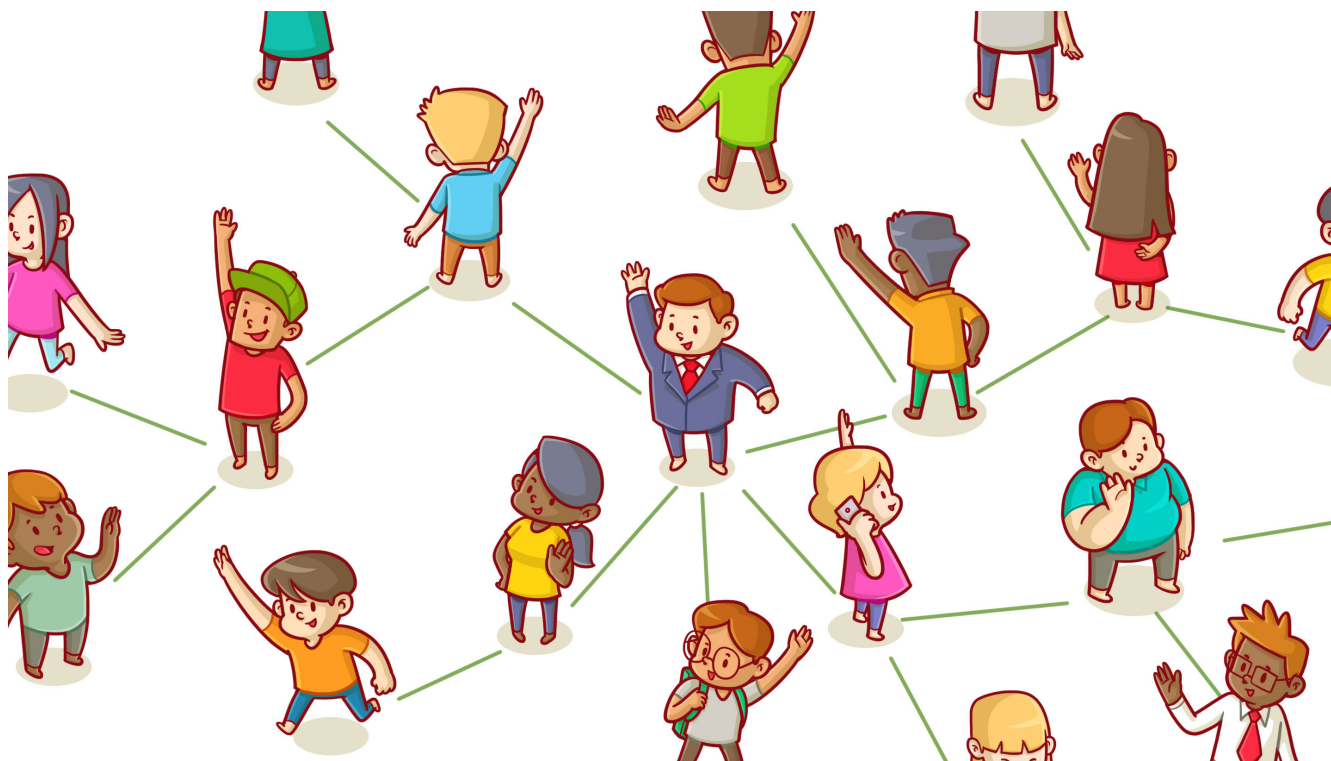Quantamagazine

How Network Math Can Help You Make Friends

By **Patrick Honner**

August 20, 2018

Studying the structure of existing friendships in your community can help you forge the best connections when forming a new circle of friends.


BIG MOUTH for Quanta Magazine

When you start at a new school or job, or move to a new city, how do you go about making new friends? You could take an active approach, forging strategic connections with the popular kids and the movers and shakers. Or you could leave things to chance, relying on random groupings and associations. Whatever your approach, understanding the structure of existing friendships in your new community can help you make the best connections, which will ultimately define your circle of friends.

Imagine moving to a strange new city, Regulartown, that has a strange rule: Everyone can have at most four friends, and everyone wants to maximize their friendships. What will the structure of friendships in Regulartown look like? To explore this question, we'll use a mathematical object called a network.

Simply put, a network is a set of objects, called "nodes," and the connections between them. Networks are mathematically versatile: They can represent computers and the wires that connect them, authors and their collaborations, or the states of a Rubik's cube and the moves that transform them — essentially any set of connections, real or abstract. To study friendships in Regulartown, we'll create a network where the nodes are people and the connections are the friendships between them.

A useful way to represent networks is to imagine the nodes as dots and the connections as line segments, which we will also call edges. This network diagram can give us a glimpse into its structure. So what will the network of friendships look like in Regulartown? At some point it may look something like this:
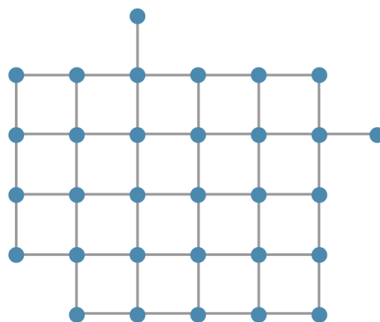
Each person will try to find their four friends, and as new people move to town, they will look for someone with fewer than four friends to connect with. In this way the network will continue to grow over time, continually expanding at the edges as new nodes are added. (It's also possible for independent cliques to form, but we'll ignore that possibility in our example.)

Diagrams of networks can be illuminating when they indicate a clear structure. But when networks get big, or don't exhibit the kind of regular structure of a Regulartown, diagrams may be less useful. It helps to develop different ways to analyze a network's structure. One way is to think about the degree distribution of the network.

In a network, the number of connections a node has is known as the "degree" of that node. A node with a high degree is connected to many other nodes; a node with a low degree is connected to few other nodes.

A node
of degree 8
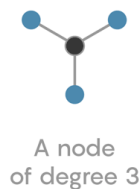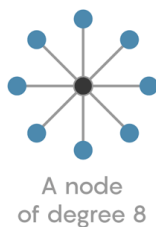
A node
of degree 3

The degree of a node is an important measurement in a network, but it is a local one: It only describes the structure of a network at a single node. But by thinking of the degrees of all the nodes simultaneously, we can create a useful tool for understanding a network's global structure.

In our friendship network, the degree of each node is the number of friends each person has. In Regulartown most people will have four friends, so most nodes will have degree 4. No residents will have more than four friends, but some will have fewer, so there will be nodes with degree 3, 2 or 1. We can summarize the distribution of degrees like this:
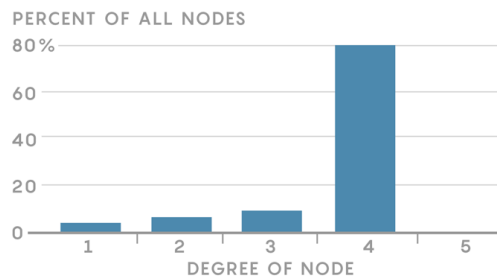
This histogram conveys important information about the structure of our network. In this simple example, it may not tell us as much as our network diagram, but we'll see how degree distributions can be powerful tools for understanding different kinds of networks.

Let's move on to a new town. In Randomville, friendships happen randomly. Since randomness can be tricky business, let's be clear about what we mean: We'll imagine every person in town as a node in a network, which makes every possible edge a possible friendship. To generate a random friendship, we will choose one of those possible edges at random and draw it in, establishing a connection between those two nodes and thus a friendship between those two people.

What would Randomville's network look like? Assuming we start with a bunch of nodes and randomly add a bunch of edges, the picture may look like this:
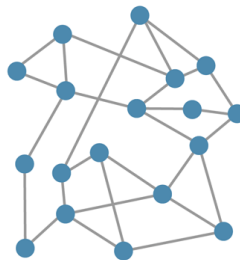

It may be hard to see structure in this diagram. But the degree distribution of this network is illuminating. While it's not easy to compute directly, we can reason through some important properties using a simple example.

Imagine you are one of 10 people in Randomville. How many possible friendships are there? Each of the 10 people could be connected to the other nine, so it seems like you could potentially draw 10 × 9 = 90 edges. But this actually counts each possible friendship twice: Once for each friend. So the total number of possible friendships is really 90 divided by 2, or 45.

Now let's say we randomly choose a friendship, that is, we randomly select one of the 45 possible edges in our network. What's the probability it connects to you? Well, there are nine possible edges extending from you to each of the other nine nodes. Since nine of the 45 edges connect to you, the probability that a randomly selected edge will connect to you is $\frac{9}{45} = \frac{1}{5}$, or 20 percent.

But this same argument applies to everyone in Randomville, so every node has a 20 percent chance of being connected to the randomly selected edge. Now, as edges (and nodes) are added, these probabilities will change slightly, but in the long run they will remain roughly the same. This means that friendships will be pretty evenly distributed around Randomville. There will be some slight variation here and there, but having very few friends, or very many friends, will be unlikely. In Randomville, almost everyone is likely to end up with something close to an average number of friends.

These familiar features are embodied in the "binomial" degree distribution of a typical random network.
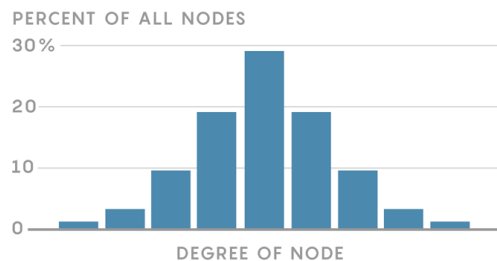
By looking only at the degree distribution of this network, we can infer a particular kind of uniformity: When it comes to connectivity, most nodes are average and very few are extreme. This is useful information when it comes to understanding the structure of the network. (As nodes are added, say, when new people come to town, the distribution will change slightly, but the general features will persist.)

Now, neither of these two examples — the at-most-four-friends rule of Regulartown or the randomly selected friendships of Randomville — are realistic models of friendship. People can have more than four friends, and having a lot of friends isn't as unusual as the binomial distribution suggests. So what is a more realistic model of friendship?

As you build connections with friends and friends of friends, the structure of your friendships will likely share features common to other real-world networks like food webs, protein interactions and the internet. These features characterize so-called "scale-free" networks, a model of connectivity that has come to dominate network science over the past 20 years. Researchers from mathematics, physics, economics, biology and the social sciences have all seen the telltale signs of scale-free networks in their disparate fields.

A complex scale-free network representing the metadata from a social network.

Martin Grandjean

The structure of scale-free networks depends on the simple principle of "preferential attachment." Preferential attachment is a rich-get-richer rule of network growth: A node with many existing connections is more likely to get new connections than a node with few existing connections. New connections show a preference for high-degree nodes.

Does this make sense in the context of friendship formation? Generally speaking, it seems reasonable to argue that someone with many friends will be more likely to make new ones. Since they are already connected to more people, they are more likely to meet new people through those existing connections. Having more friends creates more opportunities to make new friends. And the

fact that they already have a lot of friends suggests they may have some kind of capacity or affinity for making friends. This will likely draw others to them, just as popular websites draw links from other sites and blogs, and established cities invite new rail lines and air routes.

While there are multiple factors at work in the development of scale-free networks, preferential attachment is seen by many as the most fundamental. And it has a fascinating consequence on a network's degree distribution.

PERCENT OF ALL NODES
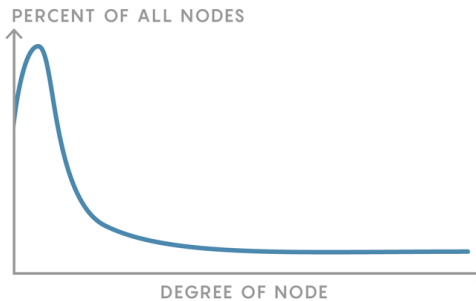
DEGREE OF NODE

Preferential attachment predicts a "fat-tailed" degree distribution. The majority of nodes in the network will be low degree, but there will be nodes of increasingly higher degree. This contrasts with the friendship networks of Regulartown and Randomville, which had few or no nodes of high degree.

These nodes of high degree, which act like hubs, are a critical feature of scale-free networks. They are the social butterflies of friendship networks, the banks at the center of economies, the centralized routers that regional internet lines run through, the Kevin Bacons of the acting world. Hubs can bring a small-world feel to an enormous network — for example, any two randomly selected users from the two billion people on Facebook are less than four friends apart on average. And the number and diversity of hubs also gives scale-free networks resilience against certain kinds of breakdowns: For example, even if many internet connections fail, messages can still get through, partly because there will still be many ways to get to and from the many hubs.

While there seems to be agreement on the utility of scale-free networks and their high-level features, this area of study is not without controversy. The precise mathematical characteristics of these degree distributions can be difficult to interpret. In his book *Linked: The New Science of Networks*, the network science pioneer and physicist Albert-László Barabási argued that networks exhibiting preferential attachment will have degree distributions that essentially follow a "power law." Power law distributions are seen in many physical situations, like the inverse square laws of gravitation and electric fields. They can be represented as functions of the form $f(x) = \frac{a}{x^k}$, and their graphs typically look like this:

PERCENT OF ALL NODES

DEGREE OF NODE
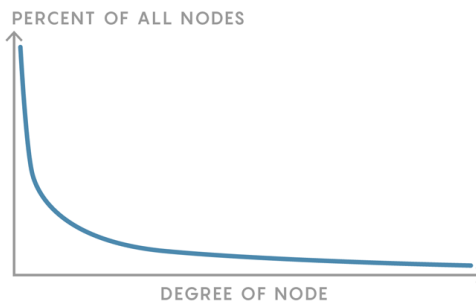
Power law distributions indeed have fat-tails. But just how fat? That is, how many hubs of each degree should we expect in such a network? A study published earlier this year analyzed 1,000 real-world networks and concluded that only a third had degree distributions that could be reasonably described by a power law distribution. Many of the networks had degree distributions that could be more accurately described using "exponential" and "log-normal" distributions. They may have the high-level features

characteristic of scale-free networks, but without the expected degree distribution, can they really be considered scale-free? And does it really matter?

It matters if we want to connect our theories to our data. Is preferential attachment really the primary factor in scale-free network formation? Or are there other factors that also play substantial roles, factors that may push degree distributions in different directions? Answering these questions, and figuring out the right questions to ask next, are part of fully understanding the nature and the structure of networks, how they develop and how they evolve.

And the controversy also reminds us that, like our networks, mathematics itself is a set of evolving connections. Contemporary research is challenging 20-year-old conjectures in the relatively young field of network science. As new ideas join the network, they connect us all to the mathematics of the past as well as the future. So when it comes to mathematics, just as in friendships, you'll do well to find the hubs and maximize your degree.

Exercises

1. What would a friendship network look like if every person had exactly two friends?
2. In Regulartown each person can have up to four friends. It's possible for cliques to form in Regulartown, small groups in which each person has exactly four friends. How many people could be in such a clique? (Hint: The answer is related to a Platonic solid.)
3. Our friendship networks rely on friendship being a symmetric relationship — that is, if A is friends with B then B is friends with A. How might we adjust our network model to accommodate a non-symmetric notion of friendship, where A could be friends with B but B not be friends with A?
4. In Friendville, everyone is friends with everyone else. If there are n people in Friendville, how many friendships are there?

Correction August 23, 2018: A previous version of this column incorrectly described exponential and log-normal distributions as having fat tails. Log-normal distributions have heavy tails and exponential distributions have thinner tails.