An Analysis of Cybercrime Activity within an Underground Gaming Forum

Jack Hughes

Cambridge Cybercrime Conference
11th July 2019

joh32@cam.ac.uk

UNIVERSITY OF CAMBRIDGE

Background

- Research into the role of gaming as an entry point into cybercrime is growing

- Example: **DDoS attacks-as-a-service** can be used by gamers with little technical knowledge to gain an advantage over opponents

- Exposure to, and use of, these services is believed to be a pathway into more serious cybercrime

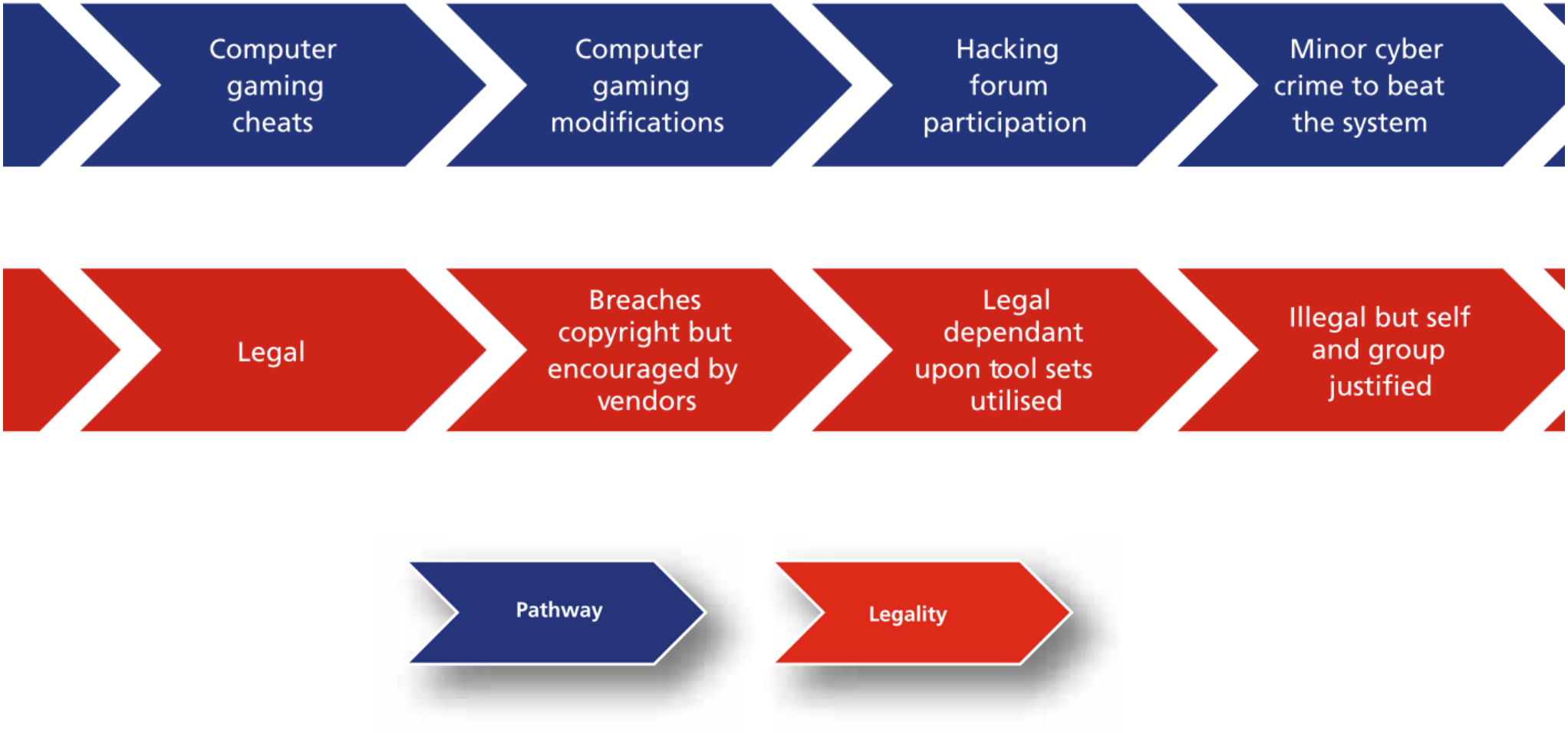


Figure from: National Crime Agency. (2015). Identify, Intervene, Inspire: Helping young people to pursue careers in cyber security, not cyber crime, 6.

Related Work

- Previous work by Pastrana et al.[1]:
  - Analysed **Hack Forums**, for predicting future key actors
  - Produced open-source research tools for analysis

- Hack Forums is a **general-purpose** underground hacking forum
- MPGH is specifically for **multiplayer games**

- Both forums are available on the open web
  - Also available in the CrimeBB dataset, available for research use from the Cambridge Cybercrime Centre

[1] Pastrana S., Hutchings A., Caines A., Buttery P. (2018) Characterizing Eve: Analysing Cybercrime Actors in a Large Underground Forum. In: Bailey M., Holz T., Stamatogiannakis M., Ioannidis S. (eds) Research in Attacks, Intrusions, and Defenses. RAID 2018. Lecture Notes in Computer Science, vol 11050. Springer, Cham

Ethics

- This work has received approval from the Department of Computer Science & Technology's ethics committee

- Only carrying out analysis of collective behaviour, rather than identifying individuals

Studying MPGH

- Aim is not to carry out “predictive policing”, but towards identifying possible intervention points

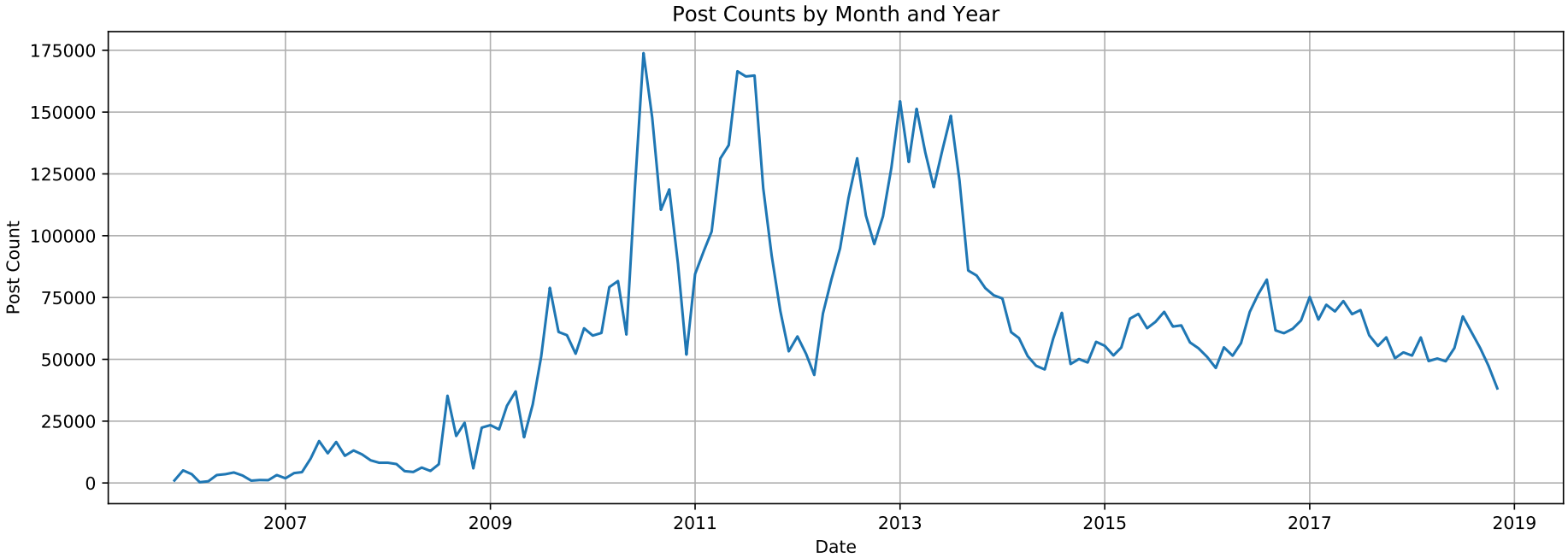
- This work combines prediction techniques to identify characteristics of key actors

Key Actors

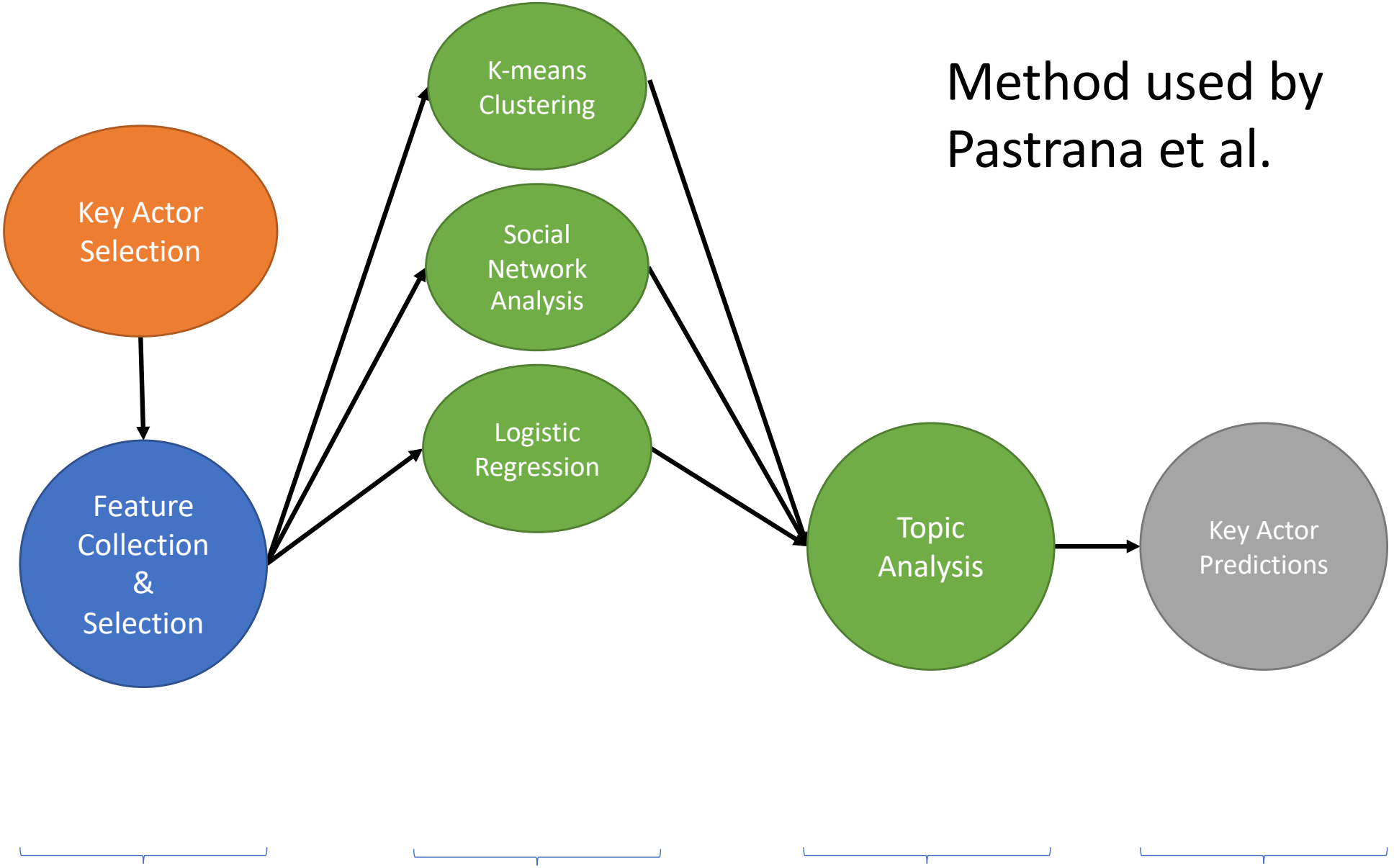
Individuals who have released tools and tutorials on the forum, or have advertised cybercrime related services such as DDoS-for-hire.

MPGH Dataset

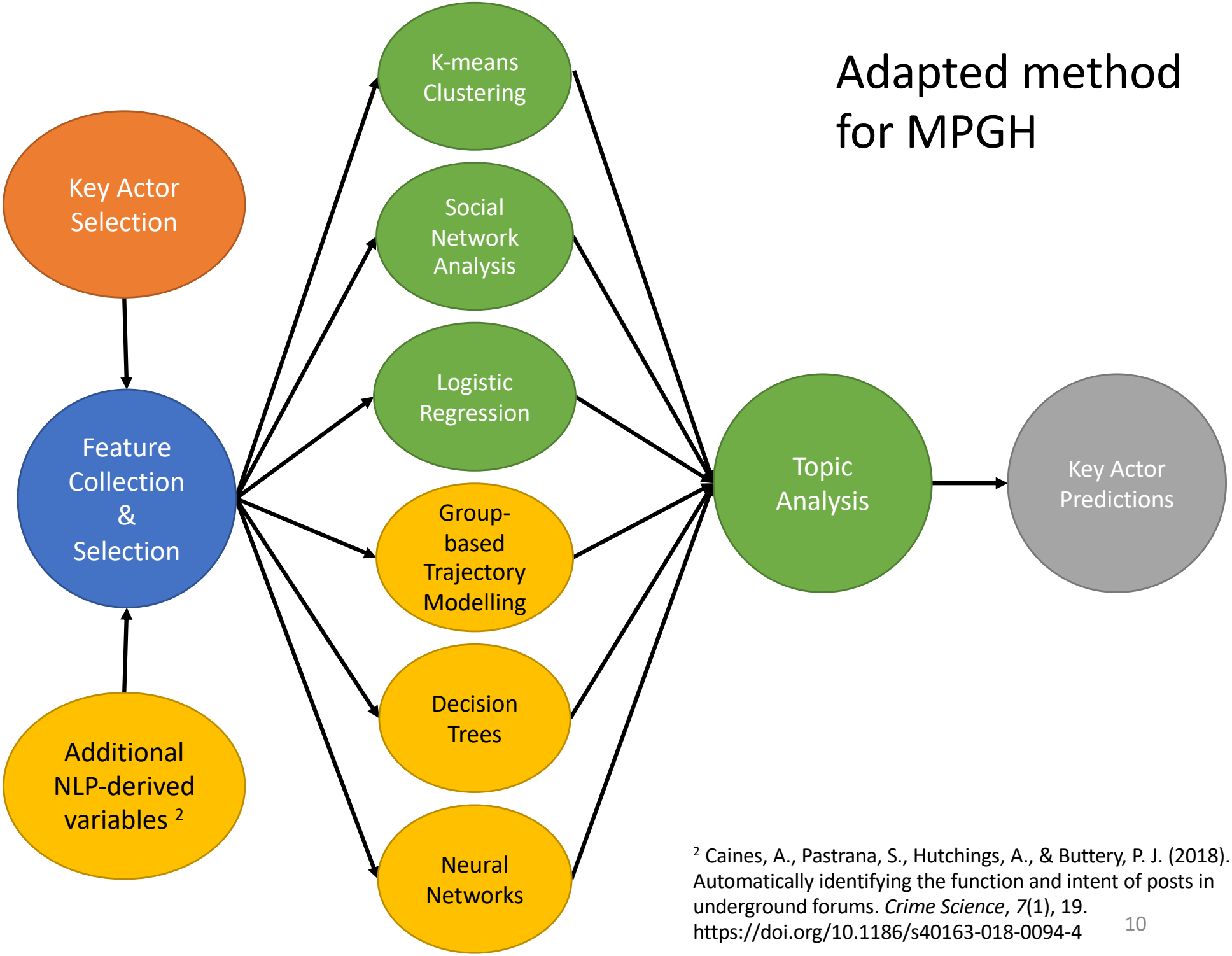
- Snapshot of forum activity
- 764k threads, 9.36m posts, 132k members with >5 posts



Method used by Pastrana et al.



Adapted method for MPGH



[2] Caines, A., Pastrana, S., Hutchings, A., & Buttery, P. J. (2018). Automatically identifying the function and intent of posts in underground forums. *Crime Science*, *7*(1), 19. https://doi.org/10.1186/s40163-018-0094-4

Key Actor Selection

- Manually selected 87 key actors, including:
  - Those who have **released tools and tutorials** on cracking, gaming and hacking forums
  - Those who have advertised **DDoS-for-hire (booter/stresser) services**
  - Those who are **strongly connected** to other key actors, and are involved in similar activities to key actors
- No information relating to any arrests or offending are available for this forum
  - Therefore a manual selection process was used

Feature Collection

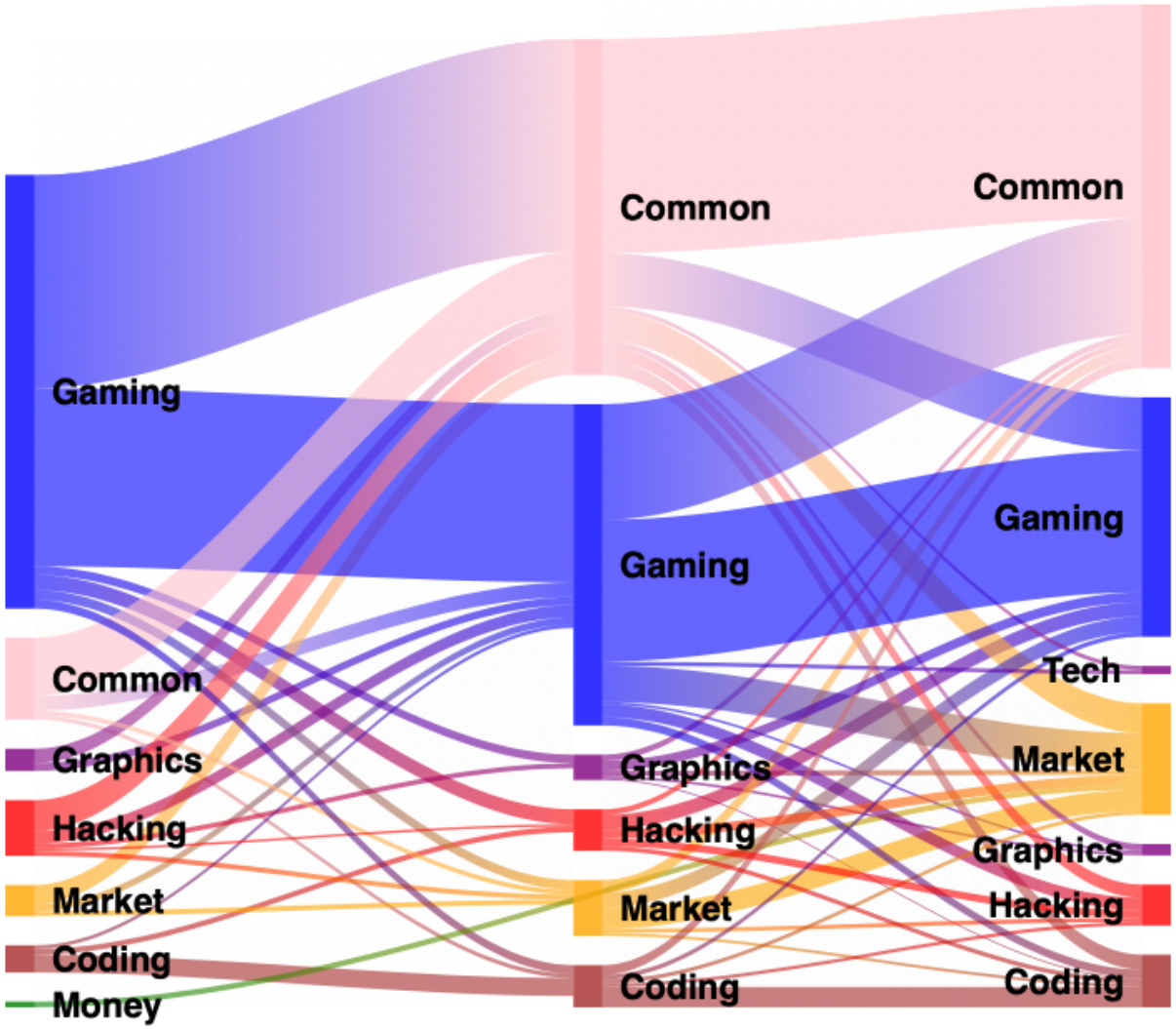
- Initial features include:
  - Social network analysis (eigenvector centrality, …)
  - Activity counts (thread count on marketplace, …)
  - Activity metrics (days spent on forum, …)
  - Interaction metrics (number of citations, …)
  - Impact metrics (h-index, i-10 index, …)
- Additional features from NLP tools include (averaged over user's posts):
  - Sentiment (quantitative measure of emotion)
  - Post types (information request, social, tutorial, …)
  - Post intents (positive, negative, aggressive, …)
  - Addressee types

Feature Selection

- Only members with more than 5 posts (‘active members’) are considered for analysis (~17% of all)
- Features are iteratively removed until correlations are less than 80%
  - Some techniques and analysis rely on low multicollinearity of features
- Features are scaled
  - Some techniques rely on normalised distances of features
- Dataset is split into train-test-validation sets

Key Actor Insights

Changing Interests Over Time



Lifetime of Key Actor on the Forum

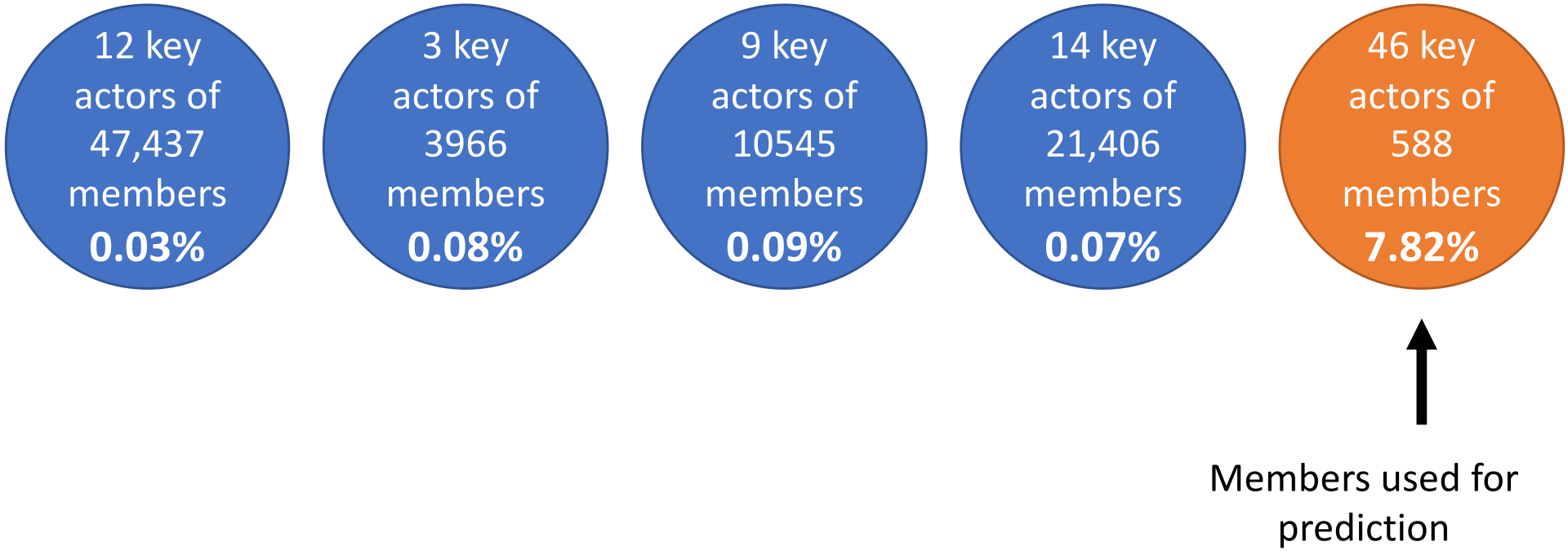
Logistic Regression

		B	S.E.	Wald	Sig	Exp(B)	95% C.I. for Exp(B)	
							Lower	Upper
Step 13	POSTS_CODE	.001	.000	19.598	.000	1.001	1.001	1.002
	POSTS_GAME	.000	.000	20.875	.000	1.000	1.000	1.001
	POSTS_WEB	.146	.060	5.954	.015	1.158	1.029	1.302
	THREAD_GAME	-.013	.004	11.876	.001	.987	.980	.995
	THREAD_HACK	.072	.023	10.209	.001	1.075	1.028	1.123
	THREAD_MARKET	.014	.006	5.746	.017	1.014	1.002	1.025
	CURRENCY_EX	-.034	.022	2.399	.121	.967	.926	1.009
	NUM_REPLIES	.000	.000	14.294	.000	1.000	1.000	1.000
	H_INDEX	.164	.026	41.206	.000	1.178	1.120	1.238
	I_100	-.141	.035	15.909	.000	.868	.810	.931
	Constant	-8.297	.204	1650.400	.000	.000		

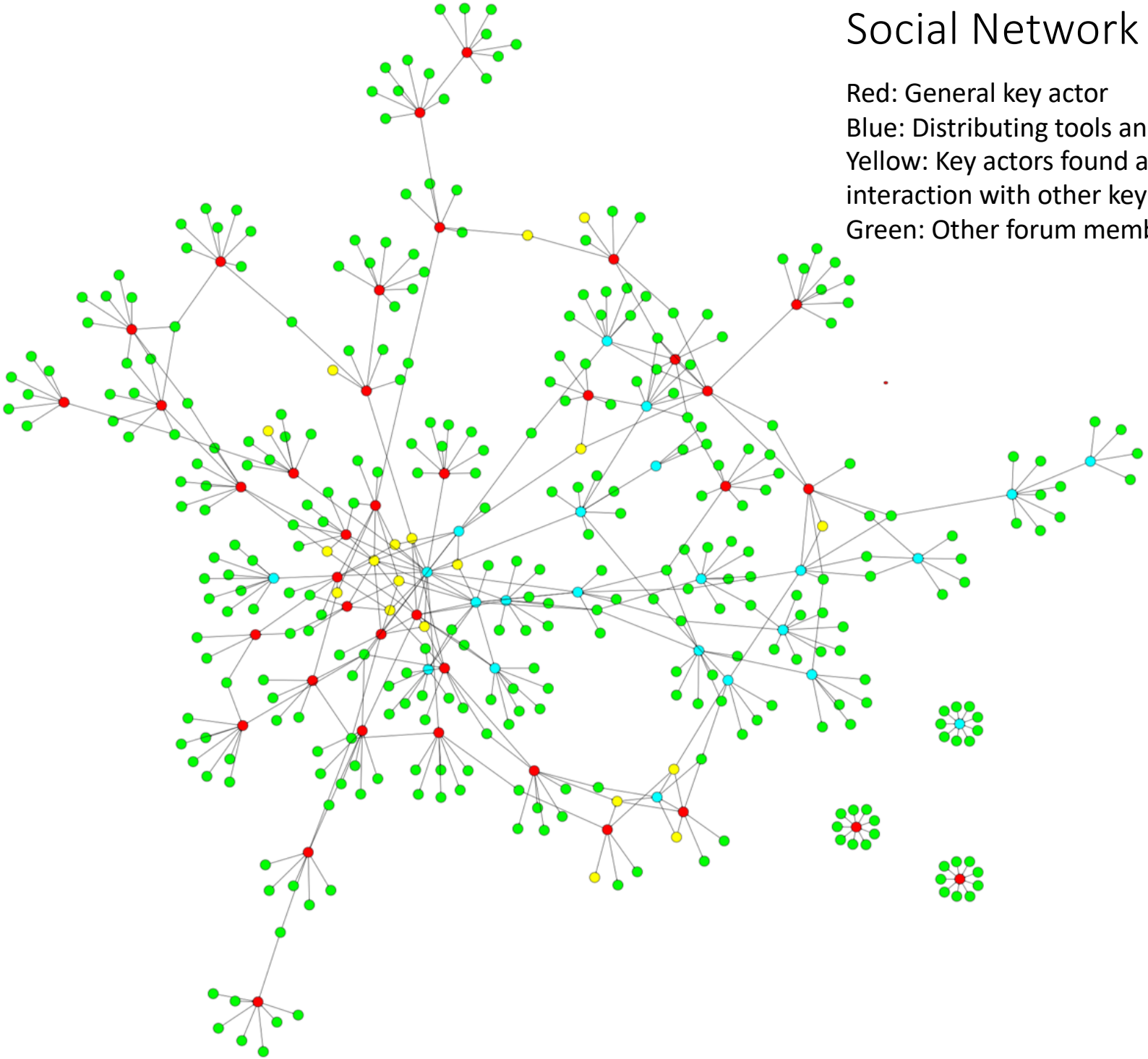
Potential Key Actor Predictions

K-means Clustering (All Members)

- Placing members into (k=)5 groups
- Proportion of key actors per group:

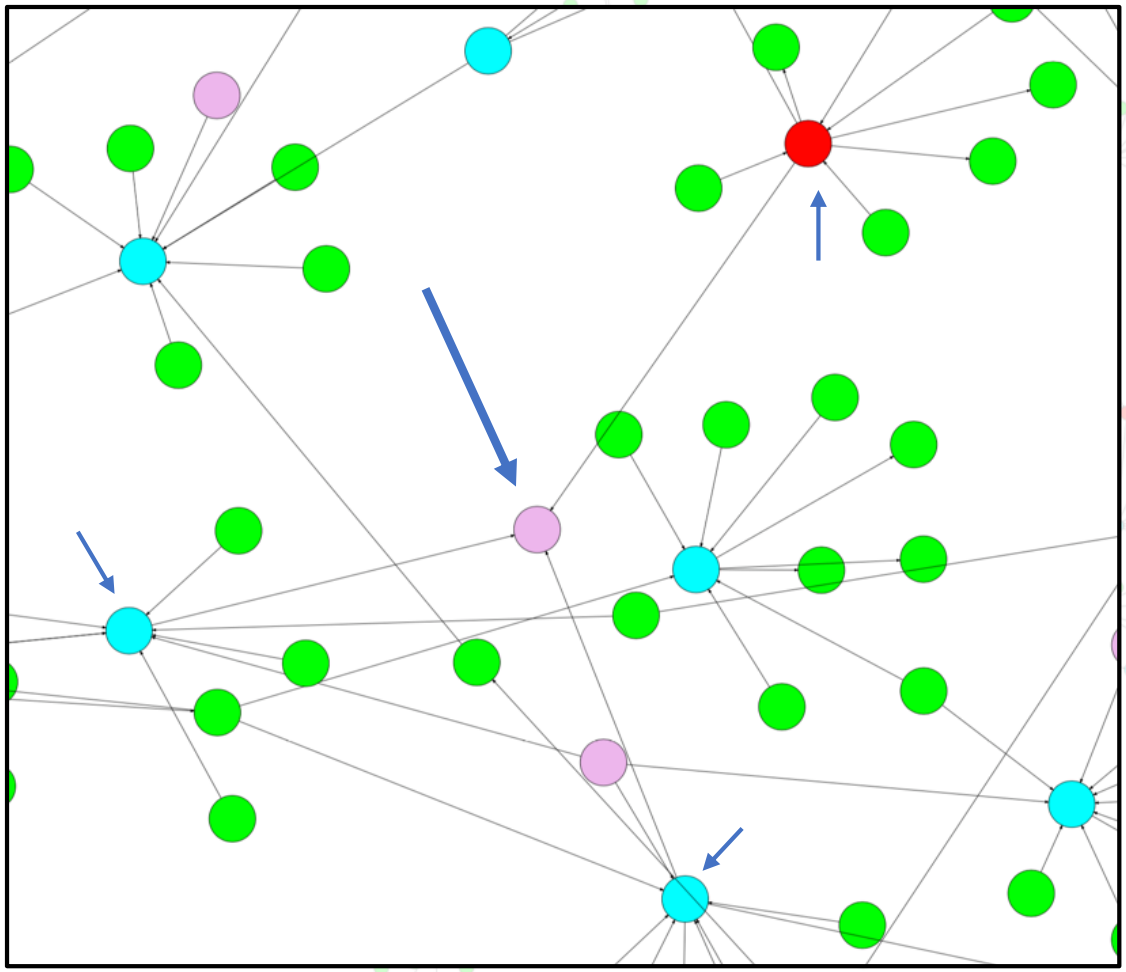


Social Network Analysis



Red: General key actor
Blue: Distributing tools and tutorials
Yellow: Key actors found after interaction with other key actors
Green: Other forum members

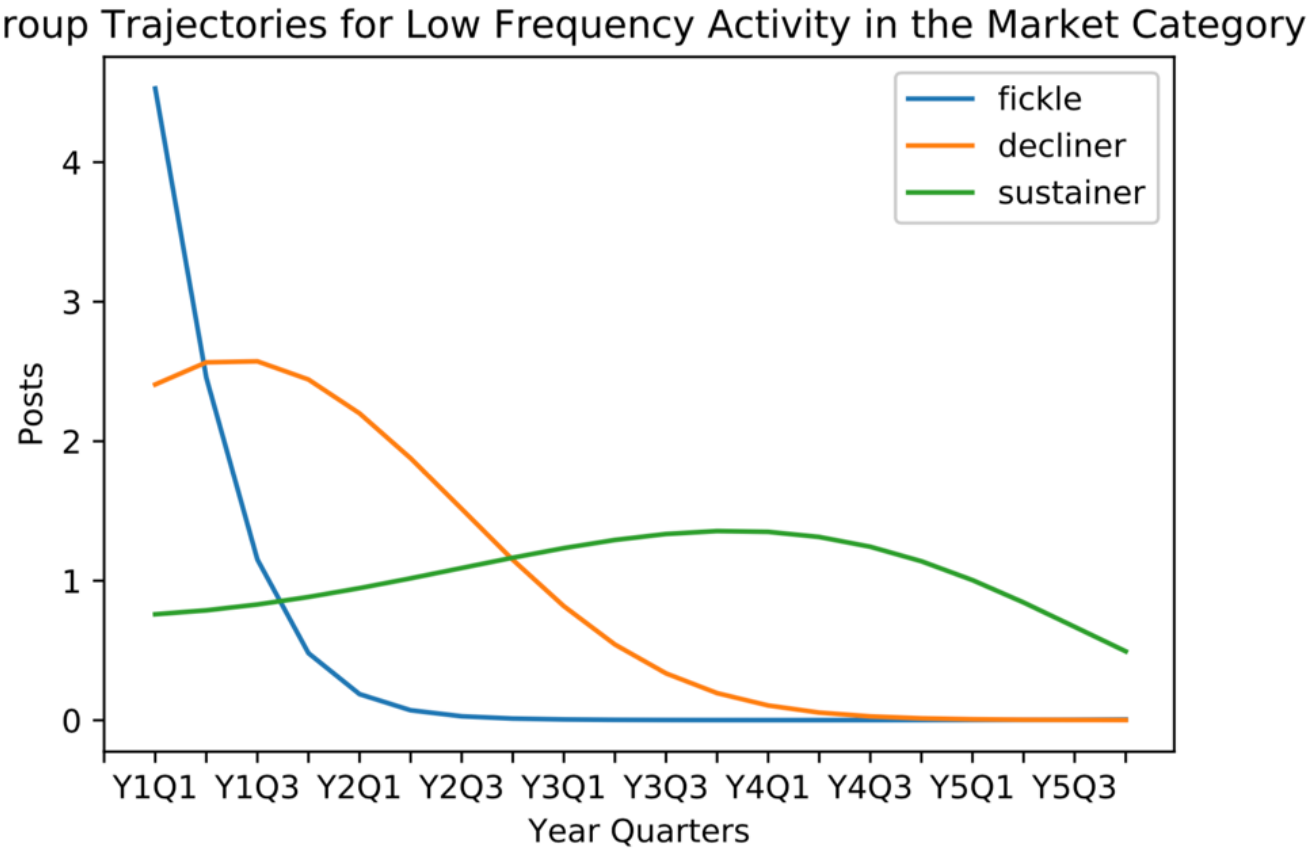
Social Network Analysis



Red: General key actor
Blue: Distributing tools and tutorials
Yellow: Key actors found after interaction with other key actors
Green: Other forum members

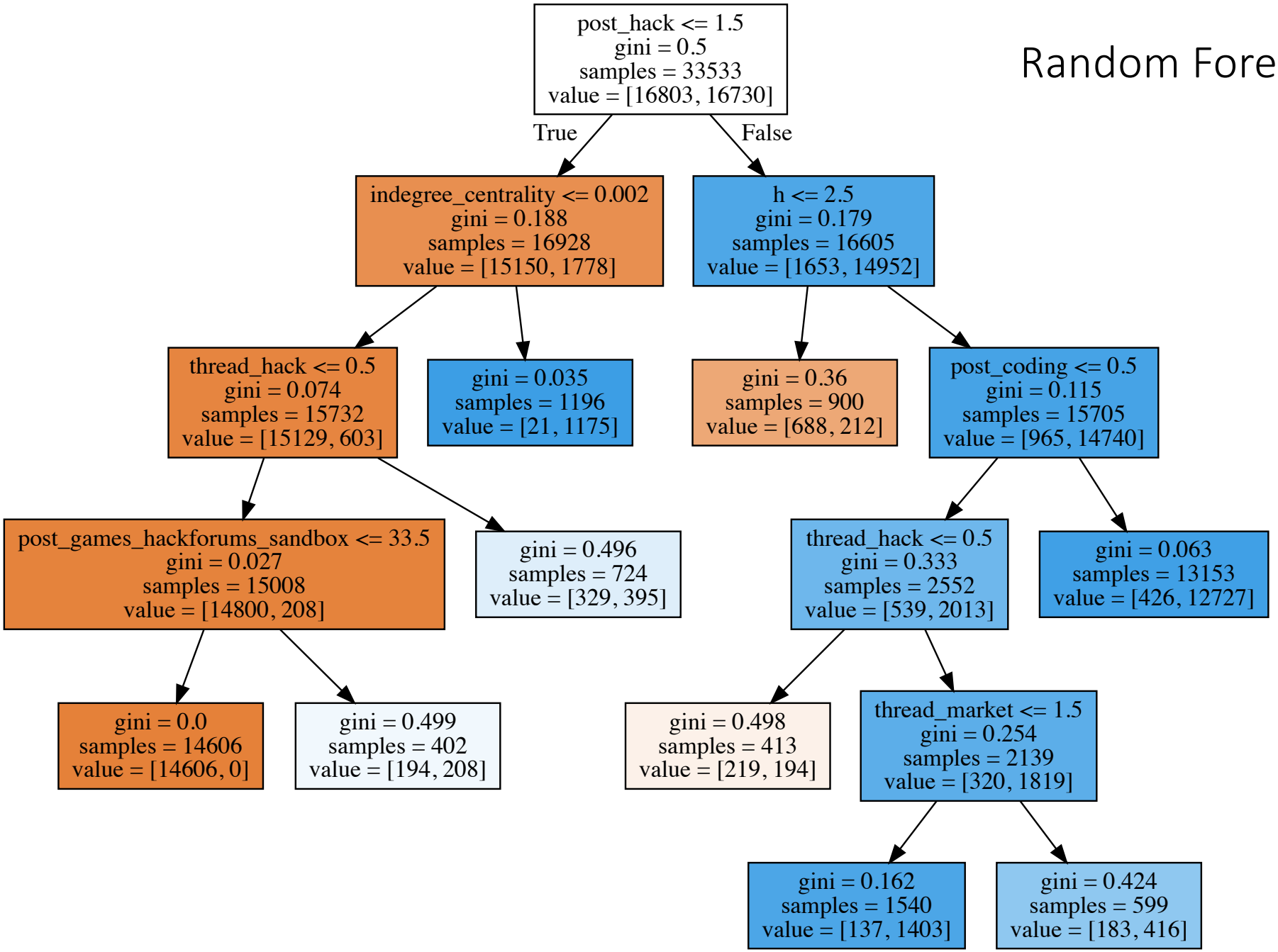
Pink: Predicted key actors

Group-based trajectory modelling

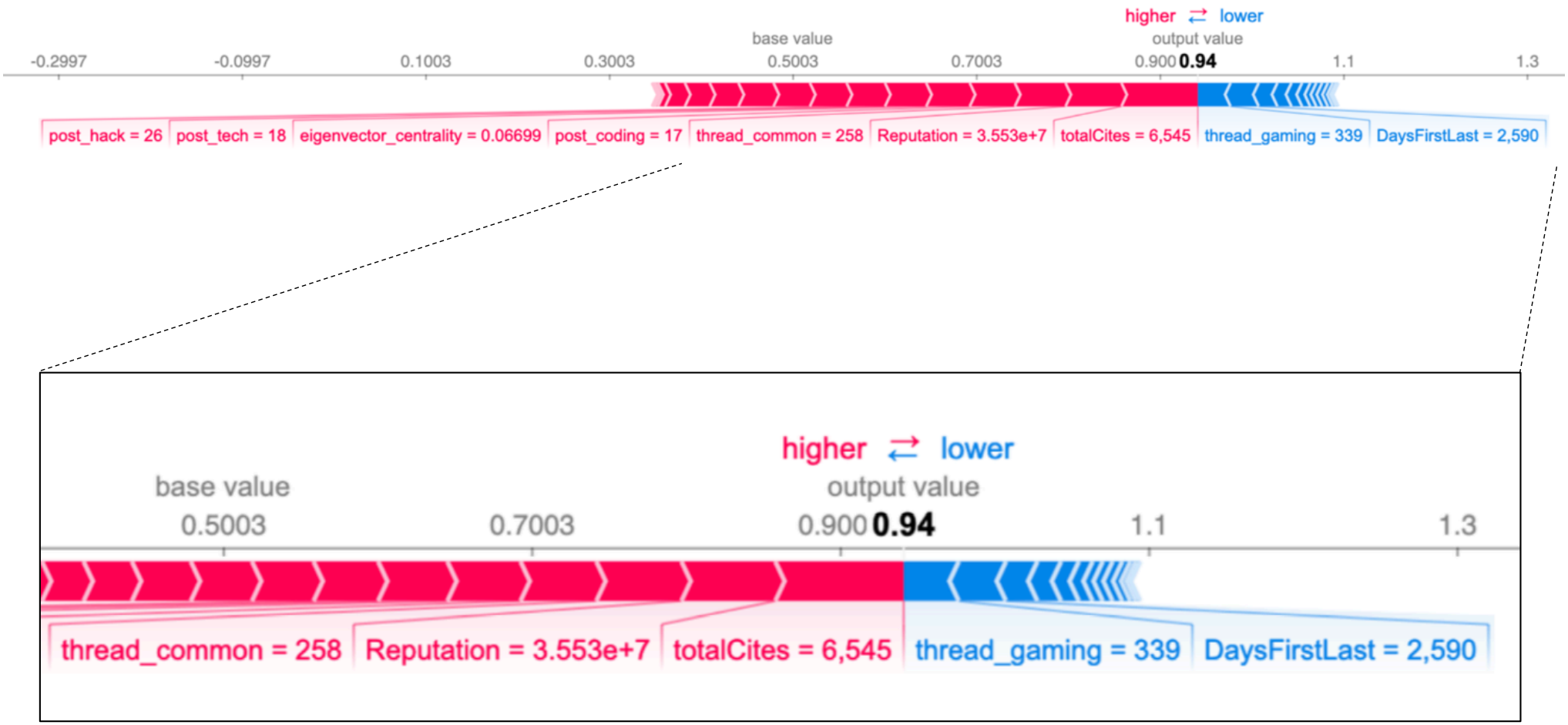


This sustainer trajectory **contains 28% of all key actors**, and is used for prediction

Random Forest



Inspecting Random Forest and Neural Network Models



SHAP diagram explaining the prediction of one member

Topic Analysis

- Computationally expensive to compute for all members, but is used to verify prediction results

one (85), time (85), thanks (84), html (84), help (83), game (83), mpgh (83), work (83), post (82), name (81), thing (81), way (81), **code (80)**, people (80), got (80), thank (79), c (79), lol (79), **hack (78)**, day (78), computer (78), version (78), thread (78), copy (78), **account (78)**, **virus (78)**, man (78), **program (77)**, ip (77), **scan (76)**, section (75), information (75), money (74), pm (74), end (74), **ban (73)**, method (73), **key (73)**, pc (73), part (73), player (73), **password (72)**, image (72), info (71), case (71), **cheat (70)**, **source (70)**, year (70), function (69), **mod (68)**, address (68), service (68), haha (68), class (68), keyboard (67), music (67), build (67), order (67), window (66), browser (66), laptop (65), news (65), card (65), **injector (65)**, weapon (65), contact (65), bump (64), **block (64)**, **aimbot (64)**, paypal (63), skype (63), mouse (63), **hacker (63)**, price (62), skill (61), range (61), flash (60), gun (60), cpu (60), troll (59), gain (59), ram (58), graphic (54), performance (54), market (53), nexon (52), board (51), string (51), trading (50), refund (48), giveaway (48), cooler (46), nx (41), budget (39), predator (38), currency (38), symmetrical (37), bitcoin (34), coin (31), xml (30), integer (27), btc (27), dim (26), eth (24), crypto (20), byval (15), congratulation (15), c++ (12), bch (7), tether (5), usdt (2)

Terms related directly to cybercrime, or to the creation of tools used for cybercrime

Key Actor Predictions

SNA	Clust	LogReg	NLPClust	NLPLogReg	RF	NN	GBTM	Predicted/Total	Avg. Distance	Farthest	Closest
	✓		✓		✓	✓		1/2	0.68	0.64	0.71
	✓		✓		✓		✓	6/9	0.66	0.54	0.79
	✓		✓	✓	✓			2/3	0.58	0.46	0.64
	✓		✓			✓	✓	1/1	0.64	0.64	0.64
	✓		✓	✓			✓	1/2	0.68	0.54	0.82
	✓	✓	✓		✓	✓		5/5	0.58	0.42	0.64
	✓	✓	✓		✓		✓	2/2	0.7	0.68	0.71
	✓	✓	✓	✓	✓			6/6	0.63	0.43	0.79
✓	✓	✓	✓			✓		1/1	0.64	0.64	0.64
	✓	✓	✓			✓	✓	0/1	0.48	0.48	0.48
	✓	✓	✓	✓		✓		4/4	0.67	0.43	0.79
✓	✓	✓	✓	✓				3/3	0.74	0.68	0.79
	✓		✓		✓	✓	✓	4/4	0.62	0.53	0.75
✓	✓		✓	✓	✓			0/1	0.61	0.61	0.61
	✓		✓	✓	✓		✓	3/3	0.69	0.64	0.79
✓	✓	✓	✓		✓	✓		1/1	0.64	0.64	0.64
	✓	✓	✓		✓	✓	✓	0/1	0.68	0.68	0.68
	✓	✓	✓	✓	✓	✓		5/8	0.61	0.5	0.75
	✓	✓	✓	✓	✓		✓	1/1	0.64	0.64	0.64
✓	✓	✓	✓	✓		✓		0/1	0.61	0.61	0.61
	✓	✓	✓	✓		✓	✓	1/1	0.71	0.71	0.71
✓	✓	✓	✓	✓	✓	✓		2/2	0.7	0.64	0.75
7	49	31	49	28	38	25	19				

SNA	Clust	LogReg	NLPClust	NLPLogReg	RF	NN	GBTM
7	49	31	49	28	38	25	19

49 members are predicted as key actors

Summary: Key Actor Behaviour

- Different techniques begin to explain the behaviour of key actors, showing they:
  - Have a higher h-index
  - Have been active on the forum for longer
  - Mostly well-connected with other key actors, and have high eigenvector centrality
  - Sustain low-frequency post activity on the marketplace, and high-frequency post activity in the gaming category

Summary: Techniques

- Techniques should be **combined** to produce better predictions and insights of potential key actors
  - Individual features used for prediction, including *reputation*, are not good indicators of key actors

Wider Context

- Finding common characteristics of key actor activities are useful in understanding behaviours
- These can later be used to identify points of intervention, to deter and prevent individuals from progressing further into cybercrime
- This could include law enforcement activity having a presence on the forum
  - Could include disrupting low-level sustaining activity on the marketplace

Jack Hughes
joh32@cam.ac.uk

Data used is available from the Cambridge Cybercrime Centre:
https://www.cambridgecybercrime.uk/process.html

References

[1] Pastrana S., Hutchings A., Caines A., Buttery P. (2018) Characterizing Eve: Analysing Cybercrime Actors in a Large Underground Forum. In: Bailey M., Holz T., Stamatogiannakis M., Ioannidis S. (eds) Research in Attacks, Intrusions, and Defenses. RAID 2018. Lecture Notes in Computer Science, vol 11050. Springer, Cham

[2] Caines, A., Pastrana, S., Hutchings, A., & Buttery, P. J. (2018). Automatically identifying the function and intent of posts in underground forums. *Crime Science*, *7*(1), 19. https://doi.org/10.1186/s40163-018-0094-4