Catherine King
cking2@cmu.edu

Peter Carragher
pcarragh@cmu.edu

Dr. Kathleen M. Carley
kathleen.carley@cs.cmu.edu

Citation Network Analysis of Misinformation Interventions

Social media misinformation is a serious societal problem

- Social media misinformation has been shown to
  - Undermine democracy [1]
  - Increase extremism [2]
  - Lower the uptake of public health measures [3]
- Research in this domain is across disciplines and can be challenging because of
  - Lack of data access from social media platforms
  - Ethical challenges associated with sharing data or running direct experiments
  - Costs associated with running large experiments or surveys

Existing review articles often miss the bigger picture

- Many focus on specific categories like media literacy [4] or content moderation [5]
- Others look more broadly on what interventions have been over or understudied [6]
- Most analyze effectiveness but exclude user acceptance or political feasibility

Research Questions

1. Which countermeasures are being under or over-studied in the literature?
2. What types of impacts is the literature studying? Impacts include effectiveness, user acceptance, and political feasibility
3. Are researchers working separately or collaborating across disciplines?

Method: Literature Review

142 Papers were Selected

1. From bibliography of a broad intervention literature review [6]
2. First two pages of Google Scholar results for specific keywords (shown below)
3. CitationGecko on selected papers - forward and backward citation mapping

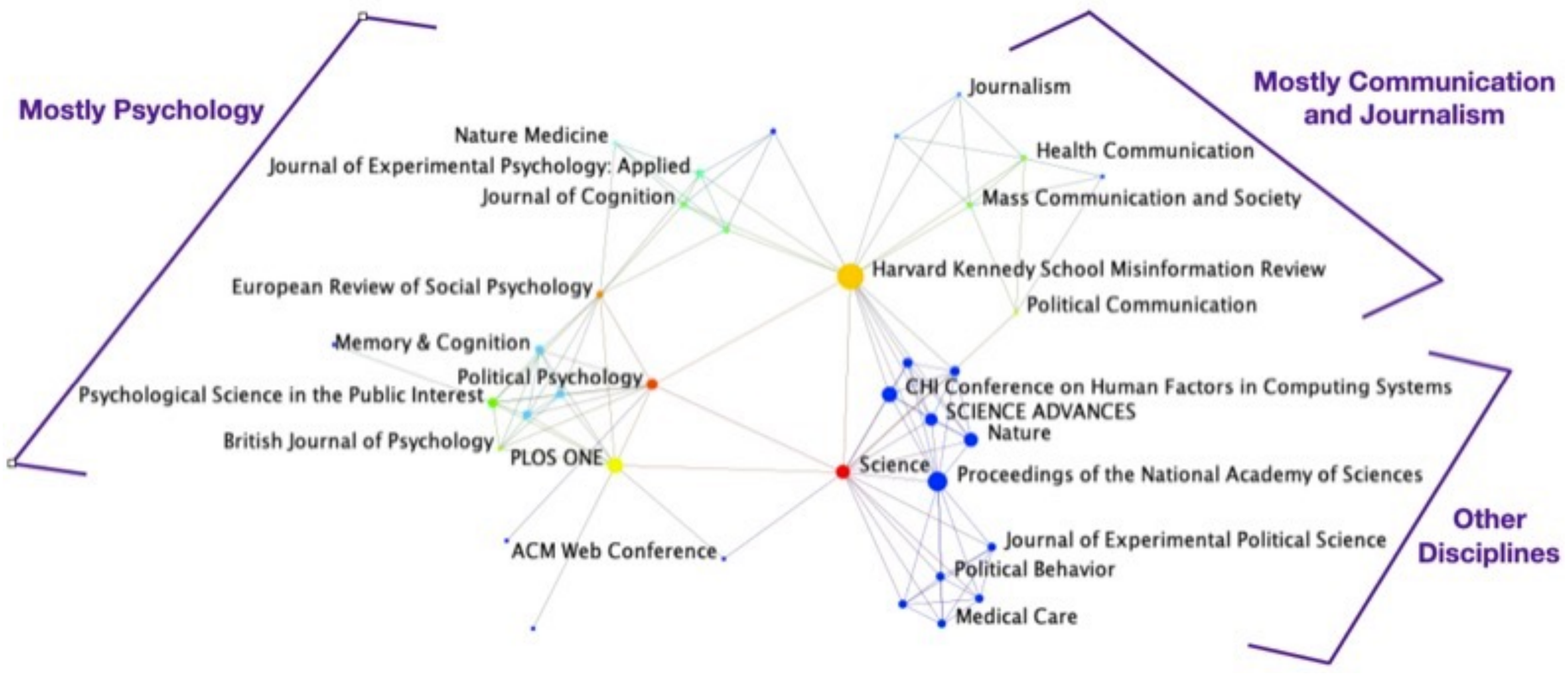
See paper for more details and inclusion criteria

Misinformation countermeasures, Countering misinformation, Countering fake news, Fact-checking, Deplatforming, Algorithmic downranking, Regulation social media, Government regulation social media, Content moderation, Social media advertising policy, Media literacy social media, Content labeling social media, Local news social media, Social media data sharing

Table 1: Google Scholar Keywords

Results: Co-Publication Venue Network

- The Co-Publication Venue network shows the co-authorship among venues
- Almost half of venues (41) are isolates, indicating how disjointed the literature is
  - Figure shows the main component (39 venues), sized by total degree centrality
  - Colored by betweenness (Red higher betweenness, Blue lower)


Mostly Psychology
Nature Medicine
Journal of Experimental Psychology: Applied
Journal of Cognition
European Review of Social Psychology
Memory & Cognition
Political Psychology
Psychological Science in the Public Interest
British Journal of Psychology
PLOS ONE
ACM Web Conference
Journalism
Mostly Communication and Journalism
Health Communication
Mass Communication and Society
Harvard Kennedy School Misinformation Review
Political Communication
CHI Conference on Human Factors in Computing Systems
SCIENCE ADVANCES
Nature
Science
Proceedings of the National Academy of Sciences
Journal of Experimental Political Science
Political Behavior
Medical Care
Other Disciplines

Countermeasures Categories

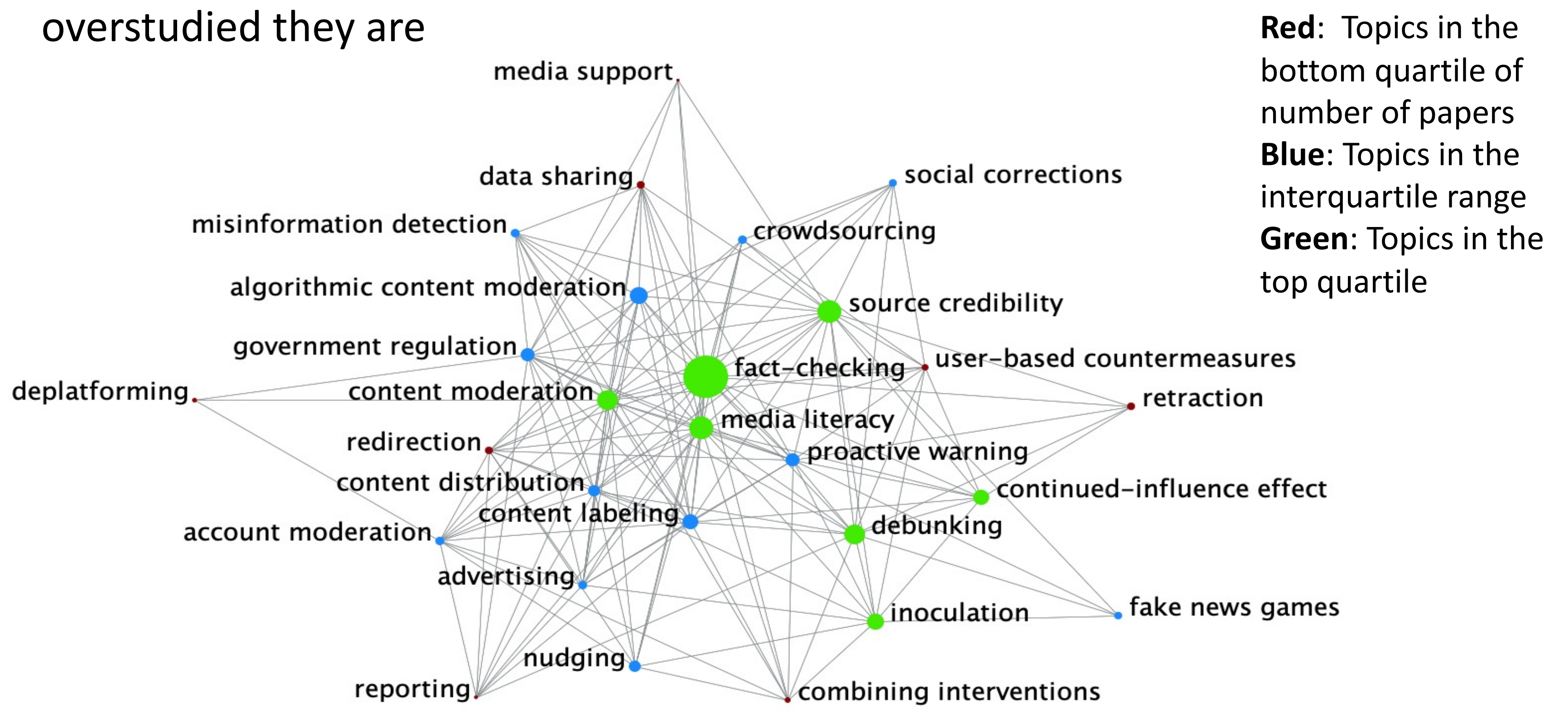
- Drawing from the literature [6,7,8,9], developed 8 categories of interventions
- Derived a comprehensive list of 27 unique labels from these categories

Category	Labels
Content Distribution	Distribution, redirection, nudging
Content/Account Moderation	Content moderation, fact-checking, debunking, misinformation detection, algorithmic content moderation, continued-influence effect, account moderation, deplatforming
Content Labeling	Labeling, crowdsourcing, source credibility
Media Support	Investing in/promoting local news
Media Literacy and Awareness	Media literacy, fake news games, inoculation, proactive warning, data sharing
Advertising	Advertising policy
User-based Countermeasures	General user-based countermeasures, reporting, social corrections, retraction
Other	Government regulation, combining interventions

Other labels: *Review Article, Meta-Analysis, Studies Acceptance, Studies Effectiveness*

Results: Co-Topics Network

- The Co-Topics network shows which topics are often studied together
- Nodes are sized by Total Degree Centrality and colored by how relatively under or overstudied they are


Red: Topics in the bottom quartile of number of papers
Blue: Topics in the interquartile range
Green: Topics in the top quartile

Discussion and Future Work

1. **User acceptance is greatly understudied relative to effectiveness –** 85 papers (60%) analyzed effectiveness, 11 acceptance (8%), and just 2 on both
2. **There are several under and over-studied interventions –** many high impact and frequently interventions are understudied (redirection, user-based ,media support, data sharing, and combining interventions)
3. **Few cross-disciplinary journals –** Harvard Misinformation Review bridges the gap
4. **Lack of consensus –** during the literature review, found several interventions lack consensus on effectiveness (fake news games, debunking, nudging)

Next Steps: network analysis on authors/universities, in-depth analysis of finding #4

References

[1] Tucker et al. 2017. "From Liberation to Turmoil: Social Media and Democracy." *Journal of Democracy*
[2] Warner and Neville-Shepard 2014. "Echoes of a Conspiracy: Birthers, Truthers, and the Cultivation of Extremism". *Communication Quarterly.*
[3] Olesky et al. 2021. "Content matters. Different predictors and social consequences of general and government-related conspiracy theories on COVID-19." *Personality and Individual Differences.*
[4] Jeong et al. 2012. "Media Literacy Interventions: A Meta-Analytic Review." *The Journal of Communication.*
[5] Jiang et al. 2023. "A Trade-off-centered Framework of Content Moderation." *ACM Transactions on Computer-Human Interaction.*
[6] Courchesne et al. 2021. "Review of social science research on the impact of countermeasures against influence operations." *HKS Misinfo Review.*
[7] Gwiaździński et al. 2023. "Psychological interventions countering misinformation in social media: A scoping review." *Frontiers in Psychiatry.*
[8] Helmus and Kepe 2021. "A Compendium of Recommendations for Countering Russian and Other State-Sponsored Propaganda." Tech Report, *Rand.*
[9] Yadav 2021. "Platform Interventions: How Social Media Counters Influence Operations." Tech Report, *Carnegie Endowment for Int'l Peace.*

This work was supported in part by the Office of Naval Research (ONR) Award N00014-21-12749, the Knight Foundation, the Center for Informed Democracy and Social-cybersecurity (IDeaS), and the Center for Computational Analysis of Social and Organization Systems (CASOS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of ONR or the U.S. government.