Sosyal Ağ Analizi

Yapay Zekâ ve Veri Dersleri Rehberi · Kısım IV: Veri

Ağ ve çizge | Merkezilik | Topluluk bulma | Ağ modelleri | Yayılım | Komşu dersler

Sercan KÜLCÜ · Giresun Üniversitesi · Bilgisayar Mühendisliği

© Sercan Külcü. Tüm hakları saklıdır. · Sunumun etkileşimli HTML sürümü: https://sercankulcu.github.io/files/ai_courses/sunumlar/Bolum_30_Sosyal_Ag_Analizi.html

GİRİŞ

Ders Hangi Soruyu Sorar?

Kişiler ve ilişkilerden oluşan ağlar nasıl çözümlenir?

Sosyal ağ analizi kişileri düğüm, ilişkileri bağlantı sayarak oluşan **çizgeyi** inceler. Soru tek tek kişilerin özelliği değil, ilişkilerin yapısıdır: kim merkezde, hangi gruplar var, bir söylenti ya da hastalık ağda nasıl yayılır.

Bir bakışta

YAKLAŞIM çizge kuramı ve istatistik	**GİRDİ** düğümler ve bağlantılar	**ÇIKTI** merkezilik, topluluk, yayılım
BAŞARI ÖLÇÜTÜ modülerlik, yorumlanabilirlik	**MATEMATİK YÜKÜ** orta: çizge, matris	**VERİ İHTİYACI** ağ verisi

GİRİŞ

Kısa Tarihçe

1736
Euler, Königsberg köprüleri problemini çözer; çizge kuramı doğar.

1934
Moreno, ilişkileri sosyogram adını verdiği çizimlerle gösterir.

1967
Milgram'ın mektup deneyi "altı adım" düşüncesini yayar.

1973
Granovetter "zayıf bağların gücü"nü gösterir: iş fırsatları uzak tanıdıklardan gelir.

1998
Watts ve Strogatz küçük dünya ağ modelini yayımlar.

2008
Louvain yöntemi, milyonlarca düğümlü ağlarda hızlı topluluk bulmayı mümkün kılar.

DERS

Temel Kavramlar

Düğüm ve bağlantı
Yönlü ya da yönsüz, ağırlıklı ya da ağırlıksız.

Derece dağılımı
Az sayıda çok bağlı düğüm, çok sayıda az bağlı.

Merkezilik
Derece, arasındalık, yakınlık ve özvektör ölçüleri.

Kümelenme katsayısı
Arkadaşlarımın birbirini tanıma oranı.

Küçük dünya
Yüksek kümelenme ama kısa ortalama yol.

Topluluk
İçeride sık, dışarıyla seyrek bağlı düğüm grubu.

Modülerlik
Bir topluluk bölünmesinin ne kadar iyi olduğu.

Yayılım modelleri
Bilgi ya da salgının ağda yayılması (SIR).

DERS

Tipik İş Akışı



Örnek: Bir araştırmacı bir okulun arkadaşlık anketinden ağı kurar, derece dağılımına bakar, arasındalığı yüksek "köprü" öğrencileri bulur, sınıfları aşan arkadaş gruplarını Louvain ile çıkarır ve sonucu çizerek öğretmenlerle tartışır.

DERS

Derste Neler İşlenir?

1. Çizge temelleri: düğüm, bağlantı, komşuluk matrisi
2. Ağ verisi toplama: anket, iz verisi, ego ağları
3. Derece dağılımı, yol uzunluğu ve çap
4. Merkezilik: derece, arasındalık, yakınlık
5. Özvektör merkeziliği ve PageRank
6. Kümelenme, üçlü kapanma ve zayıf bağlar
7. Homofili ve derece uyumu
8. Rastgele çizgeler ve küçük dünya modeli
9. Ölçekten bağımsız ağlar ve tercihli bağlanma
10. Topluluk bulma: modülerlik ve Louvain
11. Yayılım: SIR modeli ve bilgi yayılması
12. Bağlantı tahmini ve ağ görselleştirme

Kazandırdığı beceriler

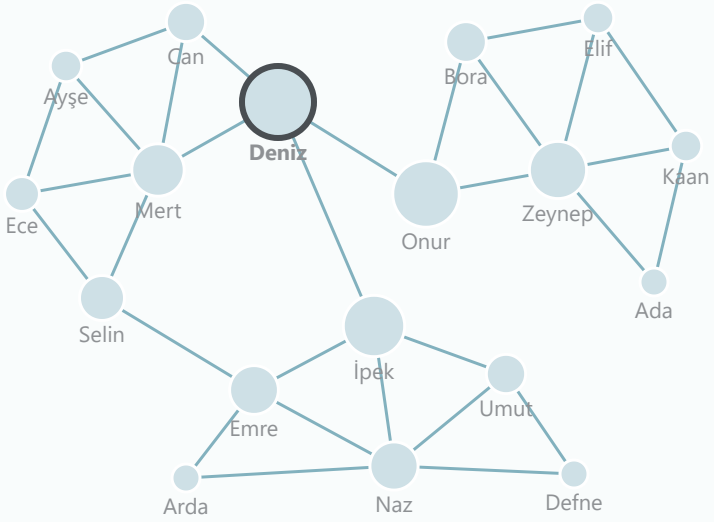
NetworkX ile ağ çözümlemek · Gephi ile ağ görselleştirmek · merkezilik ölçülerini yorumlamak · topluluk bulmak ve sınamak · yayılım benzetimi yazmak

Tipik bir dönemin akışıdır; sıra ve kapsam üniversiteden üniversiteye değişir.

DENE

Kampanya Kimden Başlasın?

Bir kulüp, 18 üyesinin arkadaşlık ağında duyurusunu bir kişiye verecek. Kişiye tıklayıp merkeziliğine bakın, toplulukları bulun, sonra haberin o kişiden nasıl yayıldığını adım adım izleyin.



Boyut: Derece **Arasındalık**

Toplulukları bul | Yayılmayı başlat

Haberi iletme olasılığı p = **0,4**

Deniz: derece **4** (4. sıra), arasındalık **0,58** (1. sıra)

Arasındalık, en kısa yolların kaçının bu kişiden geçtiğini sayar; köprü kişiler küçük çevreyle de yüksek çıkar.

Derecede Mert gibi kalabalık çevreliler öne çıkar, ama gruplar arası köprü Deniz'dir (arasındalık). Haberi köprüden başlatmak onu birden çok topluluğa taşır; etkileyici seçimi böyle yapılır.

KONUM

Ön Koşullar

Bu dersten önce bilinmesi yararlı olan dersler ve neden gerektikleri.

Ders	Neden gerekir
Doğrusal Cebir	komşuluk matrisi, spektral yöntemler
Veri Yapıları	komşuluk listesi ve matrisi: çizgeyi bellekte tutmak
Algoritmalar	BFS, DFS ve en kısa yol: merkezilik hesaplarının temeli
Ayrık Matematik	çizge kuramı, kümeler ve sayma

KONUM

Bu Ders Sana Göre mi?

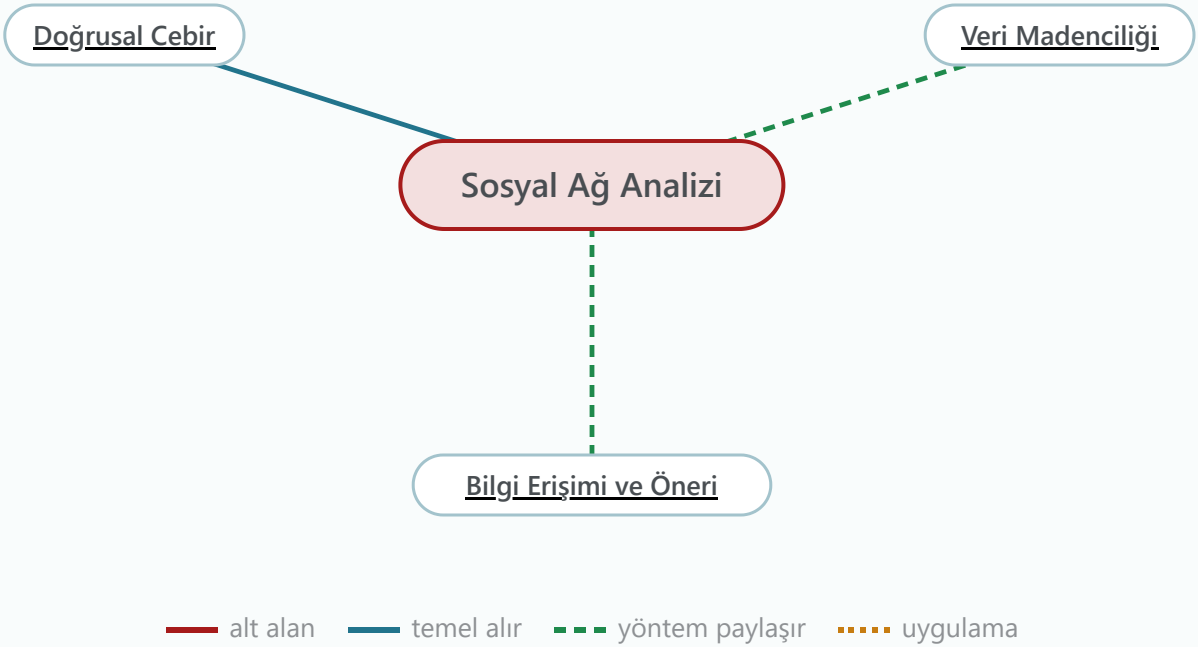
Kendine sor; "hayır" dediklerin için sağdaki sütun nereden başlayacağını gösterir.

Kendine sor	Hayırsa önce
BFS ile bir çizgede iki düğüm arasındaki en kısa yolu bulabiliyor musun?	Algoritmalar dersindeki çizge aramaları
Bir matrisin özvektörünün ne anlama geldiğini açıklayabiliyor musun?	Doğrusal Cebir
Bir dağılımın ortalamasını ve varyansını hesaplayabiliyor musun?	Olasılık ve İstatistik
Python'da sözlük kullanarak bir kenar listesi dosyasını okuyabiliyor musun?	Programlamaya giriş, ardından NetworkX'e kısa bir giriş

Çoğuna "evet" diyebiliyorsan derse hazırsın; bir ya da iki "hayır" dönem içinde kapanabilir.

KONUM

Dersin Komşuları



ÖRTÜŞME

Örtüştüğü Dersler

İlişki	Ders	Ortak konu
temel alır	Bu ders Doğrusal Cebir dersini temel alır	komşuluk matrisi, spektral yöntemler
yöntem paylaşır	Veri Madenciliği ile yöntem paylaşır	topluluk bulma
yöntem paylaşır	Bilgi Erişimi ve Öneri Sistemleri ile yöntem paylaşır	PageRank, bağlantı çözümlemesi

Ayrıştığı Noktalar (1/2)

Ölçüt	Sosyal Ağ Analizi	Veri Madenciliği	Bilgi Erişimi ve Öneri
Temel soru	Kişiler ve ilişkilerden oluşan ağlar nasıl çözümlenir?	Büyük veri yığınlarındaki gizli ve işe yarar örüntüler nasıl bulunur?	Kullanıcıya en ilgili belge ya da ürün milyonlarca aday içinden nasıl getirilir?
Yaklaşım	çizge kuramı ve istatistik	keşfe dayalı çözümleme	eşleme ve sıralama
Girdi	düğümler ve bağlantılar	büyük tablo ya da işlem verisi	sorgu, belgeler, kullanıcı geçmişi
Çıktı	merkezilik, topluluk, yayılım	kural, küme, aykırı değer	sıralı sonuç listesi
Başarı ölçütü	modülerlik, yorumlanabilirlik	ilginçlik, destek, güven	kesinlik, duyarlılık, NDCG
Matematik yükü	orta: çizge, matris	orta	orta: vektör uzayı
Veri ihtiyacı	ağ verisi	çok	çok etkileşim verisi

- **Veri Madenciliği:** Veri madenciliği tablo halindeki kayıtlarda örüntü arar; bu ders kayıtlar arasındaki bağlantıların yapısını inceler.
- **Bilgi Erişimi ve Öneri:** Bilgi erişimi PageRank'i sıralama sinyali olarak kullanır; bu ders merkeziliğin ağ hakkında ne söylediğini sorar.

AYRIŞMA

Ayrıştığı Noktalar (2/2)

Ölçüt	Sosyal Ağ Analizi	Doğrusal Cebir	Çok Etmenli Sistemler
Temel soru	Kişiler ve ilişkilerden oluşan ağlar nasıl çözümlenir?	Vektörler ve matrisler veriyi ve dönüşümleri nasıl temsil eder?	Birden çok özerk etmen birlikte ya da birbirine karşı nasıl karar verir?
Yaklaşım	çizge kuramı ve istatistik	vektör uzayları ve dönüşümler	etkileşen özerk etmenler
Girdi	düğümler ve bağlantılar	vektör, matris	etmenler, hedefleri, iletişim kuralları
Çıktı	merkezilik, topluluk, yayılım	çözüm, ayrıştırma, dönüşüm	ortak davranış, denge, anlaşma
Başarı ölçütü	modülerlik, yorumlanabilirlik	doğruluk, sayısal kararlılık	toplam fayda, kararlılık
Matematik yükü	orta: çizge, matris	kuramın kendisi	orta: oyun kuramı
Veri ihtiyacı	ağ verisi	gerekmez	az; benzetim

- **Doğrusal Cebir:** Doğrusal cebir komşuluk matrisinin özdeğerlerini hesaplatır; bu ders onları topluluk ve önem olarak yorumlar.
- **Çok Etmenli Sistemler:** Çok etmenli sistemler etmenlerin nasıl karar verdiğini modeller; bu ders gözlenen ilişki ağını çözümler.

UYGULAMA

Uygulama Alanları

Sosyal medya
Etkileyici kişiler ve yankı odaları.

Salgın bilimi
Temas ağında hastalığın yayılması.

Sahtekârlık
Bağlantılı sahte hesap halkaları.

Bilim ağları
Atıf ve ortak yazarlık çözümlemesi.

Kurum içi ağlar
Bilginin şirkette nasıl dolaştığı.

Arkadaş önerisi
Ortak tanıdığa göre bağlantı tahmini.

UYGULAMA

Riskler ve Etik

Paylaşılmamış bilginin sızması

Arkadaşlık ağı, kişinin hiç paylaşmadığı siyasi görüş ya da cinsel yönelim gibi bilgileri tahmin ettirebilir.

Adı silmek yetmez

Adları silinmiş bir ağda bile bağlantı örüntüsü kişiyi tanıtabilir; yapısal yeniden tanımlama saldırıları bunu gösterdi.

Rıza ve veri kazıma

Cambridge Analytica olayında bir kişilik testi uygulaması, katılanların arkadaşlarının verisini de rızaları olmadan topladı.

Eksik ağ, yanlış sonuç

Ankette unutulan ya da platformda görünmeyen bağlar, merkezilik sıralamasını ve toplulukları ciddi biçimde değiştirebilir.

Ayrıntılı tartışma: Açıklanabilir Yapay Zekâ ve Yapay Zekâ Etiği.

UYGULAMA

Araçlar ve Örnek Veriler

Araçlar ve kütüphaneler

NetworkX | igraph | Gephi | graph-tool | Neo4j | PyTorch Geometric

Tanınmış veri setleri

Zachary karate kulübü | Les Misérables karakter ağı | Enron e-postaları | SNAP ağ koleksiyonu | DBLP ortak yazarlık ağı | Facebook ego ağları (SNAP)

UYGULAMA

Sık Karıştırılanlar

Yaygın düşünce	Doğrusu
En çok arkadaşı olan kişi en etkilidir.	Derece tek ölçüdür; iki grubu birleştiren köprü kişi (arasındalık) daha kritik olabilir.
Ağ çizimine bakmak ağı anlamak için yeter.	Büyük ağların çizimi "kıl yumağı"dır; sayısal ölçüler ve sınama gerekir.
Topluluk bulmanın tek doğru cevabı vardır.	Farklı yöntemler farklı bölünmeler verir; sonuç alan bilgisiyle yorumlanır.
Bağlantılı kişiler birbirini etkilemiştir.	Benzerler zaten birbirini seçer (homofili); etki ile seçimi ayırmak zordur.

SONRASI

Bu Dersten Sonra

Bu dersi temel alan ya da genişleten dersler

Bu ders çoğunlukla uç noktadır; öteki dersleri temel alır.

İlgili meslekler

Ağ bilimci | Veri bilimci | Sosyal medya analisti | Sahtekârlık analisti | Hesaplamalı sosyal bilimci

Bu dersten sonra çizge sinir ağlarıyla ağ üzerinde öğrenmeye, bilgi erişimiyle bağlantı temelli sıralamaya, salgın bilimiyle yayılım modellerine geçilebilir. Hesaplamalı sosyal bilimler, bu dersin yöntemlerini toplum sorularına uygulayan disiplinler arası bir alandır.

SONRASI

Kaynaklar

- Wasserman ve Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994.
- Easley ve Kleinberg, *Networks, Crowds, and Markets*, Cambridge University Press, 2010.
- Newman, *Networks*, 2. baskı, Oxford University Press, 2018.
- Barabási, *Network Science*, Cambridge University Press, 2016.

Son

Bölüm 30: Sosyal Ağ Analizi

← Önceki bölüm · Alanın haritası · Sonraki bölüm →