Compliance Graph Analysis Techniques using Network Theory Approaches

Presentation by Noah L. Schrick for the University of Tulsa's CS-7863 Network Theory course final

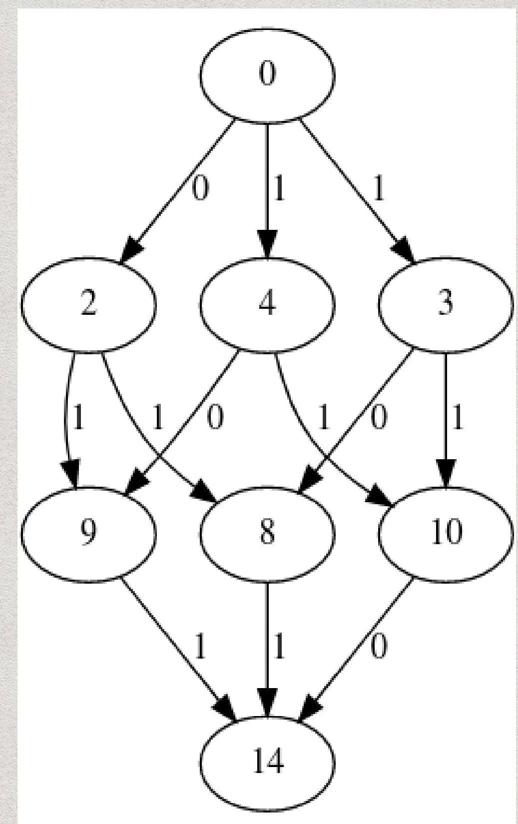
Presentation Overview

- Definitions and Introduction
- Experimental Networks
- Centralities and their Applications to Compliance Graphs
- Transitive Closure
- Dominator Tree
- Results and Result Analysis
- Conclusions and Future Work

Definitions and Introduction (1)

- Compliance Graphs:
  - Examine a system or set of systems in regards to SOX, HIPAA, GDPR, PCI DSS, etc.
  - Use for compliance checking of cyber-physical systems (critical infrastructure, IoT)
  - Quantify risks and violations in terms of fines, legal sanctions, mandatory shutdowns, and other costs of compliance violation

Definitions and Introduction (2)



Definitions and Introduction (3)

- Analysis Difficulties for DAGs:
  - Asymmetric adjacency matrices
  - Eigenvector centralities are not applicable [6]
  - Graph Laplacian is often undefined [7]
  - Complex eigenvalues [7]
  - Adjacency matrices are unable to be diagonalized [7]

Definitions and Introduction (4)

- Related Work:
  - Centralities for Attack Graphs [8]
    - Katz, K-path Edge, PageRank
  - Hermitian adjacency matrix [9]
    - Avoid complex eigenvalues
  - Directed graph spectra [10] combined with Schur’s Theorem [7]
    - Bounded eigenvalues and explicit computation

Experimental Networks (1)

- Three Networks:
  - Vehicle Maintenance Network
  - HIPAA Compliance Network
  - PCI DSS Network

Network	Nodes	Edges	Connectivity (%)
Car	2491	12968	0.209
HIPAA	2321	8063	0.150
PCI DSS	61	163	4.381

Table 1: Network Properties for the Three Networks Utilized

Centralities (1)

- Degree
  - Trivial, localized
  - High in-degree centrality: greater probability the system may enter this state
  - High out-degree centrality: creates further opportunity for violation

Centralities (2)

- Betweenness
  - Importance related to information flow
  - Shortest pair between all pairs of nodes in a network
  - High betweenness centrality: quickest way that systems may fall out of compliance

$$\sum_{i\neq i\neq v} \frac{\sigma_{ij}(v)}{\sigma_{ij}}$$

Centralities (3)

- Katz [14]
  - Importance related to all paths in a network
  - High Katz centrality: prevent opportunity of future compliance violation that is reachable, but may be many steps ahead

$$C_{\mathrm{Katz}}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (A^k)_{ji}$$

$$\vec{x} = (I - \alpha A)^{-1} \vec{\beta}$$

Centralities (4)

- K-path Edge [15]
  - Importance is related to information flow, but is localized and constrained to k-steps from a given node
  - High K-path Edge centrality: identifies a short chain of changes that may result in violation
    - Prevent states where the system is near a violation

$$L^{(K)}(m) = \sum_{n=1}^{N} \frac{\delta_n^{(K)}(m)}{\delta_n^{(K)}}$$

Centralities (5)

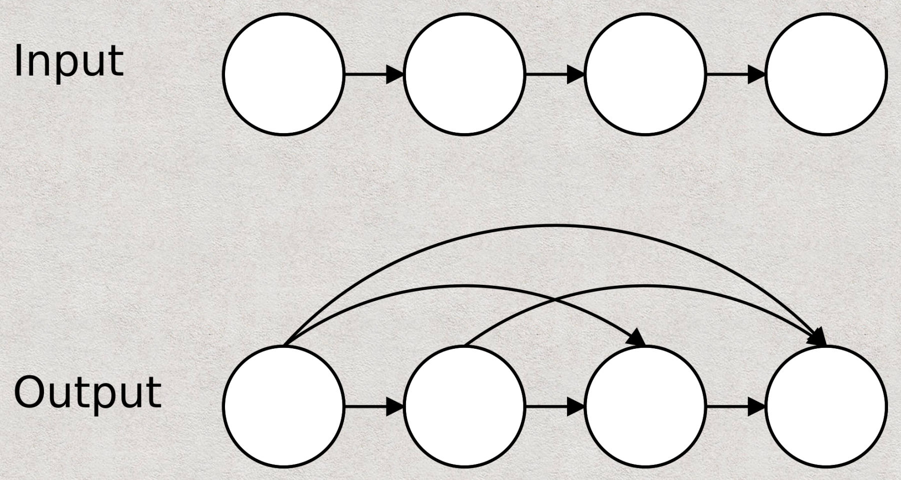
- PageRank [16], [17]
  - Importance related to number and quality of connections
  - High PageRank centrality: determine nodes that are likely to be visited, and apply correction schemes near these nodes

$$x_i = \frac{1-\gamma}{n} + \gamma \sum_{j=1}^{n} \frac{A_{ij}}{k_j} x_j$$

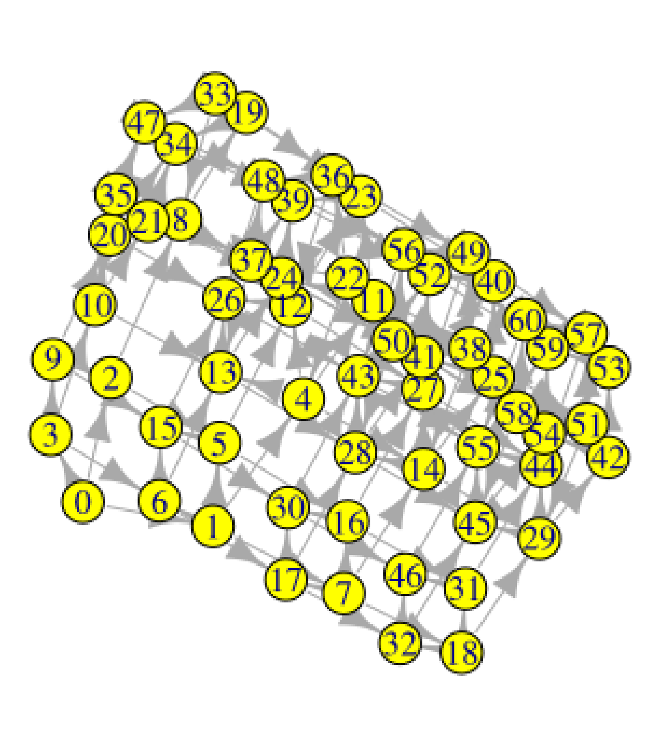
$$(I - \gamma A D)\vec{x} = \frac{1-\gamma}{n}\mathbb{1}$$

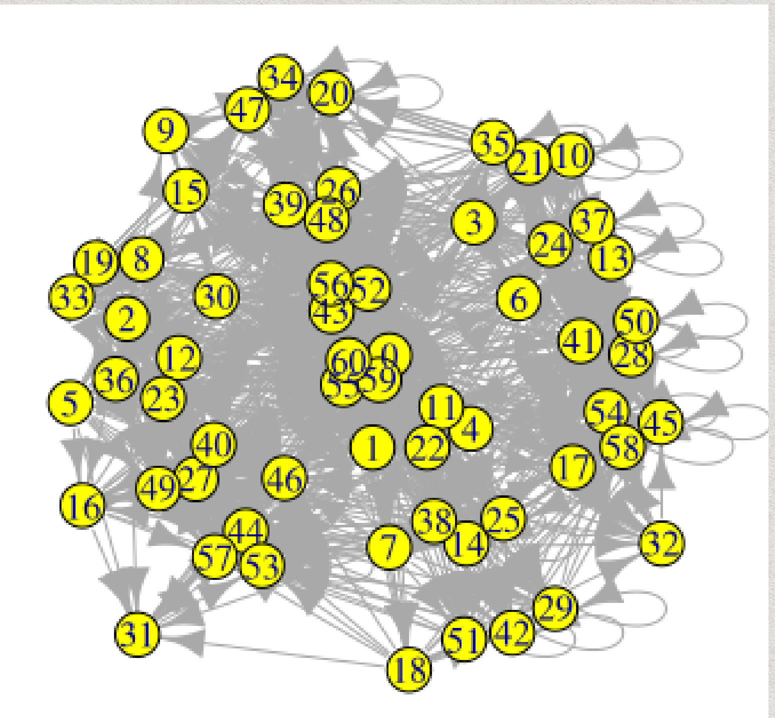
Transitive Closure (1)

- Transitive relation on a given binary set
- Determine reachability of a given network
- Adversarial actions with unlimited time and resources
- No prior knowledge of the network, and all changes are equally likely



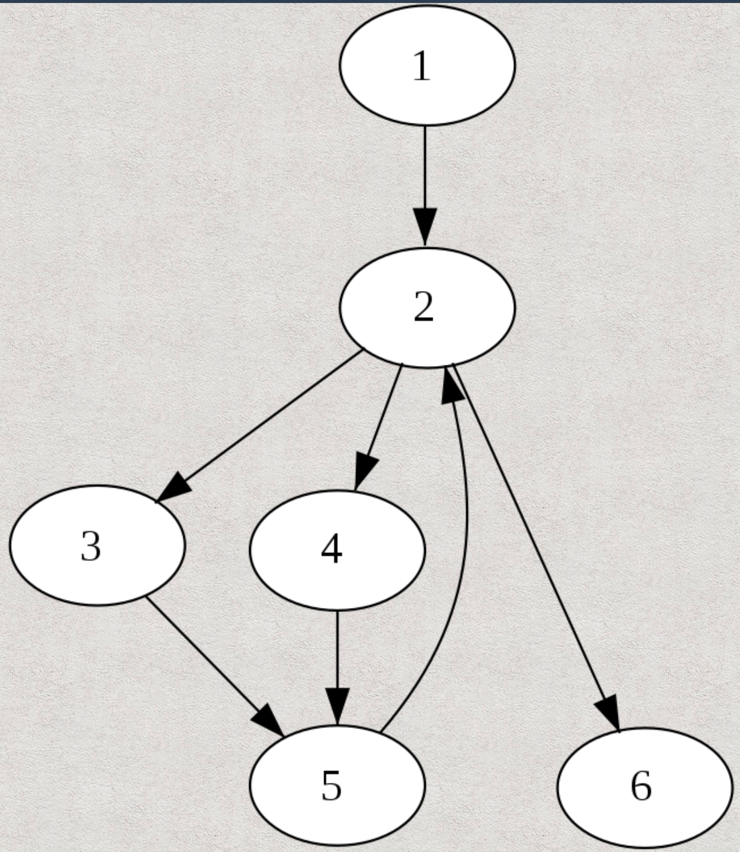
Transitive Closure (2)





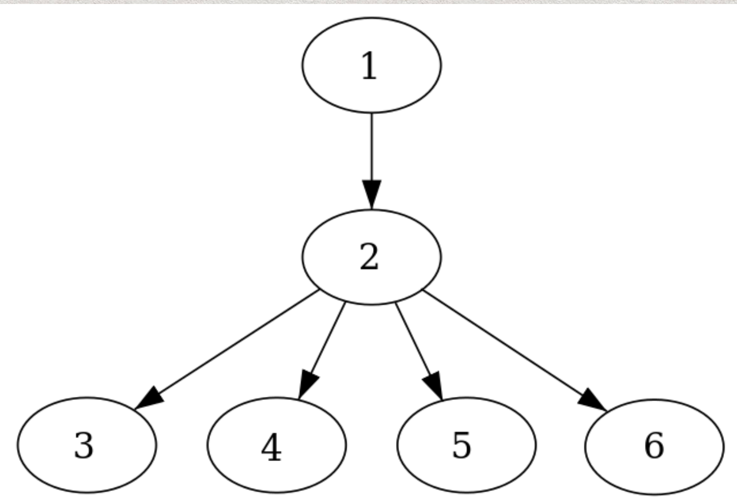
Dominator Tree (1)

- Dominance [20]
  - Node that is in every path to another node
  - Node 2 dominates 2, 3, 4, 5, and 6

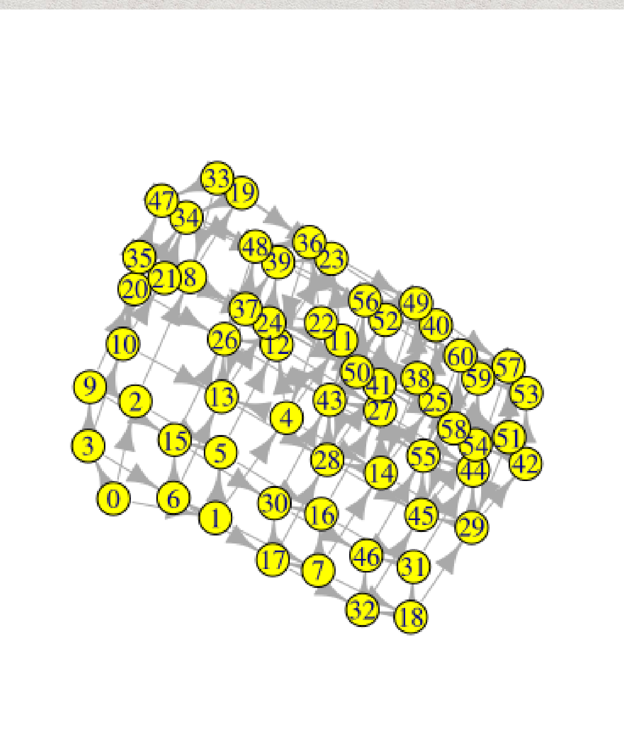


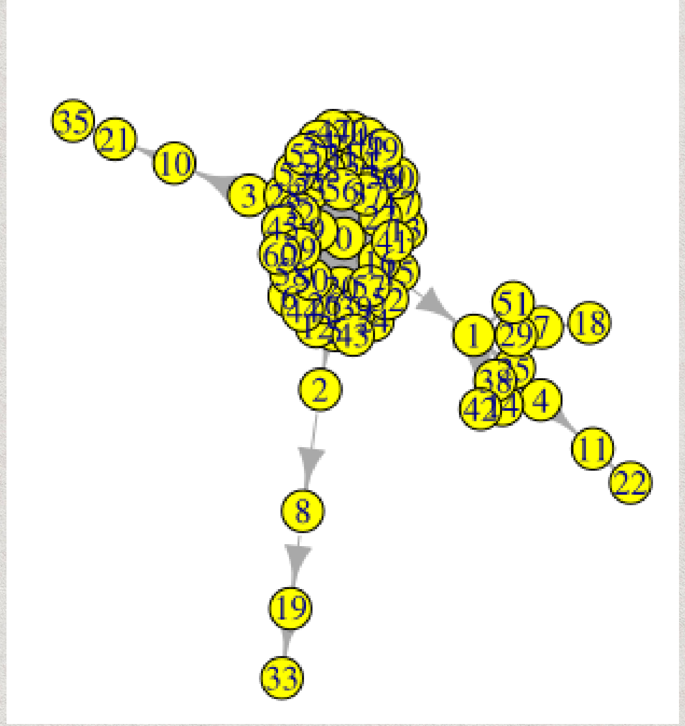
Dominator Tree (2)

- Dominator Tree
  - Parent nodes immediately dominate its children
  - Restructure compliance graphs in terms of information flow



Dominator Tree (3)





Results (1)

Base		Transitive Closure		Dominant Tree	
Node	Value	Node	Value	Node	Value
314	11	0	2490	1	1246
346	10	1	2489	3	934
362	10	3	2487	7	156
370	10	7	2479	42	115
374	10	15	2463	314	31
376	10	27	2447	0	1
377	10	42	2431	15	1
378	10	60	2367	27	1
379	10	87	2303	60	1
380	10	130	2239	87	1
381	10	187	2175	130	1
382	10	250	2111	187	1
398	9	314	2047	250	1
406	9	2	1244	2	0
410	9	4	1243	4	0

Table 2: Top 15 Nodes with Degree Centrality

Base		Transitive Closure		Dominant Tree	
Node	Value	Node	Value	Node	Value
42	9067.205	0	0	1	2489
27	8442.166	1	0	3	2486
60	8279.62	2	0	7	927
87	7580.359	3	0	42	906
15	7578.523	4	0	27	760
130	6868.21	5	0	15	612
7	6482.031	6	0	314	372
187	6111.862	7	0	250	352
50	5950.928	8	0	187	330
70	5822.054	9	0	130	306
104	5683.944	10	0	87	280
156	5474.525	11	0	60	252
1467	5299.985	12	0	0	0
250	5296.964	13	0	2	0
115	5196.398	14	0	4	0

Table 6: Top 15 Nodes with Betweenness Centrality

Results (2)

Base		Transitive Closure		Dominant Tree	
Node	Value	Node	Value	Node	Value
2490	0.0827	2490	0.1992	314	0.001655
1004	0.01506	2479	0.0158	250	0.001479
1467	0.00969	2480	0.0158	187	0.001272
2479	0.00948	2481	0.0158	130	0.001028
2480	0.00948	2482	0.0158	42	0.001025
2481	0.00948	2483	0.0158	87	0.00074
2482	0.00948	2484	0.014	27	0.00074
2483	0.00948	2485	0.014	1	0.00074
667	0.00919	2486	0.0139	378	0.00044
2484	0.0083	2487	0.0139	379	0.00044
2485	0.0083	2488	0.0139	380	0.00044
2486	0.0083	2489	0.0139	381	0.00044
2487	0.0083	2424	0.0029	382	0.00044
2488	0.0083	2425	0.0029	470	0.00044
2489	0.0083	2426	0.0029	471	0.00044

Table 5: Top 15 Nodes with PageRank Centrality

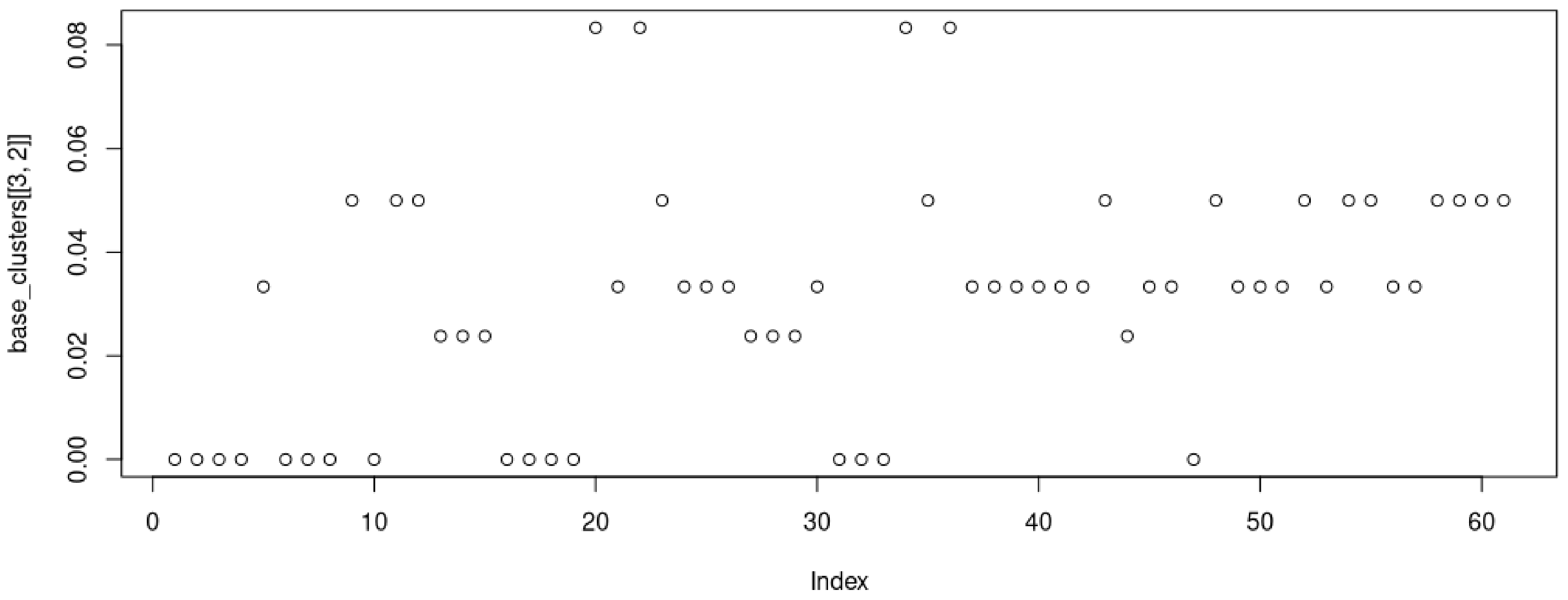
Conclusions

- Each centrality measure provides various information regarding network correction schemes
- Unique rankings can be identified from transitive closures and dominator trees

Future Work

- Artificially implement correction schemes based on centrality measures to observe the effects
- Implement user-defined data matrices for Katz or PageRank
- Edge weights in terms of probability of change
- Further research into transitive closures and dominator trees for compliance graph representations
- Clustering investigation with methods applicable to DAGs

Clemente and Grassi



Q&A