Characterizing System Architectures Using Network Data

Zhenglin Wei

George Washington University

2019

CSER

conference on systems engineering research

Characterizing System Architectures Using Network Data

By

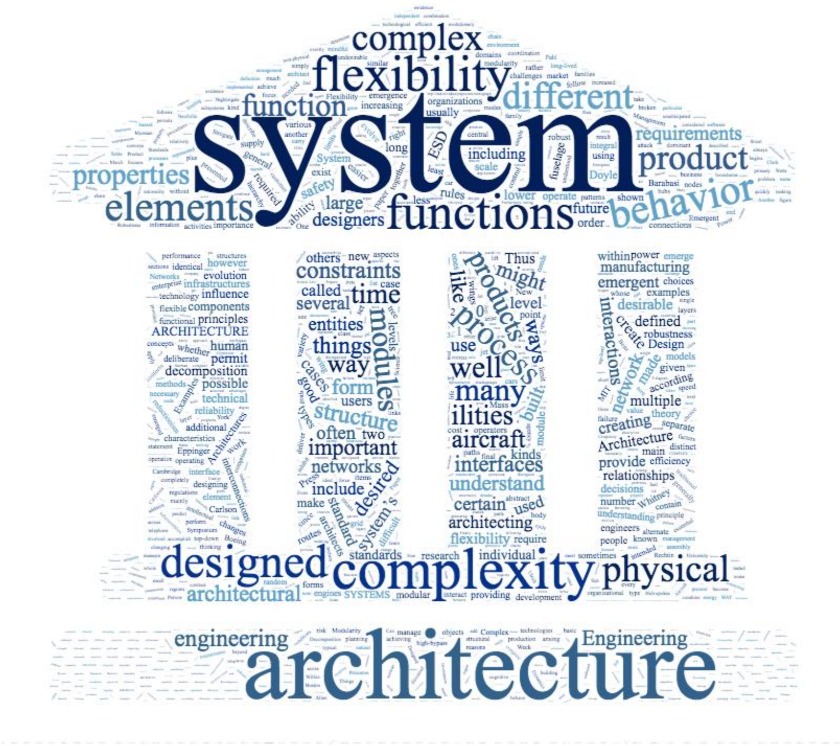
Zhenglin Wei and David A. Broniatowski

17th Annual Conference on Systems Engineering Research

April 4th, 2019

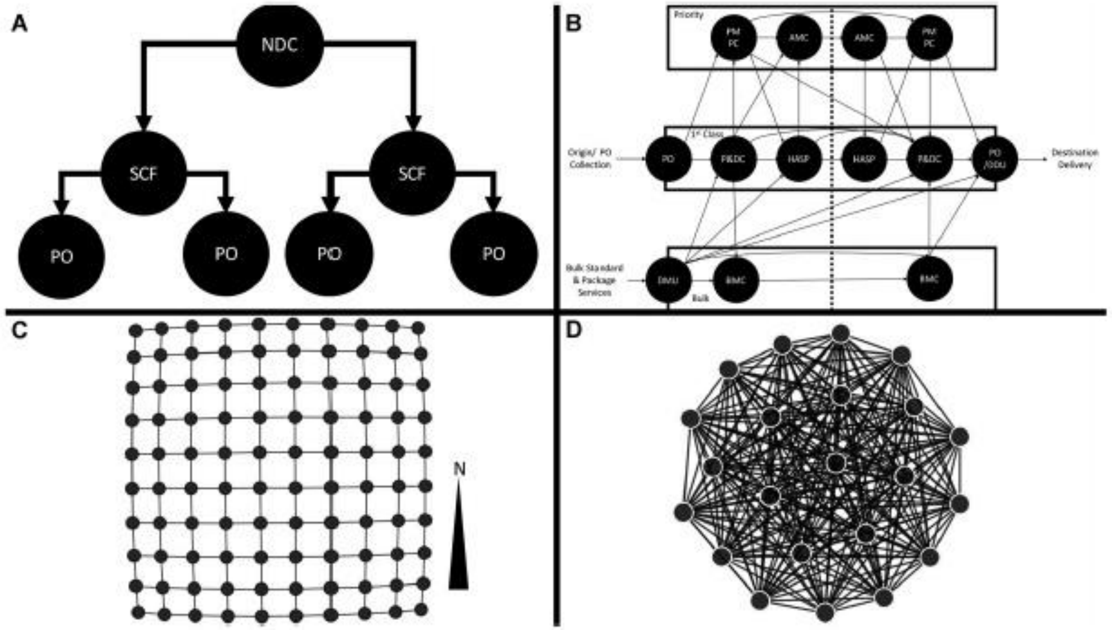
Washington DC

Purpose



- Systems' architectures drive their lifecycle properties (e.g. flexibility, controllability etc.)
- However, measuring the architectures of real complex systems remains a challenge
- Characterize network representations of real-world systems in multiple contexts by their *laterality* and *verticality*, introduced in the generic architecture theory
- Propose a new set of metrics to measure the concordance of real-world systems with four generic structures

Generic Architecture Theory



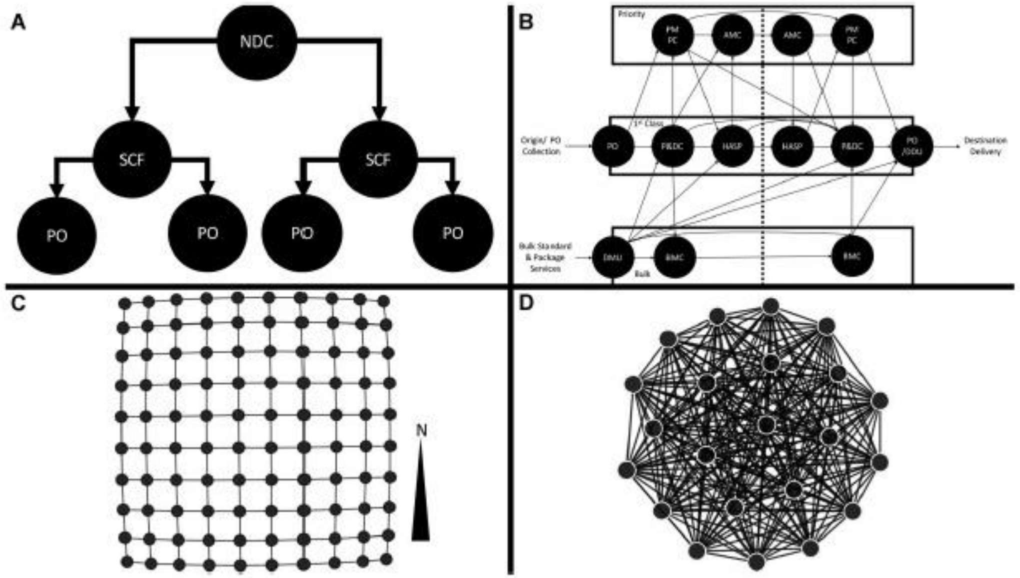
A: Tree Structure
B: Layered Structure
C: Grid Structure
D: Team Structure

"Generic" architectures proposed and developed by J. Moses and D. Broniatowski

- If we understand the behavior of the "generic" architecture of the system
- Then the real world systems can be treated as mixture of these systems in different ways

Broniatowski, D. A., & Moses, J. (2016). Measuring Flexibility, Descriptive Complexity, and Rework Potential in Generic System Architectures: METRICS FOR GENERIC SYSTEM ARCHITECTURES. *Systems Engineering*, *19*(3), 207–221. https://doi.org/10.1002/sys.21351

Generic Architecture Theory



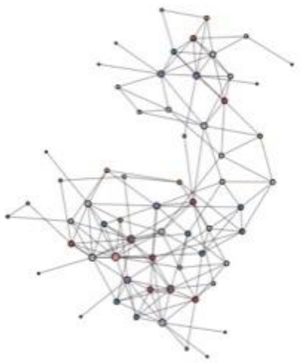
- Node: elements/components of the system
  - Entities, Decision Makers etc.
- Edges: interactions between nodes
  - Information Flow, decision alternatives

A: Tree Structure
B: Layered Structure
C: Grid Structure
D: Team Structure

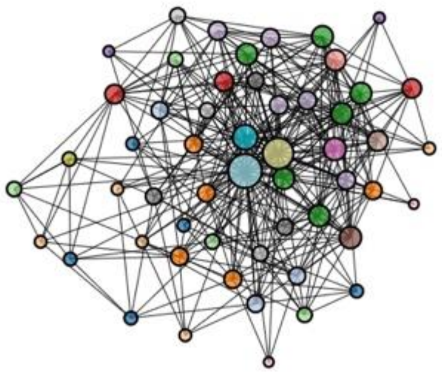
- Verticality: Hierarchy
- Laterality: Interconnectedness

System Structure(Network) Data

- Using topological characteristics to measure laterality and verticality in real-world systems (represented as directed networks)
- Data: 67 networks representing systems in different contexts are collected from different sources

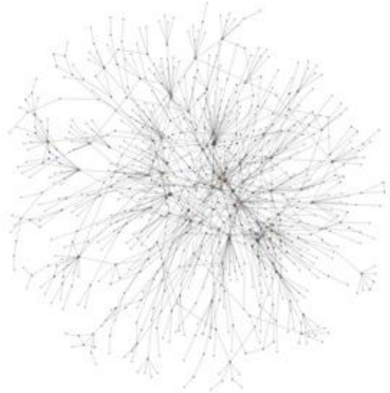


SOC-DOLPINES
N/E: 62/159
Source: UCINet



ECO-STMARKS
(FoodWeb)
N/E: 54/353
Source: UCINet

ROAD-EU
N/E: 1.2K/1.4k
Source: Network Repository



RT-OBAMA
N/E: 3.2K/2.4k
Source: Network Repository

System Structure(Network) Data

- Using topological characteristics to measure laterality and verticality in real-world systems (represented as directed networks)
- Data: 67 networks representing systems in different contexts are collected from different sources

Nodes: 9 ~ 12k
Edges: 8 ~ 100K
Contexts: Social Network, Infrastructures, Power Systems(Power Grids), Tech Products (Routers), Information Systems, Biological Systems, Ecological Systems etc.

Topological Characteristics

Set--- I : 26 Network-level cohesion features

- Commonly used in social network analysis to characterize networks
- Density, Degree Centrality, Closure etc.
- Focus on networks' overall properties instead of particular nodes or edges
- Unsupervised Learning – PCA
- K-means Clustering analysis

Set--- 2 : Network Motifs (3-node subgraphs)

- Motifs*: "simple building blocks" of a network
- NLP: "Bag of words" – "Bag of motifs"
- Motif-Frequency, Inverse Network Frequency (analog to TF-IDF)
- Singular Value Decomposition to find latent factor
- K-means Clustering analysis

Set-1: Network Cohesion Features

1. Number of Nodes	2. Number of Edges	3. Edge/Node	4. Average Degree	5. Indegree H-Index	6. Degree Centralization	7. Outdegree-Centrality	8. Indegree-Centrality
9. Density	10. Number of Components	11. Component Ratio	12. Connectedness	13. Fragmentation	14. Closure	15. Average Distance	16. Std. Deviation of the Distance
17. Diameter	18. Wiener Index	19. Dependency Sum	20. Breadth	21. Compactness	22. Mutual edges	23. Asymmetric Edges	24. Null Edges
25. Arc Reciprocity	26. Dyad Reciprocity						

Table R1: 26 Network Cohesion Features

Dimensionality Reduction: Principal Component Analysis

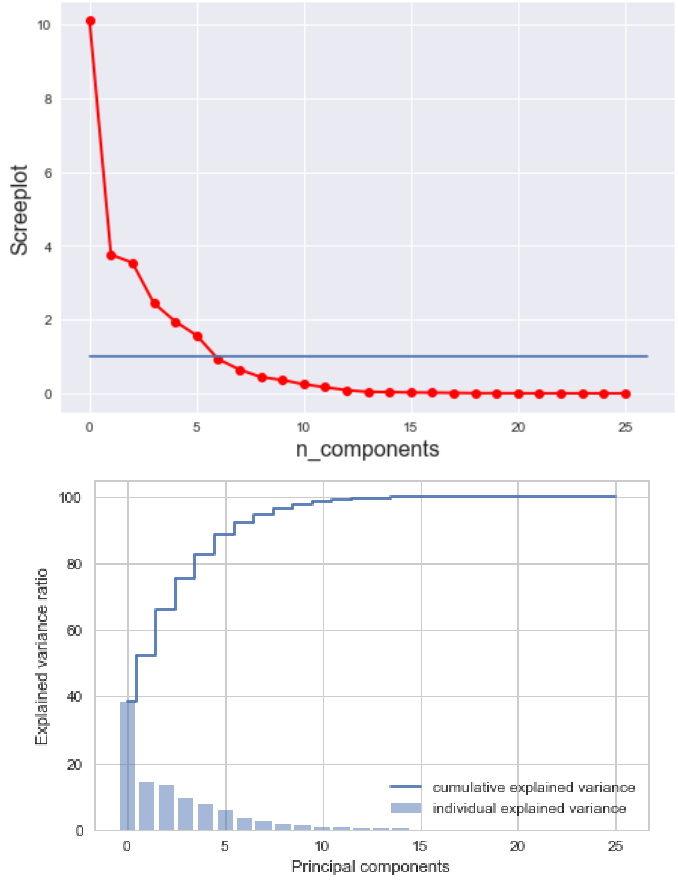


Figure R1: Scree-plot of PCA analysis

Set-1: Network Cohesion Features

Principal Component-2:

- Wiener Index
  - ρ= 0.79
  - The sum of distances between all pairs of vertices of G
- Dependency Sum
  - ρ= 0.77
  - The sum of dependency of all nodes
- Indegree H-Index
  - ρ= 0.58
  - The largest number x such that there are x vertices of degree at least x in the underlying graph

PC2
Hierarchy
||
Verticality

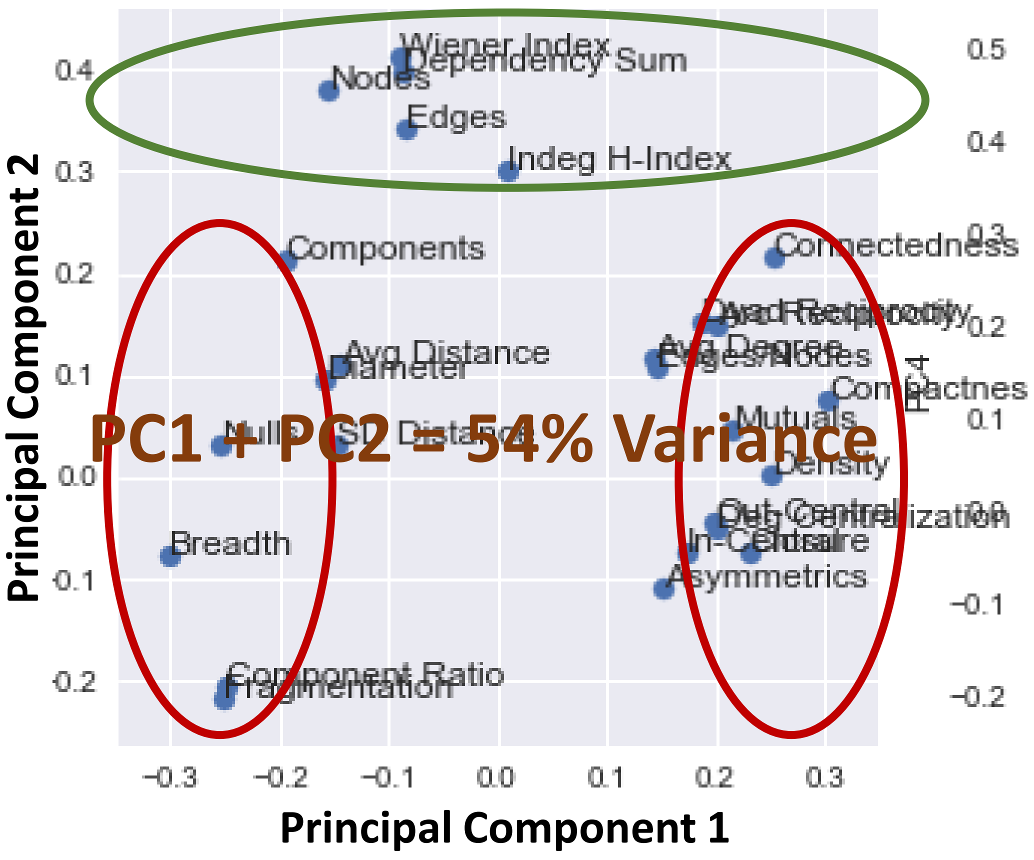


Figure R2: Loading Plot of the first 2 PCs

Principal Component-1:

- Compactness
  - (ρ= 0.95, 1-Breadth)
  - The mean of all the reciprocal distances
- Fragmentation
  - (ρ= -0.80, 1-Connectedness)
  - Proportion of pairs of vertices that are unreachable
- Density (ρ= 0.79)
  - Number of edges divided by the maximum number possible
- Closure (ρ=0.73)
  - The number of non-vacuous transitive triples divided by number of paths of length 2
- Degree Centralization (ρ=0.63)
  - The number of links incident upon a node

PC1:
Interconnectedness
||
Laterality

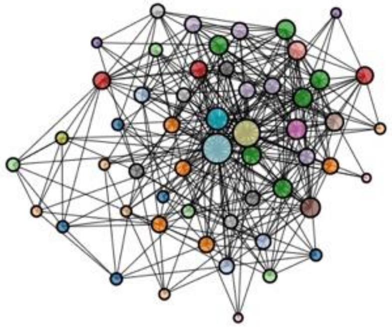
conference on systems engineering research 2019

Clusters Based on PCs

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Cluster-1
(high lateral)

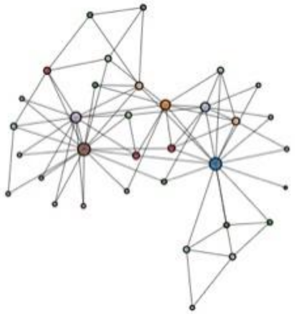
a
*



b
*

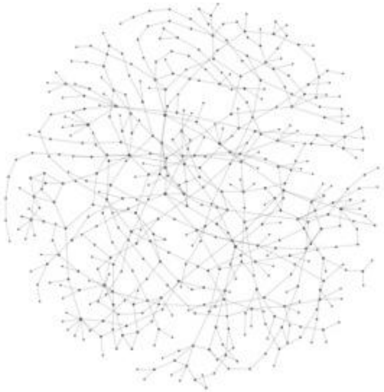
c

d



a-ECO-STMARKS, b- CELEGAN, c-DOLPHIN, d-KARATE

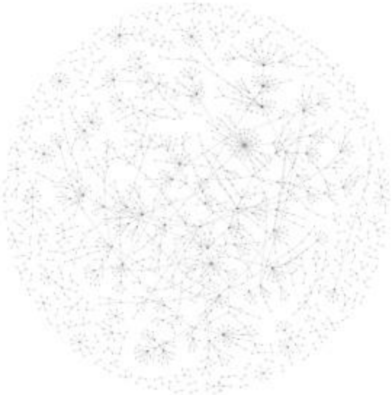
a
*



b
*

c

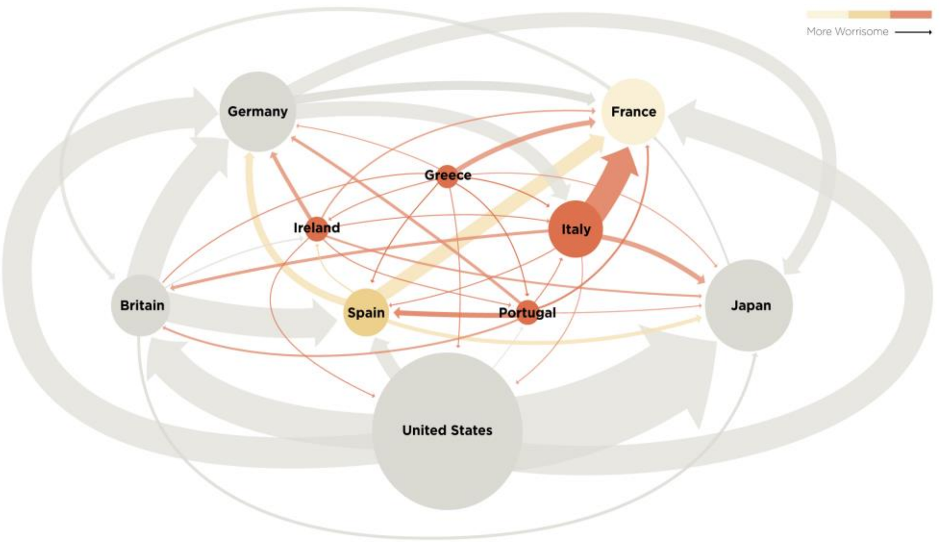
d

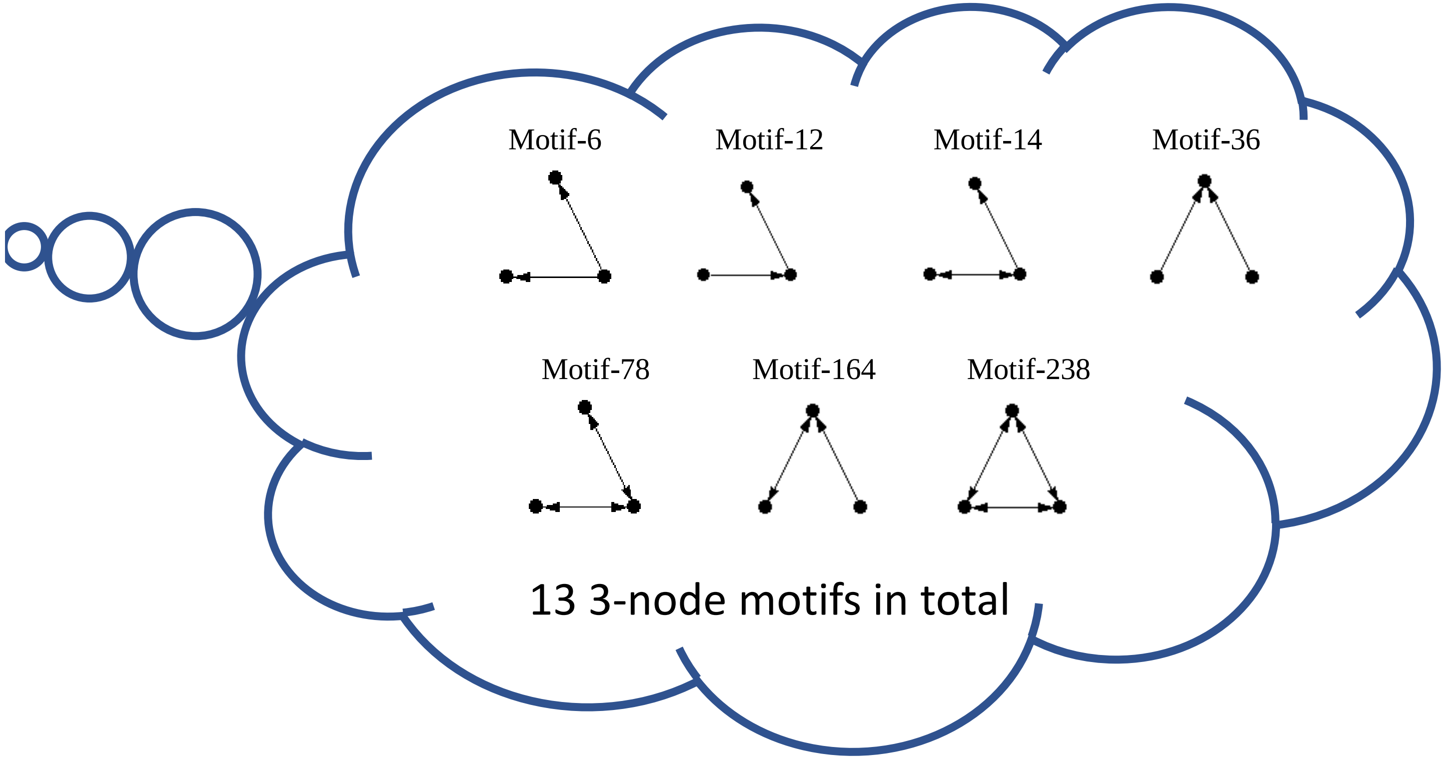


Cluster-2
(low lateral)

a-POW-BUS, b-ROAD-EU, c-RT-OBAMA, d-CSPHD

Set-2: Network Motif Characteristics



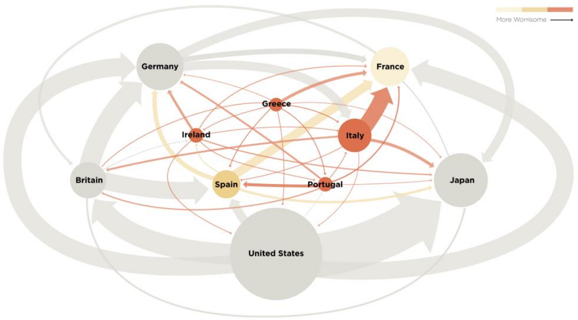


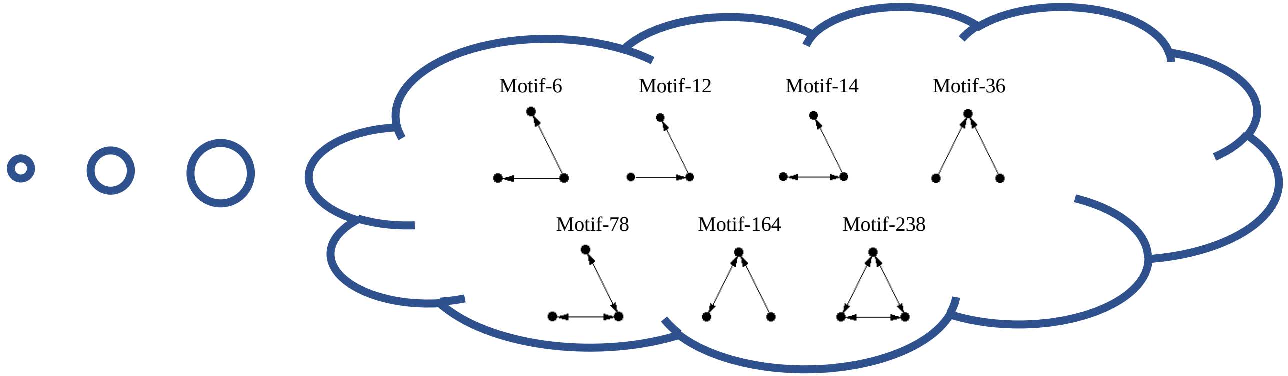
Motifs: “simple building blocks”

Barabási, Albert-László, and Márton Pósfai. 2016. *Network Science*. 1 edition. Cambridge, United Kingdom: Cambridge University Press.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, *298*(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

Set-2: Network Motif Characteristics





"Bag of motifs"

Motif-Frequency, Inverse Network Frequency (MF-INF, analog to TF-IDF)

$$MF_{i,j} = \frac{c_{i,j}}{\sum_k c_{k,j}} \qquad INF_{i,j} = \log \frac{|N|}{|\{j : c_{i,j} \neq 0\}|}$$

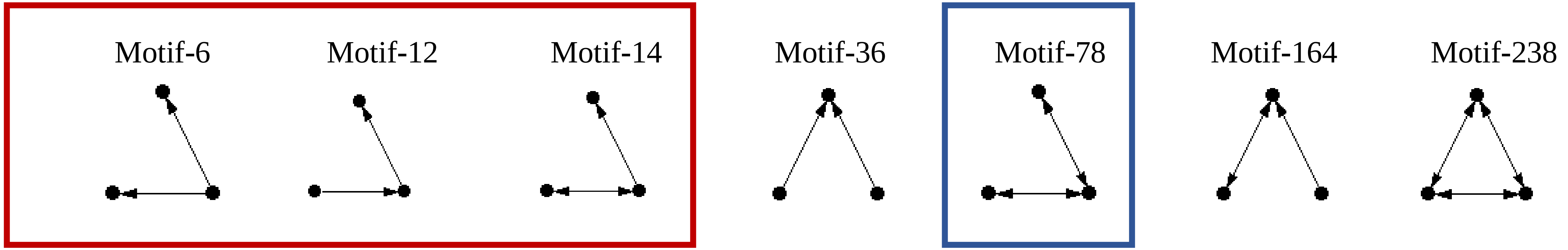
Singular Value Decomposition – Latent Factor

$$MF\text{-}INF_{i,j} = MF_{i,j} \times INF_{i,j}$$

Barabási, Albert-László, and Márton Pósfai. 2016. *Network Science*. 1 edition. Cambridge, United Kingdom: Cambridge University Press.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, *298*(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

Set-2: Latent Factor Analysis

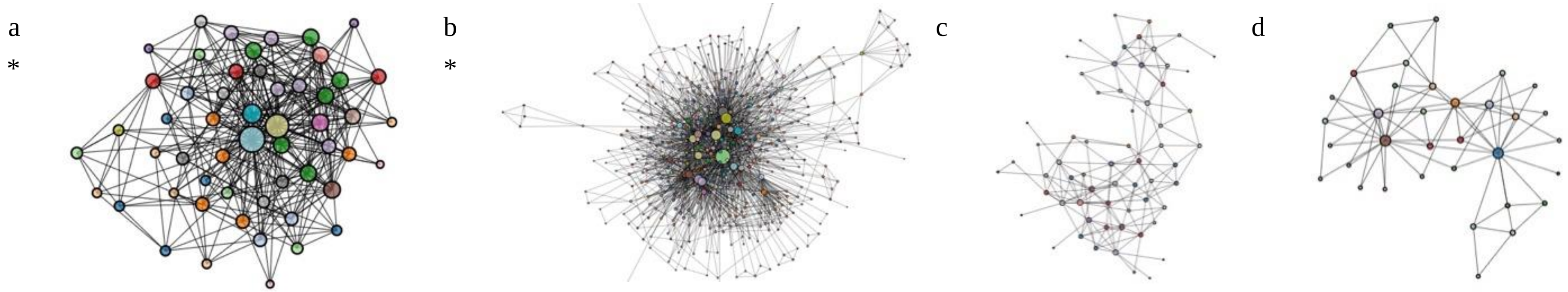


First Latent Factor
-
Verticality

Second Latent Factor
-
Laterality

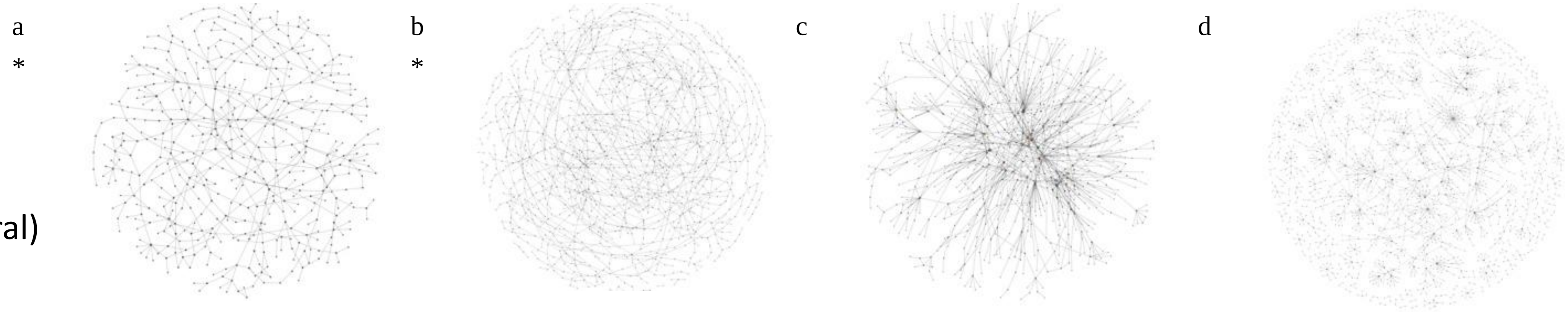
Clusters Based on PCs

Cluster-1
(high lateral)



a-ECO-STMARKS, b- CELEGAN, c-DOLPHIN, d-KARATE

Cluster-2
(low lateral)



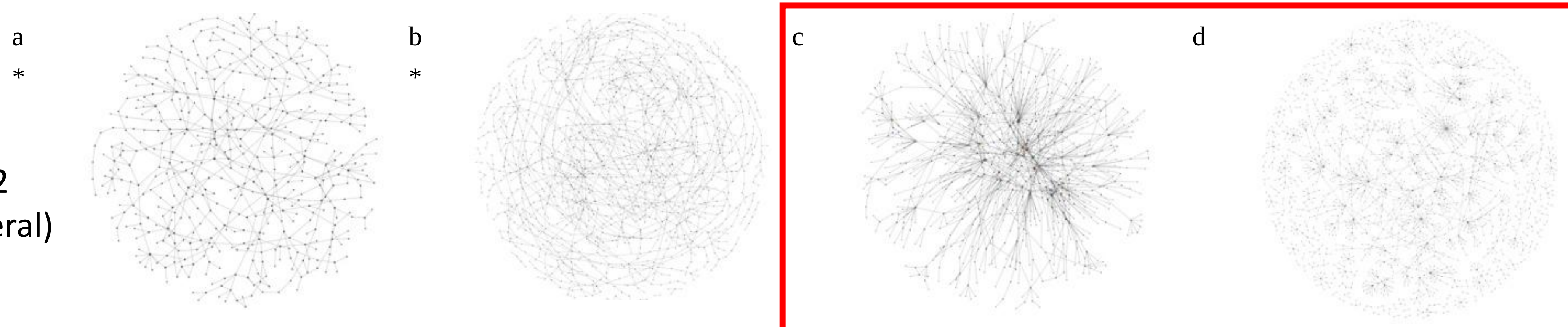
a-POW-BUS, b-ROAD-EU, c-RT-OBAMA, d-CSPHD

Clusters Based on Motifs

Cluster-1
(high lateral)

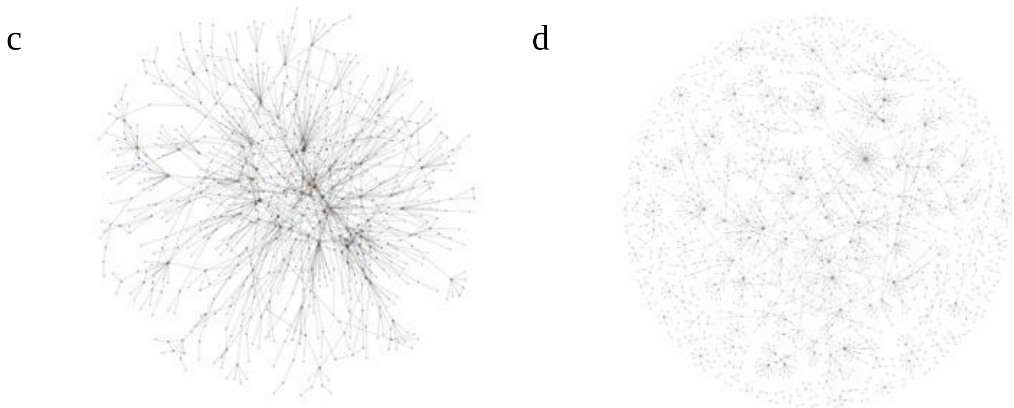
a-ECO-STMARKS, b- CELEGAN, c-DOLPHIN, d-KARATE

Cluster-2
(low lateral)

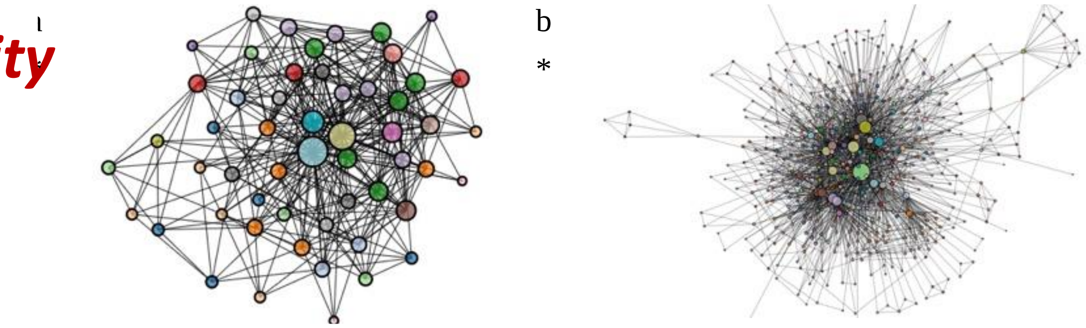


a-POW-BUS, b-ROAD-EU, c-RT-OBAMA, d-CSPHD

High Verticality

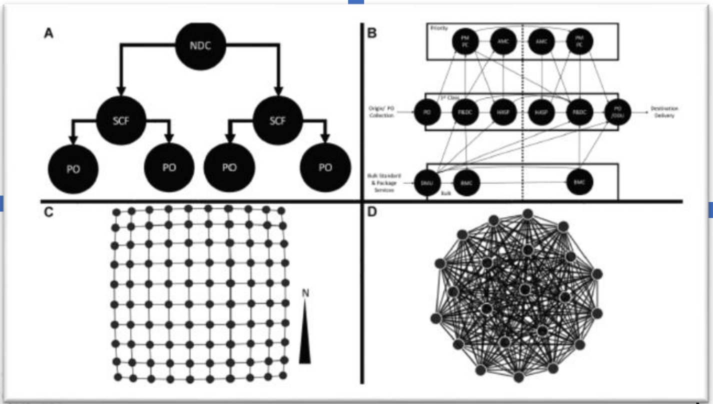


c-RT-OBAMA, d-CSPHD
(tree-like, family tree)
A: Tree Structure

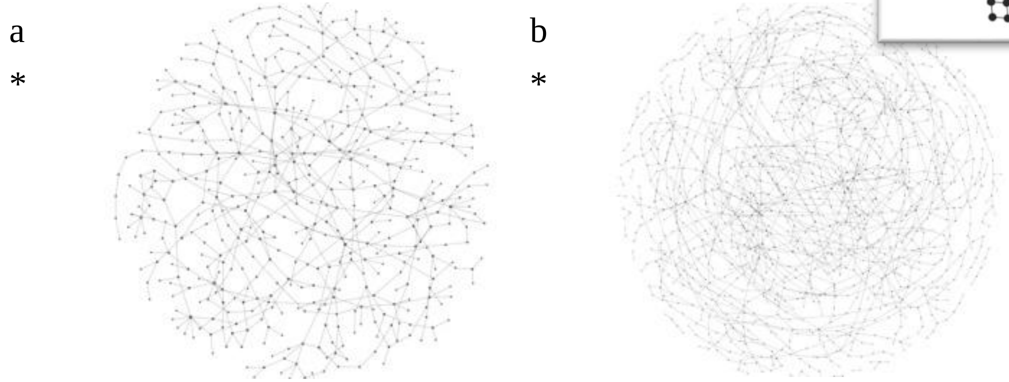


a-ECO-STMARKS, b- CELEGAN
(canonical layered structures)
B: Layered Structure

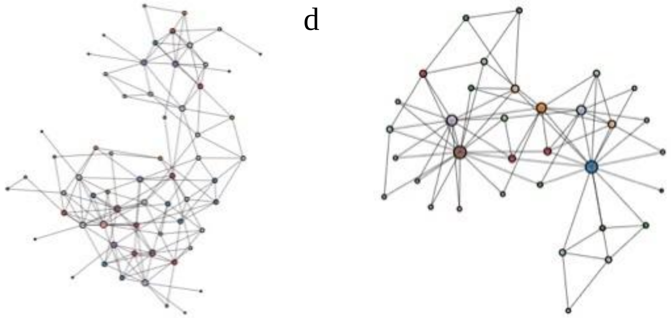
Low Laterality



High Laterality



a-POW-BUS, b-ROAD-EU
C: Grid Structure



c-DOLPHIN, d-KARATE
D: Team Structure

Low Verticality

Conclusion

1) laterality and verticality are meaningful characteristics of systems that may be used to assess similarity to generic architectures
2) network cohesion features and network motifs may both be used to distinguish laterality and verticality and group systems

Limitations:

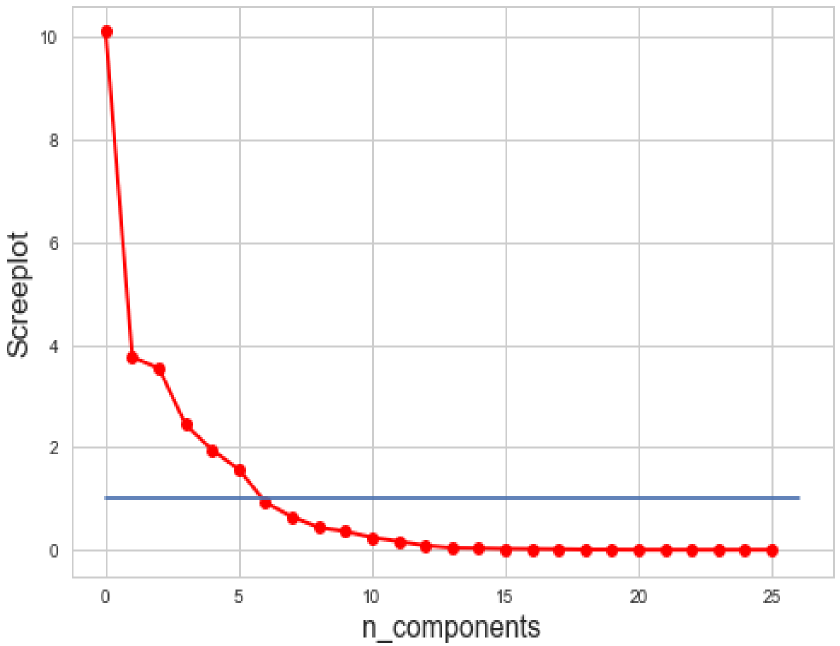
- relatively small sample size and unbalanced networks by domains
- only consider 3-node motifs due to limited computational power

Future work:

- explore more varied network data, utilizing faster algorithms, examining how the choice of network representation interacts with systems structures, and exploring the association between these measures and system lifecycle properties.

Backup Slides

Feature Names	Description
Nodes	Number of nodes
Edges	Number of edges
Edges/Nodes	Ratio of N/E
Avg Degree	The average degree in the underlying graph
Indeg H-Index	The largest number x such that there are x vertices of degree at least x in the underlying graph
Deg Centralization	The number of links incident upon a node
Out-Central	The number of ties directed to the node
In-Central	The number of ties that the node directs to others
Density	Number of edges divided by the maximum number possible
Components	Number of weak components
Component Ratio	Number of components minus one divided by the number of actors minus one
Connectedness	1 minus the fragmentation
Fragmentation	Proportion of pairs of vertices that are unreachable
Closure	The number of non-vacuous transitive triples divided by number of paths of length 2
Avg Distance	Average geodesic distance amongst reachable pairs
SD Distance	Standard deviation of the geodesic distances amongst reachable pairs
Diameter	Length of the longest geodesic
Wiener Index	The sum of distances between all pairs of vertices of G
Dependency Sum	The sum of dependency of all nodes
Breadth	1 minus the compactness
Compactness	The mean of all the reciprocal distances
Mutual	The probability of mutual relation
Asymmetric	The probability of Asymmetric relation
Nulls	The probability of nulls
Arc Reciprocity	The proportion of reciprocated arcs of all outgoing arcs
Dyad Reciprocity	The proportion of symmetric dyads of all dyads



Left : scree plot of principle components of cohesion features; Right: Loading table

C.1-Laterality

C.2-Verticality

Interconnectedness

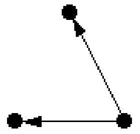
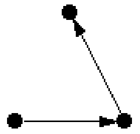
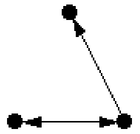
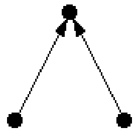
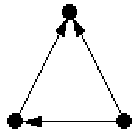
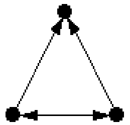
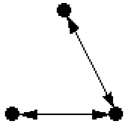
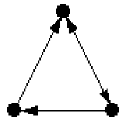
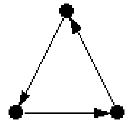
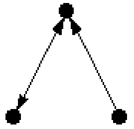
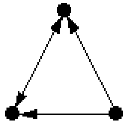
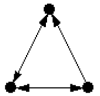
Difficulty of reaching every node in one direction

	C.1	C.2	C.3	C.4	C.5	C.6
Breadth	-0.95					
Compactness	0.95					
Connectedness	0.80					
Fragmentation	-0.80					
Null Edges	-0.80					
Density	0.79				0.50	
Component Ratio	-0.79					
Closure	0.73					
Mutual edges	0.68				0.62	
Degree Centralization	0.63					
Outdegree-Centrality	0.62					
Components	-0.62					
Indegree-Centrality	0.55					
Wiener Index		0.79				
Dependency Sum		0.77				
Number of Nodes		0.73				
Number of Edges		0.66				
Dyad Reciprocity	0.59		-0.61			
Indegree H-Index		0.58	0.60			
Edges/Nodes			0.58			
Average Degree			0.53			0.51
SD Distance				0.77		
Average Distance				0.73		
Diameter				0.74		
Asymmetric Edges						

Results of Principle Component Analysis

C1+C2 = 54% variance

Motifs

No.	ID	
1(Boss)	M-6	
2(Order)	M-12	
3(Division)	M-14	
4(Competition)	M-36	
5(Feedforward)	M-38	
6(Consultant)	M-46	
7(Grid)	M-78	
8(Family)	M-102	
9(Loop)	M-140	
10(Leader)	M-164	
11(Controller)	M-166	
12(Acquisition)	M-174	
13(Team)	M-238