



HAL
open science

Génération multi-agents de réseaux sociaux

Audren Bouadjio Boulic

► **To cite this version:**

Audren Bouadjio Boulic. Génération multi-agents de réseaux sociaux. Modélisation et simulation. Université Toulouse 1 Capitole (UT1 Capitole), 2021. Français. NNT: . tel-03400140

HAL Id: tel-03400140

<https://ut3-toulouseinp.hal.science/tel-03400140>

Submitted on 24 Oct 2021

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.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 1 Capitole (UT1 Capitole)*

Présentée et soutenue le *(15/03/2020)* par :

Audren Bouadjio-Boulic

Génération multi-agents de réseaux sociaux

JURY

GAUDOU BENOÎT
TUBARO PAOLA
BERTELLE CYRILLE
THIRIOT SAMUEL
FREDERIC AMBLARD

Maître de conférence
Directrice de recherche
Professeur d'Université
Chargé de recherche
Professeur d'Université

Président du Jury
Rapporteur
Rapporteur
Membre du Jury
Membre du Jury

École doctorale et spécialité :

MITT : Domaine STIC : Intelligence Artificielle

Unité de Recherche :

IRIT (UMR 5505)

Directeur(s) de Thèse :

Chihab Hanachi et Frédéric Amblard

Rapporteurs :

Tubaro Paola et Bertelle Cyrille

Résumé

Nous nous intéressons dans cette thèse à la reproduction de réseaux sociaux par une simulation multi-agents. Deux approches distinctes sont traditionnellement utilisées pour générer des réseaux. La première cherche à reproduire mathématiquement les propriétés topologiques de réseau. La seconde mime les comportements sociaux amenant la formation d'un réseau. L'avantage de cette dernière approche est qu'elle est compréhensible et maîtrisable par des utilisateurs issus de n'importe quel domaine de recherche. L'inconvénient est que les réseaux obtenus n'ont pas de propriétés topologiques semblables aux réseaux réels. A contrario, la première approche permet d'obtenir des réseaux fidèles topologiquement. Les méthodes faciles d'utilisation proposent des résultats qualitatifs ; obtenir des résultats quantitatifs demande l'usage de modèles plus difficiles d'utilisation. Nous positionnons notre modèle de génération de réseaux entre ces deux grands courants. La génération de réseaux est le résultat d'une simulation multi-agents ayant suivi des règles interprétables socialement, et la validation du résultat se fait sur des critères topologiques du réseau obtenu. Cette démarche nous inscrit dans l'étude des systèmes complexes : les phénomènes microscopiques d'interactions entre agents amènent des résultats macroscopiques que sont les réseaux obtenus. Cette thèse propose une approche de la reproduction de réseaux sociaux par une simulation multi-agents. Dans ce cadre, nous proposons une définition théorique d'un Framework général pour la génération de réseaux ainsi que les concepts et les mécanismes communs sur lesquels se basent nos modèles de génération de réseau.

Deux modèles distincts sont développés depuis le Framework. Chacun fixe une instanciation particulière des mécanismes mettant, ce faisant, en avant la généricité de notre approche. Afin de trouver les meilleurs paramètres de simulation pour reproduire un réseau donné, un algorithme génétique est utilisé. Chacun de ces modèles ont des avantages et inconvénients en terme de réseaux générables. Nous confrontons le résultat des deux modèles de génération à différents types de réseaux sociaux. Nous pensons avoir proposé une méthode de génération de réseau novatrice, visant à la fois une facilité d'utilisation et une précision dans les résultats obtenus, allié à une forte modularité.

Abstract

In this thesis, we study the reproduction of social networks using an Agent-Based simulation. Two distinct approaches are often used in order to generate social networks. The first consists in mathematically replicating the topological properties of the network. The other is to recreate the social behaviors leading to network creation. The advantage of the latter is that it is understandable and masterable by users from any domain. The drawback of this approach is that the networks generated may not have realistic topologies. On the other hand, the first approach leads to good topological results. Unfortunately, the easy-to-use models leads to quantitative results whereas the qualitative results requires the use of harder models. Our model stands between these two approaches. The generation of the network is the result of an agent-based simulation where agents follow social-interpretable rules. The validation of the result is performed on the generated network topology. This method is related to complex systems : microscopic interaction phenomena leads macroscopic results which are generated networks. First, a general framework is theoretically defined, offering concepts and common mechanisms shared by model instances. Experiments are performed to analyze agent behavior and the resulting network topologies. Two instances of the model are then created and finally the parameter space of the model is explored using a genetic algorithm. Each of these model offer advantages and drawbacks concerning network generation. We compare the results of those models with different social networks. In this work, we offer an original way to generate networks which is both easy to use and precise in the topologies generated.

remerciement

Je tiens avant tout à remercier l'ensemble des membres du jury, et en particulier Paola Tubaro et Cyrille Bertelle, rapporteur de cette thèse, pour leur lecture critique et la qualité de leurs retours.

Concernant la vie professionnelle, j'aimerais remercier toutes les personnes ayant pu m'assister de près ou de loin pendant ces années. Mes encadrants, Chihab Hanachi et Frédéric Amblard ; Les membres du projets Genstar, en particulier Kevin Chapuis et Samuel Thiriot pour ce fameux coding camp ; Les membres d'OpenMole, Romain Reuillon et Paul Chapron, qui m'ont proposé les outils pour explorer mes modèles ; les membres du projet Atimodis et en particulier Philippe Ramirez avec qui il a été très agréable et enrichissant de travailler ; les membres de l'équipe SMAC, en particulier Benoit Gaudou.

Concernant la vie personnelle quotidienne, Carlos, compagnon d'infortune et plus encore, Les membres de l'openspace "qui n'est pas une salle de TP", Caro, Jean, Denis, Medhi, Jiefu, Mickael, Raphaëlle.

Ceux à la croisée des deux, qui m'ont soutenu indirectement par des conversations enrichissantes, leur bonne humeur ou leur lumière sur le chemin tortueux de l'administration, le duo infernal Nathalie Valles et Franck Ravat, Yves Duthen, Chantal Soulé-Dupuy, Ronan Tournier, Michèle Cuesta. les membres du projet Monoil, en particulier Nicolas Maestripieri et Medhi Saqualli.

En ce qui concerne des choses bien plus personnelles, mes amis proches, pour leur soutien ou non, me permettant de retrouver motivation ou au contraire de me changer les idées.

Lounes l'un des derniers qui partage ma vision du monde ; au même titre que Cxx, chercheur de vérité ; Nico et son expertise d'une certaine App ; L'amitié de Vlad ; Tristan et sa pédagogie ; Alex et son insouciant nihilisme ; Jo le philosophe ; Yo et nos conversations à base scientifique ; Tout les habitants du Marais Salant : H, Ben et Marc, leur boss, le milliardaire le commissaire et le père ; Guillaume, relecteur attentif ; Jules & Colas pour leur attention (bien trop) constante sur mon avancée ; et ofc, mais cela depuis des années plus anciennes encore, les disciples du Roi, Rilles, Mot, Muzu le gris, Fab, Ling x3 et Chalban ; ainsi que mes amitiés profondes et antédiluviennes, Matthias, Etienne, Matthieu, Sylvain x2.

Bien entendu, le trépied familiale au sein duquel nous partageons un amour inconditionnel.

Table des matières

1	Introduction	1
1.1	Des systèmes complexes aux réseaux	2
1.2	Contexte de la thèse	3
1.3	Motivations	4
1.4	Problématique de la thèse	6
1.5	Plan du manuscrit	6
2	Etat de l’art	7
2.1	Introduction	8
2.2	Systèmes complexes	9
2.2.1	Définition d’un système complexe	9
2.2.2	L’émergence dans un système complexe	10
2.2.3	Dynamique d’un système complexe : de la stabilité au chaos	12
2.2.4	Auto organisation des entités constituantes	14
2.2.5	La modélisation de systèmes complexes	15
2.2.6	Conclusion sur les systèmes complexes	18
2.3	Des graphes aux réseaux sociaux	19
2.3.1	Le réseau en tant que représentation d’un système	19
2.3.2	Les réseaux sociaux : un cas particulier de réseau	23
2.3.3	Conclusion sur les réseaux sociaux	29
2.4	Génération de réseaux sociaux	30
2.4.1	Introduction aux paradigmes de génération	30
2.4.2	Contexte	30
2.4.3	Familles de méthodes de génération de réseaux	32
2.4.4	Les méthodes de génération de réseaux sociaux	35
2.4.5	Conclusion sur la génération de réseaux sociaux	57
3	Le Framework général	58
3.1	Les objectifs du modèle	60
3.1.1	Introduction	60
3.1.2	Objectif du Framework & des modèles	60
3.1.3	Élaboration d’un patron de solution	63
3.1.4	Résumé des objectifs	66
3.2	Le Framework	67

3.2.1	Le simulateur	67
3.2.2	Les dynamiques de construction de réseau	70
3.2.3	La recherche de réseau	73
3.3	Indicateurs du Framework	74
3.3.1	Propriétés observées	74
3.3.2	Éléments mis en avant par les observations	78
3.4	Expérimentations sur le Framework	79
3.4.1	Objectif des expérimentations	79
3.4.2	Test du simulateur	80
3.4.3	Reproduction d'un réseau sans échelle	89
3.4.4	Etude des dynamiques en fonction des paramètres de propagation	98
3.4.5	Questions sur la symétrie des actions	99
3.4.6	Reproduction d'un réseau Small-World	103
3.4.7	Conclusion	112
3.5	Conclusion du Framework et de ses expérimentations	113
3.5.1	Retour sur les objectifs	113
3.5.2	Instanciation en modèles	114
3.6	Résumé du chapitre	115
4	Modèle I	116
4.1	Définition du modèle I	118
4.1.1	Les complexe-actions	118
4.1.2	Propagation	119
4.1.3	Scoring	119
4.2	Résultats du modèle I	123
4.2.1	Configurations et réseau à reproduire	123
4.2.2	Résultats des simulations	124
4.2.3	Retour sur les objectifs	136
4.2.4	Discussions et élaboration du modèle II	138
5	Modèle II	140
5.1	Définition du modèle II	142
5.1.1	Les complexe-actions	142
5.1.2	Propagation	142
5.1.3	Scoring	143
5.2	Résultat du modèle II	145
5.2.1	Configuration et réseaux à reproduire	145
5.2.2	Résultats des simulations	146
5.2.3	Discussion sur les résultats	153
5.2.4	Retour sur les objectifs	153

6 Conclusion et perspectives	155
6.1 Conclusion	155
6.1.1 Résumé de la contribution	155
6.1.2 Conclusion sur les objectifs	156
6.1.3 Conclusion technique	157
6.2 Perspectives et ouvertures	157
6.2.1 Perspectives	157
6.2.2 Ouvertures	158

Génération multi-agents de réseaux sociaux

Audren Bouadjio-Boulic

8 février 2021

Chapitre 1

Introduction

Sommaire

1.1	Des systèmes complexes aux réseaux	2
1.2	Contexte de la thèse	3
1.3	Motivations	4
1.4	Problématique de la thèse	6
1.5	Plan du manuscrit	6

Dans le domaine des systèmes complexes, modéliser permet d'appréhender un système, un phénomène ou un processus complexe. Dans ce cadre, les réseaux sont une représentation compacte d'objets et de leurs relations. Les réseaux sociaux peuvent alors modéliser de manière adéquate des relations sociales entre individus. Ces structures deviennent les supports de processus qui, lors de simulations, permettent entre autres de tester des hypothèses.

Les données concernant les réseaux de la vie réelle peuvent être manquantes ou incomplètes, il peut donc être utile de pouvoir en générer.

Nous proposons ici une approche visant à reproduire une structure de réseau, sans se soucier du réalisme ou de l'interprétation des processus de création et destruction de liens sous-jacents. Une simulation multi-agents est lancée, et de l'application de règles sociales par les entités émerge un réseau. Les règles sociales disponibles pour les entités se propagent pendant la simulation, ce qui permet un espace de paramètres initiaux faible et l'intégration flexible de nouveaux agents en cours de simulation. Les paramètres initiaux déterminent les actions disponibles et leurs probabilités de propagation. La topologie du réseau obtenue à la fin de la simulation est évaluée par comparaison de certaines de ses propriétés topologiques avec celles du réseau à reproduire. Cette évaluation permet à un algorithme génétique de choisir les meilleures configurations pour la reproduction du réseau cible.

1.1 Des systèmes complexes aux réseaux

L'homme, doué de conscience, cherche à expliquer son environnement. L'une des méthodes pour y parvenir est la méthode scientifique : on cherche alors à décrire un phénomène observé par des lois universelles, appliquées au contexte particulier. Des outils développés, abandonnés ou constamment améliorés sont disponibles, à l'instar des mathématiques. Afin de pouvoir les utiliser, une étape de théorisation du "réel" observé subjectivement est nécessaire. Une fois théorisé, une modélisation peut y être appliquée.

Un champ de la science est dévolu à la compréhension des systèmes dits complexes, pour lesquels les résultats des interactions qui s'y jouent ne sont pas triviaux. Pour les comprendre, les scientifiques individuels s'appliquent à les théoriser puis à les modéliser. Si les résultats obtenus par les expériences sur cette modélisation se rapprochent de l'observation du phénomène réel, on peut alors considérer que la théorie est un candidat explicatif valide.

Évoluant a contrario de la vision réductionniste cartésienne, l'étude des systèmes complexes s'intéresse autant aux entités qu'à leur mode d'interaction, avec comme phrase emblématique que "Le tout est plus que la somme des parties". Les réseaux interviennent alors comme un outil de représentation : chaque nœud est une entité et les liens sont le support de leurs interactions ponctuelles, définissant alors un type de relation.

La simulation multi-agents est un paradigme de modélisation intéressant pour représenter les systèmes complexes : il permet une transcription "naturelle" des entités du système en agents. Une fois les règles d'interactions bien définies, le système simulé reproduit le comportement du système en question. Il est alors possible de tester des hypothèses en modifiant les conditions et/ou les paramètres.

Les réseaux peuvent alors encore intervenir, avec comme rôle de définir les interactions possibles entre les agents de la simulation : il s'agit alors de réseaux sociaux.

1.2 Contexte de la thèse

Cette thèse s'est effectuée dans le cadre du projet GenStar ([Drogoul, 2013](#)), avec un co-financement de la Région Occitanie et de l'Agence Nationale de la Recherche¹. Dans le domaine des simulations sociales à base d'agents, le réalisme des résultats demande des populations synthétiques cohérentes. Le projet vise à fournir un Framework (outils et méthodes) permettant la génération de populations synthétiques et réalistes, ayant vocations à être utilisées dans des simulations sociales à base d'agents.

Ces populations synthétiques sont ensuite utilisées dans des simulations, pouvant aider à la décision pour des systèmes socio-environnementaux, systèmes par nature complexes, constitués de nombreux éléments en interactions. Ces éléments sont des agents sociaux (de l'individu à l'institution), dont les comportements sont influencés par leurs attributs, leurs interactions avec les autres agents et avec l'environnement.

Dans ce contexte, afin de réaliser une modélisation réaliste d'un système complexe, les points importants sont :

- représenter de façon convenable les éléments constitutifs du système,
- la façon dont interagissent ces éléments,
- et enfin l'incorporation de données réelles ou réalistes dans le cadre de la reconstruction de données.

L'incorporation de données réelles, *data-driven approach*, est une approche relativement récente ([Hassan, 2009](#)). Elle permet de prendre en compte des données dont la disponibilité va en augmentant, en qualité et quantité, mais demande une architecture particulière pour les Frameworks. Le projet se propose de fournir pour ces nouveaux paradigmes une solution composée de 3 phases complémentaires.

- la génération de population synthétique,
- sa spatialisation,
- et la création du réseau social correspondant.

Les données réelles interviennent pour la création de populations ainsi que leur spatialisation.

Générer des populations synthétiques depuis des données empiriques. Les données à disposition sont la plupart du temps des données agrégées : re-

1. projet Genstar ANR-13-MONU-0006

censement de la population, enquêtes... Il va s'agir ici de désagréger ces données par l'usage de statistiques afin d'obtenir une instanciation cohérente d'une population d'agents, respectant la forme agrégée d'entrée.

Spatialisation Les simulations sont pour la plupart spatiales, le Framework inclut donc une phase de spatialisation de la population précédemment générée. La difficulté de cette étape réside dans le positionnement cohérent de la population sur la carte, avec des densités correspondant au profil des zones géographiques étudiées.

Génération d'un réseau social Afin de relier entre eux les individus de la population, une couche réseau social est à ajouter. C'est dans cette étape que la thèse s'inscrit. Le projet ne s'interdit pas d'intégrer plusieurs modèles de génération de réseaux. Contrairement aux étapes précédentes qui sont terminales, la création d'un réseau peut être poursuivie par un processus d'évolution en cas de perturbation. Cela pourrait par exemple correspondre au déménagement d'individus dans la population générée ou à des évolutions démographiques que l'on souhaite prendre en compte.

1.3 Motivations

Deux paradigmes de génération de réseau Les approches pour générer les réseaux peuvent être divisées en deux : celles qui abordent le réseau en tant qu'objet mathématique dont on ne considère que l'aspect topologique et celles qui le voient comme le résultat d'un processus de formation. Le terme réaliste prend alors un sens différent en fonction de l'approche utilisée. Est réaliste le réseau qui est généré par des comportements plausibles d'individus, considérés comme fondateurs pour la littérature en sciences sociales. De l'autre côté, un réseau réaliste sera celui possédant des propriétés topologiques conformes aux autres réseaux observés du même type.

Différents degrés de précision Ces deux approches peuvent ensuite être affinées selon un degré de précision. On retrouve la distinction entre les paradigmes de modélisation KISS et KIDS. L'un, Keep It Simple Stupid, propose une modélisation simple, stylisée, donnant des résultats qualitatifs, révélateurs de tendances. L'autre, Keep It Descriptive Simple, préfère une approche descriptive poussée, détaillée ([Edmonds and Moss, 2005](#)).

Les approches sociales L'avantage de l'approche sociale est de fournir un ensemble de règles compréhensibles et maîtrisables par l'utilisateur. Les mécanismes sont intelligibles et ont directement du sens. L'inconvénient est que le réseau obtenu n'est pas nécessairement en accord avec

ceux observés dans la réalité, et que les approches plus descriptives (KIDS) entraînent souvent un grand nombre de paramètres qui sont difficiles à configurer de manière rigoureuse.

Les approches mathématiques L'autre approche, mathématique et statistique, apporte des résultats plus réalistes sur le plan topologique. Les versions les plus abstraites génèrent des réseaux cohérents qualitativement, et les versions poussées obtiennent de très bons résultats sur de nombreux critères.

Néanmoins, chacun de ces extrêmes comporte des défauts non négligeables. Les topologies abstraites miment certaines propriétés considérées comme fondatrices pour les réseaux sociaux. Malgré la véracité de ces observations, ces éléments seuls sont loin de garantir des comportements de simulation fiables, même qualitativement.

À l'autre extrême, les modèles très poussés sont difficiles d'utilisation ou ont des résultats difficiles à interpréter et à intégrer dans le cadre d'une simulation sociale.

Un autre type d'approche Nous allons essayer de fournir une approche conciliant les avantages vus précédemment : interprétable, à la topologie quantitativement cohérente et facile d'utilisation. Nous nous positionnons à la jonction des mathématiques et des sciences sociales. L'aspect configuration poussée, inhérent à des résultats quantitativement viables, sera masqué par l'usage d'un algorithme génétique.

La méthode de génération est donc orientée processus, associée à une évaluation topologique des réseaux.

La problématique soulevée par la thèse peut être synthétisée sous la forme de deux questions principales :

- "Comment générer des réseaux réalistes ?"
- "Comment prendre en compte leur évolution ?"

La première question demande un modèle de génération de réseaux, la seconde la possibilité d'adapter le réseau à des changements au cours de la simulation. La façon dont nous avons choisi de répondre à ce sujet est de trouver un processus dynamique de création de réseaux. Cela permet, pour faire évoluer le réseau, de continuer à faire jouer la dynamique de création.

Nous souhaitons être capables de réaliser des réseaux quelque soit le domaine d'application de la simulation. Autrement dit, il faut pouvoir générer des réseaux avec aussi peu de connaissance métier que possible.

1.4 Problématique de la thèse

Il s'agira donc dans le cadre de cette thèse de définir des méthodes et des outils permettant au modélisateur multi-agents de générer des réseaux sociaux utilisables au sein de ses modèles. Les processus de création de réseaux proposés sont endogènes et individus-centrés et reposent sur l'application, par les entités de la simulation, de règles simples de construction et de suppression des relations. Ce type de modèle novateur au sein de la littérature en simulation sociale, présente un certain nombre de difficultés (atteignabilité des configurations, stabilité du système en particulier), qui ont amené à mettre en place des méthodes d'études particulières de ces réseaux.

La principale problématique de ce travail de thèse est de fournir une méthode générique de génération de réseau capable d'atteindre des réseaux aux topologies singulières. Les dynamiques de construction de réseaux peuvent rapidement devenir chaotique ou tendre vers des formes aléatoires. Il faut réussir à contraindre suffisamment les dynamiques pour éviter cet écueil, tout en gardant suffisamment de liberté pour pouvoir générer une multitude de réseau.

1.5 Plan du manuscrit

Dans la première partie, **l'état de l'art** présente les notions importantes associées à cette thèse : les systèmes complexes, les réseaux en tant que représentation et enfin les méthodes de génération de réseaux existantes.

En ce qui concerne le travail réalisé, l'approche plébiscitée est expérimentale. Nous avons créé un cadre de simulation permettant la génération d'une grande variété de réseaux, tout en permettant la reproduction des résultats de manière consistante. La deuxième partie décrit la partie **Framework**, se concentrant sur le développement incrémental de ce cadre.

Ensuite, les parties trois **-modèle I-** et quatre **-modèle II-** sont deux instantiations du Framework en modèles. Le deuxième modèle cherche à pallier des défauts du premier au prix d'une augmentation de la complexité et du contrainte du système. Nous finirons par une **conclusion** générale qui synthétisera la contribution en soulignant de manière critique ses intérêts et ses limites mais également ses potentialités.

Chapitre 2

Etat de l'art

Sommaire

2.1	Introduction	8
2.2	Systèmes complexes	9
2.2.1	Définition d'un système complexe	9
2.2.2	L'émergence dans un système complexe	10
2.2.3	Dynamique d'un système complexe : de la stabilité au chaos	12
2.2.4	Auto organisation des entités constituantes	14
2.2.5	La modélisation de systèmes complexes	15
2.2.6	Conclusion sur les systèmes complexes	18
2.3	Des graphes aux réseaux sociaux	19
2.3.1	Le réseau en tant que représentation d'un système	19
2.3.2	Les réseaux sociaux : un cas particulier de réseau	23
2.3.3	Conclusion sur les réseaux sociaux	29
2.4	Génération de réseaux sociaux	30
2.4.1	Introduction aux paradigmes de génération	30
2.4.2	Contexte	30
2.4.3	Familles de méthodes de génération de réseaux	32
2.4.4	Les méthodes de génération de réseaux sociaux	35
2.4.5	Conclusion sur la génération de réseaux sociaux	57

2.1 Introduction

L'objectif de la thèse est de générer et faire évoluer des réseaux sociaux, représentant des individus et leurs relations. La méthode de génération de graphe, objet mathématique associé aux réseaux, se base sur une modélisation à base d'agents.

L'état de l'art cherche à apporter les éléments nécessaires pour permettre la compréhension du travail réalisé. Dans une première section, nous définirons les *systèmes complexes* et leur modélisation ; puis nous en détaillerons la représentation sous forme de graphe, les *réseaux*. Enfin, nous détaillerons les modèles de *génération de réseaux* les plus utilisés et les plus proches du travail réalisé.

2.2 Systèmes complexes

*Dans cette section, l'attention va être portée sur les systèmes complexes. Après une définition des principaux concepts, leurs propriétés importantes, émergence et auto-organisation, seront vues en détails. La section se termine par les deux paradigmes de modélisation les plus utilisés pour représenter ces systèmes. Les systèmes complexes ont été définis de nombreuses manières. Toutes ces définitions ont en commun d'intégrer les notions d'entités, d'interactions entre ces entités, et une difficulté pour l'observateur de prédire le comportement global du système en se basant sur ses éléments constitutifs. Les entités composant le système peuvent avoir tendance à s'auto-organiser localement. Des interactions entre les entités résultent l'émergence d'un phénomène. La notion d'émergence implique une difficulté à décrire le résultat à partir des simples comportements des éléments constitutifs. Ce résultat est observable après changement d'échelle : l'observateur observe l'émergence d'un phénomène **macroscopique** à partir des interactions **microscopiques**.*

Nous nous intéressons à ce domaine pour deux raisons. Tout d'abord, la formation de réseaux découle d'interactions entre des entités dans un environnement donné, sujet d'études des systèmes complexes. Ensuite, l'outil que nous utilisons afin de générer des réseaux est la simulation multi-agents. Cette dernière est une représentation "naturelle" d'un système complexe. Nous sommes en particulier attentif aux dynamiques menées dans les systèmes complexes. il s'agit de l'un des verrous particuliers liés à notre travail : ne pas arriver dans des dynamiques chaotiques entraînant une génération de réseau aléatoire, non contrôlée.

2.2.1 Définition d'un système complexe

Il existe dans la littérature de nombreuses définitions d'un système complexe. Les définitions données par différents auteurs vont ici être mises en relation afin d'en faire ressortir les éléments communs.

(Sayama, 2015) définit les systèmes complexes comme un ensemble de composants qui interagissent les uns avec les autres, de manière non linéaire. Un tel système peut évoluer par auto-organisation, de sorte qu'il ne soit ni complètement aléatoire ni complètement organisé. Cela permet l'apparition de comportements émergents, visibles aux échelles macroscopiques.

Pour (Kirer and Cirpici, 2016), un système est qualifié de complexe quand il est composé d'un grand nombre d'éléments. Ces éléments interagissent et s'organisent spontanément en structures élaborées, changeantes dans le temps.

(Weaver, 1947) quant à lui, définit l'étude des systèmes complexes comme s'insérant entre les deux vues traditionnelles d'un système :

- Aux composants parfaitement indépendants, comme les gaz parfaits,
- Aux composants parfaitement couplés, comme l'étude de corps rigides.

Ainsi, l'approche statistique usitée dans le premier cas, ou l'approche par un modèle à quelques paramètres dans le deuxième cas, ne sont pas envisageables pour décrire ou prédire le comportement d'un système complexe.

Ces définitions sont assez générales mais couvrent l'idée d'entités qui vont interagir ensemble, et s'organiser localement. Par organisation est entendue le fait qu'une entité privilégie certaines connections. Faisant suite aux actions des composants du système, des structures vont apparaître. Ces dernières sont observables à une échelle supérieure à celles des entités.

Si l'on s'intéresse aux systèmes complexes sociaux, on peut trouver une définition de la complexité sociale, donnée par (Heppenstall et al., 2012) :

Social complexity [...] is a condition whereby social behavior cannot be understood simply as a scaled-up replication of the behavior of the individuals comprising the society. The interplay of social processes as an outcome of socially embedded individuals gives rise to the social behavior, which [...] cannot be forecast on the basis of individuals' characteristics and predictions alone.

Une représentation des différents sujets d'études du domaine des systèmes complexes est donnée par (Sayama, 2015) en figure 2.1.

Malgré cette diversité d'approche des systèmes complexes, certains éléments restent communs.

Un système complexe est donc constitué d'entités qui interagissent de manière organisée. De ces interactions émergent des organisations particulières des composants. Une particularité de ces systèmes est que ce résultat macroscopique n'est pas prédictible depuis la connaissance seule de l'ensemble des éléments constituants et de leurs règles d'interactions.

Nous allons maintenant détailler les notions importantes propres aux systèmes complexes.

2.2.2 L'émergence dans un système complexe

Suite à ces définitions, nous allons détailler les principales notions qui s'y rattachent.

L'émergence qualifie un résultat que l'on observe dans un système à l'échelle macroscopique, lorsqu'il est la conséquence d'interactions des entités constitutives du système. Les propriétés émergentes ne sont pas possédées par les entités de bas niveau constituant le système (Aziz-Alaoui and Bertelle, 2009). Si on transpose l'émergence à la société humaine, Geller et Moss nous proposent la définition suivante :

"Un certain nombre de phénomènes sociaux ont des caractéristiques qui peuvent être qualifiées de complexes, situations dans lesquelles les actions des agents et leurs interactions sociales se combinent pour créer des résultats macroscopiques, qui ne peuvent être prédits par la seule connaissance des ces actions et interactions."

[...] a variety of social phenomena exhibit characteristics which can be labeled as complex, a condition in which agent behavior and social interaction combine to generate macrolevel outcomes that could not

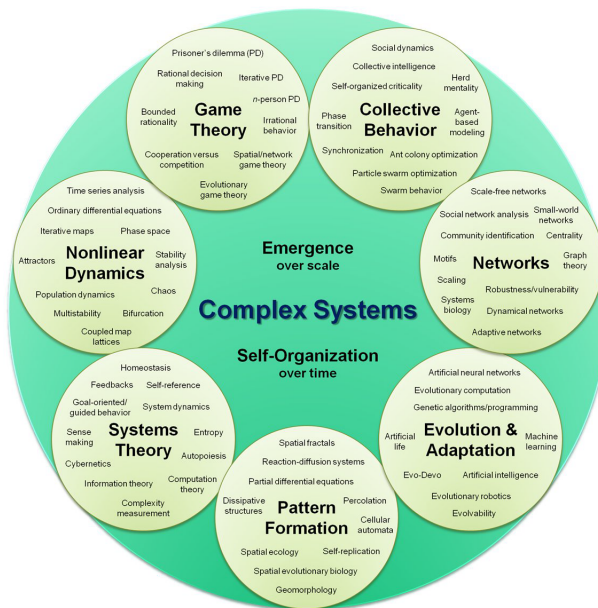


Figure 2.1: La recherche sur les systèmes complexes divisée en sept sections. Les trois sections sur la gauche (Dynamiques non-linéaires, théorie des systèmes et théorie des jeux) sont les premiers éléments historiques constituant des recherches sur les systèmes complexes. Les quatre autres sont des sujets de recherche plus récents. Illustration issue de (Sayama, 2015)

be predicted from knowledge of the behavior and nature of interactions alone” (Geller and Moss, 2008).

L'émergence qualifie une observation d'une organisation macroscopique conséquence des interactions entre entités microscopiques. D'un point de vue plus "philosophique", le phénomène émergent est vu par les yeux d'un observateur, suivant une métrique ou un objectif qui lui est propre.

2.2.3 Dynamique d'un système complexe : de la stabilité au chaos

L'émergence est issue du processus d'évolution du système étudié, de sa dynamique temporelle. Elle est menée par la succession d'interactions entre les entités (Calvez, 2009). Afin de représenter un système et son évolution, on peut se servir d'un espace de phase. Il s'agit d'un repère où chaque variable du système est représentée par une dimension. Un point dans l'espace représente une combinaison de valeurs des variables à un instant donné. La succession de ces points indiquent les changements que le système subit dans le temps : on parle alors de trajectoire (Collet, 2015). Cette représentation est intuitive pour des espaces aux dimensions inférieures à quatre (figure 2.2).

La notion de *bassin d'attraction* est utile lorsqu'on s'intéresse aux états finaux atteignables par un système. Un bassin d'attraction qualifie un sous-ensemble de l'espace de phase, dans lequel la trajectoire d'un système sera attirée. Une fois l'état du système "pris" dans le bassin, la trajectoire se dirige vers son centre, et n'en sortira pas sans événement exogène. On peut distinguer les états finaux atteignables suivants :

- Etat final stationnaire - Le système se place sur point fixe et y demeure.
Ex : mouvement d'un pendule avec frottement qui s'immobilisera.
- Etat final cyclique - la trajectoire du système boucle dans un sous-espace donné. *Ex : satellite lancé depuis la Terre et mis en orbite.*

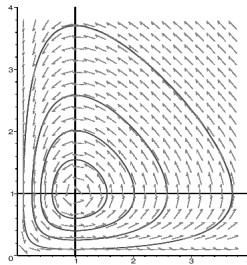


Figure 2.2: Représentation dans un espace d'état d'un système possédant deux variables, leurs valeurs correspondant aux axes. Les vecteurs indiquent le sens de progression du système dans le temps.

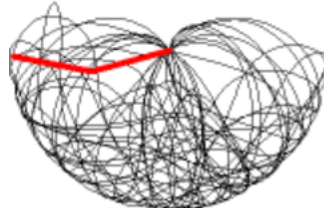


Figure 2.3: Pendule à deux bras, dont le mouvement du bout du second bras est suivi. En rouge, les bras du pendule, en noir, la trajectoire du bout du second bras. Malgré la simplicité du dispositif, leur mouvement est très difficile à prédire, et des changements infimes dans les paramètres initiaux entraînent des résultats très différents. (Source : wikipedia-Chaos Theory).

- Etat final chaotique - aucune stabilité n'est atteinte par le système. *Ex : système météorologique terrestre.*

Un système peut posséder aucun ou plusieurs attracteurs. La trajectoire d'un système et sa capture dans un bassin d'attraction dépendent de la nature et l'ordre des actions entreprises par les entités. En effet, les premières interactions définissent la trajectoire du système dans l'espace des réalisations possibles, et sont donc souvent déterminantes.

Les dynamiques chaotiques

La théorie du chaos étudie l'hyper-sensibilité aux conditions initiales d'un système dynamique. Comme l'explique (Baranger, 2000), la plupart des systèmes réels sont chaotiques - proportion mal représentée par les sujets d'études de la physique jusqu'au 20ème siècle. D'infimes différences dans les conditions initiales se retrouvent exponentiellement décuplées au cours de l'évolution du système. Le pendule à deux branches en est un exemple simple 2.3. Malgré le déterminisme du système étudié, il est extrêmement difficile de prédire l'aspect final et les évolutions du système autrement que par la simulation ou l'expérimentation.

Une métaphore répandue est "l'effet papillon" : le battement d'aile d'un papillon pourrait déclencher un ouragan à l'autre bout du monde. On doit cette image à Lorenz, météorologue, qui a grandement participé à développer l'idée de sensibilité d'un système aux conditions initiales. Cet exemple est aussi intéressant pour illustrer le phénomène d'émergence : une action d'une entité amène un résultat visible à un niveau d'ordre supérieur, après une cascade d'actions-conséquences. En plus de la sensibilité aux paramètres initiaux, l'évolution d'un système complexe peut être très sensible à l'ordre dans lequel sont réalisées les premières actions. Ces dernières vont déterminer la trajectoire initiale du système dans l'espace de phase, et potentiellement le bassin d'attraction vers lequel il se dirige (Easley and Kleinberg, 2010) : il s'agit de la notion de dépendance de sentier *path dependence*.

les états stables

Lorsque, après des actions prises, le système arrive dans un état qui se modifiera peu par la suite, on considère alors qu'il a atteint un état stable.

La notion de stabilité peut avoir plusieurs définitions dépendant du domaine scientifique d'étude. Pour un écosystème, cette notion est proche de celle de résilience, capacité qu'a un système à retrouver un état stable après perturbation.

Une image proposée par (Scheffer et al., 2001) permet de lier les deux notions intuitivement : La résilience d'un écosystème est représentée par une balle située dans un bassin, la forme du bassin qualifiant la qualité de la résilience. Un bassin plat induit une résilience faible, tandis qu'un bassin profond entraîne une forte résilience. Un système stable occupera la partie la plus basse du bassin. Un système à haute résilience peut occuper plusieurs points dans ce bassin tout en y restant stable.

"The resilience of ecosystems has been depicted as a ball residing in a basin (Scheffer et al. 2001), where the shape of the basin discerns the extent of resilience. A flat basin indicates low resilience, whereas a cup-shaped basin indicates higher resilience. A system with high stability would reside at the lowest point of the basin, whereas a system with high resilience may occupy various points of the basin."

Même dans des systèmes aux éléments simples, la dynamique temporelle peut être chaotique et ne jamais atteindre d'état stable, stationnaire (Frontier, 2014). L'une des raisons pour qu'un système parvienne à atteindre un état stable est que les entités se soient organisées entre elles. Cela apporte une résilience s'opposant aux perturbations amenées par l'environnement.

Nous allons nous intéresser au phénomène d'auto-organisation qui peut survenir entre les composants d'un système complexe.

2.2.4 Auto organisation des entités constituantes

Un autre élément que l'on retrouve dans la définition des systèmes complexes est l'organisation locale, décentralisée et non aléatoire des entités constituantes. Les entités agissent, organisent leurs interactions localement, sans supervision. Individuellement, chacune d'entre elles stabilise sa situation, permettant au système d'être à la fois "organisé et organisant" (Moigne, 1977).

L'auto-organisation induit un agencement particulier des entités. Ces dernières possèdent des attributs qui leurs sont propres et qui les singularisent, influençant leur façon d'interagir. La conséquence est que l'on obtient des poches de comportements fonction de la répartition non homogène des entités dans l'espace des attributs (pouvant être aussi simple que la position spatiale). L'auto organisation est ici vue comme amenant une amélioration d'un système au sens fonctionnel : en fonction d'un objectif, ou suivant une métrique d'entropie - la stabilité. Les êtres vivants tendent par exemple à s'organiser en communautés.

L'action locale pour les individus est de se lier aux personnes déjà validées par leurs pairs, la conséquence globale est la création d'ensemble dense de relations, les communautés (Giorgas, 2000).

Les entités, de par leur homogénéité de comportements ou d'attributs, s'organisent en créant des interactions privilégiées. Cela peut amener une résilience aux perturbations pour le système dans son ensemble.

2.2.5 La modélisation de systèmes complexes

Les sous-sections précédentes ont permis de décrire les propriétés importantes dans le fonctionnement d'un système complexe. Sa théorisation puis modélisation vont s'atteler à trouver une représentation plausible de la réalité, dépendant de l'utilisation prévue, qu'il s'agisse de :

- Prédire l'évolution du système dans l'ensemble ou de ses variables
- Tester des hypothèses afin d'observer l'adaptation du système
- Reproduire des propriétés statistiques

(Heath et al., 2009) voit une gradation dans les utilisations : décrire, expliquer, prédire. Le modélisateur part de sa vision du système "réel", composé d'une quantité d'éléments parfois indénombrables et/ou aux interactions à grande complexité. Avant de pouvoir étudier le système ou des hypothèses le concernant, il va être nécessaire de réduire cette complexité, le nombre d'éléments et leurs formes d'interactions, afin d'en permettre l'appréhension et la modélisation (Ashby, 1970) (Stanislaw, 1986).

(Stanislaw, 1986) décrit différentes étapes pour la modélisation d'un système. La première étape est d'établir une théorie sur l'objet que l'on souhaite représenter. C'est lors de ce processus que des informations seront perdues (Ashby, 1970), les théories étant des abstractions de la réalité. La difficulté consiste, malgré cela, à rester le plus proche possible de l'objet réel. Le contexte de l'usage voulu pour le modèle définit la dégradation de qualité acceptable.

Autrement dit, les étapes de modélisation impliquent de faire des choix sur la façon dont l'abstraction est réalisée. Ces choix vont influencer les utilisations possibles pour un modèle spécifique. Par conséquent, un système modélisé ne sera jamais une représentation parfaite (Bonabeau, 2002). Les choix de modélisation dépendent du but que l'on poursuit par ce procédé, de l'usage ultérieur du modèle.

On peut citer d'autres utilisations, telles que les jeux de rôle ou scénarios explicatifs, de négociation ou ludo-éducative. Elles sont abordées dans la partie modélisation à base d'agents.

Méthodes de modélisation

les premières modélisations scientifiques de phénomènes réels se sont faites sous la forme d'équations mathématiques. Des outils tels que les intégrales ou

les exponentielles ont permis de se rapprocher de notions telles que la continuité ou les croissances non linéaires. La complexité des outils allant de paire avec celle des phénomènes représentés.

L'approche basée sur des équations recherche des formes mathématiques permettant d'obtenir une description du système au niveau global, macroscopique. Les différents types d'entités sont traités comme des éléments homogènes, leurs évolutions étant représentées par une équation. L'ensemble des équations définit alors la dynamique du système. Au contraire, la modélisation basée sur des entités formalise les comportements et interactions de ces premières. De leur simulation résulte alors la dynamique du système.

Nous allons voir plus en détail l'approche basée sur des entités.

Modélisation à base d'agents La modélisation à base d'agents (ABM - Agent-Based Modelling) est un paradigme de représentation des systèmes complexes. L'attention est portée sur la spécification des composants constitutifs et de leurs interactions. Cette forme de modélisation ne nécessite pas une spécification du système dans sa globalité, contrairement à l'approche basée sur les équations. La modélisation basée sur des agents est une approche ascendante (*bottom-top*). Elle peut être qualifiée de naturelle : il est possible de faire correspondre à chaque entité du système à modéliser un agent dans la simulation (Parunak et al., 1998). Cela entraîne les possibilités suivantes (Bonabeau, 2002) :

- prendre facilement en compte les attributs des agents, entre autre la spatialisation
- doter chaque agent de règles de décision, ou d'un système d'apprentissage
- intégrer des données au niveau microscopique ou macroscopique. L'aspect macro peut par exemple correspondre à l'introduction de politiques publiques, ou autres facteurs exogènes aux agents (Chapuis, 2016)

On retrouve dans les dynamiques de la simulation orientée agents les mêmes mécanismes qui font partie intégrante d'un système complexe réel : auto-organisation des entités et émergence d'observables au niveau global (Edmonds, 2003). Le but est souvent de faire correspondre les dynamiques de la simulation avec celle du système simulé (Kirer and Cirpici, 2016). Ce type de modélisation est donc particulièrement indiqué dans les domaines possédant une forte hétérogénéité des entités constitutives - sciences sociales, système géographique, marché financier...

Un autre avantage est la possibilité de transposer directement le système à modéliser, en en représentant chaque élément. La compréhension du modèle est alors simplifiée. Cela permet d'utiliser la modélisation comme un support explicatif et participatif : (Barreteau et al., 2001) et (Guyot and Drogoul, 2005), dans une optique "jeu de rôle", remplacent certains agents par des humains ; (Le Page, 2017) l'utilisent avec des plateformes explicatives permettant aux participants de tester des règles de gestion et d'en voir les conséquences. De plus, le découplage des entités et interactions est utile lorsque plusieurs domaines scientifiques travaillent de concert. Par exemple, MAELIA (Gaudou et al., 2014) est un projet réunissant géographes, agronomes, hydrologues, et informaticiens

pour un modèle agents de gestion de l'eau dans le bassin versant de la Garonne. Chacun peut développer son module avant d'en réunir les éléments dans la simulation.

Avantages et inconvénients Le fait que chacune des entités du système réel soit représentée permet de prendre en compte leurs spécificités : de leurs attributs à leurs relations, en passant par leur éventuel modèle de prise de décision. Il est donc possible d'intégrer des données au niveau microscopique - sur les entités -, ainsi qu'au niveau macroscopique - en tant que contraintes environnementales.

Les principaux inconvénients de cette méthode de modélisation concernent les difficultés de validation et de configuration. Chaque agent, ainsi que ses interactions, peuvent être le jeu de plusieurs paramètres, à tel point que les modèles résultants requièrent des heuristiques pour l'exploration des paramètres initiaux (Calvez, 2009).

La validation est aussi un point complexe. Des données ne sont pas toujours disponibles pour vérifier les résultats de simulation. Obtenir des résultats conformes au système simulé permet de conclure que le modèle étudié est un bon candidat, mais ne le certifie pas pour autant. Concernant les ressources informatiques nécessaires à la simulation, les systèmes multi-agents peuvent rapidement être coûteux, chaque entité étant représentée. Cette limitation est néanmoins repoussée par l'avancée de la taille des mémoires et des puissances de calcul.

Un autre problème vient de l'aspect stochastique de la simulation. Deux simulations aux paramètres initiaux identiques peuvent donner des résultats différents. Cela peut être dû à une sensibilité du système aux conditions initiales (chaotique), à l'influence forte des premières actions entreprises, ou simplement à la nature aléatoire du système.

La modélisation des systèmes complexes demande avant tout une théorisation du système modélisé, au prix d'un choix sur les éléments considérés comme importants.

L'approche basée sur des agents permet une hétérogénéité, pouvant prendre en compte attributs ou règles de décision. Les modèles sont plus faciles à réaliser, mais plus difficiles à valider, et nécessitent plus de ressources informatiques pour être simulés.

2.2.6 Conclusion sur les systèmes complexes

Nous définissons un système complexe par ses *entités* constituantes et leurs *interactions*. Des organisations particulières entre les entités et des interactions préférentielles vont survenir au cours de l'évolution du système. Il s'agit d'*auto-organisation*. Cela va souvent amener des changements observables à un ordre supérieur : on parle alors d'*émergence*. Les dynamiques d'un système peuvent entre autre l'amener à un état de *stabilité*, ou à l'autre extrême dans un état de fluctuations importantes permanentes, *chaotiques*. Afin de pouvoir étudier ces systèmes ou d'y tester des hypothèses, il va être nécessaire de se les représenter. La modélisation d'un système complexe passe d'abord par une phase de théorisation, processus visant à abstraire le réel. Deux paradigmes principaux de modélisation s'appliquent alors pour représenter une théorie : la modélisation à *base d'équations* et celle à *base d'agents*.

La modélisation à base d'équations est une approche descendante : Il faut être capable de décrire le système dans son ensemble. La modélisation à base d'agents, au contraire, propose une approche ascendante, où la définition des entités et de leurs interactions suffisent pour définir le système. La difficulté est alors reportée sur l'étape de *validation*.

2.3 Des graphes aux réseaux sociaux

Une représentation possible des entités d'un système et de leurs interactions (ou relations) est le réseau. Nous allons le décrire en tant qu'objet mathématique, alors appelé graphe. On peut alors en calculer des propriétés ; les réseaux sociaux ont, de par leurs processus et contraintes de formation, des valeurs spécifiques pour certaines de ces propriétés.

Nous observons donc les propriétés spécifiques aux réseaux sociaux, qui nous serviront d'objectif à reproduire lors de notre processus de génération de réseau sociaux. Les données sont abondantes pour certains types de réseaux informatisés. C'est moins le cas lorsque le support d'interaction ne l'est pas, comme pour les réseaux sociaux de la vie réelle.

2.3.1 Le réseau en tant que représentation d'un système

Un réseau est une représentation d'un système : entités qui le constituent et leurs interactions (ou relations) (Menezes and Roth, 2014), (Newman, 2003), (Kirer and Cirpici, 2016).

Chaque nœud représente une entité, et les arcs leurs relations, support des interactions. Les types de réseaux - sociaux, biologiques, infrastructures ...- résultent de processus de formation et de contraintes d'environnements particuliers. Ces particularités sont qualifiables par des valeurs de propriétés particulières.

On peut discerner deux niveaux d'éléments impactés par le type de système étudié :

- *Les attributs du graphe* Il s'agit d'éléments qualifiant le type mathématique du graphe : Avoir un graphe dirigé, des liens pondérés... Les liens entre les individus sur Facebook sont par exemple toujours réciproques. Les liens du graphe les représentant sont donc non orientés.
- *Les propriétés sur le graphe* : Il s'agit des valeurs, observées ou calculées, spécifiant un graphe en particulier. On y retrouve le nombre de liens, la moyenne des poids sur ces liens, la taille moyenne des chemins entre deux nœuds pris au hasard..

Les réseaux sont particulièrement utilisés dans les disciplines suivantes :

Biologie - réseaux d'interactions de protéines et molécules (Farkas et al., 2003), réseaux trophiques (*foodweb*) (Jordano et al., 2002).

Physique - Interactions entre particules électriquement chargées - Modèle d'Ising (Herrero, 2002)), étude de vibration (Estrada et al., 2012).

Sciences sociales - relations entre individus - Amitié (Rapoport and Horvath, 2007), mariage (Padgett and Ansell, 1993), relations d'affaires entre professionnels (Galaskiewicz, 1985).

Géographie - Réseaux de transport (Rodrigue et al., 2017), d'aéroports (Li and Cai, 2004), de routes (Makse et al., 1998), trajets et flux (Levinson and Wu, 2005).

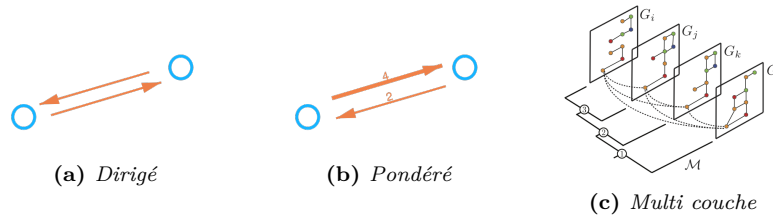


Figure 2.4: Types de graphe possibles (attributs).

La partie suivante s'intéresse aux réseaux en tant qu'objets mathématiques, alors appelé graphes.

Représentation mathématique

Un graphe $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ est composé d'un ensemble de sommets (*node* \mathcal{N}), ainsi que d'arcs reliant ces sommets (*edge* \mathcal{E}). Nous utiliserons de façon équivalente les termes nœud // sommet, ainsi que lien // arc.

Les notions de graphe et de réseau sont assez similaires. Le graphe est l'objet mathématique représentant un réseau, qui peut être (fig 2.4) :

Dirigé Un réseau dirigé prend en compte le sens des arcs entre deux nœuds. Chaque paire de nœuds peut donc avoir deux arcs entre eux.

Pondéré un réseau pondéré est un réseau dont chaque arc possède une valeur associée. Cette valeur numérique est appelée poids de l'arc.

Multi couches Un réseau multi couche (ou multiplexe) est un réseau qui peut avoir plusieurs types d'arcs entre ses sommets. On appelle couche un sous-réseau porteur d'un certain type d'arcs entre les nœuds.

Dynamique Un réseau dynamique est un réseau qui évolue, temporellement. Il peut s'agir de changement structurel, ajout ou retrait de liens ou de nœuds. Il peut aussi être question de changements de propriétés des nœuds ou des liens.

Propriétés

En tant qu'objet mathématique, des propriétés peuvent être calculées sur un graphe, permettant de le caractériser et d'en extraire des informations agrégées. Ces propriétés peuvent être classifiées suivant qu'elles concernent le graphe de manière "locale", au niveau du nœud, ou global, sur l'ensemble du graphe.

Localement,

Degré Le degré d'un nœud est le nombre de liens qu'il possède. Dans le cas d'un graphe dirigé, la distinction est faite entre le degré des liens entrants et celui des liens sortants.

Clustering Le coefficient de clustering d'un nœud est un indice indiquant à quel point ses voisins sont interconnectés. Il s'agit du rapport entre le nombre de connections effectives entre voisins divisé par le nombre de connections totales possibles entre ces derniers. Plus précisément, le coefficient de clustering c_i pour un nœud i , possédant un nombre d_i de voisins, est défini par le nombre de liens entre les voisins e_i divisé par le nombre maximum de liens possibles entre ces voisins $d_i(d_i - 1)/2$ (on se place ici dans le cas d'un graphe non dirigé)

$$c_i = \frac{2e_i}{d_i(d_i - 1)}.$$

Ce facteur varie de 0 à 1. 0 correspond à la situation où aucune des acointances du nœud i ne possède de liens entre elles. Un facteur à 1 correspond au cas où tous les voisins du nœud i sont aussi connectés entre eux. En terme de topologie, on peut se représenter le clustering comme le nombre de triangles fermés parmi les voisins d'un nœud i (fig 2.5b).

De manière globale

Densité La densité d'un réseau est le rapport entre le nombre de liens dans le réseau divisé par le nombre maximum de liens possibles dans le réseau. Si le réseau est dirigé, et que N est le nombre de nœuds, il existe ainsi $N(N-1)$ connections possibles (deux fois moins pour un réseau non dirigé).

$$\frac{|E|}{n(n-1)}$$

Distribution de degré La distribution de degré des nœuds d'un réseau compte le nombre de nœuds possédant un degré en particulier, et ceci pour tous les degrés. Elle est souvent représentée sous la forme d'un tableau (fig 2.5a).

Coefficient de clustering moyen Le coefficient de clustering moyen d'un graphe est défini comme la moyenne des coefficients de clustering de tous les nœuds du graphe. Ce coefficient mesure donc à quel point les nœuds du réseau sont interconnectés entre eux. Une notion proche, mais définie de façon plus informelle, est la communauté. On cherche une partition d'un graphe telle que les structures appelées communautés contiennent des nœuds plus connectés à l'intérieur de la communauté qu'à l'extérieur. Elle donne donc la probabilité pour que deux nœuds j et k , connectés au même nœud i , soient connectés entre eux [Vázquez \(2003\)](#).

Assortativité L'assortativité des nœuds d'un réseau mesure la similarité entre entités connectées. On s'intéresse, pour des nœuds sans attribut particulier associé, à une similarité en terme de degré. Il existe plusieurs mesures pour ce facteur. Nous prenons la définition de ([Newman, 2002](#)), variant de -1 pour les réseaux aux nœuds complètement mélangés, à 1

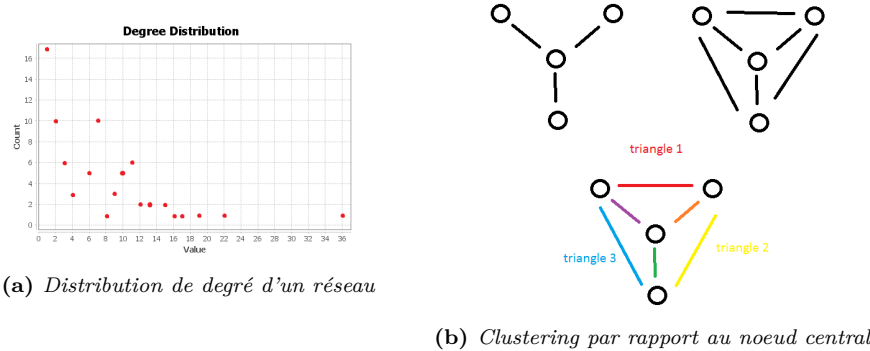


Figure 2.5: Illustration de certaines propriétés sur les réseaux. Distribution de degré par l'outil Gephi

pour des liens seulement entre des nœuds identiques en degré. On qualifie donc un réseau de non assortatif (dissassortatif) lorsque sa valeur d'assortativité est négative. (L'assortativité est en fait un coefficient de Pearson pour les couples de nœuds connectés).

Chemin Un chemin entre deux nœuds A et B est l'ensemble des arcs et nœuds qu'il faut parcourir pour aller de A à B. On s'intéresse souvent aux chemins les plus courts. On définit alors le **diamètre** comme le maximum des plus courts chemins. On va plus souvent s'intéresser au **diamètre effectif**, prenant la taille de chemin moyen qu'il faut à un nœud A pour atteindre 90% des nœuds du réseau. Cela permet d'éviter de comptabiliser les situations extrêmes données par les chaînes les plus longues (Kunegis, 2011). On peut aussi considérer la **Taille moyenne de chemin** (*Average Path Length, APL*) : il s'agit de la moyenne de la taille des plus courts chemins entre tous les nœuds du graphe.

Centralités Les indicateurs de centralité capturent l'importance d'un nœud dans un graphe. La centralité d'intermédiarité (*betweenness*) s'intéresse au nombre de chemins les plus courts passant par un nœud donné (Boccalletti et al., 2006). Cette mesure peut aussi être appelée la charge, terme explicite lorsque l'on s'intéresse à des flux, par exemple dans un réseau de pipeline.

Les propriétés globales des réseaux sont souvent des informations agrégées : On procède donc à une perte d'information afin de pouvoir fournir une représentation plus compacte. Le choix sur des propriétés définies comme représentatives est primordial. Par exemple, la densité nous informe sur la proportion de liens présents, sans nous donner la répartition de ces derniers entre les nœuds. Il faut s'intéresser à une mesure d'ordre supérieur, comme la distribution de degré, pour avoir plus d'informations sur la disposition du réseau amenant une telle densité (figure 2.6).

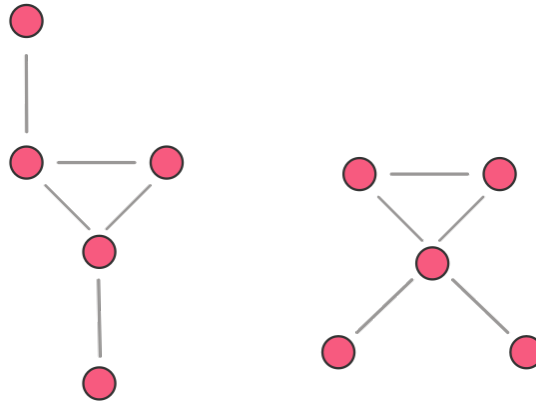


Figure 2.6: Exemple de deux réseaux possédant certaines propriétés en commun : Même nombre de nœuds, de liens, de taille de chemin moyen, de coefficient de clustering... mais avec certaines autres propriétés différentes, comme la distribution de degrés des nœuds. Le choix des propriétés pour définir le profil d'un réseau est donc essentiel.

Particularité en terme de propriétés et d'attributs des types de réseaux

les valeurs des propriétés sont différentes en fonction du type de système représenté par le réseau. Cela est dû aux processus ou contraintes de formation. A titre d'exemple, mettons que l'on choisisse de représenter un cours d'eau à l'aide d'un graphe, les nœuds étant les intersections et les arcs les bras les reliant (Rodríguez-Iturbe and Rinaldo, 2001), le graphe obtenu sera donc dirigé, sans cycle et les arcs potentiellement pondérés par le débit d'eau.

Les réseaux sociaux ont eux aussi des attributs et propriétés particuliers, avec notamment des propriétés prenant valeur sur des plages caractéristiques. C'est ce que nous allons voir dans la partie suivante.

Les réseaux permettent de modéliser les entités et relations entre ces dernières. L'objet mathématique est un graphe, et ce graphe a des caractéristiques dépendant du système représenté (orientation, poids sur les liens..).

2.3.2 Les réseaux sociaux : un cas particulier de réseau

Un réseau social est un ensemble d'individus interagissant entre eux de manière organisée (Newman, 2003). En fonction du système étudié, ses liens peuvent représenter différentes choses.

Les réseaux sociaux sont utilisés en particulier dans les simulations qui se basent sur les interactions possibles entre des agents. Pour citer certains domaines d'utilisation, on retrouve la simulation sociale, l'économie ou encore la

géographie. Dans le cas spécifique de la simulation sociale, on peut nommer les problématiques de :

- Formation d'opinion (Grabowski, 2009)
- Ségrégation (Edmonds, 2003)
- Recherche de conseil (Wong, 2008) et réseau d'influence (Lazega et al., 2012)
- Échange d'information pour la résolution de problème (Lazer and Friedman, 2007)
- Choix d'une activité (Ronald et al., 2012)
- Mariage (Billari et al., 2007)
- Diffusion de maladie (Ben-Zion et al., 2010), de savoir et d'innovations (Abrahamson and Rosenkopf, 1997), (Manzo et al., 2018).

Quel que soit le domaine, la structure du réseau sert de support pour les interactions entre agents (Holzhauer, 2015).

Particularités des réseaux sociaux

La formation d'un réseau social est conséquence de l'environnement et des comportements des individus - dépendant donc des choix des acteurs et du milieu dans lequel ils évoluent. Malgré la diversité des supports d'interactions sociales, informatisés ou non, certains archétypes de comportements se retrouvent.

- *Homophilie* - L'homophilie désigne la tendance qu'ont les individus à préférer interagir avec ceux qui leur ressemblent (Scholz, 2012) ; cela peut en particulier s'appliquer à des similarités d'opinions (McPherson et al., 2001).
- *Distribution de Pareto* - La répartition des ressources (argent, influence, célébrité) est non aléatoire et donc non uniforme. La distribution de Pareto désigne une répartition pour laquelle une large partie des ressources est détenue par une petite partie des entités. Elle a initialement été observée dans le domaine social, au niveau de la distribution de richesse dans la société (Pareto, 1965).
- *Nombre de connaissances limité* - Le nombre de connaissances d'un individu est limité, et de surcroît faible comparé au nombre de connections possibles (Hamill and Gilbert, 2009). Cela est entre autre dû au coût, en terme de temps ou autres ressources y compris cognitives pour maintenir les relations (Dunbar, 1992).
- *Inégalité de la qualité des relations* - En plus d'un nombre restreint de connaissances par individu, il est possible de distinguer l'importance de leurs relations. (Boissevain, 1975) en définissent cinq gradations, du plus fort au plus faible. Les recherches en sciences sociales se sont surtout concentrées sur les liens les plus forts (Hamill, 2010), dont le nombre varie de 3 - cercle de confidents (Marsden, 1987) à 40 - connaissances "noyau" et "significatives" (Boase et al., 2006). Ces liens forts s'opposent aux liens faibles, à la qualité sociale moindre mais bien plus nombreux (de l'ordre de plusieurs centaines).

Ces comportements singuliers se répercutent sur les propriétés du graphe

représentant le système. On juge de ces particularités en comparant les propriétés à celles d'un réseau aléatoire ou aux réseaux d'autres domaines.

- L'homophilie peut impliquer une assortativité de degré positive des nœuds. Les réseaux aléatoires ont une assortativité nulle, les réseaux d'autres domaines (biologiques, technologiques, informations) souvent négative (Newman, 2003).
- Pour des raisons sociales diverses, un individu va avoir tendance à découvrir de nouvelles connaissances par l'intermédiaire de ses amis. De proche en proche, on obtient des groupes d'individus plus fortement connectés entre eux qu'au reste du réseau. Les propriétés du réseau associées à cet élément sont le coefficient de clustering et les communautés.
- Le nombre limité de connaissances pour un individu entraîne une densité de réseau faible (Leskovec and Krevl, 2014) :
 - facebook : 0.541 %
 - twitter : 0.026 %
 - "réseau" du monde : ordre de $1e^{-10}$ %, notre nombre de connaissances par rapport à la population terrestre.
- La distribution de Pareto se retrouve dans la distribution de degrés des nœuds. Peu d'entre eux possèdent beaucoup de connections aux autres ; à l'inverse un grand nombre de nœuds sont faiblement connectés. Cela entraîne des structures en étoile dans le réseau (Kumar et al., 2006). Plus précisément, la distribution de degrés des nœuds d'un réseau social suit (de façon plus ou moins stricte) une loi de puissance, et peut aller jusqu'à être sans échelle *Scale-Free*. Lorsque l'on s'intéresse à la distribution de degrés dans un réseau aléatoire, sa forme est gaussienne, centrée sur la densité du réseau.
- Malgré une tendance à se regrouper en communautés, l'existence de liens faibles permet aux individus d'avoir des connaissances hors de celles-ci. Cela lui donne "accès" à d'autres communautés en peu de transition. Grâce à cela, les réseaux ont une taille de chemin moyen (APL) court entre deux individus pris au hasard (Newman, 2003).
- Un coefficient de clustering élevé associé à un APL faible est appelé effet Small-World (Vázquez, 2003).

La question de la distribution de degrés d'un réseau social est encore discutée dans la communauté des réseaux. Les réseaux définis comme *Scale-Free* au sens strict ont leur distribution de degrés qui suit une loi de puissance,

$$P(k) \propto k^{(-\gamma)} \quad (2.1)$$

avec un γ compris entre 2 et 3 (Newman, 2016). Seuls certains réseaux peuvent correspondre à cette définition (principalement les réseaux biologiques dont le processus de formation est fractal ou les réseaux technologiques). (Broido and Clauset, 2019) classifie plus de 900 réseaux issus des systèmes complexes, et détermine à quel point ces derniers ont une distribution sans échelle (figure 2.8). En remarquant que 49% des réseaux ne montrent pas de structure *Scale-Free* et que pour 88% d'entre eux, leur distribution de degré est au moins aussi bien

Réseau	r	cc	d	APL ou DE
Aléatoire	0	-	-	-
Réseau trophique	-0.263	0.16	3.3e-2	APL :2.05
Interaction de protéines	-0.156	0.072	5e-4	APL :6.80
Co-auteurs physique	0.363	0.45	8.8e-5	APL : 6.2
Facebook	0.22	0.6055	5.4e-3	DE : 4.7
Twitter		0.5653	2.67e-4	DE :4.5

Figure 2.7: Valeurs des propriétés pour différents graphes. r : Assortativité, cc coefficient de clustering, d densité, APL/DE Average Path Length / Diamètre efficient; Mesure pour différents types de réseaux. Réseau aléatoire, Co-auteurs d'articles en Physique (Neuman, 2003), Cycle alimentaire marin (foodweb) (Huxham et al., 1996), interactions de protéines pour le Yeast (Jeong et al., 2001), Facebook listes d'amis sur 4000 noeuds, Twitter cercles sur 80.000 noeuds (Leskovec and Krevl, 2014). En coloré pour les réseaux dit sociaux. Les valeurs de cc , d et APL d'un réseau aléatoire dépendent de la densité choisie en paramètre pour la construction de ce réseau.

décrite par une loi log-normal. La conclusion est que le phénomène Scale-Free est loin d'être aussi répandu que pensé dans les réseaux réels, et que sa définition strict n'est pas nécessairement un marqueur si important.

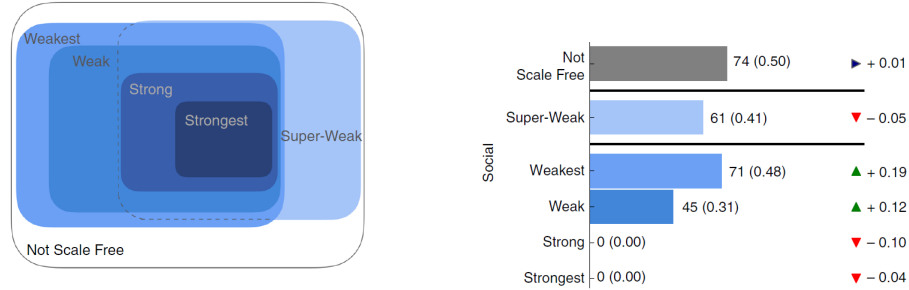
Dynamique et réseaux sociaux

Dynamique des réseaux sociaux (Leskovec, 2008) et (Kunegis, 2011) ont tous les deux étudié l'évolution temporelle de réseaux sociaux en ligne de très grande taille (MSN avec 240 millions de nœuds, Twitter avec 42 millions). Les deux ont remarqué la tendance qu'ont les réseaux à se densifier avec l'arrivée de nouveaux nœuds. Une conséquence immédiate de ce phénomène étant la réduction du diamètre du réseau.

Dynamique sur les réseaux sociaux Qu'il soit lui-même changeant (dynamique) ou immobile (statique), le réseau est alors un support pour un processus particulier. Les réseaux sociaux sont utilisés pour simuler des phénomènes de propagation ou autres interactions entre les entités en présence (Kirer and Cimpici, 2016), (Johansson, 2017), (Barrett et al., 2009), (Holzhauer, 2015), (Boccaletti et al., 2006). Dans chacune de ces applications la structure du réseau est déterminante, influençant la manière dont la dynamique va se jouer. La figure 2.9 illustre de manière simple l'influence sur l'aspect vitesse de propagation du processus joué dessus.

Le problème des données

Les données concernant les réseaux sociaux en ligne sont désormais abondantes (Leskovec, 2008) (Kumar et al., 2006). De nombreux datasets sont mis à disposition sur internet. Néanmoins, ces données ne concernent qu'un certain



(a) Les zones "strong" et "strongest" correspondent à la définition classique que l'on donne des réseaux Scale-Free (l'exposant γ compris entre 2 et 3) pour respectivement 50% et 90% des nœuds du réseau étudié. (b) Détails des degrés de réseaux scale-free dans le domaine social.

Figure 2.8: Répartition des réseaux étudiés parmi les distributions de degrés plus ou moins sans échelle. Cette répartition est observée quelque soit le domaine du réseau (social, technologique ou biologique). Illustrations issues de l'article (Broido and Clauset, 2019).

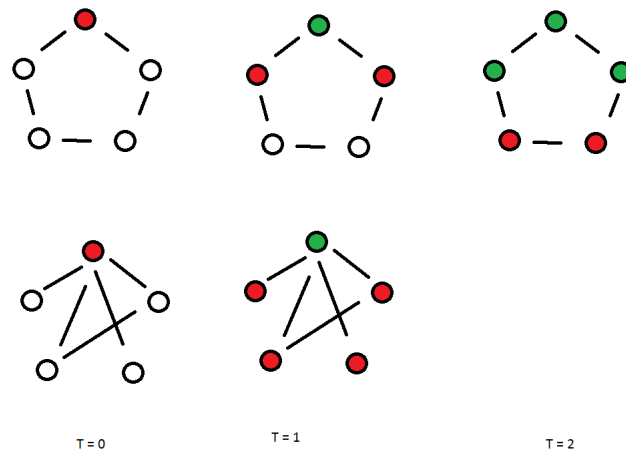


Figure 2.9: Exemple de propagation sur des réseaux aux topologies différentes mais au même nombre de nœuds et de liens. On y considère le scénario d'une propagation de maladie, où chaque personne malade en rouge, contamine son entourage sain en vert à chaque pas de temps. La forme du réseau influe sur la vitesse de propagation de la maladie, qui serait complète en trois pas pour le réseau supérieur, ou deux dans le cas du réseau du dessous.

type de réseaux, excluant souvent les réseaux hors ligne. Les réseaux personnels d'un individu sont plus difficiles à obtenir car nécessitent des interviews et sondages, et sont réalisés à petite échelle. Ils sont souvent incomplets, ou aux informations non adaptées à une utilisation en simulation. (Newman, 2003). Les données sont aussi manquantes pour des études historiques ou sociologiques, cherchant à simuler des populations pour lesquelles l'information concernant les réseaux sociaux est inexistante. Il va donc se poser la question de la génération de réseaux pertinents pour des usages en simulation.

Du fait des processus de formation des réseaux sociaux, les graphes ont en commun certaines propriétés considérées comme caractéristiques. Les plus connues sont :

- l'effet Small-World, regroupant un clustering important et une courte taille de chemin moyen entre deux nœuds,*
- et une distribution de degrés se rapprochant d'une distribution sans échelle (Scale-Free) des nœuds du réseau.*

Ces propriétés entraînent des comportements particuliers lorsque des processus (de diffusion par exemple) sont joués sur le réseau.

2.3.3 Conclusion sur les réseaux sociaux

Les réseaux sont une représentation formalisée d'un système. Ils mettent en avant les interactions entre les entités constituant ce système. La création et l'évolution des systèmes réels dépendent des contraintes d'environnement et de la façon dont ses éléments interagissent ensemble. Les réseaux représentant un même type de système exhibent des propriétés communes, qui les caractérisent donc vis-à-vis des autres types de réseaux.

Les réseaux sociaux ont, eux aussi, une série de propriétés topologiques qui leur sont particulières :

- Un clustering important, conséquence d'un regroupement en communauté
- Une assortativité positive, dûe au phénomène d'homophilie dans le choix des relations
- Le chemin entre deux individus pris aléatoirement est court en nombre de sauts, de part les liens faibles des connaissances
- La distribution de degrés des nœuds du réseau est proche une loi de puissance, à associer à la distribution de Pareto

Malgré une énorme quantité de données pour les réseaux dits en ligne, les autres réseaux souffrent d'un manque, voir d'une absence de données.

2.4 Génération de réseaux sociaux

2.4.1 Introduction aux paradigmes de génération

Différentes situations peuvent nécessiter la génération de réseaux synthétiques. On peut entre autres citer le manque ou l'inexistence de données, ou la volonté d'obtenir des réseaux semblables à certains déjà existants afin d'y jouer des simulations.

Les approches pour la génération de réseaux sociaux se divisent en deux paradigmes, qui sont parfois utilisés conjointement. Le premier tend à reproduire les comportements identifiés comme générateurs des liens entre entités. Cette approche requiert une connaissance du domaine, depuis les entités en présence jusqu'à leurs interactions pertinentes.

Le second paradigme consiste à trouver des modèles pour reproduire les caractéristiques mathématiques des réseaux réels. La fidélité à l'original fait office de validation, mais de bons critères de comparaison doivent être retenus. Les méthodes donnant des résultats quantitativement bons sont souvent complexes à mettre en œuvre, avec des liens entre paramètres et comportements sociaux non triviaux à expliciter. De plus, la plupart de ces méthodes se concentrent sur la reproduction d'une photographie d'un réseau à un instant donné, prenant rarement en compte les dynamiques d'évolution.

Nous allons donc détailler les modèles de génération de réseau les plus utilisés, afin de positionner notre méthodes parmi l'existant.

2.4.2 Contexte

La nécessité de générer des réseaux est apparue dans différentes disciplines, amenant des approches distinctes. En résumé, on peut distinguer trois courants :

- **Comportemental** - une approche par simulation sociale, qui essaye de reproduire les comportements des individus étudiés ([Jahanbazy et al., 2015](#))
- **Topologique qualitatif** - une approche statistique, qui cherche des modèles probabilistes reproduisant quantitativement des observations mesurables sur les réseaux réels ([Holland and Leinhardt, 1981](#))
- **Topologique quantitatif** - une approche abstraite, qui choisit et reproduit qualitativement des propriétés identifiées comme fondamentales du point de vue du domaine d'étude ([Watts and Strogatz, 1998](#)) ([Barabasi and Albert-László, 1999](#))

L'approche par simulation sociale travaille le plus souvent avec un petit nombre d'entités (de l'ordre de la centaine ou du millier), aux comportements distincts et identifiables. Les entités possèdent des attributs qui leurs sont propres, par exemple leur position dans un espace ou leur âge. Le modélisateur y intègre des règles d'interactions, représentant les comportements des individus du système "réel". La simulation est alors lancée, les actions successives des entités générant un réseau.

A titre d'exemple, une famille de modèles basiques sont ceux basés sur une dis-

tance entre les entités pour déterminer l'existence de lien (Hamill and Gilbert, 2009). Dans le cas d'une répartition régulière des entités dans l'espace, et d'un lien créé pour une distance constante, on se retrouve dans le cas extrêmement simple d'une grille régulière.

De l'autre côté, les modèles statistiques sont utilisés pour des systèmes aux entités nombreuses (Leskovec et al., 2010). Ces approches proposent un modèle de génération de données, cherchant les bons paramètres internes pour mimer la topologie du réseau à reproduire. La recherche de ces paramètres passe par des méthodes telles que les régressions linéaires ou les algorithmes génétiques (le modèle de régression symbolique (Menezes and Roth, 2014), ERGM (Holland and Leinhardt, 1981)). La fonction à optimiser est une distance entre données générées et données cibles, et les paramètres des modèles définissent donc les données générées.

L'étude des systèmes et réseaux complexes se développant, des méthodes spécifiques et des adaptations de l'existant ont été développées. Dans les années 2000, deux modèles se proposent de mimer de manière qualitative les propriétés les plus évidentes des réseaux sociaux. Le premier, l'attachement préférentiel (*Preferential attachment*), génère un réseau en reproduisant l'aspect propriété sans échelle de la distribution de degré des réseaux sociaux (Barabasi and Albert-László, 1999). Dans la même période, un deuxième modèle, le *Small World*, permet d'obtenir un réseau avec un clustering et une taille de chemin moyen qualitativement satisfaisants (Watts and Strogatz, 1998). Ces modèles peuvent être qualifiés d'abstraites, au sens où leur ressemblance avec des réseaux réels reste très qualitative, et qu'ils ne reproduisent que peu de propriétés à la fois.

Trois approches cohabitent pour la génération de réseaux. Issue des sciences sociales, une approche se base sur la reproduction des comportements des individus considérés comme générateurs de l'observé. Une autre approche cherche par des méthodes statistiques les meilleurs paramètres d'un modèle de génération pour une correspondance des propriétés (souvent topologiques) avec l'observé. Et enfin, la dernière approche reproduit qualitativement certaines propriétés considérées comme des signatures des réseaux sociaux.

2.4.3 Familles de méthodes de génération de réseaux

En reprenant la distinction faite précédemment, les méthodes de générations applicables ou spécifiques aux réseaux sociaux peuvent se diviser en deux grands ensembles : processus ou topologiques.

- Les approches basées sur les processus se concentrent sur les mécanismes, socialement plausibles et essentiellement macroscopiques, de construction d'un réseau
- Les méthodes basées sur la topologie cherchent à reproduire les propriétés d'un réseau réel. Ces dernières se divisent en deux sous-ensembles
 - les méthodes abstraites, qui reproduisent quelques propriétés à gros traits (mono-qualitatif)
 - les méthodes statistiques, cherchant à reproduire au mieux des propriétés clés mesurées sur les réseaux réels (multi-quantitatif)

Les processus cherchent un ensemble d'actions cohérentes dans le contexte de leur sujet d'étude. Les approches topologiques recherchent des modèles de génération permettant de reproduire un cliché pris à un moment donné d'un réseau cible.

Méthodes centrées sur les processus

Les méthodes centrées sur les processus se basent sur les comportements supposés générateurs de liens entre individus. Le but est de trouver un ensemble d'actions à allouer à une population qui va mener à la création d'un réseau les connectant entre eux. La difficulté réside dans le fait de trouver des actions cohérentes sur le plan social, générant tout de même des réseaux plausibles sur le plan topologique. Bien souvent, les hypothèses au sujet des processus sociaux semblent plus choisies à posteriori au modèle plutôt que par souci de véracité en terme de processus sociaux générateurs. La création et/ou l'évolution d'un réseau est souvent indissociable de la simulation sociale dans laquelle il est intégré.

Par exemple, dans le modèle de mariage de (Billari et al., 2007), la possibilité pour que deux agents se marient va dépendre de règles définies préalablement telles que l'intérêt porté ou la proximité, et des interactions qui évoluent au cours de la simulation.

Dans ce cadre, la plupart des modèles de génération de réseaux sont ad-hoc, définis spécifiquement pour une simulation aux objectifs plus larges que l'unique génération d'un réseau. (Johansson, 2017) qualifie ces modèles de spécialisés, pour lesquels la structure du réseau n'est pas indépendante des processus qui s'y jouent. Dissocier la génération de réseau et son utilisation dans une simulation amène à posteriori le problème de correspondance (mapping) entre nœuds et entités (Menezes and Roth, 2014).

Validation des méthodes par processus La validation des méthodes de génération de réseau par processus est un sujet assez délicat en ceci qu'il n'a pas forcément vocation à être fidèle à l'observé mais plus aux éléments à mettre

en avant qui ont été choisis. On peut distinguer une gradation des objectifs de la modélisation du test d'hypothèses à la prédiction du comportement du système modélisé. Entre ces différents cas, le processus de validation peut varier. Lorsque l'on manque de données, il se concentre sur des éléments attenants : un soin particulier est donné aux hypothèses d'élaboration ou à l'avis d'experts (Heath et al., 2009). En présence de données, la comparaison entre système réel et simulé est envisageable (Menezes and Roth, 2014).

Méthodes centrées sur les propriétés topologiques

Les méthodes centrées sur la topologie s'intéressent au réseau en tant qu'objet mathématique. Il est vu comme une représentation du système étudié. La qualité d'un réseau est réduite à sa similarité topologique avec les réseaux du même domaine, la précision et le nombre de propriétés retenues étant variables. Ces méthodes ne s'intéressent pas aux processus réels ayant construits le réseau, et sont donc applicables à un grand nombre de situations. En contrepartie, ils ne font que reproduire un réseau cible à un moment précis, ne garantissant pas son adéquation lorsque des dynamiques temporelles sont étudiées.

Validation des méthodes topologiques Le fait de travailler sur des propriétés de réseau permet de faciliter les processus de validation. Ils peuvent se faire de plusieurs manières :

- Lorsque la disponibilité des données est assez importante, il est possible de comparer les résultats d'un modèle avec le réel. C'est par exemple le cas en économie (Sallans et al., 2003), (Bianchi et al., 2007).
- Pour d'autres cas, la validation peut se contenter de résultats qualitatifs. C'est la solution qu'empruntent souvent les modèles usant d'approche processus mais cherchant à valider les propriétés topologiques obtenues (Hamill, 2010)
- la validation peut aussi être indirecte ; on peut ainsi comparer le résultat de processus lancés sur les réseaux, comme par exemple le résultat de phénomènes de propagation (Cointet and Roth, 2007).

Lorsque l'on choisit de valider en se basant sur la comparaison avec le réseau réel que l'on cherche à reproduire, on peut comparer une série de propriétés (Menezes and Roth, 2014).

Au sujet de la validation

La validation concerne tout les niveaux d'une modélisation. Elle touche à :

- la théorie, en tant qu'abstraction du réel
- aux hypothèses, dans leur rôle de réduction de complexité
- la modélisation, en tant que transcription de la théorie en modèle
- l'implémentation du modèle

Nous n'abordons ici que la validation spécifique à la génération d'un réseau, donc du réseau en lui-même sur le plan topologique. Quelque soit l'approche usitée, certaines méthodes de validation peuvent être utilisées de manière "générique".

En cas de manque de données, le croisement des modèles permet de vérifier que des modèles implémentés de différentes manières donnent les mêmes résultats ([Axtell et al., 1996](#)).

2.4.4 Les méthodes de génération de réseaux sociaux

Nous allons détailler une sélection de méthodes pour la génération de réseaux sociaux, sur le critère de la pertinence par rapport à notre sujet. Le *principe* de base de chacune des méthodes choisies est présenté, ainsi que les données ou connaissances *requises*. Le fonctionnement sera ensuite *détaillé*, avant de finir par la présentation des *résultats*.

Méthodes orientées processus

Nous nous intéressons dans un premier temps aux méthodes orientées processus, c'est à dire cherchant à reproduire les actions réelles des entités génératrices de réseaux.

La plupart de ces méthodes sont intégrées dans des simulations. Ces dernières n'ont pas seulement vocation à générer des réseaux, mais aussi à jouer des dynamiques dépendant de ces réseaux. Autrement dit, la formation d'un réseau par interactions des entités est rarement décorrélée de son utilisation au sein de la simulation. Nous présentons une méthode de génération spécifique à un modèle (Su et al., 2014), puis une des rares méthodes génériques (à notre connaissance) : Yang (Thiriot and Kant, 2008).

Modèle de génération spécifique orienté processus

Nous allons nous intéresser à un travail utilisant une méthode ad-hoc de génération de réseau, y faisant jouer conjointement des dynamiques de modèle. Dans le contexte de la formation d'opinion, (Su et al., 2014) s'intéressent à la co-évolution des opinions d'individus et d'un réseau dynamique d'interactions entre ces individus.

Principe Le modèle répète deux étapes jusqu'à convergence : la mise à jour des opinions et de celle du réseau en lui-même. La suppression des liens peut survenir quand la différence d'opinions de deux agents dépasse un certain seuil ; leur création survient aléatoirement.

Pré-requis Aucune donnée n'est nécessaire pour pouvoir utiliser ce modèle.

Détails Les paramètres généraux du modèle sont un niveau de confiance ϵ , et une probabilité de création de lien dirigé q . Chaque entité est dotée d'une valeur d'opinion aléatoire P et d'un nombre de liens avec les autres entités.

Le modèle commence avec un réseau aléatoire dirigé. La première étape consiste à mettre à jour une liste de confiances w pour chaque agent A_{acting} : S'il est connecté à un agent A_{other} , et que leur différence d'opinions est inférieure à epsilon- ϵ , A_{other} sera ajouté à w .

Ensuite, les liens du réseau sont mis à jour.

- Si aucun lien n'existe entre deux entités, il est créé avec une probabilité q
- Si un lien existe, mais que la différence d'opinions au temps t d_t est supérieure au niveau de confiance ϵ , le lien est détruit avec une probabilité qui augmente en fonction de la valeur de la différence et d' ϵ suivant la formule

$$P(\text{suppression}) = 1 - e^{1 - \frac{d_t}{\epsilon}}$$

La dernière étape consiste en une mise à jour de l'opinion des entités. Chaque A_{acting} va faire la moyenne des opinions de la liste w avec sa propre opinion. Cette nouvelle valeur de P sera celle de l'agent pour le temps $t+1$.

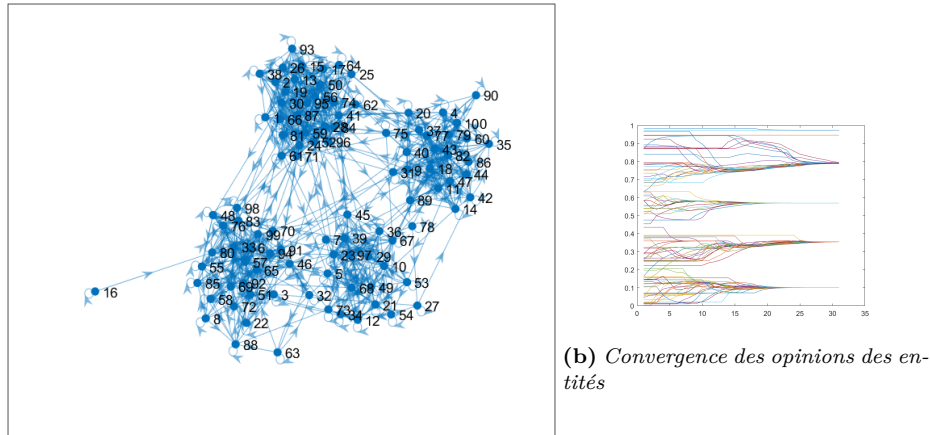
Résultats Le niveau de confiance ϵ est le paramètre clé du modèle. Pour une valeur élevée, le réseau tend vers un réseau aléatoire et la formation d'un unique cluster d'opinions. Au contraire, une valeur faible favorise de multiples clusters au sein des entités et un réseau associé qui possède un clustering particulier, plus élevé que dans le cas d'un réseau aléatoire.

Dans les exemples utilisés pour illustration, le modèle est initialisé avec un nombre de nœuds à 500, un nombre de liens par nœud de 10, et la probabilité q de création spontanée de lien est à 0.01. Nous étudions le résultat pour deux valeurs d' ϵ , [0.1;0.5].

Les résultats numériques sont donnés dans le tableau 2.1. dans les deux cas, la distribution de degrés se rapproche d'une gaussienne. La création de clusters

ϵ	densité	APL	Coeff. clust.
0.1	0.07	3	.24
0.5	0.12	.217	.145

Table 2.1: Résultats des dynamiques du modèle pour deux valeurs d'epsilon.

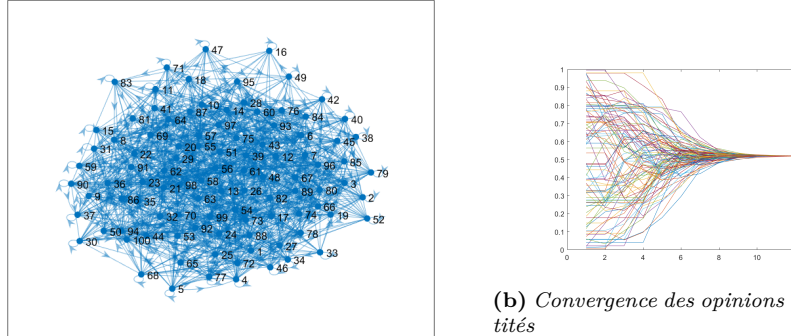


(a) Réseau obtenu, on remarque quatre communautés fortement connectées

Figure 2.10: Résultat après 500 itérations, 100 noeuds, $\epsilon=0.1$. Plusieurs clusters d'opinions se sont formés, phénomène que l'on retrouve au niveau du réseau en lui-même. Le coefficient de clustering est de 0.24, quatre fois supérieur à celui qu'aurait un réseau aléatoire de même taille (0.07); La densité de réseau est à 0.07.

d'opinions et de liens dans le cas d' ϵ à 0.1 permet d'obtenir un coefficient de clustering élevé par rapport à la densité, ce qui l'éloigne d'un réseau aléatoire sur ce point. Ce n'est pas le cas de l'autre configuration, entraînant un réseau proche de l'aléatoire.

Le modèle permet d'obtenir un coefficient de clustering particulier (par rapport à la densité du réseau) dans le cadre d'une dynamique conjointe de construction et diffusion d'opinions sur un réseau. Il reste simple d'utilisation avec un seul paramètre. Les résultats cessent d'être probants lorsque l'on essaye d'obtenir des densités de réseau plus importantes (> 0.1).



(a) Graphe obtenu, proche d'un réseau aléatoire.

Figure 2.11: Résultat après 50 itérations, 100 noeuds, $\epsilon=0.5$. Plusieurs clusters d'opinions se sont formés, phénomène que l'on retrouve au niveau du réseau en lui-même. Le coefficient de clustering est de 0.145, et la densité est à 0.12. On est donc très proche d'un réseau aléatoire.

Yang

Principe Yang - Yet Another Network Generator (Thiriot and Kant, 2008) - est un générateur de réseau qui se base sur un réseau bayésien. Des règles de création de liens sont à donner manuellement, et le réseau bayésien va générer un réseau respectant règles et observations. Ces observations permettent d'intégrer des données "éparses" (*scattered*) sur la population (pyramide des âges, revenus des agents..).

Pré-requis Une expertise sur les processus générateurs des réseaux (*mœurs sociales, habitudes*) et toute autre donnée pouvant servir d'attribut aux agents (*pyramides des âges, natalité*), qui seront prises en compte par les règles de création de lien définies.

Détails Le but de ce modèle est de pouvoir fournir un générateur de réseaux sociaux, en y intégrant le plus possible d'attributs et propriétés connus. Le générateur permet de :

- Prendre en compte les attributs des agents pour la création de liens,
- Définir des règles de création sur ces liens, comme la réciprocité, l'homophilie, ou d'autres mécanismes plus complexes,
- définir plusieurs types de liens entre les agents.

L'utilisateur définit en premier lieu les types de liens qui peuvent exister entre les nœuds. Il définit ensuite les attributs des agents. Un attribut peut être numérique (*age*), nominal (*sexe*), ou représenter un lieu de fréquentation (*école par exemple*).

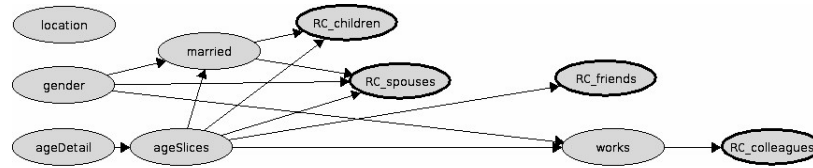


Figure 2.12: Réseau bayésien d'un agent pour le cas d'application Kényan. Un lien entre deux nœuds correspond à un lien de probabilité entre les variables correspondantes. Les nœuds en gras sont le nombre de liens à créer pour chaque type de liens. Le nœud "age detail" permet par exemple l'intégration de données réelles, ici la probabilité d'âge issu de la pyramide des âges

Les variables des agents étant souvent dépendantes (nombre d'enfants en fonction de l'âge et du statut matrimonial), ces relations sont définies par des probabilités conditionnelles et sont représentées au sein d'un réseau bayésien. A ce stade, il est possible d'intégrer des données externes, comme par exemple la pyramide des âges pour la distribution de l'attribut associé.

La dernière étape consiste à définir la probabilité d'avoir un lien entre deux agents, fonction de la valeur de leur attribut. Le réseau bayésien est alors résolu, générant une population avec des attributs, puis générant des liens entre les individus. Sur l'exemple proposé dans l'article ([Thiriot and Kant, 2008](#)), le cas d'application est un réseau social dans un village Kényan, s'intéressant plus précisément à l'influence sociale dans le choix d'utilisation de contraceptifs. Les types de liens dépendants des attributs sont [époux ;mère ;collègues ;amis], ceux créés en deuxième passe car dépendants des premiers sont [père ;fratrie].

Les attributs des entités alors considérés comme influençant la création des liens sont [statut marital ;age ;genre ; travail ;localisation].

Est alors créé un réseau bayésien définissant les probabilités de liens en fonction de la valeur des attributs et des autres liens.

Résultats Le réseau résultant est une population d'individus dotés d'attributs, et de liens de différents types entre eux. Une fois atteint les 10.000 agents, le seuil d'erreur statistique est dépassé et les résultats deviennent stables. La densité est en dessous de 1%, la taille de chemin le plus court autour de 5, et la transitivité (proche du clustering) est forte comparée à la densité du réseau (~ 0.21 pour une densité de 0.01) (figure 2.15). Deux types d'erreurs peuvent survenir, dues soit à la nature discrète des agents couplée à une taille de population pas assez importante, soit à des définitions des probabilités d'attributs et de liens non solubles, auquel cas il sera nécessaire de les redéfinir :

- Erreur de distribution statistique : signifiant que tous les individus répondant aux contraintes spécifiées n'ont pas pu être générés. Cette erreur est mesurée par la différence entre les probabilités théoriques et celles effectivement trouvées dans le réseau généré (ligne pointillée du bas dans la figure 2.14).

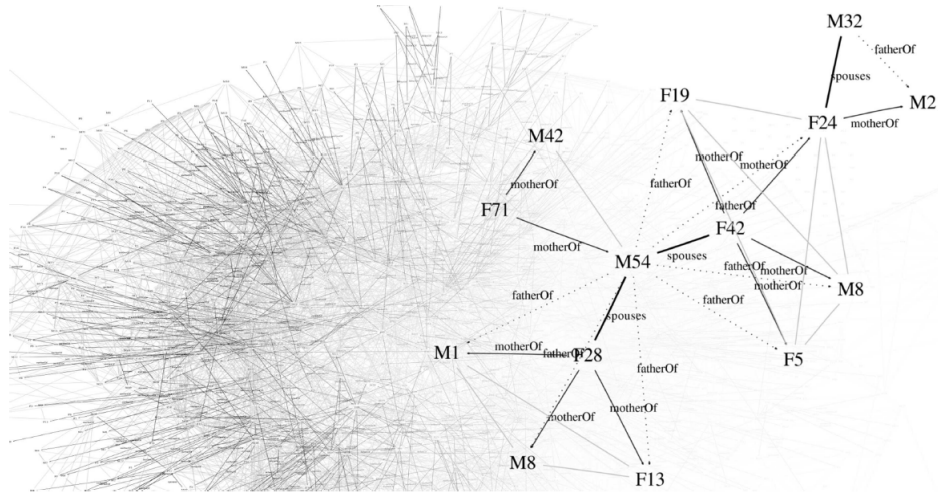


Figure 2.13: Gauche : réseau de relations obtenu avec 50.000 agents. Droite : Détail sur un agent. M ou F pour "Male" et "Female", suivi de l'âge. L'homme de 54 ans est donc marié à 2 femmes, F28 et F42, et ils ont 7 enfants.

- Erreur d'affiliation (matching) : tous les liens décidés lors de la résolution du réseau bayésien n'ont pu être affectés aux agents en présence.

Ce modèle nous semble plus approprié pour des utilisations descriptives fortes : des contraintes plus élevées entraînent des résultats plus stables, mais nécessitent des populations de relativement grande taille. Les propriétés topologiques du réseau obtenu vont dépendre des contraintes définies par le modélisateur et sont difficilement maîtrisables.

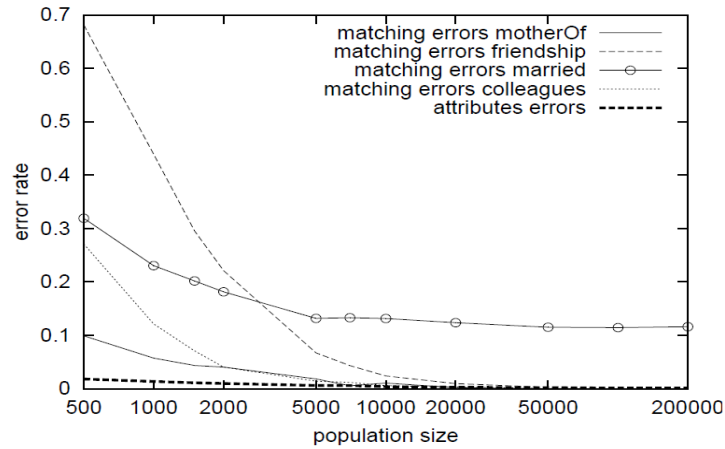


Figure 2.14: Taux d'erreur d'utilisation de Yang, en fonction des tailles de population. On remarque que ce taux d'erreur diminue lorsque le nombre d'entités augmente, permettant d'arriver dans des zones statistiques confortables où le coût d'une d'erreur est moins important. Avec le réseau bayésien défini dans le cas d'utilisation de la population Kényane, une population minimum de 10.000 agents est nécessaire pour de bons résultats.

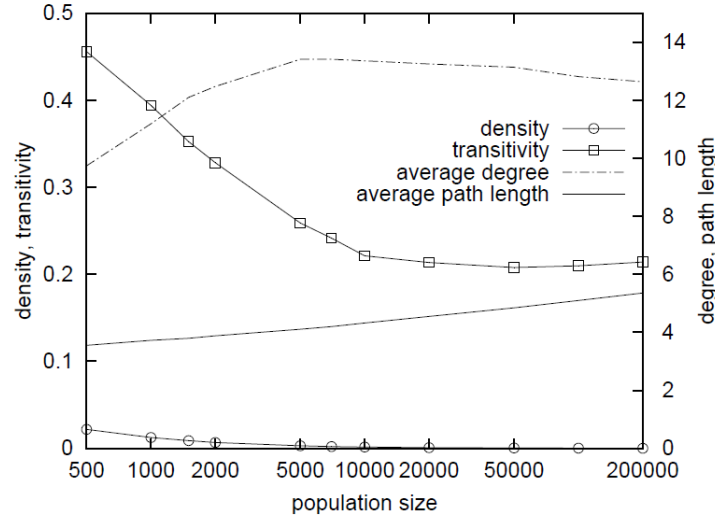


Figure 2.15: Propriétés du réseau obtenu en fonction de la taille de population. La densité est faible dû au cas d'application, les relations rentrant en compte pour les choix de contraception étant moins nombreuses que l'ensemble des relations d'un individu. La transitivity est une propriété assimilable au clustering et significativement élevée. La taille de chemin moyen, autour des 4.8, correspond aux valeurs trouvées dans les réseaux Small-world.

Nous allons maintenant nous intéresser aux méthodes de génération basées sur une approche topologique - A commencer par les méthodes abstraites, stylisées, auxquelles feront suite des méthodes plus précises.

Méthodes orientées topologie - abstraites

Nous allons détailler les méthodes abstraites de génération - visant à reproduire qualitativement une ou deux propriétés des réseaux réels. Ces méthodes donnent des résultats topologiques assez pauvres, mais possèdent une simplicité de processus et ont peu de paramètres. Elles font parties des modèles les plus utilisés par la communauté des sciences sociales (Amblard et al., 2015) (Holzhauer, 2015).

Small-World

Le phénomène Small-World fait référence à l'idée popularisée par l'expérience sociale de (Milgram, 1967) que chaque individu est connecté à l'ensemble de la population par une chaîne de connaissance de six individus. Le modèle éponyme de (Watts and Strogatz, 1998) est celui de référence pour générer ce type de réseau : il reproduit qualitativement les propriétés de clustering et taille de chemin moyen mesurées sur les réseaux sociaux (Vázquez, 2003). Le modèle commence par un réseau en grille régulière, et procède à des rebranchements aléatoires de liens, jusqu'à atteindre l'ordre de grandeur souhaité sur les propriétés susnommées.

Principe Le réseau de départ avant modification est une grille régulière (*regular lattice*), où chaque nœud est connecté à ses k plus proches voisins. Ce réseau possède un haut clustering et une moyenne de taille de chemins importante. Le rebranchement aléatoire de certains liens de la grille va permettre la création de raccourcis entre les clusters de nœuds, sans en faire baisser significativement le coefficient moyen. À terme, l'application répétée de ce procédé fait tendre la grille régulière vers un réseau aléatoire, possédant alors un coefficient de clustering bas (figures 2.16).

Pré-requis Le seul pré-requis est un ordre de grandeur concernant la taille de chemin moyen entre des nœuds et le clustering que l'on souhaite atteindre.

Détails Le réseau de départ avant modification est une grille régulière, ou chaque nœud est connecté à ses k plus proches voisins. Dans ce réseau, le coefficient de clustering est élevé, défini par

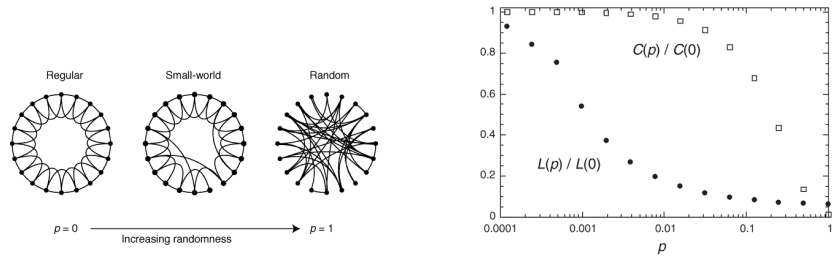
$$\frac{3k(k-1)}{2k(2k-1)}$$

et la longueur moyenne de chemin entre noeuds $\frac{N}{2k}$.

À l'autre extrême, un réseau aléatoire de densité d possède un coefficient de clustering de d et un APL d'ordre $\log(N)$ (Fronczak et al., 2004).

La plage de valeurs que peut prendre la probabilité de reroutage pour transformer la grille régulière en réseau Small-World est relativement étendue :

$$0.001 < p < 0.1$$



(a) Transition de la topologie

(b) Transition des propriétés de clustering et taille de chemin moyen entre nœuds (échelle log).

Figure 2.16: Transition depuis une grille régulière vers un réseau aléatoire, en passant par une plage de réseaux dit Small-World : Clustering élevé et longueur de chemin moyen entre deux nœuds aléatoires courte. (de l'ordre de $\log(N)$). Illustrations issues de (Watts and Strogatz, 1998)

Une valeur très faible est suffisante à créer les raccourcis faisant diminuer significativement la taille de chemin moyen entre deux nœuds pris au hasard. À l'autre extrême, la quantité de liens rebranchés ne doit pas dépasser un certain seuil afin de conserver une propriété de clustering importante.

Résultats Les valeurs de clustering et d'APL pour un réseau de densité 4% sont données dans le tableau 2.2, pour différentes valeurs de probabilité de rebranchement. On remarque que le modèle permet d'obtenir un clustering très important tout en faisant varier l'APL de son maximum

à son minimum dans le cas d'un réseau aléatoire

$$\propto \log(N)$$

Le modèle est particulièrement simple d'utilisation. Sur les propriétés topologiques d'APL et de clustering, il génère des réseaux en accord avec ceux observés dans le monde réel, qualitativement.

\mathcal{P}	Clustering	APL
0.0 (grille régulière)	0.7105	12.954
0.001	0.7057	6.1037
0.01	0.6738	3.722
0.1	0.3951	2.67
1 (réseau aléatoire)	0.0408	2.389

Table 2.2: Valeur d'APL et de clustering pour des réseaux générés depuis une grille régulière de 500 nœuds, avec une connectivité $k=10$, entraînant une densité de 0.04. Valeurs données pour trois valeurs du paramètre p (probabilité de reroutage) pour le modèle de Watts-Strogatz ([Watts and Strogatz, 1998](#)).

Attachement préférentiel

Principe Le modèle de création de réseau par attachement préférentiel (([Barabasi and Albert-László, 1999](#))) cherche à reproduire la propriété sans échelle (Scale-Free) de la distribution de degrés des nœuds. Cette propriété a été observée dans les réseaux technologiques (réseaux de routeurs, infrastructures électriques (*power grid*)), biologiques (réseaux de protéines), et de manière moins régulière -souvent plus proche d'une distribution de Pareto- dans les réseaux sociaux (collaboration, citation) ([Barabási and Pósfai, 2016](#)).

Le modèle se base sur un processus d'ajout progressif de nœuds. Chaque nouveau nœud entrant favorise une connection aux entités plus fortement connectées.

Pré-requis Le paramètre du modèle définit le nombre de liens m auxquels va se connecter chaque nouvel entrant. Le réseau initial peut être vide ou comporter certains nœuds et liens afin de pouvoir obtenir un réseau final avec une densité précise.

- Détails** Le réseau commence vide ou avec un certain nombre de nœuds
- Un nouveau nœud entrant choisit aléatoirement un nœud parmi ceux existants, avec une probabilité proportionnelle à leur degré
 - Un lien est créé

L'ajout d'un lien est répété jusqu'à ce que le nœud crée m connections. le processus continue jusqu'à ce que tous les nœuds soient ajoutés.

Résultats Le processus entraîne une préférence pour les nœuds possédant une plus forte connectivité. Quel que soit le paramètre m , le réseau est invariant dans le rapport entre les différents degrés de sa distribution. Le coefficient γ , qui détermine la probabilité d'avoir un nœud d'un degré k donné, est définie par :

$$P(k) \propto k^{(-\gamma)}$$

les valeurs de γ sont comprises entre 2 et 3 pour la définition stricte d'un réseau sans échelle. Le modèle de ([Barabasi and Albert-László, 1999](#)) donne, à partir

d'un certain degré, un γ à 3. Pour être précis, la distribution de degrés est donnée par (Newman, 2003)

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$$

qui se réduit pour de grandes valeurs de k à γ , devenant indépendante du paramètre m .

Il s'agit du modèle le plus simple permettant d'obtenir une distribution de degrés sans échelle. Néanmoins, les résultats sont peut-être, même pour une approche qualitative, trop éloignés des observations pour les réseaux sociaux.

Résultats des méthodes abstraites Ces méthodes sont de loin les plus utilisées pour la génération de réseaux en simulation sociale. L'étude ([Amblard et al., 2015](#)) se base sur l'analyse de 15 ans d'utilisation de réseaux sociaux dans un journal de simulation sociale (The Journal of Artificial Societies and Social Simulation - JASSS). On constate que 67% des simulations utilisent au moins une de ces méthodes pour générer des réseaux sur lesquels travailler. Néanmoins, les réseaux générés par ces modèles reproduisent qualitativement certaines propriétés, ce qui n'assure pas des comportements cohérents lors de leur utilisation. Ce point est élaboré dans la partie *Discussion sur les méthodes de génération*.

Méthodes topologiques plus complexes

Nous allons maintenant nous intéresser aux méthodes plus complexes de génération de réseaux basées sur la topologie.

ERGM

Principe L'*Exponential Random Graph Model* (Holland and Leinhardt, 1981) présente une famille de méthodes statistiques, proposant deux processus

- Lors du premier processus, l'utilisateur va fournir une liste d'attributs topologiques considérés importants (comme par exemple ceux donnés en 2.17). *ERGM* va déterminer leurs facteurs d'implication statistique dans la topologie d'un réseau donné en modèle à reproduire.
- La deuxième étape permet de générer des réseaux semblables à celui fourni en tant que modèle, en se basant sur les estimations de paramètres réalisées dans la première étape.

Pré-requis Il est nécessaire d'avoir un réseau réel à reproduire, ainsi que faire le choix des formes topologiques de réseau que l'on considère comme importantes et caractérisantes.

Détails ERGM est une famille de modèles statistiques, connaissant un intérêt grandissant et des améliorations depuis 50 ans (Holland and Leinhardt, 1981) (Frank and Strauss, 1986) (Snijders et al., 2010). Le principe est de réaliser des régressions linéaires sur certaines propriétés, sans avoir l'hypothèse d'indépendance entre les liens du réseau qui était une limitation des premiers modèles.

Le modélisateur choisit les formes topologiques (patterns ou structures) de réseaux qui lui semblent fondatrices pour le réseau cible observé. Ces formes peuvent être choisies depuis une bibliothèque des plus courantes (figure 2.17, fournie notamment par la librairie Statnet (Handcock et al., 2003) pour le logiciel R (R Core Team, 2013)), ou de définir les siennes, en prenant par exemple en compte des propriétés particulières associées aux nœuds par ses soins dans son modèle. Ces structures peuvent par exemple être "la propension pour des individus de même âge de se lier", ou encore "se lier si des amis en communs existent".

Pour le premier processus, le modèle va chercher l'influence de ces patterns dans la topologie du réseau, en comparaison d'un réseau aléatoire de même nombre de nœuds et de liens. On peut y voir une régression linéaire dépourvue de l'hypothèse d'indépendance entre les variables.

L'issue de cette première partie est une série de facteurs, associés à chaque pattern, pouvant être négatifs (ce qui signifierait alors que l'absence du pattern en question est statistiquement significative dans la topologie du réseau observé). À l'issue de l'obtention de ces facteurs, on peut alors tirer des conclusions sur l'influence qu'on eu ces patterns dans la construction du réseau dans le cadre d'étude sociologique (par exemple (Lazega et al., 2012) sur l'influence des normes sociales pour la construction de réseau). On décrit *ERGM* comme une famille de modèles, car un jeu précis de ces facteurs permet de générer une classe de réseaux issue du réseau cible. Plus spécifiquement, la probabilité d'obtenir un réseau y particulier est donnée par

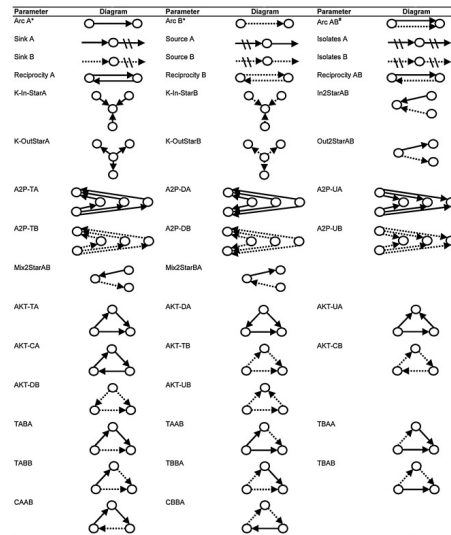


Figure 2.17: Exemples de patterns dont on peut chercher la prévalence dans un réseau avec la méthode ERGM

$$h(y) = \exp\{\eta(\theta) \cdot g(y)\} / c(\theta)$$

où $g(y)$ est un vecteur des structures préalablement sélectionnées, $\eta(\theta)$ les facteurs qui leurs sont associés, et $c(\theta)$ un facteur de normalisation.

Il est ensuite possible de générer des réseaux de même classe, à partir des facteurs trouvés précédemment. Le modèle va utiliser des chaînes de Markov, partant d'un réseau vide, ajoutant un nœud et regardant si le réseau généré se rapproche plus du réseau cible. Le modèle continue alors jusqu'à trouver un réseau correct.

Résultats Les résultats de la première étape permettent, sous l'hypothèse d'avoir choisi une bonne grille de lecture, i.e. de bons patterns, de déterminer leur importance et donc les processus qui peuvent les expliquer. Par exemple, dans un réseau social, une fermeture de triangle équivaut à dire qu'un ami en commun présente deux personnes, qui deviennent à leur tour amis. Si le facteur associé au pattern de triangle fermé est important, cela peut signifier que la plupart des connaissances se font par l'intermédiaire d'amis communs. Néanmoins certains patterns peuvent se "recouvrir" en terme de topologie, rendant plus difficile la détermination de l'importance relative de ces derniers.

Quant à la deuxième étape, il est possible d'évaluer la génération d'un réseau par sa distance au réseau qui a permis de déterminer les facteurs des patterns.

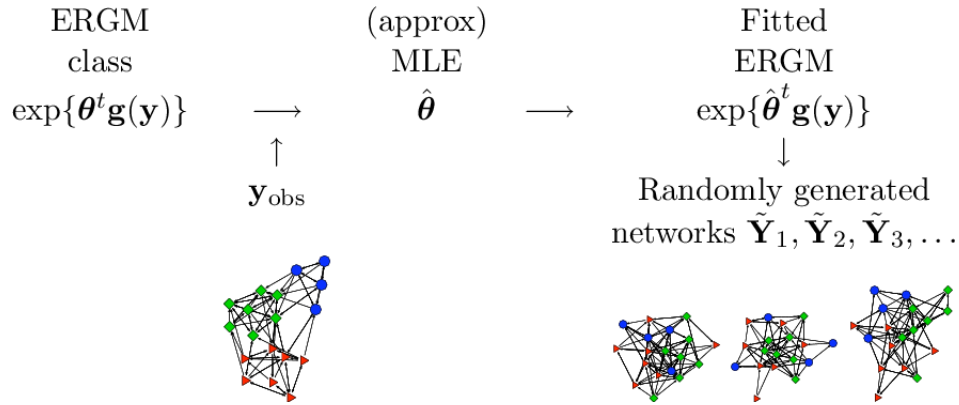


Figure 2.18: Illustration du fonctionnement d'ERGM. $\theta^t g(y)$ représente les patterns choisis et leurs facteurs. Y_{obs} un réseau cible. Une méthode est alors utilisée pour déterminer ces facteurs. (ici, **Maximum Likelihood Estimation**). À l'issue de quoi un vecteur de facteurs donne l'importance relative des patterns $g(y)$ pour le réseau Y_{obs} . Il est alors possible de générer des réseaux aléatoires de même famille. Illustration issue de (Hunter et al., 2008).

La puissance du modèle ERGM requiert pour être utilisée des connaissances assez fortes en statistiques. Le choix des patterns n'est pas trivial et demande une certaine compréhension du phénomène ayant amené à la création du réseau observé.

Variable	Nom	Dirigé ou non
i,j	identifiant des noeuds	les deux
k	degré entrant / degré	les deux
kj	degré sortant	dirigé
d	distance simple	les deux
dD	distance directe	dirigé
dr	distance inversée	dirigé

Table 2.3: *Attributs des nœuds pris en compte pour les formules.*

Regression Symbolique (Menezes and Roth, 2014)

Principe La méthode consiste à trouver automatiquement une formule permettant de générer un réseau semblable à un réseau choisi comme cible à reproduire. Cette formule détermine la probabilité de générer un arc entre deux nœuds, en fonction d'attributs de ces derniers. La recherche de cette formule est réalisée par un algorithme génétique.

Pré-requis Le modèle nécessite un réseau cible en entrée, et fonctionne avec sept paramètres pour l'algorithme génétique, dont les valeurs par défaut n'ont pas été changées lors des tests sur six réseaux à copier. Ces paramètres ne conditionnent pas la qualité des résultats du modèle, mais la vitesse pour les trouver.

Détails Le modèle commence avec une formule aléatoire de générateur de lien. Un réseau est alors généré avec cette formule, avec comme condition d'arrêt le même nombre de liens que le réseau réel. Il s'agit donc d'un processus de génération croissant à l'instar d'un attachement préférenciel. Le score du réseau obtenu est évalué, puis l'algorithme génétique va faire muter la formule qui définit la probabilité d'avoir un lien entre les nœuds du réseau. Cette formule utilise des opérateurs mathématiques (tableau 2.4). Une fois une nouvelle formule obtenue, plusieurs réseaux sont générés, puis on calcule la distance moyenne des réseaux obtenus au réseau cible. On évalue cette formule, et elle remplacera la précédente si elle est meilleure. Les opérateurs mathématiques prennent en compte les attributs des nœuds (tableau 2.3); la fonction de fitness est la distance au réseau cible suivant cinq critères(2.5). L'attribut *index* possédé par les nœuds, leur attribuant un numéro identifiant, permet de ne définir qu'une formule pour générer le réseau, tout en permettant par exemple des regroupements locaux de ces derniers à l'aide de la formule mathématique modulo.

Afin de trouver des formules de taille raisonnable, deux d'entre elles sont maintenues en même temps, l'une plus efficace et l'autre plus courte.

Type d'opérateur	opérateurs
Arithmétique simple	+, -, /, *
Fonction mathématique	xy, ex, log, abs, min, max
Expression conditionnelle	!, =, !=
Fonction d'affinité	phi (opérateur de modulo)

Table 2.4: *Regression Symbolique : Opérateurs à la disposition de la formule de création de lien.*

k - degree centrality distributions
PR - PageRank centralities
Triadic profiles
Topological distance

Table 2.5: *Critères de score entre réseau cible et réseau généré.*

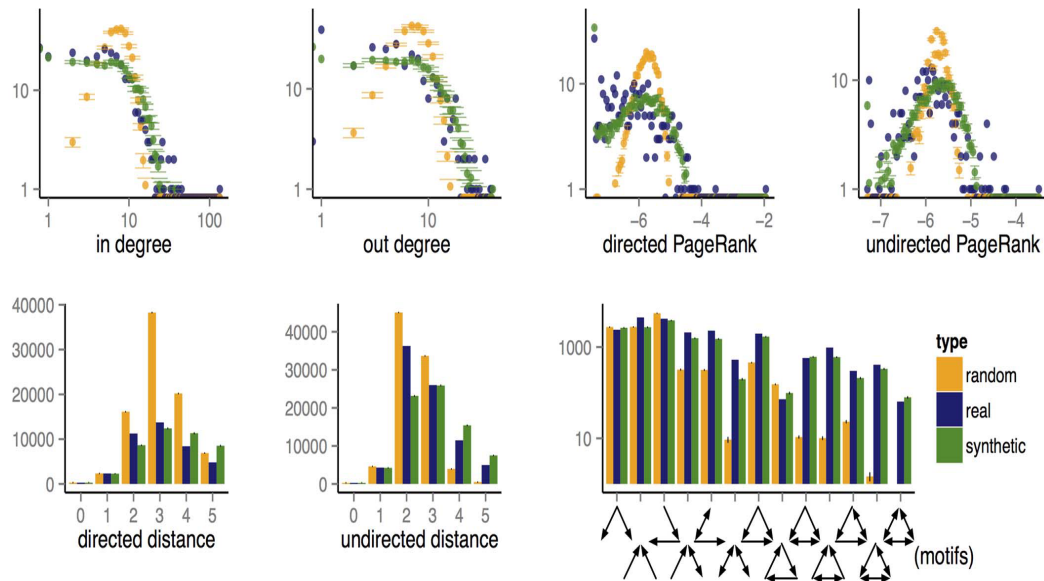


Figure 2.19: *Résultats sur les métriques prises en compte pour le score pour la reproduction d'un réseau neuronal de C.Elegans. Les motifs correspondent au critère de profils de triade (Triadic profiles).*

Résultat

Ce modèle n'utilise que des actions au niveau du nœud pour faire émerger des réseaux aux propriétés semblables. Malgré l'aspect stochastique des règles de création de liens, les réseaux générés avec une même formule ont des résultats proches suivant les métriques utilisées. L'algorithme converge.

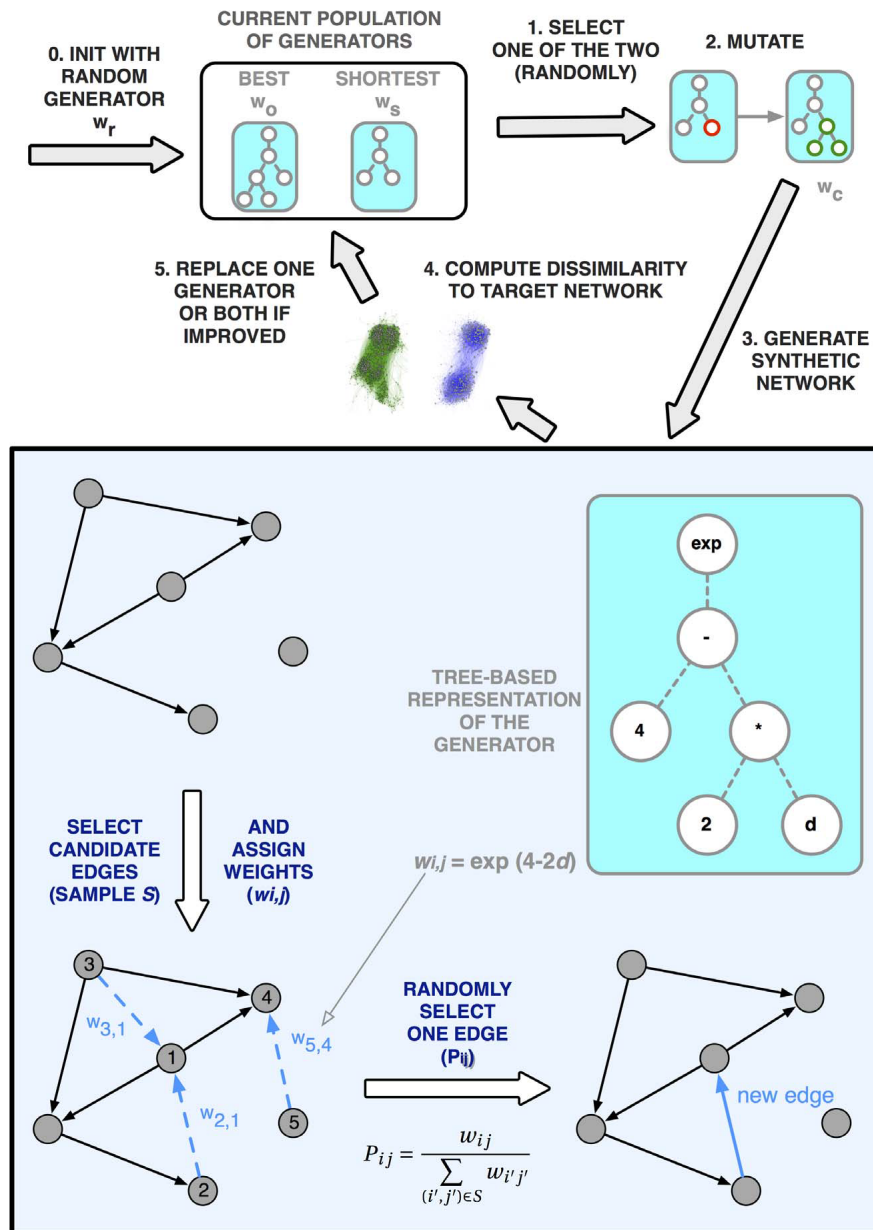


Figure 2.20: Schéma général de la méthode de génération de formule pour la probabilité de lien entre noeuds. La méthode travaille avec deux formules simultanément dans le but d'en extraire une efficace et réduite en taille. 0- Initialisation avec une formule aléatoire. 1- Choix de l'une des formules. 2- Mutation de la formule. 3- Génération d'un réseau synthétique 4- Calcul d'un score de distance 5- Remplacement des formules de génération si meilleurs résultats. Schéma issu de l'article (Menezes and Roth, 2014).

Discussion sur les méthodes de génération

Une utilisation des modèles les plus simples... Malgré l'extrême simplicité de certaines méthodes de génération (réseau aléatoire, réseau spatial), ou la grande simplicité d'autres (Small-World, Attachement Préférentiel), et de leurs résultats très qualitatifs, il s'avère qu'il s'agit des méthodes les plus utilisées. 70% de la communauté de la simulation sociale représentée dans un journal spécialisé (JASSS) les utilisent ([Amblard et al., 2015](#)).

Les raisons sont multiples. En l'absence d'information sur le processus sous-jacent à l'élaboration d'un réseau, le réseau aléatoire propose un modèle sans hypothèse. S'il s'agit davantage d'entités spatialisées, la grille régulière remplit ce rôle. Si les données sont fortement manquantes, voir quasi inexistantes, il est tout de même possible d'utiliser des modèles abstraits, qui assurent certaines caractéristiques communes à tous les réseaux sociaux. Cela permet de générer des réseaux de base. C'est le cas du modèle d'attachement préférentiel qui possède un ordre de grandeur correct de distribution de degré, et du modèle Small-world qui propose un clustering et une moyenne de taille de chemin qualitativement satisfaisants.

Une autre raison peut être que le modélisateur peut chercher à tester des hypothèses simples, et souhaite seulement avoir une idée qualitative des dynamiques.

Enfin, les académiques utilisent ces modèles depuis maintenant un certain temps, et leur intérêt pour des modèles plus poussés n'apparaît pas forcément.

...mais pas nécessairement pertinente Malgré l'apparente logique à utiliser des réseaux aux propriétés simples pour y jouer des processus simples, un contre-argument peut y être apporté. Dans le cas d'une approche topologique, les propriétés du réseau déterminantes pour le résultat d'une dynamique jouée ne sont pas nécessairement toutes maîtrisées ([Thiriot, 2010](#)) ([Menezes and Roth, 2014](#)).

Ainsi, ([Cointet and Roth, 2007](#)) ont joué une dynamique de propagation simple sur différents types de réseaux et ont comparé les résultats. Les réseaux générés par *Preferential Attachment* et *Small-World* ont des résultats bien plus proches du réseau aléatoire que ceux d'un modèle consistant à remélanger les liens en gardant les structures de clustering. Cointet et Roth avancent comme explication que ce phénomène est dû à des propriétés topologiques importantes pour le processus de propagation qui n'ont pas été prises en compte pour la génération des réseaux artificiels simples. En l'occurrence, le coefficient de fermeture de diamant, structure semblable aux triangles mais pour 4 nœuds, serait en cause.

Des modèles plus complexes Lorsque l'on commence à s'intéresser aux méthodes topologiques plus poussées, de nouvelles difficultés apparaissent. Ces méthodes peuvent être accompagnées d'une grande complexité d'utilisation.

méthode	Type d'approche	Précision des résultats	Complexité d'utilisation	Données nécessaires
Yang	Processus	\emptyset	++	+++
SW	Topologique abstrait	+	+	+
PA	Topologique abstrait	+	+	+
ERGM	Topologique statistique	+++	+++	+++
RS	Topologique probabiliste	++	++	+++

Table 2.6: *Résumé des principales méthodes de génération. Les appréciations de précisions et complexités sont données à titre indicatif. La précision des résultats de Yang dépend des règles utilisées.*

ERGM (Exponential Random Graph Model (Frank and Strauss, 1986) (Snijders et al., 2010)) est un outil statistique très élaboré, dont les résultats peuvent être très précis. Mais sa mise en pratique est complexe, demandant de solides connaissances en statistiques. Jusqu'à dernièrement, la méthode originale d'ERGM souffrait de dysfonctionnement qui la rendait non viable pour de nombreux cas réels. Malgré son amélioration récente pour être utilisée pour les réseaux sociaux, la méthode demande toujours de bonnes connaissances statistiques mais aussi de bonnes connaissances du modèle en lui même. Un cas particulier est le modèle de Régression Symbolique, qui permet de trouver automatiquement une formule de réseau possédant de bonnes propriétés. Ce modèle est simple à utiliser et donne de bons résultats topologiques. Néanmoins, à notre connaissance, il n'a fait que peu d'émules. Malgré un procédé ingénieux assurant des formules les plus succinctes possible, ces dernières restent assez verbeuses. Il est possible, même s'il ne s'agit pas d'un procédé visant à reproduire les actions des agents, que la difficulté d'interprétation puisse rebuter des utilisateurs ayant à cœur de "posséder" leurs outils.

Globalement, une plus grande précision sur les propriétés du réseau synthétique s'obtient au prix d'une plus grande difficulté d'utilisation du modèle de génération.

2.4.5 Conclusion sur la génération de réseaux sociaux

Deux approches sont disponibles pour générer des réseaux sociaux. L'approche processus identifie les actions des individus qui ont mené la création d'un réseau entre eux. Elles sont alors reproduites de façon plus ou moins détaillées dans une simulation. Cette approche demande une expertise du domaine concerné, et les réseaux obtenus sont rarement fidèles topologiquement. Ce type de modèle est souvent intégré dans une simulation sociale, leurs dynamiques se jouant simultanément.

L'autre approche propose de mimer topologiquement un réseau cible, sans nécessairement s'intéresser au processus de création sous-jacent. Pour peu que les propriétés choisies à reproduire soient pertinentes, les réseaux obtenus sont de bons candidats pour y jouer des dynamiques. Néanmoins, ces méthodes copient un réseau photographié à un moment précis, ne garantissant pas que les résultats soient toujours valables lorsqu'on s'intéresse aux évolutions temporelles.

Ces méthodes ne viennent pas sans inconvénient.

Concernant la généralité, Les approches processus sont souvent des solutions ad-hoc, ce qui empêche entre autre les comparaisons direct des résultats entre modèles. D'un autre côté, les approches topologiques génèrent essentiellement des réseaux de manière décorréliées des simulations ayant vocation à les utiliser. Cela qui pose alors un problème de correspondance (*mapping*) entre agents de la simulation et nœuds du réseau.

Un autre élément spécifique aux modèles topologiques est qu'ils fournissent un réseau statique. S'intéresser à un réseau après passage du temps ou autre perturbation demande donc de repasser par toutes les étapes du processus de génération.

Nous proposons une approche dynamique qui peut permettre l'étude de l'évolution d'un réseau après perturbation. La méthode est basée sur une simulation multi-agents, permettant une correspondance directe entre entités de simulation et nœuds du réseau. Les processus restent génériques, permettant l'usage de la méthode sur des réseaux de n'importe quel domaine. La validation du réseau généré est cependant topologique, simplifiant grandement cette étape.

Chapitre 3

Le Framework général

Sommaire

3.1	Les objectifs du modèle	60
3.1.1	Introduction	60
3.1.2	Objectif du Framework & des modèles	60
3.1.3	Élaboration d'un patron de solution	63
3.1.4	Résumé des objectifs	66
3.2	Le Framework	67
3.2.1	Le simulateur	67
3.2.2	Les dynamiques de construction de réseau	70
3.2.3	La recherche de réseau	73
3.3	Indicateurs du Framework	74
3.3.1	Propriétés observées	74
3.3.2	Éléments mis en avant par les observations	78
3.4	Expérimentations sur le Framework	79
3.4.1	Objectif des expérimentations	79
3.4.2	Test du simulateur	80
3.4.3	Reproduction d'un réseau sans échelle	89
3.4.4	Etude des dynamiques en fonction des paramètres de propagation	98
3.4.5	Questions sur la symétrie des actions	99
3.4.6	Reproduction d'un réseau Small-World	103
3.4.7	Conclusion	112
3.5	Conclusion du Framework et de ses expérimentations	113
3.5.1	Retour sur les objectifs	113
3.5.2	Instanciation en modèles	114
3.6	Résumé du chapitre	115

Maintenant que les notions importantes ont été présentées dans l'état de l'art, les trois chapitres suivants sont dévolus à la description de la solution proposée. **Le chapitre Framework va, depuis les objectifs auxquels nous souhaitons répondre, élaborer un patron de solution.** Les choix concernant son paramétrage seront pris lors de l'instanciation de ce Framework en deux modèles fonctionnels, dans les deux derniers chapitres.

Ce chapitre commence par **détailler les résultats attendus pour les modèles instanciés**, explicitant les contraintes définies par le sujet ainsi que par le cadre d'utilisation. **La construction d'un patron de solution, le Framework, est détaillée pas-à-pas.** Une fois ce cadre général établi, **des expérimentations y sont menées.** Les résultats de ces dernières vont guider l'instanciation du Framework pour obtenir deux modèles de génération de réseau, modèles qui seront détaillés dans les deux derniers chapitres (*chapter 4:Modèle I et chapter 5:Modèle II*).

- Le cadre d'utilisation et objectifs des modèles vont être décrits dans la *section 3.1:Les objectifs du modèle*
- le Framework -AL- qui en propose un patron de résolution est défini dans la *section 3.2:Le Framework*
- les expérimentations dans -AL- qui permettront ensuite son paramétrage et instanciation en modèle sont rapportées dans *section 3.4:Expérimentations sur le Framework*

3.1 Les objectifs du modèle

```

Framework
├─ Objectif : Explicitation des objectifs
├─ Mécanismes
├─ Indicateurs
└─ Expérimentations

```

3.1.1 Introduction

Le modèle a pour objectif de générer un réseau synthétique, reproduisant un réseau cible fourni par l'utilisateur. La construction dynamique du graphe se réalise lors d'une simulation multi-agents. Des règles simples de construction topologique, ajout et retrait de lien non directionnel, sont fournies et appliquées par les agents. Ces règles se diffusent parmi eux par un mécanisme de propagation. Deux dynamiques sont donc jouées conjointement : la construction du réseau par les règles et la diffusion de ces dernières.

Le fonctionnement de ces deux mécanismes sont dépendants de paramètres. Une recherche automatique des meilleurs paramètres est réalisée à l'aide d'un algorithme génétique, la fonction objectif étant un critère de similarité entre le réseaux cible et celui généré. Le détail de ce score de similarité est basé sur un ensemble de mesures topologiques (fig. 3.8). Au delà de la reproduction d'un réseau donné, on veut pouvoir étudier à terme l'évolution topologique après perturbations extérieures, survenant sous la forme d'ajouts ou de retraits de liens ou d'entités.

Dans un premier temps, nous allons voir dans le détail les objectifs du travail à réaliser.

3.1.2 Objectif du Framework & des modèles

Nous allons suivre une bonne pratique parfois utilisé dans le domaine du génie logiciel : diviser la structure d'un logiciel en deux parties. La première, le Framework, contient tous les éléments qui seront communs aux différentes instanciations. La spécification de ce Framework pour une utilisation donnée est un modèle. Le Framework -AL- vise à fournir un cadre général à partir duquel seront instanciés des modèles pour des utilisations particulières. -AL- définit donc un cadre, des méthodes et règles de fonctionnement génériques. Le modèle propose un paramétrage de ces méthodes et des actions spécifiques associées aux entités de la simulation.

L'objectif est de fournir un modèle, instancié depuis le Framework, permettant la génération de réseaux ayant :

- une expressivité forte, en terme de quantité de réseaux atteignables,
- tout en étant suffisamment stable pour qu'une configuration précise lancée plusieurs fois produise des résultats similaires.

L'expressivité du modèle concerne les types de réseaux reproductibles. L'idée est de pouvoir copier des réseaux précis - sans pour autant chercher à reproduire les processus réels sous-jacents à leur construction. Dans la réalité, les raisons exactes de l'existence d'un réseau sont souvent complexes, et multiples. Avoir une approche descriptive (KIDS - (Edmonds and Moss, 2005)), tout en souhaitant qu'elle puisse être appliquée à plusieurs types de réseaux semble très difficile. L'objectif de -AL- et de ses modèles est de reproduire n'importe quel réseau, en y appliquant la même méthode. Deux approches contraires ont été envisagées :

1. Réaliser une abstraction des processus sous-jacents à la création des réseaux réels, c'est à dire reproduire de manière abstraite les comportements sociaux identifiés comme générateur de réseau entre les individus
2. Utiliser une approche basée sur un processus de génération de réseau décorrélié des processus réels

Nous avons choisi de suivre la deuxième approche : le processus de création de réseau est propre au framework -AL- et a ses modèles, et n'entend pas reproduire une quelconque réalité. Cela permet entre autre plus de liberté et de contrôle sur ces processus. Une interprétation des résultats est ainsi laissée à la discrétion de l'utilisateur.

À ce besoin d'expressivité s'ajoute l'objectif de reproductibilité des résultats. On désire, à paramétrage égal, obtenir un réseau de même famille en terme de propriété.

Dans le cadre d'un système complexe comme un système multi-agents, des effets chaotiques peuvent rapidement survenir. En effet, une approche stochastique est communément utilisée par ces systèmes décentralisés. Afin de permettre la reproduction des résultats, il est nécessaire de minimiser ces aspects. L'idéal est de travailler dans un système continu au sens où une petite modification des paramètres initiaux amène une modification de même amplitude pour les propriétés du réseau obtenu.

On souhaite donc une faible sensibilité aux paramètres sans pour autant exclure l'existence d'effet de seuil. C'est par exemple le cas dans les modèles de percolation où la probabilité de connexion entre nœuds admet une valeur critique à partir de laquelle une composante géante se forme dans le réseau (Albert and Barabási, 2002).

Idéalement, le changement d'échelle doit aussi être possible. Un bon nombre des propriétés mesurées sur un réseau ne dépend pas du nombre de nœuds total (densité, coefficient de clustering, ...). On souhaiterait donc conserver les valeurs de ces propriétés lorsqu'une même configuration est jouée sur un réseau possédant plus de nœuds.

Les modèles instanciés depuis -AL- doivent aussi être facile d'utilisation : les utilisateurs de Genstar (Drogoul, 2013) venant de différents domaines, les pro-

jets de simulation sociale peuvent regrouper différents acteurs aux spécialités diverses (géographes, ethnologues, historiens...). Les utilisateurs de nos modèles ne doivent donc pas avoir besoin :

- de connaissances métier du domaine dont est issu le réseau à reproduire,
- de connaissances approfondies en informatique et
- de connaissances en mathématiques

Les données que les utilisateurs fournissent sont réduites à un fichier décrivant les nœuds et sommets d'un réseau cible à reproduire. L'indépendance aux données métiers est facilitée par une approche processus décorrélée du processus réel ayant dicté la construction du réseau cible.

En plus de l'utilisation d'un réseau, le projet Genstar s'intéresse aussi à l'étude de l'adaptation de ce dernier après perturbation. Dans ce cadre, une fois le modèle défini, il faudrait pouvoir perturber le système par une perte ou ajout de liens ou de nœuds, et observer la dynamique de reformation d'un réseau. Cela fournirait alors une hypothèse sur la façon dont le système représenté par le réseau cible se comporterait.

Enfin, sur un plan technique, la conception et le développement du Framework se veulent les plus modulables et génériques possible, afin entre autres de pouvoir affiner sa spécification en fonction des expériences menées.

Pour résumer, nous cherchons à créer un Framework et à implémenter un modèle depuis ce premier réunissant les propriétés suivantes :

A-Répétition des résultats possibles permettant qu'une configuration lancée plusieurs fois engendre les mêmes familles de réseaux en résultat.

B-Espace des réseaux atteignables large afin d'assurer la possibilité de générer un large panel de réseaux.

C-Peu de configuration côté utilisateur pour s'assurer une utilisation ne nécessitant aucune connaissance technique poussée.

D-Espace des paramètres internes au modèle réduit afin de permettre une recherche automatisée efficace de résultats.

E-Continuité c'est-à-dire avoir une monotonie de l'évolution des formes de réseaux obtenues en fonction de légères modifications dans les paramètres.

F-Indépendance aux premières actions prises que le résultat final ne soit pas dépendant de l'ordre des actions.

G-Etude de la résilience que le modèle donne la possibilité d'étudier l'évolution du réseau après perturbation.

Maintenant que les objectifs ont été établis, nous allons détailler l'élaboration d'un patron de solution.

3.1.3 Élaboration d'un patron de solution

Le choix des mécanismes de simulation de la plateforme -AL- visent à résoudre au mieux les contraintes. Par exemple, la contrainte d'étude de l'évolution après ajout ou retrait de nœuds dans le réseau va être adressée par l'utilisation d'un phénomène de propagation.

Une fois les principes des mécanismes définis, leurs réglages définitifs se situent côté modèles : un modèle particulier est alors une instanciation possible du Framework.

Remplir les objectifs définis dans la section précédente demande des choix de construction et des paramétrages à différents niveaux : dans la définition des mécanismes généraux du Framework (section 3.2) mais aussi dans leur paramétrage au niveau des modèles (chapitres suivants, *chapter 4:Modèle I* et *chapter 5:Modèle II*). On peut donc ici distinguer deux types de paramètres :

- les paramètres du modèle : il s'agit des choix de configuration du Framework, permettant son instanciation en modèle (nombre d'itérations d'une simulation, détails des mécanismes de propagation...)
- la configuration initiale d'une simulation : il s'agit des paramètres d'entrée pour ce modèle qui, une fois le processus de simulation appliqué, génèrent un réseau synthétique. C'est sur ces paramètres qu'est conduite une recherche automatique afin de générer un réseau synthétique semblable à celui fourni en paramètre par l'utilisateur. Concrètement, ces paramètres concernent la définition des complexe-actions et de leurs probabilités de propagation.

Des objectifs au patron d'une solution Nous allons ici déterminer les principaux mécanismes de fonctionnement du Framework. La méthode de génération de réseau que l'on conçoit recherche l'équilibre entre deux tendances :

Ouverture : pour générer un grand nombre de formes de réseaux,

Contrainte : pour reproduire les résultats d'une génération à l'autre.

Système multi-agents La modélisation agent permet de faire **une correspondance directe entre les agents et les nœuds du réseau à générer**. L'ajout et le retrait de liens et de nœuds pour la création et perturbation du réseau sont transcrites directement par les comportements agents (*G-Etude de la résilience*). Les agents de la simulation portent des actions qui vont leur permettre d'interagir entre eux.

Propagation des comportements La création d'un réseau au cours de la simulation va dépendre d'actions réalisées par les agents. Chacun d'entre eux

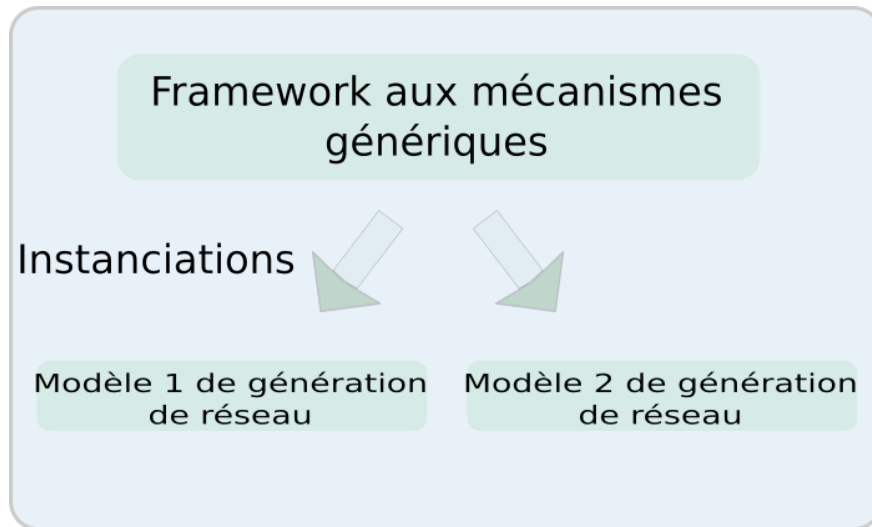


Figure 3.1: *Instanciation des mécanismes et paramètres du Framework pour l'obtention d'un modèle particulier de génération de réseau.*

dispose d'un ensemble d'actions possibles.

Un problème que l'on retrouve fréquemment dans les systèmes complexes - et à fortiori les multi-agents - est l'importance des premières actions entreprises (effet cascade, comportement chaotique). Afin de limiter l'importance de l'ordre des premières actions réalisées par les agents, **l'objectif est de concevoir un système qui tend vers une forme dynamique d'équilibre**, un état final stable.

Afin d'obtenir d'autres formes de réseaux que les aléatoires (*B-Espace des réseaux atteignables large*), **des comportements singuliers doivent être joués**. Cela peut se traduire par une répartition dissymétrique d'actions, ou par des actions aux effets dissymétriques.

Définir les comportements de chaque agent ouvre un espace combinatoire d'ordre exponentiel en nombre de nœuds dans le réseau. **Un mécanisme de propagation de comportements est donc mis en place** afin de respecter (*D-Espace des paramètres internes au modèle réduit*). Une configuration initiale de simulation peut alors se résumer à un ensemble d'actions disponibles et à leurs conditions de propagation.

Les mécanismes & interactions, à paramétrage identique, doivent permettre de tendre vers la construction de réseaux similaires (*F-Indépendance aux premières actions prises* & *A-Répétition des résultats possibles*). On pourra ensuite perturber l'équilibre atteint pour étudier la résilience aux perturbations (*G-Etude de la résilience*). Le mécanisme de propagation des actions entre les agents nous assure aussi qu'en cas d'ajout de nœuds dans le réseau, ces derniers acquerront aussi des comportements à exécuter.

On pourrait être tenté de définir un nombre maximum d'apprentissage d'ac-

tions pour les agents. La stabilité du réseau final serait atteinte plus rapidement car seules les dynamiques d'applications d'actions seraient encore jouées. Néanmoins, on se retrouve alors dans le cas où les actions initiales sont les plus appliquées et les plus apprises. Autrement dit, la sensibilité aux premières actions entreprises serait trop forte (*F-Indépendance aux premières actions prises*).

Patron pour les paramètres Dans la mesure du possible, il faut que les paramètres soient définis sur des valeurs continues et non sur des valeurs discrètes comme les booléens, étant plus propices aux effets de seuil (*E-Continuité*).

Patron pour les actions Afin de faciliter l'expressivité des réseaux atteignables, des actions modulables ou facilement modifiables seraient idéales (*B-Espace des réseaux atteignables large*).

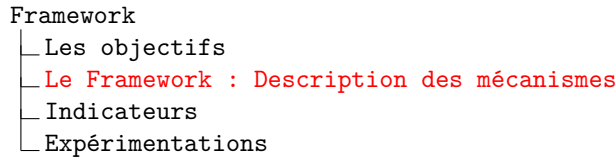
3.1.4 Résumé des objectifs

En partant du sujet de thèse et du cadre d'utilisation, **un ensemble de contraintes ont été établies**. L'environnement doit être suffisamment "ouvert" pour permettre la génération de réseaux aux profils variés, mais suffisamment **contraint** pour qu'une même configuration simulée plusieurs fois donne des résultats du même ordre. Nous nous proposons d'y répondre en deux étapes : **La création d'un patron de solution**, le Framework, **puis l'instanciation** de ce Framework en modèles de génération de réseaux spécifiques. La réponse apportée aux contraintes se trouve répartie entre

- les concepts et mécanismes de fonctionnement du Framework et
- la paramétrisation de ces derniers au niveau modèle.

La prochaine section va décrire le Framework -AL-.

3.2 Le Framework



Cette partie va décrire le Framework -AL-, **mécanismes généraux de la solution élaborée**. Il s'agit donc du cadre général dans lequel les expériences du chapitre suivant (*Expérimentations sur le Framework*) seront menées, permettant la définition ultérieure d'un modèle.

L'élaboration d'-AL- est guidée par les contraintes vues en *Les objectifs du modèle*, et entend adresser un maximum d'entre elles avant de laisser l'instanciation en modèle finir d'y répondre. La plupart des éléments qui s'y trouvent ont pu être déterminés par construction théorique. L'instanciation en modèle fonctionnel consistera alors en un paramétrage et des réglages fins de ce cadre. **Les choix définitifs sont donc réalisés et explicités par la mise en modèle de -AL- dans les chapitres suivants.**

3.2.1 Le simulateur

Le principe général du système est de **simuler des interactions entre agents, amenant la création d'un réseau synthétique**. Les agents -ou entités- représentent les nœuds du réseau par association directe. De même, les liens entre ces premiers sont des liens du réseau entre les sommets correspondants. Les agents sont dotés de **complexe-actions** ; **l'application de ces derniers permettent la création ou suppression des liens**. Les complexe-actions s'appliquent en fonction de contraintes sur les attributs des entités et peuvent se propager entre les agents. Ces éléments sont vus plus en détail dans la partie *Les complexe-actions*. **Un score de distance entre le réseau** à reproduire fourni par l'utilisateur et le réseau en cours de construction dans la simulation multi-agents est calculé, et est pris en compte lors d'une recherche automatique des meilleures configurations.

Avant de voir dans le détail les différentes parties composant le simulateur, nous donnons une idée générale de son fonctionnement.

1. *Sélection d'une entité*
2. *Sélection d'un complexe-action porté par cette entité*
3. *Recherche d'une entité répondant aux critères pour l'application de l'action portée par le complexe-action*
4. *Application de l'action d'ajout ou de retrait de lien*

La spécification de nombreux éléments est propre au modèle instancié depuis le Framework : réseau de départ, actions disponibles pour les entités, la méthode de propagation, condition d'arrêt...

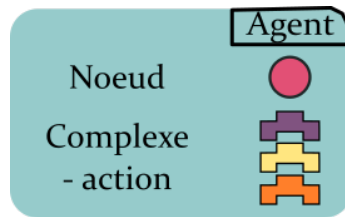


Figure 3.2: Un agent, associé à un nœud et porteur d'un ensemble de complexe-action

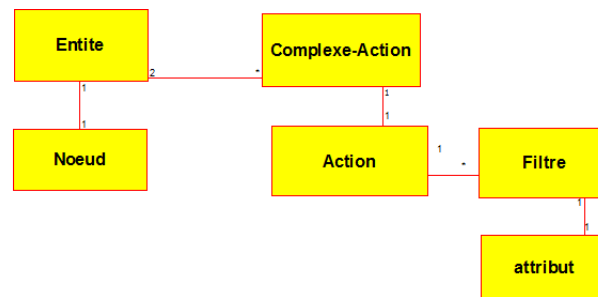


Figure 3.3: Diagramme de classe pour les entités et complexe-actions.

Ordonnanceur

La simulation commence avec un nombre d'agents identique au nombre de nœuds dans le réseau cible. Les complexe-actions disponibles dans la simulation sont distribués aléatoirement parmi les agents. A chaque étape, le système va choisir un agent au hasard, de manière équiprobable. **Un complexe-action porté par cet agent est choisi et est appliqué.** Le système réitère ce procédé jusqu'à ce que l'arrêt de la simulation soit décidé. Le réseau obtenu peut alors être évalué en terme de distance avec un réseau défini comme objectif.

Un agent

Les nœuds du réseau sont les agents du système. Ils peuvent être porteurs d'attributs, calculés (degré) ou non (age, ..). Ces attributs doivent être comparables pour une utilisation par les complexe-actions (figure 3.2).

Les complexe-actions

Les complexe-actions sont détenus par les agents du réseau, leur permettant d'agir dans la simulation. Un complexe-action est composé de plusieurs éléments (figure 3.6). Une *action*, et un ensemble de *filtres* portant sur un *attribut*. Le diagramme 3.3 résume cela. Les entrées et sorties des filtres sont une liste d'agents. L'action portée par le complexe-action sera appliquée sur l'ensemble des agents restants après l'application de tous les filtres. Cet ensemble peut être vide, l'action ne sera dans ce cas appliquée à aucun agent.

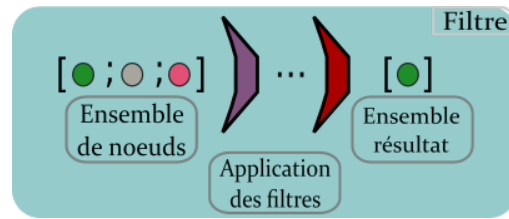


Figure 3.4: Application de filtres sur des ensembles de noeuds



Figure 3.5: Symbolisation d'action de modification topologique du réseau

Attributs Une entité peut porter n'importe quel nombre d'attributs. Ces derniers doivent être comparables, afin de pouvoir y appliquer des relations d'ordres ($>$, $=$, \dots), utilisée par les filtres. Les attributs peuvent aussi être des propriétés calculées, telles que le degré du nœud.

Filtres Le principe des filtres est de renvoyer un sous-ensemble des entités reçues en paramètres (figure 3.4). Un filtre \mathcal{F} prend en paramètres :

- une entité à partir de laquelle faire les comparaisons,
- une liste d'entités sur lesquelles appliquer les comparaisons,
- un opérateur mathématique de comparaison (supérieur à, égale à..) et
- un attribut sur lequel porte l'opérateur mathématique.

\mathcal{F} renvoie un sous-ensemble des entités de la liste en entrée. Plusieurs filtres peuvent être appliqués successivement.

Actions Les actions sont des éléments de modification topologique du réseau (figure 3.5) : ajout ou retrait de lien, ajout ou retrait de nœuds.

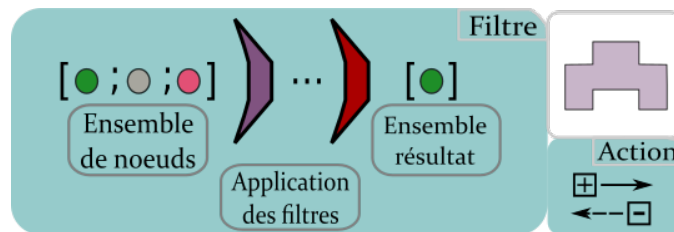


Figure 3.6: Complexe action, composé d'une série de filtres appliqués sur un attribut. L'application des filtres réduit l'ensemble des agents sur lesquels l'action d'ajout ou de retrait de liens va être jouée. Les agents restants après application de tous les filtres se voient appliquer l'action

Mécanisme de propagation

Afin de **favoriser les dynamiques d'adaptation** aux perturbations du réseau **tout en réduisant la taille de l'espace de configuration** initial, un mécanisme de propagation des complexe-actions est utilisé (figure 3.7). **Un complexe-action est transmis de l'entité agissante vers celle sur laquelle est appliquée l'action** selon certaines conditions. Les variations possibles pour ce mécanisme sont nombreuses, les détails seront vus dans les *section 3.4:Expérimentations sur le Framework* et dans les chapitres suivants.

Les mécanismes de propagation et d'application d'actions sont joués de manière concomitantes : le but est de coupler leurs dynamiques pour tendre vers des solutions stables.

Les paramètres du Framework

On distingue deux types de paramètres :

- les paramètres qui permettent d'**instancier le Framework en modèle** : le type de mécanisme de propagation, la condition d'arrêt pour les simulations, les propriétés prises en compte pour définir la distance entre deux réseaux...
- **la configuration pour une simulation donnée** : liste des complexe-actions présents, leurs probabilités de transmission.

Maintenant que les éléments du Framework ont été présentés, nous allons nous intéresser aux dynamiques qui s'y jouent.

3.2.2 Les dynamiques de construction de réseau

La dynamique de construction du réseau peut être divisée en trois axes, intervenant de manière concomitante :

- **la propagation des actions sur l'ensemble** des entités du réseau,
- une **dynamique de répartition des actions** parmi les entités et
- **la construction topologique du réseau** issu de l'application des actions.

Le but de l'outil que nous concevons est de générer des réseaux grâce à des dynamiques topologiques. Passée l'étape de propagation des comportements sur l'ensemble des entités, leurs applications feront tendre le réseau vers une forme topologique, idéalement stable. Théoriquement, il est possible que les dynamiques jouées arrivent à un point fixe, s'arrêtant sur un réseau pertinent. Une fois ce réseau atteint, plus aucune action ou propagation n'y serait jouée. En pratique, à l'exception des cas limites des réseaux vides et complets, ces situations sont très difficilement atteignables et reproductibles. De plus, un état final statique pourraient empêcher l'étude des perturbations dans les réseaux.

Par ailleurs, en fonction des règles de propagation, certaines entités pourraient ne jamais être atteintes par des complexe-actions. Cette situation est à éviter car elle pourrait amener une grande sensibilité aux premières actions entreprises par les entités.

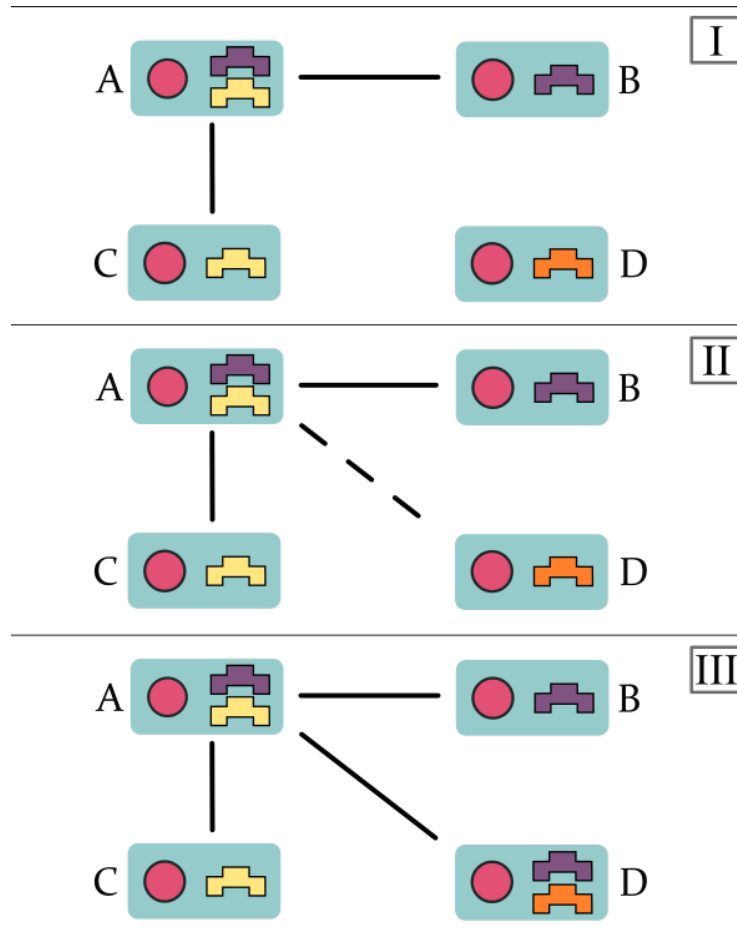


Figure 3.7: Exemple d'application de complexe-action I - Sur un réseau constitué de quatre entités-noeuds [A;B;C;D]. II - l'entité A agit, applique son complexe-action et III - la transmet à l'entité D.

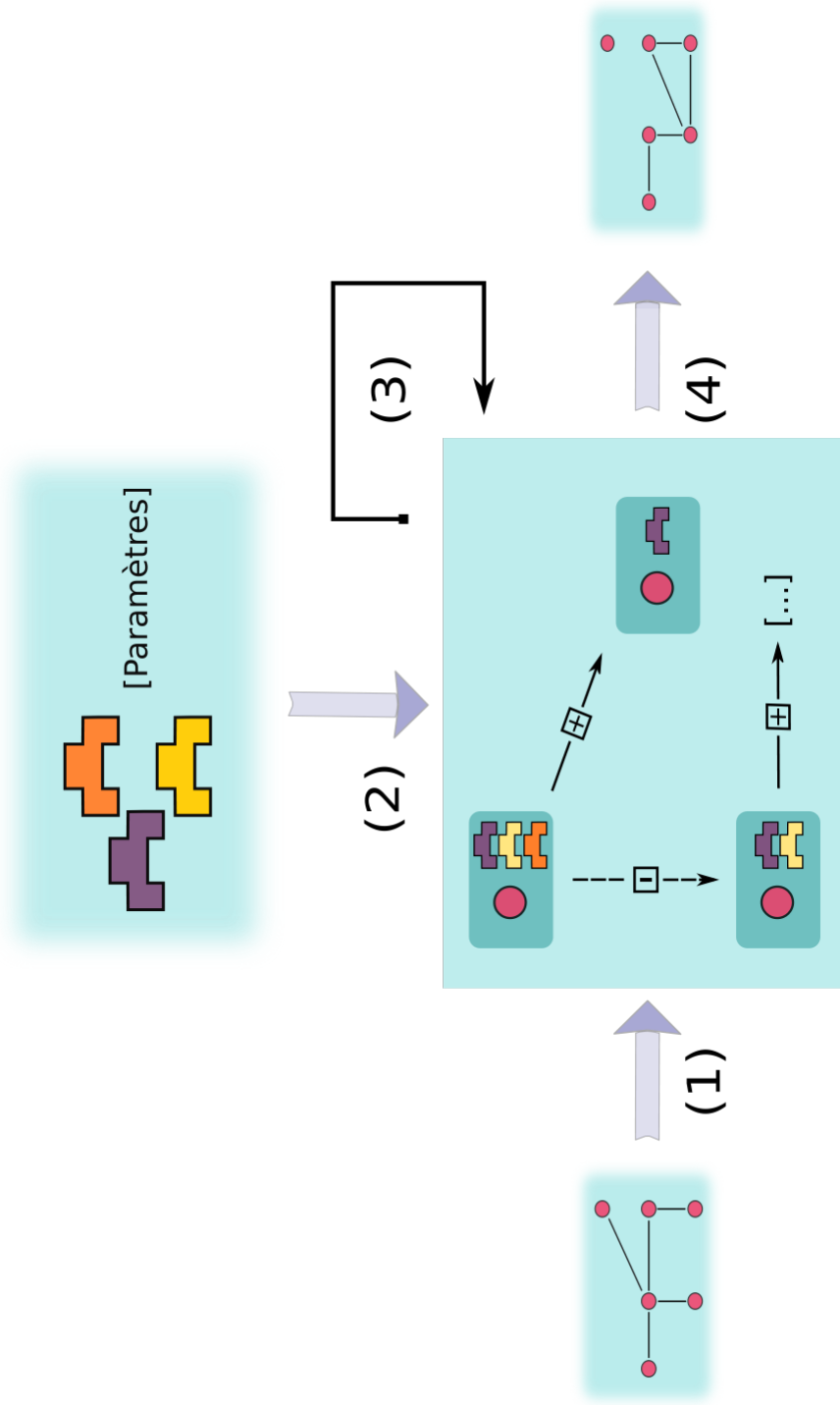


Figure 3.8: Fonctionnement général d'une simulation (1) Entrée utilisateur : Réseau à reproduire. (2) Entrée du modèle : configuration. (3) - Simulation jusqu'à stabilité ou sur un nombre d'itérations donné. (4) - Sortie : Réseau obtenu avec cette configuration.

Pour ces raisons, **on cherche à atteindre une dynamique stable de propagation de comportement.**

De part le choix d'un système de propagation (cf. *section 3.2.1: Mécanisme de propagation*), l'application des actions et la répartition des complexe-actions sont intrinsèquement liées. A priori, les dynamiques les plus prometteuses semblent être celles avec une relative stabilité de la répartition des actions sur les entités.

La sous-section suivante détaille les mécanismes permettant la recherche de solutions dans le cadre de ce Framework.

3.2.3 La recherche de réseau

Les configurations initiales des simulations définissent la forme des réseaux atteignables. Une méthode de recherche automatique de ces paramètres est utilisée.

Fin d'une simulation

Avant de pouvoir rechercher automatiquement la configuration initiale optimale, il faut être en mesure de déterminer si la dynamique de construction du réseau possède un point de convergence, et si elle l'a atteint. L'idée initiale était d'attendre la stabilisation de la dynamique de construction du réseau. L'usage (*section 3.4: Expérimentations sur le Framework*) montrera que ça ne sera pas suffisant. **Est donc fixé un nombre d'actions réalisées** au delà duquel la simulation est arrêtée.

Score

Afin de permettre une évaluation des réseaux obtenus par la simulation, un **score de distance entre le réseau cible en entrée et le réseau synthétique obtenu est calculé**. L'objectif de recherche automatique est de trouver la configuration d'une simulation minimisant ce score. Quelque soit l'ensemble des propriétés observées, le score de chacune d'entre elles sur les deux réseaux est normalisé. Cela va permettre d'en agréger n'importe quel nombre en leur donnant la même importance.

Méthode de recherche automatique

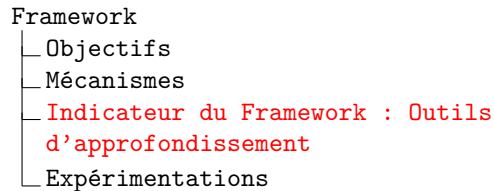
Plusieurs méthodes de recherche automatique de configurations ont été mises en place.

- Recherche exhaustive dans l'espace de paramètres
- Recherche aléatoire sur ce dernier
- Recherche par l'utilisation d'un **algorithme génétique**, prenant le score de distance entre les réseaux comme fonction d'évaluation (*fitness*).

Les détails seront vus dans le chapitre suivant.

La section suivante détaille les outils qui ont été développés afin de comprendre et qualifier les dynamiques et réseaux synthétiques obtenus.

3.3 Indicateurs du Framework



Nous allons nous intéresser aux indicateurs utilisés lors des simulations pour en comprendre les résultats.

Afin de pouvoir analyser les résultats des simulations, plusieurs outils et indicateurs ont été développés. Certaines observations sont réalisées à un moment précis, d'autres sur des intervalles de temps. Les propriétés du réseau peuvent être observées :

- de manière ponctuelle, pour qualifier l'état du réseau à un moment précis,
- de manière dynamique, afin d'en étudier l'évolution temporelle,
- de façon répartie sur plusieurs itérations d'une configuration, observant les moyennes et écarts-types de valeurs.

La sous-section *Propriétés observées* va détailler les propriétés observées par les indicateurs, et la sous-section *Éléments mis en avant par les observations* les interprétations au sujet du système.

3.3.1 Propriétés observées

Propriétés du graphe

Il est possible de calculer sur le réseau (ou graphe en tant qu'objet mathématique) un ensemble de propriétés. Ces mesures peuvent être observées à un instant donné afin de qualifier le réseau ponctuellement, ou à intervalles réguliers afin d'étudier des phénomènes dynamiques.

Les propriétés topologiques du graphe auxquelles nous nous intéressons sont les suivantes :

- **Densité** : une stabilité de densité indique un volume stable de créations de liens par rapport à leur destruction.

$$\frac{\text{Nombre de liens du graphe}}{\text{Nombre de liens maximum possible}}$$

- **Distribution de degré** : le tableau de distribution informe sur la répartition des liens entre les nœuds (figure 3.9). Certains types de réseaux ont des signatures particulières. Si les créations des liens du réseau suivent une même loi de probabilité, la répartition de leur degré est une loi normale, dont la moyenne est la densité du réseau. Il s'agit alors de réseau aléatoire.

Les réseaux sans échelle (Scale-Free) ont une distribution caractérisée

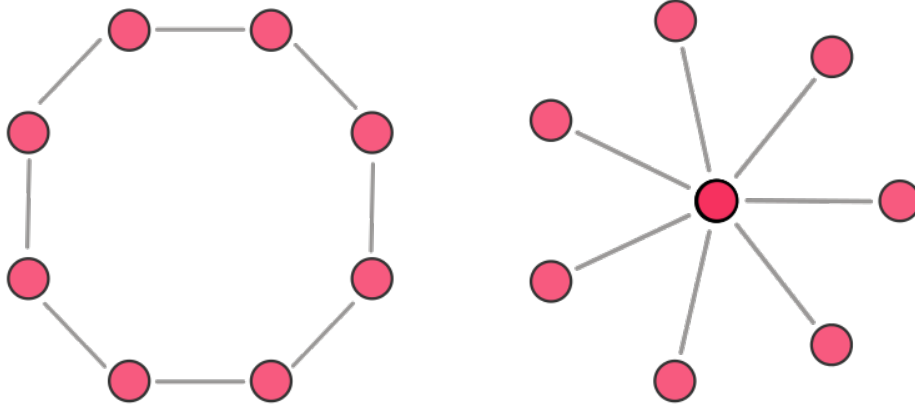


Figure 3.9: Ces deux réseaux possèdent le même nombre de liens mais des distributions de degrés très différentes.

par une asymétrie positive à lourde queue *right skewed distribution with heavy tail*. Il est possible de s'intéresser à la manière dont diminue le nombre de nœuds ayant un degré de plus en plus grand. Le coefficient γ définit le rapport de proportionnalité entre degré et nombre de nœuds ayant ce degré par la formule $P(k) \propto k^{(-\gamma)}$. La valeur de γ observée dans les réseaux sans échelle varie entre 2 et 3. Le tableau 3.1 nous montre les bornes du nombre de nœuds par degré dans un réseau à 1000 nœuds.

- **Le coefficient de clustering moyen** est la moyenne du clustering de chaque nœud du réseau. Dans un réseau aléatoire, ce coefficient est très proche de la densité, mais des formes particulières d'arrangements topologiques amènent des différenciations entre ces deux grandeurs.
- **La moyenne des plus courts chemins APL-Average Path Length** est un indicateur très important pour les réseaux de type Small-World. Pour ce type de réseaux sociaux, l'APL est faible et évolue en $\log(\text{Nombre de nœuds})$. Cela est dû aux raccourcis créés entre les groupes de nœuds fortement connectés entre eux.

Indicateurs associés à ces propriétés

Les actions

L'évolution topologique du graphe dépend des actions qui y sont jouées. **On distingue l'appel d'un complexe-action et l'application effective dans le graphe de l'action portée par ce premier.** La dynamique topologique dépend uniquement de la réussite d'application des actions portées par le complexe-action. Le détail de fonctionnement des actions est explicité en

Degree	$p(k) \propto k^{(-2)}$ normalisé	Nb nœuds pour réseau 1000 nœuds	$k^{(-3)}$ normalisé	Nb nœuds réseau 1000 nœuds
1	0.649	649.45	0.835	836
2	0.16	162.4	0.10	104.47
3	7.22E-2	72.16	3.09E-2	30.95
4	4.06E-2	40.59	1.31E-2	13.06
5	2.6E-2	25.98	6.68E-3	6.69
6	1.80E-2	18.04	3.87E-3	3.87
7	1.32E-2	13.25	2.44E-3	2.44
8	1.015E-2	10.15	1.6E-3	1.6
9	8.02E-3	8.02	1.15E-3	1.15
1000		[...]		

Table 3.1: Nombre de noeuds par degré pour des réseaux sans échelle théoriques à 1000 nœuds. La distribution est donnée pour les valeurs γ limites de 2 et 3.

section 3.2.1: Les complexe-actions.

Appel des complexe-actions Le choix de l'entité qui agit et du complexe-action porté par cette dernière est **aléatoire et équiprobable**. (figure 3.10)

La répartition des actions Le choix de l'entité agissante est aléatoire. Par conséquent, **la répartition des complexe-actions parmi les entités est le seul facteur qui influence la probabilité de l'appel** d'une action en particulier. Observer leur répartition à un moment donné ainsi que l'évolution de cet indicateur est donc fondamental (figure 3.11).

Échecs des applications des actions En plus des informations sur les appels aux complexe-actions, les échecs d'applications de ces dernières permettent de mieux comprendre la dynamique du réseau. En fonction des *filtres* portés, il n'est pas certain qu'il soit possible d'appliquer l'action d'un complexe-action. À titre d'exemple, une entité connectée à toutes les autres entités du réseau n'aura pas de cible pour jouer l'action "ajout de lien".

Indicateurs associés aux actions Les successions d'appel des complexe-actions permettent d'expliquer les formes topologiques obtenues. Les indicateurs sont les suivants :

Pour l'appel aux complexe-actions :

- Ratio des appels de chaque complexe-action depuis le début de la simulation, ainsi que sur fenêtre glissante.
- Ratio des applications des actions associées (ajout de lien, retrait) depuis le début de la simulation, ainsi que sur fenêtre glissante.

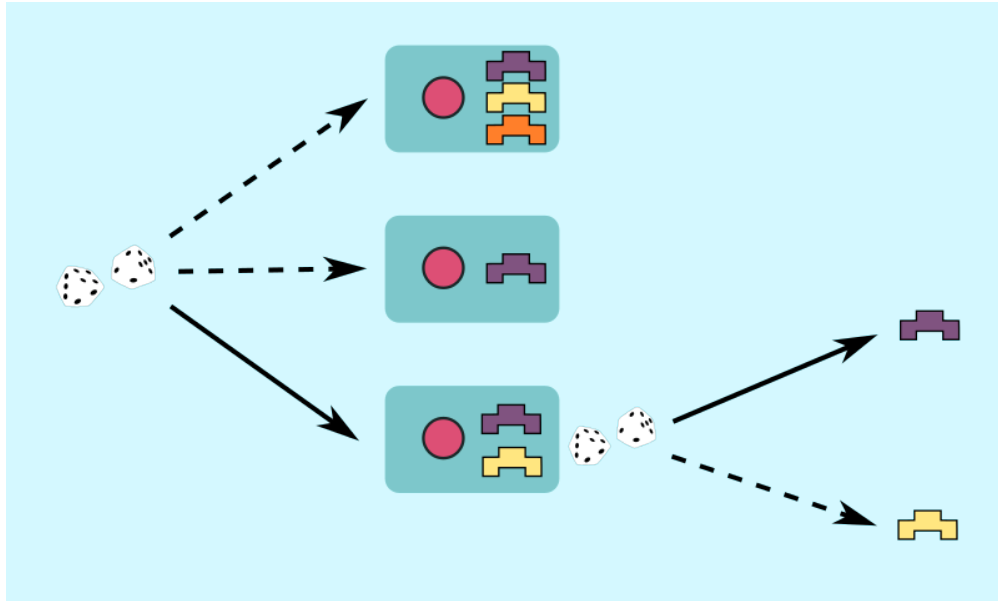


Figure 3.10: *Processus de sélection d'un complexe-action à appliquer, répété à chaque itération de la simulation. Le choix d'une entité, ainsi que celui du complexe-action à appliquer est aléatoire.*

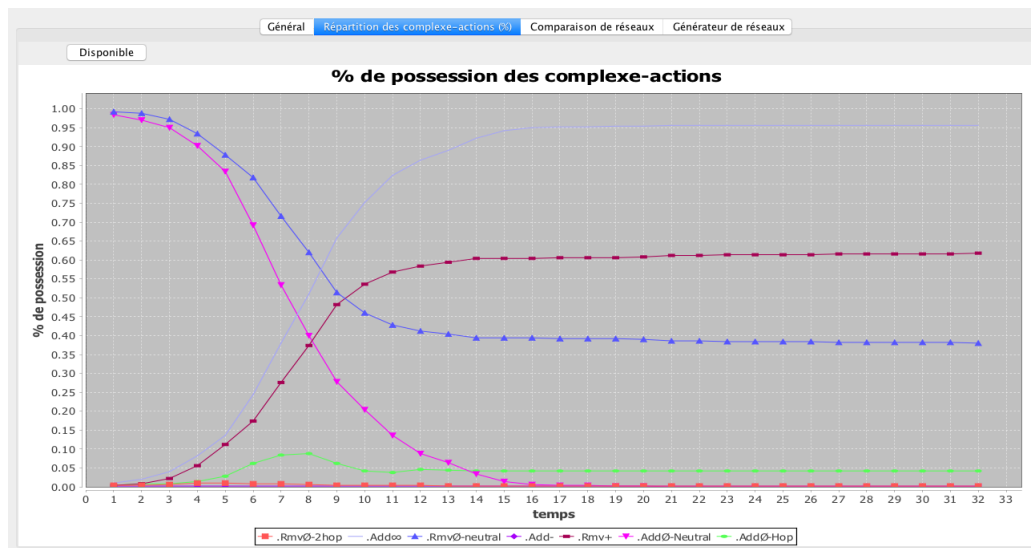


Figure 3.11: *Indicateur représentant l'évolution de la répartition des complexe-actions parmi les entités de la simulation*

Concernant l'appel aux actions :

- Compte et succession des actions appelées depuis le début de la simulation.
- Compte et succession des actions appelées sur une fenêtre glissante des X dernières actions appliquées.

Pour les échecs d'application des actions :

- Les échecs d'application, par action et type d'action (ajout et retrait).

3.3.2 Éléments mis en avant par les observations

Ces indicateurs permettent de mettre en lumière le fonctionnement d'un modèle (en tant qu'instance du Framework -AL-) sur différents points.

Espace des réseaux atteignables Les indicateurs concernant les propriétés du réseau synthétique obtenu permettent de qualifier les réseaux atteignables. Il s'agit d'une photographie au moment où la simulation s'arrête.

Stabilité Les indicateurs présentant l'évolution de certaines propriétés telles que la densité ou la répartition des complexe-actions permettent de vérifier si la forme du réseau obtenu à la fin de la simulation est stable.

Reproductibilité Le fait de lancer plusieurs fois la même configuration et de s'intéresser aux moyennes et aux écarts-types des propriétés obtenues permet de connaître la reproductibilité d'une configuration donnée.

Continuité des paramètres & comportements non chaotiques Faire varier progressivement la valeur des paramètres initiaux et observer la variation des résultats nous permet d'estimer la sensibilité du modèle aux conditions initiales.

La prochaine section va décrire les expérimentations qui ont été menées dans le cadre du Framework.

3.4 Expérimentations sur le Framework

Framework
_ Objectifs
_ Mécanismes
_ Indicateurs
_ Expérimentations sur le Framework

Dans cette partie, nous implémentons le Framework en modèle temporaire afin de tester certains paramétrages. Les expérimentations qui vont y être menées permettront ensuite de faire des choix d’instanciation du Framework en un modèle particulier. Elles vont se dérouler en deux étapes.

Premièrement, **nous allons chercher à reproduire certains réseaux** issus de modèles abstraits, tel qu’attachement préférentiel ou Small-World. Cela va amener à l’élaboration de complexe-actions permettant l’apparition de propriétés :

- Attachement préférentiel : distribution de degrés suivant une loi de puissance
- Small-World : haut clustering et plus court chemin moyen (APL) entre nœuds d’ordre $\log(\text{Nombre de nœuds})$.

Ensuite, nous utiliserons un algorithme génétique afin d’explorer rapidement les meilleures configurations pour la génération d’autres types de réseaux.

Dans la mesure du possible, les expérimentations sont présentées dans l’ordre où elles ont été menées, permettant de suivre la logique d’élaboration.

3.4.1 Objectif des expérimentations

Ces expérimentations permettent d’acquérir des intuitions sur le fonctionnement du Framework et de ses futurs modèles, de développer et d’affiner les *Indicateurs du Framework*, et de calibrer certains paramètres majeurs :

- *Création des actions à disposition* On cherche un **ensemble réduit de complexe-actions** permettant d’obtenir une grande variété de réseaux. Cela évite l’augmentation inutile de la complexité, qui pénaliserait les recherches automatiques de configurations initiales.
- *Condition de propagation des actions* Le mécanisme de propagation des complexe-actions entre les entités peut être instancié de nombreuses manières. On cherche une méthode de propagation permettant d’atteindre un grand nombre de réseaux, tout en assurant une reproductibilité des résultats.
- *Condition d’arrêt d’une simulation avant calcul des propriétés du réseau obtenu.* La convergence vers une forme de réseau stable malgré les dynamiques qui s’y jouent est assez rare : la plupart des configurations n’aboutissent pas à un état final stable, et ont des variations de propriétés fortes. En plus de l’état de dynamique stable pris dans un attracteur que nous recherchons, une troisième possibilité existe. Il s’agit d’un état statique

où plus aucune action ne peut être jouée - situation retrouvée le plus souvent avec des réseaux vides ou complets -.

Dans ces trois cas, il faut être en mesure d'automatiser la décision de fin de simulation.

- *Propriétés à prendre en compte pour le calcul de la distance entre deux réseaux.* Il faut définir un ensemble de propriétés du réseau cible pertinent à reproduire. De ce choix va dépendre le bon comportement du réseau lors de ses utilisations ultérieures.

Par souci de clarté, et parce que nous jugeons ces éléments non pertinents, le développement et l'évolution des outils de mesures ne seront pas détaillés. Ces derniers ont donc entièrement été présentés dans la section précédente, *section 3.3:Indicateurs du Framework*, dans leur version aboutie.

Plan des expérimentations Nous commencerons par tester le simulateur sans complexe-action particulier, afin de s'habituer aux outils et de dégager une condition d'arrêt satisfaisante.

Ensuite, nous chercherons à reproduire un réseau sans échelle. Cela va engager l'utilisation de plusieurs complexe-actions portant la même action à appliquer, en l'occurrence l'ajout d'un lien. L'intérêt de l'utilisation de différentes probabilités de propagation apparaît alors. De plus, la nécessité de l'utilisation de complexe-actions temporaire pendant la simulation fait jour.

Après une interrogation sur l'aspect symétrique de certains complexe-actions, nous nous attellerons à reproduire un réseau Small-world, qui mettra en avant la difficulté de sortir de l'ensemble des réseaux aléatoires.

3.4.2 Test du simulateur

Dans un premier temps, nous allons nous familiariser avec les outils du Framework, les indicateurs à disposition, et configurer certains éléments, comme la **condition d'arrêt de la simulation**. À cette fin, une simulation avec les deux complexe-actions les plus basiques, [*ajout aléatoire de lien*] et [*retrait aléatoire d'un lien*], sans filtre particulier pour appliquer ces actions, va être lancée. La méthode de propagation est la plus simple : lorsqu'une entité A applique à une autre entité B un complexe-action CA , CA est transmis à l'entité B . Le nombre de nœuds dans la simulation est de 100, afin d'assurer un temps de simulation court. Pour tous les tests, les conclusions sont vérifiées sur des réseaux avec un nombre de nœuds plus élevé.

Condition initiale

Les actions présentes au début de la simulation sont détaillées dans le tableau 3.2.

La simulation commence par un réseau vide. Chacune des deux actions disponibles est placée sur deux entités distinctes. Le choix de l'entité agissante lors de la simulation est équiprobable. Chaque entité ne peut avoir qu'une paire d'actions : ajout et/ou retrait. Le choix entre ces actions est équiprobable.

Nom	Diminutif	Action	Attribut	Filtres	Prob. de transmission
Ajout aléatoire	addø	Ajout de lien	-	Non connecté	1
Retrait aléatoire	rmvø	Retrait de lien	-	Connecté	1

Table 3.2: Actions disponibles dans la simulation pour le test à vide. Le seul filtre présent est celui strictement nécessaire pour l'application des actions : la vérification préalable de l'existence d'un lien. La transmission de l'action est automatique (Probabilité égale à 1). Les diminutifs sont les noms de ces actions utilisés dans le logiciel.

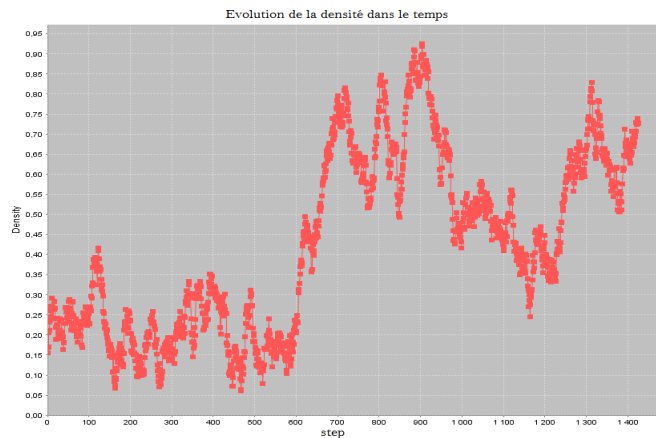


Figure 3.12: Évolution de la densité du réseau avec les actions d'ajout et de retrait aléatoires. $N = 100$ nœuds. L'axe des abscisses correspond au pas de relevé de la densité et non au nombre d'actions réalisées par les entités.

Progression de la simulation

Test initial Après avoir lancé la simulation, on observe l'évolution des différentes propriétés. On s'intéresse en premier lieu à la progression de la densité du réseau, que l'on souhaite utiliser comme marqueur de stabilité et donc de fin de simulation (figure 3.12).

La densité varie énormément, prenant presque l'ensemble du spectre des valeurs possibles. Cette grande variation dépend de la succession des actions d'ajout et de retrait de liens. Sur une durée de simulation arbitrairement longue, la probabilité d'apparition d'évènements tels que :

"Avoir 70 actions d'ajout et 30 de retraites sur les 100 jouées"

augmente. **L'asymétrie d'application des actions sur une certaine fenêtre**

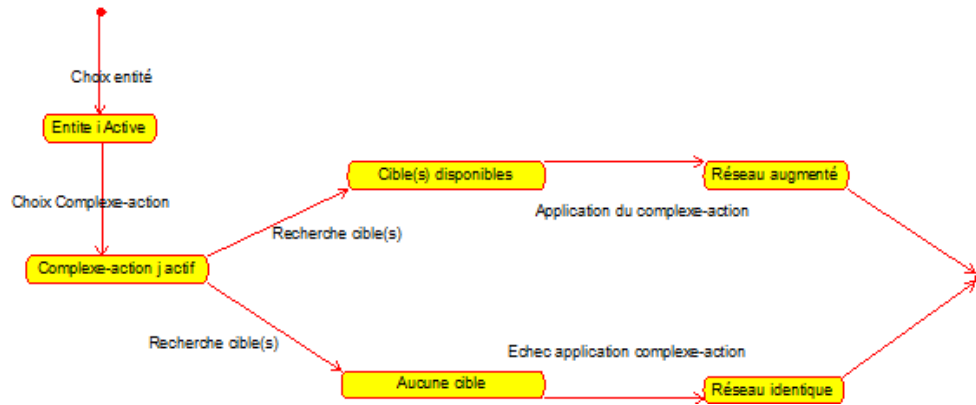


Figure 3.13: Diagramme d'état représentant un tour pour l'application d'un complexe-action d'ajout de lien.

temporelle amène une variation de la densité. Cette variation est d'autant plus notable lorsque le nombre de nœuds dans la simulation est bas, le nombre de liens étant rapporté au nombre de nœuds au carré. Soit L le nombre de liens, N le nombre de nœuds total, le réseau étant non dirigé, la densité

$$D = \frac{2 * L}{N(N - 1)}$$

Pour deux situations distinctes dans le même réseau, L_1 et L_2 le nombre de nœuds correspondant, la variation de la densité est

$$\Delta D = \frac{2 * (L_1 - L_2)}{N(N - 1)}$$

Dans le cas où 50 liens ont été ajoutés entre L_1 et L_2 , sur un réseau possédant 100 nœuds, la variation représente une augmentation de la densité : $\frac{100}{9900} = 1\%$ Si le réseau possède 1000 nœuds, cette même variation correspond à $\frac{100}{999000} = 0.001\%$ d'augmentation.

Il n'est donc pas étonnant d'observer de grandes variations de densité pour des réseaux de petite taille, y compris dans la situation simple où l'ajout et le retrait de liens sont joués aléatoirement.

Nous allons nous intéresser au départ des dynamiques d'évolution d'un réseau, par le biais de trois indicateurs :

- (1) Le ratio entre $\frac{\text{Nombre de tentatives d'ajout de lien}}{\text{Nombre de tentatives de retrait de lien}}$
- (2) Le ratio $\frac{\text{Nombre d'échecs d'applications d'ajout}}{\text{nombre d'échecs d'applications de retrait}}$
- (3) Le ratio $\frac{\text{Nombre d'échecs d'applications d'actions}}{\text{Nombre de réussites}}$

Les échecs d'applications d'une action surviennent lorsqu'une entité, après avoir choisi une action à réaliser, n'est pas capable de trouver une cible pour sa réalisation : par exemple l'ajout d'un lien alors que l'entité agissante est déjà connectée à tous les nœuds du réseau.

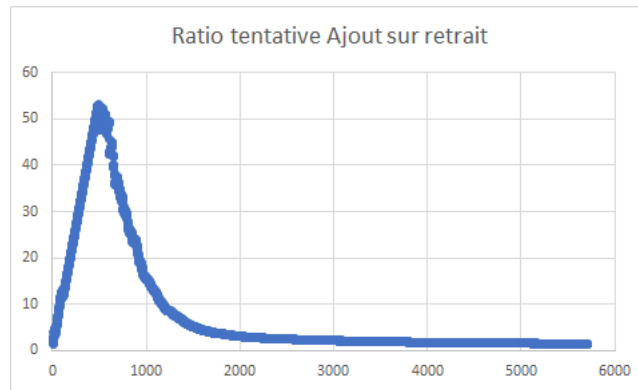


Figure 3.14: *Évolution du ratio entre appels aux actions d'ajouts et les appels aux actions de retraits. La dynamique commence avec bien plus d'appels au complexe-action d'ajout de lien. En effet, les complexe-actions de retrait de lien ne sont au début pas joués puisque le réseau commence sans lien.*

Un exemple rend plus parlant l'utilisation de ces compteurs :

l'entité A est choisie pour appliquer son action d'ajout. Augmentation du compteur *Tentative Ajout* (1). il ne trouve pas de cible valide. *Échec application ajout++* (2), & *échec d'application d'action++* (3).

Dans la figure 3.14, on remarque une évolution singulière du ratio entre le nombre d'appels à l'action d'ajout et le nombre d'appels de l'action de retrait. Cela est dû au fait qu'en partant d'un réseau vide, les seules actions pouvant être appliquées sont celles d'ajout. En effet, la possibilité d'appliquer une action de retrait dépend de l'existence d'un lien. De ce fait, **les premiers complexe-actions à se propager sont ceux portant l'action d'ajout**. De plus, la probabilité de sélectionner un complexe-action d'ajout ne fait qu'augmenter jusqu'à un plus grand nombre d'entités en sont dotées.

Les autres indicateurs n'évoluent que très peu :

- Aucune des actions d'ajout n'a échoué à s'appliquer.
- Une petite dizaine d'actions de retrait ont échoué, au début de la simulation. Ce cas survient lorsqu'une entité à qui le dernier lien a été retiré essaye d'appliquer elle aussi l'action de retrait.
- Le ratio entre échecs et réussites tend rapidement vers 0.

Excepté en début de simulation, il n'arrive que rarement qu'une action tentée échoue. En effet, les entités ont rarement un lien vers toutes les autres, condition pour laquelle l'ajout n'est plus possible. Respectivement, pour le retrait et la connectivité nulle.

Vérification sur un réseau avec N=500 Le réseau passe désormais à 500 nœuds. **Les conclusions sont identiques**, et les observations conformes

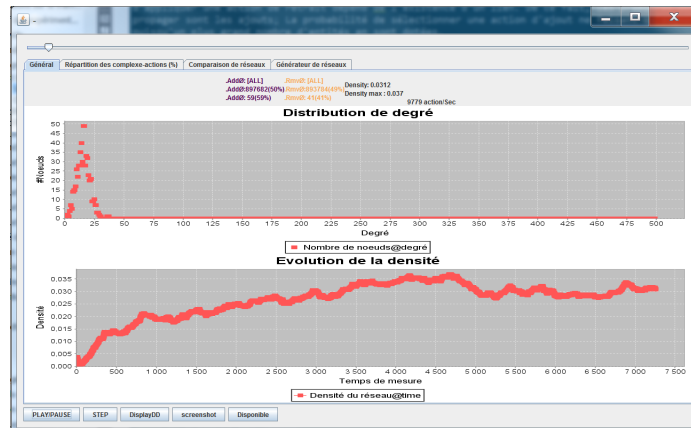


Figure 3.15: Variation de densité avec un plus grand nombre de nœuds. On remarque que les variations semblent plus régulières une fois sorties de la zone où des actions ne peuvent plus être appliquées,

aux dynamiques évoquées plus haut (variation de la densité en figure 3.15). Trois étapes distinctes de dynamique sont observées :

1. Le réseau commence vide, **les premiers complexe-actions à se propager correspondent à l'ajout de lien**. Ces liens existants entre les entités vont permettre au complexe-action de retrait de se transmettre à l'ensemble des nœuds du réseau
2. Passée cette phase de lancement, le ratio d'ajout de lien est supérieur à celui de retrait. En effet, dans certaines situations un nœud ne possédant aucun lien va tout de même être sélectionné pour essayer d'en supprimer un. Le réseau va être poussé vers une situation où tous les nœuds possèdent des liens.
3. Une fois cette situation atteinte, **la variation de densité va se stabiliser**. En moyenne, le nombre d'ajouts équivaut au nombre de retraits.

Observations réalisées sur le test initial Cette configuration, aux actions les plus simples, fait comprendre que la stabilité de la densité du réseau n'est pas un marqueur suffisant d'une fin de simulation. Cela est d'autant plus le cas lorsque le nombre de nœuds en présence est faible. La solution générique la plus simple va être adoptée : **un nombre d'actions maximum**.

Amélioration Il va être nécessaire d'établir une condition d'arrêt en rapport au nombre d'actions réalisées par les entités : la simulation s'arrêtera une fois cette condition atteinte. **Il faut donc trouver un ordre de grandeur cohérent pour cette valeur**.

Dans un premier temps, l'attention va se porter sur le seul élément dont la progression atteint une fin : l'acquisition par les entités de leur premier couple de complexe-actions. **On s'intéresse au nombre d'actions nécessaires pour qu'elles soient entièrement réparties.** À la suite de quoi sera observée l'éventuelle corrélation entre cette valeur et le nombre d'entités dans la simulation. Cela va nous donner un semblant d'ordre de grandeur pour la condition d'arrêt, ainsi que la corrélation de son évolution avec celle du nombre de nœuds du réseau.

On vérifiera pour les résultats des expériences à venir qu'aucun résultat particulier ne survient lorsque la simulation est laissée un nombre d'actions d'ordre 10 fois supérieur.

Progression du temps de propagation des actions en fonction du nombre de nœuds Dans un premier temps, l'attention est portée sur le nombre d'actions nécessaires pour que toutes les actions soient réparties parmi les entités. Sur 100 simulations :

Propriétés	Moyenne du nombre d'action	Écart type
Propagation totale des actions	1370	240

Pour une simulation à 100 nœuds, les ≈ 1500 premiers complexe-actions appliqués permettent leur propagation complète. C'est à la suite de cette étape d'initialisation que des dynamiques spécifiques pourront éventuellement amener la simulation dans un état stable.

Nous nous intéressons maintenant à **l'évolution de cette valeur en fonction de l'augmentation du nombre de nœuds** dans la simulation. Le nombre d'entités va varier de 100 à 1000, par pas de 100. Chaque simulation est lancée 30 fois, la moyenne et l'écart type du nombre d'actions nécessaires pour propagation complète y sont alors relevés (tableau 3.3).

Un des indicateurs pour étudier la corrélation entre deux variables est le coefficient de Pearson (tableau 3.3). Ici, il est de 0.9, ce qui indique une covariance positive quasi maximale. **Le nombre d'actions nécessaires pour répartir tous les complexe-actions augmente donc linéairement par rapport au nombre de nœuds dans la simulation.** On choisira donc de faire évoluer le nombre d'actions maximum avant la fin de simulation linéairement avec le nombre de nœuds dans le graphe. Il est intéressant de constater que ce nombre d'actions nécessaire augmente de façon linéaire en moyenne, mais qu'il subit d'énormes variations d'une simulation à l'autre (Cf. les valeurs d'écart-type dans le tableau 3.3).

Le choix du nombre d'actions avant l'arrêt d'une simulation est délicat :

Une valeur trop basse peut entraîner que les dynamiques n'aient pas encore amenées le système dans un éventuel attracteur. Au contraire, une valeur trop grande diminuerait la rapidité des recherches automatiques des meilleures configurations.

Un multiple du nombre d'actions nécessaires à la transmission de

Nombre de nœuds	nb moyen d'application d'actions pour propagation totale	Écart type	Valeur théorique
100	1802.5	381.2	1353
200	4247.5	667.3	4200
300	6995	873.1	7047
400	9417.5	799.7	9895
500	12465	890.4	12742
600	16057.5	970.7	15589
700	18125	1047.5	18435
800	20905	1152.4	21283
900	24160	1195.7	24130
1000	27482.5	1223.4	26977

Table 3.3: Évolution du nombre moyen d'applications d'actions pour permettre la propagation de tous les complexe-actions dans le réseau. Le nombre de nœuds dans le réseau évolue de 100 à 1000 par pas de 100, et 30 répliques sont réalisées par configuration. Le coefficient de Pearson entre le nombre de nœuds dans le réseau et le nombre d'actions nécessaires pour leur propagation totale est de 0.999, avec $a = 28.47$ et $b = -1493$. a et b obtenus par la méthode des moindres carrés. Les valeurs théoriques sont les valeurs obtenues avec ces paramètres.

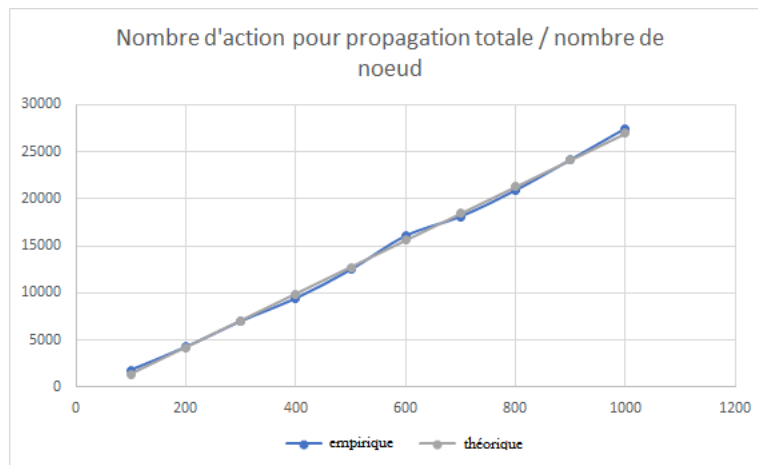


Figure 3.16: Nombre d'itérations nécessaires pour la propagation complète des actions en fonction du nombre de nœuds ; comparaison entre valeurs théoriques (gris) et réelles (bleu). Les valeurs théoriques correspondent aux valeurs calculées si on suppose une corrélation parfaite entre les deux variables. Les coefficients étant donnés dans le tableau 3.3

Propriétés	Moyenne	Écart type
Densité	.213	.09
Nb actions propagation	1923	400

Table 3.4: Valeur de densité et du nombre d'actions nécessaires pour la propagation des comportements, ainsi que l'écart type sur 100 itérations, dans un réseau à 100 nœuds.

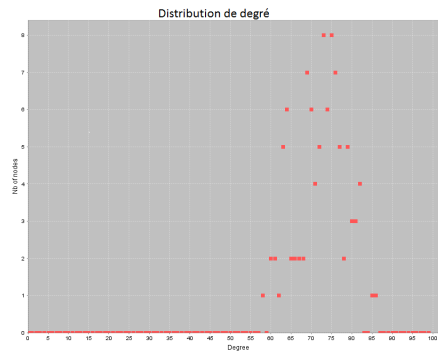


Figure 3.17: Simulation des complexe-actions [ajout aléatoire et retrait aléatoire]. La distribution de degrés est une gaussienne (loi normale) centrée sur la densité (le degré moyen des nœuds sur le nombre maximum possible). Cette densité varie en fonction de la succession d'actions d'ajout ou de retrait prises sur les dernières actions.

toutes les actions est choisi arbitrairement, et nous resterons attentifs au comportement des configurations passé ce seuil. A cette fin, pour vérifier qu'une organisation particulière des entités ne survienne pas passée la limite du nombre d'action, chaque configuration ne sera arrêtée qu'après un nombre d'actions 10 fois supérieur au seuil.

Modification apportée Pour les conditions d'arrêt d'une simulation, un nombre d'actions maximum est ajouté en complément de la stabilité de la densité.

Nouveaux résultats Les conditions initiales sont les mêmes : sur 100 simulations, complexe-actions [*ajout aléatoire*] et [*retrait aléatoire*], condition d'arrêt 250.000 actions (basée sur le facteur de corrélation linéaire entre nombre de nœuds et actions nécessaires pour une propagation totale, le multiplicateur ayant été établi empiriquement).

Propriétés des réseaux obtenus Grâce à la définition stricte d'un critère d'arrêt de simulation, une valeur finale de densité de réseau est observable (Tableau 3.4), en plus de la distribution de degré (figure 3.17) dont la forme en gaussienne déjà observée est l'une des caractéristiques des réseaux aléatoires.

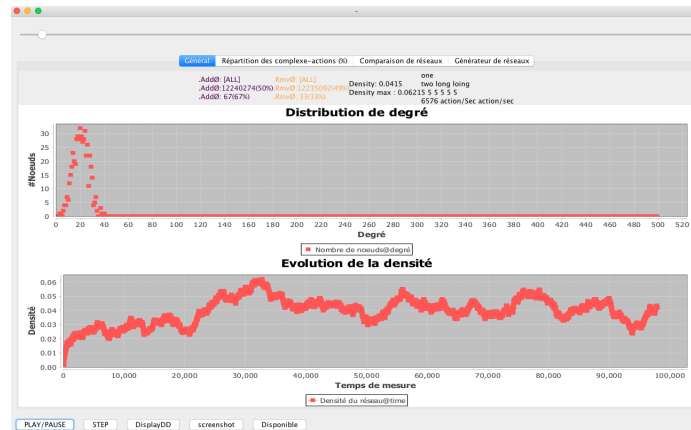


Figure 3.18: Simulation jouée sur 10 fois plus d'action pour la configuration ajout et retrait aléatoire, en partant d'un réseau vide, avec 500 nœuds. Aucune nouvelle dynamique n'apparaît.

Évolution de la propagation des actions Après propagation totale des actions, toutes les entités possèdent le couple d'actions [*ajout de lien aléatoire*] et [*retrait de lien aléatoire*].

Comportement aux limites Afin de s'assurer que la condition de fin de simulation, qui dépend d'un nombre d'actions maximum, n'empêche pas l'apparition de réseaux particuliers, chaque simulation est laissée un temps d'ordre 10 fois supérieur. Le nombre de nœuds est de 500 afin de minimiser les variations de propriétés principalement dues au faible nombre de nœuds. Les résultats sont représentés dans la figure 3.18 : **aucun comportement inattendu ne survient.**

Conclusion établie à partir des tests à vide

Malgré la configuration simple, aucune stabilité topologique du réseau n'est atteinte. Il est nécessaire d'ajouter une condition d'arrêt indépendante des propriétés du réseau en construction. Le critère est placé sur le nombre d'actions réalisées par les entités lors d'une simulation.

Nom	Code	Action	Attribut	Filtres	Probabilité de transmission
Ajout degré sup.	add+	Ajout lien	degré	Non connecté ; degré supérieur à l'agissant	1
Retrait aléatoire	rmvø	Retrait de lien	-	Connecté	1

Table 3.5: Première tentative : configuration initiale pour l'obtention d'un réseau sans échelle. Le réseau est pourvu de deux complexe-actions aux effets dissymétriques : l'ajout de lien doit mener à quelques nœuds plus fortement connectés que la plupart des autres.

3.4.3 Reproduction d'un réseau sans échelle

L'objectif va être de reproduire un réseau sans échelle (Scale-Free) à 500 nœuds, généré par un modèle d'attachement préférentiel (Preferential Attachment, PA ([Barabasi and Albert-László, 1999](#))). Les réseaux de type sans échelle ont une distribution particulière : quelques nœuds ont un degré élevé, et la grande majorité en ont un très faible.

L'idée va être de s'inspirer du modèle d'Attachement Préférentiel, utilisé pour générer des réseaux sans échelle. La différence est que ce modèle utilise un mécanisme de croissance ; une fois qu'un nœud est ajouté au réseau, il n'agit plus. Dans notre cas, **nous essayons de trouver une dynamique convergeant vers une même distribution en loi de puissance et une même densité de réseau.**

Sans échelle - V1 : nouveaux filtres

Afin d'atteindre ce type de réseau, nous allons retranscrire l'attractivité pour les entités de degré supérieur via la création d'un filtre combinable dans les complexe-actions.

Les **filtres** $>$ et $<$ sont des conditions comparant un attribut du nœud agissant, à ceux des nœuds candidats pour l'application de l'action. Le $>$ renvoie les nœuds possédant une valeur d'attribut supérieure à celle du nœud agissant, et inversement pour le $<$.

L'attribut associé à ces comparaisons est le *degré*.

Le couple de complexe-actions choisi dans le tableau 3.5 vise donc à transcrire le mécanisme du modèle d'attachement préférentiel en dynamique stable. L'ajout de lien se fait vers un nœud de degré supérieur auquel l'entité agissante n'est pas encore connectée. Un complexe-action de retrait y est ajouté, afin de conserver une dynamique de construction toujours active, et de ne pas tendre vers un réseau complet. Aucun filtre (condition) n'est associé à cette action de retrait ([*Retrait de lien aléatoire*]).

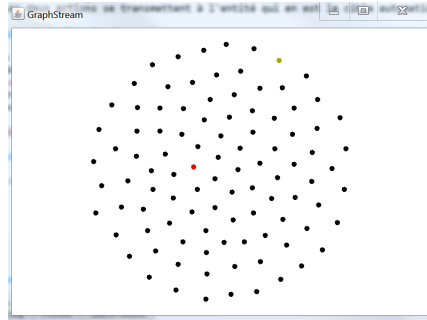


Figure 3.19: Résultat de simulation depuis un réseau vide, pour les complexe-actions [Ajout degré $>$; retrait aléatoire]. Capture d'écran réalisée sur un réseau à 100 nœuds par souci de visualisation. La couleur des nœuds représente les complexe-actions portés. La couleur rouge correspond au complexe-action [retrait aléatoire], le marron à [Ajout degré \geq] et le noir à aucun complexe-action.

Résultats A priori, commencer la simulation depuis un réseau vide semble adapté. Seuls les complexe-actions détermineraient alors la forme du réseau obtenu. Dans ce cas précis, le résultat est donné par la figure 3.19 : **le réseau est vide.**

La première action pouvant être entreprise est l'ajout de nœud vers un autre nœud de degré supérieur. Mais aucun nœud ne correspond à cette description. Pour résoudre cette situation, il est possible de :

- baisser la contrainte d'ajout pour qu'elle puisse se faire à un nœud de degré supérieur ou égal ou
- choisir de commencer d'un réseau aléatoire, et non pas vide.

Sans échelle - V2 : relaxation de la contrainte

La première solution va être éprouvée (figure 3.20). La contrainte sur les conditions de l'ajout baissée, le premier nœud actif peut appliquer son action, crée un lien avec sa cible, et lui transmet l'action. Une fois cela effectué, les deux se retrouvent être les nœuds de degré maximum. Plus aucune action n'est possible :

- les deux seuls nœuds possédant une action d'ajout sont déjà connectés et
- le nœud possédant l'action de retrait n'est pas connecté, il ne peut donc pas jouer son action.

Il est à noter qu'une série d'évènements bien particuliers permet le déploiement des actions sur le réseau. Soient $Agent_1$ l'entité possédant l'action d'ajout, $Agent_2$ celle possédant l'action de retrait.

1. Application du complexe-action [ajout de lien] depuis $Agent_1$ vers $Agent_2$.
La sélection de l' $Agent_1$ est automatique car il est le seul agent actif sur le réseau. Il a à sa disposition les $N - 1$ autres agents.

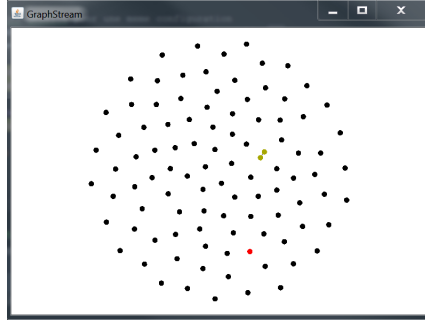


Figure 3.20: Version 2- Résultats de simulation depuis un réseau vide, pour les complexe-actions [Ajout degré \geq ; retrait aléatoire] (contrainte sur le filtre $>$ relaxée). Capture d'écran réalisée sur un réseau à 100 nœuds par souci de visualisation. La couleur des nœuds représente les complexe-actions portés. La couleur rouge correspond au complexe-action [retrait aléatoire], le marron à [Ajout degré \geq] et le noir à aucun complexe-action.

$$\mathbb{P}(A) = \frac{1}{N} \quad (3.1)$$

2. Lorsque l'action de retrait sera appliquée, elle se transmettra depuis $Agent_2$ vers $Agent_1$.
3. $Agent_1$ possède alors les actions d'ajout et de retrait, l'ensemble se diffusera à tout le réseau.

Cette solution n'est pas retenue car les résultats ne se reproduisent pas à paramètres égaux.

Nous allons donc appliquer la deuxième possibilité. **La simulation commence à partir d'un réseau aléatoire, où chaque nœud est connecté en moyenne à la moitié du réseau.**

Sans échelle - V3 : modifications du réseau de départ

Le réseau de départ est un réseau aléatoire de densité 50%, $N=500$. Les nouveaux résultats sont donnés par la figure 3.21 :

Grâce aux liens présents dès le début de la simulation, les actions de retrait de lien peuvent se propager à tout le réseau. Ce dernier va donc brusquement baisser en densité. **Le complexe-action d'ajout de lien est propagé plus lentement** : le filtre associé rend l'application de l'action plus difficile, certains nœuds ne pouvant pas trouver de cible adéquate.

Après répétition de la configuration sur 100 simulations de 100 nœuds, la répartition des complexe-actions parmi les nœuds est donnée dans le tableau 3.6¹. Pour comprendre pourquoi les complexe-actions d'ajout ne se propagent pas sur toutes les entités du réseau, nous allons nous attarder sur le fichier d'échecs

1. Les résultats sont aussi validés sur des réseaux à 500 nœuds

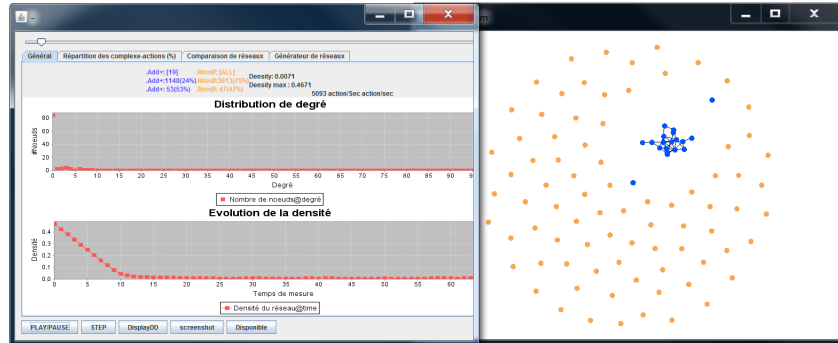


Figure 3.21: Version 3- Résultats de simulation depuis un réseau aléatoire avec une densité de 50%. La contrainte sur le filtre associé à l'action d'ajout est relaxée. Les nœuds de couleur sont ceux possédant l'action de retrait seule, les bleus possèdent à la fois les actions d'ajout et de retrait. Capture d'écran réalisée sur un réseau à 100 nœuds par souci de lisibilité.

Complexe-action	Moyenne	Écart type
Add+	16.9	7.6
Rmvø	100	0

Table 3.6: Version 3- Répartition finale des complexe-actions parmi les entités, pour 100 nœuds en partant d'un réseau aléatoire de densité 50%. Couple de complexe-action [ajout >] & [retrait aléatoire] Configuration lancée 100 fois.

d'application des actions.

Dans les 30 premières actions appelées, l'action d'ajout a obtenu 11 échecs sur 16 appels. A contrario, le complexe-action portant l'action de retrait de lien est toujours applicable, le réseau de départ étant un réseau aléatoire portant suffisamment de liens. Rapidement, de plus en plus de nœuds possèdent ce complexe-action. Les nœuds étant choisis de façon uniforme pour jouer une action, le complexe-action de retrait est plus souvent appelé. De nombreux nœuds se retrouvent non connectés au réseau avant d'avoir pu apprendre le complexe-action d'ajout. Au vu de la contrainte d'ajout portée, ils n'auront pas l'occasion de se voir appliquer l'action d'ajout et donc de l'apprendre pour le reste de la simulation.

Sans échelle - V4 : combinaison des modifications précédentes

La solution précédente souffre d'un défaut majeur : une fois qu'une entité se retrouve déconnectée du reste du réseau sans action d'ajout avec elle, il lui est impossible de le réintégrer. Combiner les deux solutions devrait permettre d'obtenir l'effet recherché.

Après simulation (figure 3.22), l'ensemble des complexe-actions se propagent bien aux entités du réseau.

Sans échelle - V4 alternative : complexe-actions de "fluidité"

Malgré le fait que la solution précédente atteigne l'objectif, **le moyen d'y parvenir ne semble pas satisfaisant**. On peut par exemple penser à un cas où un complexe-action d'ajout de lien possède une probabilité de propagation très faible ; il n'aurait pas le temps de se propager à l'ensemble du réseau avant que tous les liens ne soient supprimés par une action de retrait. Nous allons essayer une solution alternative.

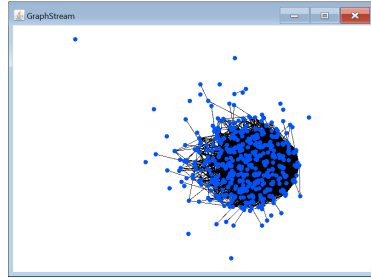
L'objectif est de permettre la propagation des complexe-actions aux entités de la simulation en évitant la possibilité de blocage. L'application, et donc la propagation des complexe-actions, dépendent de l'existence de liens entre entités ; on va établir une dynamique topologique qui disparaîtra au fur et à mesure de la propagation des complexe-actions aux entités du réseau.

Chaque nœud va initialement être doté des complexe-actions [ajout aléatoire] et [retrait aléatoire]. Ces derniers ne peuvent pas être transmis, et seront donc remplacés au fil de la simulation. Cela assure que les liens du réseau soient toujours existants pendant la phase de propagation des actions initiales.

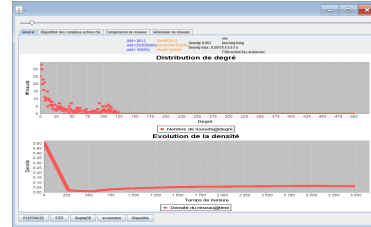
Comparaison des méthodes "fluidité" et "relaxé + réseau aléatoire"

On s'intéresse à comparer les résultats donnés par les trois méthodes :

1. relaxation de contrainte de comparaison sur le filtre et réseau de départ à 50%,
2. complexe-actions de fluidité et



(a) Réseau obtenu.

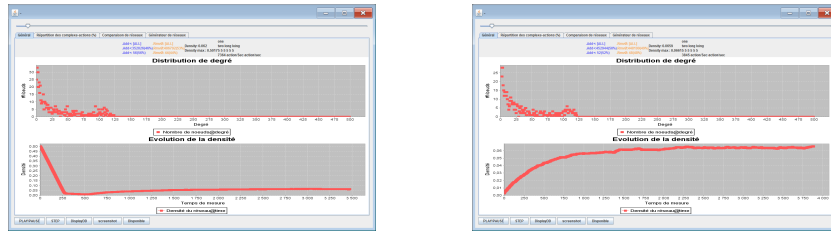


(b) Évolution de la densité et distribution de degré.

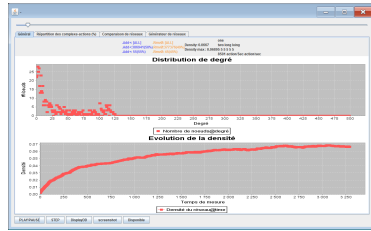
Figure 3.22: Version 4- Résultat du modèle partant d'un réseau aléatoire de densité 50% et utilisant la relaxation sur la contrainte du filtre (\geq). Couple de complexe-actions [ajout \geq] & [retrait aléatoire]. $N=500$ nœuds. La couleur bleu des nœuds sur le réseau correspond à la possession des deux complexe-actions.

Nom	Code	Action	Attribut	Filtres	Probabilité transmission
Ajout degré sup.	add+	Ajout lien	degré	Non connecté & degré supérieur à l'agissant	1
Retrait aléatoire	rmvø	Retrait de lien	-	Connecté	1
Ajout aléatoire	addø	Ajout de lien	-	Non connecté	0
Retrait aléatoire	rmvø	Retrait de lien	-	Connecté	0

Table 3.7: Configuration initiale pour l'obtention d'un réseau sans échelle. Le réseau est pourvu de deux complexe-actions de "fluidité", ajout et retrait, qui disparaîtront lors de la propagation des autres actions.



(a) Cas solution mixte : contrainte relaxée (\geq) (b) Cas complexe-actions fluidité et contrainte stricte ($>$)



(c) Cas complexe-actions fluidité et contrainte relaxée (\geq)

Réseau	Nb lien	Gamma
Mixed	8079	0.78
Fluidité stricte	8170	0.94
Fluidité Relaxee	8024	0.84

(d) Comparaison du nombre de nœuds et coefficient gamma pour les réseaux générés.

Figure 3.23: Résultats de génération de réseaux pour l’obtention d’un réseau Scale-Free, avec comme complexe-actions de base [Ajout degré $>$; retrait aléatoire]. Comparaison de trois méthodes évitant les situations initiales de blocage.

3. complexe-actions de fluidité et relaxation de contrainte sur le filtre.

La figure 3.23 compare les résultats obtenus sur les différentes approches.

Les résultats issus des trois solutions semblent équivalents. On va préférer la solution avec les actions de fluidités. Fondamentalement, les méthodes proposent la même chose : fournir des liens sur le réseau afin de permettre la propagation des complexe-actions. La différence est que la méthode de fluidité continue la création et suppression de liens dynamiquement. Certaines configurations initiales ont besoin de ces liens pour progresser. Quant à la différence entre une comparaison stricte ou relaxée, on restera attentif aux différences de résultats apportées par ce choix.

Étude approfondie des résultats

On va étudier plus en détail les résultats obtenus avec la configuration suivante : 500 nœuds, complexe-actions de fluidité et contrainte relaxée pour le filtre associé à l’action d’ajout de liens (\geq). On commence par un réseau vide, avec les deux actions d’ajout et de retrait habituelles (figure 3.7), et la méthode de fluidité.

L’ensemble initial des complexe-actions s’est réparti sur le graphe (figure 3.24). **La distribution de degrés du réseau obtenu a une ”bonne forme” pour être sans échelle :** beaucoup de nœuds ont peu de connexions et quelques-uns en ont beaucoup. Néanmoins, pour déterminer si un réseau est vraiment de type

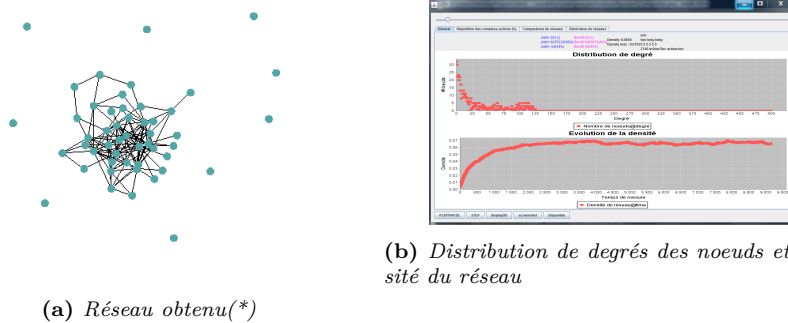


Figure 3.24: Résultats de simulation depuis un réseau vide, avec deux complexactions permettant l'ajout et un permettant le retrait. La distribution de degrés représente une distribution de Pareto mais n'est pas assez appuyée pour être considérée comme Scale-Free. (*) Capture d'écran du réseau avec 50 nœuds pour sa visualisation.

sans échelle, il faut s'intéresser au coefficient γ , déterminant la distribution de degrés des nœuds par la relation

$$P(k) \propto k^{(-\gamma)}$$

donnant la répartition des nœuds de degré k . Ce coefficient doit être compris entre 2 et 3 pour correspondre à la définition stricte d'un réseau de type sans échelle.

Afin de fournir un réseau cible respectant cette propriété, nous allons générer un réseau avec le modèle d'attachement préférentiel. Le paramètre m définissant le nombre de connexions par nœud entrant du modèle d'attachement préférentiel n'influence pas l'exposant γ du réseau généré. Néanmoins, cela influence le degré minimum que l'on peut retrouver parmi les nœuds du réseau. On va donc se baser sur un réseau d'attachement préférentiel généré avec un paramètre $m=3$. En effet, il nous est plus difficile de reproduire une grande quantité de nœuds possédant un degré de 1.

Certaines valeurs de propriétés des réseaux sans échelle sont données par (Albert and Barabási, 2002), nous gardons ici des réseaux de taille comparable, et calculons ce facteur pour un réseau à 1000 nœuds, généré par l'algorithme d'attachement préférentiel (tableau 3.8).

On remarque que le coefficient gamma n'est pas dans la bonne plage de valeurs, et que **notre réseau est trop dense**. Le tableau 3.9 nous renseigne sur les détails des écarts du nombre de nœuds nécessaires par degré.

Modifications

Malgré l'aspect sans échelle du réseau obtenu, la densité y est trop forte. Il existe trop de liens entre les entités de degré faible.

Nous allons essayer de trouver d'autres actions pour rendre la décroissance en

Numero	Réseau	Taille	Degré moyen : densité	Gamma
1	Routeur internet	3888	2.57 : 8.9 e-4	2.48
2	Protein S.Cerev	1870	2.39 : 1.2 e-3	2.4
3	SF par PA	1000	NP : 2 e-3	2.44
4	Réseau synthétique	1000	NP : 5 e-2	0.94

Table 3.8: Valeur du facteur γ pour différents réseaux. 1- réseau de routeur internet, 2- réseau de protéines 3 - Réseau sans échelle généré par un modèle d'attachement préférentiel avec comme paramètre $m=3$, 4 - réseau généré par notre méthode avec les complexe-actions [**A**jout degré >], [**R**etrair aléatoire], [**A**jout aléatoire] et [**R**etrair aléatoire] comme éléments de fluidité.

Degré	Nb nœuds PA $m=1$	$m=2$	$m=3$	Nb nœuds synthétique
0	0	0	0	0
1	662	0	0	38
2	171	510	0	53
3	66	193	401	28
4	39	109	203	39
5	18	65	111	30
6	6	27	69	29
7	10	14	50	31
8	9	12	34	21

Table 3.9: Détails sur le nombre de nœuds des premières valeurs de degrés pour les réseaux sans échelle générés par attachement préférentiel, en fonction de leur paramètre m . Comparaison avec le nombre de nœuds par degré obtenu avec notre méthode de génération. Les densités des réseaux sont PA $m=1$: 4.0 e-3, $m=2$: 1.0 e-2, $m=3$: 1.2 e-2, notre méthode : 6.5 e-2.

Nom	Code	Action	Filtres	Attribut	Probabilité de transmission
Ajout aléatoire	add \emptyset	Ajout lien	-	-	1
Retrait sup.	rmv+	Retrait de liens	Mon attribut \geq à la cible	Degré	Variable
Retrait inf.	rmv-	Retrait de liens	Mon attribut \leq à la cible	Degré	Variable

Table 3.10: Configuration initiale pour l'étude de l'influence des probabilités de propagation. Une action d'ajout aléatoire avec une probabilité de propagation à 1, et deux complexe-actions de retrait, leurs filtres portant sur la différence de degrés (\geq et \leq) entre l'entité agissante et ses cibles.

répartition de degrés des nœuds plus abrupte. Il va donc nous falloir introduire de nouveaux complexe-actions portant des actions similaires à celles déjà présentes (ajout ou retrait). Nous allons, en premier lieu, étudier l'influence du paramètre de probabilité de propagation sur la répartition finale.

3.4.4 Etude des dynamiques en fonction des paramètres de propagation

On se place dans le cas où plusieurs complexe-actions portent la même action (ajout ou retrait de lien). On cherche alors à déterminer l'influence de leurs probabilités de propagation sur la distribution finale.

Condition initiale La configuration de simulation qui va servir d'objet d'étude est donnée dans le tableau 3.10 : Deux opérations de retrait de liens pour un seul d'ajout. Le réseau de départ est aléatoire, avec une densité de 50%. Aucune situation de blocage ne peut survenir : l'ajout de liens sans condition assure que le réseau ne tende pas vers un réseau vide, les actions de retrait peuvent alors toujours être jouées. La comparaison sur le degré de la cible n'est pas stricte. Les combinaisons de simulations suivantes sont explorées :

Retrait de lien depuis un nœud de degré inférieur, probabilité de transmission variant de 0.2 à 1 avec des pas de 0.2
X
Retrait de lien depuis un nœud de degré supérieur, probabilité de transmission variant de 0.2 à 1 avec des pas de 0.2

	rmv+	0.2	0.4	0.6	0.8	1
Rmv-	0.2	31.5[22.7]	25.2[19]	31.0[17.6]	25.6[15.8]	21.3[17.1]
	0.4	74.7[11.1]	59.6[7.7]	49.2[6.1]	41.4[5.2]	38.2[13.8]
	0.6	84.1[9]	66.2[7]	60.4[8.7]	53.4[8]	50[13.2]
	0.8	87.7[8.6]	73.4[6.3]	62.3[5.6]	59[7]	58.1[11.8]
	1	89[6.8]	78.3[6.1]	70.1[5.8]	62.5[7.2]	61.7[5.3]

Figure 3.25: Distribution des complexe-actions *Rmv-* après 300.000 actions, dans une simulation contenant les complexe-actions *rmv- ;rmv+ ;add random*. Les moyennes et écarts-types sont relevés sur 20 itérations par configuration. Les probabilités de transmission varient de 0.2 à 1, avec des intervalles de 0.2.

Ces 25 jeux de paramètres sont répétés 20 fois dans un réseau à 100 nœuds, afin d’avoir une idée significative sur les écarts-types des résultats.

Résultats La densité moyenne de l’ensemble des réseaux de la simulation est de 0.98, avec un écart type moyen de 0.008 (l’écart-type sur la moyenne des écarts type : 0.005). Le réseau obtenu tend donc vers un réseau complet pour l’ensemble des configurations.

En ce qui concerne la répartition des complexe-actions après 300.000 actions jouées, on constate que malgré des tendances, les écarts-types, surtout en situation de propagation faible, sont grands (tableau 3.25). La figure 3.26 présente la répartition du complexe-action *rmv+* en fonction des probabilités de transmission de ce dernier. La figure 3.27 représente les résultats pour les mêmes paramètres excepté la taille du réseau à 500 nœuds. Les résultats sont similaires, avec des variations inférieures sur les écarts types de répartition.

Conclusion La probabilité de transmission d’un complexe-action influence bien sur la distribution parmi les entités, mais les écarts-types ont des valeurs assez importantes. Le faible nombre de nœuds dans la simulation est l’une des causes de cela, l’augmentation du nombre de nœuds amoindrit cet effet.

3.4.5 Questions sur la symétrie des actions

Outre l’influence de la probabilité de transmission sur la répartition finale des complexe-actions, on peut aussi se poser la question de l’équivalence des combinaisons d’actions. Le complexe-action [*Ajout degré >*] donne-t-il les mêmes résultats qu’un [*Retrait degré <*] ?

Nous allons comparer deux configurations simples :

1. la combinaison de [*ajout degré >*] avec [*retrait aléatoire*] et
2. [*retrait degré <*] avec [*ajout aléatoire*]

dans un réseau à 500 nœuds, en gardant la comparaison stricte (une comparaison relaxée aurait des cas d’ajout de lien en commun ce qui pourrait fausser la

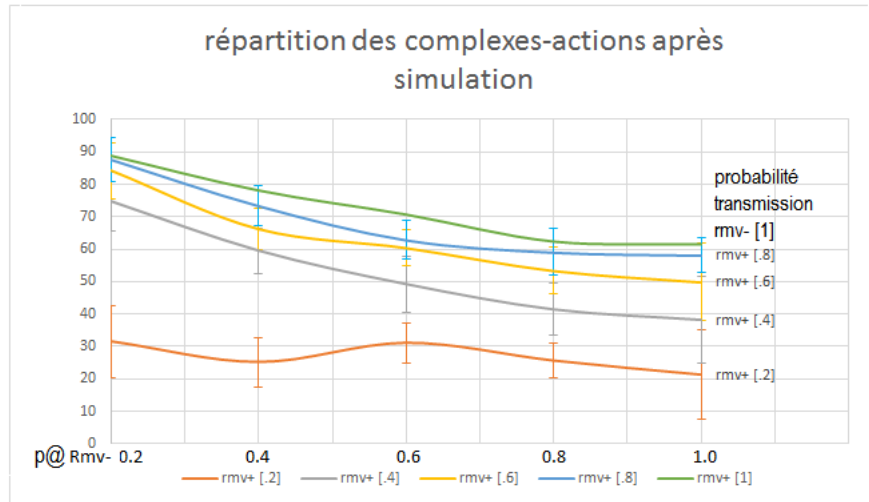


Figure 3.26: Variation de la répartition des complexe-actions $rmv+$ après simulation, en fonction de leur probabilité de transmission, avec $N=100$.

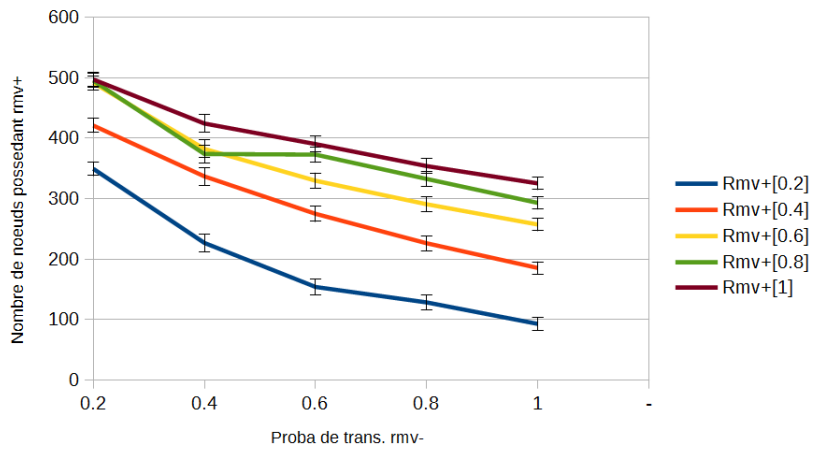


Figure 3.27: Reproduction de la variation de la répartition des complexe-actions $rmv-$ et $rmv+$ après simulation, en fonction de leur probabilité de transmission, cette fois avec 500 nœuds.

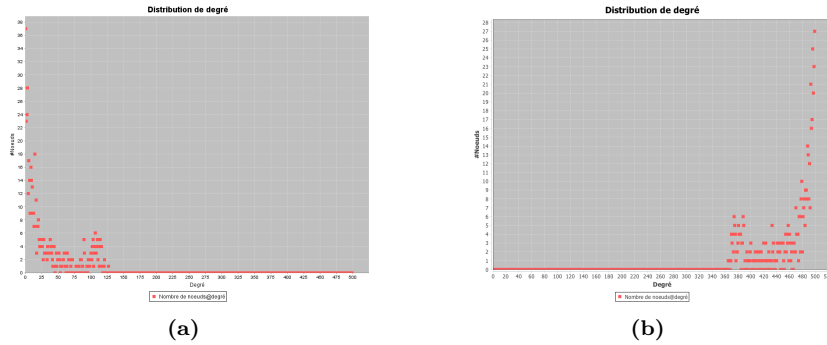


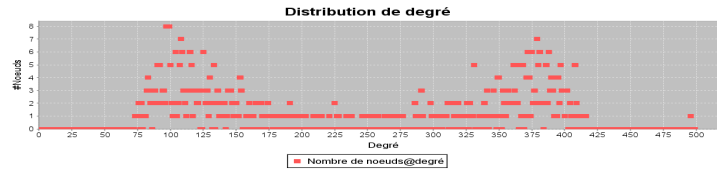
Figure 3.28: Comparaison de la distribution de degrés pour deux configurations.
 (a) - Distribution de degrés obtenue avec la configuration [ajout degré >], [retrait aléatoire]. La densité du réseau est de 0.065.
 (b) - Distribution de degrés obtenue avec la configuration [retrait degré <], [ajout aléatoire]. La densité du réseau est de 0.935.
 Les distributions de degrés et densité sont symétriques par rapport à leur valeurs moyennes.

comparaison des deux configurations). Les résultats concernant la distribution de degré et la densité semblent **symétriques, plutôt qu'équivalents** (figure 3.28). Cela s'explique par le fait que les complexe-actions [*Ajout aléatoire*] et [*retrait aléatoire*] peuvent toujours être appliqués, tandis que leurs contreparties conditionnelles essuient des échecs d'applications. Par conséquent, le réseau tend vers une forme vide ou complète, respectivement. Les densités de degré du réseau tournent autour de 0.065 et 0.935.

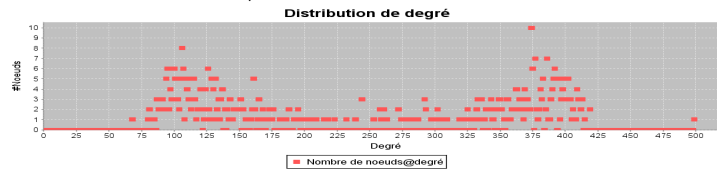
Une autre question peut porter sur la comparaison des couples d'actions

- [*Ajout degré >*] et [*Retrait degré <*],
- avec [*Ajout degré <*] et [*Retrait degré >*].

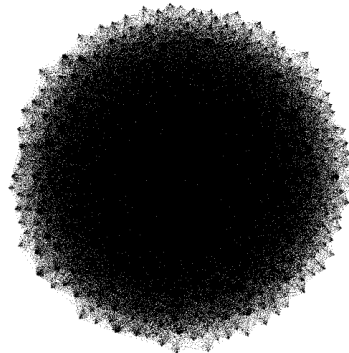
Les résultats sont beaucoup plus étonnants. Malgré une densité de 0.5% commune aux deux configurations, les manières d'y parvenir sont différentes. Nous implémenterons donc pour chaque filtre des complexe-actions l'initial et son opposé sémantique, par exemple ajouter un lien vers un degré supérieur et supprimer un lien depuis un degré inférieur.



(a) Distribution de degrés obtenue avec la configuration [ajout degré >], [retrait degré <]. La densité varie autour de 0.48.



(b) Distribution de degrés obtenue avec la configuration [ajout degré \geq], [retrait degré \leq]. La densité finale varie autour de 0.5.



(c) Une partie des nœuds plus densément connectés au centre, et les nœuds formant l'autre gaussienne de distribution de degrés dans la couronne.

Figure 3.29: Résultats obtenus avec les configurations 1- [ajout degré >] & [retrait degré <] 2-[ajout degré \geq] & [retrait degré \leq]. Les résultats sont très similaires, à ceci près que la configuration aux contraintes relaxées atteint une densité moyenne (0.5) tandis que la configuration stricte y tend sans dépasser 0.48. Dans les deux cas, les nœuds se répartissent en deux groupes, un noyau fortement connecté et un autre plus légèrement.



(a) Distribution de degrés obtenue avec la configuration [ajout degré >], [retrait degré <]. La densité finale varie autour de 0.5.
 (b) Distribution de degrés obtenue avec la configuration [ajout degré <], [retrait degré >]. Ce résultat est très lent à être atteint.

Figure 3.30: Distribution de degré obtenue pour les deux configurations symétriques, avec des contraintes strictes en ce qui concerne les filtres de comparaison entre degrés.

3.4.6 Reproduction d'un réseau Small-World

Nous allons maintenant confronter le Framework à des réseaux Small-World, générés par le modèle de (Watts and Strogatz, 1998). Ils ont deux caractéristiques principales :

- les entités possèdent un coefficient de clustering bien supérieur à celui d'un réseau aléatoire de même taille et
- la taille moyenne des plus courts chemins (*APL-Average Path Length*) est faible, grâce à des liens "faibles" entre les groupes d'entités, et évolue en $\log(N)$.

Ces définitions restent cependant qualitatives. Nous allons nous intéresser à ces propriétés pour des réseaux théoriques et réels, afin d'obtenir un point de comparaison pour les réseaux générés.

En ce qui concerne les réseaux Small-World réels, (Leskovec and Krevl, 2014) fournissent les données pour, entre autres, Facebook, Google+ et Twitter.

Le modèle de (Watts and Strogatz, 1998) génère quant à lui un réseau théorique de type Small-world en rebranchant les liens entre les nœuds d'un réseau de grille régulière. Deux paramètres sont nécessaires pour ce modèle : la densité du réseau initial et la probabilité de rebranchement d'un lien. Nous choisissons une densité de 0.028, cohérente pour ce type de réseaux, en se basant sur celles relevées sur Facebook (.01, .04), Google+ (0.007) et Twitter (.04). En ce qui concerne la probabilité p de rebranchement, elle doit être comprise entre $0.001 < P < 0.1$ (Barabási and Pósfai, 2016). Une illustration des réseaux obtenus en fonction de la valeur de ce paramètre se retrouve dans la figure 3.31. Les valeurs de propriétés pour ces réseaux sont résumées dans les tableaux en 3.11.

P	Clustering	APL
0.001	0.689	9.44
0.01	0.673	4.05
0.1	0.519	3.34

Réseau	Clustering	Diamètre effectif
Facebook	0.6055	4.7
Google+	0.491	3
Twitter	0.5653	4.5

(a) Valeurs des coefficients de clustering et d'APL pour des réseaux Small-World théoriques. Les réseaux générés par le modèle de Watts-Strogatz (*Watts and Strogatz, 1998*) sont rebranchés depuis un réseau aléatoire à densité de 2.8%, avec $N=500$ nœuds. Le nombre k de voisins par nœud de la grille régulière est donc de 7.

(b) Valeur des coefficients de clustering et d'APL pour des réseaux small-world réels, issus de (*Leskovec and Krevl, 2014*).

Table 3.11: Valeurs de clustering et APL pour différents types de réseaux Small-World.

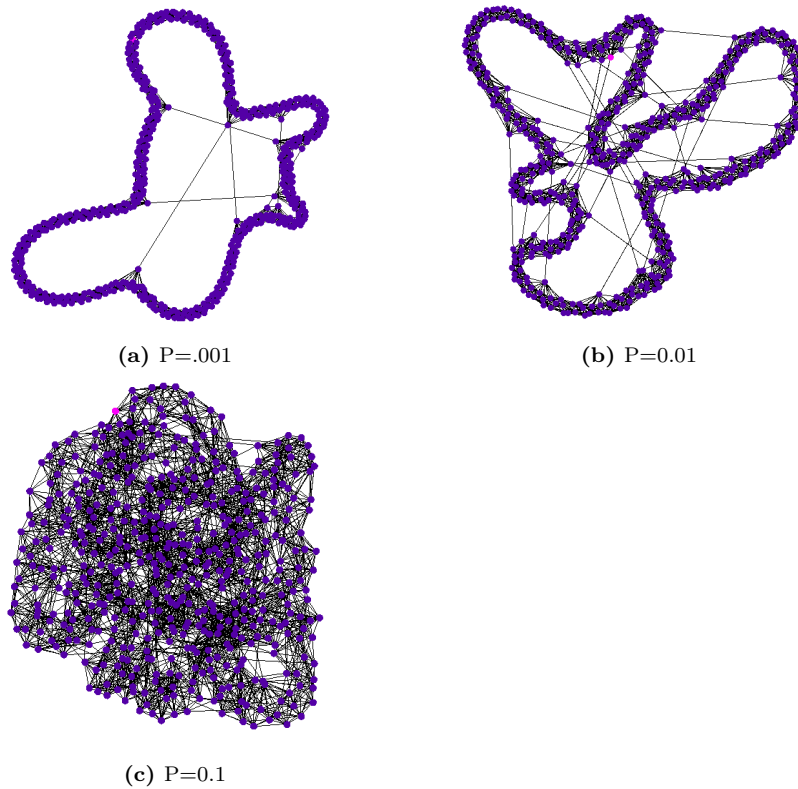


Figure 3.31: 3 réseaux générés par le modèle de (*Watts and Strogatz, 1998*), 500 nœuds, 7000 liens pour une densité de 2.8%. Les probabilités P de re-routage des liens depuis la grille régulière sont de 0.001, 0.01 et 0.1.

Première approche

Lorsque l'on se place du point de vue d'une entité E_1 , le clustering définit la proportion d'entités E_i connectées à E_1 aussi connectées entre elles. **On va donc introduire un nouveau filtre pour complexe-action, le *triangle*.**

Le filtre *triangle* est une condition portant sur la distance de connectivité entre un nœud actif N_a et ceux auxquels il est indirectement connecté. Le filtre *triangle* renvoie tous les nœuds N_{d2} (à distance 2) connectés par l'intermédiaire des nœuds N_{d1} directement connectés à N_a . Aucun attribut n'est donc pris en compte. Lorsque ce filtre est associé à une action d'ajout de lien, cela permet de représenter la **transitivité**.

La figure 3.32 illustre le fonctionnement du filtre. À partir de ce filtre est constitué un complexe-action qui ajoute un lien vers un nœud à distance 2, *add-2hop*. Cela doit permettre d'obtenir un coefficient de clustering supérieur à celui d'un réseau aléatoire. On y associe un complexe-action de retrait de lien aléatoire pour les mêmes raisons que précédemment : obtenir une dynamique qui ne tend pas vers un réseau complet.

La configuration initiale est un réseau vide de 500 nœuds, avec les complexe-actions de fluidité, fermeture de triangle et retrait aléatoire (tableau 3.12).

Résultats Le réseau résultat est sans lien. Les actions de fluidités aident à la répartition totale des complexe-actions sur l'ensemble du réseau. Une fois cet état atteint, le réseau tend vers un réseau vide. En effet, le complexe-action de retrait aléatoire peut s'appliquer sans condition, tandis que celui d'ajout demande que deux nœuds soient déjà connectés indirectement par un troisième nœud. Les actions d'ajouts ne peuvent donc pas être appliquées systématiquement, ce qui entraîne le réseau vers un état sans lien.

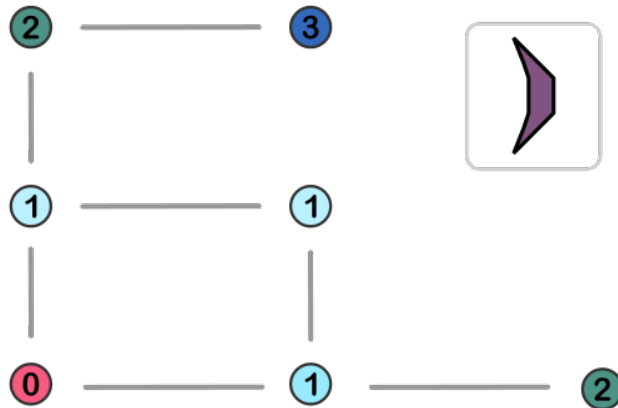


Figure 3.32: *Filtre de triangle.* Le filtre renvoie les éléments à une distance de 2 de l'agent actif. En cas d'ajout de lien entre deux de ces nœuds, une "fermeture de triangle" augmentant le coefficient de clustering des entités agissantes aura lieu. Ici, le nœud agissant est le nœud à distance "0".

Nom	Code	Action	Attribut	Filtres	Probabilité de transmission
Ajout fermeture de triangle	add-2hop	Ajout lien	degré	Non connecté ; fermeture de triangle	1
Retrait aléatoire	rmvø	Retrait de lien	-	Connecté	1
Fluidite	-	-	-	-	0

Table 3.12: *Première tentative : configuration initiale pour l'obtention d'un réseau Small World.* Le réseau est pourvu de deux complexe-actions aux effets dissymétriques : l'ajout de lien doit mener à quelques nœuds plus fortement connectés que la plupart des autres. On y ajoute aussi les actions de fluidité.

Version 2 - modifications apportées : ajout aléatoire

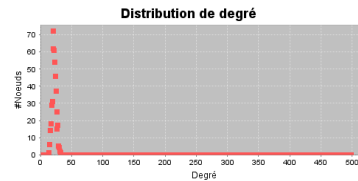
Le principal problème de cette configuration survient lorsque tous les complexe-actions d'ajout de lien aléatoire sont remplacés par ceux de fermeture de triangle. Il n'existe plus d'action qui permette de créer des liens avec les entités coupées du réseau. Nous allons donc essayer de **rajouter une action d'ajout aléatoire**. On part donc d'un réseau vide, $N=500$, et les actions données dans le tableau 3.13. On rajoute aussi une notion de porteuse initiale, l'entité possédant un complexe-action à l'initialisation de la simulation. Afin que les complexe-actions ne se perdent pas, les porteuses initiales ne peuvent remplacer leur complexe-action par un autre au cours de la simulation.

Nom	Code	Action	Attribut	Filtres	Probabilité de transmission
Ajout fermeture de triangle	add-2hop	Ajout lien	degré	Non connecté ; fermeture de triangle	1
Retrait aléatoire	rmvø	Retrait de lien	-	Connecté	1
Ajout aléatoire	addø	Ajout de lien	-	Connecté	1
Fluidite	-	-	-	-	0

Table 3.13: Version 2- Complexe-actions choisies pour l'obtention d'un réseau de type Small-World. Des complexe-actions de fluidité y sont aussi ajoutés.

Résultats L'addition à la simulation d'un [ajout aléatoire] permet effectivement de ne plus avoir d'entité déconnectée au réseau (figure 3.34). Néanmoins, les deux complexe-actions d'ajout pouvant désormais s'appliquer sans échec, cette configuration rencontre un autre problème : **il n'y a pas de stabilité sur leur répartition parmi les entités** (fig 3.34b). En terme de propriété, le réseau obtenu est très proche d'un réseau aléatoire, avec néanmoins un clustering légèrement plus important, et une distribution de degrés plus "étalée" (fig 3.33). Cela est dû au fait que l'ajout aléatoire est plus souvent appliqué et réduit grandement l'effet escompté par l'application de fermeture de triangle.

Réseau	Clustering	APL
Généré	0.071	2.091
aléatoire	0.042	2.316
SW	0.6541	4.63



(a) Propriétés de clustering et d'APL pour le réseau généré, un réseau aléatoire, et un réseau SW avec $P=0.01$ (b) Distribution de degrés du réseau aléatoire de même densité que le réseau $N=500 ; [addø][add-2hop][rmvø]$, à titre de comparaison avec celui obtenu figure 3.34c

Figure 3.33: Version 2- Comparaison du résultat avec un réseau aléatoire et un réseau Small-World de même densité (0.045), avec 500 nœuds

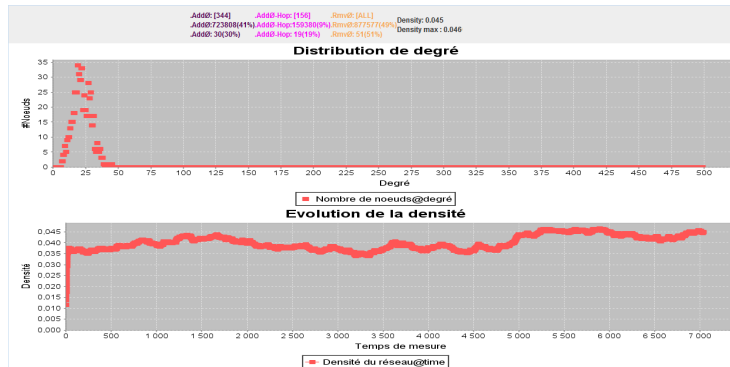
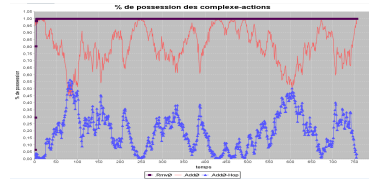
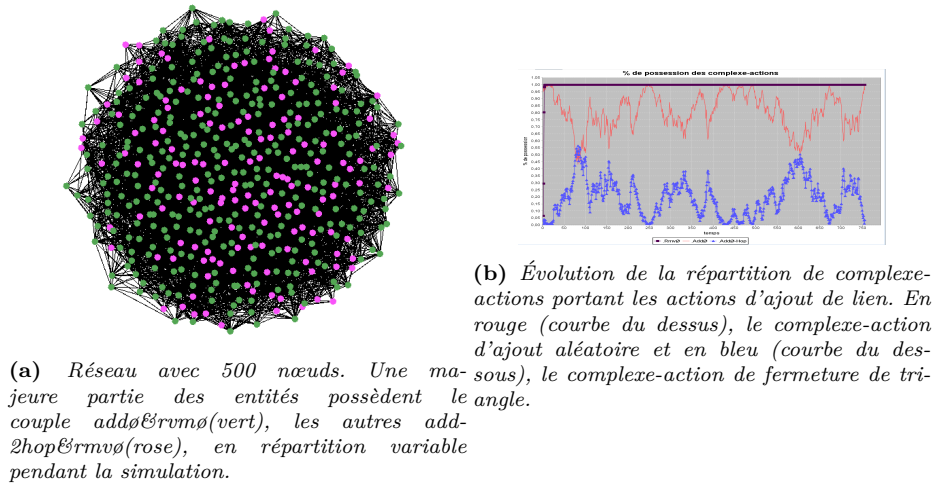


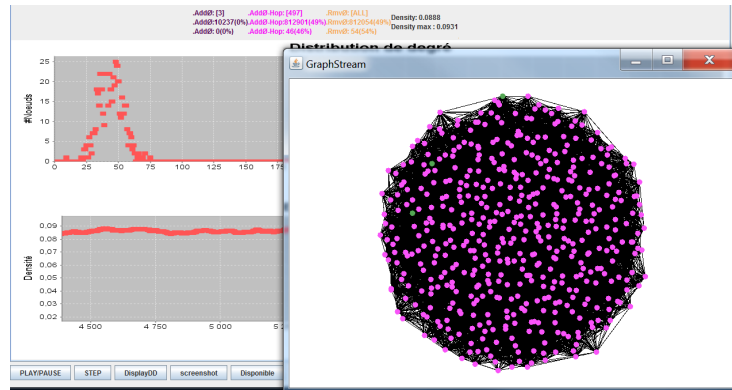
Figure 3.34: Version 2- Réseau généré avec les complexe-actions [ajout fermeture de triangle], [ajout aléatoire] & [retrait aléatoire]. $N=500$ nœuds. Densité :0.045, clustering moyen 0.071, APL 2.091. Les complexe-actions portant les actions d'ajout aléatoire sont globalement plus possédés et appliqués par les entités.

Version 3- Modifications apportées : modification de la probabilité de transmission

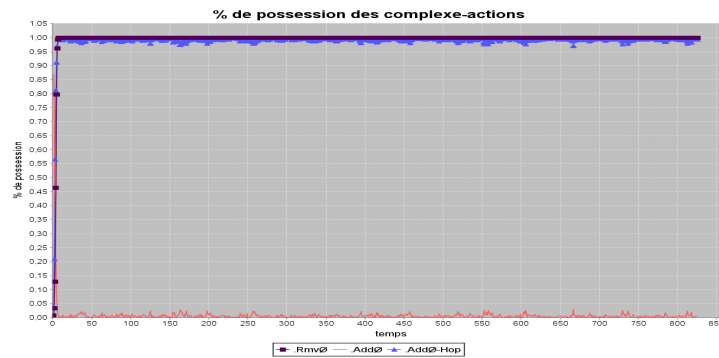
Afin d'avoir un plus grand contrôle sur la distribution finale des complexe-actions d'ajouts de liens, **nous allons jouer sur la probabilité de transmission de l'[ajout aléatoire]**. Les autres paramètres restent identiques (tableau 3.14). Le fait de modifier la probabilité de transmission permet effectivement d'influencer la répartition finale des complexe-actions. Il faut déterminer cette probabilité en prenant en compte que :

- une trop forte valeur rend instable la distribution finale de complexe-actions d'ajout et
- au contraire une valeur trop faible réduirait le nombre d'ajouts aléatoires et le comportement de la dynamique serait similaire à celui de l'expérience précédente.

Expérimentalement, une valeur autour de $P=0.6$ est trouvée. La répartition des complexe-actions d'ajouts reste stable, et aucun nœud ne se retrouve coupé du réseau (figure 3.35). **Néanmoins, les valeurs de clustering restent toujours aussi faibles.** Cela remet en doute l'efficacité du seul filtre de fermeture de triangle pour trouver une configuration de type Small-World.



(a) Réseau et distribution de degrés. Seules quelques entités possèdent le complexe-action d'ajout aléatoire.



(b) Répartition des complexe-actions : seules trois entités possédant l'ajout aléatoire suffisent à empêcher la dégénérescence du réseau vers un réseau vide.

Figure 3.35: Version 3- Réseau généré avec les complexe-actions [ajout fermeture de triangle[1]], [ajout aléatoire[0.6]] & [retrait aléatoire[1]]. $N=500$ nœuds. Densité : 0.089, clustering moyen 0.095, APL 1.93

Nom	Code	Action	Attribut	Filtres	Probabilité de transmission
Ajout degré sup.	add+	Ajout lien	degré	Non connecté ; degré supérieur à l'agissant	1
Retrait aléatoire	rmvø	Retrait de lien	-	Connecté	1
Ajout aléatoire	addø	Ajout de lien	-	Connecté	0.6
Fluidité	-	-	-	-	0

Table 3.14: *Version 3- Complexe-actions choisies pour l'obtention d'un réseau de type Small-World, en troisième approche. Des complexe-actions de fluidité y sont aussi ajoutés.*

Maintenant que nous nous ouvrons à fixer les probabilités de transmission dans les configurations initiales, le nombre de possibilités explose. Nous allons donc fixer un modèle, et nous l'explorerons à l'aide d'algorithmes génétiques.

3.4.7 Conclusion

Les expérimentations ont mis en avant :

- la nécessité de mettre en place certains mécanismes afin d'**éviter des situations de blocage** dans la simulation (complexe-actions de fluidité, nombre d'actions maximum par simulation),
- **la sensibilité des résultats** - réseaux atteignables et reproductibilité - aux choix de paramétrages et
- **l'effet "bootstrap"** : la propagation des complexe-actions est la première phase de la simulation. À cela succède la dynamique de construction de réseau, qui se stabilise à la condition d'une stabilisation des transmissions de comportements.

3.5 Conclusion du Framework et de ses expérimentations

3.5.1 Retour sur les objectifs

Nous allons revenir sur les objectifs initiaux afin de mettre en avant la façon dont ils ont été résolus au mieux dans les choix du Framework.

A-Répétition des résultats possibles & F-Indépendance aux premières actions prises *A- Possibilité de générer, pour plusieurs simulations de configurations équivalentes, des réseaux synthétiques très similaires en terme de propriétés mesurées.*

F- Que le résultat final ne soit pas dépendant de l'ordre des actions.

La même configuration est répétée plusieurs fois, et l'écart-type des propriétés est pris en compte dans le score associé afin de privilégier les configurations les plus stables. Au niveau des mécanismes du Framework, aucun n'est borné en nombre d'itérations : les agents peuvent apprendre un nombre de fois illimité... Seul le nombre d'actions dans la simulation est limité, pour le cas où aucune dynamique amenant une stabilité ne soit atteinte.

B-Espace des réseaux atteignables large *Pouvoir reproduire un grand nombre de réseaux*

Aucun choix fait au niveau du Framework aide sur ce point, mais sa conception technique permet d'ajouter facilement de nouveaux comportements.

C-Très peu de configuration *Garantir une utilisation simple sans connaissance experte en mathématiques, informatique, ou du domaine métier dont est issu le réseau à reproduire.*

Interface graphique simple d'utilisation.

D-Espace de paramètre réduit *Doit permettre une recherche automatisée efficace des résultats.*

Usage de propagation qui ne demande pas de définir des comportements pour tous les agents de la simulation.

E-Continuité *Monotonie de l'évolution des formes de réseaux obtenues en fonction de légères modifications dans les paramètres.*

Les effets de seuil dans le fonctionnement interne sont à éviter. Ainsi, les paramètres sont des réels, de nature continue. Les seuls paramètres à effet de seuil

sont les booléens définissant l'existence dans la simulation d'un complexe-action particulier.

G-Etude de la résilience *Que le modèle donne la possibilité d'étudier l'évolution du réseau après perturbation.*

La construction du réseau se fait dynamiquement lors de la simulation. Il est donc facile d'en modifier les éléments et de laisser les interactions le reconstruire. La méthode de propagation des comportements assure que d'éventuels nouveaux nœuds sont rapidement intégrés à la dynamique de reconstruction du réseau.

3.5.2 Instanciation en modèles

Les deux chapitres suivants vont détailler l'instanciation du Framework en deux modèles particuliers. Ces modèles vont proposer des variations dans la mise en pratique de certains mécanismes.

3.6 Résumé du chapitre

En partant des objectifs du travail de thèse, nous avons élaboré un Framework, -AL-, définissant le cadre le plus précis possible pour répondre aux besoins. Les expérimentations qui y ont été menées vont permettre de définir les derniers paramètres afin d'instancier le Framework en modèle fonctionnel. Les difficultés rencontrées concernent l'équilibrage d'un environnement à la fois suffisamment contraint afin de pouvoir reproduire des résultats, mais aussi ouvert afin de pouvoir atteindre une grande variété de réseaux.

Chapitre 4

Modèle I

Sommaire

4.1	Définition du modèle I	118
4.1.1	Les complexe-actions	118
4.1.2	Propagation	119
4.1.3	Scoring	119
4.2	Résultats du modèle I	123
4.2.1	Configurations et réseau à reproduire	123
4.2.2	Résultats des simulations	124
4.2.3	Retour sur les objectifs	136
4.2.4	Discussions et élaboration du modèle II	138

Dans ce chapitre, **nous allons définir une première instanciation du Framework en modèle fonctionnel et en étudier les résultats**. La première section, *Définition du modèle I*, va décrire les choix de mécanismes. La partie suivante, *Résultats du modèle I*, évaluera les résultats obtenus par l'utilisation de ce modèle pour copier des réseaux Small-World, Scale-Free et aléatoire.

4.1 Définition du modèle I

Modèle I	
├	Définition du modèle I
└	Résultats du modèle I

Fournir un modèle capable de remplir les objectifs (résumés dans la partie *section 3.1:Les objectifs du modèle*) passe par une mise en place de certains mécanismes et par leurs configurations précises, proposant ainsi une instanciation particulière.

Cette section va détailler les choix pris à différents niveaux.

4.1.1 Les complexe-actions

Le nombre d'actions que peuvent porter les entités **est limité à deux** : une place pour une action d'ajout de lien et une pour une action de retrait de lien. Si les entités avaient à disposition tous les complexe-actions, il faudrait définir une méthode de sélection assurant une non-homogénéité permettant l'apparition de topologies particulières.

Attribut Le seul attribut utilisé pour comparer deux nœuds est leur **degré**. Il s'agit d'un attribut générique, qui ne concerne que la topologie du réseau en cours de construction. L'hypothèse prise, à l'instar du modèle Dk-graph ([Mahadevan et al., 2006](#)), est que reproduire les structures en s'attachant uniquement aux degrés des nœuds à un niveau de complexité suffisamment élevé permet de générer un réseau de famille similaire.

Actions associées Les actions disponibles pour être intégrées aux complexe-actions sont **l'ajout et le retrait de liens**. Ces deux actions sont les éléments les plus simples, nécessaires et suffisants pour la construction du réseau. Une extension possible serait l'ajout et le retrait de nœuds, afin d'essayer de reproduire des dynamiques plus complètes de construction de réseau avec évolution des populations d'entités concernées.

Filtres Nous proposons un ensemble de filtres, issus des expérimentations sur le Framework *section 3.4:Expérimentations sur le Framework* pour reproduire des formes particulières de réseaux observés dans la littérature, notamment Small-World et sans échelle (*Scale-Free*) ([Watts and Strogatz, 1998](#)) ([Barabasi and Albert-László, 1999](#)). Les filtres prennent en entrée une liste d'entités et en renvoient un sous-ensemble en fonction d'une condition. Cette dernière est souvent une comparaison entre l'attribut de l'entité qui appelle le complexe-action et de ceux de l'ensemble de départ.

Les filtres disponibles sont :

- **Aléatoire** : renvoie aléatoirement une entité unique depuis la liste en entrée.

- Lien **connecté** [non connecté] : renvoie le sous-ensemble des entités connectées à l'entité appliquant le complexe-action.
- Attribut **inférieur** [supérieur-égal-différent] : renvoie les entités possédant une valeur d'attribut inférieure à celle de l'entité agissante.
- Attribut **maximum** [minimum] : renvoie les entités possédant la valeur d'attribut maximum parmi celles de la liste.
- **Distance N** : renvoie la liste des entités atteignables en un minimum de N entités intermédiaires depuis l'agissante. Lorsque N=2, il s'agit du filtre "triangle".

Les complexe-actions sont dotés de zéro à plusieurs filtres disposés dans un ordre précis.

Liste des complexe-actions Un complexe-action est donc composé d'une combinaison :

- d'actions de modification des liens du réseau et
- d'un ensemble, potentiellement vide, de filtres, pouvant porter sur des attributs ou non.

Nous créons une série de complexe-actions qui nous semblent potentiellement pertinent pour des modifications de forme topologique. Rajouter des possibilités ne fera que pénaliser le temps de recherche pour l'algorithme génétique mais augmentera l'expressivité du modèle. Les complexe-actions pour le modèle I sont donnés dans le tableau 4.1.

4.1.2 Propagation

Le mécanisme de propagation permet aux dynamiques de construction du réseau de se répartir sur toutes les entités.

- La propagation utilisée est direct, une entité *A* appliquant un complexe-action à une entité *B* lui transmettra avec une probabilité portée par le complexe-action.
- **L'entité sur laquelle est placée un complexe-action à l'initialisation ne pourra pas le remplacer au cours de la simulation.** Cela évite que des comportements s'appliquant moins fréquemment disparaissent.
- Il n'y a pas de nombre d'apprentissages maximum pour une entité.

4.1.3 Scoring

Les propriétés auxquelles nous nous intéressons sur les réseaux sont numériques et quantitatives. La définition d'un score de distance entre les propriétés du réseau cible fourni et celui obtenu par simulation est donc assez directe. **L'objectif est d'agrèger le score de plusieurs propriétés**, sans survaloriser certaines d'entre elles. Cela implique :

- de normaliser les scores obtenus pour chaque propriété, pour qu'aucune n'ait plus d'importance que les autres et

Nom	Action & [filtres]	Explications
$Add\emptyset$	ajout [Non lié \rightarrow Aléatoire]	Ajout d'un lien vers une nœud non liée
$Rmv\emptyset$	retrait [Lié \rightarrow Aléatoire]	Suppression de l'un des liens
$Add+$	Ajout [non lié \rightarrow Degré sup. \rightarrow Aléatoire]	Ajout d'un lien vers un nœud de degré supérieur
$Rmv+$	Retrait [lié \rightarrow Degré sup. \rightarrow Aléatoire]	Retrait d'un lien depuis un nœud de degré supérieur
$Add-$	Ajout [non lié \rightarrow Degré inf. \rightarrow Aléatoire]	Ajout vers nœud de degré inférieur
$Rmv-$	retrait [lié \rightarrow Degré inf. \rightarrow Aléatoire]	Retrait depuis nœud de degré inférieur
$AddEq$	Ajout [non lié \rightarrow Degré inf. \rightarrow Aléatoire]	Ajout vers nœud de degré égal
$RmvEq$	retrait [lié \rightarrow Degré inf. \rightarrow Aléatoire]	Retrait depuis nœud de degré égal
Add°	Ajout [Degré min. \rightarrow Aléatoire]	Ajout si non existant vers un nœud de degré maximum sur le réseau
Rmv°	retrait [Degré min. \rightarrow Aléatoire]	Retrait si existant depuis un nœud du réseau de degré minimum
$Add^\circ!$	Ajout [Non lié \rightarrow Degré min. \rightarrow Aléatoire]	Ajout d'un lien vers un nœud de degré minimum parmi les nœuds non encore connectés
$Rmv^\circ!$	retrait [Lié \rightarrow Degré min. \rightarrow Aléatoire]	Retrait d'un lien depuis un nœud de plus faible degré parmi ceux lié au nœud agissant.
$Add\infty$	Ajout [Degré max. \rightarrow Aléatoire]	Ajoute un lien si non existant vers le nœud de plus haut degré du réseau
$Rmv\infty$	Ajout [Degré max. \rightarrow Aléatoire]	Ajoute un lien si non existant vers le nœud de plus haut degré du réseau
$Add\infty!$	Ajout [Non lié \rightarrow Degré max \rightarrow Aléatoire]	Ajoute un lien vers le nœud de plus haut degré parmi ceux non connecté
$Rmv\infty!$	Ajout [Lié \rightarrow Degré max]	Ajoute si non existant vers le nœud de plus haut degré du réseau
$ADD\emptyset - 2Hop$	ajout [Distance 2 \rightarrow Non lié \rightarrow Aléatoire]	Ajout d'un lien vers un nœud à une distance de 2. "Fermeture de triangle"
$ADD\emptyset - 3Hop$	ajout [distance 3 \rightarrow Non lié \rightarrow Aléatoire]	Ajout d'un lien vers un nœud à une distance de 3
$RMV\emptyset - 2Hop$	retrait [lié \rightarrow Triangle \rightarrow Aléatoire]	Retrait d'un lien depuis un nœud N_1 , connecté à la fois au nœud agissant N_a et à un nœud N_2 , N_a et N_2 étant aussi connectés entre eux
$RmvChain$	Retrait [lié \rightarrow distance 2 \rightarrow Aléatoire]	retrait d'un lien si le nœud agissant possède un degré ≥ 2 . Permet de limiter la création de noeuds isolés

Table 4.1: Ensemble des complexe-actions créés pour le modèle instancié. De nombreux filtres de complexe-actions se terminent par l'application d'un filtre aléatoire afin de n'appliquer l'action qu'à une seule entité.

- d'éviter de prendre plusieurs fois en compte des propriétés qui se recoupent. Typiquement, la densité et le nombre de nœuds dans le réseau sont deux propriétés directement corrélées.

Soient deux réseaux R_1 et R_2 . La normalisation prend la différence des valeurs d'une propriété sur ces deux réseaux, et est divisée par la valeur maximum possible de cette différence, afin d'être entre 0 et 1. Ce résultat est alors multiplié par 100.

$$\frac{100 * |ValeurR_1 - valeurR_2|}{valeurmaximum}$$

Le plus grand écart d'une propriété correspond souvent à sa comparaison dans un réseau vide et un réseau complet.

Les propriétés considérées sont les suivantes :

- **Densité.** Maximum pour normalisation = 1 (cas R_1 complet et R_2 vide)
- **Distance inter-quartile** (1er et 3er) de la distribution de degrés. Maximum = (Nombre de nœuds - 1) (cas R_1 ayant la moitié des nœuds au degré maximum et l'autre au degré minimum, et R_2 vide)
- **Clustering moyen.** Maximum = 1 (cas R_1 complet et R_2 vide)
- **APL.** Maximum = $(n+1) / 6$; (cas R_1 chaîne et R_2 complet)
- **Similarité de forme** de la distribution de degré.

La somme de ces scores est divisée par le nombre d'attributs pris en compte et correspond donc à un score moyen obtenu sur les différentes dimensions.

La similarité de forme entre les distributions de degrés consiste à centrer l'une des distributions sur la moyenne de l'autre, puis de faire la moyenne des différences degré par degré.

Par ailleurs, nous avons rajouté une pénalité lorsque le réseau généré contenait des nœuds non connectés. Cette situation réduit le nombre de nœuds actifs dans le réseau, sans pour autant être capturée par la propriété "nombre de nœuds". Une autre possibilité aurait été de retirer en fin de simulation tous les nœuds non connectés, afin que cette propriété soit directement prise en compte par le dispositif de scoring. Néanmoins, nous considérons que le nombre de nœuds actifs n'est pas un élément vers lequel on veut tendre mais que l'on souhaite avoir précisément, et nous avons donc décidé d'en faire une pénalité appliquée à posteriori sur le score obtenu. Nous multiplions le score par le nombre de nœuds isolés.

L'intégration du calcul du coefficient de la loi de puissance γ de la distribution de degrés des réseaux de type Scale-Free est un peu délicate : cette propriété n'est par exemple pas pertinente pour des réseaux à la distribution gaussienne. Elle ne pourra être vérifiée qu'après l'obtention d'un réseau et ne sera donc pas intégrée aux métriques prises en compte lors d'une recherche automatisée des meilleurs paramètres initiaux.

Une même configuration est relancée vingt fois, afin de calculer une moyenne et un écart type sur les scores obtenus. Le score final associé à chaque configuration

est le produit mathématique de ces dernières valeurs. Cela permet d'intégrer la reproductibilité des résultats.

Illustration de l'usage de scoring le tableau 4.2 donne un exemple de la prise en compte de différentes propriétés pour définir le score associé à un réseau synthétique donné.

	Densité	InterQ	Cluster	APL
Réseau cible	0.1	4	.4	2.8
Réseau synthétique	0.15	3	.1	7
valeur maximale	1	199	1	33
Score réseau * 100	$\frac{0.15-0.1}{1} (*100) = 5$	$\frac{1}{199}=0.5$	$\frac{.3}{1}=30$	$\frac{4.2}{33} = 12.7$

Table 4.2: Illustration fictive d'un calcul de score pour une établir une distance entre deux réseaux possédant 200 nœuds. Le score du réseau généré est donc la moyenne des scores des propriétés choisies. $\frac{5+.5+30+12.7}{4}=12.05$. Le score associé à la configuration ayant abouti à ce réseau prendra en compte l'écart-type sur les scores de plusieurs réseaux générés.

Les maximums pouvant être atteints sont donnés dans le paragraphe précédent subsection 4.1.3:Scoring. #lien : nombre de liens dans le réseau, interQ : distance interquartile ($3^{eme}-1^{er}$), cluster=coefficient de clustering, APL=Average Path Length.

4.2 Résultats du modèle I

Modèle I	
└ Définition du modèle I	
└ Résultats du modèle I	

4.2.1 Configurations et réseau à reproduire

Nous allons dans un premier temps nous concentrer sur la reproduction de réseaux abstraits issu de la littérature afin de vérifier la possibilité pour le modèle I de reproduire des propriétés fondamentales des réseaux sociaux, que sont :

- un **clustering important** associé à une densité faible,
- une taille de chemin entre nœuds (**APL**) courte et
- une distribution de degrés en **loi de puissance**.

Nous nous sommes donc focalisés sur les réseaux *Small World*, *Preferential Attachment* et *Random* :

- **Random** - Nombre de nœuds [300] - Densité [0.02]
- **Small-World** - Nombre de nœuds [300] - Probabilité [0.01]
- **Preferential Attachment** - Nombre de nœuds [300] - nombre de liens par nœud [2]

Outils utilisés Maintenant que l'utilisation de la probabilité de transmission a été mise en place, le nombre de configurations initiales possibles explose. Nous utiliserons un algorithme génétique pour la recherche des meilleures configurations. L'algorithme génétique est fourni par le framework OpenMOLE¹ (Reuillon et al., 2013), et les simulations sont distribuées sur les clusters OSIRIM² mis à disposition par l'IRIT.

La visualisation des graphes est faite avec Graphstream³, et l'un de leurs algorithmes est utilisé pour le calcul de la propriété de chemin moyen (APL).

La visualisation des graphiques est faite avec JFreeChart⁴.

1. openmole.org
 2. <https://osirim.irit.fr/site/en>
 3. graphstream-project.org
 4. <http://www.jfree.org/jfreechart/>

4.2.2 Résultats des simulations

Le score prend en compte :

- la densité,
- la distance interquartile (3^{eme} - 1^{er}),
- la différence des distributions de degrés, degré par degré, après centrage sur l'une des moyennes,
- le coefficient de clustering moyen et
- la distance *Average Path Length* (APL).

L'algorithme génétique travaille sur 150 générations. Chaque complexe-action disponible est représenté par deux éléments : un booléen signifiant sa présence dans la simulation et un double compris entre 0 et 1 pour sa probabilité de transmission. La condition d'arrêt pour une configuration précise avant lecture du réseau obtenu est un nombre d'actions appelées, fixé à 10^6 .

Réseaux aléatoires Nous allons dans un premier temps nous intéresser aux réseaux aléatoires. Le but est de vérifier que le modèle est capable de reproduire différentes densités. Les deux réseaux à atteindre sont des réseaux aléatoires aux densités de 0.02 (ordre de grandeur pour les réseaux sociaux) et 0.3.

réseau aléatoire de densité 0.02 Le tableau 4.3 donne les configurations retenues après 150 générations, avec leur score moyen et écart-type associé. Le détail de ces scores est donné dans le tableau 4.4, pour chaque attribut pris en compte.

Pour les deux configurations, 9 complexe-actions sont sélectionnés pour générer un réseau. Néanmoins, passée l'étape de propagation à toutes les entités, 3 des complexe-actions deviennent largement majoritaires, aussi bien en répartition sur les entités qu'en nombre d'applications.

La première configuration donne de fortes variations de répartition de deux complexe-actions d'ajout de lien (figure 4.1 et 4.2).

L'autre configuration possède un meilleur score et est bien plus stable (figure 4.3).

Lorsque l'on s'intéresse au détail des scores obtenus, la propriété la moins bien reproduite est la similarité de forme de la courbe de distribution de degré. **Les distributions de degré obtenues sont plus étroites que celle requise** (figure 4.4). Ces résultats se confirment malgré des temps de simulations cent fois supérieurs, c'est à dire après 10^8 actions.

Config.	CA constituant et probabilité de transmission	score $[\sigma]$
1	[Add- :0.766] [Rmv ∞ ! :0.762] [Rmv+ :0.668] [Add \emptyset :0.5768] [Add $^\circ$:0.0445] [Rmv $^\circ$:0.368] [Add \emptyset -Hop :0.248] [Add ∞ :0.0] [Rmv ∞ :0.0]	11.4 [5.65]
2	[Add ∞ :1.0] [Rmv+ :1.0] [Add $^\circ$:1.0] [RmvChain :0.74] [Add+ :0.65] [Rmv \emptyset :0.575] [Rmv ∞ :0.49] [Rmv $^\circ$:0.47] [Rmv $^\circ$! :0.027]	9.7 [1.0]

Table 4.3: *Random 2% : Configurations trouvées pour atteindre un réseau aléatoire avec une densité de 0.02. Score moyen associé et écart-type sur 20 simulations par configuration. L'algorithme génétique d'OpenMOLE (Reuillon et al., 2013) réalise 150 générations, le critère d'arrêt étant 10^6 actions entreprises (avec succès ou non).*

Propriété	cible	config #1		config #2	
		valeur	score	valeur	score
interq.	2	1.2[0.5]	4.45[2.6]	1.8[0.5]	1.73[2.65]
clust. moy.	0.015	0.036[0.078]	2.59[7.63]	0.05[0.008]	3.63[0.83]
nb. lien	1800	3688[6953]	10.1[11.8]	1616.2[123.6]	5.64[0.98]
APL	3.38	3.29[0.74]	7.03[2.78]	4.19[0.22]	8.82[1.58]
Distrib deg.	-	-	33.4[7.3]	-	32.7[5.6]

Table 4.4: *Random 2% : Valeurs et scores des propriétés prises en compte pour le score de distance entre le réseau cible et les réseaux générés par les 2 meilleures configurations trouvées. Écart-type de ces valeurs sur 20 répliquions de simulation par configuration. Les valeurs des distribution de degrés n'y sont pas représentées, n'étant pas un nombre unique.*

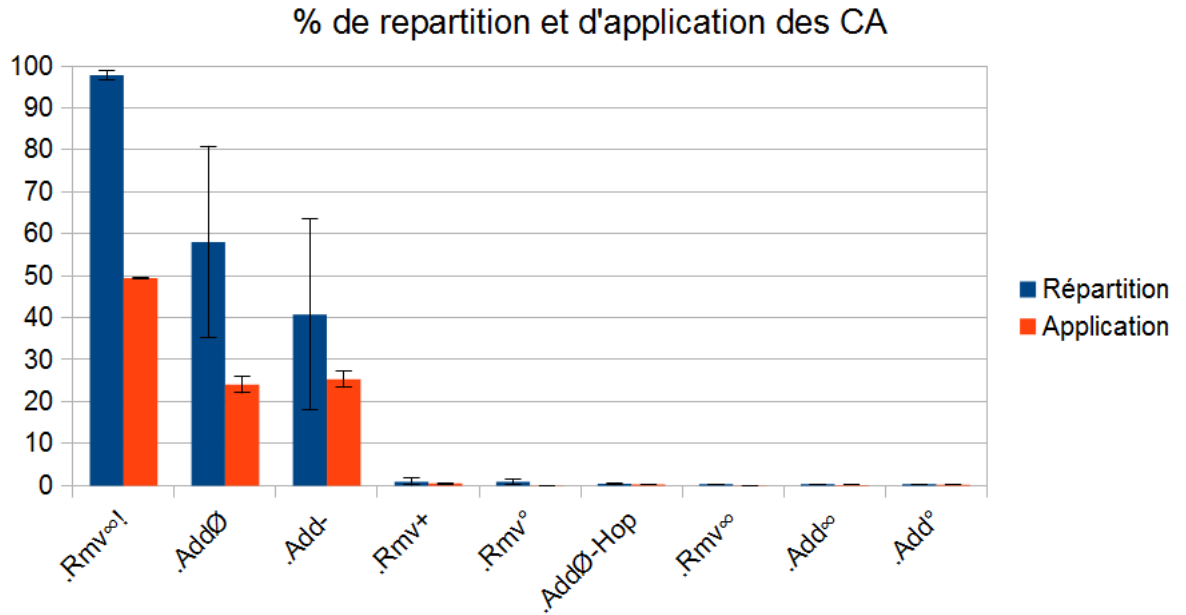


Figure 4.1: *Random 2%, configuration 1 : Taux de répartition et application des complexe-actions pour la 1er configuration, avec les écarts-types associés. La grande variation de répartition entre les complexe-actions Add \emptyset et Add- indique que la dynamique de cette configuration doit les voir alterner en terme de possession. Les applications somment à 100 tandis que les répartitions somment à 100 par catégorie d'action portée (ajout et retrait).*

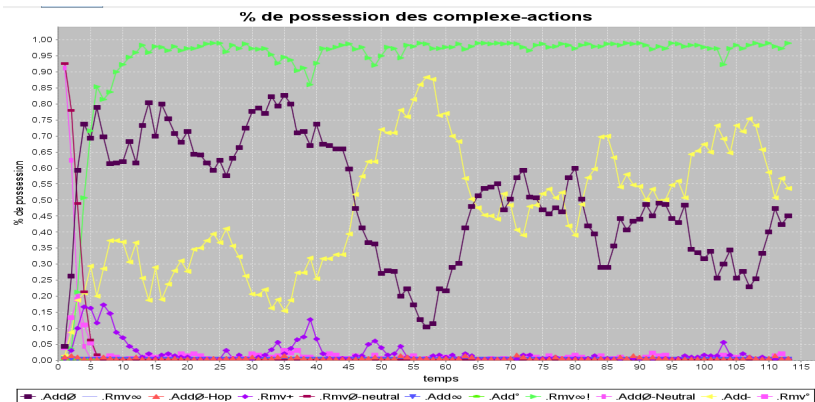


Figure 4.2: *Random 2%, configuration 1 : L'observation de la répartition des complexe-actions explique l'écart-type important pour celles de Add \emptyset et Add-. Les deux complexe-actions subissent de grandes variations de répartitions et se remplacent mutuellement au cours du temps.*

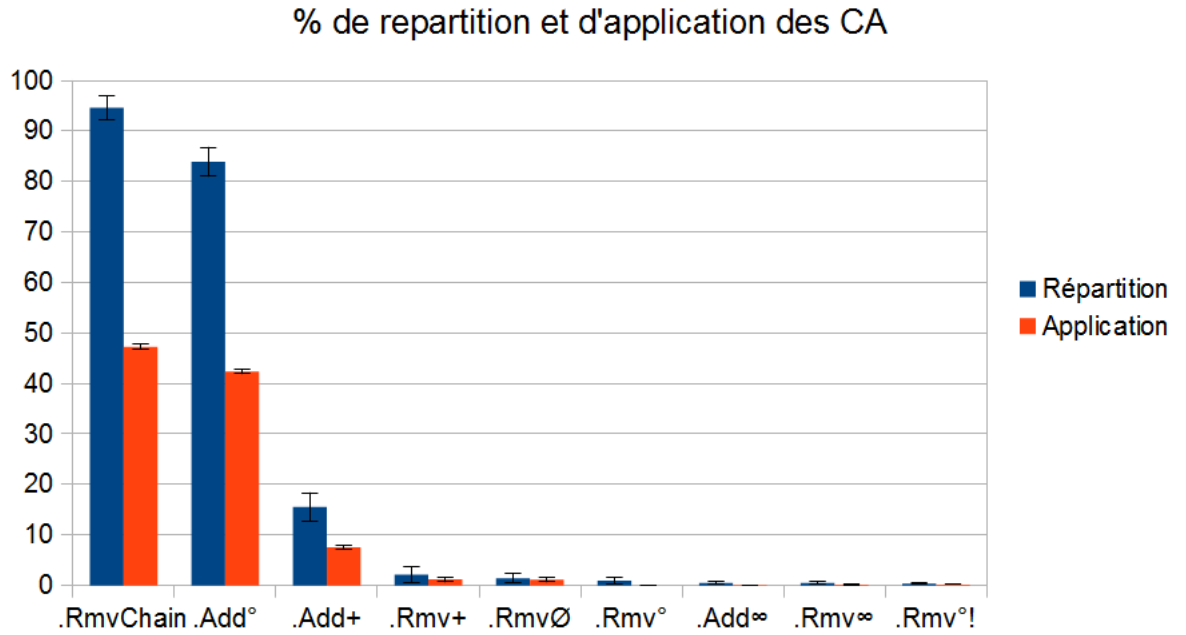


Figure 4.3: *Random 2%, configuration 2 : Taux de répartition et application des complexe-actions pour la 2e configuration, avec leurs écarts-types. On observe une plus grande stabilité dans leur répartition, et 3 d'entre d'eux écrasent en usage et présence les 5 autres.*

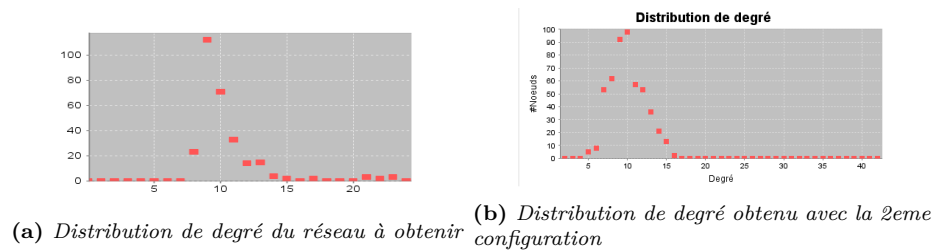


Figure 4.4: *Random 2% : distributions de degré à obtenir et obtenue. Celle du réseau à reproduire est plus étalée autour de la moyenne.*

Réseau aléatoire de densité 0.3 Nous allons chercher à reproduire un réseau aléatoire de densité 0.3, afin de vérifier si le modèle peut copier de façon consistante des réseaux à la densité plus importante.

A priori, les mêmes observations sont faites sur les configurations trouvées (tableau 4.5), avec un score amoindri par des formes de distribution de degré différentes (tableau 4.6), et où un petit nombre de complexe-actions supplantent les autres. **Néanmoins, une fois la limite du nombre d'actions levée, le nombre de configurations efficaces diminue grandement.** Sur ces temps de simulation plus long, la configuration 1 garde ses propriétés, tandis que la configuration 2 continue à tendre vers un réseau de densité 0.7 (figure 4.5). L'algorithme génétique a donc trouvé ici une configuration atteignant la bonne densité de réseau aléatoire dans le nombre d'actions imparties, mais ne fonctionnant pas si cette limite est dépassée.

Config.	CA constituant et probabilité de transmission	score $[\sigma]$
1	[.Add $^\circ$:0.98][Rmv $^\circ$:0.81][RmvEq :0.77][Rmv $^\circ$! :0.77][Rmv \emptyset -2hop :0.71][Rmv- :0.59][Add \emptyset -3Hop :0.57][Rmv \emptyset :0.43][Rmv ∞ :0.18]	14.45 [0.77]
2	[Add \emptyset -3Hop :0.54][Add $^\circ$:0.54][RmvEq :0.88][Rmv-Chain :0.07][Rmv+ :0.2][AddEq :0.82][Add+ :0.94][Add \emptyset -Hop :0.93][Rmv ∞ :0.75]	12.7 [1.1]

Table 4.5: *Random 30% : configurations trouvées pour atteindre un réseau aléatoire avec une densité de 0.02. Score moyen associé et écart-type sur 20 simulations par configuration. L'algorithme génétique d'OpenMOLE (Reuillon et al., 2013) réalise 150 générations, le critère d'arrêt est un nombre d'actions limite d'un million.*

Propriété	cible	config #1		config #2	
		valeur	score	valeur	score
Interq.	8	1.5[0.5]	2.17[0.17]	2.8[0.6]	1.739[0.201]
Clust. moy.	0.3	0.263[0.012]	3.709[1.164]	0.05[0.008]	3.63[0.83]
Nb Lien	27000	24538.8[1032]	2.7 [1.15]	32620.4[2582]	6.27[2.88]
APL	1.7	1.73[0.012]	0.03[0.012]	1.627[0.031]	0.07[0.03]
Distrib deg.	-	-	62.97[2.68]	-	47.3[1.49]

Table 4.6: *Random 30% : valeurs et scores des propriétés prises en compte pour le score de distance entre le réseau cible et les réseaux générés par les 2 meilleures configurations trouvées. Écart-type de ces valeurs sur 20 répliques de simulation par configuration. Les valeurs des distribution de degrés n'y sont pas représentées, n'étant pas un nombre unique.*

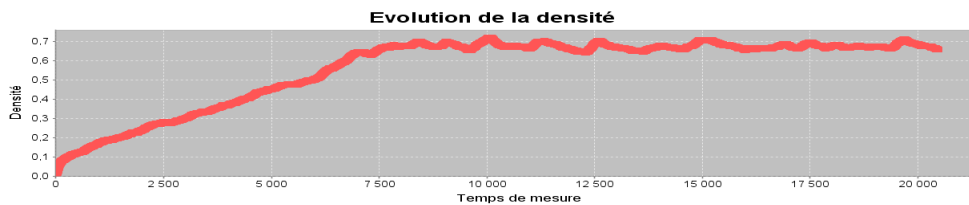


Figure 4.5: *Random 30%, configuration 2 : Sur des temps de simulation plus long, la configuration tend vers un réseau à densité de 0.7*

Résultats sur les réseaux aléatoires Excepté pour la forme de la distribution de degré, les résultats pour reproduire les réseaux aléatoires sont plutôt bons. Néanmoins, pour les réseaux aléatoires de densité 0.3, certaines des configurations trouvées permettent de reproduire ces réseaux par un état transitoire vers des états de plus forte densité. Il va donc falloir être attentif aux comportements des configurations trouvées aux limites.

Scale Free Nous allons maintenant nous intéresser aux réseaux sans échelle, générés par le modèle d'attachement préférenciel (avec paramètre $m=2$ - commençant depuis un réseau vide, chaque nœud créant deux liens en "arrivant" sur le réseau).

Les deux meilleures configurations que nous trouvons sont données dans le tableau 4.7, et dans le détail de leurs scores dans le tableau 4.8. Concernant la domination de certain complexe-actions, on retrouve le même cas de figure que pour les réseaux aléatoires (figures 4.6 et 4.7). **Le coefficient de clustering des réseaux obtenus est trop faible, et s'approche de la densité, marque des réseaux aléatoires.** Lorsque l'on s'intéresse dans le détail à la configuration ayant obtenue le meilleur score (figure 4.8), le réseau obtenu est assez intéressant. **Il est composé de deux sous réseaux, l'un fortement connecté et l'autre moins dense.** Une fois de plus, un petit nombre de complexe-actions dominent les répartitions et applications. Comme présenté dans la figure 4.8, la configuration propose un réseau quelque peut trop dense, avec deux sous réseaux formés. Lorsque l'on s'intéresse au calcul du gamma, ce dernier est de 3.2, ce qui est au dessus de la limite pour le considérer comme un réseau Scale-Free strict.

Config.	CA constituant et probabilité de transmission	score $[\sigma]$
1	[RmvØ-2hop :0.72] [RmvChain :0.70] [Add° :0.65] [Rmv∞! :0.62][Add∞ :0.45] [RmvEq :0.4][AddEq :0.12] [Rmv° :0.08]	14 [2.5]
2	[Add°! :1.0] [Add- :1.0] [RmvChain :1.0] [Rmv- :0.82] [Add∞ :0.526] [Rmv∞! :0.25] [RmvØ :0.238] [Add° :0.18] [AddØ-3Hop :0.0]	10.6 [1.13]

Table 4.7: SF : configurations trouvées pour atteindre un réseau sans échelle. Score associé et écart type sur 20 simulations par configuration. L'algorithme génétique d'OpenMOLE réalise 150 générations (Reuillon et al., 2013), le nombre d'actions limitée est de 900000.

Propriété	cible	config #1		config #2	
		valeur	score	valeur	score
interq.	2	1.15[0.36]	4.916[2.065]	1.9[0.3]	0.578[1.735]
clust. moy.	0.058	0.011[0.004]	4.7[0.425]	0.013[0.004]	4.498[0.437]
densite	0.013	0.014[0.002]	3.605[1.423]	0.015[0.001]	4.109[1.606]
APL	3.53	4.245[0.35]	8.163[2.098]	4.011[0.218]	6.697[1.702]
Distrib deg.	-	-	48.63[10.15]	-	37.05[4.577]

Table 4.8: *SF* : valeurs et scores des propriétés présent en compte pour le score de distance entre réseau cible et réseaux générés par les 2 meilleures configurations trouvées. Écart-type de ces valeurs sur 20 itérations de simulation par configuration. Les valeurs de distribution de degré n'y sont pas représentées.

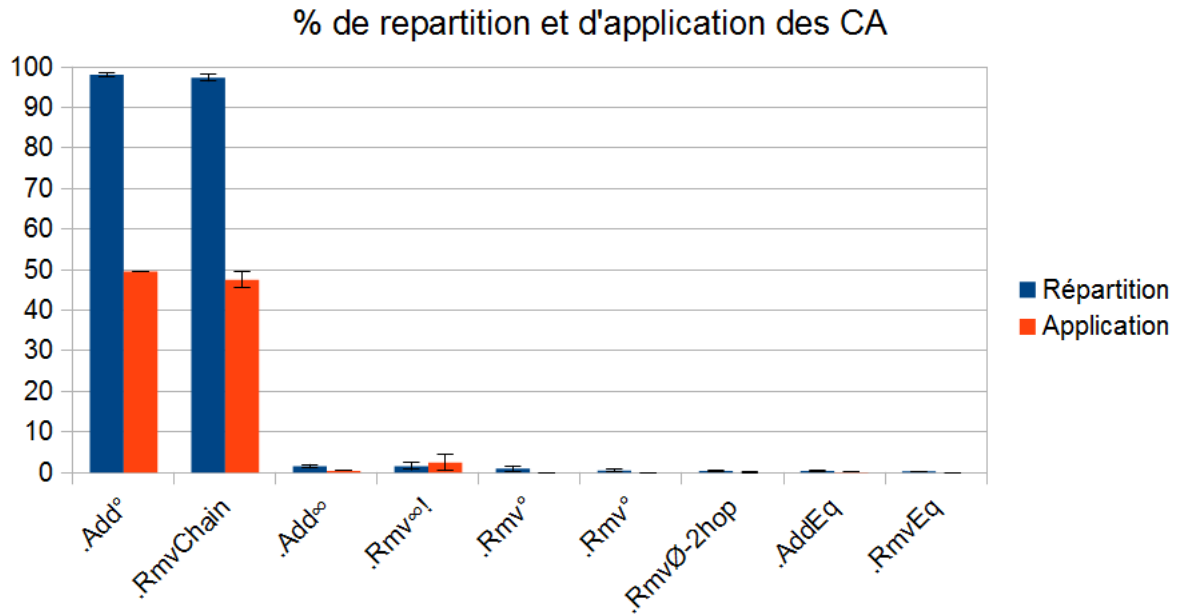


Figure 4.6: *SF*, configuration 1 : Les applications somment à 100 tandis que les répartitions somment à 100 par catégorie d'action portée (Ajout et retrait).

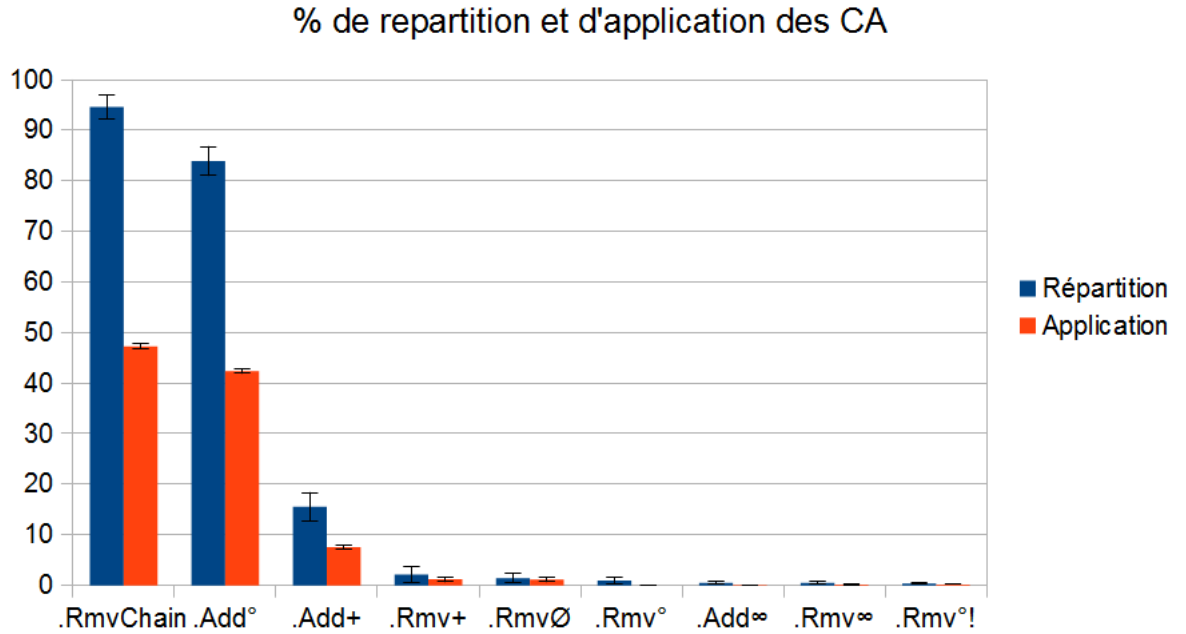
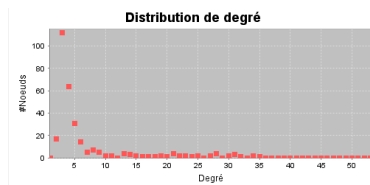
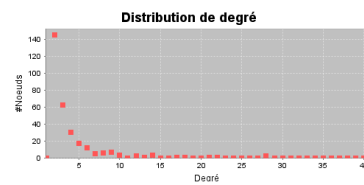


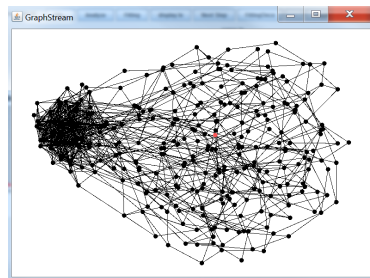
Figure 4.7: SF, configuration 2 : Taux de répartition et application des complex-ions pour la 2eme configuration, avec leurs écarts-types.



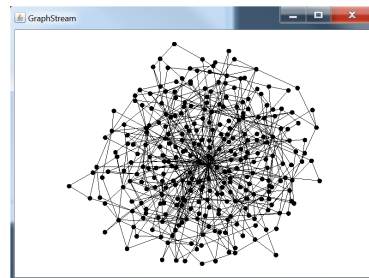
(a) Distribution de degrés de la 2eme configu-ration



(b) Distribution de degrés du réseau SF à re-produire



(c) Réseau associé à la deuxième configuration



(d) Réseau SF à obtenir

Figure 4.8: SF, configuration 2 : réseau et sa distribution de degrés ayant obtenu les meilleurs résultats pour la reproduction d'un réseau Scale-Free

Small World Le réseau Small-World que nous prenons pour objectif est issu d'une grille régulière, chaque nœud étant connecté à 7 autres nœuds, afin d'obtenir une densité de réseau à 0.046 lorsque 300 nœuds sont présents. Ce choix est en accord avec les densités observées dans les réseaux sociaux en ligne, comme discuté dans le chapitre Framework.

Le tableau 4.9 propose deux configurations aux scores équivalents. Lorsqu'on les regarde dans le détail (tableau Small World), on constate que la manière d'y parvenir est bien différente : **l'un des réseaux propose une densité conforme et un coefficient de clustering trop faible, l'autre l'inverse. Comme pressenti avec les résultats pour le réseau Scale-free, le modèle ne trouve pas de configuration dissociant grandement les valeurs de densité et de coefficient de clustering**; les réseaux générés sont encore trop proches des réseaux aléatoires. Nous allons donc nous concentrer sur la recherche d'une configuration possédant à la fois une densité de réseau très faible et un clustering très important, en modifiant le calcul du score associé à un réseau.

Config.	CA constituant et probabilité de transmission	score $[\sigma]$
1	[RmvChain :0.61] [Rmv $^\circ$! :0.53] [Rmv- :0.38] [Rmv+ :0.326] [Rmv $^\infty$! :0.32] [Add $^\circ$! :0.28] [Add- :0.11] [Add \emptyset -3Hop :0.1]	26.16 [1.45]
2	[Rmv $^\infty$! :1.0] [RmvEq :0.98] [Add \emptyset :0.92] [Add $^\infty$:0.9] [Rmv \emptyset -2hop :0.53] [Add \emptyset -Hop :0.42] [Add $^\circ$:0.013] [Add $^\circ$! :0]	27.8 [1.3]

Table 4.9: *SW* : configurations trouvées pour atteindre un réseau Small-World. Score associé et écart-type sur 20 simulations par configuration. L'algorithme génétique d'OpenMOLE réalise 150 générations (Reuillon et al., 2013), le nombre d'actions limite est de 900000.

Propriété	cible	config #1		config #2	
		valeur	score	valeur	score
interq.	0	2.7 [0.46]	0.9 [0.153]	5.7 [1.95]	1.9 [0.65]
clust. moy.	0.65	0.5 [0.02]	15.76 [2.25]	0.06 [0.02]	59.3 [1.9]
densite	0.047	0.5 [0.03]	45.2 [2.24]	0.07 [0.02]	3.3 [0.9]
APL	3.98	1.5 [0.02]	2.47 [0.02]	2.19 [0.32]	1.8 [0.32]
Distrib deg.	-	-	56 [2.5]	-	73 [6.44]

Table 4.10: *SW* : valeurs et scores des propriétés prises en compte pour le score de distance entre réseau cible et réseaux générés par les 2 meilleures configurations trouvées. Écart-type de ces valeurs sur 20 itérations de simulation par configuration. Les valeurs de distribution de degré n'y sont pas représentées.

Densité et clustering d'un réseau Small World Les résultats donnés jusqu'à présent ne proposent pas de configuration entraînant un coefficient de clustering significativement supérieur à la densité du réseau obtenu. Nous allons isoler ces deux propriétés pour la définition du score.

Le réseau cible reste un Small-World, généré par l'algorithme de (Watts and Strogatz, 1998), mais les propriétés prises en compte pour le score de distance entre le réseau généré et celui à reproduire se limitent au coefficient de clustering et à la densité. On ne peut se contenter de prendre en compte que le clustering, car un réseau aléatoire possède un coefficient de clustering égal à sa densité, et une solution évidente serait alors un réseau de densité équivalente au clustering nécessaire.

Après 300 générations par algorithme génétique, trois configurations ressortent 4.11. Aucune d'entre elles n'obtient une topologie éloignée de celle des réseaux aléatoires : le clustering et la densité du réseau restent très semblables. Il est intéressant de constater que l'algorithme a trouvé trois manières d'obtenir le même score moyen :

1. une bonne reproduction d'une valeur élevée du clustering (configuration 2), au prix d'une valeur de densité trop élevée ;
2. un compromis entre les deux, clustering et densité moyens ;
3. un clustering faible, impliquant une bonne reproduction de la densité (configuration 3).

Config.	CA constituant et probabilité de transmission	score[σ]
1	[Add- :1.0] [Add $^\circ$:0.83] [Add \emptyset -Hop :0.79] [Add ∞ ! :0.72] [Add \emptyset :0.44] [Add \emptyset -3Hop :0.43] [Rmv- :0.43] [Add ∞ :0.36] [RmvEq :0.35] [RmvChain :0.29] [Rmv ∞ :0.27] [Rmv \emptyset -2hop :0.10] [Add+ :0.10] [Rmv $^\circ$:0.05]	29.5 [0.1]
2	[Add \emptyset :1.0] [Rmv $^\circ$:1.0] [Rmv ∞ :1.0] [Rmv $^\circ$! :0.87] [Rmv+ :0.84] [Rmv \emptyset :0.73] [Add+ :0.53] [RmvChain :0.51] [Add ∞ :0.49] [Rmv \emptyset -2hop :0.02] [RmvEq :0.002] [Add \emptyset -Hop :0.001] [Add- :0.0]	20.8 [1.2]
3	[Rmv \emptyset :0.73] [Add- :0.71] [Add \emptyset :0.69] [Add+ :0.55] [Add ∞ :0.52] [Rmv $^\circ$! :0.43] [Rmv $^\circ$:0.42] [Rmv \emptyset -2hop :0.01] [Rmv ∞ :0.32] [Add \emptyset -Hop :0.26]	29.9 [0.22]

Table 4.11: Configurations trouvées pour atteindre les propriétés de densité et coefficient de clustering d'un réseau small-World. Score associé et écart type sur 20 simulations par configuration. L'algorithme génétique d'OpenMOLE réalise 300 générations (Reuillon et al., 2013), le nombre d'actions limite est de 900000.

Propriété	cible	config #1		config #2		config #3	
		valeur	score	valeur	score	valeur	score
clust. moy.	.65	0.35 [0.02]	30.0 [1.95]	0.5 [0.02]	14.9 [2.28]	0.11 [0.01]	54.06 [1.72]
densite	.047	0.34 [0.02]	28.9 [2.07]	0.3 [0.01]	26.7 [0.87]	0.10 [0.02]	5.70 [1.76]

Table 4.12: SW score réduit : valeurs et scores des propriétés prises en compte pour le score de distance entre réseau cible et réseaux générés par les 2 meilleures configurations trouvées. Écart-type de ces valeurs sur 20 itérations de simulation par configuration. Les valeurs de distribution de degré n'y sont pas représentées.

Observations sur les résultats obtenus Avant même de s'intéresser dans le détail aux aspects reproductibilité et changement d'échelle qu'offre le modèle I, **on observe que les réseaux obtenus ne sont pas intéressants concernant le rapport entre coefficient de clustering et densité.** Effectivement, les réseaux générés sont toujours proches des réseaux aléatoires en ce qu'il ne parviennent pas à obtenir un coefficient de clustering important sans pour autant avoir une densité faible.

Deux éléments expliquent cette homogénéisation du réseau :

- quelques complexe-actions prennent le dessus sur les autres, nuisant à "l'expressivité" de nombreuses configurations, et
- les complexe-actions ne restent pas nécessairement associés à la même entité, par conséquent cette dernière ne garde que difficilement une place singulière dans le réseau, comme par exemple celle de plus haut degré.

Il apparaît donc que ce modèle ne nous permet pas en l'état d'atteindre des topologies de réseau particulières et propres aux réseaux sociaux.

4.2.3 Retour sur les objectifs

Nous allons revenir sur les objectifs initiaux afin de résumer la façon dont ils ont été adressés, en précisant l'apport du modèle et du Framework.

A- Reproductibilité des résultats & F-Indépendance aux premières actions prises

A- Possibilité de générer, pour plusieurs simulations de configurations équivalentes, des réseaux synthétiques très similaires en terme de propriétés mesurées.

F- Que le résultat final ne soit pas dépendant de l'ordre des actions.

Un complexe-action de chaque type (ajout et retrait) maximum par entité; pas d'oubli de la part des entités possédant initialement un complexe-action

B-Espace des réseaux atteignables large

Avoir la possibilité de générer un grand nombre de réseaux avec le Framework, afin d'assurer un large panel de réseaux reproductibles.

Grande diversité des complexe-actions disponibles.

C-Très peu de configuration

Garantir une utilisation simple sans connaissance experte mathématique, informatique, ou du domaine métier dont est issu le réseau cible.

Automatisation de la recherche des meilleurs paramètres initiaux par un algorithme génétique.

D-Espace des paramètres réduit

Doit permettre une recherche automatisée efficace de résultats.

La multiplication des paramètres et des combinaisons d'interactions implique une configuration plus lourde. Un seul couple de complexe-actions par entité restreint cela.

E-Continuité

Monotonie de l'évolution des formes de réseaux obtenus en fonction de légères modifications dans les paramètres.

Non adressé par le modèle.

G-Etude de la résilience

Que le modèle donne la possibilité d'étudier l'évolution du réseau après perturbation.

Non adressé par le modèle.

4.2.4 Discussions et élaboration du modèle II

Du fait que les actions ne sont affectées à aucune entité en particulier et ne cessent d'être échangées, les nœuds du réseau ne restent que difficilement au centre de situations topologiques singulières. Il est donc **difficile d'atteindre des réseaux hétérogènes**.

Idéalement, il faudrait que toutes ou une partie des entités conservent leurs complexe-actions, ou que les transmissions d'actions soient plus restreintes. Cela devrait permettre aux dynamiques de construction de réseau de "s'organiser" autour ces entités particulières.

Afin d'estimer l'ordre de grandeur de la complexité de l'espace des paramètres du modèle I, on va supposer une précision des probabilités de propagation à 1%. Chacun des 20 complexe-actions compte donc pour 100 possibilités. Les booléens de présence de chaque complexe-action s'ajoutent aussi, pour une complexité de l'espace des paramètres initiaux de :

$$100(\text{proba})^{20(\text{ComplexeActions})} + 2^{20} \sim 10^{40}$$

Nous allons chercher des solutions dont l'espace des paramètres soient dans le même ordre de grandeur. Les possibilités auxquelles nous avons pensées sont :

— **Repartir sur une nouvelle base.**

Fixer les complexe-actions sur les nœuds et retirer le mécanisme de propagation. Afin de répondre à la possibilité de fournir des complexe-actions à de nouveaux nœuds, on associerait alors aléatoirement un couple de complexe-actions au nouveau nœud entrant, probabilité proportionnelle à la répartition en présence sur le réseau. Le retrait de nœud demanderait probablement un mécanisme de redistribution des complexe-actions pour éviter de voir certains complexe-actions disparaître. Le nombre de couples possibles, composés par 10 complexe-actions d'ajout et 10 de retrait, est de 100.

La complexité de l'espace de configuration initial serait alors :

$$100^{300(\text{nbdnoeuds})}$$

On peut préférer définir pour chaque couple de complexe-actions des proportions de présence sur le réseau. Les proportions se somment donc à 1. Si on leur suppose une précision de 1% :

$$C_{20}^{100} + 2^{100} \sim 10^{30}$$

— **Améliorer l'existant.**

On peut chercher à modifier le modèle I afin de pallier à ses défauts, permettre à certains nœuds de prendre des "rôles" dans le réseau

— **Réduire les échanges de complexe-actions** entre les entités et

— **Éviter "l'écrasement" de complexe-actions par d'autres.**

(Nous nous intéresserons à l'espace de paramètre de cette solution dans le chapitre suivant, partie *Configuration et réseaux à reproduire*).

Les deux solutions semblent raisonnables. Afin de pouvoir continuer à répondre aux objectifs du modèle, et pour pouvoir garder un espace des paramètres relativement faible, on cherche une solution assez proche du modèle actuel.

Chapitre 5

Modèle II

Sommaire

5.1	Définition du modèle II	142
5.1.1	Les complexe-actions	142
5.1.2	Propagation	142
5.1.3	Scoring	143
5.2	Résultat du modèle II	145
5.2.1	Configuration et réseaux à reproduire	145
5.2.2	Résultats des simulations	146
5.2.3	Discussion sur les résultats	153
5.2.4	Retour sur les objectifs	153

Le modèle précédent à obtenu de bon résultat pour reproduire des réseaux aléatoires mais **n'a pas été en mesure de générer des réseaux aux topologies plus variées**, principalement pour la propriété de coefficient de clustering. Nous postulons que cela est dû au fait que les complexe-actions peuvent se propager à toutes les entités sans discernement. Aucune entité ne prend alors un "rôle" particulier sur la durée, et le réseau se dirige vers une structure moyenne, aléatoire. **Nous allons essayer de pallier à cela en définissant des "profils" de comportement**, ainsi qu'en contraignant plus fortement leurs conditions de propagation. La première section, *Définition du modèle II*, va décrire les choix de mécanismes. La partie suivante, *Résultat du modèle II*, évaluera les résultats obtenus par l'utilisation de ce modèle pour copier des réseaux Small-World, Scale-Free, aléatoire, et un réseau réel, Jazz.

5.1 Définition du modèle II

Modèle II	
├	Définition du modèle II
└	Résultats du modèle II

Deux modifications vont être apportées pour contraindre les propagations possibles.

En premier lieu, **nous allons travailler avec des profils de comportement.**

L'unité complexe-action est désormais un couple composé :

- d'un complexe-action possédant une action d'ajout de lien et
- d'un complexe-action possédant une action de retrait de lien.

Le but est d'avoir des comportements permettant à certains nœuds d'atteindre des connectivités particulières par rapport au reste du réseau.

Par ailleurs, nous allons **limiter les possibilités de transmission** des dit couples, en prenant en compte la proximité en terme de degré entres nœuds concernés pour une éventuelle propagation.

Nous espérons ainsi voir apparaitre des ensembles de nœuds plus connectés entre eux, réunis à la fois par similarité de degré et de complexe-actions possédés.

5.1.1 Les complexe-actions

Afin de définir des profils de comportements, l'unité insécable concernant les actions est désormais un couple de complexe-actions. L'un d'entre eux est associé à une action d'ajout de lien, l'autre de retrait ; les deux possédant leurs propres filtres.

Attribut On conserve l'hypothèse qu'il est possible de reproduire des familles similaires de réseaux en n'utilisant comme attribut que le degré.

Actions associées La diversité des complexe-actions du premier modèle est conservée, mais ces derniers sont désormais associés en couple. Nous disposions précédemment de 10 complexe-actions d'ajout et autant de retrait. Les combinaisons possibles sont donc au nombre de 100. Nous rajoutons l'action vide ("void"), ne faisant rien, permettant l'apparition de couple de complexe-actions n'ayant en réalité qu'une seule action possible. Ainsi, le couple composé de :

- [**AddEq**] Création d'un lien vers un nœud de degré égal et
- [**RmvVoid**] aucune action de retrait

ne pourra en étant appelé qu'appliquer effectivement [**AddEq**].

Le nombre de couples disponibles est donc de 121.

5.1.2 Propagation

La propagation concerne désormais le couple dans son ensemble ; l'application d'une action d'ajout ou de retrait entre deux entités peut permettre la propagation de l'ensemble du couple.

Nom	action & [filtres]	explications
AddVoid	Aucune action, aucun filtre	Ne fait aucune action
RmvVoid	Aucune action, aucun filtre	Ne fait aucune action

Table 5.1: En plus des complexe-actions définis pour le modèle I, deux autres ne faisant rien sont ajoutés. Ils permettent l'apparition de couples composés d'un seul élément actif.

diff. de degré	0	1	2	3
diviseur	$\frac{P_{CCA}}{1}$	$\frac{P_{CCA}}{4}$	$\frac{P_{CCA}}{9}$	$\frac{P_{CCA}}{16}$

Table 5.2: Résumé de la réduction de la probabilité P_{CCA} portée par un couple de complexe-actions par rapport à la valeur absolue de la différence de degré entre les deux entités concernées par une éventuelle transmission.

Afin de limiter les entités pouvant se faire transmettre un couple de complexe-actions, **la probabilité de propagation prend en compte la distance en terme de degré entre les nœuds concernés**. Dans le même temps, les couples de complexe-actions portent encore une probabilité de propagation. Soit une entité E_1 au degré $d_1 = 10$, possédant un couple de complexe-action C_{CA1} avec la probabilité de transmission $P_{CCA1} = 0.6$. Soit E_2 l'entité sur laquelle est appliqué C_{CA1} , au degré $d_2 = 9$. La probabilité de transmission p du couple CCA depuis E_1 vers E_2 est définie par :

$$P = \frac{P_{CCA1}}{(1 + |d_1 - d_2|)^2}$$

dans le cas présent,

$$P = \frac{0.6}{2^2} = 0.15$$

Le tableau 5.2 donne un aperçu des probabilités de transmission pour les premières valeurs de différences de degré. Il est important de diminuer rapidement la probabilité de transmission de complexe-actions car les réseaux sociaux sont composés de nœuds aux degrés très faibles par rapport au degré maximum possible.

Par ailleurs, le nombre de configurations initiales augmente significativement depuis le modèle I. Nous allons donc limiter la valeur que peut prendre une probabilité de propagation, allant de 0 à 1 par palier de 0.1.

5.1.3 Scoring

Pas de changement dans le mécanisme ainsi que dans les propriétés présent en compte pour l'établissement d'un score pour un réseau obtenu vis-à-vis d'un réseau à reproduire. Les propriétés à disposition sont les suivantes

- Densité
- Distance inter quartile
- Clustering moyen

- APL
- Similarité de la forme de la distribution de degré

Une même configuration est relancée plusieurs fois, afin de calculer une moyenne et un écart type sur son score. Le score final associé à chaque configuration est le produit de ces dernières valeurs. Cela permet de prendre en compte la répétabilité des résultats.

5.2 Résultat du modèle II

Modèle II	
└ Définition du modèle II	
└ Résultats du modèle II	

5.2.1 Configuration et réseaux à reproduire

Nous gardons les mêmes objectifs que pour le premier modèle. On veut particulièrement être en mesure de trouver des réseaux possédant :

- un clustering important associé à une densité faible
- une taille de chemin entre nœuds courte
- une distribution de degrés en loi de puissance

Les réseaux étudiés restent donc :

- Random - Nombre de nœuds [300] - Densité [0.02]
- Small-World - Nombre de nœuds [300] - Probabilité [0.01]
- Preferential Attachment - Nombre de nœuds [300] - nombre de liens par nœud [2]

Il s'agit du même ensemble de réseaux que pour le modèle I. Nous y rajoutons aussi un réseau réel, le réseau de musiciens de Jazz ([Gleiser and Danon, 2003](#)), choisi pour son nombre de nœuds (200) qui est de l'ordre de grandeur avec lequel nous travaillons dans les modèles.

Lors des résultats précédents, nous avons remarqué qu'un petit nombre de complexe-actions prennent le pas sur d'autres qui ne sont ni présents ni appliqués. Nous réduisons donc le nombre de complexe-actions présents dans la simulation, et le fixons à 4 couples.

L'ordre de taille de l'espace des paramètres est donc :

Combinaisons possibles de sélection de couple de complexe-actions (I) * combinaisons des probabilités de de propagation (II)

$$C_{121}^4(I) * 10^4(II) \sim 10^{13}$$

Outils utilisés Les outils restent les mêmes :

- l'algorithme génétique est fourni par le Framework OpenMOLE ([Reuillon et al., 2013](#)),
- les simulations sont distribuées sur les clusters OSIRIM¹,
- la visualisation des graphiques est réalisée par JFreeChart²,
- celle des graphes grâce à Graphstream³ et
- le calcul d'APL est réalisé grâce à l'un des algorithmes implémentés par cette même librairie.

1. <https://osirim.irit.fr/site/en>
 2. <http://www.jfree.org/jfreechart/>
 3. graphstream-project.org

5.2.2 Résultats des simulations

Réseaux aléatoires Nous allons nous intéresser aux réseaux à densité de 30% afin de voir si les nouvelles configurations trouvées pallient au problème de la dégénérescence des résultats lorsque la configuration est laissée plus longtemps. Les meilleures configurations trouvées sont données dans le tableau 5.3. Dans les trois cas, les configurations trouvées ne proposent que des réseaux passant par la densité souhaitée, augmentant lorsque l'on augmente le nombre d'actions avant arrêt de la simulation. **il n'y a donc pas d'amélioration par rapport au premier modèle pour les réseaux aléatoires.**

Config.	CA constituant et probabilité de transmission	score $[\sigma]$
1	[Add- :0.766] [Rmv ∞ ! :0.762] [Rmv+ :0.668] [Add \emptyset :0.5768] [Add $^\circ$:0.0445] [Rmv $^\circ$:0.368] [Add \emptyset -Hop :0.248] [Add ∞ :0.0] [Rmv ∞ :0.0]	11.4 [5.65]
2	[Add ∞ :1.0] [Rmv+ :1.0] [Add $^\circ$:1.0] [RmvChain :0.74] [Add+ :0.65] [Rmv \emptyset :0.575] [Rmv ∞ :0.49] [Rmv $^\circ$:0.47] [Rmv $^\circ$! :0.027]	9.7 [1.0]

Table 5.3: *Random 30% : configurations trouvées pour atteindre un réseau aléatoire avec une densité de 0.3. Score moyen associé et écart type sur 20 simulations par configuration. L'algorithme génétique d'OpenMOLE (Reuillon et al., 2013) réalise 150 générations, le critère d'arrêt étant 10^6 actions entreprises (avec succès ou non).*

Scale-Free Nous allons chercher à reproduire un réseau généré par l'algorithme d'attachement préférentiel, avec comme paramètre $m=2$.

Deux configurations sont trouvées (tableau 5.4 et dans le détail 5.4). Ces configurations ne proposent que des distributions de degrés de réseau aléatoire, en gaussienne.

Config.	CA constituant et probabilité de transmission	score $[\sigma]$
1	[Add-&RmvChain :0.8] [Add°&Rmv∞! :0.5]	8.9 [0.1]
2	[Add°&Rmv°! :0.4]	6.9 [2.9]

Table 5.4: *SF* : configurations trouvées pour atteindre un réseau sans échelle. L'algorithme génétique d'OpenMOLE réalise 150 générations (Reuillon et al., 2013), le nombre d'actions limite est d'un million.

Propriété	cible	config #1		config #2	
		valeur	score	valeur	score
interq.	2	1.333[0.471]	0.223[0.158]	1.333[0.471]	0.223[0.158]
clust. moy.	0.058	0.015[0.005]	4.264[0.523]	0.009[0.002]	4.929[0.242]
densite	0.013	0.019[0.003]	0.548[0.273]	0.009[0.001]	0.392[0.095]
APL	3.53	3.568[0.289]	0.239[0.165]	5.948[0.655]	2.408[0.653]
Distrib deg.	-	-	39.333[1.089]	-	27.0[13.425]

Table 5.5: *SF* : valeurs et scores des propriétés présent en compte pour le score de distance entre réseau cible et réseaux générés par les 2 meilleures configurations trouvées. Écart-type de ces valeurs sur 20 itérations de simulation par configuration. Les valeurs de distribution de degré n'y sont pas représentées.

Small World Sur les trois meilleures configurations trouvées, deux d'entre elles n'offrent pas les résultats attendus en terme de coefficient de clustering. La dernière offre un résultat intéressant (tableau 5.6). Dans le détail, nous arrivons enfin à **une configuration possédant un rapport entre densité et coefficient de clustering proche du réseau cible** (tableau 5.7), malgré le fait que la densité soit supérieure à celle qui est recherchée. Le réseau obtenu est donné en figure 5.1. On va s'intéresser au détail des liens entre les nœuds afin de vérifier si les modifications amenées dans le modèle II ont un impact sur la distribution des complexe-actions.

La figure 5.3 est un graphe qui représente la répartition de liens entre les nœuds regroupés par couple possédé. Le nœud [*Add° !.Rmv° !*] de la figure représente par exemple l'ensemble des entités de la simulation possédant ce couple d'actions. Un lien entre les nœuds de la figure nous donne la proportion de liens depuis un groupe de nœuds vers un autre. On constate que des clusters de nœuds se sont créés, deux des ensembles étant autant connectés entre eux qu'à l'extérieur. Les nœuds possédant le couple [*Add∞ !.Rmv Void*] servent de *hub* au reste du réseau, notamment parce que leurs actions de retrait de lien est vide. On retrouve ces profils dans les degrés élevés de la distribution en question (figure 5.2).

Une fois passée l'étape de propagation des couples à l'ensemble des entités du réseau, nous n'observons plus qu'une transmission toutes les 10.000 actions réalisées par les entités. Cela signe **une stabilisation du nombre d'échanges, confirmant que les cluster d'entités regroupé par degré sont ancrés**. A titre de comparaison, une configuration stable en terme de variation de densité du modèle I procède à 3000 propagations de complexe-actions en moyenne pour le même nombre d'actions entreprises.

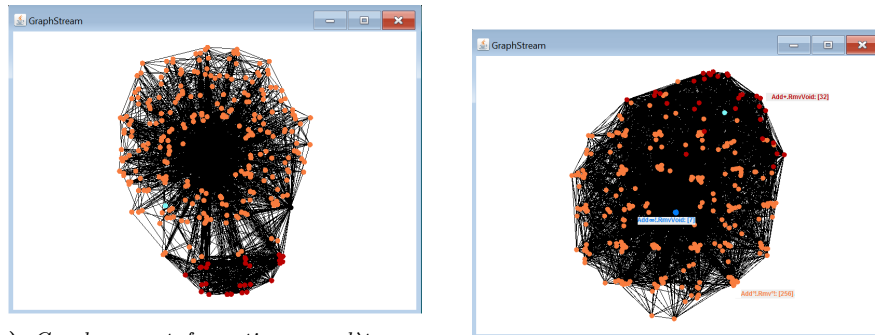
Cette configuration finit tout de même par dégénérer vers une moins bonne topologie une fois le nombre d'actions limite dépassé (figure 5.4). Cela met encore plus en avant les groupes de nœuds se réunissant par couple de complexe-actions portés, en trois endroits de la distribution de degré.

Config.	CA constituant et probabilité de transmission	score $[\sigma]$
1	[Add \emptyset -3Hop.Rmv ∞ :0.5] [Add $^\circ$!.Rmv $^\circ$! :0.5] [Add ∞ !.RmvVoid :0.2] [Add+.RmvVoid :0.6]	24 [1.3]

Table 5.6: *SW* : configuration trouvée pour atteindre un Small-World. L’algorithme génétique d’OpenMOLE réalise 150 générations (Reuillon et al., 2013), le nombre d’actions limite est d’un million.

Propriété	cible	config #1	
		valeur	score
interq.	0	11.0[8.556]	3.679[2.861]
clust. moy.	0.65	0.642 [0.118]	8.791[7.912]
densite	0.047	0.123 [0.054]	7.635[5.39]
APL	3.98	1.877[0.054]	2.094[0.054]
Distrib deg.	-	-	97.867[1.343]

Table 5.7: *SW* : valeurs et scores des propriétés prises en compte pour le score de distance entre réseau cible et réseaux générés par les 2 meilleures configurations trouvées. Écart-type de ces valeurs sur 20 itérations de simulation par configuration. Les valeurs de distribution de degré n’y sont pas représentées.



(a) Graphe avant formation complète, pour une meilleure visualisation des groupes de complexe-actions.

(b) graphe final obtenu.

Figure 5.1: *SW* : réseau obtenu obtenant un bon rapport entre densité et coefficient de clustering. 3 groupes de nœuds portant des complexe-actions apparaissent. Rouge : [Add+.RmvVoid], bleu foncé : [Add \emptyset -3Hop.Rmv ∞], orange : [Add ∞ !.RmvVoid] et bleu clair : [Add ∞ !.RmvVoid].

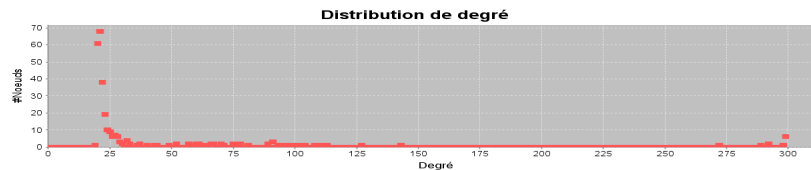


Figure 5.2: *SW* : distribution de degré associé au réseau. Les "hubs" expliquent la présence de degrés très élevés.

Couple	Nb noeud	Degré Moyen	degré moyen des noeuds connectés
[Add ∞ !.RmvVoid]	2	28735	41.74
[Add \emptyset -3Hop.Rmv ∞]	12	263	41.16
[Add $^\circ$!.Rmv $^\circ$!]	248	34.2	157
[Add+.RmvVoid]	38	80.7	116.7

Table 5.8: SW : pour chaque couple présent dans la simulation, le nombre de nœuds qui les portent, le degré moyen de ces nœuds, et le degré des nœuds auxquels sont connectés ces premiers nœuds.

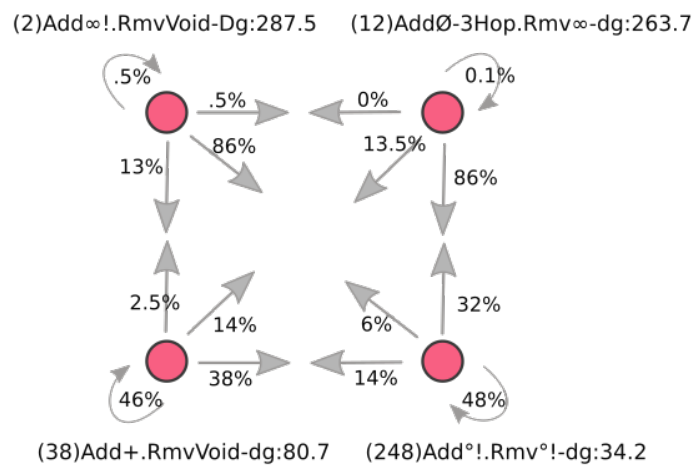


Figure 5.3: SW : quantité de liens reliant des noeuds regroupés par couple de complexe-actions possédé. (nombre de noeuds)[Complexe-actions]-Dg : (degré moyen des noeuds)



Figure 5.4: SW : état du réseau lorsque le nombre d'actions limite est dépassé. Trois groupes de noeuds se forment de manière significative, chacun avec des degrés moyens distincts.

Réseau Jazz Maintenant que nous avons constaté que le modèle II est capable d’offrir un coefficient de clustering satisfaisant pour les réseaux sociaux, nous allons chercher à reproduire un réseau réel. Notre choix s’est porté sur un réseau de musiciens de jazz (Gleiser and Danon, 2003).

Sur les 4 meilleures configurations renvoyées par l’algorithme génétique, **une est particulièrement intéressante** (tableau 5.9). Dans le détail des paramètres, la densité et le coefficient de clustering ont des valeurs du bon ordre de grandeur (tableau 5.10). Une fois de plus, le modèle use d’une manière particulière pour obtenir un coefficient de clustering élevé (figure 5.5) : certains nœuds [**Add.RmvEq :0.6**] servent de hubs et sont connectés à tous les autres nœuds beaucoup plus nombreux [**AddEq.Rmv- :0.2**], comme on le voit dans la figure 5.5. Ainsi, ces derniers nœuds, en étant simplement connectés entre eux, augmentent radicalement le coefficient de clustering. La conséquence sur la comparaison des distributions de degré est la présence pour le réseau généré de quelques nœuds à très haut degré, nœuds absents du réseau réel (figure 5.6).

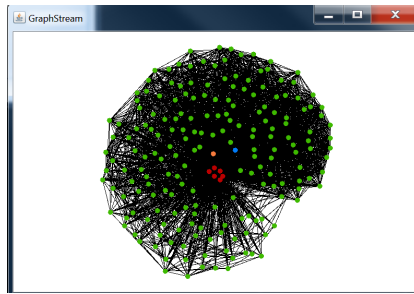
Cette configuration garde les mêmes propriétés sur des temps de simulation supérieurs.

Config.	CA constituant et probabilité de transmission	score[σ]
1	[AddEq.Rmv- :0.2] [Add ∞ .RmvEq :0.5] [AddVoid.Rmv ∞ :0.1] [Add-.RmvEq :0.6]	11.7 [0.9]

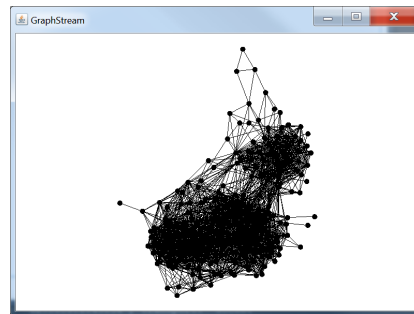
Table 5.9: Jazz : Configuration trouvée pour reproduire le réseau social de musicien de Jazz. L’algorithme génétique d’OpenMOLE (Reuillon et al., 2013) réalise 200 générations, le nombre d’actions limite est d’un million.

Propriété	cible	config #1	
		valeur	score
interq.	23	21.8[1.47]	0.711[0.65]
clust. moy.	0.61	0.429 [0.073]	18.82[7.255]
densite	0.14	0.162 [0.018]	2.143[1.765]
APL	2.2	1.84[0.019]	0.594[0.028]
Distrib deg.	-	-	36.162[2.547]

Table 5.10: Jazz : Détail des propriétés topologiques du réseau Jazz et du réseau synthétique trouvé.



(a) Réseau obtenu avec le modèle II.



(b) Réseau réel à reproduire.

Figure 5.5: Jazz : Réseau cible et réseau obtenu avec le modèle.

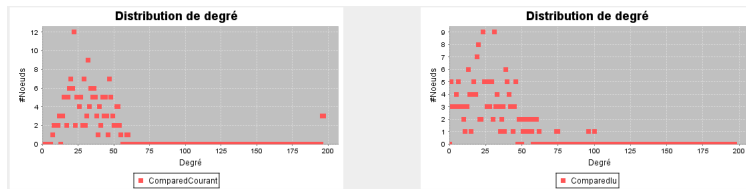


Figure 5.6: Jazz : distribution de degré associé aux réseaux. Les "hubs" expliquent la présence de degrés très élevés pour le réseau synthétique.

5.2.3 Discussion sur les résultats

Nous avons travaillé avec deux instanciations distinctes du Framework. Le premier s'est révélé trop libre. Nous avons donc cherché à contraindre plus fortement les possibilités d'interactions entre les entités, afin de faire émerger des topologies particulières, éloignées des réseaux aléatoires. Vis-à-vis des propriétés topologiques avec lesquelles nous travaillons, le marqueur le plus important est le rapport entre la densité du réseau et son coefficient de clustering. **Le second modèle s'est révélé être en mesure de générer des réseaux possédant une bonne valeur pour ce rapport.** Cela a pu être possible par une définition plus stricte des complexe-actions qui sont portées par les entités, ainsi que par une restriction des conditions de propagation y étant associées.

Il est intéressant de constater la manière trouvée pour y parvenir : peu de nœuds sont connectés à tous les autres nœuds de degré inférieur du réseau. Ainsi, quelques liens entre ces nœuds de degré inférieur font monter l'indice de clustering sans trop "coûter" en terme de densité. Hélas, la présence de ces hubs porte préjudice aux réseaux obtenus si l'on s'intéresse aux phénomènes de propagation ; ces derniers s'opéreraient trop rapidement.

5.2.4 Retour sur les objectifs

Nous allons parcourir les objectifs que nous avons défini, et la façon dont le modèle II y répond en particulier.

A-Répétition des résultats possibles & F-Indépendance aux premières actions prises *A- Possibilité de générer, pour plusieurs simulations de configurations équivalentes, des réseaux synthétiques très similaires en terme de propriétés mesurées.*

F- Que le résultat final ne soit pas dépendant de l'ordre des actions.

Une paire de complexe-actions de chaque type(ajout et retrait); pas d'oubli de la part des entités fournissant une paire au réseau.

B-Espace des réseaux atteignables large *Avoir la possibilité de générer un grand nombre de réseaux avec le Framework, afin d'assurer un large panel de réseaux reproductibles.*

Mécanisme de propagation favorisant la naissance de "clusters" d'entités possédant le même couple de complexe-actions.

C-Très peu de configuration *Garantir une utilisation simple sans connaissance experte mathématique, informatique, ou du domaine métier dont est issu le réseau échantillon.*

Automatisation de la recherche des meilleurs paramètres initiaux par un algorithme génétique.

D-Espace de paramètre réduit *Doit permettre une recherche automatisée efficace de résultats.*

Limitation des probabilités de transmission possibles.

E-Continuité *Monotonie de l'évolution des formes de réseaux obtenus en fonction de légères modifications dans les paramètres.*

Non adressé par le modèle.

G-Etude de la résilience *Que le modèle donne la possibilité d'étudier l'évolution du réseau après perturbation.*

Non adressé par le modèle.

Chapitre 6

Conclusion et perspectives

6.1 Conclusion

6.1.1 Résumé de la contribution

L'approche utilisée pour la reproduction de réseau est originale : il s'agit d'une approche dynamique orientée processus locaux, avec une validation topologique des résultats macroscopiques obtenus. Cela permet de rendre intelligible le processus de construction, tout en assurant une validation quantitative du résultat. L'utilisation d'un algorithme génétique permet de chercher les meilleures configurations initiales afin de permettre, depuis des comportements d'ordre microscopiques, d'obtenir une émergence macroscopique d'un réseau aux propriétés topologiques désirées.

Les processus réels ayant amené la construction du réseau cible sont remplacés par les processus stylisés du cadre du Framework. Cela permet une forte généralité : la méthode peut être appliquée à n'importe quel réseau. Il suffit à l'utilisateur de choisir les propriétés topologiques à prendre en compte pour l'évaluation d'un réseau généré, dépendant d'un domaine d'application particulier.

L'aspect dynamique de la construction du réseau offre la possibilité d'ajouter ou de retirer des entités une fois l'équilibre atteint pour en étudier les conséquences. La plus grande difficulté rencontrée concerne la diversité des réseaux atteignables. Nous avons établi dans les objectifs qu'il fallait un système à la fois contraint et ouvert, pour respectivement pouvoir répéter des résultats et atteindre une grande variété de réseaux. Il a été nécessaire d'augmenter les contraintes appliquées sur le système, afin de permettre de générer des réseaux aux topologies plus hétérogènes.

Ainsi, le deuxième modèle vise à contraindre plus fortement le mécanisme de propagation, afin de favoriser l'apparition d'ensembles de nœuds plus fortement connectés entre eux, portant les mêmes complexe-actions.

6.1.2 Conclusion sur les objectifs

Nous allons repasser sur l'ensemble des objectifs et discuter leurs éventuelles résolutions.

A- Reproductibilité des résultats & F-Indépendance aux premières actions prises

A- Possibilité de générer, pour plusieurs simulations des configurations équivalentes, des réseaux synthétiques très similaires en terme de propriétés mesurées.

F- Que le résultat final ne soit pas dépendant de l'ordre des actions.

Ces deux objectifs ont été relativement bien résolus. En fonction des configurations, l'écart-type sur les propriétés topologiques après plusieurs itérations est faible.

B-Espace des réseaux atteignables large

Avoir la possibilité de générer un grand nombre de réseaux avec le Framework, afin d'assurer un large panel de réseaux reproductibles.

Ce point a été la principale difficulté du travail. Les résultats, particulièrement pour le couple de propriétés propres aux réseaux Small-World, n'ont pas été complètement probants avec le premier modèle. Le modèle II y a mieux répondu, mais est passé par une topologie de réseaux différente, ce qui peut porter préjudice en cas de processus de simulation joués comme par exemple un processus de diffusion.

Cependant, la modularité du Framework laisse penser qu'il est possible d'inventer d'autres complexe-actions plus adéquats pour ce type de réseaux.

C-Très peu de configuration

Garantir une utilisation simple sans connaissance experte mathématique, informatique, ou du domaine métier dont est issu le réseau cible.

Globalement, le Framework peut être utilisé sans aucune configuration lorsqu'il est utilisé de concert avec un algorithme génétique. Il est néanmoins possible de modifier facilement les propriétés prises en compte pour le calcul de distance entre le réseau synthétique et le réseau objectif. De même, l'implémentation du Framework permet d'ajouter facilement de nouveaux complexe-actions aux modèles.

D-Espace des paramètres réduit

Doit permettre une recherche automatisée efficace de résultats.

L'espace des paramètres est relativement important mais il est possible de considérer les probabilités de propagation comme une variable discrète.

E-Continuité

Monotonie de l'évolution des formes de réseaux obtenus en fonction de légères modifications dans les paramètres.

Aucun effet de seuil n'a été remarqué au cours de l'exploration manuelle des modèles.

G-Etude de la résilience

Que le modèle donne la possibilité d'étudier l'évolution du réseau après perturbation.

Cette aspect n'a pas été étudié en profondeur, mais a été un élément fondamental dans la l'élaboration de la solution.

6.1.3 Conclusion technique

L'approche technique a suivi la méthodologie de séparation conceptuelle entre Framework et modèles. Ainsi, L'implémentation a été réalisée en langage objet (JAVA), avec un soin porté aux concepts de généricité et flexibilité. Cela permet l'instanciation rapide du Framework conceptuel en différents modèles concrets. L'ensemble du code réalisé est disponible sur les serveurs Github¹.

6.2 Perspectives et ouvertures

6.2.1 Perspectives

Plusieurs améliorations sont envisageables pour perfectionner le travail réalisé. En ce qui concerne l'existant, une étude des comportements en cas de changement d'échelle serait intéressante. Quelques étapes évidentes pourraient être de :

- Étudier le comportement de la génération de réseau pour une configuration trouvée, en changeant le nombre de nœuds initial et
- Comparer les configurations trouvées par algorithme génétique pour deux réseaux similaires en terme de propriétés, mais en nombre de nœuds différent.

Travailler avec un nombre de nœuds significativement plus grand pourrait demander une optimisation du code. La contre-partie de la généricité et de la flexibilité sont des syntaxes et constructions plus lourdes en mémoire et processeur. Il serait possible de faire d'un modèle une version "définitive", plus optimisée.

Dans l'optique d'une utilisation réelle par un public varié, l'interface utilisateur devrait être améliorée, et un algorithme génétique serait à développer et à intégrer pour que l'ensemble des étapes soient réalisables depuis l'application.

Enfin, il serait intéressant qu'un utilisateur puisse non seulement donner en entrée le réseau à reproduire, mais aussi un ensemble d'agents et leurs propriétés. Il serait alors possible d'utiliser la référence de ces agents pour créer le réseau, ce qui résoudrait la question de la correspondance entre nœuds du

1. <https://github.com/Rakhx/Giteri>

réseau et entités. Cela faciliterait l'utilisation du réseau généré dans un autre environnement de simulation.

6.2.2 Ouvertures

Une des évolutions serait de s'atteler à reproduire non plus des réseaux mais des dynamiques de création de réseau, avec la prise en compte des ajouts et retraits de nœuds. Cela nécessiterait :

- l'addition d'actions d'ajouts et de retraits de nœuds dans le réseau
- prendre en entrée plusieurs photographies du réseau cible à des moments distincts
- une manière de combiner les résultats obtenus sur chaque étape : optimisation multi-objectifs ou dans un premier temps l'addition des scores sur les différents réseaux synthétiques.

Il pourrait aussi être intéressant d'étoffer plus amplement le calcul de score entre deux réseaux, en y ajoutant le résultat de processus joués, par exemple un phénomène de propagation.

À terme, la construction générique du Framework permettrait aussi de prendre en compte d'autres attributs que le degré des entités, tel qu'une position géographique ou un âge. On pourrait par exemple représenter les phénomènes d'homophilie observés dans les réseaux réels.

Bibliographie

- Eric Abrahamson and Lori Rosenkopf. Social Network Effects on the Extent of Innovation Diffusion : A Computer Simulation. *Organization Science*, 8(3) : 289–309, June 1997. ISSN 1047-7039, 1526-5455. doi : 10.1287/orsc.8.3.289. URL <http://pubsonline.informs.org/doi/abs/10.1287/orsc.8.3.289>.
- Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1) :47, 2002. URL <http://journals.aps.org/rmp/abstract/10.1103/RevModPhys.74.47>.
- Frédéric Amblard, Audren Bouadjio-Boulic, Carlos Sureda Gutierrez, and Benoit Gaudou. Which models are used in social simulation to generate social networks? : a review of 17 years of publications in JASSS. In *Proceedings of the 2015 Winter Simulation Conference*, pages 4021–4032. IEEE Press, 2015. URL <http://dl.acm.org/citation.cfm?id=2889153>.
- W Ross Ashby. Analysis of the system to be modeled. *The process of model building in the behavioral sciences*, pages 94–114, 1970. bibtex : ashby1970analysis.
- Robert Axtell, Robert Axelrod, Joshua M. Epstein, and Michael D. Cohen. Aligning simulation models : A case study and results. *Computational and Mathematical Organization Theory*, 1(2) :123–141, February 1996. ISSN 1381-298X, 1572-9346. doi : 10.1007/BF01299065. URL <http://link.springer.com/10.1007/BF01299065>.
- Moulay Aziz-Alaoui and Cyrille Bertelle. *From system complexity to emergent properties*. Springer Science & Business Media, 2009.
- Albert Barabasi and Réka Albert-László. Emergence of Scaling in Random Networks. *Science*, 286(5439) :509–512, October 1999. ISSN 00368075, 10959203. doi : 10.1126/science.286.5439.509. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.286.5439.509>.
- Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, United Kingdom, 2016. ISBN 978-1-107-07626-6. OCLC : ocn910772793.

- Michel Baranger. Chaos, complexity, and entropy. *New England Complex Systems Institute, Cambridge*, 2000.
- Olivier Barreteau, François Bousquet, and Jean-Marie Attonaty. Role-Playing Games for Opening the Black Box of Multi-Agent Systems : Method and Lessons of Its Application to Senegal River Valley Irrigated Systems. *Journal of Artificial Societies and Social Simulation*, 4(2) :1–5, 2001. URL <https://ideas.repec.org/a/jas/jasssj/2000-18-1.html>.
- Christopher L. Barrett, Richard J. Beckman, Maleq Khan, V. S. Anil Kumar, Madhav V. Marathe, Paula E. Stretz, Tridib Dutta, and Bryan Lewis. Generation and analysis of large synthetic social contact networks. pages 1003–1014. IEEE, December 2009. ISBN 978-1-4244-5770-0. doi : 10.1109/WSC.2009.5429425. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5429425>.
- Yossi Ben-Zion, Yahel Cohen, and Nadav M. Shnerb. Modeling epidemics dynamics on heterogenous networks. *Journal of Theoretical Biology*, 264(2) : 197–204, May 2010. ISSN 00225193. doi : 10.1016/j.jtbi.2010.01.029. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022519310000494>.
- Carlo Bianchi, Pasquale Cirillo, Mauro Gallegati, and Pietro A. Vagliasindi. Validating and Calibrating Agent-Based Models : A Case Study. *Computational Economics*, 30(3) :245–264, September 2007. ISSN 0927-7099, 1572-9974. doi : 10.1007/s10614-007-9097-z. URL <http://link.springer.com/10.1007/s10614-007-9097-z>.
- F.C. Billari, A. Prskawetz, B.A. Diaz, and T. Fent. The "wedding-ring" : an agent-based marriage model based on social interaction. *Demographic Research*, 17(3), 2007. URL <https://www.demographic-research.org/Volumes/Vol17/3/17-3.pdf>.
- Jeffrey Boase, John Horrigan, Barry Wellman, and Lee Rainie. The strength of internet ties. *Pew Research Center's Internet & American Life Project*, 2006.
- S Boccaletti, V Latora, Y Moreno, M Chavez, and D Hwang. Complex networks : Structure and dynamics. *Physics Reports*, 424(4-5) :175–308, February 2006. ISSN 03701573. doi : 10.1016/j.physrep.2005.10.009. URL <http://linkinghub.elsevier.com/retrieve/pii/S037015730500462X>.
- Jeremy Boissevain. JEREMY BOISSEVAIN. Friends of Friends : Networks, Manipulators and Coalitions. Pp. 285. Oxford, England : Basil Blackwell, 1974. No price. *The ANNALS of the American Academy of Political and Social Science*, 417(1) :193–194, January 1975. ISSN 0002-7162, 1552-3349. doi : 10.1177/000271627541700164. URL <http://journals.sagepub.com/doi/10.1177/000271627541700164>.
- Eric Bonabeau. Agent-based modeling : Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl

- 3) :7280–7287, 2002. URL http://www.pnas.org/content/99/suppl_3/7280.short.
- Anna D. Broido and Aaron Clauset. Scale-free networks are rare. *Nature Communications*, 10(1), December 2019. ISSN 2041-1723. doi : 10.1038/s41467-019-08746-5. URL <http://www.nature.com/articles/s41467-019-08746-5>.
- Benoît Calvez. Le calibrage de modèles à base d’agents pour la simulation de systèmes complexes. page 173, 2009.
- Kevin Chapuis. *Modelisation multi-agents de la satisfaction au travail.pdf*. PhD thesis, January 2016.
- Jean-Philippe Cointet and Camille Roth. How Realistic Should Knowledge Diffusion Models Be? *Journal of Artificial Societies and Social Simulation*, 10(3) :5, 2007. ISSN 1460-7425. URL <http://jasss.soc.surrey.ac.uk/10/3/5.html>. bibtex : cointet2007.
- P Collet. Quelques notions et résultats sur les systèmes dynamiques, 2015.
- Alexis Drogoul. GenStar Project, 2013. URL <https://anr.fr/Project-ANR-13-MONU-0006>.
- R. I. M. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6) :469 – 493, 1992. ISSN 0047-2484. doi : [https://doi.org/10.1016/0047-2484\(92\)90081-J](https://doi.org/10.1016/0047-2484(92)90081-J). URL <http://www.sciencedirect.com/science/article/pii/004724849290081J>.
- David Easley and John Kleinberg. Networks, Crowds, and Markets : Reasoning about a Highly Connected World. 2010. URL <http://www.jstor.org/stable/1885620>.
- Bruce Edmonds. How are physical and social spaces related? - cognitive agents as the necessary ”glue”, 2003. URL <http://bruce.edmonds.name/hdssapsr/>.
- Bruce Edmonds and Scott Moss. From KISS to KIDS – An ‘Anti-simplistic’ Modelling Approach. In Paul Davidsson, Brian Logan, and Keiki Takadama, editors, *Multi-Agent and Multi-Agent-Based Simulation*, pages 130–144. Springer Berlin Heidelberg, 2005. ISBN 978-3-540-32243-6.
- Ernesto Estrada, Naomichi Hatano, and Michele Benzi. The physics of communicability in complex networks. *Physics Reports*, 514(3) :89–119, May 2012. ISSN 03701573. doi : 10.1016/j.physrep.2012.01.006. URL <http://linkinghub.elsevier.com/retrieve/pii/S0370157312000154>.
- I. Farkas, H. Jeong, T. Vicsek, A.-L. Barabási, and Z.N. Oltvai. The topology of the transcription regulatory network in the yeast, *Saccharomyces cerevisiae*. *Physica A : Statistical Mechanics and its Applications*, 318(3-4) :601–612,

- February 2003. ISSN 03784371. doi : 10.1016/S0378-4371(02)01731-4. URL <http://linkinghub.elsevier.com/retrieve/pii/S0378437102017314>.
- Ove Frank and David Strauss. Markov Graphs. *Journal of the American Statistical Association*, 81(395) :832, September 1986. ISSN 01621459. doi : 10.2307/2289017. URL <http://www.jstor.org/stable/2289017?origin=crossref>.
- Agata Fronczak, Piotr Fronczak, and Janusz A. Hołyst. Average path length in random networks. *Physical Review E*, 70(5), November 2004. ISSN 1539-3755, 1550-2376. doi : 10.1103/PhysRevE.70.056110. URL <https://link.aps.org/doi/10.1103/PhysRevE.70.056110>.
- Serge Frontier. *Écosystèmes : structure, fonctionnement, évolution*. Dunod, Paris, 2014. ISBN 978-2-10-051935-4. OCLC : 1026143356.
- Joseph Galaskiewicz. *Social organization of an urban grants economy : a study of business philanthropy and nonprofit organizations*. Academic Press, Orlando, 1985. ISBN 978-0-12-273860-9 978-0-12-273861-6.
- Benoit Gaudou, Christophe Sibertin-Blanc, Olivier Therond, Frédéric Amblard, Yves Auda, Jean-Paul Arcangeli, Maud Balestrat, Marie-Hélène Charron-Moirez, Etienne Gondet, Yi Hong, Romain Lardy, Thomas Louail, Eunat Mayor, David Panzoli, Sabine Sauvage, José-Miguel Sánchez-Pérez, Patrick Taillandier, Nguyen Van Bai, Maroussia Vavasseur, and Pierre Mazzega. The MAELIA Multi-Agent Platform for Integrated Analysis of Interactions Between Agricultural Land-Use and Low-Water Management Strategies. In Shah Jamal Alam and H. Van Dyke Parunak, editors, *Multi-Agent-Based Simulation XIV*, volume 8235, pages 85–100. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. ISBN 978-3-642-54782-9 978-3-642-54783-6. doi : 10.1007/978-3-642-54783-6_6. URL http://link.springer.com/10.1007/978-3-642-54783-6_6.
- Armando Geller and Scott Moss. Growing qawm : An evidence-driven declarative model of Afghan power structures. *Advances in Complex Systems*, 11(02) :321–335, 2008. URL <http://www.worldscientific.com/doi/abs/10.1142/S0219525908001659>.
- Dimitria Giorgas. Community formation and social capital in Australia. In *7th Australian Institute of Family Studies Conference, Sydney*, volume 25. Citeseer, 2000.
- Pablo M. Gleiser and Leon Danon. COMMUNITY STRUCTURE IN JAZZ. *Advances in Complex Systems*, 06(04) :565–573, December 2003. ISSN 0219-5259, 1793-6802. doi : 10.1142/S0219525903001067. URL <https://www.worldscientific.com/doi/abs/10.1142/S0219525903001067>.

- Andrzej Grabowski. Opinion formation in a social network : The role of human activity. *Physica A : Statistical Mechanics and its Applications*, 388(6) :961–966, March 2009. ISSN 03784371. doi : 10.1016/j.physa.2008.11.036. URL <http://linkinghub.elsevier.com/retrieve/pii/S0378437108009746>.
- Paul Guyot and Alexis Drogoul. Multi-agent Based Participatory Simulations on Various Scales. In Toru Ishida, Les Gasser, and Hideyuki Nakashima, editors, *Massively Multi-Agent Systems I*, pages 149–160, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg. ISBN 978-3-540-31889-7.
- Lynne Hamill. *Communications, Travel and Social Networks since 1840 - A Study Using Agent-based Models*. PhD thesis, University of Surrey, 2010.
- Lynne Hamill and Nigel Gilbert. Social circles : A simple structure for agent-based social network models. *Journal of Artificial Societies and Social Simulation*, 12(2) :3, 2009. URL <http://jasss.soc.surrey.ac.uk/12/2/3.html>.
- Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. *statnet : Software tools for the Statistical Modeling of Network Data*. Seattle, WA, 2003. URL <http://statnetproject.org>.
- Samer Hassan. Mentat : A Data-Driven Agent-Based Simulation of Social Values Evolution. *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, MABS 2009 : Multi-Agent-Based Simulation X :135–146, 2009. ISSN 0302-9743.
- Brian Heath, Raymond Hill, and Frank Ciarallo. A survey of agent-based modeling practices (January 1998 to July 2008). *Journal of Artificial Societies and Social Simulation*, 12(4) :9, 2009. URL <http://jasss.soc.surrey.ac.uk/12/4/9.html>.
- Alison J. Heppenstall, Andrew T. Crooks, Linda M. See, and Michael Batty, editors. *Agent-Based Models of Geographical Systems*. Springer Netherlands, Dordrecht, 2012. ISBN 978-90-481-8926-7 978-90-481-8927-4. doi : 10.1007/978-90-481-8927-4. URL <http://link.springer.com/10.1007/978-90-481-8927-4>.
- Carlos P. Herrero. Ising model in small-world networks. *Physical Review E*, 65(6), June 2002. ISSN 1063-651X, 1095-3787. doi : 10.1103/PhysRevE.65.066110. URL <http://arxiv.org/abs/cond-mat/0206079>. arXiv : cond-mat/0206079.
- Paul W. Holland and Samuel Leinhardt. An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76(373) :33, March 1981. ISSN 01621459. doi : 10.2307/2287037. URL <https://www.jstor.org/stable/2287037?origin=crossref>.
- Sascha Holzhauser. Which network shall we choose? An evaluation framework to select social network modelling approaches for agent-based simulation. 2015. URL <https://www.researchgate.net/>

- [profile/Sascha_Holzhauer/publication/281966335_Which_network_shall_we_choose_An_evaluation_framework_to_select_social_network_modelling_approaches_for_agent-based_simulation/links/55ffcf9608aeba1d9f840eda.pdf](https://www.jstatsoft.org/profile/Sascha_Holzhauer/publication/281966335_Which_network_shall_we_choose_An_evaluation_framework_to_select_social_network_modelling_approaches_for_agent-based_simulation/links/55ffcf9608aeba1d9f840eda.pdf).
- David R. Hunter, Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. *ergm* : A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of Statistical Software*, 24(3), 2008. ISSN 1548-7660. doi : 10.18637/jss.v024.i03. URL <http://www.jstatsoft.org/v24/i03/>.
- M. Huxham, S. Beaney, and D. Raffaelli. Do Parasites Reduce the Chances of Triangulation in a Real Food Web? *Oikos*, 76(2) :284, June 1996. ISSN 00301299. doi : 10.2307/3546201. URL <https://www.jstor.org/stable/3546201?origin=crossref>.
- Marzieh Jahanbazy, Christopher Frantz, Maryam Purvis, and Martin Purvis. Building an Artificial Primitive Human Society : An Agent-Based Approach. In *Coordination, Organizations, Institutions, and Norms in Agent Systems X*, volume 9372, pages 89–96. Springer International Publishing, Cham, 2015. ISBN 978-3-319-25419-7 978-3-319-25420-3. doi : 10.1007/978-3-319-25420-3_6. URL http://link.springer.com/10.1007/978-3-319-25420-3_6.
- H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833) :41–42, May 2001. ISSN 0028-0836, 1476-4687. doi : 10.1038/35075138. URL <http://www.nature.com/articles/35075138>.
- Tobias Johansson. Generating Artificial Social Networks. 2017. URL <http://www.diva-portal.org/smash/record.jsf?pid=diva2:1065198>.
- Pedro Jordano, Jordi Bascompte, and Jens M. Olesen. Invariant properties in coevolutionary networks of plant-animal interactions : Invariant properties in coevolutionary networks. *Ecology Letters*, 6(1) :69–81, December 2002. ISSN 1461023X, 14610248. doi : 10.1046/j.1461-0248.2003.00403.x. URL <http://doi.wiley.com/10.1046/j.1461-0248.2003.00403.x>.
- Hale Kirer and Yasemin Cirpici. A Survey of Agent-Based Approach of Complex Networks. *Ekonomik Yaklasim*, 27(98) :1, 2016. ISSN 1300-1868. doi : 10.5455/ey.35900. URL <http://www.scopemed.org/?mno=163685>.
- Ravi Kumar, Jasmine Novak, and Andrew Tomkins. Structure and Evolution of Online Social Networks. page 7, 2006.
- Jerome Kunegis. *On the spectral evolution of large Networks*. PhD thesis, University of Koblenz–Landau, 2011.
- Emmanuel Lazega, Lise Mounier, Tom Snijders, and Paola Tubaro. Norms, status and the dynamics of advice networks : a case study. *Social Networks*, 34 :

- 323–332, 2012. URL <https://hal.archives-ouvertes.fr/hal-01800221/document>.
- David Lazer and Allan Friedman. The Network Structure of Exploration and Exploitation. *Administrative Science Quarterly*, 52(4) :667–694, December 2007. ISSN 0001-8392, 1930-3815. doi : 10.2189/asqu.52.4.667. URL <http://journals.sagepub.com/doi/10.2189/asqu.52.4.667>.
- Christophe Le Page. Simulation multi-agent interactive : engager des populations locales dans la modélisation des socio-écosystèmes pour stimuler l’apprentissage social, 2017. URL <http://agritrop.cirad.fr/583938/>.
- Jure Leskovec. *Dynamics of Large Networks.pdf*. PhD thesis, 2008.
- Jure Leskovec and Andrej Krevl. *SNAP Datasets : Stanford Large Network Dataset Collection*. June 2014. URL <http://snap.stanford.edu/data>.
- Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs : An approach to modeling networks. *The Journal of Machine Learning Research*, 11 :985–1042, 2010. URL <http://dl.acm.org/citation.cfm?id=1756039>. bibtex : leskovec.kronecker_2010.
- David Levinson and Yao Wu. The rational locator reexamined : Are travel times still stable? *Transportation*, 32(2) :187–202, March 2005. ISSN 0049-4488, 1572-9435. doi : 10.1007/s11116-004-5507-4. URL <http://link.springer.com/10.1007/s11116-004-5507-4>.
- W. Li and X. Cai. Statistical analysis of airport network of China. *Physical Review E*, 69(4), April 2004. ISSN 1539-3755, 1550-2376. doi : 10.1103/PhysRevE.69.046106. URL <https://link.aps.org/doi/10.1103/PhysRevE.69.046106>.
- Priya Mahadevan, Dmitri Krioukov, Kevin Fall, and Amin Vahdat. Systematic topology analysis and generation using degree correlations. In *ACM SIGCOMM Computer Communication Review*, volume 36, pages 135–146. ACM, 2006. URL <http://dl.acm.org/citation.cfm?id=1159930>.
- Hernán A. Makse, José S. Andrade, Michael Batty, Shlomo Havlin, and H. Eugene Stanley. Modeling urban growth patterns with correlated percolation. *Physical Review E*, 58(6) :7054–7062, December 1998. ISSN 1063-651X, 1095-3787. doi : 10.1103/PhysRevE.58.7054. URL <https://link.aps.org/doi/10.1103/PhysRevE.58.7054>.
- Gianluca Manzo, Simone Gabbriellini, Valentine Roux, and Freda Nkirote M’Mbogori. Complex Contagions and the Diffusion of Innovations : Evidence from a Small-N Study. *Journal of Archaeological Method and Theory*, 25(4) :1109–1154, 2018. doi : 10.1007/s10816-018-9393-z. URL <https://hal.sorbonne-universite.fr/hal-02187036>. Publisher : Springer Verlag.

- Peter V. Marsden. Core Discussion Networks of Americans. *American Sociological Review*, 52(1) :122, February 1987. ISSN 00031224. doi : 10.2307/2095397. URL <http://www.jstor.org/stable/2095397?origin=crossref>.
- Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a Feather : Homophily in Social Networks. *Annual Review of Sociology*, 27(1) :415–444, August 2001. ISSN 0360-0572, 1545-2115. doi : 10.1146/annurev.soc.27.1.415. URL <http://www.annualreviews.org/doi/10.1146/annurev.soc.27.1.415>.
- Telmo Menezes and Camille Roth. Symbolic regression of generative network models. *Scientific Reports*, 4 :6284, September 2014. ISSN 2045-2322. doi : 10.1038/srep06284. URL <http://www.nature.com/articles/srep06284>.
- Stanley Milgram. The small World Problem. *Psychology Today*, Vol. 2 :60–67, 1967. URL http://barabasilab.neu.edu/courses/phys5116/content/milgram_smallworld.pdf.
- Jean-Louis le Moigne. *La théorie du système général : théorie de la modélisation*. Presses universitaires de France, Paris, 1977. ISBN 978-2-13-046515-7. OCLC : 951565693.
- Mark E. J. Newman. *Networks : an introduction*. Oxford University Press, Oxford, reprinted edition, 2016. ISBN 978-0-19-920665-0. OCLC : 1001287500.
- Mark EJ Newman. Assortative mixing in networks.pdf. 2002. URL <http://arxiv.org/pdf/cond-mat/0205405v1.pdf>.
- Mark EJ Newman. The structure and function of complex networks. March 2003. URL <https://arxiv.org/pdf/cond-mat/0303516v1.pdf>.
- John F. Padgett and Christopher K. Ansell. Robust Action and the Rise of the Medici, 1400-1434. *American Journal of Sociology*, 98(6) :1259–1319, May 1993. ISSN 0002-9602, 1537-5390. doi : 10.1086/230190. URL <https://www.journals.uchicago.edu/doi/10.1086/230190>.
- Vilfredo Pareto. Pareto (Vilfredo) - Cours d'économie politique. *Revue économique*, 16(5) :811–812, 1965. ISSN 0035-2764. URL https://www.persee.fr/doc/reco_0035-2764_1965_num_16_5_407682_t1_0811_0000_001.
- H. Van Dyke Parunak, Robert Savit, and Rick L. Riolo. Agent-based modeling vs. equation-based modeling : A case study and users' guide. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, pages 10–25. Springer, 1998. URL http://link.springer.com/chapter/10.1007/10692956_2.
- R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL <http://www.R-project.org/>.

- Anatol Rapoport and William J. Horvath. A study of a large sociogram. *Behavioral Science*, 6(4) :279–291, January 2007. ISSN 00057940, 10991743. doi : 10.1002/bs.3830060402. URL <http://doi.wiley.com/10.1002/bs.3830060402>.
- Romain Reuillon, Sebastien Rey-Coyrehourcq, and Matthieu Leclaire. OpenMOLE, a workflow engine specifically tailored for the distributed exploration of simulation models. *Future Generation Computer Systems*, 29(8) :1981 – 1990, 2013. URL <http://www.openmole.org/files/FGCS2013.pdf>.
- Ignacio Rodríguez-Iturbe and Andrea Rinaldo. *Fractal river basins : chance and self-organization*. Cambridge Univ. Press, Cambridge, 1. paperback ed edition, 2001. ISBN 978-0-521-00405-3 978-0-521-47398-9. OCLC : 47637708.
- Jean-Paul Rodrigue, Claude Comtois, and Brian Slack. *The geography of transport systems*. Routledge, Taylor & Francis Group, London ; New York, 4th edition edition, 2017. ISBN 978-1-138-66956-7 978-1-138-66957-4.
- Nicole Ronald, Theo Arentze, and Harry Timmermans. Modeling social interactions between individuals for joint activity scheduling. *Transportation Research Part B : Methodological*, 46(2) :276–290, February 2012. ISSN 01912615. doi : 10.1016/j.trb.2011.10.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S0191261511001433>.
- Brian Sallans, Alexander Pfister, Alexandros Karatzoglou, and Georg Dorffner. Simulation and Validation of an Integrated Markets Model. *Journal of Artificial Societies and Social Simulation*, 6(4), 2003. URL <http://jasss.soc.surrey.ac.uk/6/4/2.html>.
- Hiroki Sayama. *Introduction to the modeling and analysis of complex systems*. January 2015. ISBN 978-1-942341-06-2 1-942341-06-7 978-1-942341-08-6 1-942341-08-3 978-1-942341-09-3 1-942341-09-1. URL <http://textbooks.opensuny.org/introduction-to-the-modeling-and-analysis-of-complex-systems>. OCLC : 918567125.
- Marten Scheffer, Steve Carpenter, Jonathan A. Foley, Carl Folke, and Brian Walker. Catastrophic shifts in ecosystems. *Nature*, 413(6856) :591–596, October 2001. ISSN 0028-0836. doi : 10.1038/35098000. URL <http://www.nature.com/doifinder/10.1038/35098000>.
- Matthias Scholz. Node similarity as a basic principle behind connectivity in complex networks. *arXiv preprint arXiv :1010.0803*, 2012. URL <http://arxiv.org/abs/1010.0803>.
- Tom AB Snijders, Gerhard G. Van de Bunt, and Christian EG Steglich. Introduction to stochastic actor-based models for network dynamics. *Social networks*, 32(1) :44–60, 2010. URL <http://www.sciencedirect.com/science/article/pii/S0378873309000069>.

- Harold Stanislaw. Tests of computer simulation validity : what do they measure ? *Simulation & Games*, 17(2) :173–191, 1986. bibtex : stanislaw1986tests.
- Jiongming Su, Baohong Liu, Qi Li, and Hongxu Ma. Coevolution of Opinions and Directed Adaptive Networks in a Social Group. *Journal of Artificial Societies and Social Simulation*, 17(2), 2014. ISSN 1460-7425. doi : 10.18564/jasss.2424. URL <http://jasss.soc.surrey.ac.uk/17/2/4.html>.
- Samuel Thiriot. Small world is not enough : Criteria for network choice and conclusiveness of simulations, 2010. URL http://samuelthiriot.res-ear.ch/IMG/pdf/thiriot_2010_1.pdf.
- Samuel Thiriot and Jean-Daniel Kant. Generate country-scale networks of interaction from scattered statistics. page 12, 2008.
- Alexei Vázquez. Growing network with local rules : Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(5), May 2003. ISSN 1063-651X, 1095-3787. doi : 10.1103/PhysRevE.67.056104. URL <http://link.aps.org/doi/10.1103/PhysRevE.67.056104>.
- Duncan J. Watts and Steven H. Strogatz. Collective dynamics of /‘small-world/’ networks : Article : Nature, 1998. URL <http://www.nature.com/nature/journal/v393/n6684/full/393440a0.html>.
- Warren Weaver. Science and complexity. In *Facets of Systems Science*, pages 449–456. Springer, 1947. URL http://link.springer.com/chapter/10.1007/978-1-4899-0718-9_30.
- Sze-Sze Wong. Judgments about knowledge importance : The roles of social referents and network structure. *Human Relations*, 61(11) :1565–1591, November 2008. ISSN 0018-7267, 1741-282X. doi : 10.1177/0018726708096638. URL <http://journals.sagepub.com/doi/10.1177/0018726708096638>.