OPTIMISING SEEDING STRATEGIES BY INCORPORATING EMPIRICAL EVIDENCE INTO INFLUENCE MAXIMISATION MODELS

Radu Tanase, Manuel S. Mariani, René Algesheimer

URPP Social Networks, University of Zurich

Spreading a new cookbook

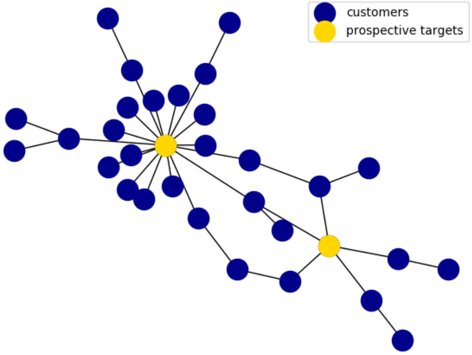
- *CookHero* is a manufacturer of kitchen appliances.
- Manages an **online community of customers** who can create and discover recipes, and **establish social connections**.

Spreading a new cookbook

- *CookHero* is a manufacturer of kitchen appliances.
- Manages an **online community of customers** who can create and discover recipes, and **establish social connections**.

*CookHero wants to promote a **cookbook on sustainable recipes** on the platform. **Which customers to target first?***

Which customer to target?



Related literature and open questions

Evidence supporting seeding the social hubs.

- Simulation-based and theoretical studies [Valente and Davis (1999); Kempe et al. (2003); Kitsak et al. (2010); Libai et al. (2013); Peres (2014)].
- Observational evidence [Goldenberg et al. 2009].
- Field experiments [Hinz et al. (2011); Banerjee et al. (2013)].

Open questions.

- Behavioral characteristics affect adoption choices [e.g., Bearden et al. (1989); Aral and Walker (2012)].
  Can we improve seeding performance by incorporating empirical, individual-level behavioral characteristics?
- Arbitrary assumption at the individual-level in simulation studies. → Misleading diffusion predictions.
  How to calibrate diffusion simulations with empirical individual-level behavior?

Related literature and open questions

Evidence supporting seeding the social hubs.

- Simulation-based and theoretical studies [Valente and Davis (1999); Kempe et al. (2003); Kitsak et al. (2010); Libai et al. (2013); Peres (2014)].
- Observational evidence [Goldenberg et al. 2009].
- Field experiments [Hinz et al. (2011); Banerjee et al. (2013)].

Open questions.

- Behavioral characteristics affect adoption choices [e.g., Bearden et al. (1989); Aral and Walker (2012)].
  Can we improve seeding performance by incorporating empirical, individual-level behavioral characteristics?
- Arbitrary assumption at the individual-level in simulation studies. $\rightarrow$ Misleading diffusion predictions.
  How to calibrate diffusion simulations with empirical individual-level behavior?

Threshold model of new product diffusion

- **Adoption rule.** Individual i adopts only if at least a **threshold fraction** τ_i of her social contacts already adopted.
- Vast-cross disciplinary literature [Granovetter (1979); Watts (2002); Kempe et al. (2003); Centola and Macy (2007); Watts and Dodds (2007)].
- Lack of empirical evidence on individuals' adoption threshold levels [Peres et al. (2010)].
- Misspecified thresholds $\rightarrow$ misleading predictions of diffusion success.

> *How to estimate individual-level thresholds?*

Threshold model of new product diffusion

- **Adoption rule.** Individual i adopts only if at least a **threshold fraction** τ_i of her social contacts already adopted.
- Vast-cross disciplinary literature [Granovetter (1979); Watts (2002); Kempe et al. (2003); Centola and Macy (2007); Watts and Dodds (2007)].
- Lack of empirical evidence on individuals' adoption threshold levels [Peres et al. (2010)].
- Misspecified thresholds $\rightarrow$ misleading predictions of diffusion success.

> *How to estimate individual-level thresholds?*

Threshold model of new product diffusion

- **Adoption rule.** Individual i adopts only if at least a **threshold fraction** τ_i of her social contacts already adopted.
- Vast-cross disciplinary literature [Granovetter (1979); Watts (2002); Kempe et al. (2003); Centola and Macy (2007); Watts and Dodds (2007)].
- Lack of empirical evidence on individuals' adoption threshold levels [Peres et al. (2010)].
- Misspecified thresholds $\rightarrow$ misleading predictions of diffusion success.

How to estimate individual-level thresholds?

Individual-level choice modeling

- Modeling *i*'s choice: Whether to adopt product α or not [Train (2009); Miller, Hofstetter et al. (2011)].
- Not adopting: $v_{i\alpha}^{(0)}$.
- Adopting: $v_{i\alpha}^{(1)} = v_{i\alpha}^{(A)} + v_{i\alpha}^{(S)}$.
- **Attribute utility**: $v_{i\alpha}^{(A)}$ (linear combination of α's attributes).
- **Social utility**: $v_{i\alpha}^{(S)} = \beta_i^{(S)} s_{i\alpha}$, where $s_{i\alpha}$ social signal received by i about α.

> *i's threshold $\tau_{i\alpha}$: minimum s such that $v_{i\alpha}^{(1)} > v_{i\alpha}^{(0)}$.*
> *Choice data $\rightarrow$ Hierarchical Bayes $\rightarrow$ Estimated thresholds.*

Individual-level choice modeling

- Modeling i's choice: Whether to adopt product α or not [Train (2009); Miller, Hofstetter et al. (2011)].
- Not adopting: $v_{i\alpha}^{(0)}$.
- Adopting: $v_{i\alpha}^{(1)} = v_{i\alpha}^{(A)} + v_{i\alpha}^{(S)}$.
- **Attribute utility**: $v_{i\alpha}^{(A)}$ (linear combination of α's attributes).
- **Social utility**: $v_{i\alpha}^{(S)} = \beta_i^{(S)} s_{i\alpha}$, where $s_{i\alpha}$ social signal received by i about α.

> *i's threshold $\tau_{i\alpha}$: minimum s such that $v_{i\alpha}^{(1)} > v_{i\alpha}^{(0)}$.*
> *Choice data $\rightarrow$ Hierarchical Bayes $\rightarrow$ Estimated thresholds.*

Individual-level choice modeling

- Modeling *i*'s choice: Whether to adopt product α or not [Train (2009); Miller, Hofstetter et al. (2011)].
- Not adopting: $v_{i\alpha}^{(0)}$.
- Adopting: $v_{i\alpha}^{(1)} = v_{i\alpha}^{(A)} + v_{i\alpha}^{(S)}$.
- **Attribute utility**: $v_{i\alpha}^{(A)}$ (linear combination of α's attributes).
- **Social utility**: $v_{i\alpha}^{(S)} = \beta_i^{(S)} s_{i\alpha}$, where $s_{i\alpha}$ social signal received by i about α.

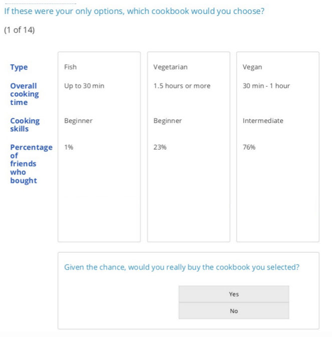
> *i's threshold $\tau_{i\alpha}$: minimum s such that $v_{i\alpha}^{(1)} > v_{i\alpha}^{(0)}$.*
> *Choice data $\rightarrow$ Hierarchical Bayes $\rightarrow$ Estimated thresholds.*

Individual-level choice modeling

- Modeling *i*'s choice: Whether to adopt product α or not [Train (2009); Miller, Hofstetter et al. (2011)].
- Not adopting: $v_{i\alpha}^{(0)}$.
- Adopting: $v_{i\alpha}^{(1)} = v_{i\alpha}^{(A)} + v_{i\alpha}^{(S)}$.
- **Attribute utility**: $v_{i\alpha}^{(A)}$ (linear combination of α's attributes).
- **Social utility**: $v_{i\alpha}^{(S)} = \beta_i^{(S)} s_{i\alpha}$, where $s_{i\alpha}$ social signal received by *i* about α.

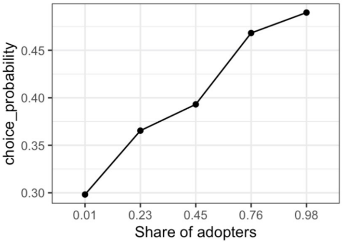
> *i's threshold $\tau_{i\alpha}$: minimum s such that $v_{i\alpha}^{(1)} > v_{i\alpha}^{(0)}$.*
> *Choice data $\rightarrow$ Hierarchical Bayes $\rightarrow$ Estimated thresholds.*

Choice-based conjoint experiment: Cookbook choice



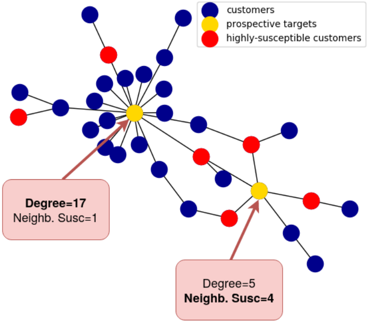
- Type (meat / fish / vegetarian / vegan)
- Overall cooking time (Up to 30 min / 30 min - 1h / 1.5h)
- Recommended cooking skills (beginner / medium / advanced)
- **Social signal.** Percentage of friends who bought the cookbook (1%, 23%, 45%, 76%, 98%).

Choice-based conjoint experiment: Pilot study results ($n = 303$)



- The choice probability increases with the social signal.
- **Accurate prediction of out-of-sample choices**.
  Out-of-sample accuracy: 70%.
- Estimated thresholds and susceptibility scale [Bearden et al. (1989)]: $r = -0.19 (p < 0.05)$.

Seeding policies



- **Centrality-based: Degree**. Top-5% by number of social contacts.
- **Behavior-based: Neighborhood susceptibility**. Top-5% by number of **low-threshold** social contacts, according the estimated thresholds.

Policy evaluation

Ground-truth properties

Ground-truth utilities [Hein et al. (2020)] $\rightarrow$ Simulated experiments and choice data $\rightarrow$ **Estimated thresholds for seeding policies.**

Simulations:

- 64 products $\times$ 10 network structures.
- **Seed set** (5%) determined by the degree or neighborhood susceptibility.
- For each network, product, and seed set, we run simulations with the ground-truth thresholds $\rightarrow$ **Ground-truth diffusion success.**
- ROI $=$ ground-truth diffusion success/cost
- **Cost function** [Bakshy et al. (2011)]: cost $= \alpha + \beta$ degree

Policy evaluation

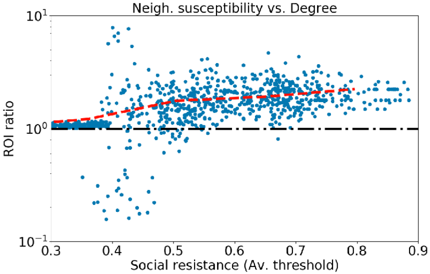
Ground-truth properties

Ground-truth utilities [Hein et al. (2020)] $\rightarrow$ Simulated experiments and choice data $\rightarrow$ **Estimated thresholds for seeding policies**.

Simulations:

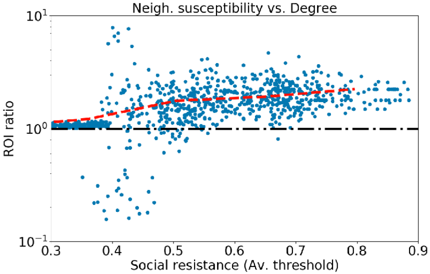
- 64 products $\times$ 10 network structures.
- **Seed set** (5%) determined by the degree or neighborhood susceptibility.
- For each network, product, and seed set, we run simulations with the ground-truth thresholds $\rightarrow$ **Ground-truth diffusion success**.
- ROI $=$ ground-truth diffusion success/cost
- **Cost function** [Bakshy et al. (2011)]: cost $= \alpha + \beta$ degree

The neighborhood susceptibility outperforms the centrality



Integrating behavioral information is especially beneficial for high-resistance products.

The neighborhood susceptibility outperforms the centrality



Integrating behavioral information is especially beneficial for high-resistance products.

Managerial implications

+ **Social network data.**
  Social network data about all customers: difficult.
  Instead, identify **highly-susceptible (low-threshold) customers**, and only collect social network data for those.

+ **Macro vs micro influencers.**
  Selecting the right "niche influencers" far from trivial [Haenlein et al. (2020)].
  Possible approach: look at the **susceptibility (thresholds)** of a prospect influencer's audience.

Managerial implications

- <u>**Social network data.**</u>
  Social network data about all customers: difficult.
  Instead, identify **highly-susceptible (low-threshold) customers**, and only collect social network data for those.

- <u>**Macro vs micro influencers.**</u>
  Selecting the right "niche influencers" far from trivial [Haenlein et al. (2020)].
  Possible approach: look at the **susceptibility (thresholds)** of a prospect influencer's audience.

Managerial implications

+ **Social network data.**
  Social network data about all customers: difficult.
  Instead, identify **highly-susceptible (low-threshold) customers**, and only collect social network data for those.

+ **Macro vs micro influencers.**
  Selecting the right "niche influencers" far from trivial [Haenlein et al. (2020)].
  Possible approach: look at the **susceptibility (thresholds)** of a prospect influencer's audience.

Managerial implications

- **Social network data.**
  Social network data about all customers: difficult.
  Instead, identify **highly-susceptible (low-threshold) customers**, and only collect social network data for those.

- **Macro vs micro influencers.**
  Selecting the right "niche influencers" far from trivial [Haenlein et al. (2020)].
  Possible approach: look at the **susceptibility (thresholds)** of a prospect influencer's audience.

Contributions

1. Experimental method to **estimate individual-level thresholds**.
2. Integrating **micro-level** experimental thresholds into agent-based simulations of **macro-level** diffusion.
3. **Seeding policies** that combine the estimated individual-level thresholds and network structure can improve seeding performance.

Thank you.

Radu Tanase, Manuel S. Mariani, René Algesheimer

URPP Social Networks, University of Zurich

manuel.mariani@business.uzh.ch

http://www.socialnetworks.uzh.ch

Integrating individual-level and collective descriptions of adoption behavior

- **Focus on the individual.** Choice models and experiments, Consumer behavior.
- **Focus on the collective dynamics.** Diffusion of innovations, Agent-based models.
- Integrating the two approaches follows recent suggestions.
  - *"Intensive research is still needed to build firm theoretical and empirical infrastructures for **choice-based growth models**"* [Peres et al. (2010)].
  - *"Researchers who conduct their research with human subjects in controlled laboratory settings should consider using ABM techniques to analyze the **"macro" implications of their "micro" findings**"* [Smith and Rand (2018)].

Integrating individual-level and collective descriptions of adoption behavior

- **Focus on the individual**. Choice models and experiments, Consumer behavior.
- **Focus on the collective dynamics**. Diffusion of innovations, Agent-based models.
- Integrating the two approaches follows recent suggestions.
  - *"Intensive research is still needed to build firm theoretical and empirical infrastructures for **choice-based growth models**"* [Peres et al. (2010)].
  - *"Researchers who conduct their research with human subjects in controlled laboratory settings should consider using ABM techniques to analyze the **"macro" implications of their "micro" findings**"* [Smith and Rand (2018)].

Difference compared to conventional simulation-based studies

+ **Traditional simulation-based arguments.**
  Diffusion models / agent-based simulations not calibrated with individual-level behavior.
  E.g., Viral spreading models, Bass model.
+ Arbitrary assumptions at the individual level → Misleading aggregate predictions.

Our approach

+ Estimate individual-level behavioral characterstics (susceptibility) in **choice experiments**.
+ Use the results to **calibrate agent-based simulations.**

Difference compared to conventional simulation-based studies

- **Traditional simulation-based arguments.**
  Diffusion models / agent-based simulations not calibrated with individual-level behavior.
  E.g., Viral spreading models, Bass model.
- Arbitrary assumptions at the individual level $\rightarrow$ Misleading aggregate predictions.

Our approach

- Estimate individual-level behavioral characterstics (susceptibility) in **choice experiments.**
- Use the results to **calibrate agent-based simulations.**

Difference compared to conventional simulation-based studies

- **Traditional simulation-based arguments.**
  Diffusion models / agent-based simulations not calibrated with individual-level behavior.
  E.g., Viral spreading models, Bass model.
- Arbitrary assumptions at the individual level $\rightarrow$ Misleading aggregate predictions.

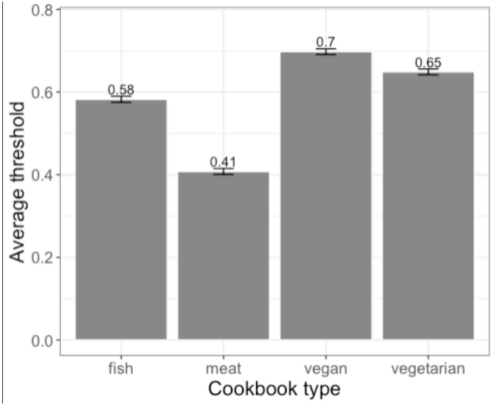
Our approach

- Estimate individual-level behavioral characterstics (susceptibility) in **choice experiments**.
- Use the results to **calibrate agent-based simulations**.

The obtained thresholds correlate negatively with the susceptibility scale and predict unseen choices

Average Importances	Average Importances	Standard Deviation
Type	47.17824	18.59446
Overall cooking time	14.28970	8.10268
Cooking skills	15.49785	11.27420
Percentage of friends who bought	23.03421	14.57720
In-sample prediction	85%	
Out of sample prediction	70%	
Fixed question 1	94%	
Fixed question 2	83%	
Correlation consumer susceptiblity scale	- 0.19 ***	

More sustainable cookbooks tend to face higher resistance



Future directions

At the individual level:

- Study tasks where there is a ground truth $\rightarrow$ participant performance.
- Understand how the threshold depends on the risk associated with the choice.
- Estimate individuals' threshold in observational data.

At the collective level:

- Laboratory experiments to validate the conclusions by agent-based simulations.

Simulated experiment

- Adopting product α:

$$u_{i\alpha} = v_{i\alpha}^{(S)} + v_{i\alpha}^{(A)} + \epsilon_{i\alpha}$$

- Noise: σ_ϵ^2.
- Hierarchical Bayes $\rightarrow$ Estimated utilities and thresholds $\rightarrow$ Estimation error

$$E = \overline{|\text{Estimated th.} - \text{Ground-truth th.}|}$$

Preliminary result

- Low noise ($\sigma_\epsilon^2 = 1$). $E \simeq 0.10$
- High noise ($\sigma_\epsilon^2 = 2$). $E \simeq 0.12$

Simulated experiment

- Adopting product α:

$$u_{i\alpha} = v_{i\alpha}^{(S)} + v_{i\alpha}^{(A)} + \epsilon_{i\alpha}$$

- Noise: σ_ϵ^2.
- Hierarchical Bayes $\rightarrow$ Estimated utilities and thresholds $\rightarrow$ Estimation error

$$E = \overline{|\text{Estimated th.} - \text{Ground-truth th.}|}$$

Preliminary result

- Low noise ($\sigma_\epsilon^2 = 1$). $E \simeq 0.10$
- High noise ($\sigma_\epsilon^2 = 2$). $E \simeq 0.12$

Thank you.

Radu Tanase, Manuel S. Mariani, René Algesheimer

URPP Social Networks, University of Zurich

manuel.mariani@business.uzh.ch

http://www.socialnetworks.uzh.ch