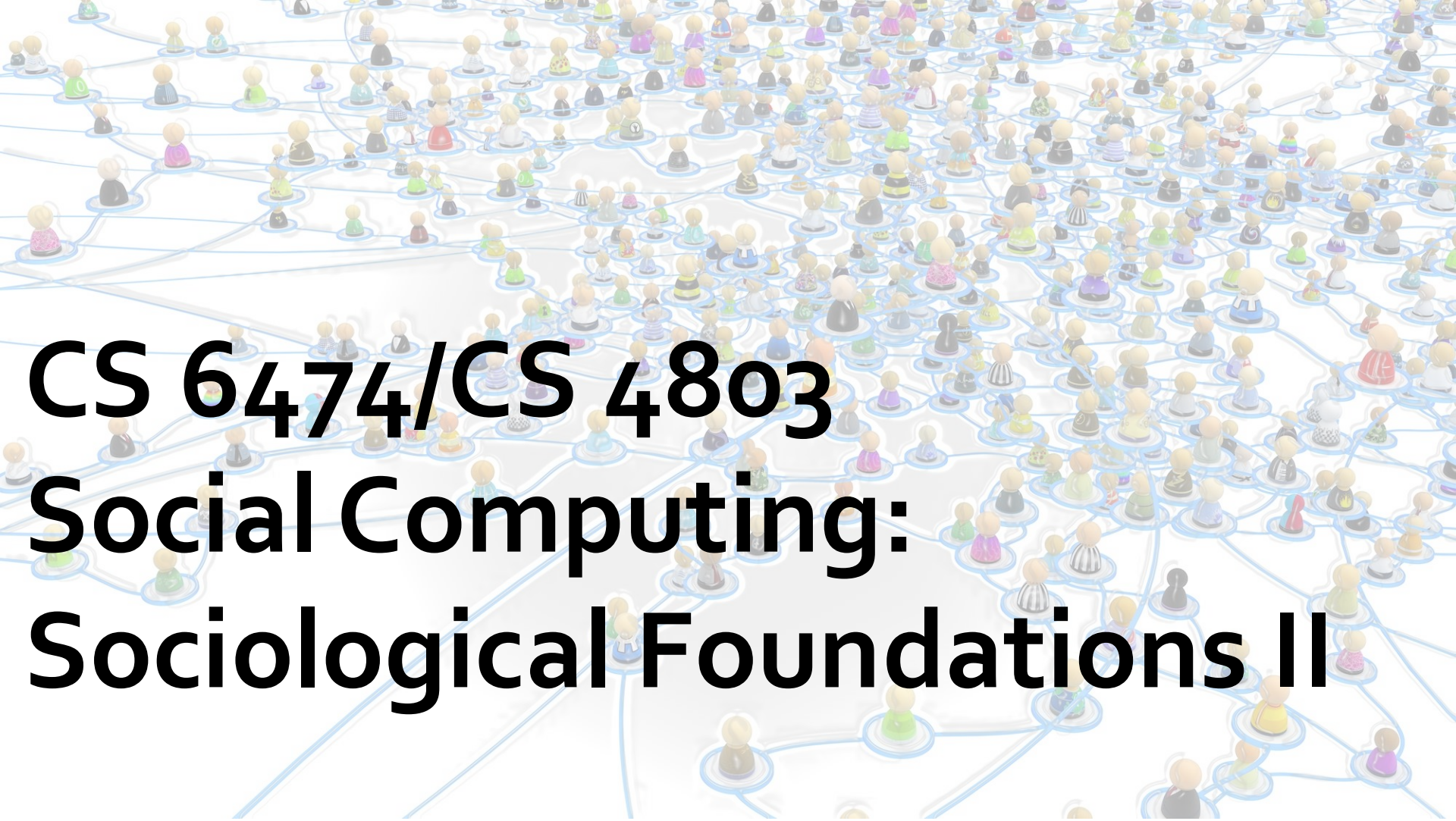


CS 6474/CS 4803 Social Computing: Sociological Foundations II

Munmun De Choudhury

munmund@gatech.edu

Week 2 | January 21, 2026

Example Class Projects

Data and Behavioral Characterization

- Comparison of public and critics sentiment around movies (Twitter vs. Rotten Tomatoes + IMDB)
  - Focus on most watched and recent movies
  - Limited dataset
- Identifying hateful comments on YouTube
  - Mixed methods approach
  - Limited human annotations
  - Context matters

Community Studies

- Differences between reputation mechanisms of Quora and StackOverflow – what leads to better reputation?
  - Rich dataset
- Understanding the structure and dynamics of pro-eating disorder communities
  - Resulted in a publication (CSCW)

Prediction

- Predicting emotionality and loneliness of Flickr photos by analyzing visual features
  - Computer vision application
  - Loneliness is a complex construct
- Prediction of workplace stress by analyzing corporate email content (Enron)
  - Mixed method study
  - Interesting NLP application
  - Connection to real world events (stock market indices)

Tools and Visualization

- Classification and exploration of political bias in blogs via a visualization
  - Left and right wing blogs shown in a network visualization
- Content summarization tool to provide positive social support to mental health help seekers on Reddit

Domain-Specific Exploration (1)

- Inferring a "health score" of the eating activities of college students by analyzing content of Yelp reviews
- Assessing citizens' affect from tweets in different urban settings
- "Social listening" – using social media sentiment to infer stock market trends
- "Social listening" – using social media sentiment around the 2024 Presidential election in the Usw
- "Social listening" – using social media sentiment around the two ongoing global conflicts – the Ukraine-Russia and the Israel-Gaza conflicts

Domain-Specific Exploration (2)

- Discovering the grievances, emotional expression, and topics shared in an online community of incarcerated individuals
- Employing Facebook pages as a platform for community policing (disappearances of individuals in Mexico)
- Analyzing and Predicting Student Success Based on Course Interaction in an Introductory Computer Science Class

Domain-Specific Exploration (3)

- What messages elicit less or more social support?
  - Reddit mental health communities
  - Up- and downvotes as a proxy for social support
  - NLP and psycholinguistic analysis

- Online community (Reddit) discussions about COVID-19 and their relation to offline COVID-19 diagnostic testing volume.
  - City level analysis
  - External data source – city-level Covid testing records

Ethics and Equity – Paper (1)

- Large language models are now the standard to develop state-of-the-art solutions for text detection and classification tasks.

- The adoption of large language models for building approaches for tasks aimed at detecting humanitarian information, fake news, and emotion can lead to systematically lower performance on non-English languages when compared against the performance on English.
- The project empirically showed that including images via multimodal learning bridges this performance gap.

Ethics and Equity (2)

- A project examined racial bias and authenticity in yelp restaurant reviews
- Content analysis of misogyny and sexism in Instagram comments

Start forming teams for term project

Team matchmaking: Jan 26
Proposal specs available on course website: Jan 28
Proposals due: Feb 11

Last class: *human social networks have unique characteristic structures*

Structuralist Approach

A collection of human beings does not become a society because each of them has an objectively determined or subjectively impelling life-content. It becomes a society only when the vitality of these contents attains the form of reciprocal influence; only when one individual has an effect, immediate or mediate, upon another, is mere spatial aggregation or temporal succession transformed into society. (Simmel, 1908 [1971], pp. 24-25)

Georg Simmel

Georg Simmel

Born	1 March 1858 Berlin, Kingdom of Prussia
Died	26 September 1918 (aged 60) Strassburg, German Empire
Nationality	German
Alma mater	University of Berlin
Era	19th-century philosophy
Region	Western philosophy
School	Neo-Kantianism *Lebensphilosophie*[1]
Institutions	University of Berlin University of Strasbourg
Notable students	György Lukács
Main interests	Philosophy, sociology
Notable ideas	Formal sociology, social forms and contents, the tragedy of culture,[2] web of group affiliation

Structuralist Approach

- Looking at network causes of phenomenon of interest
  - Today
- Looking at network effects of phenomenon of interest
  - Next class

Social structures, creativity, and innovation

Structural Holes and Good Ideas

Summary

- Role of social network structure on access to social resources
- Burt's observations:
  - Opinions and thoughts within groups are homogenous
  - People who extend themselves across the 'structural holes' between groups are exposed to diverse ways of thinking

- Brokerage across structural holes between groups can lead to greater accumulation of "social capital" – quantitatively defining the *network constraint* measure, that uses the size, density, and hierarchy measures of an individual's egonetwork
  - Hypothesis is tested with a case study of the network structure of managers in a supply chain company

Network constraint measures the extent to which a person's network is like a straitjacket around them, limiting their vision of alternative ideas and sources of support.

Network Constraint

- Measure of the extent to which the people a respondent knows are tied to each other
- High constraint means the network is redundant and recycles information
- Low constraint = bridging between groups = good ideas

Summary

- Managers asked to come up with an idea to improve the supply chain
- Then asked:
  - whom did you discuss the idea with?
  - whom do you discuss supply-chain issues with in general
  - do those contacts discuss ideas with one another?

- 673 managers (455 (68%) completed the survey)
- ~ 4000 relationships (edges)

Structural Holes

(Figure 1 from Burt 2004)

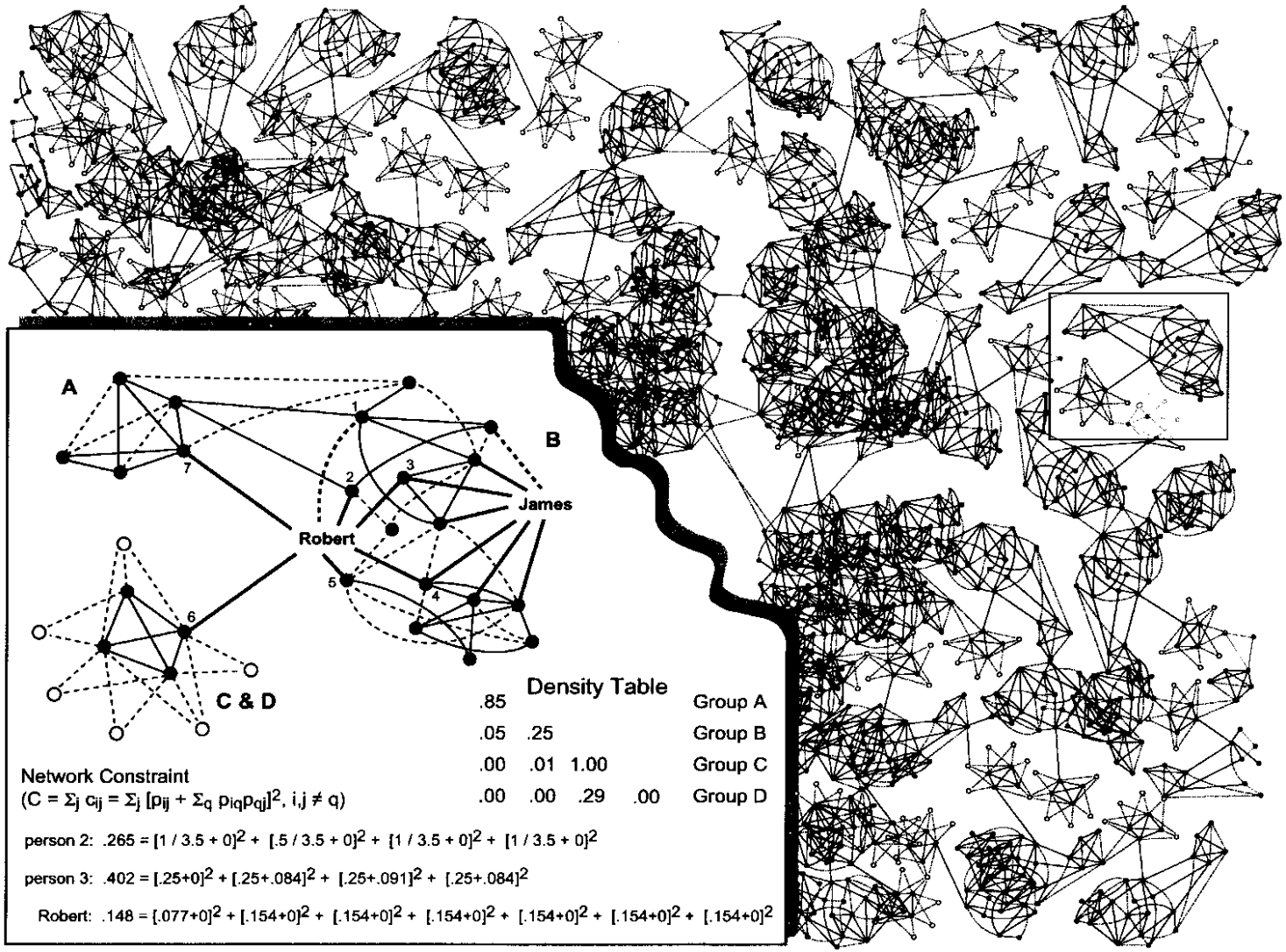


TABLE 1

PREDICTING PERFORMANCE

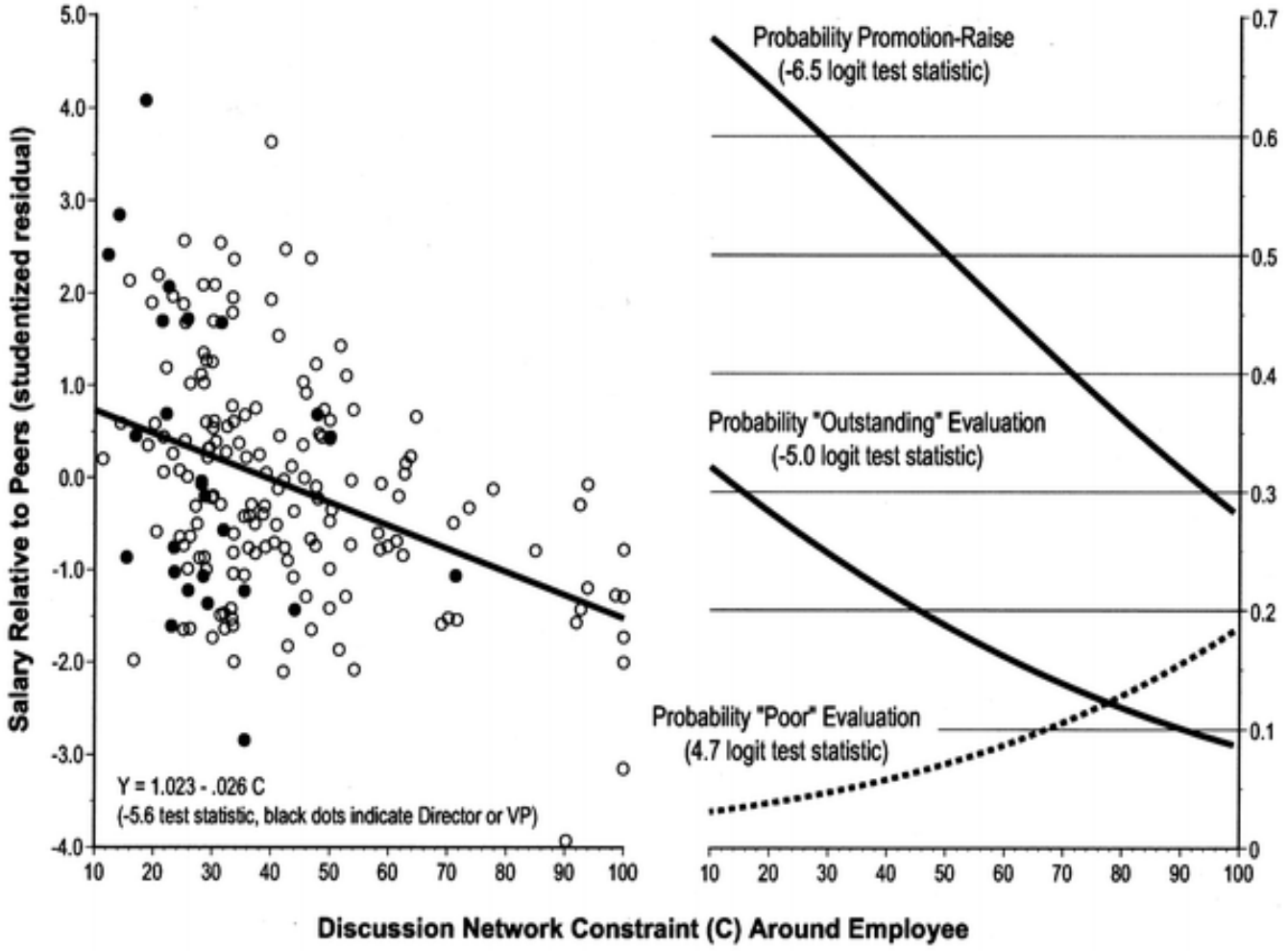
	1 Salary		2 Salary		3 Evaluation		4 Promotion	
Manager 1	−31,099**	(2,882)	−35,707**	(3,498)	−.973	(.678)	.689	(.670)
Manager 2	−16,652**	(2,745)	−19,892**	(3,479)	−.863	(.631)	1.165	(.648)
Manager 3 (reference)	. . .		. . .		. . .		. . .	
Sr. manager	19,638**	(3,782)	15,484**	(4,143)	.116	(.843)	−.635	(.885)
Executive	65,394**	(4,522)	61,930**	(4,835)	.423	(1.01)	.221	(1.08)
Purchasing	754	(1,351)	1,811	(1,884)	.410	(.313)	.478	(.345)
Age	338**	(52)	300**	(71)	−.085**	(.013)	−.084**	(.013)
Bachelor	1,610	(1,003)	200	(1,401)	−.211	(.237)	.118	(.240)
Graduate	734	(864)	−451	(1,155)	−.208	(.203)	.182	(.204)
Hightech	3,516**	(880)	3,150*	(1,189)	.087	(.209)	.162	(.210)
Lowtech	−6,927**	(1,481)	−6,607*	(2,375)	−.351	(.342)	−.409	(.378)
Urban 1	3,613**	(1,046)	3,947**	(1,456)	.423	(.247)	−.152	(.252)
Urban 2	5,049**	(1,010)	5,585*	(1,427)	−.564	(.238)	−.052	(.243)
Network constraint	−7	(25)	−1	(38)	−.014**	(.004)	−.022**	(.006)
Mgr2 × constraint	−19	(35)	−47	(58)	.004	(.008)	−.008	(.009)
Mgr3 × constraint	−47	(38)	−159*	(59)	−.007	(.009)	.003	(.009)
SrMgr × constraint	−214*	(75)	−216*	(84)	−.005	(.017)	.010	(.019)
Executive × constraint	−681**	(124)	−697**	(132)	−.011	(.028)	.024	(.030)
N	673		398		673		638	

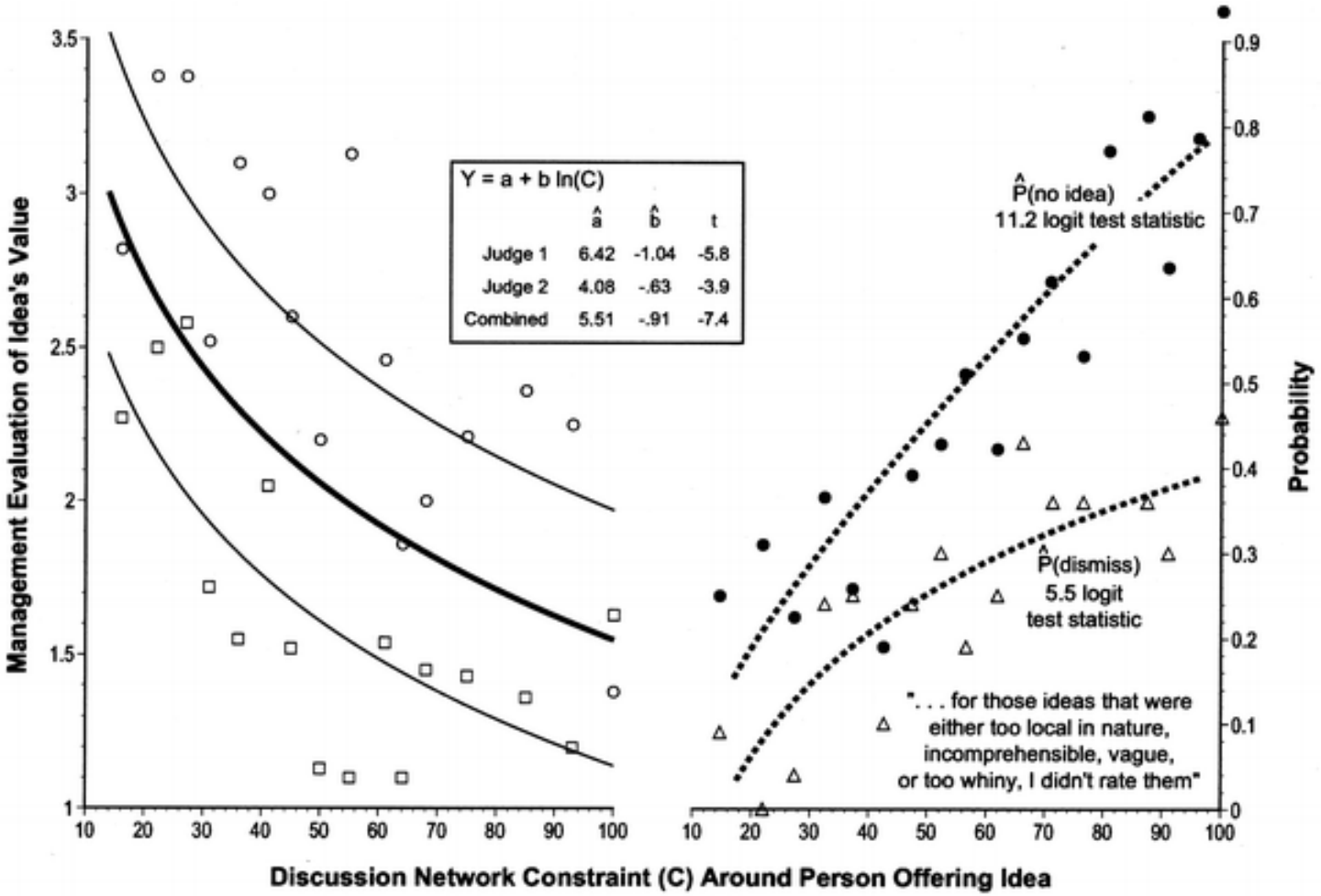
NOTE.—Coefficients in models 1 and 2 are change in salary dollars with a unit increase in row variable (respectively .80 and .83 squared multiple correlations; network effect plotted in fig. 4). Coefficients in model 3 predict three levels of evaluation for an ordinal logit model (114.8 χ^2 with 17 *df*; network effects are plotted in fig. 4 holding age constant). Coefficients in model 4 are for a logit model predicting whether the employee was promoted in the year after the network survey or received an above average raise (100.5 χ^2 with 17 *df*; network effect is plotted in fig. 4 holding age constant). SEs are given in parentheses.

* $P < .05$.

** $P < .001$.

The results show a strong effect of network constraint on salary, evaluation and promotion, independent of the job/age characteristics related to human capital explanations.




Management Evaluation of Idea's Value
Probability
Discussion Network Constraint (C) Around Person Offering Idea
Y = a + b ln(C)
â b̂ t
Judge 1 6.42 -1.04 -5.8
Judge 2 4.08 -.63 -3.9
Combined 5.51 -.91 -7.4
P̂(no idea)
11.2 logit test statistic
P̂(dismiss)
5.5 logit
test statistic
". . . for those ideas that were either too local in nature, incomprehensible, vague, or too whiny, I didn't rate them"

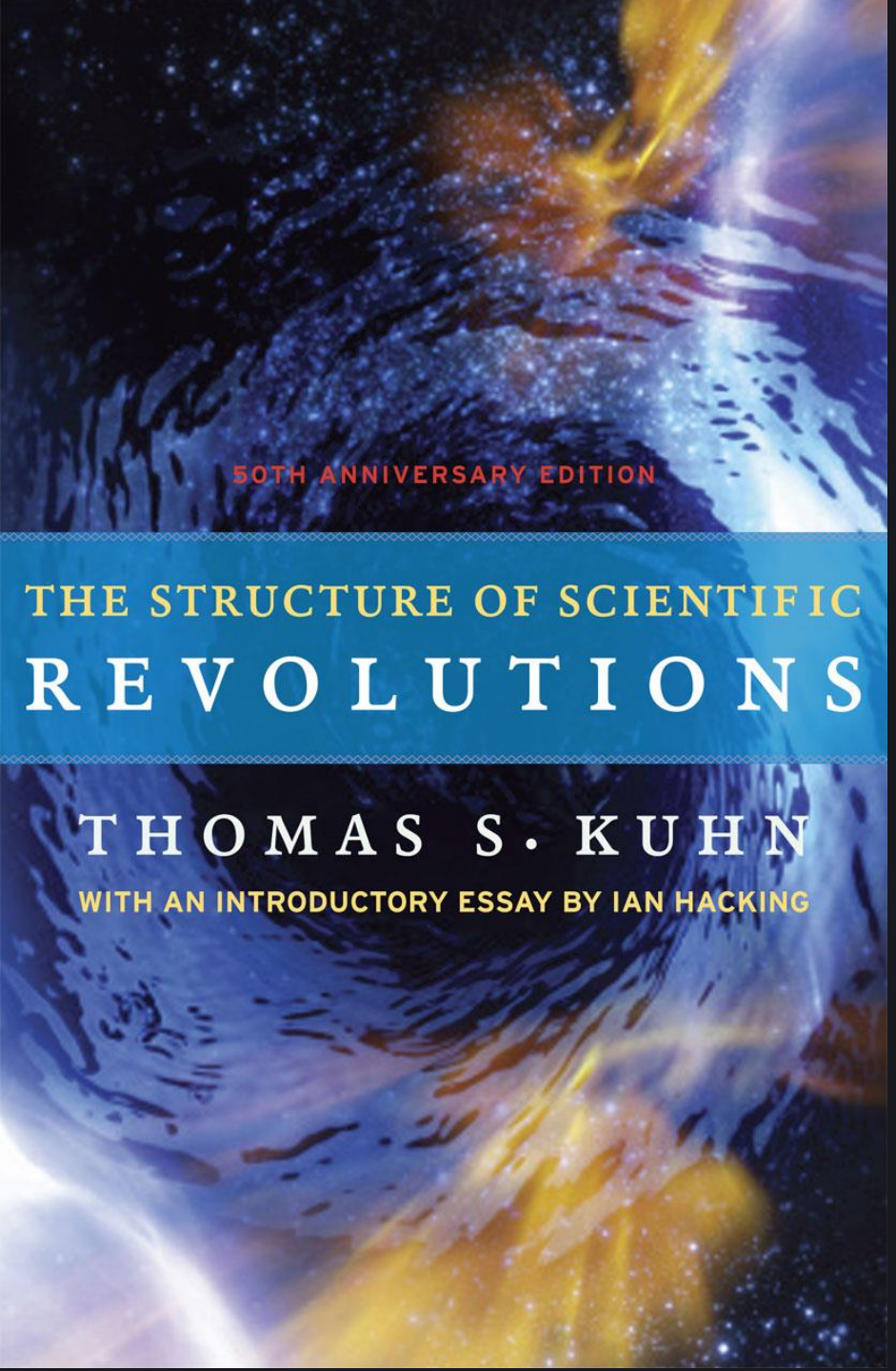
Four levels of brokerage

Summary

- Main finding – interconnected groups give rise to "better ideas" compared to densely intra-connected groups
- Other findings –
  - 1) organizations that collaborate with partner firms show greater financial growth;
  - 2) higher ranked, high tech, and managers in urban settings came up with better ideas;
  - 3) managers who brokered connections across structural holes were rewarded for brokerage in terms of compensation, performance evaluations, and promotions

To what extent are the findings on the importance of brokerage and structural holes affected by the case studies considered?

What are some of the variables that should have been considered/controlled for in the study?



"Almost always the men who achieve [these] fundamental inventions of a new paradigm have been either very young or very new to the field whose paradigm they change. And perhaps that point need not have been made explicit, for obviously these are the men who, being little committed by prior practice to the traditional rules of normal science, are particularly likely to see that those rules no longer define a playable game and to conceive another set that can replace them."

—*Thomas S. Kuhn, The Structure of Scientific Revolutions*

THINK TANK; Where to Get a Good Idea: Steal It Outside Your Group

By MICHAEL ERARD MAY 22, 2004

Got a good idea? Now think for a moment where you got it. A sudden spark of inspiration? A memory? A dream?

Most likely, says Ronald S. Burt, a sociologist at the University of Chicago, it came from someone else who hadn't realized how to use it.

"The usual image of creativity is that it's some sort of genetic gift, some heroic act," Mr. Burt said. "But creativity is an import-export game. It's not a creation game."

Mr. Burt has spent most of his career studying how creative, competitive people relate to the rest of the world, and how ideas move from place to place. Often the value of a good idea, he has found, is not in its origin but in its delivery. His observation will undoubtedly resonate with overlooked novelists, garage inventors and forgotten geniuses who pride themselves on their new ideas but aren't successful in getting them noticed. "Tracing the origin of an idea is an interesting academic exercise, but it's largely irrelevant," Mr. Burt said. "The trick is, can you get an idea which is mundane and well known in one place to another place where people would get value out of it."

Mr. Burt, whose latest findings will appear in the American Journal of Sociology this fall, studied managers in the supply chain of Raytheon, the large electronics company and military contractor based in Waltham, Mass., where he worked until last year. Mr. Burt asked managers to write down their best ideas about how to improve business operations and then had two executives at the company rate their quality. It turned out that the highest-ranked ideas came from managers who had contacts outside their immediate work group. The reason, Mr. Burt said, is that their contacts span what he calls "structural holes," the gaps between discrete groups of people.

Structural holes always help? Well it depends

New Contract Revenue — **Coefficients[a]**

	Unstandardized Coefficients			
	B	Std. Error	Adj. R^2	Sig. F Δ
(Base Model)			0.40	
Size Struct. Holes	13770***	4647	0.52	.006
Betweenness	1297*	773	0.47	.040

a. Dependent Variable: **Bookings02**
b. Base Model: YRS_EXP, PARTDUM, %_CEO_SRCH, SECTOR(dummies), %_SOLO.

Contract Execution Revenue — **Coefficients[a]**

	Unstandardized Coefficients			
	B	Std. Error	Adj. R^2	Sig. F Δ
(Base Model)			0.19	
Size Struct. Holes	7890*	4656	0.24	.100
Betweenness	1696**	697	0.30	.021

a. Dependent Variable: **Billings02**
b. N=39. *** p<.01, ** p<.05, * p<.1

Bridging diverse communities is more significant for *landing* new contracts.

Being in the thick of information flows is more significant for contract *execution*.

Source: M. van Alstyne, S. Aral. Networks, Information & Social Capital (formerly titled 'Network Structure & Information Advantage'), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=958158

Structural holes always help? Well it depends

New Contract Revenue — **Coefficients[a]**

	Unstandardized Coefficients			
	B	Std. Error	Adj. R^2	Sig. F Δ
(Base Model)			0.40	
Best structural pred.	12604.0***	4454.0	0.52	.006
Ave. E-Mail Size	-10.7**	4.9	0.56	.042
Colleagues' Ave. Response Time	-198947.0	168968.0	0.56	.248

a. Dependent Variable: **Bookings02**
b. Base Model: YRS_EXP, PARTDUM, %_CEO_SRCH, SECTOR(dummies), %_SOLO.

Contract Execution Revenue — **Coefficients[a]**

	Unstandardized Coefficients			
	B	Std. Error	Adj. R^2	Sig. F Δ
(Base Model)			0.19	
Best structural pred.	1544.0**	639.0	0.30	.021
Ave. E-Mail Size	-9.3*	4.7	0.34	.095
Colleagues' Ave. Response Time	-368924.0**	157789.0	0.42	.026

a. Dependent Variable: **Billings02**
b. N=39. *** p<.01, ** p<.05, * p<.1

Sending *shorter* e-mail is positively related to both new contracts and contract execution.

Faster response from colleagues is positively related to contract execution revenues.

Source: M. van Alstyne, S. Aral. Networks, Information & Social Capital (formerly titled 'Network Structure & Information Advantage'), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=958158

What structural holes are not…

	Revenue $	$ for completed searches	Completed searches	Multitasking	Duration	Duration controlling for multitasking
Size of rolodex (Q50)	-10.2 (60.3)	-22.9 (32.6)	0.000 (0.001)	0.000 (0.001)	-0.013 (0.021)	-0.013 (0.016)

*** p < 0.10, ** p < 0.05, *** p < 0.01, Standard err in paren.**

Instead, a larger private rolodex is associated with:

- Less information sharing
- Less DB proficiency
- Lower % of e-mail read
- Less learning from others
- Less perceived credit for ideas given to colleagues
- More dissembling on the phone

Recruiters with larger personal rolodexes generate no more or less output

Source: M. van Alstyne, S. Aral. Networks, Information & Social Capital (formerly titled 'Network Structure & Information Advantage'), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=958158

Your thoughts

Brokerage led to promotions, salary hikes, and positive performance evaluations of managers. If brokerage improves "performance" in an online setting, what form of "performance" can it be? On a related note, what would it mean to replicate Burt's findings in online social networks?

Class Exercise

Pick one familiar online platform, for example:

Reddit
TikTok
Discord
Instagram
Stack Overflow

1. What are the "clusters" on this platform?
2. Who or what bridges that gap?
3. What counts as a "good idea" on this platform and who is structurally positioned to produce it?
4. Does brokerage increase visibility, credibility, or risk on this platform?
5. Is brokerage user-driven or system-driven here?

Assignment I is now available on the course website (Due: Jan 28, 2026)