



Impact des dimensions spatiale et temporelle dans la modélisation d'un phénomène collectif de type “ free-riding ”

Tomas Navarrete Gutierrez, Julien Siebert, Laurent Ciarletta, Vincent Chevrier

► To cite this version:

Tomas Navarrete Gutierrez, Julien Siebert, Laurent Ciarletta, Vincent Chevrier. Impact des dimensions spatiale et temporelle dans la modélisation d'un phénomène collectif de type “ free-riding ”. 18èmes Journées Francophones des Systèmes Multi-Agents - JFSMA'10, Oct 2010, Mahdia, Tunisie. 2010. <inria-00534600>

HAL Id: inria-00534600

<https://hal.inria.fr/inria-00534600>

Submitted on 10 Nov 2010

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.

Impact des dimensions spatiale et temporelle dans la modélisation d'un phénomène collectif de type « free-riding »

Tomás Navarrete Gutiérrez^a
tomas.navarrete@loria.fr

Julien Siebert^a
julien.siebert@loria.fr

Laurent Ciarletta^a
laurent.ciarletta@loria.fr

Vincent Chevrier^a
vincent.chevrier@loria.fr

^aLORIA - Campus Scientifique
BP 239 - 54506 Vandoeuvre-les-Nancy Cedex

Résumé

Dans cet article, nous présentons une comparaison de cinq modèles différents construits sur la même hypothèse comportementale individuelle d'un phénomène collectif présent dans les réseaux d'échanges de fichiers peer-to-peer : le « free-riding ». Nous étudions un modèle analytique global et quatre modèles à base d'agents. Les modèles multi-agent intègrent les dimensions temporelle et spatiale rarement présentes dans les modèles agrégés du phénomène retrouvés dans la littérature. Nous discutons les conditions *a priori* et expérimentales pour lesquelles les modèles sont équivalents. Nous montrons aussi qu'un même algorithme de décision individuelle peut amener des informations contradictoires.

Mots-clés : Simulation multi-agents, modèles de comportement, multiples modèles, phénomènes collectifs, free-riding, peer-to-peer

Abstract

We present a comparison of five different models built upon the same individual behavior hypothesis of a collective phenomenon present in peer-to-peer file exchange networks: "free-riding". We study a global analytical model and four agent based models. Multi-agent models include the space and time dimensions rarely seen in the literature discussing aggregated models of the collective phenomenon in question. We discuss the *a priori* and the experimental conditions under which the models are equivalent. We demonstrate that one individual decision algorithm can lead to contradictory information.

Keywords: multi-agent simulation, behavioral models, multiple models, collective phenomena, free-riding, peer-to-peer

1 Introduction

Les réseaux pair-à-pair (p2p) sont un moyen avantageux d'offrir et de partager des services. Ces systèmes sont peu coûteux, tolérants aux

pannes et aux déconnexions, faciles d'utilisation et passent aisément à l'échelle. Ces avantages sont dus à un concept fort : la décentralisation des services. Chaque utilisateur est à la fois client et fournisseur d'un service.

Cependant, il arrive que des utilisateurs de ces réseaux ne « jouent » pas le jeu, c'est à dire, qu'ils se contentent d'être uniquement clients du service. C'est le cas pour le partage de fichiers par des utilisateurs qui téléchargent des fichiers sans en retour en offrir au téléchargement. Ce comportement, appelé « free-riding », remet en cause le principe de base à l'origine des réseaux p2p et peut entraîner une dégradation des performances du réseau et le rendre inutilisable. La question qui se pose alors est de savoir à partir de quelle proportion de free-rider un réseau devient inutilisable.

Pour y répondre dans cet article, nous faisons l'hypothèse d'un comportement individuel rationnel et envisageons différentes modélisations intégrant différentes dimensions telles que l'espace ou le temps. Nous discutons des conditions *a priori* et expérimentales pour lesquelles ces modèles sont équivalents. Nous montrons qu'un même algorithme de décision individuelle peut amener des informations contradictoires selon le point de vue choisi pour considérer le phénomène collectif.

2 Modélisation de phénomènes collectifs

2.1 Activité de modélisation

Un modèle est une représentation généralement mathématique¹ qui explicite les relations entre les éléments d'un système. Mais comme le dit Minsky [10], un modèle existe dans la mesure où nous pouvons répondre à certaines questions sur l'objet modélisé en observant ou manipulant

¹ mais aussi bien physique ou graphique par exemple.

le modèle. Le rôle du créateur des modèles est donc aussi important que le modèle lui-même, car avant sa création, il faut savoir quelles questions lui seront posées.

Plus particulièrement, au sujet de phénomènes collectifs, deux catégories de modèles peuvent être utilisés :

Modèle analytique. Les modèles analytiques (ou encore basés équations, agrégés) proposent une description du phénomène à partir de grandeurs macroscopiques. Ils décrivent la variation de ces grandeurs en fonction du temps. Ils ont donc pour vocation à expliciter la dynamique globale du phénomène en se concentrant exclusivement sur un comportement moyen et global.

Selon la forme des équations, celles-ci peuvent être résolues mathématiquement (par exemple en cas de convergence des valeurs vers un point fixe) ou leur solution peut être calculée de manière approchée avec le calcul numérique.

Modèle multi-agent. Ce paradigme est approprié [4] pour simuler des systèmes distribués complexes tels que les phénomènes collectifs en biologie [6, 16, 17] ou en sociologie [2].

Dans cette approche, la dynamique globale d'un système, au niveau macroscopique, n'est pas donnée d'avance (à l'inverse des modèles précédemment cités) mais résulte de l'interaction des comportements de chaque agent au niveau microscopique. Les intérêts de cette approche sont de pouvoir représenter et simuler des systèmes ouverts ; de prendre en compte, dès la modélisation, les caractères dynamiques et hétérogènes des comportements individuels ; et de pouvoir analyser l'impact d'un comportement local sur le fonctionnement global du système [12]. De par son essence, cette forme de modélisation impose, pour des conditions initiales données, de recourir à la simulation pour connaître le comportement global du système, et par conséquent un coût de calcul.

2.2 Le phénomène de free-riding

Il s'agit du phénomène collectif que nous souhaitons modéliser.

Définition du free-riding. Dans un système qui prétend remplir ses objectifs grâce à la coopération organisée parmi tous les participants, un free-rider est un participant qui ne coopère pas. Le free-riding dans un réseau p2p de partage de

fichiers se caractérise par le téléchargement de fichiers sans en partager en retour. La présence importante de free-riders sur un réseau peut dégrader sa performance au point de le rendre inutilisable, d'où l'intérêt de combattre son apparition.

Modèles de free-riding dans des réseaux p2p d'échanges de fichiers. [13] et [5] sont parmi les premiers à s'être intéressés aux effets du free-riding sur la performance du réseau bien que l'existence du phénomène a été identifié pour la première fois en 2000 [1]. La plupart des modèles existants sont des modèles basés sur la théorie des jeux comme on peut le vérifier : [18, 9, 7]

Il existe une caractéristique commune à tous ces modèles (de par leur nature globale) : ils négligent les aspects en relation avec le réseau tels que la topologie ou la latence, et oublient les dimensions de temps et d'espace. Il est difficile de modéliser des utilisateurs hétérogènes dont le comportement évolue avec le temps car, comme nous l'avons souligné, les approches de type analytique ou basées sur des équations modélisent directement le fonctionnement global du système. Les modèles multi-agents ont déjà été utilisés dans la modélisation de systèmes p2p [11, 14] mais pas dans le cadre de notre travail.

2.3 Question abordée

L'hypothèse d'un utilisateur rationnel est très présente dans les modèles présentés en section 2.2 et plus particulièrement dans [3]. Cependant, les dimensions spatiales et temporelles sont peu abordées dans ces modèles. Or, si l'on est intéressé à savoir *si* et *quand*, dans un réseau p2p, *le système s'écroule* faute de « partageurs » *ou s'il est viable*, il ne suffit pas simplement de proposer un modèle mais également de cerner ses limites. Cela permettra de savoir sous quelles conditions la réponse qu'il donne s'applique.

Un modèle d'un phénomène ne peut se concevoir comme l'unique explication de celui-ci mais plutôt comme un moyen de conforter ou d'infirmer des hypothèses relatives à la structure du phénomène étudié, ou d'en prédire le comportement [17]. Dans notre cas d'étude (le free-riding dans les systèmes p2p), nous souhaitons évaluer ce qu'impliquent ces hypothèses relatives aux dimensions spatiale et temporelle vis à vis du comportement du système. Pour cela, nous allons utiliser, outre un modèle analytique,

des modèles multi-agents en nous inspirant directement de [12].

Nous allons comparer cinq modèles construits à partir d'un même comportement des utilisateurs. Pour ce faire nous nous sommes inspirés d'un modèle analytique existant et nous avons y ajouté graduellement d'autres dimensions telles que le temps et l'espace. Nous proposerons plusieurs modèles multi-agents dans la section 3.

3 Modélisation du free-riding

3.1 Hypothèse commune

Tous les modèles que nous proposons s'appuient sur un comportement rationnel de l'utilisateur. Nous empruntons à [3] le modèle comportemental de l'utilisateur dans lequel un usager de réseau P2P partagera s'il trouve un intérêt à le faire. Schématiquement, un utilisateur (i) est caractérisé par sa générosité (θ_i), il partagera si le coût de contribution (caractéristique du réseau) est en dessous de celle-ci.

Le coût de contribution est inversement proportionnel au niveau de contribution du système, le niveau de contribution étant défini comme la proportion de participants qui partagent des fichiers dans le système.

Nous insistons sur le fait que, même si nous ajoutons de nouvelles dimensions, le mécanisme de décision demeurera identique (c.f. algorithme 1) tout au long des nouveaux modèles. Par contre, la perception du niveau de contribution sera globale ou locale (dimension spatiale) et la prise de la décision décalée ou non dans le temps.

Algorithme 1 Décision d'un agent

```

si  $\theta_i \geq \text{coût}$  alors
   $\text{partage} \leftarrow \text{Vrai}$ 
sinon
   $\text{partage} \leftarrow \text{Faux}$ 
fin si

```

3.2 Modèle analytique

Feldman *et al.* [3] proposent un modèle avec un utilisateur rationnel i qui partagera des fichiers si $\theta_i \geq \frac{1}{X}$ où X est le niveau de contribution du réseau et sa réciproque est le coût de contribution. Si l'utilisateur ne partage pas il fera du free-riding.

En supposant i) une répartition particulière de la générosité dans le réseau (la distribution de la valeur θ est l'uniformité : $\theta \sim U(0, \theta_{max})$), ii) une décision synchrone des agents et iii) une perception complète du réseau ; il est possible de résoudre exactement l'équation correspondante². Par exemple, si 20% des utilisateurs partagent ($x_{init} = 0.2$) et la générosité du réseau est uniformément distribué entre 0 et une valeur maximal $\theta_{max} = 10$, le coût de contribution sera $\frac{1}{x_{init}} = 5$. Cela est représenté dans la figure 1 par la flèche allant de $(0, 0.2)$ à $(5, 0.2)$. La question qui se pose maintenant est de savoir combien d'utilisateurs partageront étant donnée un coût de contribution de 5. La réponse sera donné par la probabilité $P(\theta_i \geq 5)$ c'est à dire le complément de la densité de probabilité d'une variable suivant une distribution uniforme. Nous avons alors $P(\theta_i \geq 5) = 1 - \frac{1}{x_{init} * \theta_{max}} = 0.5$. Le niveau de contribution sera maintenant de 50%.

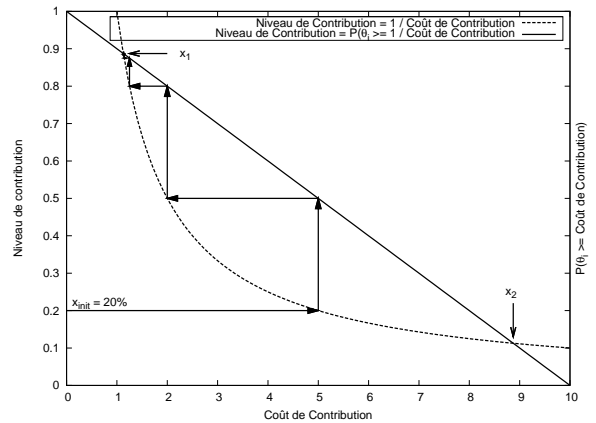


FIG. 1: Évolution du niveau de contribution d'après le modèle analytique avec $\theta_{max} = 10$ et $x_{init} = 20\%$.

En effet, cette proposition analytique donne une dynamique de point fixe avec deux points particuliers : x_1 et x_2 . Le niveau de contribution final X_f du système avec un niveau de contribution initial x_{init} est donné par :

$$X_f = \begin{cases} x_1 & \text{si } x_{init} \geq x_2, \\ 0 & \text{si } x_{init} < x_2. \end{cases}$$

Bien que le modèle permette de décrire la dynamique globale du système (convergence vers une valeur connue en fonction de conditions initiales), nous en avons identifié deux limites :

1. **Il implique une synchronisation des actions entre tous les utilisateurs.** Les effets

²Pour en savoir plus voir [3]

des actions des utilisateurs sont toujours immédiats et tous les utilisateurs réagissent en même temps. Avec l'utilisation d'internet comme moyen de communication et la nature ouverte des réseaux p2p, il est évident que la supposition préalable à propos du synchronisme est éloignée de la réalité.

2. **Tous les utilisateurs ont une connaissance « parfaite » et « globale » du réseau.** Le niveau de contribution exact est connu par chaque utilisateur quand il décide de partager ou non. Nous croyons qu'en réalité, un utilisateur ne peut acquérir directement ce type d'information. Au mieux, l'utilisateur peut tenter de l'estimer à partir de l'information donnée par son logiciel client mais il ne pourra en connaître la valeur exacte.

Les modèles suivants vont intégrer ces dimensions spatiale et temporelle en conservant la même fonction de décision pour les agents et la même répartition initiale de la générosité³.

3.3 Modèles multi-agents

Nous proposons quatre nouveaux modèles multi-agents qui permettent la prise en compte des deux dimensions suivantes :

1. *Temps.* Un modèle qui a une dimension temporelle peut décrire des utilisateurs qui agissent à différents moments.
2. *Espace.* Un réseau qui représente les interactions entre les utilisateurs peut décrire les utilisateurs estimant le niveau de contribution basé sur leurs perceptions locales.

Dans ces modèles, les agents représentent les utilisateurs du réseau. Ils ont des interactions au sein de l'environnement (le réseau p2p) qu'ils perçoivent au travers de l'information qu'ils obtiennent de leurs logiciels clients. Une approche de type multi-agents nous permet de comprendre l'impact des comportements locaux dans l'état global du système tels qu'une prise de décision asynchrone ou l'utilisation d'information locale dans la prise de décision.

En ajoutant les dimensions de temps et d'espace dans les nouveaux modèles, nous serons en mesure de manipuler le moment où chaque utilisateur change son état ainsi que la perception du niveau de contribution de chaque utilisateur.

³ Contrainte due au modèle analytique insoluble dans le cas général.

Modèle Itéré. Dans un premier temps, nous allons convertir le modèle analytique en un modèle multi-agents. Nous avons créé un modèle qui imite le comportement des utilisateurs tel qu'il a été décrit dans le modèle analytique. Ainsi les utilisateurs ont une connaissance complète du niveau de contribution du système à n'importe quel moment et changent leur état seulement après que tous les utilisateurs aient pris leur décision. Nous avons N agents avec un mécanisme de décision identique et qui décident de manière synchrone. Chaque itération correspond au temps de prise de décision (c.f. algorithme 2).

Algorithme 2 Modèle itéré.

```

pour tout pasDeTemps faire
   $X \leftarrow \text{calculerContributionGlobale}()$ 
  coût  $\leftarrow \frac{1}{X}$ 
  pour tout agent faire
    agent.decider()
  fin pour
fin pour

```

Modèle asynchrone. Dans cette modélisation, tous les participants d'un réseau p2p ne réagissent pas en même temps. Les modèles précédents n'incluent pas de dimension temporelle et ne font aucune référence à propos de l'ordre dans lequel les participants du système interagissent. La manière initiale d'introduire une dimension de temps à la modélisation consiste à permettre seulement à une partie des agents de décider s'ils contribuent ou non à chaque pas de temps.

Pour matérialiser l'inclusion d'une dimension « temporelle » dans ce modèle, nous avons permis seulement à un pourcentage d'agents, sélectionnés d'une manière uniformément aléatoire, de changer leur état à chaque pas de temps. Une fois que ces agents ont mis à jour leur état, nous avons recalculé le niveau global de contribution et dans le pas de temps suivant, nous avons permis à nouveau à un pourcentage d'utilisateurs sélectionnés de manière uniformément aléatoire de changer d'état (c.f. algorithme 3).

Algorithme 3 Modèle asynchrone.

```

pour tout pasDeTemps faire
   $X \leftarrow \text{calculerContributionGlobale}()$ 
  coût  $\leftarrow \frac{1}{X}$ 
  ensemble  $\leftarrow \text{ensembleAléatoireDAgents}()$ 
  pour tout agent  $\in$  ensemble faire
    agent.decider()
  fin pour
fin pour

```

Modèle topologique. Ici, nous remettons en cause le fait que tous les participants connaissent parfaitement et globalement le système. Dans ce modèle les utilisateurs n’ont pas accès au niveau de contribution « réel » et seront obligés de l’estimer en se basant sur leurs interactions avec d’autres utilisateurs. Pour mettre en œuvre cela, nous ajoutons une dimension spatiale au travers d’une organisation des utilisateurs en réseau de type Small-World suivant l’algorithme décrit en [19]. L’algorithme commence avec le graphe d’un treillis de N noeuds. Chaque noeud i a des liens vers ses k voisins les plus proches : à gauche $i - 1, i - 2, \dots, i - \frac{k}{2}$ et à droite $i + 1, i + 2, \dots, i + \frac{k}{2}$. Le dernier noeud a comme voisins à droite les noeuds $0, 1, \dots, \frac{k}{2}$, ce qui rend le treillis circulaire. Après, on prend les liens à droite de chaque noeud et avec une probabilité p on va rediriger le lien vers un autre noeud, sélectionné de manière aléatoire, ne faisant pas partie de l’ensemble original des voisins du noeud.

Il existe des indices [8, 15] suggérant que certains réseaux p2p ont une structure de type Small-World. De plus, nous avons choisi ce type de réseau parce qu’il permet d’avoir facilement des réseaux complètement structurés, comme un treillis circulaire, aussi bien que des réseaux complètement aléatoires (c.f. figure 2). L’intérêt de cette caractéristique est de pouvoir évaluer la pertinence de nos expériences par rapport à la connaissance de la structure du réseau. Dans le

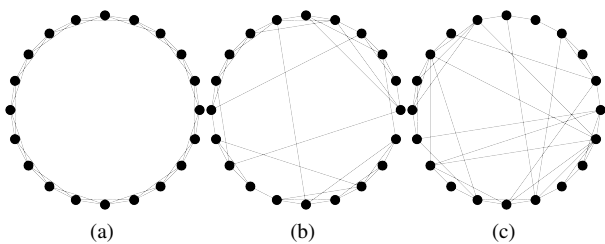


FIG. 2: Exemples de graphes générés avec l’algorithme Small-World [19]. (a) $p = 0$ (Treillis circulaire), (b) $p = 0.25$, (c) $p = 1$ (réseau aléatoire)

graphe du réseau les sommets représentent les utilisateurs et les arcs des interactions abstraites ou génériques entre les utilisateurs.

Dans ce modèle, avant de prendre la décision de partager ou pas, l’agent estime le niveau global de contribution grâce au niveau de contribution que nous appelons « local » et qui correspond à la proportion de ses voisins dans le graphe qui

partagent des fichiers. Quand un agent estime le niveau de contribution uniquement à partir de ses voisins directs, nous disons qu’il a réalisé une estimation de profondeur $d = 1$. Une estimation de profondeur $d = 2$ signifie qu’en plus de ses voisins directs, l’agent a pris en considération, pour faire son calcul du niveau local de contribution, les voisins directs de ses voisins. Et ainsi de suite pour d’autres valeurs de d (c.f. algorithme 4).

Algorithme 4 Modèle Topologique.

```

pour tout pasDeTemps faire
  pour tout agent faire
     $X \leftarrow \text{calculerContributionLocale}(\text{profondeur } d)$ 
     $\text{coût} \leftarrow \frac{1}{X}$ 
    agent.décider()
  fin pour
fin pour

```

Modèle asynchrone et topologique. Dans cette modélisation, les agents perçoivent de manière locale (du point de vue topologique) les éléments d’information nécessaires à la prise de décision et uniquement une partie des usagers réagit à chaque pas de temps des simulations.

Comme pour le modèle topologique, nous avons utilisé un réseau de type Small-World aux caractéristiques identiques. Comme pour le modèle asynchrone, à chaque pas de temps, seule une partie des agents, sélectionnée de manière aléatoire, sera planifiée pour estimer le niveau de contribution et décider de partager ou pas.

3.4 Conditions d’équivalence à priori

En considérant les définitions des cinq modèles décrits auparavant, nous pouvons identifier les conditions d’équivalence de construction sous lesquelles les résultats des modèles seront *a priori* équivalents.

Si nous utilisons le modèle *itéré*, nous pouvons nous attendre aux mêmes résultats que ceux obtenus avec le modèle analytique, avec l’imprécision entraînée par le calcul numérique.

Pour le cas du modèle *topologique*, il peut y avoir deux formes d’équivalence lorsque l’agent a une estimation parfaite du niveau global de contribution :

- si un agent peut d’interagir avec tous les autres agents,
- si un agent peut estimer la contribution avec la profondeur maximale

Plus généralement, plus la profondeur augmentera ou plus le graphe sera totalement connecté

plus les résultats doivent s'approcher de ceux des modèles itéré ou analytique.

Le modèle *asynchrone* sera équivalent aux modèles itéré ou analytique si tous les agents décident de partager ou pas lors d'un pas de temps. Plus généralement, plus la proportion d'agents qui sont actifs à un pas de temps est grande, plus les résultats doivent s'approcher de ceux des modèles itéré ou analytique

Enfin, le modèle asynchrone et topologique nécessite deux conditions qui doivent se remplir en même temps pour qu'il y ait une équivalence avec les autres modèles : d'un côté, une grande partie d'agents doit pouvoir entrer en action et de l'autre, les agents doivent pouvoir interagir avec beaucoup d'autres agents.

4 Simulation

Dans cette partie, nous comparons à partir de résultats analytiques et expérimentaux le comportement prédit par les modèles. Ce qui nous intéresse c'est de comparer, selon le comportement en terme de stabilisation ou d'extinction, la valeur où le système se stabilise. Plus particulièrement, ce qui nous intéresse dans le cas expérimental c'est la dispersion autour de la valeur de stabilisation et la quantité temps nécessaire pour que le niveau de contribution reste stable.

4.1 Paramètres communs des simulations

- Une quantité de $N = 1000$ utilisateurs⁴.
- $\theta \sim U(0, \theta_{max})$ où $\theta_{max} = 10$. Ce qui implique que les utilisateurs avec une valeur de $\theta_{max} = 10$ partageront si le niveau de contribution du système est d'au moins 10%.
- La solution au modèle analytique donne les points d'attraction suivants : $x_1 = 0.88729833$ et $x_2 = 0.11270167$. Cela signifie que le niveau de contribution du système se stabilisera à 88.7% quand le niveau de contribution initial x_{init} est supérieur ou égal à 11.2% et que le niveau de contribution du système tombera à 0 si le niveau de contribution initial x_{init} est inférieur à 11.2%.

Dans les figures qui reportent les résultats des simulations les *points d'attraction* x_1 et x_2 , prédits par le modèle analytique, ont été identifiés

⁴ Des expériences ont été faites avec des valeurs de N allant jusqu'à neuf millions, mais par souci d'espace et considérant que les résultats observés dans ces autres expériences étaient essentiellement les mêmes nous ne présentons ici que pour $N = 1000$.

dans les cas pertinents avec une ligne respectivement horizontale et verticale. Pour chaque point, nous présentons aussi la valeur minimale et la valeur maximale observée tout au long de ces séries de simulations.

Nous avons considéré comme critère de fin de simulation soit le 200^{ème} pas de temps, soit la stabilisation du niveau de contribution global : le niveau obtenu est identique à celui du pas de temps précédent.

Dans les modèles qui incluent un graphe de type Small-World, nous avons toujours utilisé une valeur de $k = 10$ pour garantir l'obtention d'un graphe complet.

4.2 Modèle Itéré

Les expériences ont reproduit les résultats prédits par le modèle analytique à quelques exceptions près. Parfois, autour du point d'instabilité x_2 , le niveau final de contribution n'a pas été celui attendu comme nous pouvons l'observer dans la figure 3. Nous supposons que cela est dû à des questions d'arithmétique (arrondi) lors des calculs. Par conséquent, nous considérons que le système se comporte de la même manière que les prédictions du modèle analytique mais avec une marge d'erreur induite par la simulation numérique.

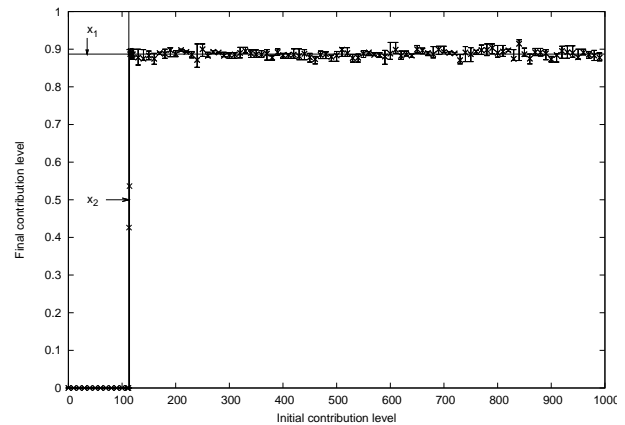


FIG. 3: Modèle itéré. Valeurs moyennes, minimales et maximales du niveau de contribution final.

La valeur moyenne de pas de temps pour arriver à la stabilisation du niveau de contribution a été de 4.5 avec un écart type de 1.56. La figure 4 montre cela en regroupant les résultats par niveau de contribution initial x_{init} .

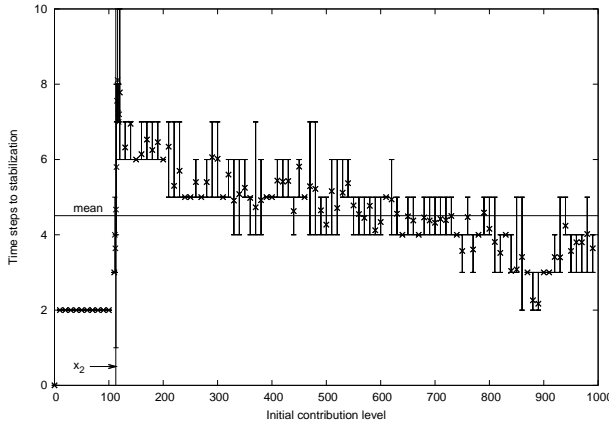


FIG. 4: Modèle itéré. Valeurs moyennes, minimales et maximales des pas de temps nécessaires à la stabilisation.

4.3 Modèle asynchrone

Les expériences faites avec ce modèle consistent à faire varier la quantité d'agents actifs par pas de temps entre 5% et 95% par tranche de 5%.

Nous pouvons observer dans la figure 6 que à partir de 25% d'agents actifs dans le système, le niveau de contribution a la même dynamique que les modèles itéré et analytique. Par contre, en observant la figure 5, qui reporte les résultats du modèle asynchrone avec une faible quantité d'agents actifs par pas de temps, nous pouvons constater que plus le pourcentage est faible, plus le niveau de contribution sera proche de 80%.

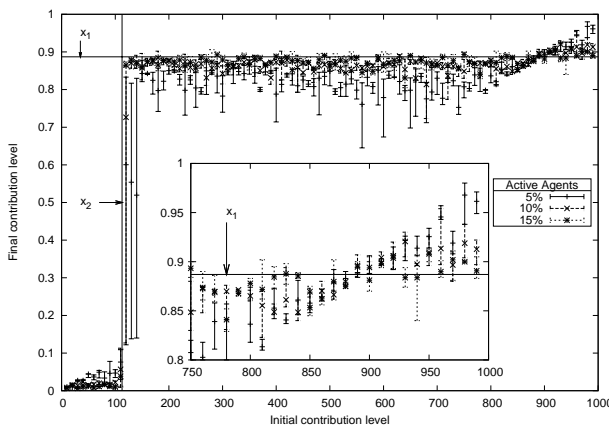


FIG. 5: Modèle asynchrone. Valeurs moyennes, minimales et maximales du niveau de contribution final 5%, 10% et 15% des agents actifs par pas de temps dans.

Dans la figure 7, nous pouvons constater que quand le niveau de contribution initial est inférieur à environ 11% (valeur correspondante au

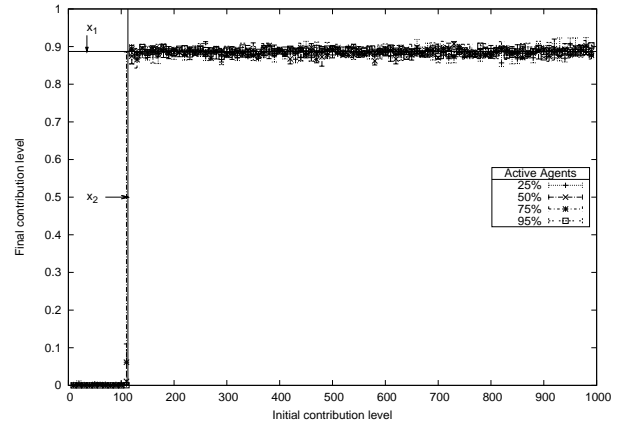


FIG. 6: Modèle asynchrone. Valeurs moyennes, minimales et maximales du niveau de contribution final avec 25%, 50%, 75% et 95% des agents actifs par pas de temps.

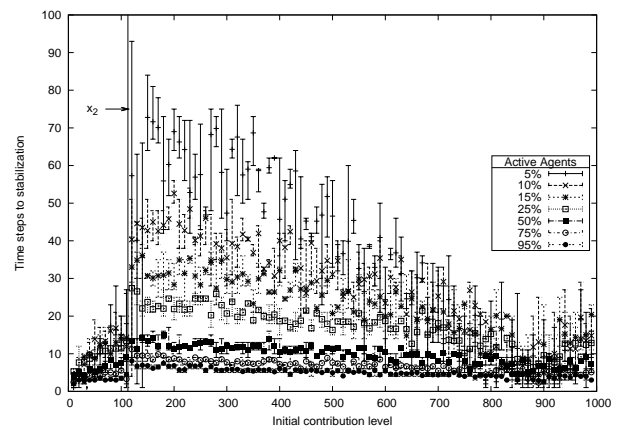


FIG. 7: Modèle asynchrone. Valeurs moyennes, minimales et maximales des pas de temps nécessaires à la stabilisation.

point d'attraction x_2) la quantité de pas de temps nécessaires pour la stabilisation est moins importante que pour un niveau de contribution initial supérieur à 11%. On notera aussi que plus le pourcentage d'agents actifs par pas de temps sera faible, plus la quantité de pas de temps nécessaires pour la stabilisation sera importante pour toutes les valeurs du niveau de contribution initial.

4.4 Modèle topologique

Avec le modèle topologique nous avons mené des expériences consistant à faire varier d'une part, la profondeur d des voisins que chaque agent utiliserait pour estimer le coût de partager, et d'autre part, la probabilité p qui rend les réseaux utilisés dans le modèle plus ou moins

structurés selon que p tend vers 0 ou 1. Les valeurs utilisées pour la profondeur sont 1, 2 et 3 et les valeurs utilisées pour la probabilité p sont 0, 0.25, 0.50, 0.75 et 1. Nous présentons seulement les résultats de la simulation avec une profondeur $d = 1$ et $d = 3$ dans les figures 8 et 9 car ils sont suffisamment représentatifs. Pour le réseau configuré comme un treillis circulaire ($p = 0$), le niveau de contribution final commence à se stabiliser à une valeur supérieure à 0 à partir d'un niveau de contribution initial x_{init} supérieur à 3%. Pour le reste des configurations avec $p > 0$ la stabilisation du niveau de contribution à une valeur supérieure à 0 commence à partir d'une valeur autour de 5%. Dans tous ces cas, les valeurs de stabilisation supérieures à 0 retrouvées quand $x_{init} < 7\%$ ne sont pas proches du point d'attraction prédit par le modèle analytique x_1 . Par contre, quand $x_{init} > 7\%$ les valeurs de stabilisation sont proches du point d'attraction x_1 .

Nous allons cependant souligner que quand la valeur de profondeur $d = 3$, le système a le même comportement que celui du modèle itéré *sauf* dans le cas du réseau configuré comme un treillis circulaire $p = 0$ (c.f. figure 9).

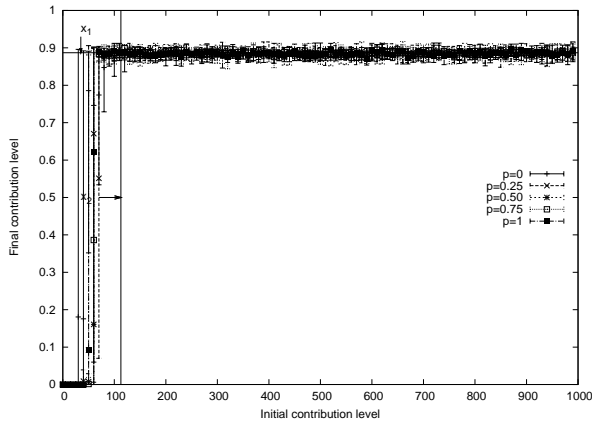


FIG. 8: Modèle topologique. Valeurs moyennes, minimales et maximales du niveau de contribution final pour une profondeur $d = 1$.

Pour les pas de temps nécessaires à la stabilisation, nous présentons seulement le cas où $d = 3$ dans la figure 10 car elle est représentative des deux autres cas $d = 1$ et $d = 2$. Pour les trois profondeurs $d = 1, 2$ et 3 , le treillis circulaire (le réseau configuré avec $p = 0$) a les valeurs moyennes les plus élevées de pas de temps nécessaires à la stabilisation quand le niveau de contribution initial se trouve entre 5% et 11%.

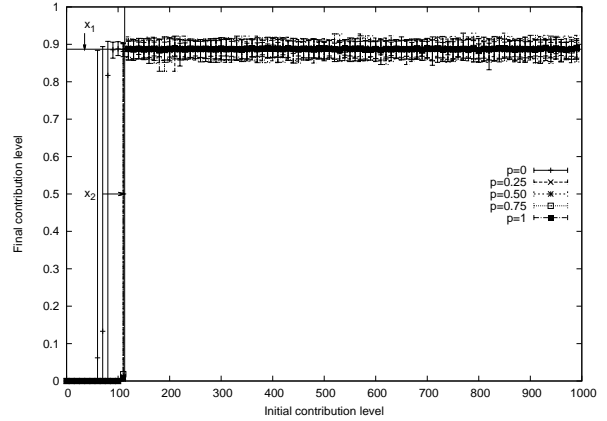


FIG. 9: Modèle topologique. Valeurs moyennes, minimales et maximales du niveau de contribution final pour une profondeur $d = 3$.

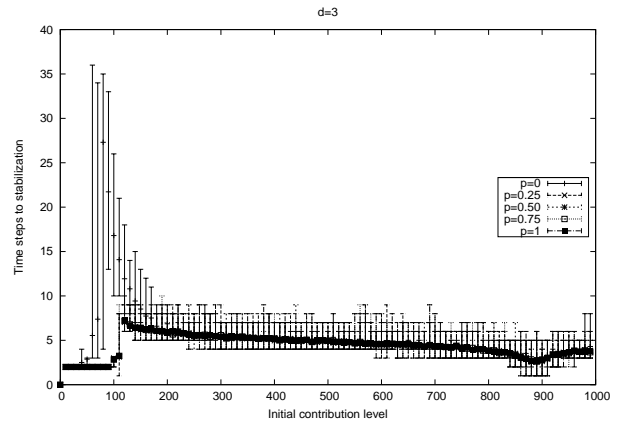


FIG. 10: Modèle topologique. Pas de temps nécessaires pour la stabilisation du modèle topologique avec profondeur $d = 3$

4.5 Modèle asynchrone et topologique

Pour les expériences menées avec ce modèle, nous avons fait varier la profondeur d entre 1, 2 et 3 et la quantité d'agents actifs par pas de temps entre 5% et 95% par tranche de 5%. Nous présentons seulement les résultats de la simulation avec une profondeur $d = 1$ dans la figure 11 car ils sont suffisamment représentatifs des deux autres profondeurs.

On peut observer dans la figure 11 que le niveau de contribution final devient 0 quand la quantité d'agents actifs par pas de temps est inférieure à 25%. La même figure nous montre que quand le niveau de contribution initial est supérieur à 25% nous avons une stabilisation du niveau de contribution à des valeurs qui se rapprochent de

80% quand la quantité d'agents actifs par pas de temps se rapproche de 95%. On pourra noter :

- dans la figure 6 le modèle asynchrone avec 25% d'agents actifs se stabilise sur la valeur prédite par le modèle analytique,
- dans la figure 8 le modèle topologique (mais avec 100% d'agents actifs) se stabilise sur la valeur prédite par le modèle analytique,
- dans la figure 11 le modèle asynchrone et topologique avec 25% d'agents actifs et une topologie équivalente à celle de la figure 8 se stabilise entre 0.2 et 0.4 et **non** sur la valeur prédite .

La façon de modéliser le collectif a donc un impact sur le comportement du système alors que le comportement individuel n'a pas changé. Finalement, on a observé dans ces expérimentations que plus le réseau utilisé est aléatoire (p se rapproche de 1) plus les valeurs de stabilisations seront importantes par rapport à la quantité d'agents actifs par pas de temps.

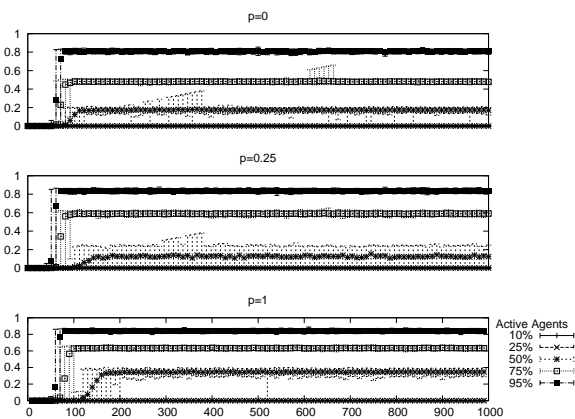


FIG. 11: Modèle asynchrone et topologique. Valeurs moyennes, minimales et maximales du niveau de contribution final pour une profondeur $d = 1$ avec 10%, 25%, 50%, 75% et 95% d'agents actifs par pas de temps.

4.6 Discussion

La question motivant cette étude a été d'évaluer les implications d'hypothèses supplémentaires relatives aux dimensions spatiales et temporelles vis à vis du comportement du système (lequel comportement constitue la raison d'être de nos modèles). Il est possible de répondre à cette question selon plusieurs points de vue.

Un premier concerne le comportement quantitatif du système, à savoir, le réseau se stabilise-t-il à la même proportion de partageurs quel que soit le modèle utilisé ? ou le comportement qualitatif. Ce dernier peut revêtir différentes formes :

le système est-il actif ou éteint ? Le système se stabilise-t-il (éventuellement à des proportions différentes de partageurs selon les modèles) ?

Les résultats des simulations montrent que, bien que nous puissions identifier des équivalences entre les modèles dans certains cas, celles-ci n'apparaissent que sous certaines conditions. Par exemple, aussi bien le modèle asynchrone que le modèle topologique montrent le même comportement qualitatif et quantitatif que les modèles analytique ou itéré quand $x_{init} \geq x_2$ (stabilisation à une valeur fixée), mais il n'en est pas de même lorsque $x_{init} < x_2$ (il n'y a pas systématiquement extinction du système). Le modèle asynchrone et topologique a le comportement contraire : il coïncide avec les modèles analytique ou itéré pour $x_{init} < x_2$ (extinction) mais quand $x_{init} \geq x_2$, il se stabilise plus ou moins avec une certaine amplitude autour de la valeur définitive moyenne.

Un autre point de vue concerne le temps de résolution. Avec le modèle analytique, nous avons une réponse immédiate à la question de la dynamique du phénomène sans alimenter le modèle avec le nombre d'utilisateurs du système modélisé. Par contre, avec les modèles multi-agents, nous sommes obligés de donner des valeurs aux paramètres avant de les simuler. Dans notre cas d'étude cela implique, par exemple, d'avoir une estimation de la quantité d'utilisateurs du système, du type de réseau, du taux de connexion/déconnexion d'utilisateurs ou de leur fréquence d'action. De plus, le comportement du système sera obtenu en agrégeant les valeurs du niveau de contribution individuelles.

Dans la mesure où nous avons des équivalences, il est possible de déterminer quel modèle utiliser selon les conditions initiales. En particulier, nous pourrions dire que pour les modèles qui reproduisent le comportement du modèle analytique quand $x_{init} \geq x_2$, nous n'avons pas besoin de faire tourner des simulations pour savoir la valeur de stabilisation finale, et encore moins de spécifier les paramètres de construction (notamment du graphe).

Cependant, si nous sommes intéressés par le processus qui conduit au résultat, à la variabilité de la valeur finale, seul les modèles multi-agents permettent de répondre à ces questions.

De manière plus générale, les modèles multi-agents sont plus riches que le modèle analytique parce qu'ils permettent de modifier le temps pendant lequel les utilisateurs seront en action, mais aussi, de modifier l'espace dans lequel ils

se trouvent et par conséquent leurs perceptions de l'environnement. Ils sont donc beaucoup plus facilement utilisables pour répondre à une famille de questions nécessitant des hypothèses larges. Ainsi, passer du modèle tel que proposé par [3], à un modèle où la répartition des valeurs de la générosité suit une autre loi que celle uniforme, est aisé avec le modèle itéré mais peut être rapidement (très) problématique avec un modèle analytique.

5 Conclusions et Perspectives

On s'aperçoit que les modèles analytiques et les modèles individu-centrés peuvent être complémentaires [6]. On retrouve deux (généralement) attracteurs dans la dynamique du système (soit la contribution se stabilise, soit elle devient 0). Cependant le modèle analytique ne peut pas spécifier la valeur exacte des attracteurs (ni sa variabilité) pour chacune des configurations du réseau que ce soit dans sa dimension temporelle ou spatiale (intégration de caractéristiques locales). Ils n'arrivent pas non plus à décrire le processus sous-jacent à la dynamique mais le modèle analytique a réussi à capturer l'essence de la dynamique (stabilisation).

Un même algorithme/mécanisme de décision au niveau de l'individu dans les modèles, mais avec une implémentation de la dimension espace/temps différente, peut donner des informations contradictoires (aussi bien sur le plan qualitatif que sur le plan quantitatif). C'est le cas pour les résultats obtenus avec le modèle asynchrone et topologique : il est clair qu'il existe un impact sur le niveau de contribution final lors que il y a peu d'agents actifs et que ces dernières ont peu de connaissance du réseau. Il est donc important de savoir dans quelles limites un modèle s'exprime. Nous avons trouvé les critères d'équivalences entre les modèles ainsi que les rangs de valeurs sous lesquels nous pouvons utiliser un modèle avec un point de vue global (modèle analytique) pour répondre à des questions qui traitent de paramètres locaux. Cependant, nous avons besoin de connaître de manière relativement précise les conditions initiales pour pouvoir sélectionner le modèle qui correspond le plus à nos intentions.

Nous aimerions également affiner la façon selon laquelle la dimension temps est ajoutée afin de pouvoir clarifier les effets d'addition de la dimension espace. La durée des actions est un des aspects spécifiques que nous voulons étudier. Nous aimerions pouvoir considérer l'effet

de modéliser chaque usager avec un temps de réaction, ainsi que pouvoir considérer l'effet de sélectionner une partie des usagers dans chaque pas de temps et utiliser les résultats dans un contexte de contrôle/prédiction.

Remerciements

Nous remercions les relecteurs pour leurs remarques qui ont permis d'améliorer l'article.

Références

- [1] E. Adar and B. A. Huberman. Free riding on gnutella. *First Monday*, 5(10), 2000. [online].
- [2] F. Amblard and D. Phan. *Modélisation et simulation multi-agents : Applications pour les Sciences de l'Homme et de la Société*. Hermès science publications, 2006.
- [3] Michal Feldman, Christos H. Papadimitriou, John Chuang, and Ion Stoica. Free-riding and whitewashing in peer-to-peer systems. *IEEE Journal on Selected Areas in Communications*, 24(5) :1010–1019, 2006.
- [4] J. Ferber. *Les systèmes multi-agents : vers une intelligence collective*. InterEditions Paris, 1995.
- [5] Z. Ge, DR Figueiredo, S. Jaiswal, J. Kurose, and D. Towsley. Modeling peer-peer file sharing systems. In *IEEE INFOCOM 2003*.
- [6] V. Grimm and S.F. Railsback. *Individual-based modeling and ecology*. Princeton Univ Pr, 2005.
- [7] L. Jian and J.K. MacKie-Mason. Why share in peer-to-peer networks? In *Proceedings of the 10th international conference on Electronic commerce*. ACM New York, NY, USA, 2008.
- [8] M. Jovanović, F. Annexstein, and K. Berman. "Modeling peer-to-peer network topologies through small-world models and power laws". In *TELFOR 2001*, 2001.
- [9] R. Krishnan, M.D. Smith, Zhulei Tang, and R. Telang. The impact of free-riding on peer-to-peer networks. In *System Sciences*, 2004.
- [10] M. Minsky. Matter, mind and models. In *Proceedings of International Federation of Information Processing Congress*, volume 1, pages 45–49, 1965.
- [11] M. Morge and P. Mathieu. Conception de mécanisme décentralisé pour les systèmes P2P d'échange de fichiers. In *JFSMA'2006*, pages 145–158, Annecy, France, 2006.
- [12] H.V.D. Parunak, R. Savit, and R.L. Riolo. Agent-Based Modeling vs Equation-Based Modeling : A Case Study and Users' Guide. *Lecture Notes in Computer Science*, 1534 :10–25, 1998.
- [13] L. Ramaswamy and L. Liu. Free riding : A new challenge to peer-to-peer file sharing systems. In *Proceedings of the Hawaii International Conference on Systems Science*, 2003.
- [14] J. Siebert, V. Chevrier, and L. Ciarletta. Impact du comportement des utilisateurs dans les réseaux pair-à-pair (P2P) : modélisation et simulation multi-agents. In *JFSMA 2008*, pages 129–138, Brest, France, 10 2008.
- [15] D. Stutzbach, R. Rejaie, and S. Sen. Characterizing unstructured overlay topologies in modern p2p file-sharing systems. *IEEE/ACM Trans. Netw.*, 16(2) :267–280, 2008.
- [16] V. Thomas, C. Bourjot, and V. Chevrier. Du collectif pour la satisfaction individuelle : un modèle d'auto-organisation. In *JFSMA'2003*, pages 261–265, Hammamet, Tunisia, 2003. none.
- [17] V. Thomas, C. Bourjot, V. Chevrier, and D. Desor. Démarche incrémentale pour l'évaluation d'un modèle multi-agent en éthologie. In *ARCo'07 : Cognition – Complexité – Collectif*, pages 177–188, Nancy France, 2007.
- [18] DK Vassilakis and V. Vassalos. Modelling real p2p networks : The effect of altruism. In *IEEE P2P 2007*, pages 19–26, 2007.
- [19] DJ Watts and SH Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393 :440–442, 1998.