BRAINWORKS

Mohammad M. Ghassemi, Ph.D.

Assistant Professor, Computer Science, Michigan State University
Data and Technology Advancement National Service Scholar, NIH

Career straddles academia, industry, and entrepreneurship

Interested in both the theoretical development, and practical deployment, of technology

Mohammad Ghassemi, Ph.D.

Is a assistant professor of computer science at MSU with extensive national and international consulting experience. He holds graduate degrees from MIT, and Cambridge (UK). He was a director of data science at S&P Global, and a strategic consultant with BCG. He has over ten years of technical and strategic consulting experience for many of the world's largest brands.

CV Website Email

BCG BOSTON CONSULTING GROUP

S&P Global

NIH

ghamut

UNIVERSITY OF CAMBRIDGE

Massachusetts Institute of Technology

MICHIGAN STATE UNIVERSITY Founded 1855

Human Augmentation and AI Lab

develop tools and systems that combine human and machine intelligence (A.I.) to solve problems that neither humans nor machines can solve as effectively alone. More specifically, we develop new theoretical knowledge and practical tools for *Augmented Intelligence (A-I)*: the enhancement of individual or collective cognitive function through the use of technology and social/environmental factors.

Data scholar project supports mission of the BRAIN Initiative

BRAIN supports technologies that promote a dynamic understanding of the brain

Deliver a functional proof-of-concept of the "*BRAIN initiative Workspace to ORganize the Knowledge Space*" platform ("BRAIN WORKS" hereafter) – a tool for the discovery of more comprehensive theories of brain function through knowledge integration.

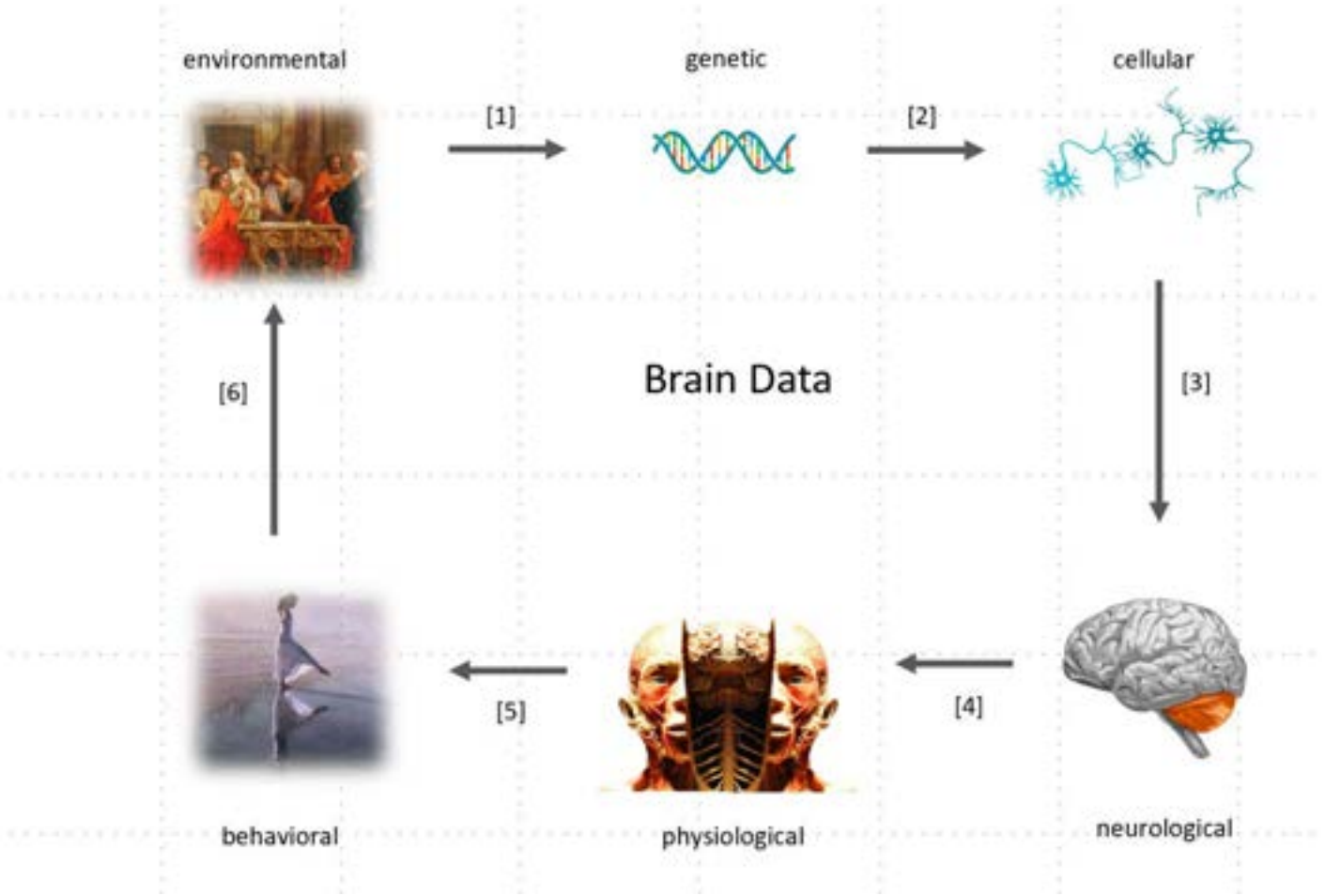
BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

Data Science is Pivotal: by automating tasks, augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function, and definition of new research horizons.



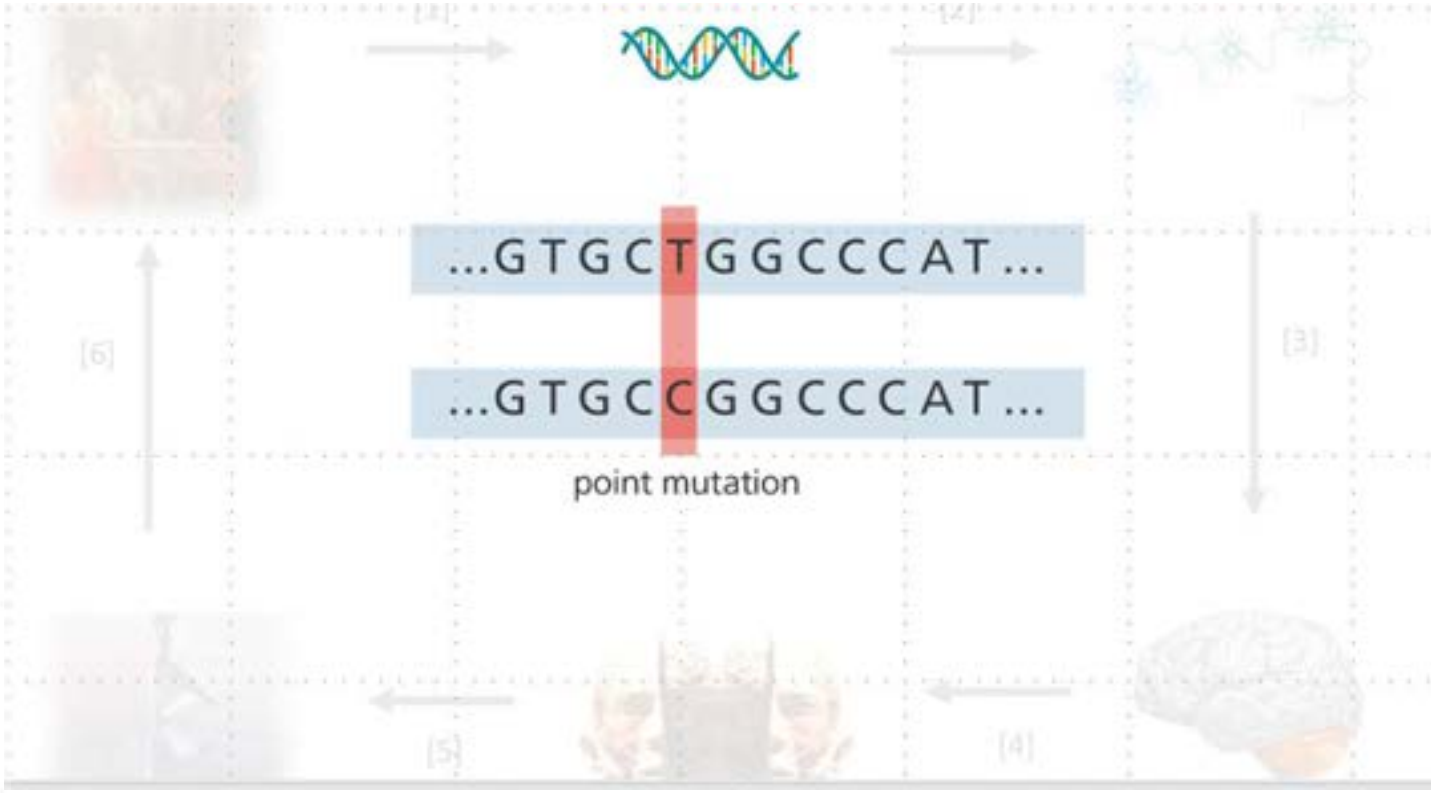
BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

Data Science is Pivotal: by automating tasks, augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function, and definition of new research horizons.



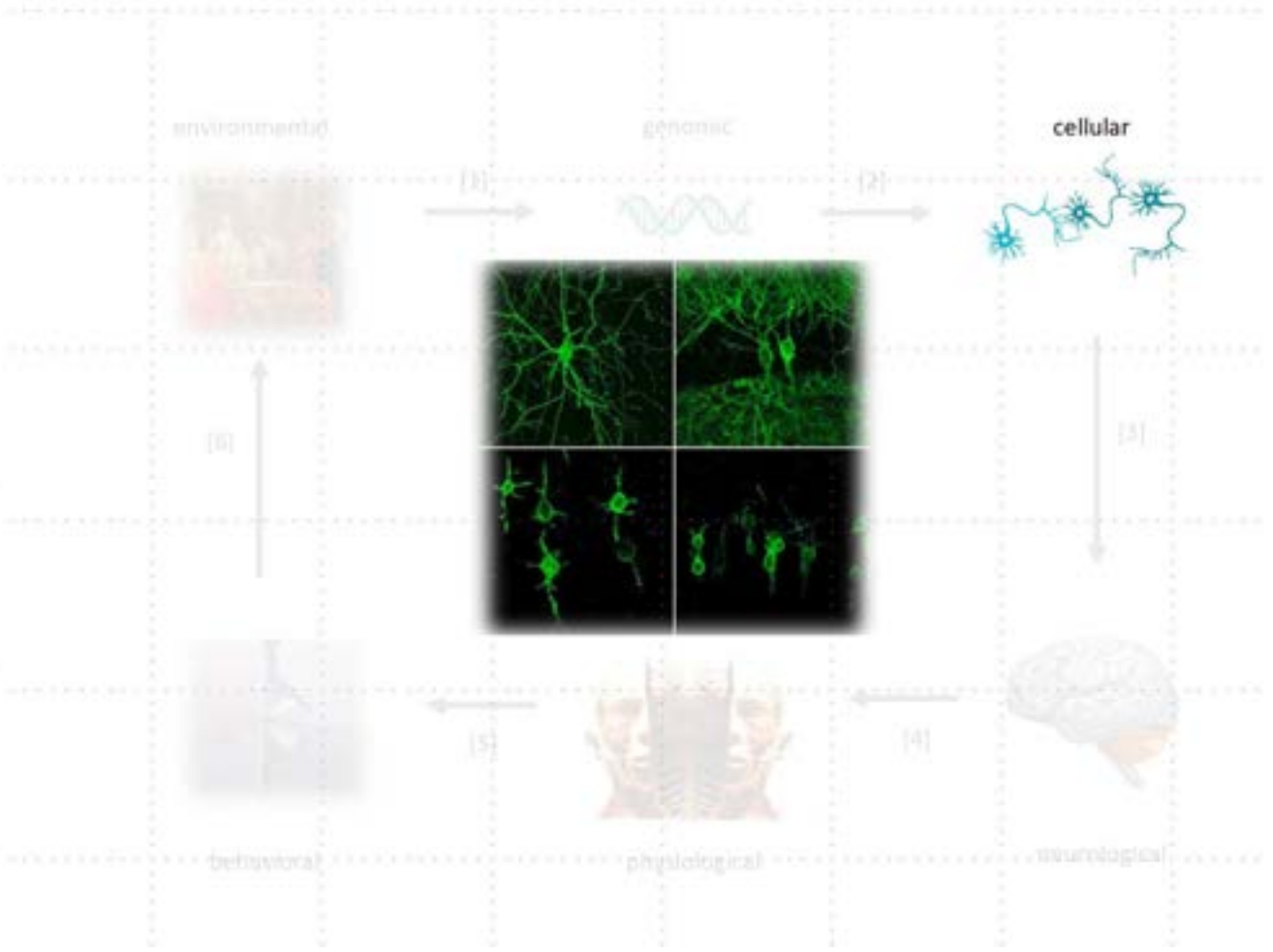
BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

Data Science is Pivotal: by automating tasks, augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function, and definition of new research horizons.



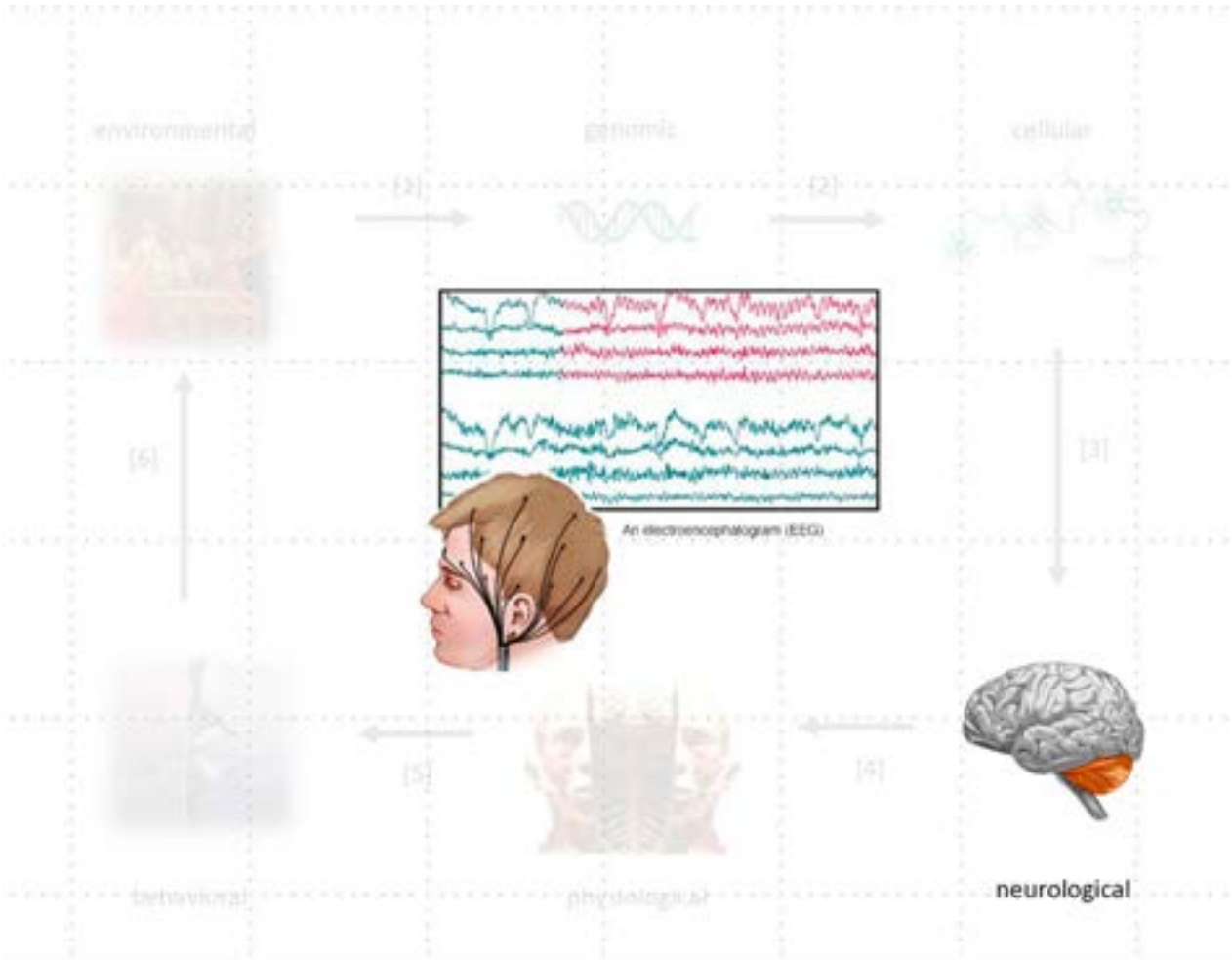
BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

Data Science is Pivotal: by automating tasks, augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function, and definition of new research horizons.



BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

Data Science is Pivotal: by automating tasks, augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function, and definition of new research horizons.

BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

Data Science is Pivotal: by automating tasks, augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function, and definition of new research horizons.

behavioral

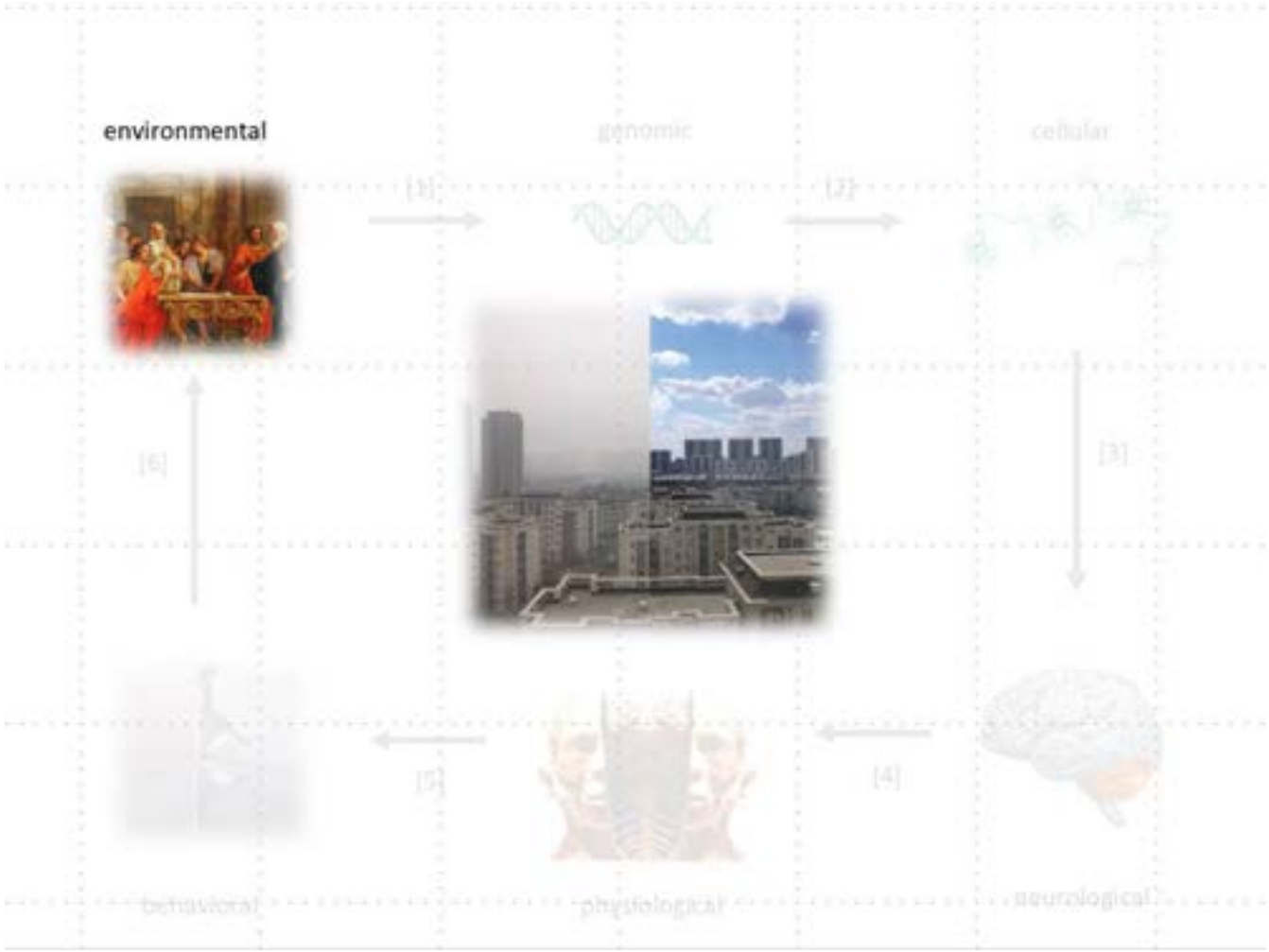
BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

Data Science is Pivotal: by automating tasks, augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function, and definition of new research horizons.



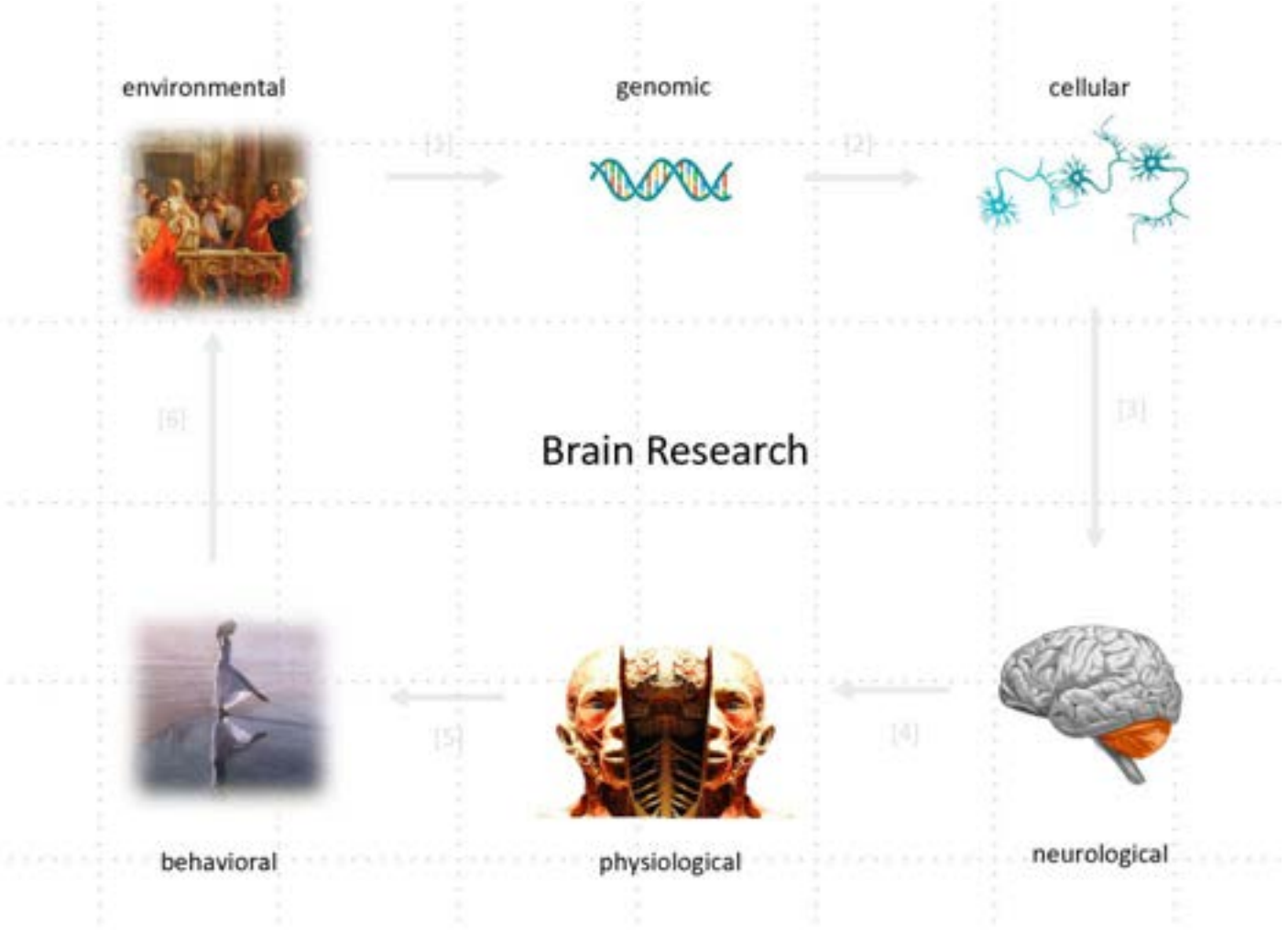
BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

Data Science is Pivotal: by automating tasks, augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function, and definition of new research horizons.



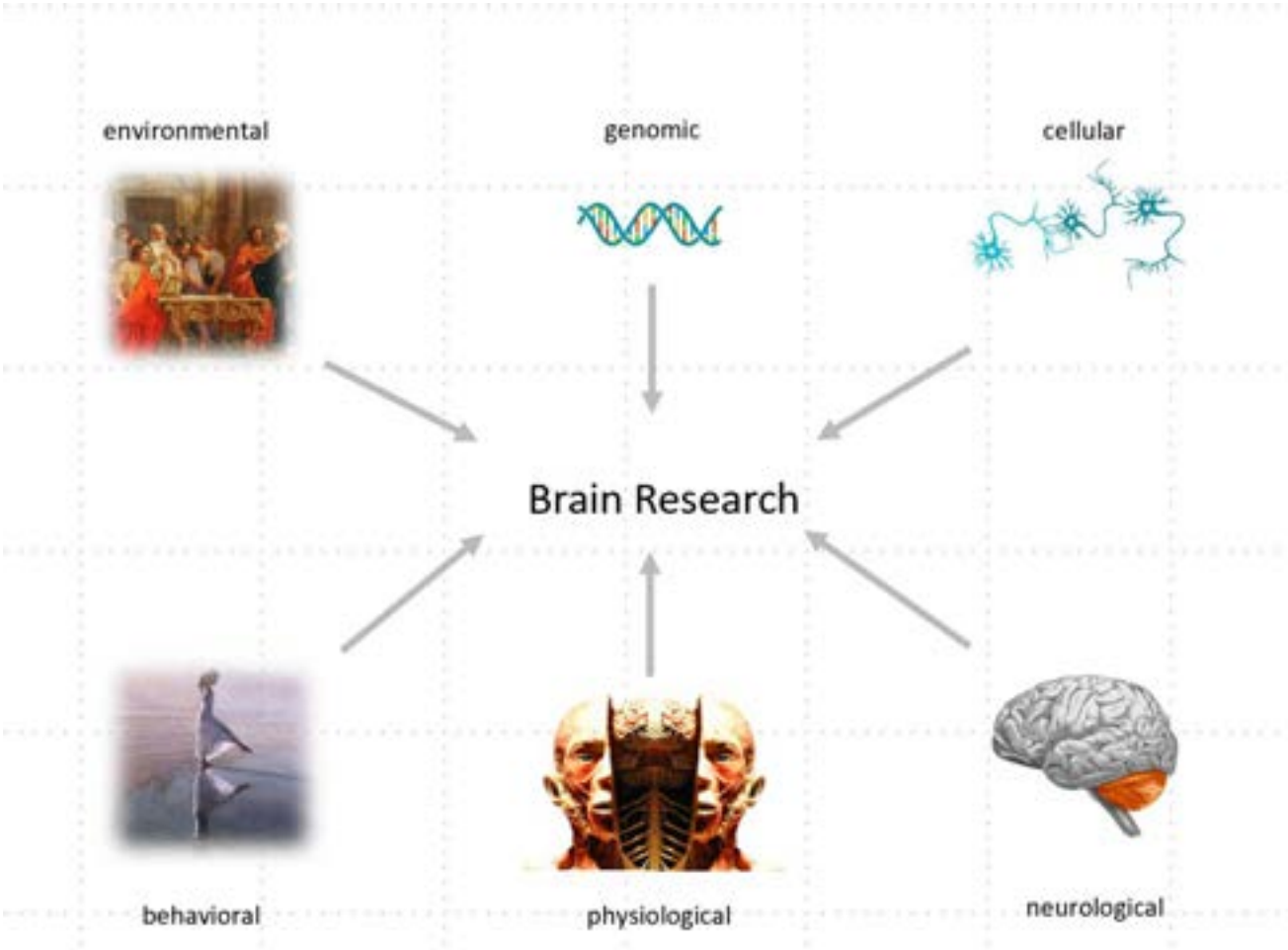
BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

Data Science is Pivotal: by automating tasks, augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function, and definition of new research horizons.



[1-6]: See References Slide

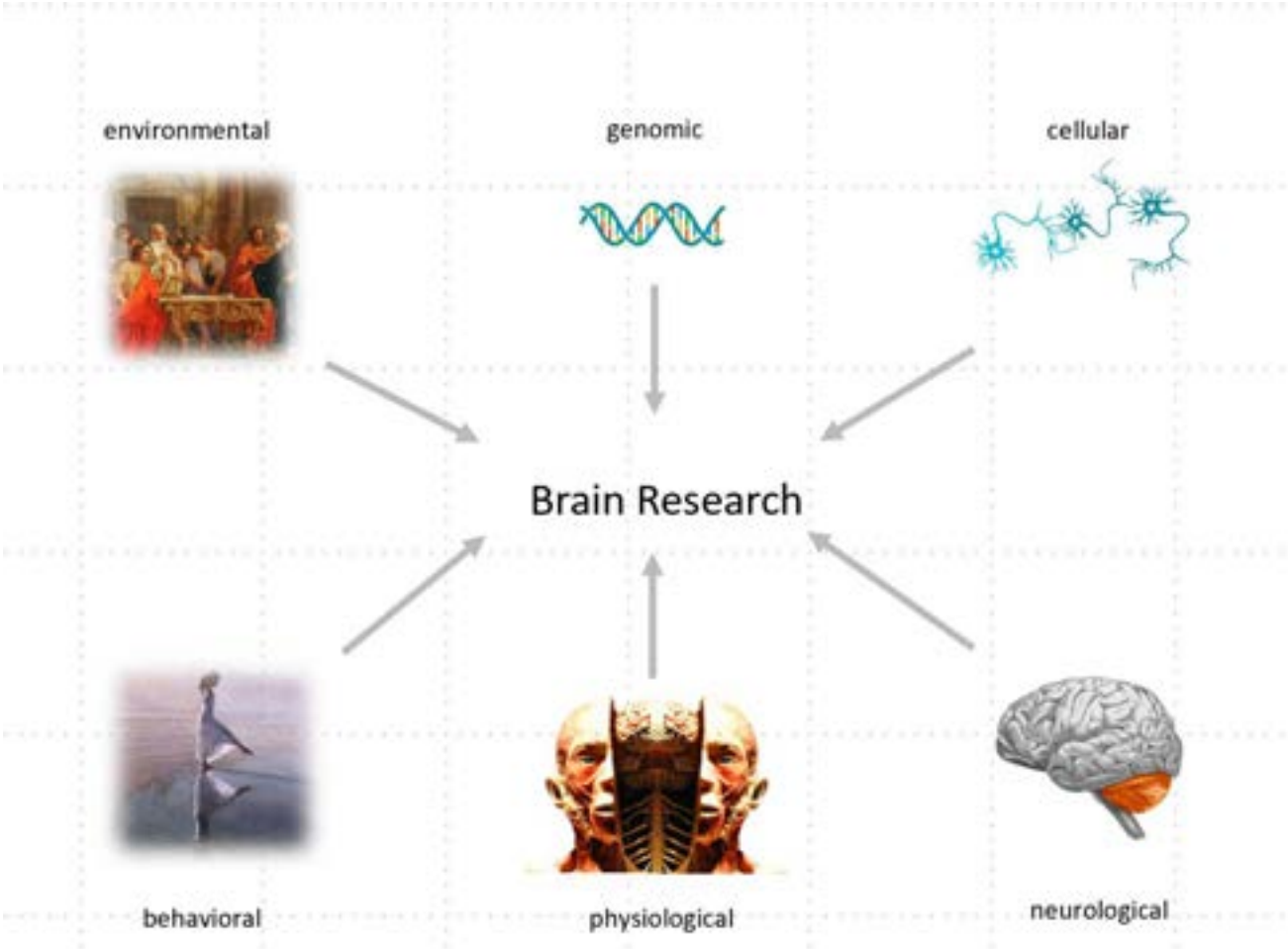
BRAINWORKS will support discovery of new brain theories

Discovery through integration of multi-modal, multi-scale brain data and knowledge

Brain Data is Heterogeneous: exists in multiple modalities, scales and levels of resolution. The heterogeneity of brain data requires thoughtful approaches to data storage, analysis and representation.

Discovery Requires Integration: next generation brain theories require a holistic approach to the brain where the plurality of contexts that impact brain structure and function are accounted for.

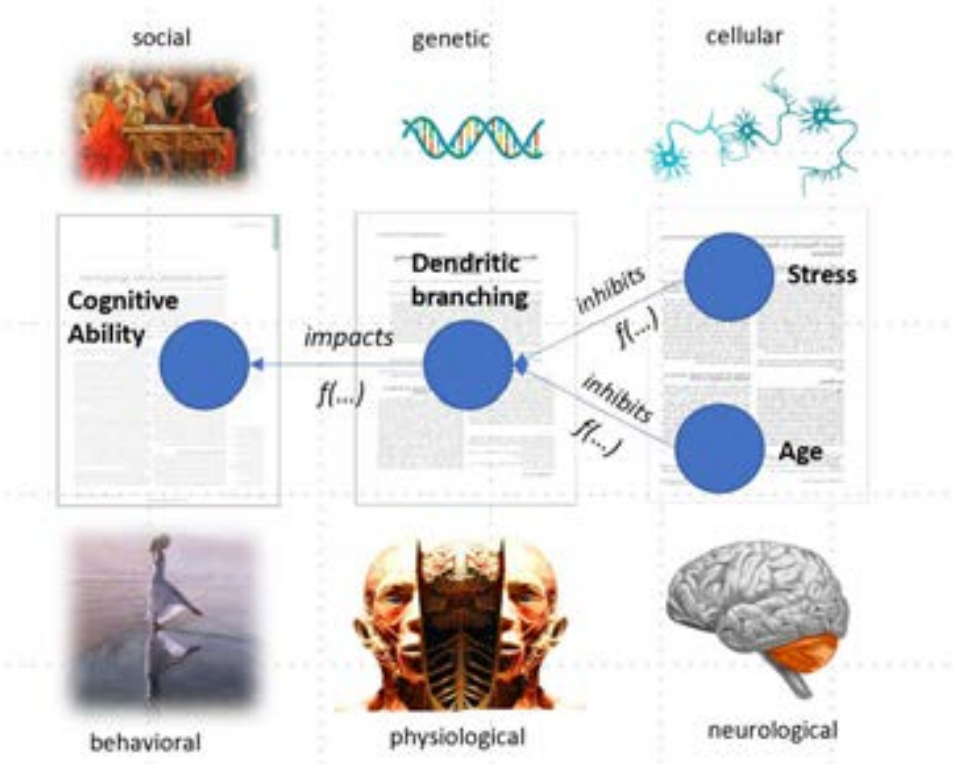
Data Science is Pivotal: by automating tasks and augmenting investigator capabilities we may assist with the discovery of holistic theories of brain function and define new research horizons.



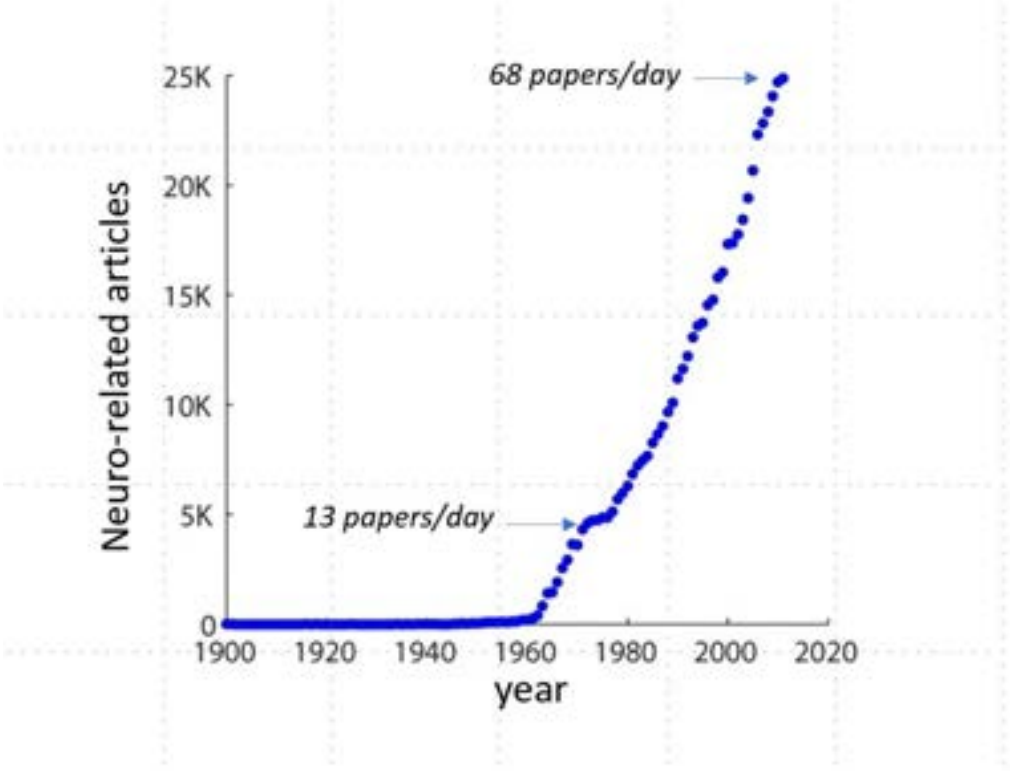
BRAINWORKS converts scientific papers into knowledge graphs

Allows for exploration of scientific literature, and integration of findings across papers

BRAINWORKS: is an web application being developed in 2021 that uses AI to organize the neuroscience literature as an intuitive and interactive knowledge graph.



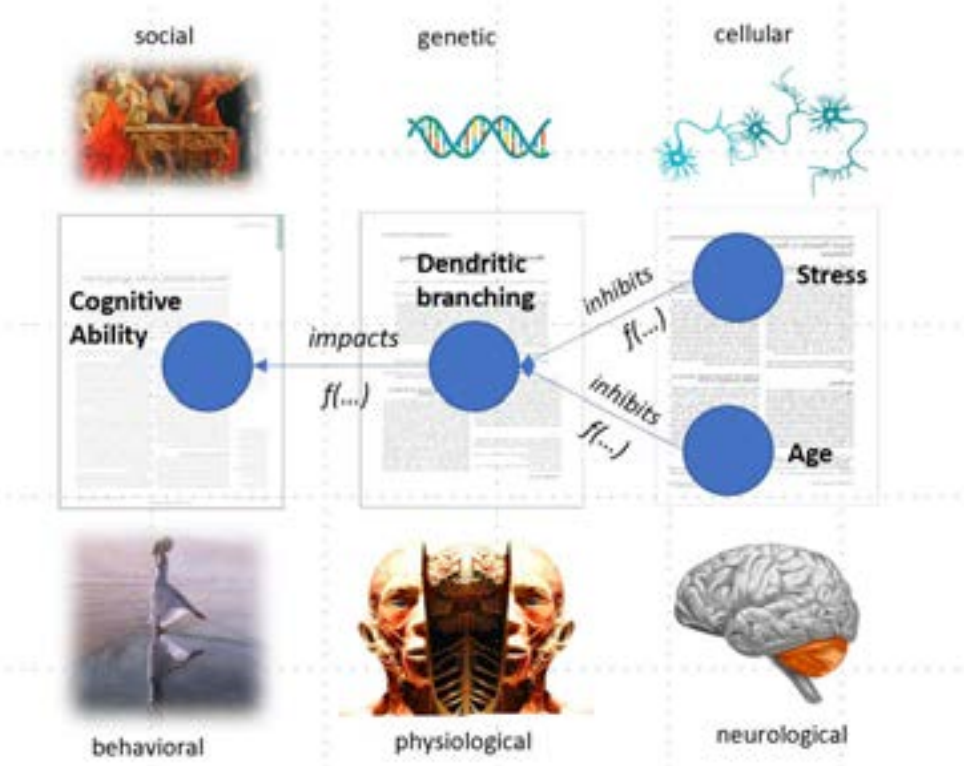
Motivation: Neuroscience theories and knowledge are distributed across a complex, rapidly evolving scientific landscape that no one person can fully master.



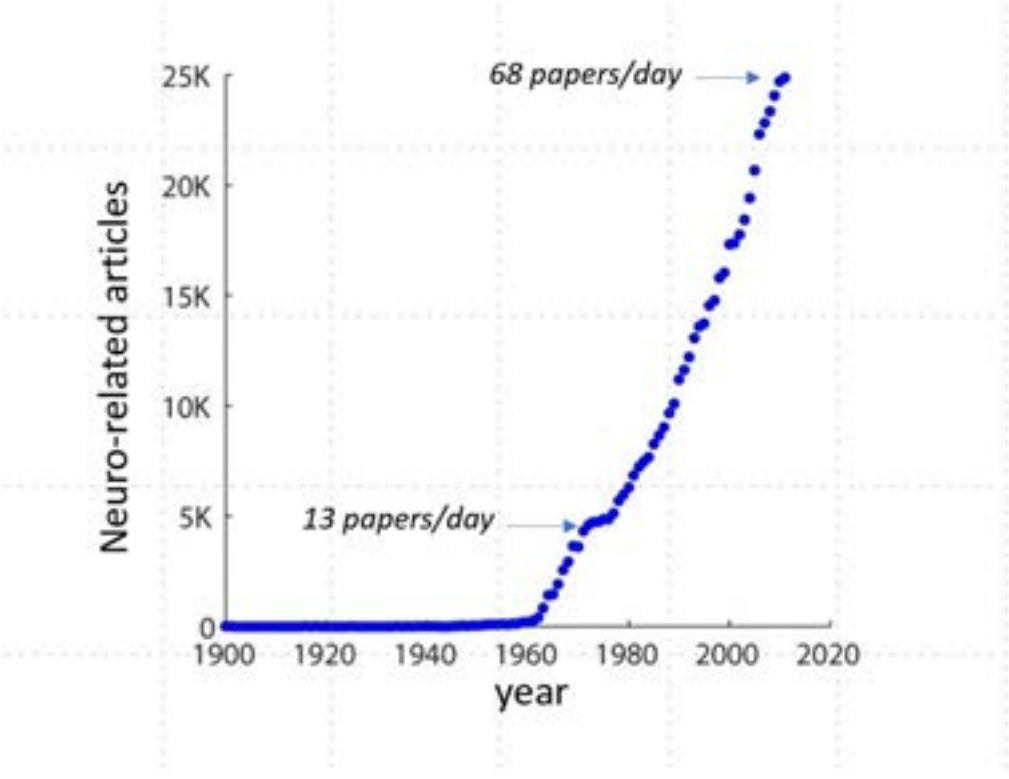
BRAINWORKS converts scientific papers into knowledge graphs

Allows for exploration of scientific literature, and integration of findings across papers

BRAINWORKS: is an web application being developed in 2021 that uses AI to organize the neuroscience literature as an intuitive and interactive knowledge graph.

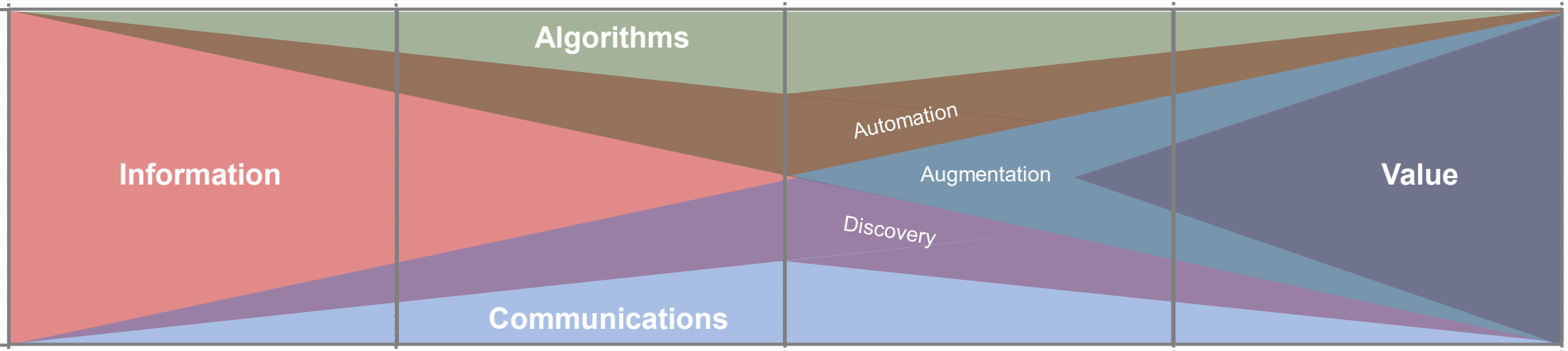


Motivation: Neuroscience theories and knowledge are distributed across a complex, rapidly evolving scientific landscape that no one person can fully master.



Project developed in five phases over a twelve-month period

Initial efforts focus on data preparation transitioning to outcome generation and value demonstration.

1. Specification *January, 2021*	2. Data to Information *February – May, 2021*	3. Algorithms and Comms. *June – August, 2021*	4. Augmentation *September – November, 2021*	5. Value *December, 2021*
				
1.1 Define value 1.2 Define outcomes 1.3 Define Foundations	2.1 Data characterization 2.2 Data centralization 2.3 Data cleansing	3.1 Automation outcomes 3.2 Communication outcomes 3.3 Performance Evaluation	4.1 Augmentation outcomes 4.2 Integration of novel data 4.3 Performance Evaluation	5.1 Demonstration 5.2 Documentation 5.3 Hand-off

To start, we spoke with several internal and external experts

Definition of the value, data science outcomes, and required foundations

1. Specification

January, 2021

1.1 Define value

1.2 Define outcomes

1.3 Define Foundations

Requirements captured by synthesizing needs of NIH & Scientific community.

Acknowledgements

Grace Peng
Program Director
NIBIB

Jim Gnadt
Program Director
NINDS

Karen David
Program Director
NINDS

Christina Fang
Program Manager
NINDS

Michele Ferrante
Program Director
NIMH

Amy Poremba
Program Director
NIDCD

Susan Wright
Program Director
NIDA

Tatiana Pasternik
Scientific Review Officer
NINDS

Elizabeth Buffalo
Professor
Washington

Ilya Nemenman
Professor
Emory

Edmund Talley
Program Director
NINDS

Olivier Burton
Program Director
NIDA

Susan Volman
Program Director
NIDA

Martha Flanders
Program Director
NEI

Merav Sabri
Program Director
NCCIH

Holly Moore
Program Officer
NIDARossi, Andrew

John Ngai
BRAIN Director
NINDS

Elizabeth Powell
Program Director
NIAAA

David Kleinfeld
Professor
UCSD

Mark Reimers
Professor
MSU

William Lytton
Professor
SUNY Downstate

Fidel Santamaria
Professor
UTSA

Carina Curto
Professor
Penn State

Brent Dorion
Professor
Pittsburgh

Larry Abbot
Professor
Columbia

Marc Howard
Professor
BU

Greg DeAngelis
Professor
Rochester

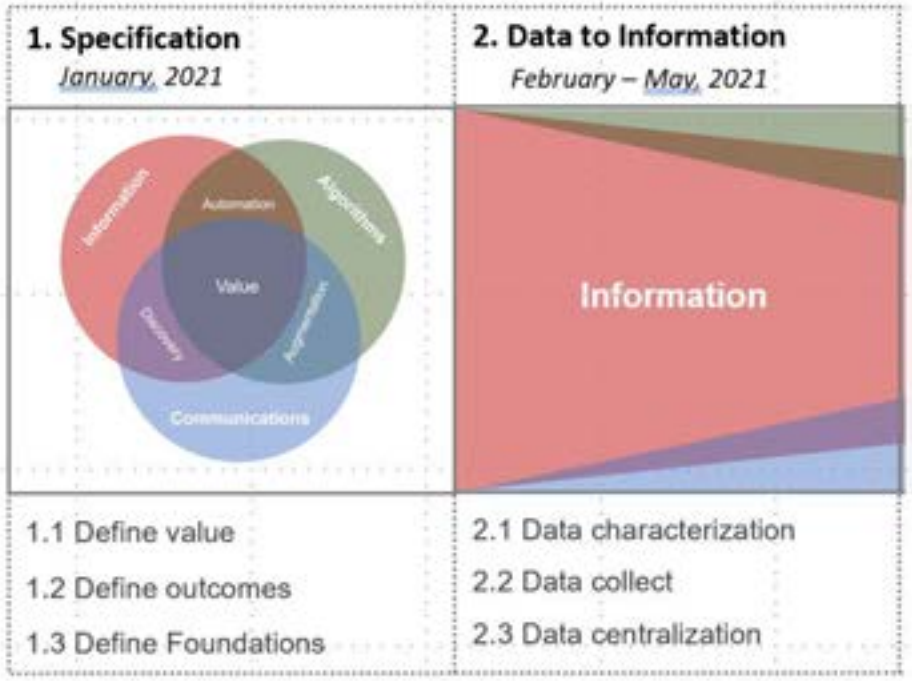
Ivan Soltesz
Professor
Stanford

Jan Drugowitsch
Professor
Harvard

Florian Engert
Professor
Harvard

Our next step was to collect data to build the knowledge graph

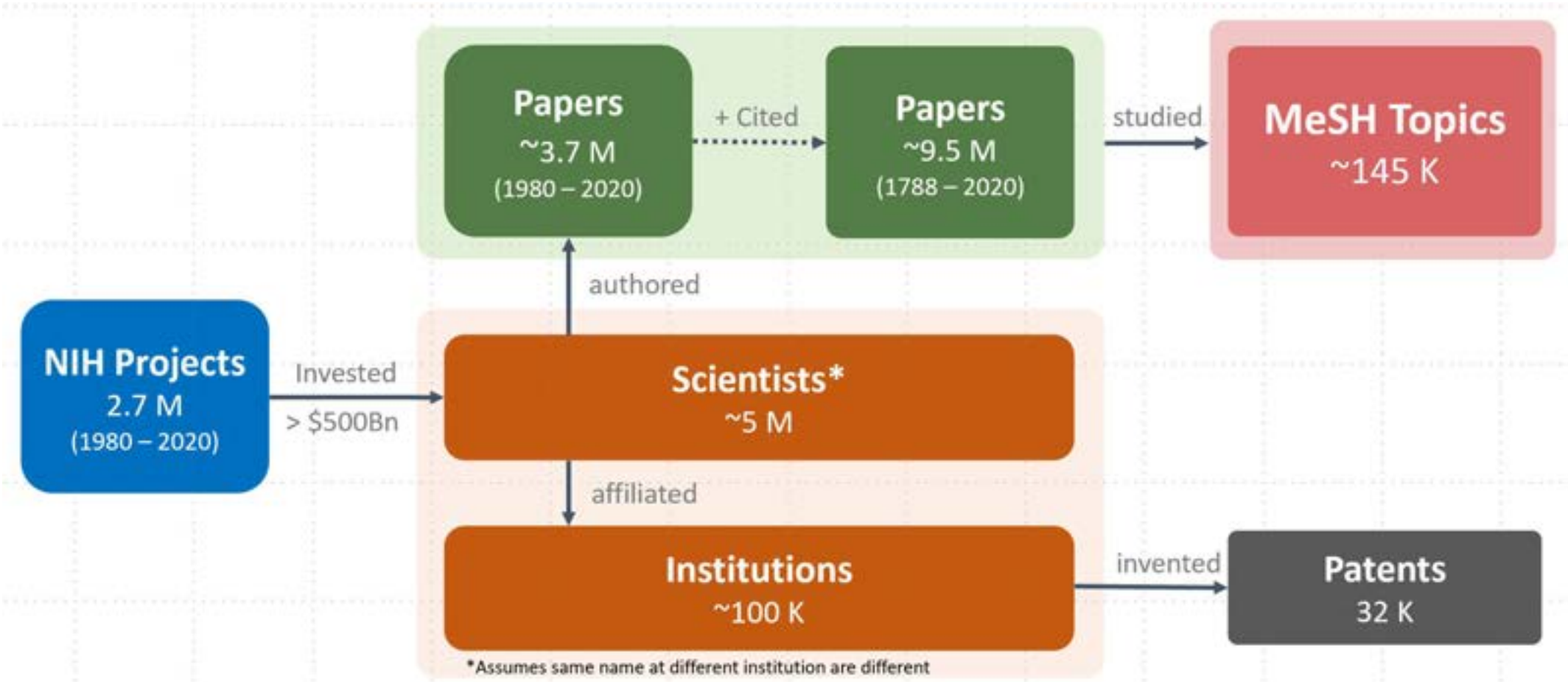
We focused on data cleansing, characterization, and centralization



Collect data to enable the platform realization.

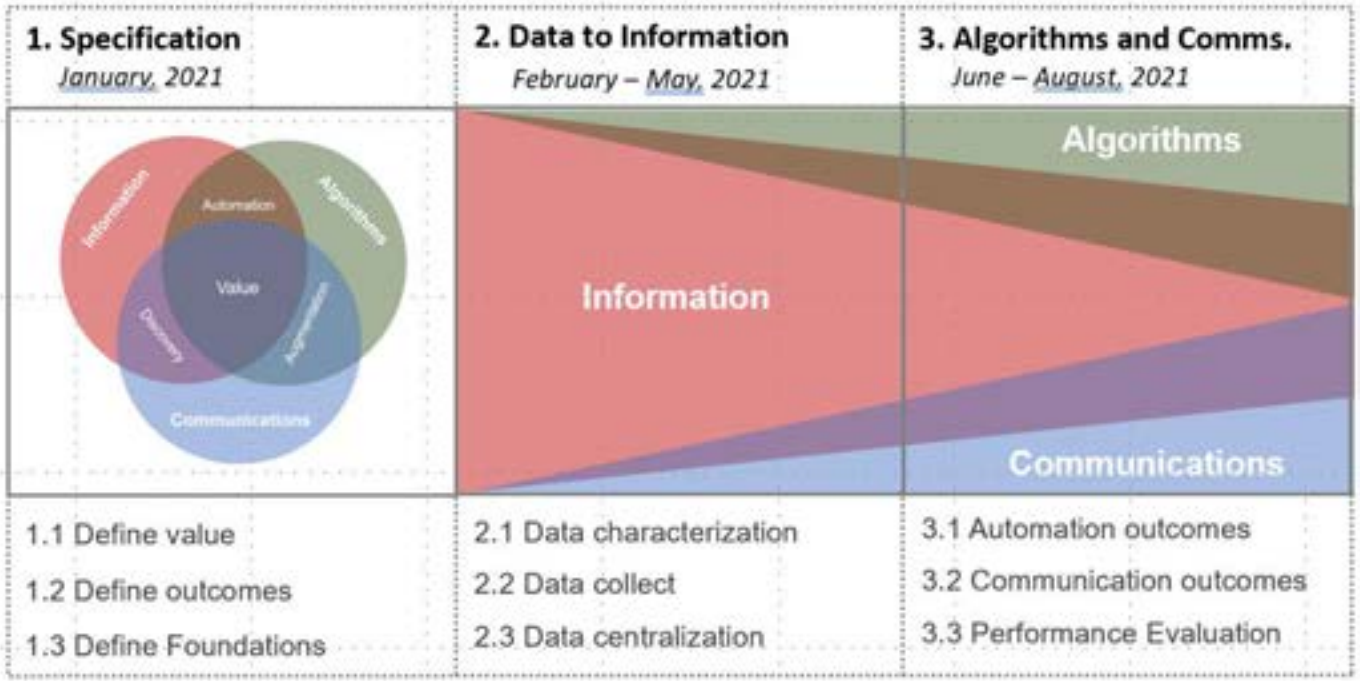
Combined data provides view of science in more complete context

Unites information on grants, papers, topics, authors, institutions, and patents



Now we are developing tools to analyze and visualize the data

Specifically, we are focusing on NLP processing of documents, and dynamic networks



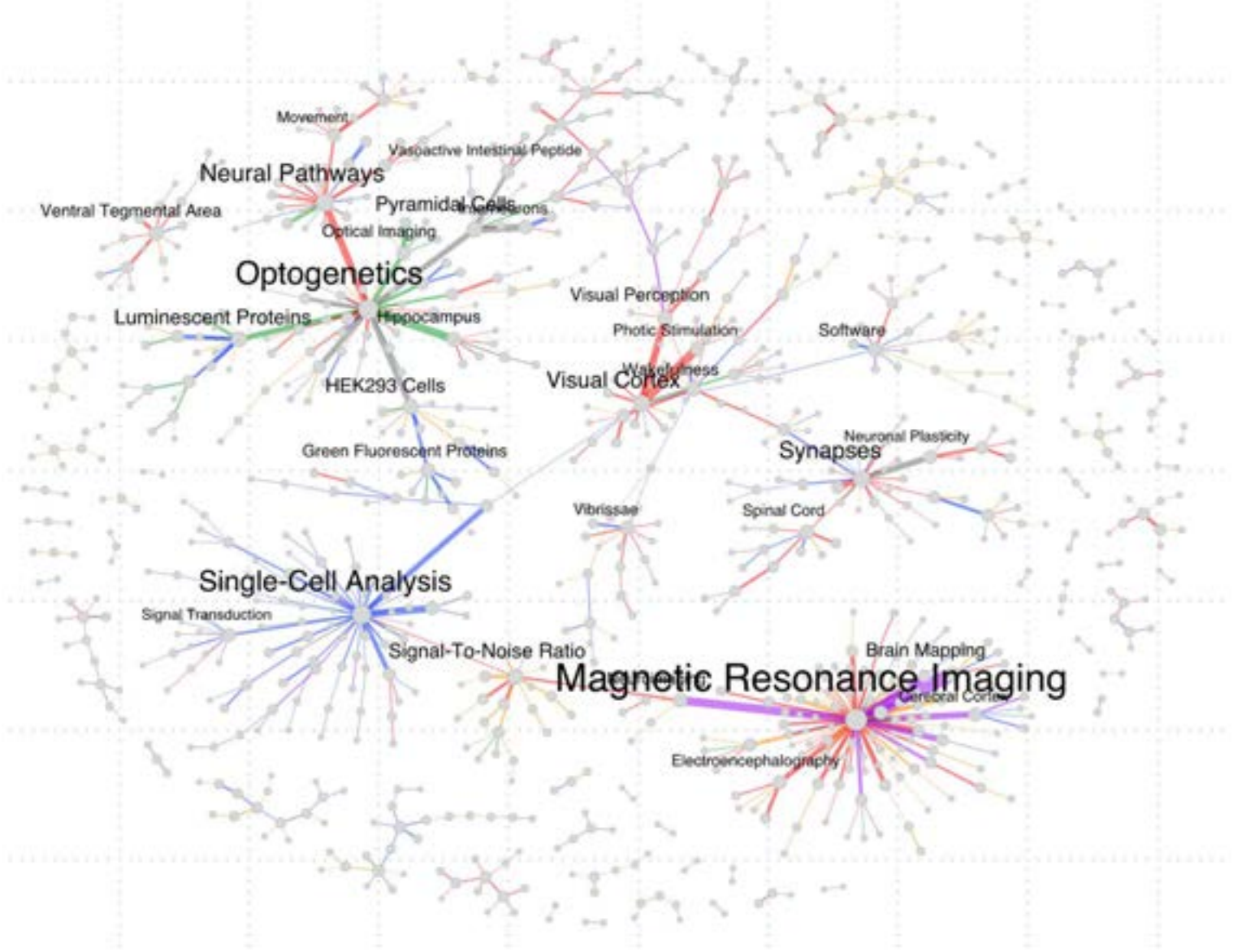
Generate Dynamic Networks
using the collected data

Publication topic graph of BRAIN awardees (2014 – 2020)

Nodes: topics in papers
- Size: number of papers published

Edges: topic co-occurrence in papers
- Color: NIH BRAIN Team
- Gray: increased inter-group output

Exclusion Criteria:
- Edges**:** only rank 1 edges are shown.

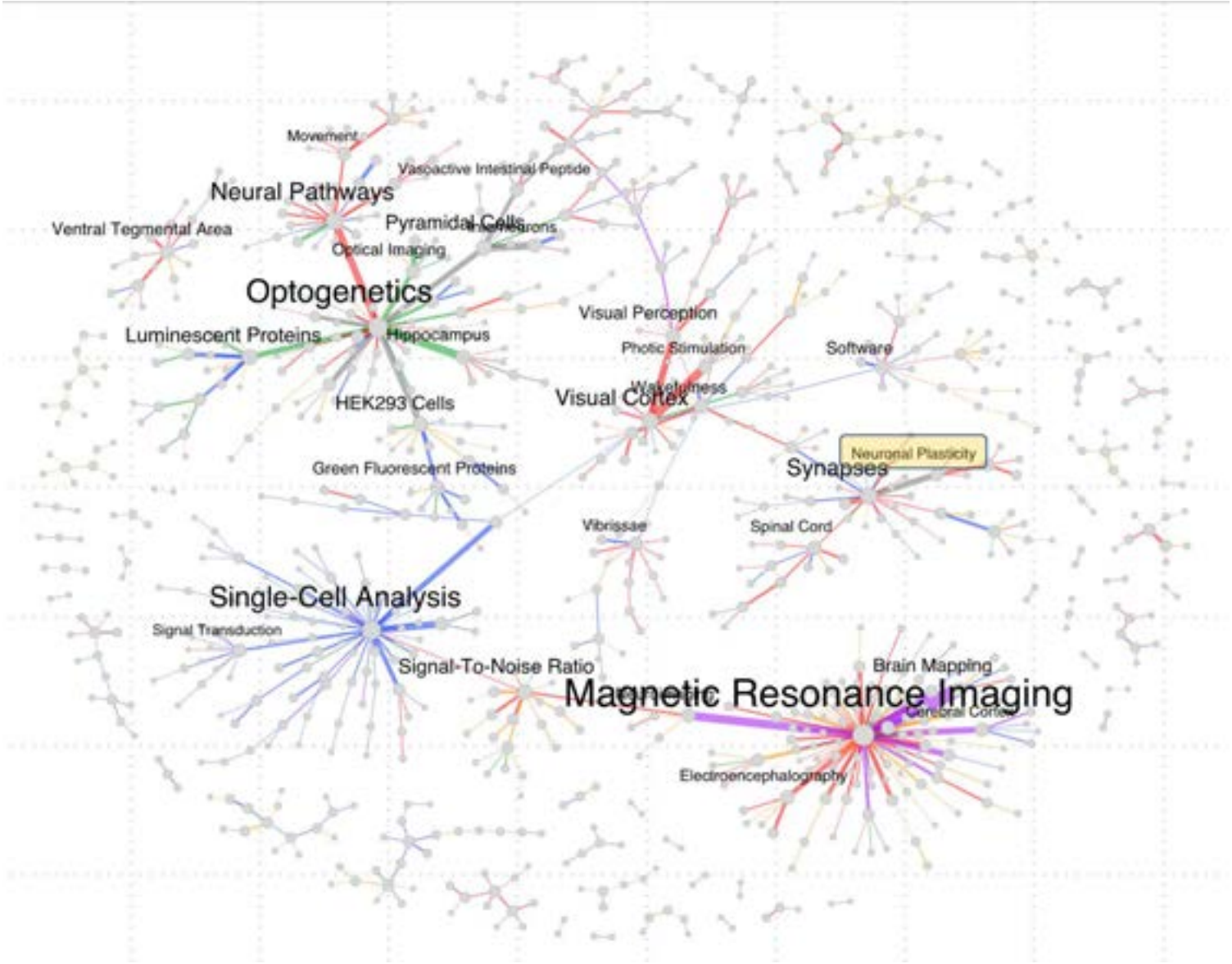


Publication topic graph of BRAIN awardees (2014 – 2020)

Nodes: topics in papers
- Size: number of papers published

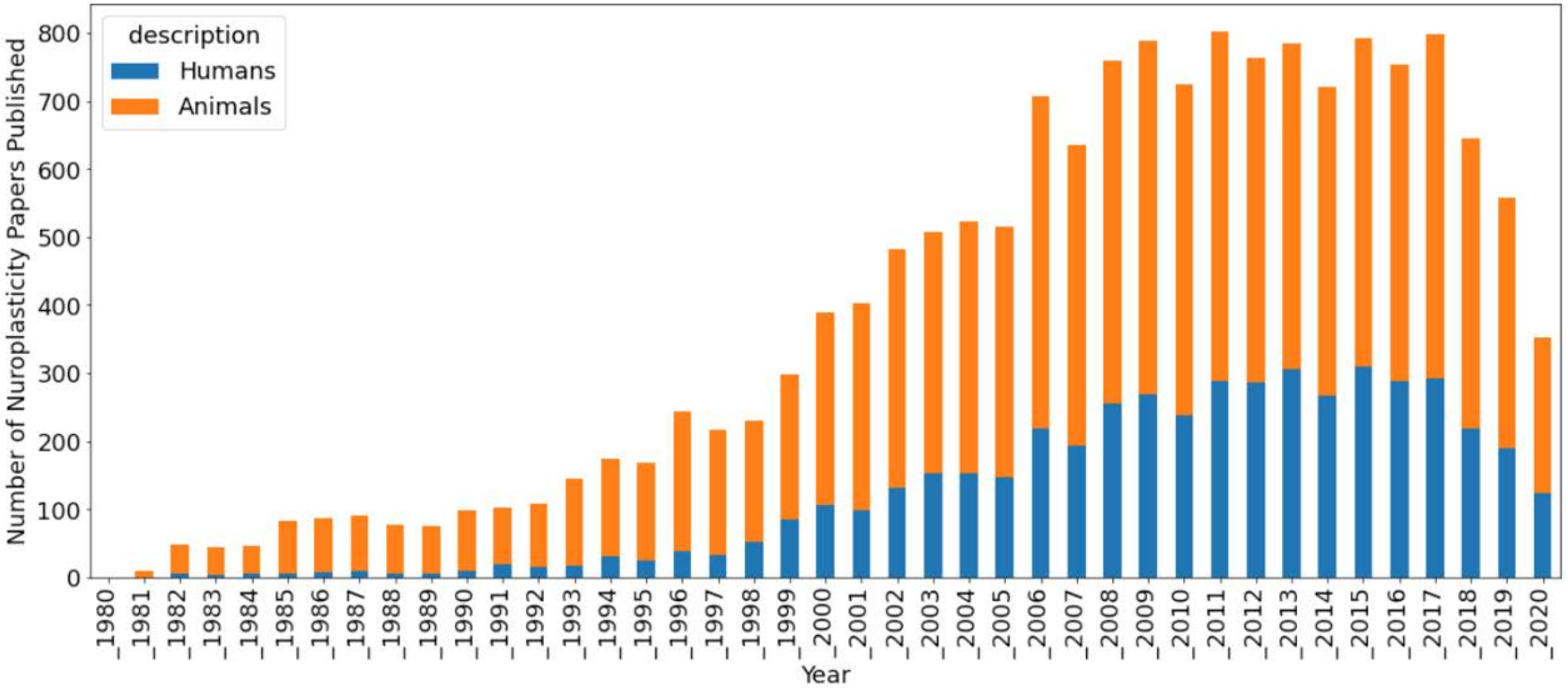
Edges: topic co-occurrence in papers
- Color: NIH BRAIN Team
- Gray: increased inter-group output

Exclusion Criteria:
- Edges**:** only rank 1 edges are shown.



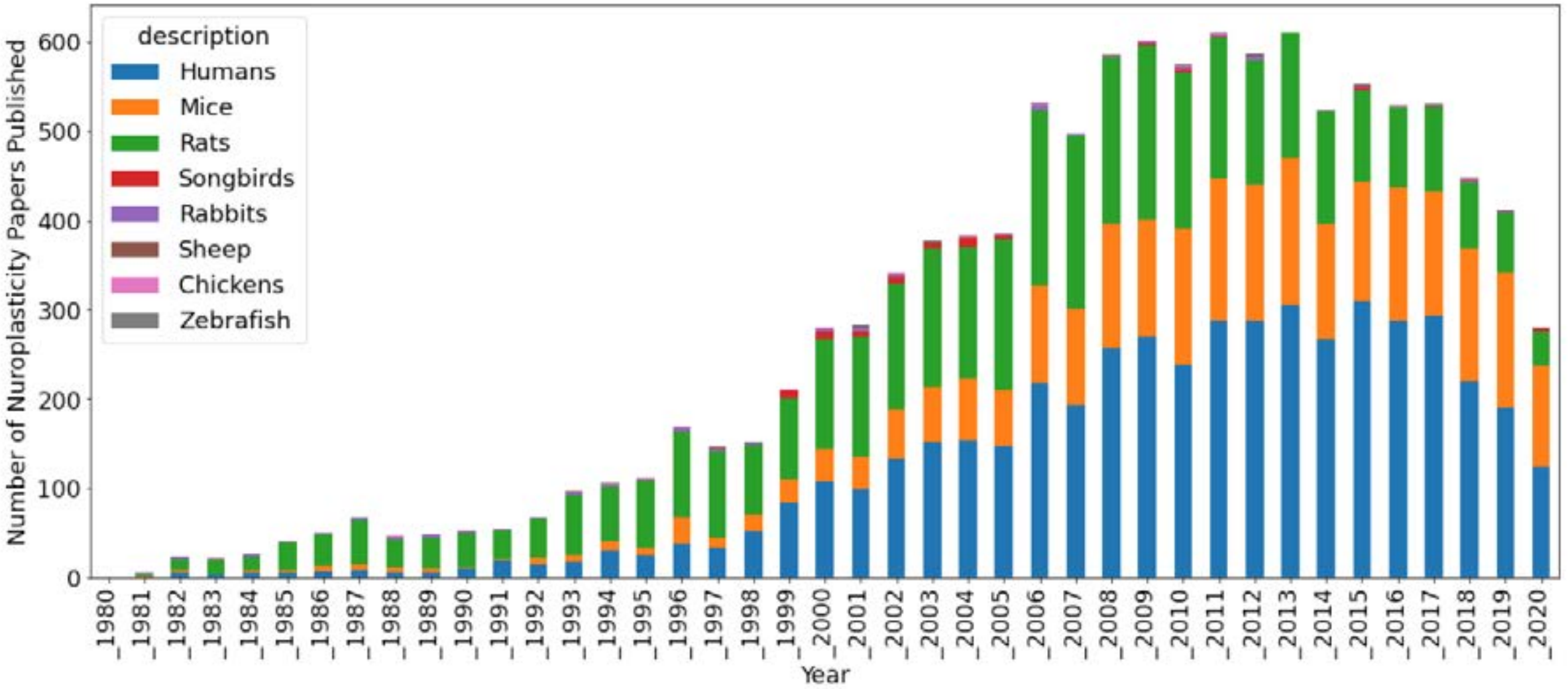
Co-occurrence of topics in publications provides research trends

Trends in ‘Neuroplasticity’ research may be understood by topics that covary with it



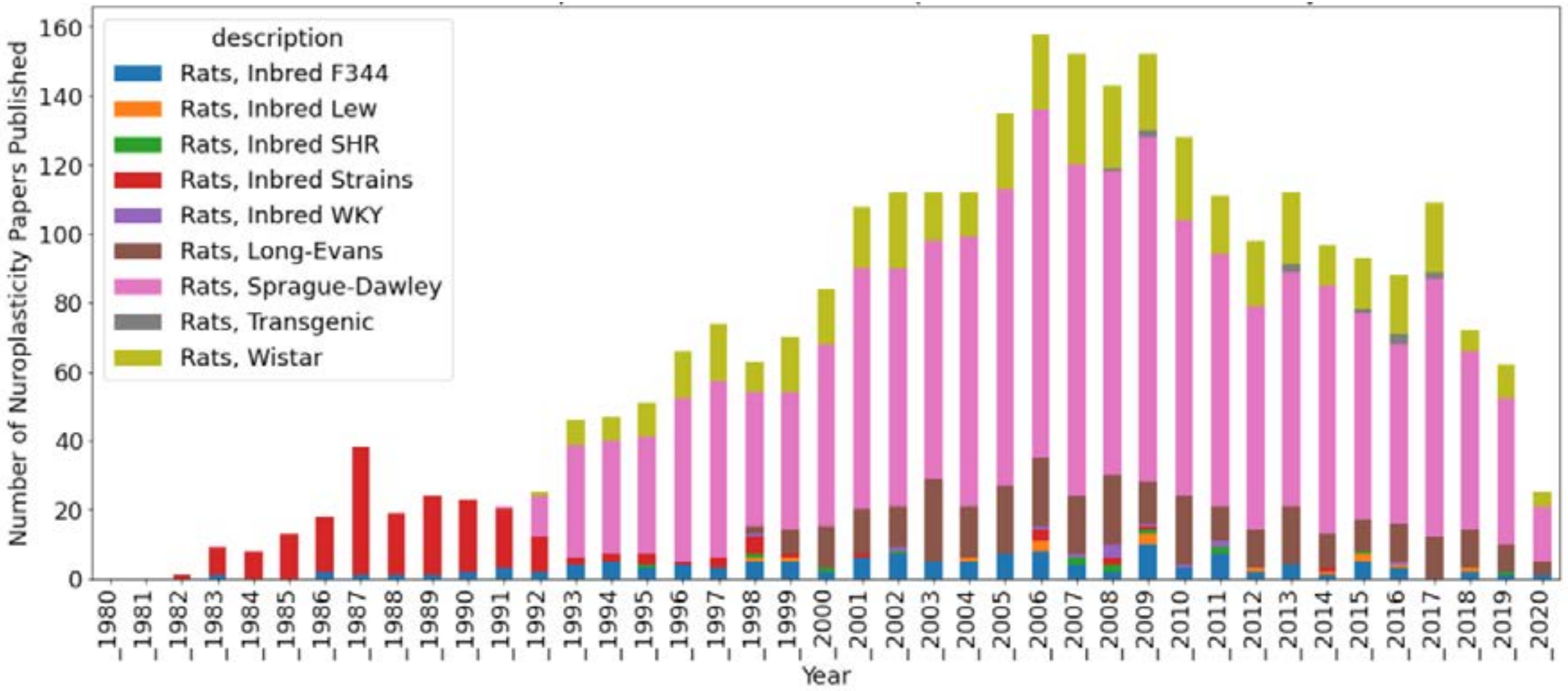
Co-occurrence of topics in publications provides research trends

Trends in ‘Neuroplasticity’ research may be understood by topics that covary with it



Co-occurrence of topics in publications provides research trends

Trends in ‘Neuroplasticity’ research may be understood by topics that covary with it

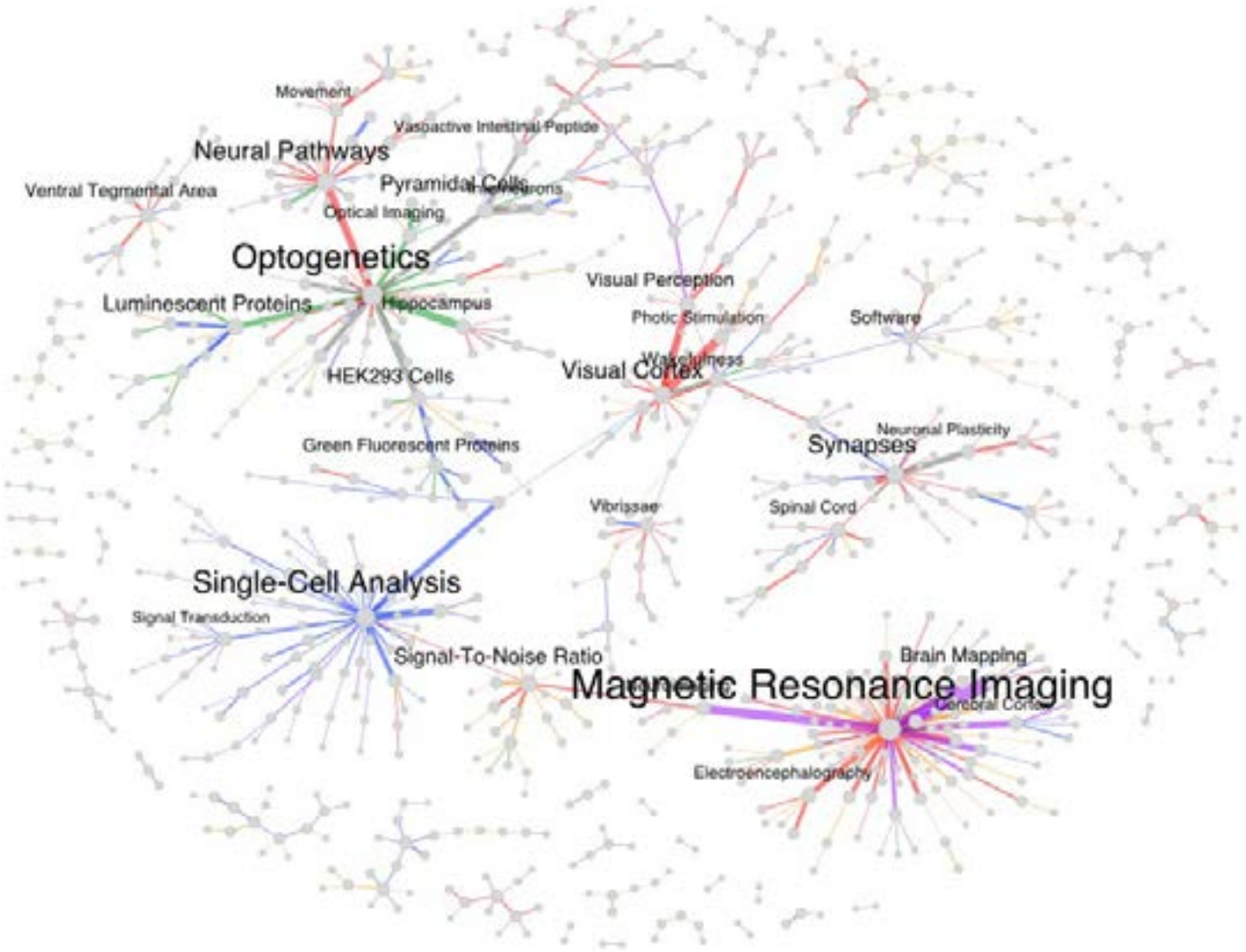


Publication topic graph of BRAIN awardees (2014 – 2020)

Nodes: topics in papers
- Size: number of papers published

Edges: topic co-occurrence in papers
- Color: NIH BRAIN Team
- Gray: increased inter-group output

Exclusion Criteria:
- Edges**:** only rank 1 edges are shown.



How do collaborative projects impact author networks and paper

- Retrospective population:
  - ~1,500 Researchers affiliated with ~60 lab:
  - Intervention: ~30 labs received an award in 2017
  - Control: ~30 labs received an award in 2019 or later

- Study period:
  - 2015 – 2016: 2 years before the intervention group received their award
  - 2017 – 2018: 2 years following the intervention group receiving their award

- Characteristics of interest
  - *Changes* in publication output and collaborations

Changes in Neuroscience collaborations (2015 – 2019)

Control Group

Nodes: individual authors

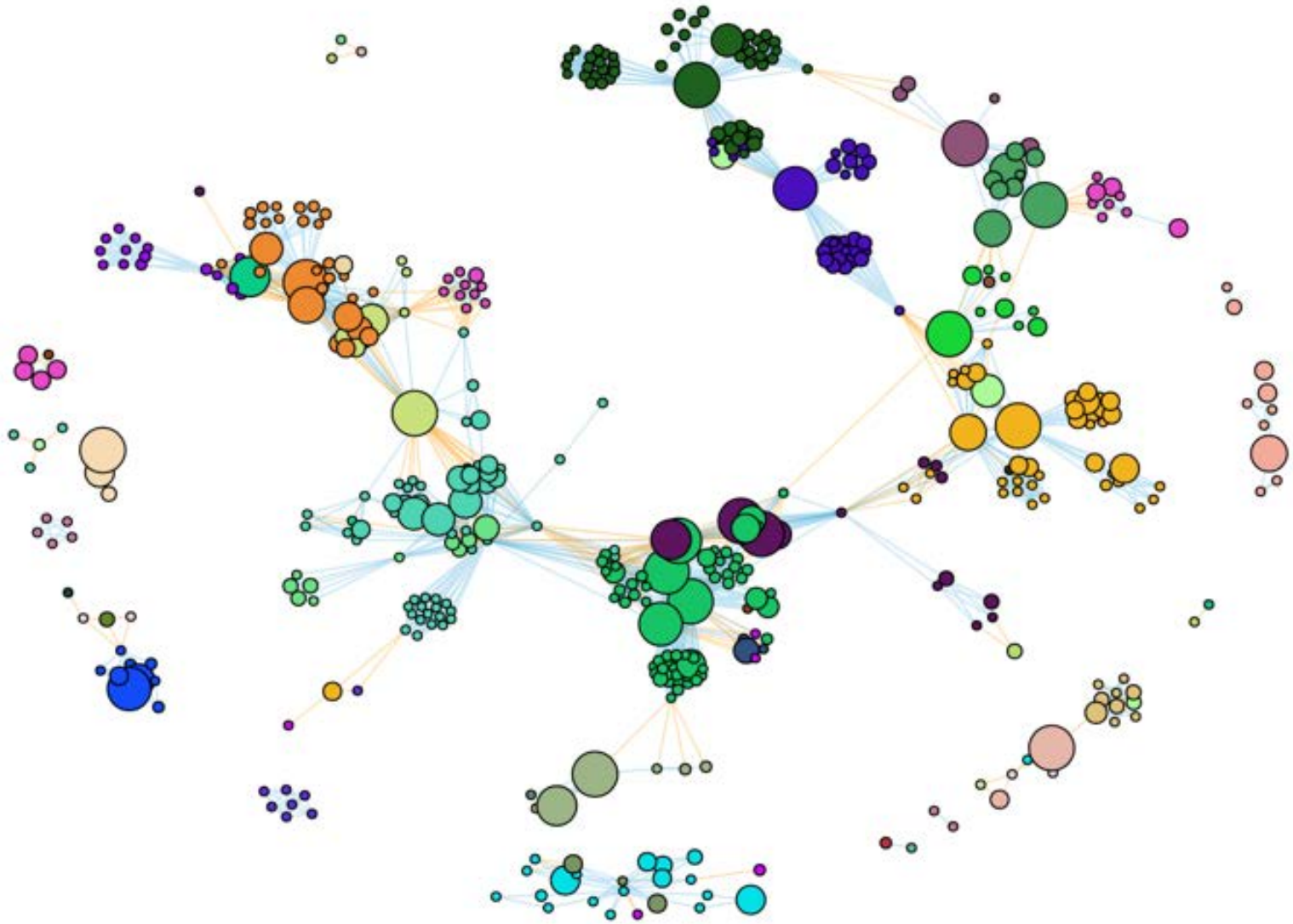
- Color: investigator group
- Size: increase in publication output from [2016/2017] to [2018/2019]

Edges: author collaboration

- Orange: increased intra-group output
- Blue: increased inter-group output

Exclusion Criteria:

- Nodes: authors with decreased or stable collaborations not shown.
- Edges: reductions in collaborations not shown.



Changes in Neuroscience collaborations (2015 – 2019)

Control Group

Nodes: individual authors

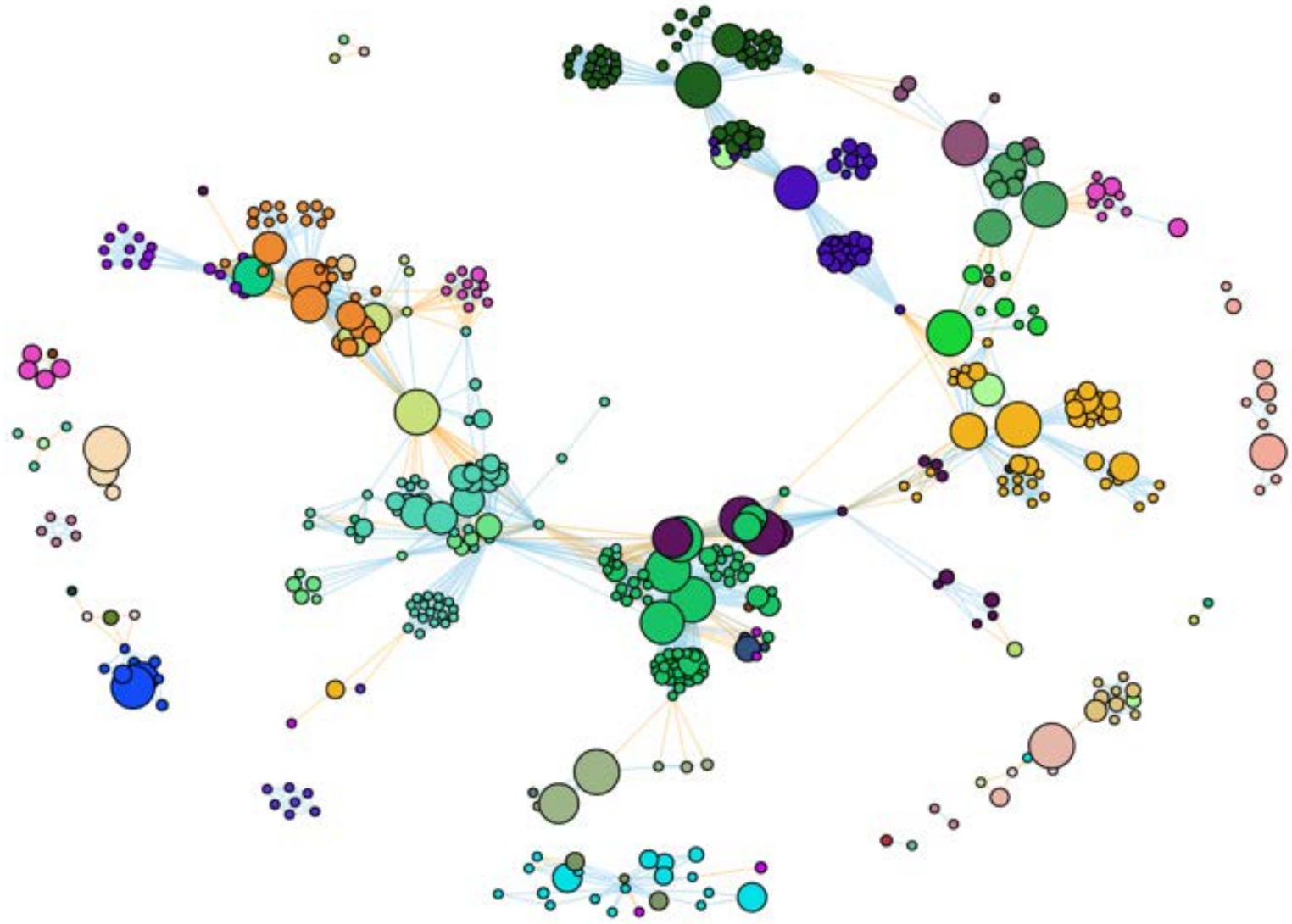
- Color: investigator group
- Size: increase in publication output from [2016/2017] to [2018/2019]

Edges: author collaboration

- Orange: increased intra-group output
- Blue: increased inter-group output

Exclusion Criteria:

- Nodes: authors with decreased or stable collaborations not shown.
- Edges: reductions in collaborations not shown.



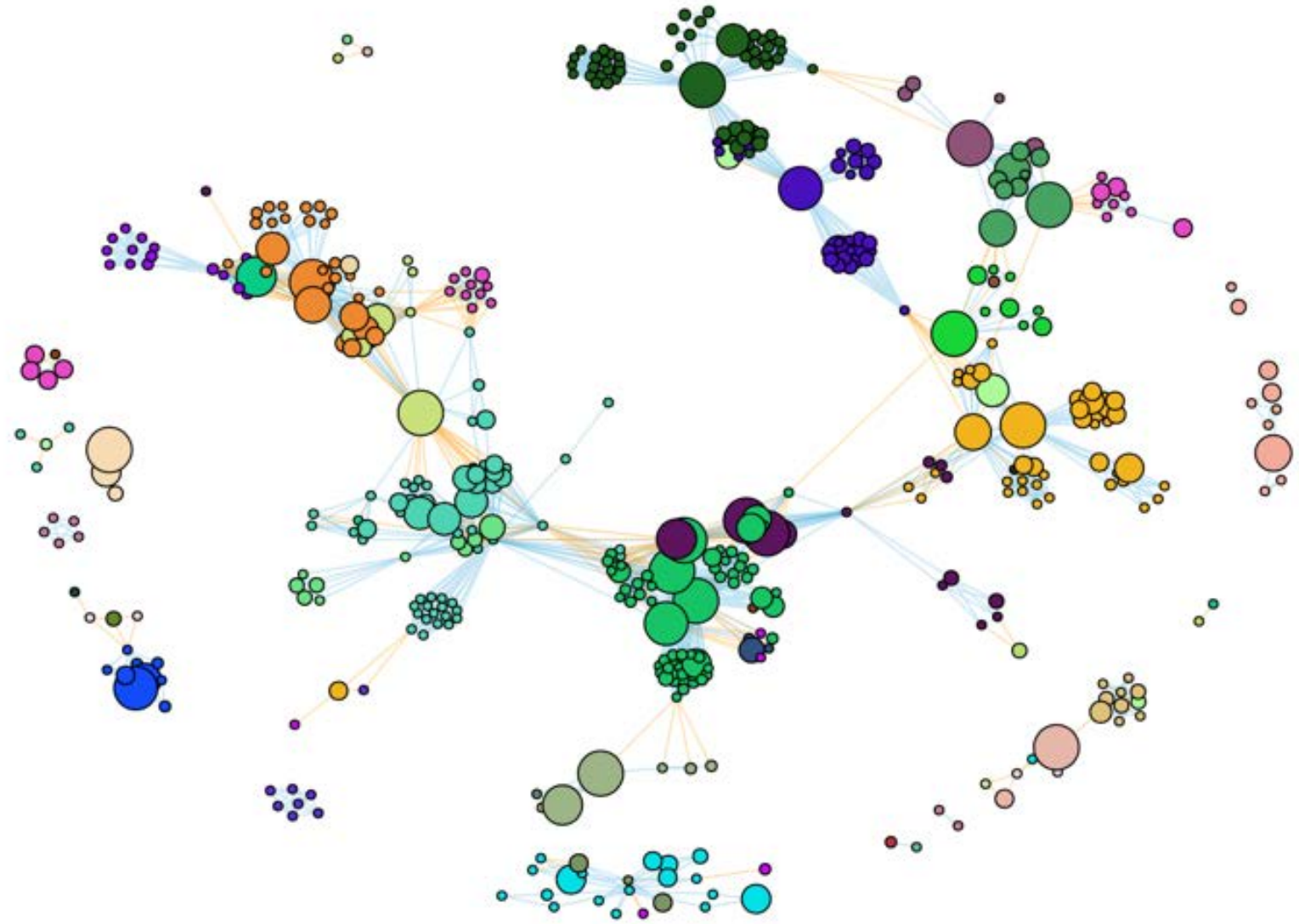
Changes in Neuroscience collaborations (2015 – 2019)

Control Group

Nodes: individual authors
- Color: investigator group
- Size: increase in publication output from [2016/2017] to [2018/2019]

Edges: author collaboration
- Orange: increased intra-group output
- Blue: increased inter-group output

Exclusion Criteria:
- Nodes: authors with decreased or stable collaborations not shown.
- Edges: reductions in collaborations not shown.



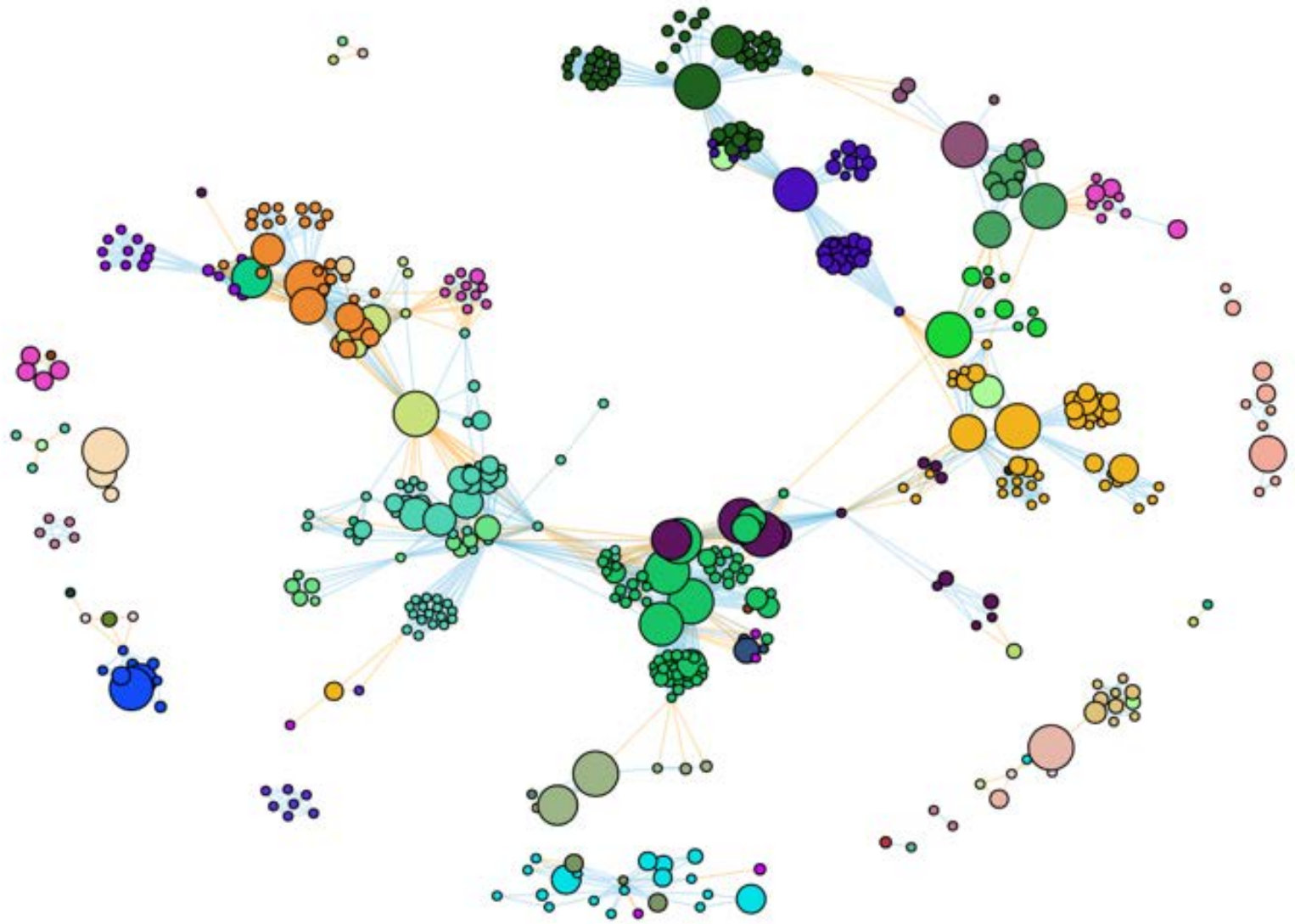
Changes in Neuroscience collaborations (2015 – 2019)

Control Group

Nodes: individual authors
- Color: investigator group
- Size: increase in publication output from [2016/2017] to [2018/2019]

Edges: author collaboration
- Orange: increased intra-group output
- Blue: increased inter-group output

Exclusion Criteria:
- Nodes: authors with decreased or stable collaborations not shown.
- Edges: reductions in collaborations not shown.



Changes in Neuroscience collaborations (2015 – 2019)

Intervention Group

Nodes: individual authors

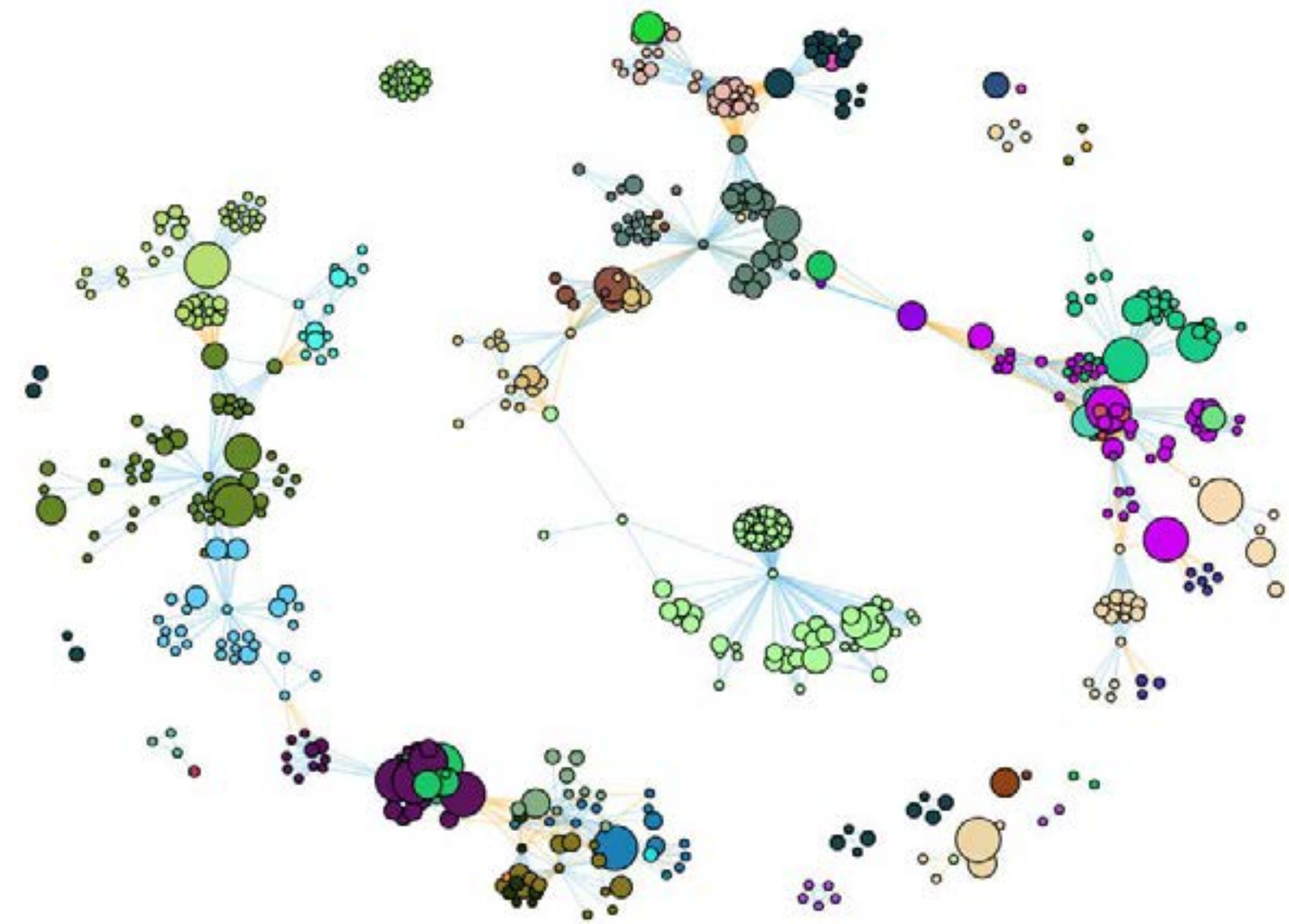
- Color: investigator group
- Size: increase in publication output from [2016/2017] to [2018/2019]

Edges: author collaboration

- Orange: increased intra-group output
- Blue: increased inter-group output

Exclusion Criteria:

- Nodes: authors with decreased or stable collaborations not shown.
- Edges: reductions in collaborations not shown.



Changes in Neuroscience collaborations (2015 – 2019)

Intervention Group

Nodes: individual authors

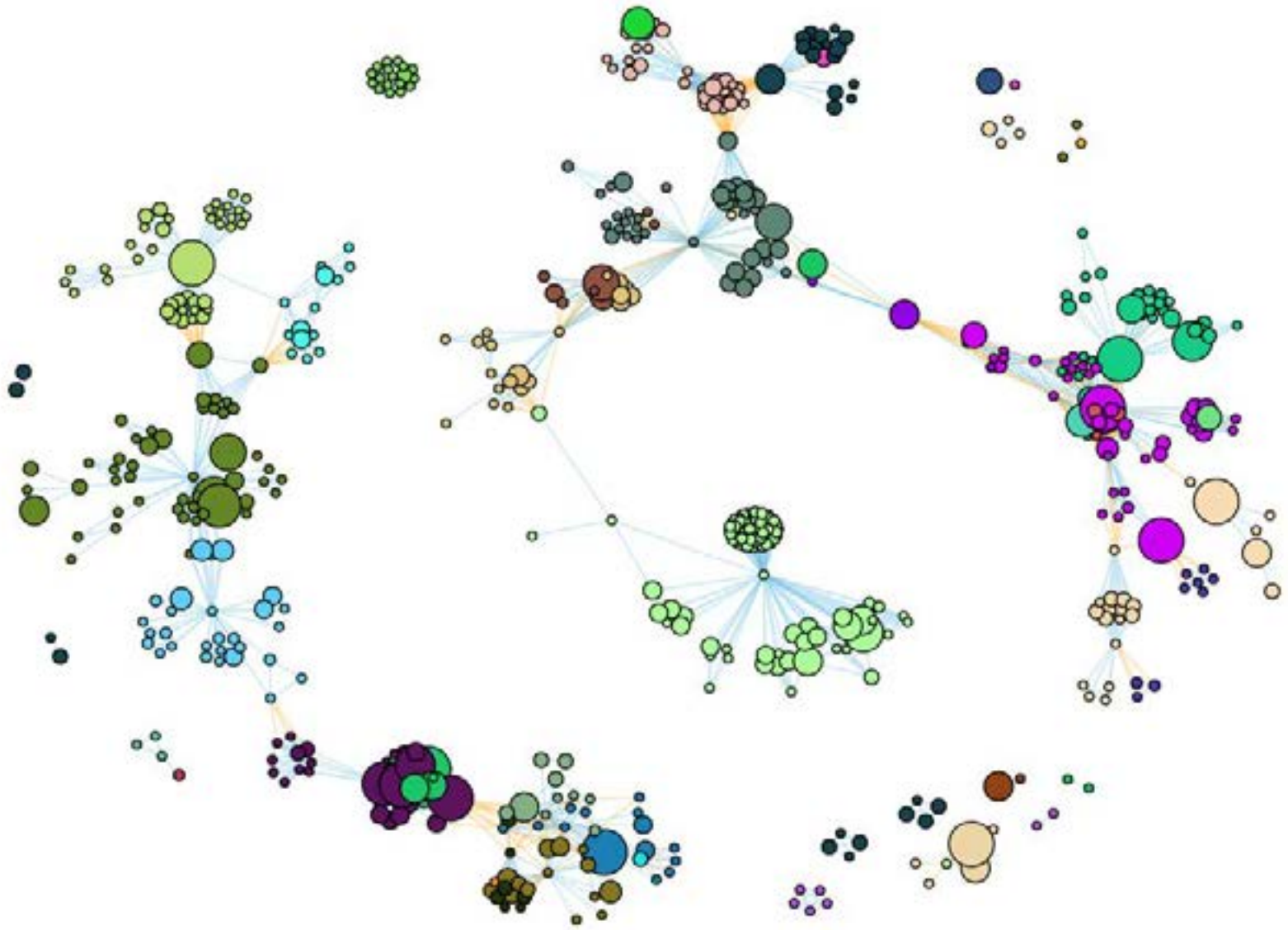
- Color: investigator group
- Size: increase in publication output from [2016/2017] to [2018/2019]

Edges: author collaboration

- Orange: increased intra-group output
- Blue: increased inter-group output

Exclusion Criteria:

- Nodes: authors with decreased or stable collaborations not shown.
- Edges: reductions in collaborations not shown.



Changes in Neuroscience collaborations (2015 – 2019)

Intervention Group

Nodes: individual authors

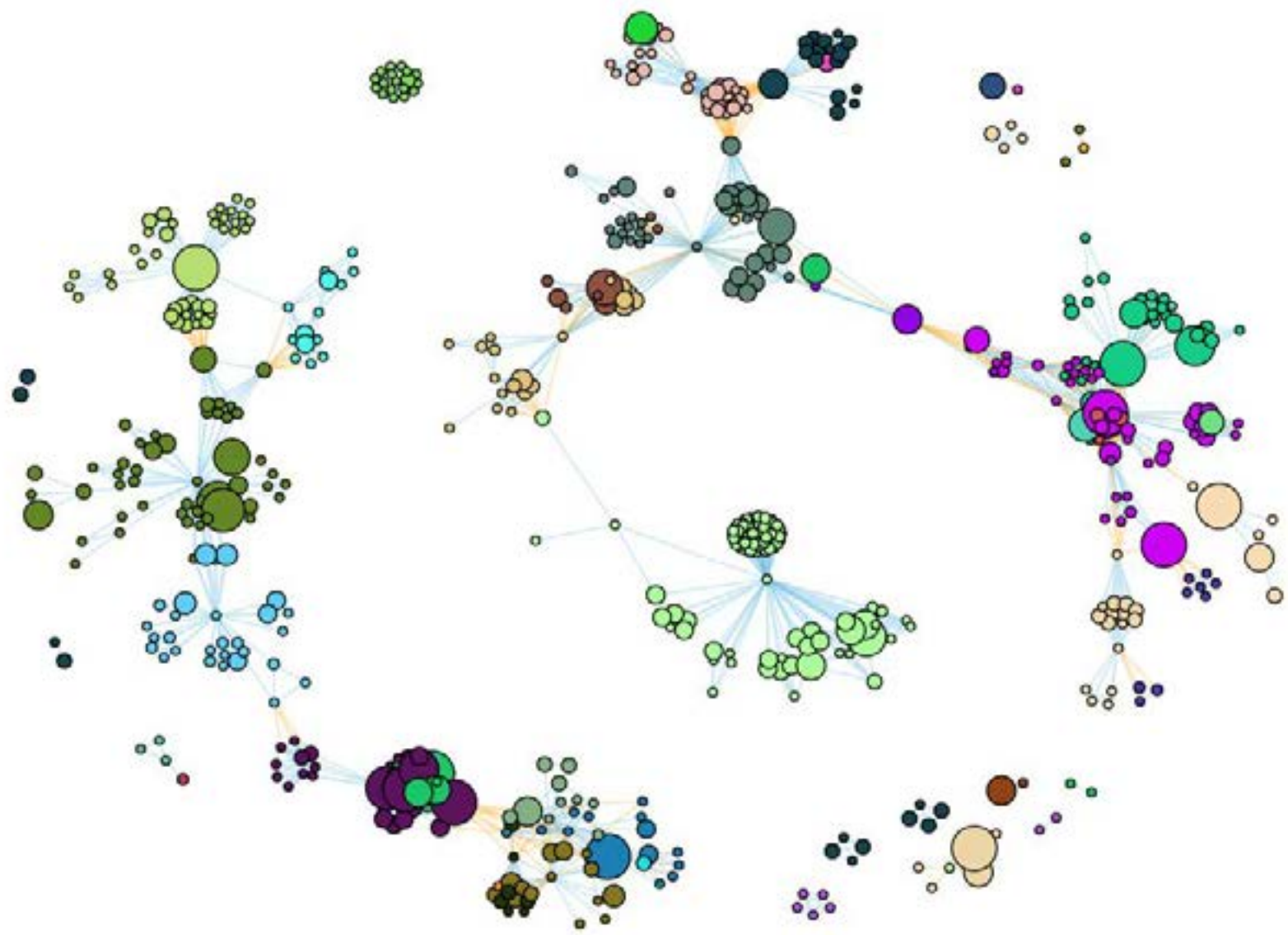
- Color: investigator group
- Size: increase in publication output from [2016/2017] to [2018/2019]

Edges: author collaboration

- Orange: increased intra-group output
- Blue: increased inter-group output

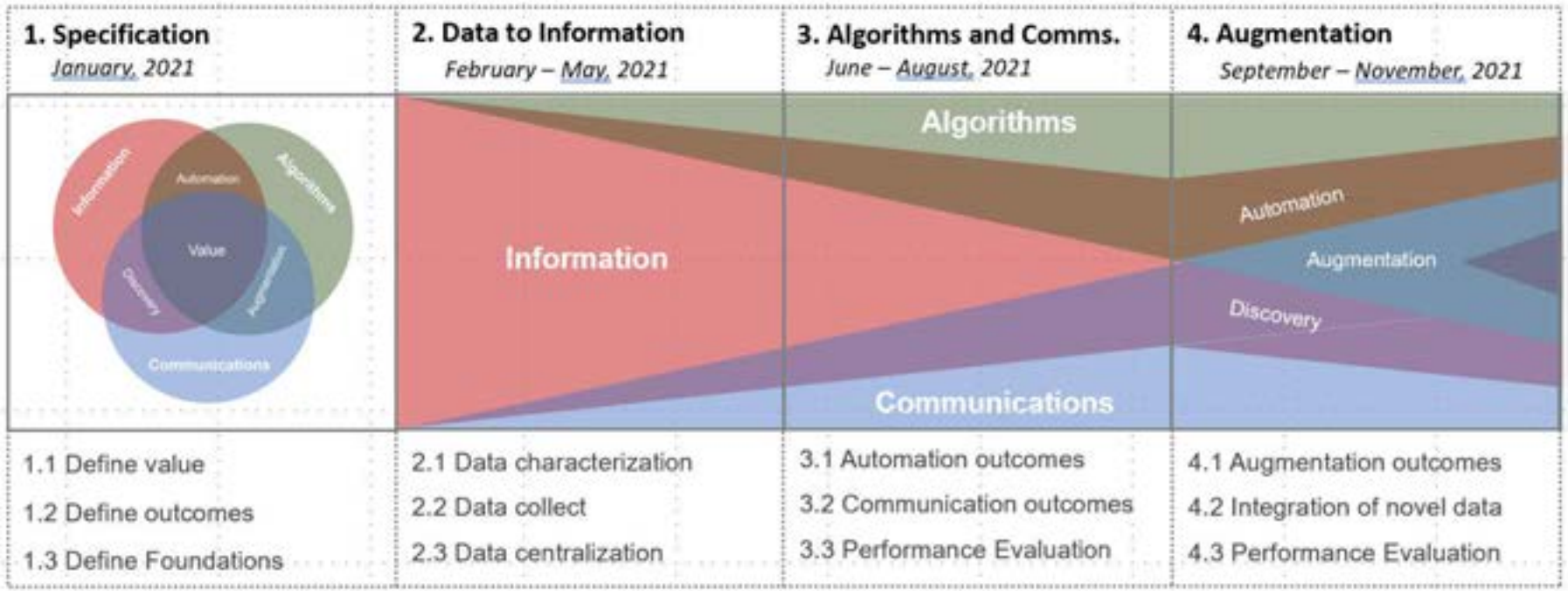
Exclusion Criteria:

- Nodes: authors with decreased or stable collaborations not shown.
- Edges: reductions in collaborations not shown.



Next steps are to develop user interface to allow data exploration

Specifically, basic search functionality, graph representation, and group comparison



Contact Information

Mohammad Ghassemi, PhD
mohammad.ghassemi@nih.gov