

## A Appendix

We include here extended results for each of the stages of work that we covered in the main paper: findings for inferring word-level gender associations (Tables 5 and 6), findings from clustering and cluster labeling (Tables 7 and 8), and results from our two evaluation tasks (Tables 4 and 9). Tables 5–9 are on the following three pages.

### A.1 Word-Level Results

Tables 5 and 6 provide the top 25 most significant female-associated and male-associated nouns, verbs, and adjectives in both domains. Gender association is determined using the method described in Section 4.1.

### A.2 Clustering and Cluster Labeling Results

Tables 7 and 8 provide the clustering and cluster labeling results for the top- $n$  clusters in each domain, ranked by ascending normalized sum of squared errors. During clustering, we set  $k$  to  $\frac{N}{50}$ , where  $N$  was the number of embeddings being clustered. In the celebrity domain, this resulted in 45 clusters; in Table 7, we show the top 12. In the professor domain, this resulted in 16 clusters; in Table 8, we show the top 6.

### A.3 Human Evaluation Results

Table 4 provides more detailed results for the Word Intrusion task. As in the main paper, we provide the results in each domain for all clusters and for just the top 8, determined by ascending normalized sum of squares.

Domain	Level	Precision	Fleiss' $\kappa$
Celeb	Top-8	.725	.530
Celeb	Overall	.527	.314
Prof.	Top-8	.600	.216
Prof.	Overall	.488	.212

Table 4: Results for Word Intrusion task. All results significantly outperform the random baseline of .20 ( $p \leq 0.0001$ ).

Table 9 displays more detailed results for the cluster labeling task. As discussed in the main paper, we would like to see that the rate at which in-cluster words fall under the cluster label is much higher than the rate at which out-of-cluster words fall under. We tried the top 4 predicted labels, compared against the centroid as a baseline label.

For each cluster, we tested labels against 10 in-cluster words and 3 out-of-cluster words; thus, we tested 65 questions (13 words x 5 labels) per cluster. Due to the high number of questions per cluster, we only tested 15 clusters – the top 10 labeled in the celebrity domain and the top 5 labeled in the professor domain. We report results over all tested clusters, but broken down by label type (e.g. centroid or 2<sup>nd</sup> predicted label).

Table 5: Top 25 most gender-associated nouns, verbs, and adjectives in the celebrity domain. Words are listed in order of decreasing significance, but all words fall under  $p \leq 0.05$ , with Bonferroni correction.

	<b>Female-Associated</b>	<b>Male-Associated</b>
Nouns	girl, eye, cover, star, sister, shoulder, husband, lady, baby, issue, wedding, actress, reality, daughter, carpet, gown, fashion, mom, pregnancy, photo, top, hair, look, outfit, mother	movie, actor, president, wife, dad, death, film, host, news, statement, father, girlfriend, man, assault, allegation, attorney, claim, investigation, misconduct, behavior, lawsuit, action, guy, director, harassment
Verbs	caption, wear, look, keep, love, date, match, rock, use, accessorize, style, pair, feel, share, show, shop, fit, plunge, model, apply, reveal, flaunt, open, tone, color	say, deny, accord, claim, accuse, allege, would, apologize, publish, fire, hear, play, come, continue, ask, involve, pass, replace, investigate, charge, win, rap, pay, state, sue
Adj.	red, pink, pregnant, sexy, sheer, white, black, beautiful, stylish, nude, stunning, chic, natural, blonde, new, blue, sweet, glam, loose, hot, oversized, casual, gorgeous, toned, little	sexual, consensual, inappropriate, furious, guilty, alleged, oral, presidential, late, republican, deadpool, many, financial, bad, political, public, false, russian, comic, non, dead, detailed, numerous, untrue, criminal

Table 6: Top 25 most gender-associated nouns, verbs, and adjectives in the professor domain. Words are listed in order of decreasing significance, but all words fall under  $p \leq 0.05$ , with Bonferroni correction, aside from the last three terms listed for Female-Associated Verbs and the last four terms listed for Male-Associated Verbs.

	<b>Female-Associated</b>	<b>Male-Associated</b>
Nouns	class, work, teacher, person, woman, assignment, email, credit, week, step, direction, attitude, instruction, lady, instructor, date, response, discussion, sweetheart, communication, husband, point, group, mail, time	material, concept, professor, exam, sense, topic, joke, math, stuff, programming, world, department, lecture, prof, note, humor, guy, story, lecturer, curve, tangent, man, dude, genius, industry
Verbs	take, would, respond, email, recommend, help, love, submit, miss, grade, treat, follow, complete, accept, hat, need, send, reply, turn, correct, communicate, check, feel, wait, finish	learn, curve, teach, lay, talk, crack, know, write, lecture, challenge, want, relate, sleep, solve, ramble, push, start, entertain, joke, put, bore, mumble, base, cover, explain
Adj.	easy, helpful, nice, rude, online, due, wonderful, extra, sweet, kind, busy, caring, hot, late, strict, timely, annoying, horrible, mean, quick, disorganized, pleasant, responsive, beautiful, lovely	good, knowledgeable, great, smart, interesting, intelligent, difficult, real, funny, hilarious, boring, cool, dry, entertaining, favorite, challenging, brilliant, arrogant, interested, hard, passionate, old, tough, practical, excellent

Table 7: Top 12 clusters out of 45 overall in the celebrity domain. Predicted labels are included if applicable – we were only able to predict labels for clusters that contained nouns, since our clustering labeling algorithm relied on the noun taxonomy in WordNet. In the **Sample Words in Cluster** column, italics indicate female-associated terms, and non-italics indicate male-associated. **F:M** refers to the ratio of female-associated to male-associated words in the cluster.

<b>Sample Words in Cluster</b>	<b>F:M</b>	<b>Centroid</b>	<b>Top 3 Pred. Cluster Labels</b>
<i>month</i> , <i>day</i> , year, decade, hour	2:3	month	time period, fundamental quantity, measure
<i>engagement</i> , <i>marriage</i> , <i>divorce</i> , <i>relationship</i> , <i>split</i> , <i>breakup</i>	6:0	breakup	state, union, separation
<i>woman</i> , man, guy, people, someone, anyone, one, person	1:7	man	person, man, adult
<i>photo</i> , <i>post</i> , <i>pic</i> , <i>snap</i> , <i>selfie</i> , <i>image</i> , <i>snapshot</i> , <i>kiss</i> , <i>photograph</i> , tribute	10:2	snapshot	photograph, representation, picture
<i>reveal</i> , <i>gush</i> , <i>tell</i> , <i>explain</i> , <i>dish</i> , <i>admit</i> , say, recall, reply	7:3	explain	N/A
<i>trophy</i> , win, category, nominate, nomination, finalist	1:5	finalist	collection, condition, award
<i>exclusive</i> , <i>peek</i> , <i>sneak</i> , <i>glimpse</i> , <i>preview</i>	5:0	preview	look, sensing, screening
<i>difference</i> , allegation, claim, accusation, truth, lie	1:5	accusation	claim, assertion, statement
<i>accessorize</i> , <i>sneaker</i> , <i>toe</i> , <i>jewel</i> , <i>floral</i> , <i>embroider</i> , <i>pink</i> , <i>hat</i> , <i>veil</i> , <i>turtleneck</i> , <i>sequin</i> , <i>tiara</i> , <i>bodysuit</i>	123:1	halter	artifact, garment, covering
<i>girl</i> , <i>sister</i> , <i>baby</i> , <i>daughter</i> , <i>mom</i> , <i>pregnancy</i> , dad, father, brother, son	14:6	dad	person, parent, mother
<i>husband</i> , <i>boyfriend</i> , <i>beau</i> , <i>hubby</i> , <i>fiancee</i> , wife, girlfriend, finacee	15:7	fiance	lover, person, relative
<i>shoulder</i> , <i>gown</i> , <i>top</i> , <i>dress</i> , <i>pant</i> , <i>skirt</i> , <i>diamond</i> , <i>neckline</i> , <i>waist</i> , <i>bra</i>	25:0	blouse	covering, cloth covering, clothing

Table 8: Top 6 clusters out of 16 overall in the professor domain. Same details as Table 7 apply.

<b>Sample Words in Cluster</b>	<b>F:M</b>	<b>Centroid</b>	<b>Top 3 Pred. Cluster Labels</b>
<i>teacher</i> , <i>woman</i> , <i>lady</i> , professor, prof, guy, lecturer, man, dude	5:7	prof	man, adult, woman
<i>everything</i> , material, concept, topic, stuff, tangent, subject, content	1:8	content	content, idea, cognition
<i>respond</i> , <i>email</i> , <i>response</i> , <i>responsive</i> , <i>stuff</i> , <i>reply</i> , <i>contact</i> , <i>answer</i>	13:0	respond	statement, message, communication
<i>familiar</i> , <i>literate</i> , <i>savvy</i> , genius, nerd, background, scientist, geek	5:10	geek	person, expert, anomaly
<i>communicate</i> , learn, teach, know, solve, explain, introduce, convey	1:18	simplify	N/A
<i>compassionate</i> , <i>lovely</i> , <i>amazing</i> , <i>sweet</i> , <i>understanding</i> , <i>hot</i> , <i>nice</i> , passionate, funniest, smart, cool, inspirational, intelligent, likable	21:44	outgoing	N/A

Label Type	In-Cluster Rate	Out-of-Cluster Rate	Difference	Fleiss' $\kappa$
Centroid	.597	.178	.419	.322
1 <sup>st</sup> pred. label	.621	.156	.465	.258
2 <sup>nd</sup> pred. label	.540	.200	.340	.256
3 <sup>rd</sup> pred. label	<b>.653</b>	<b>.044</b>	<b>.609</b>	.353
4 <sup>th</sup> pred. label	.452	.111	.341	.261

Table 9: Results for cluster labeling task. The 3<sup>rd</sup> predicted label has a significantly lower out-of-cluster rate than the centroid and all the other predicted labels ( $p \leq 0.02$ ). The same label also slightly outperforms the centroid on the in-cluster rate, thus producing a much larger gap between rates than the centroid.