Talia Wise
260659717
LLCU 255
December 11, 2017

Programmer – Man + Woman = **?**

Associations Between Gender and Technology on Anonymous Internet Forums

In 2012, Ellen Pao, a partner in a Silicon Valley venture capital firm, filed a discrimination lawsuit against her employer for not promoting her because of her gender. Her lawsuit has been characterized as the event that brought gender discrimination in the tech world to the public eye.[1] The question of whether biases against women and minorities exist in the tech world and whether these biases are harmful has been heavily discussed over the past half-decade.[2] Google Trends shows that the number of searches for 'women in tech' remained steady from 2004 to 2012 but after lawsuits by women in tech companies and events such as the Google 'memo' last August, the number of searches for this term quadrupled since 2012.[3]

Although it is undoubtedly true that women are less represented in technical roles, are payed less, and are less likely to be in upper management positions in tech companies, the causes of these imbalances are far more difficult to measure. To explain the overrepresentation of white men in tech companies, Silicon Valley has been recently promoting the notion of 'implicit bias': the idea that every one of us holds unconscious biases towards certain population groups.[4] Due to such implicit bias, when we look at a male and female applicant for a technical position, even though we might not be aware of it, our gut reaction will be that the man is more suited to the role because he more closely matches our implicit notion of a technical worker. This probabilistic and associative bias is captured in language as well. It has been demonstrated to exist in word-embeddings and is even amplified by common machine learning algorithms.[5] In this paper, I seek to explore how gender bias is encoded in discussions about technology and the tech world in

---

[1] "Ellen Pao is not done fighting" https://nyti.ms/2gR6Z0i

[2] "The tech industry's gender-discrimination problem" https://www.newyorker.com/magazine/2017/11/20/the-tech-industrys-gender-discrimination-problem

[3] https://trends.google.com/trends/explore?q=women%20in%20tech

[4] http://kirwaninstitute.osu.edu/research/understanding-implicit-bias/

[5] Bolukbasi, Tolga, et al. "Quantifying and reducing stereotypes in word embeddings." *arXiv preprint arXiv:1606.06121* (2016).

anonymous internet forums. From a technological perspective, it is important to know how bias is encoded in order to prevent amplifying this bias in systems that use word-embeddings such as resume rankings and algorithms that search for potential employees.

I trained Word2Vec on 1.6 million recent comments from Hacker News, a popular anonymous technology forum frequented by Silicon Valley employees. I then trained two more Word2Vec neural networks on recent comments relating to technology from two popular subreddits, AskMen and AskWomen, in order to compare the discursive space of the (supposedly) non-gendered Hacker News' associations between gender and technology with the associations between gender and technology in these two explicitly gender-oriented and even gender-segregated subreddits.

I used Trevor Martin's Subreddit Algebra algorithm in order to test associations between gender-oriented and technologically-oriented spaces on Reddit by measuring subreddit similarity.[6] Since Reddit is an anonymous forum, it is impossible to count how many users of each gender participate in any given subreddit, but it is possible to compare the ratios of participation between gender-focused subreddits and other types of subreddits. Using Subreddit Algebra I found a clear difference between gendered participation in AskMen and AskWomen. Despite being identically structured, AskMen is only the 59th most similar subreddit to AskWomen according to cosine similarity of commonly co-occurring subreddits. On the other hand, AskWomen is the third most similar subreddit to AskMen, which means that an AskMen member is more likely to be connected to AskWomen than an AskWomen member to AskMen. Furthermore, the top hundred subreddits similar to AskMen are not all male oriented while most of the top hundred similar subreddits to AskWomen are clearly female oriented so it is fair to assume that AskWomen is a more female

---

[6] https://github.com/fivethirtyeight/data/tree/master/subreddit-algebra and https://fivethirtyeight.com/features/dissecting-trumps-most-rabid-online-following/

space than AskMen is a male space. Adding or subtracting technology-focused subreddits from these two subreddits shows no clearly discernable difference between the two subreddits' connectedness to technological subreddits. Both subreddits show roughly the same amount of technology related comments (20,000 and 18,000 respectively in 2015).[7]

Vector algebra on subreddit vectors such as 'Tech – AskMen + AskWomen' returns nearly the same result on both the male and female subreddits, because the gendered subreddits are far more related to each other than they are to technology focused subreddits. This phenomenon can be seen in the graph below, which shows the 124 most commonly subscribed subreddits in k-cliques. AskMen and AskWomen as well as other gendered subreddits appear in the black box on top, while all the technology focused subreddits are in in the purple box in the middle. It is notable that AskMen and AskWomen are on different sides of the gender box but neither is significantly closer to the technology area. (This graph may be imagined as a network graph where both color and proximity are measures of connectedness.)
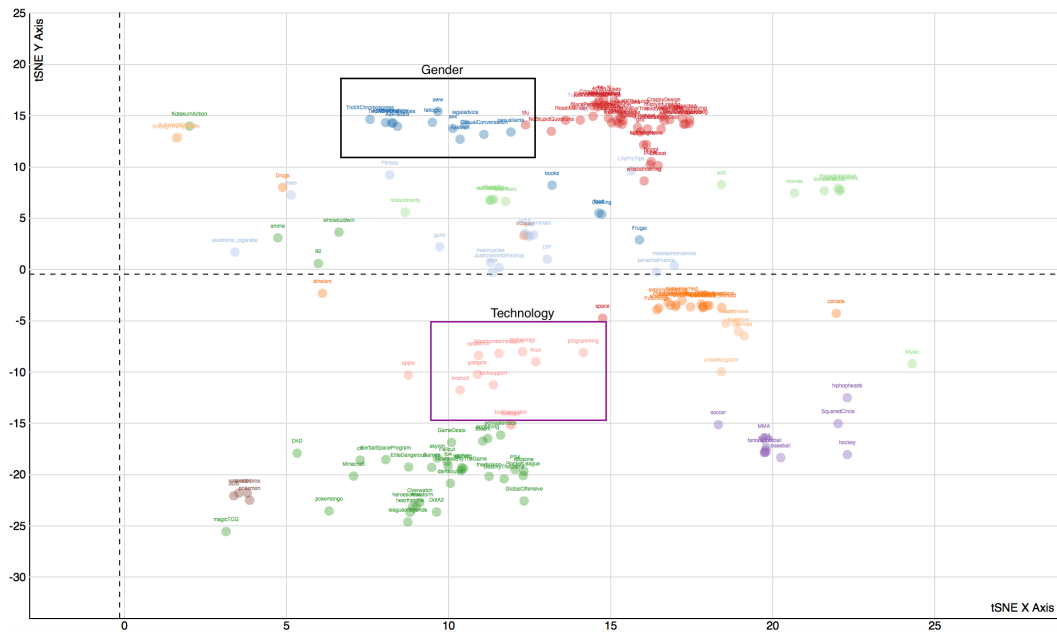
[7] I extracted comments with at least one technology-related word (technology, tech, computer, programmer, developer, etc.) from the two subreddits.

The reason I did not compare associations between gender and technology in technologically oriented subreddits is that, as shown in the graph above, the technologically oriented subreddits are very strongly connected, so subreddit algebra on them always yields other technology, gaming or science subreddits. Furthermore, most of the computer science and technology subreddits have strict rules against discussing non-technological topics, so there is no reference to gender or anything social in these subreddits at all.

Because Hacker News is a less strictly moderated forum and more business and socially oriented, it presents a more interesting space for an analysis of gender and tech discourse. The word vector for 'man' contains a fairly unexpected list of top ten words: {man boy woman hugged Pharisees straw jesus girl rapist}. 'Man', 'boy', 'woman' and 'girl' are all coherently connected, and upon searching the Hacker News site it is clear that 'straw' refers to a strawman argument. The fact that the neural network learned that 'Jesus' and 'Pharisees' are connected to 'man' is puzzling. Upon examining their vectors, it seems that these words appear because of a combination of evangelization and religious jokes. One might also attribute this to overfitting. 'Hugged' seems to be connected to a happier masculinity with words like 'protector' and 'boyhood' (in contrast with the very negative association to 'rapist'). The vector for 'woman' is more coherent, including: {woman girl male female raped manly men husband objectifying}.

Interestingly, the plural 'men' and 'women' appear in a different discourse than 'man' and 'woman', likely because of their generalizing capacity. These plural forms especially appear in connection to feminism and identity politics. The most similar vectors to 'men + women' are: {men women females males male girls female sex heterosexual feminists minorities stereotypes blacks boys underrepresented adults rapists hetero marriages gender}.[8] In order to further explore these

---

[8] I chose to use the notation 'men + women' or 'programmer – google' in order to represent instances of vector algebra.

differences, I also subtracted singular forms from the plural forms of gender and found that 'female – woman' generates qualities and actions that commenters describe as female, including: {varying skillsets essayists classifications preoccupations applicants disciplines backgrounds}. However, 'male - man' and 'men – man' simply returns the same vectors of gender and identity politics as does 'men + women' (described above). 'Wom*en* – wom*a*n' returns places and groups that women might be part of: {groups academics communities universities colleges underrepresented multipliers professions categories innovations schools essayists disciplines nonprofits}. While neither the word vectors generated by 'female – woman' or 'women – woman' are particularly negative towards women, and in fact most can be seen as positive, the contrast between femaleness and maleness in this dataset shows that femaleness is often used as a qualifier for women's participation or presence while male-ness is only brought up in conversations about gender. While maleness is occasionally connected to very negative attributes such as 'rapist,' these vectors are connected specifically to discussions of gender.

Returning now to the title of this paper, we may use the vector calculation 'programmer – man + woman' to find words associated with 'woman' in the same way that 'programmer' is associated with 'man'. The table below shows notable words from each resulting vector:

Female		Male	
Programmer – man + woman	Programmer – men + women	Programmer – woman + man	Programmer – women + men
Nurse	Employable	Craftsman	Professional
Unemployable	Hirable	Newbie	Interpreter
Competent	Autodidacticism	Interpreter	Dweeb
Mentor	Mediocre	Sysadmin	Naïve
Generalist	Skillset	Polyglot	Codemonkey
Salesperson	Unemployable	Naïve	Typist
Teacher	Tinkerer		Craftsman
Overqualified	Generalist		Smalltalker
Recruiter			Neophyte
Employable			

'Nurse', 'salesperson', 'teacher' and 'recruiter' are to 'woman' as 'programmer' is to 'man', while 'craftsman', 'sysadmin' and 'typist' are to 'man' as 'programmer' is to 'woman'. Male programmers are also viewed more readily on a journey from neophyte to professional, while female programmers are evaluated as either hirable or unemployable. Altogether, this does not paint, however, an entirely negative picture for women over men and, unfortunately, there seems to be bias towards male programmers as well. Male programmers may be perceived as 'dweebs', 'naïve' and 'codemonkeys', while their positive attributes are 'professionalism', 'craftsmanship' and 'polyglot'. Women, on the other hand, are not evaluated negatively based on personal qualities, but rather on 'employability' and they are positively perceived as 'generalists', 'competent', 'autodidactic' and 'tinkerers'.

By contrast, in both AskMen and AskWomen, there is little difference between man, male women, etc. All these gendered words return vectors that have some element of identity politics but mostly have other gendered words and words that suggest people are speaking about their male or female family and friends. The vector for programmer in AskMen added to the vector for programmer in AskWomen generates a vector with words related to technology and occupation: {programmer mainframe researcher machinist engineer mathematician developer citizen chemist consultant}.[9] 'Programmer – man + woman' or 'programmer – woman + man' also return lists of occupations in both subreddits. The lists are fairly similar and do not have obvious gender biases, so I have not included them, but let us happily note that occupation does not seem to be nearly as gendered in these forums.

---

[9] Subtracting the vector for programmer in AskMen from the vector for programmer in AskWomen and vice-versa gives different results when the main model matches the subtractor and subtractee vector. I suspect that this sort of cross-model vector algebra might actually allow for interesting analyses of words in different contexts but I don't currently have any understanding of what this sort of cross-model vector algebra might be doing so I hesitate to include it in this paper. I do intent to explore this as a possibility in the future because especially in a literary context knowing the difference between associations to a word in different discourses is hugely useful and being able to combine this with other aspects of vector algebra might result in interesting results.

Another way in which Hacker News differs from Reddit is in the correlation of attributes of excellence with attributes of gender. When plotting 75 words most similar to 'good' and 'bad' based on gender affiliation, I found that Hacker News commenters do indeed value women less than men while this bias is less prevalent in the two subreddits. The graph below shows the Hacker News data. The x-axis is female on the right and male on the left and the y-axis tracks 'good' on top and 'bad' on the bottom. While the positive attributes veer towards the male side, with 'brilliant' and 'clever' being the most male positive attributes, and 'decent' and 'tough' the most female ones, negative attributes are mainly on the female side (and most of the negative attributes in the male side do not seem like actually negative words).
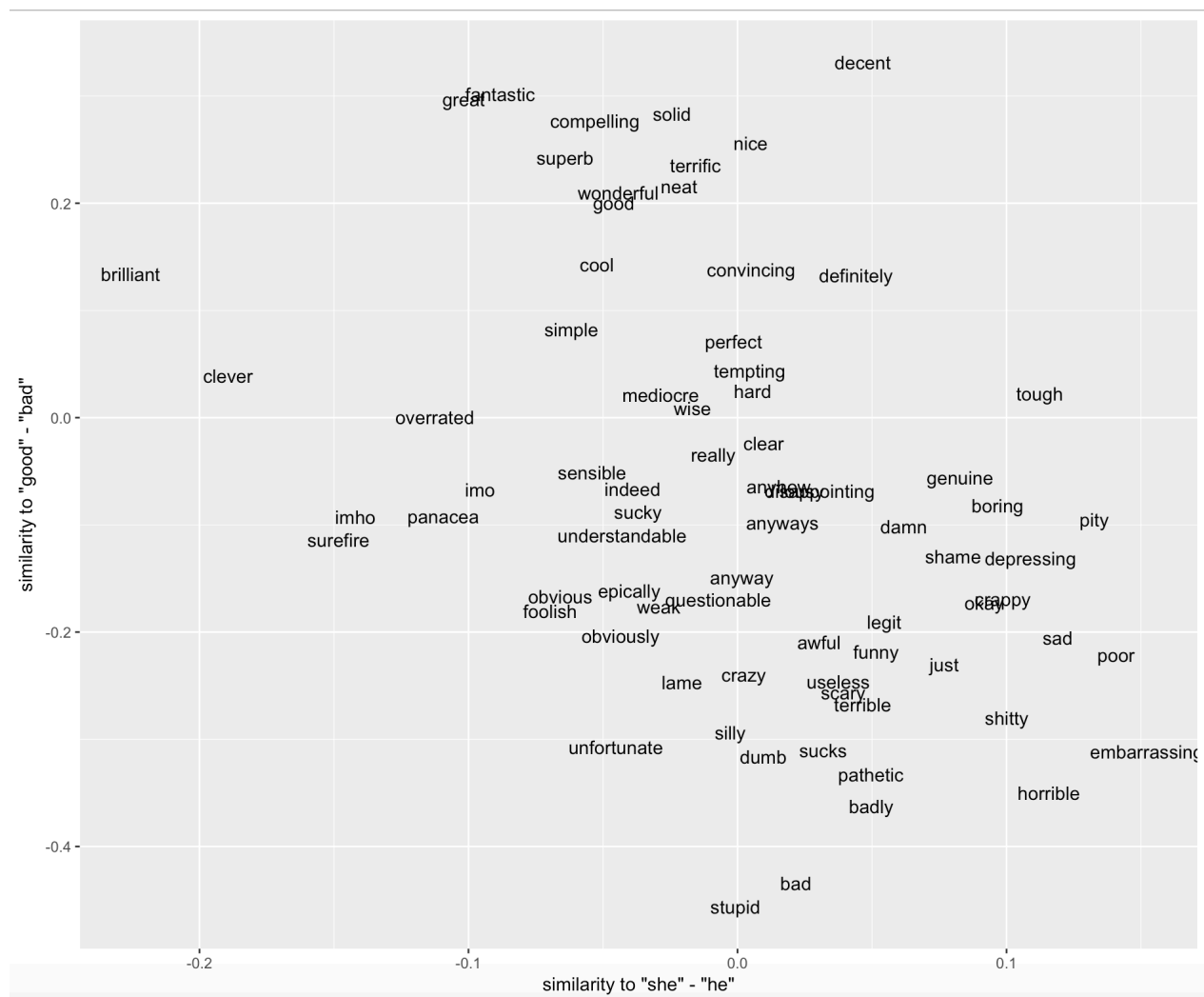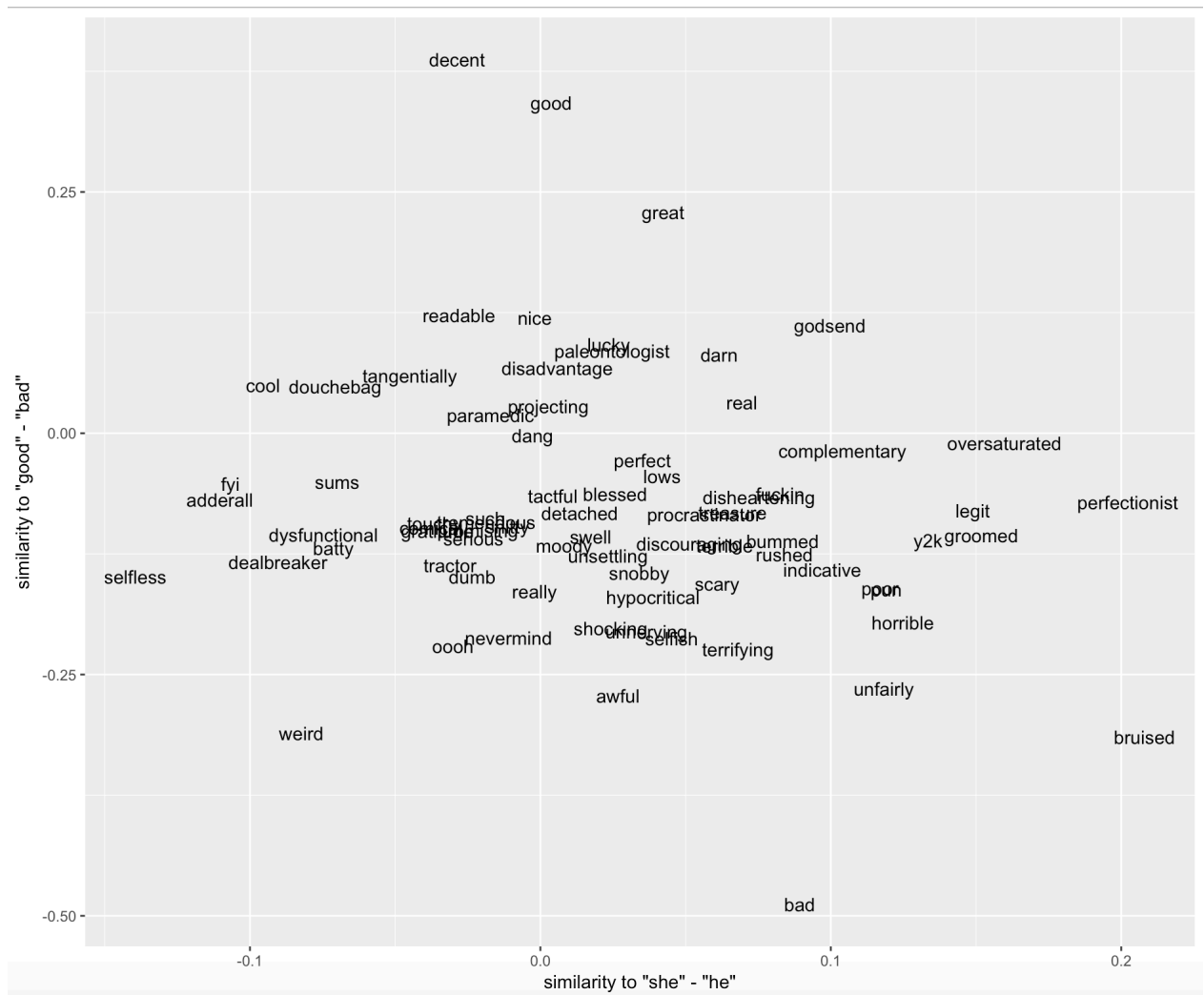
The same graphs in both subreddits do not show the same gender biased associations (see below). In both graphs, quality words are more evenly spread between the four quadrants and while some denote obvious biases ('smart' is slightly male in AskMen and 'perfectionism' is female in AskWomen), in general these biases are not coherent groups of words as in the Hacker News data.

The difference between the Hacker News and Reddit graphs shows that not all internet forums are equally biased against women. In particular, in technologically-oriented anonymous internet conversation, Hacker News shows quite a bit of bias in terms of its users' gendered perception of skillsets, while the subreddits show very little if any bias. However, this study also promisingly shows that even the Hacker News biases are not absolutely gendered. Of course, as with anything connected to language we must ask ourselves if people are really saying what they think. The research on implicit bias suggests that even those of us who actively disavow stereotyping and bias still show bias when asked to instantaneously respond to questions connecting gender and occupation. It could very well be that AskMen and AskWomen members are nearly as biased against women as Hacker News commenters, but that they are simply moderating their speech because of political correctness. Nonetheless, even this speech moderation caused the neural network not to learn the same gender biases for occupations that it learnt on the Hacker News data. And moreover, just as Neural Networks learn from examples, we as people also learn and internalize biases because of the examples we see. This paper thus serves as both a warning against training word embeddings for NLP tasks on biased data, and as a message that less biased data exists on the internet and is detectable.

Additionally, I think that in order to compare biases in different data sets, if cross dataset vector algebra can be reworked to generate meaningful connections, it will probably be a more valuable tool than just comparing the same vector algebra in both datasets as I did in this paper. Cross-dataset vector algebra could also be a strong tool for using word embeddings in literary

research because it would allow comparison of the same word or concept across groups, creating a quantitative measurement of difference for symbolism.

similarity to "good" - "bad"

0.25

0.00

-0.25

-0.50

decent

good

great

readable   nice   godsend
                paleontologist   darn
                lucky
cool douchebag tangentially disadvantage
                projecting   real
        paramedic
            dang
                        perfect   complementary   oversaturated
                        lows
fyi   sums   disheartening fucking
adderall   tactful blessed   procrastinate stature   legit   perfectionist
            dysfunctional   tougher such   detached   y2k groomed
                batty   gratifying serious swell   discouragingle bummed
        dealbreaker   moody unsettling   rushed
selfless   tractor   snobby   scary   indicative
                dumb   really   hypocritical   poon
                                                    horrible
        oooh nevermind   shocking crying semen terrifying

                            awful   unfairly

weird   bruised

bad

similarity to "she" - "he"

-0.1   0.0   0.1   0.2

AskMen