quantcast

Quantcast Methodology Overview
Delivering An Actionable Audience Service

Quantcast Corporation

201 3rd Street, 2nd Floor
San Francisco, CA 94103

Phone: 415.738.4755
Email: contact@quantcast.com

Our Business is in Transition

For decades, media buyers and sellers have used panel-based audience measurement as the basis for defining target segments, evaluating media properties and projecting the efficiency of advertiser investments. In a pure analog or unit-based economy (e.g. TV, radio and print) this common language served the industry well; content choices were few in number as were the media consumption options for consumers. A relative 'trust in the numbers' approach enabled publishers and marketers to focus on brand goals and performance.

The evolution of digital media has brought about not only content and audience fragmentation but also measurement fragmentation. Today, digital audience measurement services frequently report data that is out of sync with publishers' own internal traffic measures due to varying methodological approaches. Adding complexity — marketers typically possess back–channel data from their advertising campaigns as well as other proprietary analytical sources and tools.

While heightening the expectation of accountability, the existence of so many disparate and disconnected data sources leaves the industry with a lack of common language for the planning, buying, selling and evaluation of digital media investments. Accuracy issues aside, digital marketers are forced to rely on one set of measures for strategy and planning (audience-based, panel-centric) while using other systems to track campaign activity and impact (impression-based, directly-measured). Discussions centering on brand goals and performance are typically clouded by a collective 'disagreement' on which numbers matter most.
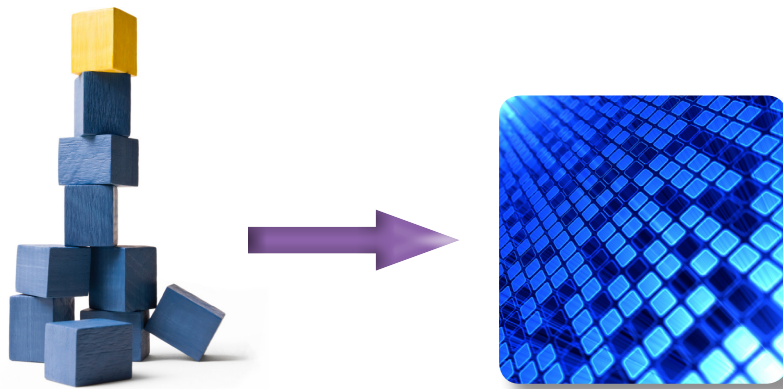
The media industry is moving from a *unit*-based to an addressable *impression*-based economy.

For audience measurement to be relevant in an addressable media economy, it must represent what is being transacted. The industry's shift away from a unit-based economy (30-second spots, pages in magazines, etc.) to one increasingly based on dynamically allocated impressions, only compounds the need to address the shortcomings of existing measurement approaches. By solving this problem, we as an industry will move closer to realizing the promise of digital media and provide more relevant advertising and content offerings for consumers, resulting in decreased waste in media spending and delivering improved yield for publishers and advertisers.

Quantcast's Common Language Solution

Digital media measurement should play the same role as its analog cousin — it should deliver data that empowers publishers and marketers to transact in trusted, meaningful ways. Unlike analog forms of media, digital offers the exciting potential for full addressability, supporting real-time decision-making and significant improvements in advertising effectiveness. To get there, it is therefore critical that any digital measurement standard be highly actionable, specifically in real-time.

Embracing the inherent accountability and data richness of the digital media landscape, the Quantcast approach provides a foundation to unite previously disjointed measurement paradigms to deliver a consistent, insightful and actionable audience service.

In September 2006, Quantcast launched a no-cost direct-measurement solution for digital media participants. Technology and content agnostic, the Quantcast Quantified Publisher service is able to capture virtually all forms of media consumption activity — whether destination based, or distributed in nature. Websites, blogs, widgets, games, audio, video and even ad campaigns can be easily, and comprehensively, measured. Today the program delivers audience profiles inferred from directly-measured traffic in real-time from over 80,000 participating publishers. Together, participating publishers' directly-measured properties represent over 10 million distinct web destinations.

Resulting data can be utilized throughout the entire digital marketing ecosystem from strategy development and target definition/ segmentation, to media negotiation, implementation, optimization and performance evaluation. Post campaign, Quantcast-driven insights can be applied to future planning, sales and media delivery activities.

Publishers benefit by participating in the Quantcast Publisher program because the data collection approach ensures that *all* media consumption is being counted regardless of its origin (work/home/university) or access platform (PC, Mac, iPhone, etc). A key benefit of this service is that publishers have an easy way of validating the accuracy of traffic data — Quantcast cookie counts and page views can be easily compared to a publisher's internal web analytics measures — and, as the only service that also provides people counts, the approach offers transparency giving both buyer and seller confidence in the metrics that are being transacted against.

Publishers can also choose to organize their digital media assets based on the way they package and sell, instead of relying solely on third-party panel-based aggregations which many times are not sensitive or responsive enough to provide real-time, granular audience profiles.

Marketers benefit because the resulting audience data (people-based demographics, behavioral affinities, etc.) reflect what a seller can actually deliver. This provides marketers the opportunity to plan, buy and evaluate advertising opportunities that tie directly to their targeting objectives and overall business goals.

Most importantly, open participation by both publishers and marketers in the Quantcast audience service promotes an environment of control, transparency and accountability while providing the foundation for real-time decision-making that will define the future of addressable advertising. By providing services that restore the collective 'trust in the numbers,' Quantcast enables publishers and marketers to return to discussions about what really matters: brand goals and performance.

How Quantcast's Methodology Works

The Quantcast methodology represents an innovative new way of understanding audiences. In a major advancement over the panel-based audience measurement approaches that have dominated the industry for decades, Quantcast's breakthrough approach couples machine learning with massive quantities of directly measured data to deliver detailed audience information that can be leveraged in real-time for addressable advertising.

The Visit Graph

At the heart of Quantcast's unique direct measurement audience service is an inference model based on a proprietary representation of digital media consumption that is called the *visit graph* (see figure 1). The visit graph analyzes the dynamic relationship between Internet media audiences and their interactions with a comprehensive set of digital media assets, which includes websites, blogs, video, widgets and advertising campaigns.
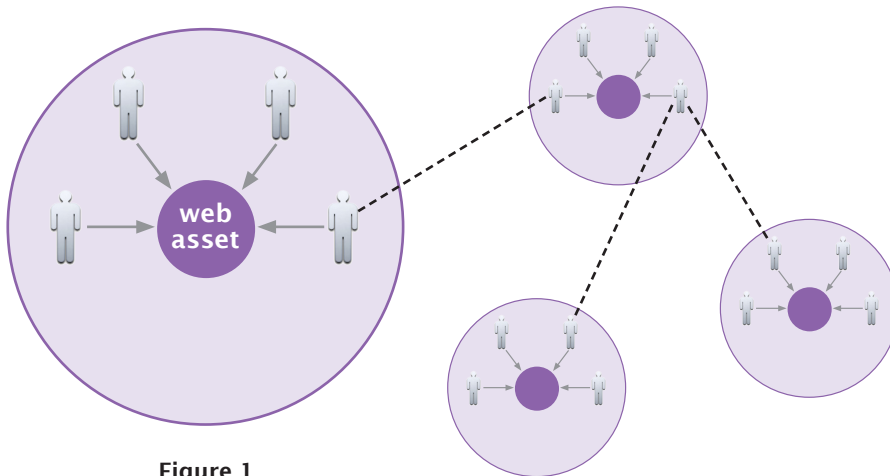

Figure 1

Search engines use similar models based on the linkages between websites. The nature of these linkages forms evolving link graphs that directly impact the natural search results of keyword queries. In a similar fashion, Quantcast developed its visit graph model to characterize the various types of audience behavior in relation to a wide range of media assets based on directly measured data from partner publishers and marketers.

Quantcast's direct measurement methodology is technology, content and behavior agnostic offering publishers and marketers the ability to measure user activities such as:

- · viewing a web page
- · downloading an MP3
- · responding to an ad
- · launching a widget

- · watching a video
- · purchasing a product
- · chatting on IM
- · playing an online game

Quantcast currently captures the audience and behavioral activity of over 10 million distinct media assets from over 80,000 partner publishers worldwide. Our 8 data centers located around the globe capture and process over 4 billion new media consumption events daily which contribute to an evolving model of media consumption based on more than 1 trillion direct observations.

This direct measurement approach supplies the visit graph with petabytes of census-level traffic data and provides our model with a unique foundation of comprehensive audience visibility against which we model audience size and composition.

Visit Graph Data Inputs

The visit graph model accommodates directly measured data and multiple reference data sets:

- **Directly Measured Data** forms the core of the Quantcast analytical methodology. Quantcast launched the Quantified Publisher program in September 2006 and through it we collect comprehensive census records of media consumption from over 10 million distinct web destinations each month. This data set grows by more than 125 billion records every month, representing the media consumption activity of over 200 million people in the US and more that 800 million people worldwide.

- **Multiple Reference Points** supplement the directly measured data within the visit graph. Reference sources include data sets that would traditionally be used in panel approaches comprising click-stream and non-PII (Personally Identifiable Information) user data. These data sets are sourced from **multiple parties** including market research companies, ISPs (broadband and dial-up) and toolbar vendors and cover in excess of 2 million individuals (1.5 million in the US). Quantcast also provides support for Quantified Publishers to share non-PII data via the direct measurement solution and this provides cookie-level data on many tens of millions of individuals. Quantcast is continually adding new reference data sources to validate, establish and improve the inference capabilities of its models.

Quantcast's combination of directly measured data and multiple contributing reference data sets is unique in the industry. It is important to note that the visit graph model is **not** a pixel-panel intersect, but rather a holistic mathematical inference process that intelligently combines the data. Pixel-panel intersect models suffer from all of the same sample size limitations and biases of traditional panel based approaches.

The visit graph construction process uses proprietary disambiguation methods to account for multiple users of single machines in both directly measured and reference point data sets.

Reconciling Reach Measurement

The subject of reach is one that has long been debated in Internet measurement. Many publishers believe that single-source panel estimates poorly capture the true scale of their audience due to the inherent biases (or lack of them) within any given panel (e.g. at work users — especially Fortune 1000, students, Mac OS, industry verticals, etc). Panel measurement vendors re-weight their sample to adjust for bias, however, this can only be achieved for the biases that are well understood, and the unique draw of different media properties may encompass audience niches and biases that cannot be accounted for adequately by traditional panel normalization techniques.

On the other side of the measurement divide, direct measurement approaches such as web analytics services are counting cookies, which due to issues of cookie deletion and multiple machine usage may not be representative of the number of people consuming media.

Quantcast's inference approach bridges this measurement divide by constructing models that deal with both cookie deletion and multiple machine usage. The key to this approach is to compare directly measured data with reference points that provide known truths (for example web destinations that require registration) and to use this to calibrate models. The resulting Quantcast model takes into account numerous factors including the frequency of visitation and the respective balance between work and home access to build a translation of cookies to people that is unique to each web property. The model was introduced in June 2008 and is detailed in our white paper, 'Cookie Corrected Audience Data' available for download on the Quantcast website.

Calculating Audience Composition

The calculation of audience composition is the most ground-breaking aspect of Quantcast's audience service. Bypassing the sampling error limitations of single-source panels, Quantcast's visit graph inference model enables the characterization of digital media audiences of nearly any size.

Panel approaches are only able to provide estimates for properties that are of a certain minimum size due to the

inherent limitations of sample based statistics, the result being that most services only provide data on 20-30k properties. Quantcast's composition inference approach is built on the foundation of direct measurement and provides rich audience profiles for millions of digital media assets, with as few as 500 unique viewers.

How Quantcast's audience composition inference models work:

Quantcast directly observes more than 130 billion user media exposures on over 10 million media assets each month. The daily aggregation of all this media activity from partner publishers and marketers forms the underlying structure of the visit graph (figure 1).

In considering the audience estimate for a particular media asset, the Quantcast model takes into account data that may be available from reference sources as well as the audience estimates for all members of the sub-graph of adjacent audience members and properties in relation to the relative strength of the connection between those nodes (figure 2).
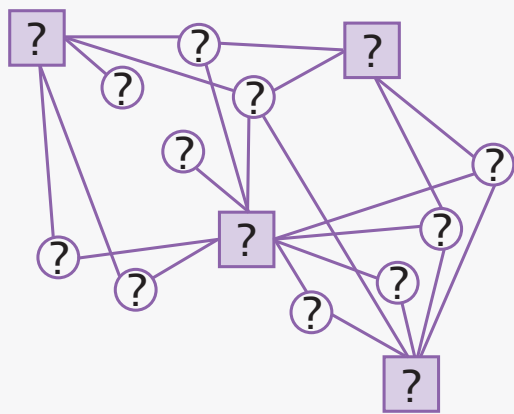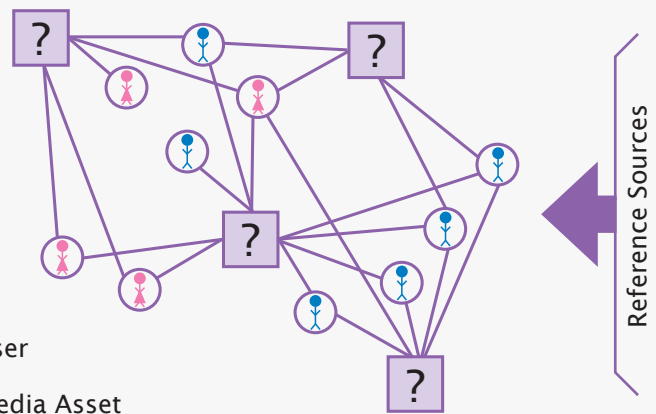
Figure 1

Figure 2

○ User

□ Media Asset

Reference Sources

The audience estimate of any quantified media asset can be then be expressed as a function of the audience characteristics of the entities that consume this media (a visit within the visit graph). Similarly, the audience dimensions of each of these entities can be expressed as a function of the characteristics of the media assets that it consumes. Quantcast applies mathematical models that operate iteratively on sections of the visit graph of relevance to a given digital media asset (figure 3).
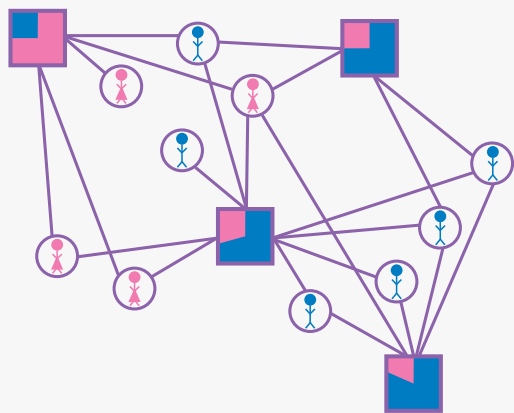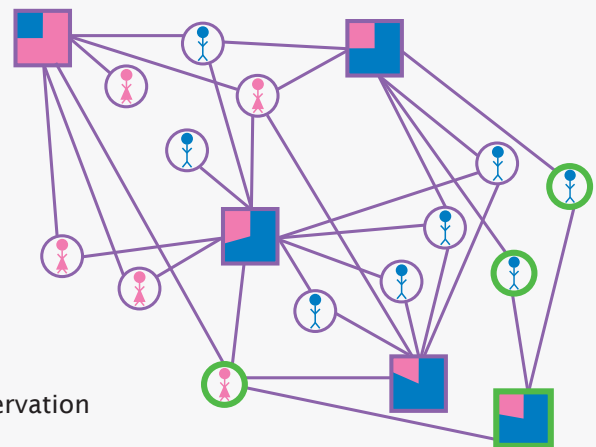
Figure 3

Figure 4

■ Male

■ Female

■ New observation

This machine-learning inference model just gets smarter over time. As additional media assets and related exposures are continually added to the model, a *ripple effect* occurs. Real-time updates to the estimates are made not just to the added observations, but also to all relevant exposures and web assets along adjacent parts of the visit graph (figure 4).

Validating the Quantcast Audience Service Approach

Quantcast's approach is different from that employed to deliver traditional media measurement. Rather than working solely from a sample using a constructivist modeling approach, Quantcast operates within an environment of massive data volumes and, critically, reliable dependent variables of known attributes. This allows for traditional model validation approaches to be leveraged and Quantcast makes extensive use of these for both reach modeling and inference associated with audience composition.

Industry acceptance of Quantcast's ground-breaking media measurement solution is our top priority and we have initiated the MRC audit process. Our pre-audit is complete and we anticipate completion of the first full phase of the audit in 2009.

A Methodology Grounded in Actionable Data & Control

The Quantcast mission is to enable a liquid marketplace for digital advertising, driven by a deep understanding and rapid delivery of audience insights.

We are focused on delivering innovative solutions that leverage the potential of directly-measured data as the underlying foundation for dynamic advertising models.

Our service is predicated on fundamentals that differentiate us from traditional audience measurement services:

- More data is better than less

- An increasing variety of inputs can be used to continually validate and refine Quantcast inference models

- Scalable computing can resolve the challenges of an increasingly dynamic and fragmented media landscape

- Because digital media and related marketing are delivered in real-time, audience services should be delivered in real-time as well, not weeks or even months later

- Contributors of directly measured data should have appropriate control over how that data can be accessed or subsequently used

Quantcast Corporation

201 3rd Street, 2nd Floor
San Francisco, CA 94103

Phone: 415.738.4755
Email: contact@quantcast.com