CSE 190 – Lecture 9

Data Mining and Predictive Analytics

Temporal Modeling of Reviewer Expertise

Why do Americans like Pale Ale?

Hop cone

Lupulin

American Pale Ales:
- Hopsecutioner
- Hoptimus Prime
- Smooth Hoperator
- Red Hoptober
- Hoppy ending
- Hoptopus
- Hopsickle
- Tricerahops

Lupulin threshold shift:

People become accustomed to hops over time, and can recognize more subtle flavors

Users and products evolve over time

"Classics" are rated better (Koren, 2010); new products cause users to change focus (Koller & Malouf, 2007)

Age of the **product**

?

e (development) of the **user**

Users influence each other (Ma et al., 2011); communities shift over time (Xiong et al., 2010)

Age (zeitgeist) of the **community**

How can we effectively characterize **acquired tastes** or **expertise**?

Data

3M reviews, **100K** beers, **40K** users

1.5M reviews, **60K** beers, **30K** users

2M reviews, **500K** wines, **45K** users

Models of user and community evolution

Replace the 'standard' model

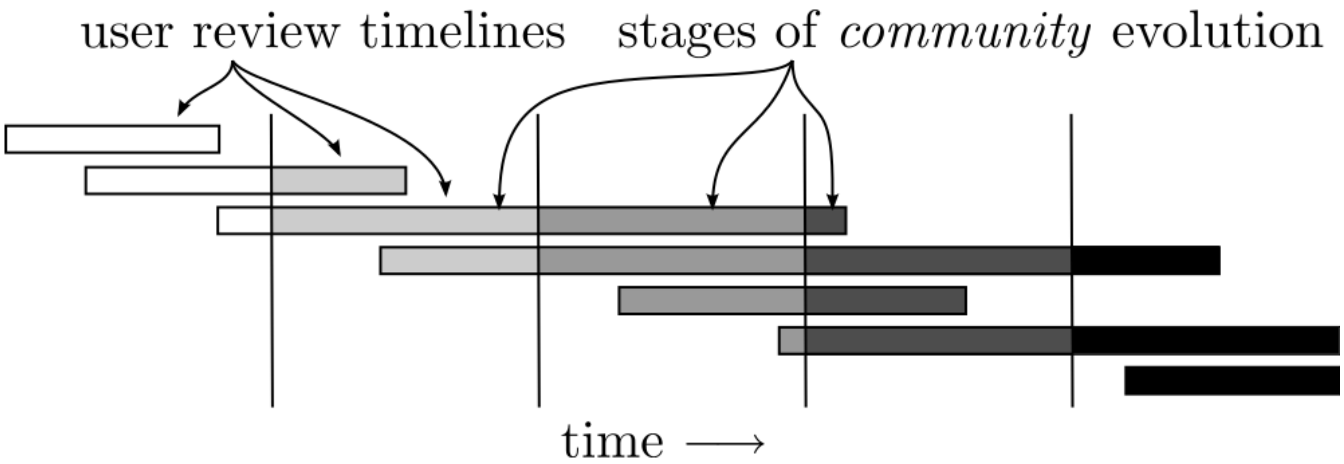
$$rec(u,i) = \alpha + \beta_u + \beta_i + \gamma_u \cdot \gamma_i$$

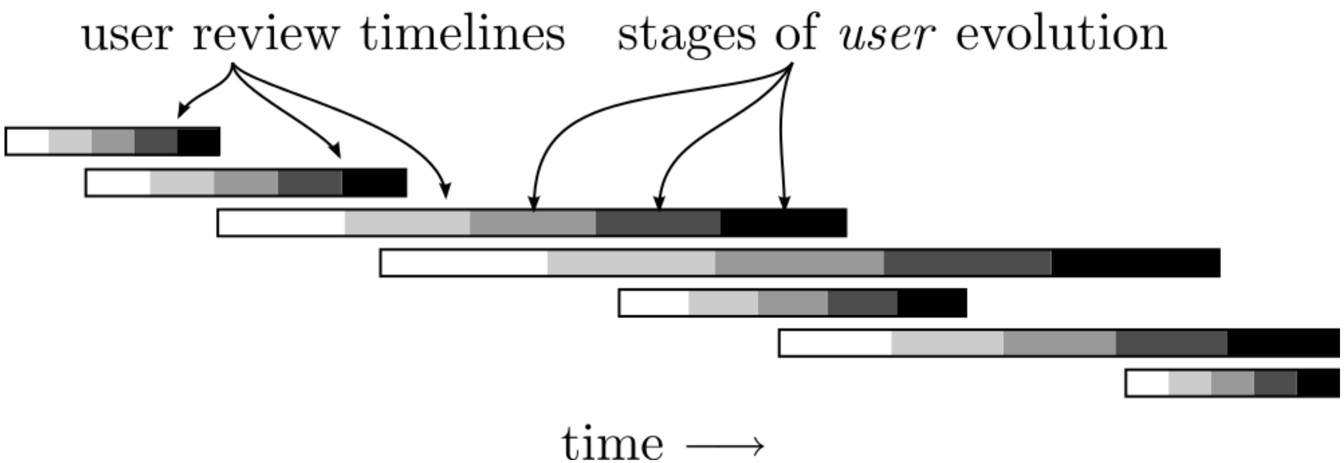
With one whose parameters change as a function of time (t)

$$rec_t(u,i) = \alpha(t) + \beta_u(t) + \beta_i(t) + \gamma_u(t) \cdot \gamma_i(t)$$

How we define *t* determines what type of evolution we model

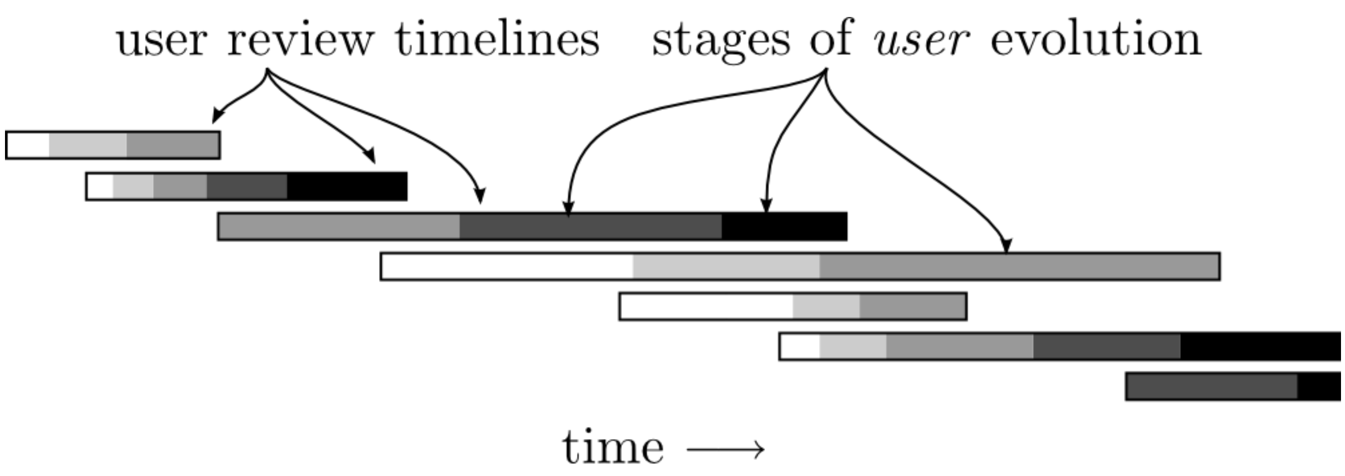
Models of user and community evolution





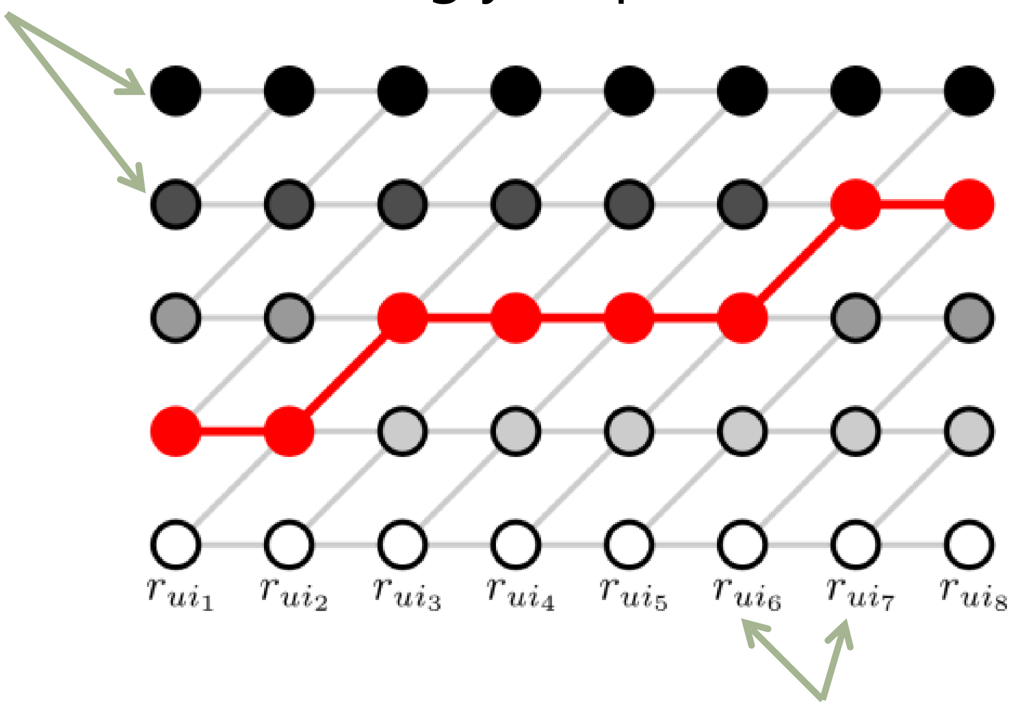
Models of user and community evolution

Observation:
People evolve and develop at different rates.
We must **learn** the rate of development for each user



Models of user and community evolution

rows: models of increasingly "experienced" users



columns: review timeline for one user

Each user's evolution can be thought of as a **monotonic** path through a graph

Optimization problem & fitting

Model:

$$rec_{e_{u,i}}(u,i) = \alpha(e) + \beta_u(e) + \beta_i(e) + \gamma_u(e) \cdot \gamma_i(e)$$

experience at time of review

offset, bias (user/item), and latent factors

Optimization problem:

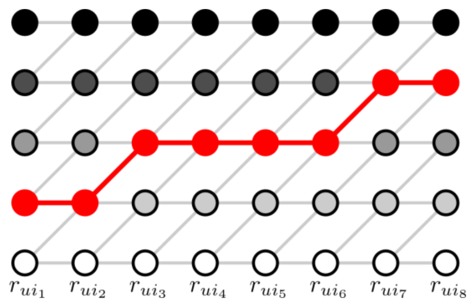
$$\arg\min_{\Theta,\mathcal{E}} \frac{1}{|\mathcal{T}|} \sum_{r_{u,i} \in \mathcal{T}} \underbrace{(rec_{e_{u,i}}(u,i) - r_{u,i})^2}_{\text{rating error}} + \underbrace{\Omega(\Theta)}_{\text{regularizer}}$$

model & experience parameters

smoothness & l2 regularity

Optimization problem & fitting

Repeat steps (1) and (2) until convergence:



(solved using dynamic programming)

Step 1:
fit expertise progression

$$\arg\min_{\Theta} \frac{1}{|\mathcal{T}|} \sum_{r_{u,i} \in \mathcal{T}} (rec_e(u,i) - r_{u,i})^2 + \Omega(\Theta)$$

solved via gradient ascent using L-BFGS
(see e.g. Koren & Bell, 2011)

Step 2:
fit rating models for each expertise level

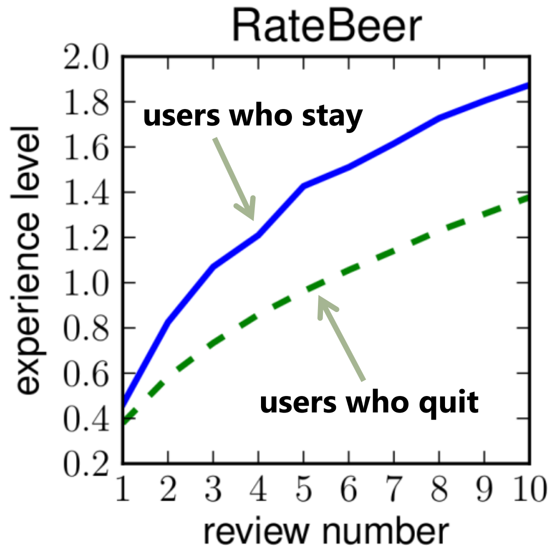
Outcomes – applications

Rating prediction:

- Beer: 6% improvement over state-of-the-art
- Wine: 13% improvement
- Movies (Amazon): 23% improvement

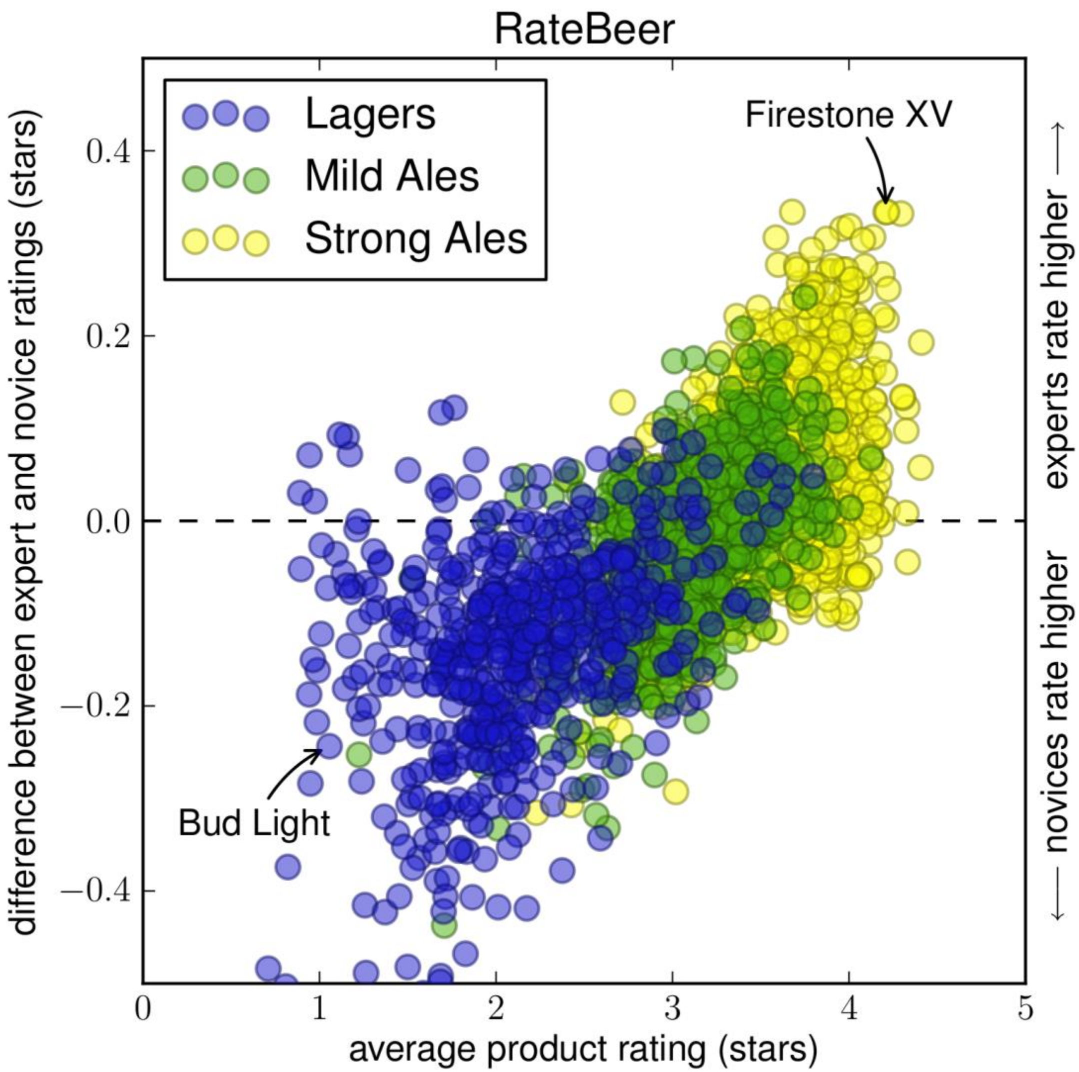
User retention:

What happens to users who **fail** to acquire taste for a product?



Users who acquire tastes slowly are more likely to quit the community

Outcomes – understanding



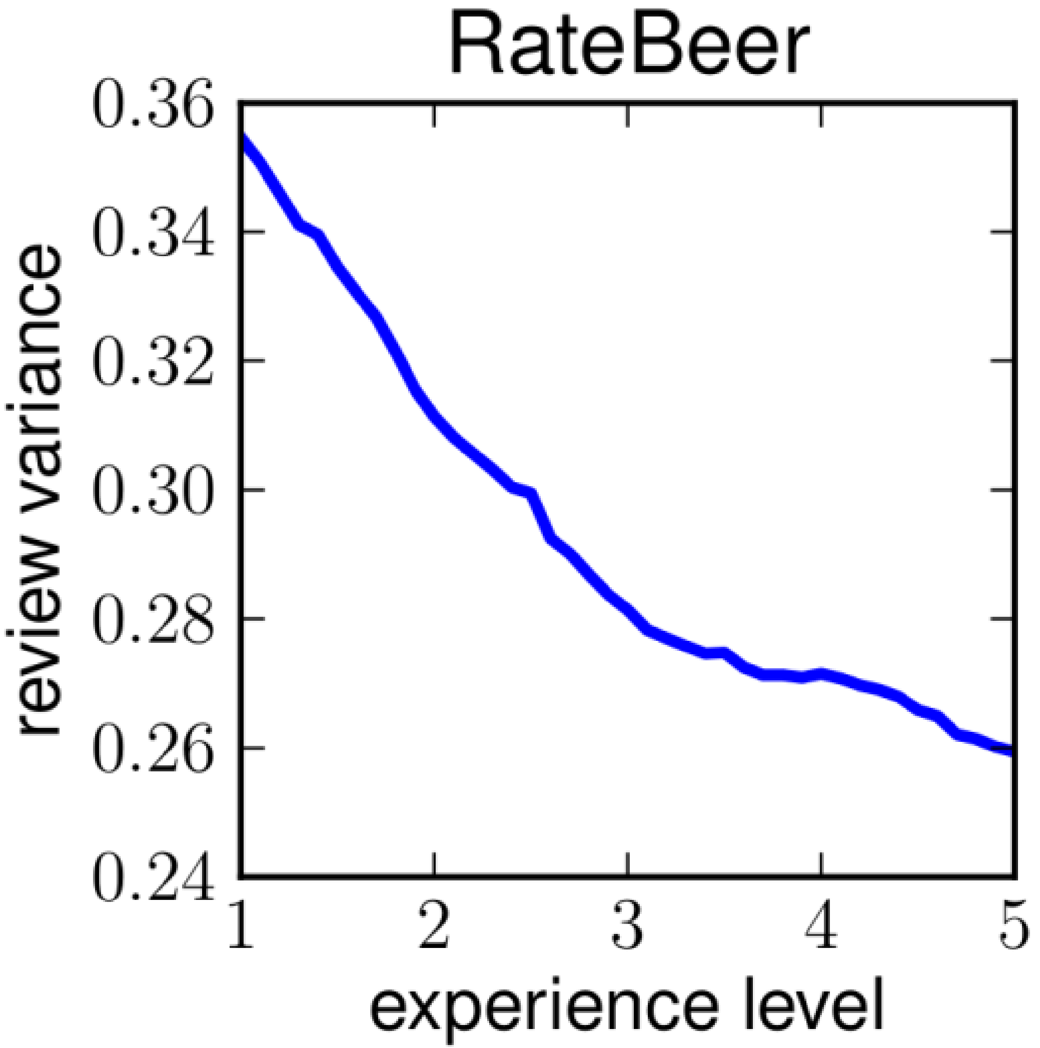
1. Entire categories of products tend to be preferred by experts or beginners

2. Experts rate the top products more generously, and the bottom products more harshly

Outcomes – understanding

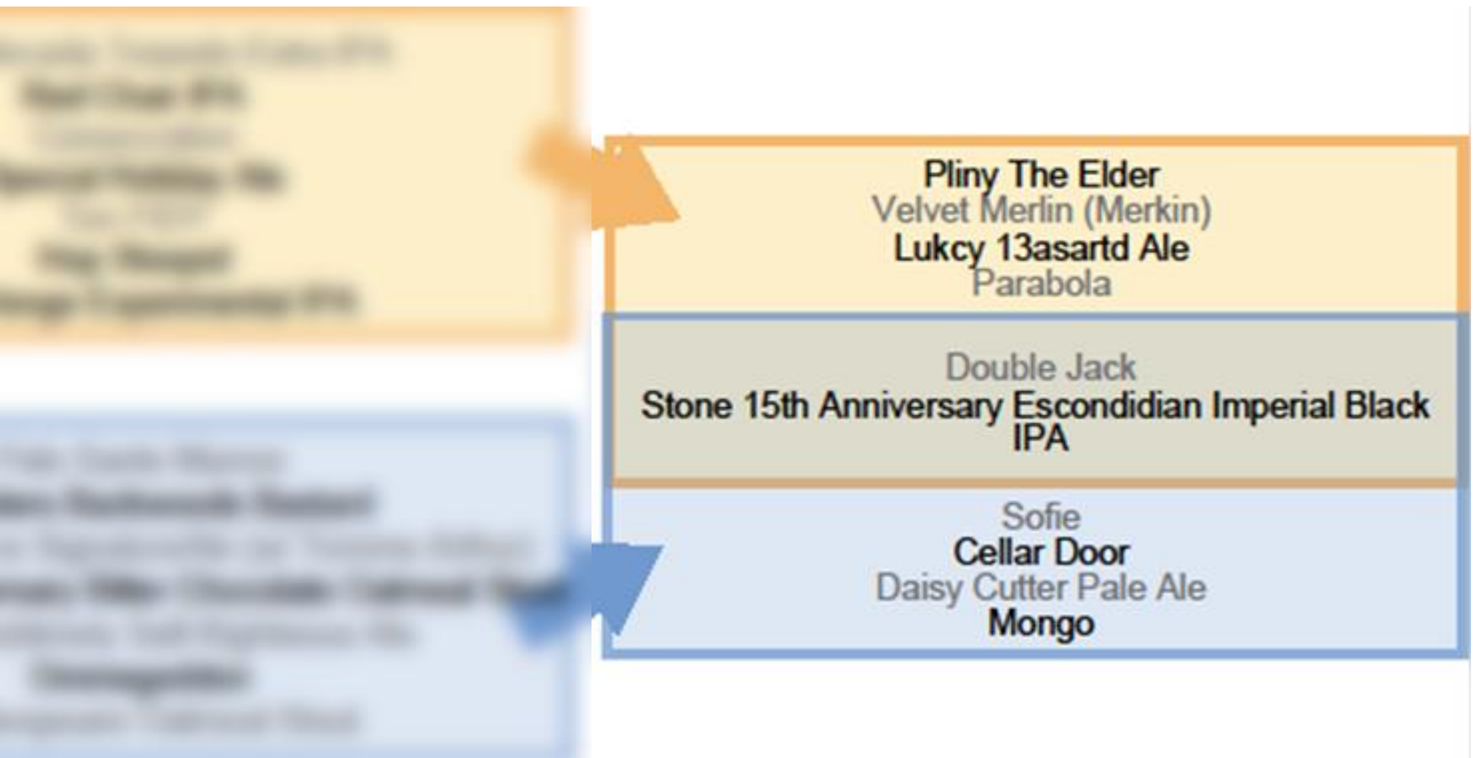
What are experts?
Predictability and agreement are two necessary conditions (Einhorn, 1974)

Experts are more predictable than beginners. They are also more inclined to agree with each other (right).



Extensions: Multiple progression classes

1. We extended our model to allow for multiple "classes" of progression:



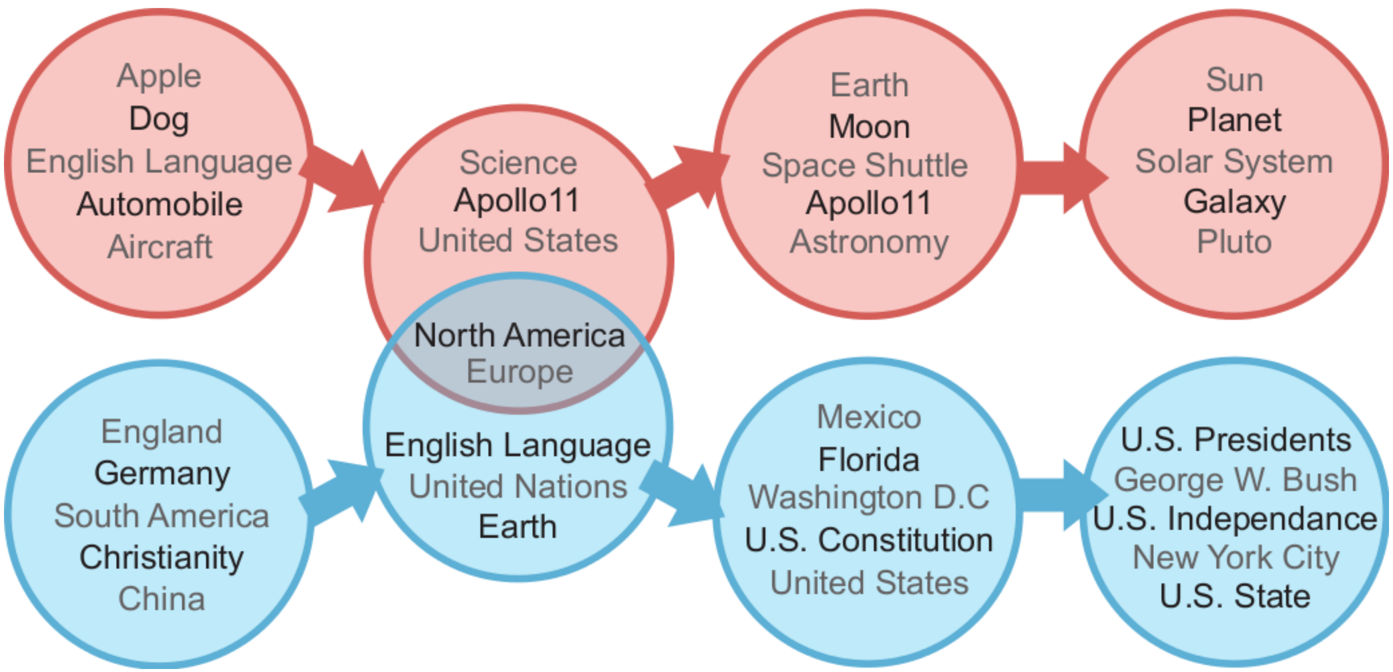
"Class 1" users

"Class 2" users

WWW 2014 (w/ Yang, Leskovec, LePendu & Shah)

Extensions: Other types of data

2. We modeled to data from other domains, including web navigation traces

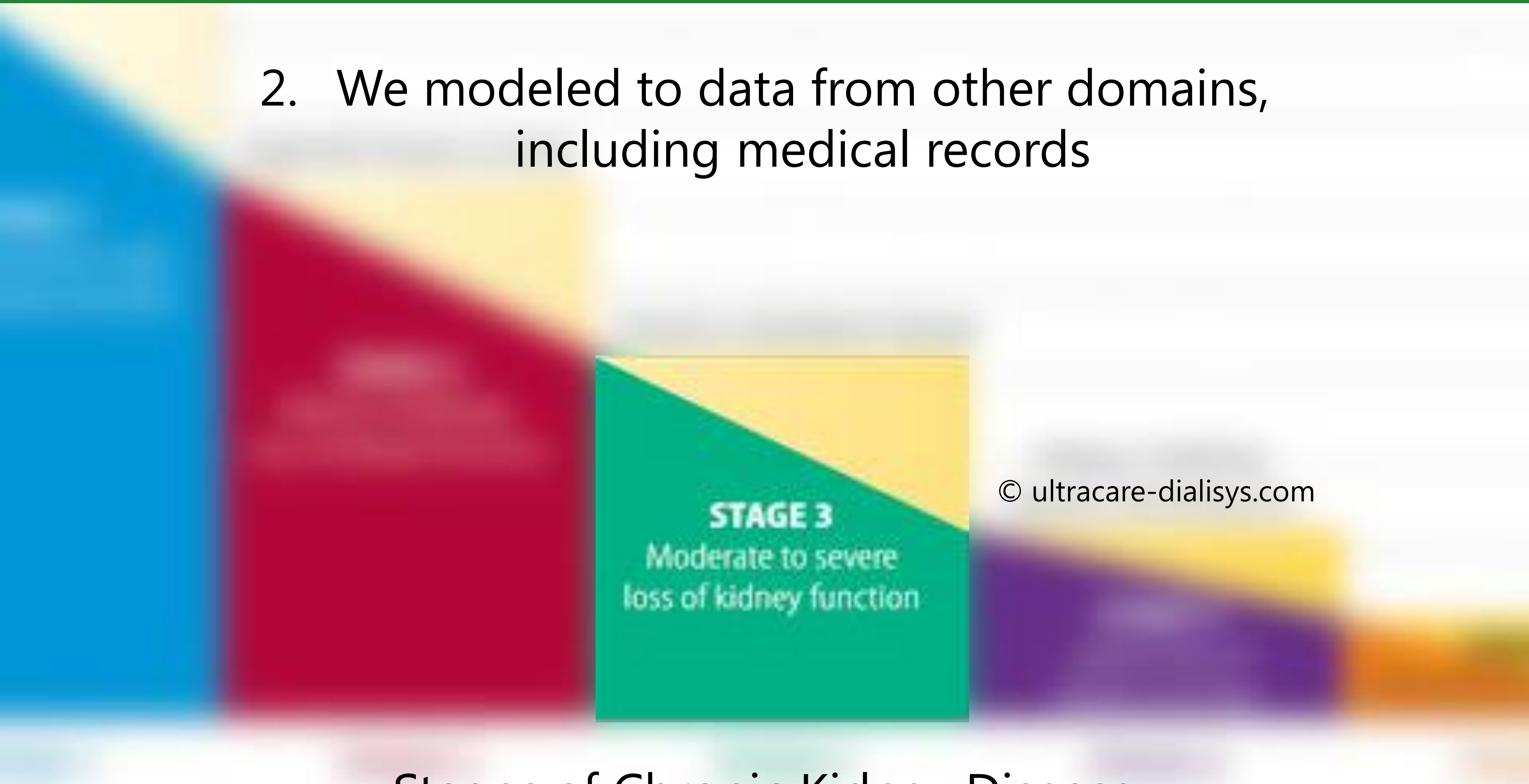


Browsing strategies for two classes of articles (from "Wikispeedia")

WWW 2014 (w/ Yang, Leskovec, LePendu & Shah)

Extensions: Other types of data

2. We modeled to data from other domains, including medical records



© ultracare-dialisys.com

Stages of Chronic Kidney Disease

WWW 2014 (w/ Yang, Leskovec, LePendu & Shah)

Questions?