Navigating a Small World

Q14.1 Why does each triad closure have three different connected triples?

Take a look at Illustration 37. Basically, we can obtain a *unique* connected triple by removing any of the three links `(A,C)`, `(B,C)`, and `(A,B)` one at a time. Removing the first leaves A connected to B, and B to C. So, we obtain `(A,B,C)`. Removing the second and third we obtain `(B,A,C)` and `(A,C,B)`, respectively. These are the only unique triples in the triad: `(B,C,A)` uses the same links as `(A,C,B)`, and `(C,A,B)` uses the same as `(B,A,C)`. We only want to count each of them once.

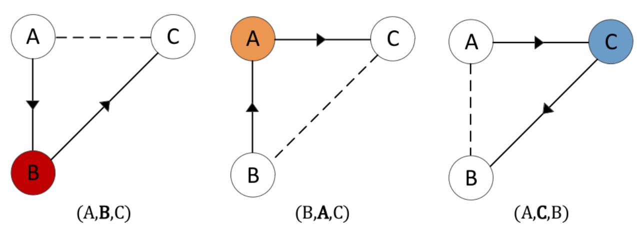

Illustration 37: Explanation of why there are three connected triples in each triad closure. We can remove any one of the three segments, and obtain a unique connection among the three nodes A, B, and C. On the left, we remove `(A,C)` and keep `(A,B)` and `(B,C)`, making B the midpoint of the connected triple `(A,B,C)`. Similarly, we have `(B,A,C)` in the center and `(A,C,B)` on the right. We have kept the center point colored/shaded in each case.

Q14.2 What is the formula for the clustering coefficient of a regular graph?

In general, the clustering coefficient can be found as

$$\frac{3 \times (\text{links/node} - 2)}{4 \times (\text{links/node} - 1)}.$$

We can try two test cases to verify this: with 2 links/node, we get zero. This is the smallest case possible where as we said, we only have a circle of links, so indeed the clustering coefficient should be 0. With

4 links/node, it jumps up to $(3 \times 2)/(4 \times 3) = 1/2$, as we calculated in the book. As the links per node gets very large, this equation will go to 3/4, as we expected too.

Q14.3 How can social search be modeled?

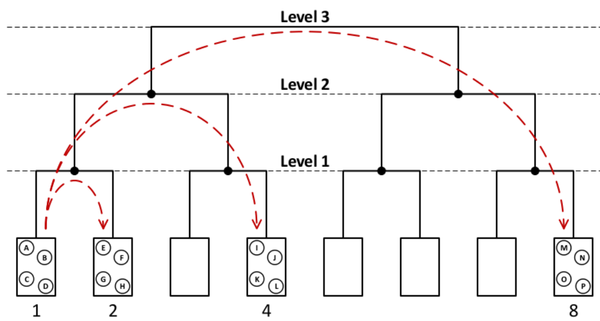

Illustration 38: An example of the Watts-Dodds-Newman model. At the bottom of the tree, we have eight leaves. Each leaf has complete, direct connections between all people inside, *e.g.*, A, B, C, and D are all connected in Leaf 1. Each person also has a certain chance of having a link to those outside of its direct leaf, but the probability of this being the case decreases as the first level of common ancestry between the nodes gets higher. For instance, A and E's first level of common ancestry is Level 1, A and J's is 2, and A and M's is 3. There's a greater chance of A and E being connected than A and J, and this is in turn more likely than A and M.

There have been several models for social search in the past decade beyond the original Watts-Strogatz model. We will examine one here in particular: the **Watts–Dodds–Newman Model**. Here, people live in different "leaf nodes" of a **binary tree**, as depicted in Illustration 38. In this tree, there are eight leaf nodes (labeled 1 – 8) at the bottom, and within each, all people are connected directly to one another. Moving up the tree, we have increasing levels of *ancestry*, which include more and more of these leaf nodes. There are four ancestry nodes at level 1, each of which contains two groups of leaf nodes (1 & 2, 3 & 4, and so on). There are two at level 2, both of which contain four groups of leaf nodes (1 – 4 & 5 – 8). Finally, the third level encapsulates all the leaves. Adding more and more leafs, which must be done in a multiple of two to preserve symmetry, just adds ancestry levels. So 16 nodes requires 4 levels, 32 requires 5, and so on.

All the people within the same leaf are directly connected, as we said. Thinking in terms of Milgram's experiment, we expect them to all be close in terms of both geographical and occupational proximity, and clustered accordingly. To analyze how far two leaves are from each other, we have to look at the level of *first common* ancestry. Person A (in Leaf 1) and Person E (in Leaf 2) are first connected at Level 1. Person A and Person I (in Leaf 4) are first connected at Level 2, and so on. As this level increases, the probability that the two nodes know each other gets smaller. So A and B, being in the same leaf, are guaranteed to know each other. A and E are not, but are more likely than A and I, which is in turn more likely than A and M (in Leaf 8). The higher this level becomes, the *further* the people are from one another, in terms of geography and occupation. But it's still possible that they know one another: this encompasses the random-adding of links in Watts-Strogatz.

This model sounds quite intuitive, but we still have to show that the path discovered by greedy social search exhibits small world. And what exactly does this mean? Well, without going into mathematical detail, it suffices to say that the discoverable path length must grow *very slowly* as the number of nodes in the graph increases. So as we increase the number of leaf nodes from 8 to 16, 32,...,1024, 2048 (powers of two) to millions and millions, this path length must remain relatively small.

As it turns out, if we specify how the probability of people knowing each other will decay as the level of common ancestry goes up, we can achieve this. In particular, the probability must decrease by *exactly one half* between one level and the next. For example, if there are 3 levels, the probabilities at 1 - 3 must be roughly 57.1%, 28.5%, and 14.3%, since $28.5 = 0.5 \times 57.1$, and $14.3 = 0.5 \times 28.5$ (note they must add up to 100%). From our three-level graph in Illustration 38, this means that the probability A and E are linked must be 57.1%, that of A and I 28.5%, and that of A and M 14.3%. If we changed the probability of decrease from 50% to, say, 60%, these three would change to 64.1%, 25.7%, and 10.3%, again since $25.7 = (1-0.6) \times 64.1$, and $10.3 = (1-0.6) \times 25.7$.

In Illustration 39, we show just how important this one-half decrease is.

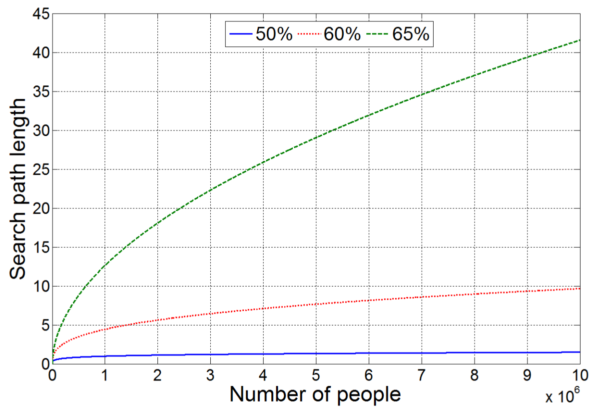

Illustration 39: Impact of changing how the probability of people knowing each other goes down with increasing ancestry, in the Watts-Dodds-Newman model. For 50%, we see the search path length barely increases as the number of people increases to 10 million. For 60%, the increase is perhaps still tolerable, but at 65%, it increases drastically. This highlights the importance of keeping it at 50%.

In this example, we assume that there are 50 people in each leaf-node, and that each person is, on average, connected directly to 100 people (so 50 in their leaf, and 50 outside). We plot three cases: the ones when the decreases are 50%, 60%, and 65%. In each, we have plotted the search path length as the total number of people increases to 10 million. For 50%, the search path length only increases to about 1.5 at 10 million people, which is close to nothing. For 60%, it increases to about 9.7, which a lot higher than 1.5. And for 65%, it goes all the way up to 41.6, which is just way too large.

In this numerical example, as long as we keep the decrease at 50%, we can guarantee that the greedy search length will exemplify small world for the Watts-Dodds-Newman model. We can thus conclude that Milgram's observed six degrees is indeed discoverable in realistic networks.