

# CSE 258 – Lecture 17

## Web Mining and Recommender Systems

More temporal dynamics

## Temporal models

This week we'll look back on some of the topics already covered in this class, and see how they can be adapted to make use of **temporal** information

1. **Regression** – sliding windows and autoregression
2. **Classification** – dynamic time-warping
3. **Dimensionality reduction** - ?
4. **Recommender systems** – some results from Koren

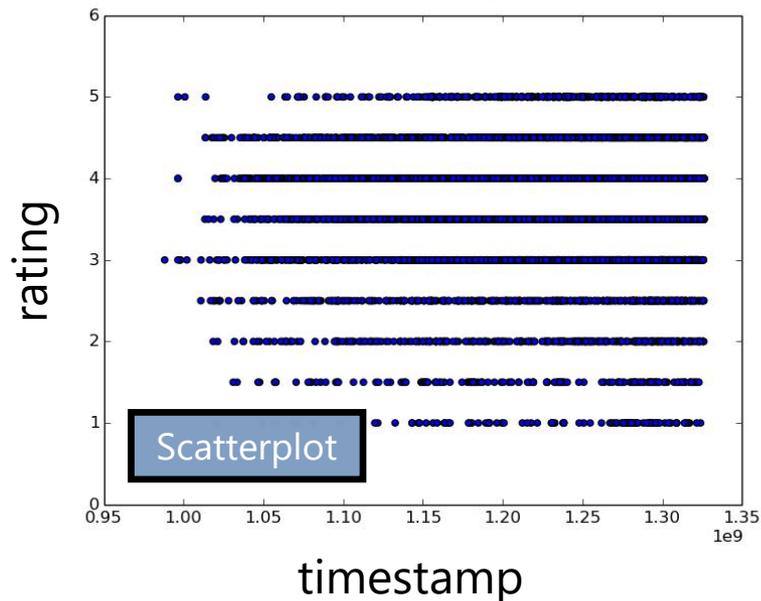
Today:

1. **Text mining** – “Topics over Time”
2. **Social networks** – densification over time

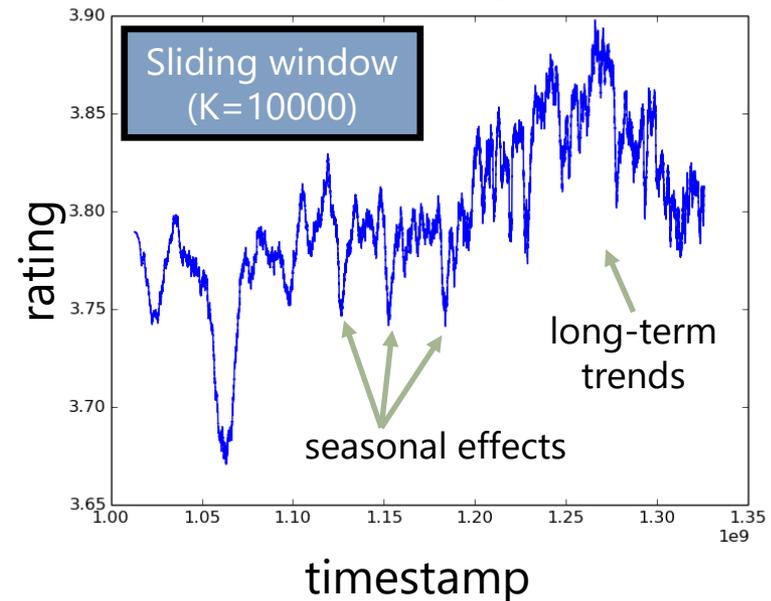
# Monday: Time-series regression

## Also useful to plot data:

BeerAdvocate, ratings over time



BeerAdvocate, ratings over time



Code on:

<http://jmcauley.ucsd.edu/code/week10.py>

# Monday: Time-series classification

As you recall...

The longest-common subsequence algorithm is a standard dynamic programming problem

	-	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>T</b>
-	0	0	0	0	0	0
<b>G</b>	0	↙↘0	↖1	←1	←1	←1
<b>A</b>	0	↖1	↙↘1	↖1	↖2	←2
<b>C</b>	0	↑1	↙↘1	↖2	↙↘2	↖2

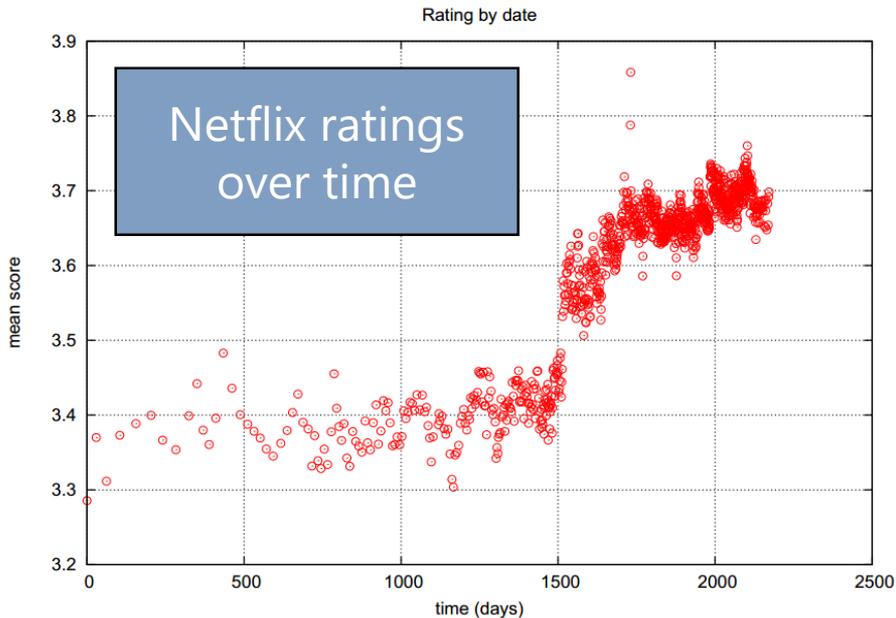
1<sup>st</sup> sequence

2<sup>nd</sup> sequence

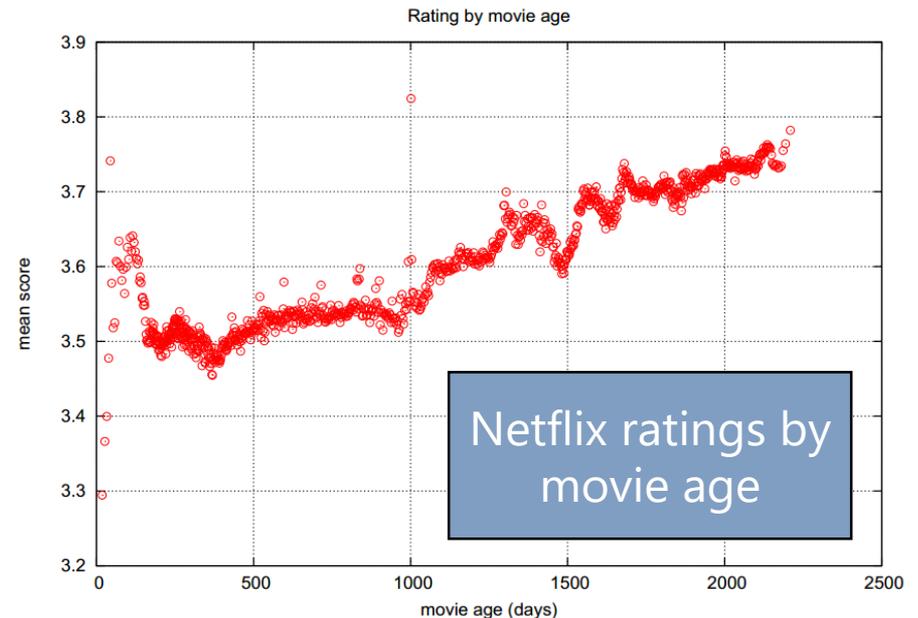
- ← = optimal move is to delete from 1<sup>st</sup> sequence
- ↑ = optimal move is to delete from 2<sup>nd</sup> sequence
- ↙↘ = either deletion is equally optimal
- ↖ = optimal move is a match

# Monday: Temporal recommendation

To build a reliable system (and to win the Netflix prize!) we need to account for **temporal dynamics**:



(Netflix changed their interface)



(People tend to give higher ratings to older movies)

# Week 5/7: Text

yeast and minimal red body thick light a Flavor  
 sugar strong quad. grape over is molasses lace  
 the low and caramel fruit Minimal start and  
 toffee. dark plum, dark brown Actually, alcohol  
 Dark oak, nice vanilla, has brown of a with  
 presence. light carbonation. bready from  
 retention. with finish. with and this and plum  
 and head, fruit, low a Excellent raisin aroma  
 Medium tan  
 Bags-of-Words

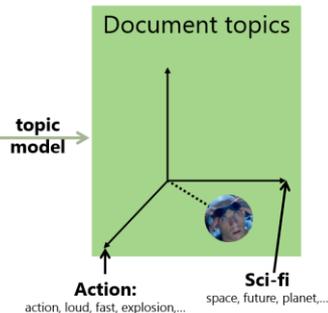
## What we would like:

87 of 102 people found the following review helpful  
 ★★★★★ You keep what you kill. December 27, 2004  
 By [Schtinsky "Schtinsky"](#) (Washington State) - [See all my reviews](#)  
 vine-vine

This review is from: [The Chronicles of Riddick \(Widescreen Unrated Director's Cut\) \(DVD\)](#)  
 Even if I have to apologize to my Friends and Favorites, and my family, I have to  
 admit that I really liked this movie. It's a Sci-Fi movie with a "Mad Max" appeal  
 that, while changing many things, left Riddick from "Pitch Black" to be just Riddick.  
 They did not change his attitude or soften him up or bring him out of his original  
 character, which was very pleasing to "Pitch Black" fans like myself.

First off, let me say that when playing the DVD, the first selection to come up is  
 Convert or Fight, and no explanation of the choices. This confused me at first, so  
 I will mention off the bat that they are simply different menu formats. Select  
 either one and continue with the movie.

(review of "The Chronicles of Riddick")

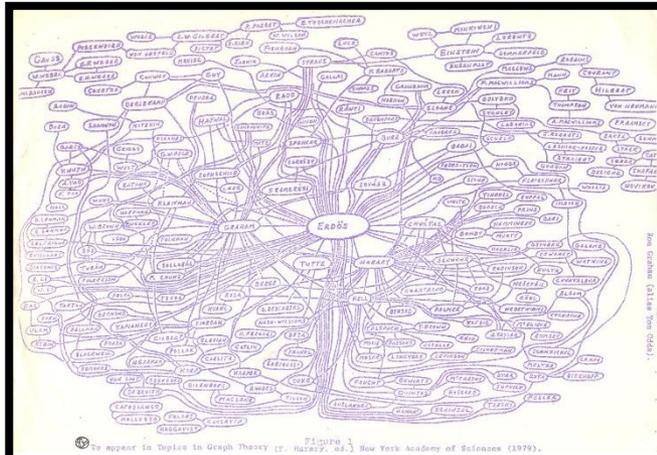


## Topic models

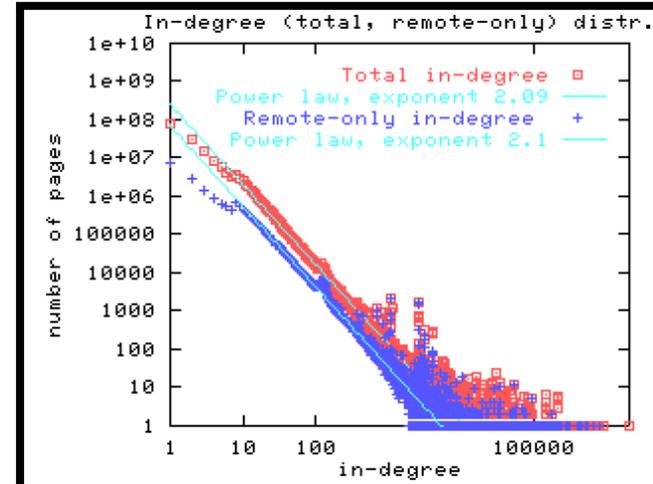


## Sentiment analysis

# 8. Social networks



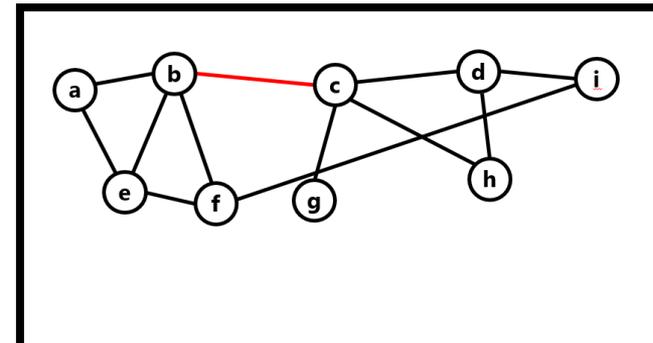
Hubs & authorities



Power laws

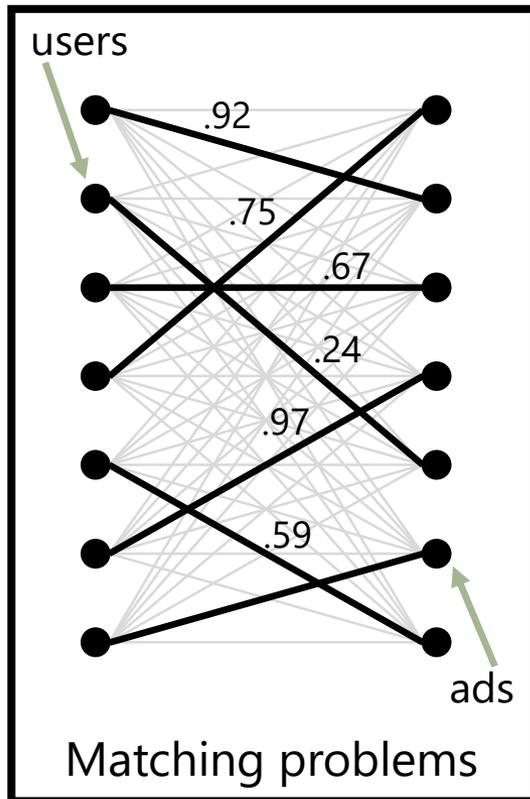


Small-world phenomena

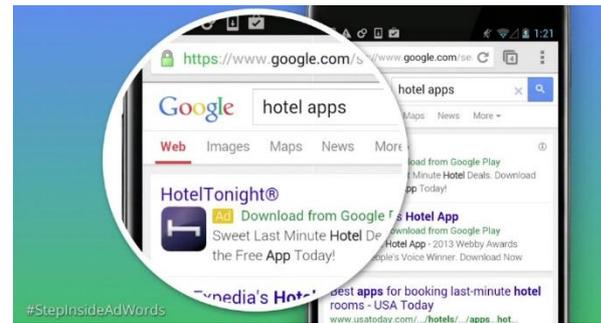


Strong & weak ties

# 9. Advertising



## AdWords



## Bandit algorithms

# CSE 258 – Lecture 17

Web Mining and Recommender Systems

Temporal dynamics of text

# Bag-of-Words representations of text:

## The Peculiar Genius of Bjork

CULTURE | BY EMILY WITT | JANUARY 23, 2015 11:30 AM

*Solo musician or master collaborator? For her new album, Bjork has merged the two sides of her artistry to create a new experience of music – again.*



$F_{\text{text}} = [150, 0, 0, 0, 0, 0, \dots, 0]$

a

aardvark

zoetrope

musician, who creates her music in an emotional cocoon, tinkering with technologies, concepts and feelings; and Bjork the producer and curator, who seeks out



# Latent Dirichlet Allocation

In week 5/7, we tried to develop low-dimensional representations of documents:

## What we would like:

87 of 102 people found the following review helpful

★★★★★ **You keep what you kill**, December 27, 2004

By [Schtinky "Schtinky"](#) (Washington State) - [See all my reviews](#)

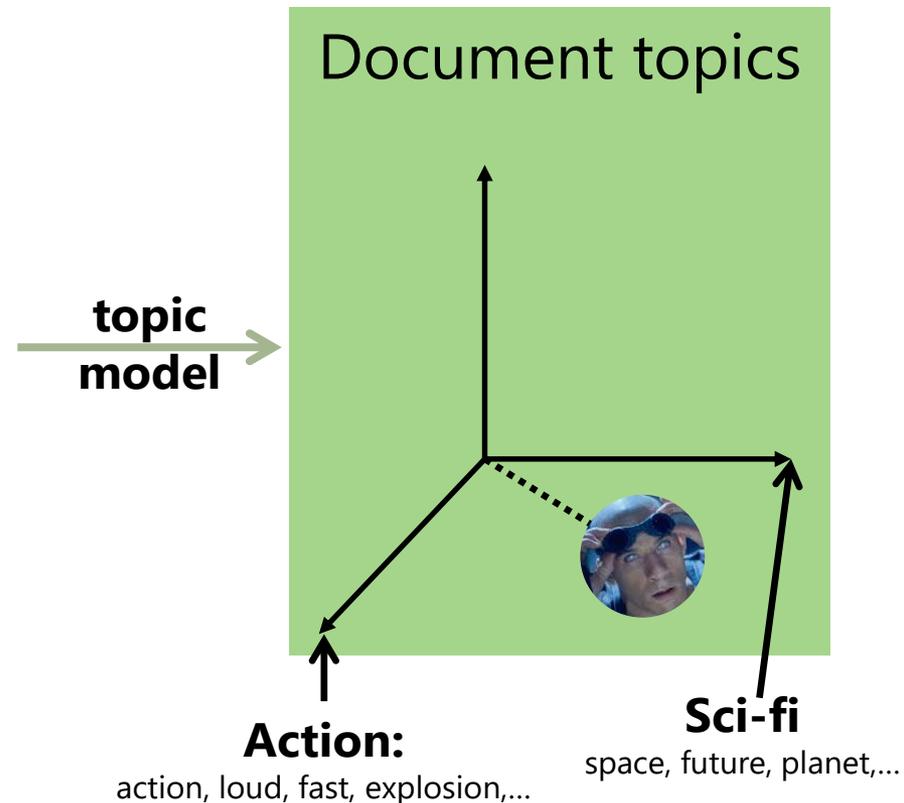
VINE™ VOICE

This review is from: [The Chronicles of Riddick \(Widescreen Unrated Director's Cut\) \(DVD\)](#)

Even if I have to apologize to my Friends and Favorites, and my family, I have to admit that I really liked this movie. It's a Sci-Fi movie with a "Mad Maxx" appeal that, while changing many things, left Riddick from 'Pitch Black' to be just Riddick. They did not change his attitude or soften him up or bring him out of his original character, which was very pleasing to 'Pitch Black' fans like myself.

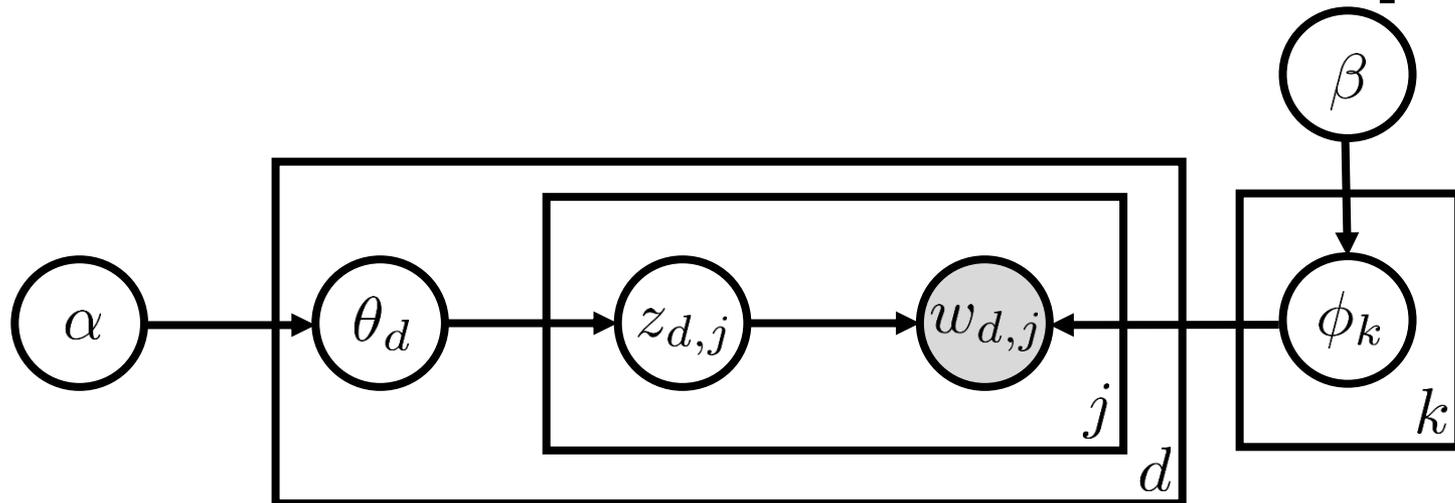
First off, let me say that when playing the DVD, the first selection to come up is Convert or Fight, and no explanation of the choices. This confused me at first, so I will mention off the bat that they are simply different menu formats, that each menu has the very same options, simply different background visuals. Select either one and continue with the movie.

(review of "The Chronicles of Riddick")



# Latent Dirichlet Allocation

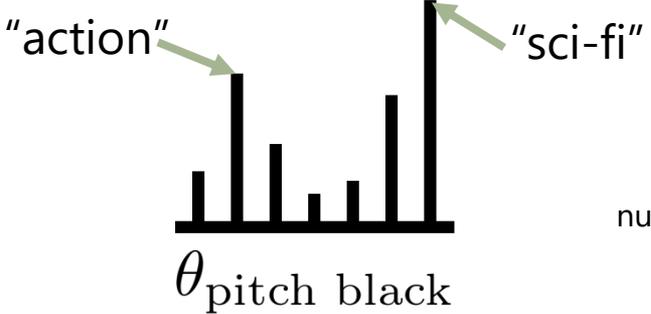
We saw how **LDA** can be used to describe documents in terms of **topics**



- Each document has a **topic vector** (a stochastic vector describing the fraction of words that discuss each topic)
  - Each topic has a **word vector** (a stochastic vector describing how often a particular word is used in that topic)

# Latent Dirichlet Allocation

Topics and documents are **both** described using stochastic vectors:



Each document has a **topic distribution** which is a mixture over the topics it discusses

number of topics  $\rightarrow$

$$\theta_d \in \Delta^K \text{ i.e., } \forall_d \sum_k \theta_{d,k} = 1$$


Each topic has a **word distribution** which is a mixture over the words it discusses

number of words  $\rightarrow$

$$\phi_k \in \Delta^D \text{ i.e., } \forall_k \sum_w \phi_{k,w} = 1$$

# Latent Dirichlet Allocation

**Topics over Time** (Wang & McCallum, 2006) is an approach to incorporate temporal information into topic models

e.g.

- The topics discussed in conference proceedings progressed from neural networks, towards SVMs and structured prediction (and back to neural networks)
- The topics used in political discourse now cover science and technology more than they did in the 1700s
- Within an institution, e-mails will discuss different topics (e.g. recruiting, conference deadlines) at different times of the year

# Latent Dirichlet Allocation

**Topics over Time** (Wang & McCallum, 2006) is an approach to incorporate temporal information into topic models

The ToT model is similar to LDA with one addition:

1. For each topic  $K$ , draw a word vector  $\phi_k$  from  $\text{Dir}(\beta)$
2. For each document  $d$ , draw a topic vector  $\theta_d$  from  $\text{Dir}(\alpha)$
3. For each word position  $i$ :
  1. draw a topic  $z_{di}$  from multinomial  $\theta_d$
  2. draw a word  $w_{di}$  from multinomial  $\phi_{z_{di}}$
  3. **draw a timestamp  $t_{di}$  from  $\text{Beta}(\psi_{z_{di}})$**

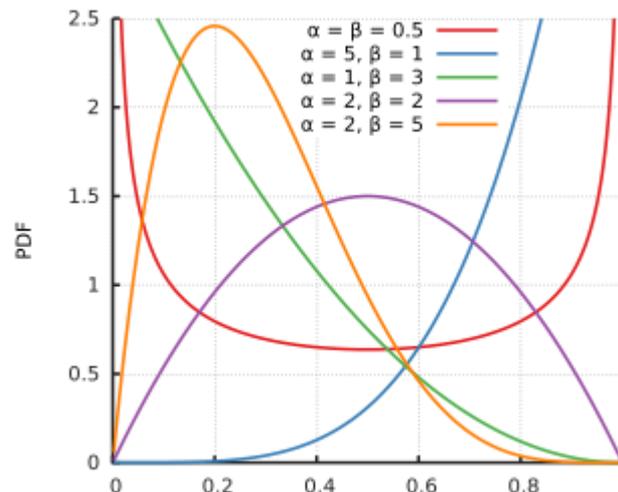
# Latent Dirichlet Allocation

**Topics over Time** (Wang & McCallum, 2006) is an approach to incorporate temporal information into topic models

**3.3. draw a timestamp  $t_{\{di\}}$  from  $\text{Beta}(\psi_{\{z_{\{di\}}\}})$**

- There is now one Beta distribution **per topic**
- Inference is still done by Gibbs sampling, with an outer loop to update the Beta distribution parameters

Beta distributions are a flexible family of distributions that can capture several types of behavior – e.g. gradual increase, gradual decline, or temporary “bursts”

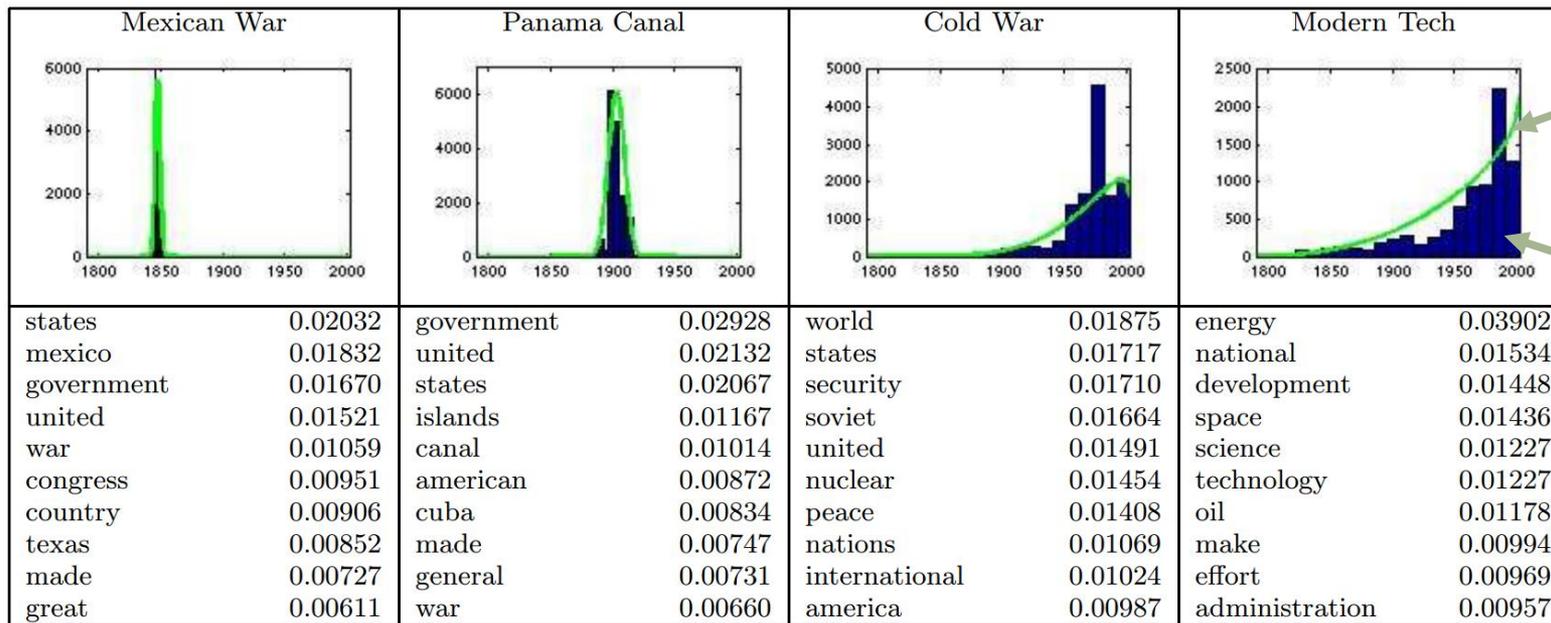


$$\text{p.d.f.:} \\ \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)}$$

# Latent Dirichlet Allocation

## Results:

Political addresses – the model seems to capture realistic “bursty” and gradually emerging topics

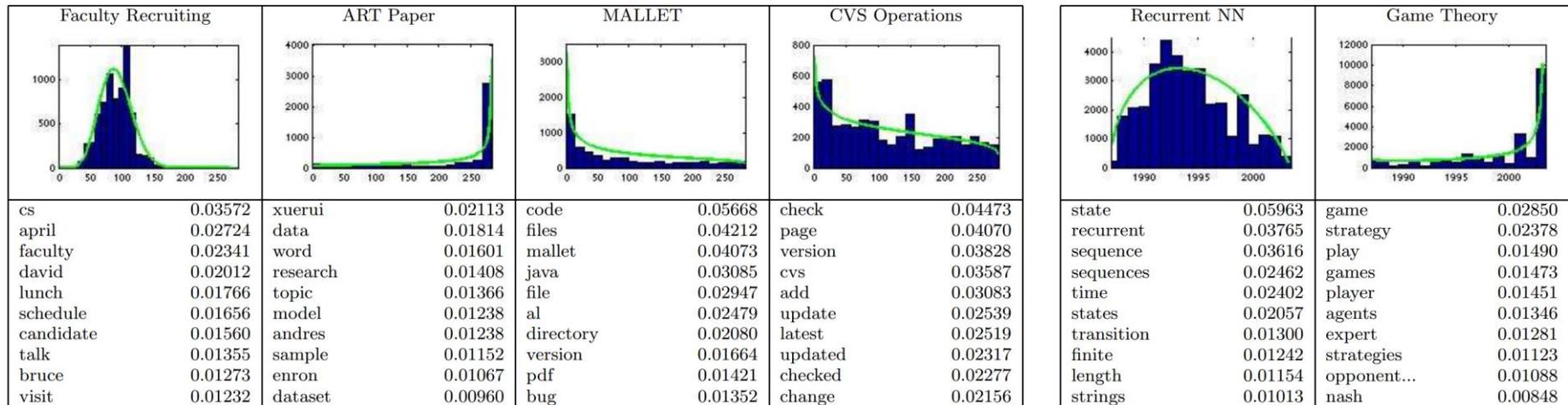


fitted Beta distribution

assignments to this topic

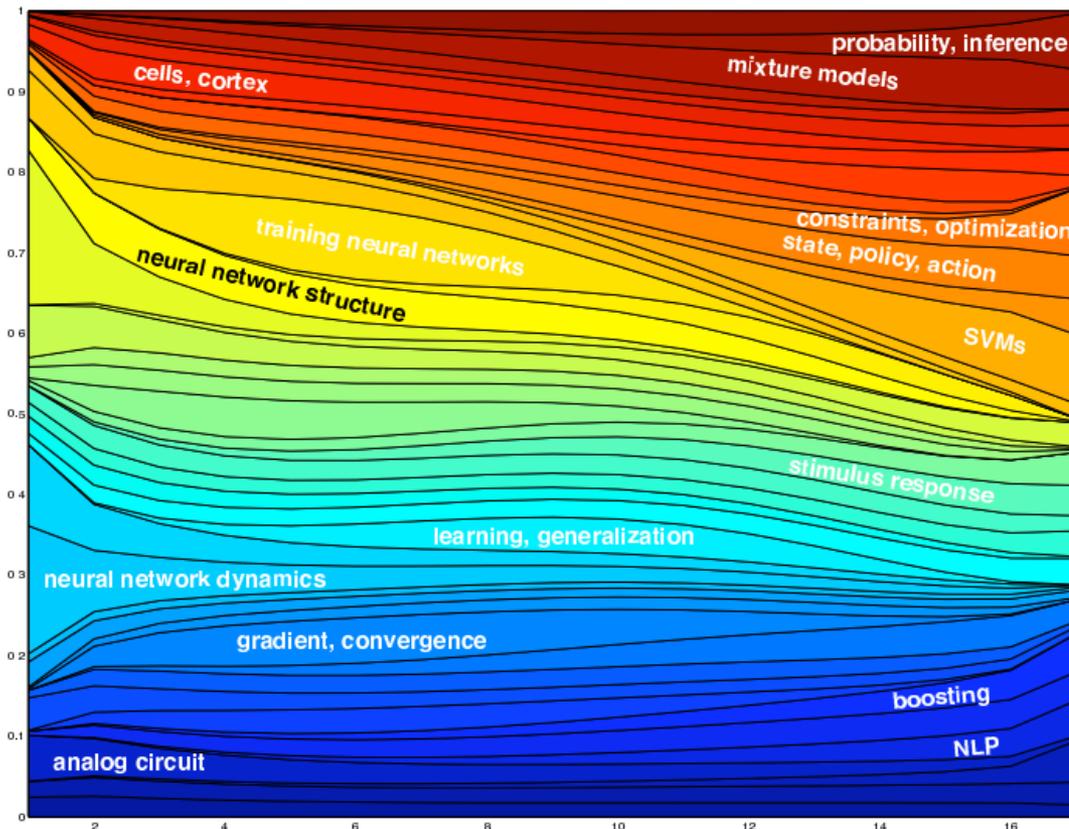
# Latent Dirichlet Allocation

## Results: e-mails & conference proceedings



# Latent Dirichlet Allocation

**Results:**  
conference proceedings (NIPS)



Relative weights  
of various topics  
in 17 years of  
NIPS proceedings

# Questions?

Further reading:

“Topics over Time: A Non-Markov  
Continuous-Time Model of Topical  
Trends”

(Wang & McCallum, 2006)

<http://people.cs.umass.edu/~mccallum/papers/tot-kdd06.pdf>

# CSE 258 – Lecture 17

Web Mining and Recommender Systems

Temporal dynamics of social networks

## How can we **characterize, model, and reason about** the structure of social networks?

1. Models of network structure
2. Power-laws and scale-free networks, “rich-get-richer” phenomena
3. Triadic closure and “the strength of weak ties”
4. Small-world phenomena
5. Hubs & Authorities; PageRank

# Temporal dynamics of social networks

Two weeks ago we saw some processes that model the generation of social and information networks

- Power-laws & small worlds
- Random graph models

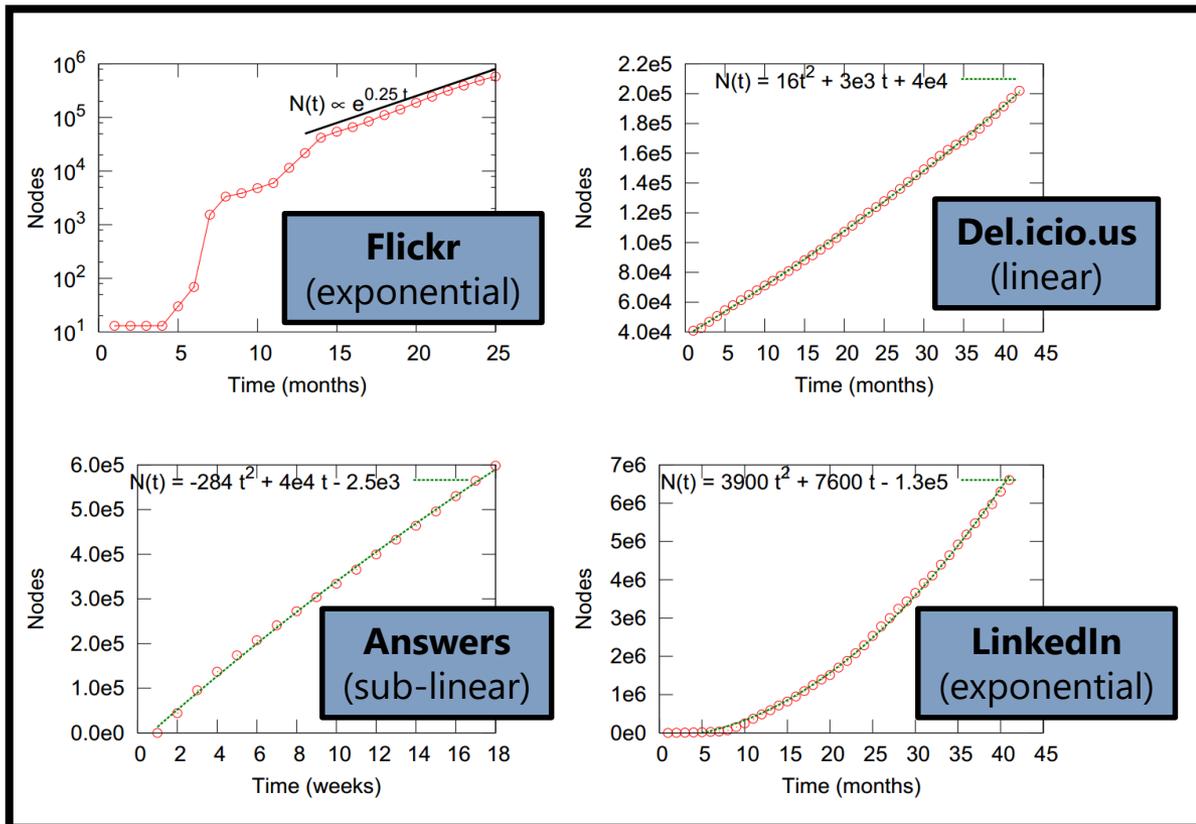
These were all defined with a “static” network in mind.

But if we observe the **order** in which edges were created, we can study how these phenomena change as a function of time

First, let’s look at “microscopic” evolution, i.e., evolution in terms of individual nodes in the network

# Temporal dynamics of social networks

**Q1:** How do networks grow in terms of the number of nodes over time?



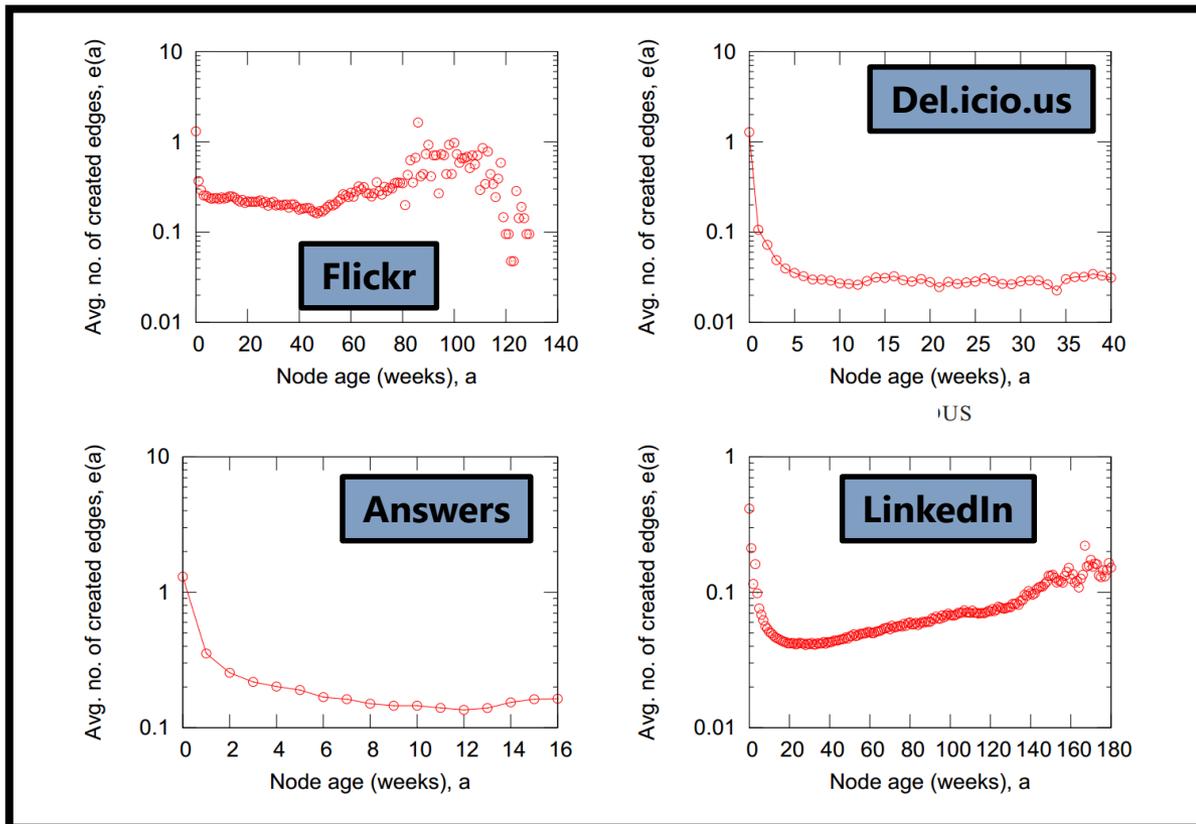
(from Leskovec, 2008 (CMU Thesis))

**A:** Doesn't seem to be an obvious trend, so what **do** networks have in common as they evolve?

# Temporal dynamics of social networks

## Q2: When do nodes create links?

- x-axis is the age of the nodes
- y-axis is the number of edges created at that age

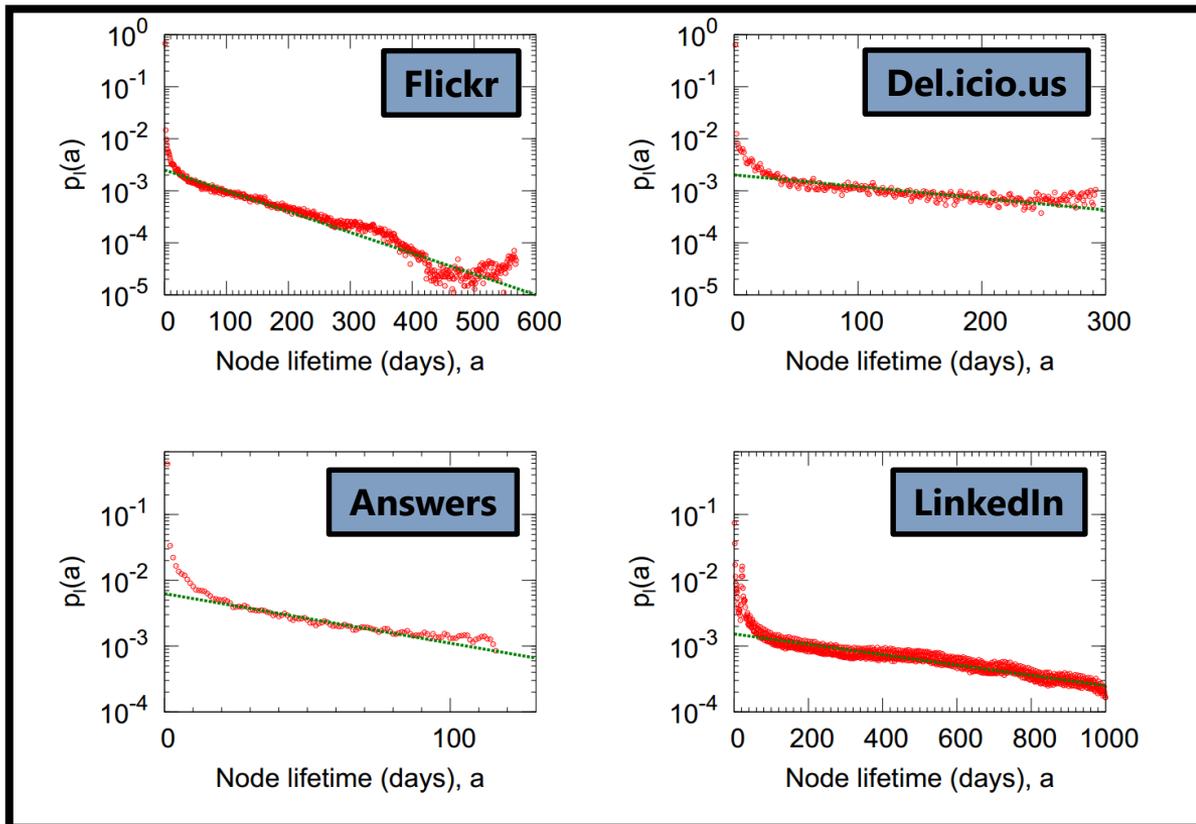


**A:** In most networks there's a "burst" of initial edge creation which gradually flattens out. Very different behavior on LinkedIn (guesses as to why?)

# Temporal dynamics of social networks

## Q3: How long do nodes "live"?

- x-axis is the diff. between date of last and first edge creation
  - y-axis is the frequency

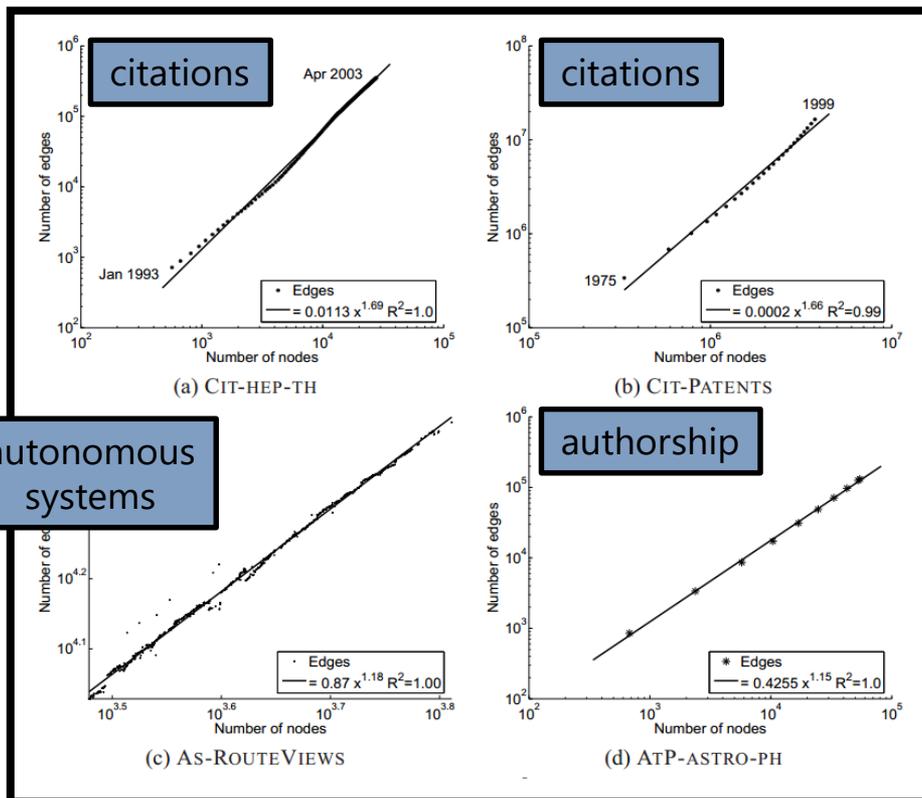


**A:** Node lifetimes follow a power-law: many many nodes are shortlived, with a long-tail of older nodes

# Temporal dynamics of social networks

What about “macroscopic” evolution, i.e., how do global properties of networks change over time?

**Q1:** How does the # of nodes relate to the # of edges?



- A few more networks: citations, authorship, and autonomous systems (and some others, not shown)
- **A:** Seems to be linear (on a log-log plot) **but** the number of edges grows **faster** than the number of nodes as a function of time

# Temporal dynamics of social networks

**Q1:** How does the # of nodes relate to the # of edges?

**A:** seems to behave like

$$E(t) \propto N(t)^a$$

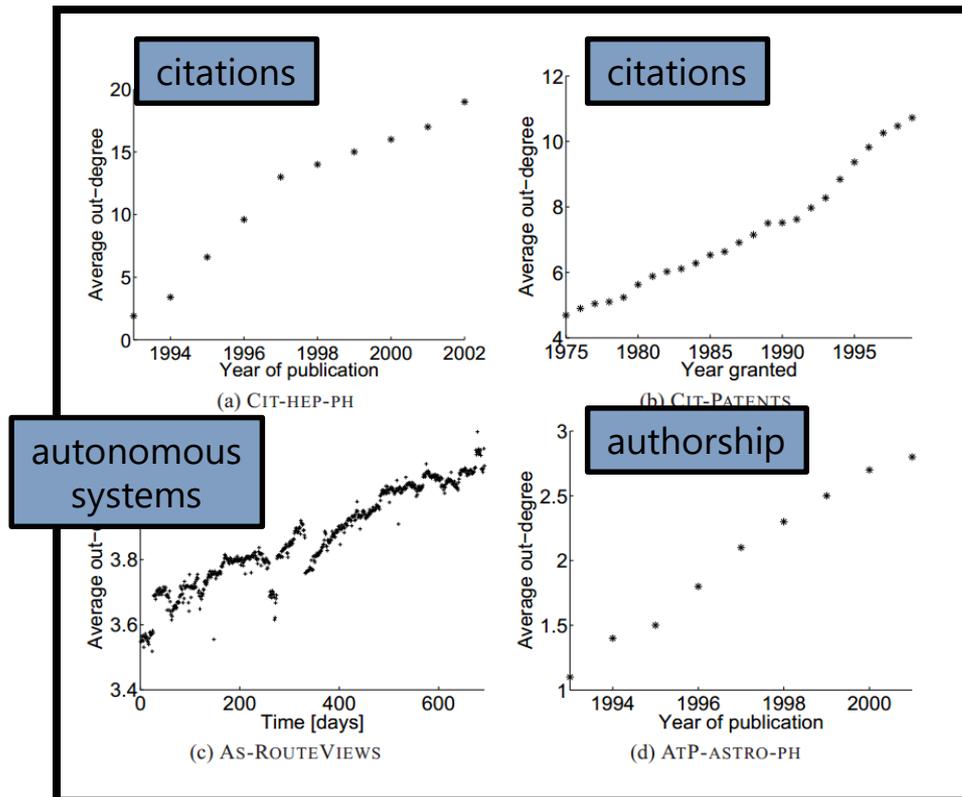
where

$$1 \leq a \leq 2$$

- $a = 1$  would correspond to **constant** out-degree – which is what we might traditionally assume
- $a = 2$  would correspond to the graph being fully connected
  - What seems to be the case from the previous examples is that  $a > 1$  – the number of edges grows faster than the number of nodes

# Temporal dynamics of social networks

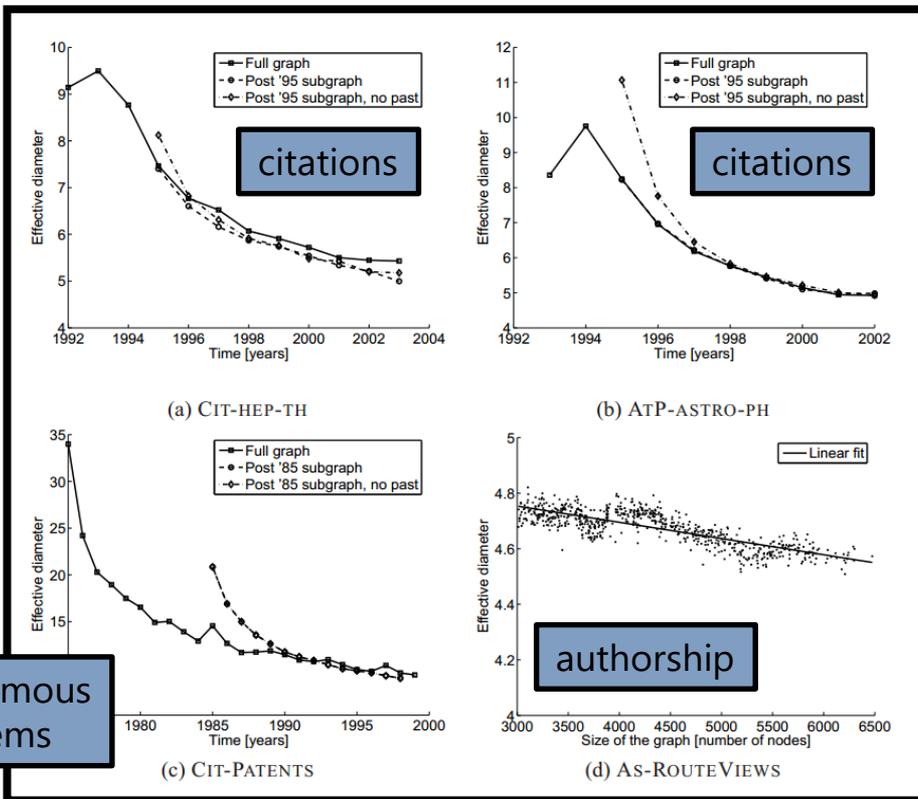
**Q2:** How does the degree change over time?



- **A:** The average out-degree **increases** over time

# Temporal dynamics of social networks

**Q3:** If the network becomes **denser**, what happens to the (effective) diameter?

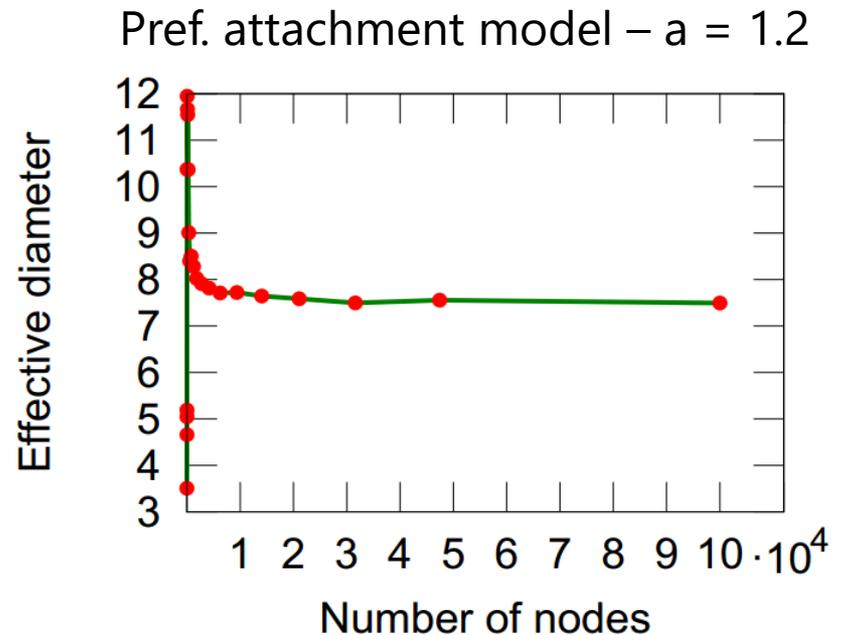
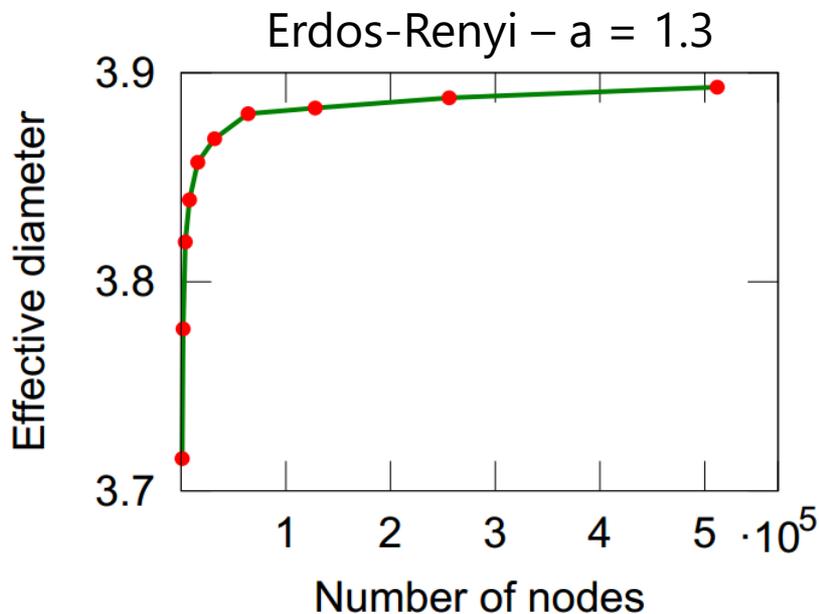


- **A:** The diameter seems to decrease
- In other words, the network becomes **more** of a small world as the number of nodes increases

autonomous systems

# Temporal dynamics of social networks

**Q4:** Is this something that **must** happen – i.e., if the number of edges increases faster than the number of nodes, does that mean that the diameter must decrease?  
**A:** Let's construct random graphs (with  $a > 1$ ) to test this:

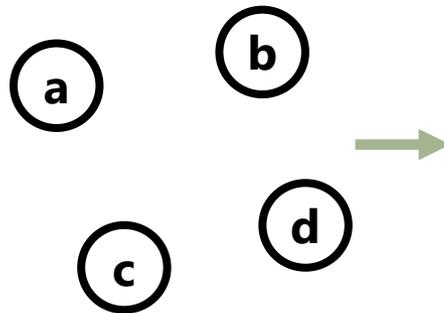


# Temporal dynamics of social networks

So, a decreasing diameter is **not** a “rule” of a network whose number of edges grows faster than its number of nodes, though it is consistent with a preferential attachment model

**Q5:** is the degree distribution of the nodes sufficient to explain the observed phenomenon?

**A:** Let's perform **random rewiring** to test this

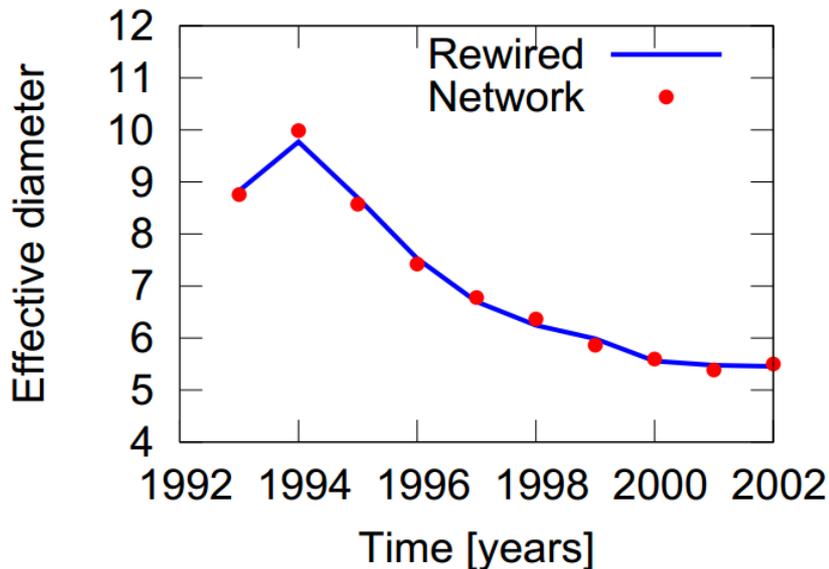


random rewiring preserves the degree distribution, and randomly samples amongst networks with observed degree distribution

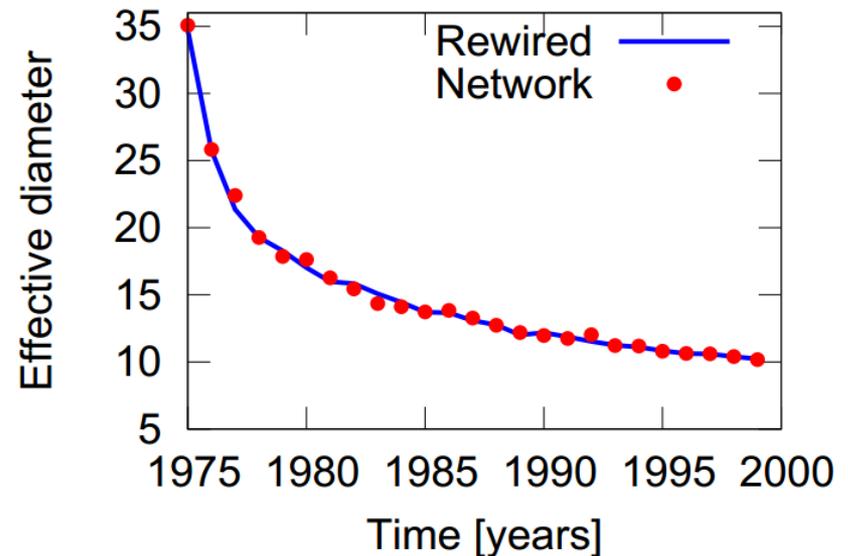
# Temporal dynamics of social networks

So, a decreasing diameter is **not** a “rule” of a network whose number of edges grows faster than its number of nodes, though it is consistent with a preferential attachment model

**Q5:** is the degree distribution of the nodes sufficient to explain the observed phenomenon?



(c) Affiliation network (ATP-ASTRO-PH)



(d) US patent citation network (CIT-PATENTS)

# Temporal dynamics of social networks

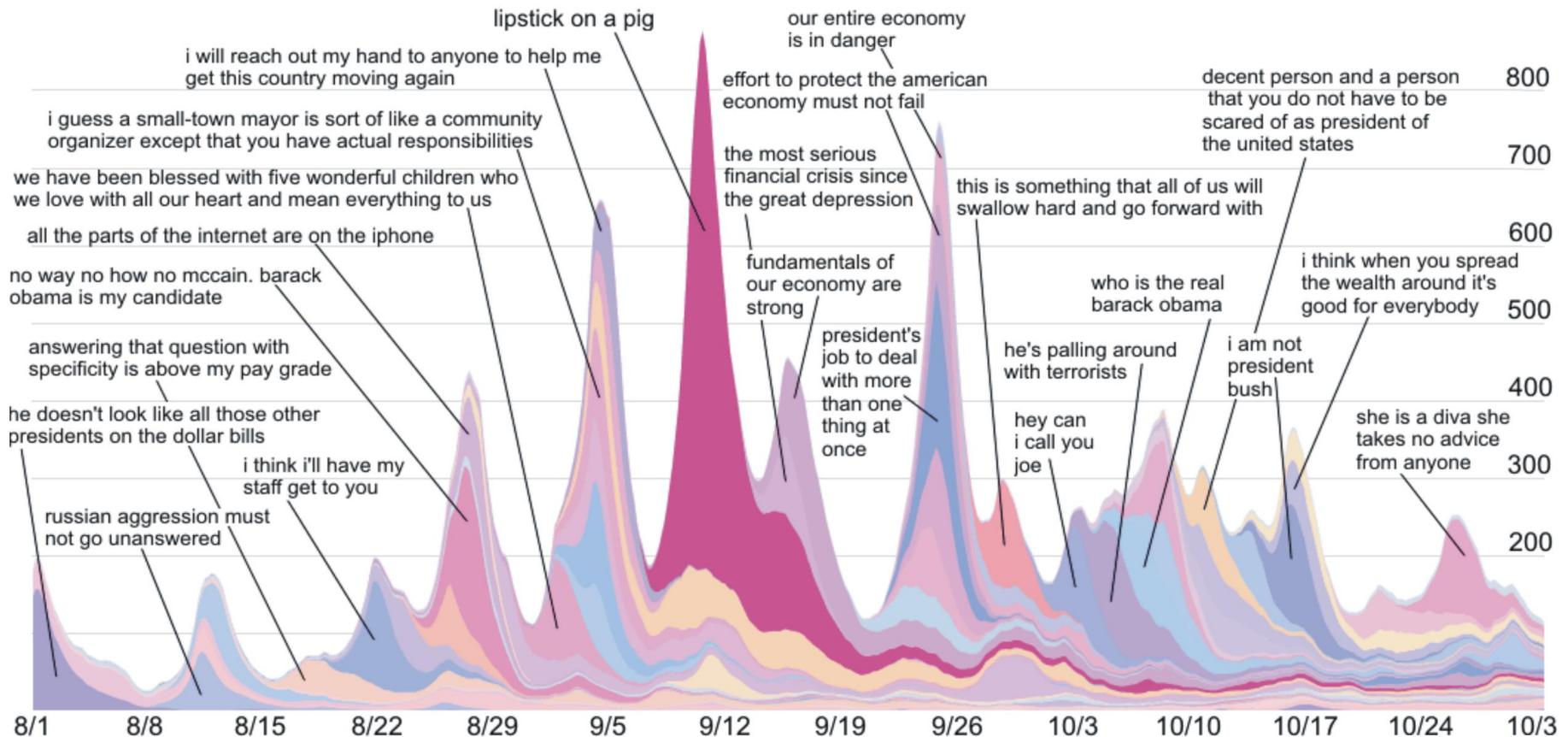
So, a decreasing diameter is **not** a “rule” of a network whose number of edges grows faster than its number of nodes, though it is consistent with a preferential attachment model

**Q5:** is the degree distribution of the nodes sufficient to explain the observed phenomenon?

**A:** Yes! The fact that real-world networks seem to have decreasing diameter over time can be explained as a result of their degree distribution **and** the fact that the number of edges grows faster than the number of nodes

# Temporal dynamics of social networks

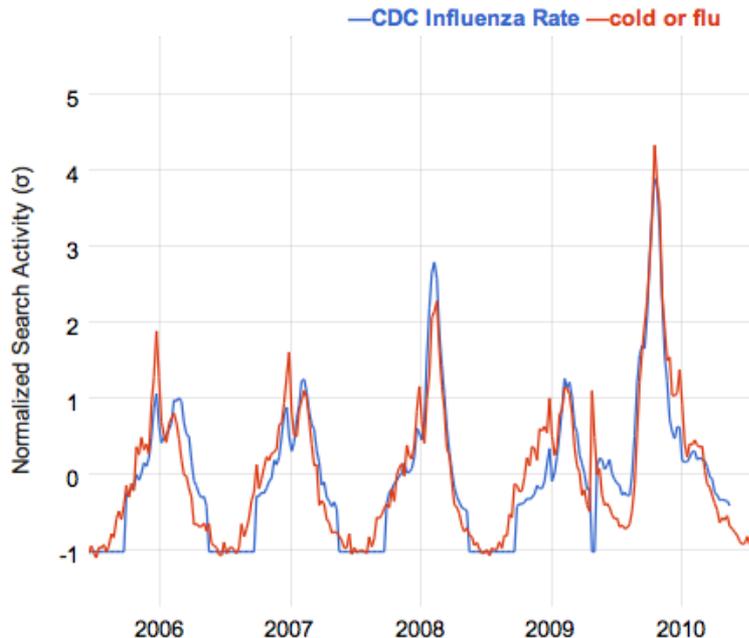
## Other interesting topics...



"memetracker"

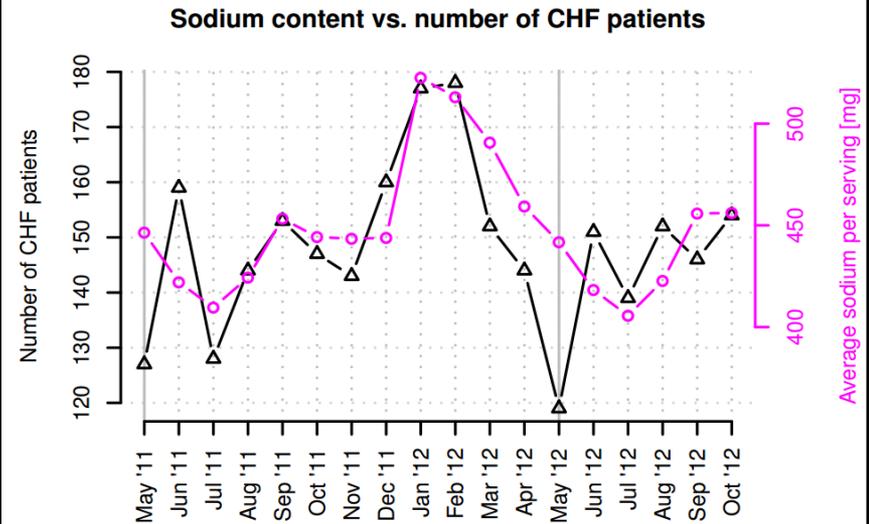
# Temporal dynamics of social networks

## Other interesting topics...



Aligning query data with disease data –  
Google flu trends:

<https://www.google.org/flutrends/us/#US>



Sodium content in recipe searches vs.  
# of heart failure patients – “From  
Cookies to Cooks” (West et al. 2013):

[http://infolab.stanford.edu/~west1/pubs/West-White-Horvitz\\_WWW-13.pdf](http://infolab.stanford.edu/~west1/pubs/West-White-Horvitz_WWW-13.pdf)

# Questions?

## Further reading:

“Dynamics of Large Networks” (most plots from here)

Jure Leskovec, 2008

<http://cs.stanford.edu/people/jure/pubs/thesis/jure-thesis.pdf>

“Microscopic Evolution of Social Networks”

Leskovec et al. 2008

<http://cs.stanford.edu/people/jure/pubs/microEvol-kdd08.pdf>

“Graph Evolution: Densification and Shrinking  
Diameters”

Leskovec et al. 2007

<http://cs.stanford.edu/people/jure/pubs/powergrowth-tkdd.pdf>

# CSE 258 – Lecture 17

## Web Mining and Recommender Systems

Some incredible assignments



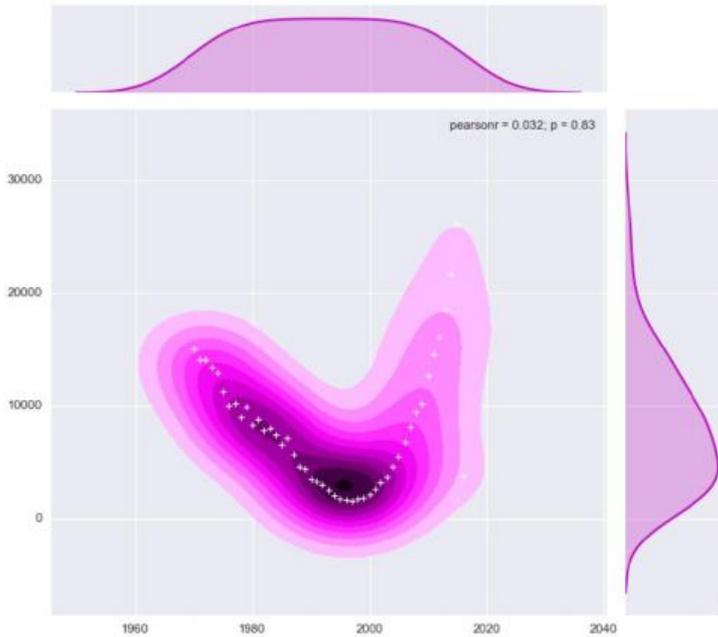




# Beer reviews



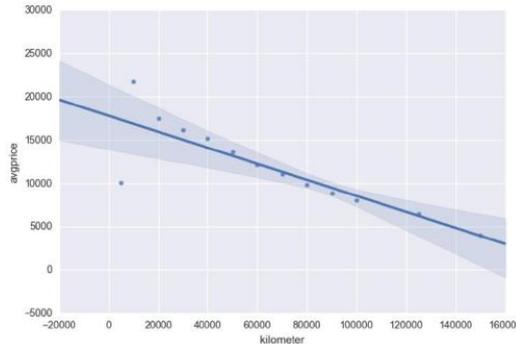
# Used car price prediction



Price vs. registration year

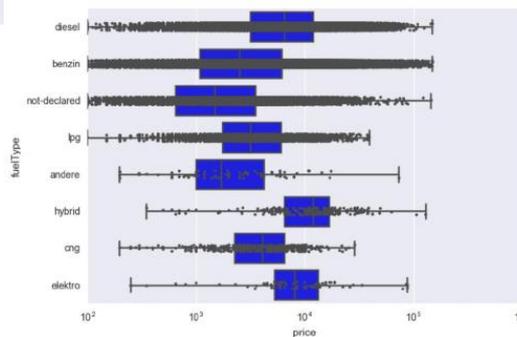
Kaggle used cars dataset (370,000 instances)

Price vs. mileage



- Type (sedan, van, etc.)
- Mileage
- Age
- PowerPS
- Damage
- Gearbox
- Fuel type

Price vs. fuel type



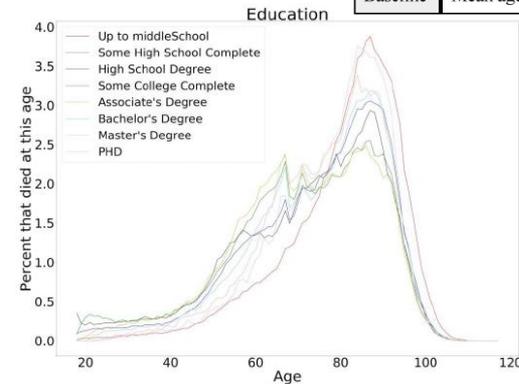
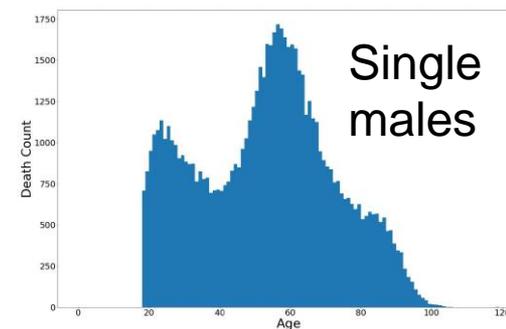
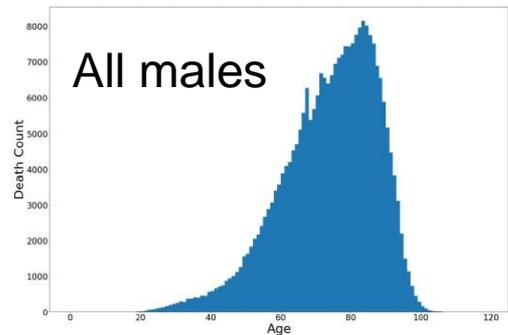
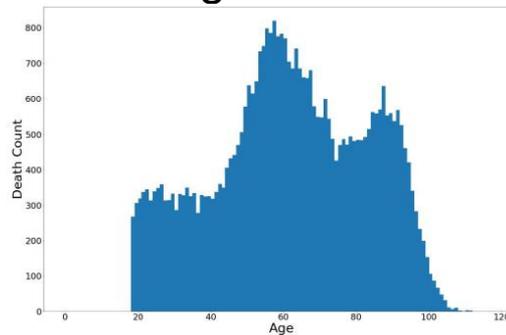
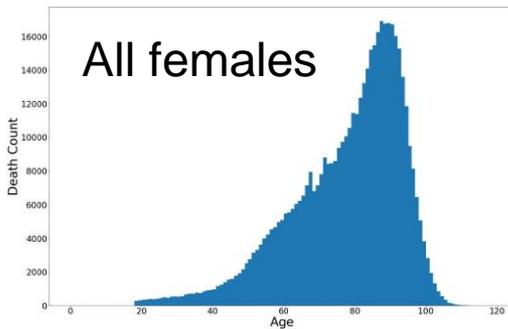
Features	Train Set Accuracy	Test Set Accuracy
E	0.62770924	0.628140622
D	0.660893404	0.661312692
B	0.685244315	0.686518303
B+E	0.689159007	0.690291888
B+E+D	0.836074827	0.802370585
B+E+D+A	0.88159571	0.830882116
B+E+D+F	0.978870343	0.775907112
B+E+D+C	0.846331237	0.803096275

# Death clock

## CDC Mortality Dataset (2.1 million instances)

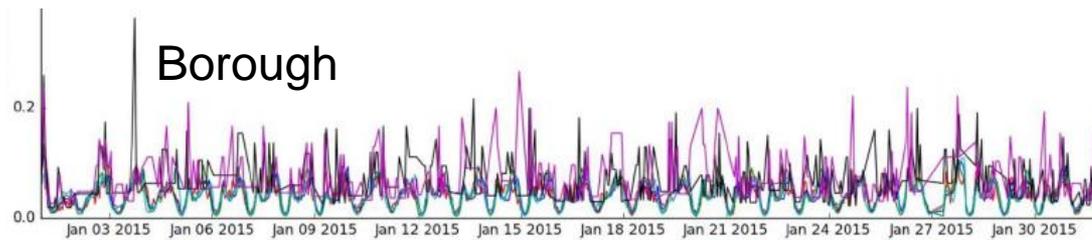
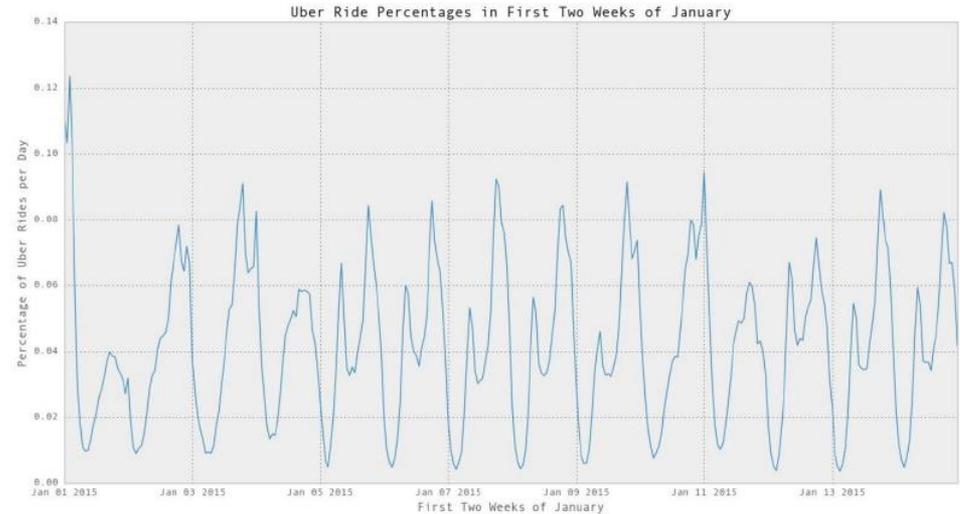
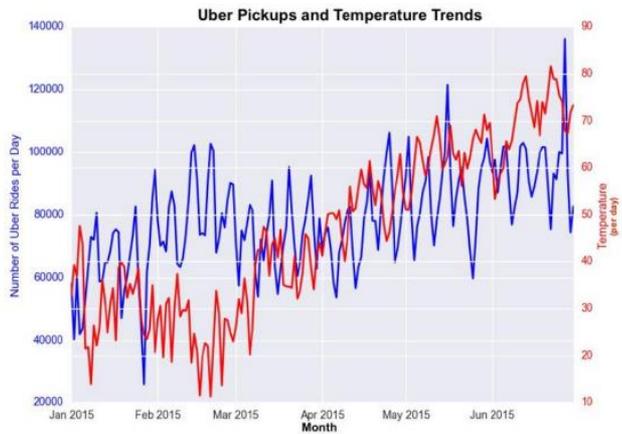
Models	Feature List	MSE
Best	Linear Regression using: {All above features} and 2,100,000 deaths	142.43
2 <sup>nd</sup>	Linear Regression using: {All above features except <i>Activity Code</i> features} and 2,100,000 deaths	142.89
3 <sup>rd</sup>	Linear Regression using: {All above features except <i>Resident Code</i> features} and 2,100,000 deaths	143.52
4 <sup>th</sup>	Linear Regression using: {All above features except <i>Education Level</i> features} and 2,100,000 deaths	144.24
5 <sup>th</sup>	Linear Regression using: {All above features except <i>Marriage</i> features} and 2,100,000 deaths	186.49
6 <sup>th</sup>	SVM using {All above features} and 50,000 deaths	178.20
Baseline	Mean age at death	270.25

### Single females



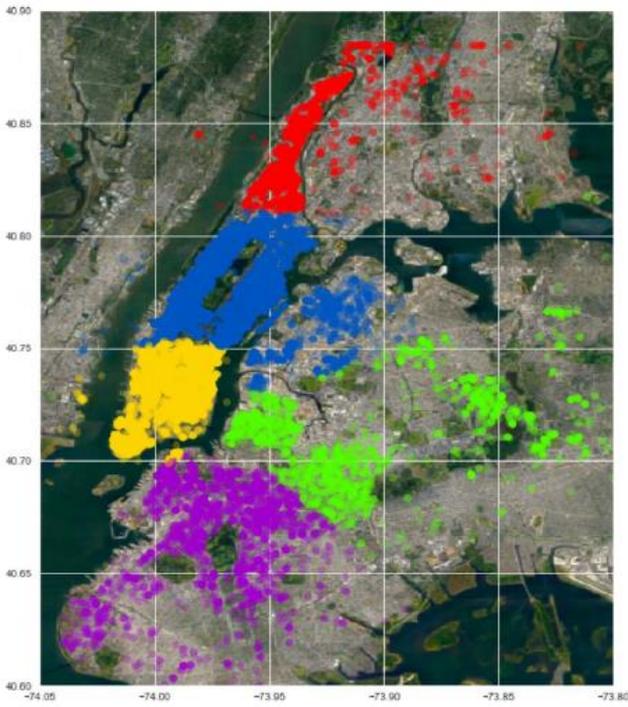
# Uber pickups

## NYC Uber Dataset (14.2 million samples)

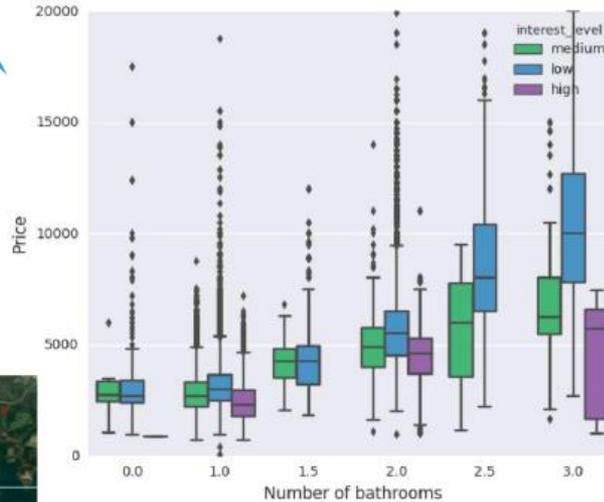


Sample Size	Baseline test MSE	Ltsq train MSE (w/ weather)	Ltsq Test MSE (w/ weather)	Ltsq train MSE (w/o weather)	Ltsq Test MSE (w/o weather)
500,000	1.90e-4	1.823e-4	1.189e-3	1.846e-4	1.005e-3
750,000	2.10 e-4	1.874e-4	1.189e-3	1.883e-4	2.070e-4
1000000	4.33 e-4	1.939e-4	1.302e-3	1.942e-4	2.204e-4

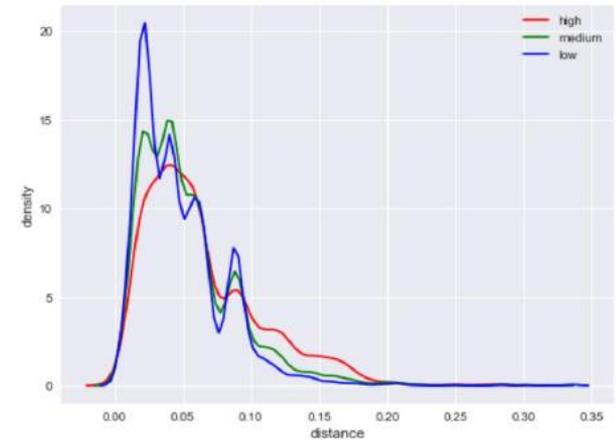
# Rental recommendations



#bathrooms



distance to city center



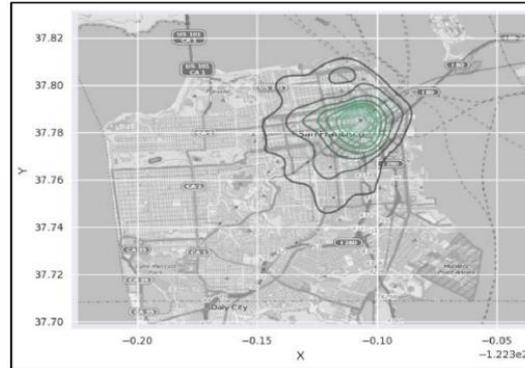
Interest level:



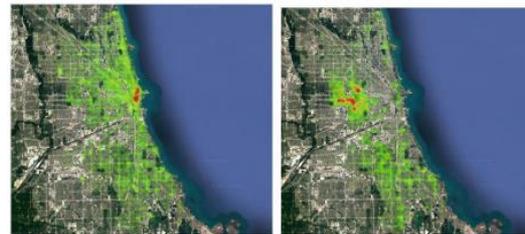
Wen Zhang, Xingbo Wang, Kaixiang Zhao, Lifan Chen  
 Shiunn An Lu, Shanyu Chuang, Hao-En Sung  
 Side Li, Yifan Xu  
 Dhruv Sharma, Keshav Sharma, Saransh Jain

# Crime prediction

Field	Description
ID	Unique identifier for the record
Case Number	The Chicago Police Department RD Number
Date	Date when the incident occurred in mm/dd/yyyy.
Block	The partially redacted address where the incident occurred.
IUCR	The Illinois Uniform Crime Reporting code
Primary Type	The primary description of the IUCR code.
Description	The secondary description of the IUCR code
Location Description	Description of the location where the incident occurred.
Arrest	Indicates whether an arrest was made.
Domestic	Indicates whether the incident was domestic-related
Beat	Indicates the beat (the smallest police geographic area) where the incident occurred.
District	Indicates the police district where the incident occurred.
Ward	The ward (City Council district) where the incident occurred.
Community Area	Indicates the community area where the incident occurred. Chicago has 77 community areas.
FBI Code	Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).
X Coordinate	The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection.
Y Coordinate	The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection.
Year	Year the incident occurred.
Updated On	Date and time the record was last updated.
Latitude	The latitude of the location where the incident occurred.
Longitude	The longitude of the location where the incident occurred.

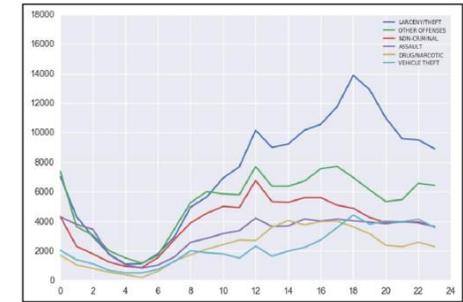


Theft by location

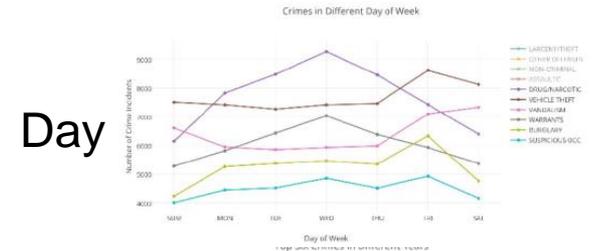


(a) Thefts

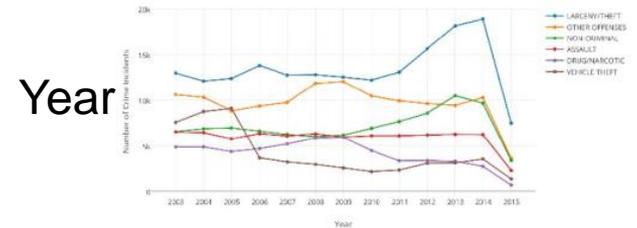
(b) Narcotics



Crime types by hour



Day



Year

Wenbin Zhu, Yuchen Wang, Wenjie Tao  
 Sahil Agarwal, Ujjwal Gulecha, Shalini Kedlaya  
 Junyang Li, Shenghong Wang

# H1B petitions

Kaggle dataset (~1 million samples)

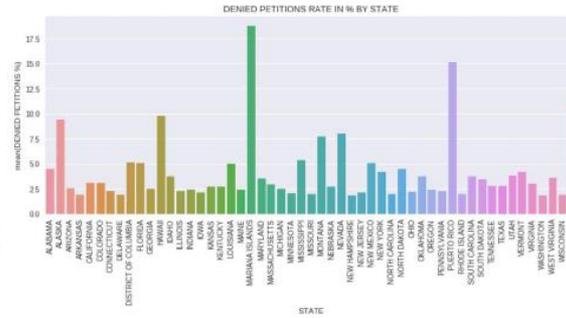
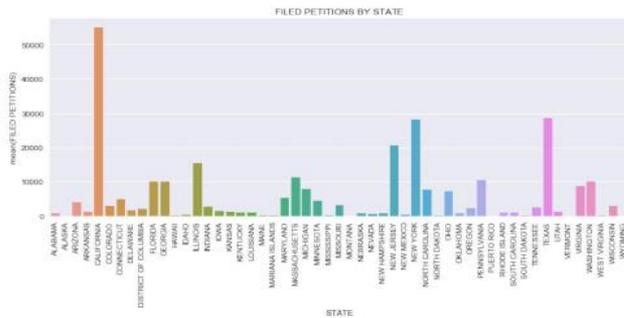


Fig. 8. State-wise Median wage

## Company

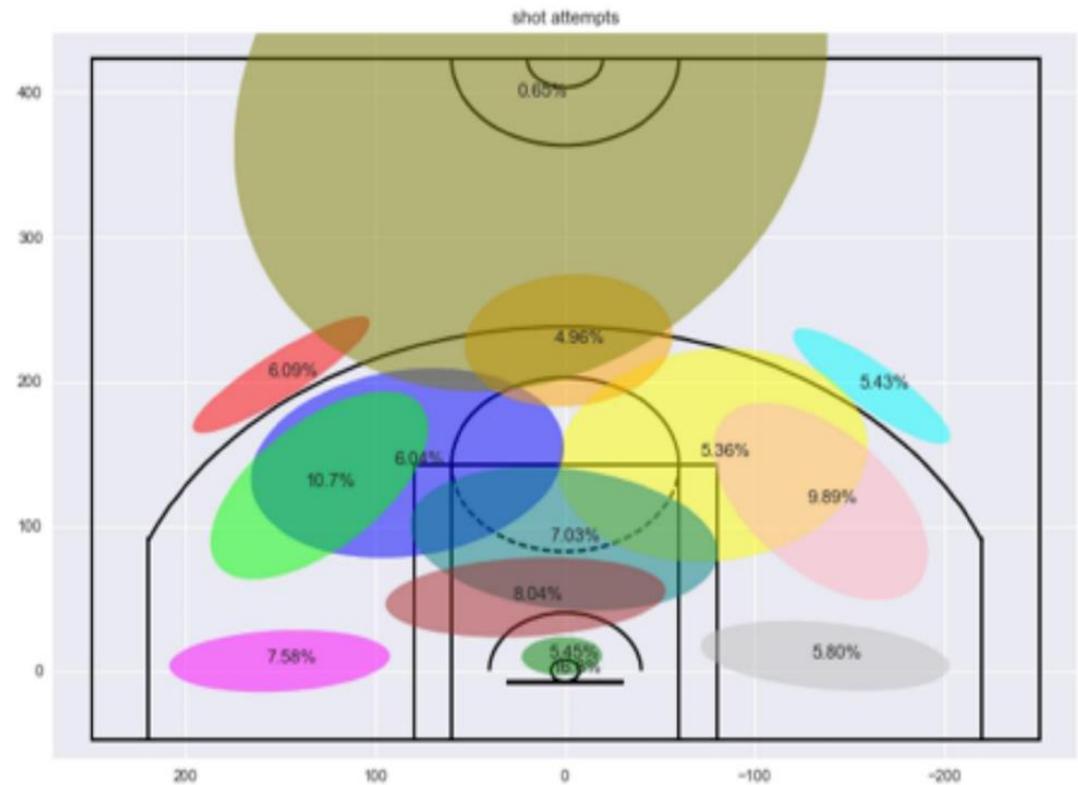
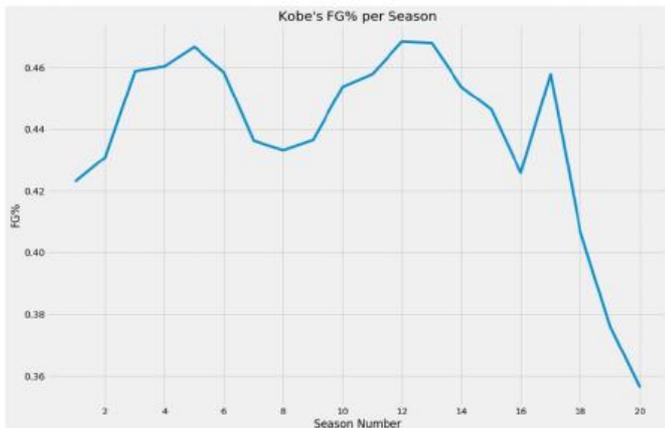
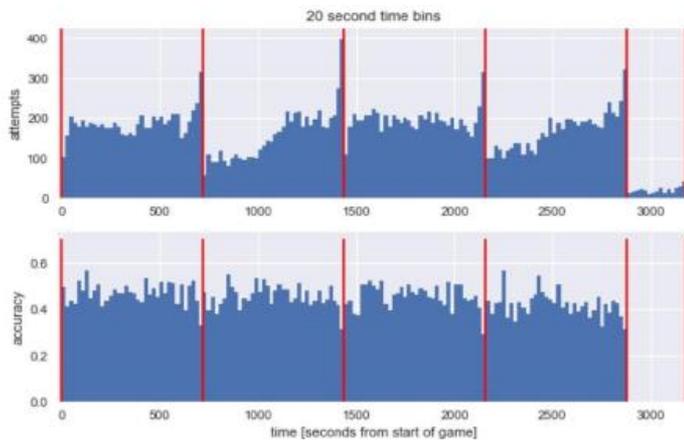


	SVM	Linear Regression	Logistic Regression
MSE(Training Set)	N\A	0.435	0.393
MSE(Validation set)	N\A	0.527	0.438
MSE(Testing set)	N\A	0.583	0.526
ER(Training set)	0.099	0.059	0.042
ER(Validation Set)	0.125	0.114	0.088
ER(Testing Set)	0.224	0.203	0.127

Yuchen Feng, Xuanzhen Xu, Jianxiong Lin  
 Prahal Arora, Rahul Vijay Dubey, Induja Sreekanthan, Jahnvi Singhal  
 Jialin Wang, Yishu Ma, Han Li

# Kobe field goals

Kaggle competition of 30,000 field-goal attempts



# Taxi tips

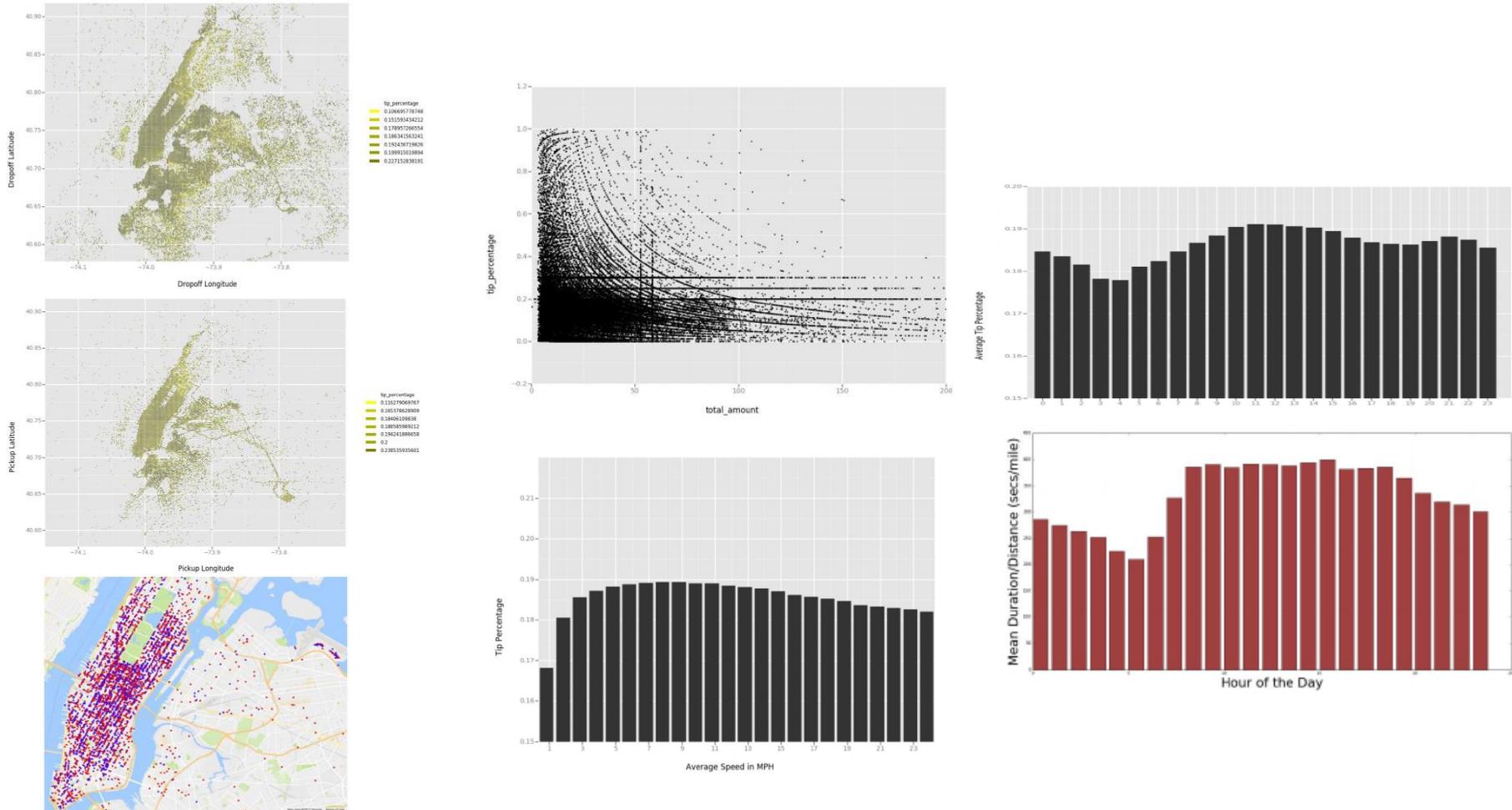


Fig. 1: Mapping of pick-up and drop-off locations

Rushil Nagda, Sudhanshu Bahety, Shubham Gupta  
 Tejas Saxena, Himanshu Jaiswal, Tushar Bansal, Prateek Ravindra Jakate

# Fill out those evaluations!

- Please evaluate the course on <https://academicaffairs.ucsd.edu/Modules/Evals?e2250306> !

Thanks!