

# ***Some Assembly Required: Networking and Team Science for 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, IIS-1249137**

**Army Research Institute W91WAW-08-C-0106/P00004;**

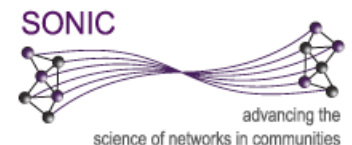
**Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053**

**& NIH/NCRR & NIH/NCRR grant to Northwestern University**

**Clinical and Translational Sciences Institute (NUCATS)**



**NORTHWESTERN  
UNIVERSITY**



## Building the Team That Built Watson



Cezir Muhammedi/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.

 RECOMMEND

 TWITTER

 LINKEDIN

 E-MAIL

 PRINT

 REPRINTS


 SHARE

  
Descendants  
now playing everywhere

David Ferrucci,  
New York Times  
1/7/2012



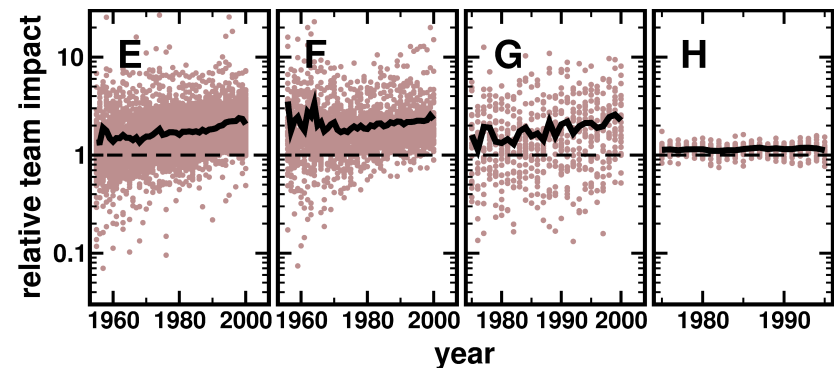
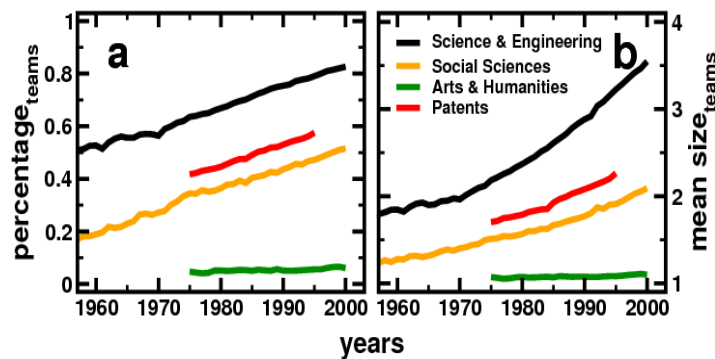
NORTHWESTERN  
UNIVERSITY

SONIC  
  
advancing the  
science of networks in communities

# Move to Teams

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.

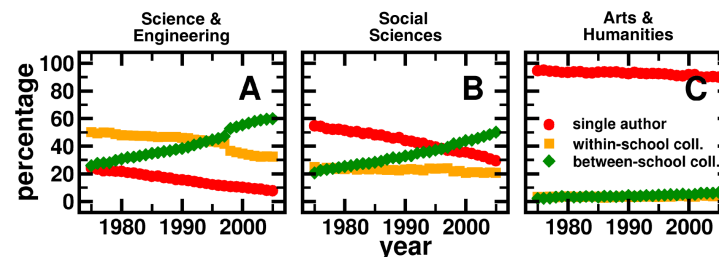
1. For virtually all fields, research is increasingly done in teams
2. Teams typically produce more highly cited research than individuals do (accounting for self-citations), and this team advantage is increasing over time.
3. Teams now produce the exceptionally high impact research, even where that distinction was once the domain of solo authors.



# Move to *Virtual* Team Science

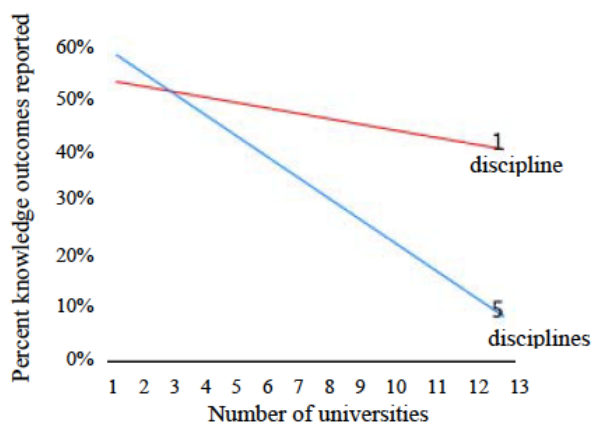
The trend toward virtual communities was ***not*** driven by a growth in teamwork by scientists working with other co-located scientists. Using the Web of Science database to analyze the collaboration arrangements of over 4,000,000 papers over a 30 year period, they found that:

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



# The Successes are by Teams, but not all Teams are Successful

Study of NSF-funded project teams finds collaborations involving more universities produced fewer patents, publications, and other knowledge outcomes, especially when more than one discipline was represented in the project.

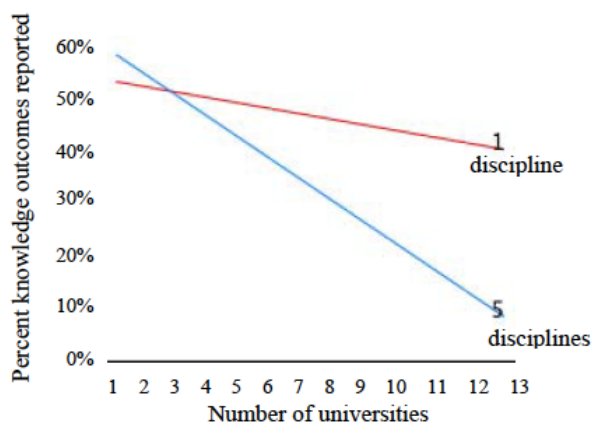


Distance & Disciplinary  
Differences Most Challenging

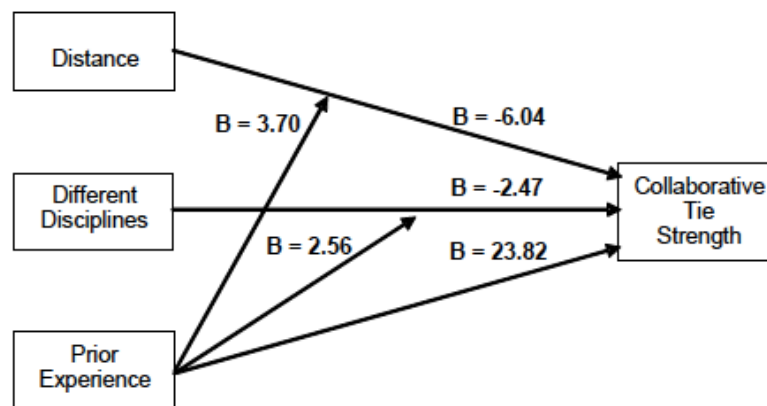


# The Successes are by Teams, but not all Teams are Successful

Study of NSF-funded project teams finds collaborations involving more universities produced fewer patents, publications, and other knowledge outcomes, especially when more than one discipline was represented in the project.



Distance & Disciplinary Differences Most Challenging



But, prior experience mitigates the harmful effects of distance and disciplinary differences



# Key Takeaways

- Understanding and enabling assembly of scientific teams 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

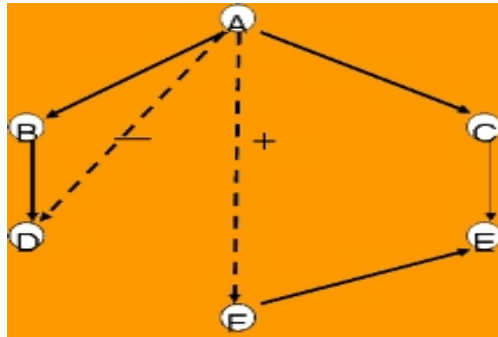
## *Sources:*

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*.

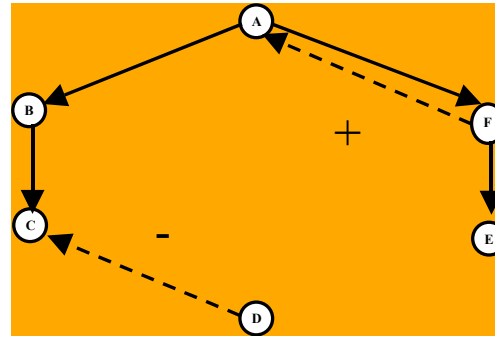
Monge, P. R. & Contractor, N. S. (2003). *Theories of Communication Networks*. New York: Oxford University Press.



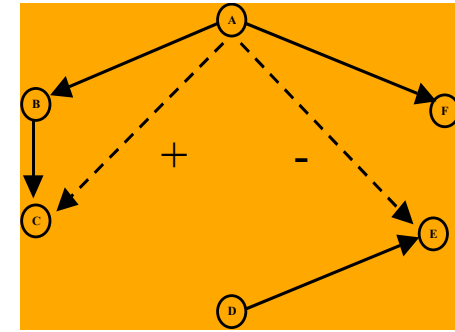
# “Structural signatures” of MTML Motivations for Team Assembly



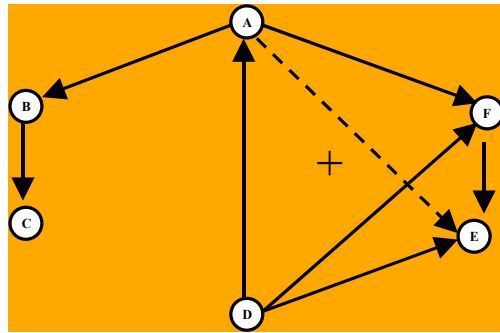
Theories of Self interest



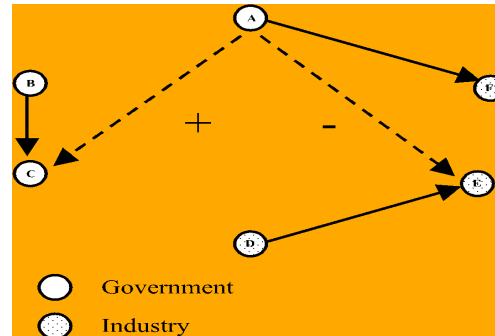
Theories of Exchange



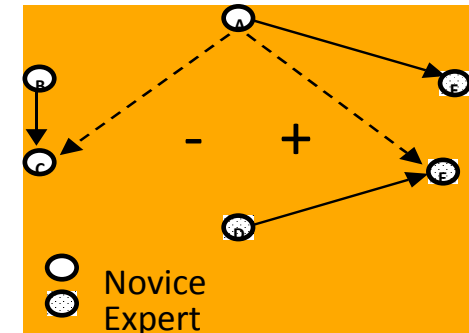
Theories of Balance



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 empirically testing, extending, and exploring theories about team assembly ...



# The Hubble telescope: \$2.5 billion



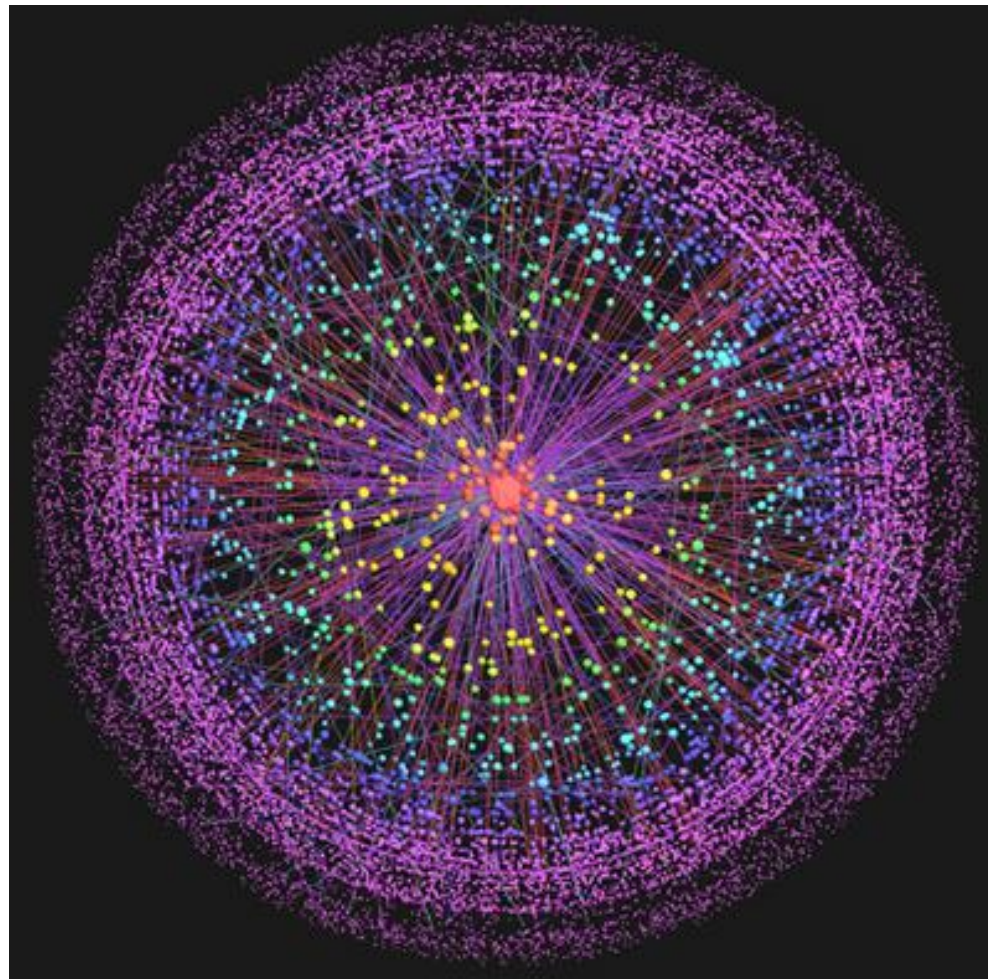
*Source: David Lazer*

# CERN particle accelerator: \$1 billion/year



*Source: David Lazer*

# The Web: priceless\*



\* *Apologies to MasterCard*



Source: David Lazer

SOCIAL SCIENCE

# Computational Social Science

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

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.

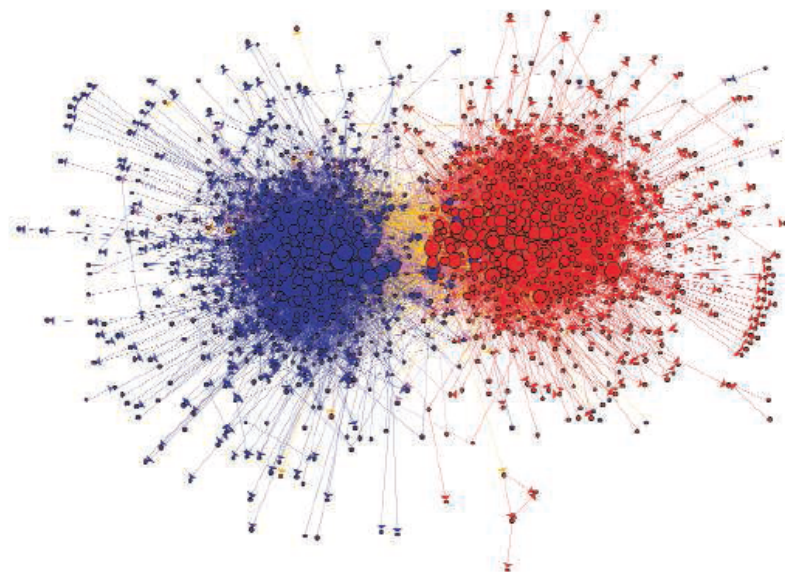
The capacity to collect and analyze massive amounts of data has transformed such fields as biology and physics. But the emergence of a data-driven “computational social science” has been much slower. Leading journals in economics, sociology, and political science show little evidence of this field. But computational social science is occurring—in Internet companies such as Google and Yahoo, and in govern-

ment 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

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.

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



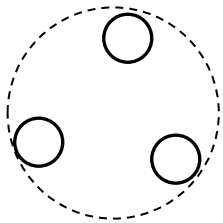
**Data from the blogosphere.** Shown is a link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number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

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



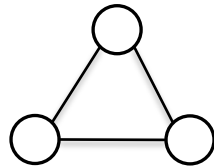
# Four Levels of Influences on Team Assembly

Compositional Level



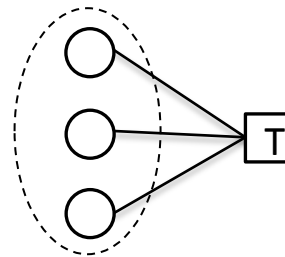
(a) Team as a collection of individuals

Relational Level



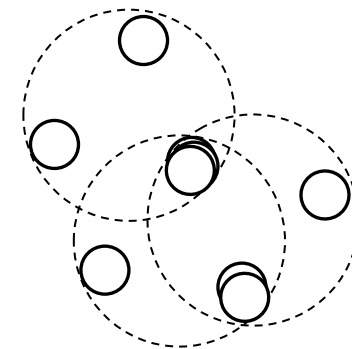
(b) Team as individuals and relations

Multimodal Network Level



(c) Team as a network of individuals and tasks

Ecosystem Level



(d) Ecosystem of teams

○ Individual

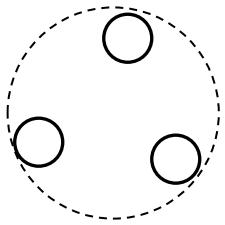
□ T Task



NORTHWESTERN  
UNIVERSITY

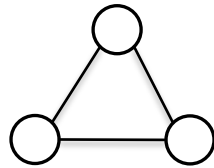
# Four Levels of Influences on Team Assembly

## Compositional Level



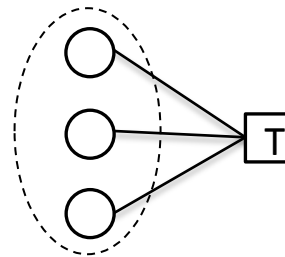
(a) Team as a collection of individuals

## Relational Level



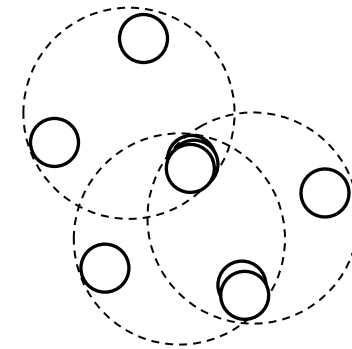
(b) Team as individuals and relations

## Multimodal Network Level



(c) Team as a network of individuals and tasks

## Ecosystem Level



(d) Ecosystem of teams

○ Individual

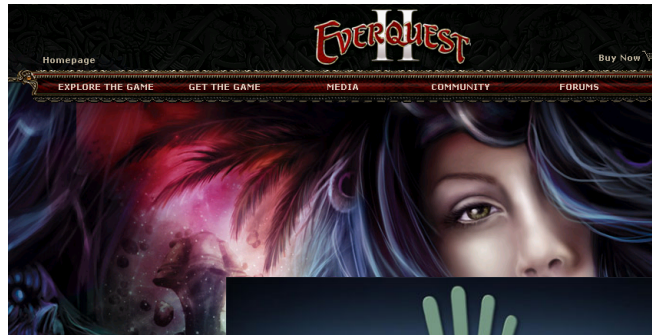
□ T Task



# Virtual World Exploratorium



NORTHWESTERN  
UNIVERSITY



[vwobservatory.org](http://vwobservatory.org)



NORTHWESTERN  
UNIVERSITY



復旦大學



advancing the  
science of networks in communities

# Combat Groups in EverQuest II

- Difficult combat tasks require collaboration of multiple players and assembly of combat groups
- From 2006-08-27 to 2006-09-11 on Antonia Bayle Server
  - 8,423 players
  - 46,393 groups
  - 9,436,741 combat related records

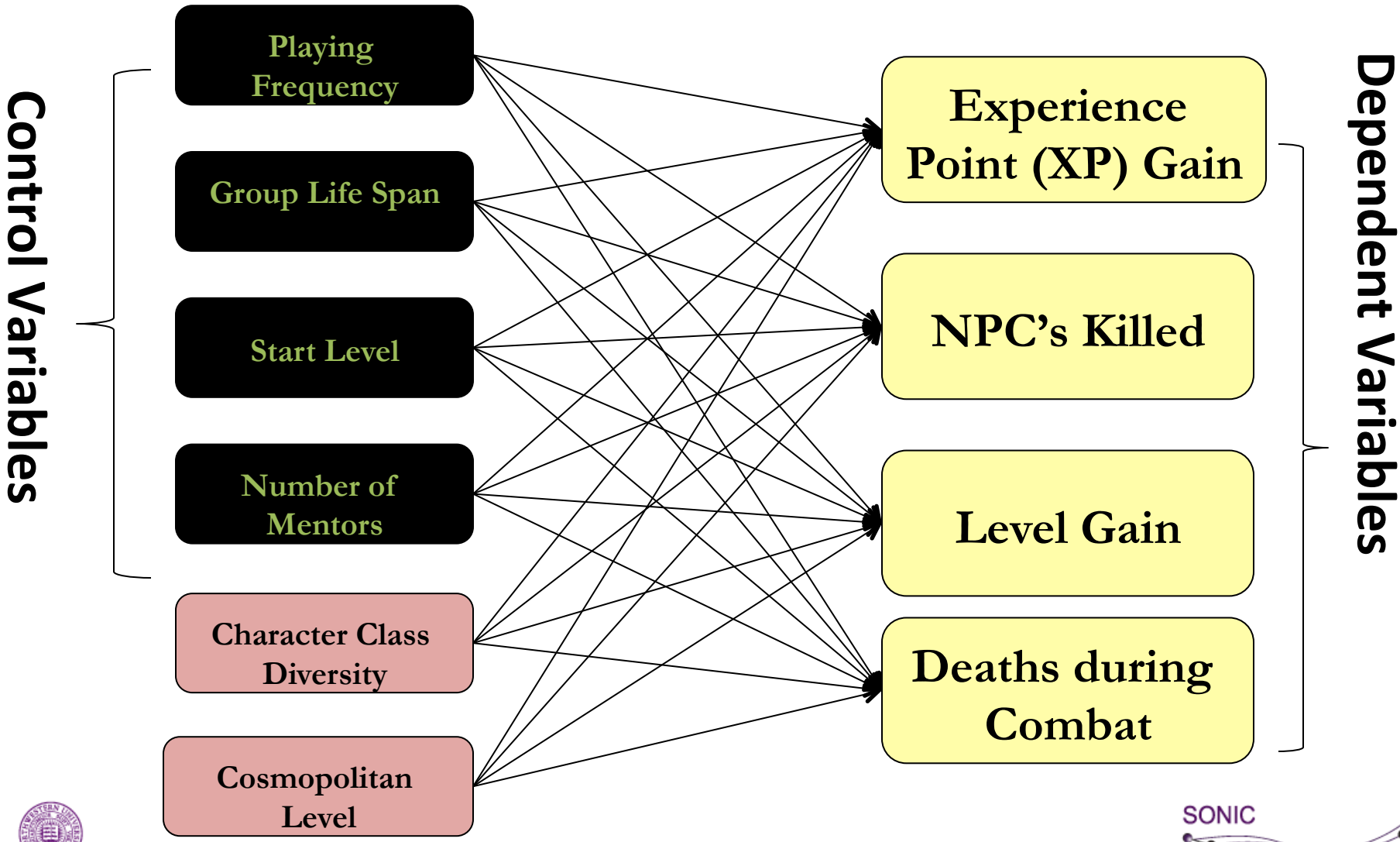


# What makes a group successful?

- **Group Diversity**
  - Four character classes in the game: Fighter, Mage, Scout and Priest, each having a different role in a group
  - Measure Group Diversity: Blau's Index
- **Group member's cosmopolitan level**
  - Group members being involved in multiple different groups



# Effects of Group Attributes on Performance Measures



## Regression Analysis Results on Combat Groups of Four Players

	XP	NPC's	Level Gain	Deaths
<b>Constant</b>	-20939.926**	-3.376 (.361)	.717**	4.011**
<b>Frequency</b>	-1553.494 (.105)	4.127**	-.010 (.816)	.601**
<b>Life Span</b>	736.797**	1.174**	.015**	.063**
<b>Num of Mentor</b>	1.000**	1.000**	1.000**	-.050**
<b>Diversity</b>	20819.998**	14.342**	.726**	-1.873 (.095)
<b>Member Cosmo.</b>	30.254 (.612)	-.025 (.698)	-.010**	-.032**
<b>R<sup>2</sup></b>	.000	0.621	0.571	0.244
<b>F</b>	595.213 (p=.000)	1368.793 (p=.000)	176.071 (.000)	96.274 (.000)

Diversity helps the groups to achieve more.

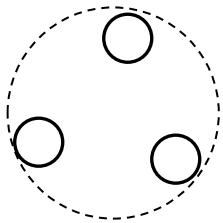
Members being cosmopolitan doesn't help with gains but helps to avoid loss.

\*\* indicates significant results at .01 level



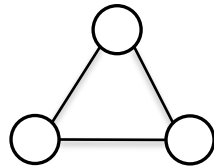
# Four Levels of Influence on Team Assembly

Compositional Level



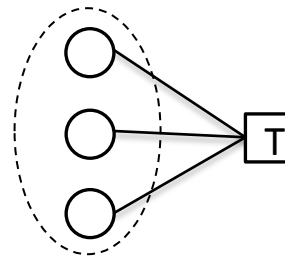
(a) Team as a collection of individuals

Relational Level



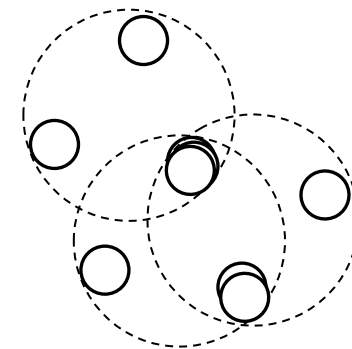
(b) Team as individuals and relations

Multimodal Network Level



(c) Team as a network of individuals and tasks

Ecosystem Level



(d) Ecosystem of teams

○ Individual

□ T Task



NORTHWESTERN  
UNIVERSITY

# Data Set

- 1,103 grant proposals submitted to NSF (both awarded and un-awarded)
- 2 interdisciplinary programs
- 3-year period
- 2,186 PIs and Co-PIs



# Who submits proposals?

- Individuals are *more* likely to submit proposals with their co-authors.
- Individuals are *more* likely to submit proposals with those they cite



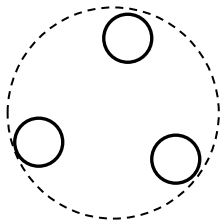
# Who submits *successful* proposals?

- Individuals are *more* likely to submit *successful* proposals with their co-authors.
- Individuals are *less* likely to submit successful proposals with those they cite



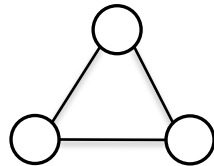
# Four Levels of Influence on nanoHUB Team Assembly

Compositional Level



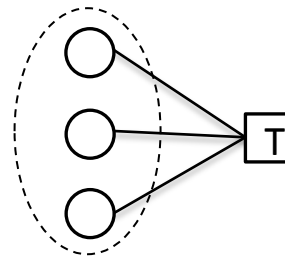
(a) Team as a collection of individuals

Relational Level



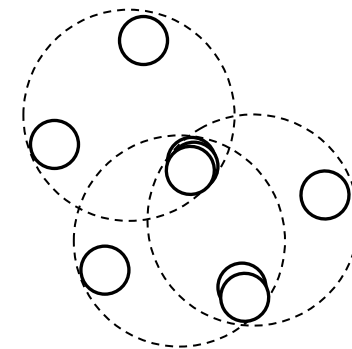
(b) Team as individuals and relations

Multimodal Network Level



(c) Team as a network of individuals and tasks

Ecosystem Level



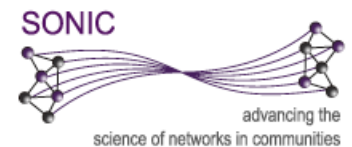
(d) Ecosystem of teams

○ Individual

□ T Task



NORTHWESTERN  
UNIVERSITY



# 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



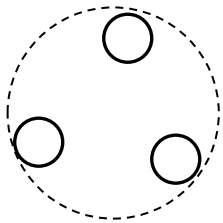
# Multimodal influences on nanoHUB Team assembly

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



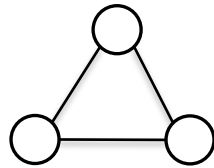
# Four Levels of Influence on Team Assembly

Compositional Level



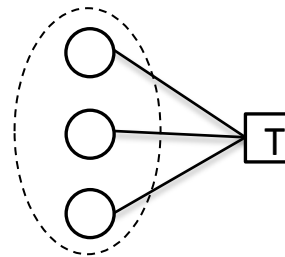
(a) Team as a collection of individuals

Relational Level



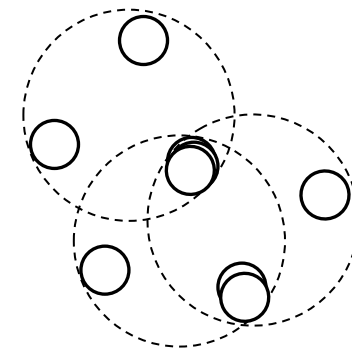
(b) Team as individuals and relations

Multimodal Network Level



(c) Team as a network of individuals and tasks

Ecosystem Level



(d) Ecosystem of teams

○ Individual

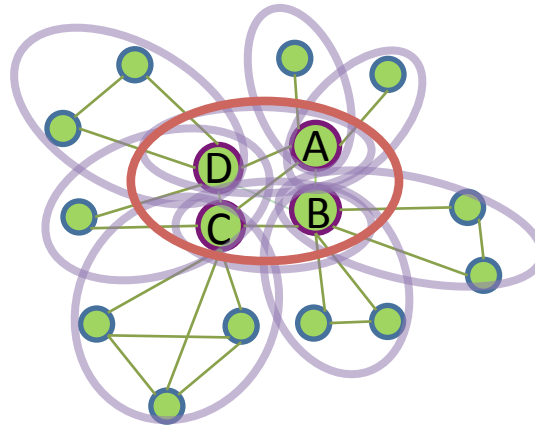
□ T Task



NORTHWESTERN  
UNIVERSITY

# 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”



- -> Team Member
- -> Past Collaborator
- -> Co-authored paper
- -> Link based on Co-authorship
- -> Newly assembled team for scientific innovation

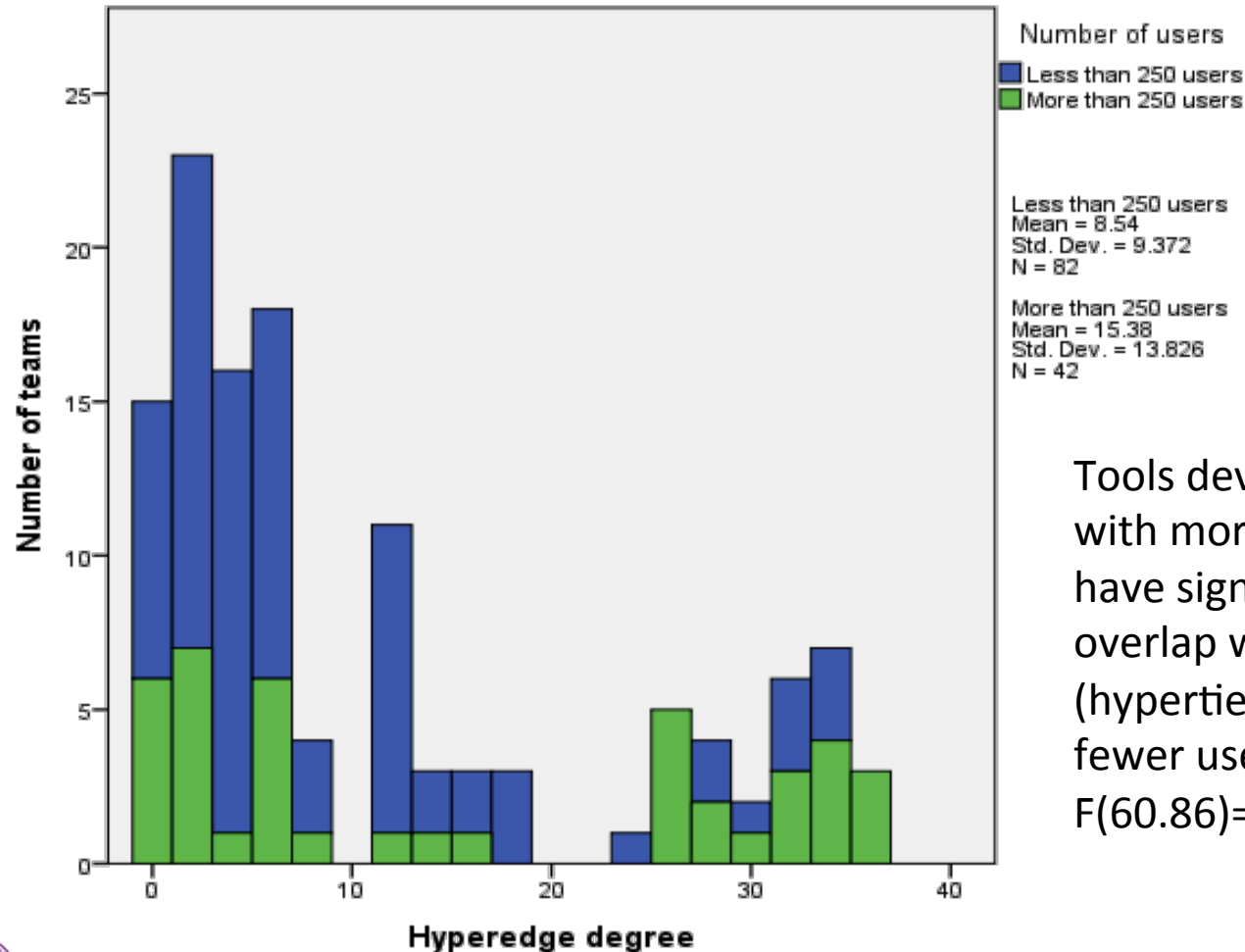


# Ecosystem influences on nanoHUB Team Assembly

- 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  
 $F(60.86)=-2.89, p=0.005.$



# Ecosystem influences on assembly of clinical and translational science teams

- Bibliographic data on teams of researchers who submitted research proposals to a grant competition funded by NUCATS Pilot Grants
- 100 research proposal teams in original data set
  - Total of 147 participants.
  - 7 teams awarded funding (i.e., “Successful” teams)
  - 47 single author proposals excluded from sample
  - 53 coauthored proposals
- 11 teams included in current analyses
  - 4 funded, 7 unfunded



---

# From Understanding to Enabling Team Assembly Or ....



# Demo

- [NUCATS Semantic C-IKNOW](#)



# Key Takeaways

- Understanding and enabling assembly of scientific teams 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.



# Acknowledgements



National Science Foundation  
WHERE DISCOVERIES BEGIN



National Cancer Institute

U.S. National Institutes of Health | [www.cancer.gov](http://www.cancer.gov)



NORTHWESTERN  
UNIVERSITY

