

Distributional Asymmetries in Online Nutrition Discourse Across Social Determinants of Health Domains on Reddit

Maria Erialnta Loufi¹, Vasiliki Bountziouka¹, Efstathios Kaloudis^{1*}

¹Computer Simulation, Genomics and Data Analysis Laboratory, Department of Food Science and Nutrition, University of the Aegean, Lemnos, Greece

INTRODUCTION & AIM

- Online nutrition discourse extends beyond explicitly health-focused communities, yet its distribution across broader social domains remains poorly quantified.
- The Social Determinants of Health (SDH) framework captures the way structural factors (healthcare access, socioeconomic status, education) shape health behaviors.
- Social Media (and Reddit in particular) offers a unique window into how nutrition information circulates across society.

Aim → to quantify the distribution of online nutrition discourse across SDH domains on Reddit using activity-weighted community analysis.

METHOD

- Starting corpus:** ~ 60000 subreddits, 2025 Reddit Pushshift dataset
- Thematic classification:** 3-level Wikipedia taxonomy
- Lexical shortlisting:** Confidence scoring & Not Safe For Work content overrides
- Nutrition detection:** Forced-choice Large Language Model single label
- SDH mapping:** Constrained outputs & refinement pass
- Weighting distribution:** Submission & comment volume

RESULTS & DISCUSSION

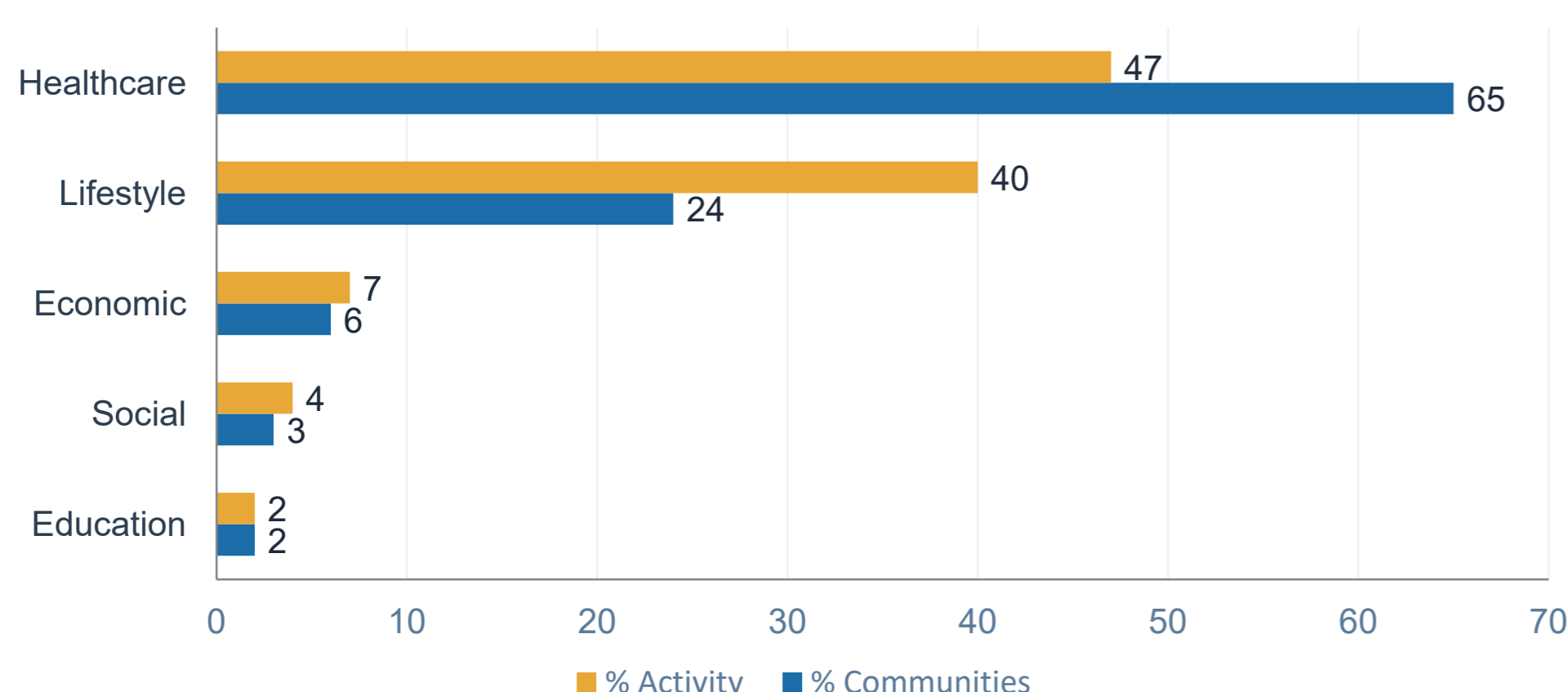


Figure 1. Community share (% of subreddits) vs. activity share (% of submissions + comments, 2025 Pushshift) by SDH domain. Healthcare leads in community count (47%), but lifestyle communities generate 40% of all activity from just 24% of communities. Economic, social, and education domains each account for <10%.

RESULTS & DISCUSSION

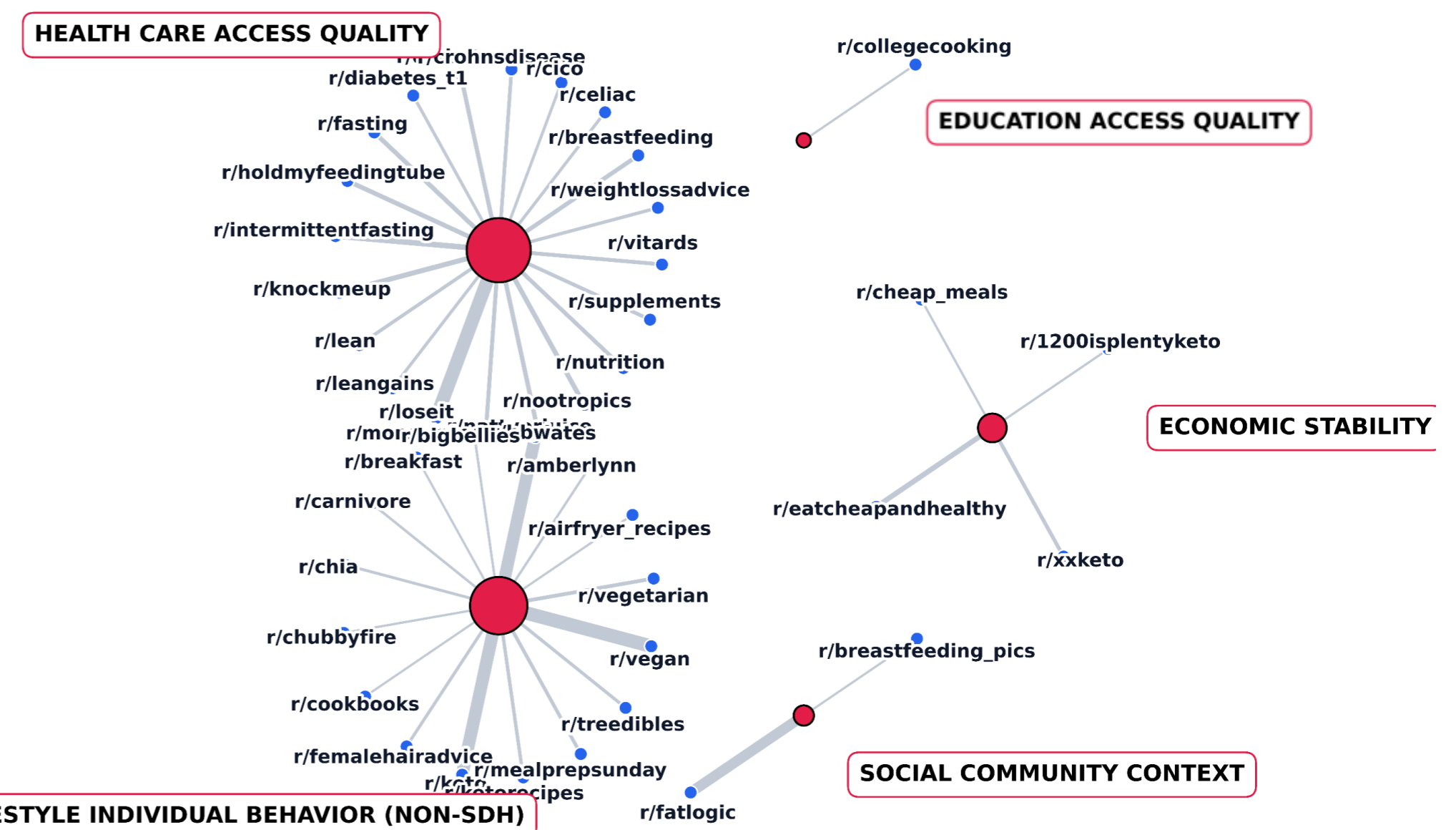


Figure 2. Weighted bipartite network mapping 43 nutrition subreddits to five SDH domains. Node size is proportional to lifetime content volume (submissions + comments); edge weights reflect classification confidence. Large lifestyle nodes (red points, left) contrast with the more numerous but smaller healthcare nodes (blue points, top-left), illustrating the engagement asymmetry between the two dominant domains.

The network reveals that healthcare communities (e.g., r/diabetes_t1, r/celiac) are the most numerous, yet their individual engagement varies widely. Lifestyle communities (such as r/intermittentfasting, r/loseit, and r/vegan) are fewer but generate substantially higher content volumes, as shown by their larger node sizes. Several high-traffic subreddits (e.g., r/nutrition, r/supplements) bridge healthcare and lifestyle domains, suggesting overlapping thematic content. Economic, social, and education nodes remain peripheral, reflecting sparse nutrition discourse in these areas on Reddit.

CONCLUSION

- ▶ Lifestyle communities represent only 24% of nutrition subreddits but generate 40% of total activity, nearly matching healthcare activity (47%).
- ▶ Activity weighting reverses healthcare's apparent dominance.
- ▶ Nutrition discourse spans all SDH domains beyond clinical settings.

FUTURE WORK / REFERENCES

- ▶ Subreddit labels may mask within-community variation.
- ▶ Reddit communities are not population-representative.
- ▶ Next: post-level analysis, misinformation detection, sentiment and temporal trends.

Conflict of interest: The authors declare no conflicts of interest.

Funding: The study received no funding.

Contact: stathiskaloudis@aegean.gr

[1] WHO Commission on SDH. (2008). Closing the gap in a generation. WHO.
 [2] Frank, J. W., et al. (2020). Int J Environ Res Public Health, 17(16), 5856. <https://doi.org/10.3390/ijerph17165856>
 [3] Schier, H. E., et al. (2024). Public Health Nutr, 27(1), e166. <https://doi.org/10.1017/S1368980024001459>
 [4] Kaloudis, E., et al. (2025). Nutrients, 17(5), 818. <https://doi.org/10.3390/nu17050818>