



Understanding Your Traffic Sources

BY JOSH SCHWARTZ, CHARTBEAT HEAD DATA SCIENTIST

A note from the Chartbeat:

Understanding Your Traffic Sources was originally a 5-part series shared on the Chartbeat blog. Due to the series's overwhelming popularity, we combined the posts into this chock-full-of-data report so that folks like you can read it through more easily – and even share it with friends more easily, too.

This report is designed to give you a solid foundation of knowledge about traffic sources and how different types of site visitors behave differently. The data indicates that some visitors are more likely to return to your site than others, and thus some types of traffic are worth paying more attention to more than others, depending on the source. It's our hope that this kind of information can help you take a more strategic approach to building your site's loyal audience.

Happy reading!

Part 1: Introduction

It goes without saying that thoroughly understanding how traffic arrives at your site is critical for audience development efforts — certain traffic sources are ideal for reaching new readers, others may put your content in front of your existing audience and encourage them to visit more often, and still others may send traffic that will never be seen again.

This section kicks off a five-part series that takes a detailed look at how audience behavior differs across traffic sources. Over the coming posts, we'll go through homepage/direct traffic, traffic from external links and social sources, and search traffic. To start, though, I want to talk about how we classify referrers and define some basic metrics through which we compare their traffic.

KNOW YOUR REFERRER TYPES

At Chartbeat, we divide visitors' traffic sources into four buckets:

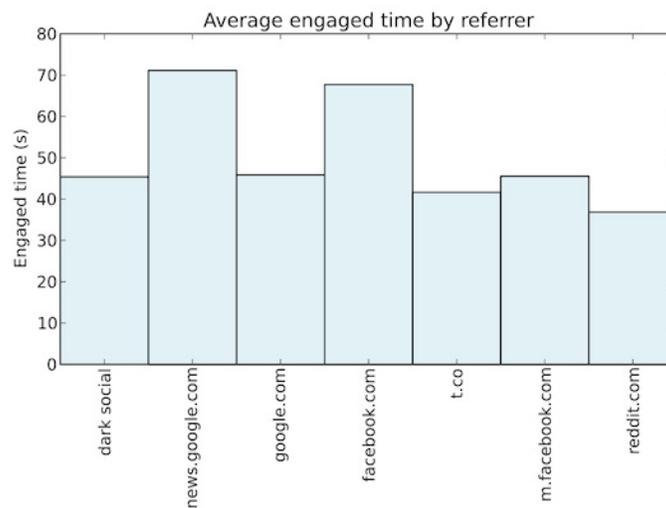
- **Direct:** people who get to your site by entering your site's name in a browser, typically landing on your homepage
- **Social:** people who arrive at your site from links provided by friends or others they follow on social media (e.g. Twitter, Facebook, Pinterest) or on email/IM
- **External:** those who arrive via links from external sites (e.g. Drudge Report, Google News)
- **Search:** visitors who come from search engines

The distribution of traffic across these buckets varies widely between sites — we see some sites with over 90% of traffic coming direct and others with over 90% of traffic from search. There’s no correct breakdown, but it’s important to keep your goals in mind when considering where to optimize: external sources typically send the most new visitors, while your most loyal audience is likely to come directly to your homepage.

MEASURING AUDIENCE ENGAGEMENT

As we’ve talked about [before](#), Engaged Time is a good measure of fit between your content and your audience. So, one of the first questions we ask when trying to assess the quality of a given traffic source for a site is: how visitors are engaging with the pages they land on?

Below we see visitors’ Average Engaged Time on article pages when coming from seven of the most common referrers:



Notably, visitors from Facebook and Google News spend more than 50% more time than those who come from other sources (Tweet this fact). One possible explanation, which we’ll delve into more in Part 3, is that these are the only two sources on the list that typically show the first lines of a story when they present a link, so visitors from these sources are more likely to be committed to actually reading the story they end up on.

UNDERSTANDING AUDIENCE RETENTION

Of course, we don’t simply care about how people behave when they land on your site — we want to know whether people who come to a site will choose to come back again. This turns out to be a question whose answer varies dramatically by traffic source. Below is a figure showing the fraction of visitors from a given referrer (the seven referrers from above, plus those who come direct to a site’s homepage) who come back sometime in the next week.

Rows of this figure represent a visitor's traffic source on one visit and columns represent where they come from the next time. Redder cells in each row represent combinations that occur more frequently, while whiter cells represent less frequent combinations.

The thing that stands out most in this figure is that people who visit your site from a given referrer are far more likely to come to your site again from the same referrer than they are to come from any other source. We term this phenomenon referrer loyalty, and it tells us that we can't rely on traffic from an outside source returning without a steady stream of links on that source.

Cells in the direct column that are bright represent what is perhaps an ideal traffic relationship — referrers that send traffic that is likely to come directly to your site the next time they visit.

MORE TO COME

This section, I hope, raised more questions than it answered. Why do rates of engagement and retention vary so much by traffic source? And, more importantly, what can we do to affect these rates over time? We'll start answering these questions in the following sections by looking at your most core audience, your homepage direct traffic.

Part 2: Direct Traffic

DEFINING "DIRECT"

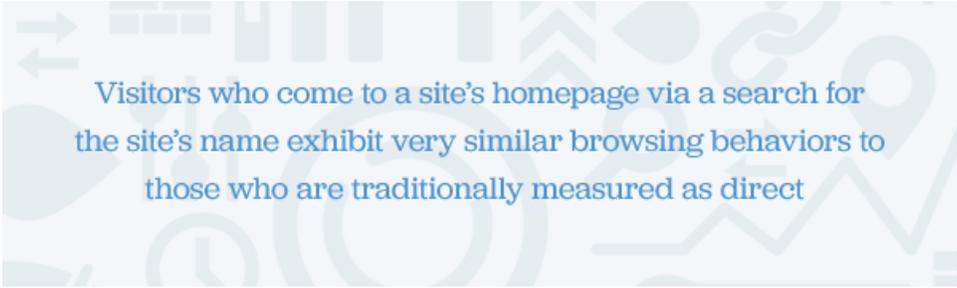
When we say "direct traffic", we're referring to those visitors who come to your site's homepage deliberately, as opposed than those who come to your site via a link to a particular news story.

Typically, people measure direct traffic by looking at the HTTP referrer for each page view: when you visit a page, your browser usually records the referrer that sent you to the page; if no referrer is recorded then it should mean that you either typed in the name of the page or visited via a bookmark.

There's a bit of subtlety to that story — many sites that use HTTPS (email sites and some apps are notable examples) don't send HTTP referrers and are frequently misclassified as direct. The phenomenon of external traffic masquerading as direct has been termed "dark social" and discussed often on the web (e.g. Alexis Madrigal's defining piece for [The Atlantic](#) and [this post](#) on the Chartbeat blog). In short, it's very unlikely that a person would truly come directly to a news article, that they'd actually have typed in the URL for that particular article, and it's much more likely that this traffic is coming from a source that doesn't send HTTP referrer information.

So, a certain amount of traffic is often over-attributed as direct when it's actually coming from external traffic sources. But what about cases where traffic is misattributed the other way, where direct traffic is classified as external?

I'd argue that this is the case with what's often called branded search, but is better termed direct search. Look at search data in your favorite analytics dashboard and you'll likely see that all of your top search terms are words related to your domain. If someone searches "Chartbeat" and clicks on a link to our homepage, I'd argue that that's essentially equivalent in spirit to them typing chartbeat.com into a browser — that person intended to visit our site. Indeed, we see that visitors who come to a site's homepage via a search for the site's name exhibit very similar browsing behaviors to those who are traditionally measured as direct, and in fact these visitors are highly likely to come back "truly" direct the next time they visit.



Visitors who come to a site's homepage via a search for the site's name exhibit very similar browsing behaviors to those who are traditionally measured as direct

DIRECT TRAFFIC'S QUALITY

Since direct visitors typically land on home pages and section front pages — pages that link to other content rather than providing content themselves — it doesn't make sense to look at engagement on the landing pages themselves because these landing pages are not designed to produce engagement. Rather, it makes more sense to ask how deeply people read upon arrival and how often they come back. From both of these perspectives, direct traffic far outperforms the norm: on a typical site, we see that visitors who come directly come 3-4 times per week and view 9 articles across the week — a drastic difference compared to the average visitor, who visits 1-2 times per week and views 2-3 pages.

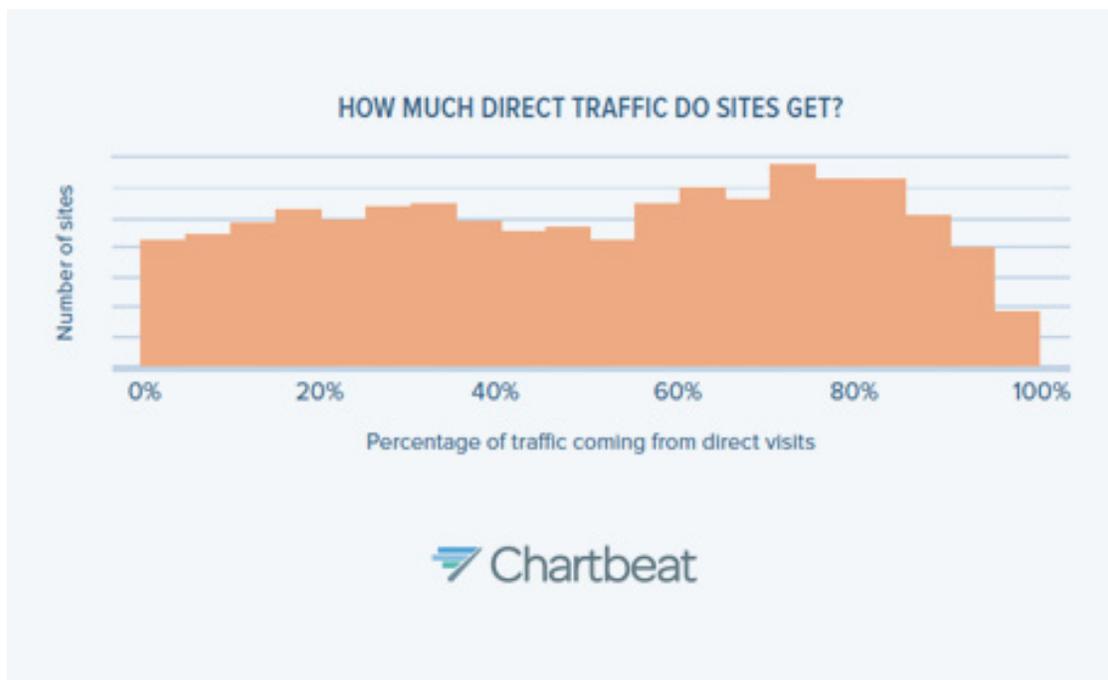


On a typical site, we see that visitors who come directly visit 3-4 times per week and view 9 articles across the week - a drastic difference compared to the average visitor, who visits 1-2 times per week and views 2-3 pages

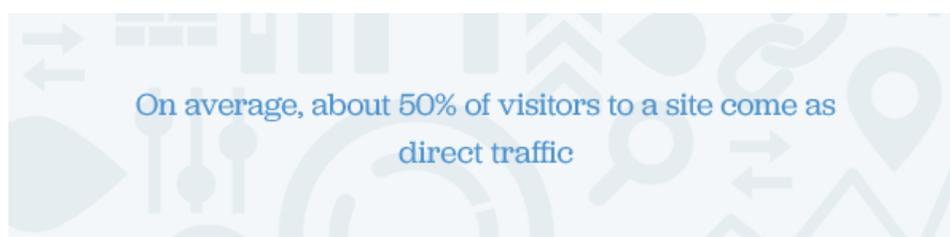
In that sense, we don't need to think about whether direct visitors are coming back or whether they're reading stories, we want to think about how frequently they're coming back and how deeply they're reading. This is, perhaps, best expressed in terms of the number of daily direct visitors versus the number of weekly or monthly direct visitors — we'd like to push this ratio as high as possible, so those who come directly are doing so every day.

HOW MUCH SHOULD YOU GET?

If direct traffic represents your most loyal audience, it's natural to ask how much direct traffic you should hope to get. The answer, unfortunately, is that there's a very wide spread. The figure below shows the breakdown of how much direct traffic sites get.



On average, about 50% of visitors to a site come as direct traffic, but the distribution has an interesting bimodal shape — there is a set of sites for which about 30% of visitors come direct and another set for which the direct rate is about 75%.



If your site is in the former camp, increasing the fraction of direct traffic should be a top priority. That means outreaching to the sources whose readers are most likely to convert to direct (see here for a figure analyzing this statistic across the web), increasing the branding on your site to make sure that side-door traffic knows what site they're on, and publishing more content to encourage visitors to come back more frequently.

A CAVEAT AND CONCLUSIONS

The picture that we've presented of direct traffic so far is a positive one — direct visitors visit often and account for a disproportionately high amount of the total content consumption on a site. There's one caveat, though, which is that direct traffic is almost by definition composed of people who are not new to your site. Those sites with the highest direct traffic have the lowest rates of new visitors. If that issue rings close to home, the first place to look for new visitors is the social web, which will be the topic of the next post in this series. I hope you'll stay tuned.

Part 3: Social Traffic

Overall, about 26% of traffic we measure comes from social sources — Facebook, Twitter, and email, for example — making social the second most significant source of traffic, next to direct. In some sense, social traffic and direct traffic represent polar opposites: Visitors who arrive via your homepage are, critically, people who intended to visit your site specifically rather than a particular piece of content. Those who come from social sources may or may not know what site they're landing on — they're coming because of an article that's been recommended to them. That's a double-edged sword. On the one hand, social visitors are more likely than other visitors to actually read the pages they land on; on the other, they're also amongst the least likely to return to your site, and when they do they're very likely to only come via the same social channel.

Social is also categorically different than other sources of traffic because it's the only channel that's easily influenced — while converting visitors to come directly to your homepage is an art and affecting search engine placement leaves much to chance, we can actively choose which articles we put on social media and when to provide those links.

DEMOGRAPHICS

Before we talk about evaluating social traffic, it's worth discussing what sort of visitors come from social sites and how they read. First off, social sources are a better than average source of new visitors: while an average of 31% of a site's traffic comes from new visitors (those who haven't visited in the past 30 days), an average of 41% of social visitors are new.



While an average of 31% of a site's traffic comes from new visitors, an average of 41% of social visitors are new.

Social traffic is also dramatically more mobile-based than all other traffic — an average of 25% of traffic is on mobile, but on many sites over 40% of social traffic is mobile. That should affect what stories you push to social media, and when you push them. We'll cover both of those topics below.

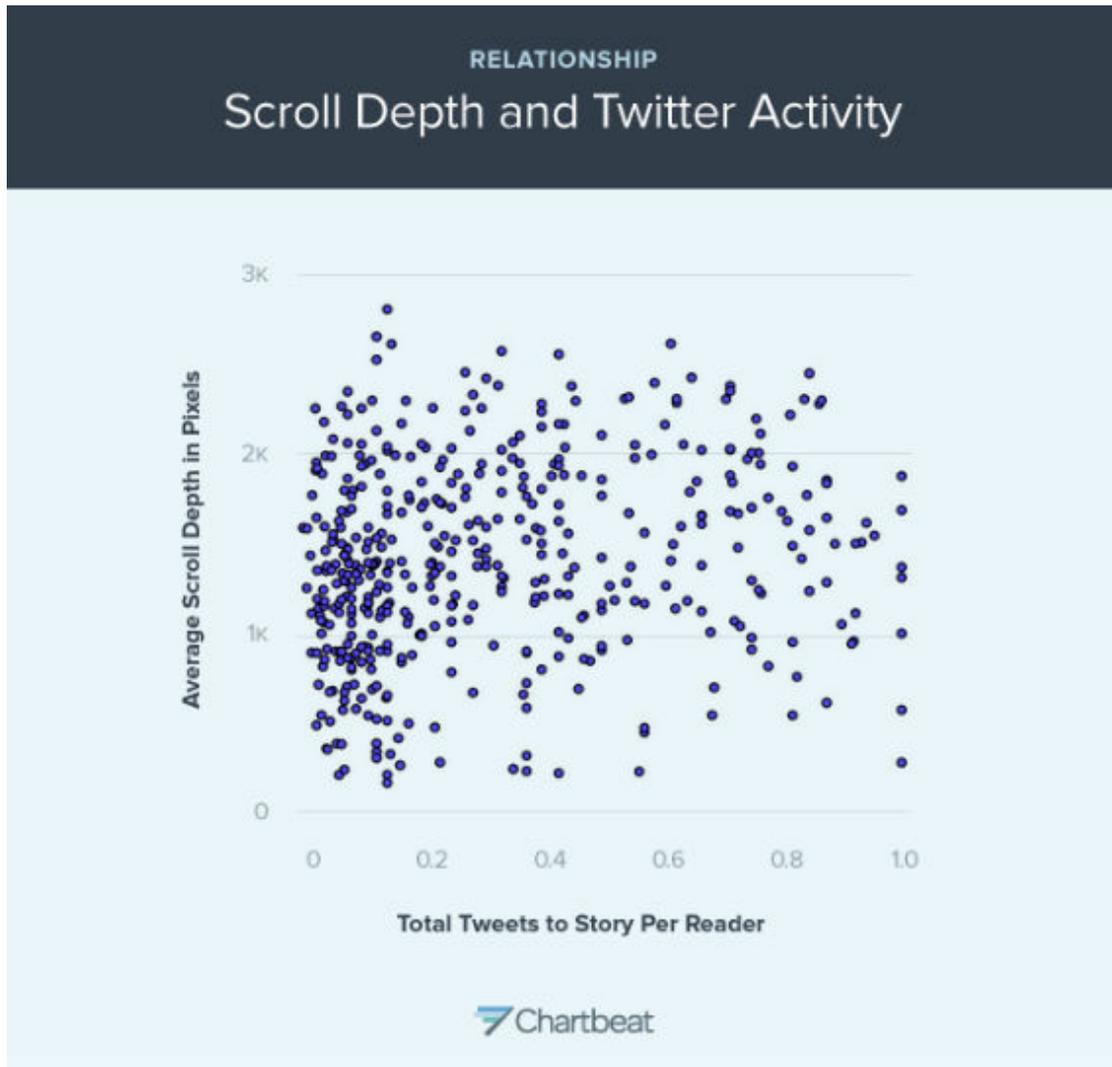


Social traffic is also dramatically more mobile-based than all other traffic - on many sites over 40% of social traffic is mobile.

SOCIAL ENGAGEMENT VERSUS ON-SITE ENGAGEMENT

People frequently take social media interactions as the de facto standard for “engagement” with a piece. The idea is that people who share a piece are likely to have enjoyed it. While there’s some kernel of truth here, our data suggests that there’s more to the engagement story than raw counts of tweets and likes.

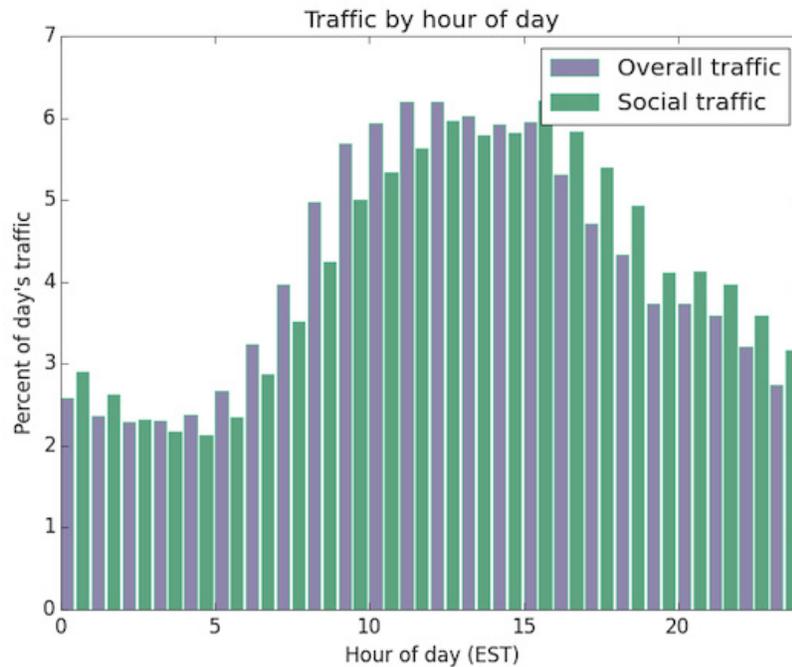
Take a look at the graph below, which was first presented in Slate:



This graph shows how fully people read an article (as measured by how far down the page they scrolled; all articles shown here were over 3000 pixels high), compared to how frequently they tweet about it. If the most engaging stories to read were the stories that were most likely to be shared, we’d expect this graph to look like a line. Instead, we see that there’s essentially no correlation between the two numbers. That doesn’t mean that social interactions are a bad way to measure engagement, but it does show that social engagement and on-site engagement are often different phenomena.

TIMING OF SOCIAL POSTS

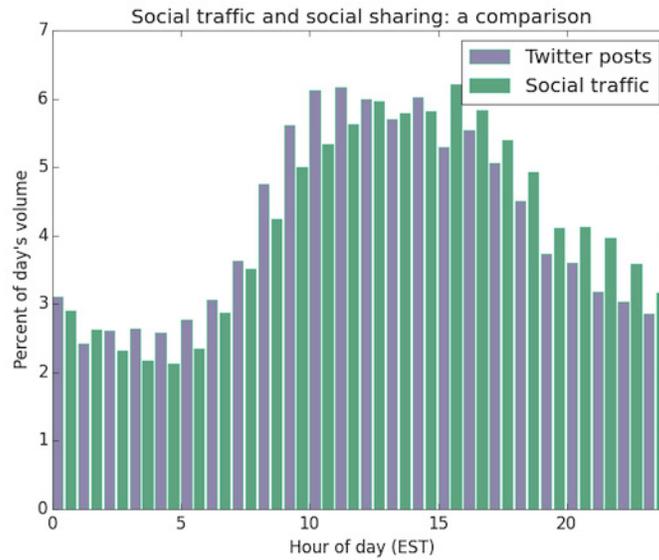
So, what makes for successful social content? There's been much written about how to write successful social posts — most recently, I read a [great study by Knight fellow Sonya Song](#) and its more concise [writeup](#) on Nieman Lab. It's beyond the scope of this post to tackle what content to put in your social posts, but one question we're frequently asked is what time of day is best for social sharing. Below is a chart showing how social traffic compares to overall traffic across for a set of sites (all of which are based in EST) across the past week.



Unsurprisingly, the shape of social traffic closely follows that of overall traffic, but it's notable that social traffic substantially underperforms overall traffic from about 5am to noon, and social substantially overperforms overall traffic from about 3pm until 1am. From the perspective of driving traffic to your site, it appears that late afternoon through night is the best time to reach your readers on social media and get them to click through to your site.

Social traffic substantially underperforms overall traffic from about 5am to noon, and social substantially overperforms overall traffic from about 3pm until 1am.

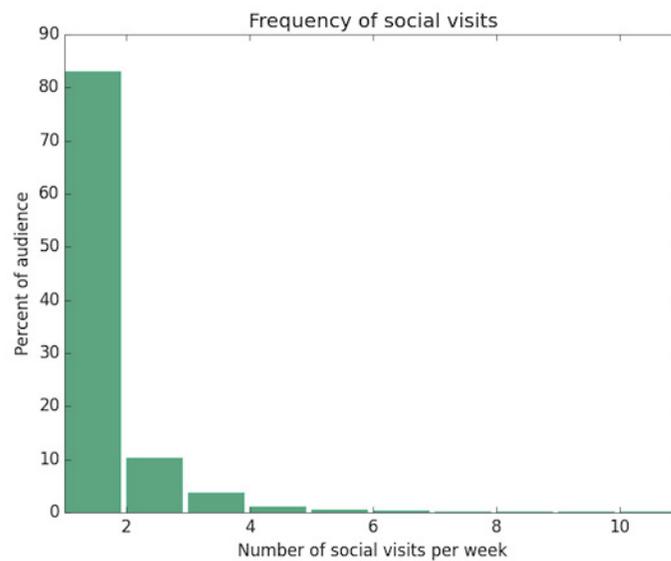
Interestingly, this trend appears to be true despite people's best efforts to the contrary. Below, we see a graph of how frequently these sites posted to Twitter, compared to their social traffic.



Posting to Twitter is strong all morning and reaches its peak just before noon, even though traffic from social is actually its strongest later in the day.

RETURN FREQUENCY

While we're discussing timing, it's worth noting that visitors who come to a site from social sources do so an average of 1.5 times per week. Below we see the distribution of how many times a visitor comes from social sources across a week.



About 82% of visitors who come from social only come once, but there's a long tail of people who come two or more times.

As mentioned above, almost 80% of visitors who come to your site from a social source will only come to your site via that source. That figure is particularly bad for visitors from Twitter, of which only about 16% will return to your site directly. These are fairly significant numbers to consider as you decide where to invest time and resources into developing your audience.



Almost 80% of visitors who come to
your site from a social source will only
come to your site via that source.

CONCLUSION

This post is only a fraction of what can be said about social media — entire companies exist to help optimize social strategy — but I hope it started you thinking about how social sharing relates to your site's overall traffic.

Part 4: External Traffic

*this section refers to data collected in November 2013

Now we're going to be talking about traffic from external links — those unpredictable pickups from sites across the internet. Unsurprisingly, external referrers provides sites' greatest volume of new visitors. They also provide sites' greatest challenge in terms of generating actual engagement.

TYPES OF EXTERNAL TRAFFIC

Broadly, we can divide external referrals into two camps: huge aggregators that send large volumes of traffic and incidental links from across the web. Let's go through each in turn.

In any discussion of external traffic sources, Google News deserves its own special treatment. As opposed to the majority of traffic sources, where links are human-curated, Google News pickups are algorithmically generated. That means that while most sites can probably only expect a few pickups from major referrers on a given day, many sites have hundreds of articles linked on Google News — over the last two weeks alone, over 616,000 distinct pages across our network received traffic from Google News.

Because of its volume of links, Google News is a significant and consistent driver of traffic — you're not going to get new pickups from most sites every day, but you can reasonably expect to get daily links from Google News. On the other hand, while a pickup from many external sites might presage a cascade of links from across the web, Google News pickups are not necessarily predictive of broader trends.

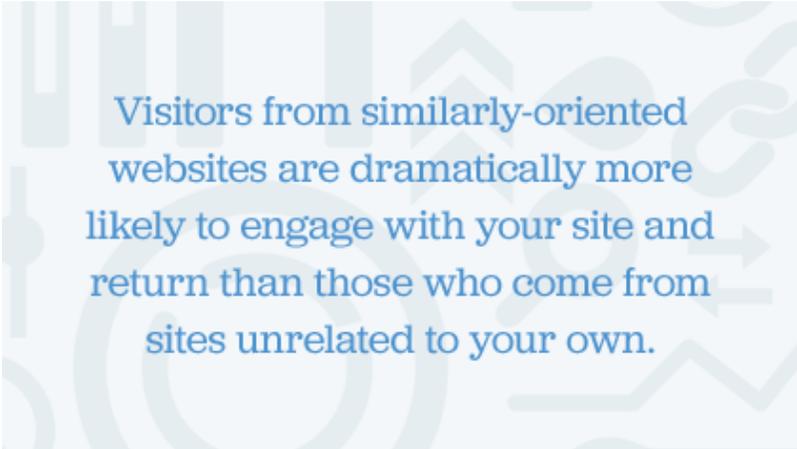
People who read Google News tend to do so frequently, which means that visitors from Google News come back substantially more frequently than average. On a typical site, over a quarter of visitors from Google News return in the next week. Very few of those Google News visitors who do return, though, come back to a site directly — fewer than 15%. That means that to attract these users back you have to concentrate on receiving a regular supply of links from Google News.



On a typical site, over a quarter of visitors from Google News return in the next week. Very few of those Google News visitors who do return, though, come back to a site directly – fewer than 15%.

Perhaps the best example of a non-automated site that sends massive amounts of traffic is Drudge Report. Just over 2000 pages had traffic from Drudge over the past two weeks, though the total volume of traffic sent was roughly comparable to that of Google News. Visitors from Drudge rarely read more than one page in their visit and are exceptionally unlikely to return to a site — fewer than 15% of Drudge visitors to a typical site return, and fewer than 15% of returners come back directly. Drudge is perhaps the most significant example, but we see similar behavior across most aggregators. Indeed, large social sites like Reddit send traffic that's typically even less likely to return to your site than that from Drudge.

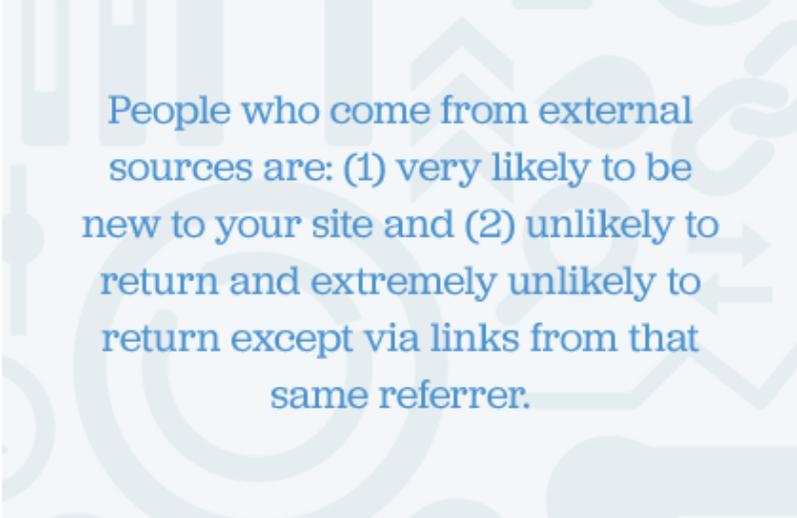
Beyond the largest sources of external traffic, there's a long tail containing all of the incidental links that occur. These links have such huge variation in traffic quality that it's difficult to sum them up. The best guiding principle we see is that, unsurprisingly, visitors from similarly-oriented websites are dramatically more likely to engage with your site and return than those who come from sites unrelated to your own. Visitors on a left-leaning political site, for example, can be twice as likely to return when coming from another left-leaning site as opposed to a right-leaning one. That means it's always important to consider external traffic spikes in context — a pickup from a referrer that's likely to send high quality traffic might be worth doubling down on, whereas a pickup from an unrelated site might be best treated as a less significant event.



Visitors from similarly-oriented websites are dramatically more likely to engage with your site and return than those who come from sites unrelated to your own.

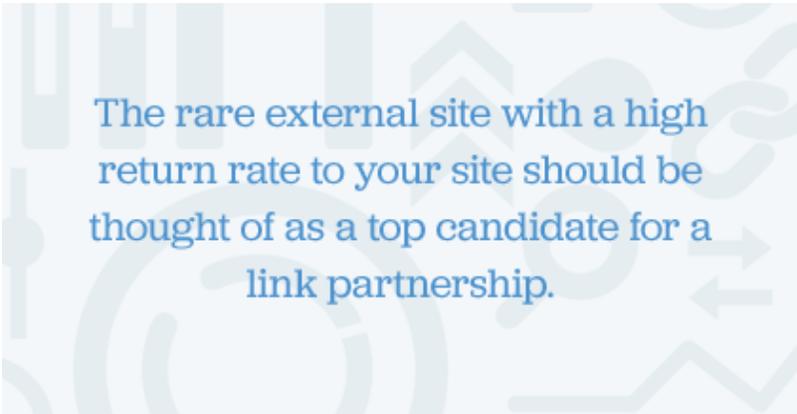
CONCERNS FOR EXTERNAL TRAFFIC

We've consistently seen that people who come from external sources are: (1) very likely to be new to your site and (2) unlikely to return and extremely unlikely to return except via links from the same referrer.



People who come from external sources are: (1) very likely to be new to your site and (2) unlikely to return and extremely unlikely to return except via links from that same referrer.

That means that you should interpret an external pickup very differently than a pickup on social or a page that's getting its traffic from the homepage. To get visitors from Twitter, for instance, to return you might push them to follow you on Twitter, but no such mechanism exists for external traffic. External links typically denote interest in a topic, as opposed to interest in your site in general. To that end, stories garnering the most external traffic should be thought of as inspirations for follow-up pieces. The rare external site with a high return rate should be thought of as a top candidate for a link partnership.



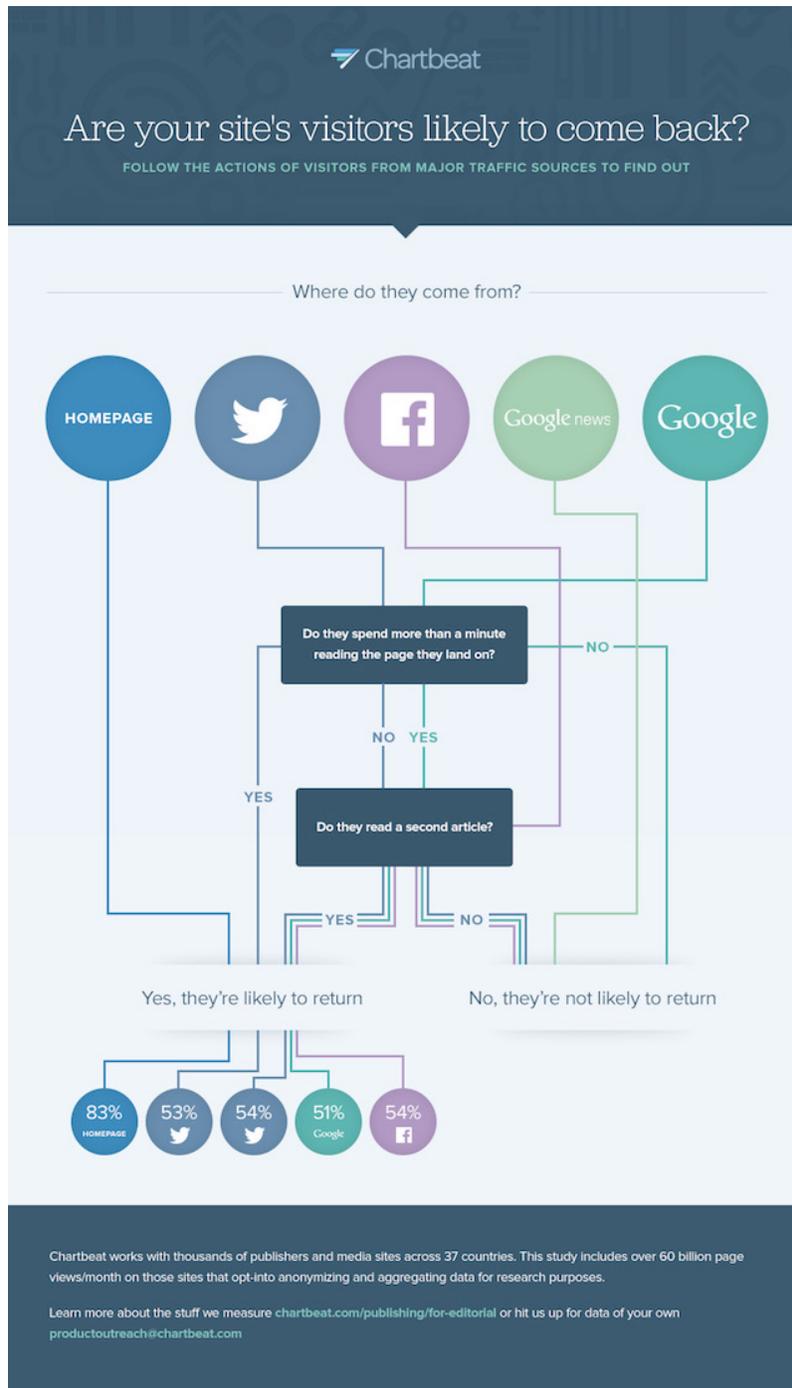
The rare external site with a high return rate to your site should be thought of as a top candidate for a link partnership.

Of course, external sources' extremely low return rates can also be taken as a challenge: if 3% of visitors from Drudge come back to your site and you can push that number to 5%, you could see dramatic growth in your audience. Compared to pushing Facebook traffic's return rate up from 30%, that challenge might be relatively easy.

Part 5: Conclusion

For the final chapter of our series on Understanding Your Traffic Sources, I wanted to go over some best practices for managing referral traffic and identify a few places where you can use Chartbeat data to support your decision-making.

But first, let's sum up the data that we've seen over the past few weeks. The graphic below shows what sort of browsing behaviors are indicative of visitors coming back to your site, based on many sites' most common traffic sources.



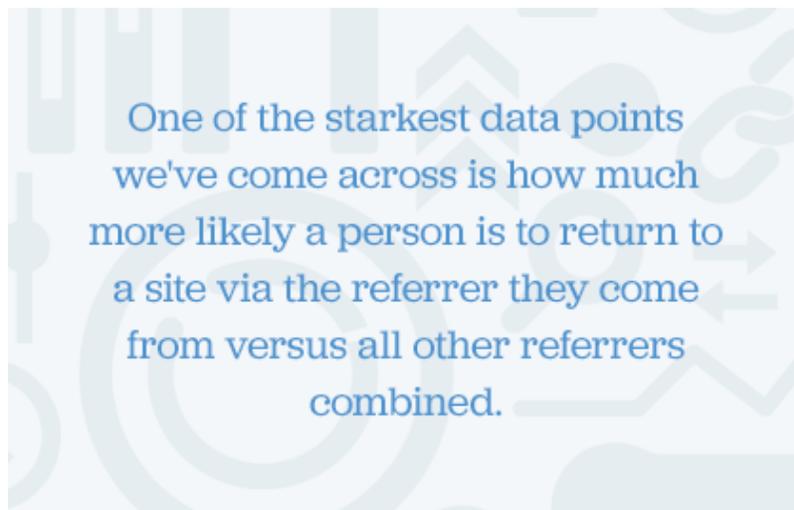
At one extreme, we have visitors who come to your site homepage direct and are always likely to return. At the other, those who come via Google News are unlikely to return, regardless of how they read. In the middle, though, we have an interesting split:

- Visitors who come from Facebook are likely to read most of the article they land on, but those who click to a second article are much more likely to return
- Visitors from Twitter and Google search, on the other hand, consuming the entire article they land on is the best indicator of a likelihood of returning

Traffic from other, smaller sources tends to behave much like Google News or Twitter traffic in this graphic. Now that we have a sense of how different kinds of referral traffic behaves, I'll dive into right into what actions you can take with this data.

WHERE – AND HOW – TO CONCENTRATE YOUR EFFORTS

One of the starkest data points we've come across is how much more likely a person is to return to a site via the referrer they come from versus all other referrers combined. Those who come from Facebook are likely to return only via Facebook, those who come from Google News are likely to return only via Google News, and so on. In that sense, the most important thing you can do to grow audience from a given referrer is maintain a steady stream of links from that referrer.



Given that, you should ask two questions. **First, what sources should we concentrate on building traffic from? Second, what can we do to build that traffic?**

The best way to decide the former, if you're a Chartbeat Publishing client, is to take a look at the "return rate" and "return direct rate" columns of your Weekly Perspectives. Those columns express, in essence, the value of links from different referrers — those with higher return rates send traffic that's more likely to return to your site.

If you don't have access to Chartbeat Publishing, the general trend that we've seen is that, unsurprisingly, visitors from social sources have the highest likelihood of returning, while sources like Google News, Reddit, and Outbrain are likely to increase your site's reach by sending new visitors, but are unlikely to meaningfully help you grow your audience in a self-sustaining way.

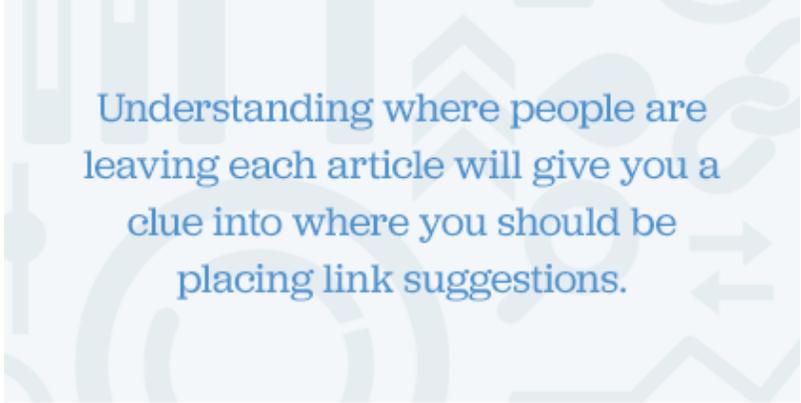
The second question, of course, is much harder to answer in broad terms. Taking each traffic source one-by-one, though:

- **Twitter:** One thing we've seen many times is that people don't promote posts nearly as often on Twitter as they should. Most sites see the majority of their Twitter traffic coming from their own tweets, and the lifetime of a tweet is incredibly short. Tweeting headlines is rarely the right choice.
- **Facebook:** Facebook traffic typically comes from organic sharing, which means it's harder to predict and control. One thing you can control is Facebook's preview text, and it's hugely important. If you don't know what text is showing up on Facebook's previews, you need to figure it out.
- **In-network sites:** If your site is part of a network, working to maintain links from your sister sites is critical. It's not uncommon to see return rates over 50% (about twice as high as for typical referrers) for in-network traffic, which is a function both of similarity of audience and of the regularity of links. Fostering these types of link partnerships is one of the best ways to sustainably build audience.
- **Google:** First off, it's critical to separate "branded" search (searches for your domain name or URL) from truly organic search and Google News. Branded search should be thought of as akin to direct traffic. Optimization for organic search is a huge topic unto itself and probably beyond the scope of this post.

A CAVEAT FOR PAYWALL SITES

One place where sites often miss out is with paywalls that are porous for traffic from external referrers, only presenting a prompt to subscribe on later pages. Under that scheme, a visitor, for instance, who always comes from Twitter and only read the article she lands on will never even be asked to subscribe. We've seen some publishers move toward differentiated paywalls for exactly this reason — traffic from some referrers is immediately asked to log in while visitors from others are allowed to read an article or two for free.

If that fine-grained control isn't in the cards, your goal should always be to get visitors to read through to a second article. Looking at "subsequent time" in your Weekly Perspectives should give you some idea of which referrers send visitors that are likely to click to a second page — concentrating on getting traffic from these referrers makes sense. And, understanding where people are leaving each article will give you a clue into where you should be placing link suggestions. Great related links at the top of an article aren't in view for visitors who read the whole page, and great links at the bottom of an article don't matter to those who never scroll down to see them.

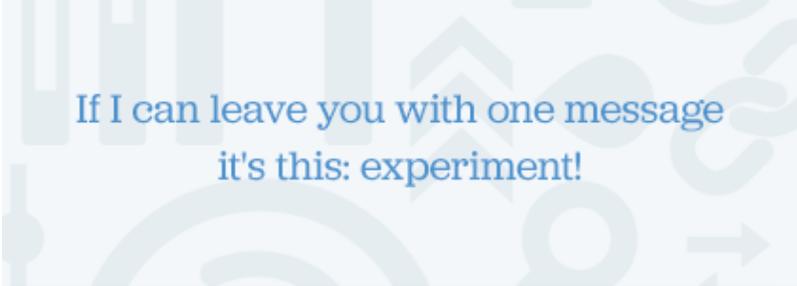


Understanding where people are
leaving each article will give you a
clue into where you should be
placing link suggestions.

WRAPPING UP

This series hardly scratched the surface of what can be said about traffic sources. Much of the most exciting data is easiest to find under the hood of your dashboard – the data that’s specific to your site, not the internet as a whole.

Thanks for reading, and if I can leave you with one message it’s this: experiment! What we’ve presented over the past five articles are broad statistics about traffic across the internet, but we regularly see sites that wildly depart from the average. If you see a return rate of 10% from a given referrer, take that as a challenge and try getting traffic to a different set of links from that referrer and see if you can push next week’s rate to 11%.



If I can leave you with one message
it's this: experiment!