Mapping Influence and Deception

Guido Zarrella

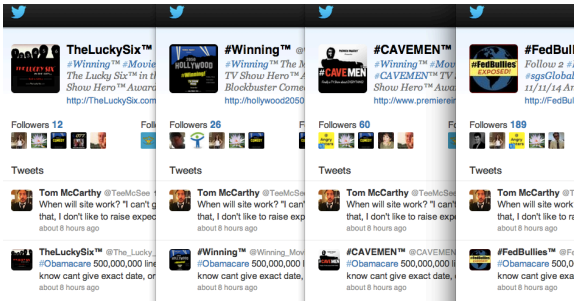
"On the internet, no one knows you're a robot!"

Problem

Identity deception is commonplace in social media. Large scale sockpuppet operations, "astroturf", and covert information campaigns bias any attempt to observe the attitudes and intentions of populations around the globe. Phishing scams and links to malware spread online from users disguised as trusted sources.

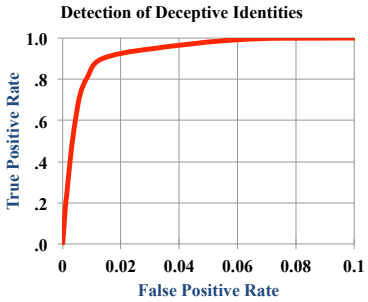
Idea

We use novel data mining techniques to detect and mitigate adversarial influence, to track the spread of ideas propagating through social networks, and to model the influences impacting the behaviors of large populations of interest.



Findings

Our *Pinocchio* software analyzes millions of authors to automate detection of deceptive identities. On Twitter, accounts we flag are 2x more likely to eventually be suspended for violating ToS. Twitter detects only 1/20th of our flagged accounts, which can make up >15% of traffic on a given topic.



Impacts

We annotate and filter data upstream of traditional analysis efforts, decluttering the social radar to enable improved intelligence gathering and strategic communication. By isolating dubiously sourced content, it is now possible to passively observe cyber adversaries' messaging campaigns.