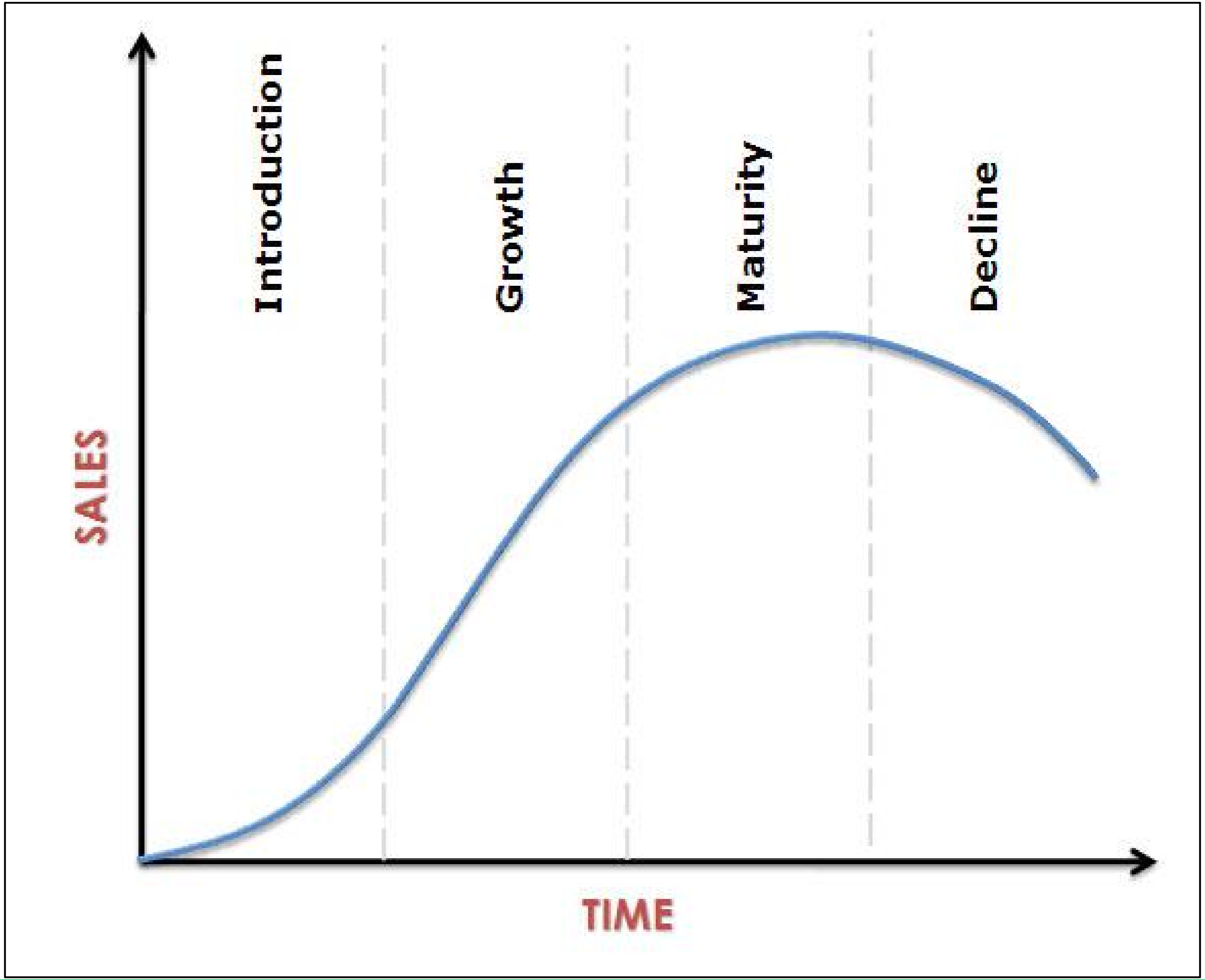
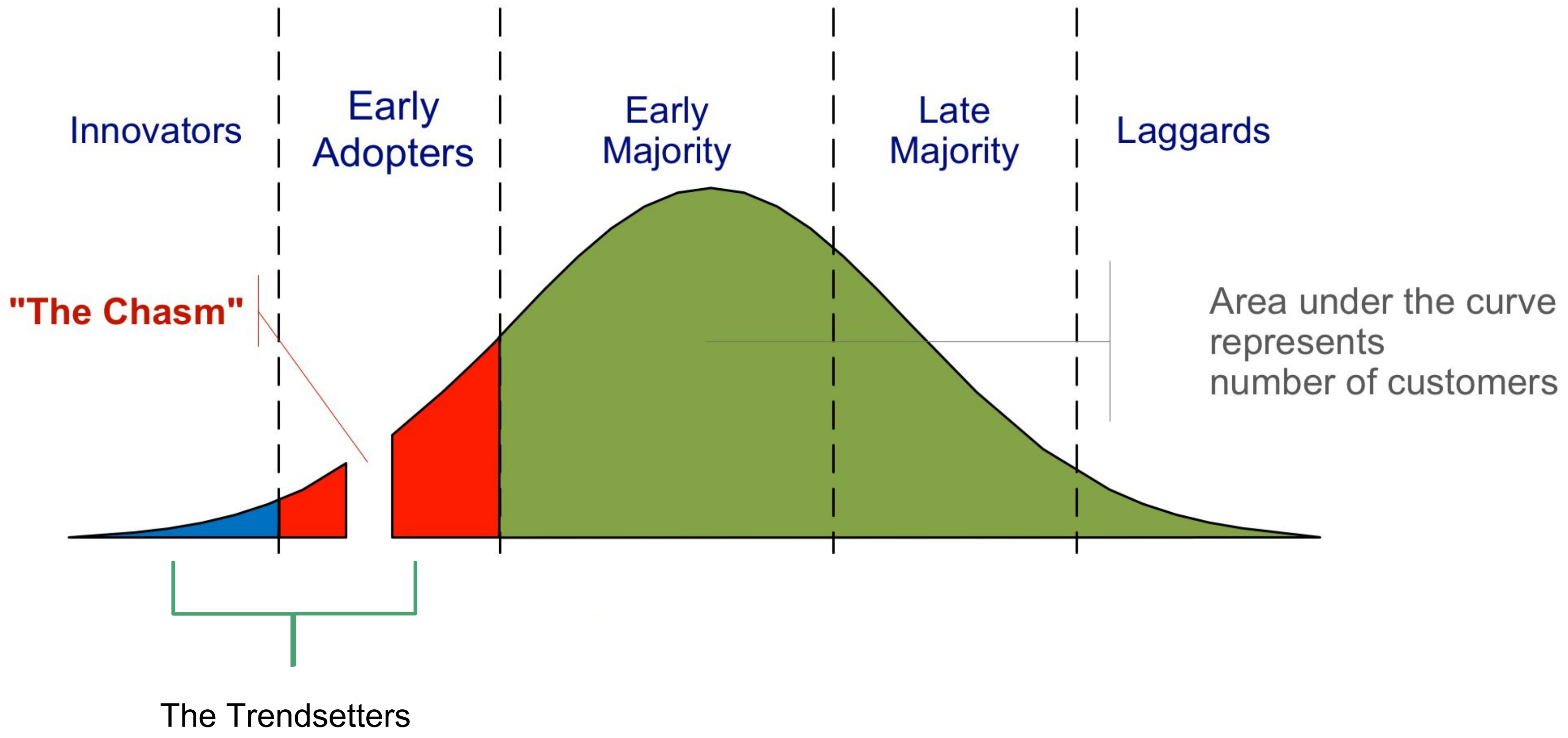
Trendiness & Influence

Identifying & Analyzing Yelp’s most trendy users and their levels of influence within their social circles

Table of Contents

1. Introduction
2. Identifying & Defining ‘Trendy’ Businesses
3. Defining & Analyzing ‘Trendy’ Users
4. Defining & Analyzing ‘Trendy’ User Influence
5. Case Study
6. Value for Businesses & Moving Forward


Introduction
Growth
Maturity
Decline
SALES
TIME


Innovators
Early Adopters
Early Majority
Late Majority
Laggards
"The Chasm"
Area under the curve represents number of customers
The Trendsetters

Not All Users Are Equal

Large Datasets

+1 Million Users

```
{
    "user_id":"encrypted user id",
    "name":"first name",
    "review_count":number of reviews,
    "yelping_since": date formatted like "2009-12-19",
    "friends":["an array of encrypted ids of friends"],
    "useful":"number of useful votes sent by the user",
    "funny":"number of funny votes sent by the user",
    "cool":"number of cool votes sent by the user",
    "fans":"number of fans the user has",
    "elite":["an array of years the user was elite"],
    "average_stars":floating point average like 4.31,
    "compliment_hot":number of hot compliments received by the use
    "compliment_more":number of more compliments received by the
    "compliment_profile": number of profile compliments received
    "compliment_cute": number of cute compliments received by the
    "compliment_list": number of list compliments received by the
    "compliment_note": number of note compliments received by the
    "compliment_plain": number of plain compliments received by th
    "compliment_cool": number of cool compliments received by the
    "compliment_funny": number of funny compliments received by th
    "compliment_writer": number of writer compliments received by
    "compliment_photos": number of photo compliments received by
    "type":"user"
}
```

+4 Million Reviews

```
{
    "review_id":"encrypted review id",
    "user_id":"encrypted user id",
    "business_id":"encrypted business id",
    "stars":star rating, rounded to half-stars,
    "date":"date formatted like 2009-12-19",
    "text":"review text",
    "useful":number of useful votes received,
    "funny":number of funny votes received,
    "cool": number of cool review votes received,
    "type": "review"
}
```

+500k Businesses

```
{
    "business_id":"encrypted business id",
    "name":"business name",
    "neighborhood":"hood name",
    "address":"full address",
    "city":"city",
    "state":"state -- if applicable --",
    "postal code":"postal code",
    "latitude":latitude,
    "longitude":longitude,
    "stars":star rating, rounded to half-stars,
    "review_count":number of reviews,
    "is_open":0/1 (closed/open),
    "attributes":["an array of strings: each array element is an attribute"]
    "categories":["an array of strings of business categories"],
    "hours":["an array of strings of business hours"],
    "type": "business"
}
```

We’ll be analyzing businesses in Cleveland, which has 3,838 businesses in our dataset

Table of Contents

1. Introduction
2. Identifying & Defining ‘Trendy’ Businesses
3. Defining & Analyzing ‘Trendy’ Users
4. Defining & Analyzing ‘Trendy’ User Influence
5. Case Study
6. Value for Businesses & Moving Forward

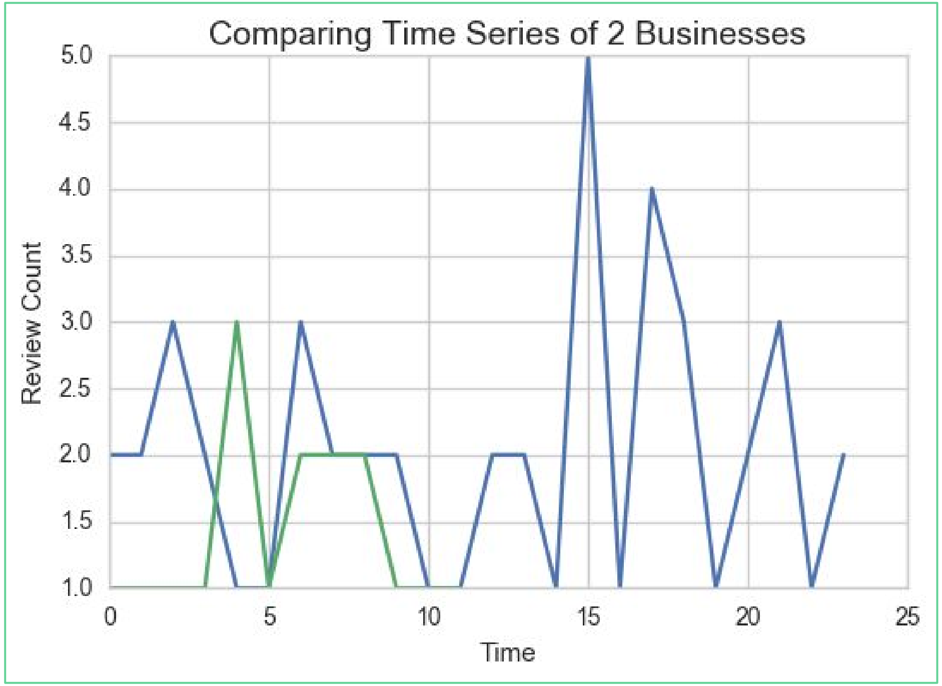
What makes a Business “Trendy”?

Classifying “Trendy” Businesses: Time Series Similarity

What does a “trendy” business look like?

1. Measure similarity between time series of monthly review counts for each business in our dataset

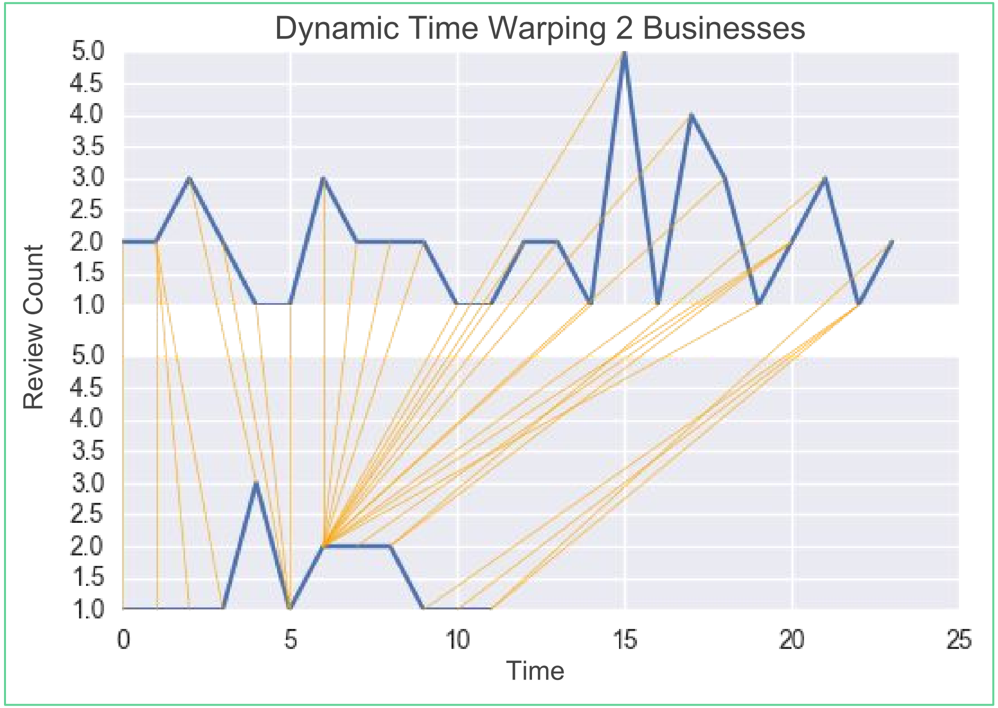
2. However, all of our time series are of different length, making it difficult to compare with basic distance measures

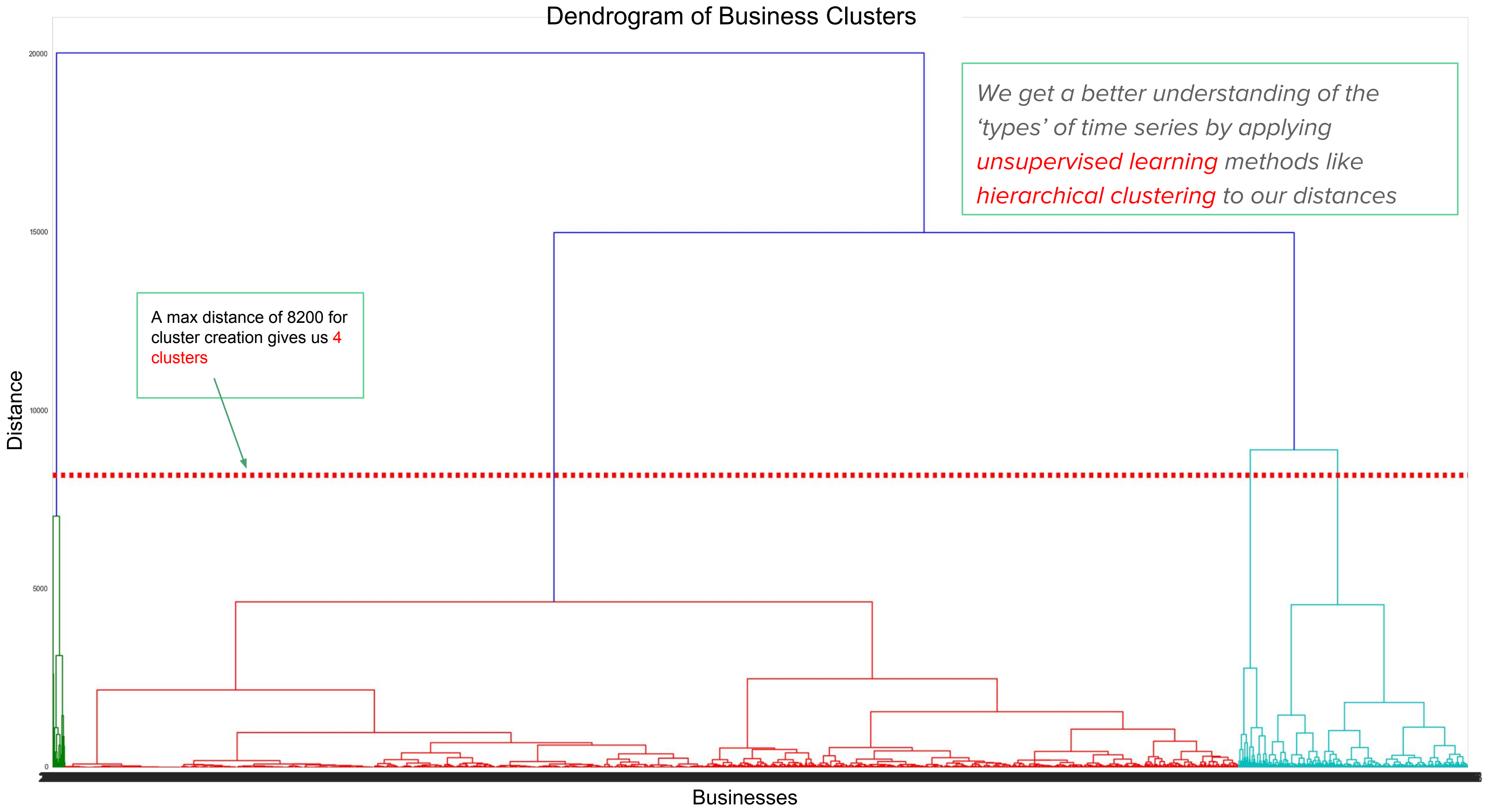


Dynamic Time Warping & Clustering

Dynamic Time Warping allows us to measure similarity of temporal series with different lengths.

Applying DTW to all of our business time series creates a matrix table of distances (similarity) between businesses monthly review count over time

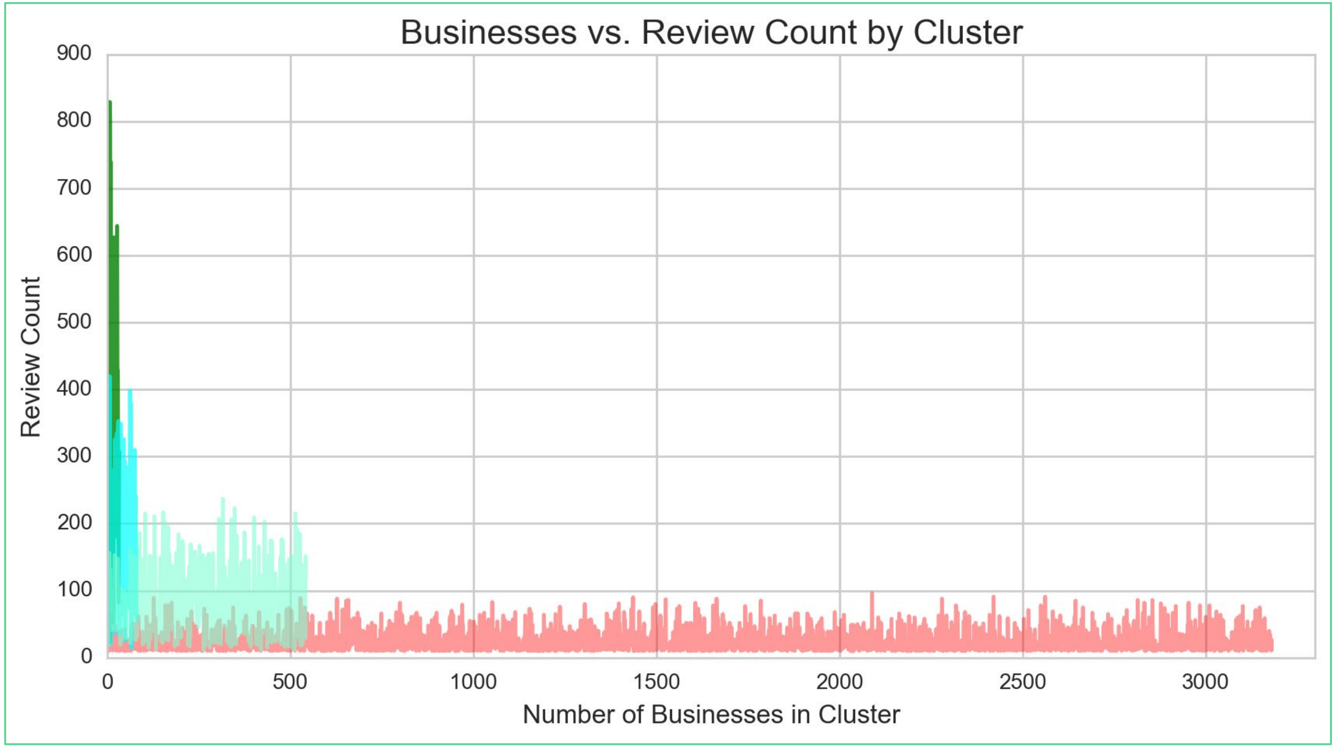




A max distance of 8200 for cluster creation gives us 4 clusters

We get a better understanding of the 'types' of time series by applying unsupervised learning methods like hierarchical clustering to our distances

Plotting Our Clusters

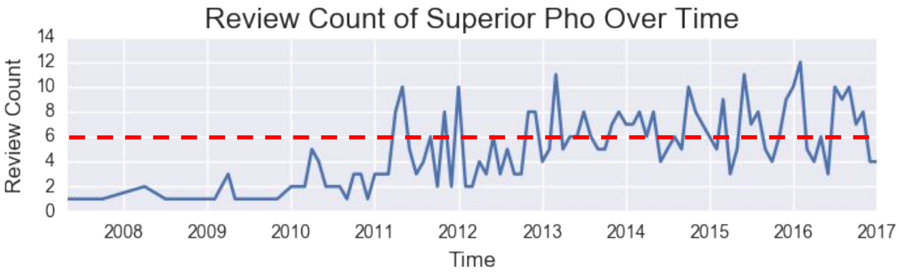


Businesses Per Cluster

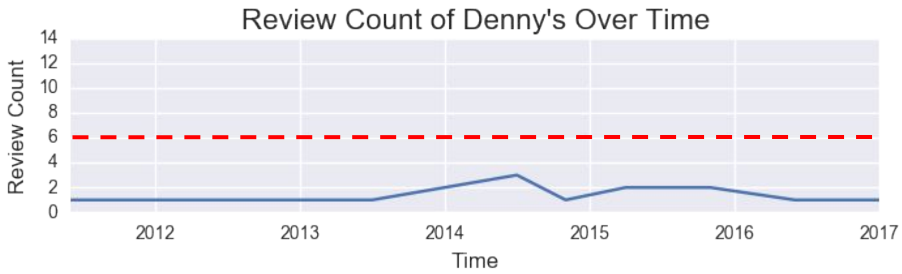
Cluster One:	33
Cluster Two:	80
Cluster Three:	543
Cluster Four:	3182

Times Series Plotting of Clusters

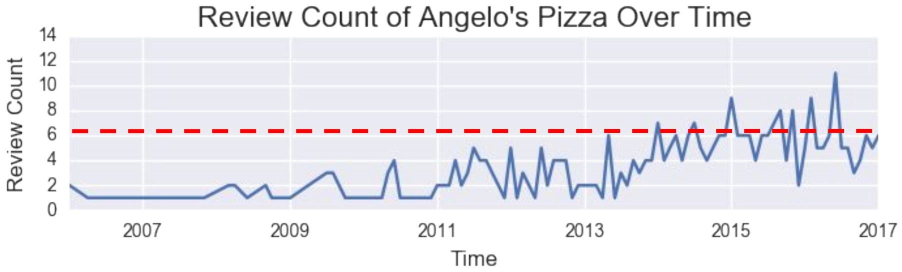
Cluster One



Cluster Three



Cluster Two



Cluster Four

A monthly review count of 6-8 seems to separate our clusters into 2 sections comfortably.

Defining "Trendy" Businesses

Based on our time-series analysis, we can now come up with a concrete definition for "trending" business.

Business Trendiness

- A business in Cleveland considered trending if it experiences at least *2 consecutive months* of *7 or more* monthly reviews.
    - Relatively strict threshold given the monthly average of 1.75

- Number of businesses that meet condition: 136
    - Cluster One + Cluster Two = 113

*Average of Cleveland Businesses

Table of Contents

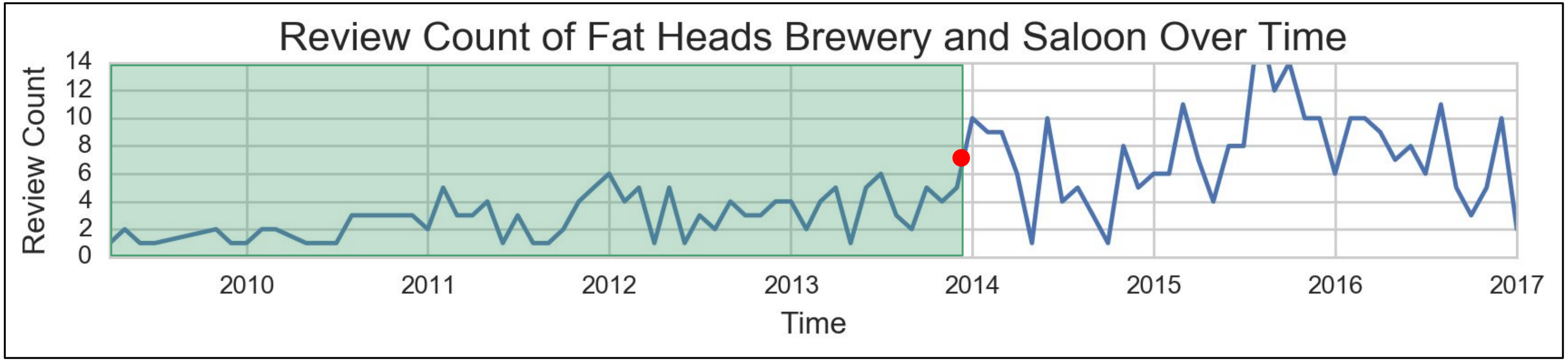
1. Introduction
2. Identifying & Defining ‘Trendy’ Businesses
3. Defining & Analyzing ‘Trendy’ Users
4. Defining & Analyzing ‘Trendy’ User Influence
5. Case Study
6. Value for Businesses & Moving Forward

Identifying “Trendy” Users

Based on our trending business condition, how do we identify ‘trendy’ users?

User Trendiness

- A user’s trendiness can be measured by the number of businesses he or she reviewed <u>*prior*</u> to the business beginning to trend.



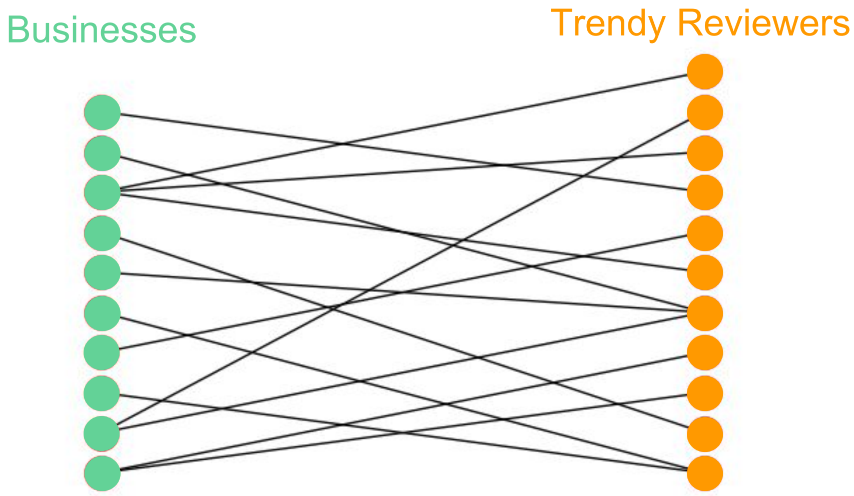
Social Network Analysis On Trendy Users

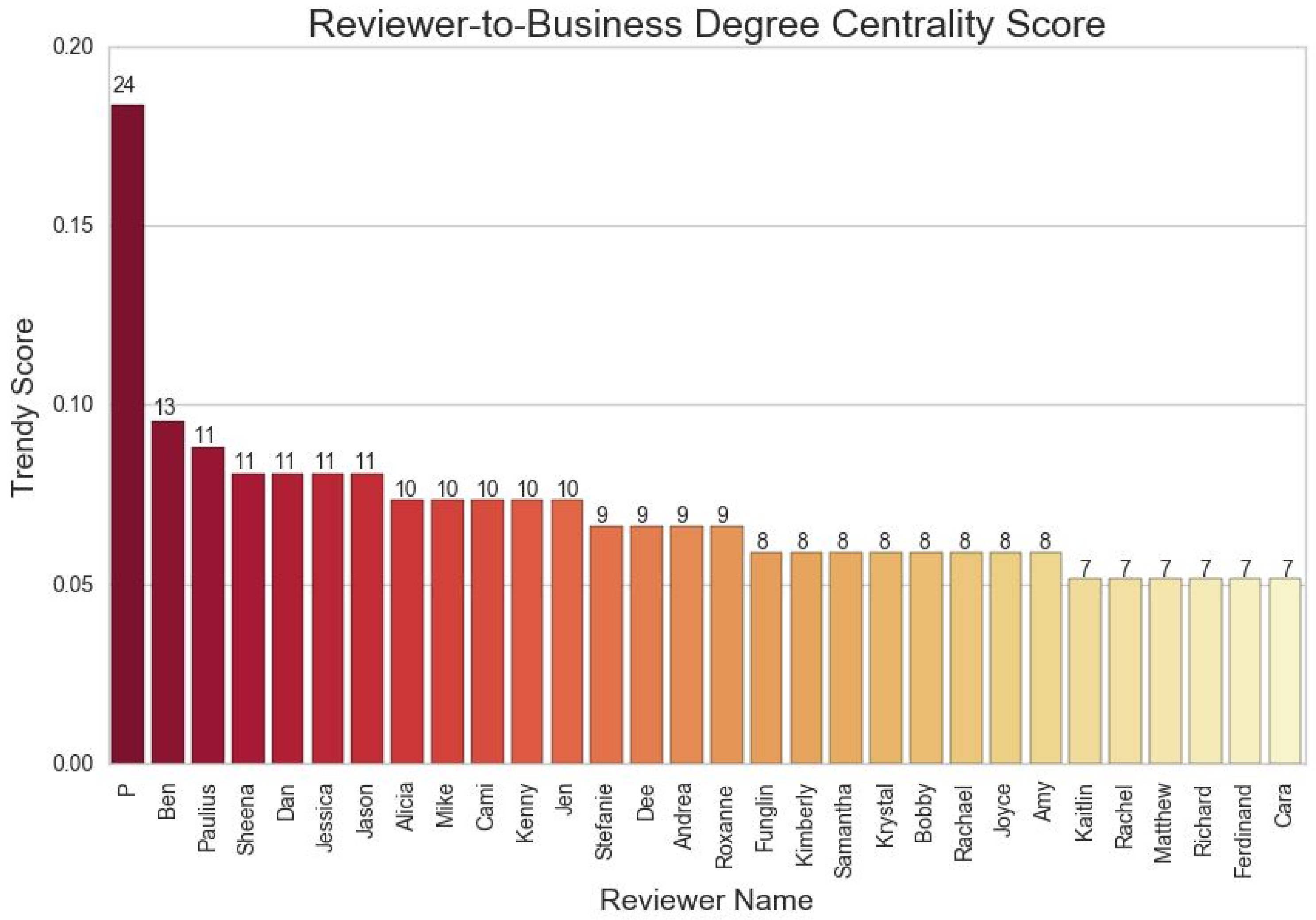
How can we measure each of our pre-trending user's 'trendiness'?

1. Create a bipartite graph with two sets of nodes: businesses & pre-trending users

2. Add edges from users to businesses based on reviews

3. Calculate each reviewer's degree centrality score





Reviewer-to-Business Degree Centrality Score
Trendy Score
0.20
0.15
0.10
0.05
0.00
24
13
11
11
11
11
11
10
10
10
10
10
9
9
9
9
8
8
8
8
8
8
8
8
7
7
7
7
7
7
P
Ben
Paulius
Sheena
Dan
Jessica
Jason
Alicia
Mike
Cami
Kenny
Jen
Stefanie
Dee
Andrea
Roxanne
Funglin
Kimberly
Samantha
Krystal
Bobby
Rachael
Joyce
Amy
Kaitlin
Rachel
Matthew
Richard
Ferdinand
Cara
Reviewer Name

Table of Contents

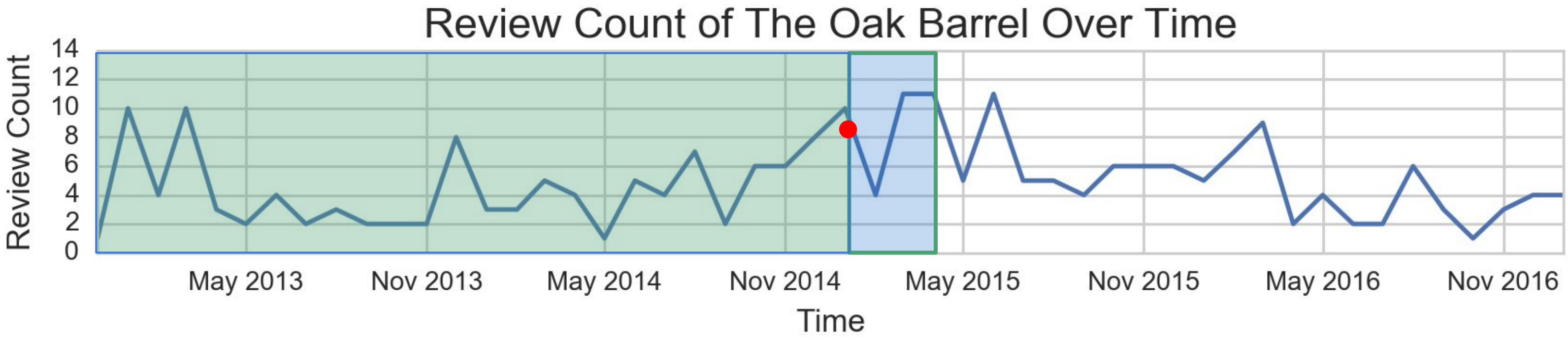
1. Introduction
2. Identifying & Defining ‘Trendy’ Businesses
3. Defining & Analyzing ‘Trendy’ Users
4. Defining & Analyzing ‘Trendy’ User Influence
5. Case Study
6. Value for Businesses & Moving Forward

Defining “Influential” User

Based on our time-series analysis, we can now come up with a concrete definition for “trending” business.

Influential User

- A user’s influence can be measured by the number of his or her friends reviewed the same business *2 months or less* after to the user’s initial review.

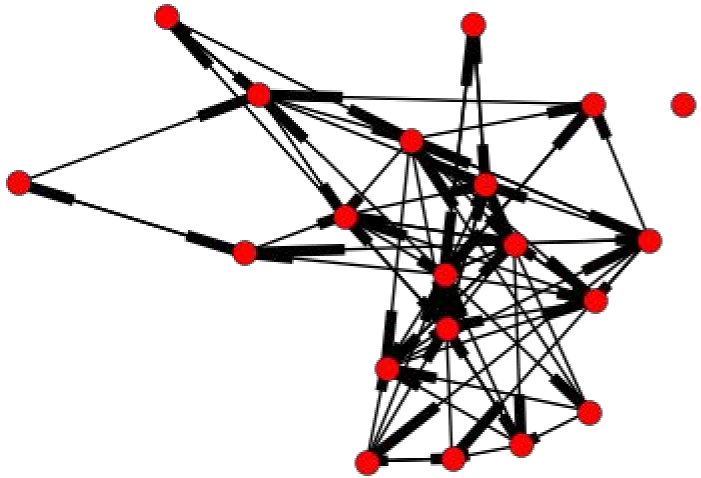


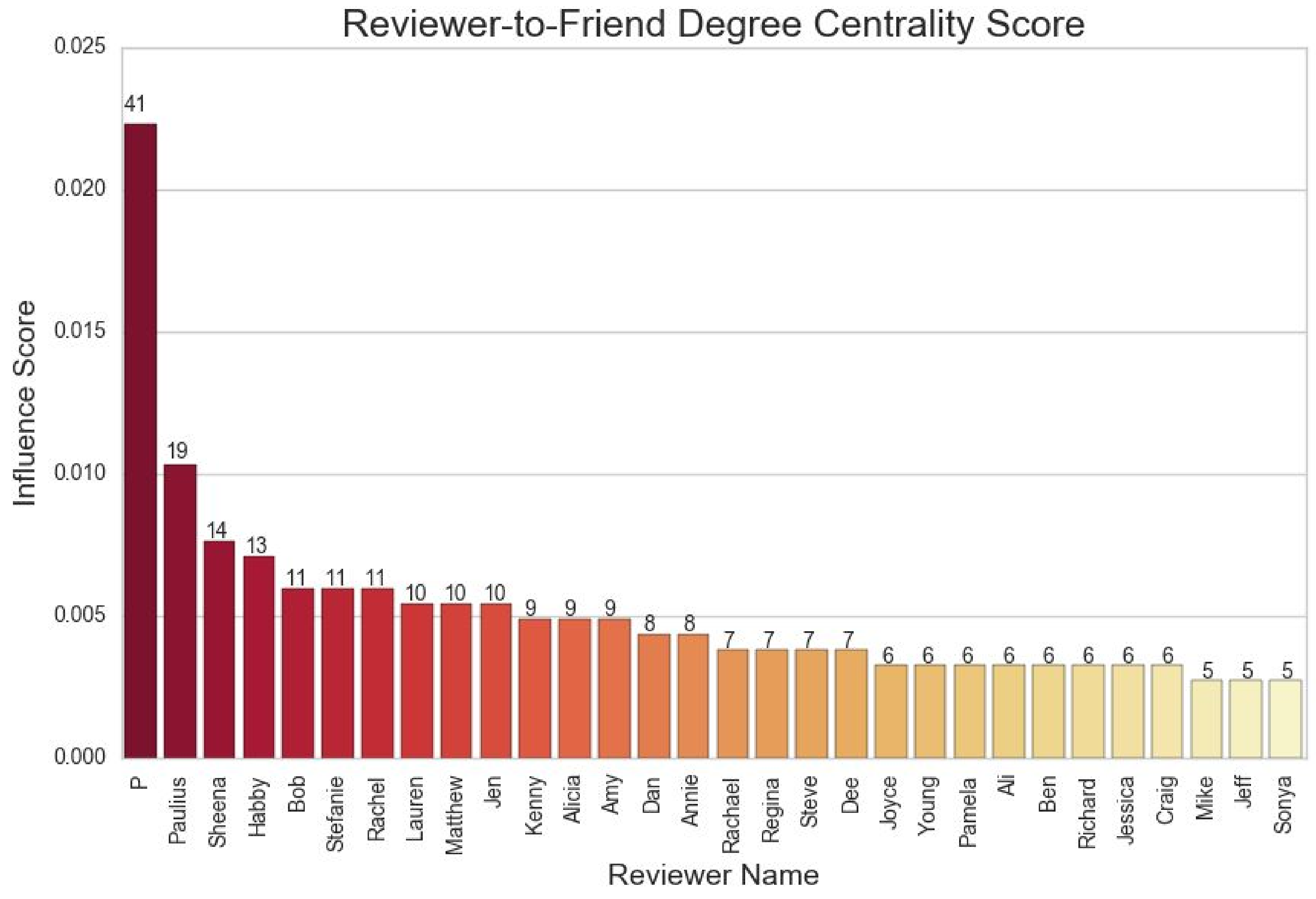
Social Network Analysis on Influence

1. Create a directed graph with reviewers & reviewer's friends influenced

2. Add directed edges from reviewers to friend reviewers influenced

3. Calculate each reviewer's degree centrality score



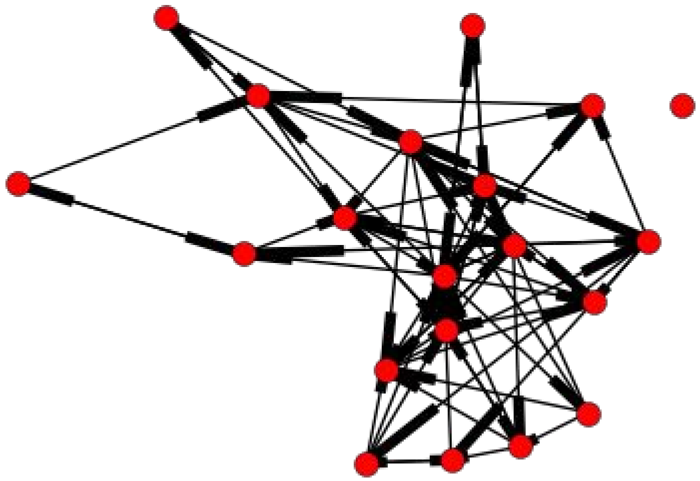


Reviewer-to-Friend Degree Centrality Score
Influence Score
0.025
0.020
0.015
0.010
0.005
0.000
41
19
14
13
11
11
11
10
10
10
9
9
9
8
8
7
7
7
7
6
6
6
6
6
6
6
6
5
5
5
P
Paulius
Sheena
Habby
Bob
Stefanie
Rachel
Lauren
Matthew
Jen
Kenny
Alicia
Amy
Dan
Annie
Rachael
Regina
Steve
Dee
Joyce
Young
Pamela
Ali
Ben
Richard
Jessica
Craig
Mike
Jeff
Sonya
Reviewer Name

Power Law Distribution within Influence Networks

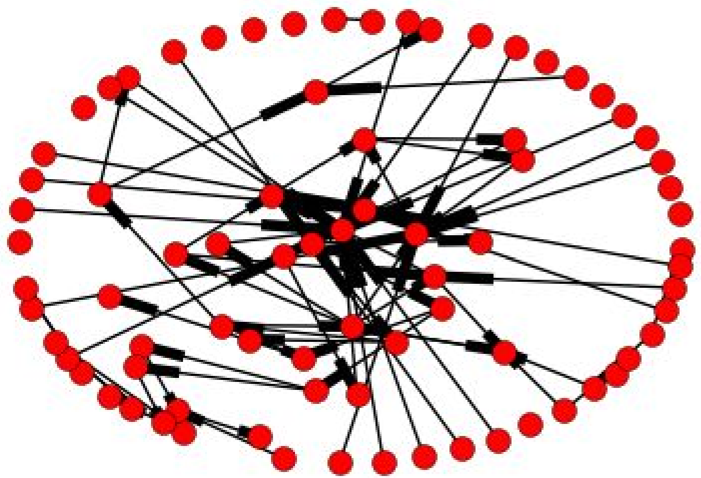
The influential networks on Yelp are skewed heavily, with a few key individuals holding much of the influential power.

Top 20 Influencer's Network



Network Density = 0.19

Influencers Ranked 21th-100th



Network Density = 0.01225

Both networks have a total of 75 influence connections, yet one has 60 less reviewers

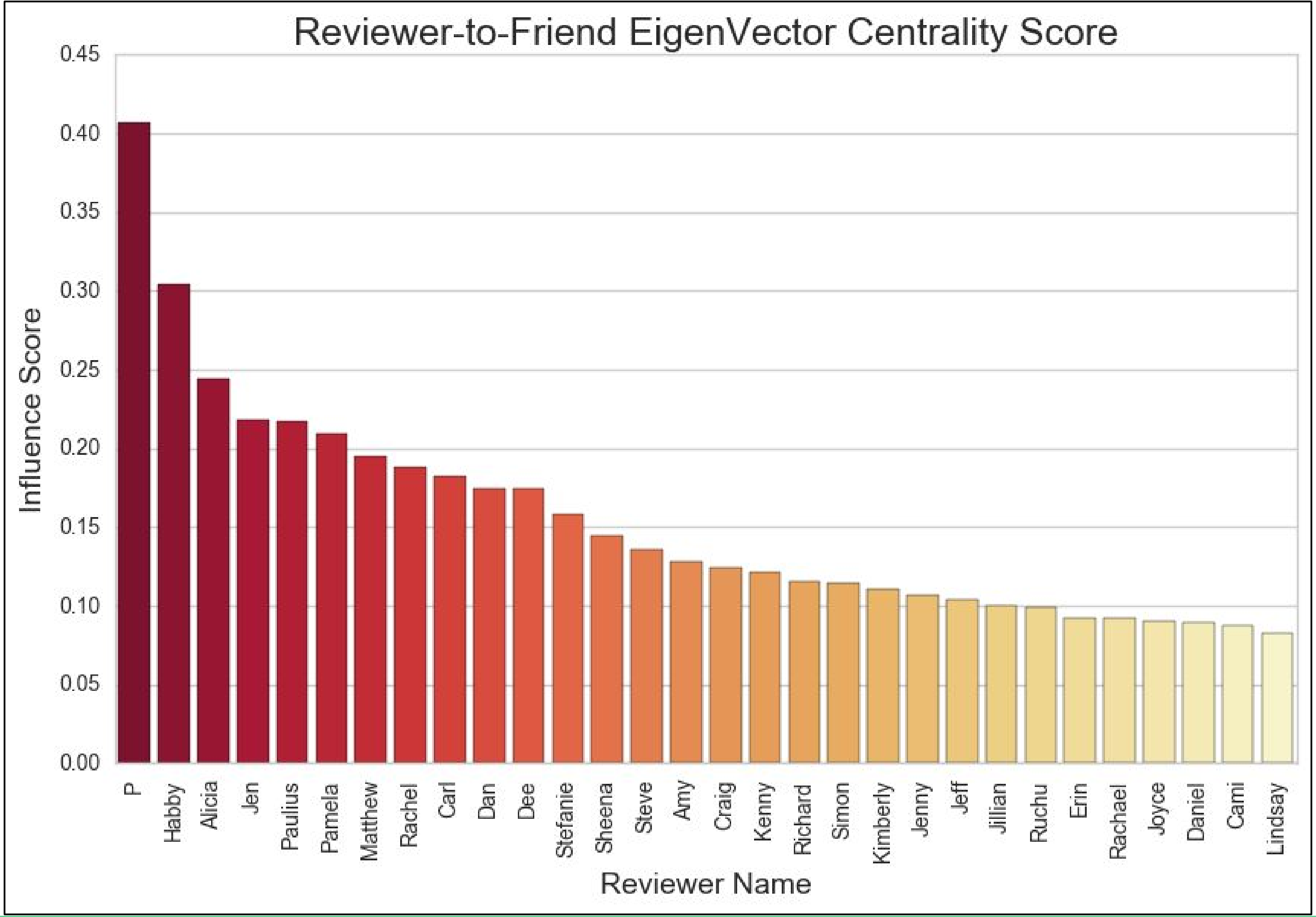

Reviewer-to-Friend EigenVector Centrality Score
Influence Score
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P
Habby
Alicia
Jen
Paulius
Pamela
Matthew
Rachel
Carl
Dan
Dee
Stefanie
Sheena
Steve
Amy
Craig
Kenny
Richard
Simon
Kimberly
Jenny
Jeff
Jillian
Ruchu
Erin
Rachael
Joyce
Daniel
Cami
Lindsay
Reviewer Name

Table of Contents

1. Introduction
2. Identifying & Defining ‘Trendy’ Businesses
3. Defining & Analyzing ‘Trendy’ Users
4. Defining & Analyzing ‘Trendy’ User Influence
5. Case Study
6. Value for Businesses & Moving Forward

Case Study: Porco Lounge and Tiki Room

Porco Lounge and Tiki Room

Claimed

249 reviews · Details

$$ · Tiki Bars · Edit

Michael V.
Chicago, IL
18 friends
293 reviews
35 photos
Elite '17

- Share review
- Embed review
- Compliment
- Send message
- Follow Michael V.

3/28/2017

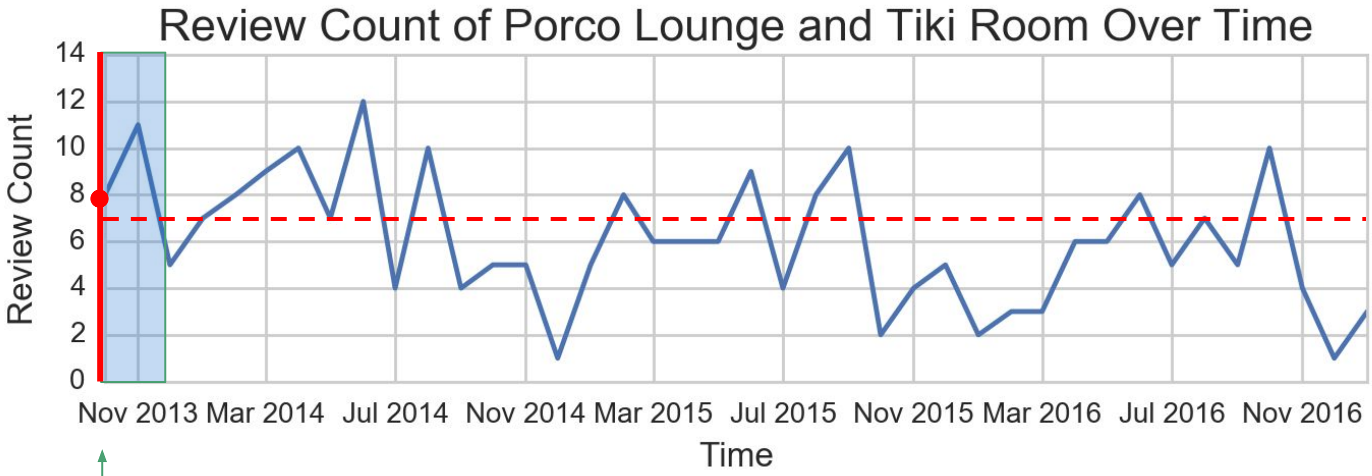
Awesome, awesome Tiki bar!

I was visiting from Chicago and my friends had been talking up Porco all weekend. We finally went on Saturday night and I have to say it exceeded all expectations. Staff was incredibly helpful and friendly. While waiting for a table we grabbed one of the standing ones and proceeded to take our name off the list, but they still came around later and offered a sit down table. Our server gave great drink suggestions to our group and had a stellar sense of humor. The hodgepodge collection of tiki stuff decorating the place really added to the experience too.

I don't really remember all the drinks I ordered but I can tell you I did get the Zombie. 5+ ounces of rum and it was still delicious. Almost ordered another one but then my sensibilities kicked in. Or my friends were leaving. Whichever.

Forget LeBron, Porco's should be the #1 attraction in Cleveland.

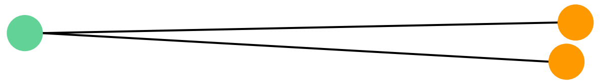
The Value of Influencers



Porco had a strong opening and trended almost immediately

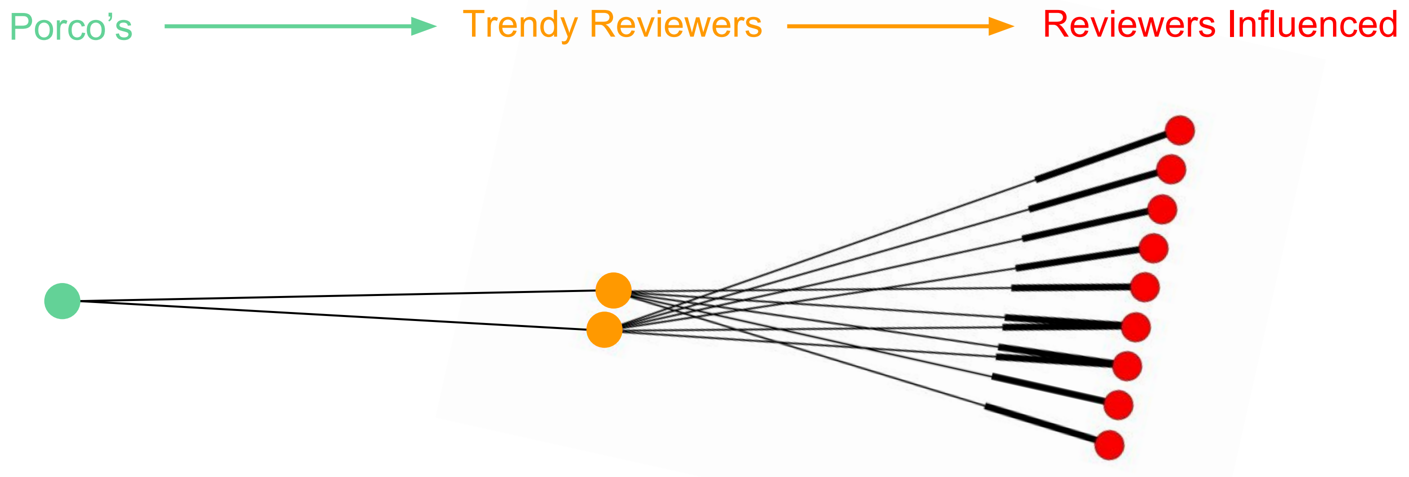
Porco's Trendy Influencer Network: Base Case

Porco's → Trendy, Non-Influential Reviewers



In a baseline scenario, we assume our reviewer's review directly influences none of his or her friend's to write a review.

Porco's Trendy Influencer Network: Influence Case



By targeting the right trendy, influential reviewers, Porco's can acquire at the minimum 4.5x the number of new customers and revenues than compared to a baseline, while also decreasing their average customer acquisition cost.

Table of Contents

1. Introduction
2. Identifying & Defining ‘Trendy’ Businesses
3. Defining & Analyzing ‘Trendy’ Users
4. Defining & Analyzing ‘Trendy’ User Influence
5. Case Study
6. Value for Businesses & Moving Forward

Value

Both Yelp & Local Businesses have much to gain from understanding who the most trendy & influential users in Yelp's network are.

Yelp

- Concretely visualize impact a user's review has on their social circle
- Increase ad conversion rate through better targeting:
  a) Command a greater premium for ads → larger profit margins

Local Businesses

- Know who to target to achieve the optimal levels of growth
- Decrease customer acquisition costs early-on by targeting the most trendy and influential users in the area

Moving Forward

1

2nd-degree connections

Identify 2nd degree connections of influencers to get better understanding of influence across different stages of business

2

Categories

Identify what categories of businesses tend to trend most often, and who their most trendy & influential users

3

City-Level Modeling

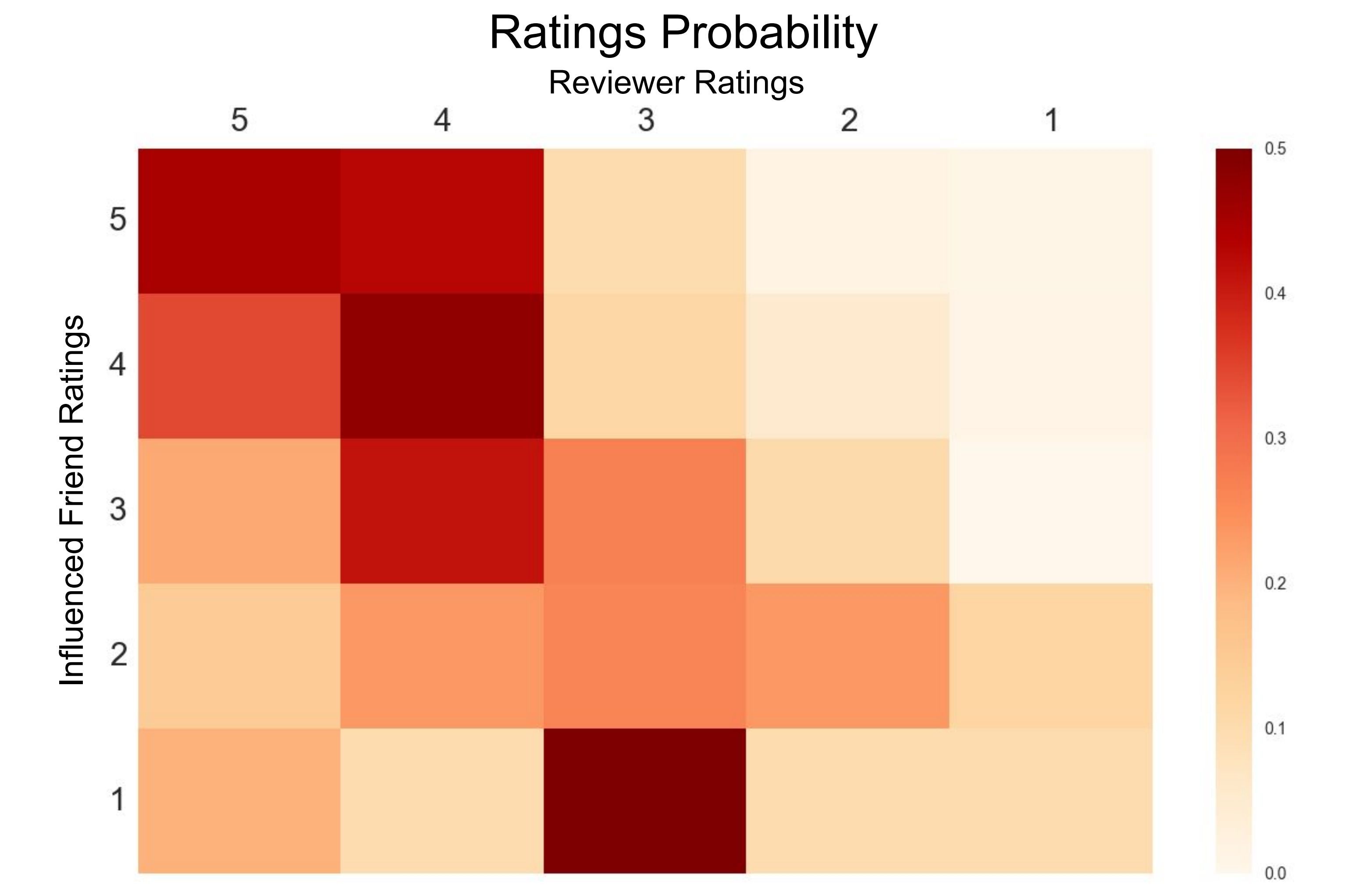
Analyze user's with highest trendy & influence scores and identify common (or different) attributes among them across multiple cities

HUGE Thanks to Everyone!

Thanks to Alex, Vrushank, and Varsha, as well as all my classmates, for all the help they provided me along the way !!!

Questions

Further Research


Ratings Probability
Reviewer Ratings
5
4
3
2
1
Influenced Friend Ratings
5
4
3
2
1
0.5
0.4
0.3
0.2
0.1
0.0