Supplementary Information: Dream Content Discovery from Social Media Using AI

Anubhab Das[1], Sanja Šćepanović[2*], Luca Maria Aiello[3,4], Remington Mallett[5], Deirdre Barrett[6], Daniele Quercia[2]

[1*]Heritage Institute of Technology, Chowbaga Road, Anandapur, Kolkata, 700107, West Bengal, India.
[2]Nokia Bell Labs, Broers Building, Cambridge, CB3 0FA, United Kingdom.
[3]IT University of Copenhagen, Rued Langgaads Vej, 7, Copenhagen, 2300, Denmark.
[4]Pioneer Centre for AI, Øster Voldgade 3, Copenhagen, 1350, Denmark.
[5]Northwestern University, 633 Clark St, Evanston, 60208, Illinois, United States.
[6]Harvard University, 352 Harvard Street, Cambridge, 02138, Massachusetts, United States.

*Corresponding author(s). E-mail(s): sanja.scepanovic@nokia-bell-labs.com;
Contributing authors: anubhabdas.ad@gmail.com; lajello@gmail.com; mallett.remy@gmail.com; deirdre_barrett@hms.harvard.edu; daniele.quercia@nokia-bell-labs.com;

S1 Data statistics (Extended)

Fig. S1, S2, S3 & S4 provide additional insights into the dream reports. We were able to geolocate 4.04% (n=1784) of all dream reports (distributed amongst 3.22% (n=1423) users) at the US state level. Fig. S2 demonstrates representative coverage of our data in all 50 US states; using the aforementioned data subset.

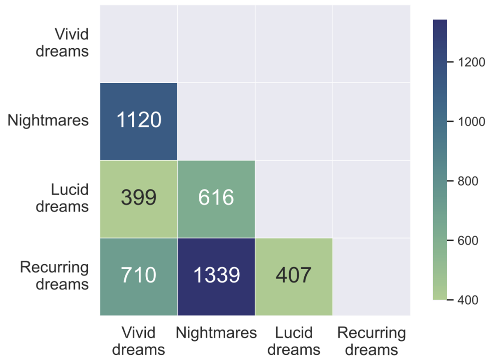

Fig. S1: **Heatmap that shows no. of dreams present across two dream types simultaneously**

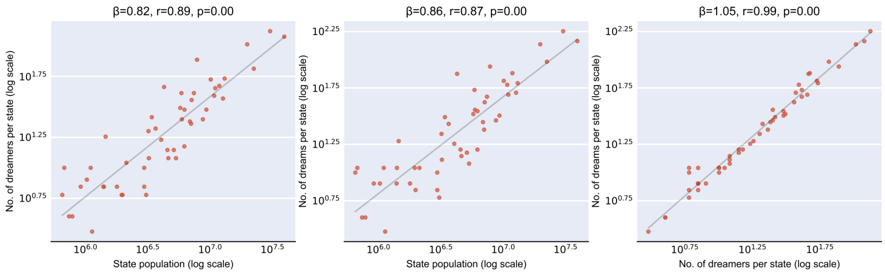

Fig. S2: **Correlation plots b/w 1. US state population, 2. No. of dreamers per state, and 3. No. of dreams per state**

S2 Methods & Implementation Details (Extended)

S2.1 Details of Topic Modelling

BERTopic embeds all the documents into vectors, projects them onto a lower dimensional space using UMAP, clusters the reduced embeddings using HDBSCAN (Hierarchical Density-based Spatial Clustering of Applications with Noise), and finally generates the topic names using a class-based variation of TF-IDF, namely c-TF-IDF.

In the application of BERTopic, we set the hyperparameters of the *all-mpnet-base-v2* sentence transformer model such that the min. no. of dream sentences required to form a topic is 100 (min_topic_size = 100) and it automatically merges similar topics together using HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) (nr_topic = 'auto').

After fitting the model, we used its functionality to update the topic representation i.e., modify the topic words to remove English stop words and take into account both unigrams and bigrams as topic words/phrases. This is particularly important for 2-word phrases like *sleep paralysis*, *recurring dreams* and, *serial killers* which we found, quite frequently mentioned in the extracted dataset of topics.

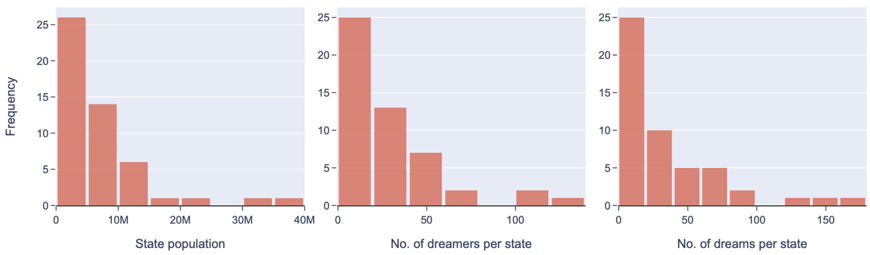

Fig. S3: **Distribution of state population & no. of dreams, dreamers per state**

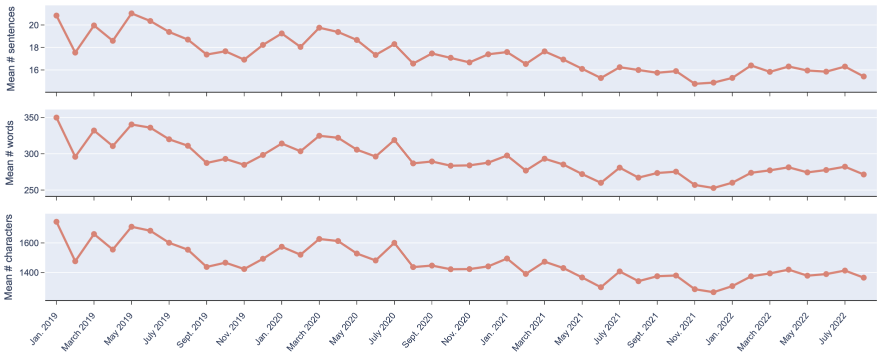

Fig. S4: **Evolution of dream length over time**

The probabilities of all topic words for each topic sum to 1; however since we removed "dream" and "dreams" topic words, we had to re-normalize the topic word probabilities to ensure they summed to 1.

Fig. S5 captures the distribution of no. of dreams reports, dreamers & the no. of dream sentences linked to a topic. There were 1982 (4.48%) dreams distributed amongst 1277 dreamers which weren't associated with any topic at all. While there were 4978 (11.26%) dreams amongst 3416 dreamers which weren't associated with any dreaming topics or themes.

Fig. S6 shows the plots for selecting the no. of clusters, post application of K-Means to topic embeddings.

S2.2 Details of building dream topics taxonomy (co-occurrence network)

Fig. S7 shows the plot used to determine the backboning threshold for visualizing the theme co-occurrence network.

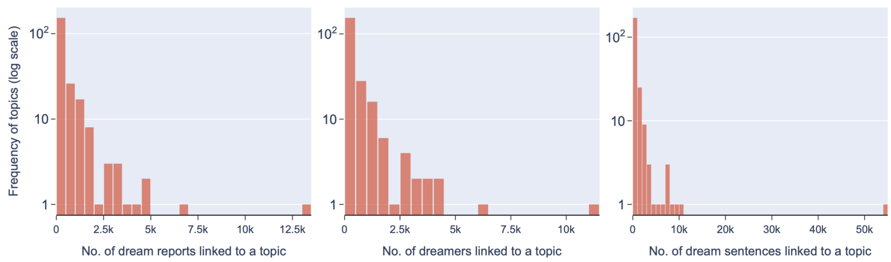

Fig. S5: **Distribution of no. of dreams, dreamers & dream sentences linked to a topic**

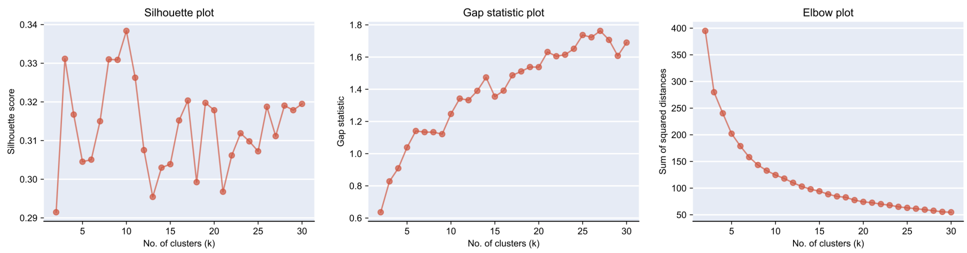

Fig. S6: **Selection of no. of clusters (k) for K-Means**

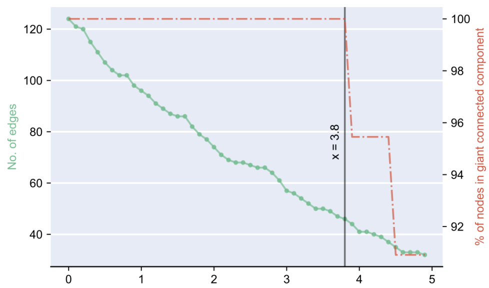

Fig. S7: **Backboning graph for theme network with backboning threshold = 3.8**

S2.3 Details of finding topics and themes through time

We did not include dreams from September 2022 (last month in our dataset) in temporal analyses as we only collected data up until 7th September 2022. Manual inspection revealed that, during 2018, the temporal curves for the raw z-scores were quite noisy to accurately infer anything, hence we further restricted our analyses from January 2019 to August 2022. For July and August 2022, since we did not have the data from the subsequent months for smoothing, we used $min\_periods = 0$ for the *rolling* function in Pandas which helped to compute the average by excluding the subsequent months' data. For calculating the smoothed z-scores of January and February 2019, we leveraged the values from November and December 2018, prior to discarding them, as discussed above.

S3 Results (Extended)

Table S1 provides an exhaustive list of dream topics and themes they belong to; as discovered by our proposed methodology.

Table S1: Full list of dream topics binned into various themes

Theme no.	Theme	Topics
0	People and relationships	0 lady, face, looked, head; 2 dream, girl-dream, dreamt, ve; 8 ex, years, dating, talk; 11 dreams-mean, staring, talk, people-dream; 16 dream-dad, dad-dream, dream-mom, mom-dream; 19 mum, aunt, mom-came, called-mom; 80 ex-boyfriend, dreams, having-dreams, relationship; 92 hugged, hug, kissed, arms; 106 names, remember-faces, characters, man-woman; 107 ll, nickname, named, mr; 115 crush, falling-love, fell-love, fall-love; 117 sex-dream, sexual, wet-dream, dream-sex; 169 versions, different-person, identity, perfect; 186 relationship, brother, mother-father, wife; 213 cheating, having-dreams, dream-boyfriend, relationship; 243 brother-dream, younger-brother, dream-younger, talking-little; 258 clowns, little-girl, creepy, costume
1	Indoor locations	5 mall, restaurant, eating, ice-cream; 7 bus, driving, cars, train-station; 13 doors, house, rooms, mansion; 36 toilet, shower, restroom, bathrooms; 37 house-dream, dream-house, apartment, childhood-home; 42 elevator, stairs, floors, escalator; 50 hospital, surgery, nurse, doctors; 60 flying, airport, helicopter, planes; 116 universes, alternate-reality, different-dimension, travel; 147 911, dial, emergency, ambulance; 170 portals, dimension, tunnel, hell; 220 roller-coaster, rollercoaster, rides, wheel; 250 house, amp-x200b, window, walking; 273 maze, hyper, tunnels, runner; 275 tier, different-levels, sam, happens-time; 277 elevators, dreams, reoccurring-dreams, recurrent
2	Violence and death	14 death, died-dream, going-die, death-dream; 43 pistol, shooting, shot-head, shotgun; 44 police, officer, officers, kidnapped; 46 nuke, war, fires, missile; 49 soldiers, hitler, russia, world-war; 76 serial-killer, dream-killing, killed-dream, killers; 83 tw, warning, nsfw, sexual-assault; 101 passed-away, cancer, away-years, father-passed; 156 rape, sexually-assaulted, abuse, nightmares; 158 jail, dream, court, police-car; 205 fight, vs, duel, eachother; 209 violent, disturbing, dreams-past, weird-dreams; 238 killed, putting-hands, fighting, beat; 268 violence, person-real, want-hurt, hate; 279 intruder, breaking, recurring, recurring-dream
3	Mental reflection and interactions	26 know-make, sure-don, know-think, know-sure; 40 yeah, like-wtf, hell, like-fuck; 52 asked-doing, tell, help, begged; 53 know, okay, yeah, know-going; 61 don-want, needed, doing-don, remember-doing; 120 said-yes, respond, didn-answer, yes-did; 122 know-knew, knew-didn, knew-don, known; 127 matter, case, know-isn, mattered; 139 lied, did-don, wish, don-just; 151 plan, care, notices, did-just; 161 confused-just, im, kinda-confused, just-really; 175 couldn-just, couldn-time, couldn-didn, know-able; 195 make-sense, makes, logic, sense-like; 217 miss, say-love, loved, dearly; 218 remember-saying, said-words, understand, mumbling; 233 oh, say, reply, asked-said; 235 doesn-exist, didn-exist, just-know, indication; 253 realised, exploring, felt-confused, bathroom-looked
4	Feelings	20 felt-real, dream-felt, real-dream, real-like; 39 woke-crying, started-crying, woke-tears, crying-dream; 64 feel-pain, painful, pain-dream, felt-pain; 75 relieved, felt-peace, happiness, felt-happy; 81 expression, tears, shock, nervous; 119 wasn-scary, scary-just, frightening, scariest; 129 ve-felt, felt-feeling, feeling-felt, haven-felt; 132 feel-right, felt-wrong, wasn-right, feel; 134 shocked, shock, surprise, disbelief; 136 crying, gasped, terrified, fear; 160 felt-sad, kind-sad, feel-sad, distraught; 172 upset, angry, livid, got-pissed; 194 guilt, felt-guilty, feel-guilty, remorse; 201 panic, panicking, panicked, started-panic; 242 really-creepy, weird-creepy, creepiest, sounds; 266 calm, calmed, tried-calm, feel-calm
5	Sights and vision	4 lights, sun, pitch-black, wearing; 87 reflection, looked-mirror, looking-mirror, mirrors; 140 blinded, vision-blurry, blur, recite; 174 disappeared, went-look, window, lost; 211 faded-away, existence, outline, suddenly-felt; 212 stared, head-looked, looked-eyes, looking-direction; 215 blurry, quality, picture, coated; 222 sight, seen, walkway, bodies; 226 sight, majestic, really-pretty, stunning; 259 mirrors, looking-mirror, bathroom-mirror, dream-looked; 264 decided-look, kept-looking, look, looking-looked
6	Animals	10 kitten, lion, birds, owl; 48 spider, maggots, batman, thanos; 88 snakes, alligator, turtle, bite; 176 horse, goats, brown, granny; 191 bear, polar, fence, started-chasing; 203 wolf, pack, grey, attack; 241 bees, stung, swarm, buzzing; 263 rat, traps, bites, swarming; 265 monkeys, enclosure, jungle, goo; 267 dinosaur, park, spawn, edges; 276 bunny, glitched, grey, hopping; 282 toad, poison, green, psychedelic; 284 deer, saw, tried-bite, road
7	Other topics (size, smell, apocalypse, etc.)	25 dream-ends, story, endings, recurring-theme; 86 huge, inches, like-size, big-small; 99 pov, 3rd-person, person-perspective, person-view; 111 room-starts, walking, family, house-starts; 112 really-cool, pretty, neat, thought-cool; 118 apocalyptic, dream-end, world-dreams, earth; 142 look-like, hair, skinny, face; 153 thousands, total, maybe-20, voodoo; 166 apocalypse, world-end, cosmic, event; 173 gay, trans, gender, feminine; 181 started-normal, normal-like, normal-looked, completely-normal; 188 smell, smelled-like, rot, disgusting; 199 describing, hard, try-best, properly; 206 coincidence, connection, thought, uncanny; 207 scene-changed, things-changed, switched, shifting; 214 chaos, disaster, mess, society; 216 dream-like, like-ve, haven-dream, like-long; 227 horrible, good-bad, bad-just, know-good; 239 knew, travelling, end-finally, didn-finish; 261 know-sounds, sounds-weird, weird-know, crazy; 262 scientists, experiments, laboratory, discover
8	Outdoor locations	12 beach, ocean, swim, river; 29 path, garden, hills, plants; 105 cave, tunnels, underground, tower; 114 forest, field, grass, dream-starts; 202 wooden, driving, bridges, unfinished; 249 islands, beings, coast, plans

Continued on next page

Table S1: Full list of dream topics binned into various themes

Theme no.	Theme	Topics
9	Personal objects	33 dressed, mask, naked, clothing; 56 phones, ringing, check-phone, battery; 74 pages, pen, ink, letters; 143 boxes, necklace, rings, cardboard-box; 145 shoes, pair, barefoot, sock; 146 crystals, statues, stone, gems; 163 backpack, bags, suitcase, packing; 164 knives, axe, pocket-knife, kitchen; 177 projector, remote, screen, television; 180 dolls, porcelain, voodoo, haunted; 193 counting, lottery, significance, noticed; 224 blankets, bed, pillow, felt; 246 paintings, art, wall, canvas; 255 eggs, ipad, scrambled, cracking; 270 keys, pocket, puzzle, box; 274 tattoos, artist, sleeve, piercing; 281 coins, drawer, wallet, rare
10	School	15 classroom, teachers, principal, student; 27 school-dream, dream-school, college, dream-high; 269 tests, failed, exam, professor
11	Movement and action	67 left, wanted-leave, time-leave, wanted-home; 85 continued-walking, continue, street, walk-home; 95 run, started-running, run-like, sprinting; 98 falling, hit-ground, fall-ground, let-fall; 124 lungs, couldn-breathe, like-couldn, heart; 128 remove, string, try-pull, tugged; 135 control, couldn-control, control-body, hopeless; 141 hid, closet, stairs, underneath; 144 escape, escaped, way-escape, managed-escape; 237 lose, round, hockey, scissors; 272 queue, line-people, waiting-line, standing-line; 280 dream-running, running-dream, run-fast, dream-run
12	Supernatural entities	28 creature, bite, mouth, face; 96 shadows, shadowy-figure, dark-figure, entity; 97 zombies, zombie-apocalypse, outbreak, like-zombie; 125 aliens, invasion, grey, race; 178 dragon, chinese, bearded, red; 184 alien, invasion, abducted, jar; 189 robots, ai, giant, colony; 192 zombie-apocalypse, zombies, outbreak, walking-dead; 197 ghosts, haunted, like-ghost, saw; 278 vampires, gang, hunter, glowing-red; 283 clones, disabled, machine, scientist
13	Sounds and lack thereof	41 singing, songs, lyrics, stage; 77 voices, heard-voice, hear-voice, voice-head; 89 footsteps, ghost, noises, ringing; 100 paralyzed, able, speak, couldn-speak; 121 silent, sound, went-silent, eerily; 182 speak, couldn-speak, mouth, whisper; 223 laughter, couldn-hear, hear-talking, couldn-understand
14	Media and tech	31 theater, anime, movie-like, tv; 57 minecraft, games, vr, game-like; 90 game-dream, dream-playing, vr, video-games; 219 cartoon, documentary, girl-guy, lisa; 236 batman, marvel, thanos, loki; 285 spongebob, episode, freeze, robots
15	Religious and spiritual	66 demons, devil, monster, demonic; 68 church, cult, lot-people, dream-god; 70 demon, devil, demons, angel; 109 religious, atheist, catholic, supernatural; 110 hell, purgatory, afterlife, heaven-hell; 137 powers, abilities, superpowers, telekinesis; 229 cult, leaders, satanic, ritual
16	Life events	59 birth, babies, pregnancy, newborn; 72 party, invited, having-party, brother; 78 giving-birth, dreamt, dream, twins; 190 party, party-dream, dream-party, dinner; 198 wedding, ceremony, aisle, venue; 252 wedding, getting-married, engaged, ceremony
17	Human body, especially teeth and blood	35 blood, skin, humanoid, arms; 108 teeth, falling, tongue, pain; 130 teeth, tooth, falling, gums; 133 blood, like-blood, covered-blood, splatter; 254 bodies, corpses, dead-bodies, body-parts
18	Work	47 new-job, boss, jobs, shift
19	Weather, especially storms	79 rain, tornado, hurricane, started-raining; 104 snow, snowing, cold, winter; 271 tornado, storms, april, category
20	Time, time travel and timelines	102 timeline, time-travel, time-skip, like-time; 165 time-travel, dream-world, universes, dream-time; 187 noon, 00am, early-morning, evening
21	Space	159 sun, earth, eclipse, phases; 221 space, space-ship, nasa, oxygen; 251 meteors, earth, asteroid, coming