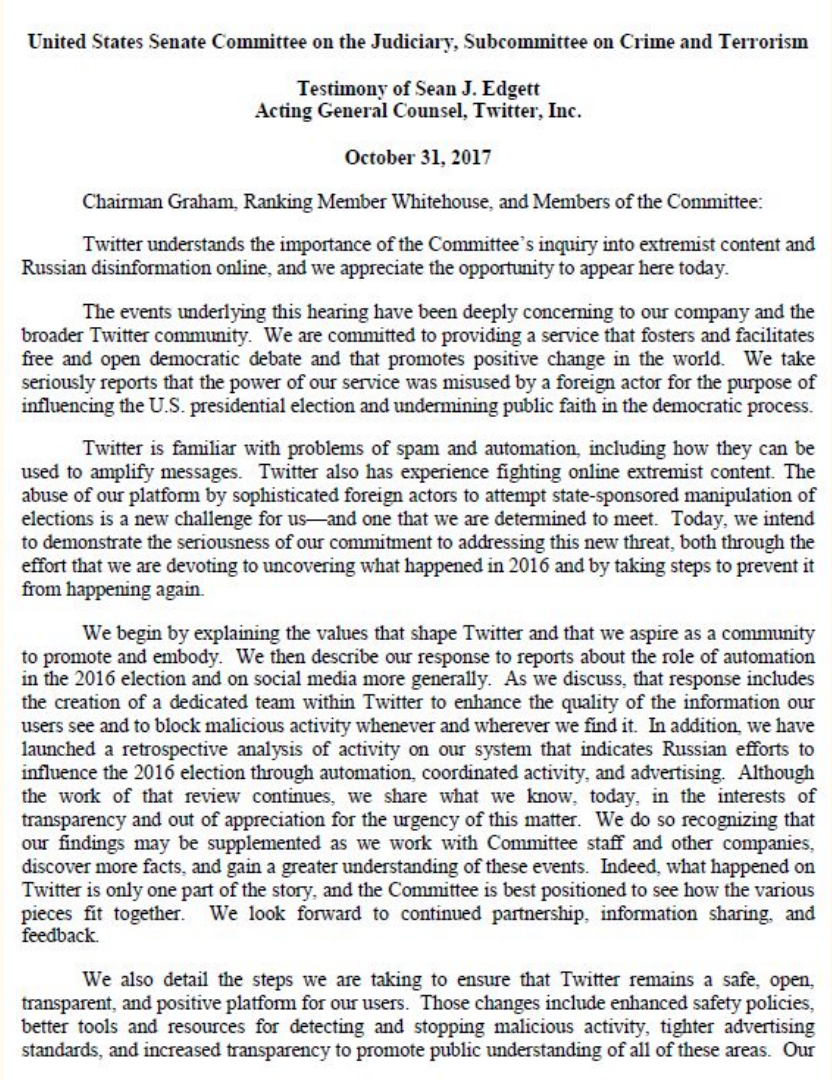
Exploratory Analysis of Communities of Interfering Agents

Nicholas Vogt
CIS 579: Artificial Intelligence | University of Michigan - Dearborn

What are Interfering Agents?



United States Senate Committee on the Judiciary, Subcommittee on Crime and Terrorism

Testimony of Sean J. Edgett
Acting General Counsel, Twitter, Inc.

October 31, 2017

Chairman Graham, Ranking Member Whitehouse, and Members of the Committee:

Twitter understands the importance of the Committee's inquiry into extremist content and Russian disinformation online, and we appreciate the opportunity to appear here today.

The events underlying this hearing have been deeply concerning to our company and the broader Twitter community. We are committed to providing a service that fosters and facilitates free and open democratic debate and that promotes positive change in the world. We take seriously reports that the power of our service was misused by a foreign actor for the purpose of influencing the U.S. presidential election and undermining public faith in the democratic process.

Twitter is familiar with problems of spam and automation, including how they can be used to amplify messages. Twitter also has experience fighting online extremist content. The abuse of our platform by sophisticated foreign actors to attempt state-sponsored manipulation of elections is a new challenge for us—and one that we are determined to meet. Today, we intend to demonstrate the seriousness of our commitment to addressing this new threat, both through the effort that we are devoting to uncovering what happened in 2016 and by taking steps to prevent it from happening again.

We begin by explaining the values that shape Twitter and that we aspire as a community to promote and embody. We then describe our response to reports about the role of automation in the 2016 election and on social media more generally. As we discuss, that response includes the creation of a dedicated team within Twitter to enhance the quality of the information our users see and to block malicious activity whenever and wherever we find it. In addition, we have launched a retrospective analysis of activity on our system that indicates Russian efforts to influence the 2016 election through automation, coordinated activity, and advertising. Although the work of that review continues, we share what we know, today, in the interests of transparency and out of appreciation for the urgency of this matter. We do so recognizing that our findings may be supplemented as we work with Committee staff and other companies, discover more facts, and gain a greater understanding of these events. Indeed, what happened on Twitter is only one part of the story, and the Committee is best positioned to see how the various pieces fit together. We look forward to continued partnership, information sharing, and feedback.

We also detail the steps we are taking to ensure that Twitter remains a safe, open, transparent, and positive platform for our users. Those changes include enhanced safety policies, better tools and resources for detecting and stopping malicious activity, tighter advertising standards, and increased transparency to promote public understanding of all of these areas. Our

Twitter gave Congress 3,814 account names it linked to Kremlin-linked interference in the 2016 Election (Edgett Testimony, 2017)

"At the request of NBC News, three sources familiar with Twitter's data systems cross-referenced the list of names released by Congress, excluding any account that Twitter later restored, to create a partial database of tweets that could be recovered from the suspended accounts."
- NBC News

Can we infer behavior from the community structure of interfering agents?

Expectations

Do Interfering Agents primarily interact with each other?

Are there online communities of interfering agents to give the impression of a unified group?

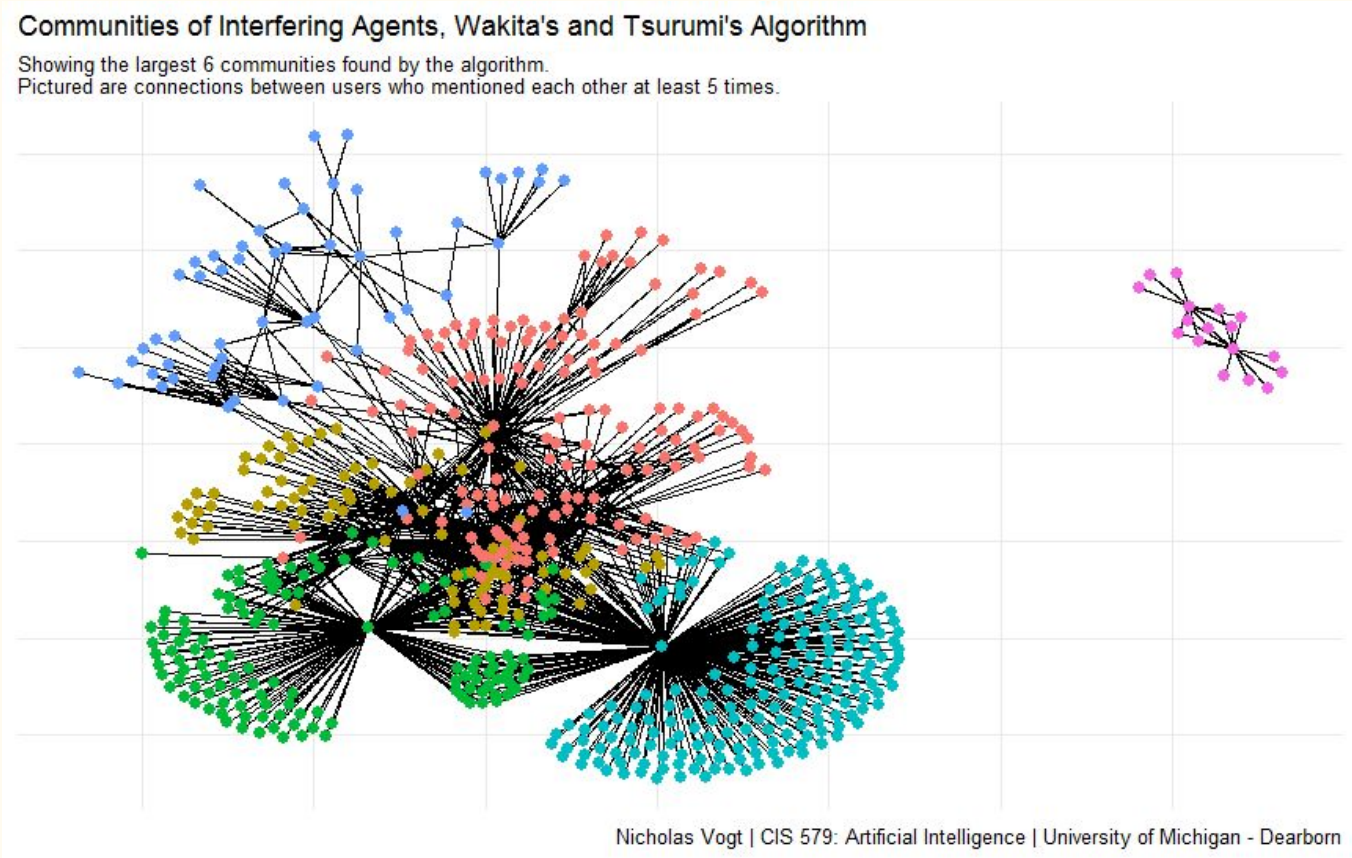
- Reddit's /r/The_Donald
- 4chan's /pol/

...or do they operate independently?

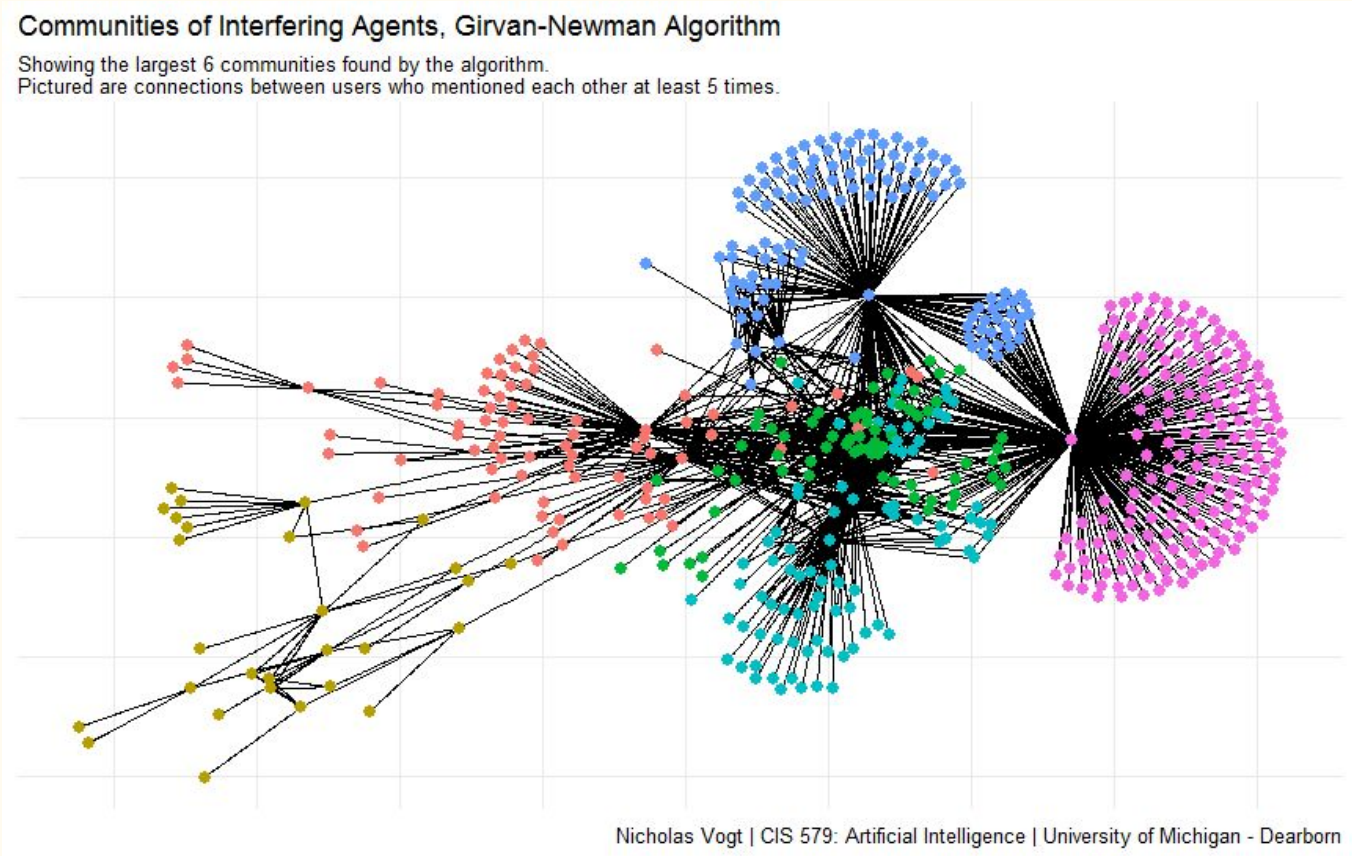
Do they infiltrate existing communities or follow individuals separately?

It's easy to spot a crowd. Individuals may be harder to identify.

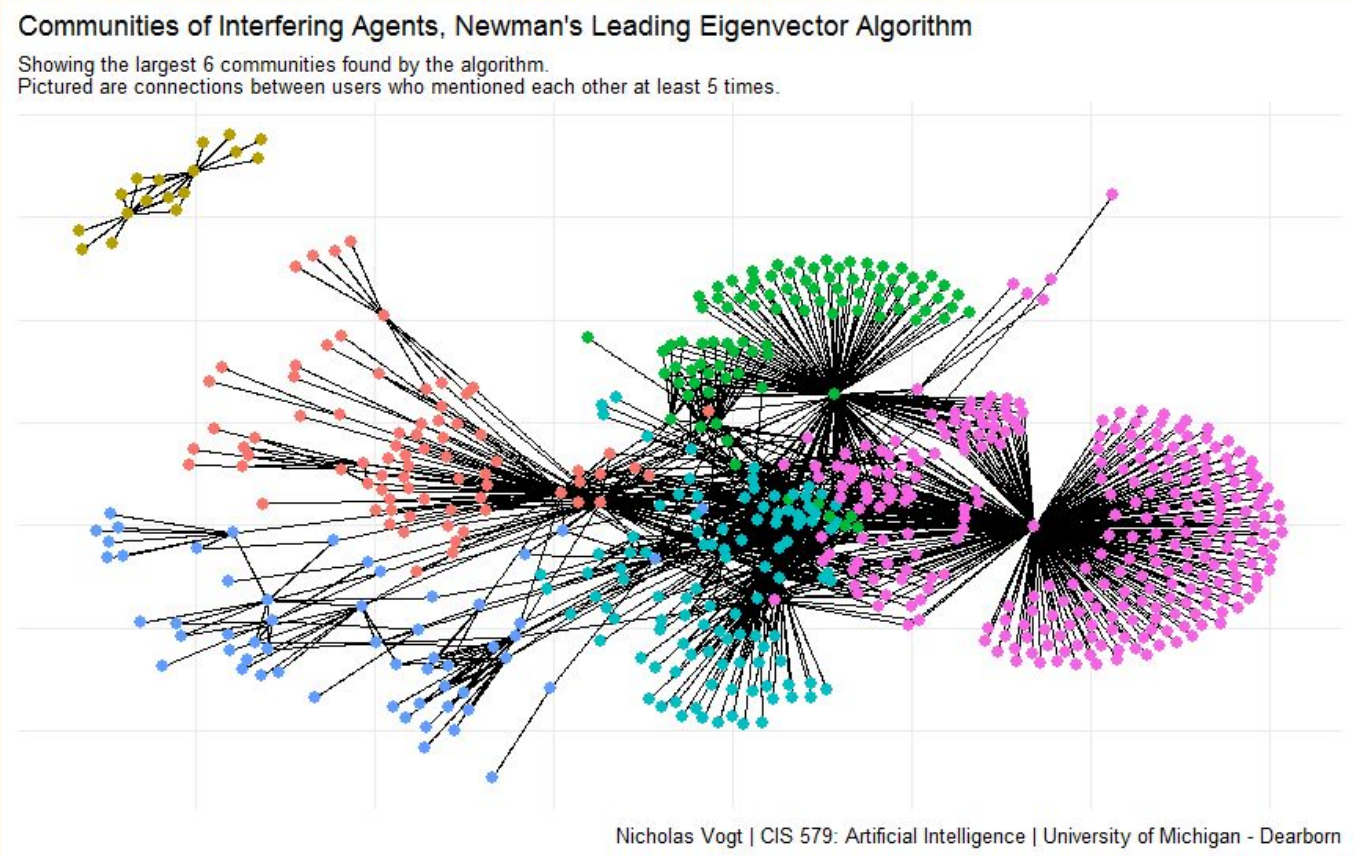
Three Community Detection Methods



Three Community Detection Methods



Three Community Detection Methods



Expectations vs Results

~~**Do Interfering Agents primarily interact with each other?**~~

~~**...or do they operate independently?**~~

They interact with common users and rarely interact directly.

This behavior is remarkably ordinary. Twitter has several high-profile users who garner the attention of disjoint communities.

Do what do these communities talk about?

Expectations

Do Interfering Agents discuss the same topics with their common interactions?

Do interfering agents have an agenda of talking points that they spread to all common interactions?

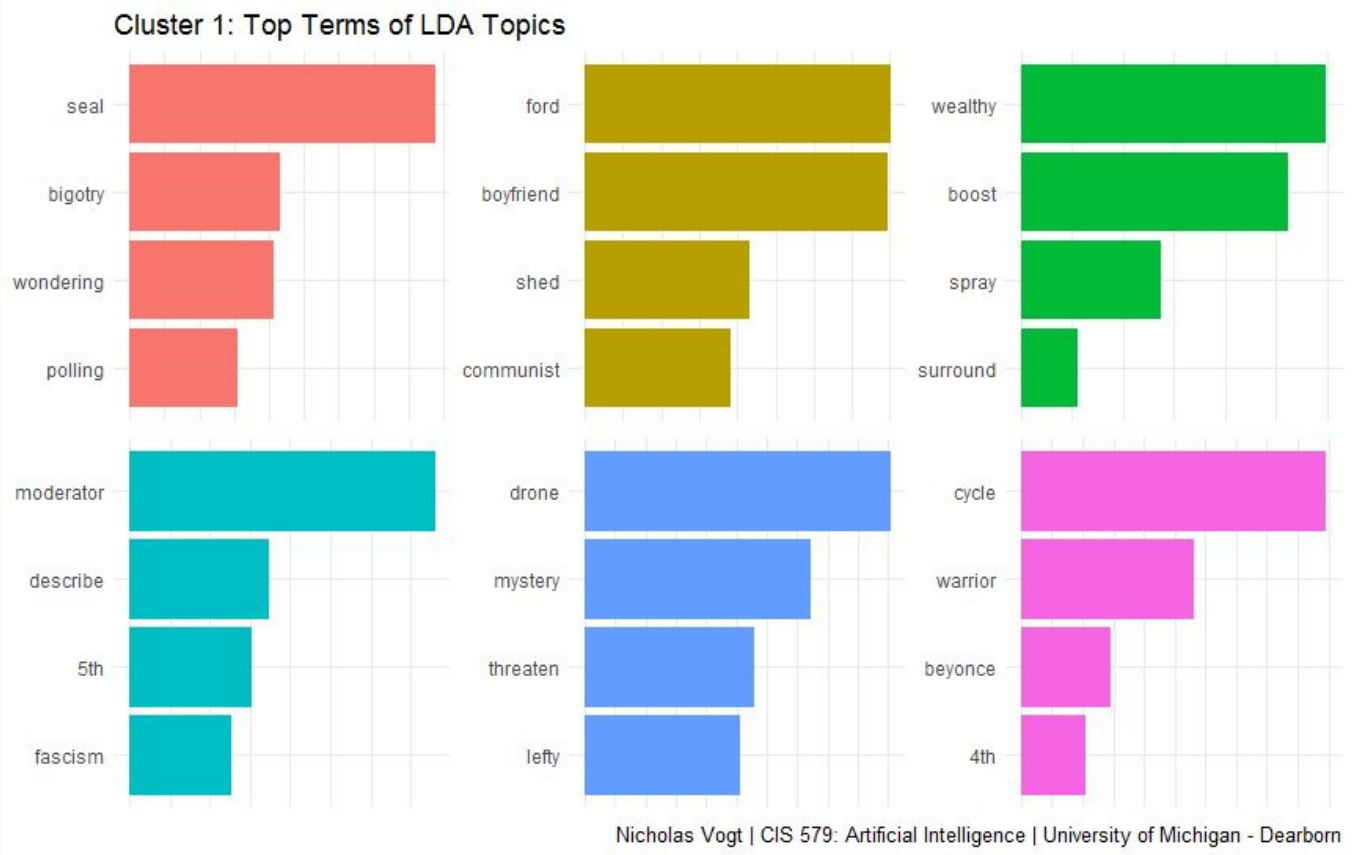
If so, that implies that they cooperate out-of-view.

...or do they discuss different topics?

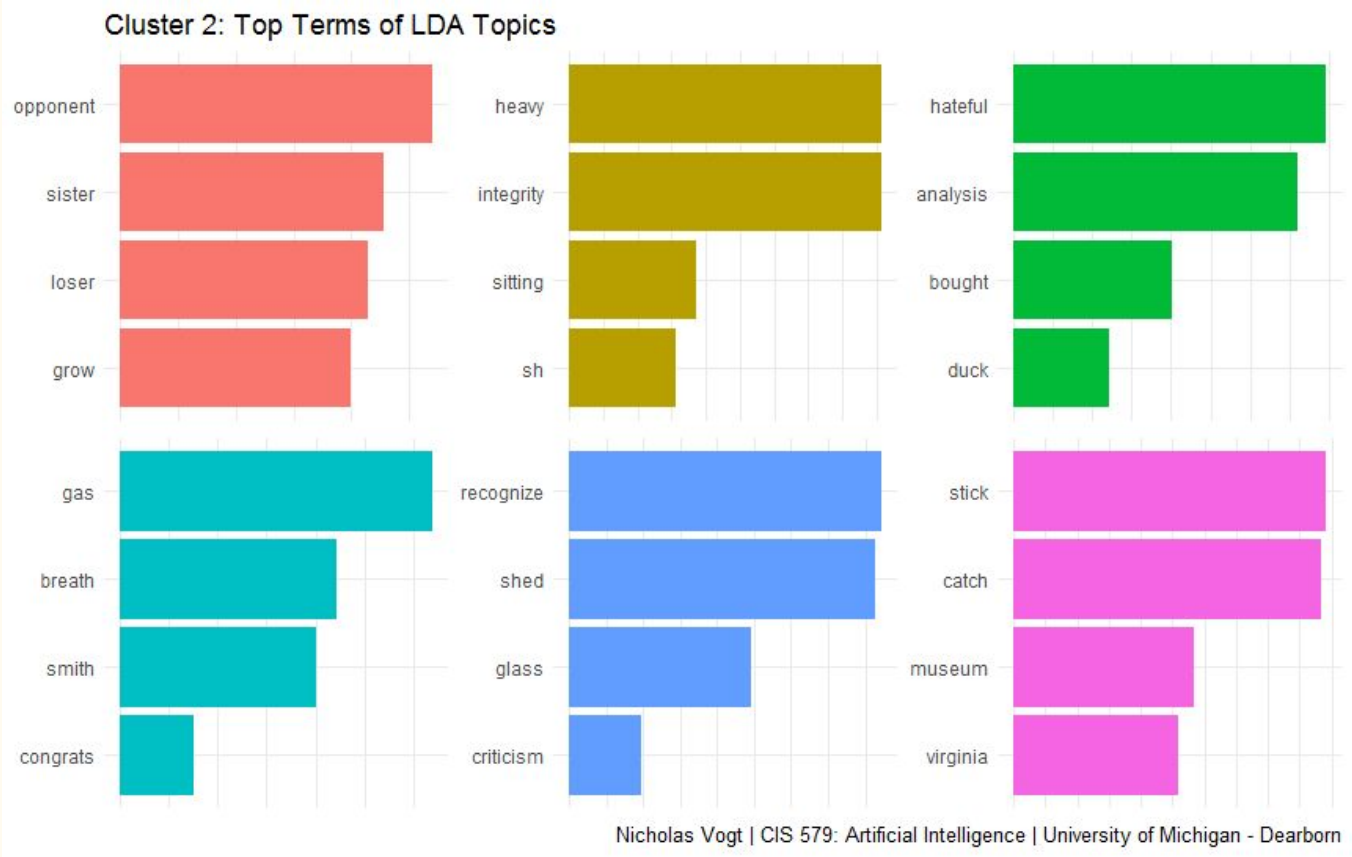
Do interfering agents talk about topics important to their common interactions?

Nancy in Texas may care about illegal immigration while Bill in Pennsylvania cares about job creation.

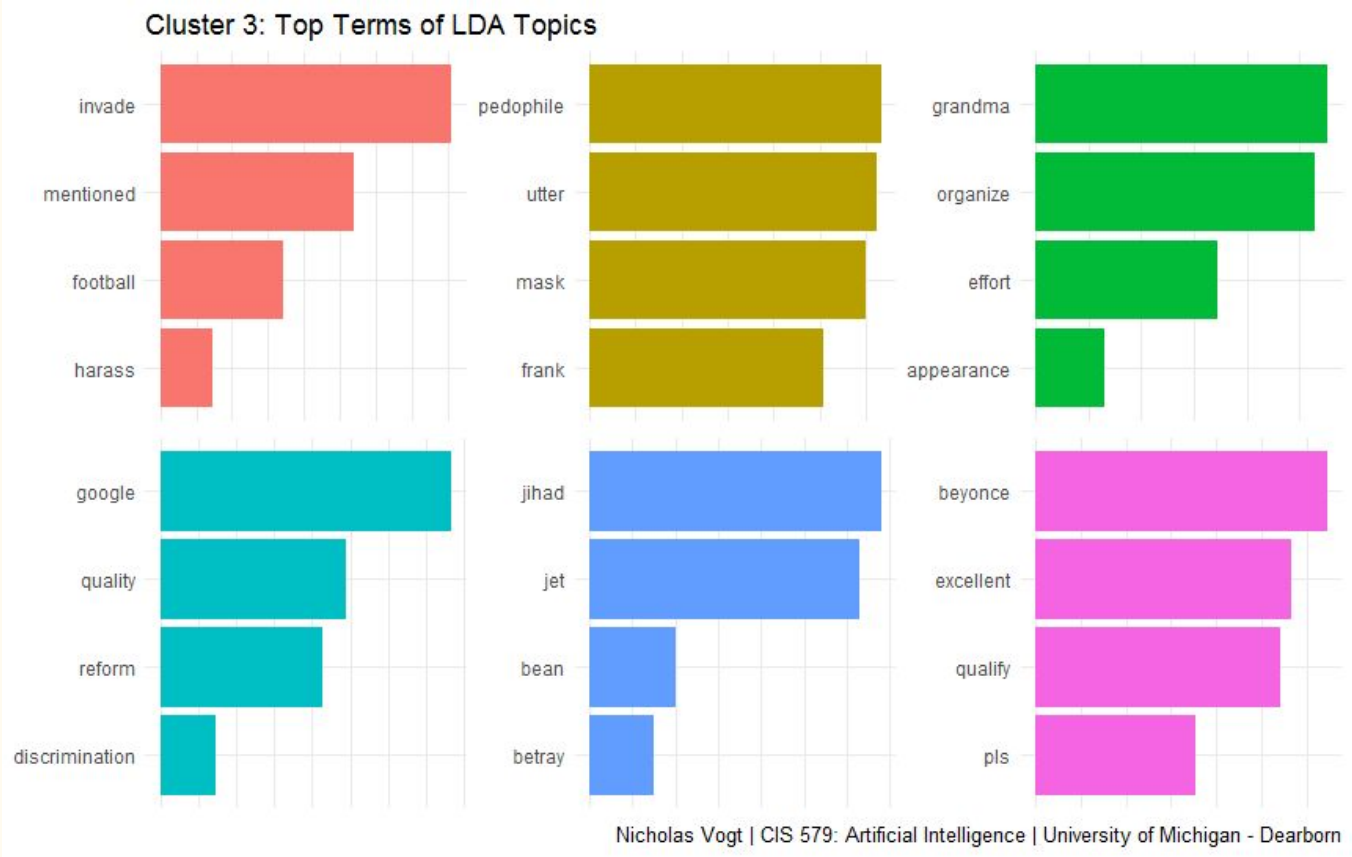
LDA Results on the Six Largest Clusters



LDA Results on the Six Largest Clusters



LDA Results on the Six Largest Clusters



Expectations vs Results

~~**Do Interfering Agents primarily interact with each other?**~~ ~~**...or do they operate independently?**~~

Results are <u>inconclusive</u>. But strange keywords provoke a deeper analysis.

LDA suffered from poor preprocessing of the data.

- misspellings (e.g. RALLY from RALY),
- abbreviations (e.g. COALITION from COAL),
- concatenations (e.g. HILLARY THE from HILLARYTHE), and
- acronyms (e.g. BLACKLIVESMATTER from BLM).

Conclusion

Community detection yielded unexpected community structure. Interfering agents

1. target common users, and
2. rarely interact directly with each other.

LDA suffered from

1. poor data cleansing,
2. a small corpus, and
3. relatively small documents.