

Supplementary information for **“Historical Dynamics and Cultural Associations of Relational Norms in American and Chinese Political Discourse”**

Ava Q. Ma de Sousa^{*1}, Srishti Thapar¹, Shangcheng W. Zhao¹, Amber X. Chen¹, Tianhao Qin¹, Margaret S. Clark², Brian D. Earp^{3,4,5}, Hongbo Yu^{*1}

¹Department of Psychological and Brain Sciences, University of California Santa Barbara, Santa Barbara CA 93106, U.S.A.

²Department of Psychology, Yale University, New Haven, CT 06520-8047, U.S.A.

³Centre for Biomedical Ethics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

⁴Department of Psychology, National University of Singapore, Singapore

⁵Department of Philosophy, National University of Singapore, Singapore

*Corresponding authors

Table S1.
Reliability checks.

Function	U.S. ICC(C,1), single wording	U.S. ICC(C,k), average of 4	China ICC(C,1), single wording	China ICC(C,k), average of 4
Care	0.44	0.76	0.45	0.77
Hierarchy	0.30	0.63	0.56	0.83
Transaction	0.53	0.82	0.47	0.78
Mating	0.79	0.94	0.63	0.87

Table S2.
Validity checks.

Corpus	Function	Spearman ρ [95% CI]	Pearson r	AUC [95% CI]
U.S.	Care	.796 [.729, .842]	.967	.996 [.989, 1.000]
U.S.	Hierarchy	.784 [.710, .837]	.940	.978 [.946, .999]
U.S.	Transaction	.816 [.761, .849]	.988	1.000 [1.000, 1.000]
U.S.	Mating	.816 [.761, .850]	.946	.992 [.981, .999]
China	Care	.715 [.635, .780]	.798	.949 [.915, .974]
China	Hierarchy	.814 [.762, .846]	.955	.989 [.968, 1.000]
China	Transaction	.817 [.762, .851]	.942	.993 [.980, 1.000]
China	Mating	.423 [.320, .506]	.490	.862 [.802, .913]

Table S3.
Bayesian Change Point Detection in Both Corpora.

Function	U.S. Change points with posterior > .70	China Change points with posterior > .70
Care	1930 (.996)	1963 (.81)
Hierarchy	1882 (.990) 1920 (.98) 1883 (.96) 1897 (.95)	1969 (1.00) 1966 (.99) 1965 (.97) 1978 (.96) 1963 (.74)
Transaction	1931 (.994) 2008 (.86) 2001 (.71)	1946 (.998) 1948 (.99) 1961 (.92) 1957 (.90)
Mating	1980 (.920) 1944 (.91) 1897 (.75)	NA (highest posterior = .54)

Figure S1.

Historical changes in prevalence of cooperative functions across relationship specific discourse (private relationships) in the American and Chinese corpora.

Colored lines represent annual scores for each function, with care in blue, hierarchy in purple, transaction in yellow, and mating in pink.

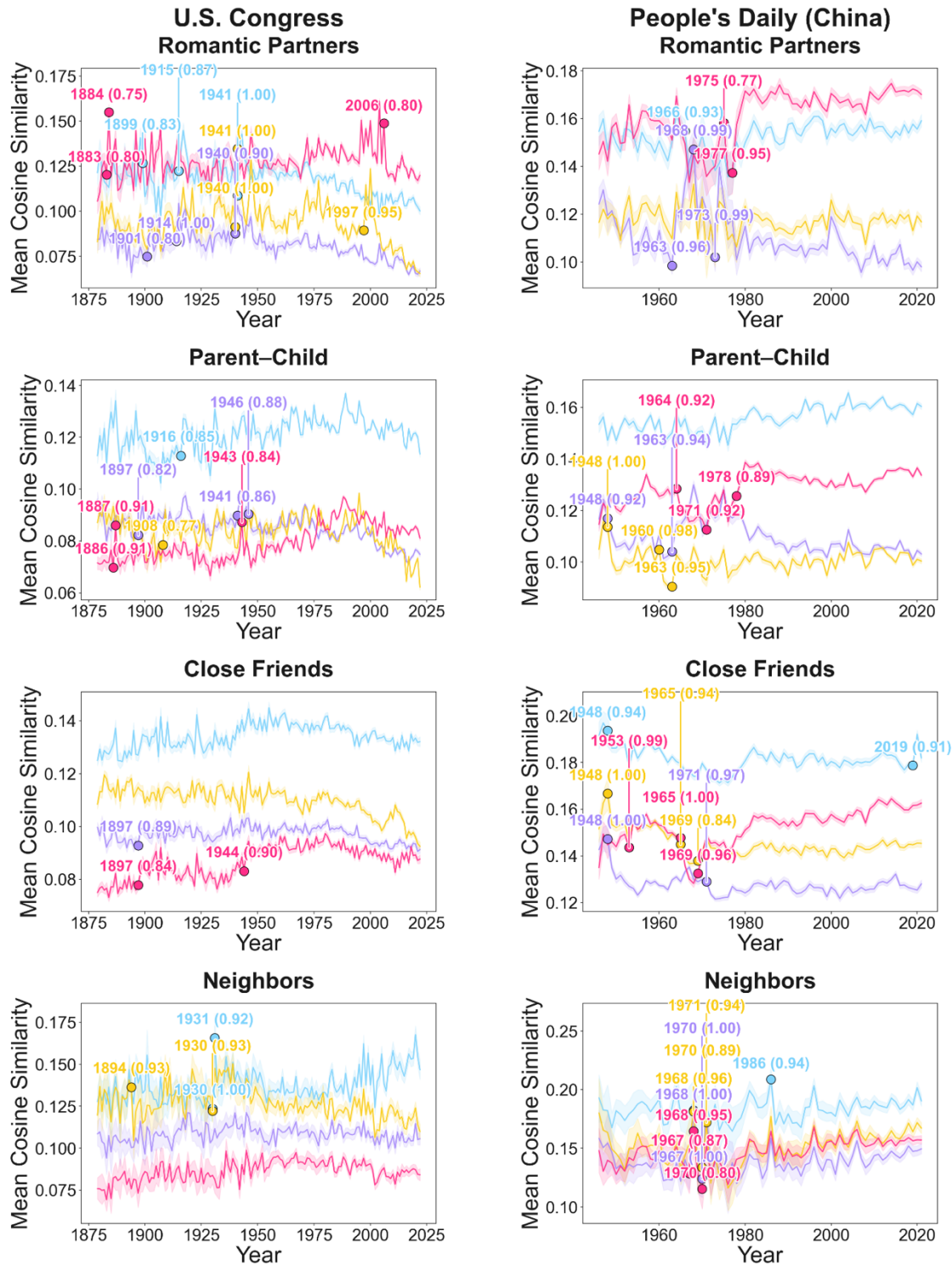


Figure S2.

Historical changes in prevalence of cooperative functions across relationship specific discourse (public relationships) in the American and Chinese corpora.

Colored lines represent annual scores for each function, with care in blue, hierarchy in purple, transaction in yellow, and mating in pink.

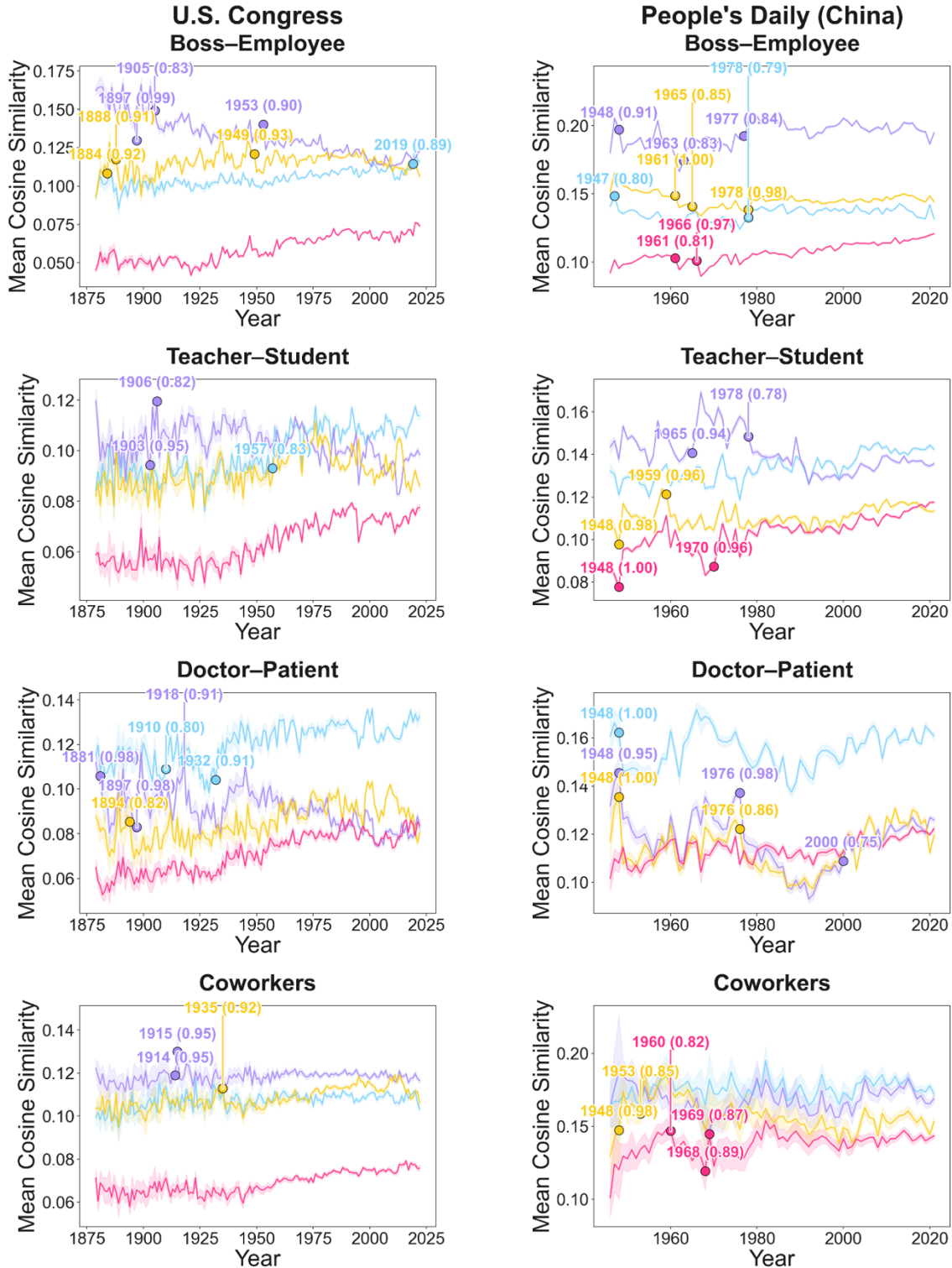


Table S4.

Significant Granger and ARIMAX associations between cultural values and relationship-specific cooperative function discourse.

Corpus	Relationship	Cooperative function	Cultural indicator	Predictive direction	Granger q	ARIMAX p	GDP-adjusted p	Peak lag (years)	β	p (β)	GDP-adjusted β	GDP-adjusted p (β)
U.S.	Parent-child	Hierarchy	Tightness	Function → cultural value	.027	.251†	.255†	3	+.018	.788	+.078	.142
U.S.	Close friends	Transaction	Collectivism	Cultural value → function	.017	< .001	< .001	7	-.016	.712	-.013	.714
U.S.	Romantic partners	Care	Collectivism†	Function → cultural value	< .001	.001†	.001†	7	-.009	.002	-.009	.002
U.S.	Romantic partners	Hierarchy	Tightness	Function → cultural value	.041	.011†	.010†	6	-.064	.229	-.064	.230
U.S.	Romantic partners	Transaction	Tightness	Function → cultural value	.015	.061†	.059†	11	-.120	.021	-.122	.021
U.S.	Romantic partners	Mating	Individualism	Function → cultural value	.044	.007†	.007†	2	+.027	.351	+.030	.306
U.S.	Romantic partners	Mating	Tightness	Function → cultural value	.015	.110†	.417†	2	+.091	.072	+.060	.273
U.S.	Romantic partners	Mating	Looseness	Function → cultural value	.005	.025†	.035†	1	+.033	.191	+.035	.130
U.S.	Boss-employee	Care	Individualism	Cultural value → function	< .001	< .001	< .001	9	+.048	.024	+.048	.028
U.S.	Boss-employee	Care	Looseness	Cultural value → function	.009	.002	.002	10	+.030	.032	+.027	.209
U.S.	Boss-employee	Hierarchy	Collectivism	Cultural value → function	.002	< .001	< .001	3	+.052	.011	+.041	.057
U.S.	Boss-employee	Mating	Tightness	Cultural value → function	.034	.004	.001	4	-.011	.773	-.029	.397
U.S.	Teacher-student	Hierarchy	Individualism	Cultural value → function	.025	.007	.007	5	-.019	.441	-.021	.406

U.S.	Teacher–student	Hierarchy	Collectivism‡	Cultural value → function	.041	.035	.005	2	+.264	.294	–.370	.270
U.S.	Teacher–student	Hierarchy	Looseness	Cultural value → function	.015	.002	.002	8	–.016	.416	–.022	.318
U.S.	Doctor–patient	Hierarchy	Collectivism‡	Function → cultural value	< .001	.003†	.002†	3	+.007	.023	+.007	.019
U.S.	Doctor–patient	Transaction	Tightness	Cultural value → function	.017	.062	.053	8	+.015	.652	+.013	.698
U.S.	Coworkers	Transaction	Collectivism‡	Cultural value → function	.033	< .001	< .001	4	–.706	< .001	–2.09	< .001
China	Parent–child	Hierarchy	Collectivism	Cultural value → function	.042	< .001	.005†	1	–.271	< .001	–.262	.005
China	Parent–child	Mating	Collectivism	Function → cultural value	.002	< .001	< .001	2	+.174	.126	+.176	.123
China	Close friends	Hierarchy	Individualism	Cultural value → function	.046	.085†	.075†	2	–.053	.540	–.069	.410
China	Close friends	Transaction	Collectivism	Function → cultural value	.018	.001	.002	1	+.283	.001	+.297	.002
China	Close friends	Mating	Collectivism	Function → cultural value	.005	.003	.002	1	+.275	.003	+.339	.002
China	Romantic partners	Care	Collectivism	Cultural value → function	.001	.063	.002	6	+.135	.058	+.163	.027
China	Romantic partners	Mating	Collectivism	Function → cultural value	.022	.012	.011	2	+.280	.005	+.287	.004
China	Teacher–student	Hierarchy	Collectivism	Cultural value → function	.001	< .001	.011†	6	–.235	.051	–.235	.014
China	Teacher–student	Mating	Collectivism	Cultural value → function	.015	< .001	< .001	1	+.055	.534	+.094	.277
China	Teacher–student	Mating	Tightness	Cultural value → function	.005	.001	.001	4	–.153	.098	–.139	.123

China	Doctor–patient	Care	Individualism	Function → cultural value	.034	.008	< .001	5	+.311	.002	+.363	< .001
China	Doctor–patient	Care	Collectivism	Function → cultural value	.034	.041	.005	3	–.178	.077	–.185	.068
China	Doctor–patient	Hierarchy	Collectivism	Function → cultural value	.030	.005	.005	3	–.327	< .001	–.323	< .001
China	Doctor–patient	Transaction	Collectivism	Cultural value → function	.028	< .001†	< .001†	3	–.267	< .001	–.224	.002
China	Coworkers	Transaction	Collectivism	Function → cultural value	.014	.001	.001	1	+.298	.002	+.321	.001
China	Coworkers	Transaction	Tightness	Function → cultural value	.004	.015	.084†	4	–.163	.015	–.122	.057
China	Coworkers	Mating	Collectivism	Cultural value → function	.019	.023	.006	6	–.216	.008	–.254	.002
China	Neighbors	Hierarchy	Tightness	Function → cultural value	.004	.032	.029	1	+.137	.009	+.130	.008

Note. ARIMAX p and GDP-adjusted p are joint Wald tests of all lags. † Ljung–Box $p < .05$ for that model's residuals. β is the standardized coefficient at the peak lag, and $p(\beta)$ tests the peak-lag single coefficient.

Pre-processing of U.S. Congressional Record

We used a preprocessed corpus of U.S. Congressional speeches from a previous study (Segun et al., 2025). The corpus comprised 8.4M non-procedural speeches spanning congressional sessions from 1879 to 2022. We conducted additional preprocessing to improve the quality of sentence-level representations used in subsequent embedding analyses.

First, we removed duplicate and near-duplicate speeches. Speeches were considered potential duplicates if they were delivered on the same day, in the same chamber, and by the same speaker. For text comparison, speeches were normalized by converting all characters to lowercase and removing punctuation and redundant whitespace. A pair of speeches was classified as duplicate if their normalized texts matched exactly. For non-identical texts, we considered pairs only if their lengths differed by no more than 5%. Within these candidate pairs, speeches were classified as duplicates if either of two text-similarity measures reached a threshold of .98: (1) word-level Jaccard similarity or (2) character-level Ratcliff–Obershelp sequence similarity, implemented using `SequenceMatcher` from Python's `difflib` library. Duplicate pairs were grouped temporarily, and one speech per group was retained, preferring the Congressional Record version, then the longer text, then the lower speech identifier.

Next, each speech was segmented into sentences. Sentence boundaries were placed after a period, question mark, or exclamation mark followed by whitespace and an uppercase letter, digit, or quotation mark; periods in common abbreviations (e.g., "Mr.", "U.S.", "H.R.", "S. Res.") and personal initials were masked beforehand so they were not treated as boundaries. For speeches that were originally drawn from the Gentzkow et al. (2018) source, which replaced commas and semicolons with periods, we first restored a comma wherever a period was followed by a lowercase word. Because periods preceding capitalized words (e.g., names and titles) could not be restored this way, we additionally merged short verbless segments (six words or fewer containing no auxiliary, modal, common procedural verb, or -ing/-ed form) into an adjacent sentence with a comma: segments at the start of a speech were merged into the following sentence and all others into the preceding sentence. This merge was not applied to Congressional Record (CREC) text, whose original punctuation was retained.

Table S5.*Congressional Record Dropped Duplicated Speeches by 20-Year Era.*

Era	Number of raw sentences	Removed sentences	Percentage of raw (%)
1879-1898	893626	684	0.08
1899-1918	1252939	440	0.04
1919-1938	1342551	459	0.03
1939-1958	1222239	1677	0.14
1959-1978	1464325	3598	0.25
1979-1998	1308779	2856	0.22
1999-2018	859466	32544	3.79
2019-2022	91844	984	1.07

Table S6.*Congressional Record Dropped Duplicated Speeches by 20-Year Era.*

Era	Number of sentences	Number of segments	Number of merged	Percent merged	Number of sentences
1879-1898	892942	4777573	488211	10.22	4289362
1899-1918	1252499	6085926	611380	10.05	5474546
1919-1938	1342092	6977157	666133	9.55	6311024
1939-1958	1220562	8804419	816665	9.28	7987754
1959-1978	1460727	15608436	1547786	9.92	14060650
1979-1998	1305923	20084618	2306159	11.48	17778459
1999-2018	826922	16459038	1735763	10.55	14723275
2019-2022	90860	1636323	0	0	1636323
Total	8392527	80433490	8172097	10.16	72261393

Note: Sentence merging was only applied to content that originated from Gentzkow et al. (2018), mostly before 2016

Table S7.
People's Daily Pre-processing.

Reason	# Removed rows	% of raw	Description
Bare date	1,332,501	2.56%	rows that are only a date (e.g., 1956-01-02, 1955年12月29日)
Column Label	1,327,940	2.55%	column-label rows: ‘专栏: ’
Repeated Header	1,327,868	2.55%	Headers were repeated twice
Page Marker	1,323,989	2.54%	Page-number rows ‘第N版 (...)’
Byline Fragmentation	1,169,966	2.25%	bare name / name-list rows: 2-4-hanzi chunks joined by 、 · or spaces, ≤20 chars, with no ending punctuation (‘刘衡, 本报记者 刘衡’)
No Chinese characters	935,498	1.80%	rows with zero Chinese characters (e.g. lone ", ..., English/number fragments)
Meta data	316,510	0.61%	rows starting [版名: X] (section-name field)
Empty strings	281,540	0.54%	rows empty after stripping a lone wrapping "
Garbage characters	199,096	0.38%	Rows with >30% of characters outside CJK/ASCII/CJK-punctuation (mojibake Z\x01\x0e:, figure marks ④, ★)
Length	149,085	0.29%	<5 chars and no ending punctuation (二, (一), (掌声), 南史说:)
Photo credit	129,351	0.25%	photo/graphics credits (‘...张三摄, 新华社记者 X 摄, 制图’)
Table row	30,853	0.06%	rows containing a tab \t, which marked table columns dumped as text
Provenance	10,045	0.02%	Source stamp rows: 本条数据为: 人民日报YYYY.MM.DD第N版
QR promo	5,105	0.01%	QR / app-download / follow-us promo lines (‘扫描二维码...看更多内容’)

Cross reference tabs	2,147	0.004%	cross-reference tags (e.g. '(上接第X版) , (下转...) , (相关报道见...))
Contact/submission lines	779	0.001%	contact/submission lines (电话: , 传真: , 来稿请寄, email)
Duplicates	2,512,452	4.82%	Drops ≥ 5 consecutive normalized-text matches at a fixed row offset, then boundary extension, drops the later copy; the earliest is kept

Table S8.

Terms used for relationship lexical filtering search.

Relationship	English (role A)	English (role B)	Chinese (role A)	Chinese (role B)
Parent-child	parent(s), mother(s), father(s), mom(s), dad(s)	child, children, son(s), daughter(s), kid(s)	父母, 母亲, 妈妈, 父亲, 爸爸	孩子, 子女, 儿子, 女儿
Siblings	brother(s), sibling(s)	sister(s), sibling(s)	兄弟, 手足, 兄弟姐妹	姐妹, 手足, 兄弟姐妹
Romantic partners	husband(s), wife/wives, spouse(s), boyfriend(s), girlfriend(s), couple(s), marriage(s), romantic	—	夫妻, 丈夫, 妻子, 太太, 情侣, 男朋友, 男友, 女朋友, 女友, 恋人, 配偶	—
Close friends	friend(s), friendship(s)	—	朋友, 友谊, 好友, 挚友	—
Neighbors	neighbor(s)	—	邻居, 邻里	—
Acquaintances	acquaintance(s)	—	熟人, 相识	—
Coworkers	co-worker(s)/coworker(s), colleague(s), workmate(s), fellow worker(s)/fellow employee(s)	—	同事, 同僚	—
Boss-employee	boss(es), employer(s), supervisor(s), manager(s)	employee(s), worker(s), staff, subordinate(s)	老板, 上司, 雇主, 经理, 领导, 首长	员工, 雇员, 下属, 工人
Teacher-student	teacher(s), professor(s), instructor(s), educator(s)	student(s), pupil(s), learner(s)	老师, 教师, 教授, 导师	学生, 弟子
Doctor-patient	doctor(s), physician(s), surgeon(s)	patient(s)	医生, 大夫, 医师	病人, 患者

Table S9.*Distribution of Sentences per Relationship.*

Relationship	U.S. total	U.S. mean/yr (SD)	U.S. min– max	China total	China mean/yr (SD)	China min– max
Parent–child	1,192,278	8,280 (9,265)	422–41,444	502,907	6,617 (3,911)	822–16,498
Romantic partners	236,566	1,643 (1,648)	132–8,589	80,033	1,053 (654)	88–2,329
Close friends	473,334	3,287 (1,774)	656–8,673	324,508	4,270 (1,627)	253–9,677
Neighbors	58,555	407 (295)	46–1,147	18,319	241 (223)	17–894
Boss–employee	1,222,003	8,486 (6,696)	182–23,089	1,803,903	23,736 (7,362)	3,860–48,012
Teacher–student	523,466	3,635 (3,946)	62–15,842	737,182	9,700 (4,452)	879–20,702
Doctor–patient	271,882	1,888 (2,010)	65–11,166	200,552	2,639 (2,042)	181–12,299
Coworkers	321,542	2,233 (2,098)	101–7,320	24,983	329 (318)	18–1,586

References

- Aroyehun, S. T., Simchon, A., Carrella, F., Lasser, J., Lewandowsky, S., & Garcia, D. (2025). Computational analysis of US congressional speeches reveals a shift from evidence to intuition. *Nature Human Behaviour*, 9(6), 1122–1133.
- Gentzkow, M., Shapiro, J. M., & Taddy, M. (2018). Congressional Record for the 43rd–114th Congresses: Parsed speeches and phrase counts [Data set]. Stanford Libraries.