Discovering Strategic Behaviors for Collaborative Content-Production in Social Networks

Yuxin Xiao, Adit Krishnan, Hari Sundaram
University of Illinois at Urbana-Champaign

Yash Saboo
Tai-Ying Chen

ABSTRACT

- **Setting**: There are always individuals who obtain greater rewards and reputation than their peers on social networking sites, no matter if the reward is transparent or opaque.
- **Challenge**: Identify the rationality behind their actions due to factors like the combinatorial strategy space, inability to determine payoffs, and resource limitations faced by individuals.
- **Problem the Paper addresses**: Can resource-limited individuals discover strategic behaviors associated with high payoffs when producing collaborative/interactive content in social networks?

ABSTRACT

- **Proposed Solution**: Dynamic Dual Attention Networks (DDAN)
  - models individuals' content production strategies through a generative process, under the influence of social interactions involved in the process
- **Findings**:
  - Different strategies give rise to different social payoffs;
  - The best performing individuals exhibit stability in their preference over the discovered strategies, which indicates the emergence of strategic behavior; and
  - The stability of a user's preference is correlated with high payoffs.

Related Work

- **Aim**: Find if individuals can successfully discover strategies with high payoffs in social networks
- Herbert A Simon (1972). “Theories of bounded rationality”
  - Introduced the idea of bounded rationality—that human beings use limited resources to make decisions.
  - Previous lectures: D. Kahneman, "A perspective on judgment and choice: Mapping bounded rationality." and Papadimitriou, C. H. and Yannakakis, M. (1994), "On complexity as bounded rationality"

Related Work

- Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. (2013), “Steering user behavior with badges”
    - Online social networks typically have an explicit mechanism that allocates rewards (usually points) that vary with users’ behaviors;
    - for example,
        - StackOverflow (explicit)
        - Twitter(implicit)
- Thodoris Lykouris, Vasilis Syrgkanis, and Éva Tardos (2016). “Learning and efficiency in games with dynamic population”
    - when agents play repeated games with strategies that guarantee low-adaptive regret, high social welfare is ensured. But the question arises that does this still hold in practice?

Problem Formulation

- Authors ask:

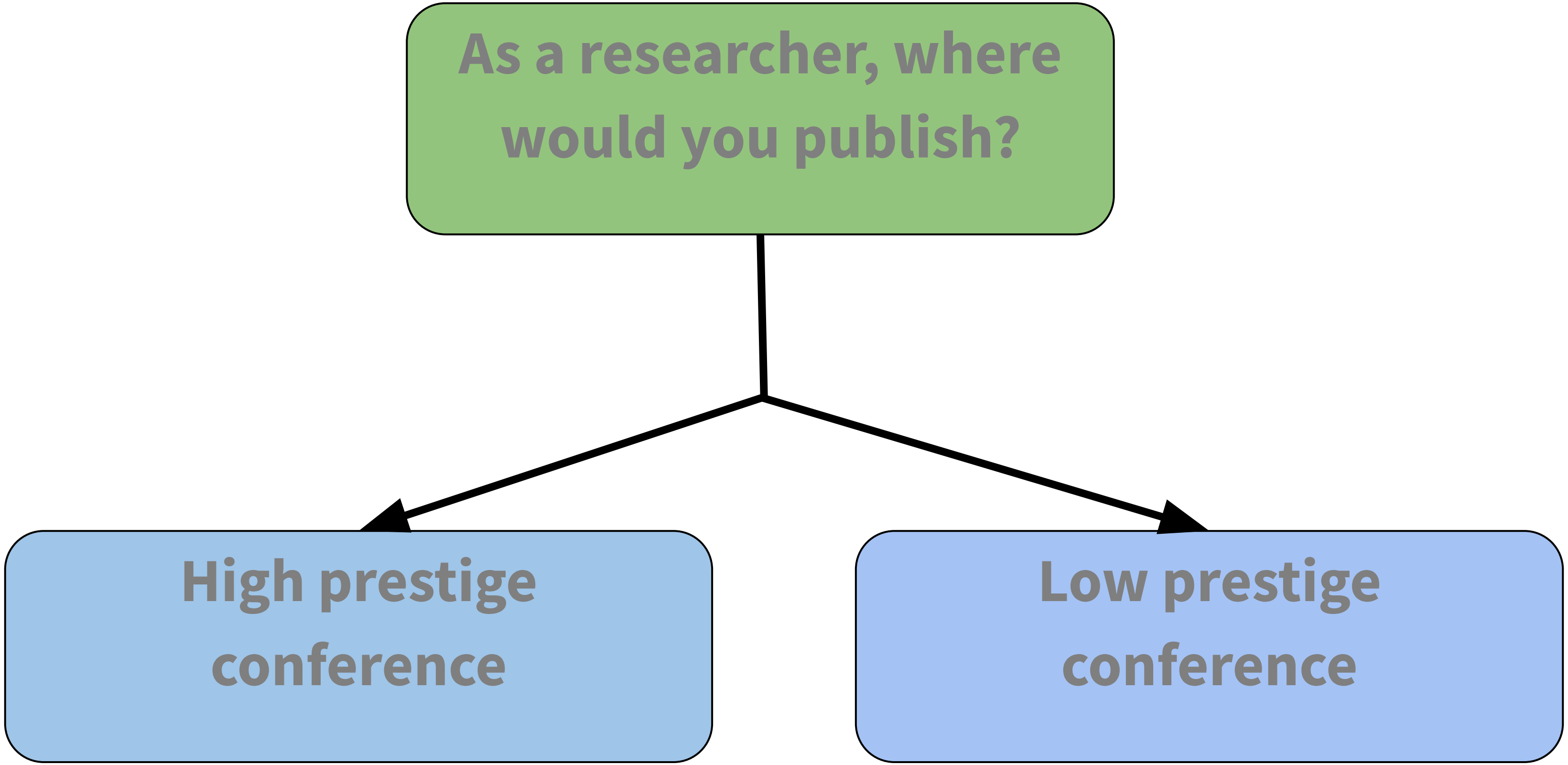
  Are resource-limited individuals in social networks able to discover content-production strategies that yield high payoffs?

- **Resource-limited**: limited time, attention
- **Content**: blog post, academic paper, question-answer forum (e.g., StackExchange)
- **Payoffs**: citations (academic paper), in-links (blog posts), up votes (QA forum)

Novelty

- Previous work was on:
  - theoretical concepts only
  - no attempt to identify strategic behaviors from data
- 1st attempt to identify strategic behaviors from empirical data
- Strong experimental findings:
  - Different strategies result in different payoffs.
  - Stability of preference is correlated with high payoffs.

Let's Brainstorm, shall we?



VLDB 2019

The Sixth IEEE International Conference on Multimedia Big Data

D·A·V·I·D·S
V·L·O·G

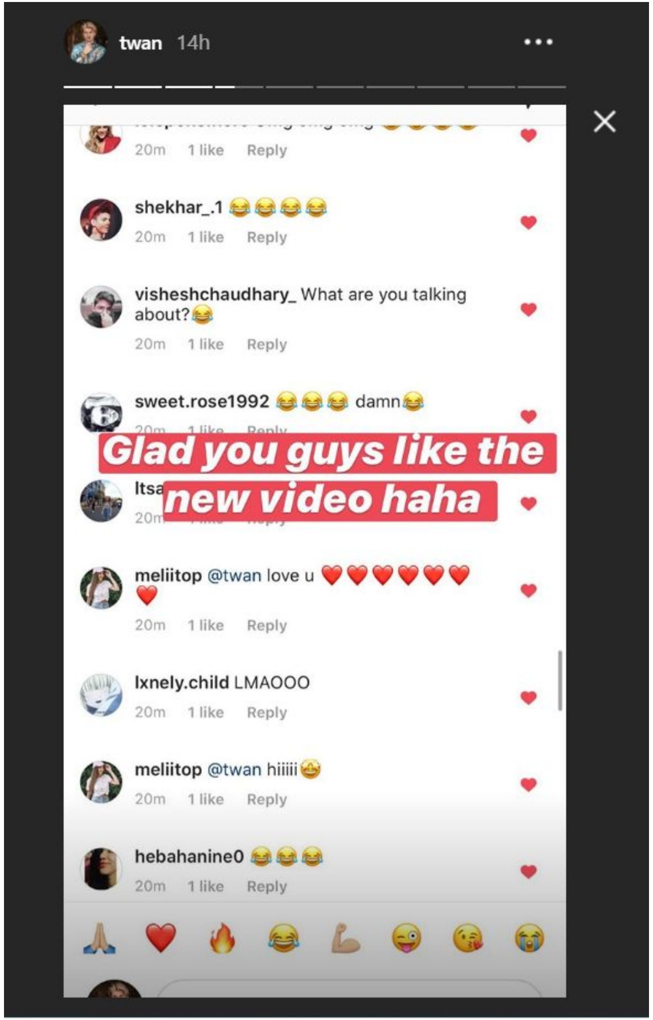
Where should David Dobrik publish his next video?

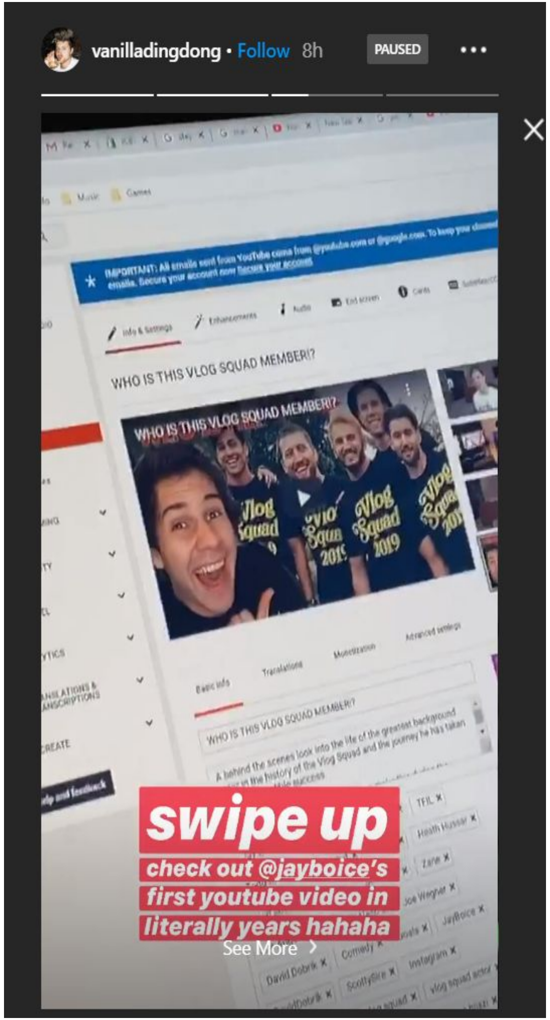
IGTV

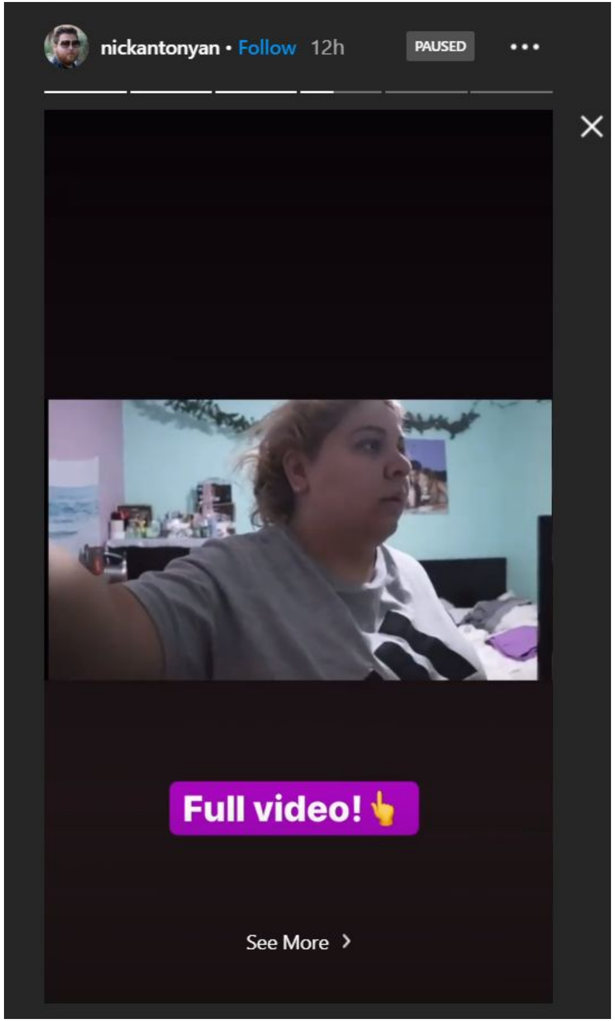
(12M)

YouTube

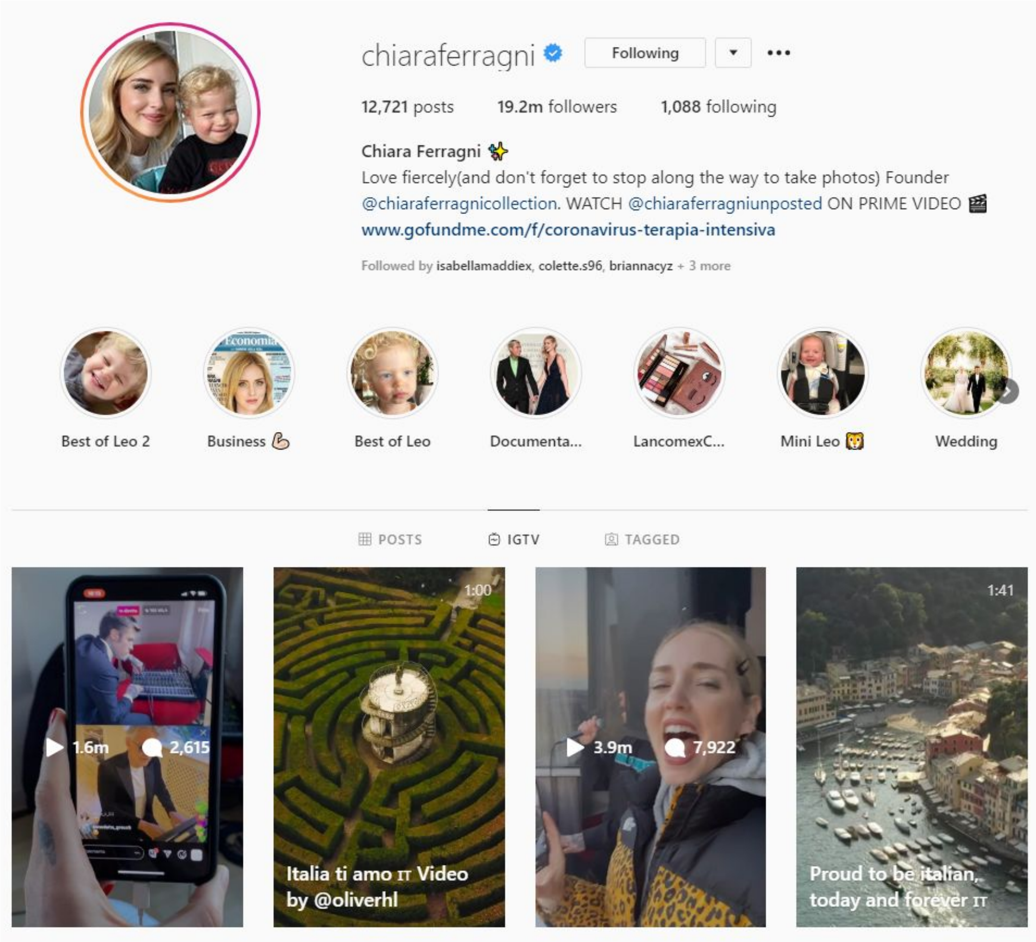
(16.6M)

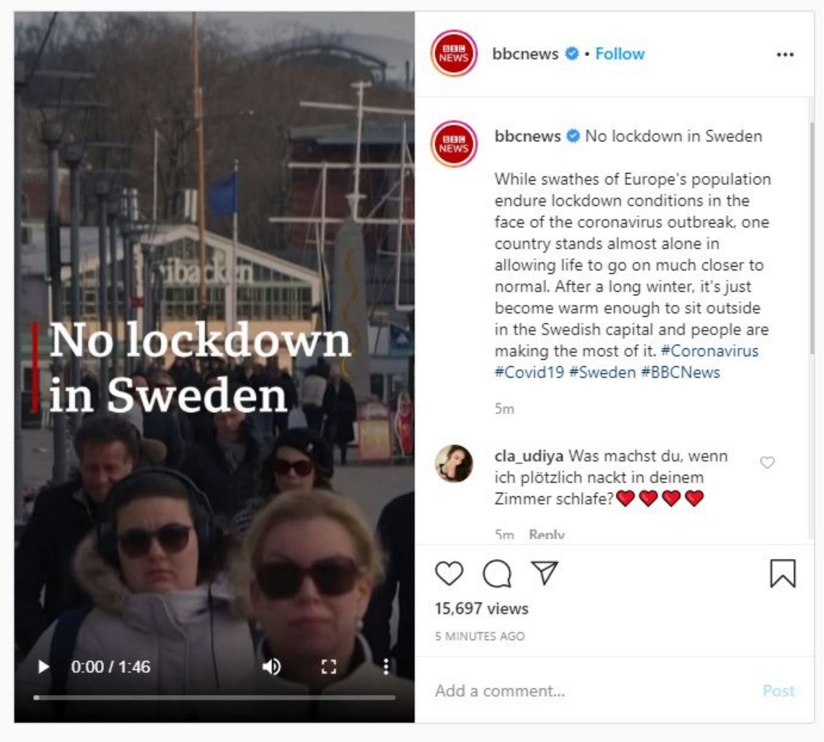

twan 14h
20m 1 like Reply
shekhar_.1 😂😂😂😂
20m 1 like Reply
visheshchaudhary_ What are you talking about?😂
20m 1 like Reply
sweet.rose1992 😂😂😂 damn😂
Glad you guys like the new video haha
meliitop @twan love u ❤️❤️❤️❤️❤️❤️❤️
20m 1 like Reply
lxnely.child LMAOOO
20m 1 like Reply
meliitop @twan hiiiii🤩
20m 1 like Reply
hebahanine0 😂😂😂
20m 1 like Reply

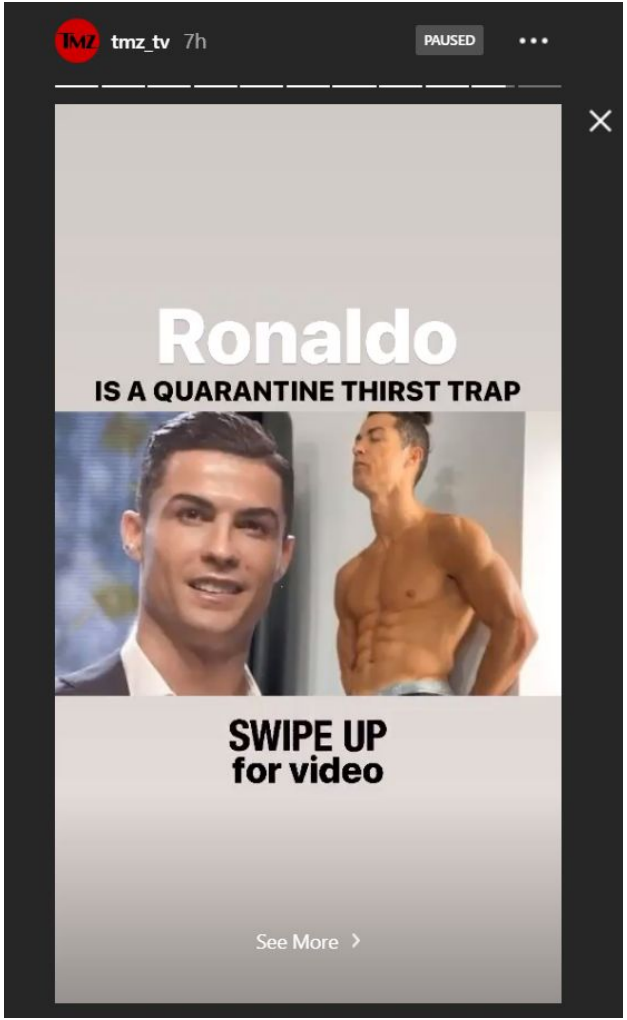

vanilladingdong • Follow 8h
PAUSED
WHO IS THIS VLOG SQUAD MEMBER!?
swipe up
check out @jayboice's first youtube video in literally years hahaha
See More

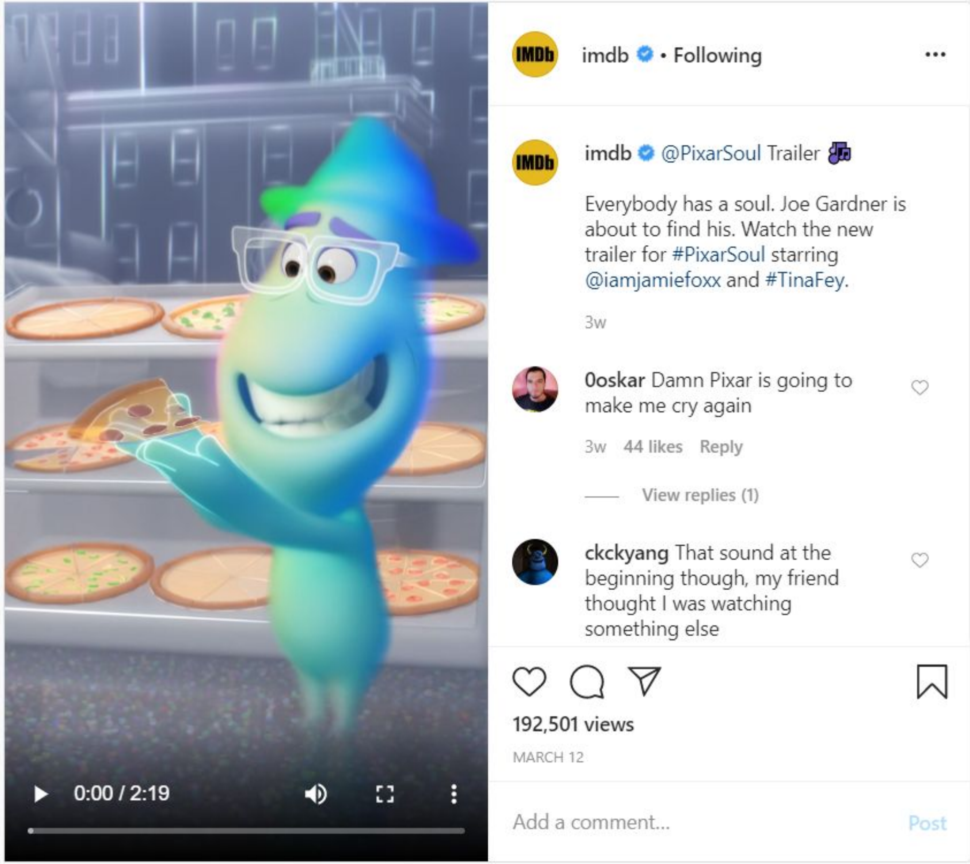

nickantonyan • Follow 12h
PAUSED
Full video! 👆
See More

Do you think people “have gamed the system”?


chiaraferragni
Following
12,721 posts
19.2m followers
1,088 following
Chiara Ferragni
Love fiercely(and don't forget to stop along the way to take photos) Founder
@chiaraferragnicollection. WATCH @chiaraferragniunposted ON PRIME VIDEO
www.gofundme.com/f/coronavirus-terapia-intensiva
Followed by isabellamaddiex, colette.s96, briannacyz + 3 more
Best of Leo 2
Business
Best of Leo
Documenta...
LancomexC...
Mini Leo
Wedding
POSTS
IGTV
TAGGED
1.6m
2,615
1:00
Italia ti amo IT Video by @oliverhl
3.9m
7,922
1:41
Proud to be italian, today and forever IT


No lockdown in Sweden
0:00 / 1:46
bbcnews • Follow
bbcnews No lockdown in Sweden
While swathes of Europe's population endure lockdown conditions in the face of the coronavirus outbreak, one country stands almost alone in allowing life to go on much closer to normal. After a long winter, it's just become warm enough to sit outside in the Swedish capital and people are making the most of it. #Coronavirus #Covid19 #Sweden #BBCNews
5m
cla_udiya Was machst du, wenn ich plötzlich nackt in deinem Zimmer schlafe?
5m Reply
15,697 views
5 MINUTES AGO
Add a comment...
Post


tmz_tv 7h
PAUSED
Ronaldo
IS A QUARANTINE THIRST TRAP
SWIPE UP
for video
See More


imdb • Following
imdb @PixarSoul Trailer
Everybody has a soul. Joe Gardner is about to find his. Watch the new trailer for #PixarSoul starring @iamjamiefoxx and #TinaFey.
3w
0oskar Damn Pixar is going to make me cry again
3w 44 likes Reply
View replies (1)
ckckyang That sound at the beginning though, my friend thought I was watching something else
192,501 views
MARCH 12
Add a comment...
Post
0:00 / 2:19

Thanks for attending Social Influencing Class 101

Deep Dive into Paper's Aim

- Can individuals with limited resources discover content production strategies with high payoffs in social networks? Let's disintegrate it:
  - 1. Does the preference order among strategies for authors stabilize over time, indicating the emergence of strategic behavior?
  - 2. If the preference order is stable, does the preference order maximize utility?
- Things to note:
  - 1. preference order stability does not imply high payoffs.
  - 2. the stability may arise due to other factors such as social norms.

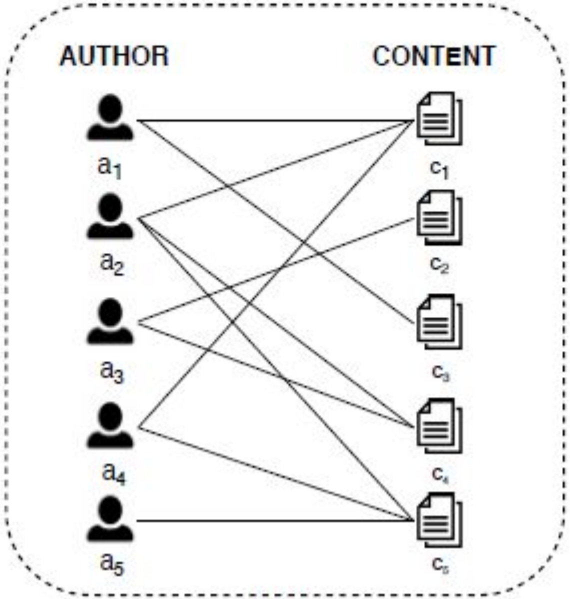
It's a non-trivial problem to solve.

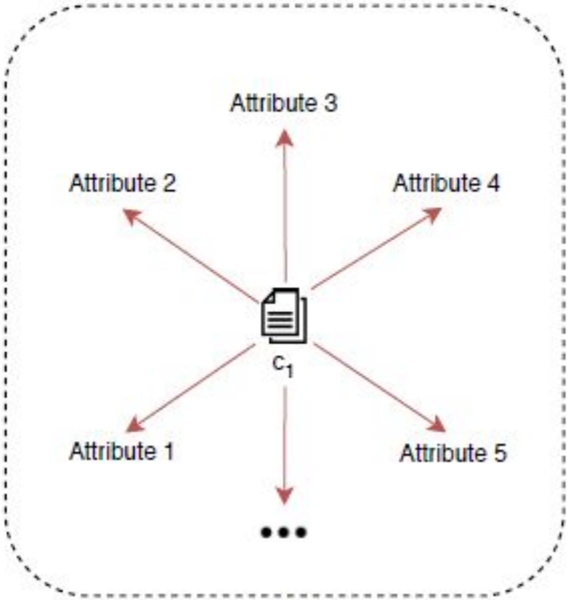
Technical Insight

- **Assumption**: Set of strategies is common to all, but each individual adopts a mixed-strategy over the set of different strategies.
- **Proposition**: Use bipartite graph (set of graph vertices decomposed into two disjoint sets such that no two graph vertices within the same set are adjacent) to conceptualize content production, where
    - Node: individuals and contents
    - Many to One relation between them: content may have multiple authors

Data Model

- set of individuals = A
    - Authors may either collaborate or work alone
- piece of content = c
    - content could be a blog, post, an academic paper, or when a group attempts on a answering Stack-Exchange question.





- construct an undirected bipartite graph where,
    - V = A ∪ C,
    - E = {(a, c) | a ∈ A, c ∈ C, a is an author of c},
- each content c ∈ C may have multiple attributes

Data Model

- attributes of c = venue of publication; time of publication; content topic; links to other entities including contents and authors.
- Author picks the attribute values strategically. Let every author use the same strategy space.
    - Strategy space = S, where |S| = m
- However, each author randomizes over them strategies independently.
    - When a group of authors collaborate, we assume that they negotiate and develop a consensus strategy.
- a's strategy distribution at time t = $D_a(t)$
    - Authors assume that the author's past strategy $D_a(t - 1)$ and her co-authors' strategy at time t influence Da(t). Authors call this as **Strategy distributions assumption**.

Data Model

Notation	Description
$G(t)$	Snapshot of the author-content graph at time t
$C_a(t)$	Set of contents created by a at time t
$\widetilde{\mathcal{A}}'(t)$	Set of authors with over 5 active contents at time t
$\boldsymbol{h}_a(t)$	Embedding vector of a at time t
$F_a(t)$	Field vector of a at time t
$D_a(t)$	Strategy distribution of a at time t
$r(a \mid c)$	a's contribution to c's strategy distribution
$\mu_a(t, k)$	Utility received by a with $D_a(t)$ over k time units
$\hat{\mu}_a(t-k)$	Normalized total utility received by a with $D_a(t-k)$
$\bar{\mu}_g(t, S)$	The global expected normalized utility for strategy S

Table 1: Notation table.

Data Model

- time-varying utility for content c = $\mu_c(k)$
- time of publishing the content = t
- time passed since content has been published = k
- utility after k time units (i.e t+k) = papers receive citations; up/down votes
  - Since each author contributes to a different extent to produce c, we assume

utility that flows back to author ∝ author's contribution

$$\mu_{a|c}(k) \ \alpha \ (\mu_c(k) \times r(a|c))$$

Authors call this as **Utility calculation assumption**.

$$\mu_a(t,k) = \sum_{c \in C_a(t)} \mu_c(k) \cdot r(a|c)$$

Refresh Paper's Aim

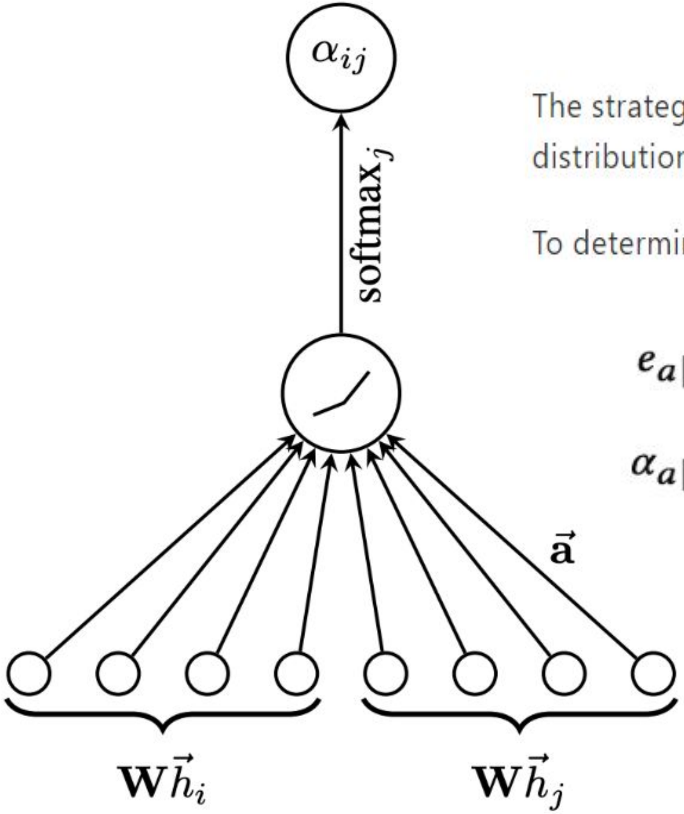
- Authors asks 2 questions:
  - How to determine the strategy distribution D_c for content c, jointly authored by a set of authors A_c?
  - Determine how the prior strategy distribution $D_a(t - 1)$ and the strategy distributions of the co-authors of a influence the strategy distribution $D_a(t)$.

Assumptions

- Vertex representation:
  - node embedding vector for content c = $h_c \in R^F$ and
  - node embedding vector for author a at time t = $h_a(t) \in R^F$
    - a time-dependent embedding vector for an author, by treating the same author at different times as separate nodes when embedding the network.
- Network snapshots:
  - Since the graph G = (V, E) grows over time, we divide the graph into snapshots.
  - If an author appears for the first time in snapshot t , we draw the prior strategy distributions $D_a(t-1)$ from a *flat Dirichlet distribution* and use an all zero vector as the prior embedding $h_a(t-1)$.

Dynamic Dual Attention Networks (DDAN)

Determining strategy for production of single content



The strategy distribution $\mathbf{D}_c$ of content c created at time t is affected by the strategy distribution $\mathbf{D}_a(t)$ of all authors $a \in \mathcal{A}_c$

To determine the contribution $\alpha_{a|c}$ of a specific author a towards $\mathbf{D}_c$

$$e_{a|c} = \sigma\left(\boldsymbol{\phi}_{c,a}^{\top} \cdot \left[\mathbf{W}_{c,a}\boldsymbol{h}_c \,||\, \mathbf{W}_{c,a}\boldsymbol{h}_a(t)\right]\right),$$

$$\alpha_{a|c} = \text{softmax}_a(e_{a|c}) = \frac{\exp(e_{a|c})}{\sum_{a' \in \mathcal{A}_c} \exp(e_{a'|c})},$$

Determining strategy for production of single content

Since $\boldsymbol{\alpha}_{a|c}$ is $\boldsymbol{a}$'s contribution to the determination of $\mathbf{D}_c$, set $\boldsymbol{r}(\boldsymbol{a}|\boldsymbol{c}) = \boldsymbol{\alpha}_{a|c}$

$$\boldsymbol{D}_c = \xi\left(\sum_{a \in \mathcal{A}_c} \alpha_{a|c} \cdot \boldsymbol{D}_a(t)\right).$$

- ξ: tanh nonlinear activation
- L1 normalization to ensure $\mathbf{D}_c$ is a valid strategy distribution

Determining an author's strategy

An authors's strategy $\mathbf{D}_a(t)$ depends on the strategy adopted for each content she authors at time t as well as her past strategy distribution $\mathbf{D}_a(t-1)$.

Examine the effect of the strategy for the production of content c where she is a co-author in $\mathcal{A}_c$.

$$e_{c|a} = \sigma\left(\boldsymbol{\phi}_{a,c}^{\intercal} \cdot \left[\mathbf{W}_{a,c}\boldsymbol{h}_a(t) \,||\, \mathbf{W}_{a,c}\boldsymbol{h}_c\right]\right),$$

$$\alpha_{c|a} = \mathrm{softmax}_c(e_{c|a}) = \frac{\exp(e_{c|a})}{\sum_{c' \in C_a(t)} \exp(e_{c'|a})}.$$

$\alpha_{c|a}$: content c's contribution on author a's strategy distribution $\mathbf{D}_a(t)$

Determining an author's strategy

To determine the contribution of a's strategy distribution at time $t-1$ on her current strategy distribution.

$$\beta_a(t) = \text{sigmoid}\left(\boldsymbol{\phi}_{a,a}^{\top} \cdot \left[\mathbf{W}_{a,a}\boldsymbol{h}_a(t) \,||\, \mathbf{W}_{a,a}\boldsymbol{h}_a(t-1)\right]\right).$$

a's strategy distribution $\mathbf{D}_a(t)$ at time t is the weighted sum of strategy distribution $\mathbf{D}_c$ for $c \in C_a(t)$ and $\mathbf{D}_a(t-1)$.

$$\boldsymbol{D}_a(t) = \xi\left(\beta_a(t)\boldsymbol{D}_a(t-1) + (1-\beta_a(t)) \sum_{c \in C_a(t)} \alpha_{c|a} \cdot \boldsymbol{D}_c\right).$$

Model for Rational Behavior

- An author engaged in rational behavior would be able to evaluate the utilities of all strategies and be able to identify the optimal strategy
- Ask:

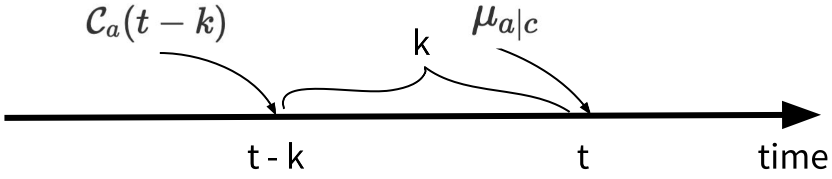
Given the utility at time t of content co-authored by a at time $t-k$.
What is the utility of author a using strategies distribution $\mathbf{D}_a(t-k)$?

where $\mu_{a|c}(k)$ is the utility that flows back to a after k time

Model for Rational Behavior

$$\hat{\mu}_a(t-k) = \underbrace{\frac{1}{k}}_{\text{time}} \times \frac{\sum_{c \in C_a(t-k)} \mu_{a|c}(t-k)}{\underbrace{|C_a(t-k)|}_{\text{count}}}$$

- $\hat{\mu}_a(t-k)$: normalized utility due to the distribution $\mathbf{D}_a(t-k)$
- $\mu_{a|c}(t-k)$: relative utility
- $c \in \mathcal{C}_a(t-k)$: participation in creation



Experiments

Datasets

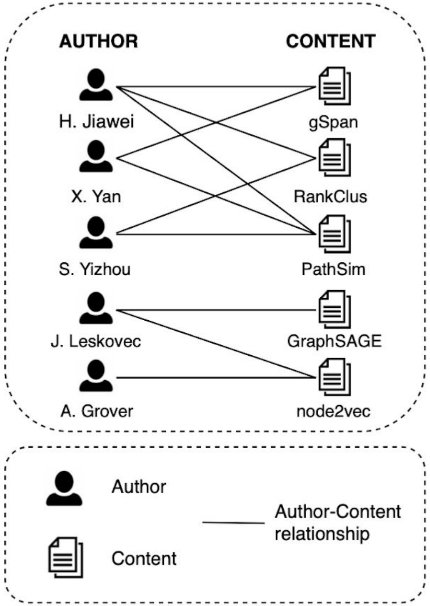
- DBLP academic dataset [33, 36]
- Goal is to discover strategic behavior associated with two paper attributes

1. Citations: Whom to cite?
2. Publication Venue: Where to publish?

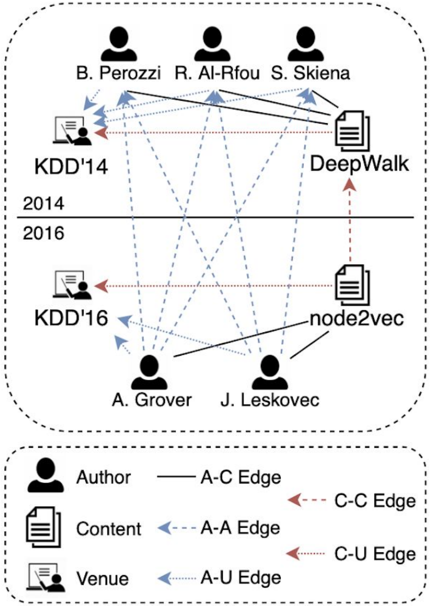
[33] Arnab Sinha, Zhihong Shen,Yang Song, Hao Ma, Darrin Eide, Bo-june Paul Hsu, and Kuansan Wang. 2015. An overview of microsoft academic service (mas) and applications. In Proceedings of the 24th international conference on world wide web. ACM, 243–246.
[36] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnet-Miner Extraction and Mining of Academic Social Networks. In KDD'08. 990–998.

Strategy Spaces

- One can create additional graphs using the attributes of each paper
  - $G_{a,c}$: represents the content production, connecting authors to the content
  - $G_{a,a}$: author-author citation graph
  - $G_{c,c}$: paper-paper citation graph
  - $G_{c,u}$: paper-location graph
  - $G_{a,u}$: author-location graph



(b)



(d)

Strategy Spaces

Identify four aspects

1. Popularity
   - Preferential attachment: The probability of citing a past paper is proportional to its citations
   - Uniform attachment
2. Similarity of field
   - Preferring similar fields
   - Preferring distinct fields
3. Familiarity
   - Preferring nodes: Cite other papers based on authorship
   - Preferring unfamiliar nodes
4. Time recency
   - Preferring small time gaps
   - Choose Random time gaps

Aspect	Strategy
Popularity	$s_{1,0}$, preferential attachment
	$s_{1,1}$, uniform attachment
Field	$s_{2,0}$, preferring similar fields
	$s_{2,1}$, preferring distinct fields
Familiarity	$s_{3,0}$, preferring familiar nodes
	$s_{3,1}$, preferring unfamiliar nodes
Time	$s_{4,0}$, preferring small time gaps
	$s_{4,1}$, choosing random time gaps

Strategy Spaces

- Consider a paper c_1 that cites c_2 and is published at location u_1, explain directed edges (c_1, c_2) and (c_1, u_1)
- Composite Strategies: Likelihood of the edge is a composite of each pure strategy
- (c_1, c_2) : popularity, field, familiarity and time recency, $2^4 = 16$ composite strategies
- (c_1, u_1) : popularity, field, familiarity, $2^3 = 8$ composite strategies
- Example of a citation strategy: $S_4^c = s_{1,0} \times s_{2,0} \times s_{3,1} \times s_{4,0}$
- Example of a location strategy: $S_6^l = s_{1,0} \times s_{2,1} \times s_{3,1}$

DDAN Training & Optimization

- Consider loss function for graph $G_{c,c}$

$$L_{c,c}(t) = \sum_{(c_i,c_j)\in\mathcal{E}_{c,c}(t)} -\log \sum_{S_i\in\mathcal{S}} P(S_i \mid D_{c_i}) \cdot \ell((c_i, c_j) \mid S_i).$$

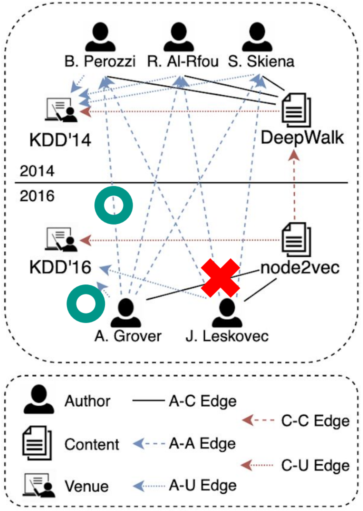
- $P(S_i|\mathbf{D}_{c_i})$: probability of picking strategy S_i given the distribution $\mathbf{D}_{c_i}$ for content c_i
- $\ell((c_i, c_j)|S_i)$: likelihood of edge (c_i, c_j) given strategy S_i

- Overall loss function

$$L(t) = L_{c,c}(t) + L_{c,u}(t) + L_{a,a}(t) + L_{a,u}(t)$$

Experiment Settings

- Apply to the task of link prediction
- Identify the set of authors with over five new contents in the current snapshot and partition each author's contents for 5-fold cross validation
- Hide the author-content edges and model aims to recover the hidden attribute edges
- Baselines
  - Logistic Regression (LR)
  - Dirichlet Multinomial Mixture Model (DMM)
  - Topic Over Time (TOT)



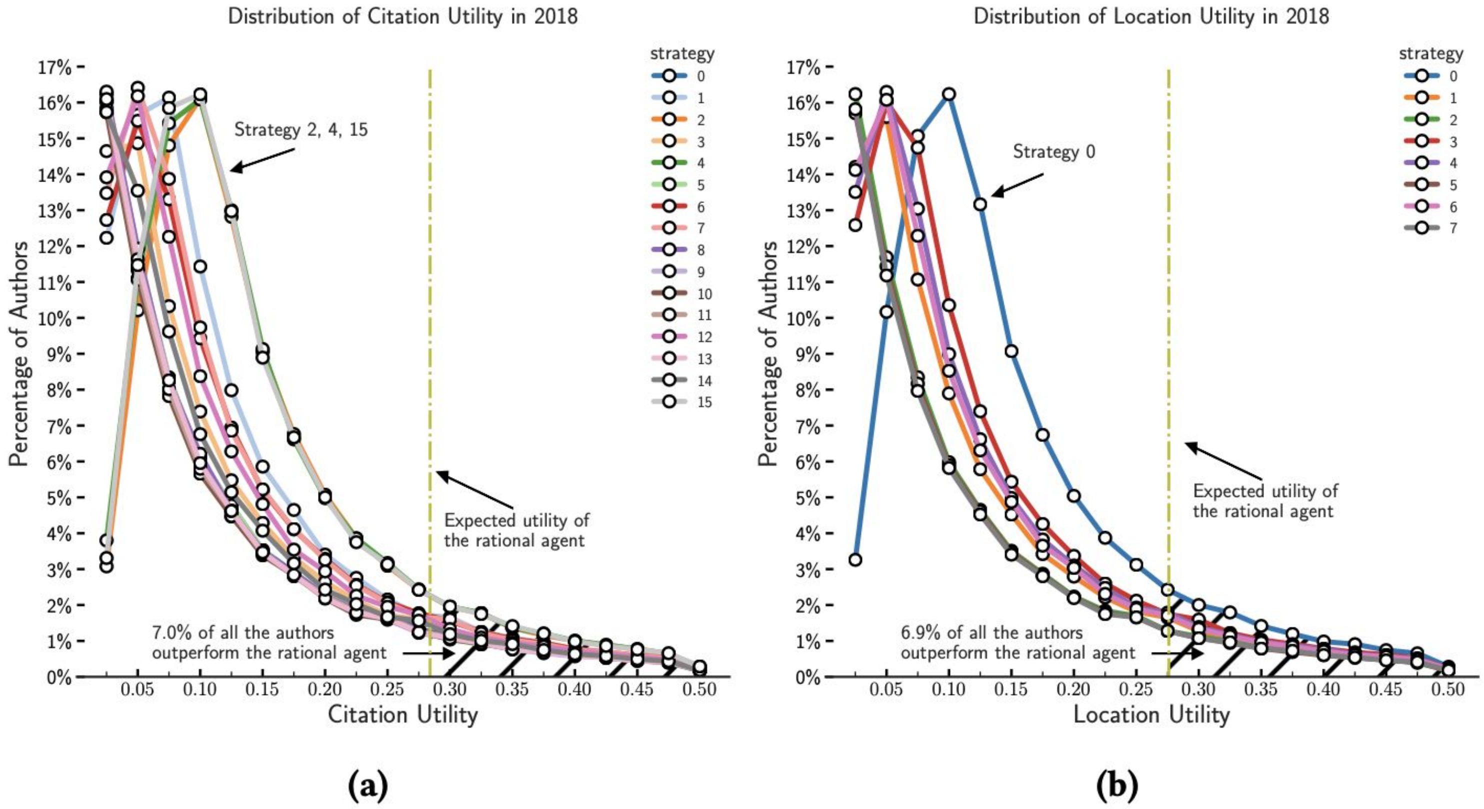
Experiment Results

Year	Strategies	LR [13]	DMM [45]	TOT [41]	DDAN
2000	Citation	0.72	0.72	0.73	**0.74**
	Publication	0.71	0.70	0.73	**0.75**
2005	Citation	0.69	0.69	0.70	**0.71**
	Publication	0.69	0.69	0.72	**0.73**
2010	Citation	0.67	0.67	0.68	**0.69**
	Publication	0.71	0.71	0.73	**0.74**
2015	Citation	0.67	0.67	0.68	**0.69**
	Publication	0.72	0.72	0.74	**0.75**
2018	Citation	0.67	0.67	0.68	**0.69**
	Publication	0.76	0.75	0.77	**0.78**

Table 4: Experiment results using Mean Average Precision (MAP) as the evaluation metrics. DDAN achieves the highest scores for both strategies in all testing snapshots.

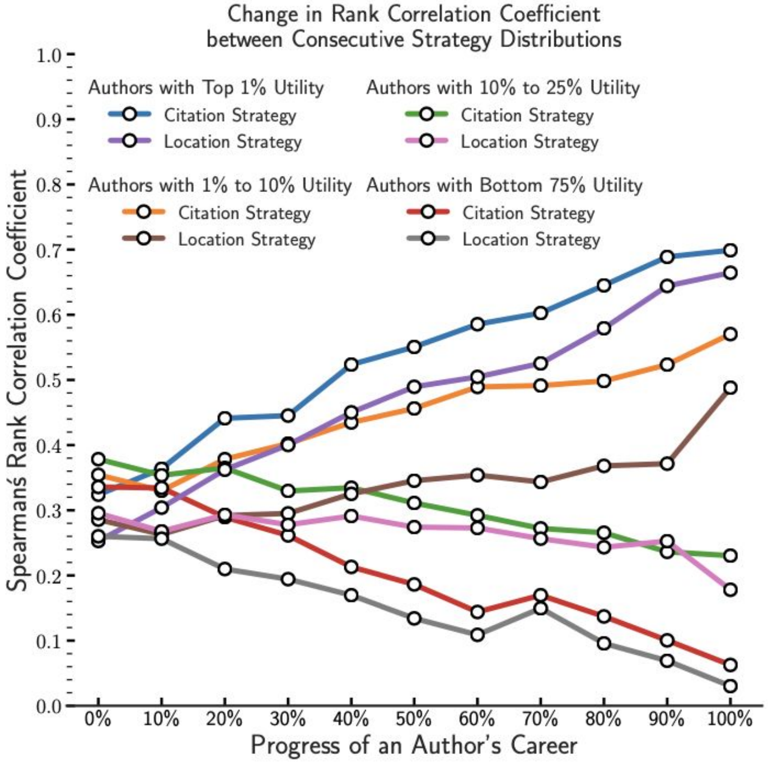
Qualitative Analysis

Do strategies matter?



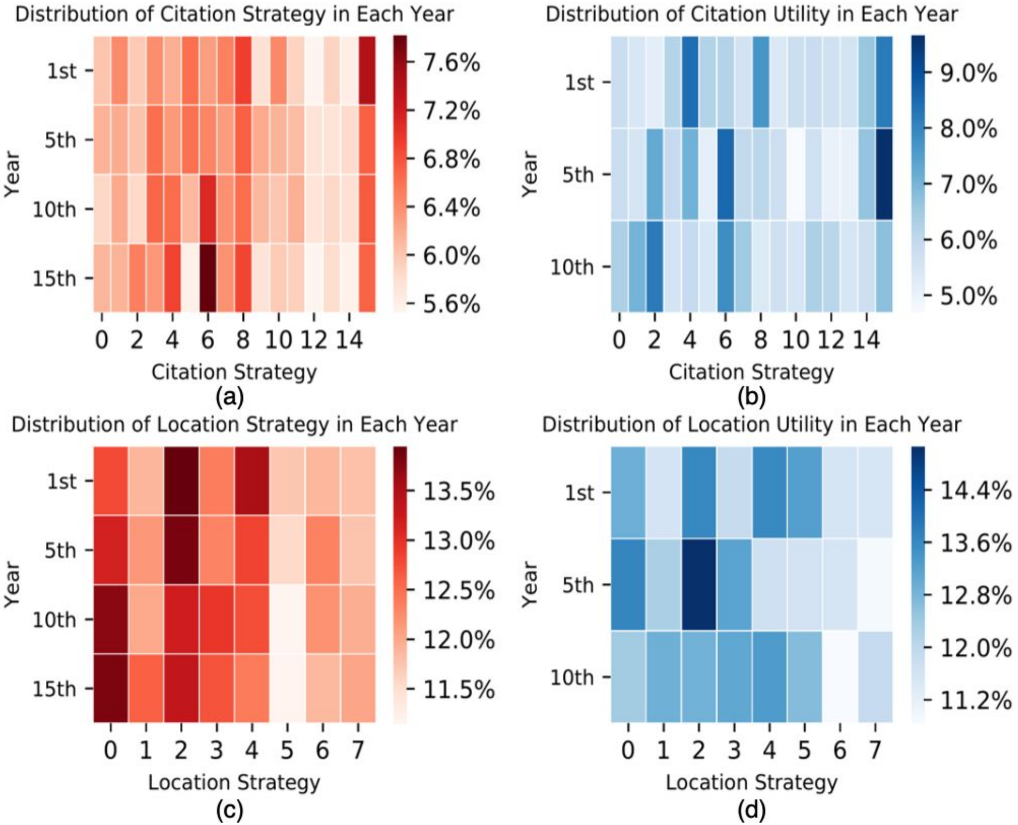
Emergence of Order

- Compute the Spearman rank correlation coefficient
- The correlations **increase** for authors with the normalized utility <u>in the top 10%</u>

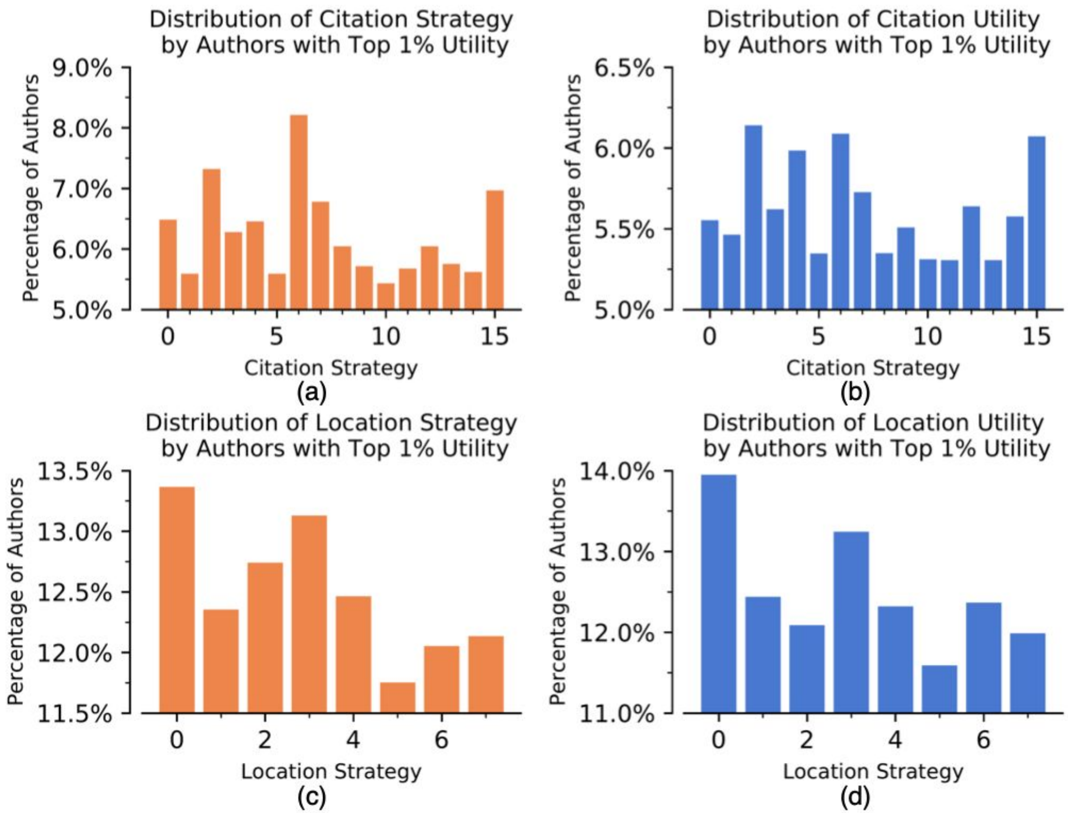


(c)

Emergence of Order



Stability and Payoffs



Discussion

- Strengths
  - Conceptualize the observed behavior as a generative process
  - Encode hand-crafted strategies gives the opportunity to interpret the network
- Weaknesses
  - Come up with a complete strategy space is not trivial
  - Rational model is myopic
  - Lack of explanations on why resources such as social norm limited the model

Questions