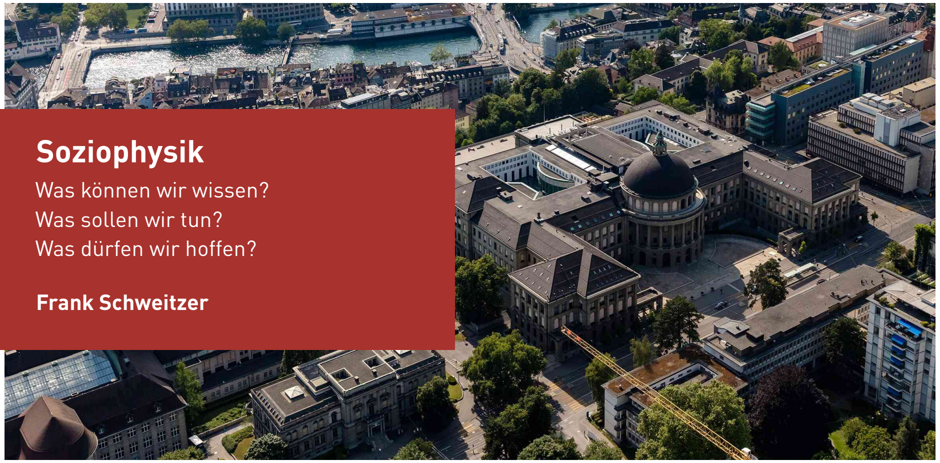
ETH zürich

Chair of Systems Design
www.sg.ethz.ch



Soziophysik

Was können wir wissen?
Was sollen wir tun?
Was dürfen wir hoffen?

Frank Schweitzer

Outline

Was können wir wissen?

Was sollen wir tun?

Was dürfen wir hoffen?

Drei Fragen an die Soziophysik

Was können wir wissen? [3]
- Beiträge der Physik zur Sozialwissenschaft
- **positiv:** Modellansätze, Forschungsfragen

Was sollen wir tun? [2]
- Respekt vor Modellen, Systemen, Paradigmen
- **skeptisch:** Fehler, falsche Erwartungen

Was dürfen wir hoffen? [1]
- Beispiele datengetriebener Modellierung
- **optimistisch:** neue Methoden, neue Probleme

Immanuel Kant (1724-1804)

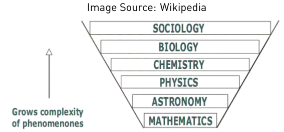

Critik der Urtheilskraft von Immanuel Kant.

[1] Kritik der reinen Vernunft (1782), [2] Kritik der reinen Vernunft (1788), [3] Kritik der Urteilskraft (1793)

Physics $\equiv$ Success

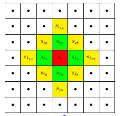
- **18th century: A new science of man**
  - David Hume (1711-1776): "A Treatise of Human Nature"
    - mathematics, physics as paragon
- **19th century: Measurements, laws**
  - Adolphe Quetelet (1796-1874): "Essays on social physics"
    - *statistical laws*: BMI, mortality
  - Auguste Comte (1798-1857): "Sociology"
    - society follows *general laws*
    - *positivism:* social science builds on experience
  - Karl Marx (1818-1883): "The Capital"
    - objective *"laws of motion"* of capitalist system
- **20th century: The "Observer"**
  - Relativity theory
  - Quantum mechanics

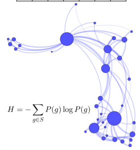
Image Source: Wikipedia

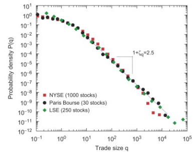


Contributions from physics

1. **Generic models for collective phenomena**
   - spin systems, lattice models
   - complex systems, emergent properties
   - agent-based models (social/biological systems ⇒ *"active matter"*)
2. **Statistical ensembles**
   - phase transitions, order/control parameters
   - grand-canonical e. ⇒ Exponential Random Graph models (ERGM)
   - entropy for social systems, quantify *potentiality*
3. **Universal scaling laws**
   - Why should different systems obey universal behavior?
   - empirical truth: "Stylized Facts" (N. Kaldor, 1961)
   - remember Quetelet: statistical regularities

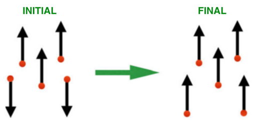




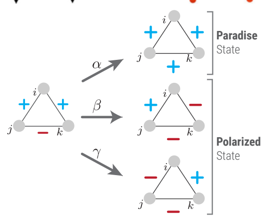


F.S., *An agent-based framework of active matter with applications in biological and social systems*, **European Journal of Physics**, vol. 40, 014003 (2019)
F.S., *The law of proportionate growth and its siblings: Applications in agent-based modeling of socio-economic systems*, In: **Complexity, Heterogeneity, and the Methods of Statistical Physics in Economics** (Eds. H. Aoyama, Y. Aruka, H. Yoshikawa), Springer (2020), 145-176
F.S., Social percolation revisited: From 2d lattices to adaptive networks, **Physica A** (2020)
C. Zingg, G. Casiraghi, G. Vaccario, F.S.: What is the Entropy of a Social Organization? **Entropy** (2019)

Example: Opinion Dynamics - Consensus or Coexistence



- **Ising model:** ferro/antiferromagnetic phases
  - 1st cellular automaton ⇒ Lattice models
- **Voter model:** random copy of another opinion
  - Axelrod model of *cultural dissemination*: homophily
- **Structural balance**
  - frustrated systems, spin glasses



Problems
- *"Do humans behave much like atoms? Sociophysics ... answers in the affirmative."* (S. Galam 2012)
- Instead of **ad hoc assumptions** we need to **understand** mechanisms behind emergent phenomena

Example: Social forces

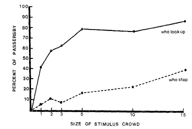
- **Social impact** (Milgram, 1969, Latane, 1981)
  - agent: **reactive**, not reflective
  - social forces: persuasion, support: $I_i(t) = I_i^p\,[P_i(t)] - I_i^q\,[Q_i(t)]$
  - transition probability: $p(-s_i|s_i) \propto \exp\{I_i/T\}$
- **Collective motion of humans**
  - *Brownian agent*: **reactive**, not reflective ⇒ Langevin dynamics

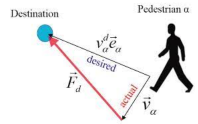
$$\frac{d\mathbf{v}_i(t)}{dt} = -\frac{1}{\tau_i}\mathbf{v}_i(t) + \mathbf{f}_i(t) + \sqrt{\frac{2\,\varepsilon_i}{\tau_i}}\,\xi_i(t)$$

  - "social force" model results in selforganized "behavior"

$$\mathbf{f}_i(t) = \frac{1}{\tau_i}v_i^0\mathbf{e}_i - \nabla_{\mathbf{r}_i}\left[V_B(|\mathbf{r}_i - \mathbf{r}_B^i|) + V_{\text{int}}(\mathbf{r}_i, t)\right]$$

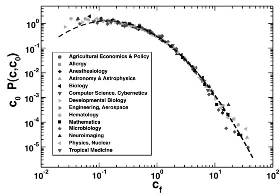
  - **applications:** evacuation scenarios, mass panic



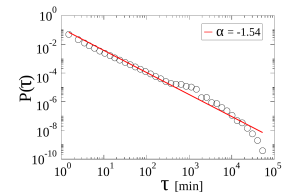


J. Holyst, K. Kacperski, F.S.: *Phase Transitions in Social Impact Models of Opinion Formation*, **Physica A** 285 (2000) 199-210
D. Helbing, F.S., J. Keltsch, P. Molnar: *Active Walker Model for the Formation of Human and Animal Trail Systems*, **Physical Review E** 56/3 (1997) 2527-2539

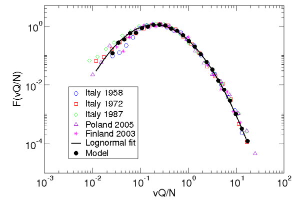
Example: Scaling laws in social systems



citations
Radicchi et al, 2008



communication
Garas et al, 2013



votes
Fortunato & Castellano, 2007

Finding: Surprise
- Complex social phenomena follow simple statistical laws (remember Quetelet)

Problem: Explanation
- Which **social mechanisms** lead to these "laws"?
- What is their **meaning?** How do they relate to **economic and social theory**?

Outline

Was können wir wissen?

Was sollen wir tun?

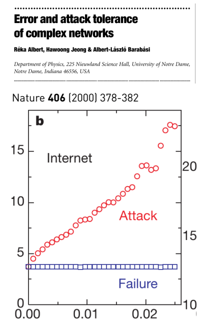
Was dürfen wir hoffen?

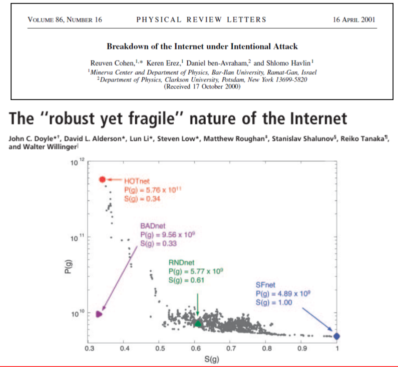
Respect!

- **Understand your models**
  - respect model *limitations*
  - *generative models* do not generate reality
  - *Null models:* what can be expected at random
- **Understand your system**
  - "Social" is what *deviates* from "physical"
  - What is the *eigendynamics*?
  - Adaptivity? Emergence of *new* systemic properties?
- **Understand your colleagues**
  - respect *different* scientific approaches
  - expect *high entry barriers* for interdisciplinary research
  - learn their language!

M. Gallegati, S. Keen, T. Lux, P. Ormerod (2006): Worring trends in econophysics, Physica A, vol. 370, 1-6
P. Ball (2006): Culture Crash, Nature, vol. 441, 686-688; P. Ball: Financial Times, October 29, 2006.
T. Di Matteo, T. Aste (2007): No Worries: Trends in Econophysics, Eur. Phys. J. B., vol 55, no 2

Example: Designed systems



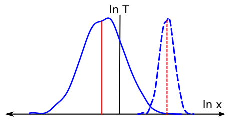


Problem: Wrong conclusions

- scale-free network → vulnerable against targeted attacks → Internet
- Ensemble based approach: **bad model** of real, engineered systems

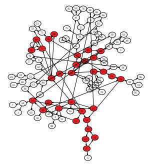
Example: Social systems are adaptive

- **Wrong Wisdom**



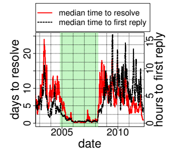
  - Wisdom of crowds: average opinion ≈ truth
  - social influence leads to consensus
  - subjectively: **right**, objectively: **wrong**

- **Collapsing core**



  - reciprocity improves core performance
  - collapse after peripheral nodes left

- **Efficiency and collapse**



  - manager improves performance
  - increased dependency leads to collapse

Unintended consequences

Systems **respond** to "improvements" in intended/unintended ways - **ALWAYS!**

Social science ≠ Social physics

- late 19th century: **Positivism** *(objective)*
  - use of quantitative methods, based on empirics
  - independence of subject (researcher) and object (phenomena)
- early 20th century: **Post-positivism** *(subjective)*
  - values, knowledge of researcher influences observation
  - remember modern physics: "The observer"
- late 20th century: **Anti-positivism** *(constructivist)*
  - social phenomena elude objective and reductionist methods (Weber, 1909)
  - focus: interpretation, discourse, *understanding*: "Verstehen" vs. "Erklären" social science/meaning ⇔ natural science/causal
  - Germany: "Positivismusstreit" (Popper/Adorno)
- early 21th century: **Akademie für Soziologie** (2017)
  - empirical/analytical sociologists break up

POSITIVISMUSSTREIT

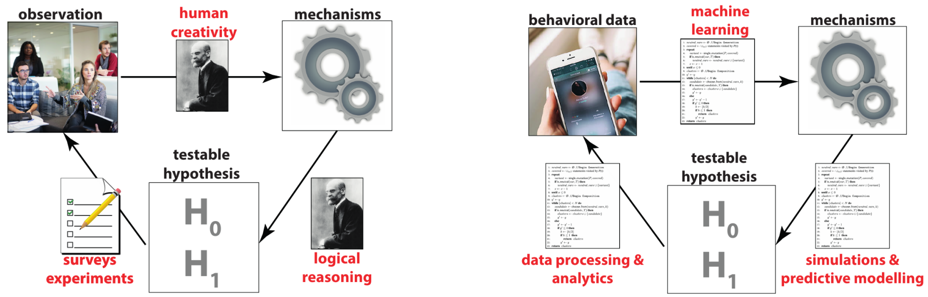
Ein Quexit in der Soziologie?

VON GERALD WAGNER

Ein neuer Positivismusstreit spaltet die Soziologie. Deuten steht gegen Messen. Der Konsens über Methodenvielfalt scheint aufgebrochen.

Max Weber (Wikipedia), G. Wagner, FAZ 2019

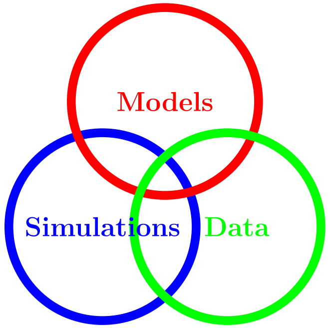
Computational social science: A change of paradigm



What goes wrong?

- **tools, no questions:** "Astronomy is not about telescopes"
- **data, no meaning:** "42 is the answer, but what was the question?"
- **observations, no theory:** "WHY do people do what they do?"

The Future of Computational Social Science



Emphasis on science ⇒ research questions

- **complements** existing (social) sciences

- not driven by available **data**
  - **aim:** *interpretation*, data has no "message"
- not driven by available **methods**
  - **aim:** *understanding*, not applying "tools"
- not driven by available **theory**
  - **aim:** *test propositions* instead of arguing

Wanted: Data-Driven Modeling

- **agent-based model** that builds on theory and is calibrated against real data
- **generative** ("analytical") modeling, not just statistical models

Outline

Was können wir wissen?

Was sollen wir tun?

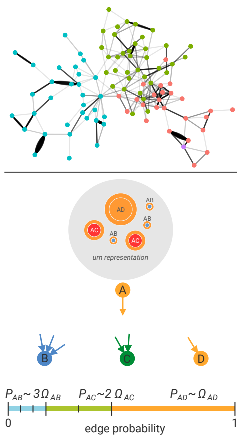
Was dürfen wir hoffen?

1. New ways to analyze social systems

- **High school contacts:** $n = 94$ students from 3 classes
  - $m = 1469$ repeated interactions, $\hat{m} = 213$ unique interactions
  - which network feature explains most of the social structure? (a) friendship (closed triangles), (b) class membership
- **ERGM** (unweighted network): $Pr(G|\Theta) = \frac{1}{Z} e^{-\sum_i \Theta_i f_i(G)}$
  - class membership: minor role, triadic closure matters most
- **Generalized hypergeometric ensemble**

$$\Pr(G|\Omega) = \left[\prod_{i,j} \binom{\Xi_{ij}}{A_{ij}}\right] \int_0^1 \prod_{i,j} \left(1 - z^{\frac{\Omega_{ij}}{S_\Omega}}\right)^{A_{ij}} dz$$

  - Class structure is more important than triadic closure
  - *repeated interactions matter!*



G. Casiraghi, V. Nanumyan, I. Scholtes, F.S., *SocInfo 2017*

2. New ways to test social theory

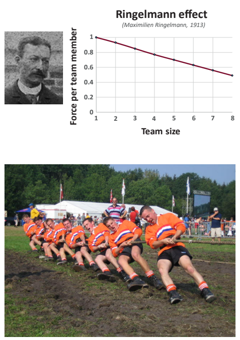


image: CC-by-SA Bart Derksen

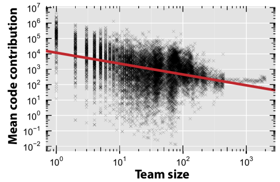
Ringelmann effect (social psychology)

- **larger teams** are **less productive**
  1. motivational factors: "social loafing"
  2. overhead of coordination

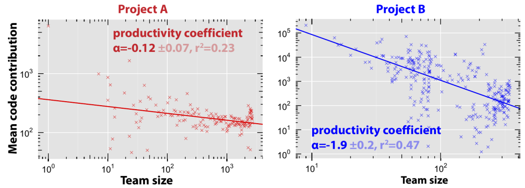
- **"The whole is greater than the sum of its parts"** (Aristotle)
  - ... but **absolute** contribution of individuals **drops** with system size
  - **economics:** *decreasing returns to scale*

I. Scholtes, P. Mavrodiev, F.S.: From Aristotle to Ringelmann: A large-scale analysis of team productivity and coordination in Open Source Software projects, Empirical Software Engineering 21 (2016) 642-683

Example: Productivity of software developer teams



- **robust log-linear regression:** $\alpha = 0.86 \pm 0.02$
  - *negative* productivity coefficient
- productivity Y **drops** with increasing team size X:
  - $X \times 2.0 \Rightarrow Y \times$ **1.1** (for **OSS** projecs!)
  - large variation across projects



- **data:** GHTorrent `gitHub` dump ($\sim$ 1.5 TB)
  - analysis of 58 most active OSS projects
  - $580'000$ commits, $30'000$ developers
  - more than 10 years of `commit history`
  - *Levenshtein distance* between `commits`

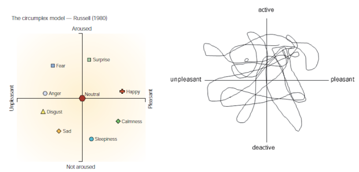
I. Scholtes, P. Mavrodiev, F.S.: From Aristotle to Ringelmann: A large-scale analysis of team productivity and coordination in Open Source Software projects, Empirical Software Engineering 21 (2016) 642-683

3. New ways to simulate social systems

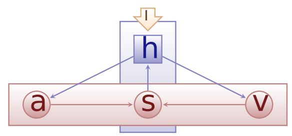
Agent-based model of emotional influence

- **Brownian Agents** with **two variables:** *arousal* $x(t)$, *valence* $y(t)$

$$\dot{x}_i = -\gamma_x\, x_i(t) + \mathcal{G}_x + A_{xi}\, \xi_y(t)$$
$$\dot{y}_i = -\gamma_y\, y_i(t) + \mathcal{G}_y + A_{yi}\, \xi_y(t)$$



- **driven variable:** *emotional expression* $s(t)$
  - depends on both arousal and valence $s_i[x_i(t), y_i(t)] = f[x_i(t)]\ \Theta[y_i(t) - \mathfrak{T}_i]$
  - emotional information stored in field h $\dot{h}_\pm = -\gamma_{h\pm} h_\pm(t) + s n_\pm(t) + I_\pm(t)$



F.S., D. Garcia: *An agent-based model of collective emotions in online communities*, European Physical Journal B, vol 77, no 4 (2010) 533-545

Toward a fully calibrated ABM: Parameter estimations

- **Methods**
  - *non-linear dynamics:* bifurcation analysis
  - *experiments:* physiological measurements
  - *text mining:* sentiment algorithms
- allow to **test hypotheses** about agent interactions
  - **Valence**

$$\frac{\Delta v(t)}{\Delta t} = -\gamma_v \left(v(t) - b\right) + h * \left(b_0 + b_1 v(t) + b_2 v(t)^2 + b_3 v(t)^3\right) + A_v \varepsilon$$

parameter	γ_v	b	b_0	b_2	b_3	R^2	N	$R^2(\xi_v)$
estimate	0.367***	0.056**	0.14***	0.057*	−0.047**	0.52	1271	0.85

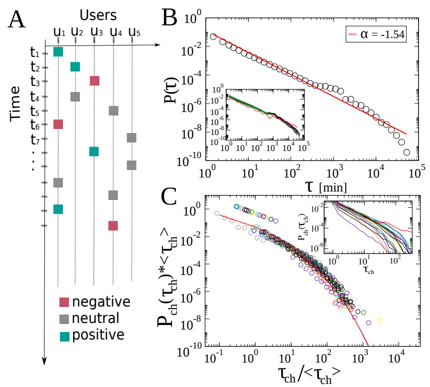
  - **Arousal**

$$\frac{\Delta a(t)}{\Delta t} = -\gamma_a \left(a(t) - d\right) + |h| * \left(d_0 + d_1 a(t) + d_2 a(t)^2 + d_3 a(t)^3\right) + A_a \varepsilon$$

parameter	γ_a	d	d_0	d_1	R^2	N	$R^2(\xi_a)$
estimate	0.414***	−0.442***	0.178***	0.14469**	0.28	1271	0.78

D. Garcia, A. Kappas, D. Kuster, F.S.: The Dynamics of Emotions in Online Interaction, Royal Society Open Science (2016) vol: 3, no: 160059

Example: Chatroom discussions

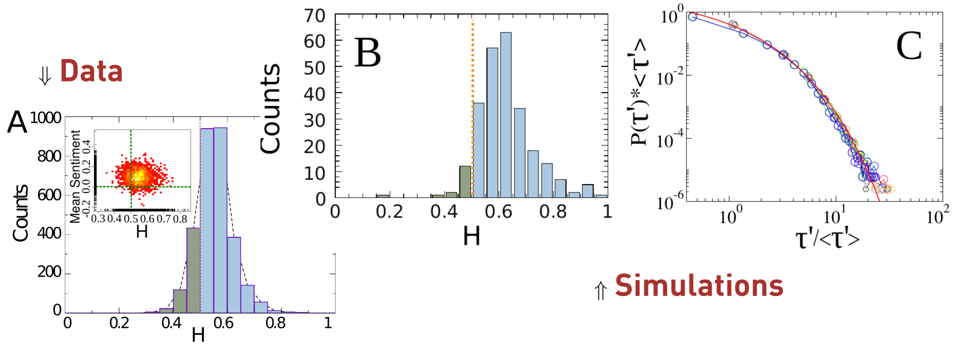


- **Chatrooms:** *no social network, costly signals*
  - bursty activities
    - B) *inter-activity time distribution*
    - C) *inter-event time distribution*

Data: 20 IRC Channels, 42 days, 2′688′760 posts, 25′166 users

A. Garas, D. Garcia, M. Skowron, F.S., *Scientific Reports* 2:402 (2012)

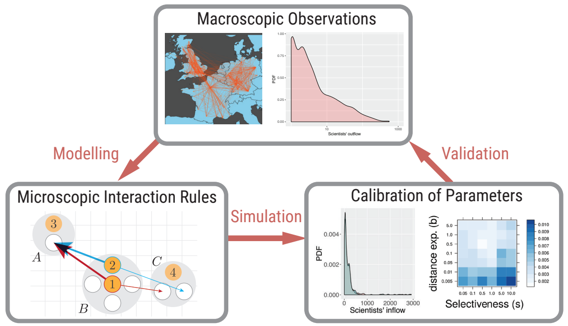
Agent-based simulations: Emotional persistence



- **Macroscopic dynamics correctly reproduced**

A. Garas, D. Garcia, M. Skowron, F.S., *Scientific Reports* 2:402 (2012)

ABM Alternative: Data-driven modeling

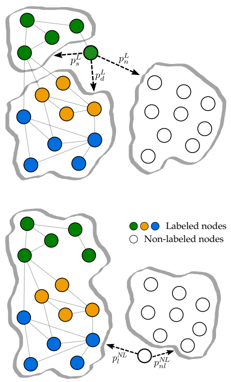


- combines *models*, *data* and *simulations*
- focus on mechanisms: *generative* ("analytical") models
- developed with *calibration* and *validation* in mind

Collaboration of Scientists: Co-Authorship Network

APS (1895-2004): $226'724$ scientists, $1'567'084$ papers

Agent based model of link formation

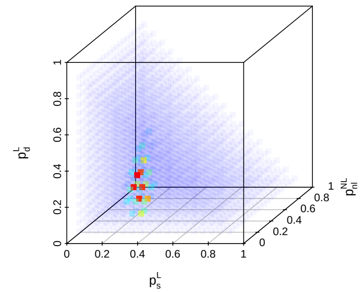


- **agent** i: two fixed properties
  - **activity** a_i: propensity to engage in a collaboration $\Rightarrow$ from data
  - **label** l_i: membership in a *circle of influence/group* ($\rightarrow$ *color*)
- **dynamics**:
  1. activation
  2. choose m *collaboration partners* (m taken from data)
     - **Incumbent:** (labeled node): $p_s^L + p_d^L + p_n^L = 1$
     - **Newcomer:** (non-labeled node): $p_{nl}^{NL} + p_l^{NL} = 1$
     - within labeled groups, partners are chosen wrt their degree
  3. form fully connected clique of size m, label propagation
- **Our task:** determine p_d^L, p_s^L, p_{nl}^{NL}

M. V. Tomasello, N. Perra, C. J. Tessone, M. Karsai, F.S.: *The Role of Endogenous and Exogenous Mechanisms in the Formation of R&D Networks*, Scientific Reports, vol. 4 (2014) 5679

Calibration 1/2

- First exploration: *network formation* parameter space
- M. L. approach $\Rightarrow$ parameter combination giving the *best match* with reality, w.r.t.:
  - average degree $\langle k \rangle$;
  - average path length $\langle l \rangle$;
  - global clustering coefficient (transitivity) C.
- *Optimal simulated network*: $p^* \equiv (p_s^{*L}, p_d^{*L}, p_n^{*L}, p_{nl}^{*NL}, p_l^{*NL})$
  - errors $\varepsilon_{\langle k \rangle}, \varepsilon_{\langle l \rangle}, \varepsilon_C$ have to be smaller than ε_0



M. V. Tomasello, G. Vaccario, F. Schweitzer: *Data-driven modeling of collaboration networks: A cross-domain analysis*, EPJ Data Science, volume: 6, pages:22 (2017)

Calibration 2/2

	ε^0	$\langle k \rangle^*$	$\langle l \rangle^*$	C^*	p_s^{*L}	p_d^{*L}	p_n^{*L}	p_l^{*NL}	p_{nl}^{*NL}
Co-authorship networks									
Quant.mech., field th., sp.relat. (PACS 03)	12%	5.83	5.58	0.392	0.85	0.05	0.10	0.45	0.55
General relativity and gravitation (PACS 04)	$>$ *30%**	*16.64*	*4.39*	*0.535*	*0.50*	*0.05*	*0.45*	*0.05*	*0.95*
Optics (PACS 42)	10%	7.60	5.79	0.451	0.60	0.05	0.35	0.35	0.65
Elect. transport in cond. matter (PACS 72)	8%	6.15	6.58	0.471	0.50	0.05	0.45	0.30	0.70
Superconductivity (PACS 74)	7%	7.51	5.51	0.465	0.55	0.05	0.40	0.35	0.65
Other applied and interd. physics (PACS 89)	8%	3.82	7.82	0.501	0.65	0.05	0.30	0.25	0.75

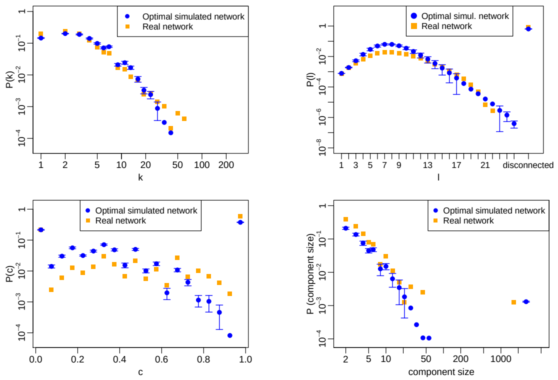
- ML estimation: $\varepsilon_0 < 12\%$ (R&D: 8%), no match for `General relativity and gravitation`
- links between *incumbents*: $p_s^{*L} + p_d^{*L} > 55\%$ (R&D: 70%)
- even stronger preference for *same circle of influence*: $p_s^{*L} \geq p_d^{*L}$ (note: $p_d^{*L} = 5\%$)
- *Newcomers* tend to link with **newcomers**: $p_l^{*NL} < p_{nl}^{*NL}$ (R&D: $p_l^{*NL} > p_{nl}^{*NL}$)

Network endogenous factors (patters already present in network, encoded in labeled agents)

- explain most of the newly formed links: $p_s^{*L} + p_d^{*L} + p_l^{*NL} > p_{nl}^{*NL} + p_{nl}^{*L}$
- rules for the formation of communities captured by label dynamics

M. V. Tomasello, G. Vaccario, F. Schweitzer: *Data-driven modeling of collaboration networks: A cross-domain analysis*, EPJ Data Science, volume: 6, pages:22 (2017)

Validation: Reproduce network distributions (PACS 89)

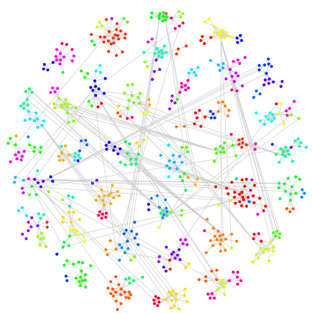


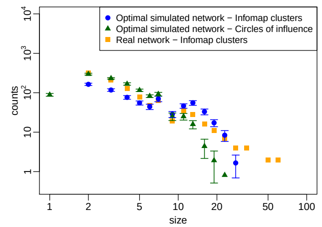
Reproduced distributions:

- degree
- path length
- local clustering coefficient
- component size

M. V. Tomasello, G. Vaccario, F. Schweitzer: *Data-driven modeling of collaboration networks: A cross-domain analysis*, EPJ Data Science, volume: 6, pages:22 (2017)

Validation: Reproduce clusters (PACS 89)

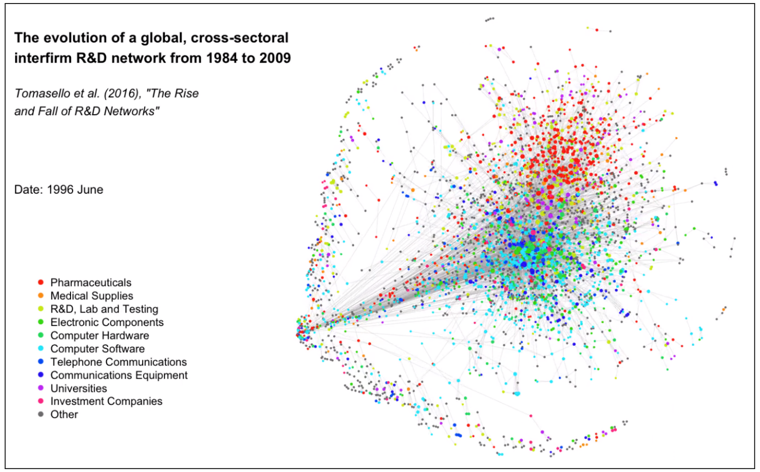




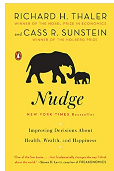
- Optimal simulated network (only the largest 30 clusters identified by Infomap). Each color=different label.
- 966 **empirically detected clusters**. I_{norm}(labels,Infomap clusters)=0.95 **remarkable overlap**.

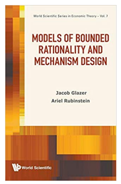
M. V. Tomasello, G. Vaccario, F. Schweitzer: *Data-driven modeling of collaboration networks: A cross-domain analysis*, EPJ Data Science, volume: 6, pages:22 (2017)

4. New ways to compare systems: Scientists vs firms

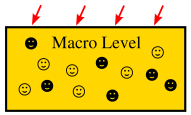


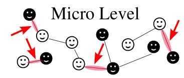
5. New ways to influence social systems





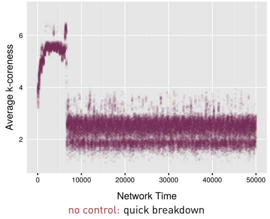
1. **Ethical problem:** *"do we manipulate people's decisions?"*
   - **Economics:** Mechanism design
     - control **rules** ⇒ *game theoretical* setting
   - **Psychology:** Nudging
     - control **information** ⇒ *behavioral* setting
2. **Technical problem:** *identify target*
   - **Marketing:** Influencers
     - whom to influence?, how to influence?
   - **Control theory:** Is the system controllable?
     - driven to a pre-defined state, optimal trajectory?
     - Which is the right element? Which is the right incentive? Timing?

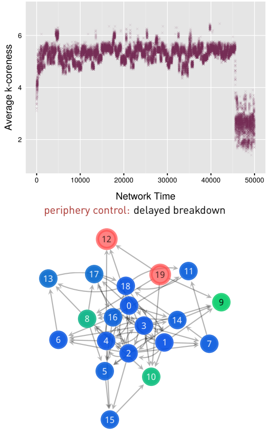


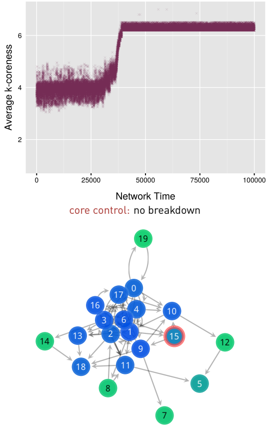


F.S., Designing systems bottom up: Facets and problems, **Advances in Complex Systems** (2020)
F.S., The Bigger Picture: Complexity Meets Systems Design, In: *Design. Tales of Science and Innovation* (Eds. G. Folkers, M. Schmid) (2019)

Network interventions: Improved robustness





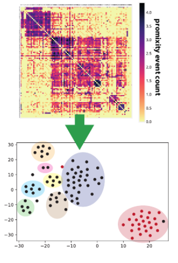


G. Casiraghi, F.S.: *Improving the robustness of online social networks: A simulation approach of network interventions.* **Frontiers in Robotics and AI** (2020)

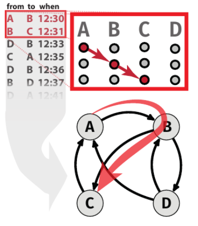
Conclusions

- **End of theory?** – Quite the opposite
  - profound disciplinary knowledge (social science)
- **Social physics?** – less physics, more data science
  - new methods, new topics, new questions, new skills

Learning in Complex Relational Data



Network Analytics for Time Series Data



Data Analytics in Social Systems