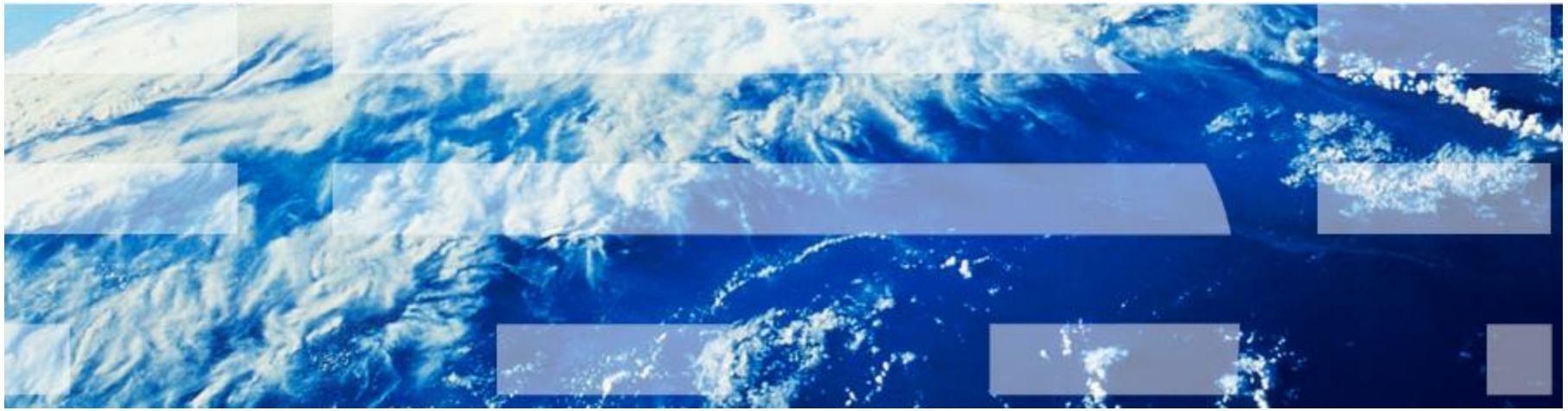
E6893 Big Data Analytics:

Exploring the Meetup.com Social World
Large Scale Event-Based Social Network Analysis

Team Members: Mengge Li (ml3695, EE), Yiwei Zhang (yz2698,EE)
Rahul Gaur (rg2930,CS), Rongyao Huang (rh2648, STAT))



December 9, 2014

Motivation

Unique value of event-based social network (EBSN)

- Both online offline social interactions
- Commercial value: industrial trends, recommendation of services/ products based on user preference

Big Fan of Meetup.com

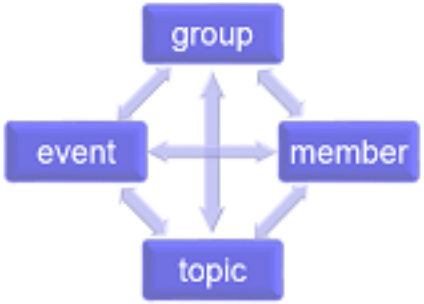
- Popularity across academia, industry and recreation
- Excellent API: user, group, event, tags – location & time

Great opportunity to apply big data techniques

- Graph database: Neo4j with Cypher
- Clustering/ Community Detection
- Large scale social network analysis

Dataset, Algorithm, Tools

Meetup Dataset



- # Users: 4,448,454 # Groups: 42,052
- # Events: 1,595,833 # Tags: 77,810
- # User-Group Pairs: 8,863,235 # User-Event Pairs: 13,553,134
- # User-Tag Pairs: 15,057,535 # Group-Tag Pairs: 144,793
- # Users with Locations: 3,741,699 # Events with Locations: 983,333

Analytics/ Modelling

- Properties of social interactions: degree, centrality, separation, clustering coef, density, etc.
- Temporal & spatial patterns of specific groups/events
- Clustering: Fiedler method, linear combination, generalized SVD
- Recommendation: user-based, item-based

Tools

- Neo4j
- Java/Python
- Gephi

Theoretical Framework for Network Analysis

$$G = (V, E)$$

$$G = < U, N^{on}, N^{off} >$$

Online and Offline User Network

$$G = < Vg, Eg >$$

Online Group Network

$$w_{Uab} = \sum \frac{1}{S_i}$$

Defined by the degree of closeness between users, where S is the size of the group which contains both user a and b.

$$w_{Gab} = \sum \frac{1}{K_i}$$

Defined by the degree of closeness between groups, where K is the number of shared members between group a and b.

Construct Weighted Social Network in Neo4j

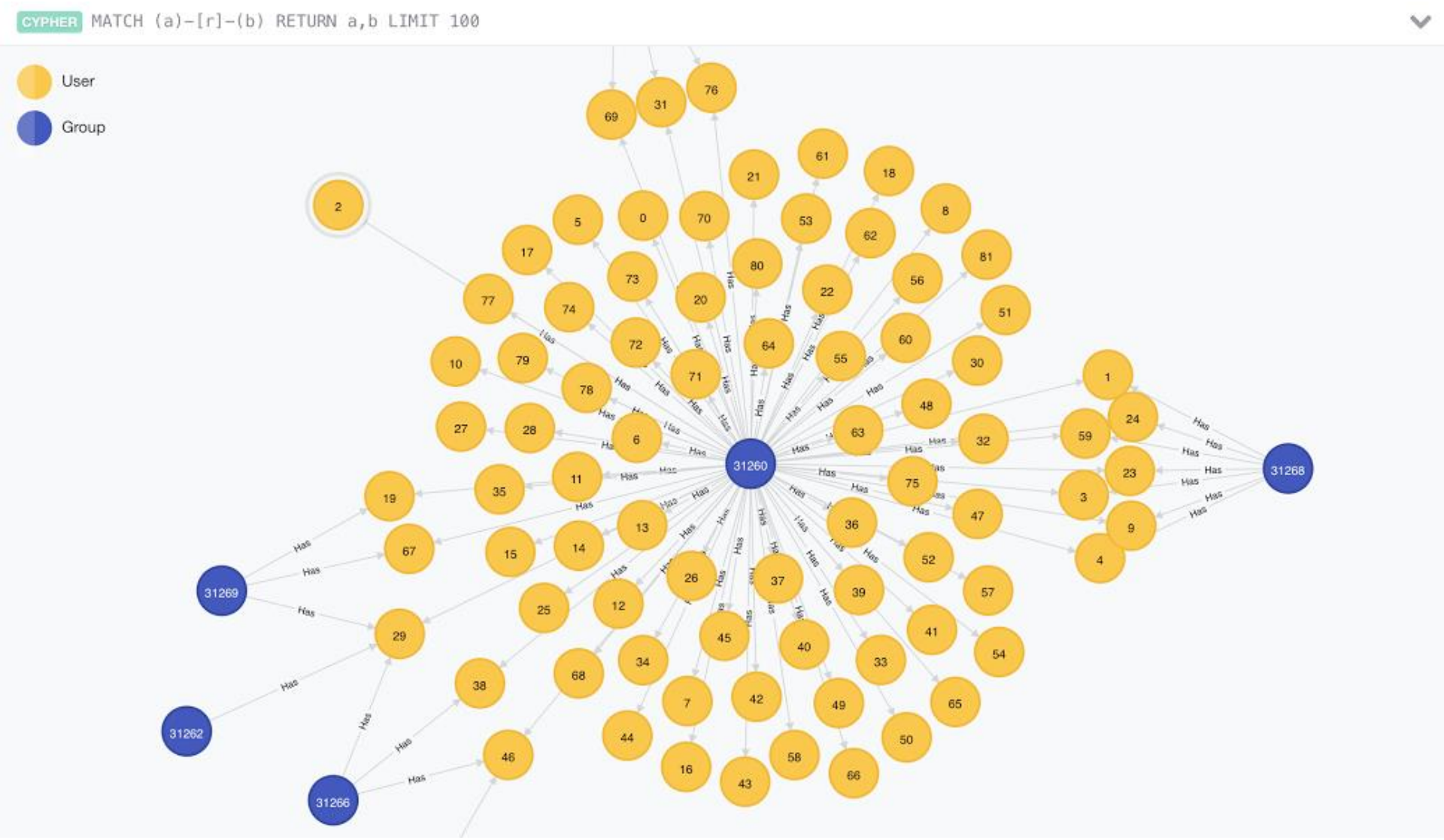
```
1 MATCH (group1)-[:Has]->(comusers)<-[:Has]-(group2)
2 return group1 AS source,group2 AS target ,count(comusers) AS
  weight
```

CYPHER MATCH (group1)-[:Has]->(comusers)<-[:Has]-(group2) return group1 AS sourc

source	target	weight
groupID 1768166	groupID 1844911	2
groupID 5194482	groupID 995599	41
groupID 5194482	groupID 5618342	19
groupID 5284602	groupID 382157	1
groupID 5194482	groupID 1797478	6

Returned 848 rows in 601 ms

Construct Weighted Social Network in Neo4j



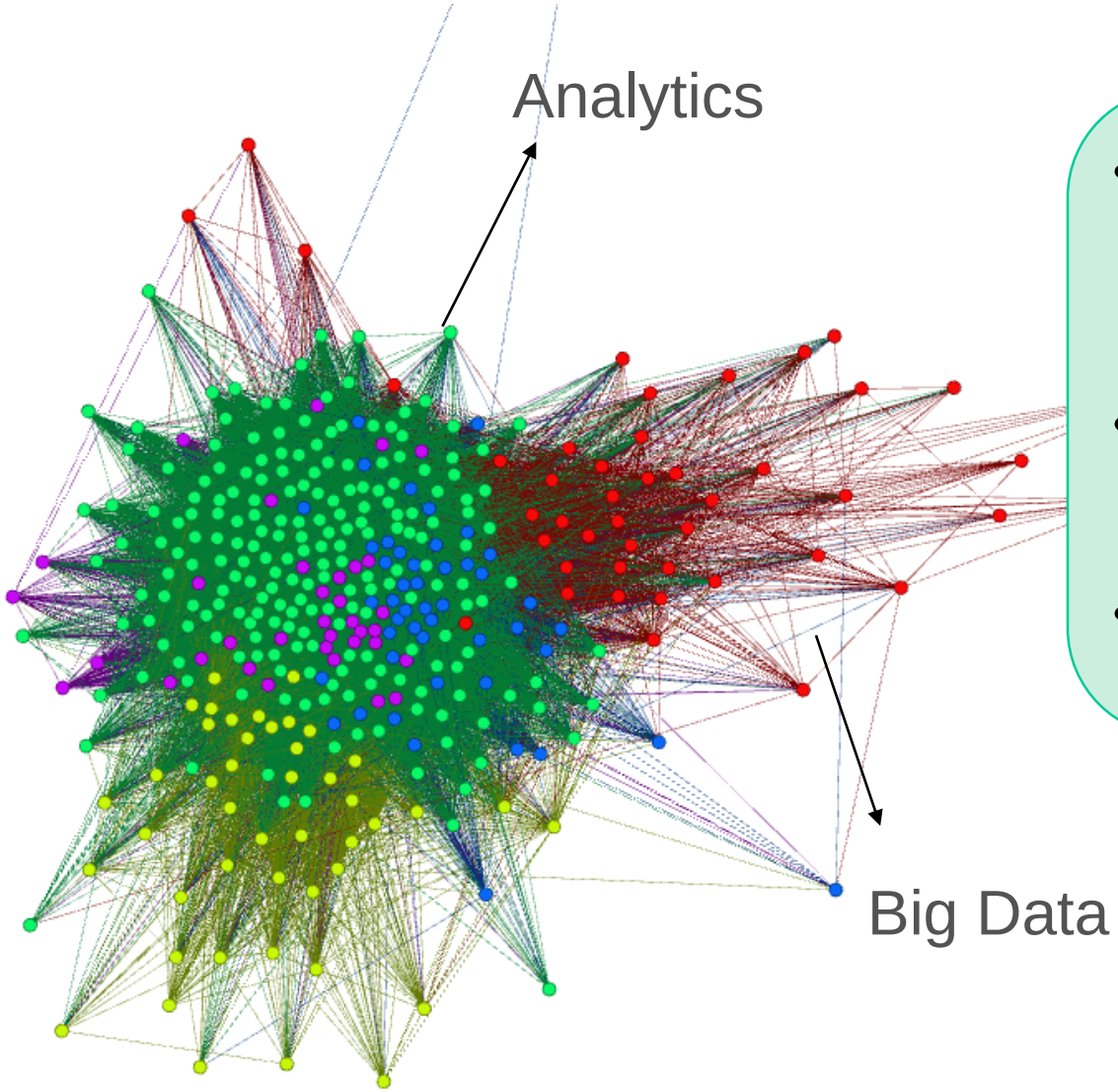
Properties of User Network

	Average Degree	Clustering Coeficient
Online Network	1660.1	0.443
Offline Network	157.3	0.246

- The online network is much larger than the offline network, which means people used to join a lot of on-line groups and know others on the internet, and fewer people choose to attend the events off-line.

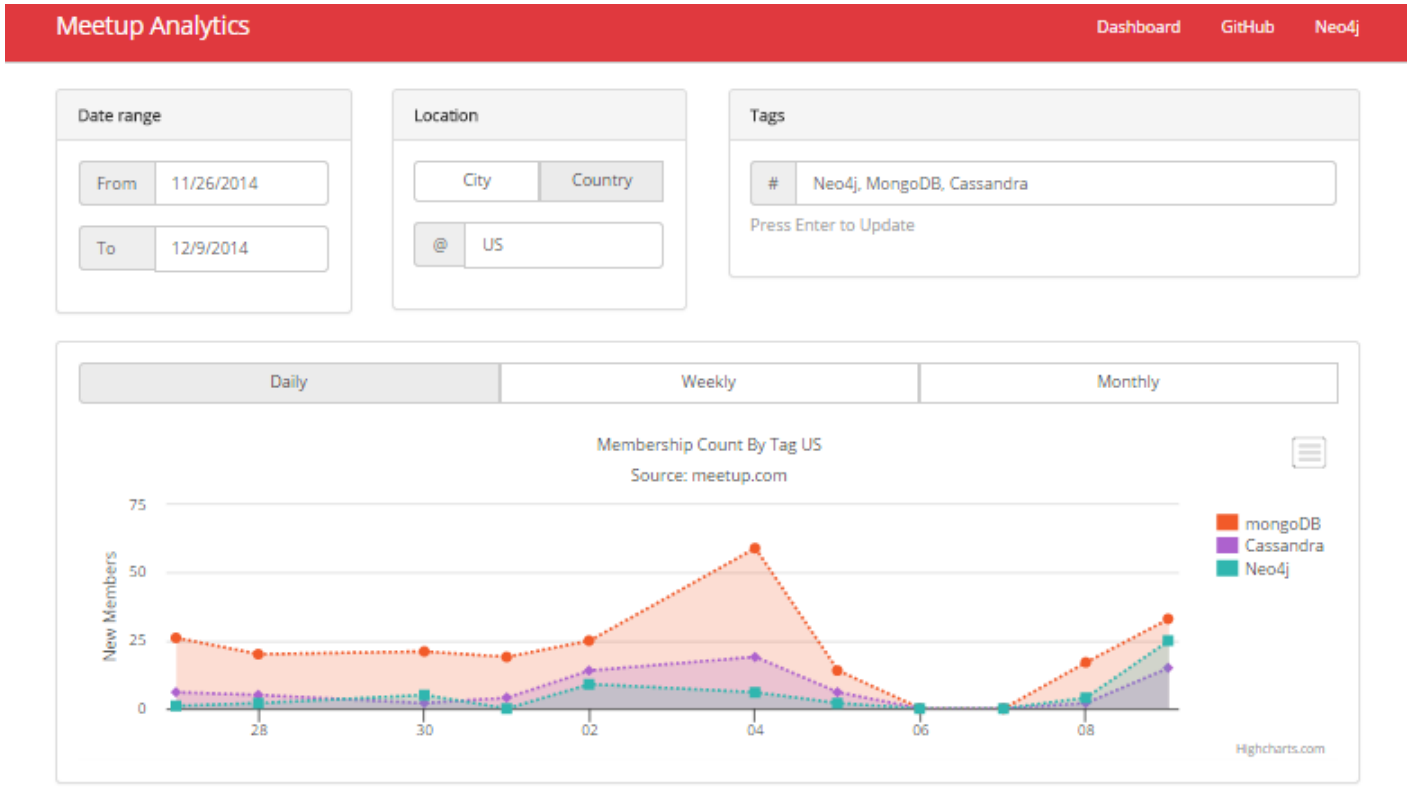
- Clustering Coef of the offline network is less than online network, there are less closed triangles in former. So the connection between users offline is much weaker and the graph is separated.

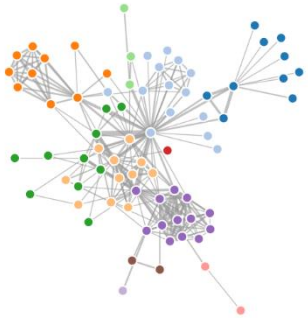
Group Network and Community Detection

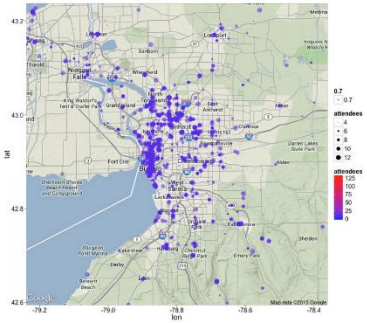


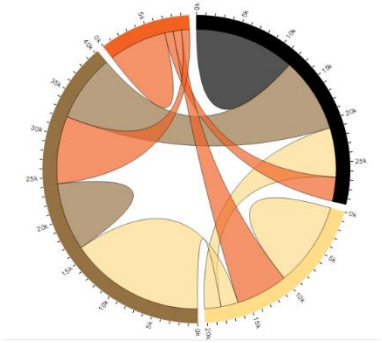
- Filter Groups using group tags that are data and analytics related.
- Clustering using the Fiedler Methods implemented in Java
- End up with 5 clusters

The Meetup Social Network Dashboard (Coming Soon)









[D3 Live Examples]