Estimating the Relative Utility of Networks for Predicting User Activities

Nina Mishra
Microsoft Research
Mountain View, CA
ninam@microsoft.com

Daniel M. Romero*
Northwestern University
Evanston, IL
d-romero@kellogg.northwestern.edu

Panayiotis Tsaparas†
University of Ioannina
Ioannina, Greece
tsap@cs.uoi.gr

ABSTRACT

Link structure in online networks carries varying semantics. For example, Facebook links carry social semantics while LinkedIn links carry professional semantics. It has been shown that online networks are useful for predicting users' future activities. In this paper, we introduce a new related problem: given a collection of networks, how can we determine the relative importance of each network for predicting user activities? We propose a framework that allows us to quantify the relative predictive value of each network in a setting where multiple networks are available. We give an ϵ-net algorithm to solve the problem and prove that it finds a solution that is arbitrarily close to the optimal solution. Experimentally, we focus our study on the prediction of ad clicks, where it is already known that a single social network improves prediction. The networks we study are implicit affiliations networks, which are based on users' browsing history rather than declared relationships between the users. We create two networks based on covisitation to pages in the Facebook domain and Wikipedia domain. The learned relative weighting of these networks demonstrates covisitation networks are indeed useful for prediction, but that no single network is predictive of all kinds of ads. Rather, each category of ads calls for a significantly different weighting of these networks.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and behavioral sciences

Keywords

affiliation networks, covisitation networks, advertising, predictive modeling

*Work partially done while the author was an intern at Microsoft Research.

†Work partially done while the author was at Microsoft Research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CIKM '13 Burlingame, CA USA
Copyright 2013 ACM 978-1-4503-2263-8/13/10 ...$15.00.
http://dx.doi.org/10.1145/2505515.2505586

1. INTRODUCTION

Over the years, many user networks have come into existence. Many of these networks carry different meanings: LinkedIn is a network of professional relationships, Twitter is a network of microblogging followers, Skype is a communications network, etc. While there is evidence that user networks can be useful in predicting users' activities, it is an open question whether each network is equally useful in predicting all kinds of activities, or whether certain networks are more powerful for predicting certain activities.

In this paper, we introduce the new problem of estimating the relative utility of networks for predicting user activities. Our goal is to quantify the value that each network provides for different kinds of activities. We propose a way to estimate this utility by appealing to existing models of user behavior.

To demonstrate that our approach works, we select one application where it is already known that social networks can be predictive, namely predicting sponsored search ad clicks [4, 23, 24]. On the other hand, it has also been shown that not all categories of ads benefit from social network information [17]. This suggests that advertising is a good domain to study the predictive value of different kinds of networks. We emphasize that the problem we are addressing is more general - our abstraction can be used to predict interests, queries, and other kinds of activities. But to illustrate the approach on a real dataset, we consider ad clicks.

On top of *explicit* networks such as social, professional, and collaboration networks, there are also several *implicit* networks defined by users engaging in common activities without necessarily interacting with each other. Commonly referred to as *affiliation networks*, these networks define a specific type of relationship, depending on the activity that ties the individuals together. For example, a tie defined between two individuals that check-in at the same restaurant may be characterized by locality and similar food preferences, while a tie defined by two people who read the same political blog may signal similar political views.

As mentioned earlier, the predictive power of explicit networks (particularly social networks) has been studied. However, less is known about the utility of implicit networks. In this paper, we are particularly interested in comparing the predictive power or covisitation networks – networks defined between web users by common visits of a website. We

claim that these networks are important because they provide signals that explicit networks do not. For example, many users visit similar pages but are not necessarily connected on social networks. Furthermore, these networks can be constructed from users' browsing history without having to determine their personal connections on explicit networks such as Facebook, Twitter, or Google Plus.

While it is possible to construct a large number of networks based on visits to different websites, where each website generates one network, we choose to compare two networks based on websites that are very different in nature and very popular. We investigate covisitation networks based on visits to Wikipedia and Facebook pages. In the Wikipedia (resp., Facebook) covisitation graph, there is an edge between two users if they both visited the same Wikipedia (resp., Facebook) page.

The choice of Facebook and Wikipedia was inspired by the social and topical nature of each site, respectively. Our goal is to demonstrate that each one of these networks can be useful in predicting different kinds of activities (ad-clicks), so it is important that each network has a different meaning to better interpret the results. We will show that the relative utility of these networks depends on the ad category, i.e., for some categories the Facebook covisitation graph is more important than the Wikipedia covisitation graph and in other categories the importance is flipped. Not only can we tell which network is more important for each category, but we can also quantify the importance of each network for predicting future click activity.

Contributions: We propose a general methodology for estimating the relative utility of networks in predicting user activities and focus on ad clicks in our experiments. Given a category of advertisements and users that belong to multiple networks, our method assesses the relative value of each network in predicting click behavior. We are not aware of prior work that provides a comparative analysis of networks for this task. The model assumes that a user's click activity is governed by four key factors: what the user previously clicked, what everyone clicks, and what their connections in each network click. A model with these assumptions has been shown to work well for predicting users' activities [6]. In our model, we extend the assumptions made in [6] to include multiple networks and to allow a user's activities to be governed by other users who are multiple hops away, not just direct neighbors. Each of these factors has a relative importance (weight) depending on the specific activity. We use a ϵ-net algorithm for learning a good combination of weights and prove that the error in our solution is arbitrarily close to the error from the optimal solution. The parameters of the model reveal the relative utility that each network has in predicting click activity.

Since the networks we consider are completely synthesized, we begin with basic tests to determine the potential utility of the networks in predicting activities. We ask if pairs of users who covisit are more similar in terms of ad clicks than pairs of users that do not. We find that indeed they are more similar in both covisitation networks. In addition, we ask whether having ties that click an ad increases a user's chance of clicking the ad themselves. We find that with more ties that click the ad, the user is herself also more likely to click the ad. This initial result demonstrates that even though these covisitation networks are not based on interactions among users, their edges still carry signals regarding the users' ad clicks. This is consistent with findings from real networks [4], but since users connected in our networks may not interact with or know each other, mechanisms such as influence or homophily do not necessarily explain these results.

Having established that these networks are indeed suitable for ad prediction, we then carry out a massive-scale experiment to learn the relative utility of these covisitation networks. We focus on predicting ad categories. For each category, we learn four weights indicating the relative importance of a user's own history, what the general population does, and what their covisitors in Wikipedia and Facebook do. We show that having these networks is better than not having them, in the sense that without them the test error increases. Surprisingly, we find that a user's own personal click history has almost no power in predicting their future click activity. Finally, we show that both networks are valuable to various degrees depending on the category, although overall the Facebook covisitation network is more valuable.

2. RELATED WORK

The use of networks for improving targeted advertising has received considerable attention in the literature. Liu and Tang [17] consider an instant messaging social network and observe that the probability of clicking an ad increases when friends have also clicked the ad. They use this observation to introduce social features to a classifier that predicts click probability for behavioral targeting. Papadimitriou et al. [23] perform a field experiment and show that people with friends exposed to an advertising campaign are more likely to pose queries related to the campaign. Recently, Bakshy et al. [4] showed that on Facebook, the probability of a user to click an ad increases when given indication that friends have clicked the same ad. Provost et al. [24] considered implicit social networks for brand advertising. They used a network of co-visitations to a social media site to identify potential audiences for brand advertising. In order to identify audiences, they use measures of closeness between potential targets and users with known affinity to the brand. We build on their work and propose a method for comparing the effectiveness of different user networks in predicting users' ad click activity, which we test on co-visitation networks. We note that prior work does not consider the comparative power of different networks, and in most cases they assume that the network is given.

The idea that networks of users can be helpful in predicting their activities is based on the principles of *homophily* and *influence*. Homophily posits that people who are socially connected tend to be similar to each other [20, 15]. This phenomenon has been observed in both offline and online social networks [21, 12, 3, 10, 13, 1]. Other work in offline settings has also found evidence that both homophily and influence play a significant role [11]. Aral et al. [3] showed that distinguishing between influence and homophily driven by other factors is important to avoid overestimating the presence of influence in social contagion processes.

There has been work on finding the mechanisms that give rise to homophily. Crandall et al. [6] developed a methodology and analyzed social networks formed in Wikipedia and LiveJournal to answer the fundamental question of whether socially connected people become similar because of the influence they exert on each other after they meet, or whether they were similar before they met and their similarity causes them to meet. To answer this question, they develop a model where users are assumed to select their activities by choosing from their previous activities, their friends' activities, and the general population's activities. We made the same assumptions in our model and build a framework that allows us to compare the marginal effectiveness of multiple networks in predicting users' behavior. Furthermore, we allow users to sample their activities, not only from their neighbors' activities, but also from their second degree neighbors.

While our methodology applies to any kind of network, we are particularly interested in implicit co-visitation networks. Implicit networks of activity online and offline have been shown to be useful for predicting people's social connections. For example, Crandall et al. [7] studied the power of real life geographical co-occurrences to predict whether people actually know each other. Their work is related to ours in that geographical co-occurrences create an implicit network of common activity in an offline setting. Furthermore, Eagle et al. [8] showed that mobile phone calling behavior can predict self-reported relational information about subjects. Finally, Romero at al. [26] showed that common usage of Twitter post labels known as hashtags can predict social relationships, and that the structure of the Twitter social network can predict the eventual popularity of the hashtags. In this work, we are also interested in the users' common activities, but our aim is not to predict their social connection but to compare the relative utility of different networks in predicting future behavior.

Similarity in online behavior among socially connected users has been observed previously. Singla el at. [28] found that there is correlation between people who chat with each other and the queries they issue. Along the same lines, Panigrahy et al. [22] consider two different measures of affinity on a graph, and show that there is correlation between the similarity in the queries that people ask on the search engine and their affinity on the Hotmail correspondence graph. In this work, our methodology does not assume that we have access to a social network. Instead, we construct it based on the web browsing activity of the users alone. Furthermore, our goal is not just to show that connected pairs are similar, but to compare the predictive power of different networks, not necessarily social networks.

Finally, there is a significant amount of work in the area of *collaborative filtering* [9, 25], where similarity of past behavior between users is utilized to predict future behavior and make recommendations. In particular, collaborative filtering has been used in different contexts to make accurate predictions of product preferences [16], ad relevance to search queries [2], movie preferences [27], etc. Recently, the information about the users' social network is also incorporated in collaborative filtering techniques by exploiting the fact that connected users are likely to behave in a similar way [19, 18]. These methods bear resemblance to our approach since they use the past user behavior and social connections to predict future ad clicks. However, our main goal in this work is *not* to develop a new method for ad click prediction, but rather to investigate and compare the signal from different networks on user behavior – the prediction model is just a means to that end. Extracting this information from collaborative filtering models is computationally intensive so instead we use a simpler model, directly quantifying the prediction value of each network.

3. METHODOLOGY

In this section we formalize the problem of network value, and present our methodology.

3.1 Problem Definition

Let $\mathcal{G} = \{G_1, \ldots, G_k\}$ be a collection of k weighted undirected graphs, corresponding to k different networks in which a web user may participate. Graph $G_i = (N_i, E_i)$ has nodes N_i and weighted edges E_i, where the weight of edge (u, v) denotes the strength of the relationship between nodes u and v. The graphs are not necessarily defined over the same set of nodes, since some users may not participate in some networks. We use $U = \bigcup_{i=1}^{k} N_i$ to denote the set of users that belong to *any* of the networks in $\mathcal{G}$, and $I = \bigcap_{i=i}^{k} N_i$ to denote the users that belong to *all* of the networks in $\mathcal{G}$. The latter set is important since it allows us to study users that are affected by all different types of networks.

Now, let $\mathcal{A}$ denote the set of all ads that may be shown to the users in U. If m is the size of $\mathcal{A}$, then for every user $u \in U$ we define an m-dimensional vector V_u, where $V_u(a)$ is the number of clicks of user u on ad a. Let $A \subseteq \mathcal{A}$ denote a set of advertisements we are interested in studying. This set may consist of a single ad, the ads of a specific domain, the ads corresponding to a campaign, or all of the ads of a specific category. In our experiments the sets of ads we consider correspond to different categories of ads. For a user $u \in U$ we define the *click activity* of user u over the set A as $C_u(A) = \frac{\sum_{a \in A} V_u(a)}{\sum_{a \in \mathcal{A}} V_u(a)}$, that is, the fraction of ad clicks of user u to the ads in the set A.

Our hypothesis is that for a given set A, the click activity of the users in I over A is affected by their membership to the networks in $\mathcal{G}$. The goal of this work is to verify this hypothesis and quantify the effect of the networks on the user click activity over different subsets (categories) of ads. That is, given a set A, and a collection of graphs $\mathcal{G}$ produce a value $f_A(G_i)$ that measures the utility of network G_i in estimating the click activity $C_u(A)$.

3.2 User behavior model

To quantify *how much* the network affects the behavior of a user, we first need a model of *how* the network affects the behavior of a user. Let $u \in I$ be a user that belongs to all k networks. Making similar assumptions made in [6], we assume that the click activity of user u over a set of ads A is affected by the following factors: (a) u's own personal history; (b) the click activity of all users in U; (c) the click activity of the users that are *close* to u in the networks $G_1, ..., G_k$.

More formally, let $C_{u,t}(A)$ denote the click activity of user u over A during time period $t = (\tau_1, \tau_2)$, and let $C_{u,T}(A)$ denote the click activity of user u over A at the time period preceding time t, i.e. $T = (\tau_0, \tau_1)$. We assume that the click activity $C_{u,t}(A)$ of user u over A during time period t is determined according to the following model, parameterized by $R = \{\psi, \alpha, \gamma_1, ..., \gamma_k\}$, where $\psi + \alpha + \sum_i \gamma_i = 1$.

- With probability ψ, the click activity of user u is determined by her own past history. In other words, $C_{u,t}(A)$ is the same as $C_{u,T}(A)$; the user continues to act as she has acted so far.
- With probability α, the click activity of user u is determined by the activity of the global population. More specifically, $C_{u,t}(A)$ is the same as $C_{v,T}(A)$, where $v \in U$ is a user chosen uniformly at random from the set of all users U. If $n = |U|$, we use $C_{U,T}(A) = \sum_{v \in U} \frac{1}{n} C_{u,T}(A)$ to denote the expected click activity of the global population.
- With probability γ_i the click activity of user u is determined by the behavior of the users in network G_i. Intuitively, user u is more influenced by users that are *close* to her in the network than those that are further away. Therefore, instead of sampling uniformly at random from all the users in the network as we did for the case of U, we will bias sampling towards nodes that are close to u in G_i. We implement this idea by performing a *random walk with restarts*. The random walk starts at node u. At each step, with probability $1 - \beta$, it follows an outgoing edge of the current node, with probability proportional to the weight of the edge. With probability β, the random walk restarts at node u. This process converges to a unique stationary distribution. Let $\pi^i_u(v)$ denote the stationary probability of ending up at node v when we start the random walk at node u. Because of the restart, the stationary distribution gives higher probability to the nodes that are only a few hops away from node u and are connected with u with multiple paths. As it is customary with random walks of this type, we set $\beta = 0.15$.

  To determine the click activity $C_{u,t}(A)$ of node u we sample a node v from network G_i according to the distribution $\pi^i_u(v)$, and set $C_{u,t}(A) = C_{v,T}(A)$. If N_i is the set of nodes in network G_i, we use $C^i_{u,T}(A) = \sum_{v \in N_i} \pi^i_u(v) C_{v,T}(A)$ to denote the expected click activity of the social circle of user u in the network G_i.

Essentially, our model stipulates that user u "adopts" the click activity of some user v in U. This user is either herself, someone from the full population, or someone close to the her in one of the networks she is a member of.

Therefore, the click activity of user u over A at time t is expressed as follows.

$$C_{u,t}(A) = \psi C_{u,T}(A) + \alpha C_{U,T}(A) + \sum_{i=1}^{k} \gamma_i C^i_{u,T}(A) \quad (1)$$

Note that the ψ, α, and the γ_i's parameters are specific to the set of ads A, and they quantify the effect of each factor on the click activity of the users over this set of ads. High γ_i value for a graph G_i implies that network G_i is conductive to the spread of the ads in A within its population. If $\gamma_i > \gamma_j$, this implies that network G_i has a stronger effect on the click activity of the users over the set A than network G_j. Therefore, network G_i has higher utility for an advertiser for this class of ads.

3.3 Estimating network value

Given the model described above, we will estimate the utility of a network for a set of ads A by finding the parameter values that best fit the model to the observed data. The observed data consists of the ad clicks of the users in U over a period of time. For a set of ads A and a user u, we compute the "true" click activity $C_{u,t}(A)$ of user u over the set A during time period t by computing the fraction of clicks of user u on an ad in A over the time interval (t_1, t_2). We estimate the history of user u prior to time t, that is the click activity $C_{u,T}(A)$, as the fraction of clicks of u over the time period (t_0, t_1) over the set A. For a given setting of parameters $R = \{\psi, \alpha, \gamma_1, \gamma_2\}$, we use $\hat{C}_{u,t,R}(A)$ to denote the click activity estimated by our model. We define the sum of squares error of our estimation as follows:

$$\text{SSE}(R) = \frac{1}{|U|} \sum_u \left(\hat{C}_{u,t,R}(A) - C_{u,t}(A) \right)^2$$

We want to find the set of parameters that minimizes the $\text{SSE}(R)$ error, that is, find the set of parameters R^*_A such that

$$R^*_A = \arg\min_R \text{SSE}(R)$$

Note that the set $R^*_A = \{\psi^*_A, \alpha^*_A, \gamma^*_{A,1}, ..., \gamma^*_{A,k}\}$ is specific to the set of ads A. The value $\gamma^*_{A,i}$ defines the relative utility of the network G_i for the set A. That is, $f_A(G_i) = \gamma^*_{A,i}$.

In order to approximate R^*_A, we construct an *ϵ-net* of the parameters R_A and search for the parameters that produce the smallest error. That is, for each parameter in R_A we search in the range $[0, 1]$ with spacing ϵ and identify the set of parameters R^ϵ_A with smallest error. We only consider sets of parameters R_A that sum to 1. While it is possible that R^*_A does not fall in the ϵ-net, we prove that the error produced by the parameters R^ϵ_A is arbitrarily close to the error produced by the optimal set of parameters R^*_A. That is, we show that we can make $|\text{SSE}(R^*_A) - \text{SSE}(R^\epsilon_A)|$ arbitrarily small by choosing a small ϵ.

CLAIM 1. *Let $\delta > 0$ and $\epsilon < \sqrt{\frac{\delta}{2(k+2)}}$. Let R^ϵ_A be the set of parameters on the ϵ-net that produce the smallest error $\text{SSE}(R^\epsilon_A)$. Then $|\text{SSE}(R^*_A) - \text{SSE}(R^\epsilon_A)| < \delta$.*

PROOF. Assume that $\epsilon < \sqrt{\frac{\delta}{2(k+2)}}$ and consider a set of parameters $\hat{R_A}$ closest to R^*_A on the ϵ-net. Note that $|\psi^* - \hat{\psi}| < \epsilon$, $|\alpha^* - \hat{\alpha}| < \epsilon$, $|\gamma^*_1 - \hat{\gamma}_1| < \epsilon$, ..., $|\gamma^*_k - \hat{\gamma}_k| < \epsilon$.

We now have:

$$
\begin{aligned}
&|\mathrm{SSE}(R_A^*) - \mathrm{SSE}(\hat{R_A})| = \\
&\quad \frac{1}{|U|}\left|\sum_u \left[(\hat{C}_{u,t,R^*}(A) - C_{u,t}(A))^2\right.\right. \\
&\quad \left.\left. - \ (\hat{C}_{u,t,\hat{R}}(A) - C_{u,t}(A))^2\right]\right| \\
&\leq \quad \frac{1}{|U|}\sum_u |[\hat{C}_{u,t,R^*}(A) - C_{u,t}(A))^2 - (\hat{C}_{u,t,\hat{R}}(A) - C_{u,t}(A)]^2| \\
&\leq \quad \frac{1}{|U|}\sum_u 2[|\hat{C}_{u,t,R^*}(A) - \hat{C}_{u,t,\hat{R}}(A)|]^2 \\
&= \quad \frac{1}{|U|}\sum_u 2[|\psi^* - \hat{\psi}|C_{u,T}(A) \\
&\qquad + |\alpha^* - \hat{\alpha}|C_{U,T}(A) + \sum_{i=1}^{k} |\gamma_i^* - \hat{\gamma}_i|C_{u,T}^i(A)]^2 \\
&\leq \quad \frac{1}{|U|}\sum_u 2[|\psi^* - \hat{\psi}| + |\alpha^* - \hat{\alpha}| + \sum_{i=1}^{k} |\gamma_i^* - \hat{\gamma}_i|]^2 \\
&\leq \quad \frac{1}{|U|}\sum_u 2[\epsilon + \epsilon + k\epsilon]^2 = 2((2+k)\epsilon)^2 \\
&< \quad \delta
\end{aligned}
$$

Since $|\mathrm{SSE}(R_A^*) - \mathrm{SSE}(\hat{R_A})| < \delta$ then $|\mathrm{SSE}(R_A^*) - \mathrm{SSE}(R_A^\epsilon)| < \delta$. Otherwise, $\hat{R_A}$ would have been chosen in the ϵ-net procedure. □

4. AFFILIATION NETWORKS

To test our methodology, we used browsing logs and constructed implicit affiliation networks defined by *covisitations* of users to common web pages. In this section we describe the process of constructing the affiliation networks and confirm that the networks we constructed constitute a useful source of information for user click activity.

4.1 Affiliation network construction

Affiliation networks are defined by creating a link between two individuals if they share a common activity. This could include co-editing of a Wikipedia page, purchasing of the same product, participation in the same board, or check-in at the same venue. Affiliation networks have been studied extensively in the literature of network analysis [14, 24] and they have several advantages as potential predictors of activities: (a) They are relatively easy to obtain; (b) The links between individuals are supported by common activity and this can indicate a stronger connection than other kinds of relationships; (c) By controlling the activity that brings the two individuals together we can control the type of relationship between the two users.

In the affiliation networks we consider we define a link between two users if they share a visit to a common web page. The weight of an edge is defined as the number of common visits between the two neighbors. We could potentially construct an affiliation network for every domain on the web. However, for our experiments, we consider the affiliation networks defined by visits to Facebook and Wikipedia web pages. Our choice was motivated by two factors. First, these are two of the most popular destinations on the Web. More importantly, we wanted to test our methodology on two networks that capture two completely different types of connections. Common visits to a Facebook profile may indicate some degree of social "closeness" between the two users since they share a common interest to a member of the social network. On the other hand, common visits to Wikipedia pages may be indicative of a topical connection between the two users, since they share a common interest to a specific topic.

We state upfront that by no means do we claim that our Facebook covisitation network approximates the true Facebook network or any other social network. Covisitation does not imply friendship. For example, some Facebook profiles are celebrities, business, organizations, etc. This means that users who are not even socially close to each other could easily co-visit a celebrity page. To partially mitigate this situation, we remove pages with a very high number of visits, which probably correspond to celebrity pages. In a similar vein, Facebook friendship does not imply covisitation. Since Facebook users receive updates about their friends on their wall, there is not always a reason to explicitly visit a Facebook page. Furthermore, when friendships disappear or become weak, users may still be Facebook friends but may not visit the same Facebook pages.

On the other hand, covisitation is an indication of social proximity. Users connected in our graph covisited a person's profile, picture, status update, etc., so they are likely to be socially close, if not actually friends. In the work of Provost et al.[24], through a similar construction of the covisitation graph, they showed that connected users are very likely to visit each other's profile in the social networking site, suggesting that they indeed share a social relationship. Given that the average Facebook user has hundreds of friends, but visits a much smaller number of profiles, a visit to a page is arguably a stronger indication of interest than an explicit Facebook friendship. Additionally, our graph could include users who are not on Facebook, yet still visit Facebook profiles. Thus, in some ways, our graph is more inclusive than the Facebook graph.

Finally, we note that any social graph is structurally different from a covisitation network. While there are some common features for both graphs, such as the power law degree distribution and the existence of a giant component, the covisitation graph consists of the union of many cliques. Indeed, every Facebook profile yields a clique among all the users who visited the profile. In this sense, the structure of the covisitation graph differs dramatically from a graph of friendships.

It is easier to argue for the topical nature of the ties in the Wikipedia covisitation graph. Web users browse Wikipedia articles to either obtain or contribute information about a topic, hence a common visit implies a common topical interest. Note that there is large variation in Wikipedia content. While some may be indicative of a very specific interest such as the Wikipedia article for a local basketball team, others are very broad such as the Wikipedia article for *Sports*. Hence, each individual covisitation to a Wikipedia article may indicate a different degree of topical similarity,

but as a whole, the set of Wikipedia covisitations of a pair of users can be a strong signal of their commonalities. Similar to the construction of the Facebook graph, we remove any Wikipedia pages that were visited many times before constructing the graph.

4.1.1 Browsing Data

The data used to construct our covisitation networks was based on two months of web browsing data collected from consenting users (Dec 2011 and Jan 2012). Note that this data contains every page visited by the user, including clicks on search results, as well as visits outside of search results. The data can viewed as triplets of the form user id, page browsed, timestamp, where the user id is unique to each browser. To remove celebrity Wikipedia and Facebook pages, we removed pages visited by more than 600 users over two months.

4.1.2 Affiliation Network Statistics

Table 1 shows some basic statistics of the Facebook and Wikipedia graphs. Both covisitation networks have on the order of millions vertices, with the Facebook network having $\tilde{3}.5$ times the vertices as the Wikipedia network. The number of nodes in the network depends on site traffic, explaining why the Facebook graph has more vertices. The median degree of the Facebook graph is 65, while the median degree of the Wikipedia graph is 258. Evidently, when a user visits an average Wikipedia page, they are connected to more users than when they visit an average Facebook page. Finally, we report the number of edges in this network where pairs of users click on the same ad domain.

In the following, for simplicity, we will refer to the Wikipedia covisitation graph as the Wikipedia graph, and the Facebook covisitation graph as the Facebook graph.

4.2 Affiliation network validation

The Wikipedia and Facebook graphs we constructed capture our notion of social and topical connection, but there is no guarantee that they are appropriate for the task we are interested in, that is, predicting the user click activity. Therefore, before applying our methodology we perform two additional experiments to validate that our graphs are suitable for the task at hand.

Ad Similarity. In the first experiment, we measure the similarity between connected users in our graph with respect to their click activity. For a user u we define A_u to be the set of ad domains clicked by u. The data we collect is sponsored search ad clicks, i.e., these are ads clicked as a result of a search query. We do not consider display or contextual ad clicks. We measure the similarity between two users u, v as the Jaccard similarity of the sets A_u and A_v, $J(u,v) = \frac{A_u \bigcap A_v}{A_u \bigcup A_v}$. For a graph G, we wish to know if connected users are more similar to disconnected ones. We compute the average Jaccard similarity among users who are connected in each network, and we normalize it by average similarity of two disconnected users (the *baseline* similarity). This ratio captures the relative increase in similarity due to the connection between the users. In order to study the effect of the number of covisitations on the similarity between users, we compute this ratio for each number of covisitations separately. More precisely, for each integer $k > 1$, we define $J_k = \frac{1}{|E_k|} \sum_{(u,v) \in E_k} J(u,v)$ where E_k is the set of all user pairs with k covisitations in the network. The value of J_k is the average Jaccard similarity among pairs with k covisitations. We define $R_k = \frac{J_k}{B}$ where B is the baseline similarity of the network. The ratio R_k tells us how many times more similar pairs with k covisitation are than pairs with no covisitations at all.

Figure 1 shows the value of R_k vs. k for the two different networks with confidence intervals of 95% confidence for each value. We note three important points. First, the curves are significantly higher than 1, which corresponds to the baseline value, for all values of k. This means that even though the connections were constructed using page covisitations, which could be a weak and noisy signal, we still observe a significantly increased level of similarity among connected users. Second, for the Wikipedia network we observe a monotonic increase in R_k as k increases. For the Facebook graph, the increasing trend is much less clear. It appears that most of the power of social ties for ad similarity is realized at a single covisitation without increasing much with additional covisitations. Finally, we observe that the curve of the Facebook graph lies above that of the Wikipedia graph for small values of k, but as the number of covisitations increases the topical curve catches up to the social one. In some sense, the number of covisitations measures the strength of the tie, hence Figure 1 suggests that socially connected users are more similar than topically connected ones only for the case of weak ties.

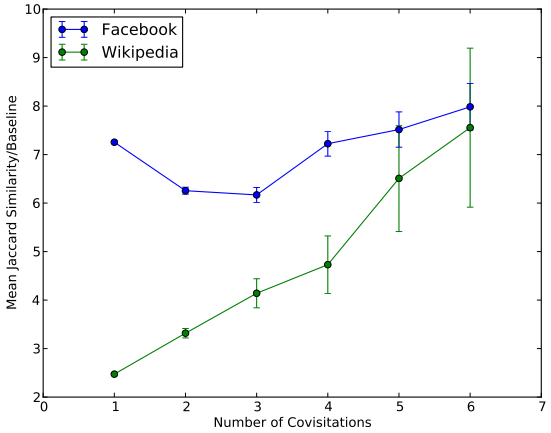

Figure 1: Average Jaccard similarity of ads over baseline as a function of number of common pages visited. Comparisons between networks based on Facebook and Wikipedia covisitations.

Ad Engagement. In the second experiment we measure how the engagement of the users with ads changes when they have ties in the network that have clicked on the ad domain. If d is an ad domain, we estimate the probability that a user will click on d given that she is connected to *exactly* k users who clicked on d. We define it by $P_k(d) = \frac{|I_k(d)|}{|X_k(d)|}$ where

Site	Number of Users	Number of Edges	Median Degree	Edges with Shared Ad Clicks
Wikipedia	7.6M	2.6B	258	368M
Facebook	26.1M	3.8B	65	442M

Table 1: Basic statistics for the Facebook and Wikipedia covisitation networks: Number of Users, number of edges, median degree and number of edges in the covisitation where users shared an ad click. Networks are based on two months of search activity.

$X_k(d)$ is the set of users connected to exactly k users who clicked on the ad domain d and $I_k(d)$ is the set of users in $X_k(d)$ who clicked d. Using the function P_k we can measure how the probability of a user clicking on an ad domain changes as the number of connections who clicked the ad domain increases. Figure 2 shows a clear increasing trend for both the Facebook and Wikipedia networks. Furthermore, we observe that topical ties provide a weaker signal for the probability of clicking on a domain ad than social ties. For these plots we opt to show the raw values instead of the increase relative to the baseline P_0. This is because the probability P_0 of a node with no connections that have clicked on the ad, to click on the ad is essentially the probability that a randomly selected node clicks on the ad, which is very small. Therefore, the ratios take very large values which are no longer informative.

Our experiments show users connected in both the Wikipedia and Facebook graphs tend to have more similar ad clicks than disconnected users, and in the case of Wikipedia the similarity increases with additional covisitations. Furthermore, for both graphs, users are more likely to click on an ad if they are connected to others who clicked on the ad. These findings suggest that the covisitation graphs are indeed suitable for our prediction task.

Note that in [4], the authors demonstrated that Facebook users are more likely to click an ad if they are explicitly told how many and which friends previously clicked the ad. In contrast, our users do not know if their ties clicked an ad, or even who their ties are. The fact that we still see the increasing trend in Figure 2 is quite remarkable.

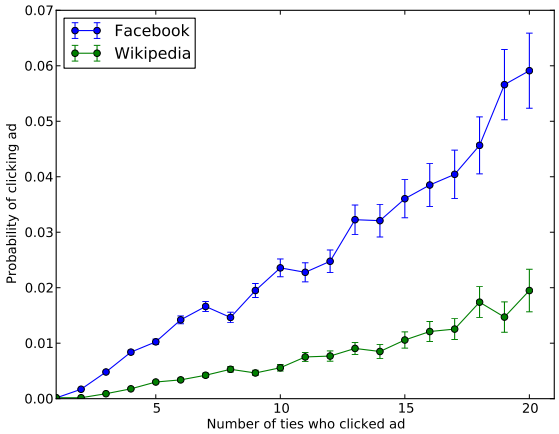

Figure 2: Probability of clicking on an ad given that k ties clicked on it.

5. EXPERIMENTS

We conducted a large-scale experiment over several months of browsing activity to learn the parameters of our model. Our experiment is directed towards predicting the distribution of ad category clicks. The key findings are (1) Affiliation networks do indeed improve our ability to predict the future distribution of ad clicks. We find this intriguing given that we are not using any explicit user network such as Facebook or Twitter, but rather completely synthesized covisitation graphs. (2) A user's own personal history of ad clicks is surprisingly unhelpful in predicting the future distribution of ad clicks and (3) Both Wikipedia and Facebook covisitation networks are helpful to various degrees depending on the category of ads, although the Facebook network is more helpful on average than the Wikipedia network.

5.1 Data Preprocessing

As mentioned previously, our experiments are based on the browsing activity of consenting users. The browsing activity records pages visited by the user including clicks on search results, clicks on sponsored search ads, and also visits to pages outside of the search results. We construct two covisitation networks based on two months of full browsing history (12/2011, 01/2012). The vertices of the graph G_W (resp., G_F) are users who visited Wikipedia (resp., Facebook) pages visited by at least one more user in the set. The weight of the edge (u, v) in G_W (resp., G_F) is the number of Wikipedia (resp., Facebook) pages that both u and v visited. The probability of walking from u to v in the random walk is the weight of (u, v) divided by the total weight of all edges emanating from u.

Ideally, we would compute the stationary distribution of the matrix corresponding to the random walk in G_W and G_F. However, because the matrix grows very large very quickly, we only computed a two-step random walk with restarts. For computational reasons, we compute this two-step walk when removing vertices of high degree, set to 200 or larger in our experiment.

For each user in the network, we also gather ad clicks. The ads considered in this study are sponsored search only, i.e., ads that surface as a result of a search query on any major search engine. As a result, we obtain a broad set of ad clicks that are not biased from the ranking, population demographics, or other characteristics of a specific search engine. Since our goal is to learn the parameters of our model for large classes of activities, we categorized the ads into the Open Directory Project (ODP) taxonomy using the categorizer proposed in [5]. The categorizer uses the text of the ad to determine the probability that a URL belongs to a category. While in principle every ad should have a cate-

gory, in reality, we do not obtain categories for all ads, e.g., the text of the ad may not be available or the categorizer may not produce any categories with sufficiently high confidence. Roughly 20% of the ads are categorized in our data. While the effectiveness of the categorizer does influence our results, the coverage is certainly non-trivial. We categorized the ads into 189 second level categories in ODP.

Our experiment is based on data collected over four time periods $t_1 < t_2 < t_3 < t_4$. During time period t_1 we create the covisitation graph. During time periods t_2, t_3 and t_4 we create the distribution of clicks over ad categories for each user. The distribution from t_2 comprises the user's history, while the distribution from t_3 is used to learn the parameters of the model, and the distribution from t_4 is used to evaluate the performance of the learned model. We separate t_1 from t_2 to be certain that the ad clicks are not somehow influenced by visits to Wikipedia or Facebook pages. The remaining time periods are separated for more customary reasons, i.e., separating training data from testing data. While t_1 is based on two months of data, t_2, t_3, t_4 are each based on one month of data that consecutively follow t_1, i.e., t_2 is 02/2012, t_3 is 03/2012 and t_4 is 04/2012.

After restricting the user population to those who clicked ads that we could categorize and those with connections in both G_W and G_F, we were left with 0.5 million users. To learn the parameters of the model, we exhaustively tried all $(\psi, \alpha, \gamma_W, \gamma_F)$ values such that $\psi + \alpha + \gamma_W + \gamma_F = 1$ on an ϵ-net of granularity 0.01 and output the choice that minimized training error, as indicated in section 3.

5.2 Findings

The output of our learning procedure is one set of the four parameters $(\psi, \alpha, \gamma_W, \gamma_F)$ for each ad category. In the findings below, we provide perspective into these values.

We begin by asking if the covisitation networks are helpful for predicting future ad clicks. In order to answer this question, we learned parameters for each ad category when only α and ψ were possible parameters, i.e., no networks were allowed by setting $\gamma_W = \gamma_F = 0$, and compared that to learning when both networks were allowed. For each category, we compare the the two models by taking the ratio r_c of the test error with the networks over the test error without the networks. The average test error with the networks is 0.00109. The average value of r_c is 0.8, indicating a 20% improvement in error. Furthermore, through a t-test we confirm that distribution of r_c values has mean significantly lower than 1 (p-val < 0.01).

Usually, the best predictor of what a user will do in the future is what they have done in the past. However, since users do not click many ads, their history is often too sparse to have any predictive power in this case. The parameters learned by our model confirm this. For 93% of the categories, the weight placed on a user's history ψ is less than 0.05.

Next, we compare the distribution of $\alpha + \psi$ with the distribution of $\gamma_W + \gamma_F$. Our goal is to understand, across all ad categories, how the importance placed on network compares with the importance placed on everyone's history. Figure 3 shows the CDF of the value $\alpha + \psi$ for all the categories. Note

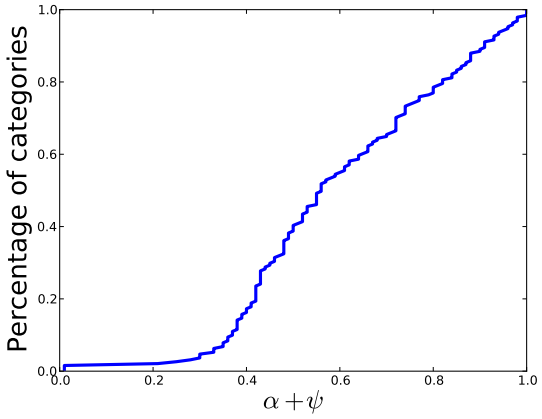

Figure 3: Cumulative distribution function of the value $\alpha + \psi$ for all of the second-level categories.

Category	γ_W	γ_F
Recreation/Aviation	0	0
Reference/Knowledge Management	0	0
Society/Genealogy	0	0
Arts/Animation	0	0
Society/Death	0.02	0
Society/Folklore	0.02	0
Arts/Online Writing	0.01	0.01
Reference/Education	0.01	0.02
Computers/Programming	0	0.03

Table 2: Categories where the sum in $\gamma_F + \gamma_W$ is minimized

that $\alpha + \psi = 1 - \gamma_W + \gamma_F$, hence one can easily infer the distribution of $\gamma_W + \gamma_F$ from the CDF of $\alpha + \psi$. We observe that there are quite a few categories where the network does not play a large role, i.e. $\alpha + \psi$ is very large. That is, for many of the categories, most of the predictive value comes from the users' own history combined with the general population's history. We provide some examples of such categories in Table 2. For these categories, ties do not provide much information to advertisers. Instead they should rely on the user's personal history, and the global population behavior.

On the other hand, for a significant percent of the categories (about 40%), the importance of the networks is larger than the importance of the users' own history and the population's history. That is, for about 40% of the categories, $\alpha + \psi$ is less than 0.5.

Regarding the relative importance of the two networks, a scatter plot of γ_W versus γ_F is shown in Figure 4. We find that that for 80% of the categories, $\gamma_F > \gamma_W$. This suggests that the social network is, on average, more predictive than the topical network. However, for most of the categories, both γ_W and γ_F tend to be relatively large, implying that both networks contribute a significant amount. For instance, for 74% of the categories both γ_W and γ_F are larger than 0.05.

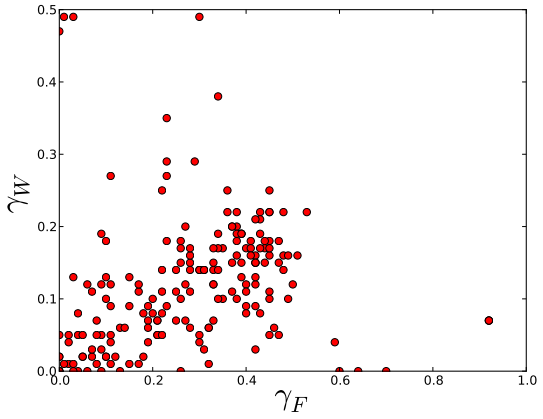

Figure 4: Scatter plot of γ_W versus γ_F.

Finally, we show some specific categories with interesting parameters. First, we show a set of categories where the difference between γ_F and γ_W is large. If $(\gamma_F - \gamma_W)$ is large, then the Facebook co-visitation network is more accurate in predicting future ad clicking activity than the Wikipedia network, and vice-versa when $(\gamma_W - \gamma_F)$ is large. Table 3 shows categories where $\gamma_F - \gamma_W$ is maximized followed by categories where $\gamma_W - \gamma_F$ is maximized. While we do not aim to explain why certain categories benefit more from one network than the other, we note that there are some interesting trends. For example, Business categories seem to mostly benefit from the Facebook covisitation network and Arts categories seem to favor mostly from the Wikipedia network. It is possible that the *social* aspect of the Facebook network and the *topical* aspect of the Wikipedia network may explain these differences. However, we leave any hypotheses that may explain these difference as future work. Identifying, which categories benefit from each network could be useful for advertisers. For example, advertisers of categories that benefit mostly from Facebook may place greater weight on a person's Facebook covisitors than Wikipedia covisitors.

Category	γ_W	γ_F
Business/Information Technology	0.07	0.92
Business/Marketing & Advertising	0.07	0.92
Computers/Computer Science	0.07	0.92
Sports/Golf	0	0.7
Science/Math	0	0.6
Business/Accounting	0.04	0.59
Society/Holidays	0.05	0.47
Shopping/Pets	0.06	0.46
Kids and Teens/People and Society	0.05	0.45
Business/Aerospace and Defense	0.03	0.42
Arts/Performing Arts	0.12	0.06
Sports/Basketball	0.18	0.1
Health/Pharmacy	0.13	0.03
Society/Government	0.19	0.09
Games/Card Games	0.35	0.23
Kids & Teens/Health	0.27	0.11
Arts/Music	0.49	0.3
Arts/Architecture	0.49	0.03
Recreation/Humor	0.47	0
Society/Religion & Spirituality	0.49	0.01

Table 3: Categories where the difference in $\gamma_F - \gamma_W$ is maximized as well as where $\gamma_W - \gamma_F$ is maximized.

6. DISCUSSION AND FUTURE WORK

With many networks in existence today, and many more on the horizon, advertisers are challenged to find ways to use these networks in beneficial ways. Our work abstracts this question into a model that predicts user behavior via the behavior of others: be it themselves, everyone, or their ties in multiple networks. Our findings enable us to understand the relative usefulness of ties in various networks for predicting future click activity.

To the best of our knowledge, we are the first to study the use of multiple, heterogeneous networks for advertising, and we believe that we are just scratching the surface of a broader and deeper problem. Any system that makes use of the network should take into account that users are part of multiple networks (both explicit and implicit) that complement, amplify, or contradict each other. It is an open question how one can make the most of these multiple networks. This question goes beyond advertising and our methodology is not restricted to ads. In principle, it could be applied to any kind of activity: search queries, browsing clicks, online purchases. However, we believe that more techniques and tools will be required to address new challenges posed by the existence of multiple networks.

In our work, we experimented with affiliation networks created by covisitations of users to common web pages. We defined our networks by common visits to pages from the Facebook and Wikipedia domains. We observed that the ties in these networks provide useful signals for advertising, and we can identify cases where one network is more helpful than the other. Given the implicit way these networks are constructed, this is a striking finding.

Furthermore, our view of the covisitation domain (Facebook and Wikipedia in our case) as giving a type to the link between two individuals (social and topical in our case) is powerful and can form the basis for studying other networks that explore different types of relationships. For example, Twitter covisitations are interesting in that users follow their friends as well as people with similar topical interests. LinkedIn covisitations could induce a professional covisitation network. Yelp, Amazon, or IMDB covisitation networks can be used for more specialized purposes. More generally, a larger scale experiment could be implemented where every website generates a network.

From an advertising perspective, there is more to be done. We did not explore the problem of predicting ad click-through rate in this paper. Ad click-through rate is a well-studied problem involving many features. A full-blown advertising experiment determining whether and how signals from ties improve ad click-through rate above and beyond existing features is still an open question. Finally, our work opens the possibility for new bidding systems where advertisers can bid on users who have ties in various affiliation networks. Exploring the economic value of ties in these networks and

pricing strategies for bidding on them is an interesting subject for future work.

7. REFERENCES

[1] L. A. Adamic and N. Glance. The political blogosphere and the 2004 U.S. election: divided they blog. In *LinkKDD*, pages 36–43, 2005.

[2] T. Anastasakos, D. Hillard, S. Kshetramade, and H. Raghavan. A collaborative filtering approach to ad recommendation using the query-ad click graph. In *Proceedings of the 18th ACM conference on Information and knowledge management*, CIKM '09, pages 1927–1930, New York, NY, USA, 2009. ACM.

[3] S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *PNAS*, 106(51):21544–21549, 2009.

[4] E. Bakshy, D. Eckles, R. Yan, and I. Rosenn. Social influence in social advertising: Evidence from field experiments. *CoRR*, abs/1206.4327, 2012.

[5] P. Bennett, K. Svore, and S. Dumais. Classification-enhanced ranking. In *WWW '10*, pages 111–120, 2010.

[6] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *KDD*, pages 160–168, 2008.

[7] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *PNAS*, (52):22436–22441, 2010.

[8] N. Eagle, A. Pentland, and D. Lazer. Inferring Social Network Structure using Mobile Phone Data. *PNAS*, 2007.

[9] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Commun. ACM*, 35(12):61–70, Dec. 1992.

[10] S. Hill, F. Provost, and C. Volinsky. Network-Based Marketing: Identifying Likely Adopters via Consumer Networks. *Statistical Science*, 21(2):256–276, May 2006.

[11] D. B. Kandel. Homophily, selection, and socialization in adolescent friendships. 84(2):427–436, Sept. 1978.

[12] G. Kossinets and D. Watts. Origins of homophily in an evolving social network. 115(2):405–50, Sept. 2009.

[13] G. Kossinets and D. J. Watts. Empirical Analysis of an Evolving Social Network. *Science*, 311(5757):88–90, Jan. 2006.

[14] S. Lattanzi and D. Sivakumar. Affiliation networks. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 427–434, New York, NY, USA, 2009. ACM.

[15] P. Lazarsfeld and R. K. Merton. Friendship as a social process: A substantive and methodological analysis. In M. Berger, T. Abel, and C. H. Page, editors, *Freedom and Control in Modern Society*, pages 18–66. 1954.

[16] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, Jan. 2003.

[17] K. Liu and L. Tang. Large-scale behavioral targeting with a social twist. In *CIKM*, pages 1815–1824, 2011.

[18] H. Ma, I. King, and M. R. Lyu. Learning to recommend with explicit and implicit social relations.

[19] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*.

[20] M. McPherson, L. S. Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444, 2001.

[21] J. Moody. Race, school integration, and friendship segregation in america. 107(3):679–716, Nov. 2001.

[22] R. Panigrahy, M. Najork, and Y. Xie. How user behavior is related to social affinity. In *WSDM*, pages 713–722, 2012.

[23] P. Papadimitriou, P. Krishnamurthy, R. Lewis, D. Reiley, and H. Garcia-Molina. Display advertising impact: Search lift and social influence. In *KDD*, 2011.

[24] F. Provost, B. Dalessandro, R. Hook, X. Zhang, and A. Murray. Audience selection for on-line brand advertising: privacy-friendly social network targeting. In *KDD*, pages 707–716, 2009.

[25] P. Resnick and H. R. Varian. Recommender systems. *Commun. ACM*, 40(3):56–58, Mar. 1997.

[26] D. M. Romero, C. Tan, and J. Ugander. On the interplay between social and topical structure. In *ICWSM*, 2013.

[27] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, WWW '01, pages 285–295, New York, NY, USA, 2001. ACM.

[28] P. Singla and M. Richardson. Yes, there is a correlation - from social networks to personal behavior on the web. In *WWW*, pages 655–664, 2008.