

VERİ MADENCİLİĞİ

Sosyal Ağlar

Yrd. Doç. Dr. Şule Gündüz Öğüdücü
http://www3.itu.edu.tr/~sgunduz/courses/verimaden/

1

Sosyal Ağ

- Sosyal ağ kişiler arasındaki ilişkilerin oluşturduğu bir yapıdır
- Sosyal ağ incelemesi: ağ yapısının, kişiler ya da gruplar (topluluklar) arasındaki ilişkilerin ve bilgi akışının incelenmesi



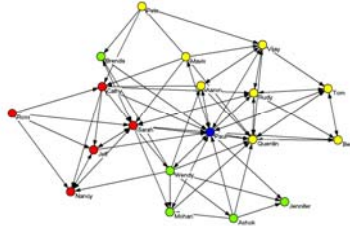
L.C. Freeman, Visualizing Social Networks. Journal of Social Structure, 2000.

http://www3.itu.edu.tr/~sgunduz/courses/verimaden/

2

Sosyal İlişkiler: Çizge

- Düğüm: Kişiler
- Ayrıtlar: sosyal ilişkiler
 - aile, arkadaş, iş
- Çizge $G(V,E)$
 - V : düğümler kümesi
 - E : Ayrıtlar kümesi
- Benzerlik Matrisi



S. Milgram (1967)

Yakınlığın Altı Derecesi: Six Degrees of Separation

http://www3.itu.edu.tr/~sgunduz/courses/verimaden/

3

Sosyal Ağların Özellikleri

- Farklı (doğal) ağlar
 - sosyal, biyolojik, teknik, içerik..
- Ortak özellikleri
 - Çok büyük, dinamik: düğümler, ayrıtlar eklenebilir/silenebilir
 - düğüm hangi düğümlerle ilişkide olacağına kendileri karar veriyorlar
 - düğümler arası etkileşim ayrıtlarla sınırlı
 - uzaklık/benzerlik için soyut bilgi: coğrafi, içerik, ilişkiler
- Sosyal ağ kuramı: link analizi
 - Farklı ağların genel özellikleri nelerdir?
 - Bu özellikler nasıl belirlenir, nasıl ölçülür?

http://www3.itu.edu.tr/~sgunduz/courses/verimaden/

4

Sosyal Ağların Özellikleri

- Uzunluk ($d_G(s,t)$): İki düğüm (s,t) arasındaki en büyük, en küçük, ortalama uzaklık
 - İki düğüm arasında bulunan yoldaki ayrıtlar sayısı
 - İki düğüm arasında bulunan yoldaki ayrıtların ağırlıklarının toplamı
- Derece: yönlü ise düğüme gelen (in-link) / düğümden çıkan (out-link) bağlantıların sayısı
- Merkez: Ağdaki diğer düğümlerin bağlı olduğu bir ya da bir kaç düğüm
- Yoğunluk: Ağdaki bağlantı sayısının olası bütün bağlantı sayısına oranı

http://www3.itu.edu.tr/~sgunduz/courses/verimaden/

5

Tanımlar

- $\sigma_{st} = \sigma_{ts}$ s ve t ($s, t \in V$) düğümleri arasındaki en kısa yol sayısı $\rightarrow \sigma_{ss} = 1$
- $\sigma_{st}(v)$: $v \in V$ düğümünün üzerinde bulunduğu s ve t düğümleri arasındaki en kısa yol sayısı

$$C_c(v) = \frac{1}{\sum_{t \in V} d_G(v, t)}$$

closeness centrality (Sabidussi, 1966)

$$C_G(v) = \frac{1}{\max_{t \in V} d_G(v, t)}$$

graph centrality (Hage and Harary, 1995)

$$C_s(v) = \sum_{s \neq v \neq t \in V} \sigma_{st}(v)$$

stress centrality (Shimbel, 1953)

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

betweenness centrality (Freeman, 1977; Anthonisse, 1971)

Ulrik Brandes, A Faster Algorithm for Betweenness Centrality, Journal of Mathematical Sociology 25(2):163-177, (2001).

http://www3.itu.edu.tr/~sgunduz/courses/verimaden/

6

Tanımlar

- Hizip (Clique):
 - seçilebilecek her düğüm çifti arasında bir bağ olan alt çizge
 - tam bağlı alt çizge
- Daha zayıflatılmış
 - N-Hizip (N-Clique): Bir düğümün içinde bulunduğu alt çizgedeki diğer tüm düğümlere olan uzaklığı en çok N olabilir
 - N-Klan (N-Clan): N-Hizipteki düğüm çiftleri arasındaki yollar üstündeki düğümler de N-Hizip üyesi
 - K-Plexes: Bir düğümün n düğümden oluşan bir N-Hizip içindeki en az n-k düğüm ile doğrudan bağlı olması
- Kesitleme noktası (Cut Points): Bağlı olan bir G çizgesinden, v düğümü ve bu düğüme bağlı olan bütün ayrıtlar çıkarıldığında oluşan G-v çizgesi bağlı değil ise v kesitleme noktasıdır.



<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

7

Sosyal Ağların Yeteneği

Özellik	Tanım	Etki
Derece (Degree)	Bir düğümün bağlantı sayısı	daha fazla seçenek
Yakınlık	Diğer düğümlere olan yolun uzunluğu	Diğer düğümlerle doğrudan etkileşim
Ara düğüm	İki düğüm arasında yer alan düğüm	Diğer iki düğüm arasında ilişkiye sağlamak/kesmek

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

8

Problemler

- Bağlı parçalar:
 - kaç parça, büyüklükleri ne, ne kadar bağlılar
- Ağ çapı:
 - en uzak – ortalama
 - Bağlı olmayan düğümler / parçalar
 - Küçük dünya özelliği
- Demetleme
 - Gruplaşan ilişkiler/düğüm
 - Örtüşen gruplar
 - Grup içi ve gruplar arası ilişkilerin oranı
 - Grup içi ve gruplar arası ilişkilerin rolü
- Ağın yapısı
 - düğümün derecesi
 - k_i : i düğümünün derecesi, N_i : i düğümünün komşular kümesi
 - ilişkilerin ağırlığının dağılımı
 - ağ içinde önemli rolü olan düğümler: iki gruba birbirine bağlayan

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

9

Web'de Sosyal Ağlar

- Facebook
 - Myspace
 - LinkedIn
 - Classmates
 - Orkut
 - Bebo
- Medya paylaşım siteleri:
- YouTube
 - Flickr

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

10



Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1), article 11. <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>

11

Sosyal Ağlar için Modeller

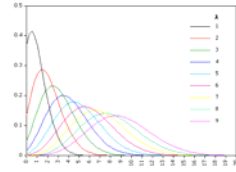
- Rassal çizgeler (Random Graphs: Erdős-Rényi models)
- Watts-Strogatz modelleri
- Scale-free Networks

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

12

Rassal Ağlar

- Erdős-Rényi (ER) Model: 1959 yılında Paul Erdős ve Alfred Rényi
- düğüm sayısı N , iki düğüm arasında ayırılma olasılığı p , $G(N,p)$
 - $p = 1/2N$, $p = 1/N$, $p = 2/N$, $p = 10/N$, $p = \log(N)/N$...
- $N(N-1)/2$ adet hileli yazı tura atma
- Derece dağılımı
 - N çok büyük olduğunda Poisson dağılımı
 - $G(N,p)$ ile bir çizge oluştur, rastgele bir u düğümü
 - $\Pr[\deg(u) = k] ?$
 - Poisson dağılımı ortalama $\lambda = p(N-1) \sim pN$
- Demetleme katsayısı (clustering coefficient) C küçük



$$f(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

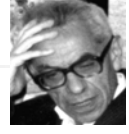
$$C_i = \frac{2|e_{jk}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{jk} \in E$$

$$\bar{C} = \frac{1}{N} \sum_{i=1}^N C_i$$

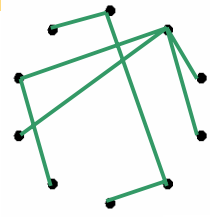
<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

13

Erdős-Rényi Model (1960)



Pál Erdős (1913-1996)



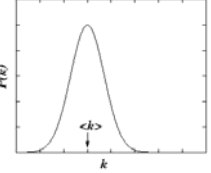
Connect with probability p

$$p = 1/6$$

$$N = 10$$

$$\langle k \rangle \sim 1.5$$

Poisson distribution



<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

Watts-Strogatz modelleri: Caveman and Solaria

- Erdős-Rényi
 - Ortak komşuları olması iki düğüm arasında ayırılma olasılığını artırmıyor
 - her ayırılma daha öncekilerden bağımsız olarak oluşuyor
- Gerçekte oluşan ağ yapısına uygun değil
 - iki kişinin tanışma olasılığı ortak arkadaşları varsa daha fazladır
 - Web de iki sayfa birbirine bağlı ise büyük olasılıkla aynı konudadırlar
- Watts Caveman:
 - ayrıtların genel olarak yoğunluğu az
 - iki düğümün ortak komşuları varsa aralarında ayırılma olasılığı büyük
- Watts Solaria
 - ayrıtların genel olarak yoğunluğu az, bir düğümün komşuları arasında ayırılma olasılığı farklı değil
 - Erdős-Rényi çizgesine benzer

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

15

α -model

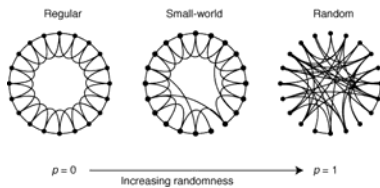
- α -model için parametreler
 - N düğüm sayısı
 - k : ortalama derece
 - p : iki düğüm arasında ayırılma olasılığı
 - α : yakın ilişkilerin olasılığını artırmak için parametre
- u, v düğüm çifti için
 - $m(u, v)$: ortak komşu sayısı
- $R(u, v)$: iki düğümün arasında ayırılma eğilimi (propensity)
 - $m(u, v) \geq k$, $R(u, v) = 1$
 - $m(u, v) = 0$, $R(u, v) = p$
 - diğer durumlarda $R(u, v) = p + (m(u, v)/k)^\alpha (1-p)$
- $\alpha \rightarrow \infty$ için Erdős-Rényi çizgelerine benziyor

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

16

Watts-Strogatz Model

- Bir daire etrafında eşit dağılmış N düğüm
- Her düğümün en yakın k komşusu arasında k ayırıl (yakın ilişki)
- p olasılığı ile bir düğüme az sayıda rastgele ayırıl ekle (uzak ilişki)
- farklı p değerleri için farklı çizgeler



Collective dynamics of 'small-world' networks
Duncan J. Watts & Steven H. Strogatz

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

17

Small Worlds - Occam's Razor

- küçük α değerleri için demetleme katsayısı büyük
 - Erdős-Rényi $\rightarrow$ çapı küçük
 - α -model $\rightarrow$ büyük demetleme katsayısı
- Occam's Razor
 - farklı özellikler için basit tek bir model
- Watt's small world:
 - çapı küçük
 - demetleme katsayısı büyük

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

18

Örnek

- üç gerçek ağ üzerinde inceleme
 - Oyuncular
 - batı bölgesindeki güç santralleri
 - C.elegans sinir sistemi

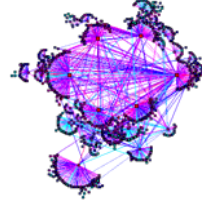
	n	k	d	c
Actors	225,226	61	3.65	0.79
Power-grid	4,941	2.67	18.7	0.08
C.elegans	282	14	2.65	0.28

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

19

Small World

- Serbest Ölçekli Ağlar (Scale free): Örnek web
 - düğüm web sayfaları
 - ayrıntlar bağlantılar



web örümcekleri/web robotlar

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

20

Serbest Ölçekli Ağlar

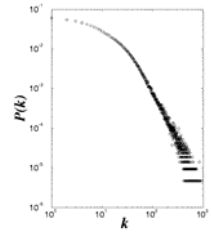
- Düğüm sayısı N sabit değil
 - Ağ sürekli yeni düğümlerin eklenmesi/silinmesi ile değişiyor
 - web: yeni sayfalar/siteler ekleniyor/siliniyor
 - yayınlar: yeni yayınlar ekleniyor
- Eklenen düğümlerle oluşan ayrıntlar üniform değil
 - çok sayıda sayfanın/sitenin bağlantı verdiği sayfaya/siteye bağlantı verme olasılığı yüksek
 - çok sayıda yayının referans gösterdiği yayının referans gösterilme olasılığı yüksek

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

21

Serbest Ölçekli Ağlar

- Pareto* veya *power law* dağılımı
- çapı küçük ($\sim \log(N)$)
- demetleme katsayısı çok büyük değil
 - yakın komşuları ile ayrıt olma olasılığı yüksek değil
 - "hub" olan düğümlerle ayrıt oluşturma olasılığı yüksek



<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

22

Topluluk Belirleme

- S. Fortunato and C. Castellano, *Community Structure in Graphs*, ArXiv e-prints
- Bir çizge içinde ortak özellikleri/görevleri olan düğümler topluluğu
 - aynı konudaki web sayfaları
 - benzer işleve sahip proteinler
 - aynı konuda çalışan insan grupları
 - aynı ilgi alanına sahip insan grupları
- Topluluk içindeki topolojik konumlarına göre düğümler sınıflandırılabilir
 - topluluğun merkezinde yer alan düğüm
 - topluluğun sınırında yer alan düğüm

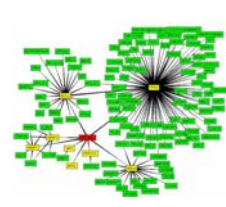
<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

23

Metabolik



Protein



Sosyal



Ekonomik



24

Topluluk Tanımı

- Dar tanım
- Geniş tanım
- D ğ m benzerliđi

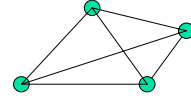
<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

25

Dar Tanım

- Sadece alt  izgedeki iliřkilere g re topluluklar  belirliyor

 rnek: Hizip, n-klan, k-plexes



<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

26

Geniř Tanım

- Topluluk  izginin yapısal bir birimi
- Alt izgeler hem kendi i indeki iliřkiler hem de  izgenin geri kalanıyla olan iliřkileri ile belirleniyor
- Null model: i inde topluluk bulunmayan  izge
 - Erd s-Renyi
 - Newman-Girvan: d ğ mlerin orjinal  izge ile aynı dereceye sahip olduđu rassal  izge

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

27

D ğ m benzerliđi

- D ğ mler birbirine "benzer" ise aynı toplulukta
 - dar
 - geniř

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

28

 izge Par alama

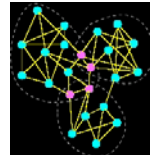
- Graph Partitioning:  izgeyi bir d ğ m bir grupta kalacak řekilde gruplara (alt izgelere) b lme
- Problem:
 -  rt řen topluluklar
 - Hiyerarřik yapı

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

29

 rt řen Topluluklar

- Ger ek d nyada bir nesne birden fazla gruba  ye olabilir



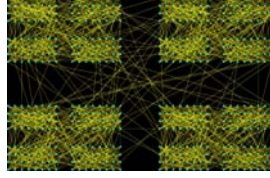
G. Palla, I. Der nyi, I. Farkas, T. Vicsek,
*Uncovering the overlapping community structure of
complex networks in nature and society*
Nature 435, 814, 2005
<http://www.cfinder.org/>

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

30

Hiyerarşik Yapı

- Altçizgeler tekrar parçalanabilir



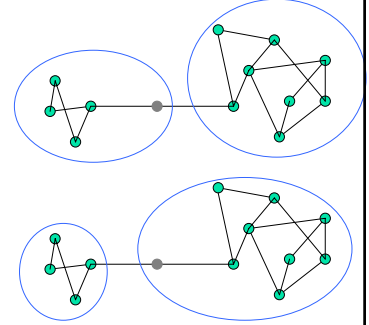
A. Clauset, C. Moore, M.E.J. Newman,
Hierarchical structure and prediction of missing links
Nature 453, 98, 2008

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

31

Çizge Parçalama

- Hangisi daha iyi



<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

32

İki Farklı Sonuç

- Üstünlüklerini karşılaştırmak için bir kriter Q
- $Q(P1) > Q(P2)$ veya $Q(P1) < Q(P2)$?

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

33

Birimsellik

- Modularity: $A_{uv} = \begin{cases} 1 & (u,v) \in E \\ 0 & \text{diğer} \end{cases}$

$$m = \sum_{u,v \in V} \frac{A_{uv}}{2}$$

$$k_u = \sum_{v \in V} A_{uv}$$

$$C = \{c_1, c_2, \dots\}, c_i \cap c_j = \emptyset (i \neq j) \cup_{c_i \in C} c_i = V$$

$$e_{ij} = \sum_{u \in c_i, v \in c_j} A_{uv} / 2m$$

$$a_i = \sum_{v \in c_i} k_v / 2m$$

$$Q(G, C) = \sum_i (e_{ii} - a_i^2)$$

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

34

Girvan-Newman Algoritması

- M. Girvan & M.E.J. Newman, *Community structure in social and biological networks*, PNAS 99, 7821-7826 (2002)
- Toplulukları birbirine bağlayan ayrıtları belirle
 - Betweenness
- bu ayrıtları silme
 - birbirinden ayrık demetler

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

35

Girvan-Newman Algoritması

- Bütün ayrıtların "betweenness" değerleri hesaplanır
- En büyük "betweenness" değerine sahip ayrıt silinir
- Kalan ayrıtların "betweenness" değerleri hesaplanır
2. adıma geri dönülür

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

36

Geliştirilmiş Girvan-Newman Algoritması

- M.E.J. Newman & M. Girvan, *Finding and evaluating community structure in networks*, Phys. Rev. E 69, 026113 (2004)
 1. Her düğüm bir demet
 2. Q değerini en büyütecek iki düğümü birleştir
 3. Bütün düğümler tek demet olana kadar işleme devam et
 4. En fazla Q değerine sahip demetlemeyi seç

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

37

Problemler

- Örtüşen topluluklar
- Yönlü çizgeler
- Ağırlıklı çizgeler
- Karmaşıklık

<http://www3.itu.edu.tr/~sgunduz/courses/verimaden/>

38