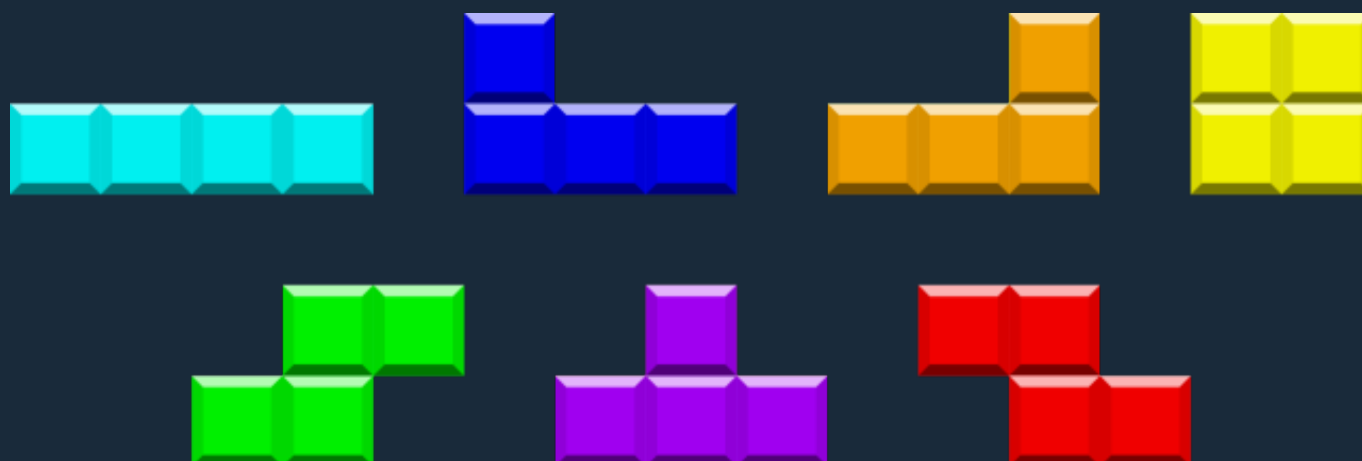
Wednesday, December 9.

Scientific Networks and Success

An (online) Satellite Workshop of CCS 2020

Successful Configurations — When Form Follows Function



Dr. Alexander M. Petersen

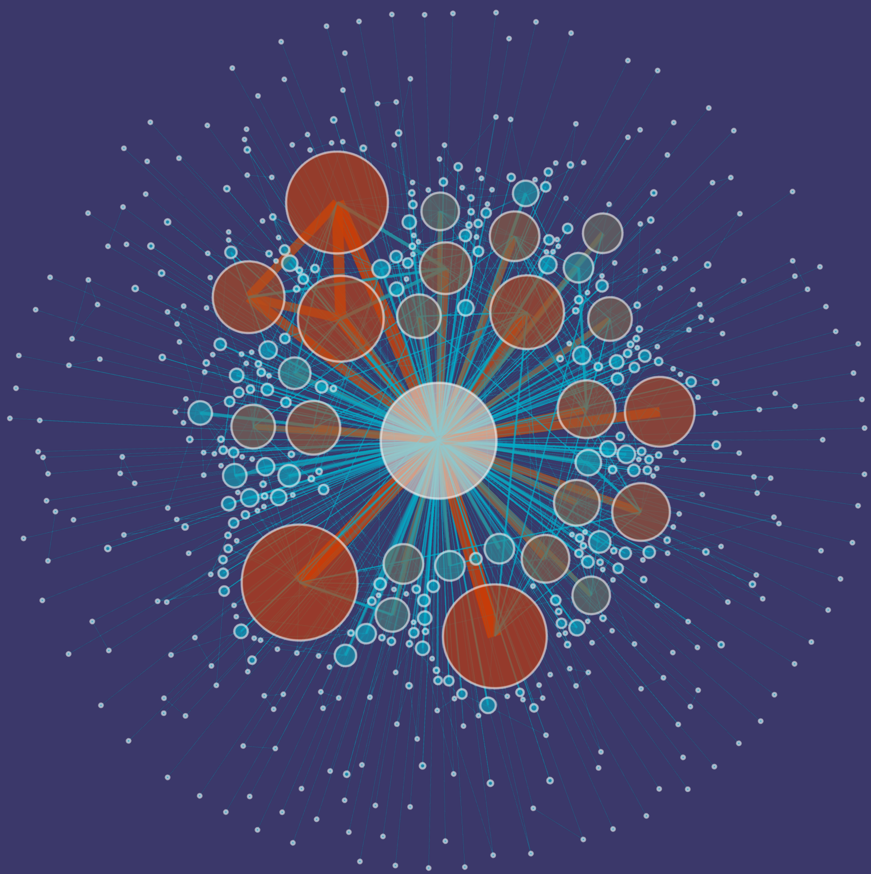
Management of Complex Systems Department

Ernest & Gallo School of Management*

UNIVERSITY OF CALIFORNIA MERCED

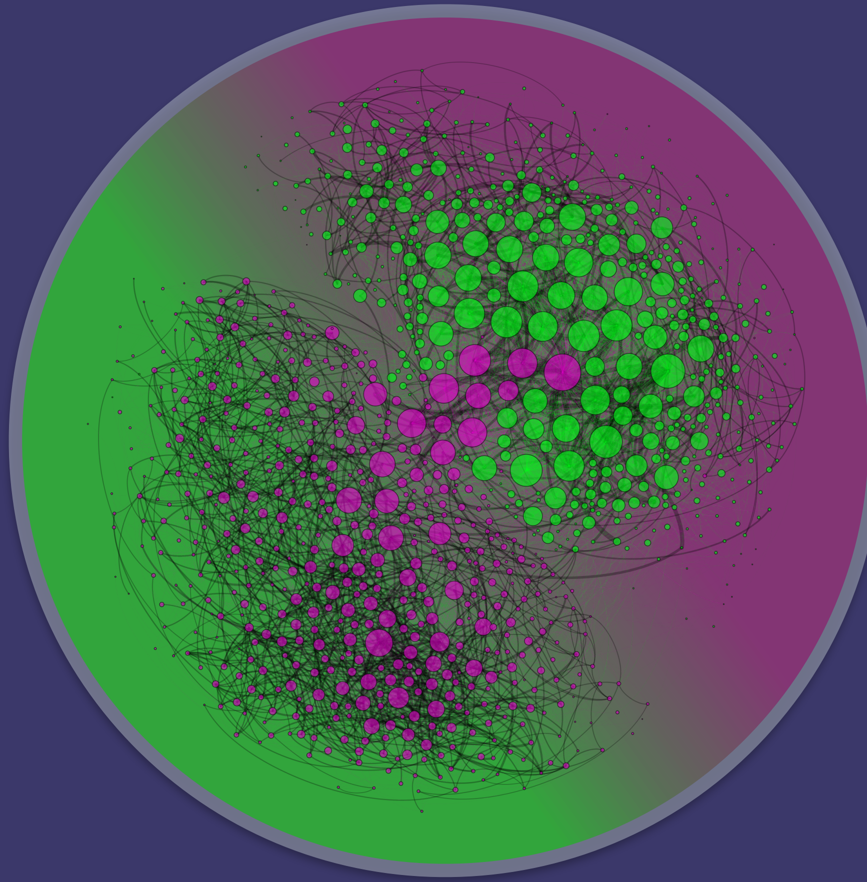
Quantifying the impact of "super ties" and cross-disciplinary configurations in scientific careers

Individual Collaboration Network
— Paul Erdös—



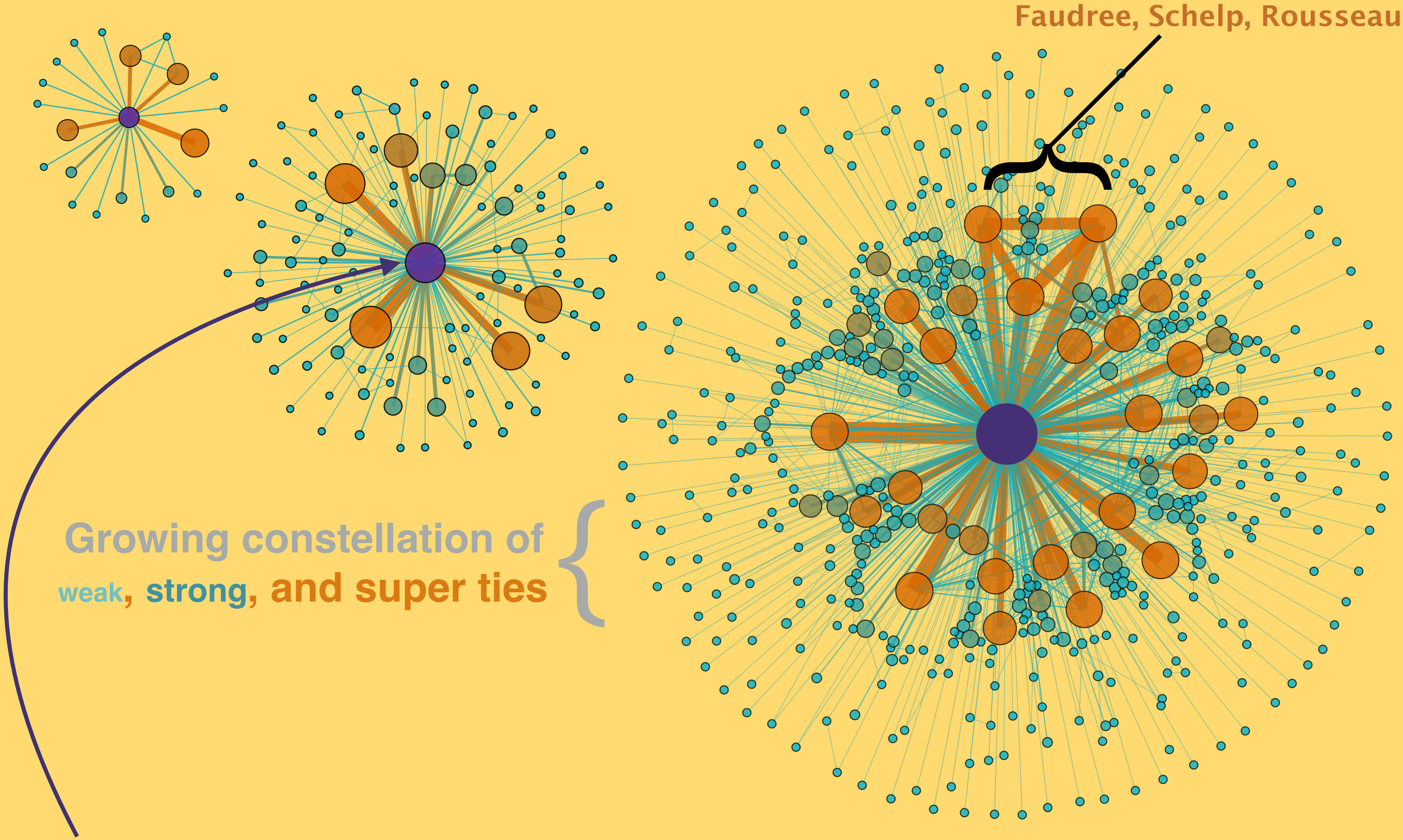
Part I

Cross-Disciplinary Collaboration Network
— Genomics Revolution —



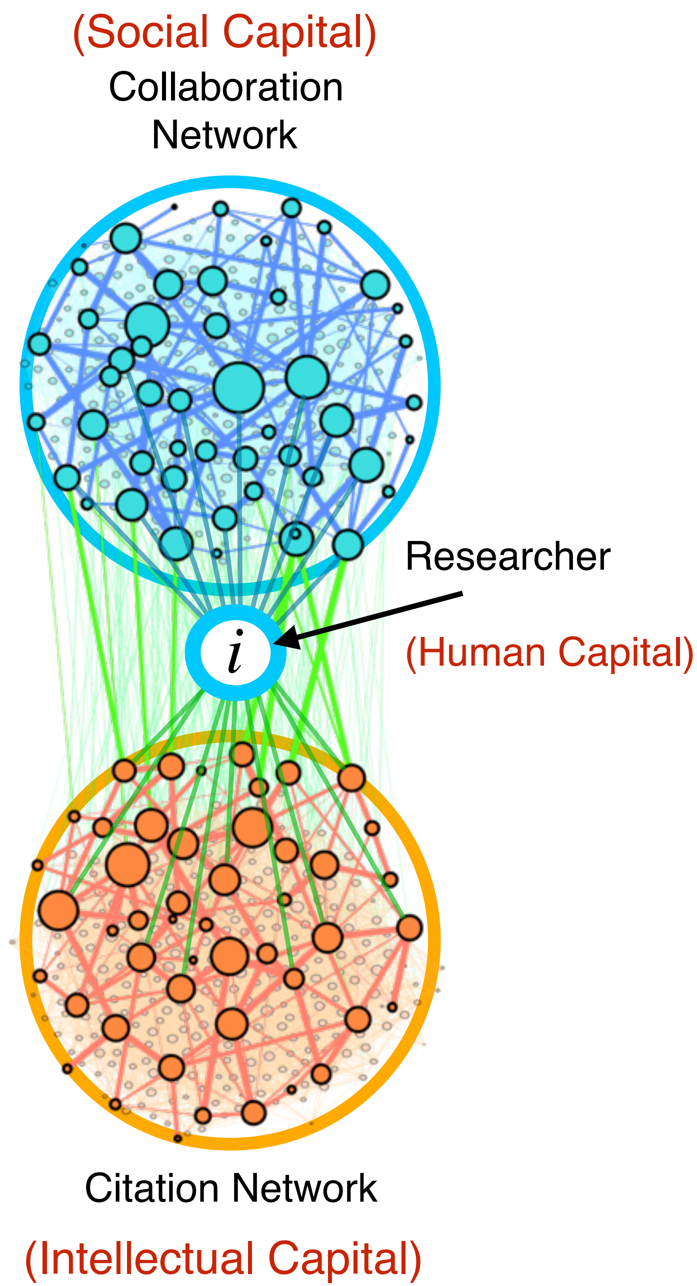
Part II

Part I — Quantifying the impact of weak, strong, and super ties in scientific careers — PNAS (2015)



Paul Erdös (1913-1996): collaboration network at career age 10, 30 years & present day*

Science careers are embedded in a co-evolving network of networks



coevolutionary system:

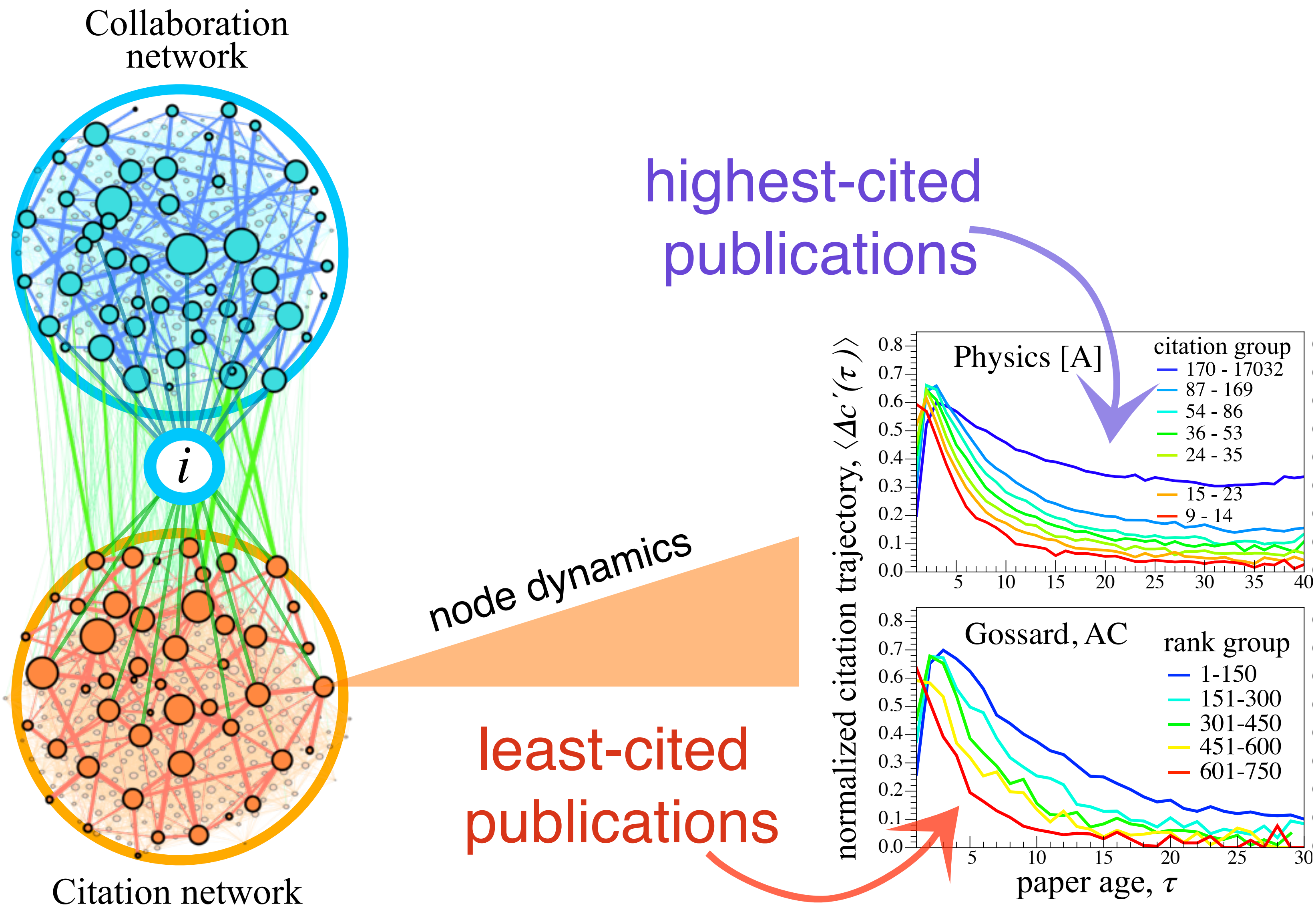
- knowledge
- institutions
- researchers

social phenomena:

- behavioral aspects
- economic incentives
- cumulative advantage mechanisms
- collaboration / competition

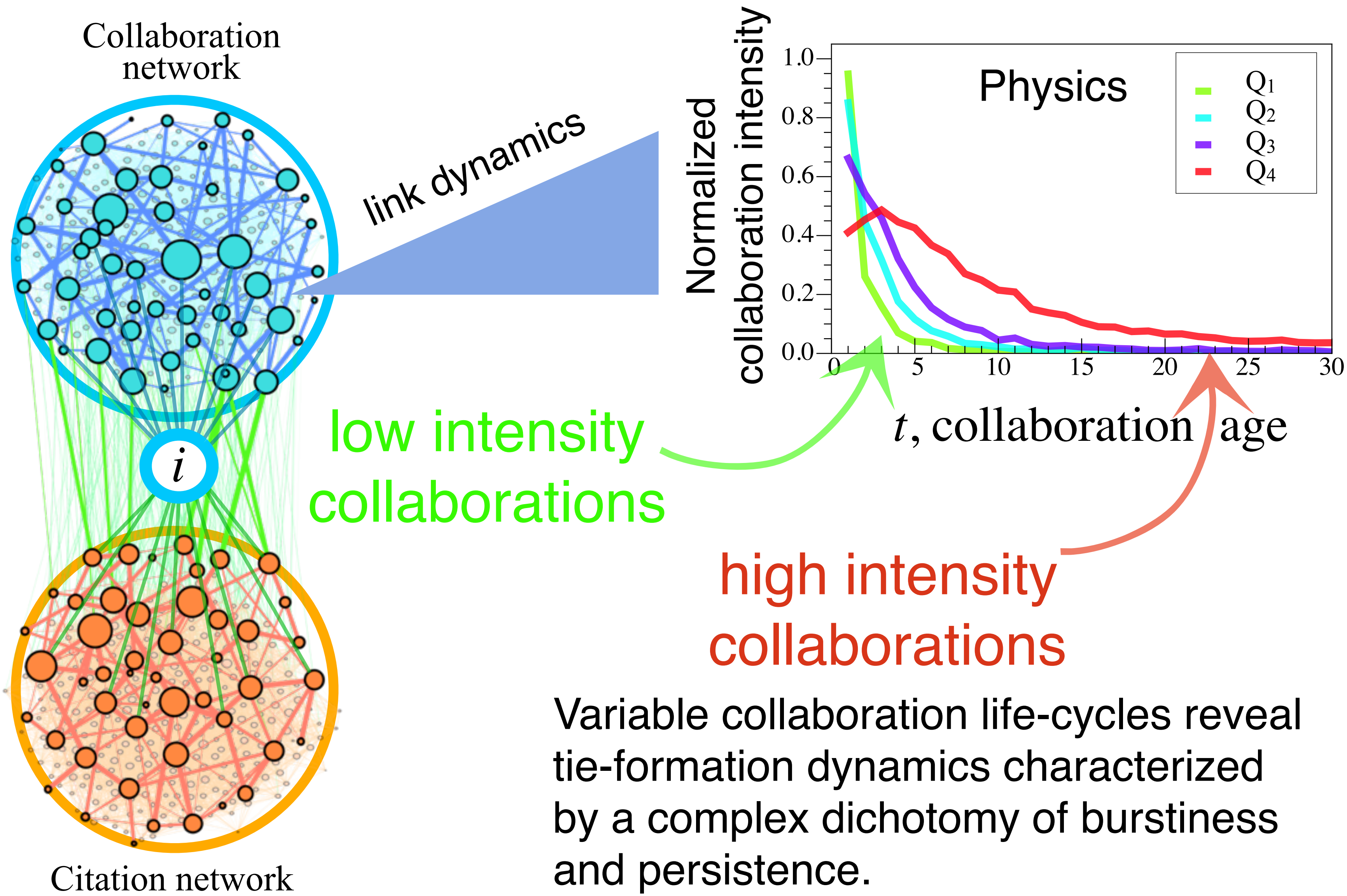
Dynamic network characterized by life-cycles

Reputation and Impact in Academic Careers — PNAS (2014)



Dynamic network characterized by life-cycles

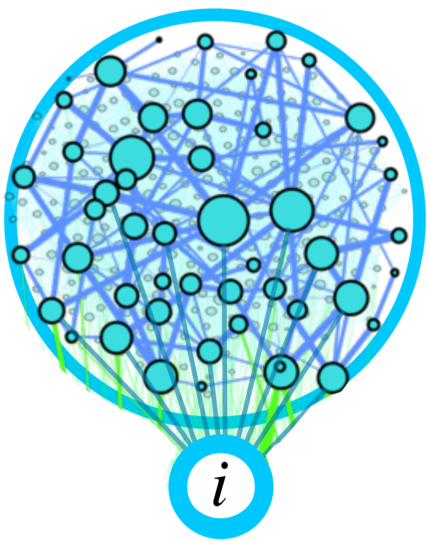
Quantifying the impact of weak, strong, and super ties in scientific careers — PNAS (2015)



Variable collaboration life-cycles reveal tie-formation dynamics characterized by a complex dichotomy of burstiness and persistence.

An ego-centric perspective reveals a wide range of collaboration strategies

Collaboration network



central author i

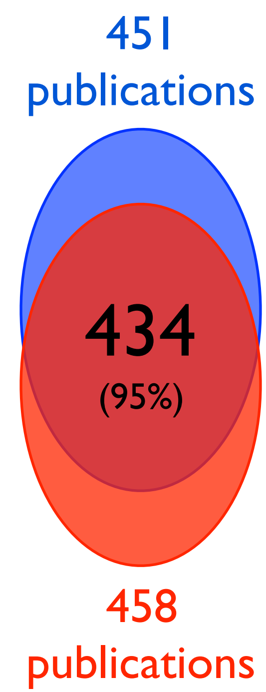
Interactions mediated by social “forces”:

- Collaboration (attractive)
- Competition (repulsive)
- Knowledge (an “exchange particle”)

Binary-star strategy:

* Michael Stuart Brown
* Joseph L. Goldstein

Recipients of the 1985 Nobel Prize in Physiology or Medicine for describing the regulation of cholesterol metabolism.

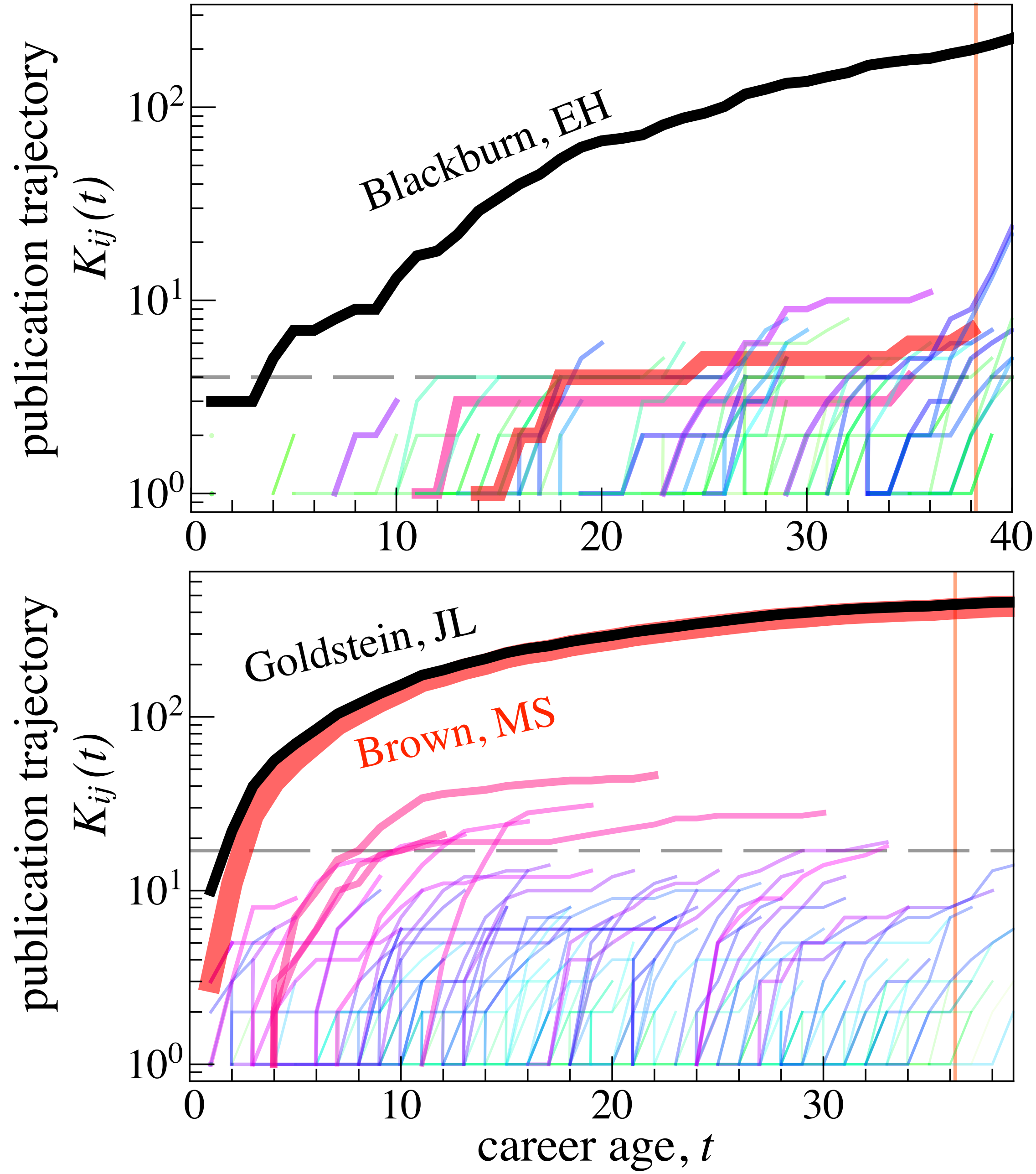


Solo-artist strategy:

* **Marilyn Kozak** (also cell biologist)

$N = 70, N_{solo} = 59$

Wide variation in the temporal collaboration profiles — even among Nobel Laureates —



Ego collaboration network: quantifying *dynamic* & *heterogenous* patterns of collaboration within scientific careers

Sir Andre K. Geim

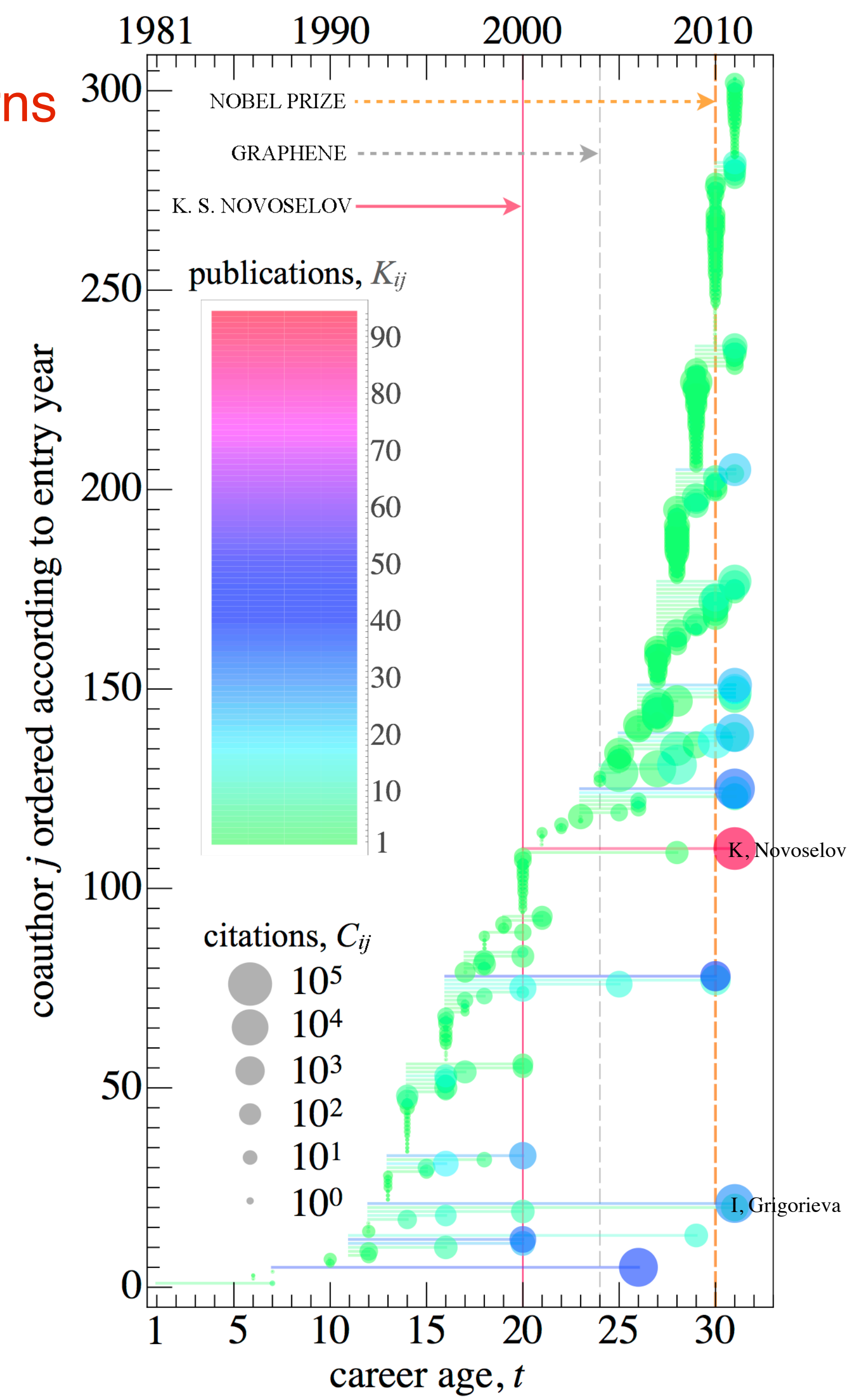
 # publications, N_i (2012) = 217

S_i = 303 coauthors

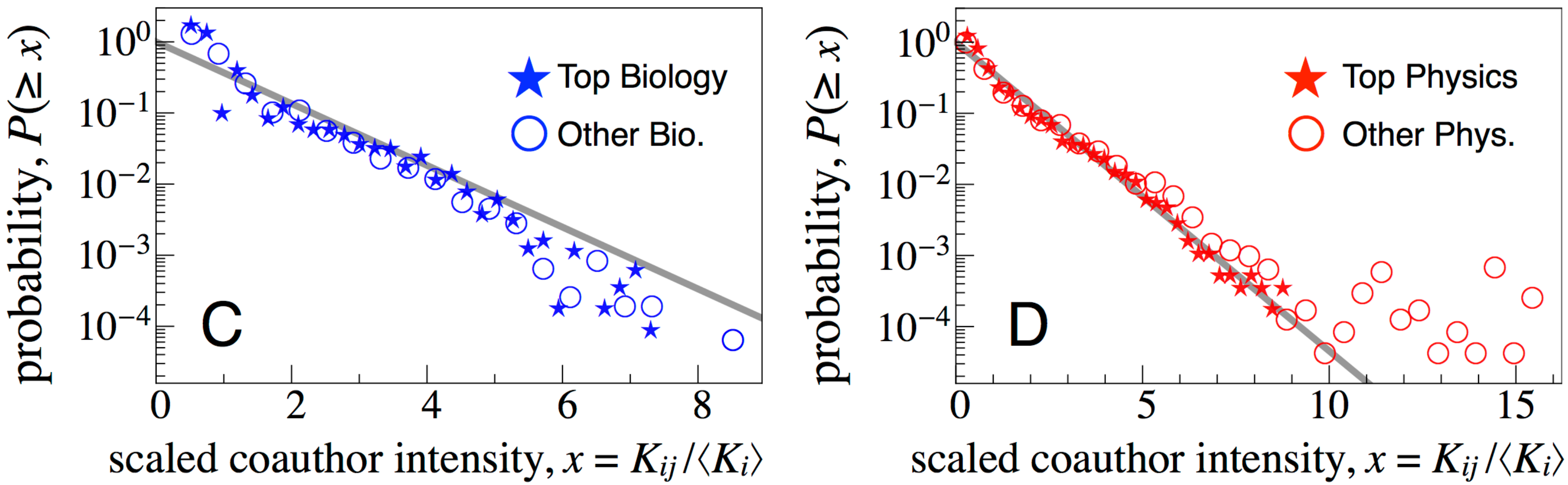
The average copublication duration

$\langle L_i \rangle$ = 2.1 years, $\langle K_i \rangle$ = 3.7 pubs.

- Measuring the duration L_{ij} of the tie (time b/w 1st and last copublication)
- Measuring the intensity K_{ij} of the tie (# of copublications)
- Measuring the net scientific impact C_{ij} of the tie (net citation tally for pubs. between i and j)



Is there a characteristic collaboration intensity scale?



In order to aggregate across careers with varying coauthorship patterns, we use the normalized variable $x = K_{ij}/\langle K_i \rangle$

$P(\geq x)$ is well-described by an exponential distribution, for which there is a closed-form solution to the extreme value equation:

$$1/S_i = \sum_{K_{ij} > K_i^c}^{\infty} P(K_{ij}) = \exp(-\kappa_i K_i^c)$$

which has the simple solution

"super tie" threshold $K_i^c = (\langle K_i \rangle\text{-}1)\ \mathrm{Ln}(S_i)$

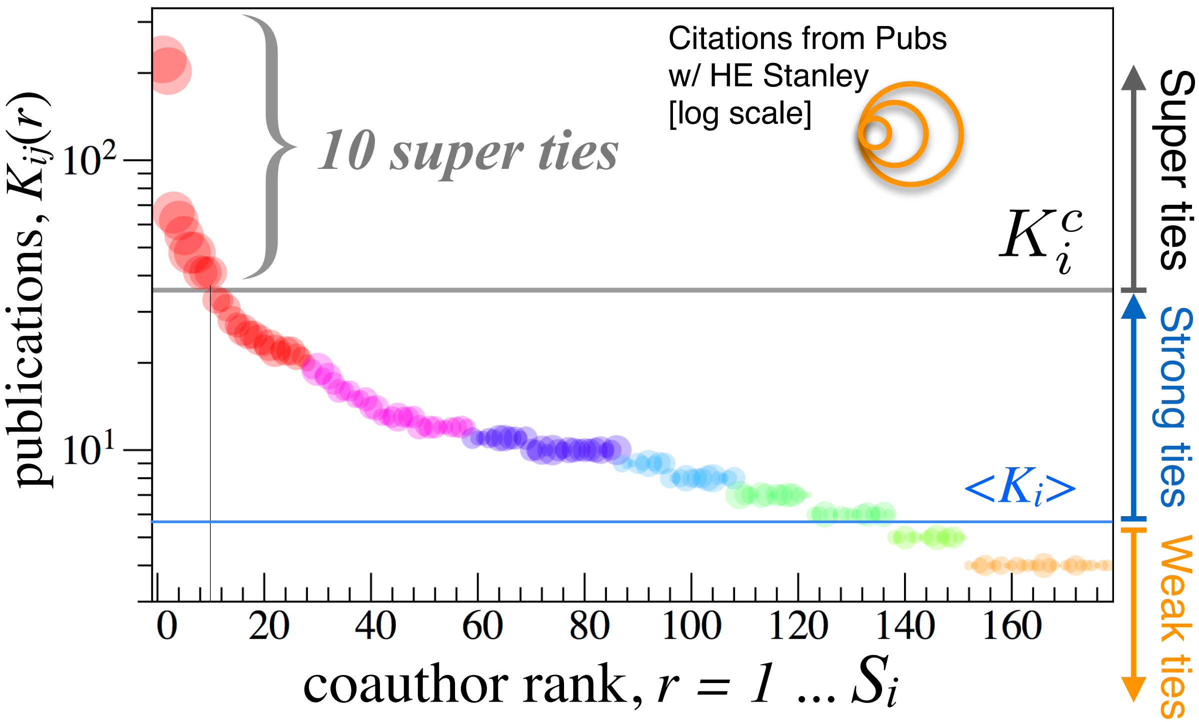
Weak ties, Strong ties, and Super ties

H. Eugene Stanley

$N_i\,(2010) = 909$ *publications*

$S_i = 541$ *coauthors*

$\langle K_i \rangle = 6.7$ *papers*



rank		K_{ij}	%
1	HAVLIN, S	223	25
2	BULDYREV, SV	203	22
3	AMARAL, LAN	66	7
4	SCIORTINO, F	62	6
5	IVANOV, PC	55	5
6	GOLDBERGER, AL	48	
7	PENG, CK	48	
8	GOPIKRISHNAN, P	41	
9	PLEROU, V	41	
10	STARR, FW	41	
11	DOKHOLYAN, NV	33	4
12	PAUL, G	33	
13	BUNDE, A	31	3
14	GIOVAMBATTISTA, N	28	
15	MAKSE, HA	27	
16	CONIGLIO, A	26	
17	URBANC, B	25	
18	CRUZ, L	25	
19	SCALA, A	24	
20	LARRALDE, H	23	
21	MANTEGNA, RN	23	
22	POOLE, PH	22	2

Extreme outlier based upon the exponential distribution:

"super tie" threshold $K_i^c = (\langle K_i \rangle - 1)\, Ln(S_i)$

Data & Measures

Data from Clarivate Analytics Web of Science: 473 researcher profiles spanning more than 15,000 career years, 94,000 publications, and 166,000 collaborators.

Researcher Profiles: split into 4 groups: top-cited biology, not top-cited biology, top-cited physics, and not top-cited physics

Collaboration Tie Measures

— Strength —

K_{ij} **Individual level**: How strong/weak is the collaboration tie?

$a_{i,p}$, $\overline{a}_{i,t}$ **Team level:** How big is the team?

$G^K_{t,i}$ **Group level**: How concentrated are the tie strengths?

— Duration —

$\langle L_i \rangle$ **Individual career level**: What is the characteristic collaboration length?

$\overline{L}_{i,t}$ **Team level**: What is the team's experience together?

Is there a citation advantage associated with Super Ties?

On average:
- 1 in 25 collaborators qualify as a super-tie
- 1 in 2 publications include a super-tie

Unit of analysis : publication p

Hierarchical "fixed effects" model : 473 researchers indexed by i

Dependent variable = $z_{i,p}$ = the citation impact $c_{i,p,y}$ of publication p normalized to baseline citation levels defined by other papers published in the same year y.

$$z_{i,p} = \frac{(\ln c_{i,p,y} - \langle \ln c_y \rangle)}{\sigma[\ln c_y]}$$

This measure maps $c_{i,p,y}$ to a stable normal distribution N(0,1) >> appropriate for comparing citation impact across time.

$R_{i,p}$ A super-tie indicator variable = 1 if at least one of the coauthors is a super tie, and 0 otherwise. 52% of publications have R=1.

$a_{i,p}$ number of coauthors ≈ proxy for coordination costs and technology level

t_p publication year of p, measured as a career age, accounting for aging and cumulative advantage effects, learning and prestige

$N_i(t_p)$ number of papers up to year t_p ≈ prestige measure

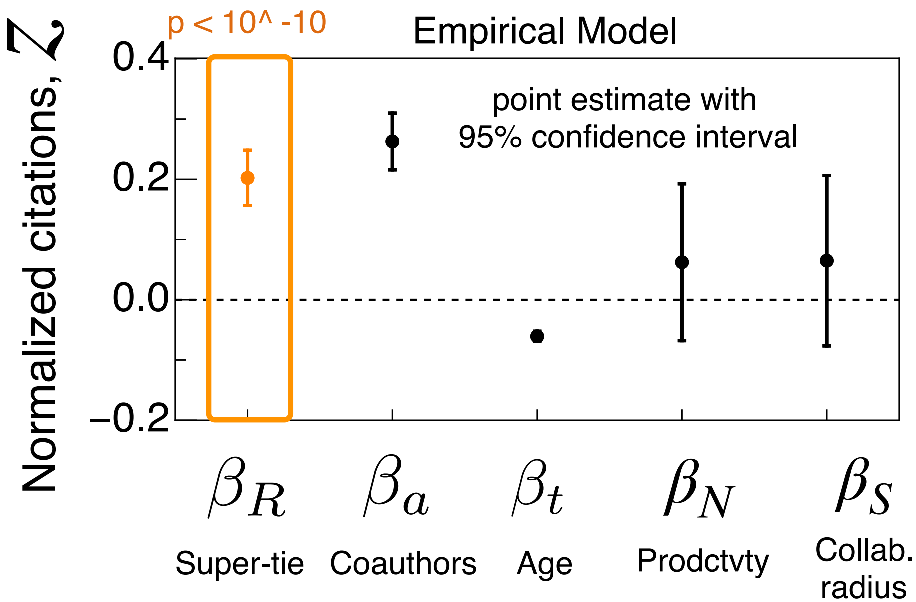
$S_i(t_p)$ number of distinct coauthors up to year t_p ≈ collaboration radius measuring access to new/old team members

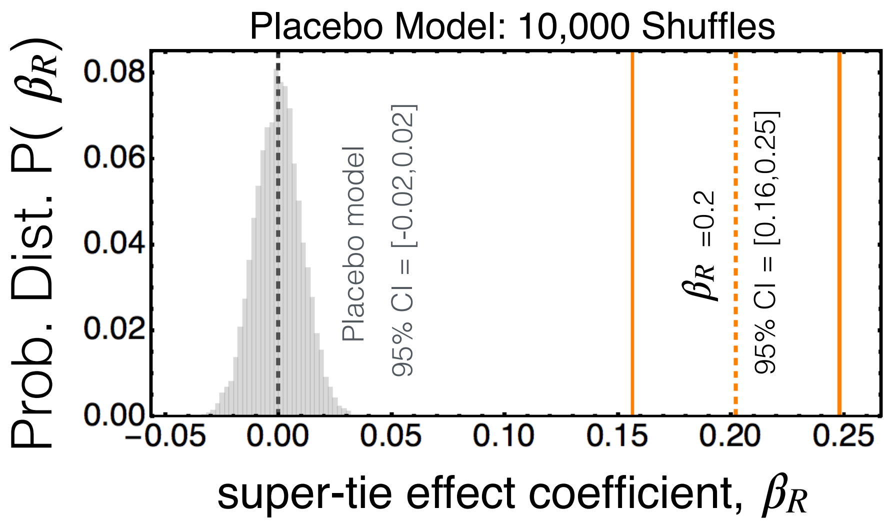
Fixed-effects model - measures each researcher against his/her baseline $z_{i,p}$

$$z_{i,p} = \beta_R R_{i,p} + \beta_a \ln a_{i,p} + \beta_t t_{i,p} + \beta_N \ln N_i(t_p) + \beta_S \ln S_i(t_p) + \beta_i + \epsilon_{i,p}$$

Strategic value of high-intensity collaborations

Emphasizes ***who*** in addition to ***how many*** coauthors

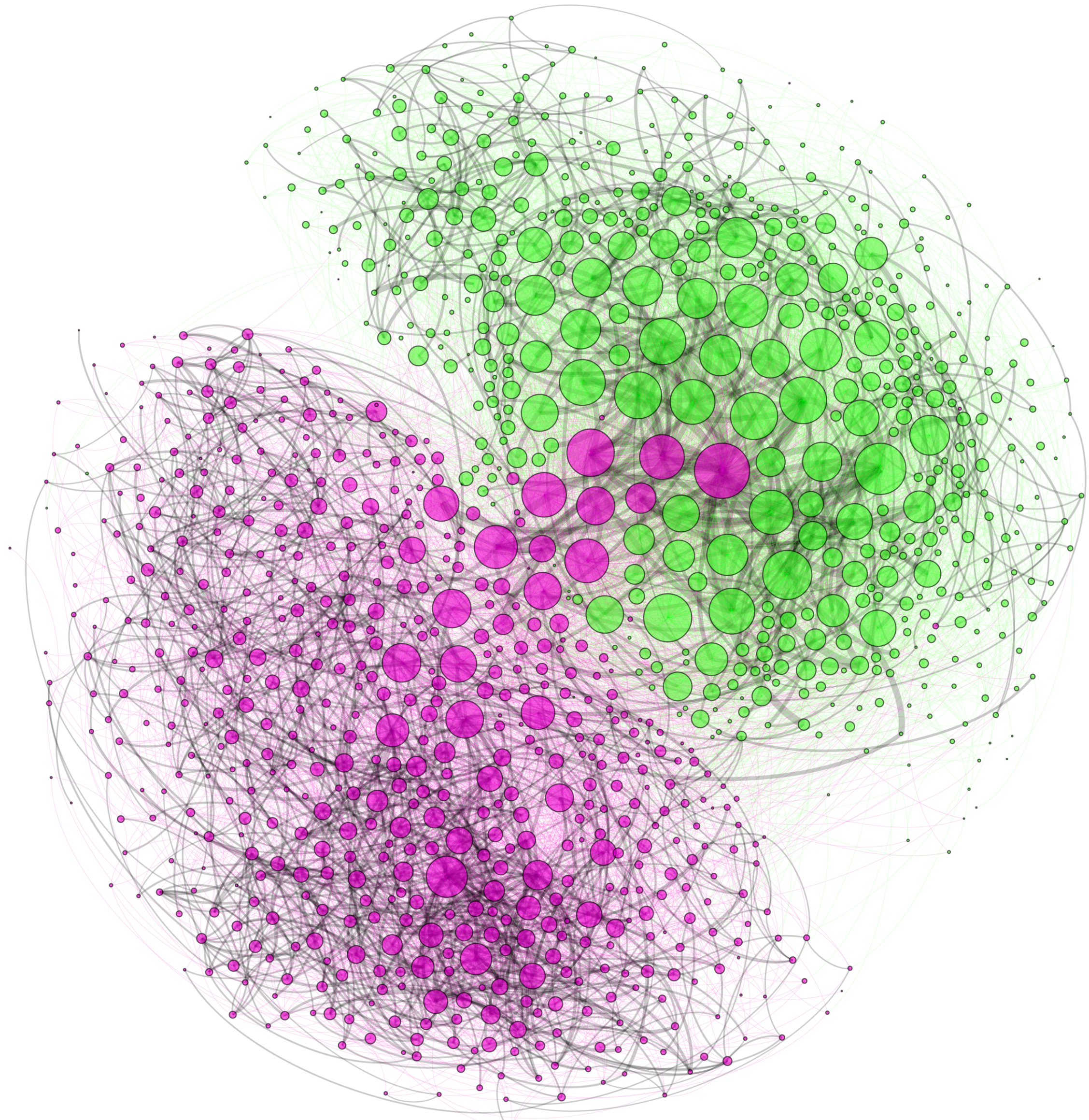




In terms of real citation impact: 100 x 0.2 σ_z corresponds to ~ 20% citation increase at the publication level (relative to the author's own mean baseline)!

Plausible explanations: compounding self-citations, reputation arising from larger formal and informal social network; added value of skill complementarity, trust, conviction, commitment, experience, collocation, moral support, risk-profit sharing

Part II — Cross-disciplinary Evolution of the Genomics Revolution — Science Advances (2018)



Data & Methods: ~80 US Biology and Computing Departments faculty directories ⇒ List of Scholars

— we then collected data from their 4,190 Google Scholar profiles, comprising 413,565 publications

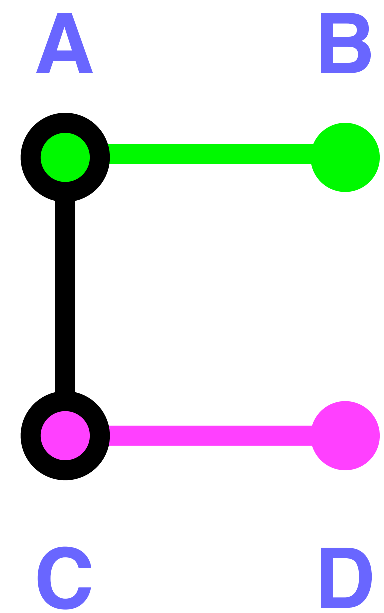
Author i	Coauthors	Department $\mathcal{F}_i$	Orientation $\mathcal{O}(\mathcal{F}_i)$
A	**B,C**	***BIO***	***XD***
B	**A**	***BIO***	***BIO***
C	**A,D**	***CS***	***XD***
D	**C**	***CS***	***CS***

Direct link: publication between scholar i and j

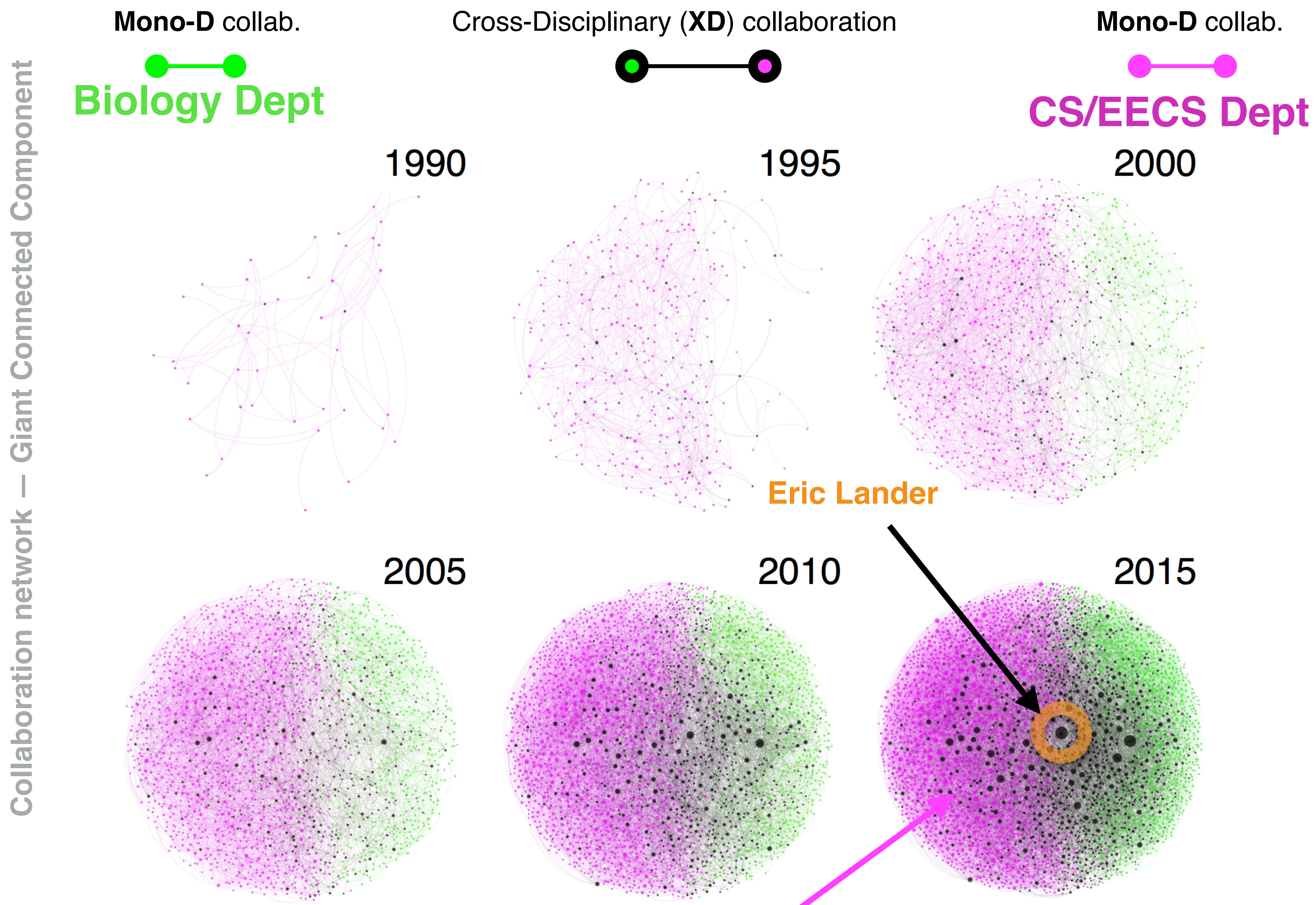
Mono-Disc. scholar : $\mathcal{O}_i(\mathcal{F}) = \mathrm{BIO}$

Mono-Disc. scholar : $\mathcal{O}_i(\mathcal{F}) = \mathrm{CS}$

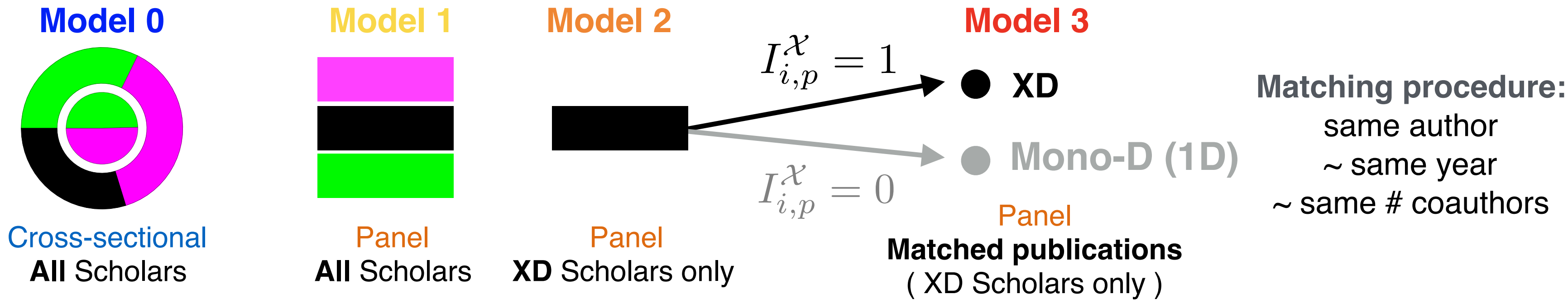
Cross-Disc. scholar : $\mathcal{O}_i(\mathcal{F}) = \mathcal{X}$



Longitudinal Case Study of the Genomics Revolution (HGP, 1990-2003)

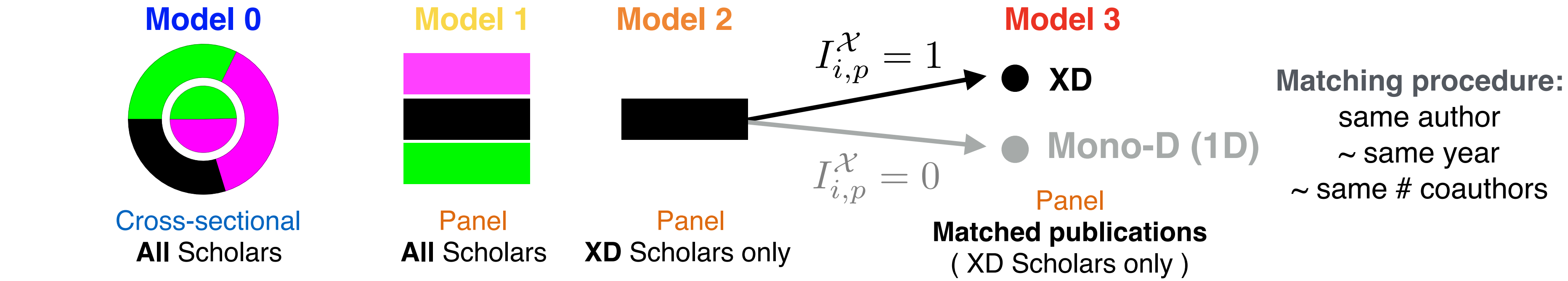


Work in collaboration with Professor Ioannis Pavlidis, Dept. of Computer Science — University of Houston



Panel model specification:
(w/ Author Fixed Effects)
Unit of Analysis = **Publication**

$$\underbrace{z_{i,p}}_{\text{Normalized Citation impact}} = \beta_i + \beta_a \underbrace{\ln a_{i,p}}_{\text{\# coauthors}} + \beta_\tau \underbrace{\tau_{i,p}}_{\text{Career year}} + \beta_I \underbrace{I^{\mathcal{X}}_{i,p}}_{\text{XD indicator}} + \underbrace{\mathrm{D}_t}_{\text{Year dummy}} + \epsilon_{i,p}$$

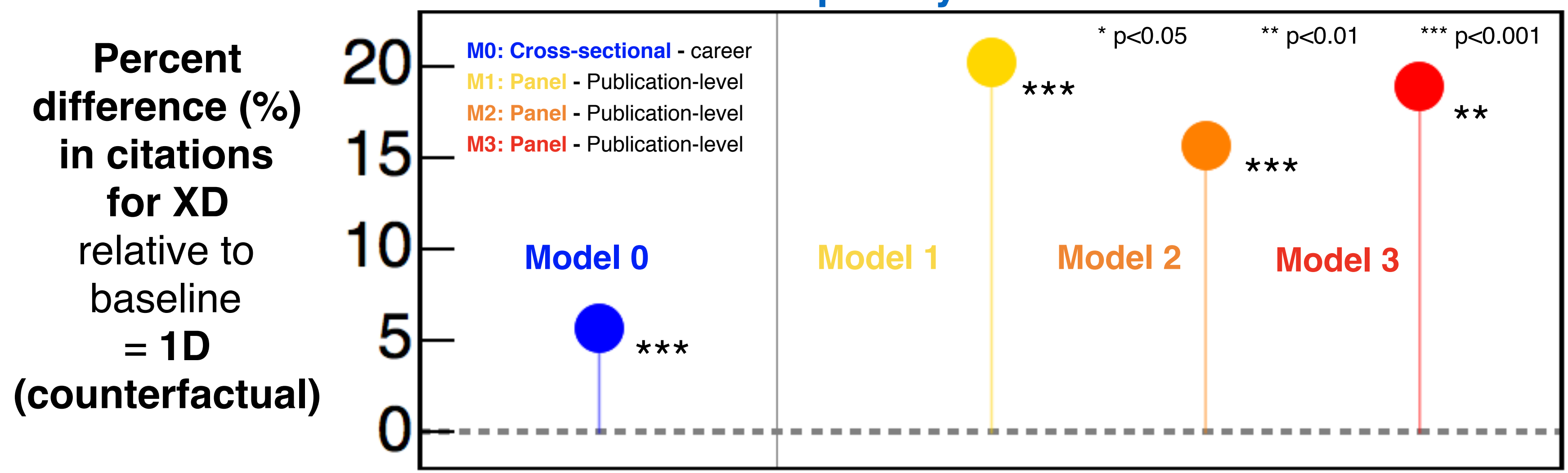


Panel model specification:
(w/ Author Fixed Effects)
Unit of Analysis = **Publication**

Normalized Citation impact; # coauthors; Career year; XD indicator; Year dummy

$$z_{i,p} = \beta_i + \beta_a \ln a_{i,p} + \beta_\tau \tau_{i,p} + \beta_I I^{\mathcal{X}}_{i,p} + \mathrm{D}_t + \epsilon_{i,p}$$

Cross-disciplinary Citation Premium



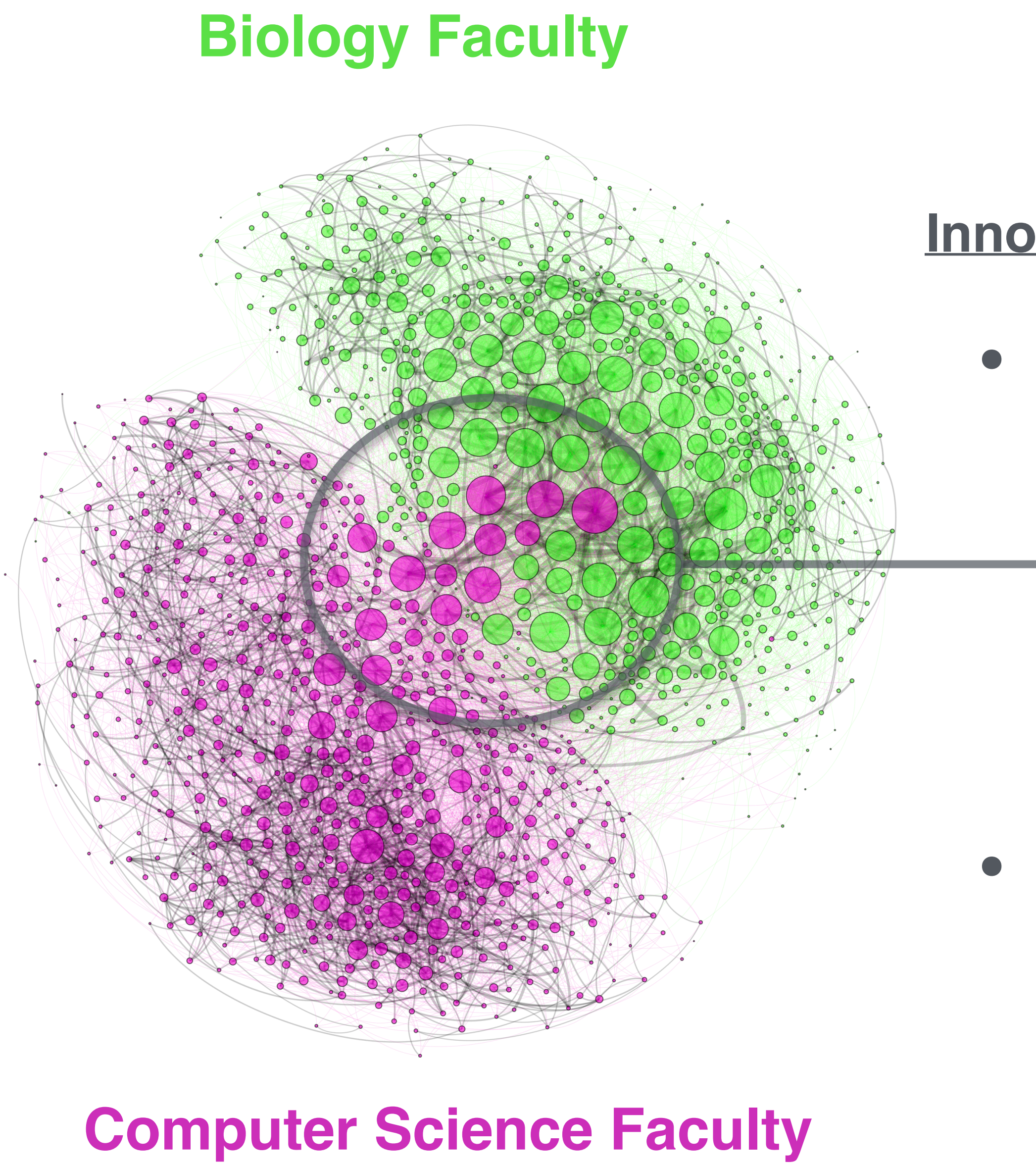
Coefficient estimates relation between: $\beta_{\mathcal{X}}$ = Career citations and Fraction of coauthors that are XD

β_I = Article citations and XD coauthors [coauthors from both BIO and CS, $I^{\mathcal{X}}_{i,p} = 1$]

Scholars with 10% XD-Collaborators are cited ~ 6% more than 1D Scholars from the same discipline

Articles featuring cross-disciplinary combination of authors are cited ~20% more than 1D articles by same author

Human Genome Project — a cross-disciplinary bridge facilitating a highly functional marriage

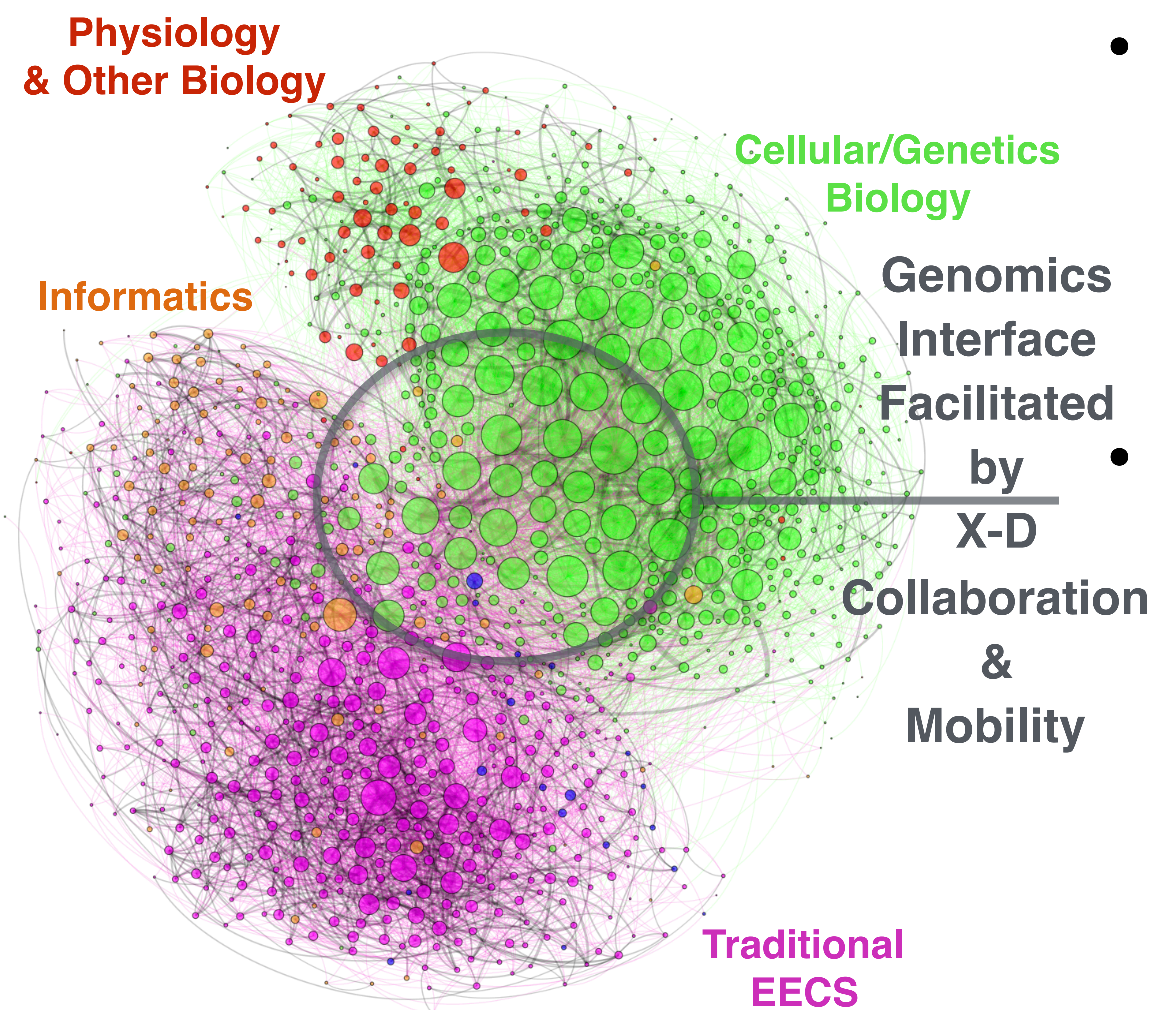


Innovation @ the genomics interface

- **Success factors:**
  - Methodological diversity leveraging common language
  - Cultural assimilation: XD collaboration facilitates XD mobility of CS into elite BIO
- **Outcomes:**
  - Transformative research
  - Flagship program model
  - Consortium model — *teams of teams*

Disciplinary Propensity revealed by Scholar-Scholar interactions

Network community structure



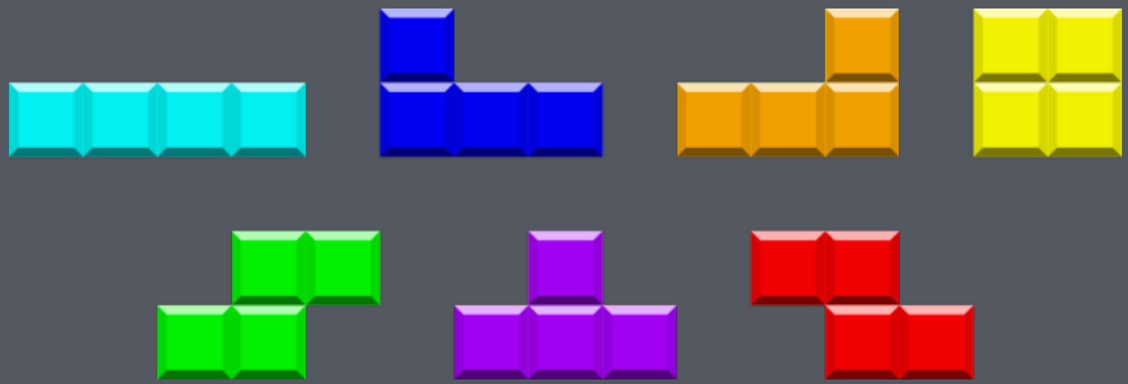
Implications for Funding Policy/Design

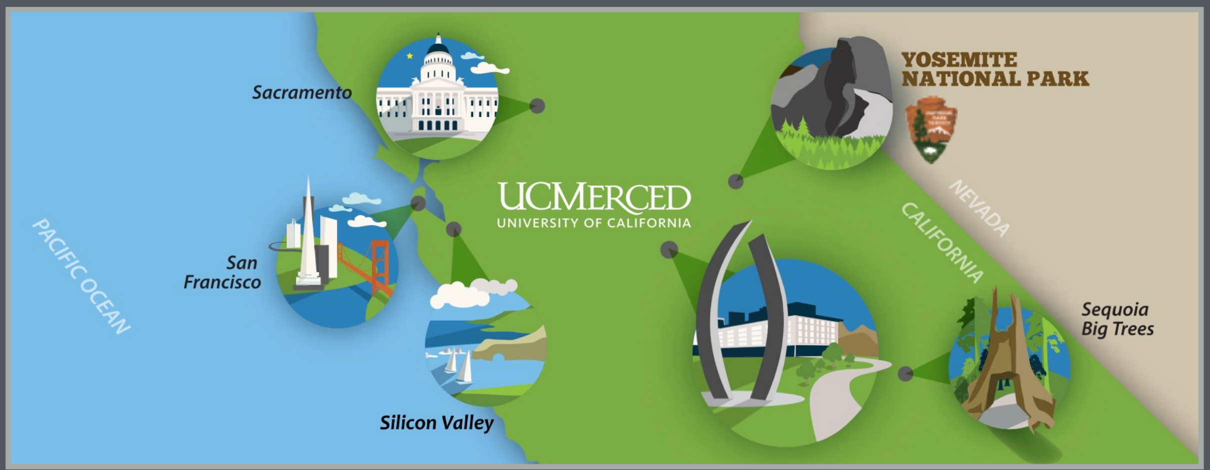
- **Flagship Programs:** funding around Grand Challenges may reduce the barriers associated with disciplinary borders, thereby **incentivizing cross-disciplinary collaboration & mobility**
- ***"Consortium Science"*:** ***teams of teams*** coalesce with common objectives, including sharing benefits equitably within and beyond institutional boundaries — **an organizational model championed by the HGP and further developed by numerous follow-up *"Omics"* consortiums**

Cross-disciplinary Evolution of the Genomics Revolution — Science Advances (2018)

Successful Configurations — when *Form Follows Function*

Warming and earlier spring increase western US forest wildfire activity.
— AL Westerling et al., Science 2006



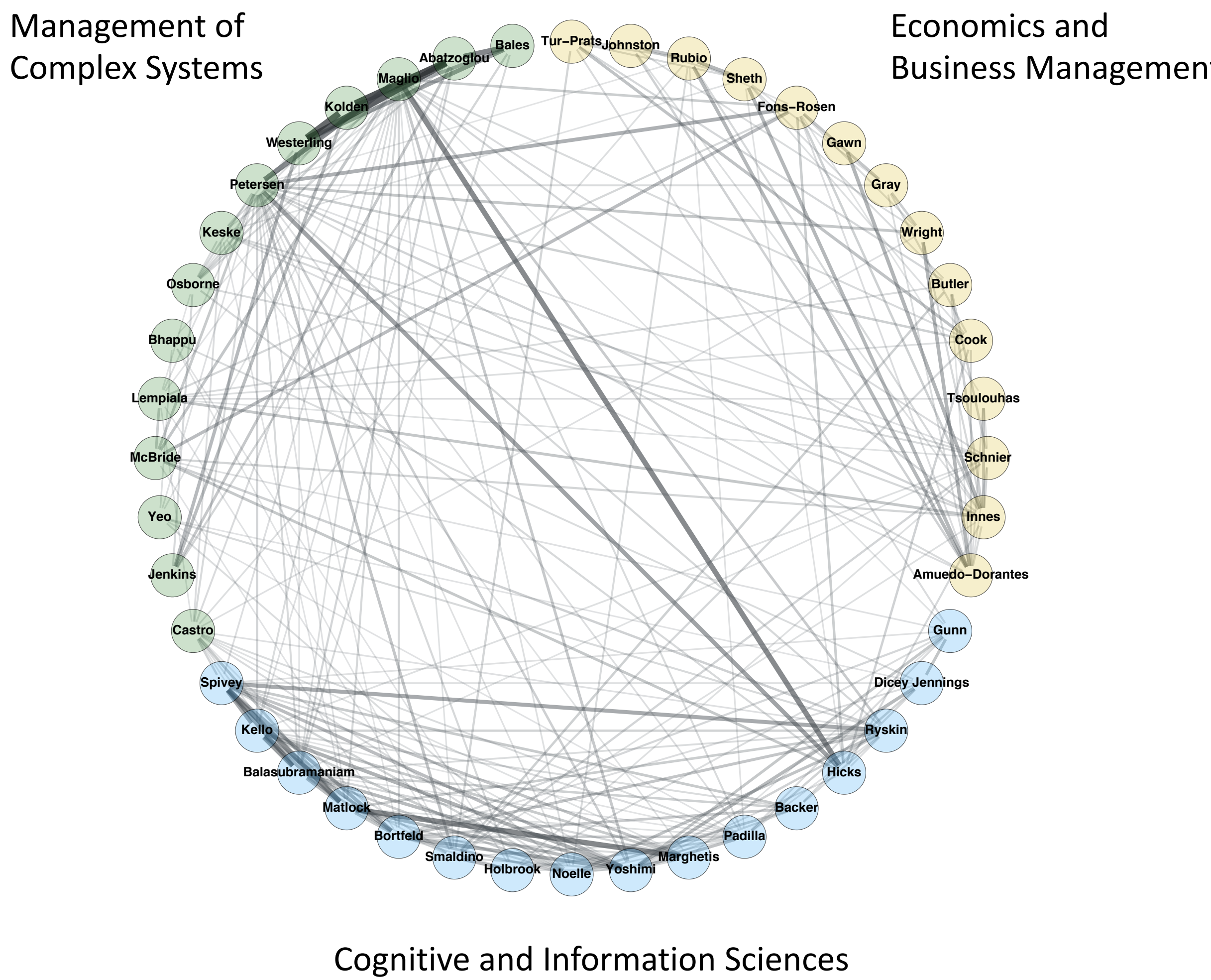


Management of Complex Systems Department — mcs.ucmerced.edu
Ernest & Gallo School of Management*

UCM Gallo Management School Initiative

A configuration designed for addressing 21st Century ***Grand Challenges***
emerging at the intersection of Coupled Human and Environmental Systems

Research



Faculty Network illustrating faculty by department (nodes), and commonality in research interests (indicated by links);
Giant Sequoia — a symbol of Resiliency & Robustness

Thanks for your attention!

and also to my esteemed collaborators in this and related work
— in particular Ioannis Pavlidis @ University of Houston —

Quantifying the impact of weak, strong, and super ties in scientific careers
PNAS (2015) — Petersen

Cross-disciplinary evolution of the genomics revolution
Science Advances (2018) — Petersen, Majeti, Kwon, Ahmed, Pavlidis

Come and explore cross-disciplinary opportunities at University of California, Merced

ERNEST & JULIO
GALLO
MANAGEMENT PROGRAM
INNOVATION • SUSTAINABILITY • TECHNOLOGY