

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

Sébastien Delarre^a, Fabien Eloire^a,
Maxime Morge^b et
Antoine Nongaillard^b

^aUniv. Lille, CNRS, UMR 8019 Clersé,
F-59000 Lille, France

^bUniv. Lille, CNRS, Centrale Lille, UMR
9189, CRISAL, F-59000 Lille, France

6 juillet 2023

Distributed Community of Practice (DCoP)

Communauté de Pratique (CoP) [Wenger et al., 2002]

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.



The screenshot shows a Stack Overflow page with a question and one answer. The question is about filtering movies by genre using DataFrames. The answer provides a Scala code snippet to filter comedies and show their content.

Stack Overflow Products

Home

PUBLIC

- Questions**
- Tags
- Users

COLLECTIVES 1

- Explore Collectives

FIND A JOB

- Jobs

1 Answer

Active Oldest Votes

scala java jvm node.js c#

We offer great

- Events and celebration
- Meaningful work in rec
- Family-centric
- Paid vacation

Learn more

You can filter it:

```
0 val comediesDF = moviesDF.filter("tagline like '%comedy%'")
```

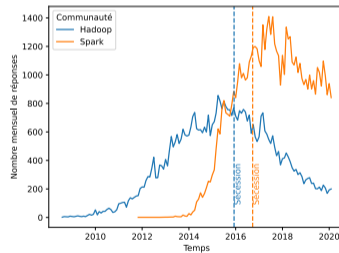
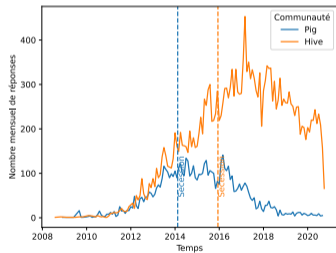
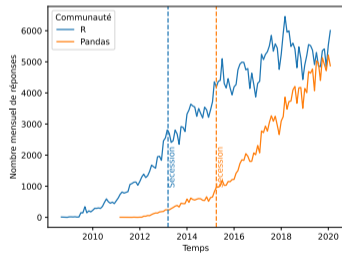
Then you can show the content:

```
✓ comediesDF.show(false)
```

Share Edit Delete Flag edited Feb 15 at 21:34 answered Feb 14 at 17:28 maxime.morge 20 • 4

Phénomène de sécession des experts

Évolution du nombre mensuel de réponses et date de sécession – à laquelle 80 des 100 réponses les mieux notées sont données – pour 6 communautés en face à face R/Pandas, Pig/Hive, Hadoop/Spark



Motivations

- **Objectif.**
Expliquer le phénomène de sécession des experts
- **Thèse.**
Les phénomènes sociaux complexes qui se jouent sur Stack Overflow sont le résultat de l'imbrication de comportements élémentaires microscopiques
- **Méthodologie.**
L'identification des processus de solidarité locale et des comportements sociaux individuels sous-jacents
- **Contributions.**
La reproduction de faits stylisés
 - ① l'activité est inégalement distribuée parmi les membres
 - ② les réponses aux questions les plus récurrentes sont rapidement capitalisées

Plan

- 1 Travaux connexes
- 2 Modèle
- 3 Simulateur
- 4 Calibration et simulation
- 5 Discussion

Travaux connexes

- Analyses statistique [Moutidis et Williams, 2021]
 - L'analyse de métriques macroscopiques montre que 34 % des utilisateurs de Stack Overflow deviennent inactifs. Pourquoi ?
- Analyses de réseaux sociaux [Bougussa et al., 2008; Yang et al., 2014]
 - PageRank et HITS ne sont adaptés aux DCoP
 - une petite proportion d'utilisateurs est responsable de la majorité des contributions
 - la disparition précoce des « hiboux » et l'arrivée continue de nouveaux « moineaux »
- Analyse temporelle [Pal, Chang et Konstan, 2012; Pal, Harper et Konstan, 2012]
 - la valeur d'une réponse est le bénéfice qu'elle apporte par rapport aux réponses antérieures
- Modélisation à base d'agents [Zhang et al., 2007a; Zhang et al., 2007b]
 - les liens résultent d'interactions
 - pas d'inégalité de l'activité d'une communauté
 - pas d'explication au phénomène de sécession des experts

Modèle centré individus

Contribution

Chaque contribution (question/réponse) c a : un identifiant $id(c)$, un horodatage $time(c)$, un auteur $author(c)$, un score $score(c)$ et éventuellement répond à une question, $reply(a)$.

Agent

Chaque agent a_i se caractérise par :

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

Population

La distribution du nombre bihebdomadaire de réponses d'un agent est de type loi de puissance :

$$\overline{nA(a_i)} = \max A \times \text{rank}^{\overline{nA}}(a_i)^{k_A} \text{ avec } k_A < 0 \quad (1)$$

Comportement individuel de demandeur et d'aidant

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 $\text{random}U(0, 1) \leq \{nQ(a_i, t)\}$ **alors**
 [$nQ(a_i, t) += 1$;

pour $j \in [1; nQ(a_i, t)]$ **faire**
 [$id(q) = |\{q \in Q\}| + 1$;
 $time(q) = t$;
 $author(q) = a_i$;

$score(q) = \max S_q \times (id(q))^{k_{S_q}}$;
 $Q \cup = q$;

retourner \underline{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 $\text{random}U(0, 1) \leq \{nA(a_i, t)\}$ **alors**

[$nA(a_i, t) += 1$;

$Q_O = \langle q_1, \dots, q_n \rangle$;

*/** où Q_O est la liste des $q_j \in \{q \in Q \mid a(q) \neq a_i\}$ ordonnée par

$\frac{k_{S_a} \times score(q_j)}{|\{a \in A \mid reply(a) = q_j\}| + 1}$ décroissant **/*

pour $j \in [1; nA(a_i, t)]$ **faire**

[$id(a) = |\{a \in A\}| + 1$;

$time(a) = t$;

$author(a) = a_i$;

$reply(a) = q_j$;

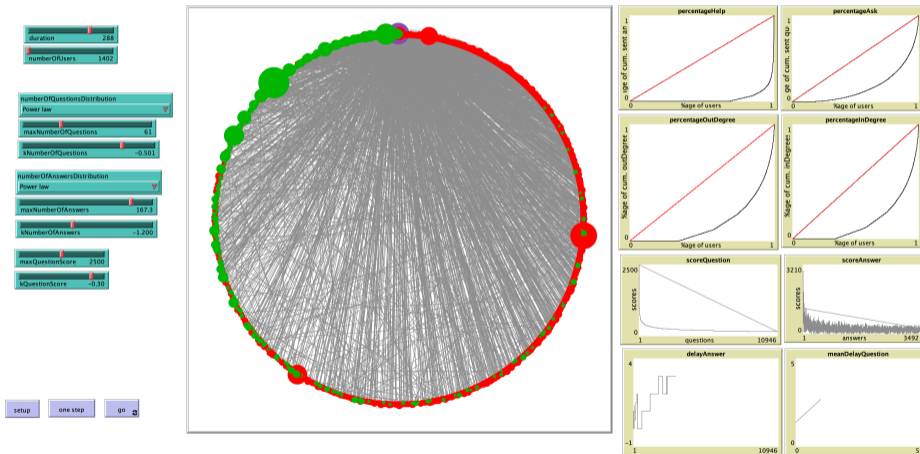
$score(a) = \mathcal{N}(\mu = \frac{k_{S_a} \times score(q_j)}{|\{a \in H \mid reply(q_j)\}| + 1}, \sigma^2 = 0.05)$;

[$A \cup = a$;

retourner \underline{H}

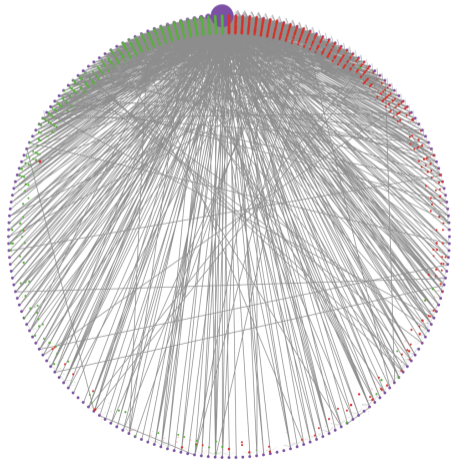
ABM4DCoP [Morge et Chaoui, 2022]

Configuration, visualisation and restitution



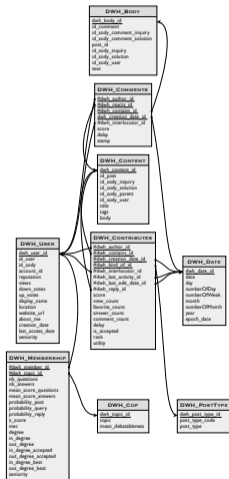
Réseau communautaire quasi biparti

Distribution spatiale de la productivité des agents



- Les agents
- Les liens sociaux
- Les contributions au cours de la simulation
 - Les questions
 - Les réponses

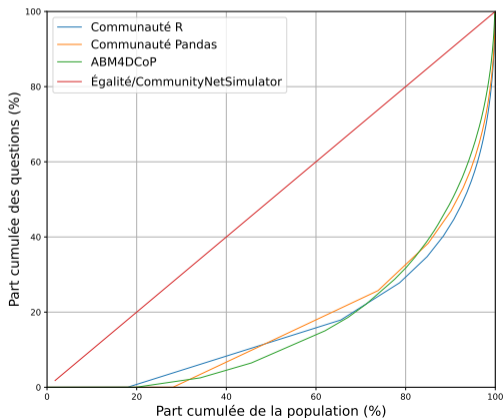
Données de calibration [Morge, 2021]



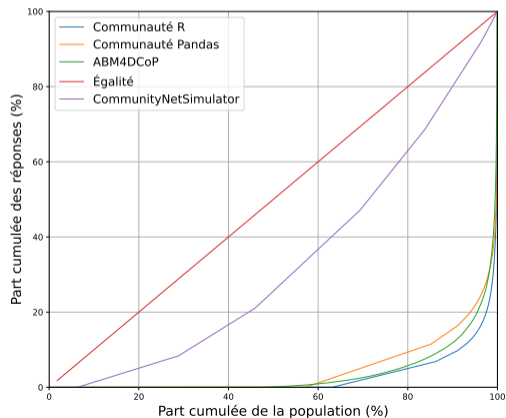
- Stack Overflow de 2008 à 2021 ~50 GiO
- 6 communautés en face à face :
Pig/Hive, Hadoop/Spark et R/Pandas.
- E.g. 1 300 000 contributions, 250 000 membres (0,15 % écartés)
- Sans l'activité après les premiers confinements (mars 2020)

Inéquité des contributions

Courbes de Lorenz qui montrent la répartition des questions et des réponses observées, simulées par CommunityNetSimulator et par ABM4DCoP au sein des communautés population.



Morge



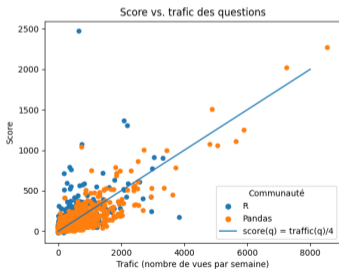
JFSMA23

Page 12

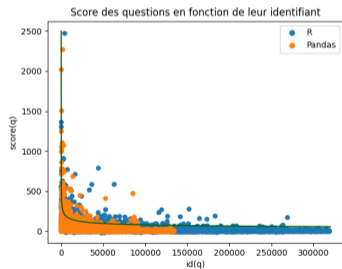
Sécession

3 hypothèses

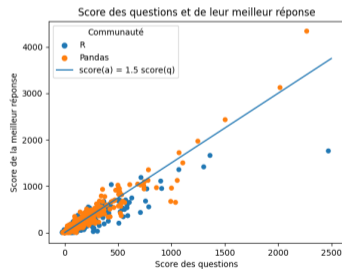
Le score d'une question mesure sa fréquence



Les questions fréquentes sont posées immédiatement

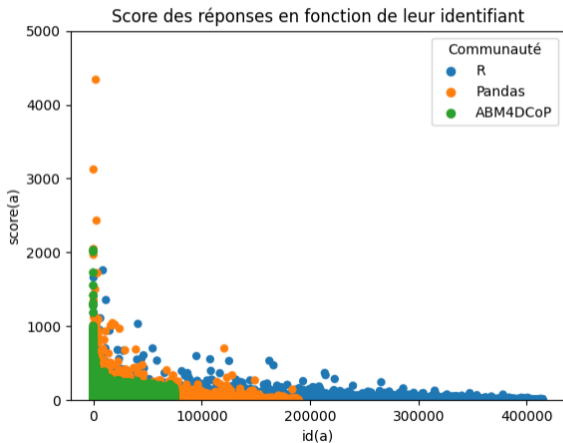


Une bonne réponse nécessite une bonne question



Sécession

Les meilleurs réponses sont précoces ou l'effondrement qualitatif des réponses



À emporter

- **Objectif.**

L'effondrement qualitatif précoce des réponses résulte de la capitalisation des réponses aux questions les plus récurrentes dans la communauté

- **Thèse.**

Les phénomènes sociaux complexes qui se jouent sur Stack Overflow sont le résultat de l'imbrication de comportements élémentaires microscopiques

- **Méthodologie.**

L'identification des processus de solidarité locale et des comportements sociaux individuels sous-jacents

- **Contributions.**

La reproduction de faits stylisés

- ① l'activité est inégalement distribuée parmi les membres
- ② les réponses aux questions les plus récurrentes sont rapidement capitalisées

- **Perspectives.**

- Le passage à l'échelle
- D'autres phénomènes mésoscopiques tels que l'efficacité ou la controverse

- Albouys-Perrois, J., Sabouret, N., Haradji, Y., Schumann, M., & Inard, C. (2019). Simulation multi-agent de l'autoconsommation collective en relation avec l'activité des foyers. Actes des JFSMA, 129-138.
- Assima, K. A. (2022). Analyse sociale aux seins des communautés de développeurs sur Stack Overflow [mém. de mast., Université de Lille].
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509-512.
- Bougoussa, M., Dumoulin, B., & Wang, S. (2008). Identifying Authoritative Actors in Question-Answering Forums : The Case of Yahoo! Answers. Proc. of the ACM Conference SIGKDD, 866-874.
- Chaoui, A. (2022). Agent-Based Model of a Community of Practice [mém. de mast., Ecole Centrale de Lille].

Bibliographie II

- Chaoui, A., Delarre, S., Fabien, E., Morge, M., & Nongaillard, A. (2022). Toward an Agent-Based Model of a Community of Practice : Demonstration. In Advances in Practical Applications of Agents, Multi-Agent Systems, and Complex System (p. 467-472, T. 13616). Springer.
- Chaoui, A., Vanderlynden, J., & Morge, M. (s. d.). CommunityNetSimulator : Agent-Based Model for Simulating a Community [Online; accessed 2023-03-01].
- Darty, K., Saunier, J., & Sabouret, N. (2015). Calibration de simulations multi-agents à l'aide d'une méthode semi-automatique d'analyse du comportement (présentation courte). In L. Vercouter & G. Picard (Éd.), Actes des JFSMA (p. 205-214). Cépaduès Éditions.
- Dewaele, A. (2022). Analyse sociale aux seins des communautés de développeurs sur Stack Overflow [mém. de mast., Université de Lille].
- Gilbert, N. (Éd.). (1994). Simulating societies : the computer simulation of social phenomena. UCL Press.

Bibliographie III

- Hoernle, N., Kehne, G., Procaccia, A. D., & Gal, K. (2022). The phantom steering effect in Q&A websites. *Knowl. Inf. Syst.*, *64*(2), 475-506.
- Kleinberg, J. M. (1999). Hubs, Authorities, and Communities. *ACM computing surveys (CSUR)*, *31*(4es), 5-es.
- Mathieu, P., & Picault, S. (2014). Calibrer les comportements d'agents à partir de données réelles. *Rev. d'Intelligence Artif.*, *28*(4), 463-484.
- Morge, M. (2021). SoDyOnStack : An Analysis of Social Dynamics on Stack Overflow [Online; accessed 2023-02-01].
- Morge, M., & Chaoui, A. (2022). ABM4DCOP : Agent-Based Modeling for Distributed Community Of Practice [Online; accessed 2023-03-01].
- Moutidis, I., & Williams, H. T. P. (2021). Community Evolution on Stack Overflow. *PLOS ONE*, *16*(6), e0253010.
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank Citation Ranking : Bringing Order to The Web (rapp. tech.). [Stanford InfoLab](#).

Bibliographie IV

- Pal, A., Chang, S., & Konstan, J. A. (2012). Evolution of Experts in Question Answering Communities. [sixth international AAAI conference on weblogs and social media](#), 6(1), 274-281.
- Pal, A., Harper, F. M., & Konstan, J. A. (2012). Exploring question selection bias to identify experts and potential experts in community question answering. [ACM Transactions on Information Systems \(TOIS\)](#), 30(2), 1-28.
- Reynaud, Q., Sempé, F., Haradji, Y., & Sabouret, N. (2017). Simuler l'activité humaine avec des approches statistiques basées sur les enquêtes emploi du temps. [Actes des JFSMA](#), 117-126.
- Sato, G. Y., de Azevedo, H. J. S., & Barthès, J.-P. A. (2012). Agent and multi-agent applications to support distributed communities of practice : a short review. [Autonomous Agents and Multi-Agent Systems](#), 25(1), 87-129.
- Schelling, T. C. (1971). Dynamic models of segregation. [Journal of mathematical sociology](#), 1(2), 143-186.

Vanderlynden, J. (2022).

Modélisation centrée « individus » de communauté de pratique : état de l'art
[mém. de mast., Université de Lille].

Wenger, E., McDermott, R. A., & Snyder, W. (2002).

Cultivating communities of practice : A guide to managing knowledge. Harvard
business press.

Wilensky, Uri. (s. d.). NetLogo [Online; accessed 2023-03-01].

Yang, J., Tao, K., Bozzon, A., & Houben, G.-J. (2014). Sparrows and owls : Characterisation of
expert behaviour in stackoverflow.

Proc. of Int. Conference on User Modeling, Adaptation, and Personalization, 266-277.

Zhang, J., Ackerman, M. S., & Adamic, L. (2007a). Community Net Simulator : Using
simulations to study online community networks. In

Proc. of Int. Conference on Communities and Technologies (p. 295-321). Springer.

Zhang, J., Ackerman, M. S., & Adamic, L. (2007b). Expertise networks in online communities : structure and algorithms.
[Proc. of the Int. Conference on World Wide Web, 221-230.](#)