Social Bootstrapping

How Pinterest and Last.fm Social Communities Benefit by Borrowing Links from Facebook

C. Zhong¹, M. Salehi², S. Shah³, M. Cobzarenco⁴, N. Sastry¹, M. Cha⁵
¹King's College London, ²University of Tehran, ³UC Berkeley, ⁴Last.fm, ⁵KAIST

10 April 2014
Social Bootstrapping

How Pinterest and Last.fm Social Communities Benefit by Borrowing Links from Facebook

C. Zhong¹, M. Salehi², S. Shah³, M. Cobzarenco⁴, N. Sastry¹, M. Cha⁵
¹King's College London, ²University of Tehran, ³UC Berkeley, ⁴Last.fm, ⁵KAIST

10 April 2014
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).
Social Bootstrapping in action: Friend Finder in Pinterest
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 → Facebook Friends of A

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

User A → Log In with Facebook → Connected User

Y

All Connected Users

Friends not using Pinterest

Friends using Pinterest

Facebook

Friends of A

Facebook
Social Bootstrapping in action: Friend Finder in Pinterest

User A

Log In with Facebook

Connected User

All Connected Users

Facebook Friends of A

Friends not using Pinterest

Friends using Pinterest

Invite friends

Invite friends to Pinterest

Bi Zhao
Invite

Suniil Shah
Invite

黎黎莎
Invite
Social Bootstrapping in action: Friend Finder in Pinterest

User A

Log in with Facebook

Connected User

All Connected Users

Facebook Friends of A

Friends not using Pinterest

Friends using Pinterest

Invite friends

Copy friends

Copied Friends of A
Social Bootstrapping in action: Friend Finder in Pinterest

End result: a subset of links are copied from Facebook to Pinterest for each connected user.

 Invite friends

 Copy friends

Friends not using Pinterest

Friends using Pinterest

User A

Log In with Facebook

Connected User

Facebook

Copied Friends of A

Y

N
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.
Collecting Pinterest Data
Collecting Pinterest Data

Facebook handle
Collecting Pinterest Data

Facebook handle
Activity info
Social Network
Pinterest Dataset

An anonymized version of the Pinterest dataset used in our WWW14 and ICWSM13 papers is being made available to the research community. If you are interested in using this data, please send us an email at netsys@kcl-dot-AC-dot-uk to get the link where you can download the data. Note that sending the email indicates that you accept our terms and conditions in the following section. Please indicate which of the following parts of the dataset you need in the email.

1. **Pinterest network**: A snowball sampled social graph of Pinterest, crawled in Apr 2013.

2. **Facebook network**: The Facebook social graphs of users who appear in the Pinterest activities dataset below.

3. **Fb-copied network**: The subset of Pinterest network that only contains links common to both Pinterest and Facebook.

4. **User information**: The basic statistics (such as the number of pins, likes, followees and followers) of Pinterest users.

5. **Pinterest activities**: Repin and like activity in Pinterest during 03-21 Jan, 2013.
Pinterest Dataset

Datasets
An anonymized version of the Pinterest dataset used in our WWW14 and ICWSM13 papers is being made available to the research community. If you are interested in using this data, please send us an email at netsys@kcl-dot-AC-dot-uk to get the link where you can download the data. Note that sending the email indicates that you accept our terms and conditions in the following section. Please indicate which of following parts of the dataset you need in the email.

1. **Pinterest network**: A snowball sampled social graph of Pinterest, crawled in Apr 2013.
2. **Facebook network**: The Facebook social graphs of users who appear in the Pinterest activities dataset below.
3. **Fb-copied network**: The subset of Pinterest network that only contains links common to both Pinterest and Facebook.
4. **User information**: The basic statistics (such as the number of pins, likes, followees and followers) of Pinterest users.
5. **Pinterest activities**: Repin and like activity in Pinterest during 03-21 Jan, 2013.
Pinterest Dataset


Datasets

An anonymized version of the Pinterest dataset used in our WWW14 and ICWSM13 papers is being made available to the research community. If you are interested in using this data, please send us an email at netsys@kcl-dot-AC-dot-uk to get the link where you can download the data. Note that sending the email indicates that you accept our terms and conditions in the following section. Please indicate which of following parts of the dataset you need in the email.

1. **Pinterest network**: A snowball sampled social graph of Pinterest, crawled in Apr 2013.
2. **Facebook network**: The Facebook social graphs of users who appear in the Pinterest activites dataset below.
3. **Fb-copied network**: The subset of Pinterest network that only contains links common to both Pinterest and Facebook.
4. **User information**: The basic statistics (such as the number of pins, likes, followees and followers) of Pinterest users.
5. **Pinterest activities**: Repin and like activity in Pinterest during 03-21 Jan, 2013.
Datasets*: Pinterest and Last.fm

- **Connected users**
  - Users that have connected with their Facebook accounts
    - Pinterest: 40m / 68m
    - Last.fm: 600k

- **Copied links**
  - Links copied from Facebook
    - Pinterest: 1b / 3.8b
    - Last.fm: 2.8m

Datasets*: Pinterest and Last.fm

- **Connected users**
  - Users that have connected with their Facebook accounts
    - Pinterest: 40m / 68m
    - Last.fm: 600k

- **Copied links**
  - Links copied from Facebook
    - Pinterest: 1b / 3.8b
    - Last.fm: 2.8m

Datasets*: Pinterest and Last.fm

- **Connected users**
  - Users that have connected with their Facebook accounts
    - Pinterest: 40m / 68m
    - Last.fm: 600k

- **Copied links**
  - Links copied from Facebook
    - Pinterest: 1b / 3.8b
    - Last.fm: 2.8m

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

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

BUT: Different networks have different purposes. Does copying create a good social community on the target website?

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
Structural Benefits

Our analyses show that:

- **Reciprocity**: Copied > Native
- **Clustering**: Copied > Native
- **Connectivity**: Giant Connected Component appears easily
Structural Benefits

Our analyses show that:

- **Reciprocity:**  *Copied > Native*
- **Clustering:**  *Copied > Native*
- **Connectivity:**  *Giant Connected Component appears easily*

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.

• Social Repins: Repins made by friends.
Copied links richer in social repins

![Graph showing cumulative distribution function of social repin ratio for Pins native and Fb-copied links.]

- The majority of copied links (50–60%) are copied from the Facebook network.
- A smaller fraction of Facebook expats (CR=1) compared to Pinterest.
- A larger fraction of natives who have zero copy ratio (CR=0) and a smaller fraction of bi-networked users (0 < CR < 1) with a mixture of native and copied friends are more active. Pinterest and Last.fm natives (CR=0) are entirely copied from Facebook and are the least active, whereas bi-networked users have a larger fraction of copied links.
- User activity or influence level is inversely proportional to copy ratio. Higher activity levels are associated with lower copy ratios.
- Users who copy a lot tend to have lower activity levels, whereas users who pin or scrobble a lot tend to have lower levels of influence on Pinterest and Last.fm.
- Investment in natively formed links increases proportionally. Users who start with the bootstrapping advantage tend to move away from reliance on the existing Facebook network and start building new connections.
- Overall, the results above indicate that as users settle down and become more active and influential, their reliance on copied links decreases, and they invest more in natively formed links.

For a directed network, representing a node's follower (resp., followee) set in the target network (i.e., Pinterest) by $\mathcal{F}_{\text{out}}$ (resp., $\mathcal{F}_{\text{in}}$), who only create links natively on the website (i.e., $\mathcal{F}_{\text{out}} \cap \mathcal{F}_{\text{in}} = \emptyset$). Denoting the set of all friends in the target network by $\mathcal{F}$, the copy ratio in a undirected network, such as Last.fm's, is defined as:

$$\text{copy ratio} = \frac{|\mathcal{F}_{\text{out}} \cap \mathcal{F}_{\text{in}}|}{|\mathcal{F}|}.$$
Copied links richer in social repins

![Graph showing the cumulative distribution function of social repin ratios for native and copied links. The graph compares the distribution of social repins for native links (Pnt-native) and copied links (Fb-copied). The copied links are shown to have a higher cumulative distribution, indicating a richer concentration of social repins compared to native links.]
Copied links richer in social repins

Copy ratios show that Facebook expats (CR=1) whose social links are entirely copied from Facebook are the least active, whereas bi-networked users (0<CR<1) with a mixture of native and copied friends are more active. Users who start with the bootstrapping advantage tend to move away from reliance on the existing Facebook network and start building new social connections.

Fig. 6a–c. All combinations of activity measures and the two social networks are explored

5.2 Active and influential users copy fewer links

For activity (or influence) as the fraction of social repins made (or received), we define a user's social repin ratio as:

\[ \text{Social Repin Ratio} = \frac{\text{social repins}}{\text{all repins}} \]

In order to study levels of copying, we introduce a measure called the copy ratio \( CR \) as:

\[ CR = \frac{|\text{native friends}|}{|\text{all friends}|} \]

For a directed network, representing a node's follower (resp., follower copy ratio \( fr \)), we define:

\[ fr_{\text{ind}} = \frac{|\text{native followers}|}{|\text{all followers}|}, \quad fr_{\text{out}} = \frac{|\text{native followers}|}{|\text{all followers}|} \]

Denoting the set of all friends in the target network (i.e., Pinterest) by \( \text{all} \), the copy ratio in a undirected network, such as Last.fm's, is defined as:

\[ CR = \frac{|\text{native friends}|}{|\text{all friends}|} \]

Fig. 5 shows the cumulative distribution of copy ratios (CR).

5.3 Influential and active users remain social, but with native rather than copied friends

As activity levels increase, for the case of pins and scrobbles in Pinterest and Last.fm respectively. This demonstrates a clear inverse relationship between the activity levels and copy ratio, with users who pin or scrobble a lot tending to have lower levels of copying—that is, higher activity levels are associated with lower copy ratios. Fig. 6f shows that this result extends to measures of social interaction.

Fig. 6d–e drill down further and examine how the copy ratios change as activity levels increase, for the case of pins and scrobbles in Pinterest and Last.fm respectively. This demonstrates a clear inverse relationship between the activity levels and copy ratio, with users who pin or scrobble a lot tending to have lower levels of copying—that is, higher activity levels are associated with lower copy ratios. Fig. 6f shows that this result extends to measures of social interaction.

Copying creates networks which is good for social interaction
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

Figure 6: (a-c) Facebook-expats (CR=1) exhibit lesser activity in terms of pins, likes and scrobbles, compared to Pinterest and Last.fm natives (CR=0) and bi-networked users (0 < CR < 1), who are the most active. (d-e) Higher activity levels measured by pins (scrobbles) are associated with lower following copy ratio in Pinterest (Last.fm). (f) Users who are influential on Pinterest, as measured by repins, tend to have lower copy ratios.

Figure 7: (a) Social repins increase proportionally as activity level, measured by pins, increases. (b) The same increasing trend is shown for influence, measured by likes received. (c) However, the fraction of social repins over links copied from Facebook decreases proportionally as influence, measured by likes received, increases.

We also focus on social repins and ask whether copied links promote social repins. We define the Facebook repin ratio for activity (or influence) as the fraction of social repins made (or received) over Fb-copied links among all social repins made (or received). Fig. 7c reveals that as activity (or influence) levels increase, social repins happening over copied links decrease.

5.4 Weaning, biases and community evolution

We conclude by asking how the nature of the target network community would evolve as users 'wean' from copying to make more native links. To understand this, we study user preferences or biases in the kinds of links they copy and the links they make natively. We also seek to understand the role that copying plays in creating more native links.

User studies in previous work on Pinterest [27] identified that Pinterest users most value the social aspect of the service that helps them find people with similar tastes in pictures. Therefore, we examine whether natively created links on Pinterest enables discovery of individuals with a more similar taste than those with copied links. Specifically, if \( I_1 \) is the set of user \( u_1 \)'s board categories, and \( I_2 \) is the set of \( u_2 \)'s, we define their similarity as

\[
s = \frac{|I_1 \\ I_2|}{|I_1|}.
\]

Fig. 8a confirms that according to this measure, users connected by native Pinterest links are more similar to each other than those connected by Facebook-copied links. Fig. 8a also shows that there is no difference in similarity between users who are copied and users who are not copied over from Facebook. This implies that users are not selecting Facebook friends to copy based on similarity. On Last.fm, native links also show higher similarity than Facebook copied links, which in turn show higher similarity than Facebook links which were deliberately not copied. We conjecture that on some websites like Pinterest, there might be social norms at play which are being used to decide which links could be copied over.
Are active users less social?

Figure 6: (a-c) Facebook-expats (CR=1) exhibit lesser activity in terms of pins, likes and scrobbles, compared to Pinterest and Last.fm natives (CR=0) and bi-networked users (0 < CR < 1), who are the most active. (d-e) Higher activity levels measured by pins (scrobbles) are associated with lower following copy ratio in Pinterest (Last.fm). (f) Users who are influential on Pinterest, as measured by repins, tend to have lower copy ratios.

Figure 7: (a) Social repins increase proportionally as activity level, measured by pins, increases. (b) The same increasing trend is shown for influence, measured by likes received. (c) However, the fraction of social repins over links copied from Facebook decreases proportionally as influence, measured by likes received, increases. Interaction continues to be increasingly essential for active (or influential) users.

We also focus on social repins and ask whether copied links promote social repins. We define the Facebook repin ratio for activity (or influence) as the fraction of social repins made (or received) over Fb-copied links among all social repins made (or received). Fig. 7c reveals that as activity (or influence) levels increase, social repins happening over copied links decrease.

5.4 Weaning, biases and community evolution

We conclude by asking how the nature of the target network community would evolve as users 'wean' from copying to make more native links. To understand this, we study user preferences or biases in the kinds of links they copy and the links they make natively. We also seek to understand the role that copying plays in creating more native links.

User studies in previous work on Pinterest [27] identified that Pinterest users most value the social aspect of the service that helps them find people with similar tastes in pictures. Therefore, we examine whether natively created links on Pinterest enables discovery of individuals with a more similar taste than those with copied links. Specifically, if $I_1$ is the set of user $u_1$'s board categories, and $I_2$ is the set of $u_2$’s, we define their similarity as $s = |I_1 \setminus I_2| / |I_1|$. Fig. 8a confirms that according to this measure, users connected by native Pinterest links are more similar to each other than those connected by Facebook-copied links. Fig. 8a also shows that there is no difference in similarity between users who are copied and users who are not copied over from Facebook. This implies that users are not selecting Facebook friends to copy based on similarity. On Last.fm, native links also show higher similarity than Facebook copied links, which in turn show higher similarity than Facebook links which were deliberately not copied. We conjecture that on some websites like Pinterest, there might be social norms at play which are being used to decide which links could be copied over.
Weaning from Facebook

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

Activity level of users (measured by likes)

FB repin ratio (Average)
Weaning from Facebook

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

Users evolve to repin less from Facebook friends and more from Native Friends.
Why do active/influential users wean from copied to native friends?

![Cumulative Distribution Function](chart.png)

- **Copied**
- **Not Copied**
- **Pnt-Native**

**Interest Similarity** vs **Cumulative Distribution Function**

- The chart illustrates the cumulative distribution of interest similarity between copied and uncopied friends.
- Users with copied friends tend to have a higher cumulative distribution function, indicating a closer interest similarity.
- The differences between copied and uncopied friends are statistically significant.

---

**Social Bootstrapping**

Including the above studies, most existing research on the design of new user-driven communities around health issues, learning environments, and social networks are known to facilitate the formation of learning communities, enhance user engagement and community interaction. Social networks are also the core of the all user experience for students in synchronous and asynchronous learning environments.

---

**6. RELATED WORK**

---

**Many websites try to incorporate a social networking aspect to build stronger bonds with their weak ties and to reach out to employees they do not know.** Another study identified that social environments ranging from professional settings to online games contribute to the motivation of users in creating social network links. One intended purpose for the copied link in the target network. i.e., copying creates the opportunity to discover and follow them, creating new native links.

---

**Pinterest users tend to use the "friend finder" tool to copy close friends they know from established source networks like Facebook, but when they discover new friends on the target network, they tend to prefer users with similar tastes.** Thus, as native links become more important and numerous, native friends are more similar to a user than copied friends, but copied and uncopied friends do not differ significantly in tastes.

---

**Distribution**

- **Interest Similarity**
- **Cumulative Distribution Function**

**Fig. 8**

(a) Native friends have similar interests. (b) Per-user CDF of closeness between copied and uncopied friends. It shows that copied friends are closer than uncopied friends. (c) The difference between closeness of copied and uncopied friends.

---

**Another aspect of this paper was the large-scale empirical study showing that native friends are more similar to a user than copied friends, but copied and uncopied friends do not differ significantly in tastes.**
Why do active/influential users wean from copied to native friends?

Figure 8: User choice in copying links (Pinterest data) (a) CDF of similarity between users linked by copied, uncopied and native links, (b) Per-user CDF of closeness between copied and uncopied friends. It shows that copied friends are closer than uncopied friends. (c) The distribution of copied and uncopied links shows that closeness between copied friends is higher than between uncopied friends. In Last.fm, however, there is no significant difference between copied and uncopied friends. In Facebook, there is a significant difference.

Interested in the content-driven website. However, copying continues to be important for the creation of communities, and users are more likely to copy links they have previously copied.

Together these results suggest that Pinterest users tend to use the "friend finder" tool to copy close friends they know from established source networks like Facebook, but when they discover new networks, they tend to prefer users with similar tastes. Thus, as native links become more important and numerous, the more friends a user copies and follows in the target network, the more they tend to prefer users with similar tastes.
Why do active/influential users wean from copied to native friends?

Because native friends match their interests more than friends copied from Facebook.
Answers

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.**
Thank you!

Our dataset is available at


Social Bootstrapping: How Pinterest and Last.fm Social Communities Benefit by Borrowing Links from Facebook

C. Zhong1, M. Salehi2, S. Shah3, M. Cobzarenco4, N. Sastry1, M. Cha5

1King's College London, 2University of Tehran, 3UC Berkeley, 4Last.fm, 5KAIST