



HAL
open science

Toward an Agent-Based Model of Community of Practice: Demonstration

Amal Chaoui, Sébastien Delarre, Fabien Eloire, Maxime Morge, Antoine
Nongaillard

► To cite this version:

Amal Chaoui, Sébastien Delarre, Fabien Eloire, Maxime Morge, Antoine Nongaillard. Toward an Agent-Based Model of Community of Practice: Demonstration. Frank Dignum; Philippe Mathieu; Juan Manuel Corchado; Fernando De La Prieta. 20th International Conference on Practical Applications of Agents and Multi-Agent Systems, Jul 2022, L'Aquila, Italy. Springer International Publishing, Advances in Practical Applications of Agents, Multi-Agent Systems, and Complex Systems Simulation, 13616, pp.467-472, 2022, The PAAMS Collection. 10.1007/978-3-031-18192-4_39 . hal-03812873

HAL Id: hal-03812873

<https://hal.science/hal-03812873v1>

Submitted on 14 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Toward an Agent-Based Model of Community of Practice: Demonstration^{*}

Amal Chaoui¹, Sébastien Delarre², Fabien Eloire², Maxime Morge ¹[0000-0003-2139-7150], and Antoine Nongaillard²[0000-0001-8551-0509]

¹ Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
`amal.chaoui@centrale.centralelille.fr`

`{antoine.nongaillard,maxime.morge}@univ-lille.fr`
² Univ. Lille, CNRS, UMR 8019 Clersé, F-59000 Lille, France
`{sebastien.delarre,fabien.eloire}@univ-lille.fr`

Abstract. Communities of Practice services enable users that share same interests and exchange knowledge in an ongoing learning process. The Stack Overflow platform supports such communities where the majority of the content are produced by a small number of highly active users. We propose an agent-based model of a Community of Practice exhibiting a social dynamics which is compliant with this insight.

Keywords: Agent-based modeling, Social network analysis, Complex system

1 Introduction

A Community of Practice (CoP) is “a group of people who share a concern, set of problems, or passion for a given topic, and who deepen their knowledge and expertise in that area by interacting on an ongoing basis” [8]. The Internet has contributed to the creation of Distributed Community of Practice (DCoP) whose members are geographically distributed and use technological means to remotely interact with each other. For instance, Stack Overflow³ is not only a simple Question and Answer (Q&A) website about programming, but also a collaborative work platform that structures DCoPs.

DCoPs are complex systems. The interleaving of microscopic elementary behaviors leads to complex mesoscopic dynamics (at the community level) and macroscopic ones (at the platform level). In particular, the identification of authoritative actors cannot be tackled by the classical algorithms **PageRank** and **HITS** for analyzing such networks whose topologies are heterogeneous and distinguish themselves from the web [2]. The individual-centered approach is more suitable since it allows to conceptualize and simulate such complex systems as self-organized systems of cooperative agents with limited perceptions and rationality, which are structured by the effect of interactions [4].

^{*} We thank Koboyoda Arthur Assima and Audrey Dewaele for their contribution on the data analysis and Jarod Vanderlynden for his work about the state-of-the-art.

³ <https://stackoverflow.com>

While multi-agents technologies supporting DCoPs are numerous and diverse [7], the agent-based models of a DCoP are much rarer. Zhang et al. adopt an individual-centered approach to generate the social network [12] and the `CommunityNetSimulator` [11] analyze it. Contrary to [1], links are not created solely on the basis of the preferential attachment principle whereby new nodes connect more easily to those having already a large number of connections, but the model considers links as the outcome of interactions (questions/answers) resulting from the individual behaviors, which are in turn guided by their expertise levels. However, the interaction model introduced by [12] does not reproduce a commonly observed insight: a small part of the members is responsible for a vast part of the community activity.

We propose here an agent-based model of a Community of Practice that exhibits a social dynamic where the activity is unfairly distributed among the members. The significance of our work consists in reproducing this phenomenon according to the principle of parsimony. By reducing the number of parameters and restricting the mechanisms to the simplest and most general ones, we propose an explainable model of this social phenomenon based on a clear understanding of the underlying local solidarity processes built upon a large real-world dataset. Our simulator `ABM4DCoP` (Agent-Based Modeling for Distributed Community of Practice) [6] is built upon `NetLogo`. Our experiments show that, our model, contrary to [11], reproduces the unfair distribution of posts observed on Stack Overflow.

2 Agent-Based Model

In order to simulate a DCoP, we model a constant population of N agents. Each simulation step t corresponds to a biweekly period. During such a period, each agent can publish a number of posts (questions or answers), bounded by a biweekly maximum number of questions (respectively answers), $maxQ$ (respectively $maxA$). The idiosyncrasy of an agent determines its actions. Concretely, the behavior of each agent a_i is specified by two individual parameters:

- $\overline{nQ(a_i)}$, i.e. the mean number of questions asked per period;
- $\overline{nA(a_i)}$, i.e. the mean number of answers replied per period.

The number of questions (respectively answers) posted at time t , denoted $nQ(a_i, t)$ (respectively $nA(a_i, t)$), is drawn from a normal distribution of mean $\overline{nQ(a_i)}$ (respectively $\overline{nA(a_i)}$).

In order to populate the simulation, we consider that the distribution of the biweekly number of posts of the agent is of power law type:

$$\overline{nA(a_i)} = maxA \times rank^{\overline{nA}(a_i)^{k_A}} \text{ with } k_A < 0 \quad (1)$$

where the degree of the law k_A is constant. As explained in [10], users motivated by gamification incentives mostly post answers. This is the reason why we generate $nQ(a_i)$ as $nA(a_i)$, and in such a way that the more answers an agent

gives, the fewer questions it asks. The simulation of a period t is divided into two phases (cf. Figure 1). During the question phase, each agent a_i asks $nQ(a_i, t)$ questions. During the responding phase, each agent reacts to $nA(a_i, t)$ questions asked by the other agents⁴.

<pre> Data: a_i, t: agent/timestep H: history Result: H': history $nQ(a_i, t) = \mathcal{N}(\mu = \overline{nQ(a_i)}, \sigma^2 = 0.02)$; if $randomU(0, 1) \leq \{nQ(a_i, t)\}$ then $nQ(a_i, t) += 1$; for $j \in [1; nQ(a_i, t)]$ do $H \cup = question(author = a_i)$; return H </pre>	<pre> Data: a_i, t: agent/timestep H: history Result: H': History $nA(a_i, t) = \mathcal{N}(\mu = \overline{nA(a_i)}, \sigma^2 = 0.02)$; if $randomU(0, 1) \leq \{nA(a_i, t)\}$ then $nA(a_i, t) += 1$; $Q = \{question \in H \mid author(question) \neq a_i\}$ for $j \in [1; nA(a_i, t)]$ do $q = random(Q)$; $H \cup = answer(author = a_i, question = q)$; return H </pre>
---	--

Fig. 1: Individual asker behavior (at left) and helper behavior (at right)

3 Simulator

Our ABM4DCoP simulator [6] is based on NetLogo [9] which is a multi-agents programming language and an integrated development environment for the design of agent-based simulations. Even if NetLogo is user-friendly, it allows the design of high-level models with complex behaviors.

As depicted in Figure 2, the interface of our simulator consists of three types of components to:

- configure the simulation parameters;
- visualize agents, interactions and links;
- analyze the evolution of the metrics which are computed and displayed in real time during the simulation.

The agents are arranged in a circle in a clockwise direction in ascending order of $\overline{nA(a_i)}$. The graphical representation of the community in Figure 3 allows us to observe the spatial distribution of the productivity of the agents (in purple) and of their links (in grey). An arc represents the help provided by the source

⁴ The variances are small enough to maintain the shape of the power law type distribution for the mean number of posts published per period.

agent to the destination one. The outdegree of a node is the number of peers helped and the indegree is the number of advisors. During the simulation, the posts accumulate on the radial segment of its author: questions are in green and answers are in red.

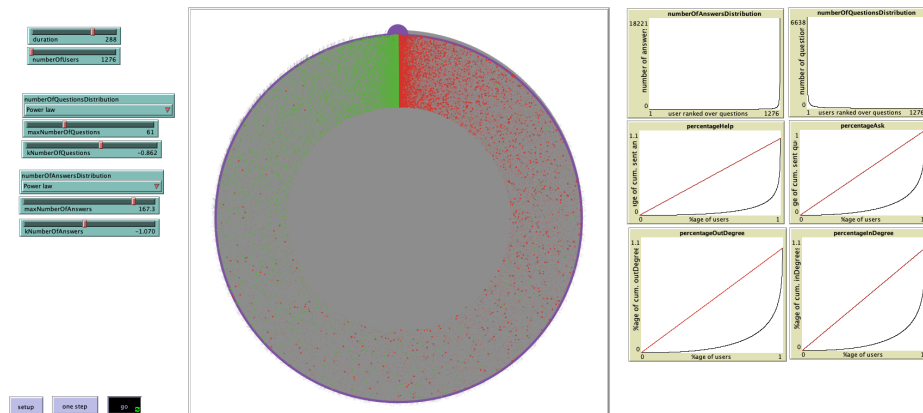


Fig. 2: Configuration (at left), visualization of agents arranged in a circle (at center) and reporting (at right) of the simulation.

4 Discussion

In order to explore our Stack Overflow study area, we have extracted, transformed and loaded the raw data produced from 2008 to 2021 (~ 50 GB) into a data warehouse [5]. Our multidimensional conceptual model allows the analysis of key performance indicators such as the number of posts, the debatability of questions or the response time. The multidimensional logic model is supplied by 6 face-to-face communities: **Hive/Pig**, **Hadoop/Spark** and **R/Pandas**. The latter two communities represent about 1,300,000 posts published by nearly 250,000 members. We have discarded 2,000 posts (0.15 %) involving 4,000 users who deleted their accounts and for whom we did not have attribute data.

Figure 4 represents the cumulative help of users from lowest to highest mean numbers of questions asked per period ($nA(a_i)$). We have reimplemented **CommunityNetSimulator** [3] with **NetLogo**. According to this model, 20% of the most active members are responsible for up to 66% of the answers. However, we observe that 20% of the most active members of the **R** community are the authors of 94% of the answers on this topic. According to our model, the unfairness of the distribution is similar if the maximum biweekly number of answers is $maxA = 167.3$ (as in the **R** community) and $k_A = -1.2$. In online communities, very few users post the majority of content. This mesoscopic phenomenon is shared by all the DCoPs that we we have studied on Stack Overflow. The

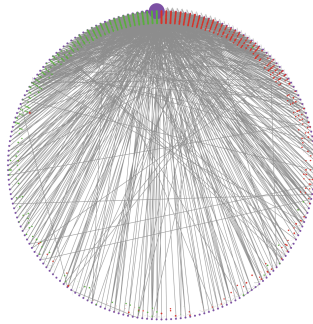


Fig. 3: Quasi-bipartite community network.

pseudo-random and sequential scheduler of our engine, which allows talkative agents to speak more than once per period, purposely reproduces the inequality of the event-driven activity.

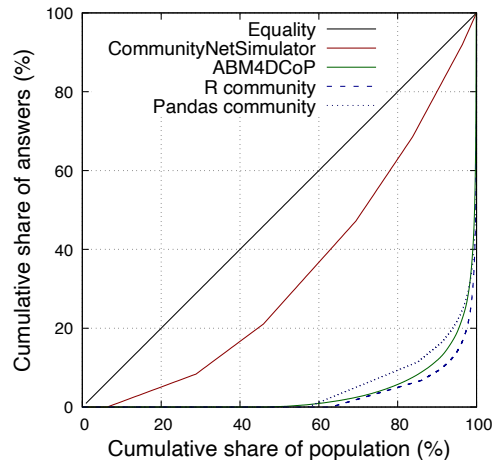


Fig. 4: Lorenz curves that show the number of answers observed (dotted line), simulated by CommunityNetSimulator (in red) and ABM4DCoP (in green) with the share of the population.

Our future work will focus on the statistical modeling of the distribution of idiosyncratic characteristics in order to calibrate the population of our simulations based on the large dataset we have collected. For now, the population of agents is homogeneous and their expertise is measured by their activities *ex post*. However, we wish to distinguish the most active individuals whose answers are not necessarily the most relevant ones (sparrows) from the domain experts who are not necessarily the most active (owls) [10]. In the longer term, we wish

study and confront other mesoscopic phenomena such as: (a) the popularity - measured by the number of views of threads, (b) controversy - measured by the debatability of questions, and (c) the efficiency with the response time.

References

1. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. *Science* **286**(5439), 509–512 (1999)
2. Bouguessa, M., Dumoulin, B., Wang, S.: Identifying Authoritative Actors in Question-Answering Forums: The Case of Yahoo! Answers. In: *Proc. of the ACM Conference SIGKDD*. p. 866–874 (2008)
3. Chaoui, A., Vanderlynden, J., Morge, M.: CommunityNetSimulator: Agent-Based Model for Simulating a Community. <https://gitlab.univ-lille.fr/mocicos/communitynetsimulator>, online; accessed 2022-05-03
4. Gilbert, N. (ed.): *Simulating societies: the computer simulation of social phenomena*. UCL Press (1994)
5. Morge, M.: SoDyOnStack: An Analysis of Social Dynamics on Stack Overflow. <https://gitlab.univ-lille.fr/mocicos/sodyonstack>, online; accessed 2022-06-01
6. Morge, M., Amal, C.: ABM4DCOP: Agent-Based Modeling for Distributed Community Of Practice. <https://gitlab.univ-lille.fr/mocicos/abm4dcop>, online; accessed 2022-06-01
7. Sato, G.Y., de Azevedo, H.J.S., Barthès, J.P.A.: Agent and multi-agent applications to support distributed communities of practice: a short review. *Autonomous Agents and Multi-Agent Systems* **25**(1), 87–129 (2012)
8. Wenger, E., McDermott, R.A., Snyder, W.: *Cultivating communities of practice: A guide to managing knowledge*. Harvard business press (2002)
9. Wilensky, Uri: Netlogo. <http://ccl.northwestern.edu/netlogo>, online; accessed 2022-06-01
10. Yang, J., Tao, K., Bozzon, A., Houben, G.J.: Sparrows and owls: Characterisation of expert behaviour in stackoverflow. In: *Proc. of Int. Conference on User Modeling, Adaptation, and Personalization*. pp. 266–277. Springer (2014)
11. Zhang, J., Ackerman, M.S., Adamic, L.: Community net simulator: Using simulations to study online community networks. In: *Proc. of Int. Conference on Communities and Technologies*, pp. 295–321. Springer (2007)
12. Zhang, J., Ackerman, M.S., Adamic, L.: Expertise networks in online communities: structure and algorithms. In: *Proc. of the Int. Conference on World Wide Web*. pp. 221–230 (2007)