Automatically Dismantling Online Dating Fraud

University of BRISTOL

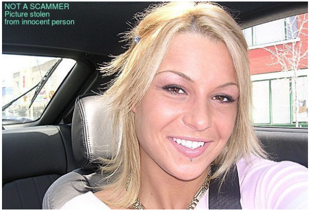
Bristol Cyber Security Group

Matthew Edwards
Guillermo Suarez-Tangil
Claudia Peersman
Gianluca Stringhini
Awais Rashid
Monica Whitty

University of Bristol
University College London
University of Warwick

October 29, 2019

Online Dating Fraud



age 32

location Winona, Minnesota.

occupation beautician

marital status single

age 64

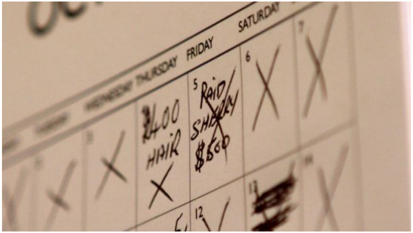
location Richmond, Virginia.

occupation military

marital status widowed

Online Dating Fraud

Emotional Harm



Financial Losses

Motivation

Scammers

- Lie on their profile
- Contact lots of users
- Move conversations off-platform

Real Users

- (White) lie on their profile
- Contact lots of users
- Move conversations off-platform

The (Limited) Data

"[...]our approach to Russian and Nigerian scam is different than on other dating services. We don't just kick them out AFTER they tried to scam somebody. We don't wait for users reports. We simply don't allow scammers to register and contact other singles in the first place." **datingnmore.com**

Here we list scammers who UNSUCCESSFULLY tried to register on our dating site, but got booted **scamdiggers.com**

Approach

Scraped in March, 2017

- **datingnmore.com**: 14,720 ordinary dating profiles
- **scamdiggers.com**: 5,402 scammer profiles

Profiles

- **Demographics**: Categorical information such as age, gender, ethnicity, etc.
- **Images**: One or more images of the user. Users are usually motivated to include pictures that illustrate their hobbies
- **Description**: Short textual self-description from the user, in which they advertise their key traits and interests

How do scammer profiles differ?

Real Profiles

- Average age: ≈ 40
- Average age (m): ≈ 40
- Average age (f): ≈ 40

Scam Profiles

- Average age: ≈ 40
- Average age (m): ≈ 50
- Average age (f): ≈ 30

How do scammer profiles differ?

Table: Profile ethnicities

Ethnicity	Real	Scam
white	0.44	0.66
hispanic	0.32	0.02
other	0.07	0.04
black	0.06	0.06
mixed	0.05	0.07
asian	0.04	0.02
native american	0.01	0.11

Table: Profile marital statuses

Status	Real	Scam
single	0.57	0.51
divorced	0.21	0.14
separated	0.09	0.01
other/none	0.06	0.05
widowed	0.04	0.28
married	0.02	0.00
in relationship	0.01	0.00

How do scammer profiles differ?

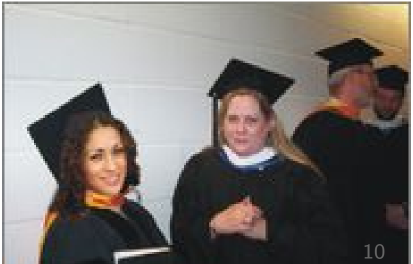
Table: Male profile occupations

Real	Freq	Scam	Freq
other	0.15	military	0.25
self	0.07	engineer	0.25
engineer	0.07	self	0.10
tech.	0.05	business	0.06
student	0.05	building	0.06
retired	0.05	other	0.04
building	0.05	contract	0.04
service	0.04	medical	0.03
transport	0.04	manager	0.02
manual	0.03	sales	0.02

Table: Female profile occupations

Real	Freq	Scam	Freq
other	0.15	student	0.21
student	0.10	self	0.16
carer	0.08	carer	0.10
service	0.06	sales	0.07
clerical	0.06	military	0.05
teacher	0.06	fashion	0.04
retired	0.05	business	0.04
self	0.04	other	0.04
medical	0.04	finance	0.03
housewife	0.03	service	0.03

How do scammer profiles differ?

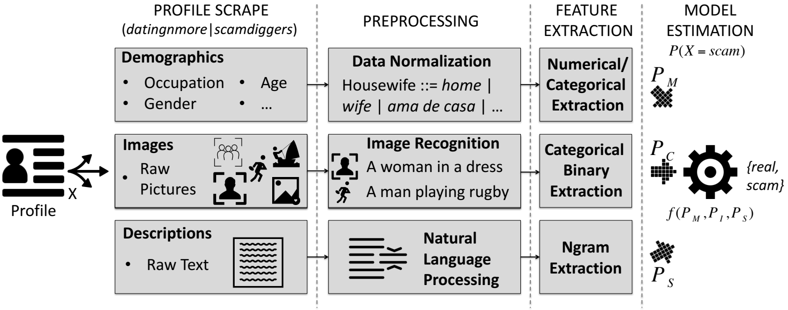


How do scammer profiles differ?

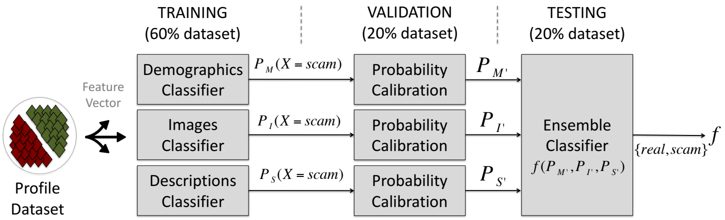
"*[...] To find a woman who loves to dance would be a miracle, but not impossible I believe, even if rare, but definitely a bonus, but not a requirement! lol I also love a man who knows how to treat a woman like a baby and a baby like a woman.*"

- Scammers produce more words per profile
- More reference to emotion – positive and negative – and topics like family
- More formal language, more certainty
- Real users talk more about motives, ambitions, work, leisure

Ensemble classifier



Ensemble classifier



Ensemble classifier

Table: Confusion matrix; precision & recall; and F1-score & accuracy

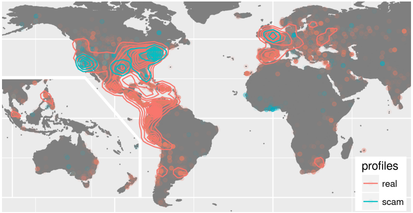
TYPE	CLASSIFIER	TN	FN	FP	TP	PRE	REC	F1	ACC
Individual	demographics	2725	196	149	903	85.8	82.2	84.0	91.3
	captions	2872	499	2	600	99.7	54.6	70.5	87.4
	description	2758	215	116	884	88.4	80.4	84.2	91.7
Ensemble	simple-vote	2870	189	4	910	99.6	82.8	90.4	95.1
	weighted-vote	2834	78	40	1021	96.2	92.9	94.5	97.0

Error analysis

Table: Comparison of overall, validation and false-negative incidence of moderator justifications for scam-classified profiles

Reason	all scams	valid.	fn	rec.
IP contradicts location	3030 (87%)	620 (87%)	44 (85%)	0.93
Suspicious language use	2499 (72%)	507 (71%)	34 (65%)	0.93
IP address is a proxy	2156 (62%)	433 (60%)	25 (48%)	0.94
Known scammer picture	1379 (40%)	299 (42%)	17 (33%)	0.94
Known scammer details	1368 (39%)	284 (40%)	13 (25%)	0.95
Self-contradictory profile	1145 (33%)	242 (34%)	12 (23%)	0.95
IP location is suspicious	968 (28%)	211 (29%)	22 (42%)	**0.90**
Mass-mailing other users	761 (22%)	168 (23%)	10 (19%)	0.94
Picture contradicts profile	261 (7%)	55 (8%)	4 (8%)	0.93

Follow it up



What can we learn about the 'suspicious locations' in online dating fraud? [1]

M. Edwards, G. Suarez-Tangil, C. Peersman, G. Stringhini, A. Rashid, and M. Whitty.
The geography of online dating fraud.
In *Workshop on Technology and Consumer Protection*. IEEE, 2018.

Where Does it Come From?

. . . and why do we care?

- Technical countermeasures at dating sites;
- Targeting campaigns for disruption and prevention;
- Assisting investigations & crime reporting.

The (Limited) Data's Limitations

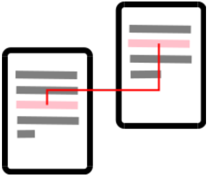
5,194 scam profiles with IP addresses

- Not a large dating site;
- Unknown biases in attraction of scammers;
- Biases in scammer origin identification efficacy.

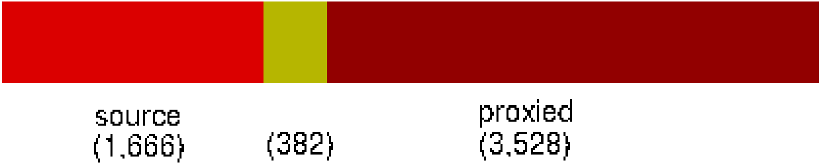
Geolocation & Resource Sharing

We examined **resource sharing** patterns in the scam dating profiles.

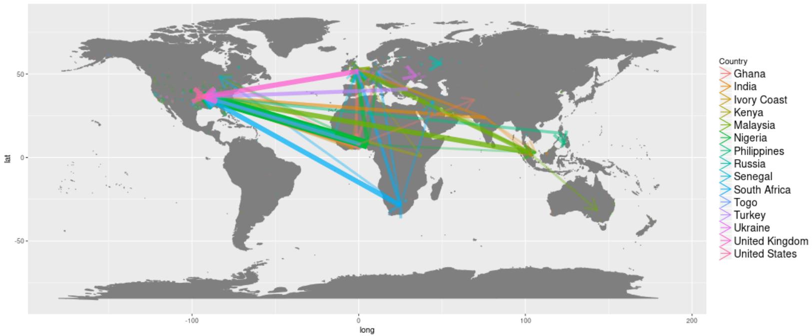
(1) Textual overlap in the descriptions



(2) Perceptual hashing of profile images used on profiles.

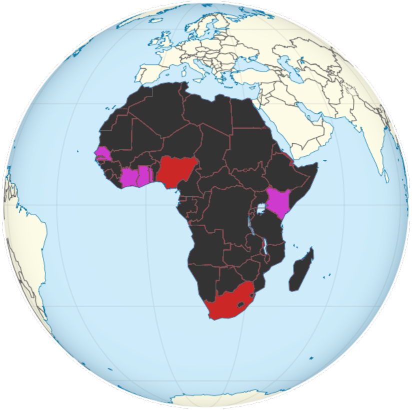


Overview



50
0
-50
lat
-100
0
100
200
long
Country
Ghana
India
Ivory Coast
Kenya
Malaysia
Nigeria
Philippines
Russia
Senegal
South Africa
Togo
Turkey
Ukraine
United Kingdom
United States

A Whistle-Stop Tour: Africa



Nigeria	Largest single origin (30% of all dating fraud). 73% male.
Ghana	Second-largest origin (13%). M/F balance. Honest locations.
South Africa	(4th, 8%). Nigerian model + widows. 81% male.
Togo	80% Female
Senegal	90% Female
Ivory Coast	70% Female
Kenya	Ghanian model?

A Whistle-Stop Tour: Asia

Malaysia	Third-largest origin. 79% male.
India	100% male. 'businessman'.
Philippines	97% female. Mixed-race. US & local presentation.

A Whistle-Stop Tour: Europe

UK	93% male. Divorced, (Many 'in the US').
Turkey	72% male. Similar to South Africa.
Ukraine	93% female. Academic.
Russia	92% female. Accountants.
Italy	89% male. Real estate?

A Whistle-Stop Tour: Americas

USA Nigerian pattern, but very dubious. There is lots of effort at imitating US location, targeting US victims, and a worrying number of the US-located profiles share resources with Nigerian/South African scam profiles.

Final Caution

- Data is inherently limited, and potentially non-representative.
- Scammers motivated to hide their origins.
- Some results are confusing:
  - Surprisingly low numbers from Russia & Ukraine.
  - UK is a global target, but profiles traced here mostly present as in the US.
  - Nigerian diaspora links.

Thank you

Questions & suggestions welcome.

Nation	N	Age		Gender		Occupation			Ethnicity			Marital Status		
		$\bar{x}$	z	$\bar{x}$	z	x	$\bar{x}$	z	x	$\bar{x}$	z	x	$\bar{x}$	z
Nigeria	488	42.61	0.95	**0.73**	**4.13**	military	0.19	1.54	white	0.60	-1.38	single	0.47	-1.73
Ghana	216	40.01	-2.81	**0.46**	**-5.34**	military	0.22	2.07	white	0.68	1.65	**single**	**0.63**	**3.70**
Malaysia	178	**46.53**	**5.33**	**0.79**	**4.29**	**engineer**	**0.30**	**5.71**	white	0.60	-0.86	single	0.46	-1.17
South Africa	140	**48.61**	**6.95**	**0.81**	**4.35**	engineer	0.22	2.15	**white**	**0.77**	**3.57**	**widow**	**0.57**	**7.47**
UK	86	**46.15**	**3.38**	**0.93**	**5.64**	**military**	**0.33**	**4.09**	white	0.66	0.71	**divorce**	**0.33**	**5.51**
USA	57	**47.33**	**3.56**	0.84	3.21	**engineer**	**0.34**	**3.71**	white	0.61	-0.20	widow	0.30	0.18
Turkey	50	46.08	2.53	0.72	1.21	military	0.26	1.82	**white**	**0.86**	**3.48**	**widow**	**0.58**	**4.66**
India	47	42.62	0.30	**1.00**	**5.17**	**business**	**0.32**	**8.14**	white	0.62	-0.14	single	0.53	0.39
Togo	41	**39.44**	**-1.56**	**0.20**	**-5.89**	**military**	**0.37**	**3.52**	white	0.39	-3.20	single	0.68	2.34
Senegal	40	**33.98**	**-4.67**	**0.10**	**-7.07**	**student**	**0.57**	**11.23**	**black**	**0.38**	**7.75**	**single**	**0.88**	**4.81**
Philippines	29	**27.66**	**-7.07**	**0.03**	**-6.75**	**sales**	**0.50**	**14.85**	**mixed**	**0.48**	**7.81**	**single**	**0.97**	**5.14**
Ukraine	28	**29.15**	**-6.11**	**0.07**	**-6.23**	**academic**	**0.22**	**9.82**	**white**	**1.00**	**4.24**	**single**	**0.89**	**4.26**
Russia	24	**29.25**	**-5.72**	**0.08**	**-5.65**	**accounts**	**0.43**	**23.18**	**white**	**0.96**	**3.51**	single	0.79	2.94
Ivory Coast	23	36.52	-2.44	**0.30**	**-3.32**	**student**	**0.30**	**3.86**	**black**	**0.48**	**8.02**	single	0.65	1.48
Kenya	22	35.73	-2.72	0.45	-1.79	**self**	**0.32**	**3.28**	white	0.55	-0.82	single	0.64	1.30
Italy	19	39.37	-1.09	0.89	2.33	**realty**	**0.63**	**23.74**	white	0.89	2.55	**single**	**0.89**	**3.59**
SOURCE	*1666*	*42.13*	-	*0.64*	-	*military*	*0.17*	-	*white*	*0.63*	-	*single*	*0.50*	-