

Prof. Giovanni Stilo

If you can read this... congratulation! you have 11/10 or more... (only for inclass student)

SOCIAL MINING

2020/21

giovanni.stilo@univaq.it

Form

**To know the needs of the class; please complete
the form linked below NOW (emails are NOT collected) :
<https://forms.gle/M8QHg5LzAYMxqHBA9>**



About this course

Teams: Social Mining 2020-21

Code: zgnm3l1

(Slides, course material,COMUNICATIONS)

Official Schedule:

- **Online**
 - *Wednesday, 12:10 – 13:50*
 - *Time format: 45 – 10 – 45*
- **Online**
 - *Thursday, 14:55 – 16:35*
 - *Time format: 45 – 10 – 45*

Course is organized as follows:

- 1~2 / 3 “standard” lectures
- 2~1 / 3 Lab:
 - **design of an IR system with Lucene,**
 - **Gathering social data**
 - **Scraping the web**
 - **Implementing a Crawler**
 - **Analyzing Social Data**

Course Topics

Part I:

- **Web information retrieval**
- **Architecture of an information retrieval system**
- **Text processing, indexing, retrieval**
- **Ranking: vector space model, latent semantic indexing**
- **Web information retrieval: browsing, scraping**
- **Web information retrieval: link analysis (HITS, PageRank)**

Part II:

- **Social Network Analysis**
- **Modeling a social network: local and global measures**
- **Community detection**
- **Graph Sampling**
- **Mining social networks:**
 - **temporal mining,**
 - **user profiling and recommenders systems**

Requirements

- **Laptop** (1 every 2 students).
 - Programming Languages:
 - **Java – R/Python (eventually)**

the most important one:

Enjoy Yourself

Homeworks

Are **not mandatory**

But they helps you

They will be considered as a plus
during the final evaluation

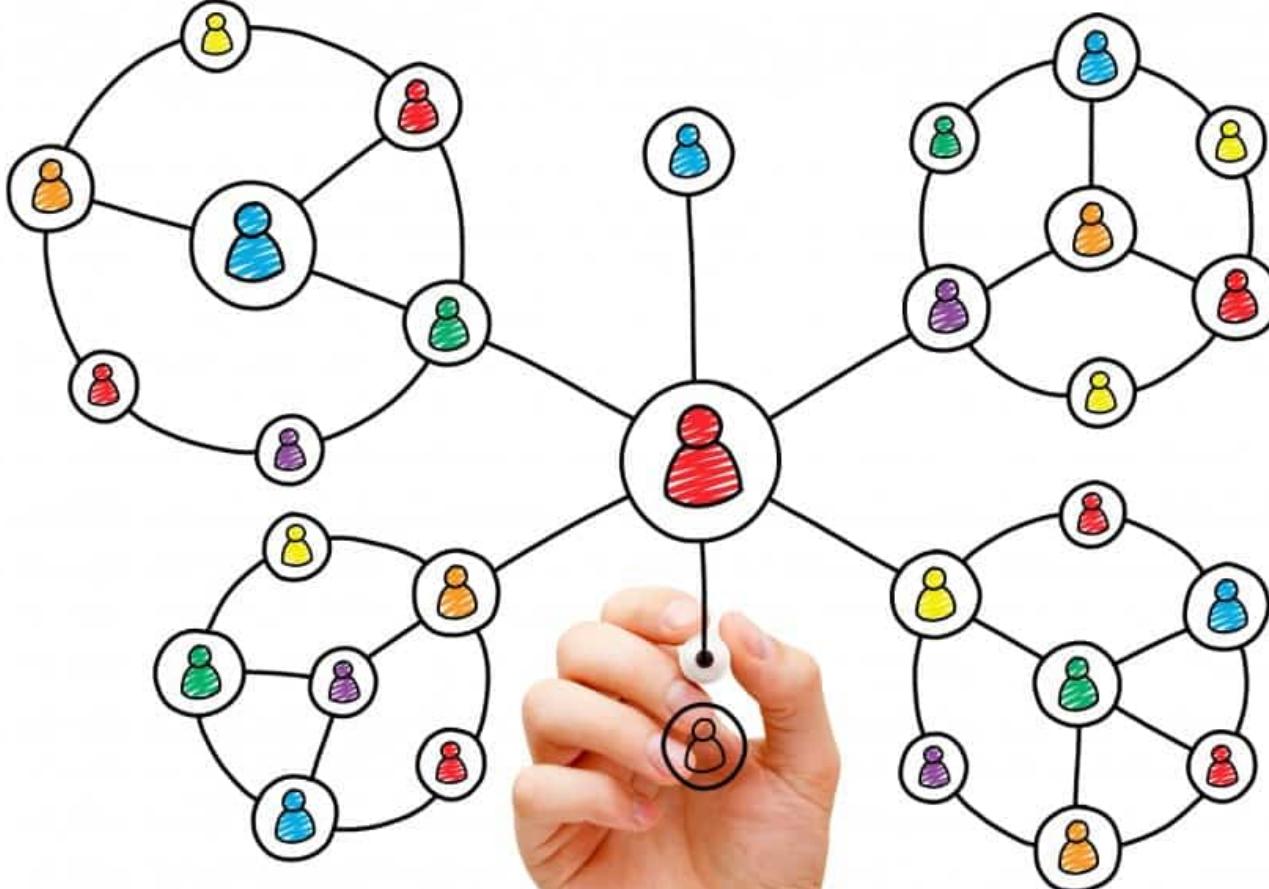
Exam:

- **Project**
 - big datasets preprocessed using java;
 - Text Stream/posts and graph algorithms;
 - Several tasks are needed to be completed;
 - *Some parts of the project are intentionally left free or underspecified to be solved providing different approaches;*
- **Oral Discussion of the project:**
 - adopted solutions;
 - encountered problems;
 - theoretical questions about algorithms used in the project;

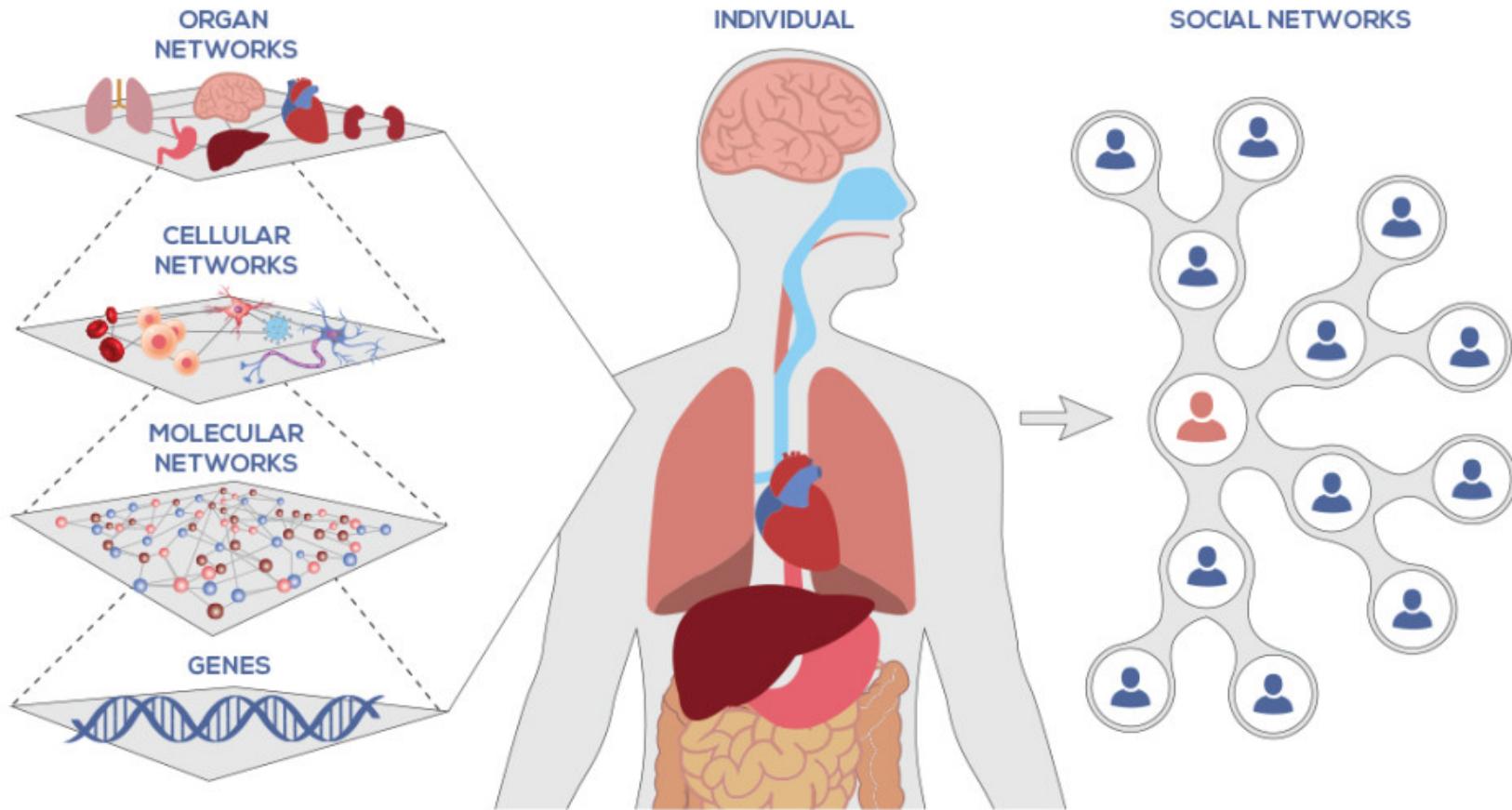
Giovanni Stilo

Social Mining

Social Interactions



Biological Interactions



Entities



Internet Users

3,3 B.



Facebook Users

1,5 B.



Twitter Users

0,5 B.



Proteins

25 K.

Unstructured Data



Tweets

180 B.



Email

3.5 T.



Blogs Post

1.3 B.



Biological Concepts

40 K.

Messages



Kat @Kowareta_Doll · 2 h

Day 5 of cold/**flu** bug: actually slept ok thanks to night nurse again. Still icky **cough** and weak. going to try venturing to the park later.



ohBulan @ahhhMunn · 4 h

Don't feel like going out bcs of **flu & cough** but i wanna meet bb



Syamsuzzaman @_abgsyam · 6 h

Flu and **cough** gtg

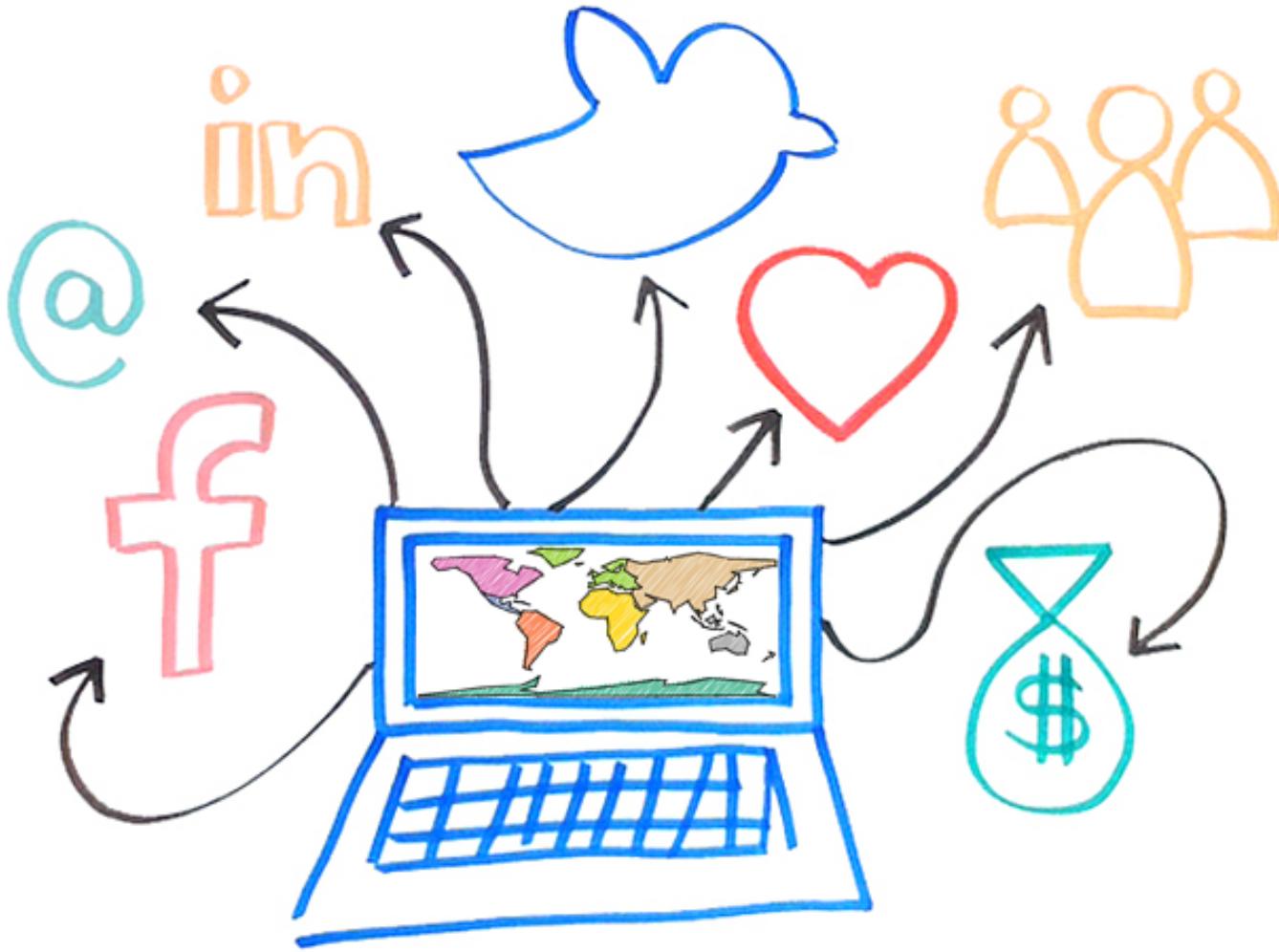


Knowledge @itzwikipedia · 6 h

Pineapple juice is 5 times more effective than **cough** syrup. It also prevents colds and the **flu**.



Purpose



Understanding



Football

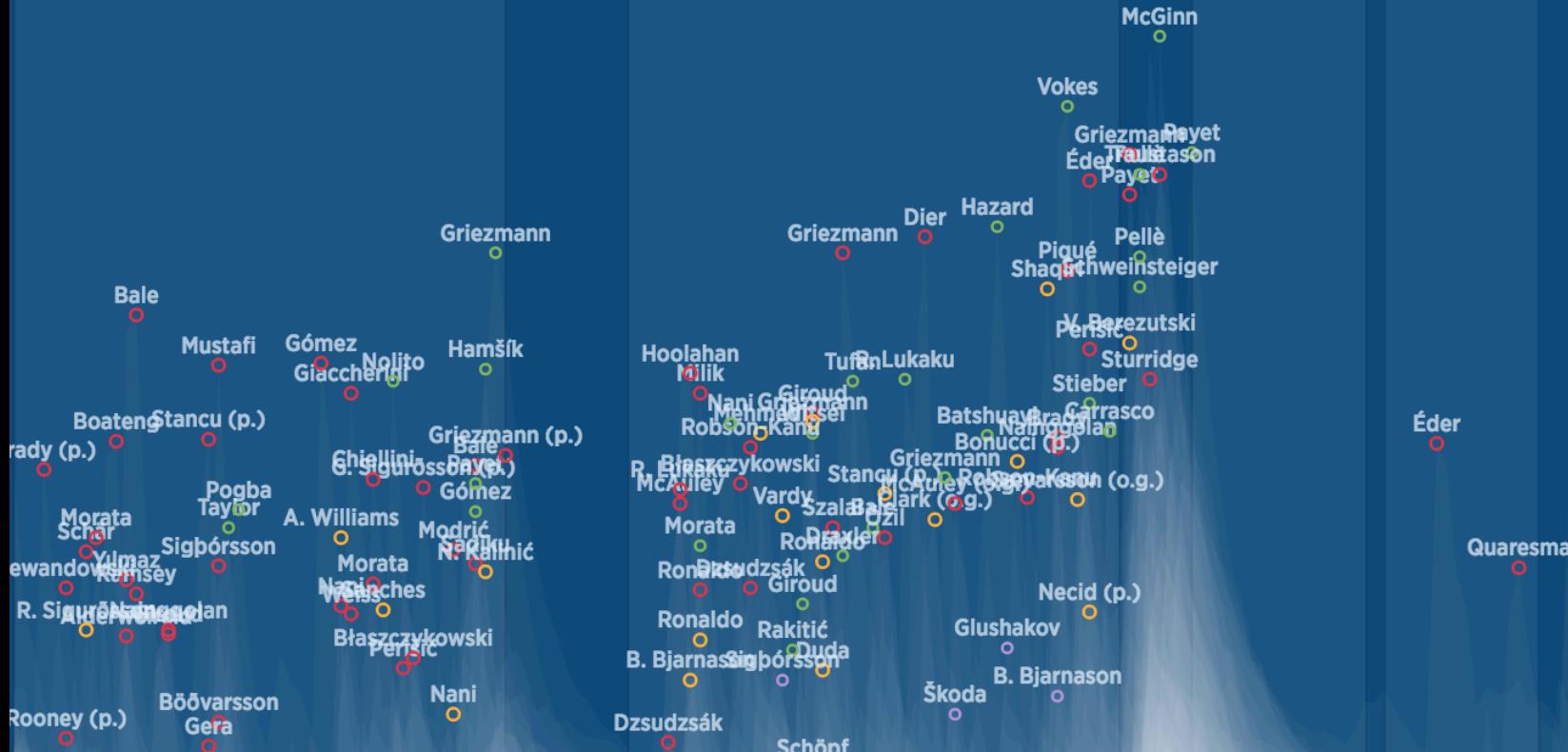
1st Half

2nd Half

1st ET

2nd ET

Pen



1st Half

2nd Half

1st ET

2nd ET

Penal

ROUND OF 16 — 27 JUNE 2016

Italy 2-0 Spain



#ITA
Chiellini
33'

#ITA
Pellè
90+1'

Terms Time Series

— happy — what — house — time — data — tree

12/4

13/4

14/4

15/4

16/4

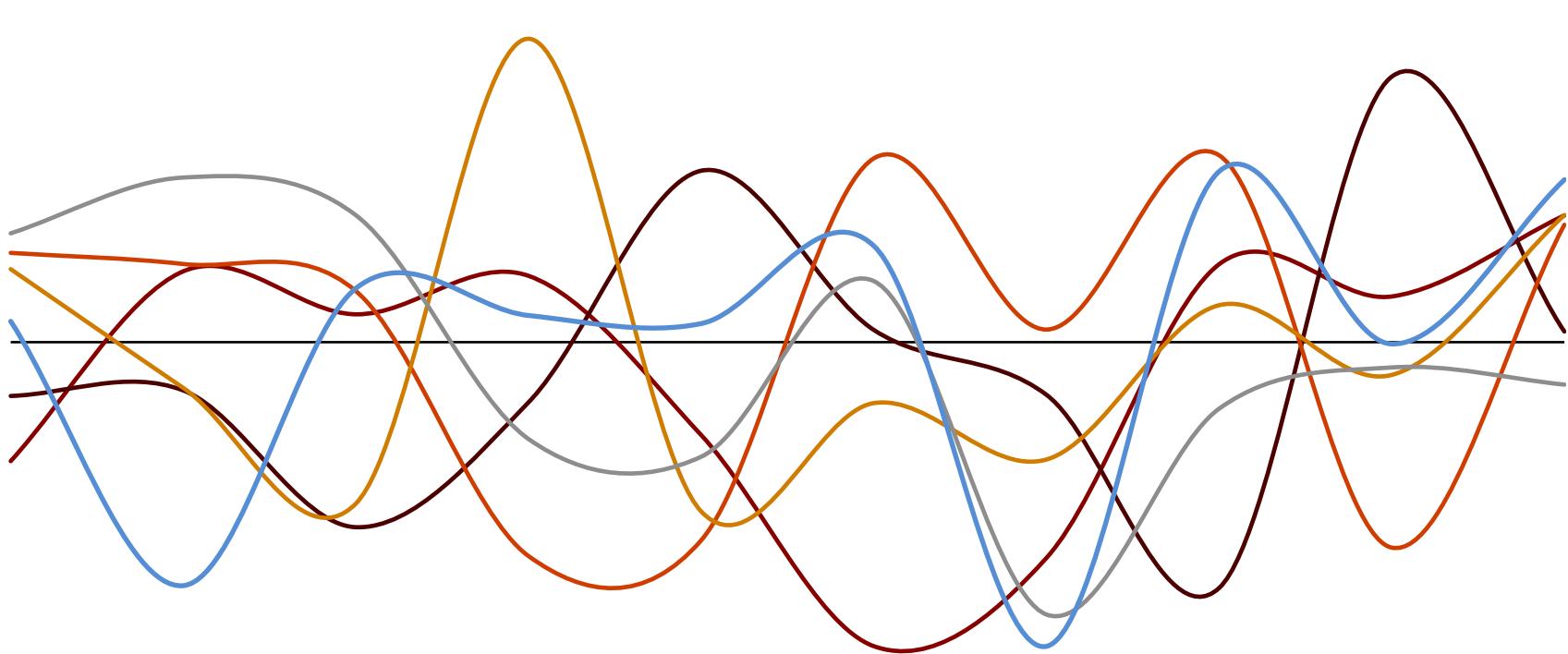
17/4

18/4

19/4

20/4

21/4



Paris Attacks

Bataclan Courage ParisAttacks Prayforparis terrorism

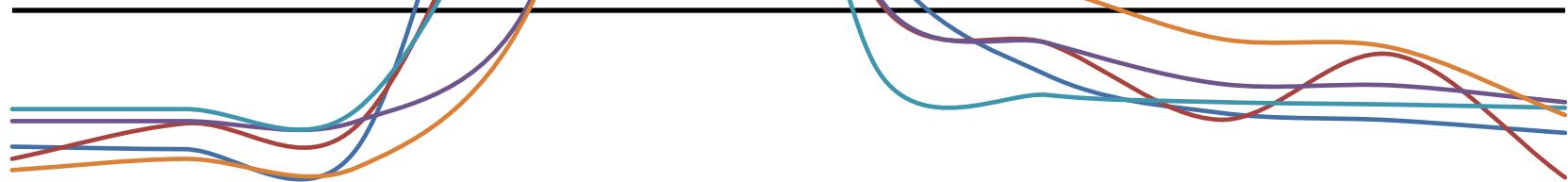
10/11

12/11

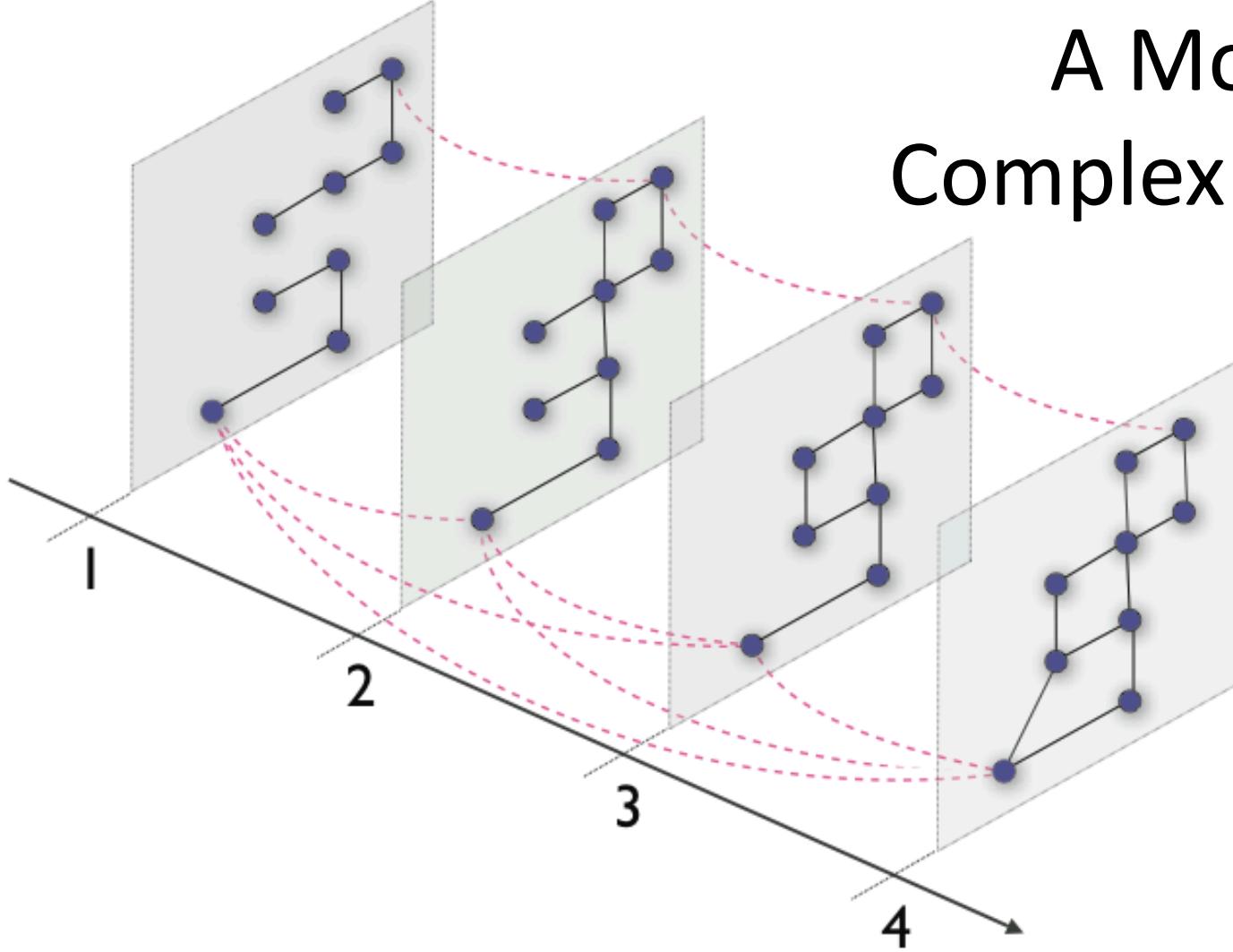
14/11

16/11

18/11



People's Signals



A More
Complex Model

improve



SOCIAL ENTERPRISE

Social Network Analysis

Personalize Network Parameters

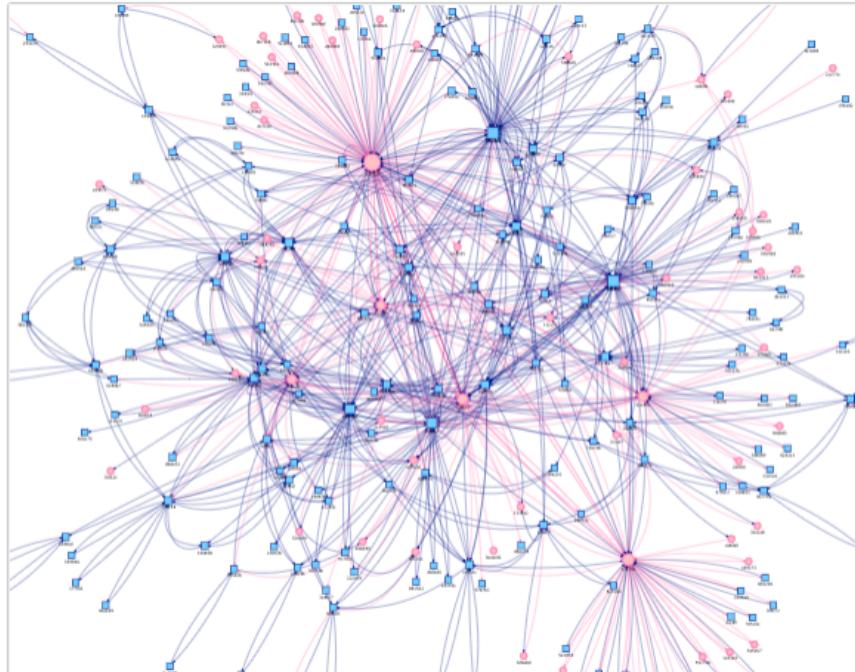
Month: 2014-01-01 Period: 1 mese Gender: female Role: 6 of 15 selected

Community Stabilized

[select all] Consultant Senior Consultant Manager Content Publisher Partner

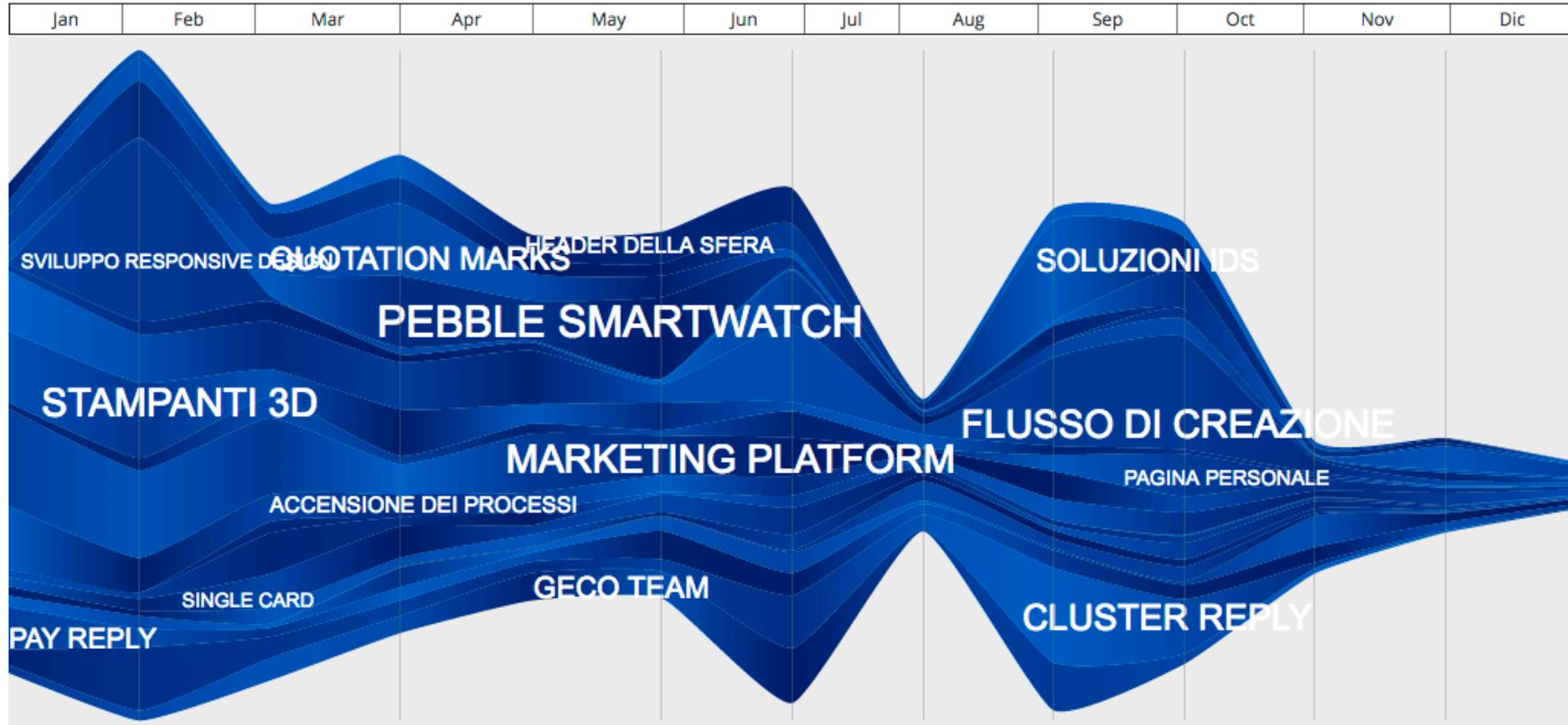
Query: Insert a query Topics: Select a Topic

Info Network
Interval: 1 month
From: 2014-01-01
To: 2014-02-01
Number nodes: 323
Number edges: 1042
Gender: all
Role: Consultant, Senior Consultant, Manager, Content Publisher, Partner, External



Ranking		
Leader Rank		Brokerage
Search: <input type="text" value="Search"/>		
Showing 1 to 5 of 323 entries		
#	USERNAME	VALUE
1	fd850304	0.499
2	62b17861	0.497
3	b80a4fcb	0.374
4	d2cad61c	0.291
5	51b78141	0.249

⚙ Personalize Streamgraph Parameters



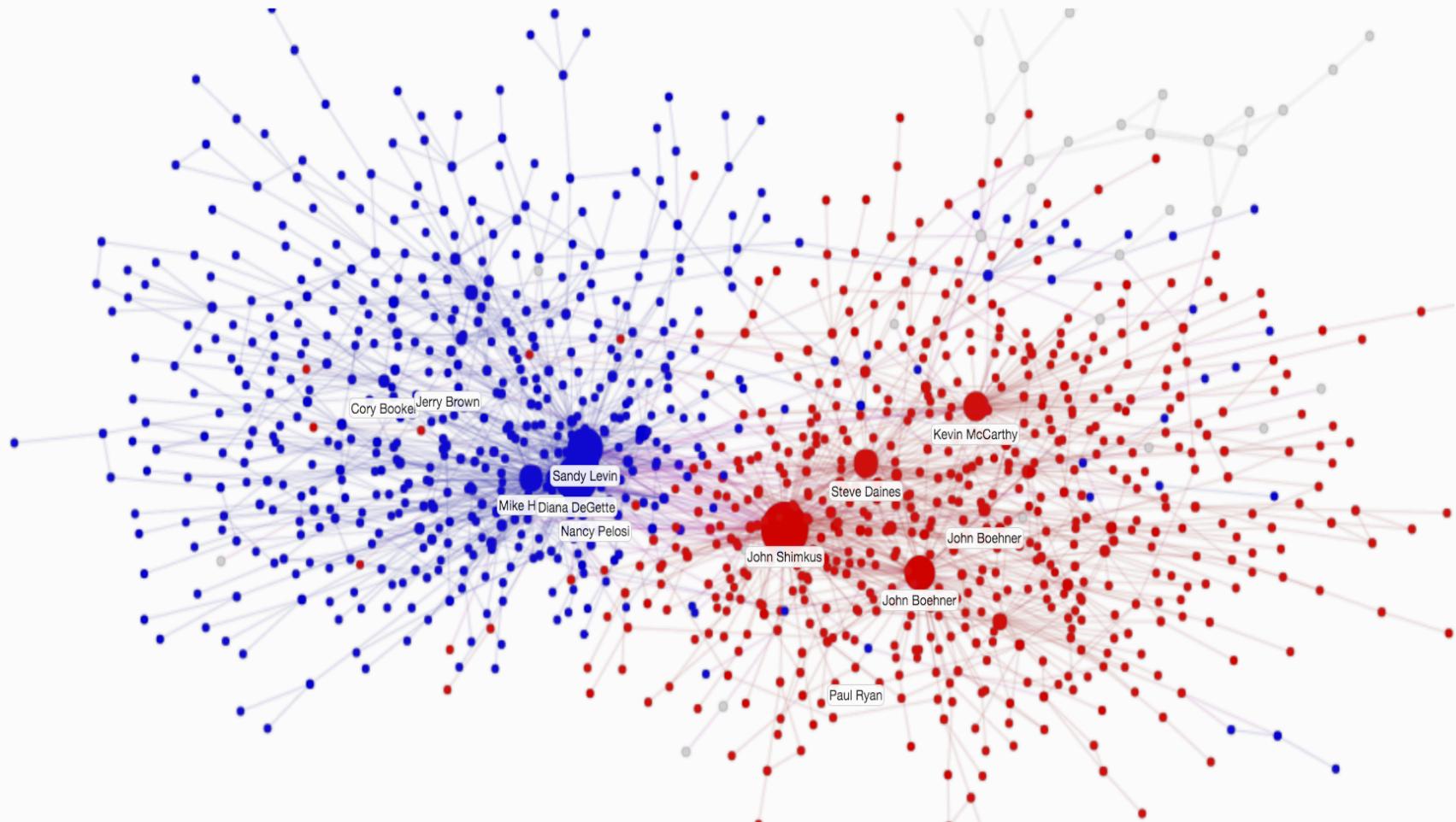
Recognition



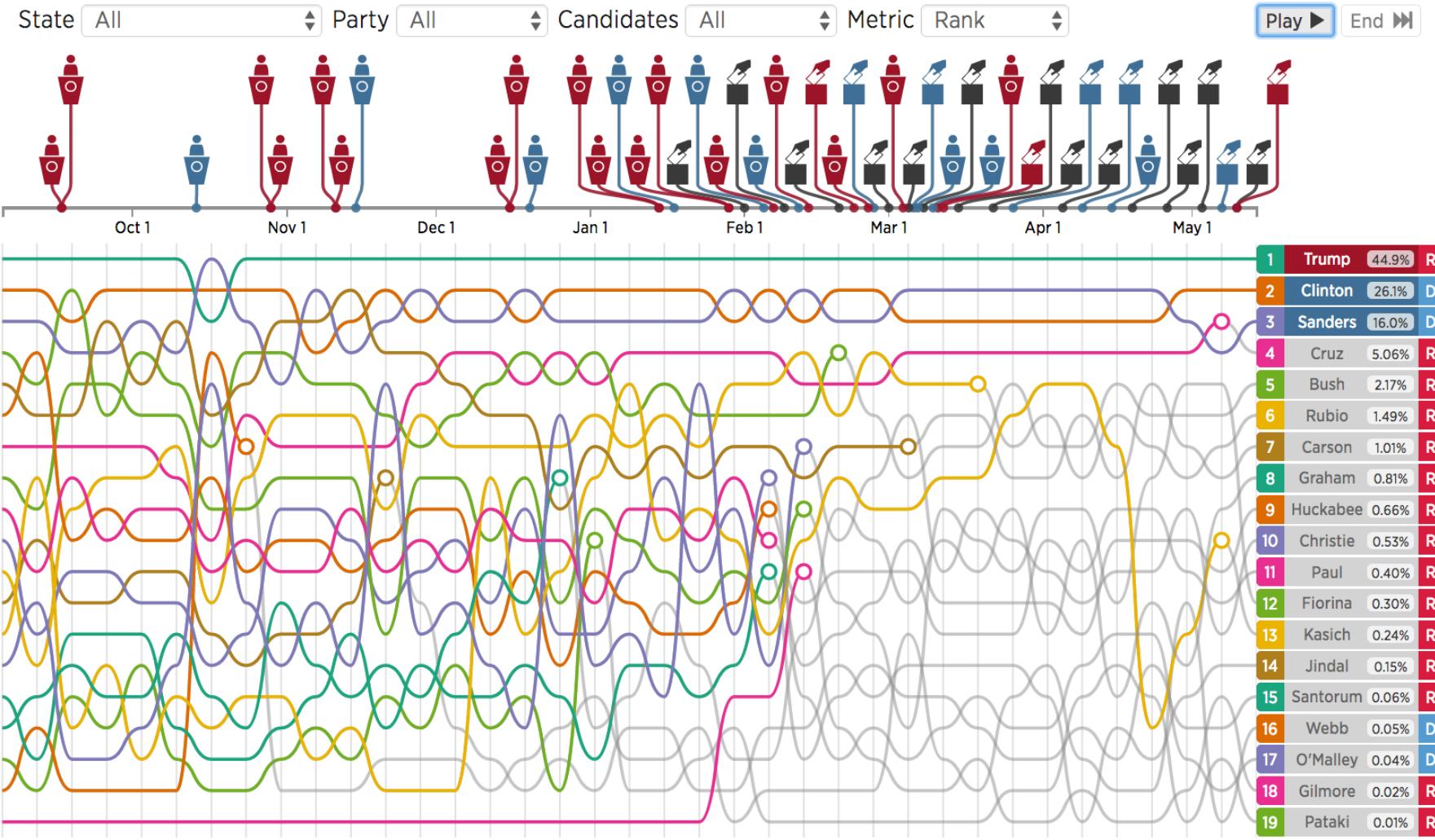
Political Parties

Color by: ● Party ○ Office

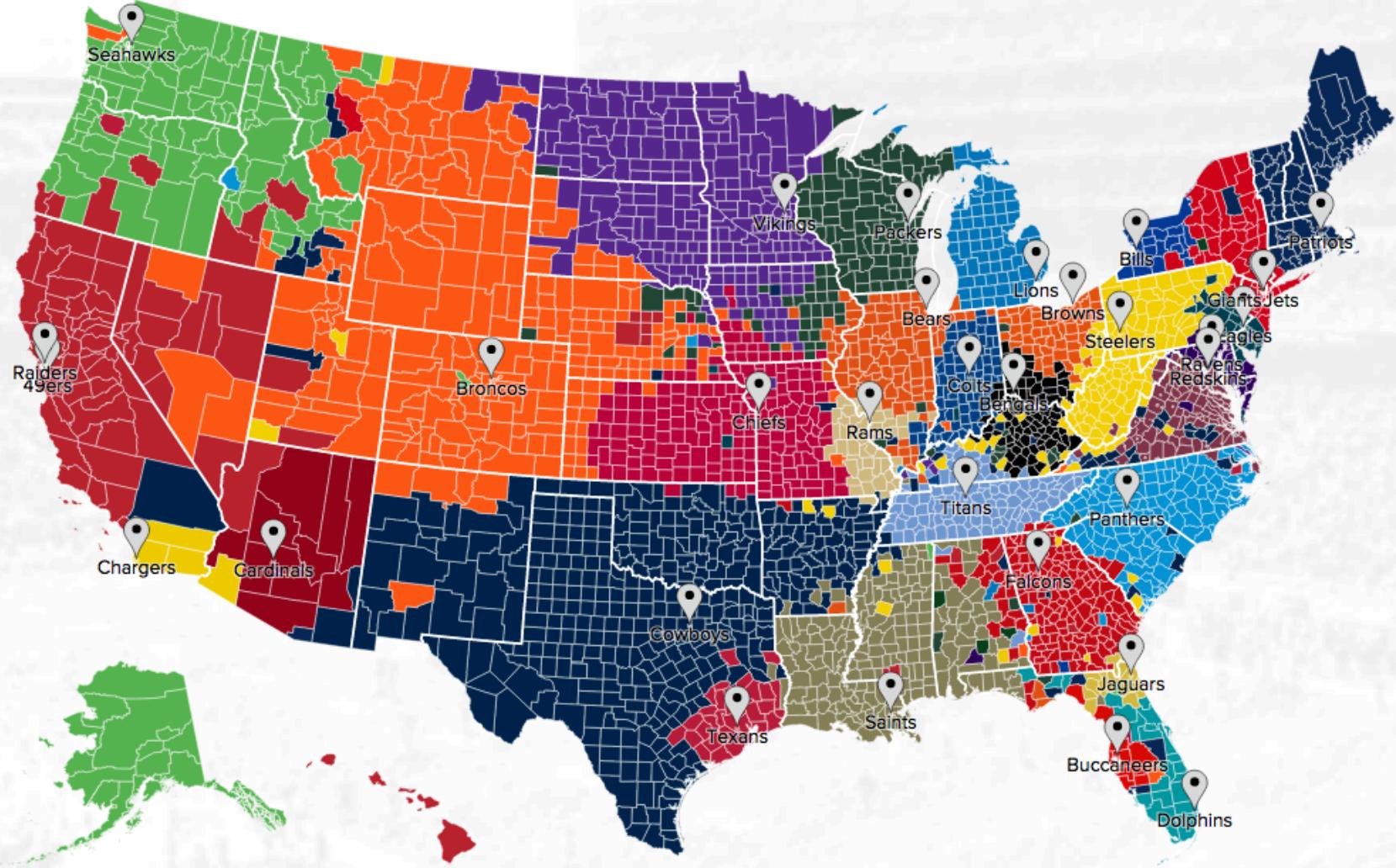
■ Democrat ■ Republican



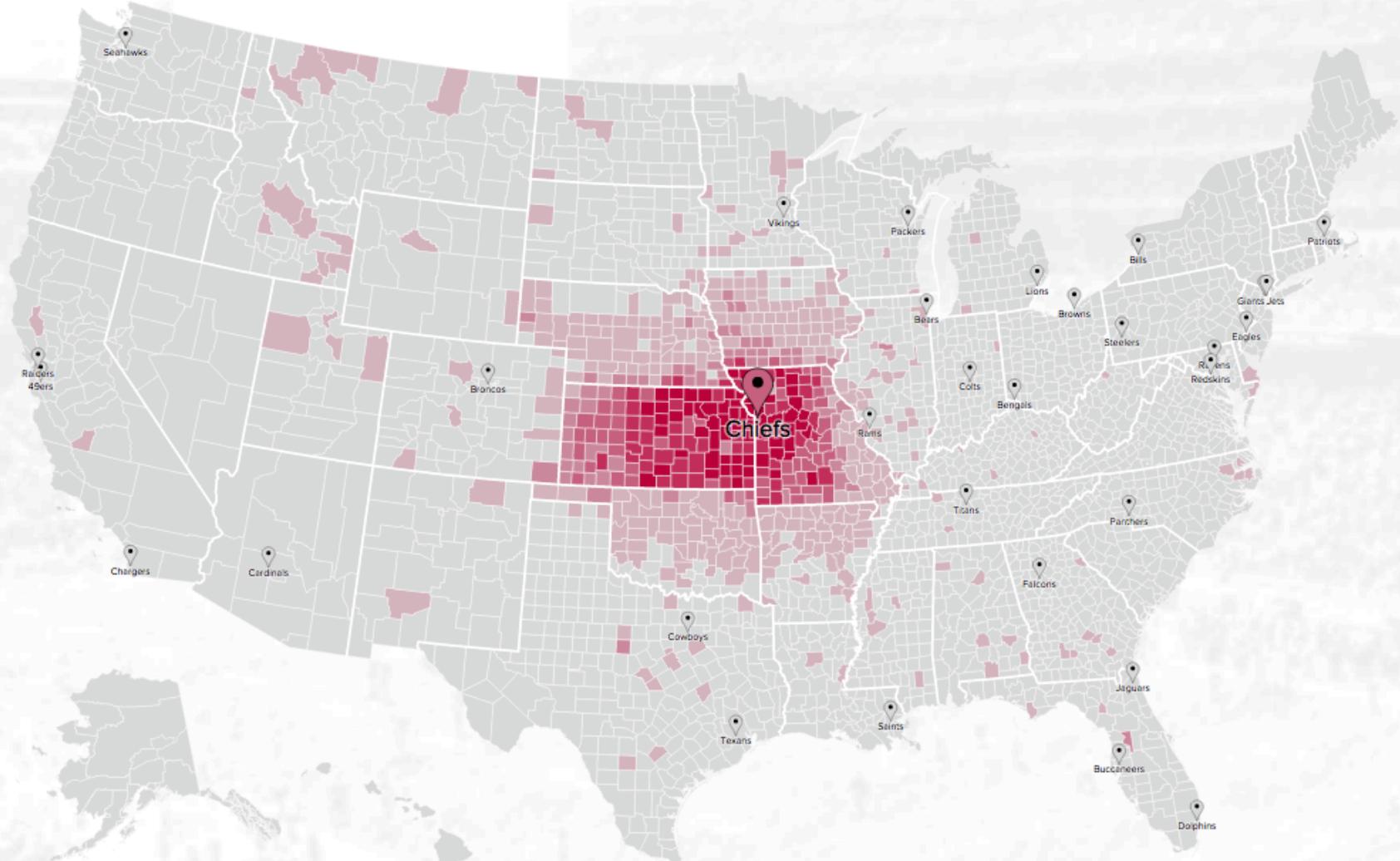
Election



Geotagged



Diffusion



Methodology



Medical Case Definition



Mapping medical terminology
with everyday language



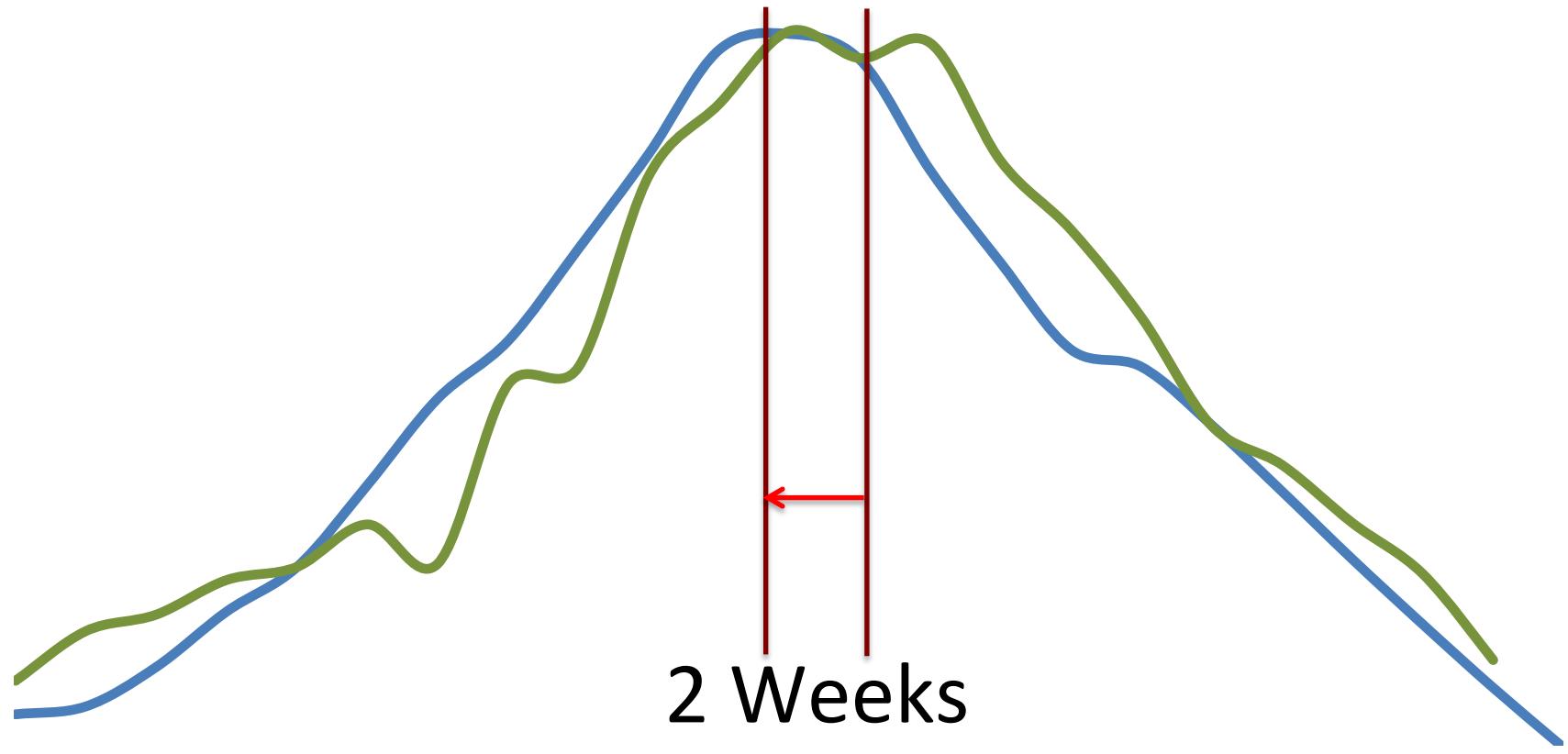
Messages Selection



Trends Production

Results

— Our System — Official System
(2 weeks later)



- **Register to Twitter as common User**
- **Goto:** <https://apps.twitter.com/>
 - Login with credentials and generate new **app.**
- **Generate all keys:**
 - Consumer Key (API Key)
 - Consumer Secret (API Secret)
 - Access Token
 - Access Token Secret

Question?



Assignment

Assignment

Describing an application scenario you have encountered in your life, which involves social interactions (not solely SN) or could benefit from an analysis of users/companies/items interactions.
(Prepare few [less than 5] slides)