Propriétés universelles des algorithmes de recherche combinatoire probabilistes, 2ème partie

Rémi Monasson
LPTENS Paris

et

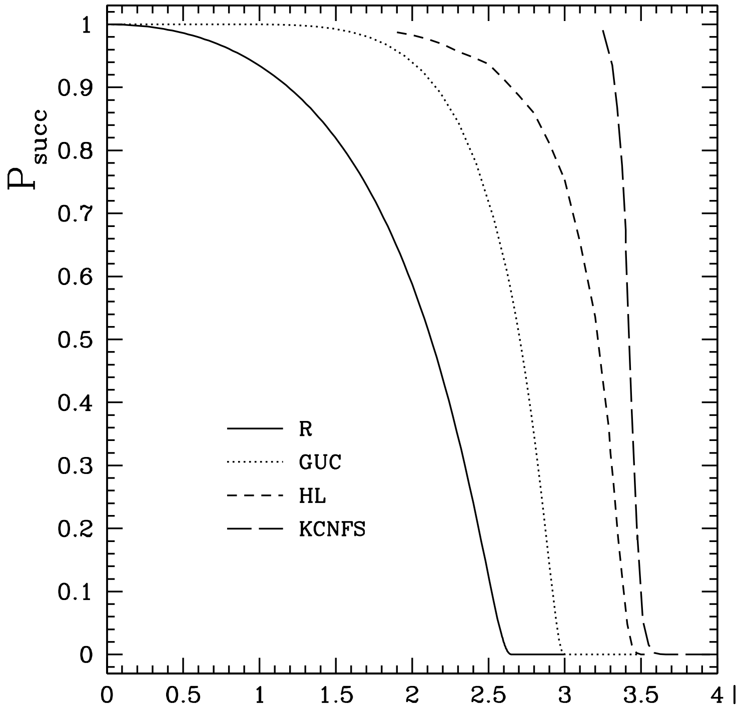
Christophe Deroulers
LPTENS Paris

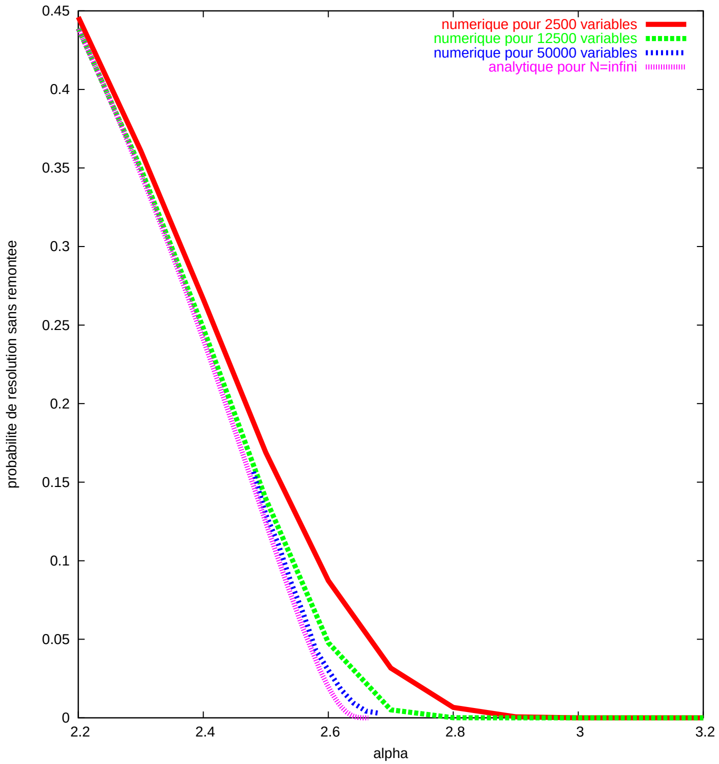
INRIA Rocquencourt, 8 novembre 2004

But de cette partie

Étude des algorithmes gloutons :

- Calcul de $P(\alpha, N = +\infty)$ pour l'algorithme `random` ;
- Calcul de $P(\alpha, N \gg 1)$ pour la classe «P.U.» au voisinage de $\alpha_{\mathrm{poly/exp}}$.





Réduction du problème

Notations

N variables, αN clauses.

$0 \leq T \leq N$: nombre de variables éliminées depuis le début de la résolution.

Un état des αN clauses lors de la résolution, après élimination de T variables sur N :

		VRAI							i	g	OU	$\overline{h}$
d	OU	$\overline{g}$	OU	j	g	OU	h	OU	$\overline{i}$	VRAI		
		VRAI					VRAI			VRAI		

Il y a C_1 1-clauses, C_2 2-clauses, C_3 3-clauses.

Une étape de l'algorithme

$\boxed{C_1 > 0}$ (clauses unitaires présentes) $\Rightarrow$ Propagation Unitaire

VRAI	i	g OU $\overline{h}$
d OU $\overline{g}$ OU j	g OU h OU $\overline{i}$	VRAI
VRAI	VRAI	VRAI

$\boxed{\text{OU}}$

$\boxed{C_1 = 0}$ $\Rightarrow$ heuristique qui dépend de l'algorithme

VRAI	e OU i	g OU $\overline{h}$
d OU $\overline{g}$ OU j	g OU h OU $\overline{i}$	$\overline{e}$ OU i OU j
VRAI	VRAI	$\overline{e}$ OU f

Première réduction

Avec cette distribution des instances,

Quand on a moyenné sur les entrées de l'algorithme, la distribution de formules à 1-, 2- et 3-clauses à un instant T donné est uniforme, conditionnée à la donnée de $\vec{C}(T) := (C_3(T), C_2(T), C_1(T))$.

Franco

Matrice d'évolution

On peut écrire de façon exacte les taux de transition de la «chaîne de Markov» $K(\vec{C} \leftarrow \vec{C}', T)$. Exemple de `random` :

$$\binom{C_3'}{C_3'-C_3}\left(1-\frac{3}{N-T}\right)^{C_3}\left(\frac{3}{2(N-T)}\right)^{C_3'-C_3}\sum_{w_2=0}^{C_3'-C_3}\binom{C_3'-C_3}{w_2}\times$$

$$\sum_{z_2=0}^{C_2'}\binom{C_2'}{z_2}\left(1-\frac{2}{N-T}\right)^{C_2'-z_2}\left(\frac{1}{N-T}\right)^{z_2}\sum_{w_1=0}^{z_2}\binom{z_2}{w_1}\delta_{C_2=C_2'+w_2-z_2}\times$$

$$\Bigg\{(1-\delta_{C_1'=0})\sum_{e_1=0}^{C_1'-1}\binom{C_1'-1}{e1}\left(\frac{1}{2(N-T)}\right)^{e_1}\left(1-\frac{1}{N-T}\right)^{C_1'-1-e_1}\times$$

$$\delta_{C_1=C_1'+w_1-e_1-1} \quad + \quad \delta_{C_1'=0}\delta_{C_1=w_1}\Bigg\}$$

Deuxième réduction

Il y a concentration de la mesure pour C_2 et C_3.

Donc on calcule juste les moyennes $\langle C_2(T) \rangle$ et $\langle C_3(T) \rangle$ plutôt que toutes les distributions.

Chao & Franco

Pour `random` :

$$
\begin{aligned}
\langle C_3(T+1) \rangle - \langle C_3(T) \rangle &= -\frac{3\langle C_3 \rangle}{N-T} \\
\langle C_2(T+1) \rangle - \langle C_2(T) \rangle &= -\frac{2\langle C_2 \rangle}{N-T} + \frac{1}{2}\frac{3\langle C_3 \rangle}{N-T}
\end{aligned}
$$

Deuxième réduction

Et pour tous les algorithmes avec Propagation Unitaire,

$$\begin{aligned}\langle C_3(T+1)\rangle - \langle C_3(T)\rangle &= f_1(\langle C_3\rangle, \langle C_2\rangle, N-T)\\ \langle C_2(T+1)\rangle - \langle C_2(T)\rangle &= f_2(\langle C_3\rangle, \langle C_2\rangle, N-T)\end{aligned}$$

car

- On le vérifie lors d'un pas de Propagation Unitaire ($C_1 > 0$),
- C'est forcément vrai si $C_1 = 0$,
- Donc c'est vrai pour la somme des deux contributions.

$\rightsquigarrow$ On peut calculer $\langle C_2(T)\rangle$ et $\langle C_3(T)\rangle$.

Temps discret $\rightarrow$ temps continu

Chao & Franco

$C_2(T), C_3(T)$ varient sur des temps T d'ordre 1, mais $\frac{C_j}{N}$ varient sur des temps T d'ordre N :
$\exists c_2(t)$ et $c_3(t)$ fonctions régulières telles que $C_j(T) = N c_j(\frac{T}{N}) + O(\sqrt{N})$

Alors (pour `random`)

$$\begin{aligned} \frac{dc_3}{dt} &= -\frac{3c_3}{1-t} \\ \frac{dc_2}{dt} &= -\frac{2c_2}{1-t} + \frac{1}{2}\frac{3c_3}{1-t} \end{aligned}$$

et $c_3(t) = \alpha(0)(1-t)^3$, $c_2(t) = \frac{3}{2}\alpha(0)t(1-t)^2$ pour $0 \leq t \leq 1$...

Chao & Franco

$C_2(T), C_3(T)$ varient sur des temps T d'ordre 1, mais $\frac{C_j}{N}$ varient sur des temps T d'ordre N :
$\exists c_2(t)$ et $c_3(t)$ fonctions régulières telles que $C_j(T) = Nc_j(\frac{T}{N}) + O(\sqrt{N})$

Alors (pour `random`)

$$\begin{aligned} \frac{dc_3}{dt} &= -\frac{3c_3}{1-t} \\ \frac{dc_2}{dt} &= -\frac{2c_2}{1-t} + \frac{1}{2}\frac{3c_3}{1-t} \end{aligned}$$

et $c_3(t) = \alpha(0)(1-t)^3$, $c_2(t) = \frac{3}{2}\alpha(0)t(1-t)^2$ pour $0 \leq t \leq 1$...

...tant qu'il n'y a pas de contradiction entre deux clauses unitaires.

Contradictions ?

Proba de ne pas remarquer de contradiction de T à $T+1$:

$$\left(1 - \frac{1}{2(N-T)}\right)^{\max(C_1 - 1, 0)}$$

Contradictions ?

Proba de ne pas remarquer de contradiction de T à $T+1$:

$$\left(1-\frac{1}{2(N-T)}\right)^{\max(C_1-1,0)}$$

d'où la proba de ne pas rencontrer de contradiction de 0 à $T = Nt$:

$$\exp\left(-\int_0^t dt\frac{\langle\max(C_1-1,0)\rangle}{2(1-t)}\right)$$

finie quand $N\to+\infty$ si C_1 est borné par $O(1)$.

$\longrightarrow$ il faut étudier la distribution de $C_1(T)$.

Étude de la distribution de C_1

Matrice de transition pour C_1

$$H_N[C_1' \leftarrow C_1; T, C_2] = \sum_{s_2, r_2=0}^{C_2} \binom{C_2}{s_2, r_2} \left(\frac{1}{N-T}\right)^{s_2+r_2} \left(1 - \frac{2}{N-T}\right)^{C_2-s_2-r_2} \times$$

$$\Bigg[\delta_{C_1=0}\delta_{C_1'=r_2} + (1-\delta_{C_1=0})\times$$

$$\sum_{s_1=0}^{C_1-1} \binom{C_1-1}{s_1} \left(\frac{1}{2(N-T)}\right)^{s_1} \left(1 - \frac{1}{N-T}\right)^{C_1-1-s_1} \delta_{C_1'=C_1-1-s_1+r_2}\Bigg]$$

Matrice de transition pour C_1

$$H_N[C_1' \leftarrow C_1; T, C_2] = \sum_{s_2, r_2 = 0}^{C_2} \binom{C_2}{s_2, r_2} \left(\frac{1}{N-T}\right)^{s_2+r_2} \left(1 - \frac{2}{N-T}\right)^{C_2 - s_2 - r_2} \times$$

$$\Bigg[\delta_{C_1=0}\delta_{C_1'=r_2} + (1-\delta_{C_1=0})\times$$

$$\sum_{s_1=0}^{C_1-1} \binom{C_1-1}{s_1} \left(\frac{1}{2(N-T)}\right)^{s_1} \left(1 - \frac{1}{N-T}\right)^{C_1-1-s_1} \delta_{C_1'=C_1-1-s_1+r_2}\Bigg]$$

- H ne dépend (explicitement) que de C_1, C_2 et T.
- Équation exacte. Contient toute l'information (sauf C_2).
- Expression valable uniquement pour la distribution «random K-SAT»
- mais pour tous les algorithmes qui appliquent la Propagation Unitaire.

C_1 reste fini si...

$P_N(C_1, T)$: proba que, au temps T, on n'ait pas remarqué de contradiction et qu'il y ait C_1 1-clauses.

C_1 reste fini si...

$P_N(C_1, T)$: proba que, au temps T, on n'ait pas remarqué de contradiction et qu'il y ait C_1 1-clauses.

Hypothèse : $P_N(C_1, T) = p_0(C_1, t) + \frac{1}{N} p_1(C_1, t) + \mathcal{O}(\frac{1}{N^2})$.

...On reporte dans la matrice de transition de C_1...

C_1 reste fini si...

$P_N(C_1, T)$: proba que, au temps T, on n'ait pas remarqué de contradiction et qu'il y ait C_1 1-clauses.

Hypothèse : $P_N(C_1, T) = p_0(C_1, t) + \frac{1}{N} p_1(C_1, t) + \mathcal{O}(\frac{1}{N^2})$.

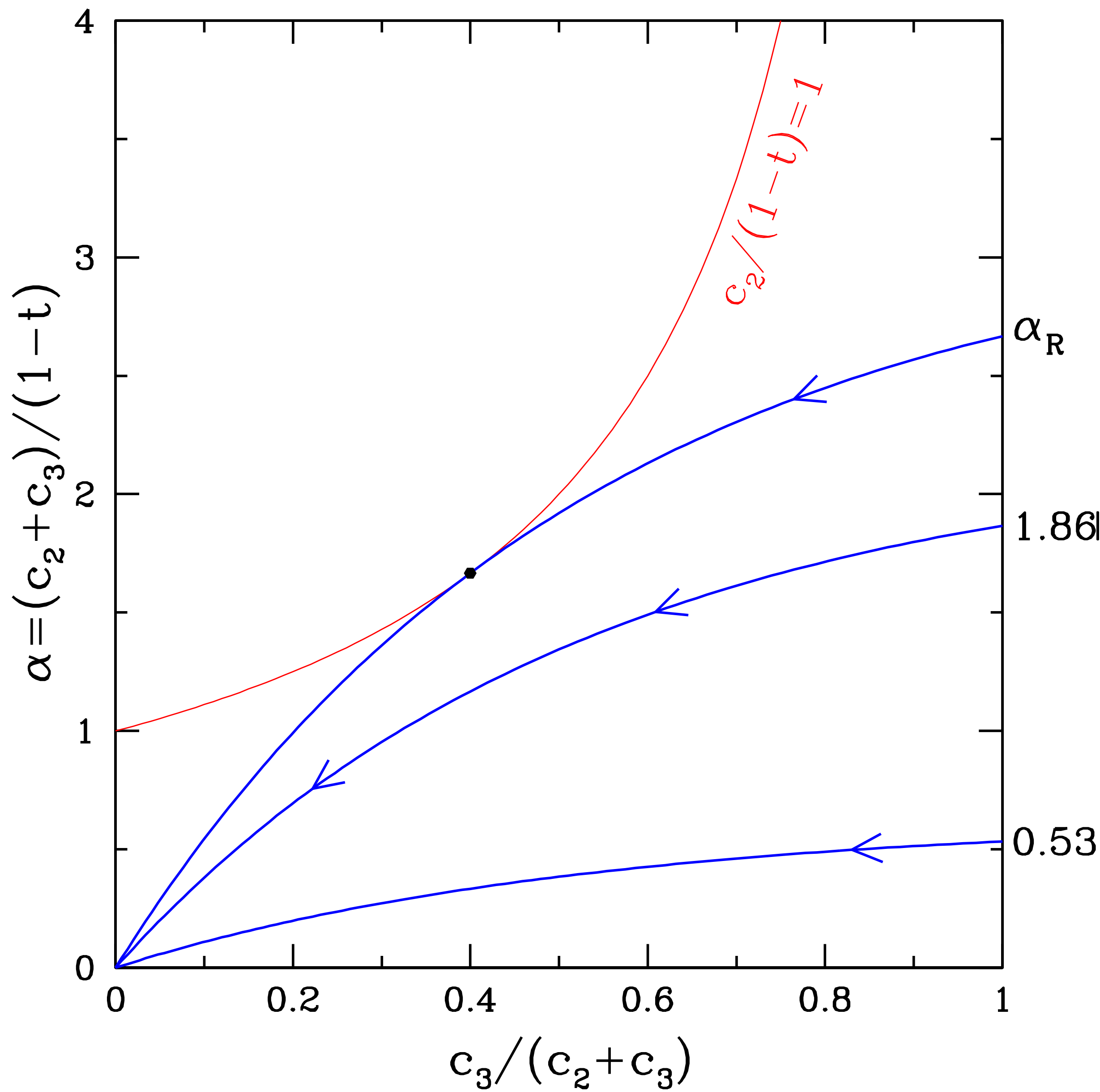
...On reporte dans la matrice de transition de C_1...

$\rightarrow$ pour C_1 grand, $p_0(C_1, t) \asymp e^{-\rho C_1}$ où

$$\rho \frac{1-t}{c_2} = 1 - e^{-\rho}$$

Tant que $\frac{c_2}{1-t} < 1$, $\rho > 0$ et C_1 est borné p.s. quand $N \rightarrow +\infty$.

Trajectoires de résolution



Probabilité de succès finie

$\rightarrow$ Tant que $\alpha < \alpha_{\texttt{random}} = \frac{8}{3}$, $C_1(T)$ est borné p.s. et on connaît sa distribution en fonction de $c_2(t)$. (Si $\alpha \geq \alpha_{\texttt{random}}$, C_1 n'est plus toujours borné.)

Probabilité de succès finie

$\longrightarrow$ Tant que $\alpha < \alpha_{\texttt{random}} = \frac{8}{3}$, $C_1(T)$ est borné p.s. et on connaît sa distribution en fonction de $c_2(t)$. (Si $\alpha \geq \alpha_{\texttt{random}}$, C_1 n'est plus toujours borné.)

$$P_{\text{succ}} = \exp\left(- \int_0^1 dt \frac{\langle \max(C_1 - 1, 0) \rangle}{2(1-t)} \right)$$

d'où pour `random`, $\alpha < \frac{8}{3}$ et $N = +\infty$:

$$\boxed{\ln(P_{\text{succ}}) = \frac{3\alpha}{16} - \frac{1}{2\sqrt{\frac{8}{3\alpha} - 1}} \arctan\left(\frac{1}{\sqrt{\frac{8}{3\alpha} - 1}} \right)}.$$

On peut faire le calcul pour d'autres algorithmes. Universalité ?

Le régime critique

Préliminaire

Fonction génératrice de C_1 :

$$p_N(T,x) := \sum_{C_1=0}^{+\infty} x^{C_1} P_N(C_1,T)$$

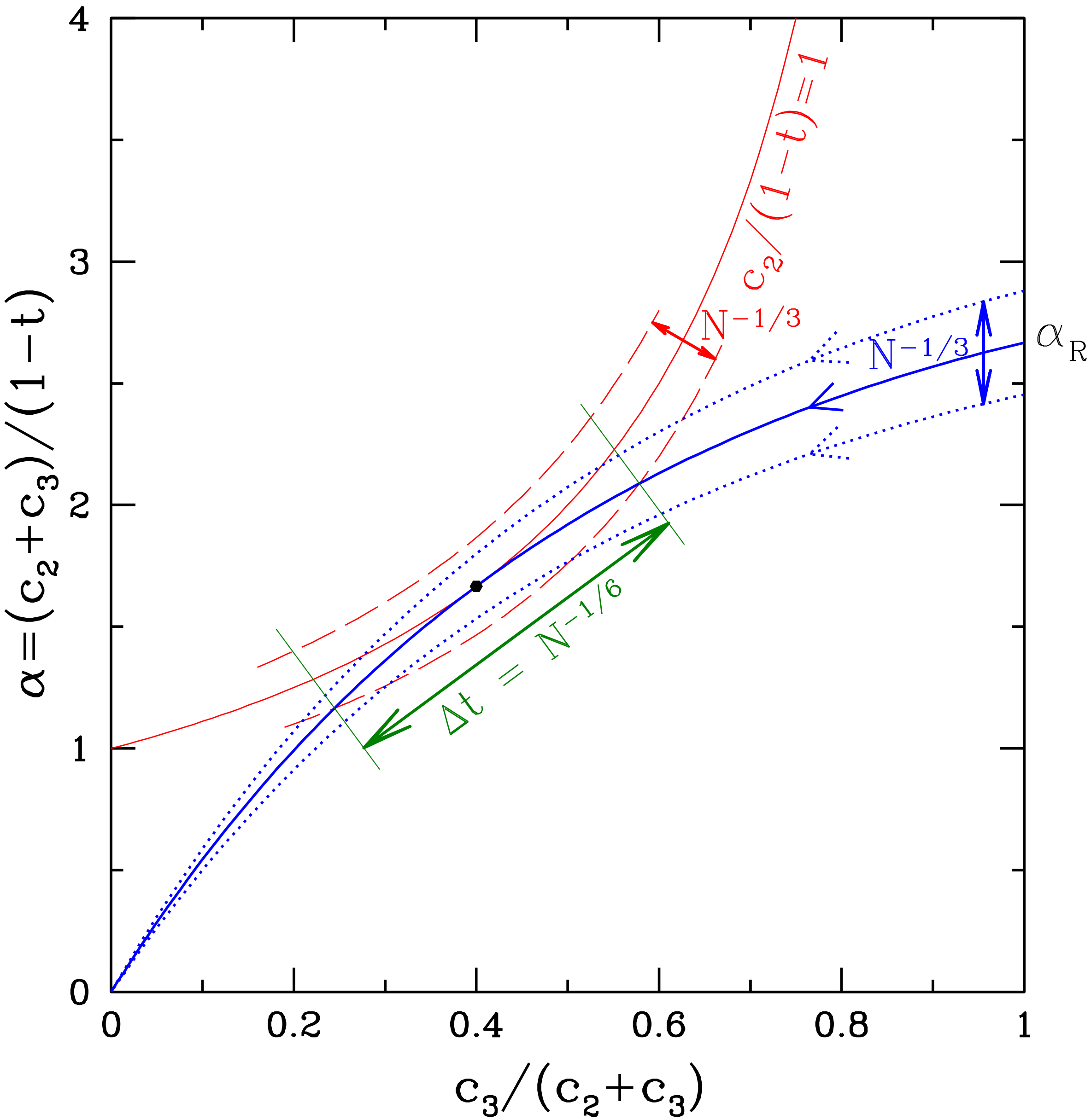
Matrice $H(C_1' \leftarrow C_1)$ devient :

$$p_N(T+1,x) = \left(1 + \frac{x-1}{N-T}\right)^{C_2(T)} \times \left(\frac{1}{A} p_N(T,A) + (1-\frac{1}{A}) p_N(T,0)\right)$$

avec $A := \frac{1}{2(N-T)} + (1 - \frac{1}{N-T})x$.

Exact, contient toute l'information sauf C_2 ; valable pour cette distribution des instances seulement mais pour tous les algorithmes avec Propagation Unitaire.

Zoom in



$\alpha = (c_2+c_3)/(1-t)$
$c_3/(c_2+c_3)$
$c_2/(1-t)=1$
$N^{-1/3}$
$N^{-1/3}$
α_R
$\Delta t = N^{-1/6}$
0
1
2
3
4
0
0.2
0.4
0.6
0.8
1

Zoom in

Zoom in autour du point de tangence $\frac{c_2}{1-t} = 1$. Dans le cas de `random` :

- $\alpha = \frac{8}{3}(1 + \epsilon_0 N^{-\theta})$
- $t = \frac{1}{2}\,(1 + t_0 N^{-\tau})$
- $C_2(T) = 4T(1 - \frac{T}{N})^2 + \mathcal{O}(\sqrt{N})$

θ, τ à déterminer pour capturer le 1er ordre non nul.

Hypothèses

Hypothèses sur la distribution de $C_1(T)$ (on ne sait pas tout démontrer ☹) :

- $P_N(C_1, T) = e^{-N^\lambda \phi(t_0)} F(C_1, t_0)$ avec F normalisée (F = distribution de C_1 conditionnée à l'absence de contradiction remarquée).

Hypothèses

Hypothèses sur la distribution de $C_1(T)$ (on ne sait pas tout démontrer ☹) :

- $P_N(C_1, T) = e^{-N^\lambda \phi(t_0)} F(C_1, t_0)$ avec F normalisée (F = distribution de C_1 conditionnée à l'absence de contradiction remarquée).

- La valeur typique de C_1 se comporte comme N^γ, $0 < \gamma < 1$:
  $F(C_1, t_0) = \frac{1}{N^\gamma} f\left(\frac{C_1}{N^\gamma}\right)$ avec $f(c)$ fonction régulière.

Hypothèses

Hypothèses sur la distribution de $C_1(T)$ (on ne sait pas tout démontrer ☹) :

- $P_N(C_1, T) = e^{-N^\lambda \phi(t_0)} F(C_1, t_0)$ avec F normalisée (F = distribution de C_1 conditionnée à l'absence de contradiction remarquée).

- La valeur typique de C_1 se comporte comme N^γ, $0 < \gamma < 1$:
  $F(C_1, t_0) = \frac{1}{N^\gamma} f\left(\frac{C_1}{N^\gamma}\right)$ avec $f(c)$ fonction régulière.

- Grande déviation : $F(C_1 = 0, t_0) = \frac{f_0(t_0)}{N^{\gamma_0}}$.

Hypothèses

Hypothèses sur la distribution de $C_1(T)$ (on ne sait pas tout démontrer ☹) :

- $P_N(C_1, T) = e^{-N^{\lambda}\phi(t_0)} F(C_1, t_0)$ avec F normalisée (F = distribution de C_1 conditionnée à l'absence de contradiction remarquée).

- La valeur typique de C_1 se comporte comme N^{γ}, $0 < \gamma < 1$: $F(C_1, t_0) = \frac{1}{N^{\gamma}} f\left(\frac{C_1}{N^{\gamma}}\right)$ avec $f(c)$ fonction régulière.

- Grande déviation : $F(C_1 = 0, t_0) = \frac{f_0(t_0)}{N^{\gamma_0}}$.

On s'intéresse aux valeurs de C_1 typiques, donc en N^{γ}, donc dans la fonction génératrice il faut prendre $x = 1 - \frac{x_0}{N^{\gamma}}$ (zoom in) :

$$p_N(T, x) = e^{-N^{\lambda}\phi(t_0)} \pi(x_0, t_0)$$

Calculs...

On remplace ces notations et hypothèses dans

$$p_N(T+1,x) = \left(1+\frac{x-1}{N-T}\right)^{C_2(T)} \times \left(\frac{1}{A}p_N(T,A) + (1-\frac{1}{A})p_N(T,0)\right)$$

avec $A := \frac{1}{2(N-T)} + (1-\frac{1}{N-T})x$.

θ et τ sont libres. Si on demande que les termes non nuls au premier ordre soient tous du même ordre :
$\boxed{\theta = \gamma = \gamma_0 = \frac{1}{3}, \lambda = \tau = \frac{1}{6}}$.

Et

$$-\partial_{t_0}\phi(t_0)\pi(x_0,t_0) = \partial_{x_0}\pi(x_0,t_0) + \pi(x_0,t_0)\left(\frac{x_0^2}{2} + (t_0^2-\epsilon_0)x_0\right) - f_0(t_0)x_0$$

En $x_0 = 0 \rightsquigarrow \partial_{t_0}\phi(t_0) = -\partial_{x_0}\pi(0,t_0) = \overline{c}(t_0)$.

EDP pour π

Donc :
$\partial_{x_0}\pi(x_0,t_0) + \pi(x_0,t_0)\left(\frac{x_0^2}{2} + (t_0^2 - \epsilon_0)x_0 + \overline{c}(t_0)\right) - f_0(t_0)x_0 = 0$
avec $\overline{c}(t_0)$ moyenne de c.

Remarque : $\pi(x_0,t_0)$, fonction génératrice, est aussi la transformée de Laplace de $f(c = \frac{C_1}{N}, t_0)$.
f régulière $\Rightarrow$ pour $x_0 \to +\infty$,

$$\pi(x_0,t_0) = 0 \times x_0 + 0 + \mathcal{O}(\frac{1}{x_0})$$

EDP pour π

Donc :

$\partial_{x_0}\pi(x_0,t_0)+\pi(x_0,t_0)\left(\frac{x_0^2}{2}+(t_0^2-\epsilon_0)x_0+\overline{c}(t_0)\right)-f_0(t_0)x_0=0$

avec $\overline{c}(t_0)$ moyenne de c.

Remarque : $\pi(x_0,t_0)$, fonction génératrice, est aussi la transformée de Laplace de $f(c=\frac{C_1}{N},t_0)$.

f régulière $\Rightarrow$ pour $x_0\to+\infty$,

$$\pi(x_0,t_0)=0\times x_0+0+\mathcal{O}(\frac{1}{x_0})$$

d'où les conditions aux limites pour f :

$$f_0(t_0)=\frac{f(0,t_0)}{2}$$

$$\partial_c f(c=0,t_0)+(t_0^2-\epsilon_0)f(c=0,t_0)=0$$

EDP pour f

$$\partial_{x_0}\pi(x_0,t_0)+\pi(x_0,t_0)\left(\frac{x_0^2}{2}+(t_0^2-\epsilon_0)x_0+\overline{c}(t_0)\right)-f_0(t_0)x_0=0$$

$\longrightarrow$ (transformée de Laplace inverse)

$$\boxed{\frac{1}{2}\frac{\partial^2 f}{\partial c^2}+(t_0^2-\epsilon_0)\frac{\partial f}{\partial c}+(c-\overline{c})f=0}$$

EDP pour f

$$\partial_{x_0}\pi(x_0,t_0) + \pi(x_0,t_0)\left(\frac{x_0^2}{2} + (t_0^2-\epsilon_0)x_0 + \overline{c}(t_0)\right) - f_0(t_0)x_0 = 0$$

$\longrightarrow$ (transformée de Laplace inverse)

$$\boxed{\frac{1}{2}\frac{\partial^2 f}{\partial c^2} + (t_0^2-\epsilon_0)\frac{\partial f}{\partial c} + (c-\overline{c})f = 0}$$

On résout pour f :

$$f(c,t_0) \propto e^{-(t_0^2-\epsilon_0)c}\mathrm{Ai}[\sqrt[3]{2}c + z(t_0^2-\epsilon_0)]$$

où $z(x)$ est la fonction inverse de $x(z) = \sqrt[3]{2}\frac{\mathrm{Ai}'(z)}{\mathrm{Ai}(z)}$.

Résultat final

Rappel : $P_N(C_1, T) = e^{-N^\lambda \phi(t_0)} F(C_1, t_0)$ avec F normalisée (F = distribution de C_1 conditionnée à l'absence de contradiction remarquée).

Donc $\ln P_{\text{succ}} = -N^\lambda \phi(t_0 = +\infty)$ et $\phi(t_0 = -\infty) = 0$.

$\partial_{t_0} \phi(t_0) = \overline{c}(t_0) \rightsquigarrow$

$$\boxed{-\ln P_{\text{succ}}((1+\epsilon_0)\alpha_{\texttt{random}}, N) = N^\lambda \Phi(\epsilon_0 N^\theta)}$$

avec

$$\boxed{\Phi(\epsilon_0) = \frac{1}{4} \int_{-\epsilon_0}^{+\infty} \frac{\mathrm{d}x}{\sqrt{\epsilon_0 + x}} \left[x^2 - 2^{2/3}\, z(x)\right]}$$

Portée de ce résultat et conclusion

Le calcul ne dépend de l'algorithme qu'à travers C_2. En fait, si on fait le zoom autour du point de tangence t_{A} (qui n'intervient pas directement) :

$$\frac{c_2}{1-t} = 1 + b(t - t_{\mathrm{A}})^2 \text{ pour } t \to t_{\mathrm{A}}$$

les calculs sont les mêmes, sauf que b a une valeur différente. On trouve

$$\Phi_{\mathrm{A}}(\epsilon_0) = r_{\mathrm{A}}^{\Phi}\, \Phi(r_{\mathrm{A}}^{\epsilon}\, \epsilon_0)$$

où les r_{A} sont fonctions de b.

Portée de ce résultat et conclusion

Le calcul ne dépend de l'algorithme qu'à travers C_2. En fait, si on fait le zoom autour du point de tangence t_A (qui n'intervient pas directement) :

$$\frac{c_2}{1-t} = 1 + b(t - t_\mathrm{A})^2 \text{ pour } t \to t_\mathrm{A}$$

les calculs sont les mêmes, sauf que b a une valeur différente. On trouve

$$\Phi_\mathrm{A}(\epsilon_0) = r_\mathrm{A}^\Phi \, \Phi(r_\mathrm{A}^\epsilon \, \epsilon_0)$$

où les r_A sont fonctions de b.

Cas non générique (possible si $K >$ 3-SAT) :

$$\frac{c_2}{1-t} = 1 + b(t - t_\mathrm{A})^{4,6,\ldots}$$

Portée de ce résultat et conclusion

Le calcul dépend de la distribution des instances : la fonction Φ peut changer. Mais les exposants sont les mêmes à cause de la robustesse de la distribution des tailles des composantes connexes d'un graphe aléatoire (cf. première partie).