ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/319987236

Agent-Based Modelling for Online Community Designers

Working Paper · July 2017

CITATIONS
0

READS
24

1 author:

Simon à Campo
Technische Universiteit Eindhoven
6 PUBLICATIONS **4** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

Online community Heuristics for crowdsourcing platforms View project

All content following this page was uploaded by Simon à Campo on 22 September 2017.

The user has requested enhancement of the downloaded file.

Agent-Based Modelling for Online Community Designers

Simon à Campo
Industrial Design, Eindhoven
University of Technology
Eindhoven, the Netherlands
s.p.h.m.a.campo@student.tue.nl

ABSTRACT
Online community designers have to make difficult design decisions with unclear outcomes. Using an agent-based model by Ren and Kraut, the behavior can be predicted. This paper adds parameters such as newcomer's rate to the model to understand how to increase the activity level of members and make the underlying networks visible. Using the new interface, the model shows that small communities tend to have a larger percentage of contributors.

Author Keywords
Agent-based modelling, online community design

INTRODUCTION
Half of the world's population is connected to the Internet and 75% of them are using social media platforms actively [1]. Social sciences help to explain the behaviour of people on these platforms by comparing a small set of variables and thus formulate theories. The theories for successful online communities is broadly described to design for better platforms [2–5]. However, community design is a balancing game, since parameters influence each other and it can be unclear how. Agent-based modelling can be used to combine theories of social sciences and recreate the complex behaviour of online communities [6,7]. Online community designers can use these models to predict the outcomes of their design decisions.

To our knowledge, only one model has combined multiple social science theories into an agent-based model, created by Ren and Kraut [8]. According to the authors, "the model simulates people's motivation to participate in and contribute to an online community" and can be used broadly to any text-based, conversationally oriented online community. The model is limited to the examination of the effect of three design decisions that are difficult for community designers—topical breadth, message volume and discussion moderation.

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Times New Roman 8-point font. Please do not change or modify the size of this text box.
Each submission will be assigned a DOI string to be included here.

Although the goal of the model is "to serve as a decision-making tool for community designers", the interface is limited to only a few parameters and many of them are hard-coded. Therefore I wish to contribute with this paper in multiple ways: 1) improve the interface of the model by changing fixed parameters to dynamic parameters for community designers; 2) identify the limitations of the model that are not described by the authors; 3) see how the models holds up when introduced to a new social science theory; 4) find new effects by changing parameters that were fixed and hard-coded by the authors.

BACKGROUND
Ren and Kraut's model uses seven established and heavenly cited social theories in its core, of which an overview can been found in Figure 1 of their paper [8]. The theories include: expectancy theory [9], collective effort model [10], resource-based theory [11], information overload theory [12], group identity [13], interpersonal bonds [14,15], and intrinsic and extrinsic motivation [16]. The authors list the key assumptions of these theories and their conflicting predictions for their model.

The authors used existing researched relationships between factors and parameter values when possible. When evidence was not present between the different theories, the authors relied on their best judgement and empirical analysis of 100 Usenet groups to estimate these parameters. They calibrated the estimations to reproduce the power-law distribution of three statistics using the training data of 12 Usenet groups. They achieved a Pearson correlation between the empirical data of 25 new Usenet groups and the simulated data between 0.90 and 0.96, proving that the model matches real community behaviour well.

To test the model to a different social science rule, I use Nielsen's 90-9-1 for participation inequality in social media and online communities [17]. The rule explains that generally 90% of the members in a community are lurkers, which are members that read or observe, but don't contribute. Then, 9% of the members contribute from time to time, but other priorities dominate their time. Lastly, 1% of the members are very active and account to 90% of the contributions on a platform. Nielsen mentions that the inequality can be even more skewed to 95–5–0.1 on blogs. The model will be used in order to see if I can recreate these divisions and at the same time I can see which parameters could help to improve the ratio's to a fairer 80–16–4 that Nielsen proposes.

METHODOLOGY

Netlogo

The model is created using Netlogo, a cross-platform multi-agent modelling environment [18]. Within the simulation, agents (members) take actions in parallel during a simulated day. All active agents can chose to read or post a message, based on the expected benefits before moving on to the next day. Messages posted the previous day are distributed to all agents the next day and used to update the decision.

The model is created in version 3.1.5 whereas the most up to date version of the software is 6.0.1 at the time of writing. Attempts to convert the model to the newest version using the transition guide [19] failed. Fixing certain errors, wouldn't guarantee an accurate match with the 3.1.5 version. Therefore, I used an older version of the software, which is still available for download by Netlogo [1].

Interface

To improve the model, I first started fixing minor errors. For example, the buttons and sliders didn't work, since their value was depended on a fixed value in the code, instead of using user input value by sliders. Next I the removed code that wasn't necessary for a community designer to make the code shorter and thus more understandable. Mostly code to export data into .txt files were removed, since these were only meant for post data analysis by the researchers. These can be easily retrieved if one needs them, since Ren and Kraut's model is uploaded to the Netlogo community.

A legend was added to explain the different colours of the agents in the model, so one doesn't need to dive into the code to understand the colour coding. The parameters that the researchers used to experiment were replace from hard-code values, to sliders. These include: the topic breadth, the comment cost level and the newcomers-rate. In the comments of the code, the authors stated what parameters they used to test their model, thus these values were used to determine the sliders scales. All the set-up parameters that could only be changed at the beginning of the model, were put on the left and all the parameters that could be changed when the model is running are on the right.

Colour coding agents

A new colour coding for the agents in the model has been added, to show who are the newcomers (green), lurkers (blue), posters (red), superposters (yellow) and inactive members(grey). A newcomer is someone who just appeared for one day in the community. A lurker is a person who is reading messages, but didn't post anything in the last 2 days. A poster is a person that has contributed by creating a reply or thread post and stays a poster for 2 days after which he becomes a lurker again. A superposter, is a person who has contributed in the last 2 days and has a reputation higher than 0.6. The reputation number is built into the model by the authors and is based on the recognition from other members in the community. An inactive member, hasn't contributed or read any of the posts in the last 15 days and will disappear after 30 days of inactivity. A simple counter was added to represent the percentage of these agents occurring in the model. In the counter, newcomers and lurkers are combined, since newcomers haven't posted anything yet and thus can be defined as a lurker. Also, inactive members are not included into the lurker group, since a lurker does view posts and thus is an active member. Using three counters, we can monitor if Nielsen's 90-9-1 rule is achieved. The choice of the amount of days when a person changes status (for example from poster to lurker after 2 days) are explained in the results.

Crosslinks

Lastly, the links between members are made visible in the case that a poster and a member have a cross link, which is in the case when the responded to each other. The goal is to see if we can find new effects or behaviour of the model that hasn't been described yet.

RESULTS

Approaching the 90-9-1- rule

A baseline for the 90-9-1 rule has been created by setting all parameters to the default values of the authors: initial-members 190, initial-message 30, topical-breadth 5, comm-cost 1.5, newcomer-rate 5 and a community-type of interest. The run-time of the model was 200, providing a balance between steady values and fast run-time. The model was run 10 times and stops automatically after 200 simulated days. These settings already approached the 90-9-1 really close with a mean of 1,14 (σ = 0,49) for the superposters, a mean of 11,23 (σ = 1,67) for the posters and a mean of 87,63 (σ = 1.49) for the lurkers.

The simulation was run again, with a poster becoming a lurker after 4 days instead of 5. The simulation was run 5 times instead 10, since the previous simulations had a 97% – 99,5% match with the 10 times simulations and it could half the time it took to do simulations. The model was run which 2 more variations in the settings of days (d) and reputations (r) of which the results can be seen in Table 1.

Settings	% super poster	% poster	% lurker
d = 4 & r = 0.5	1,10	11,67	87,23
d = 3 & r = 0.5	1,35	9,95	88,70
d = 3 & r = 0.6	0,88	10,82	88,30
d = 2 & r = 0.6	0,85	7,44	91,71

Table 1. Table captions should be placed below the table. We recommend table lines be 1 point, 25% black. Minimize use of unnecessary table lines.

The table shows that the ideal settings are between 2 and 3 days. The model doesn't allow us to choose 2.5 days, since the model updates all the agents in parallel. The setting with the closest match to the rule is chosen, which is with 2 days and a reputation threshold of 0.6 since the percentage

[1] https://ccl.northwestern.edu/netlogo/download.shtml

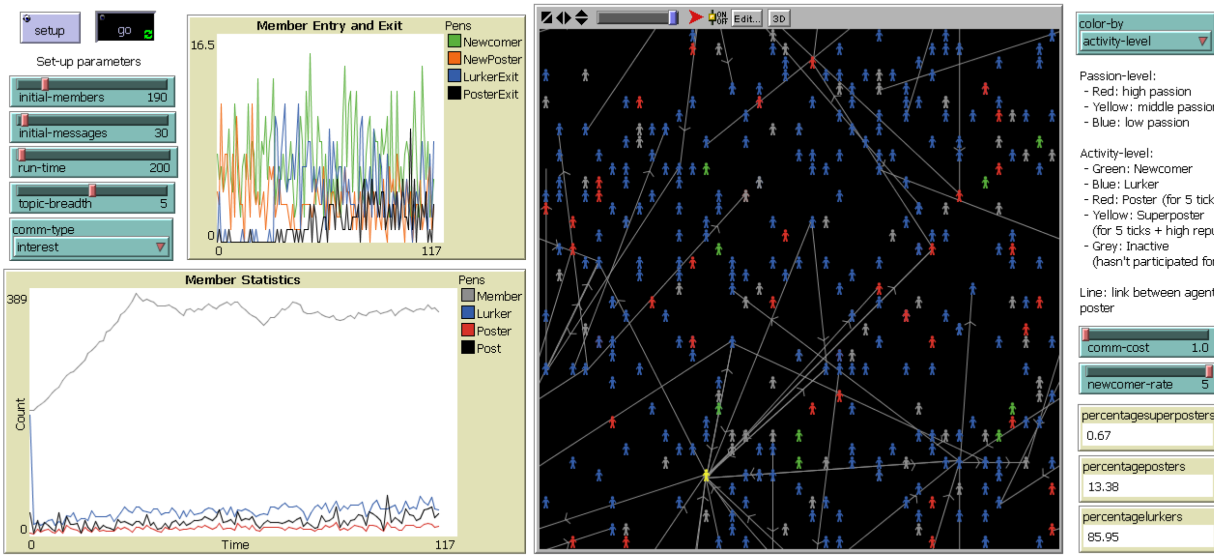

Figure 1. : The new interface of the model with added sliders to set-up a model on the right, and sliders and colour coding options that can be changed during the simulation on the right.

difference with the amount of lurkers is the lowest (1,1%). The settings will be used as a baseline to change the parameters and see the effects on the model.

Cost of contribution

Average of 5 simulations	% super poster	% poster	% lurker
comm-cost 1	0,64	16,426	82,932
comm-cost 1.5	0,85	7,436	91,712
comm-cost 2	0,454	5,338	94,208

Table 2. The effect on the percentage of superposters, posters and lurkers when changing the cost of contributing. Higher cost of contribution results in a lower amount of posts.

The cost of contribution to post a message has an influence on the percentage of posters and total amount of members. Increasing the cost doesn't have a correlation with the amount of superposters but a strong negative Pearson correlation of -0.94 (not significant with only 3 steps) and thus positive on the amount of lurkers. The total amount of members decreases when the cost is increased with a strong negative correlation of -1. The cost of contribution of 2 shows a resembles with the blog type of communities that have a 95-5-0,1 ratio.

Newcomer's rate

The newcomer's rate (Nr) has an effect on the total amount of members. When the newcomer rate is low, the contribution per member is higher in de case of Nr 1 & 2. To get a better nuance between the numbers, the simulations of Nr 3,4 & 5 have been done 10 times. A smaller community tend to have a higher percentage of posters and superposters.

New comer's rate	% super poster	% poster	% lurker	total members
Nr = 1	2,97	12,96	84,07	80,83
Nr = 2	1,07	10,99	87,94	145,00
Nr = 3	0,91	8,48	90,62	211,14
Nr = 4	0,69	8,90	90,41	271,50
Nr = 5	0,85	7,74	91,41	323,75

Table 3. The table shows the effect of the height of the newcomer's rate (Nr) on the percentage of superposters, posters, lurkers and total amount of members. The numbers are averages of 6x simulationsin the case of Nr1 &1 and 10x for Nr 3,4 & 5

The extremes

Lastly, the model is tested by setting the parameter to the extreme cases. These are not proven correlations, but just finding discovered when playing around with the model. It is encouraged to try these settings yourself to see the behaviour of the model.

For example, the desired 80-16-4 can be approached when comm-cost and newcomer rate is set to 1 and the topic-breadth to 9 and the community-type set to support. These condition trigger relationships the most which resulted in high amounts of superposters (4%) and posters (22%). At the same time, these kind of communities stayed rather small between around 100 and 150 members. Especially in the movielens type of community, high levels of posters are achieved being around 30% but at the same time the community is even smaller to around 50 members.

Crosslinks

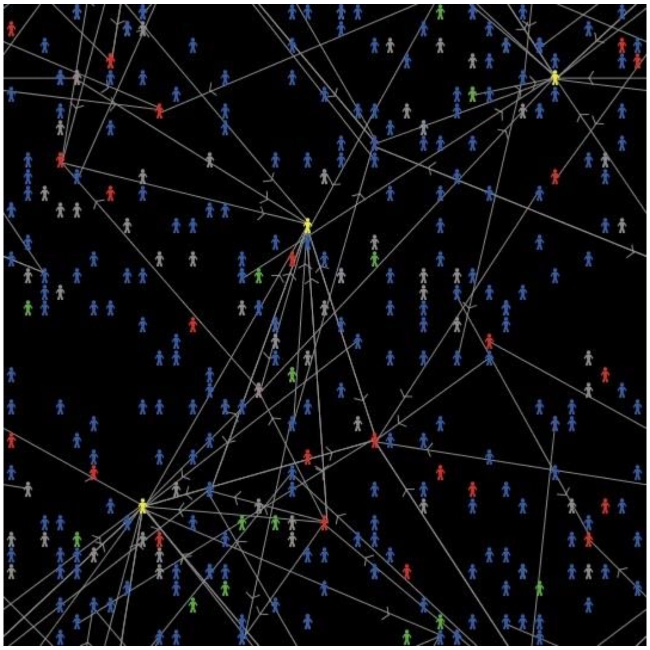

Figure 2. The crosslinks as shown by white lines connecting different agents. The superposters (yellow) are the agents with the most links.

The links between the agents reveal that superposters have many cross links and thus remain superposters for multiple ticks. The effects is the results of the benefit from bond-based attachment, which calculates the benefit from interpersonal bonds as a function of the number of other agents with whom the agent has developed a relationship through repeated interactions. A superposter with many links, remains a superposter for a long time since his network triggers him to keep posting. There seems to be no correlation between the height of the passion level and the crosslinks.

DISCUSSION

The approach of the finding the settings of the 90-9-1 is based on 10 runs of the simulation by changing two parameters. Ideally, all the parameters combinations are checked instead of fixing some based on the authors chose. The approximate to the rule could thus be further improved when more simulations are done, with a longer run-time for more stable numbers. In our research, the model was stopped after 200 simulated days. Ideally the percentage of agents is counted throughout the simulation, given a more stable number. The approximation of our paper is however close to the 90-9-1 rule, which seems that the model holds up to different social science theories with a 98,9% match with the amount of lurkers.

The model currently uses two ways to define lurkers. In our case, a lurker is someone who has read message, but hasn't contributed any of themselves in the last 15 days. Therefore a poster can become a lurker again. In the member statics graphs however, a lurker can become a poster when it has made his first post and will not be changed back to a lurker again.

In most cases, complex and adaptive systems use a small set of simple rules to create complex behaviour. The community model is however complex itself, including many theories and correlations. The result of changing certain parameters can thus not show as apparent results on the model as in more simple models that the Netlogo libraries provides.

Note for community designers

To understand the outcome of the model, there are some elements that the authors do not mention in their paper nor the interface. Community designer that want to interpret their own generated data, could benefit from the elements described in these paragraphs.

Starting with the Member Statics graph; it is important to know that after 30 ticks a change of the trend line can be seen, since at this point the members that have been inactive for the entire time will be hidden and disappear from the counters. A second important point is the 60 ticks mark, which is the amount of days the model is delayed before design interventions take effects, due to lack of data.

The model cannot be used to input values of a real community and predict how it will act. This is because the model doesn't take into account the age of the initial set amount of message. Also, the initial size of the members don't include any links between the members yet.

Member can respond to old message from turtles that have died and thus the turtle could be triggered to respond to that message again. The model runs starts running slow over time since it has an ever increasing number of turtles. Also, it make it difficult to select the right agent when wanting to see its properties, given the long list of links and agents that are already hidden. Having hundreds of agents in only a small field means that the model will stack these agents on top of each other, only revealing the colour of the latest added agent. Sometimes superposter are not visible because a newer agent is on top of the superposter. The percentage counters can thus deviate from the expected amount of agents.

Opportunities

A counter could be added for the amount of links created between the agents. For example, it seemed as if an increased number of crosslinks appeared when lowering the cost of contribution. These counter could also count the amount of links in general, showing how many of them are active, inactive and crosslinks.

The model doesn't take into account that certain messages could be spam messages that have a negative effect on the community. Lastly, the creation of subgroup is not possible in the model, which could positively influence the relationships that are built among the members as can be seen in the simulations of small communities (< 100).

CONCLUSION

Our paper proposes addition to Ren and Kraut's model of online communities, to be usable for online community designers. The model holds up to newly introduced social science theory 90-9-1 rule and thus can be seen as a robust model. In this paper we found which parameters decrease the amount of lurkers and increase the amount of posters and thus inform community designer about the design decisions they have to make. Only a limited amount of parameters are explored and the model shows a lot of potential for further research and understanding of online communities.

ACKNOWLEDGEMENTS

The paper is written for the master course Complex and Adaptive Systems at the faculty Industrial Design, Eindhoven university of Technology. I would like to thank: F.L.M. Delbressine, L.M.G. Feijs and G.W.M. Rauterberg for their inspirational lectures and feedback.

REFERENCES

[1] We Are Social, Hootsuite, DIGITAL IN 2017: GLOBAL OVERVIEW, We areSocial. (2017) 107. https://wearesocial.com/uk/blog/2017/01/digital-in-2017-global-overview (accessed June 28, 2017).

[2] J. Preece, C. Abras, D. Maloney-Krichmar, Designing and evaluating online communities: research speaks to emerging practice, Int. J. Web Based Communities. 1 (2004) 2–18. doi:10.1504/IJWBC.2004.004795.

[3] D. Gurzick, W.G. Lutters, Towards a design theory for online communities, in: Proc. 4th Int. Conf. Des. Sci. Res. Inf. Syst. Technol. DESRIST 09, ACM Press, New York, New York, USA, 2009: p. 1. doi:10.1145/1555619.1555634.

[4] R.E. Kraut, P. Resnick, S. Kiesler, Building successful online communities : evidence-based social design, MIT Press, 2011.

[5] A. Kim, Jo, Community building on the Web, Peachpit Press, 2000.

[6] J. Miller, S.E. Page, Complex adaptive systems : an introduction to computational models of social life, Princeton University Press, 2007.

[7] J.M. Epstein, Agent-based computational models and generative social science, Complexity. 4 (1999) 41–60. doi:10.1002/(SICI)1099-0526(199905/06)4:5<41::AID-CPLX9>3.3.CO;2-6.

[8] Y. Ren, R.E. Kraut, Agent-Based Modeling to Inform Online Community Design: Impact of Topical Breadth, Message Volume, and Discussion Moderation on Member Commitment and Contribution, Human–Computer Interact. 29 (2014) 351–389. doi:10.1080/07370024.2013.828565.

[9] V. Vroom, L. Porter, E. Lawler, Expectancy Theories, Organ. Behav. 1 Essent. Theor. Motiv. Leadersh. (2005) 94–113. https://books.google.nl/books?hl=nl&lr=&id=kUO5NWwaySYC&oi=fnd&pg=PA94&dq=expectancy+theories+vroom&ots=UFcGBiMlJS&sig=byYQtk-pUEJe8OPMqsDtKaMTEuA#v=onepage&q=expectancy theories vroom&f=false (accessed June 28, 2017).

[10] S.J. Karau, K.D. Williams, Social loafing: A meta-analytic review and theoretical integration., J. Pers. Soc. Psychol. 65 (1993) 681–706. doi:10.1037/0022-3514.65.4.681.

[11] B.S. Butler, Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures, Inf. Syst. Res. 12 (2001) 346–362. doi:10.1287/isre.12.4.346.9703.

[12] Q. Jones, G. Ravid, S. Rafaeli, Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration, Inf. Syst. Res. 15 (2004) 194–211. doi:10.1287/isre.1040.0023.

[13] M.A. Hogg, D.J. Terry, SOCIAL IDENTITY AND SELF-CATEGORIZATION PROCESSES IN ORGANIZATIONAL CONTEXTS., Acad. Manag. Rev. 25 (2000) 121–140. doi:10.5465/AMR.2000.2791606.

[14] N.L. Collins, L.C. Miller, Self-disclosure and liking: A meta-analytic review., Psychol. Bull. 116 (1994) 457–475. doi:10.1037/0033-2909.116.3.457.

[15] L. Festinger, Informal social communication., Psychol. Rev. 57 (1950) 271–282. doi:10.1037/h0056932.

[16] R.M. Ryan, E.L. Deci, Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions, Contemp. Educ. Psychol. 25 (2000) 54–67. doi:10.1006/ceps.1999.1020.

[17] J. Nielsen, Participation Inequality: The 90-9-1 Rule for Social Features, (2006). https://www.nngroup.com/articles/participation-inequality/ (accessed November 17, 2016).

[18] U. Wilensky, NetLogo. http://ccl.northwestern.edu/netlogo/., Cent. Connect. Learn. Comput. Model. Northwest. Univ. Evanst. 2009 (1999) Evanston, IL. http://ccl.northwestern.edu/netlogo/.

[19] Uri Wilensky, NetLogo 6.0.1 User Manual: Transition Guide, (n.d.). https://ccl.northwestern.edu/netlogo/docs/transition.html (accessed June 28, 2017).

APPENDIX

Link to download the Netlogo file:

https://www.dropbox.com/s/4r4gqi8c25vq0cd/RenKraut-SimulateOnlineCommunity-addition-by-Simon.nlogo?dl=0

Link to the original used Netlogo file by Ren and Kraut:

http://ccl.northwestern.edu/netlogo/models/community/RenKraut-SimulateOnlineCommunity

Link to a report comparing the new and old code:

https://www.diffnow.com/?report=ag79n

View publication stats