Promise and Pitfalls of Using Non-State Actor Media for Large-Scale Event Datasets

Emily K. Gade, Shahryar Minhas, Bree Bang Jensen, Kai Ping Leung, Michael Gabbay, Mohammed Hafez, Emily Thomas, Kayla Salehian, Ava Sharifi, Tara Farkouh, Ramin Farrokhi, Arica Schuett, Willa Jeffers, Bernadette Bresee, Ben Lefkowitz

April 3, 2024

I was supposed to present

- "It's Complicated": Explaining Frenemies Relaionships in Civil War"

I was supposed to present

- "It's Complicated": Explaining Frenemies Relaionships in Civil War"
- but who would've thought ... it's complicated!

frenemies tldr

- In some conflicts such as Syria we observe fluid conflict and cooperation dynamics between actors

frenemies tldr

- In some conflicts such as Syria we observe fluid conflict and cooperation dynamics between actors
- Much of the literature, however, does not model dynamics between rebel actors this way (my and Emily's work included)

frenemies tldr

- In some conflicts such as Syria we observe fluid conflict and cooperation dynamics between actors
- Much of the literature, however, does not model dynamics between rebel actors this way (my and Emily's work included)
- We sought to model the joint evolution of cooperation and conflict, a few things to make life interesting though:
  - Multivariate modeling (the joint modeling of cooperation and conflict) is not fun

frenemies tldr

- In some conflicts such as Syria we observe fluid conflict and cooperation dynamics between actors
- Much of the literature, however, does not model dynamics between rebel actors this way (my and Emily's work included)
- We sought to model the joint evolution of cooperation and conflict, a few things to make life interesting though:
  - Multivariate modeling (the joint modeling of cooperation and conflict) is not fun
  - Especially when we want to model the temporal dynamics of these relationships (i.e., why is frenemy behavior sometimes sticky and not)

frenemies tldr

- In some conflicts such as Syria we observe fluid conflict and cooperation dynamics between actors
- Much of the literature, however, does not model dynamics between rebel actors this way (my and Emily's work included)
- We sought to model the joint evolution of cooperation and conflict, a few things to make life interesting though:
  - Multivariate modeling (the joint modeling of cooperation and conflict) is not fun
  - Especially when we want to model the temporal dynamics of these relationships (i.e., why is frenemy behavior sometimes sticky and not)
  - And ... I could keep going, happy to discuss more (yes, of course we started simple and it got complicated when our model's just did not fit)

frenemies tldr

- In some conflicts such as Syria we observe fluid conflict and cooperation dynamics between actors
- Much of the literature, however, does not model dynamics between rebel actors this way (my and Emily's work included)
- We sought to model the joint evolution of cooperation and conflict, a few things to make life interesting though:
  - Multivariate modeling (the joint modeling of cooperation and conflict) is not fun
  - Especially when we want to model the temporal dynamics of these relationships (i.e., why is frenemy behavior sometimes sticky and not)
  - And ... I could keep going, happy to discuss more (yes, of course we started simple and it got complicated when our model's just did not fit)
- But bringing us back to today ... where would we have gotten the data for this?!

SAGAA: Syrian Armed Groups, Alliances, and Actions

- Ostovar & Mccants (Forthcoming): "There are hundreds if not thousands of small groups that have been active in the Syrian conflict, nearly countless of which have split or changed alliances, political outlook, or leadership over time."

SAGAA: Syrian Armed Groups, Alliances, and Actions

- Ostovar & Mccants (Forthcoming): "There are hundreds if not thousands of small groups that have been active in the Syrian conflict, nearly countless of which have split or changed alliances, political outlook, or leadership over time."
- News outlets have little incentive and perhaps ability at times to track these micro-dynamics.

SAGAA: Syrian Armed Groups, Alliances, and Actions

- Ostovar & Mccants (Forthcoming): "There are hundreds if not thousands of small groups that have been active in the Syrian conflict, nearly countless of which have split or changed alliances, political outlook, or leadership over time."
- News outlets have little incentive and perhaps ability at times to track these micro-dynamics.
- Our SAGAA dataset tries to take on this task by using rebel social media data from 28,000 posts depicting on-the-ground conflict engagmenets between 2012 and 2018

Challenges in Conflict Data Collection

- Traditional observational datasets face challenges in conflict zones: reliability, over and under-reporting, and bias are significant issues with news-based data Davenport 2009, Earl et al. 2004, Eck 2012, Weidmann 2016.

Challenges in Conflict Data Collection

- Traditional observational datasets face challenges in conflict zones: reliability, over and under-reporting, and bias are significant issues with news-based data Davenport 2009, Earl et al. 2004, Eck 2012, Weidmann 2016.
- Specifically, for civil conflicts, news media tend to emphasize:
  - larger, more letal attacks Behlendorf, Belur and Kumar 2016

Challenges in Conflict Data Collection

- Traditional observational datasets face challenges in conflict zones: reliability, over and under-reporting, and bias are significant issues with news-based data Davenport 2009, Earl et al. 2004, Eck 2012, Weidmann 2016.
- Specifically, for civil conflicts, news media tend to emphasize:
  - larger, more letal attacks Behlendorf, Belur and Kumar 2016
  - do not distinguish between subgroups Warner and Lizzo 2021

Challenges in Conflict Data Collection

- Traditional observational datasets face challenges in conflict zones: reliability, over and under-reporting, and bias are significant issues with news-based data Davenport 2009, Earl et al. 2004, Eck 2012, Weidmann 2016.
- Specifically, for civil conflicts, news media tend to emphasize:
  - larger, more letal attacks Behlendorf, Belur and Kumar 2016
  - do not distinguish between subgroups Warner and Lizzo 2021
  - and can be influenced by local political actors and context Baum and Zhukov 2015.

Potential of Rebel-Produced Media

- Rebels and other non-state actors have always used propaganda, but social media has changed the reach Aday et al. 2010, Jones and Mattiacci 2019, Loyle and Bestvater 2019, Zeitzoff 2017.

Potential of Rebel-Produced Media

- Rebels and other non-state actors have always used propaganda, but social media has changed the reach Aday et al. 2010, Jones and Mattiacci 2019, Loyle and Bestvater 2019, Zeitzoff 2017.
- Many have argued non-state actor-produced media can provide an alternative source of information about conflict Aday et al. 2010, Jones and Mattiacci 2019, Loyle and Bestvater 2019, Zeitzoff 2017.

Concerns with Rebel-Produced Media

- We all know that terrorist groups sometimes choose not to claim responsibility for attacks they have undertaken due to a number of strategic considerations Hoffman 2010, Kearns et al. 2014, Abrahms & Conrad 2017.

Concerns with Rebel-Produced Media

- We all know that terrorist groups sometimes choose not to claim responsibility for attacks they have undertaken due to a number of strategic considerations Hoffman 2010, Kearns et al. 2014, Abrahms & Conrad 2017.
- But insurgent groups might be using social media differently, specifically, rebel groups may have different incentives for producing accurate information about their behavior

Concerns with Rebel-Produced Media

- We all know that terrorist groups sometimes choose not to claim responsibility for attacks they have undertaken due to a number of strategic considerations Hoffman 2010, Kearns et al. 2014, Abrahms & Conrad 2017.
- But insurgent groups might be using social media differently, specifically, rebel groups may have different incentives for producing accurate information about their behavior
- In Iraq, Gabbay and Thirkill-Mackelprang (2010) find that posts were largely directed at rank-and-file soldiers, who had a great deal of specific local knowledge that enabled them to discern about veracity of claims

Concerns with Rebel-Produced Media

- We all know that terrorist groups sometimes choose not to claim responsibility for attacks they have undertaken due to a number of strategic considerations Hoffman 2010, Kearns et al. 2014, Abrahms & Conrad 2017.
- But insurgent groups might be using social media differently, specifically, rebel groups may have different incentives for producing accurate information about their behavior
- In Iraq, Gabbay and Thirkill-Mackelprang (2010) find that posts were largely directed at rank-and-file soldiers, who had a great deal of specific local knowledge that enabled them to discern about veracity of claims
- Whiteside 2020 notes that this was true of the Islamic State's media operations as well
- Greene and Lucas (2020) find that rebel media can accurately articulate relationships between Hezbollah and other armed actors

How we use Rebel-Produced Media in SAGAA

- **Where we expect rebels to lie**:
  - Civilian targeting
  - Casualty numbers
  - Who won or who started the fight

How we use Rebel-Produced Media in SAGAA

- **Where we expect rebels to lie**:
  - Civilian targeting
  - Casualty numbers
  - Who won or who started the fight
- **Where we think rebel media is accurate**:
  - Where they were
  - Who they were with
  - When they were there

What do Rebel-Produced Media Look Like?

Fath al-Sham Front

Yesterday, we killed dozens of regime soldiers with our rockets and guns alongside Jaysh al-Fath in Aleppo. The local battalions of the area helped us in our quest.

What do Rebel-Produced Media Look Like?

Fath al-Sham Front

Yesterday, we killed dozens of regime soldiers with our rockets and guns alongside Jaysh al-Fath in Aleppo. The local battalions of the area helped us in our quest.

Ahrar al-Sham Brigades

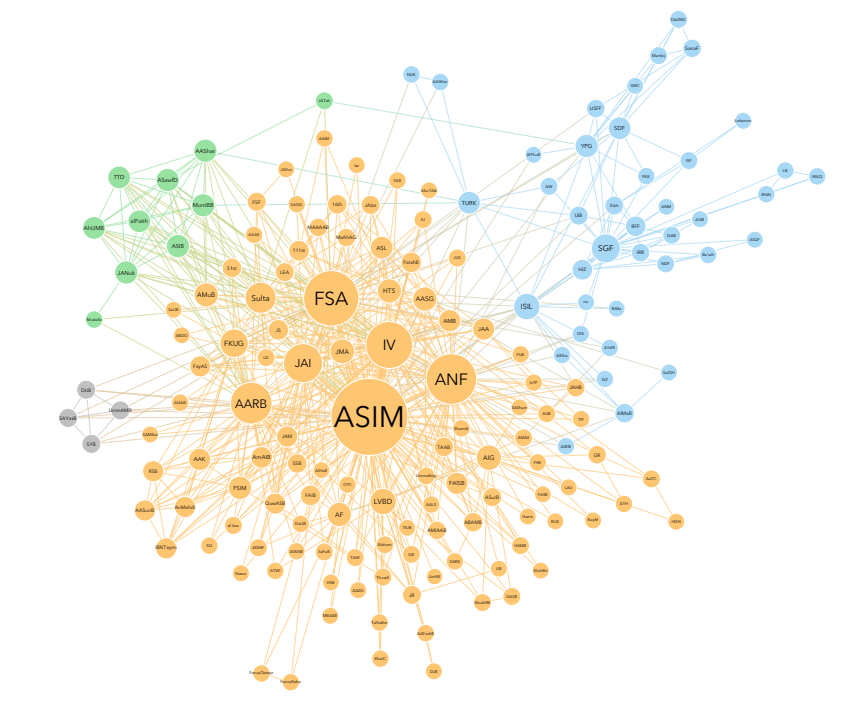
The Abu-Ubaydah Bin-al-Jarrah Battalion of the Al-Iman Brigade, of the Ahrar al-Sham Brigade, claimed to have "destroyed the pick-up vehicles of the 'traitorous security gangs' with a planted explosive device on the Old Industrial Road, killing three Shabbihas and wounding two others on board."

SAGAA Dataset Design

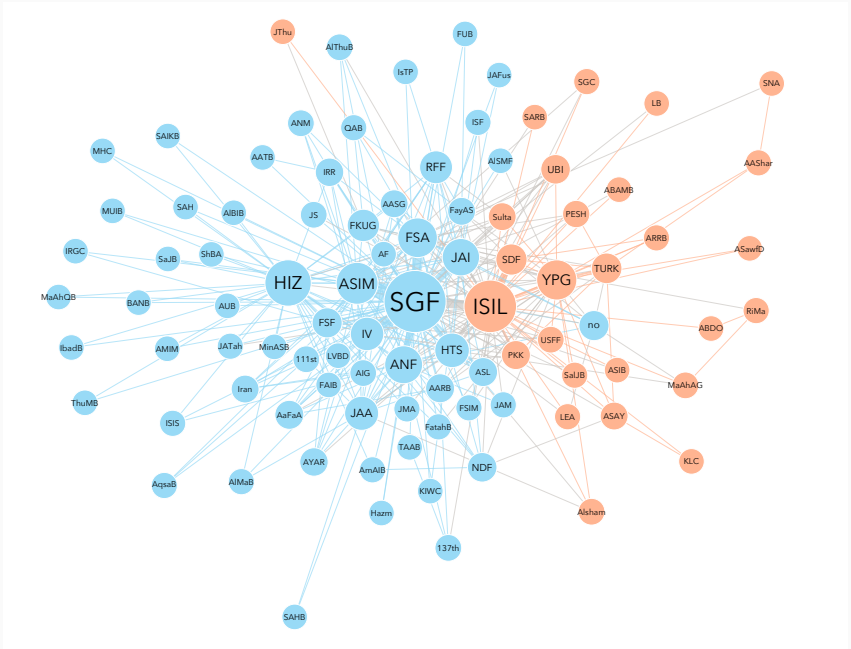
Dataset	Source	Details
Event Dataset	Translations of rebel social media posts	28,000 hand-coded social media posts, 7,000 of which were excluded because they were not claims of attack. Each of the remaining 21,000 claims of attacks were coded for the variables described in §C.
Collaborative and Conflictual Network Dataset	In-text references about relationships among groups drawn from claims of attacks	Direct textual references about collaborative and conflictual relationships among groups.
Hierarchical Dataset	In-text references about relationships among groups drawn from claims of attacks	This focused on hierarchical relationships including parent-subgroup ties, alliances, and joint operations rooms.
Group-level Covariate Dataset	Think tank reports, NGO reports, rebel's own media, news articles, google searches	We selected the 67 most prominent groups in our collaborative and conflictual networks (this cut-off was based on group degree in each network) and modestly updated that list as groups splintered and merged. Following detailed coding ontology presented in §C, we collected group-year level data for the most prominent groups in the network.

What does SAGAA Give Us?

Collaboration Network



Conflict Network



Comparison to UCDP: Actors

- We compare the SAGAA dataset to the current gold standard, UCDP Syria dataset, to evaluate the veracity of SAGAA as well as consider what rebel media might be able to tell us that conventional news cannot

Comparison to UCDP: Actors

- We compare the SAGAA dataset to the current gold standard, UCDP Syria dataset, to evaluate the veracity of SAGAA as well as consider what rebel media might be able to tell us that conventional news cannot
- UCDP's data is collected via string searches through the Dow Jones news aggregator to identify all articles which involve death or injury for each particular conflict

Comparison to UCDP: Actors

- We compare the SAGAA dataset to the current gold standard, UCDP Syria dataset, to evaluate the veracity of SAGAA as well as consider what rebel media might be able to tell us that conventional news cannot
- UCDP's data is collected via string searches through the Dow Jones news aggregator to identify all articles which involve death or injury for each particular conflict
- To compare the data sources, we first mapped the group names from UCDP Syria to SAGAA.
  - The UCDP data contains 76 unique groups for our time period (2012-2018).
  - SAGAA contains more than 400 for the same period.
  - UCDP dataset contains less than 16% of the groups listed in the SAGAA dataset.

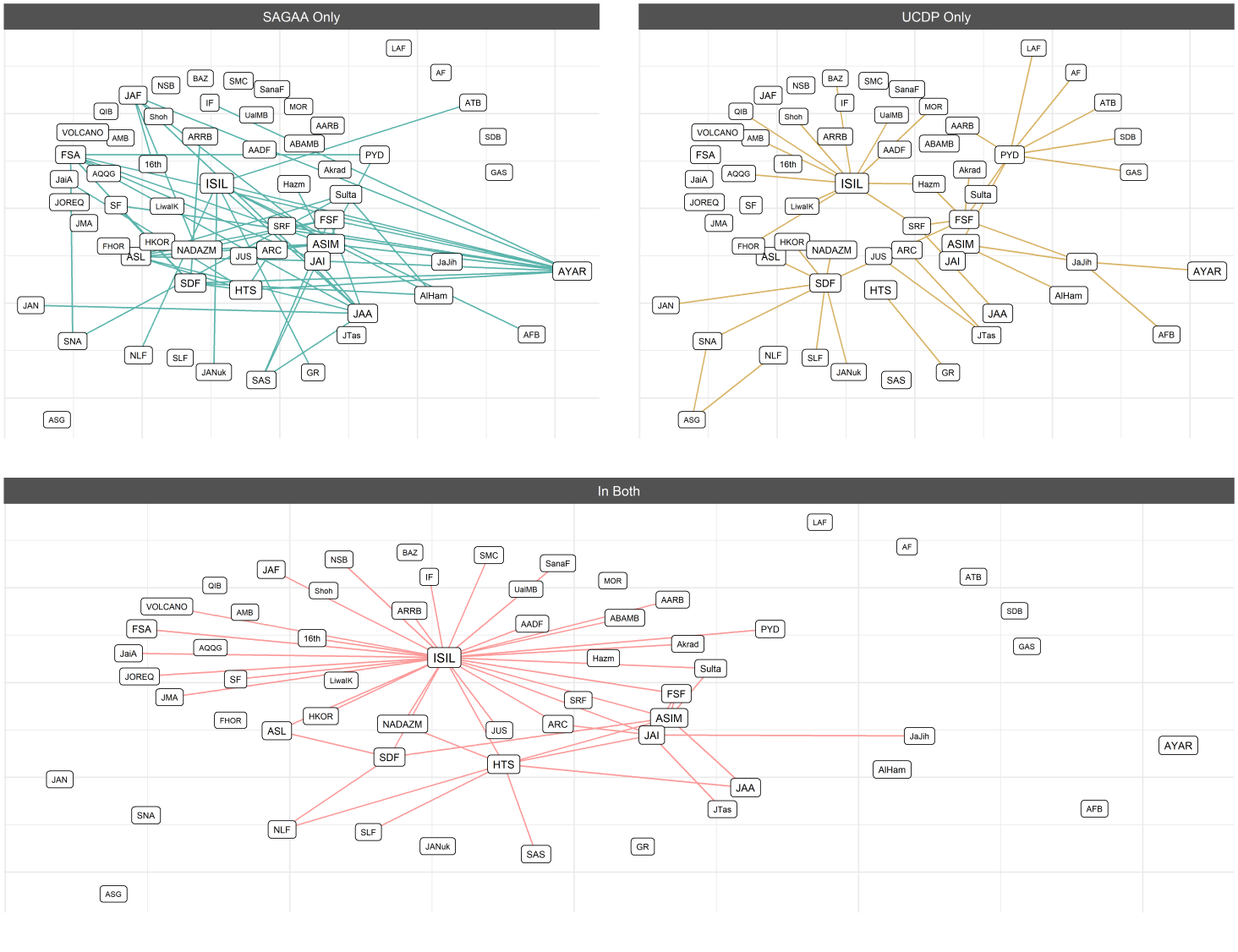
Comparison to UCDP: Events

- Between 2012-2018 UCDP contains triple the number of incidents as the SAGAA dataset, **but** 78.9% of the UCDP incidents are from a single dyad—described as "Syrian Government v. Syrian Insurgents."

Comparison to UCDP: Events

- Between 2012-2018 UCDP contains triple the number of incidents as the SAGAA dataset, **but** 78.9% of the UCDP incidents are from a single dyad—described as "Syrian Government v. Syrian Insurgents."
- A significant proportion of identified data (e.g. not "Syrian Insurgents") in the UCDP dataset is focused on government attacks and attacks by the Islamic State

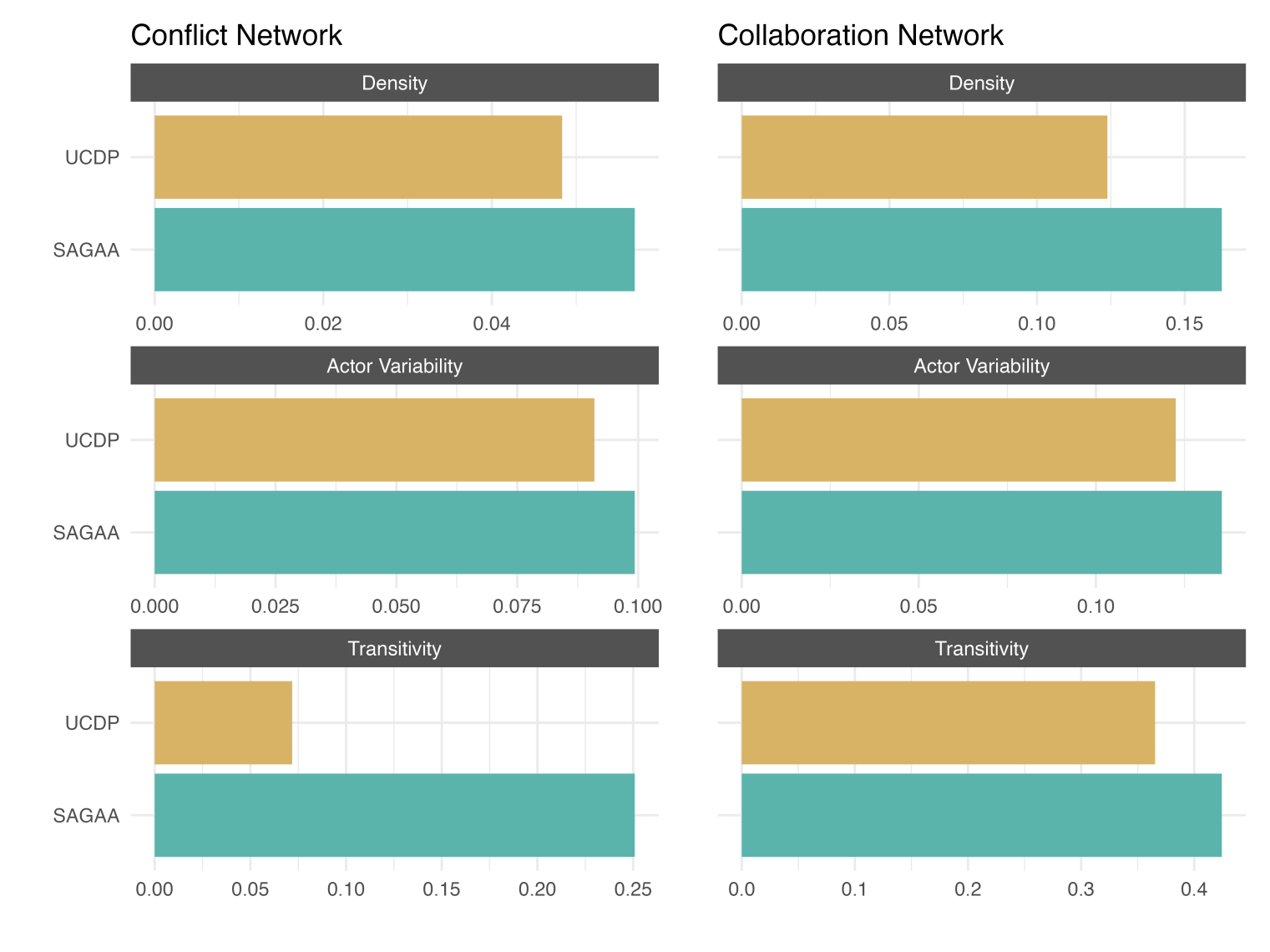
Conflict Event Comparison between SAGAA and UCDP



UCDP

SAGAA Only
HTS
ARC
SF
FSA
AARB
ATB
AF
JAI
ABAMB
JUS
JAA
ASIM
GAS
ISIL
UalMB
JAN
AYAR
JMA
AFB
JANuk
FHOR
FSF
QIB
LiwalK
NLF
NADAZM
16th
ASL
AADF
SLF
IF
Hazm
MAI
SDB
SRF
AMB
Sulta
Akrad
ARRB
BAZ
SAS
AlHam
NSB
PYD
SMC
SanaF
UCDP Only
HTS
ARC
SF
FSA
AARB
ATB
AF
JAI
ABAMB
JUS
JAA
ASIM
GAS
ISIL
UalMB
JAN
AYAR
JMA
AFB
JANuk
FHOR
FSF
QIB
LiwalK
NLF
NADAZM
16th
ASL
AADF
SLF
IF
Hazm
MAI
SDB
SRF
AMB
Sulta
Akrad
ARRB
BAZ
SAS
AlHam
NSB
PYD
SMC
SanaF
In Both
HTS
ARC
SF
FSA
AARB
ATB
AF
JAI
ABAMB
JUS
JAA
ASIM
GAS
ISIL
UalMB
JAN
AYAR
JMA
AFB
JANuk
FSF
QIB
FHOR
LiwalK
NLF
NADAZM
16th
ASL
AADF
SLF
IF
Hazm
MAI
SDB
SRF
AMB
Sulta
Akrad
ARRB
BAZ
SAS
AlHam
NSB
PYD
SMC
SanaF

Descriptive Comparison between SAGAA and UCDP



Predictive Comparison

- Next we use a network reconstruction approach to see how well the SAGAA and UCDP datasets can reconstruct each other.

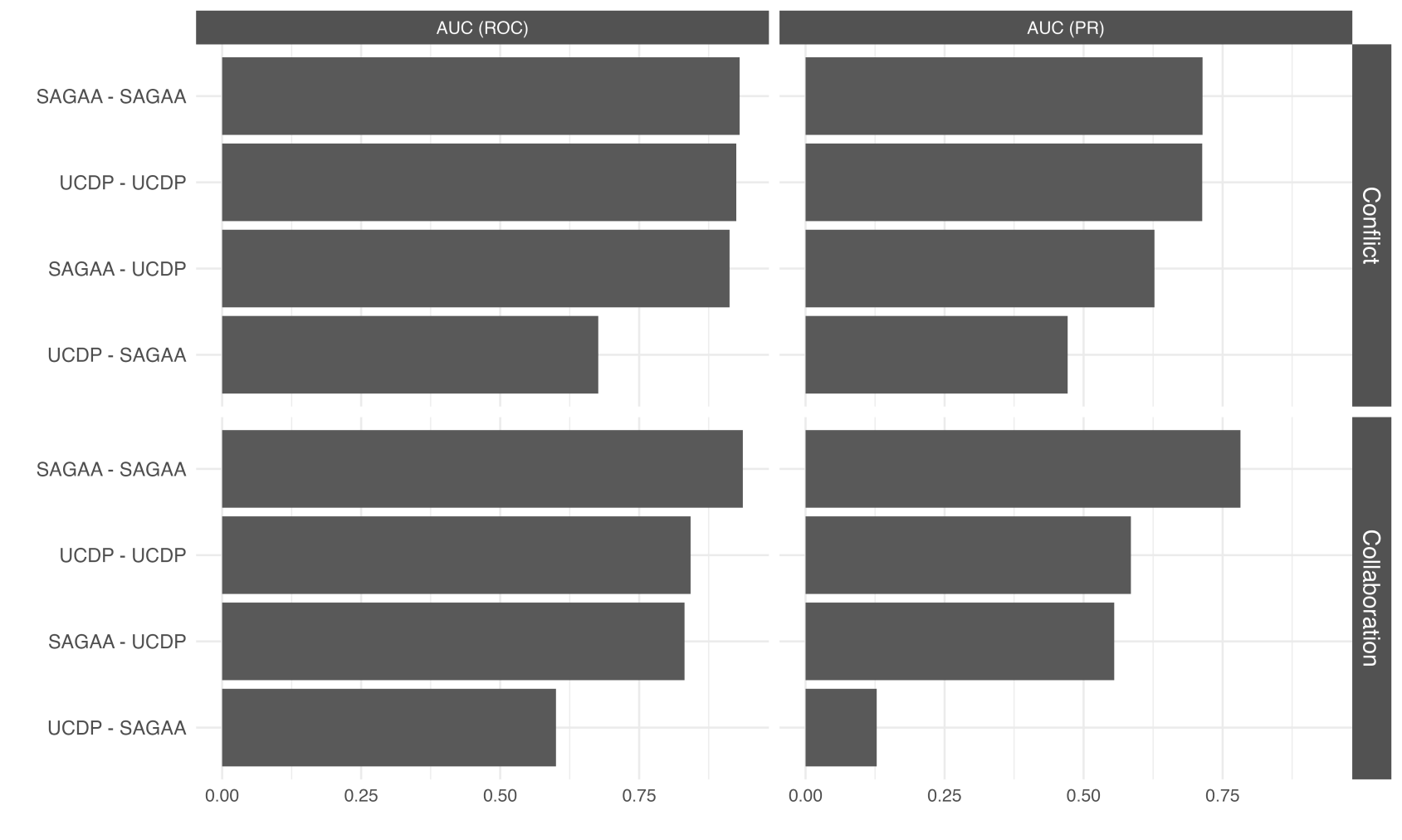
Predictive Comparison

- Next we use a network reconstruction approach to see how well the SAGAA and UCDP datasets can reconstruct each other.
- Begin with the assumption that both datasets provide incomplete representations of the conflict Demarest and Langer 2022, Donnay et al. 2019 .
  - For example, we know UCDP primarily tracks events involving the Syrian State or ISIS, while SAGAA is likely to suffer from some level of incompleteness as well

Predictive Comparison

- Next we use a network reconstruction approach to see how well the SAGAA and UCDP datasets can reconstruct each other.
- Begin with the assumption that both datasets provide incomplete representations of the conflict Demarest and Langer 2022, Donnay et al. 2019 .
  - For example, we know UCDP primarily tracks events involving the Syrian State or ISIS, while SAGAA is likely to suffer from some level of incompleteness as well
- But there is a signal in both datasets and we can use a latent factor model to extract that signal, specifically, we see how well each network can reconstruct the other and itself in an out-of-sample 10-fold cross-validation context

Predictive Comparison Results



Wrapping up

- News-based event data sources often do track the conflict at a high level in terms of at least the actors involved
- This can lead to a lack of understanding about the complexity of the system and would certainly complicate the work of anyone trying to do group-group level analysis
- Rebel media can provide a more granular view of the conflict, but it is not without its own challenges
- Gotta go back to frenemies now but feedback on this would be highly appreciated!

Thanks!