

# Combining Survey and Bibliometric Approaches in the Study of Collaborative and Professional Networks of Academic Scientists

*Julia Melkers, Associate Professor*

*Diana Hicks, Professor and Chair*

*Fang Xiao, Doctoral Student*

*Jian Wang, Doctoral Student*

**School of Public Policy**

**Georgia Institute of Technology**

*Project PI's: Julia Melkers and Eric Welch*

*Research funded by the National Science Foundation:  
(Grant #REC-0529642 & NSF Grant # DRL-0910191)*

*Program Officer: Janice Earle*



# An Overview: Guiding Research Questions

*“...differences in career trajectories for men and women are generated and reinforced by the social structures in which people are situated and by the networks of interactions in which they participate”*

*(NAS, 2006 (Bias and Barriers))*

## **Therefore, we ask:**

- ❑ How and why do professional networks matter for career advancement retention, and satisfaction in academic science?**
- ❑ Are these networks and their resources significantly different for women and underrepresented minorities? And, to what end?**

**Methodological challenge:  
how to effectively measure networks?**



# Tracking Networks in Science: Production-Based Links

- Bibliometric methods based on publication data have often been used to identify ties between academic scientists.
  - Data show production based ties which can be used to address quantity, impact, and (with Vantage Point) additional characteristics (interdisciplinarity)
- Strengths: accessible and well established data
- Weaknesses: little insight to pre-publication relationships, social aspects of collaboration, other factors



# Survey-Based Social Network Methods: Gaining Traction in the Research Community

- Survey-based social network methods allow for the collection of detailed data on the formal and informal social structures of science.
- Survey methods ask about specific relationships – can be tailored to research needs.
  - Data allow us to uncover the relationships/ties that precede the development of publication ties.
  - Which ties are closest? Most productive? Where do the ties originate?
- Strengths: Rich data on collaborative relationships
- Weaknesses: Self reported, burdensome, may not account for production accurately



# Methodological Questions

- What do bibliometric and survey-based social network approaches tell us about the professional relationships of scientists?
- How can these methods be combined for an even more robust approach to questions about the scientific workforce and research outcomes?
- How well do they align?
  - Important distinction: close vs full collaborative networks



# Multi-methodological and Longitudinal Approach

- National NSF-funded longitudinal study of academic scientists in 6 disciplines

## ■ Quantitative Core

- Longitudinal survey of academic scientists
  - Detailed ego-centric social network survey design
  - Collects detailed individual-level data regarding *SPECIFIC* professional ties plus traditional survey data
- Merge with institutional data
- Bibliometric data
- CV data



# Method 1: Survey-Based Social Network Methods



# NETWISE I Survey Themes

## ■ What is the social structure?

✓ *name generators*

- Close research collaboration networks (within and outside of one's university)
- Research discussion networks
- Advice networks (career and departmental information)
- Mentor relationships

## ■ What are the characteristics of each relationship?

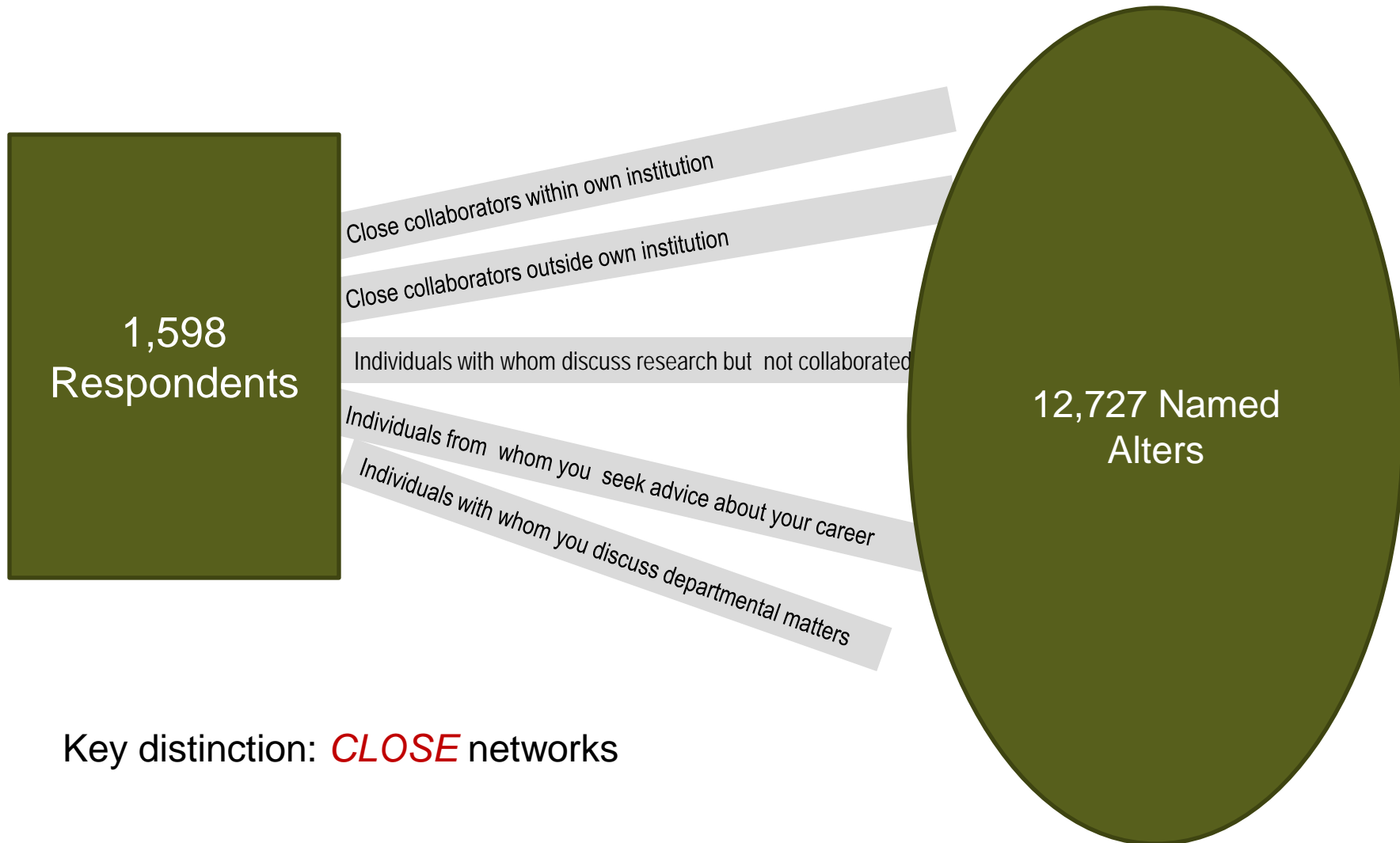
✓ *name interpreters*

- Characteristics of named alter (gender, skills)
- Origin and nature of the relationship
- Types of collaboration
- Collaborative outcomes
- Types of advice
- Career resources (introductions, nominations, advice)
- Connections between named alters





# Generating Network Data



Key distinction: **CLOSE** networks



# Aggregating Network Data to Create Relational Attribute Data: Counts, Sums, Binary Data

*Example: Female Full Professor in EAS*

	ID	nilenght
5311	40279	2
5312	40279	1
5313	40279	1
5314	40279	1
5315	40279	3
5316	40279	3
5317	40279	3
5318	40279	2
5319	40279	3
5320	40279	2
5321	40279	3
5322	40279	3
5323	40279	2
5324	40279	3



**Respondent #40279**  
*Mean length of respondent has known alter*

---

1=less than 3 years  
2= 3-6 years  
3= more than 6 years

---

**Mean Collaborative Network=2.0**

**Mean Research Talk Group= 2.50**

**Mean Advice Groups=2.75**



# Method 2: Gathering and Matching Bibliometric Data to Survey data

# Gathering Bibliometric Data

- Gathering bibliometric data for each survey respondent in 2007 and 2010, respectively.

	<b>2007 Collection</b>	<b>An Example- Ego's name: Will Smith</b>	<b>2010 Collection</b>	<b>An Example- Ego's name: Will Smith</b>
Search	Ego's last name + first initial plus Ego's last name + first initial + middle initial	Author's name: Smith, W or Smith, W E	Ego's last name + first initial plus Ego's last name + first initial + middle initial	Author's name: Smith, W or Smith, W E
Refine search results by	All institution egos have been affiliated with since they entered Ph.D. programs	Institution A (Phd 95-00) B (Postdoc 00-02) C (Asst Professor 02-05) D (Assoc Professor 05-08)	The institution egos were affiliated with when we conducted the survey plus egos' new institution if they have	Institution  D (Assoc. Professor 05-08) E (Professor 08-present)
Time Span	(The year the person received Ph.D. degree -6) to 2007	1994-2007	2006 - present	2006-2010

Survey Respondents' CVs  
or personal website



# Our Bibliometric Data

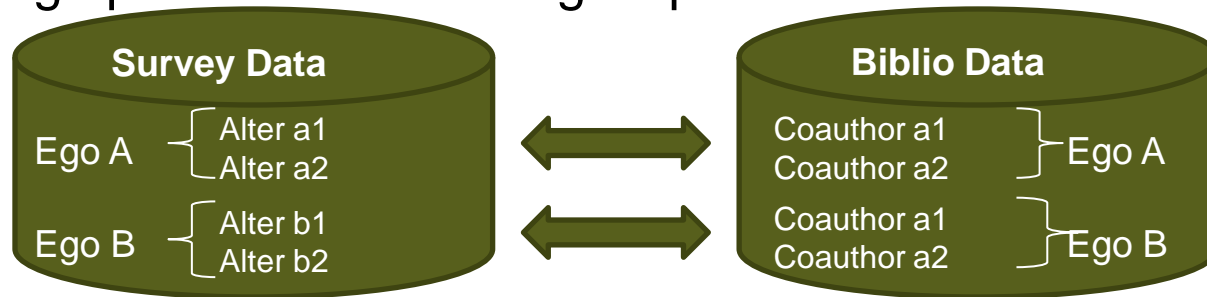
## ■ Number of publications

- 1592 respondents have 94,669 publications in WoS
- The largest number of publications by one respondent is 861.

Field	Results in the Dataset
Abstract	56,514 abstracts; 71% publications have abstracts.
Author Affiliations	92% publications have author affiliations. Organization name, city and country 18,124 organization names. But they need to be cleaned furthermore.
ISI Unique Article Identifier	83,124
Cited References	2,819,511 cited references (an average of 29.8 per article)
Publication type	Journal (92,631), Serials (1,320), Conference (718)
Times Cited	Least number of citations: 0 Largest number of citations: 6,229
Publication Year	Earliest publication year: 1963

# Matching Bibliometric Data with Survey Data

- Methodological Question: Are collaborators named by egos showing up as coauthors in egos' publications?



## Procedures

- Step 1: Clean alter names – 7306 names

EgoID	AlterID	Cleaned name	Found name	First Initial	Middle Initial	Last name
10001	1000101	Will Smith	Will E. Smith	W	E	Smith

- Step 2: Match manually unusual names with biblio author names (10%)
  - COREY S. O'HERN -> O'HERN or OHERN
  - Theodor W. Hänsch -> Hänsch or Hansch
  - Walter A. de Heer -> de Heer or Heer
  - Ann Cornell-Bell -> Cornell-Bell or Bell
- Step 3: Use Microsoft Access Program to run matching
- Step 4: Additional error checking and review



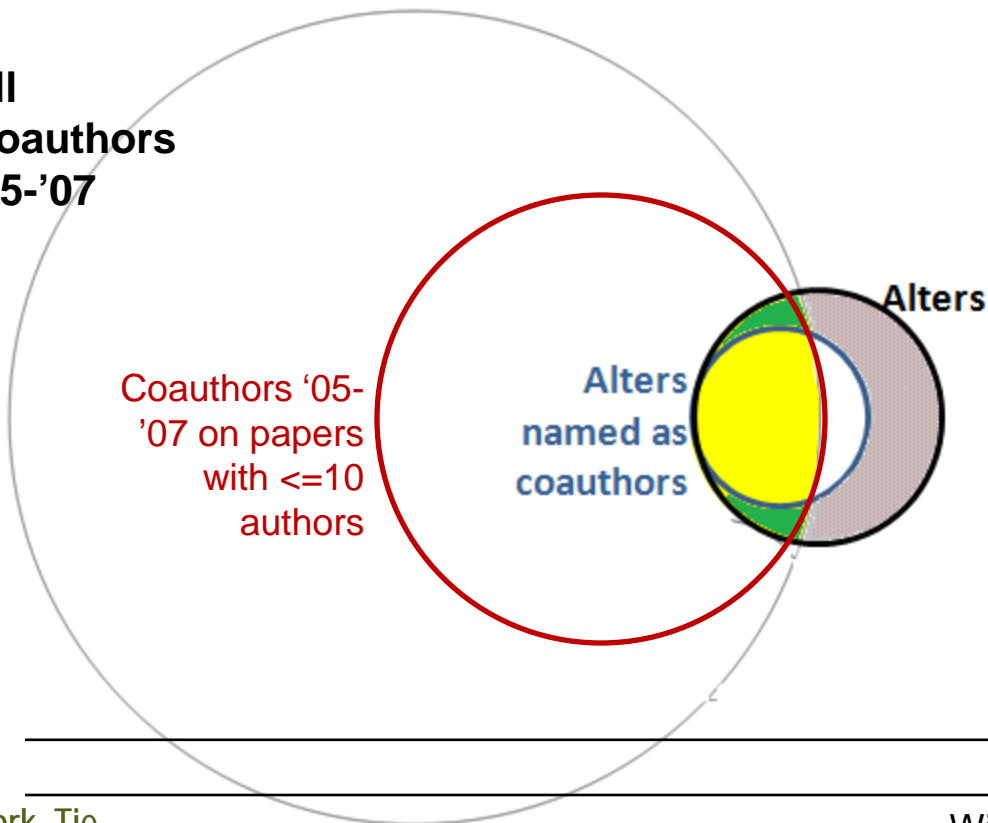
# Results: Assessing the Bibliometric-Survey/Social Network Match



## Bibliometric Data

## Survey Data

All Coauthors '05-'07



	Names show up in Biblio data	Names do NOT show up in Biblio data
Alters named as "close" coauthors	Group1 – <b>Yellow</b> 1600	Group2 – White 869
Alters not named as "close" coauthors	Group3 – <b>Green</b> 704	Group4 – Grey 2872

		Group 1	2	3	4
Network Tie Type	Within University	42.6%	40.5%	40.5%	<b>56.5%</b>
	Outside University	57.4%	59.5%	59.5%	<b>43.5%</b>
Strength of Interaction	Length of time knowing alters (1=less than 3 years, 3=more than 6 years;)	2.57	2.50	2.53	<b>2.15</b>
	Frequency of personal contact with alters (1=less often, 4=at least daily)	2.62	2.74	2.67	2.78
Other Collaboration Type	Collaboration Multiplexity Index (created by summing the collaboration types except coauthoring. min=0 and max=3)	.65	.62	<b>.73</b>	<b>.72</b>
	Unpublished working or conference papers	.51	.51	<b>.62</b>	.47
Knowledge Distance	Knowledge of collaborators expertise (1=detailed understanding, 2=working understanding, 3= little to no understanding)	1.41	1.44	1.44	<b>1.67</b>
Structural Position	Alter is senior to ego	.36	.30	.38	<b>.42</b>





# Matched alters appear to be more important than other alters

Matched alters provide more types of resources and advices, and have closer relationship with egos than other alters

	Matched alters	Unmatched
<b>Resources from alter (0-3)</b>	.84	.69
Resource Multiplexity index (mean)		
Alters reviewed egos' papers or proposals prior to submission		
Alters introduced egos to potential collaborators outside of egos' university		
Alters nominated egos for an award or as an invited speaker		
<b>Advices from alter (0-5)</b>	3.29	2.87
Multiplexity index (mean): egos seek advice from alters on		
Publishing, grant getting, overall career development strategies, interactions with colleagues, and work/family balance		
<b>Close Relationship (0-4)</b>	0.56	0.48
Multiplexity index (mean)		
created by summing the roles in which the collaborator was named ( <i>primary mentor, discuss department issues and turn to for advice and whether he or she was named as a close friend.</i> )		



# Observations thus far

- Synergies are fantastic!
- Combining detailed bibliometric data with detailed social network and survey data can add considerably to our understanding of collaborative interactions.
- Methodological assumptions and weaknesses in approaches can be addressed through combination and analysis of data.
- Results provide insight -- not all collaborations are equal.