Oh The Places You'll Share: An Affordances-Based Model of Social Media Posting Behaviors

Harmanpreet Kaur
University of Michigan
harmank@umich.edu

Cliff Lampe
University of Michigan
cacl@umich.edu

Isaac Johnson
Northwestern University
isaacj@u.northwestern.edu

Brent Hecht
Northwestern University
bhecht@northwestern.edu

Hannah J. Miller
University of Minnesota
hmiller@cs.umn.edu

Walter S. Lasecki
University of Michigan
wlasecki@umich.edu

Loren G. Terveen
University of Minnesota
terveen@cs.umn.edu

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada
ACM 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188601

Abstract
With rising use of multiple social network sites (SNSs), people now have an increasing number of options for audience, media, and other SNS features at their disposal. In this paper, our goal is to build machine learning models that can predict people's multi-SNS posting decisions, thus enabling technology that can personalize and augment current SNS use. We explore affordances—the perceived utilities of a SNS's features—for creating these models. We build an affordance-based model using data collected from a survey about people's multi-SNS posting behavior ($n = 674$). Our model predicts posting decisions that are ∼35% more accurate compared to a random baseline, and ∼10% more accurate than predictions based on SNS popularity.

Author Keywords
Social media ecosystem; affordances; usage models

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction and Background
Many people now use multiple social network sites (SNSs) instead of just one. According to the latest report by Pew Research Center, 56% of online adults use more than one of the following SNSs: Facebook, Twitter, Instagram, Pinter-

est, and LinkedIn [3]. As this multi-SNS—or *SNS ecosystem*—use becomes more prominent, there are combinatorially more options for audience, content and site features to consider. This constantly changing social media ecosystem enables new communication opportunities and behaviors that researchers have begun to study from both theoretical and practical standpoints. Prior work has reevaluated important theories of communication and media use based on this new behavior. For example, people share different things on different subsets of SNSs because of a lowest common denominator approach – only sharing content that is relevant to all audiences on a site [4]. From a practical standpoint, SNSs have now made this multi-SNS use easier, e.g., by adding features for cross-posting, and aggregating multiple SNS feeds in one place.

A key element for building applications for multi-SNS use is being able to translate theoretical understanding of this behavior into computational models. In this paper, we attempt to accomplish this goal by building models that can predict multi-SNS use. As the SNS ecosystem continues to grow, building models that can predict people's multi-SNS posting decisions can serve as an enabling technology for SNSs and other systems to better anticipate user needs. If we can model how people use their SNS ecosystems for posting content, we can build tools for augmenting people's decision-making processes to avoid errors when cognitive heuristics break (e.g., when how they usually post on an SNS no longer applies due to new features). Further, we can build ecosystem-level tools for easy sharing across subsets of SNSs (potentially even personalized per user), helping anticipate useful functionality or recommendations.

We use *affordances* as a means of creating models for explaining people's posting decisions in their SNS ecosystems. According to Gibson, "affordances are properties taken with reference to the observer [2, p.135]." He explains that all objects have features that distinguish them from other objects, but what distinguishes objects and *people's use of them* is each person's perception of the objects' features. We hypothesize that this inherent relationship between affordances and object usage translates to the online sphere: affordances can potentially explain people's choices of SNSs they use when posting something. Prior work has also advocated for taking this affordance-based approach in studying SNS ecosystems [6]. Thus, our primary research question is: *how well do affordances explain people's posting decisions in the social media ecosystem?*

We address this question using a scenario-based survey of 674 Mechanical Turk participants that elicited their posting decisions for hypothetical communication scenarios. Our participants answered two sets of questions about the same affordances: (1) the affordances they *desire* for posting content in a scenario, and (2) the affordances they *anticipate* in each SNS they use. We use data from this survey to build a SVM-Based Model that models posting decisions using machine learning classifiers to find the ideal matches of desired and anticipated affordances. Our results provide a strong initial signal that affordances can explain people's SNS posting decisions. Our model makes $\sim$35% more accurate posting decisions compared to a random baseline, and $\sim$10% more accurate than a popularity baseline.

Modelling SNS Ecosystems Using Affordances

Given our goal of building affordance-based models that can predict multi-SNS use, we needed a universal operationalization of affordances—one that explained the utilities derived from individual SNSs while simultaneously being broad enough that it could be used to explain the utilities of entire SNS ecosystems. We use the affordance literature to guide our operationalization and build a survey around it.

Scenario: **You read the latest book by your favorite author and want to share your opinion of it.** Please answer the following questions with this scenario in mind.

1. **Which of the following media would you use to post something to your social network(s) for this scenario?**
   - ❑ Text
   - ❑ Image
   - ❑ Video
   - ❑ Link
   - ❑ Other (Please Specify) _____
   - ❑ I would not post anything for this scenario

2. **What would be the ideal audience size with whom you would share the post?**
   - ○ Small
   - ○ Medium
   - ○ Large
   - ○ I would not post anything for this scenario

3. **How are you connected with the people with whom you would share the post?**
   - ❑ Friends that you knew in-person first
   - ❑ Family
   - ❑ Professional Connections
   - ❑ People you met online
   - ❑ People you don't know at all
   - ❑ Other (Please Specify) ____
   - ❑ I would not post anything for this scenario

4. **How would you select the people with whom you would share the post? Would you select:**
   - ❑ Specific individuals relevant for this scenario
   - ❑ A predefined, custom list of people from your social network
   - ❑ Everyone in your social network
   - ❑ Public
   - ❑ Other (Please Specify) _____
   - ❑ I would not post anything for this scenario

5. **Would you want the post to automatically disappear from your page after a certain amount of time?**
   - ○ Yes
   - ○ No, but this is something I might delete on my own after some time
   - ○ No, I would be okay with this being available permanently
   - ○ Other (Please Specify) ____
   - ○ I would not post anything for this scenario

Figure 1: Survey - Part One - Desired Affordances. Part Two was modeled using the same questions and affordances, worded differently to elicit information about anticipated affordances.

Operationalizing The Affordance Perspective

An extensive affordance literature provides useful definitions and categories of affordances; we rely on Treem and Leonardi's four affordance categories [5] for our operationalization because they follow the Gibsonian definition of affordances [2]. According to Gibson, features of a system are constant, and affordances are the utilities from those features as perceived by the user [2]. We follow this platform-agnostic perspective on affordances, and generate constructs for Treem and Leonardi's four affordance categories:

- **Visibility:** the "means, methods, and opportunities for presentation" allowed by SNSs, i.e. how information is presented on a SNS. Constructs: text, image, video, link or other media.
- **Editability:** the ability for a user to craft or edit their content before and after making it available on a SNS. Given that this affordance is allowed in a similar fashion by all SNSs we study, we do not include it in our survey or compare SNSs based on it.
- **Persistence:** the ephemerality of a post. Constructs: SNS-controlled ephemerality via an auto-delete option, user-controlled via self-delete, or permanent (never delete).
- **Association:** the established relationships between individuals or individuals and data. Our constructs here are thus audience-related: (1) *audience type*: friends, family, professional connections, people you meet online, people you do not know at all; (2) *audience boundary*: sharing with specific people, a custom list of people, everyone in your network, public; and (3) *audience size* (small, medium, large).

Survey

We conducted a two-part, between-subjects survey. Participants first provided basic demographic information and selected the SNSs that they had used at least once in the last 3 months, which is Pew Research Centers metric for active use of social media [3]. We included 10 SNSs in the survey: Facebook, Twitter, Instagram, Reddit, Pinterest, Google+, Snapchat, LinkedIn, Tumblr and Flickr. We selected these 10 platforms by first taking a union of SNSs studied by Pew Research Center (e.g., [3]) and those on Alexa's top 500 most visited sites list, and then removing any SNSs that did not comply with boyd and Ellison's definition of social network sites [1]. We required participants to be active users of at least 3 out of the 10 networks we study to ensure they were SNS ecosystem users.

For survey-part 1, we elicited people's mental model for posting content by providing them with a hypothetical scenario, i.e., a situation in which they might want to post something. Each hypothetical scenario was selected from a set of 48 scenarios that were generated as the cross product of three content characteristics: (1) *content type*: personal vs. impersonal; (2) *communication behavior*: sharing vs. seeking; and (3) *topic*: selected from a list of 12 topics found in social media literature. The topics included Arts/Culture/Literature, Quotes, Crisis, Politics, Holiday and Travel, Health and Medicine, Food and Cooking, Science and Technology, Nature and Weather, Style and Fashion, Big Life Events and Professional Information. Then, the participants were asked to choose how they would post something in the given scenario, i.e., the affordances they would use to post the content. These questions employed the affordance constructs mentioned in our operationalization above as multiple choice options (Figure 1). We also allowed participants to say "I would not share anything for this scenario" to ensure that our results reflected valid content sharing that participants would perform in a real situation. For part 2, to elicit information about the affordances people perceived in each SNS they actively used, we asked the same set of affordances-based questions from part 1, reframed in the context of each SNS. For example, "I use $\langle X \rangle$ network to post the following media (text, image, video, link, other)," and similarly for other affordances. For all these questions, we allowed participants to say "I dont post anything on $\langle X \rangle$ network," thus giving us a set of affordances for sharing needs rather than general browsing needs.

Participants and Data. We recruited participants through Amazon Mechanical Turk, requiring them to be located in the U.S. and have a HIT acceptance rate of at least 97%. We paid participants $0.80 for $\sim$5 minutes of their time. We received responses from 807 participants overall: 133 of them failed one or more of these validation checks: (i) they did not pass the three-SNS use requirement, (ii) they did not provide data about any of their SNSs by choosing "I do not post/share on $\langle X \rangle$ network" for all questions, or (iii) they did not answer the attention check question correctly (i.e., the same question asked twice, with a different ordering). After removing these responses, 674 participants remained in our dataset. These participants were distributed evenly across our 48 hypothetical scenarios (Mean=14, S.D.=2).

SVM-Based Model

We construct a model that uses a Support Vector Machine (SVM) classifier with a feature set comprised of both the desired and anticipated affordances using the constructs from our operationalization. We implement our SVM-Based Model using a set of 10 Support Vector Machine (SVM) classifiers (one per SNS). We use SVMs to model people's multi-SNS posting decisions over other probabilistic methods (e.g., Logistic Regression) because our goal is measure the extent to which affordances can explain Yes/No posting decisions, rather than determine how likely (percent-wise) it is for a SNS to be posted to. There are N feature vectors generated per participant, where N is the number of SNSs the participant uses. Each feature vector is composed of the affordances desired by the participant and the anticipated affordances of one out of N SNSs used by the participant. We encode each desired and anticipated affordance construct as a binary: $0/1$ depending on whether the individual selected it in the survey. Each SVM classifier outputs a binary Yes/No answer for whether the participant should post to the SNS being considered. We aggregate these per-SNS outputs for each participant to find their final set of SNSs selected for posting by the model, using a leave-one-out cross validation setup.

Evaluation Metrics

We test our model against ground truth data obtained from participants through our survey. We measure significance using two-tailed, paired t tests.

Precision at 1 (P@1). To find an upper bound on the precision that our model can achieve using affordance information (subject to non-zero recall), we calculate the percentage of *cases* for which at least one SNS that the model selects for posting is also one of the SNSs selected by the participant for posting. This is a test of our model's best-case performance in terms of precision because we choose the smallest non-trivial set of selections that our model can contribute with the highest confidence.

Precision and Recall. Our primary measures of model performance are precision and recall. Precision measures the number of instances selected by the model that are also relevant according to the participant. Recall measures how many relevant instances chosen by the participant are also selected by the model. We also calculate F1 score for each participant, which is a joint measure of both precision and recall — the harmonic mean of precision and recall values. We compute each metric per participant and report average values across participants.

Random Baseline. Evaluating the results of models against randomly generated results is a popular evaluation methodology employed in the field of machine learning. It is a proxy for how a system with no intuition about SNSs would make posting decisions. We calculate two values for this baseline: (1) randomly selecting one SNS from each participant's ecosystem for posting, a comparison point for our p@1 metric explained above; and (2) randomly selecting 53.54% of SNSs in each participants' ecosystem for posting, for comparing the precision, recall, and F1 score values. We pick 53.54% of SNSs in each participant's ecosystem since that is the average percentage of SNSs people use for posting among all SNSs in their ecosystems. For random baseline, P@1 = 53.5%, precision = 53%, recall = 28%, and F1 score = 0.35 (Table 1).

Popularity Baseline. Prior work on SNSs has highlighted SNS popularity as a common heuristic used by people in deciding where to post content. To reflect this, our second baseline is dependent on the global ranking of SNS popularity used by our participants. We first calculate this global ranking depending on how many people use each SNS in our study (order: Facebook, Twitter, Reddit, Instagram, Pinterest, Google+, LinkedIn, Snapchat, Tumblr, Flickr). Given this order, we calculate two types for values for this baseline: (1) always selecting the one most globally popular SNS also present in the individual's ecosystem, for comparison with p@1 value; and (2) selecting 53.54% of SNSs in each participants' ecosystem in the order of the most globally popular SNSs, for comparing with precision, recall, and F1 scores. For example, if a participant uses Instagram, Snapchat, Twitter, and Pinterest, the (2) part of this baseline would predict Twitter and Instagram for this participant. Our popularity baseline has: P@1 = 84.6%, precision = 68%, recall = 76%, and F1 score = 0.70 (Table 1).

Results

To ensure that we captured the right signal / type of affordance, we compared SVM classifiers with individual affordances and different combinations as features. While the anticipated permanence affordance was most predictive, most affordances were weighted approximately the same as one another, showing each contributes similarly. Critically, training a single-affordance SVM classifier was not as effective as using all of the affordances together. There was a significant improvement (all p « 0.005) of using *all* affordance features over each individually. Below, we present results of this all-affordances model (Table 1).

Precision at 1. To compare p@1 performance, we obtain the best single-choice SNS given by the SVM-Based Model. We do so by calculating the distance of each SNS from the classification boundary (decision surface) — the larger the distance, the more confident we can be in the prediction — and then pick the highest-confidence answer (or answers, in the event of a tie). We find that our single-choice SNS(s) are correct for 99.8% of participants, i.e. for 99.8% participants, we are able to correctly identify at least one of the SNSs that they would post to. The SVM-Based Model outperforms both baselines for p@1.

Precision / Recall Performance. For each user, we aggregate the results of our 10 per-SNS SVM classifiers, each of which predict "Yes" or "No" for posting to the corresponding SNS. Averaging across users, we find an average of 78% of all predicted posting decisions are correct (S.D. precision=21). The SVM-Based Model correctly selects

	SVM-Based Model	Random Baseline	Popularity Baseline
Precision at 1	**99.8**	53.5	84.6
Avg Precision	**78 ± 21**	53 ± 35	68 ± 28
Avg Recall	**83 ± 16**	28 ± 40	76 ± 23
Avg F1 Score	**0.79 ± 0.20**	0.35 ± 0.33	0.70 ± 0.24

Table 1: Comparison of our SVM-Based Model with baselines based on our evaluation metrics.

(recalls) an average of 83% of all SNSs selected by each participant, and has a F1 score value of 0.79. Compared to the random baseline, SVM-Based Model has 25% higher precision, 55% higher recall, and 0.44 higher F1 score. It also outperforms our popularity baseline by 10% higher precision, 7% higher recall, and 0.09 higher F1 score (all differences significant, p « 0.001).

Implications. With our affordances-based model, we are not only able to more precisely select the set of SNSs used for posting (78% precision, on average), but are also able to select an average of 83% of the set of SNSs considered ideal for posting by the participant. Given the improvement in performance of this model compared to all our baselines, we believe that affordances contain useful “signal” for explaining user posting behavior. The viability of this approach provides concrete evidence that affordances can predict a significant number ($\sim$0.80 F1 score) of posting decisions made by people. By creating these initial models, we introduce opportunities for automation in predicting multi-SNS usage using affordances, and present new baselines for future tools and methods.

Conclusion

In this paper, we have presented a method for modeling users’ social network *ecosystems*, and leveraged this model within a novel machine learning pipeline to predict where users will post content within their ecosystem. Our work presents the first step towards a more comprehensive understanding of posting behavior in light of the ever-increasing use of multiple social network sites. We hope that, by demonstrating the power of such approaches, our work encourages future research to consider this more holistic view. Furthermore, our method is orthogonal to content-based methods that use the text or images within a post to predict usage. Thus, future work can combine these methods to even more accurately model and predict usage.

REFERENCES

1. Nicole B Ellison and others. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication* 13, 1 (2007), 210–230.
2. James J Gibson. 2014. *The ecological approach to visual perception: classic edition*. Psychology Press.
3. Shannon Greenwood, Andrew Perrin, and Maeve Duggan. 2016. Social media update 2016. *Pew Research Center* 9 (2016).
4. Bernie Hogan. 2010. The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society* 30, 6 (2010), 377–386.
5. Jeffrey W Treem and Paul M Leonardi. 2013. Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association* 36, 1 (2013), 143–189.
6. Xuan Zhao, Cliff Lampe, and Nicole B Ellison. 2016. The Social Media Ecology: User Perceptions, Strategies and Challenges. In *CHI 2016*. ACM, 89–100.