Social Network Analysis of the Planetary Data System

Final Presentation

Kate Martin
Mark Avnet

May 9, 2006

ESD.342
Advanced System Architecture
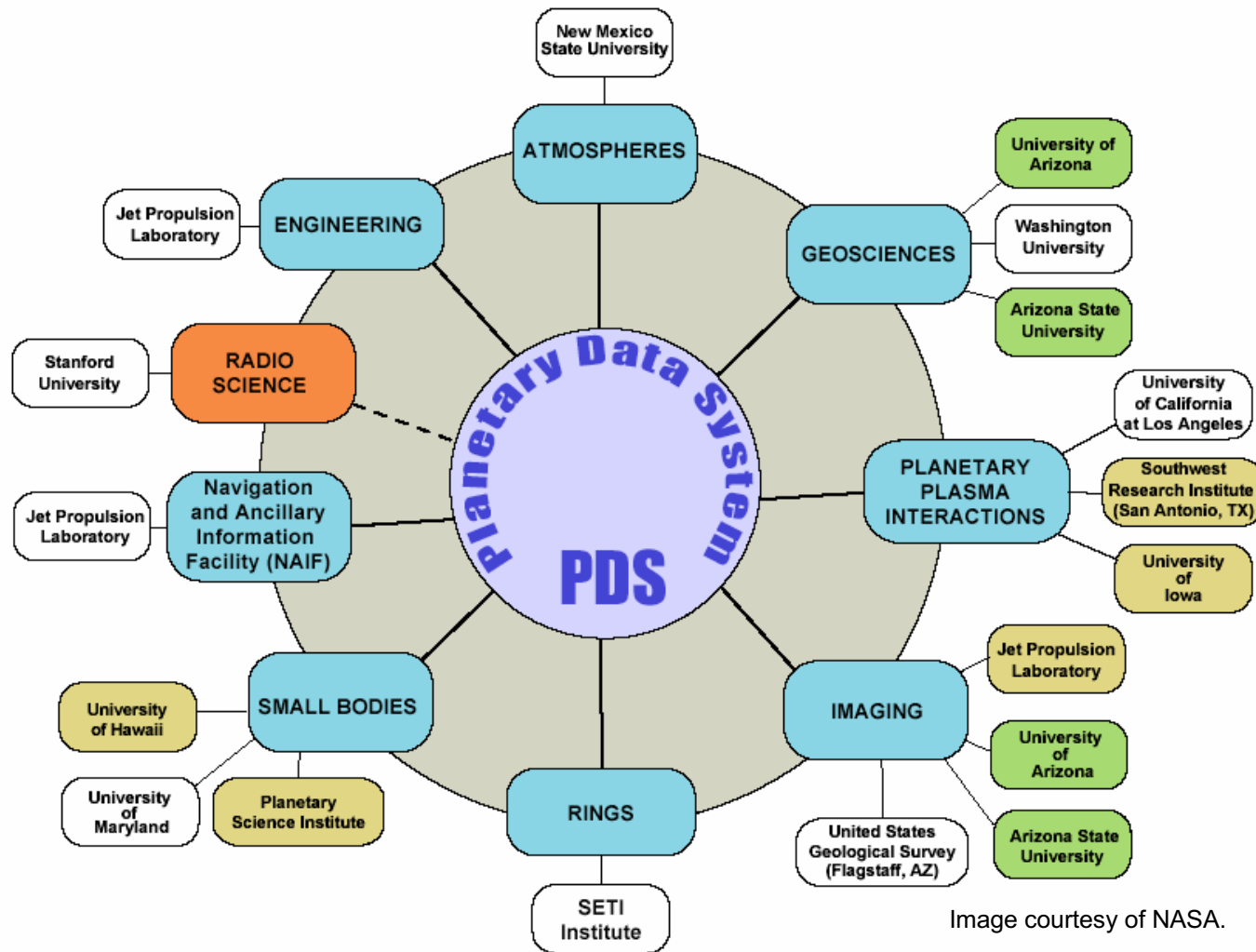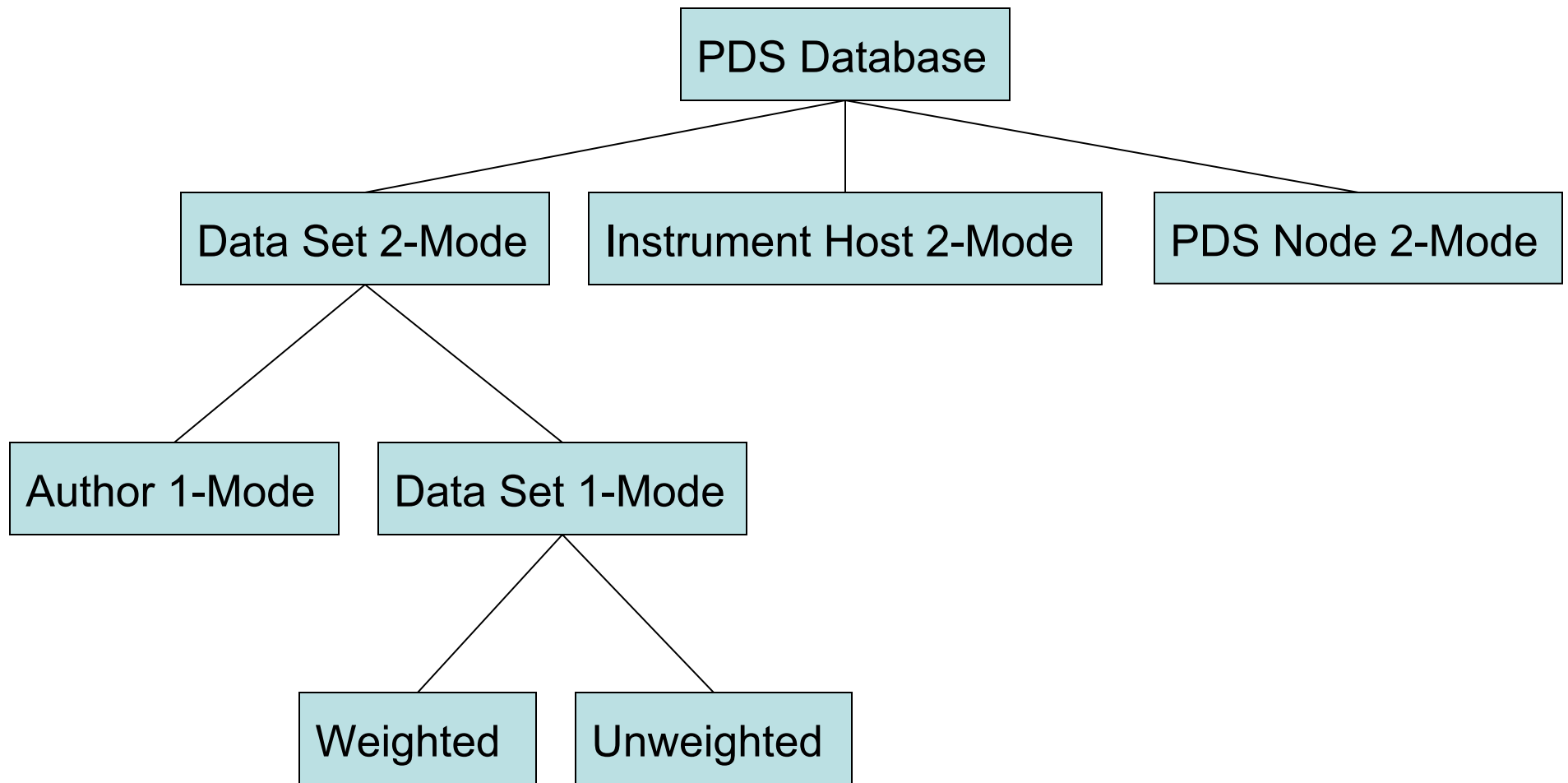
Overview Planetary Data System

Image courtesy of NASA.

Decomposing the PDS Database into 3 Bipartite Networks …

```
                        ┌──────────────────┐
                        │   PDS Database   │
                        └──────────────────┘
              ┌───────────────────┼────────────────────┐
    ┌──────────────────┐ ┌────────────────────────┐ ┌──────────────────┐
    │  Data Set 2-Mode │ │ Instrument Host 2-Mode │ │ PDS Node 2-Mode  │
    └──────────────────┘ └────────────────────────┘ └──────────────────┘
         ┌──────┴────────┐
┌──────────────────┐ ┌──────────────────┐
│  Author 1-Mode   │ │ Data Set 1-Mode  │
└──────────────────┘ └──────────────────┘
                        ┌──────┴──────┐
                ┌──────────────┐ ┌──────────────┐
                │   Weighted   │ │  Unweighted  │
                └──────────────┘ └──────────────┘
```

… Yields 12 separate 1-Mode networks for comparative analysis

Newman-Type Table for the 12 PDS Networks

Type of Node	Network	Weighted?	n	m	$<k>$	l	$\frac{\log n}{\log <k>}$	α	$C^{(2)}$	$<k>/n$	r
Authors as Nodes	PDS Nodes	No	439	27493	125.3	1.886	1.260	-0.31	0.981	0.285	0.80
	PDS Nodes	Yes	439	27527	125.4	1.886	1.259	-0.31	0.988	0.286	0.80
	Instrument Hosts	No	439	8240	37.5	2.703	1.678	-0.49	0.929	0.086	0.68
	Instrument Hosts	Yes	439	8581	39.1	2.703	1.660	-0.48	1.092	0.089	0.55
	Data Sets	No	439	3240	14.8	3.1	2.260	-0.60	0.936	0.034	0.96
	Data Sets	Yes	439	4366	19.9	3.1	2.035	-0.53	1.534	0.045	0.95
Events as Nodes	PDS Nodes	No	9	11	2.4	1.861	2.458	-1.1	0.62	0.272	-0.70
	PDS Nodes	Yes	9	28	6.2	1.861	1.202	-0.73	1.927	0.691	-0.52
	Instrument Hosts	No	103	282	5.5	2.426	2.726	--	0.715	0.053	0.039
	Instrument Hosts	Yes	103	520	10.1	2.426	2.004	--	1.675	0.098	-0.042
	Data Sets	No	1046	5820	11.1	2.761	2.886	-0.27	0.937	0.011	0.99
	Data Sets	Yes	1046	6514	12.5	2.761	2.757	-0.29	1.455	0.012	0.97

- **Network-Specific Results**
  - Small-magnitude α values, but the regions that follow a power law are small for many of the networks
  - Short path lengths and high clustering coefficients
  - Degree Correlation
    - "Social" and "information" networks have r > 0
    - "Technological" networks have r < 0 (except in one case where r << 1)

Newman-Type Table for the 12 PDS Networks

Type of Node	Network	Weighted?	n	m	\<k\>	l	log n / log \<k\>	α	C⁽²⁾	\<k\>/n	r
Authors as Nodes	PDS Nodes	No	439	27493	125.3	1.886	1.260	-0.31	0.981	0.285	0.80
	PDS Nodes	Yes	439	27527	125.4	1.886	1.259	-0.31	0.988	0.286	0.80
	Instrument Hosts	No	439	8240	37.5	2.703	1.678	-0.49	0.929	0.086	0.68
	Instrument Hosts	Yes	439	8581	39.1	2.703	1.660	-0.48	1.092	0.089	0.55
	Data Sets	No	439	3240	14.8	3.1	2.260	-0.60	0.936	0.034	0.96
	Data Sets	Yes	439	4366	19.9	3.1	2.035	-0.53	1.534	0.045	0.95
Events as Nodes	PDS Nodes	No	9	11	2.4	1.861	2.458	-1.1	0.62	0.272	-0.70
	PDS Nodes	Yes	9	28	6.2	1.861	1.202	-0.73	1.927	0.691	-0.52
	Instrument Hosts	No	103	282	5.5	2.426	2.726	--	0.715	0.053	0.039
	Instrument Hosts	Yes	103	520	10.1	2.426	2.004	--	1.675	0.098	-0.042
	Data Sets	No	1046	5820	11.1	2.761	2.886	-0.27	0.937	0.011	0.99
	Data Sets	Yes	1046	6514	12.5	2.761	2.757	-0.29	1.455	0.012	0.97

- **Network-Specific Results**
  - **Small-magnitude α values, but the regions that follow a power law are small (and possibly non-existent) for many of the networks**
  - Short path lengths and high clustering coefficients
  - Degree Correlation
    - "Social" and "information" networks have r > 0
    - "Technological" networks have r < 0 (except in one case where r << 1)

Newman-Type Table for the 12 PDS Networks

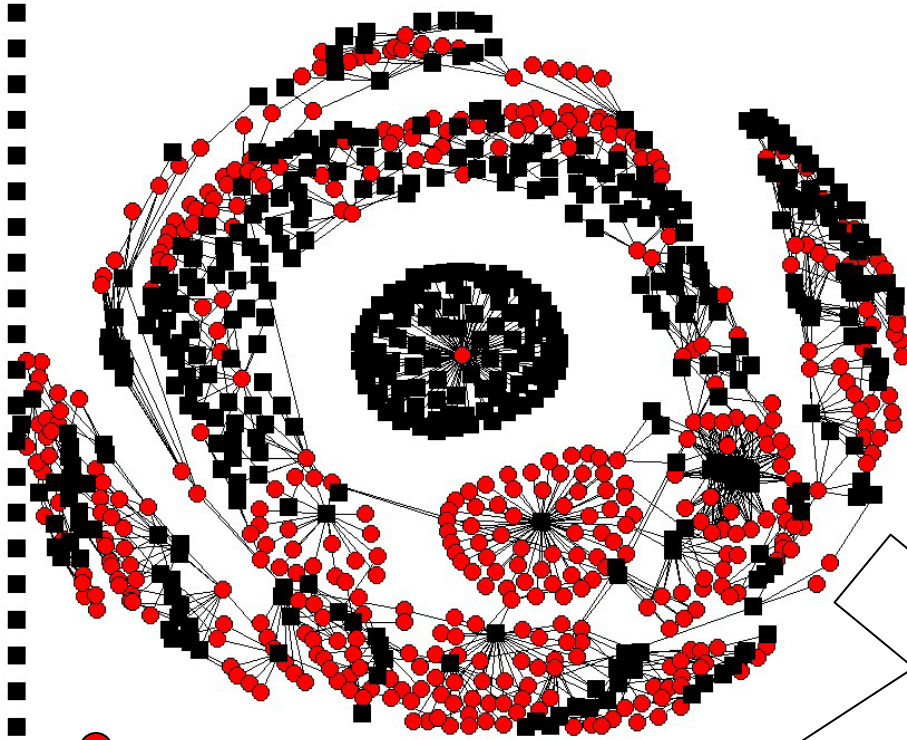| Type of Node	Network	Weighted?	n	m	$<k>$	l	$\log n / \log <k>$	α	$C^{(2)}$	$<k>/n$	r
Authors as Nodes	PDS Nodes	No	439	27493	125.3	1.886	1.260	-0.31	0.981	0.285	0.80
	PDS Nodes	Yes	439	27527	125.4	1.886	1.259	-0.31	0.988	0.286	0.80
	Instrument Hosts	No	439	8240	37.5	2.703	1.678	-0.49	0.929	0.086	0.68
	Instrument Hosts	Yes	439	8581	39.1	2.703	1.660	-0.48	1.092	0.089	0.55
	Data Sets	No	439	3240	14.8	3.1	2.260	-0.60	0.936	0.034	0.96
	Data Sets	Yes	439	4366	19.9	3.1	2.035	-0.53	1.534	0.045	0.95
Events as Nodes	PDS Nodes	No	9	11	2.4	1.861	2.458	-1.1	0.62	0.272	-0.70
	PDS Nodes	Yes	9	28	6.2	1.861	1.202	-0.73	1.927	0.691	-0.52
	Instrument Hosts	No	103	282	5.5	2.426	2.726	--	0.715	0.053	0.039
	Instrument Hosts	Yes	103	520	10.1	2.426	2.004	--	1.675	0.098	-0.042
	Data Sets	No	1046	5820	11.1	2.761	2.886	-0.27	0.937	0.011	0.99
	Data Sets	Yes	1046	6514	12.5	2.761	2.757	-0.29	1.455	0.012	0.97

- **Network-Specific Results**
  - Small-magnitude α values, but the regions that follow a power law are small for many of the networks
  - **Short path lengths** and high clustering coefficients
  - Degree Correlation
    - "Social" and "information" networks have r > 0
    - "Technological" networks have r < 0 (except in one case where r << 1)

Newman-Type Table for the 12 PDS Networks

Type of Node	Network	Weighted?	n	m	$<k>$	l	$\log n / \log <k>$	α	$C^{(2)}$	$<k>/n$	r
Authors as Nodes	PDS Nodes	No	439	27493	125.3	1.886	1.260	-0.31	0.981	0.285	0.80
	PDS Nodes	Yes	439	27527	125.4	1.886	1.259	-0.31	0.988	0.286	0.80
	Instrument Hosts	No	439	8240	37.5	2.703	1.678	-0.49	0.929	0.086	0.68
	Instrument Hosts	Yes	439	8581	39.1	2.703	1.660	-0.48	1.092	0.089	0.55
	Data Sets	No	439	3240	14.8	3.1	2.260	-0.60	0.936	0.034	0.96
	Data Sets	Yes	439	4366	19.9	3.1	2.035	-0.53	1.534	0.045	0.95
Events as Nodes	PDS Nodes	No	9	11	2.4	1.861	2.458	-1.1	0.62	0.272	-0.70
	PDS Nodes	Yes	9	28	6.2	1.861	1.202	-0.73	1.927	0.691	-0.52
	Instrument Hosts	No	103	282	5.5	2.426	2.726	--	0.715	0.053	0.039
	Instrument Hosts	Yes	103	520	10.1	2.426	2.004	--	1.675	0.098	-0.042
	Data Sets	No	1046	5820	11.1	2.761	2.886	-0.27	0.937	0.011	0.99
	Data Sets	Yes	1046	6514	12.5	2.761	2.757	-0.29	1.455	0.012	0.97

- **Network-Specific Results**
  - Small-magnitude α values, but the regions that follow a power law are small for many of the networks
  - Short path lengths and **high clustering coefficients**
  - Degree Correlation
    - "Social" and "information" networks have r > 0
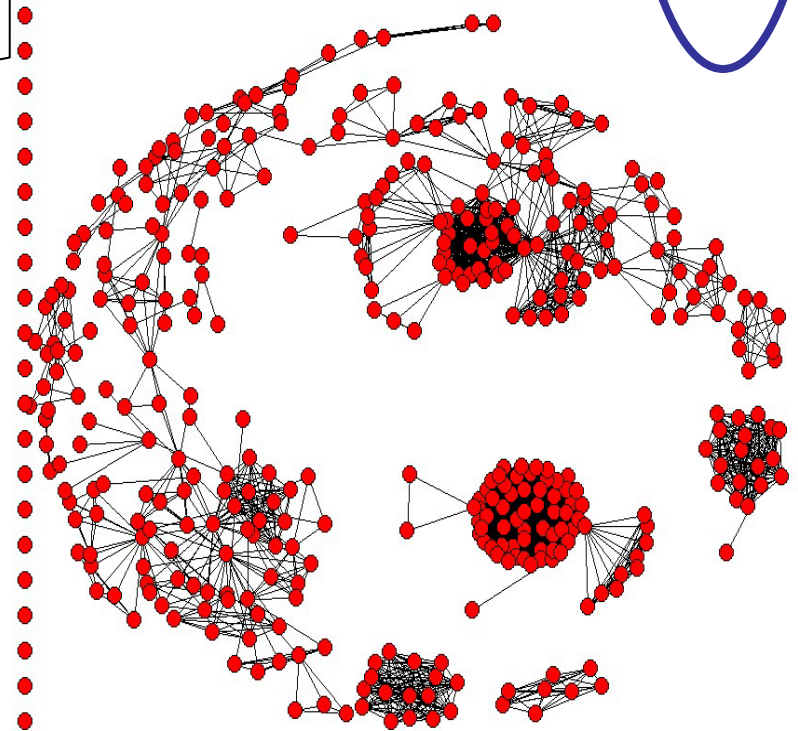   - "Technological" networks have r < 0 (except in one case where r << 1)

Type of Node	Network	Weighted?	$C^{(2)}$
Authors as Nodes	PDS Nodes	No	0.981
	PDS Nodes	Yes	0.988
	Instrument Hosts	No	0.929
	Instrument Hosts	Yes	1.092
	Data Sets	No	0.936
	Data Sets	Yes	1.534
Events as Nodes	PDS Nodes	No	0.62
	PDS Nodes	Yes	1.927
	Instrument Hosts	No	0.715
	Instrument Hosts	Yes	1.675
	Data Sets	No	0.937
	Data Sets	Yes	1.455

● Authors
■ Data Sets

Affiliation networks lead to triangle motifs and high clustering coefficients

Network of authors with data sets as edges

Newman-Type Table for the 12 PDS Networks

Type of Node	Network	Weighted?	n	m	$<k>$	l	$\frac{\log n}{\log <k>}$	α	$C^{(2)}$	$<k>/n$	r
Authors as Nodes	PDS Nodes	No	439	27493	125.3	1.886	1.260	-0.31	0.981	0.285	0.80
	PDS Nodes	Yes	439	27527	125.4	1.886	1.259	-0.31	0.988	0.286	0.80
	Instrument Hosts	No	439	8240	37.5	2.703	1.678	-0.49	0.929	0.086	0.68
	Instrument Hosts	Yes	439	8581	39.1	2.703	1.660	-0.48	1.092	0.089	0.55
	Data Sets	No	439	3240	14.8	3.1	2.260	-0.60	0.936	0.034	0.96
	Data Sets	Yes	439	4366	19.9	3.1	2.035	-0.53	1.534	0.045	0.95
Events as Nodes	PDS Nodes	No	9	11	2.4	1.861	2.458	-1.1	0.62	0.272	-0.70
	PDS Nodes	Yes	9	28	6.2	1.861	1.202	-0.73	1.927	0.691	-0.52
	Instrument Hosts	No	103	282	5.5	2.426	2.726	--	0.715	0.053	0.039
	Instrument Hosts	Yes	103	520	10.1	2.426	2.004	--	1.675	0.098	-0.042
	Data Sets	No	1046	5820	11.1	2.761	2.886	-0.27	0.937	0.011	0.99
	Data Sets	Yes	1046	6514	12.5	2.761	2.757	-0.29	1.455	0.012	0.97

- ## Network-Specific Results
  - Small-magnitude α values, but the regions that follow a power law are small for many of the networks
  - Short path lengths and high clustering coefficients
  - **Degree Correlation**
    - **"Social" and "information" networks have r > 0**
    - "Technological" networks have r < 0 (except in one case where r << 1)

Newman-Type Table for the 12 PDS Networks

Type of Node	Network	Weighted?	n	m	$<k>$	l	$\frac{\log n}{\log <k>}$	α	$C^{(2)}$	$<k>/n$	r
Authors as Nodes	PDS Nodes	No	439	27493	125.3	1.886	1.260	-0.31	0.981	0.285	0.80
	PDS Nodes	Yes	439	27527	125.4	1.886	1.259	-0.31	0.988	0.286	0.80
	Instrument Hosts	No	439	8240	37.5	2.703	1.678	-0.49	0.929	0.086	0.68
	Instrument Hosts	Yes	439	8581	39.1	2.703	1.660	-0.48	1.092	0.089	0.55
	Data Sets	No	439	3240	14.8	3.1	2.260	-0.60	0.936	0.034	0.96
	Data Sets	Yes	439	4366	19.9	3.1	2.035	-0.53	1.534	0.045	0.95
Events as Nodes	PDS Nodes	No	9	11	2.4	1.861	2.458	-1.1	0.62	0.272	-0.70
	PDS Nodes	Yes	9	28	6.2	1.861	1.202	-0.73	1.927	0.691	-0.52
	Instrument Hosts	No	103	282	5.5	2.426	2.726	--	0.715	0.053	0.039
	Instrument Hosts	Yes	103	520	10.1	2.426	2.004	--	1.675	0.098	-0.042
	Data Sets	No	1046	5820	11.1	2.761	2.886	-0.27	0.937	0.011	0.99
	Data Sets	Yes	1046	6514	12.5	2.761	2.757	-0.29	1.455	0.012	0.97

- **Network-Specific Results**
  - Small α values, but the regions that follow a power law are small for many of the networks
  - Short path lengths and high clustering coefficients
  - **Degree Correlation**
    - "Social" and "information" networks have r > 0
    - **"Technological" networks have r < 0 (except in one case where r << 1)**

Newman-Type Table for the 12 PDS Networks

Type of Node	Network	Weighted?	n	m	$<k>$	l	$\frac{\log n}{\log <k>}$	α	$C^{(2)}$	$<k>/n$	r
Authors as Nodes	PDS Nodes	No	439	27493	125.3	1.886	1.260	-0.31	0.981	0.285	0.80
	PDS Nodes	Yes	439	27527	125.4	1.886	1.259	-0.31	0.988	0.286	0.80
	Instrument Hosts	No	439	8240	37.5	2.703	1.678	-0.49	0.929	0.086	0.68
	Instrument Hosts	Yes	439	8581	39.1	2.703	1.660	-0.48	1.092	0.089	0.55
	Data Sets	No	439	3240	14.8	3.1	2.260	-0.60	0.936	0.034	0.96
	Data Sets	Yes	439	4366	19.9	3.1	2.035	-0.53	1.534	0.045	0.95
Events as Nodes	PDS Nodes	No	9	11	2.4	1.861	2.458	-1.1	0.62	0.272	-0.70
	PDS Nodes	Yes	9	28	6.2	1.861	1.202	-0.73	1.927	0.691	-0.52
	Instrument Hosts	No	103	282	5.5	2.426	2.726	--	0.715	0.053	0.039
	Instrument Hosts	Yes	103	520	10.1	2.426	2.004	--	1.675	0.098	-0.042
	Data Sets	No	1046	5820	11.1	2.761	2.886	-0.27	0.937	0.011	0.99
	Data Sets	Yes	1046	6514	12.5	2.761	2.757	-0.29	1.455	0.012	0.97

- **General Observations**
  - Results depend on the choice of how to represent the network
  - Weighting can give a more accurate picture but can also cause the metrics to lose their meaning

Newman-Type Table for the 12 PDS Networks

| Type of Node	Network	Weighted?	n	m	$<k>$	l	$\log n / \log <k>$	α	$C^{(2)}$	$<k>/n$	r
Authors as Nodes	PDS Nodes	No	439	27493	125.3	1.886	1.260	-0.31	0.981	0.285	0.80
	PDS Nodes	Yes	439	27527	125.4	1.886	1.259	-0.31	0.988	0.286	0.80
	Instrument Hosts	No	439	8240	37.5	2.703	1.678	-0.49	0.929	0.086	0.55
	Instrument Hosts	Yes	439	8581	39.1	2.703	1.660	-0.48	1.092	0.089	0.55
	Data Sets	No	439	3240	14.8	3.1	2.260	-0.60	0.936	0.034	0.96
	Data Sets	Yes	439	4366	19.9	3.1	2.035	-0.53	1.534	0.045	0.95
Events as Nodes	PDS Nodes	No	9	11	2.4	1.861	2.458	-1.1	0.62	0.272	-0.70
	PDS Nodes	Yes	9	28	6.2	1.861	1.202	-0.73	1.927	0.691	-0.52
	Instrument Hosts	No	103	282	5.5	2.426	2.726	--	0.715	0.053	0.655
	Instrument Hosts	Yes	103	520	10.1	2.426	2.004	--	1.675	0.098	-0.042
	Data Sets	No	1046	5820	11.1	2.761	2.886	-0.27	0.937	0.011	0.99
	Data Sets	Yes	1046	6514	12.5	2.761	2.757	-0.29	1.455	0.012	0.97

- **General Observations**
  - **Results depend on the choice of how to represent the network**
 - Weighting can give a more accurate picture but can also cause the metrics to lose their meaning

# Newman-Type Table for the 12 PDS Networks

Type of Node	Network	Weighted?	n	m	$<k>$	l	$\frac{\log n}{\log <k>}$	α	$C^{(2)}$	$<k>/n$	r
Authors as Nodes	PDS Nodes	No	439	27493	125.3	1.886	1.260	-0.31	0.981	0.285	0.80
	PDS Nodes	Yes	439	27527	125.4	1.886	1.259	-0.31	0.988	0.286	0.80
	Instrument Hosts	No	439	8240	37.5	2.703	1.678	-0.49	0.929	0.086	0.68
	Instrument Hosts	Yes	439	8581	39.1	2.703	1.660	-0.48	1.092	0.089	0.55
	Data Sets	No	439	3240	14.8	3.1	2.260	-0.60	0.936	0.034	0.96
	Data Sets	Yes	439	4366	19.9	3.1	2.035	-0.53	1.534	0.045	0.95
Events as Nodes	PDS Nodes	No	9	11	2.4	1.861	2.458	-1.1	0.62	0.272	-0.70
	PDS Nodes	Yes	9	28	6.2	1.861	1.202	-0.73	1.927	0.691	-0.52
	Instrument Hosts	No	103	282	5.5	2.426	2.726	--	0.715	0.053	0.039
	Instrument Hosts	Yes	103	520	10.1	2.426	2.004	--	1.675	0.098	-0.042
	Data Sets	No	1046	5820	11.1	2.761	2.886	-0.27	0.937	0.011	0.99
	Data Sets	Yes	1046	6514	12.5	2.761	2.757	-0.29	1.455	0.012	0.97

- **General Observations**
  - Results depend on the choice of how to represent the network
  - **Weighting can give a more accurate picture but can also cause the metrics to lose their meaning**
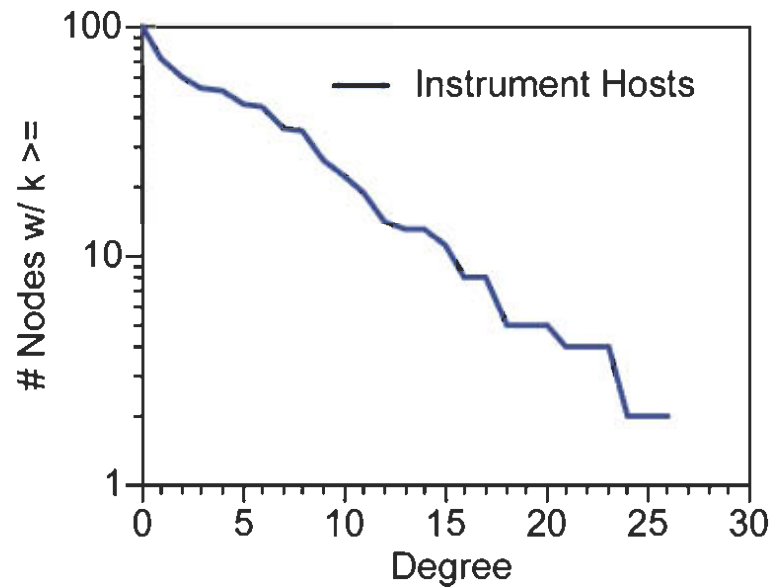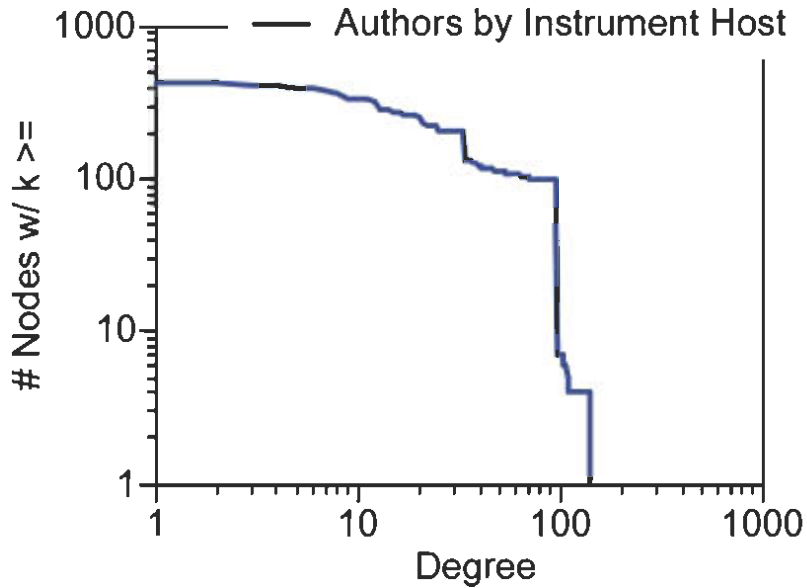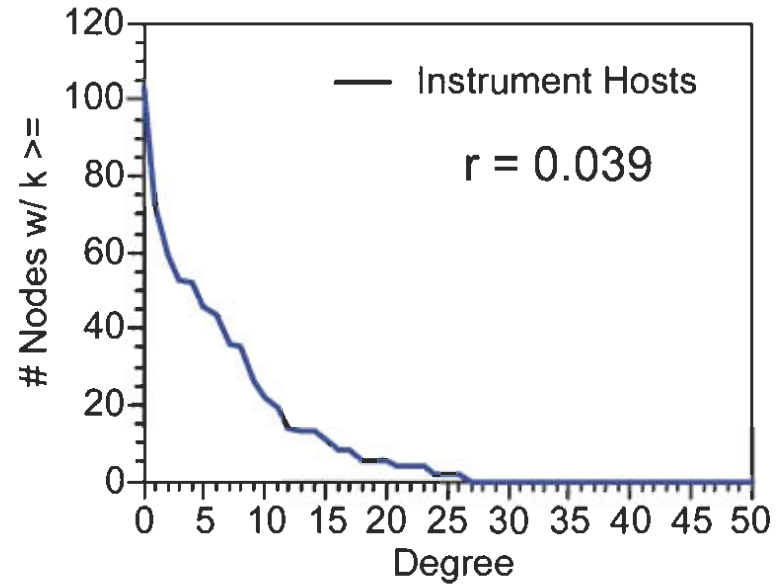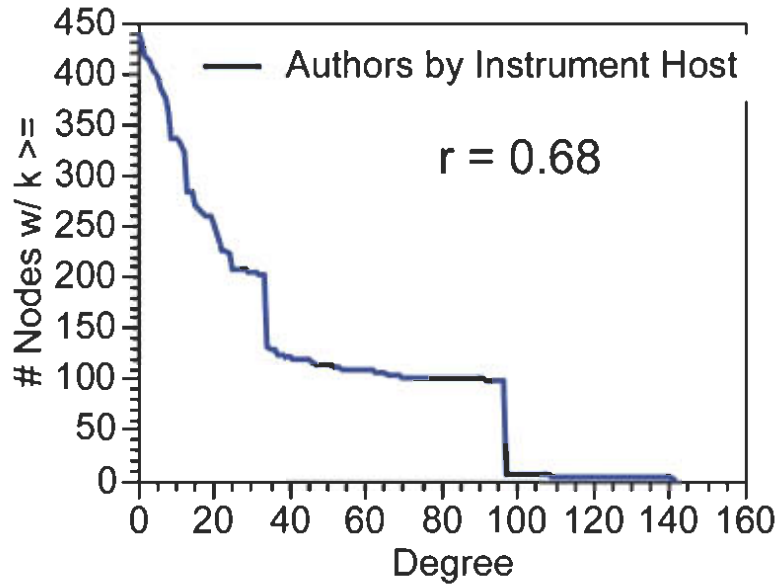
Completely Connected Areas

==> High clusteri[ng]
coefficients

and

High Pearson deg[ree]
correlation

Degree Distributions

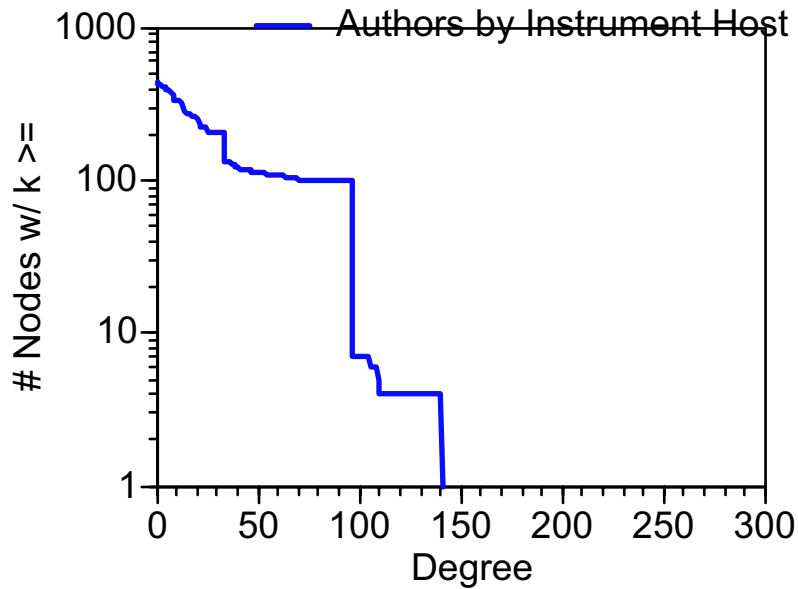Authors by Dataset

Datasets
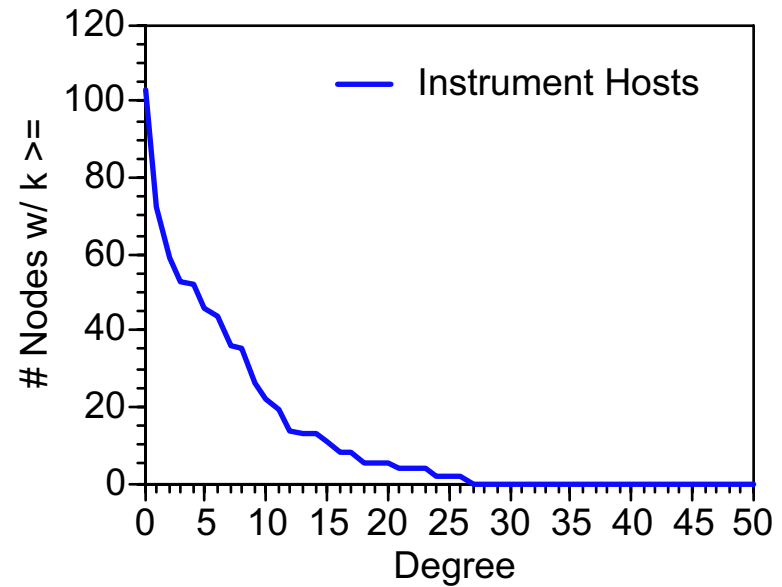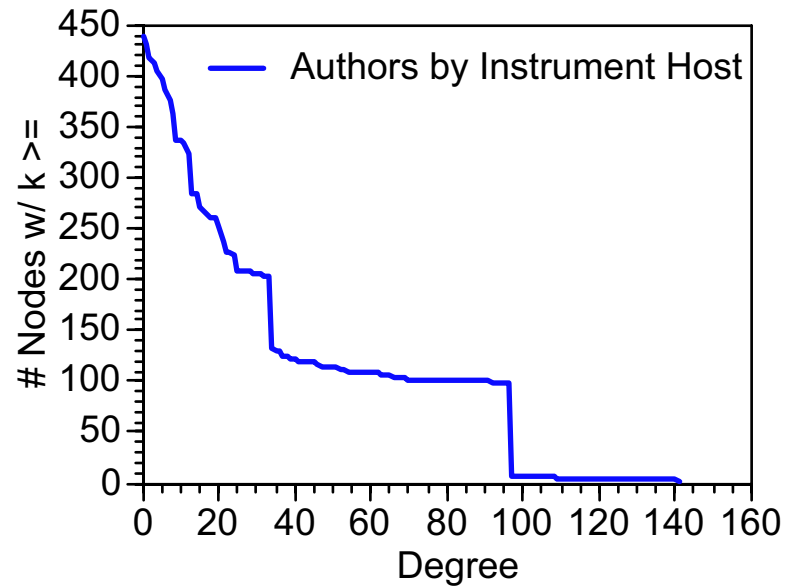
Degree Correlation

r = 0.96

r = 0.99
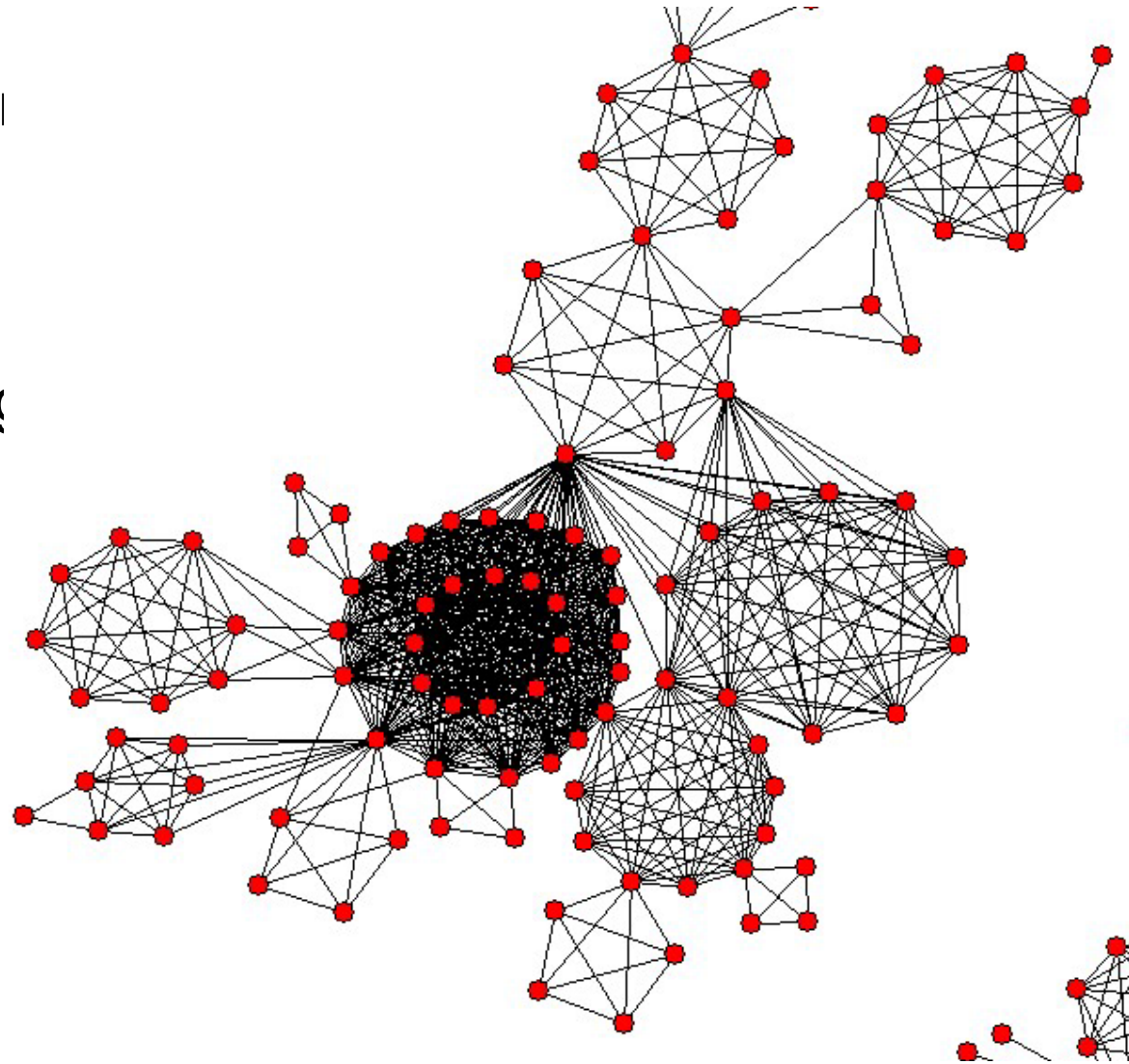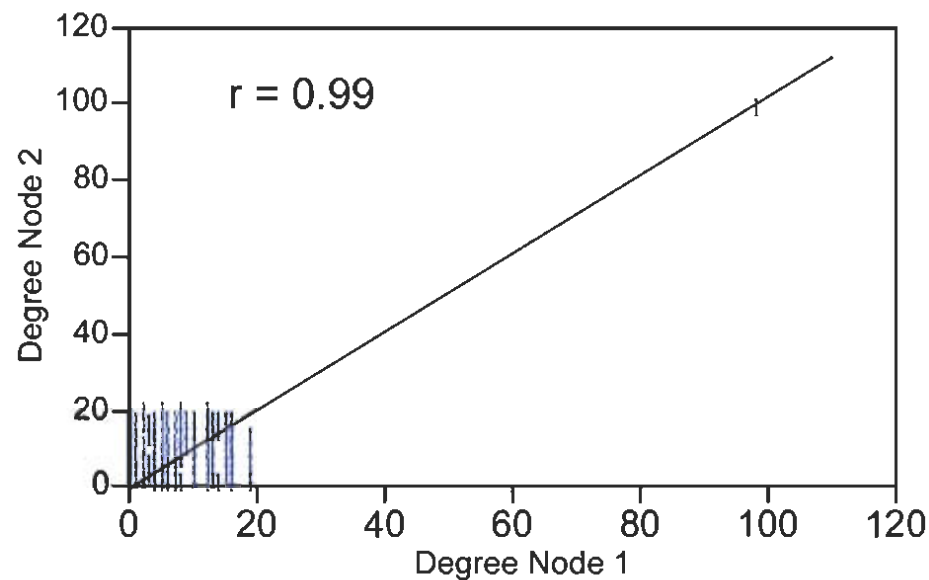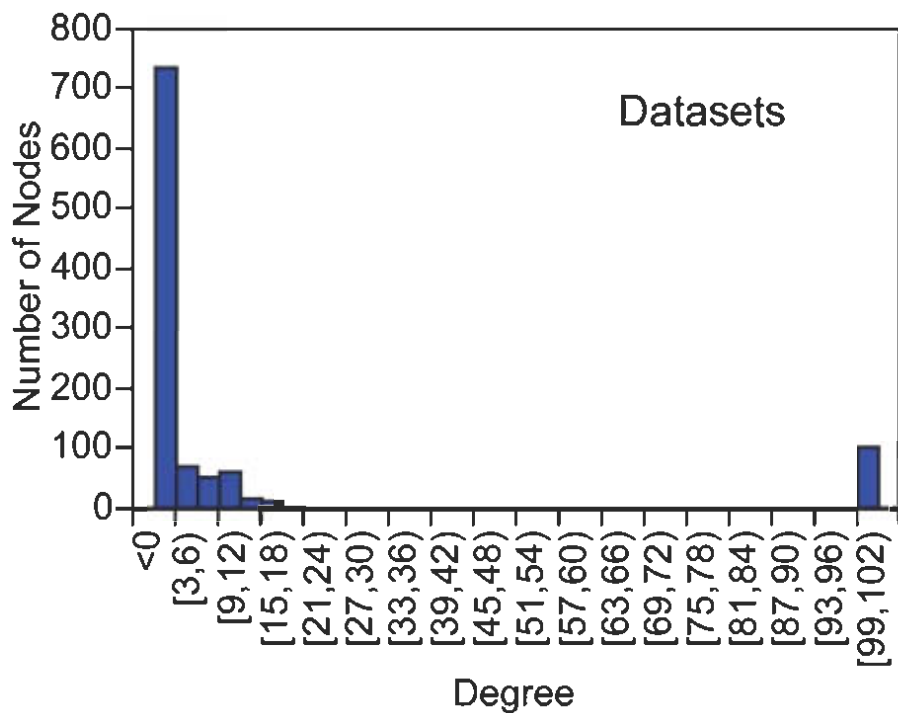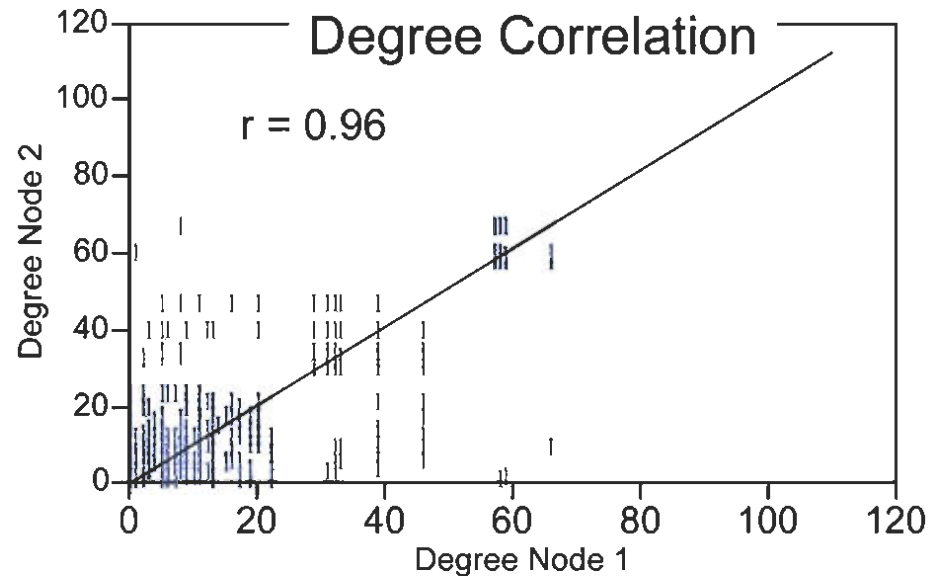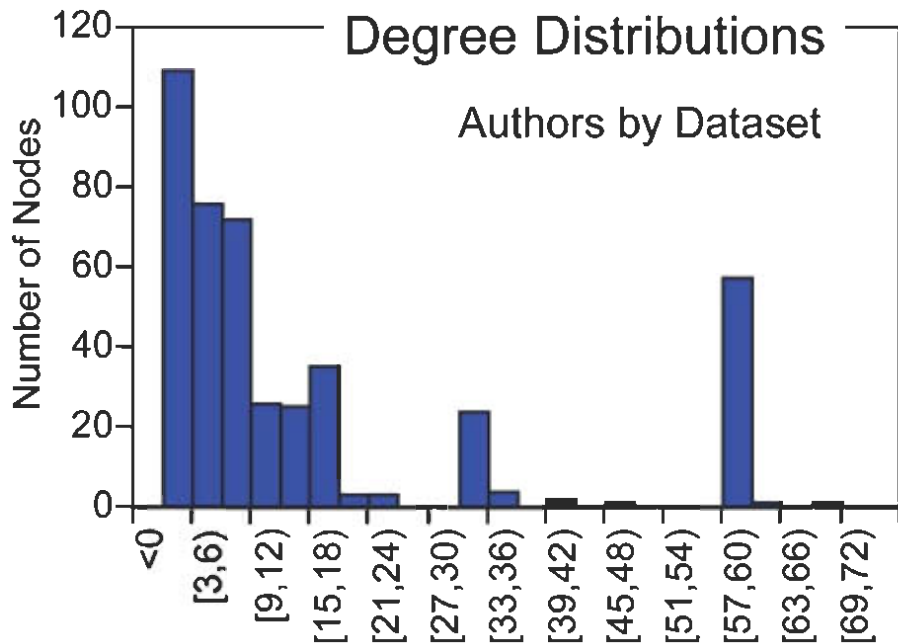
Degree Distributions

Cumulative Degree Distributions
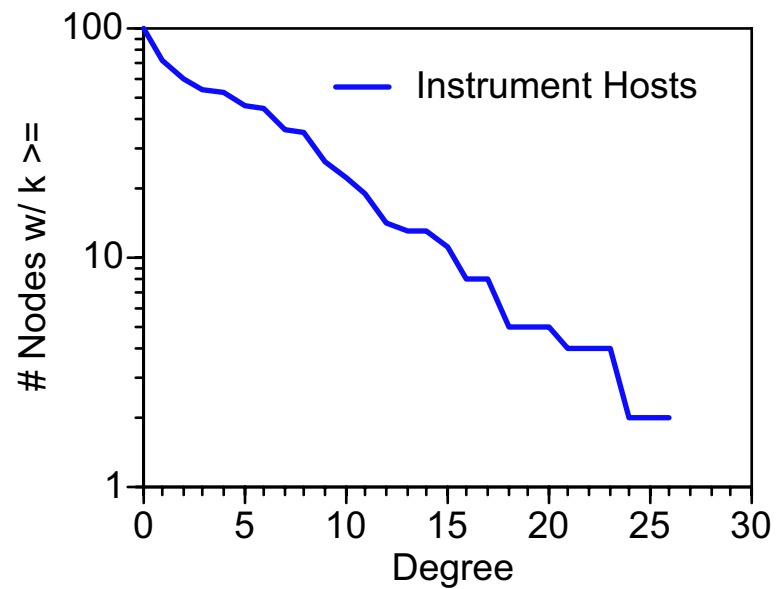
Community Structure

Do PDS nodes represent
communities? NO

Node size: Betweenness
Color: PDS Node Association
Shape: Newman-Girvan

Community Structure

Do Instrument Hosts
represent communities?
YES

Node size: Betweenness
Color: Instrument Host Association
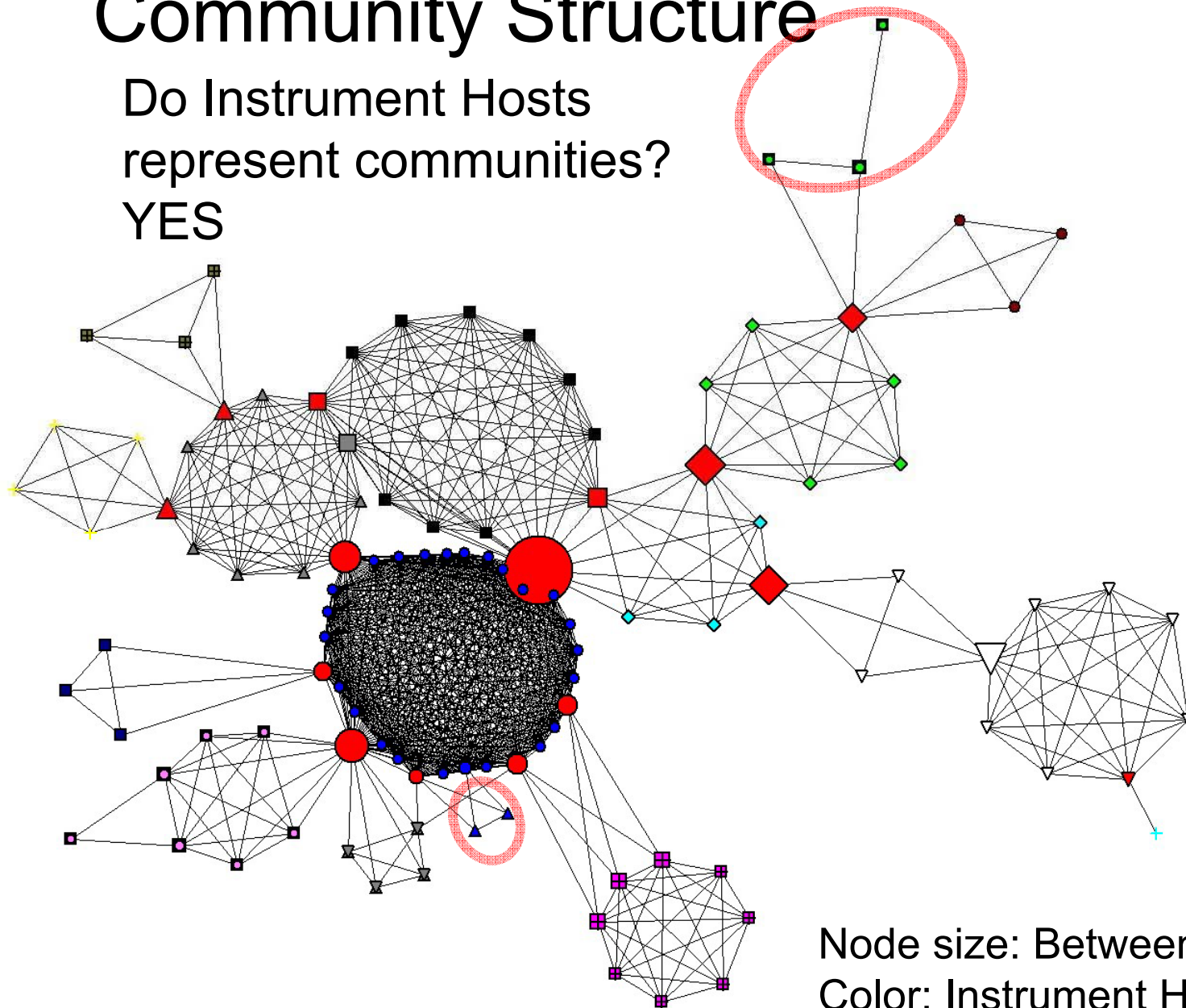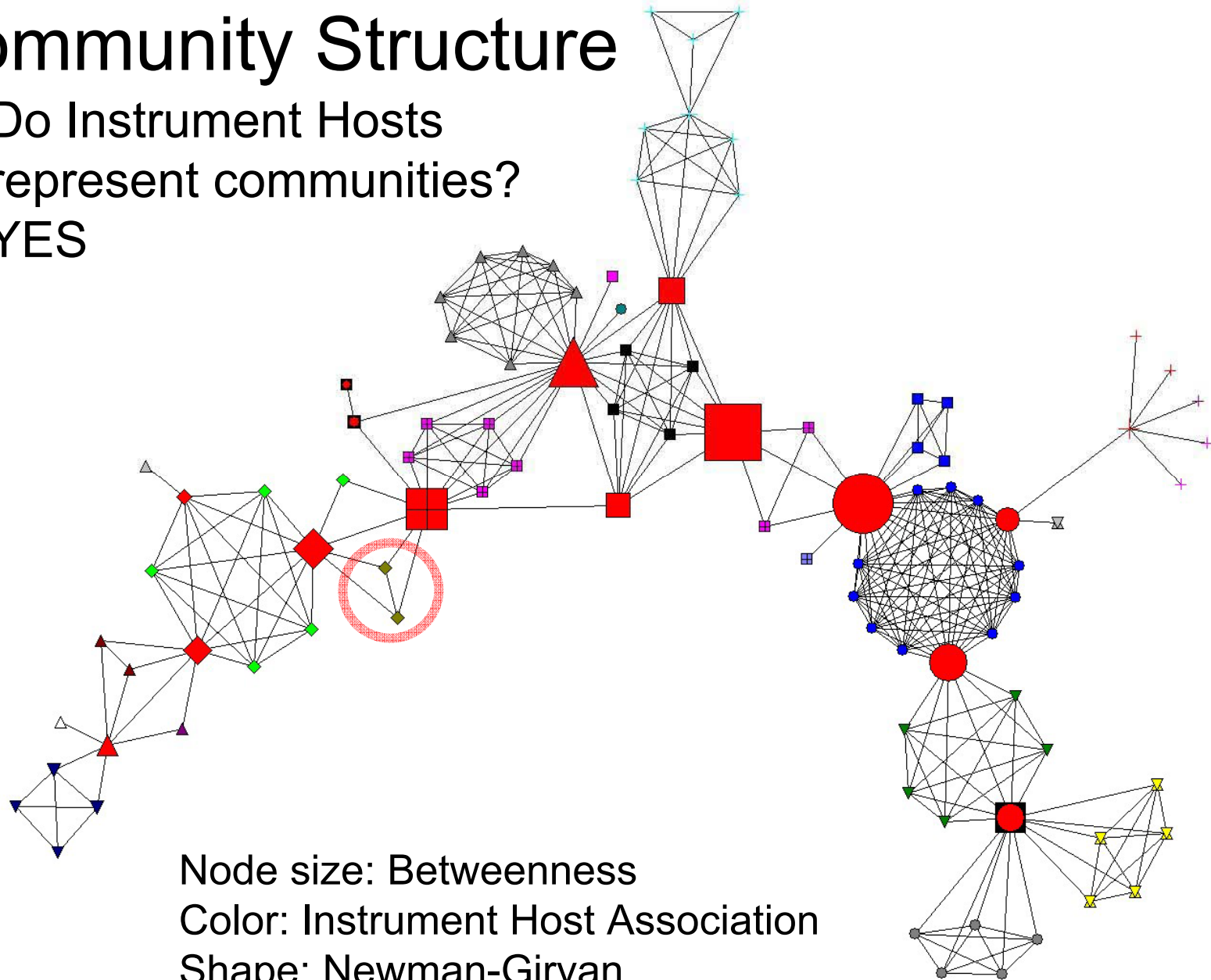Shape: Newman-Girvan

Community Structure

Do Instrument Hosts
represent communities?
YES

Node size: Betweenness
Color: Instrument Host Association
Shape: Newman-Girvan

Centrality and Network Representation

Type of Node	Network	Weighted?	Centrality Measure			
			Degree	Closeness	Betweenness	Eigenvector
Authors as Nodes	PDS Nodes	No	0.339	0.364	0.136	0.060
	PDS Nodes	Yes	0.172	0.364	0.136	0.062
	Instrument Hosts	No	0.237	Unconnected	0.147	0.145
	Instrument Hosts	Yes	0.043	Unconnected	0.147	0.276
	Data Sets	No	0.120	Unconnected	0.026	0.354
	Data Sets	Yes	0.026	Unconnected	0.026	0.014
Events as Nodes	PDS Nodes	No	0.571	0.578	0.605	0.563
	PDS Nodes	Yes	0.314	0.578	0.605	2.152
	Instrument Hosts	No	0.205	Unconnected	0.074	0.385
	Instrument Hosts	Yes	0.046	Unconnected	0.074	1.555
	Data Sets	No	0.084	Unconnected	0.004	0.132
	Data Sets	Yes	0.006	Unconnected	0.004	0.512

- As before, weighting can affect the results.
  - Higher centrality without weighting.
  - Some measures are affected, and others are not.
- Some measures are not valid for some networks.
- Centrality tends to be highest for the network of PDS Nodes (with authors as edges).

Centrality and Network Representation

Type of Node	Network	Weighted?	Centrality Measure			
			Degree	Closeness	Betweenness	Eigenvector
Authors as Nodes	PDS Nodes	No	0.339	0.364	0.136	0.060
	PDS Nodes	Yes	0.172	0.364	0.136	0.062
	Instrument Hosts	No	0.237	Unconnected	0.147	0.145
	Instrument Hosts	Yes	0.043	Unconnected	0.147	0.276
	Data Sets	No	0.120	Unconnected	0.026	0.354
	Data Sets	Yes	0.026	Unconnected	0.026	0.014
Events as Nodes	PDS Nodes	No	0.571	0.578	0.605	0.563
	PDS Nodes	Yes	0.314	0.578	0.605	2.152
	Instrument Hosts	No	0.205	Unconnected	0.074	0.385
	Instrument Hosts	Yes	0.046	Unconnected	0.074	1.555
	Data Sets	No	0.084	Unconnected	0.004	0.132
	Data Sets	Yes	0.006	Unconnected	0.004	0.512

- **As before, weighting can affect the results.**
  - **Higher centrality without weighting.**
  - Some measures are affected, and others are not.
- Some measures are not valid for some networks.
- Centrality tends to be highest for the network of PDS Nodes (with authors as edges).

Centrality and Network Representation

Type of Node	Network	Weighted?	Centrality Measure			
			Degree	Closeness	Betweenness	Eigenvector
Authors as Nodes	PDS Nodes	No	0.339	0.364	0.136	0.060
	PDS Nodes	Yes	0.172	0.364	0.136	0.062
	Instrument Hosts	No	0.237	Unconnected	0.147	0.145
	Instrument Hosts	Yes	0.043	Unconnected	0.147	0.276
	Data Sets	No	0.120	Unconnected	0.026	0.354
	Data Sets	Yes	0.026	Unconnected	0.026	0.014
Events as Nodes	PDS Nodes	No	0.571	0.578	0.605	0.563
	PDS Nodes	Yes	0.314	0.578	0.605	2.152
	Instrument Hosts	No	0.205	Unconnected	0.074	0.385
	Instrument Hosts	Yes	0.046	Unconnected	0.074	1.555
	Data Sets	No	0.084	Unconnected	0.004	0.132
	Data Sets	Yes	0.006	Unconnected	0.004	0.512

- **As before, weighting can affect the results.**
  - Higher centrality without weighting.
  - **Some measures are affected, and others are not.**

- Some measures are not valid for some networks.

- Centrality tends to be highest for the network of PDS Nodes (with authors as edges).

Centrality and Network Representation

Type of Node	Network	Weighted?	Centrality Measure			
			Degree	Closeness	Betweenness	Eigenvector
Authors as Nodes	PDS Nodes	No	0.339	0.364	0.136	0.060
	PDS Nodes	Yes	0.172	0.364	0.136	0.062
	Instrument Hosts	No	0.237	Unconnected	0.147	0.145
	Instrument Hosts	Yes	0.043	Unconnected	0.147	0.276
	Data Sets	No	0.120	Unconnected	0.026	0.354
	Data Sets	Yes	0.026	Unconnected	0.026	0.014
Events as Nodes	PDS Nodes	No	0.571	0.578	0.605	0.563
	PDS Nodes	Yes	0.314	0.578	0.605	2.152
	Instrument Hosts	No	0.205	Unconnected	0.074	0.386
	Instrument Hosts	Yes	0.046	Unconnected	0.074	1.555
	Data Sets	No	0.084	Unconnected	0.004	0.152
	Data Sets	Yes	0.006	Unconnected	0.004	0.512

- As before, weighting can affect the results.
  - Small Bodies is a much better center without weighting.
  - Some measures are affected, and others are not.
- **Some measures are not valid for some networks.**
- Centrality tends to be highest for the network of PDS Nodes (with authors as edges).

Centrality and Network Representation

Type of Node	Network	Weighted?	Centrality Measure			
			Degree	Closeness	Betweenness	Eigenvector
Authors as Nodes	PDS Nodes	No	0.339	0.364	0.136	0.060
	PDS Nodes	Yes	0.172	0.364	0.136	0.062
	Instrument Hosts	No	0.237	Unconnected	0.147	0.145
	Instrument Hosts	Yes	0.043	Unconnected	0.147	0.276
	Data Sets	No	0.120	Unconnected	0.026	0.354
	Data Sets	Yes	0.026	Unconnected	0.026	0.014
Events as Nodes	PDS Nodes	No	0.571	0.578	0.605	0.563
	PDS Nodes	Yes	0.314	0.578	0.605	2.152
	Instrument Hosts	No	0.205	Unconnected	0.074	0.385
	Instrument Hosts	Yes	0.046	Unconnected	0.074	1.555
	Data Sets	No	0.084	Unconnected	0.004	0.132
	Data Sets	Yes	0.006	Unconnected	0.004	0.512

- As before, weighting can affect the results.
    - Higher centrality without weighting.
    - Some measures are affected, and others are not.
- Some measures are not valid for some networks.
- **Centrality tends to be highest for the network of PDS Nodes (with authors as edges).**

What's Your Szego Number?

Overall Best Centers						
			Authors			
PDS Nodes	Instrument Hosts	Data Sets	By PDS Nodes	By Instrument Hosts	By Data Sets	Overall
Small_Bodies	Voyager 2	MER1-M-MI-5-	T. Z. Martin	C. Neese	K. Szego	K. Szego
Planetary_Atmospheres	Ground-Based Telescopes	MOSAIC-OPS-V1.0	R. Mehlman	L. S. Elson	J. T. Gosling	J. T. Gosling
	Public Literature		J. R. Spencer	C. H. Acton	R. F. Beebe	
				B.V.Semenov		

- Top two "best centers" determined for each of the 12 networks by each of the 4 metrics
- The most commonly appearing nodes selected as the "best centers" for each network (above)
  - Not surprisingly, two catch-alls are among the best centers (2nd and third) of instrument hosts
- K. Szego is the Erdös of the Planetary Data System

What's Your Szego Number?

Overall Best Centers						
			Authors			
PDS Nodes	Instrument Hosts	Data Sets	By PDS Nodes	By Instrument Hosts	By Data Sets	Overall
Small_Bodies	Voyager 2	MER1-M-MI-5-	T. Z. Martin	C. Neese	K. Szego	K. Szego
Planetary_Atmospheres	Ground-Based Telescopes	MOSAIC-OPS-V1.0	R. Mehlman	L. S. Elson	J. T. Gosling	J. T. Gosling
	Public Literature		J. R. Spencer	C. H. Acton	R. F. Beebe	
				B.V.Semenov		

- Top two "best centers" determined for each of the 12 networks by each of the 4 metrics
- The most commonly appearing nodes selected as the "best centers" for each network (above)
  – **Not surprisingly, two catch-alls are among the best centers (2nd and third) of instrument hosts**
- K. Szego is the Erdös of the Planetary Data System

What's Your Szego Number?

			Overall Best Centers			
				Authors		
PDS Nodes	Instrument Hosts	Data Sets	By PDS Nodes	By Instrument Hosts	By Data Sets	Overall
Small_Bodies	Voyager 2	MER1-M-MI-5-MOSAIC-OPS-V1.0	T. Z. Martin	C. Neese	K. Szego	K. Szego
Planetary_Atmospheres	Ground-Based Telescopes		R. Mehlman	L. S. Elson	J. T. Gosling	J. T. Gosling
	Public Literature		J. R. Spencer	C. H. Acton	R. F. Beebe	
				B.V.Semenov		

- Top two "best centers" determined for each of the 12 networks by each of the 4 metrics
- The most commonly appearing nodes selected as the "best centers" for each network (above)
  - Not surprisingly, two catch-alls are among the best centers (2nd and third) of instrument hosts
- **K. Szego is the Erdös of the Planetary Data System**

Future Work

- Look at the sub-network within one of the PDS nodes to determine centrality and community structure within a subfield.

- Consider PDS Data Access Statistics.
  - Only collected by NASA for access at the PDS node level by hostname (rather than individual). PDS staff may be able to collect data for data set and instrument host access, but it is not being done now.
  - We have collected the available usage statistics for some nodes (Small Bodies, Engineering, and PPI) and are in contact with all of the node managers.

- Create bipartite networks of authors and subject areas from a major planetary science conference.
  - American Astronomical Society's Division for Planetary Sciences (DPS) Annual Meeting has been suggested by R.P. Binzel (one of the authors here at MIT).
  - Compare results to those presented here. Are the community structure and centrality around instrument host and subject area specific to the PDS or more generalizable?

- Gather data on funding and political support for missions and for PDS nodes to determine the dynamics between scientists and policymakers and how that relates to the results presented here. (Possible component of Mark's research)