The 70th Annual International Communication Association Conference

USC University of Southern California

I'll Play on My Smurf Account: Sybils and Network Measures

Natalie Jonckheere, Do Own (Donna) Kim, Calvin Liu, Malika Seth, Tanay Shankar, Jinney Guo, Fred Morstatter, Dmitri Williams

We thank Wargaming for providing data access, and Jeremy Ballenger and Eugene Kislyi for their helpful feedback.

USC Annenberg School for Communication and Journalism

USC Viterbi Information Sciences Institute

USC Marshall School of Business

World of Tanks

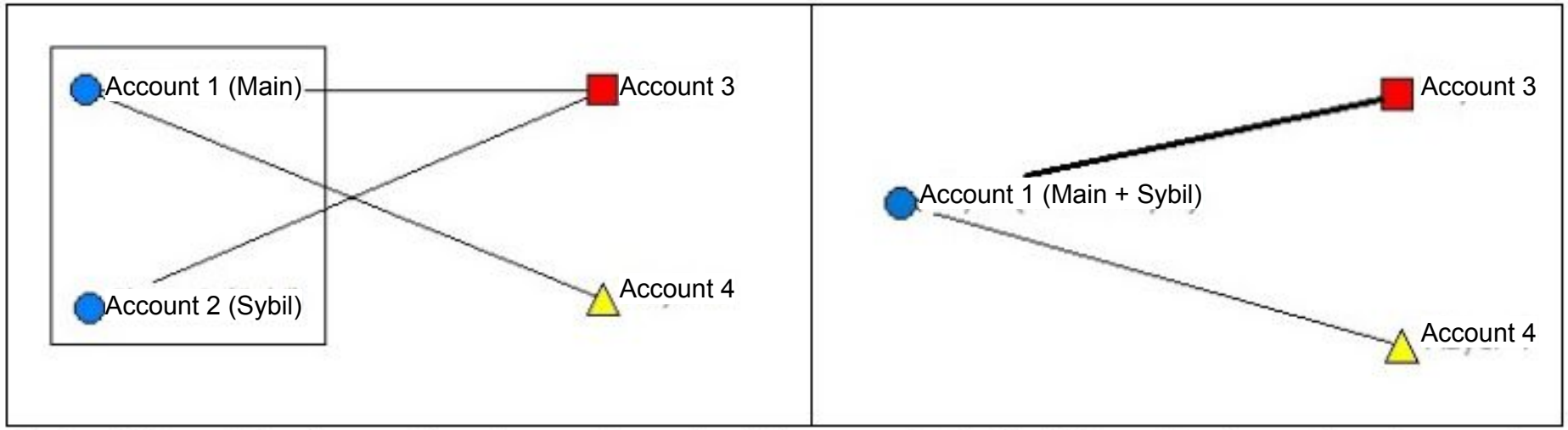
- Massively multiplayer online (MMO) team-based vehicle battle simulator developed by Wargaming.net
- More than 160,000,000 players (worldoftanks.com)
- Various battle modes/ways to connect

Sybils

Sybils

Figure 1

Example of a Sybil Account in a Network

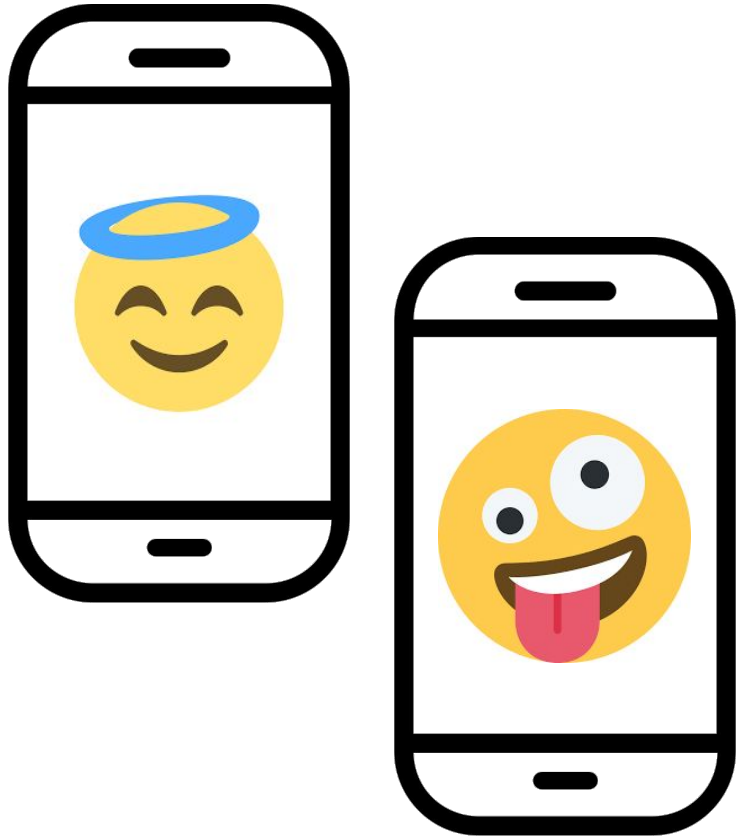


Notes. Network before collapsing main and sybil accounts (left); Network after collapsing main and sybil accounts (right).

Why play with sybil accounts?

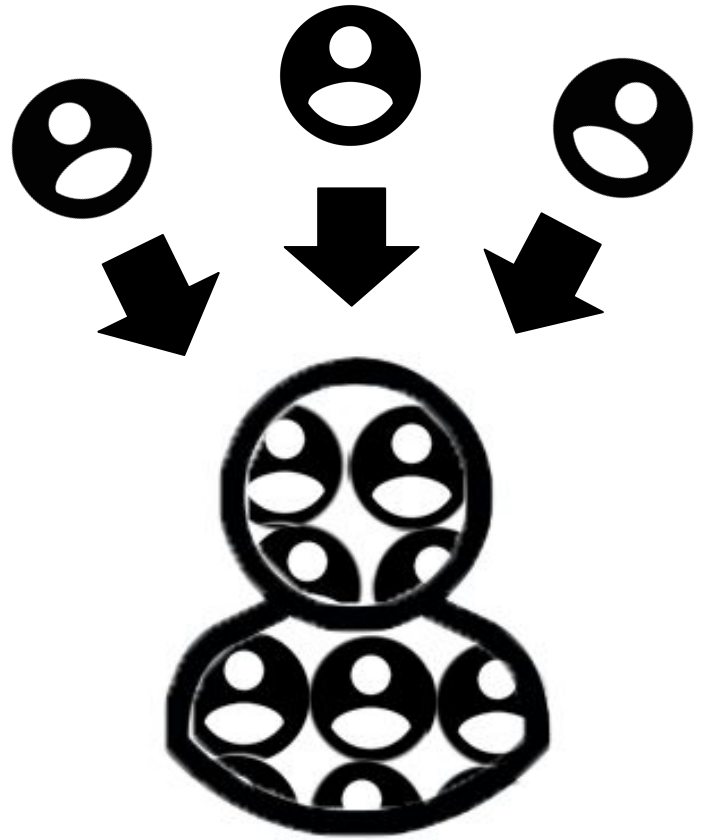
- Impression management
- Privacy
- Accessing different groups
- Gain system benefits

(Aghasian et al, 2017; Bullingham & Vasconcelos, 2013; Goffman, 1978; Kang & Wei, 2019; Marwick & boyd, 2011; Patsakis et al, 2014; Raynes-Goldie, 2010)



Sybils: Extant Approaches

- CS: sybil detection, e.g. friendship networks and clickstreams (Douceur, 2002; Gong et al., 2014; Patsakis et al., 2014; Wang et al., 2013; Yang et al., 2014; Yang et al., 2018)
- Limited exploration on effects and implications when sybils are collapsed
    - Networks based on **each account** → Networks based on **each *individual***



Research Questions

- **RQ1:** What effect does merging sybil accounts have on:
  (a) global player network;
  (b) players' ego networks?

- **RQ2:** What player behaviors predict whether an account is likely to be a sybil account?

Methods

- Wargaming.net server data:
  - a list of **sybil accounts** (using a robust proprietary methodology based on behaviors and real-world data)
  - a list of anonymized IDs for players and other players they had *battles alongside*, including *how many times* they had *actively chosen* to play alongside each other player

  ➔ transformed into **“co-play” data**

Datasets Used

Co-play Data

	SPA1	SPA2
0	00001DD103717A29AE6EAF27BC0E5DD8E03F0CAD177F2A...	0048E013B0F7958ACC52E538699ED1DE84D18E2B9620E9...
1	00001DD103717A29AE6EAF27BC0E5DD8E03F0CAD177F2A...	00DF9BA062D72A54D3C5688ED684078920922BC2583BA8...

Sybil Data

	PARENT_SPA_KEY	CHILD_SPA_KEY
0	658CDFD65103B78226828BA1EF405150A3615CF13DE410...	55250F[illegible]219DDB757D7FBD9AFB6FBB249EFAF69E9DE4A...
1	75CEA7828759FE157E838A4E8EE64190876AC50257825B...	35FE10DE6F56666254FC45D7B86D3793780723F68F53B6...

Global Network Analysis

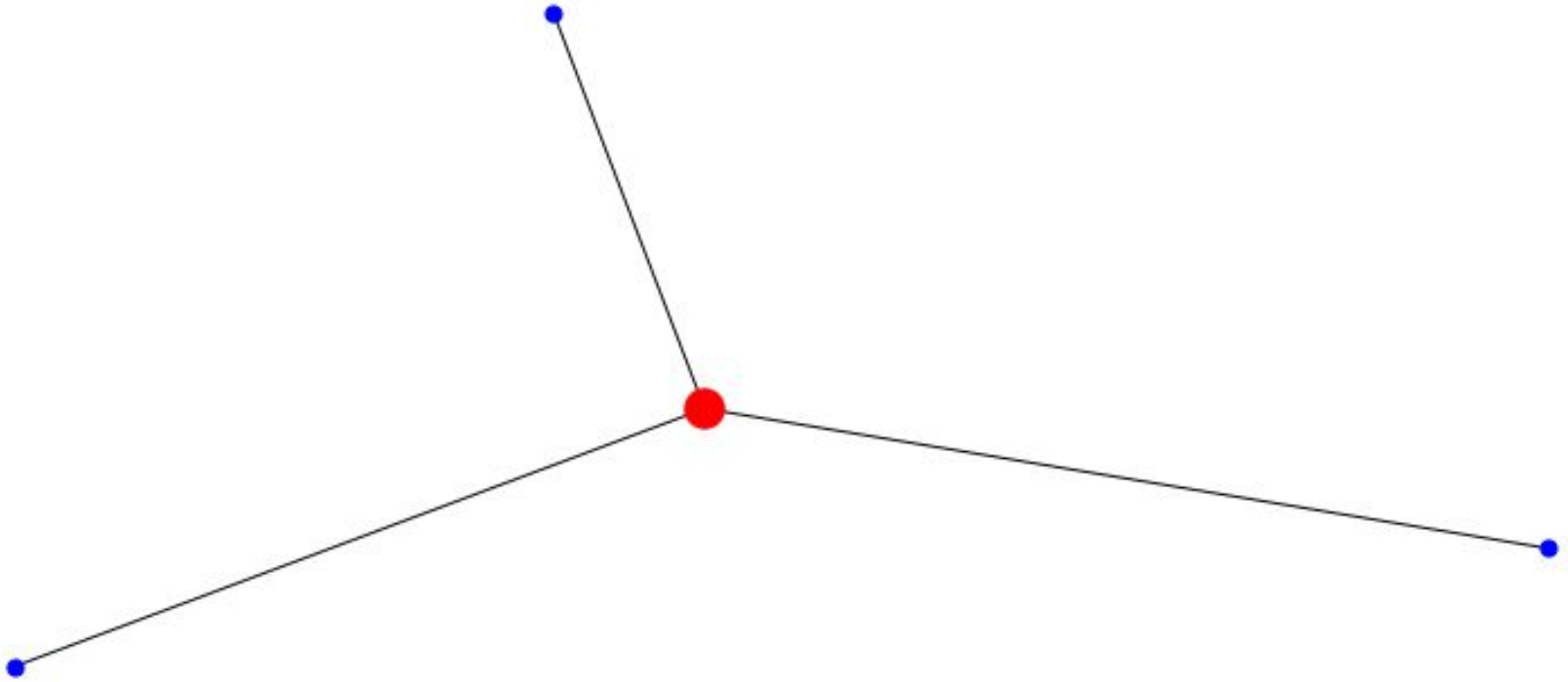
➔ Collapsed co-play network by merging sybil nodes.

Statistic	Pre-Merge	Post-Merge	Percent Change
Vertices	169,149	159,787	-5%
Edges	4,233,190	4,202,663	-0.7%
Global Clustering Coefficient	0.06893263	0.06811161	-1.2%
Density	0.0002959112	0.000329211	9.2%

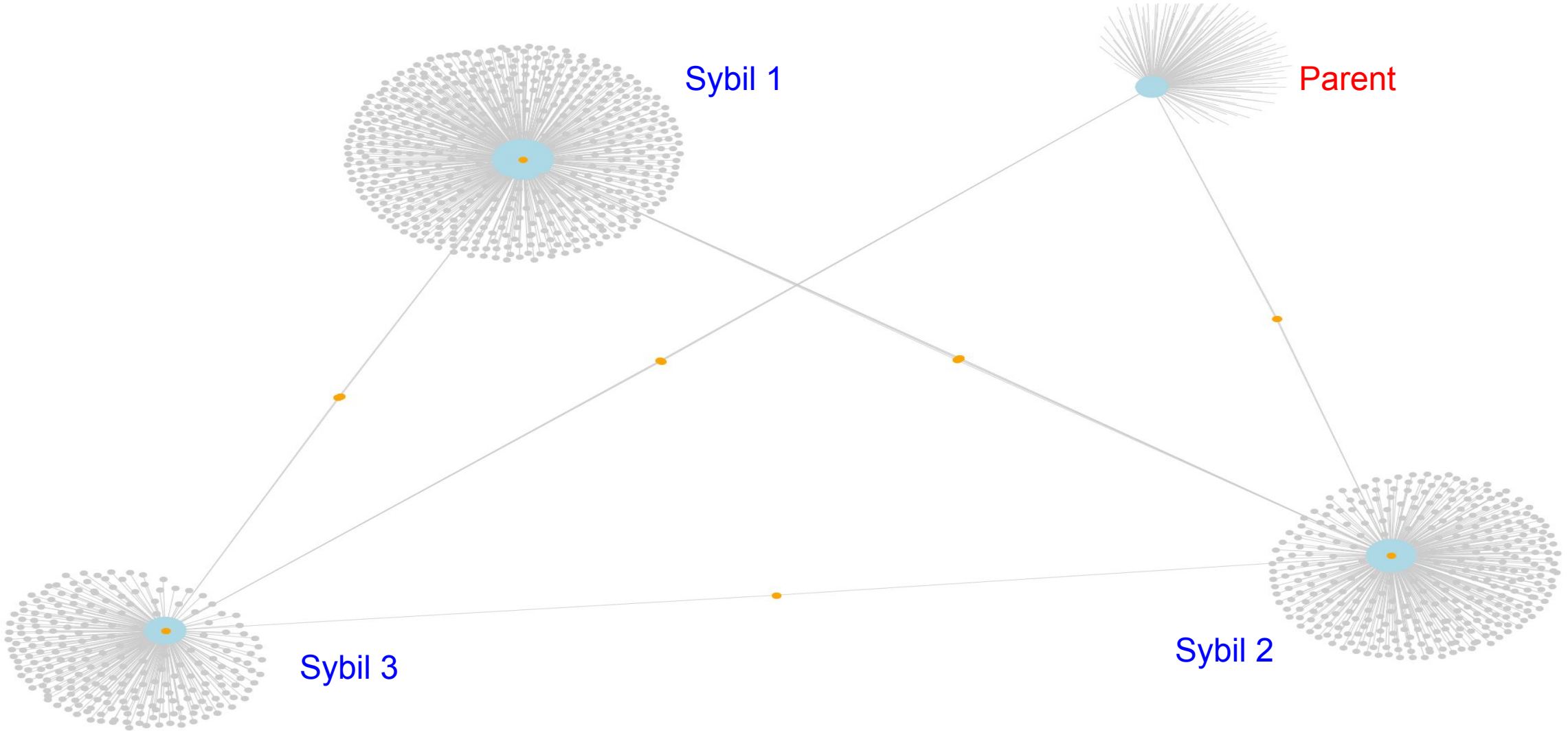
➔ Disproportionate increase in density - low overlap between the ego networks of sybil accounts shared by one person.

Ego-Level Network Investigation

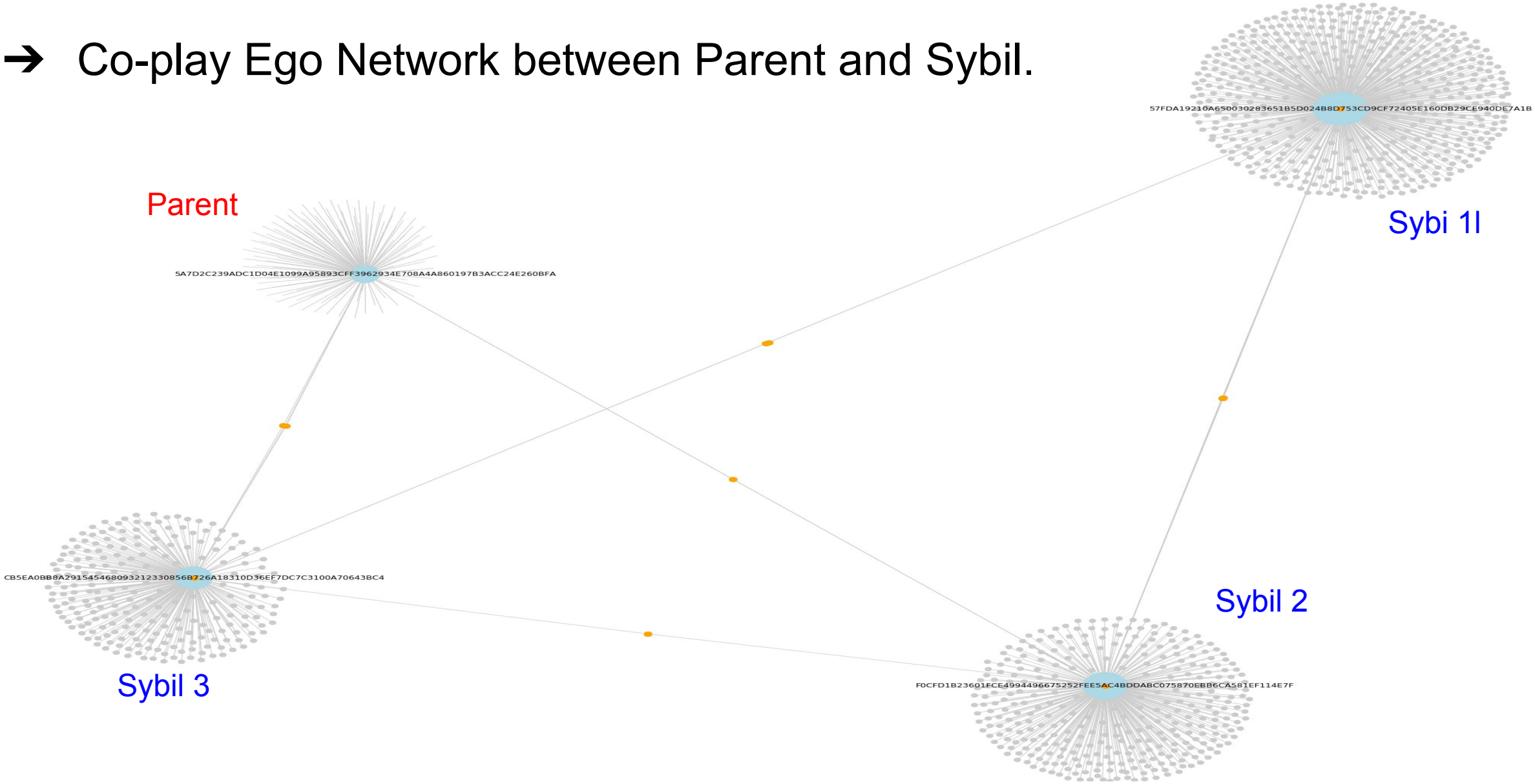
➔ Graph of Parent (red) connected to 3 Sybil accounts(blue).



➔ Co-play Ego Network between Parent and Sybil.



➔ Co-play Ego Network between Parent and Sybil.



Ego-Level Network Investigation

	Degree (before collapse)	Local Clustering Coefficient (before collapse)	Degree (after collapse)	Local Clustering Coefficient (after collapse)
Parent	143	0.014	1216	0.013
Sybil 1	511	0.04		
Sybil 2	339	0.011		
Sybil 3	236	0.012		

Behavioral Dynamics of Sybil Accounts

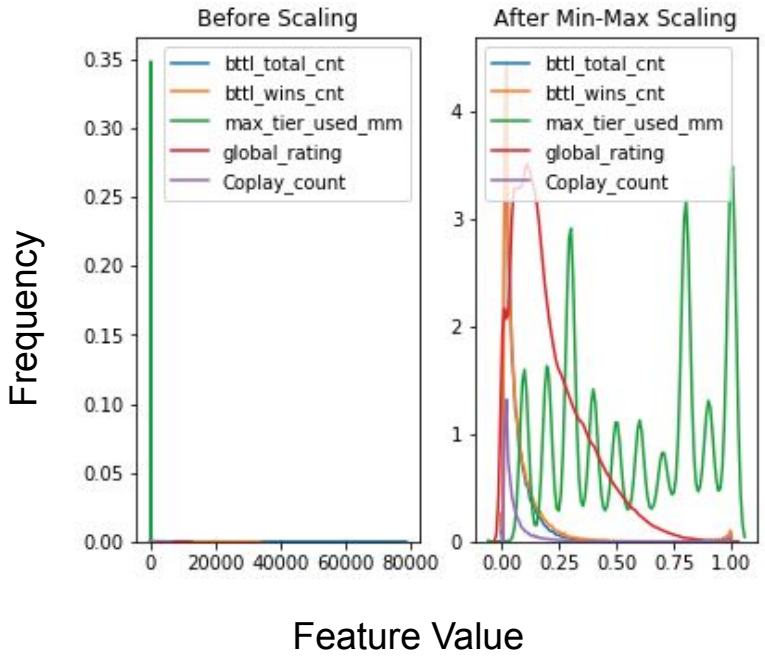
Features -

spa_key_sha256	bttl_total_cnt	bttl_wins_cnt	max_tier_used_	global_rating	Coplay_count	sybil_status
72A02EA50BD7...	18	9	3	74	72.0	0.0
3F0CAD177F2A...	7460	3656	10	4574	66.0	1.0

Behavioral Dynamics of Sybil Accounts

Imbalanced Dataset- SMOTE

- ➔ Number of Regular accounts in training data = 1,051,919
- ➔ Number of Sybil accounts in training data = 187,227
- ➔ After oversampling, number of Regular accounts = 789,006
- ➔ After oversampling, number of Sybil accounts = 789,006



Feature Scaling -

- ➔ MinMax Scaler - Transform all features into a uniform scale between 0-1

Behavioral Dynamics of Sybil Accounts

Algorithm - Random Forest Model

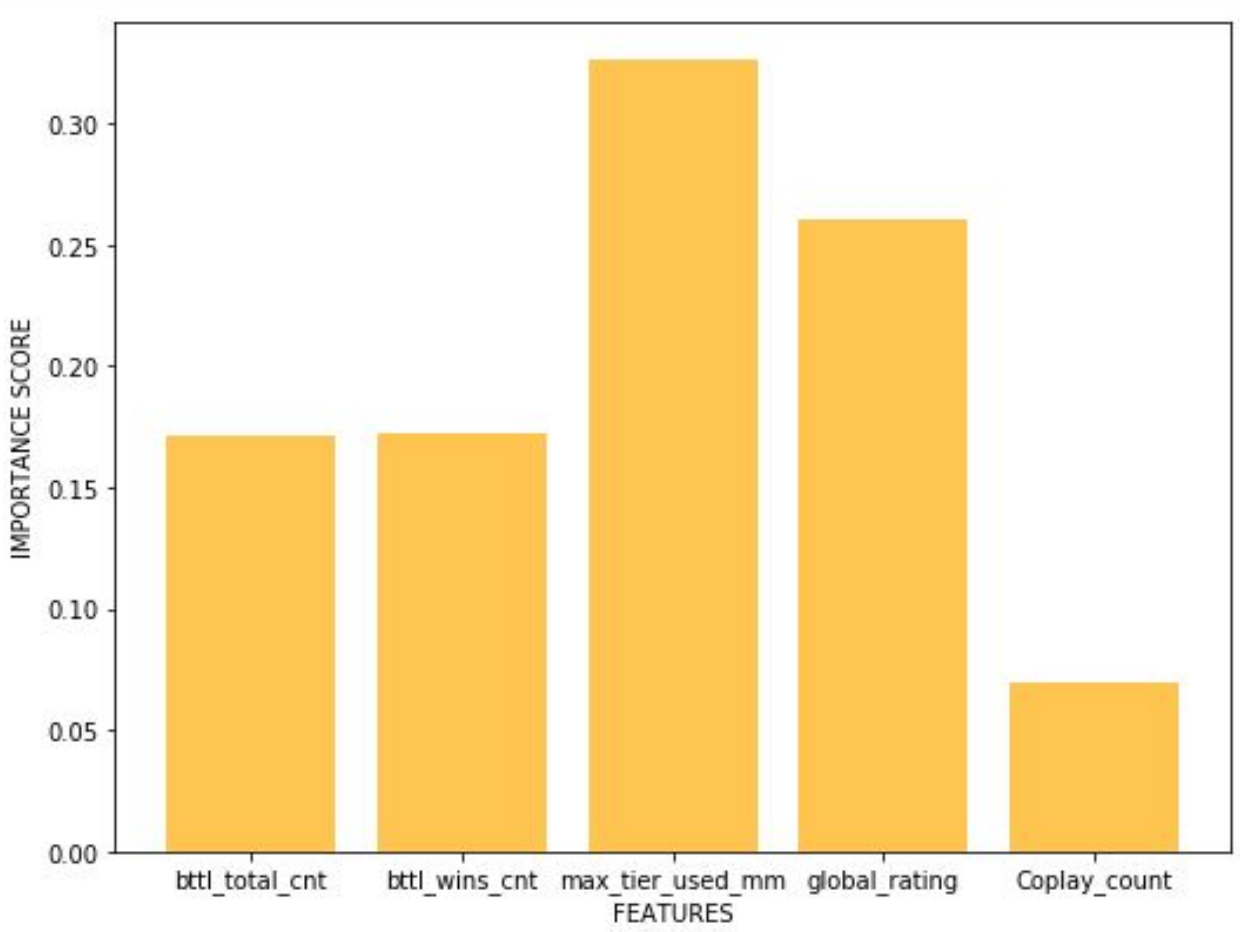
- ➔ Class 0 - Regular Account
- ➔ Class 1 - Sybil Account (Parent/Child)

	precision	recall	f1-score	support
0.0	0.9524	0.9711	0.9617	315712
1.0	0.8168	0.7267	0.7691	56032
accuracy			0.9342	371744

- ➔ Root Mean Squared Error: 0.25643

Behavioral Dynamics of Sybil Accounts

Feature Importance-



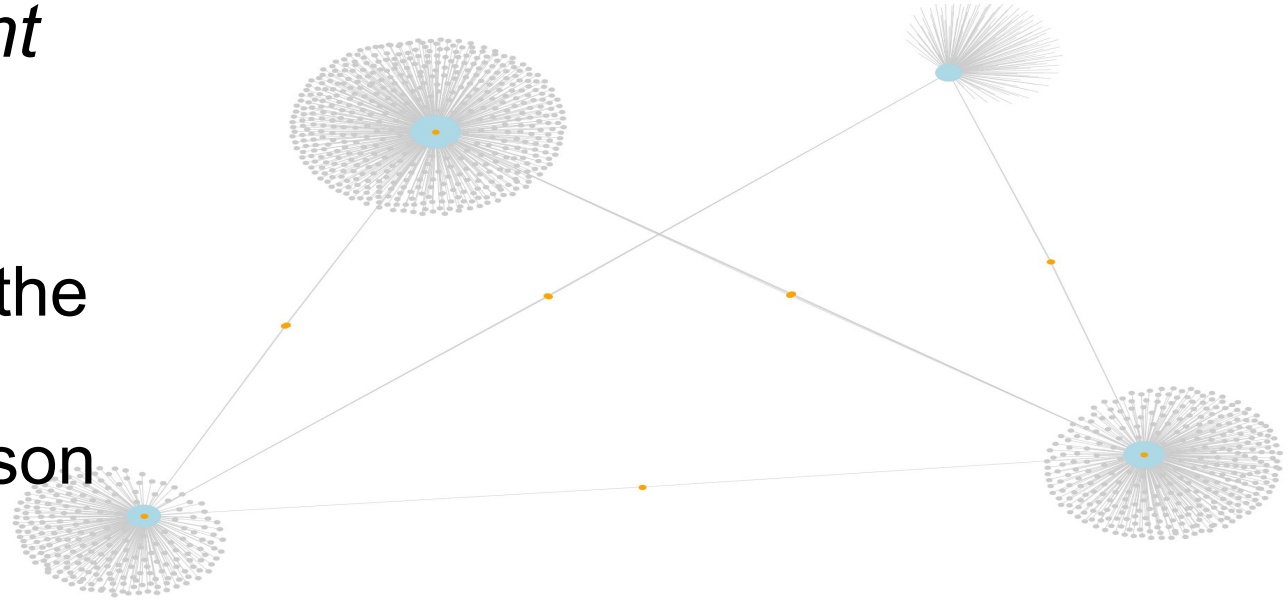
Implications/Future Directions

- Low overlap between parent and child sybil accounts.
  - Active context management: intentional “division of situations” (c.f. Rinsta/Finsta; Meyrowitz, 1985)

Statistic	Pre-Merge	Post-Merge	Percent Change
Vertices	169,149	159,787	-5%
Edges	4,233,190	4,202,663	-0.7%
Global Clustering Coefficient	0.06893263	0.06811161	-1.2%
Density	0.0002959112	0.000329211	9.2%

Implications/Future Directions

- Brokers may NOT be brokers per se, but brokers for *different* accounts of the *same* person
  - Brokerage/bridge *between* sybils (already connected to the same individual)
  - May be sybils of another person
  - May be individuals from overlapping contexts



Implications/Future Directions

- Network size management
  - Desire to keep smaller core network (parent account)
  - “Moving on” to reinvent self

	Degree (before collapse)	Local Clustering Coefficient (before collapse)	Degree (after collapse)	Local Clustering Coefficient (after collapse)
Parent	143	0.014	1216	0.013
Sybil 1	511	0.04		
Sybil 2	339	0.011		
Sybil 3	236	0.012		

Implications/Future Directions

- Features: mastery over familiarity?
  - Mastery: maximum tier, global rating > battle win
  - Familiarity: battle count, maximum tier
- Features: relative less importance of the # co-plays
  - Managing core networks: social groups are predictive, but quality trumps quantity (c.f. Dunbar's number)
  - Performance-oriented: playing/doing *better* is more important than playing/doing *with*