Quelques utilisations des arbres en combinatoire

Yann Verhoeven

L.R.I.

Université Paris-Sud XI

Plan

- Présentation des arbres de Galton-Watson
  - Définition
  - Propriété
- Le seuil de satisfaisabilité de 2-SAT aléatoire
  - Présentation
  - Résultats généraux concernant le seuil
  - L'utilisation des arbres aléatoires
- La forte connexité dans le graphe aléatoire dirigé
  - Un résultat de seuil
  - Les arbres aléatoires et la forte connexité
- Conclusion

Arbres de Galton-Watson

Nous utilisons les arbres de Galton-Watson.

- aléatoires ;
- même distribution D à chaque génération.

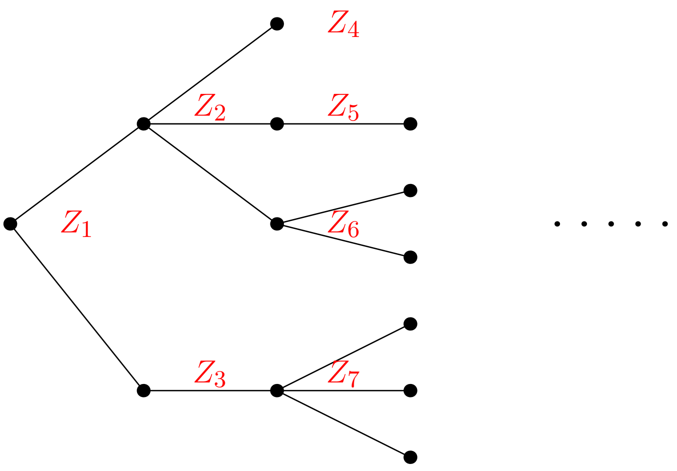
On les construit de la manière suivante (en notant Z_i le nombre d'enfants du $i^{\text{ème}}$ enfant),

- une racine a Z_1 enfants (Z_1 distribué selon D) ;
  - chacun des Z_1 enfants a un nombre aléatoire d'enfants (distribué selon D) ;
  - ...

Exemple de réalisation

On considère l'arbre :

- $Z_1 = 2$
- $Z_2 = 3$, $Z_3 = 1$
- $Z_4 = 0$, $Z_5 = 1$, $Z_6 = 2$, $Z_7 = 3$
- ...



Propriété fondamentale

Soit un arbre dont la loi de reproduction est D.

On note m l'espérance de la distribution D,

on considère la fonction génératrice f de D

$$f(s) = \sum_{k=0}^{\infty} p_k s^k,\ s \in [0,1].$$

- si $m \leq 1$, l'arbre s'éteint presque sûrement ;
- si $m > 1$, il existe un unique $q \in [0,1[$ tel que $f(q) = q$ et l'arbre survit avec probabilité $1 - q$.

2-SAT aléatoire

On considère des formules booléennes.

- Variables boolénnes ;
- opérateurs booléens : négation ($\neg$), et ($\wedge$) et ou ($\vee$).

Les formules 2-CNF : conjonctions de clauses de taille 2.

Un exemple :

on prend les variables u_1, u_2 et u_3,

$$\phi = (u_1 \vee u_2) \wedge (u_1 \vee \neg u_2) \wedge (u_2 \vee u_3) \wedge (\neg u_1 \vee \neg u_3).$$

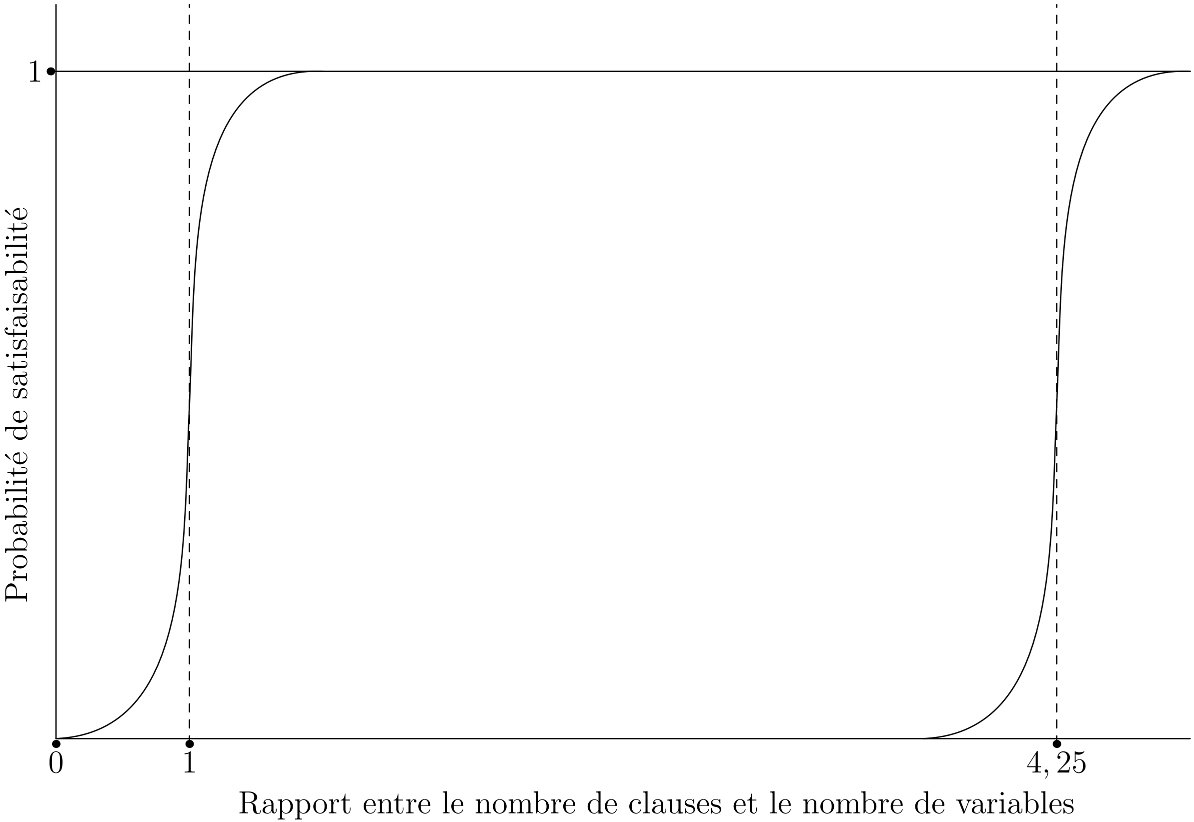
Le problème 2-SAT

entrée : une formule sous forme 2-CNF
sortie : existe-t-il une valuation satisfaisant cette formule ?

Pourquoi étudier ce problème qui est dans P ?

- Appartient à la famille de problèmes k-SAT qui est importante en informatique et dont la version probabiliste présente un phénomène de seuil pour $k \geq 2$;
- k-SAT apparaît en physique statistique.

Le phénomène de seuil



Les formules aléatoires

Deux façons de tirer des formules aléatoirement, pour n variables,

- on considère toutes les clauses possibles et chacune est choisie avec probabilité p. Ce modèle est noté $\mathcal{F}_{n,p}$;
- on choisit de manière uniforme une formule parmi celles contenant m clauses. Ce modèle est noté $\mathcal{F}_{n,m}$.

Nous exprimons nos résultats dans le modèle $\mathcal{F}_{n,p}$, la traduction dans le modèle $\mathcal{F}_{n,m}$ est facile.

Le seuil de transition

Théorème [Goerdt 92, Chvátal et Reed 92]

Soit $F = \mathcal{F}_{n,m}$ *avec* $m = m(n)$.
Alors, on a

i. *si* $\frac{m}{n} \leq c$, *pour* $c < 1$ *fixé, alors*

$$\lim_{n \to \infty} Pr[F \textit{ est satisfaisable}] = 1 \; ;$$

ii. *si* $\frac{m}{n} \geq c$, *pour* $c > 1$ *fixé, alors*

$$\lim_{n \to \infty} Pr[F \textit{ est insatisfaisable}] = 1.$$

Au dela du seuil

Les fonctions $\rho_S = \rho_S(n)$ et $\rho_I = \rho_I(n)$ sont dites admissibles si l'on a

$$\frac{m}{n} \le \rho_S \Rightarrow \Pr[F \text{ est satisfaisable}] = 1 - o(1),$$

$$\frac{m}{n} \ge \rho_I \Rightarrow \Pr[F \text{ est satisfaisable}] = o(1).$$

Alors, sont admissibles,

$\rho_S = 1 - \varepsilon$	$\rho_I = 1 + \varepsilon$	[Goerdt 92, Chvátal et Reed 92]
$\rho_S = 1 - 1/\sqrt{\ln n}$		[Goerdt 99]
	$\rho_I = 1 + \omega(n)n^{-1/4}$	[Verhoeven 99]
$\rho_S = 1 - \omega(n)n^{-1/8}$		[de la Vega 00]
$\rho_S = 1 - \omega(n)n^{-1/3}$	$\rho_I = 1 + \omega(n)n^{-1/3}$	[Bollobás, Borgs, Chayes, Kim, Wilson 01]
$\rho_S = 1 - \omega(n)n^{-1/3}$		[de la Vega et Verhoeven 01]

Le graphe associé

On transforme une formule 2-CNF F en un graphe dirigé $G(F)$ [Aspvall, Plass, Tarjan].

Pour chacune des n variables u_i on introduit dans le graphe les sommets u_i et $\neg u_i$.

Pour chaque clause $u \vee v$ de F on introduit les arcs :

$$\neg u \rightarrow v \text{ et } \neg v \rightarrow u.$$

$(\neg\neg v = v)$

Théorème [Aspvall, Plass et Tarjan 79]

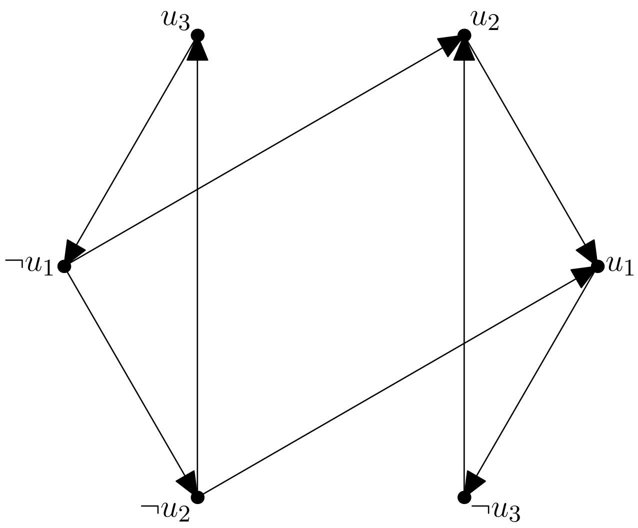
Une formule 2-CNF F est insatisfaisable si et seulement s'il existe dans $G(F)$ un circuit contenant un littéral et son complément.

Un exemple :

On considère à nouveau

$$\phi = (u_1 \vee u_2) \wedge (u_1 \vee \neg u_2) \wedge (u_2 \vee u_3) \wedge (\neg u_1 \vee \neg u_3).$$

Le graphe qui lui est associé est



Une borne supérieure du seuil

Théorème [Verhoeven 99]

Soit $F = \mathcal{F}_{n,p}$ *avec* $p \geq (1 + n^{-1/4})/2n$. *Alors*

$$\lim_{n \to \infty} Pr[F_{n,p} \text{ est insatisfaisable }] = 1.$$

Schéma de démonstration

On prend $p \geq (1 + n^{-1/4})/2n$.

Pour montrer l'insatisfaisabilité d'une formule, on cherche s'il existe une variable u tel qu'on ait dans le graphe associé :

- un chemin de u vers $\neg u$;
- un chemin de $\neg u$ vers u.

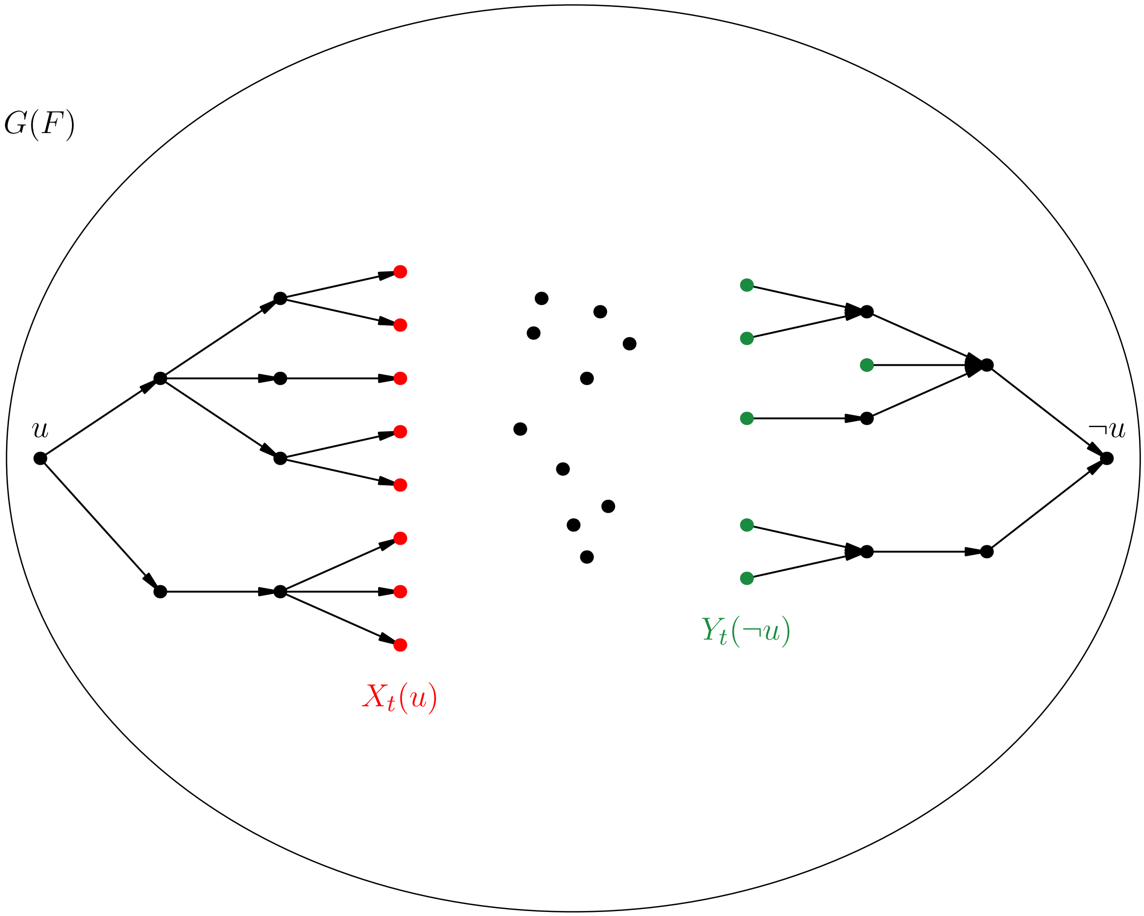
Un algorithme qui trouve de tels chemins avec probabilité $1-o(1)$ permet de conclure en vertu du théorème de [Aspvall,Plass et Tarjan].

L'algorithme

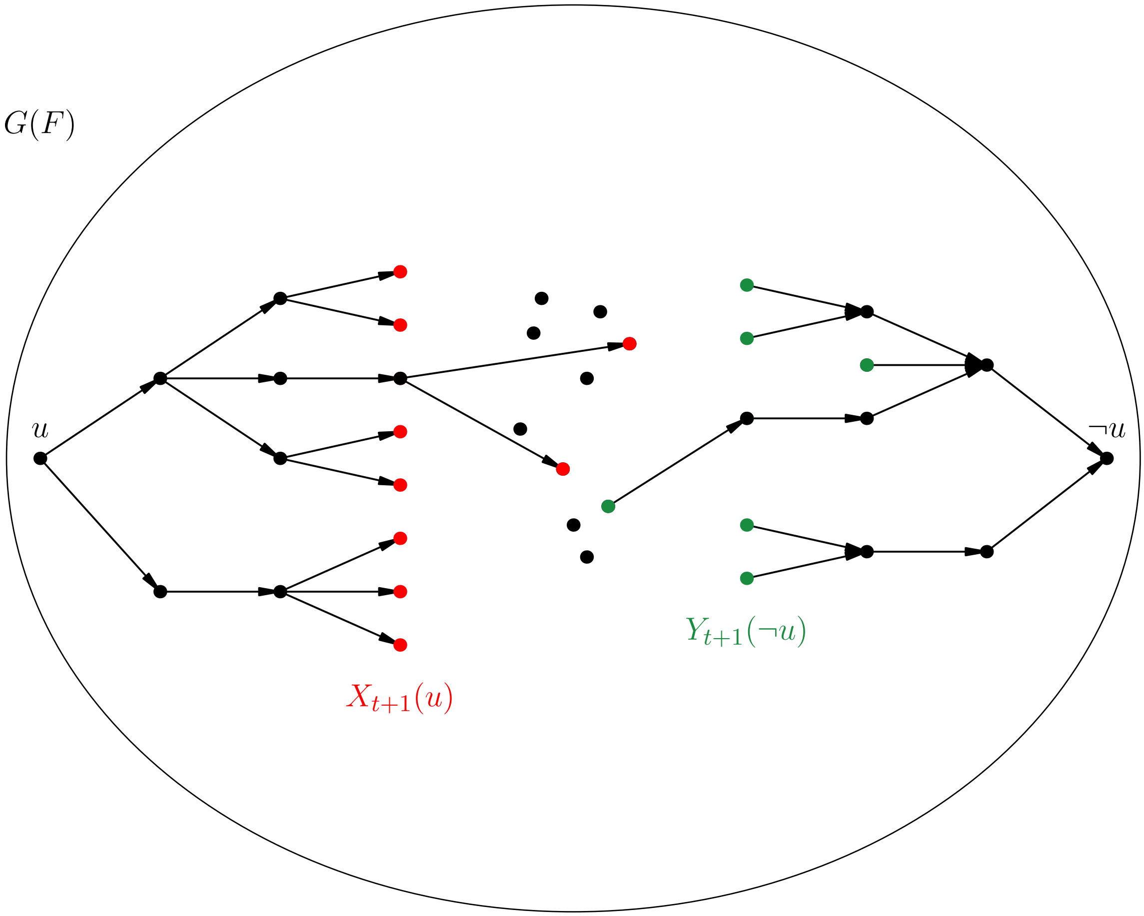
Schéma de l'algorithme :

- soit u un sommet de $G(F)$;
- on construit pas à pas l'arbre des sommets atteignables à partir de u (dont la frontière est notée $X_t(u)$) et l'arbre des sommets qui peuvent atteindre $\neg u$ (dont la frontière est notée $Y_t(\neg u)$),
  - si l'un des deux arbres meurt on choisit un sommet v parmi ceux non encore touchés et on recommence ;
  - dès que les frontières $X_t(u)$ et $Y_t(\neg u)$ atteignent la taille $\omega(n)\sqrt{2n}$ pour un certain t l'algorithme s'arrête avec succès car deux ensembles de taille $\omega(n)\sqrt{2n}$ sont reliés par un arc dans les deux sens avec probabilité $1 - o(1)$.

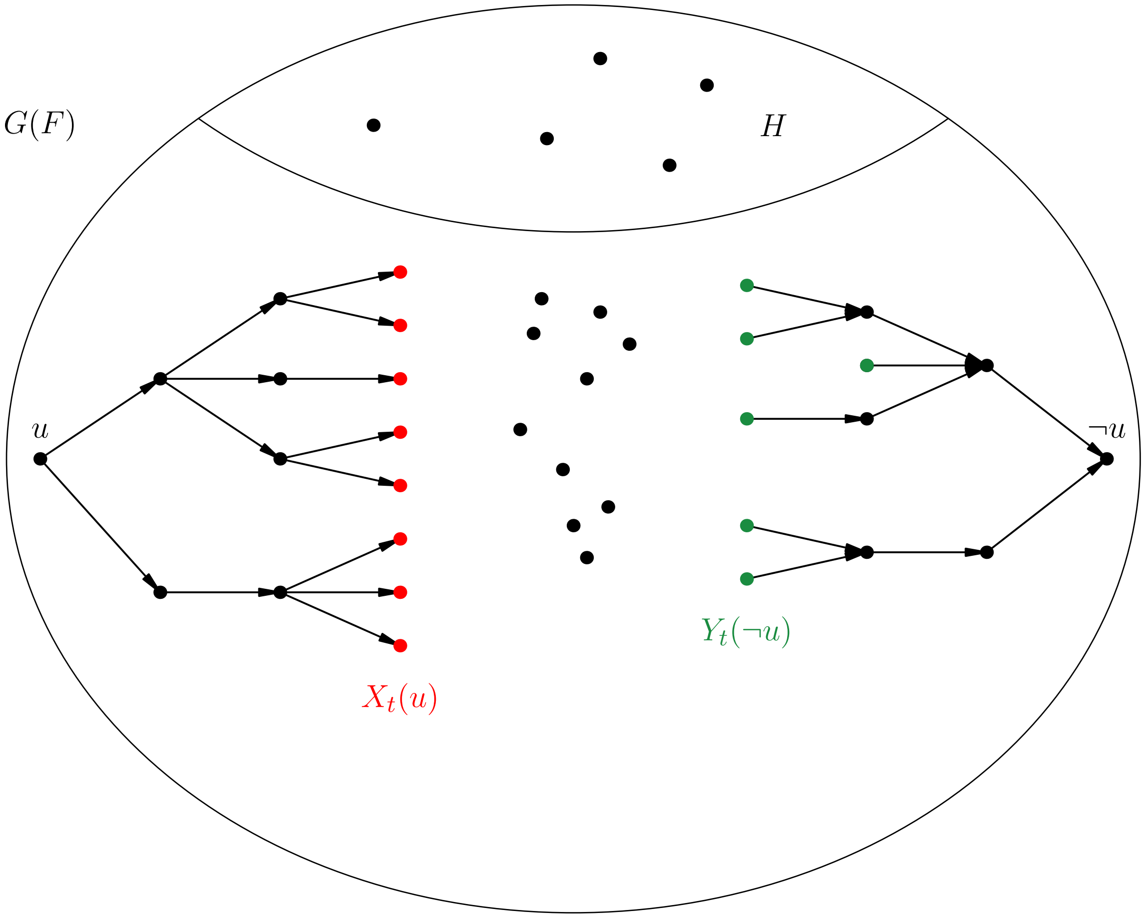
Illustration de l'algorithme



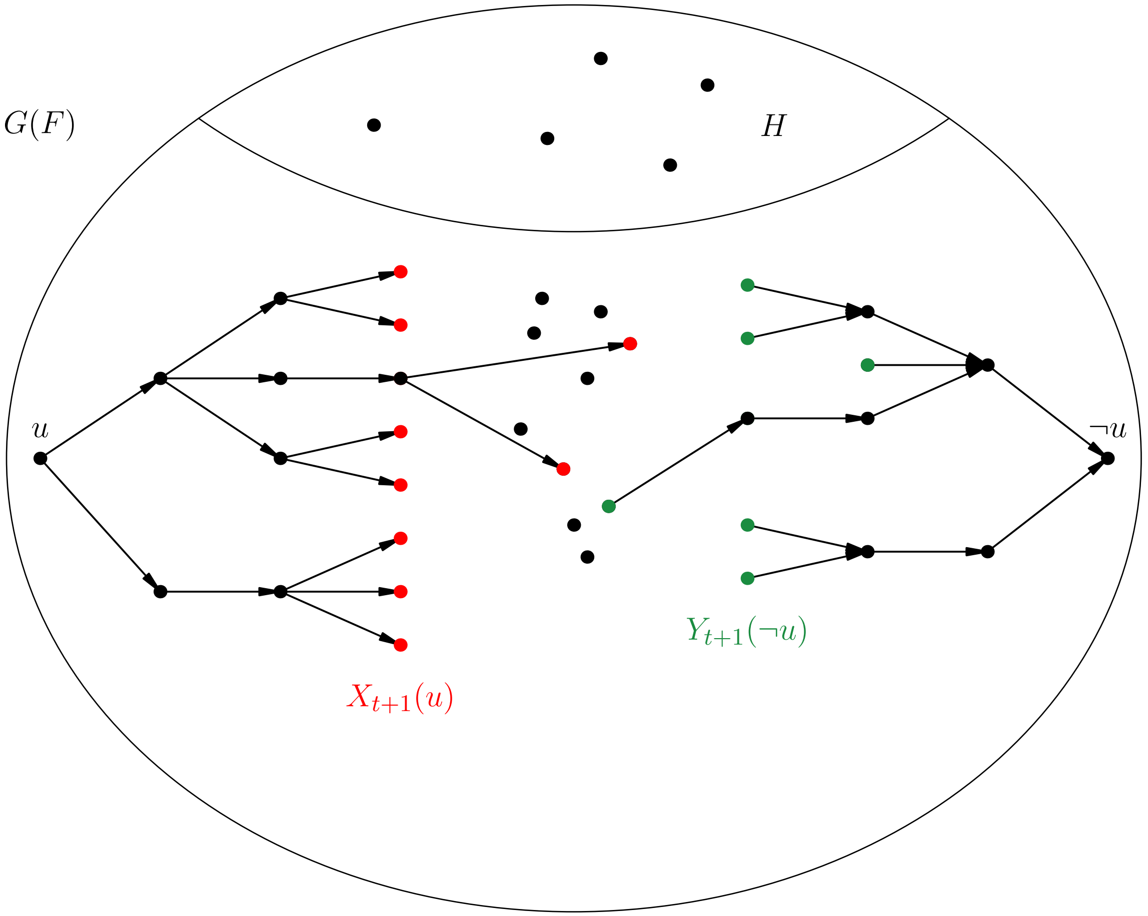
Développement d'un arbre



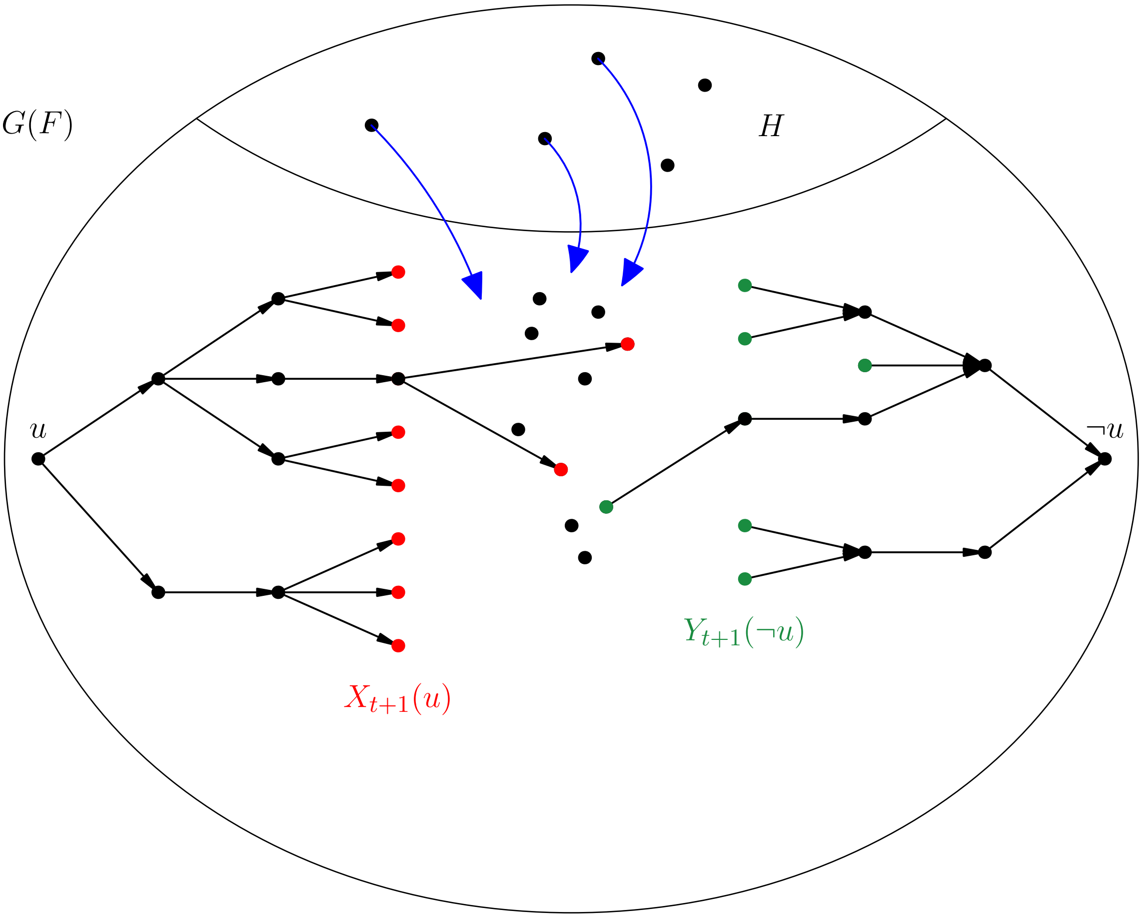
Utilisation des arbres de Galton-Watson



Utilisation des arbres de Galton-Watson (2)



Utilisation des arbres de Galton-Watson (3)



Lemmes intermédiaires, la taille de l'arbre

On suppose $X(u) = X_\infty(u)$.
On pose $p \geq (1 + n^{-1/4})/2n$.

Lemme

Soit $F = \mathcal{F}_{n,p}$ *et* u *un littéral de* F. *Alors*

$$Pr\left[|X(u)| \geq \omega(n)\sqrt{2n}\right] \geq 12n^{-1/4}.$$

Et par ailleurs,

$$E[|X(u)|||X(u)| < \omega(n)\sqrt{2n}] \leq 2n^{1/4}.$$

Ce lemme est obtenu à l'aide d'arbres de Galton-Watson.

Lemmes intermédiaires, l'algorithme réussit

- On développe deux arbres de façon indépendante ;
- chacun est suffisamment grand avec une probabilité supérieure à $12n^{-1/4}$.

Lemme

L'algorithme présenté permet de trouver deux sous-ensembles de $G(F)$ *dont les tailles des frontières sont au moins* $\omega(n)\sqrt{2n}$ *après au plus* $\omega(n)n^{1/2}/3$ *essais avec probabilité* $1 - o(1)$.

Lemmes intermédiaires, nombre de sommets utilisés

- L'algorithme comporte au plus $\omega(n)n^{1/2}/3$ étapes ;
- chaque étape qui échoue “consomme” $2n^{1/4}$ sommets.

Lemme

L'algorithme présenté n'utilise pas plus de $\omega(n)n^{3/4}$ sommets, avec probabilité $1 - o(1)$, avant de trouver deux sous-ensembles de $G(F)$ dont la frontière est de taille $\omega(n)\sqrt{2n}$.

Il suffit donc de prévoir initialement $H = \omega(n)n^{3/4}$ sommets de réserve.

Connexité forte

On considère le graphe aléatoire dirigé $D_{n,p}$ à n sommets où chaque arc est présent avec la probabilité p.

Deux sommets x et y appartiennent à la même composante fortement connexe si et seulement si

- il existe un chemin de x à y ;
- et il existe un chemin de y à x.

Motivations de ce problème,

- il est relié au problème 2-SAT ;
- il est lié à un phénomène de seuil ;
- peut être un outil de l'étude du World Wide Web.

Le seuil et l'existence de la composante géante

Comme dans le cas non orienté [Erdös-Rényi], on assiste à un phénomène de seuil pour la taille de la composante fortement connexe dans $D_{n,p}$ [Karp].

Le phénomène est le suivant,

- si $p = (1 - \varepsilon)/n$ pour $\varepsilon > 0$ fixé, les composantes connexes sont "petites" ;
- si $p = (1 + \varepsilon)/n$ pour ε fixé, il apparaît une composante fortement connexe "géante".

La composante géante

Théorème [Karp 91]

Soit $\omega(n)$ *une fonction croissante non bornée,* c *une constante strictement supérieure à* 1 *et* Θ *la racine de* $1 - x - e^{-cx} = 0$ *dans* $]0, 1[$.
Il existe une constante M *telle que, avec probabilité* $1 - o(1)$,

- *un graphe aléatoire orienté* $D_{n,c/n}$ *a exactement une composante fortement connexe contenant plus de* $M \log n$ *sommets ;*

- *la taille de cette composante est comprise dans l'intervalle*

$$[\Theta^2 n - \omega(n)\sqrt{n \log n}, \Theta^2 n + \omega(n)\sqrt{n \log n}].$$

Une taille plus précise

Théorème [Verhoeven 01]

Soit $\omega(n)$ *une fonction croissante non bornée,* c *une constante strictement supérieure à* 1 *et* Θ *la racine de* $1 - x - e^{-cx} = 0$ *dans* $]0, 1[$.
Alors,

- *la taille de la composante géante de* $D_{n,c/n}$ *est comprise dans l'intervalle*
$$[\Theta^2 n - \omega(n)\sqrt{n}, \Theta^2 n + \omega(n)\sqrt{n}].$$

Ce résultat est obtenu à l'aide d'arbres de Galton-Watson.

Conclusion

Utilisation des arbres de Galton-Watson dans l'analyse d'algorithmes de recherche :

- recherche en largeur [Verhoeven]
  - une borne supérieure du seuil de 2-SAT aléatoire ;
  - une estimation des fluctuations de la taille de la composante fortement connexe géante.
- recherche en profondeur [Ajtai, Komlós et Szemerédi] ;