



Sosyal Medya Verilerden Çizge Vektörlerinin Çıkarılması

BIL 496
Son Sunum

Hasan MEN

Proje Danışmanı: Dr. Öğretim Üyesi Burcu YILMAZ

Haziran 2018



- Proje Tanımı ve Node2Vec
- Neler Yapıldı ve Sonuçlar?
- Başarı Kriterleri
- Kaynaklar



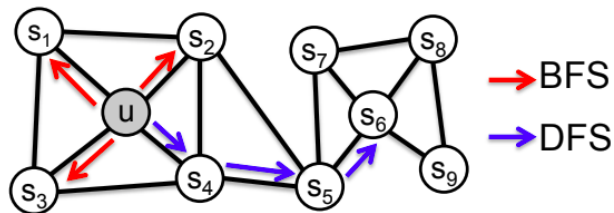
Proje Tanımı ve Node2Vec

- **Proje Nedir?**

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

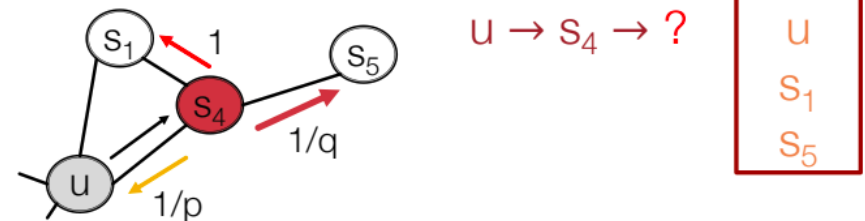
- **Node2Vec**

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



$$N_{BFS}(u) = \{s_1, s_2, s_3\} \quad \text{Local microscopic view}$$

$$N_{DFS}(u) = \{s_4, s_5, s_6\} \quad \text{Global macroscopic view}$$



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



Neler Yapıldı? - 1

- TWNERTC
 - Domain Independent, CoNLL2003
 - 100K Train, 10K Test, 10K Dev
- Karakter ve LSTM tabanlı NER modeli eğitilmiştir.
 - TR: Accuracy:%90, F1:%55
 - ENG: Accuracy: %94, F1:85

Text: Fikret Orman'dan Aziz Yıldırım'ın sözleri üzerine; "Bununla ilgili ben değil, federasyon cevap verecek"

NER: fikret B-PER, orman I-PER, aziz B-PER, yıldırım I-PER,

Text: Tiner ve zararları.... <https://t.co/MxhuZFGKYS>

NER: tiner B-PER,

Text: RT @burak_turna: Irak'ı ABD işgal ederken İtalya'nın adı hiç geçmiyordu, şimdi İtalya'dan başka adı geçen ülke yok:) <https://t.co/d5IgqTvI10>

NER: irak B-LOC, italya B-LOC, italya B-LOC,

```
%%time
words, preds = extract_ner("Hasan Türkiye'ye taşındı.")
print(words, preds)
```

```
['Hasan', 'Türkiye', 'ye', 'taşındı.'] ['B-PER', 'B-LOC', 'O', 'O']
CPU times: user 117 ms, sys: 95.8 ms, total: 213 ms
Wall time: 198 ms
```



Neler Yapıldı? - 2

- Twitter veri toplama

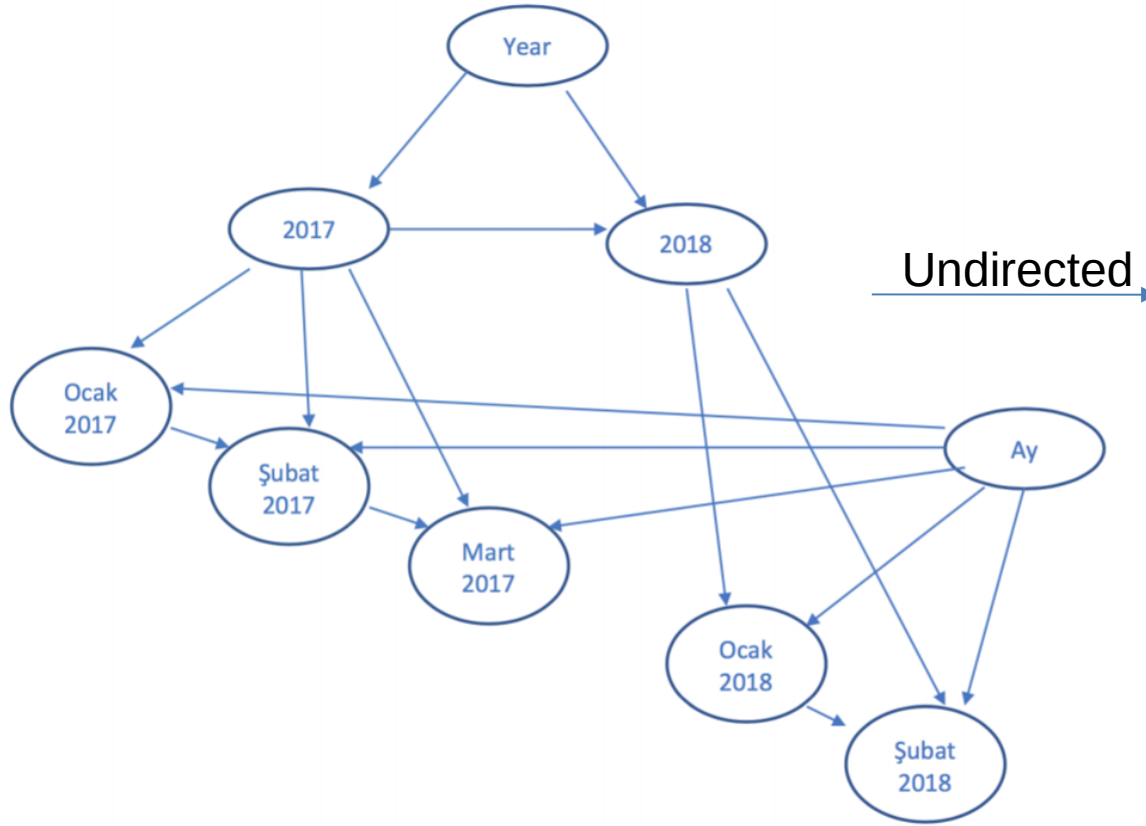
```
[ {  
  {  
    "sentByScreenName": "username1",  
    "inReplyToStatusID": "t76314116",  
    "inReplyToUserID": "u75717673",  
    "text": "TweetTextHere",  
    "userID": "u162375",  
    "tweetID": "t76314"  
  }  
}
```

```
[ {  
  {  
    "UserID": "u123",  
    "FriendID": "u234"  
  },  
  {  
    "UserID": "u123",  
    "FriendID": "u235"  
  },  
  {  
    "UserID": "u8234",  
    "FriendID": "u134234"  
  }  
}
```

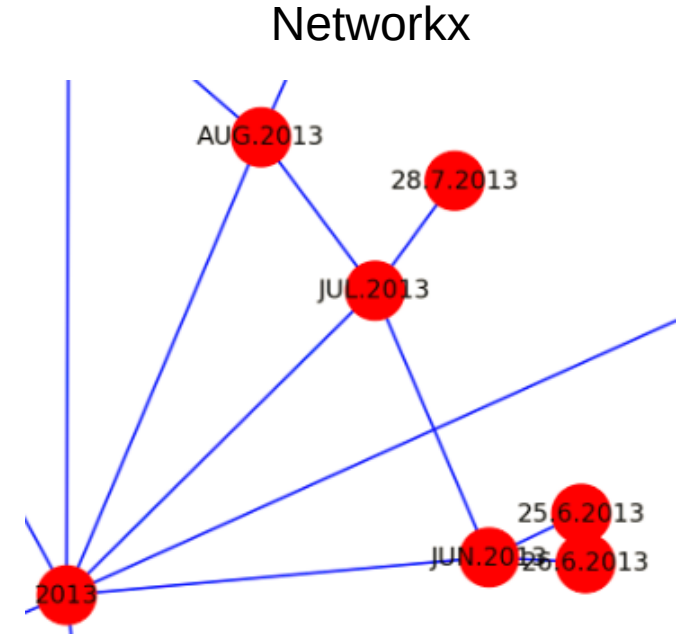
- MySQL ->JSON Format
- ~1000 Tweet



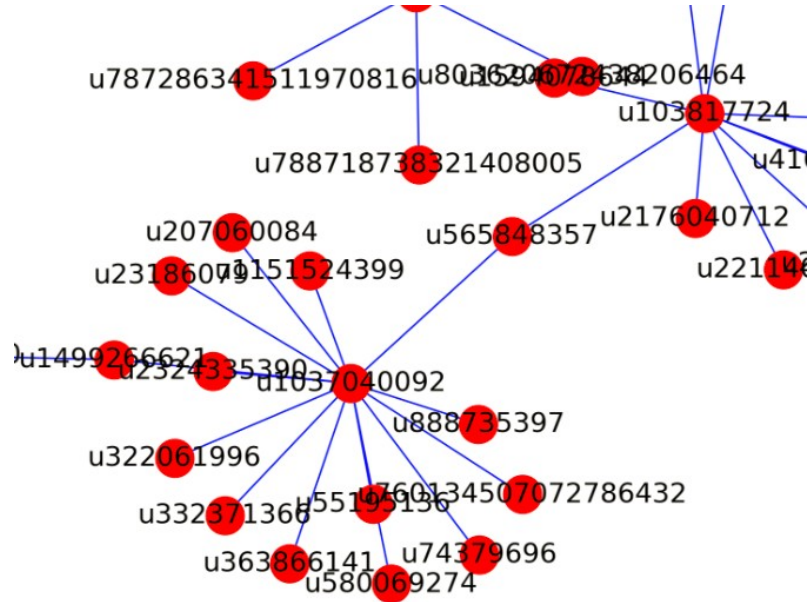
Neler Yapıldı? - 3



Undirected



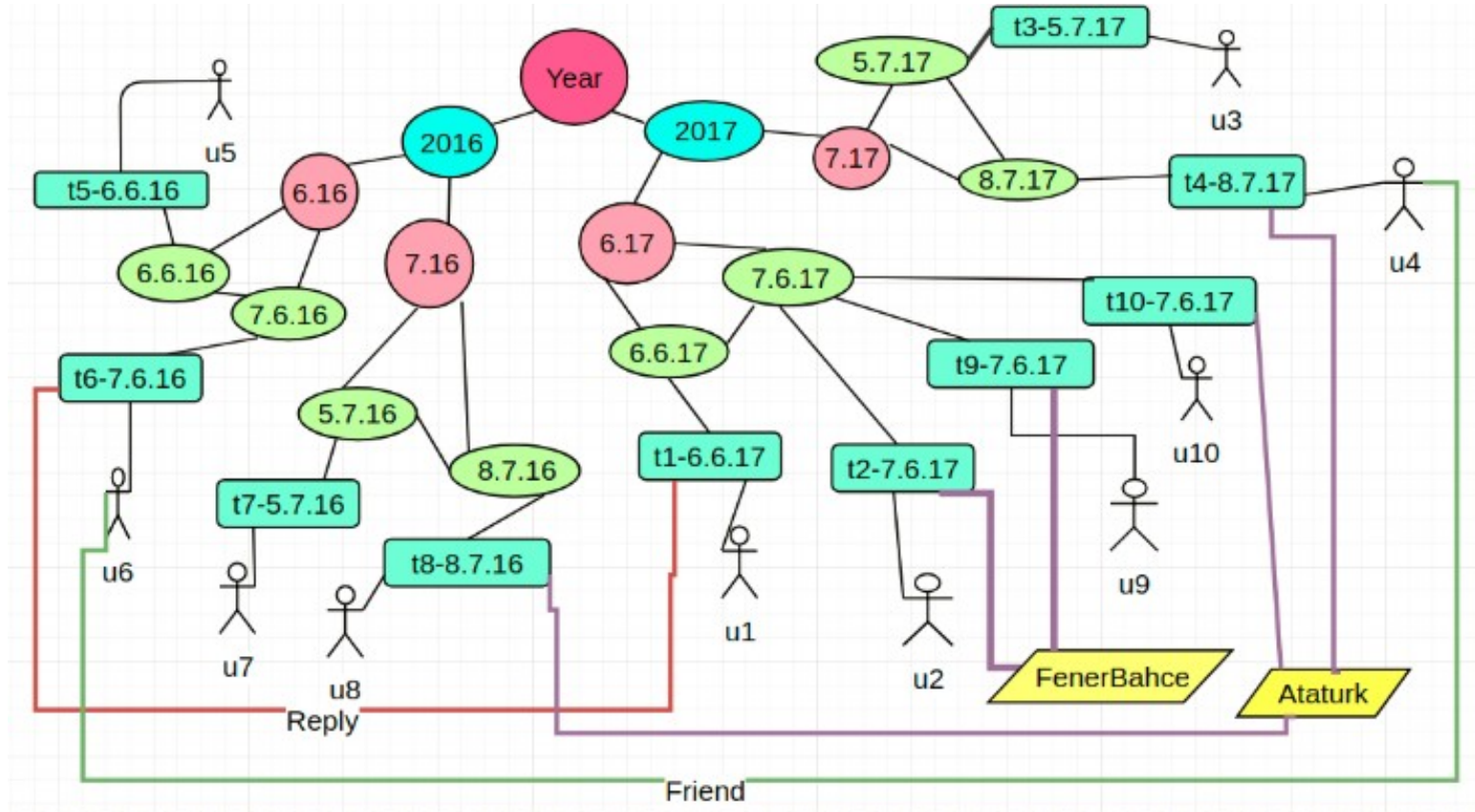
Neler Yapıldı? - 4



	u1037040092	u1499266621	u565848357	u103817724	u1028468581	u272613386
u1037040092	1.000000	0.850528	0.801281	0.511622	0.234287	0.240509
u1499266621	0.850528	1.000000	0.713349	0.441772	0.236389	0.214318
u565848357	0.801281	0.713349	1.000000	0.766105	0.117046	0.420595
u103817724	0.511622	0.441772	0.766105	1.000000	0.116033	0.591966
u1028468581	0.234287	0.236389	0.117046	0.116033	1.000000	0.256069
u272613386	0.240509	0.214318	0.420595	0.591966	0.256069	1.000000



Neler Yapıldı? - 5



Neler Yapıldı? - 6

- Kişi ve zaman kosinüs benzerlikleri

	u1623755515	u565848357	u700397072063795200	u296406287
u1623755515	1.000000	0.201955	0.242478	0.216975
u565848357	0.201955	1.000000	0.155391	0.405474
u700397072063795200	0.242478	0.155391	1.000000	0.090024
u296406287	0.216975	0.405474	0.090024	1.000000

	2016	8.2016	10.8.2016	16.8.2016	10.2016	29.10.2016	30.10.2016	2017	1.2017	12.1.2017
2016	1.000000	0.653825	0.509933	0.174226	0.754052	0.357579	0.227544	0.764847	0.571097	0.179766
8.2016	0.653825	1.000000	0.749489	0.338116	0.591549	0.300241	0.218591	0.435226	0.331710	0.184006
10.8.2016	0.509933	0.749489	1.000000	0.265180	0.472819	0.215346	0.216337	0.358742	0.355489	0.255866
16.8.2016	0.174226	0.338116	0.265180	1.000000	0.156207	0.124332	0.194046	0.088237	0.133471	0.172880
10.2016	0.754052	0.591549	0.472819	0.156207	1.000000	0.579656	0.454672	0.485708	0.348812	0.158056
29.10.2016	0.357579	0.300241	0.215346	0.124332	0.579656	1.000000	0.419442	0.248395	0.215501	0.217186
30.10.2016	0.227544	0.218591	0.216337	0.194046	0.454672	0.419442	1.000000	0.188774	0.211827	0.238516
2017	0.764847	0.435226	0.358742	0.088237	0.485708	0.248395	0.188774	1.000000	0.823800	0.402279
1.2017	0.571097	0.331710	0.355489	0.133471	0.348812	0.215501	0.211827	0.823800	1.000000	0.647309
12.1.2017	0.179766	0.184006	0.255866	0.172880	0.158056	0.217186	0.238516	0.402279	0.647309	1.000000



Neler Yapıldı? - 7

	t765562268767678464	t763141405766250496	basketbol	t765466344707989504	t765582802343657472	t765648654883512321
t765562268767678464	1.000000	0.154652	0.222707	0.350529	0.357823	0.452084
t763141405766250496	0.154652	1.000000	0.267026	0.142732	0.295929	0.140562
basketbol	0.222707	0.267026	1.000000	0.651889	0.668917	0.353605
t765466344707989504	0.350529	0.142732	0.651889	1.000000	0.636400	0.498252
t765582802343657472	0.357823	0.295929	0.668917	0.636400	1.000000	0.425720
t765648654883512321	0.452084	0.140562	0.353605	0.498252	0.425720	1.000000

t765466344707989504, t765582802343657472 ve t765648654883512321 aynı gün atılmıştır ama ilk ikisi extra olarak basketboldan konuşmuşlardır.

Makale Gönderimi:

32nd International Symposium on Computer and Information Sciences
June 6, 2018



- 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



1. 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.

