Noshir Contractor
Jane S. & William J. White Professor of Behavioral Sciences
Professor of Ind. Engg & Mgmt Sciences, McCormick School of Engineering
Professor of Communication Studies, School of Communication &
Professor of Management & Organizations, Kellogg School of Management,
Director, Science of Networks in Communities (SONIC) Research Laboratory
nosh@northwestern.edu

Supported by NSF: IIS-0838564, OCI-0753047, IIS-0729505, IIS-0535214
Aphorisms about Networks

- **Social Networks:**
  - Its not what you know, its *who* you know.

- **Cognitive Social Networks:**
  - Its not who you know, its *who they think* you know.

- **Knowledge Networks:**
  - Its not who you know, its *what they think* you know.
Cognitive Knowledge Networks

It's not who you know.
It's what who you know knows.

Well Connected  MORGAN STANLEY  DEAN WITTER
Multidimensional Networks in Team Science
Multiple types of Nodes and Multiple Types of Relationships
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
Digital Harvesting of Relational Metadata

- Bios, titles & descriptions
- Personal Web sites Google search results
- Web of Science Citation

CATPAC

uberlink

UBERLINK

IKNOW

CI-KNOW Analyses and Visualizations
SOCIAL SCIENCE

Computational Social Science

David Lazer,1 Alex Pentland,2 Lada Adamic,3 Sean Aral,4 Albert-Laszlo Barabasi,5
Deen Brown,6 Nicholas Christakis,7 Nolan Contractor,8 James Fowler,9 Myron Gutmann,9
Tony Jebara,9 Gary King,1 Michael Macy,61 Deb Roy,2 Marshall Van Alstyne2

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 transpor-
tation, 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 mass-

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 econo-

mics, sociology, and political science show

little evidence of this field. But computational
social science is occurring—in Internet compa-
nies such as Google and Yahoo, and in govern-
ment agencies such as the U.S. National Secu-

rity Agency. Computational social science could
become the exclusive domain of private com-
panies and government agencies. Alternatively,
there might emerge a privileged set of aca-
demic researchers possessing ever more data
from which they produce papers that cannot be
criticized or replicated. Neither scenario will
serve the long-term public interest of accumu-

lating, verifying, and disseminating knowledge.

What value might a computational social
science—based in an open academic environ-
ment—offer society, by enhancing understand-
ing 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 node reflects the number of other blogs that
link to it. [Reproduced from (8) with permission from the Association for Computing Machinery]
Projects Investigating Social Drivers for Teams

Science Applications

CI-Scope: Understanding & Enabling CI in Virtual Communities (NSF)

NUCATS: Clinical & Translational Science (NIH)

VOSS: NanoHub (NSF)

TSEEN: Tobacco Surveillance Evaluation & Epidemiology Network (NSF, NIH, CDC)

Business Applications

PackEdge Community of Practice (P&G)

Kraft Design Teams

Core Research

Socio-technical Drivers for Understanding & Enabling Teams

Societal Justice Applications

Mapping Climate Change Networks In Low Income Communities (City of Chicago)

Mapping Digital Media and Learning Networks (MacArthur Foundation)

Entertainment Applications

Second Life (NSF, Army Research Institute, Linden Labs)

EverQuest II (NSF, Army Research Institute, Linden Labs)

SONIC

Advancing the Science of Networks in Communities
The Assembly of Task-oriented Groups

Yun Huang, Mengxiao Zhu, Jing Wang, Brian Keegan & Noshir Contractor, Northwestern University

Nishith Pathak
University of Minnesota

Cuihua Shen, Dmitri Williams
University of Southern California

Supported by NSF IIS-0729505, Army Research Institute (W91WAW-08-C-0106), and Sony Online Entertainment
Using Digital Traces to Understand Team Assembly

- Massively-multiplayer online games (MMOGs) have over 45 million users worldwide and over $3 billion in revenue in 2008

- What does social behavior in online worlds tell us about the “real” world and vice versa?
  - Online games exhibit features that map onto real world processes:
    - Social networks, economics, groups, communication, conflict, expertise, leadership, crime, innovation, epidemics, etc.
  - Online games already capture the signatures of these behaviors in huge databases, just waiting to be analyzed
Hypotheses

Team formation mechanisms

H1: Players who have low combat ability are more likely to participate in teams than those who have high combat ability. (*Self-interest*)

H2: Players are more likely to join the same set of players multiple times. (*Reduce Coordination cost*)

H3a: Players are more likely to join teams of high expertise diversity. (*Transactive Memory*)

H3b: Players are more likely to join teams in which they can provide unique expertise. (*Transactive Memory*)
Hypotheses (cont.)

- Group outcome

- H4: Teams with many players are more likely to have member death. *(Higher Coordination cost)*

- H5: Teams with many players tend to have higher performance. *(Mutual interest)*

- H6: Teams with many players have shorter duration. *(Higher Coordination Cost)*
# Team Formation Structures

<table>
<thead>
<tr>
<th>Team membership</th>
<th>Selective team membership</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Membership in multiple teams</th>
<th>Teams with many members</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3" alt="Diagram" /></td>
<td><img src="image4" alt="Diagram" /></td>
</tr>
</tbody>
</table>

**Notes:**
- Characters whose attributes are **not** taken into consideration
- Characters whose attributes are taken into consideration
Zone Antonica
Character (size: level)
Group (size: performance)
16
### Descriptive Statistics for the Zones

- In the whole dataset, there are 2,774 characters, 3,547 group events; 15,152 group membership links.
- Divide it into 11 zones based on the game map

<table>
<thead>
<tr>
<th>Zone #</th>
<th>Zone Name</th>
<th># of char</th>
<th># of group</th>
<th>Median Level</th>
<th>Mean Group Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Thundering Steppes</td>
<td>639</td>
<td>591</td>
<td>29</td>
<td>4.15</td>
</tr>
<tr>
<td>2</td>
<td>Kingdom of Sky</td>
<td>625</td>
<td>436</td>
<td>65</td>
<td>4.80</td>
</tr>
<tr>
<td>3</td>
<td>The Enchanted Lands</td>
<td>530</td>
<td>537</td>
<td>38</td>
<td>4.48</td>
</tr>
<tr>
<td>4</td>
<td>Desert of Flames</td>
<td>499</td>
<td>518</td>
<td>53</td>
<td>4.36</td>
</tr>
<tr>
<td>5</td>
<td>Antonica</td>
<td>465</td>
<td>396</td>
<td>21</td>
<td>4.04</td>
</tr>
<tr>
<td>6</td>
<td>Commonlands</td>
<td>380</td>
<td>315</td>
<td>24</td>
<td>4.01</td>
</tr>
<tr>
<td>7</td>
<td>Nektulos Forest</td>
<td>287</td>
<td>161</td>
<td>36</td>
<td>3.92</td>
</tr>
<tr>
<td>8</td>
<td>Feerrott</td>
<td>269</td>
<td>206</td>
<td>45</td>
<td>4.45</td>
</tr>
<tr>
<td>9</td>
<td>Everfrost</td>
<td>211</td>
<td>165</td>
<td>45</td>
<td>4.36</td>
</tr>
<tr>
<td>10</td>
<td>Lavastorm</td>
<td>198</td>
<td>141</td>
<td>49</td>
<td>4.51</td>
</tr>
<tr>
<td>11</td>
<td>Zek</td>
<td>170</td>
<td>81</td>
<td>40</td>
<td>3.95</td>
</tr>
</tbody>
</table>
### Results: Antonica as An Example

<table>
<thead>
<tr>
<th>Findings</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low level players are more likely to join groups. (H1: Supported)</td>
<td>-0.01*</td>
</tr>
<tr>
<td>Players are more likely to join the same set of players for multiple times. (H2: Not supported)</td>
<td>-0.11</td>
</tr>
<tr>
<td>Players are more likely to join groups of high expertise diversity. (H3a: Supported)</td>
<td>4.24*</td>
</tr>
<tr>
<td>Players are more likely to join groups in which they can provide unique expertise (H3b: Partially supported)</td>
<td>-1.27* (Priest)</td>
</tr>
<tr>
<td>Supported for priests but the other character classes do not show such a tendency.</td>
<td>0.03 (Mage)</td>
</tr>
<tr>
<td>-0.07 (Scout)</td>
<td></td>
</tr>
<tr>
<td>Groups of larger size are more likely to have member death. (H4: Supported)</td>
<td>0.60*</td>
</tr>
</tbody>
</table>

**Notes:**
* indicate twice of standard deviation

**Green** indicates results supporting the hypotheses; **black** indicates non-significant results; **red** indicates results in the opposite direction.
## Results: Antonica as An Example

<table>
<thead>
<tr>
<th>Findings</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups with many players gain higher performance. (H5: Supported)</td>
<td>0.005*</td>
</tr>
<tr>
<td>Groups with many players have shorter duration. (H6: Supported)</td>
<td>-0.33*</td>
</tr>
<tr>
<td>Players are active in joining groups.</td>
<td>5.00*</td>
</tr>
<tr>
<td>Players tend to join multiple groups (or group events).</td>
<td>0.79*</td>
</tr>
<tr>
<td>Combat groups tend to be small.</td>
<td>-7.44*</td>
</tr>
<tr>
<td>Compared to fighters, priests are more likely to join a group, but mages or scouts are not.</td>
<td>0.92* (priest)</td>
</tr>
<tr>
<td></td>
<td>-0.05 (mage)</td>
</tr>
<tr>
<td></td>
<td>0.004 (scout)</td>
</tr>
</tbody>
</table>

**Notes:** * indicate twice of standard deviation
**Green** indicates results supporting the hypotheses; **Black** indicates non-significant results; **Red** indicates results in the opposite direction.
Results Summary

- Players are active in joining groups, especially those at lower levels.
- Players are more likely to join the groups that 1) have higher expertise diversity and 2) to which they can provide unique expertise (especially for priest and mage).
- Groups with more members tend to 1) have higher performance, 2) last a shorter time, and 3) be more likely to have member death during the combat.
- Players tend to join multiple groups, and most groups are of small size.
Projects Investigating Social Drivers for Teams

**Science Applications**
- CI-Scopes: Understanding & Enabling CI in Virtual Communities (NSF)
- NUCATS: Clinical & Translational Science (NIH)
- VOSS: NanoHub (NSF)
- TSEEN: Tobacco Surveillance Evaluation & Epidemiology Network (NSF, NIH, CDC)

**Business Applications**
- PackEdge Community of Practice (P&G)
- Kraft Design Teams

**Societal Justice Applications**
- Mapping Climate Change Networks in Low Income Communities (City of Chicago)
- Mapping Digital Media and Learning Networks (MacArthur Foundation)

**Entertainment Applications**
- Second Life (NSF, Army Research Institute, Linden Labs)
  - EverQuest II (NSF, Army Research Institute, Linden Labs)

**Core Research**
Socio-technical Drivers for Understanding & Enabling Teams
The Impacts of Co-authorship Networks and Citation Networks in “Team Science”*

By Meikuan Huang, Jordan Liu, Annie Wang, & Noshir Contractor

“Group-staffing riddle” (Huber & Lewis, 2010):

How to assemble a group to obtain both
(1) high productivity based on diversity of expertise and cognitive models &
(2) smooth coordination and communication among group members with shared cognitive models

Our goal: To discover how prior co-authorship and citation network configurations influence team formation and success in scientific research groups.

Theoretical Background

(1) Transactive memory (TM)
- A key TM dimension: Sharedness of knowledge at the group level, or the extent to which all members have similar perceptions of each other’s task responsibilities and expertise level in different knowledge areas (Brandon & Hollingshead, 2004; Huber & Lewis, 2010).

(2) Prior collaboration
- People are likely to prefer partners with whom they are already familiar from prior work on joint projects (Hinds, Carley, Krackhardt, & Wholey, 2000).

(3) Homophily
- The tendency of individuals to interact more with those to whom they are more similar (Ibarra, 1992; McPherson & Smith-Lovin, 1987).
- Reasons: Ease of communication, shared understandings and comfort (Carley, 2002).
## Hypotheses & Analysis

<table>
<thead>
<tr>
<th></th>
<th>Hypothesis (H)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>Co-authorship</td>
<td>Researchers tend to collaborate on proposal teams with those with whom they have a co-authorship relationship.</td>
</tr>
<tr>
<td>H2</td>
<td>Co-citing</td>
<td>Researchers tend to collaborate on proposal teams with those with whom they have a citation relationship.</td>
</tr>
<tr>
<td>H3</td>
<td>Co-citation</td>
<td>Researchers who cite similar publications are more likely to collaborate on proposal teams.</td>
</tr>
</tbody>
</table>

**Analysis:**

- ERGM models (Exponential Random Graph Modeling) (Frank & Strauss, 1986; Robins & Pattison, 2005; Wasserman & Pattison, 1996)
- PNet (Wang, Robins, & Pattison, 2006).
Data

- 60 Proposals
- 117 applicants, with 60 PIs and 57 Co-PIs, totally
- 37 departments in total
Tenure Distribution

- 29 Professors
- 24 Associate Professors
- 28 Assistant Professors
- 14 Research Assistant Professors
- 3 Post Docs
- 3 Others: student, research scientist, adjunct assistant professor
- 101 in data
<table>
<thead>
<tr>
<th>Department</th>
<th>Number of applicants in the department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Medicine and Rehabilitation</td>
<td>9</td>
</tr>
<tr>
<td>Surgery</td>
<td>8</td>
</tr>
<tr>
<td>Biomedical Engineering *</td>
<td>6</td>
</tr>
<tr>
<td>Cardiology</td>
<td>6</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>5</td>
</tr>
<tr>
<td>Chemistry *</td>
<td>4</td>
</tr>
<tr>
<td>Hematology Oncology</td>
<td>4</td>
</tr>
<tr>
<td>Infectious Disease</td>
<td>4</td>
</tr>
<tr>
<td>Molecular Pharmacology</td>
<td>4</td>
</tr>
<tr>
<td>All others</td>
<td>3 or less</td>
</tr>
</tbody>
</table>

* Indicates that the department is outside the medical school.
Applicant Distribution Across Schools

- School of Medicine: 73
- School of Engineering: 17
- College of Arts & Sciences: 10
- School of Communication: 3
Gender Distribution

74 males (72%)
27 females (28%)
Number of Applicants in the Proposal

<table>
<thead>
<tr>
<th>Number of Applicants</th>
<th>Total</th>
<th>Unfunded</th>
<th>Funded</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Total: 60
Unfunded: 55
Funded: 5
H1: Co-proposal & Co-authorship Network

Node size indicates the # of publications
H2: Co-proposal & co-citation network
H3: Co-proposal & Citing network
Researchers are not likely to randomly form a project collaboration relationship with each other.

Researchers are more likely to have better familiarity of and collaborate again with those they share a collaboration history (co-authorship or citing each other).

Researchers are also more likely to collaborate with those who cited similar articles in their publications.
### Funded vs. Unfunded

<table>
<thead>
<tr>
<th>Effects</th>
<th>Funded</th>
<th>SD</th>
<th>Unfunded</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N = 8)</td>
<td>(N = 93)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge (co-proposal)</td>
<td>-3.28</td>
<td>1.07</td>
<td>-4.33</td>
<td>0.13</td>
</tr>
<tr>
<td>Co-author</td>
<td>6.95</td>
<td>7.14</td>
<td>0.34</td>
<td>1.06</td>
</tr>
<tr>
<td>Cite one another</td>
<td>7.32</td>
<td>4.61</td>
<td>-2.93</td>
<td>4.37</td>
</tr>
<tr>
<td>Cite same sources</td>
<td>6.61</td>
<td>7.99</td>
<td>-4.17</td>
<td>15.83</td>
</tr>
</tbody>
</table>
3D Strategy for Enabling Team Science

- **Discovery**: Effectively and efficiently foster network links from people to other people, knowledge, and artifacts (data sets/streams, analytic tools, visualization tools, documents, etc.)
  - “If only NSF knew what NSF knows”.

- **Diagnosis**: Assess the “health” of internal and external networks - in terms of scanning, absorptive capacity, diffusion, robustness, and vulnerability to external environment

- **Design**: Model or re-wire networks using social and organizational incentives (based on social network research) and network referral systems to enhance emergent and mature teams.
Design Examples:
Mapping & Enabling Networks in ...

Tobacco Research: TobIG Demo

Computational Nanotechnology: nanoHUB Demo

Cyberinfrastructure: CI-Scope Demo

Oncofertility: Onco-IKNOW
Summary

- The Science of Team Science is well poised to make a quantum intellectual leap by facilitating collaboration that leverages recent advances in:

  - Theories about the social motivations for creating, maintaining, dissolving and re-creating network ties within teams

  - Developments in cyberinfrastructure and Web 2.0 that provide the technological capability to capture and analyze relational metadata needed to more effectively understand and enable teams.

  - Statistical techniques to make theoretically grounded team assembly recommendations that go beyond the Lovegety and SNIF

  - Petascale computational infrastructure to execute the statistical and optimization algorithms
Acknowledgements
Research Team @ SONIC

Yun Huang                  Annie Wang                          Mengxiao Zhu                     David Huffaker     Brian Keegan
Post-doc                          Post-doc                      IEMS Doctoral candidate      SoC  Doctoral candidate      Doctoral Candidate

Jingling Li               Jeffrey  Treem                  York Yao                     Zack Johnson
Programmer          Doctoral candidate             Programmer                  SoC Undergraduate

Advancing the Science of Networks in Communities