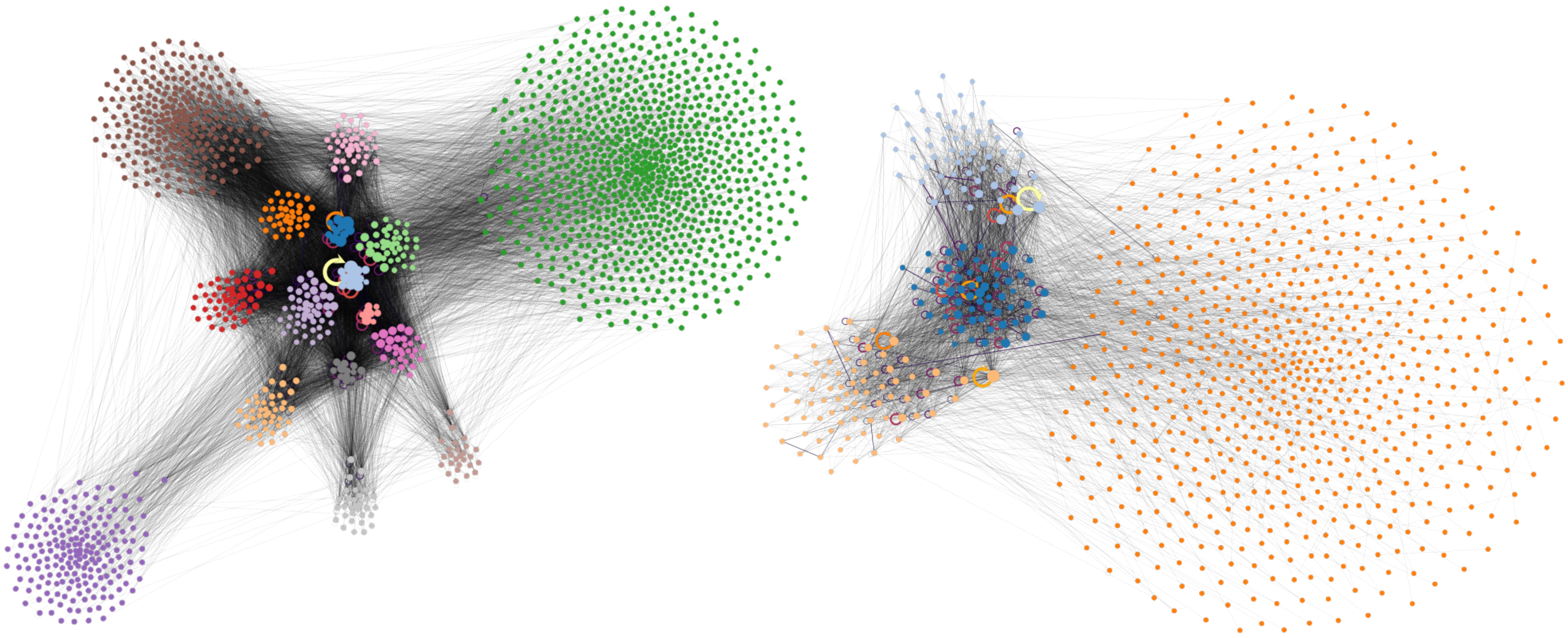
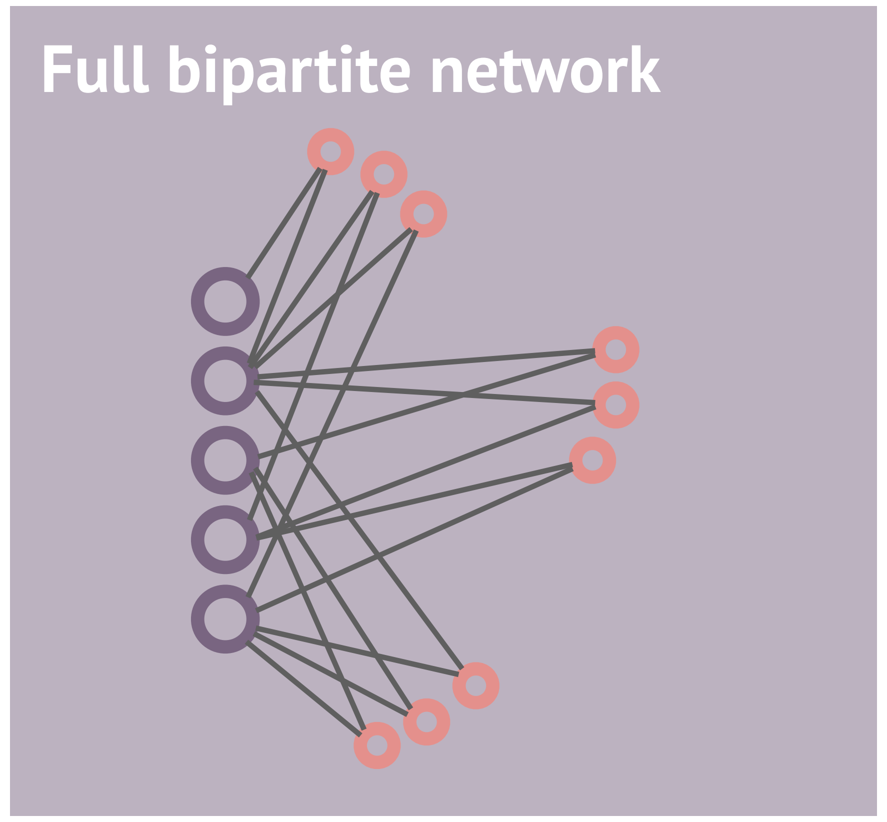


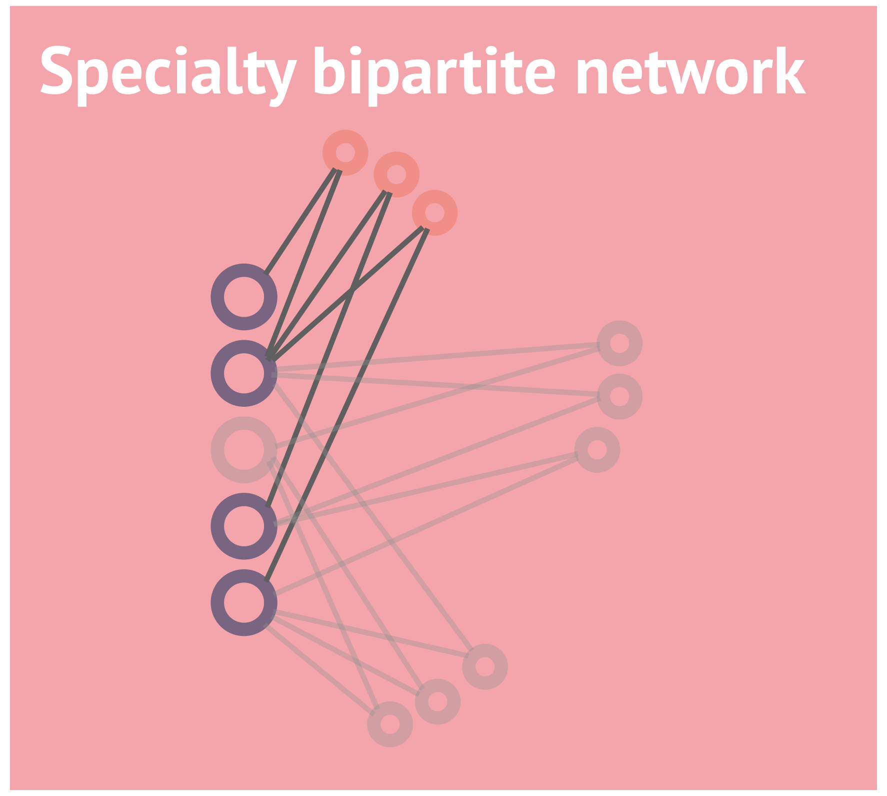
Training trajectories of physicians in the US

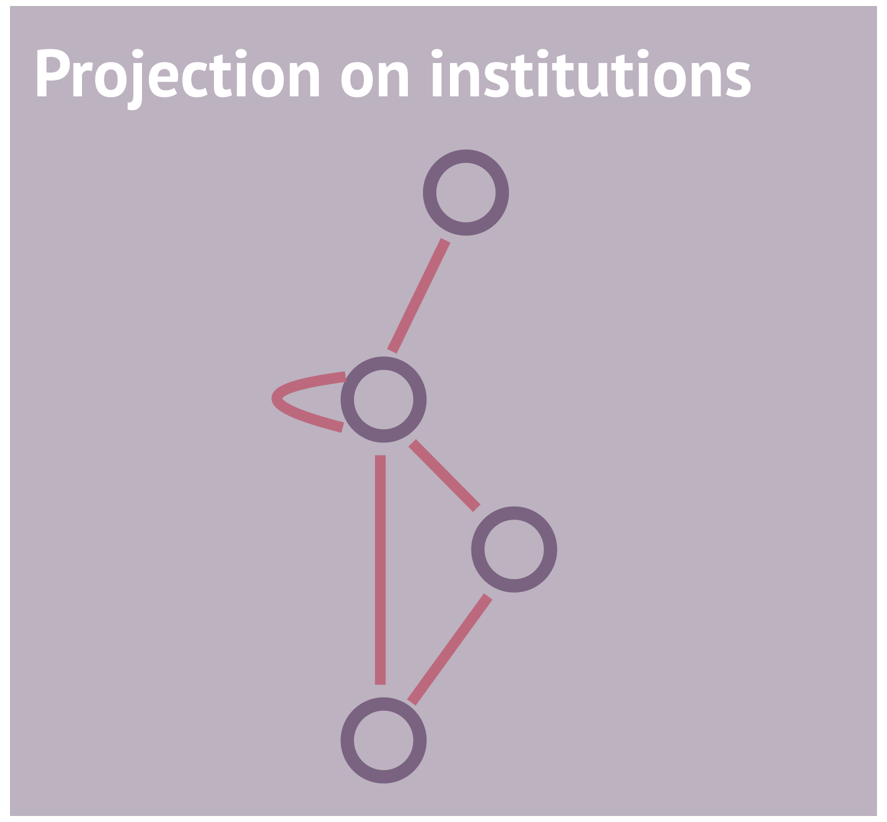
Nima Moghaddas | Network Science PhD student

My network

- Constructed from scraped data from 15 different specialties:
  - Allergy & Immunology
  - Cardiology
  - Child neurology
  - Colorectal surgery
  - Dermatology
  - Gastroenterology
  - Geriatrics
  - Hematology
  - Interventional radiology
  - Medical genetics
  - Neurosurgery
  - Obstetrics and gynecology
  - Ophthalmology
  - Pediatrics
  - Psychiatry

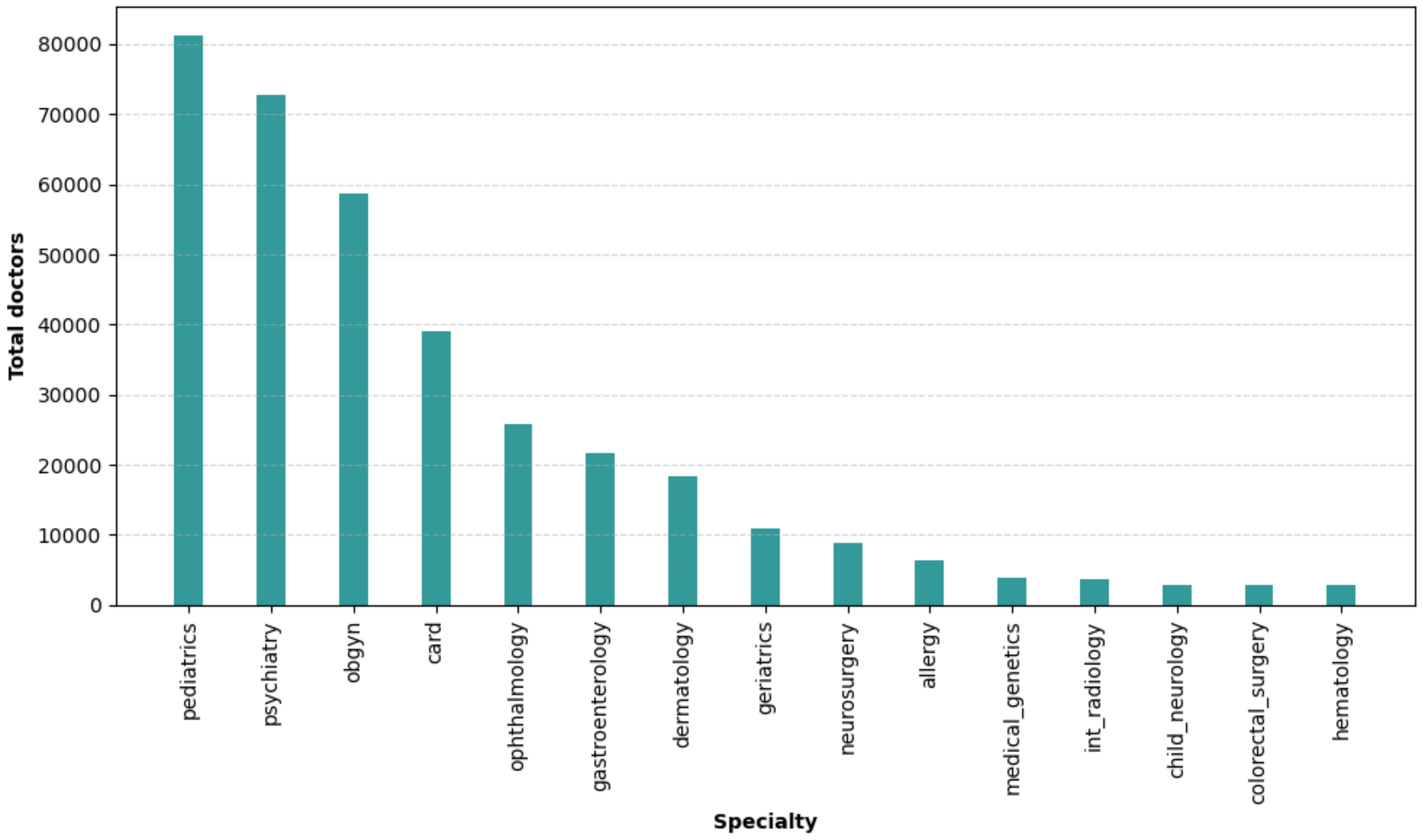






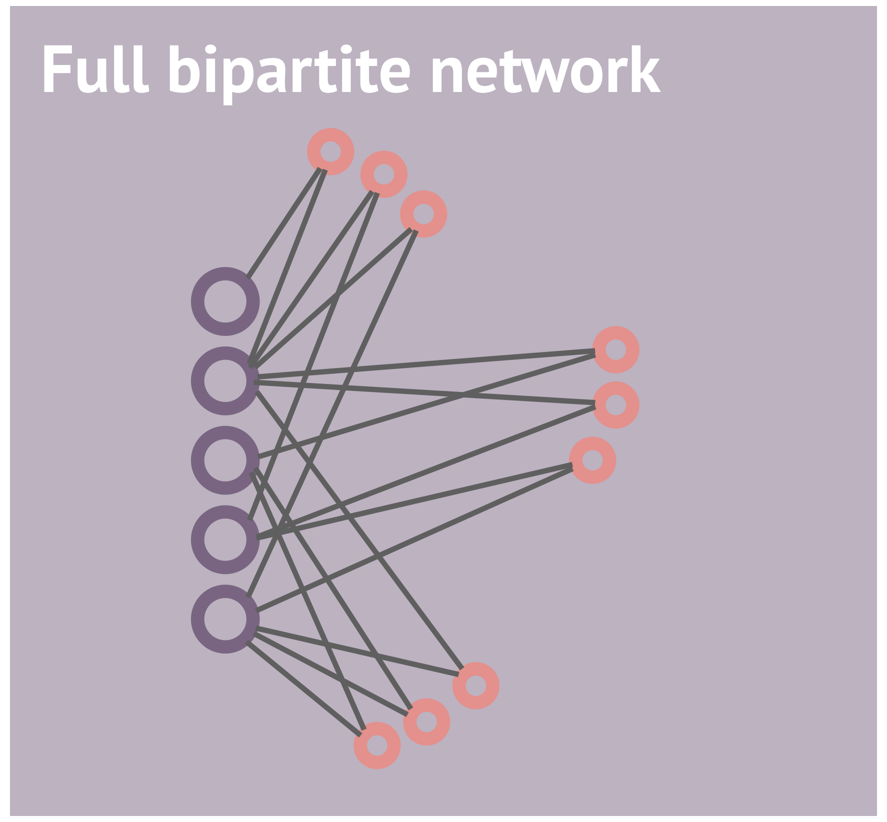
My network

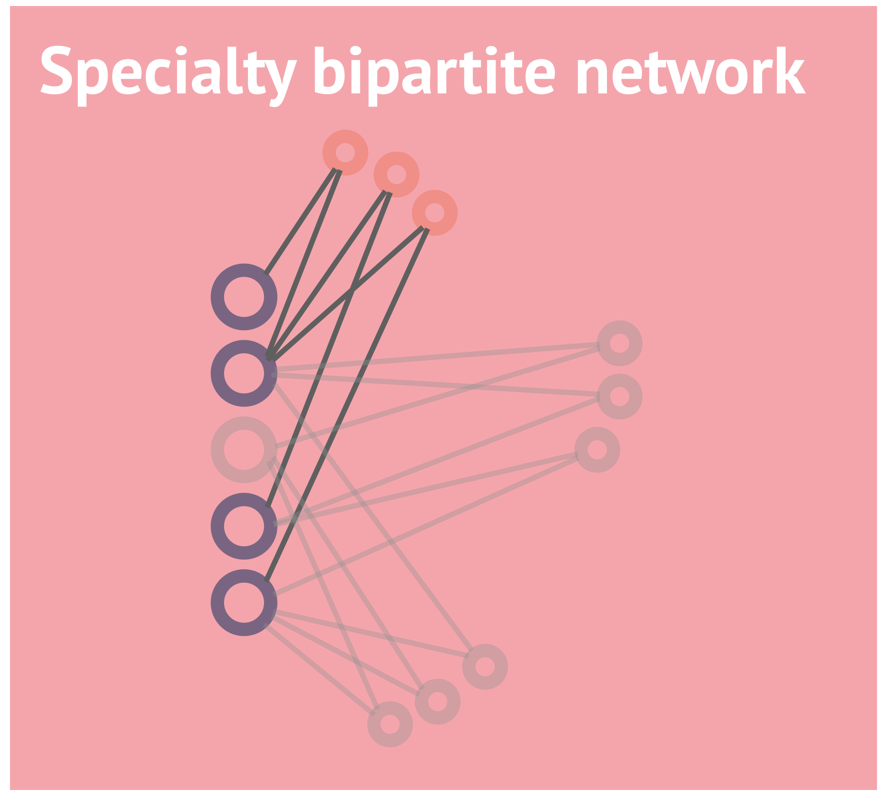
- Constructed from scraped data from 15 different specialties:
  - Allergy & Immunology
  - Cardiology
  - Child neurology
  - Colorectal surgery
  - Dermatology
  - Gastroenterology
  - Geriatrics
  - Hematology
  - Interventional radiology
  - Medical genetics
  - Neurosurgery
  - Obstetrics and gynecology
  - Ophthalmology
  - Pediatrics
  - Psychiatry

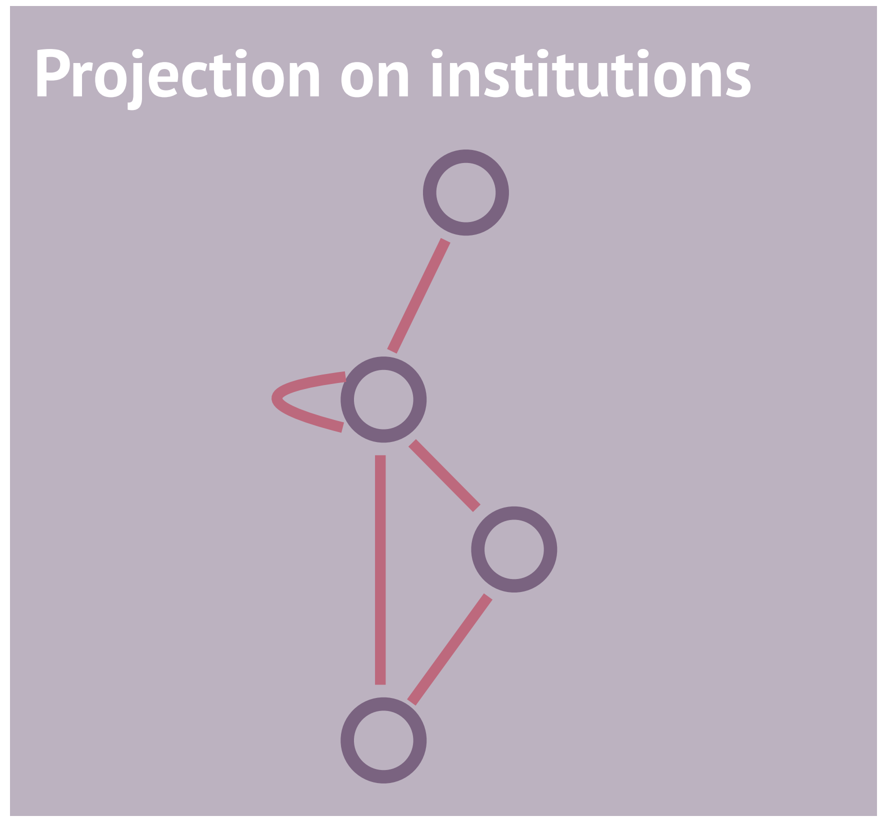


360,155 Doctors

5,051 Unique Institutions

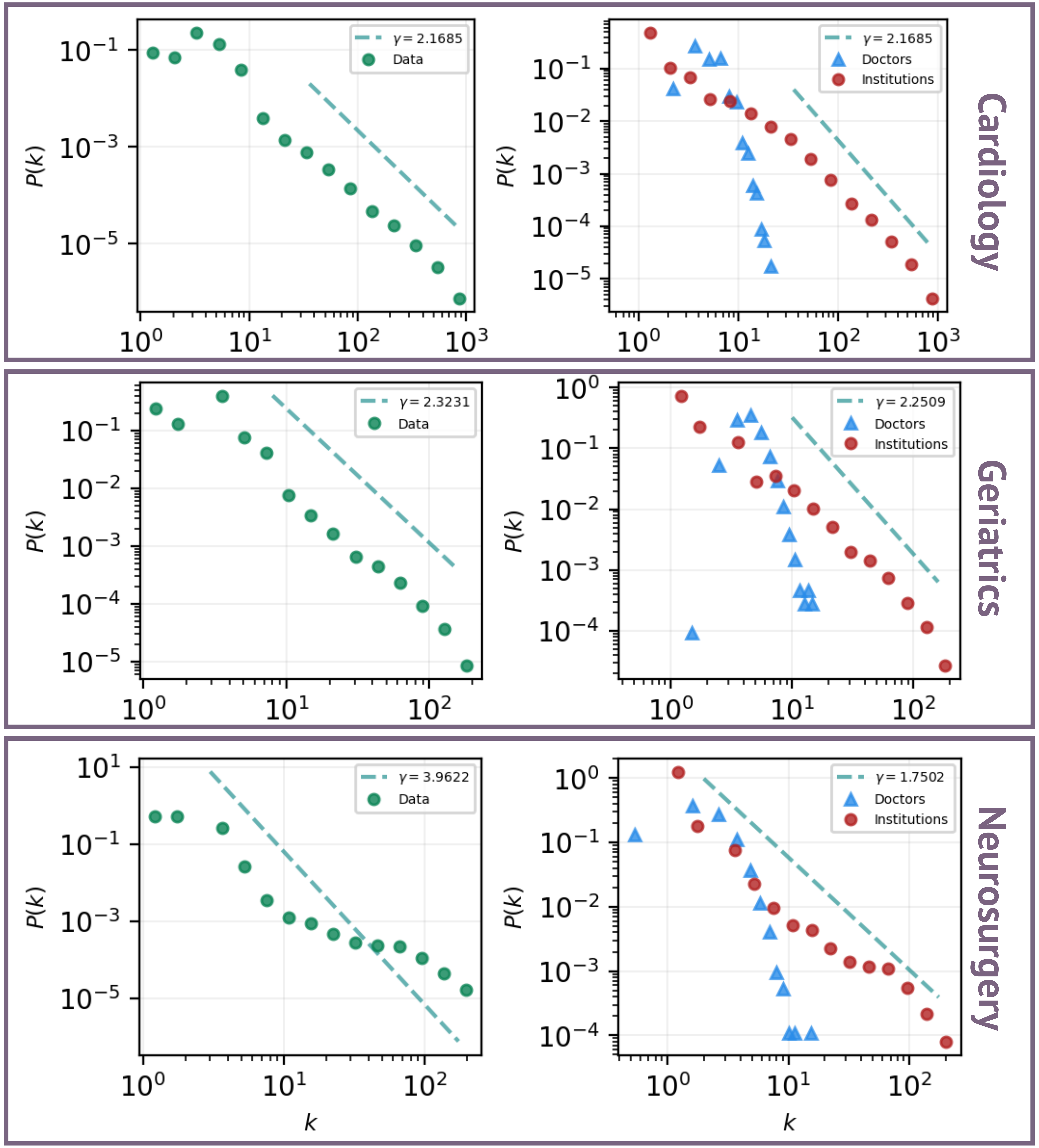






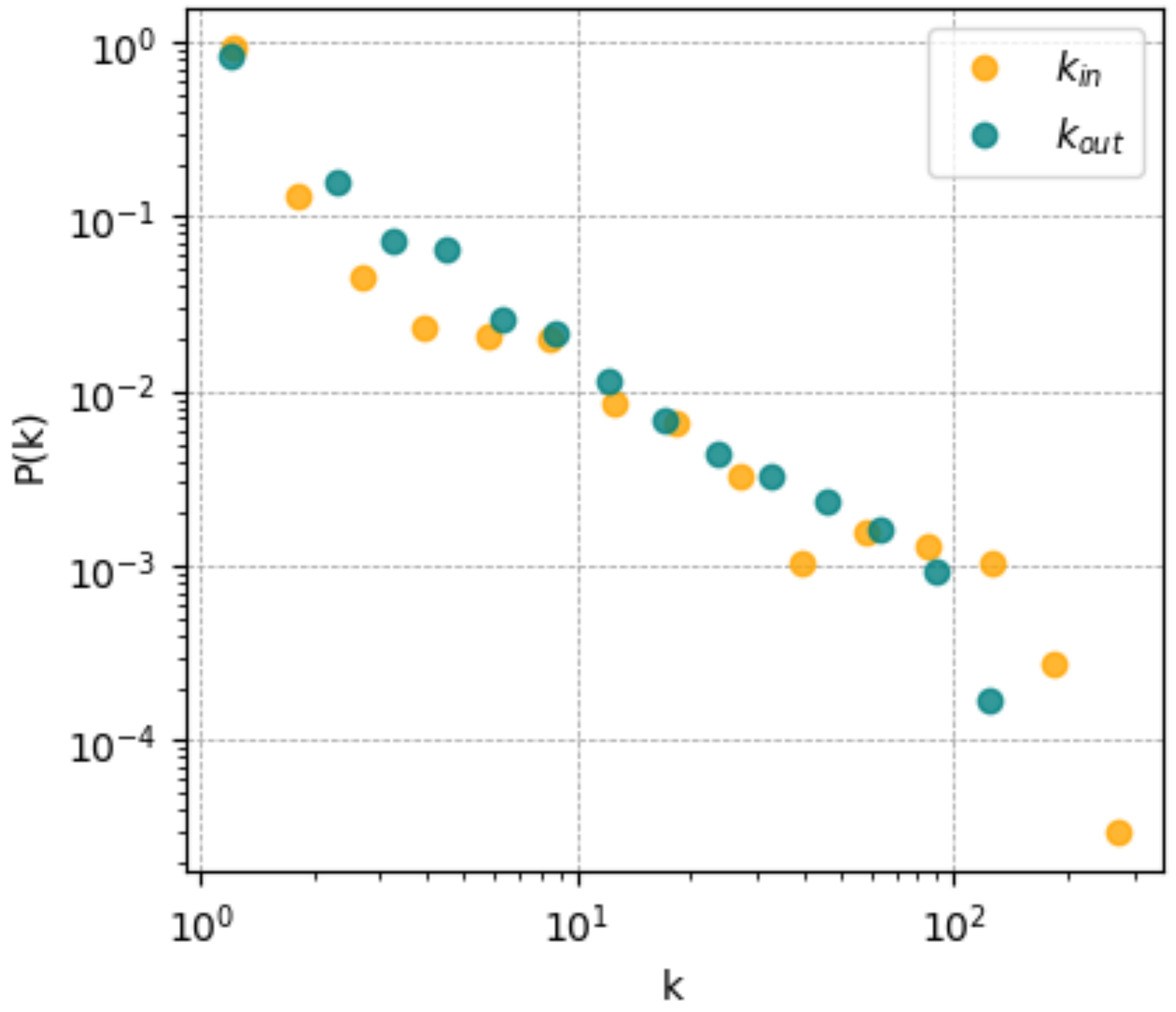
Degree distribution (bipartite)

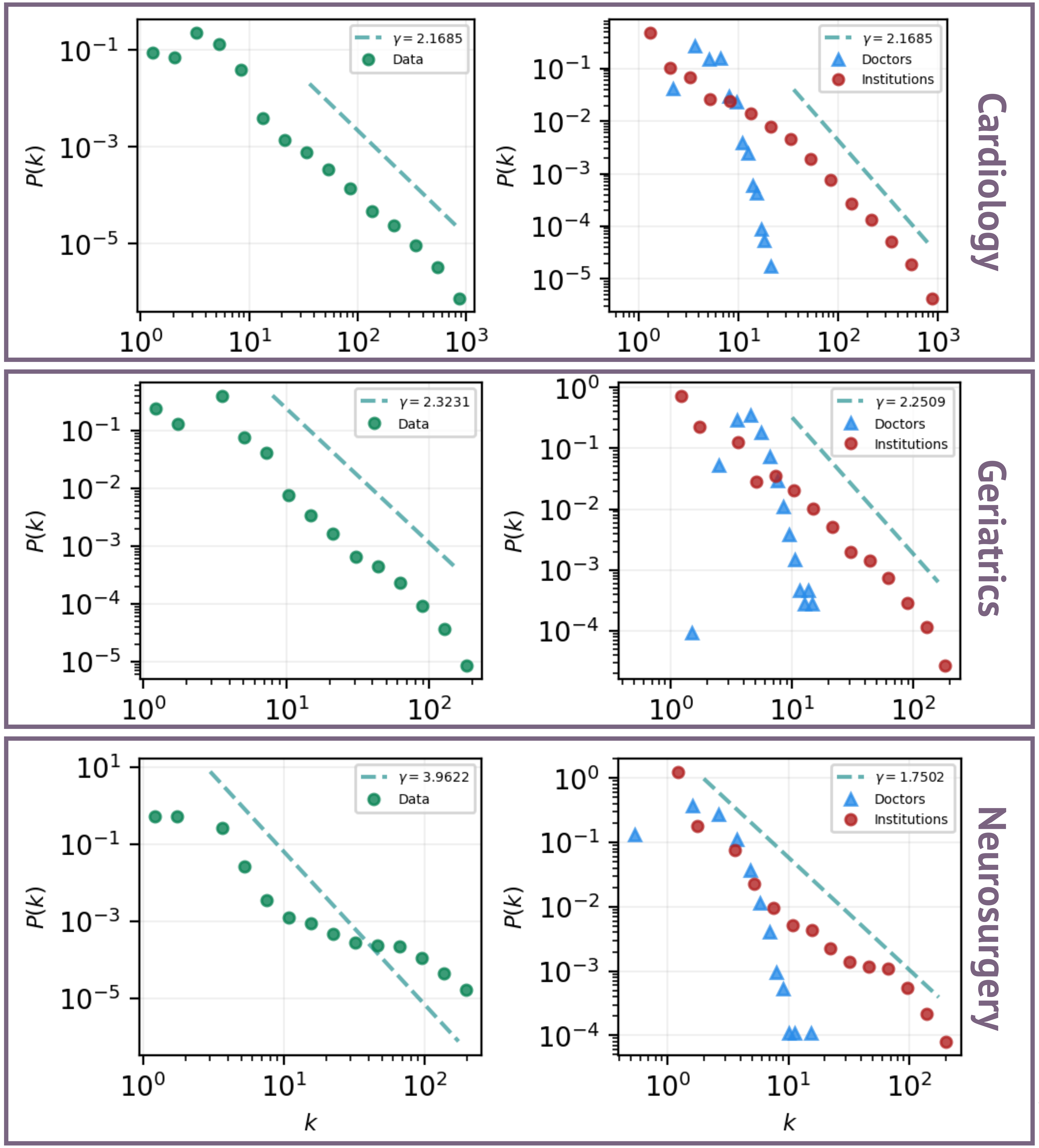
- Degree distributions of bipartite network by specialty
- Degrees of doctors and institutions follow different distributions
  - Institutions follow power law
  - Doctors follow exponentially bounded distribution



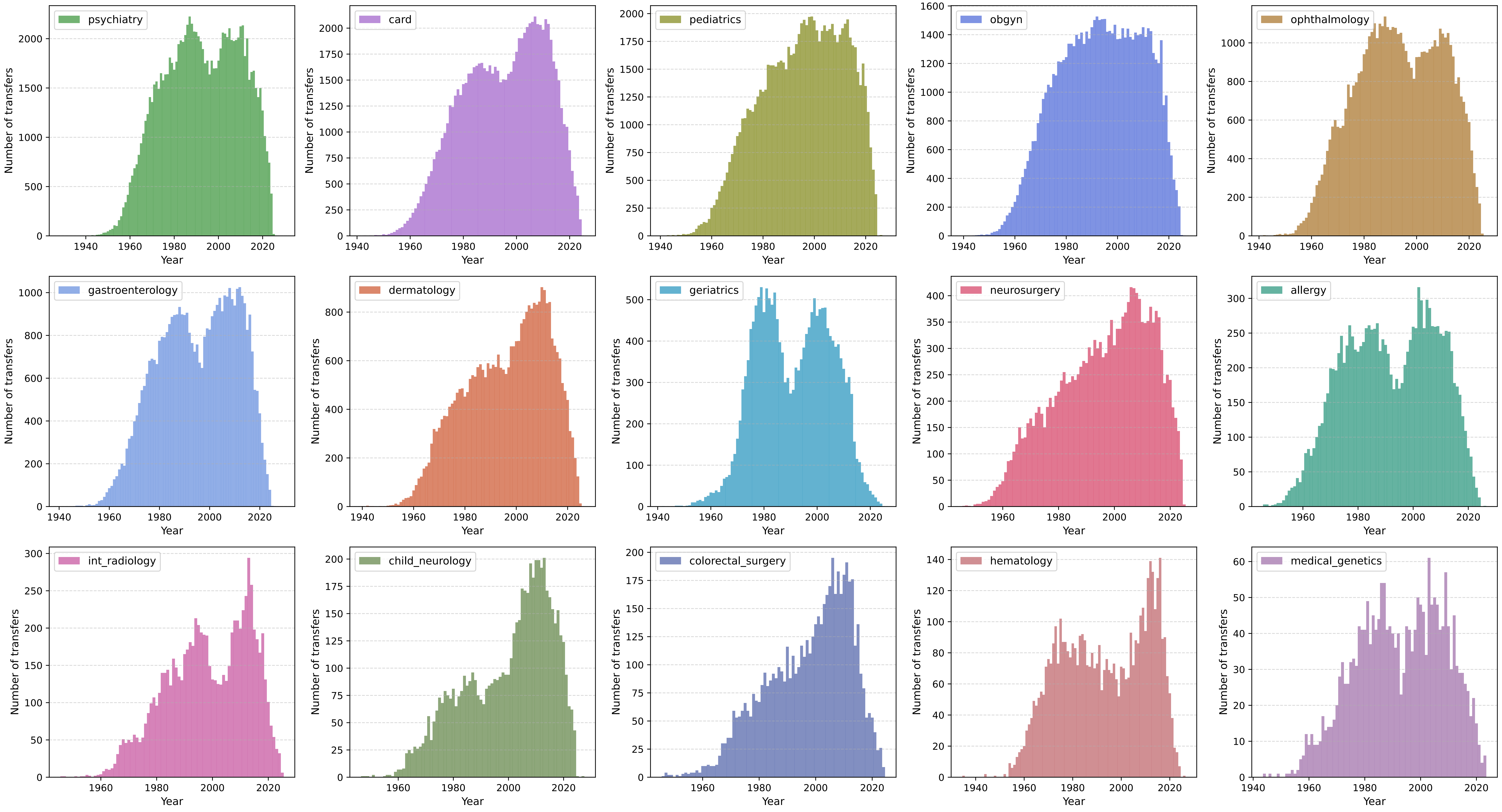
Degree distribution (bipartite)

- Degree distributions of bipartite network by specialty
- Degrees of doctors and institutions follow different distributions
  - Institutions follow power law
  - Doctors follow exponentially bounded distribution

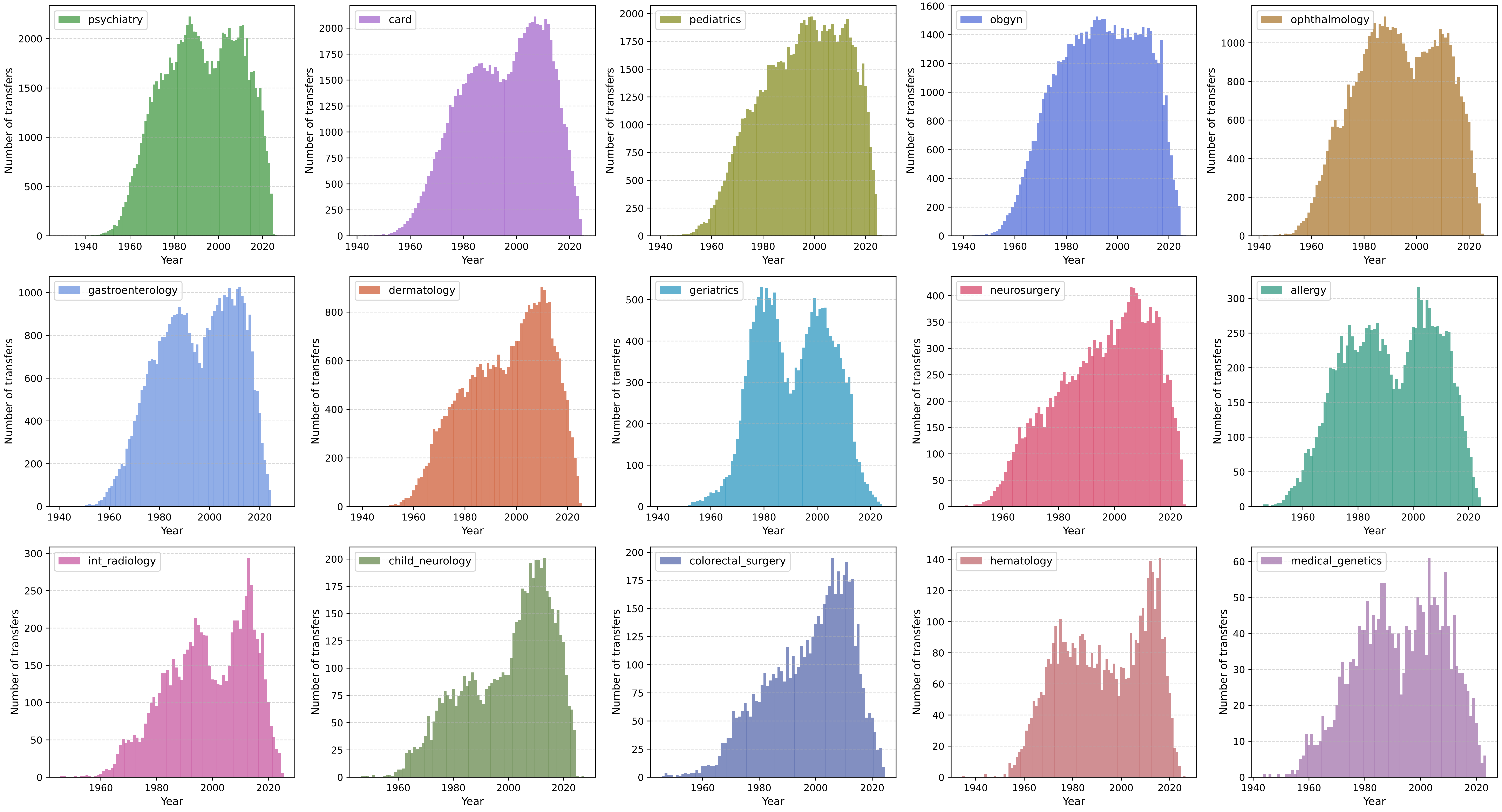


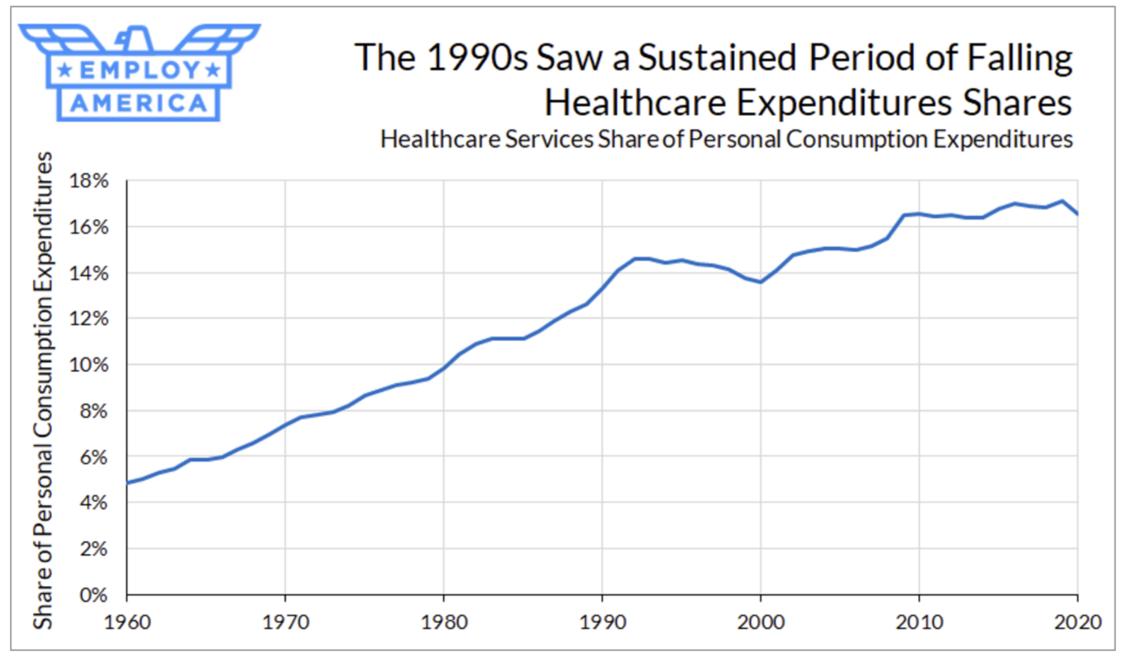


Temporal transfer network



Temporal transfer network





Source: Bureau of Economic Analysis, Author's Calculations

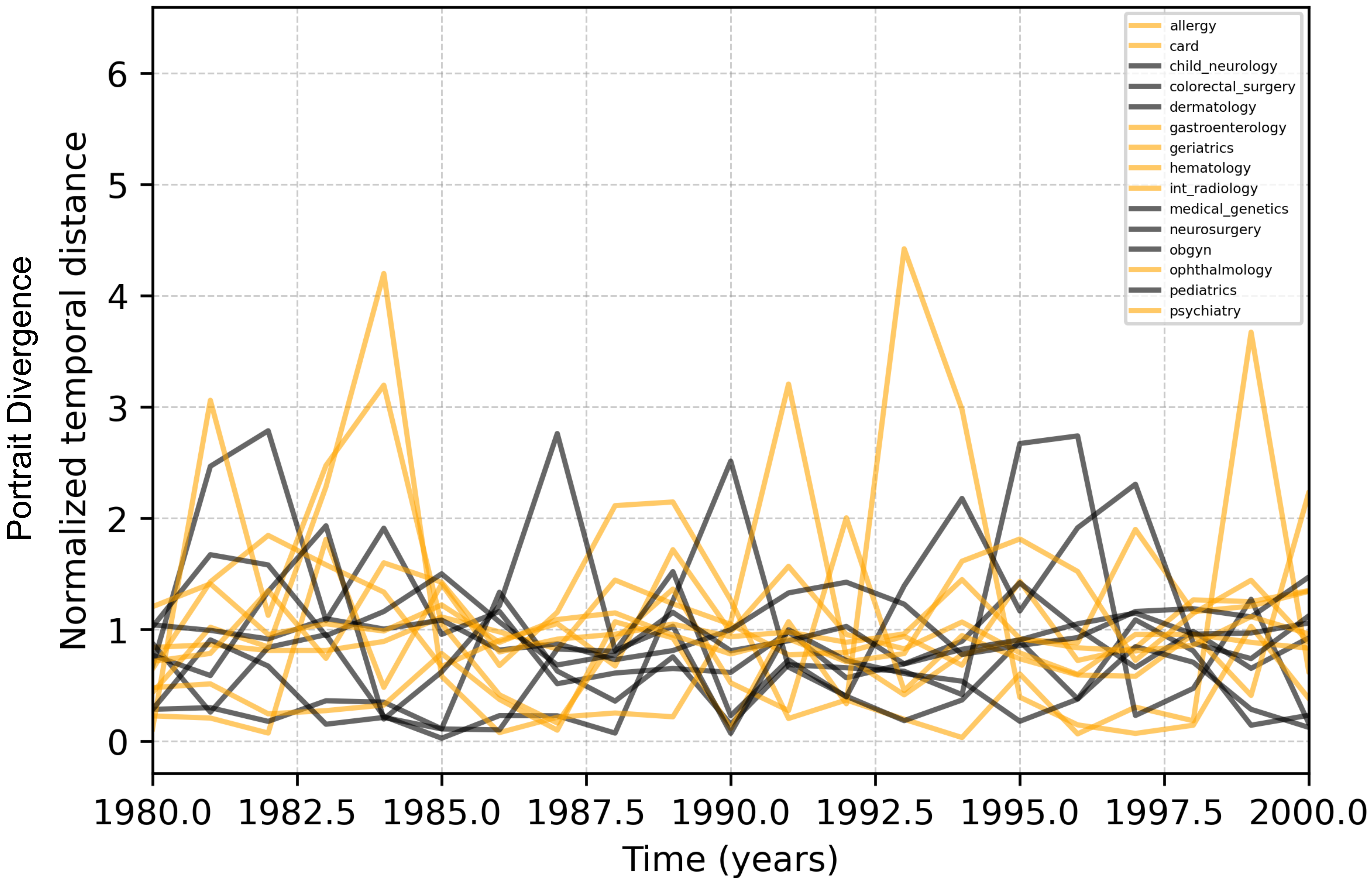
"In particular, there was one extended period where healthcare spending slowed substantially: the 1990s. Between 1993 and 2000, healthcare's share of PCE fell from 14.6% to 13.6%. In the post-war era, this is the only period of a sustained decline in healthcare's share of PCE."

- For the remainder of the analysis I focus on transfers that occurred between 1980-2010

5015 Nodes

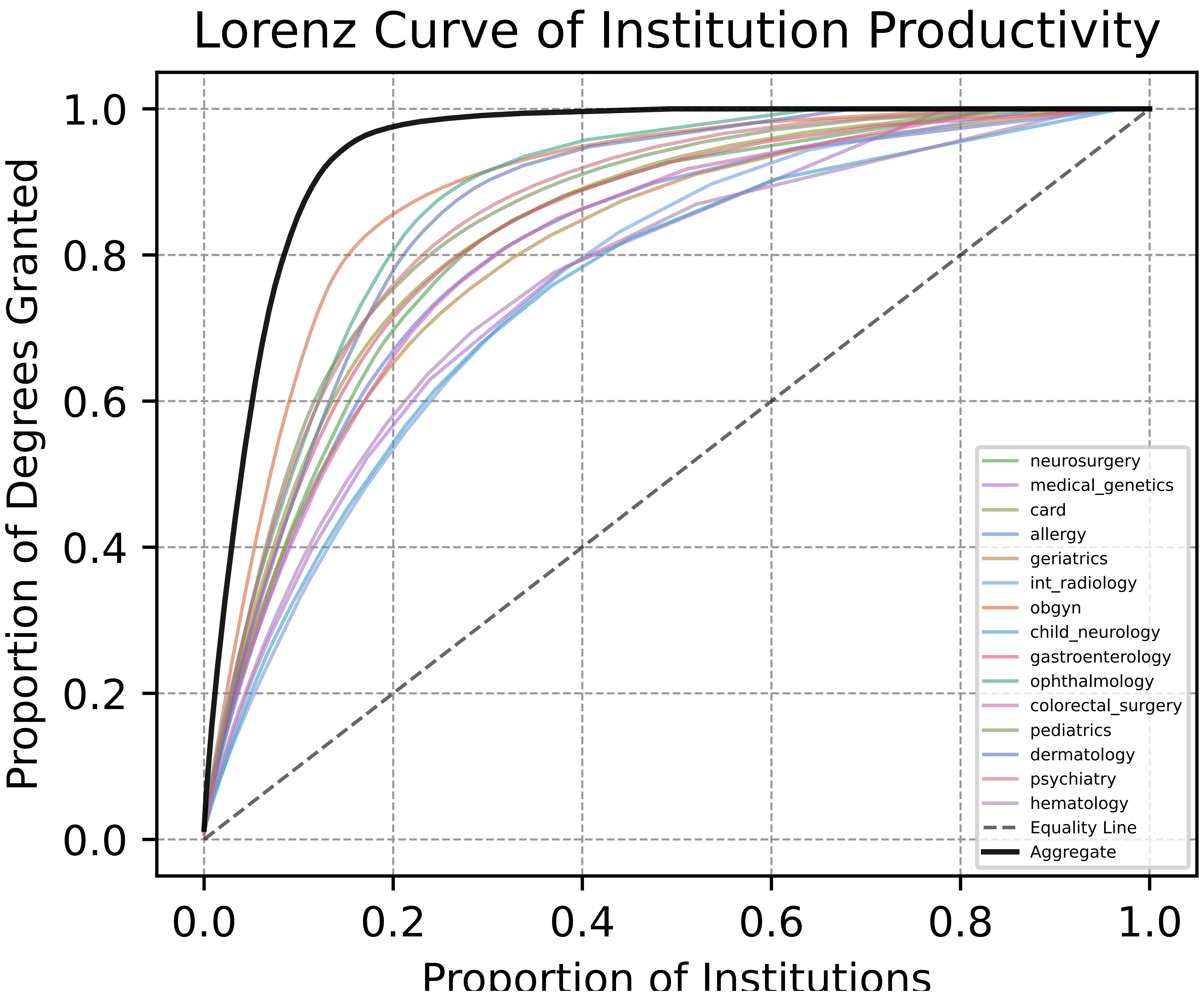
560,368 Edges

Normalized temporal distance



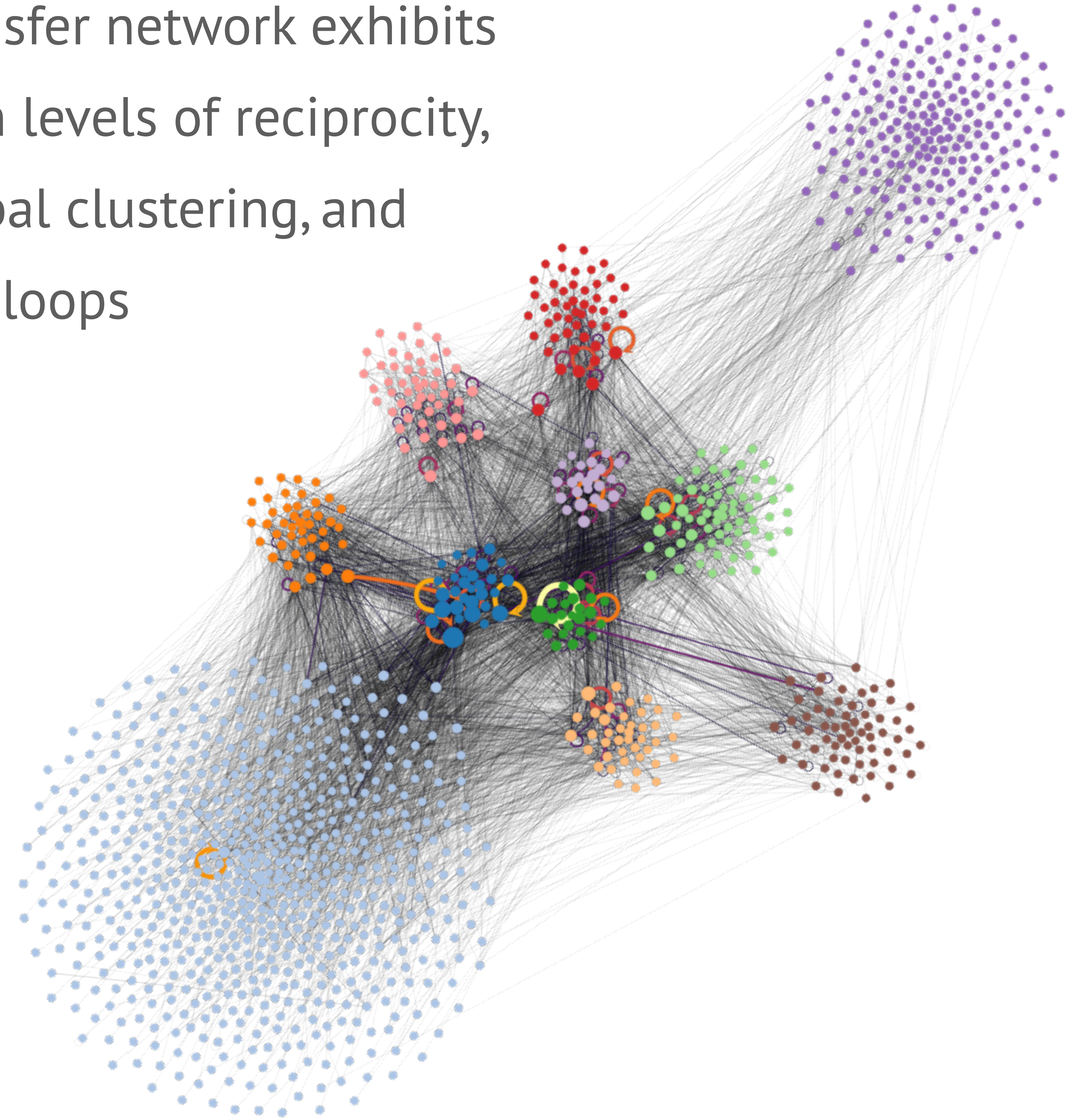
Productivity

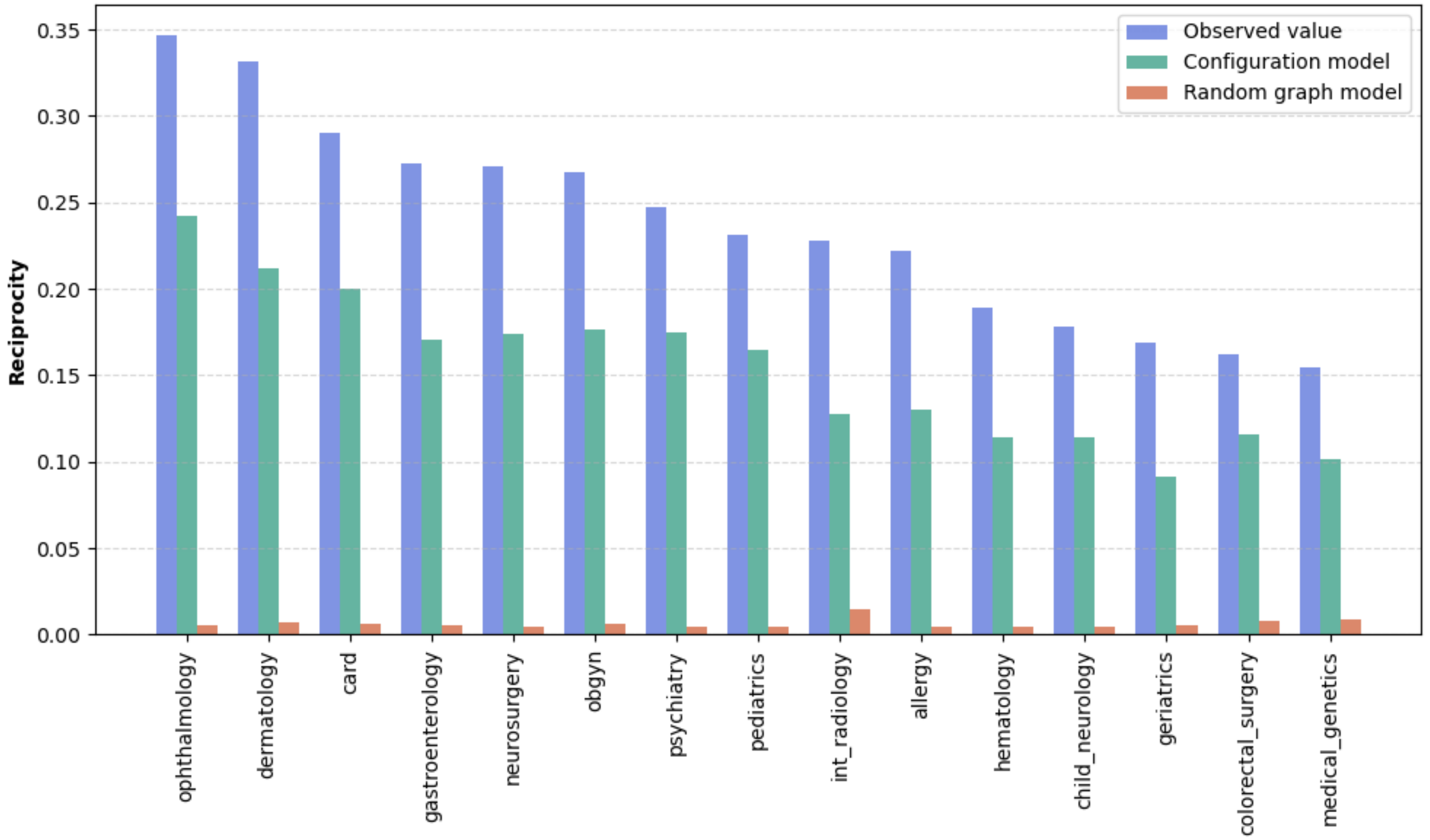
- Productivity of medical institutions is highly unequal
- Across specialties, the top 20% of most productive institutions award 50-85% of degrees
- In aggregate, the 20% most productive institutions awarded >95% of degrees

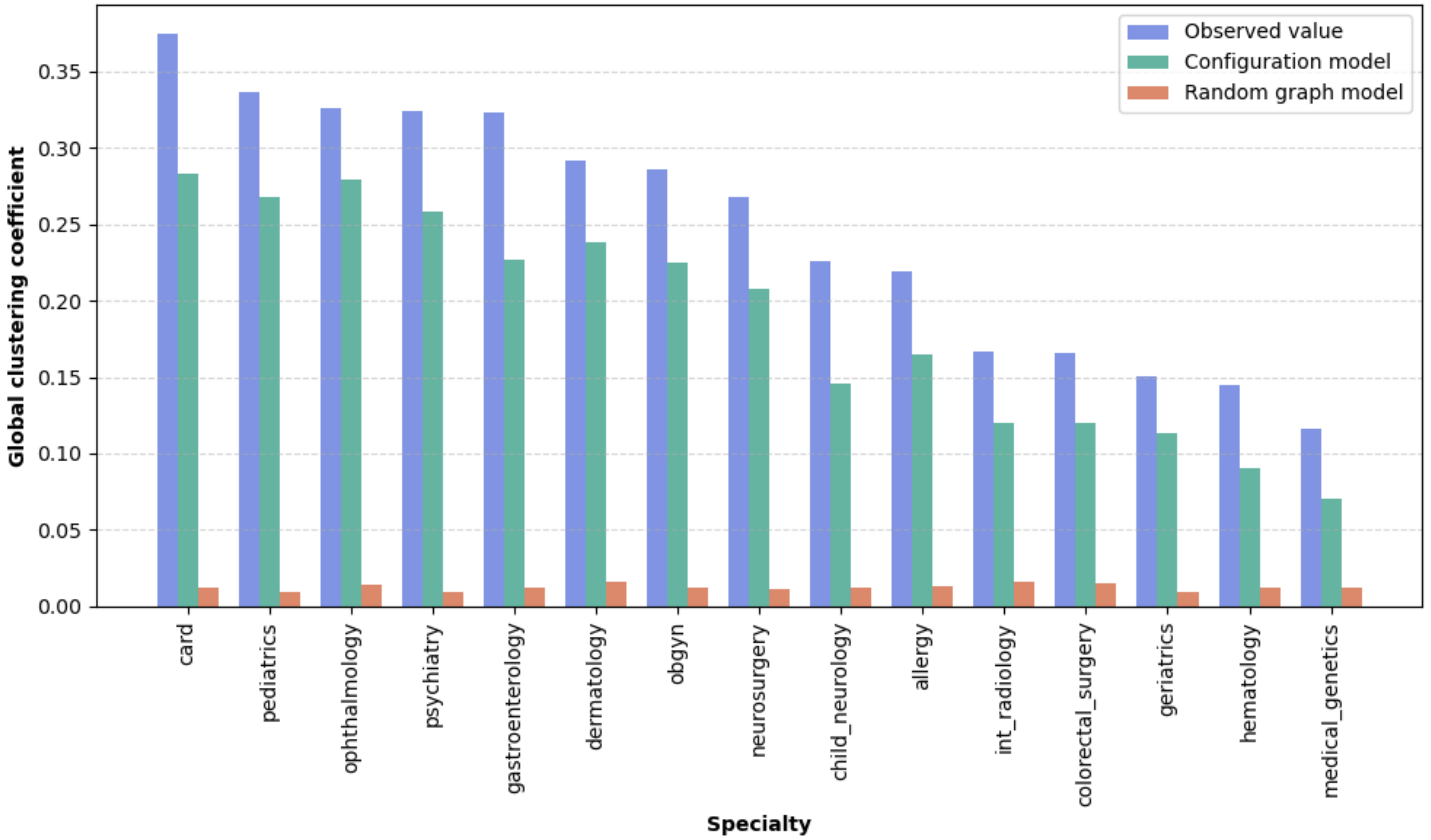


Selectivity

- Transfer network exhibits high levels of reciprocity, global clustering, and self loops







Ranking

- Starting with a weighted directed network where $i -> j$ represents the transfer of a graduate from university i to university j
- We can assume that such a transfer is evidence that institution j endorses institution i because they are satisfied with the quality of graduate(s) from i
- We can model the network as a physical system where each edge is mapped to a spring

$$H(\mathbf{s}) = \sum_{i,j=1}^{N} A_{ij} H_{ij} = \frac{1}{2} \sum_{i,j} A_{ij} (s_i - s_j - 1)^2$$

$$P_{ij}(\beta) = \frac{e^{-\beta H_{ij}}}{e^{-\beta H_{ij}} + e^{-\beta H_{ji}}} = \frac{1}{1 + e^{-2\beta(s_i - s_j)}}$$


SCIENCE ADVANCES | RESEARCH ARTICLE

NETWORK SCIENCE

A physical model for efficient ranking in networks

Caterina De Bacco[1,2*†], Daniel B. Larremore[2,3,4*†], Cristopher Moore[2†]

We present a physically inspired model and an efficient algorithm to infer hierarchical rankings of nodes in directed networks. It assigns real-valued ranks to nodes rather than simply ordinal ranks, and it formalizes the assumption that interactions are more likely to occur between individuals with similar ranks. It provides a natural statistical significance test for the inferred hierarchy, and it can be used to perform inference tasks such as predicting the existence or direction of edges. The ranking is obtained by solving a linear system of equations, which is sparse if the network is; thus, the resulting algorithm is extremely efficient and scalable. We illustrate these findings by analyzing real and synthetic data, including data sets from animal behavior, faculty hiring, social support networks, and sports tournaments. We show that our method often outperforms a variety of others, in both speed and accuracy, in recovering the underlying ranks and predicting edge directions.

INTRODUCTION

In systems of many individual entities, interactions and their outcomes are often correlated with these entities' ranks or positions in a hierarchy. While in most cases these rankings are hidden from us, their presence is nevertheless revealed in the asymmetric patterns of interactions that we observe. For example, some social groups of birds, primates, and elephants are organized according to dominance hierarchies, reflected in patterns of repeated interactions in which dominant animals tend to assert themselves over less powerful subordinates (*1*). Social positions are not directly visible to researchers, but we can infer each animal's position in the hierarchy by observing the network of pairwise interactions. Similar latent hierarchies have been hypothesized in systems of endorsement in which status is due to prestige, reputation, or social position (*2*, *3*). For example, in academia, universities may be more likely to hire faculty candidates from equally or more prestigious universities (*3*).

In all these cases, the direction of the interactions is affected by the status, prestige, or social position of the entities involved. But it is often the case that even the existence of an interaction, rather than its direction, contains some information about those entities' relative prestige. For example, in some species, animals are more likely to interact with others who are close in dominance rank (*4*–*8*); human beings tend to claim friendships with others of similar or slightly higher status (*9*); and sports tournaments and league structures are often designed to match players or teams on the basis of similar skill levels (*10*, *11*). This suggests that we can infer the ranks of individuals in a social hierarchy using both the existence and the direction of their pairwise interactions. It also suggests assigning real-valued ranks to entities rather than simply ordinal rankings, for instance, to infer clusters of entities with roughly equal status with gaps between them.

Here, we introduce a physically inspired model that addresses the problems of hierarchy inference, edge prediction, and significance testing. The model, which we call SpringRank, maps each directed edge to a directed spring between the nodes that it connects and finds real-valued positions of the nodes that minimize the total energy of these springs. Because this optimization problem requires only linear algebra, it can be solved for networks of millions of nodes and edges in seconds.

We also introduce a generative model for hierarchical networks in which the existence and direction of edges depend on the relative ranks of the nodes. This model formalizes the assumption that individuals tend to interact with others of similar rank, and it can be used to create synthetic benchmark networks with tunable levels of hierarchy and noise. It can also predict unobserved edges, allowing us to use cross-validation as a test of accuracy and statistical significance. Moreover, the maximum likelihood estimates of the ranks coincide with SpringRank asymptotically.

We test SpringRank and its generative model version on both synthetic and real data sets, including data from animal behavior, faculty hiring, social support networks, and sports tournaments. We find that it infers accurate rankings, provides a simple significance test for hierarchical structure, and can predict the existence and direction of as-yet unobserved edges. In particular, we find that SpringRank often predicts the direction of unobserved edges more accurately than a variety of existing methods, including popular spectral techniques, minimum violation ranking (MVR), and the Bradley-Terry-Luce (BTL) method.

Related work

Ranking entities in a system from pairwise comparisons or interactions is a fundamental problem in many contexts, and many methods have been proposed. One family consists of spectral methods such as eigenvector centrality (*12*), PageRank (*13*), rank centrality (*14*), and the method of Callaghan *et al.* (*15*). These methods propose various types of random walks on the directed network and therefore produce real-valued scores. However, by design, these methods tend to give high ranks to a small number of important nodes, giving us little information about the lower-ranked nodes. In addition, they often require explicit regularization by adding a small term to every element of the adjacency matrix if the graph of comparisons is not strongly connected.

A second family focuses on ordinal rankings, that is, permutations, that minimize various penalty functions. This family includes MVR (*16*–*18*), SerialRank (*19*), and SyncRank (*20*). MVR imposes a uniform penalty for every violation or "upset," defined as an edge that has a direction opposite to the one expected by the rank difference between the two nodes. Nonuniform penalties and other generalizations are often referred to as agony methods (*21*). For common choices of the penalty function, minimization can be computationally difficult (*17*, *22*), forcing us to use simple heuristics that find local minima.

SerialRank constructs a matrix of similarity scores between each pair of nodes by examining whether they produce similar outcomes when compared with the other nodes, thereby relating the ranking problem to

[1]Data Science Institute, Columbia University, New York, NY 10027, USA. [2]Santa Fe Institute, Santa Fe, NM 87501, USA. [3]Department of Computer Science, University of Colorado, Boulder, CO 80309, USA. [4]BioFrontiers Institute, University of Colorado, Boulder, CO 80303, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: cdebacco@santafe.edu (C.D.B.); daniel.larremore@colorado.edu (D.B.L.); moore@santafe.edu (C.M.)

De Bacco, C., Larremore, D.B., Moore, C., 2018. A physical model for efficient ranking in networks. Sci. Adv. 4, eaar8260.

Ranking

QS world rankings

Top Ranked	Institution
1	Harvard University
2	Stanford University
3	Yale University
4	Columbia University
5	University of Pennsylvania
6	Duke University
7	University of Washington
8	Cornell University

Ranking

QS world rankings

Top Ranked	Institution
1	Harvard University
2	Stanford University
3	Yale University
4	Columbia University
5	University of Pennsylvania
6	Duke University
7	University of Washington
8	Cornell University

Degree centrality

Top Ranked	Institution
1	Duke University
2	Washington University in St. Louis
3	Drexel University
4	University of Pennsylvania
5	Aston University
6	University of Miami
7	Stanford University
8	University of Florida

Ranking

QS world rankings

Top Ranked	Institution
1	Harvard University
2	Stanford University
3	Yale University
4	Columbia University
5	University of Pennsylvania
6	Duke University
7	University of Washington
8	Cornell University

Degree centrality

Top Ranked	Institution
1	Duke University
2	Washington University in St. Louis
3	Drexel University
4	University of Pennsylvania
5	Aston University
6	University of Miami
7	Stanford University
8	University of Florida

Eigenvector centrality

Top Ranked	Institution
1	Duke University
2	Washington University in St. Louis
3	University of Iowa
4	University of Florida
5	Stanford University
6	Texas A&M University
7	University of Washington
8	University of Virginia

Ranking

QS world rankings

Top Ranked	Institution
1	Harvard University
2	Stanford University
3	Yale University
4	Columbia University
5	University of Pennsylvania
6	Duke University
7	University of Washington
8	Cornell University

Degree centrality

Top Ranked	Institution
1	Duke University
2	Washington University in St. Louis
3	Drexel University
4	University of Pennsylvania
5	Aston University
6	University of Miami
7	Stanford University
8	University of Florida

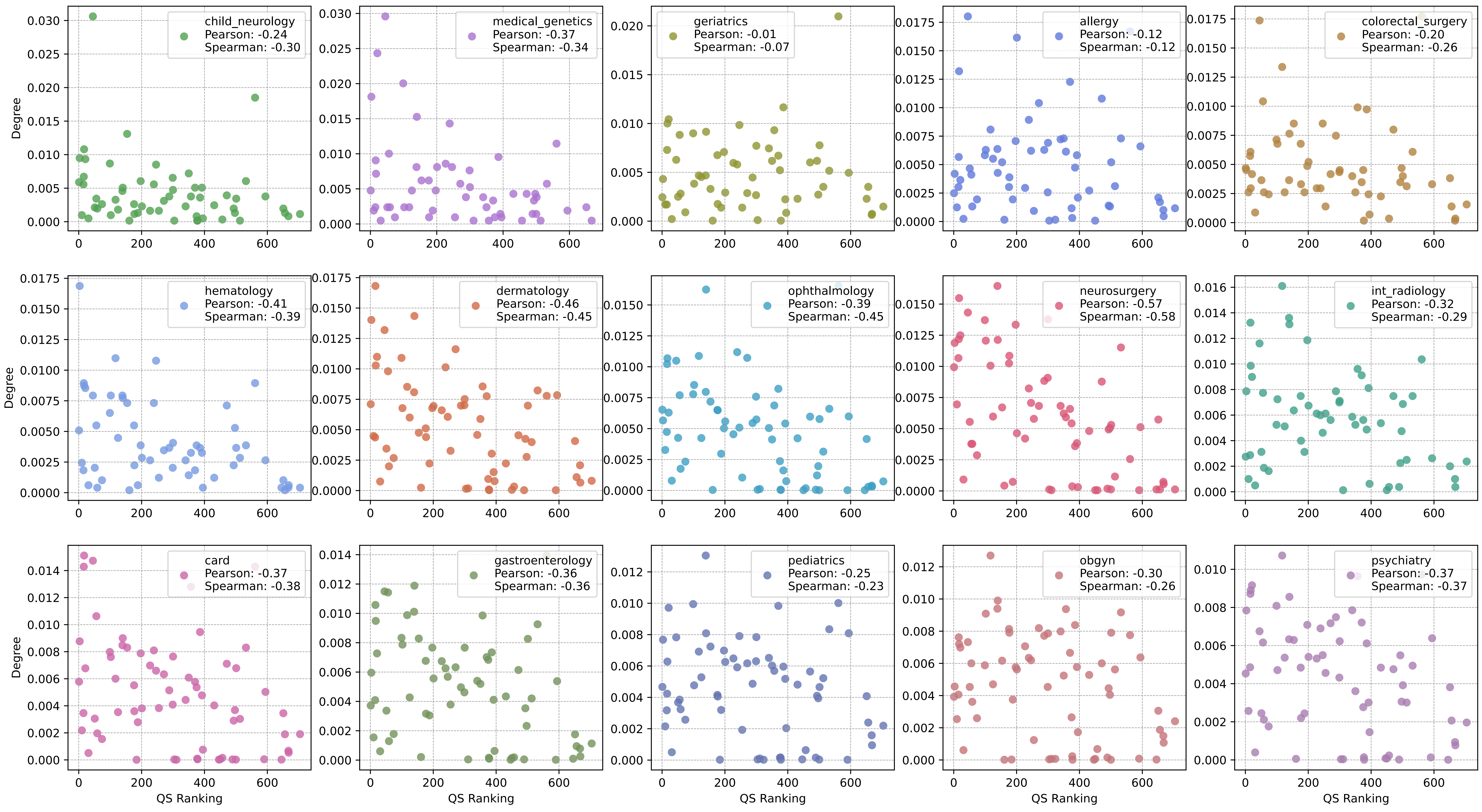
Eigenvector centrality

Top Ranked	Institution
1	Duke University
2	Washington University in St. Louis
3	University of Iowa
4	University of Florida
5	Stanford University
6	Texas A&M University
7	University of Washington
8	University of Virginia

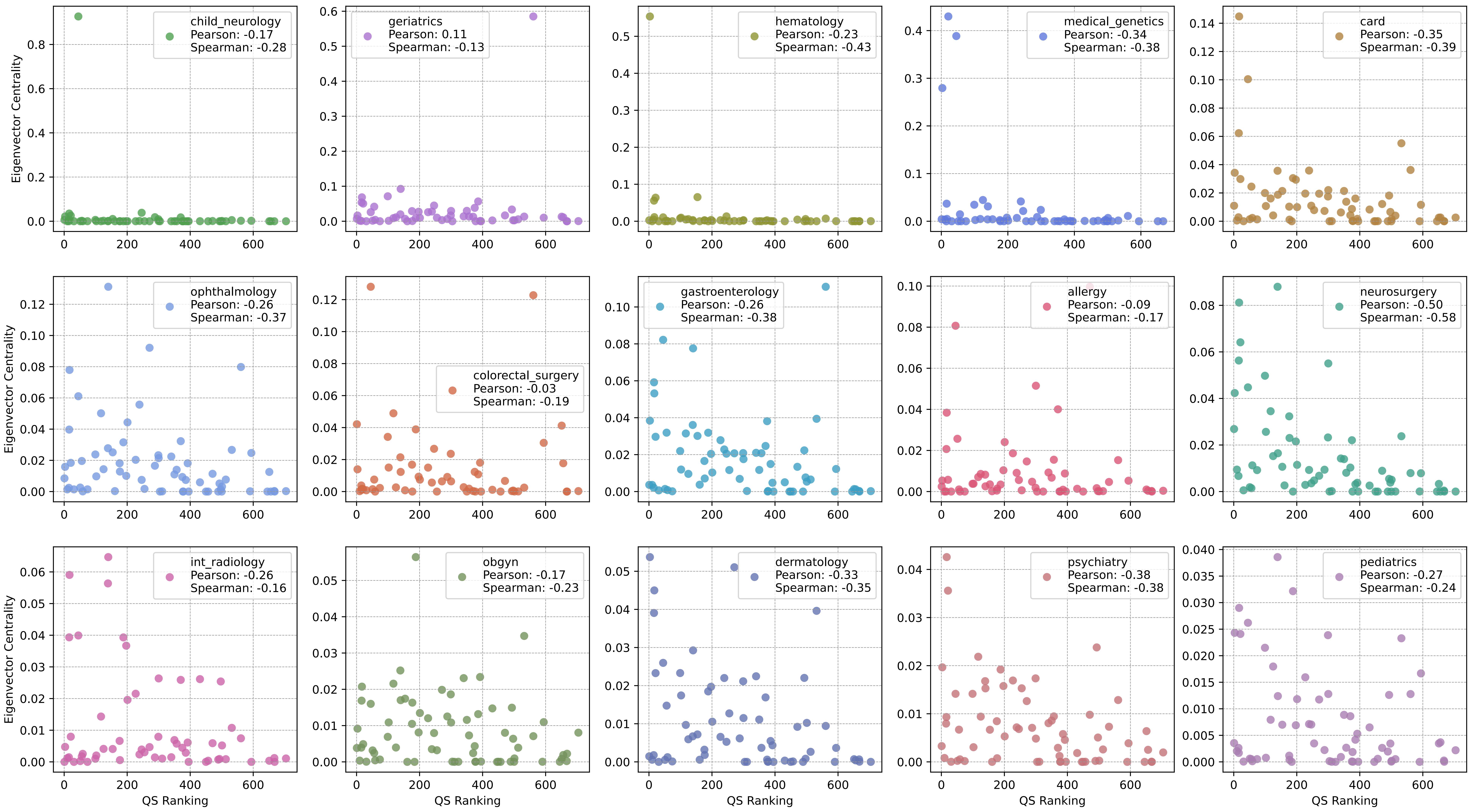
SpringRank

Top Ranked	Institution
1	Harvard University
2	Northwestern University
3	Columbia University
4	Wayne State University
5	Yale University
6	University of Pittsburgh
7	University of South Carolina
8	University of Kansas

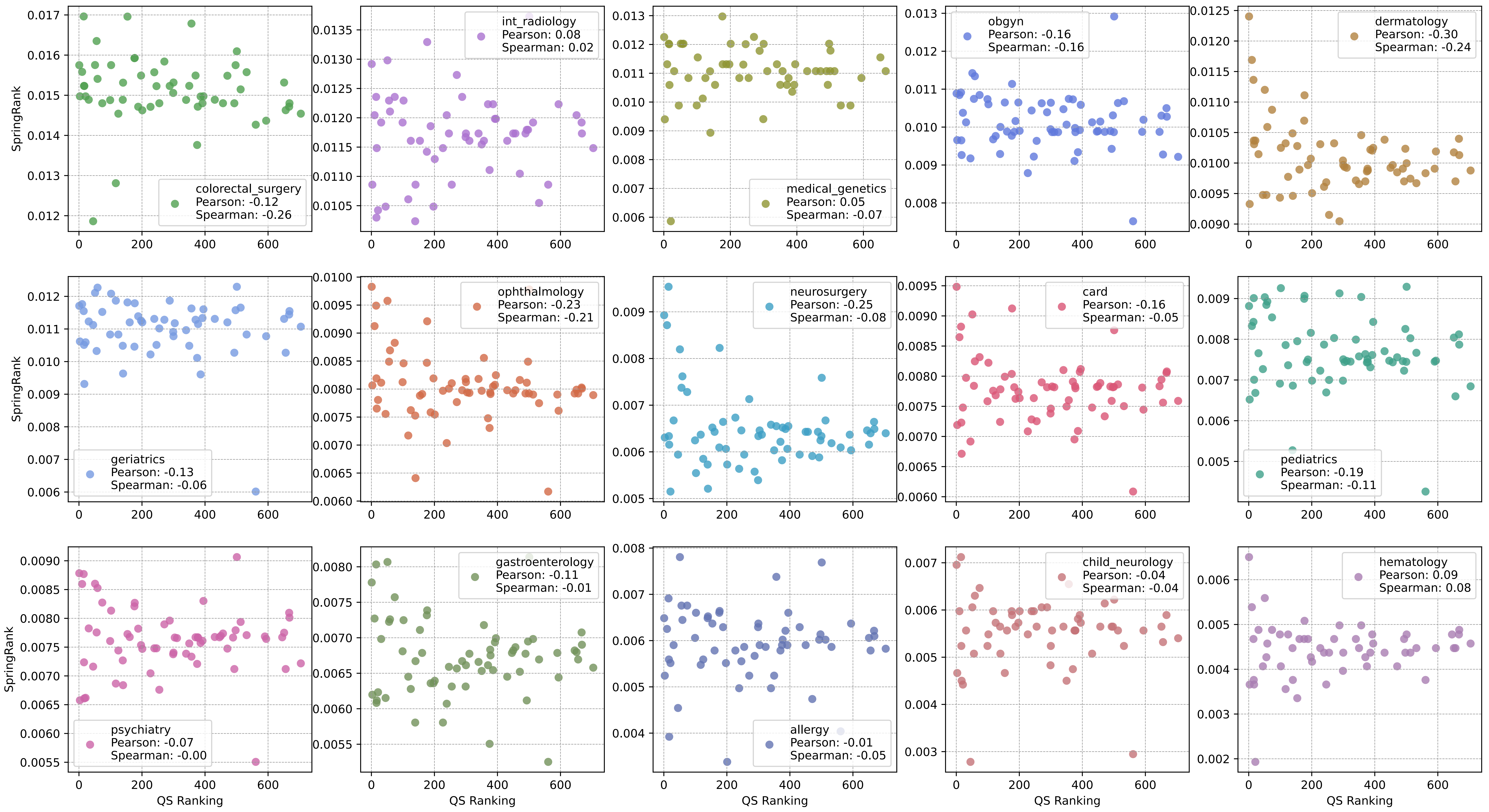
Ranking (degree centrality)



Ranking (eigenvector centrality)



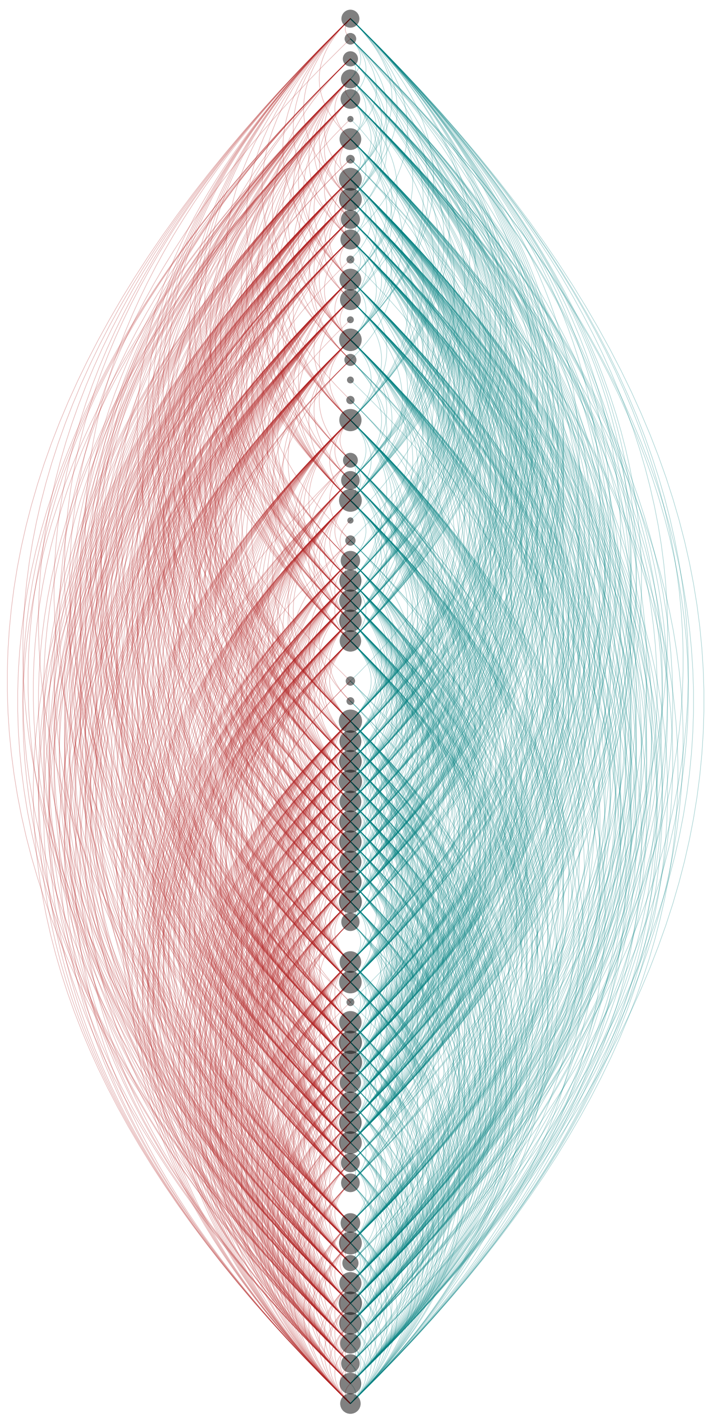
Ranking (SpringRank)



Ranking

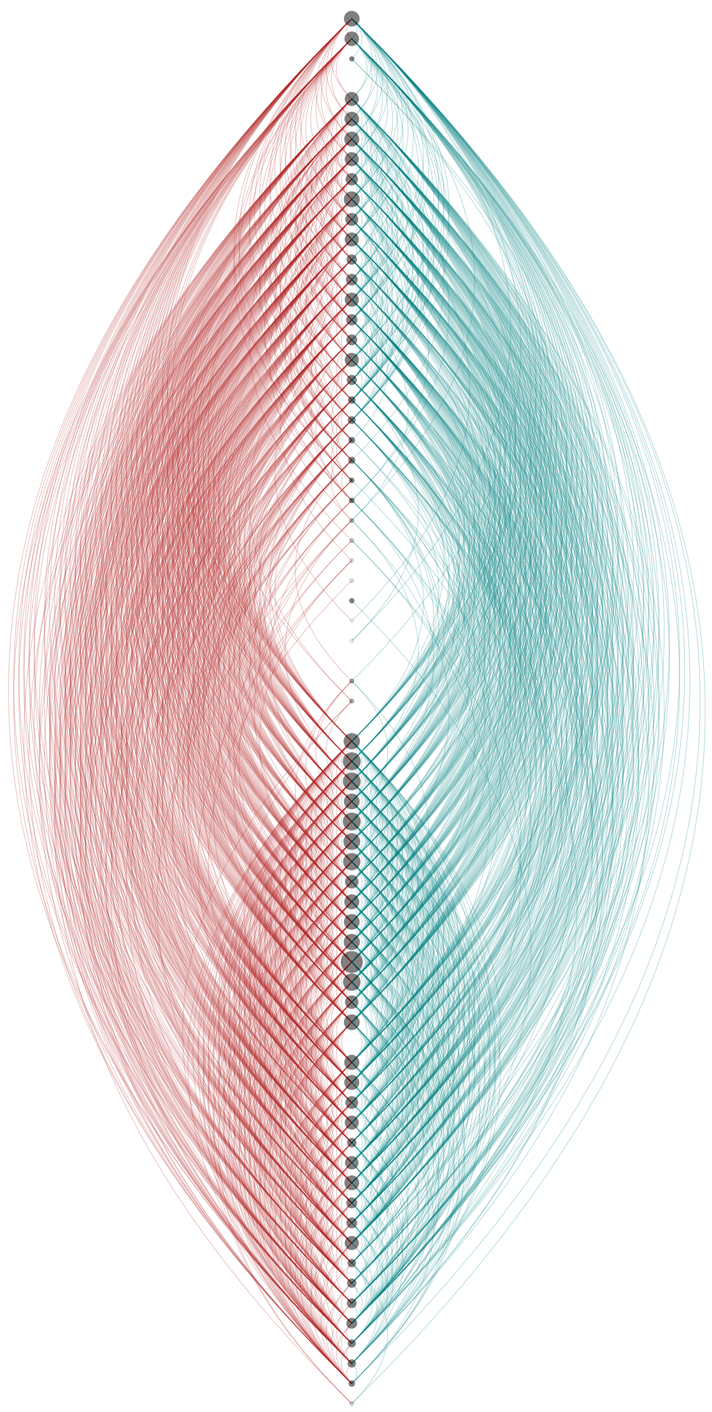
QS world rankings

Fraction of edges going up: 0.49
Fraction of edges going down: 0.51



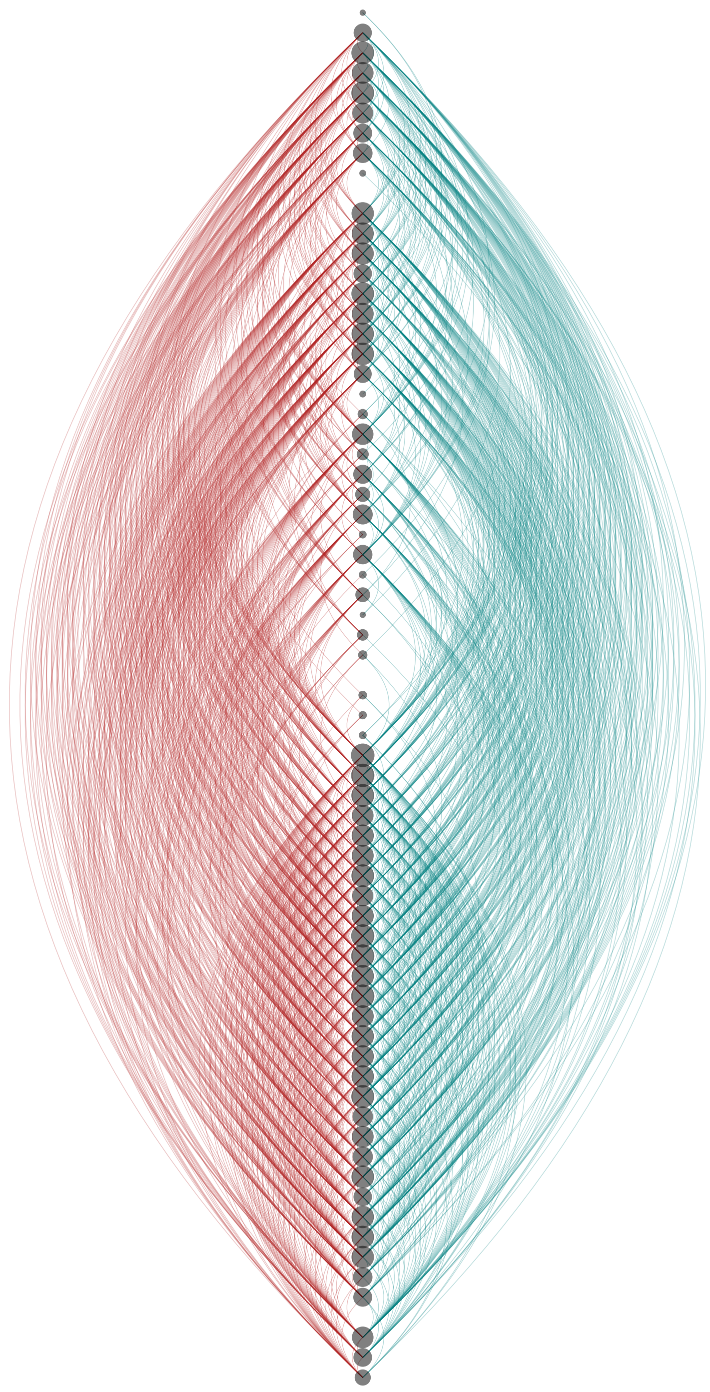
Degree centrality

Fraction of edges going up: 0.54
Fraction of edges going down: 0.46



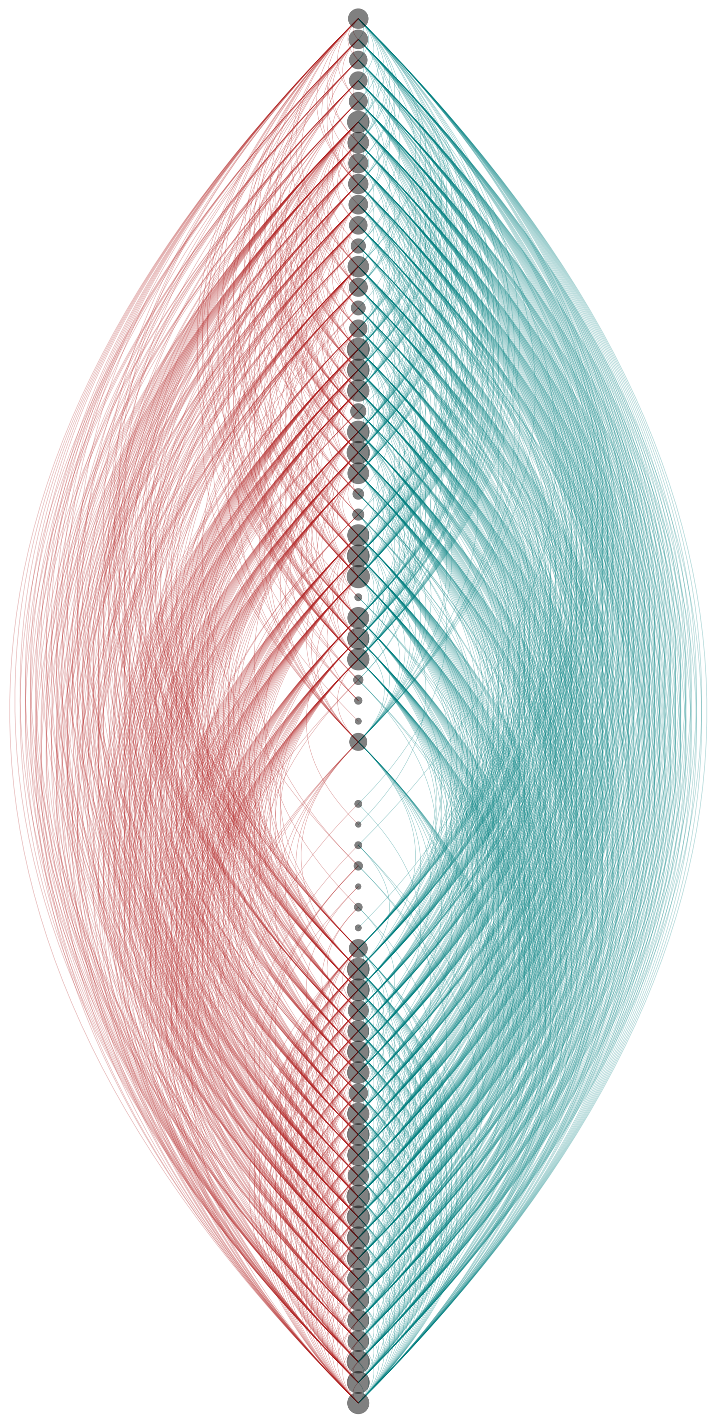
Eigenvector centrality

Fraction of edges going up: 0.53
Fraction of edges going down: 0.47



SpringRank

Fraction of edges going up: 0.43
Fraction of edges going down: 0.57



Whole network:

$f_{up} = 0.63$
$f_{down} = 0.37$

$f_{up} = 0.69$
$f_{down} = 0.31$

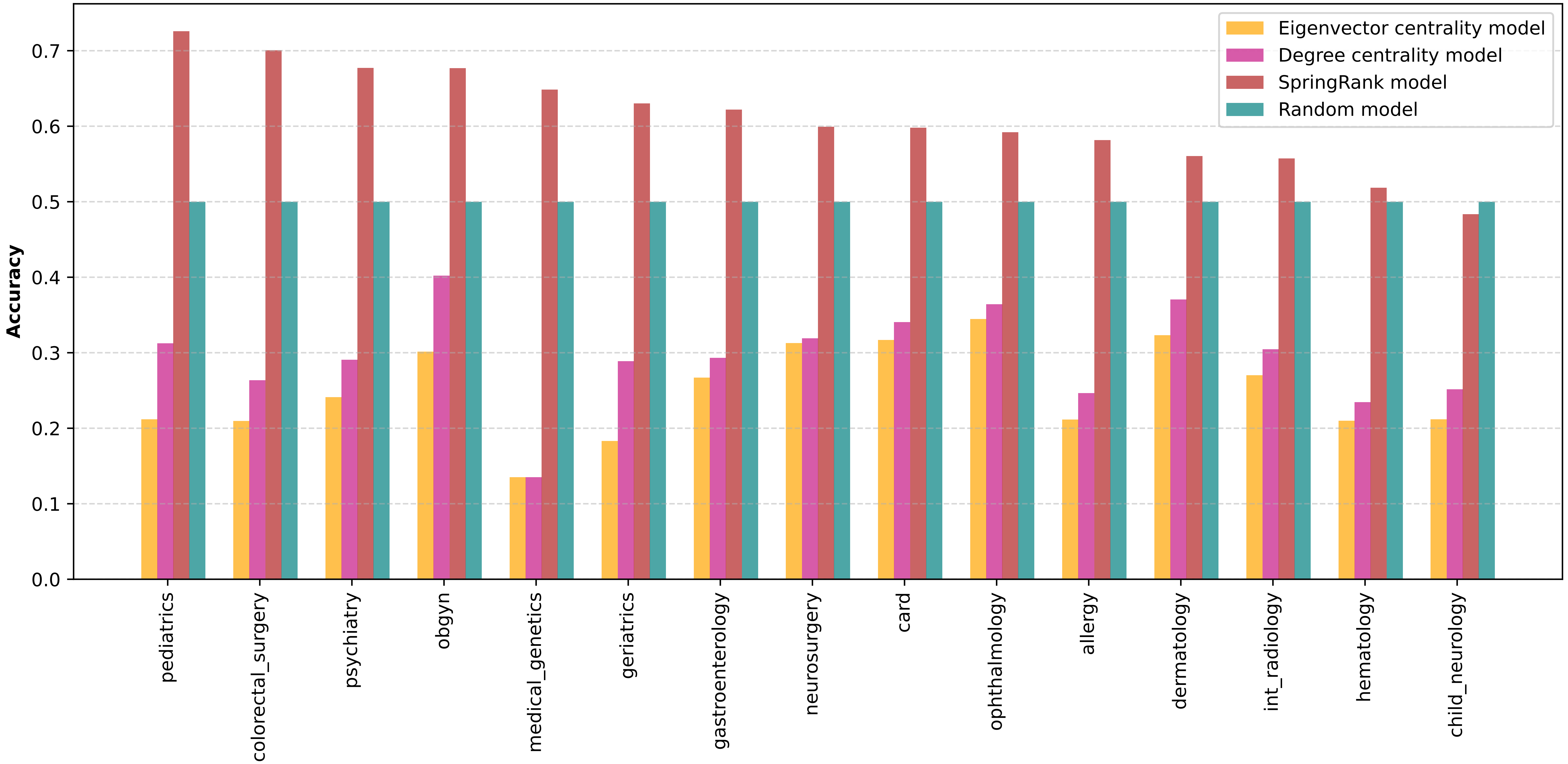
$f_{up} = 0.33$
$f_{down} = 0.67$

Link prediction

- Question: Given a set of training data, can we predict the direction of unseen edges?
- Approach:
  - Calculate ranks based on transfer data from 1980 to 2009
  - For each edge in 2010, given node i and node j, predict whether the edge goes from $i -> j$ or $j -> i$.

Link prediction

- Question: Given a set of training data, can we predict the direction of unseen edges?
- Approach:
    - Calculate ranks based on transfer data from 1980 to 2009
    - For each edge in 2010, given node i and node j, predict whether the edge goes from $i \rightarrow j$ or $j \rightarrow i$.



Takeaways

- Training network of US medical professionals is highly unbalanced
- Training network exhibits high levels of reciprocity, self-loops and has more triangles than expected from the degree sequence alone
- Network based rankings are only weakly correlated with public rankings
- Spring rank can better predict the direction of edges than random and compared to degree based and eigenvector based rankings