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

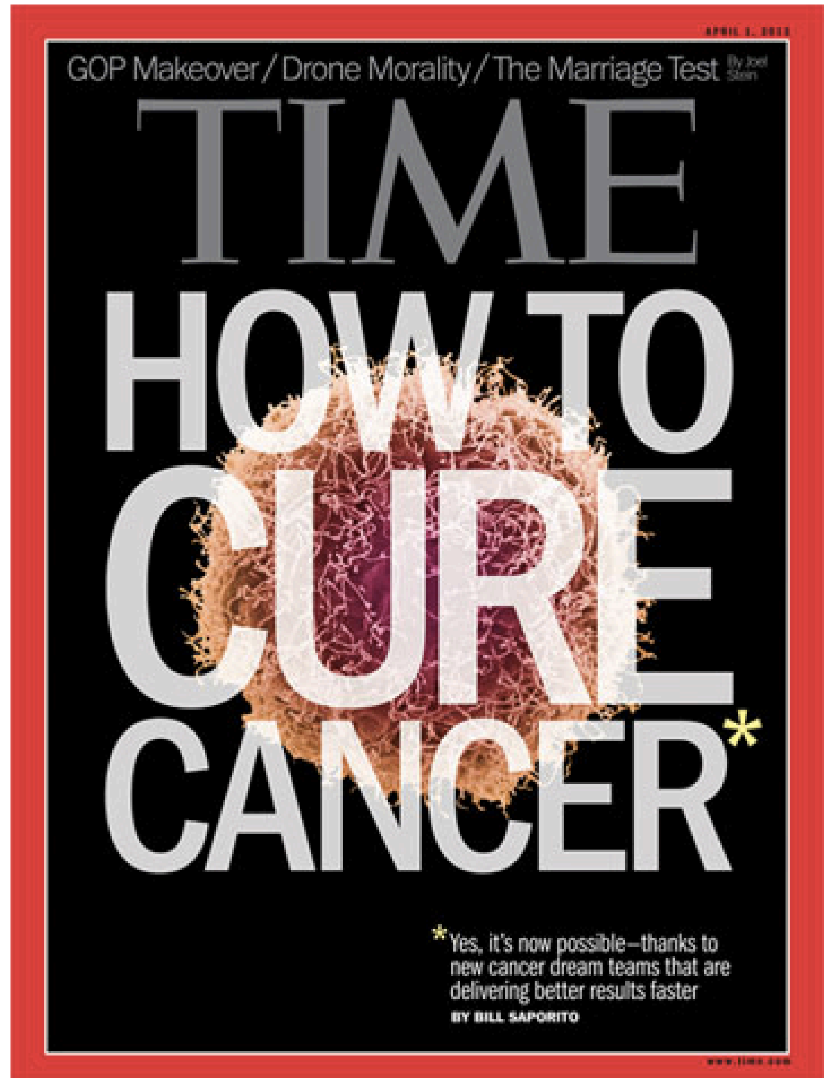
The Conspiracy To End Cancer

By Bill Saporito | Monday, Apr. 01, 2013

The hero scientist who defeats cancer will likely never exist.

No exalted individual, no victory celebration, no Marie Curie or Jonas Salk, who in 1955, after he created the first polio vaccine, was asked, So what's next? Cancer?--as if a doctor finished with one disease could simply shift his attention to another, like a chef turning from the soup to the entrée.

Cancer doesn't work that way. It's not just one disease; it's hundreds, potentially thousands. And not all cancers are caused by just one agent--a virus or bacterium that can be flushed and crushed. Cancer is an intricate and potentially...



Battiere Effect

The No-Stats All-Star

Robert Seale for The New York Times

Statistical Anomaly His greatness is not marked in box scores or at slam-dunk contests, but on the court Shane Battier makes his team better, often much better, and his opponents worse, often much much worse.

NORTHWESTERN UNIVERSITY

New York Times, Feb 15, 2009

SONIC

advancing the science of networks in communities

Tasks don't always come before Teams

Journal of Applied Statistics
Vol. 32, No. 5, 461–474, July 2005

Routledge
Taylor & Francis Group

The Most-Cited Statistical Papers

THOMAS P. RYAN* & WILLIAM H. WOODALL**

*National Institute of Standards and Technology, Gaithersburg, Maryland, USA, **Department of Statistics, Virginia Tech, Blacksburg, Virginia, USA*

(19) With 2,529 citations (120 per year),

Box, G. E. P. & Cox, D. R. (1964) An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, 26, pp. 211–243 (discussion pp. 244–252).

DeGroot (1987) provided some interesting background on this paper from an interview with Professor Box. Box recounted, for example, that he and Cox were on a committee of the Royal Statistical Society and several people suggested that they collaborate. Their motivation and the idea of the paper sprung, to some extent, from the similarities of their family names.

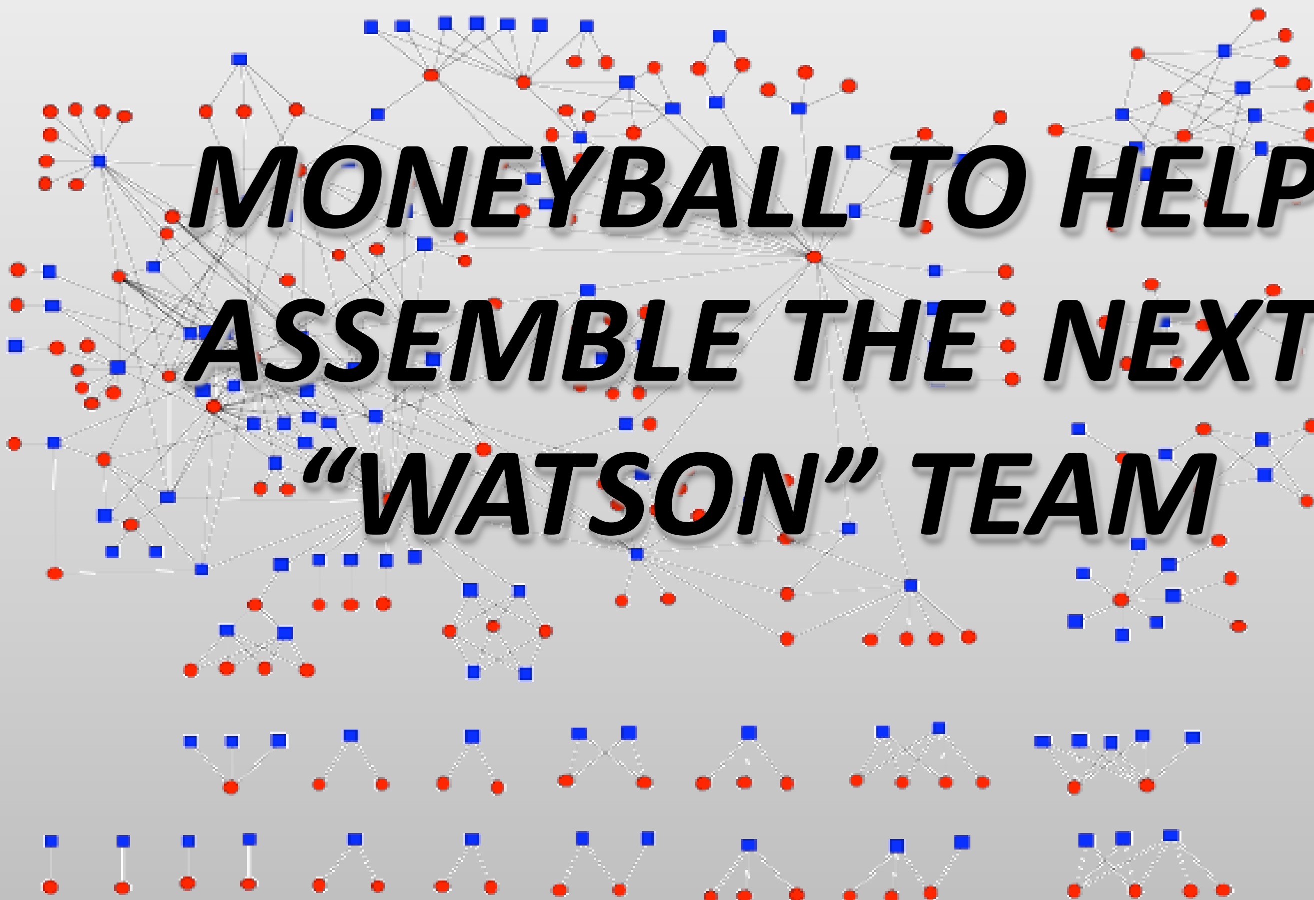
Box & Cox (1964) presented a very useful family of power transformations that have typically been used to transform the dependent variable in a regression model so as to try to meet the assumptions of homoscedasticity and normality of the error terms. The right side of the model can then be transformed in the same manner so as to retrieve the quality of the fit before the dependent variable was transformed.

NORTHWESTERN UNIVERSITY

DeGroot, M. H. (1987) A conversation with George Box, *Statistical Science, 2*, pp. 239 – 258

advancing the science of networks in communities

MONEYBALL TO HELP ASSEMBLE THE NEXT "WATSON" TEAM



"Your goal shouldn't be to buy players. Your goal should be to buy wins. In order to buy wins, you need to buy runs." (Bakshi & Miller, 2011).

NORTHWESTERN UNIVERSITY

SONIC

advancing the science of networks in communities

SOCIAL SCIENCE

Computational Social Science

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]

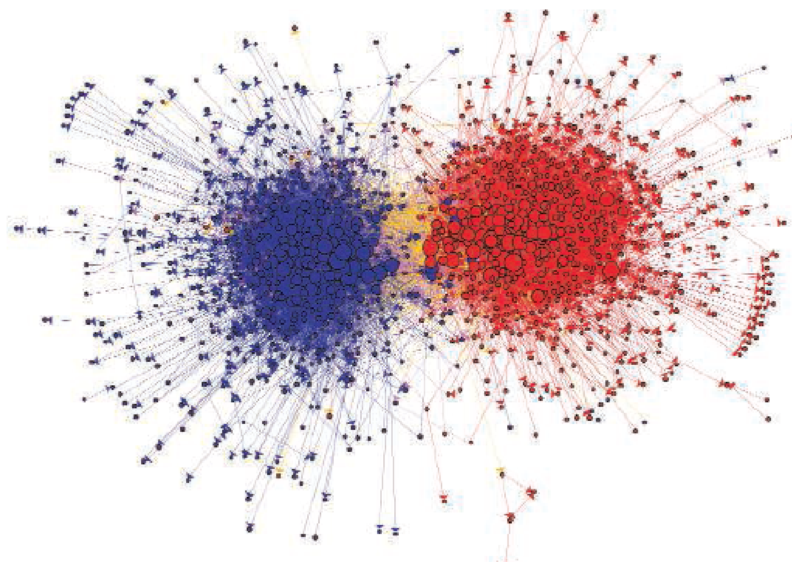
A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

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

[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.



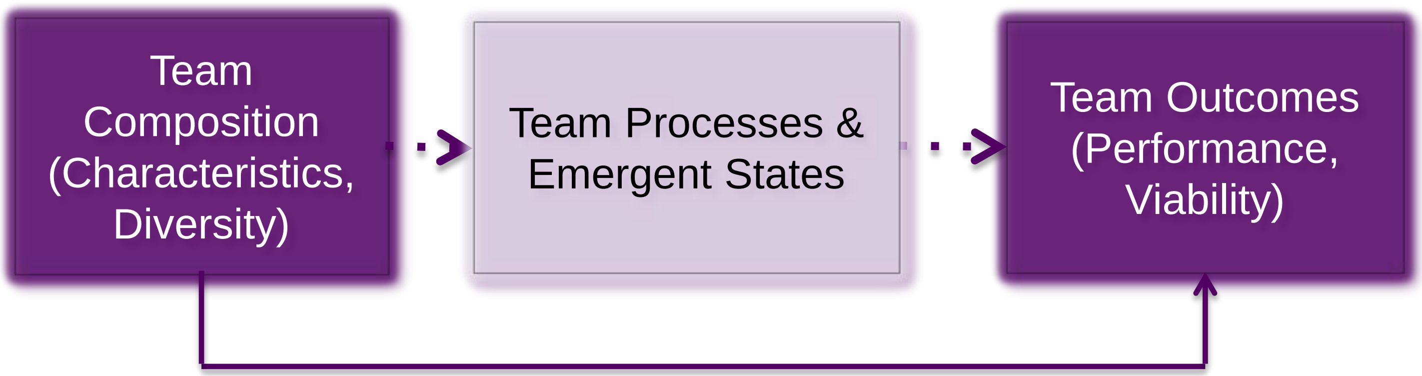
Data from the blogosphere. Shown is a link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number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

Published by AAAS

NORTHWESTERN UNIVERSITY

advancing the science of networks in communities

Current Model for Team Assembly (aka Staffing)



Key takeaway:

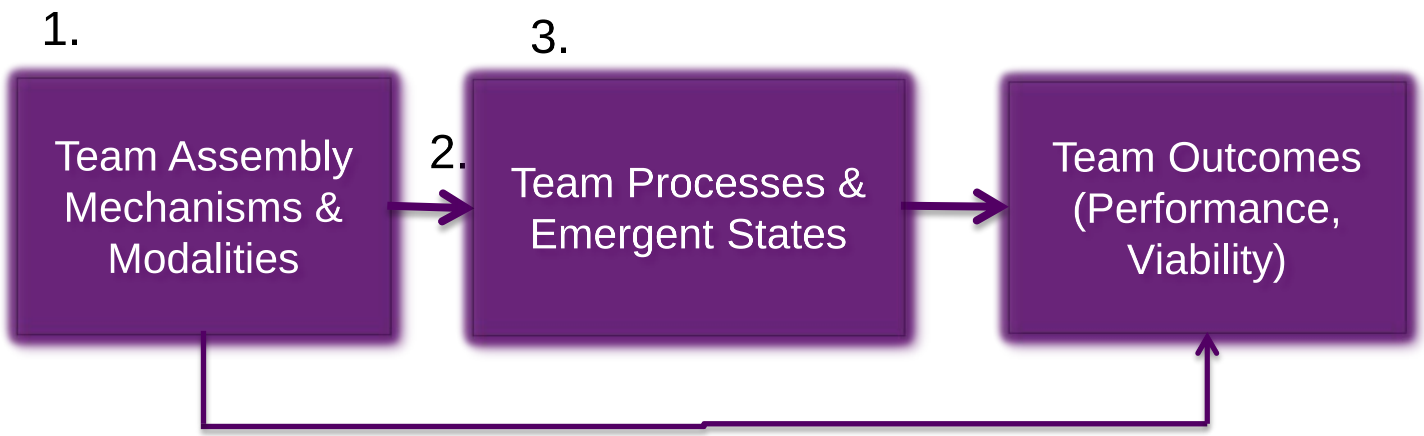
Put smart people in a team, they tend to perform better

Table. Summary of Published Meta-Analyses Relating Team Composition/Diversity to Team Performance

Source: Wax, A. (2013). *Impact of Social and Informational Faultlines on Patterns of Trust and Coordination in Teams*. Unpublished Masters Thesis, Georgia Institute of Technology.

Meta-Analysis	*k*	Predictor(s)	Criteria	Effect Size(s)
Bowers, Pharmer, & Salas, 2000	57	Team composition (ability, personality, and gender)	Performance	None
Devine & Philips, 2001	24	Cognitive ability mean (lab/field moderator)	Performance	.29
Webber & Donahue, 2001	37	Diversity	Cohesion, Performance	None
Peeters, van Tuijl, Rutte, & Reyman, 2006	9	Extraversion	Performance	.04
	10	Conscientiousness	Performance	.21
	9	Emotional stability	Performance	.04
	6	Extraversion variability	Performance	.06
	6	Agreeableness variability	Performance	-.12
	6	Conscientiousness variability	Performance	-.24
	4	Openness to experience variability	Performance	-.01
Stewart, 2006	38	Aggregate (personality, cognitive ability, expertise)	Performance	.27
	20	Personality	Performance	.26
	10	Cognitive ability	Performance	.40
	14	Expertise	Performance	.16
	26	Team size	Performance	.04
	38	Extraversion	Performance	.09
Horowitz & Horowitz, 2007	15	Task-related diversity	Performance	.13
Bell, Villado, Lukasik, Belau, & Briggs, 2011	31	Functional background	Performance	.10
	17	Organizational tenure mean	Performance	.08
	15	Team tenure mean	Performance	.09
	31	Race	Performance	-.11
	38	Sex	Performance	-.06
Roth, Purvis, & Bobko, 2012	61	Gender Differences	Performance	None

Current Model for Team Assembly (aka Staffing): Three Deficiencies



~~Team Composition (Characteristics, Diversity)~~

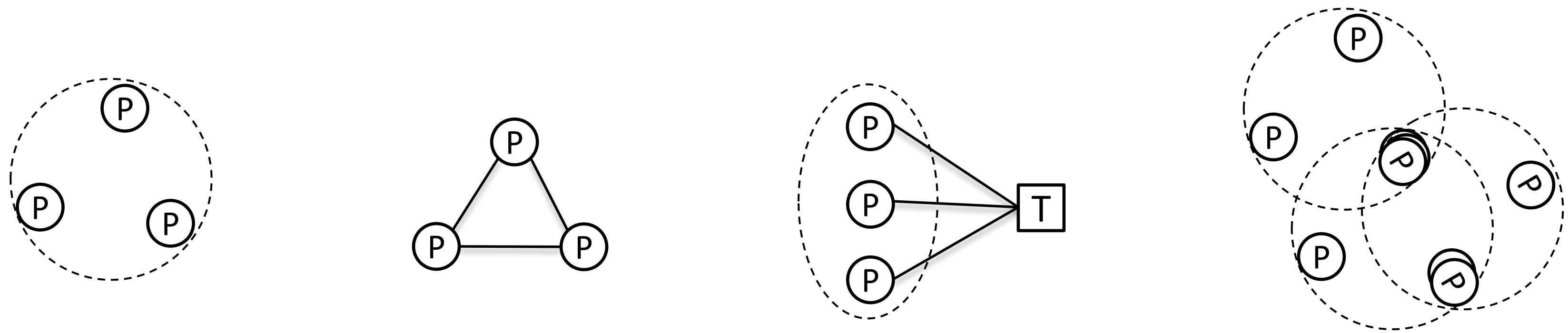
1. Individual Inputs (cka Team Composition) are much more than "combinations of characteristics" – **Team Assembly**
2. Team Assembly affects processes, states, and performance – **Model the Mechanisms**
3. Detecting effects on team processes & states requires relational level analysis – **Relational Level**

Two Dimension of Team Assembly Modalities

	Teams are self-organized	Teams are assigned
Structured Information	Data-driven Self-Organization	Data-driven Assignment
Unstructured Information	*Team/Crowd Science* Ad hoc Self-Organization	Ad hoc Assignment

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

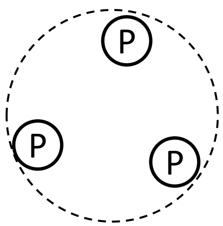
(d) Ecosystem of teams

P Individual

T Task

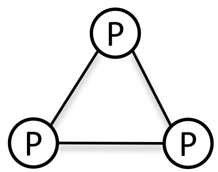
Four Levels of Influences on Team Assembly

Compositional Level



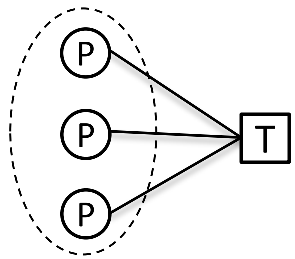
(a) Team as a collection of individuals

Relational Level



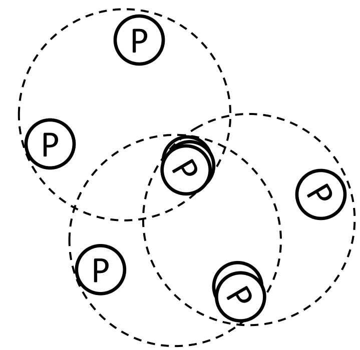
(b) Team as individuals and relations

Multimodal Network Level



(c) Team as a network of individuals and tasks

Ecosystem Level



(d) Ecosystem of teams

P Individual

T Task

NORTHWESTERN UNIVERSITY

SONIC

advancing the science of networks in communities

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

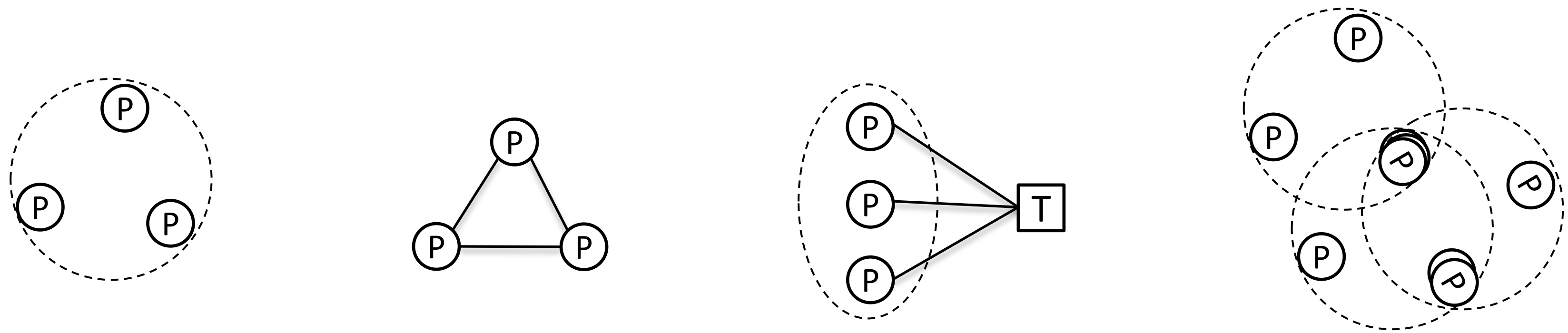
Note: * p<.10, ** p<.05, *** p<.01

NORTHWESTERN UNIVERSITY

advancing the science of networks in communities

Four Levels of Influence 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

P Individual

T Task

Relational Influences on Software Development Teams:

nanoHUB
online simulation and more
an NCN project

- 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

NORTHWESTERN UNIVERSITY

SONIC
advancing the science of networks in communities

Relational Influences on nanoHUB Team Assembly

Co-contribution in …	Successful Teams (>250 users)	Unsuccessful 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

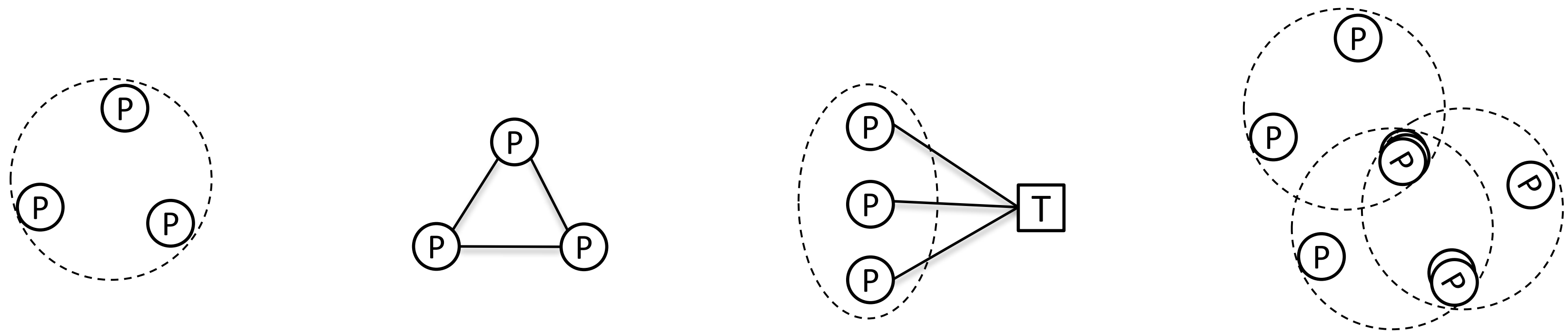
Note: * p<.10, ** p<.05, *** p<.01

NORTHWESTERN UNIVERSITY

advancing the science of networks in communities

Four Levels of Influence 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

P Individual

T Task

NORTHWESTERN UNIVERSITY

SONIC

advancing the science of networks in communities

Multimodal Influences on Team Assembly: nanoHUB

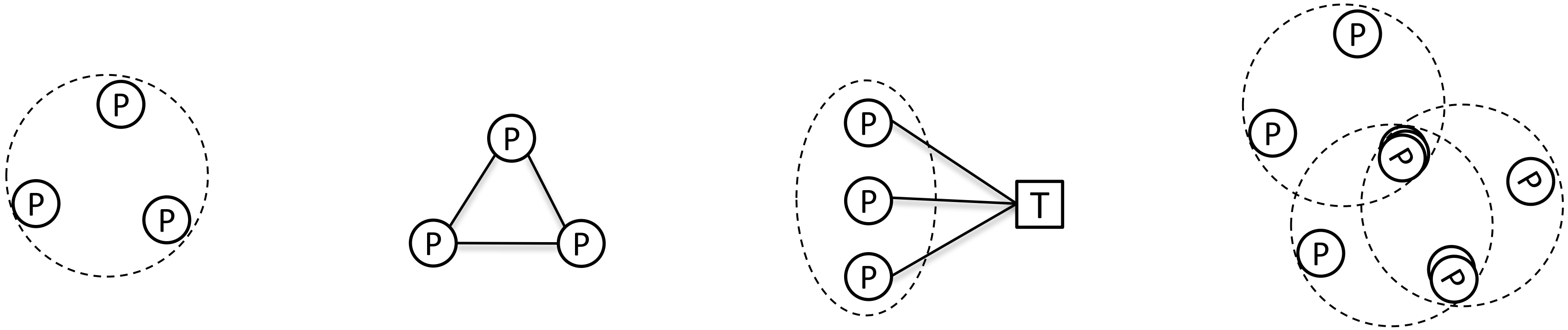
- Outcome variables
  - Team affiliation network(s)
- Explanatory variables
  - Contributor attributes
  - Team attributes
  - Network structures
  - Positions in co-authorship and citation networks
- Methods: p*/BPnet

Multimodal Influences on Assembly of Software Development Teams: nanoHUB

	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 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

P Individual

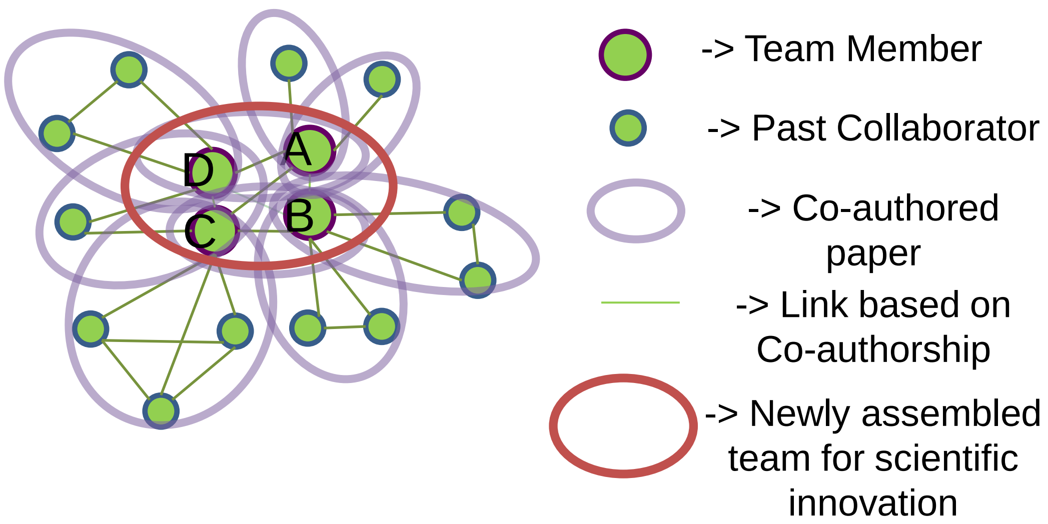
T Task

NORTHWESTERN UNIVERSITY

SONIC
advancing the science of networks in communities

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"



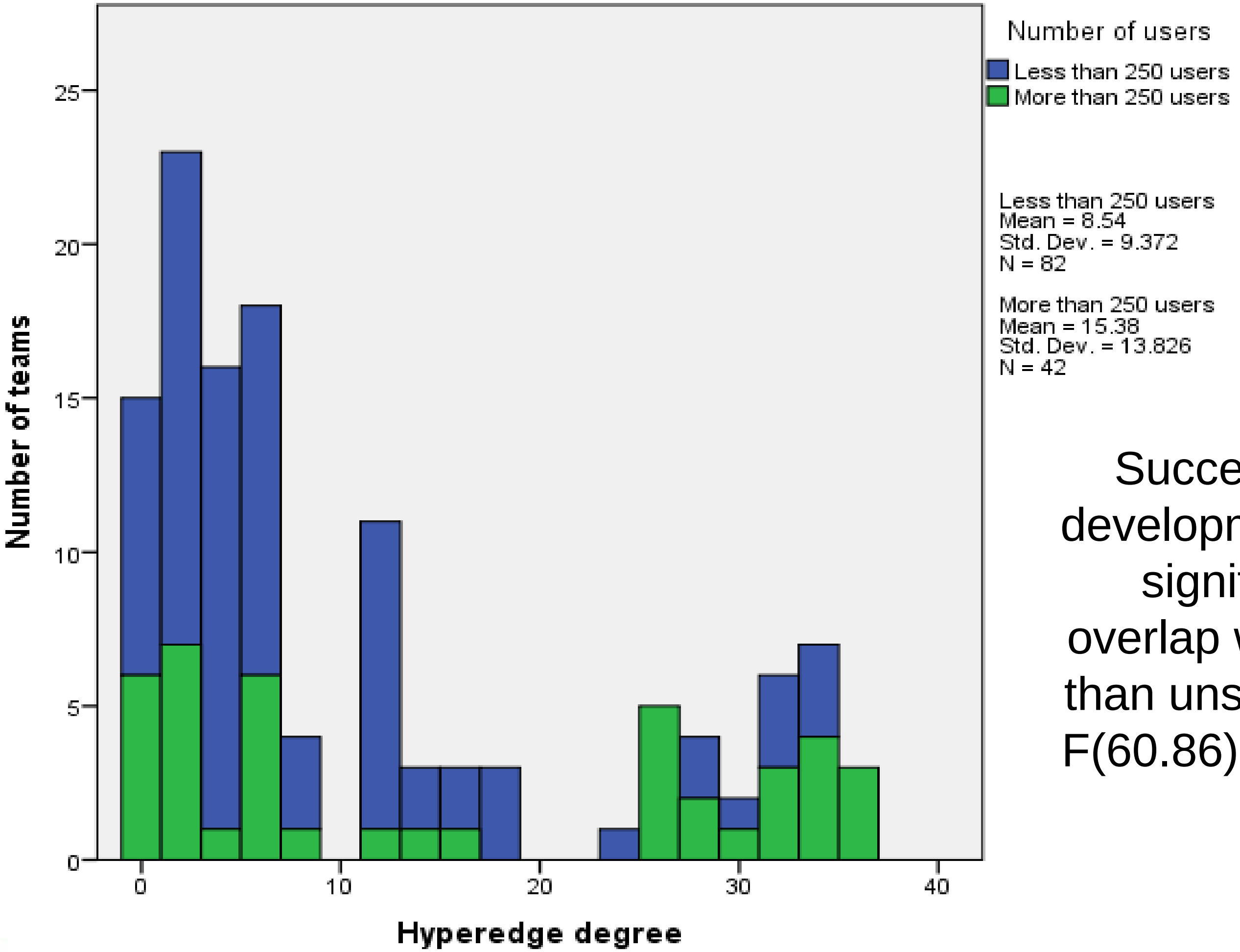
- Are there certain characteristics of the scientific ecosystem that lead to team assembly?
- Do variations in these ecosystem characteristics predict team performance?

NORTHWESTERN UNIVERSITY

SONIC

advancing the science of networks in communities

Ecosystem influence on nanoHUB Team Assembly



Successful software development teams have significantly more overlap with other teams than unsuccessful teams $F(60.86)=-2.89$, $p=0.005$.

Demo

- [Intra-university Research Networking: NUCATS Semantic C-IKNOW]