

Sosyal Medya Verilerden Çizge Vektörlerinin Çıkarılması

BIL 496 1. izleme

Hasan MEN

Proje Danışmanı: Yrd. Doç. Dr. Burcu YILMAZ

Nisan 2018



İçerik



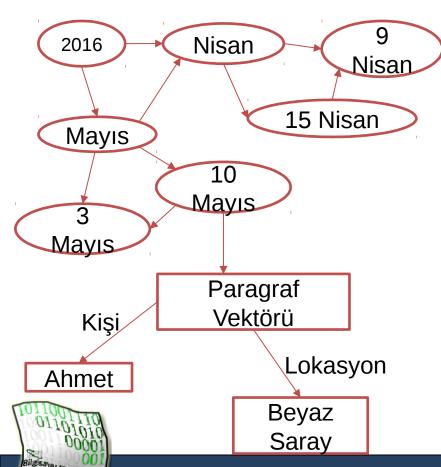
- Projenin Şeması ve Tanımı
- Neler Yapıldı?
- Neler Yapılacak?
- Başarı Kriterleri
- Kaynaklar



Proje Şeması ve Tanımı



	#	SentByScreenName	TextText	DateCreatedAt
ľ	1	abdulhamitgul	Hayırlı olsun. Türkiye yoluna hızlı ve emin adımlarla	Fri Apr 01 00:07:35 EEST 2016
i	2	ahmet_iyimaya	KESİNTİSİZ ZİKİR HÂLİNDE OLAN ORMAN ZEN	Fri Apr 01 04:35:37 EEST 2016
	3	hakanchelik	Cumhurbaşkanı Erdoğan, Beyaz Saray'da ABD Baş	Fri Apr 01 05:58:24 EEST 2016
	4	hakanchelik	Cumhurbaşkanı Erdoğan ve ABD Başkanı Obama g	Fri Apr 01 06:59:16 EEST 2016

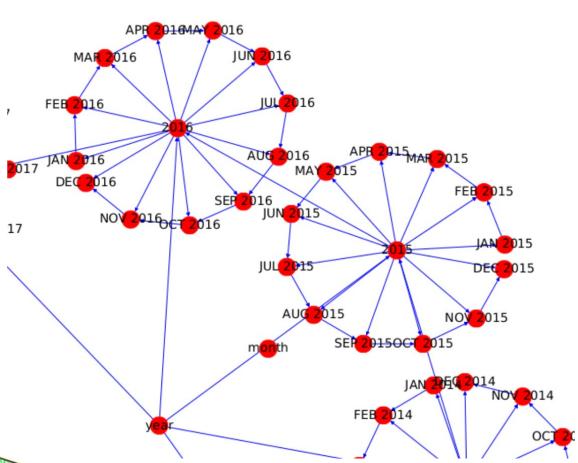


Proje Nedir?

* Sosyal medya verilerinden çizge vektörlerinin çıkarılması ve düğümler arasındaki yakınlıkların node2vec vektörleriyle gösterilmesi



Ay bazında çizgeler ve vektörleri çıkarıldı.



GTÜ - Bilgisayar Mühendisliği Bölümü

```
G_time2_model.wv.most_similar("2014")

[('DEC 2014', 0.4804452657699585),
    ('NOV 2014', 0.473899781703949),
    ('OCT 2014', 0.46294546127319336),
    ('AUG 2014', 0.44025206565856934),
    ('JUN 2014', 0.439254492521286),
    ('JUL 2014', 0.4392107129096985),
    ('SEP 2014', 0.43653786182403564),
    ('MAY 2014', 0.4274585247039795),
    ('DEC 2015', 0.4250481426715851),
    ('APR 2014', 0.42327630519866943)]
```



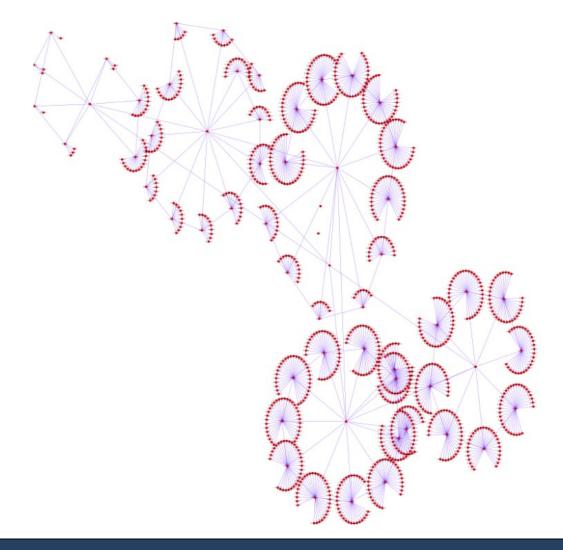
Ay bazında cizgeler ve vektörleri cıkarıldı,







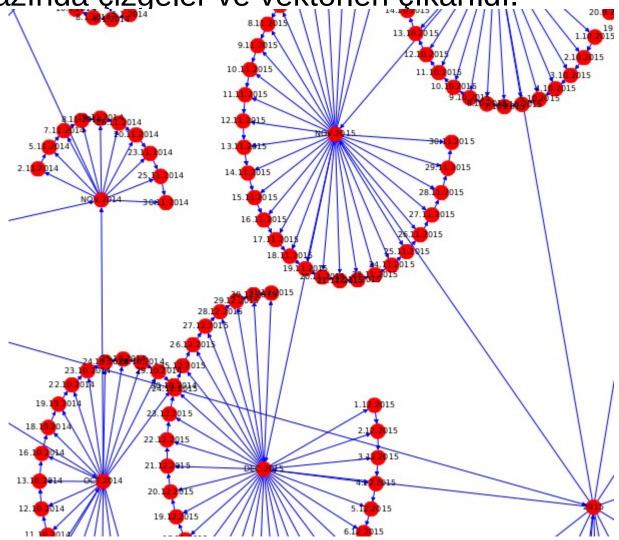
Gün bazında çizgeler ve vektörleri çıkarıldı.







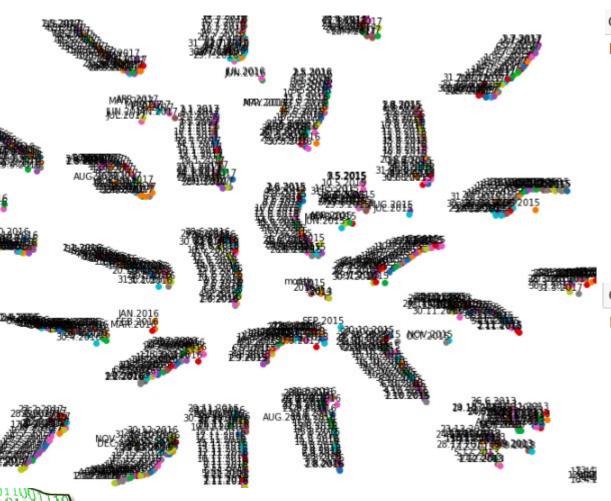
Gün bazında çizgeler ve vektörleri çıkarıldı.







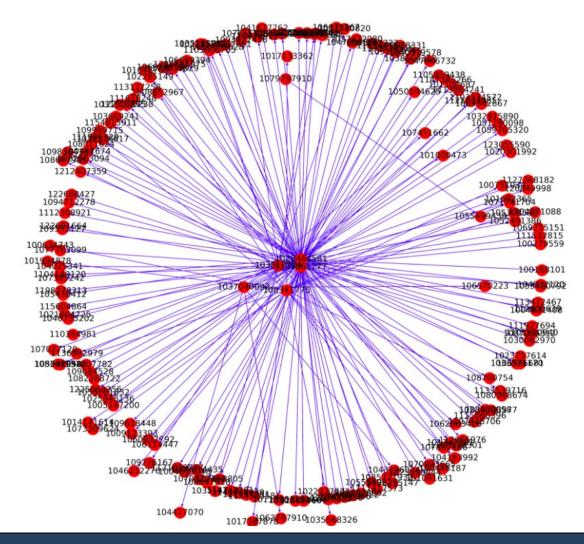
Gün bazında çizgeler ve vektörleri çıkarıldı.



```
G time3 model.wv.most similar("JUN.2014")
[('JUL.2014', 0.8733973503112793),
 ('MAY.2014', 0.8347429633140564),
('14.6.2014', 0.7849587202072144),
 ('12.6.2014', 0.78019118309021),
 ('AUG.2014', 0.7594411373138428),
 ('21.6.2014', 0.7541152834892273),
 ('15.6.2014', 0.7511141300201416),
 ('28.6.2014', 0.748937726020813),
 ('2.7.2014', 0.7451514601707458),
('29.6.2014', 0.7447819709777832)]
G time3v1 model.w Tmost similar("JUN.2014"
[('12.6.2014', 0.8578124046325684),
 ('14.6.2014', 0.8463558554649353),
 ('JUL.2014', 0.8231008052825928),
 ('29.6.2014', 0.8161718249320984),
 ('26.6.2014', 0.8142330050468445),
 ('21.6.2014', 0.8108713626861572),
 ('MAY.2014', 0.8063054084777832),
 ('30.6.2014', 0.8048838973045349),
 ('28.6.2014', 0.7968608140945435),
 ('15.6.2014', 0.7966266870498657)]
```



Arkadaşlık çizgeleri ve vektörleri çizildi.







Arkadaşlık çizgeleri ve vektörleri çizildi.





- Char based NER (CharNER)
 - Eğitimler tamamlanmadığı için tweetler üzerinde NER çizgesi çıkarılamadı.
 - 30000'den fazla cümlelerde RAM sorunu

```
processed 45947 tokens with 6261 phrases; found: 4748 phrases; correct: 1980.
accuracy: 84.65%; precision: 41.70%; recall: 31.62%; FB1: 35.97
             LOC: precision: 55.90%; recall: 48.54%; FB1: 51.96 1220
            MISC: precision: 49.62%; recall: 19.91%; FB1: 28.42 1433
             ORG: precision: 10.91%; recall: 4.20%; FB1: 6.06 55
             PER: precision: 28.48%; recall: 50.88%; FB1: 36.52 2040
tst
     10
             8.2529e-02
                           72,1386
                                      0.1501
                                                41,7000
                                                           31,6200
                                                                     35.9700
                                                                                35,9700
                                                                                               10
```



11



- Paragraf vektörleri çıkarıldı.
- Konu bakımından yakınlıklar var.
 - Edebi tweetler

```
model.docvecs.most similar("1")
[('430', 0.8913657665252686),
 ('138', 0.8643275499343872),
 ('400', 0.8520273566246033),
 ('58', 0.8425953388214111).
 ('398', 0.8381558656692505),
 ('257', 0.8364518284797668),
 ('196', 0.835493803024292),
 ('156', 0.8344122171401978),
 ('223', 0.8268957138061523),
 ('433', 0.8143468499183655)]
print(1,tweets[1].text)
print(430, tweets[430].text)
print(58,tweets[58].text)
print(61.tweets[61].text) # <- BURASI FARKLI
1 Kibir, bele bağlanmış taş gibidir. Onunla ne yüzülür ne de uçulur. (Hacı Bayram-ı Vel
430 Ne güzel ümmetin olmak..Ne güzel seni görmeden sevmek Ya Resulallah (S.A.V.) Kandil
t.co/KJq7HYrjJ6
58 Gel de birbirimizin kıymetini bilelim. Çünkü ansızın ayrılacağız birbirimizden. //
61 Önce Trabzonspor, sonra Fenerbahçe şimdi ise Beşiktaş ! haydı hayırlısı...
Şaka gibi...
```



Neler Yapılacak?



- NER tamamlanacak.
- Çizge vektörlerine Named Entity eklenecek.
- Çıkarılan parafraf vektörleride çizgelere eklenerek çizge vektörleri tekrardan oluşturulacak.

•



Başarı Kriterleri



- Ardışık gün vektörlerinin diğer günlere göre daha yakın olması
- Arkadaş kişi vektörlerinin diğer kişilere göre daha yakın olması
- Benzer konu&lokasyon(NER) vektörlerinin daha yakın olması



Kaynaklar



- Node2Vec: http://snap.stanford.edu/node2vec/ [Ziyaret Tarihi: 15/03/18]
- 2. Bag of Words: https://www.wikiwand.com/en/Bag-of-words_model [Ziyaret Tarihi: 15/03/18]
- 3. Networkx: https://networkx.github.io/ [Ziyaret Tarihi: 15/03/18]
- 4. Paragraph Vectors: https://github.com/inejc/paragraph-vectors [Ziyaret Tarihi: 15/03/18]
- 5.

