NATURE | NEWS: Q&A

The mathematics behind Internet virality

Computational social scientist Sharad Goel studies the spread of memes such as #TheDress.

Davide Castelvecchi

05 March 2015

Sharad Goel

Computational social scientist Sharad Goel devised a mathematical way to define virality.

For a few days in late February, the world stopped to argue over a photograph. Everyone had a strong opinion about the true colour of a dress, and the debate quickly became tribal. Social media were swamped by arguments between those who perceived the dress as blue and black, and those who perceived it as white and gold.

But what does a phenomenon such as #TheDress tell us about how information spreads and what makes it go viral? In a paper due to appear in *Management Science*[1], Sharad Goel and his collaborators propose a mathematical definition of virality that quantifies the extent to which a concept is spreading between friends as opposed to via popular news outlets. *Nature* asked Goel, a computational social scientist at Stanford University in California, how his work applies to #TheDress.

What did you think of how the photograph of #TheDress spread?
As a casual observer, it's pretty amazing. People often use the phrase 'go viral' but don't really know how something is spreading. The dress is a viral phenomenon in the way we formulated in our paper, but at the same time it has lots of non-viral aspects.

What does 'viral' mean?
When people say viral they can mean a lot of different things. It's often a synonym for popular. People will say, "Look at this viral video", when really it was something released by Taylor Swift or something like that.

Something that's a little bit closer to what I think of as viral, is something that's not being promoted by a celebrity and that you wouldn't ordinarily think is going to become very popular. Closer still — and this is what we get at in the paper — is something that it is diffusing mostly person-to-person, rather than through broadcasts such as those from the *New York Times* or other mainstream popular news outlets. We call this person-to person process structural virality.

There are lots of other features that might go into a more general notion of virality — for example the speed, the unpredictability — that are not included in this definition. The definition that we made I think encapsulates a lot of the core intuition behind virality but it certainly doesn't capture all of it.

How do you determine whether something has 'structural virality'?

We don't say that something is or is not viral, but rather we measure virality on a continuous scale that goes from broadcast on one end to fully person-to-person on the other. In technical terms, the instances in which people shared a particular piece of content — by tweeting the link to a picture, say — forms a tree that traces out the spread of the content. To measure the level of virality, we calculate the average distance between pairs of sharing events in the tree. At one extreme, the pure broadcast case, all events are just two hops away from each other [because they can all be traced back to the same broadcast], so the average is two; at the other extreme, the average distance grows larger with the size of the tree.

Empirically, it's very hard to measure these things. In our paper, we looked at more than a billion of events on Twitter's 'Firehose' [the company's database of raw data on all tweets] and tried to assess how often things go viral.

Was the #TheDress structurally viral?

Yes, but there were also lots of broadcasts. I first found out about it from an article in the *New York Times*. I don't spend a lot of time on Facebook or Twitter, so I heard about it not through some viral means, but through a traditional broadcast source. The relative contribution of broadcast versus pure peer-to-peer diffusion is not clear.

#TheDress also became a topic of discussion offline, in real life. Was face-to-face interaction a crucial component in the spread?

The reaction of others does seem to have been a driver. But I don't know if it had to be face to face. Sometimes you might share a picture because you are looking for a reaction.

What made this particular image so compelling?

Part of it lies in the fact that it wasn't designed to be an optical illusion. If you realize that something was designed to trick you, I think it has a different effect. This was just a person who stumbled upon this. I don't think anyone could have reasonably guessed how popular it would be.

If this had been released on a different day or if different people had seen it, then a lot of things could have gone differently and not as many people would have seen it. There is an element of luck that often people don't realize in these phenomena. We are used to thinking deterministically.

Why study virality?

Despite our collective, cultural fascination with viral events, we know surprisingly little about what drives them, how often they occur or even what the term 'viral' actually means. We now have both the data and the computational tools to shed light on these fundamental questions, which I find exciting.

Nature | doi:10.1038/nature.2015.17046

References

1. Goel, S., Anderson, A., Hofman, J. & Watts, D. J. *Manage. Sci.* (in the press). Available at http://jakehofman.com/inprint/twiral.pdf