Rozwiązanie zad. 3 (kolokwium z RPiS, 6 grudnia 2008)

Zadanie 3 (25 punktów) *Z kliki o n wierzchołkach wybieramy podgraf $G = G(n,p)$ w ten sposób, że każda krawędź jest niezależnie wybrana z prawdopodobieństwem p lub nie wybrana z prawdopodobieństwem $1-p$. Dla powstałego grafu G:*

1. *Oblicz prawdopodobieństwo, że ustalony wierzchołek jest izolowany (2 punkty).*
2. *Oblicz wartość oczekiwaną liczby wierzchołków izolowanych (5 punktów).*
3. *Wykaż, że wariancja liczby izolowanych wierzchołków jest równa*

$$n(1-p)^{n-1} + n(1-p)^{2n-3}(np-1)$$

   (8 punktów).

4. *Wykaż, że gdy $p = c \cdot \frac{ln(n)}{n}$ dla pewnej stałej $c < 1$, to prawdopodobieństwo, że G ma wierzchołek izolowany dąży do 1, gdy n dąży do nieskończoności (10 punktów).*

   Wskazówka: Skorzystaj z wyniku poprzedniego podpunktu i odpowiedniej nierówności.

Rozwiązanie

1. Ustalony wierzchołek jest izolowany wtedy i tylko wtedy, gdy żadna z krawędzi z nim incydentych nie została wylosowana. Tych krawędzi jest $n-1$ a prawdopodobieństwo, że pojedyńcza krawędź nie zostanie wylosowana wynosi $1-p$. A zatem, prawdopodobieństwo, że ustalony wierzchołek jest izolowany wynosi $(1-p)^{n-1}$.

2. Dla każdego wierzchołka v grafu $G(n,p)$, niech X_v będzie zmienną indykującą wydarzenie, że v jest wierzchołkiem izolowanym. Każda ze zmiennych X_v ma rozkład dwupunktowy o nośniku $\{0,1\}$ oraz wartości oczekiwanej równej $(1-p)^{n-1}$. Przez I_n oznaczmy sumę tych zmiennych losowych. A zatem, I_n interpretuje się jako liczbę izolowanych wierzchołków. Jej wartość oczekiwana wynosi

$$E[I_n] = E[X_1 + X_2 + \ldots + X_n] = n(1-p)^{n-1},$$

   na mocy liniowości wartości oczekiwanej.

3. Obliczymy teraz wariancję zmiennej I_n.

$$\begin{aligned}
Var[I_n] &= Var\Big[\sum_v X_v\Big] \\
&= E\Big[\big(\sum_v X_v - E[\sum_v X_v]\big)^2\Big] \\
&= E\Big[\sum_{u,v}\big(X_u - E[X_u]\big)\big(X_v - E[X_v]\big)\Big] \\
&= \sum_{u,v} E\Big[\big(X_u - E[X_u]\big)\big(X_v - E[X_v]\big)\Big] \\
&= \sum_u E\Big[\big(X_u - E[X_u]\big)^2\Big] + \sum_{u\neq v} E\Big[\big(X_u - E[X_u]\big)\big(X_v - E[X_v]\big)\Big] \\
&= n\ Var(X_1) + n(n-1)\ E\Big[\big(X_1 - E[X_1]\big)\big(X_2 - E[X_2]\big)\Big], \qquad (1)
\end{aligned}$$

gdzie zmienne u, v przebiegają po zbiorze wierzchołków grafu oraz 1 i 2 są dowolnymi dwoma spośród nich. Dowolność wyboru tych wierzchołku jest uzasadniona tym, że sposób losowania grafu jest niezmienniczy ze względu na permutacje wierzchołków.

Sprawdzimy, że

$$Var[X_1] = (1-p)^{n-1} - (1-p)^{2n-2} \qquad (2)$$

oraz, że

$$E\Big[\big(X_1 - E[X_1]\big)\big(X_2 - E[X_2]\big)\Big] = (1-p)^{2n-3}\cdot p. \qquad (3)$$

Faktycznie:

$$Var[X_1] = E[X_1^2] - E[X_1]^2$$

przy czym $X_1^2 = X_1$ gdyż zmienna ta ma wartości 0 lub 1. Tak więc,

$$Var[X_1] = E[X_1] - E[X_1]^2 = (1-p)^{n-1} - (1-p)^{2n-2}.$$

Z kolei

$$\begin{aligned}
E\Big[\big(X_1 - E[X_1]\big)\big(X_2 - E[X_2]\big)\Big] &= E[X_1\cdot X_2] - E[X_1]E[X_2] \\
&= P[X_1 = 1 \wedge X_2 = 1] - \big((1-p)^{n-1}\big)^2 \\
&\overset{(*)}{=} (1-p)^{2n-3} - (1-p)^{2n-2} \\
&= (1-p)^{2n-3}\big(1-(1-p)\big) \\
&= p(1-p)^{2n-3}.
\end{aligned}$$

W przejściu $(*)$ skorzystaliśmy z obserwacji, że $X_1 = 1 \wedge X_2 = 1$ wtedy i tylko wtedy, gdy nie wylosowano żadnej z $2n-3$ krawędzi wychodzących z wierzchołków 1 lub 2, a prawdopodobieństwo tego wydarzenia to $(1-p)^{2n-3}$. (Tych krawędzi jest $2n-3$ bo z wierzchołka 1 wychodzi $n-1$ krawędzi, a z wierzchołka 2 też wychodzi $n-1$ krawędzi, ale jedna jest wspólna).

Po wstawieniu równości (2) oraz (3) do (1), otrzymujemy żądany wynik.

4. Na mocy nierówności Czebyszewa (to zastosowanie nosi nazwę „metody drugiego momentu"),

$$\begin{aligned} P[I_n = 0] &= P\big[I_n - E[I_n] = -E[I_n]\big] \\ &\leq P\big[|I_n - E[I_n]| \geq E[I_n]\big] \\ &\leq \frac{Var[I_n]}{E[I_n]^2}. \end{aligned}$$

Aby wykazać zbieżność lewej strony do 0, wystarczy nam wykazać, że prawa strona nierówności zbiega do 0.

Gdy $p = c\frac{\ln n}{n}$ i $c < 1$, nietrudno jest wykazać następujące zbieżności przy $n \to \infty$:

$$n\frac{Var[X_1]}{E[I_n]^2} = n\frac{(1-p)^{n-1} - (1-p)^{2n-2}}{(n(1-p)^{n-1})^2} \to 0$$

oraz

$$n(n-1)\frac{E\Big[\big(X_1 - E[X_1]\big)\big(X_2 - E[X_2]\big)\Big]}{E[I_n]^2} = n(n-1)\frac{(1-p)^{2n-3} \cdot p}{(n(1-p)^{n-1})^2} \to 0.$$

A zatem, ze wzoru (1), $\frac{Var[I_n]}{E[I_n]^2} \to 0$, więc tym bardziej $Pr[I_n = 0] \to 0$ gdy $n \to \infty$, stąd, z prawdopodobieństwem dążącym do 1 istnieje izolowany wierzchołek.