

THE SMALL-WORLD NETWORK OF SQUASH

BY MICHAEL KEARNS AND RYAN RAYFIELD

Not all social networks are built in front of glowing monitors with a Mountain Dew and a bag of Cheetos at hand. There are some social networks in which participation is outright good for your health—like squash. Using tools from the emerging field of network science we will investigate the specialized social network in which each node is a squash player, and there is a link between any pair of players who have played a match before.

The source data for our study was all US Squash singles matches recorded over a recent multi-year period. The number of players in this network was 26,503 and the number of matches was 240,446. The average number of matches played per player was 18.4 and the maximum was 210 (by Gabriel Bassil of Brooklyn). Like virtually all large-scale social networks, the squash network is sparse, meaning that the number of matches actually played was only a tiny fraction of those possible—less than 7 hundredths of 1 percent. It was also the case that a small number of the most active players account for a disproportionate fraction of the total matches; in network science parlance, the distribution of the number of matches across players is heavy-tailed.

To understand the global shape or structure of our network, we need to examine the *connected components*, which are the islands of connectivity.

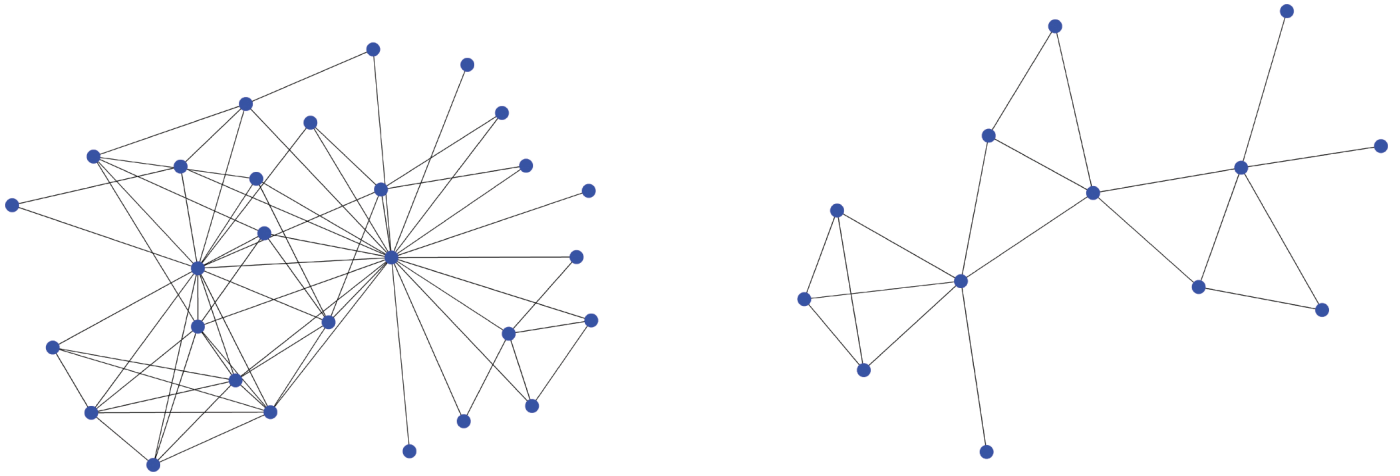
Let's consider two players as living in the same island if there is any chain of matches that connects them. So if Alice played Bob, and Bob played Charlie, and Charlie played Dana, then Alice and Dana are in the same connected component (or "island") by virtue of this chain, even if they have never played each other.

Network science predicts that in any real social network, there should be a giant component—a mainland which contains the vast majority of the population—along with an archipelago of much smaller islands with no links to the mainland. This was the case with our data. The largest component of the squash network contained almost 99% of the players. Intuitively it's hard for two large components to coexist: all it takes is one match between a player from each island and the two merge to become one.

What about the 1% of players in the archipelago, which consists of 77 additional components? What do these tiny islands look like? Unlike Facebook, playing squash requires physical proximity, so it is not surprising that many of the tiny components had a strongly geographic flavor. For instance, the second largest component had only twenty-eight players, all of whom live in Raleigh, NC, while the third largest consisted exclusively of players in San Antonio. Many of the other small components were lonely, isolated pairs of players who had only played each other. We encourage them to play more squash and join the giant component.

Not all the players in the giant component are

Visualization of the "mainland" of the US Squash network.



Visualization of the Raleigh and San Antonio connected components or "islands".

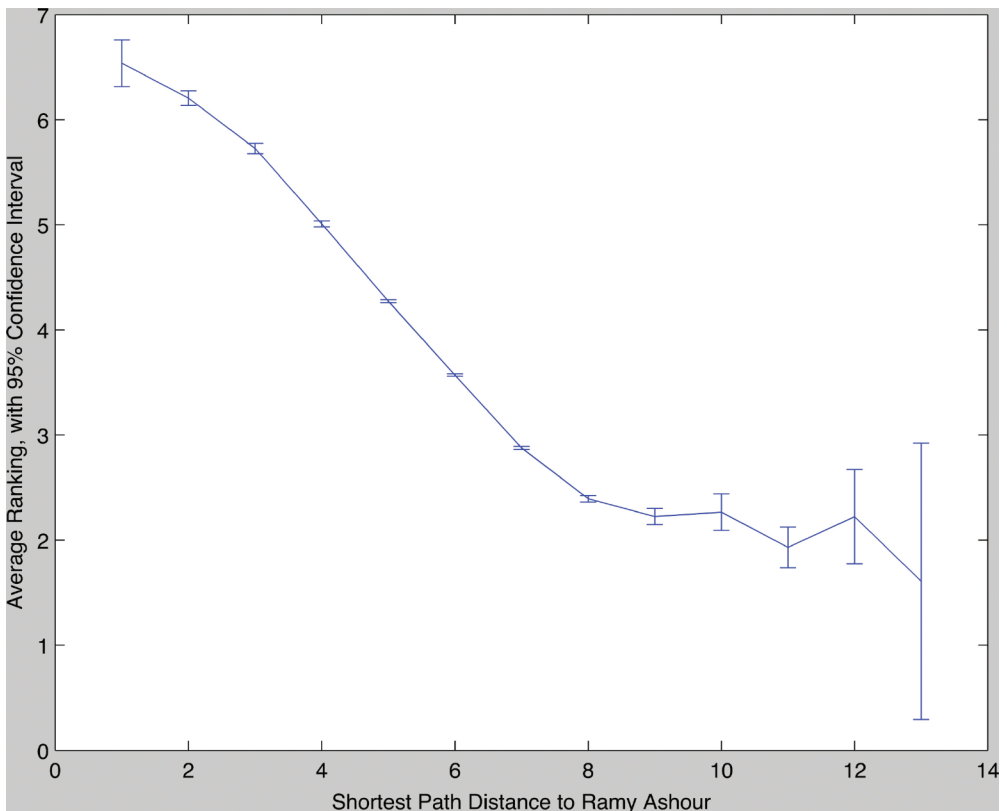
necessarily connected by short chains of matches. Indeed, the longest shortest chain between two players in the giant component was the nineteen-match chain that connected Simon Anderson of Madison, Wisconsin to Ari Wolgin of Philadelphia. But overall, the small world property does indeed hold for the squash network: on average, a chain of only 4.8 matches (or degrees of separation) connected a typical pair of players in the giant component.

Speaking of distances, it can be illuminating to consider shortest chains to particular players of interest. Some readers may be familiar with the popular parlor game "Six Degrees of Kevin Bacon," but since Kevin is not (yet) a squash player, we'll instead use Ramy Ashour. Let's define your Ashour number to be the length of the shortest chain of matches connecting you to the great Egyptian world champion. So if you're Ramy Ashour, your Ashour number is zero. If you have played a match with Ramy Ashour, your Ashour number is one. If you haven't played a match with him, but have played someone who has, your Ashour number is two. And so on.

For example, coauthor Kearns has a twelve-year old son named Gray with an Ashour number of seven, via the following chain of matches: Gray Kearns—Ben Stewart—Auggie Bhavsar—Jimmy Li—Ryan Rayfield—Chris Hanson—Alan Clyne—Ramy Ashour. Note that in another demonstration of small worlds, coauthor Rayfield appears along this path.

It turns out that player ratings increase steadily with

Distance to Ramy Ashour vs. average player rating.

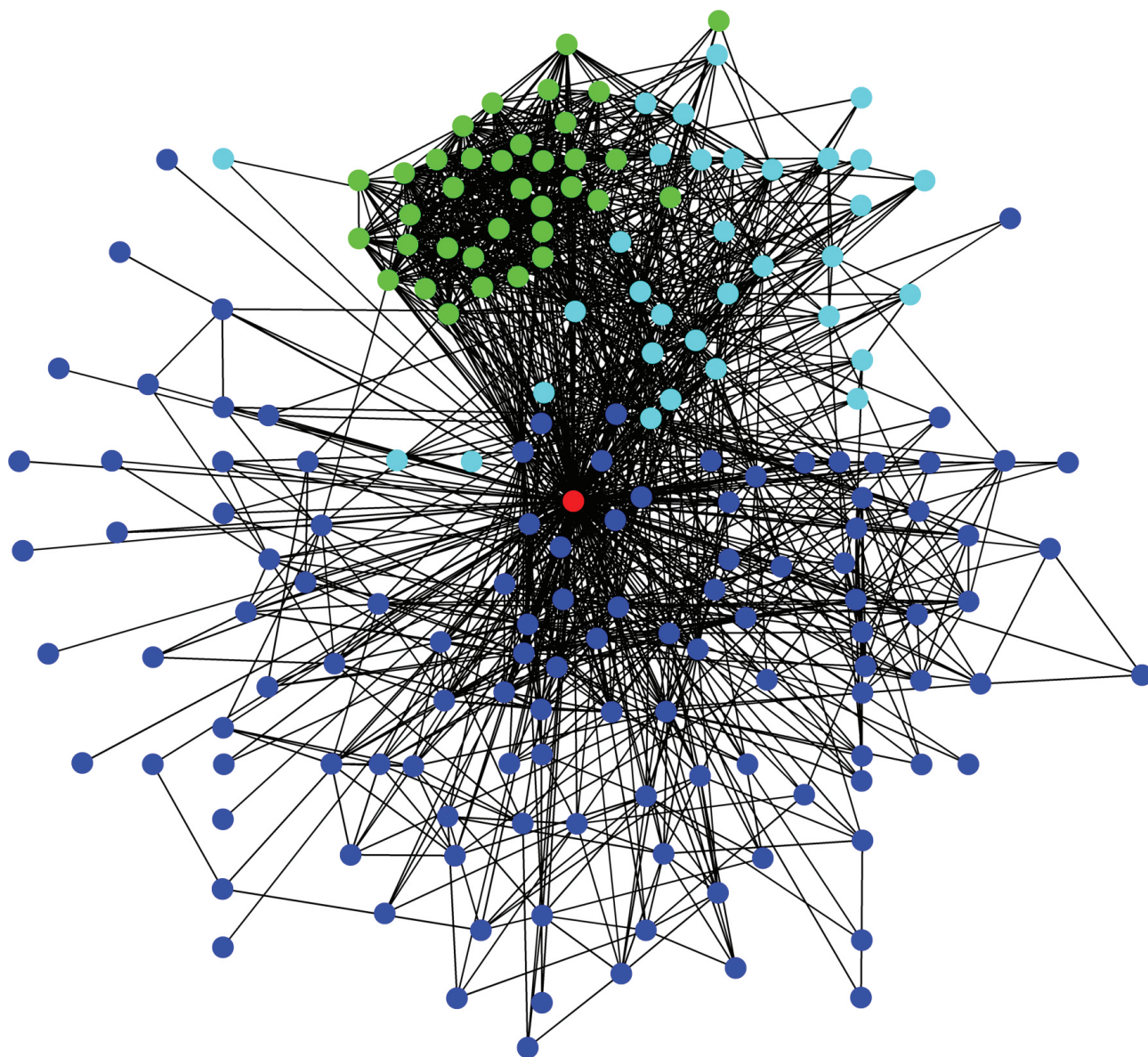


each hop along this chain from Gray to Ramy. This is far from a fluke. On average, the higher your Ashour number, the lower your rating. (In precise statistical terms, the correlation between Ashour numbers and ratings is strongly negative, -0.76 ; a correlation of -1 would mean your Ashour number completely determines your rating.) In other words, Ashour numbers are actually already a pretty good rating system *even though they entirely ignore the outcome of any matches* and only measure a kind of social distance. This is a consequence of the broad fact that our network strongly exhibits what sociologists call homophily: the concept that birds of a feather flock together. In our case this means that there is a strong bias in the network towards similarly skilled players playing each other.

While Ramy Ashour is certainly an important player, in network terms he lives in an elite and remote neighborhood that few of us will ever visit. But there are other notions of importance in network science that capture being in the middle of the network rather than in an enclave. One of these notions is known as *betweenness centrality*. This measures how many chains between other pairs of players pass through you, and thus illuminates the extent to which you are in the middle of the network, or a hub of traffic.

Unlike the players with very small Ashour numbers, players with higher centrality tend to have moderately strong rather than stratospheric ratings. These players tend to play a lot but, more importantly, they seem to play a diverse collection of opponents, both in ratings and geography.

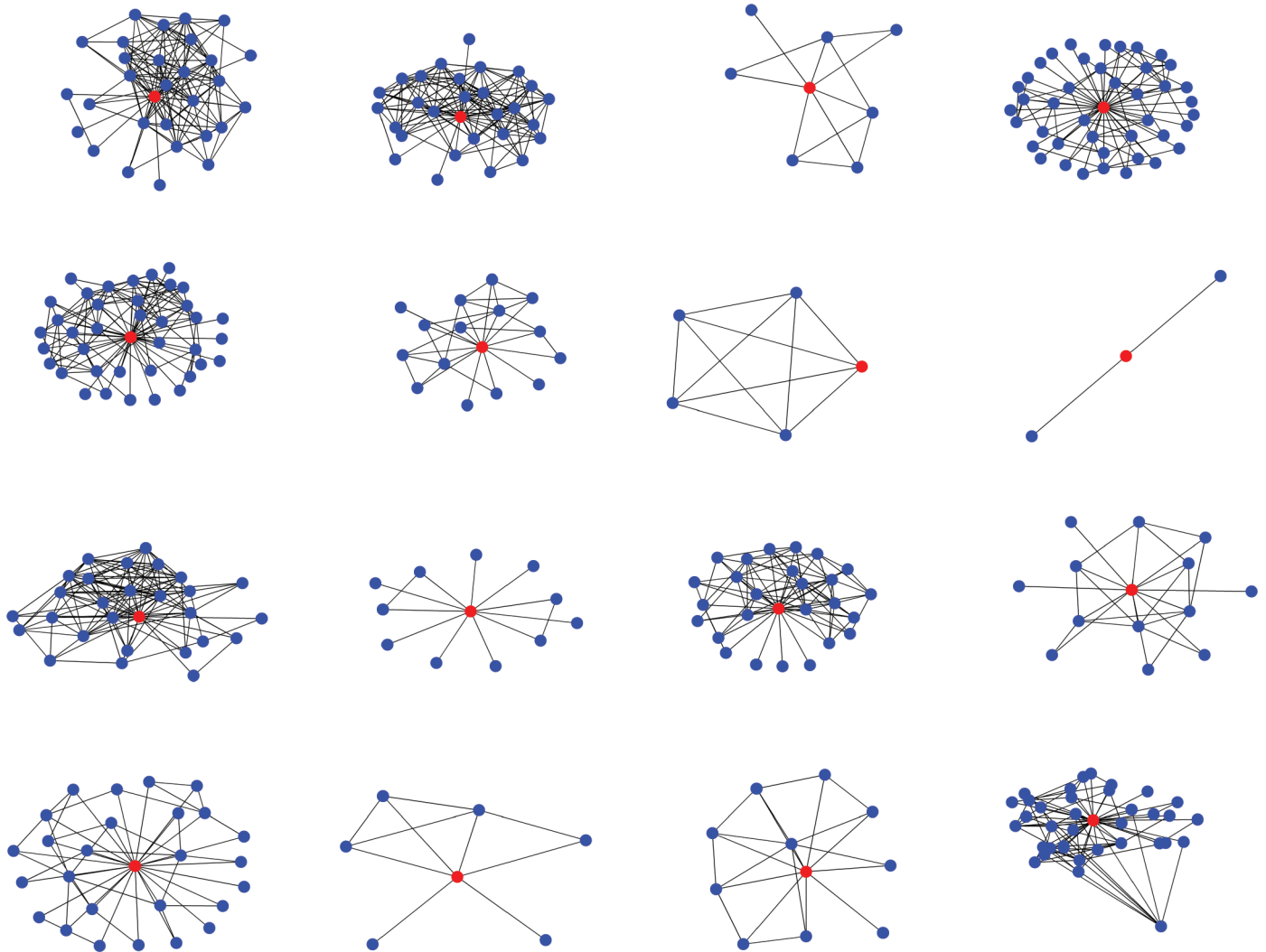
The MCP (Most Central Player) Award went to Dillon Huang, a junior player from Fremont, CA. About 2.5% of all shortest paths in the entire network passed through Dillon—roughly 650 times what you'd expect if the network consisted of entirely randomly chosen matches. In the accompanying visualization of his local or "ego" network, Dillon is shown as a red node in the center of his past opponents. We ran a standard clustering algorithm on Dillon's network, dividing his opponents into color-coded groups that played amongst themselves a great deal. The algorithm found three relatively distinct clusters, reflecting the fact that Dillon was a top junior who tended to enter tournaments for higher age brackets, as well as adult open tournaments. Who knows—as the squash network evolves, perhaps in a few years we will be discussing Huang numbers rather than Ashour numbers.



Local network for most central player Dillon Huang. Colors indicate groups of Dillon's opponents who have played against each other frequently.

What do the squash neighborhoods of mere mortals look like? Below we show the local networks for a handful of randomly selected players. In each case the sampled player is shown as a red node, all of their past opponents as blue nodes, and all matches between opponents are shown as links.

Perhaps our study and the Ashour number will inspire you to alter your network structure by playing more matches and playing with opponents you might not have otherwise—maybe even with Ramy Ashour himself.



The variety of ego network structures reflects the great diversity of player types. In addition to the obvious variation in the number of partners, we see that some players lie at the center of a squash neighborhood that is very dense (lots of matches between their opponents), and at the other extreme, there are players that seem to be the hub of a group of players, few of whom have played each other. A variety of other highly symmetric formations appear, such as a pentagon circumscribing a star. These are the crop circles of the squash universe. (But lest we become too mystical about such structures, a branch of mathematics known as Ramsey Theory predicts that any sufficiently large random network will reliably produce them.)

Local networks for randomly sampled players

Michael Kearns is a Professor of Computer and Information Science at the University of Pennsylvania and an avid recreational squash player.

Ryan Rayfield is Senior Director of Technology and Strategy at US Squash, and a former undergraduate student of Kearns.