Rapport de Stage - Heuristiques pour la brûlure de Graphe

Lilian Berruet

Juin 2020

1 Introduction

1.1 Présentation

Je m'appel **Lilian Berruet**, je suis élève à l'**Université d'Orléans**. J'ai effectué, dans le cadre de mon diplôme de licence Informatique, un projet d'étude qui consiste en l'implémentation d'heuristiques pour des problèmes de graphes. L'un des problèmes de graphe est la notion du **Burning number**.

1.2 Définitions

Definition 1.1. *Une **heuristique** est une méthode de calcul qui fournit rapidement une solution réalisable, pas nécessairement optimale, pour un problème d'optimisation.*

Definition 1.2. *Un **graphe** est un réseau de sommet (ou noeud) relié par des arêtes. Un graphe est un couple G = (V,E) , avec V un ensemble de sommets (ou noeud) , et E un ensemble d'arêtes qui sont des paires de noeuds,* $E \subseteq \{\{x, y\} | (x, y) \in V^2 \wedge x \neq y\}$.

On travaillera dans ce rapport uniquement sur des graphes simple non-orienté, simple car il ne peut exister qu'une seule arête entre deux noeuds et non orienté car une arête (x , y) induit une arête (y , x).

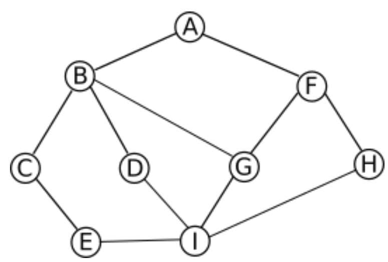

Figure 1: Exemple de graphe.

Ici l'ensemble V est {A,B,C,D,E,F,G,H,I} et E est {(A,B),(A,F),(B,C),(B,D) ,(B,G),(F,G),(F,H),(C,E),(D,I),(G,I),(H,I),(E,I)}, le graphe étant non orienté il suffit de représenter qu'un seul sens de l'arête.

Definition 1.3. *Le **Burning number** est un indice sur la vitesse de la brûlure d'un graphe.*

1.3 Précision sur le Burning number

Pour comprendre ce qu'est le **Burning number**, il faut d'abord savoir comment l'on brûle un graphe. La **brûlure** d'un graphe se déroule en plusieurs étape, prenons le graphe Figure 2 :

Étape 1 : Tous les sommets sont intacts.

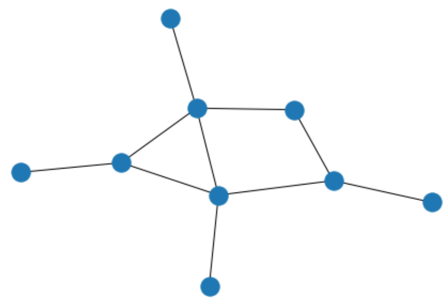

Figure 2: État initial du graphe.

Étape 2 : On choisis un sommet non-brûlé et on le brûle.

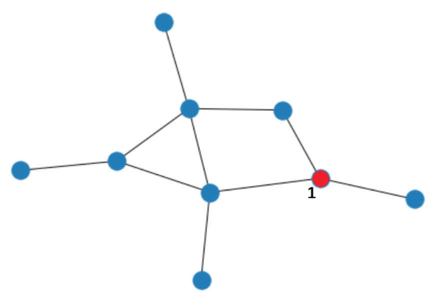

Figure 3: On choisis le sommet indexé 1 comme foyer d'un feu.

Étape 3 : Tous les sommets brûlés mettes feu à leurs voisins. Car le feu se répand sur les arètes.

Étape 4 : On retourne à l'étape 2 jusqu'à que tous les sommets soient brûlés.

Comme tous les noeuds ne sont pas brûlés à l'étape 4, on choisit un nouveau sommet indexé 2, on répète l'étape 3, puis on choisit un troisième sommet indexé 3. Tout le graphe est brûlé, on arrête l'algorithme.

La liste des sommets choisis à l'**Étape 2** est appelé la **Burning sequence** et la taille de la plus petite Burning sequence est appelé le **Burning number**. Donc le Bunring number est le nombre d'itération des étapes 2 3 et 4 minimale pour arriver à un graphe complètement brûlé.

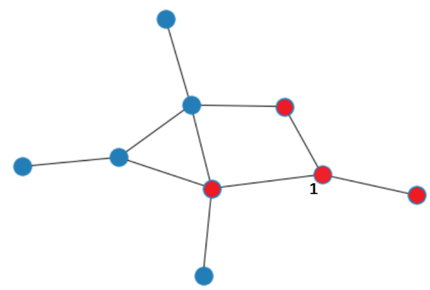

Figure 4: Le feu se répand aux voisins de 1.

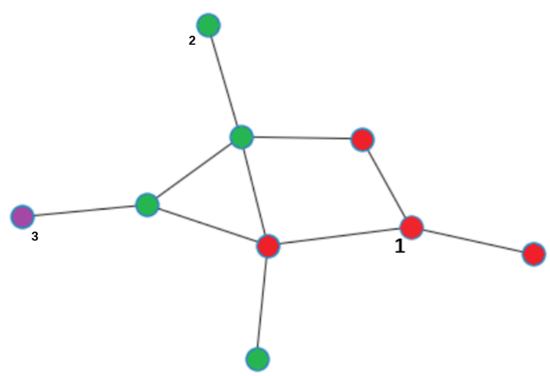

Figure 5: Graphe entièrement brûlé.

La brûlure d'un graphe schématise la **contagion** d'une information dans un **réseau**. Si l'on imagine le feu comme une information, les sommets comme des personnes et les arêtes le contact entre deux personne. Le **Burning number** est donc un indice sur la vitesse à laquelle l'information se répand.

1.4 Lien avec les heuristiques

Pour obtenir le **Burning number**, L'une des solutions serait de tester tous les **Burning sequence** possible. Mais cette solution demanderait énormément de temps dans des cas de graphe très grand. On essayeras donc de réduire ce nombre en s'approchant le plus possible de l'optimal. Ainsi l'**heuristique**, dans ce problème, sera la méthode par laquelle on choisis le prochain sommet a brûlé afin d'obtenir la **Burning sequence** la plus petite possible.

2 Heuristiques des Études

Il m'a été demandé d'observer des études traitant du **Burning number**, fournis par mes responsables de stage, et également d'essayer de reproduire leurs méthodes de choix de sommet.

2.1 Brûlure par Diamètre

Cette heuristique est proposée dans une étude du 23 Mars 2020l[lien étude]. Pour comprendre ce qu'est le **diamètre** d'un graphe, j'ai besoin d'introduire de nouvelles définitions :

Definition 2.1. *La* ***distance*** *d'un sommet avec un autre sommet et le nombre minimum d'arête qui les sépare.*

Definition 2.2. *L'**excentricité** d'un sommet est le **maximum** parmi les distances qui le sépare des autres sommets du graphe.*

Definition 2.3. *Le* ***diamètre*** *est le maximum parmi les* ***excentricités*** *du graphe.*

Definition 2.4. *Le* ***chemin le plus court*** *d'un sommet vers un autre sommet est le chemin avec la plus petite* ***distance*** *qui sépare ces deux sommets.*

Pour cette **heuristique** nous prenons le chemin le plus court du graphe de la taille du **diamètre** (le plus long des plus court chemin), et on le brûle. On peut le brûler en $\sqrt{diam(Graphe)}$ étapes, preuve dans cette étude [lien étude]. Une fois le chemin complètement brûlé, on choisit les prochains sommets **aléatoirement**.

On essaye ici de repérer la colonne vertébrale du graphe et de la brûler le plus vite possible. Seulement l'algorithme est efficace dans des cas de graphe bien particulier basé sur des cycles ou des graphe basé sur un très long diamètre. Les expériences montrent l'heuristique utile sur des graphes générées de façon artificielle qui favorise sa réussite.

2.2 Brûlure par Centralité de Vecteur Propre

La **centralité** de valeur propre (ou eigen-vector centrality) mesure de quelle manière un sommet est **connecté** aux autres sommets fortement connectés du graphe. On se base sur le fait qu'une arête avec un sommet peu connecté vaut moins qu'une arête avec un sommet très connecté.

Cette **heuristique** est proposée dans cette étude [lien étude]. On la modifiera pour qu'elle se contente de brûler les sommets avec les plus grandes **centralité**.

L'une des premières intuition que nous avons en se penchant sur le burning number est que plus un noeud est central dans le graphe plus il atteindra de sommet, donc il faudrait brûler les noeuds les plus centraux. Cette heuristique applique cette idée.

2.3 Brûlure par Distance Maximum & Distance Intermédiaire

Dans cette heuristique, proposé dans cette étude [lien de l'étude], on choisit le premier sommet dans le **centre** du graphe.

Definition 2.5. *Un sommet est dans le centre du graphe si son **excentricité** est égale au rayon du graphe.*

Definition 2.6. *Le **rayon** d'un graphe est la plus petite excentricité du graphe.*

On calcul ensuite le **time_to_burn**, le nombre d'étape que va mettre le sommet pour brûler si on n'ajoute pas de nouveau foyer. C'est donc pour un sommet non-brûlés la **distance minimum** qui le sépare d'un sommet en feu. Si on prends le plus grand time_to_burn, qu'on appellera **t_max**, il s'agit du temps que mettra l'entièreté du graphe à brûler si l'on ajoute pas de nouveau foyer.

Dans l'heuristique de distance maximum, le sommet suivant est sélectionné si son **time_to_burn** est égal à **t_max -1** (si il restes uniquement des noeuds isolés ils sont tous à t_max-1 on pourra donc brûler n'importe qu'elle noeud); dans celle de distance intermédiaire si il est égal à **t_max/2**.

Cette heuristique est également très intuitif. On commence par brûler le graphe en son centre pour y répandre un feu très rapidement, puis on cherche de nouveau foyer au endroit isolé qui mettront du temps à atteintes par le feu central.

3 Heuristiques Simples

3.1 Brûlure par Degré

Cette **heuristique** consiste a brûlé le sommet avec le plus haut degré.

Definition 3.1. *Le **degré** d'un sommet est son nombre de voisin.*

On brûle donc en priorité les sommets fortement connectés du graphe. On se rapproche fortement d'une notion de centralité. Si l'on brûle un noeud de degré 10, il brûlera 10 noeuds docn en prenant toujours le degré le plus grand on brûle le graphe rapidement.

3.2 Brûlure par Centralité d'Intermédiarité

Il existe plusieurs centralité. On s'intéresse ici à la **centralité intermédiarité**. Elle est égale au nombre de fois qu'un sommet est sur le **chemin le plus court** entre deux autres sommets quelconques du graphe. Un sommet possède une grande intermédiarité s'il a une grande influence sur les transferts de données du graphe.

Definition 3.2.

$$\sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} = center(v)$$

où σ_{st} est le nombre de plus courts chemins de s à t et $\sigma_{st}(v)$ est le nombre de ces chemins passant par v.

L'heuristique consiste donc a brûlé le sommet avec la plus grande **centralité**. De la même façon que pour la centralité de vecteur propre.

4 Notion de Communauté

Beaucoup de **systèmes** du monde réel peuvent être représenter sous la forme de **réseau**, et un réseau sous la forme d'un **graphe**. On peut représenter des échanges **sociaux**, des structures **biologiques**, des tables **lexicales**. On peut penser aussi à Google Maps basé fortement sur une structure de graphe. Dans beaucoup de ses exemples on peut y ajouter la notion de **communauté**. Une communauté se retranscrit part un groupe de sommet fortement connecté entre eux. Il serait intéressant de lier notre problème de **brûlure** de graphe avec cette notion pour se rapprocher des **graphes du réels**.

4.1 Méthode Louvain

Avant de se demander comment interagir avec les communautés il faut d'abord les détecter dans un graphe. Un des algorithmes de détection est celle de la méthode Louvain qui a l'avantage d'être très peu complexe de l'ordre *O(n log n)*, la complexité est une mesure du temps d'exécution d'un algorithme.

4.2 Méta-Graphe

Le méta-graphe est un graphe qui simplifie un autre graphe, les sommets représentent une communauté, les arêtes possèdent une valeur qu'on appel un poids qui corresponds à l'interaction entre ces deux communautés.

Sur l'image, la communauté des noeuds est indiqué par une couleur. Si l'on devait dessiner le Méta-Graphe, il ressemblerait à ceci:

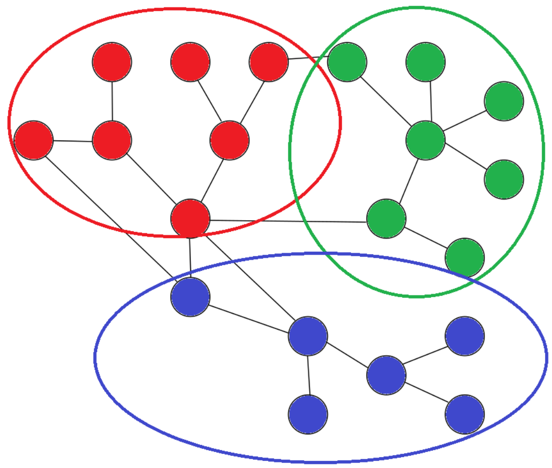

Figure 6: Exemple graphe avec ses communautés.

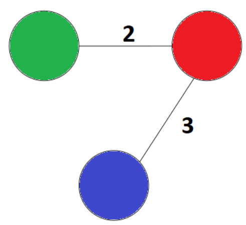

Figure 7: Exemple Méta-Grpahe de la figure 6.

Le chiffre sur les arêtes correspondent à son poids, donc aux nombres d'arêtes que partageait les communautés.

4.3 Node Predominance

La Node Prédominance sert à évaluer la connectivité d'un sommet dans sa communauté. Plus elle est grande plus le sommet est connecté dans sa communauté.

Definition 4.1.

$$NP(u) = \frac{d_{C_u}(u)}{d_{C_u}}$$

où C_u est la communauté de u, $d_{C_u}(u)$ est le degré du sommet u dans C_u, et d_{C_u} est le nombre d'arêtes dans la communauté C_u.

4.4 Page Rank

Le PageRank d'un sommet est une variante de la centralité de Vecteur Propre. On se base sur l'idée que plus les voisins d'un sommet sont important et nombreux, plus le sommet est important.

Definition 4.2.

$$PR(u) = (1-d) + d \sum_{v \in N(u)} \frac{PR(v)}{|N(v)|}$$

où $d \in \{0{,}1\}$, N(u) les voisins de u, $|N(v)|$ le nombre de voisin de v.

5 Heuristiques de Communauté

5.1 Node Predominance & Excentricité

Dans cette heuristique, on liera les notions d'excentricité et celle de node predominance.

Étape 1 : Détecter les communautés.

Étape 2 : Calculer la Node Predominance de chaque noeud du graphe pour chacune de ces communautés. On obtient un vecteur de taille nombre de communautés pour chaque noeud du graphe, on l'appel "np".

Étape 3 : Calculer le Méta-Graphe des communautés.

Étape 4 : Calculer l'excentricité de chaque noeud du Méta-Graphe. On obtient un vecteur de taille nombre de communautés, on l'appel "ecc".

Étape 5 : Pour chaque noeud du graphe, faire le produit scalaire du vecteur de Node Predominance du noeud avec l'excentricité des communautés. *Pour tous les noeuds n, prod[n] = ecc . v[node].*

Étape 6 : On brûle le noeud avec le plus grand produit scalaire, on l'appel "node".

Étape 7 : On modifie le vecteur d'excentricité. *Pour toutes les communautés i, ecc[i] = ecc[i] - ecc[i] * v[node][i].**

Étape 8 : Si le graphe n'est pas entièrement brûlé on réitère 5,6 et 7.

$*0 \leq v[node][i] \geq 1, 0 \leq ecc[i]$.

On met l'excentricité du Méta-Graphe au carré pour accentuer l'importance de la distance entre les communautés dans le calcul du produit scalaire.

En ajoutant la notion de Méta-Graphe, on cherche à se rapprocher des graphes du réel. On choisis une communauté excentrée et on brûle le noeud avec la plus grande Node Predominance de cette communauté, donc un noeud qui potentiellement brûlera très vite sa communauté. Puis on diminue l'importance de la communauté en feu dans le choix de la nouvelle communauté à brûler.

5.2 Node Predominance & PageRank

Cette heuristique est similaire à celle ci-dessus, on remplace le vecteur d'excentricité par un vecteur de PageRank.

Étape 1 : Détecter les communautés.

Étape 2 : Calculer la Node Predominance de chaque noeud du graphe pour chacune de ces communautés. On obtient un vecteur de taille nombre de communautés pour chaque noeud du graphe, on l'appel "np".

Étape 3 : Calculer le Méta-Graphe des communautés.

Étape 4 : Calculer le PageRank de chaque noeud du Méta-Graphe. On obtient un vecteur de taille nombre de communautés, on l'appel "pg".

Étape 5 : Pour chaque noeud du graphe, faire le produit scalaire du vecteur de Node Predominance du noeud avec le PageRank des communautés. *Pour tous les noeuds n, prod[n] = pg . v[node].*

Étape 6 : On brûle le noeud avec le plus grand produit scalaire, on l'appel "node".

Étape 7 : On modifie le vecteur d'excentricité. *Pour toutes les communautés i, pg[i] = pg[i] - pg[i] * v[node][i].**

Étape 8 : Si le graphe n'est pas entièrement brûlé on réitère 5,6 et 7.

$*0 \leq v[node][i] \geq 1, 0 \leq pg[i]$.

On choisis une communauté centrale et on brûle le noeud avec la plus grande Node Predominance de cette communauté, donc un noeud qui potentiellement brûlera très vite sa communauté. Puis on diminue l'importance de la communauté en feu dans le choix de la nouvelle communauté à brûler.

5.3 Connectivité

On peut se demandait si on peut pas relier ces heuristiques à la connectivité de la communauté. Les arêtes du Méta-Graphe représente la connectivité des communauté entre elles, on suppose donc que la somme des poids des arêtes d'une communauté dans le Méta-Graphe est un indice sur la communication d'une communauté avec celles qui l'entoure. On peut donc l'ajouter à nos heuristiques pour favoriser les communautés fortement connectés.

Il suffit donc de modifié le calcul des vecteurs pg et ecc, en y ajoutant la connectivité de la communauté.

Soit somme[i] la somme des arêtes de la communauté i dans le Méta-Graphe, ecc[i] = excentricité[i]^2 *+* $\sqrt{somme[i]}$ (resp. pour pg[i]).

On ajoute la racine de la somme pour diminuer sont importance dans le produit scalaire et laisser le facteur de l'excentricité ou du PageRank majoritaire.

6 Références

Zahra Rezai Farokh, Maryam Tahmasbi, Zahra Haj Rajab Ali Tehrani et Yousof Bual New Heuristics For Burning Graphs 2020.How To Brun A Graphe Heuristics for Spreading Alarm throughout a Network