Copy content, Copy friends: Understanding and Augmenting the Power of Social Sharing through Pinterest

Nishanth Sastry
King’s College London
Social sharing provides a **alternate, trusted** source of information (and can start revolutions)

Social sharing is **quick**

It took 30 seconds for the August 23rd earthquake to travel from Washington DC to New York.

Twitter Ad: https://www.youtube.com/watch?v=0UFsJhYBxzY
Social sharing is **quick**

It took 30 seconds for the August 23rd earthquake to travel from Washington DC to New York.

Twitter Ad: [https://www.youtube.com/watch?v=0UFsJhYBxzY](https://www.youtube.com/watch?v=0UFsJhYBxzY)
Websites derive significant value from social sharing

- >46% of first-day views (and 25% of all views) to YouTube videos are “social” referrals and embeds (Broxton et al, 2011)

- 31.2% of traffic in Q4 2014 to Shareaholic’s network sites.

Websites derive significant value from social sharing

- >46% of first-day views (and 25% of all views) to YouTube videos are “social” referrals and embeds (Broxton et al, 2011)

- 31.2% of traffic in Q4 2014 to Shareaholic’s network sites.


$80.00 •

AVERAGE ORDER

http://www.shopify.co.uk/infographics/pinterest
Benefit of social: **crowdsourced** content discovery
Pinterest: A good case study

- Content Curation:
  - 50 billion pins (images) collected by people onto more than 1 billion boards (@Pinterest*)

- Social Bootstrapping
  - ~60% (40m/68m) of users connected with Facebook**

* https://twitter.com/Pinterest/status/582960872093556736
** According to our dataset
We appear to be at an inflection point...

We appear to be at an inflection point...

But is this a step forward or backward?
“Searching” Historical Parallels
People have been making lists from the beginning
This gargantuan guide offers reviews and listings of thousands of websites... over 10,000 entries. Reviews include a synopsis of the website, the website address and occasional screen shots of the site. Each section also includes the editors top pick ... included CD-ROM also contains the website's addresses and reviews.

Some of the topics covered in the World Wide Webpages include animals, business, education, gardening, health, music, parenting, relationships, sports and travel....A handy book to have in front of you while you are surfing the Net.
Idea: Make a Website catalog, not a book — No more editions!
Idea: Make a Website catalog, not a book — No more editions!
Idea: Make a Website catalog, not a book — No more editions!
Idea: Make a Website catalog, not a book — No more editions!
... And then, search killed the need for making lists!
So, why is manual recommendation still popular?

**Research Questions**

- Why do users put in the manual effort? [ICWSM12,13]
- What is the value of social recommendations? [WWW14]
  - How do “real” friends compare with “friends” on website?
- Can we automate the manual effort involved? [WWW15]
Outline

• What type of content is curated?
• Why do users curate?

[ICWSM12,13]

• Can we automate content curation?

[WWW15]

• Can social bootstrapping create a good community?

[WWW14]
Outline

• What type of content is curated?
• Why do users curate?

• Can social bootstrapping create a good community? [WWW14]

• Can we automate content curation? [WWW15]
What kind of content are curated?
What kind of content are curated?
What kind of content are curated?

Curation Ranking
1.
2.
3.
4.
5.
6.
7.
8.

Watercolour
Beauty
Pinboard
Pinterest
Image feed
User 1
Images
Hahaha
Hats
User 2
Cute cats & dogs
Kittens & puppies
Animals
Art
Fashion
Humour
What kind of content are curated?

<table>
<thead>
<tr>
<th>Curation Ranking</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>website 1</td>
</tr>
<tr>
<td>2</td>
<td>website 2</td>
</tr>
<tr>
<td>3</td>
<td>website 3</td>
</tr>
<tr>
<td>4</td>
<td>website 4</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
What kind of content are curated?

<table>
<thead>
<tr>
<th>Curation Ranking</th>
<th>Google PageRank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>website 1</td>
</tr>
<tr>
<td>2</td>
<td>website 2</td>
</tr>
<tr>
<td>3</td>
<td>website 3</td>
</tr>
<tr>
<td>4</td>
<td>website 4</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
What kind of content are curated?

<table>
<thead>
<tr>
<th>Curation Ranking</th>
<th>Website</th>
<th>Google PageRank</th>
<th>Alexa Traffic ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>website 1</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>website 2</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>website 3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>website 4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
What kind of content are curated?

<table>
<thead>
<tr>
<th>Curation Ranking</th>
<th>Google PageRank</th>
<th>Alexa Traffic ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>website 1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>website 2</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>website 3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>website 4</td>
<td>5</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
What kind of content are curated?

Globally unpopular but **niche interest** content is highlighted. “Curation comes up when search stops working”

- Clay Shirky
Niche interest content is important to many
Niche interest content is important to many
Niche interest content is important to many
Niche interest content is important to many
Tail likes vs. head likes

- Tail likes are geographically more diverse
  - Likers may not have offline context
  - No common traditional media either (given different countries)
- Viral propagation more common
Tail likes vs. head likes

- Tail likes are geographically more diverse
  - Likers may not have offline context
  - No common traditional media either (given different countries)
- Viral propagation more common

Social network support important for the tail!
Why do people take the effort to curate manually?

Online survey: 33 Pinterest users (270 Last.fm users)

- 85% for personal scrapbooking
- 48% for displaying content to others
Why do people take the effort to curate manually?

Online survey: 33 Pinterest users (270 Last.fm users)

- 85% for personal scrapbooking
- 48% for displaying content to others

Many curate for personal/private reasons! (Not necessarily social)
All the same, there may be social side-effects
Manual social recommendation has pros and cons

• Niche but interesting content highlighted.

• Content are well organised and personalised for each user.

• It is a manual process.

• People curate for personal reasons, so is it useful to others?? (mixed picture)
Outline

- What type of content is curated?
- Why do users curate?
- Can social bootstrapping create a good community?
- Can we automate content curation?
Example
Social Curation on Pinterest

https://www.pinterest.com/pin/287386019946917492/
Pinterest Curation Process

Pinterest Image feed

Images

Pinboard

Cute cats & dogs  Watercolour  Beauty

User 1

Kittens & puppies  Hahaha  Hats

User 2
Pinterest Curation Process

Pinboards are not comparable!
Poll  How will you categorise this image?

• Animals
• Art
• Fashion
• Film, Music & Books
• Geek
• Science & Nature
Observation

When does majority category emerge?
When does majority category emerge?

Observation

![Graph showing the percent of images over repin steps, indicating an increase with more repin steps.](image-url)
Observation

When does majority category emerge?

Majority category appears before 5th step for >90% images.
Observation

Category vs. Pinboard

1vs1  1vs2  1vs3

Category
Pinboards
Observation

Category vs. Pinboard

1vs1  
1vs2  
1vs3

Category  
Pinboards
Category vs. Pinboard

Observation

1vs1

1vs2

1vs3

Category

Pinboards
Observation

Category vs. Pinboard

1vs1 1vs2 1vs3

Category
Pinboards

1vs1 1vs2

0 0.25 0.5 0.75 1
Observation

Category vs. Pinboard

1vs1

1vs2

1vs3

Category

Pinboards
Observation Category vs. Pinboard

Category -> Pinboard

1vs1 1vs2 1vs3

Category  Pinboards
Evaluation

Prediction cascade

User

Image/Pin

Task 1: Attention Prediction

Repin vs. Noaction 1 of 32 category

Task 2: Category Prediction

Personalisation

Task 3: Pinboard

Random Forest Classifier
We look at images pinned to Pinterest in January, 2013 and obtained more than 5 repins.

Features

- Profiles
e.g. Activity count, follower count

- Category preference
e.g. I like fashion not technology.

- Object preference (based on deep object detection)
e.g. I like dog not cat.

- Objects recognised by deep learning
  e.g. It is a cat.

- Deep Learning Features
  i.e. Features from the layer right before the final

- Image Quality
  e.g. Contrast, Sharpness, Simplicity

Image/Pin (5110)

User (1038)

Crowd (5)

Crowd features
i.e. The majority category among first 5 repiners.
(when majority category of 90% of images appears)
Task #1

Attention Prediction

Feature importance

- Object prefs.
- Deep learning
- Objects
- Category prefs. (U)

Task 1: Attention Prediction
Task 2: Category Prediction
Task 3: Pinboard

Repin vs. Noaction
1 of 32 category

Without

Changtao Zhong
Microsoft Research Asia, 10 Feb 2015
Task #1

Attention Prediction

- Object prefs.
- Deep learning
- Objects
- Category prefs. (U)

Feature importance

Without

0 0.1 0.2 0.3 0.4
Task #1

Attention Prediction

<table>
<thead>
<tr>
<th></th>
<th>Without object prefs.</th>
<th>With object prefs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.66</td>
<td>0.675</td>
</tr>
<tr>
<td>Precision</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Recall</td>
<td>0.64</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Feature importance

- Object prefs.
- Deep learning
- Objects
- Category prefs. (U)

User

Image/Pin

Task 1: Attention Prediction

Repin vs. Noaction

1 of 32 category

Task 3

Pinboard

Personalisation

End-to-End Prediction

Changtao Zhong

Microsoft Research Asia, 10 Feb 2015
### Task #1: Attention Prediction

<table>
<thead>
<tr>
<th>Metric</th>
<th>Without Object Prefs</th>
<th>With Object Prefs</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.66</td>
<td>0.77</td>
<td>+0.11</td>
</tr>
<tr>
<td>Precision</td>
<td>0.7</td>
<td>0.83</td>
<td>+0.13</td>
</tr>
<tr>
<td>Recall</td>
<td>0.64</td>
<td>0.69</td>
<td>+0.05</td>
</tr>
</tbody>
</table>

**Feature Importance**

- Object Prefs: Deep learning = 0.3, Objects = 0.2, Category Prefs (U) = 0.1

---

**Diagram:**
- **User** (Pinboard)
- **Task 1:** Attention Prediction
- **Task 2:** Category Prediction
- **Task 3:** Personalisation

**Pinboard:**
- **Task 1:** Repin vs. Noaction
- **1 of 32 category**

**End-to-End Prediction**

- Changtao Zhong | Microsoft Research Asia, 10 Feb 2015
Task #1

Attention Prediction

Accuracy

Precision

Recall

Without object prefs.

With object prefs.

0.66 + 0.11 = 0.77

0.7 + 0.13 = 0.83

0.64 + 0.05 = 0.69

Feature importance

Object prefs.

Deep learning

Objects

Category prefs. (U)

With

0.0

0.1

0.2

0.3

0.4
Task #2

Category Prediction

User

Image/Pin

Attention Prediction

Task 1

Task 2

Category Prediction

Repin vs. Noaction

1 of 32 category

Personalisation

Task 3

Pinboard
Task #2

Category Prediction

Task 1
Attention Prediction

Task 2
Category Prediction

Task 3
Pinboard

Personalisation

User
Image/Pin

Repin vs. Noaction
1 of 32 category
**Task #2**

### Category Prediction

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0</td>
</tr>
<tr>
<td>User</td>
<td>0.225</td>
</tr>
<tr>
<td>Image + User</td>
<td>0.45</td>
</tr>
<tr>
<td>Crowd + Image + User</td>
<td>0.675</td>
</tr>
<tr>
<td>Crowd + User</td>
<td>0.9</td>
</tr>
</tbody>
</table>

---

**Additional Notes**

- 1 of 32 category
- Repin vs. Noaction
- Personalisation
- Pinboard

Changtao Zhong
Microsoft Research Asia, 10 Feb 2015
Task #2

Category Prediction

<table>
<thead>
<tr>
<th>Category</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.19</td>
</tr>
<tr>
<td>User</td>
<td></td>
</tr>
<tr>
<td>Image+User</td>
<td>0.225</td>
</tr>
<tr>
<td>Crowd+Image+User</td>
<td>0.45</td>
</tr>
<tr>
<td>Crowd+User</td>
<td></td>
</tr>
</tbody>
</table>

End-to-End Prediction

Changtao Zhong
Microsoft Research Asia, 10 Feb 2015

Task 1
Attention Prediction

Task 2
Category Prediction

Task 3
Pinboard

Personalisation

Repin vs. Noaction
1 of 32 category
Task #2

Category Prediction

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.19</td>
</tr>
<tr>
<td>Image + User</td>
<td>0.225</td>
</tr>
<tr>
<td>Crowd + Image + User</td>
<td>0.45</td>
</tr>
<tr>
<td>Crowd + User</td>
<td>0.675</td>
</tr>
<tr>
<td>User category preference</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Changtao Zhong
Microsoft Research Asia, 10 Feb 2015
Category Prediction

Task #2

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.19</td>
</tr>
<tr>
<td>User</td>
<td>0.42</td>
</tr>
<tr>
<td>Image+User</td>
<td>0.225</td>
</tr>
<tr>
<td>Crowd+Image+User</td>
<td>0.45</td>
</tr>
<tr>
<td>Crowd+User</td>
<td>0.675</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
</tr>
</tbody>
</table>

User category preference

Accuracy

Changtao Zhong
Microsoft Research Asia, 10 Feb 2015

End-to-End Prediction

Attention Prediction

Pinboard

Repin vs. Noaction

1 of 32 category

Personalisation
Task #2

Category Prediction

- Random: 0.19
- User: 0.42
- Image + User: 0.225
- Crowd + Image + User
- Crowd + User

Accuracy:
- 0
- 0.225
- 0.45
- 0.675
- 0.9

User category preference:
- Deep learning features

1 of 32 category

Personalisation

Repin vs. Noaction
### Task #2

**Category Prediction**

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.19</td>
</tr>
<tr>
<td>User Deep Learning Features</td>
<td>0.42</td>
</tr>
<tr>
<td>Image + User</td>
<td>0.77</td>
</tr>
<tr>
<td>Crowd + Image + User</td>
<td></td>
</tr>
<tr>
<td>Crowd + User</td>
<td></td>
</tr>
</tbody>
</table>

**Accuracy Scores:**

- 0
- 0.225
- 0.45
- 0.675
- 0.9

**Changtao Zhong**

Microsoft Research Asia, 10 Feb 2015
Category Prediction

Task #2

- Random
- User category preference
- User
- Deep learning features
- Image + User
- Crowd + Image + User
- Crowd + User

First 5 votes

Accuracy

0.19
0.42
0.77

0 0.225 0.45 0.675 0.9

Changtao Zhong
Microsoft Research Asia, 10 Feb 2015
Task #2

Category Prediction

First 5 votes

Random: 0.19
User: 0.42
Image + User: 0.77
Crowd + Image + User: 0.88

User category preference
Deep learning features
Task #2

Category Prediction

Random 0.19
User category preference
User 0.42
Deep learning features
Image+User 0.77
Crowd+Image+User 0.88
First 5 votes
Crowd+User 0.85

Accuracy

0 0.225 0.45 0.675 0.9

Task 1: Attention Prediction
Task 2: Category Prediction
Task 3: Pinboard
Personalisation
Repin vs. Noaction
1 of 32 category
Task #2

Category Prediction

Accuracy

Random 0.19
User category preference 0.42
Deep learning features 0.77
Image+User 0.77
Crowd+Image+User 0.88
Crowd+User 0.85

First 5 votes

Task 1: Attention Prediction
Task 2: Category Prediction
Task 3: Personalisation

Repin vs. Noaction
1 of 32 category
Evaluation

Prediction cascade

- Task 1: Attention Prediction
  - Repin vs. Noaction
  - 1 of 32 category
- Task 2: Category Prediction
- Task 3: Pinboard
  - Personalisation

Random Forest Classifier
Prediction cascade

Task 1: Attention Prediction (77%)
- Repin vs. Noaction
- 1 of 32 category

Task 2: Category Prediction
- Personalisation

Task 3: Pinboard
- Task 1
- Task 2

Random Forest Classifier
Prediction cascade

Task 1: Attention Prediction (77%)
Task 2: Category Prediction (88%)
Task 3: Pinboard Personalisation

Repin vs. Noaction (1 of 32 category)

Random Forest Classifier
Evaluation

Prediction cascade

Task 1: Attention Prediction
- Repin vs. Noaction
- 1 of 32 category
- 77%

Task 2: Category Prediction
- 88%

Task 3: Pinboard
- Personalisation
- 73%

Random Forest Classifier
Evaluation

Prediction cascade

User

Image/Pin

Attention Prediction

Task 1
77%

Repin vs. Noaction
1 of 32 category

Random Forest Classifier

Task 2
Category Prediction

88%

Task 3
Personalisation

Pinboard

73%

Changtao Zhong
Microsoft Research Asia, 10 Feb 2015
Prediction cascade

User

Image/Pin

Task 1
Attention Prediction
77%

Task 2
Category Prediction
88%

Task 3
Pinboard
73%

Personalisation

Random Forest Classifier

Acc@5=75%

Repin vs. Noaction
1 of 32 category
Outline

• What type of content is curated? Niche content

• Why do users curate? For personal reasons.

• Can we automate content curation?
  • Yes, using a mix of deep learning and crowdsourcing.

[ICWSM13]

• Can social bootstrapping create a good community?
  • Copying is useful to initiate social interaction
  • active/influential users tend to move away from copied to native friends.

[WWW14]
Outline

- What type of content is curated? Niche content
- Why do users curate? For personal reasons.  
  [ICWSM13]
- Can we automate content curation? 
  - Yes, using a mix of deep learning and crowdsourcing.  
  [WWW15]
- Can social bootstrapping create a good community? 
  - Copying is useful to initiate social interaction
  - active/influential users tend to move away from copied to native friends.  
  [WWW14]
The Dilemma for New Websites: How to construct social network?

- Option 1:
  - Create entirely new social network

- Option 2:
  - Social Bootstrapping
Social Bootstrapping

- The process of **copying** links from established social networks (source network) onto a third-party website (target network).

![Diagram of social networks and login options](image-url)
Social Bootstrapping in action: Friend Finder in Pinterest
Social Bootstrapping in action: Friend Finder in Pinterest
Social Bootstrapping in action: Friend Finder in Pinterest

User A

Log In with Facebook

Connected User

N
Y
Social Bootstrapping in action: Friend Finder in Pinterest
Social Bootstrapping in action: Friend Finder in Pinterest

User A

Connected User

All Connected Users

Facebook Friends of A

Friends not using Pinterest

Friends using Pinterest

Facebook
Social Bootstrapping in action: Friend Finder in Pinterest

User A → Log In with Facebook → Connected User

All Connected Users → Friends not using Pinterest

Facebook Friends of A → Friends using Pinterest

Invite friends to Pinterest

Friends not using Pinterest

Friends using Pinterest
Social Bootstrapping in action: Friend Finder in Pinterest
Social Bootstrapping in action: Friend Finder in Pinterest

End result: a subset of links of the user are copied from Facebook to Pinterest.
Analytical Model: Link Bootstrapping Sampling

- **Node sampling:**
  - Users in target network connect to their accounts in source network.

- **Link sampling:**
  - Connected users import friends from source network to target network.
Analytical Model: Link Bootstrapping Sampling

• **Node sampling**:  
  • Users in target network connect to their accounts in source network.

• **Link sampling**:  
  • Connected users import friends from source network to target network.
Analytical Model: Link Bootstrapping Sampling

- **Node sampling**: Users in target network connect to their accounts in source network.

- **Link sampling**: Connected users import friends from source network to target network.
Datasets*: Pinterest

- **Connected users**
  - Users that have connected with their Facebook accounts
    - 40m / 68m

- **Copied links**
  - Links copied from Facebook
    - 1b / 3.8b

Social Bootstrapping has advantages on paper

✓ Can instantly bootstrap from a mature network.
  • Facebook has 10 years of history; Twitter 8 years.¹

✓ Not “yet another” network fighting for user attention
  • 71% of online adults are now Facebook users²

Different networks have different purposes

General-purpose Social Networks

Facebook
Twitter
LinkedIn

Interest-based Social Networks

Google+
Pinterest
Spotify
Vimeo
Last.fm
Spotify

42
Different networks have different purposes

![Facebook](image1.png)  ![Twitter](image2.png)  ![Google+](image3.png)  ![LinkedIn](image4.png)  ![Spotify](image5.png)

![Pinterest](image6.png)  ![Vimeo](image7.png)  ![Last.fm](image8.png)

Does copying create a good social community on target website?

- General-purpose Social Networks
- Interest-based Social Networks
Research Questions/Outline

• *Q*: *Does copying create a good social community on target website?*

• **Structural Benefits**: Copying helps users get started with “good” structure with more social interactions

• **“Weaning”**: Beyond Bootstrapping, active and influential users wean from Facebook to create new links natively
Structural Benefits: Copied network vs. native network

Our dataset show that:

- **Reciprocity**: Copied > Native
- **Clustering**: Copied > Native
- **Connectivity**: Giant Connected Component appears in copied networks quickly (according to our analytical model)
Structural Benefits: Copied network vs. native network

Our dataset show that:

- **Reciprocity**: Copied > Native
- **Clustering**: Copied > Native
- **Connectivity**: Giant Connected Component appears in copied networks quickly (according to our analytical model)

Copying links results in a **stronger** and **denser** social structure.
Structural benefits → Social interaction?

• **Repin** (the most popular activity on Pinterest):
  • Put images published by others into one’s own collections.
  • Define **Social Repins**: Repins made by friends.
  • Check **Are social repins more on copied or native links?**
Copied links richer in social repins
Copied links richer in social repins

Copying creates networks good for social interaction

Image: CDF of the fractions of users’ natively created and copied links which are sampled by the repin network. Copied links tend to have more repins.

Figure 4: How the social repin network samples the Pinterest graph (0-valued points not shown): (a) CDF of fraction of users’ reciprocated links (resp., CR=1) and can be termed unreciprocated (directed) links, which are included in the repin network. A greater fraction of reciprocated links than directed links are the most active. Pinterest and Last.fm natives (CR=0) networked users (0 < CR < 1) with a mixture of native and copied links are the least active, whereas bi-networked users (CR=1) that is, higher activity levels are associated with lower copy ratios. Fig. 6f shows that this result extends to measures of influence on Pinterest. We find that users who are influential, measured by repins, tend to have lower copy ratios. (In the case of influence, investment in natively formed links increases proportionally. Overall, the results above indicate that as users settle down sure.) Overall, the results above indicate that as users settle down.
Research Questions/Outline

• Q: Does copying create a good social community on the target website?

• **Structural Benefits**: Copying helps users get started with “good” structure with more social interactions

• “Weaning”: Beyond Bootstrapping, active and influential users wean from Facebook to create new links natively
Active/influential users copy less

Activity level of users (measured by pins)

Copy ratio (average)
Is it because active users are less social?
Is it because active users are less social? NO!
Weaning from Facebook

\[
\text{FB Repin Ratio} = \frac{\text{Repins of Facebook friends}}{\text{Repins of Facebook and Native friends}}
\]

Activity level of users (measured by pins)
Weaning from Facebook

$\text{FB Repin Ratio} = \frac{\text{Repins of Facebook friends}}{\text{Repins of Facebook and Native friends}}$

Users evolve to repin **less** from Facebook friends and **more** from Native Friends.

**Activity level of users (measured by pins)**
Why do active/influential users wean from copied to native friends?

![Graph showing CDF of copied and uncopied friends](image)
Why do active/influential users wean from copied to native friends?

Native friends match user's interests more than friends copied from Facebook.
Answers to Research Questions

Q: Does social bootstrapping by copying links from Facebook create a good social community on the target website?

• Copying is **useful to initiate social interaction**

• Taking a long-term view, **active/influential users tend to move away** from copied social links and build social relationships natively.
Q: Does social bootstrapping by copying links from Facebook create a good social community on the target website?

- Copying is **useful to initiate social interaction**
- Taking a long-term view, **active/influential users tend to move away** from copied social links and build social relationships natively.

Copying and building links natively are both equally important to the success of target website.
Outline

Social Bootstrapping

Content Curation

Organised content

Online content (images)

images

Outline

Social Bootstrapping

Content Curation

Organised content

Online content (images)

images

Outline

Social Bootstrapping

Content Curation

Organised content

Online content (images)

images
Outline

- What type of content is curated? Niche content
- Why do users curate? For personal reasons.
  - [ICWSM13]
- Can we automate content curation?
  - Yes, using a mix of deep learning and crowdsourcing.
  - [WWW15]
Outline

<table>
<thead>
<tr>
<th>What type of content is curated?</th>
<th>Niche content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why do users curate?</td>
<td>For personal reasons.</td>
</tr>
<tr>
<td>Can we automate content curation?</td>
<td>Yes, using a mix of deep learning and crowdsourcing.</td>
</tr>
</tbody>
</table>

Can social bootstrapping create a good community?

- Copying is useful to initiate social interaction
- Active/influential users tend to move away from copied to native friends.

[ICWSM13]

[WWW14]

[WWW15]
Thank you!

[ICWSM12] How to tell Head from Tail in User-Generated Content Corpora?
[WWW14] Social Bootstrapping: How Pinterest and Last.fm Social Communities Benefit by Borrowing Links from Facebook.

Our dataset is available at http://bit.ly/pinterest-dataset