Lecture 11.

Equilibria and Butterflies

CS 222: AI Agents and Simulations

Stanford University

Joon Sung Park

Announcement

- Submit your team name and project description (we will do this at the end of the class):
  - http://bit.ly/3Ah4dKP

- We've released AgentBank-CS222. Please check the Announcements section on Canvas.

So far this quarter, we have covered...

- The architecture and implementation of generative agents
- How to evaluate and leverage generative agents to model individuals (models of individuals) and populations (effect sizes)
- How to build the foundations of GABM—models in which many generative agents "interact" with each other

But how do you evaluate and leverage GABM?

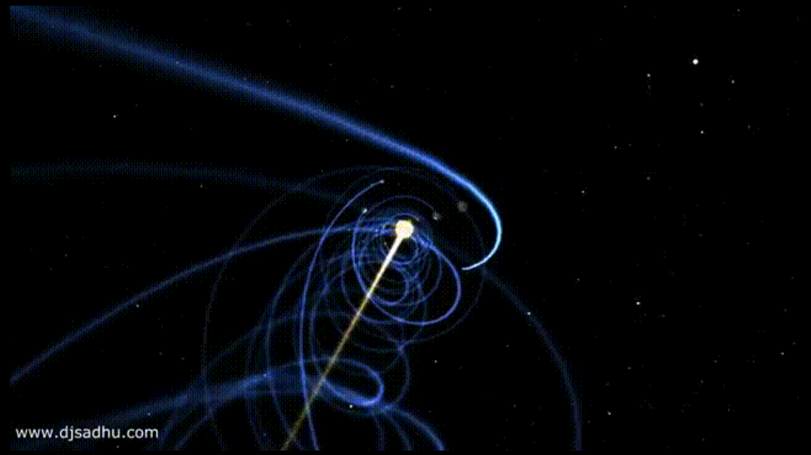
I have taught you everything I (and the field) *know*.

Today (and Wednesday), I will share my conjectures on where I *think* we are headed.

Complex systems and the butterfly effect

Complex system: a system composed of many interconnected components that interact in dynamic and often nonlinear ways, producing collective behaviors that are difficult to predict from the behavior of individual parts.

We are surrounded by complex systems in nature



Planetary orbits

Cloud formations

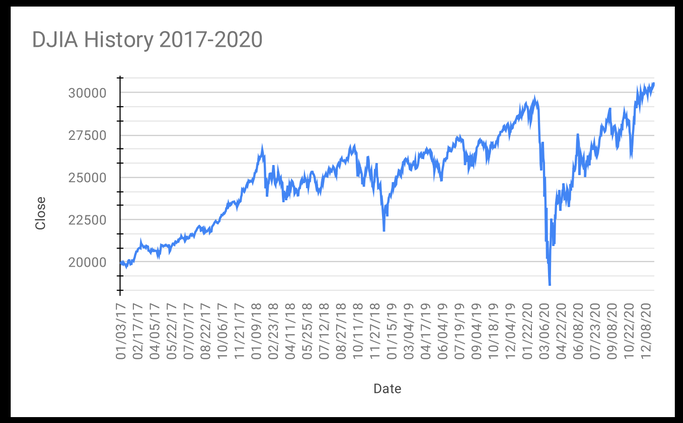
Ocean waves

Flock of birds

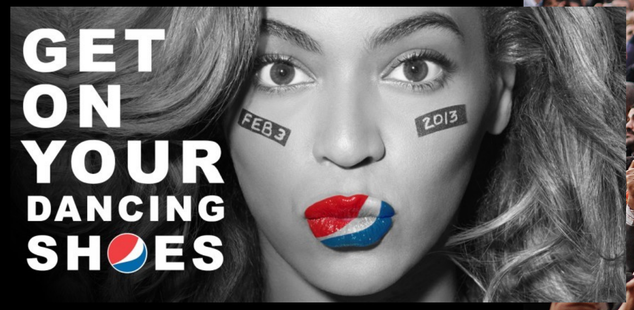
Heart rhythms

And in our social lives

Phantom traffic jams

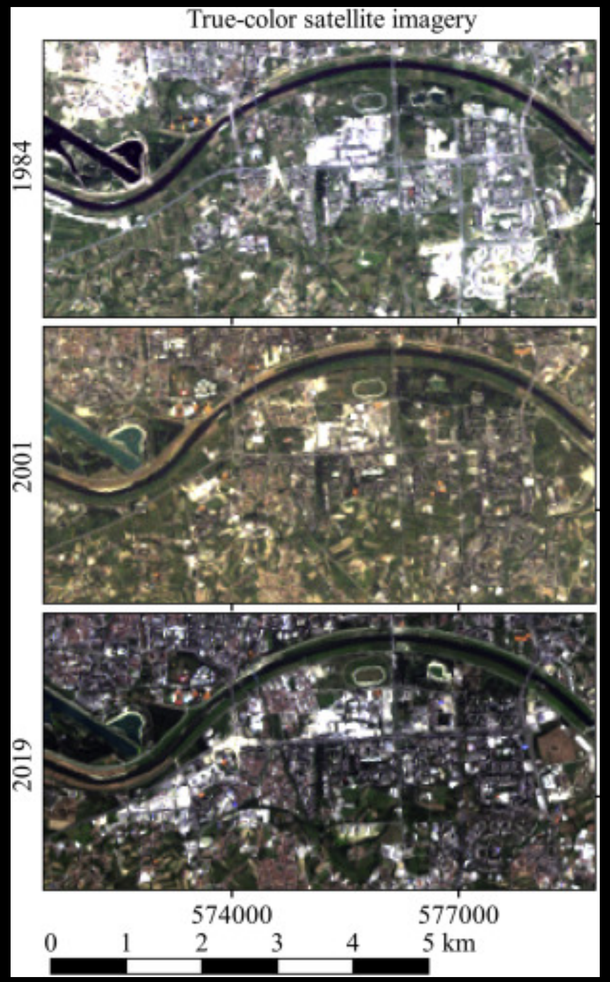


Market crash

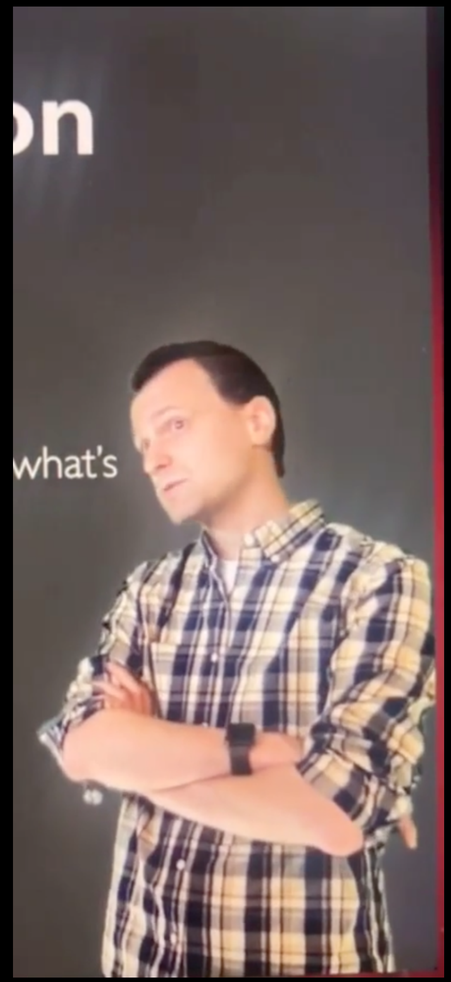


Consumer behavior

Social movement

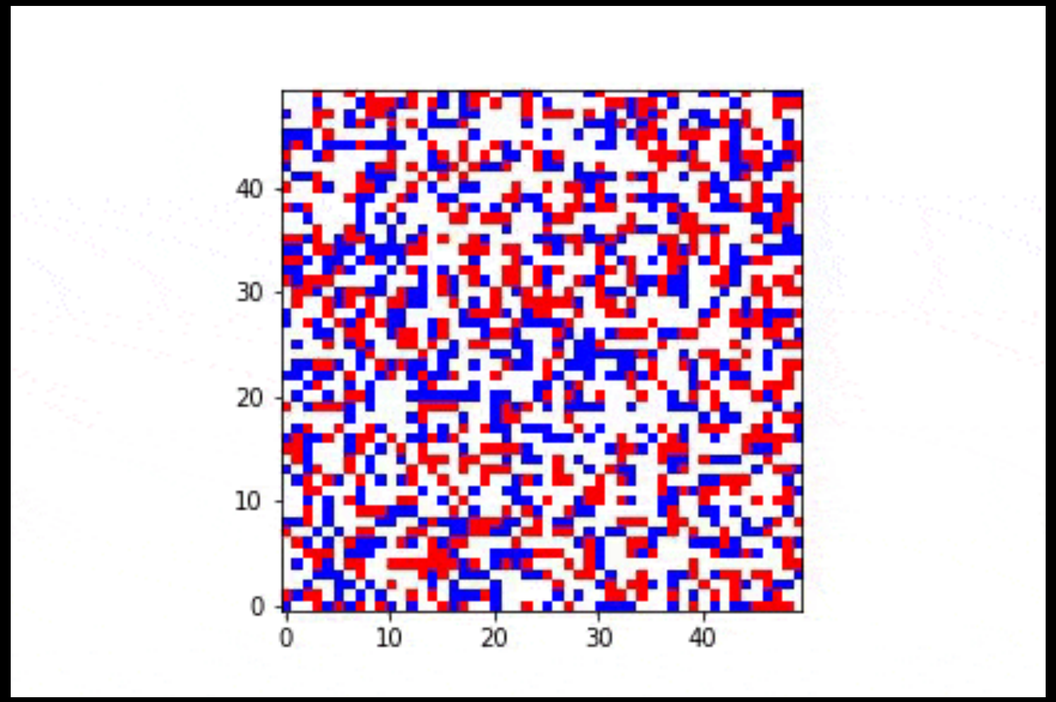


Urban growth



Viral content

Agent-based models (and generative agent-based models) are complex.



Model of segregation

Smallville

T. C. Schelling, Micromotives and Macrobehavior (W.W. Norton & Company, 1978).
T. C. Schelling, Dynamic models of segregation. J. Math. Sociol. 1, 143–186 (1971).
J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).

Chaos is prevalent in complex systems: tiny variations in the initial conditions of a system can lead to vastly different outcomes, to the point where the outcome seems random

Q: Imagine Sam won the election in our simulation. What might cause the outcome to differ in real life?

https://pollev.com/helenav330

Is our world inherently unpredictable?

And if so, what can we learn from simulations?

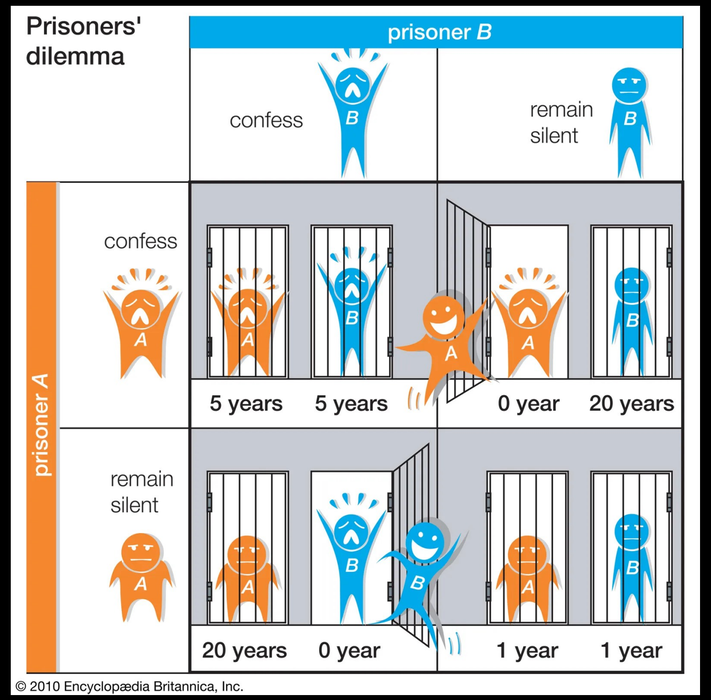
Equilibrium in complex systems

Equilibria are states in which a system remains balanced, with no net change in the absence of external disturbances, as opposing forces or influences are in a stable relationship.

Nash equilibria refer to situations in which each player in a game has chosen a strategy, and no one can benefit by changing their strategy while the others keep theirs unchanged.

C. A. Holt, A. E. Roth, The Nash equilibrium: A perspective. Proc. Natl. Acad. Sci. U.S.A. (2004).
Nash, J. F. Non-cooperative games. Ann. Math. 54, 286-295 (1951)

Prisoner's dillema



Two prisoners are given a choice to confess or stay silent. If both stay silent, they get minimal sentences. If one confesses and the other stays silent, the confessor goes free while the silent one gets a heavy sentence. If both confess, they both get moderate sentences.

Nash Equilibrium: Both confess. Neither prisoner can improve their situation by changing their choice alone, even though mutual silence would yield a better outcome for both.

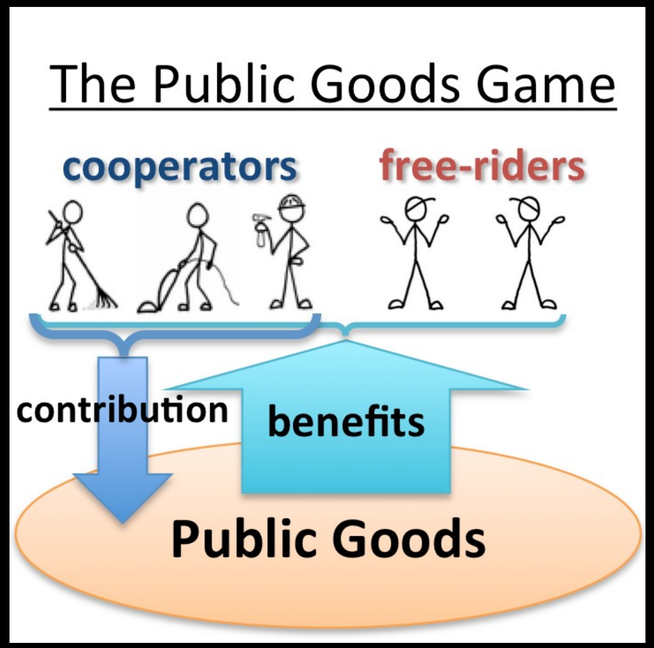
Tucker, A. W. A two-person dilemma. (Stanford University, 1950)

Rock-Paper-Scissors

In this game, players choose rock, paper, or scissors, with each option winning, losing, or tying against another in a cyclical pattern.

Nash Equilibrium: Each player randomly chooses rock, paper, or scissors with equal probability (1/3 for each choice). No player can improve their outcome by changing their strategy alone, as any predictable pattern would be exploited by the other player.

Public Goods Game



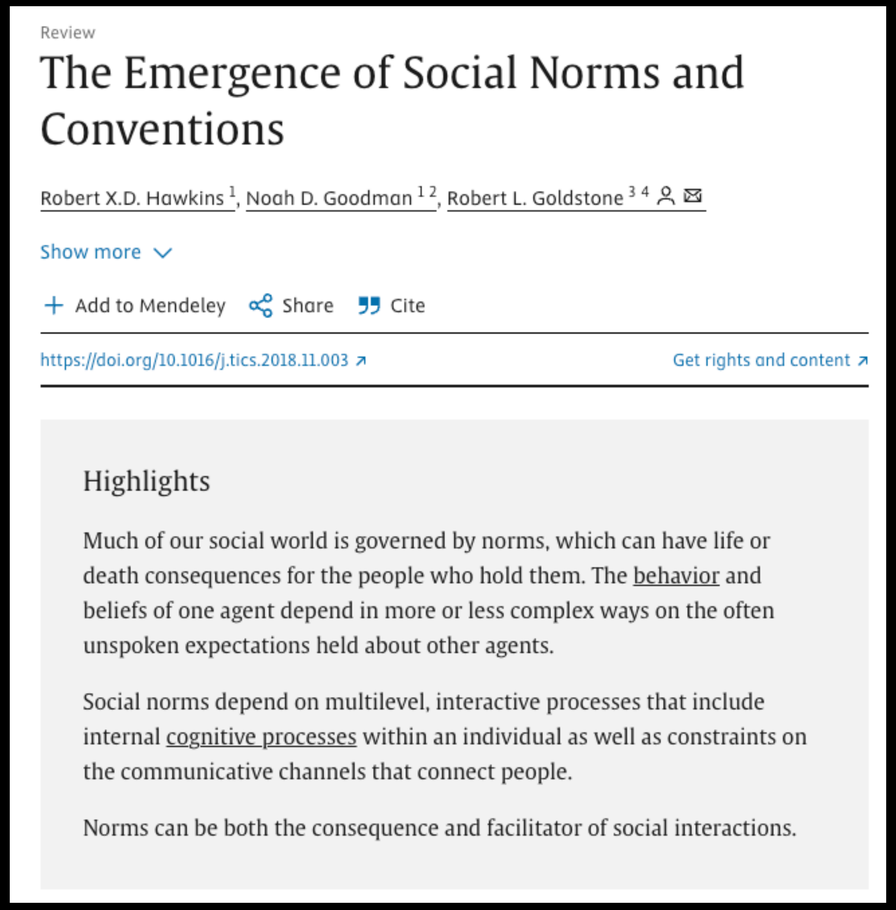
Individuals in a group decide how much of their resources to contribute to a public good. Everyone benefits from the public good, but individual contributions reduce personal resources.

Nash Equilibrium: Each individual contributes nothing if they believe others won't contribute enough to make a difference. In this equilibrium, the public good isn't provided, illustrating a "free-rider problem."

Marwell, G. & Ames, R. E. Economists free ride, does anyone else?: Experiments on the provision of public goods, IV. J. Public Econ. 15, 295-310 (1981)

In large, complex social systems, we often observe the rise of equilibria or emergent phenomena.

Case study 1: social norms



Review

The Emergence of Social Norms and Conventions

Robert X.D. Hawkins [1], Noah D. Goodman [1 2], Robert L. Goldstone [3 4]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.tics.2018.11.003 Get rights and content

Highlights

Much of our social world is governed by norms, which can have life or death consequences for the people who hold them. The behavior and beliefs of one agent depend in more or less complex ways on the often unspoken expectations held about other agents.

Social norms depend on multilevel, interactive processes that include internal cognitive processes within an individual as well as constraints on the communicative channels that connect people.

Norms can be both the consequence and facilitator of social interactions.

Social norms are a type of equilibrium where behaviors stabilize around commonly accepted rules, such as queuing in lines or greeting people in certain ways.

Hawkins, R. X. D., Goodman, N. D., Goldstone, R. L. The emergence of social norms and conventions. Trends Cogn. Sci. 23, 158-169 (2019)

Case study 2: political polarization

ANNUAL REVIEWS Publications A-Z Journal Information

Home / A-Z Publications / Annual Review of Political Science / Volume 11, 2008 / Article

ANNUAL REVIEW OF POLITICAL SCIENCE Volume 11, 2008

Review Article

Political Polarization in the American Public

Morris P. Fiorina[1], and Samuel J. Abrams[2]

View Affiliations

Vol. 11:563-588 (Volume publication date June 2008) | https://doi.org/10.1146/annurev.polisci.11.053106.153836

© Annual Reviews

Info Sections Get Access Tools Share

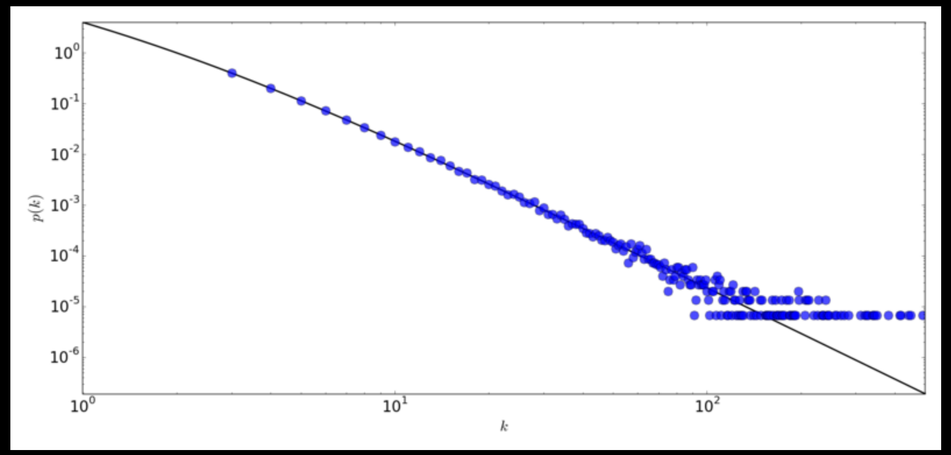
For more than two decades political scientists have discussed rising elite polarization in the United States, but the study of mass polarization did not receive comparable attention until fairly recently. This article surveys the literature on mass polarization. It begins with a discussion of the concept of polarization, then moves to a critical consideration of different kinds of evidence that have been used to study polarization, concluding that much of the evidence presents problems of inference that render conclusions problematic. The most direct evidence—citizens' positions on public policy issues—shows little or no indication of increased mass polarization over the past two to three decades. Party sorting—an increased correlation between policy views and partisan identification—clearly has occurred, although the extent has sometimes been exaggerated. Geographic polarization—the hypothesized tendency of like-minded people to cluster together—remains an open question. To date, there is no conclusive evidence that elite polarization has stimulated voters to polarize, on the one hand, or withdraw from politics, on the other.

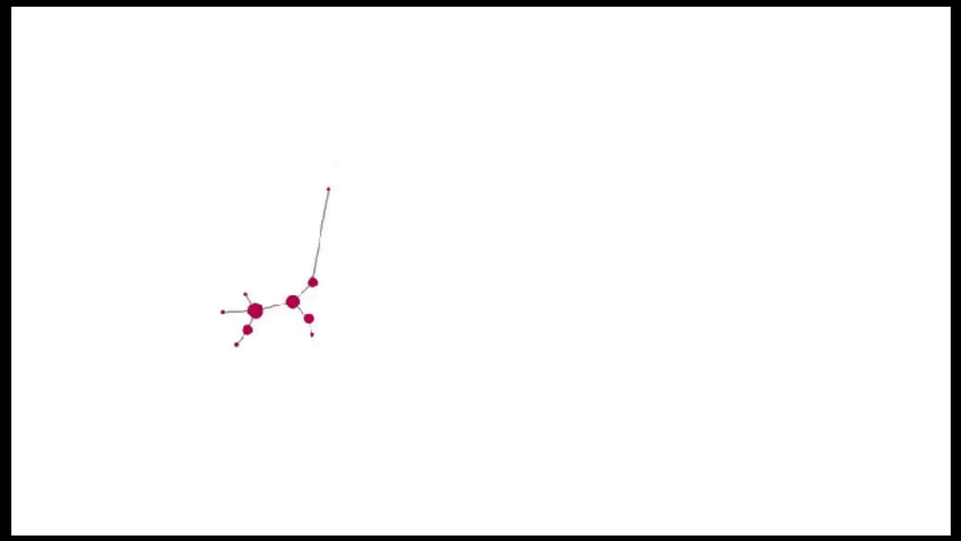
Keyword(s): culture war, elite polarization, mass polarization, party sorting

In some societies, political opinions settle into a polarized equilibrium, where two main viewpoints dominate. Even if there are shifts in public opinion or specific issues, the overall structure remains stable, with opposing views balancing each other and preventing any single perspective from dominating.

Fiorina, M. P., Abrams, S. J. Political polarization in the American public. Annu. Rev. Polit. Sci. 11, 563-588 (2008)

Case study 3: emergence of scale free network





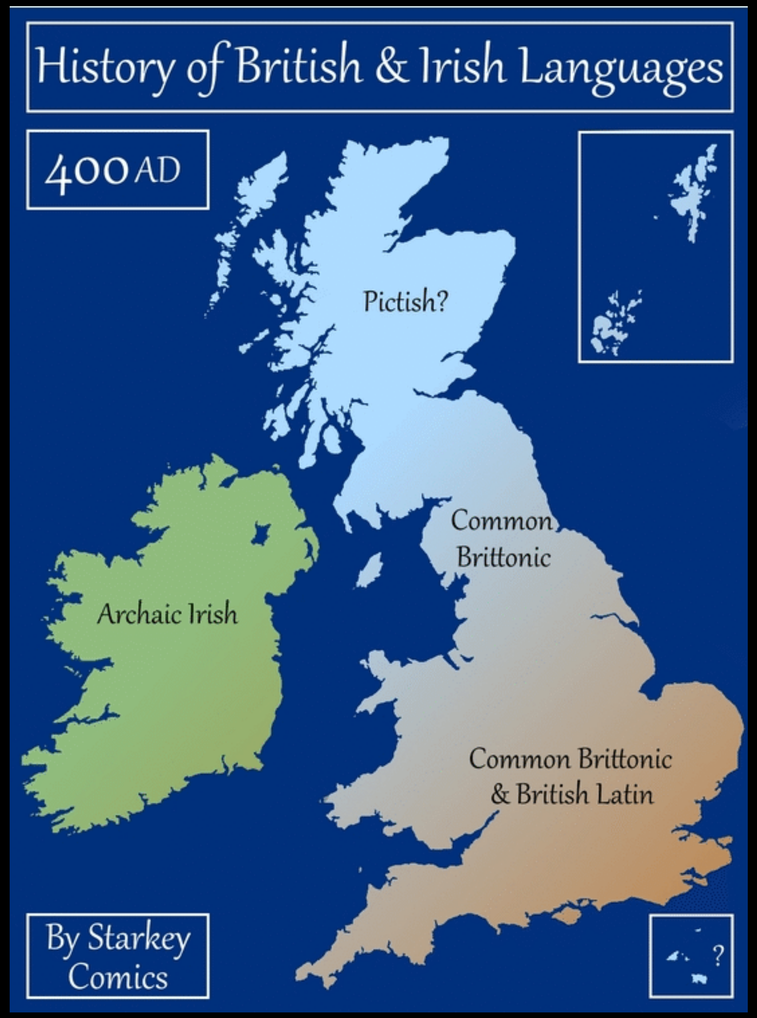
In many social networks, some nodes (or vertices) have a significantly higher number of connections (or edges) than others, following a power-law distribution. This means that a small number of nodes, called "hubs," have many connections, while most nodes have relatively few.

Barabási, A.-L., Albert, R. Emergence of scaling in random networks. Science 286, 509-512 (1999)

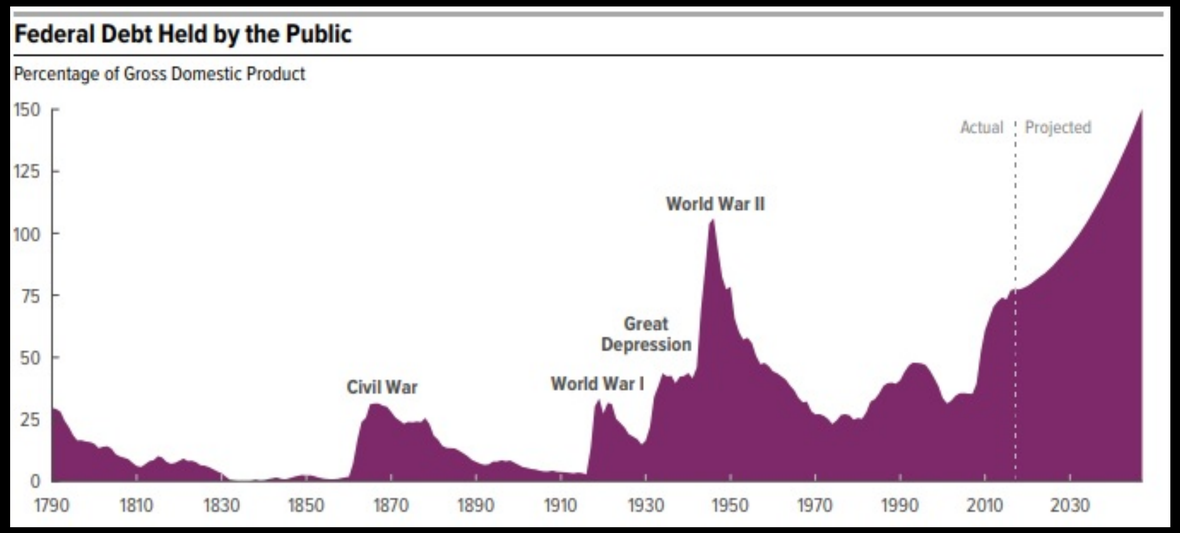
But they vary in their degrees of stability.

In complex systems, there is no such thing as “happily ever after.”

Nations rise, fall, and rise again.



Equilibria in complex systems are often *unstable*



Economic Bubbles: Financial markets can enter unstable equilibria during asset bubbles, where prices rise rapidly due to speculative investment, becoming disconnected from the asset's intrinsic value. A small trigger, such as a disappointing earnings report or a shift in interest rates, can burst the bubble, leading to sharp market declines and economic crises.

Kindleberger, C. P. Manias, Panics, and Crashes: A History of Financial Crises (Basic Books, New York, 1978)
Shiller, R. J. Irrational Exuberance (Princeton Univ. Press, ed. 3, 2015)

Extracting insights from GABM

Anecdotal insights

Social Simulacra: Creating Populated Prototypes for Social Computing Systems

Joon Sung Park
Stanford University
Stanford, USA
joonspk@stanford.edu

Lindsay Popowski
Stanford University
Stanford, USA
popowski@stanford.edu

Carrie J. Cai
Google Research
Mountain View, CA, USA
cjcai@google.com

Meredith Ringel Morris
Google Research
Seattle, WA, USA
merrie@google.com

Percy Liang
Stanford University
Stanford, USA
pliang@cs.stanford.edu

Michael S. Bernstein
Stanford University
Stanford, USA
msb@cs.stanford.edu

ABSTRACT

Social computing prototypes probe the social behaviors that may arise in an envisioned system design. This prototyping practice is currently limited to recruiting small groups of people. Unfortunately, many challenges do not arise until a system is populated at a larger scale. Can a designer understand how a social system might behave when populated, and make adjustments to the design before the system falls prey to such challenges? We introduce *social simulacra*, a prototyping technique that generates a breadth of realistic social interactions that may emerge when a social computing system is populated. Social simulacra take as input the designer's description of a community's design—goal, rules, and member personas—and produce as output an instance of that design with simulated behavior, including posts, replies, and anti-social behaviors. We demonstrate that social simulacra shift the behaviors that they generate appropriately in response to design changes, and that they enable exploration of "what if?" scenarios where community members or moderators intervene. To power social simulacra, we contribute techniques for prompting a large language model to generate thousands of distinct community members and their social interactions with each other; these techniques are enabled by the observation that large language models' training data already includes a wide variety of positive and negative behavior on social media platforms. In evaluations, we show that participants are often unable to distinguish social simulacra from actual community behavior and that social computing designers successfully refine their social computing designs when using social simulacra.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing systems and tools**.

KEYWORDS

social computing, prototyping

ACM Reference Format:

Joon Sung Park, Lindsay Popowski, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2022. Social Simulacra: Creating Populated Prototypes for Social Computing Systems. In *The 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22), October 29-November 2, 2022, Bend, OR, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3526113.3545616

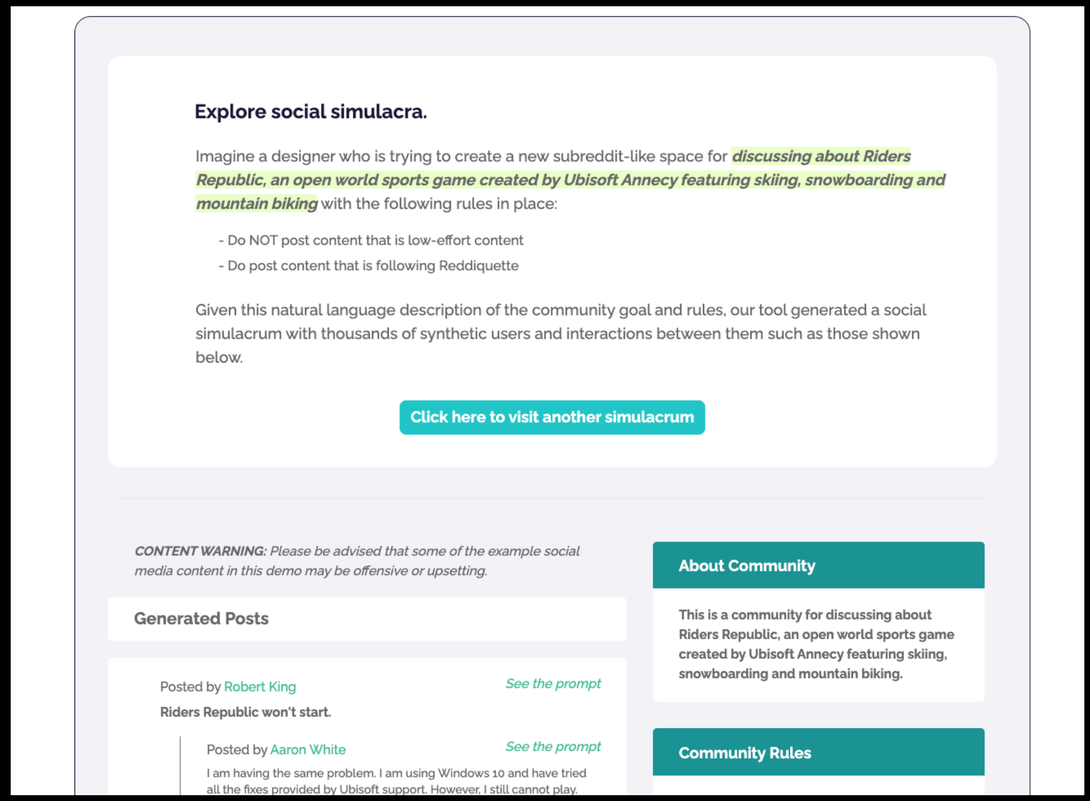
UIST '22, October 29-November 2, 2022, Bend, OR, USA

ACM ISBN 978-1-4503-9320-1/22/10…$15.00
https://doi.org/10.1145/3526113.3545616

1 INTRODUCTION

How do we anticipate the interactions that will arise when a social computing system is populated [4, 23]? In social computing, design decisions such as a community's goal and rules can give rise to dramatic shifts in community norms, newcomer enculturation, and anti-social behavior [45]. Success requires that the designer make informed decisions to shape these socio-technical outcomes. Yet, despite decades of progress in research and practice, understanding the effects of these design decisions remains challenging; as a result, designers are regularly surprised by the behaviors that arise when their spaces are fully populated.

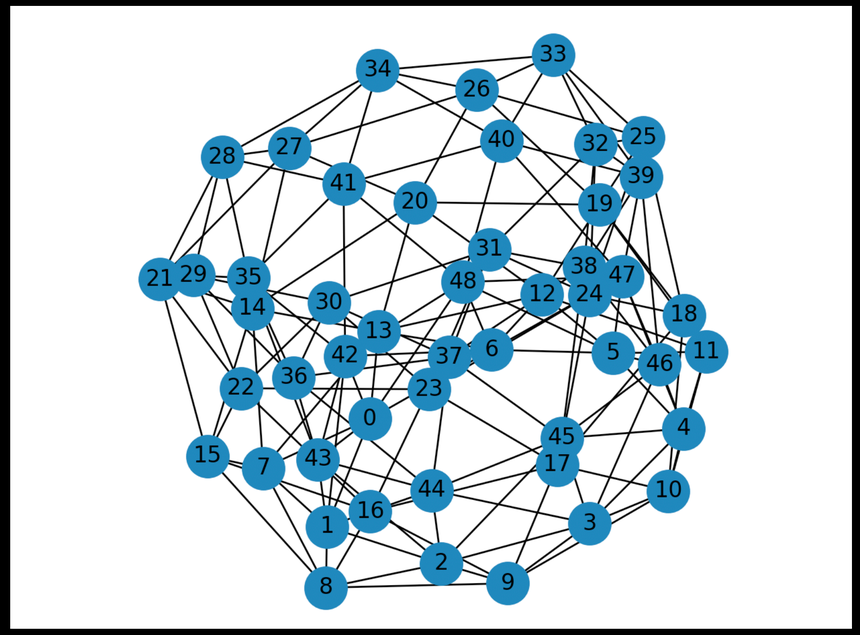
To design pro-social spaces, designers need *prototyping* techniques that enable them to reflect on social behaviors that may result from their design choices, then iterate [69]. Prototypes in social computing typically take the form of experience prototypes where the designer recruits a small group of people to use the system [7, 22]. However, there remains a large gap between the behaviors that arise in a small set of test users and the behaviors that arise in a socio-technical system when it is fully populated: for example, anti-social behaviors may not arise within a tight-knit group [45]; small homogeneous groups overlook the breadth of users or content that may arise in the system [24, 42, 74]; rules and moderation strategies may not need to be spelled out explicitly or enforced [41]. Barring actually launching our systems at scale, designers currently have no way of starting to explore these questions to reflect on the social dynamics of their designs. This need becomes only more urgent as social computing reckons with the harms it can engender [23] at the same time as designers fashion new computationally-mediated social spaces in forms both familiar (e.g., a new subreddit or Discord server) and novel (e.g., a new workspace platform).



Social simulacra

Social Simulacra: Creating Populated Prototypes for Social Computing Systems. UIST 2022.

Equilibria and emergence



Network simulation

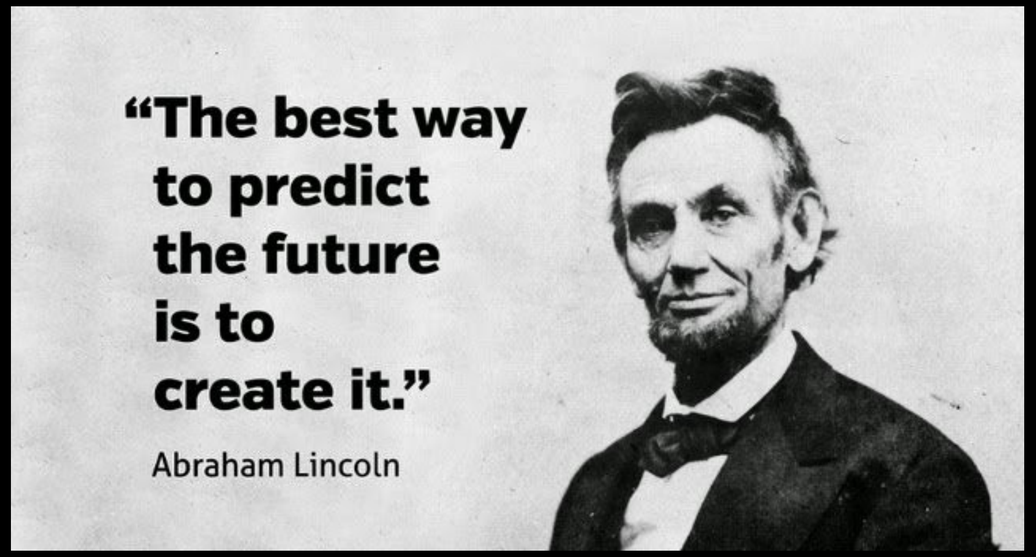
do	an	do		
an	do	do	an	an
an	do	an		do
do	do	do		an
an			an	

do	do	an	an	do
do	do		an	an
do	do	an		an
do	do	an		an
		an		

Model of segregation

Interventions

Can simulations provide a step-by-step guideline for interventions to shape the future?



Closing thought:

Will GPT-X solve wicked problems?

Hypothesis: While LLMs will serve as cognitive CPUs, simulations will function as cognitive GPUs in the era of generative AI.

The grand challenges of our generation do not require a complex central reasoning unit.

Rather, they require relatively simple cognitive units that come together to form complex phenomena.

Brief work session for the final project

- Submit your team name and project description (we will do this at the end of the class):
  - http://bit.ly/3Ah4dKP

References

- T. C. Schelling, Micromotives and Macrobehavior (W.W. Norton & Company, 1978).
- T. C. Schelling, Dynamic models of segregation. J. Math. Sociol. 1, 143–186 (1971).
- J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).
- C. A. Holt, A. E. Roth, The Nash equilibrium: A perspective. Proc. Natl. Acad. Sci. U.S.A. (2004). Nash,
- J. F. Non-cooperative games. Ann. Math. 54, 286-295 (1951)
- Tucker, A. W. A two-person dilemma. (Stanford University, 1950)
- Marwell, G. & Ames, R. E. Economists free ride, does anyone else?: Experiments on the provision of public goods, IV. J. Public Econ. 15, 295-310 (1981)
- Hawkins, R. X. D., Goodman, N. D., Goldstone, R. L. The emergence of social norms and conventions. Trends Cogn. Sci. 23, 158-169 (2019)

References

- Fiorina, M. P., Abrams, S. J. Political polarization in the American public. Annu. Rev. Polit. Sci. 11, 563-588 (2008)
- Barabási, A.-L., Albert, R. Emergence of scaling in random networks. Science 286, 509-512 (1999)
- Kindleberger, C. P. Manias, Panics, and Crashes: A History of Financial Crises (Basic Books, New York, 1978)
- Shiller, R. J. Irrational Exuberance (Princeton Univ. Press, ed. 3, 2015)
- Social Simulacra: Creating Populated Prototypes for Social Computing Systems. UIST 2022.

CS 222: AI Agents and Simulations

Stanford University

Joon Sung Park