

## INF224 Veri Yapısı ve Algoritmalar Dönem Sonu Projesi

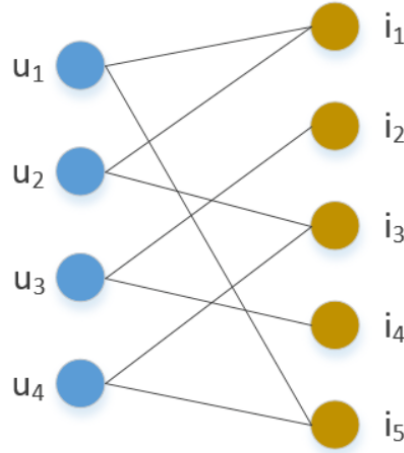
### Proje Tanıtımı

Bu projede öğrenciler bipartite graph yapısını C veya C++ programlama dillerinden biri ile modelleyecekler. Ardından çeşitli algoritmalarından faydalanarak bu veri üzerinde veri keşfi gerçekleştirecek ve basit öneri sistemleri tasarlayacaklar.

### Bipartite Graph Nedir?

Bipartite graph, düğümlerin iki ayrı kümeye ayrıldığı ve kenarların sadece farklı kümelerdeki düğümler arasında bulunduğu özel bir graph türüdür. Bu projede:

- İlk küme kullanıcıları (users)
- İkinci küme ürünleri (items) temsil edecektir.

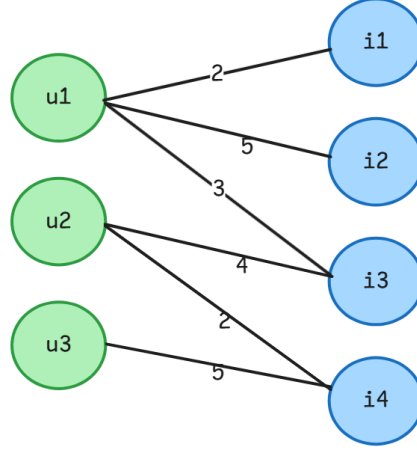


Bu veri yapısının kısıtları şunlardır:

- Yalnızca iki tip düğüm içerecek.
- Aynı tip düğümler arasında bağlanma olmayacak.
- Farklı tip düğümler arasındaki bağlantılar yönsüz olacak.

Veri yapısındaki bazı tanımlamalar aşağıdaki gibidir:

- Bir düğümün derecesi o düğümün toplamda bağlı olduğu düğüm sayısını ifade eder.
- $u$  ve  $i$  şeklindeki iki düğüm arasındaki yol uzunluğu  $u$ 'dan başlayıp  $i$ 'de biten bağlantı serisinin uzunluğuna eşittir.
- $u$  ve  $i$  düğümleri arasındaki en kısa yol uzunluğuna jeodezik mesafe adı verilir.
- $u$  ve  $i$  şeklindeki iki düğüm arasındaki ağırlıklı yol uzunluğu,  $u$  düğümünden başlayıp  $i$  düğümünde biten bağlantı serisinin ağırlıklarının toplamına eşittir. Yani, yol boyunca geçen kenarların ağırlıkları toplanarak bu yolun ağırlıklı uzunluğu hesaplanır.



Örnek Kullanıcı - ürün uzaklık tablosu

	i1	i2	i3	i4
u1	2	5	3	9
u2	9	12	4	2
u3	16	19	11	5

### Veri Dosyası Formatı

interactions.txt: Her satırda bir user-item etkileşimi içerir.

- Format: userId itemId rating
- Her satır bir kullanıcının bir ürünle etkileşimini ve değerlendirme puanını gösterir
- Rating değeri 1-5 arasında bir sayı

Projeyi geliştirirken MovieLens-100K verisetini kullananınız  
(<https://grouplens.org/datasets/movielens/>).

### Projenin Desteklemesi Gereken Temel Öneri Sistemi Yaklaşımları

1. Hiçbir kriteri önceliklemeyen kullanıcının daha önce izlemediği filmler arasından tamamen rassal bir film öneri algoritması.
2. Kullanıcının daha önce izlemediği filmler arasından derecesi en yüksek olan filmleri öneren bir algoritma.
3. Hedef kullanıcının en fazla ortak film izlediği kullanıcı ile benzerlik kurarak, bu kullanıcının en çok beğendiği filmleri öneren bir algoritma.
4. Hedef kullanıcıya en yakın ağırlıklı uzaklığa sahip filmleri öneren bir algoritma.
5. Öğrenci kendi geliştirdiği ek bir öneri algoritmasını da gerçekleştirmelidir.

**Projenin Notlandırılması**

Yukarıda tanımı ve özellikleri verilen yapıyı C ya da C++ dilinde tasarlamanız ve kodlamanız gerekmektedir. Bunun yanında verilen öneri yaklaşımlarının hepsini de yine kodlayarak oluşturmanız gerekmektedir. **Kodların özgünlüğü, çalışabilirliği, performansı ve tasarımı notlandırma sırasında göz önünde bulundurulacaktır.** Projenin notlamasında

1. Proje raporu %20
2. Proje kodları %20
3. Proje sözlüsü sırasında kodlarınızı anlatabilmeniz ve sorulan sorulara cevap verebilmeniz %60

etkili olacaktır.

**Bu dosyada verilen direktifler dışında**

1. Eğer jeodezik mesafe yerine rassal yürüyüş mesafesi hesabı da yaparsanız,
2. Çıkardığınız önerilerin başarı performans ölçümünü de projenize dahil edip farklı yaklaşımları hem model başarısı hem zaman hem de mekan karmaşıklığı açısından detaylı biçimde değerlendirir ve raporlarsanız

Proje notunuz yüksek olacaktır. Bu iki madde için gelecek puanları projenin bonus puanları olarak değerlendireceğiz.