

Distribution multimodale de la taille du sous-graphe des plus courts chemins dans un graphe aléatoire

Sergey Kirgizov

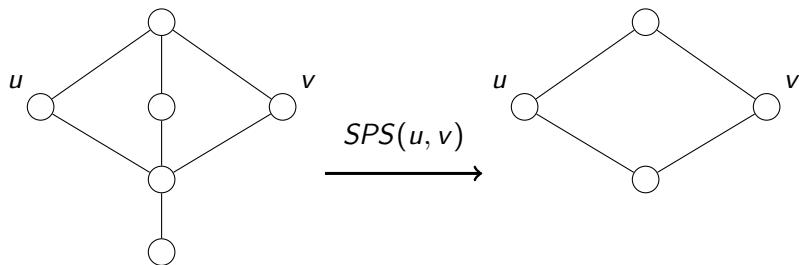
Complex Networks $\subset$ LIP6 $\subset$ (UPMC $\times$ CNRS)

Journées Graphes et Algorithmes
Novembre 2013

Plan

- ① Notion de sous-graphe des plus courts chemins
- ② Dans des graphes aléatoires
- ③ Dans des graphes aléatoires creux

Sous-graphe des plus courts chemins



$SPS(u, v)$: sous-graphe de tous les plus courts chemins entre u et v .

$S(u, v)$: nombre de sommets dans $SPS(u, v)$.

$d(u, v)$: distance entre u et v .

$$S(u, v) \geq d(u, v) + 1$$

Graphes aléatoires

$G(n, p)$ est un graphe aléatoire à n sommets. Chacune des $n(n - 1)/2$ arêtes est présente avec probabilité p .

1951–1952 *Solomonoff et Rapoport*

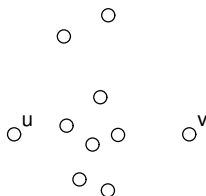
1959 *Gilbert, Erdős et Rényi*

Graphes aléatoires

$G(n, p)$ est un graphe aléatoire à n sommets. Chacune des $n(n - 1)/2$ arêtes est présente avec probabilité p .

1951–1952 *Solomonoff et Rapoport*

1959 *Gilbert, Erdős et Rényi*

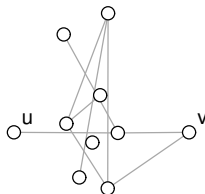


Graphes aléatoires

$G(n, p)$ est un graphe aléatoire à n sommets. Chacune des $n(n - 1)/2$ arêtes est présente avec probabilité p .

1951–1952 *Solomonoff et Rapoport*

1959 *Gilbert, Erdős et Rényi*

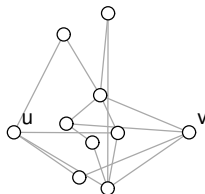


Graphes aléatoires

$G(n, p)$ est un graphe aléatoire à n sommets. Chacune des $n(n - 1)/2$ arêtes est présente avec probabilité p .

1951–1952 *Solomonoff et Rapoport*

1959 *Gilbert, Erdős et Rényi*



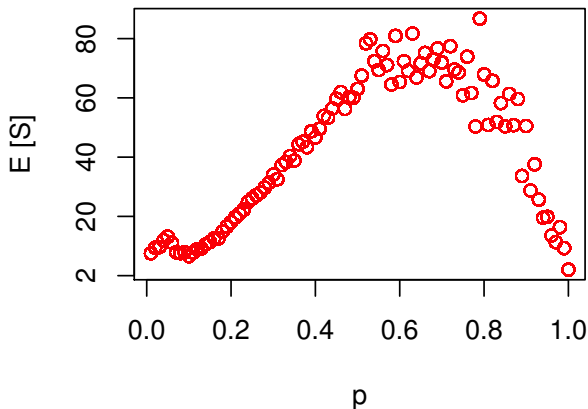
Taille moyenne de *SPS* dans des graphes aléatoires

Fixons n
 $p \mapsto \mathbb{E}[S]$

?

Taille moyenne de *SPS* dans des graphes aléatoires

Fixons n
 $p \mapsto \mathbb{E}[S]$



Moyenne et tailles réelles de *SPS* dans des graphes aléatoires

Fixons n

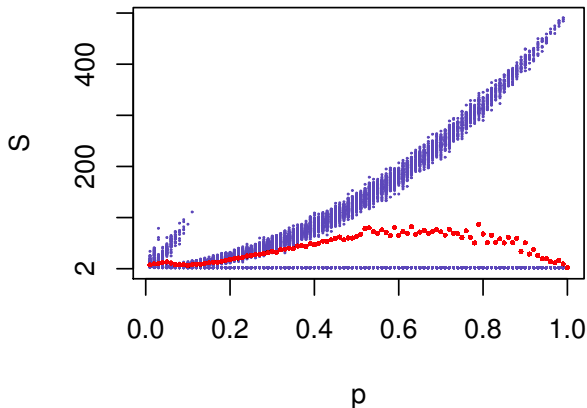
$$p \mapsto \{S\}$$



Moyenne et tailles réelles de *SPS* dans des graphes aléatoires

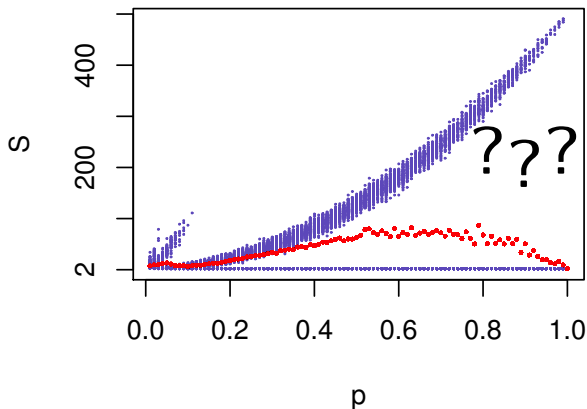
Fixons n

$$p \mapsto \{S\}$$



Moyenne et tailles réelles de *SPS* dans des graphes aléatoires

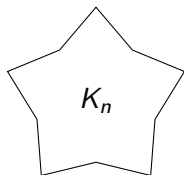
Fixons n
 $p \mapsto \{S\}$



Un “trou” ?

Graphe complet : K_n

$$p = 1$$

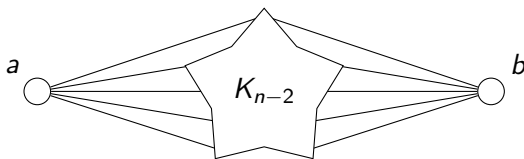


$$SPS(u, v) = \begin{array}{c} u \qquad v \\ \circ \text{---} \circ \end{array}$$

$$S(u, v) = 2$$

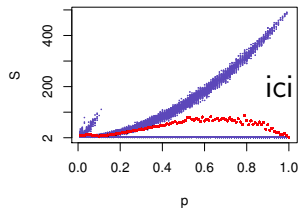
Graphe quasi-complet : $K_n - ab$

$$p = 1 - \varepsilon$$



$$S(u, v) = \begin{cases} n & \text{si } \{u, v\} = \{a, b\}, \\ 2 & \text{sinon.} \end{cases}$$

Si $n > 3$, alors il y a un “trou”.



Notations

$$f_d = \Pr[d(u, v) = d]$$

$$f_{>d} = \Pr[d(u, v) > d]$$

$$d(u, v) = x \iff \widetilde{uv}^x$$

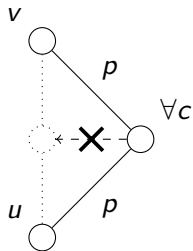
Les grands graphes aléatoires ont un diamètre 2

Théorème

Considérons un graphe aléatoire $G(n, p)$. Soit $p > 0$ et $n \rightarrow \infty$, alors on a $f_1 = p$, $f_2 \rightarrow 1 - p$, $f_{>2} \rightarrow 0$

Idée de la preuve

Par la définition de $G(n, p)$, on a $f_1 = p$ et $f_{>1} = 1 - p$.



$$f_{>2} = f_{>1}(1 - p^2)^{n-2}$$

Fixons $p > 0$, alors $f_{>2} \xrightarrow[n \rightarrow \infty]{} 0$.

SPS dans des grands graphes aléatoires

$f_{>2} \xrightarrow[n \rightarrow \infty]{} 0 \Rightarrow$ on n'a que deux cas : $\widetilde{uv}^1$ et $\widetilde{uv}^2$

SPS dans des grands graphes aléatoires

$f_{>2} \xrightarrow[n \rightarrow \infty]{} 0 \Rightarrow$ on n'a que deux cas : $\widetilde{uv}^1$ et $\widetilde{uv}^2$

$\widetilde{uv}^1$



$S = 2$

SPS dans des grands graphes aléatoires

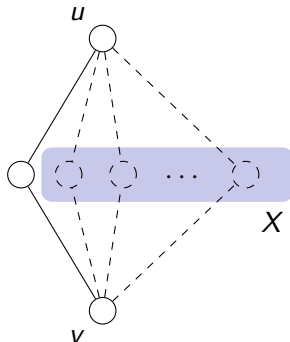
$f_{>2} \xrightarrow[n \rightarrow \infty]{} 0 \Rightarrow$ on n'a que deux cas : $\widetilde{uv}^1$ et $\widetilde{uv}^2$

$\widetilde{uv}^1$



$S = 2$

$\widetilde{uv}^2$



$S = 3 + X$

X , c'est combien ?

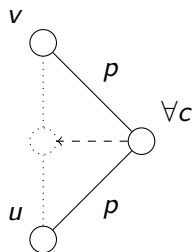
X , c'est combien ?

Théorème

Si la distance entre u et v est égal à 2, alors X suit une loi binomiale.

$$S - 3 = X \sim B(n - 3, p^2).$$

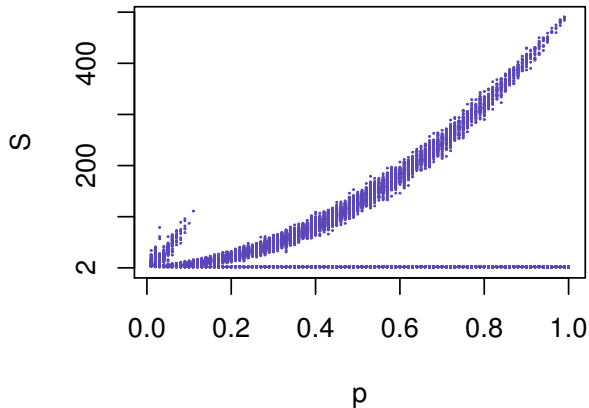
Idée de la preuve



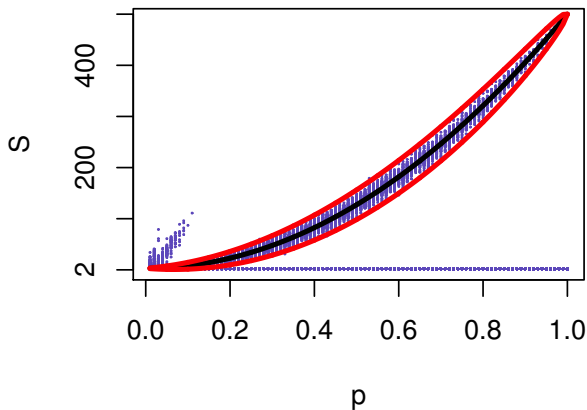
Formules

$$\mathbb{E} \left[S \mid \widetilde{uv}^2 \right] = 3 + (n - 3)p^2,$$
$$\text{Var} \left[S \mid \widetilde{uv}^2 \right] = (n - 3)p^2(1 - p^2).$$

Taille de *SPS* dans des graphes aléatoires

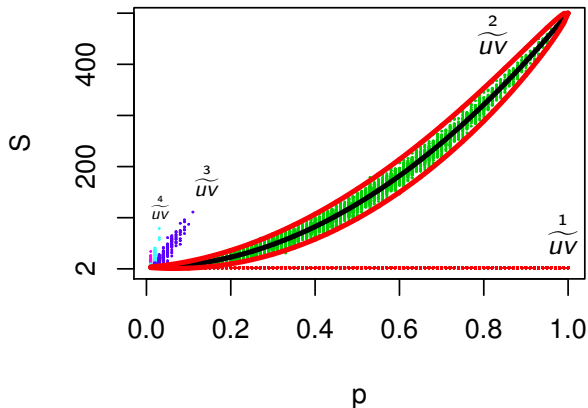


Taille de *SPS* dans des graphes aléatoires



$$\mathbb{E}[S | \widetilde{uv}^2] \pm 3\sqrt{\text{Var}[S | \widetilde{uv}^2]}$$

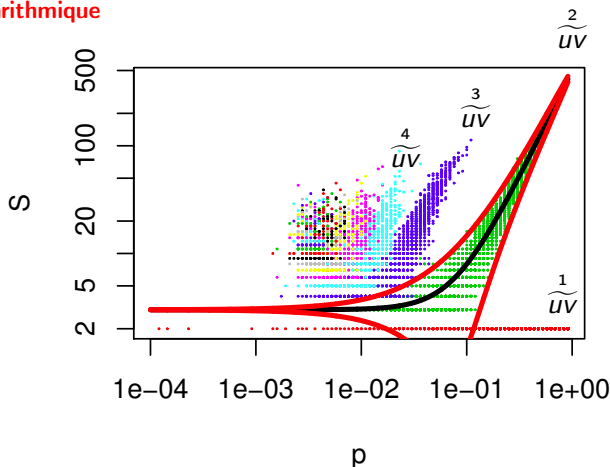
Taille de SPS dans des graphes aléatoires



$$\mathbb{E}[S | \widetilde{uv}^2] \pm 3\sqrt{\text{Var}[S | \widetilde{uv}^2]}$$

Taille de *SPS* dans des graphes aléatoires

échelle logarithmique



$$\mathbb{E}[S | \widetilde{UV}^2] \pm 3\sqrt{\text{Var}[S | \widetilde{UV}^2]}$$

Plan

- ① Notion de sous-graphe des plus courts chemins
- ② Dans des graphes aléatoires
- ③ Dans des graphes aléatoires creux

Distribution des distances internodaux

Blondel, Guillaume, Hendrickx, and Jungers en 2007 ont montré que f_d peut être approchée par une formule récurrente :

$$f_d \approx F_{d-1} - F_d, F_0 = 1 - \frac{1}{n},$$

$$F_d = \left(1 - \frac{1}{n}\right) (1 - p)^{(1-F_{d-1})n}.$$

Théorème

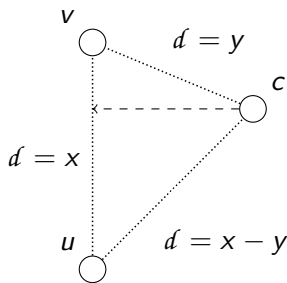
Si la distance entre u et v est égal à x , on a :

$$\mathbb{E} \left[S \mid \widetilde{uv}^x \right] \approx x + 1 + (n - x - 1) \sum_{y=1}^{x-1} f_y f_{x-y}.$$

Idée de l'approximation

$$\mathbb{E} \left[S \mid \widetilde{uv}^x \right] = x + 1 + (n - x - 1) \Pr [c \in SPS(u, v)]$$

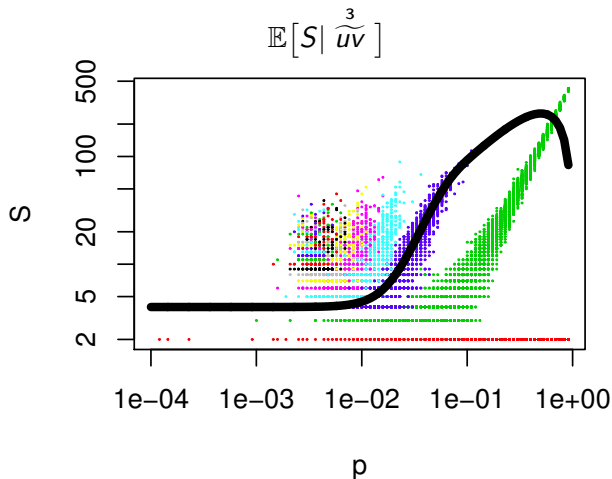
$$c \in SPS(u, v) \iff d(u, v) = d(u, c) + d(c, v)$$



$$\Pr [c \in SPS(u, v)] \approx \sum_{y=1}^{x-1} f_y f_{x-y}$$

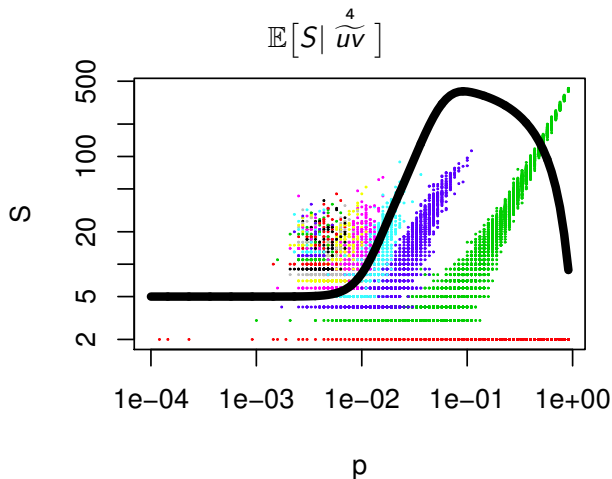
Résultat de l'approximation

échelle logarithmique

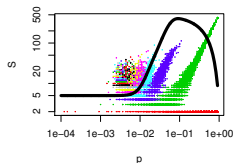
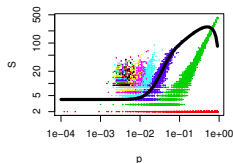
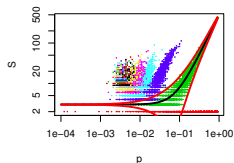


Résultat de l'approximation

échelle logarithmique



Conclusion

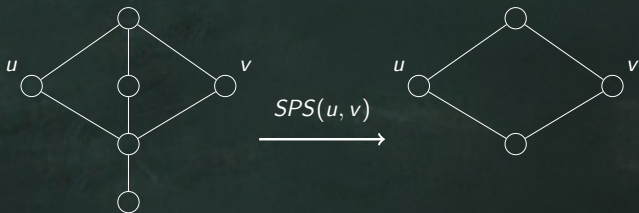


$$S = \begin{cases} 2 & \text{si } \widetilde{uv}^1, \\ 3 + X & \text{si } \widetilde{uv}^2 \text{ (où } X \sim B(n-3, p^2) \text{)}, \\ i + X_i & \text{si } \widetilde{uv}^{i-1}. \end{cases}$$

$$\mathbb{E}[X_i] \approx (n-i) \sum_{y=1}^{i-2} f_y f_{x-y}.$$

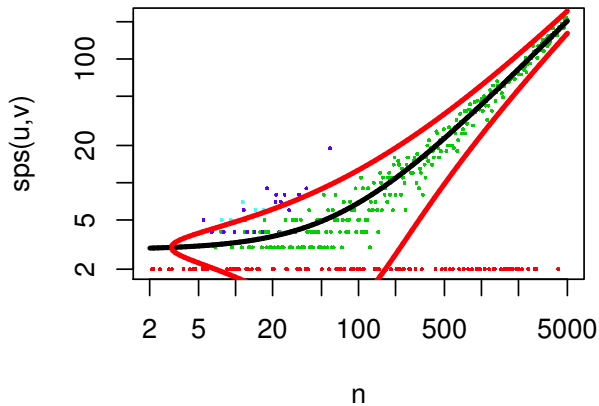
Perspectives

- Formules exactes
- *SPS* dans les graphes réels
- Autres applications de *SPS*
 - Détection de communautés
 - *SPS*-décomposition et *SPS*-reconstruction



Questions?

$p = 0.2, n \in [2, 5000]$



Grands graphes aléatoires ont un diamètre 2

Théorème (Grands graphes aléatoires ont un diamètre 2)

Considérons un graphe aléatoire $G(n, p)$. Soit $p > 0$ et $n \rightarrow \infty$, alors on a $f_1 = p$, $f_2 \rightarrow 1 - p$, $f_{>2} \rightarrow 0$

Démonstration.

- Par la définition de $G(n, p)$, on a $f_1 = p$ et $f_{>1} = 1 - p$.
- La distance entre deux sommets non adjacents est plus grand que 2 si et seulement si ils n'ont pas de voisins communs. Pour un sommet, ça arrive avec une probabilité $1 - p^2$, et ça doit être vrai pour les $n - 2$ sommets restants. Alors

$$f_{>2} = f_{>1}(1 - p^2)^{n-2}.$$

- Fixons $p > 0$, alors

$$f_{>2} \xrightarrow[n \rightarrow \infty]{} 0.$$

