The Empirical Analysis of Weighted Directed Networks: An Application to the World Trade Web

G. Fagiolo [1] J. Reyes [2] S. Schiavo [3]
giorgio.fagiolo@sssup.it
https://mail.sssup.it/~fagiolo

[1]Sant'Anna School of Advanced Studies, Pisa, Italy

[2]University of Arkansas, USA

[3]University of Trento, Italy

Net Workshop, 12-13 April 2007, Milan (Italy)

My Homepage

https://mail.sssup.it/~fagiolo/welcome.html

Motivations and Goals

- **Illustrating some statistical tools for the analysis of empirically-observed networks**
  - Three simple but important methodological points
  - Leading example: network of trade among World countries

- **Methodological points**
  1. If appropriate, empirical analyses should be carried out in the framework of weighted networks
  2. Often, but not always, directed network analyses should be preferred to undirect ones
  3. The empirical analysis of weighted, directed networks requires new statistical tools

Motivations and Goals

- **Illustrating some statistical tools for the analysis of empirically-observed networks**
  - Three simple but important methodological points
  - Leading example: network of trade among World countries

- **Methodological points**
  1. If appropriate, empirical analyses should be carried out in the framework of weighted networks
  2. Often, but not always, directed network analyses should be preferred to undirect ones
  3. The empirical analysis of weighted, directed networks requires new statistical tools

Background Papers

- Fagiolo, G. (2006a), “Directed or Undirected? A New Index to Check for Directionality of Relations in Socio-Economic Networks", *Economics Bulletin*, 3: 1-12.
- Fagiolo, G. (2006b), “Clustering in Complex Directed Networks”, Working Paper, arXiv:physics/0612169v2.
- Fagiolo, G., Reyes, J. and Schiavo, S. (2007), “The Evolution of the World Trade Web: A Weighted Network Approach”, LEM Working Paper, Sant’Anna School of Advanced Studies, forthcoming.

From Social to Hard Sciences…

- **Networks of interpersonal relations**
  - Old idea in sociology: “relevant others” (Miller, 1963)
  - Explaining patterns of interactions among people of groups
  - Friendship (Rapoport & Horvath, 1961; Milgram, 1967)
  - Marriage (Padgett & Ansell, 1993)
  - Job-market interactions (Granovetter, 1974)

- **Statistical analysis of network topology**
  - Albert & Barabási (2002), Newman (2003), Pastor-Satorras & Vespignani (2004), Dorogovtsev & Mendes (2003), …
  - Properties of real-world technological, biological and information networks
    - WWW and the Internet, peer-to-peer networks, power grids, train routes, airline connections, electronic circuits, metabolism, protein interactions, neural networks

From Social to Hard Sciences. . .

- **Networks of interpersonal relations**
  - Old idea in sociology: “relevant others” (Miller, 1963)
  - Explaining patterns of interactions among people of groups
  - Friendship (Rapoport & Horvath, 1961; Milgram, 1967)
  - Marriage (Padgett & Ansell, 1993)
  - Job-market interactions (Granovetter, 1974)

- **Statistical analysis of network topology**
  - Albert & Barabási (2002), Newman (2003), Pastor-Satorras & Vespignani (2004), Dorogovtsev & Mendes (2003), . . .
  - Properties of real-world technological, biological and information networks
    - WWW and the Internet, peer-to-peer networks, power grids, train routes, airline connections, electronic circuits, metabolism, protein interactions, neural networks

. . . And Back to Social Sciences (Econophysics)

- **Empirical analysis of social and economic networks**
  - Socio-economic systems as networks?
  - A non-exhaustive list of applications
    - scientific co-authorship (Newman, 2001) and citation (Redner, 1998)
    - telephone calls (Aiello et al., 2000)
    - email exchanges (Kossinets and Watts, 2006)
    - sexual relationships (Liljeros et al., 2001)
    - knowledge spill-overs among firms in industrial clusters (Giuliani and Bell, 2001)
    - market investment (Battiston and Catanzaro, 2004)
    - patent citation and innovation networks (Breschi and Lissoni, 2001; Ahuja, 2000)
    - firm alliance formation (Gulati, 1998; Garcia Pont and Nohria, 2002)
    - R&D teams and other within-firm networks (Reagans and co-authors, 2001, 2004)
    - social capital (Walker, Kogut, and Shan, 1997)
    - company ownership and control (Garlaschelli et al., 2005)
    - financial networks (Kullman et al., 2001)
    - bank-firm relationships (De Masi et al., 2007)
    - and also . . .

. . . And Back to Social Sciences (Econophysics)

- **Empirical analysis of social and economic networks**
  - Socio-economic systems as networks?
  - A non-exhaustive list of applications
    - scientific co-authorship (Newman, 2001) and citation (Redner, 1998)
    - telephone calls (Aiello et al., 2000)
    - email exchanges (Kossinets and Watts, 2006)
    - sexual relationships (Liljeros et al., 2001)
    - knowledge spill-overs among firms in industrial clusters (Giuliani and Bell, 2001)
    - market investment (Battiston and Catanzaro, 2004)
    - patent citation and innovation networks (Breschi and Lissoni, 2001; Ahuja, 2000)
    - firm alliance formation (Gulati, 1998; Garcia Pont and Nohria, 2002)
    - R&D teams and other within-firm networks (Reagans and co-authors, 2001, 2004)
    - social capital (Walker, Kogut, and Shan, 1997)
    - company ownership and control (Garlaschelli et al., 2005)
    - financial networks (Kullman et al., 2001)
    - bank-firm relationships (De Masi et al., 2007)
    - and also . . .

The network of international trade flows (1/3)

- **Main idea**
  - Web of trade relations among countries as a networks
  - Countries = nodes
  - Links = existence of trade relationship (import/export)
- **Any value added?**
  - Standard empirics: imports-exports as country-specific variables
  - Network analysis: flows as relational variables
  - Topological structure, higher-order trade structure
  - Network structure and macroeconomic dynamics

The network of international trade flows (1/3)

- **Main idea**
  - Web of trade relations among countries as a networks
  - Countries = nodes
  - Links = existence of trade relationship (import/export)
- **Any value added?**
  - Standard empirics: imports-exports as country-specific variables
  - Network analysis: flows as relational variables
  - Topological structure, higher-order trade structure
  - Network structure and macroeconomic dynamics

The network of international trade flows (2/3)

- **Old tradition in political sciences**
  - Relational variables are more important than country characteristics to explain international trade patterns
  - Focus on core-periphery and world dependency theories
  - Snyder & Kick (1979), Breiger (1981), Nemeth & Smith (1985), Schott (1986), Smith & White (1992), Sacks et al. (2001), Kim & Shin (2002), Kastelle et al. (2006), Mahutga (2006)

- **Econophysics enters the stage**
  - Serrano and Boguña (2003), *Physical Review E*
  - Li, Jin, and Chen (2003), *Physica A*
  - Garlaschelli and Loffredo (2004, 2005), *Physical Review Letters*

The network of international trade flows (2/3)

- **Old tradition in political sciences**
  - Relational variables are more important than country characteristics to explain international trade patterns
  - Focus on core-periphery and world dependency theories
  - Snyder & Kick (1979), Breiger (1981), Nemeth & Smith (1985), Schott (1986), Smith & White (1992), Sacks et al. (2001), Kim & Shin (2002), Kastelle et al. (2006), Mahutga (2006)

- **Econophysics enters the stage**
  - Serrano and Boguña (2003), *Physical Review E*
  - Li, Jin, and Chen (2003), *Physica A*
  - Garlaschelli and Loffredo (2004, 2005), *Physical Review Letters*

The network of international trade flows (3/3)

- **Basic ingredients of the analysis**
  - Take N countries and T time-periods (years)
  - Collect statistics on e_{ij}^t = exports from country i to country j in year $t = 1, \ldots, T$
  - In each t, build a $N \times N$ adjacency matrix A^t, where $a_{ij}^t = 1$ iff $e_{ij}^t > \underline{e}$
  - Critical point: thresholds and GDP scaling

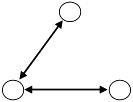
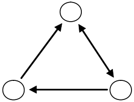
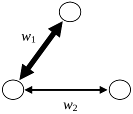
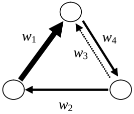
- **Problems**
  - Directed or undirected analysis?
  - Using appropriate tools for directed analysis
  - Disregarding heterogeneity of link importance

The network of international trade flows (3/3)

- **Basic ingredients of the analysis**
  - Take N countries and T time-periods (years)
  - Collect statistics on e_{ij}^t = exports from country i to country j in year $t = 1, \ldots, T$
  - In each t, build a $N \times N$ adjacency matrix A^t, where $a_{ij}^t = 1$ iff $e_{ij}^t > \underline{e}$
  - Critical point: thresholds and GDP scaling
- **Problems**
  - Directed or undirected analysis?
  - Using appropriate tools for directed analysis
  - Disregarding heterogeneity of link importance

A Taxonomy of Network Classes

- **Two Dimensions: Links can be**
  - binary or weighted
  - undirected or undirected

Links	**Undirected**	**Directed**
Binary	 BUN	 BDN
Weighted	 WUN	 WDN

A Taxonomy of Network Classes

- **Two Dimensions: Links can be**
  - binary or weighted
  - undirected or undirected

Links	**Undirected**	**Directed**
Binary	Adjacency Matrix $A = \{a_{ij}\}$ Symmetric: $a_{ij} = a_{ji}$	Adjacency Matrix $A = \{a_{ij}\}$ Asymmetric
Weighted	Weight Matrix $W = \{w_{ij}\}$ Symmetric: $w_{ij} = w_{ji}$	Weight Matrix $W = \{w_{ij}\}$ Asymmetric

Binary Undirected Networks (BUNs)

- **Standard BUN statistics**
  - Node degree (d_i)
  - Average nearest-neighbor degree (*ANND*)
  - Clustering coefficient
  - Average shortest-distance path
  - Betweenness centrality
- **When is a BUN analysis appropriate?**
  - Suppose we can disregard link directionality
  - Nature of relationships must be binary
  - No heterogeneity among links
  - Example: Marriage (Padgett and Ansell, 1993)
  - What about airline traffic, the Internet, scientific citations, the WTW?
  - Intensity or importance of links may strongly differ!

Binary Undirected Networks (BUNs)

- **Standard BUN statistics**
  - Node degree (d_i)
  - Average nearest-neighbor degree (*ANND*)
  - Clustering coefficient
  - Average shortest-distance path
  - Betweenness centrality
- **When is a BUN analysis appropriate?**
  - Suppose we can disregard link directionality
  - Nature of relationships must be binary
  - No heterogeneity among links
  - Example: Marriage (Padgett and Ansell, 1993)
  - What about airline traffic, the Internet, scientific citations, the WTW?
  - Intensity or importance of links may strongly differ!

What if these conditions are not met?

- **Employing a weighted undirected network (WUN) approach**
  - From (symmetric) adjacency matrix to (symmetric) weight matrix
  - Need for a generalization of BUN statistics
  - Barrat et al. (2004, *PNAS*); Barthélemy et al. (2005, *Physica A*)

- **Two crucial necessary conditions**
  1. WUN analysis should bring some value added: heterogeneity must be relevant
  2. Results must not depend on the weighting setup

What if these conditions are not met?

- **Employing a weighted undirected network (WUN) approach**
  - From (symmetric) adjacency matrix to (symmetric) weight matrix
  - Need for a generalization of BUN statistics
  - Barrat et al. (2004, *PNAS*); Barthélemy et al. (2005, *Physica A*)

- **Two crucial necessary conditions**
  1. WUN analysis should bring some value added: heterogeneity must be relevant
  2. Results must not depend on the weighting setup

WUN Indicators: A brief tutorial (1/2)

- **Node Strength**

$$s_i = \sum_j w_{ij} = W_{(i)}\mathbf{1}.$$

- **Average Nearest-Neighbors Strength (ANNS)**

$$anns_i = d_i^{-1} \sum_j a_{ij} s_j = d_i^{-1} \sum_j \sum_h a_{ij} w_{jh} = \frac{A_{(i)} W \mathbf{1}}{A_{(i)} \mathbf{1}}.$$

- **Weighted Average of Nearest-Neighbors Degree (WANND)**

$$wannd_i = s_i^{-1} \sum_j w_{ij} d_j = s_i^{-1} \sum_j \sum_h w_{ij} a_{jh} = \frac{W_{(i)} A \mathbf{1}}{W_{(i)} \mathbf{1}}.$$

WUN Indicators: A brief tutorial (1/2)

- **Node Strength**

$$s_i = \sum_j w_{ij} = W_{(i)}\mathbf{1}.$$

- **Average Nearest-Neighbors Strength (ANNS)**

$$anns_i = d_i^{-1} \sum_j a_{ij} s_j = d_i^{-1} \sum_j \sum_h a_{ij} w_{jh} = \frac{A_{(i)} W\mathbf{1}}{A_{(i)}\mathbf{1}}.$$

- **Weighted Average of Nearest-Neighbors Degree (WANND)**

$$wannd_i = s_i^{-1} \sum_j w_{ij} d_j = s_i^{-1} \sum_j \sum_h w_{ij} a_{jh} = \frac{W_{(i)} A\mathbf{1}}{W_{(i)}\mathbf{1}}.$$

WUN Indicators: A brief tutorial (1/2)

- **Node Strength**

$$s_i = \sum_j w_{ij} = W_{(i)}\mathbf{1}.$$

- **Average Nearest-Neighbors Strength (ANNS)**

$$anns_i = d_i^{-1} \sum_j a_{ij} s_j = d_i^{-1} \sum_j \sum_h a_{ij} w_{jh} = \frac{A_{(i)} W\mathbf{1}}{A_{(i)}\mathbf{1}}.$$

- **Weighted Average of Nearest-Neighbors Degree (WANND)**

$$wannd_i = s_i^{-1} \sum_j w_{ij} d_j = s_i^{-1} \sum_j \sum_h w_{ij} a_{jh} = \frac{W_{(i)} A\mathbf{1}}{W_{(i)}\mathbf{1}}.$$

WUN Indicators: A brief tutorial (2/2)

- **Node Disparity (Herfindahl Concentration Index)**

$$h_i = \frac{(N-1)\sum_j \left(\frac{w_{ij}}{s_i}\right)^2 - 1}{N-2} = \frac{(N-1)\frac{1}{s_i^2}\sum_j w_{ij}^2 - 1}{N-2} = \frac{(N-1)\frac{W_{(i)}^{[2]}\mathbf{1}}{(W_{(i)}\mathbf{1})^2} - 1}{N-2}$$

- **Binary Clustering Coefficient (CC)**

$$C_i(A) = \frac{\frac{1}{2}\sum_{j\neq i}\sum_{h\neq(i,j)} a_{ij}a_{ih}a_{jh}}{\frac{1}{2}d_i(d_i-1)} = \frac{(A^3)_{ii}}{d_i(d_i-1)}.$$

- **Weighted Clustering Coefficient (WCC)**

$$\tilde{C}_i(W) = \frac{\frac{1}{2}\sum_{j\neq i}\sum_{h\neq(i,j)} w_{ij}^{\frac{1}{3}} w_{ih}^{\frac{1}{3}} w_{jh}^{\frac{1}{3}}}{\frac{1}{2}d_i(d_i-1)} = \frac{(W^{[\frac{1}{3}]})^3_{ii}}{d_i(d_i-1)},$$

WUN Indicators: A brief tutorial (2/2)

- **Node Disparity (Herfindahl Concentration Index)**

$$h_i = \frac{(N-1)\sum_j \left(\frac{w_{ij}}{s_i}\right)^2 - 1}{N-2} = \frac{(N-1)\frac{1}{s_i^2}\sum_j w_{ij}^2 - 1}{N-2} = \frac{(N-1)\frac{W_{(i)}^{[2]}\mathbf{1}}{(W_{(i)}\mathbf{1})^2} - 1}{N-2}$$

- **Binary Clustering Coefficient (CC)**

$$C_i(A) = \frac{\frac{1}{2}\sum_{j\neq i}\sum_{h\neq(i,j)} a_{ij}a_{ih}a_{jh}}{\frac{1}{2}d_i(d_i-1)} = \frac{(A^3)_{ii}}{d_i(d_i-1)}.$$

- **Weighted Clustering Coefficient (WCC)**

$$\tilde{C}_i(W) = \frac{\frac{1}{2}\sum_{j\neq i}\sum_{h\neq(i,j)} w_{ij}^{\frac{1}{3}} w_{ih}^{\frac{1}{3}} w_{jh}^{\frac{1}{3}}}{\frac{1}{2}d_i(d_i-1)} = \frac{(W^{[\frac{1}{3}]})^3_{ii}}{d_i(d_i-1)},$$

WUN Indicators: A brief tutorial (2/2)

- **Node Disparity (Herfindahl Concentration Index)**

$$h_i = \frac{(N-1)\sum_j \left(\frac{w_{ij}}{s_i}\right)^2 - 1}{N-2} = \frac{(N-1)\frac{1}{s_i^2}\sum_j w_{ij}^2 - 1}{N-2} = \frac{(N-1)\frac{W_{(i)}^{[2]}\mathbf{1}}{(W_{(i)}\mathbf{1})^2} - 1}{N-2}$$

- **Binary Clustering Coefficient (CC)**

$$C_i(A) = \frac{\frac{1}{2}\sum_{j\neq i}\sum_{h\neq(i,j)} a_{ij}a_{ih}a_{jh}}{\frac{1}{2}d_i(d_i-1)} = \frac{(A^3)_{ii}}{d_i(d_i-1)}.$$

- **Weighted Clustering Coefficient (WCC)**

$$\tilde{C}_i(W) = \frac{\frac{1}{2}\sum_{j\neq i}\sum_{h\neq(i,j)} w_{ij}^{\frac{1}{3}} w_{ih}^{\frac{1}{3}} w_{jh}^{\frac{1}{3}}}{\frac{1}{2}d_i(d_i-1)} = \frac{(W^{[\frac{1}{3}]})^3_{ii}}{d_i(d_i-1)},$$

Data

- **International trade data**
  - Gleditsch (2002) database
  - See http://ibs.colorado.edu/~ksg/trade/
- **Data structure**
  - We employ a panel of 159 countries
  - Time period: 1981-2000 ($T = 20$ years)
  - Baseline observation
    - $\{e_{ij}^t\}$: Exports from country i to country j in year t
    - GDP_i^t and $pcGDP_i^t$ of country i in year t
  - Data in current US$ (deflated)
- **Important remark**
  - Suppose export flows are sufficiently symmetric. . .

Data

- **International trade data**
  - Gleditsch (2002) database
  - See `http://ibs.colorado.edu/~ksg/trade/`
- **Data structure**
  - We employ a panel of 159 countries
  - Time period: 1981-2000 ($T = 20$ years)
  - Baseline observation
    - $\{e_{ij}^t\}$: Exports from country i to country j in year t
    - GDP_i^t and $pcGDP_i^t$ of country i in year t
  - Data in current US$ (deflated)
- **Important remark**
  - Suppose export flows are sufficiently symmetric…

Data

- **International trade data**
  - Gleditsch (2002) database
  - See `http://ibs.colorado.edu/~ksg/trade/`
- **Data structure**
  - We employ a panel of 159 countries
  - Time period: 1981-2000 ($T = 20$ years)
  - Baseline observation
    - $\{e_{ij}^t\}$: Exports from country i to country j in year t
    - GDP_i^t and $pcGDP_i^t$ of country i in year t
  - Data in current US$ (deflated)
- **Important remark**
  - Suppose export flows are sufficiently symmetric. . .

Links and Weights

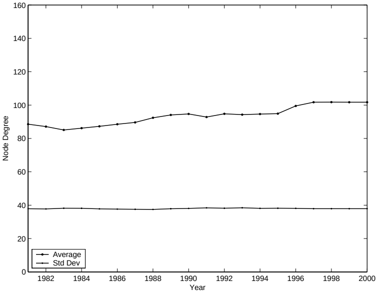
- **Adjacency matrix A_i^t**
  - We follow commodity flow (rows: exporters)
  - A link ij exists if $e_{ij}^t = e_{ji}^t > 0$
- **Weight matrix W_i^t**
  - We use the baseline definition
    - Exports from i to j divided by exporter's GDP (i)
  - But we experiment with many alternatives:
    - Exports from i to j divided by importer's GDP (j)
    - Exports from i to j (not scaled)
    - Same as above but now divided by total exports
  - Weights are renormalized s.t. $w_i^t \in [0, 1]$

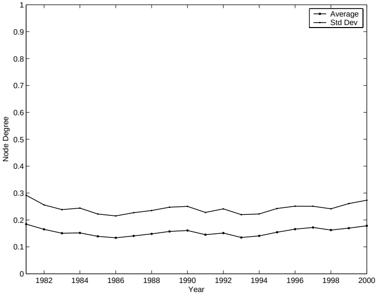
Links and Weights

- **Adjacency matrix** A_i^t
  - We follow commodity flow (rows: exporters)
  - A link *ij* exists if $e_{ij}^t = e_{ji}^t > 0$
- **Weight matrix** W_i^t
  - We use the baseline definition
    - Exports from *i* to *j* divided by exporter's GDP (*i*)
  - But we experiment with many alternatives:
    - Exports from *i* to *j* divided by importer's GDP (*j*)
    - Exports from *i* to *j* (not scaled)
    - Same as above but now divided by total exports
  - Weights are renormalized s.t. $w_i^t \in [0, 1]$

WTW Connectivity: Average and Standard Deviation

Highly connected BUN vs. Weakly connected WUN



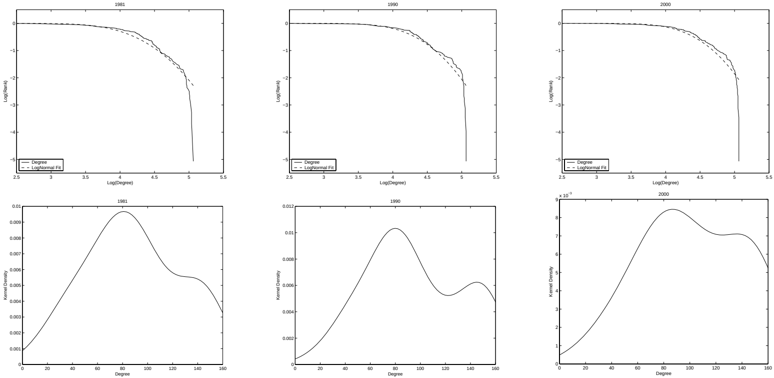


Node Degree (Ave/StDev)

Node Strength (Ave/StDev)

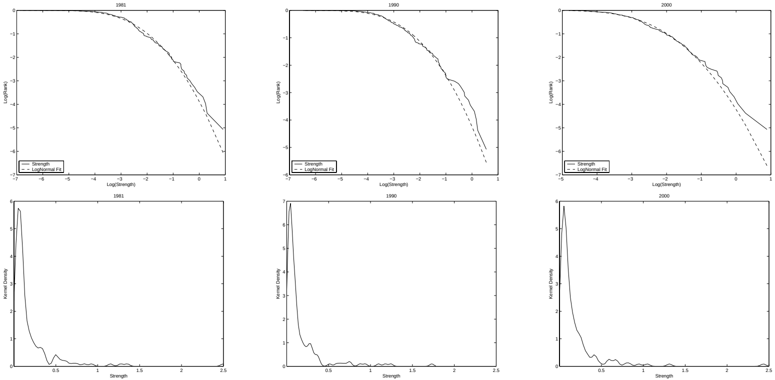
WTW Connectivity: Shape of Degree Distributions

Weak skewness; not lognormal/power-law; bimodality



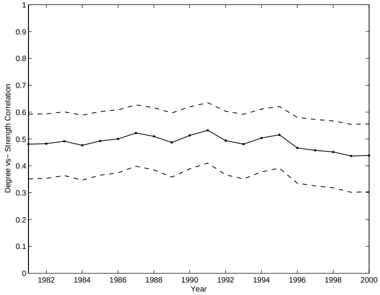
WTW Connectivity: Shape of Strength Distributions

High skewness; more lognormal/power-law; no bimodality

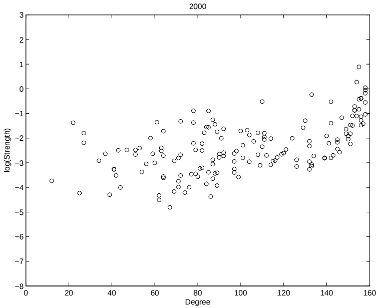


WTW Connectivity: Degree-Strength Correlation

Positive but not very strong correlation



Deg-Str Correlation



Deg vs. Str in t=2000

Remark: WUN analysis, a first value added

- **WUN Connectivity**
  - A picture substantially different from BUN
  - Trade link heterogeneity matters

- **Degree-Strength Distributions**
  - Degree: Bimodality
  - Strength: Skewed distributions, quasi scale-free, core-periphery structure

Remark: WUN analysis, a first value added

- **WUN Connectivity**
  - A picture substantially different from BUN
  - Trade link heterogeneity matters
- **Degree-Strength Distributions**
  - Degree: Bimodality
  - Strength: Skewed distributions, quasi scale-free, core-periphery structure

WTW Assortativity

- **Do strongly-connected countries trade with strongly-connected partners?**
  - In terms of node degree (BUN) and ANND
  - In terms of node strength (WUN) and ANNS-WANND
- **Networks can be**
  - Assortative: Positive correlation
  - Disassortative: Negative correlation

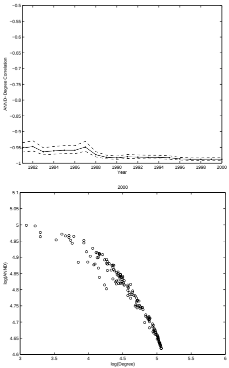
WTW Assortativity

- **Do strongly-connected countries trade with strongly-connected partners?**
  - In terms of node degree (BUN) and ANND
  - In terms of node strength (WUN) and ANNS-WANND
- **Networks can be**
  - Assortative: Positive correlation
  - Disassortative: Negative correlation

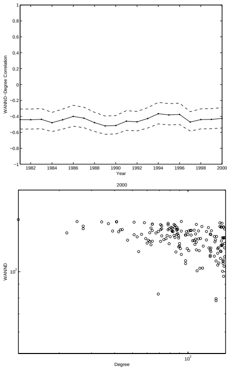
WTW Assortativity: Correlation patterns

BUN: Strongly disassortative; WUN: Weakly disassortative

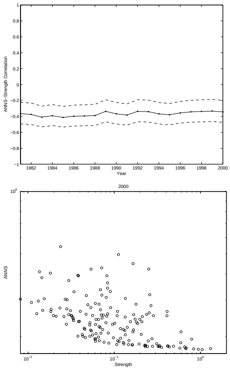
ANND-DEG



WANND-DEG



ANNS-STR



Remark: WUN analysis, a second value added

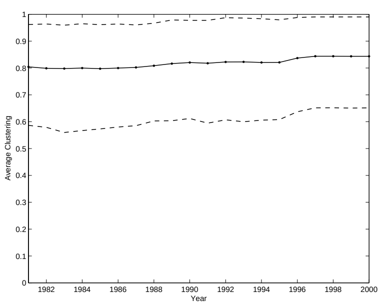
- **WUN Connectivity**
  - Core-periphery (quasi scale-free) structure
- **WUN Assortativity**
  - Poorly-connected trading with highly-connected
  - But: Emergence of intermediate periphery
  - Medium-highly connected trade with highly-connected

Remark: WUN analysis, a second value added

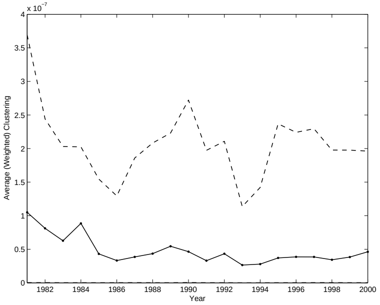
- **WUN Connectivity**
  - Core-periphery (quasi scale-free) structure
- **WUN Assortativity**
  - Poorly-connected trading with highly-connected
  - But: Emergence of intermediate periphery
  - Medium-highly connected trade with highly-connected

WTW Clustering: Average Levels

Do countries hold many/intense trade relationships with countries that intensively trade with each other?



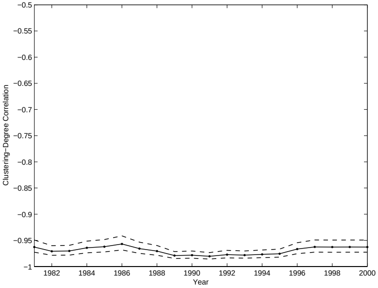
Average Node CC (Expected $\sim$ 0.6)



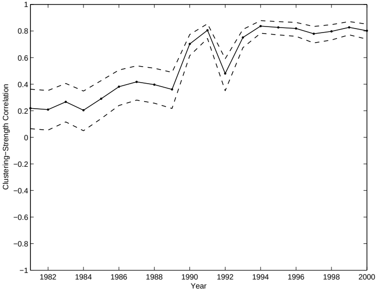
Average Node WCC (Expected $\sim$ 0.4)

WTW Clustering: Correlation with Degree/Strength

Are better connected countries more clustered?



BUN, CC-Degree: Strongly Negative



BUN, WCC-Strength: Positive

Remark: WUN analysis, a third value added

- **BUN Clustering**
  - Highly clustered on average
  - Highly-connected countries hold trade partners that do not trade with each other
  - Poorly-connected countries do not trade among them but are connected to the hubs
- **WUN Clustering**
  - Poorly clustered on average
  - Countries holding intense trade relationships are typically involved in highly-interconnected triples
  - A sort of “rich club phenomenon”?

Remark: WUN analysis, a third value added

- **BUN Clustering**
  - Highly clustered on average
  - Highly-connected countries hold trade partners that do not trade with each other
  - Poorly-connected countries do not trade among them but are connected to the hubs
- **WUN Clustering**
  - Poorly clustered on average
  - Countries holding intense trade relationships are typically involved in highly-interconnected triples
  - A sort of "rich club phenomenon"?

Binary vs. Weighted Analysis: Summary

- **We have shown that**
  - Link heterogeneity matters a lot in studying WTW
  - If link heterogeneity is not taken into account we are disregarding a lot of information: a very different picture may emerge
  - A WUN approach is able to provide more and better insights
- **Do results depend on weighting setup?**
  - Not at all!
  - All previous results hold under alternative weighting schemes
  - If we do not scale by GDP, larger positive correlation between strength and WCC!

Binary vs. Weighted Analysis: Summary

- **We have shown that**
  - Link heterogeneity matters a lot in studying WTW
  - If link heterogeneity is not taken into account we are disregarding a lot of information: a very different picture may emerge
  - A WUN approach is able to provide more and better insights
- **Do results depend on weighting setup?**
  - Not at all!
  - All previous results hold under alternative weighting schemes
  - If we do not scale by GDP, larger positive correlation between strength and WCC!

What happens if networks are directed?

- **Maintained Assumption: Undirected Networks**
  - Majority of real-world networks are intrinsically directed
  - Example: WTW
  - Directed or undirected analysis?
  - General Rule
    - Undirected: If network is intrinsically symmetric (marriage)
    - Directed: Must statistically detect if empirically-observed network is sufficiently asymmetric
- **Why can't we simply employ a directed analysis?**
  - If network is directed (binary or weighted) appropriate tools and indicators must be employed
  - Many papers: analyze directed networks with undirected-network tools
  - Directed networks indicators are more complicated but they extract much more information

What happens if networks are directed?

- **Maintained Assumption: Undirected Networks**
  - Majority of real-world networks are intrinsically directed
  - Example: WTW
  - Directed or undirected analysis?
  - General Rule
    - Undirected: If network is intrinsically symmetric (marriage)
    - Directed: Must statistically detect if empirically-observed network is sufficiently asymmetric
- **Why can't we simply employ a directed analysis?**
  - If network is directed (binary or weighted) appropriate tools and indicators must be employed
  - Many papers: analyze directed networks with undirected-network tools
  - Directed networks indicators are more complicated but they extract much more information

WDN Indicators: Strength and ANNS

- **In, Out, Total Strength**
  - $s_i^{in} = \sum_j w_{ji} = W_{(i)}^T \mathbf{1}$
  - $s_i^{out} = \sum_j w_{ij} = W_{(i)} \mathbf{1}$
  - $s_i^{tot} = s_i^{in} + s_i^{out} = (W^T + W)_{(i)} \mathbf{1}$
- **Average Nearest-Neighbor Strength**
  - $anns_i^{out-out}$: Average out-strength of i's out-neighbors
  - $anns_i^{out-in}$: Average in-strength of i's out-neighbors
  - $anns_i^{in-out}$: Average out-strength of i's in-neighbors
  - $anns_i^{in-in}$: Average in-strength of i's in-neighbors

WDN Indicators: Strength and ANNS

- **In, Out, Total Strength**
  - $s_i^{in} = \sum_j w_{ji} = W_{(i)}^T \mathbf{1}$
  - $s_i^{out} = \sum_j w_{ij} = W_{(i)} \mathbf{1}$
  - $s_i^{tot} = s_i^{in} + s_i^{out} = (W^T + W)_{(i)} \mathbf{1}$
- **Average Nearest-Neighbor Strength**
  - $anns_i^{out-out}$: Average out-strength of i's out-neighbors
  - $anns_i^{out-in}$: Average in-strength of i's out-neighbors
  - $anns_i^{in-out}$: Average out-strength of i's in-neighbors
  - $anns_i^{in-in}$: Average in-strength of i's in-neighbors

WDN Indicators: Strength and ANNS

- **More formally**
  - $anns_i^{out-out} = (d_i^{out})^{-1} \sum_j a_{ij} s_j^{out} = \frac{A_{(i)} W^T \mathbf{1}}{A_i \mathbf{1}}$
  - $anns_i^{out-in} = (d_i^{out})^{-1} \sum_j a_{ij} s_j^{in} = \frac{A_{(i)} W \mathbf{1}}{A_i \mathbf{1}}$
  - $anns_i^{in-out} = (d_i^{in})^{-1} \sum_j a_{ji} s_j^{out} = \frac{A_{(i)}^T W^T \mathbf{1}}{A_i^T \mathbf{1}}$
  - $anns_i^{in-in} = (d_i^{in})^{-1} \sum_j a_{ji} s_j^{in} = \frac{A_{(i)}^T W \mathbf{1}}{A_i^T \mathbf{1}}$
- **Decomposition of Total ANNS**

$$anns^{tot} = \frac{(A^T + A)_{(i)}(W^T + W)\mathbf{1}}{(A^T + A)_{(i)}\mathbf{1}}$$

$$(A^T + A)_{(i)}(W^T + W) = A_{(i)}^T W^T + A_{(i)}^T W + A_{(i)} W^T + A_{(i)} W.$$

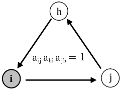
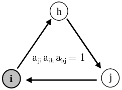
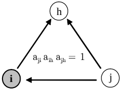
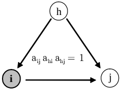
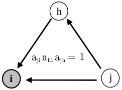
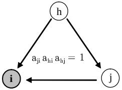
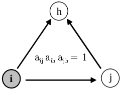
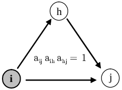
WDN Indicators: Strength and ANNS

- **More formally**
  - $anns_i^{out-out} = (d_i^{out})^{-1} \sum_j a_{ij} s_j^{out} = \frac{A_{(i)} W^T \mathbf{1}}{A_i \mathbf{1}}$
  - $anns_i^{out-in} = (d_i^{out})^{-1} \sum_j a_{ij} s_j^{in} = \frac{A_{(i)} W \mathbf{1}}{A_i \mathbf{1}}$
  - $anns_i^{in-out} = (d_i^{in})^{-1} \sum_j a_{ji} s_j^{out} = \frac{A_{(i)}^T W^T \mathbf{1}}{A_i^T \mathbf{1}}$
  - $anns_i^{in-in} = (d_i^{in})^{-1} \sum_j a_{ji} s_j^{in} = \frac{A_{(i)}^T W \mathbf{1}}{A_i^T \mathbf{1}}$
- **Decomposition of Total ANNS**

$$anns^{tot} = \frac{(A^T + A)_{(i)} (W^T + W) \mathbf{1}}{(A^T + A)_{(i)} \mathbf{1}}$$

$$(A^T + A)_{(i)} (W^T + W) = A_{(i)}^T W^T + A_{(i)}^T W + A_{(i)} W^T + A_{(i)} W.$$

Clustering in WDNs (Fagiolo, 2006b)

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		$\frac{(A+A^T)^3_{ii}}{2}$	$d_i^{tot}(d_i^{tot} - 1) - 2d_i^{\leftrightarrow}$	$C_i^D = \frac{(A+A^T)^3_{ii}}{2T_i^D}$	$\tilde{C}_i^D = \frac{(\hat{W}+\hat{W}^T)^3_{ii}}{2T_i^D}$

Two crucial issues

- **How can we decide whether a directed analysis should be preferred?**
  - Computing percentage of reciprocated links
  - Correlation between upper and lower diagonal entries (see Garlaschelli and Loffredo, 2004, *Physical Review Letters*)
  - More robust statistical checks?
- **Values added of a directed analysis?**
  - As happens for WUNs, we must show that a directed analysis, when appropriate, adds something to our understanding of the properties of the network under study

Two crucial issues

- **How can we decide whether a directed analysis should be preferred?**
  - Computing percentage of reciprocated links
  - Correlation between upper and lower diagonal entries (see Garlaschelli and Loffredo, 2004, *Physical Review Letters*)
  - More robust statistical checks?
- **Values added of a directed analysis?**
  - As happens for WUNs, we must show that a directed analysis, when appropriate, adds something to our understanding of the properties of the network under study

A Simple Index (Fagiolo, 2006a)

- **Idea**
  - The more the network is undirected, the smaller $\|\tilde{W} - \tilde{W}^T\|$ (appropriately normalized)
- **Technical assumption**

$$Q = \{q_{ij}\} = \tilde{W} - (1 - \tilde{W})I_N$$

- **Define**

$$\widetilde{S}(Q) = \frac{\|Q - Q^T\|_F^2}{\|Q\|_F^2 + \|Q^T\|_F^2} = \frac{\|Q - Q^T\|_F^2}{2\|Q\|_F^2} = \frac{1}{2}\left[\frac{\|Q - Q^T\|_F}{\|Q\|_F}\right]^2$$

A Simple Index (Fagiolo, 2006a)

- **Idea**
  - The more the network is undirected, the smaller $\|\tilde{W} - \tilde{W}^T\|$ (appropriately normalized)
- **Technical assumption**

$$Q = \{q_{ij}\} = \tilde{W} - (1 - \tilde{W})I_N$$

- **Define**

$$\widetilde{S}(Q) = \frac{\|Q - Q^T\|_F^2}{\|Q\|_F^2 + \|Q^T\|_F^2} = \frac{\|Q - Q^T\|_F^2}{2\|Q\|_F^2} = \frac{1}{2}\left[\frac{\|Q - Q^T\|_F}{\|Q\|_F}\right]^2$$

A Simple Index (Fagiolo, 2006a)

- **Idea**
  - The more the network is undirected, the smaller $\|\tilde{W} - \tilde{W}^T\|$ (appropriately normalized)
- **Technical assumption**

$$Q = \{q_{ij}\} = \tilde{W} - (1 - \tilde{W})I_N$$

- **Define**

$$\widetilde{S}(Q) = \frac{\|Q - Q^T\|_F^2}{\|Q\|_F^2 + \|Q^T\|_F^2} = \frac{\|Q - Q^T\|_F^2}{2\|Q\|_F^2} = \frac{1}{2}\left[\frac{\|Q - Q^T\|_F}{\|Q\|_F}\right]^2$$

A Simple Index (Fagiolo, 2006a)

- **Expanding the sums...**

$$\widetilde{S}(Q) = 1 - \frac{\sum_i \sum_j q_{ij} q_{ji}}{\sum_i \sum_j q_{ij}^2}.$$

- **To get an index in $[0, 1]$, define:**

$$S(Q) = \frac{N+1}{N-1} \widetilde{S}(Q),$$

- **We can find $(m_W(N), s_W(N))$ such that:**

$$S_W(Q) = \frac{S(Q) - m_W(N)}{s_W(N)} \sim N(0, 1)$$

A Simple Index (Fagiolo, 2006a)

- **Expanding the sums...**

$$\widetilde{S}(Q) = 1 - \frac{\sum_i \sum_j q_{ij} q_{ji}}{\sum_i \sum_j q_{ij}^2}.$$

- **To get an index in $[0, 1]$, define:**

$$S(Q) = \frac{N+1}{N-1}\widetilde{S}(Q),$$

- **We can find $(m_W(N), s_W(N))$ such that:**

$$S_W(Q) = \frac{S(Q) - m_W(N)}{s_W(N)} \sim N(0, 1)$$

A Simple Index (Fagiolo, 2006a)

- **Expanding the sums...**

$$\widetilde{S}(Q) = 1 - \frac{\sum_i \sum_j q_{ij} q_{ji}}{\sum_i \sum_j q_{ij}^2}.$$

- **To get an index in $[0, 1]$, define:**

$$S(Q) = \frac{N+1}{N-1} \widetilde{S}(Q),$$

- **We can find $(m_W(N), s_W(N))$ such that:**

$$S_W(Q) = \frac{S(Q) - m_W(N)}{s_W(N)} \sim N(0, 1)$$

A Simple Index (Fagiolo, 2006a)

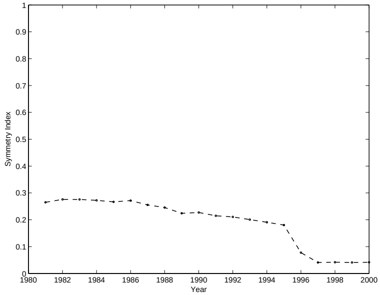
- **Using the index**
  - Define $Q = \{q_{ij}\} = \tilde{W} - (1 - \tilde{W})I_N$
  - Compute the index $S_W(Q)$
  - Fix a threshold k (in term of standard deviations)
  - If $S_W(Q) > k$ the graph is asymmetric

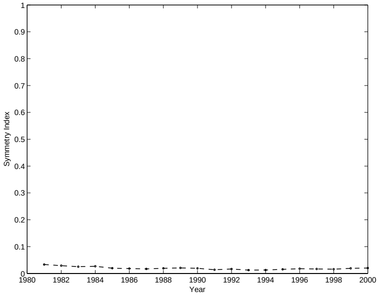
TABLE I: The index S and its standardized version $S_{\{*\}}, \{*\} = \{B(inary), W(eigthed)\}$ for social networks studied in [3], cf. Chapter 2.5.

	Social Network	N	S	S_*
1	Advice relations btw Krackhardt's hi-tech managers	21	0.521327	0.491228
2	Friendship relations btw Krackhardt's hi-tech managers	21	0.500813	0.004610
3	"Reports-to" relations btw Krackhardt's hi-tech managers	21	0.536585	0.860033
4	Business relationships btw Padgett's Florentine families	16	0.000000	-9.232823
5	Marital relationships btw Padgett's Florentine families	16	0.000000	-9.232823
6	Acquaintanceship among Freeman's EIES researchers (Time 1)	32	0.109849	-10.025880
7	Acquaintanceship among Freeman's EIES researchers (Time 2)	32	0.094968	-11.143250
8	Messages sent among Freeman's EIES researchers	32	0.014548	-17.181580
9	Country Trade Flows: Basic Manufactured Goods	24	0.260349	-6.643695
10	Country Trade Flows: Food and Live Animals	24	0.311966	-5.217508
11	Country Trade Flows: Crude Materials (excl. Food)	24	0.272560	-6.306300
12	Country Trade Flows: Minerals, Fuels, Petroleum	24	0.403336	-2.692973
13	Country Trade Flows: Exchange of Diplomats	24	0.080208	-11.620970

Results for the WTW

- **The WTW is extremely symmetric**
  - Binary vs. weighted: no differences
  - Symmetric under all weighting schemes
  - Procedure employed above was appropriate





WTW: A Weighted Directed Network Analysis

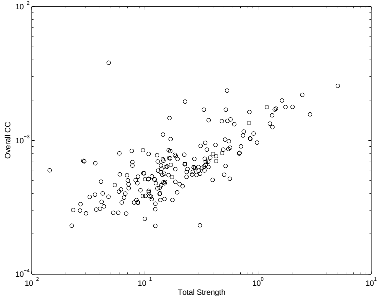
- **Is it worthwhile anyway?**
  - Despite strong symmetry, does a WDN bring any value added?
- **Clustering associated to different triangle patterns**
  - Triangles and their meaning in terms of export/import
  - Heterogeneity: CC ranges from 0.0004 to 0.0013
  - Cycles only 18% of all triangles, other 27% each; due to economic redundancy?

WTW: A Weighted Directed Network Analysis

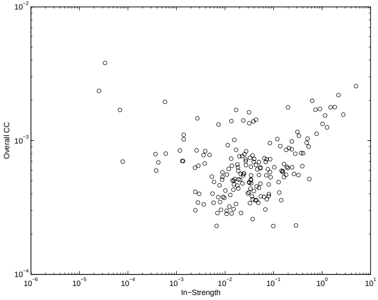
- **Is it worthwhile anyway?**
  - Despite strong symmetry, does a WDN bring any value added?
- **Clustering associated to different triangle patterns**
  - Triangles and their meaning in terms of export/import
  - Heterogeneity: CC ranges from 0.0004 to 0.0013
  - Cycles only 18% of all triangles, other 27% each; due to economic redundancy?

WTW: Clustering-Strength Correlation (1/2)

- **How do Directed WCC correlate with strength?**
  - WCC vs. tot-strength: positive correlation
  - WCC vs. in-strength: U-shaped
  - Low clustering level of weakly connected countries mainly depends on their weakly exporting relationships



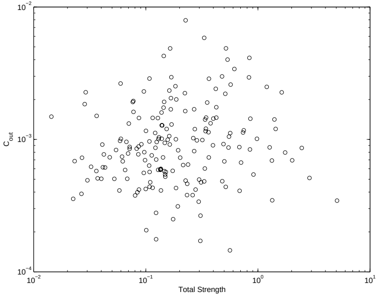
Tot-Str



In-Str

WTW: Clustering-Strength Correlation (2/2)

- **WCC for different triangles vs. strength**
  - WCC for cyc, mid, in are positively correlated with strength
  - WCC for out not correlated with strength
  - Countries hold exporting relationships with connected pairs of countries independently of total strength



Directed vs. Undirected Analysis

- **We have shown that**
  - Apart from extreme cases, deciding whether to employ a directed or an undirected analysis is an empirical issue
  - It is possible to introduce an index to check for network symmetry/asymmetry
  - This index is not a hypothesis test but has nice statistical properties
- **Values added of a directed network analysis**
  - Even in the case the network looks extremely symmetric (as in the WTW case), a directed analysis can provide interesting insights
  - Need to extend network indicators to the WDN case!
  - Example: Betweenness centrality

Directed vs. Undirected Analysis

- **We have shown that**
  - Apart from extreme cases, deciding whether to employ a directed or an undirected analysis is an empirical issue
  - It is possible to introduce an index to check for network symmetry/asymmetry
  - This index is not a hypothesis test but has nice statistical properties
- **Values added of a directed network analysis**
  - Even in the case the network looks extremely symmetric (as in the WTW case), a directed analysis can provide interesting insights
  - Need to extend network indicators to the WDN case!
  - Example: Betweenness centrality