Note that in the below calculations, "log" refers to the natural log with base e.

Let's start at the beginning with the Elo rating system, which was designed for two-player matches. We'll call those players Player A and Player B. Player A enters the match with a rating of R_A and Player B enters the match with a rating of R_B.

Next, each player is assigned an "expected score". Statistically speaking, this is the expected value of the outcome of the match; in other words, a player with a score of 0.75 is expected to win 75% of the time (more accurately, as draws are a possibility, in an infinite number of trials, a player with an expected score of 0.75 would earn 0.75 on average out of all of those trials). Note that the scores of both Player A and Player B must equal 1; hence, 1 will be awarded to the winner, 0 will be awarded to the loser, and 0.5 will be awarded to both players in the event of a draw. The exact calculation of those scores is as follows. Player A's expected score, or E_A, is:

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

And similarly, Player B's expected score, or E_B, is:

$$E_B = \frac{1}{1 + 10^{(R_A - R_B)/400}}$$

Finally, Player A scores S_A points while Player B scores S_B points; as aforementioned, these must be equal to 1. Player A's new rating, R'_A, is:

$$R'_A = R_A + K(S_A - E_A)$$

And the same for Player B. In other words, Player A gains or loses K times the number of points that they earned above or below expectation. (For instance, if both players entered with the same rating, Player A would be expected to win 0.5 points; if they won the match, earning 1 point, they would earn 0.5*K toward their rating.) K serves as a "value" for the match; we'll address how we got K in a moment.

The Power Ranking rates each competitor against each other competitor in the competition. For instance, let's say an athlete finishes 5th out of 20 competitors in their division. They'll be compared against all 19 other competitors, having effectively beaten 15 and lost against 4. In total, there would be 19 matchups per competitor and 20 competitors for a total of 380 matchups. Don't worry, a computer will do this in the blink of an eye!

Due to the volatility of the sport of ninja, we found that this system, left to its own devices, too harshly penalized athletes for below-expectation performances. All ninja athletes have fallen earlier than expected in their career; this is simply a part of ninja. In order to counter this, if an athlete loses points, their rating will only be subtracted by 80% of what they otherwise would have lost. Put mathematically,

$$R'_A = R_A + 0.8K(S_A - E_A) \; if \; K(S_A - E_A) < 0$$

Let's go back to K from the above equation. Because some competitions are larger than others, the larger competitions would have the potential to be too heavily weighted if all matchups had the same value. That said, it is obviously more difficult to win a larger competition than a smaller competition, and we wanted the Power Rankings to reflect this. As a result, we came up with this value for K:

$$K = \frac{100\,((\log N) + 1)}{N}$$

Where N is the number of participants in a competition.

Finally, the adjustment factor affects the points greater than or less than 1500. (For instance, if your unadjusted rating is 1700 and the adjustment factor is 2, your score would be 1500 + 200/2 = 1600.) The adjustment factor starts at 1 after the athlete's first competition and increases by 0.05 for each subsequent competition. Put mathematically,

$$A = 0.95 + 0.05C$$

Where A is the adjustment factor and C is the number of competitions the athlete has participated in.

While certainly not the simplest mathematically, we tested a number of different formats and decided that using an existing protocol in the Elo rating system while modifying it slightly to better suit the needs of ninja as a sport got us exactly what best reflected our desires for the Power Rankings.