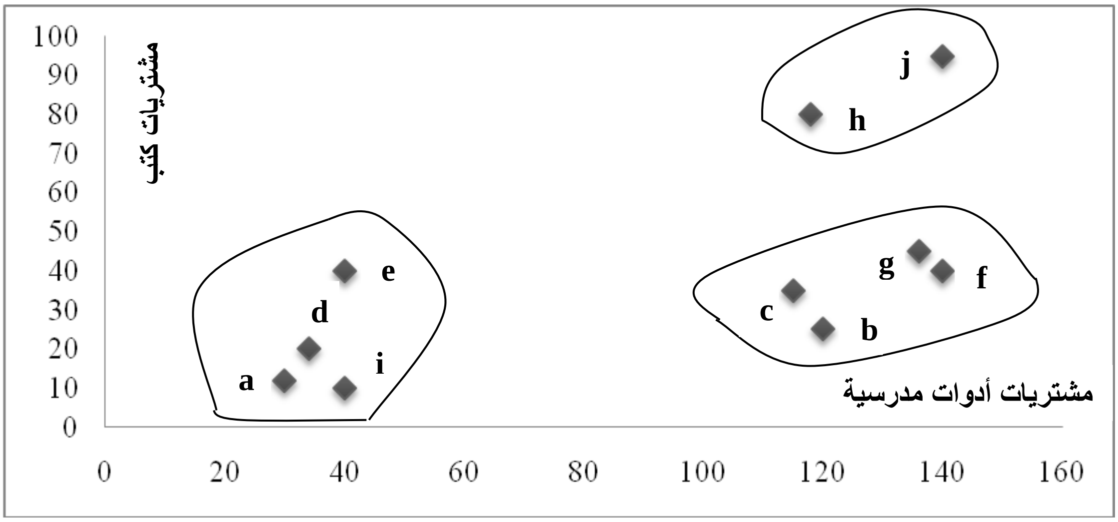
سلسلة تمارين رقم (01) حول التحليل العنقودي

حل التمرين الأول:

1. التمثيل البياني لمشتريات العملاء من الأدوات المدرسية والكتب.

نمثل كل عميل على الرسم البياني بثنائية نقطية (X_1,X_2)، حيث الفاصلة X_1 تمثل مشتريات العميل من الأدوات المدرسية، الترتيبة X_2 يمثل مشتريات العميل من الكتب.

المفردات هي مجموعة العملاء (a, b, c, d, e, f, g, h, i, j)، ومنه: المجتمع الاحصائي يتكون من n=10 مفردة أو مشاهدة. أما المتغيران الإحصائيان X_1 وX_2 فهما متغيران كميان (مشتريات: مبالغ نقدية)



2. التحليل:

نلاحظ أن مواقع العملاء على الرسم حسب مشترياتهم من الأدوات والكتب تشكل ثلاث مجموعات منفصلة هي:

- المجموعة G_1(a, d, e, i)، حيث المشتريات منخفضة من كل من الأدوات والكتب، يجب على المؤسسة دراسة أسباب انخفاض مشترياتهم، والبحث عن أساليب تسويقية لمعالجتها (توفير الأدوات والكتب المرغوبة)
- المجموعة G_2(b, c, f, g)، حيث المشتريات مرتفعة من الأدوات المدرسية، ومنخفضة من الكتب. يجب على المؤسسة البحث عن أسباب انخفاض مشتريات هؤلاء العملاء من الكتب وعلاجها، والحفاض وتدعيم مشترياتهم من الأدوات المدرسية بالوسائل السويقية الملائمة.
- المجموعة G_3(h, j)، حيث المشتريات مرتفعة من الأدوات المدرسية والكتب معا، هذه المجموعة تمثل مصدر النقدية الأساسي للمؤسسة، لذا يجب على المؤسسة الاهتمام برغباتهم (توفير احتياجاتهم نوعا وكما، منحهم إئتمان تجاري أكبر، منحهم تخفيضات أكبر...).

إن تصنيف هؤلاء العملاء حسب ومشتريات الأدوات X_1 ومشتريات الكتب X_2 ضمن مجموعات(عناقيد) متشابية داخلياً ومتباينة خارجياً، يتطلب اتباع منهجية العنقدة، والتي تقتضي حساب المسافات المختلفة بين مواقع هؤلاء العملاء للحصول على مصفوفة التباعد، ثم تجميعهم حسب الأقرب فالأقرب في فئات متشابهة من حيث المشتريات بشكل عام.

حل التمرين الثاني:

لإعداد مصفوفة التباعد D، نستخدم طريقة التجميع، لذا نفترض أولاً أن كل واحد من العملاء الخمسة يشكل لوحده عنقوداً خاصا، ثم نقوم بحساب المسافات الإقليدية بين كل زوج منهم (j, k) ، حيث j عميل وk عميل آخر بالعلاقة:

$$d_{jk}= [(x_{1j} – x_{1k})^2+(x_{2j}- x_{2k})^2]^{1/2}$$

وهكذا نجد أن المسافة بين العميلين (2 , 1) هي: $d_{12}= [(2 – 8)^2+(4- 2)^2]^{1/2}= 6.325$

وكذلك نجد أن المسافة بين العميلين (3 , 1) هي: $\mathbf{d_{13}= [(2 – 9)^2+(4- 3)^2]^{1/2}= 7.071}$

وبمتابعة حساب هذه المسافات للأزواج المختلفة الأخرى من العملاء(j,k)، بنفس الطريقة نحصل على مصفوفة التباعد D لهؤلاء العملاء التالية:

$$D=\begin{array}{c} \\ 1\\ 2\\ 3\\ 4\\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left[\begin{array}{ccccc} 0 & 6,325 & 7,071 & 1,414 & 7,159 \\ & 0 & 1,414 & 7,616 & 1,116 \\ & & 0 & 8,246 & 2,062 \\ & & & 0 & 8,500 \\ & & & & 0 \end{array}\right] \end{array}$$

نلاحظ أن مصفوفة التباعد هي مصفوفة مربعة (5×5)، وعناصر القطر معدومة: $d_{jj}=0$، وبما أن $d_{jk}=d_{kj}$، فهي مصفوفة مثلثية، لذا نكتفي بالعناصر فوق القطر فقط، ونقول أنها مصفوفة مثلثية عليا العناصر تحت القطر لا تظهر).

نلاحظ أن أصغر عناصر هذه المصفوفة هو (1,116) وهو يقابل العميلين 2 و5، لأنهما الأكثر تشابهاً في المشتريات، لذلك يمكن أن ننشأ منهما العنقود الأول.

إن عناصر مصفوفة التقارب S_{jk} ترتبط مع عناصر مصفوفة التباعد d_{jk} من خلال العلاقة التالية

$$S_{jk}= \frac{1}{1+ d_{jk}} \qquad 0\leq S_{jk}\leq 1$$

فمثلا: $s_{12}= 1/(1+6,325)= 0,136$ ؛ $s_{13}= 1/(1+7,071)= 0,124$ ؛ $s_{14}= 1/(1,414)=0,414$ ؛..... وهكذا بالنسبة لبقية العناصر، ونجد مصفوفة التقارب S .

وحتى نميز مصفوفة التقارب S عن مصفوفة التباعد D، فإن العناصر المحسوبة نضعها تحت القطر وليس فوقه، فالعنصر s_{jk} لا نضعه عند تقاطع السطر j والعمود k، بل عند تقاطع السطر k والعمود j، فمثلا العنصر $s_{12}=0,136$ لا نضعه عند تقاطع السطر الأول والعمود الثاني كما في مصفوفة التباعد، بل بالعكس عند تقاطع السطر الثاني والعمود الأول، فيظهر تحت القطر وليس فوقه.

$$S=\begin{array}{c} \\ 1\\ 2\\ 3\\ 4\\ 5 \end{array}\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left[\begin{array}{ccccc} 1 & & & & \\ 0,136 & 1 & & & \\ 0,124 & 0,414 & 1 & & \\ 0,414 & 0,116 & 0,109 & 1 & \\ 0,126 & 0,473 & 0,327 & 0,105 & 1 \end{array}\right] \end{array}$$

نلاحظ أن مصفوفة التقارب أو التشابه S مربعة (5×5)، وعناصر القطر تساوي 1، لأن كل عنصر متماثل مع نفسه ($s_{jj}=1$)، وبما أن $s_{jk}=s_{kj}$، فهي مصفوفة مثلثية، ولتمييزها عن مصفوفة التباعد D نسجل القيم تحت السطر فقط (مثلثية من الأسفل).

حل التمرين الثالث:

نقوم بتحويل كل من هذه المتغيرات إلى متغيرات ثنائية، ونعرف منها المتغيرات الثنائية التالية: $X_1', X_2', X_3', X_4', X_5', X_6'$ كما يلي:

$$X'_1=\begin{cases}1: & X_1\geq 72\\ 0: & X_1< 72\end{cases} \qquad X'_2=\begin{cases}1: & X_2\geq 150\\ 0: & X_2<\end{cases} \qquad X'_3=\begin{cases}1: & X_3=\text{بني}\\ 0: & X_3=\text{غير ذلك}\end{cases}$$

$$X'_4=\begin{cases}1: & X_4=\text{أشقر}\\0: & X_4=\text{أسود}\end{cases}\qquad X'_5=\begin{cases}1: & X_5=\text{يمنى}\\0: & X_5=\text{يسرى}\end{cases}\qquad X'_6=\begin{cases}1: & X_6=\text{أنثى}\\0: & X_6=\text{ذكر}\end{cases}$$

وبعد تفريغ قيم هذه المتغيرات، نحصل على جدول جديد للمتغيرات الثنائية يأخذ الشكل التالي:

جدول نتائج خصائص الطلبة بدلالة المتغيرات الثنائية

المتغيرات/المفردات	X_1'	X_2'	X_3'	X_4'	X_5'	X_6'
1	0	0	0	1	1	1
2	1	1	1	0	1	0
3	0	1	0	1	1	0
4	0	0	1	0	1	1
5	1	1	1	0	0	0

ولإنشاء مصفوفة التقارب S لهؤلاء الطلاب، علينا أولاً أن نقوم بإيجاد جداول التوافق لكل زوج منهم (وعددها $C^2_5=10$)، وذلك بناء على بيانات الجدول السابق، فنجد مثلاً أن جدول التوافق للطالبين (1) و(2) يأخذ الشكل التالي:

		قيم الطالب (2)		المجموع
		1	0	1
قيم الطالب (1)	1	1	2	3
	0	3	0	3
المجموع		4	2	6

ومنه نحسب مقياس التقارب بينهما، وذلك من خلال المقياس الأوللتقارب المعرف بالعلاقة التالية:

$$S_{12}= (a+d)/P = (1+0)/6= 1/6$$

وكذلك نجد أن جدول التوافق للطالبين (1) و (3) يأخذ الشكل التالي:

		قيم الطالب (3)		المجموع
		1	0	1
قيم الطالب (1)	1	2	1	3
	0	1	2	3
المجموع		3	3	6

$$S_{13}= (a+d)/P = (2+2)/6= 4/6$$

وبمتابعة حساب بقية عناصر مصفوفة التقارب S_{jk} بين هؤلاء الطلاب وباستخدام نفس المقياس، نحصل على مصفوفة متناظرة من الرتبة (5×5) وهو عدد المفردات، وتأخذ الشكل التالي:

$$S=\begin{array}{c} \\ 1\\2\\3\\4\\5\end{array}\begin{array}{c}\begin{array}{ccccc}1&2&3&4&5\end{array}\\ \left[\begin{array}{ccccc}1&&&&\\1/6&1&&&\\4/6&3/6&1&&\\4/6&3/6&2/6&1&\\0&5/6&2/6&2/6&1\end{array}\right]\end{array}$$

ومن خلال هذه المصفوفة نلاحظ مباشرة أن أكبر عنصر في المصفوفة S هو العنصر التكراري $S_{52}=5/6$، الذي يقابل الطالبين (2) و(5)، وهذا يعني أن هذين الطالبين أقرب إلى بعضهما (حسب المتغيرات المستخدم) من أي طالبين آخرين، لذلك يمكننا أن نشكل منهما مجموعة أولى تمثل العنقود الأول.

كما نلاحظ أن أبعد طالبين عن بعضهما هما الطالبين (1) و(5)، لأنهما يقابلان أصغر عنصر في المصفوفة هو $S_{51}= 0$، وهناك أزواج تقع بين هاتين الحالتين.

واذا أردنا تقسيم الطلاب إلى مجموعتين جزئيتين متجانستين نسبياً بناء على بيانات مصفوفة التقارب، يمكننا أن نشكل مجموعتين جزئيتين مؤلفتين من هؤلاء الطلاب: $G_1=(2,5)$ و $G_2= (1, 3, 4)$.

حل التمرين الرابع:

لتصنيف الطلاب الـ (6) ضمن عناقيد متشابهة، نقوم بتطبيق خطوات الخوارزمية السابقة فنجد أن:

الخطوة (1): نعتبر أن كل طالب (مفردة) يشكل عنقوداً مستقلا، ثم نقوم بدراسة عناصر المصفوفة D، فنجد أن أصغر عنصر في المصفوفة D هو العنصر (3) المقابل للطالبين (3) و(5)، لذلك نقوم بدمج هذين الطالبين في عنقود واحد، ونرمز له بـ: (3, 5)، ونحذف العمودين (3) و(5) والسطرين (3) و(5) من المصفوفة D، ثم نظيف عموداً خاصاً وسطرا خاصاً للعنقود الجديد (3, 5)، ونضعه مكان العمود (3) والسطر (3)، فنحصل على المصفوفة التالية:

$$D_1= \begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & (3,5) & 4 & 6 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ (3,5) \\ 4 \\ 6 \end{array} & \begin{bmatrix} 0 & 4 & d_{1(3,5)} & 24 & 8 \\ & 0 & d_{2(3,5)} & 22 & 10 \\ & & 0 & d_{(3,5)4} & d_{(3,5)6} \\ & & & 0 & 18 \\ & & & & 0 \end{bmatrix} \end{array}$$

وهنا نلاحظ أن عملية الدمج لا تؤثر على المفردات أو العناقيد الأخرى، ويتم حساب عناصر العمود (3,5)، ثم عناصر السطر (3,5) من المصفوفة الأساسية D، علما أن: $d_{jk}= d_{kj}$ وفق العلاقات التالية:

$d_{1(3,5)}= \min(d_{13}, d_{15})= \min (13, 12)= 12$

$d_{2\,(3,5)} = \min (d_{23}, d_{25})= \min (10, 11)= 10$

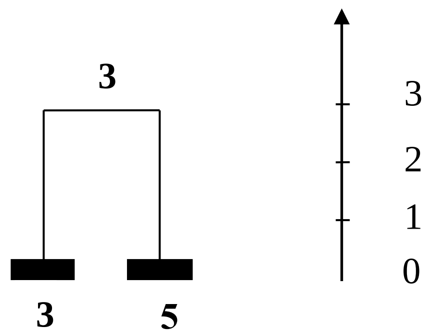
$d_{(3,5)4}= \min (d_{34}, d_{54})= \min(7, 6)= 6$

$d_{(3,5)6}= \min (d_{36}, d_{56})= \min(9, 8.5)= 8.5$

وبذلك نحصل على مصفوفة المسافات الجديدة التالية:

$$D_1= \begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & (3,5) & 4 & 6 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ (3,5) \\ 4 \\ 6 \end{array} & \begin{bmatrix} 0 & 4 & 12 & 24 & 8 \\ & 0 & 10 & 22 & 10 \\ & & 0 & 6 & 8.5 \\ & & & 0 & 18 \\ & & & & 0 \end{bmatrix} \end{array}$$

(1)



الخطوة (2): ولقد رسمنا العنقود إلى جانب المصفوفة D_1 للتوضيح، ولمتابعة إجراء العنقدة، نقوم بتطبيق الخوارزمية مرة أخرى، فنلاحظ أن أصغر عنصر في المصفوفة الأخيرة D_1 هو العنصر (4) المقابل للمفردتين (1) و(2)، لذلك نقوم بدمج المفردتين (1) و(2) ضمن عنقود آخر نرمز له بـ: (1, 2)، ونحذف العمودين (1) و(2) والسطرين (1) و(2)، ثم نخصص عموداً واحدا (1,2) وسطرا (1,2) للعنقود الجديد ونضيفهما في مكان العمود (1) والسطر (1)، فنحصل على المصفوفة التالية:

$$D_2=\begin{array}{cc} & \begin{array}{cccc}(1,2) & (3,5) & 4 & 6\end{array}\\ \begin{array}{c}(1,2)\\(3,5)\\4\\6\end{array} & \begin{bmatrix}0 & d_{(1,2)(3,5)} & d_{(1,2)4} & d_{(1,2)6}\\ & 0 & 6 & 8.5\\ & & 0 & 18\\ & & & 0\end{bmatrix}\end{array}$$

ولحساب عناصر العنقود (1,2) الواقعة في السطر(1,2) نحسبها من المصفوفة D_1، وذلك بتطبيق العلاقات التالية:

$d_{(1,2)(3,5)}= \min [d_{1(3,5)}, d_{2(3,5)}]=\min[12, 10]= 10$

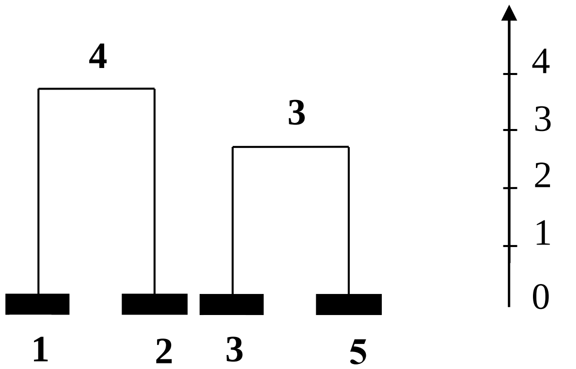
$d_{(1,2)4}= \min [d_{14}, d_{24}]=\min[24, 22]= 22$

$d_{(1,2)6}= \min [d_{16}, d_{26}]=\min[8, 10]= 8$

وبذلك نحصل على مصفوفة المسافات D_2 التالية:

$$D_2=\begin{array}{cc} & \begin{array}{cccc}(1,2) & (3,5) & 4 & 6\end{array}\\ \begin{array}{c}(1,2)\\(3,5)\\4\\6\end{array} & \begin{bmatrix}0 & 10 & 22 & 8\\ & 0 & \mathbf{6} & 8.5\\ & & 0 & 18\\ & & & 0\end{bmatrix}\end{array}$$

(2)



الخطوة (3): ولمتابعة عملية العنقدة نقوم بتطبيق الخوارزمية السابقة من جديد، فنجد أن أصغر عنصر في المصفوفة D_2 هو (6) المقابل للمفردة (4) والعنقود (3,5)، لذلك ندمجها في عنقود واحد ونرمز له بـ [(3,5),4]، ونحذف العمودين (4) و(3,5) والسطرين (4) و (3,5)، ثم نضيف عمودا جديدا [(3,5),4] وسطرا جديدا [(3,5),4]، ونضعهما مكان العمود والسطر المحذوفين، فنحصل على المصفوفة التالية:

$$D_3=\begin{array}{cc} & \begin{array}{ccc}(1,2) & [(3,5),4] & 6\end{array}\\ \begin{array}{c}(1,2)\\ [(3,5),4]\\ 6\end{array} & \begin{bmatrix}0 & d_{(1,2)[(3,5),4]} & 8\\ & 0 & d_{[(3,5),4],6}\\ & & 0\end{bmatrix}\end{array}$$

ولحساب عناصر العنقود الجديد [(3,5),4]، نجد من المصفوفة D_3 أن :

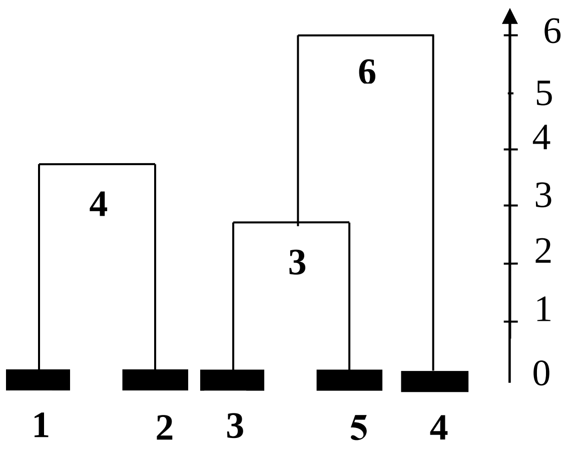
$d_{(1,2)[(3,5),4]}= \min[d_{(1,2)(3,5)}, d_{(1,2)4}]= \min[10, 22]= 10$

$d_{[(3,5)4]6}= \min[d_{(3,5)6}, d_{46}]= \min[8.5 , 18]= 8.5$

وبذلك نحصل على المصفوفة التالية D_3:

$$D_3=\begin{array}{cc} & \begin{array}{ccc}(1,2) & [(3,5),4] & 6\end{array}\\ \begin{array}{c}(1,2)\\ [(3,5),4]\\ 6\end{array} & \begin{bmatrix}0 & 10 & \mathbf{8}\\ & 0 & 8.5\\ & & 0\end{bmatrix}\end{array}$$

(3)



الخطوة (4): ولمتابعة عملية العنقدة، نلاحظ أن أصغر عنصر في المصفوفة الأخيرة D_3 هو (8) المقابل للمفردة (6) وللعنقود (1,2)، لذلك ندمجها في عنقود جديد نرمز له بـ [(1,2),6]، وبعد إجراء العمليات والحسابات اللازمة نحصل على المصفوفة النهائية D_4 التالية:

$$D_4=\begin{array}{c}\\ [(1,2),6] \\ [(3,5),4]\end{array}\begin{array}{cc}[(1,2),6] & [(3,5),4] \\ \left[\begin{array}{c} 0 \\ \\ \end{array}\right. & \left.\begin{array}{c} d_{[(1,2),6][(3,5),4]} \\ 0 \end{array}\right]\end{array}$$

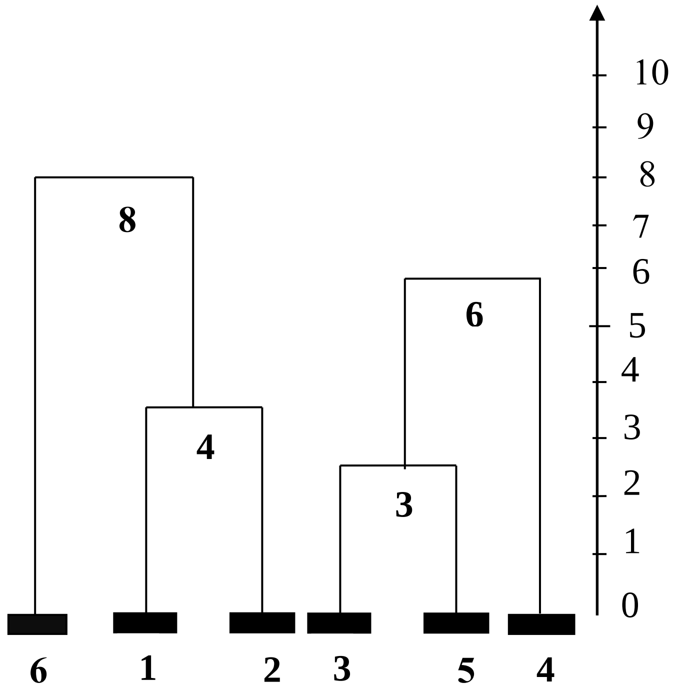
ولحساب عناصر العنقود الجديد [(1,2),6]، نجد من المصفوفة D_3 أن :

$$d_{[(1,2),6][(3,5),4]}= \min [d_{[(1,2)[(3,5),4]}, d_{[(3,5),4]6}]= \min[10, 8.5]= 8.5$$

وبذلك نحصل على مصفوفة المسافات D_4 التالية:

$$D_4=\begin{array}{c}\\ [(1,2),6] \\ [(3,5),4]\end{array}\begin{array}{cc}[(1,2),6] & [(3,5),4] \\ \left[\begin{array}{c} 0 \\ \\ \end{array}\right. & \left.\begin{array}{c} 8.5 \\ 0 \end{array}\right]\end{array}$$

(4)



وبذلك نحصل على مصفوفة D_4 تتألف من عنقودين: [(3,5),4] و[(1,2),6]، نقوم بدمجهما في عنقود واحد لتشكيل العنقود الأخير فنحصل على المخطط التالي، ونلاحظ أن أطوال الأغصان هي المسافات بين العناقيد.