Some Assembly Required

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

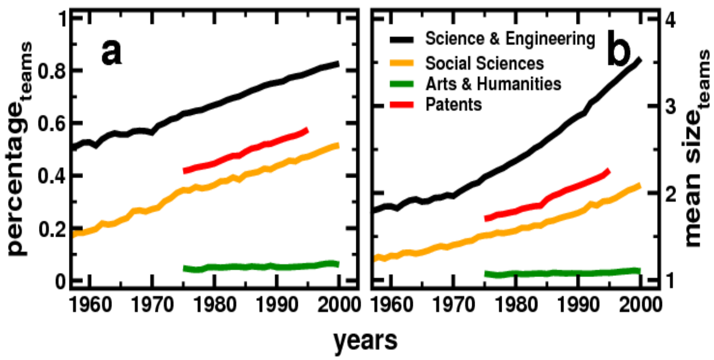
Ascendance of Teams

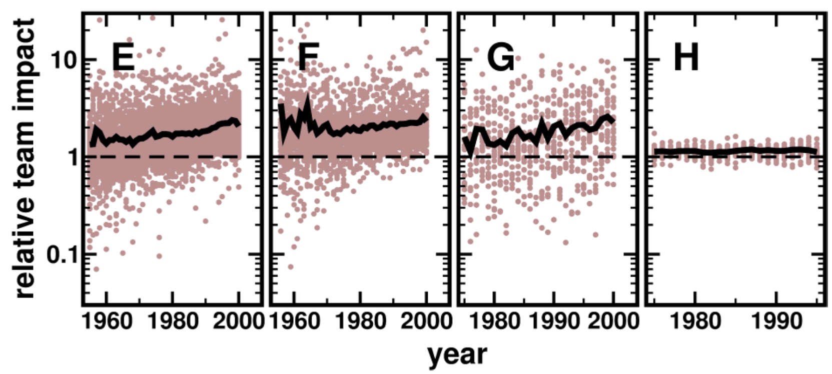
Studies of 19.9 million research articles over 5 decades as recorded in the Web of Science database, and an additional 2.1 million patent records from 1975-2005 found three important facts.

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.





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

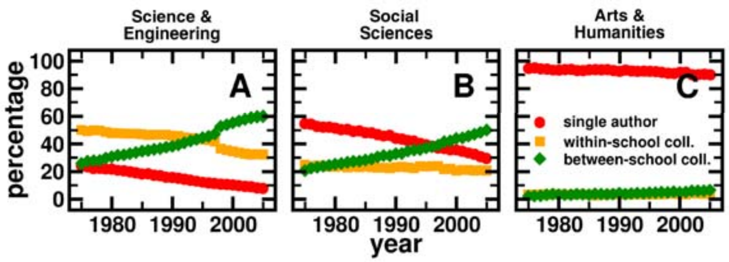
Ascendance of *Virtual* Teams

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



Source: Jones, Wuchty, Uzzi, 2008

Key Takeaways

- Web Science is well poised to make a leap in understanding and enabling team assembly by facilitating recent advances in:
  - Theories: Theories about the social motivations for creating, maintaining, dissolving and re-creating networks
  - 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 networks.
  - Methods: An ensemble of qualitative and quantitative methods (exponential random graph modeling (*p**) techniques to understand and enable theoretically grounded network recommendations.
  - Computational infrastructure: Cloud computing and petascale applications are critical to face the computational challenges in analyzing the data

Multi-theoretical Multilevel (MTML) Motivations for Team Assembly

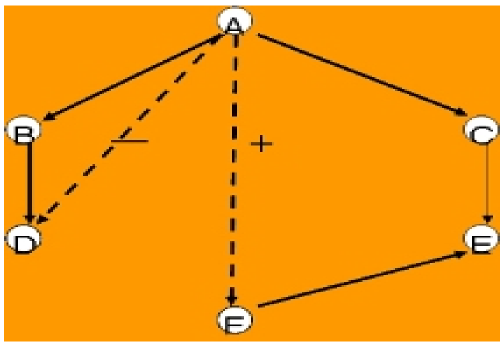
- Theories of self-interest
- Theories of social and resource exchange
- Theories of mutual interest and collective action
- Theories of contagion
- Theories of balance
- Theories of homophily
- Theories of proximity

Sources:

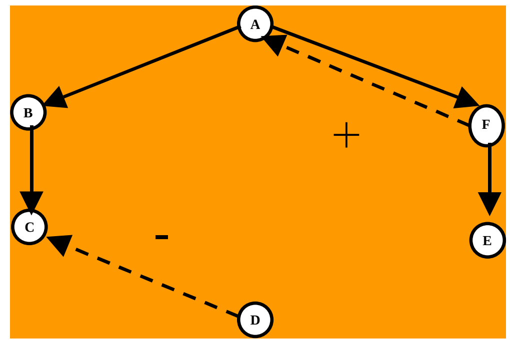
Contractor, N. S., Wasserman, S. & Faust, K. (2006). Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review*.

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

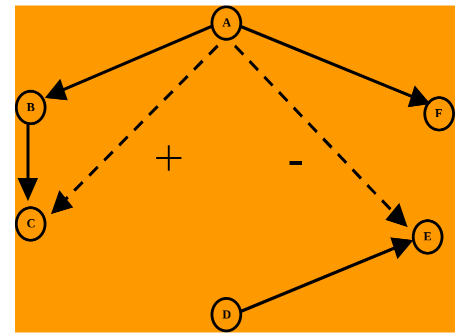
“Structural signatures” of MTML
Motivations for Team Assembly



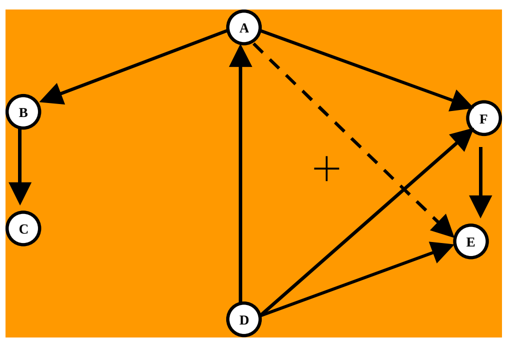
Theories of Self interest



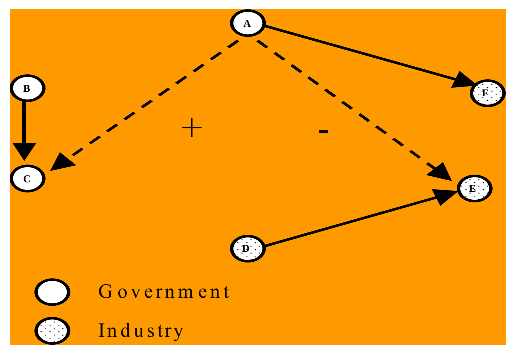
Theories of Exchange



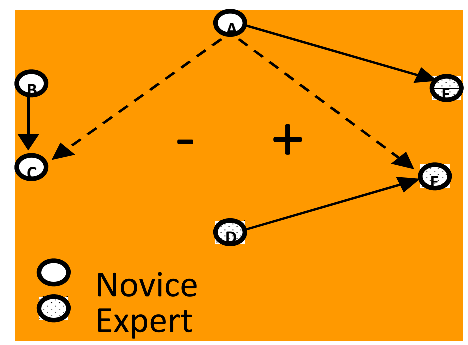
Theories of Balance



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

p*/Exponential Random Graph Models

- Analysis of network data with Interdependencies – endogenous correlation among the relations
- ERGMs are a class of stochastic models:

$$P(\mathbf{X}=\mathbf{x})=\kappa^{-1}\exp\left(\sum_{A\subseteq N_D}\lambda_A z_A(\mathbf{x})\right),$$

where:

- (i) the summation is over structural signatures of types A;
- (ii) λ_A is the parameter corresponding to structural signatures of type A;
- (iii) $z_A(\mathbf{x})$ is the *network statistic* corresponding to structural signature A
- (iv) κ is a normalizing quantity to ensure that (1) is a proper probability distribution

Frank & Strauss, 1986; Pattison & Wasserman, 1999; Robins, Pattison, & Wasserman, 1999; Wasserman & Pattison, 1996; Hunter, 2007; Robins, Snijders, Wang, Handcock, & Pattison, 2007

Challenges of <u>empirically</u> testing, extending, and exploring theories about emergence of networks
… until now

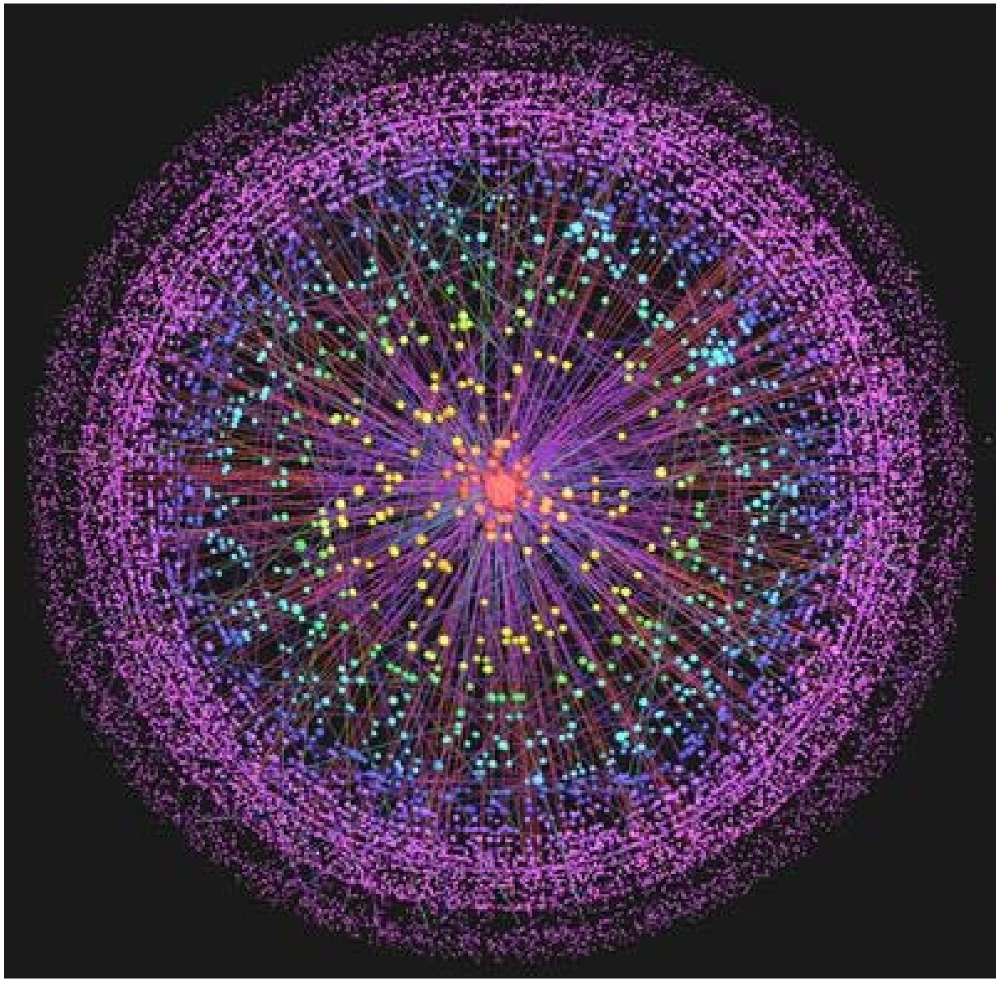
The Hubble telescope: $2.5 billion

Source: David Lazer

CERN particle accelerator: $1 billion/year

Source: David Lazer

The Web: priceless*



Source: David Lazer

* Apologies to MasterCard

NORTHWESTERN UNIVERSITY

SONIC
advancing the
science of networks in communities

Exemplars

- I. Team assembly for interdisciplinary NSF proposals
- II. Virtual World Exploratorium (two exemplars)
- III. Recommender system for enabling team assembly in clinical and translational science

Exemplar I

Team Assembly for Interdisciplinary NSF proposals

with …..

Alina Lungeanu
Ph.D. Candidate

Yun Huang
Postdoctoral Fellow

NORTHWESTERN UNIVERSITY

SONIC
advancing the science of networks in communities

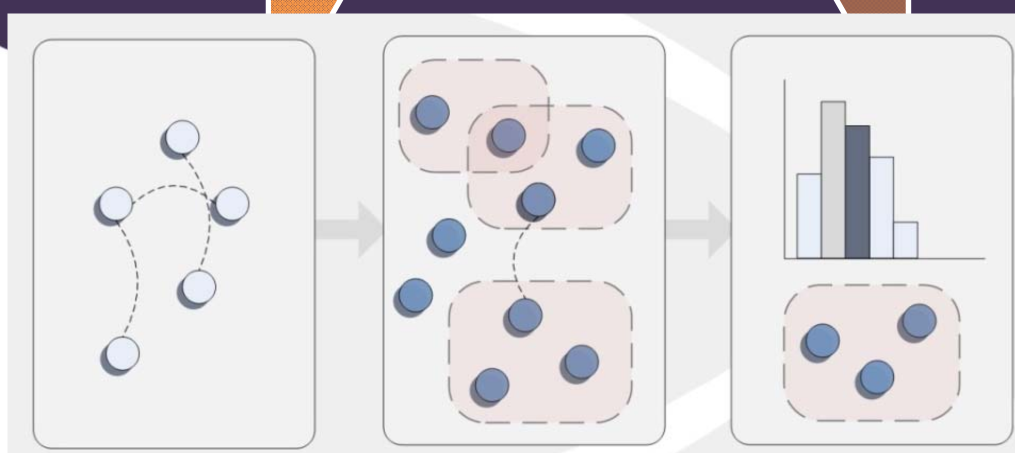
“Group Staffing Riddle”

(Huber & Lewis, 2010)

High productivity based on diversity of expertise and cognitive models

How to assemble a team to obtain both

Smooth coordination and communication among team members with shared cognitive models



Previous collaboration

Current collaboration

Team performance

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

Overview

- Motivation:
  - To better understand the assembly and outcomes of interdisciplinary scientific teams (NSF funding) to improve team success

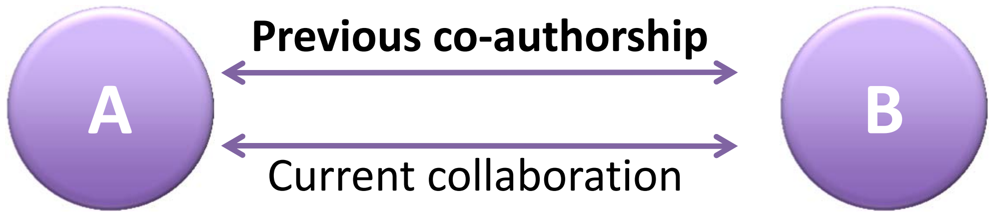
- Research Questions:
  - How do interdisciplinary scientific teams assemble?
  - What are the factors that influence the performance of the interdisciplinary scientific teams?

Data Set

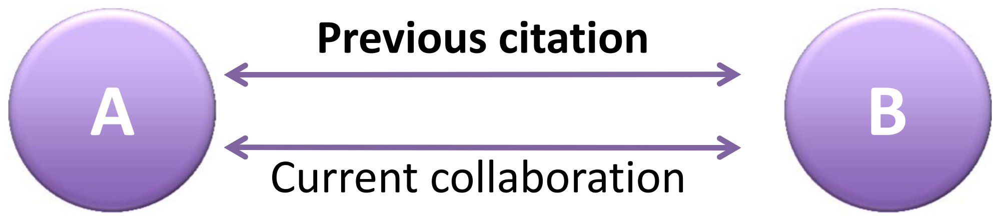
- 1,103 grant proposals submitted to NSF (both awarded and un-awarded)
- 2 interdisciplinary programs
- 3-year period
- 2,186 PIs and Co-PIs

Hypotheses

→ **H1**



→ **H2**



Model for Understanding Team Assembly

	Effects	Full model (N=2,186)
Control	Edge (DV: co-proposal)	-6.751*
Control	Isolates	5.447*
Control	Joining large teams	4.623*
Control	Institution Tier *Reference: low-tier university*	-0.098*
Control	Gender *Reference: male*	0.021
Control	Tenure	0.002*
Control	H-index	-0.014*
H1	Co-authorship	2.431*
H2	Citation relation	1.132 *

*: $p<0.05$

Model for Understanding T

Researchers are not likely to randomly form a project collaboration relationship with each other. The numbers of single author proposals and big research teams are larger than random chances.

	Effects	Full mo (N=2,18
Control	Edge (DV: co-proposal)	-6.751*
Control	Isolates	5.447*
Control	Joining large teams	4.623*
Control	Institution Tier *Reference: low-tier university*	-0.098*
Control	Gender *Reference: male*	0.021
Control	Tenure	0.002*
Control	H-index	-0.014*
H1	Co-authorship	2.431*
H2	Citation relation	1.132 *

*: $p<0.05$

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

Model for Understanding Team Assembly

	Effects	Full model (N=2,186)
Control	Edge (DV: co-proposal)	-6.751*
Control	Isolates	5.447*
Control	Joining large teams	4.623*
Control	Institution Tier *Reference: low-tier university*	-0.098*
Control	Gender *Reference: male*	0.021
Control	Tenure	0.002*
Control	H-index	-0.014*
H1	Co-authorship	2.431*
H2	Citation relation	1.132 *

Researchers from top-tier institutions are less likely to collaborate.

*: $p<0.05$

Model for Understanding Team Assembly

	Effects	Full model (N=2,186)
Control	Edge (DV: co-proposal)	-6.751*
Control	Isolates	5.447*
Control	Joining large teams	4.623*
Control	Institution Tier *Reference: low-tier university*	-0.098*
Control	Gender *Reference: male*	0.021
Control	Tenure	0.002*
Control	H-index	-0.014*
H1	Co-authorship	2.431*
H2	Citation relation	1.132 *

Researchers with high tenure are more likely to collaborate .

*: p<0.05

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

Model for Understanding Team Assembly

	Effects	Full model (N=2,186)
Control	Edge (DV: co-proposal)	-6.751*
Control	Isolates	5.447*
Control	Joining large teams	4.623*
Control	Institution Tier *Reference: low-tier university*	-0.098*
Control	Gender *Reference: male*	0.021
Control	Tenure	0.002*
Control	H-index	-0.014*
H1	Co-authorship	2.431*
H2	Citation relation	1.132 *

Researchers with a high H-index are less likely to collaborate.

*: p<0.05

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

Model for Understanding Team Assembly

	Effects	Full model (N=2,186)
Control	Edge (DV: co-proposal)	-6.751*
Control	Isolates	5.447*
Control	Joining large teams	4.623*
Control	Institution Tier *Reference: low-tier university*	-0.098*
Control	Gender *Reference: male*	0.021
Control	Tenure	0.002*
Control	H-index	-0.014*
H1	Co-authorship	2.431*
H2	Citation relation	1.132 *

*: p<0.05

Researchers are more likely to collaborate with those with whom they have co-authored or with those they cite.

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

Comparing Funded & Unfunded proposals

	Effects	Full model (N=2,186)	Funded model (N=445)	Un-funded model (N=1,834)
Control	Edge (DV: co-proposal)	-6.751*	-5.341*	-6.571*
Control	Isolates	5.447*	10.138*	4.477*
Control	Joining large teams	4.623*	8.908*	3.779*
Control	Institution Tier *Reference: low-tier university*	-0.098*	-0.098*	-0.104*
Control	Gender *Reference: male*	0.021	0.119*	-0.009
Control	Tenure	0.002*	0.001	0.002*
Control	H-index	-0.014*	-0.005*	-0.009*
H1	Co-authorship	2.431*	1.386*	0.914*
H2	Citation relation	1.132 *	-0.147 *	-0.008

*: p<0.05

Comparing Funded & Unfunded proposals

	Effects	Full model (N=2,186)	Funded model (N=445)	Un-funded model (N=1,834)
Control	Edge (DV: co-proposal)	-6.751*	-5.341*	-6.571*
Control	Isolates	5.447*	[illegible]	[illegible]
Control	Joining large teams	4.623*	[illegible]	[illegible]
Control	Institution Tier *Reference: low-tier university*	-0.098*	-0.098*	[illegible].104*
Control	Gender *Reference: male*	0.021	0.119*	-0.009
Control	Tenure	0.002*	0.001	0.002*
Control	H-index	-0.014*	-0.005*	-0.009*
H1	Co-authorship	2.431*	1.386*	0.914*
H2	Citation relation	1.132 *	-0.147 *	-0.008

Females are more likely to collaborate on awarded proposals!

*: p<0.05

Comparing Funded & Unfunded proposals

	Effects	Full model (N=2,186)	Funded model (N=445)	Un-funded model (N=1,834)
Control	Edge (DV: co-proposal)	-6.751*	-5.341*	-6.571*
Control	Isolates	5.447*		
Control	Joining large teams	4.623*		
Control	Institution Tier *Reference: low-tier university*	-0.098*		
Control	Gender *Reference: male*	0.021	…19*	-0.009
Control	Tenure	0.002*	0.001	0.002*
Control	H-index	-0.014*	-0.005*	-0.009*
H1	Co-authorship	2.431*	1.386*	0.914*
H2	Citation relation	1.132 *	-0.147 *	-0.008

The odds of collaborating with a previous co-author on an awarded proposal is 4 times more than collaborating with someone else.

*: p<0.05

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

Comparing Funded & Unfunded proposals

	Effects	Full model (N=2,186)	Funded model (N=445)	Un-funded model (N=1,834)
Control	Edge (DV: co-proposal)	-6.751*	[illegible]	[illegible]
Control	Isolates	5.447*	[illegible]	[illegible]
Control	Joining large teams	4.623*	[illegible]	[illegible]
Control	Institution Tier *Reference: low-tier university*	-0.098*	[illegible]	[illegible]
Control	Gender *Reference: male*	0.021	[illegible]	[illegible]
Control	Tenure	0.002*	0.001	[illegible]2*
Control	H-index	-0.014*	-0.005*	-0.0[illegible]9*
H1	Co-authorship	2.431*	1.386*	0.914*
H2	Citation relation	1.132 *	-0.147 *	-0.008

The odds of collaborating with a previous co-author on <u>awarded</u> proposal is 4 times more than collaborating with someone else.

The odds of collaborating with a previous co-author on an un-awarded proposal is only 2.5 times more.

*: p<0.05

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

Comparing Funded & Unfunded proposals

	Effects	Full model (N=2,186)	Funded model (N=445)	Un-funded model (N=1,834)
Control	Edge (DV: co-proposal)	-6.751*	-5.341*	-6.571*
Control	Isolates	5.447*	10.138*	4.477*
Control	Joining large teams	4.623*	8.908*	3.779*
Control	Institution Tier ↓	[illegible]	[illegible]	[illegible]
Control	[illegible]	[illegible]	[illegible]	[illegible]
Control	[illegible]	[illegible]	[illegible]	[illegible]
Control	H-index	-0.014*	-0.005*	-0.009*
H1	Co-authorship	2.431*	1.386*	0.914*
H2	Citation relation	1.132 *	-0.147 *	-0.008

The odds of collaborating with someone that you cite are 3 times more than collaborating with someone else.

However, the odds of collaborating with someone that you cite on an awarded proposal is 0.14 times less!

*: $p<0.05$

NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

Comparing Funded & Unfunded proposals

	Effects	Full model (N=2,1	Funded model	Un-funded model
Control	Edge (DV: co-proposal)	-6.75		
Control	Isolates	5.44		
Control	Joining large teams	4.62		
Control	Institution Tier *Reference: low-tier university*	-0.09		
Control	Gender *Reference: male*	0.021	.119*	-0.009
Control	Tenure	0.002*	0.001	0.002*
Control	H-index	-0.014*	-0.005*	-0.009*
H1	Co-authorship	2.431*	1.386*	0.914*
H2	Citation relation	1.132 *	-0.147 *	-0.008

Therefore, researchers are more likely to collaborate on awarded proposals if they have collaborated before (co-authors) and if they come from different research areas (not citing each other).

*: p<0.05

Exemplar II.1 & II.2
Motivations for creating teams in massively multiplayer online games

NORTHWESTERN UNIVERSITY

SONIC
advancing the science of networks in communities

Virtual World Exploratorium

vwobservatory.org

NSF

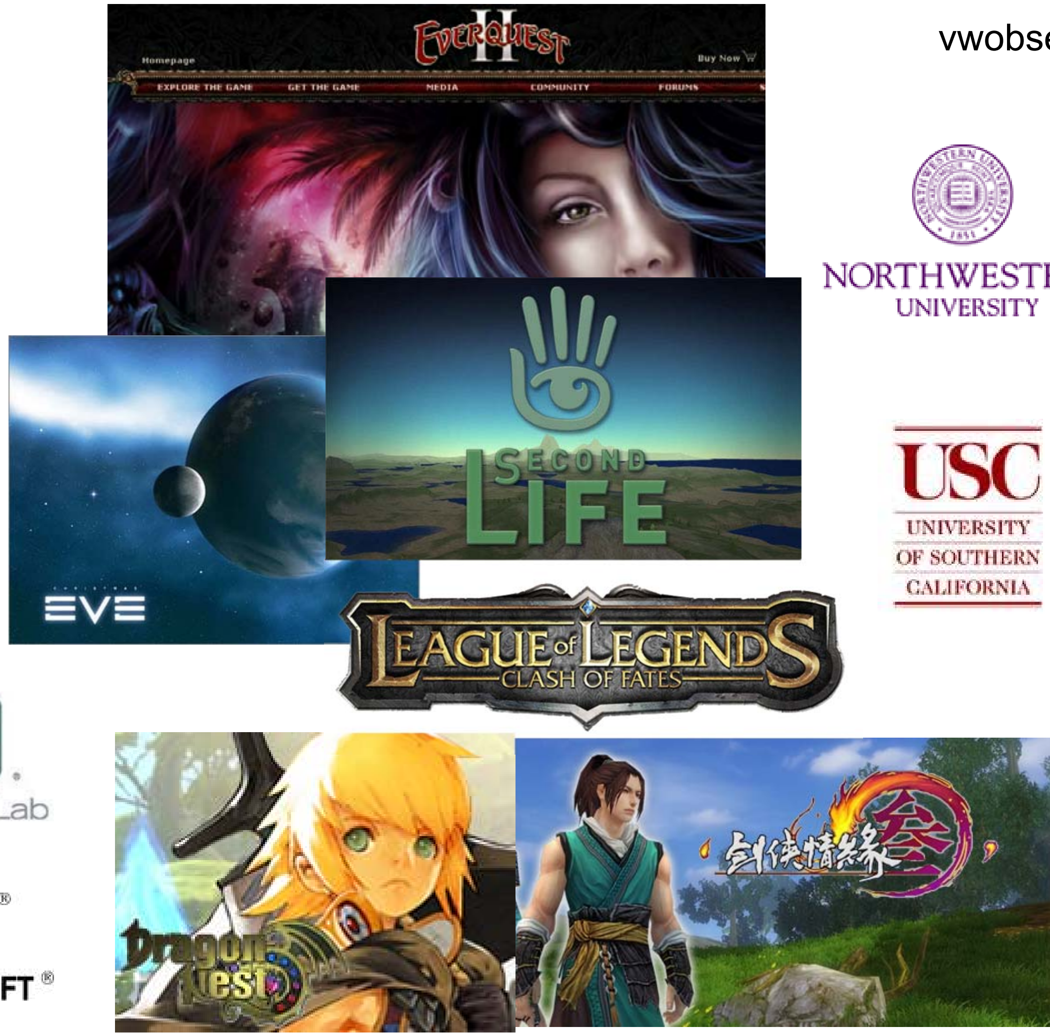
AIR FORCE RESEARCH LABORATORY

SONY ONLINE ENTERTAINMENT

Linden Lab

SHANDA GAMES

KINGSOFT



NORTHWESTERN UNIVERSITY

REGENTS OF THE UNIVERSITY OF MINNESOTA

USC UNIVERSITY OF SOUTHERN CALIFORNIA

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

FUDAN UNIVERSITY 1905
復旦大學

NORTHWESTERN UNIVERSITY

advancing the science of networks in communities

VWE Principal Investigators

Noshir Contractor
Northwestern University
Social networks

Jaideep Srivastava
University of Minnesota
Data Mining

Dmitri Williams
University of Southern California
Online games

Scott Poole
University of Illinois
Group Communication

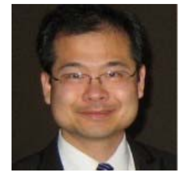
http://vwobservatory.org/

Why Study These Things?

- MMOGs are of interest in their own right
  - Over 1.185 billion registered virtual world accounts and approximately $22.5 billion/year in revenue in 2011
  - Psychological, social, and economic impacts
- What does social behavior in virtual worlds tell us about the “real” world and vice versa?
  - MMOGs may be a mirror of the real world
    - Networks, economics, group processes, conversation, conflict, learning, expertise, leadership, crime, innovation, epidemics, etc.
  - Online games already capture the signatures of these behaviors in huge databases, just waiting to be analyzed

Exemplar II.1
Motivations for creating teams

with ….



Yun Huang
Post doctoral Fellow

Mengxiao Zhu
Ph.D. Candidate

Brian Keegan
Ph.D. Candidate

Jeff Treem
Ph.D. Candidate

NORTHWESTERN
UNIVERSITY

SONIC
advancing the
science of networks in communities

Four Types of Relations in EQ2

Partnership: Two players play together in combat activities;

Instant messaging: Two players exchange messages through Sony universal chat system

Player trade: Players meet "face-to-face" in EQ2 and one gives items to another;

Mail: One player sends a message and/or items to others by in-game mail

	Synchronous	**Asynchronous**
Interpersonal interaction	Partnership, Instant messaging	
Transactional interaction	Player trade	Mail

Data Description

3140 players from Aug 25 to Aug 31 2006, in Antonia Bayle

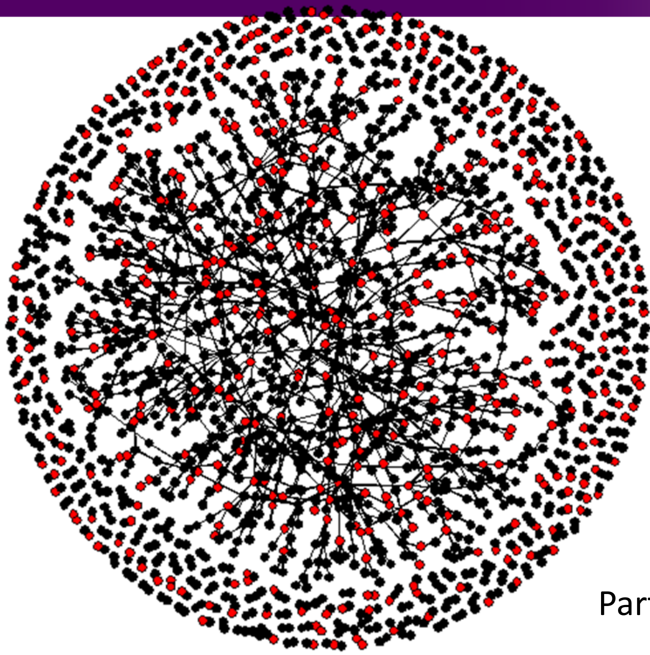
2998 US, 142 CA ; 2447 male, 693 female

- Demographic information
  - Gender, age, and account age (years played Sony games)
  - Zip code, state, and country

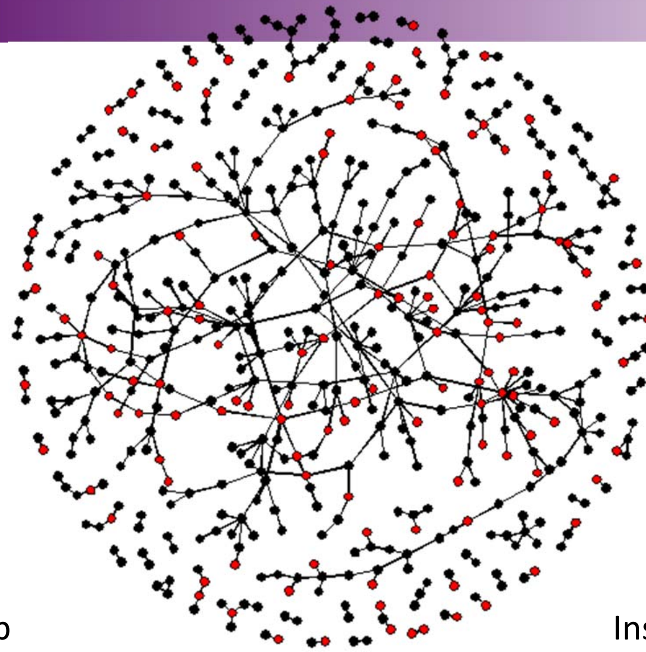
NORTHWESTERN UNIVERSITY

SONIC advancing the science of networks in communities

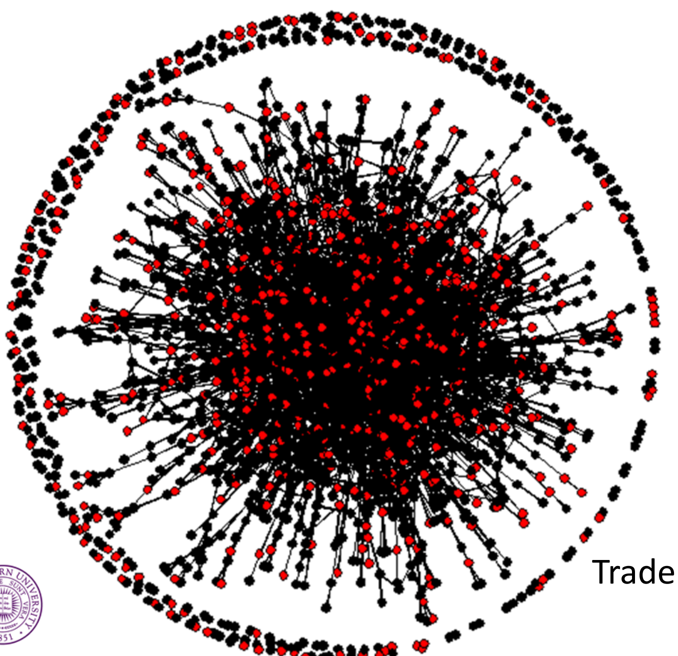
Black: male
Red: female



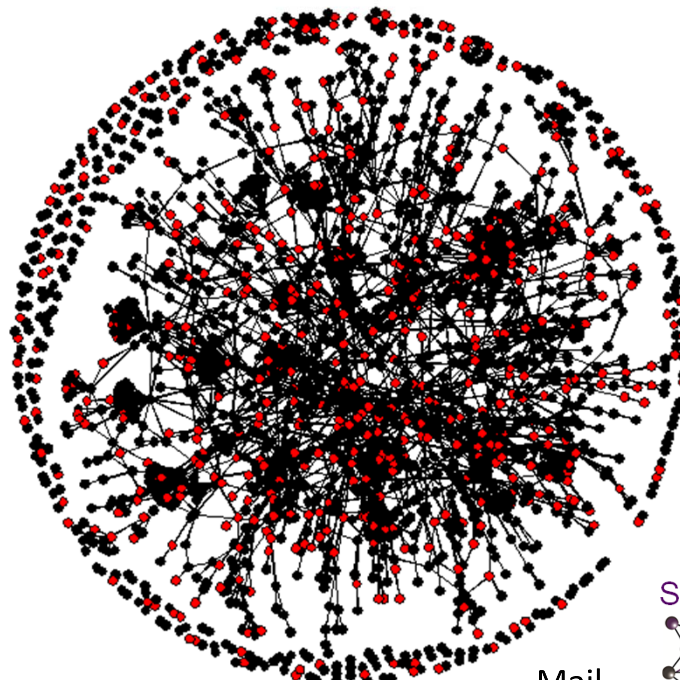
Partnership



Instant messaging



Trade



Mail

Results

Selectivity and transitivity (friend of a friend) exists in all online relations.

Homophily of age and game experience is supported in all four relations.

Distance matters but short distances are more important. Individuals living within 50 Km are 22.6 times more likely to be partners than those who live between 50 and 800 Km.

Time zones impacts gaming and trading but not IM and mail. Individuals in the same time zone are 1.25 times more likely to be game partners than the individuals with one hour difference (but no time zone effect for

Gender homophily is not supported for all relations and female players are more likely to interact with the male players.

Exemplar II.2

How does virtual team assembly influence outcomes?

with…

Mengxiao Zhu
Ph.D. Candidate

NORTHWESTERN UNIVERSITY

SONIC
advancing the science of networks in communities

Combat Groups in EverQuest II

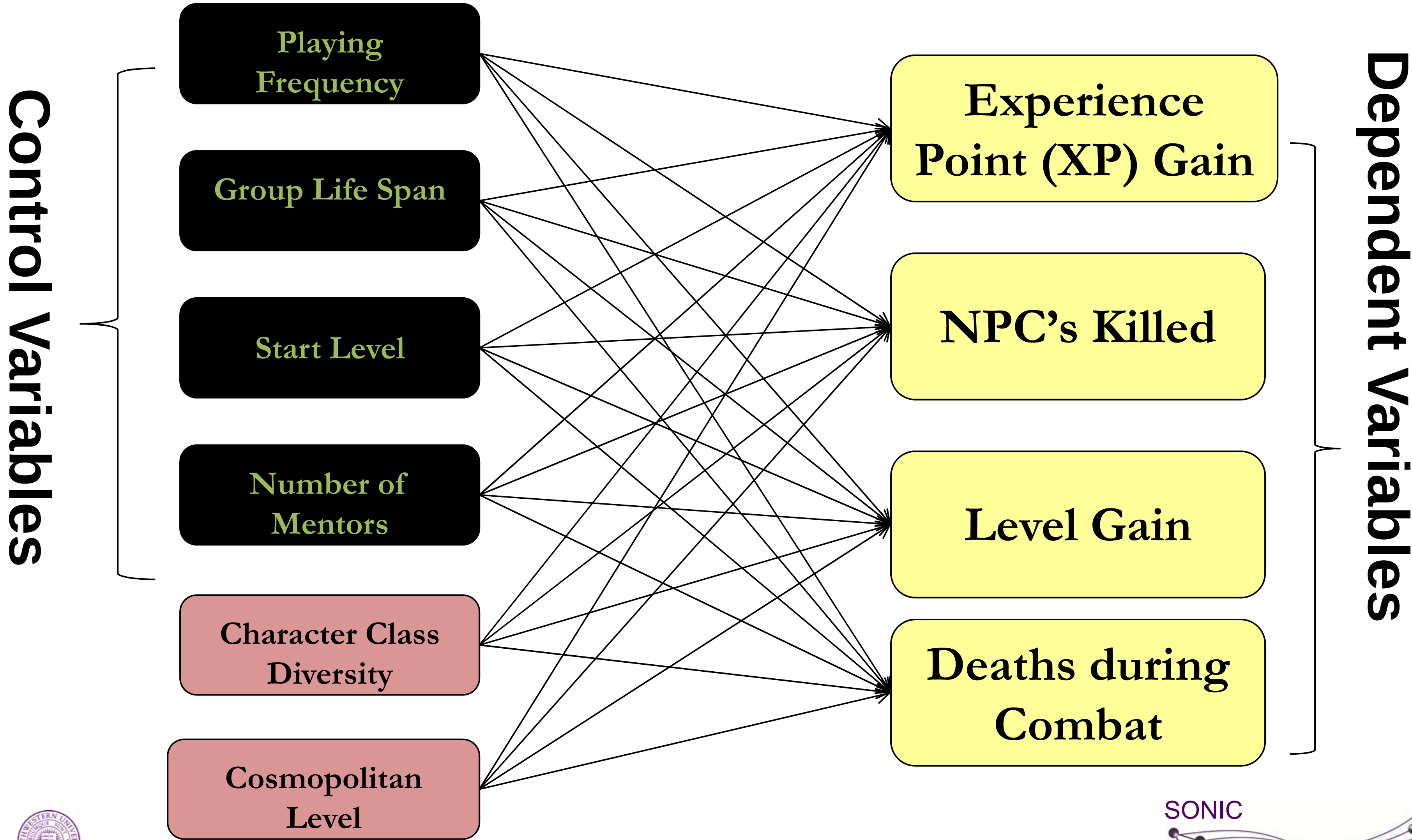
- Difficult combat tasks require collaboration of multiple players and assembly of combat groups
- From 2006-08-27 to 2006-09-11 on Antonia Bayle Server
  - 8,423 players
  - 46,393 groups
  - 9,436,741 combat related records

http://everquest2.station.sony.com/screenshots.vm

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
Constant	-20939.926**	-3.376 (.361)	.717**	4.011**
Frequency	-1553.494 (.105)	4.127**	-.010 (.816)	.601**
Life Span	736.797**	1.174**	.015**	.063**
[illegible]	[illegible]	[illegible]	[illegible]3**	-.050**
Num of Mentor	[illegible]	[illegible]	[illegible]2**	-.647**
Diversity	20819.998**	14.342**	.726**	-1.873 (.095)
Member Cosmo.	30.254 (.612)	-.025 (.698)	-.010**	-.032**
R^2	[illegible]	0.821	0.371	0.244
F	595.213 (p=.000)	1368.793 (p=.000)	176.071 (.000)	96.274 (.000)

** indicates significant results at .01 level

Diversity helps the groups to achieve more.

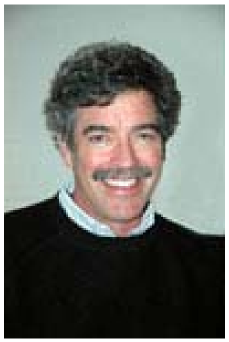
Members being cosmopolitan doesn't help with gains but helps to avoid loss.

Exemplar III

From Understanding to Enabling Team Assembly … or

"Match-making" in Science

Members in the Project



Stanley Wasserman
Indiana University

Hugh Devlin
Northwestern University

Maryam Fazel-Zarandi
University of Toronto

Meikuan Huang
California State University at Stanislaus

Yun Huang
Northwestern University

Alina Lungeanu
Northwestern University

Chuang Zhang
Beijing University of Posts and Telecommunication

Zhe Zhang
Northwestern University

NORTHWESTERN UNIVERSITY

SONIC
advancing the science of networks in communities

Link Predictions for Recommendation Systems

- Social network context
  - Product recommendation: collaborative filtering
  - Expert recommenders
- Other contexts
  - Biology
    - Protein-protein interactions
    - Gene-protein interactions
    - Ecology: predictor-prey relations
  - Information retrieval
  - Record linkage problem
    - Author matching

Expertise Finding as a Two Phase Process

- **Two steps in finding expertise (McDonald and Ackerman,1998):**
  - Expertise **identification**: Problem of knowing what information or special skills other individuals have
  - Expertise **selection**: Problem of appropriately choosing among people with the required expertise

Link Prediction Approaches

1. Node-wise similarity approaches
   - Define or learn a measure of similarity between two nodes to determine link existence
   - Example: Feature vector based, Statistical Relational Learning (SRL), Collaborative filtering, Content-based filtering

2. Network topology based approaches
   - Exploit topological pattern, ranging from local patterns around the nodes to the global patterns covering the entire social network
   - Example: Katz, PageRank.

Liben-Nowell and Kleinberg, 2007; Xiang, 2008

Link Prediction Approaches (cont.)

3. Probabilistic model based approaches
    - Abstract the underlying structure from the observed data network to a compact probabilistic model. Regenerate the unobserved part of the network using the learned model.
    - Example: Probabilistic Relational Models (PRM) framework, Directed Acyclic Probabilistic Entity Relationship (DAPER) framework

- Exponential Random Graph models (p*/ERGM) can be considered as a probabilistic approach based on Markov random graphs.

Predict Link Probability in *p**

- Likelihood of a random network with link *(i,j)* given an observed network x:

$$Pr(\, X_{ij} = 1 \mid \; X = x)$$

where X_{ij} is an element of a random network X: X_{ij} =1 if (i, j) ϵ L; X_{ij} =0 otherwise

- Calculate the probability using the statistics of network configurations

$$\Pr\left(X_{ij} = 1 \middle| \mathbf{X}^{c}_{ij}\right) = \frac{p(\mathbf{X} = \mathbf{x}^{+}_{ij})}{p(\mathbf{X} = \mathbf{x}^{+}_{ij}) + p(\mathbf{X} = \mathbf{x}^{-}_{ij})} = \frac{\exp\{\boldsymbol{\theta}'\mathbf{g}(\mathbf{x}^{+}_{ij})\}}{\exp\{\boldsymbol{\theta}'\mathbf{g}(\mathbf{x}^{+}_{ij})\} + \exp\{\boldsymbol{\theta}'\mathbf{g}(\mathbf{x}^{-}_{ij})\}}$$

where X^{c}_{ij} is the rest of the observed network other than the link x_{ij},
x^{+}_{ij} and x^{-}_{ij} are the network realized by fixing x_{ij}=1 and x_{ij}=0, respectively

Benchmarking p*/ERGM recommendations

Northwestern University Clinical and Translational Science (NUCATS)

Enabling assembly of scientific teams to reduce the delays in translational science from “bench to bedside and back”

Evaluation

- Three benchmark data mining approaches
  1. Node-wise Similarity-based Approach
  2. The Katz Method (Katz 1953)
  3. Relational Bayesian Networks (Jaeger 1997)

- Accuracy comparison based on Average Rank of the Correct Recommendation (ARC) (Burke 2005)

Outcome Comparison

Methods	Average Rank of the Correct Recommendation (std. dev.)			
	Nodal features	Network structures	Dyadic covariates	All variables
p* model	4587 (3231)	803 (1370)	3751 (2855)	603 (1148)
Node-wise similarity	5155 (3243)	1671 (2483)		
Katz		1652 (2469)	1834 (1551)	1551 (1311)[1]
RBN			3381 (3217)	

Similar to the findings in Liben-Nowell and Kleinburg, 2007, the Katz method has the best performance among the benchmark models. Network structures provide the essential information for the predictions and feature similarity only has a small marginal contribution.

Outcome Comparison

Methods	Average Rank of the Correct Recommendation (std. dev.)			
	Nodal features	Network structures	Dyadic covariates	All variables
p* model	4587 (3231)	803 (1370)	3751 (2855)	603 (1148)
Node-wise similarity	5155 (3243)	1671 (2483)		
Katz		1652 (2469)	1834 (1551)	1551 (1311)[1]
RBN			3381 (3217)	

p* models have better performance both in terms of the average rank and consistency of predictions compared to the benchmark models using similar variables. The final p* model utilizes all variables and achieves the best performance.

Upper Bound Comparison

- Ideal way to combine multiple methods

Test cases	Method A	Method B	Method C	Combine(A,B,C)
Validate x_{ij}	$Rank_A(x_{ij})$	$Rank_B(x_{ij})$	$Rank_C(x_{ij})$	$Min(Rank_A(x_{ij}), Rank_B(x_{ij}), Rank_C(x_{ij}))$
Validate x_{ik}	$Rank_A(x_{ik})$	$Rank_B(x_{ik})$	$Rank_C(x_{ik})$	$Min(Rank_A(x_{ik}), Rank_B(x_{ik}), Rank_C(x_{ik}))$
…				

- p* vs. three methods using all variables
  - Combine(Node-wise similarity with nodal features, Katz with network structures, BRN) → ARC 411 (p* ARC 603)
- Can the p* model bring more prediction power?
  - Combine(p*, Node-wise similarity with nodal features, Katz with network structures, BRN) → ARC 284

Demo

- NUCATS Semantic C-IKNOW

Key Takeaways

- Web Science is well poised to make a leap in understanding and enabling team assembly by facilitating recent advances in:
  - Theories: Theories about the social motivations for creating, maintaining, dissolving and re-creating networks
  - 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 networks.
  - Methods: An ensemble of qualitative and quantitative methods (exponential random graph modeling (*p**) techniques to understand and enable theoretically grounded network recommendations.
  - Computational infrastructure: Cloud computing and petascale applications are critical to face the computational challenges in analyzing the data

Get involved with ..

- NetSci 2012 @ Northwestern:
  - Workshops: June 18, Monday – June 19, Tuesday
  - Conference: June 19, Tuesday – June 22, Friday

- ACM WebSci 2012 @ Northwestern:
  - Workshops: June 21, Thursday
  - Conference: June 22, Friday – June 24, Sunday

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