How Network Theory Predicts Bias

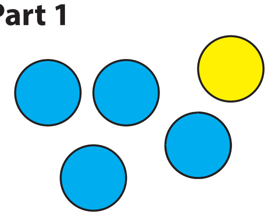
Let all nodes be either blue or yellow. Nodes evaluate other nodes that are either blue or yellow at random. *Yellow nodes ignore **all** blue nodes -- accepting and passing through only nodes of their same color. Blue nodes accept nodes of **any** color and pass them through to the next level of the group.*

Nodes here are used to represent people. The color yellow represents a single arbitrary category of bias (ex. racism, sexism, classism) that is held only by a subset of group. The group is defined to contain "qualified individuals within a given organization or social network." All individuals in the groups examined (including the biased ones) are assumed to be qualified.

We are also not assuming scarcity or competition; e.g. there are enough slots for everyone; the yellow nodes do not "lose" by having more blue nodes participate.

It is also important to note we are making one large assumption (**only 1 in 5 members of a group are biased**) and that we are looking only at the effect of selection bias in networks over time. **We are *not* making any claims as to the relative merit of any particular node, but simply assuming that all nodes in the group meet the base criteria for approval.**

Part 1



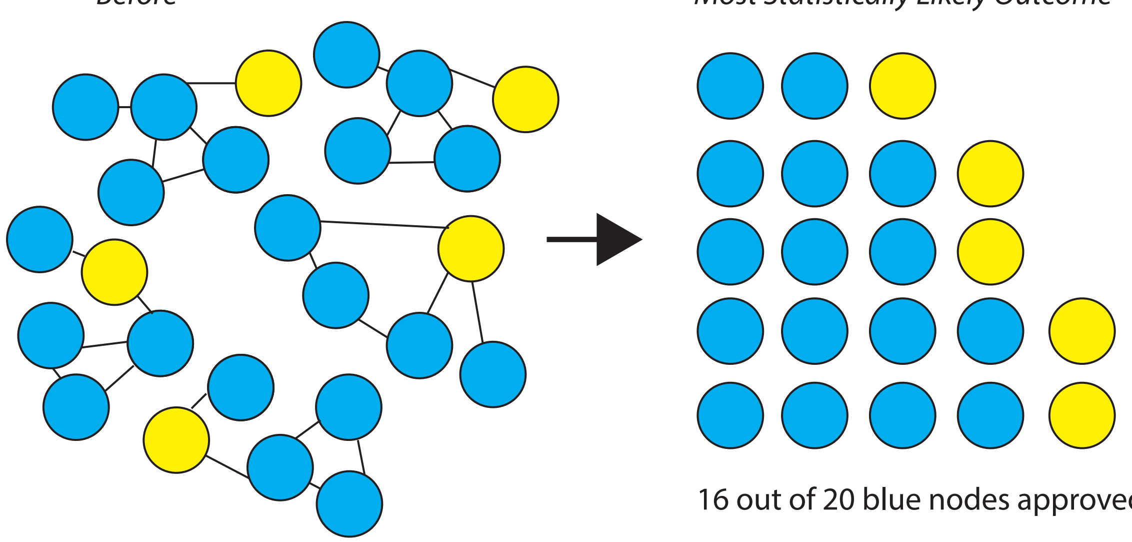
Five Nodes.
Each node must be randomly approved
by one other node to pass through to next round.

Probability of an individual yellow node getting approved: 100% (4 out of 4)
Probability of any given blue node getting approved: 75% (3 out of 4)
Probability of at least two blue nodes getting approved: 94.92%
Probability of at least three blue nodes getting approved: 73.83%
Probability of all four blue nodes getting approved: 31.64% (¾ · ¾ · ¾ · ¾)
Probability of no blue nodes at all getting approved: .0039% (¼ · ¼ ·¼ ·¼)

Part 2

Now assume that we have five such clusters of nodes (5 clusters of 5 nodes each) all sharing the same distribution and behavior of blue and yellow nodes.

Before — *Most Statistically Likely Outcome*



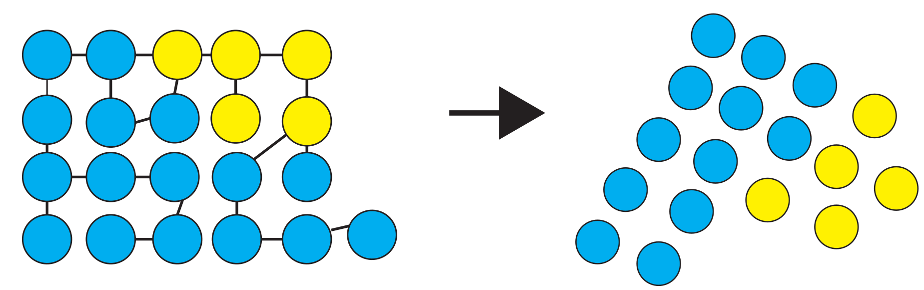
16 out of 20 blue nodes approved.

25 Nodes Total, 80% Blue

New Total: 16 Blue Nodes, 5 Yellow
21 Nodes Total. 76.19% Blue

Part 3

What happens if you run the same selection process through this new group?



Probability of all 5 yellow nodes getting approved: 100% (20 out of 20)
Probability of any given blue node getting approved: 75% (15 out of 20)
Probability of all qualified blue nodes getting approved: 1% (3/4 ^ 16)
Most likely statistical outcome: 12 out of 16 blue nodes approved. (22.19%)

17 Nodes Total. Now, only 57.14% of nodes are blue.

The proportion of blues in each round drops. 80% > 76% > 57%

In just two selection rounds, we have lost almost one third of all nodes.

These results come from a system where all individuals are defined to be qualified and **where the majority of actors are non-biased.** The model does not make moral judgments, or even speculate on what type of bias is in play. It simply observes the effects of bias in a given proportion of the population.

Blue nodes cannot avoid interactions with a a biased node, nor can they conclusively prove bias or retaliate in a future round. Bias is either not detected or not acknowledged by other group members. This is very different from the scenario of tribalism, where all group members are assumed to be biased and to act out their bias.

Hidden bias can be difficult to identify, even when its effects are severe. Alarmingly, we see evidence that open, decentralized systems are uniquely vulnerable to these effects.

Alternate Models

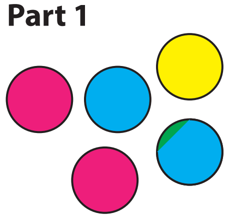
What if the group overall is less biased? If we assume that only 1 member in 10 of a group is biased, we still are likely to lose at least one blue node during the first selection round. During the second round, the probability of any given blue node getting approved remains about the same.

So far we have only looked at networks with only two categories of individual: biased and non-biased. These models do not make any distinction between types of bias (e.g. one cluster might contain an individual who is sexist and homophobic; a second cluster might contain an individual who is biased against members of a certain religious group). It only looks at two categories: those who accept all individuals based on their qualifications, and those exclude a certain individuals for reasons not related to their qualifications. We believe this is a relevant and realistic set of behavior patterns. Still, we recognize that the model may feel a little abstract for some.

What if we seek to model hidden bias specifically in terms of gender?* To create a system that more closely approximates gender bias in the real world, we introduce a third category, breaking down our model of qualified individuals into the following:

- Those who are biased. When given the chance to "vote," they approve all yellow or blues. They never approve a pink. (Biased Men)
- Those who are not biased and do not themselves experience the effects of bias. When given the chance to "vote," they approve yellows, other blues, and pinks. (Non-Biased Men)
- Those who are not biased and experience the effects bias.* When given the chance to "vote," they approve yellows, other pinks, and blues. (All Women)

Part 1



Five Nodes.
Each node must be randomly approved by one other node to pass through to next round. Proportion of biased nodes is the same. Proportion of nodes who may experience bias now 2 out of 5.

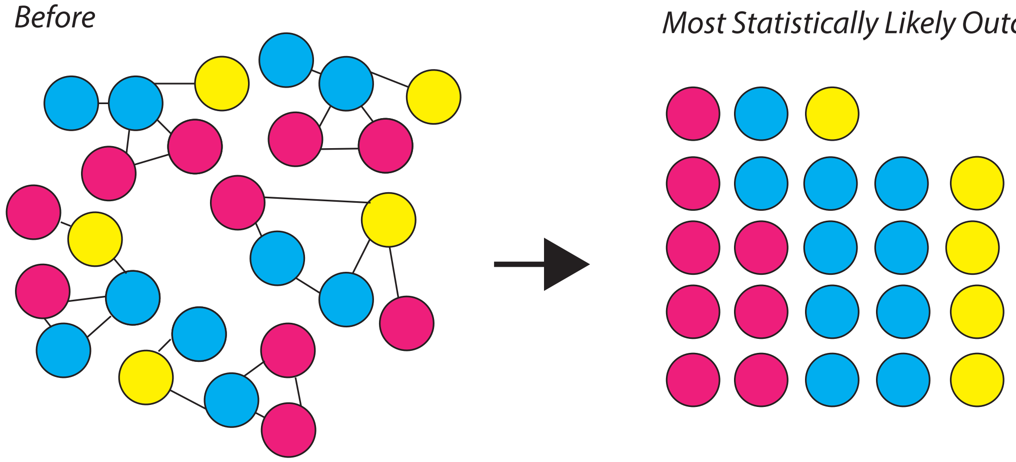
Probability of all yellow and blue nodes getting approved: 100% (4 out of 4)
Probability of one pink node getting approved: 75% (3 out of 4)
Probability of both pink nodes getting approved: 56.25% (**¾ · ¾**)
Probability of no pink nodes at all getting approved: 6.25% (¼ · ¼**)**

So far, so good. We have the same probability that one node will be approved, and a higher probability of no nodes getting dropped. But what happens if we look at more nodes?

Part 2

We assume that we have five such clusters of nodes (5 clusters of 5 nodes each) all sharing the same distribution and behavior of pink, blue, and yellow nodes.

Before — *Most Statistically Likely Outcome*

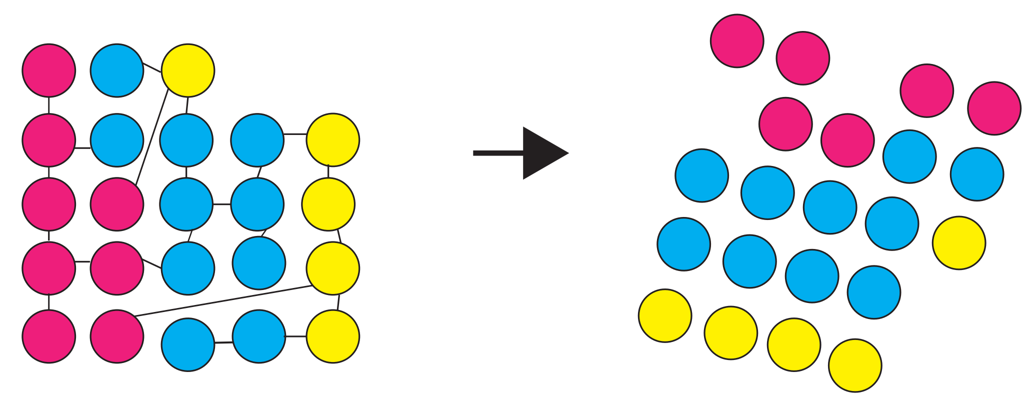


25 Nodes Total, 40% Pink
10 Pink, 10 Blue, 5 Yellow

23 Nodes Total, 34.78% Pink
8 Pink, 10 Blue, 5 Yellow

Part 3

What happens if you run the same selection process through the new group?



Probability of all 5 yellow and blue nodes getting approved: 100% (22 out of 22)
Probability of any given pink node getting approved: 77.27% (17 out of 22)
Probability of all qualified pink nodes getting approved: 12.71% (17/22 ^ 8)
Most likely statistical outcome: 6 out of 8 pink nodes approved. (30.79%)

21 Nodes Total. Now, only 28.57% of nodes are pink. This doesn't look so good either. **40% > 34.78% > 28.57% In just two selection rounds, we have lost nearly half of all pink nodes.** And there were not a lot to begin with.

Next Steps

Before proceeding further, we need to look at iterations involving larger sample sizes:

- Look at effects for N rounds when voting cluster size stays small (5 nodes or less) and each round contains more than one cluster. The composition of clusters in progressive rounds will be determined based on the statistical distribution predicted from voting patterns in the first round. Assumes a limited number of nodes whose future interactions can be traced back to the initial selection round. (*Workplace Model*)
- Look at effects for N rounds with increasingly large voting clusters. The composition of each new "supercluster" will reflect that of the most statistically likely outcome. Assumes a limitless supply of nodes, all conforming to the same distribution and behavior patterns. (*Twitter Model*)
- Look at the effects of reducing the percent biased to 10% from 20%.

* We do not discount the possibility that members of a group experiencing bias may at times discriminate against members of the same group. We will not examine this scenario further, except to say that in all cases, this outcome is worse for groups and individuals who wish to avoid bias.

Additional Background and Context

A less intuitive, but more mathematically precise way to explain the models on the previous pages would be to say that each node casts a total of 4 votes, one for each other node, but the order in which they reach the other nodes is random. All votes are counted, even if the node casting the votes is excluded in the following round.

> **If a node receives a positive vote, which the yellow nodes in all cases will, they are "in."**
>
> **If a node receives even one negative vote before they get a positive vote, they are "out."**

This work is a response to James D'Amore, aka "The Google Guy" and his spurious claims that women are biologically unsuited to be programmers.

Here is a hypothetical workplace example:

For instance, **Agreeable Alice** is in a workplace situation. She attends her first meeting. Somebody who is influential (**Biased Brian**) delivers negative feedback. She gets passed over for the opportunity to work on a more challenging project.

Meanwhile **Blameless Bob** presents at the next meeting. **Biased Brian** says only good things about Bob. Bob gets the opportunity to work on the more challenging, high visibility project. Because Bob, like Alice, is highly qualified, he quickly moves through the organization's ranks.

Alice, meanwhile, finds herself in a role with little opportunity for advancement or personal growth. Frustrated, she leaves the company. If Alice had gotten feedback from somebody other than Biased Brian, she would have been fine. But because Biased Brian spoke up first, she never got that opportunity.

*We hope that the models and examples contained herein are interesting and useful as starting points for further research and examination. They are not meant to be definitive, but we believe they shed light on some important phenomena -- **namely, how can a system in which the majority of actors are non-biased consistently produce biased results?***

We feel that network and probability theory provides a more convincing explanation than either biological differences or the blanket statement that all members of a privileged class are prejudiced.

*We do not wish to disparage open, decentralized models or suggest that they are not effective. However, preliminary evidence **strongly** suggests that when the nodes passing on information and making decisions represent people, **a totally distributed social organization will tend to favor more exclusionary, polarized viewpoints over time.** We feel that this phenomenon is worthy of attention above and beyond admittedly important questions of gender equity.*

This poster and associated preliminary models by Tess Gadwa, with input and consultation from Joe Benson, Ph.D., and Ron Breukelaar, Ph.D. Thanks to the crew at Matrix.org for being an informal test audience!