

Sosyal Medya Verilerden Çizge Vektörlerinin Çıkarılması

BIL 496 1. İzleme

Hasan MEN

Proje Danışmanı: Yrd. Doç. Dr. Burcu YILMAZ Nisan 2018



İçerik



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



Proje Tanımı ve Node2Vec

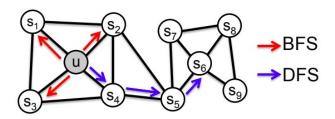


Proje Nedir?

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

Node2Vec

- Random Walk Based(Unsupervised) Feature Learning
- Node -> word, path -> sentence

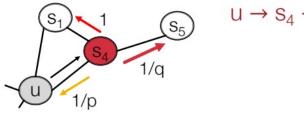


$$N_{BFS}(u) = \{s_1, s_2, s_3\}$$

Local microscopic view

$$N_{DFS}(u) = \{ s_4, s_5, s_6 \}$$

Global macroscopic view



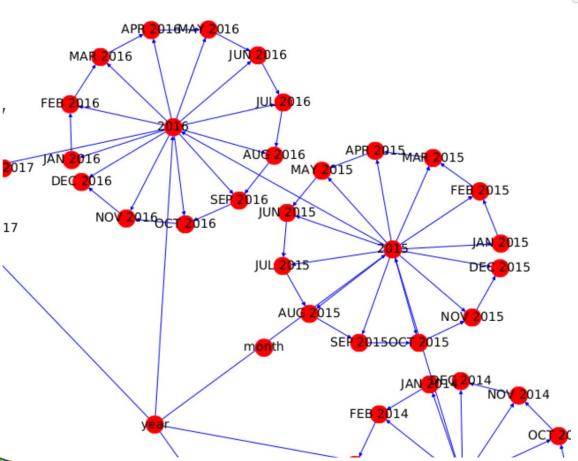
$$u \rightarrow s_4 \rightarrow ?$$

p-> Bir önceki düğüme geri dön q-> BFS, DFS





Ay bazında çizgeler ve vektörleri çıkarıldı.



```
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)]
G time2 model.wv.most similar("2016")
 [('DEC 2016', 0.46964654326438904),
 ('AUG 2016', 0.447937548160553).
  ('NOV 2016', 0.44665172696113586),
  ('APR 2016', 0.4372538626194),
  ('FEB 2016', 0.4347067177295685),
  ('JAN 2016', 0.42998284101486206),
  ('OCT 2016', 0.4287787079811096),
  ('SEP 2016', 0.42597827315330505),
  ('MAR 2016', 0.4255003035068512),
  ('JUN 2016', 0.4234979748725891)]
```



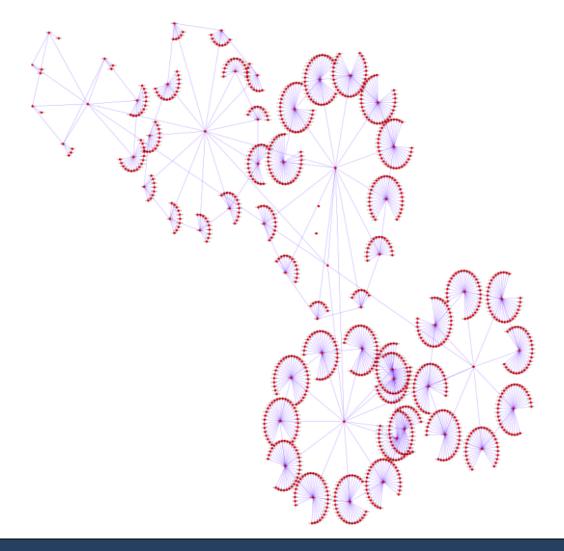
Ay bazında çıkarılan vektörlerin t-SNE çizimi







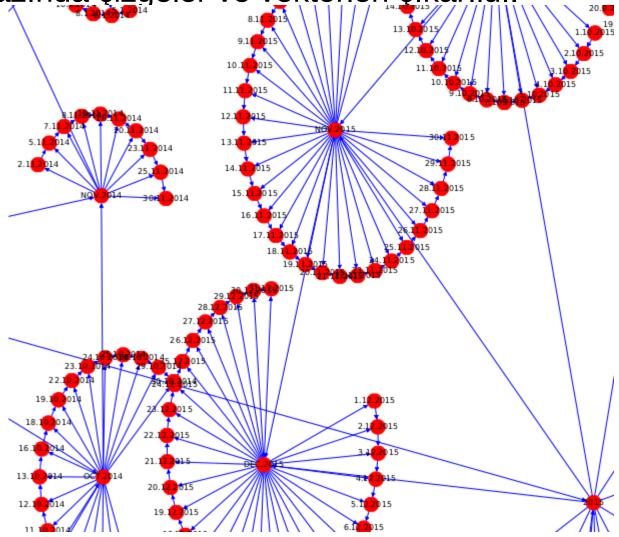
Gün bazında çizgeler ve vektörleri çıkarıldı.







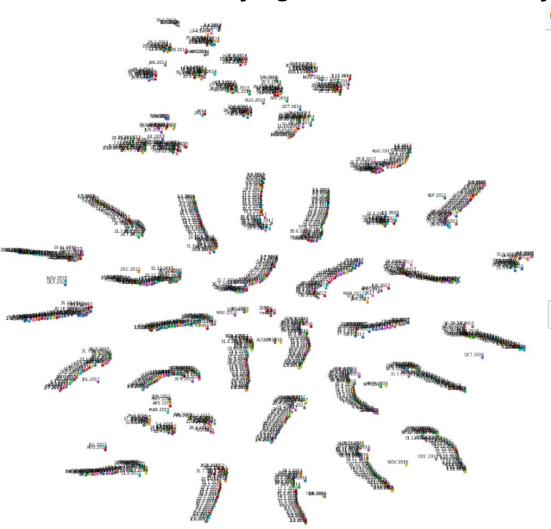
Gün bazında çizgeler ve vektörleri çıkarıldı.







Gün bazında çizgeler ve vektörleri çıkarıldı.



```
G_time3v1_model.wv.most_similar("JUN.2014")

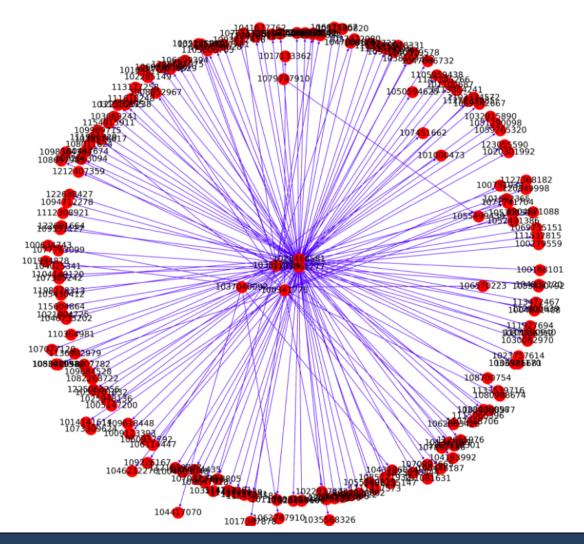
[('12.6.2014', 0.8578124046325684),
    ('14.6.2014', 0.8463558554649353),
    ('JUL.2014', 0.8231008052825928),
    ('29.6.2014', 0.8161718249320984),
    ('26.6.2014', 0.8142330050468445),
    ('21.6.2014', 0.8108713626861572),
    ('MAY.2014', 0.8063054084777832),
    ('30.6.2014', 0.8048838973045349),
    ('28.6.2014', 0.7968608140945435),
    ('15.6.2014', 0.7966266870498657)]
```

```
G_time3_model.wv.most_similar("AUG.2015")

[('SEP.2015', 0.6852082014083862),
    ('JUL.2015', 0.6794805526733398),
    ('4.8.2015', 0.5286612510681152),
    ('21.8.2015', 0.5271968245506287),
    ('9.8.2015', 0.5268226265907288),
    ('8.8.2015', 0.5258182287216187),
    ('6.8.2015', 0.5237710475921631),
    ('5.8.2015', 0.5236572027206421),
    ('3.8.2015', 0.5222306251525879),
    ('7.8.2015', 0.5213732719421387)]
```



Arkadaşlık çizgeleri ve vektörleri çizildi.







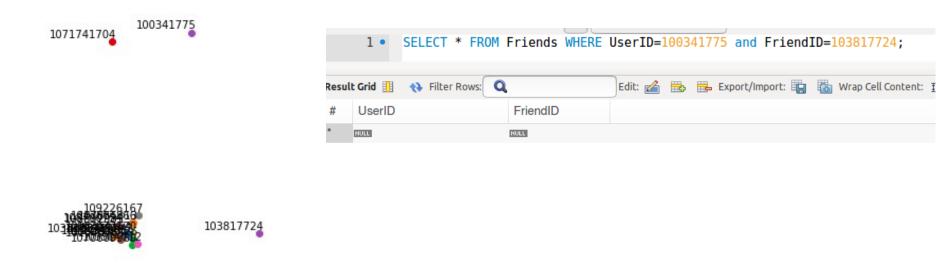
Arkadaşlık çizgeleri ve vektörleri çizildi.

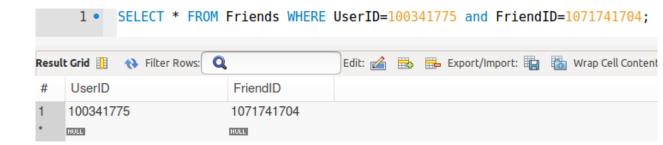






Arkadaşlık çizgeleri ve vektörleri çizildi.







102146672



- Char based NER (CharNER)
 - Eğitimler tam olarak tamamlanmadığı için tweetler üzerinde NER çizgesi çıkarılamadı.
 - 30000 lik veri ile test edildi.

GTÜ - Bilgisayar Mühendisliği Bölümü

```
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
```



Neler Yapılacak?



- NER eğitimlerini tamalama
- Çizge vektörlerine Named Entity ekleme

•



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.

