Modèle explicatif de la sécession des experts dans les communautés de pratiques

Sébastien Delarre, Fabien Eloire, Antoine Nongaillard, Maxime Morge

Modèle explicatif de la sécession des experts dans les communautés de pratiques

Sébastien Delarre[a] sebastien.delarre@univ-lille.fr
Fabien Eloire[a] fabien.eloire@univ-lille.fr
Antoine Nongaillard[b] antoine.nongaillard@univ-lille.fr
Maxime Morge[b] maxime.morge@univ-lille.fr

[a]Univ. Lille, CNRS, UMR 8019 Clersé, F-59000 Lille, France

[b]Univ. Lille, CNRS, Centrale Lille, UMR 9189, CRIStAL, F-59000 Lille, France

Résumé

Les sites de questions-réponses permettent aux utilisateurs, qui partagent les mêmes intérêts, d'échanger des connaissances dans le cadre d'un processus d'apprentissage continu. En particulier, la plateforme `Stack Overflow` *soutient des communautés de pratique où : (a) la majorité du contenu est produit par un petit nombre de membres très actifs ; (b) les réponses aux questions les plus récurrentes sont rapidement capitalisées. Nous proposons un modèle centré individus d'une communauté de pratique présentant une dynamique sociale conforme à ces faits stylisés et qui explique le phénomène de sécession des experts.*

Mots-clés : *Simulation multi-agents, Analyse des réseaux sociaux, Système complexe, Explicabilité*

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 : (a) the majority of the content are produced by a small number of highly active users ; (b) the answers to the most common questions are quickly capitalized on. We propose an agent-based model of a Community of Practice exhibiting a social dynamics that is compliant with these insights and which explains the phenomenon of secession of experts.*

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

1 Introduction

Une communauté de pratique - en anglais *Community of Practice* (CoP) - est « un groupe de personnes qui partagent une préoccupation, un ensemble de problèmes ou une passion pour un sujet donné, et qui approfondissent leurs connaissances et leur expertise dans ce domaine en interagissant de manière continue » [17]. L'Internet a contribué à la création de communautés de pratique distribuées - en anglais *Distributed Community of Practice* (DCoP) - dont les membres sont géographiquement répartis et utilisent des moyens technologiques pour interagir à distance les uns avec les autres. Par exemple, `Stack Overflow`[1] n'est pas seulement un simple site de questions-réponses sur la programmation, mais aussi une plateforme de travail collaboratif qui structure des DCoPs.

La formation et la structuration des DCoPs résultent des phénomènes sociaux complexes. Nous nous intéressons ici tout particulièrement au phénomène de sécession des experts. Au-delà du fait que chaque année un tiers des utilisateurs des `Stack Overflow` deviennent inactifs [12], on observe que, pour la plupart des communautés sur cette plateforme, 80 des 100 réponses les mieux notées sont données au cours de la phase de formation de la communauté (cf. figure 1). Les phénomènes sociaux complexes qui se jouent sur `Stack Overflow` sont le résultat de l'imbrication de comportements élémentaires microscopiques qui conduit à des dynamiques complexes mésoscopiques (au niveau de la communauté) et macroscopiques (au niveau de la plateforme). Nous soutenons que l'approche centrée individus est appropriée pour l'étude des DCoPs car elle permet de conceptualiser, de calibrer et de simuler ces phénomènes complexes à l'aide de systèmes auto-organisés d'agents coopératifs aux perceptions et à la rationalité limitées, qui sont structurés par l'effet des interactions. Le contrôle de tous les paramètres de cette communauté artificielle permet d'observer *ceteris paribus* les changements induits [7]. Le modèle visé doit être explicatif en

1. `https://stackoverflow.com`

proposant une restitution visuelle de ses propriétés et prescriptif en permettant d'introspecter les choix de conception de la plateforme pour favoriser l'innovation et l'intelligence collective.

Nous proposons ici un modèle centré individus d'une DCoP [2] qui présente une dynamique sociale où : (a) l'activité est inégalement distribuée parmi les membres ; (b) les réponses aux questions les plus récurrentes sont rapidement capitalisées. Nos résultats expérimentaux montrent que notre modèle est non seulement capable de reproduire ces faits stylisés mais également que la capitalisation explique le phénomène de sécession des experts qui se traduit par un effondrement qualitatif des réponses. Le principal verrou scientifique auquel nous nous sommes confrontés réside dans la calibration à partir de données ouvertes et massives pour la validation du modèle multi-agents de par son caractère multi-niveaux, l'aspect individuel des comportements et les boucles de rétro-action [1, 6, 9, 15]. C'est la raison pour laquelle nous nous sommes conformés au principe de parcimonie. En réduisant le nombre de paramètres et en limitant les mécanismes aux plus simples et aux plus généraux, nous proposons un modèle explicatif fondé sur l'identification des processus de solidarité locale et des comportements sociaux individuels sous-jacents.

Après un aperçu des travaux connexes dans la section 2, nous formalisons notre modèle centré individus, en particulier les comportements de demandeur et d'aidant (cf. section 3). La section 4 décrit notre implémentation de ce modèle. La section 5 présente les résultats obtenus par le modèle à partir de sa calibration sur des données réelles et massives. Enfin, la section 6 résume nos contributions, identifie les limites de notre modèle et présente nos perspectives.

2 Travaux connexes

L'étude et la modélisation des DCoPs n'est pas novateur. Ces dernières ont fait l'objet d'analyses statistiques, d'analyses de réseaux sociaux ou d'analyses temporelles mais de peu de modélisation à base d'agents.

Moutidis et Williams proposent la première grande analyse de l'évolution des communautés sur `Stack Overflow` [12]. Cette étude statistique de l'apparition, de la perpétuation et de la transformation des DCoPs s'appuient sur des algorithmes de détection de communautés d'utilisateurs en fonction de la similitude des étiquettes sur leurs contributions. À l'inverse, dans notre modèle, le réseau social réifie les interactions (questions/réponses) résultantes du comportement des individus. À l'aide de métriques macroscopiques comme le nombre de contributions ou le score total des contributions par période, cette étude observe que chaque année 34 % des utilisateurs deviennent inactifs sans pour autant expliquer ce phénomène.

L'analyse des réseaux sociaux pour les DCoPs vise à automatiser l'identification des figures d'autorité dans les communautés. Conçu pour le web, les algorithmes classiques tels que `PageRank` et `HITS` s'avèrent inadaptés aux réseaux sociaux, en particulier les DCoPs dont les topologies sont hétérogènes et se distinguent du web [3]. Bouguessa et *al.* observent que le nombre de contributions identifiées comme la meilleure réponse à une question est la métrique mésoscopique la plus appropriée pour évaluer le niveau d'autorité de l'aidant. L'analyse des réseaux sociaux permet également l'identification de stéréotypes parmi les utilisateurs. Yang et *al.* observent qu'une petite proportion d'utilisateurs est responsable de la majorité des contributions [19]. Ils mesurent également la qualité des réponses à l'aide d'une métrique microscopique, i.e. la contribution experte moyenne. Cela leur permet de distinguer deux stéréotypes de comportement individuel parmi les 10 % des aidants les plus actifs :

- les « moineaux » qui sont majoritaires, très réactifs vis-à-vis des questions simples mais avec des réponses qui ne sont pas nécessairement les plus pertinentes ;
- les « hiboux » qui sont minoritaires, pas nécessairement les plus réactifs mais dont les réponses sont les plus pertinentes et dont les questions sont difficiles.

Yang et *al.* interprète la disparition précoce des « hiboux » et l'arrivée continuelle de nouveaux « moineaux » comme le résultats des incitations à la gamification de plateforme. Toutefois, les résultats d'une récente enquête [8] suggèrent un effet limité des badges et des privilèges de réputation sur la motivation des utilisateurs. Notre modèle montre que ces phénomènes sont le résultat de la capitalisation des réponses aux questions les plus récurrentes.

L'analyse temporelle des DCoPs étudie l'évolution des membres, des contributions et des interactions. Pal et *al.* propose un modèle probabiliste de sélection des questions auxquelles répondre [13, 14]. Afin de maximiser la probabi-

2. Cet article est une version étendue en français de [4].

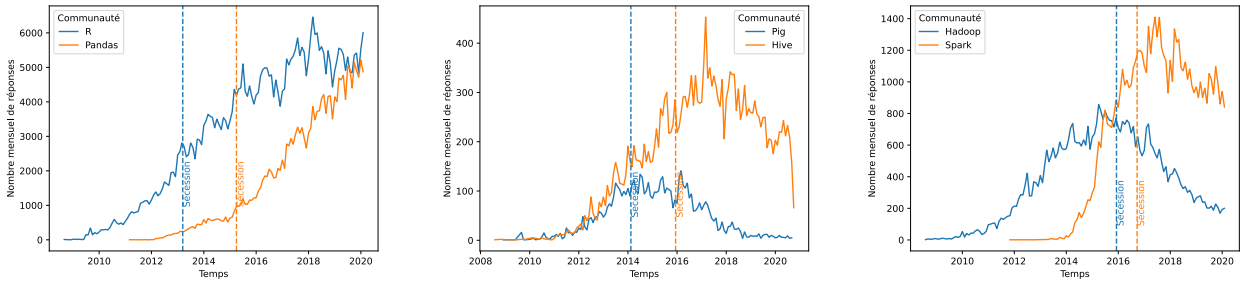


FIGURE 1 – Pour 5 des 6 communautés/thématiques représentées ici en face à face (`R/Pandas`, `Pig/Hive`, `Hadoop/Spark`), on observe que la date de sécession des experts, i.e. la date à laquelle 80 des 100 réponses les mieux notées sont données, intervient dans la phase de formation, i.e. quand le nombre mensuel de réponse croit

lité que leur réponse soit bénéfique pour la communauté, les aidants, qui sont solidaires, considèrent la valeur d'une question comme la valeur globale des réponses à cette question et la valeur d'une réponse comme le bénéfice qu'elle apporte par rapport aux réponses antérieures. C'est sur ce principe que s'appuie le mécanisme de sélection de questions des aidants dans notre modèle. Il est important de noter que cette étude conclut que l'explication du phénomène de sécession des experts est une question ouverte à laquelle notre modèle de capitalisation répond.

Alors que les technologies multi-agents soutenant les DCoPs sont nombreuses et variées [16], les modèles à base d'agents pour les DCoPs sont beaucoup plus rares. Zhang et *al.* adoptent une approche centrée individus pour générer le réseau social [21] et le `CommunityNetSimulator` [20] pour l'analyser. Contrairement à [2], les liens ne sont pas créés uniquement sur la base du principe d'attachement préférentiel selon lequel les nouveaux noeuds se connectent plus facilement à ceux qui ont déjà un grand nombre de connexions comme c'est le cas dans la plupart des réseaux sociaux, mais le modèle considère les liens comme résultants d'interactions (questions/réponses) entre des comportements individuels qui sont à leur tour guidés par les niveaux d'expertise. Afin de peupler la simulation, Zhang et *al.* considèrent, sans calibration, que la distribution de l'expertise est de type loi de puissance. Dans ce modèle, seuls les novices posent des questions et la probabilité de répondre à une question croit de manière exponentielle avec la différence de niveau d'expertise entre le demandeur et l'aidant. Contrairement à ce dernier, notre modèle numérique est calibré à partir des données réelles massives. À la différence de notre modèle, celui introduit par [21] ne permet pas de reproduire pas le phénomène d'inéquité de l'activité d'une communauté parmi les membres et il n'explique pas le phénomène de sécession des experts.

3 Modèle

Afin de simuler une DCoP, nous modélisons une population constante de N agents. Chaque pas de simulation t correspond à une période bihebdomadaire. Au cours d'une période, chaque agent peut publier un certain nombre de contributions, des questions et/ou des réponses.

Chaque contribution c a un identifiant $id(c)$ – i.e. une clé artificielle distincte pour les questions et les réponses, un horodatage $time(c)$ – i.e. le pas de simulation au cours duquel elle est émise, un auteur $author(c)$ et un score $score(c)$ – i.e. une évaluation numérique de sa fréquence (pour une question) ou de sa pertinence (pour une réponse). De plus, chaque réponse aborde une question, $reply(a)$.

L'idiosyncrasie d'un agent détermine ses actions. Concrètement, le comportement de chaque agent a_i est spécifié par deux paramètres individuels :

- $\overline{nQ(a_i)}$, i.e. le nombre moyen de questions par période ;
- $\overline{nA(a_i)}$, i.e. le nombre moyen de réponses par période.

Le nombre de questions publiées à l'instant t, noté $nQ(a_i, t)$, est issu d'une distribution normale de moyenne $\overline{nQ(a_i)}$ et il est borné par un nombre maximum de questions par pas de simulation, $maxQ$. Le nombre de réponses données à l'instant t, noté $nA(a_i, t)$, est issu d'une distribution normale de moyenne $\overline{nA(a_i)}$ et il est

borné par un nombre maximum de réponses par pas de simulation, $maxA$.

Afin de peupler la simulation, nous considérons, comme le laisse supposer nos données de calibration (cf. section 5.1), que la distribution du nombre bihebdomadaire de réponses d'un agent est de type loi de puissance :

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

où $rang^{\overline{nA}}(a_i)$ dénote le rang de l'agent a_i par nombre moyen de réponses bihebdomadaires décroissant et le degré de la loi k_A est constant. Comme [19], notre étude de *Stack Overflow* montre que les aidants, qui posent également des questions, sont rares et peu actifs. C'est la raison pour laquelle nous générons $\overline{nQ(a_i)}$ comme $\overline{nA(a_i)}$, et de telle sorte que plus un agent donne de réponses, moins il pose de questions. Formellement, $rang^{\overline{nQ}}(a_i) = N + 1 - rang^{\overline{nA}}(a_i)$ où $rang^{\overline{nQ}}(a_i)$ dénote le rang de l'agent a_i par nombre moyen de questions bihebdomadaires décroissant.

Afin de déterminer le score d'une question, nous considérons, comme le laisse supposer nos observations (cf. section 5.3), une distribution qui est de type loi de puissance :

$$score(q) = maxS_q \times (id(q))^{k_{S_q}} \text{ avec } k_{S_q} < 0 \quad (2)$$

où le score maximal d'une question $maxS_q$ et le degré de la loi k_{S_q} sont constants.

La simulation d'une période t est divisée en deux phases (cf. figure 2). Pendant la phase de question, chaque agent a_i pose $nQ(a_i, t)$ questions. Pendant la phase de réponse, chaque agent réagit à $nA(a_i, t)$ questions posées par les autres agents [3]. Comme [13, 14], nous supposons que les agents visent à maximiser le bénéfice de leur réponse par rapport aux réponses antérieures. C'est la raison pour laquelle les agents choisissent de répondre aux questions les plus fréquentes. Parce que nous supposons qu'une réponse est d'autant plus pertinente que la question est fréquente, le score de la première réponse est proportionnel au score de cette question (k_{S_a}). Comme nous supposons qu'une réponse vient compléter à la marge les réponses précédentes à la même question, le score décroît de manière exponentielle à chaque nouvelle réponse (e_{S_a}).

3. Les variances sont suffisamment faibles pour maintenir la forme de la distribution de type loi de puissance pour le nombre moyen de messages publiés par période.

4 Simulateur

Notre simulateur `ABM4DCoP` [11] est fondé sur `NetLogo` [18] qui est un langage de programmation multi-agents et un environnement de développement intégré pour la conception de simulations à base d'agents. Même si `Netlogo` est convivial, il permet la conception de modèles de haut niveau avec des comportements complexes.

Comme le montre la figure 3, l'interface de notre simulateur se compose de trois types de composants pour :

- configurer les paramètres de la simulation ;
- visualiser les agents, les interactions et les liens ;
- analyser l'évolution des métriques qui sont calculées et affichées en temps réel pendant la simulation.

Les agents sont disposés en cercle dans le sens des aiguilles d'une montre par ordre croissant de $\overline{nA(a_i)}$. La représentation graphique de la communauté dans la figure 4 nous permet d'observer la distribution spatiale de la productivité des agents (en violet) et de leurs liens (en gris). Un arc représente l'aide apportée par l'agent source à l'agent destinataire. Le degré sortant d'un noeud est le nombre de pairs aidés et le degré entrant est le nombre de conseillers. Au cours de la simulation, les publications s'accumulent sur le segment radial de son auteur : les questions sont en vert et les réponses en rouge.

5 Calibration et simulation

Après une brève description du jeu de données utilisé pour la calibration, nous présentons nos résultats de simulation qui montrent que : (a) notre modèle permet de reproduire l'inégale répartition de l'activité au sein d'une communauté ; (b) la capitalisation des réponses aux questions les plus récurrentes explique l'éffrondrement qualitatif des réponses.

5.1 Données de calibration

Afin d'explorer notre terrain d'étude `Stack Overflow`, nous avons extrait, transformé et chargé les données brutes produites de 2008 à 2021 ($\sim$50 GiO) dans un entrepôt de données [10]. Notre modèle conceptuel multidimensionnel permet d'analyser des indicateurs clés de performance tels que le nombre de publications, la discutabilité des questions ou le temps de réponse. Le modèle logique multidimensionnel est alimenté par 6 communautés en face à

Données : a_i, t : agent/pas de temps
$H = (Q, A)$: historique des questions/réponses
Résultat : H' : historique
$nQ(a_i, t) = \mathcal{N}(\mu = \overline{nQ(a_i)}, \sigma^2 = 0.02)$;
si $randomU(0,1) \leq \{nQ(a_i, t)\}$ **alors**
  $nQ(a_i, t)$ += 1 ;
pour $j \in [1; nQ(a_i, t)]$ **faire**
  $id(q) = \left| \{q \in Q\} \right| + 1$;
  $time(q) = t$;
  $author(q) = a_i$;
  $score(q) = maxS_q \times (id(q))^{k_{S_q}}$;
  $Q \cup = q$;
retourner H

Données : a_i, t : agent/pas de temps
$H = (Q, A)$: historique de questions/réponses
Résultat : H' : historique
$nA(a_i, t) = \mathcal{N}(\mu = \overline{nA(a_i)}, \sigma^2 = 0.02)$;
si $randomU(0,1) \leq \{nA(a_i, t)\}$ **alors**
  $nA(a_i, t)$ += 1 ;
$Q_O = \langle q_1, \ldots q_n \rangle$;
/* où Q_O est la liste des $q_j \in \{q \in Q \mid author(q) \neq a_i\}$ ordonnée par $\frac{k_{S_a} \times score(q_j)}{e_{S_a}^{|\{a \in A | reply(a) = q_j\}|+1}}$ décroissant */
pour $j \in [1; nA(a_i, t)]$ **faire**
  /* Pour les questions les plus fréquentes Q_O */
  $id(a) = \left| \{a \in A\} \right| + 1$;
  $time(a) = t$;
  $author(a) = a_i$;
  $reply(a) = q_j$;
  /* Le score d'un réponse est proportionnel au score de la question et décroît de manière exponentielle à chaque nouvelle réponse */
  $score(a) = \mathcal{N}(\mu = \frac{k_{S_a} \times score(q_j)}{e_{S_a}^{|\{a \in H | reply(q_j)\}|+1}}, \sigma^2 = 0.05)$;
  $A \cup = a$;
retourner H

FIGURE 2 – Comportement individuel de demandeur (à gauche) et d'aidant (à droite)

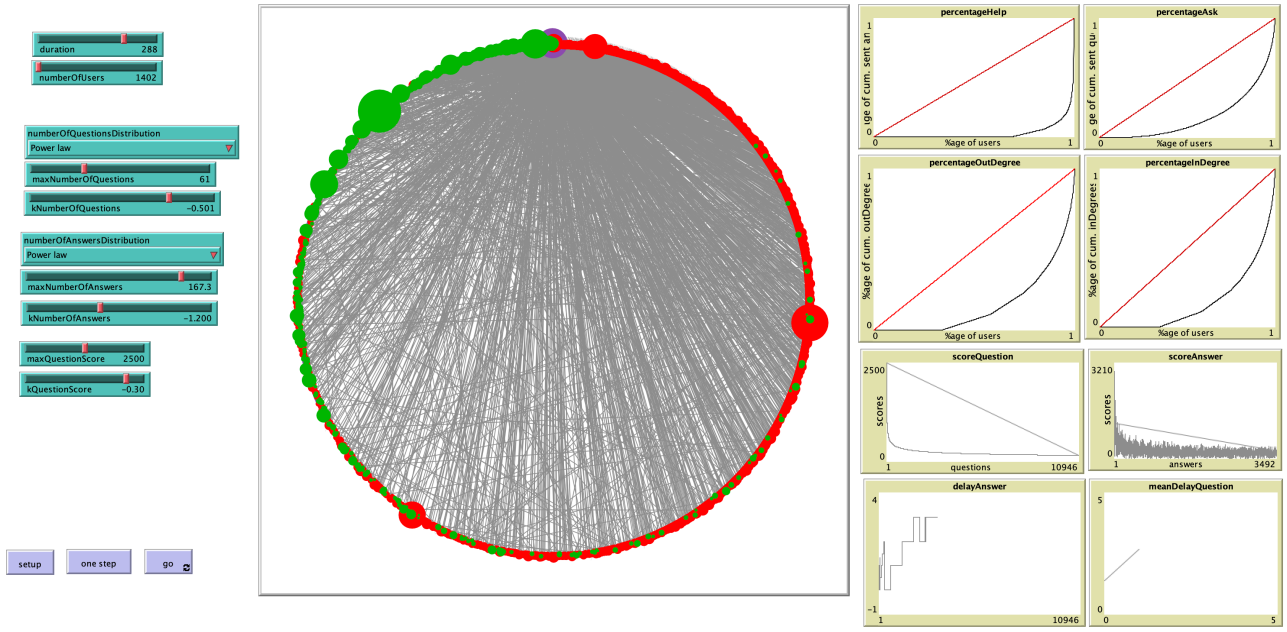


FIGURE 3 – Configuration (à gauche), visualisation des agents disposés en cercle (au centre) et restitution (à droite) de la simulation

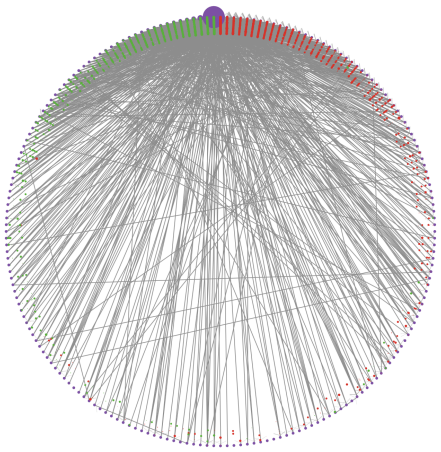


FIGURE 4 – Réseau communautaire quasi biparti

face : `Pig/Hive`, `Hadoop/Spark` et `R/Pandas`. En particulier, ces deux dernières communautés représentent environ 1 300 000 de messages publiés par près de 250 000 membres. Nous avons écarté 2000 publications (0,15 %) impliquant 4000 utilisateurs qui ont supprimé leur compte et pour lesquels nous ne disposions pas de données attributaires. De plus, l'activité après les premiers confinements (mars 2020) est exclue.

Afin de calibrer notre modèle sur la communauté `R` à partir de nos données réelles, nous considérons que le nombre maximum de questions (resp. réponses) bihebdomadaires est $maxQ = 61$ (resp. $maxA = 167,3$) et $k_Q = -0,5$ (resp.$k_A = -1,2$). Ainsi, notre modèle coïncide avec les données de calibration comme l'illustre la figure 5 qui représente le nombre moyen de réponses par période et par membre.

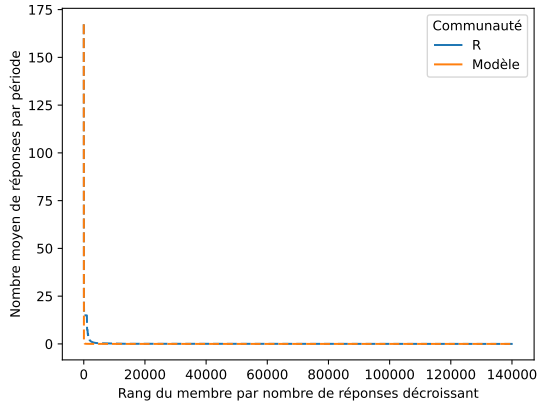


FIGURE 5 – Modèle calibré à partir de données réelles

5.2 Inéquité des contributions

La figure 6 représente la demande et l'aide cumulée des utilisateurs du plus petit au plus grand nombre moyen de questions (resp. réponses) par période, $\overline{nQ(a_i)}$ (resp. $\overline{nA(a_i)}$). Nous avons réimplémenté `CommunityNetSimulator` [5] avec `NetLogo`. Selon ce modèle, 20 % des membres les plus actifs sont responsables de plus de 66 % des réponses. Cependant, nous observons que 20 % des membres les plus actifs de la communauté `R` sont les auteurs de 94 % des réponses sur ce sujet. Selon notre modèle, l'inéquité de la distribution est similaire.

Dans les communautés en ligne, très peu de membres publient la majorité du contenu. Ce phénomène mésoscopique est partagé par tous les DCoPs que nous avons étudiés sur `Stack Overflow`. L'ordonnanceur pseudo-aléatoire et séquentiel de notre moteur, qui permet aux agents bavards de parler plus d'une fois par période, reproduit à dessein l'inégalité de l'activité événementielle.

5.3 Sécession

Notre modèle suppose ici que le score d'une question, qui est le résultat de votes en sa faveur (*upvote*) ou en sa défaveur (*downvote*), est proportionnelle à sa fréquence. Cette hypothèse est confirmée par la figure 7a qui représente pour chaque question posée dans les communautés `R` et `Pandas`, son score ainsi que le trafic web généré par son fil de discussion. On peut noter que, pour ne pas pénaliser les questions les plus récentes, nous considérons le nombre de vues par semaine d'ancienneté.

Dans les DCoPs que nous avons étudiés sur `Stack Overflow`, nous observons que les questions les mieux notées, c'est-à-dire les plus récurrentes, sont rapidement posées dans la phase de formation de la communauté. La figure 7b, qui représente le score des questions classées par ordre d'apparition, illustre cette observation. De plus, elle représente la distribution du score des questions de notre modèle (en vert) qui, à dessein, est de type loi de puissance.

Notre modèle suppose que le score de la meilleure réponse à une question, qui est le résultat de votes, est proportionnelle au score de la question à laquelle elle répond. Cette hypothèse est confirmée par la figure 7c qui représente pour chaque question à laquelle les communautés `R` et `Pandas` ont répondu, son score ainsi que celui

de sa meilleur réponse. C'est la raison pour laquelle notre modèle a été calibré avec $k_{S_a} = 1,5$. De manière similaire, nous avons calibré notre modèle avec $e_{S_a} = 2$.

Le phénonème de sécession des experts correspond à l'effrondrement qualitatif rapide des réponses. La figure 7d qui représente le score des réponses classées par ordre d'apparition illustre ce phénomène dans les communautés `R` et `Pandas`. On y observe également que notre modèle, avec 1402 agents (1% de la communauté `R`) simulé pendant 2 ans, est capable de reproduire ce même fait stylisé. Notre modèle explicatif décrit ce phénomène comme le résultat de la capitalisation des réponses aux questions les plus récurrentes. Les questions les plus fréquentes sont rapidement posées dans la phase de formation de la communauté. Pour maximiser le score de leur réponse, le comportement individuel des agents consiste à répondre en premier lieu à ces questions. Dès que ces dernières ont une réponse, la pertinence des conseils décroît rapidement.

6 Discussion

Nous avons proposé ici un modèle explicatif de la sécession des experts dans les communautés de pratiques. Selon notre modèle centré individus, l'effrondrement qualitatif précoce des réponses résulte de la capitalisation des réponses aux questions les plus récurrentes dans la communauté. Les membres, qui sont solidaires, évaluent la valeur d'une réponse proportionnelle à la valeur de la question et inversement proportionnelle au nombre de réponses pour cette même question. Leur comportement consiste à donner des réponses perspicaces aux questions les plus récurrentes qui sont rapidement posées puis complètent ces réponses pour finalement donner des réponses étroites aux questions les plus spécifiques. C'est la raison pour laquelle nous pensons que les mécanismes ludiques de gamification de la plateforme `Stack Overflow` devrait en premier lieu inciter les utilisateurs à poser de bonnes questions.

Bien que calibré à partir des données réelles massives, notre modèle se veut parcimonieux. En réduisant le nombre de paramètres et en limitant les mécanismes aux plus simples et aux plus généraux, nous proposons un modèle explicatif du phénomène de sécession.

Afin de calibrer la population de nos simulations, la population d'agents est homogène, i.e. modélisée par une loi de puissance, et leur expertise est mesurée par leurs activités *ex post*. Cette modélisation statistique de la distribution des caractéristiques idiosyncrasiques pourrait être raffinée en distinguant les « moineaux » (les individus les plus actifs dont les réponses ne sont pas nécessairement les plus pertinentes) des « hiboux » (les experts du domaine qui ne sont pas nécessairement les plus actifs). Le verrou technologique auquel nous sommes désormais confrontés réside dans le passage à l'échelle du modèle. Jouer 2 ans d'échanges d'une communauté constituée de 1402 agents (1% de la communauté `R`) nécessite plusieurs dizaines d'heure de simulation.

À plus long terme, nous souhaitons étudier et confronter d'autres phénomènes mésoscopiques tels que l'efficacité avec le temps de réponse et la controverse mesurée par le caractère discutable des questions.

Remerciements. Nous adressons nos remerciements aux relecteurs qui par leurs remarques nous ont permis de clarifier notre modèle.

Références

[1] Jérémy Albouys-Perrois, Nicolas Sabouret, Yvon Haradji, Mathieu Schumann et Christian Inard. « Simulation multi-agent de l'autoconsommation collective en relation avec l'activité des foyers ». In : *Actes des Journées Francophones sur les Systèmes Multi-Agents (JFSMA)*. Cépaduès, 2019, p. 129-138.

[2] Albert-László Barabási et Réka Albert. « Emergence of scaling in random networks ». In : *Science* 286.5439 (1999), p. 509-512.

[3] Mohamed Bouguessa, Benoit Dumoulin et Shengrui Wang. « Identifying Authoritative Actors in Question-Answering Forums : The Case of Yahoo ! Answers ». In : *Proc. of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. 2008, p. 866-874.

[4] Amal Chaoui, Sébastien Delarre, Eloire Fabien, Maxime Morge et Antoine Nongaillard. « Toward an Agent-Based Model of a Community of Practice : Demonstration ». In : *Advances in Practical Applications of Agents, Multi-Agent Systems, and Complex Systems Simulation. The PAAMS Collection*. T. 13616. LNAI. Springer, 2022, p. 467-472.

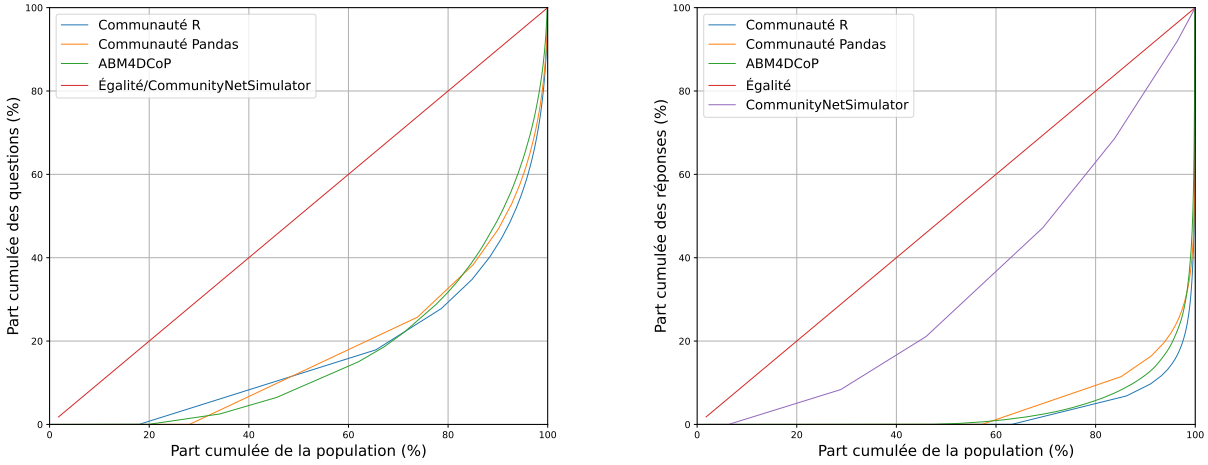
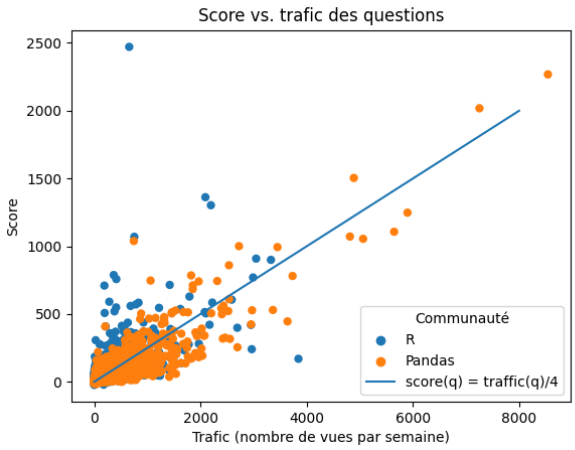
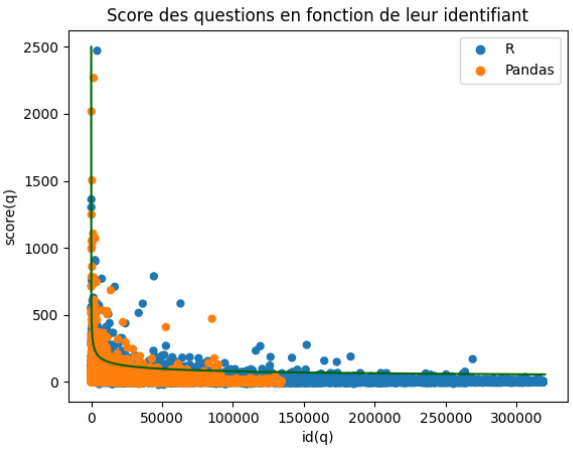


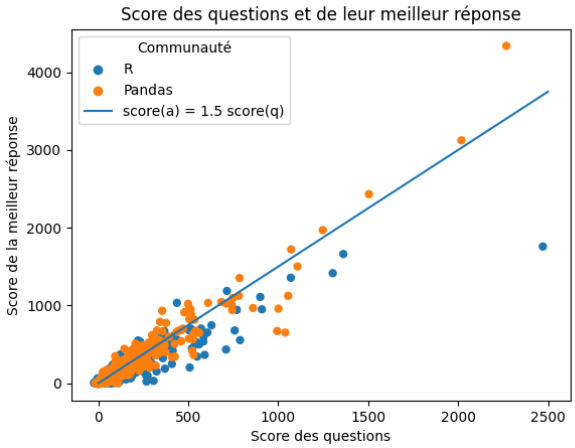
FIGURE 6 – Courbes de Lorenz qui montrent la répartition des questions (à gauche) et des réponses (à droite) au sein des communautés observées, simulées par `CommunityNetSimulator` et par `ABM4DCoP`



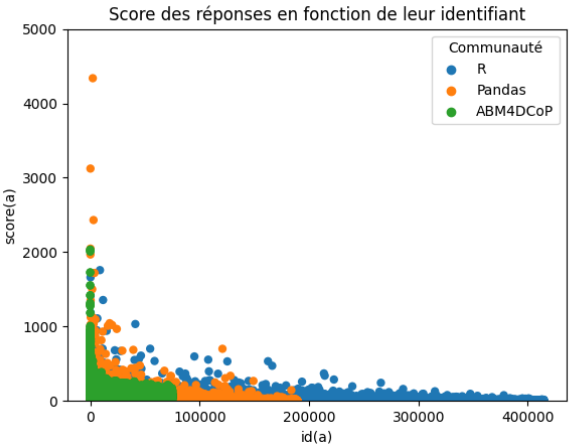
(a) Le score d'une question mesure sa fréquence



(b) Les questions fréquentes sont posées immédiatement



(c) Une bonne réponse nécessite une bonne question



(d) Les meilleurs reponses sont précoces

FIGURE 7 – Le phénomène de sécession des experts résulte de la capitalisation des réponses aux questions les plus récurrentes

[5] Amal Chaoui, Jarod Vanderlynden et Maxime Morge. *CommunityNetSimulator : Agent-Based Model for Simulating a Community*. https://gitlab.univ-lille.fr/mocicos/communitynetsimulator. Online ; accessed 2023-03-01.

[6] Kévin Darty, Julien Saunier et Nicolas Sabouret. « Calibration de simulations multi-agents à l'aide d'une méthode semi-automatique d'analyse du comportement (présentation courte) ». In : *Actes des Journées Francophones sur les Systèmes Multi-Agents (JFSMA)*. Cépaduès Éditions, 2015, p. 205-214.

[7] Nigel Gilbert, éd. *Simulating societies : the computer simulation of social phenomena*. UCL Press, 1994.

[8] Nicholas Hoernle, Gregory Kehne, Ariel D. Procaccia et Kobi Gal. « The phantom steering effect in Q&A websites ». In : *Knowledge and Information Systems* 64.2 (2022), p. 475-506.

[9] Philippe Mathieu et Sébastien Picault. « Calibrer les comportements d'agents à partir de données réelles ». In : *Revue d'Intelligence Artificielle* 28.4 (2014), p. 463-484.

[10] Maxime Morge. *SoDyOnStack : An Analysis of Social Dynamics on Stack Overflow*. https://gitlab.univ-lille.fr/mocicos/sodyonstack. Online ; accessed 2023-02-01.

[11] Maxime Morge et Amal Chaoui. *ABM4DCOP : Agent-Based Modeling for Distributed Community Of Practice*. https://gitlab.univ-lille.fr/mocicos/abm4dcop. Online ; accessed 2023-03-01.

[12] Iraklis Moutidis et Hywel T. P. Williams. « Community Evolution on Stack Overflow ». In : *PLOS ONE* 16.6 (17 juin 2021), e0253010. issn : 1932-6203.

[13] Aditya Pal, Shuo Chang et Joseph A Konstan. « Evolution of Experts in Question Answering Communities ». In : *sixth international AAAI conference on weblogs and social media*. T. 6. 1. 2012, p. 274-281.

[14] Aditya Pal, F Maxwell Harper et Joseph A Konstan. « Exploring question selection bias to identify experts and potential experts in community question answering ». In : *ACM Transactions on Information Systems (TOIS)* 30.2 (2012), p. 1-28.

[15] Quentin Reynaud, François Sempé, Yvon Haradji et Nicolas Sabouret. « Simuler l'activité humaine avec des approches statistiques basées sur les enquêtes emploi du temps ». In : *Actes des Journées Francophones sur les Systèmes Multi-Agents (JFSMA)*. Cépaduès Éditions, 2017, p. 117-126.

[16] Gilson Yukio Sato, Hilton José Silva de Azevedo et Jean-Paul A Barthès. « Agent and multi-agent applications to support distributed communities of practice : a short review ». In : *Autonomous Agents and Multi-Agent Systems* 25.1 (2012), p. 87-129.

[17] Etienne Wenger, Richard Arnold McDermott et William Snyder. *Cultivating communities of practice : A guide to managing knowledge*. Harvard business press, 2002.

[18] Wilensky, Uri. *NetLogo*. http://ccl.northwestern.edu/netlogo. Online ; accessed 2023-03-01.

[19] Jie Yang, Ke Tao, Alessandro Bozzon et Geert-Jan Houben. « Sparrows and owls : Characterisation of expert behaviour in stackoverflow ». In : *Proc. of International Conference on User Modeling, Adaptation, and Personalization*. Springer. 2014, p. 266-277.

[20] Jun Zhang, Mark S Ackerman et Lada Adamic. « Community Net Simulator : Using simulations to study online community networks ». In : *Proc. of International Conference on Communities and Technologies*. Springer, 2007, p. 295-321.

[21] Jun Zhang, Mark S Ackerman et Lada Adamic. « Expertise networks in online communities : structure and algorithms ». In : *Proc. of the International Conference on World Wide Web*. 2007, p. 221-230.