



HAL
open science

Modélisation et Analyse des réseaux complexes : Application à la sûreté nucléaire

Mouna Rifi, Mohamed Hibti, Rushed Kanawati, Younès Bennani

► To cite this version:

Mouna Rifi, Mohamed Hibti, Rushed Kanawati, Younès Bennani. Modélisation et Analyse des réseaux complexes : Application à la sûreté nucléaire. Congrès Lambda Mu 22 “ Les risques au cœur des transitions ” (e-congrès) - 22e Congrès de Maîtrise des Risques et de Sûreté de Fonctionnement, Institut pour la Maîtrise des Risques, Oct 2020, Le Havre (e-congrès), France. hal-03453580

HAL Id: hal-03453580

<https://hal.science/hal-03453580>

Submitted on 28 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Modélisation et Analyse des réseaux complexes : Application à la sûreté nucléaire Complex networks modeling and analysis : Application to nuclear safety

Mouna Rifi
EDF R&D

7 boulevard Gaspard Monge
91120 Palaiseau, France
mouna.rifi@edf.fr

Mohamed Hibti
EDF R&D

7 boulevard Gaspard Monge
91120 Palaiseau, France
mohamed.hibti@edf.fr

Rushed Kanawati

Université Sorbonne Paris nord
LIPN, UMR CNRS 7030 - Institut Galilée
99, avenue Jean-Baptiste Clément
93430
Villetaneuse, France
rushed.kanawati@sorbonne-paris-nord.fr

Younès Bennani

Université Sorbonne Paris nord
LIPN, UMR CNRS 7030 - Institut Galilée
99, avenue Jean-Baptiste Clément
93430
Villetaneuse, France
younes.bennani@sorbonne-paris-nord.fr

Résumé— L'objectif de cet article est d'explorer les Etudes Probabilistes de Sûreté nucléaire (EPS) par des approches basées sur la théorie des graphes afin de leur apporter un éclairage complémentaire. Pendant les phases de design où la réalisation des EPS est impossible, cet éclairage est important.

Abstract— This paper aims to explore nuclear Probabilistic Safety Assessments studies (PSA) by graph-based theory approaches in order to provide additional insights. Particularly in the design phases of the systems, where the realization of the EPS is impossible.

Mots clés— Réseaux Complexes, Sûreté nucléaire, Centralités, Similarité de graphes, Facteur d'Accroissement du Risque, Corrélation de Kendall, Clustering de graphes.

Keywords— Complex Networks, nuclear safety, centrality, graph similarity, Risk Increase Factor, Kendall correlation, graph-clustering.

I. INTRODUCTION

Les réseaux complexes sont de plus en plus utilisés dans plusieurs domaines pour étudier les liens entre différents objets qu'ils modélisent, la structure sous-jacente de ces systèmes et éventuellement certains patterns qu'ils peuvent contenir.

Plusieurs travaux se sont intéressés à l'étude des réseaux complexes. Ils sont regroupés en 4 catégories selon [1] :

- Des réseaux sociaux :
Modélisant les relations entre des directeurs d'entreprises [2, 3], les liens d'amitié entre écoliers [4], ou encore des graphes d'appels téléphoniques [5, 6].
- Des réseaux biologiques :
Modélisant les interactions entre protéines [7, 8], le réseau de régulation génétique [9] ou le système nerveux [2, 10].
- Des réseaux technologiques :
Modélisant le trafic aérien [11], les réseaux de routes [12] ou encore ceux des chemins de fer [13, 14].
- Réseaux d'information :
Modélisant les citations scientifiques [15], le réseau Web [16, 17, 18, 19] ou les réseaux de préférences (pour les systèmes de recommandations) [20, 21, 22]

Ruiz-Martin et al. [23] utilisent les réseaux complexes pour analyser la résilience organisationnelle ¹ (contre la défaillance des différents canaux de communication) dans l'exploitation nucléaire en Espagne. Ce travail exploite le Plan d'Urgence Nucléaire externe issu du journal officiel espagnol 1985 et 2009.

¹ La résilience organisationnelle, selon l'article, est « la capacité d'une organisation à faire face à une épreuve (ou un incident) et à la surmonter dans des conditions qui peuvent être défavorables ».

Quelques travaux ont été réalisés en rapport avec les Etudes Probabilistes de Sûreté (EPS) notamment dans [24] et [25]. Des algorithmes basés sur la théorie des graphes sont utilisés pour rechercher des combinaisons minimales de défaillances, ou des coupes minimales conduisant à un événement indésirable à partir d'un événement initiateur.

Dans [26] un réseau est construit à partir d'une séquence d'événements qui décrit les scénarios accidentels suivant l'occurrence d'un événement initiateur. Ce travail met en évidence l'existence d'une forte relation entre la mesure d'intermédiarité (élevée) et la fréquence d'occurrence dans la liste des ensembles de coupes de la séquence correspondante. Ainsi, il propose une expérimentation de métriques de la théorie des graphes pour analyser des modèles probabilistes de sûreté des centrales nucléaires.

Le corps de cet article se compose de 4 parties suivies de conclusions et perspectives.

La partie II définit les notions de la théorie des réseaux complexes ainsi que celles des EPS nécessaires à la lecture de cet article.

La partie III décrit la méthode de construction du réseau complexe d'un initiateur dans un état donné.

La partie IV regroupe deux applications des mesures de centralités à la sûreté : le premier paragraphe présente les différents composants et systèmes identifiés comme importants par ces mesures; le second consiste à utiliser ces mesures de centralité pour prédire le RIF (Risk Increase Factor), qui est une mesure d'importance dans les EPS.

La dernière partie propose une méthode de comparaison structurelle de systèmes complexes en représentant chaque réseau correspondant par un coefficient vectoriel synthétique de centralité basé sur les interdépendances entre les vecteurs de centralité. La comparaison se fait par Clustering Ascendant Hiérarchique appliqué à ces coefficients vectoriels synthétiques de centralité.

II. QUELQUES DEFINITIONS

Dans cette partie, nous présentons quelques définitions des concepts et éléments de la théorie des réseaux complexes puis celles des Etudes probabilistes de Sûreté (EPS).

A. Terminologie des réseaux complexes

1) Principales propriétés des graphes

- **Sommet**
Il s'agit de l'unité fondamentale d'un réseau, il est également appelé site en physique, nœud en informatique ou acteur en sociologie.
- **Arête (réseaux simples) ou arc (réseaux dirigés)**
Elle relie deux sommets dans un réseau. Elle est aussi appelé liaison en physique, lien en informatique ou relation en sociologie.
- **Réseau complexe**
Il s'agit d'un graphe qui modélise les objets d'un système réel et dépendances ou liens entre ces objets. La plus part des réseaux complexes ont des caractéristiques communes, nous en citons quelques-unes :

▪ **Distribution des degrés**

La distribution des degrés $P(k)$ est la probabilité qu'un sommet ait k sommets voisins dans le réseau (centralité de degré). Les réseaux complexes ont une distribution des degrés qui suit une loi de puissance, ce qui signifie que très peu de sommets ont un nombre élevé de liens.

▪ **Diamètre**

Il s'agit de la valeur maximale obtenue par les longueurs plus courts chemins entre chaque paire de sommets du réseau. La longueur d'un chemin est le nombre de liens du chemin. Le diamètre de la plupart des réseaux complexes est faible par rapport au nombre de sommets.

▪ **Densité**

La densité d'un réseau est définie comme étant la proportion de liens réellement présents par rapport au nombre maximal de liens possibles. Les réseaux complexes sont connus par leur faible densité, ce qui signifie qu'ils sont très clairsemés (sparse networks).

▪ **Coefficient de Clustering**

Cette mesure est plutôt adaptée aux graphes simples (où les liens ne sont pas orientés), ce qui n'est pas le cas des réseaux complexes modélisant des séquences accidentelles ou des systèmes de sûreté nucléaire que nous étudions.

Dans de nombreux réseaux réels, on observe souvent que deux sommets connectés à un même sommet ont tendance à être également connectés. Dans le contexte des réseaux de relations sociales, cela revient à dire que l'ami de mon ami est également susceptible d'être mon ami. Cette propriété est appelée Clustering, elle est mesurée par le coefficient de Clustering local qui correspond à la probabilité que les voisins du sommet étudié soient connectés.

2) Mesures de centralités :

- **Degré entrant (In degree) [27]**
Elle correspond au nombre d'arcs entrants vers le sommet étudié, cette valeur est souvent normalisée.
- **Degré sortant (Out degree) [27]**
Cette mesure correspond au nombre d'arcs sortants du sommet étudié, elle est souvent normalisée.
- **Proximité entrante (In closeness) [27]**
La proximité est basée sur la moyenne des plus courts chemins reliant le sommet en question aux autres sommets du graphe. Elle prend en compte dans le calcul des plus courts chemins l'orientation des liens (liens entrants au sommet). Ainsi, un sommet avec une proximité entrante élevée est un sommet vers lequel plusieurs plus courts chemins mènent.

- **Proximité sortante (Out closeness) [27]**
Son calcul prend en compte l'orientation des liens des plus courts chemins. Ici, nous nous intéressons aux liens sortants du sommet.
Ainsi, un sommet avec une proximité sortante élevée est un sommet dont plusieurs plus courts chemins sortent.
- **Centralité d'intermédiarité (Betweenness) [27]**
Cette mesure se calcule par le nombre des plus courts chemins du graphe passant par chaque sommet. Une valeur élevée de cette centralité n'est pas forcément corrélée avec un degré important du sommet.
Les sommets avec des intermédiarités importantes sont des « points de passages indispensables » pour relier rapidement deux sommets du graphe.
- **Page Rank**
Cette mesure est utilisée par l'algorithme de Google pour indexer les pages web [28].
Le Page Rank d'un sommet dépend de celui de tous les autres sommets, donc le calcul du Page Rank d'un sommet implique le calcul de ceux de tous les autres sommets.
Typiquement, un sommet important a des liens qui proviennent de sommets eux aussi importants.
- **Hub et Autorité**
L'algorithme des hubs et autorités a été développé par [29] et initialement utilisé pour caractériser des pages web. Les hubs (sommets avec des valeurs élevées de la métrique Hub score) sont supposés contenir des catalogues avec un nombre important de liens sortants. Les autorités sont les sommets qui reçoivent plusieurs liens de ces hubs. Ainsi, l'autorité d'un sommet est proportionnelle à la somme des Hub scores des sommets ayant des liens entrants vers ce dernier. Ainsi, par symétrie le Hub score d'un sommet est proportionnel aux valeurs d'autorité des sommets qui reçoivent des liens provenant de lui.

B. Terminologie des EPS

- **Etude Probabiliste de Sûreté**
Cette méthode, apparue avec le rapport WASH-1400 de Rasmussen [30], permet d'analyser de manière systématique et de quantifier la fréquence d'un évènement indésirable. Elle fournit un aperçu quantitatif et global en intégrant à la fois la gravité d'un évènement et sa fréquence d'occurrence obtenue grâce aux calculs probabilistes.
- **Evènement Initiateur**
Il s'agit d'un aléa (incendie, rupture de tuyauterie, explosion, etc.) interne ou externe qui peut causer un évènement indésirable (par exemple : les défaillances intrinsèques de matériels, la fusion du cœur d'un réacteur nucléaire).

- **Séquence accidentelle**
Une séquence accidentelle est une succession de succès et d'échecs de missions de sauvegarde mises en place pour mitiger les effets d'un évènement initiateur et conduire l'installation nucléaire à un état sûr. Une séquence accidentelle peut conduire à une conséquence soit acceptable ou inacceptable. Elle peut aussi renvoyer vers une séquence accidentelle partagée avec un autre initiateur.

- **Risk Increase Factor (RIF)**

Il s'agit d'une des mesures d'importance utilisées en sûreté nucléaire [31]. Elle permet d'évaluer l'importance d'un composant de point de vue de l'évaluation de sûreté en estimant l'impact des dégradations sévères d'un composant sur l'accroissement du risque. Elle correspond au facteur d'accroissement du risque obtenu en supposant que le composant étudié est entièrement défaillant ou indisponible par rapport au risque de référence R_{base} . Le calcul du RIF d'un composant x_i est obtenu par la formule suivante :

$$RIF(x_i) = \frac{R(x_i = 1)}{R_{base}}$$

- **Design équilibré**

Le caractère bien équilibré du design d'une installation est défini comme la propriété qu'aucune classe particulière d'accidents ou caractéristique n'apporte de contribution disproportionnée au risque global [32]. En d'autres termes, «aucun composant seul, système, phénomène ou tout autre facteur n'est prédominant dans l'évaluation de risque et que la part des risques difficiles à gérer est aussi bas que possible» [33].

III. METHODE DE CONSTRUCTION DE RESEAUX COMPLEXES D'UN SYSTEME OU D'UNE SEQUENCE ACCIDENTELLE

Dans des travaux antérieurs [26, 34], une méthode de construction de graphes dirigés et attribués à partir d'un ensemble de séquences accidentelles a été proposée et appliquée à des cas réels. Nous proposons dans cette partie de rappeler cette méthode de construction et de caractériser le réseau complexe construit à partir d'un initiateur.

A. Méthode de construction des réseaux complexes

On s'intéresse à la modélisation en réseau de séquences accidentelles relatives à un initiateur dans un état du réacteur. Les analyses fonctionnelles des séquences accidentelles qui permettent de caractériser les missions de sauvegarde à mettre en œuvre pour assurer la sûreté de l'installation sont nommées Analyses Qualitatives de Séquences (AQS).

Prenons l'exemple simplifié (voir Figure 1) relatif à un évènement initiateur (IE), incluant 3 missions de sauvegarde et modélisant 2 conséquences : acceptable (AC) et inacceptable (UC).

Les transitions en vert représentent les succès des missions tandis que celles en rouge correspondent aux échecs.

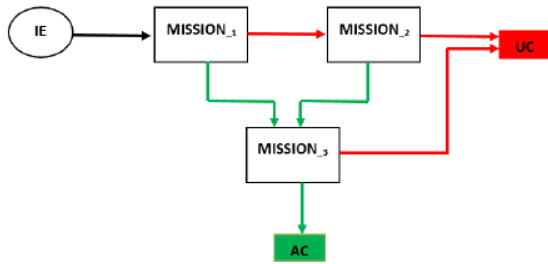


Figure 1 : Exemple simplifié de séquences accidentelles d'un initiateur IE

Un macro-réseau (ou réseau squelette) est produit en considérant comme macro-sommets les différentes missions, et comme liens les transitions d'échecs et de succès menant à une conséquence inacceptable.

Ensuite, pour chaque mission, l'idée est de construire le réseau correspondant au système dont elle dépend puis de le réduire aux composants qui sont impliqués dans la réalisation de cette mission. Les missions peuvent se réduire à un facteur humain.

Ces systèmes sont soit frontaux, c'est-à-dire impliqués directement dans la réalisation du scénario de sauvegarde, soit supports (électriques, de contrôle commande, ventilation) qui permettent d'alimenter ou commander les différents composants des systèmes frontaux.

L'extraction des éléments de ces systèmes est réalisée par du parsing des fichiers Figaro 0 de ces systèmes à partir du logiciel de sûreté de fonctionnement KB3 [36].

Les macro-sommets sont remplacés par les réseaux des missions correspondantes qui sont obtenus en réduisant les réseaux des systèmes pour n'en garder que ce qui est impliqué dans la réalisation de la mission.

Les réseaux des missions supports sont connectés aux sommets (composants) qu'elles alimentent ou commandent. Ainsi, les sommets du réseau complexe d'un évènement initiateur à un état donné du réacteur est un réseau dirigé, puisque les liens sont orientés.

Les sommets de ces réseaux complexes sont les différents composants des systèmes de sauvegarde et de support de types pompes, vannes, clapets, diesel, tableau électrique, disjoncteurs, acquisitions, etc. 55 types de composants sont pris en compte dans notre modélisation.

Les liens représentent les différents flux (hydrauliques, électriques, signaux, puis échecs ou succès de missions). Ces réseaux complexes sont également attribués car ils modélisent les caractéristiques propres aux composants et flux tels que leurs types. La modélisation prévoit également l'ajout d'autres attributs tels que ceux de localisation.

B. Application à un évènement initiateur

Le réseau complexe de l'initiateur étudié à une étape intermédiaire du projet a été obtenu en appliquant la méthode de construction de réseaux complexes décrite dans le paragraphe précédent. Il est représenté dans la Figure 2. Les couleurs représentent les types de composants des missions impliqués dans ces séquences accidentelles.

Il s'agit d'un réseau dirigé connexe et acyclique.

Il se compose d'environ 1700 sommets et de 2700 liens, ce qui fait de lui un très petit réseau comparé à ceux étudiés

dans la littérature tel que celui des utilisateurs de Facebook ou encore des liens entre les pages de Google (10^6 pages). La densité de ce réseau est faible (0.009), son diamètre est moyennement grand (33), c'est-à-dire que le transit dans le graphe entre toute paire de sommets nécessite au maximum 33 pas. Il présente une distribution des degrés en loi de puissance ce qui confirme son appartenance à la classe réseau complexe.

Les mesures de centralités permettront d'apporter plus d'information sur sa structure, d'autant plus qu'elles ne sont pas complexes à calculer vu la taille du réseau.

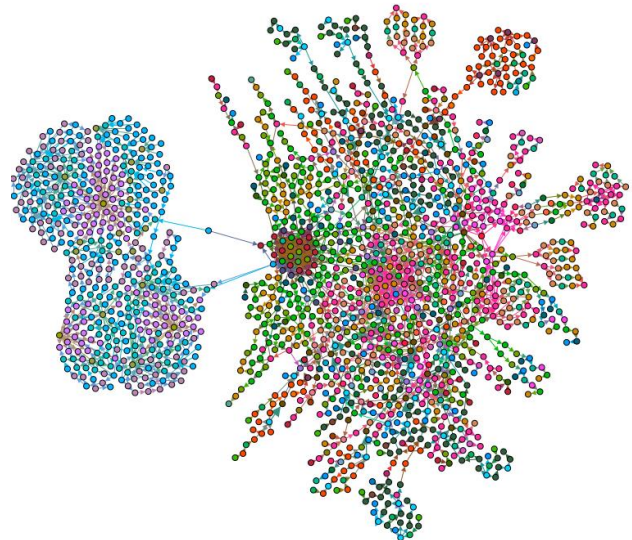


Figure 2 Réseau complexe d'un initiateur dans une étape intermédiaire du projet.

IV. MESURES DE CENTRALITES APPLIQUEES A LA SURETE

Cette partie répond à deux objectifs : Le premier est d'explorer la structure du réseau complexe, décrit dans la section précédente, par les mesures de centralités présentées dans le paragraphe II-A-2.

Le second consiste à utiliser ces mesures de centralité pour prédire le RIF, qui est une mesure complexe à calculer car l'obtention de sa valeur pour chaque composant est basée sur le calcul du risque de l'évènement redouté en supposant la défaillance du composant étudié.

Un travail antérieur [26] met en évidence l'existence d'une forte relation entre la mesure d'intermédiarité (élevée) et la fréquence d'occurrence dans la liste des ensembles de coupes de la séquence étudiée dans cet article. En effet, les composants avec des intermédiarités élevées sont très présents dans la liste des coupes minimales. L'analyse a été réalisée sur un cas d'étude différent de celui utilisé dans cette partie.

A. Analyse du réseau complexe par les mesures de centralité.

Nous avons appliqué chacune des mesures de centralités sur les sommets du réseau complexe de l'initiateur étudié précédemment. Rappelons qu'une mesure de centralité est un score compris entre 0 et 1 et donc un classement d'importance.

Les distributions des valeurs de centralités sont représentées dans les histogrammes de la Figure 3. Nous remarquons qu'elles sont différemment distribuées et que les degrés entrants et sortants sont bien en loi de puissance.

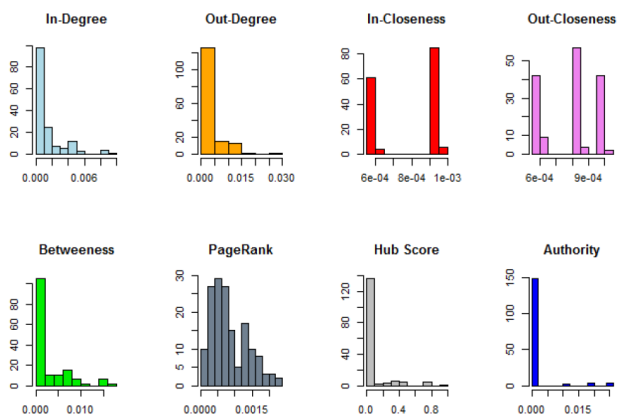


Figure 3 : Distributions des valeurs de centralités des sommets du réseau complexe

Afin d'identifier les types de composants et leurs systèmes révélés comme importants par chaque mesure de centralité, nous sélectionnons les 20 sommets (composants) les mieux classés par chaque mesure de centralité [34]. Ainsi nous remarquons que :

- Les sommets de degrés entrants élevés appartiennent tous au système SIS (l'injection de sécurité) tandis que ceux avec des degrés sortants élevés sont en premier lieu ceux du système I&C (de contrôle commande), ensuite viennent des composants de systèmes hydrauliques SIS, CVCS et CWS et ayant des types différents (vannes, pompes, et autres).
- Concernant la proximité entrante, il s'avère que tous les composants les mieux classés proviennent de systèmes hydrauliques SIS, CVCS et RCP. La proximité sortante, quant à elle, désigne comme importants uniquement des composants du système électrique.
- L'intermédierité et le Page Rank valorisent uniquement des composants de systèmes hydrauliques en premier lieu celui de SIS puis CHRS pour l'intermédierité et RCP pour le Page Rank.
- Le Hubs score classe en tête les clapets et vannes du système SIS puis des composants divers du contrôle commande. L'autorité retient comme importants les composants du système de contrôle commande de type logique spécifique ainsi que très peu de composants hydrauliques du système SIS.

Ainsi, chaque centralité permet d'avoir un angle de vision différent d'un réseau complexe, et donc des séquences accidentelles modélisées et plus généralement de tout système complexe.

Ces analyses sont intéressantes, même en phase de design, dans la mesure où elles permettent de vérifier la cohérence des différents niveaux de défense en profondeur introduits par les concepteurs de systèmes grâce aux différents niveaux de redondance ou d'interconnexions. Il s'agit donc d'un nouveau point de vue.

B. Prédiction du RIF par les mesures de centralités.

L'objectif des travaux [37, 38] était de prédire le RIF, une mesure d'importance très utilisée dans les études probabilistes de sûreté, mais qui est complexe à calculer. Cette prédiction utilise comme variables les centralités des réseaux complexes.

Le cas d'étude ici est le réseau complexe de l'initiateur étudié précédemment.

Etant donné le coût de calcul du RIF, l'étude s'est faite sur un échantillon décrit dans le paragraphe précédent. Il comprend pour rappel l'ensemble des composants identifiés comme importants par chaque centralité (TOP 20).

La variable d'intérêt est le RIF et les variables potentiellement prédictives sont les 8 différentes centralités. Nous proposons de catégoriser la variable RIF en deux classes. La première correspond aux valeurs faibles et la seconde celles des valeurs élevées du RIF qui pointent les composants critiques du point de vue sûreté, c'est-à-dire dont la défaillance implique une forte augmentation du risque.

Nous souhaitons prédire la seconde classe qui est minoritaire car elle ne représente que 5% de l'échantillon, il s'agit donc d'un problème de classification déséquilibrée.

Les méthodes de Machine Learning utilisées sont la régression logistique et les ensembles méthodes basées sur des arbres de décision. Elles sont bien décrites dans l'un des livres de référence en apprentissage statistique [39].

Chacune des deux méthodes maximise un des paramètres d'évaluation de la qualité du modèle, à savoir la précision et le rappel.

La régression logistique serait utile si l'objectif est de fournir à un analyste de risque une liste exhaustive de composants potentiellement critiques de point de vue sûreté à étudier avec attention. Néanmoins, il ne pourra pas s'en servir pour justifier la criticité certaine des composants car la précision du modèle est faible (33%).

Cependant, si l'objectif est d'être précis, comme lorsqu'il est question de prendre des décisions, la méthode Gradient Boosted Machine permet d'avoir de meilleurs résultats à savoir une précision parfaite (100%), mais un rappel moins parfait que celui de la régression logistique (67%).

Notons que la seconde méthode fournit un meilleur compromis entre précision et rappel.

Malgré le fort déséquilibre de classes, les résultats de prédiction obtenus restent satisfaisants.

V. COMPARAISON STRUCTURELLE DE SYSTEMES

Ce travail propose une nouvelle mesure de similarité structurelle pour les systèmes complexes.

Cette comparaison se fait indépendamment de leurs tailles, nombre de relations (liens, dépendances), densité ou autres mesure caractérisant les réseaux complexes modélisant ces systèmes.

En effet, elle se base uniquement sur les différents classements des sommets de ces réseaux par chacune des mesures de centralité.

Chacune des centralités étant une représentation différente du réseau complexe, le fait de combiner toutes ces représentations permet d'obtenir une représentation plus riche de la structure globale de chaque système.

A. Coefficient vectoriel synthétique de centralité et mesure de similarité

Un coefficient vectoriel synthétique de centralité est défini dans le but de résumer l'information concernant les variations des distributions des différentes centralités dans un même réseau.

Cela se fait en capturant les interdépendances entre les mesures de centralité en utilisant la corrélation non-paramétrique de Kendall.

Ainsi, comparer deux réseaux revient tout simplement à comparer leurs vecteurs synthétiques de centralités respectifs.

Le coefficient vectoriel synthétique de centralité a comme coordonnées les coefficients de corrélation de Kendall entre tous les vecteurs de centralité pris par paires dans l'ordre bien entendu.

Pour un réseau dirigé, nous calculons pour les sommets de sa plus grande composante connexe, 7 vecteurs de centralité (Degrés entrant et sortant, Proximités entrante et sortante, Intermédierité, Hub score et Autorité). Le coefficient vectoriel synthétique de centralité possède donc 21 coordonnées.

Pour un réseau simple, il ne sera calculé que 4 vecteurs de centralité (Degré, Proximité, Intermédierité et Hub score).

Le coefficient vectoriel synthétique de centralité a donc 6 coordonnées. La mesure de similarité entre réseaux est définie comme suit :

$$Sim(G_1, G_2) = dist(Kcor[G_1], Kcor[G_2])$$

Où G_1 et G_2 sont deux réseaux tous deux dirigés ou simples. $dist()$ est une distance de \mathbb{R}^2 .

B. Validation sur des jeux de données de graphes de benchmark

Afin de tester les performances et valider l'efficacité de cette nouvelle mesure de similarité, nous l'avons appliquée pour effectuer un Clustering sur différents types de graphes théoriques. L'étude a été réalisée sur 5 jeux de données très diversifiés. Les résultats produits sont satisfaisants. Cette méthode a montré son efficacité pour discriminer les différents types de réseaux dans chaque jeu de données. Pour plus de détails, le lecteur peut se référer au manuscrit de la thèse [40].

C. Application aux systèmes de sûreté d'une installation nucléaire

L'objectif est de pouvoir comparer les structures des différents systèmes de l'installation nucléaire étudiée.

23 systèmes ont été modélisés en réseaux selon la méthodologie décrite précédemment.

Les 23 réseaux complexes ont tous des diamètres proches de 13, des densités autour de 0.1%, des distributions des degrés en loi de puissance ce qui confirme leur appartenance à la classe des réseaux complexes. Ils ont des caractéristiques topologiques semblables et nous souhaitons les départager en groupes structurellement semblables.

Dans un premier temps il est question de construire le coefficient vectoriel synthétique de centralité $Kcor(G)$ pour chaque réseau G .

Rappelons que cela se fait sur la plus grande composante connexe du réseau G s'il n'est pas connexe lui-même.

Ensuite il est question de calculer les vecteurs de centralité sur les sommets de cette plus grande composante connexe, puis les corrélations entre les vecteurs de centralité pris par paires pour obtenir le coefficient vectoriel synthétique de centralité.

Un Clustering est ensuite réalisé en utilisant la méthode de CAH (Clustering Ascendant Hiérarchique) [39] après avoir repéré les composantes principales par ACP (Analyse en Composantes Principales).

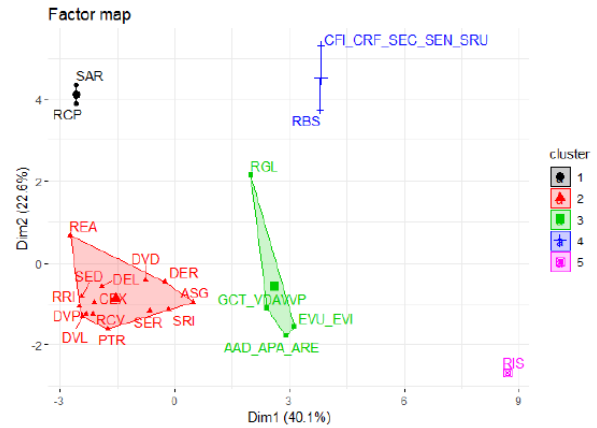


Figure 4 : Représentation des réseaux des systèmes de sûreté nucléaire sur le premier plan factoriel

La Figure 4 représente la projection des différents réseaux complexes sur le premier plan factoriel de l'ACP. Ce plan représente environ 63% de l'information. Nous remarquons que le réseau RIS est isolé et éloigné des autres. Les réseaux représentés par chaque couleur sont proches entre eux et loin des autres. Les groupes en rouge et en vert sont légèrement proches. La fusion de groupes peut être envisagée ou non.

Le dendrogramme² du Clustering est représenté dans la Figure 5.

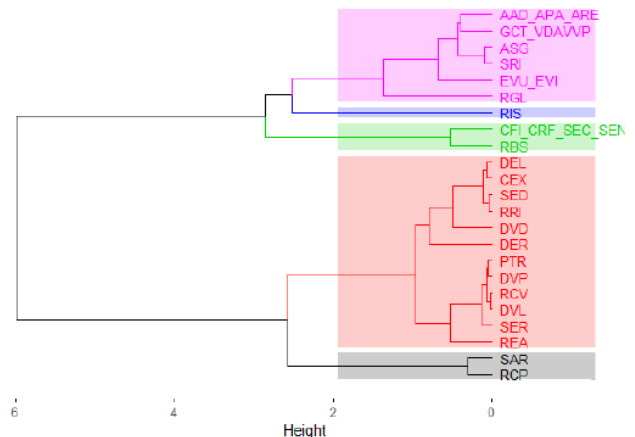


Figure 5 : Dendrogramme du Clustering des réseaux complexes modélisant les systèmes de sûreté nucléaire

Les systèmes affectés au même groupe sont des systèmes structurellement similaires. Deux systèmes éloignés dans cet

² La CAH rassemble les individus de manière itérative afin de produire un dendrogramme ou arbre de classification.

arbre le sont car très différents par la mesure de similarité structurelle définie.

Cette approche a été utilisée dans [41] pour étudier le caractère équilibré du design en se basant sur les réseaux complexes. En effet, les systèmes et composants ont été étudiés d'un point de vue structurel et indépendamment des données de fiabilité.

Les classifications des systèmes, des séquences et des conséquences ont pour but de renforcer la démonstration du design équilibré de l'installation nucléaire en plus des autres approches dites probabilistes (la contribution de chaque système ou séquence accidentelle au risque global CDF (Core Damage Frequency) ou LERF (Large Early Release Frequency), etc.).

Il s'agit à ce jour de la seule méthode, à notre connaissance, qui peut faire la preuve du caractère équilibré dans une phase très précoce sans que l'on puisse disposer d'EPS proprement dite.

VI. CONCLUSIONS ET PERSPECTIVES

Dans ces travaux de thèse, l'objectif était d'explorer les EPS par les approches de réseaux complexes.

Nous avons proposé, dans un premier temps, une méthode de construction des réseaux complexes d'un système ou d'une séquence accidentelle en réseau dirigé et attribué.

L'analyse du réseau complexe d'un initiateur réel par les mesures de centralité a permis de mettre en évidence les différents types de composants et systèmes identifiés comme importants par chaque mesure de centralité.

Les mesures de centralité ont été utilisées pour prédire le RIF qui est une mesure d'importance très utilisée dans les EPS. Les méthodes de Machine Learning appliquées pour la prédiction sont les méthodes d'ensembles Learning à base d'arbres de classification et la régression logistique. Malgré le fort déséquilibre de classes, les résultats obtenus sont satisfaisants.

D'autre part, une méthode de comparaison structurelle de systèmes complexes indépendamment de leurs tailles ou d'autres propriétés topologiques a été mise en place.

Elle a été validée dans un premier temps sur des réseaux de benchmark, puis appliquée au Clustering des systèmes de sûreté nucléaire. Une des applications identifiées de ce Clustering est l'aide à la démonstration du caractère équilibré du design. Nous souhaitons réaliser ultérieurement le Clustering de plusieurs initiateurs.

Comme perspectives de ces travaux, nous souhaitons étendre les analyses réalisées sur des cas d'études à plusieurs initiateurs pour fournir des enseignements plus globaux, et identifier ceux qui sont plus spécifiques à des initiateurs.

D'autre part, l'exploration des phénomènes de propagation dans les réseaux complexes permettront d'étudier : les propagations d'agressions internes ou externes telles que les incendies, inondations ou autres; des simulations des réparations.

Les aspects probabilistes relatifs à la contribution au risque des composants, systèmes et séquences accidentelles (en particulier les données de fiabilité) ne sont pas encore pris en compte dans cette approche, ils pourraient être intégrés

par une extension qui prend en compte des sommets de type pondéré.

VII. REFERENCES

- [1] Newman, M., « *The structure and function of complex networks.* » SIAM review 45.2, 2003, pp 167-256.
- [2] Amaral, L. A. N., Scala, A., Barthélémy, M and Stanley, H. E. « *Classes of small-world networks* », Proc. Natl. Acad. Sci. USA 97, 2000, pp 11149–11152.
- [3] Watts, D. J. and Strogatz, S. H., « *Collective dynamics of 'small-world' networks* », Nature 393, 440–442 (1998)..
- [4] Fararo, T. J. and Sunshine, M., « *A Study of a Biased Friendship Network* », Syracuse University Press, Syracuse (1964).
- [5] Aiello, W., Chung, F., and Lu, L., « *A random graph model for massive graphs* », in *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pp.171–180, Association of Computing Machinery, New York (2000).
- [6] Aiello, W., Chung, F., and Lu, L., « *Random evolution of massive graphs*, in J. Abello, P. M. Pardalos, and M. G. C. Resende (eds.) », *Handbook of Massive Data Sets*, pp. 97–122, Kluwer, Dordrecht (2002).
- [7] Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y., « *A comprehensive two-hybrid analysis to explore the yeast protein interactome* », Proc. Natl. Acad. Sci. USA 98, 4569–4574 (2001).
- [8] Maslov, S. and Sneppen, K., « *Specificity and stability in topology of protein networks* », Science 296, 910–913 (2002).
- [9] Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadmodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M., « *A comprehensive analysis of protein–protein interactions in saccharomyces cerevisiae* », Nature 403, 623–627 (2000).
- [10] White, J. G., Southgate, E., Thompson, J. N., and Brenner, S., « *The structure of the nervous system of the nematode C. Elegans* », Phil. Trans. R. Soc. London 314,1–340 (1986).
- [11] Adamic, L. A., « *The small world web*, in *Lecture Notes in Computer Science*», vol. 1696, pp. 443–454, Springer, New York (1999).
- [12] Kalapala, V. K., Sanwalani, V., and Moore, C., « *The structure of the United States road network*», Preprint, University of New Mexico (2003).
- [13] Latora, V. and Marchiori, M., « *Is the Boston subway a small-world network?* », Physica A 314, 109–113 (2002).
- [14] Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., and Manna, S. S., « *Small-world properties of the Indian railway network* », Preprint cond-mat/0208535 (2002).
- [15] Egghe, L. and Rousseau, R., « *Introduction to Informetrics* », Elsevier, Amsterdam (1990).
- [16] Albert, R., Jeong, H., and Barabasi, A.-L., « *Diameter of the world-wide web* », Nature 401, 130–131 (1999).
- [17] Barabasi, A.-L., Albert, R., and Jeong, H., « *Scale-free characteristics of random networks: The topology of the World Wide Web* », Physica A 281, 69–77 (2000).
- [18] Kleinberg, J. M., Kumar, S. R., Raghavan, P., Rajagopalan, S., and Tomkins, A., « *The Web as a graph: Measurements, models and methods* », in *Proceedings of the International Conference on Combinatorics and Computing*, no. 1627 in Lecture Notes in Computer Science, pp. 1–18, Springer, Berlin (1999).
- [19] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J., « *Graph structure in the web* », Computer Networks 33, 309–320 (2000).
- [20] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D., « *Using collaborative filtering to weave an information tapestry* », Comm. ACM 35, 61–70 (1992).
- [21] Resnick, P. and Varian, H. R., « *Recommender systems* », Comm. ACM 40, 56–58 (1997).
- [22] Shardanand, U. and Maes, P., « *Social information filtering: Algorithms for automating "word of mouth"* », in *Proceedings of ACM Conference on Human Factors and Computing Systems*, pp. 210–217, Association of Computing Machinery, New York (1995).

- [23] Ruiz-Martin, C., Paredes, A. L., and Wainer, G. A., « *Applying complex network theory to the assessment of organizational resilience* », *IFAC-PapersOnLine*, 48(3) :1224–1229, (2015).
- [24] Ait-Ferhat, D., Friedlhuber, T., and Hibti, M., « *An andromeda extension for network based safety assessment* », In *American Nuclear Society Winter meeting on Risk Management for Complex Socio-technical Systems*, 2013.
- [25] Hibti, M., « *What if we revisit evaluation of psa models with network algorithm* », PSA, International Topical Meeting on Probabilistic Safety Assessment and Analysis, (2013).
- [26] Hibti, M., Marechal, A., and Oudjit, « *A. Exploring relations between graph metrics and importance measures in pra sequences* », International Conference on Probabilistic Safety Assessment and Management (PSAM 13), (2016).
- [27] Freeman, L.C., « *Centrality in Social Networks I: Conceptual Clarification. Social Networks* », Vol. 1 , (1979), 215–239
- [28] Brin, S. and L. Page, L., « *The Anatomy of a Large-Scale Hypertextual Web Search Engine* », In *Seventh International World-Wide Web Conference*. <http://ilpubs.stanford.edu:8090/361> (1998).
- [29] Kleinberg, J. M., « *Authorative sources in a hyperlinked environment* », J. ACM (1999).
- [30] Rasmussen, N., « *Wash-1400 (reactor safety study). An assessment of accident risks in US Commercial Nuclear Power Plants* », vol. 8, (1975).
- [31] Schoonakker, H. and Van der Borst, M., « *An overview of PSA importance measures* », *Reliability Engineering & System Safety* 72, (2001), 241–245
- [32] Kang, H. G. et al., « *An overview of risk quantification issues for digitalized nuclear power plants using a static fault tree* », *Nuclear Engineering and Technology*, 41(6), (2009).
- [33] Stuk, « *Systems design for nuclear powerplants* », guide yvl 2.0, (2002).
- [34] Rifi, M., Hibti, M. and Kanawati, R., « *Exploring network metrics for accident scenarios: a case of study of the Uncontrolled Level Drop* », PSA, International Topical Meeting on Probabilistic Safety Assessment and Analysis (2017).
- [35] Friedlhuber, T., Hibti, M. and Rauzy, A., « *Reinforcement of qualitative risk assessment-proposals from computer science* », *Proceedings of PSAM Topical, Conference in Tokyo* (2013).
- [36] Bouissou, M., Villatte, N.; Bouhadana, H. and Bannelier, M., « *Knowledge modelling and reliability processing : presentation of the figaro language and associated tools* », Technical report, Electricite de France, (1991).
- [37] Rifi, M., Hibti, M. and Kanawati, R., « *A complex network analysis approach for risk increase factor prediction in nuclear power plants* », Complexis, 3rd International Conference on Complexity, Future Information Systems and Risk (2018).
- [38] Rifi, M., Hibti, M. and Kanawati, R., « *Réseaux complexes pour la classification du facteur d'accroissement du risque dans une centrale nucléaire* », Conférence Internationale Francophone des Sciences de Données (2018).
- [39] Hastie, T., Tibshirani, R. and Friedman, J., « *The Elements of Statistical Learning* », Springer, (2001).
- [40] Rifi, M., « *Modélisation et Analyse des Réseaux Complexes : Application à la sûreté nucléaire* », Manuscrit de thèse soutenue le 03/05/2019 , Ecole doctorale Galilée, Université Paris 13.
- [41] Rifi, M., Hibti, M., Vermuse, S., Bennani, Y. and Kanawati, R., « *A complex network analysis for balanced design verification* », PSA, International Topical Meeting on Probabilistic Safety Assessment and Analysis (2019).