Mapping Collaborations Across Fields

EUGENE HWANG, AMIT TALAPATRA, MARISSA WIENER, WENDY ZHANG, WEI ZHENG

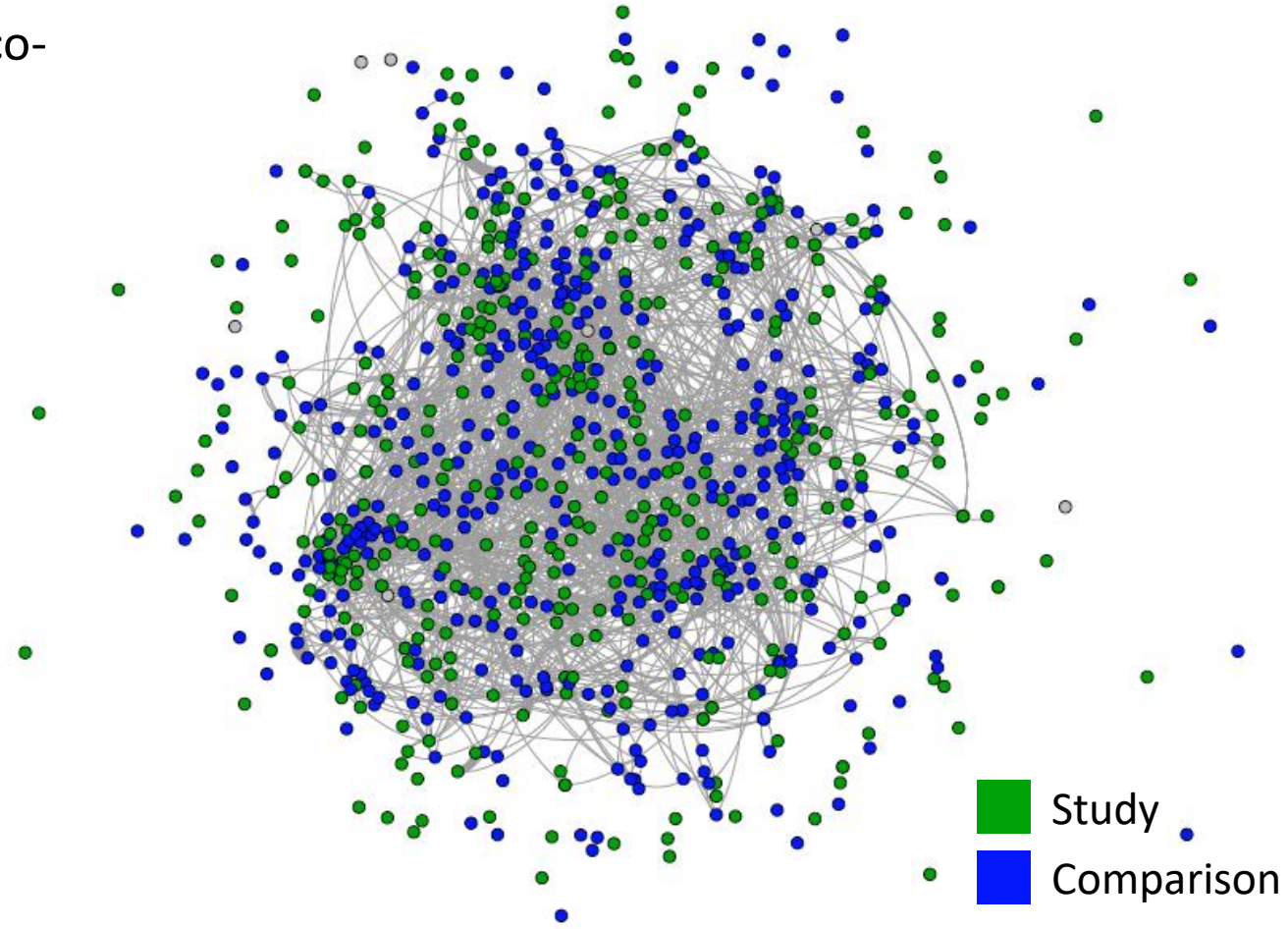
4/28/17

Problem Statement

- Investigating collaboration behavior in a network consisting of two groups of researchers:
  - Study Group: Researchers continually supported by NIGMS from 2001-2015.
  - Comparison Group: Researchers continually supported by NIH from 2001-2015.
- Data from NIH RePORTER, Elsevier Scopus, and Thomson Reuters Web of Science was compiled and filtered to identify the authors, publications, and projects relevant to our scope.
- Tools used: R, Python, igraph, Gephi.
- Study Questions:
  1. Do NIGMS-supported researchers collaborate more often with other NIGMS-supported researchers?
  2. Does the type of grant awarded influence collaboration behavior?
  3. Does variety in funding sources influence collaboration behavior?
  4. Does the scientific discipline influence collaboration behavior?

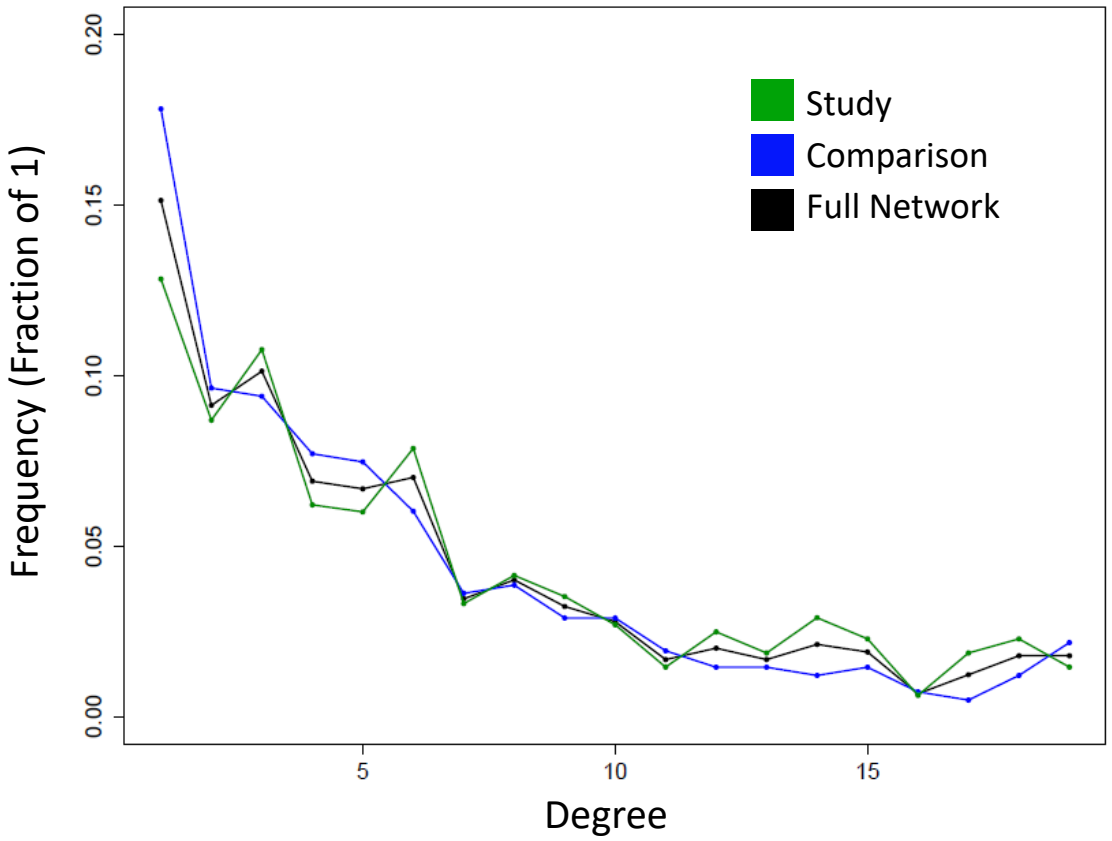
Network Overview

- A network of researchers, connected by co-authorship on publications.
- 898 Researchers (Nodes)
  - Study Group: 483
  - Comparison Group: 415
- 4004 Unique Publications, 4927 Edges
- Average Degree: 10.91

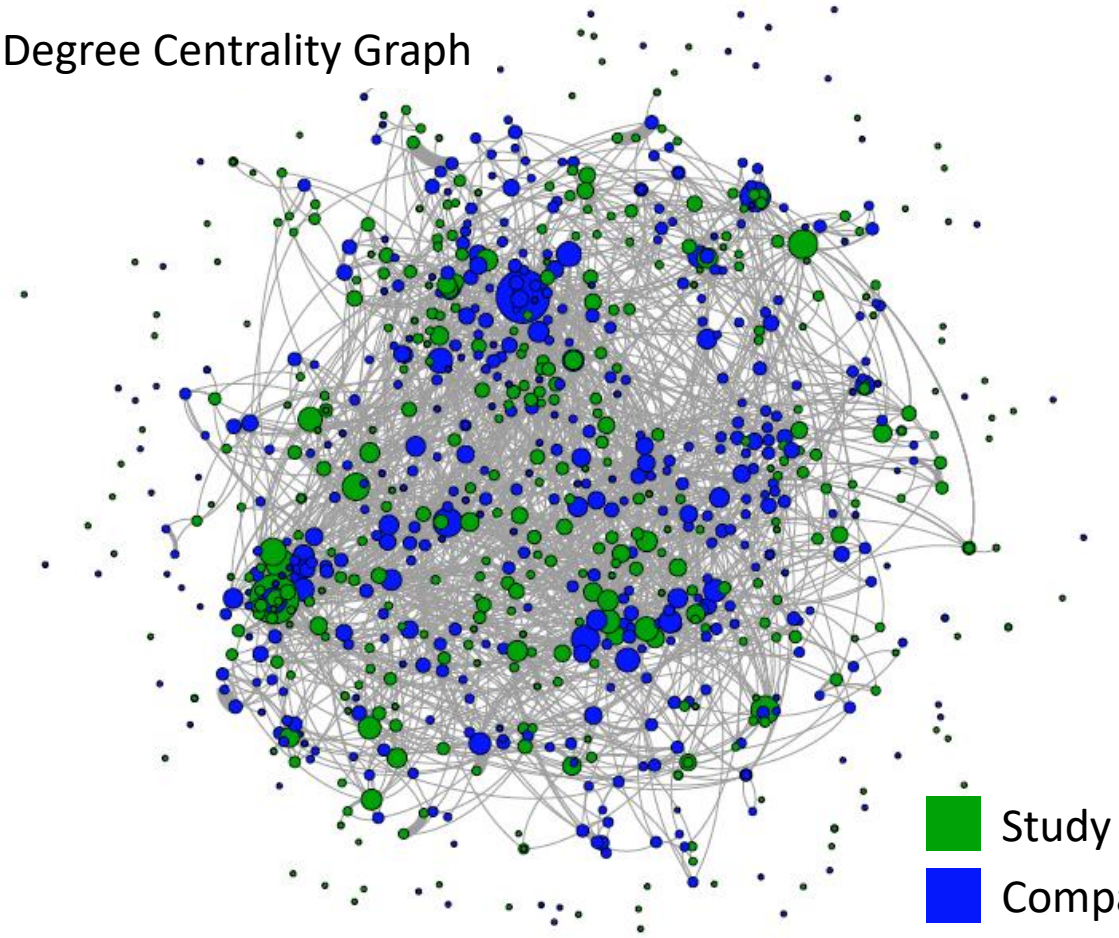


Study Question 1: Do NIGMS-supported researchers collaborate more often with other NIGMS-supported researchers?

Degree Distribution



Degree Centrality Graph

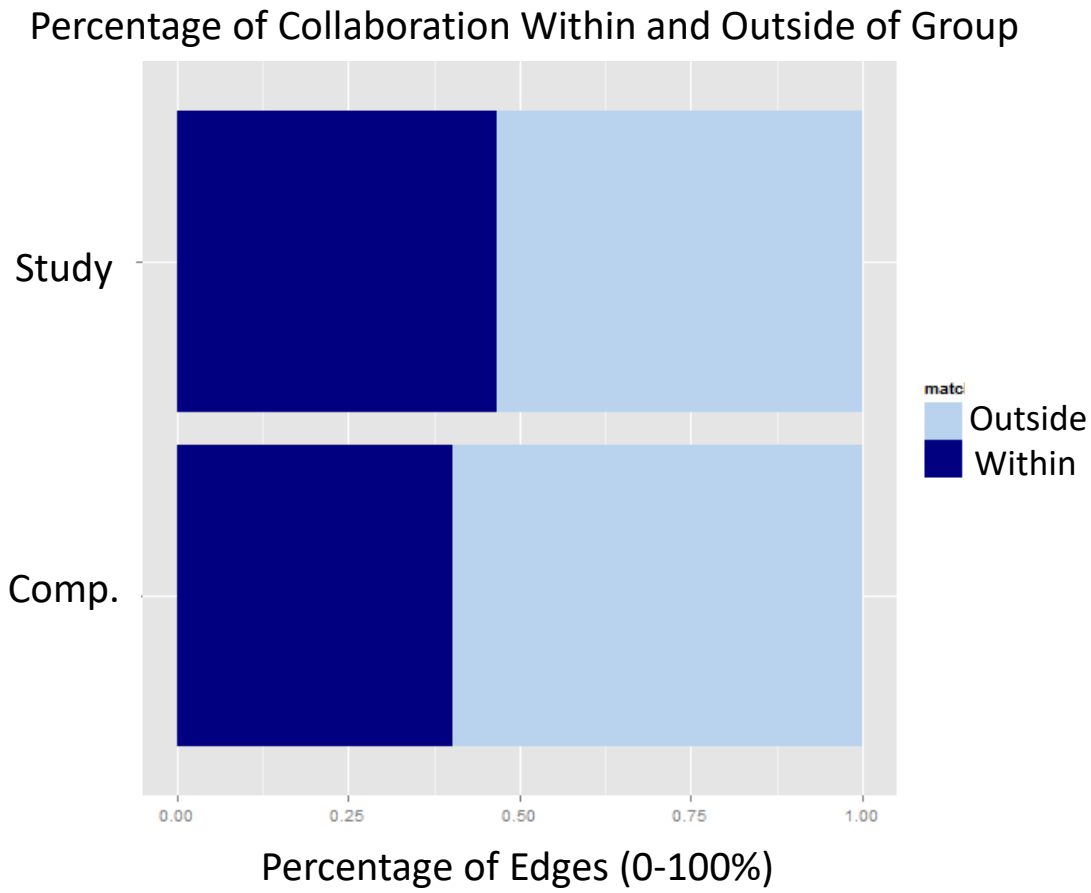


Study Question 1: Do NIGMS-supported researchers collaborate more often with other NIGMS-supported researchers?

- When grouping by study or comparison group, the assortativity was 0.21. This is positive, indicating that there is homophily by study or comparison group.
- Breakdown of connections between the groups:

Publications **Within** Study Group	Publications **Within** Comparison Group	Publications **Across** Groups
34%	**29%**	**37%**

- Complete triads:
  - Study Group: 200 (group of 483 nodes)
  - Comparison Group: 83 (group of 415 nodes)

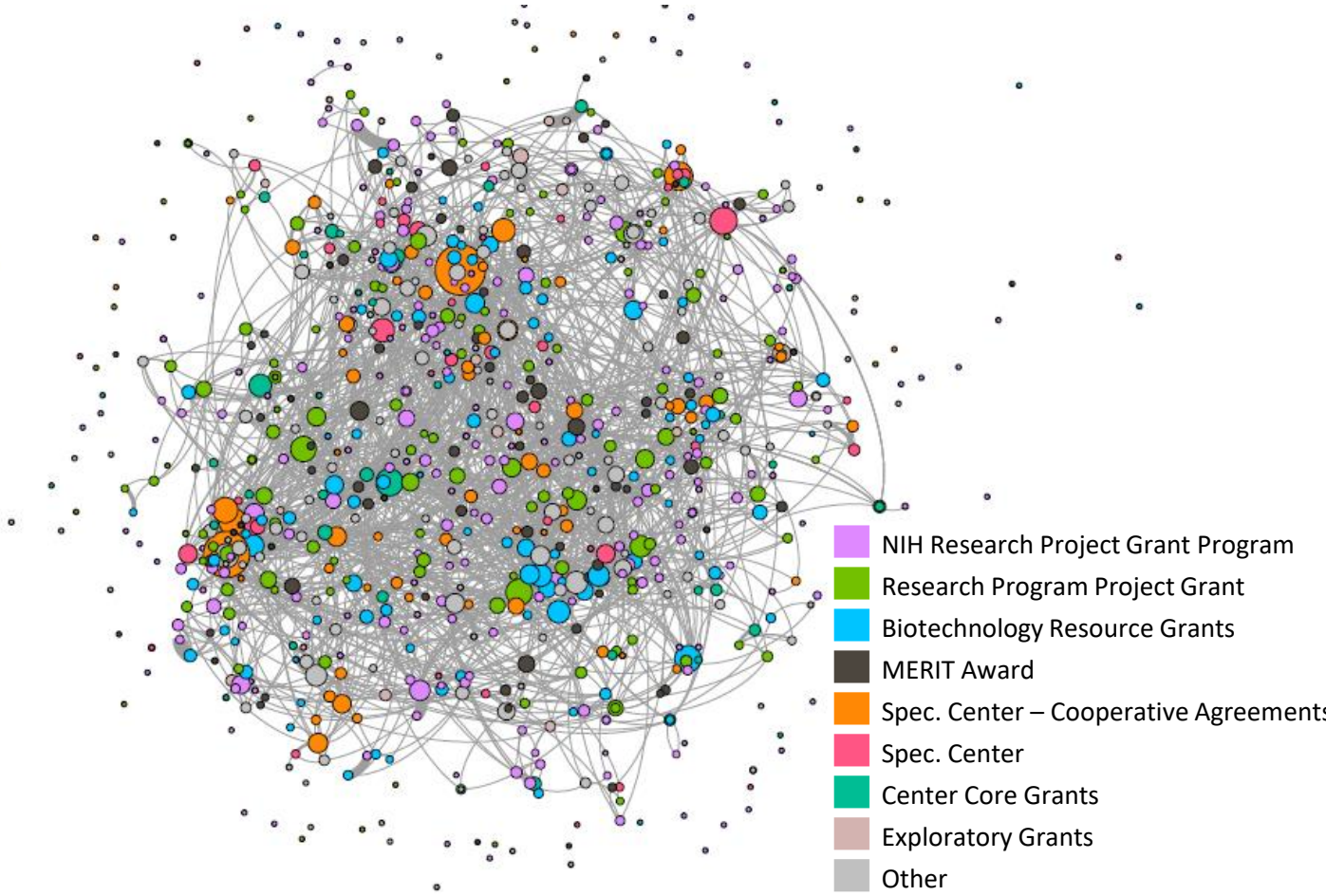


Study Question 1 Findings

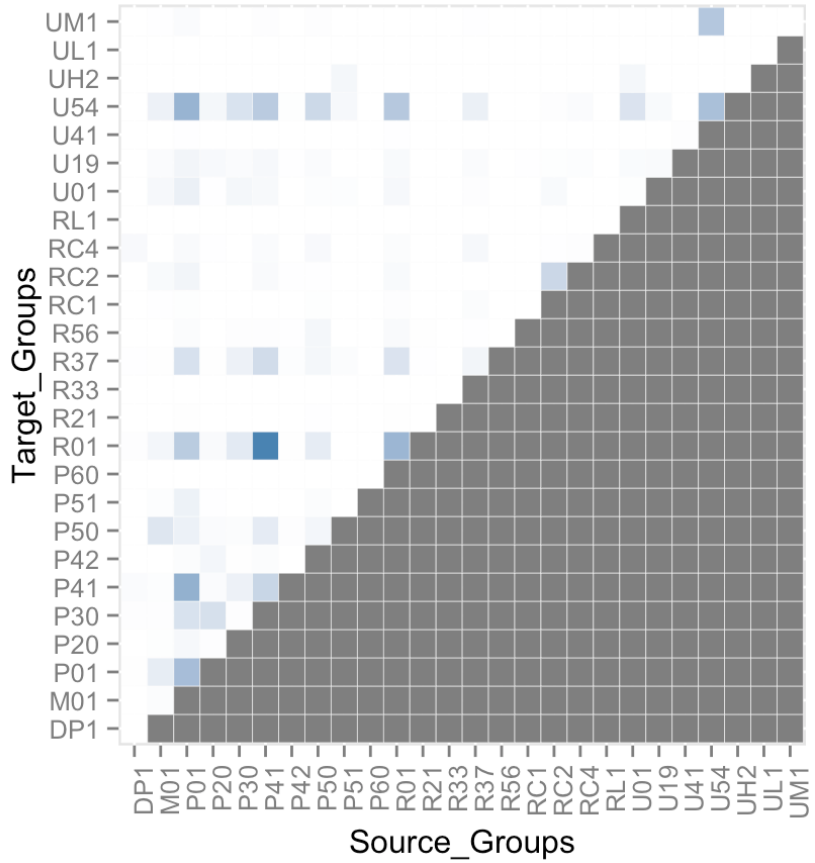
- Study Question 1: Do NIGMS-supported researchers collaborate more often with other NIGMS-supported researchers?
  - The network statistics show us that there is relatively more collaboration within groups than across groups.
  - Clustering is relatively higher within the study group than the comparison group (based on triads).
  - Conclusion: NIGMS-supported (study group) researchers collaborate more often with other NIGMS-supported researchers.

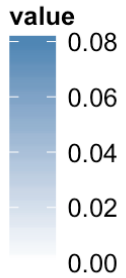
Study Question 2: Does the type of grant awarded influence collaboration behavior?

- To make use of node-level statistics, an activity code was applied to each researcher based on the amount of grant money attached to their projects.
- When grouping nodes by their most significant grant type, the assortativity was 0.10. This is positive, indicating that there is homophily by grant type groupings.

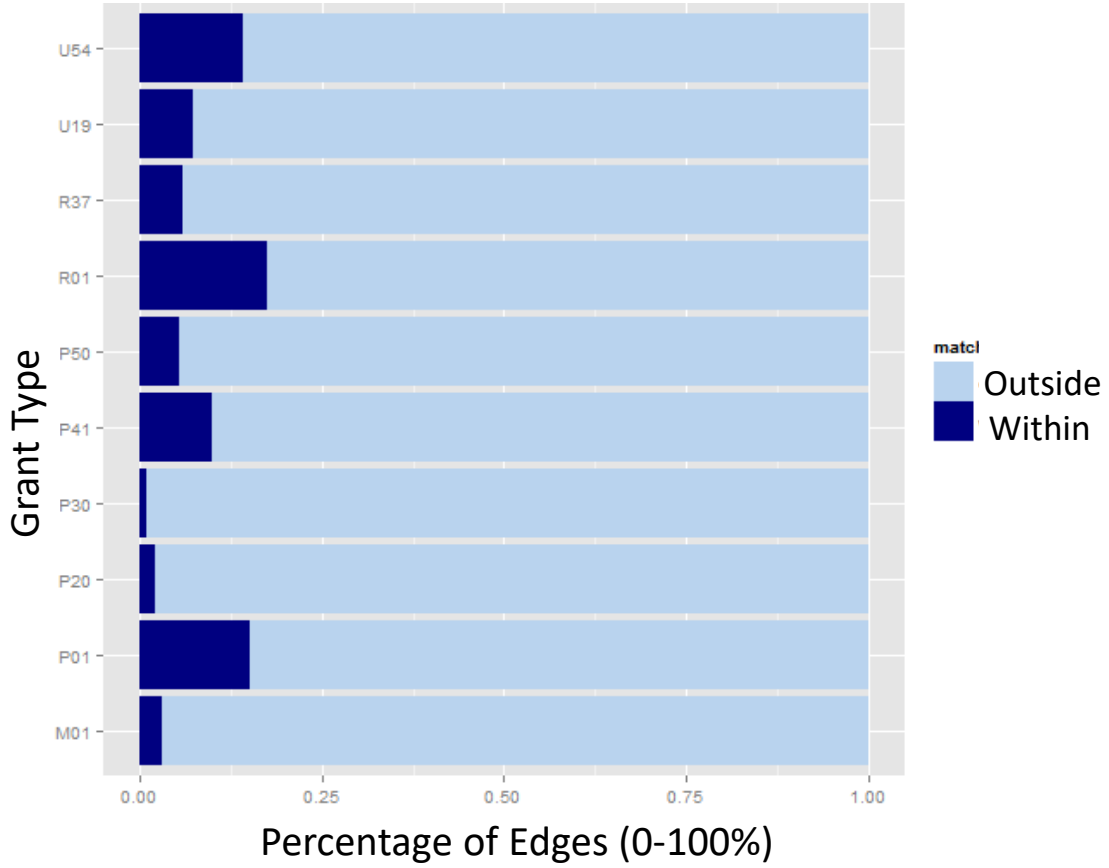


Study Question 2: Does the type of grant awarded influence collaboration behavior?



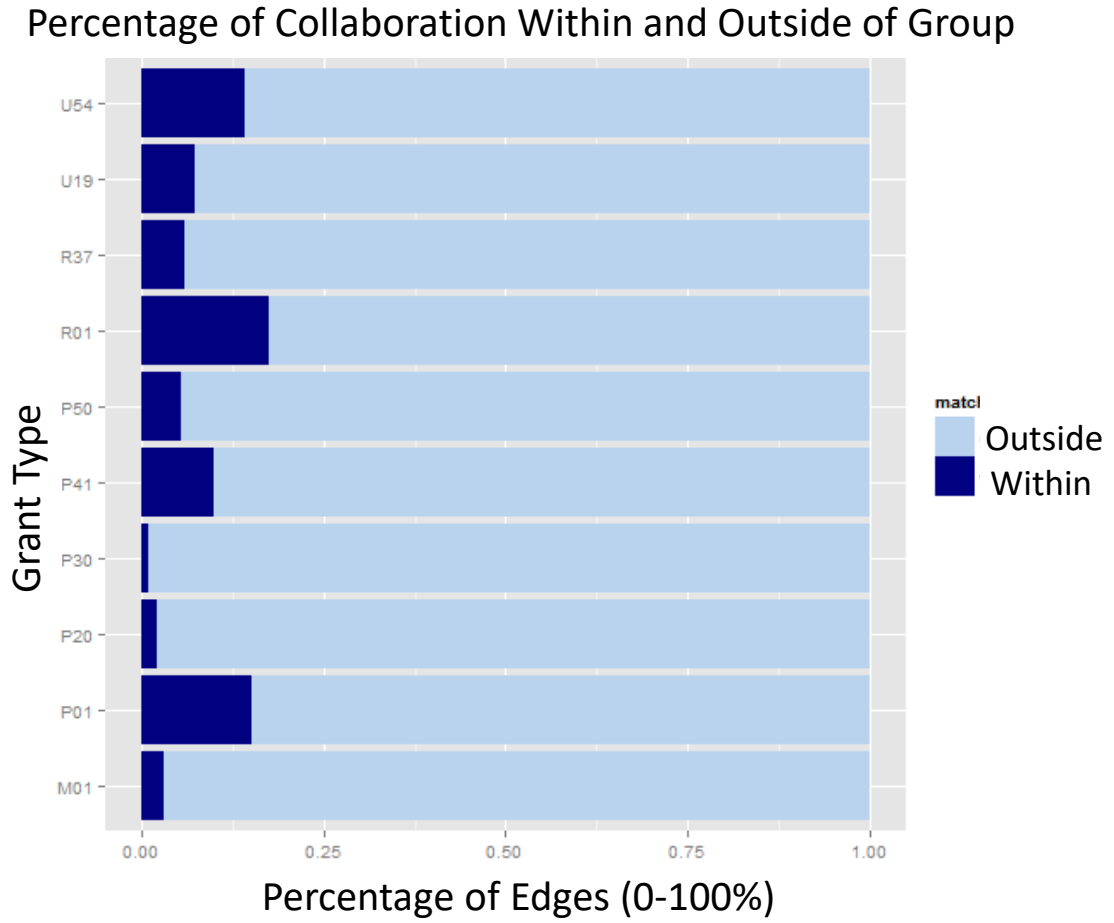


Percentage of Collaboration Within and Outside of Group



Study Question 2: Does the type of grant awarded influence collaboration behavior?

- As compared to the other grant types, researchers classified as the following tend to collaborate more within their respective grant type than with other grant types:
  - R01 - NIH Research Project Grant Program
  - P01 - Research Program Projects
  - U54 - Spec. Center - Cooperative Agreements
  - P41 - Biotechnology Resource Grants
- Researchers classified as the following tend to collaborate more outside of their own primary grant type.
  - P30 - Center Core Grants
  - P20 - Exploratory Grants
  - M01 - General Clinical Research Centers Program
  - P50 - Spec. Center
  - R37 - MERIT Award
  - U19 - Research Program--Cooperative Agreements



Study Question 2: Does the type of grant awarded influence collaboration behavior?

- Average Degree:

U54	U01	P41	M01	P50	P30	P01	P20	U19	R37	R01	R56
22.32	16.94	13.22	12.87	12.65	12.15	10.77	9.80	9.76	6.28	6.26	3.55

- Weighted Degree:

U54	U01	M01	P50	P30	U19	P41	P01	P20	R37	R01	R56
8.39	6.48	5.18	4.33	2.92	2.32	2.31	2.01	1.75	0.87	0.69	0.48

- Triad Count/Node Count:

U54	P41	U01	M01	P01	P20	R01	R37	P50	P30	U19	R56
0.28	0.07	0.06	0.04	0.04	0.04	0.01	0.00	0.00	0.00	0.00	0.00

- Transitivity:

M01	P50	P01	U01	U54	P30	U19	P20	P41	R37	R01	R56
0.24	0.23	0.20	0.20	0.19	0.19	0.18	0.18	0.17	0.16	0.16	0.13

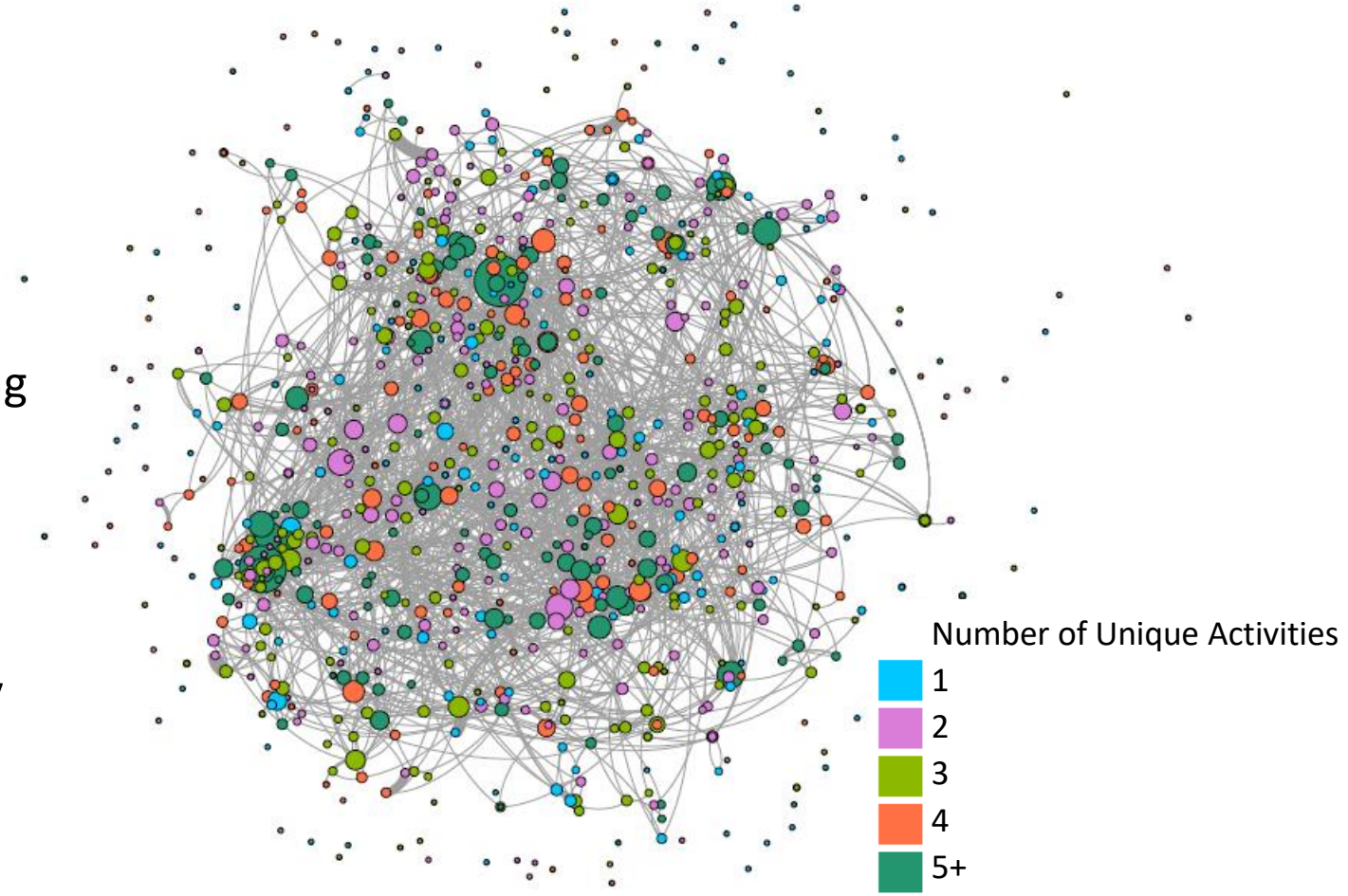
M01	General Clinical Research Centers Program
P01	Research Program Projects
P20	Exploratory Grants
P30	Center Core Grants
P41	Biotechnology Resource Grants
P50	Spec. Center
R01	NIH Research Project Grant Program
R37	MERIT Award
R56	High Priority, Short Term Project Award
U01	Research Project--Cooperative Agreements
U19	Research Program--Cooperative Agreements
U54	Spec. Center - Cooperative Agreements

Study Question 2 Findings

- Study Question 2: Does the type of grant awarded influence collaboration behavior?
  - Overall, grant types are an indicator of homophily.
  - Cooperative agreements tend to foster more collaboration.
    - ‘U54’ and ‘U01’ are the top performing agreements.
    - ‘U54’ (Specialized Center) performs better than ‘U01’ (Research Project) in terms of collaboration.
    - ‘U54’ has more collaboration per researcher, while ‘U01’ shows more collaboration within the same groups of researchers (transitivity).
  - ‘R01’, despite being the most common grant type, has low levels of collaboration.
  - ‘R56’ (high-priority, short-term project award) has extremely low collaboration.

Study Question 3: Does variety in funding sources influence collaboration behavior?

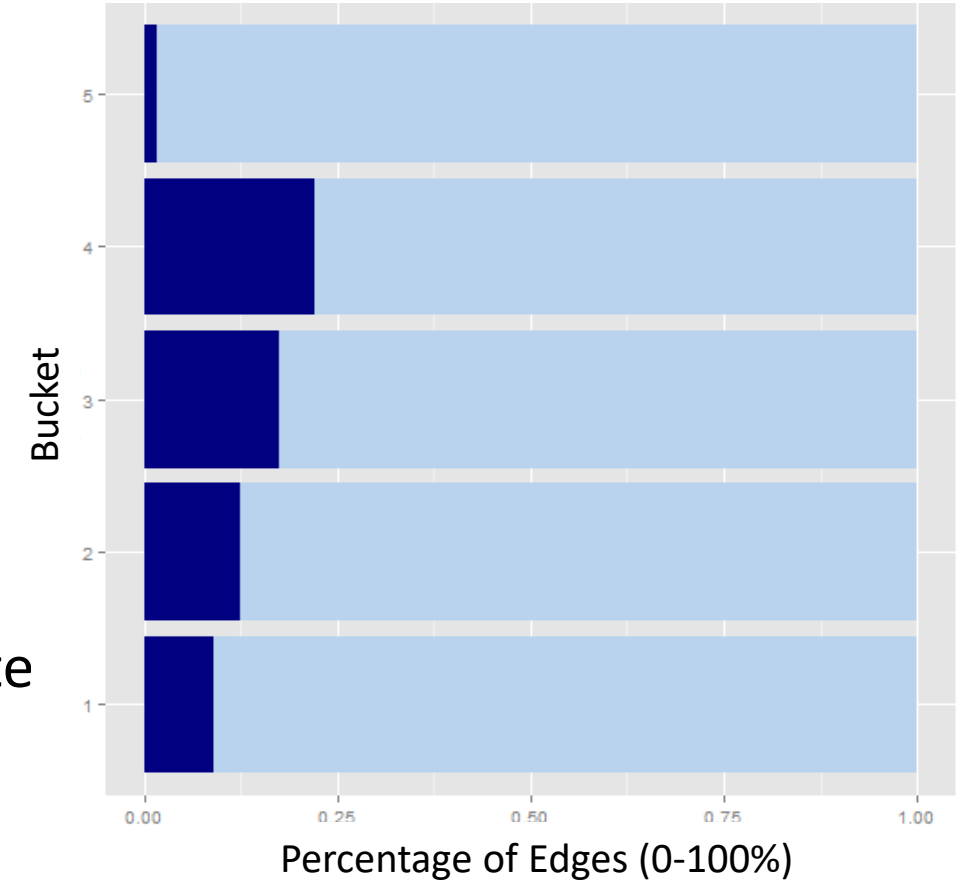
- The graph to the right shows degree centrality based on the number of unique activities (grant types) attached to each researcher within the study-comparison network.
- For network statistics 'variety in funding sources' was defined as the number of unique funding sources per project.
- When grouping by quantiles, the assortativity was 0.03. This is positive, indicating that there is some homophily between nodes with similar variety in funding sources.



Study Question 3: Does variety in funding sources influence collaboration behavior?

- There is a positive correlation between number of projects a researcher has and number of unique of grant types
- To account for this, calculated the number of unique grants per total projects for each researcher and bucketed the result into quantiles
- Researchers with a similar level of variety in funding tend to collaborate more and at an increasing rate

Percentage of Collaboration Within and Outside of Group



Bucket	Different Grant Types per 10 Projects
5	>10
4	7– 10
3	5 – 6+
2	4 – 5
1	1 – 4

Study Question 3: Does variety in funding sources influence collaboration behavior?

- Average Degree:

Bucket:	3	4	2	1	5
Average Degree:	14.63	10.95	9.46	9.34	6.88

- Transitivity:

Bucket:	5	4	3	2	1
Transitivity:	0.22	0.21	0.19	0.17	0.14

- Triad Count/Node Count:

Bucket:	3	2	4	1	5
Triad Count/Node Count:	0.03	0.03	0.02	0.01	0.00

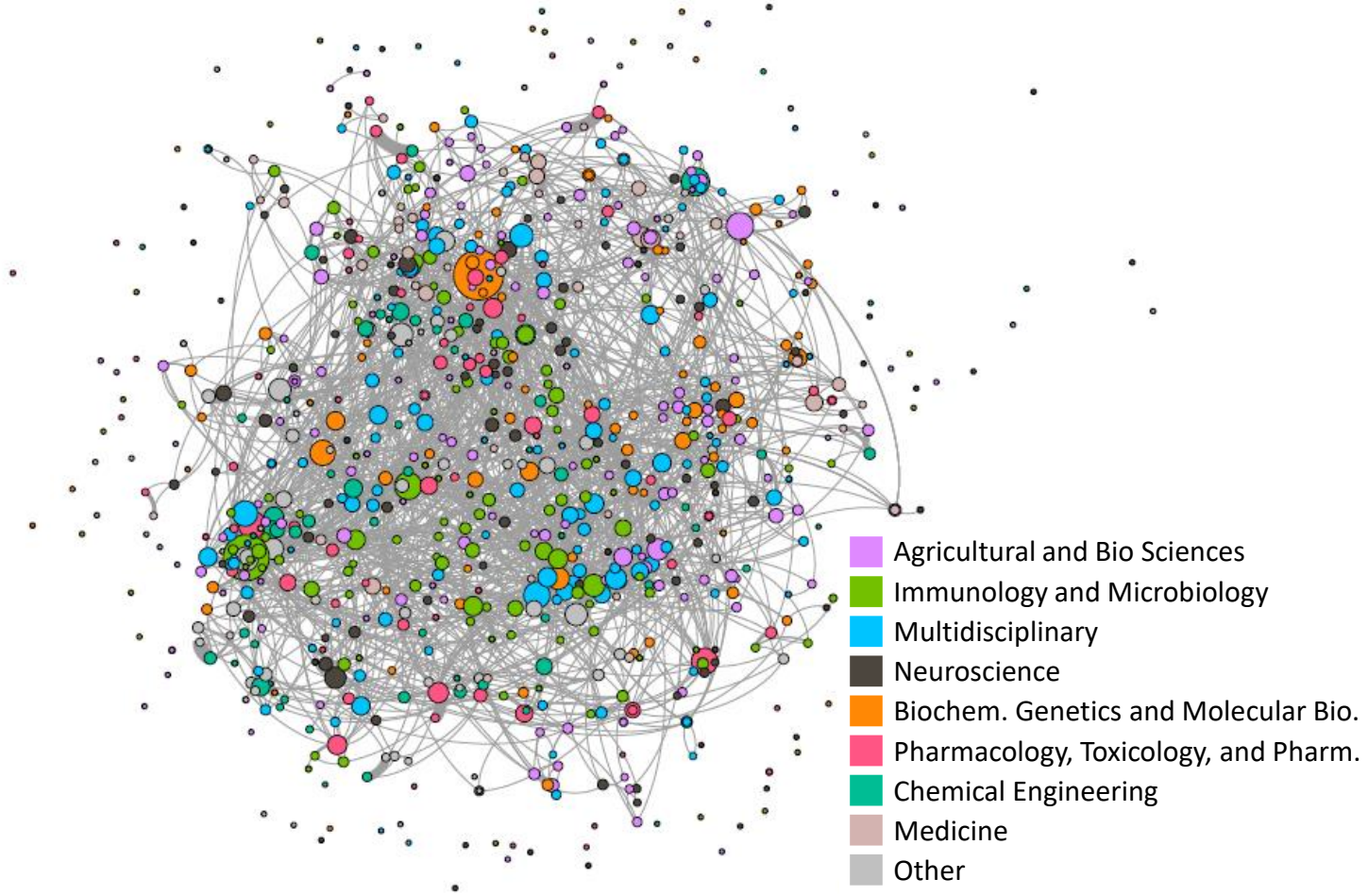
Bucket	Different Grant Types per 10 Projects
1	1 – 4
2	4 – 5
3	5 – 6+
4	7 – 10
5	> 10

Study Question 3 Findings

- Study Question 3: Does variety in funding sources influence collaboration behavior?
  - Overall, variety in funding sources is an indicator of homophily.
  - On average, as the variety in funding sources increases, the likelihood of researchers to collaborate within the same groups increases (based on transitivity and edges within/outside buckets).
  - Researchers in the bucket representing low variety (1-4 grant types) show low collaboration by measure of transitivity, but do not have significantly less connections than the other buckets (based on average degree).

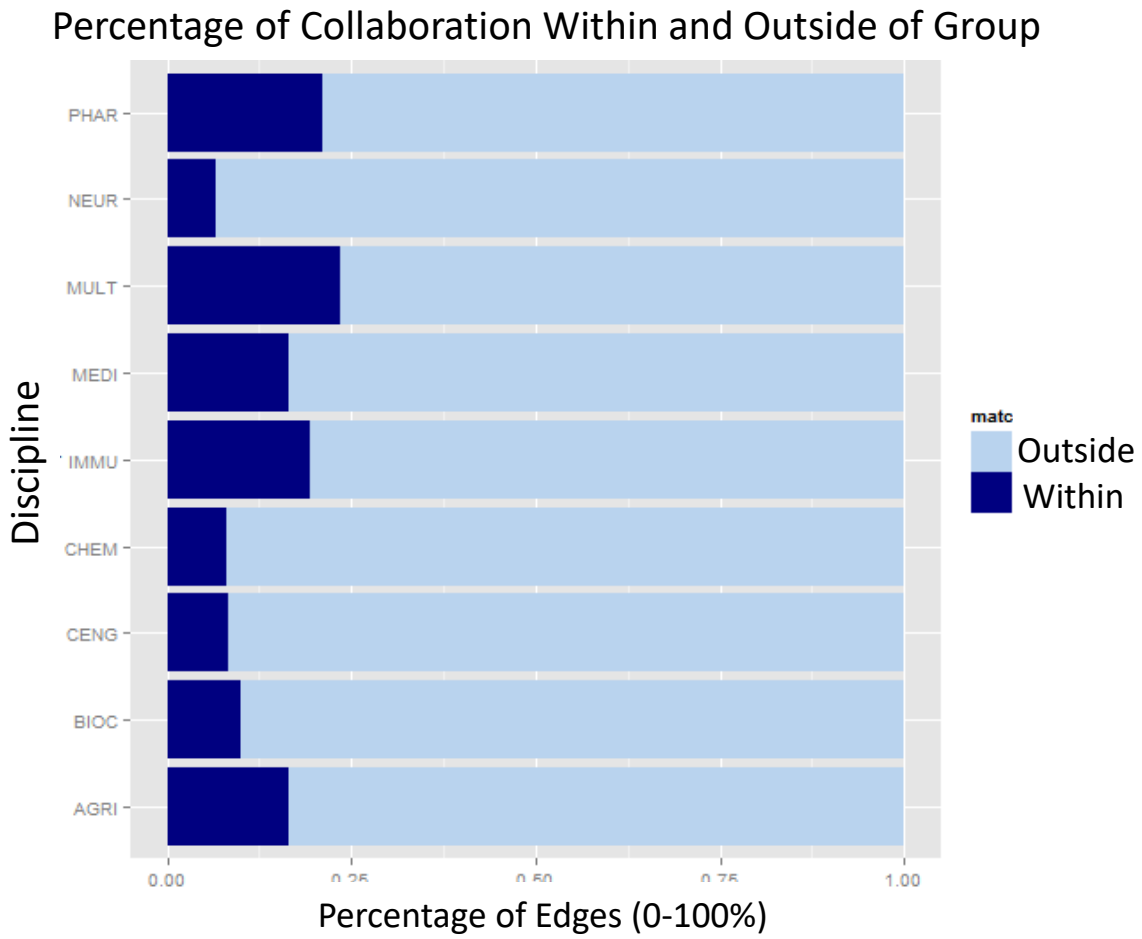
Study Question 4: Does the scientific discipline influence collaboration behavior?

- The graph to the right shows degree centrality based on the primary scientific discipline attached to each researcher within the study-comparison network.
- When grouping nodes by scientific discipline, the assortativity was 0.16. This is positive, indicating that there is homophily by scientific discipline.



Study Question 4: Does the scientific discipline influence collaboration behavior?

- As compared to the other disciplines, researchers classified as the following tend to collaborate more within their respective discipline than with other disciplines:
  - Pharmacology, Toxicology, and Pharmaceutics
  - Medicine
  - Immunology and Microbiology
  - Agricultural and Bio Sciences
- Researchers classified as the following tend to collaborate more outside of their own discipline
  - Neuroscience
  - Chemistry
  - Chemical Engineering and Biochemistry
  - Genetics and Molecular Biology



Study Question 4: Does the scientific discipline influence collaboration behavior?

- Average Degree:

CHEM	CENG	PHAR	MULT	MEDI	IMMU	BIOC	AGRI	NEUR
15.56	14.83	13.90	13.80	12.95	10.34	9.01	7.39	6.20

- Weighted Degree:

CHEM	CENG	PHAR	MULT	MEDI	IMMU	NEUR	BIOC	AGRI
3.89	3.84	3.75	3.43	2.79	2.57	1.71	1.68	1.54

- Transitivity:

MEDI	PHAR	IMMU	CHEM	CENG	MULT	AGRI	BIOC	NEUR
0.34	0.19	0.19	0.18	0.18	0.17	0.16	0.14	0.13

- Triad Count/Node Count:

MULT	IMMU	PHAR	CENG	BIOC	MEDI	AGRI	CHEM	NEUR
0.15	0.12	0.11	0.08	0.07	0.04	0.02	0.02	0.00

AGRI	Agricultural and Bio Sciences
BIOC	Biochemistry, Genetics, and Molecular Biology
CENG	Chemical Engineering and Biochemistry
CHEM	Chemistry
IMMU	Immunology and Microbiology
MEDI	Medicine
MULT	Multidisciplinary
NEUR	Neuroscience
PHAR	Pharmacology, Toxicology, and Pharmaceutics

Study Question 4 Findings

- Study Question 4: Does the scientific discipline influence collaboration behavior?
  - Overall, scientific discipline is an indicator of homophily.
  - Disciplines with more interdisciplinary collaboration (chemistry and chemical engineering) have higher average degrees than other disciplines.
  - ‘Medicine’ is a tightly-knit network (high transitivity), and has a relatively high amount of internal collaboration.

Conclusions

- NIGMS-supported researchers tend to collaborate more with other NIGMS-supported researchers.
- Cooperative grant types effectively foster collaboration among researchers and form tightly-knit communities.
- Discipline is a strong indicator of collaboration level.
- The metrics presented can be used to target qualities that represent opportunities for developing collaboration.

Acknowledgements

- NIH: Anna Calcagno, Andrew Miklos, and Claire Schulkey
- GWSB: Professor Shivraj Kanungo

Appendix

Process for Identifying the Study and Comparison Groups

1. Filter out non-NIH entities.
2. Use activity codes to filter down to funding mechanisms of interest to our scope.
3. Filter down to only 2001-2015 projects.
4. Split up PI lists to identify individual PIs and split up grant money shared among PIs.
5. Identify the PIs with >$50,000 of support for every year in the study period (2001-2015).

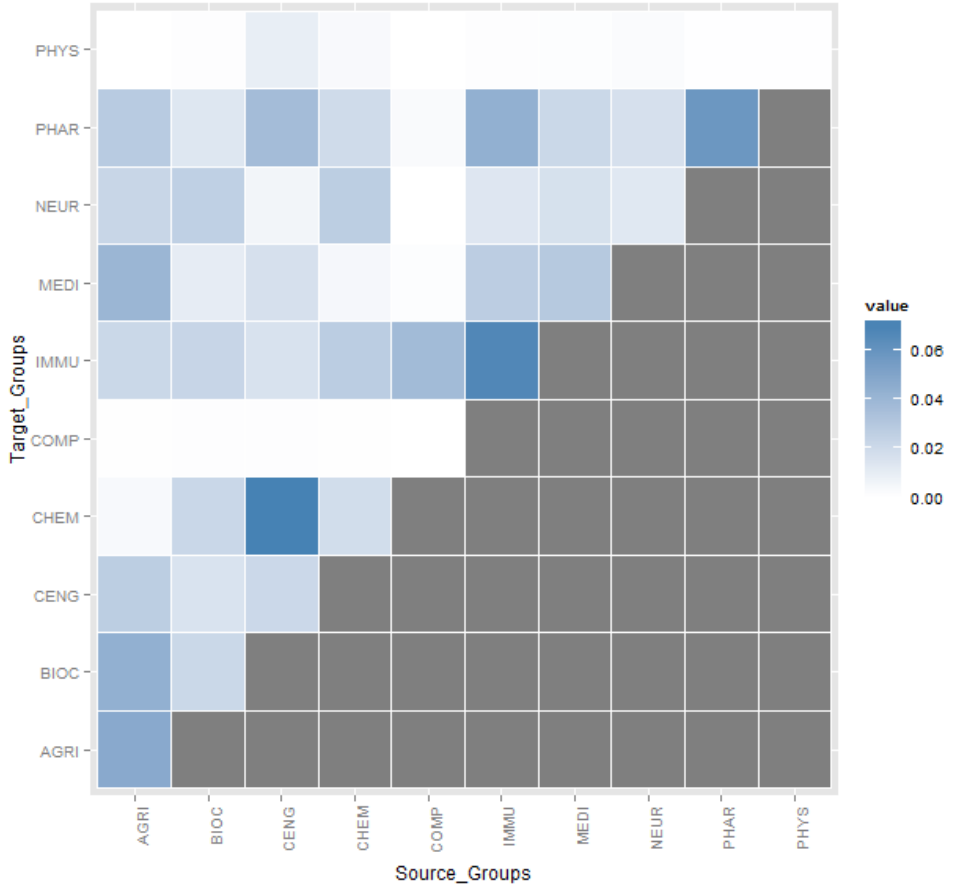
Graph Statistics Definition

1. Node
   a. A visual representation of a researchers.
2. Edge
   a. A visual representation of a publication between two researchers.
3. Homophily
   a. “Birds of a feather stick together”
   b. Individuals tend to congregate with other like-minded individuals to form tightly knit clusters.
4. Assortativity
   a. Provides a measure of homophily, ranging from -1 to 1.
   b. A positive assortativity indicates that each group tends to collaborate within themselves.

Graph Statistics Definition Cont'd

5. Degree Centrality
   a. Measures the number of direct connections (degrees) one node has to other nodes.
   b. The number of connections is a key measure of importance or influence within the network.
6. Transitivity
   a. A measure of social process such that friends of friends will become friends.
   b. Dyad consists of two connected nodes and triad consists of three connected nodes.
   c. Transitivity = $\frac{3 * number\ of\ triangles\ (number\ of\ closed\ triplets)}{number\ of\ connected\ triplets\ of\ vertices}$
   d. A high transitivity measure indicates a tightly knit collaborative environment (clique), while a low transitivity measure indicates that neighboring researchers do not collaborate with each other (star).

Discipline Collaboration Heat Map



Discipline Collaboration

Discipline Combinations		Collaboration Count
• Chemistry	• Chemical Engineering	133
• Computer Science	• Immunology and Microbiology	128
• Chemical Engineering	• Chemistry	109
• Chemical Engineering	• Pharmacology, Toxicology and Pharmaceutics	107
• Immunology and Microbiology	• Pharmacology, Toxicology and Pharmaceutics	94
• Agricultural and Biological Sciences	• Medicine	89
• Pharmacology, Toxicology and Pharmaceutics	• Agricultural and Biological Sciences	88
• Agricultural and Biological Sciences	• Biochemistry, Genetics and Molecular Biology	87
• Biochemistry, Genetics and Molecular Biology	• Neuroscience	63
• Biochemistry, Genetics and Molecular Biology	• Agricultural and Biological Sciences	62

Grant Type Collaboration

Grant Type		Collaboration Count
• P41	• R01	219
• R01	• P41	185
• UM1	• U54	172
• U54	• P01	143
• P01	• P41	127
• P41	• P01	122
• P01	• U54	95
• U54	• R01	93
• R01	• P01	88
• U54	• P41	87

Unique Activity Distribution

Grant Type Combinations	Researcher Count
R01_R37	80
P01_R01	41
R01_R21	37
P41_R01	28
R01_R56	14

Grant Type Combinations	Researcher Count
P41_R01_R37	21
P01_R01_R37	15
P01_R01_R21	14
P30_R01_R37	10
P01_P41_R01	10

Grant Type Combinations	Researcher Count
P01_P41_R01_R37	10
P01_P30_R01_R21	5
P01_R01_R37_U54	4
P01_R01_R21_R37	4
P01_P41_R01_U01	4

Methodology for Node Discipline Assignment

1. Sparse Matrix with Indicators of Subject

	BIOC	CHEM	AGRI	OTHER
PMID_1	0	1	0	0
PMID_2	1	0	0	1
PMID_3	1	0	1	0
PMID_4	0	0	0	1
PMID_5	1	0	0	0
PMID_6	0	1	1	1
PMID_7	1	0	1	0
PMID_8	1	0	1	0
Total	5	2	4	3
Distribution	36%	14%	29%	21%
Factor (1/Dist)	2.8	7	3.5	4.666667

2. Expand with PI Information

		BIOC	CHEM	AGRI	OTHER
PIID 1	PMID_1	0	1	0	0
PIID 3	PMID_1	0	1	0	0
PIID 2	PMID_2	1	1	0	1
PIID 8	PMID_2	1	1	0	1
PIID 9	PMID_2	1	1	0	1
PIID 3	PMID_3	1	0	0	0
PIID 4	PMID_3	1	0	0	0
PIID 1	PMID_3	1	0	0	0
PIID 7	PMID_3	1	0	0	0
PIID 4	PMID_3	1	0	0	0
PIID 4	PMID_4	0	0	0	1
PIID 5	PMID_5	1	0	0	0
PIID 6	PMID_5	1	0	0	0
PIID 6	PMID_6	0	1	1	1
PIID 1	PMID_6	0	1	1	1
PIID 3	PMID_6	0	1	1	1
PIID 7	PMID_7	1	0	1	0
PIID 2	PMID_7	1	0	1	0
PIID 8	PMID_8	1	0	1	0
PIID 9	PMID_8	1	0	1	0
PIID 5	PMID_8	1	0	1	0

3. Apply Weights

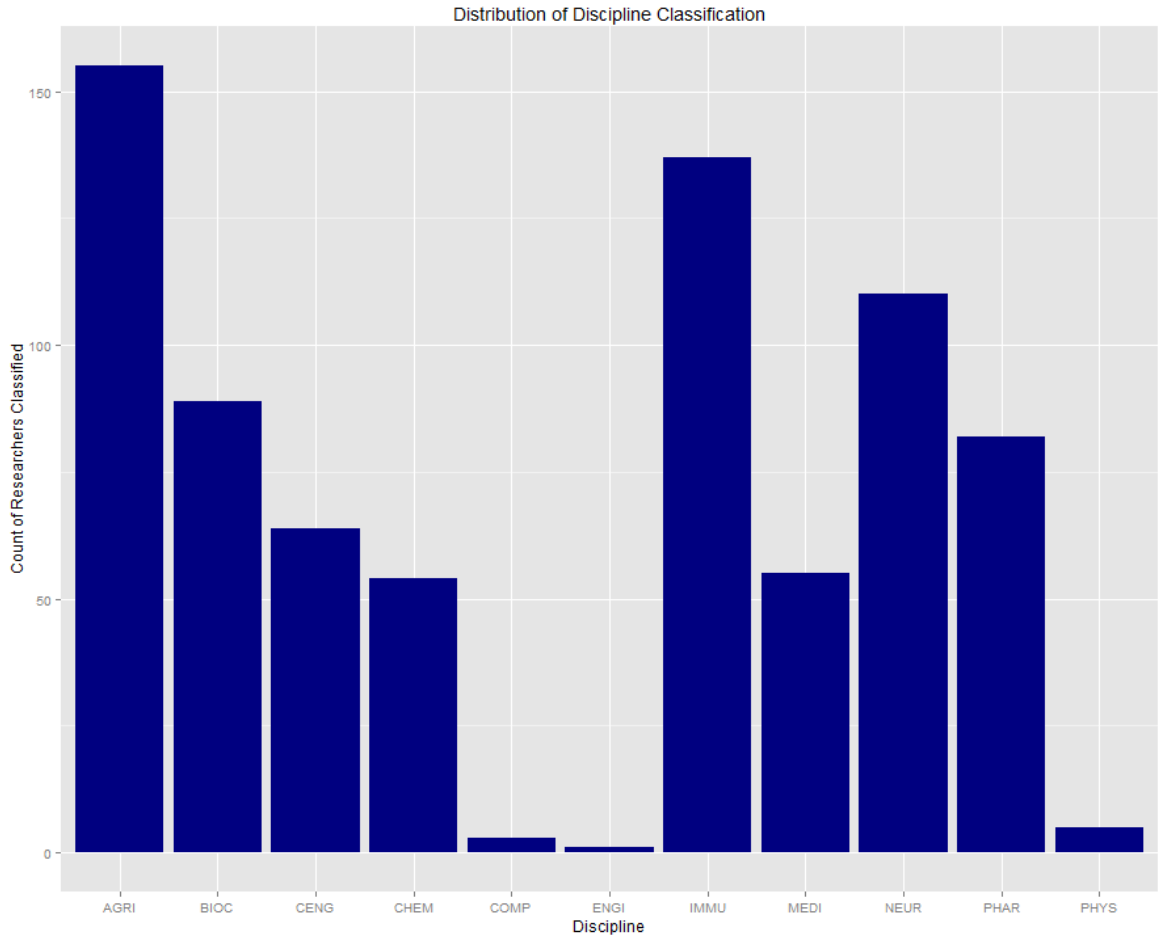
		BIOC	CHEM	AGRI	OTHER
PIID 1	PMID_1	0	7	0	0
PIID 3	PMID_1	0	7	0	0
PIID 2	PMID_2	2.8	7	0	4.666667
PIID 8	PMID_2	2.8	7	0	4.666667
PIID 9	PMID_2	2.8	7	0	4.666667
PIID 3	PMID_3	2.8	0	0	0
PIID 4	PMID_3	2.8	0	0	0
PIID 1	PMID_3	2.8	0	0	0
PIID 7	PMID_3	2.8	0	0	0
PIID 4	PMID_3	2.8	0	0	0
PIID 4	PMID_4	0	0	0	4.666667
PIID 5	PMID_5	2.8	0	0	0
PIID 6	PMID_5	2.8	0	0	0
PIID 6	PMID_6	0	7	3.5	4.666667
PIID 1	PMID_6	0	7	3.5	4.666667
PIID 3	PMID_6	0	7	3.5	4.666667
PIID 7	PMID_7	2.8	0	3.5	0
PIID 2	PMID_7	2.8	0	3.5	0
PIID 8	PMID_8	2.8	0	3.5	0
PIID 9	PMID_8	2.8	0	3.5	0
PIID 5	PMID_8	2.8	0	3.5	0

Methodology for Node Discipline Assignment – continued

4. Resort and collapse					
		BIOC	CHEM	AGRI	OTHER
PIID 1	PMID_1	0	7	0	0
PIID 1	PMID_3	2.8	0	0	0
PIID 1	PMID_6	0	7	3.5	4.666667
PIID 2	PMID_2	2.8	7	0	4.666667
PIID 2	PMID_7	2.8	0	3.5	0
PIID 3	PMID_1	0	7	0	0
PIID 3	PMID_3	2.8	0	0	0
PIID 3	PMID_6	0	7	3.5	4.666667
PIID 4	PMID_3	2.8	0	0	0
PIID 4	PMID_3	2.8	0	0	0
PIID 4	PMID_4	0	0	0	4.666667
PIID 5	PMID_5	2.8	0	0	0
PIID 5	PMID_8	2.8	0	3.5	0
PIID 6	PMID_5	2.8	0	0	0
PIID 6	PMID_6	0	7	3.5	4.666667
PIID 7	PMID_3	2.8	0	0	0
PIID 7	PMID_7	2.8	0	3.5	0
PIID 8	PMID_2	2.8	7	0	4.666667
PIID 8	PMID_8	2.8	0	3.5	0
PIID 9	PMID_2	2.8	7	0	4.666667
PIID 9	PMID_8	2.8	0	3.5	0

5. Choose final classification as max of normalized weights					
	BIOC	CHEM	AGRI	OTHER	Discipline
PIID 1	2.8	14	3.5	4.666667	CHEM
PIID 2	5.6	7	3.5	4.666667	CHEM
PIID 3	2.8	14	3.5	4.666667	CHEM
PIID 4	5.6	0	0	4.666667	BIOC
PIID 5	5.6	0	3.5	0	BIOC
PIID 6	2.8	7	3.5	4.666667	CHEM
PIID 7	5.6	0	3.5	0	BIOC
PIID 8	5.6	7	3.5	4.666667	CHEM
PIID 9	5.6	7	3.5	4.666667	CHEM

Discipline Classification Node Distribution



Activity Code Classification Node Distribution