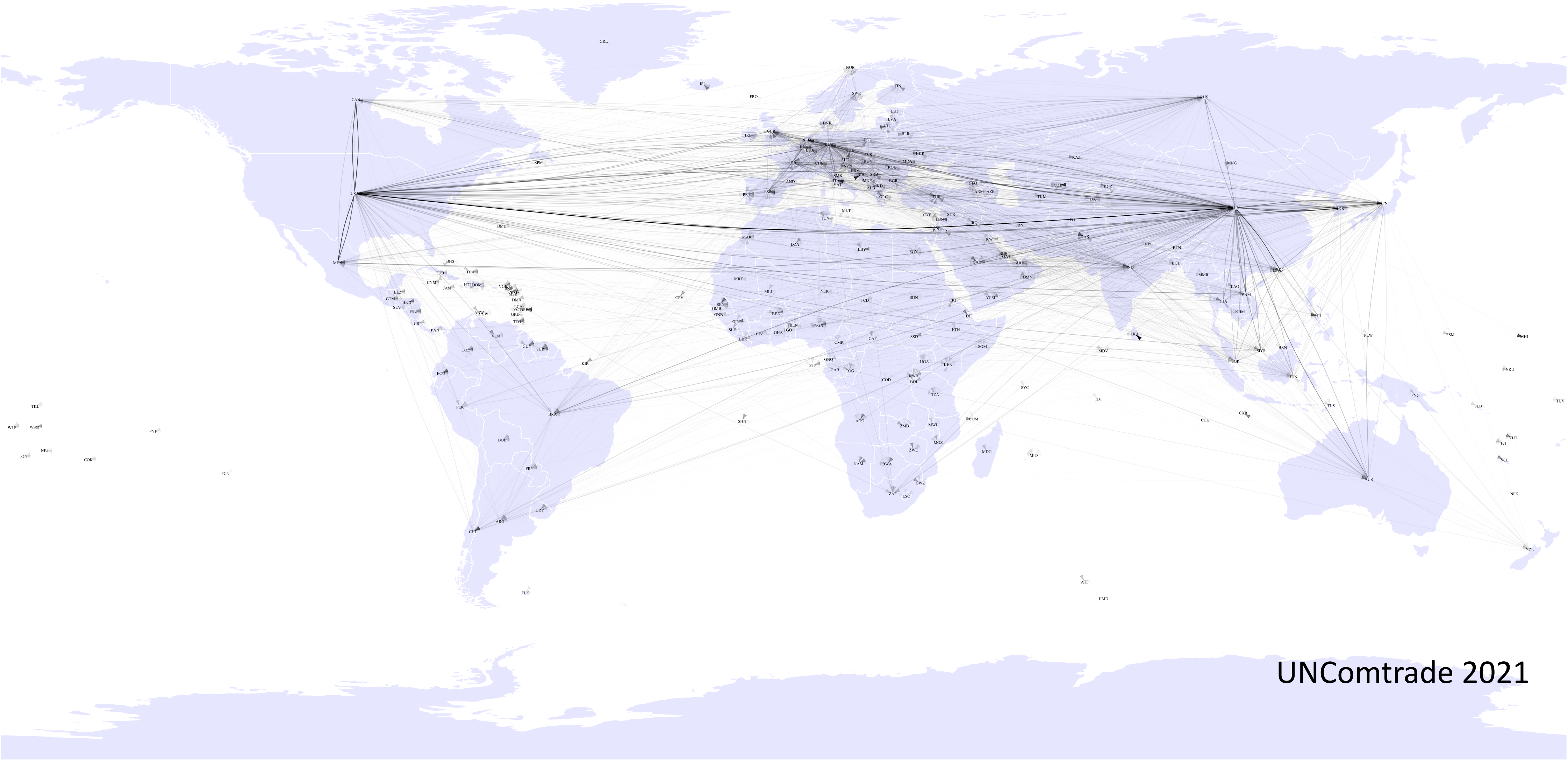
Higher order assortativity for directed weighted networks

Alberto Arcagni
MEMOTEF Department, Sapienza University of Rome, Italy

Roy Cerqueti
Department of Economic and Social Sciences, Sapienza University of Rome , Italy
GRANEM, Université d'Angers, SFR CONFLUENCES, F-49000 Angers, France

Rosanna Grassi
Department of Statistics and Quantitative Methods, University of Milano - Bicocca , Italy

Motivating application: world trade network



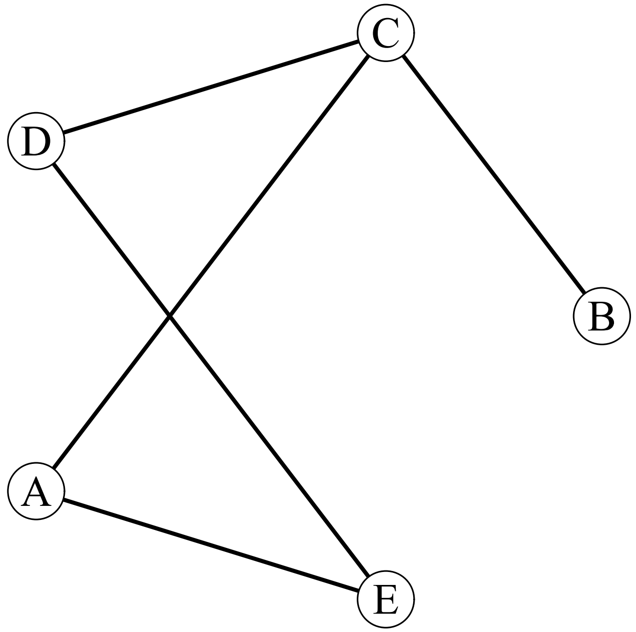
UNComtrade 2021

Assortativity

- Assortativity is a graph metric and describes the tendency of high degree nodes to be directly connected to high degree nodes and low degree nodes to low degree nodes.
- Newman (2002) index for an undirected and unweighted networks

$$r = \frac{\sum_i e_{ii} - \sum_i q_i^2}{1 - \sum_i q_i^2}$$

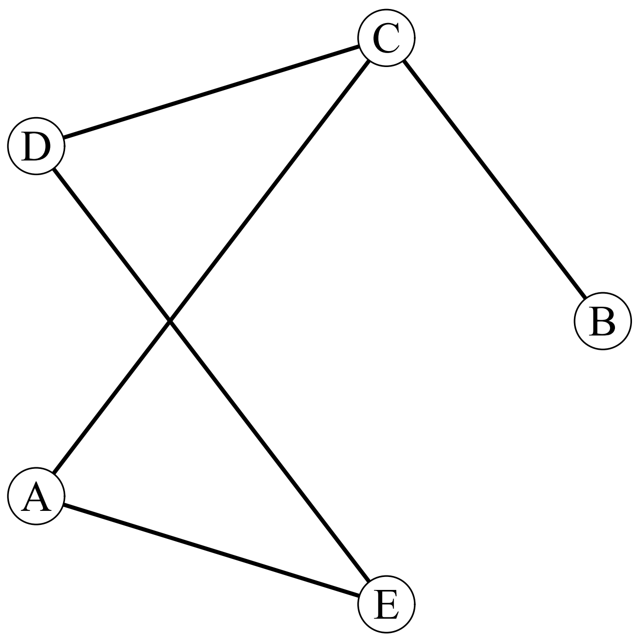
Adjacency



	A	B	C	D	E	
A			1		1	**2**
B			1			**1**
C	1	1		1		**3**
D			1		1	**2**
E	1			1		**2**
	2	**1**	**3**	**2**	**2**	$2 \cdot 5$

$$n = \|\mathrm{A}\|_{1,1} = 2 \cdot |E|$$
$$\boldsymbol{d} = \mathrm{A}\mathbf{1}$$
$$\mathrm{E} = \mathrm{A}/n$$
$$\boldsymbol{q} = \mathrm{E}\mathbf{1} = \boldsymbol{d}/n$$

Degree centrality



Vertex	Degree
A	2
B	1
C	3
D	2
E	2

Covariance with a joint distribution

X\Y	y_1	$\cdots$	y_j	$\cdots$	y_c	
x_1	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	n_{10}
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
x_i	$\cdots$	$\cdots$	n_{ij}	$\cdots$	$\cdots$	n_{i0}
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
x_r	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	n_{r0}
	n_{01}	$\cdots$	n_{0j}	$\cdots$	n_{0c}	n

$$
\begin{aligned}
Cov(X,Y;\mathrm{N}) &= \frac{1}{n}\sum_{ij}(x_i-\bar{x})(y_j-\bar{y})n_{ij} = \\
&= \frac{1}{n}\sum_{ij} x_i y_j n_{ij} - \bar{x}\bar{y} = \\
&= \frac{1}{n}\boldsymbol{x}'\mathrm{N}\boldsymbol{y} - \frac{1}{n}\boldsymbol{x}'\boldsymbol{n}_X \cdot \frac{1}{n}\boldsymbol{n}_Y'\boldsymbol{y} = \\
&= \frac{1}{n}\boldsymbol{x}'\left(\mathrm{N} - \frac{\boldsymbol{n}_X\boldsymbol{n}_Y'}{n}\right)\boldsymbol{y}
\end{aligned}
$$

$$\boldsymbol{n}_X = \mathrm{N}\mathbf{1}, \boldsymbol{n}_Y = \mathrm{N}'\mathbf{1}, n = \|\mathrm{N}\|_{1,1}$$

Covariance with a joint distribution

X\Y	y_1	$\cdots$	y_j	$\cdots$	y_c	
x_1	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	q_{10}
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
x_i	$\cdots$	$\cdots$	e_{ij}	$\cdots$	$\cdots$	q_{i0}
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
x_r	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	q_{r0}
	q_{01}	$\cdots$	q_{0j}	$\cdots$	q_{0c}	1

$$Cov(X, Y; \mathrm{N}) = \boldsymbol{x}'(\mathrm{E} - \boldsymbol{q}_X \boldsymbol{q}_Y')\boldsymbol{y}$$

$$\mathrm{E} = \frac{\mathrm{N}}{n}$$

$$\boldsymbol{q}_X = \mathrm{E}\mathbf{1} = \frac{\boldsymbol{n}_X}{n}$$

$$\boldsymbol{q}_Y = \mathrm{E}'\mathbf{1} = \frac{\boldsymbol{n}_Y}{n}$$

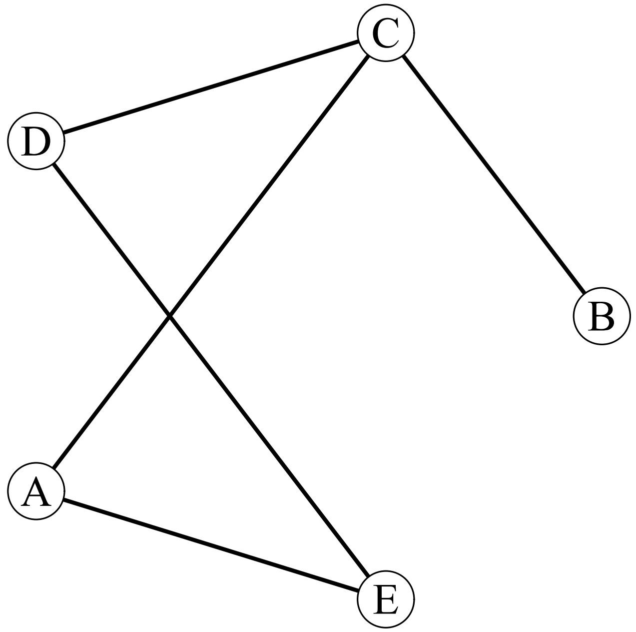
Matrix notation of variance

X\X	x_1	$\cdots$	x_i	$\cdots$	x_r	
x_1	n_{10}					$\boldsymbol{n_{10}}$
$\vdots$		$\ddots$				$\vdots$
x_i			n_{i0}			$\boldsymbol{n_{i0}}$
$\vdots$				$\ddots$		$\vdots$
x_r					n_{r0}	$\boldsymbol{n_{r0}}$
	$\boldsymbol{n_{10}}$	$\cdots$	$\boldsymbol{n_{i0}}$	$\cdots$	$\boldsymbol{n_{r0}}$	

$$Var(X;\boldsymbol{n}_X) = \frac{1}{n}\boldsymbol{x}'\left(\mathrm{D}_{\boldsymbol{n}_X} - \boldsymbol{n}_X\boldsymbol{n}_X'\right)\boldsymbol{x}$$

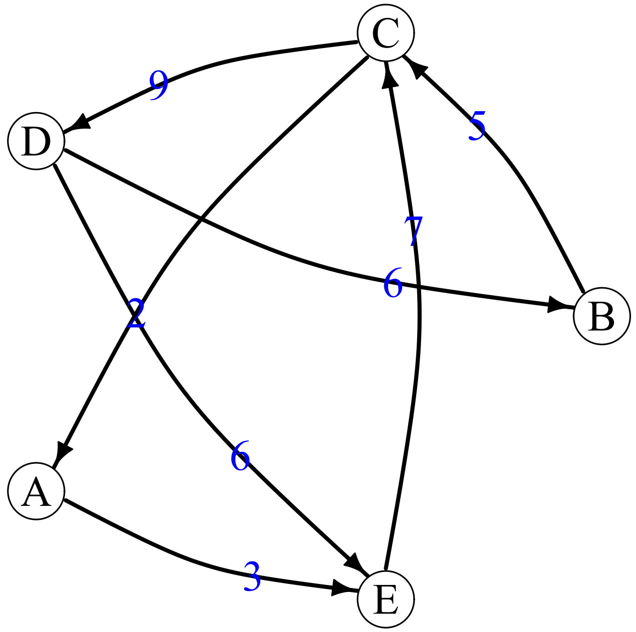
$$Var(X;\boldsymbol{n}_X) = \boldsymbol{x}'\left(\mathrm{D}_{\boldsymbol{q}_X} - \boldsymbol{q}_X\boldsymbol{q}_X'\right)\boldsymbol{x}$$

Matrix notation of Newman's index



$$r = \frac{\boldsymbol{d}'(\mathrm{E} - \boldsymbol{q}\boldsymbol{q}')\boldsymbol{d}}{\boldsymbol{d}'\left(\mathrm{D}_{\boldsymbol{q}} - \boldsymbol{q}\boldsymbol{q}'\right)\boldsymbol{d}}$$

Directed and weighted newtworks



	A	B	C	D	E	out
A					3	3
B			5			5
C	2			9		11
D		6			6	12
E			7			7
in	2	6	12	9	9	*t=38*

$$\mathrm{A} = [a_{ij} : w_{ij} > 0 \; \forall ij \in 1, .., n]$$

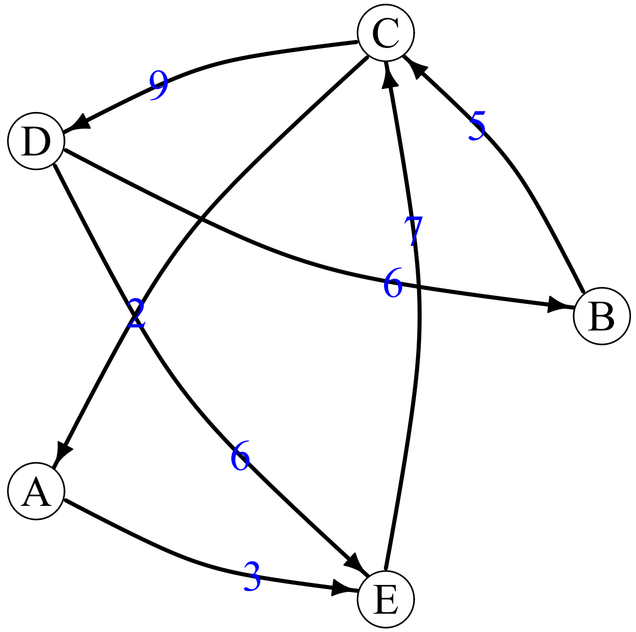
$$t = \|\mathrm{W}\|_{1,1}$$

$$\boldsymbol{s}_{out} = \mathrm{W}\mathbf{1}, \boldsymbol{s}_{in} = \mathrm{W}'\mathbf{1}$$

$$\mathrm{E_w} = \frac{\mathrm{W}}{\mathrm{t}}$$

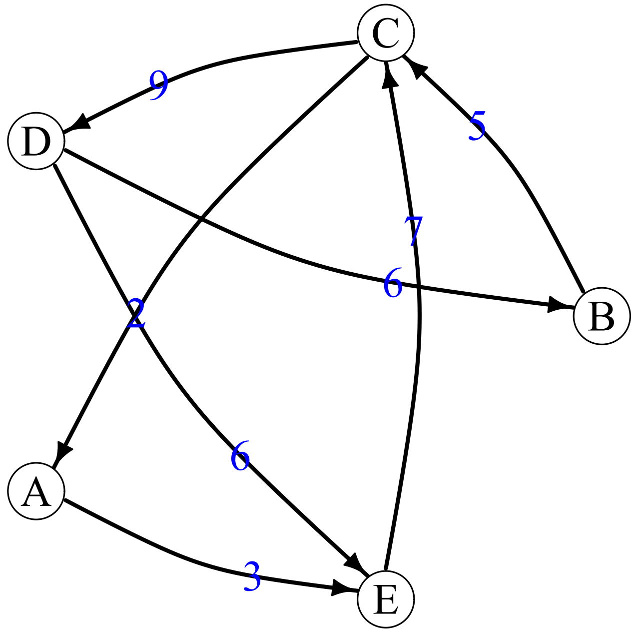
$$\boldsymbol{p}_{out} = \mathrm{E_W}\mathbf{1} = \frac{\boldsymbol{s}_{out}}{t}, \boldsymbol{p}_{in} = \mathrm{E}'_{\mathrm{W}}\mathbf{1} = \frac{\boldsymbol{s}_{in}}{t}$$

In-out degree and strength centrality



Vertex	d_{in}	d_{out}	s_{in}	s_{out}
A	1	1	2	3
B	1	1	6	5
C	2	2	12	11
D	1	2	9	12
E	2	1	9	7

Newman's index for weighted and directed networks



$$r(\boldsymbol{d}_{out}, \boldsymbol{d}_{in}|\mathrm{A}) = \frac{\boldsymbol{d}'_{out}(\mathrm{E} - \boldsymbol{q}_{out}\boldsymbol{q}'_{in})\boldsymbol{d}_{in}}{\sqrt{Var(\boldsymbol{d}_{out}|\boldsymbol{d}_{out}) \cdot Var(\boldsymbol{d}_{in}|\boldsymbol{d}_{in})}}$$

$$r(\boldsymbol{s}_{out}, \boldsymbol{s}_{in}|\mathrm{A}) = \frac{\boldsymbol{s}'_{out}(\mathrm{E} - \boldsymbol{q}_{out}\boldsymbol{q}'_{in})\boldsymbol{s}_{in}}{\sqrt{Var(\boldsymbol{s}_{out}|\boldsymbol{d}_{out}) \cdot Var(\boldsymbol{s}_{in}|\boldsymbol{d}_{in})}}$$

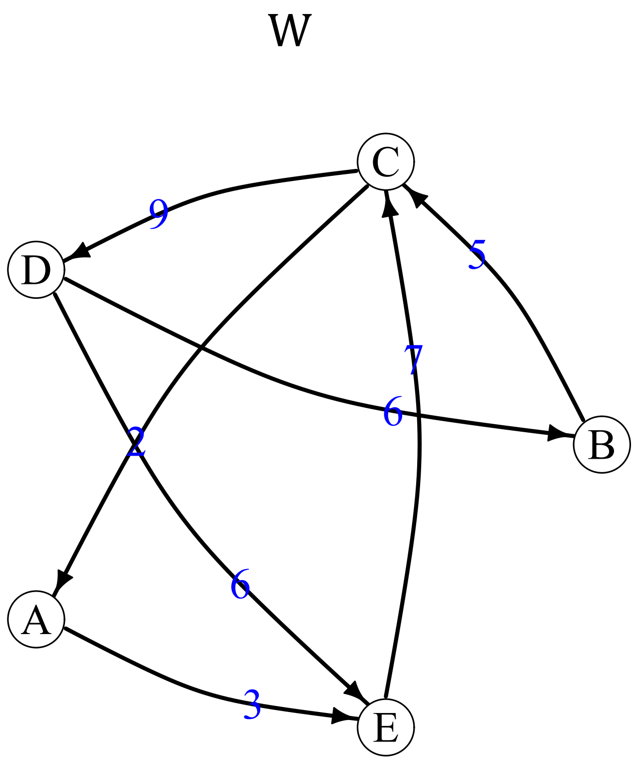
$$r(\boldsymbol{d}_{out}, \boldsymbol{d}_{in}|\mathrm{W}) = \frac{\boldsymbol{d}'_{out}(\mathrm{E_W} - \boldsymbol{p}_{out}\boldsymbol{p}'_{in})\boldsymbol{d}_{in}}{\sqrt{Var(\boldsymbol{d}_{out}|\boldsymbol{s}_{out}) \cdot Var(\boldsymbol{d}_{in}|\boldsymbol{s}_{in})}}$$

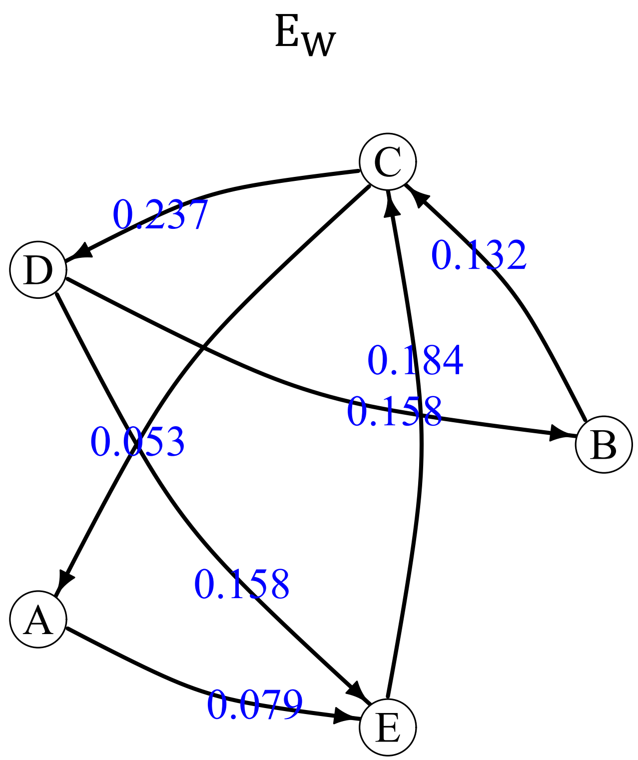
$$r(\boldsymbol{s}_{out}, \boldsymbol{s}_{in}|\mathrm{W}) = \frac{\boldsymbol{s}'_{out}(\mathrm{E_W} - \boldsymbol{p}_{out}\boldsymbol{p}'_{in})\boldsymbol{s}_{in}}{\sqrt{Var(\boldsymbol{s}_{out}|\boldsymbol{s}_{out}) \cdot Var(\boldsymbol{s}_{in}|\boldsymbol{s}_{in})}}$$

…

Note: $r(\boldsymbol{s}_{out}, \boldsymbol{s}_{in}|\mathrm{W}) \neq r(\boldsymbol{s}_{in}, \boldsymbol{s}_{out}|\mathrm{W})$

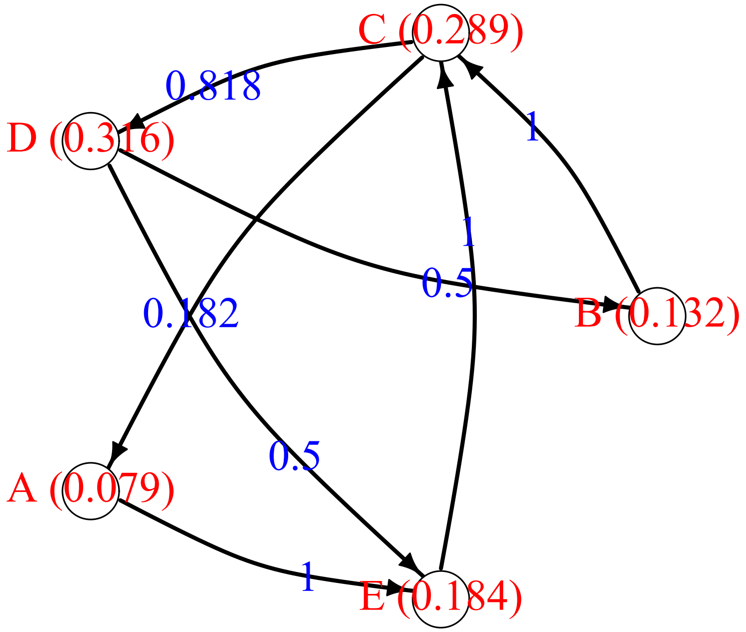
Edge sampling





Random walk

$$E_W = D_{p_0} P$$



$$\boldsymbol{p}_{out} := \boldsymbol{p}_0$$

$$E_1 = E_W = D_{\boldsymbol{p}_0} P$$

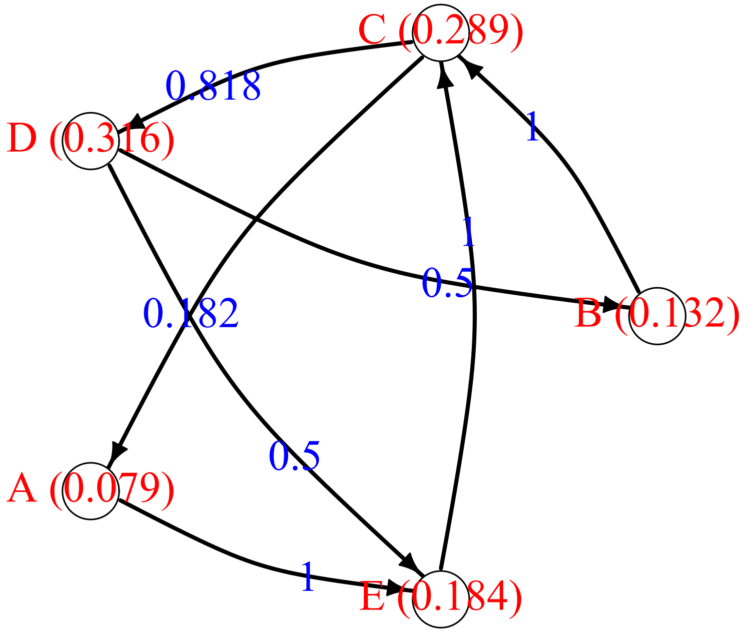
$$E_h = D_{\boldsymbol{p}_0} P_h = D_{\boldsymbol{p}_0} P^h$$

$$P^h \mathbf{1} = \mathbf{1}$$

$$E_h \mathbf{1} = D_{\boldsymbol{p}_0} P^h \mathbf{1} = D_{\boldsymbol{p}_0} \mathbf{1} = \boldsymbol{p}_0$$

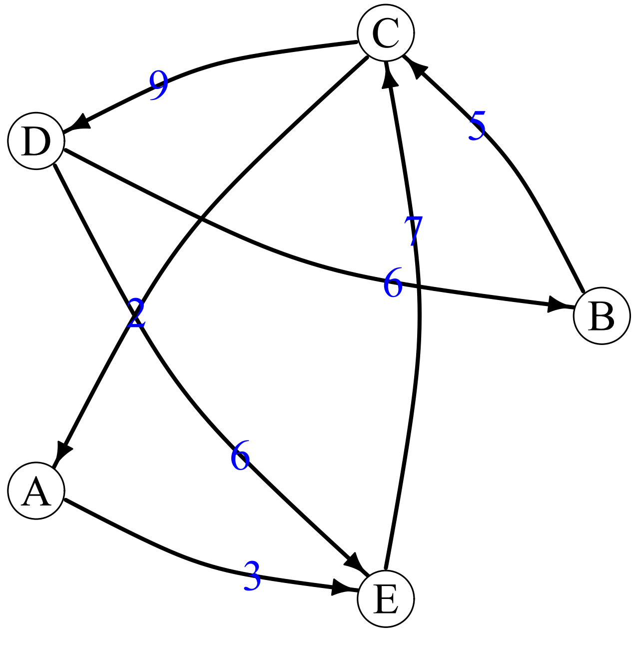
$$E_h' \mathbf{1} = \left(P^h\right)' \boldsymbol{p}_0 := \boldsymbol{p}_h ; \boldsymbol{p}_h' \mathbf{1} = 1$$

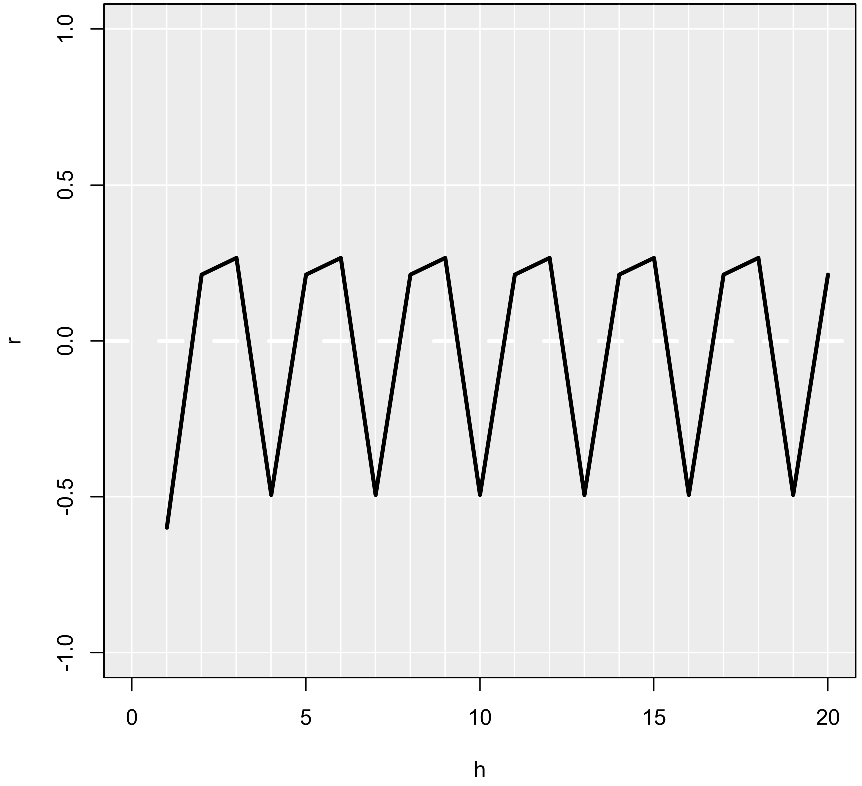
Higher order assortativity



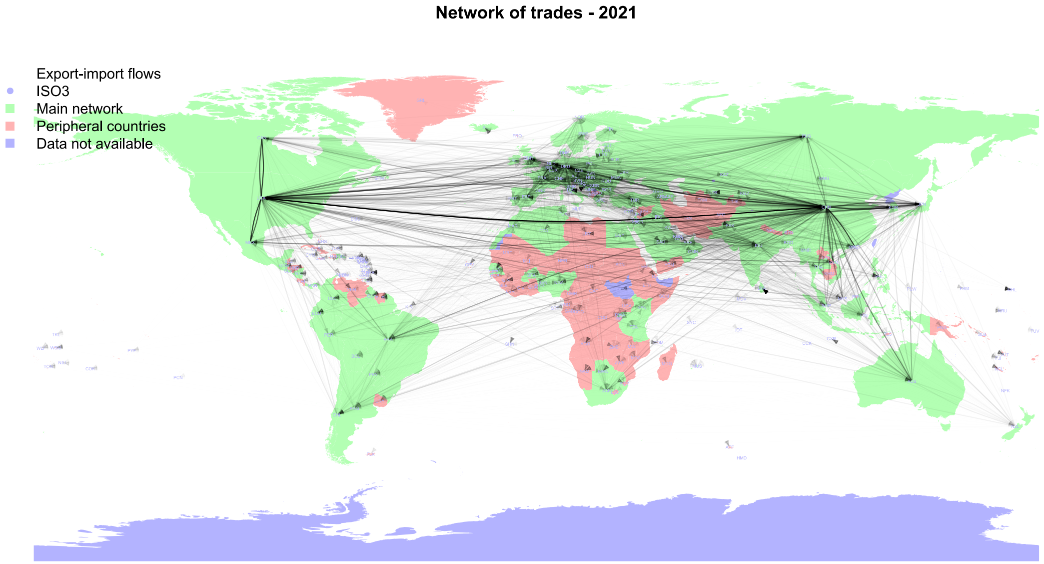
$$r_h(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{x}'\left(\mathrm{D}_{\boldsymbol{p}_0}\mathrm{P}^h - \boldsymbol{p}_0\boldsymbol{p}_h'\right)\boldsymbol{y}}{\sqrt{Var(\boldsymbol{x}; \boldsymbol{p}_0)Var(\boldsymbol{y}; \boldsymbol{p}_h)}}$$

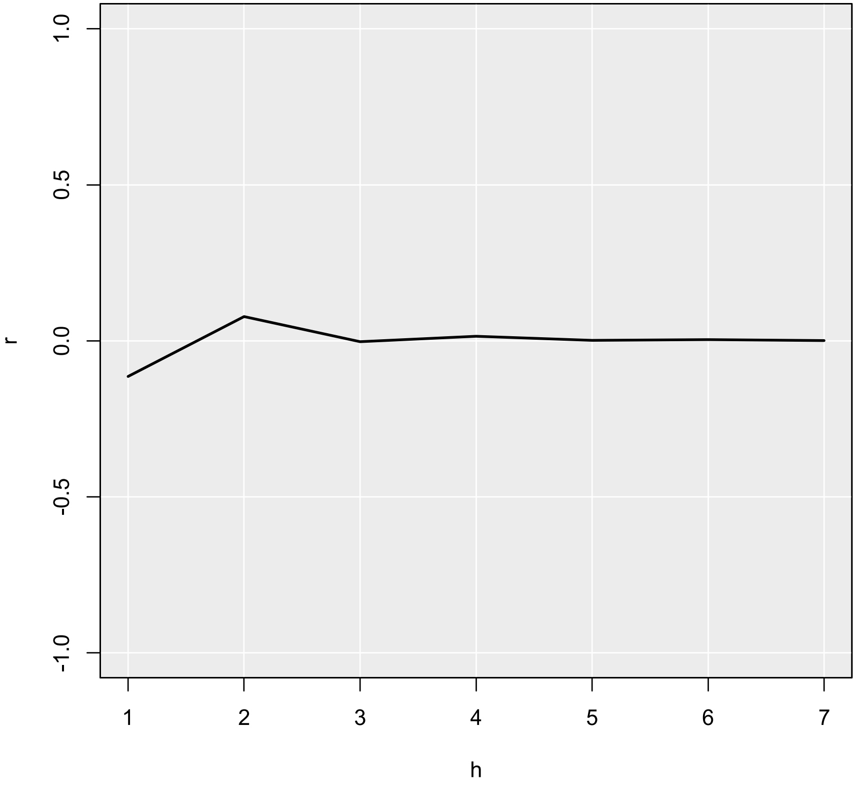
Higher order assortativity: $\boldsymbol{x} = \boldsymbol{s}_{out}, \boldsymbol{y} = \boldsymbol{s}_{in}$





Motivating application: $\boldsymbol{x} = \boldsymbol{s}_{out}, \boldsymbol{y} = \boldsymbol{s}_{in}$





Main references

- Arcagni, A., Grassi, R., Stefani, S., & Torriero, A. (2017). Higher order assortativity in complex networks. *European Journal of Operational Research*, 262(2):708–719
- Arcagni, A., Grassi, R., Stefani, S., & Torriero, A. (2019). Extending assortativity: An application to weighted social networks. *Journal of Business Research*
- Arcagni, A., Cerqueti, R., & Grassi, R. (2023). Higher order assortativity for directed weighted networks and Markov chains. ***arXiv preprint*** *arXiv:2304.01737*
- Newman, M. E. J. (2002). Assortative Mixing in Networks. *Physical Review Letters*, 89(20):208701
- UN COMTRADE (2022). International trade statistics database. Data retrieved from International Trade Statistics Database, https://comtrade.un.org