Some Assembly Required: Organizing in the 21st Century

Noshir Contractor
Jane S. & William J. White Professor of Behavioral Sciences
Northwestern University

Twitter: @noshir

Supported by following National Science Foundation grants:
IIS-0838564, IIS-0841583, BCS- 0940851, OCI-0904356, CNS-1010904
Army Research Institute W91WAW-08-C-0106/P00004**;**
Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053
& NIH/NCRR & NIH/NCRR grant to Northwestern University
Clinical and Translational Sciences Institute (NUCATS)

NORTHWESTERN UNIVERSITY

SONIC
advancing the science of networks in communities

Building the Team That Built Watson

Ozier Muhammad/The New York Times

David Ferrucci led the team behind Watson, the victorious "Jeopardy" computer. "For the scientist in me," he says, "it was an irresistible challenge.".

By DAVID A. FERRUCCI
Published: January 7, 2012

THE assignment was one of the biggest challenges in the field of artificial intelligence: build a computer smart enough to beat grand champions at the game of "Jeopardy."

Related

Smarter Than You Think: What Is I.B.M.'s Watson? (June 20, 2010)

Computer Wins on 'Jeopardy!': Trivial, It's Not (February 17, 2011)

When I stepped up to lead the team at I.B.M. that would create this computer, called Watson, I knew the task would be formidable. The computer would have to answer an unpredictable variety of complex questions with confidence, precision and speed. And we would put it to the test in a publicly televised "human versus machine" competition against the best players of all time.

It was not easy finding people to join the Watson team in the mid-1990s. Most scientists I approached favored their own individual projects and career tracks. And who could blame them? This was an effort that, at best, would mingle the contributions of many. At its worst it would fail miserably, undermining the credibility of all involved.

David Ferucci,
New York Times
1/7/2012

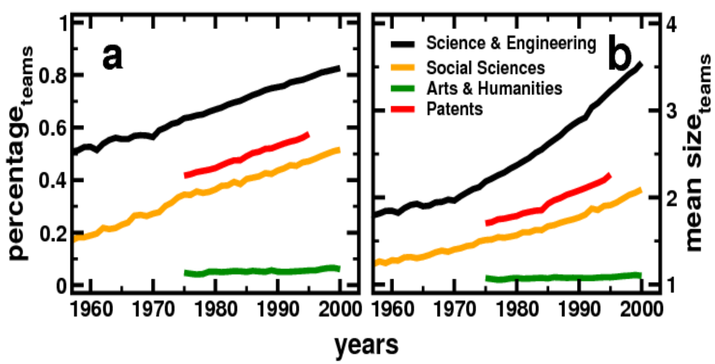
Move to Team Science

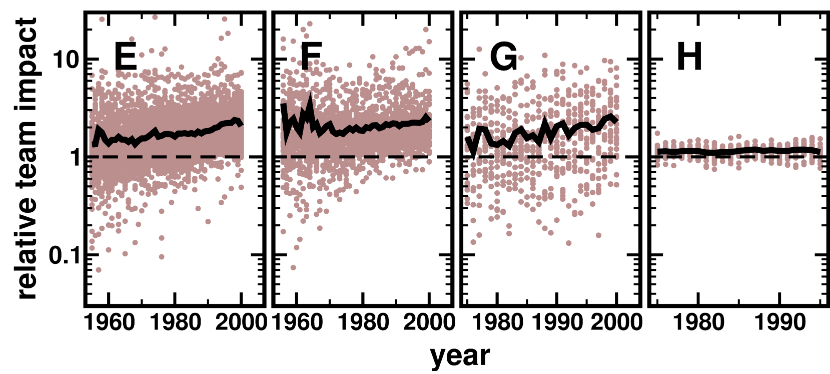
Studies of 19.9 million research articles over 5 decades as recorded in the Web of Science database, and an additional 2.1 million patent records from 1975-2005 found three important facts.

1. For virtually all fields, research is increasingly done in teams

2. Teams typically produce more highly cited research than individuals do (accounting for self-citations), and this team advantage is increasing over time.

3. .Teams now produce the exceptionally high impact research, even where that distinction was once the domain of solo authors.





NORTHWESTERN UNIVERSITY Sources: Wuchty, Jones, and Uzzi, 2007a, 2007b

SONIC
advancing the science of networks in communities

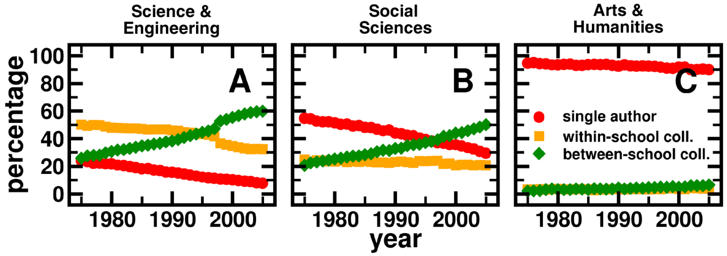
Move to *Virtual* Team Science

The trend toward virtual communities was ***not*** driven by a growth in teamwork by scientists working with other co-located scientists. Using the Web of Science database to analyze the collaboration arrangements of over 4,000,000 papers over a 30 year period, they found that:

1. Team science is increasingly composed of co-authors located at different universities.

2. These “virtual communities of scholars” produce higher impact work than comparable co-located teams or solo scientists.

3. This change is true for all fields and team sizes, as well as for research done at elite universities



NORTHWESTERN UNIVERSITY

Source: Jones, Wuchty, Uzzi, 2008

SONIC
advancing the science of networks in communities

Key Takeaways

- Understanding and enabling team assembly is well poised to make an intellectual leap by leveraging recent advances in:
  - Theories: Theories about the socio-technical motivations for creating, maintaining, dissolving and re-creating links to engage in team assembly
  - Data: Developments in Semantic Web/Web 2.0 provide the technological capability to capture, store , merge, and query relational metadata needed to more effectively understand and enable team assembly.
  - Methods: An ensemble of qualitative and quantitative methods techniques (such as exponential random graph modeling or *p**) to understand and enable theoretically grounded network recommendations for team assembly
  - Computational infrastructure: Cloud computing and petascale applications are critical to face the computational challenges in understanding and enabling team assembly.

Multi-theoretical Multilevel (MTML) Motivations for Team Assembly

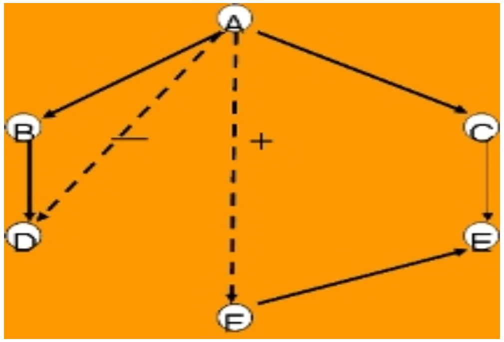
- Theories of self-interest
- Theories of social and resource exchange
- Theories of mutual interest and collective action
- Theories of contagion
- Theories of balance
- Theories of homophily
- Theories of proximity

Sources:

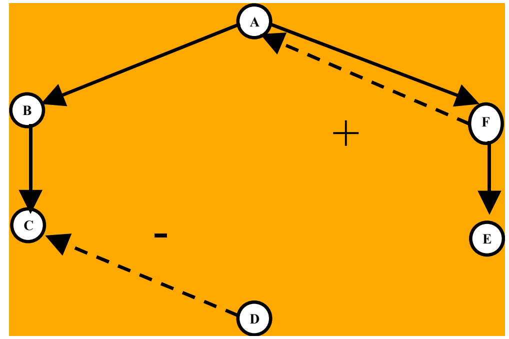
Contractor, N. S., Wasserman, S. & Faust, K. (2006). Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review*.

Monge, P. R. & Contractor, N. S. (2003). *Theories of Communication Networks*. New York: Oxford University Press.

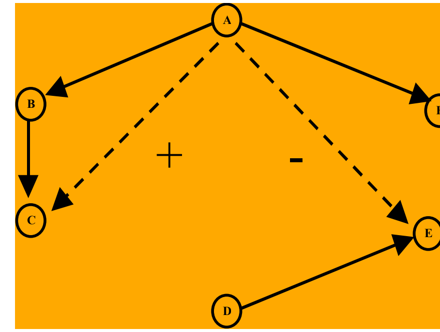
“Structural signatures” of MTML Motivations for Team Assembly



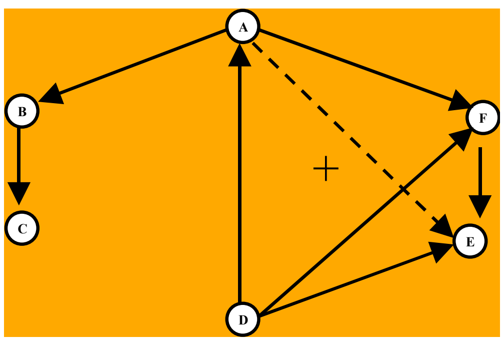
Theories of Self interest



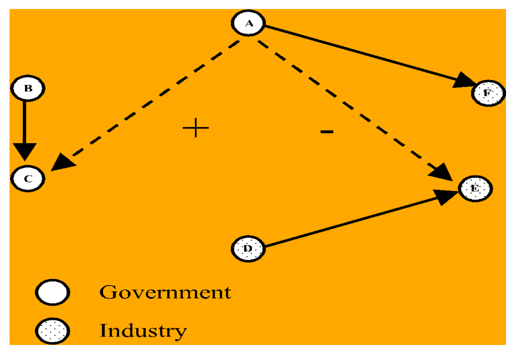
Theories of Exchange



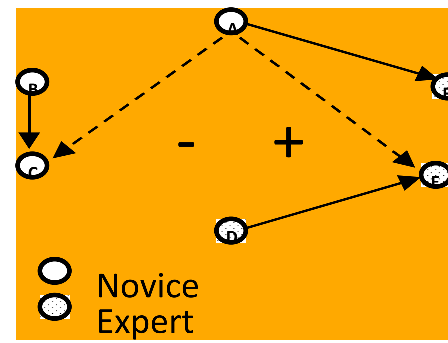
Theories of Balance



Theories of Collective Action



Theories of Homophily



Theories of Cognition

Statistical "MRI" for Structural Signatures

- p*/ERGM: Exponential Random Graph Models
- Statistical "Macro-scope" to detect structural motifs in observed networks
- Move from exploratory to confirmatory network analysis to understand multi-theoretical multilevel motivations for why we create social and information networks

Challenges of <u>empirically</u> testing, extending, and exploring theories about team assembly …

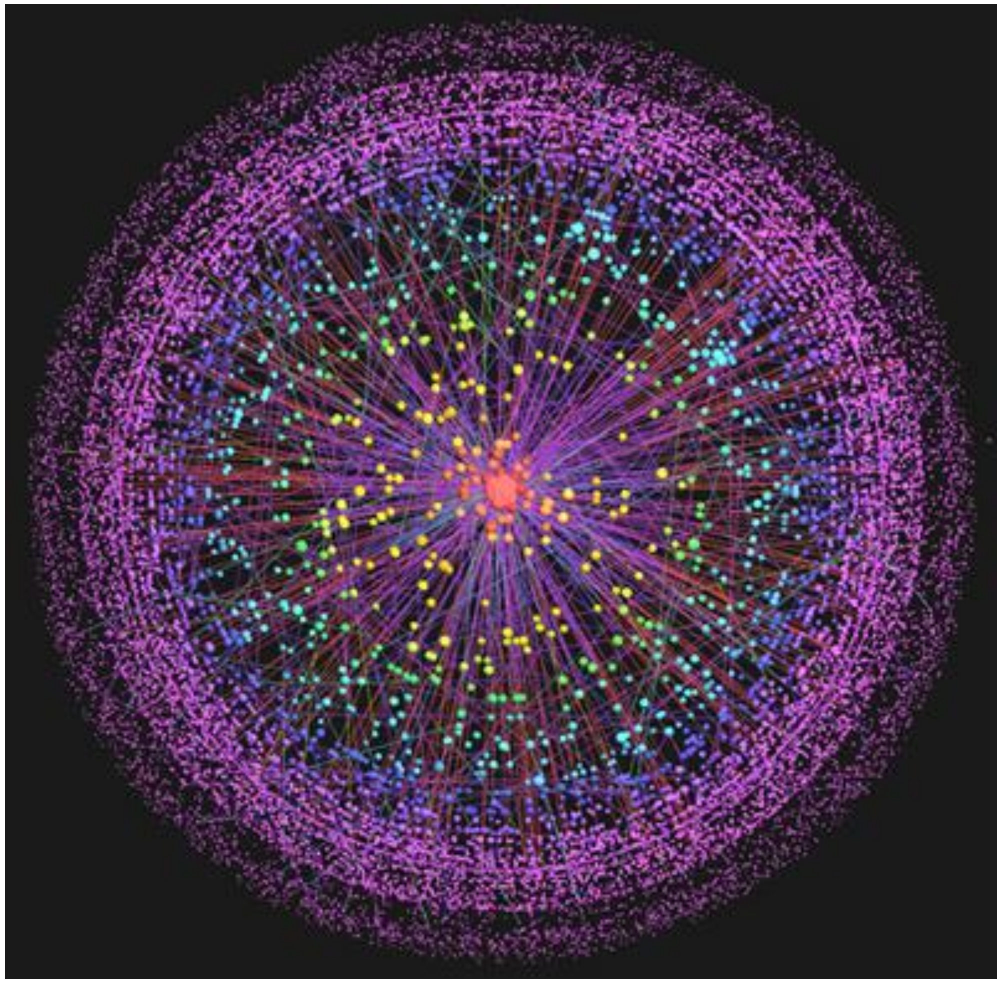
The Hubble telescope: $2.5 billion

Source: David Lazer

CERN particle accelerator: $1 billion/year

Source: David Lazer

The Web: priceless*



Source: David Lazer

** Apologies to MasterCard*

Multidimensional Networks for Team Assembly
Multiple Types of Nodes and Multiple Types of Relationships

SOCIAL SCIENCE

Computational Social Science

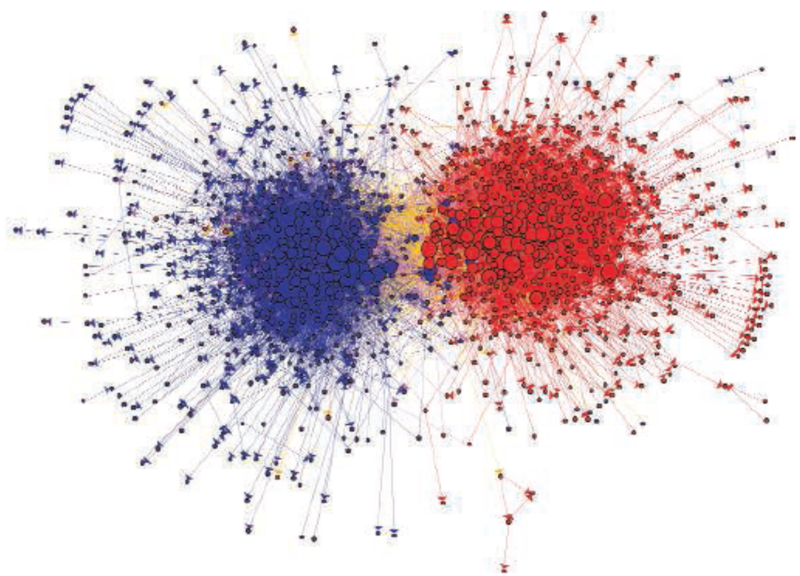
A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5] Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3] Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

The capacity to collect and analyze massive amounts of data has transformed such fields as biology and physics. But the emergence of a data-driven "computational social science" has been much slower. Leading journals in economics, sociology, and political science show little evidence of this field. But computational social science is occurring—in Internet companies such as Google and Yahoo, and in government agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of academic researchers presiding over private data from which they produce papers that cannot be critiqued or replicated. Neither scenario will serve the long-term public interest of accumulating, verifying, and disseminating knowledge.

What value might a computational social science—based in an open academic environment—offer society, by enhancing understanding of individuals and collectives? What are the



Data from the blogosphere. Shown is a link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it. [Reproduced from (8) with permission from the Association for Computing Machinery]

[1]Harvard University, Cambridge, MA, USA. [2]Massachusetts Institute of Technology, Cambridge, MA, USA. [3]University of Michigan, Ann Arbor, MI, USA. [4]New York University, New York, NY, USA. [5]Northeastern University, Boston, MA, USA. [6]Interdisciplinary Scientific Research, Seattle, WA, USA. [7]Northwestern University, Evanston, IL, USA. [8]University of California–San Diego, La Jolla, CA, USA. [9]Columbia University, New York, NY, USA [10]Cornell University, Ithaca, NY, USA. [11]Boston University, Boston, MA, USA. E-mail: david_lazer@harvard.edu. Complete affiliations are listed in the supporting online material.

Published by AAAS

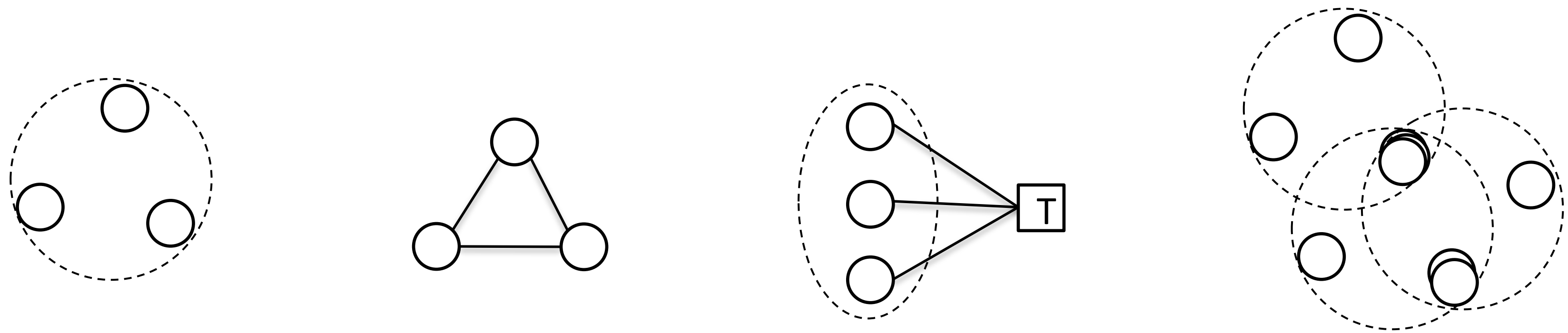
NORTHWESTERN UNIVERSITY

IC

advancing the science of networks in communities

Four Levels of Influences on Team Assembly

Compositional Level | Relational Level | Multimodal Network Level | Ecosystem Level



(a) Team as a collection of individuals

(b) Team as individuals and relations

(c) Team as a network of individuals and tasks

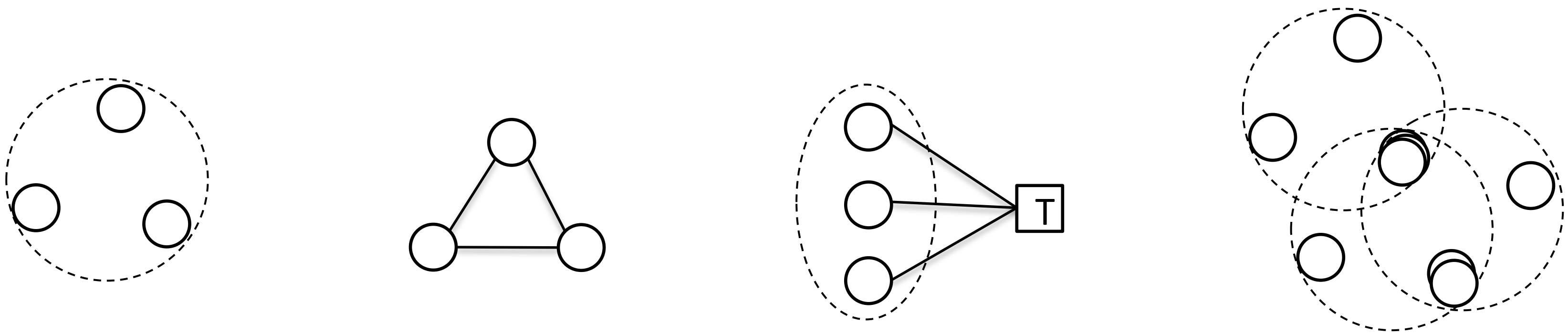
(d) Ecosystem of teams

○ Individual

T Task

Four Levels of Influences on nanoHUB Team Assembly

Compositional Level | Relational Level | Multimodal Network Level | Ecosystem Level



(a) Team as a collection of individuals

(b) Team as individuals and relations

(c) Team as a network of individuals and tasks

(d) Ecosystem of teams

Individual

Task

Compositional Influences on nanoHUB Team Assembly

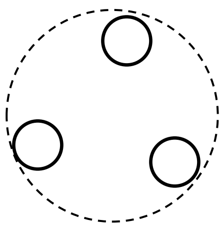
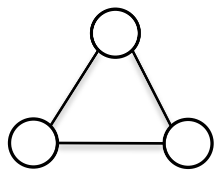
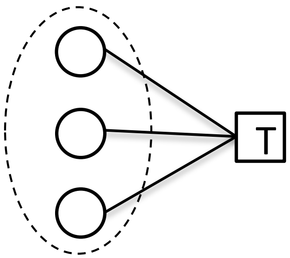
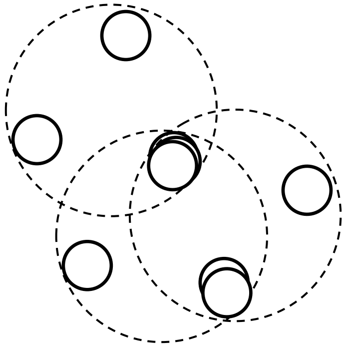
- Outcome variables
  - Tool ratings, citations, & users
- Explanatory variables
  - Team size
  - Contributor diversities: gender, affiliation, country, and publication.
  - Tool attributes: difficulty, open source, versions, and online duration.
- Methods
  - Logit regression

Compositional Influences on nanoHUB Team Assembly

	Ratings (+)	Citations	Users (>250)
Team size	0.18 (.28)	0.24 (.31)	-0.47 (.33)
Number of females	-0.75 (.50)	-0.91 (.57)	-0.58 (.49)
Num country origin	-0.27 (.30)	0.27 (.33)	**0.59*** (.34)
Num of universities	0.32 (.33)	**0.73*** (.33)	0.53 (.38)
Max H-index	0.02 (.02)	**0.05**** (.02)	**0.04**** (.02)
H-index diversity	0.08 (1.16)	-1.26 (1.44)	0.70 (1.48)
Publication diversity	-0.13 (1.28)	-0.58 (1.54)	0.31 (1.62)
Tool controls:			
Tool difficulty	-0.03 (.31)	-0.02 (.38)	**-0.65*** (.36)
Open source	0.72 (.92)	**2.08*** (1.08)	1.65 (1.06)
Number of versions	**0.21**** (.09)	0.03 (.05)	**0.21**** (.10)
Log likelihood	-47.19	-58.09	-61.40
Cox & Snell R^2	0.14	0.25	0.25

Note: * p<.10, ** p<.05, *** p<.01

Four Levels of Influence on nanoHUB Team Assembly

Compositional Level	Relational Level	Multimodal Network Level	Ecosystem Level
			
(a) Team as a collection of individuals	(b) Team as individuals and relations	(c) Team as a network of individuals and tasks	(d) Ecosystem of teams

○ Individual

T Task

Relational Influences on nanoHUB Team Assembly

- Outcome variables
  - Co-contribution network(s)
- Explanatory variables
  - Contributor attributes
  - Network structures
  - Covariate networks (co-authorship and citation)
  - Positions in co-authorship and citation networks
- Methods: p*/Exponential random graph model

Relational Influences on nanoHUB Team Assembly

Co-contribution in ...	Teams (>250 users)	Teams (<250 users)
Female	0.16 (.20)	0.17 (.21)
Same country origin	-0.01 (.21)	0.17 (.17)
Same university	**0.86***** (.10)	**1.59***** (.14)
H-index	**-0.04***** (.01)	**-0.05**** (.02)
H-index difference	**0.04***** (.02)	**0.10***** (.03)
Publication difference	-0.002 (.002)	**-0.009***** (.003)
Co-author relation (Ln)	**1.69***** (.39)	**1.39***** (.53)
Citation relation (Ln)	0.36 (.29)	**1.46***** (.37)
Control:		
Purdue	**-0.39***** (.09)	**-0.26***** (.10)
NCN	**0.57***** (.14)	**1.16***** (.20)
Edge	**-3.69***** (.50)	**-2.05***** (.53)
Alternating stars	**-1.51***** (.12)	**-2.14***** (.18)
Alternating triangles	**3.62***** (.21)	**3.13***** (.18)
N	87	118

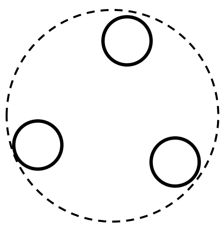
Note: * p<.10, ** p<.05, *** p<.01

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

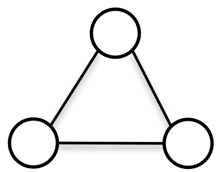
Four Levels of Influence on nanoHUB Team Assembly

Compositional Level



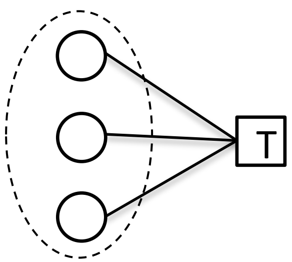
(a) Team as a collection of individuals

Relational Level



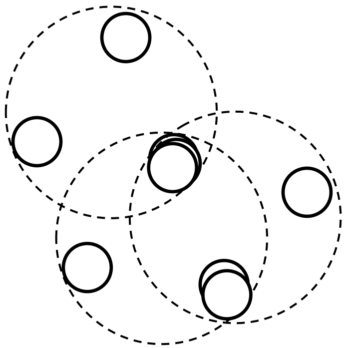
(b) Team as individuals and relations

Multimodal Network Level



(c) Team as a network of individuals and tasks

Ecosystem Level



(d) Ecosystem of teams

Individual

T Task

Multimodal influences on nanoHUB Team Assembly

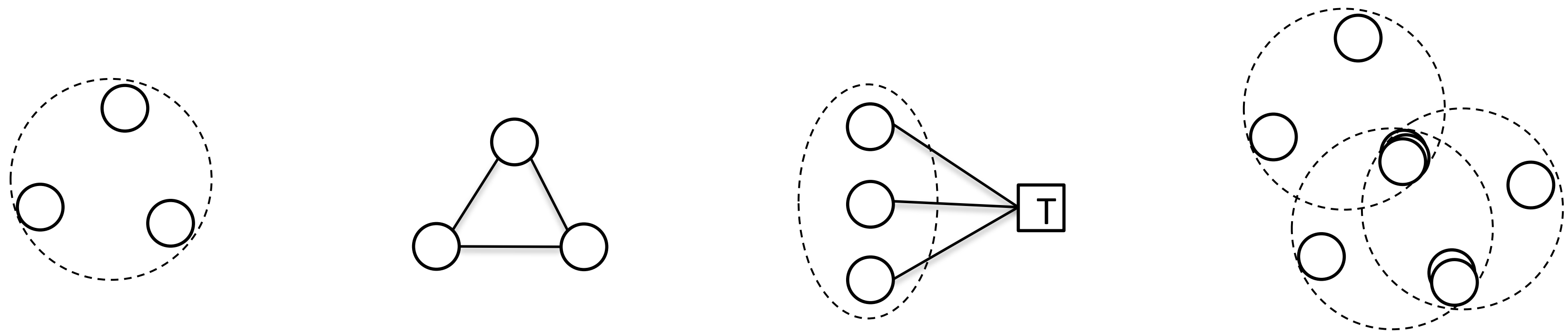
- Outcome variables
  - Team affiliation network(s)
- Explanatory variables
  - Contributor attributes
  - Team attributes
  - Network structures
  - Positions in co-authorship and citation networks
- Methods: p*/BPnet

Multimodal influences on nanoHUB Team assembly

	Teams (>250 users)	Teams (<250 users)
Female	-0.24 (.48)	-0.18 (.33)
Same country origin	-0.07 (.13)	**0.20**** (.10)
Different university	**-0.53***** (.09)	**-1.57***** (.13)
H-index	-0.01 (.01)	0.006 (.02)
H-index difference	0.007 (.008)	0.01 (0.01)
Publication difference	-0.001 (.001)	-0.003 (.002)
Team:		
Tool difficulty	0.05 (.18)	**0.39**** (.16)
Open source	**-1.57***** (.53)	-0.71 (.67)
Ratings (Binary)	0.15 (.27)	0.02 (.21)
Num citations (Ln)	**0.67***** (.18)	-0.06 (.27)
Num users (Ln)	-0.27 (.23)	0.001 (.12)
Control:		
Purdue	**-1.01***** (.28)	**-1.22***** (.16)
NCN	**2.89***** (.45)	**2.51***** (.33)
Edge	0.31 (2.01)	0.17 (1.04)
Contributor stars	**-0.96***** (.30)	**-0.97***** (.22)
Team stars	-0.06 (.61)	**-1.12**** (.53)

Four Levels of Influence on nanoHUB Team Assembly

Compositional Level | Relational Level | Multimodal Network Level | Ecosystem Level



(a) Team as a collection of individuals

(b) Team as individuals and relations

(c) Team as a network of individuals and tasks

(d) Ecosystem of teams

Individual

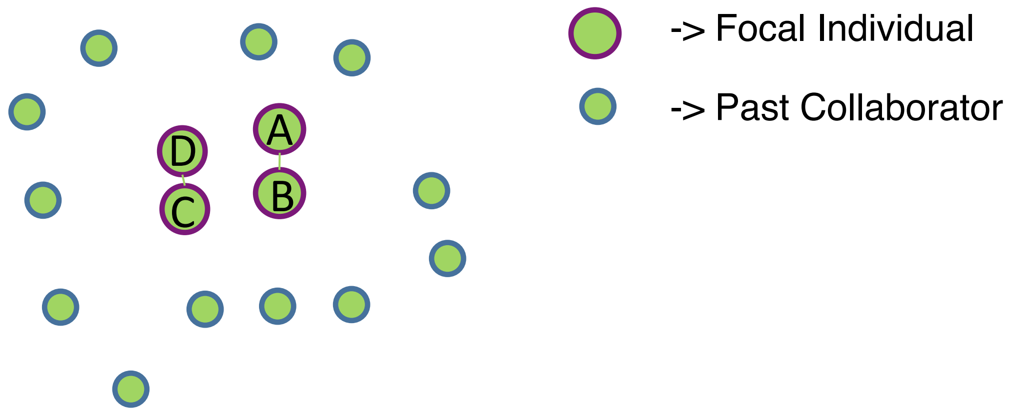
Task

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

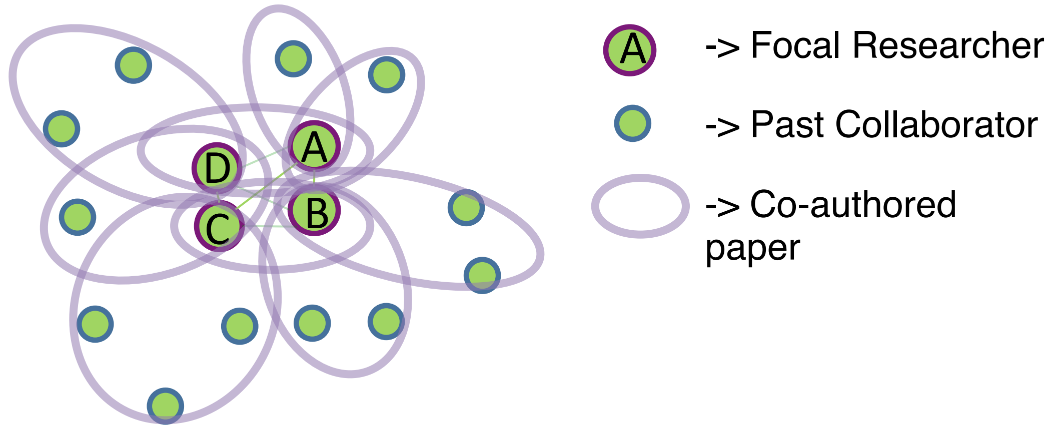
Scientific Ecosystem

- Teams do not assemble in a "vacuum"
- Teams emerge from networks of prior collaborations in a particular space
  - An "ECOSYSTEM"



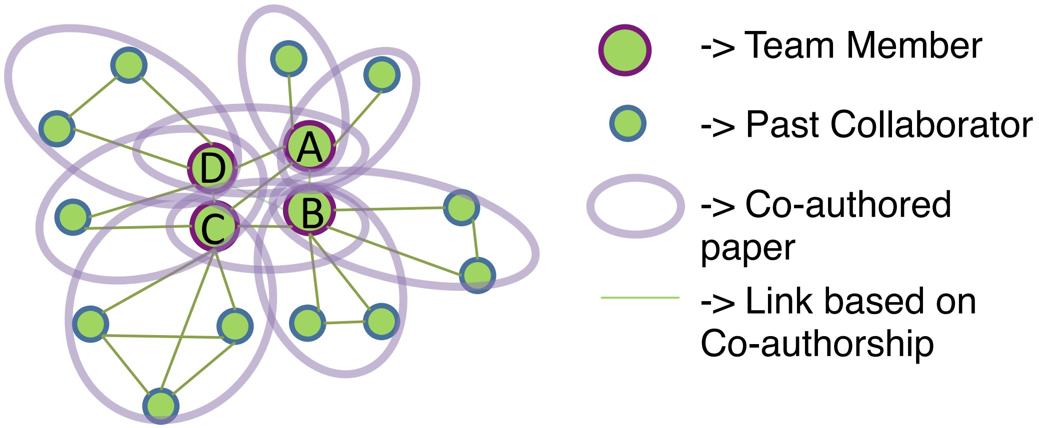
Scientific Ecosystem

- Teams do not assemble in a "vacuum"
- Teams emerge from networks of prior collaborations in a particular space
  - An "ECOSYSTEM"



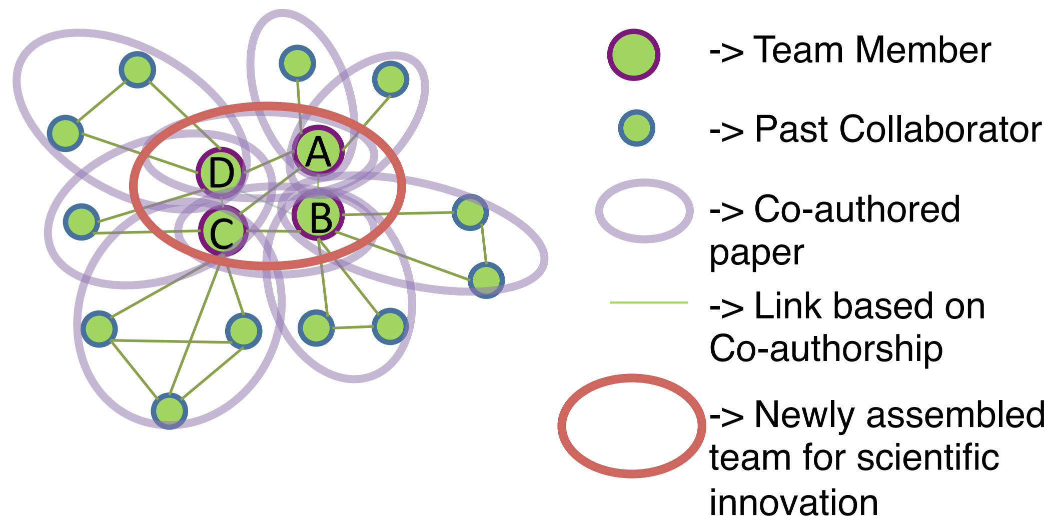
Scientific Ecosystem

- Teams do not assemble in a "vacuum"
- Teams emerge from networks of prior collaborations in a particular space
  - An "ECOSYSTEM"



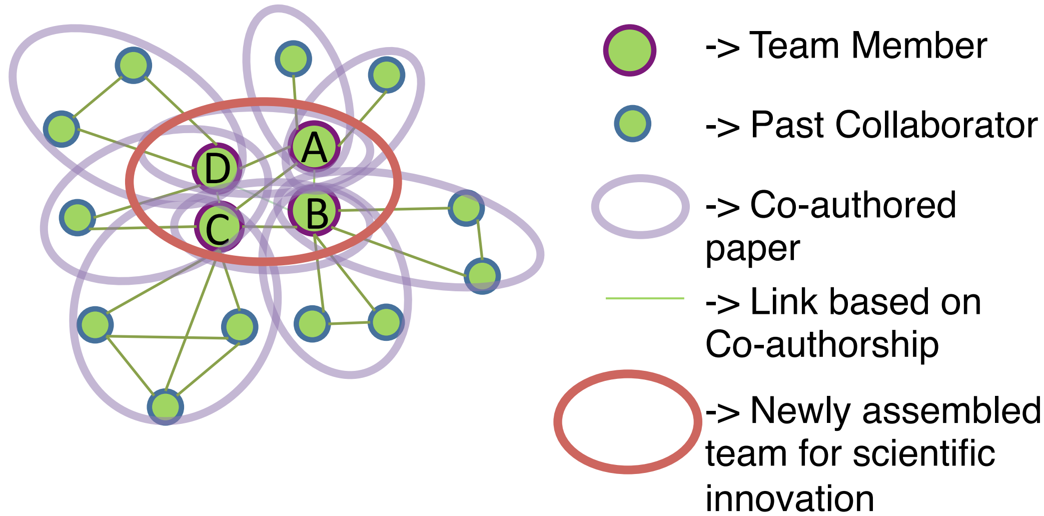
Scientific Ecosystem

- Teams do not assemble in a "vacuum"
- Teams emerge from networks of prior collaborations in a particular space
  - An "ECOSYSTEM"



Scientific Ecosystem as Antecedent of Team Assembly and Performance

- Teams do not assemble in a "vacuum"
- Teams emerge from networks of prior collaborations in a particular space
  - An "ECOSYSTEM"

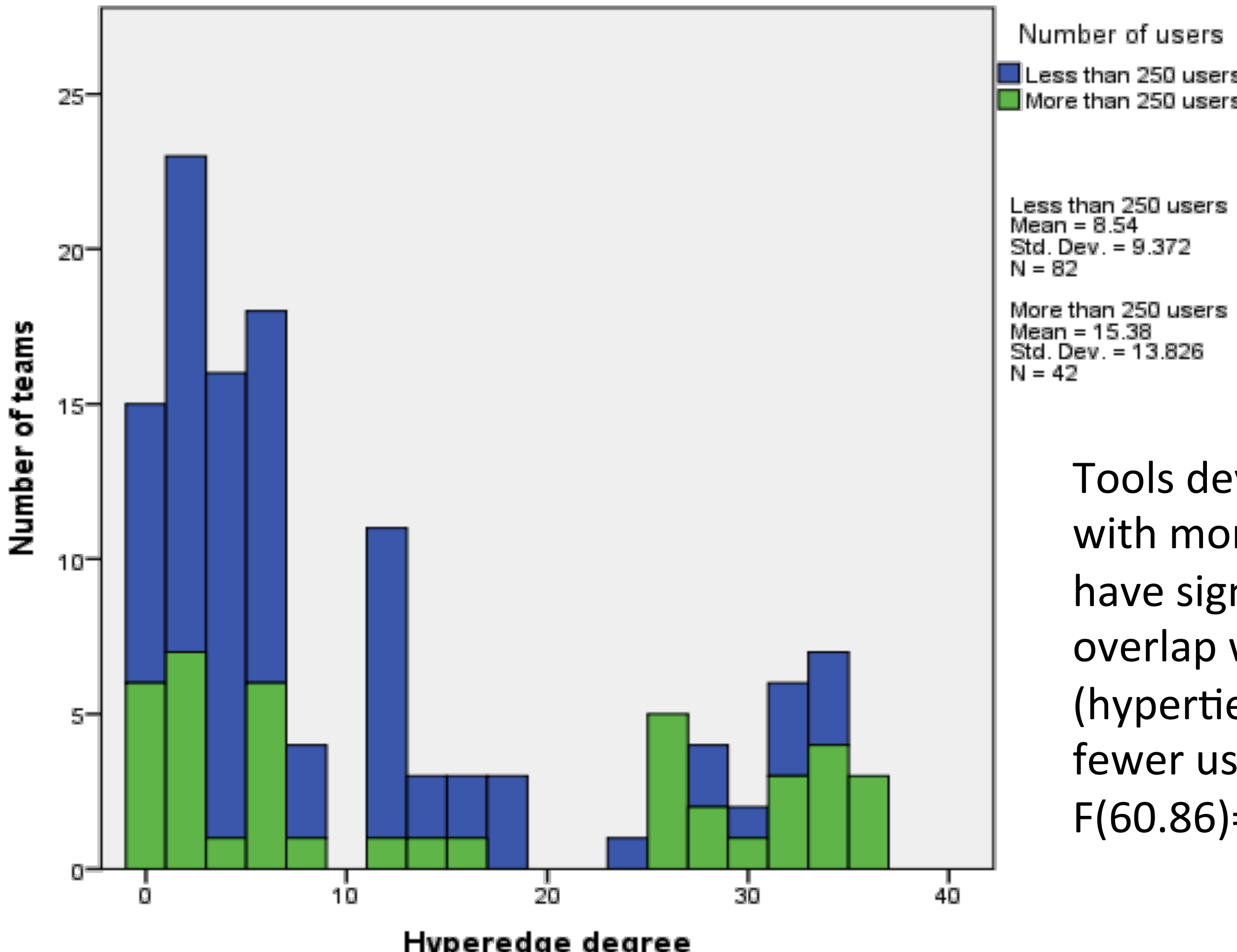


- Are there certain characteristics of the scientific ecosystem that lead to team assembly?
- Do variations in these ecosystem characteristics predict team performance?

Ecosystem influences on nanoHUB Team Assembly

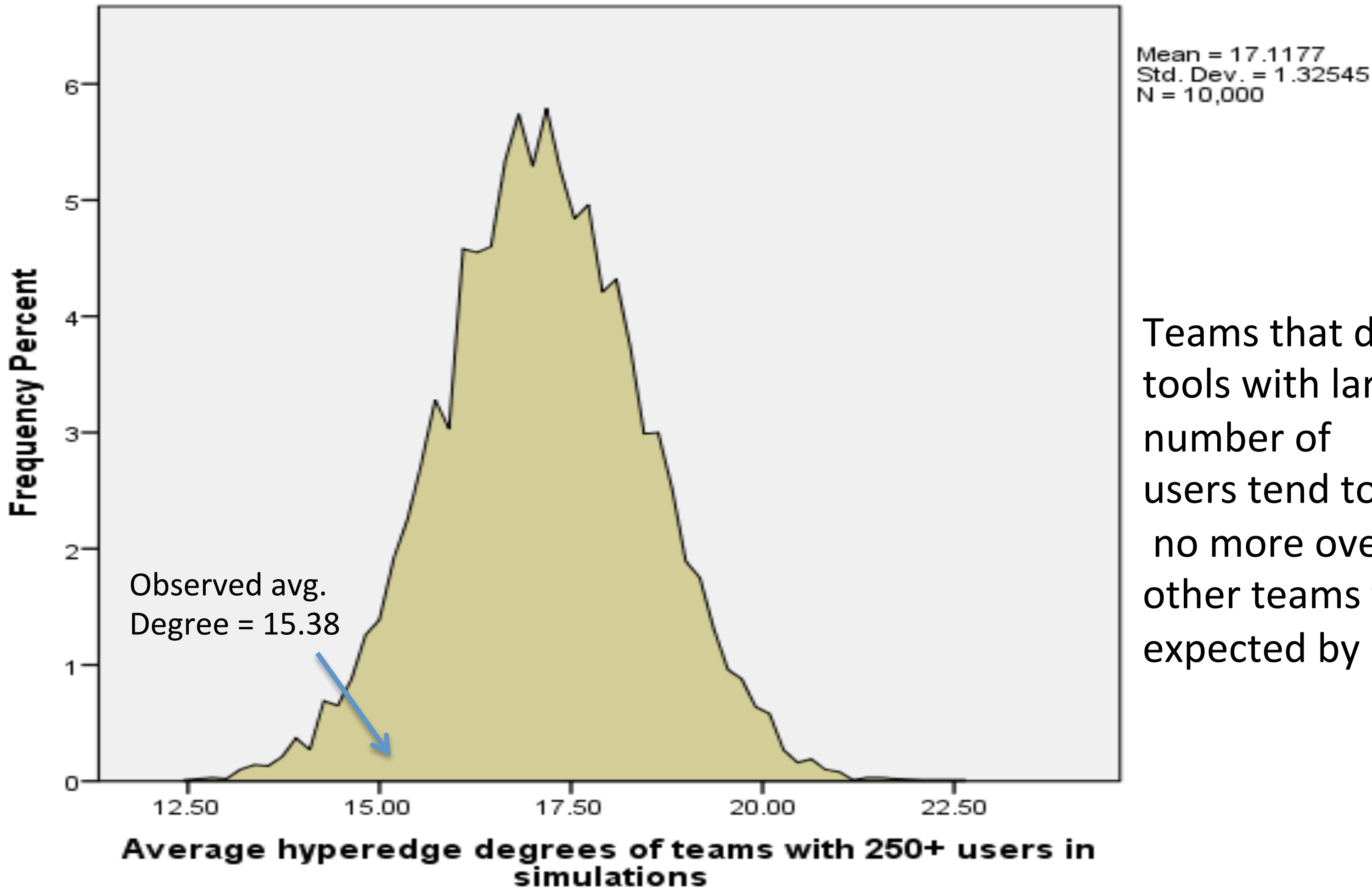
- Target network statistics
  - Team hyperties
    - Number of teams with which the focal team has overlapping members (i.e. hypertie degree).
  - Closure of team hyperties
    - Ratio of overlap among teams with which focal team has overlapping members. (i.e. mean clustering coefficients)
- Methods: Estimation by comparison to distribution generated by simulating hypergraphs

Ecosystem influence on nanoHUB Team Assembly



Tools developed by teams with more than 250 users have significantly more overlap with other teams (hyperties) than tools with fewer users $F(60.86)=-2.89$, $p=0.005$.

Observed vs Simulated Overlap for Teams w/ 250+ Users

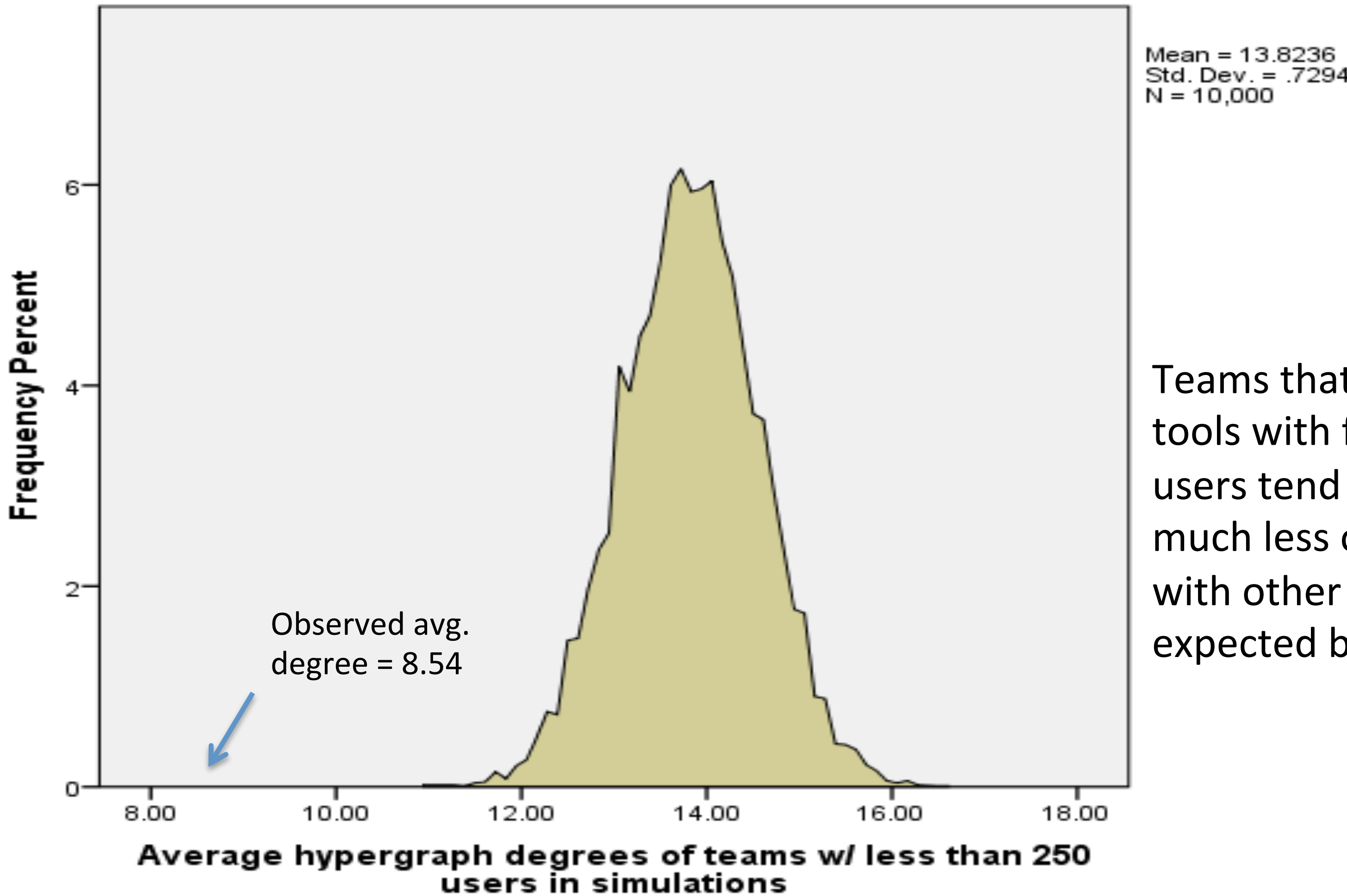


Teams that develop tools with larger number of users tend to have no more overlap with other teams than expected by chance.

NORTHWESTERN UNIVERSITY

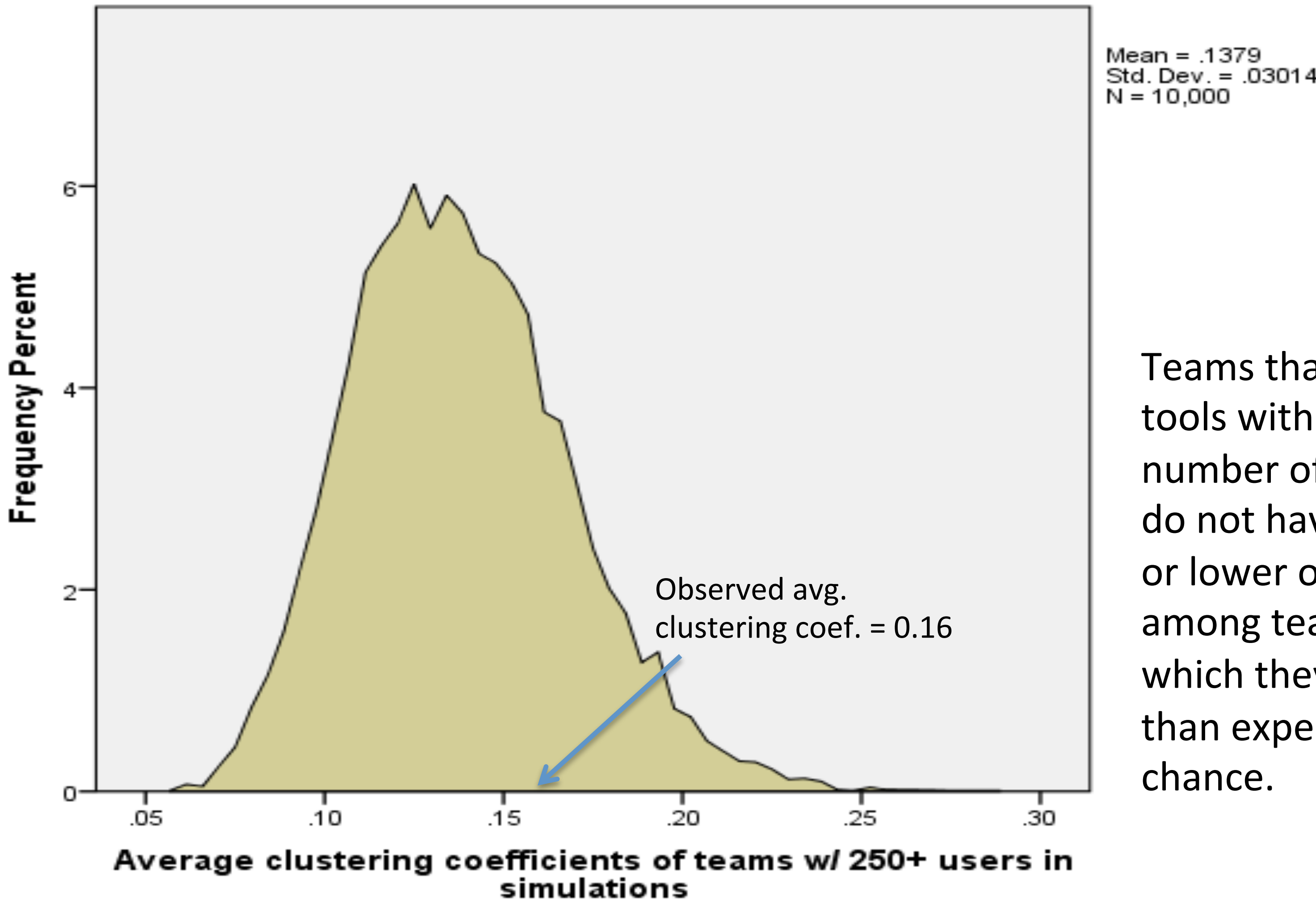
advancing the science of networks in communities

Observed vs Simulated Overlap for Teams w/ < 250 Users



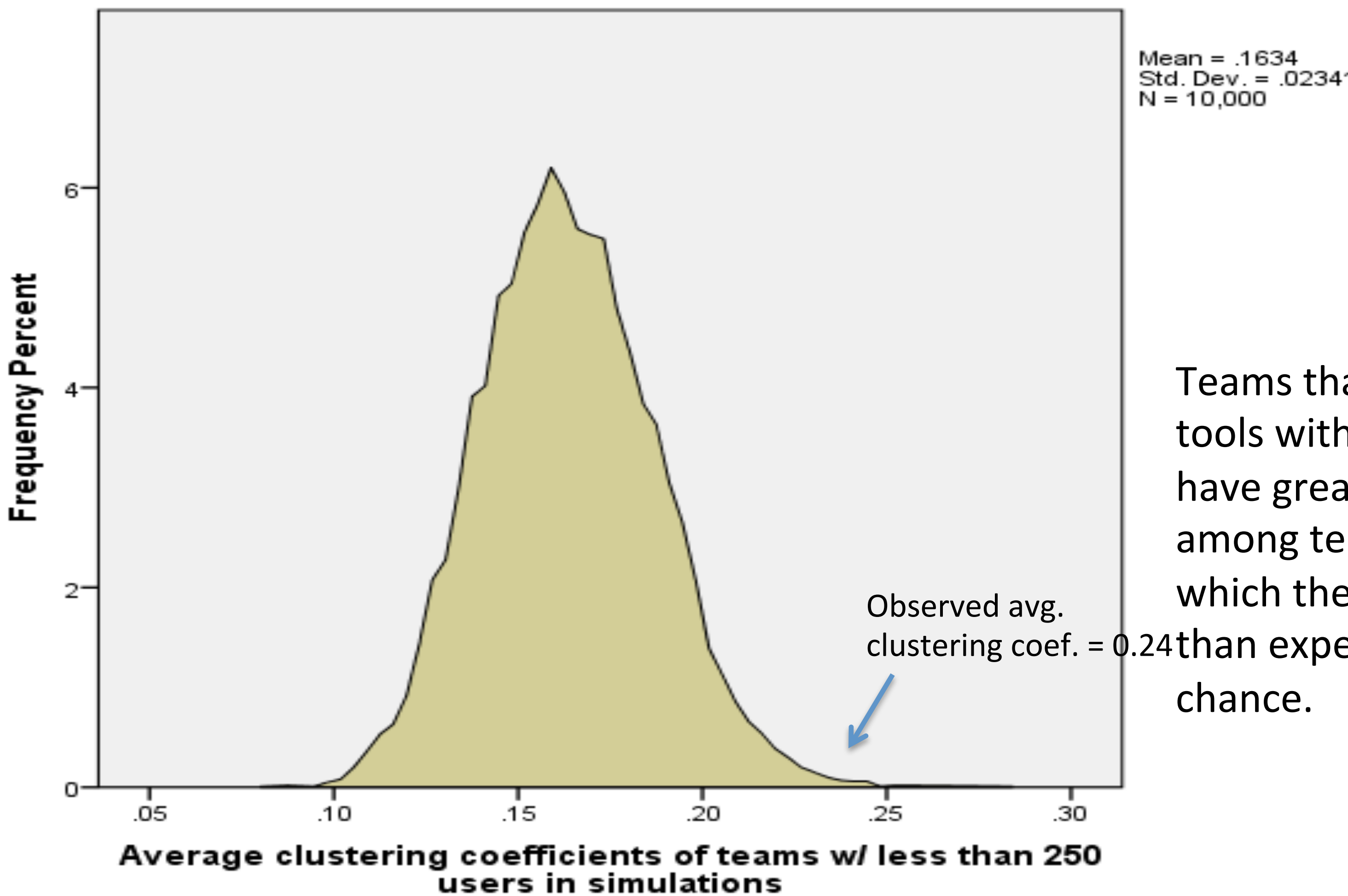
Teams that develop tools with fewer users tend to have much less overlap with other teams than expected by chance.

Clustering Coefficients of Teams w/ 250+ Users



Teams that develop tools with higher number of users do not have greater or lower overlap among teams with which they overlap than expected by chance.

Clustering Coefficients of Teams w/ less than 250 Users



Teams that develop tools with fewer users have greater overlap among teams with which they overlap than expected by chance.

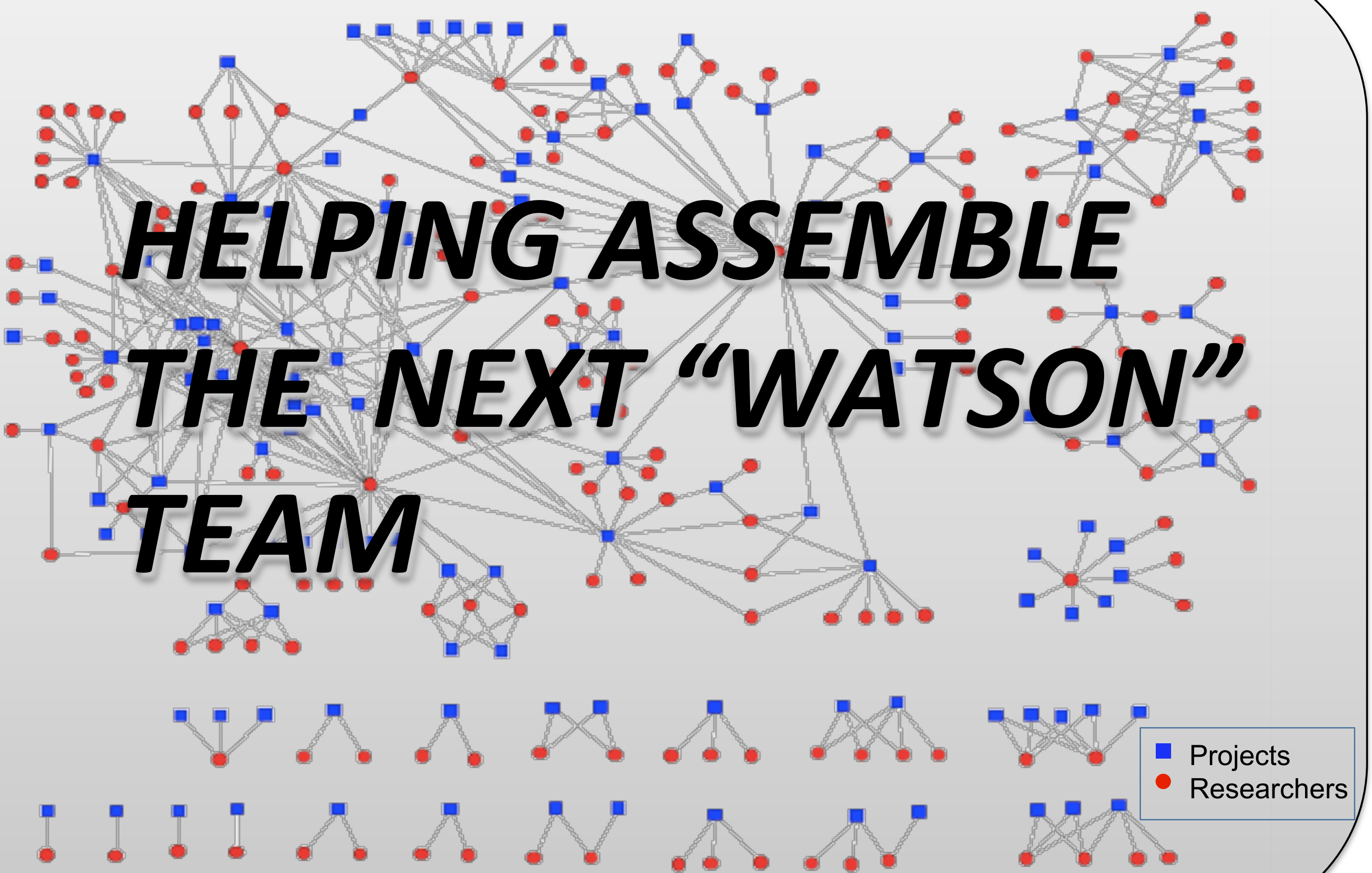
NORTHWESTERN UNIVERSITY

advancing the science of networks in communities

Exemplar IV

From Understanding to Enabling Team Assembly Or ….

HELPING ASSEMBLE THE NEXT "WATSON" TEAM



NORTHWESTERN UNIVERSITY

SONIC
advancing the science of networks in communities

Key Takeaways

- Understanding and enabling team assembly is well poised to make an intellectual leap by leveraging recent advances in:
  - Theories: Theories about the socio-technical motivations for creating, maintaining, dissolving and re-creating links to engage in team assembly
  - Data: Developments in Semantic Web/Web 2.0 provide the technological capability to capture, store , merge, and query relational metadata needed to more effectively understand and enable team assembly.
  - Methods: An ensemble of qualitative and quantitative methods (exponential random graph modeling (*p**) techniques to understand and enable theoretically grounded network recommendations for team assembly
  - Computational infrastructure: Cloud computing and petascale applications are critical to face the computational challenges in understanding and enabling team assembly.

Acknowledgements

National Science Foundation
WHERE DISCOVERIES BEGIN

National Cancer Institute
U.S. National Institutes of Health | www.cancer.gov

AIR FORCE RESEARCH LABORATORY

ARMY RESEARCH INSTITUTE

ARL

NORTHWESTERN UNIVERSITY

SONIC
advancing the science of networks in communities