University of Kent | Institute of Cyber Security for Society (iCSS)

The 9th International Symposium on Security and Privacy in Social Networks and Big Data (SocialSec 2023), Canterbury, UK, August 14–16, 2023

Cyber Security Researchers on Online Social Networks: From the Lens of the UK's ACEs-CSR on Twitter

Mohamad Imad Mahaini & Shujun Li

Institute of Cyber Security for Society (iCSS) & School of Computing, University of Kent, Canterbury, UK

{mim, S.J.Li}@kent.ac.uk

Agenda

- Introduction / Motivation
- Related Work
- Contributions
- Methodology
- Results
- Limitation & Future Work
- Conclusion

Introduction / Motivation

- 4.7 B active users on OSNs, more than half of the world population [1].
- Identifying and finding users who form different online communities has become an interesting research topic.
- Studying such users' communities can reveal useful insights: e.g., their memberships, people's opinions, intentions and motivations of online users' activities.
- Cyber security experts and criminals also use OSNs:

Cyber security experts	Cyber security criminals
knowledge exchange	reach out to victims
cyber security awareness	boast about their past "achievements"
offering help to people and organizations	talk about their future attacking plans

Introduction / Motivation

- The cyber security domain is becoming increasingly complex (vast advancements in technology, computing equipment and IT infrastructure).
- A wide range of people are involved (professionals, researchers, cyber criminals, journalists, activists, government agents, etc).
- The activities of those OSN users have been found to be a good source of information for many purposes: cyber threat intelligence and understanding behaviours of cyber criminals and related groups [1-3].

Related Work

- Nowadays, researchers have access to different types of users on OSN.
- When considering the users' relation to the cyber security domain, there are a lot of studies about the following categories of OSN users:
  - Cyber criminal groups [2,15,34]
  - Activist and hacktivist groups [13,24]
  - Non-expert cyber security users [25,30]
- To the best of our knowledge, there has been no previous work on studying cyber security researchers using a data-driven approach based on OSN data.

Research Objective & Questions

- Study the cyber security researchers on OSNs.
- Taking ACEs-CSR network on Twitter as a case study.
  - The Academic Centres of Excellence in Cyber Security Research (ACEs-CSR) scheme is sponsored by the National Cyber Security Centre (NCSC) and UK Research and Innovation.

- **RQ1**: How to identify cyber security research related accounts on Twitter?
- **RQ2**: What is the social structure of a typical cyber security research community on Twitter?
- **RQ3**: What topics do cyber security research related users discuss on Twitter?
- **RQ4**: What is the general sentiment towards the ACE-CSR program and the ACEs-CSR?

Contributions

- We tested the performance of the machine learning (ML) classifiers reported in [18] for detecting cyber security related accounts in a real-world setting.
- We developed a new ML classifier to detect cyber security research related accounts with good performance.
- Using graph-based analysis and community detection algorithms, our study showed that such methods can produce useful insights about cyber security research communities on Twitter.
- Using topic modelling, we identified a wide range of topics discussed by cyber security researchers on Twitter, including some less related to cyber security.
- By applying sentiment analysis, we observed a generally positive sentiment on the ACE-CSR programme and the ACEs-CSR.

Data Collection

- Seed Accounts
  - Created a list of 19 Twitter accounts, each corresponds to an ACE-CSR, selected as follows.
    I. ACE-CSR Official account on Twitter, or the ACE-CSR lead's account.
    II. The most well-known cyber security in the corresponding ACE-CSR.
- Friends and Followers of the Seed Accounts
  - For each seed account (Lv1), we fetched its friends and followers, denoted as Lv2.
  - We got 42,028 accounts for further analysis (19 in Lv1 and 42,009 in Lv2).
- Account Timelines
  - We used the Twitter API to obtain these timelines (up to 3,250 tweets per account).

RQ1: ML Classifiers

- Studying the ACEs-CSR network on Twitter required identifying accounts that are both cyber security and research related.

  → Thus, two classifiers were needed.

- Additionally, we needed a classifier to detect whether a Twitter account belongs to an individual or non-individual (e.g., group, organisation, government, NGO, news channel).

- Regarding the cyber security related and individual classifiers, we used two classifiers we developed in 2021, reported in [18].

- As for Research related classifier, we built a new classifier.

Baseline & Individual Classifiers

- Before using these two classifiers for prediction, we re-trained them and evaluated their performances using our ACEs-CSR dataset (42k accounts).
- We utilised the same original labelled datasets and followed the same steps for the feature extraction phase from [18].
- Then, we selected the best-performing feature sets according to the reported results.

Feature Extraction

Original features were arranged into 5 groups:
Profile (P), Behavioural (B), Content Statistics (C), Linguistic (L), Keyword-based (K)

Profile Features (P)			Behavioral Features (B)			Content Statistics Features (C)		
Screen Name	F01	LEN (screen name)	Tweets Statistics	F26	CNT (Tweets)	Cyber Security Keywords Statistics	F48	CNT (Keywords)
	F02	CNT (Alphabetic char)		F27	CNT (Original tweets)		F49	CNT (Keywords) [Original tweets]
	F03	CNT (Lowercase char)		F28	CNT (Retweets)		F50	CNT (Unique keywords)
	F04	CNT (Uppercase char)		F29	CNT (Replies)		F51	CNT (Unique keywords) [Original tweets]
	F05	CNT (Numerical char)		F30	CNT (Tweets with mentions)		F52	CNT (Tweets with keywords)
	F06	CNT (Special char)		F31	Ratio (Original tweets to all)		F53	Ratio (Tweets with keywords to all)
Description	F07	LEN (description)		F32	Ratio (Retweets to all)	Readability & Diversity	F54	Flesch-Kincaid Score
	F08	CNT (Alphabetic char)		F33	AVG (Number of mentions)		F55	SMOG Index
	F09	CNT (Lowercase char)		F34	AVG (Number of hashtags)		F56	Lexical Diversity
	F10	CNT (Uppercase char)		F35	AVG (Number of URLs)			
	F11	CNT (Numerical char)	Network	F36	CNT (Tweets received likes)	**Linguistic Features (L)**		
	F12	CNT (Special char)		F37	CNT (Tweets were retweeted)	LIWC	F57	Measures L{93}
	F13	CNT (Control char)		F38	CNT (Mentioned users)			
	F14	CNT (Words)		F39	CNT (Replied-to users)	**Keyword-based Features (K)**		
	F15	CNT (Keywords)		F40	CNT (Likes given)	Keywords Frequencies	F58	Weirdness Score
Network	F16	CNT (Friends)		F41	CNT (Likes received)		F59	Prototypical Words
	F17	CNT (Followers)		F42	CNT (Retweets received)		F60	TF-IDF Score
	F18	Followers/Friends	Activity	F43	AVG (Daily Tweets)		F61	User Count (UC)
Misc	F19	Profile Image used?		F44	AVG (Weekly Tweets)		F62	Hybrid Metric UC-IDF
	F20	Profile Theme used?		F45	AVG (Monthly Tweets)		F63	Hybrid Metric UC-TFIDF
	F21	Location provided?		F46	AVG (time between tweets)			
	F22	CNT (Lists)		F47	STD (time between tweets)			
	F23	Account protected?						
	F24	URL provided?						
	F25	Account Age						

LEN	Length
CNT	Count
AVG	Average
STD	Standard Deviation

Research Related Classifier

- We considered a data sample as a positive case if it is involved with any research work or activity related to research.
- This is judged based on the account's description and timeline.
- This makes any cyber security researcher a positive case, even if they does not work in academia or is not associated with any research organisation. This is the significant difference between our Research classifier and the Academia classifier reported in [18].

Research Related Classifier (Features)

- Besides the features we extracted for the Baseline and Individual classifiers, we introduced new features (R group).

- Connectivity with seed accounts (number of followers/friends with seeds)
- Researcher Keywords (a compiled list of 27 keywords that can be found in the Twitter Display Name and Description fields and can refer to an account that is related to research (Professor, Academic, Lecturer, University, PHD…).
- Verified: a binary value corresponding to the Verified profile attribute in Twitter.
- Website category: derived from the “Website” field of the account's profile. We used three categories: Research, Mixed, and Other.

Research Related Classifier (Training Dataset)

- After using the Baseline classifier to predict the labels of the 42k accounts, we kept only the accounts that were predicted as cyber security related accounts.

- The manual labelling process was done in iterations until we got a balanced dataset of 1k data samples.

ML Classifiers Training Results

- For the prediction of the cyber security research related accounts, we selected the trained classifier built using the R feature set and the SVM-R model (F1-score = 83%, Precision = 96%)

Table 1: Experimental results of all the machine learning classifiers

Task	Featurs	#F	#S	Decision Tree			Random Forest			Extra Trees		
				F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec
Baseline	PBCL	149	1974	0.88	0.88	0.89	0.91	0.90	0.95	0.91	0.91	0.94
Individual	PBCL	149	957	0.84	0.84	0.84	0.89	0.91	0.87	0.88	0.93	0.84
Academia	K:UCIDF	200	245	0.81	0.68	1.00	0.90	0.82	1.00	0.92	0.85	1.00
Research	R	46	1003	0.78	0.94	0.67	0.81	0.94	0.72	0.81	0.94	0.71

Logistic Reg.			XGBoost			SVM (Linear)			SVM (RBF)		
F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec
0.90	0.91	0.91	0.91	0.90	0.94	0.91	0.91	0.92	0.90	0.91	0.91
0.89	0.90	0.88	0.91	0.92	0.90	0.89	0.91	0.87	0.87	0.91	0.83
0.00	0.00	0.00	0.82	0.69	1.00	0.00	0.00	0.00	0.43	0.71	0.58
0.82	0.97	0.72	0.81	0.94	0.72	0.82	0.97	0.72	0.83	0.96	0.73

ML Classifiers Prediction Results

- Research classifier was applied after the Baseline classifier, so we only considered the positive samples (9,377) predicted by the Baseline classifier as the input for this classifier.
- Finally, we got 1,684 positive samples and 7,693 negative samples.
- After manual verification, the selected nodes became 1,817.

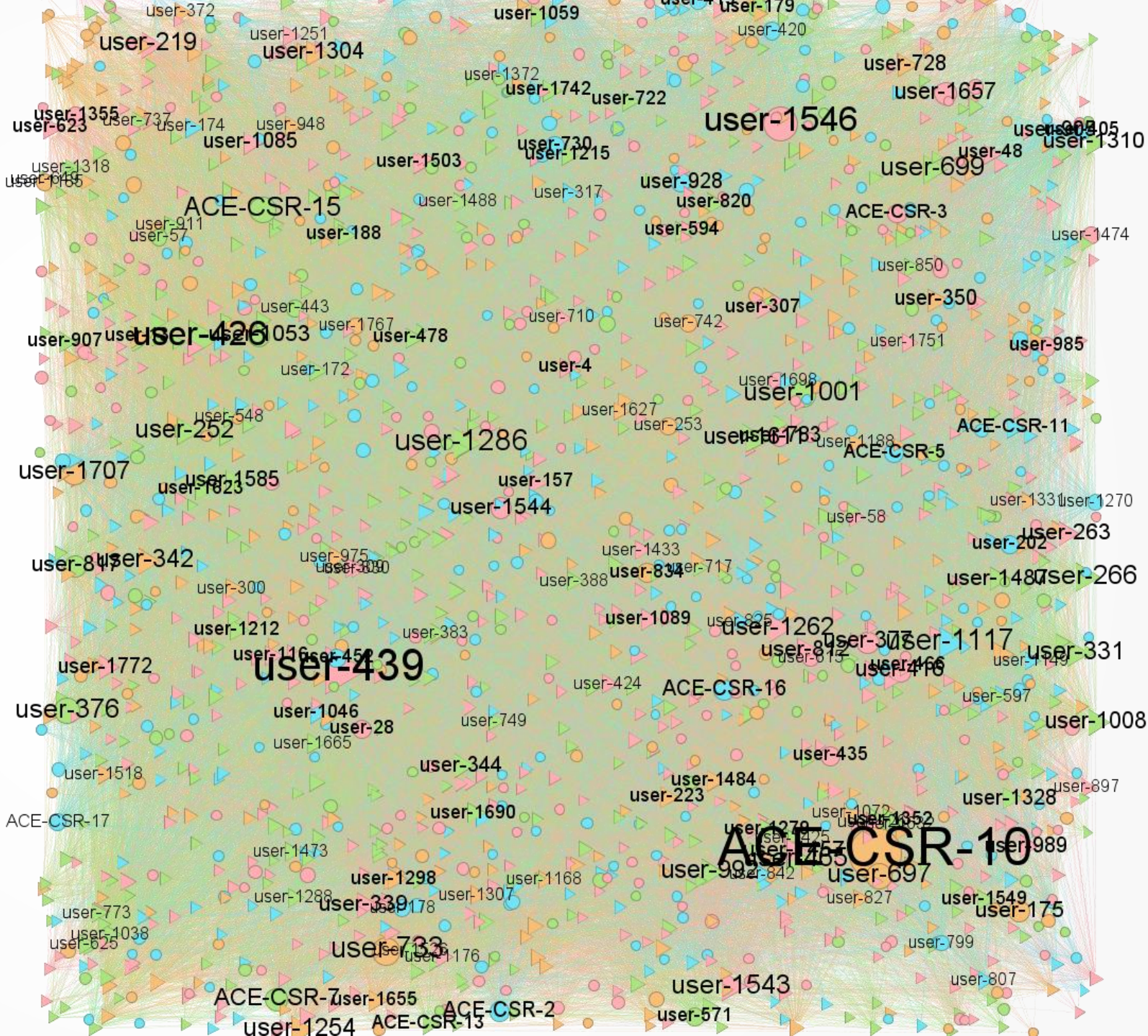
Table 2: The prediction results of the used machine learning classifiers

Task	Features	Model	#(Samples)	Prediction Samples	Positive	Negative
Baseline	PBCL	RF	42,028	42,028	9,377	32,651
Individual	PBCL	RF	42,028	9,377	4,795	4,582
Research	R	SVM-R	42,028	9,377	1,684	7,693

RQ2:

Social Structural Analysis

- To construct the social graph of the ACEs-CSR network, we had to identify the nodes and their edges.
- As a result, we built a directed graph with 1,817 nodes and 64,826 edges.
- Dataset was anonymized, and the node names were changed.
- **ACE-CSR-i**: represents ACE-CSR node.
- **User-j**: represents non-ACE-CSR node.

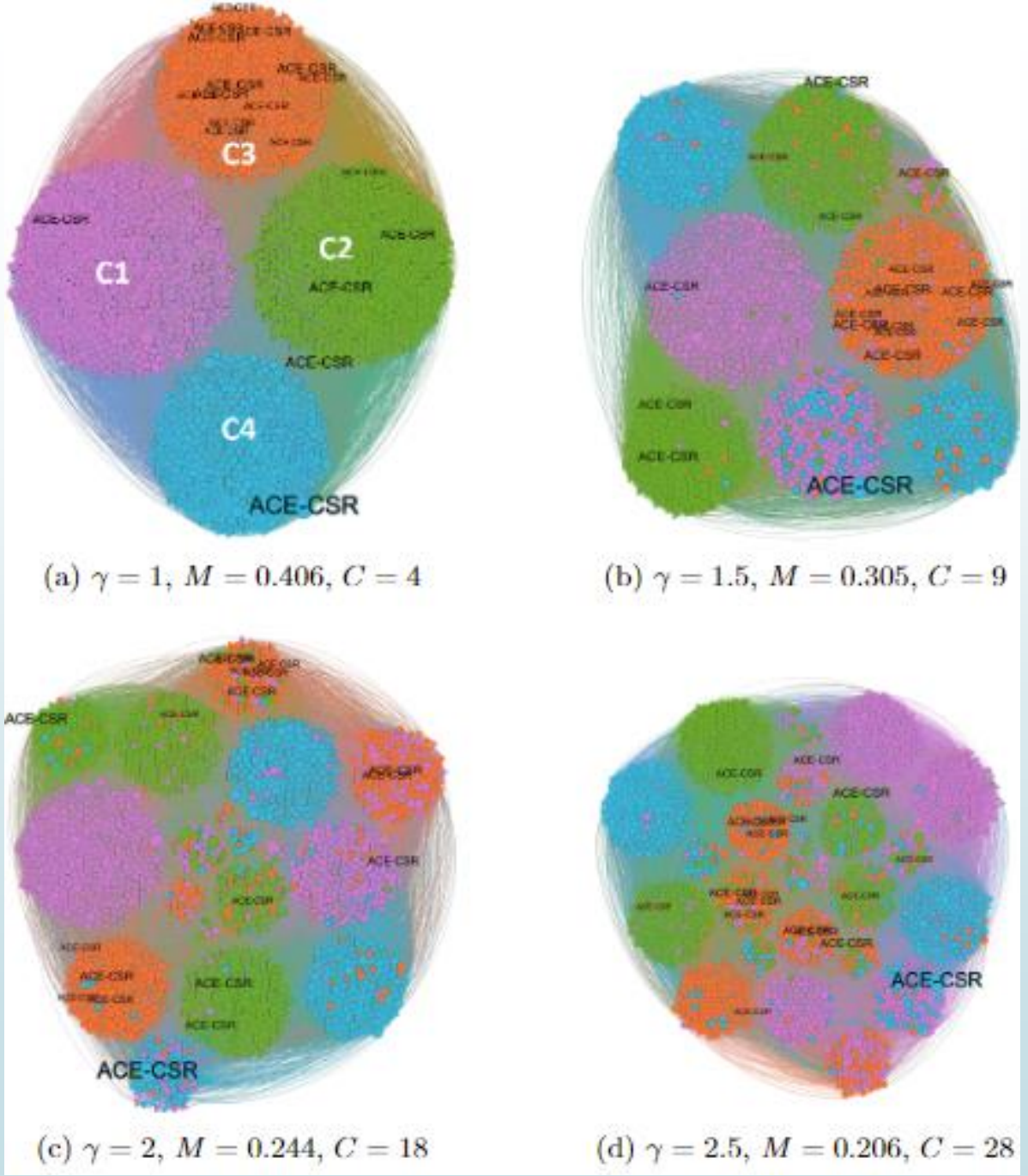


Communities Detection & Analysis

- To study the big ACE-CSR graph, we broke it down into sub-graphs, where each graph represents a community or a group of users that have something in common.
- Since the graph nodes had no ground truth labels of any characteristic, using supervised classifiers was impossible. Thus, we used unsupervised clustering techniques to divide the graph nodes into clusters (i.e., communities).
- We tested several community detection algorithms that are widely adopted in the literature, DBSCAN [31], Girvan-Newman algorithm [10] and modularity-optimisation-based algorithms [22] such as the Louvain algorithm [7] and Leiden algorithm [35].

Communities Detection: (Resolution ?)

- We adopted the Leiden algorithm [35] at the end as its results were the best compared to other methods.
- Modularity-based algorithms use a resolution parameter γ which controls the size of the detected communities. Increasing it results in more communities, while reducing it does the opposite.
- We showed the results of using the following resolutions: 1.0, 1.5, 2.0 and 2.5.



(a) $\gamma = 1,\ M = 0.406,\ C = 4$

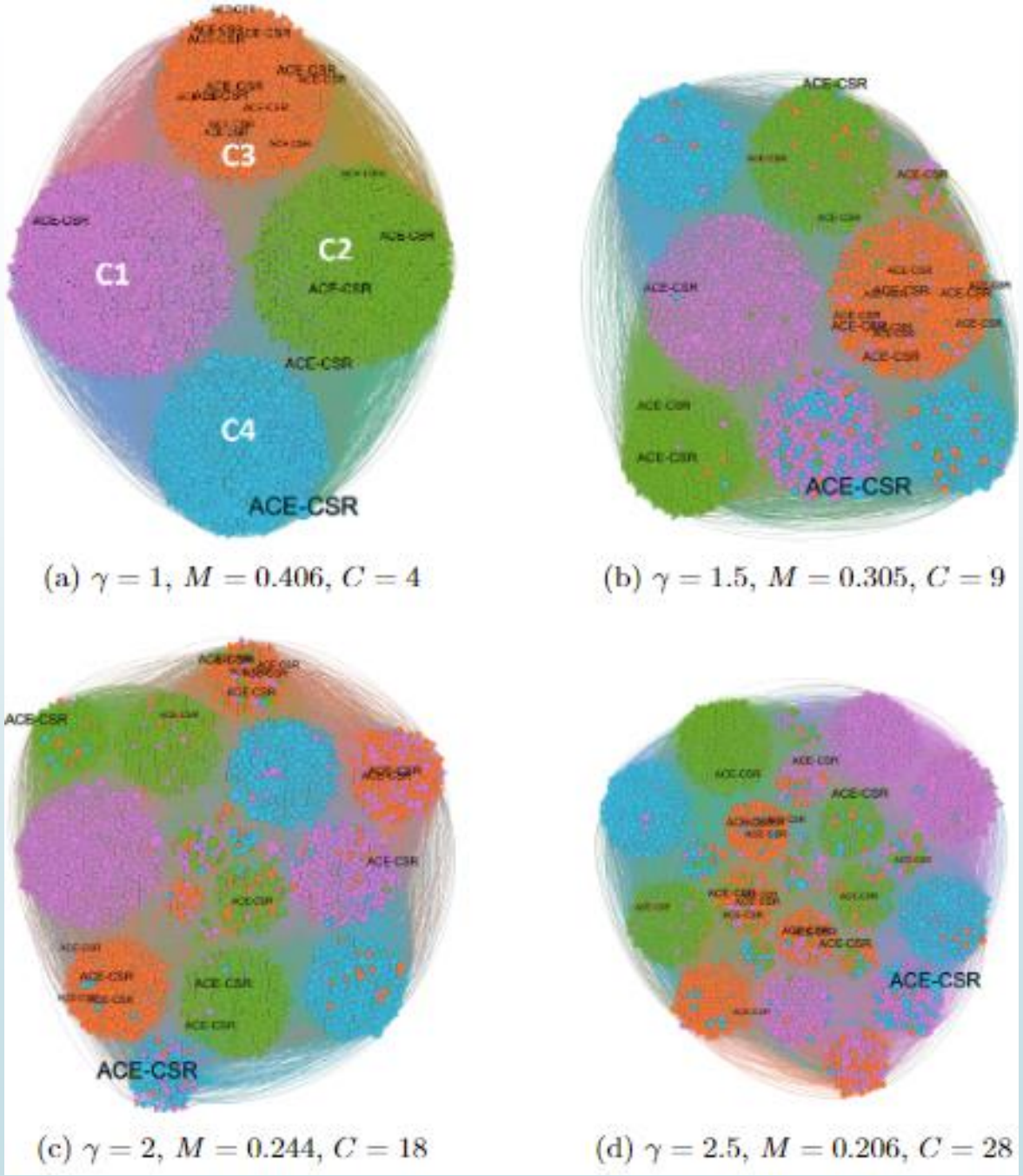
(b) $\gamma = 1.5,\ M = 0.305,\ C = 9$

(c) $\gamma = 2,\ M = 0.244,\ C = 18$

(d) $\gamma = 2.5,\ M = 0.206,\ C = 28$

Communities Analysis: (Initial Insights)

- We expected each ACE-CSR Twitter account to have a strong community around its node, but this was not the case for a few of them!
- Some ACE-CSR nodes always appear in the same cluster regardless of the chosen resolution.
- Using different values for resolution and checking the resulted communities each time, we observed some clusters that do not have any ACE-CSR nodes.



(a) $\gamma = 1,\ M = 0.406,\ C = 4$

(b) $\gamma = 1.5,\ M = 0.305,\ C = 9$

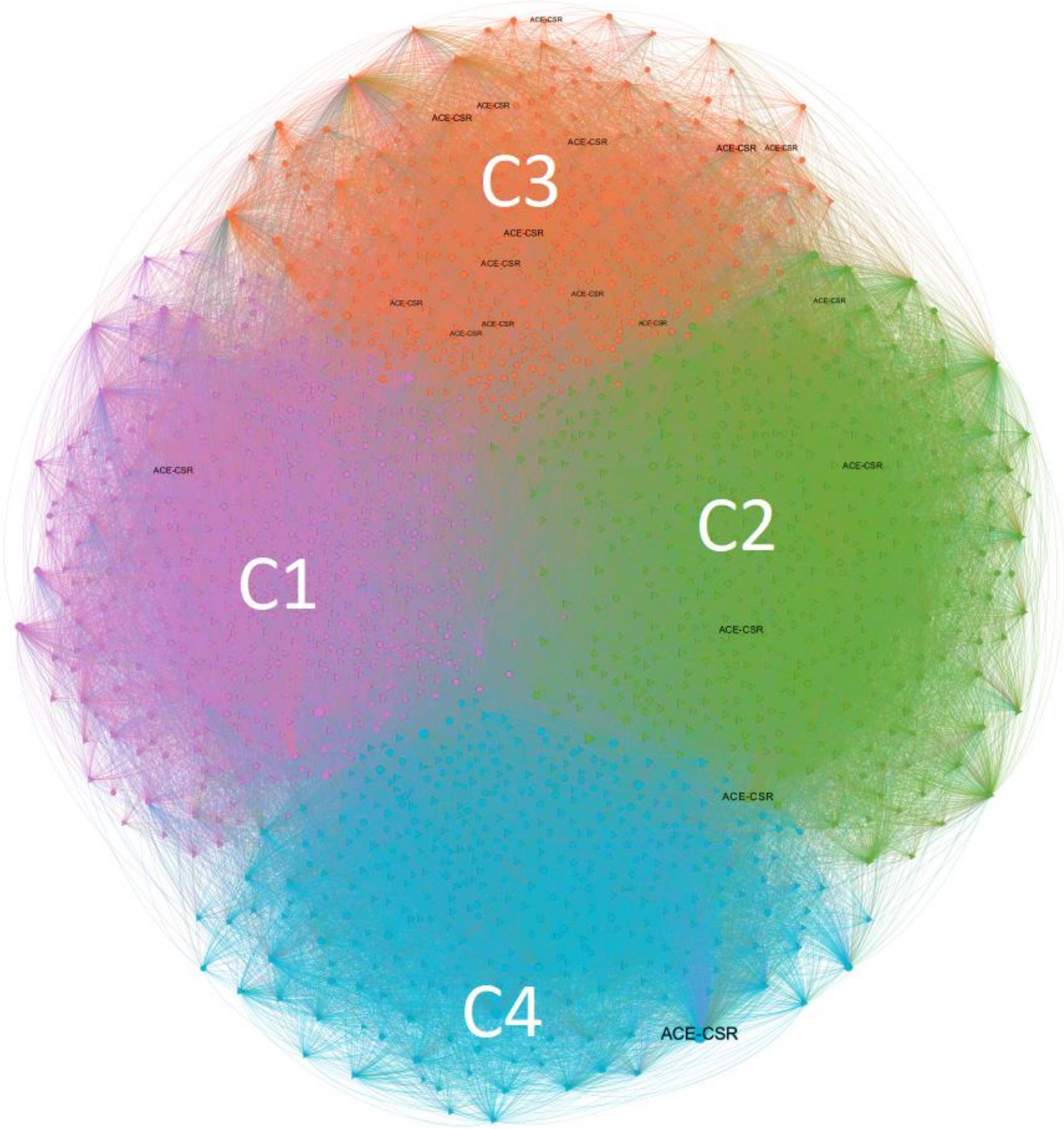
(c) $\gamma = 2,\ M = 0.244,\ C = 18$

(d) $\gamma = 2.5,\ M = 0.206,\ C = 28$

Communities Analysis

- Selecting the right resolution depends on how many communities we want to work with.

- For simplicity and explainability, we carried out some additional analysis focusing only on the communities corresponding to γ = 1



Communities Analysis (Individual Members)

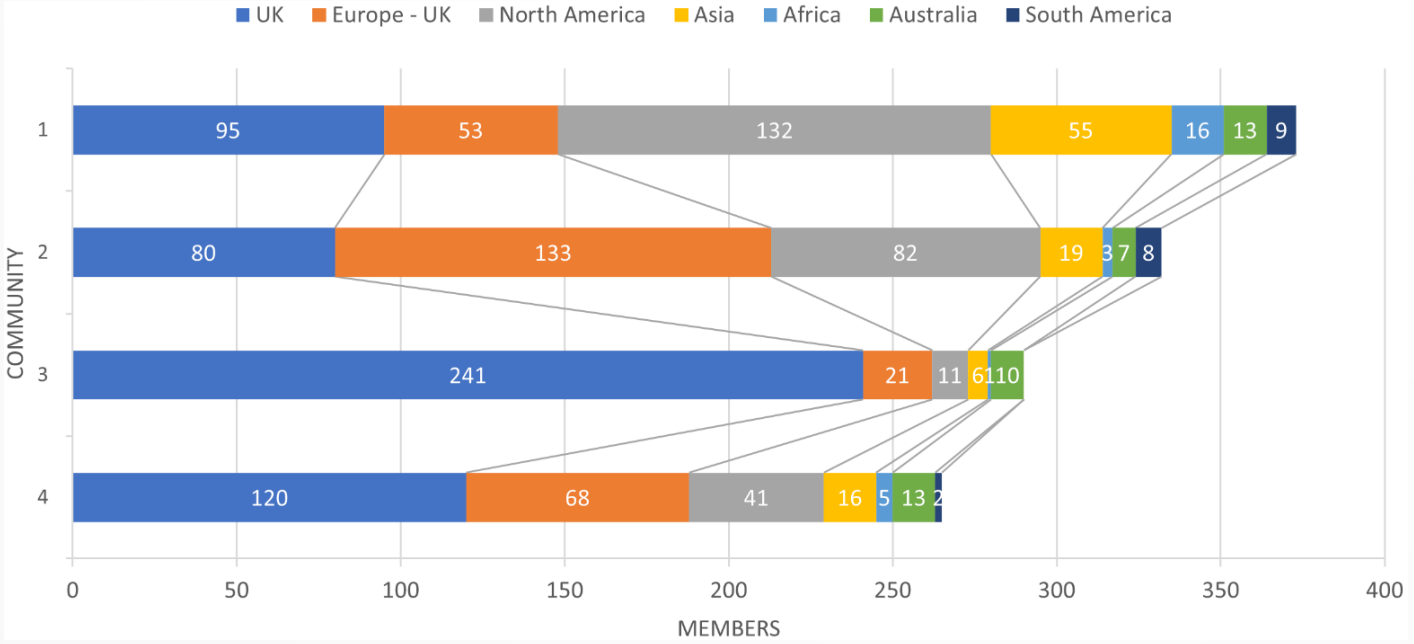
- Knowing the percentage of individuals in the ACEs-CSR network is interesting.
- We used the predicted labels from the Individual classifier for this analysis.
- The overall individual/non-individual percentages were 69.40%, and 30.60%, respectively.
- The individual percentage reached 79.14% for Community C2, which is higher than other communities.
- Upon inspecting C2, we found that individuals in this community are often well-known researchers and figures in the cyber security research domain.

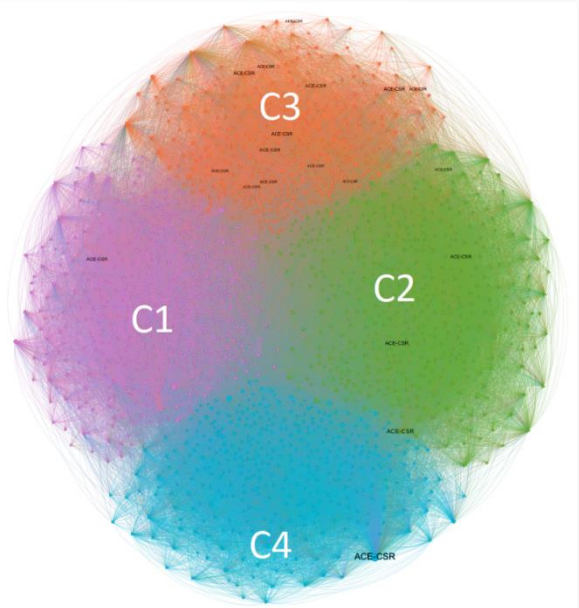
Table 3: Statistics of discovered communities ($\gamma = 1$)

Community	Colour	Members	Size	Individual Accounts	Non-individual Accounts
C1	Purple	595	32.75%	72.61%	27.39%
C2	Green	465	25.59%	79.14%	20.86%
C3	Orange	382	21.02%	51.83%	48.17%
C4	Blue	375	20.64%	70.13%	29.87%

Communities Analysis (Location)

- The percentage of the accounts with the information provided in the whole data we collected is 61.41%, while it is 77.55% for the ACEs-CSR network.
- C3 seems a more UK-centric one, but the other three are highly international.
- C1 and C2 are dominated by non-UK accounts, and C1 seems to be the most international cluster.
- The percentage of Asian accounts in C1 is substantially higher than the other communities, indicating it may be the one representing the UK-Asia links.
- Considering UK vs non-UK accounts, C4 looks like a more balanced cluster with an approximately 1:1 ratio.





RQ3: Topic Modelling (TM) Analysis (Training)

- We used TM to automatically identify topics discussed by the cyber security research related accounts in the ACEs-CSR network.
- The data used for this analysis was the Twitter account timelines after a preprocessing.
- We used the LDA algorithm [6], one of the most widely used TM algorithms in the literature [2,25].
- LDA is an unsupervised method for clustering N documents into k categories (topics).
- LDA has 2 parameters: k, the number of topics, and r, the maximum number of iterations.
- We tried to set the values for the parameters automatically by training the model using a series of values for each parameter and then training the model and assessing the results using the coherence score UCI.
- Finally, we set k to 10 and r to 200.

Topic Modelling Analysis (Results)

- Using the inter-topic distance map (pyLDAvis), we can notice that the correlation between topics is minimum.
- Apart from T4, all the other topics are relatively balanced in size, ranging from 6.4% to 10.6% with average of 8.4%.
- Several topical themes: research, privacy, education, technical, and politics.
- Ignoring T4, the top discussed topic was T5 (“Cyber Security for Students’”, 10.6%), followed by T6 (“Data Protection Laws”', 10%).
- Interestingly, politics-related and cyber conflict discussions in T7 also had a good share with 8.4%.

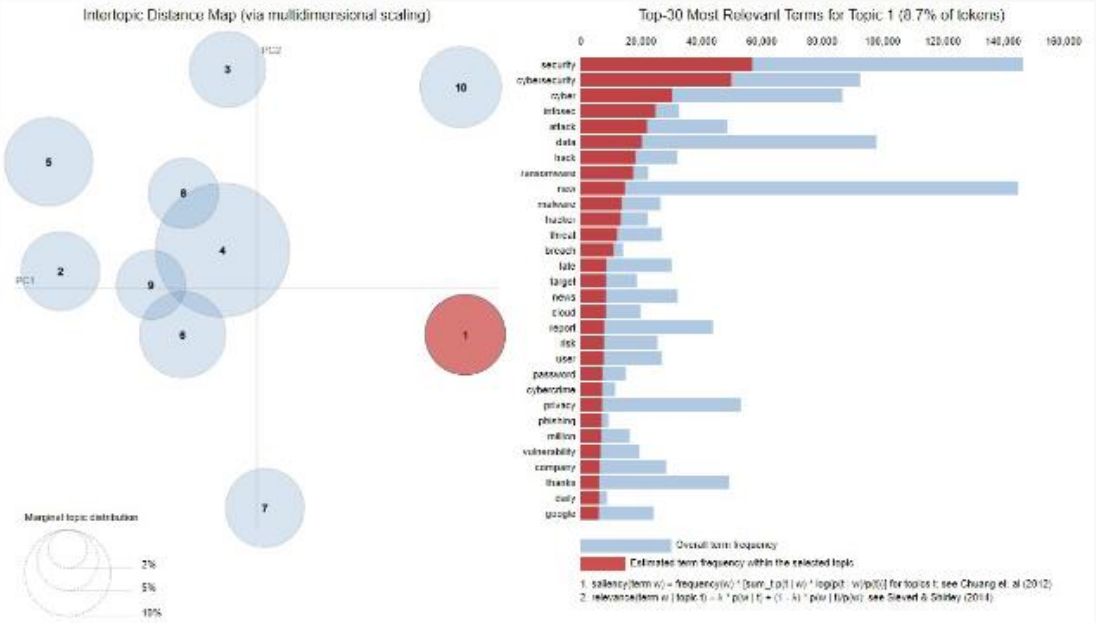


Fig. 3: Visualisation of the estimated topics by the LDA algorithm

Table 4: LDA topics with top 15 keywords, ranked in descending order by size

ID	Topic Name	Size (%)	Top Keywords
4	General Terms	24.2	like, people, think, time, good, work, know, need, look, year, thing, day, great, want, way
5	Cyber Security for Students	10.6	student, today, great, day, new, cyber, work, look, event, research, talk, join, team, uk, year
6	Data Protection Laws	10	data, privacy, law, new, right, digital, eu, ai, internet, tech, work, protection, facebook, online, gdpr
10	Vulnerabilities & Threats	8.9	new, security, malware, attack, tool, vulnerability, release, exploit, code, hack, blog, use, android, linux, update
1	Cyber Security Incidents	8.7	security, cybersecurity, cyber, infosec, attack, data, hack, ransomware, new, malware, hacker, threat, breach, late, target
2	Security Research & Education	8.4	research, new, work, security, social, read, join, look, digital, data, online, study, report, project, researcher
7	Cyber Conflict & Politics	8.4	cyber, state, russia, new, russian, china, war, ukraine, government, attack, world, country, intelligence, military, report
3	Cryptography & Privacy Research	7.9	paper, security, work, research, new, privacy, talk, crypto, open, program, phd, bitcoin, student, computer, blockchain
8	Cyber Security Events	6.6	cybersecurity, security, cyber, join, learn, new, register, ic, today, check, day, event, talk, team, course
9	ICT Industry	6.4	ai, iot, technology, data, learn, new, business, tech, future, digital, market, innovation, report, industry, world

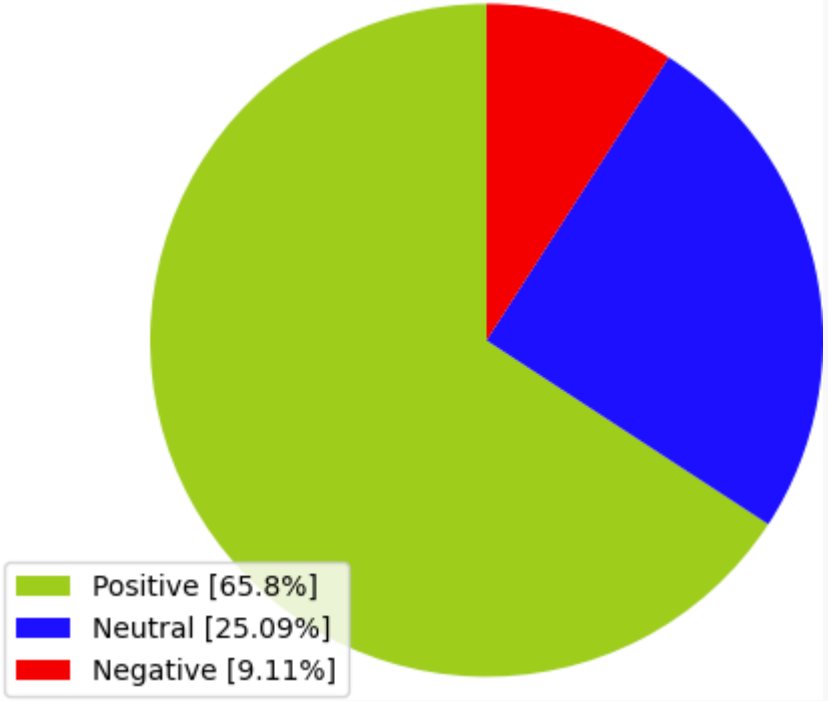
RQ4: Sentiment Analysis (Training)

- The ACE-CSR programme started almost a decade ago, and such an analysis can provide useful insights about what to do in the future with the ACE-CSR programme.
- We created a dataset of tweets for this analysis by:
  - Filtering the 42,028 accounts' timelines by searching for tweets related to the ACE-CSR program or ACEs-CSR.
  - Adding also tweets that mentioned any of the 19 seed accounts.
  - Tweets created by the seed accounts were excluded.
- A total of 21,374 tweets were obtained.
- The tweets were pre-processed.
- VADER sentiment analyser was used.

Sentiment Analysis (Results)

- 65.8% of all tweets are classified as positive, 25.09% as neutral, and only 9.11% as negative.
- The results of each sub-group are largely aligned with the main results for all.
- However, the percentage of the positive sentiment in Community C2 (the more “European” community) dropped to 61.25%, while the negative percentage increased to 10.29%.
- The more UK-centric Community C3 saw the lowest negative sentiment (8.74%) across the 4 communities.

		Positive		Neutral		Negative	
Accounts Group	Tweets	Count	%	Count	%	Count	%
Non Research related	13,915	9,306	66.88	3,377	24.27	1,232	8.85
Research related C1	608	406	66.78	134	22.04	68	11.18
Research related C2	1,613	988	61.25	459	28.46	166	10.29
Research related C3	4,485	2,888	64.39	1,205	26.87	392	8.74
Research related C4	753	476	63.21	188	24.97	89	11.82
All accounts	21,374	14,064	65.8	5,363	25.09	1,947	9.11



Limitation & Future Work

The performance of our Research classifier has an F1-score of 83%

→ This can be further improved by considering more candidate features and building a bigger dataset so that other hybrid ML models can be used, such as deep learning based ones.

Our work is based on a single OSN platform (Twitter)

→ Consider other data sources to enlarge the diversity and richness of the data, such as LinkedIn and the websites of universities and research organisations.

Other data sources?

→ Considering a wider range of data sources will allow covering a more representative subset of the targeted research community and their activities

→ We can also consider using scientific data services such as Google Scholar, ResearchGate and DBLP to explore potential correlations between online activities and scientific ones of researchers.

Questions

University of Kent | Institute of Cyber Security for Society (iCSS)

Thank you for listening :)

Cyber Security Researchers on Online Social Networks:
From the Lens of the UK's ACEs-CSR on Twitter

Mohamad Imad Mahaini & Shujun Li

Institute of Cyber Security for Society (iCSS) & School of Computing, University of Kent, Canterbury, UK

{mim, S.J.Li}@kent.ac.uk