



Entity Resolution Boosts Web Traffic 6X

Meredith needed to disambiguate anonymous users on its media properties. Using a community detection algorithm, Meredith turned 14B anonymous users into 163M unique profiles, creating a better user experience and leading to 612% more web traffic.

BY THE NUMBERS

30B nodes

35B relationships

6X more web traffic

PLATFORM

Neo4j Enterprise Edition with
Neo4j Graph Data Science (GDS)

INDUSTRY

Media

USE CASE

Entity Resolution

OBJECTIVE

Understand consumption patterns to deliver personalized content and advertising

CHALLENGE

Increasingly untraceable users and the sheer volume of data to disambiguate

SOLUTION

Use Neo4j GDS to resolve anonymous user data into rich audience segments for personalized experiences and increased conversion

RESULTS

- Increased web traffic by 612%
- Enabled commercialization of first-party audience data

The Company

Meredith Corporation is a media conglomerate whose digital presence reaches more than 180 million users monthly across their multimedia platforms, amounting to over \$3 billion in annual revenue. Through their 40+ powerhouse publications from Better Homes & Gardens to People, InStyle to Allrecipes, Meredith is dedicated to giving users specific and personalized experiences, delivering the right content at the right time.

The Challenge

Meredith aims to provide users with content based on their interests, which proves to be challenging when users are not logged in and the lifetime of a cookie is limited. Since users often delete their cookies or browsers sometimes block cookies automatically, data scientists were unable to view users' content usage over time, a crucial pattern to understand in order to effectively serve content and advertising.

On the technical side, streams of data were isolated from one another in Meredith's traditional database, leading to a narrow view of the customer. Conflicts of timestamps, cookies mismatching, and of course the computational lift required to marry multiple data streams and make sense of them, made it impossible to reconcile user profiles. Nevertheless, the team at Meredith relied on a relational database management system (RDBMS) for years, doing what they could to understand their users.

Further complicating the process, Meredith had to incorporate third-party data provided by vendors who have their own proprietary identity graphs. Media companies like Meredith send these vendors their anonymous user data, and vendors return enriched user profiles, in a somewhat opaque process. As Senior Data Scientist Ben Squire admits, "It can be difficult to validate and verify the accuracy of these products, especially when a large portion of the traffic that you send is anonymous."

The Solution

Meredith's initial foray into graph data science was an experiment to see if they could extend the lifetime of a cookie. By connecting their previously isolated data streams, they were able to see a pattern right away: cookies that were meant to identify unique users were sometimes repeated across



"Instead of 'advertising in the dark,' we now better understand our customers, which translates into significant revenue gains and better-served consumers."

**Ben Squire, Senior Data Scientist,
Meredith**

different data streams, meaning they were mislabeling users as multiple profiles instead of consolidating them into one more descriptive profile. As Squire puts it, "The most surprising result was really seeing how connected the data was. I used to think that we knew this data really well when we looked at it individually from each different data stream, but when you combine it all together and you actually look at the datasets as a whole, it makes you realize that it's like trying to solve a Rubik's Cube by only looking at one side."

The Meredith team was also able to visualize third-party customer data in a graph. When they saw suspicious patterns, such as hyperdense node structures, they were able to evaluate the accuracy of vendor data and cut the cost of using unreliable data.

They've since graduated from querying their graph for local connections and visually identifying anomalies to running graph-global analyses to make sense of their highly connected cookie co-occurrence graph, without needing to know specific patterns for which to look. Meredith began leveraging a community detection algorithm called Weakly Connected Components (WCC) to find subgraphs within their multibillion node dataset that can be attributed to distinct profiles. They use these more accurate profiles to create audience segments, the holy grail of any media property's advertising business.

"We basically have increased our understanding of a customer by 20 or 30 percent, just looking at how the data connects over time, rather than just looking at individual cookies," Squire comments.

The Results

Previously, identifying user profiles from first-party cookies was, at best, a guessing game, and Meredith was forced to blindly trust third-party vendor data. But with Neo4j GDS, Meredith has become the expert of their first-party data and empowered to vet third-party data for vendors. As a result, Meredith has resolved 14 billion anonymous data points into 163 million unique user profiles, which has helped increase web traffic by 612%.

With commercialization of their audience segments on the horizon, Meredith will soon be able to measure just how effective their identity graph and audience segments are. And their user profiles only become richer, as they add more data to their pipeline every day.

Moving forward, Meredith will be experimenting with more GDS algorithms to uncover new insights and optimizing their current data science workflow even further. With Neo4j GDS, Meredith is able to scalably analyze billions of customer data points, revealing insights they wouldn't even know to look for.