What a massive family tree can show us
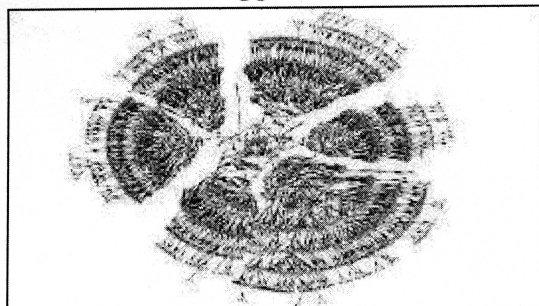
When did people stop marrying cousins? How far did they travel to find mates?

It's not the biggest family tree in the world, but it's close.

Armchair genealogists and a team of computers scientists have assembled a massive family tree that includes 13 million individual members and spans an average of 11 generations.

A study describing the tree, published this week in Science, also details some of what we can learn from this crowdsourced data. For example, it reveals when people stopped marrying their cousins, whether men or women traveled farther from home for marriage, and provides clues about how longevity is inherited.

The tree is based on data assembled by roughly 3 million genealogy enthusiasts who have identified the familial relationships of more than 86 million individuals on the website Geni.com.

Kevin Bacon is in there. So is Donald Trump.

You'll find me on the Geni site too. (Thanks, second cousin Scott!)

Not everyone in the website's database is included in the current study, however. For this work, the authors only used data from profiles that users had agreed to make public.

Before assembling the mega family tree, lead author Yaniv Erlich, a computer scientist at Columbia University, and his team first had to make sure the entries were accurate and worth using.

This took a lot of time, Erlich said. The data were not pristine, but perhaps not as faulty as you might expect considering they were provided by millions of contributors.

The researchers found that on average there was a 2% error when listing a person's father, and a 0.3% error for a mother. They also found that about 0.3% of profiles included clear mistakes such as a person having more than two parents, or someone being the parent and offspring of the same person.

To correct these errors the team developed computer programs that "pruned" the tree, removing invalid relationships. After doing that, they generated 5.3 million disjointed family trees — the largest of which included 13 million individuals.

By comparing people in the system with 80,000 death records from Vermont spanning from 1985 to 2000, the authors also found that the people included in their family tree were not any more likely to be rich or poor than the general population. They were, however, much more likely to be white.