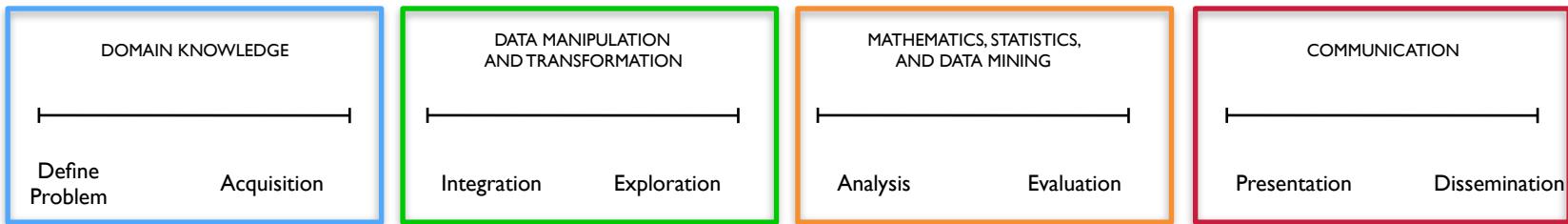




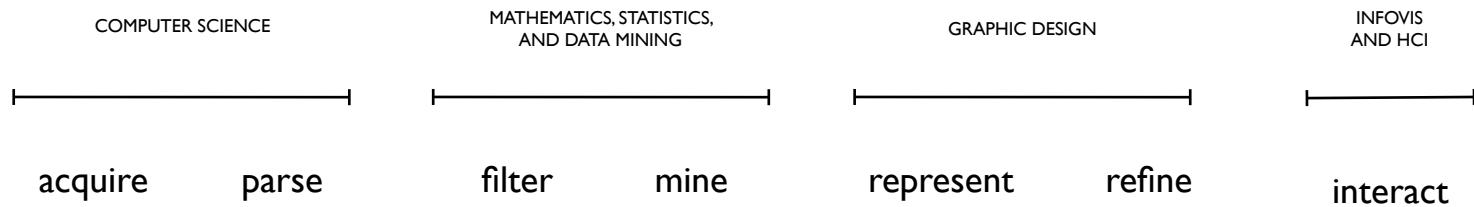
Lesson 2: The Data Science Process —Building your First Application



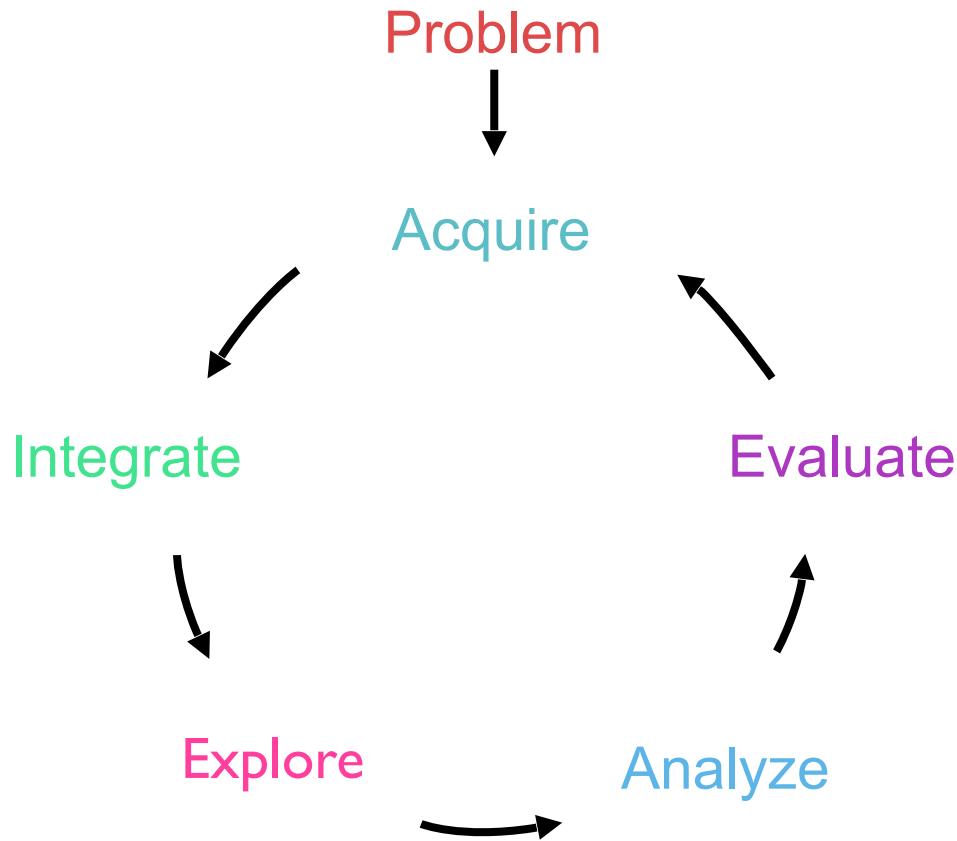
Process



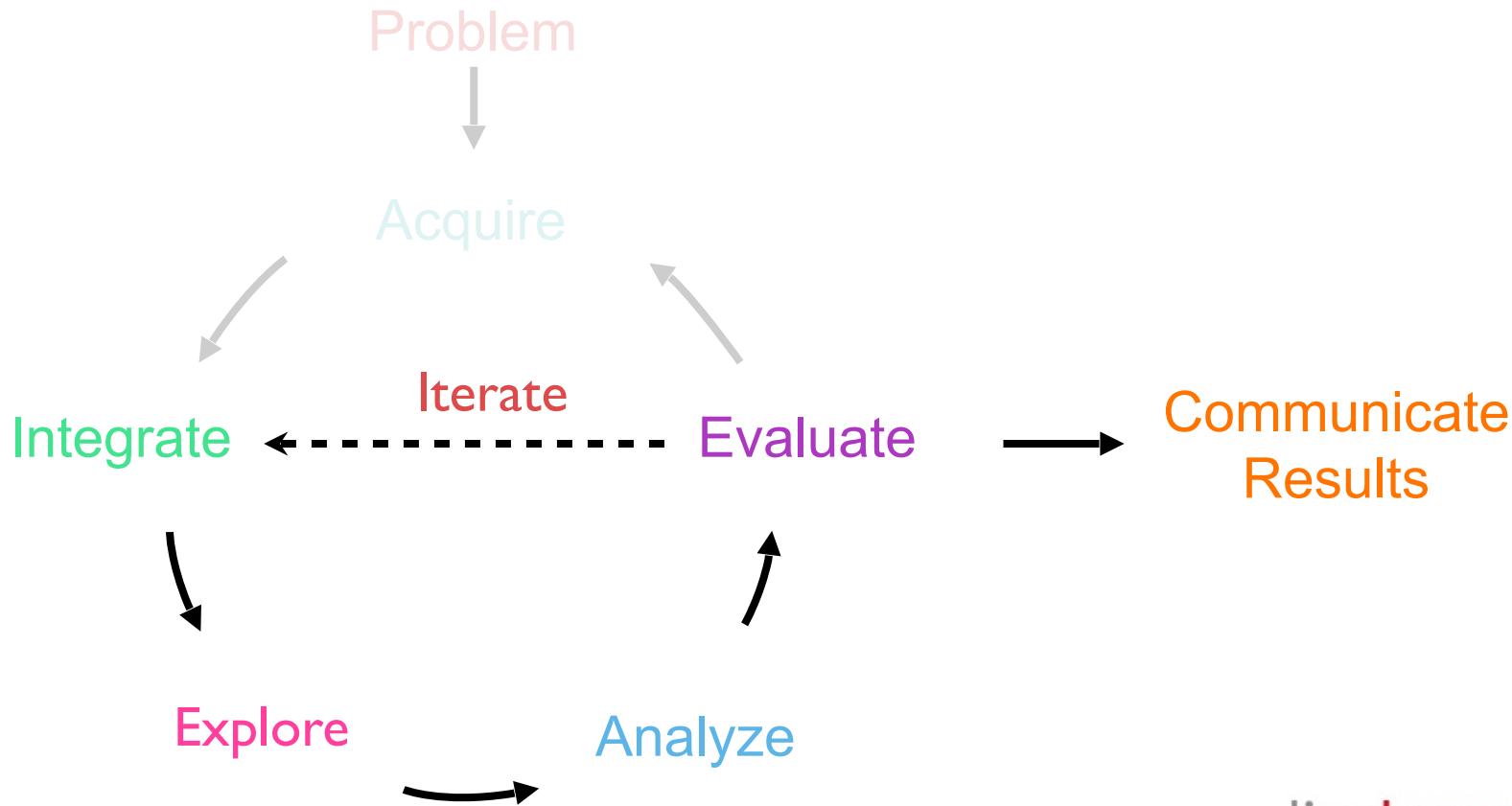
Process



Process + Iteration



Process + Iteration



*** This is merely a suggestion**

Cheaper Beds, Better Breakfasts



Search for people, jobs, companies, and more...



Advanced



28



5



Home

Profile

Connections

Jobs

Interests

Business Services

Try Premium for free



People You May Know

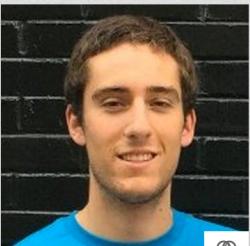


Ethan

Director, IT at Kite Pharma

24

Connect

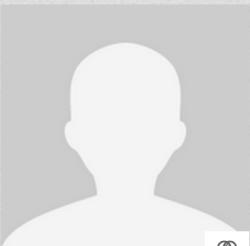


Gideon

Data Scientist

17

Connect



Alana

Strategic Partnerships

9

Connect



johnny Bigwig

owner at TLC tree cutting

1

Connect



Alice

7



Cian

Data Scientist at Stitch Fix

3



Jessica

Instructional Coach at General Assembly

3



Yuriy

Data Scientist

5

[Home](#)[Moments](#)[Notifications](#)[Messages](#)

Search Twitter

[Tweet](#)**Hopeless Optimism**

@jonathanhandinu

TWEETS
3,627FOLLOWING
158FOLLOWERS
737

Trends · Change

Paul Ryan

16.6K Tweets

#CEM7

4,263 Tweets

#forecast2016**#LoveIsLove**

319K Tweets

#WearOrange

74.9K Tweets

#ThursdayThoughts

45K Tweets

#ThinkContent

4,736 Tweets

Nick Denton

2,152 Tweets

#wine

11.4K Tweets

#channelvisionaries

What's happening?



Carnegie Mellon Retweeted

CMU Silicon Valley @cmusv · 1h

It was great having @CarnegieMellon Provost Farnam Jahanian visit us in #SiliconValley!



1



2

**Partially Derivative @partiallyd** · 3mNo joke, someone made a rap about linear equations! Of course we're excited
partiallyderivative.com/podcast/2016/0... h/t @MonaChalabi**S2E10: I'm Not Lying, I'm a Line**

The data of the world around us

partiallyderivative.com

Who to follow · Refresh · View all



Andreas M Kavountzis @am...



Jim Halberg @jimhalberg



Shaun Kane @shaunkane



Find people you know

Import your contacts from Gmail

Connect other address books

© 2016 Twitter About Help Terms Privacy
 Cookies Ads info Brand Blog Status Apps
 Jobs Businesses Media Developers

Advertise with Twitter

[Home](#)[Moments](#)[Notifications](#)[Messages](#)

Search Twitter



Tweet

**Hopeless Optimism**

@jonathanhandinu

TWEETS
3,627FOLLOWING
158FOLLOWERS
737

Trends · Change

Paul Ryan

16.6K Tweets

#CEM7

4,263 Tweets

#forecast2016**#LoveIsLove**

319K Tweets

#WearOrange

74.9K Tweets

#ThursdayThoughts

45K Tweets

#ThinkContent

4,736 Tweets

Nick Denton

2,152 Tweets

#wine

11.4K Tweets

#channelvisionaries

What's happening?



Carnegie Mellon Retweeted

CMU Silicon Valley @cmusv · 1h

It was great having @CarnegieMellon Provost Farnam Jahanian visit us in #SiliconValley!



1



2

**Partially Derivative @partiallyd** · 3mNo joke, someone made a rap about linear equations! Of course we're excited
partiallyderivative.com/podcast/2016/0... h/t @MonaChalabi**S2E10: I'm Not Lying, I'm a Line**

The data of the world around us

partiallyderivative.com

Who to follow · Refresh · View all

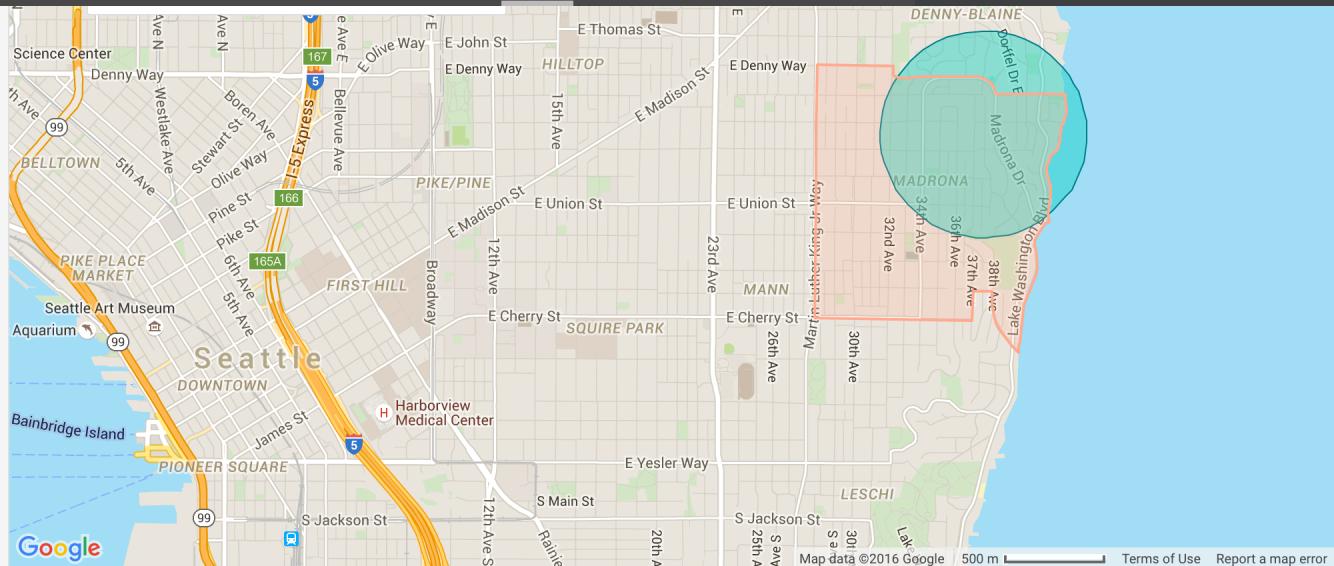
**Andreas M Kavountzis @am...****Jim Halberg @jimhalberg****Shaun Kane @shaunkane****Find people you know**

Import your contacts from Gmail

Connect other address books

© 2016 Twitter About Help Terms Privacy
 Cookies Ads info Brand Blog Status Apps
 Jobs Businesses Media Developers

Advertise with Twitter



Similar Listings



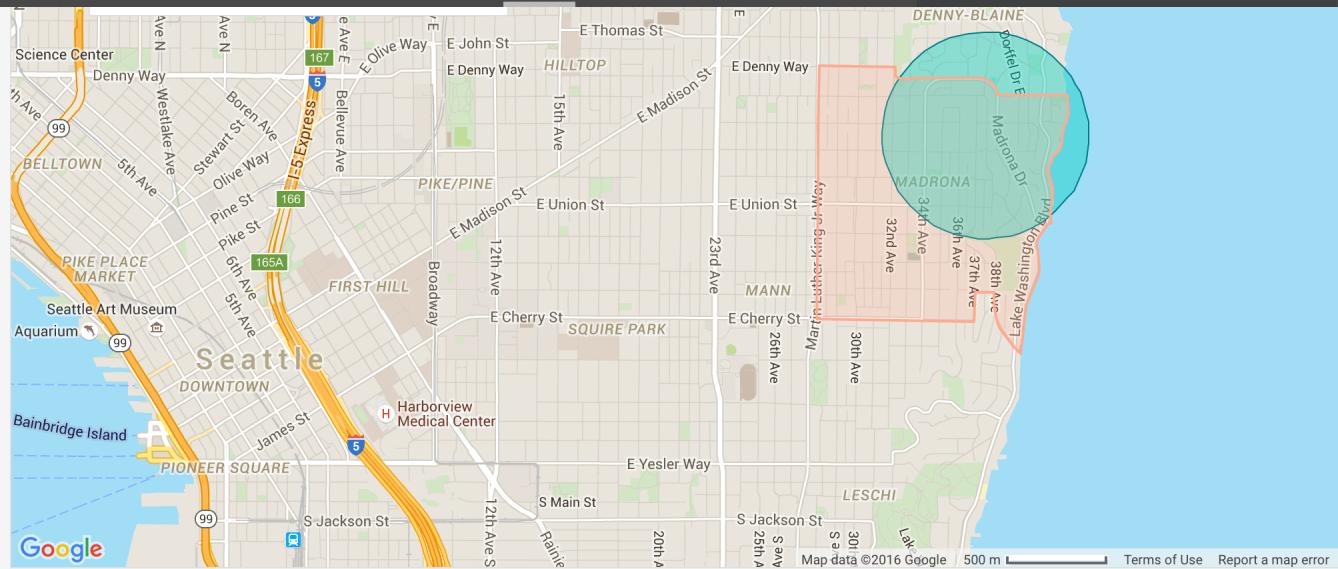
Guest House in Madison Vall...
Entire home/apt — 0.7 miles away



Private Apartment - Queen A...
Entire home/apt — 3.7 miles away



Perfect Location by Lake Wa...
Entire home/apt — 1.3 miles away



Recommendations!



Similar Listings



Guest House in Madison Vall...
Entire home/apt — 0.7 miles away



Private Apartment - Queen A...
Entire home/apt — 3.7 miles away



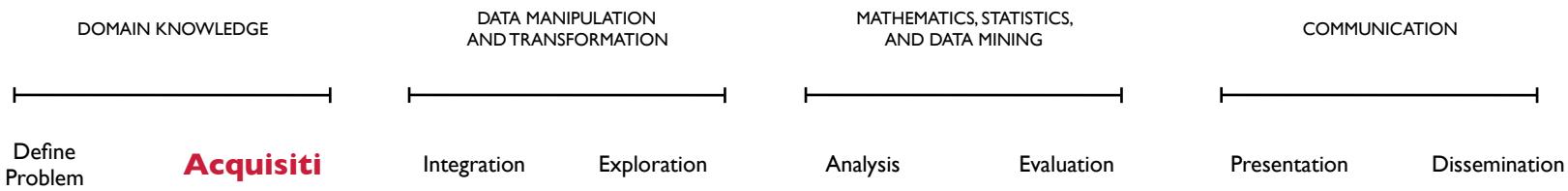
Perfect Location by Lake Wa...
Entire home/apt — 1.3 miles away

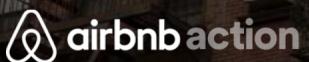
**Increase the amount of reservations made on the
AirBnB platform**

Aid users in their discovery of relevant listings

Aid users in their discovery of relevant listings

Process





MENU

DEC 1, 2015

DATA ON THE AIRBNB COMMUNITY IN NYC

[Support Home Sharing](#)

Your Email Address *

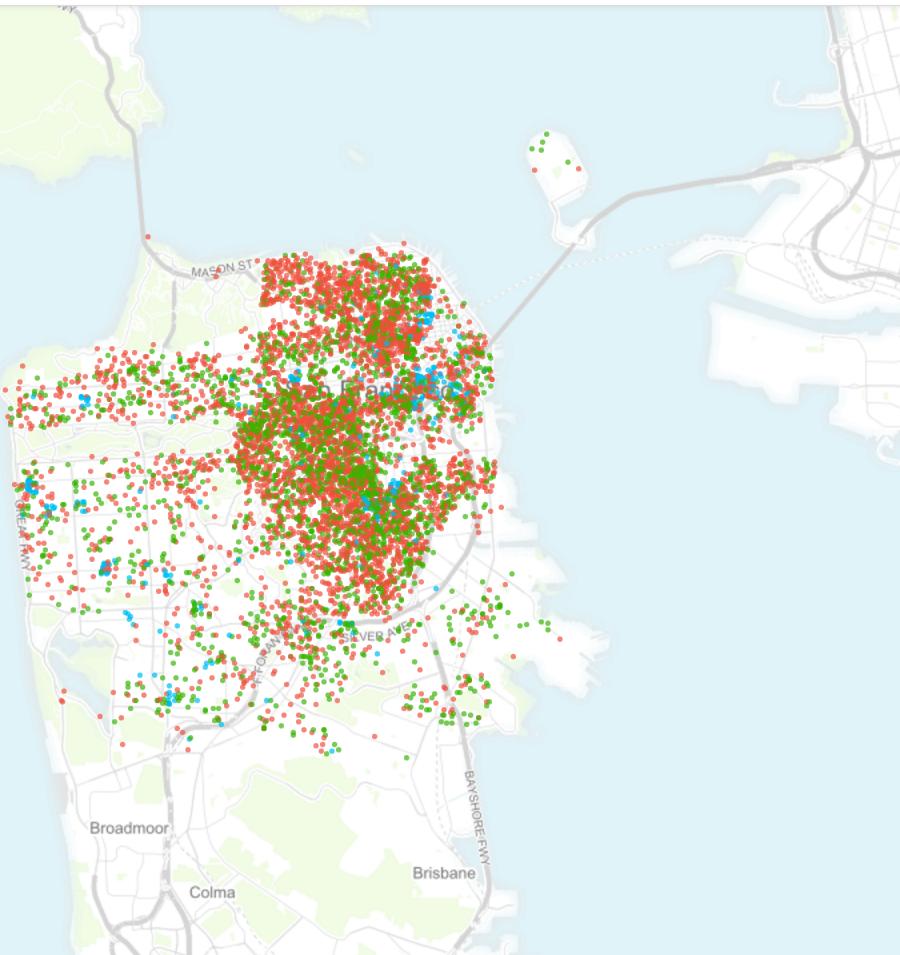
United States ▾

[Join Us](#)

* indicates a required field

Summary

Airbnb has always been focused on helping travelers belong anywhere and giving hosts the chance to welcome people from around the world and earn a little extra money to pay the bills. We want more people to understand who we are, what we're all



San Francisco

Filter by:

San Francisco

7,029
out of 7,029 listings (100%)

[About Airbnb in San Francisco](#)

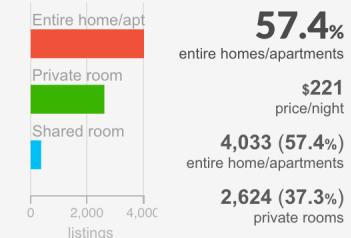
How is Airbnb really being used in and affecting your neighborhoods?

Room Type

Only entire homes/apartments

Airbnb hosts can list entire homes/apartments, private or shared rooms.

Depending on the room type, [availability](#), and [activity](#), an airbnb listing could be more like a hotel, disruptive for neighbors, taking away housing, and [illegal](#).



Activity

Only [recent](#) and [frequently](#) booked

Airbnb guests may leave a review after their stay, and these can be used as an indicator of airbnb activity.

The minimum stay, price and number of reviews have been used to estimate the [occupancy rate](#), the percentage of time a listing is available for rent.

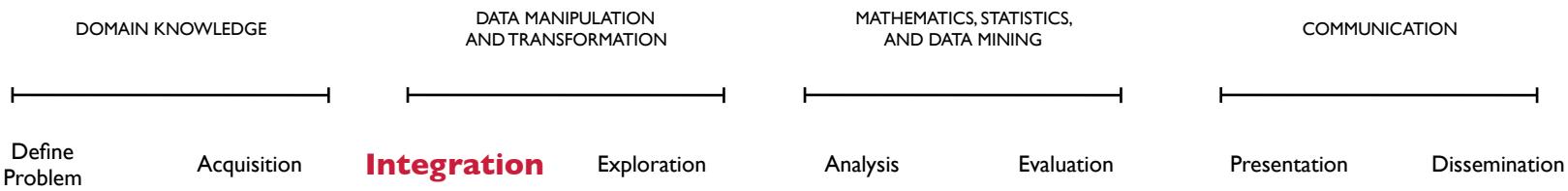
138

estimated nights/year

1.5

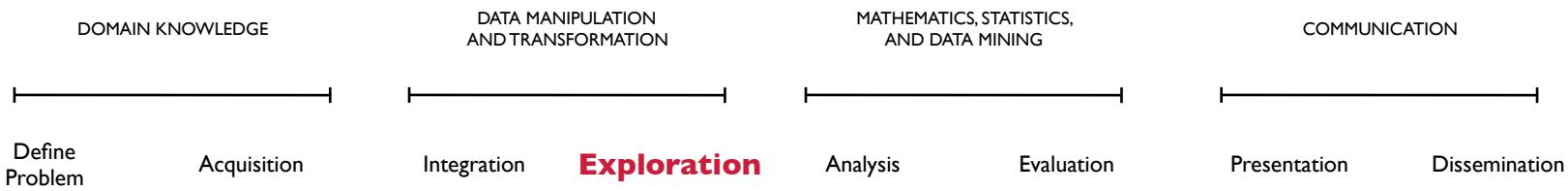
reviews/listing/month

Process



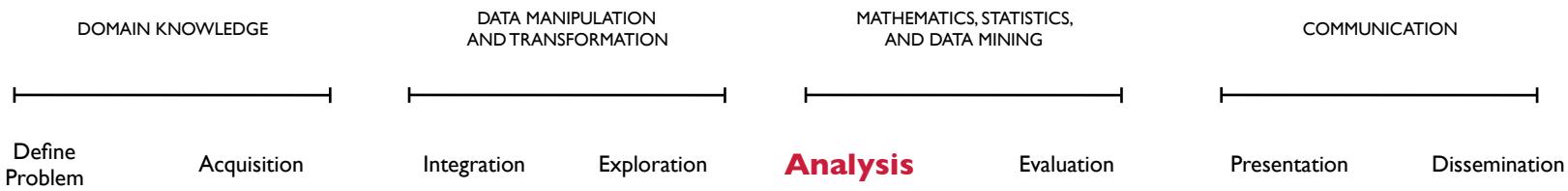
Aid users in their discovery of relevant listings

Process



Aid users in their discovery of relevant listings

Process



Aid users in their discovery of relevant listings



Search for people, jobs, companies, and more...



Advanced



28



5



Home

Profile

Connections

Jobs

Interests

Business Services

Try Premium for free



People You May Know

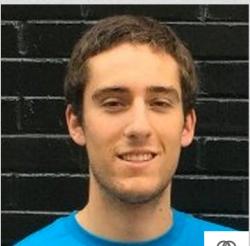


Ethan

Director, IT at Kite Pharma

24

Connect

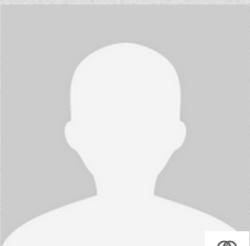


Gideon

Data Scientist

17

Connect



Alana

Strategic Partnerships

9

Connect



johnny Bigwig

owner at TLC tree cutting

1

Connect



Alice

7



Cian

Data Scientist at Stitch Fix

3



Jessica

Instructional Coach at General Assembly

3



Yuriy

Data Scientist

5



Search for people, jobs, companies, and more...



Advanced



28



5



Home

Profile

Connections

Jobs

Interests

Business Services

Try Premium for free



People You May Know



Ethan

Director, IT at Kite Pharma

24

Connect



Gideon

Data Scientist

17

Connect



Alana

Strategic Partnerships

9

Connect



johnny Bigwig

owner at TLC tree cutting

1

Connect



Alice

7



Cian

Data Scientist at Stitch Fix

3



Jessica

Instructional Coach at General Assembly

3



Yuriy

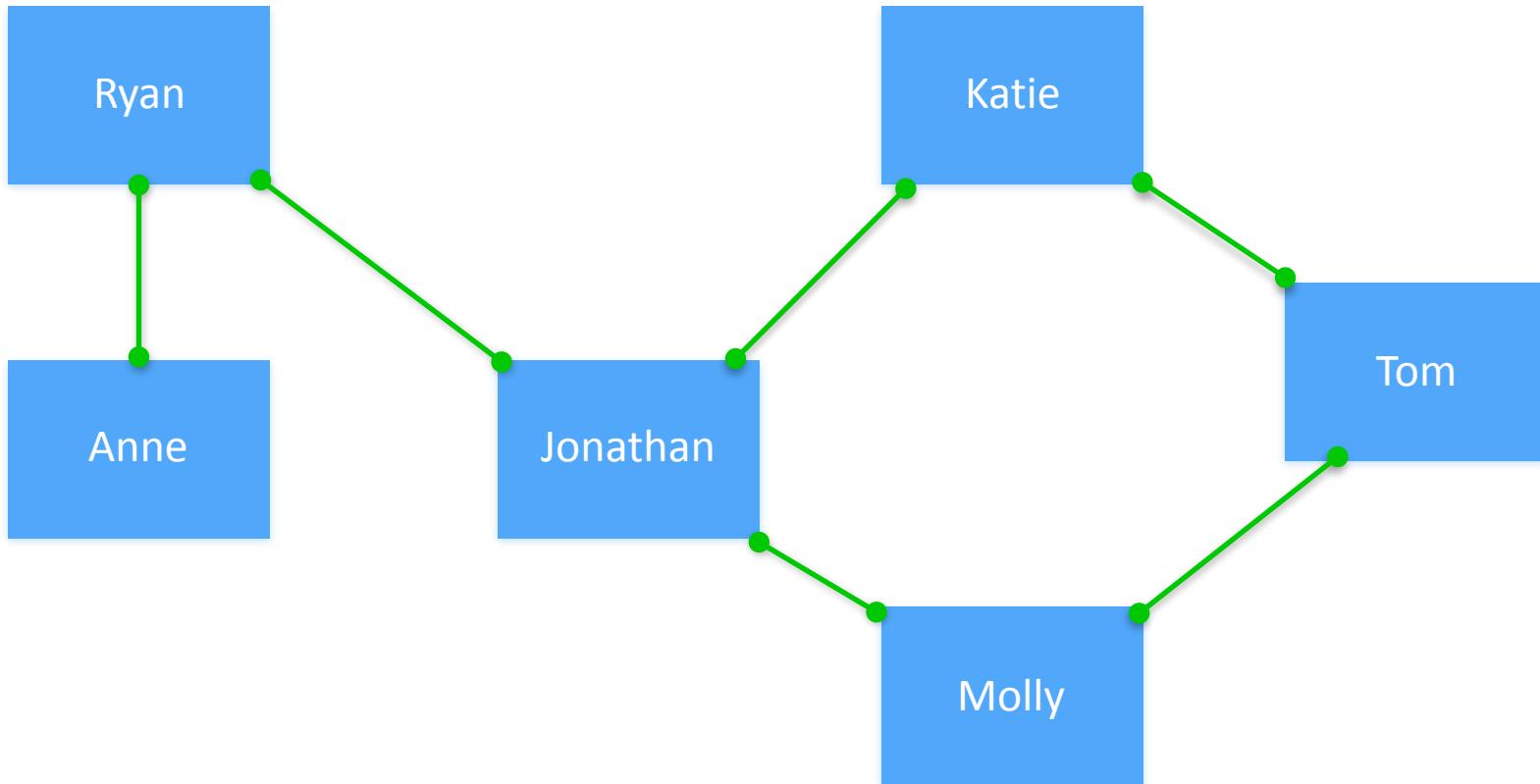
Data Scientist

5

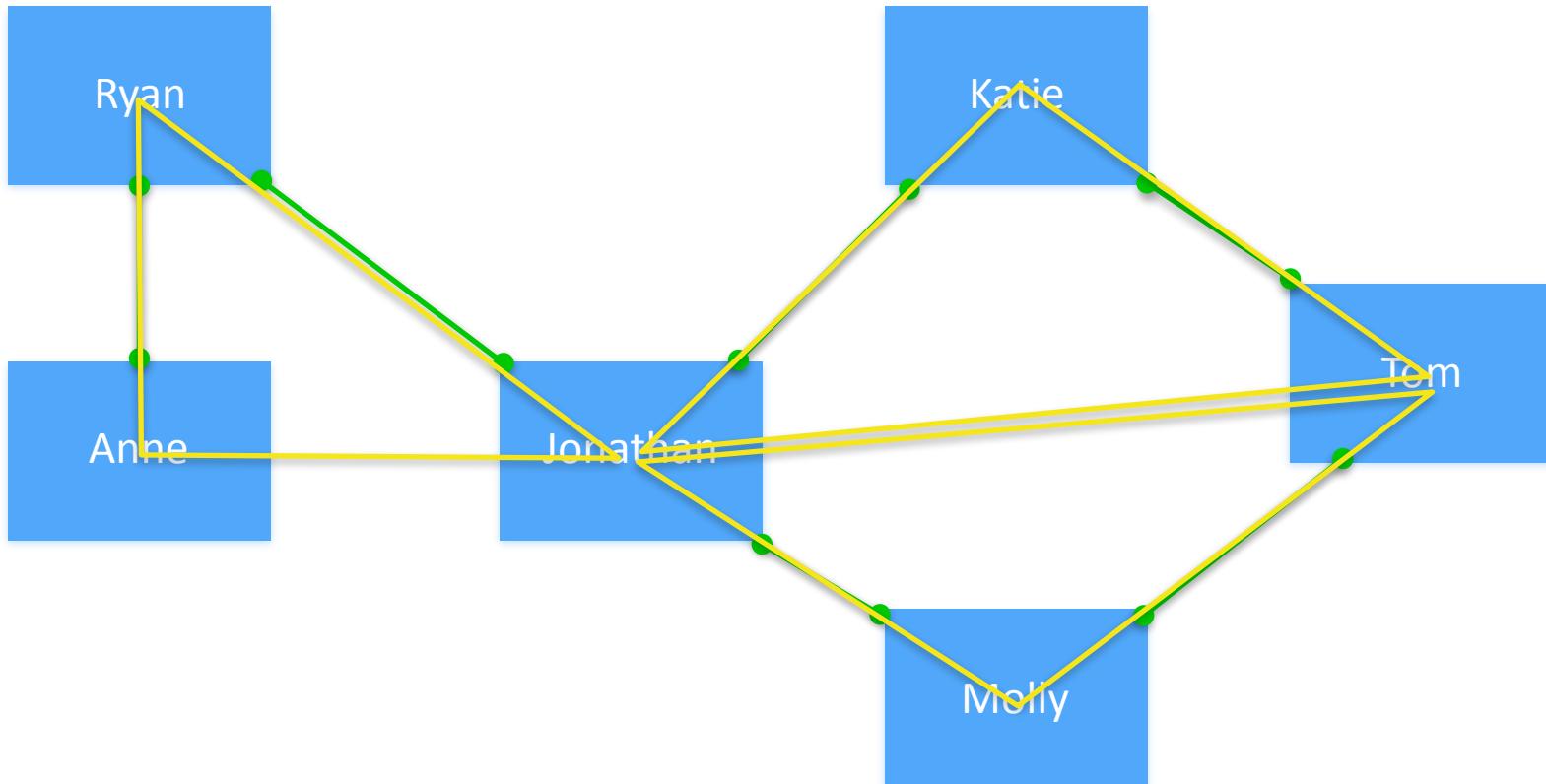
Triangle Closing

A friend of my friends will **likely** be my friend

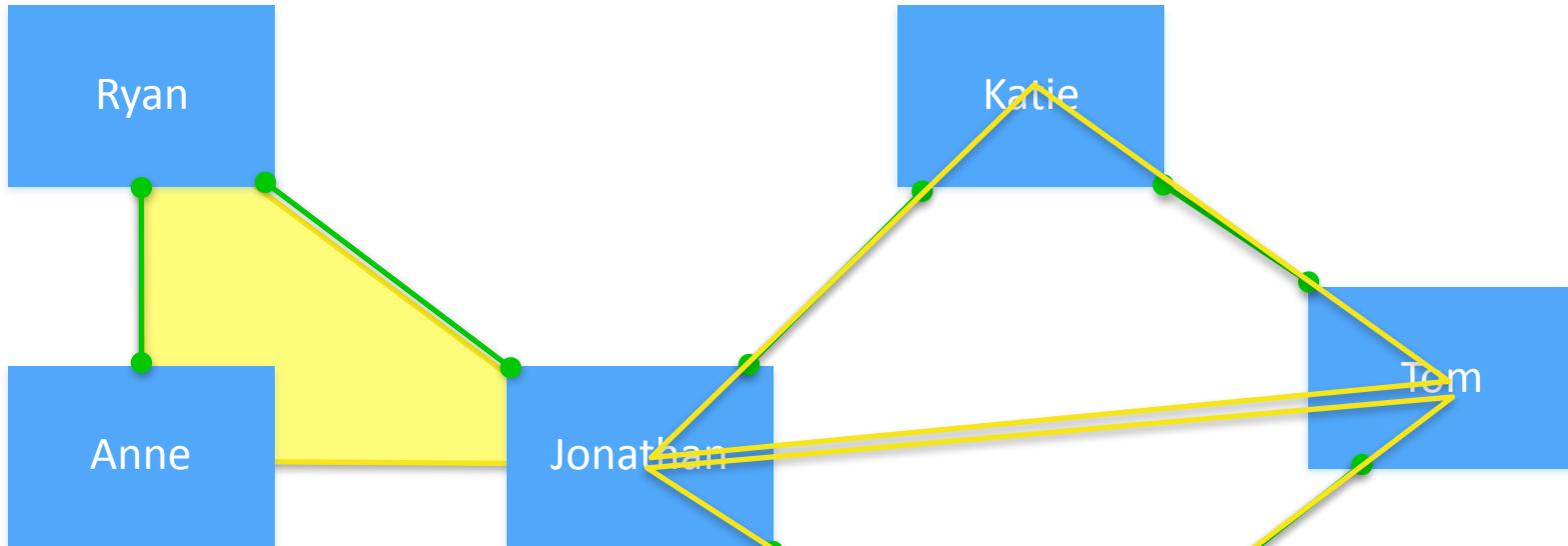
In Pictures



Connect Triangles

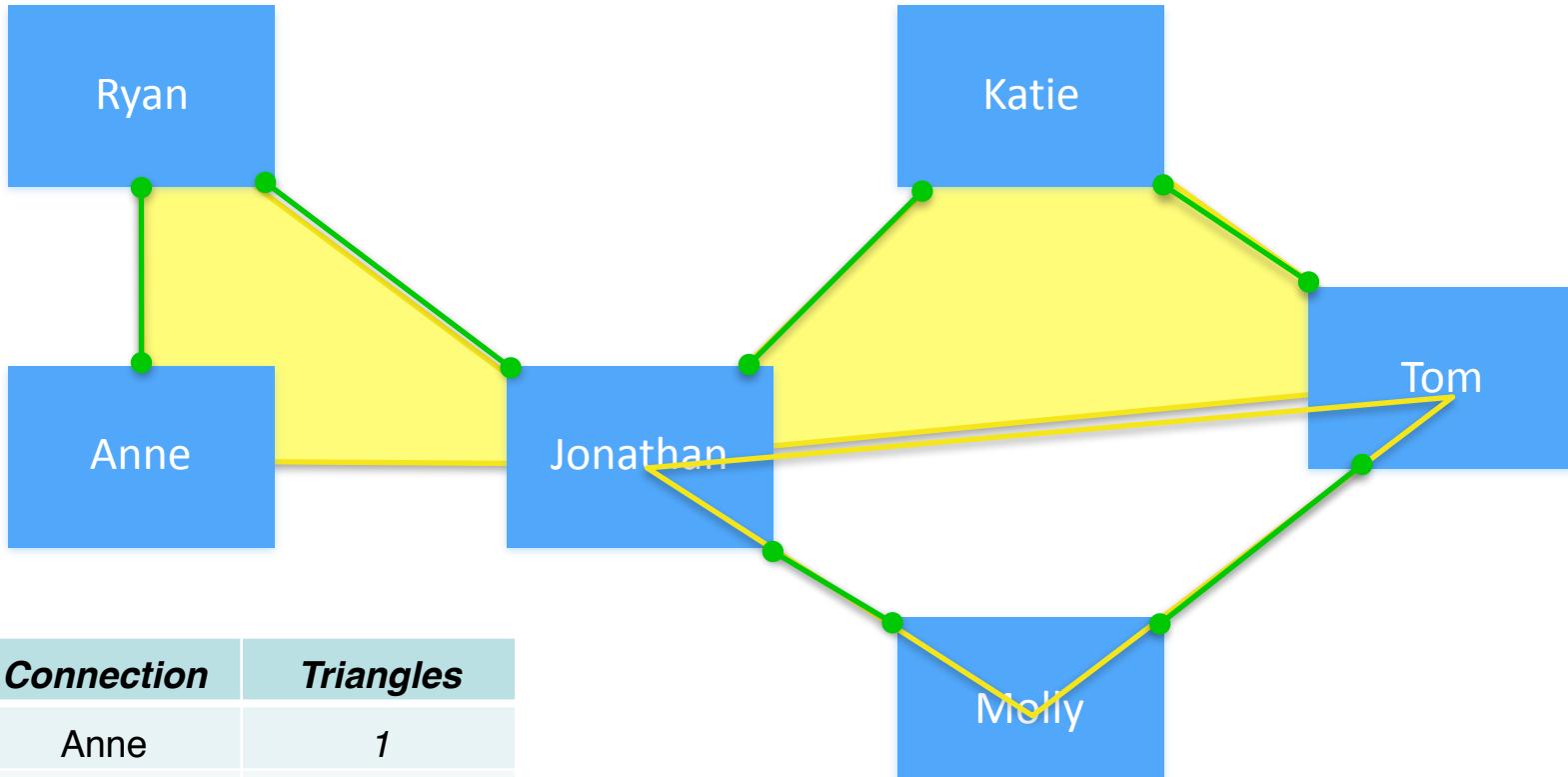


Count Triangles

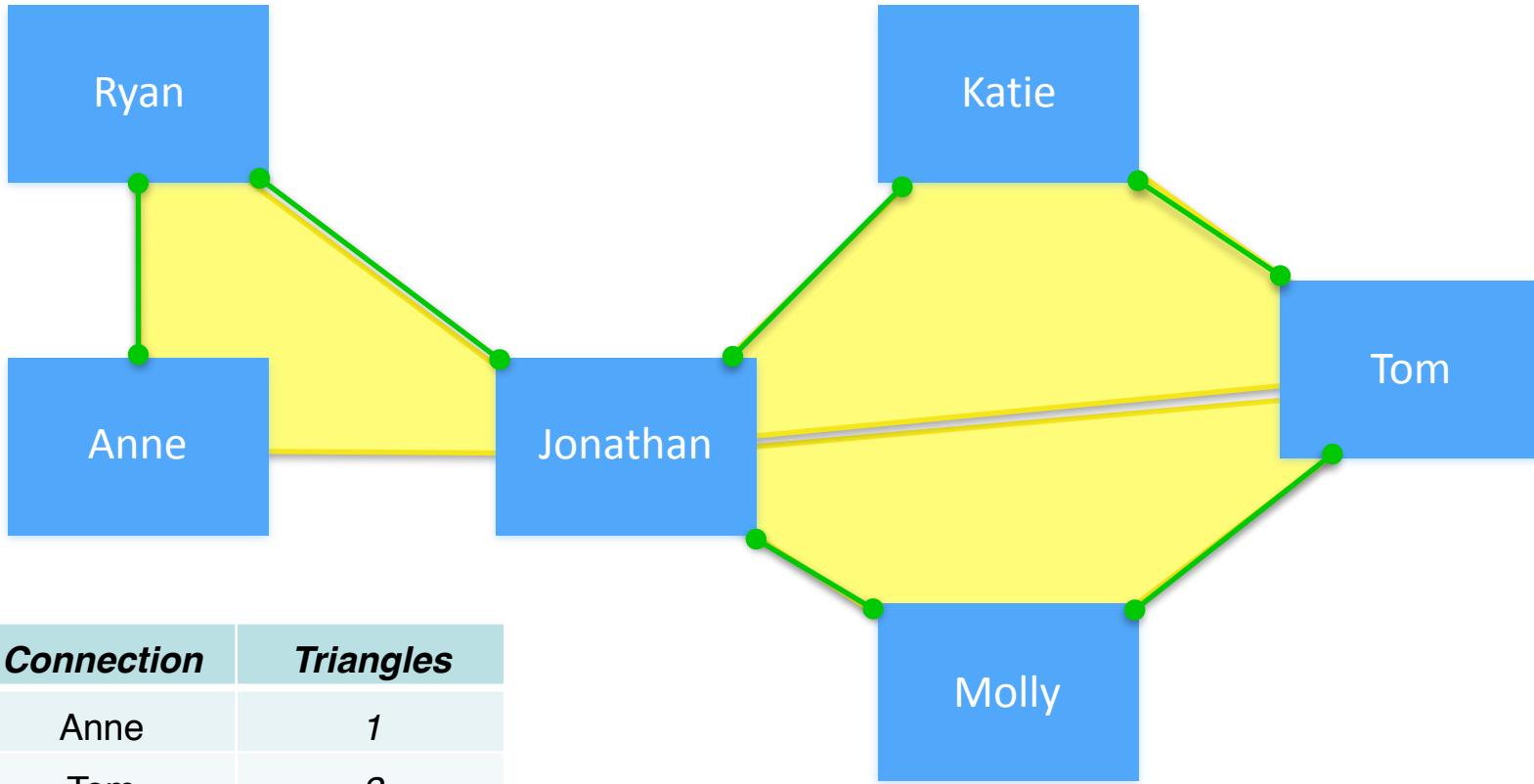


<i>Connection</i>	<i>Triangles</i>
Anne	1
Tom	0

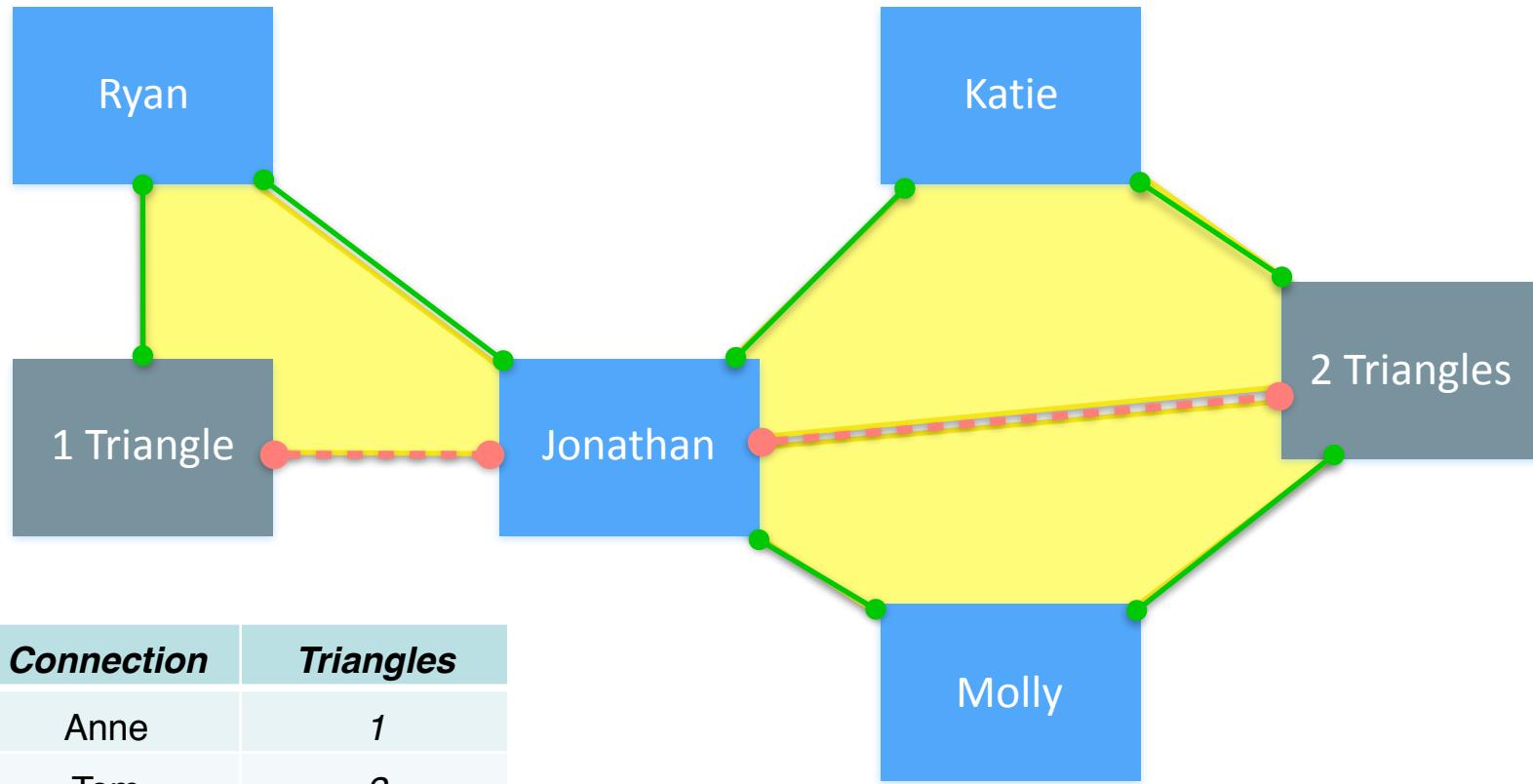
Count Triangles



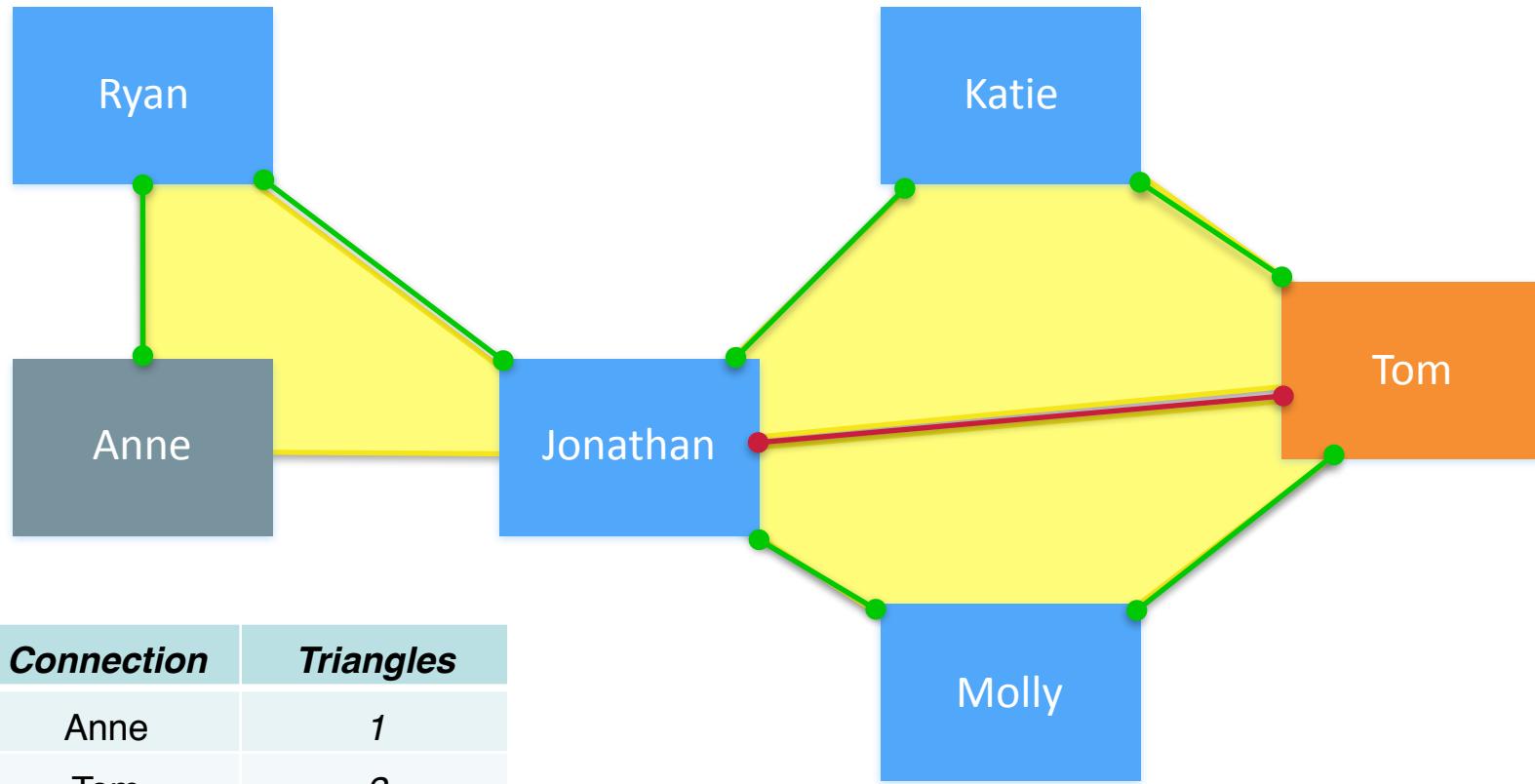
Count Triangles



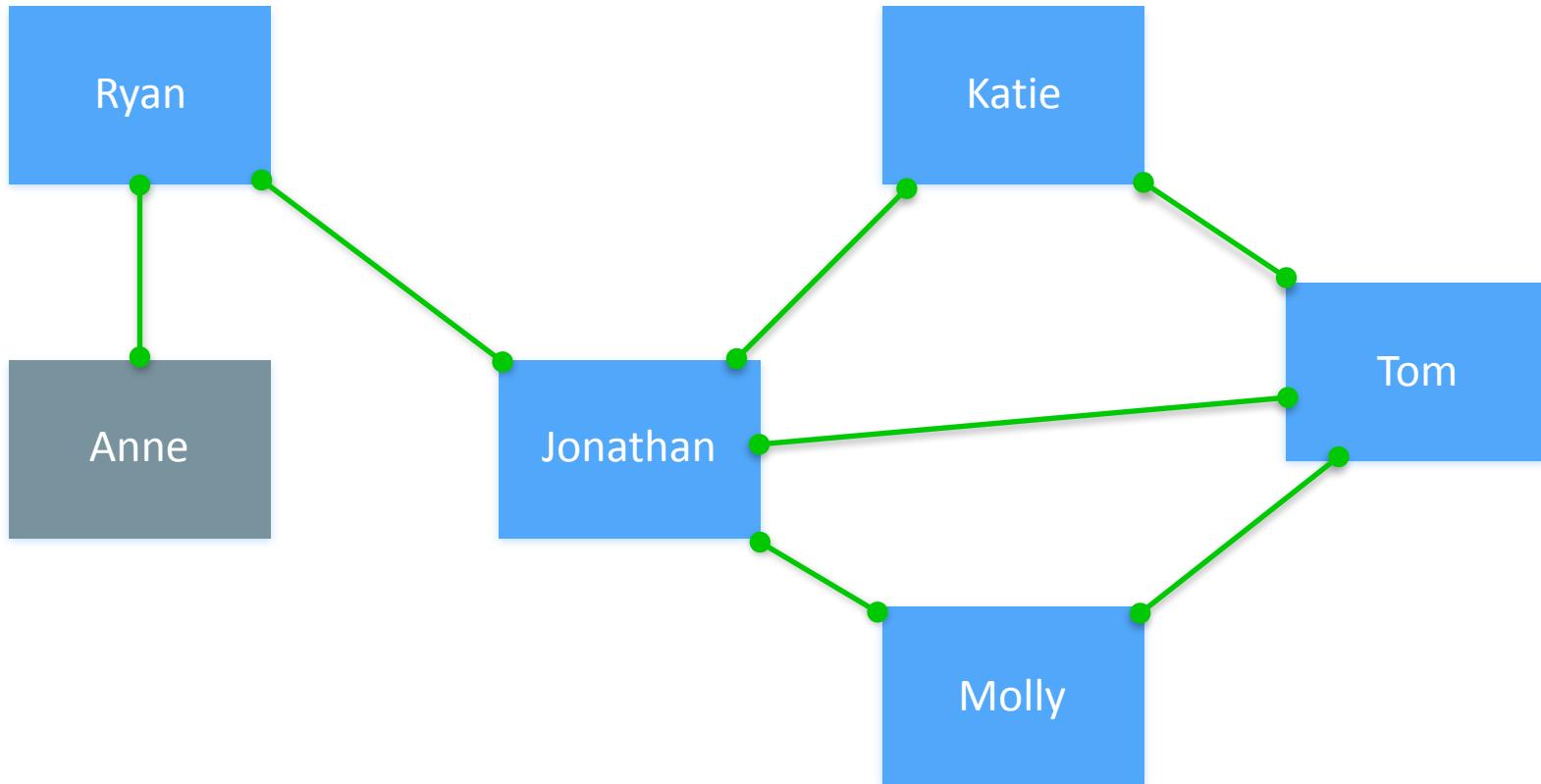
Compare Counts



Top Count as Recommendation



Create Edge



Triangle Closing

A friend of my friends will **likely** be my friend

Triangle Closing

A **listing** of my friends will **likely** be my friend

Triangle Closing

A listing of my bookings will likely be my friend

Triangle Closing

A listing of my bookings will likely be my booking

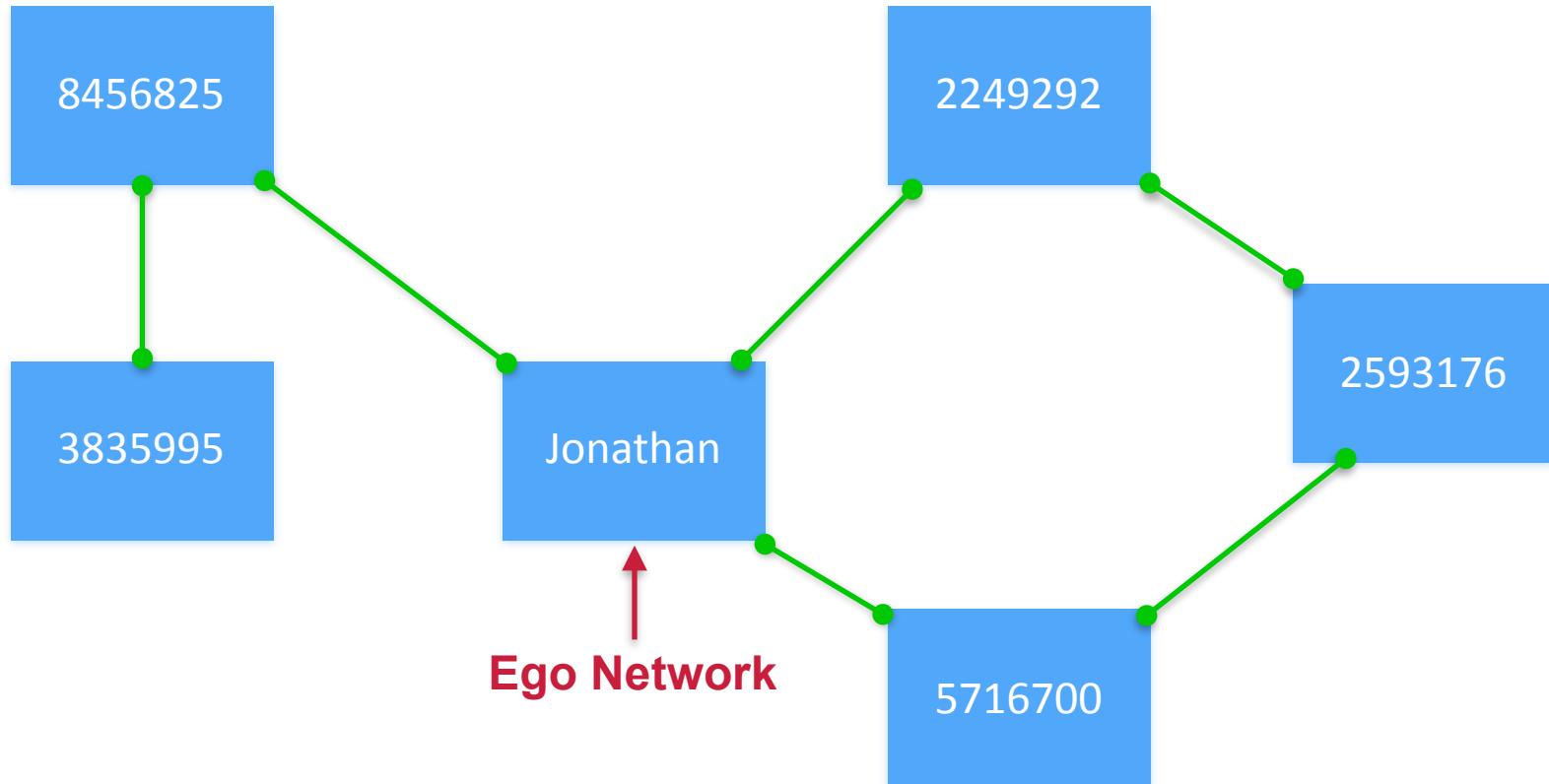
Triangle Closing

A **listing** (with the same guests) as my **bookings** will
likely be my **booking**

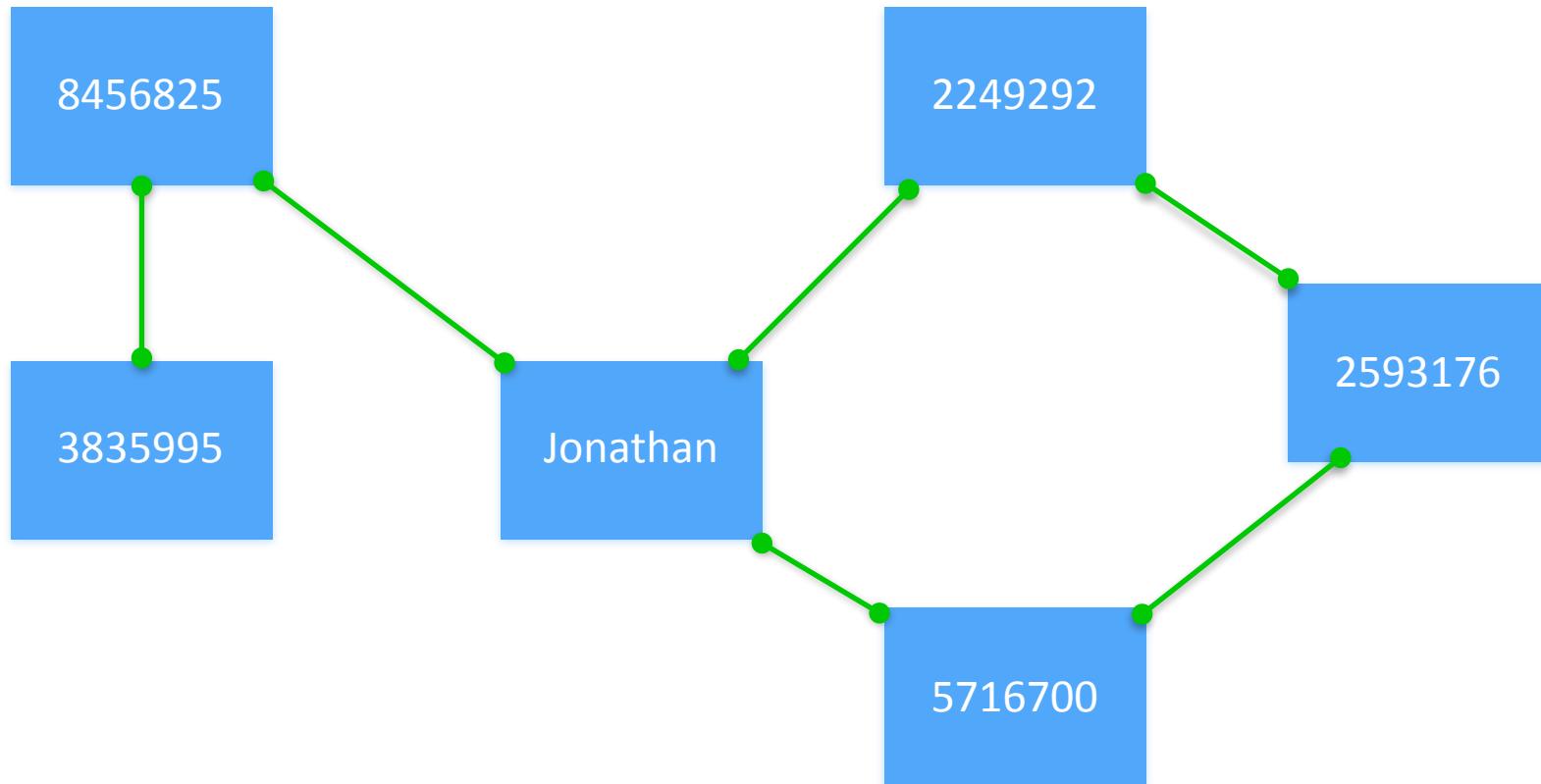
Triangle Closing

A **listing** (with the same guests) as my **bookings** will
likely be my (future) **booking**

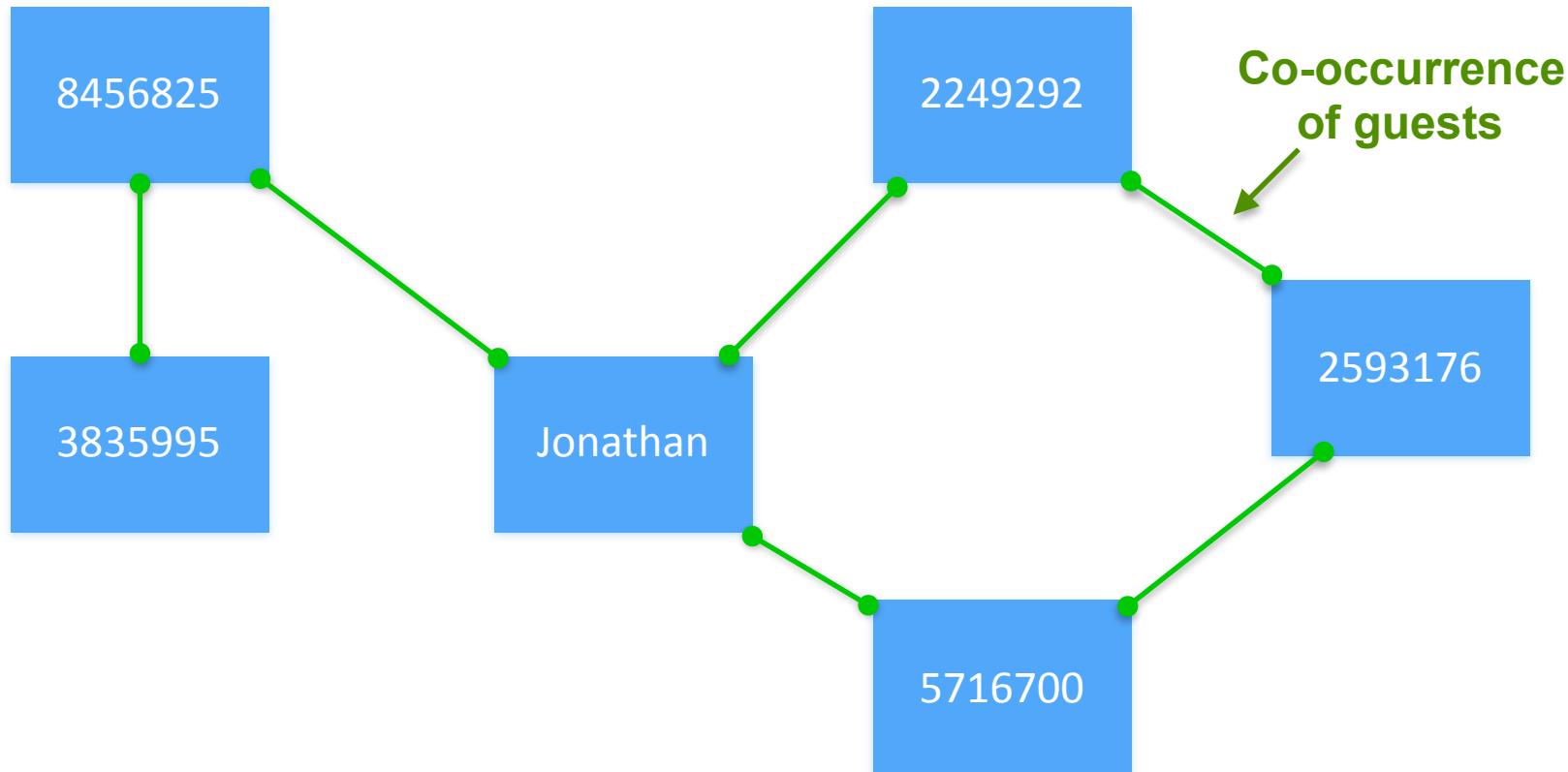
Airbnb Reviews



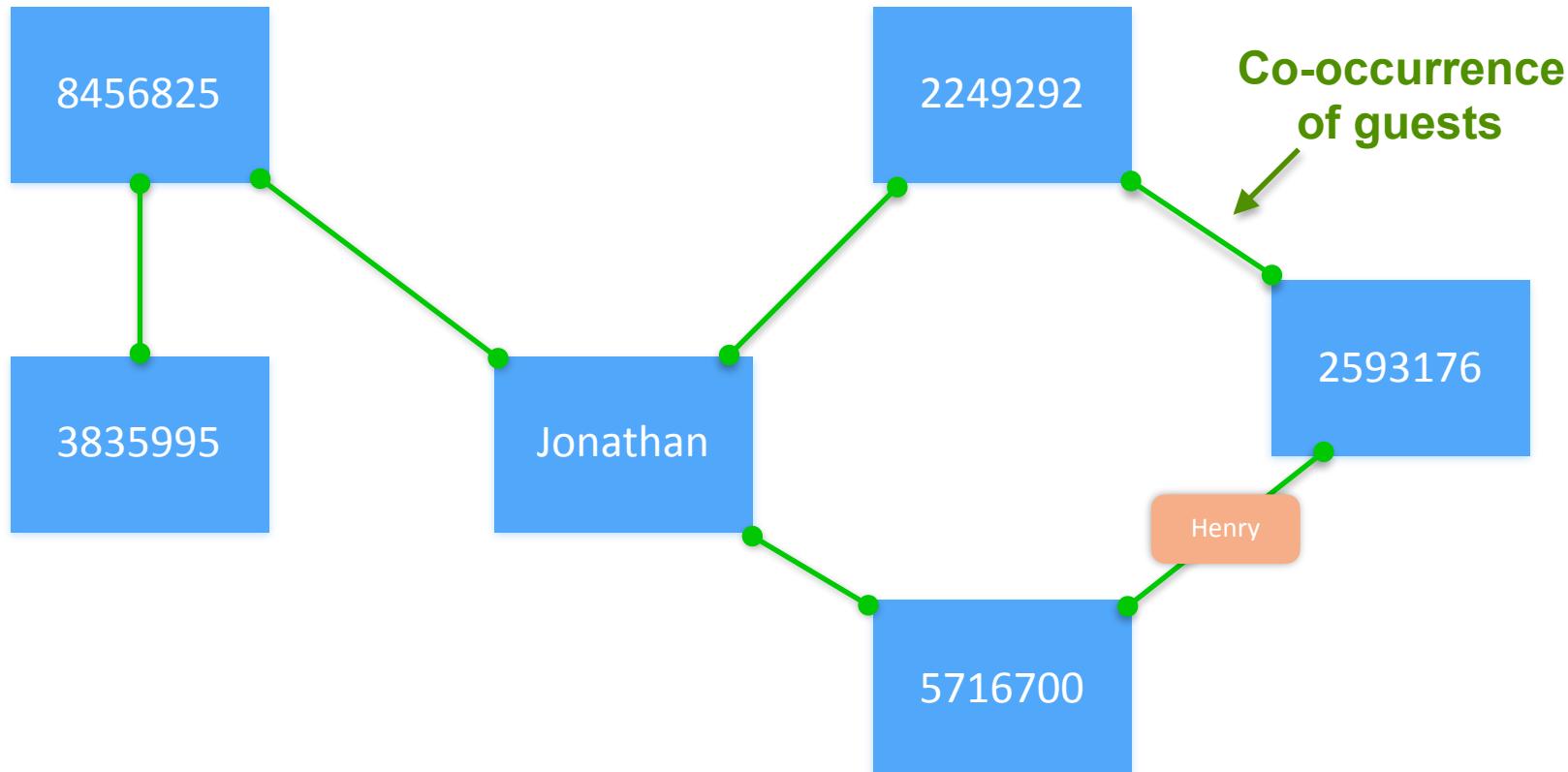
Airbnb Reviews



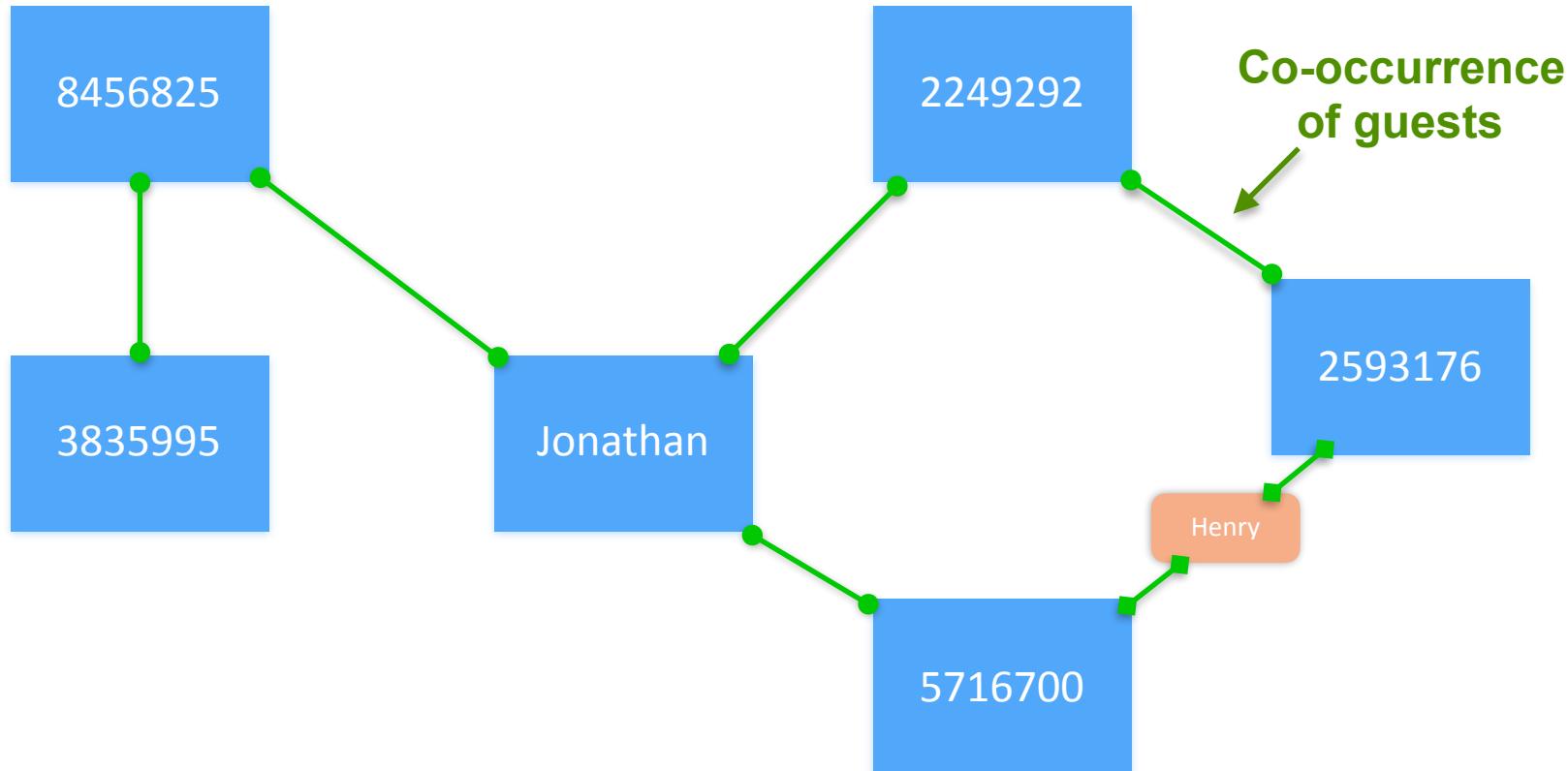
Airbnb Reviews



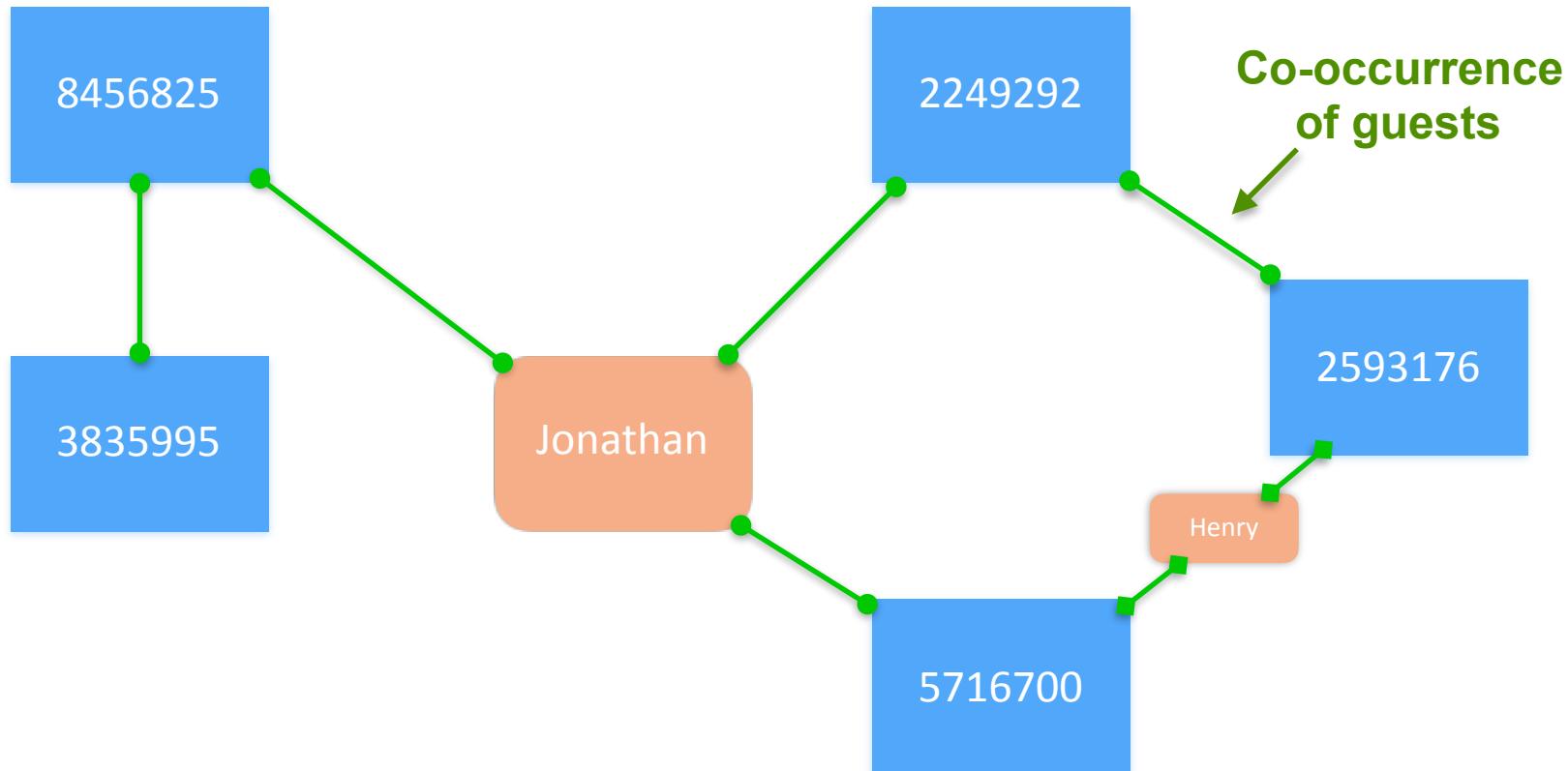
Airbnb Reviews



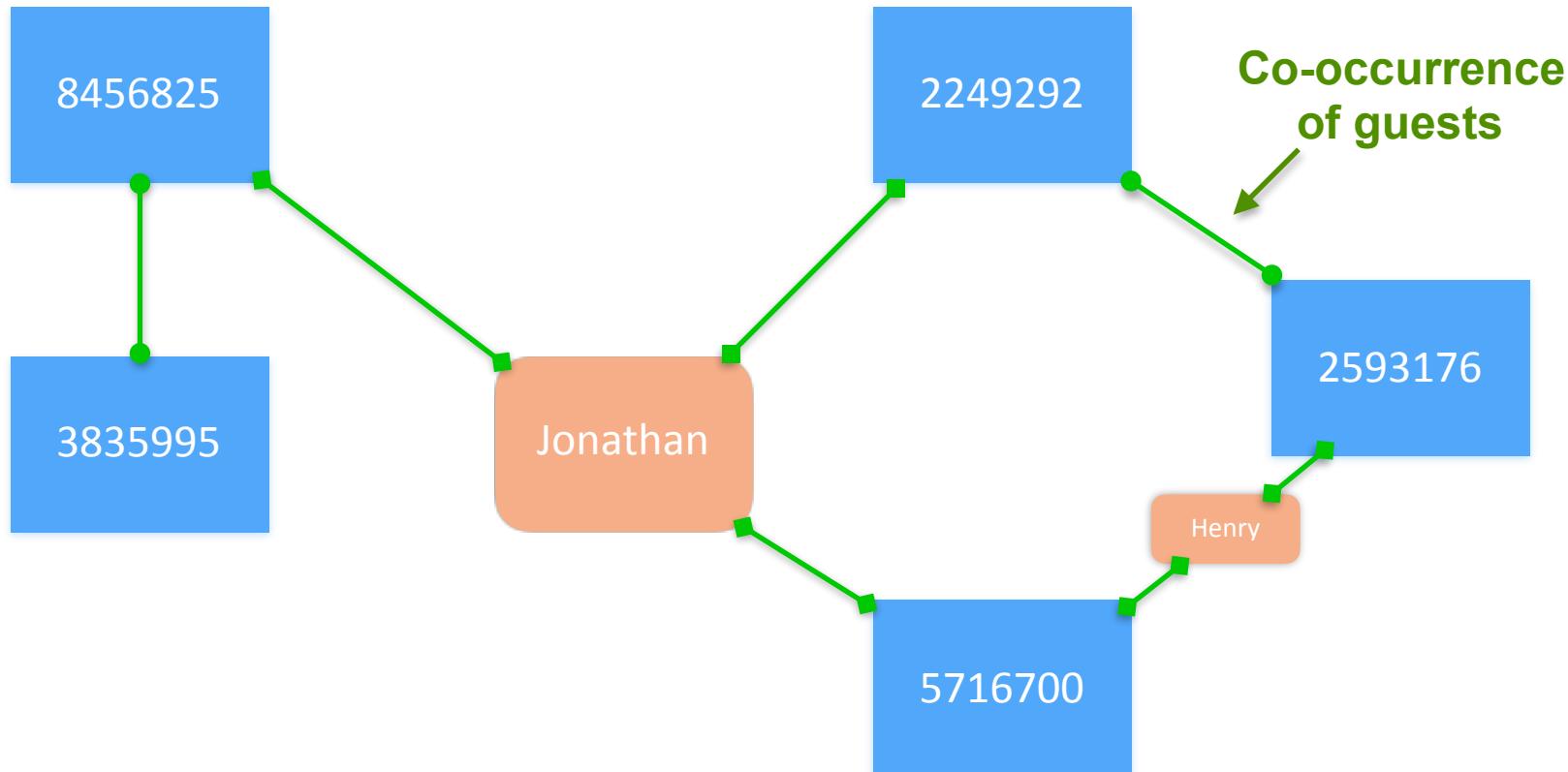
Airbnb Reviews



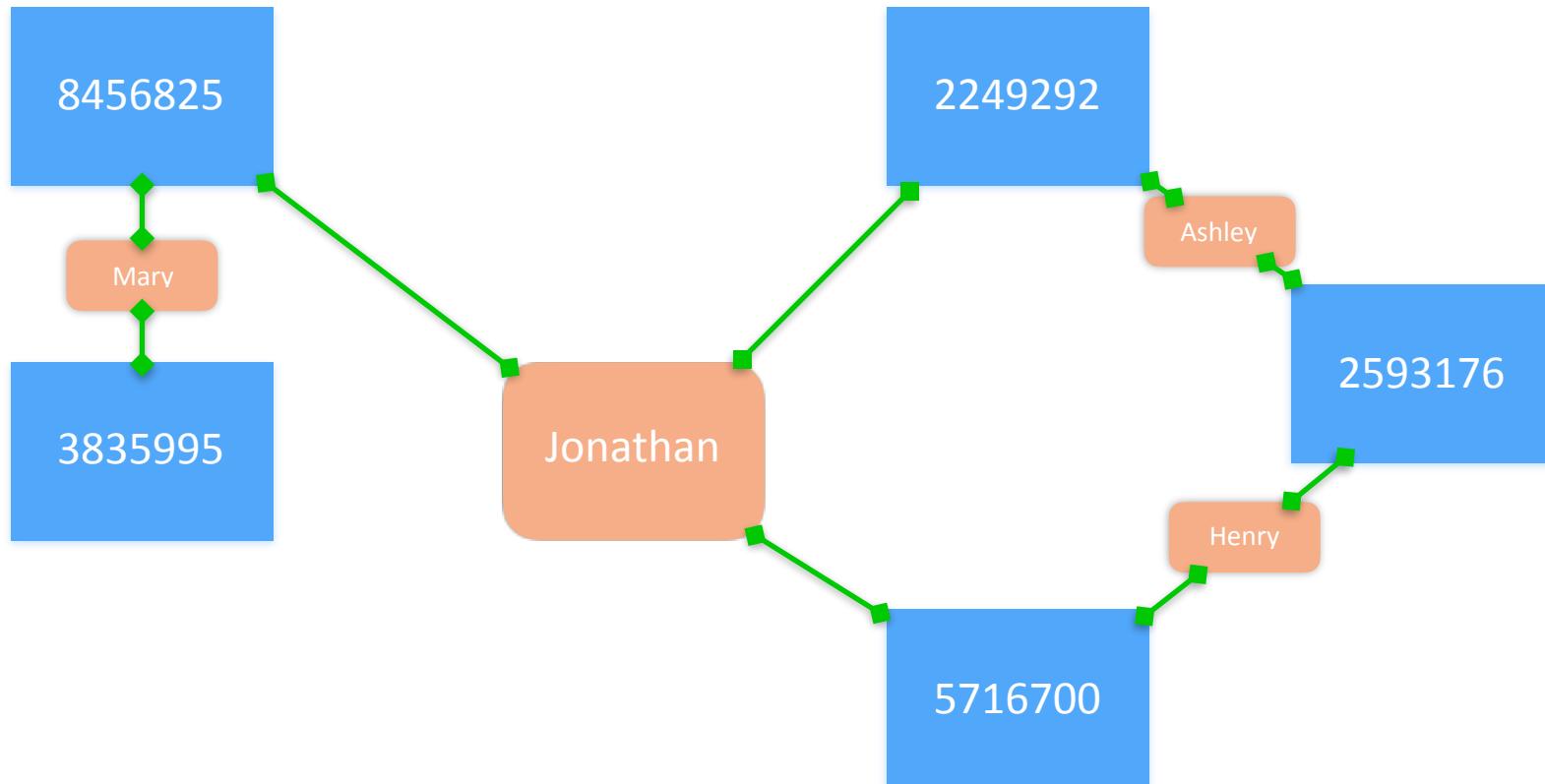
Airbnb Reviews



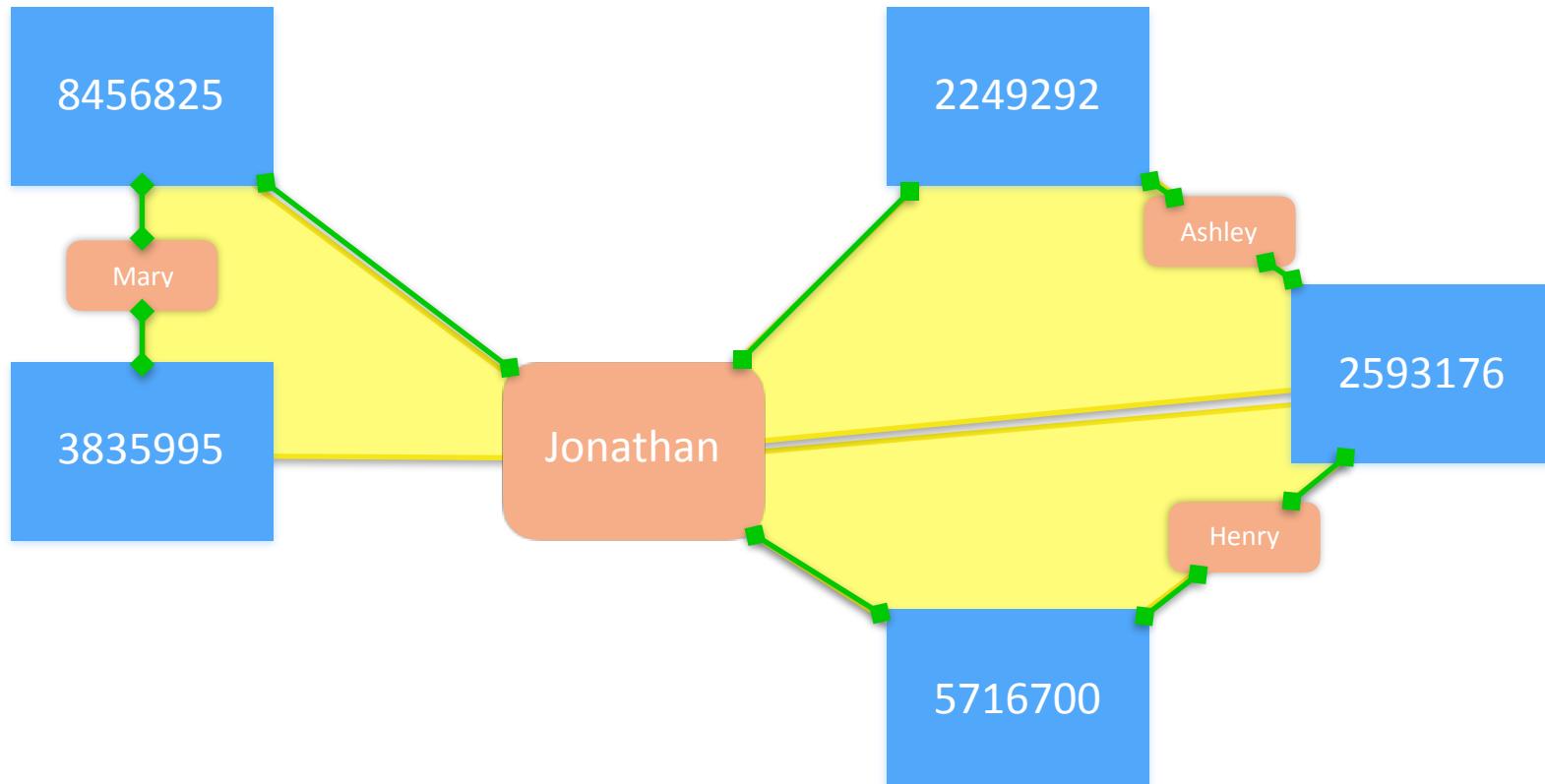
Airbnb Reviews



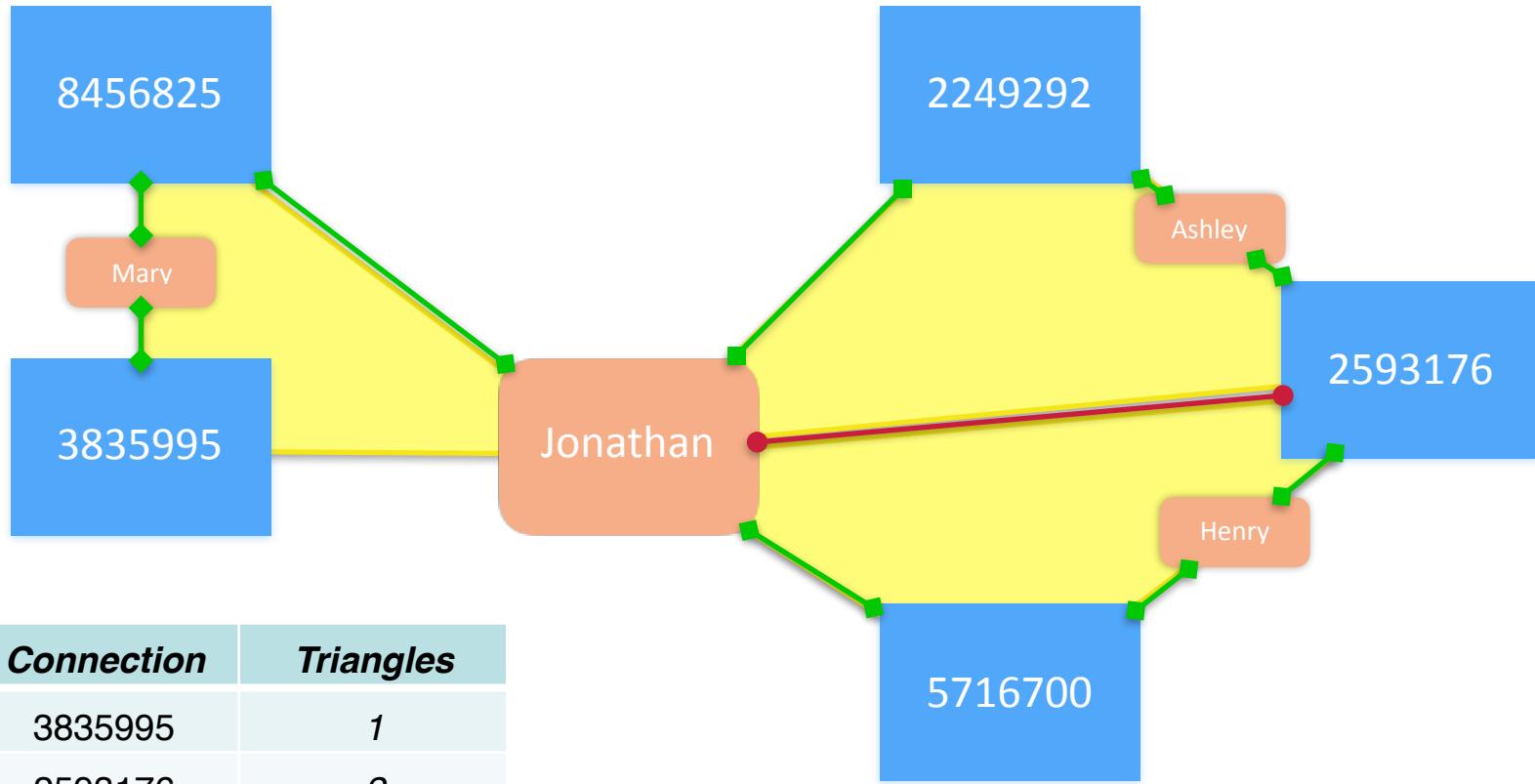
Airbnb Reviews



Airbnb Reviews



Airbnb Reviews



In (pseudo)code

```
def count_triangles(listings, user):
    network = []
    triangles = []

    for listing in listings:
        # for each listing I have stayed at
        if user.visited(listing):
            # find me all users who have also stayed there
            for guest in listing.guests:
                network.add(guest)

    # find other listings that guests in my network have been to
    for listing in listings:
        if listing.isin(network):
            # add listing to triangles
            triangles.append(listing)

    # count the listings
    listing_counts = triangles.counts()

    # return the top 10 most visited listings (to go next)
    return listing_counts.most_common(10)
```

In (pseudo)code

```
def count_triangles(listings, user):
    network = []
    triangles = []

    for listing in listings:
        # for each listing I have stayed at
        if user.visited(listing):
            # find me all users who have also stayed there
            for guest in listing.guests:
                network.add(guest)

    # find other listings that guests in my network have been to
    for listing in listings:
        if listing.isin(network):
            # add listing to triangles
            triangles.append(listing)

    # count the listings
    listing_counts = triangles.counts()

    # return the top 10 most visited listings (to go next)
    return listing_counts.most_common(10)
```

In (pseudo)code

```
def count_triangles(listings, user):
    network = []
    triangles = []

    for listing in listings:
        # for each listing I have stayed at
        if user.visited(listing):
            # find me all users who have also stayed there
            for guest in listing.guests:
                network.add(guest)

    # find other listings that guests in my network have been to
    for listing in listings:
        if listing.isin(network):
            # add listing to triangles
            triangles.append(listing)

    # count the listings
    listing_counts = triangles.counts()

    # return the top 10 most visited listings (to go next)
    return listing_counts.most_common(10)
```

In (pseudo)code

```
def count_triangles(listings, user):
    network = []
    triangles = []

    for listing in listings:
        # for each listing I have stayed at
        if user.visited(listing):
            # find me all users who have also stayed there
            for guest in listing.guests:
                network.add(guest)
```

```
# find other listings that guests in my network have been to
for listing in listings:
    if listing.isin(network):
        # add listing to triangles
        triangles.append(listing)
```

```
# count the listings
listing_counts = triangles.counts()
```

```
# return the top 10 most visited listings (to go next)
return listing_counts.most_common(10)
```

In (pseudo)code

```
def count_triangles(listings, user):
    network = []
    triangles = []

    for listing in listings:
        # for each listing I have stayed at
        if user.visited(listing):
            # find me all users who have also stayed there
            for guest in listing.guests:
                network.add(guest)

    # find other listings that guests in my network have been to
    for listing in listings:
        if listing.isin(network):
            # add listing to triangles
            triangles.append(listing)

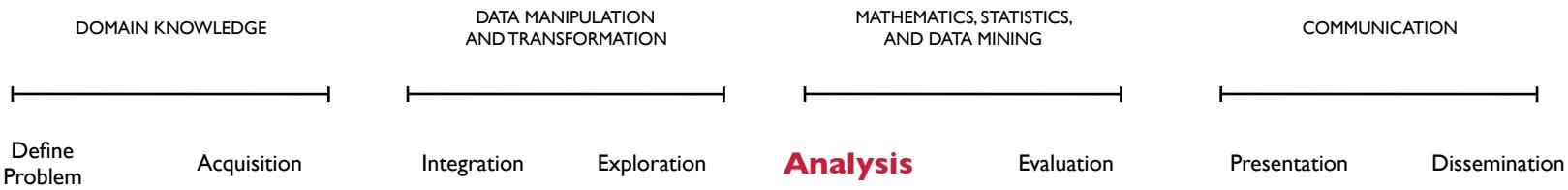
# count the listings
listing_counts = triangles.counts()

# return the top 10 most visited listings (to go next)
return listing_counts.most_common(10)
```

Data Science is...

Data Science is... just counting!

Process

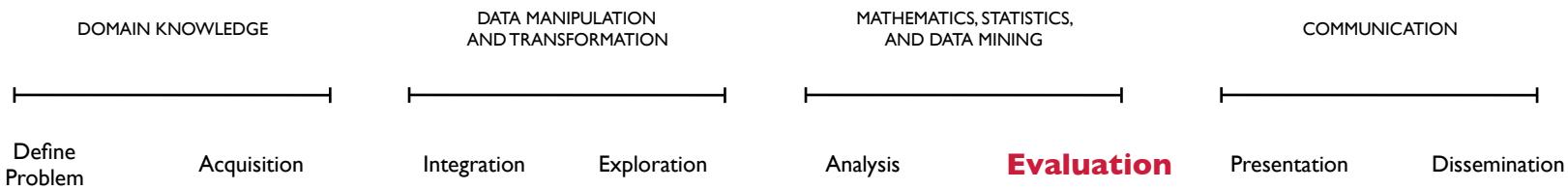


Aid users in their discovery of relevant listings

Triangle Closing

A **listing** (with the same guests) as my **bookings** will
likely be my (future) **booking**

Process



Aid users in their discovery of relevant listings

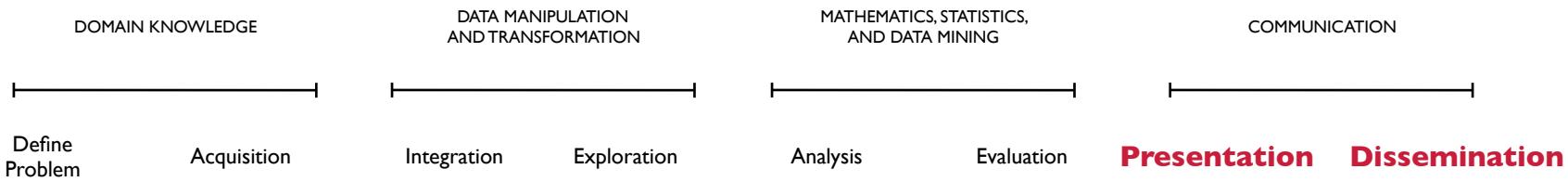
Challenges of Social Recommendation

- Cold Start Problem (i.e. not enough data)
- Data Sparsity
- Scalability
- The Long Tail (i.e. uncommon or unknown items)
- Gray Sheep (i.e. fickle user preferences)

Issues Evaluating Recommendations

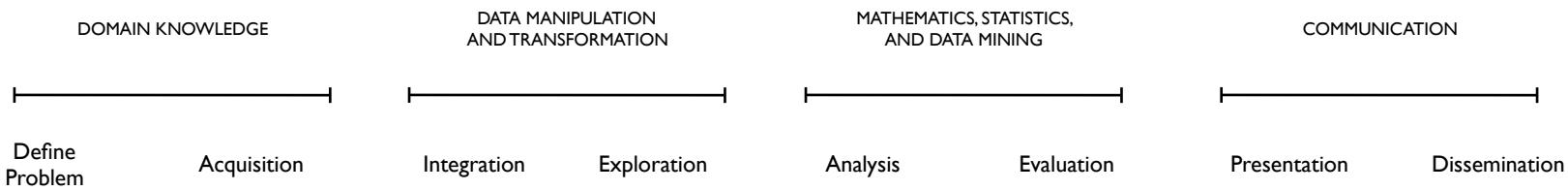
- No Gold Standard
- Temporal Variability (i.e. changing preferences)
- Limited Evaluation data
- Inherent Subjectivity

Process

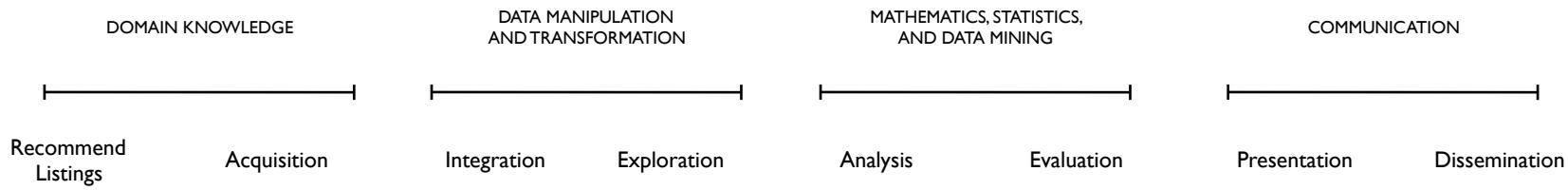


Aid users in their discovery of relevant listings

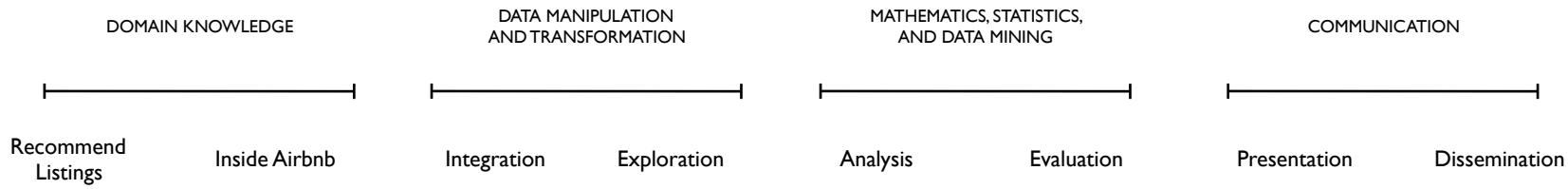
Process



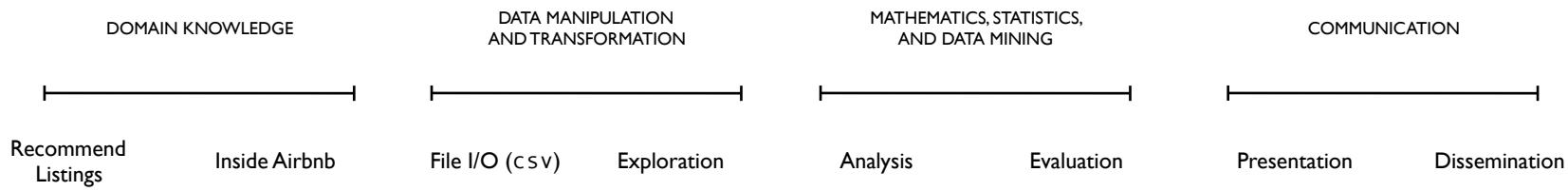
Process



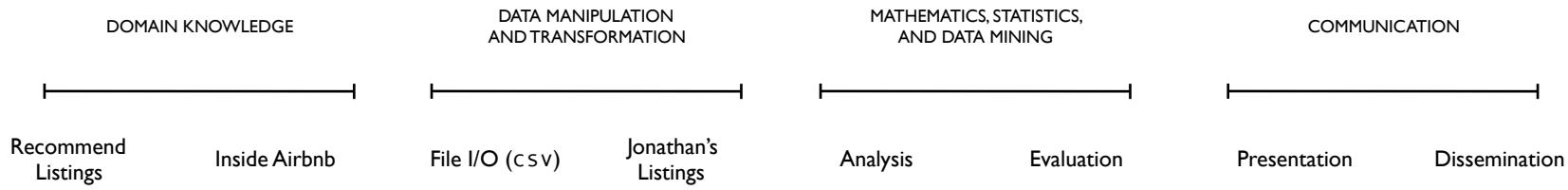
Process



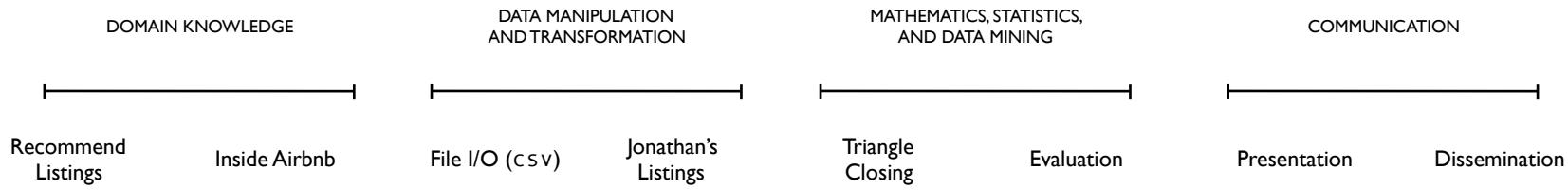
Process



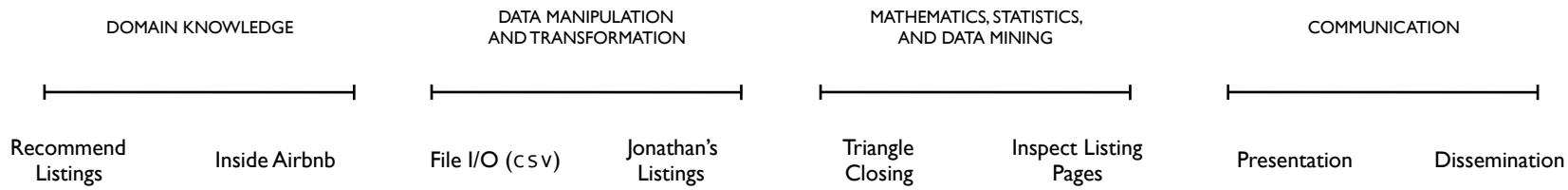
Process



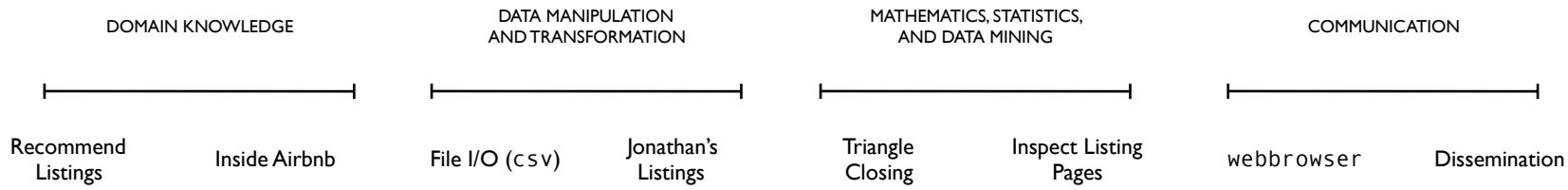
Process



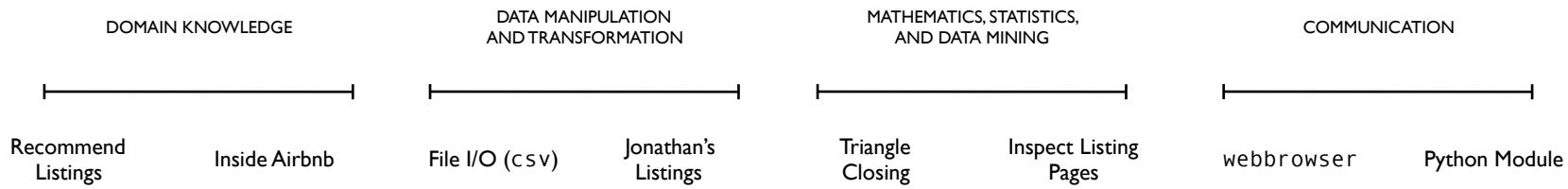
Process



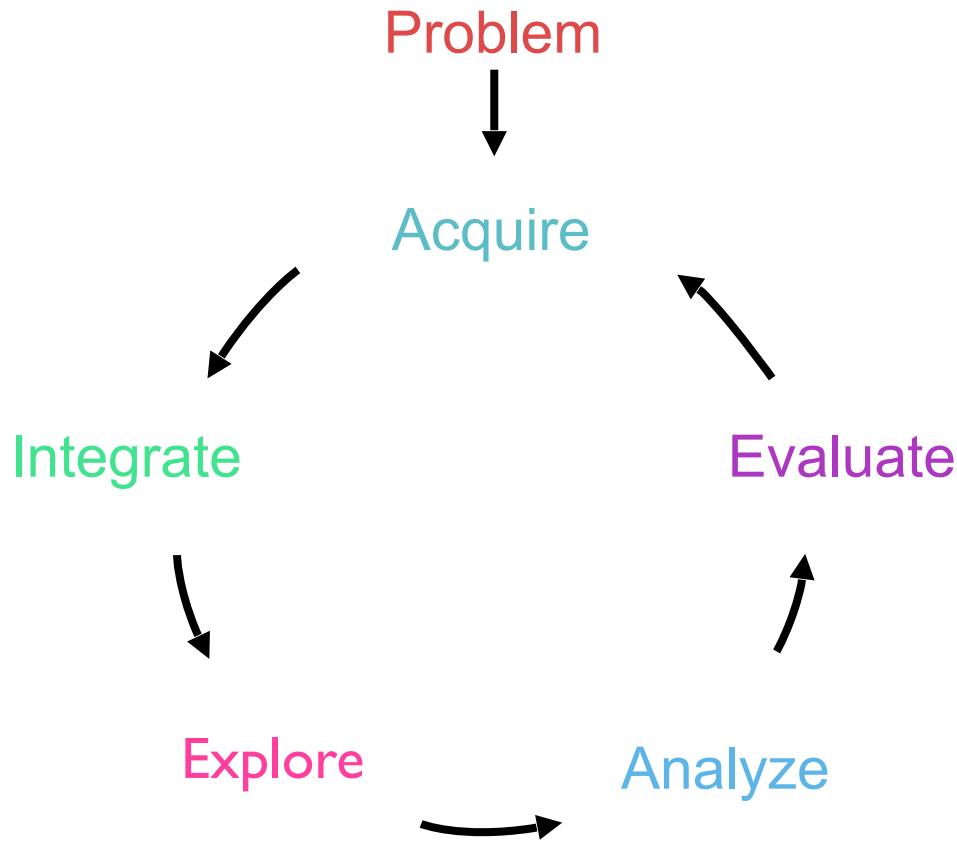
Process



Process



Process + Iteration



Process + Iteration

