Couverture par Ensembles Flous et Maximum d'Influence en Marketing Viral

MARAMI 2015

14-16 octobre 2015

Stefan Janaqi – LGI2P-EMA

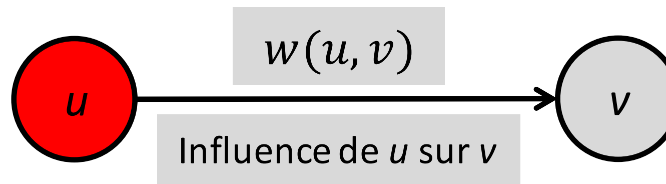
Couverture par Ensembles Flous et Maximum d'Influence

Plan

1. Définitions et problème
2. Complexité
3. Influence et couverture par ensembles flous
4. Résultats

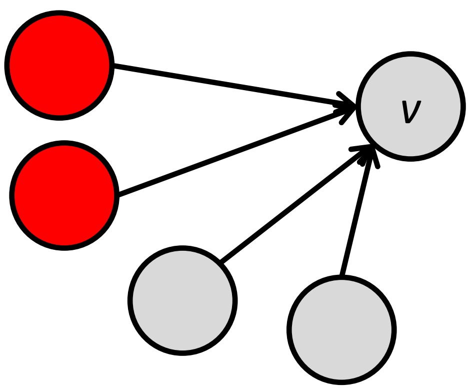
Graphe d'Influence

$$G = (V, E, w)$$
$$w: E \rightarrow R$$



- Actif

- Inactif



$t(v)$ = Seuil d'activation de v.

Modèle de Diffusion

D = Linear Threshold

$$\textbf{si} \sum_{u-actif} w(u,v) \geq t(v)$$

$$\textbf{alors} \;\; v \; devient \; actif$$

Couverture par Ensembles Flous et Maximum d'Influence

1. Définitions et problème

$Soit\ I \subseteq V$ un ensemble initial de sommets **actifs**

$p(v \mid I) \equiv p(v \mid I, D, G)$ la probabilité pour v de devenir actif à partir de I, par le modèle de diffusion D, dans le réseau G.

Variables Aléatoires $x(v), v \in V$

v	**Inactif**	**Actif**
$x(v)$	0	1
$p(x(v))$	$1 - p(v \mid I)$	$p(v \mid I)$

$X = \sum_{v \in V} x(v)$ Cardinalité d'ensemble actif final. Variable aléatoire.

$$f(I) = E(X) = \sum_{v \in V} E\big(x(v)\big) = \sum_{v \in V} p(v \mid I)$$

Problème. Trouver I maximisant $f(I)$

NP – Difficile
Heuristique Glouton

Couverture par Ensembles Flous et Maximum d'Influence

2. Complexité

EnsembleInitialViralGlouton(G, D, k)

Entrée: G **– graphe d'influence ;** D **– méthode de diffusion**
k **– cardinalité d'ensemble initial;**

Retourne: I **– ensemble viral initial**

1. $I \leftarrow \emptyset$
2. tantque $|I| < k$
3. **trouver** $u^* \leftarrow argmax_{u \in V \setminus I}(f(I \cup u))$
4. $I \leftarrow I \cup u^*$
5. fintantque

Sous modularité, Monotonie

$(i) f(\emptyset) = 0$

$(ii) A \subseteq B \subseteq V \rightarrow f(A) \leq f(B)$

$(iii) A, B \subseteq V,$
$f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$

Intérêt. [Nem 78] Si $f(I)$ est sous modulaire et monotone alors

$$\left(1 - \frac{1}{e}\right) f(Iopt) \leq f(I) \leq f(Iopt)$$

Linear Threshold [GRA 78, KEM 03] – f est sous modulaire et monotone

RUC [LAG 13a] – f n'est pas sous modulaire.

Couverture par Ensembles Flous et Maximum d'Influence

2. Complexité

LinearThresholdDiffusion (G, b, R, S, A)

$\tilde{f}(A) \leftarrow 0$
faire pour $r \leftarrow 1 \dots R$
    $t \leftarrow rand(1, n)$ // seuil aléatoire uniforme de chaque sommet
    $A_s \leftarrow A$
    faire pour $s \leftarrow 1 \dots S$
        faire pour $u \in V \setminus A_s$
            $infl(u) \leftarrow \sum_{v \in N^-(u) \cap A_s} w(v, u)$
        finpour
        faire pour $u \in V \setminus A_s$
            si $infl(u) \geq t(u)$ **alors** $A_s \leftarrow A_s \cup u$
        finpour
        $\tilde{f}(A) \leftarrow \tilde{f}(A) + |A_s|/R$
    finpour
finpour

1. Calcul de $f(I)$ difficile;
2. Approximation par Monte Carlo
3. Vitesse $1/\sqrt{R}$
4. $N_{LT} \leq R(nm + 2(n+1)^2 - 1) \sim O(Rnm)$
5. Calcul de I par l'algorithme glouton

$$H \times N_{LT}, H = \sum_{k=0}^{|I|-1} (n - k)$$

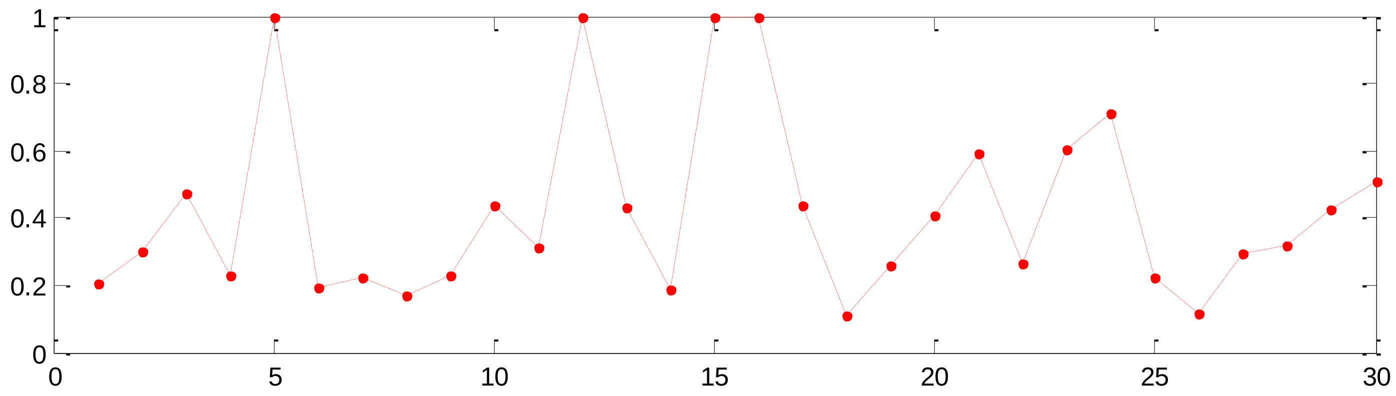
$$T_{LT} \sim O(|I|\, R\, n^2\, m)$$

$R \approx 10^4, n, m \approx 10^5 \rightarrow T_{LT} \sim O(10^{20})$!!!

Couverture par Ensembles Flous et Maximum d'Influence

3. Influence et couverture par ensembles flous

1. La recherche du premier meilleur sommet nécessite le calcul de $f(u), u \in V$;
2. Soit $\boldsymbol{P}$ la matrice $n \times n$ avec $\boldsymbol{P}(v,u) = p(v|\{u\})$;
3. Information d'ordre 1;
4. Calcul de $\boldsymbol{P}$: $n \times N_{LT}$ opérations;
5. Recherche d'un « bon » ensemble initial à partir de $\boldsymbol{P}$;
6. Chaque colonne de $\boldsymbol{P}$ peut être vue comme un degré d'appartenance d'un ensemble flou;

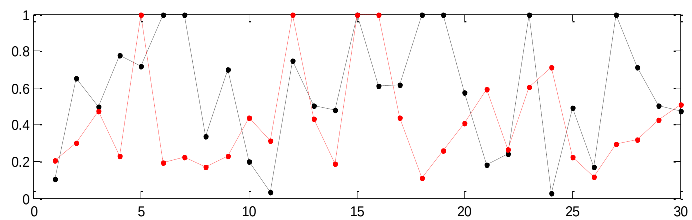


Couverture par Ensembles Flous et Maximum d'Influence

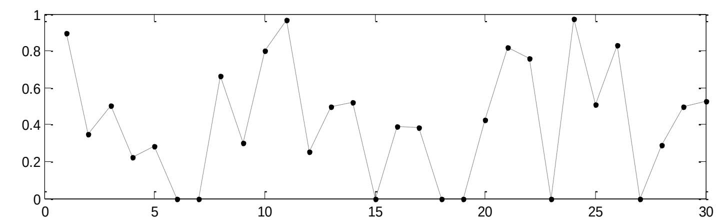
3. Influence et couverture par ensembles flous

Opérations standards avec ensembles flous $\mathbf{p},\mathbf{q}$;

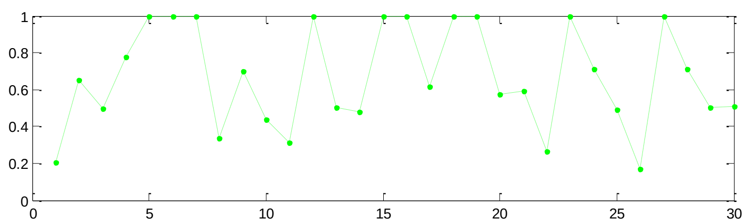
1. $NOT_{flou}(\mathbf{p}) = 1 - \mathbf{p}$; n opérations;
2. $OR_{flou}(\mathbf{p},\mathbf{q}) = \max(\mathbf{p},\mathbf{q})$; n opérations;
3. $AND_{flou}(\mathbf{p},\mathbf{q}) = \min(\mathbf{p},\mathbf{q})$; n opérations;
4. $CARD_{flou}(\mathbf{p}) = \sum_{v \in V} \mathbf{p}(v)$; n opérations;



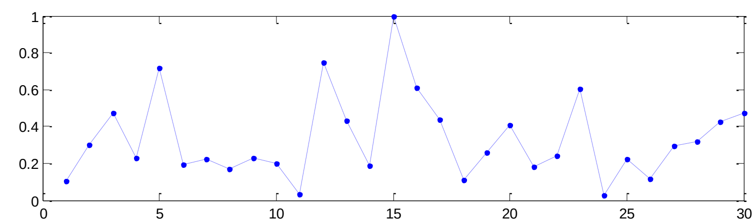
2 ensembles flous



NOT flou



OR flou



AND flou

Couverture par Ensembles Flous et Maximum d'Influence

3. Influence et couverture par ensembles flous

Intérêt des ensembles flous.

Propriété 1. Pour tout $I \subseteq V, p(v\,|I) \geq OR_{flou}\{p(v\,|\{i\}), i \in I\}$;

Propriété 2. Pour tout $I \subseteq V, f(I) \geq g(I) = \sum_{v \in V} OR_{flou}\{p(v\,|\{i\}), i \in I\}$ et cette borne est **exacte**.

Propriété 3. La fonction $g(I)$ est sous modulaire et monotone croissante, **indépendamment** de la méthode de diffusion ayant généré la matrice $\boldsymbol{P}$.

Conséquence

1. Chercher $I \subseteq V$ maximisant $g(I) \leq f(I)$;
2. NP – Difficile, cas particulier du SET COVER PROBLEM ;
3. Algorithme glouton avec garantie pour $g(I)$;
4. Faible complexité.

Couverture par Ensembles Flous et Maximum d'Influence

3. Influence et couverture par ensembles flous

GloutonCouvertureEnsembleFlou $(\boldsymbol{P}, k)$
Entrée: $\boldsymbol{P}$ – Colonnes $\boldsymbol{P}_u = p(v \mid \{u\})$; k – cardinalité d'ensemble initial;
Retourne: I – ensemble viral initial

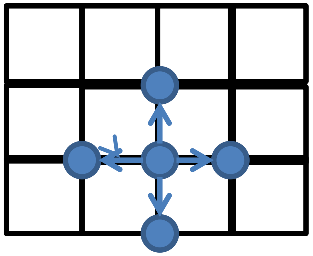
1. $I \leftarrow \emptyset$; $\mathbf{v} \leftarrow \mathbf{1}_{n\times 1}$; $\mathbf{c} \leftarrow \mathbf{0}_{n\times 1}$
2. **tantque** $|I| < k$
3. **trouver** $u^* \leftarrow argmax\left(CARD_{flou}\left(AND_{flou}(\mathbf{v}, \boldsymbol{P}_u)\right)\right), u \in V \setminus I$;
4. $I \leftarrow I \cup u^*$;
5. $\mathbf{c} \leftarrow OR_{flou}(\mathbf{c}, \boldsymbol{P}_{u*})$
6. $\mathbf{v} \leftarrow AND_{flou}\left(\mathbf{v}, NOT_{flou}(\mathbf{c})\right)$
7. **fintantque**

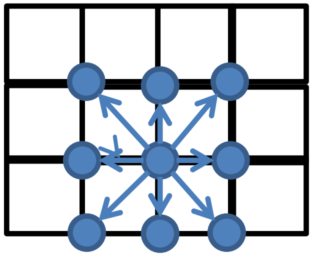
Gain en complexité

1. Complexité $O\left(|I|n^2\right)$
2. Gain $(H-n)N_{LT}$, où, $H-n=\sum_{k=1}^{|I|-1}(n-k) \sim O(|I|\ n)$ et $N_{LT} \sim O(R\ m\ n)$
3. Gain $\sim O(R\ m)$

Couverture par Ensembles Flous et Maximum d’Influence

4. Résultats, Expérimentations





Graphes Jouets

1. Graphes planaires $G = (V, E, w)$, $G\_a \times b\_C\_4\_De\_44$

Graphe de Co Auteurs

1. $\mathrm{AUTH\_82999} = (V, E, w)$, $|V| = 82999$;
2. Pour chaque auteur A la liste $L(A)$ (les indices) de ses publications est fournie (43000 publications au http://www.ncbi.nlm.nih.gov/pubmed)
3. l’influence d’un auteur A sur un auteur B

$$w(A, B) = k(A)\frac{|L(A) \cap L(B)|}{|L(B)|}$$

Couverture par Ensembles Flous et Maximum d’Influence

4. Résultats et Expérimentations

Espérance $f(I)$ pour $|I| = 30$ avec le modèle de diffusion LT. I étant choisi par une des 5 méthodes : Random; MaxDeg; MinDist; Glouton f; Glouton g.

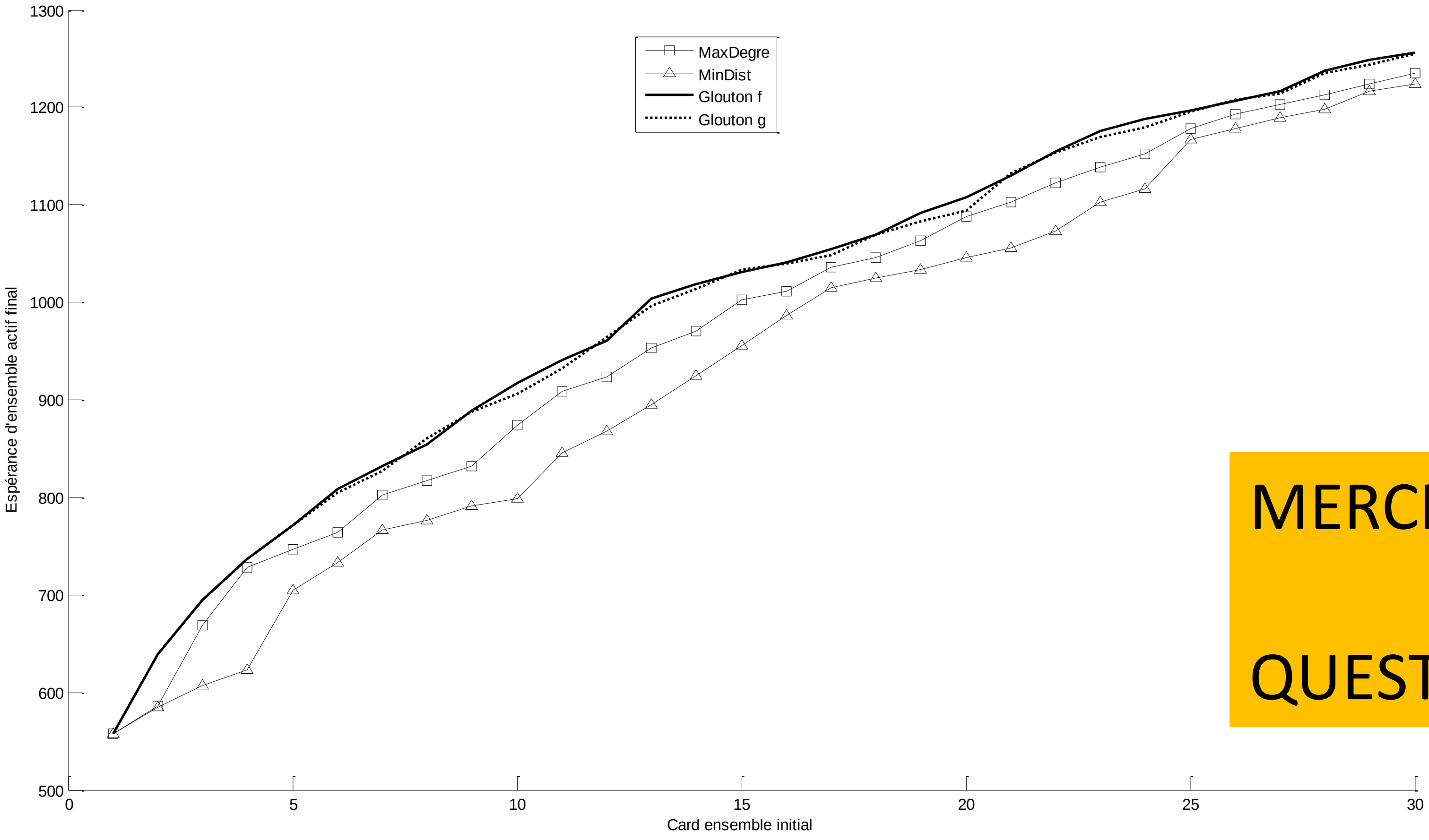
$\|I\| = 30$	Random	MaxDeg	MinDist	Glouton $f(I)$	Glouton $g(I)$
G_10x10_C_4_D_44	33	45	38	61	59
G_10x10_C_8_D_44	38	46	37	60	59
G_20x20_C_4_D_44	49	56	43	89	89
G_20x20_C_8_D_44	61	85	45	114	110
AUTH_82999	493	1235*	1224*	1256*	1255*

* - les 30 sommets de I sont sélectionnés parmi une Top_List de 500 sommets prometteurs.

Couverture par Ensembles Flous et Maximum d'Influence

4. Résultats et Expérimentations

Comparaison des espérances $f(I)$ pour $|I| = 1, \dots, 30$



MERCI !

QUESTIONS ?