A Bounded Confidence Approach to Understand User Participation in Peer Production Systems

Giovanni Luca Ciampaglia

University of Lugano

Third Intl Conference on Social Informatics (SocInfo 2011)
Singapore, October 8, 2011.

Social Informatics

One key idea of social informatics research is that the "social context" of information technology development plays a significant role in influencing the ways that people use information and technologies, and thus influences their consequences for work, organizations, and other social relationships.

(Kling 1999)

Social computing

- Commons-based peer production (Benkler 2002)
- **0-th law of Wikipedia**: "*The problem with Wikipedia is that it only works in practice. In theory, it can never work.*"[1]
- Example: user participation to peer production.

[1] http://en.wikipedia.org/wiki/User:Raul654/Raul's_laws

Social computing

- Commons-based peer production (Benkler 2002)
- **0-th law of Wikipedia**: "*The problem with Wikipedia is that it only works in practice. In theory, it can never work.*"[1]
- Example: user participation to peer production.

[1] `http://en.wikipedia.org/wiki/User:Raul654/Raul's_laws`

Social computing

- Commons-based peer production (Benkler 2002)
- **0-th law of Wikipedia**: "*The problem with Wikipedia is that it only works in practice. In theory, it can never work.*"[1]
- Example: user participation to peer production.

[1] `http://en.wikipedia.org/wiki/User:Raul654/Raul's_laws`

An example: Wikipedia's NPOV policy

A short course of wiki writing:[2]

- "*Abortion is wrong*" – Wrong!
- "*The pro-life movement holds that abortion is wrong, or occasionally that it is only justified in certain special cases*" – Right!

[2] `http://en.wikipedia.org/wiki/Wikipedia:NPOV/Examples`

An example: Wikipedia's NPOV policy

A short course of wiki writing:[2]

- "*Abortion is wrong*" – Wrong!
- "*The pro-life movement holds that abortion is wrong, or occasionally that it is only justified in certain special cases*" – Right!

[2] http://en.wikipedia.org/wiki/Wikipedia:NPOV/Examples

An example: Wikipedia's NPOV policy

A short course of wiki writing:[2]

- "*Abortion is wrong*" – Wrong!
- "*The pro-life movement holds that abortion is wrong, or occasionally that it is only justified in certain special cases*" – Right!

[2] http://en.wikipedia.org/wiki/Wikipedia:NPOV/Examples

An example: Wikipedia's NPOV policy

A short course of wiki writing:[2]

- "*Abortion is wrong*" – Wrong!
- "*The pro-life movement holds that abortion is wrong, or occasionally that it is only justified in certain special cases*" – Right!

[2] `http://en.wikipedia.org/wiki/Wikipedia:NPOV/Examples`

An example: Wikipedia's NPOV policy

A short course of wiki writing:[2]

- "*Abortion is wrong*" – Wrong!
- "*The pro-life movement holds that abortion is wrong, or occasionally that it is only justified in certain special cases*" – Right!

[2] `http://en.wikipedia.org/wiki/Wikipedia:NPOV/Examples`

User participation and norm adoption

- Groups are characterized by norms
- Norms: approved behaviors to follows, implicit knowledge, shared beliefs
- Participation to peer production community requires certain norms to be adopted

User participation and norm adoption

- Groups are characterized by norms
- Norms: approved behaviors to follows, implicit knowledge, shared beliefs
- Participation to peer production community requires certain norms to be adopted

User participation and norm adoption

- Groups are characterized by norms
- Norms: approved behaviors to follows, implicit knowledge, shared beliefs
- Participation to peer production community requires certain norms to be adopted

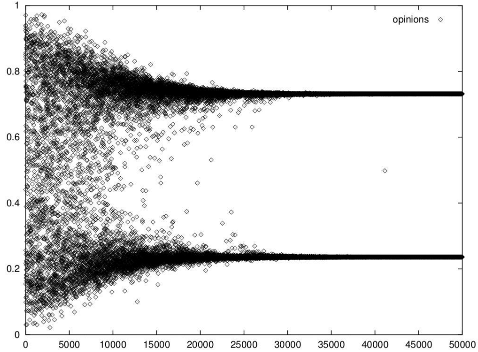
A theory of norm adoption

- Social judgement Theory (Sherif and Hovland 1961)
- Self-categorization Theory (Turner 1989)
- Formalization: **bounded confidence (BC)** principle
- If $||x(t) - y(t)|| < \varepsilon$:

$$\begin{aligned} x(t+1) &= x(t) + \mu\big(y(t) - x(t)\big) \\ y(t+1) &= y(t) + \mu\big(x(t) - y(t)\big) \end{aligned} \tag{1}$$

Norm adoption and coordination

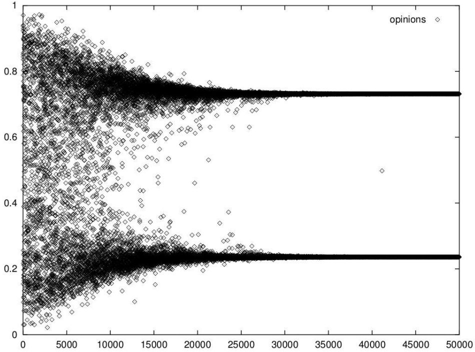
- Models of group formation
  (continuous opinion dynamics under bounded confidence)
- Group Consensus, polarization
- NOT tested empirically



(Deffuant et *al.* 2001)

Norm adoption and coordination

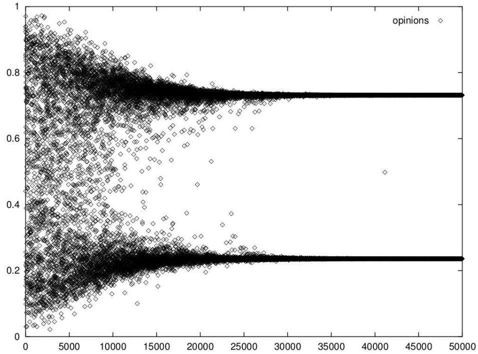
- Models of group formation (continuous opinion dynamics under bounded confidence)
- Group Consensus, polarization
- NOT tested empirically



(Deffuant et *al.* 2001)

Norm adoption and coordination

- Models of group formation (continuous opinion dynamics under bounded confidence)
- Group Consensus, polarization
- NOT tested empirically



(Deffuant et *al.* 2001)

Main ingredients of the model

- Dynamic population of users (arrival rate λ_u)
- New pages are created (creation rate λ_p)
- Users interact (interaction rate λ_e) with pages using BC rule
  - Users have initial motivation c
  - $r\left(t\right) = \frac{s(t)+c}{n(t)+c}$ fraction of “successful” edits
  - If attitude change: $r\left(t\right) \leftarrow r\left(t\right) + 1$
- Probability to abandon at time t:

$$\lambda_d\left(t\right) = \frac{r\left(t\right)}{\tau_0} + \frac{1 - r\left(t\right)}{\tau_1} \quad (2)$$

- Time scales: $\tau_0 > \tau_1$ (usually by orders of magnitude)

Main ingredients of the model

- Dynamic population of users (arrival rate λ_u)
- New pages are created (creation rate λ_p)
- Users interact (interaction rate λ_e) with pages using BC rule
  - Users have initial motivation c
  - $r(t) = \frac{s(t)+c}{n(t)+c}$ fraction of "successful" edits
  - If attitude change: $r(t) \leftarrow r(t) + 1$
- Probability to abandon at time t:

$$\lambda_d(t) = \frac{r(t)}{\tau_0} + \frac{1 - r(t)}{\tau_1} \tag{2}$$

- Time scales: $\tau_0 > \tau_1$ (usually by orders of magnitude)

Main ingredients of the model

- Dynamic population of users (arrival rate λ_u)
- New pages are created (creation rate λ_p)
- Users interact (interaction rate λ_e) with pages using BC rule
  - Users have initial motivation c
  - $r(t) = \frac{s(t)+c}{n(t)+c}$ fraction of "successful" edits
  - If attitude change: $r(t) \leftarrow r(t) + 1$
- Probability to abandon at time t:

$$\lambda_d(t) = \frac{r(t)}{\tau_0} + \frac{1 - r(t)}{\tau_1} \tag{2}$$

- Time scales: $\tau_0 > \tau_1$ (usually by orders of magnitude)

Main ingredients of the model

- Dynamic population of users (arrival rate λ_u)
- New pages are created (creation rate λ_p)
- Users interact (interaction rate λ_e) with pages using BC rule
  - Users have initial motivation c
  - $r(t) = \frac{s(t)+c}{n(t)+c}$ fraction of "successful" edits
  - If attitude change: $r(t) \leftarrow r(t) + 1$
- Probability to abandon at time t:

$$\lambda_d(t) = \frac{r(t)}{\tau_0} + \frac{1 - r(t)}{\tau_1} \quad (2)$$

- Time scales: $\tau_0 > \tau_1$ (usually by orders of magnitude)

Main ingredients of the model

- Dynamic population of users (arrival rate λ_u)
- New pages are created (creation rate λ_p)
- Users interact (interaction rate λ_e) with pages using BC rule
  - Users have initial motivation c
  - $r(t) = \frac{s(t)+c}{n(t)+c}$ fraction of “successful” edits
  - If attitude change: $r(t) \leftarrow r(t) + 1$
- Probability to abandon at time t:

$$\lambda_d(t) = \frac{r(t)}{\tau_0} + \frac{1 - r(t)}{\tau_1} \qquad (2)$$

- Time scales: $\tau_0 > \tau_1$ (usually by orders of magnitude)

Main ingredients of the model

- Dynamic population of users (arrival rate λ_u)
- New pages are created (creation rate λ_p)
- Users interact (interaction rate λ_e) with pages using BC rule
  - Users have initial motivation c
  - $r(t) = \frac{s(t)+c}{n(t)+c}$ fraction of “successful” edits
  - If attitude change: $r(t) \leftarrow r(t) + 1$
- Probability to abandon at time t:

$$\lambda_d(t) = \frac{r(t)}{\tau_0} + \frac{1 - r(t)}{\tau_1} \quad (2)$$

- Time scales: $\tau_0 > \tau_1$ (usually by orders of magnitude)

Main ingredients of the model

- Dynamic population of users (arrival rate λ_u)
- New pages are created (creation rate λ_p)
- Users interact (interaction rate λ_e) with pages using BC rule
  - Users have initial motivation c
  - $r(t) = \frac{s(t)+c}{n(t)+c}$ fraction of “successful” edits
  - If attitude change: $r(t) \leftarrow r(t) + 1$
- Probability to abandon at time t:

$$\lambda_d(t) = \frac{r(t)}{\tau_0} + \frac{1 - r(t)}{\tau_1} \qquad (2)$$

- Time scales: $\tau_0 > \tau_1$ (usually by orders of magnitude)

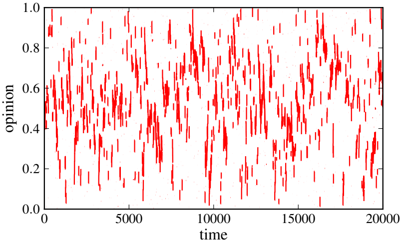
Main ingredients of the model

- Dynamic population of users (arrival rate λ_u)
- New pages are created (creation rate λ_p)
- Users interact (interaction rate λ_e) with pages using BC rule
  - Users have initial motivation c
  - $r(t) = \frac{s(t)+c}{n(t)+c}$ fraction of "successful" edits
  - If attitude change: $r(t) \leftarrow r(t) + 1$
- Probability to abandon at time t:

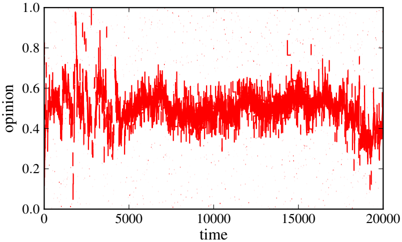
$$\lambda_d(t) = \frac{r(t)}{\tau_0} + \frac{1 - r(t)}{\tau_1} \tag{2}$$

- Time scales: $\tau_0 > \tau_1$ (usually by orders of magnitude)

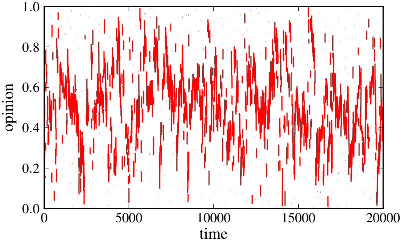
Norm adoption dynamics



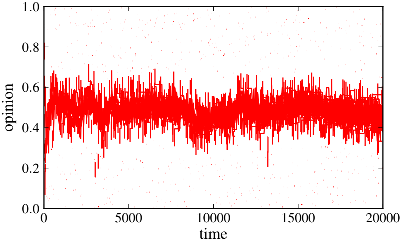
$\varepsilon = 0.1$



$\varepsilon = 0.15$



$\varepsilon = 0.125$



$\varepsilon = 0.175$

Analysis of the model

- **Question**: what are the most important parameters? (factor screening)
- **Solution**: sensitivity analysis

  parameters $\rightarrow$ model $\rightarrow$ average lifetime

- Decomposition of response variance:
  - Main interaction: fraction of variance accounted by one parameter only.
  - Total interaction: fraction of variance accounted by a parameters in conjunction with others.
- Computational issue: using surrogate model instead of simulator.

Analysis of the model

- **Question**: what are the most important parameters? (factor screening)
- **Solution**: sensitivity analysis

$$\text{parameters} \rightarrow \text{model} \rightarrow \text{average lifetime}$$

- Decomposition of response variance:
  - Main interaction: fraction of variance accounted by one parameter only.
  - Total interaction: fraction of variance accounted by a parameters in conjunction with others.
- Computational issue: using surrogate model instead of simulator.

Analysis of the model

- **Question**: what are the most important parameters? (factor screening)
- **Solution**: sensitivity analysis

$$\text{parameters} \rightarrow \text{model} \rightarrow \text{average lifetime}$$

- Decomposition of response variance:
  - Main interaction: fraction of variance accounted by one parameter only.
  - Total interaction: fraction of variance accounted by a parameters in conjunction with others.
- Computational issue: using surrogate model instead of simulator.

Analysis of the model

- **Question**: what are the most important parameters? (factor screening)
- **Solution**: sensitivity analysis

$$\text{parameters} \rightarrow \text{model} \rightarrow \text{average lifetime}$$

- Decomposition of response variance:
  - Main interaction: fraction of variance accounted by one parameter only.
  - Total interaction: fraction of variance accounted by a parameters in conjunction with others.
- Computational issue: using surrogate model instead of simulator.

Analysis of the model

- **Question**: what are the most important parameters? (factor screening)
- **Solution**: sensitivity analysis

$$\text{parameters} \rightarrow \text{model} \rightarrow \text{average lifetime}$$

- Decomposition of response variance:
  - Main interaction: fraction of variance accounted by one parameter only.
  - Total interaction: fraction of variance accounted by a parameters in conjunction with others.
- Computational issue: using surrogate model instead of simulator.

Analysis of the model

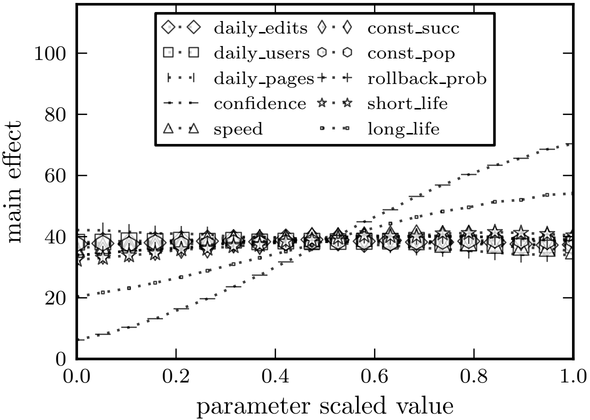
- **Question**: what are the most important parameters? (factor screening)
- **Solution**: sensitivity analysis

$$\text{parameters} \rightarrow \text{model} \rightarrow \text{average lifetime}$$

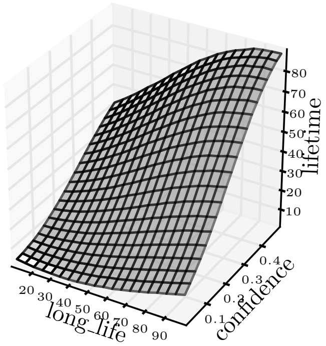
- Decomposition of response variance:
  - Main interaction: fraction of variance accounted by one parameter only.
  - Total interaction: fraction of variance accounted by a parameters in conjunction with others.
- Computational issue: using surrogate model instead of simulator.

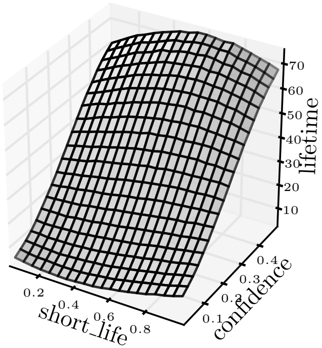
Main interaction plot

- Gaussian Process based on Latin hypercube design with 64 points
- Decomposition of variance computed with winding stairs method, 10000 samples



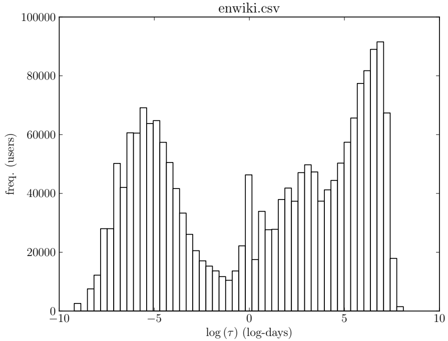
2-way interaction effects

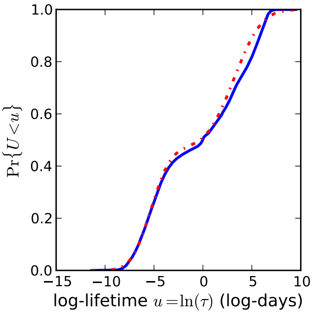




User activity lifespan

- User lifespan in blogs is exponential (Leskovec et *al.*, 2007)
- Bimodal distr. in blogs (Guo et *al.*, 2009)
- Wikipedia: mixture of lognormals (Ciampaglia and Vancheri 2010)

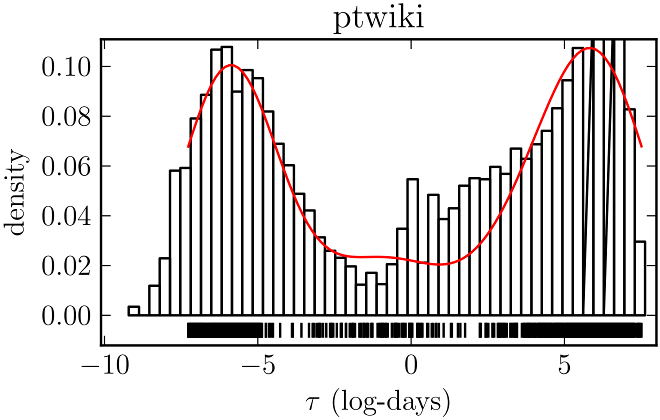




English Wikipedia, data from August 2010

Follow-up work

- Computational model fitting using indirect inference.
- Need to introduce Poissonian cascade model.



Conclusions

1. Confidence (i.e. tolerance to attitude change ε) most important parameter in explaining user participation
2. Other notable factors (e.g. initial motivation c) not important
3. New methodology for agent-based modeling, based on empirical data

References

- Kling 1999. What is social informatics and why does it matter? *D-Lib Mag.* 5 (1). `http://www.dlib.org/dlib/january99/kling/01kling.html`
- Benkler 2002. Coase's penguin: Linux and the nature of the firm. *Yale Law J.* 112 369–446.
- Deffuant et *al.* 2001. Mixing beliefs among interaction agents. *Adv. Comp. Sys.* 3, 87–98.
- Sherif and Hovland 1961. Assimilation and constrast effects in communication and attitude change. Yale University Press.
- Turner 1989. Rediscovering the social group: a self-categorization theory. Blackwell Publishers.
- Leskovec et *al.* 2007. Microscopic evolution of social networks. Proc. of KDD '08.
- Guo et *al.* 2009. Analyzing patterns of user content generation in online social networks. Proc. of KDD '09.
- Ciampaglia and Vancheri 2010. Empirical analysis of user participation in online communities: the case of Wikipedia. Proc. of ICWSM 2010.

Questions?