

# Computational Methods for Comparing Discourses

Zachary K. Stine<sup>1</sup>, James E. Deitrick<sup>2</sup>, & Nitin Agarwal<sup>1</sup>

zkstine@ualr.edu, deitrik@uca.edu, nxagarwal@ualr.edu

<sup>1</sup> Department of Information Science, University of Arkansas at Little Rock

<sup>2</sup> Department of Philosophy and Religion, University of Central Arkansas

## Overview

- Discourses around sociocultural identities comprise two levels of language usage:
  - Deliberately used language that ties the discourse to an identity (cultural lexicon).
  - Language that reflects the broader perspectives encoded within the discourse (cultural grammar).
- In comparative studies, deeper similarities and differences between discourses must be identified beneath the surface of superficial distinctions, often in a highly subjective way.
- Goal:** use computational methods to compare these deeper, structural aspects of discourses in a rigorous way that mitigates pitfalls associated with comparative studies.
- The discourses we analyze are the discussion histories from several English-language Reddit communities devoted to discussing religious and spiritual identities.

## Publications

ZK Stine, JE Deitrick, N Agarwal. 2020. Comparative religion, topic models, and conceptualization: Towards the characterization of structural relationships between online religious discourses *CHR 2020* [\[link\]](#)

ZK Stine, JE Deitrick, N Agarwal. 2021. Using Information Divergence to Differentiate Deep from Superficial Resemblances among Discourses. *C&C 2021* (part of *HCII 2021*). [\[link\]](#)

## Acknowledgements

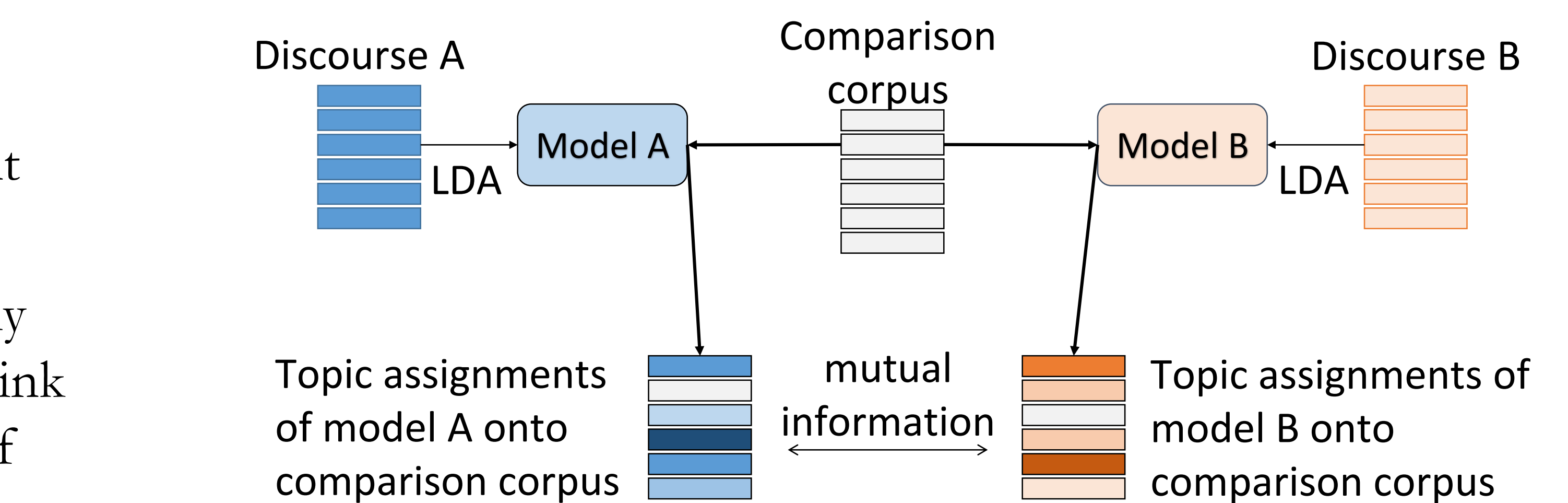
This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540, N00014-21-1-2121), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-17-S-0002, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094) to the third co-author, Nitin Agarwal. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations.



## Method 1. Comparing how discourses conceptualize text

- Case study comparing the discourses of *r/Buddhism* and *r/Christianity*.
- We train a probabilistic topic model using latent Dirichlet allocation (LDA) on each discourse in order to see how consistently the different models organize information.
- Each discourse-specific model can take an unseen document and apply its own distributional semantic features to the document. Useful to think of this as the model conceptualizing the unseen document in terms of its own semantic features.

<i>r/Buddhism</i> topic	<i>r/Christianity</i> topic	Pointwise Mutual Info
B16 Relationships	C15 Relationships	3.095
B24 Vegetarianism	C18 Abortion	2.797
B05 Repeated Text	C27 Repeated Text: Moderators	2.761
B05 Repeated Text	C10 Repeated Text: Verse Bot	2.743
B21 Intl. Politics, Conflict	C08 US Politics, Race	2.670
B12 Text Quotes	C23 Bible Verses	2.665
B25 Precepts	C25 Sex & Morality	2.617

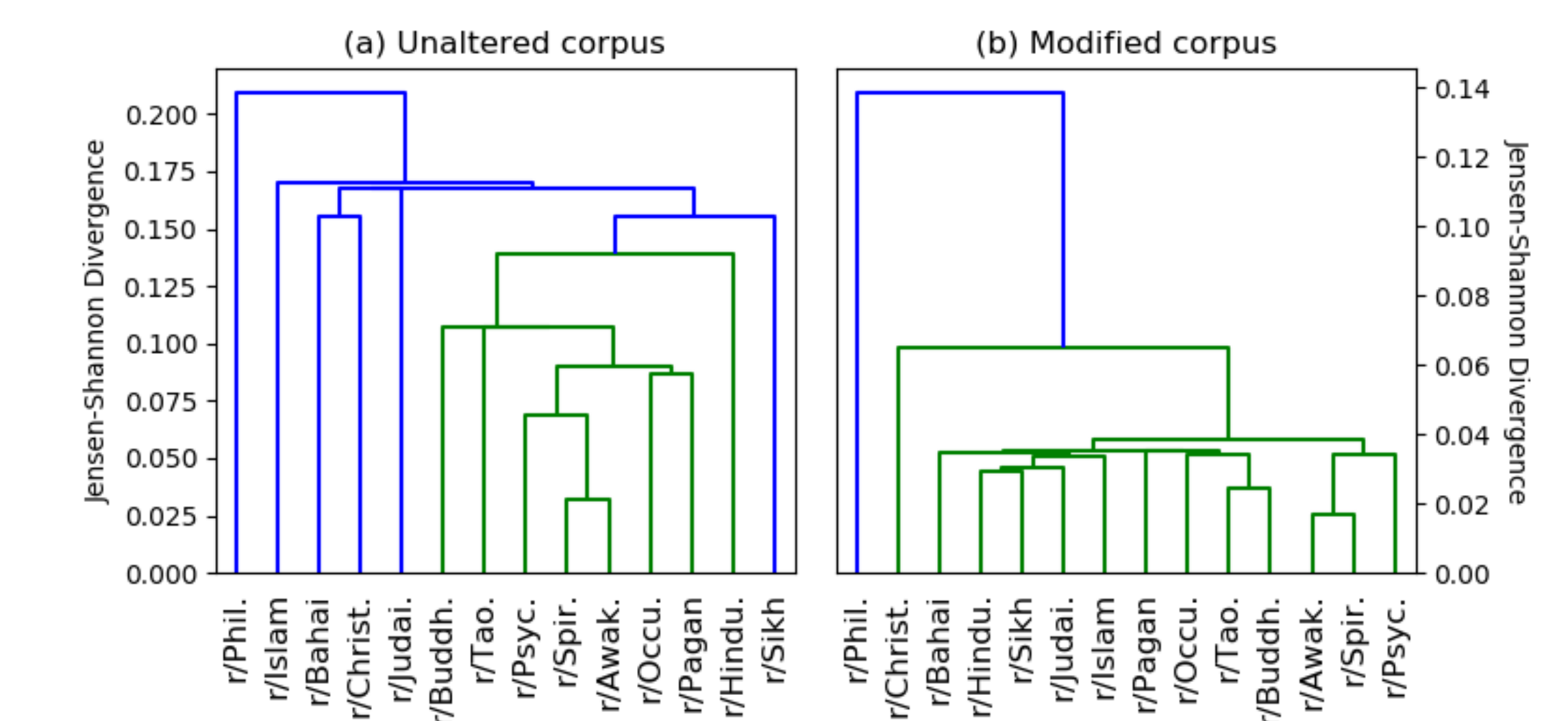


Suggests that, beneath the surface-level differences between the topics of vegetarianism and abortion, these two topics share structural similarities across discourses. While these two topics are marked by distinctive identity-specific concerns, the ways in which those concerns are discussed are quite similar.

ZK Stine, JE Deitrick, N Agarwal. 2020. Comparative religion, topic models, and conceptualization. *CHR 2020* [\[link\]](#)

## Method 2. Removing discourse-specific terms for comparisons

- Identify distinguishing terms that strongly signal one discourse relative to the others using word-level information divergence. Distributional semantic structure of discourses is resilient to such terms being removed.
- Compare discourses in two topic spaces, one learned from original text and one learned from text after removing highly distinguishing terms. Relationships between discourses may be quite different (*r/Sikh*) or similar (*r/Philosophy*) under the two modes of comparison.



### Examples of most distinguishing terms of discourses:

*r/Buddhism*: *buddhism buddha buddhist practice suffering meditation dhamma*

*r/Christianity*: *god jesus church bible christ christian sin christians christianity*

*r/Hinduism*: *hinduism hindu krishna shiva gita hindus brahman vedas vishnu*

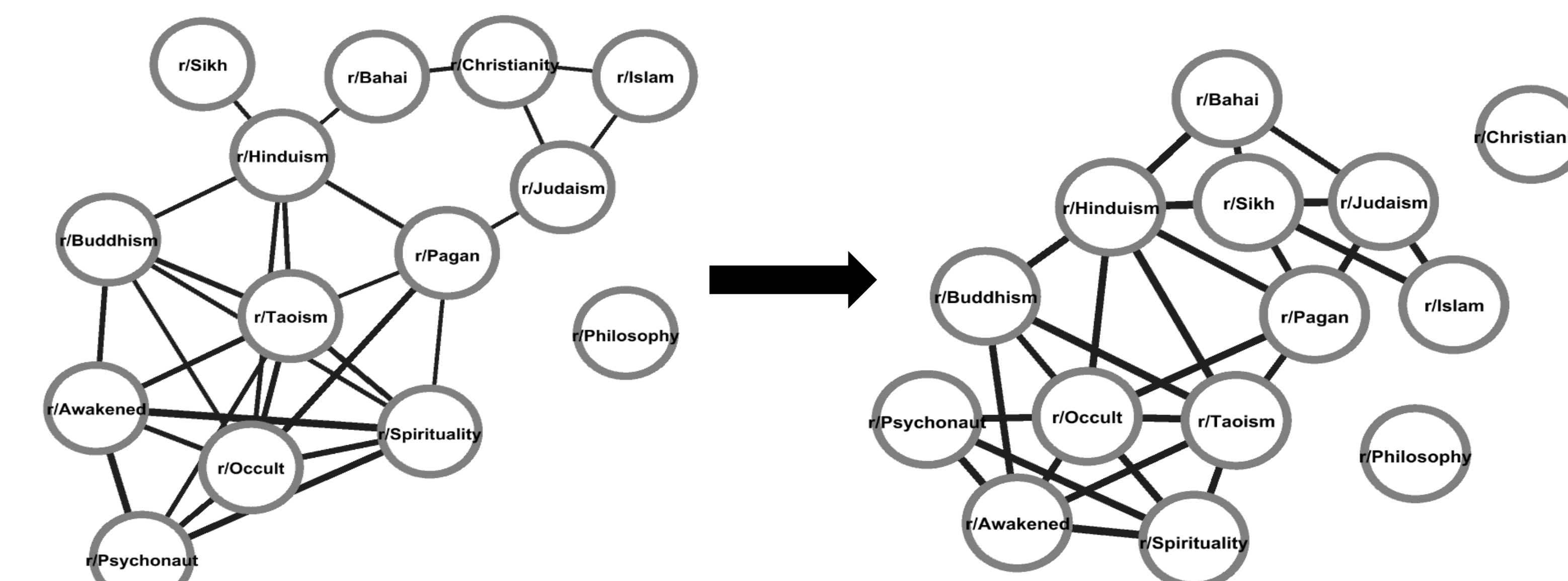
*r/Pagan*: *pagan paganism gods pagans celtic wicca deities norse goddess*

*r/Philosophy*: *philosophy moral argument philosophers unread philosophical morality*

*r/Psychonaut*: *trip psychedelics lsd dmt tripping psychedelic experience shrooms*

*r/Spirituality*: *life spiritual love feel energy spirituality yourself soul self feeling*

*r/Taoism*: *tao taoism taoist dao daoism daoist ching chinese zhuangzi wei*



ZK Stine, JE Deitrick, N Agarwal. 2021. Using Information Divergence to Differentiate Deep from Superficial Resemblances among Discourses. *C&C 2021*. [\[link\]](#)