Link Prediction on Hacker Networks

Samantha Lee
Given a network with links in one time step, can we predict the links in the next time step?

---

Given a network of hackers joined by links if they attacked the same victim, can we predict if a hacker will be linked with another hacker in the next time step?
Why do you care about it?

- Application of network science
- Increase understanding of attacker behavior – if a hacker is linked to another hacker, they could be sharing information
- Identify attackers who will collaborate
- Take preventive / offensive action
The Data

---

Zone-H http://www.zone-h.org/ cybercrime archive

10 years of data: 2005–2014

Total of 9,528,165 defacements

Website defacement is an attack on a website that changes the visual appearance of the site or a webpage.
466,226 defacements

A: country of attacker
B: Date defacement was entered into Zone-H
C: hacker name
D: victim website
E: victim IP address
F: victim operating system
G: victim server/software
H: reason for attack
I: Type of vulnerability
J: mirror site address with actual images
K: type of defacement - mass or regular
L: did the attacker re-deface the same site?
M: Published or not.
N: primary or secondary
O: unique defacement ID
P: Physical location of the server that was defaced.
Interesting Attributes
---

A: country of attacker
B: Date defacement was entered into Zone-H
C: hacker name
E: victim IP address
H: reason for attack
P: Physical location of the server that was defaced
2005 Network
2006 Network
Link Prediction - who is linked next?

---

Given: Graph $G = (E, V)$; edge set $E$; node set $V$; $e = (u, v) \in E$ with $t(e)$ timestamp of the edge

Let $G[t, t']$ be the subgraph with edges $t < t(e) < t'$

Output: Edges not in $G[t_0, t'_0]$ predicted to be in $G[t_1, t'_1]$ where $t_0 < t'_0 < t_1 < t'_1$

Note: core nodes
\[ G = (V, E) \]

\[ \text{t0 to t0'} \]

\[ \text{link prediction} \]

\[ \text{t1 to t1'} \]
3 kinds of features to say how “close” 2 nodes are:

---

Network factors (independent of actual data)

Domain specific: Country, Reason for attack, Victim ip addresses & location of victim servers

Individual attributes: total number of defacements
More on Network factors

- Shortest path distance
- Common neighbors
- Jaccard's coefficient
- Preferential attachment
- Katz
- Hitting time
- Page rank
To Predict Links: My Features

- Common neighbors
- Country of the hacker or hacker group - hackers from the same country have a higher probability of linking
- Reason for attack - common goal unites
Upcoming Work

Use subsets of features to do link prediction.

Break my data into two sets, use the first part to try link prediction, use the second part to see how it went.
Sources

https://en.wikipedia.org/wiki/Website_defacement
http://be.amazd.com/link-prediction/
The link-prediction problem for social networks. David Liben-Nowell, Jon Kleinberg
http://www.siam.org/meetings/sdm06/workproceed/Link%20Analysis/12.pdf
Cytoscape
Questions?