

Bradley Voytek, Ph.D.  
UC San Diego

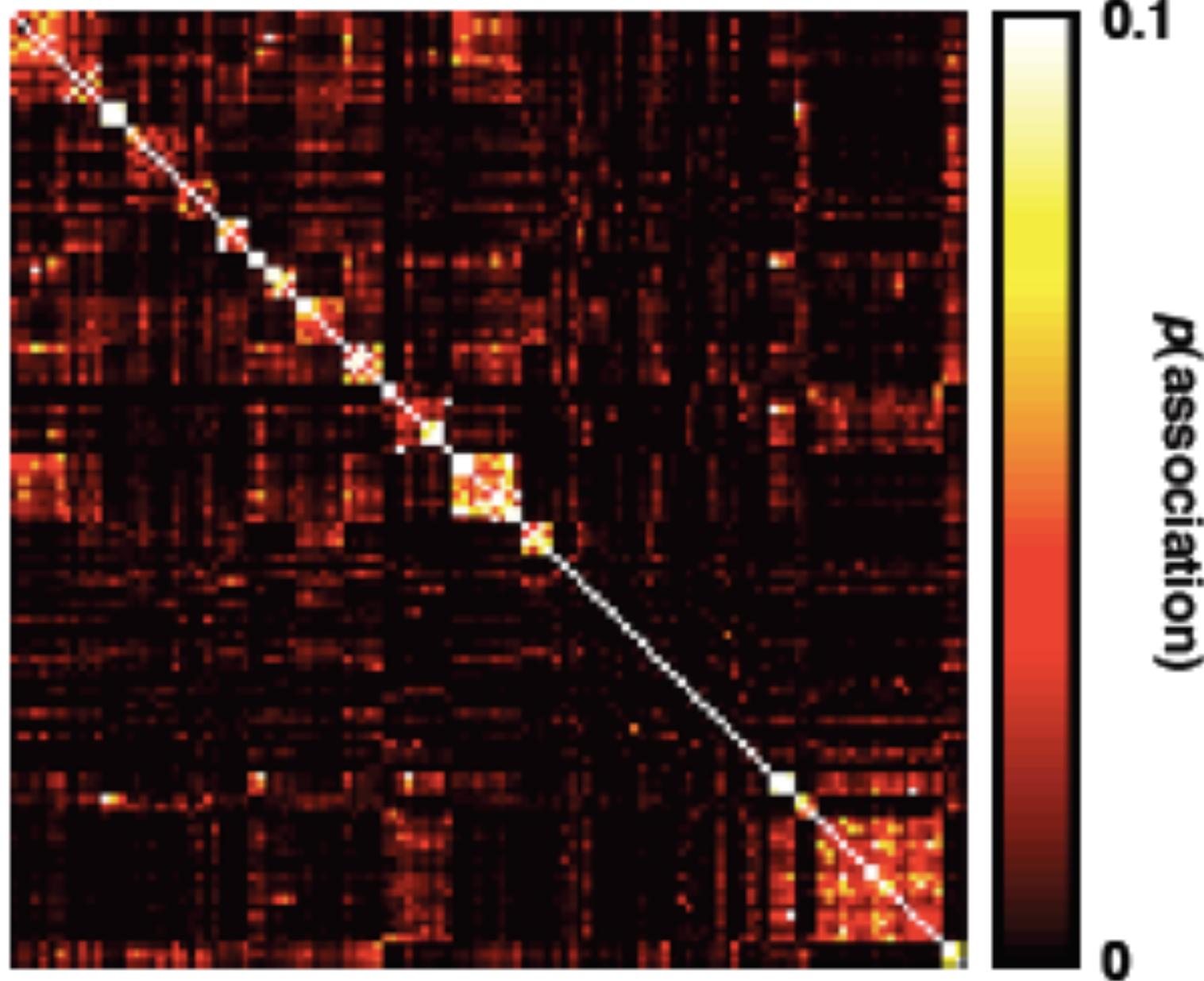
Department of Cognitive Science  
Neurosciences Graduate Program  
Halicioglu Data Science Institute

bvoytek@ucsd.edu  
@bradleyvoytek

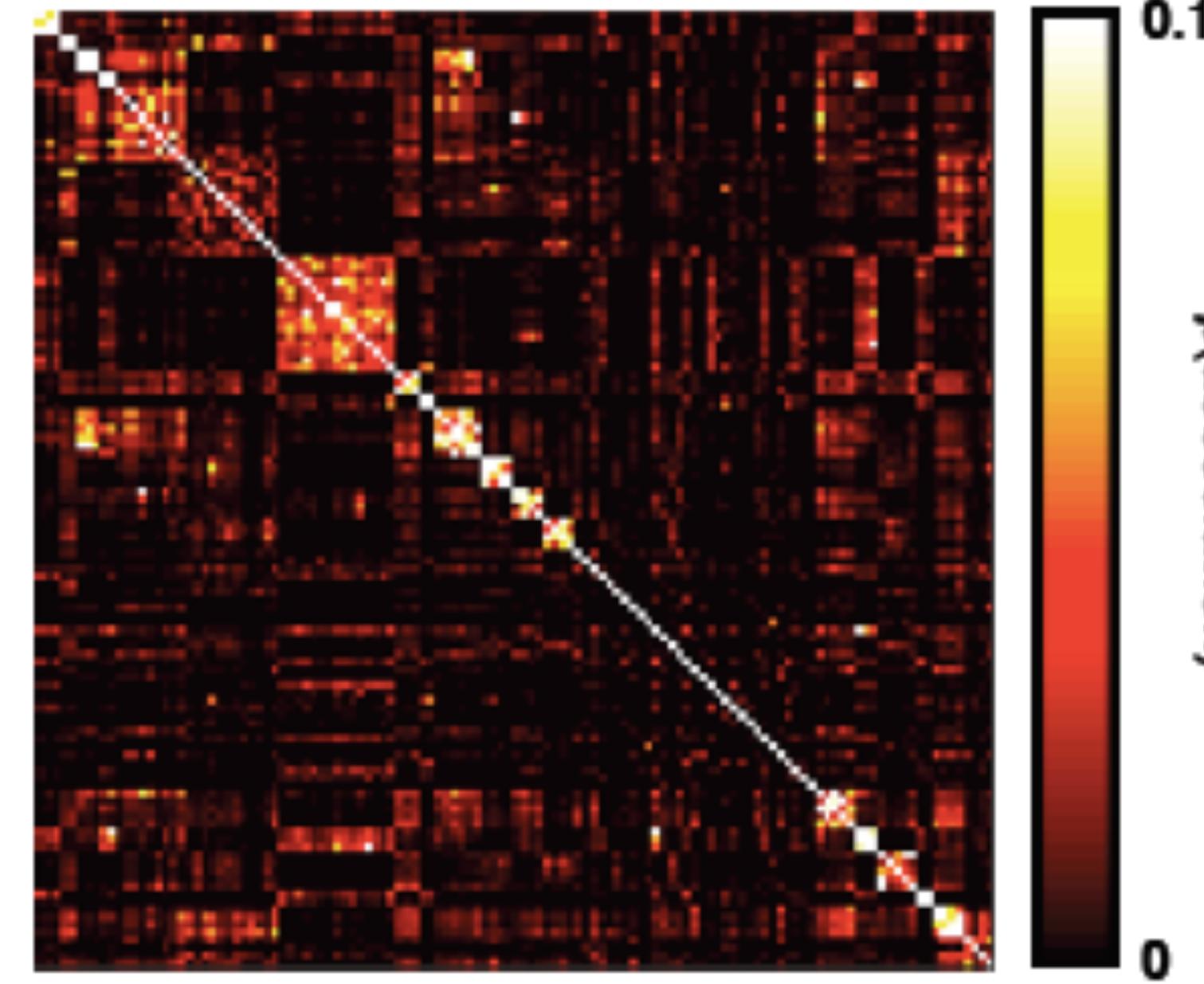
UC San Diego

# The human cognome

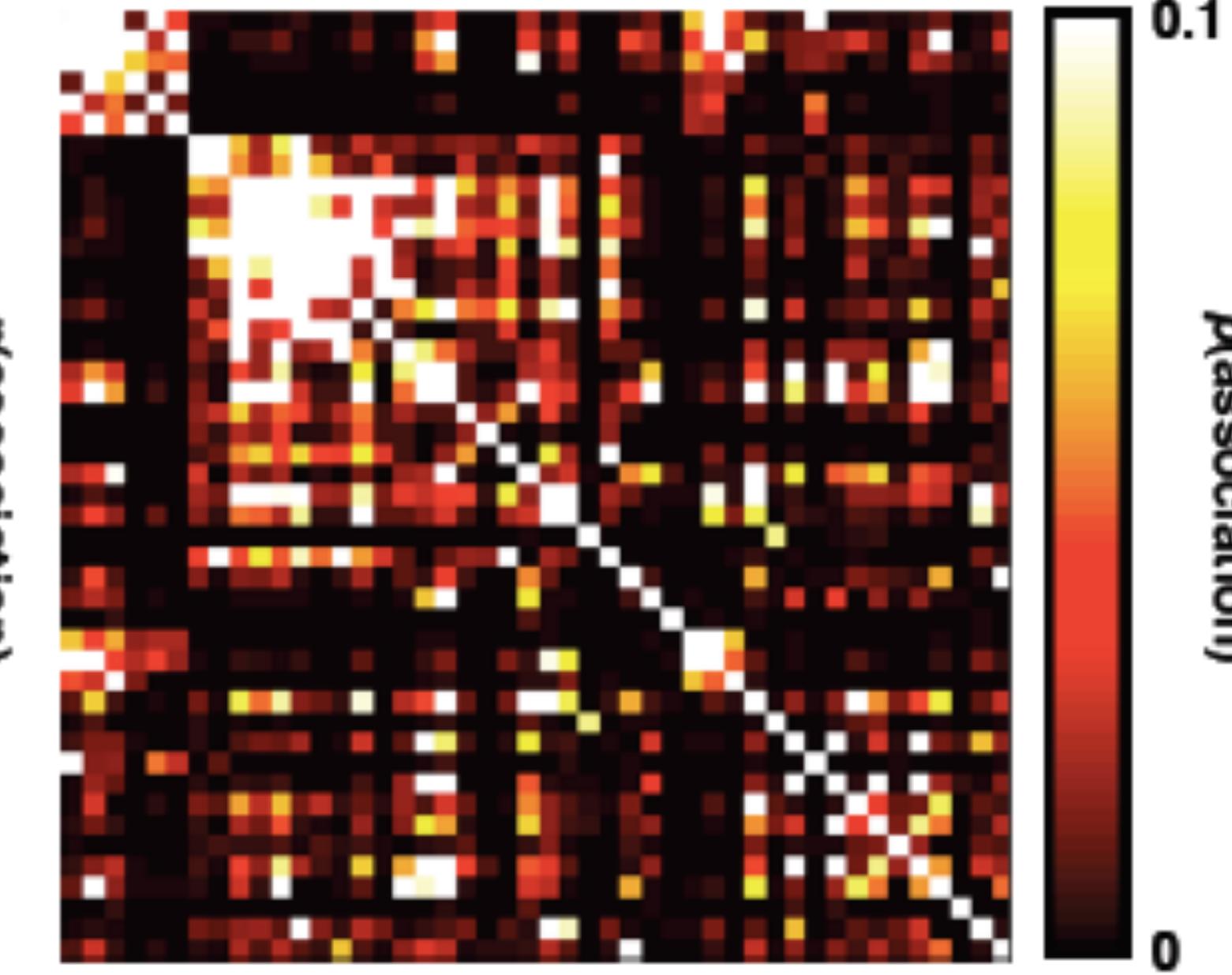
**Structures**



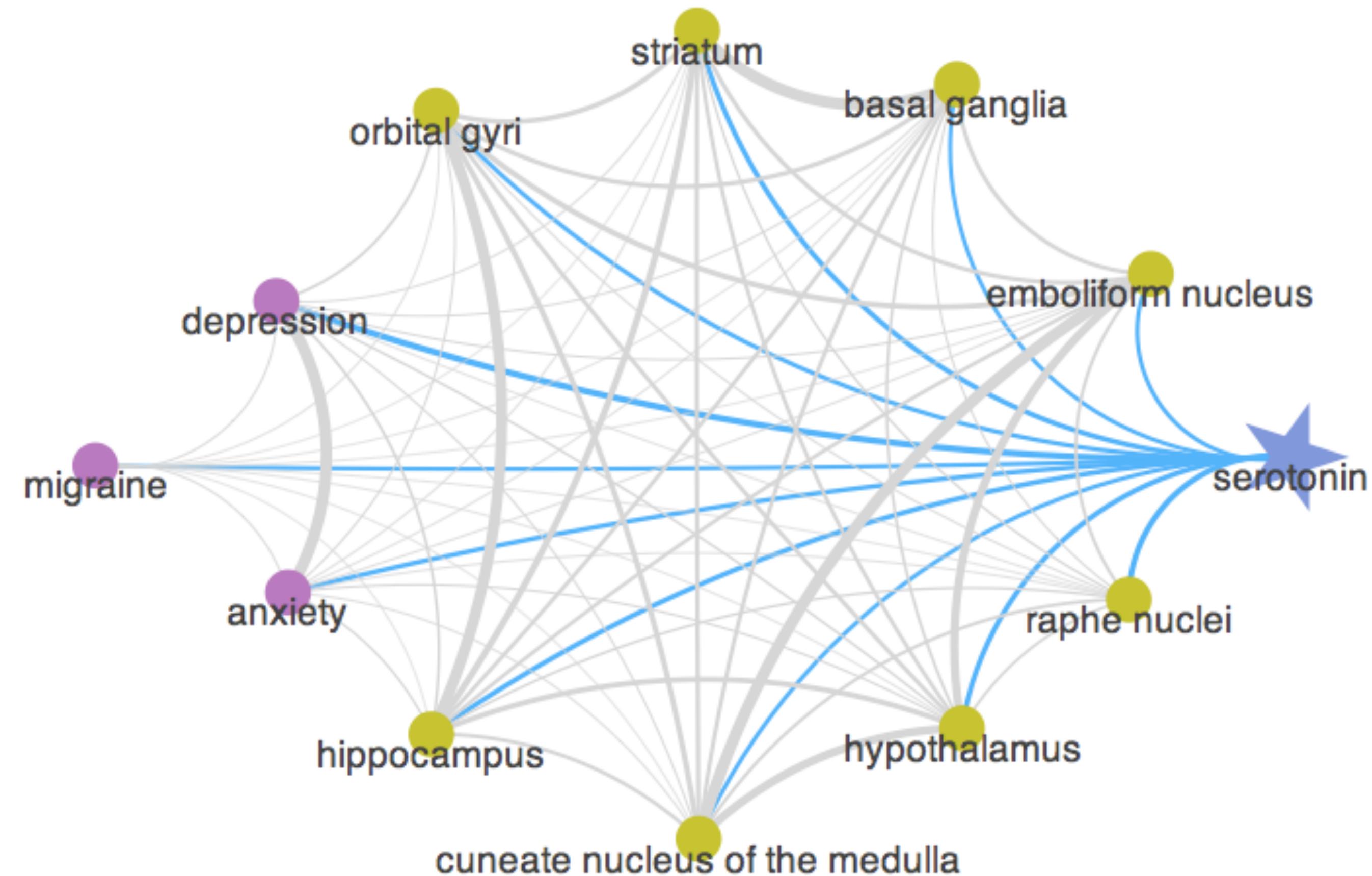
**Functions**

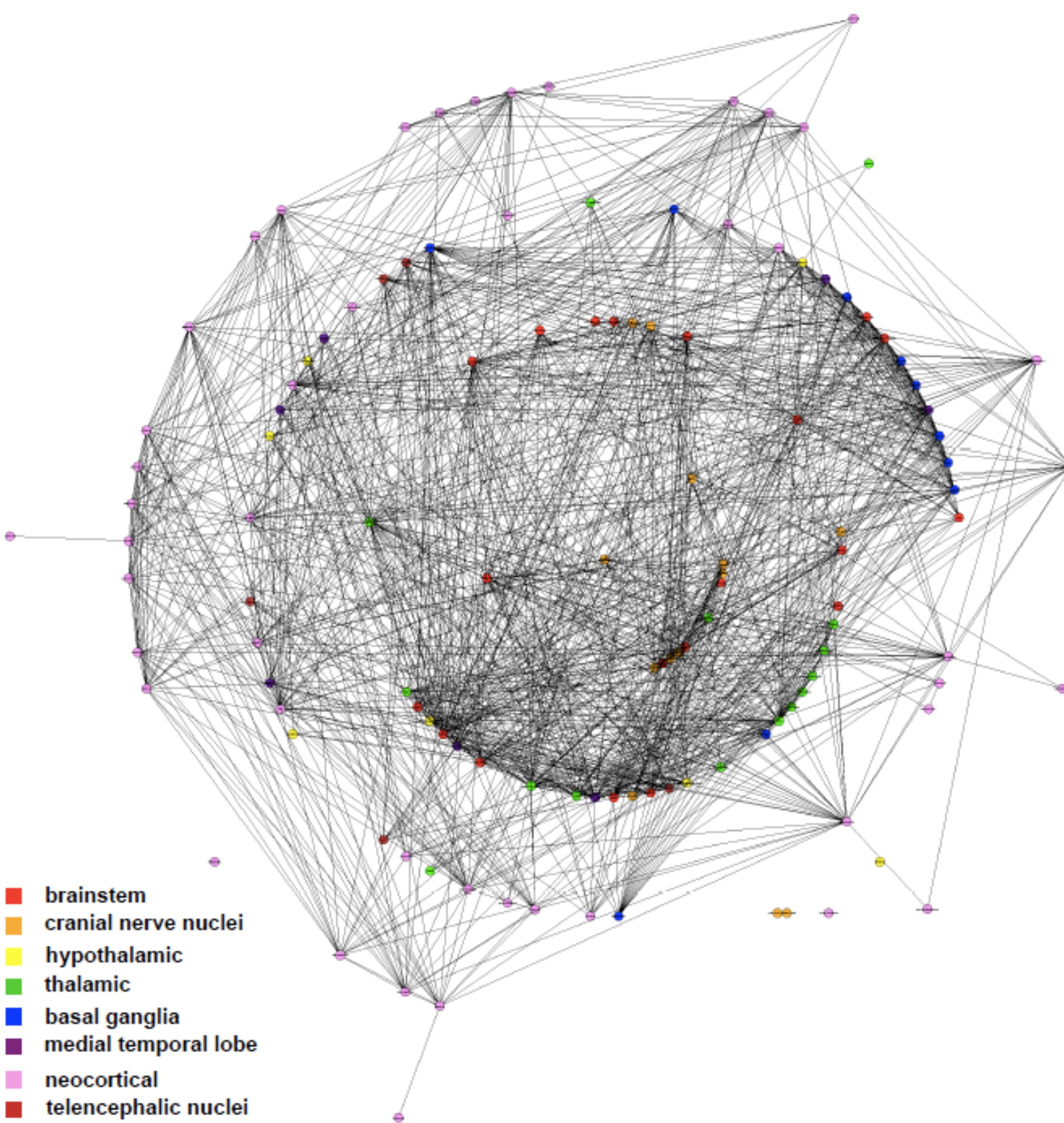


**Diseases**

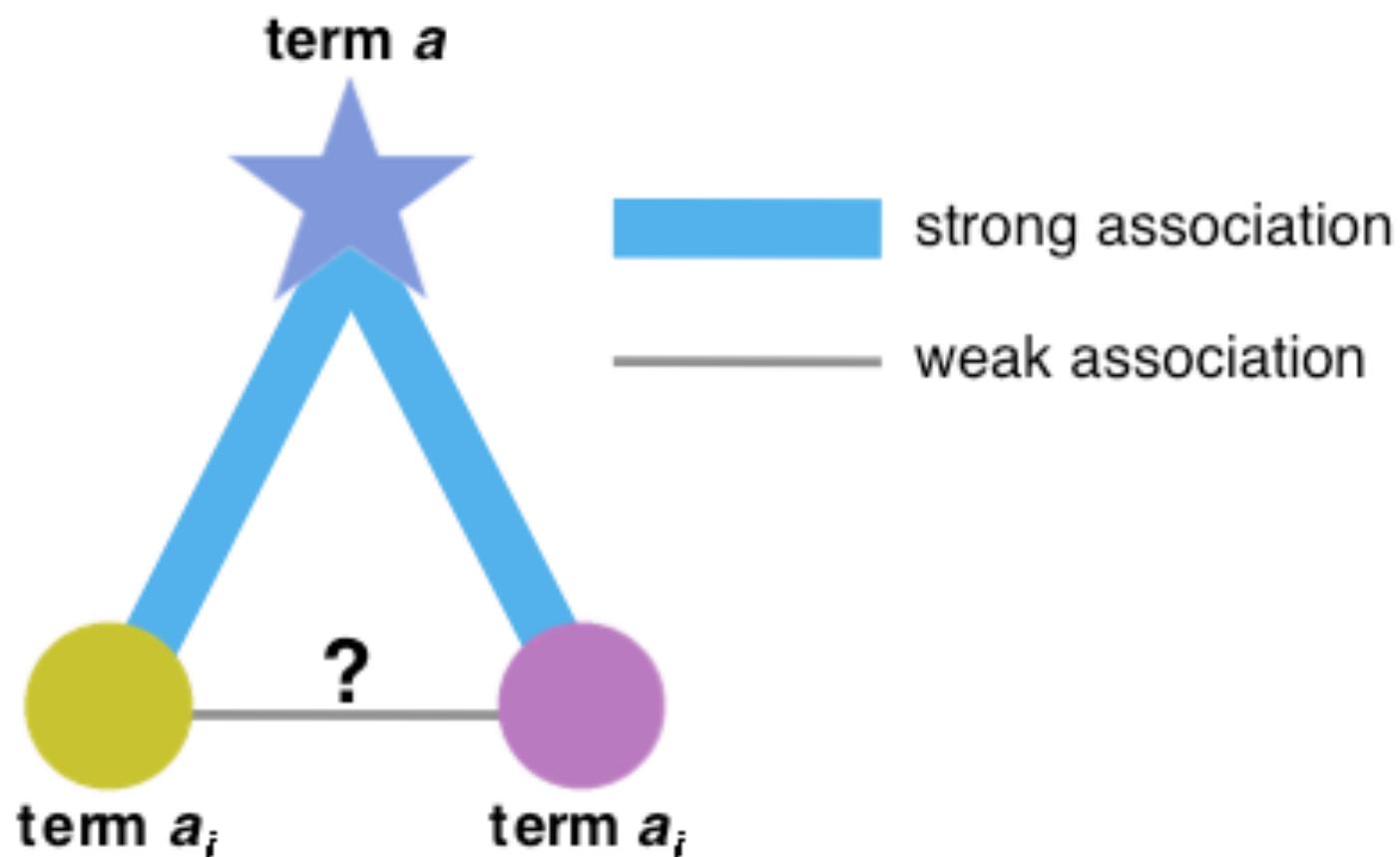


# brainSCANr

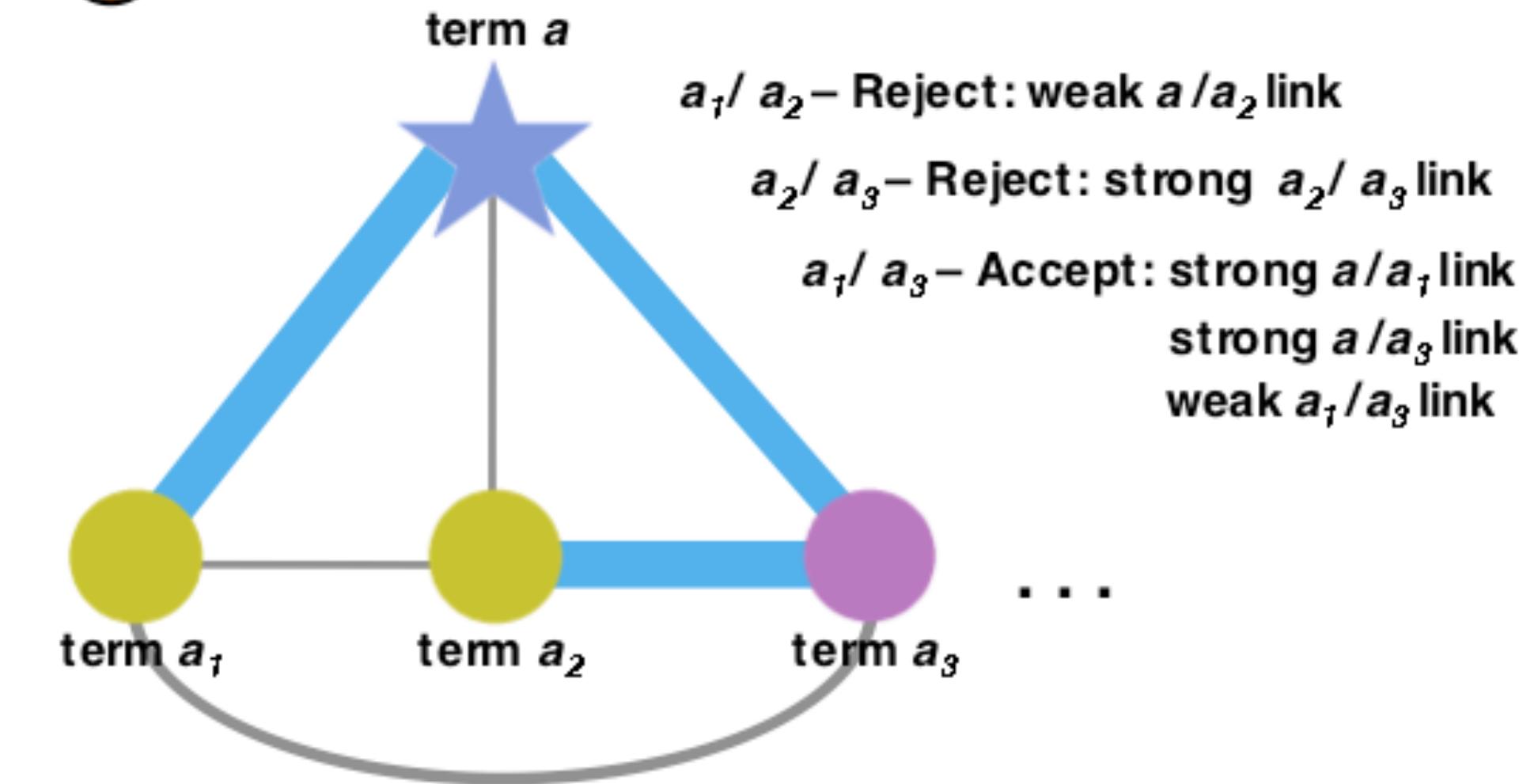




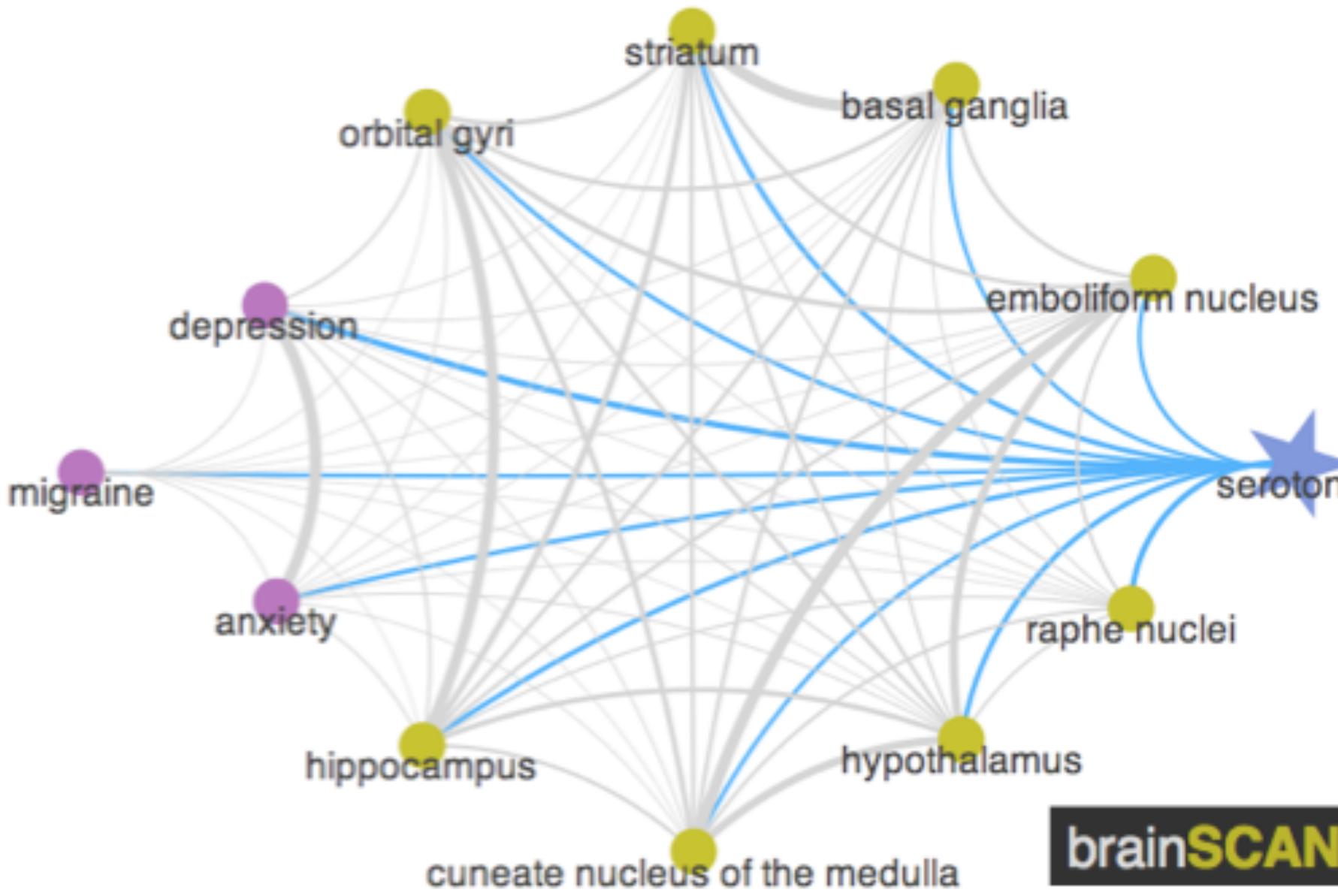
### A Hypothesis-generation model



### B Algorithmically generate hypotheses

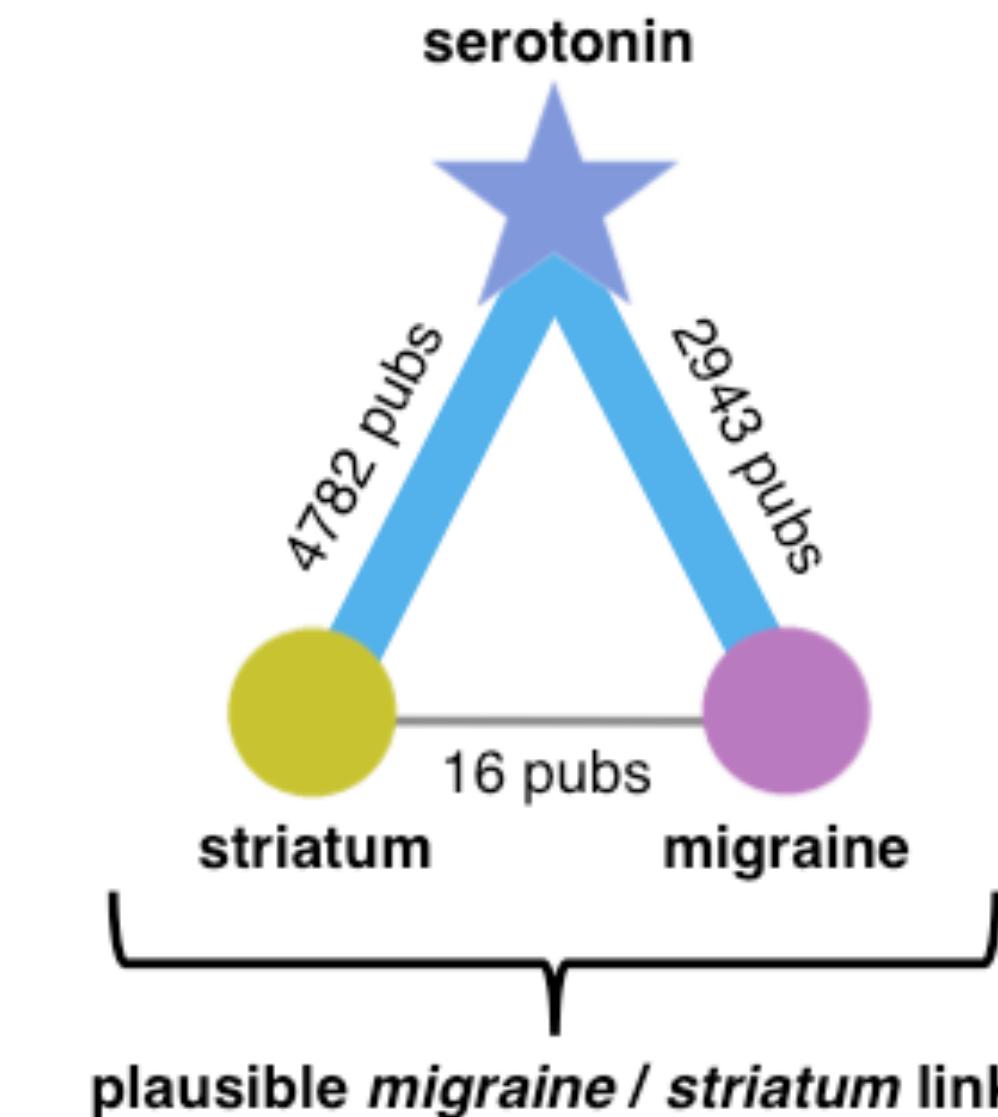


### C Visualize topic network

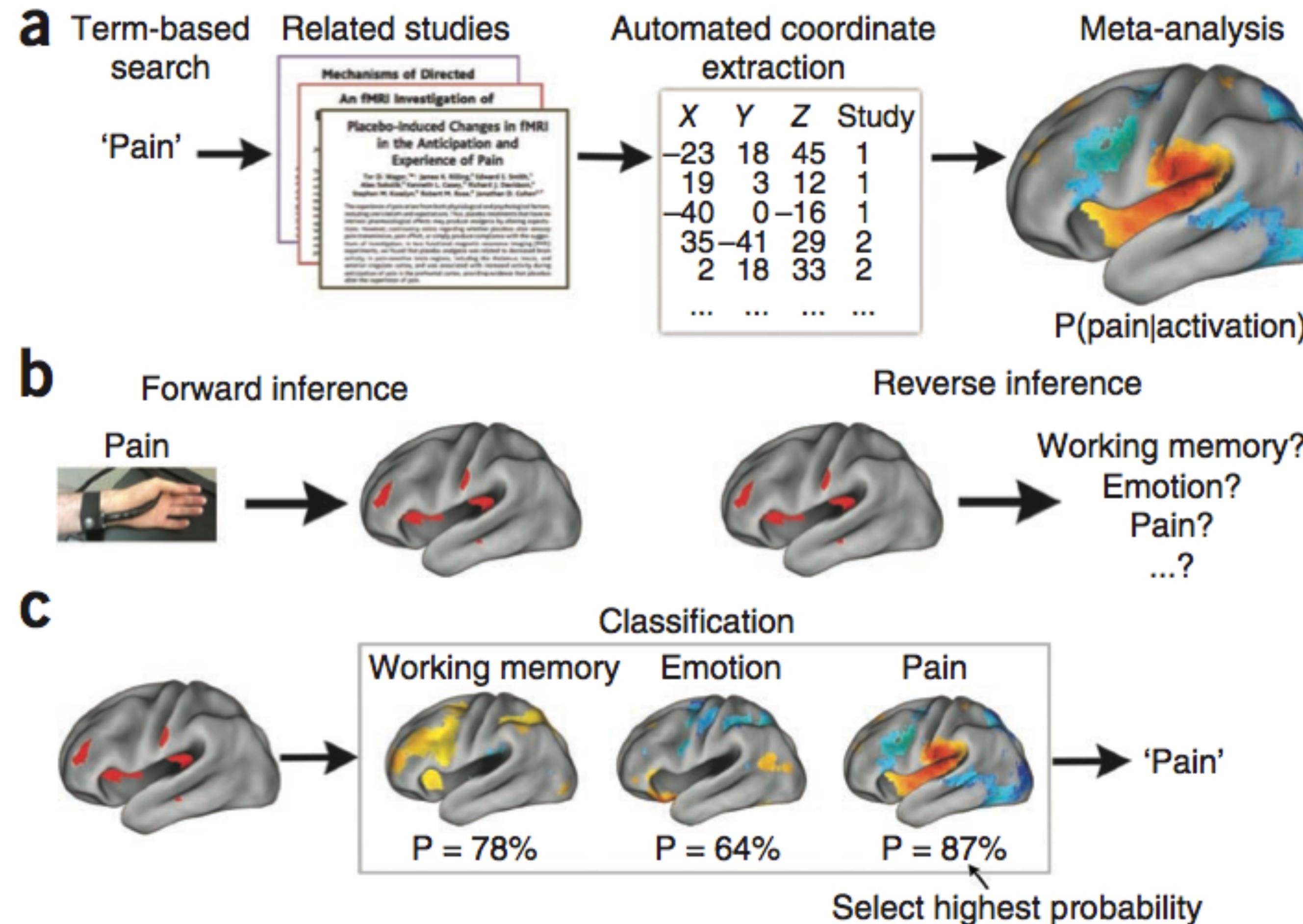


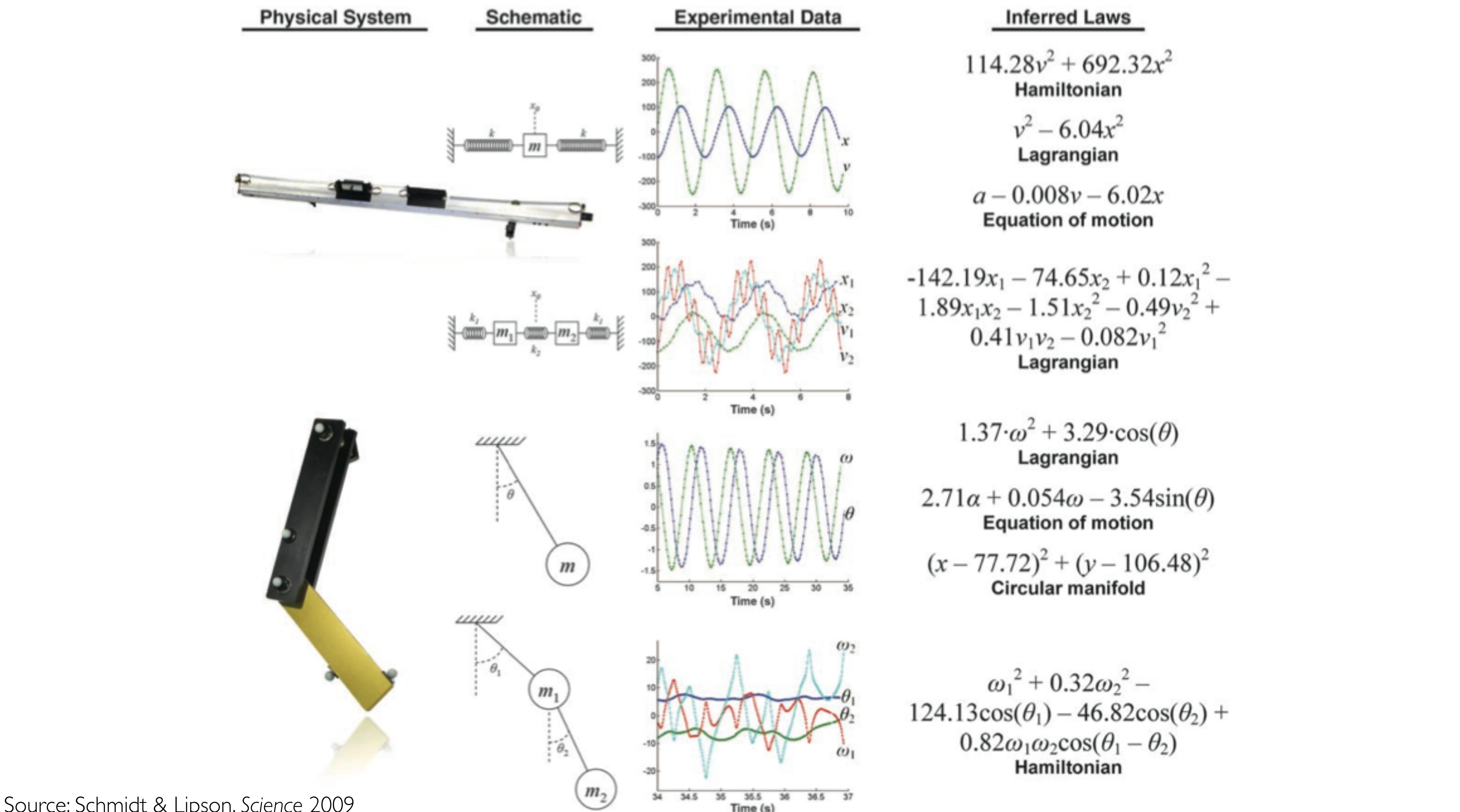
brainSCAnr

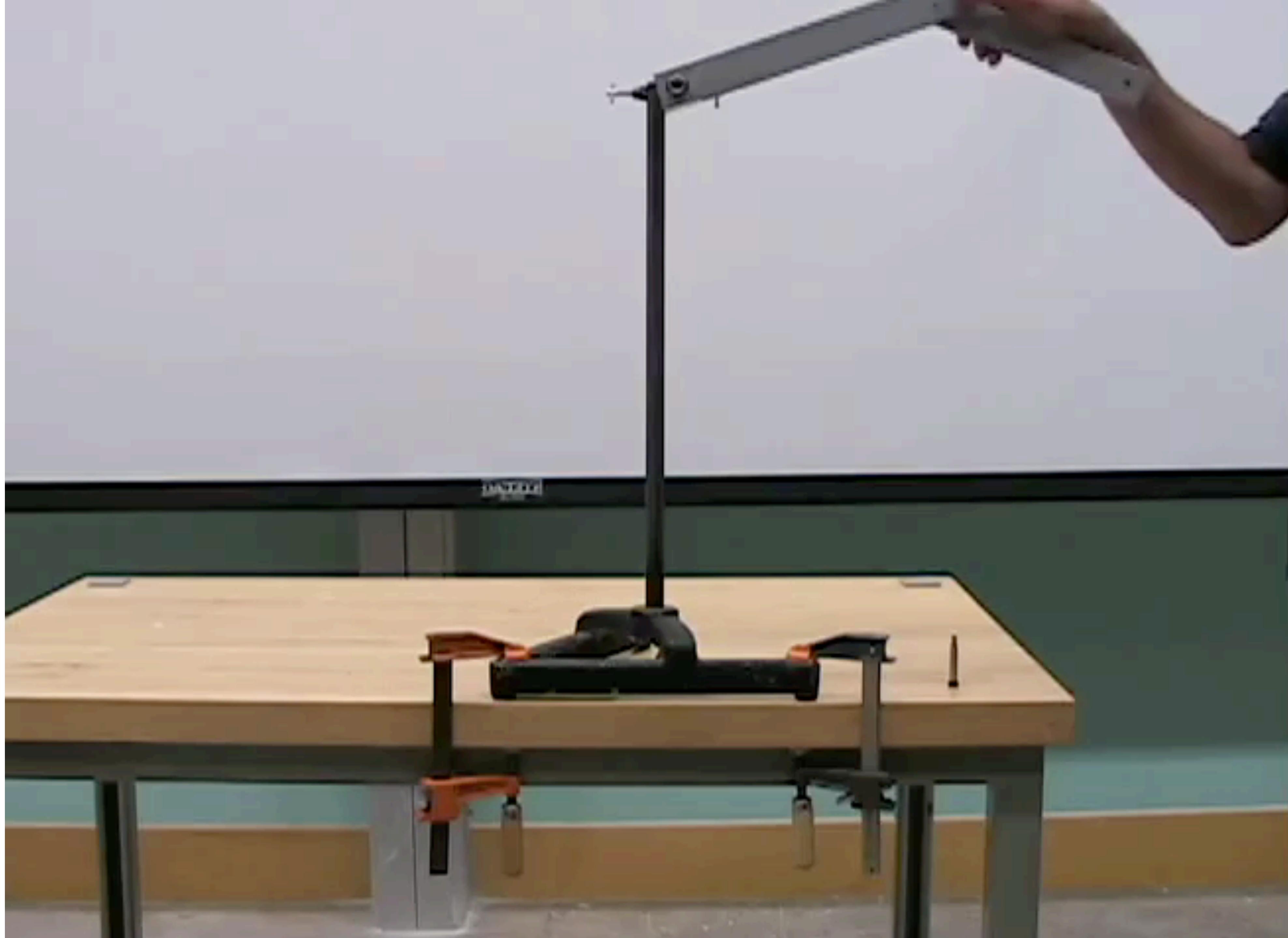
### D Assess relative topic weights



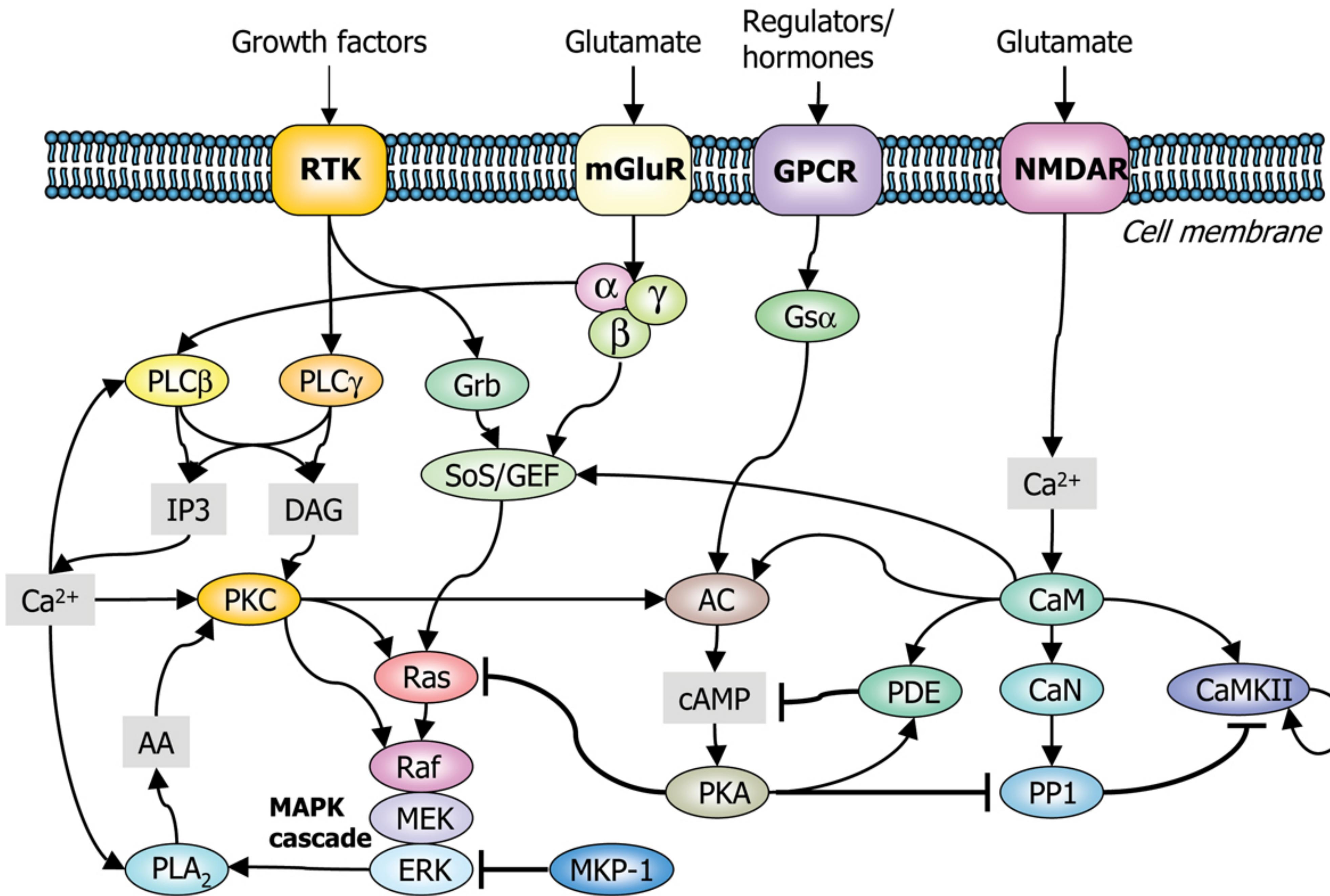
# NeuroSynth



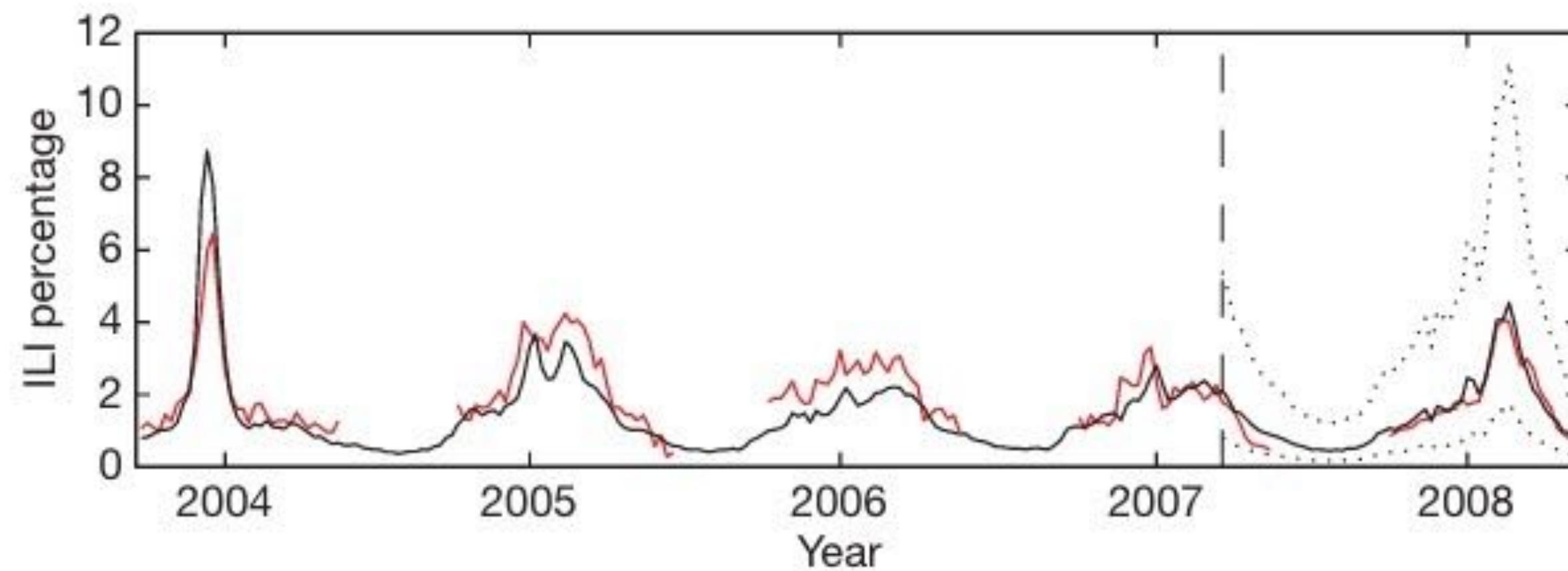




# Signaling network in neurons



# Google Flu Trends



**Figure 2 | A comparison of model estimates for the mid-Atlantic region (black) against CDC-reported ILI percentages (red), including points over which the model was fit and validated.** A correlation of 0.85 was obtained over 128 points from this region to which the model was fit, whereas a correlation of 0.96 was obtained over 42 validation points. Dotted lines indicate 95% prediction intervals. The region comprises New York, New Jersey and Pennsylvania.

# Verb regularization

**Table 1 | The 177 irregular verbs studied**

Frequency	Verbs	Regularization (%)	Half-life (yr)
$10^{-1}-1$	be, have	0	38,800
$10^{-2}-10^{-1}$	come, do, find, get, give, go, know, say, see, take, think	0	14,400
$10^{-3}-10^{-2}$	begin, break, bring, buy, choose, draw, drink, drive, eat, fall, fight, forget, grow, hang, <b>help</b> , hold, leave, let, lie, lose, <b>reach</b> , rise, run, seek, set, shake, sit, sleep, speak, stand, teach, throw, understand, <b>walk</b> , win, <b>work</b> , write	10	5,400
$10^{-4}-10^{-3}$	arise, <b>bake</b> , bear, beat, bind, bite, blow, <b>bow</b> , burn, burst, carve, chew, climb, cling, creep, <b>dare</b> , dig, <b>drag</b> , flee, <b>float</b> , <b>flow</b> , fly, <b>fold</b> , freeze, grind, leap, lend, <b>lock</b> , melt, <b>reckon</b> , ride, <b>rush</b> , <b>shape</b> , shine, shoot, shrink, <b>sigh</b> , sing, sink, slide, slip, <b>smoke</b> , spin, spring, <b>starve</b> , steal, <b>step</b> , <b>stretch</b> , strike, <b>stroke</b> , <b>suck</b> , <b>swallow</b> , swear, sweep, swim, swing, tear, wake, <b>wash</b> , weave, weep, <b>weigh</b> , wind, <b>yell</b> , <b>yield</b>	43	2,000
$10^{-5}-10^{-4}$	bark, <b>bellow</b> , bid, blend, braid, brew, cleave, cringe, crow, dive, <b>drip</b> , fare, fret, glide, gnaw, grip, heave, knead, low, milk, mourn, mow, prescribe, reddens, reek, row, scrape, <b>seethe</b> , shear, shed, <b>shove</b> , slay, slit, <b>smite</b> , sow, span, spurn, sting, stink, strew, stride, swell, <b>tread</b> , uproot, wade, <b>warp</b> , wax, wield, wring, <b>writhe</b>	72	700
$10^{-6}-10^{-5}$	bide, chide, delve, flay, <b>hew</b> , rue, shrive, slink, <b>snip</b> , spew, sup, wreak	91	300

177 Old English irregular verbs were compiled for this study. These are arranged according to frequency bin, and in alphabetical order within each bin. Also shown is the percentage of verbs in each bin that have regularized. The half-life is shown in years. Verbs that have regularized are indicated in red. As we move down the list, an increasingly large fraction of the verbs are red; the frequency-dependent regularization of irregular verbs becomes immediately apparent.

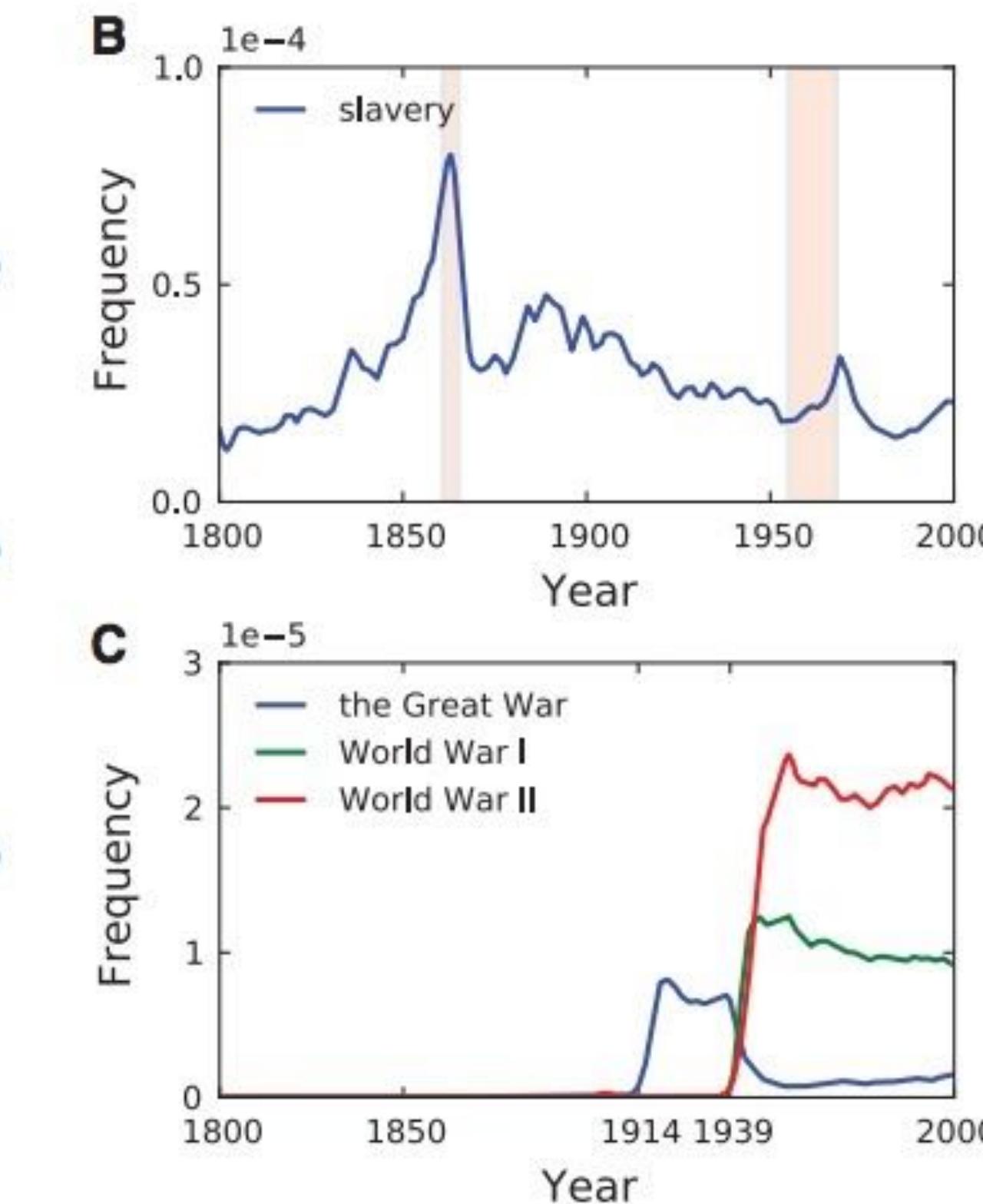
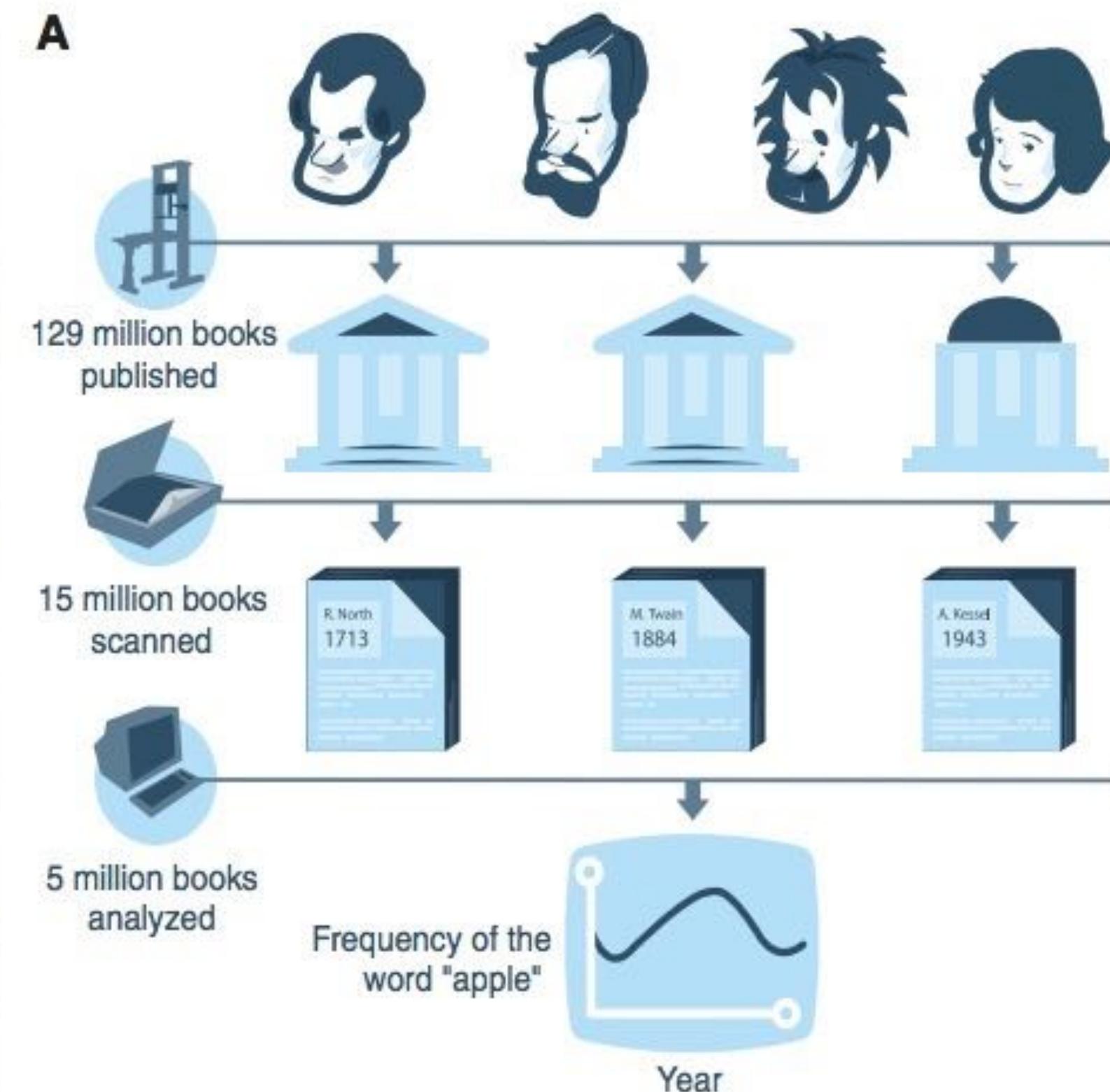
# Plurals



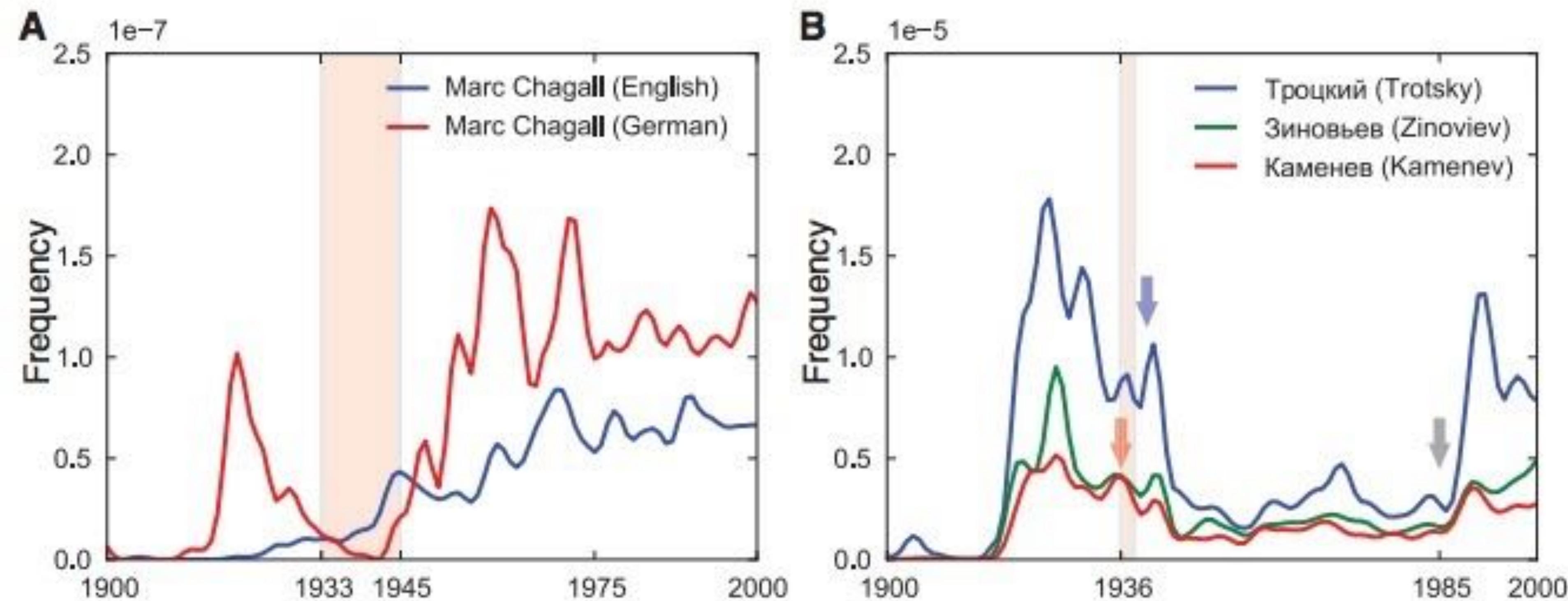
# Culturomics

**Fig. 1.** Culturomic analyses study millions of books at once. (A) Top row: Authors have been writing for millennia; ~129 million book editions have been published since the advent of the printing press (upper left). Second row: Libraries and publishing houses provide books to Google for scanning (middle left). Over 15 million books have been digitized. Third row: Each book is associated with metadata. Five million books are chosen for computational analysis (bottom left). Bottom row: A culturomic time line shows the frequency of “apple” in English books over time (1800–2000).

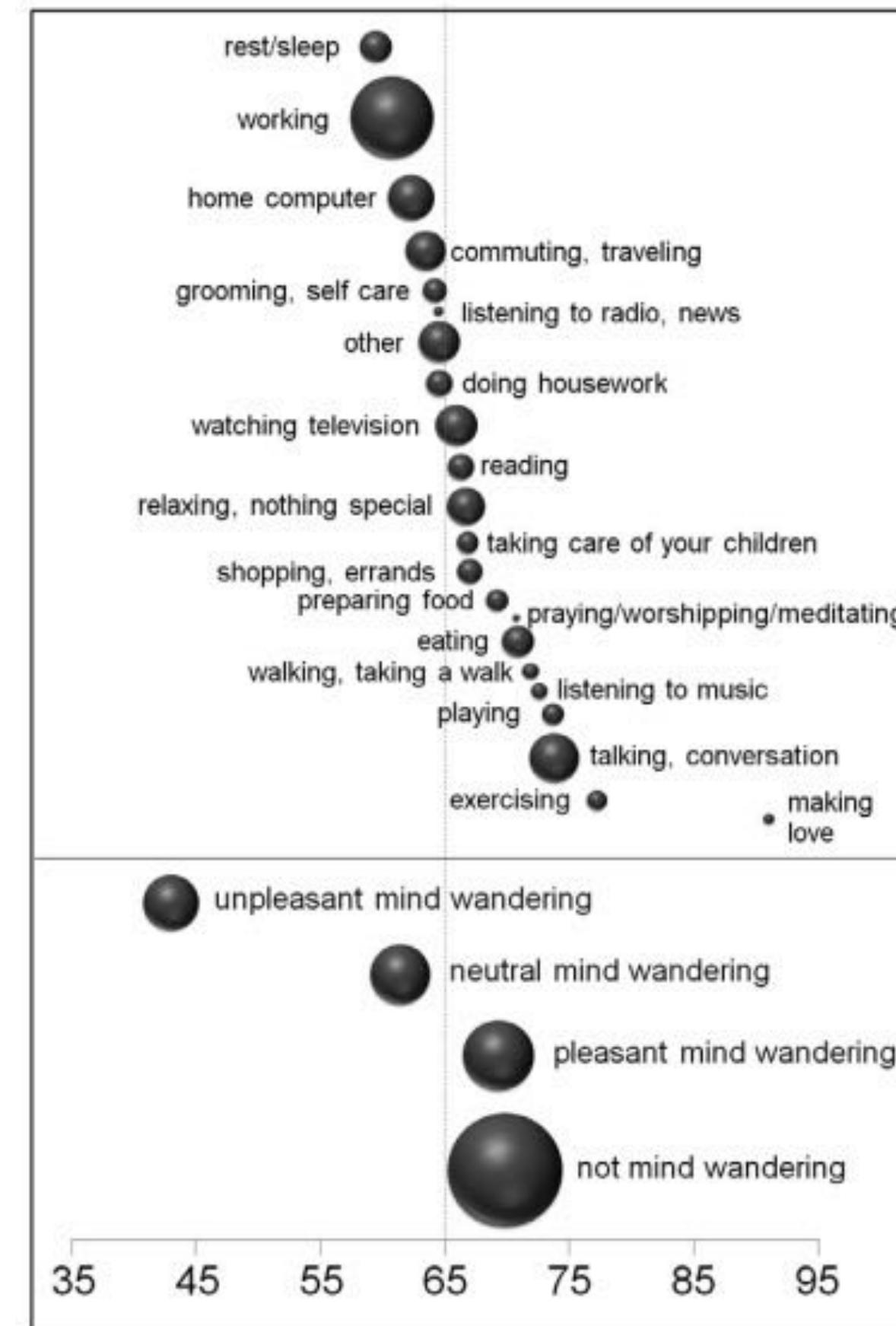
(B) Usage frequency of “slavery”. The Civil War (1861–1865) and the civil rights movement (1955–1968) are highlighted in red. The number in the upper left ( $1e-4 = 10^{-4}$ ) is the unit of frequency. (C) Usage frequency over time for “the Great War” (blue), “World War I” (green), and “World War II” (red).



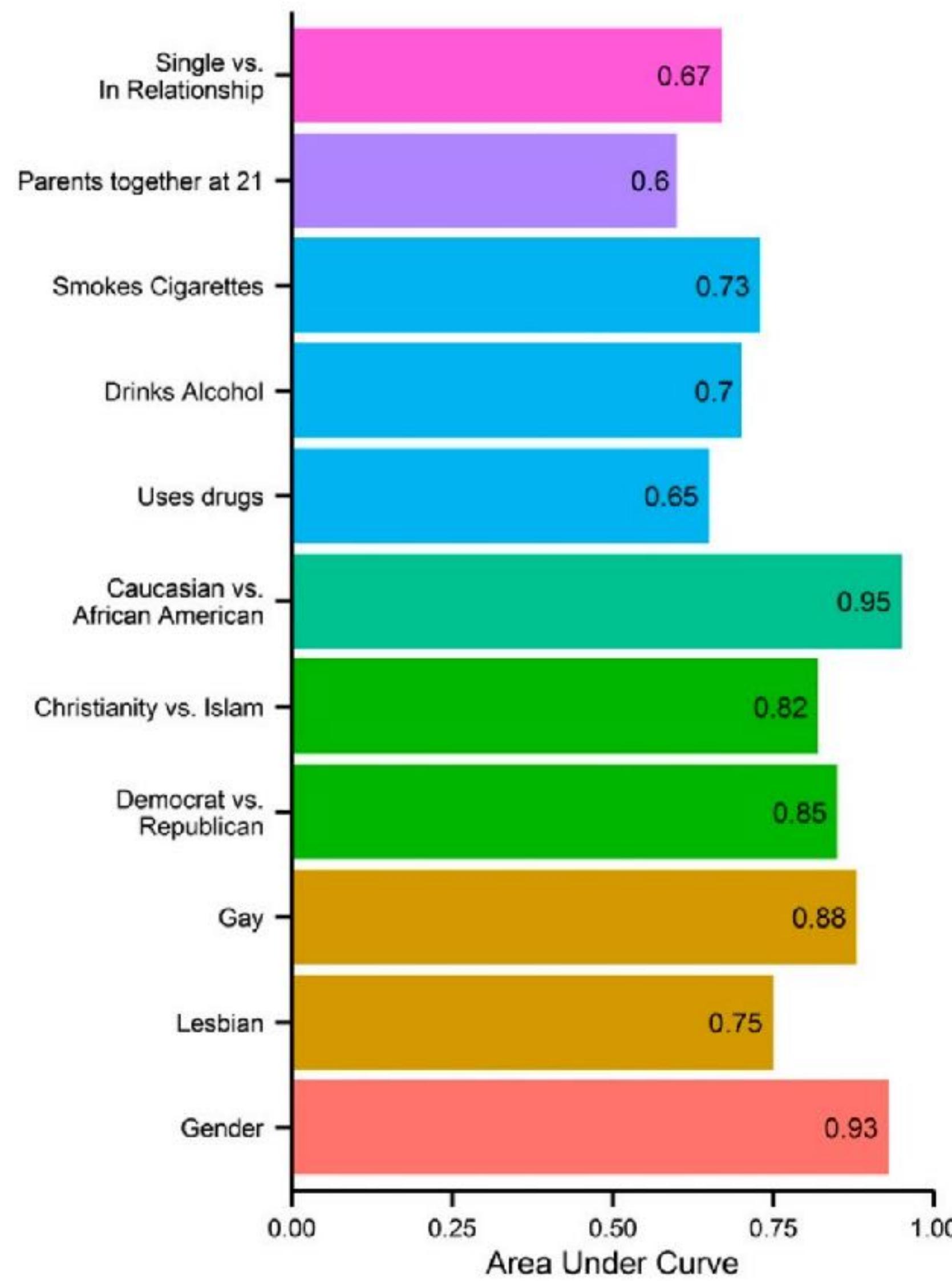
# Culturomics



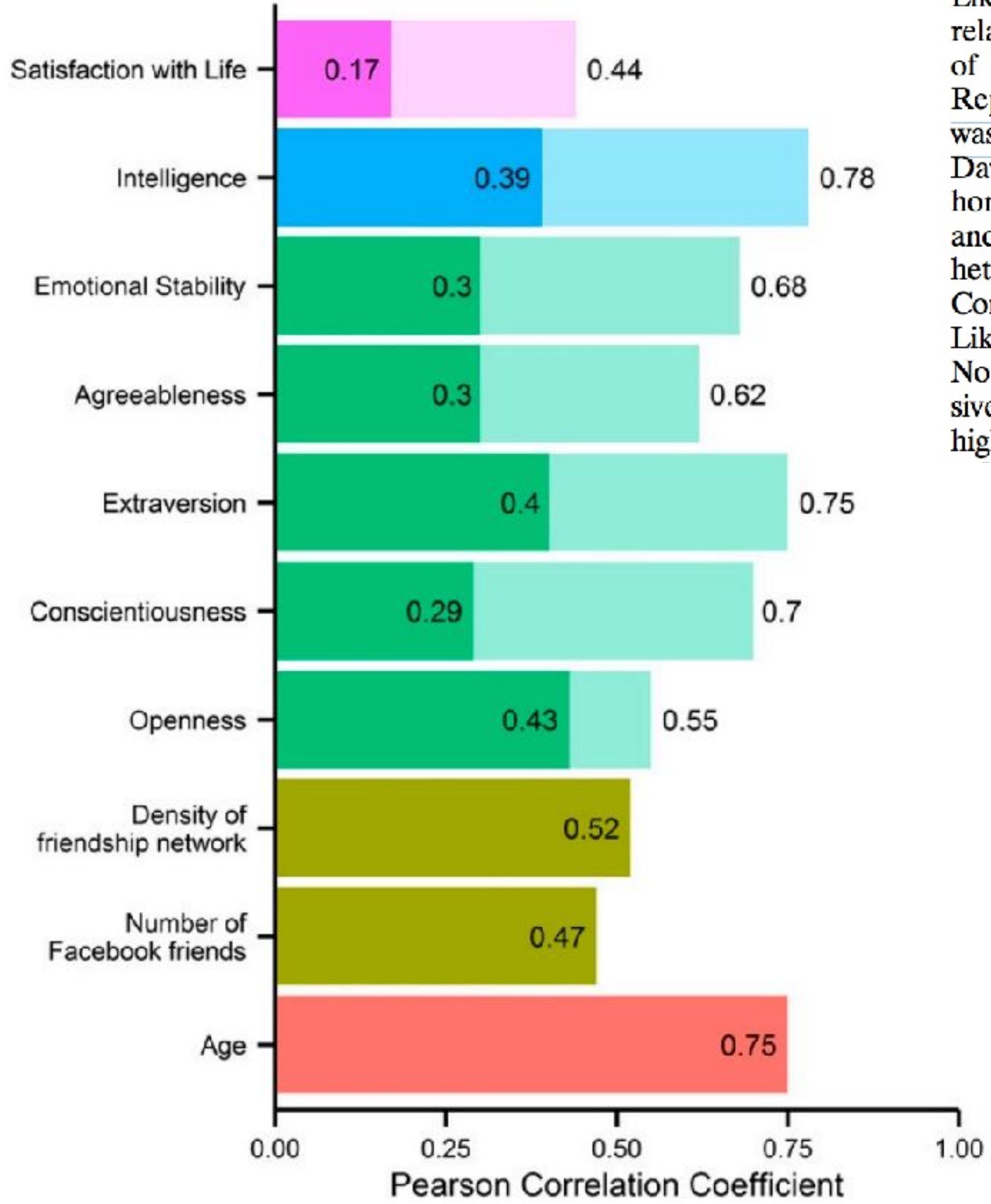
# Texting and Mood



**Fig. 1.** Mean happiness reported during each activity (top) and while mind wandering to unpleasant topics, neutral topics, pleasant topics or not mind wandering (bottom). Dashed line indicates mean of happiness across all samples. Bubble area indicates the frequency of occurrence. The largest bubble ("not mind wandering") corresponds to 53.1% of the samples, and the smallest bubble ("praying/worshipping/meditating") corresponds to 0.1% of the samples.



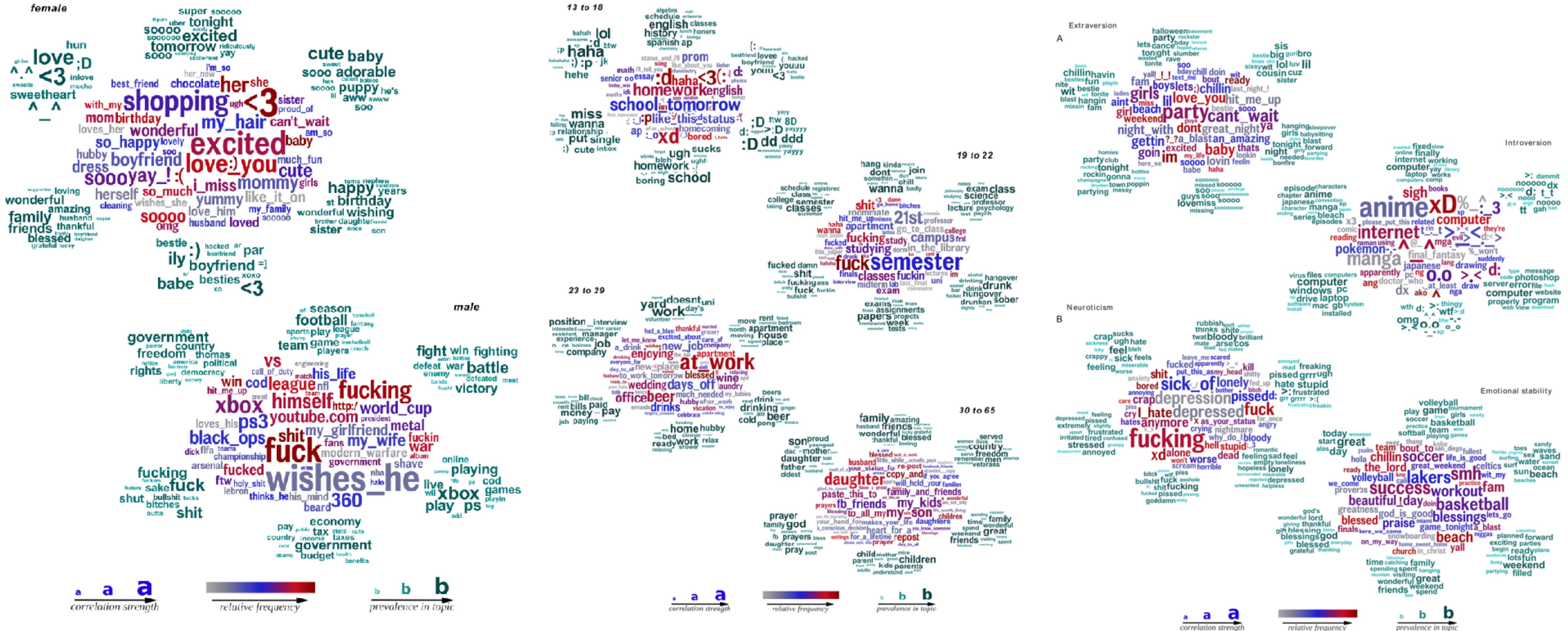
**Fig. 2.** Prediction accuracy of classification for dichotomous/dichotomized attributes expressed by the AUC.



**Fig. 3.** Prediction accuracy of regression for numeric attributes and traits expressed by the Pearson correlation coefficient between predicted and actual attribute values; all correlations are significant at the  $P < 0.001$  level. The transparent bars indicate the questionnaire's baseline accuracy, expressed in terms of test-retest reliability.

**Predictive Power of Likes.** Individual traits and attributes can be predicted to a high degree of accuracy based on records of users' Likes. **Table S1** presents a sample of highly predictive Likes related to each of the attributes. For example, the best predictors of high intelligence include "Thunderstorms," "The Colbert Report," "Science," and "Curly Fries," whereas low intelligence was indicated by "Sephora," "I Love Being A Mom," "Harley Davidson," and "Lady Antebellum." Good predictors of male homosexuality included "No H8 Campaign," "Mac Cosmetics," and "Wicked The Musical," whereas strong predictors of male heterosexuality included "Wu-Tang Clan," "Shaq," and "Being Confused After Waking Up From Naps." Although some of the Likes clearly relate to their predicted attribute, as in the case of No H8 Campaign and homosexuality, other pairs are more elusive; there is no obvious connection between Curly Fries and high intelligence.

# Facebook

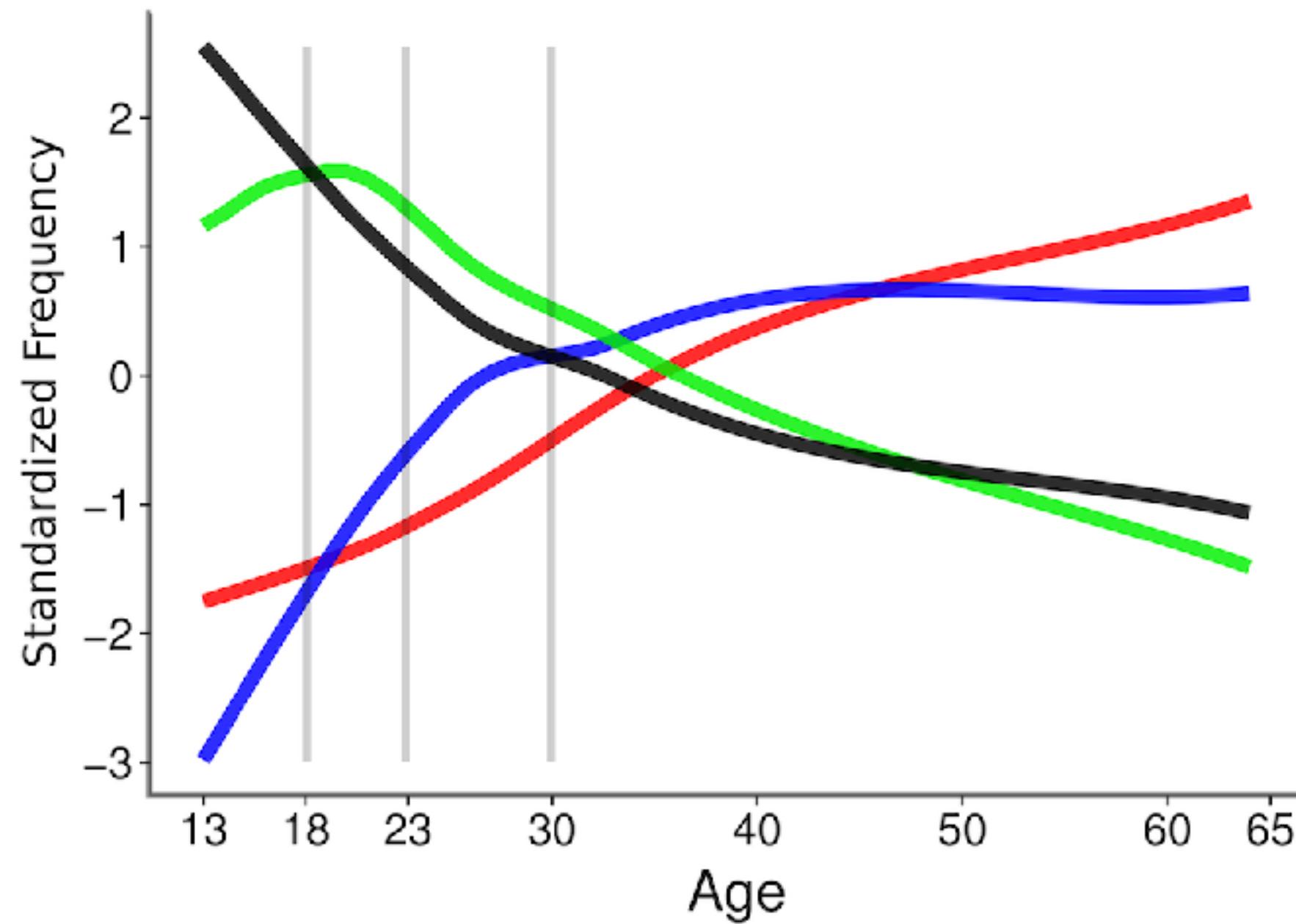


Source: Schwartz et al., PLoS ONE 2013

# Facebook identification

**A**

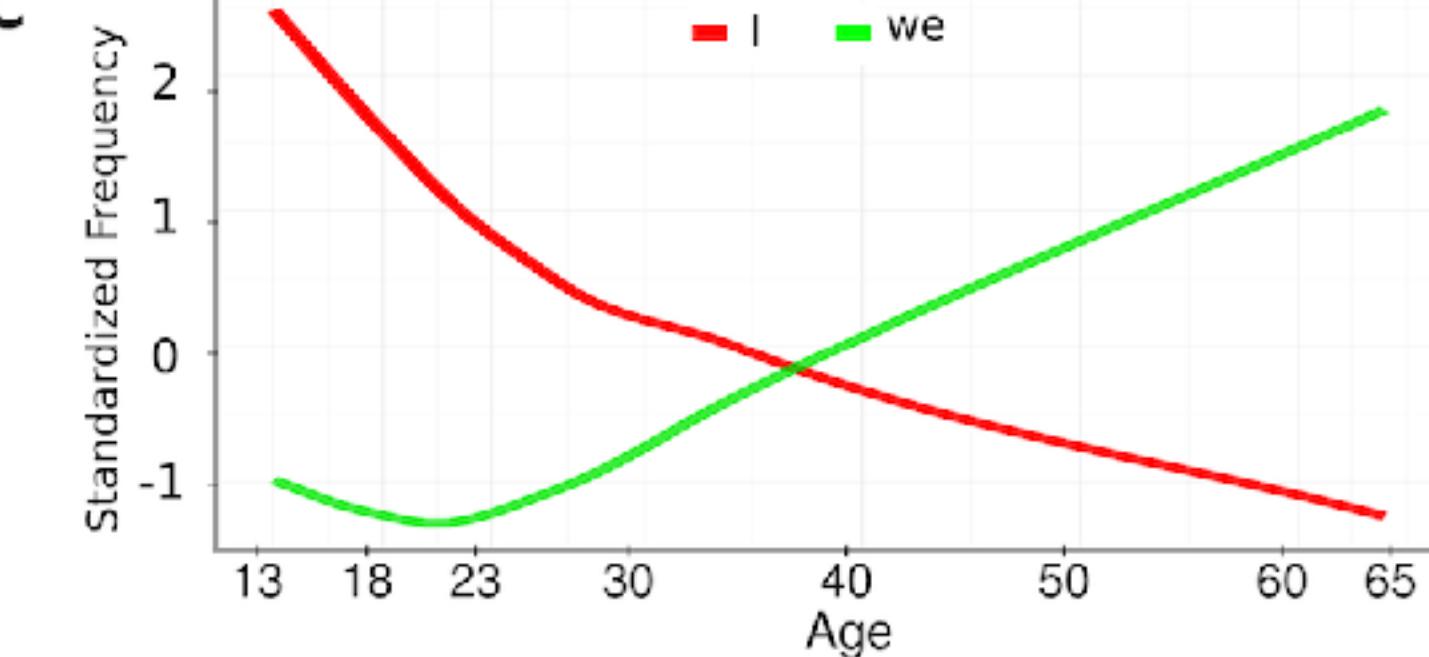
- (30 to 65) ■ son daughter father mother proud oldest data youngest  
(23 to 29) ■ job position company manager interview experience office assistant  
(19 to 22) ■ classes semester class college schedule summer registered taking  
(13 to 18) ■ haha lol :p :D ;) hehe jk ;p



**B**



**C**



Bradley Voytek, Ph.D.  
UC San Diego

Department of Cognitive Science  
Neurosciences Graduate Program  
Halicioglu Data Science Institute

bvoytek@ucsd.edu  
@bradleyvoytek

UC San Diego