

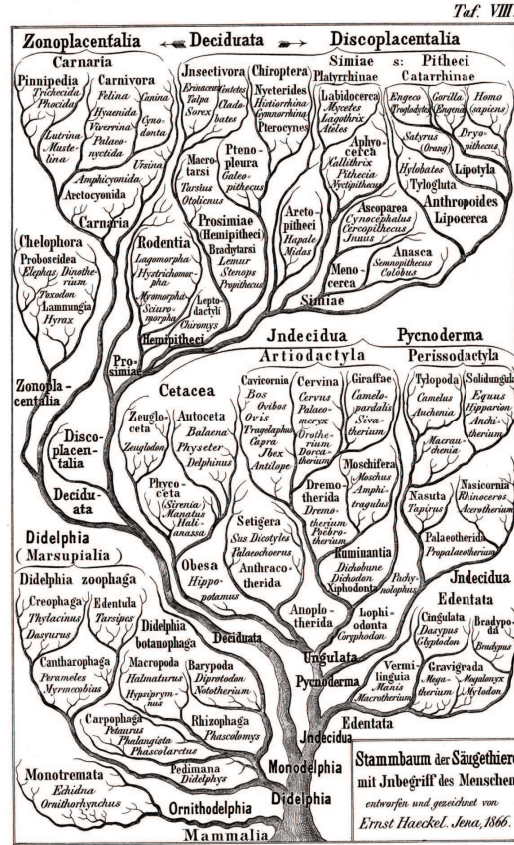
# Evrim Ağaçları ve Dört-Nokta Teoremi

Gökçe ÇAKMAK \*

Ali DENİZ †

Şahin KOÇAK ‡

Canlıların sınıflandırılması biyologlar için her zaman en önemli konulardan biri olmuştur. Yüzyıllardır morfolojik kriterlere dayalı olarak yapılan bu karşılaştırma ve düzenleme işlemi, "Sistematik" adıyla biyolojinin önemli bir dalını oluşturur. Evrim olgusunun kavranmasından sonra, sistematik dalının amacı, büyük evrim ağacını gerçeğe en yakın biçimde inşa etmeye çalışmak olmuştur. Şekil 1'de 1866'da bu "hayat ağaçları"nın ilk örneklerini oluşturan Ernst Haeckel'in bir çizimini görüyorsunuz. (İlginç bir not olarak, yüz küsur sene önce, iki genç Osmanlı münevverinin bu panteist Alman biyoloğunun bir eserini çevirmiş olduklarını belirtelim [3].) Hayat ağaçlarının (ve başka evrimsel ağaçların, örneğin dillerin dallanmasının) daha yeni örneklerini internette bolca bulabilirsiniz.



Şekil 1: Ernst Haeckel'in bir çizimi, 1866

Son dönemde, organizmaların DNA-dizilerinin çözülmesiyle, sistematikçilerin eline umulmadık yeni bir araç geçmiştir. Bu nükleotid dizilerinin (çoğu kez onların bazı özel parçalarının) karşılaştırılmasıyla, iki organizmaya sayısal bir "uzaklık" ölçüsü karşılık getirmek mümkün olmuş ve böylece matematikçileri

\*Eskişehir Teknik Üniversitesi. e-mail: gokcecakmak@eskisehir.edu.tr

†Eskişehir Teknik Üniversitesi. e-mail: adeniz@eskisehir.edu.tr

‡Anadolu Üniversitesi. e-mail: skocak@anadolu.edu.tr

de şaşırtacak (ve onları göreve davet edecek!) biçimde, elemanları organizmalar olan sonlu metrik uzaylar ortaya çıkmıştır.

Bir organizma kümesi verildiğinde, bunların ortak bir atadan zaman içinde mutasyon ile dallanarak oluştuğu düşünülürse, bunlar arasında gerçekçi bazı uzaklıkların tanımlanması halinde, ortaya çıkacak sonlu metrik uzayın, matematikte zaten uzun süredir başka amaçlarla da kullanılmakta olan, "metrik bir ağaç" formunda olmasını beklemek doğaldır. Daha açık ifade etmek gerekirse, organizmalardan oluşan böyle bir metrik uzay, organizmalar köşelere resmedilecek şekilde metrik bir ağaca gömülebilmelidir. Bunun böyle olup olmadığına nasıl karar verilebilir? Eğer böyleyse, organizmik uzayın gömülebileceği bir metrik ağaç nasıl bulunur? Aşağıda bu sorulara cevap vermeye çalışacağız. Şüphesiz, öncelikle, bahsi geçen bazı kavramları netleştirmemiz iyi olacak.

## 1 Metrik Ağaçlar

Çizge kavramını artık herkes biliyor. Matematğin bu en basit kavramı, belki de en önemli kavramdır, çünkü bu kavram bir bakıma matematiğin kendisi için bir metaforudur. Matematik, nesnelere ve onların arasındaki ilişkilerin bilimiye, çizge de zaten tam öyle bir şeydir: Köşeler ve onların bazılarını ilişkilendiren kenarlar. Hilbertvâri bir yaklaşımla nesnelere de unutup, sadece ilişkilere bakarsak, geriye sadece kenarlar kalır ve gerçekten de köşelerde nelerin oturduğu önemli değildir; çizgenin yapısını kenarların birbirlerine nasıl el verdiği belirler.

Çizgelerimizin köşe sayısının sonlu olmasını ve iki köşe arasında en fazla bir kenar olmasını istiyoruz (bir köşeden tekrar kendisine giden kenarlara da izin vermiyoruz). Bizi bu yazıda ilgilendiren, çok yerinde bir ifade ile ağaç adı verilen özel çizgeler olacak. Bunlar bağlantılı ve döngüleri olmayan çizgelerdir, yani herhangi bir köşeden diğerine kenarlar üzerinden yürüyerek ulaşabiliriz ve bir köşeden yola çıkıp, ardışık (ve farklı) kenarlar üzerinden yürüyerek tekrar aynı köşeye dönemeyiz.

Bir çizgenin köşelerine  $v_1, v_2, \dots, v_N$  dersek,  $i \neq j$  olmak üzere,  $v_i$  ve  $v_j$  köşeleri arasındaki (muhtemel) bir kenarı  $e_{v_i v_j}$  veya  $e_{ij}$  ile göstereceğiz ve  $v_i, v_j$  köşelerine bu kenarın uçları (veya köşeleri) diyeceğiz. (Tabii her  $i \neq j$  için bir  $e_{ij}$  kenarı olmayabilir.)

Bir ağacı şöyle de tanımlayabiliriz:

$v_i, v_j$  gibi herhangi iki farklı köşe verildiğinde, farklı kenarlardan oluşan tek bir  $e_{i_1 i_2} e_{i_2 i_3} \dots e_{i_{k-1} i_k} e_{i_k j}$  yolu vardır. (Çizge kuramcılarının terminolojisinde böyle bir yol için başka bir terim de kullanılıyor olabilir, bu nedenle bağışlanmayı diliyoruz.)

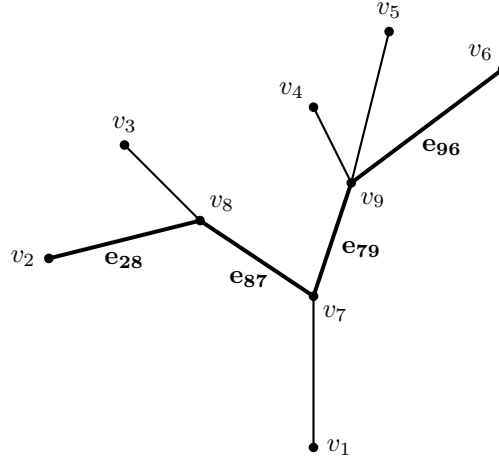
Bir ağacı bir diyagram yardımıyla görselleştirirken, köşeleri temsilen noktalar koyuyoruz ve kenarları temsilen de, bir  $e_{ij}$  kenarı için,  $v_i$  ve  $v_j$  köşelerini temsil eden noktaları (düz veya kavisli) bir çizgi ile birleştiriyoruz. Şekil 2'de 9 köşeli bir ağaç görüyoruz. Örneğin,  $e_{28} e_{87} e_{79} e_{96}$  yolu,  $v_2$  köşesi ile  $v_6$  köşesi arasındaki (tek) yoldur.

Bir ağacın köşelerinin kümesi üzerinde (bu kümeye  $V$  diyelim) doğal bir metrik tanımlanabilir:  $v_i$  ve  $v_j$  gibi iki (farklı) köşe arasındaki uzaklık olarak, bunları birleştiren yolu oluşturan kenarların sayısını atayabiliriz. ( $v_i$  ve  $v_i$  arasındaki uzaklığı sıfır olarak tanımlayabiliriz.) Örneğin Şekil 2'de  $v_2$  ile  $v_6$  arasındaki uzaklık 4'tür. (Bu suretle  $V$  üzerinde bir metrik uzay ortaya çıktığını görmek, küçük, şirin bir alıştırmadır.)

Bu metrik uzayın tek kusuru şu ki, uzaklıklar olarak hep doğal sayılar karşımıza çıkıyor. Bu nedenle, biraz daha hassas işler için,  $V$  köşeler kümesi üzerinde başka bir metrik tanımlamak istiyoruz. Bu amaçla kenarlara, bir tür ağırlık olarak pozitif sayılar atamak istiyoruz; yani, kenarlar kümesine  $E$  diyecek olursak, bir  $w : E \rightarrow \mathbb{R}^+$  fonksiyonunu göz önüne alıyoruz ve bunun yardımıyla  $d^w(v_i, v_j)$  uzaklığını,  $v_i$  ile  $v_j$  köşeleri arasındaki yolu oluşturan kenarların ağırlıklarının toplamı olarak tanımlıyoruz. Böylece  $V$  köşeler kümesi üzerinde bir  $d^w$  metriği tanımlanmış olur.

Ağacımıza şeklen bir  $(V, E)$  ikilisi gözüyle bakarsak,  $(V, E, w)$  üçlüsüne de bir metrik ağaç diyeceğiz. Yani bir metrik ağaç, kenarları pozitif sayılarla ağırlıklandırılmış bir ağaçtır; bu suretle de  $V$  köşeler kümesi üzerinde  $w$  ağırlık fonksiyonu yardımıyla tanımlanmış bir metrik ortaya çıkar.

Aslında birazcık daha ileri gidip, ağacın görselleştirilmesinde kenarları temsil eden çizgilere gerçeklik kazandırabilirdik; o şekilleri resmileştirip ağacı bir topolojik uzaya dönüştürebilir, ağırlık fonksiyonu yardımıyla da bir metrik uzaya dönüştürebilirdik. Literatürde metrik ağaç denilince çoğu kez bu anlaşılıyor. Ama bizim amaçlarımız için, sadece köşeler arasındaki uzaklığın tanımlandığı yukarıdaki basit model yeterli olacak.



Şekil 2

## 2 Sonlu Metrik Uzayları Metrik Ağaçlara Gömmek

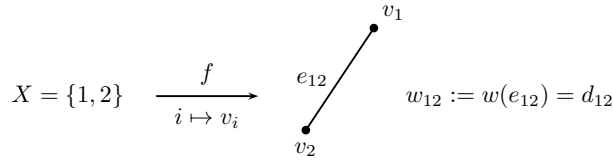
Sonlu sayıdaki organizmadan oluşan bir küme üzerinde, herhangi iki organizma arasında uygun bir uzaklık tanımlayarak bu kümeyi bir metrik uzaya dönüştürürnce, evrim olgusu, bu sonlu metrik uzayı metrik bir ağaca gömme sorusunu kendiliğinden önümüze getiriyor. Bu aynı zamanda matematiksel açıdan da çok ilginç bir soru. Organizmik metrik uzayları bir kenara bırakıp, şu masum soruyu sorabiliriz:

Herhangi bir sonlu  $(X, d)$  metrik uzayı verildiğinde, hangi koşullar altında bu uzayı bir  $(V, E, w)$  metrik ağacına gömebiliriz? Burada gömme derken şüphesiz  $x, y \in X$  için  $d(x, y) = d^w(f(x), f(y))$  olacak şekilde bir  $f : X \rightarrow V$  fonksiyonunun varlığını anlıyoruz.

Filogenetikçiler genellikle bu gömme işleminde  $f(X) \subset V$  görüntü kümesinin  $(V, E, w)$  ağacının yapraklarından oluşan alt küme olmasını istiyorlar (tek bir kenarın ucu olan köşelere yaprak deniyor);  $X$ 'in gömüldüğü ağacın diğer köşelerini evrim çatallanmasının olduğu fosil organizmalar olarak düşünüyorlar. Buna aslında tam katılmıyoruz, çünkü iki organizmanın ortak bir atası hala yaşıyor olabilir, ama gene de çizmeden yukarı çıkmayıp, bu detaylara girmeyeceğiz ve gömme deyince yukarıdaki tanımı esas alacağız.

$N$  noktalı bir metrik uzayı  $X = \{1, 2, \dots, N\}$  kümesi üzerinde düşüneceğiz ve  $i, j \in X$  arasındaki uzaklığı  $d_{ij}$  ile göstereceğiz. Gömme problemini anlamak için önce en minik uzaylara bakmak istiyoruz.

$N = 1$  ve  $N = 2$  durumları aşikârdır;  $X = \{1\}$  uzayını tek köşeli (ve kenarsız) bir ağaca;  $(X = \{1, 2\}, d_{12})$  uzayını da Şekil 3'teki iki köşeli ve tek kenarlı bir metrik ağaca gömebiliriz:



Şekil 3

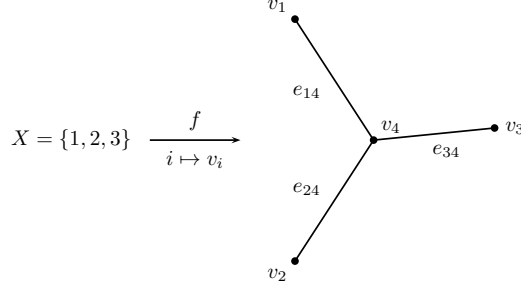
$N = 3$  durumu ilginç sayılabilir.  $(X = \{1, 2, 3\}, d = (d_{ij}))$  metrik uzayını Şekil 4'teki 4 köşeli ve 3 kenarlı

bir metrik ağaca aşağıdaki gibi gömebiliriz:

$$w_{14} := w(e_{14}) = \frac{1}{2}(d_{12} + d_{13} - d_{23}),$$

$$w_{24} = \frac{1}{2}(d_{21} + d_{23} - d_{13}),$$

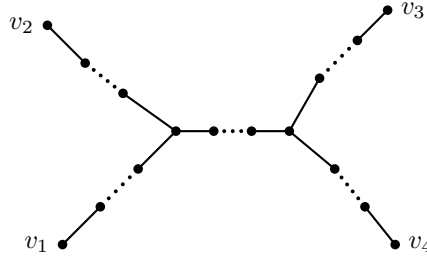
$$w_{34} = \frac{1}{2}(d_{31} + d_{32} - d_{12}),$$



Şekil 4

### 3 Dört-Nokta Koşulu

$N = 4$ 'e gelince işler biraz karışmaya başlıyor! 4 noktalı her metrik uzayı bir metrik ağaca gömemiyoruz. Örneğin  $(X = \{1, 2, 3, 4\}, d_{12} = d_{23} = d_{34} = d_{41} = 1, d_{13} = d_{24} = 2)$  metrik uzayını bir metrik ağaca gömmeye çalışsanız, belli döngülere zorlandığınızı göreceksiniz. Bunun nedenini anlamak için, herhangi bir  $(X = \{1, 2, 3, 4\}, d = (d_{ij}))$  metrik uzayının bir  $(V, E, w)$  metrik ağacına gömülebildiğini varsayalım.  $i \in X$  elemanının ağaçtaki görüntüsü  $v_i \in V$  köşesi olsun. Gömme gereği,  $d^w(v_i, v_j) = d_{ij}$  olur. Şimdi, ağaç içinde,  $i \neq j$  ( $i, j = 1, 2, 3, 4$ ) için  $v_i$  ve  $v_j$  köşeleri arasındaki yolları düşünelim. Karşımıza esas itibariyle şu resim çıkacaktır (sadece köşelerin konumlanma düzeni farklı olabilir):



Şekil 5

Burada şöyle bir durumla karşılaşyoruz:

$$d^w(v_1, v_2) + d^w(v_3, v_4) \leq d^w(v_1, v_3) + d^w(v_2, v_4)$$

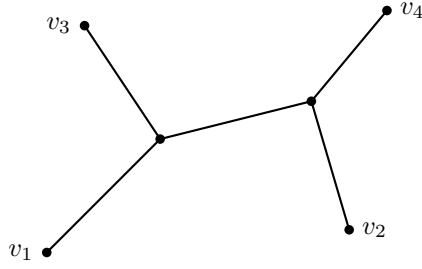
ve

$$d^w(v_1, v_2) + d^w(v_3, v_4) \leq d^w(v_1, v_4) + d^w(v_2, v_3);$$

ya da (sağdaki toplamlar eşit olduğundan)

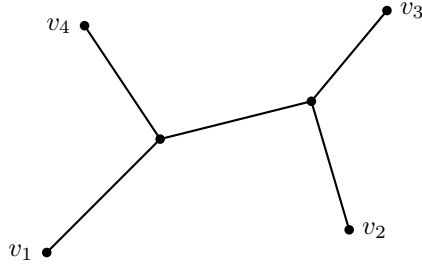
$$d_{12} + d_{34} \leq d_{13} + d_{24} = d_{14} + d_{23}.$$

Köşelerin konumlanma düzeni (şematik olarak) Şekil 6'daki gibi olsa,



Şekil 6

$d_{13} + d_{24} \leq d_{12} + d_{34} = d_{14} + d_{23}$ ; Şekil 7'deki gibi olsa,

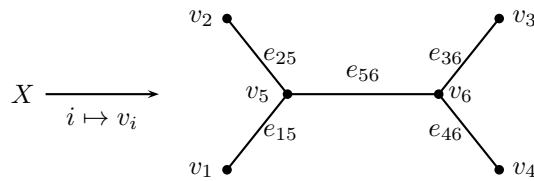


Şekil 7

$d_{14} + d_{23} \leq d_{12} + d_{34} = d_{13} + d_{24}$  ilişkileri geçerli olur.

Yani her durumda,  $d_{12} + d_{34}$ ,  $d_{13} + d_{24}$ ,  $d_{14} + d_{23}$  sayılarının en az ikisi maksimum olmakta, diğeri onlardan eşit-küçük olmaktadır. Buna dört-nokta koşulu diyoruz. Yukarıdaki ( $X = \{1, 2, 3, 4\}$ ;  $d_{12} = d_{23} = d_{34} = d_{41} = 1$ ,  $d_{13} = d_{24} = 2$ ) metrik uzayımıza bakarsak,  $d_{12} + d_{34} = 2$ ,  $d_{13} + d_{24} = 4$  ve  $d_{14} + d_{23} = 2$  olup, bu koşul sağlanmamaktadır! Bu nedenle bu dört noktalı metrik uzay bir metrik ağaca gömülemez.

Diğer yandan, bu koşul sağlanıyorsa, ( $X = \{1, 2, 3, 4\}$ ,  $d = (d_{ij})$ ) metrik uzayını bir metrik ağaca gömebiliriz: Örneğin,  $d_{12} + d_{34} \leq d_{13} + d_{24} = d_{14} + d_{23}$  ise,



Şekil 8

$$\begin{aligned}
w_{15} &:= w(e_{15}) = \frac{1}{2}(d_{12} + d_{14} - d_{24}), \\
w_{25} &= \frac{1}{2}(d_{21} + d_{24} - d_{14}), \\
w_{36} &= \frac{1}{2}(d_{32} + d_{34} - d_{24}), \\
w_{46} &= \frac{1}{2}(d_{41} + d_{43} - d_{13}), \\
w_{56} &= \frac{1}{2}(d_{13} + d_{24} - d_{12} - d_{34})
\end{aligned}$$

gönderimi,  $(X, d)$ 'nin bir metrik ağaca gömmesidir. Durumu şöyle özetleyebiliriz:

**Dört-Nokta Koşulu:**  $(X, d = (d_{ij})_{i,j=1}^4)$  dört noktalı metrik uzayı için dört-nokta koşulunun sağlanması demek,  $d_{12} + d_{34}$ ,  $d_{13} + d_{24}$  ve  $d_{14} + d_{23}$  sayılarından ikisinin eşit olması ve diğerinin de onlardan eşit-küçük olması demektir.

**Yardımcı Teorem.** *Dört noktalı bir  $(X, d = (d_{ij})_{i,j=1}^4)$  metrik uzayının bir metrik ağaca gömülebilmesi için gerek ve yeter koşul, dört-nokta koşuludur.*

Artık dört-nokta teoremini ifade edebiliriz:

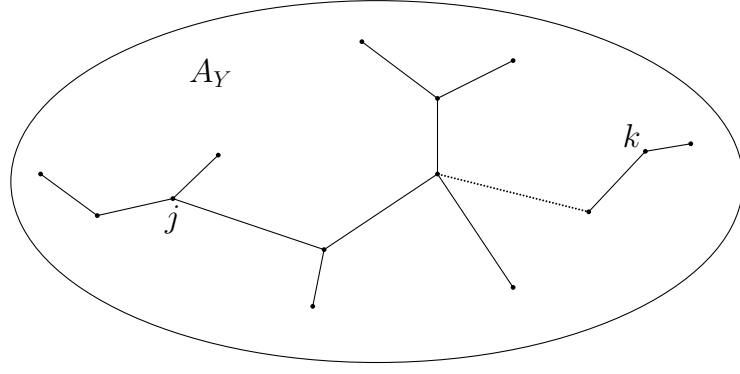
**Teorem (Dört-Nokta Teoremi).** *Sonlu bir  $(X, d)$  metrik uzayının  $(|X| \geq 4)$  bir metrik ağaca gömülebilmesi için gerek ve yeter koşul, bu uzayın dört noktalı her alt-uzayının dört-nokta koşulunu sağlamasıdır.*

Demek ki,  $|X| \geq 5$  için, sonlu bir  $(X, d)$  metrik uzayının bir metrik ağaca gömülebilmesi için gerek ve yeter koşul, onun dört noktalı her alt uzayının bir metrik ağaca gömülebilmesidir. Yani burada bir tür lokalden globale geçiş fenomeni söz konusudur. Teorem, dört-noktalı alt uzayların gömüldüğü birçok "küçük" ağacın uyumlu bir şekilde büyük bir ağacın içine yerleştirilebileceğini söylemektedir.

Son elli yılda en az dört-beş defa "keşfedilmiş" olan bu güzel teoremin literatürdeki ispatları ister-istemez biraz fazla teknik ([1, 2, 4]'e bakınız). Güzel teoremlerin ispatları da güzel olsa ve teoremin neden doğru olduğunu da kavratsa ne güzel olurdu. Biz gene de aşağıda, [4]'teki ispatı biraz değiştirerek ve açarak anlatmaya çalışacağız ([4]'teki yazarlar, bir yaprağın bağlı olduğu kenarın ağırlığının sıfır olmasına izin veriyorlar. Ama o zaman "yaprak" kavramının içi boşaldığı için bu yaklaşım pek hoşumuza gitmedi). İspatın kolay izlenmesi için şekilleri galiba biraz abartıp, çizgi filme dönüştürdük; umarız affolunur.

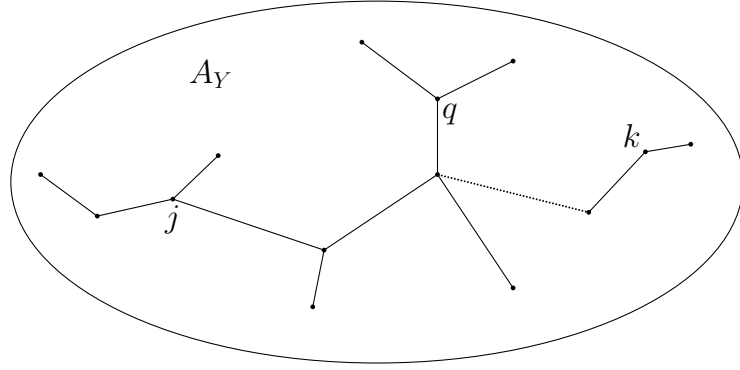
**Dört-Nokta Teoreminin İspatı.**  $(X = \{1, 2, \dots, N\}, d = (d_{ij}))$ ,  $N \geq 4$ , bir metrik uzay olsun ve dört-nokta koşulunu sağlasın. İspatı tümevarımla yapacağız.  $N = 4$  için iddiamız Yardımcı Teorem nedeniyle doğrudur. Şimdi  $N \geq 5$  olsun ve  $N - 1$  noktalı bir metrik uzayın, dört-nokta koşulunu sağlaması halinde, bir metrik ağaca gömülebileceğini kabul edelim. Bu kabulü,  $X$ 'den bir nokta atarak elde edeceğimiz bir alt uzaya uygulayacağız; ama bu noktayı çok usturuplu bir biçimde seçip atmamız gerekiyor!  $X$  içinde,  $d_{ik} + d_{jk} - d_{ij}$  ifadesi maksimum olacak şekilde bir  $(i, j, k)$  üçlüsü ( $i \neq j \neq k \neq i$ ) seçelim ve  $X$ 'den  $i$  noktasını atalım! (Bu da nereden çıktı demeyin; bu aslında Buneman'ın ilhamı [1]. Belki siz daha doğal başka bir seçim yapabilirsiniz ve bu harika olurdu!)

$(Y = X \setminus \{i\}, d_Y = d|_Y)$  metrik uzayı  $N - 1$  noktaya sahip olup, dört-nokta koşulunu da açıkça sağladığından, kabulümüze göre bir metrik ağaca gömülebilir. Bu ağaca  $A_Y$  diyelim. Bu gömmeyi Şekil 9'da temsili olarak görüyorsunuz.  $i$  köşesini henüz dışarıda ve elimizde tutuyoruz. Onu ileride bir kenarla uygun şekilde  $A_Y$  ağacına bağlayıp,  $X$  uzayının gömülebildiği bir  $A_X$  ağacı oluşturacağız.



Şekil 9

Şimdi  $X$  kümesinde,  $i, j$  ve  $k$  dışında herhangi bir  $q$  noktası göz önüne alalım (bkz. Şekil 10).



Şekil 10

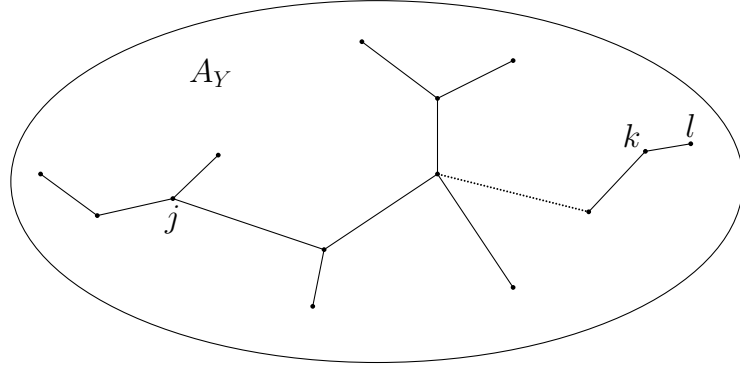
Maksimallik koşulumuz nedeniyle,

$$\begin{aligned}
 d_{qk} + d_{jk} - d_{qj} &\leq d_{ik} + d_{jk} - d_{ij} \\
 &\Rightarrow d_{qk} + d_{ij} \leq d_{ik} + d_{qj} \\
 d_{ik} + d_{qk} - d_{iq} &\leq d_{ik} + d_{jk} - d_{ij} \\
 &\Rightarrow d_{qk} + d_{ij} \leq d_{jk} + d_{iq}
 \end{aligned}$$

eşitsizlikleri geçerli olur.  $X$  içindeki  $i, j, k$  ve  $q$  noktaları dört-nokta koşulunu sağladığından,

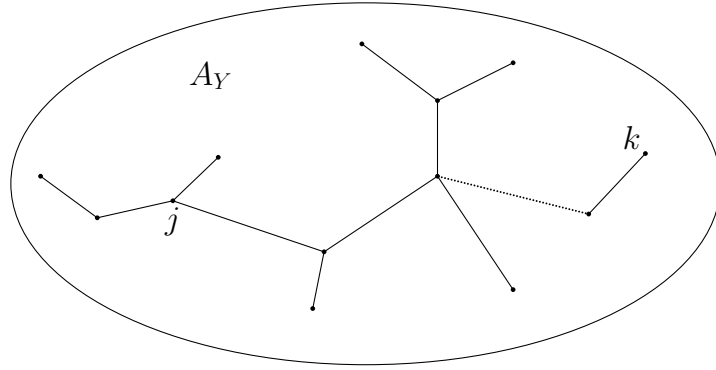
$$d_{ik} + d_{qj} = d_{jk} + d_{iq} \quad (1)$$

elde edilir. Şimdi  $A_Y$  ağacı içinde  $j$  köşesinden  $k$  köşesine giden en kısa yolu düşünelim (gömmelerde bir eleman ile onun görüntüsü için aynı sembolü kullanıyoruz). Bu yolun "devamında"  $k$  köşesinden sonra gelen ve  $Y$  kümesine ait bir köşe olamaz! Bunu görmek için  $k$ 'dan sonra bir  $l \in Y$  olduğunu varsayalım (bkz. Şekil 11).



Şekil 11

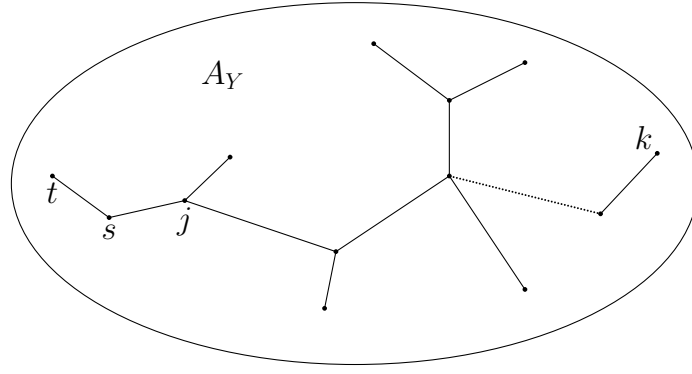
(1)'de  $q = l$  alırsak,  $d_{ik} + d_{jl} = d_{jk} + d_{il}$  eşitliğinden,  $d_{jl} > d_{jk}$  olduğundan,  $d_{il} > d_{ik}$  elde edilir. Bu durumda,  $d_{il} + d_{jl} - d_{ij} > d_{ik} + d_{jk} - d_{ij}$  olur ki, bu da  $d_{ik} + d_{jk} - d_{ij}$  ifadesinin maksimalliği ile çelişir. Demek ki, gömmemizi Şekil 12'deki gibi tasavvur edebiliriz (yani istersek  $A_Y$  ağacının  $jk$  hattı üzerinden ulaşılan kısmını kesip atabiliriz):



Şekil 12

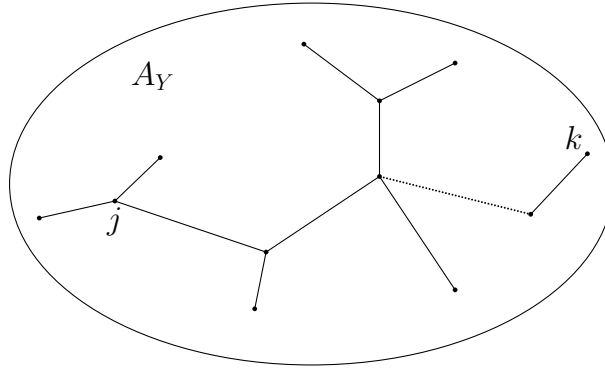
Şimdi biraz da  $jk$  en kısa yolunun  $j$  tarafına bakalım. Bu istikamette  $j$ 'ye iki kenarla bağlanan (Şekil 13'deki  $t$  gibi) ve  $Y$ 'ye ait bir nokta da var olamaz. Çünkü böyle bir durumda  $d_{tk} + d_{sk} - d_{ts} = 2d_{sk} > 2d_{jk}$  olurdu. Oysa  $X$ 'teki üçgen eşitsizliği nedeniyle  $d_{ik} + d_{jk} - d_{ij} \leq 2d_{jk}$  olup, bu durum  $d_{ik} + d_{jk} - d_{ij}$  ifadesinin maksimalliği ile çelişirdi.





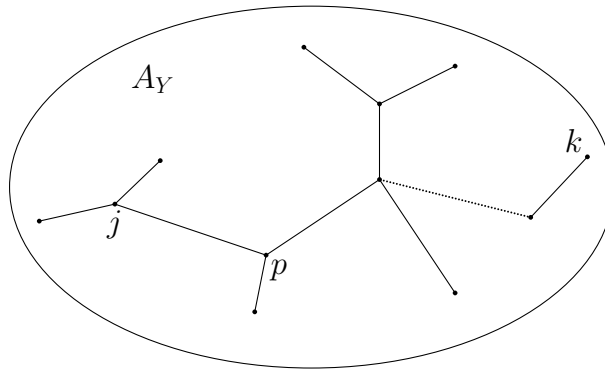
Şekil 13

Demek ki gömmemizi Şekil 14'teki gibi düşünebiliriz.



Şekil 14

Şimdi tekrar  $jk$  en kısa yolunu düşünelim ve bu yol üzerinde  $j$  köşesinden sonra gelen ve  $Y$ 'ye ait olan ilk köşeye  $p$  diyelim (bkz. Şekil 15).



Şekil 15

$X$  uzayının gömülebildiği bir  $A_X$  ağacı oluşturmak için,  $i$  köşesini,  $j$  ile  $p$  arasındaki uygun bir  $r$  noktasma

yeni bir kenarla bağlayacağız. Bu amaçla şu ağırlıkları tanımlayalım:

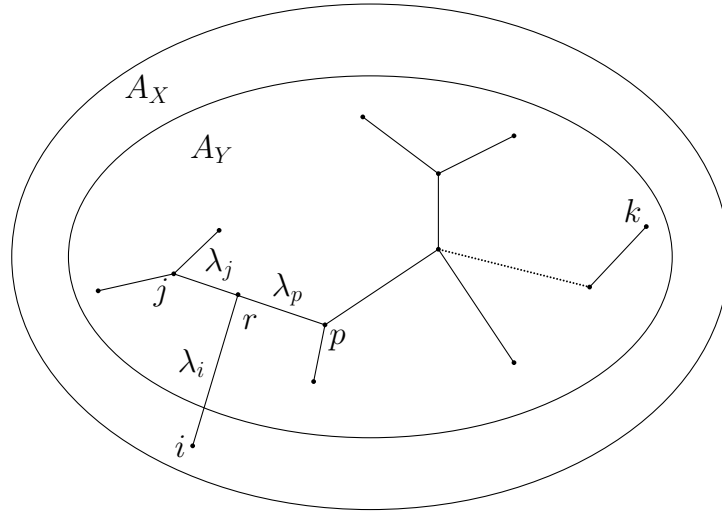
$$\begin{aligned}\lambda_i &= \frac{1}{2}(d_{ij} + d_{ik} - d_{jk}), \\ \lambda_j &= \frac{1}{2}(d_{ij} + d_{jk} - d_{ik}), \\ \lambda_p &= \frac{1}{2}(d_{ip} + d_{jp} - d_{ij}).\end{aligned}$$

Bu ağırlıkların hepsi, üçgen eşitsizliği nedeniyle pozitif veya en az sıfırdır. Ayrıca,  $\lambda_i + \lambda_j = d_{ij}$ ,  $\lambda_i - \lambda_j = d_{ik} - d_{jk}$  olup,  $\lambda_j + \lambda_p = \frac{1}{2}(d_{ip} + d_{jp} + d_{jk} - d_{ik})$  olur. (1) eşitliğinde  $q$  yerine  $p$  alacak olursak,

$$d_{jk} + d_{ip} = d_{ik} + d_{jp}$$

olup, buradan  $\lambda_j + \lambda_p = d_{jp}$  elde edilir. (Bir özel durum olarak,  $p = k$  olsa,  $d_{jk} + d_{ip} = d_{ik} + d_{jp}$  eşitliği şüphesiz yine geçerlidir.)

Şimdi Şekil 16'da belirtildiği gibi,  $A_Y$  ağacına  $i$  ve  $r$  köşelerini ve  $\lambda_i$  ağırlıklı  $e_{ir}$  kenarını ekleyerek (ve mevcut kenarı da belirtildiği gibi ikiye bölerek) bir  $A_X$  ağacı oluşturalım.



Şekil 16

$X$  uzayının  $i$  dışındaki noktaları (yani  $Y$  uzayı), izometrik olarak  $A_Y$  ağacında yerlerini almışlardır.  $i$  noktası da  $A_X$  ağacında yerini almış bulunuyor. Bu gönderimin izometrik olduğunu görmek için artık sadece  $i$  noktasının başka bir  $q$  noktasına  $A_X$  ağacı içindeki  $d_{iq}^{A_X}$  uzaklığının  $d_{iq}$  olduğunu göstermek yeterli olacaktır.

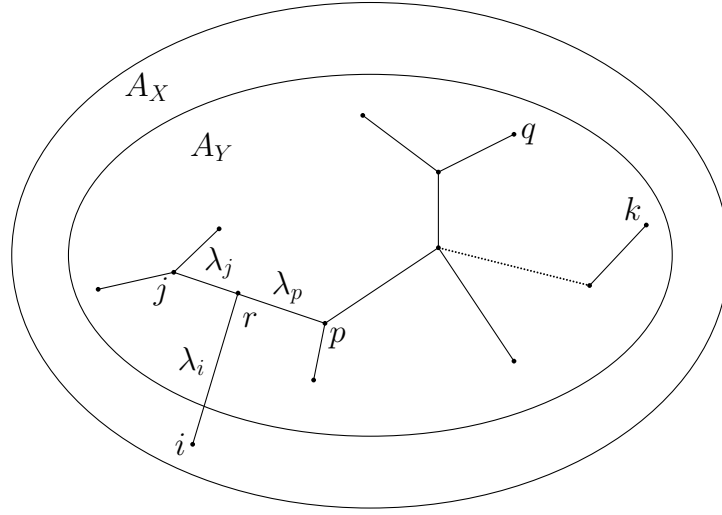
$q = j$  için  $d_{ij}^{A_X} = \lambda_i + \lambda_j$  olup, bunun  $d_{ij}$ 'ye eşit olduğuna yukarıda işaret etmiştik.  $q = k$  için,

$$d_{ik}^{A_X} = \lambda_i + (d_{jk} - \lambda_j) = (\lambda_i - \lambda_j) + d_{jk} = d_{ik} - d_{jk} + d_{jk} = d_{ik}$$

olur. Demek ki,  $d_{ij}$  ve  $d_{ik}$  uzaklıkları  $A_X$  ağacı içinde korunmaktadır.

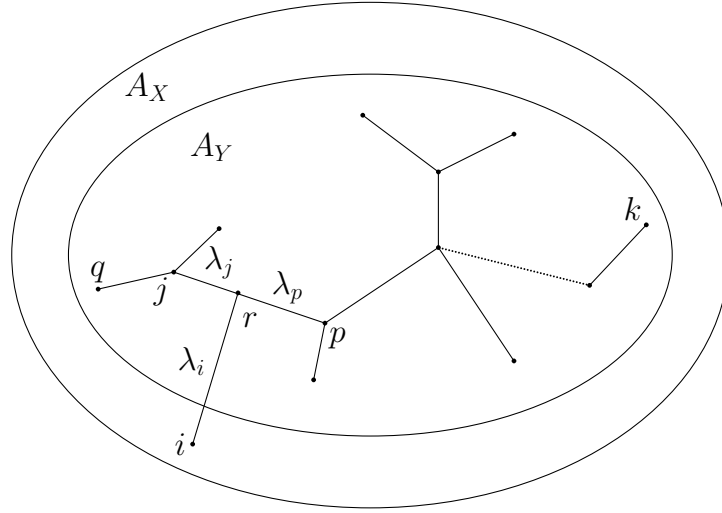
Şimdi  $q$ ,  $A_Y$  ağacı içinde,  $Y$ 'den gelen ve  $j$  ile  $k$ 'dan farklı, herhangi bir köşe olsun.  $q$  için iki durum sözkonusu olabilir:  $A_X$  ağacı içinde  $i$  köşesinden  $q$  köşesine ulaşırken  $r$  köşesinden geçmemiz gerekir ve bundan sonra  $k$  veya  $j$  istikametinde ilerleyebiliriz.

İlk durumda (bkz. Şekil 17)  $d_{iq}^{A_X} = \lambda_i + (d_{jq} - \lambda_j) = d_{jq} + (\lambda_i - \lambda_j) = d_{jq} + d_{ik} - d_{jk}$  olup, (1) nedeniyle  $d_{iq}^{A_X} = d_{iq}$  bulunur. Yani bu durumda da  $A_X$  ağacına gömme esnasında  $d_{iq}$  uzaklığı korunur.



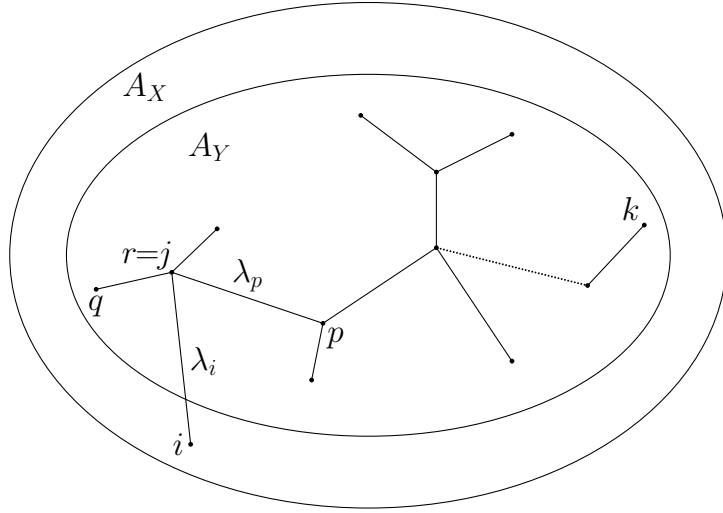
Şekil 17

Şimdi ikinci durumu ele alalım, yani  $q$  köşesi  $i$  köşesinden hareketle,  $r$  geçildiğinde,  $j$  istikametinde kalsın (bkz. Şekil 18).



Şekil 18

Bu durumda önce  $d_{ik} = d_{ij} + d_{jk}$  olacağını, dolayısıyla  $\lambda_i = d_{ij}$  ve  $\lambda_j = 0$  olup,  $r = j$  olacağını göstermek istiyoruz (bkz. Şekil 19):



Şekil 19

$q, j$  ve  $k$  köşeleri bir yol üzerinde olduğundan

$$d_{qk} = d_{qj} + d_{jk} \quad (2)$$

yazabiliriz. O halde  $d_{jk} + d_{qk} - d_{qj} = 2d_{jk}$  olur. Diğer yandan, üçgen eşitsizliği nedeniyle  $d_{ik} \leq d_{ij} + d_{jk}$  olur. Bunu da,  $d_{ik} + d_{jk} - d_{ij} \leq 2d_{jk}$  şeklinde yazalım. Maksimallik varsayımımız nedeniyle,

$$2d_{jk} = d_{jk} + d_{qk} - d_{qj} \leq d_{ik} + d_{kj} - d_{ij} \leq 2d_{jk}$$

olup, buradan  $d_{jk} + d_{qk} - d_{qj} = d_{ik} + d_{kj} - d_{ij}$  yani  $d_{qk} - d_{qj} = d_{ik} - d_{ij}$  ve (2) nedeniyle  $d_{jk} = d_{ik} - d_{ij}$ , yani  $d_{ij} + d_{jk} = d_{ik}$  elde edilir. Dolayısıyla,  $\lambda_i = d_{ij}$  ve  $\lambda_j = 0$  olmaktadır. Demek ki  $j$  tarafında tek bir  $q$  dahi varsa,  $r = j$  olmak zorundadır (bkz. Şekil 19).

Artık  $i$  ve  $q$  noktalarının  $A_X$  ağacı içindeki uzaklığını hesaplayabiliriz:

$$d_{iq}^{A_X} = \lambda_i + d_{jq}^{A_Y} = d_{ij} + d_{jq}.$$

Peki,  $d_{ij} + d_{jq} = d_{iq}$  diyebilir miyiz? Bunu görmek için,  $(i, k, q)$  üçlüsünü düşünelim. Maksimallik varsayımı nedeniyle,

$$\begin{aligned} d_{ik} + d_{kq} - d_{iq} &\leq d_{ik} + d_{kj} - d_{ij}, \\ d_{kq} + d_{ij} &\leq d_{kj} + d_{iq}, \end{aligned}$$

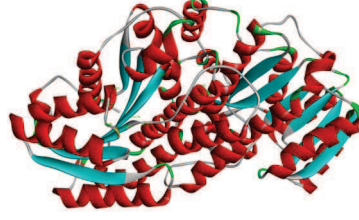
ve (2) nedeniyle,

$$\begin{aligned} d_{qj} + d_{jk} + d_{ij} &\leq d_{kj} + d_{iq}, \\ d_{qj} + d_{ij} &\leq d_{iq}, \end{aligned}$$

yazılabilir. Üçgen eşitsizliği nedeniyle  $d_{iq} \leq d_{ij} + d_{jq}$  olduğundan, buradan  $d_{ij} + d_{jq} = d_{iq}$  elde edilir. Demek ki gerçekten  $d_{iq}^{A_X} = d_{iq}$  olup, bu durumda da uzaklık korunmaktadır. Böylece, küçük maratonumuzu tamamlamış olduk.  $\square$

Bu teorem matematiksel olarak tabii çok zarif ve biz aynı zarafette bir ispat verememiş olsak da, bu teoreme hayranız. Alıştırma olarak şüphesiz kolayca basit (ama sun'i) örnekler verebilirdik, ama biz gerçek bir örnek vermek istedik ve bu münasebetle hakiki uygulamaların ne kadar zor ve problemlili olabileceğini yaşayarak gördük. Eskişehir Teknik Üniversitesi Biyoloji Bölümünde moleküler biyoloji konusunda çalışan bir arkadaşımızdan (Burak Berber), bize bir enzimin 5-6 farklı organizmadaki amino asit dizilimlerini verip

veremeyeceğini sorduk. Burak bize Şekil 20'de gördüğümüz enzimin, fare, tavşan, insan, koyun ve domuz organizmalarındaki dizilimlerini çıkardı.



Şekil 20:  $11\beta$  - *HSD1* enziminin 3 boyutlu yapısı.

Aynı "harf"lerden oluşturulan (bizim durumumuzda "harf"ler amino asitler oluyor ve bu "alfabe"nin eleman sayısı 20) ve aynı uzunluğa sahip (bizim durumumuzda 292) iki sonlu dizi (veya "kelime" ya da biyologların ifadesiyle "dizilim" ya da "sekans") verildiğinde, bunlar arasındaki "Hamming" adı verilen uzaklık, farklı harflere sahip pozisyonların sayısı olarak tanımlanıyor. Örneğin 10'ar harf uzunluğundaki

*MAVMKKNYLLP*  
*MAFMKKYLLP*

dizilimleri arasındaki uzaklık 2 dir, çünkü bu dizilimlerde 3. ve 6. pozisyonadaki harfler farklıdır (ama diğer konumlardaki harfler aynıdır). Bu kelimelerin harfleri, şu noktada bizi fazla ilgilendirmeyen bazı amino asitlerin sembolleridir.

Biz üşenmeden, hatta epeyce bir heyecanla, Burak'tan aldığımız 5 dizilim arasındaki Hamming uzaklıklarını bulduk. Her biri 292 harf uzunluğunda olan 5 dizilimi Tablo 2'de ve bunlar arasındaki uzaklıkları Tablo 1'de görebilirsiniz. Böylece {fare, tavşan, insan, koyun, domuz} kümesini bir metrik uzaya dönüştürmüş olduk! (Küçük, güzel bir alıştırma olarak, Hamming uzaklığının, belli bir alfabe üzerinde, belli bir uzunluktaki bazı kelimelerden oluşan bir küme üzerinde bir metrik tanımladığını gösterebilirsiniz.)

	Fare	Tavşan	İnsan	Koyun	Domuz
Fare	0	72	63	70	63
Tavşan	72	0	57	75	66
İnsan	63	57	0	64	49
Koyun	70	75	64	0	43
Domuz	63	66	49	43	0

Tablo 1

5 elemanlı bir organizmalar kümesi üzerindeki bu uzaklıklar tablosunu daha yakından incelediğimizde büyük bir hayal kırıklığı yaşadık. Bu metrik dört-nokta koşulunu sağlamıyordu! Örneğin daha ilk baktığımız {fare = 1, tavşan = 2, insan = 3, koyun = 4} dörtlüsünde,

$$d_{12} + d_{34} = 72 + 64 = 136$$

$$d_{13} + d_{24} = 63 + 75 = 138$$

$$d_{14} + d_{23} = 70 + 57 = 127$$

olup, bu uzaklıklar dört-nokta koşulunu sağlamıyordu!

Belki bir uygulamacı, "Hemen hemen sağlanıyor" da diyebilir ("136, 138 ve 127 sayılarının büyük olan ikisi hemen hemen eşit" diyebilir). Ama matematiğin masum ve berrak (ama belki biraz mızımız) dünyasına alışmış bizler şüphesiz bunu kabul edemeyiz. Bu örnek matematiğin basit dünyasıyla, gerçek uygulamanın karmaşık dünyası arasındaki farkı göstermesi açısından bizim için çok öğretici oldu. Biraz düşününce şunu kabul etmek zorunda kaldık ki, organizmalar arasında anlamlı bir evrimsel uzaklık tanımlamak gibi zor bir mesele bu kadar hafife alınmamalıydı. Milyarlarca harflik nükleotid dizilerinin söz konusu olduğu bu organizmalardan, birkaç yüz harflik bir aminoasit dizisine bakmak, tabii ki hassas bir uzaklık ölçüsü veremiyordu...

Diğer yandan çok daha uzun dizilimleri bile alsak, acaba Hamming uzaklığı doğru bir seçim miydi? Belki bazı pozisyonlardaki harf farklılıkları, diğer bazılarına göre çok daha önemli olabilirdi. Burada artık yavaş yavaş "Matematiksel Filogenetik" gibi yeni bir bilim dalının nasıl doğmaya başladığını görüyoruz. Bu konularda bir merak uyandırabilirdik, bu küçük yazımız amacına ulaşmış olacak.

Dört-nokta teoremine geri dönersek, özellikle verdiğimiz ispatı örneklemek için, yapay da olsa bir örnek vermekten gene de kendimizi alamadık. Kafadan bir örnek atabilirdik, ama bu konuda doktora yapmış İsveçli Alice Lesser'in tezindeki bir örneği görmenizi istiyoruz [5]. Tablo 3'teki bu örneğin kafadan atılmış olduğu besbelli. İhtimal Alice de gerçek örneklerle kolay baş edilemeyeceğini görüp bu sayıları dört-nokta koşulunu sağlayacak şekilde uydurmuş olmalı.

	İnsan	Goril	Köpek	Kedi	Böcek
İnsan	0	3	10	10	14
Goril	3	0	9	9	13
Köpek	10	9	0	2	10
Kedi	10	9	2	0	10
Böcek	14	13	10	10	0

Tablo 3

Alice'in tablosunun beş noktalı bir metrik uzay tanımladığını ve bu uzayın dört noktalı her alt kümesinin dört-nokta koşulunu sağladığını tabi ki hemen kontrol edebilirsiniz. Örneğin; {insan, goril, köpek, kedi} alt kümesi üzerinde

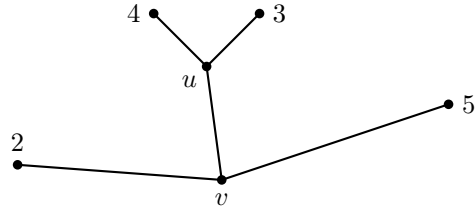
$$\begin{aligned}
d_{\text{insan, goril}} + d_{\text{köpek, kedi}} &= 3 + 2 = 5 \\
d_{\text{insan, köpek}} + d_{\text{goril, kedi}} &= 10 + 9 = 19 \\
d_{\text{insan, kedi}} + d_{\text{goril, köpek}} &= 10 + 9 = 19
\end{aligned}$$

olup, bu alt küme dört-nokta koşulunu sağlamaktadır.

Dört-nokta teoremine göre, bu beş noktalı metrik uzayın bir metrik ağaca gömülebilmesi gerekiyor. Yazım kolaylığı açısından sırasıyla insan, goril, köpek, kedi, böcek yerine 1, 2, 3, 4 ve 5 yazalım ve  $X = \{1, 2, 3, 4, 5\}$  üzerindeki  $(X, d)$  metrik uzayını düşünelim. Bu metrik uzayı gömebileceğimiz bir ağaç bulmak için, ispatımıza göre,  $X$  içinde  $d_{ik} + d_{jk} - d_{ij}$  ifadesini maksimum yapan  $(i, j, k)$ ,  $(i \neq j \neq k \neq i)$  üçlülerini bulalım. Bunlar,  $(1, 2, 5)$  ve  $(2, 1, 5)$  üçlüleridir (Bunlar için  $d_{ik} + d_{jk} - d_{ij}$  değeri 24'tür).  $(1, 2, 5)$  üçlüsünü seçelim ve  $i = 1$  olmak üzere  $(Y = X \setminus \{1\}, d|_Y)$  metrik uzayını bir  $A_Y$  ağacına gömelim. Bu ağaç

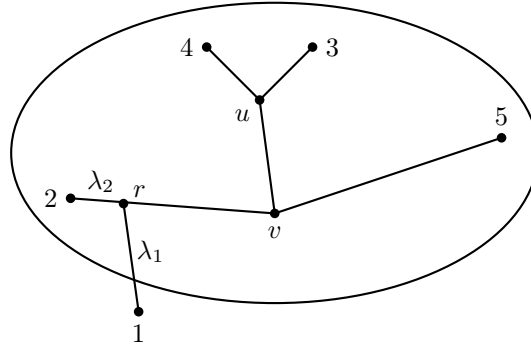
$$\begin{aligned}
w_{3u} &= \frac{1}{2}(d_{34} + d_{35} - d_{45}) = 1, \\
w_{4u} &= \frac{1}{2}(d_{34} + d_{42} - d_{23}) = 1, \\
w_{5v} &= \frac{1}{2}(d_{52} + d_{54} - d_{24}) = 7, \\
w_{2v} &= \frac{1}{2}(d_{25} + d_{24} - d_{45}) = 6, \\
w_{uv} &= \frac{1}{2}(d_{23} + d_{45} - d_{34} - d_{25}) = 2
\end{aligned}$$

olmak üzere, Şekil 21'deki gibi olacaktır.



Şekil 21

$j = 2$  köşesinden  $k = 5$  köşesine giden en kısa yol üzerindeki  $Y$ 'ye ait ilk köşe  $p = k = 5$  olacaktır (Bu yol üzerindeki  $v$  köşesi  $Y$ 'ye ait değildir). Şimdi de  $i = 1$  köşesini  $j = 2$  ile  $p = k = 5$  arasındaki uygun bir  $r$  noktasma yeni bir kenar ile bağlamalıyız.



Şekil 22

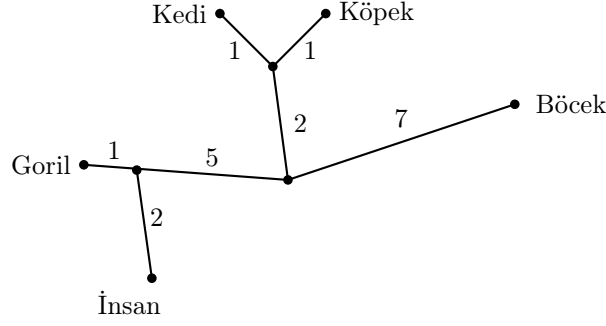
İspatımızda tanımlanan yardımcı ağırlıklar

$$\lambda_1 = \frac{1}{2}(d_{12} + d_{15} - d_{25}) = 2,$$

$$\lambda_2 = \frac{1}{2}(d_{12} + d_{25} - d_{15}) = 1,$$

$$\lambda_5 = \frac{1}{2}(d_{15} + d_{25} - d_{12}) = 12$$

olup, bu durumda elde edilen ağaç Şekil 22'deki gibi olacaktır. Bu ağacı asıl kahramanlarımızla çizersek Şekil 23'teki ağacı elde ederiz.



Şekil 23

Gerçekten de bu ağaç üzerindeki en kısa yol metriği Tablo 3'ü bize geri verecektir.





















Matematikselsel Filogenetiği sevdiyseniz, şu kaynaklara da bir göz atabilirsiniz: [6, 7, 8].

Yazımızı bitirirken, gerçek enzim dizilimleri diye tutturup epey uğraştırdığımız biyolog arkadaşımız Burak Berber'e de çok teşekkür etmek istiyoruz.

## Kaynakça

- [1] P. Buneman, *A note on the metric properties of trees*, Journal of Combinatorial Theory, Series B, 17(1),48-50, (1974).
- [2] J. M. S. Simões Pereira, *A note on the tree realizability of a distance matrix*, Journal of Combinatorial Theory, 6(3), 303-310, (1969).
- [3] E. Haeckel, *Vahdet-i mevcûd; Bir tabiat âliminin dini*, Çev. Baha Tevfik, Ahmet Nebil, Çizgi Kitabevi Yayınları, (2014).
- [4] D. Maclagan, B. Sturmfels, *Introduction to Tropical Geometry*, American Mathematical Society, Graduate Studies in Mathematics Vol: 161, (2015).
- [5] A. Lesser, *Optimal and Hereditarily Optimal Realizations of Metric Spaces*, PhD dissertation, Matematiska institutionen, Uppsala, (2007).
- [6] E. S. Allman, J. A. Rhodes, *Mathematical Models in Biology: An Introduction*, Cambridge University Press, (2003).
- [7] C. Semple, M. Steel, *Phylogenetics*, Oxford Lecture Series in Mathematics and Its Applications, (2003).
- [8] A. Dress, K. T. Huber, J. Koolen, V. Moulton, A. Spillner, *Basic Phylogenetic Combinatorics*, Cambridge University Press, (2011).



 MAVMKNYLLPILVLFLLAYYYYSTNEEFRPEMLQGGKKVIVTTGASKGIGREMAHYHLSKMGAHVVLTARSEEG LKK	 MAFMKKYLLPLGLFLMAYYYYSANEFFRPEMLQGGKKVIVTTGASKGIGREMAHYHLSKMGAHVVLTARSKETLQE	 MAFMKKYLLPILGLFLMAYYYYSANEFFRPEMLQGGKKVIVTTGASKGIGREMAHYHLSKMGAHVVLTARSKETLQK	 MAFMKKYLLPILGLFLAYYYYSANEFFRPEMLRGGKRVIVTTGASKGIGREMAHYHLSKMGAHVVLTARSEESLKKK	 MAFMKKYLLPILGLFLAYYYYSANEFFRPEMLRGGKRVIVTTGASKGIGREMAHYHLSKMGAHVVLTARSEETLKKK	 VVSRCLELGAASAHYIAGTME DM TFAEEQFIVKAGKLMGGGLDMLILNHITQTSLSLFFHDDIHSVRRVMEVNF LLS	 VVAHCLKLGAASAHYIAGTME DM TFAEEQFVAKAGKLMGGGLDMLILNHITNASL M F F N D I H H V R K E M E V N F L S	 VVSHCLELGAASAHYIAGTME DM TFAEEQFVAKAGKLMGGGLDMLILNHITNTSLNLFHDDIHVRKSM EVNF LLS	 VVSRCLELGAASAHYVAGTME DM TFAEEQFVAKAGELVGGGLDMLILNHINYTPLRVFN D I H L L R R S L E V N L L S	 VVSHCLELGAASAHYVAGTME DM TFAEEQFVAKAGKLLGGGLDMLILNHITHASMT P F S D D I H L V R R S M E V N F L S	 YVVMSTAAPMLKQSN G S I A V I S L A G K M T Q P M I A P Y S A S K F A L D G F F S T I R T E L Y I T K V N V S I T L C V L G L I D	 YVVLTVAAPMLKQSN G S I V V S S L A G K I A H P L I A P Y S A S K F A L D G F F S A I R K E H A L T N V N V S I T L C V L G L I D	 YVVLTVAAPMLKQSN G S I V V S S L A G K V A Y P M V A A Y S A S K F A L D G F F S S I R K E Y S V R V N V S I T L C V L G L I D	 YVVLSTAAPMLKQTS G S I V V S S V A G K I A C P L A A A Y S A S K F A L D G F F S S L R T E Y E A T K V N V S I T L C I L G L I D	 YVVLVAALPMLKQSN G S I V V S S Q A G K M A N P L V A P Y S A S K F A L D G F F S S I R K E Y S V T K V N V S I T L C I L G L I D	 TETAMKEISGIINAQASPK E E C A L E I I K G T A L R K S E V Y Y D K S P L T P I L L G N P G R K I M E F F S L R Y N K D M F V S N	 TDTAMKEVSGKIDMKAAPK E E C A L E I I K G G A L R Q D E V Y Y G N L Q W T P L L L G N P G K R L I E F L H L R K F D I S K L V N N	 TETAMKAVSGIVHMQAAPK E E C A L E I I K G G A L R Q E E V Y Y D S S L W T T L L I R N P C R K I L E F L Y S T S Y N M D R F I N K	 TDTAMKAVAGIYNAAEASPK E E C A L E I I K G G A L R Q D E V Y Y D N S I L T S L L L K N P G R K I M E F L S L K K Y N M E R F I N N	 TDTAMTAVAGILNVQPSPK E E C A L E I I K G A A L R Q E E V Y Y D S S I L T S L L L G N P G R K I L E F L S L R H Y N M E R F T N N
--	--	---	---	---	--	---	---	---	--	---	---	---	---	---	---	---	---	--	---

Tablo 2:  $11\beta$  - *HSD1* enziminin farklı türlere ait aminoasit dizilimleri