



# École Pratique des Hautes Études

---

## Thèse

présentée par **Coralie PETERMANN**

pour l'obtention du grade de

## Docteur

Spécialité : Informatique

### **Contribution à la modélisation de systèmes socio-techniques complexes : cas du Smart Grid, analyse structurelle et simulation**

soutenue publiquement devant le jury composé de :

#### Rapporteurs

Dr Antoine JOUGLET      MCF, HDR à l'Université de Compiègne  
Pr Jalel BEN-OTHTMAN    Professeur à l'Université de Paris 13

#### Examineurs

Pr Michel LAMURE      Professeur à l'Université Lyon I  
Dr Isis TRUCK            MCF, HDR à l'Université Paris 8  
Pr Ivan LAVALLÉE        Professeur à l'Université Paris 8

#### Directeur

Pr Marc BUI              Professeur et Directeur d'Etudes Cumulant à L'EPHE

#### Co-Directeur

Dr Soufian BEN AMOR    MCF, HDR à l'Université de Versailles

---



*"On n'est pas en thèse, on entre en thèse comme on entre en religion."*  
Ivan Lavallée.



# Table des matières

<b>Remerciements</b>	<b>vi</b>
<b>Résumé</b>	<b>xi</b>
<b>Introduction</b>	<b>1</b>
<b>1 Systèmes complexes et Smart Grids</b>	<b>5</b>
1.1 Systèmes complexes . . . . .	7
1.2 Propriétés structurelles des systèmes complexes . . . . .	7
1.2.1 Classification par domaine d'application . . . . .	7
1.2.2 Classification théorique . . . . .	11
1.3 Les Smart Grids : l'avenir du réseau électrique . . . . .	16
1.3.1 Présentation du concept de Smart Grid . . . . .	16
1.3.2 Une évolution vitale . . . . .	16
1.3.3 Une évolution progressive . . . . .	18
1.3.4 Le compteur intelligent : élément essentiel de l'intelligence du Smart Grid . . . . .	18
1.3.5 Le Smart Grid : uniquement des avantages pour le consommateur ?	20
1.4 Anatomie d'un Smart Grid . . . . .	20
1.4.1 Définition et détermination des acteurs . . . . .	21
1.4.2 Définition de l'environnement . . . . .	22
1.5 Le Smart Grid est un système complexe . . . . .	23
1.6 Optimisation dans les Smart Grids . . . . .	24
1.6.1 Enjeux liés aux technologies de l'information et de la communication (TIC) . . . . .	24
1.6.2 Enjeux économiques tarifaires et d'efficacité du réseau . . . . .	27
1.6.3 Enjeux environnementaux . . . . .	27
1.6.4 Enjeux d'optimisation de la gestion électrique . . . . .	28
<b>2 Réseaux socio-techniques complexes : quelle algorithmique innovante ?</b>	<b>29</b>
2.1 Les défis de représentation et d'analyse des systèmes socio-techniques . . . .	31
2.2 Espaces topologiques et théorie des graphes . . . . .	31
2.3 Analyse des réseaux d'influence dans les systèmes socio-techniques . . . . .	32
2.3.1 Définition de communauté . . . . .	33
2.3.2 Notre méthode . . . . .	33

2.3.3	Résultats . . . . .	36
2.4	Extension à la prétopologie . . . . .	41
<b>3</b>	<b>Modélisation intégrative de Smart Grids</b>	<b>43</b>
3.1	Introduction . . . . .	45
3.2	Modélisation du Smart Grid par les multi-agents . . . . .	45
3.2.1	Pourquoi utiliser une modélisation par multi-agents ? . . . . .	45
3.2.2	Exemples de modélisation de systèmes complexes . . . . .	47
3.3	Négociations autour de l'énergie : une modélisation à travers la théorie des jeux . . . . .	48
3.3.1	Jeux coopératifs . . . . .	48
3.3.2	Jeux de stratégies . . . . .	48
3.3.3	Jeux d'actions . . . . .	49
3.3.4	Jeux à mémoire parfaite et imparfaite . . . . .	50
3.3.5	Jeux séquentiels et jeux simultanés . . . . .	50
3.3.6	Approche bottom-up et top-down par les réseaux de jeux pour le partage de l'énergie . . . . .	50
3.4	Résilience et dynamique du réseau par la prétopologie . . . . .	51
3.4.1	Rappels et définitions . . . . .	51
3.4.2	Famille de prétopologies pour la représentation des différents critères et leurs intrications . . . . .	53
3.4.3	$\chi$ -connexités en prétopologie pour analyser la connectivité du réseau . . . . .	56
3.5	Analyse de la propagation de l'énergie et percolation . . . . .	57
3.5.1	Détermination du seuil de connectivité assurant l'efficacité du réseau . . . . .	58
3.5.2	Prévention des pannes et analyse du phénomène de brownout . . . . .	59
3.6	Déroulement de notre modèle . . . . .	60
<b>4</b>	<b>Simulation multi-agents autonome pour l'optimisation au sein de Smart Grids</b>	<b>65</b>
4.1	Le Smart Grid : un système socio-technique autonome . . . . .	67
4.2	Du modèle intégratif vers la simulation multi-agents autonome . . . . .	67
4.2.1	Cas d'utilisations de notre simulation . . . . .	68
4.2.2	Modélisation des agents . . . . .	68
4.2.3	Granularités spatiale et temporelle . . . . .	70
4.2.4	Niveaux conceptuels . . . . .	70
4.2.5	Politique d'intégration des énergies renouvelables . . . . .	72
4.2.6	Niveaux modulaires . . . . .	73
4.3	Simulation : conception, implémentation et optimisation . . . . .	80
4.3.1	Besoins liés à la complexité du système et choix d'une plateforme de simulation . . . . .	80
4.3.2	Vérification et validation des résultats . . . . .	81
<b>5</b>	<b>Perspectives</b>	<b>87</b>
5.1	Proposition de définition mathématique de Smart Grids . . . . .	88
5.1.1	Graphe dynamique et structure des lignes électriques . . . . .	88

5.1.2	Problème de flots pour la vision globale du Smart Grid . . . . .	89
5.1.3	Sac à dos multi-objectifs temporalisé et définition du problème de partage de l'énergie . . . . .	89
5.2	Améliorations de notre modélisation de partage de ressources . . . . .	90
5.2.1	Une étude approfondie de la théorie des jeux . . . . .	90
5.2.2	Modélisation de la perte énergétique par la prétopologie . . . . .	91
5.2.3	Diminution de la complexité temporelle de la simulation . . . . .	91
5.2.4	Extension du modèle à d'autres problématiques . . . . .	91
5.3	Introduction aux outils d'analyse pour les Smart Grids . . . . .	92
5.3.1	La géostatistique pour l'analyse de flux dans les Smart Grids . . . .	92
<b>Conclusion générale</b>		<b>97</b>
<b>Bibliographie</b>		<b>103</b>
<b>Webographie</b>		<b>103</b>
<b>Publications</b>		<b>105</b>
<b>Annexes</b>		<b>108</b>
<b>Apprentissage automatique dans des Smart Grids</b>		<b>109</b>
5.4	Défi fouille de textes 2011 . . . . .	110
5.5	Défi fouille de textes 2012 . . . . .	117
<b>Nos publications en cours</b>		<b>125</b>
5.6	Article soumis . . . . .	126
5.7	Intelligences journal . . . . .	132



# Remerciements

Je voudrais en premier lieu remercier tous les membres du jury pour m'avoir fait l'honneur d'accepter de juger ce travail de thèse :

- Michel Lamure, Professeur à l'Université Lyon I, président du jury et examinateur.
- Antoine Jouglet, Maître de conférences à l'Université de Compiègne, rapporteur de thèse.
- Jalel Ben-Othman Professeur à l'Université Paris 13, rapporteur de thèse.
- Isis Truck, Maître de conférences à l'Université Paris 8, examinateur de thèse.
- Ivan Lavallée, Professeur à l'Université Paris 8, examinateur de thèse.
- Soufian Ben Amor Maître de conférences à l'Université de Versailles, co-directeur de thèse et examinateur.
- Marc Bui, Professeur à l'Université Paris 8 et directeur d'études cumulant à l'EPHE, directeur de thèse et examinateur.

Je tiens à remercier en particulier mon directeur de thèse le professeur Marc Bui qui a toujours cru en mes capacités et a su me montrer la voie à suivre. Il m'a prodigué des conseils précieux tout au long de mes recherches qui ont été d'une grande aide. Qu'il sache que je le considère comme un de mes mentors et que je lui voue une vive estime.

Merci également à Soufian Ben Amor, mon co-directeur de thèse pour son soutien. Il a toujours été disponible lorsque j'avais besoin d'un avis ou de conseils et a su diriger mes recherches tout en me laissant une grande part de liberté.

Je voudrais remercier le professeur Alain Bui qui a su me montrer la voie à suivre, et sans qui je n'aurais sûrement jamais entrepris mon travail de thèse.

J'ai eu le plaisir de participer au projet interdisciplinaire Nouveau-Mendel qui m'a permis d'ouvrir mon esprit à d'autres disciplines. Je remercie notamment François Queyrel pour les travaux que nous avons effectués en collaboration.

Mes remerciements vont bien sûr à tous ceux qui ont participé de près ou de loin à la réussite de cette thèse. Je remercie ainsi les membres des laboratoires LaISC, Prism, et CHArt ; en particulier mes collègues et amis pour leur bonne humeur et leur soutien : Murat, Omar, Vincent, Yann, François, Simon, Karim, Khalil, Ke, Pierre.

Merci également à celui qui partage ma vie depuis tant d'années pour son soutien inconditionnel.

Enfin, mes remerciements les plus profonds vont à mes parents, qui n'ont pas eu la chance de réaliser les études qu'ils souhaitaient, et qui m'ont soutenue sans jamais me pousser durant mes nombreuses années d'études. Ils m'ont appris que dans la vie tout est possible pourvu qu'on le veuille vraiment et qu'on s'investisse totalement. Qu'ils sachent qu'ils sont mon modèle de réussite, et que j'espère un jour leur ressembler.

Et à tous ceux qui étaient là, avec qui j'ai tant partagé. Merci.





CONTRIBUTION À LA MODÉLISATION DE SYSTÈMES SOCIO-TECHNIQUES COMPLEXES :  
CAS DU SMART GRID, ANALYSE STRUCTURELLE ET SIMULATION

**Résumé :**

Les systèmes technologiques actuels comprennent une composante sociétale due à l'impact des utilisateurs. Ces systèmes, dits socio-techniques complexes, comportent des sous-systèmes dont les acteurs ont des objectifs divergents parfois antagoniques. Le Smart Grid est une parfaite illustration de système socio-technique. Considéré comme l'évolution de la grille électrique actuelle, la vision que nous pouvons avoir du Smart Grid est celle d'un réseau complexe où les flux bidirectionnels d'informations influent sur la distribution de l'électricité afin d'obtenir un optimum global du partage et de l'utilisation de l'énergie. Nouvel élément au cœur du Smart Grid, le consommateur est un consommateur prenant pleinement part à l'évolution de ce système. Cette évolution vers le Smart Grid n'est-elle pas un réseau d'information au sens large où le consommateur devient consommateur ?

Les défis du Smart Grid sont liés d'une part à cette hétérogénéité d'acteurs ayant des intérêts divergents et d'autre part à une complexité multi-échelles spatiales et temporelles. Les acteurs du Smart Grid sont en accès concurrentiel sur une ressource commune. Les consommateurs expriment des demandes en énergie et désirent minimiser leurs dépenses, alors que les producteurs ont des seuils de production rendant la gestion des pics de consommation difficiles et souhaitent maximiser leurs bénéfices. Les relations entre entités de cette grille hiérarchique sont difficiles à modéliser car la connexité n'est plus suffisante pour représenter la notion de proximité entre entités. Des dynamiques complexes telles que la tarification de l'énergie, les pertes énergétiques, ou les productions variables des énergies renouvelables sont à prendre en compte. D'un point de vue global, la demande en énergie connaît des variations importantes avec des pics de consommation influant fortement sur le système par le phénomène visible de baisse de tension (brownout). Arrivé à un certain seuil, les producteurs ne peuvent plus faire face à la demande des consommateurs et ceux-ci subissent des pannes partielles pouvant conduire à une panne totale si le système n'arrive pas à se rétablir de lui-même. On y retrouve la notion du rapport local-global.

Nous proposons donc une modélisation intégrative pour aborder ce problème : chaque composante de notre modélisation permet de prendre en compte un aspect du Smart Grid et c'est leur intrication qui réalise une modélisation efficace du système socio-technique étudié. Ainsi, une composante modélise les négociations autour du partage de l'énergie, couplée à des méthodes d'optimisation combinatoire. Les relations entre entités prennent en compte les dynamiques complexes, et un superviseur global permet la surveillance du réseau pour détecter les seuils critiques pouvant mener à des baisses de tension, et ainsi assurer la prévention des pannes. Cependant, le modèle est impossible à appréhender tel quel, il est donc nécessaire de mettre en place une simulation informatique pour explorer les différents paramètres et tester différents scénarios pour illustrer la pertinence des méthodes. Cette étape permet de tester l'influence du paramétrage du réseau de distribution, le calibrage des algorithmes d'optimisation, et la surveillance globale et locale du système.

Nous proposons également une définition mathématique du concept de Smart Grid, qui jusqu'à maintenant était variable selon l'angle de vue. Cette définition est issue de nos réflexions liées à la dialectique entre modélisation et simulation.

**Mots-clés :** Système complexe, algorithmique, réseau énergétique, Smart Grid, simulation, optimisation.

---

ANALYSE AND MULTI-AGENTS MODELING OF COMPLEX SOCIO-TECHNICAL SYSTEM :  
CASE OF SMART GRID

**Abstract :**

Current technological systems include a societal component due to the impact of users. These systems, called socio-technical complex systems, contain numerous sub-systems which actors have conflicting even antagonistic objectives. The Smart Grid is a perfect example of socio-technical system. Considered as the evolution of the current power grid, the vision that we have of the Smart Grid is a complex network where the bidirectional flow of information influences on the distribution of electricity to obtain a global optimum of share and use of energy. New main element of the Smart Grid, the consumer is a prosumer fully involved in the development of this system. This change to the Smart Grid isn't it an information network where the consumer becomes prosumer ?

Challenges of Smart Grid are related firstly to the heterogeneity of actors with different interests and secondly to a spatial and temporal multiscale complexity. Smart Grid players are in a competitive access to a shared resource. Prosumers express energy requirements and want to minimize their expenses, while producers have production thresholds making the management of consumption peaks difficult and want to maximize their profits. Relations between entities in this hierarchical grid are difficult to model because the connexity is no longer sufficient to represent the notion of proximity between entities. Complex dynamics such as energy pricing, energy losses, or the variable production of renewable energy are taken into account. From a global point of view, the demand for energy supports significant changes with consumption peaks which strongly influence the system through the visible phenomenon of undervoltage ( called *brownout*). At a certain threshold, producers can no longer cope with the demand and prosumers suffer from partial failures which can lead to total failure if the system fails to recover by itself. It is the notion of local-global report.

We propose an integrative model to solve this problem : each component of our model allows to consider one aspect of the Smart Grid and it is their intricacy that achieves an efficient modeling of the socio-technical system studied. Thus, a component models negotiations on the sharing of power, coupled with methods of combinatorial optimization. Relations between entities take into account the complex dynamics and an overall supervisor allows network monitoring to detect critical thresholds that could lead to brownouts, and ensure the prevention of failures. However, the model is impossible to apprehend like this, it is necessary to develop a computer simulation to explore the various parameters and test different scenarios to illustrate the relevance of the methods. This step will test the influence of parameters of the distribution network, the calibration of optimization algorithms, and the global and local monitoring of the system.

We also propose a mathematical definition of Smart Grid, which until now was variable depending on the context. This definition comes from our discussions related to the dialectic between modeling and simulation.

**Keywords :** Complex system, algorithmic, energy network, Smart Grid, simulation, optimization.





# Introduction

La "science des systèmes complexes adaptatifs" étudie les systèmes comportant un grand nombre de composants par nature indépendants, fortement interconnectés et en interaction. Ces interactions adaptatives entre les parties du système peuvent révéler des phénomènes d'émergence et d'auto-organisation, dans lesquels des processus d'apprentissage et d'évolution apparaissent. Ces systèmes sont dits complexes par rapport à la complexité à les représenter et leur comportement est difficile à prédire.

Les systèmes technologiques actuels comprennent une composante sociétale due à l'impact des utilisateurs. Ces systèmes, dits sociotechniques complexes, comportent des sous-systèmes dont les acteurs ont des objectifs divergents parfois antagonistes. La dynamique de ces réseaux technologiques amène de la complexité et les rend difficiles voire impossibles à prédire sur la seule analyse des sous-parties du système. Les outils et méthodes pour la modélisation des systèmes sociotechniques sont relativement pauvres. Nous proposons dans ce travail, plusieurs contributions pour l'analyse et l'optimisation de tels systèmes.

Le Smart Grid est une parfaite illustration de système sociotechnique complexe. Considéré comme l'évolution de la grille électrique actuelle, la vision que nous pouvons avoir du Smart Grid est celle d'un réseau complexe où les flux bidirectionnels d'informations et d'électricité permettent un ajustement en temps réel à la demande, de façon à intégrer de manière optimale les énergies renouvelables et contrôler en tout point le réseau. La mise en place d'un système technologique de surveillance continue du réseau est une des évolutions prévues de la grille électrique. D'ici 2030, certains scénarios prédisent une augmentation jusqu'à 70% de la demande en électricité, tandis que la principale source de production - l'énergie nucléaire - pourrait diminuer en raison de projets de lois européens. Pour palier cette baisse, d'autres sources de production vertes vont devoir la remplacer. L'intégration stratégique des énergies renouvelables sur la grille va devenir un enjeu primordial. De plus, nouvel élément au cœur du Smart Grid, le consommateur devient consommateur en prenant pleinement part à la dynamique de ce système.

Le réseau de distribution de l'électricité se rapproche en fait plus d'un réseau d'information au sens large plutôt qu'un réseau technologique tel que le présentait Mark Newman. En particulier, le réseau social du consommateur peut influencer sur les sous-systèmes de consommation et de production d'énergie. Cette évolution vers le Smart Grid n'est-elle donc pas un réseau d'information où le consommateur progresse vers le consommateur ?

Cette thèse a eu lieu dans le cadre du projet EPIT 2.0<sup>1</sup> qui vise à optimiser la gestion

---

1. Le projet EPIT 2.0 est un projet de recherche qui regroupe Avob, Dbt, Dotvision, Pulse-Air, Zamiren, Sagemcom, Alstom / Bouygues, Renault, CEA LIST, CEA LITEN, Econoving, Fondaterra, Supelec

d'un Smart District. Cette optimisation passe par une gestion intelligente de l'énergie dans l'éco-quartier au niveau de la production, du stockage, et de la consommation à l'aide de systèmes d'informations innovants.

Les défis du Smart Grid sont liés d'une part à cette hétérogénéité d'acteurs ayant des intérêts divergents et d'autre part à une complexité multi-échelles spatiales et temporelles.

D'un côté, les défis statiques du Smart Grid regroupent les problèmes de modélisation ponctuels tels que la prévision de la production des énergies nouvelles. Les énergies renouvelables ont en effet des productions irrégulières dépendant de conditions externes aléatoires. Il faut donc envisager des solutions pour stocker l'énergie dans les périodes de surproduction. La détection et la gestion des pannes est aussi un objectif du Smart Grid. Le réseau de capteurs mis en place doit pouvoir détecter des seuils d'alerte indiquant une panne et savoir réagir en conséquence. D'autre part, les défis dynamiques du Smart Grid sont liés à l'adaptabilité du système et à sa complexité structurelle. La topologie du réseau électrique évolue en temps réel au gré des congestions et des pannes, et ne se base pas uniquement sur des notions métriques. Les relations entre consommateurs et producteurs sont complexes car elles comprennent des objectifs économiques divergents. Des dynamiques complexes telles que la tarification de l'énergie ou les pertes énergétiques sont également à prendre en compte. Nous devons donc développer une algorithmique originale adaptée à ces problèmes, afin de garantir le bon fonctionnement du système et assurer une qualité continue aux consommateurs.

Nous proposons ici une modélisation intégrative et autonome ie. un modèle adaptable dynamique dont chaque composante permet de prendre en compte un aspect du Smart Grid. C'est l'intrication des composantes qui réalise une modélisation efficace du système sociotechnique étudié.

Les différents consommateurs sont en perpétuelle concurrence autour de l'énergie. Des négociations sont nécessaires pour partager la ressource commune. Ces mécanismes sont proches de la théorie des jeux, qui met en relation des joueurs ayant chacun une stratégie afin de gagner un jeu. Chaque entité du réseau vise à maximiser sa satisfaction en énergie.

Il faut donc fournir un optimum global par des optimisations locales de l'utilisation de la ressource partagée. La recherche opérationnelle classique fournit des algorithmes locaux d'optimisation combinatoire, permettant par exemple de maximiser l'utilisation d'une ressource en respectant une ou plusieurs contraintes. Le partage de l'électricité assignée à un foyer entre tous les appareils électriques consommateurs est semblable à un problème de sac-à-dos où les objets sont les sources de consommation et la contrainte le seuil de l'énergie fournie.

Néanmoins, d'autres contraintes sont à prendre en compte. La structure du réseau influe sur les relations entre entités. Le caractère intermittent de certaines sources d'électricité induit une dynamique complexe de la connectivité des sources et du flux d'énergie. Les problèmes de pertes énergétiques et de congestion ne sont pas directement modélisables par des seuls critères quantitatifs fixés. Nous sommes contraints de modéliser des

critères qualitatifs associés à des dynamiques complexes par une théorie adaptée.

Enfin, la seule modélisation des contraintes et de la structure du réseau ne permet pas de garantir son fonctionnement au niveau global. Les propriétés émergentes du réseau ne sont pas prévisibles d'un point de vue local et nécessitent donc une surveillance globale. Ceci est particulièrement valable dans le cadre de la détection et de la prévention des pannes. Nous proposons une approche hybride centralisée et distribuée via un superviseur global du réseau pour détecter les seuils critiques pouvant mener à des baisses de tension locales, et ainsi assurer la prévention des pannes.

Cependant, le modèle décrit ci-dessus est impossible à appréhender tel quel, il est nécessaire de mettre en place une simulation informatique pour explorer les différents paramètres et tester différents scénarios pour illustrer la pertinence des méthodes. Cette simulation permet de tester l'influence du paramétrage du réseau de distribution, le calibrage des algorithmes d'optimisation, et la surveillance globale et locale du système. Elle peut devenir, selon son utilisation, un réel outil d'aide à la décision.

Ce mémoire de thèse est composé de cinq chapitres :

- **Le chapitre 1** est un état de l'art des systèmes complexes dont le Smart Grid fait partie. On y présentera les définitions des systèmes complexes et leurs principales propriétés. Puis, nous présenterons le concept de Smart Grid, et expliquerons pourquoi le Smart Grid est un système complexe.
- **Le chapitre 2** propose une analyse fonctionnelle et théorique du système sociotechnique complexe étudié. Nous y présentons notre première contribution concernant l'exploitation des  $\chi$ -connexités prétopologiques pour l'analyse et le dimensionnement de chemins multi-critères [18]. Nous proposons un algorithme de clusterisation par les composantes unilatéralement connexes de diamètre maximal défini [12]. Cet algorithme pourrait nous permettre de catégoriser les entités du Smart Grid selon leur niveau de consommation avec prise en compte des défauts de données.
- **Le chapitre 3** est une modélisation intégrative du Smart Grid. Nous proposons l'agrégation de plusieurs théories permettant chacune de répondre à un problème du Smart Grid [15]. C'est la synergie entre ces théories qui permet une modélisation efficace. Notre modélisation prétopologique multi-critères est basée sur la théorie des jeux et la percolation pour l'analyse et l'optimisation temps réel des échanges entre les entités du Smart Grid [14].
- **Le chapitre 4** présente le développement de la simulation informatique de notre modèle théorique. Cette simulation paramétrable est généralisable à tout problème de partage de ressources et permet de tester différents scénarios d'optimisation du système [8, 10].
- **Le chapitre 5** est consacré aux perspectives et conclusions de ce travail. Nous y proposons notamment une définition mathématique théorique du concept de Smart Grid, issue de nos réflexions liées à la dialectique entre modélisation et simulation. Nous présentons également la théorie de la géostatistique comme solution pour l'approximation de données lorsque les limites des modèles prédictifs classiques sont atteintes lors de situations non prévues. Enfin, la conclusion permet de faire le point sur le travail présenté, et donne également les nouvelles directions qu'il faudrait

suivre dans les recherches futures.

# Chapitre 1

## Systèmes complexes et Smart Grids

Dans ce chapitre, nous introduisons la définition d'un système complexe, et les classes de systèmes complexes selon leur domaine d'application. Nous verrons qu'une nouvelle classe de système complexe fait son apparition et se trouve actuellement au coeur de l'actualité. Cette classe, appelée systèmes socio-techniques, comprend l'ensemble des systèmes complexes comportant une surcouche liée à l'interaction humaine sur le réseau. Nous verrons que la grille électrique du future, appelée communément Smart Grid même si ce terme est encore flou, est un parfait exemple de système socio-technique complexe.

Mais d'où vient cette complexité ? Sur quoi repose l'évolution du réseau électrique ? Quels vont être les nouveaux défis liés à cette complexité ? Telles sont les questions auxquelles nous allons tenter d'apporter des réponses dans ce chapitre.

### Sommaire

<b>1.1</b>	<b>Systèmes complexes</b> . . . . .	<b>7</b>
<b>1.2</b>	<b>Propriétés structurelles des systèmes complexes</b> . . . . .	<b>7</b>
1.2.1	Classification par domaine d'application . . . . .	7
	Réseaux sociaux . . . . .	8
	Réseaux d'informations . . . . .	8
	Réseaux technologiques . . . . .	8
	Réseaux biologiques . . . . .	9
	Réseaux socio-techniques . . . . .	10
1.2.2	Classification théorique . . . . .	11
	Les mesures mathématiques issues de la théorie des graphes . . . . .	11
	Réseaux invariants d'échelle . . . . .	13
	Réseaux petit monde . . . . .	13
<b>1.3</b>	<b>Les Smart Grids : l'avenir du réseau électrique</b> . . . . .	<b>16</b>
1.3.1	Présentation du concept de Smart Grid . . . . .	16
1.3.2	Une évolution vitale . . . . .	16
1.3.3	Une évolution progressive . . . . .	18
1.3.4	Le compteur intelligent : élément essentiel de l'intelligence du Smart Grid . . . . .	18
1.3.5	Le Smart Grid : uniquement des avantages pour le consommateur ?	20

<b>1.4 Anatomie d'un Smart Grid . . . . .</b>	<b>20</b>
1.4.1 Définition et détermination des acteurs . . . . .	21
Les producteurs . . . . .	21
Nucléaire . . . . .	21
Thermique . . . . .	21
Energies renouvelables . . . . .	22
Le consommateur . . . . .	22
1.4.2 Définition de l'environnement . . . . .	22
Le rôle du régulateur . . . . .	22
Les agrégateurs . . . . .	23
Les intégrateurs . . . . .	23
Les paramètres extérieurs . . . . .	23
<b>1.5 Le Smart Grid est un système complexe . . . . .</b>	<b>23</b>
<b>1.6 Optimisation dans les Smart Grids . . . . .</b>	<b>24</b>
1.6.1 Enjeux liés aux technologies de l'information et de la communi- cation (TIC) . . . . .	24
Communication, connectivité et normalisation des protocoles .	25
Sécurité et confidentialité des données personnelles . . . . .	25
Difficultés techniques . . . . .	25
Gestion des données et stockage . . . . .	27
1.6.2 Enjeux économiques tarifaires et d'efficacité du réseau . . . . .	27
1.6.3 Enjeux environnementaux . . . . .	27
1.6.4 Enjeux d'optimisation de la gestion électrique . . . . .	28

---

## 1.1 Systèmes complexes

La complexité est une notion utilisée dans de nombreux domaines tels que la physique, l'écologie, ou encore l'informatique, et dont la définition est différente selon le domaine d'application. Dans notre cas, nous nous intéressons à la notion de complexité dans le cadre d'une mesure de richesse : elle apporte plus de propriétés aux systèmes étudiés.

Un système complexe est un système composé d'un grand nombre d'entités en interactions locales et simultanées dont le comportement global ne peut être déduit de celui de ses composants [?], autrement dit, le comportement global du système n'est pas déductible de celui de ses parties ou sous-systèmes ni de leurs propriétés [?]. D'après le NECSI <sup>1</sup>

*Complex Systems is a new field of science studying how parts of a system give rise to the collective behavior of the system, and how the system interacts with its environment.*

Un système qui consiste en une population d'agents connectés, est dit complexe s'il existe une dynamique d'émergence au niveau global résultant de l'action simultanée des agents, plutôt que d'un contrôleur central.

La complexité est également souvent associée au concept d'émergence. Par exemple, dans le domaine de la biologie, un système est dit complexe s'il présente des phénomènes d'émergence, c'est-à-dire que le système dans sa globalité est plus riche et plus complexe que l'ensemble des sous-systèmes qui le composent. Cette situation peut être décrite par l'inéquation suivante :

$$H(x, y) > H(x) + H(y)$$

avec  $x$  et  $y$  deux sous-systèmes et  $H$  une fonction mathématique qui exprime la richesse potentielle d'un système [?]. En physique statistique, cette fonction est appelée entropie du système [?]. L'émergence est un phénomène reconnaissable dans les domaines physiques, biologiques, écologiques, socio-économiques, linguistiques et plus généralement dans tout système dynamique comportant des phénomènes d'auto-organisation [?]. Au delà d'un certain seuil de complexité, de nouvelles propriétés peuvent apparaître dans ces systèmes, elles sont dites propriétés émergentes.

## 1.2 Propriétés structurelles des systèmes complexes

La structure des systèmes complexes peut être représentée par un réseau lui-même complexe, qui peut être classé selon deux critères. Le premier concerne la catégorisation par domaine. Le second est une classification théorique selon les propriétés mathématiques du réseau.

### 1.2.1 Classification par domaine d'application

On peut classer les réseaux complexes dans 4 catégories distinctes : réseaux sociaux, réseaux d'informations, réseaux technologiques et réseaux biologiques [?].

---

1. New England Complex Systems Institute, <http://necsi.edu>

**Réseaux sociaux** Un réseau social est un ensemble de personnes ou de groupes avec plusieurs liens ou interactions entre eux [?, ?, ?]. Ils peuvent représenter les relations d'amitiés [?, ?], les partenariats entre entreprises [?, ?], etc... Le processus de diffusion d'une épidémie parmi une population tel que présenté dans la figure 1.1, illustre la complexité d'un réseau social, avec des interactions entre individus variantes prenant en compte la dynamique des déplacements et des rencontres, pouvant ainsi mener à des propagations au delà des foyers familiaux et professionnels. On remarque également la présence de super noeuds, qui sont des éléments de diffusion importants. Dans le cadre d'une épidémie, la détection des super noeuds permet de connaître les cibles prioritaires pour les campagnes de vaccination afin de limiter la diffusion de la maladie.

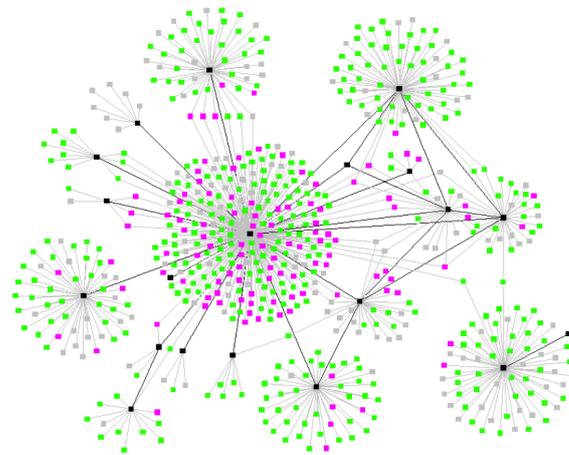


FIGURE 1.1 – Propagation d'une épidémie parmi une population d'individus.

**Réseaux d'informations** La seconde catégorie de réseaux complexes concerne les réseaux d'informations. Parfois aussi appelés "réseaux de connaissances" [?], les réseaux d'informations sont par exemple les réseaux de citations dans les publications scientifiques [?] dans lesquels les noeuds représentent les articles et les arêtes représentent les citations (une arête orientée de A vers B indique que l'article A cite l'article B). La figure 1.2 illustre bien la double contrainte temporelle et relationnelle dans les réseaux d'informations de type citations d'articles.

**Réseaux technologiques** La troisième catégorie de réseaux est constituée par les réseaux technologiques, réseaux construits par l'homme pour la distribution d'une commodité ou ressource, telles que l'électricité ou l'information [?]. Le réseau électrique est un bon exemple.

Beaucoup de réseaux de distribution ont été étudiés au cours du temps : les routes aériennes [?, ?], le réseau routier [?], le réseau ferroviaire [?], le trafic pédestre [?], mais aussi les réseaux électriques [?, ?]. Comme illustré par la figure 1.3, la structure logique d'internet est également un bon exemple grâce à toutes ses lignes de communication de

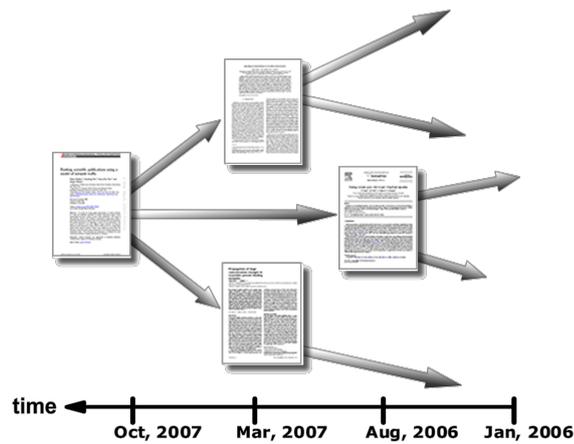


FIGURE 1.2 – Réseau de citation des articles.

différents types. Fibre optique, câbles, réseaux satellites, etc. permettent de conserver la connexion d'internet à l'aide d'une intercommunication entre différents supports.

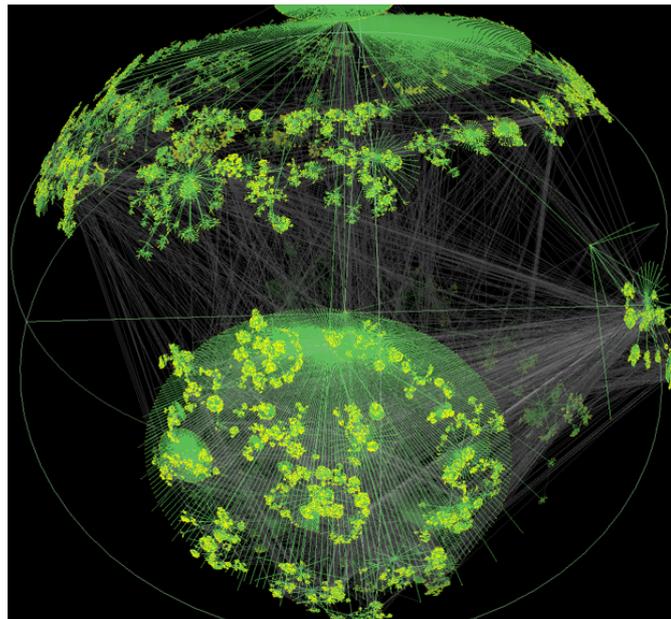


FIGURE 1.3 – Réseau technologique : la structure logique d'internet.

**Réseaux biologiques** La dernière catégorie des réseaux complexes est constituée par les réseaux biologiques [?]. Beaucoup de systèmes biologiques peuvent être représentés sous forme de réseaux. Par exemple, une protéine est un réseau d'acides aminés en interaction. Les acides aminés eux-mêmes peuvent être représentés comme un réseau d'atomes tels que carbone, azote et oxygène... La figure 1.4 présente 2 visualisations différentes d'une même protéine, selon les informations que l'on souhaite mettre en avant dans ce réseau complexe.

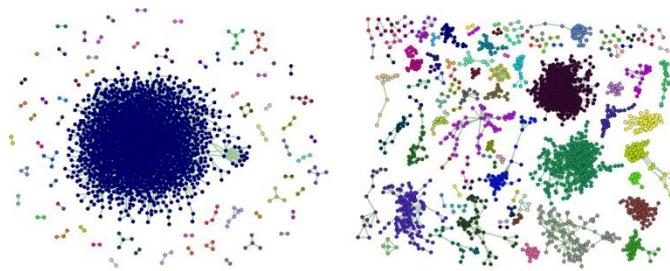


FIGURE 1.4 – Réseau biologique.

**Réseaux socio-techniques** Nous venons d'énumérer les 4 classes principales de réseaux complexes présentées par Mark Newman [?]. Cependant, les 4 catégories précédentes sont en fait des simplifications car dans la réalité nous avons souvent des situations plus compliquées. C'est pourquoi nous estimons qu'il est nécessaire d'ajouter une 5ème catégorie de réseaux complexes qui concerne les réseaux socio-techniques.

En effet, nous appelons réseau socio-technique, un réseau social sur lequel se greffe une couche de technologie. Cette classe de réseau complexe, comme l'illustre la figure 1.5, représente l'intrication dans un environnement commun entre un réseau social et un réseau technologique utilisé par des humains. La complexité est donc accrue et ce type de réseau représente de nombreux processus actuels tels que la grille électrique et l'utilisation et la gestion des appareils consommateurs.

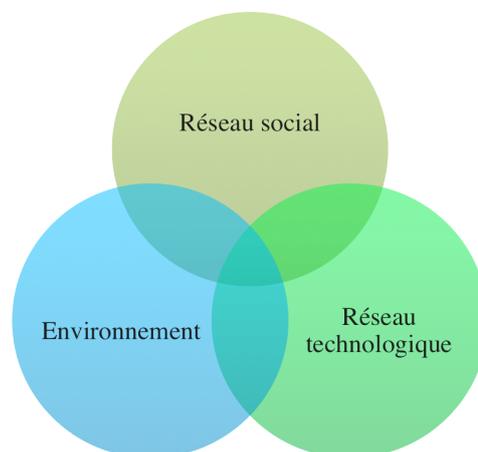


FIGURE 1.5 – Réseau socio-technique.

Nous avons vu les différents types de réseaux complexes réels dans divers domaines. Dans le prochain paragraphe nous présentons leurs propriétés mathématiques.

## 1.2.2 Classification théorique

Dans le cadre de nos recherches, nous allons travailler sur les réseaux électriques. Nous présentons dans cette partie une classification théorique des réseaux complexes en vue d'examiner les principales propriétés nécessaires au bon fonctionnement d'un réseau de distribution d'énergie.

Nous avons vu qu'un système complexe est un ensemble composé d'agents en interaction locales et globales. En particulier pour les réseaux électriques, il est nécessaire de modéliser le réseau physique de distribution d'énergie par la théorie des graphes. Cette théorie permet notamment d'analyser et de comparer les réseaux à l'aide d'outils algébriques.

### Les mesures mathématiques issues de la théorie des graphes

Plusieurs mesures algébriques permettent de décrire et définir certaines propriétés des graphes.

- **Le coefficient d'agglomération** : cette mesure représente la probabilité que deux noeuds voisins d'un 3eme noeud soient eux-mêmes voisins. Le coefficient global se calcule ainsi :

$$C = \frac{\text{Nombre de triplets fermés}}{\text{Nombre de triplets}} \quad (1.1)$$

- **La cohésion** : la cohésion est une mesure qui représente les liens directs entre chaque couple de sommets du graphe.
- **La densité** : la densité d'un graphe correspond au rapport entre le nombre de liens existants par rapport au nombre de liens possibles. Si ce rapport vaut 1, alors aucun autre lien ne peut être ajouté et on dit du graphe qu'il est complet.

$$d = \frac{\text{Nombre d'arêtes}}{\frac{n(n-1)}{2}} = 1 \text{ si graphe complet, compris entre 0 et 1 sinon.} \quad (1.2)$$

- **Le diamètre** : le diamètre d'un graphe correspond au plus long des plus courts chemins entre tout couple de sommets.
- **Le chemin moyen** : le chemin moyen est la distance moyenne entre tous les noeuds du graphe.
- **La distribution du degré** : la distribution du graphe est la courbe représentant le nombre de noeuds en fonction du degré.

La recherche théorique sur les réseaux a établi que les réseaux complexes peuvent être présentés à l'aide de graphes aléatoires particuliers. Un graphe aléatoire est un graphe dans lequel les arêtes sont aléatoirement distribuées [?]. Le premier modèle proposé, en 1959 par Paul Erdős et Alfréd Rényi [?], permet de générer mathématiquement un graphe aléatoire à partir d'un processus aléatoire qui consiste à ajouter des arêtes au hasard entre les noeuds avec une probabilité  $p$ .

### Algorithm 1

```

Erdős-Rényi(float p)
/* C :ensemble, n : nombre de noeuds */
G ← newGraph(C)
Pour i allant de 1 à n Faire
  Pour j allant de i+1 à n Faire
    Si random()/RAND_MAX ≤ p Alors
      G ← (i, j)
    Fin Si
  Fin Pour
Fin Pour

```

## Algorithm 2

```

Erdős-Rényi(integer m)
/* C :ensemble, n : nombre de noeuds, p() : fonction de probabilité */
nblinks ← 0)
Tant que nblinks < m Faire
  src ← random()
  dest ← random()
  pairp(min(src, dest), max(src, dest))
  Si src! = dest Alors
    G ← (src, dest)
    nblinks ++
  Fin Si
Fin Tant que

```

Dans ce modèle, dit Erdős-Rényi, un sommet est dit de degré  $k$  s'il est connecté à  $k$  autres sommets du graphe ; ce qui se traduit par la probabilité :

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \underset{N \rightarrow +\infty}{\sim} \frac{z^k e^{-z}}{k!} \quad (1.3)$$

où  $z = p(N-1) \underset{N \rightarrow +\infty}{\sim} pN$  est le degré moyen de chaque sommet. La proposition suivante est directement liée à cette définition :

**Théorème 1.1** *La distribution des degrés d'un graphe aléatoire issu du modèle d'Erdős-Rényi suit une loi de Poisson de paramètre  $z = pN$ .*

Dans le cadre de la modélisation des systèmes complexes, nous pouvons distinguer deux types de graphes aléatoires particuliers utilisés : les réseaux invariants d'échelle et les réseaux petit monde. Ces deux types de graphes sont intéressants dans la caractérisation des réseaux électriques. Selon Watts et Strogatz, la longueur des chemins et le coefficient d'agglomération des réseaux électriques sont caractéristiques des réseaux petits mondes [?]. Les réseaux de transport, et le transport de l'énergie en est un, peuvent être caractérisés par la propriété petit monde [?]. D'autres rapports mettent en évidence des similitudes avec les réseaux à invariance d'échelle [?]. Ces résultats s'avèrent en fait différents selon les pays étudiés. La granularité géospatiale joue même un rôle primordial car les résultats à l'échelle d'une région ne sont pas les mêmes qu'à l'échelle nationale [?, ?].

### Réseaux invariants d'échelle

Un réseau invariant d'échelle (également appelé *scale-free*) est un graphe dans lequel la distribution des degrés suit une loi de puissance. Un réseau invariant d'échelle est un réseau aléatoire comportant un nombre identifiable de noeuds dont le degré dépasse largement le degré moyen du graphe. Ces noeuds sont appelés super noeuds, colorés en gris sur la figure 1.6.

Ces super noeuds sont à la fois un atout et une faiblesse dans ces réseaux. En effet, la suppression de noeuds au hasard dans le réseau ne le fera pas s'effondrer. De même, la suppression d'un seul super noeud peut généralement être comblée par des chemins parallèles. La structure des réseaux invariants d'échelle permet une certaine tolérance aux pannes. Cependant, si les suppressions de super noeuds se multiplient, on atteint un seuil critique où le réseau n'est plus connexe et deviendra un ensemble de graphes isolés. Ce seuil est brutal et difficile à prédire.

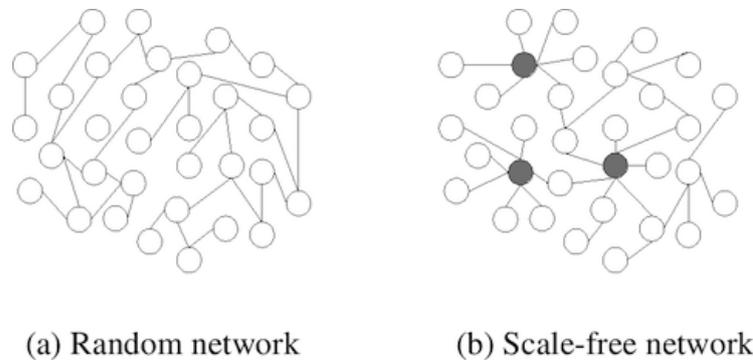


FIGURE 1.6 – Comparaison entre graphe aléatoire et réseau invariant d'échelle.

La distribution des degrés du graphe aléatoire est représentée par une courbe en forme de cloche 1.7, dont le sommet correspond au degré moyen du graphe, et les degrés les plus extrêmes sont atteints par une décroissance rapide, indiquant que ces graphes sont statistiquement homogènes. A l'inverse, la distribution des degrés du réseau invariant d'échelle suit une loi de puissance : elle diminue lorsque le degré moyen augmente.

Une autre caractéristique des réseaux invariants d'échelle concerne la distance moyenne entre 2 noeuds du réseau. En effet, le diamètre d'un réseau invariant d'échelle, même si celui-ci évolue et s'agrandit, reste relativement constant en pratique.

La propriété d'invariance d'échelle est présente dans beaucoup de domaines scientifiques, incluant la topologie des pages web, et le réseau électrique de l'ouest des Etats-Unis [?].

### Réseaux petit monde

Un réseau petit monde (appelé également *small-world*) est un type de graphe dans lequel la plupart des noeuds peuvent être atteints par tous les autres en un petit nombre d'étapes (voir la figure 1.8).

Stanley Milgram, un sociologue à Harvard, a conduit une expérience simple en 1967 ([?]). Il a développé une méthode simple pour étudier comment les individus sont connectés

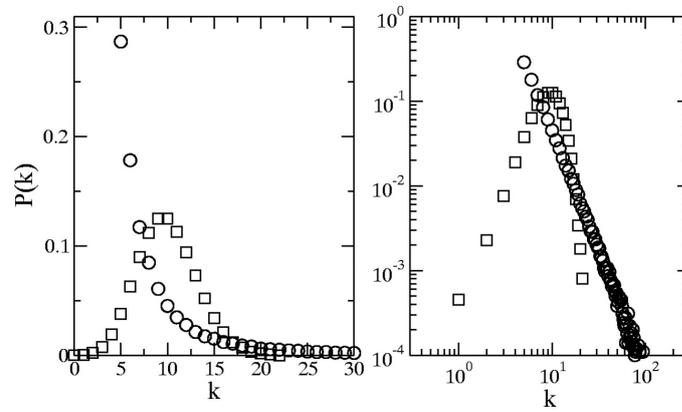


FIGURE 1.7 – Comparaison entre les degrés de distribution d’un réseau invariant d’échelle et d’un graphe aléatoire.

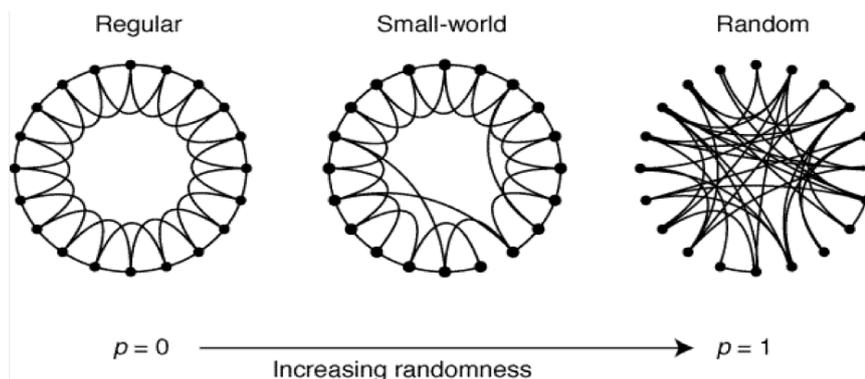
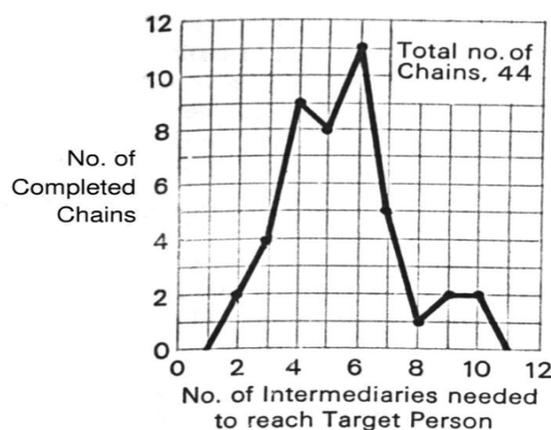


FIGURE 1.8 – Réseau Small World.

aux autres en donnant une simple enveloppe à une personne source devant l'envoyer à une personne cible. Si la source ne connaissait pas directement la cible, elle devait remettre l'enveloppe à quelqu'un parmi ses contacts, qui devrait à son tour transmettre l'enveloppe jusqu'à ce que la cible soit atteinte. Milgram pouvait alors déterminer la proximité de 2 personnes dans une population.

Dans l'étude originale, les sources du Kansas devaient faire parvenir l'enveloppe à la femme d'un étudiant de Cambridge, Massachusetts, et les sources du Nebraska devaient quant-à elles envoyer l'enveloppe à un courtier de Sharon, Massachusetts. Dans l'étude du Nebraska, le nombre médian de liens entre sources et cibles était de 5 intermédiaires, comme l'illustre la figure 1.9.



**In the Nebraska Study the chains varied from two to 10 intermediate acquaintances with the median at five.**

FIGURE 1.9 – Etude du Nebraska de Milgram.

Les graphes aléatoires obtenus à partir de la méthode d'Erdős et Rényi admettent un faible chemin moyen et un faible coefficient d'agglomération. Cependant, Watts et Strogatz ont mesuré que de nombreux réseaux réels présentaient un coefficient d'agglomération beaucoup plus important, caractéristique de la propriété petit monde. Ils ont donc proposé un nouveau modèle de génération, appelé modèle de Watts et Strogatz qui permet la modélisation de graphes ayant un faible plus court chemin et un fort coefficient d'agglomération. Ce modèle consiste à générer un réseau régulier de connectivité  $K$ , puis à remplacer les liens existants avec une probabilité  $p$  en choisissant les noeuds de façon uniforme et en évitant les boucles et les redondances.

**Définition 1.1 (Algorithme de génération selon Watts et Strogatz)** Soit  $N$  le nombre de noeuds voulus, le degré moyen  $K$  (entier pair), et un paramètre spécial  $\beta$ , satisfaisant  $0 \leq \beta \leq 1$  et  $N \gg K \gg \ln(N) \gg 1$ , le modèle construit un graphe non orienté avec  $N$  noeuds et  $\frac{NK}{2}$  arêtes selon le processus suivant :

Construire un anneau régulier avec  $N$  noeuds de connectivité  $K$ . Ainsi, soient les sommets  $n_0 \dots n_{N-1}$ , il y a un arc  $(n_i, n_j)$  si et seulement si  $0 < |i - j| \bmod n \leq \frac{K}{2}$ . Pour chaque noeud  $n_i = n_0, \dots, n_{N-1}$  prendre chaque arête  $(n_i, n_j)$  avec  $i < j$ , et remplacer

avec une probabilité  $\beta$ . La modification des arêtes est faite en remplaçant  $(n_i, n_j)$  avec  $(n_i, n_k)$  où  $k$  est choisi avec une probabilité uniforme parmi toutes les valeurs possibles ce qui empêche les boucles ( $k \neq i$ ) et le duplicata des liens (il n'y a pas d'arête  $(n_i, n_{k'})$  avec  $k' = k$ ).

## 1.3 Les Smart Grids : l'avenir du réseau électrique

Actuellement, le terme Smart Grid désigne un concept flou. Il existe plusieurs définitions en fonction des angles de vue, ainsi la vision du Smart Grid d'un législateur est différente de la vision d'un producteur électrique.

Globalement, le Smart Grid peut être défini comme la vision du réseau électrique du futur tel que présenté par Jeremy Rifkin [?], où les flux bidirectionnels d'informations et d'électricité permettent un ajustement en temps réel à la demande, de façon à intégrer de manière optimale les énergies renouvelables et contrôler en tout point le réseau [?].

Dans cette section, nous présentons une définition générale et indépendante du contexte des Smart Grids, et expliquons en quoi cette évolution est nécessaire.

### 1.3.1 Présentation du concept de Smart Grid

Le Smart Grid est composé d'équipements hétérogènes, depuis les compteurs intelligents<sup>2</sup> et les panneaux solaires, jusqu'aux équipements des stations électriques et les capteurs installés sur les lignes basses, moyennes et hautes tension. En même temps, la ressource électrique est produite par de multiples procédés : d'un côté une production stable par les centrales nucléaires et thermiques assurant un "fond de roulement", et de l'autre une production instable et dépendante de conditions extérieures avec l'intégration des énergies renouvelables. Le stockage de l'énergie est une solution envisagée comme nouvelle source d'énergie décentralisée via l'utilisation des véhicules électriques [?]. Un réseau de distribution et transport a été créé au fil des années mais sans être réellement maîtrisé ni optimisé. Notre but est de rendre la grille électrique plus efficace en intégrant de manière optimale les énergies renouvelables et en tirant parti des nouvelles technologies de l'information et de la communication.

### 1.3.2 Une évolution vitale

Quotidiennement, les appareils électriques se multiplient et les nouveaux besoins en énergie qu'ils génèrent font augmenter rapidement la demande globale en électricité (usage de plus en plus fréquent de la climatisation, chauffage électrique...) [?].

L'apparition prochaine de nouvelles technologies, parmi lesquelles nous pouvons prédire l'arrivée massive des voitures électriques, va stimuler la demande. Le graphique 1.11 présente les évolutions possibles de la consommation électrique en France selon différents scénarios envisagés. Les scénarios "haut" et "bas" illustrent les cas extrêmes. Le scénario "référence" est l'hypothèse de base pour les autres scénarios. Le scénario "MDE renforcée"

---

2. dit également *Smart Meter*, un compteur intelligent permet d'identifier de manière précise et en temps réel la consommation électrique d'un foyer



FIGURE 1.10 – Augmentation de la demande en électricité.

prend en compte l'accélération de la maîtrise de la demande d'énergie (MDE) en incluant des modifications des comportements de consommation, des optimisations de gestion des bâtiments, etc...

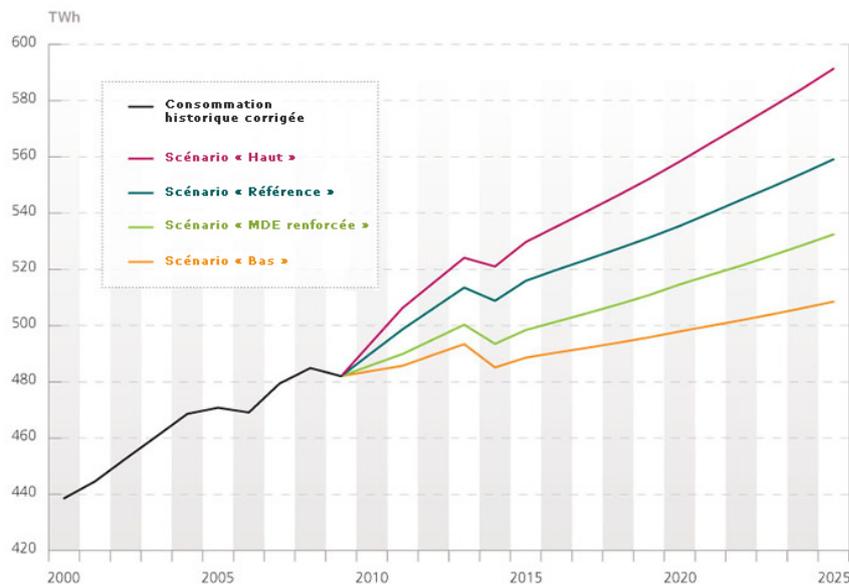


FIGURE 1.11 – Prévisions d'augmentation de la demande en électricité (source RTE).

D'ici 2030, la consommation mondiale d'électricité devrait augmenter de 76% par rapport à celle de 2007 [?].

Divers scénarios prédisent aussi un manque à venir de pétrole d'ici environ 20 ans. Ce manque devra être compensé par l'augmentation de la ressource électrique.

La grille actuelle va donc devenir obsolète. En effet, il est difficile de greffer de nouveaux équipements face à de nouvelles demandes sur un réseau déjà existant et qui n'a pas été

pensé pour cet usage. C'est par exemple le cas pour la gestion des énergies renouvelables. Le problème de ces énergies est leur intermittence de fonctionnement qui dépend des conditions climatiques et météorologiques. Il est donc difficile d'assurer une fourniture stable d'électricité aux consommateurs en se basant uniquement sur des énergies de ce type. De plus, les heures de production ne correspondent pas toujours à des demandes en électricité (pic de consommation vers 19h). L'électricité produite, sans solution de stockage est perdue [?]. Il est donc nécessaire de mettre en place des processus de décision et d'optimisation pour le partage et la gestion de l'énergie. Cela permettra notamment de raccorder les sources d'énergies renouvelables au réseau de manière plus efficace.

De plus, la grille actuelle est basée sur une infrastructure vieillissante : les transformateurs et les lignes principales ont plus de 40 ans, et les stations plus de 20 ans. La perte d'énergie et d'information lors des processus de transmission et distribution n'est pas quantifiable ni contrôlable [?]. Les flux d'informations sont pauvres : la plupart des compteurs électriques ne fournissent pas d'information en temps réel et le relevé des données n'est effectuée que 2 fois par an.

Enfin, face à l'enjeu mondial du réchauffement climatique, l'Union Européenne a mis en place des objectifs à tenir d'ici 2020 en vue de diminuer les émissions de gaz à effet de serre, optimiser l'intégration des énergies renouvelables, et baisser la consommation énergétique *per capita*. Ces mesures ne sont possibles qu'après mise en place d'un réseau intelligent capable de connaître en tout point l'état du système et de le gérer de manière efficace.

### 1.3.3 Une évolution progressive

Pour comprendre le besoin de créer de nouveaux réseaux intelligents qui évoquent le concept de Smart Grid, reprenons l'exemple du réseau français en 1990. La production était alors essentiellement basée sur le nucléaire, et les communications sur le réseau ne se réalisaient que dans un sens : du producteur vers le consommateur. Dans un premier temps, avec l'arrivée des énergies renouvelables, les sources de production se sont multipliées, et cela a entraîné des modifications sur le réseau de distribution (ajout de points communicants).

Afin de gérer ce réseau devenu complexe il est nécessaire de procéder à des changements importants sans en perturber le fonctionnement. Cette évolution importante liée aux besoins de la science se rapproche de la vision de Thomas Kuhn [?]. Cela se traduit dans le cadre du Smart Grid par la mise en place d'un réseau de capteurs et compteurs intelligents pour communiquer les données de production et consommation afin d'assurer l'équilibre entre l'offre et la demande. Le réseau intelligent (Smart Grid) sera alors communicant et interactif.

### 1.3.4 Le compteur intelligent : élément essentiel de l'intelligence du Smart Grid

Élément clé du Smart Grid, le compteur dit *intelligent* est un ensemble de capteurs qui permettent de fournir en temps réel la consommation énergétique d'une zone définie (foyer,

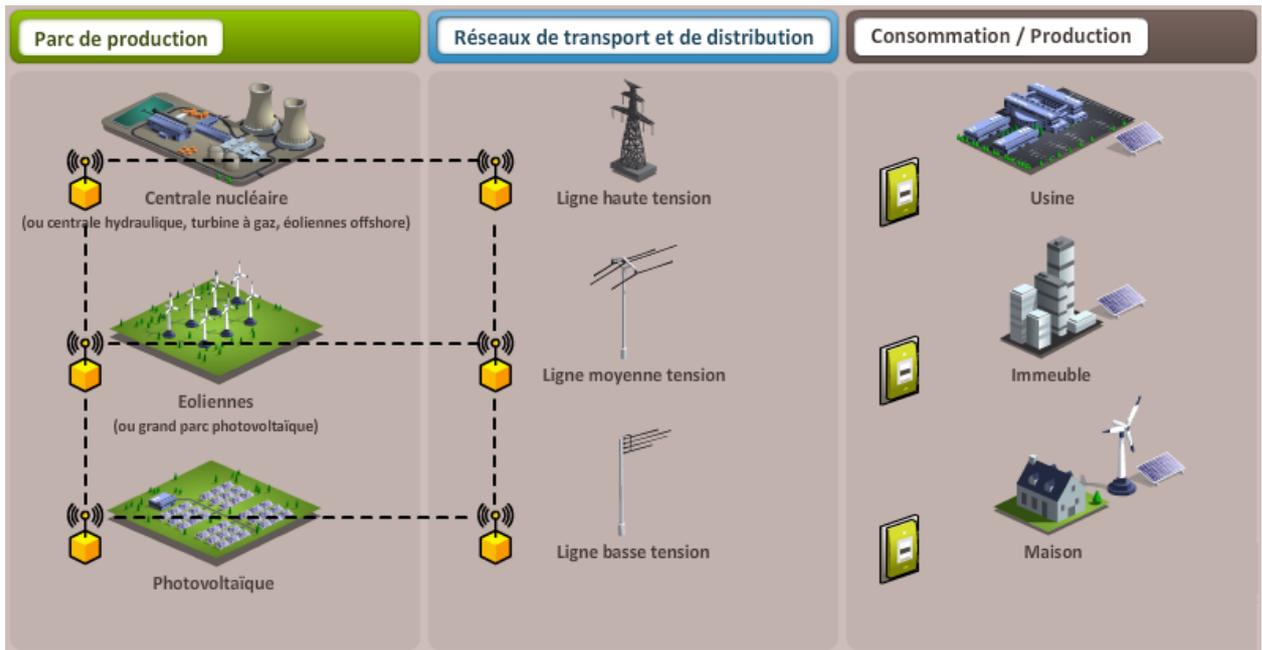


FIGURE 1.12 – Dernière étape dans la mise en place du Smart Grid : ajout de capteurs et compteurs intelligents.

entreprise, bâtiment ...). Comme illustré également par la figure 1.12 précédente, l'élément au cœur de la complexité du réseau électrique du futur sera le compteur intelligent. Sans ce fournisseur de données en temps réel, l'optimisation et la gestion automatique de la grille électrique n'est pas possible.

Deux types de compteurs co-existent : les compteurs dits AMR (*Automated Meter Reading*) qui consistent à transmettre uniquement les données de consommation ; et les compteurs AMM (*Automated Meter Management*) qui permettent en plus des actions de gestion en temps réel sur la consommation du foyer.

En France, ERDF, principal distributeur d'énergie, a lancé le projet *Linky* en 2007 qui a pour but d'équiper 35 millions de clients d'ici 2021 de compteurs de type AMM. L'installation du compteur Linky dans les foyers aurait des avantages pour tous les acteurs du réseau selon ERDF. D'une part, Linky réduirait les coûts de gestion pour les gestionnaires de réseaux. D'autre part, ces compteurs réduiraient les fraudes et les réclamations auprès des fournisseurs, et autoriseraient une diversification des offres proposées aux clients grâce à la richesse d'informations supplémentaires. D'un autre côté, pour les producteurs, un tel équipement rendrait possible une meilleure gestion des pics de consommation et une simplification de l'installation d'énergies renouvelables. Enfin, pour les clients, les factures seront établies directement d'après leur consommation réelle.

Cependant, cette optimisation en temps réel présente-t-elle uniquement des avantages pour les clients ?

### 1.3.5 Le Smart Grid : uniquement des avantages pour le consommateur ?

Grâce à l'évolution du système électrique, le consommateur va devenir **consommacteur**, c'est à dire qu'il va participer pleinement à la gestion de sa consommation, et pourra même devenir producteur.

En effet, les éoliennes et panneaux photovoltaïques, jusque-là réservés à la propre consommation du particulier, vont être injectés directement dans le Smart Grid, transformant le particulier en producteur décentralisé pouvant suppléer les producteurs principaux en cas de pic de consommation. Cette production chez les particuliers, pourrait leur assurer des revenus complémentaires en plus de leur assurer une fourniture en énergie continue. En vérité, même si certains particuliers arrivent à devenir autonomes par leur propre production d'électricité, il s'avère que selon le contexte, l'installation de tels équipements peut ne pas s'avérer rentable avant plusieurs décennies. La durée de vie du matériel peut aussi devenir problématique dans le calcul de rentabilité. De plus, ces installations dépendant de conditions extérieures, il ne faut pas compter uniquement sur elles pour s'alimenter en électricité. Une simple feuille posée sur un panneau photovoltaïque monté en série stoppe toute production d'énergie. De même pour l'éolien, un soleil de plomb sans une pointe de vent ne fait pas tourner les hélices et n'actionne donc pas la turbine de production. Le consommacteur, même suréquipé de sources d'énergies renouvelables, peut donc se retrouver dépendant du réseau électrique de base.

L'autre avantage annoncé pour le consommacteur concerne la gestion de sa consommation. Grâce aux compteurs intelligents, le consommacteur connaîtra en temps réel sa consommation et donc sa facture énergétique. Il passera ainsi d'un système prévisionnel avec vérification biannuelle d'un technicien, à un système précis en temps réel. Les surprises sur la note d'électricité n'existeront plus. De plus, les compteurs intelligents permettront une information précise sur l'utilisation faite par la domotique. Il sera ainsi aisé d'effectuer des bilans énergétiques personnalisés en vue de déterminer les changements d'habitudes pour réaliser des économies d'énergie. Les fournisseurs d'énergie pourront également proposer des offres adaptées au profil de consommation (diversification tarifaire) : un couple sans enfant aura une consommation électrique généralement plus faible qu'une famille nombreuse. Cependant, les nouveaux processus de tarification de l'énergie pourraient aussi être ambigus et s'avérer plus avantageux et rentables pour les producteurs plutôt que pour les consommateurs.

Cependant, grâce aux TIC, il sera plus facile de prédire et gérer les pics de consommation et donc d'éviter les pannes dues à une surcharge. Les opérations de maintenance pourront aussi s'effectuer à distance pour une rapidité accrue. La qualité de service pour les consommateurs sera donc grandement améliorée.

Dans la section suivante, nous présentons les divers enjeux liés au Smart Grid.

## 1.4 Anatomie d'un Smart Grid

Dans cette partie, nous nous attachons à définir d'un point de vue conceptuel les différentes parties d'un Smart Grid. Dans un premier temps, nous déterminons les différents

acteurs qu'ils soient humains ou matériels, puis dans une seconde partie, nous définissons l'environnement global.

### 1.4.1 Définition et détermination des acteurs

Le Smart Grid est l'évolution de la grille actuelle qui est elle-même composée d'un grand nombre d'entités. Le coeur de ce système repose bien sûr sur les producteurs de la ressource à partager.

#### Les producteurs

Les producteurs d'électricité sont au coeur du système. A eux seuls ils représentent un gros facteur de complexité. En effet, plusieurs types de producteurs impliquent plusieurs types de gestion et de fluctuation de l'énergie.

Comme l'indique la figure suivante, l'énergie française est produite majoritairement par 3 types d'énergies.

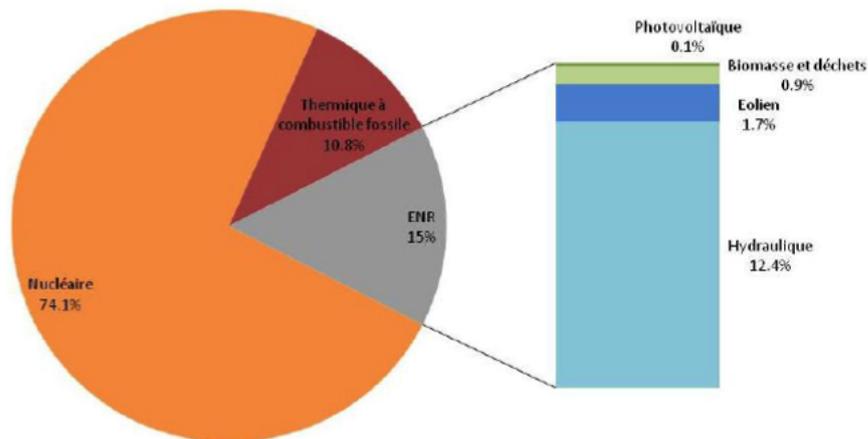


FIGURE 1.13 – La répartition de la production énergétique française.

**Nucléaire** La production nucléaire représente presque 75% de la production en France. Elle se caractérise par une production stable et à fort rendement.

**Thermique** La production thermique à combustion fossile produit également une grosse partie de l'électricité. Sa production est forte mais soumise à quelques fluctuations régulières.

**Energies renouvelables** Les énergies renouvelables ont une place grandissante dans la production. Elles sont variables et leur mode de production change radicalement d'un type à un autre. Certaines comme l'énergie solaire et les éoliennes dépendent directement des conditions climatiques externes. D'autres comme la production hydraulique, permettent de produire une grosse quantité d'énergie en très peu de temps, mais il faudra tout autant de temps pour réajuster le stockage, et cette production ne peut plus être augmentée de façon significative en France.

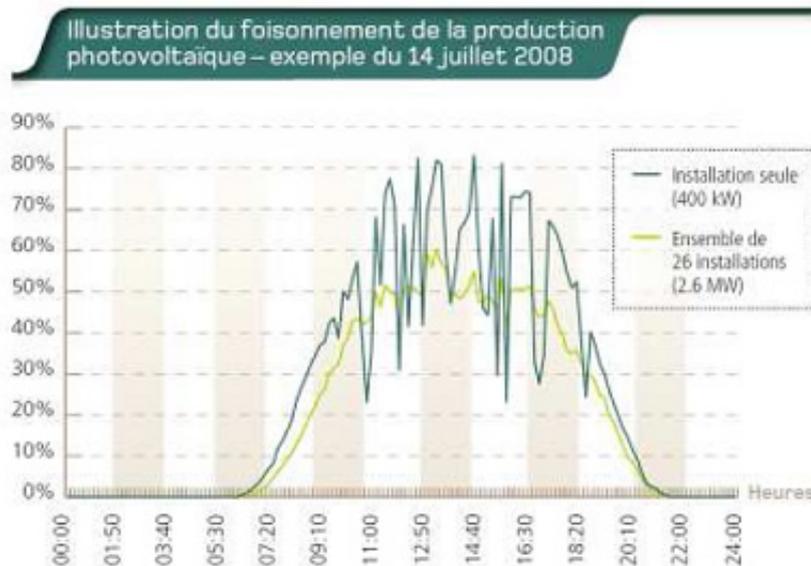


FIGURE 1.14 – Courbe de production journalière de l'énergie photovoltaïque.

## Le consommacteur

Avec l'évolution de la grille vers le Smart Grid, on ne parle plus de consommateur, mais de consommacteur. Les consommateurs deviennent responsables de leur consommation et le concept de participation à la réduction de consommation et à la production ou au stockage de l'énergie est grandissant. Les particuliers installent des panneaux photovoltaïques sur leurs toits, et certaines industries instaurent des parcs de voitures électriques au service de leurs employés pour stocker l'énergie excédentaire.

En France, la consommation d'énergie est réalisée en grande partie par les résidentiels et les industries qui totalisent à eux 2 plus de 90% de la consommation totale.

### 1.4.2 Définition de l'environnement

L'environnement du Smart Grid concerne tout ce qui a une influence externe sur les acteurs du système.

**Le rôle du régulateur** En France, la commission de régulation de l'énergie (CRE) va jouer un rôle central dans les Smart Grids. Cet organisme est chargé de la régulation du

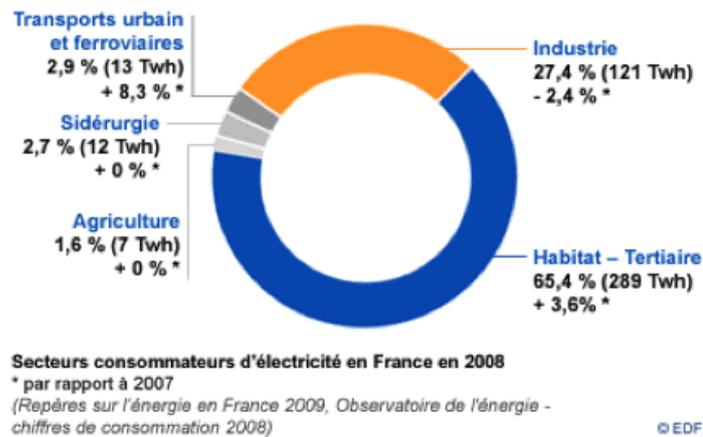


FIGURE 1.15 – Les différents consommateurs français.

secteur de l'énergie.

Le régulateur sera affecté par l'évolution du réseau en Smart Grid car il devra s'adapter au nouveau système de tarifs de l'électricité qui sera plus complexe que l'actuel en raison de la décentralisation de la production et de la diversification des producteurs. Il permettra donc de faciliter le déploiement des Smart Grid en adaptant la régulation d'énergie.

**Les agrégateurs** Les agrégateurs demandent aux industries, en les rémunérant, de baisser ponctuellement leur consommation électrique. Les industries étant de grandes consommatrices d'énergie, ce service permet de lisser ponctuellement la courbe de consommation.

**Les intégrateurs** Les intégrateurs représentent en fait des sociétés qui résolvent les problèmes d'interopérabilité en regroupant les services et produits de différents fournisseurs afin de proposer une solution globale.

**Les paramètres extérieurs** Enfin, comme tout système réel, le Smart Grid est régi par un certain nombre de critères stochastiques non prévisibles tels que le climat, les probabilités de pannes du système, ou encore le comportement humain.

## 1.5 Le Smart Grid est un système complexe

Comme vu dans la section précédente, le Smart Grid est composé d'un grand nombre d'entités hétérogènes également appelées agents (producteurs, consommateurs, puissance publique...) qui interagissent simultanément avec des objectifs divergents. Il peut s'agir d'enjeux économiques comme vendre au meilleur coût pour les producteurs, d'enjeux politiques comme la diminution des émissions de dioxyde de carbone, ou encore d'enjeux techniques de communication et sécurité.

De plus, la structure d'un Smart Grid est complexe. Le Smart Grid est donc un parfait exemple de système complexe.

De plus, la structure d'un système complexe est représentée par un réseau complexe, qui peut être vu comme la combinaison de plusieurs graphes. Dans le cadre des Smart Grids, le réseau de distribution et de communication représente un réseau complexe. D'un côté, le réseau de distribution est un ensemble de lignes électriques de différentes tensions qui permet à l'électricité de parcourir la grille pour être distribuée. D'un autre côté, le réseau de communication permet la communication entre les entités et collecte les statistiques issues des compteurs intelligents et des capteurs.

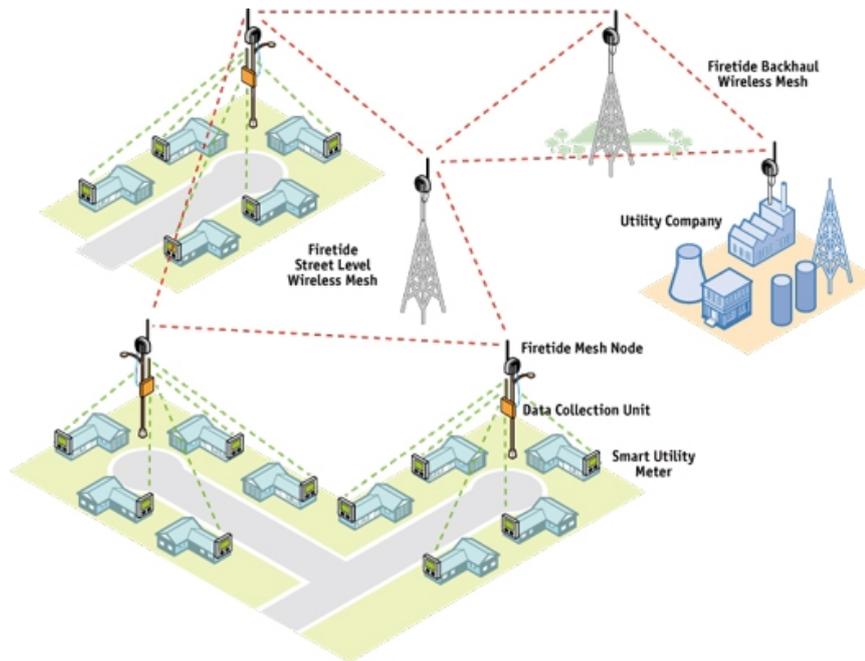


FIGURE 1.16 – Réseaux complexes des Smart Grids.

## 1.6 Optimisation dans les Smart Grids

Dans cette section, nous présentons l'ensemble des enjeux sous tendus par le concept des Smart Grids, en particulier ceux liés à l'utilisation des nouvelles technologies de l'information et de la communication, et aux impacts économiques et environnementaux.

### 1.6.1 Enjeux liés aux technologies de l'information et de la communication (TIC)

Les technologies de l'information et de la communication dites TIC vont jouer un rôle primordial dans la mise en place d'un Smart Grid, en assurant la liaison entre les divers agents pour les protocoles de communication des données issues des compteurs intelligents (linky...). Elles vont également apporter des solutions en terme d'infrastructures du réseau, sécurisation et gestion des données, etc...

Nous distinguons trois niveaux pour modéliser les Smart Grids à l'aide des TIC :

- la couche infrastructure : qui concerne les infrastructures du réseau électrique (lignes, transformateurs, distribution de l'énergie,...),
- la couche communication : pour les infrastructures de communication (fibre optique, GPRS, CPL, etc. ) pour collecter les données issues des capteurs,
- la couche applications et services pour piloter et gérer l'ensemble (systèmes de dépannage à distance, programme autonome de réponse à la demande en temps réel, etc.).

Nous présentons dans cette section les divers domaines couverts par les TIC dans le cadre des Smart Grids.

### **Communication, connectivité et normalisation des protocoles**

Le premier domaine géré par les TIC va être bien sûr la communication. L'apport des TIC va concerner la normalisation des protocoles de communication afin d'obtenir une harmonie au niveau des échanges entre agents. La mise en place de cette connectivité est nécessaire du fait de la multiplicité des différents agents intervenant dans le processus de gestion et production de l'énergie. De plus, avec l'apparition d'autres opérateurs, il est nécessaire de mettre en place des normes afin de limiter les efforts (et les pertes énergétiques) lors d'un changement d'opérateur si les protocoles ne peuvent pas communiquer.

A terme, ces protocoles devraient être également capable de s'interfacer avec la gestion des autres énergies et services des foyers (gaz, eau, déchets...).

### **Sécurité et confidentialité des données personnelles**

Avec la mise en place du Smart Grid et des TIC, de nombreuses données confidentielles risquent de circuler sur le réseau (relevés de compteurs, structure de la consommation, nom et adresse...). Le piratage d'un terminal de communication pourrait amener des individus malveillants à récupérer des informations confidentielles des consommateurs ou producteurs d'énergie, avec par exemple pour but de l'espionnage industriel (producteurs), ou encore le repérage des départs en vacances pour un cambriolage (particuliers).

Le Smart Grid doit donc permettre une sécurité efficace via une auto-surveillance constante. Le Smart Grid doit ainsi être résistant aux fraudes, attaques et autres cybermenaces [?].

De plus, conformément à la législation en vigueur, les données confidentielles doivent être cryptées et anonymisées pour assurer la protection des personnes physiques ou morales.

### **Difficultés techniques**

Les difficultés techniques de mise en place d'un Smart Grid regroupent plusieurs thématiques.

La première concerne les problèmes liés à la mise en place de nouvelles fonctionnalités sur une grille obsolète qui n'avait pas été pensée pour un tel usage. En effet, les énergies renouvelables présentent de fortes variations horosaisonnnières et géographiques. Il faut donc

mettre en place des politiques de gestion et partage de l'énergie afin d'assurer un équilibre entre l'offre et la demande, sans jamais occasionner de coupure chez les consommateurs.

La seconde thématique concerne la mise en place ou la création de normes. En effet, pour un réseau d'une telle ampleur avec des changements profonds, la multiplicité des processus et des techniques va engendrer une nécessité de normalisation pour permettre l'interopérabilité. Les Smart Grid regroupent un grand nombre de domaines et d'applications tels que le transport, la communication, les capteurs, les véhicules électriques... Chaque domaine va devoir faire l'objet de normes. Sur le plan uniquement des réseaux électriques, en 2010, la commission européenne a lancé la mise en place sur 2 ans du mandat de normalisation M/490. Très prochainement, un plan d'action devrait être publié afin de valider des normes techniques pour le secteur de l'électricité. Pour cela, une architecture de référence a été choisie par des experts, afin d'en identifier et formaliser les cas d'usages. Cela devrait permettre de fournir un ensemble de normes et de besoins prioritaires à court terme. Sur le plan plus généralisé des Smart Grids, c'est un groupe stratégique nommé SG 3, membre de la commission électrotechnique internationale (CEI), qui est en charge de la normalisation et coordination des réseaux électriques intelligents. Ce groupe a pour but de rédiger ou modifier des normes pour répondre aux besoins identifiés des Smart Grids. Enfin, sur le plan de la sécurité des données, la France est impliquée dans le groupe de travail JTC1 de l'ISO/IEC.

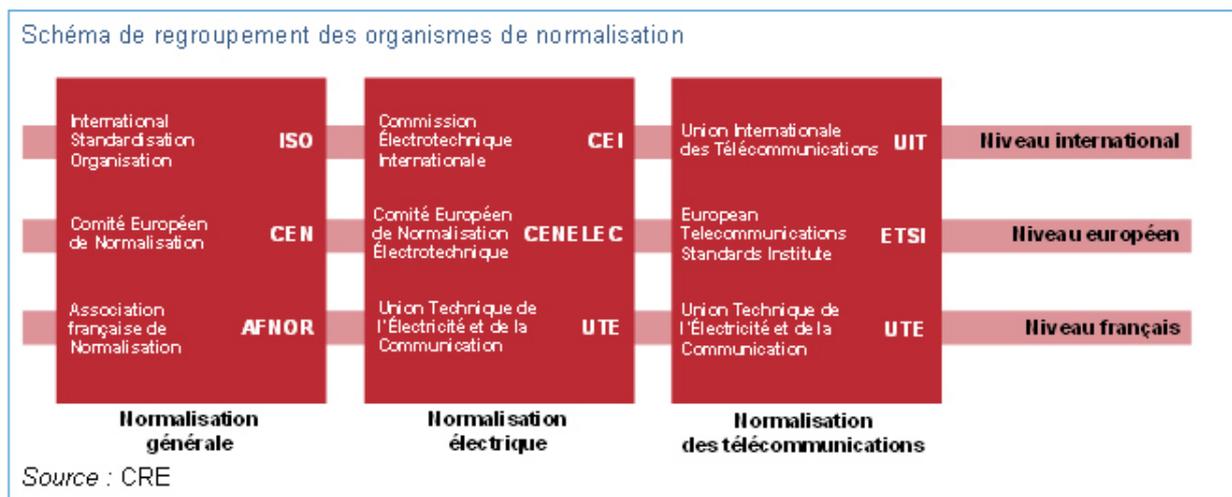


FIGURE 1.17 – Les différents organismes de normalisation.

La troisième difficulté concerne le choix des équipements à intégrer sur le réseau. Quel type de compteur intelligent choisir face au choix opulent sur le marché ? Quel réseau utiliser pour la communication des données (courants porteurs en ligne, réseau wifi...) ?

Enfin, le dernier point concerne l'évolution de la grille actuelle sans impacter les consommateurs. Comment changer radicalement un réseau sans occasionner de coupure de courant ? En effet, nous allons passer d'un réseau centralisé et unidirectionnel du producteur vers le consommateur à un réseau réparti bidirectionnel. C'est un changement sans précédent qui n'est pas réalisable ni optimal en renforçant uniquement le réseau

actuel. Pour le mettre en place de manière efficace, il faut intégrer les technologies de l'information et de la communication (TIC).

### Gestion des données et stockage

Dans le contexte de l'évolution de la grille actuelle vers un Smart Grid, les données vont devenir la clé du système. Avec l'avènement du Smart Grid, les dispositifs de compteurs intelligents, et autres capteurs, vont générer une grande quantité de données. Ces données vont être échangées continuellement sur le réseau, et il va être obligatoire de trouver des solutions de stockage efficaces.

En effet, lors des processus de décision, traiter les informations de manière rapide et efficace va faire toute la différence entre un système fiable et un système inutile.

#### 1.6.2 Enjeux économiques tarifaires et d'efficacité du réseau

Le Smart Grid implique également plusieurs enjeux économiques. Les pertes d'énergie dans les réseaux électriques, et en particulier lors de la distribution, représentent un coût important. En France, on estime que les pertes énergétiques représentent 5% de l'énergie consommée. Ces pertes sont dues à la longueur des lignes de distribution, aux conditions climatiques. L'évolution de la grille actuelle vers un Smart Grid est aussi le moyen d'augmenter l'efficacité du réseau.

Un des autres enjeux des Smart Grids concerne la gestion des tarifs de l'énergie. En effet, la création de tarifs spéciaux visant à réduire les pics de consommation ont eu l'effet contraire. Les pics de consommation ont été déplacés ou bien d'autres ont été créés. C'est le cas par exemple pour le tarif de périodes d'utilisations statiques (*static time-of-use*, TOU), où le prix de l'énergie est moins cher la nuit. Il a résulté la création d'un nouveau pic de consommation juste avant la fin de la période lorsque les utilisateurs souhaitent profiter du tarif bas avant sa coupure. De même, la tarification des seuils critiques (*critical peak pricing* ou CPP), appliquée sur la côte ouest des Etats-Unis, visait à couper les climatiseurs pendant les périodes critiques a créé de nouveaux pics lors de l'allumage de tous les appareils en fin de période critique. A la lumière de ces expériences, des chercheurs ont proposé un nouveau type de tarification en temps réel (*real time pricing* ou RTP). Le prix de l'énergie est alors mis à jour toutes les 30 minutes. Ce tarif associé à des agents de gestion intelligents, capable de réagir en fonction de l'état du réseau électrique et d'apprendre permettra de mieux gérer le Smart Grid.

#### 1.6.3 Enjeux environnementaux

Le climat devient un enjeu mondial. Pour pallier au réchauffement climatique l'Union Européenne a mis en place des directives en décembre 2008. Elles imposent 3 objectifs dits 3 fois 20 dans la ligne de conduite de la gestion de l'énergie à réaliser d'ici 2020 :

- L'intégration des énergies renouvelables : 20% de la consommation globale doit provenir des EnR.
- L'optimisation de l'efficacité énergétique : à production égale, la consommation doit diminuer de 20%.

- Baisse des émissions de gaz à effet de serre : les émissions de  $CO^2$  doivent être réduites de 20% minimum (par rapport aux valeurs de référence de 1990).

La France, dont ces directives sont gérées par le Grenelle, s'est également engagée à diviser par 4 les émissions de gaz à effet de serre.

Ces objectifs politiques vont donc fortement influencer l'utilisation et la gestion des énergies en Europe.

#### 1.6.4 Enjeux d'optimisation de la gestion électrique

Il est nécessaire d'optimiser l'intégration des énergies renouvelables, mais aussi assurer une gestion globale. Les enjeux d'optimisation visent à gérer plus efficacement l'équilibre entre l'offre et la demande, mais également de mettre en place des stratégies de diminution de la consommation globale. (effacement de la pointe, report de charge... )

Les véhicules électriques vont jouer un rôle important dans cette optique. Ils vont non seulement contribuer à diminuer les émissions de gaz à effet de serre en remplacement des véhicules traditionnels, mais vont servir de stockage temporaire des surplus de production d'énergie grâce à leur batterie. Ils peuvent ainsi participer à des lissages de courbes en restituant leur énergie lors des pics de consommation, et stockant en heure creuse. Ils représenteront donc une source d'énergie distribuée.

Nous venons de voir que le Smart Grid est un parfait exemple de système socio-technique complexe.

Dans le chapitre suivant nous présentons une algorithmique innovante pour traiter et analyser de tels systèmes.

## Chapitre 2

# Réseaux socio-techniques complexes : quelle algorithmique innovante ?

Dans ce chapitre, nous présentons les méthodes possibles et les approches algorithmiques pour traiter et étudier les systèmes socio-techniques complexes par l'appariement d'entités selon des critères voulus. En effet, l'analyse de tels systèmes requiert des outils prenant en compte les dimensions pratiques et opérationnelles du système étudié, mais également ses dimensions abstraites et formelles.

Le but de ce chapitre est de présenter une réflexion autour des théories adaptées aux problèmes pratiques des systèmes socio-techniques, qui mélangent données quantitatives et qualitatives. Nous présentons notre réflexion autour des espaces topologiques, puis une analyse plus générale et représentative des systèmes par la théorie des graphes. Enfin, nous terminons par la présentation d'une algorithmique innovante dans les réseaux complexes afin d'identifier des motifs - appelés également *patterns* - et de dimensionner les zones d'influence d'un groupe de sommets. Ce chapitre se termine par la discussion de nos résultats, et par une ouverture sur la prétopologie, une théorie permettant l'inclusion de critères plus généraux et abstraits.

### Sommaire

---

<b>2.1</b>	<b>Les défis de représentation et d'analyse des systèmes socio-techniques . . . . .</b>	<b>31</b>
<b>2.2</b>	<b>Espaces topologiques et théorie des graphes . . . . .</b>	<b>31</b>
	Définitions . . . . .	32
	Connexités . . . . .	32
<b>2.3</b>	<b>Analyse des réseaux d'influence dans les systèmes socio-techniques</b>	<b>32</b>
2.3.1	Définition de communauté . . . . .	33
2.3.2	Notre méthode . . . . .	33
2.3.3	Résultats . . . . .	36
	Méthodes d'évaluations . . . . .	37
	Information mutuelle normalisée [?] . . . . .	37
	Jaccard Index [?] . . . . .	38
	Adjusted Rand Index [?] . . . . .	38
	F-measure [?] . . . . .	38

Résultats sur des graphes générés . . . . .	38
Graphe standard . . . . .	39
Graphe avec communautés mal définies . . . . .	39
Graphe clairsemé . . . . .	39
Discussion autour de la complexité réelle . . . . .	40
Graphes générés . . . . .	40
Réseaux réels . . . . .	40
<b>2.4 Extension à la prétopologie . . . . .</b>	<b>41</b>

---

## 2.1 Les défis de représentation et d'analyse des systèmes socio-techniques

Les systèmes socio-techniques complexes, tels que décrits en chapitre 1, présentent de nombreux défis de représentation, notamment pour l'expression mathématique du système.

Il faut représenter les entités du système, avec leurs caractéristiques propres, ainsi que les interactions existantes. Une des représentations les plus évidentes, est de modéliser le système par un ensemble d'éléments associé à une définition des relations entre entités. Pour cela, nous avons étudié d'une part les fonctions multivoques, et d'autre part la théorie des graphes.

Nous présentons par la suite notre réflexion à partir des espaces topologiques.

## 2.2 Espaces topologiques et théorie des graphes

Dans cette section, nous présentons notre réflexion à partir de l'étude du livre de Claude Berge : *espaces topologiques et fonction multivoques* [?]. En effet, nous avons débuté notre recherche par l'étude des espaces topologiques, puis au fil de la lecture, nous avons choisi d'utiliser les graphes qui se sont avérés plus adaptés.

Selon Berge, un espace topologique est constitué d'un ensemble  $X$  donné, assorti d'une fonction numérique appelée *distance* ou *métrique* et notée  $d(x, y)$ , définie sur  $X * X$  telle que :

1.  $d(x, y) > 0$
2.  $d(x, y) = 0 \Rightarrow x = y$
3.  $d(x, y) = d(y, x)$
4.  $d(x, y) + d(y, z) > d(x, z)$

Les fonctions multivoques, appelées également *fonctions multivaluées*, *multiformes* ou *multifonctions*, définissent une relation qui à un élément d'un ensemble, associe un ou plusieurs éléments d'un autre ensemble.

Ce livre, "*Espaces topologiques et fonctions multivoques*", évoque également dans le chapitre 2 la définition des graphes [?] :

On appelle graphe, le couple formé par un ensemble  $X$  et une application  $\Gamma$  de  $X$  dans  $X$  ; si deux éléments  $x_1$  et  $x_2$  de  $X$  sont tels que  $x_2 \in \Gamma(x_1)$ , on dit que  $x_2$  est lié à  $x_1$  par la relation  $\Gamma$ , ce que l'on représente en joignant ces deux points par une ligne orientée de  $x_1$  vers  $x_2$ .

Les espaces socio-complexes mettant en jeu des acteurs divergents, leur dynamique et leur structure se rapprochent plus de la définition ci-dessus. Nous avons donc étudié d'avantage l'univers des graphes tels que définis par Berge, afin de retrouver les différentes propriétés et notations de connexités que nous avons jugées intéressantes dans le cadre de notre problématique.

**Définitions** Un graphe  $G = (V, A)$  est une paire composée de [?] :

1. un ensemble  $V = \{x_1, x_2, \dots, x_n\}$  appelé *sommets* (en référence aux polyèdres) ou *noeuds* (en référence à la loi des noeuds).
2. une famille  $A = (a_1, a_2, \dots, a_n)$  d'éléments du produit Cartésien  $V \times V = \{(x, y) / x \in V, y \in V\}$  appelés *arcs* (cas d'un graphe orienté) ou *arêtes* (cas d'un graphe non orienté).

En général, on note  $n$  le nombre de noeuds (aussi noté  $|V(G)|$ ) et  $m$  le nombre d'arcs (aussi noté  $|A(G)|$ ). Un chemin  $P$  est composé de  $k$  arcs tels que  $P = (a_1, a_2, \dots, a_i, \dots, a_k)$  où pour chaque arc  $a_i$ , la fin coïncide avec le début de  $a_{i+1}$ . Plusieurs notations équivalentes peuvent être utilisées :  $P = ((x_1, x_2), (x_2, x_3), \dots) = [x_1, x_2, \dots, x_k, x_{k+1}] = P[x_1, x_{k+1}]$ . Une chaîne est l'équivalent d'un chemin dans le cadre non orienté.

Un graphe peut être représenté par des matrices :

- la matrice d'incidence représente les relations entre arêtes et sommets, dites relations d'incidence.
- la matrice d'adjacence, qui représente les relations d'adjacences (si deux sommets sont reliés par une arête ils sont adjacents), définie par :  $a_{ij} = \begin{cases} 1 & \text{si } (x_i, x_j) \in E \\ 0 & \text{sinon.} \end{cases}$

**Connexités** En théorie des graphes, plusieurs connexités sont définies :

- Un ensemble  $E$  est fortement connexe (SCC) s'il existe un chemin entre tout élément de  $E$ .
- Un ensemble  $E$  est unilatéralement connexe (UCC), si pour tout  $a$  et  $b$  de  $E$ , il existe un chemin de  $a$  vers  $b$  ou il existe un chemin de  $b$  vers  $a$ .
- Un ensemble  $E$  est faiblement connexe (WCC) s'il existe une chaîne entre tout élément de  $E$ .

L'étude des Smart Grids nécessite de savoir dimensionner les zones d'études. Une décomposition en communautés représentatives de taille définie est donc parfois nécessaire. Nous proposons un algorithme de détection de communauté basé sur les connexités des graphes. Nous présentons une application du cas des composants fortement connexes appelés *composant fortement  $p$ -connecté* par Wasserman [?] qui est lié à la notion de  *$l$ -edge-connectivity* [?] (nous utilisons la notation de  *$p$ -connecté* notation au lieu de  *$n$ -connecté* pour éviter les confusions) :

- un *composant fortement  $p$ -connecté*  $p$ -SCC d'un digraphe est un sous-graphe où :  $\forall x, y \in p\text{-SCC}$ , il y a un chemin de longueur  $p$  ou moins entre  $x$  et  $y$ , et il y a un chemin de longueur  $p$  ou moins entre  $y$  et  $x$ , avec  $p \geq 2$ .

Cette définition peut être facilement généralisable aux autres types de connexités.

## 2.3 Analyse des réseaux d'influence dans les systèmes socio-techniques

Dans le cadre des systèmes socio-techniques, la détermination des communautés peut s'apparenter à une recherche de réseaux d'influence. Les notations de connexités de-

viennent alors indicateurs d'influence dans les réseaux. Nous proposons ici une algorithmique innovante pour appréhender de tels systèmes.

### 2.3.1 Définition de communauté

En ce qui concerne les réseaux non orientés, beaucoup de travaux ont été fait principalement dans le partitionnement de graphe [?, ?], le clustering [?, ?, ?, ?, ?] et les marches aléatoires [?, ?, ?, ?, ?]. Certains de ces algorithmes ont été adaptés aux réseaux orientés avec des degrés variables de succès [?].

Si la définition d'une communauté varie selon le domaine d'application [?], les communautés semblent être des ensembles relativement similaires de nœuds fortement interdépendants et plus faiblement associés avec le reste du réseau. On peut aussi imaginer une communauté en terme de réseau de communication dense. Sur la base de cette idée, on peut définir une communauté comme un groupe d'éléments où chaque élément peut fortement communiquer avec d'autres éléments : cela reflète le fait qu'il y a toujours un flux bidirectionnel qui connecte deux éléments d'une communauté. Cette vision peut être liée à la définition des composantes fortement connexes (SCC) que nous avons en théorie des graphes. Comme les SCC d'un graphe sont maximales, ces composantes ne sont souvent pas assez précises pour faire apparaître de véritables communautés.

Nous répondons à ce problème en utilisant des composantes fortement p-connexe (p-SCC) comme des subdivisions de la première SCC. Nous pouvons établir une classification hiérarchique des composantes correspondant à une classification hiérarchique des communautés (Fig. 2.1). Ainsi, nous considérons la détection de communautés comme la découverte de p-SCC.



FIGURE 2.1 – Classification hiérarchique des composants.

### 2.3.2 Notre méthode

L'une des méthodes les plus connues pour trouver les composantes SCC est l'algorithme de Tarjan [?] qui est basée sur un parcours en profondeur (DFS). Comme nous recherchons les composantes p-SCC, l'exploration du graphe est délimitée par  $p$  (la longueur maximale du chemin dans la composante) en terme d'arêtes. Cette opération est réalisée par un parcours en largeur (BFS). Notre méthode de calcul des p-SCC peut être résumée comme suit :

1. Choisir un sommet de départ au hasard  $x_s$ .
2. Réaliser un parcours BFS commençant à  $x_s$ .
3. Si, durant le parcours BFS, un circuit avec  $x_s$  est trouvé, tous les noeuds du circuit sont ajoutés au composant courant.
4. Arrêter le parcours BFS si la longueur du chemin courant atteint  $p$ , et ajouter le composant courant à la liste des p-SCC.

5. Retirer du graphe les derniers noeuds ajoutés au composant.
6. Retourner à l'étape 1 tant que le graphe n'est pas vide.

La méthode utilise une liste FIFO (i.e., First In, First Out) comme dans un parcours standard BFS. Un exemple d'exécution de notre algorithme est donné par la figure 2.2.

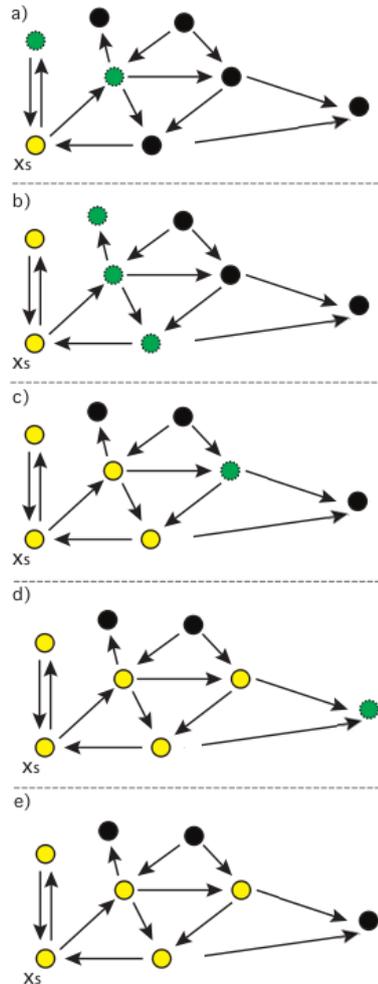
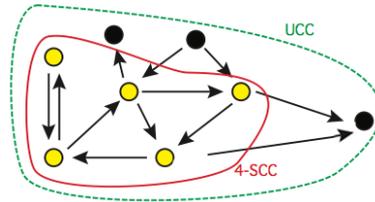


FIGURE 2.2 – Exemple d'exécution de notre algorithme avec  $p = 4$ .

Dans la figure 2.2, nous recherchons les 4-SCC. A l'étape *a*), le voisinage de  $x_s$  est exploré. Ces noeuds sont contenus dans des chemins qui sont des potentiels circuits (noeuds en vert). A l'étape suivante *b*), nous trouvons un circuit de longueur 2. L'étape *c*) découvre un circuit de longueur 3 et désélectionne une feuille. L'étape *d*) trouve un circuit de longueur 4. L'étape *e*) termine l'algorithme car nous atteignons un chemin de longueur maximale  $p$ .

Comme notre méthode démarre son exploration depuis un sommet aléatoire, les résultats peuvent être différents d'une exécution à l'autre, mais cela s'avère être un détail peu important comme nous le prouverons dans la partie démonstration. Un problème plus important concerne la discontinuité des sommets : l'algorithme peut fournir des communautés-singletons. Dans le précédent exemple de la figure 2.2 à l'étape *e*), trois

FIGURE 2.3 – Phase de suppression des communautés trop petites : p-SCC  $\rightarrow$  UCC.

noeuds restent isolés. Pour résoudre ce problème, une phase finale regroupe les communautés ayant des tailles trop petites (définies selon un paramètre précis). En utilisant nos communautés p-SCC comme communautés de base, nous associons toutes les "petites" communautés à de plus grandes, s'il y a un arc en commun entre elles, ce qui étend certains des p-SCC à UCC (Fig. 2.3). Cette phase réduit les différences d'une exécution à l'autre liées à la sélection aléatoire du sommet de départ.

Le premier algorithme (Alg. 3) trouve une composante p-SCC en commençant par un noeud donné  $x_s$ . Cette méthode est appelée tant que le graphe n'est pas vide (Alg. 4) et l'opération d'association est alors appliquée sur la composante p-SCC trouvée (Alg. 5).

### Algorithm 3

```

Method pSCC(graph G, vertex  $x_s$ , integer p)
/* C :set, paths :list, P :path, A(x) :neighborhoods of x */
C  $\leftarrow$   $\emptyset$ 
paths  $\leftarrow$   $\emptyset$ 
paths.add([ $x_s$ ])
Tant que paths  $\neq$   $\emptyset$  Faire
  P  $\leftarrow$  paths.removeFirstPath()
  Si  $x_s = P.x_{k+1}$  Alors
    C  $\leftarrow$  C  $\cup$  P
  Sinon
    Pour each y  $\in$  A( $P.x_{k+1}$ ) Faire
      PP  $\leftarrow$  P
      PP.add(y)
      Si |PP|  $\leq$  p + 1 Alors
        paths.add(PP)
      Fin Si
    Fin Pour
  Fin Si
Fin Tant que

```

### Algorithm 4

```

Method AllpSCC(graph G, integer p)
/* C :set, SCC :family of sets */
SCC  $\leftarrow$   $\emptyset$ 
Tant que V(G)  $\neq$   $\emptyset$  Faire
  x  $\leftarrow$  random node of V(G)
  C  $\leftarrow$  pSCC(G, x, p)
  V(G)  $\leftarrow$  V(G) - C
  SCC  $\leftarrow$  SCC  $\cup$  C
Fin Tant que
Retourner SCC

```

### Algorithm 5

```

Method MergeSCC(graph G ,family SCC, integer minCsize)
/* C,smallC :sets, smallSCC :family of sets, A(x) :neighborhoods of x */
smallSCC ← ∅
Pour each C ∈ SCC Faire
  Si |C| ≤ minCsize Alors
    smallCC ← smallCC ∪ C
  Fin Si
Fin Pour
SCC ← SCC − smallSCC
Tant que smallSCC ≠ ∅ Faire
  smallC ← smallSCC.remove()
  Pour each x ∈ smallC Faire
    Pour each C ∈ SCC Faire
      Si A(x) ∩ A(C) Alors
        C ← C ∪ {x}
      Fin Si
    Fin Pour
  Fin Pour
Fin Tant que
Retourner SCC

```

Note : si le graphe n'est pas faiblement connexe, nous appliquons notre méthode sur chaque composant faiblement connexe du graphe.

Pour l'instant, la complexité de notre algorithme est exponentielle selon la longueur maximale des chemins. Le pire des cas est obtenu en l'appliquant sur un graphe complet  $K_n$  (chaque paire de noeuds  $(x, y)$  a un arc de  $x$  à  $y$  et de  $y$  à  $x$ ). Comme la méthode  $pSCC$  garde en mémoire tous les chemins courants, en partant d'un noeud aléatoire, nous avons  $n - 1$  chemins initiaux. Nous explorons alors exhaustivement les voisinages de chaque fin de chemin. Le nombre d'opérations est égal à :  $(n - 1)^p$ . Le pire des cas en temps de complexité peut être écrit  $O(n^p)$  avec  $p$  la longueur maximum des chemins. Bien que la complexité soit exponentielle, elle est paramétrable par la longueur maximum des chemins. Elle peut être reliée à l'algorithme de Tarjan [?] qui énumère tous les circuits dans un digraphe avec une complexité temporelle de  $O((n \cdot m)(c + 1))$  ( $c$  est le nombre de circuits élémentaires et est exponentiel). Dans les réseaux réels, le pire des cas ne devrait pas se rencontrer car ils sont habituellement clairsemés [?], et nos expériences montrent que la meilleure valeur de  $p$  est 4 avec une complexité expérimentale faible. La complexité spatiale suit les mêmes règles.

### 2.3.3 Résultats

Pour tester l'efficacité de notre méthode, nous utilisons des références dédiées à la détection de communauté. Les générateurs de graphe ont été introduits pour autoriser les tests complexes de graphes réalistes non-orientés [?] et orientés [?] avec un ensemble de communautés de référence. Cependant, ces générateurs de graphes ne prétendent pas modéliser des phénomènes résultants de réseaux réels, ils tentent seulement d'imiter quelques unes de leurs propriétés en imposant une structure à la communauté. Nous réalisons que la structure des communautés des réseaux réels est certainement plus complexe que celle des graphes générés. Le problème est qu'il n'y a pas de grand graphe réel pour lequel nous connaissons le découpage en communautés. Donc nous testons notre méthode sur des graphes orientés générés, et nous estimons le temps pratique d'exécution sur des ensembles réels.

La mesure de similarité la plus commune de la littérature pour comparer deux partitions est l'indice NMI (Information Mutuelle Normalisée) [?] qui a été généralisé aux partitions recouvrantes (GNMI) par [?]. Cette mesure est très proche du standard NMI même lorsque les communautés ne sont pas recouvrantes. Cependant, l'utilisation d'un unique test de mesure n'est pas suffisant. L'évaluation d'une méthode de classification est un champs important du domaine de l'apprentissage automatique, et un grand nombre de méthodes d'évaluation différentes pour comparer 2 classifications existent. En plus de NMI, nous avons mesuré la performance de notre algorithme de recherche de communautés à l'aide de l'index Jaccard ajusté, l'index Jaccard, et la F-mesure. Nous comparons nos scores à la méthode de classification InfoMap [?] qui donne de bons résultats dans les graphes non-orientés [?] et orientés [?]. Bien entendu, nous nous comparons uniquement à la version orientée de leur algorithme.

### Méthodes d'évaluations

Dans une méthode d'évaluation de classification, il existe une correspondance entre les communautés de références et celles prédites, exprimée par la *matrice de confusion*. La matrice de confusion peut être considérée comme la liste des fréquences d'occurrences communes entre les communautés de références et celles prédites.

		Communauté prédite	
		Inclus	Non inclus
Communauté de référence	Inclus	95	5
	Non inclus	3	97

TABLE 2.1 – Matrice de confusion.

Dans l'exemple 2.1, chaque colonne représente le nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle. Ainsi, on lit sur cet exemple que sur 100 sommets de la communauté de référence, 95 ont été inclus dans la communauté estimés, mais 5 n'ont pas été inclus. Par contre, la communauté estimée comprend 3 sommets de plus qui n'auraient pas dus être inclus selon la référence.

Cette matrice sert à estimer des statistiques telles que l'entropies marginale, jointe et conditionnelle, ou encore l'information mutuelle.

Nous utilisons les notations suivantes :  $I(X;Y)$  est l'information mutuelle de  $X$  et  $Y$  avec  $I(X;Y) = H(X) - H(X|Y)$ ,  $H(X)$  et  $H(Y)$  sont les entropies marginales,  $H(X|Y)$  et  $H(Y|X)$  sont les entropies conditionnelles, et  $H(X,Y)$  est l'entropie jointe de  $X$  et  $Y$  qui sont deux variables discrètes aléatoires. Toutes les mesures sont entre 0 (pire des cas) et 1 (meilleur des cas).

**Information mutuelle normalisée [?]** L'information mutuelle normalisée est définie comme l'information mutuelle entre l'affectation des clusters et un marquage pré-existant de l'ensemble de données normalisées par :

$$NMI(X,Y) = \frac{I(X;Y)}{\sqrt{H(X)H(Y)}}$$

**Jaccard Index [?]** L'index Jaccard est utilisé en statistiques pour comparer la similarité et la diversité entre des échantillons. Etant donnés 2 ensembles  $A$  et  $B$ , l'index Jaccard est exprimé comme suit :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Une interprétation par la théorie des ensembles montre que  $I(X; Y)/H(X, Y)$  est équivalent à l'index Jaccard.

**Adjusted Rand Index [?]** La formule ajustée de l'index Rand, dite *Adjusted Rand Index*, qui est utilisée pour augmenter la sensibilité de l'index, est définie par :

$$AdjustedIndex = \frac{Index - ExpectedIndex}{MaxIndex - ExpectedIndex}$$

Elle peut être liée à la mesure ajustée pour l'information mutuelle :

$$AI(X; Y) = \frac{I(X; Y) - E\{I(X; Y)\}}{\max\{H(X), H(Y)\} - E\{I(X; Y)\}}$$

**F-measure [?]** La formule pour la F-mesure est définie dans [?]. Le rappel, la précision et la F-mesure pour une classe de référence  $K_i$  et une communauté prédite  $C_j$  sont définis par :  $Rappel(K_i, C_j) = n_{ij}/|K_i|$ ,  $Precision(K_i, C_j) = n_{ij}/|C_j|$  et :

$$F(K_i; C_j) = \frac{2 \times Rappel(K_i, C_j) \times Precision(K_i, C_j)}{Rappel(K_i, C_j) + Precision(K_i, C_j)}$$

où  $n_{ij}$  correspond à une cellule de la matrice de confusion.

## Résultats sur des graphes générés

Nos tests ont été effectués sur plusieurs digraphes générés par [?] et ayant différents paramètres :

- $n$  le nombre de noeuds du graphe
- $k_{avg}$  le degré moyen des noeuds,  $k_{max}$  le degré maximum des noeuds
- $min_c, max_c$  tailles min. et max. des communautés
- $\mu$  le mixing parameter *i.e.* pour chaque vertex le ratio entre le nombre d'arêtes sortantes et entrantes de chaque communauté.

Comme réalisé dans [?], nous changeons un critère pour chaque test (degré, mixing parameter, et taille minimale de communauté), mais la différence est que nous avons réalisé nos tests sur des graphes de tailles différentes, de  $n = 1000$  à  $n = 10000$ . Chaque score correspond à une moyenne de 50 générations de graphes. Les paramètres fixes utilisés pour notre méthode sont  $p = 4$  (longueur maximale des chemins) et  $minCsize = 3$  (seuil minimal pour la taille des communautés) qui s'avèrent donner de meilleurs résultats sur tout type de graphe. Tous nos tests utilisent la librairie graphique java JUNG [?] et ont été lancé sur un ordinateur portable Macbook pro. Notre méthode est notée *Méthode LP* (initiales Levorato - Petermann) dans les tests suivants.

**Graphe standard** Nous paramétrons le générateur de graphes avec les valeurs de référence suivantes :  $k_{avg} = 15$ ,  $k_{max} = 50$ ,  $min_c = 20$ ,  $max_c = 50$ ,  $\mu = 0.1$ . Les résultats de mesures sont notés dans le tableau 2.2. Il apparait que notre algorithme donne toujours de bons résultats en augmentant le nombre de noeuds du graphe.

Algorithme	Mesures	Nombre de noeuds					
		1000	2000	3000	4000	5000	10000
Méthode LP	ARI	0.95	0.97	0.98	0.98	0.98	0.99
	Jaccard	0.91	0.95	0.96	0.96	0.97	0.98
	F-Mesure	0.97	0.98	0.99	0.99	0.99	0.99
	NMI	0.93	0.95	0.97	0.97	0.98	0.99
InfoMap	ARI	0.62	0.64	0.65	0.67	0.67	0.69
	Jaccard	0.47	0.48	0.49	0.50	0.50	0.53
	F-Mesure	0.69	0.71	0.72	0.74	0.75	0.77
	NMI	0.41	0.46	0.48	0.50	0.51	0.52

TABLE 2.2 – Test d'efficacité en augmentant  $n$ .

**Graphe avec communautés mal définies** Nous avons établi les paramètres du générateur de graphes comme valeurs de graphes standards et avec  $n = 5000$  en ne changeant que le mixing parameter de  $\mu = 0.1$  à  $\mu = 0.5$ . Les résultats des mesures figurent dans le tableau 2.3. L'efficacité de l'algorithme diminue avec l'augmentation de  $\mu$ , ce qui a du sens car chaque noeud a beaucoup de liens à l'intérieur et à l'extérieur de sa propre communauté, et il devient donc de plus en plus difficile de discerner les communautés.

Algorithme	Mesures	Mixing parameter				
		0.1	0.2	0.3	0.4	0.5
Méthode LP	ARI	0.98	0.92	0.79	0.59	0.33
	Jaccard	0.97	0.86	0.66	0.43	0.20
	F-Mesure	0.99	0.96	0.89	0.78	0.62
	NMI	0.98	0.92	0.81	0.66	0.45
InfoMap	ARI	0.67	0.48	0.33	0.19	0.13
	Jaccard	0.50	0.32	0.20	0.11	0.07
	F-Mesure	0.75	0.62	0.51	0.38	0.31
	NMI	0.51	0.30	0.15	0.03	0.00

TABLE 2.3 – Efficacité de notre méthode et augmentation du paramètre  $\mu$ .

**Graphe clairsemé** Nous avons établi les paramètres du générateur de graphes comme valeurs de graphes standards en ne changeant que le degré moyen des nœuds  $k_{avg} = 7$  et  $n = 5000$ . Les résultats des mesures figurent dans le tableau 2.4. Les paramètres du graphe généré donnent un nombre moyen de 155 communautés. En moyenne, on trouve le même nombre de communautés avec  $minCsize = 10$  et avons le meilleur score NMI. Les autres mesures diminuent lorsque  $minCsize$  et le score NMI augmentent. On peut voir ici l'intérêt d'utiliser plusieurs mesures d'évaluation : dans les autres tests, toutes les mesures sont plus ou moins corrélées alors que dans le cas du graphe clairsemé, le choix de la taille minimale d'une communauté est important.

	ARI	Jaccard	F-Mesure	NMI	C
LP ( <i>minCsize</i> = 3)	0.68	0.51	0.76	0.56	≈ 260
LP ( <i>minCsize</i> = 5)	0.60	0.43	0.73	0.57	≈ 195
LP ( <i>minCsize</i> = 10)	0.48	0.39	0.69	0.62	≈ <b>155</b>
LP ( <i>minCsize</i> = 15)	0.23	0.13	0.62	0.61	≈ 130
InfoMap	0.59	0.42	0.64	0.37	≈ 265

TABLE 2.4 – Test d'efficacité sur graphes clairsemés.

## Discussion autour de la complexité réelle

**Graphes générés** Les graphes générés pour nos tests sont composés de 1000 à 250 000 noeuds. Les résultats sont affichés dans le tableau 2.5, et montrent que notre algorithme a de bonnes performances même avec 250 000 noeuds (3 min 20 s).

Noeuds	1000	5000	10 000	50 000	100 000	250 000
Temps	2s	4s	7s	36s	1min 10s	3min 20s

TABLE 2.5 – Complexité temporelle pratique pour les graphes générés.

**Réseaux réels** Les ensembles de données<sup>1</sup> utilisés sont :

- *Réseau social Epinions* : C'est un réseau social de "qui-a-confiance-en-qui" provenant du site d'avis consommateurs *Epinions.com* avec 75 879 noeuds et 508 837 arcs.
- *Réseau des communications email EU* : Le réseau a été généré en utilisant les données emails d'une grande institution européenne de recherche. Il contient 265 214 noeuds et 420 045 arcs.

Le premier réseau est traité par notre algorithme en moins de 5 minutes. Le second est traité en 58 minutes. La différence des temps de calcul peut être expliquée par l'opération finale de fusion des petites communautés. Pour *Epinions*, nous obtenons 58 026 communautés avant fusion, et 1572 après fusion. Pour le réseau d'email, nous obtenons 256 962 communautés avant fusion, et 174 après fusion, ce qui représente 99% de communautés fusionnées lors de l'étape finale de notre algorithme.

Cette observation est corrélée aux résultats que nous obtenons dans les graphes clairsemés. La densité du réseau d'email est plus petite que celle du réseau *Epinions*, et le choix du seuil concernant la taille des plus petites communautés est primordial.

Notre approche permet donc la détection de communautés en utilisant les composantes p-connexes. Elle donne de très bons résultats sur les graphes générés, et a un ratio qualité / temps d'exécution acceptable. Notre contribution est un algorithme simple et efficace pour détecter les communautés en utilisant les propriétés des graphes fortement et unilatéralement connexes. Elle peut être appliquée directement sur tout type de critères quantitatifs, cependant, dans le cadre des réseaux complexes, nous sommes confrontés à de nombreux critères qualitatifs. Nous envisageons donc d'étendre nos travaux à la prétopologie, qui peut être vue comme une extension de la théorie des graphes.

1. source : Stanford Large Network Dataset Collection (<http://snap.stanford.edu/data/>)

## 2.4 Extension à la prétopologie

Dans le cadre des Smart Grids, nous avons besoin d'algorithmes innovants pour la structure et l'analyse de tels systèmes. Notre méthode peut ainsi contribuer à classer les entités du réseau par niveau de consommation ou de production proches, et ainsi dresser des profils d'utilisation de l'énergie par communauté. Autre application possible, la détection des réseaux d'influence peut fournir des informations précises pour la mise en place des stratégies de gestion des pannes.

Cependant, la connexité géographique n'est pas forcément respectée. Nous proposons donc une extension de notre méthode par la prétopologie car cette théorie permet une représentation de relations abstraites et construit une topologie sur les graphes. Elle est donc vue comme une extension de la théorie des graphes. C'est pourquoi la majorité des algorithmes issus des graphes peuvent être implémentés par la prétopologie. Cette théorie sera présentée en détail dans le chapitre suivant. L'usage de la prétopologie permet de définir les limites de l'influence d'une communauté par la fonction d'ouverture, tandis que la fonction de fermeture définit les influenceurs. La prétopologie structure ainsi les notions d'influence sortante et entrante.

Par exemple, si la grille détecte une insuffisance en fourniture électrique sur une zone, elle pourrait faire appel à notre méthode pour définir jusqu'à quelle excentricité elle peut compter sur les producteurs d'énergies renouvelables.

Cependant, cette partie relève d'optimisations dans un travail futur car l'inclusion des critères qualitatifs du Smart Grid pour la détection de réseaux d'influence nécessite des données précises que nous n'avons pas à disposition pour le moment.



## Chapitre 3

# Modélisation intégrative de Smart Grids

Dans ce chapitre, nous proposons une modélisation intégrative pour les Smart Grids, c'est-à-dire un modèle dont chaque composante permet de prendre en compte un aspect du Smart Grid et dont l'intrication réalise une modélisation efficace du système socio-technique étudié. Le concept de flux d'énergie devient ici multi-niveaux et nécessite des théories adaptées pour le modéliser, définir un voisinage et pour réguler entre niveaux de façon optimisée.

### Sommaire

<b>3.1</b>	<b>Introduction</b>	<b>45</b>
<b>3.2</b>	<b>Modélisation du Smart Grid par les multi-agents</b>	<b>45</b>
3.2.1	Pourquoi utiliser une modélisation par multi-agents?	45
	Répondre efficacement à la demande en temps réel	46
	Défaillance du Smart Grid et robustesse des SMA	46
	Flexibilité et évolutivité	46
	Réseau réel, qualité de service et expérimentations	46
	Intégrativité et évolution du modèle	46
	Des acteurs-agents hétérogènes pour mieux représenter la réalité	46
3.2.2	Exemples de modélisation de systèmes complexes	47
<b>3.3</b>	<b>Négociations autour de l'énergie : une modélisation à travers la théorie des jeux</b>	<b>48</b>
3.3.1	Jeux coopératifs	48
3.3.2	Jeux de stratégies	48
3.3.3	Jeux d'actions	49
3.3.4	Jeux à mémoire parfaite et imparfaite	50
3.3.5	Jeux séquentiels et jeux simultanés	50
3.3.6	Approche bottom-up et top-down par les réseaux de jeux pour le partage de l'énergie	50
<b>3.4</b>	<b>Résilience et dynamique du réseau par la prétopologie</b>	<b>51</b>

3.4.1	Rappels et définitions . . . . .	51
3.4.2	Famille de prétopologies pour la représentation des différents critères et leurs intrications . . . . .	53
3.4.3	$\chi$ -connexités en prétopologie pour analyser la connectivité du réseau . . . . .	56
	Fortement connexe . . . . .	56
	Unilatéralement connexe . . . . .	56
	Hyperconnexe . . . . .	56
	Apo-connexe . . . . .	56
	Connexe . . . . .	57
<b>3.5</b>	<b>Analyse de la propagation de l'énergie et percolation . . . . .</b>	<b>57</b>
3.5.1	Détermination du seuil de connectivité assurant l'efficacité du réseau . . . . .	58
3.5.2	Prévention des pannes et analyse du phénomène de brownout . . . . .	59
<b>3.6</b>	<b>Déroulement de notre modèle . . . . .</b>	<b>60</b>

---

## 3.1 Introduction

La modélisation est une discipline visant à expliciter les interactions d'une partie d'un système, et du système dans sa globalité. Il est important de bien distinguer la phase de modélisation de celle de simulation. Un modèle est une représentation mathématique d'un système en un point particulier du temps ou de l'espace, en vue d'aider à comprendre le système réel. Une simulation est une manipulation du modèle dans le but d'analyser les interactions émergentes du modèle, qui n'auraient pas pu être connues par la simple connaissance du modèle statique dans le temps ou l'espace. Une simulation devraient contenir le plus de détails possible, alors qu'un modèle doit être simplifié au maximum pour être facilement généralisable à d'autres systèmes [?].

Actuellement, les méthodes classiques de modélisation arrivent à une certaine limite : on ne peut traiter les problèmes issus des systèmes complexes que d'une manière interdisciplinaire ou avec une approche globale. Dans cette section, nous présentons les différentes théories qui permettent chacune de prendre en compte un aspect de la problématique de modélisation de systèmes complexes. Réunies ensemble de façon modulaire, elles permettent d'obtenir une modélisation plus réaliste, à l'aide de comportements d'agents mieux représentés.

## 3.2 Modélisation du Smart Grid par les multi-agents

Les systèmes multi-agents sont apparus comme un sous-domaine de l'intelligence artificielle distribuée, qui est elle-même un cas particulier de l'intelligence artificielle. De nos jours, le terme système multi-agents réfère à tout type de système composé de multiples agents (semi) autonomes [?]. Dans cette section, nous présentons en détails les principes et avantages de la modélisation multi-agents, avant de nous attarder sur quelques définitions et exemples concrets.

### 3.2.1 Pourquoi utiliser une modélisation par multi-agents ?

Un système multi-agent (SMA) est un système composé d'un ensemble d'éléments interactifs autonomes (dits agents), d'un environnement, d'un ensemble de relations entre les entités et d'un ensemble d'opérations qui permettent aux agents de percevoir, produire, consommer, transformer et manipuler les objets de l'environnement. Les agents sont un cas particulier des objets représentant les entités actives du système. Les systèmes multi-agents sont utilisés pour solutionner des problèmes difficiles voire impossibles à solutionner à l'aide d'un agent central ou d'un système monolithique.

Techniquement, un système multi-agent consiste en un ensemble de processus se déroulant en parallèle. La clé centrale de ce type de système consiste dans la coordination entre les agents. Un agent est une entité de calcul, comme un programme d'ordinateur, un humain ou un robot, qui est capable de percevoir et d'agir par elle-même dans un environnement défini.

La modélisation de système complexe est plus réaliste à l'aide d'agents intelligents interagissant ensemble car chaque agent peut être un modèle mathématique propre [?].

De plus, les modélisations multi-agents sont très efficaces, et avec un très bon potentiel grâce à la parallélisation des tâches pour des applications en temps réel. C'est pour cela que les SMA sont une des approches les plus évidentes dans le cadre de la modélisation de systèmes complexes. L'approche par les multi-agents est aussi utilisée dans beaucoup de sujets de recherche comme la modélisation de la pollution [?], les structures sociales... Les *Grids* et les agents ont par exemple beaucoup de points communs [?]. Nous allons voir que c'est aussi le cas dans le cadre des Smart Grids.

**Répondre efficacement à la demande en temps réel** En agissant de manière parallèle, l'ensemble des agents permet d'effectuer les tâches plus rapidement qu'un système monolithique avec un agent central unique. C'est aussi plus réaliste puisque les intervenants du Smart Grid agissent et interagissent en simultané.

**Défaillance du Smart Grid et robustesse des SMA** En modélisant un système complexe à l'aide d'agents distincts agissant parallèlement avec une individualisation possible des comportements, la modélisation multi-agents permet un modèle plus proche de la réalité. Les agents sont alors partiellement autonomes et dans le cas d'agents intelligents, leur comportement dépend partiellement de leur expérience. Cela permet d'augmenter la robustesse et la fiabilité du modèle.

**Flexibilité et évolutivité** Il est aisé de modifier le nombre total d'agents afin de réaliser une simulation à une autre échelle, pour changer la granularité des résultats. De plus, les agents étant programmés sous forme de modules, il est possible de les faire évoluer par l'ajout d'un module dans le modèle.

**Réseau réel, qualité de service et expérimentations** En permettant de lancer des simulations *in silico* dans un contexte informatisé, les modélisations multi-agents sont à moindre coût. Le contexte virtuel étant sans risque, il permet de tester l'impact de changements de paramètres, tout en étudiant le comportement global du système. On peut ainsi lancer plusieurs scénarios à moindre coût. C'est un des atouts majeurs des SMA. Dans le cadre des Smart Grid, il est impossible de réaliser des tests réels si ceux-ci peuvent impacter la qualité ou la continuité du service. L'utilisation d'une simulation informatique est alors optimale et sans risque.

**Intégrativité et évolution du modèle** Les modèles étant des systèmes généralisés d'un problème, programmés sous forme de module, il est possible d'en réutiliser tout ou une partie dans le cadre d'un autre problème. Il est également possible de rajouter des modules supplémentaires au modèle existant.

**Des acteurs-agents hétérogènes pour mieux représenter la réalité** Un modèle multi-agents est dit homogène si tous les agents sont construits sur le même modèle (par exemple une colonie de fourmis), et hétérogène si les agents sont construits avec différents modèles de différents niveaux hiérarchiques (par exemple l'organisation d'une société).

Dans notre cas de modélisation de Smart Grid, nous allons utiliser des agents hétérogènes pour modéliser les différents types d'entités du système. Ce système socio-technique complexe peut être décomposé en plusieurs sous-systèmes en interaction, comme présenté dans la figure 3.1, chaque sous-système comportant lui même divers types d'agents.

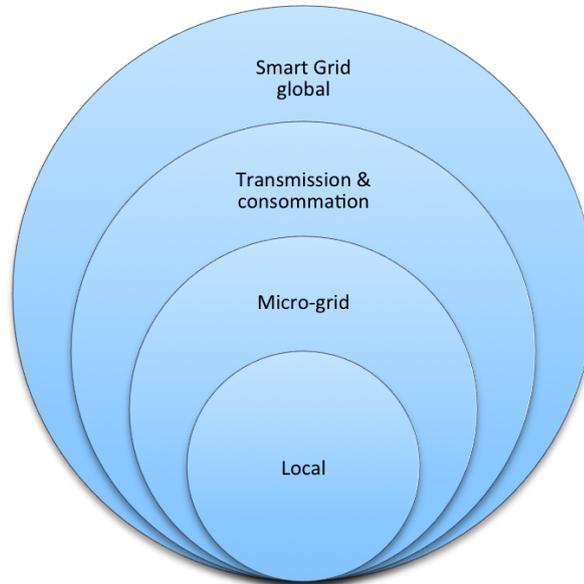


FIGURE 3.1 – Sous-systèmes identifiés du Smart Grid.

### 3.2.2 Exemples de modélisation de systèmes complexes

Dans les problèmes de partage de ressources, les modélisations multi-agents ont déjà montré leur efficacité, par exemple dans le cadre de l'utilisation des terres et le partage de l'eau dans la zone du Sahel [?].

L'approche par les multi-agents a également très souvent été utilisée dans le cadre des problèmes de fabrication et optimisation dans les chaînes des usines (dits "*manufacturing problems*") [?]. Or, il existe des similarités certaines entre ces problèmes et ceux liés à la génération d'électricité.

Cependant, la seule théorie des multi-agents ne nous suffit pas pour résoudre les systèmes de Smart Grids car il reste des problèmes théoriques non pris en compte. Dans la partie suivante, nous présentons en détail les différentes théories qui nous permettent de modéliser plus efficacement un Smart Grid. D'abord, nous présenterons la théorie des jeux, car le Smart Grid n'est autre qu'un réseau complexe avec des comportements relevant de l'optimisation et des négociations, ce qui est usuellement modélisé par la théorie des jeux. La structure du réseau et les négociations étant basées sur des notions de relations de proximité, nous présenterons ensuite la théorie des graphes, puis la prétopologie qui peut être vue comme une extension de la théorie des graphes pour les modélisations réalistes de notions de proximité. Enfin, nous détaillerons la percolation en tant que théorie permettant de valider notre modèle. Elle nous permettra de vérifier si à tout moment le réseau est correctement connecté, c'est à dire que tous les clients sont fournis en énergie.

### 3.3 Négociations autour de l'énergie : une modélisation à travers la théorie des jeux

Dans cette partie, nous présentons les idées issues de la théorie de jeux permettant de résoudre les problèmes de négociation dans la distribution de l'énergie.

Les Smart Grids impliquent des acteurs hétérogènes présentant des intérêts divergents : l'Etat, les consommateurs, les producteurs, etc. sont en perpétuel conflit d'un point de vue économique. Ce comportement se rapproche des principes de la théorie des jeux [?, ?] : c'est en effet un problème stratégique dans lequel les actions de chaque joueur influent sur le comportement des autres.

La théorie des jeux est une méthode de mathématiques appliquées utilisée pour étudier les comportements humains et animaux en modélisant les comportements liés aux interactions entre agents concurrents. Bien que cette théorie ait été initialement développée dans le cadre de l'économie pour comprendre un vaste ensemble de comportements économiques, l'utilisation de la théorie des jeux dans les sciences sociales s'est étendue, et a été appliquée aux comportements politiques, sociologiques, et psychologiques. Les applications incluent de façon non exhaustive, les modélisations de systèmes de vote [?], les comportements économiques [?], l'économie politique [?], les calculs économiques basés sur des agents [?].

La théorie des jeux peut donc être définie comme l'étude de modèles mathématiques de conflits et coopérations entre des agents intelligents. Elle fournit des techniques générales pour analyser des situations dites "jeux", où le succès d'une entité dépend des choix des autres entités du système [?].

Dans notre cas, nous utilisons la théorie des jeux pour modéliser les conflits et coopérations multi-niveaux entre les agents de notre système, et le jeu réfère à la globalité du Smart Grid.

#### 3.3.1 Jeux coopératifs

Lorsqu'un jeu comporte plus de 2 joueurs, ceux-ci peuvent choisir de coopérer afin de former des coalitions. La théorie des jeux coopératifs vise à exprimer ce genre de situation :  $n$  joueurs cherchent une distribution équitable d'un bénéfice sachant que tous les joueurs aimeraient autant que possible des avantages [?].

Dans le cadre du Smart Grid, les joueurs sont les éléments technologiques ou humains du réseau et la coopération consiste à partager la ressource électrique de manière équitable, sachant que chacun aimerait obtenir la totalité de l'énergie dont il a besoin. Les microgrids forment des coalitions, et chaque consommateur va agir au sein de sa coalition selon ses propres règles définies, on parle alors de stratégies.

#### 3.3.2 Jeux de stratégies

Au sein de chaque microgrid, des stratégies permettent de définir la distribution de l'énergie.

**Définition 3.1** *Un jeu de stratégie est défini par :*

- un ensemble  $N = 1; \dots; n$  de joueurs ;
- pour chaque joueur  $i$  un ensemble de stratégies  $S_i = s_1; \dots; s_{n_i}$  ;
- pour chaque joueur  $i$  une fonction de valuation  $\mu_i : S_1 * \dots * S_n \rightarrow R$ , qui à chaque ensemble de stratégies associe les gains du joueur  $i$ .

**Définition 3.2** On notera  $s$  un profil de stratégies  $s_1; \dots; s_n$  où  $\forall i, s_i \in S_i$ . On note  $s_{-i}$  le profil  $s$  des stratégies autres que celles du joueur  $i$  :  $s_{-i} = s_1; \dots; s_{i-1}; s_{i+1}; \dots; s_n$ . On note  $S$  l'espace des stratégies.

Un jeu est dit résoluble par élimination itérative des stratégies dominées, si on obtient un profil unique en éliminant successivement des stratégies (strictement) dominées. Tous les jeux ne sont pas résolubles par cette méthode.

**Définition 3.3** Une stratégie  $s_i$  est (strictement) dominée pour le joueur  $i$  si il existe une stratégie  $s'_i$  telle que pour tous les profils  $s_{-i} : \mu_i(s'_i; s_{-i}) > \mu_i(s_i; s_{-i})$   
 Une stratégie  $s_i$  est faiblement dominée pour le joueur  $i$  si il existe une stratégie  $s'_i$  telle que pour tous les profils  $s_{-i} : \mu_i(s'_i; s_{-i}) \geq \mu_i(s_i; s_{-i})$ .

La fonction de meilleure réponse du joueur  $i$  est la fonction  $B_i$  qui associe à chaque combinaison de stratégies des autres joueurs  $s_{-i}$  les stratégies du joueur  $i$  qui maximise son utilité. Un équilibre de Nash est un profil  $s^*$  tel que la stratégie du joueur  $i$  est une meilleure réponse :  $s^* \cdot i \in B_i(s^*_{-i})$  pour tout  $i \in N$ . Un profil (unique) obtenu par élimination itérative de stratégies (strictement) dominées est un équilibre de Nash.

**Définition 3.4** Un équilibre de Nash est un profil de stratégies  $s^* = s^*_1; \dots; s^*_n$  tel que pour tout joueur  $i$ , pour toute stratégie  $s' \in S_i : \mu_i(s^*_1; s^*_{-i}) \geq \mu_i(s'; s^*_{-i})$

Les jeux de stratégies permettent de gérer le partage de l'énergie au niveau local au sein de chaque microgrid, cependant, la sous-station responsable du microgrid est un jeu prenant en compte des paramètres économiques et sociaux. Par un système de priorisations, elle doit prévoir une attribution de l'énergie équitable. Nous utilisons alors des jeux d'enchères pour modéliser cette étape.

### 3.3.3 Jeux d'actions

Les jeux d'enchères peuvent être vus comme un dispositif pour équilibrer l'offre et la demande, ou pour fournir un mécanisme puissant pour allouer des ressources, en particulier lorsque celles-ci ont une valeur incertaine ou non standard [?].

L'analyse des jeux d'enchères peut être difficile, surtout dans un contexte stochastique, comme la fourniture et le transport de ressources. Dans le cadre de la production électrique, les producteurs et les stations électriques sont confrontés à des décisions complexes, qui découlent de paramètres économiques (coût de transport, bénéfices) mais également d'interdépendance entre les entités du système (répartition de l'offre et de la demande, politique de prix des producteurs, etc.).

Les jeux d'enchères ont été largement étudiés par les économistes, conduisant à des progrès récents dans la compréhension théorique des différents types d'enchères. Ces modèles

ont été principalement axés sur les enchères ponctuelles réalisées par des agents neutres, sans prise de risque. Cependant, dans le cas d'enchères séquentielles, devant répondre à plusieurs offres et demandes simultanées, le problème reste intraitable de manière globale [?].

Une autre source de complexité découle de la nécessité de connaître en temps réel et en chaque point du système le coût de transport de l'énergie afin d'en ajuster sa valeur. C'est un problème NP difficile<sup>1</sup>, qui ne peut généralement être résolu de manière optimale dans un environnement de taille réelle, dynamique et stochastique. Lorsque la complexité du problème des jeux d'enchères exclut la mise en œuvre d'une approche théorique globale du jeu, l'utilisation d'agents informatiques permet de simplifier ou d'adapter le problème de décision. Ces simplifications ou modifications locales pour simuler le comportement global des jeux d'enchères définissent le comportement rationnel limité [?]. Ainsi, un agent ne choisira pas forcément l'optimum parmi l'ensemble des solutions globales, mais il sélectionnera la première solution locale optimale trouvée, afin d'éviter de consommer trop de temps.

### 3.3.4 Jeux à mémoire parfaite et imparfaite

Les jeux à mémoires parfaites sont des jeux où chaque joueur mémorise chaque décision à chaque étape du jeu. Ils peuvent ainsi se souvenir de leurs choix à tout moment pour adapter leur stratégie. Les jeux à mémoire imparfaite sont basés sur une amnésie des joueurs.

Le Smart Grid est un système qui se doit d'être hybride. A la base, les joueurs du Smart Grid sont à mémoire parfaite. Ils connaissent pendant un intervalle de temps maximal défini leurs décisions, leurs consommations/productions, etc. Cependant, une part d'amnésie est nécessaire, soit par rapport aux contraintes de conservation et exploitation des données au-delà d'un certain temps, soit à cause de pannes matérielles ou structurelles sur le réseau.

### 3.3.5 Jeux séquentiels et jeux simultanés

Dans un jeu séquentiel, chaque étape du jeu est effectuée l'une après l'autre. Ainsi, la décision d'un joueur peut modifier la stratégie d'un joueur suivant. Au contraire, dans un jeu simultané, les décisions sont prises en même temps par tous les joueurs.

Le Smart Grid nous impose de travailler en temps réel pour répondre à la demande des consommateurs, il est donc nécessaire de mêler jeux séquentiels et simultanés afin de mettre en place un système efficace et souple.

### 3.3.6 Approche bottom-up et top-down par les réseaux de jeux pour le partage de l'énergie

Nous proposons une approche hybride top-down et bottom-up par un réseau de jeux pour le partage de l'énergie.

---

1. un problème est dit NP difficile s'il est au moins aussi difficile qu'un problème NP-complet

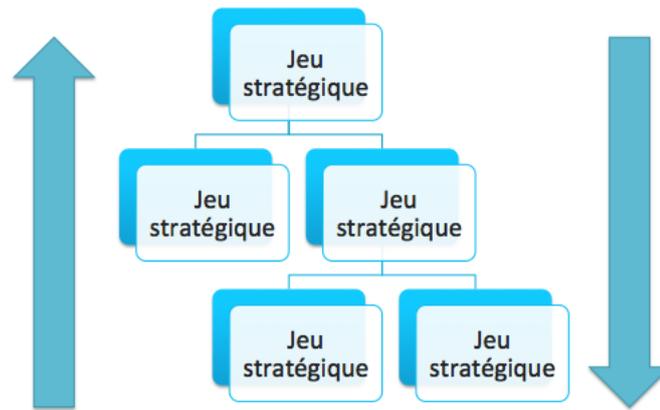


FIGURE 3.2 – Les réseaux de jeux pour modéliser les négociations autour de l'énergie.

L'approche bottom-up permet la remontée des données au coeur du système de production pour ajuster la production à la demande, puis la rétro-action top-down permet une distribution équitable par un équilibrage et un partage entre les consommateurs. Le système est découpé en jeux stratégiques exécutés tour à tour dans les chemins top-down ou bottom-up et parallèles sur un même niveau hiérarchique.

Cependant, les négociations entre agents sont basées sur des notions de proximité. Or, si nous utilisons des distances métriques, cet usage ne sera pas approprié car il ne modélisera qu'un seul critère. La topologie conventionnelle est trop lourde à implémenter pour modéliser tous les critères intervenant dans les notions de proximité entre agents. Nous avons donc choisi d'utiliser la théorie de la prétopologie car c'est la plus appropriée dans ce cas et elle apporte beaucoup d'améliorations au modèle existant.

## 3.4 Résilience et dynamique du réseau par la prétopologie

Nous proposons l'intégration de la prétopologie pour la modélisation des notions de proximité entre agents de notre modèle. Cela va nous permettre de manipuler plusieurs critères simultanément en utilisant une agrégation booléenne, et cette théorie modélisera de façon plus réaliste le comportement du Smart Grid. De plus, dans le cas de la modification d'un seul critère, nous aurons seulement à changer la fonction correspondante.

### 3.4.1 Rappels et définitions

La prétopologie est une théorie mathématique ayant une axiomatique plus faible que la topologie classique, et qui permet d'exprimer les transformations structurelles d'ensembles composés d'éléments en interaction tels que la constitution de coalitions parmi une population, les phénomènes d'alliance, les processus de tolérance, d'acceptabilité et l'émergence de comportements collectifs[?, ?]. La topologie est un cas particulier de la prétopologie.

**Définition 3.5** Nous appelons adhérence définie sur  $E$ , toute fonction  $a(.)$  de  $P(E)$  dans  $P(E)$  telle que :

- $a(\emptyset) = \emptyset$
- $\forall A \subset E, A \subset a(A)$

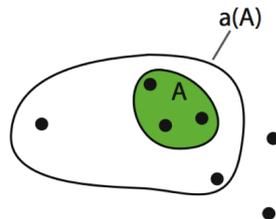


FIGURE 3.3 – Adherence de A.

**Définition 3.6** Nous appelons intérieur défini sur  $E$ , toute fonction  $i(.)$  de  $P(E)$  dans  $P(E)$  telle que :

- $i(E) = E$
- $\forall A \subset E, i(A) \subset A$

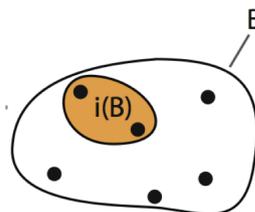


FIGURE 3.4 – Interieur de A.

Ces fonctions permettent de modéliser les opérations de base dans les réseaux complexes. En effet, comme le montre la figure 3.5, les réseaux socio-techniques complexes tels que le Smart Grid sont dynamiques et évoluent à chaque intervalle de temps étudié. Il est donc important de pouvoir les modéliser par des fonctions prenant en compte cette dynamique.

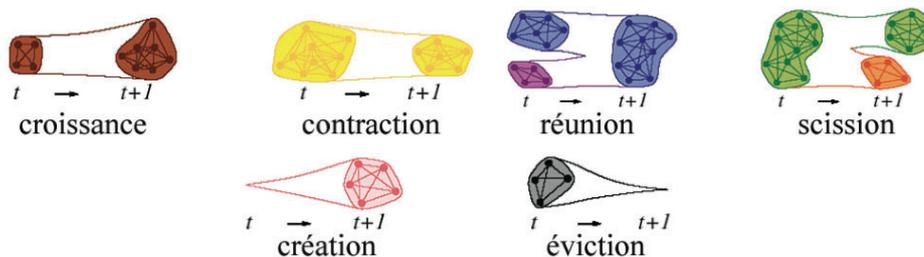


FIGURE 3.5 – Opérations présentant la dynamique des réseaux[?].

**Définition 3.7** *Etant donné, un ensemble  $E$ ,  $a(\cdot)$  et  $i(\cdot)$ , le couple  $s = (a(\cdot), i(\cdot))$  est appelé une structure prétopologique sur  $E$  et le 3-uple  $(E, a(\cdot), i(\cdot))$  est appelé espace prétopologique.*

Différents types d'espaces prétopologiques sont définis en rajoutant une propriété supplémentaire aux propriétés de base de la fonction d'adhérence :

- les espaces prétopologiques de type  $V$  ;
- les espaces prétopologiques de type  $V_d$  ;
- les espaces prétopologiques de type  $V_s$  ;

**Définition 3.8** *Par défaut, un espace topologique est un espace prétopologique de type  $V_d$ , tel que :  $\forall A \subseteq E, a(a(A)) = a(A)$*

**Définition 3.9 (Espace prétopologique de type  $V$ )** *Soit  $(E, i, a)$  un espace prétopologique. On dit qu'il est de type  $V$  si et seulement si :*

- $\forall A, B; A \subseteq E, B \subseteq E, \text{ et } A \subset B \text{ alors } a(A) \subset a(B)$
- $\forall A, B; A \subseteq E, B \subseteq E, \text{ et } A \subset B \text{ alors } i(A) \subset i(B)$

**Définition 3.10 (Espace prétopologique de type  $V_d$ )** *Soit  $(E, i, a)$  un espace prétopologique. On dit qu'il est de type  $V_d$  si et seulement si :*

- $\forall A, B; A \subseteq E, B \subseteq E, \text{ et } A \subset B \text{ alors } a(A \cup B) = a(A) \cup a(B)$
- $\forall A, B; A \subseteq E, B \subseteq E, \text{ et } A \subset B \text{ alors } i(A \cap B) = i(A) \cap i(B)$

**Définition 3.11 (Espace prétopologique de type  $V_s$ )** *Soit  $(E, i, a)$  un espace prétopologique. On dit qu'il est de type  $V_s$  si et seulement si :  $\forall A, B, A \subseteq E, B \subseteq E, \text{ et } A \subset B$  alors  $a(A) = \bigcup_{x \in A} a(\{x\})$*

**Définition 3.12 (Base de voisinage)** *Soit  $B(x)$  la base  $V(x)$  définie comme suit :  $\forall V, V \in V(x), \exists B, B \in B(x), B \subseteq V$ .*

Il est possible de définir la base de voisinage d'un élément de plusieurs manières selon la problématique. Par exemple, un espace prétopologique de type  $V$  permet de définir la notion de famille de voisinage d'un point :  $\forall x \in E, V(x) = \{V \subset E / x \in i(V)\}$ . Ou encore, un espace muni d'une relation binaire  $R_i$  aura une base de voisinage définie comme telle :  $B_i(x) = \{y \in E / x R_i y\}$ .

### 3.4.2 Famille de prétopologies pour la représentation des différents critères et leurs intrications

En prétopologie, un réseau complexe est vu comme une famille de prétopologies sur un ensemble donné  $E$  [?].

L'avantage de la prétopologie est de pouvoir séparer chaque critère dans un espace prétopologique propre afin de simplifier la modélisation. La fonction adhérence globale est alors définie comme une agrégation booléenne des adhérences de tous les espaces composant notre ensemble de données. De cette manière, une modification d'un critère est instantanée et la représentation de chaque critère est plus réaliste.

Les espaces prétopologiques peuvent inclure des relations de divers types :

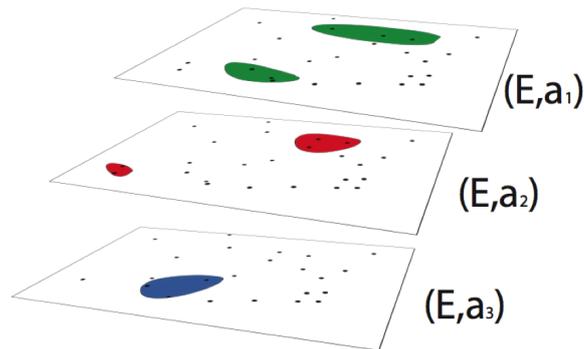


FIGURE 3.6 – Familles

- Espace métrique : les éléments sont liés par des distances.
- Espace binaire : les liens sont de type binaire comme en théorie des graphes.
- Espace valué : Lorsque les espaces binaires et métriques ne sont pas suffisants, on peut calculer la valeur de chaque arête par une fonction d'adhérence définie.

Le Smart Grid étant un système complexe, il est régi par un certain nombre de critères qualitatifs et quantitatifs. La tension des lignes électriques en est un exemple, mais on peut également prendre en compte la dispersion électrique en fonction de la longueur des lignes, l'usure des installations, les conditions climatiques...

Ces critères se modélisent de façon plus réaliste et facilement à l'aide de la prétopologie.

Nous proposons une modélisation par 3 espaces prétopologiques de type  $V_S$  :

- un espace binaire représentant la structure physique des lignes
- un espace valué pour la capacité des lignes électriques
- un espace valué pour la production énergétique en temps réel

**Définition 3.13** *Soit*

- un ensemble  $V = \{x_1, x_2, \dots, x_n\}$  représentant les entités physiques du Smart Grid (stations, sous-stations, consommateurs, producteurs...).
- une famille  $E = (a_1, a_2, \dots, a_n)$  d'éléments du produit Cartésien  $V * V = \{(x, y) / x \in V, y \in V\}$  représentant les arcs entre les agents voisins, c'est à dire les lignes électriques.
- une fonction d'adhérence pour modéliser la connectivité de la ligne :

$$a^1 : E \rightarrow \mathbb{N}$$

$$(x, y) \rightarrow a^1(x, y) = 0 \text{ ou } 1$$

L'espace prétopologique  $(V, a^1)$  permet la modélisation de la structure physique du réseau électrique.

**Définition 3.14** *Soit*

- un ensemble  $V = \{x_1, x_2, \dots, x_n\}$  représentant les entités physiques du Smart Grid (stations, sous-stations, consommateurs, producteurs...).
- une famille  $E = (a_1, a_2, \dots, a_n)$  d'éléments du produit Cartésien  $V * V = \{(x, y) / x \in V, y \in V\}$  représentant les arcs entre les agents voisins, c'est à dire les lignes électriques.

- une fonction adhérence pour modéliser la capacité des lignes :

$$a^2 : E \rightarrow \mathbb{R}$$

$$(x, y) \rightarrow a^2(x, y) = \text{capacité de transport maximale de l'élément } x \text{ au temps } t$$

L'espace prétopologique  $(V, a^2)$  modélise la capacité de transport énergétique.

**Définition 3.15** Soit

- un ensemble  $V = \{x_1, x_2, \dots, x_n\}$  représentant les entités physiques du Smart Grid (stations, sous-stations, consommateurs, producteurs...).
- une famille  $E = (a_1, a_2, \dots, a_n)$  d'éléments du produit Cartésien  $V * V = \{(x, y) / x \in V, y \in V\}$  représentant les arcs entre les agents voisins, c'est à dire les lignes électriques.
- une fonction adhérence pour modéliser la production dynamique :

$$a^3 : E \rightarrow \mathbb{R}$$

$$(x, y) \rightarrow a^3(x, y) = \text{production électrique de l'élément } x \text{ au temps } t$$

L'espace prétopologique  $(V, a^3)$  modélise la production énergétique et l'intégration des énergies à production variable.

Nous obtenons donc une modélisation du réseau par agrégation booléenne des critères d'une famille prétopologique. Cette modélisation est de complexité linéaire.

**Définition 3.16** Soit  $p$  niveaux différents.

Sur chaque niveau  $i$ ,  $i = 1, \dots, p$ , on dispose pour tout  $x$  de  $E$  ( $E$  l'ensemble des sommets du niveau) d'un concept de base de voisinage  $B_i(x)$ . En réalité il s'agit d'une famille de voisinages  $B_i(x) = \{B_i(x)\}$ .

La fonction d'adhérence globale est définie par :  $\forall A \subset E, a(A) = \{y \in E / \exists i, i = 1, \dots, p, B_i(y) \cap A \neq \emptyset\}$  Ceci revient à dire que  $y$  est adhérent à  $A$  si au moins une relation de voisinage est vérifiée.

$$\text{Soit } y \in a(A \cup B) \Rightarrow y \text{ est tel que } \exists i, i = 1, \dots, p, / B_i(x) \cap [A \cup B] \neq \emptyset$$

$$\text{Donc } \exists i, i = 1, \dots, p, / [B_i(x) \cap A] \cup [B_i(x) \cap B] \neq \emptyset$$

$$\Rightarrow \exists i, i = 1, \dots, p, / B_i(x) \cap A \neq \emptyset \text{ ou } B_i(x) \cap B \neq \emptyset$$

$$\Rightarrow y \in a(A) \cup a(B).$$

$$\text{Réciproquement } y \in a(A) \cup a(B) \Rightarrow \exists i_0 / B_{i_0}(x) \cap A \neq \emptyset \text{ ou } \exists i_1 / B_{i_1}(x) \cap B \neq \emptyset$$

$$\Rightarrow \exists i_0 / B_{i_0}(x) \cap [A \cup B] \neq \emptyset \text{ ou } \exists i_1 / B_{i_1}(x) \cap [A \cup B] \neq \emptyset.$$

$$\Rightarrow y \in a(A \cup B)$$

On a donc

$$a(A \cup B) = a(A) \cup a(B)$$

On a un espace de type  $V_D$ .  $E$  étant fini par définition on a donc un espace de type  $V_S$ . Ceci est intéressant du point de vu calcul car on peut écrire :

$$\forall A \subset E, a(A) = \bigcup_{x \in A} a(\{x\})$$

Ce qui implique un calcul sommet par sommet et donc une complexité linéaire.

### 3.4.3 $\chi$ -connexités en prétopologie pour analyser la connectivité du réseau

La connectivité est un concept intéressant pour les problèmes de modélisation liés à l'homogénéité, ou à la structuration des éléments. Cependant les axiomes de la topologie classique sont souvent en contradiction avec les conditions réelles de terrain, en particulier dans les sciences sociales [?].

La prétopologie définit différents types de connexités. Dans cette section, nous considérons un ensemble  $E$ , défini par une structure prétopologique  $s = (i(\cdot), a(\cdot))$  de type  $V$ . Nous notons  $\sigma = (O(\cdot), F(\cdot))$  les opérations d'ouverture et de fermeture de  $E$ . Il existe différents types de connexité, que l'on nomme les  $\chi$ -connexités avec  $\chi$  correspondant à :

- SC : fortement connexe ;
- UC : unilatéralement connexe ;
- HC : hyperconnexe ;
- AC : apo-connexe ;
- C : connexe ;

#### Fortement connexe

**Définition 3.17**  $\forall A \subseteq E, A \neq \emptyset, F(A) = E$  (le seul fermé non nul est  $E$ ).

$E$  est SC si et seulement si  $\forall A \in P(E), \forall B \in P(E)$ , il existe un chemin de  $A$  vers  $B$ .

#### Unilatéralement connexe

**Définition 3.18**  $E$  est unilatéralement connexe (UC) si et seulement si :

$\forall A \subseteq E, A \neq \emptyset, F(A) = E$  sinon  $\forall B \subseteq E, B \neq \emptyset, B \subseteq E - F(A) \Rightarrow A \subseteq F(B)$  ;

Cela revient à dire que  $E$  est UC si et seulement si pour tout  $A \in P(E)$  et pour tout  $B \in P(E)$  il existe un chemin de  $A$  vers  $B$  ou de  $B$  vers  $A$  (les fermés sont emboîtés).

#### Hyperconnexe

**Définition 3.19**  $\forall A \subseteq E, A \neq \emptyset, F(A) = E$ . Sinon  $\exists B \neq \emptyset, B \subseteq E - F(A)$ , tel que  $A \subseteq F(B)$

$E$  est HC si et seulement si  $\forall A \in P(E), \forall B \in P(E)$ , il existe un chemin de  $A$  vers  $B$  ou il existe un chemin de  $B$  vers  $E - F(B)$ , partie qui rencontre  $A$  (la fermeture de tout ouvert non vide vaut  $E$ ).

#### Apo-connexe

**Définition 3.20**  $\forall A, \forall B$  deux sous-ensembles non vides de  $E$ , on a toujours  $F(A) \cap F(B) \neq \emptyset$

$E$  est AC si et seulement si  $\forall A \in P(E), \forall B \in P(E)$ ,  $\exists M \in P(E)$  avec un chemin de  $M$  vers  $A$  et un de  $M$  vers  $B$ .

**Connexe** E est C si et seulement si  $\forall A \in P(E), \forall B \in P(E)$ , il existe un chemin de A vers B ou  $\exists M \in P(E)$  avec un chemin de M vers B et un de M vers E-F(B), partie qui rencontre A.

On a alors :

- E est SC  $\Rightarrow$  E est UC
- E est UC  $\Rightarrow$  E est HC et E est AC
- E est HC  $\Rightarrow$  E est C
- E est AC  $\Rightarrow$  E est C

Grâce à la prétopologie, nous pouvons donc enrichir notre modèle avec des relations multi-variées qualitatives entre les entités. Cependant, les Smart Grids sont des systèmes dynamiques donc nous avons besoin d'une méthode pour vérifier que notre modèle permet au système de fonctionner correctement. La percolation va nous aider sur ce point : nous allons vérifier que le réseau est toujours connecté en terme de distribution d'énergie et donc que les clients sont tous satisfaits. Nous présentons cette théorie en détail dans la section suivante.

### 3.5 Analyse de la propagation de l'énergie et percolation

La seule modélisation des contraintes et de la structure du réseau complexe ne permet pas de garantir son fonctionnement au niveau global. Les propriétés émergentes du système nécessitent une surveillance globale. C'est particulièrement valable dans le cadre de la détection et prévention des pannes. Nous proposons une approche hybride centralisée/distribuée via un superviseur global du réseau pour détecter les seuils critiques pouvant mener à des baisses de tension locales, et ainsi assurer la prévention des pannes. Le module du superviseur se rapproche d'un système autonome *ie.* une composante auto-adaptable dynamique.

La percolation est une théorie mathématique utilisée pour exprimer les processus de diffusion déterministe sur une structure stochastique [?]. La théorie de la percolation permet d'expliquer de nombreux phénomènes complexes en analysant les comportements de transition de phase de ces systèmes. Elle permet de mettre en valeur certains mécanismes sous-jacents. Le nom "percolation" provient du latin percolatio, signifiant filtration. Il réfère en général au concept d'agglomération et de propagation dans des supports aléatoires partiellement interconnectés [?].

**Définition 3.21** *Les problèmes physiques tels que les défis soulevés par les Smart Grids peuvent être modélisés par un réseau de points ( $n$  consommateurs) et les connections (ou arêtes) entre chaque voisin peuvent être ouvertes (permettant à l'électricité de passer) avec une probabilité  $p$ , ou fermées avec une probabilité  $(1 - p)$ , et nous définissons qu'elles sont indépendantes[?]. Il existe donc une proportion  $(1 - p)$  de liens qui peuvent être aléatoirement désactivés, sans affecter le réseau électrique de distribution, ni provoquer de panne électrique visible par les consommateurs, ni n'interfère entre une station et une autre, car il est toujours possible de passer par des stations relai si  $p$  est plus grand qu'un seuil critique, appelé seuil de percolation  $p_c$ . En dessous de cette valeur critique  $p_c$ , la*

*probabilité que deux stations puissent communiquer directement ou indirectement est très faible.*

Un modèle comportant des sommets ouverts ou fermés est appelé "percolation de sites" alors que le modèle décrit ci-dessus est appelé percolation de liens, mais beaucoup de variantes ont également été étudiées [?] : par exemple, un modèle où l'aléatoire concerne à la fois les sommets et les liens est appelé percolation mixte.

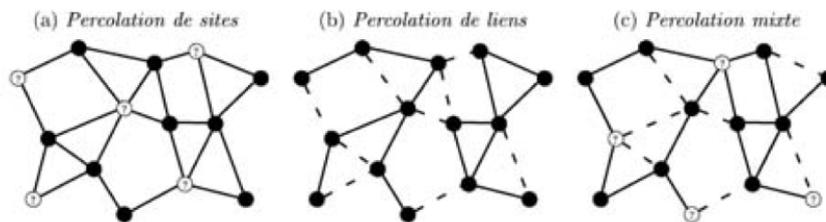


FIGURE 3.7 – Les différents types de percolation.

Cette théorie permet notamment d'expliquer le phénomène de transition de phase pendant les périodes de pointes précédées par des demandes en énergie presque nulles sans transition. La théorie de la percolation relève en effet de la famille des transitions de phase car la transition d'un état vers un autre est brutal en changeant la valeur d'un paramètre du modèle.

Ce modèle mathématique a été originellement développé pour l'étude de phénomènes physiques [?], tels que l'étude des flux de liquide dans un milieu poreux, les phénomènes de connexion/déconnexion dans un réseau de résistances, l'analyse de matériaux composites conducteurs-isolants, les transitions de glaciation, ou encore les phénomènes de polymérisation. Aujourd'hui, la percolation est appliquée dans beaucoup d'autres champs variés autres que la physique [?]. En effet, cette théorie a contribué, par exemple, à la modélisation de systèmes complexes en économie [?], marketing [?], sociologie [?], informatique [?], écologie [?], ou mathématiques [?].

Dans notre cas, nous avons choisi d'intégrer la percolation à notre modèle en tant que superviseur global pour la modélisation de deux processus de la grille électrique relevant de la transition de phase.

### 3.5.1 Détermination du seuil de connectivité assurant l'efficacité du réseau

La première contribution de la théorie de la percolation dans notre modèle concerne la détermination du seuil de diffusibilité du réseau de distribution électrique. Les pannes aléatoires des arêtes n'affectent pas la connectivité du réseau jusqu'à un certain seuil. Au delà de ce seuil, la diffusion de la ressource électrique est altérée et tous les consommateurs ne sont pas satisfaits.

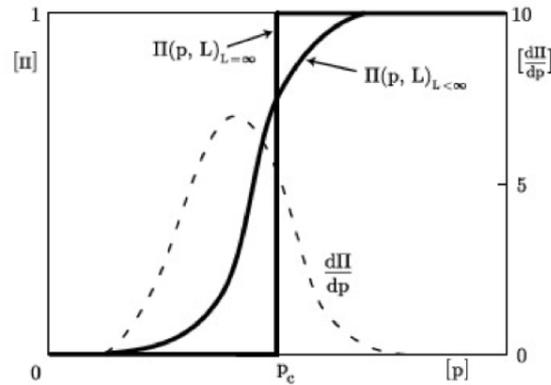


FIGURE 3.8 – Définition du seuil de diffusibilité du réseau.

### 3.5.2 Prévention des pannes et analyse du phénomène de brownout

La deuxième contribution de la percolation dans notre modèle intervient dans la détection et la prévention des pannes. Le brownout est un phénomène électrique qui se caractérise par une panne partielle : il peut y avoir des micro-coupures électriques pouvant conduire à une panne totale appelée blackout, voire même à une série de pannes en cascade [?] si le problème persiste. Ce phénomène est indépendant de la structure du réseau. Il peut être lié à des problèmes de congestion, de surconsommation, ou de sous-production. En particulier, l'intégration des énergies renouvelables génère des creux de production par leur production intermittente qu'il faut combler par l'utilisation d'autres sources ou de réserves énergétiques. Ces irrégularités peuvent être également à l'origine de brownouts [?].

Un des enjeux de notre modélisation est d'optimiser l'intégration des énergies renouvelables. Nous devons donc mettre en place un superviseur global capable de prévenir le phénomène de brownout. D'un point de vue électrique, la figure 3.9 prouve que le phénomène de brownout est très proche du phénomène de transition de phase par la percolation.

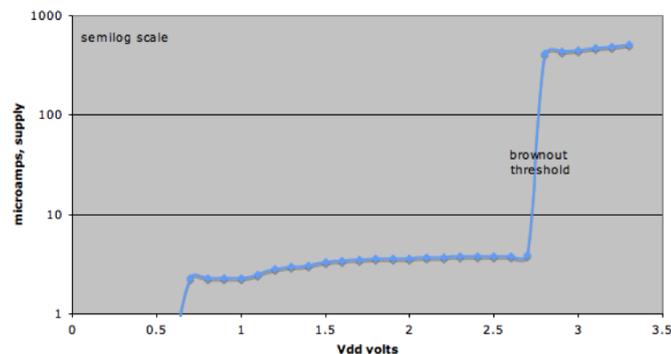


FIGURE 3.9 – Définition du seuil de brownout du réseau.

Notre deuxième contribution de la percolation dans notre modèle consiste donc à déterminer le seuil de brownout du système afin d'en minimiser l'impact sur le réseau.

### 3.6 Déroulement de notre modèle

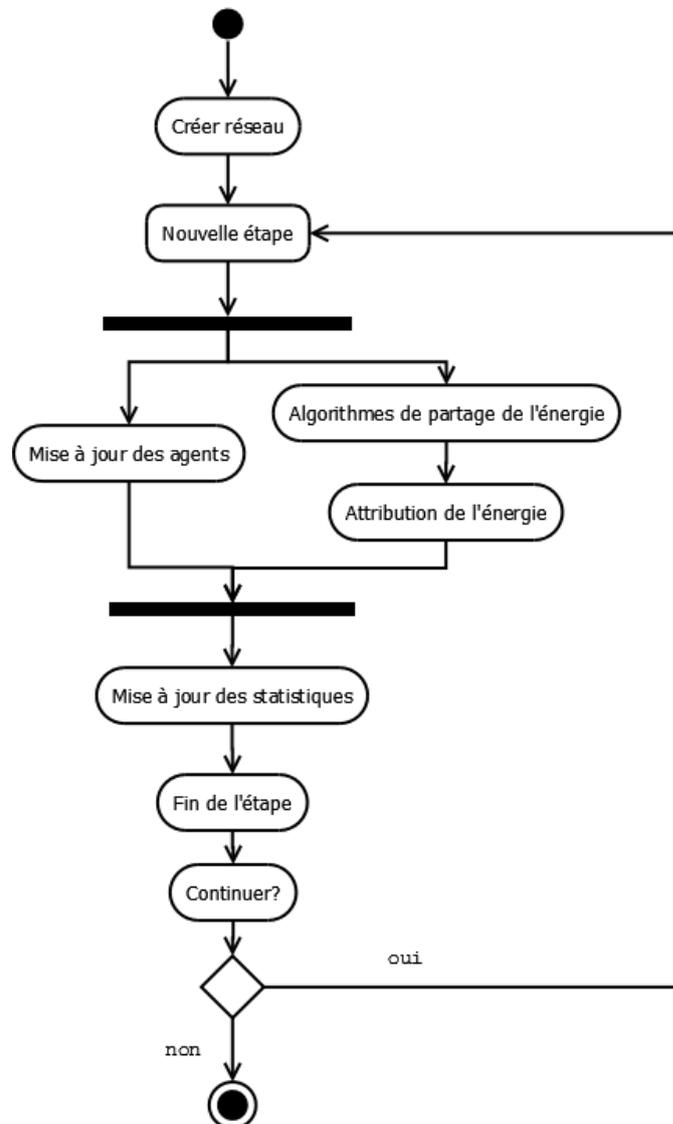


FIGURE 3.10 – Diagramme d'activités UML simplifié du modèle.

Le début d'une nouvelle itération commence avec la mise à jour des données de consommation, de production et de la distribution.

Les actions du niveau local sont décrites dans la figure 3.11. Chaque niveau local commence par mettre à jour ses données de consommation et les priorités de ses appareils électriques. L'utilité de chaque appareil est calculée en fonction de la priorité, de la nécessité propre à l'appareil (un hôpital ne peut pas subir de coupure d'électricité par exemple), et de sa consommation. Ensuite, le niveau local compare ses valeurs de consommation par rapport aux pronostiques calculés aux temps précédents en étape 1. Si les données indiquent que ce foyer est en surconsommation par rapport aux pronostiques, nous utilisons la méthode du sac à dos pour déterminer la répartition optimale

de l'énergie entre les appareils les plus utiles au foyer. Un système de récompenses et de punitions permet de remercier les consommateurs économisant l'énergie par rapport aux prévisions, et d'inciter les autres à y participer également.

Après obtention de leur consommation idéale, chaque niveau local envoie ses données au niveau supérieur : le microgrid. Cette étape marque le passage à l'étape 2 de notre algorithme.

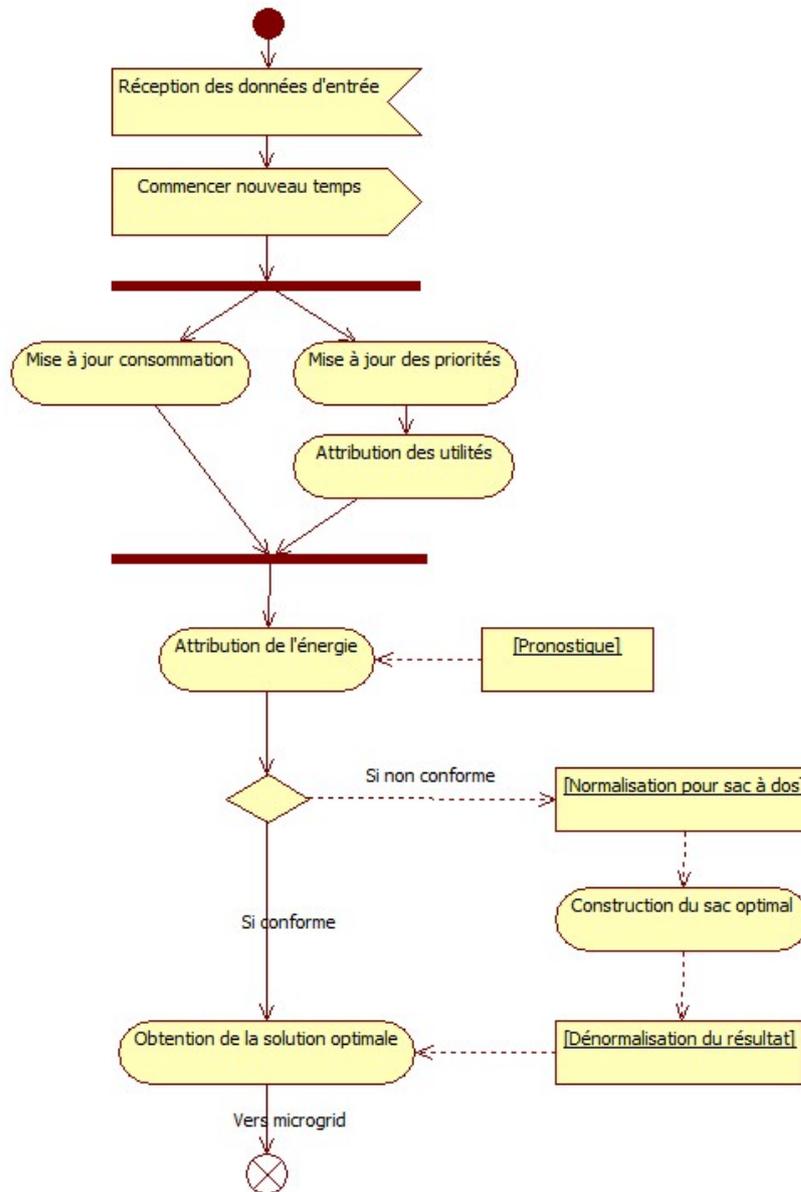


FIGURE 3.11 – Diagramme d'activités UML des algorithmes locaux.

Le microgrid, au bout d'un intervalle de temps défini, effectue des enchères en fonction de la consommation de chaque niveau local, puis somme et envoie son résultat à la station parente (étape 3). Cette étape vise à répartir au mieux sa ressource électrique

prévisionnelle en fonction des informations reçues par chaque niveau local.

A l'étape 4, le réseau de transmission et distribution met à jour son routage à l'aide du graphe résiduel. Puis, à l'étape 5, la station prend connaissance de l'énergie reçue par le routage. Un système de rétroaction entre le microgrid et la station, lors de l'étape 6, permet d'équilibrer les enchères par rapport à l'énergie réelle pouvant être desservie. Les données de la rétroaction permettent d'obtenir des pronostiques, par pondération, pour la prochaine itération.

Une fois le consensus trouvé, ou équilibre, l'énergie est distribuée à l'étape 7 au niveau local. L'énergie non allouée, s'il en reste, est redistribuée.

L'ensemble des relations entre les 3 niveaux du réseau est représenté dans la figure 3.12.

Finally, we will obtain a decision aid tool that will allow us to define in an optimal way the value of certain variables and to modify the value of parameters using theories of modeling enriched by classical mathematical theories. Our objective is to put in place a very effective model of management of the complexity of Smart Grids in order to control it in an optimal way and to improve its performance.

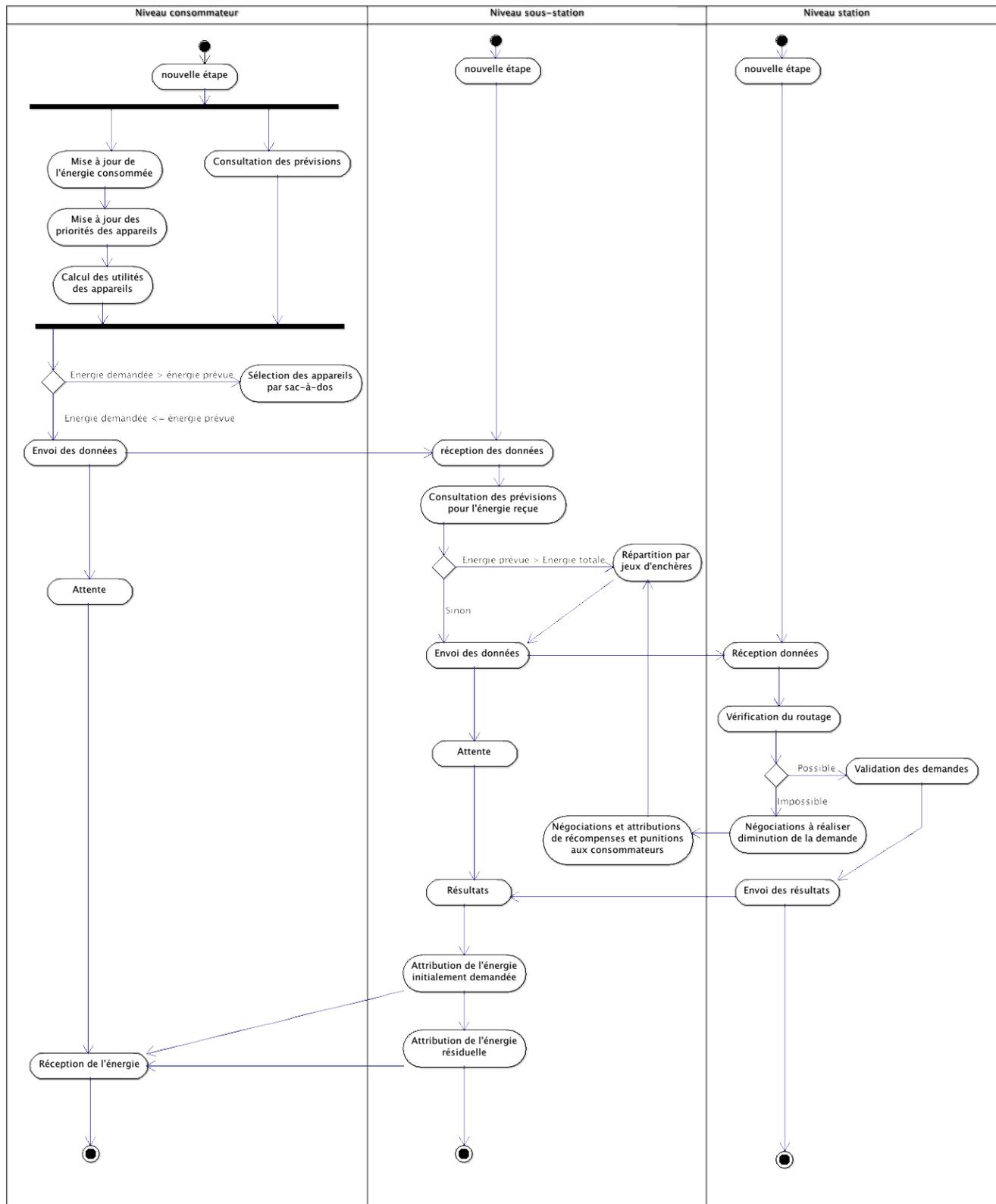


FIGURE 3.12 – Diagramme d’activités UML du système complet selon les niveaux.



## Chapitre 4

# Simulation multi-agents autonome pour l'optimisation au sein de Smart Grids

La simulation d'un système complexe a pour but de prédire une des évolutions possibles du système. Dans cette partie, nous présentons la théorie de la simulation, puis les différentes plateformes logicielles existantes. Enfin, nous présentons une application de notre modèle, via une simulation d'un éco-quartier. Notre approche permet également d'être un outil d'aide à la décision via cette simulation qui permet aux décideurs de tester divers scénarios *in silico*.

### Sommaire

---

<b>4.1</b>	<b>Le Smart Grid : un système socio-technique autonome . .</b>	<b>67</b>
<b>4.2</b>	<b>Du modèle intégratif vers la simulation multi-agents autonome . . . . .</b>	<b>67</b>
4.2.1	Cas d'utilisations de notre simulation . . . . .	68
4.2.2	Modélisation des agents . . . . .	68
4.2.3	Granularités spatiale et temporelle . . . . .	70
4.2.4	Niveaux conceptuels . . . . .	70
	Niveau producteurs : système complexe autonome de production	72
	Niveau transmission et distribution : système de liaison et d'optimisation . . . . .	72
	Niveau consommation : micro-systèmes complexes stables . . .	72
4.2.5	Politique d'intégration des énergies renouvelables . . . . .	72
4.2.6	Niveaux modulaires . . . . .	73
	Module réseau de transport . . . . .	73
	Module consommation et production . . . . .	75
	Module distribution de l'énergie . . . . .	77
	Module gestion des pannes . . . . .	78
	Module réseau de communication . . . . .	79
<b>4.3</b>	<b>Simulation : conception, implémentation et optimisation . . .</b>	<b>80</b>

4.3.1	Besoins liés à la complexité du système et choix d'une plateforme de simulation . . . . .	80
4.3.2	Vérification et validation des résultats . . . . .	81
	Vérification de la structure du modèle . . . . .	81
	Validation des algorithmes . . . . .	81
	Performance de la simulation . . . . .	84
	Architecture de la simulation . . . . .	85

---

## 4.1 Le Smart Grid : un système socio-technique autonome

Les systèmes dits autonomiques sont capables d'adapter leurs comportements et leurs ressources plusieurs centaines de fois par seconde pour décider automatiquement de la meilleure manière d'atteindre un but fixé en dépit des changements de données environnementales et des demandes [?]. Ces systèmes peuvent s'autoréguler sans intervention humaine.

Le terme autonome a d'abord été utilisé en biologie. En anatomie, le système nerveux autonome contrôle les réflexes inconscients, c'est à dire des fonctions corporelles qui ne nécessitent pas notre attention, par exemple les ajustements automatiques tels que la taille de la pupille, les fonctions digestives, le débit et la profondeur de la respiration, ou encore la dilatation et la constriction des vaisseaux sanguins.

Le calcul autonome tend à intervenir dans les systèmes informatiques de la même manière que son homologue en biologie. L'expression calcul autonome a été pour la première fois utilisée par IBM en 2001 pour décrire les systèmes munis de propriétés d'autogestion tels que l'auto-configuration, l'auto-optimisation, l'auto-réparation et l'auto-protection [?].

L'auto-configuration consiste à donner au système des buts, sans spécifier comment le réaliser, et un système autonome est capable de choisir des actions et des paramètres adaptés afin d'atteindre l'objectif [?]. Un tel système est également capable de détecter et diagnostiquer tous types de pannes aussi bien bas niveau que haut niveau. S'il en est capable, le système tente alors une réparation sans intervention externe. C'est ce qu'on appelle auto-réparation. La propriété d'auto-protection concerne les attaques malveillantes de pirates informatiques mais également les utilisations aberrantes d'utilisateurs normaux. Les systèmes autonomiques doivent assurer la sécurité du système et la protection des données.

Le Smart Grid est un parfait exemple de système autonome. En effet, il doit satisfaire à toutes les propriétés d'autogestion et doit également faire preuve de réactivité en temps réel. En cas de panne, le Smart Grid doit la localiser précisément et rapidement afin de proposer des solutions alternatives le plus vite possible. Si possible, le Smart Grid tente une réparation s'il s'agit d'un dérèglement informatique ou électronique, sinon il analyse au mieux la panne afin de fournir un maximum d'informations aux intervenants humains.

## 4.2 Du modèle intégratif vers la simulation multi-agents autonome

Les modélisations de systèmes complexes posent des problèmes au niveau simulation [14]. La décomposition des sous-systèmes est nécessaire pour une programmation modulaire. Il faut extraire les sous-systèmes stables. Cependant, cette décomposition est complexe et ne correspond pas forcément à la décomposition modulaire du modèle mathématique.

Dans le cadre des Smart Grid en particulier, pour simuler un système aussi complexe, il est nécessaire de respecter une certaine hiérarchie. En effet, le Smart Grid regroupe différents niveaux d'échelle de complexité (temporels et spatiaux), et comprend également une intrication entre la structure et la dynamique du système. Ainsi, nous avons envisagé des niveaux conceptuels, et des niveaux modulaires de programmation, sans qu'il y ait forcément de parallélisme exact entre les niveaux conceptuels du modèle et les niveaux techniques de la simulation [10].

### 4.2.1 Cas d'utilisations de notre simulation

La première étape de l'implémentation de la simulation consiste à décrire les cas d'utilisation. Ce diagramme permet de visualiser les sections importantes de notre simulation et de proposer une décomposition logique.

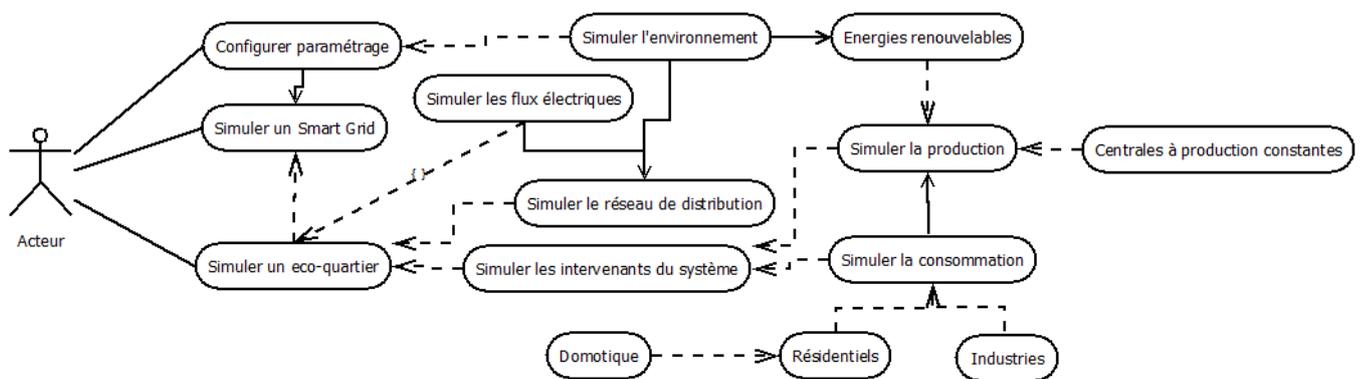


FIGURE 4.1 – Diagramme des cas d'utilisation.

### 4.2.2 Modélisation des agents

Pour représenter au mieux la réalité des Smart Grids, nous avons choisi d'intégrer à notre simulation plusieurs types d'agents.

Chaque agent est capable de communiquer avec ses voisins, qui peuvent à leur tour transmettre un message de proche en proche. Ainsi, chaque agent n'a qu'une connaissance locale du réseau, nous ne mémorisons pas l'intégralité des routes possibles, qui serait impossible à l'échelle d'un Smart Grid national.

Les agents de notre modèle sont des entités dynamiques capables de produire, consommer, ou transférer de l'énergie. Chaque agent a donc son propre modèle mathématique de gestion de l'énergie, qui peut comprendre un algorithme de production, de consommation, ou un protocole de transmission vers une cible définie.

Pour les modélisations de création ou consommation d'énergie, nous avons intégré des données réelles issues du responsable du réseau de transport français RTE. L'environnement de notre simulation définit un ensemble de paramètres externes influant le comportement des agents.

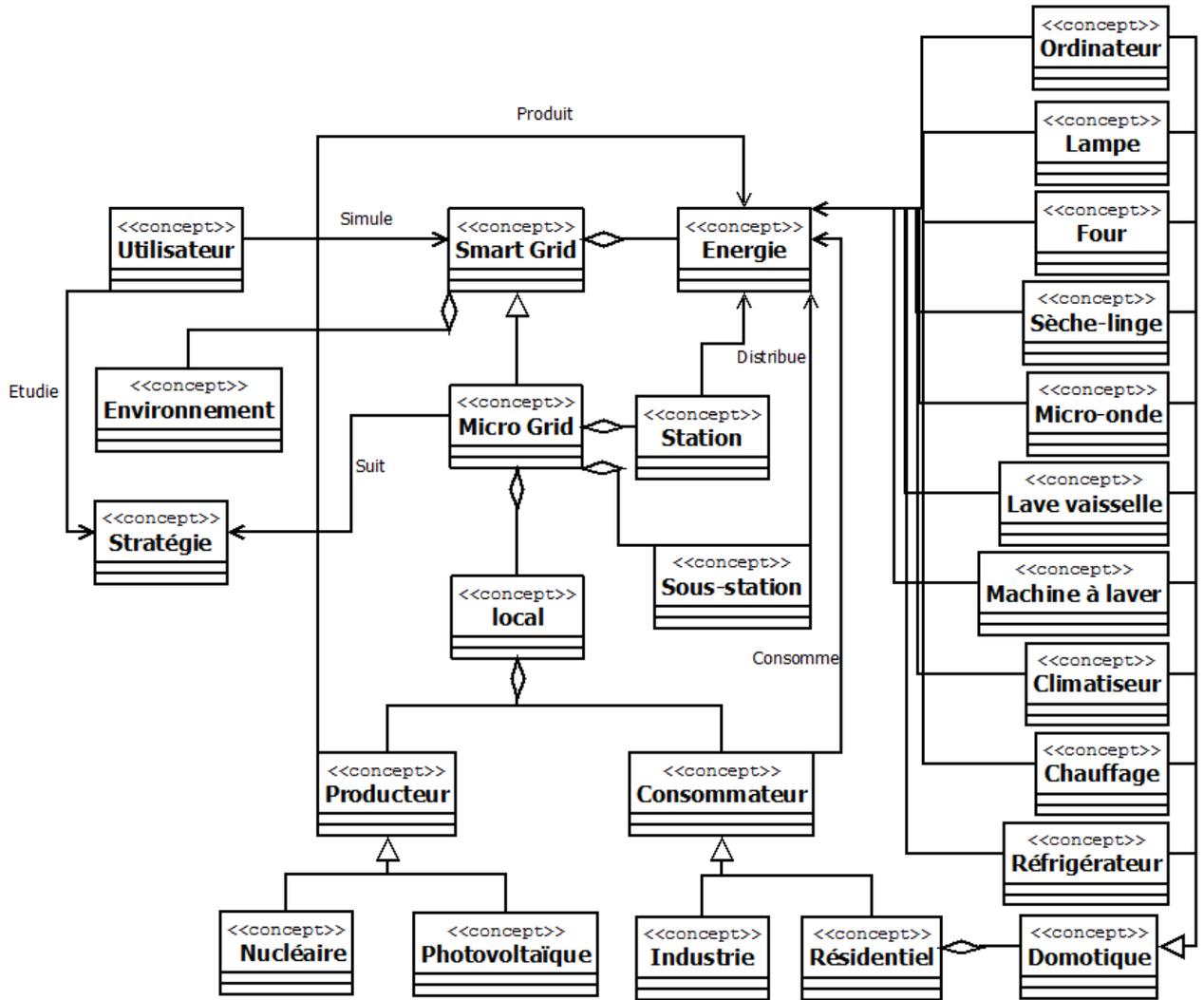


FIGURE 4.2 – Diagramme d'ontologie.

### 4.2.3 Granularités spatiale et temporelle

Notre modèle peut être envisagé avec différentes granularités, qu'elles soient spatiales ou temporelles.

D'un point de vue temporel, notre modélisation peut simuler diverses échelles qui permettent ainsi de mettre en relief les phénomènes d'émergence. Nous utilisons ainsi des pas de 5 minutes à 1 jour pour mettre en avant certains phénomènes complexes visibles uniquement sur certaines périodes temporelles (voir la figure 4.3).

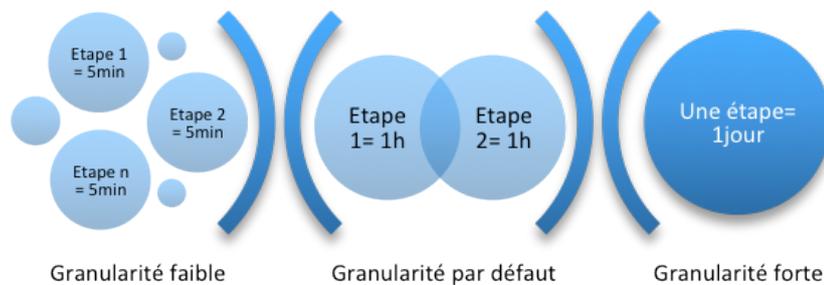


FIGURE 4.3 – Granularité temporelle de la simulation.

Par exemple, un pic ponctuel de consommation va être visible avec une échelle temporelle de 5 minutes, mais ne sera pas visible avec un pas de 1 heure. De la même manière, les pics de consommation saisonniers ne sont visibles que sur une échelle temporelle importante.

D'un point de vue spatial, un agent du modèle peut modéliser différentes entités selon le niveau d'échelle : un agent de type consommateur peut représenter une maison à l'échelle d'un éco-quartier (figure 4.4), ou alors un ensemble de maisons - ou microgrid - à l'échelle nationale (figure 4.5).

### 4.2.4 Niveaux conceptuels

Les niveaux conceptuels sont liés à la structure du réseau. Nous avons distingué 4 niveaux conceptuels interconnectés.

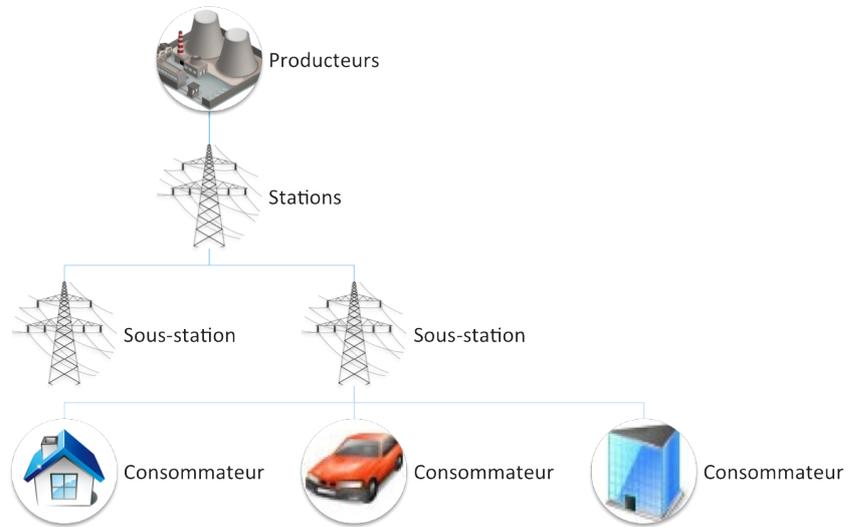


FIGURE 4.4 – Granularité spatiale de la simulation : un agent pour chaque entité.

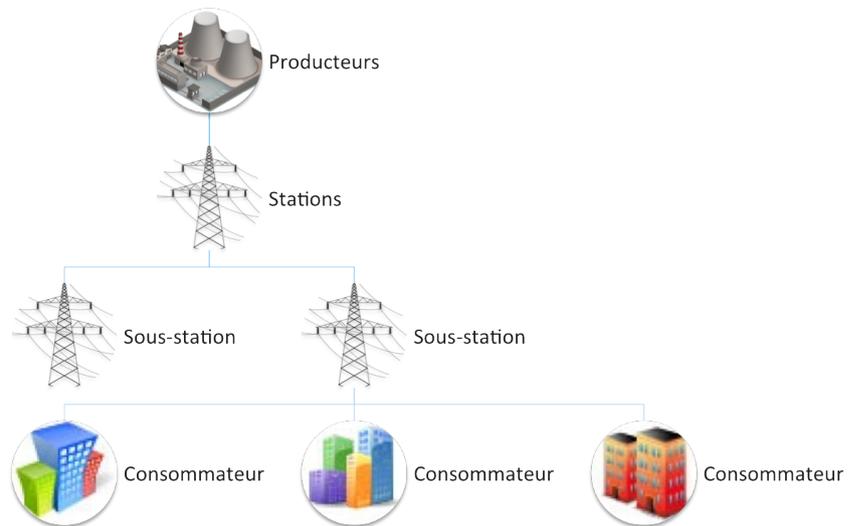


FIGURE 4.5 – Granularité spatiale de la simulation : un agent pour modéliser un ensemble d'entités.

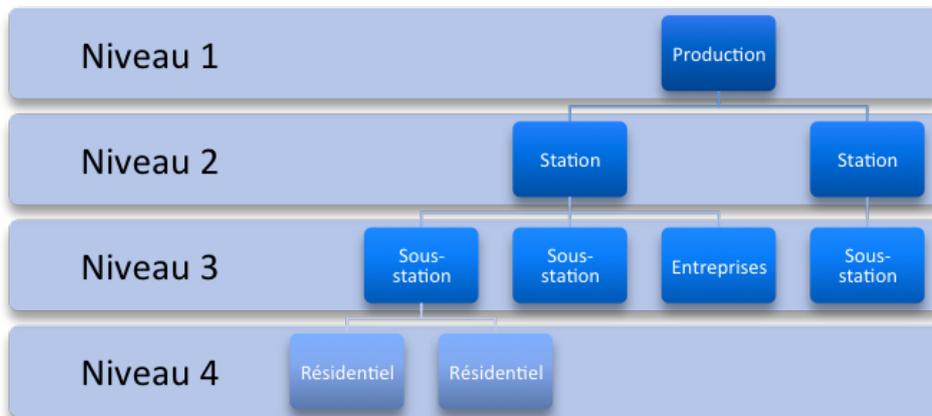


FIGURE 4.6 – Les différents niveaux conceptuels.

### Niveau producteurs : système complexe autonome de production

Le coeur du système symbolise le niveau médullaire. Il comprend la production électrique qui est modélisée par un système autonome de production d'énergie dans notre modèle.

### Niveau transmission et distribution : système de liaison et d'optimisation

Les deux niveaux intermédiaires comportent les agents de liaison qui sont également les postes d'optimisation. Ces niveaux peuvent être vus comme des niveaux logiques, représentant le passage de la haute à la moyenne tension (stations) et de la moyenne à la basse tension (sous-stations).

### Niveau consommation : micro-systèmes complexes stables

Les feuilles du réseau modélisent les consommateurs qui sont eux-mêmes des micro-systèmes complexes.

## 4.2.5 Politique d'intégration des énergies renouvelables

Les différents types d'énergies renouvelables à production intermittente pouvant être intégrés à plusieurs niveaux du réseau, nous avons mis en place une politique d'intégration composée de 3 stratégies. Ces stratégies ont l'avantage d'être très simples à mettre en oeuvre et de s'adapter à tout type d'énergie renouvelable.

La figure 4.7 illustre la relation entre stratégie d'intégration et positionnement du producteur d'énergie renouvelable :

- Si l'EnR est intégrée au niveau producteur sur le réseau haute tension, elle s'intègre en tant qu'agent supplémentaire dans notre modèle (agent producteur).
- Si l'EnR est intégrée au niveau distribution sur le réseau moyenne tension, elle est rattachée par défaut à un microgrid de référence qui prendra en compte sa production électrique lors de l'étape des enchères pour le partage de l'énergie.
- Enfin, si l'EnR est intégrée au niveau consommation, par exemple comme c'est souvent le cas pour les panneaux photovoltaïques installés sur les toits des consommateurs, elle servira à la consommation propre de l'agent sur lequel elle est installée.

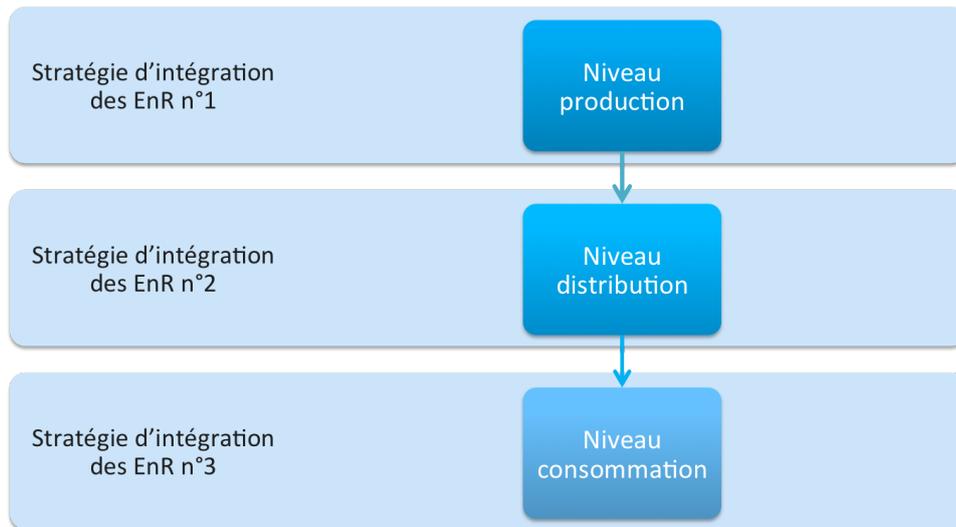


FIGURE 4.7 – Les stratégies d'intégration des énergies renouvelables.

Cette gestion a l'avantage d'être très souple en cas de création d'un nouveau type d'énergie renouvelable. De plus, comme pour notre niveau producteurs, notre modèle ne tient pas compte du profil type de production des EnR : il s'adapte totalement à l'énergie fournie par les différents producteurs pour la répartir au mieux.

#### 4.2.6 Niveaux modulaires

En raison de la complexité du modèle, nous l'avons décomposé en 5 modules que nous présentons dans la partie suivante.

Les modules de notre simulation peuvent être vus comme des thèmes de développement. Nous avons ainsi séparé en 5 parties notre simulation, afin de clarifier la tâche de développement.

##### Module réseau de transport

Tous nos agents sont interconnectés via le réseau de distribution, qui est un réseau complexe basé sur la propriété scale-free.

Grâce à une analyse de la littérature industrielle, nous avons détaillé ce réseau à l'aide d'un maillage sur 3 niveaux afin de correspondre au mieux à la réalité du réseau français. Le cœur de la production et les stations haute tension forment un réseau maillé



FIGURE 4.8 – Les différents modules composant notre modèle.

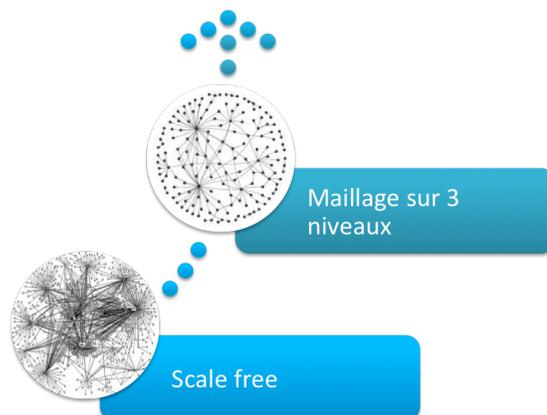


FIGURE 4.9 – Les différentes étapes du module réseau de transport.

de propriété 2-connexe (2-core). Dans la littérature industrielle, ce maillage est souvent appelé "maillage n-1" en référence à la propriété 2-connexe : la perte d'une arête n'a pas de conséquences sur la connectivité du réseau.

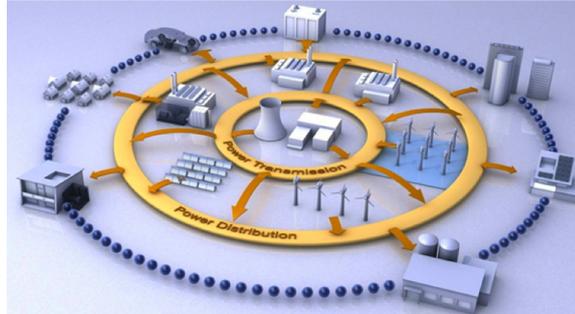


FIGURE 4.10 – Les 3 niveaux du réseau de transmission et distribution.

Chaque station haute tension est connectée à au moins une sous-station. Cette partie représente le réseau de transmission et le passage de haute à moyenne tension. La distribution en basse tension est ensuite effectuée à tous les consommateurs via un réseau linéaire ou scale free selon la granularité voulue.

Les énergies renouvelables peuvent être connectées n'importe où dans le réseau.

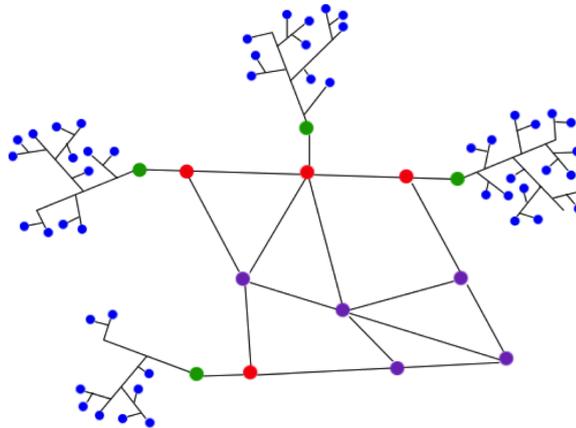


FIGURE 4.11 – Un exemple de réseau simulé.

### Module consommation et production

Le module de consommation et production a pour but de reproduire les mécanismes liés à la production et à la demande en énergie par l'inclusion de profils types de consommation et production.

Il est difficile d'obtenir des informations précises heure par heure de la consommation et l'utilisation de l'énergie. Cependant, nous avons connaissance de la répartition des consommateurs d'énergie et nous savons ainsi que la majorité de la consommation est réalisée par les résidentiels.

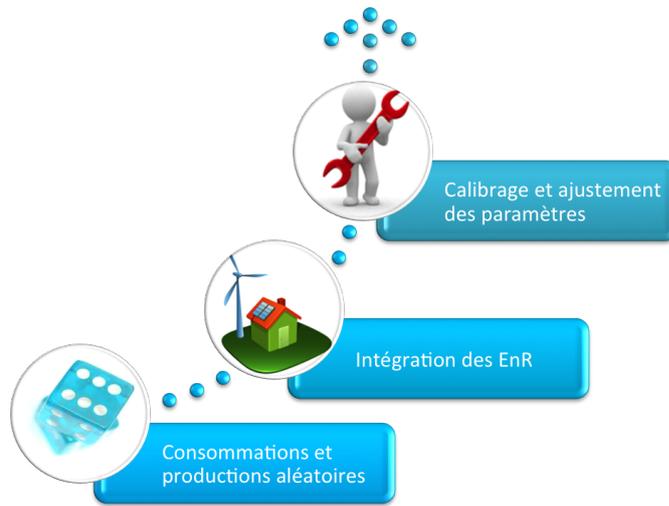


FIGURE 4.12 – Les différentes étapes du module de simulation de la production et de la consommation.

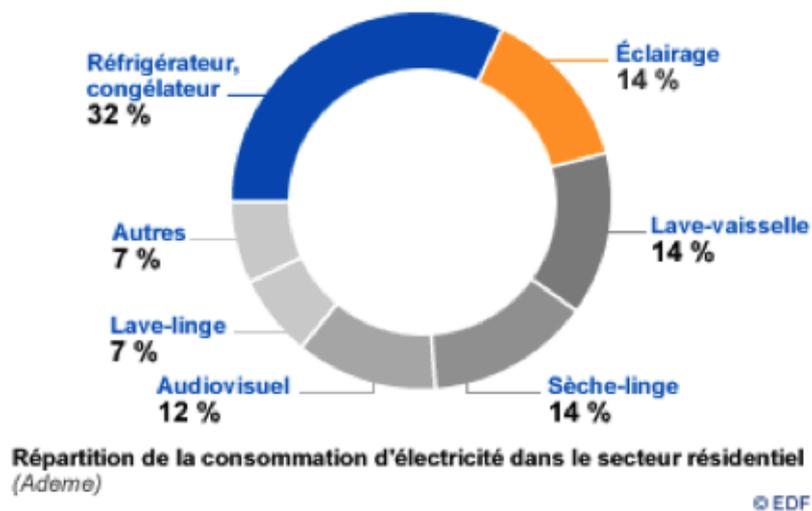


FIGURE 4.13 – Les utilisations énergétiques du consommateur.

Il est donc capital de modéliser au mieux les agents de type résidentiels. Pour cela, nous avons analysé l'utilisation de la ressource au sein de leur foyer. Nous savons par exemple que le réfrigérateur génère à lui seul un tiers de la consommation de la domotique.

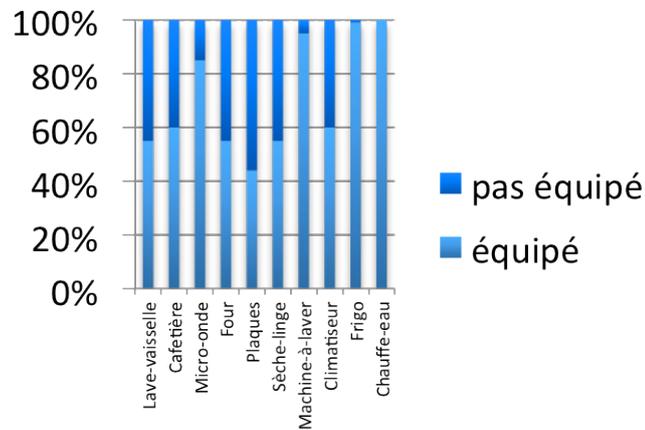


FIGURE 4.14 – La répartition des appareils dans les foyers français.

Nous avons utilisé les données d'utilisation quotidienne des appareils et de leur consommation horaire pour notre simulation.

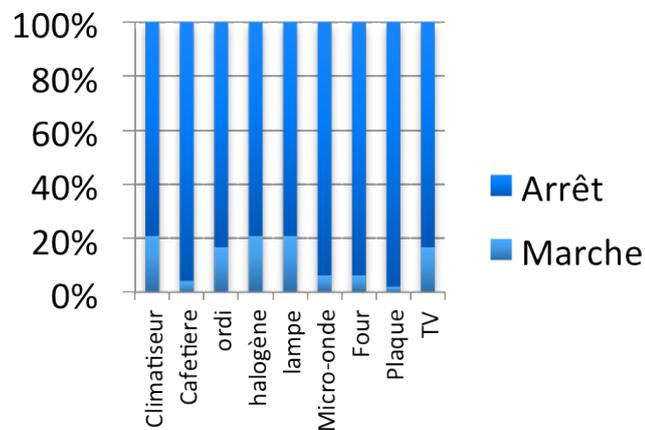


FIGURE 4.15 – La durée de marche quotidienne des appareils d'un foyer.

Grâce à ces probabilités de distribution, nous avons implémenté des profils types pour la consommation d'électricité.

### Module distribution de l'énergie

Le module distribution de l'énergie est en fait le coeur de notre modèle. C'est dans ce module que nous intégrons tout notre travail de modélisation du partage de l'énergie du Smart Grid par des algorithmes basés sur les théories des jeux, la prétopologie et la percolation.

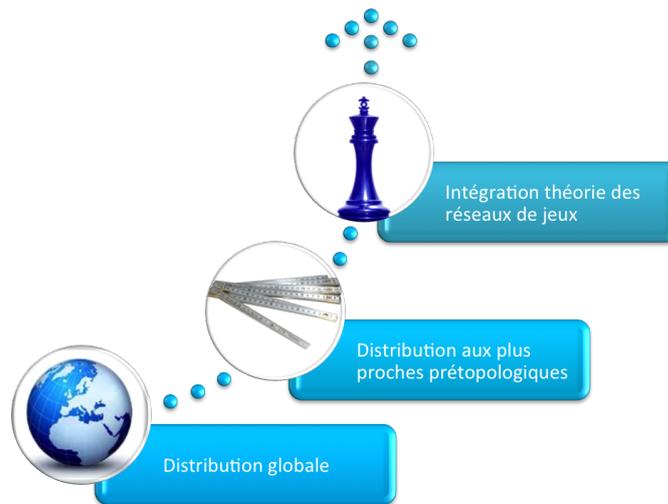


FIGURE 4.16 – Les différentes étapes du module de distribution et partage.

Nous avons conceptualisé ce module en 3 étapes comme représenté dans la figure 4.16.

Dans la première version de notre simulation, nous ne regardions que les consommations et productions globales du système, afin de vérifier si la production peut répondre à la demande d'un point de vue quantitatif et global. La deuxième étape permet l'intégration de la prétopologie pour la représentation réaliste des critères du réseau, et nécessite la mise en place de la dynamique de communication multi-niveaux et inter-agents. Enfin, la dernière étape repose sur l'inclusion de la théorie des jeux afin de déterminer un équilibre global du système. Dans ce cadre, nous avons développé au sein de notre laboratoire des algorithmes de négociations et de partage de l'énergie [8]. Nous utilisons la méthode du sac à dos afin de déterminer un optimum local de demande d'énergie en fonction des pronostiques. Le sac-à-dos est résolu par programmation dynamique, sa complexité est alors de  $O(nC)$ , avec  $n$  le nombre d'appareil et  $C$  l'énergie reçue. Nous utilisons la programmation dynamique car le nombre d'appareils et la taille du sac ne sont pas grands.

### Module gestion des pannes

Afin de modéliser un système réaliste, l'intégration des pannes est nécessaire. Nous avons décrit l'évolution de ce module en 3 étapes (cf. figure 4.17).

Dans une première étape, notre simulation fonctionne parfaitement en toutes circonstances et permet notamment la validation de nos algorithmes de partage. Aucune panne ne survient, ni sur le matériel, ni sur l'infrastructure.

Dans la 2ème étape, nous intégrons des pannes aléatoires structurelles. Les arêtes peuvent ainsi disparaître ou réapparaître au gré des pannes et réparations stochastiques, sans stratégie de gestion particulière. Cette étape permet d'une part de prouver la robustesse du réseau de distribution et d'autre part de démontrer que notre algorithme de routage est capable de faire face de manière autonome à la situation.

Enfin, dans la 3ème étape, nous ajoutons les stratégies de gestion connues, qui sont notamment préconisées par ERDF dans le cadre des résolutions de congestion. Cette étape



FIGURE 4.17 – Les différentes étapes du module de gestion des pannes.

finale permet de rendre le système plus proche de la réalité. Nous ajoutons également à ce niveau le module de percolation, permettant une supervision des pannes. Cette dernière étape pourrait ainsi superviser le réseau de manière globale pour détecter et gérer les pannes matérielles. Actuellement, nous n'avons pas les données industrielles nécessaires pour développer ce point. Cependant, il pourrait être ajouté comme module à interfacier avec le réseau de distribution, permettant la détection de comportements anormaux du matériel, et la mise en quarantaine suite à la confirmation d'une panne. L'étude de Kramer et Magee [?] concernant l'architecture d'un système capable de s'auto-gérer couplée à celle de Hagg [?] sur la détection de fautes dans les modélisations multi-agents pourraient servir de base à ce module.

### Module réseau de communication

Le module réseau de communication a pour but de mettre en relation les différentes entités communicantes de notre modèle.

Comme pour tous nos autres modules, nous l'avons découpé en étapes successives d'implémentation (cf figure 4.18).

Au début de la simulation, nous utilisons un agent central (sous forme d'une classe Java unique), qui se chargeait de lire et coordonner les informations des entités du système. A l'étape 2, nous avons intégré des stratégies de communication par échange de messages entre voisins. Ainsi, 2 entités non connectées ne peuvent pas communiquer directement. Grâce à la prétopologie, nous pouvons paramétrer des notions de proximités fines qui représentent bien la complexité de la réalité. Nous avons également instauré un seuil pour le délai de réception des messages, de sorte que, si un message est perdu sur le réseau pour une raison quelconque, cela ne génère pas d'interblocage sur le système. Enfin, à la

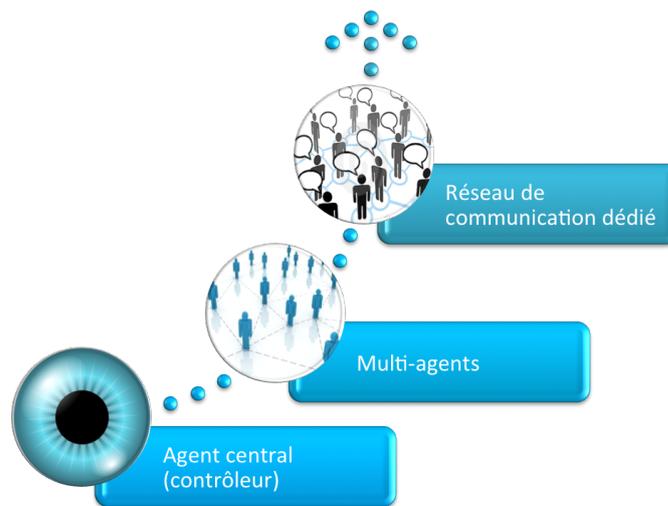


FIGURE 4.18 – Les différentes étapes du module réseau de communication.

dernière étape que nous envisageons dans les travaux futurs, nous allons étudier l'impact de l'utilisation d'un réseau de communication dédié. Quelques études ont déjà été faites démontrant les avantages et les inconvénients des différents supports de communication : wifi, courants porteurs en ligne (CPL), internet, technologie propriétaire... Le Smart Grid impose beaucoup de contraintes au niveau de la sécurité des données et de la vitesse de propagation des informations. Développer un module dans notre simulation pour tester les impacts du réseau de communication permettrait de tester in vitro l'efficacité de la technologie testée.

## 4.3 Simulation : conception, implémentation et optimisation

Dans cette section, nous présentons notre application de simulation d'un éco-quartier, basée sur notre modélisation de Smart Grid. Nous présentons notre environnement de développement et les raisons de son choix, puis nous présentons tous les paramètres nécessaires à notre simulation, avant d'en présenter les résultats.

### 4.3.1 Besoins liés à la complexité du système et choix d'une plateforme de simulation

Dans le cadre de l'implémentation de notre modèle, nous avons été confrontés à l'hétérogénéité des acteurs de notre modèle, et à leurs paramètres propres. Nous avons étudié les différentes solutions d'implémentations possibles. Après quelques tests sous Anylogic et Jade, en raison des délais imposés par le projet dans lequel ce travail s'inclue, nous avons choisi de développer notre propre programme de simulation en Java, tout en s'inspirant des possibilités offertes par Anylogic et Jade. En effet, la modularité offerte par JADE et

Anylogic n'était pas suffisante pour nous permettre d'intégrer immédiatement nos différents agents et leurs paramètres. L'intégration de la librairie logicielle PretopoLIB [?] nous obligeait par exemple à travailler sur des structures de données non gérées par Anylogic. Travailler avec un de ces deux outils aurait été possible, mais le temps passé à rajouter des surcouches nous permettant d'intégrer notre modèle était équivalent voire supérieur au temps nécessaire au développement d'un programme à part. Vu les délais impartis, nous avons donc choisi de développer entièrement notre programme de simulation.

### 4.3.2 Vérification et validation des résultats

Une des étapes les plus importantes et les plus difficiles dans la construction d'un modèle est la vérification et la validation de la simulation du modèle. Le but du processus de validation est double [?] :

- d'une part, de produire un modèle qui représente le véritable comportement du système, et suffisamment réaliste pour que le modèle soit utilisé comme substitut du système actuel pour analyser et réaliser des tests de performance en modifiant le paramétrage *in silico*
- d'autre part, d'augmenter la crédibilité du modèle à un niveau acceptable pour qu'il soit utilisé par des managers et des décideurs.

Le processus de vérification et de validation repose sur 2 étapes [?] :

- La vérification concerne l'architecture et le développement informatique du modèle et vise à repérer les erreurs de la simulation [?]. Elle consiste à comparer le modèle conceptuel à la représentation informatique implémentée lors de la conception. La vérification répond aux questions : Le modèle est-il correctement implémenté dans la simulation ? les paramètres d'entrée et la structure logique du modèle sont-ils correctement représentés ?
- La validation concerne l'exactitude du modèle. Elle vise à confirmer que le modèle est une représentation pertinente du système réel. La validation est en général réalisée via la calibration du modèle, un processus itératif de comparaison au système réel permettant d'ajuster et améliorer le modèle. Ce processus est répété jusqu'à obtention d'une précision des résultats jugée acceptable.

#### Vérification de la structure du modèle

Nous avons implémenté la structure de la simulation (figure 4.20) en nous basant directement sur la structure du modèle (figure 4.19). Ainsi, la structure logique du modèle est parfaitement respectée.

#### Validation des algorithmes

Une des étapes de validation repose sur l'exactitude des résultats des algorithmes implémentés. Pour cela, plusieurs méthodes sont possibles, basées sur la comparaison entre les résultats du système réel à ceux du modèle simulé.

La comparaison quantitative compare les résultats issus de la simulation à ceux du système réel, à paramètres d'entrée équivalents. Notre simulation permet un suivi pas à

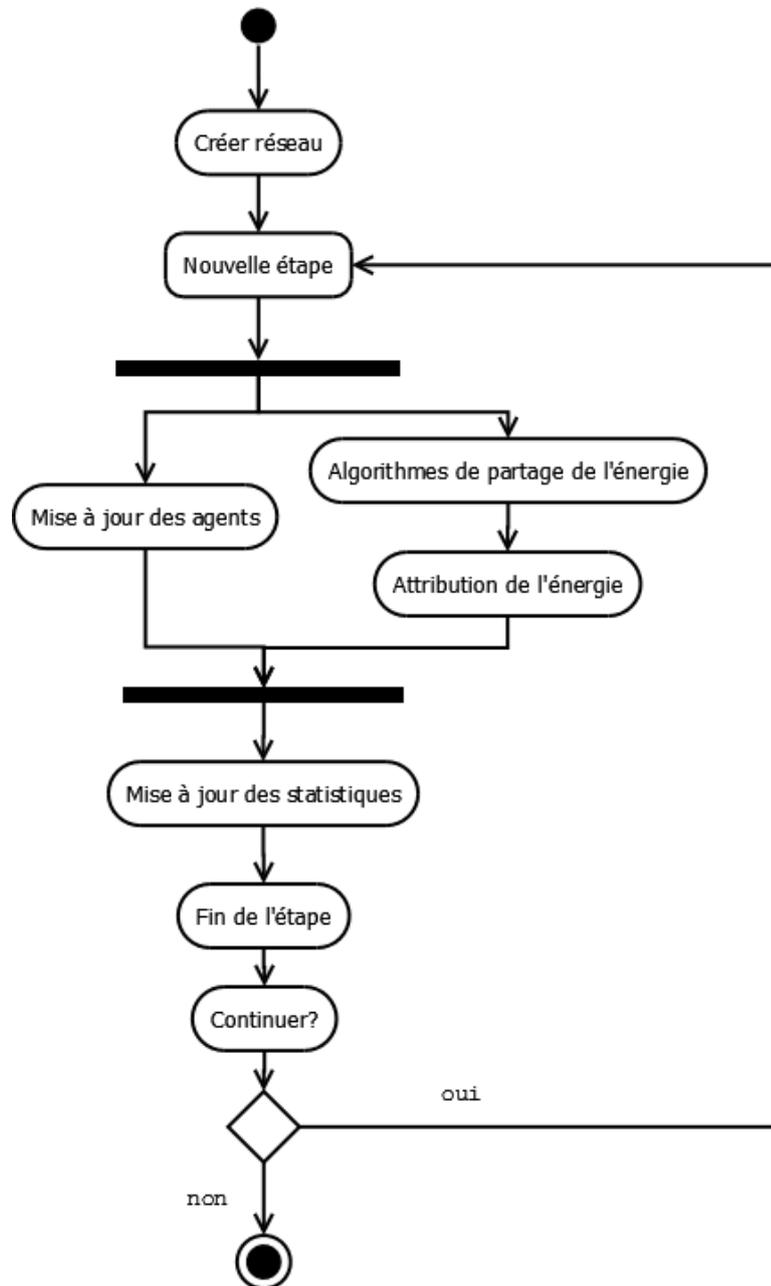


FIGURE 4.19 – Structure du modèle.

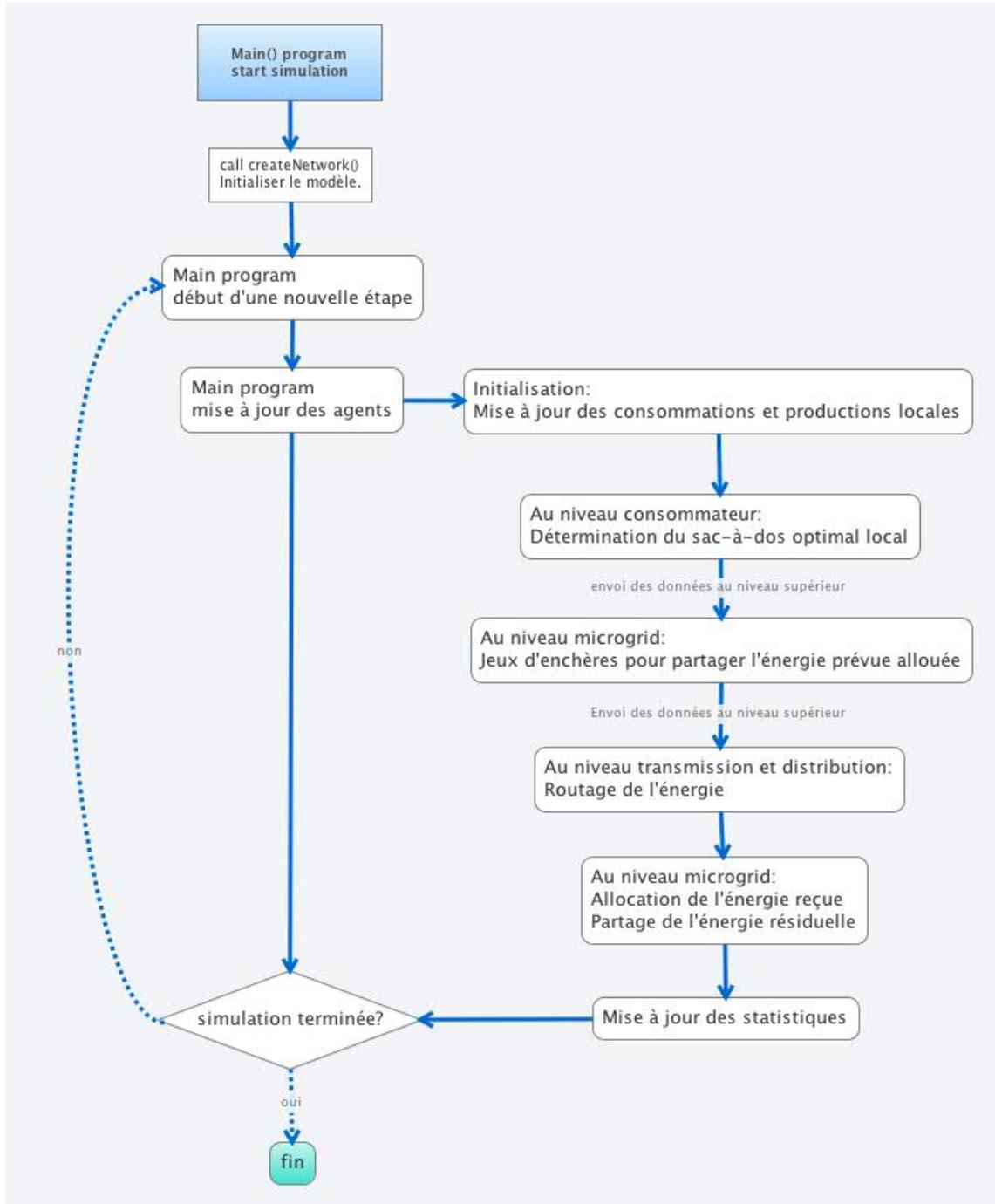


FIGURE 4.20 – Structure de notre simulation Java.

pas de l'évolution du système (figure 4.21), elle permet donc une comparaison précise des résultats en tout point (figure 4.22).

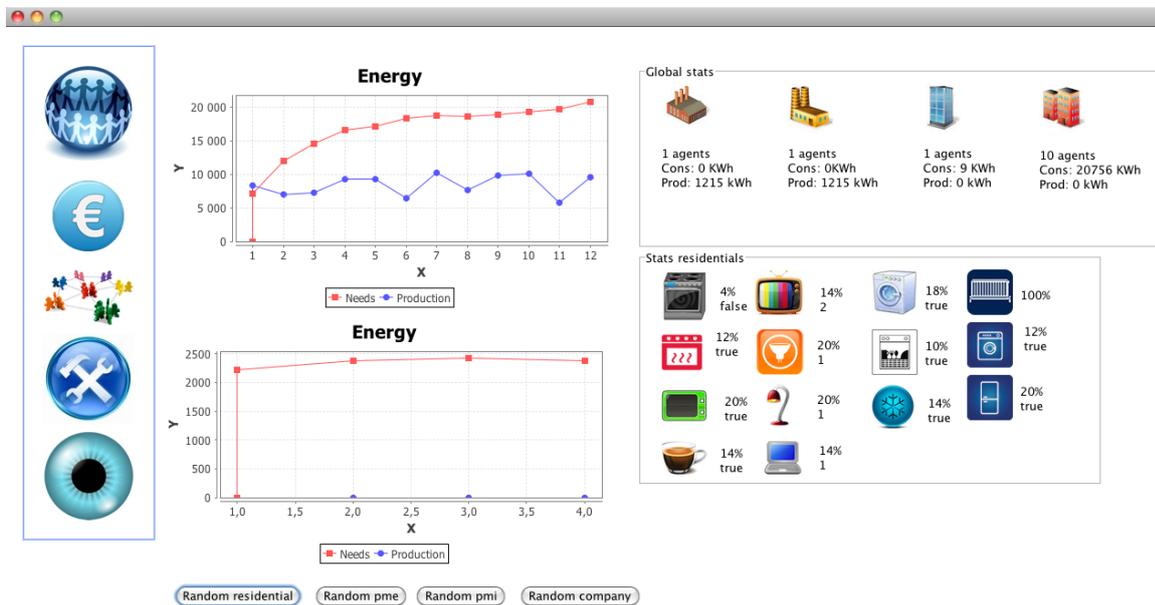


FIGURE 4.21 – Suivi des consommations et productions globales et locales.

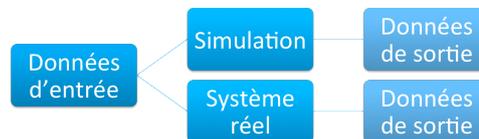


FIGURE 4.22 – Comparaison quantitative des résultats et validation.

La comparaison qualitative consiste à analyser les représentations des paramètres et des comportements du système. Par exemple, pour simuler un comportement utilisateur, il se peut que des heuristiques aient été implémentées dans la simulation. Dans notre cas, nous avons implémenté la possibilité de paramétrer le nombre d'agents de chaque type dans le réseau, tout en proposant une répartition par défaut proportionnelle à celle du réseau électrique français (figure 4.23).

### Performance de la simulation

La complexité au sens large est une mesure des ressources nécessaires pour analyser, simuler et développer le modèle [?]. D'un point de vue informatique, les performances sont mesurées principalement par le temps d'exécution et les ressources mémoire et processeur nécessaires. Il faut trouver un compromis entre la complexité du modèle et le temps

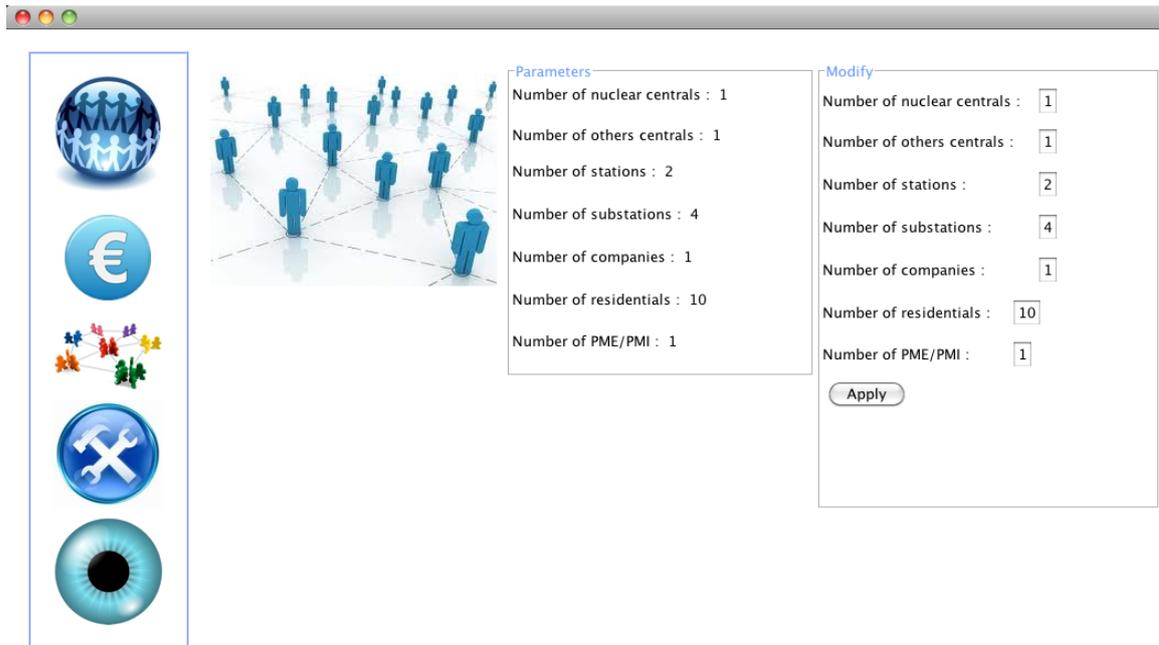


FIGURE 4.23 – Paramétrage de la structure du réseau au sein de la simulation.

d'exécution de la simulation car ces 2 paramètres sont directement liés à la précision des résultats obtenus.

Le Smart Grid impose une réactivité en temps réel. Le temps maximal de réaction reconnu par les industriels suite à une demande subite d'énergie est de l'ordre de 5 minutes.

Dans le cas d'une simulation multi-agents, chaque entité n'a qu'une connaissance locale du réseau. Ainsi, notre modèle repose sur des interactions locales simultanées, et sur des actions locales bornées (nombre maximal d'itération défini, etc.), ce qui nous assure une réactivité optimale et une convergence à chaque action.

### Architecture de la simulation

La figure 4.24 est le diagramme de classes de notre simulation java. Nous avons implémenté une architecture pseudo-MVC (modèle-vue-contrôleur), reposant sur une classe objet principale : la classe agent. Cette classe instancie les principaux paramètres et fonctions liés aux entités du système, puis nos autres classes objet héritent de cette classe avec des spécificités supplémentaires : consommateurs, stations, sous-stations, etc...

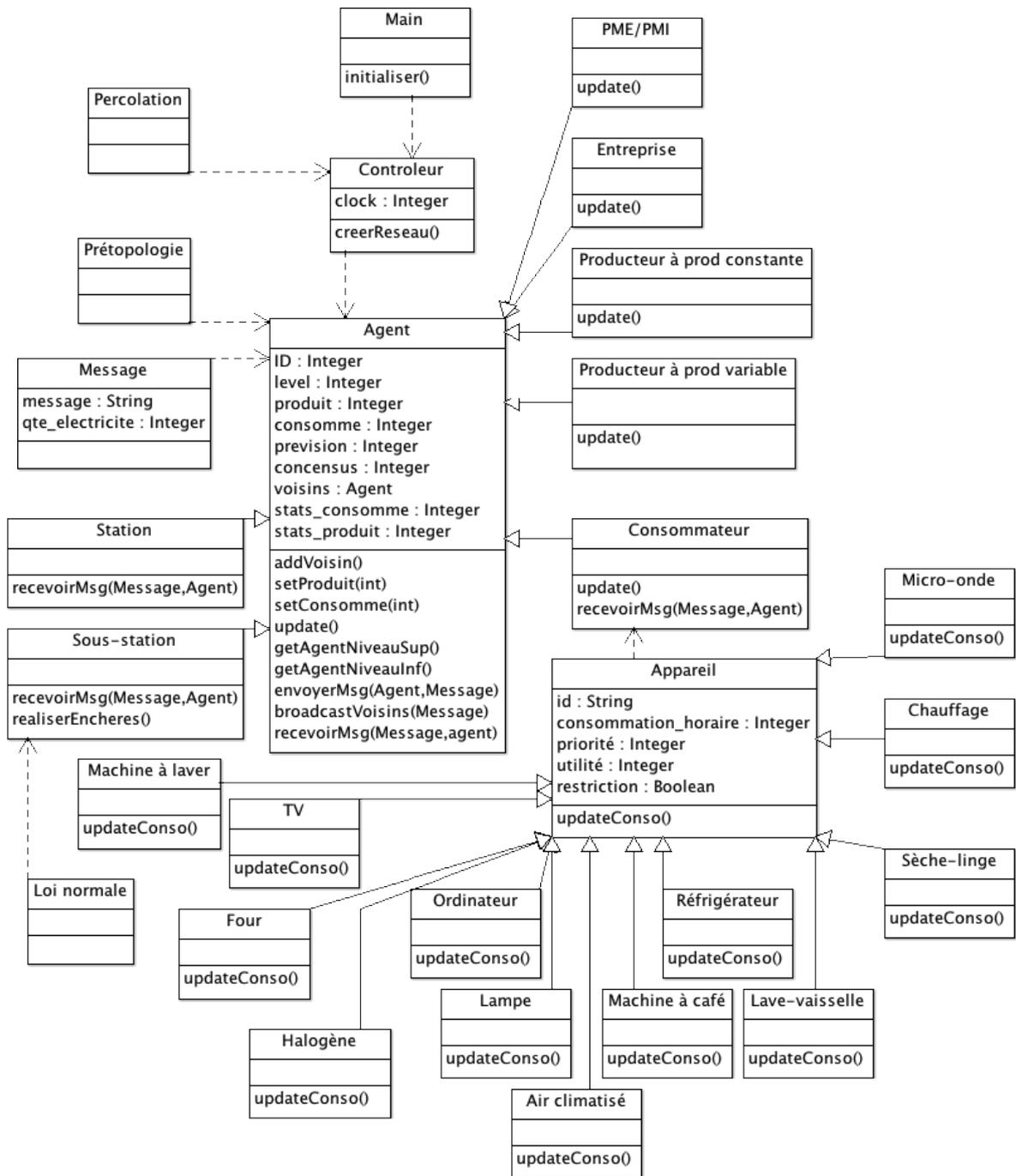


FIGURE 4.24 – Diagramme de classes de la simulation.

# Chapitre 5

## Perspectives

Ce chapitre est consacré aux perspectives de notre travail. Nous y proposons notamment une définition mathématique théorique du concept de Smart Grid, issue de nos réflexions liées à la dialectique entre modélisation et simulation. Nous présentons également la théorie de la géostatistique comme solution complémentaire à notre algorithme de clustering pour l'approximation de données, la génération de jeux de données réalistes, et comme détection des pannes indétectables par la seule théorie de la percolation.

### Sommaire

---

<b>5.1</b>	<b>Proposition de définition mathématique de Smart Grids . . .</b>	<b>88</b>
5.1.1	Graphe dynamique et structure des lignes électriques . . . . .	88
5.1.2	Problème de flots pour la vision globale du Smart Grid . . . . .	89
5.1.3	Sac à dos multi-objectifs temporalisé et définition du problème de partage de l'énergie . . . . .	89
<b>5.2</b>	<b>Améliorations de notre modélisation de partage de ressources</b>	<b>90</b>
5.2.1	Une étude approfondie de la théorie des jeux . . . . .	90
5.2.2	Modélisation de la perte énergétique par la prétopologie . . . . .	91
5.2.3	Diminution de la complexité temporelle de la simulation . . . . .	91
5.2.4	Extension du modèle à d'autres problématiques . . . . .	91
<b>5.3</b>	<b>Introduction aux outils d'analyse pour les Smart Grids . . . .</b>	<b>92</b>
5.3.1	La géostatistique pour l'analyse de flux dans les Smart Grids .	92
	Variable aléatoire et variogramme pour représenter les niveaux de production ou consommation . . . . .	93
	Krigeage et cokrigeage pour interpoler la consommation ou la production en cas d'échec des méthodes de prévision classiques . . . . .	94

---

## 5.1 Proposition de définition mathématique de Smart Grids

Comme vu précédemment, le Smart Grid est un système complexe et hétérogène. Il n'existe pas de définition pratique universelle car sa vision dépend des différents acteurs du système et il existe des différences entre la vision générale du Smart Grid et celle des industriels. Nous proposons une définition théorique mathématique, basée sur la définition pratique générale, que nous avons développée en chapitre 1. Cette définition n'a pas de but de modélisation. Utilisée en tant que modélisation, la simulation d'un tel modèle est trop complexe pour respecter les conditions de réactivité et d'autogestion du système. Elle est juste une expression mathématique simple du système.

### 5.1.1 Graphe dynamique et structure des lignes électriques

Le Smart Grid consiste en une évolution de la grille électrique actuelle. Nous allons donc passer d'un flux électrique unidirectionnel à des flux communicants hétérogènes bidirectionnels.

La représentation du réseau par un système de matrices n'est pas envisageable en raison de la multiplicité des données. L'utilisation de la théorie des graphes est donc une meilleure option. En effet, grâce à cette théorie nous allons pouvoir représenter les relations bidirectionnelles entre les différentes entités du Smart Grid. L'utilisation d'un graphe orienté permettra de gérer les portions du réseau où le flux serait encore unidirectionnel.

De plus, il est nécessaire de connaître le type de ligne électrique du réseau (haute tension, moyenne tension, etc.), la valuation du graphe est alors obligatoire afin de pouvoir gérer la tension maximale assignée à chaque portion de ligne.

Le Smart Grid peut être soumis à des pannes. Par conséquent, des arêtes peuvent aléatoirement être rompues pour un temps variable. De plus, certaines sources d'énergie telles que les énergies renouvelables ont une production instable dépendant de conditions externes non prévisibles. Par exemple, les panneaux photovoltaïques étant montés en série, une simple feuille morte tombée sur un panneau coupe la production de l'ensemble de l'installation. Les arêtes peuvent donc apparaître ou disparaître selon les pannes (cas des arêtes de transport et distribution) ou selon la production d'énergie (cas des arêtes sortantes des producteurs).

**Définition 5.1 (Structure électrique du Smart Grid)** Soit  $G=(V,E,v)$  un graphe valué dynamique représentant la structure du Smart Grid, composé de :

- un ensemble  $V = \{x_1, x_2, \dots, x_n\}$  représentant les entités physiques du Smart Grid (stations, sous-stations, consommateurs, producteurs...).
- une famille  $E = (a_1, a_2, \dots, a_n)$  d'éléments du produit Cartésien  $V \times V = \{(x, y) / x \in V, y \in V\}$  représentant les arcs entre les agents voisins ie les lignes électriques.
- une fonction de valuation  $v$  pour modéliser la tension maximale des lignes telle que :  
 $v : E \rightarrow \mathbb{R}$   
 $(x, y) \rightarrow v(x, y)$
- une fonction de probabilité  $p$  pour modéliser la probabilité des arêtes telle que :  
 $p : E \rightarrow \mathbb{R}$

$$(x, y) \rightarrow p(x, y)$$

### 5.1.2 Problème de flots pour la vision globale du Smart Grid

Au niveau global, le Smart Grid pour être vu comme un problème de flots : la distribution de l'énergie est alors assimilée à une recherche de flot max pour maximiser l'utilisation des arêtes. Nous pouvons définir le problème comme suit :

$$0 \leq f(s, v) \leq u(s, v) \quad (5.1)$$

$$\sum_{v \in V} f(s, v) = \sum_{v \in V} f(v, s) \quad (5.2)$$

avec  $f(s, v)$  le flux de  $s$  à  $v$  et  $u$  la fonction de valuation des arêtes.

Pour modéliser les capacités des producteurs électriques, on introduit un sommet supplémentaire  $s$  qui sera la source du réseau, et des arcs entre  $s$  et chaque producteurs électriques ayant une capacité égale à la production électrique du sommet cible pour assurer la propriété de conservation des flux. Si on veut mesurer l'énergie distribuée à chaque consommateur, on introduit un sommet supplémentaire  $p$ , qui sera le puits, et des arcs de chacun des consommateurs vers  $p$ . Pour imposer que les demandes soient satisfaites, on munit ces arcs d'une capacité inférieure ou égale à la demande. Cette structure permet de mettre en place un réseau sur lequel il s'agit de déterminer un flot maximisant l'utilisation des capacités des arêtes.

### 5.1.3 Sac à dos multi-objectifs temporalisé et définition du problème de partage de l'énergie

Le partage local de l'énergie peut être défini comme un problème de sac-à-dos. Cependant, le sac-à-dos classique de recherche opérationnelle ne suffit pas à définir le partage de l'énergie du Smart Grid. En effet, le Smart Grid est dynamique, et chaque flux d'énergie est conditionné de façon temporelle et physique, *ie* il a une date de début, de fin, et la valeur d'attribution finale de l'énergie dépend du seuil d'énergie transportable sur le réseau.

Nous définissons donc le partage de l'énergie sur le Smart Grid par un sac-à-dos multi-objectifs, auquel nous rajoutons une variable de temps. La figure 5.1 représente une contrainte du sac à dos temporalisé : le partage de l'énergie locale reçue.

Il existe de multiples façons de remplir le sac à dos. Pour décrire l'une d'elles il faut indiquer pour chaque élément s'il est pris ou non. On peut utiliser un codage binaire : l'état du  $i$ -ème élément vaudra  $x_i = 1$  si l'élément est mis dans le sac, ou  $x_i = 0$  s'il est laissé de côté. Une façon de remplir le sac est donc complètement décrite par un vecteur, appelé vecteur contenu, ou simplement contenu :  $X = (x_1, x_2, \dots, x_n)$ ; et le poids associé, ainsi que la valeur associée, à ce remplissage, peuvent alors être exprimés comme fonction du vecteur contenu. Pour une liste de flux  $X$  donnée, la valeur totale contenue dans le sac est naturellement :  $z(X) = \sum_{i, x_i=1} p_i = \sum_{i=1}^n x_i p_i$

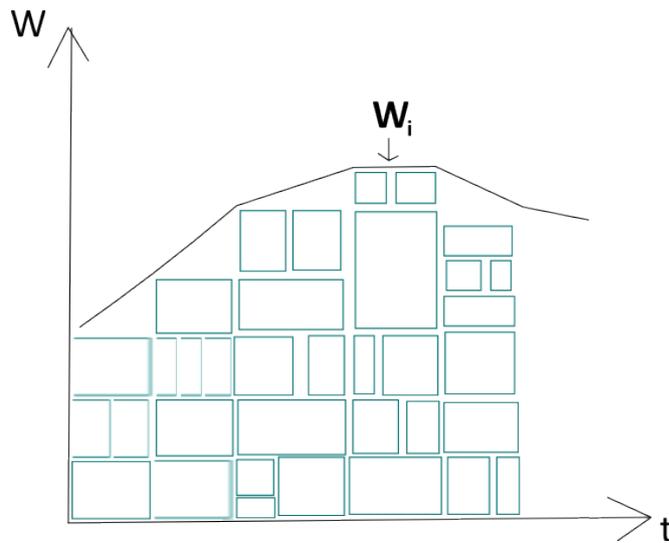


FIGURE 5.1 – Sac à dos bi-dimensionnel : partage local de l'énergie allouée.

### Définition 5.2 (Sac à dos multi-objectifs temporalisé et partage de l'énergie)

Les flux électriques sont numérotés par l'indice  $i$  variant de 1 à  $n$ . Soient  $w_i$  et  $p_i$  respectivement le poids et la valeur du flux numéro  $i$ , et  $W$  la capacité du sac. Les arêtes du système sont numérotées par l'indice  $k$  variant de 1 à  $m$ . Le problème peut être formulé comme la recherche d'un vecteur contenu (les composantes valant 0 ou 1), réalisant le maximum de la fonction valeur totale, en tout  $T$  du système tel que  $t_{i_{debut}} \leq T \leq t_{i_{fin}}$ , sous les contraintes suivantes, :

- contrainte liée à la limitation de chaque arête du système :  $\sum_{i=1, k=1}^{n, m} \frac{x_i w_i}{t_i} \leq W_k$  soit  $p_k(X) \leq W_k$
- contrainte de partage au niveau de chaque consommateur :  $\sum_{i=1, r=1}^{n, m} \frac{x_i w_i}{t_i} \leq W_r$  soit  $p_r(X) \leq W_r$

Cependant, cette définition permet de décrire uniquement le fonctionnement global du Smart Grid, mais ne peut pas être utilisée seule pour simuler le système : elle ne prend en compte qu'une vision simplifiée du Smart Grid, sans prise en compte de l'humain ou de critères complexes tels que les pertes énergétiques sur les lignes haute tension. C'est pourquoi notre modélisation multi-agents est nécessaire. Par contre, nos algorithmes locaux sont basés sur la définition locale de sac à dos pour le partage de l'énergie à chaque niveau du Smart Grid [8].

## 5.2 Améliorations de notre modélisation de partage de ressources

### 5.2.1 Une étude approfondie de la théorie des jeux

La simulation actuelle ne calcule pas d'équilibre de Nash pour le partage de l'énergie. Dans la continuité de ces recherches, nous pourrions envisager un calcul parallèle

d'équilibre dans le cadre des réseaux de jeux du Smart Grid.

### 5.2.2 Modélisation de la perte énergétique par la prétopologie

Notre modèle intègre la prétopologie par l'intégration de critères booléens ou valués dans des ensembles prétopologiques dédiés. Or, comme l'indique la figure 5.2.2 la capacité des lignes de transport de l'énergie n'est pas constante. Dans le cadre de la modélisation d'éco-quartier, ce phénomène n'est pas perceptible mais l'impact de la perte énergétique sur les lignes haute tension est préjudiciable : 70% de la capacité initiale est perdue sur 400km. Les pertes représentent environ 33 TWh annuels sur le réseau électrique français [?].

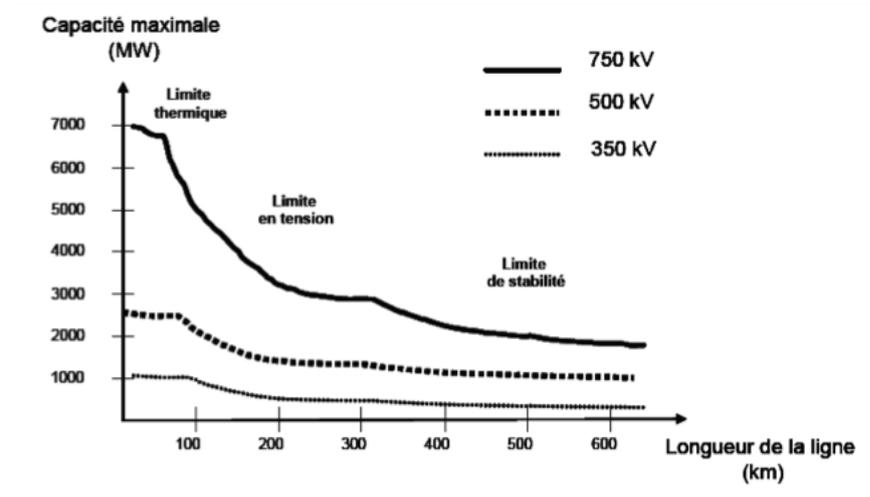


FIGURE 5.2 – Longueur des lignes et pertes énergétiques.

L'ajout d'un espace prétopologique pour la modélisation de cette perte énergétique permettrait de créer un modèle plus proche de la réalité pour la modélisation de systèmes à grande échelle.

### 5.2.3 Diminution de la complexité temporelle de la simulation

Concernant la simulation, une programmation distribuée permettrait de simuler le modèle à plus grande échelle. La figure 5.2.3 présente un exemple de parallélisation possible, en associant un *thread* à chaque agent du système.

### 5.2.4 Extension du modèle à d'autres problématiques

Nous avons proposé ce modèle dans le cadre de l'étude d'un Smart Grid, mais il pourrait également s'appliquer à d'autres systèmes socio-techniques complexes. Par exemple, il pourrait être un modèle dans le domaine de la santé où la ressource partagée abstraite

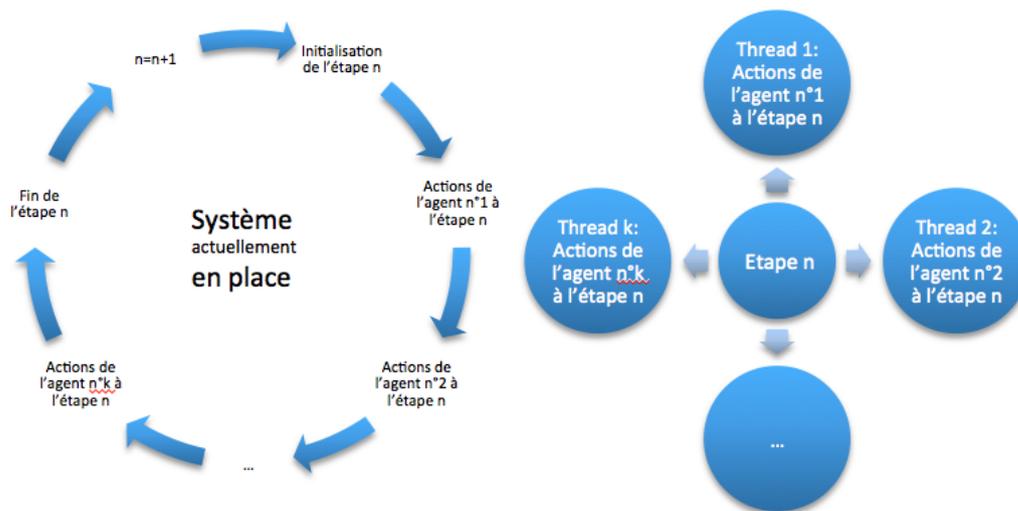


FIGURE 5.3 – A gauche notre simulation, et à droite une parallélisation possible.

serait les soins, et les informations et communications entre agents seraient les médicaments, les rendez-vous... Une telle modélisation permettrait d'étudier la proportion et la répartition des médicaments génériques, ou encore d'analyser des informations relevant de la propriété d'émergence du réseau.

## 5.3 Introduction aux outils d'analyse pour les Smart Grids

Dans cette section, nous discutons autour de l'usage de théories comme outils d'analyse applicables au cadre des Smart Grids. Nous présentons en particulier une application de la théorie des géostatistiques pour l'analyse des flux dans les Smart Grids.

### 5.3.1 La géostatistique pour l'analyse de flux dans les Smart Grids

La géostatistique est une théorie couplant mathématiques et sciences de la Terre. Son utilisation principale est l'estimation des gisements miniers, mais ses domaines d'applications sont multiples. En réalité, tout élément spatialisé peut être étudié via la géostatistique. Ainsi, la géostatistique a des applications en géologie, en biologie, en santé, en climatologie... et nous allons montrer qu'il est également possible de l'appliquer dans le cadre des Smart Grids. En effet, tous ces domaines de recherche ont en commun le fait que les données sont géospatialisées et ne sont ni indépendantes, ni identiquement distribuées.

### Variable aléatoire et variogramme pour représenter les niveaux de production ou consommation

Dans notre thème Smart Grid, une variable aléatoire peut représenter un niveau de consommation, ou de production d'un élément du Smart Grid. En effet, nous connaissons la courbe moyenne des flux électriques, mais nous ne pouvons pas connaître la mesure précise d'un flux en un point précis du Smart Grid avant de l'avoir mesuré. La fonction de masse représente alors les probabilités de distribution des valeurs connues de consommation et production. Dans le cadre de la géologie et de la climatologie, une variable aléatoire peut représenter des valeurs comme la teneur de cuivre d'une carotte de 1m, l'épaisseur d'une veine minéralisée, la concentration d'un polluant dans l'eau souterraine ou encore le pH de l'eau de pluie.

**Définition 5.3** *Variable aléatoire (v.a.) : fonction dont les résultats possibles sont connus mais dont le résultat final ne peut être déterminé, à priori, avant d'effectuer la mesure.*

Sans connaître la valeur que prendra le résultat final, on peut parfois connaître la probabilité qu'une variable aléatoire prenne chacun des résultats possibles. Les probabilités sont décrites par une fonction dite de densité pour les variables aléatoires continues, ou fonction de masse pour les variables aléatoires discrètes.

Propriété :

–  $f_x(x) \geq 0$ , toute probabilité est positive.

–  $\int_{-\infty}^{+\infty} f_x(x) dx = 1$  l'intégrale de la fonction de densité donne 1

–  $\int_a^b f_x(x) dx = P(a \leq X \leq b)$  probabilité que  $x$  prenne une valeur comprise entre  $[a$  et  $b]$ .

La théorie de la géostatistique introduit également la notion de mesures de tendance centrale.

**Définition 5.4** *Mesure de la tendance centrale :*

– *mode* :  $x$  tel que  $f_x(x)$  est maximum

– *médiane* :  $x$  tel que  $P(X < x) = 0.5$

– *moyenne (ou espérance mathématique)* :  $\mu_X$  ou  $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$ .

Dans le cadre des Smart Grids la tendance centrale mode représente la valeur de consommation ou production la plus représentée dans la zone étudiée, la tendance centrale médiane est la valeur médiane des flux d'énergie électrique, et la tendance centrale moyenne est la moyenne des demandes ou productions électriques du système.

La géostatistique introduit également une notion de mesures de dispersion, applicable à l'analyse des valeurs de consommation ou production d'un Smart Grid.

**Définition 5.5** *Mesure de dispersion :*

– *Variance* :  $\sigma_X^2 = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$

- *Ecart-type* :  $\sigma_X = \sqrt{\sigma_X^2}$
- *Asymétrie* :  $E\left[\left(\frac{X - E[X]}{\sigma_X}\right)^3\right]$
- *Aplatissement* :  $E\left[\left(\frac{X - E[X]}{\sigma_X}\right)^4\right]$

Toutes ces quantités sont a priori inconnues. On doit donc les estimer à partir d'un ensemble d'observations (appelé échantillon) à partir duquel on peut construire des estimateurs.

**Définition 5.6** *Estimateurs :*

- *de la moyenne* :  $\frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$
- *de la variance* :  $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \bar{\sigma}^2$  ou  $\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2$
- *de la fonction de densité* : *histogramme*
- *de la fonction de densité cumulative* : *courbe des fréquences cumulées*  $F_x(x) = P(X \leq x)$  estimée par :  $\text{rang}(x_i)/n$

Un des éléments importants des estimateurs est d'être sans biais : ils ont la même espérance mathématique que la quantité qu'ils cherchent à évaluer.

À partir des variables aléatoires, la géostatistique définit le variogramme qui permet de décrire une valeur sur une distance de dispersion  $\epsilon$  définie autour du point de mesure.

**Définition 5.7** *Variogramme* : Soit deux points  $x$  et  $x+h$  séparés d'une distance  $h$ . Le flux en  $x$  est une variable aléatoire  $Z(x)$ . Le flux en  $x+h$  aussi,  $Z(x+h)$ . La différence entre les valeurs prises par ces deux variables aléatoires est  $Z(x)-Z(x+h)$ . On appelle variogramme la demi-variance de cette différence :  $\gamma(x, x+h) = 0,5 * \text{Var}(Z(x) - Z(x+h))$

Chaque phénomène géologique, et chaque type de portion de Smart Grid possède un variogramme qui lui est propre. Ainsi, en géologie, un gisement d'or présente un variogramme erratique avec un fort effet de pépité et une faible portée. Un gisement de cuivre a un variogramme linéaire à l'origine, avec faible effet de pépité et grande portée. Si on fait les mêmes rapprochements dans le cadre d'un réseau de distribution d'électricité, le maillage de production central présentera un fort effet de pépité avec une faible portée, la "pépité" étant assimilée à un agent ponctuel de type centrale nucléaire produisant de grandes quantités d'énergies. Les résidentiels auront un faible effet de pépité et une portée relativement grande comparativement au maillage de production. En effet, un résidentiel est souvent intégré au sein d'un quartier résidentiel, ce qui explique sa portée. Il en est de même pour les parcs à éoliennes.

Lorsque le point de mesure est trop éloigné du point à estimer, le variogramme ne peut pas efficacement décrire la valeur, il faut alors l'estimer grâce au krigeage.

**Krigeage et cokrigeage pour interpoler la consommation ou la production en cas d'échec des méthodes de prévision classiques**

Le krigeage est une méthode d'estimation linéaire, sans biais, basée sur l'interpolation spatiale de valeurs connues qui sont la variance et l'espérance mathématique. Dans le cadre

de la géologie, elle est utilisée pour estimer la teneur en minerai d'une zone géographique, à partir des mesures effectuées dans les régions voisines.

Le krigeage est basé sur l'interpolation du variogramme de plusieurs points spatialement délimités. Il permet, par l'application d'une méthode linéaire non-biaisée, d'éliminer dans une série les valeurs improbables ou incohérentes en se basant sur les valeurs des points voisins.

**Définition 5.8** *Krigeage* : le krigeage est défini comme un modèle de régression classique :  $Z(s) = \mu(s) + \epsilon(s)$  où  $\mu(s)$  désigne une structure déterministe pour l'espérance de  $Z$  et  $\epsilon(s)$  est une fonction aléatoire, stationnaire, d'espérance nulle et de structure de dépendance spatiale connue. [?]

Dans le cadre des Smart Grid, nous envisageons deux contributions de cette méthode. Premièrement, le krigeage peut être utilisé pour remplacer des données manquantes à partir des données de compteurs dits Smart meters (compteurs intelligents qui fournissent l'information en temps réel). En effet, la grille électrique connaît régulièrement des situations exceptionnelles où les méthodes classiques de prévision de consommation atteignent leurs limites. C'est par exemple le cas lors des vagues de chaleur ou de froid exceptionnelles et non prévues qui induisent des surconsommations d'électricité exceptionnelles. Il est alors possible par cette méthode d'estimer la valeur de consommation d'un résidentiel n'étant pas équipé d'un compteur, à partir d'une interpolation par krigeage de la valeur de consommation de ses voisins. Cette méthode peut également être utilisée pour la prévision de production des énergies renouvelables vu le caractère intermittent et non prévisible de celles-ci. Deuxièmement, le krigeage peut également permettre la détection des dysfonctionnements des compteurs. Si la valeur envoyée par le compteur est aberrante par rapport à ses voisins, l'application du modèle mathématique défini par le krigeage permet de vérifier s'il s'agit d'un vrai effet pépète ou d'une panne du compteur.



## Conclusion générale

Nous avons exposé dans ce mémoire de thèse le principe et les notions importantes liées aux systèmes socio-techniques complexes. Nous avons notamment introduit les problématiques rencontrées dans le système socio-technique complexe qu'est le Smart Grid, et proposé une solution à plusieurs d'entre elles.

Le premier apport de cette thèse concerne la notion d'influence par les propriétés de connexités et par l'usage d'une notion de proximité plus réaliste que la distance métrique. En effet, nous avons introduit une mesure de proximité universelle traitant le contenu absolu en information des données. Ensuite, nous avons démontré que cette mesure peut s'adapter à une méthode de catégorisation et d'analyse efficace.

La deuxième contribution concerne l'élaboration de nouvelles solutions théoriques pour la modélisation de Smart Grid. Nous proposons une modélisation intégrative stigmergique basée sur l'intégration de plusieurs théories permettant de prendre en compte chacun des aspects des problèmes de modélisation des smart Grids.

La troisième contribution consiste à l'implémentation d'une simulation autonome à partir du modèle proposé. Nous y présentons la dialectique entre modélisation et simulation et les adaptations nécessaires liées au développement informatique du modèle.

Enfin notre dernier apport concerne une proposition de définition mathématique du concept de Smart Grid. Cette définition est proposée dans notre chapitre d'ouverture car il s'agit d'une réflexion issue de nos recherches et réalisations. Dans les travaux futurs, nous envisageons dans un premier temps de proposer des extensions de notre modélisation par l'intégration de la théorie de la percolation comme superviseur global de l'état du réseau et comme détecteur de panne; puis la calibration de certains paramètres comme la prise en compte de la perte énergétique sur les lignes électriques.



## Liste des tableaux

2.1	Matrice de confusion. . . . .	37
2.2	Test d'efficacité en augmentant $n$ . . . . .	39
2.3	Efficacité de notre méthode et augmentation du paramètre $\mu$ . . . . .	39
2.4	Test d'efficacité sur graphes clairsemés. . . . .	40
2.5	Complexité temporelle pratique pour les graphes générés. . . . .	40



# Table des figures

1.1	Propagation d'une épidémie parmi une population d'individus. . . . .	8
1.2	Réseau de citation des articles. . . . .	9
1.3	Réseau technologique : la structure logique d'internet. . . . .	9
1.4	Réseau biologique. . . . .	10
1.5	Réseau socio-technique. . . . .	10
1.6	Comparaison entre graphe aléatoire et réseau invariant d'échelle. . . . .	13
1.7	Comparaison entre les degrés de distribution d'un réseau invariant d'échelle et d'un graphe aléatoire. . . . .	14
1.8	Réseau Small World. . . . .	14
1.9	Etude du Nebraska de Milgram. . . . .	15
1.10	Augmentation de la demande en électricité. . . . .	17
1.11	Prévisions d'augmentation de la demande en électricité (source RTE). . . . .	17
1.12	Dernière étape dans la mise en place du Smart Grid : ajout de capteurs et compteurs intelligents. . . . .	19
1.13	La répartition de la production énergétique française. . . . .	21
1.14	Courbe de production journalière de l'énergie photovoltaïque. . . . .	22
1.15	Les différents consommateurs français. . . . .	23
1.16	Réseaux complexes des Smart Grids. . . . .	24
1.17	Les différents organismes de normalisation. . . . .	26
2.1	Classification hiérarchique des composants. . . . .	33
2.2	Exemple d'exécution de notre algorithme avec $p = 4$ . . . . .	34
2.3	Phase de suppression des communautés trop petites : p-SCC $\rightarrow$ UCC. . . . .	35
3.1	Sous-systèmes identifiés du Smart Grid. . . . .	47
3.2	Les réseaux de jeux pour modéliser les négociations autour de l'énergie. . . . .	51
3.3	Adherence de A. . . . .	52
3.4	Interieur de A. . . . .	52
3.5	Opérations présentant la dynamique des réseaux[?]. . . . .	52
3.6	Familles . . . . .	54
3.7	Les différents types de percolation. . . . .	58
3.8	Définition du seuil de diffusibilité du réseau. . . . .	59
3.9	Définition du seuil de brownout du réseau. . . . .	59

3.10	Diagramme d'activités UML simplifié du modèle. . . . .	60
3.11	Diagramme d'activités UML des algorithmes locaux. . . . .	61
3.12	Diagramme d'activités UML du système complet selon les niveaux. . . . .	63
4.1	Diagramme des cas d'utilisation. . . . .	68
4.2	Diagramme d'ontologie. . . . .	69
4.3	Granularité temporelle de la simulation. . . . .	70
4.4	Granularité spatiale de la simulation : un agent pour chaque entité. . . . .	71
4.5	Granularité spatiale de la simulation : un agent pour modéliser un ensemble d'entités. . . . .	71
4.6	Les différents niveaux conceptuels. . . . .	72
4.7	Les stratégies d'intégration des énergies renouvelables. . . . .	73
4.8	Les différents modules composant notre modèle. . . . .	74
4.9	Les différentes étapes du module réseau de transport. . . . .	74
4.10	Les 3 niveaux du réseau de transmission et distribution. . . . .	75
4.11	Un exemple de réseau simulé. . . . .	75
4.12	Les différentes étapes du module de simulation de la production et de la consommation. . . . .	76
4.13	Les utilisations énergétiques du consommateur. . . . .	76
4.14	La répartition des appareils dans les foyers français. . . . .	77
4.15	La durée de marche quotidienne des appareils d'un foyer. . . . .	77
4.16	Les différentes étapes du module de distribution et partage. . . . .	78
4.17	Les différentes étapes du module de gestion des pannes. . . . .	79
4.18	Les différentes étapes du module réseau de communication. . . . .	80
4.19	Structure du modèle. . . . .	82
4.20	Structure de notre simulation Java. . . . .	83
4.21	Suivi des consommations et productions globales et locales. . . . .	84
4.22	Comparaison quantitative des résultats et validation. . . . .	84
4.23	Paramétrage de la structure du réseau au sein de la simulation. . . . .	85
4.24	Diagramme de classes de la simulation. . . . .	86
5.1	Sac à dos bi-dimensionnel : partage local de l'énergie allouée. . . . .	90
5.2	Longueur des lignes et pertes énergétiques. . . . .	91
5.3	A gauche notre simulation, et à droite une parallélisation possible. . . . .	92

# Bibliographie

- [1] Murat Ahat, Soufian Ben Amor, Marc Bui, and Marie-Françoise Courel. Pollution modeling and simulation with multi-agent and pretopology. In *Proceedings of the International Conference on Complex Sciences : Theory and applications (Complex'2009)*, Shanghai, Chine, Février 2009.
- [2] Réka Albert, István Albert, and Gary L. Nakarado. Structural vulnerability of the north american power grid. *Physical Review E*, 69, 2004.
- [3] L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA* 97, pages 11149–11152, 2000.
- [4] S. Massoud Amin. Securing the electricity grid. *The bridge : national academy of engineering*, 40(1) :13–20, 2010.
- [5] Soufian Ben Amor. *Percolation, prétopologie et multialéatoires, contributions à la modélisation des systèmes complexes : exemple du contrôle aérien*. PhD thesis, 2008.
- [6] Soufian Ben Amor and Marc Bui. *Complex System Approach in Modeling Airspace Congestion Dynamics*, volume 3. Journal of Air Transport Studies, January 2012.
- [7] Soufian Ben Amor, Vincent Levorato, and Ivan Lavallée. Generalized percolation processes using pretopology theory. In *5<sup>th</sup> IEEE International Conference of Research, Innovation and Vision of the Future, (RIVF)*, pages 130–134, Hanoï, Vietnam, March 2007. IEEE Computer Society.
- [8] R. Aumann and S. Hart. *Handbook of Game Theory with Economic Applications*. 1994.
- [9] Alassane Bah, Murat Ahat, and Marc Bui. An agent-based model to understand the uses of land and water resources in sahelian pastoral area. In *Première conférence Internationale Watarid*, Urumqi, Chine, octobre 2006.
- [10] Jerry Banks, John S. Carson II, Barry L. Nelson, and David M. Nicol. *Discrete-event system simulation - fourth edition*. Pearson, 2005.
- [11] ZT. Belmandt. *Basics of pretopology*. Number 978 27056 8077. Hermann, 2011.
- [12] Claude Berge. *Espaces topologiques : fonctions multivoques (2ème édition)*. 1966.
- [13] Claude Berge. *Graphes et Hypergraphes*. Dunod, Paris, 1970.
- [14] Nino Boccara. *Modeling complex system*. Number ISBN 0-387-40462-7. Springer, 2004.
- [15] A. Bunde and S. Havlin. *Percolation I*. Springer-Verlag, 1991.

- [16] David P. Chassin and Christian Posse. Evaluating north american electric grid reliability using the barabasi-albert network model. *Physica A : Statistical Mechanics and its Applications*, 355(2-4) :p. 667–677, 2005.
- [17] K. Chen, P. Yeh, H. Hsieh, and S. Chang. Communication infrastructure of smart grid. in communications. In *Control and Signal Processing (ISCCSP), 2010 4th International Symposium on*, pages 1–5, 2010.
- [18] S. Clar, K. Schenk, and F. Schwabl. Phase transition in a forest-fire model. *Physical Review E*, 55 :pp. 2174–2183, 1997.
- [19] Clauset, Newman, and Moore. Finding community structure in very large networks. 2004.
- [20] Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics : Theory and Experiment*, 2005(9) :P09008.
- [21] Groupe de Travail "Pertes sur les réseaux électriques". Les dispositifs de couverture des pertes d'énergie des réseaux publics d'électricité. Technical report, Commission de régulation de l'énergie (CRE), 2010.
- [22] U.S. Department-Of-Energy. Grid 2030 : A national vision for electricity's second 100 years. Technical report, Department of Energy, 2003.
- [23] R. Diestel. *Graph Theory*, volume 4th ed. Springer-Verlag Berlin, 2010.
- [24] Ian Dobson, Benjamin A. Carreras, Vickie E. Lynch, and David E. Newman. Complex systems analysis of series of blackouts : Cascading failure, critical points, and self-organization. *Chaos*, 17(2), 2007.
- [25] Donetti and Munoz. Improved spectral algorithm for the detection of network communities. 2005.
- [26] Dongen. *Graph Clustering by Flow Simulation*. PhD thesis, 2000.
- [27] Steven N. Durlauf and Lawrence E. Blume. *The New Palgrave Dictionary of Economics*, volume Second Edition. 2008.
- [28] Paul Erdős and Alfréd Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6 :290–297, 1959.
- [29] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *ACM Conference SIGCOMM*, pages 251–262, 1999.
- [30] Miguel A. Figliozzi, Hani S. Mahmassani, and Patrick Jaillet. *Managing complexity : insights, concepts applications.*, chapter Repeated Auction games and learning dynamics in electronic logistics marketplaces : complexity, bounded rationality, and regulation through information. Springer, 2008.
- [31] Santo Fortunato. Community detection in graphs. 2010.
- [32] Santo Fortunato, Vito Latora, and Massimo Marchiori. Method to find community structures based on information centrality. *physical review*, 70(5), 2004.
- [33] Benjamin C. M. Fung, Ke Wang, and Martin Ester. Hierarchical document clustering using frequent item sets. In *SIAM International Conference on Data Mining (SDM 2003)*, pages 59–70, 2003.

- [34] Chowell G., Hyman J. M., and Eubank S. Analysis of a real world network : The city of portland. Technical report, Technical Report BU-1604-M, Department of Biological Statistics and Computational Biology, Cornell University, 2002.
- [35] Edith Gabriel. Introduction à la statistique spatiale. In *42èmes Journées de Statistique*, 2010.
- [36] J. Gao, Y. Xiao, J. Liu, W. Liang, and C. Chen. A survey of communication/networking in smart grids. In *Future Generation Computer Systems*, volume 28, pages 391–404, 2012.
- [37] Girvan and Newman. Community structure in social and biological networks. *PNAS*, 99(12) :7821–7826, June 2002.
- [38] J. Goldenberg, B. Libai, S. Solomon, N. Jan, and D. Stauffer. Marketing percolation. *Physica A*, 284(1-4) :pp. 335–347, 2000.
- [39] Jeffrey Goldstein. Emergence as a construct : History and issues. *Journal of Complexity Issues in Organizations and Management*, 1(1) :49–72, 1999.
- [40] G. Grimmet. *Percolation*. Springer-Verlag, 1999.
- [41] P. Gupta and P.R. Kumar. Critical power for asymptotic connectivity in wireless networks. *Stochastic Analysis, Control, Optimization and Applications*, pages pp. 547–566, 1998.
- [42] Staffan Hägg. A sentinel approach to fault handling in multi-agent systems. In *Second Australian Workshop on Distributed AI, in conjunction with Fourth Pacific Rim International Conference on Artificial Intelligence (PRICAI'96)*, 1996.
- [43] L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2(1) :193–218, 1985.
- [44] Markus C. Huebscher and Julie A. McCann. A survey of autonomic computing - degrees, models, and applications. *ACM Computing Surveys*, 40(3), 2008.
- [45] International Energy Agency (IEA). World energy outlook 2009 fact sheet. Tech. report, International Energy Agency, 2009.
- [46] Scott J. *Social Network Analysis : A Handbook*. Sage Publications, 2000.
- [47] Nicholas R. Jennings, Katia Sycara, and Michael Wooldridge. A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, 1 :7–38, 1998.
- [48] Clement Jonquet, Pascal Dugenie, and Stefano A. Cerri. Service-based integration of grid and multi-agent systems models. In *SOCASE'08 : International Workshop on Service-Oriented Computing : Agents, Semantics, and Engineering*, 2008.
- [49] V.K. Kalapala, V. Sanwalani, and C. Moore. The structure of the united states road network. *Preprint University of New Mexico*, 2003.
- [50] Jeffrey O. Kephart and David M. Chess. The vision of autonomic computing. *Computer*, 36 :41–50, 2003.
- [51] Kernighan and Lin. An efficient heuristic procedure for partitioning graphs. *The Bell system technical journal*, pages 291–307, 1970.

- [52] Jeff Kramer and Jeff Magee. Self-managed systems : an architectural challenge. In *IEEE conference of Future of Software engineering (FOSE'07)*, 2007.
- [53] V. Krishna. *Auction theory*. Academic Press, 2002.
- [54] Thomas Samuel Kuhn. *La Structure des révolutions scientifiques*. 1962.
- [55] Egghe L. and Rousseau R. *Introduction to Informetrics*. Elsevier, 1990.
- [56] Andrea Lancichinetti and Santo Fortunato. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E*, 80(1) :1–8, 2009.
- [57] Andrea Lancichinetti and Santo Fortunato. Community detection algorithms : A comparative analysis. *PHYSICAL REVIEW*, 80, 2009.
- [58] Andrea Lancichinetti, Santo Fortunato, and J. Kertesz. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3) :033015, 2009.
- [59] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *PHYSICAL REVIEW*, 78, 2008.
- [60] V. Latora and M. Marchiori. Is the boston subway a small-world network? *Physica A*, 314 :109–113, 2002.
- [61] Vincent Levorato. *Contribution à la modélisation des réseaux complexes : prétopologie et applications*. PhD thesis, université paris 8, 2008.
- [62] Vincent Levorato. Modeling groups in social networks. In *Proceedings of the 25th European Conference on Modelling and Simulation (ECMS'2011)*, Poland, 2011.
- [63] Vincent Levorato and Murat Ahat. Modélisation de la dynamique des réseaux complexes associée à la prétopologie. In *9eme congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision, ROADEF08*, Clermont-Ferrand, France, 2008.
- [64] Vincent Levorato and Soufian Ben Amor. Pretopolib : la librairie java de la pretopologie. 2010.
- [65] Macal. Emergent structures from trust relationships in supply chains. In *Conference on social dynamics : interaction, reflexivity and emergence*, pages 743–760, 2004.
- [66] J. Madadhain, D. Fishera, P. Smyth, S. White, and Y. B. Boey. Analysis and visualization of network data using jung. *Journal of Statistical Software*, pages 1–25, 2005.
- [67] Martina Maggio, Henry Hoffman, Alessandro V. Papadopoulos, Jacopo Panerati, Marco D. Santambrogio, Anant Agarwal, and Alberto Leva. Comparison of decision-making strategies for self-optimization in autonomic computing systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 7(4), 2012.
- [68] Michael Mesterton-Gibbons. *An introduction to game-theoretic modelling*. American Mathematical Society, second edition ; isbn 0-8218-1929-1 edition, 2001.
- [69] Stanley Milgram. The small world problem. *Psychology Today*, 1 :61–67, 1967.

- [70] G. W. Milligan, S. C. Soon, and L. M. Sokol. The effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure. *Pattern Analysis and Machine Intelligence IEEE Transactions*, PAMI-5(1) :40–47, 1983.
- [71] Mark S. Mizruchi and G. William Domhoff. *The American Corporate Network 1904-1974*. Sage library of social research, 1982.
- [72] M.Mézard. La propagation des convictions. *Pour La Science*, Decembre 2003.
- [73] J. L. Le Moigne. *La modélisation des systèmes complexes*. Dunod, 1990.
- [74] Oskar Morgenstern and John von Neumann. *The Theory of Games and Economic Behavior*. Princeton University Press, 1953.
- [75] Roger B. Myerson. *Game Theory : Analysis of Conflict*. Harvard University Press, 1991.
- [76] E. Navarro and R. Cazabet. Détection de communautés, étude comparative sur graphes réels. In *1ères Journées Modèles et l'Analyse des Réseaux : Approches Mathématiques et Informatique*, 2010.
- [77] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2) :167–256, 2003.
- [78] Mariolis P. Interlocking directorates and control of corporations : The theory of bank control. *Social Science Quarterly*, 56 :425–439, 1975.
- [79] Stéphane Pajot. *Percolation et économie*. PhD thesis, thèse de Doctorat en Economie de l'Université de Nantes, 2001.
- [80] M. Pipattanasomporn, H. Feroze, and S. Rahman. Multi-agent systems in a distributed smart grid : Design and implementation. In IEEE/PES, editor, *Power Systems Conference and Exposition (PSCE'09)*, pages 7794–7799, 2009.
- [81] P. Pons and M. Latapy. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2) :191–218, 2006.
- [82] Radichi, Castellano, Cecconi, and Loreto. Defining and identifying communities in networks. *PNAS*, 101(9) :2658–2663, March 2004.
- [83] RAE. Electric vehicles : Charged with the potential. Technical report, The Royal Academy of Engineering, 2010.
- [84] Sarvapali Ramchurn, Perukrishnen Vytelingum, Alex Rogers, and Nick Jennings. Agent-based homeostatic control for green energy in the smart grid. In *ACM Transactions on Intelligent Systems and Technology*, volume 2, 2011.
- [85] Sarvapali Ramchurn, Perukrishnen Vytelingum, Alex Rogers, Nicholas R, and Nick Jennings. Putting the 'smarts' into the smart grid : A grand challenge for artificial intelligence. *Communications of the ACM*, 55(4) :86–97, 2012.
- [86] A. Rapoport and W.J. Horvath. A study of a large sociogram. *Behavioral Science*, 6 :279–291, 1961.
- [87] J. Ricard. La complexité biologique. *Pour La Science*, décembre 2003.
- [88] Jeremy Rifkin. *La troisième révolution industrielle. Comment le pouvoir latéral va transformer l'énergie, l'économie et le monde*. Les Liens qui libèrent, 2012.

- [89] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. In *Proceedings of the National Academy of Sciences USA*, pages 1118–1123, 2008.
- [90] J. Rousseny. Percolation. 17 :pp. 838–840, 1992.
- [91] P. Sen, S. Dasgupta, A. Chatterjee, P.A. Sreeram, G. Mukherjee, and S.S. Manna. Small-world properties of the indian railway network. *Physical Review E*, 67, 2003.
- [92] Martin Shubik. Game theory : Economic applications. *International Encyclopedia of Statistics*, 1978.
- [93] Herbert A. Simon. *Models of Man : Social and Rational*. New York : John Wiley and Sons, Inc., 1957.
- [94] R. E. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2) :146–160, 1972.
- [95] R. E. Tarjan. Enumeration of the elementary circuits of a directed graph. *SIAM J. Comput*, 2(3) :211–216, 1973.
- [96] Leigh Tesfatsion. *Handbook of Computational Economics*, chapter Agent-Based Computational Economics : A Constructive Approach to Economic Theory, pages p. 831–880. 2006.
- [97] Volterra V. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118 :558–560, 1929.
- [98] C. J. van Rijsbergen. *Information Retrieval*, volume 2nd ed. London : Butterworth, 1979.
- [99] Vijay Vittal. The impact of renewable resources on the performance and reliability of the electricity grid. *The bridge : national academy of engineering*, 40(1) :5–12, 2010.
- [100] Stanley Wasserman and Katherine Faust. *Social Network Analysis*. Cambridge University Press, 1994.
- [101] D. J. Watts and S. H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393 :p. 440–442, 1998.
- [102] G. Weisbuch and D. Stauffer. Hits and flops dynamics. *Physica A*, 287(3-4) :pp. 563–576, 2000.
- [103] Luh Yen, Denis Vanvyve, Fabien Wousters, François Fouss, Michel Verleysen, and Marco Saerens. Clustering using a random walk based distance measure. In *ESANN'2005 European Symposium on Artificial Neural Networks*, pages 317–324, 2005.
- [104] W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33 :452–473, 1977.
- [105] Richard Zallen. *The Physics of Amorphous Solids*. 1983.
- [106] Bernard P. Zeigler, Herbert Praehofer, and Tag Gon Kim. *Theory of modeling and simulation : integrating discrete event and continuous complex dynamic systems*, volume Second edition. academic press, 2000.

- [107] H. Zhou and R. Lipowsky. Network brownian motion : A new method to measure vertex-vertex proximity and to identify communities and subcommunities. In Lecture Notes in Computer Science, editor, *Proceedings of International Conference on Computational Science (ICCS)*, volume 3038, pages 1062–1069, 2004.



## Webographie

- [108] Communications acm - <http://cacm.acm.org/>.
- [109] Edf - <http://france.edf.com>.
- [110] site de erdf, électricité réseau distribution france - <http://www.erdfdistribution.fr/accueil>.
- [111] Site du département de l'énergie américain - <http://www.smartgrid.gov/>.
- [112] Site du groupe alstom - <http://www.alstom.com/grid/fr/smart-grid/>.
- [113] Smart grids cre, site d'information sur les smart grids - <http://www.smartgrids-cre.fr/>.
- [114] Smartgrid ieee - <http://smartgrid.ieee.org/>.



---

## Publications

- [115] Murat Ahat, Alain Bui, Marc Bui, Soufian Ben Amor, Guillaume Guérard, and Coralie Petermann. Smart grid and optimization. *American Journal of Operations Research*, 3(1A) :196–206, janvier 2013.
- [116] Murat Ahat, Coralie Petermann, Yann Vigile Hoareau, Soufian Ben Amor, and Marc Bui. Algorithme automatique non supervisé pour le deft 2012. In *DÉfi Fouille de Textes (DEFT 2012)*, pages 73–80, Grenoble, France, 2012.
- [117] Alain Bui, Soufian Ben Amor, and Coralie Petermann. Systèmes complexes et recherche opérationnelle : application à une simulation de smart grid. In *Recherche Opérationnelle et Aide à la DEcision Française (ROADEF'13)*, 2013.
- [118] Yann Vigile Hoareau, Murat Ahat, Coralie Petermann, and Marc Bui. Couplage d'espaces sémantiques et de graphes pour le deft 2011 : une approche automatique non supervisée. In *DÉfi Fouille de Textes (DEFT 2011)*, 2011.
- [119] Vincent Levorato and Coralie Petermann. Detection of communities in directed networks based on strongly p-connected components. In *3rd IEEE International Conference on Computational Aspects of Social Networks (CASoN'2011)*, pages 211–216, Salamanca, Spain, 2011.
- [120] Coralie Petermann, Soufian Ben Amor, and Alain Bui. Approches théoriques pour la modélisation efficace de smart grids. In *Recherche Opérationnelle et Aide à la DEcision Française (ROADEF'12)*, pages 460–461, Anger, France, 2012.
- [121] Coralie Petermann, Soufian Ben Amor, and Alain Bui. A complex system approach for a reliable smart grid modeling. *Frontiers in Artificial Intelligence and Applications*, Advances in Knowledge-Based and Intelligent Information and Engineering Systems(Volume 243) :149–158, 2012.
- [122] Coralie Petermann, Soufian Ben Amor, and Alain Bui. A pretopological multi-agents based model for an efficient and reliable smart grid simulation. In *14th International Conference on Artificial Intelligence(ICAI'12)*, volume 1, pages 354–360, Las Vegas, Nevada (USA), July 16-19 2012.
- [123] Coralie Petermann, Soufian Ben Amor, Alain Bui, and Marc Bui. Autonomic multi-agent simulation of a complex socio-technical system. In *article soumis en attente de relecture*, 2013.
- [124] Coralie Petermann, Soufian Ben Amor, Alain Bui, and Marc Bui. Optimisation de smart grid : d'un modèle intégratif vers une simulation multi-agents autonome. In *Modélisation Agents pour les Systèmes Complexes (MASyCo'13)*, 2013.

- [125] Coralie Petermann, Vincent Levorato, and Marc Bui. Structuration unilatéralement-connexe prétopologique : une extension de la librairie pretopolib. In *Veille Stratégique, Scientifique et Technologique (VSST '12)*, Ajaccio, France, 2012.
- [126] Coralie Petermann, Vincent Levorato, and Marc Bui. Unilateral pretopological classification : an extension of the library pretopolib. *Intelligences Journal*, 2, novembre 2013.

# Apprentissage automatique dans des Smart Grids

Dans le cadre de collaborations à des projets impliquant mon laboratoire de thèse, j'ai participé par deux fois au défi fouille de textes.

Cette collaboration a pour but de développer une méthodologie d'analyse et d'extraction d'information automatique de données textuelles. Cette thématique relève du *business intelligence*, et pourrait être appliquée à la construction de réseau de compétences d'après analyse de la littérature.

Ces articles ont été rédigés dans le cadre des défis fouille de textes en 2011 et 2012.

## **Couplage d’espaces sémantiques et de graphes pour le Deft 2011 : une approche automatique non supervisée**

Yann Vigile Hoareau Murat Ahat Coralie Petermann Marc Bui  
CHArt, 41 rue Gay Lussac, 75005 Paris

hoareau@lutin-userlab.fr, murat.ahat@etu.ephe.sorbonne.fr, coralie.peterman@lasc.net,  
marc.bui@ephe.sorbonne.fr

**Résumé.** Nous décrivons l’approche mise en oeuvre dans le cadre du Défi de Fouille de Texte 2011 pour la piste 2 qui consistait à identifier, pour un article scientifique donné, le résumé qui lui correspond parmi un ensemble de résumés possibles. Cette approche est basée sur le couplage entre les méthodes d’espaces sémantiques pour la représentation des connaissances sémantiques d’une part, et les graphes pour la décision sur l’affectation d’un résumé à un article, d’autre part. La méthode proposée est entièrement automatique, sans phase de paramétrage, non-supervisée et ne nécessite aucune ressource externe.

**Abstract.** We describe our approach in Deft 2011 for track 2, which is to identify a corresponding summary, for a given scientific paper, from a set of possible abstracts. The approach is based on coupling the methods on the one hand, of semantic space for the representation of semantic knowledge, and, on the other hand, graphs for the decision on the allocation of a resume to a document. The proposed method is fully automatic, without any particular tuning, unsupervised and requires no external resources.

**Mots-clés :** Espace sémantique, Graphe, Random Indexing.

**Keywords:** Semantic Space, Graph, Random Indexing.

## Introduction

Dans cette édition du Defi de Fouille de Texte 2011, nous avons appliqué une approche originale qui consiste à mixer deux méthodes de représentation des connaissances : les espaces sémantiques qui sont des espaces vectoriels à grandes dimensions et les modèles de graphes. L'intérêt du couplage des deux approches est de bénéficier d'une part des propriétés d'apprentissage non-supervisé ainsi que des propriétés sémantiques latentes associés aux espaces sémantiques et, d'autre part de la sophistication des mathématiques sous-jacentes à la théorie des graphes. Pour ce faire, la première contrainte à respecter est de produire un graphe ayant les mêmes propriétés que l'espace sémantique en ce qui concerne la représentation des relations sémantiques latentes entre les mots ou les documents (Louwerse *et al.*, 2006). Cette contrainte satisfaite, des applications peuvent alors être réalisées directement à partir du graphe. Un exemple d'application de cette approche mixte est celui de la visualisation des relations sémantiques latentes entre documents au sein de grandes bases de données textuelles (Hoareau *et al.*, 2011).

Dans la suite de cet article, nous décrivons comment nous avons appliqué cette méthode mixte pour la tâche 2 du Defi 2011. Cette méthode a été instanciée de telle sorte à représenter la relation sémantique entre chaque résumé et chaque article dans un graphe construit à partir d'un espace sémantique, puis à transformer ce graphe complet en un graphe biparti dans lequel chaque article est associé à un et un seul résumé.

L'article est organisé de la façon suivante. Dans la première section nous décrivons les espaces sémantiques et ainsi que l'approche qui consiste à représenter les documents sous la forme d'un graphe ayant les mêmes propriétés que l'espace sémantique. Dans la deuxième section, nous décrivons la chaîne de traitement mise en oeuvre pour implémenter notre méthode. Dans la troisième section, nous présentons très brièvement les résultats de notre approche pour les tâches 1 et 2 de la piste 2. Enfin, nous concluons l'article en présentant les perspectives de recherche qui pourraient prolonger le présent travail.

## 1 Le couplage espace sémantique et graphe

### 1.1 Les espaces sémantiques

Les modèles de représentation vectorielle de la sémantique des mots sont une famille de modèles qui représentent la similitude sémantique entre les mots en fonction de l'environnement textuel dans lequel ces mots apparaissent. La distribution de co-occurrence de mots est rassemblée, analysée et transformée en espace sémantique dans lequel les mots ou les concepts sont représentés comme des vecteurs dans un espace vectoriel de grandes dimensions. *Latent Semantic Analysis* (LSA) (Landauer & Dumais, 1997), *Hyper-space Analog to Language* (HAL) (Lund & Burgess, 1996) et *Random Indexing* (RI) (Kanerva *et al.*, 2000) en sont quelques exemples. Ces modèles sont basés sur l'hypothèse distributionnelle de Harris (1968) qui affirme que les mots qui apparaissent dans des contextes semblables ont des significations semblables. La définition de l'unité de contexte est un sujet commun à tous ces modèles, même si sa nature dépend du modèle. Par exemple, LSA construit une matrice mot-document dans laquelle chaque cellule  $a_{ij}$  contient la fréquence d'un mot donné  $i$  dans une unité de contexte  $j$ . HAL définit une fenêtre flottante de  $n$  mots qui parcourt chaque mot du corpus, puis construit une matrice mot-mot dans laquelle chaque cellule  $a_{ij}$  contient la fréquence à laquelle un mot  $i$  se retrouve avec un mot  $j$  en fonction d'une fenêtre flottante donnée. Différentes méthodes mathématiques et statistiques permettant d'extraire la signification des concepts sont appliquées à la distribution des fréquences stockées dans la matrice mot-document ou mot-mot. Le premier objectif de ces traitements mathématiques est d'extraire la tendance centrale des variations de fréquences et d'éliminer ce qui peut être considéré comme du « bruit » provoqué par la part d'utilisation spécifique de la langue associée à chaque scripteur. LSA emploie une méthode générale de décomposition linéaire d'une matrice en composantes principales indépendantes : la décomposition de valeur singulière (SVD). Dans HAL la dimension de l'espace est réduite en maintenant un nombre restreint de composantes principales de la matrice de co-occurrence. Des représentations vectorielles sont employées pour le stockage et la manipulation de la signification de concepts. À la fin du processus, la similitude entre deux mots peut être calculée selon différentes méthodes. Classiquement, la valeur du cosinus de l'angle entre deux vecteurs correspondant à des mots ou un groupe de mots est calculée afin d'approximer leur proximité sémantique. Une autre méthode équivalente est la distance euclidienne pondérée.

## 1.2 Random Indexing

*Random Indexing* (Kanerva *et al.*, 2000) est un modèle d'espace sémantique qui a les mêmes propriétés que les modèles d'espaces sémantiques précédemment décrits comme LSA ou HAL. La différence avec ces deux modèles est que RI ne s'appuie pas sur des méthodes de réduction matricielle mais sur les projections aléatoires. La méthode de construction d'un espace sémantique avec RI est la suivante :

- créer une matrice  $A$  ( $d \times N$ ), contenant des *vecteurs-index*, où  $d$  est le nombre de documents ou de contextes correspondant au corpus et  $N$ , le nombre de dimensions ( $N > 1000$ ) défini par l'expérimentateur. Les vecteurs-index sont creux et aléatoirement générés. Ils consistent en un petit nombre de (+1) et de (-1) et de centaines de 0;
- créer une matrice  $B$  ( $M \times N$ ) contenant les *vecteurs-termes*, où  $M$  est le nombre de termes différents dans le corpus. Pour commencer la compilation de l'espace, les valeurs des cellules doivent être initialisées à 0;
- parcourir chaque document du corpus. Chaque fois qu'un terme  $\tau$  apparaît dans un document  $d$ , il faut *accumuler* le vecteur-index correspondant au document  $d$  au vecteur-terme correspondant au terme  $\tau$ .

À la fin du processus, les vecteurs-termes qui sont apparus dans des contextes (ou documents) similaires, auront accumulé des vecteurs-index similaires.

Le modèle a démontré des performances comparables (Kanerva *et al.*, 2000) et parfois même supérieures (Karlgrén & Sahlgrén, 2001) à celles de LSA pour le test de synonymie du TOEFL (Landauer & Dumais, 1997). *RI* a été aussi appliqué à la catégorisation d'opinion (Sahlgrén & Cöster, 2004).

## 1.3 Le couplage Espace Sémantique–Graphe

Cette section décrit le processus de construction (i) d'un graphe complet représentant les propriétés sémantiques d'un espace sémantique, puis (ii) d'un graphe biparti à partir d'un graphe complet. Le procédé consiste à calculer la distance euclidienne pondérée entre chaque document de l'espace sémantique afin de construire une matrice de connexité. Cette matrice de connexité correspond alors à une représentation de l'espace sémantique sous la forme d'un graphe à  $N$  noeuds et  $N^2$  arcs. L'intérêt de cette méthode très simple est de générer automatiquement un graphe qui a les mêmes propriétés que l'espace sémantique et de permettre ainsi d'y appliquer les méthodes issues de la théorie des graphes (Hoareau *et al.*, 2011).

L'algorithme décrit ci-après a pour objectif de construire un graphe biparti à partir du graphe complet construit à partir d'un espace sémantique. Il prend en entrée un ensemble d'articles ou un ensemble de résumés pour les représenter dans un espace sémantique. Une matrice  $m$  "article – résumé" est construite. Cette matrice contient dans chaque cellule  $m_{i,j}$ , la valeur de la distance euclidienne pondérée entre les vecteurs de l'article  $i$  et du résumé  $j$ . À partir de cette matrice, un graphe  $g$  est produit. Ce graphe  $g$  peut être ambigu au sens où un résumé donné peut correspondre à plusieurs articles et *vice versa*. Un processus de désambiguïsation est appliqué à ce graphe complet afin de produire un graphe biparti où, à un résumé ne correspond qu'un seul article et *vice versa*. Dans le cas d'une ambiguïté, la distance entre le résumé et l'article est initialisée à 0 et le prochain résumé le plus proche de l'article est recherché. Ce processus est itéré jusqu'à obtenir un graphe biparti. Il est illustré dans la section 2.4.

```

Procedure main()
  Var
    A as Article Set;
    R as Resume Set;
    N as number of articles or resumes;
    m as Matrix Article Resume
    g as graph (article --> resume)

  Begin
    spaceSemantic = RandomIndexing(A, R)

    For (i:=1 to N)
      artVector = spaceSemantic(A[i]);
      For (j:=1 to N)
        resVector = spaceSemantic(R[j]);

```

```

        m[i, j] = cosine(artVector, resVector);
    End For; //j
End For; //i

g = createGraph(m);
resolveAmbiguity(g)
End Procedure //main()

Procedure createGraph(m);
Var
    m as Matrix Article Resume
    g as graph (article --> resume)
Begin
    g = emptyGraph();
    For (i:= 1 to N)
        j = Max(m[i, :])
        g.add(i, j);
    End
    Return g;
End Procedure //createGraph()

Procedure resolveAmbiguity(g)
While ambiguityExist(g)
Do
    For (k:=1 to N)
        If (g.degree(resNode[k]) > 1) Then
            ambSet = all articles nodes connected to resNode[k];
            For (l:=1 to size(ambSet))
                If (m[ambSet[l], k] is not maximum) Then
                    m[ambSet[l], k] = 0;
                    g.delete(ambSet[l], k);
                    newJ = Max(m[ambSet[l], :]);
                    g.delete(ambSet[l], newJ);
                End
            End For
        End If
    End For
End While
End Procedure

```

## 2 La chaîne de traitement

### 2.1 Extraction des documents vers une Base de Données

La première étape a été l'indexation de l'ensemble des documents du corpus dans une base de données. Pour cela, une base de données relationnelle a été construite afin de stocker toutes les informations fournies en gardant leur structure et leurs liens. L'idée était de pouvoir facilement accéder à l'ensemble des données et modifier les unités de contextes utilisées pour l'apprentissage de l'espace sémantique (Rehder *et al.*, 1998; Bestgen, 2004). Un ensemble de scripts en langage PHP a été développé afin de réaliser ces tâches. Le système de gestion de base de données est MySql pour sa facilité de connexion avec les divers langages utilisés (Php, Java, etc.).

## 2.2 Indexation de la base de données

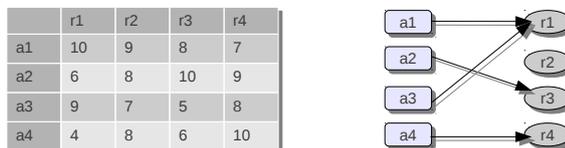
Afin d'automatiser entièrement notre chaîne de traitement, nous avons choisi d'intégrer LuSql<sup>1</sup> pour l'indexation du corpus. LuSql est une application java en ligne de commande permettant de construire des index Lucene à partir de bases de données relationnelles. Il permet de sélectionner précisément les données à indexer en passant en argument de la ligne de commande la requête SQL utilisée. De plus, dans son mode par défaut, il utilise le *multithreading* pour s'exécuter sur plusieurs processeurs et donc optimiser les temps d'indexation. Toujours en vue d'automatiser nos outils, nous avons développé une interface java intégrant LuSql.

## 2.3 Construction de l'espace sémantique

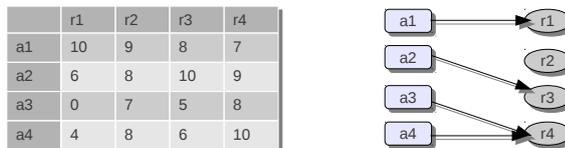
L'espace sémantique a été construit en utilisant la méthode Random Indexing implémenté par la librairie Java SemanticVectors (Widdows & Cohen, 2010). Les unités de contexte servant à l'apprentissage sont les documents tels que définis dans le corpus d'apprentissage : l'article et le résumé. Le corpus n'a fait l'objet d'aucun traitement de lemmatisation (Hoareau *et al.*, 2007; Lemaire, 2008). Les mots vides ont été supprimés. L'espace sémantique RI a été construit avec 1000 dimensions. L'ensemble des traitements a été réalisé sur un ordinateur portable à 2 Go de RAM, intel core2 duo cpu à 2.00ghz.

## 2.4 Construction du graphe biparti

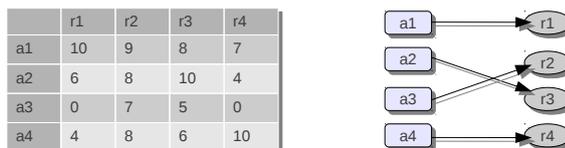
Dans cette étape, nous avons appliqué l'algorithme de la section 1.3. Le graphe construit à partir de l'espace sémantique n'avait que peu d'ambiguïtés. Nous attribuons cela d'une part à la qualité des données et d'autre part, à la robustesse de *Random Indexing*. Le processus de résolution des ambiguïtés a permis construire le graphe biparti avec chaque noeud à un degré de 1. Dans la Figure 1, nous illustrons l'ensemble du processus de la construction du graphe et de résolution des ambiguïtés.



(a) Etape 1 : Graphe avec ambiguïté



(b) Etape 2 : Graphe avec ambiguïté



(c) Etape 3 : Graphe biparti sans ambiguïté

FIGURE 1 – Exemple de construction d'un graphe biparti

1. <http://lab.cisti-icist.nrc-cnrc.gc.ca/cistilabswiki/index.php/LuSql>

### 3 Les résultats

Les performances de la méthode pour les pistes 1 et 2 représentées dans le Tableau 1 sont très satisfaisantes. La méthode réalise un score de 1 pour la tâche 1 et de 0,995 pour la tâche 2.

	Tâche 1 (articles complets)	Tâche 2 (contenu)
<b>Exécution</b>	<b>1,000</b>	<b>0,995</b>
Moyenne	0,981	0,956
Médiane	0,990	0,959
Ecart-type	0,027	0,042

TABLE 1 – Scores pour les tâches 1 et 2

### 4 Conclusion

La méthode proposée dans le cadre de notre participation au Deft 2011 repose sur le couplage entre les espaces sémantiques et les graphes. Le faible nombre de documents disponibles pour l'apprentissage constituait une contrainte forte pour notre méthode entièrement basée sur une approche distributionnelle. Les résultats indiquent que la méthode n'a pas souffert de cette contrainte.

Les méthodes d'espaces sémantiques telles que LSA, utilisant des techniques de réductions matricielles nécessitent un nombre important de documents lors de l'apprentissage afin que les processus de réduction puissent s'appliquer de façon efficace. Ainsi, il nous semble fort probable que la robustesse de notre approche serait la conséquence de l'utilisation de *Random Indexing* qui est basé sur les projections aléatoires.

De prochaines expériences seront réaliser afin de tester cette hypothèse car ses implications pourraient s'avérer importantes pour la recherche appliquée : il serait alors possible de tirer profit des propriétés des espaces sémantiques à partir d'un corpus même très limité en nombre.

### Références

- BESTGEN Y. (2004). Analyse sémantique latente et segmentation automatique des textes. In G. PURNELLE, C. FAIRON & A. DISTER, Eds., *Proceedings of the 9th International Conference on the Statistical Analysis of Textual Data (JADT'08)*, volume 1 of *Cahier du Cental*, p. 171–181, Louvain : Presse Universitaire de Louvain.
- HARRIS Z. (1968). *Mathematical Structures of Language*. New York : John Wiley and Son.
- HOAREAU Y., GANDON F., GIBOIN A., DENHIÈRE G., JHEAN-LAROSE S., LENHARD W. & BAIER H. (2007). Similarity measurement applied to information research and indexing. In F. WILD, M. KALZ, J. VAN BRUGGEN & R. KOPER, Eds., *Proceedings of the First European Workshop on Latent Semantic Analysis in Technology Enhanced Learning*, p. 5–6, Herleen (Holand).
- HOAREAU Y. V., AHAT M., MEDERNACH D. & BUI M. (2011). Un outil de navigation dans un espace sémantique. In A. KHENCHAF & P. PONCELET, Eds., *Extraction et gestion des connaissances (EGC'2011)*, volume RNTI-E-20 of *Revue des Nouvelles Technologies de l'Information*, p. 275–278 : Hermann-Éditions.
- KANERVA P., KRISTOFERSON J. & HOLST A. (2000). Random Indexing of Text Samples for Latent Semantic Analysis. In L. GLEITMAN & A. JOSH, Eds., *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, Mahwah : Lawrence Erlbaum Associates.
- KARLGRÉN J. & SAHLGRÉN M. (2001). From Words to Understanding. In Y. UESAKA, P. KANERVA & H. ASOH, Eds., *Foundations of Real-World Intelligence*. Stanford : CSLI Publications.
- LANDAUER T. & DUMAIS S. (1997). A Solution to Plato's Problem : The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. *Psychological Review*, **104**(2), 211–240.

- LEMAIRE B. (2008). Limites de la lemmatisation pour l'extraction de significations. In G. PURNELLE, C. FAIRON & A. DISTER, Eds., *Proceedings of the 9th International Conference on the Statistical Analysis of Textual Data (JADT'08)*, Cahier du Cental, p. 725–732, Louvain : Presse Universitaire de Louvain.
- LOUWERSE M., CAI Z., HU X., VENTURA M. & JEUNIAUX P. (2006). Cognitively inspired natural-language based knowledge representations : Further explorations of latent semantic analysis. *International Journal of Artificial Intelligence Tools*, **15**, 1021–1039.
- LUND K. & BURGESS C. (1996). Producing high-dimensional semantic space from lexical co-occurrence. *Behavior research methods, instruments & computers*, **28**(2), 203–208.
- REHDER B., SCHREINER M., WOLFE M., LAHAM D., LANDAUER T. & KINTSCH W. (1998). Using Latent Semantic Analysis to assess knowledge : Some technical considerations. *Discourse Processes*, **25**(2), 337–354.
- SAHLGREN M. & CÖSTER R. (2004). Using bag-of-concepts to improve the performance of support vector machines in text categorization. In *COLING '04 : Proceedings of the 20th international conference on Computational Linguistics*, p. 487, Morristown, NJ, USA : Association for Computational Linguistics.
- WIDDOWS D. & COHEN T. (2010). The semantic vectors package : New algorithms and public tools for distributional semantics. *International Conference on Semantic Computing*, **0**, 9–15.

# Algorithme automatique non supervisé pour le Deft 2012

Murat Ahat <sup>1</sup>   Coralie Petermann <sup>1, 2</sup>   Yann Vigile Hoareau <sup>3</sup>   Soufian Ben Amor <sup>1</sup>   Marc Bui <sup>2</sup>

(1) Prism, Université de Versailles Saint-Quentin-en-Yvelines, 35 avenue des Etats-Unis, F-78035 Versailles.

(2) LaISC, Ecole Pratique des Hautes Etudes, 41 rue Gay-Lussac, F-75005 Paris.

(3) CHArt, 41 rue Gay-Lussac, F-75005 Paris.

murat.ahat@prism.uvsq.fr, coralie.petermann@laisc.net,  
hoareau@lutin-userlab.fr, soufian.ben-amor@uvsq.fr, marc.bui@ephe.sorbonne.fr

## RÉSUMÉ

---

Nous décrivons l'approche mise en oeuvre dans le cadre du Défi de Fouille de Texte 2012 pour la piste 1 qui consistait à identifier, pour un article scientifique et son résumé donnés, la liste des mots clés qui lui correspondent parmi un ensemble de mot clés possibles. Cette approche est basée sur le couplage entre les méthodes d'espaces sémantiques pour la représentation des connaissances sémantiques d'une part, et les graphes pour la décision sur l'affectation d'un mot clé à un article, d'autre part. La méthode proposée est entièrement automatique, sans phase de paramétrage, non-supervisée et ne nécessite aucune ressource externe.

## ABSTRACT

---

### **Automatic unsupervised algorithm for Deft 2012**

We describe our approach in Deft 2012 for track 1, which consist in identifying a corresponding list of key word, for a given scientific paper and summary, from a set of possible key words. The approach is based on the one hand, semantic space for the representation of semantic knowledge, and, on the other hand, graphs for the decision on the allocation of a key word to a document. The proposed method is fully automatic, without any particular tuning, unsupervised and requires no external resources.

---

**MOTS-CLÉS :** Espace sémantique, Graphe, Random Indexing.

**KEYWORDS:** Semantic Space, Graph, Random Indexing.

---

# 1 Introduction

Dans cette édition 2012 du Défi Fouille de Texte, nous avons appliqué notre méthode déjà présentée lors du Deft 2011 (Hoareau *et al.*, 2011b), qui consiste à mixer deux méthodes de représentation des connaissances : les espaces sémantiques qui sont des espaces vectoriels à grandes dimensions et les modèles de graphes. L'intérêt du couplage des deux approches est de bénéficier d'une part des propriétés d'apprentissage non-supervisé ainsi que des propriétés sémantiques latentes associés aux espaces sémantiques et, d'autre part de la sophistication des mathématiques sous-jacentes à la théorie des graphes. Pour ce faire, la première contrainte à respecter est de produire un graphe ayant les mêmes propriétés que l'espace sémantique en ce qui concerne la représentation des relations sémantiques latentes entre les mots ou les documents (Louwerse *et al.*, 2006). Cette contrainte satisfaite, des applications peuvent alors être réalisées directement à partir du graphe. Un exemple d'application de cette approche mixte est celui de la visualisation des relations sémantiques latentes entre documents au sein de grandes bases de données textuelles (Hoareau *et al.*, 2011a).

L'an passé, le challenge consistait à associer un article à son résumé. Dans la suite de cet article, nous allons voir si cette année, toujours sans paramétrage ni apprentissage, notre méthode produit d'aussi bon résultats pour la tâche 1 qui consiste à apparier un article et son résumé à une liste de mots clés. Cette méthode a été instanciée de telle sorte à représenter la relation sémantique entre chaque mot clé et chaque couple article/résumé dans un graphe construit à partir d'un espace sémantique, puis à utiliser ce graphe complet pour associer à chaque article un ou plusieurs mot clé.

L'article est organisé de la façon suivante. Dans la première section nous décrivons le cadre théorique de notre algorithme en présentant les espaces sémantiques et les algorithmes de création de tels espaces à partir d'un quelconque contenu, ainsi que les bases de la théorie des graphes nécessaires à notre approche, afin de représenter les documents sous la forme d'un graphe ayant les mêmes propriétés que l'espace sémantique construit. Dans la deuxième section, nous décrivons notre algorithme. Dans la troisième section, nous présentons brièvement les résultats de notre approche et les comparons avec les résultats obtenus l'an passé. Enfin, nous concluons l'article en présentant les perspectives de recherche qui pourraient prolonger le présent travail.

## 2 Cadre théorique

### 2.1 Les espaces sémantiques

La théorie des espaces sémantiques est un ensemble de méthodes algébriques permettant de représenter des documents de tout type selon leur contenu. Plusieurs méthodes permettent de modéliser des espaces sémantiques. Elles admettent toutes l'hypothèse distributionnelle suivante : les mots ayant un sens proche apparaissent dans des documents similaires. Mais toutes reposent sur la sémantique vectorielle : les corpus sont analysés et modélisés sous forme de vecteurs à grandes dimensions, rassemblés dans une matrice de co-occurrences. Cette matrice peut être construite de deux manières selon les algorithmes :

– matrice mots-documents, utilisée par exemple dans LSA et RI, qui compte le nombre d'occur-

- rences de chaque mot dans chaque document
- matrice mots-mots, utilisée par HAL, qui regroupe les probabilités de co-occurrences pour chaque groupe de mots

Etant donnée une représentation vectorielle d'un corpus de documents, on peut introduire une notion d'espace vectoriel permettant de mettre en place la notion mathématique de proximité entre documents. En introduisant des mesures de similarité adaptées, on peut quantifier la proximité sémantique entre différents documents. Les mesures de similarité sont choisies en fonction de l'application.

Une mesure très utilisée est la similarité cosinus, qui consiste à quantifier la similarité entre deux documents en calculant le cosinus de l'angle entre leurs vecteurs. Ainsi, un cosinus nul, signe de l'orthogonalité des deux vecteurs, indiquera que ces 2 documents n'ont aucun mot en commun. L'avantage de cette méthode est que la longueur des documents n'influe en rien le résultat obtenu.

Une autre mesure possible est la distance de Manhattan (appelée aussi city-block), qui elle, prend en compte la longueur des documents comparés.

## 2.2 Random Indexing

*Random Indexing* (Kanerva *et al.*, 2000) est un modèle d'espace sémantique basé sur des projections aléatoires.

La méthode de construction d'un espace sémantique avec RI est la suivante :

- Créer une matrice  $A$  ( $d \times N$ ), contenant des *vecteurs-index*, où  $d$  est le nombre de documents ou de contextes correspondant au corpus et  $N$ , le nombre de dimensions ( $N > 1000$ ) défini par l'expérimentateur. Les vecteurs-index sont creux et aléatoirement générés. Ils consistent en un petit nombre de (+1) et de (-1) et de centaines de 0 ;
- Créer une matrice  $B$  ( $M \times N$ ) contenant les *vecteurs-termes*, où  $M$  est le nombre de termes différents dans le corpus. Pour commencer la compilation de l'espace, les valeurs des cellules doivent être initialisées à 0 ;
- Parcourir chaque document du corpus. Chaque fois qu'un terme  $\tau$  apparaît dans un document  $d$ , il faut *accumuler* le vecteur-index correspondant au document  $d$  au vecteur-terme correspondant au terme  $\tau$ .

À la fin du processus, les vecteurs-termes qui sont apparus dans des contextes (ou documents) similaires, auront accumulé des vecteurs-index similaires.

Cette méthode a démontré des performances comparables (Kanerva *et al.*, 2000) et parfois même supérieures (Karlgrén et Sahlgrén, 2001) à celles de LSA pour le test de synonymie du TOEFL (Landauer et Dumais, 1997). *RI* a été aussi appliqué à la catégorisation d'opinion (Sahlgrén et Cöster, 2004).

## 2.3 Théorie des graphes

La théorie des graphes est une théorie informatique et mathématique. Cette théorie est largement utilisée dans tous les domaines liés à la notion de réseau (réseau social, réseau informatique, télécommunications, etc.) et dans bien d'autres domaines (génétique, transports...).

Un graphe  $G = (V,A)$  est une paire composée de (Berge, 1970) :

1. un ensemble  $V = \{x_1, x_2, \dots, x_n\}$  appelé *sommets* (en référence aux polyèdres) ou *noeuds* (en référence à la loi des noeuds).
2. une famille  $A = (a_1, a_2, \dots, a_n)$  d'éléments du produit Cartésien  $V \times V = \{(x, y)/x \in V, y \in V\}$  appelés *arcs* (cas d'un graphe orienté) ou *arêtes* (cas d'un graphe non orienté).

En général, on note  $n$  le nombre de noeuds (aussi noté  $|V(G)|$ ) et  $m$  le nombre d'arcs (aussi noté  $|A(G)|$ ).

Un chemin  $P$  est composé de  $k$  arcs tels que  $P = (a_1, a_2, \dots, a_i, \dots, a_k)$  où pour chaque arc  $a_i$ , la fin coïncide avec le début de  $a_{i+1}$ . Une chaîne est l'équivalent d'un chemin dans le cadre non orienté.

Un graphe est simple si au plus une arête relie deux sommets et s'il n'y a pas de boucle sur un sommet. Dans les cas où une arête relie un sommet à lui-même (une boucle), ou plusieurs arêtes relient deux mêmes sommets, on appelle ces graphes des multigraphes.

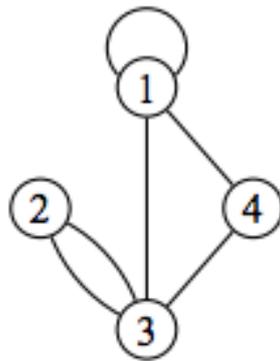
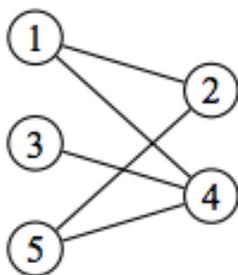


FIGURE 1 – Multigraphe.

Un graphe est biparti si ses sommets peuvent être divisés en deux ensembles  $X$  et  $Y$ , de sorte que toutes les arêtes du graphe relient un sommet dans  $X$  à un sommet dans  $Y$  (dans l'exemple ci-dessous, on a  $X = 1,3,5$  et  $Y = 2,4$ ).



**Graphe biparti**

$$V = \{1, 2, 3, 4, 5\}$$

$$E = \{\{1, 2\}, \{1, 4\}, \{2, 5\}, \{3, 4\}, \{4, 5\}\}$$

FIGURE 2 – Graphe biparti.

Dans notre méthode, nous utiliserons des graphes simples bipartis, afin d'associer chaque article à une liste de mots clés. Dans la section suivante, nous présentons notre algorithme en détails.

### 3 Notre algorithme

Cette section décrit le processus de construction (i) d'un graphe complet représentant les propriétés sémantiques d'un espace sémantique, puis (ii) d'un graphe biparti à partir d'un espace sémantique.

Notre méthode débute par une étape de prétraitements qui consiste à supprimer les mots vides de sens tels que les conjonction de coordination, articles indéfini, pronoms...

Le procédé consiste ensuite à générer notre espace sémantique à l'aide de la méthode RI puis à calculer la distance euclidienne pondérée entre chaque document et chaque mots clés de l'espace sémantique afin de construire un graphe biparti complet. L'intérêt de cette méthode très simple est de générer automatiquement un graphe biparti et de permettre ainsi d'y appliquer les méthodes issues de la théorie des graphes (Hoareau *et al.*, 2011a).

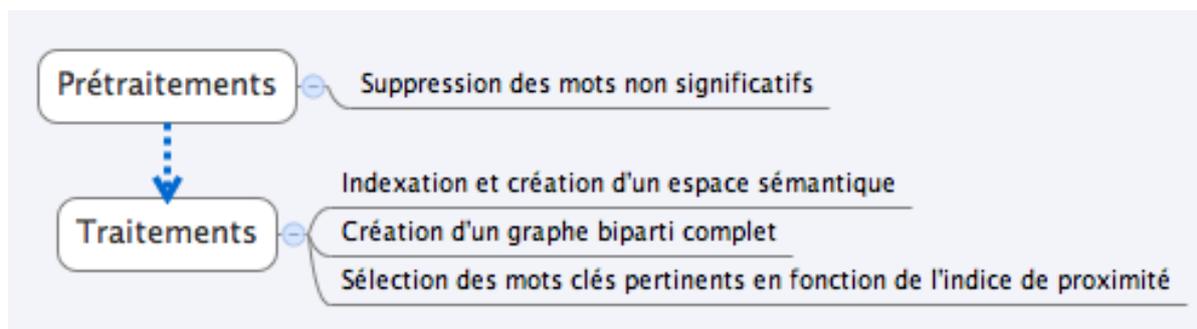


FIGURE 3 – Notre algorithme.

L'algorithme décrit ci-après a pour objectif de construire un graphe biparti à partir d'un espace sémantique. Il prend en entrée un ensemble d'articles . Une matrice  $m$  "article – mots clés" est construite. Cette matrice contient dans chaque cellule  $m_{i,j}$ , la valeur de la distance euclidienne pondérée entre les vecteurs de l'article  $i$  et du mots clés  $j$ . À partir de cette matrice, un graphe biparti complet  $g$  est produit. Un processus de filtre est appliqué à ce graphe afin de produire un graphe biparti où à un article est connecté à ces mots clés.

```
Procedure main()
  Var
    A as Article Set;
    K as Kew word Set;
    N as number of articles;
    M as number of keywords;
    m as Matrix Article Key word;
    g as graph (article --> key word);

  Begin
    spaceSemantic = RandomIndexing(A)

    For (i:=1 to N)
      artVector = spaceSemantic(A[i]);
```

```

    For (j:=1 to M)
        keyVector = spaceSemantic(R[j]);
        m[i,j] = cosine(artVector, resVector);
    End For; //j
End For; //i

g = createGraph(m);
End Procedure //main()

Procedure createGraph(m);
    Var
        m as Matrix Article Key word;
        g as graph (article --> key word);
        knumList as number of keywords for articles;
Begin
    g = emptyGraph();
    For (i:= 1 to N)
        templist = Max(m[i,:],knumbList);
        g.add(i,templist);
    End
    Return g;
End Procedure //createGraph()

```

## 4 Résultats et discussion

Pour le défi Deft 2012, nous avons soumis deux groupes de résultats, obtenus à l'aide de deux espaces sémantiques différents. Le premier est créé à partir des documents de test et d'apprentissage, alors que le second est créé uniquement à partir des documents de test. La librairie utilisée implémentant random indexing est semantic vectors, et la dimension des vecteurs a été paramétrée à 2048 avec un cycle d'entraînement. Même si notre algorithme nous a fourni de bons résultats pour Deft 2011 en nous hissant sur la première place ex aequo du podium (Hoareau *et al.*, 2011b), les résultats obtenus cette année ne sont pas satisfaisants (voir le tableau suivant).

Run	Précision	Rappel	F-score
1	0,0428	0,0428	0,0428
2	0,0242	0,0242	0,0242

TABLE 1 – Scores pour les tâches d'appariement du DEFT 2012

Nous avons tenté en vain d'améliorer nos résultats avec des paramétrages différents des espaces sémantiques pour tester des dimensions jusqu'à 6000 et jusque 5 cycle d'entraînement. Nous assumons alors que la méthode de random indexing peut être une des causes de cet échec. Nous poursuivons donc nos recherches sur ce sujet, en testant diverses méthodes de constructions d'espaces sémantiques et divers outils concernant les espaces sémantiques.

## 5 Conclusion et perspectives

La méthode proposée dans le cadre de notre participation au Deft repose sur le couplage entre les espaces sémantiques et les graphes. Le faible nombre de documents disponibles pour l'apprentissage constituait une contrainte forte pour notre méthode entièrement basée sur une approche distributionnelle. En 2011, nous avons obtenu de bons résultats mais la tâche du Deft 2012 a montré les limites de notre méthode.

De prochaines expériences seront réalisées afin de comparer notre méthode, et améliorer son paramétrage.

## Références

- BERGE, C. (1970). *Graphes et Hypergraphes*. Dunod, Paris.
- HOAREAU, Y. V., AHAT, M., MEDERNACH, D. et BUI, M. (2011a). Un outil de navigation dans un espace sémantique. In KHENCHAF, A. et PONCELET, P., éditeurs : *Extraction et gestion des connaissances (EGC'2011)*, volume RNTI-E-20 de *Revue des Nouvelles Technologies de l'Information*, pages 275–278. Hermann-Éditions.
- HOAREAU, Y. V., AHAT, M., PETERMANN, C. et BUI, M. (2011b). Couplage d'espaces sémantiques et de graphes pour le deft 2011 : une approche automatique non supervisée. In *DÉfi Fouille de Textes (DEFT 2011)*, Montpellier, France.
- KANERVA, P., KRISTOFERSON, J. et HOLST, A. (2000). Random Indexing of Text Samples for Latent Semantic Analysis. In GLEITMAN, L. et JOSH, A., éditeurs : *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, Mahwah. Lawrence Erlbaum Associates.
- KARLGREN, J. et SAHLGREN, M. (2001). From Words to Understanding. In UESAKA, Y., KANERVA, P. et ASOH, H., éditeurs : *Foundations of Real-World Intelligence*. CSLI Publications, Stanford.
- LANDAUER, T. et DUMAIS, S. (1997). A Solution to Plato's Problem : The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. *Psychological Review*, 104(2):211–240.
- LOUWERSE, M., CAI, Z., HU, X., VENTURA, M. et JEUNIAUX, P. (2006). Cognitively inspired natural-language based knowledge representations : Further explorations of latent semantic analysis. *International Journal of Artificial Intelligence Tools*, 15:1021–1039.
- SAHLGREN, M. et CÖSTER, R. (2004). Using bag-of-concepts to improve the performance of support vector machines in text categorization. In *COLING '04 : Proceedings of the 20th international conference on Computational Linguistics*, page 487, Morristown, NJ, USA. Association for Computational Linguistics.



## Nos publications en cours

Nous présentons ici 2 publications à paraître qui s'inscrivent dans la continuité de nos travaux.

# Autonomic multi-agent simulation of a complex socio-technical system

Coralie Petermann

Université de Versailles Saint-Quentin-en-Yvelines  
CaRO - Versailles, France  
Email: coralie.petermann@uvsq.fr

Alain Bui

Université de Versailles Saint-Quentin-en-Yvelines  
CaRO - Versailles, France  
Email: alain.bui@uvsq.fr

Soufian Ben-Amor

Université de Versailles Saint-Quentin-en-Yvelines  
CaRO - Versailles, France  
Email: soufian.ben-amor@uvsq.fr

Marc Bui

Ecole Pratique des Hautes Etudes (EPHE)  
ChART - Paris, France  
Email: marc.bui@ephe.sorbonne.fr

**Abstract**—In this paper, we provide a multi-agent based simulation to optimize and analyze socio-technical complex systems. We specifically focus on Smart Grid, which is a typical *complex socio-technical system* due to the heterogeneity of actors, economic issues and material aspects. Our simulation takes into account the heterogeneity of components, links them with a generalized proximity concept, and guarantee an optimal smooth functioning of the global system.

## I. INTRODUCTION

This work is part of a research project called Energy Positive IT 2.0 (EPIT2.0) that combines our laboratory PRiSM to industrial partners such as Alstom, CEA, Renault, etc. ... EPIT is a 3 year research program dedicated to develop and validate on a test site a complete system to control energy. This will enable intelligent and optimum management of energy by controlling the production, storage and consumption using innovative information systems and simulation technology at the cutting edge of innovation.

In this paper, we present our integrative modeling approach for Smart Grid optimization and management. We have defined a model based on the combination of several theories to represent and solve every aspect or problem of the system. Finally, we present our simulation that allows successive iterations to predict the system overall behavior and enable the analysis and optimization.

## II. SMART GRID: THE FUZZY FUTURE OF THE CONTEMPORARY POWER GRID

Currently, the term Smart Grid refers to a fuzzy concept. There are various definitions depending on the area. Briefly, the vision of the Smart Grid from a legislator is different from the vision of a power producer. Overall, the Smart Grid can be defined as the vision of the evolution of our actual power grid, where a two-way flow of information and power can be adjusted in real-time demand, to optimally integrate renewable energy and control electricity at any point in the network [5]. In our work, we focus on a general and context-free definition of Smart Grids.

The Smart Grid is composed of heterogeneous devices, from smart meters <sup>1</sup> and solar panels, to equipment of power stations and sensors for high-voltage lines. At the same time, the electrical resource is produced by multiple processes: on one hand a stable production by nuclear and thermal power plants, and on the other hand, unstable and outdoor conditions dependent production with the integration of renewable energies. Distribution and transport networks were created over years but actually are not really mastered nor optimized. Our goal is to make the power grid more efficient by optimally integrating renewable energy and taking advantage of new information and communication technologies.

### A. Analysing the Smart Grid complex structure

The Smart Grid is composed of a large number of heterogeneous entities that we can model using agents (producers, consumers, government ...) that interact simultaneously with conflicting goals. There may be economic issues such as selling at the best cost concerning producers, political issues such as reducing emissions of carbon dioxide, or technical issues of communication and security.

In addition, the structure of a Smart Grid is based on complex distribution and communication networks: on one hand, the distribution network is a set of power lines with different voltages that allows electricity to travel the grid for distribution, while on the other hand, the communication network allows communication between entities and collects statistics from smart meters and sensors (see fig. 1).

A complex system is a system composed of a large number of entities in local and simultaneous interactions which overall behavior can't be derived from that of its components [11], *ie*, the overall behavior of the system is not deductible from that of its parts or subsystems or their properties [10].

Smart Grid is then a perfect representation of complex socio-technical system with human and material entities, called

<sup>1</sup>a smart meter is an electric device that can identify precisely in real-time the power consumption of a home

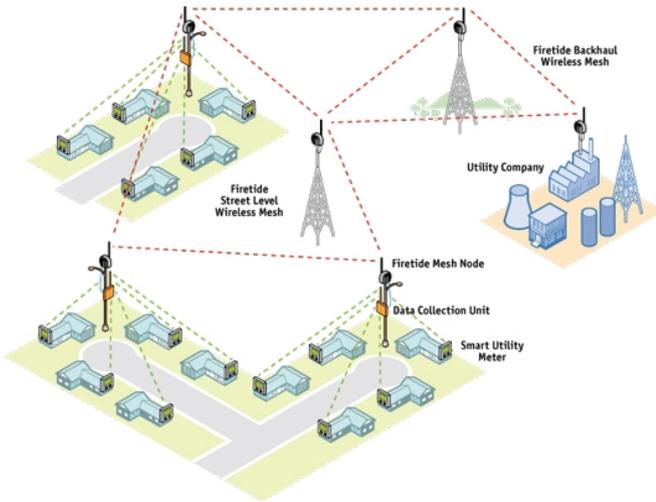


Fig. 1. Smart Grid is a complex network.

agents. In the next part, we briefly define a conceptual point of view of the different parts and agents of the Smart Grid. As a first step, we determine the different actors whether human or material, then in a second part, we define the global environment.

### B. Definition of the socio-technical entities

Electricity producers are at the core of the system. They represent a major complexity factor on their own. Indeed, several types of producers involve several types of management of energy. The French energy is produced mainly using nuclear, thermal and renewables.

Nuclear generation represent almost 75 % of the french production. It is characterized by stable production and high yield.

The thermal generation from fossil fuel also produces a large amount of electricity. Its production is high and subject to little fluctuation.

Renewable energies have a growing role in the production and it will quickly become essential in the aim of reducing carbon dioxide emission imposed by the European Union. Their production is variable and drastically changes from one type to another. Some, such as solar and wind energies, depend directly on external climatic conditions. Others, such as hydropower, can produce a large amount of energy in a very short time, but it will take as much time to adjust storage to be able to produce again.

In France, the energy consumption is achieved largely by residential and industries which together represent more than 90 % of the total consumption. However, with the evolution of the power grid to the Smart Grid, we no longer speak of consumers, but prosumers. Consumers become responsible for their own consumption and the concept of participation in the reduction of consumption and production or storage of energy is growing. Individuals install solar panels on their roofs, and some industries establish electric cars parks at the service of their employees to store excess energy : we say that

energy production is becoming decentralized. However, these activities are still marginal and in fact, the legislation does not currently allow people to sell their own energy. Then, self-sufficient prosumers won't appear in our project as having little influence on the system, but we will represent industrial and residential consumers.

### C. Definition of the environment

The Smart Grid environment concerns all that have an external influence on the actors of the system. These influences are not necessarily represented as agents, but as parameters or modules of influence.

In France, the Regulatory Commission (CRE) will play a central role in the Smart Grid. This agency is responsible for regulating the energy sector. It will be affected by the evolution of Smart Grid network because it must adapt to the new system of electricity prices, which will be more complex than the current due to the decentralization and diversification of production. It will therefore try to facilitate the deployment of Smart Grid adapting energy regulation.

Another important point of influence in the French network concerns aggregators that require industries, by paying them, to lower their power consumption occasionally. Industries are large consumers of energy, this service can occasionally smooth consumption curve.

Finally, as any real system, Smart Grid is governed by a number of unpredictable stochastic factors such as climate, probability of system failure, or human behavior. Actually, we can barely integrate human behaviors in our simulation due to its complexity but we can integrate climate changes and failures as external parameters.

In the next section, we present the mathematical model proposed to address the problems of simulation and optimization of Smart Grid.

## III. INTEGRATIVE MODEL FOR SMART GRID AUTONOMIC SYSTEMS

In this section, we summarize our model, as it has already been presented in detail in a previous article [14]. However, to understand our simulation, it is necessary to briefly justify the choice of mathematical theories.

Autonomic systems are capable of adapting their behavior and resources hundreds of times per second to automatically decide the best way to achieve a goal despite changing environmental data and applications [9]. These systems can regulate themselves without human intervention.

The term autonomic was first used in biology. In anatomy, autonomic nervous system controls unconscious reflexes, ie bodily functions that do not require attention, including automatic adjustments such as pupil size, digestion, rate and depth of breathing, or dilation and constriction of blood vessels. Autonomic computing tends to interfere with computer systems in the same way as its counterpart in biology. The term autonomic computing was first used by IBM in 2001 to describe systems with self-management properties such as self-configuration, self-optimization, self-healing and self-protection [6].

Auto-configuration assign a goal to the system, without specifying how to do it, and autonomic system is able to choose actions and appropriate settings to achieve the goal. Such a system is also able to detect and diagnose all types of faults as well as low level failures. If it can, then the system attempts a repair without external intervention (self-repair). The self-protection property protects the system from malicious hacker attacks but also from aberrant uses of normal users. Finally, autonomic systems must ensure system security and data protection.

The Smart Grid is a perfect example of autonomic system. Indeed, it must satisfy all the properties of self-management and self regulation while showing real-time responsiveness. If one of these properties is not filled, consumers will not be satisfied, the system will then be considered as unstable and unreliable.

One of the most obvious modeling method to best represent a complex socio-technical system such as the Smart Grid is multi-agent modeling. A multi-agent system (MAS) is a system composed of a set of multiple interacting intelligent entities (agents), an environment, a set of relations between entities and a set of operations that allows agents to perceive, produce, consume, transform and manipulate objects [8]. Agents are specific objects representing the active entities of the system. Multi-agent systems can be used to solve problems that are difficult or impossible for an individual agent or a monolithic system to solve. Intelligence may include some methodic, functional, procedural or algorithmic search, find and processing approach.

The design of complex system is more realist with intelligent agents interacting together, because each agent can act as a separate mathematical model [16]. That is why agent-based modeling is one of the most evident approach for modeling complex system. MAS approach is used in lots of research topics such as pollution modeling [1], social structures... One of the major benefits of MAS is to allow to study in silico the generic problem in risk-free space, while studying the global view of the system and setting up important parameters. Then, we can test the impact of various parameters in various scenari at very low-cost.

However the only theory of multi-agent does not allow us to solve Smart Grids systems because there are theoretical problems [2]. First, Smart Grids involves heterogeneous actors with different interests: the state, consumers, and companies are in perpetual conflict in the economic field. This behavior very close to game theory [13], [12]: it is indeed a strategic problem in which the actions of each player will influence the others. The negotiations are based on notions of proximity between agents, but if we use metric distances that won't be appropriate and it will manage only one criterion. Conventional topology would be too cumbersome to be implemented so pretopology, a mathematical theory weaker than classical topology, was chosen to express the structural transformation of sets composed of interacting elements, such as emergence of collective behavior [3], [7], using multivariate relationships between entities. However, the Smart Grid is dynamic so we

need a method to verify that our model generates an overall system working properly: that is, it distributes electricity regularly to every customer. The percolation is then the most appropriate theory to model brown out and black out electric phenomenon.

Eventually we will get a tool for decision support with a constructive approach [17] that will help to refine variables, change settings using modeling theories enriched by classical mathematical theories. We aim to achieve a very effective model to manage Smart Grid complexity in order to control it or enhance the performance.

In the next section, we present our simulation based on our model.

#### IV. AUTONOMIC MULTI-AGENT SIMULATION FOR SMART GRID OPTIMISATION

The simulation of complex systems aims to predict the possible evolutions of such systems. In this section, we present our approach in the design of the simulation from our model.

##### A. Use-cases of our Smart Grid simulation

The first step in the implementation of the simulation is to describe use cases. This chart allows us to view important parts of our simulation and propose a logical decomposition.

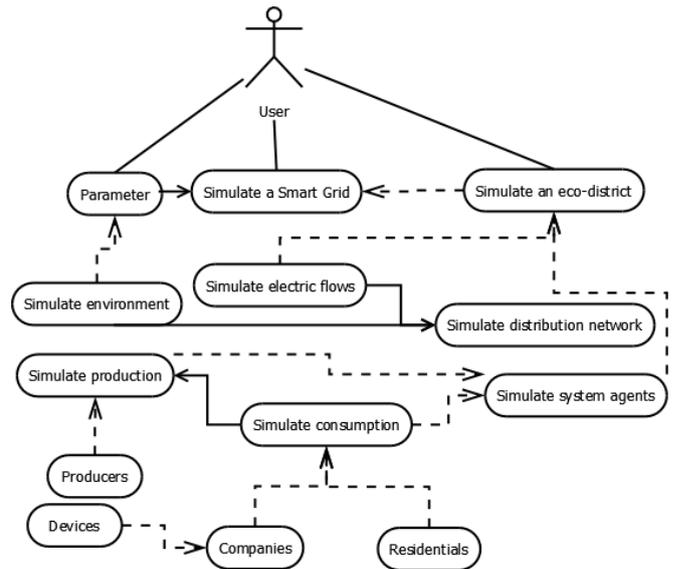


Fig. 2. Use case diagram.

##### B. Agents modeling

To take into account the heterogeneity of Smart Grid entities, we chose to integrate in our simulation several types of agents. Each agent is able to communicate with its neighbors, which in turn can send messages step by step. Thus, each agent has only local knowledge of the network, we do not memorize all the possible routes, which would be impossible on the scale of a national Smart Grid.

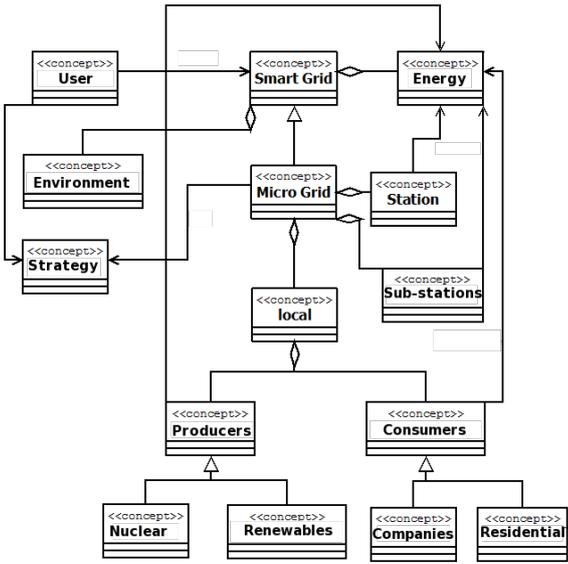


Fig. 3. Ontological diagram.

Agents of our model are dynamic entities capable of producing, consuming, or transferring energy. So each agent has its own mathematical model for energy management, which may include an algorithm for production, consumption, or a transmission protocol for a specified target.

To model creation or consumption, we incorporated real data from the French transport network RTE (*Réseau de transport français*). One system parameter adjusts the granularity (see figure 4) and then a time step can represent an interval from 5 minutes to 1 entire day. This setting allows us to highlight some complex phenomena only visible on certain time periods .

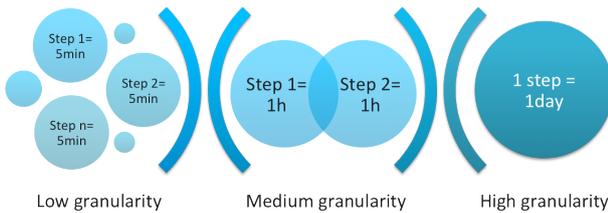


Fig. 4. Simulation granularity.

The environment of our simulation define a set of external parameters influencing the behavior of agents.

The modeling of complex systems are problematic for simulation [15]. The decomposition of stable subsystems is necessary for modular programming. However, this decomposition is complex and does not necessarily correspond to the

modular decomposition of the mathematical model.

In the context of Smart Grids in particular, to simulate a complex system, it is necessary to follow a certain hierarchy. In fact, the Smart Grid includes different levels of complexity scale (spatial and temporal), and also includes an intricacy between structure and dynamics of the system. Thus, we considered conceptual levels, and modular programming levels, without having necessarily an exact parallel between conceptual levels of design and technical levels of simulation [4].

### C. Conceptual levels

Conceptual values of the Smart Grid are due to its network structure. In order to best match the reality, we distinguished four interconnected conceptual levels (see figure 5):

- The core of the production: with nuclear and thermal agents
- Main power stations: material agents which represents branch lines and changes in voltage power lines.
- Sub-stations: same as stations at a lower level
- Main consumers: residential human agents



Fig. 5. Our conceptual levels.

## V. RESULTS

Our simulation is operational and we aim to add the percolation module for failure detection in a future work, as we currently don't have enough data to develop it.

Our simulation achieve to auto-control and optimize energy sharing. We can have an overall view of the system owing to the dashboard (see figure 6), which displays energy consumption, production for the global system and for specific points in the network.

Each new iteration begins with a data update concerning consumption, production and distribution. Each level update their data consumption and the priorities of its appliances. The value of each device is updated according to the priority of the material specific need (a hospital can not be subjected to a black out, for example), and level of consumption. Then, the local level compare consumption value relative to the previous prognostic at step 1. If the data indicate that this point

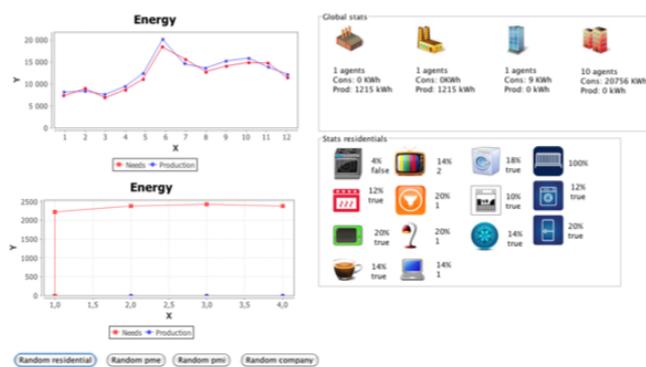


Fig. 6. Our simulation.

in the network over-consume, we use a knapsack method to determine the optimal household consumption.

After obtaining their ideal consumption, each local level sends its data to the next level, called microgrid. This step marks the transition to step 2 of our algorithm.

The microgrid, hanging on a time threshold define, conducts auction based on the consumption of each local level, then sum and sends the result to the parent station (step 3).

In step 4, transmission and distribution network updates its routing table using the residual graph. Then, in step 5, stations become aware of the possible amount of energy received from producers. A feedback system between microgrid and station, at step 6, balances auctions in relation to the actual energy being served. Data feedback allow to obtain better prognostics, by weighting, for the next iteration. Once consensus is reached, or balance, the energy is distributed in step 7. Energy unallocated, if remaining, is redistributed by bottom-up action.

The overall relationship between the three levels of the network is shown in figure 7.

## VI. CONCLUSION

To conclude, step by step we are defining and developing, through the dialectic between model and simulation, an approach for optimization and analysis of socio-technical systems. In particular, in the case of Smart Grid, our simulation aims to find optimum strategies and parameters to propose a stable and efficient intelligent power grid.

As there is little previous work on Smart Grid global simulation, we cannot directly compare our results to another work, but our algorithms distribute properly the energy owing to several simulations tests.

In a future work, we aims to integrate the percolation theory to model electric phenomena such as black out, which is very close to phase transition.

## REFERENCES

- [1] Murat Ahat, Soufian Ben Amor, Marc Bui, Michel Lamure, and Marie-Françoise Courel. Pollution modeling and simulation with multi-agent and pretopology. *Complex Sciences*, Volume 4(Part 1):225–231, 2009.
- [2] Murat Ahat, Alain Bui, Marc Bui, Soufian Ben Amor, Guillaume Guérard, and Coralie Petermann. Smart grid and optimization. *American Journal of Operations Research*, 3(1A):196–206, janvier 2013.

- [3] ZT. Belmandt. *Basics of pretopology*. Hermann, 2011.
- [4] Alain Bui, Soufian Ben Amor, and Coralie Petermann. Systèmes complexes et recherche opérationnelle : application à une simulation de smart grid. In *Recherche Opérationnelle et Aide à la DECision Française (ROADEF'13)*, 2013.
- [5] U.S. Department-Of-Energy. Grid 2030: A national vision for electricity's second 100 years. Technical report, Department of Energy, 2003.
- [6] Jeffrey O. Kephart and David M. Chess. The vision of autonomic computing. *Computer*, 36:41–50, 2003.
- [7] Vincent Levorato and Murat Ahat. Modélisation de la dynamique des réseaux complexes associée à la prétopologie. In *9eme congrès de la Société Française de Recherche Opérationnelle et d'Aide a la Décision, ROADEF08*, Clermont-Ferrand, France, 2008.
- [8] Macal and North. Tutorial on agent-based modelling and simulation. *Journal of simulation*, 4:151–162, 2010.
- [9] Martina Maggio, Henry Hoffman, Alessandro V. Papadopoulos, Jacopo Panerati, Marco D. Santambrogio, Anant Agarwal, and Alberto Leva. Comparison of decision-making strategies for self-optimization in autonomic computing systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 7(4), 2012.
- [10] M.Mézard. La propagation des convictions. *Pour La Science*, Decembre 2003.
- [11] J. L. Le Moigne. *La modélisation des systèmes complexes*. Dunod, 1990.
- [12] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.
- [13] John von Neumann Oskar Morgenstern. *The Theory of Games and Economic Behavior*. Princeton University Press, 1953.
- [14] Coralie Petermann, Soufian Ben Amor, and Alain Bui. A complex system approach for a reliable smart grid modeling. *Frontiers in Artificial Intelligence and Applications, Advances in Knowledge-Based and Intelligent Information and Engineering Systems(Volume 243)*:149–158, 2012.
- [15] Coralie Petermann, Soufian Ben Amor, and Alain Bui. A pretopological multi-agents based model for an efficient and reliable smart grid simulation. volume 1, pages 354–360, Las Vegas, Nevada (USA), July 16-19 2012.
- [16] M. Pipattanasomporn, H. Feroze, and S. Rahman. Multi-agent systems in a distributed smart grid: Design and implementation. In *IEEE/PES, editor, Power Systems Conference and Exposition (PSCE'09)*, pages p 1–8, 2009.
- [17] Alexis Tsoukiàs. De la théorie de la décision à l'aide à la décision. *Concepts et Méthodes pour l'Aide à la Décision*, 2006.

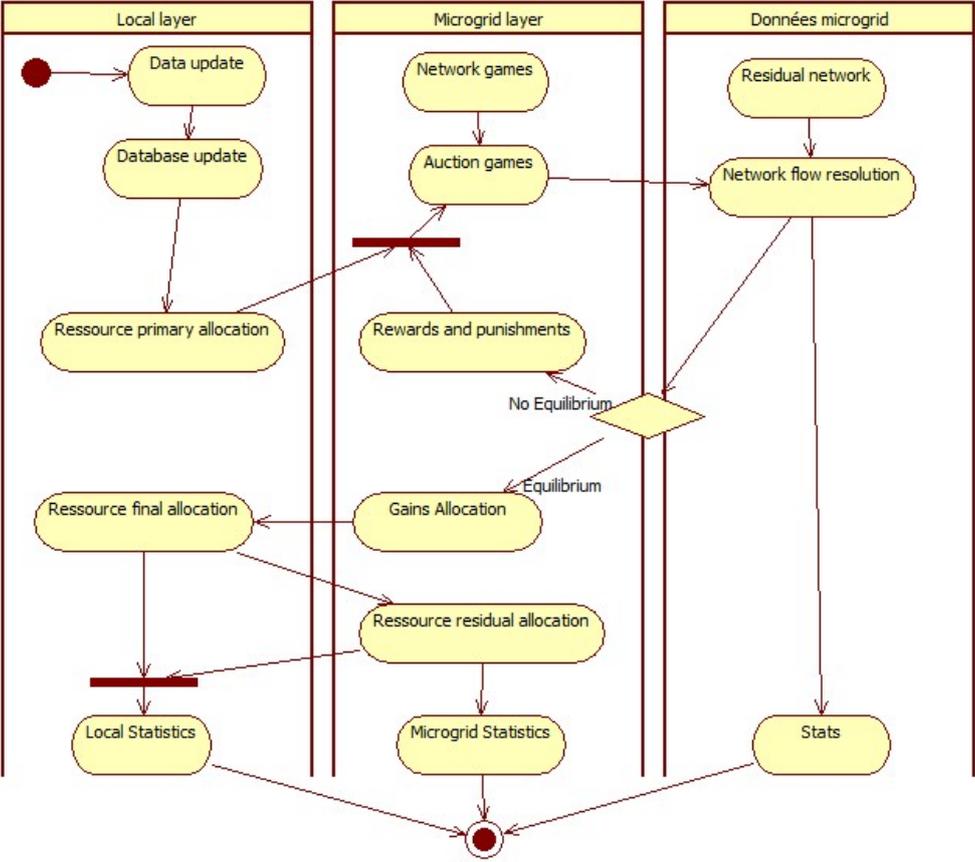


Fig. 7. Optimization steps for resource sharing.

# Unilateral pretopological classification : an extension of the PretopoLIB library.

Coralie PETERMANN (\*), Vincent LEVORATO (\*\*), Marc BUI (\*\*\*)  
[coralie.petermann@prism.uvsq.fr](mailto:coralie.petermann@prism.uvsq.fr), [vlevorato@cesi.fr](mailto:vlevorato@cesi.fr), [marc.bui@laisc.net](mailto:marc.bui@laisc.net)

(\*) [Prism](#), Université de Versailles St-Quentin-en-Yvelines, FR-78035 Versailles,

(\*\*) CESI, 959 rue de la Bergeresse, F-45160 Olivet,

(\*\*\*) [LaISC](#), 41 rue Gay-Lussac, F-75005 Paris.

**Keywords:** classification, graph theory, pretopology, pretopoLIB.

## Abstract

In this paper, we solve the issue of unilateral connectedness used in structuring components of directed graphs and networks. To achieve this, we define an algorithm to detect the components unilaterally related using pretopology, and we integrate this algorithm in the Java library PretopoLIB.

# 1 Introduction

Biological, social, technological, and information networks are often studied using graphs [6], and graph analysis is essential to understand the characteristics of these systems.

One of the most relevant features of graphs representing real systems is community structure. This problem is not yet solved despite the large number of proposed algorithms [3].

In this context, we introduced an algorithm called LP11 in order to classify oriented networks. This algorithm used the strongly connected components of maximum diameter  $p$  and obtained very satisfactory results [2]. In this article, we present the continuation of this work, because we are now focusing on other types of connectivities. We especially focus on the unilateral-connectivity used in pretopology. Indeed, we propose to store and use any information of complex networks, representing them as a pretopological space using oriented pretopological relations.

In this article, we present our algorithm detecting unilaterally related components, and the extension of PretopoLib library integrating it. In the first part we present the pretopology which is the mathematical theory used in our work. Then, we discuss the concepts of  $\chi$ -connectivity which include unilateral connectivity. Then we present our detection algorithm using unilaterally related components. Finally, we conclude with the presentation of the PretopoLIB and some demonstrations.

## 2 Pretopology

The pretopology is a mathematical theory with axioms weaker than classical topology. Pretopology allows to express structural transformations of sets with interacting components such as the formation of coalitions among a population, alliance phenomena, processes of tolerance, acceptance and the emergence of collective behavior [7,8].

We call closure defined on  $E$ , any function  $a(\cdot)$  of  $\mathcal{P}(E)$  in  $\mathcal{P}(E)$  such that:

- $a(\emptyset) = \emptyset$
- $\forall A \subset E, A \subset a(A)$

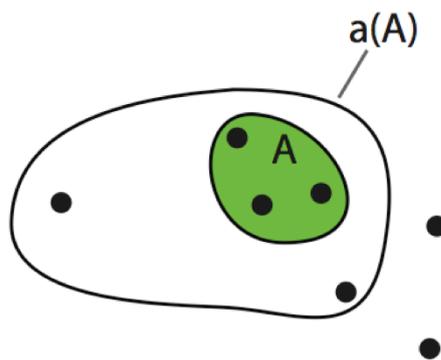


Figure 1 : closure  $a(A)$

We call interior defined on  $E$  any function  $i(\cdot)$  of  $\mathcal{P}(E)$  in  $\mathcal{P}(E)$  such that:

- $i(E) = E$
- $\forall A \subset E, i(A) \subset A$

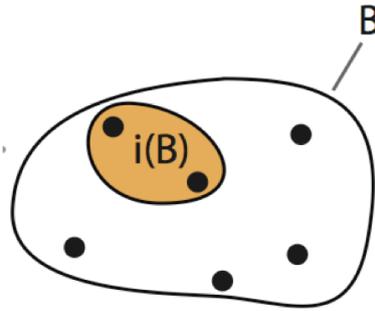


Figure 2 : interior  $i(B)$

Given a set  $E$ ,  $a(\cdot)$  and  $i(\cdot)$ , the couple  $s = (a(\cdot), i(\cdot))$  is called a pretopological structure  $E$  and the 3-tuple  $(S, a(\cdot), i(\cdot))$  is called a pretopologic space. Pretopology in a complex network is seen as a pretopological family on a given set  $E$  [9].

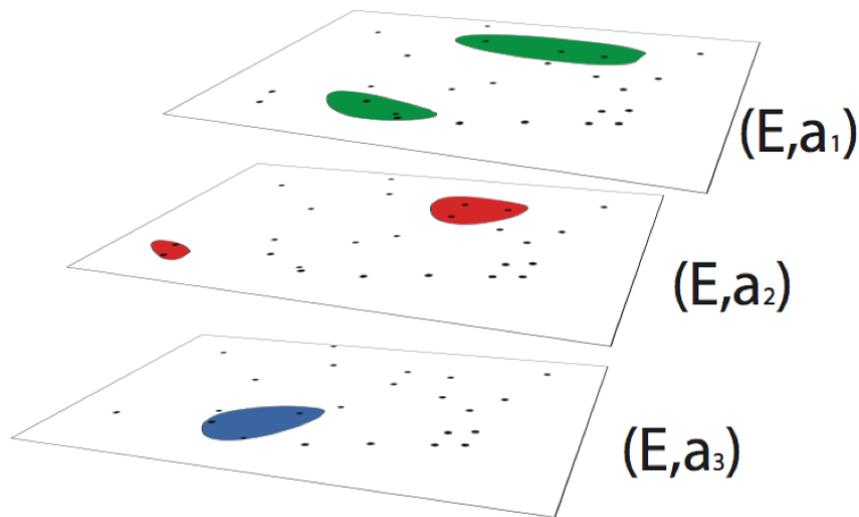


Figure 3 : a pretopological family with 3 sets

Then, we can use different pretopological spaces on the same set to better model a phenomena. By this way, an accurate setting can be done easily and efficiently, and changing the value of one criteria is instantaneous.

## 2.1 $\chi$ -connectivities

Connectivity is an interesting concept for modeling problems related to the homogeneity or the structure of elements. However, the axioms of classical topology are often in conflicts with the real-world representation, particularly in the domain of social sciences [7].

Pretopology defines several types of connectivities. In this section, we consider a set  $E$ , defined by the pretopological structure  $s = (i(\cdot), a(\cdot))$  of  $V$  type. We note  $\sigma = (O(\cdot), F(\cdot))$  the operations of opening and closure on  $E$ . There are different types of connectivities, known as the  $\chi$ -connectivities with  $\chi$  corresponding to :

- SC : strongly connected ;
- UC : unilaterally connected ;
- HC : hyperconnected ;
- AC : apo-connected ;
- C : connected;

### 2.1.1 Strongly connected

#### Definition

$\forall A \subseteq E, A \neq \emptyset, F(A) = E$  (the only non-zero closed space is E)

E is SC iff  $\forall A \in P(E), \forall B \in P(E)$ , there is a path from A to B.

### 2.1.2 Unilaterally connected

**Definition**

E is unilaterally connected (UC) iff :  $\forall A \subseteq E, A \neq \emptyset, F(A) = E$  else  $\forall B \subseteq E, B \neq \emptyset, B \subseteq E - F(A) \Rightarrow A \subseteq F(B)$  ;

That is to say that E is UC if and only if for all  $A \in P(E)$  and for all  $B \in P(E)$  there is a path from A to B or from B to A.

### 2.1.3 Hyperconnected

**Definition**

$\forall A \subseteq E, A \neq \emptyset, F(A) = E$ . Else  $\exists B \neq \emptyset, B \subseteq E - F(A)$ , such as  $A \subseteq F(B)$ .

E is HC if and only if  $\forall A \in P(E), \forall B \in P(E)$ , there is a path from A to B or there is a path from B to  $E - F(B)$ , part of E which meets A (the closure of any non-empty open set equals E).

### 2.1.4 Apo-connected

**Definition**

$\forall A, \forall B$  two non-empty subsets of E, we always have  $F(A) \cap F(B) \neq \emptyset$ .

E is AC if and only if  $\forall A \in P(E), \forall B \in P(E), \exists M \in P(E)$  with a path from M to A and a path from M to B.

### 2.1.5 Connected

E is C if and only if  $\forall A \in P(E), \forall B \in P(E)$ , there is a path from A to B or  $\exists M \in P(E)$  with a path from M to B and a path from M to  $E - F(B)$ , part of E which meets A.

Then :

- E is SC  $\Rightarrow$  E is UC
- E is UC  $\Rightarrow$  E is HC and E is AC
- E is HC  $\Rightarrow$  E is C
- E is AC  $\Rightarrow$  E is C

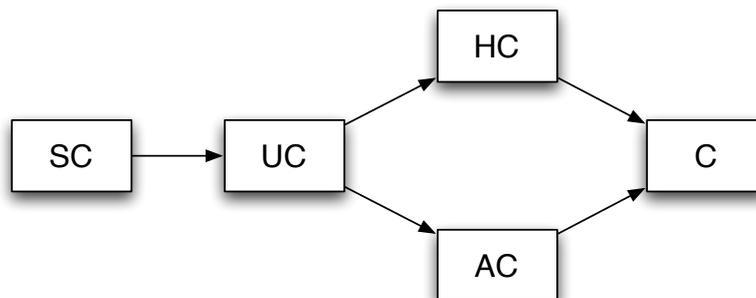
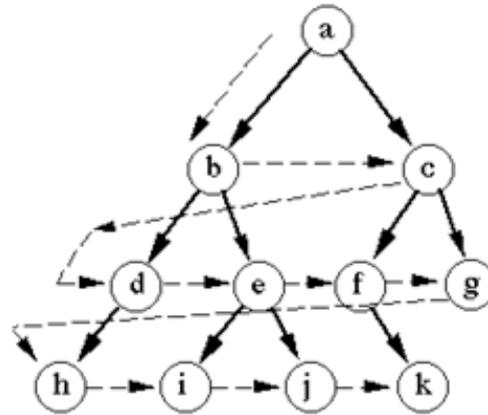


Figure 4 : X-connectivities

In the next section, we describe precisely our algorithm detecting unilaterally connected components.

### 3 Our algorithm

Our method objective is to find unilaterally connected components using a bottom-up approach. We use a breadth-first search looking for each node in order to aggregate them together.



Breadth-first search

Figure 5 : BFS

#### Algorithm

```

Method getAllUnilateralConnectedComponents(set E)
{UC : family of sets, stack : set, Tmp : arraylist, w,x,y : nodes, closure(x) : closure of x} UC ← ∅ T
mp ← ∅ stack ← E
for each x ∈ E do
    Tmp.put(x, closure(x))
end for
for k = stack.size() - 1; k >= 0; k -- do
    C ← ∅
    x ← stack.get(k)
    C ← x
    for i = 0; i < stack.size(); i ++ do
        y ← stack.get(i)
        if x ≠ y then
            if intersection(Tmp.get(y), C).size() == C.size() then
                C ← y
            else
                ok ← true
                itr ← C.iterator()
                while itr.hasNext() & ok do
                    w ← itr.next()
                    ok ← Tmp.get(w).contains(y)
                    if ok then
                        C ← y
                    end if
                end while
            end if
        end if
    end for
    UC.add(C)
end for

```

*end for*  
*return UC*

## 4 PretopoLIB

PretopoLib is a JAVA library that implements the concepts of pretopology [5]. Its interest lies in the representation of data structures allowing data manipulation with set operations. Even if libraries implementing graph theory already exist, pretopoLIB is the only library that implements pretopological concepts. It provides a framework for developing efficient algorithms for data mining, structuring and modeling complex systems.

PretopoLib is written in JAVA by Marc Bui and Vincent Levorato [11]. The objective was to provide a reusable and distributable tool. It contains two main sub-modules, PretopoNotion that implements mathematical concepts and PretopoVisual which is used as a wrapper for data visualization.

In the next section, we present some illustrations of our proposed extension to the PretopoLIB.

### 4.1 Demonstrations

The PretopoLib visualization module is based on two external frameworks : Prefuse and JUNG, originally developed to represent graphs. Each framework can represent different areas according to our needs.

To test and analyze our method, we used a graph generator that allows complex generations of undirected or oriented realistic graphs. These generated data can easily be used as pretopological structures with oriented binary relations. We chose the LFR benchmark generator that Lancichinetti and Radicchi proposed in 2008 [4] with the following configuration:

- $n$  the number of nodes in the graph ( $n=50$ )
- $k_{avg}$  the mean degree of nodes ( $k_{avg} = 5$ )
- $k_{max}$  the maximum degree of nodes ( $k_{max} = 10$ )
- $minc, maxc$  the minimum and maximum size of communities ( $minc = 2, maxc = 5$ )
- $\mu$ , the mixing parameter ie for each node, the ratio between the number of outgoing edges and the number of incoming edges in the community ( $\mu = 0.1$ )

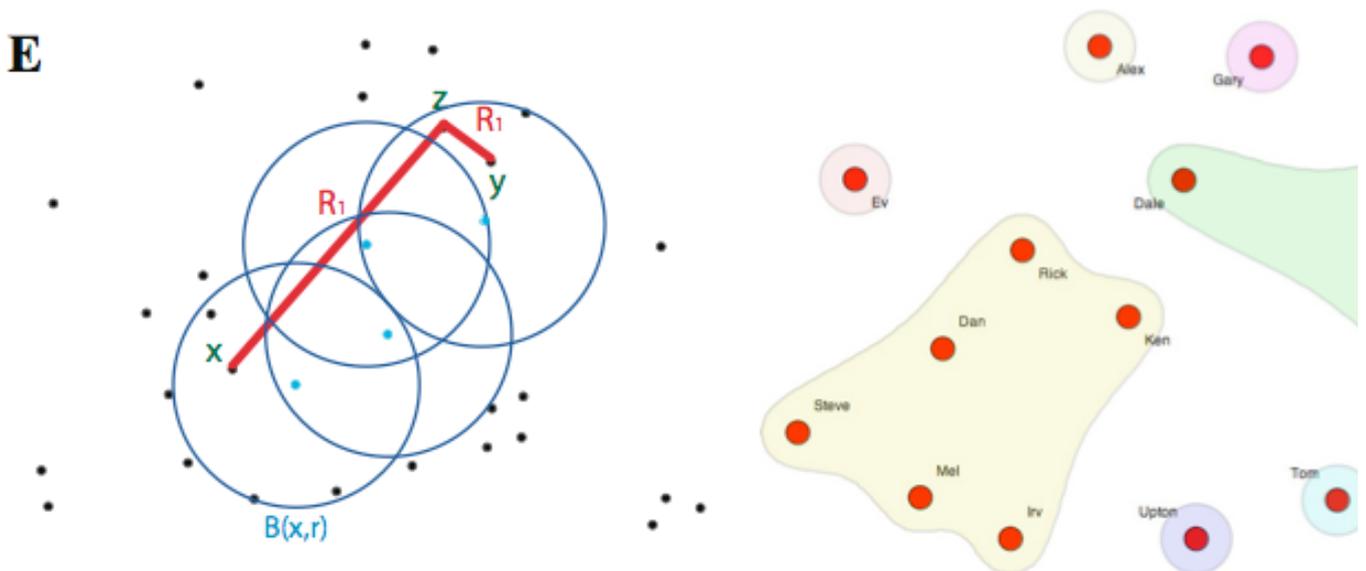


Figure 6 : Example of mixed pretopological space (left) and visualization (right).

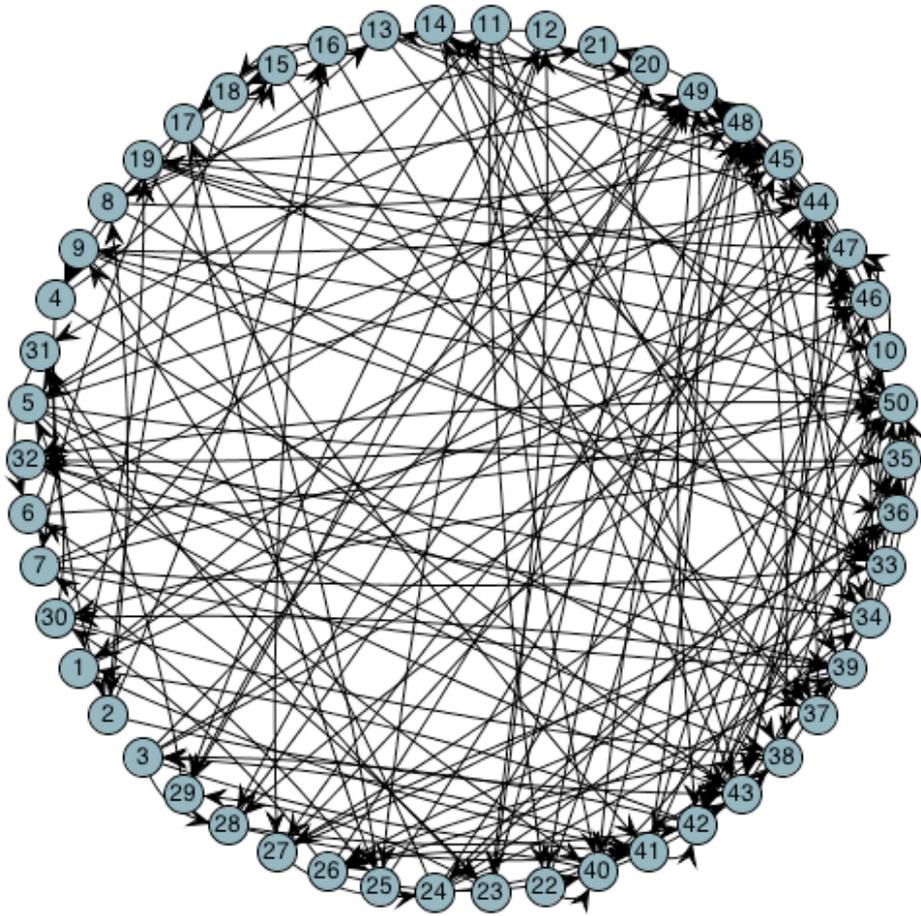


Figure 7 : pretopological space consisting of 50 nodes and oriented binary relations.

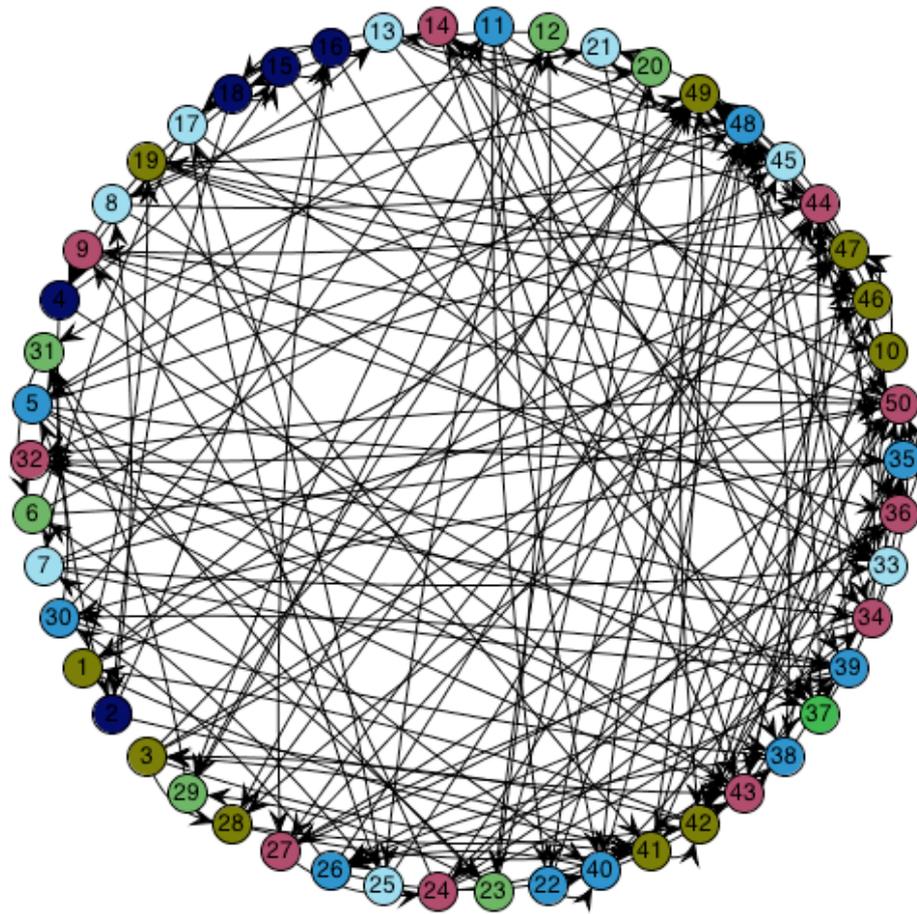


Figure 8 : structuring with unilaterally connected components (7 components)

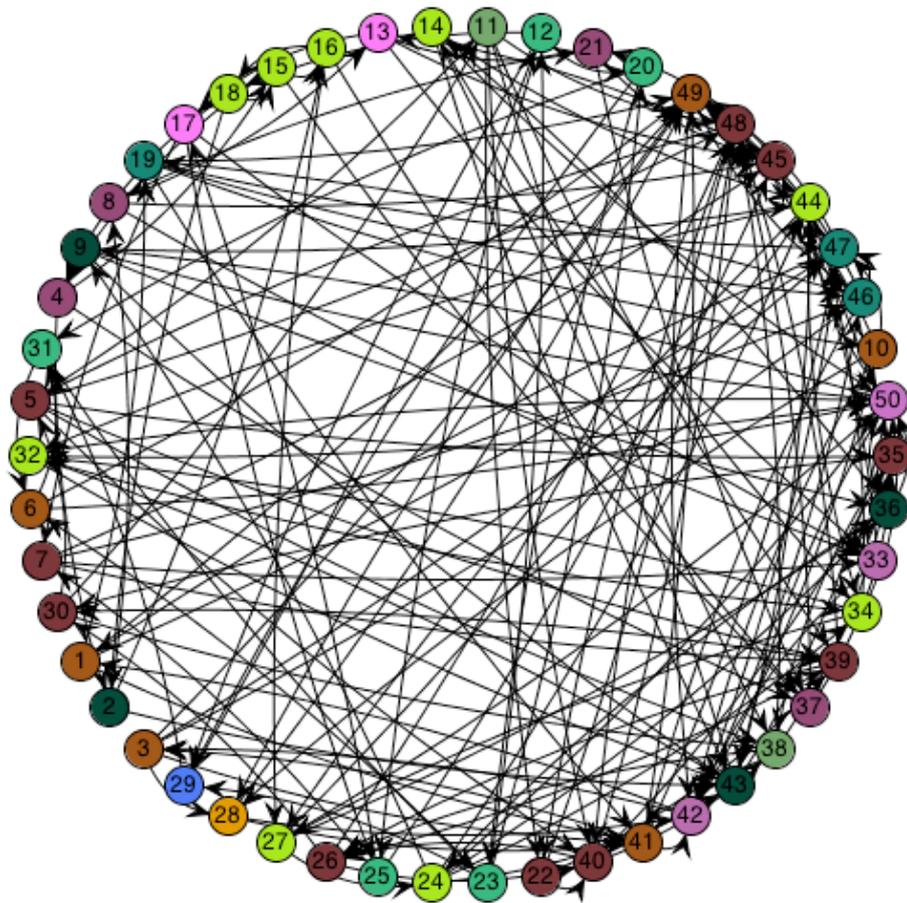


Figure 9 : structuring with strongly connected components (13 components)

As shown in the previous figures, we obtain different structures using unilaterally connected components or strongly connected components. Our algorithm allows different structures on the same pretopological space.

## 5 Conclusion

In this article, we focused on the detection of unilaterally connected components. Our main contribution relies on the extension of the Java library PretopoLIB through the addition of an algorithm to detect these components.

We still work to optimize our algorithm, in order to decrease the complexity. The accuracy of results obtained is interesting in the context of large networks strongly connected. It would be interesting to speed up processing times to perform a serie of tests on networks with a high average degree and a high number of nodes.

Eventually, we would continue our work concerning network structuring using communities and extend our previous classification algorithm to UC structuration [2].

## 6 Bibliography

[1] Berge, C. (1970). Graphes et Hypergraphes. Dunod, Paris.

[2] Vincent Levorato, Coralie Petermann. Detection of Communities in Directed Networks based on Strongly p-Connected Components 3rd IEEE International Conference on Computational Aspects of Social Networks (CASoN'2011), Salamanca, Spain ; 2011.

- [3] A. Lancichinetti and S. Fortunato, Community detection algorithms : A comparative analysis, Phys. Rev. E, vol. 80, p. 056117, 2009.
- [4] A. Lancichinetti, S. Fortunato, and F. Radicchi, Benchmark graphs for testing community detection algorithms, Phys. Rev. E, vol. 78, no. 4, p. 046110, 2008.
- [5] V. Levorato, S. Ben Amor, PretopoLib : la librairie JAVA de la Prétopologie, Extraction et Gestion des Connaissances (EGC), 2010.
- [6] M. E. J. Newman. The Structure and Function of Complex Networks. SIAM Review, Vol. 45, No. 2. (2003), pp. 167-256.
- [7] ZT. Belmandt. Basics of pretopology. Hermann, 155 pages, ISBN : 978 27056 8077, 2011.
- [8] Vincent Levorato and Murat Ahat. Modélisation de la dynamique des réseaux complexes associée à la prétopologie. 9eme congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision, ROADEF08, Clermont-Ferrand, France, February 2008.
- [9] V. Levorato. Modeling Groups in Social Networks, In Proceedings of the 25th European Conference on Modelling and Simulation (ECMS'2011), pages 129-134, Krakow, Poland, 2011.
- [10] V. Levorato. Contributions à la Modélisation des Réseaux Complexes : Prétopologie et Applications. PhD thesis, Université de Paris 8, 2008.
- [11] V. Levorato, M. Bui (2008). Pretopolib. Dépôt de logiciel à l'Agence pour la Protection des Programmes. IDDN.FR.001.470004.000.R.P.2008.000.20600.





