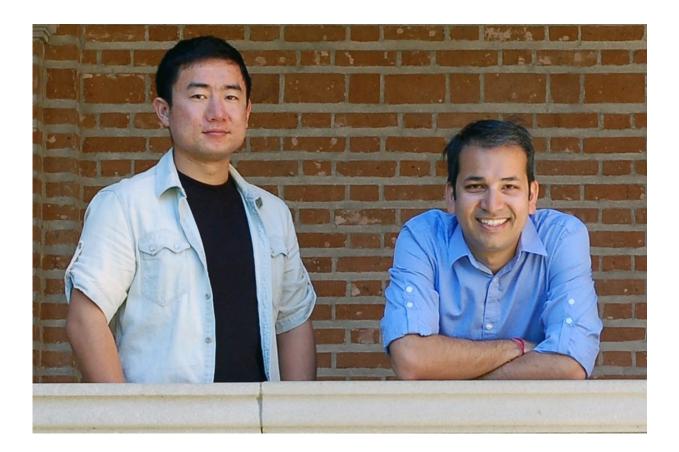# Scientists crunch social media data to explain how communities affect friendships

September 26 2018, by Jade Boyd

Chen Luo and Anshumali Shrivastava. Credit: Jeff Fitlow/Rice University

Your chances of forming online friendships depend mainly on the number of groups and organizations you join, not their types, according to an analysis of six online social networks by Rice University data

scientists.

"If a person is looking for friends, they should basically be active in as many communities as possible," said Anshumali Shrivastava, assistant professor of computer science at Rice and co-author of a peer-reviewed study presented last month at the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining in Barcelona, Spain. "And if they want to become friends with a specific person, they should try to be a part of all the groups that person is a part of."

The finding is based on an analysis of six online social networks with millions of members, and Shrivastava said its simplicity may come as a surprise to those who study friendship formation and the role communities play in bringing about friendships.

"There's an old saying that 'birds of a feather flock together,'" Shrivastava said. "And that idea—that people who are more similar are more likely to become friends—is embodied in a principal called homophily, which is a widely studied concept in friendship formation."

One school of thought holds that because of homophily, the odds that people will become friends increase in some groups. To account for this in computational models of friendship networks, researchers often assign each group an "affinity" score; the more alike group members are, the higher their affinity and the greater their chances of forming friendships.

Prior to social media, there were few detailed records about friendships between individuals in large organizations. That changed with the advent of social networks that have millions of individual members who are often affiliated with many communities and subcommunities within the network.

"A community, for our purposes, is any affiliated group of people within the network," Shrivastava said. "Communities can be very large, like everyone who identifies with a particular country or state, and they can be very small, like a handful of old friends who meet once a year."

Finding meaningful affinity scores for hundreds of thousands of communities in online social networks has been a challenge for analysts and modelers. Calculating the odds of friendship formation is further complicated by the overlap between communities and subcommittees. For instance, if the old friends in the above example live in three different states, their small subcommunity overlaps with the large communities of people from those states. Because many individuals in social networks belong to dozens of communities and subcommunities, overlapping connections can become dense.

In 2016, Shrivastava and study co-author Chen Luo, a graduate student in his research group, realized that some well-known analyses of online friendship formation failed to account for any factors arising out of overlap.

"Let's say Adam, Bob and Charlie are members of the same four communities, but in addition, Adam is a member of 16 other communities," Shrivastava said. "The existing affiliation model says the likelihood of Adam and Charlie being friends only depends on the affinity measures of the four communities they have in common. It doesn't matter that each of them are friends with Bob or that Adam's being pulled in 16 other directions."

That seemed like a glaring oversight to Luo and Shrivastava, but they had an idea of how to account for it based on an analogy they saw between the overlapping subcommunities and the overlapping similarities between webpages that must be taken into account by internet search engines. One of the most popular measures for internet

search is the Jaccard overlap, which was pioneered by Google scientists and others in the late 1990s.

"We used this to measure overlap between communities and then checked to see if there was a relationship between overlap and friendship probability, or friendship affiliation, on six well-studied social networks," Shrivastava said. "We found that on all six, the relationship more or less looked like a straight line."

"That implies that friendship formation can be explained merely by looking at overlap between communities," Luo said. "In other words, you don't need to account for affinity measures for specific communities. All that extra work is unnecessary."

Once Luo and Shrivastava saw the linear relationship between Jaccard overlap of communities and friendship formation, they also saw an opportunity to use a data-indexing method called "hashing," which is used to organize web documents for efficient search. Shrivastava and his colleagues have applied hashing to solve computational problems as diverse as indoor location detection, the training of deep learning networks and accurately estimating the number of identified victims killed in the Syrian civil war.

Shrivastava said he and Luo developed a model for friendship formation that "mimicked the way the mathematics behind the hashing work."

The model offers a simple explanation of how friendships form.

"Communities are having events and activities all the time, but some of these are a bigger draw, and the preference for attending these is higher," Shrivastava said. "Based on this preference, individuals become active in the most preferred communities to which they belong. If two people are active in the same community at the same time, they have a

constant, usually small, probability of forming a friendship. That's it. This mathematically recovers our observed empirical model."

He said the findings could be useful to anyone who wants to bring communities together and enhance the process of friendship formation.

"It seems that the most effective way is to encourage people to form more subcommunities," Shrivastava said. "The more subcommunities you have, the more they overlap, and the more likely it is that individual members will have more close friendships throughout the organization. People have long thought that this would be one factor, but what we've shown is this is probably the only one you have to pay attention to."

Provided by Rice University