On-line Brand Advertising using Social Networks based on User-generated Content

Foster Provost\textsuperscript{1}, Brian Dalessandro\textsuperscript{2}, Rodney Hook\textsuperscript{3}, Xiaohan Zhang\textsuperscript{1}

\textsuperscript{1}Stern School of Business, NYU
\textsuperscript{2}Media6degrees
\textsuperscript{3}Coriolis Ventures

Extended Abstract

This paper is about brand advertising on the web, and about a method for taking advantage of user-generated content on social-networking sites (and beyond) to improve on-line advertising.

Different from most sponsored search advertising and click-driven display advertising, on-line brand advertising is less straightforward to assess, and perhaps for that reason has received much less attention in the academic literature and in practice. Brand advertising generally is difficult to assess; however, the on-line environment provides unique opportunities, because as we describe below certain brand actions become visible and measurable on-line. Furthermore, brand advertising has a huge opportunity for growth. Even though the vast majority of on-line ads are display ads, sponsored search advertising is responsible for the majority of advertising revenue (and profits). Well-designed brand advertising may be much more appropriate for visits to nytimes.com or myspace.com, since the user has not come for the purpose of clicking on a returned link. Improving brand advertising on-line may also have a substantial impact on social welfare: the access to a large amount of free content on-line is due largely to sponsorship by advertising. ComScore [comScore 2008] recently reported a clear correlation between seeing brand advertising and increasing purchases, both on-line and off-line, and well into the future (beyond the reach of current view-through conversion measurement technology).

The first contribution of this paper is to address one key question for brand advertisers: How can on-line brand audiences be assessed? We present a new framework for assessing on-line brand advertising audiences. The framework is not meant to supplant traditional brand-advertising evaluations, but to complement them with an on-line-specific evaluation. The idea is as follows. For some browsers, on-line advertising networks can observe certain brand actions, such as visiting a brand’s web site, buying a brand’s product on-line, entering a brand-specific search, etc. We adapt a machine-learning-style hold-out evaluation to brand actions: for any technique that claims to identify a “good” audience for a brand, we can compare the density of brand actors in an identified subset of the population, and compare it to the population as a whole (or a competing technique). Our guiding assumption is that a “better” audience for a brand will exhibit a higher density of brand actors. This is specifically in contrast to an evaluation based on the likelihood of clicking or of converting directly.

The second contribution of this study is to provide a new answer to a second key question for brand advertisers: How can we find a good audience? Existing answers include an adaptation of the traditional, vertical television/magazine model: associate brand advertisements with content. Our method takes advantage of the ability on-line to do horizontal targeting, identifying browsers of interest and targeting them across the web.

Specifically, we present a privacy-friendly method of targeting social network neighbors with brand advertising. One of the most influential results of social theory is the notion of homophily [McPherson et al. 2001], that similar people tend to have social relationships. This has been shown to be directly useful for targeted direct marketing: Hill et al. [2006] show that social-network neighbors of existing customers are several times more likely to respond to an offer for a telecommunications product than consumers who do not have an existing customer as a social-network neighbor.
Space limits here prevent a detailed description of the method, but in a nutshell: The first step is to use user-generated content to define an anonymous, (quasi-) social network among browsers, based on data from a large ad network. Ad networks serve massive numbers of ads to massive numbers of browsers, and keep track of what browsers visit what content. When this content is user pages from a social-networking site, the page visitations create a quasi-social network among users, induced from the bipartite affinity network between users and UGC. Frequencies of visitation become strengths for the individual network links. It is likely, although we do not show this, that true friends have strong connections in this quasi-social network. We anonymize all browsers and all content, so we do not know who are true friends and who just share a strong affinity but don’t actually know each other (the distinction is not important for our purposes).

The second step is to use this social network to identify brand audiences. Define the brand action of interest to be visiting a brand’s web site; to be concrete, but completely hypothetical, let’s call this the iPhone site. We define the audience of interest to be the (close) social network neighbors of brand actors – say, prior iPhone site visitors. Then, using the holdout evaluation methodology describe above, we can assess the (future) brand action density of the audience. Specifically, we take the brand actors from one time period to define the network-neighbor audience (not including the prior brand actors), and then use a future time period to see what percentage of the audience exhibited brand activity. A good audience will have a brand-action density significantly larger than the overall population (or equivalently, a randomly chosen subset).

Our results verify strongly that the audience defined by the social-network neighbors of brand actors indeed does have a substantially higher density of future brand actors than the population as a whole. These results hold for a large number of well-known brands—to a greater or lesser extent depending on the brand and seemingly how “tribal” it is, but the general result holds for every brand we have experimented with. Furthermore, the closer a neighbor is in this social network [Bernstein et al. 2003] to a brand actor, the higher is the likelihood that it will exhibit brand activity. The figures below show some sample results for one of the best-known brands.

The status of the manuscript is that we are writing it now. The results are strong and very exciting, but have not yet appeared anywhere. There is a lot of future work to be done. For example, can more sophisticated measures of network proximity find much higher-lift audiences?

![Comparison of Ranking Performance over Random](image)

Figure 1. Network neighbor audiences show 2-3x lift in brand-actor density over the population as a whole, based on those 10% and 25% neighbors closest to a prior brand actor, for several proximity measures (maximum cosine similarity, the maximum linkage to an action taker, the total number of links to action takers, and the minimum Euclidean distance.)
Figure 2. For the best similarity measure, closest network neighbors show lifts in excess of 3x.

Figure 3. ROC curve for network-neighbor audience ranked by proximity to brand actor in the social network, where proximity is calculated as the (max) cosine similarity.

References

