

# Social Network Analysis

## Modeling Attacks and Visualizing Privacy

- ➔ IFIP/Primelife'09, Nice, France — 09.09.2009
- © Dominik Birk, Felix Gröbert, Christoph Wegener
- ☎ [felix@groeibert.org](mailto:felix@groeibert.org)
- ✂ [creativecommons.org/licenses/by-nc-nd/3.0/de](http://creativecommons.org/licenses/by-nc-nd/3.0/de)

# Value of Data

Information	Price in US\$
Internetbanking Login Credentials	10 - 1000
Custom Banking-Trojan	1000
E-Mail Addresses	~ 0,35 / MB
Creditcards with CCV2	0,50 - 12
Cash-out Service	8-50% share
Identity Profiles	0,90—25

Sources: Panda Labs: The Business of Cybercrime, Symantec Report on the Underground Economy

*why would you buy an  
identity profile?*

# ... to do identity theft!

- Identity theft
  - ❖ Doing fraud in the name of someone else
- Types of ID theft:
  - ❖ Financial ➔ existing & new accounts
  - ❖ Criminal
  - ❖ Medical
  - ❖ ...

# Results of ID Theft

- Illegal physical access
  - ❖ Immigration
  - ❖ Company-site
- Illegal virtual access
  - ❖ Bank, shopping accounts
  - ❖ Communication accounts: E-Mail, IM, Web 2.0
- Blackmailing and other reasons to impersonate others

*so,  
why would you buy an  
identity profile*

*when you can get it  
for free on the internet?*

# Our Attack

- Conduct ID fraud in an automated fashion
  - ❖ Using Social Network Site (SNS) data
  - ❖ To establish trust
  - ❖ And lure victims into reacting on a message
- Divided into four phases:
  - ❖ Phase 1: gather data on many victims
  - ❖ Phase 2: correlate data to refine it
  - ❖ Phase 3: select potential victims based on data
  - ❖ Phase 4: attack victims using sophisticated phishing messages

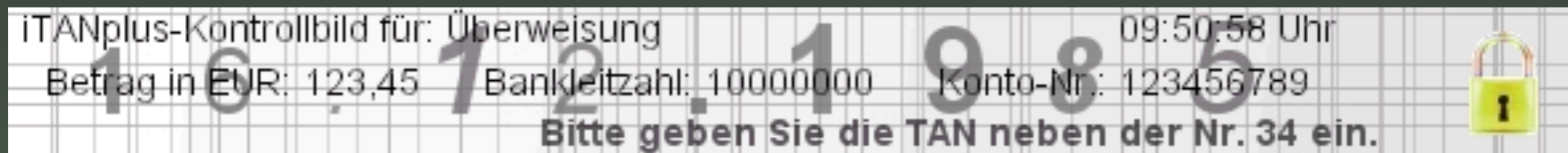
# Phase 1: Data Acquisition



- Parallel methods:
  - ❖ Crawling
  - ❖ Cracking
  - ❖ Buying

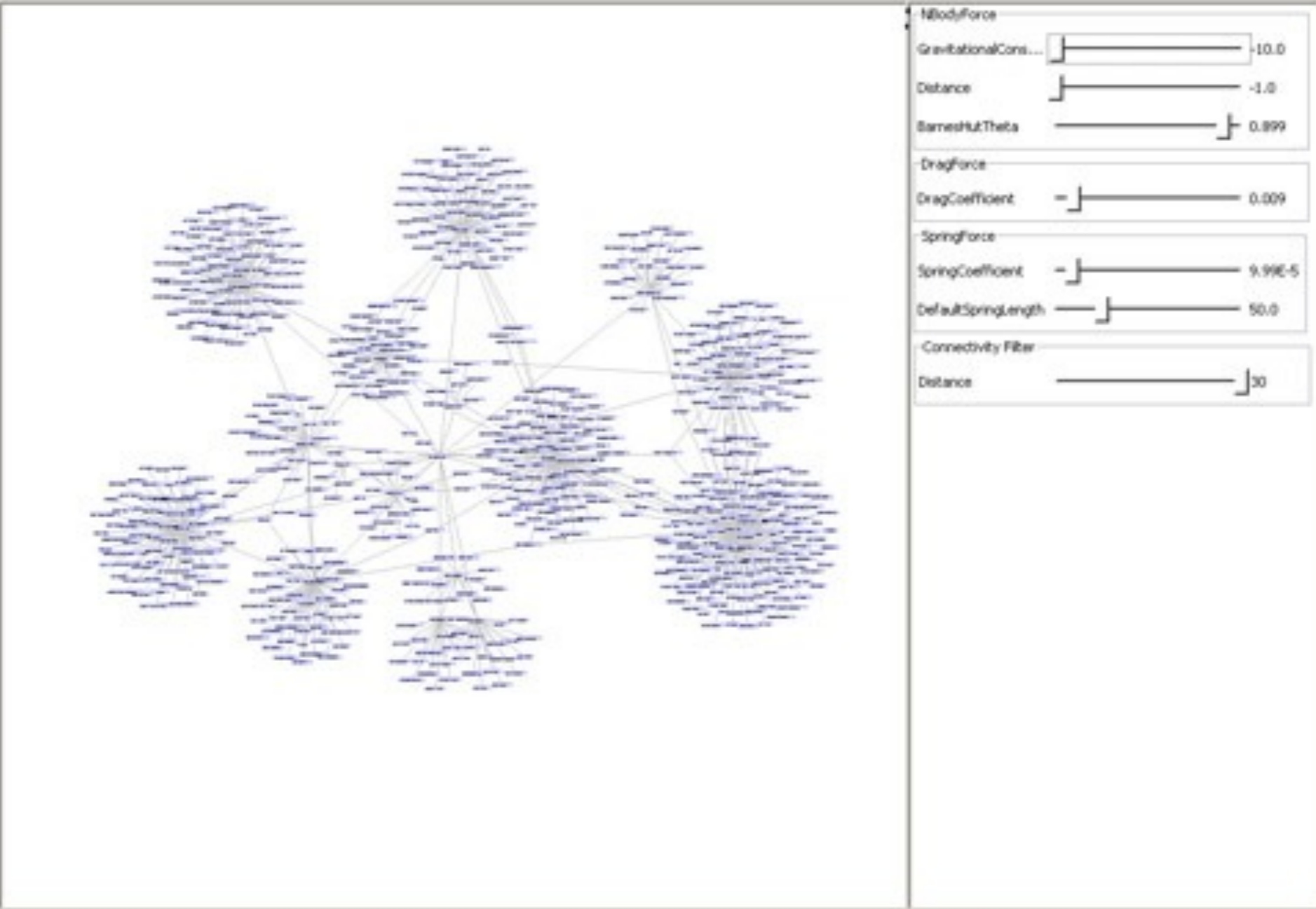
# Acquisition: Crawling

- Crawling
  - ❖ Iterate over profile.php?id=1001, profile.php?id=1002, ...
  - ❖ Uniform HTML file structure
  - ❖ Countermeasure: Captchas
    - Arms-Race
    - Not widely deployed



prefuse | graphview

Data



Control Panel Parameters:

- NBodyForce**
  - GravitationalConst...: -10.0
  - Distance: -1.0
  - BarnesHutTheta: 0.899
- DragForce**
  - DragCoefficient: 0.009
- SpringForce**
  - SpringCoefficient: 9.99E-5
  - DefaultSpringLength: 50.0
- Connectivity Filter**
  - Distance: 30

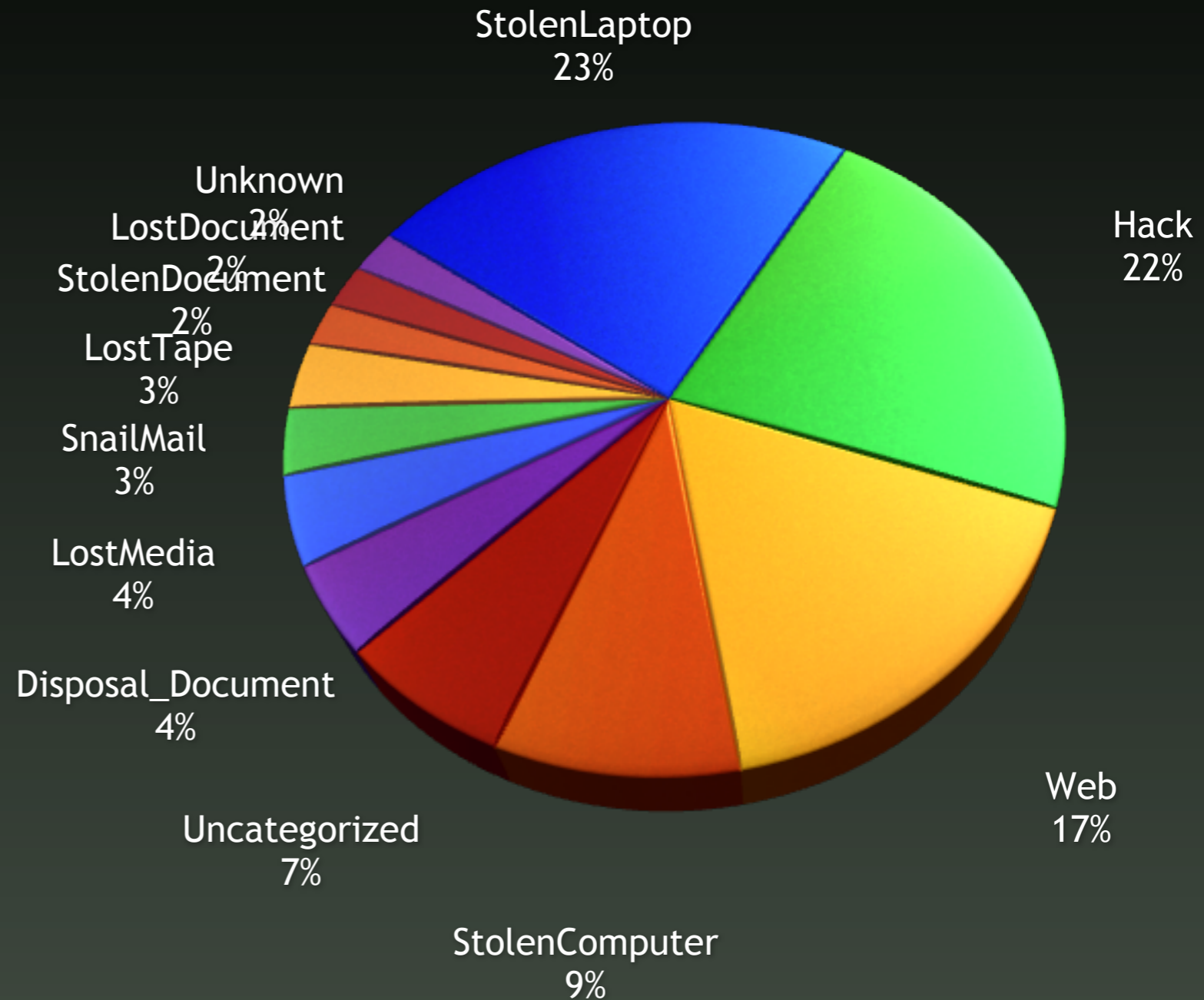
Windows Taskbar: Start, Postein..., Deathr..., 212. Pin..., FlashFIP, Lokaler..., Java 2..., prefuse, Java - ..., Downlo..., prefu... 16:37

Source: <http://icepic.org/wiki/doku.php?id=studivz-analyse-tool>

# Acquisition: Cracking

- Cracking

- ❖ Wide attack surface of Web 2.0 apps with a multitude of technologies
- ❖ SQL Injection most often reported CVE in 2008
- ❖ Dataloss Foundation: from 01/2008 to 07/2008 reported 25M identity breaches



Incidents by Breach Type, Source:Open Security Foundation's DataLossDB

# Acquisition: Buying

- Illegal purchase
  - ❖ Boards, IM, IRC
- Legal purchase
  - ❖ Data traders
  - ❖ Call Center Incidents in Germany



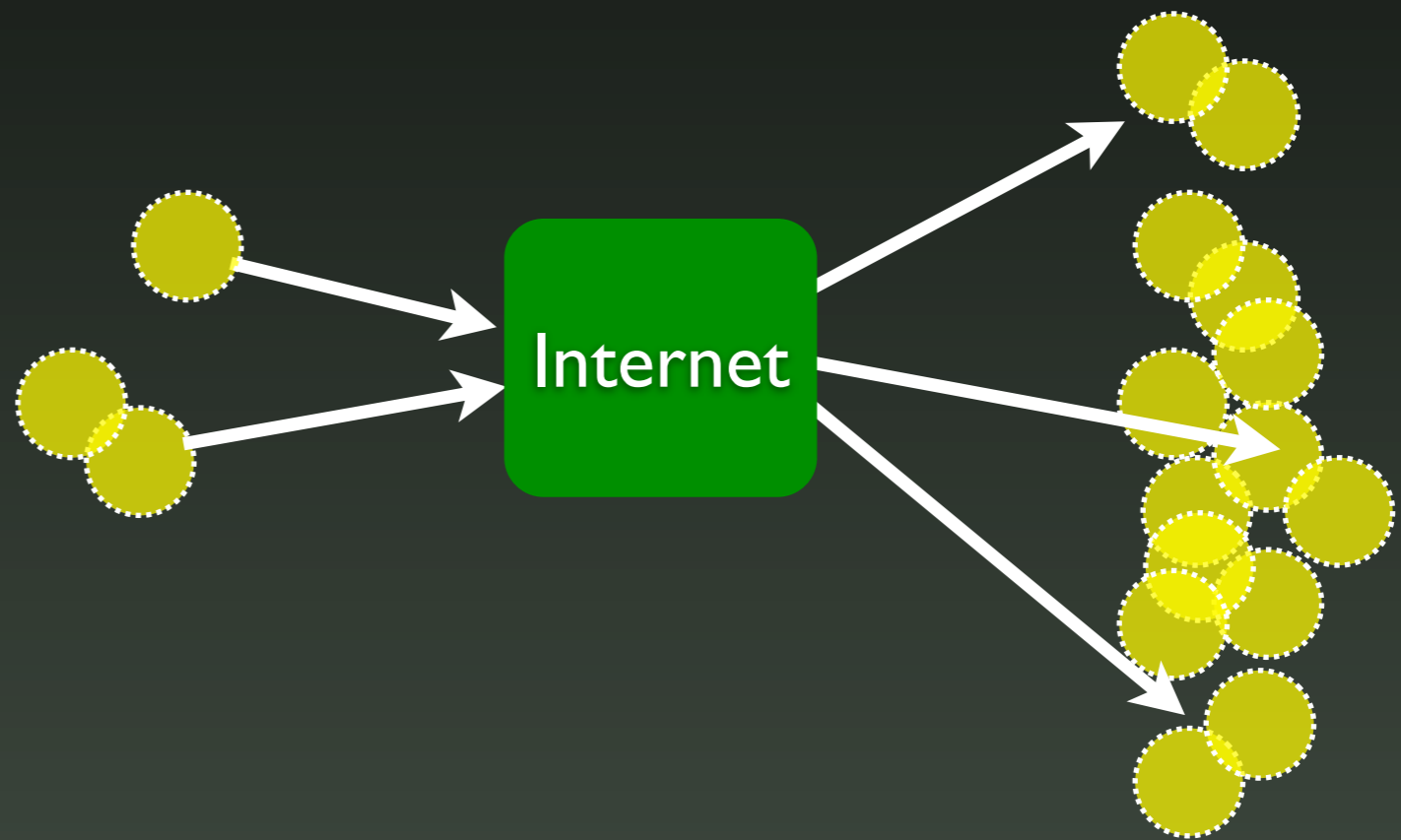
An advertisement for 'CC MARKET'. The logo 'CC MARKET' is displayed in a white box on the left. The main text reads 'Kreditkarten kaufen wie im Supermarkt' (Buy credit cards like in a supermarket). At the bottom, there are links for 'Länderübersicht | FAQ | Statistiken | Aufladung'. The top right corner of the ad has 'Home | Logout'.

# *the web 2.0 leverage*

# Web 1.0

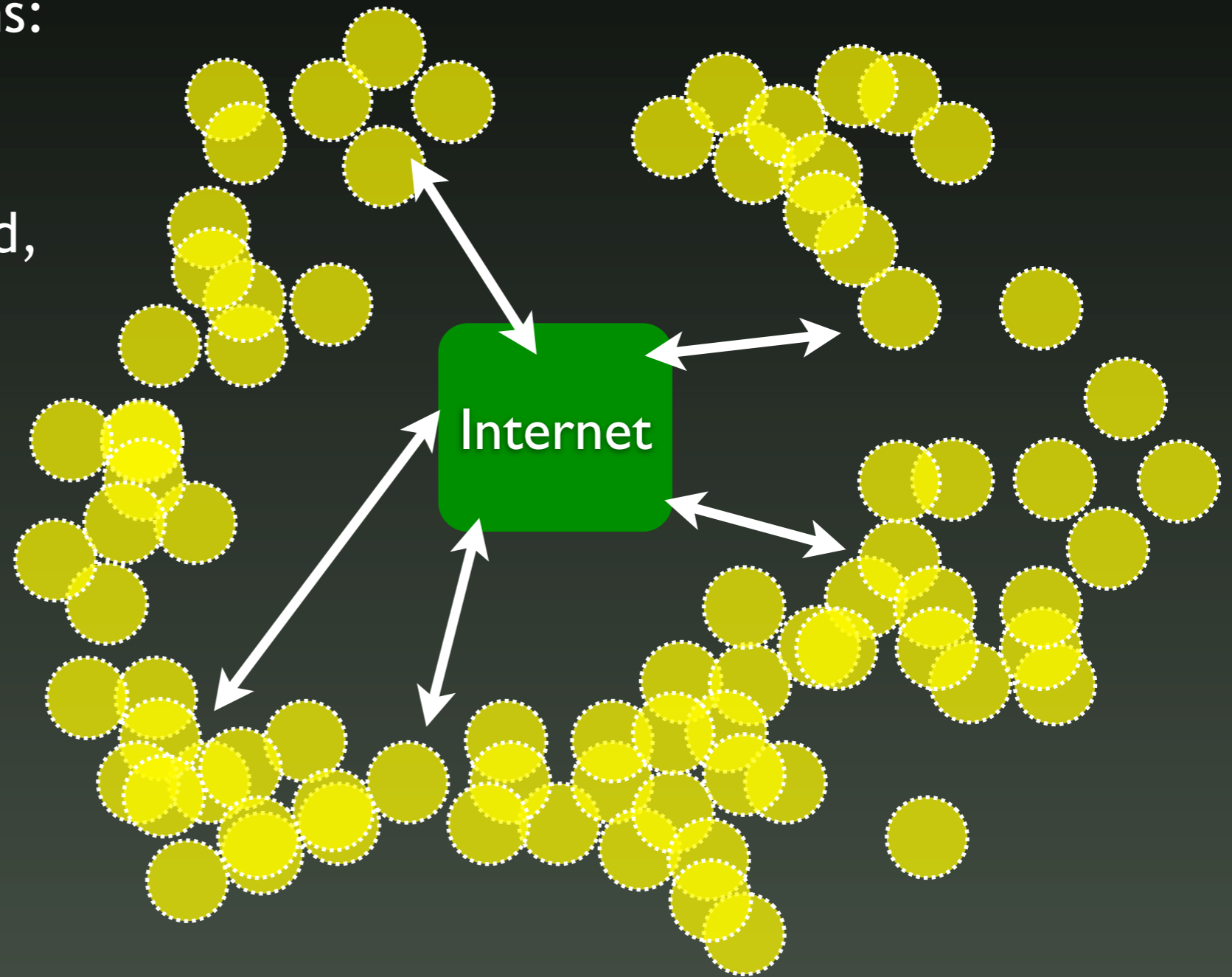
E-Mail

B2C  
E-Commerce



# Web 2.0

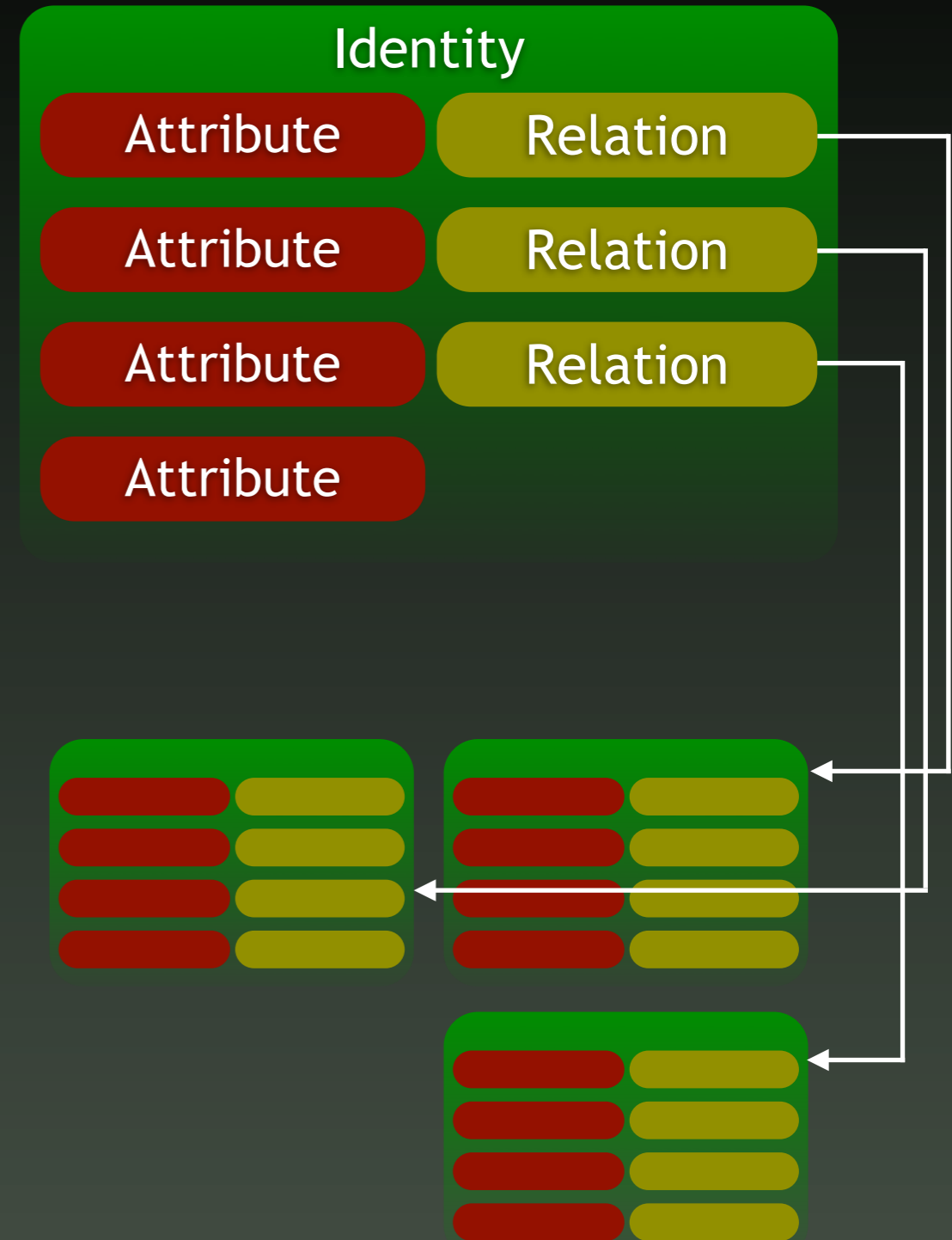
- Interlaced, supporting prerequisites for security implications:
  - ❖ Open APIs
  - ❖ Multimedia-based, personal data



*how to refine the data?*

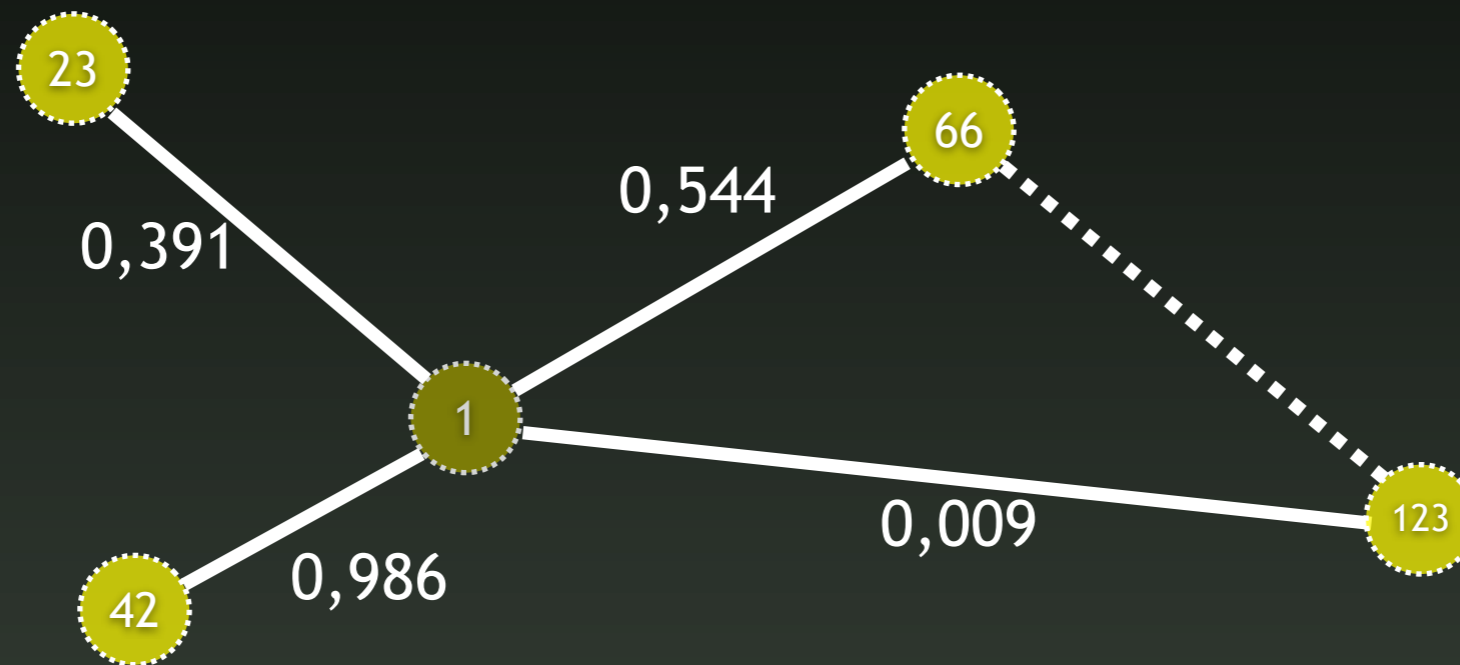
# Identity

- $( i, A, R ) =$  describes ID
- Attributes:  
(Name, Address, Hobbies, Birthday, E-Mail, Employee)
- Relations:



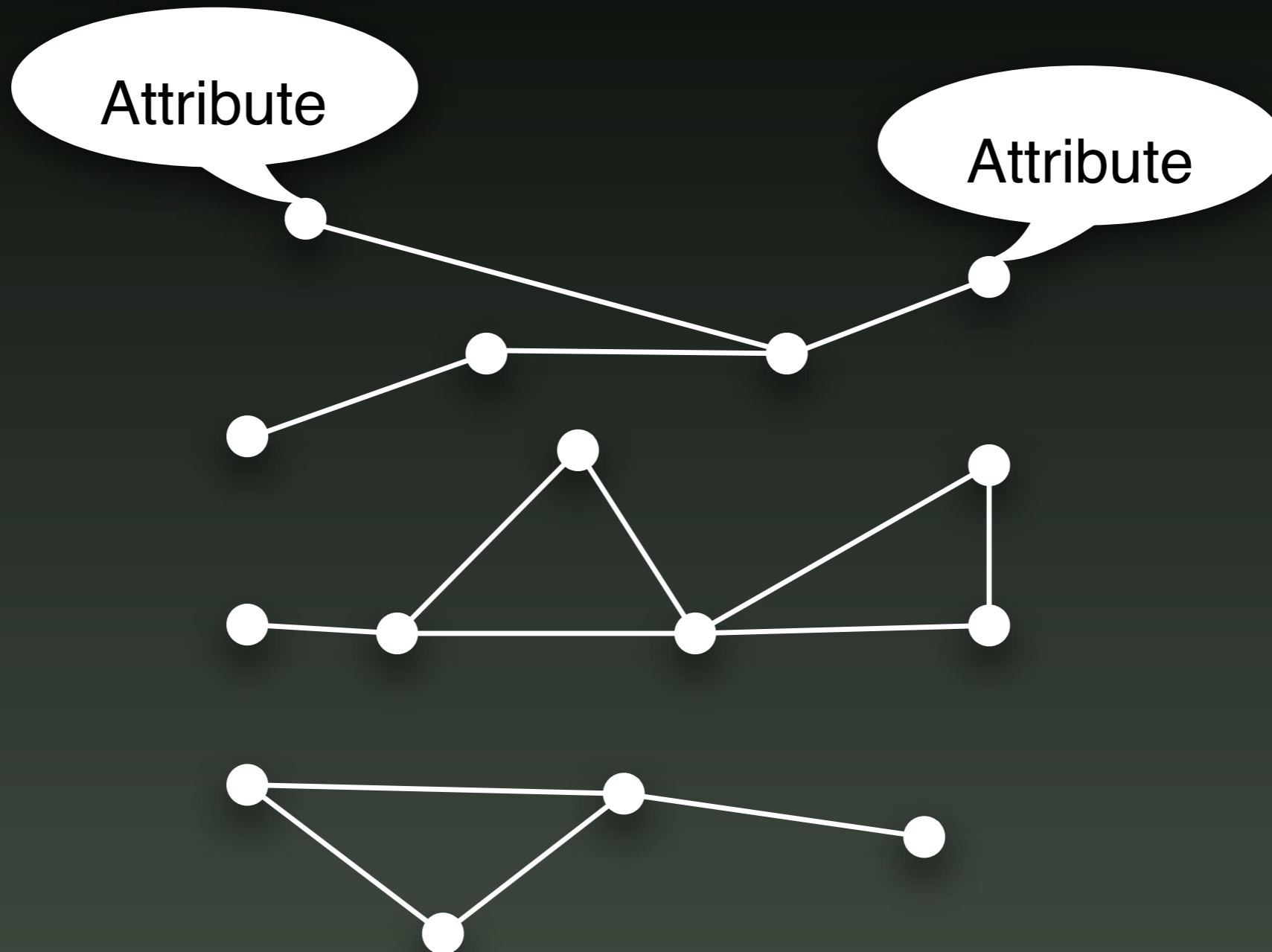
$i_n$	23	42	66	123
$w_n$	0,391	0,986	0,544	0,009

# Network

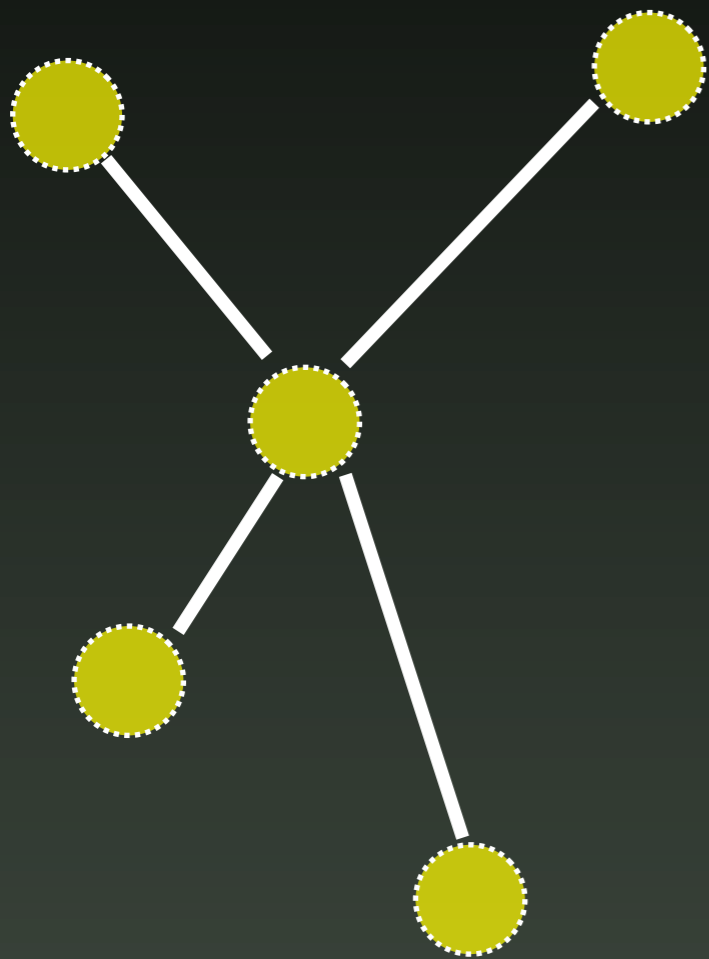


$i_n$	23	42	66	123
$w_n$	0,391	0,986	0,544	0,009

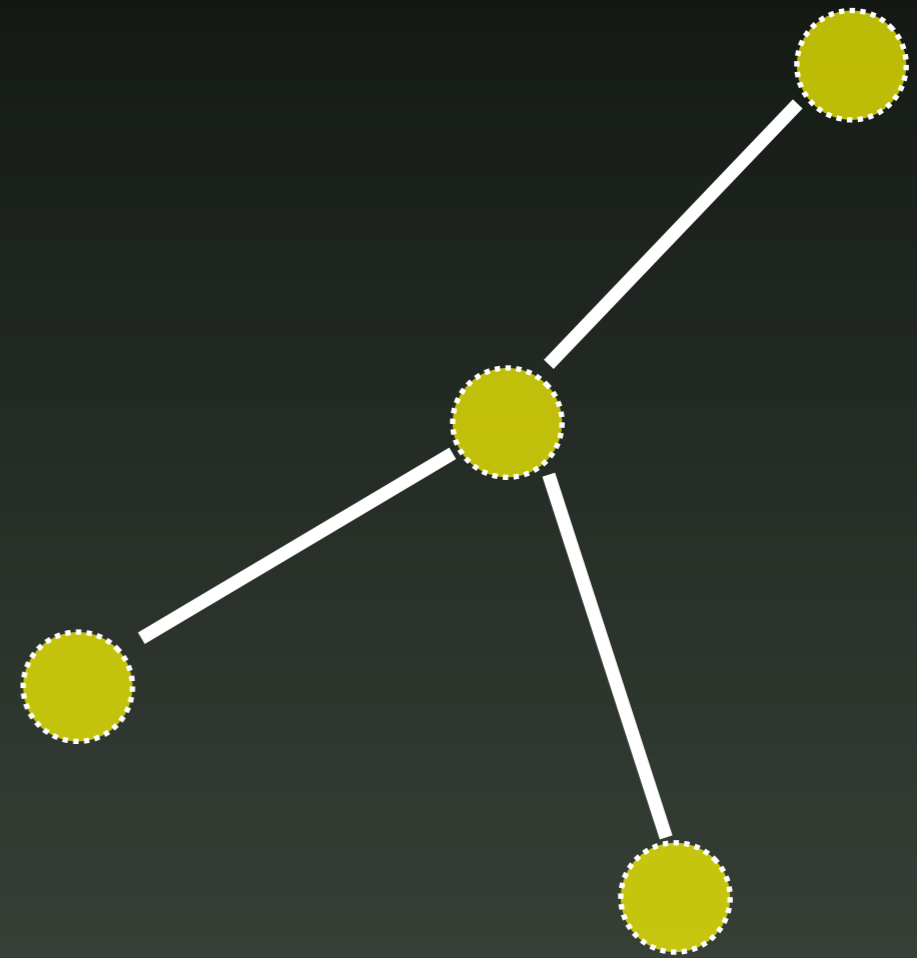
# Aggregation



# Correlation

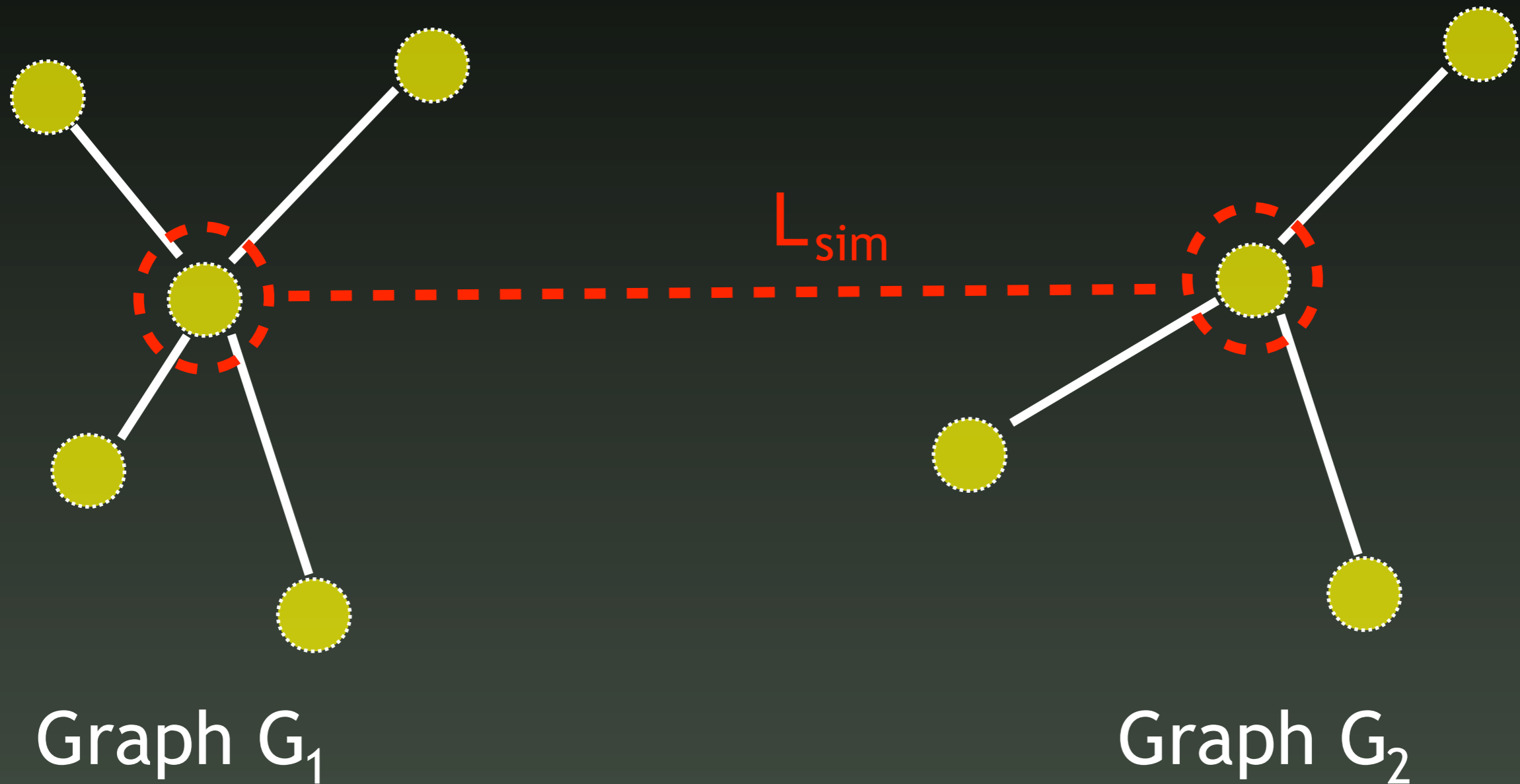


Graph  $G_1$

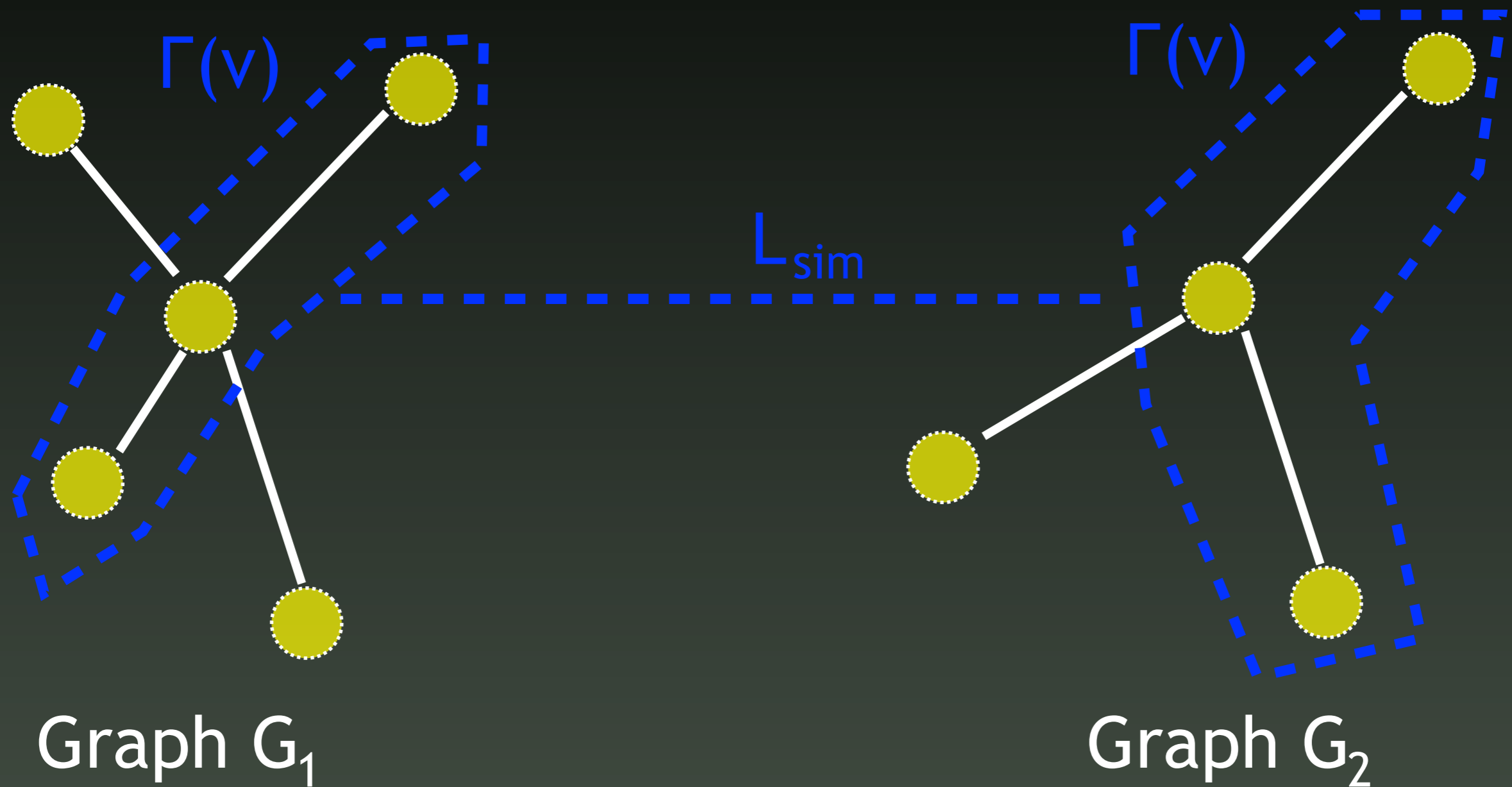


Graph  $G_2$

# Correlation



# Correlation

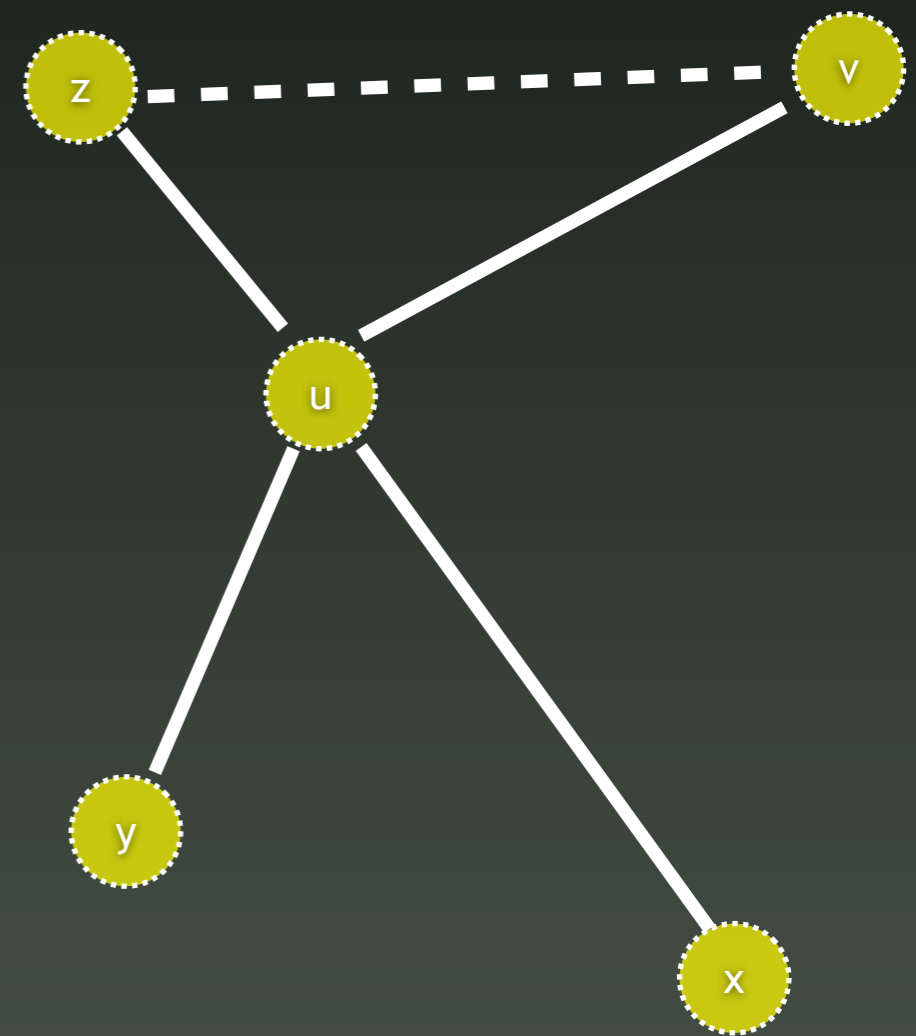


# Correlation

- Graph-based Data Mining
  - ❖ Neighborhood of target vertex
  - ❖ Is Graph  $G_1$  isomorphic to a sub-graph of  $G_2$  ?
  - ❖ If  $L_{\sim} > \text{threshold } \pi$   
then profile linkage is possible

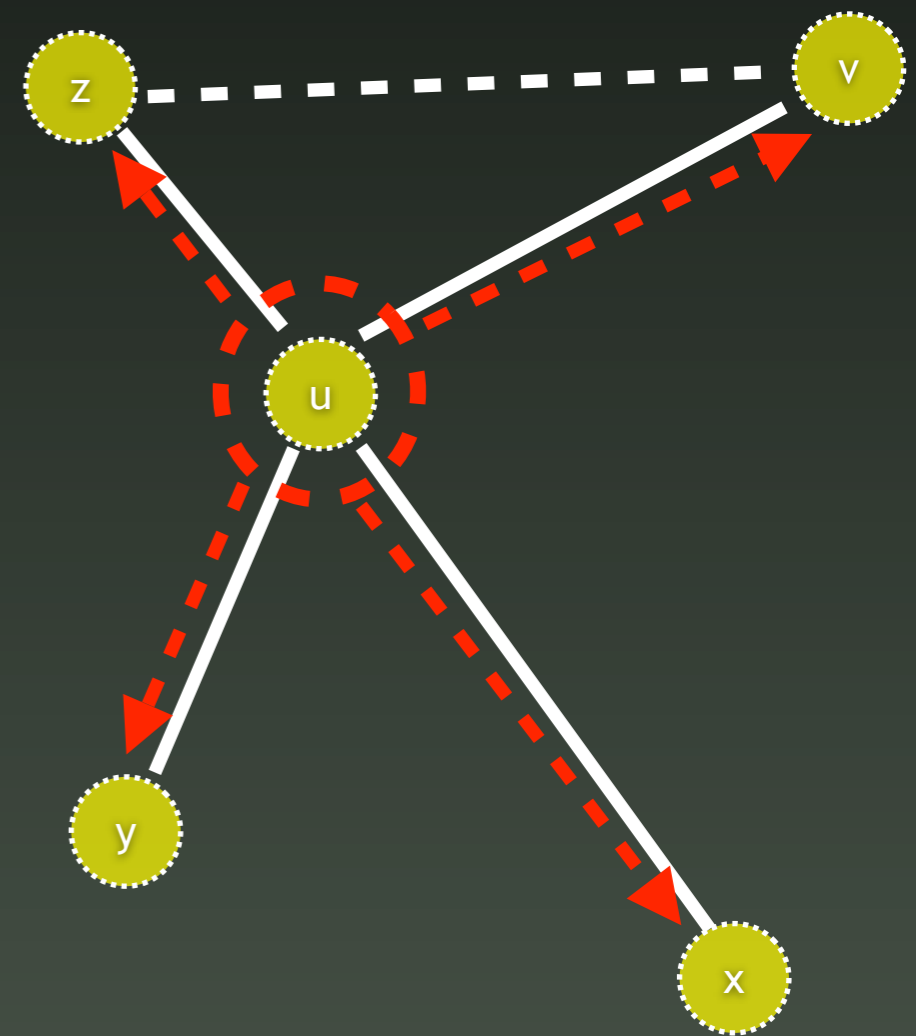
# Attack Vector

- Abuse of the trust-relation between identities
- Reference to common and personal attributes
  - ❖ „Hello [v, x, y, z], it's me, u“



# Attack Vector

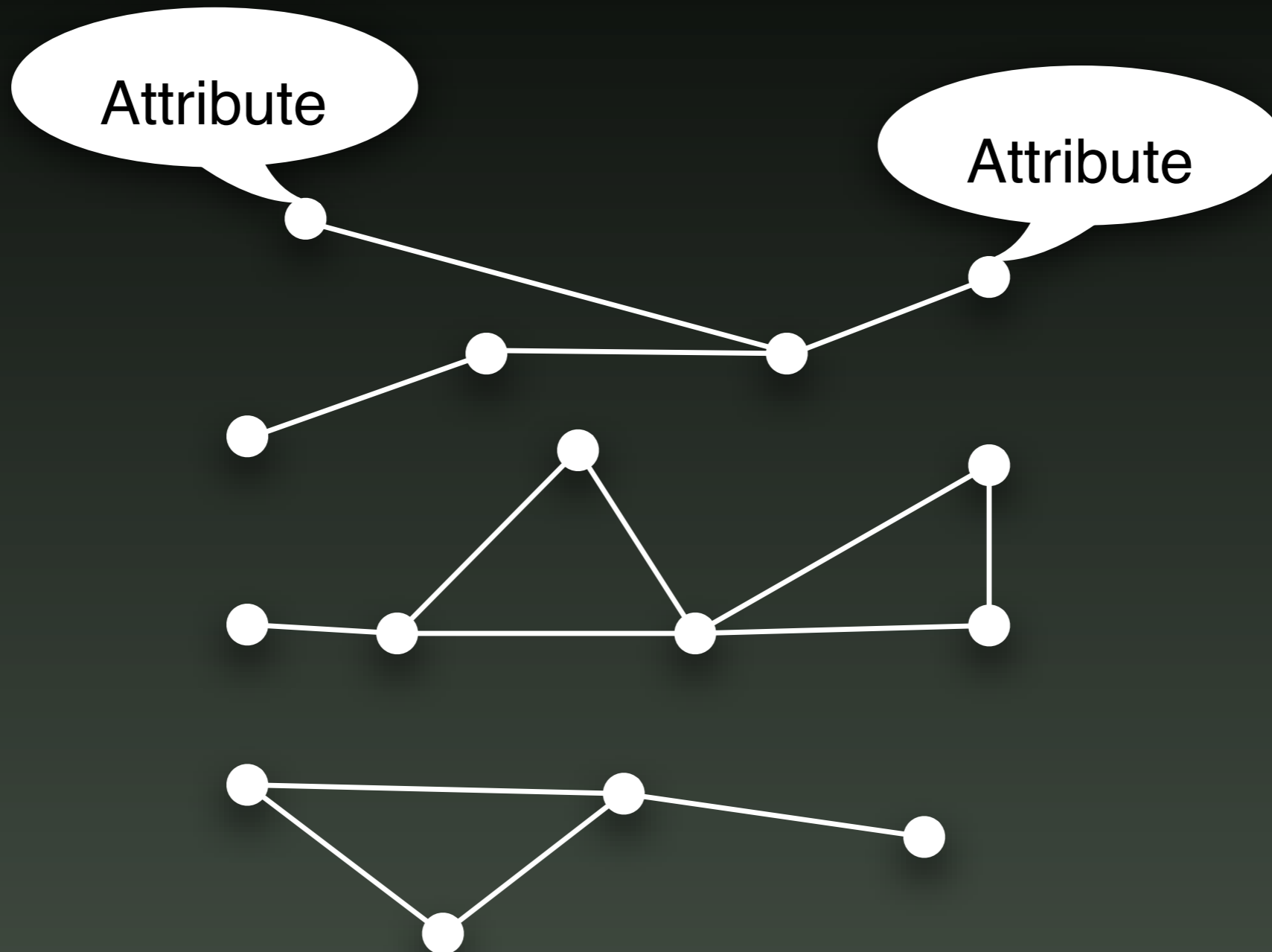
- Abuse of the trust-relation between identities
- Reference to common and personal attributes
  - ❖ „Hello [v, x, y, z], it's me, u“



# Selection Methods (still experimental)

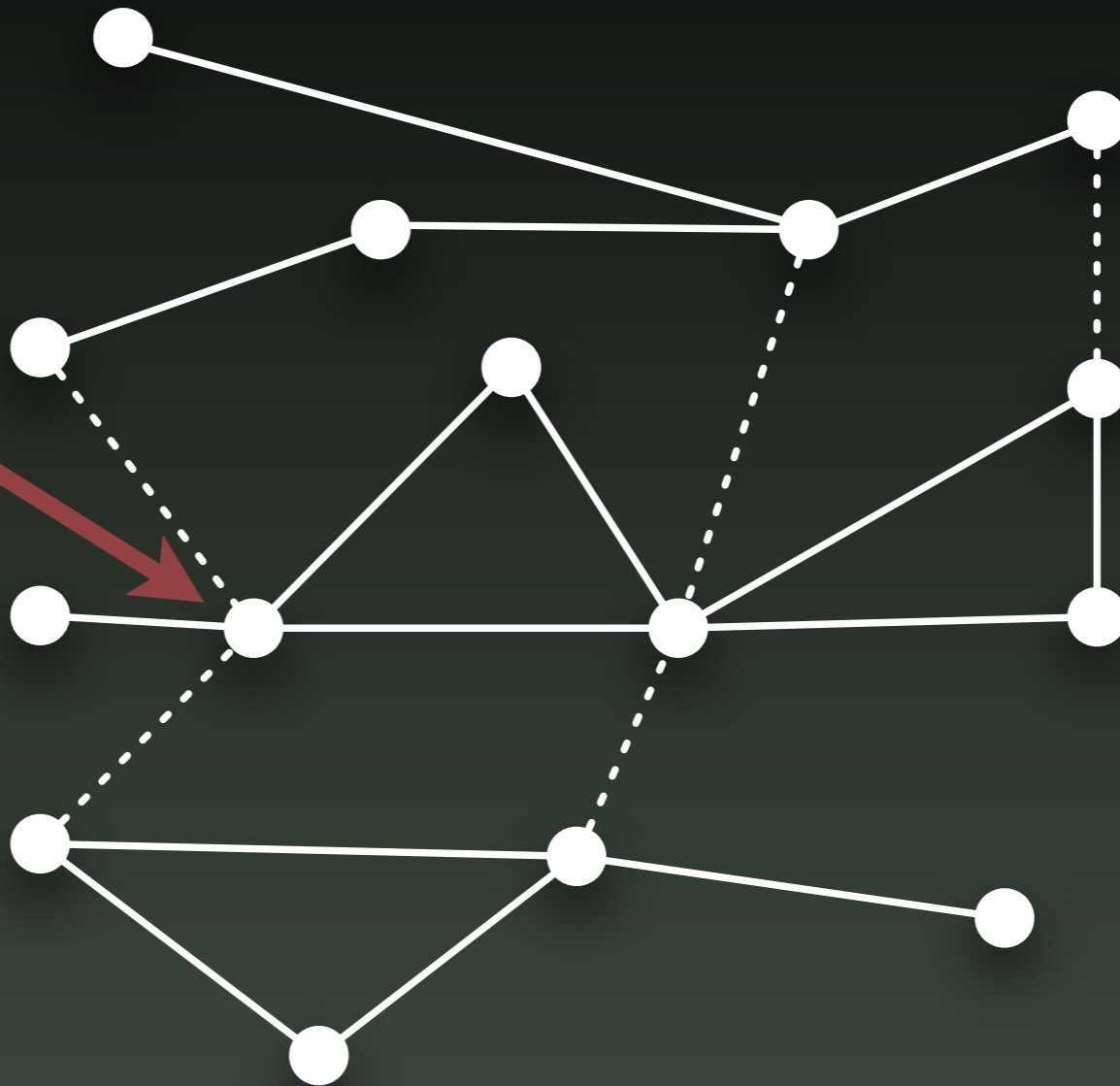
- Maximal information about identity available?
  - ❖  $(\text{deg}(A, R) \rightarrow \text{max})$
- Broker in sub graph?
  - ❖ Neuralgic points, if removed graph falls into subgraphs
- Complete sub graph?
  - ❖ Clique
- Maximum (local) cluster coefficient?
  - ❖ Implicates distinctive neighborhood

# Aggregation

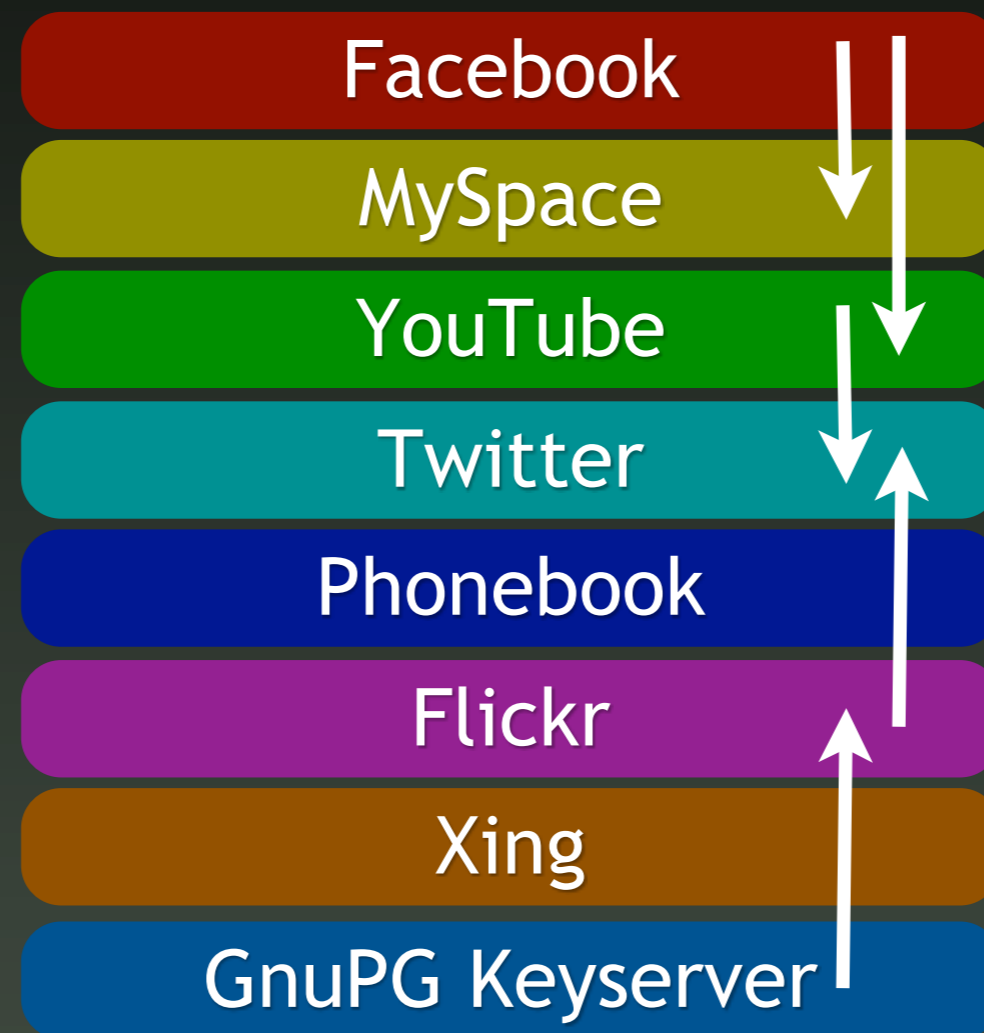


# Correlation

Same Person in  
three different SNS

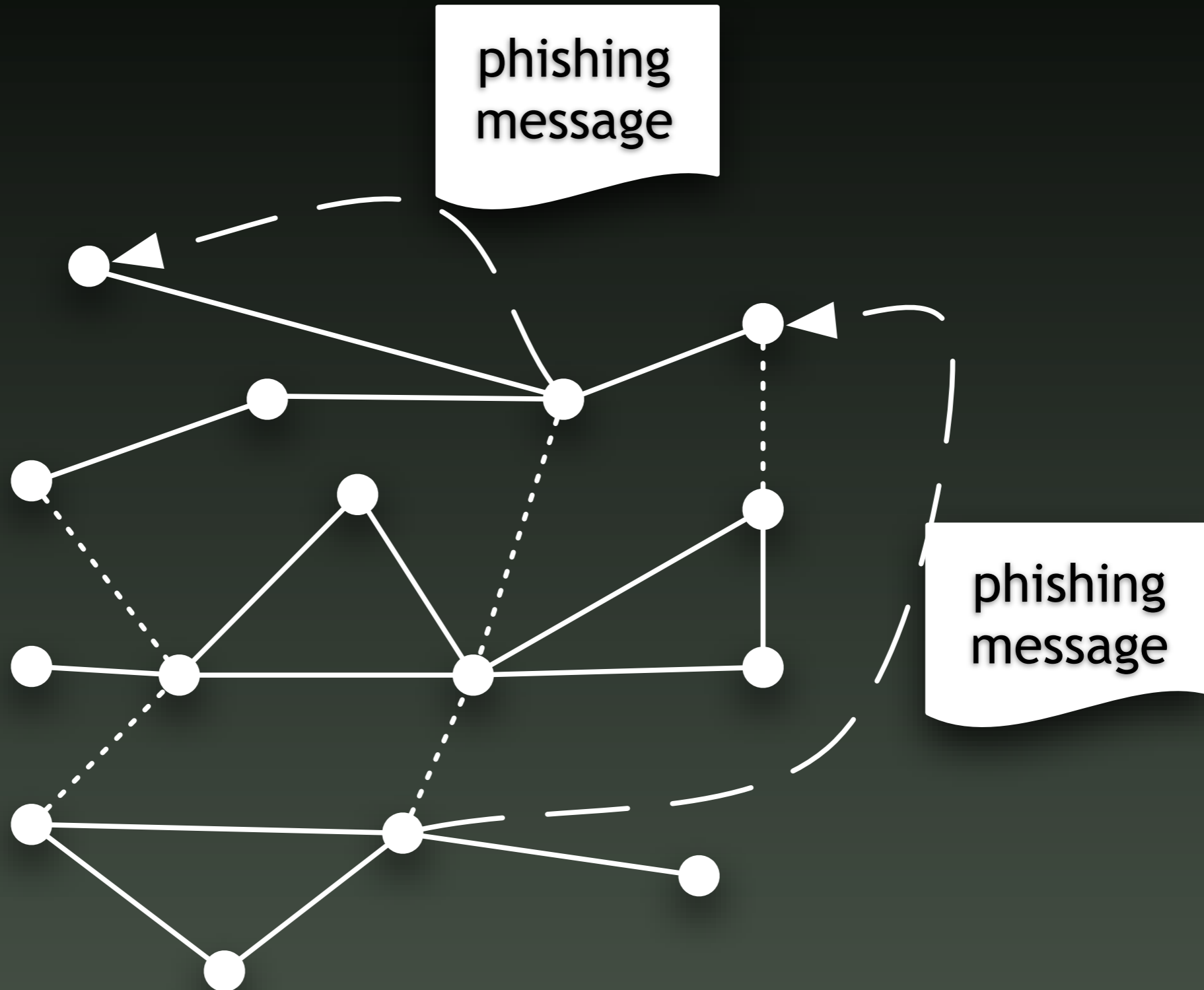


# Refinement



↑ Relation

# Attackphase

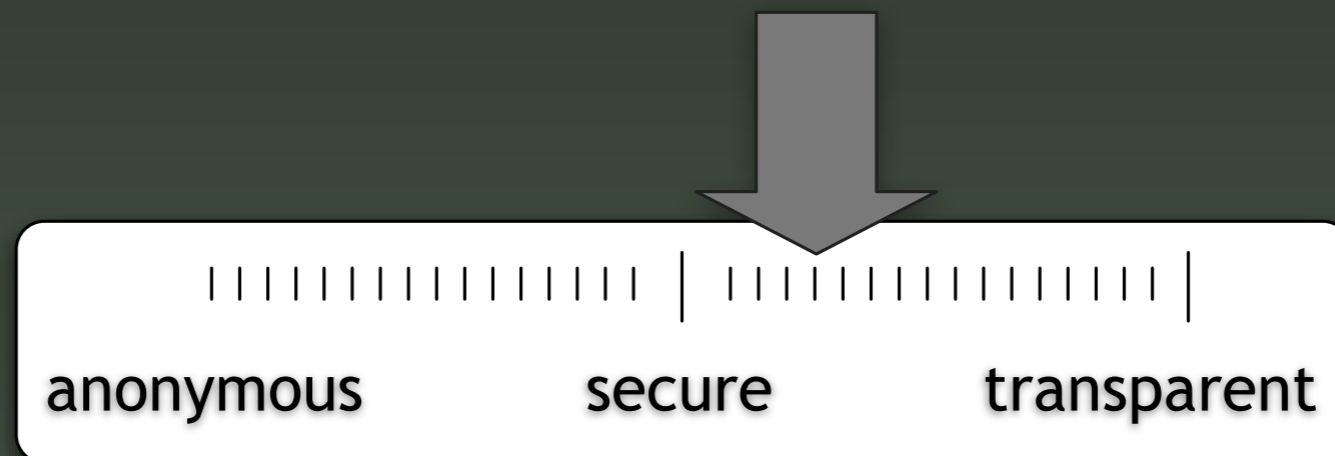


# Results

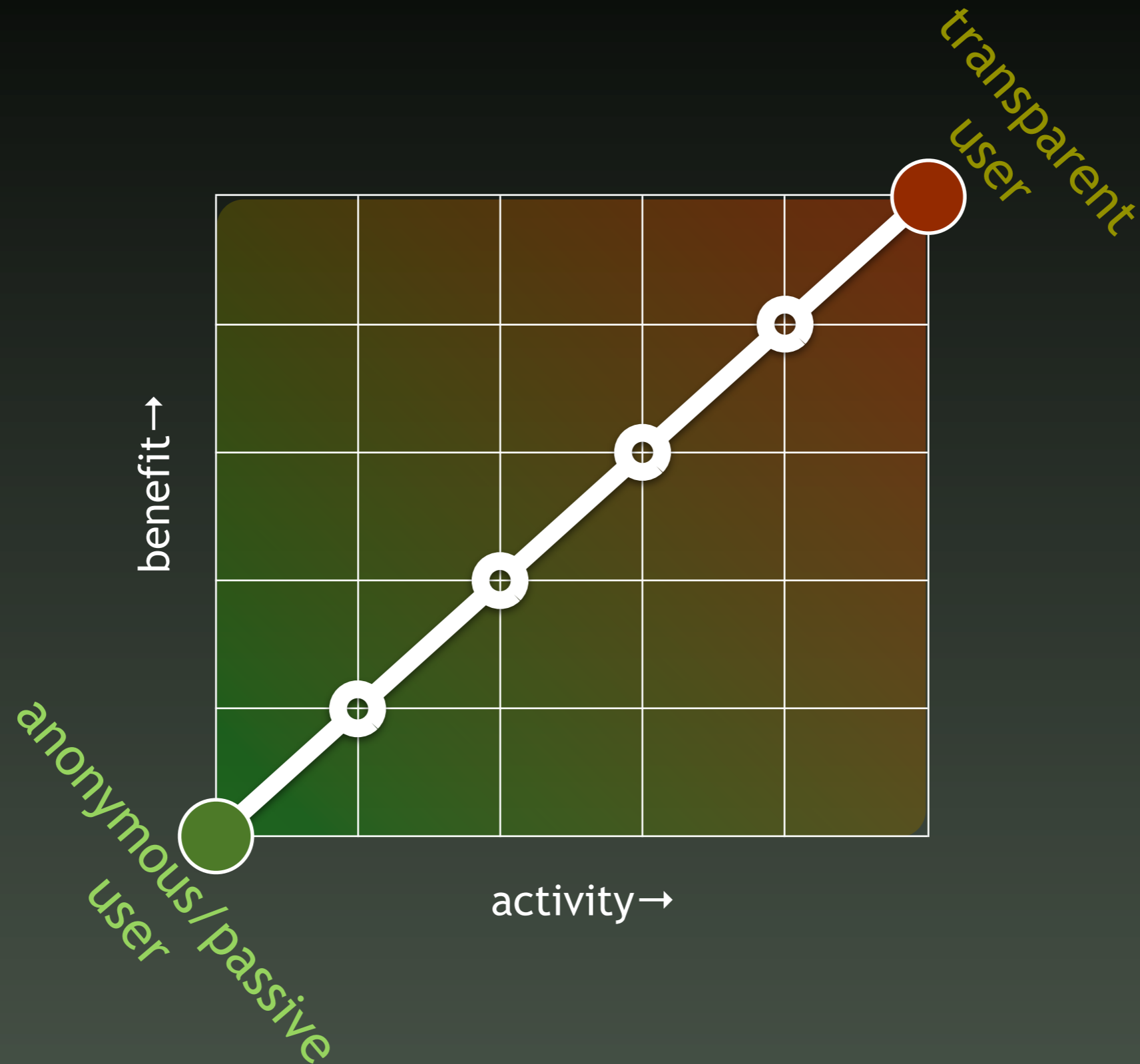
- Reacting to a link may result into
  - ❖ Stolen credentials by sophisticated phishing
    - ⦿ Establishing trust by directly addressing victim, referencing to friends and “personal/private” data
  - ❖ Browser-based infection by exploiting insecure browsers and plugins
  - ❖ Browser-based Cross-site Request Forgery (XSRF)
  - ❖ ...
- Like Spam, massive scale factor makes it profitable for the adversary
  - ❖ Achieve through automation

# Countermeasures

- Different layers
  - ❖ Pseudonymous user accounts (?)
  - ❖ Anti-crawling methods by the service provider
  - ❖ Security analysis of Webservices & -applications
- Increase users feeling for data-austerity
- Educate users about privacy settings of platforms



# Basic Problem: Trade-Off



# Calculating Threat

$$\kappa = \sum_{n=1}^s \frac{\left( \frac{\sum_{i=1}^r \delta_i \cdot g_i}{r} \right) \cdot w_n}{s} = \left( \sum_{i=1}^r \frac{\delta_i \cdot g_i}{r} \right) \cdot \left( \sum_{n=1}^s \frac{w_n}{s} \right)$$

Var.	Description
$r$	amount of personal attributes available
$\delta$	binary value (1,0) which defines if a personal attribute $a$ is known to the attacker
$g$	weighting coefficient which testifies the relevance of the personal attribute with regards to the privacy of the specific identity
$s$	amount of relationships the specific identity obtains
$w_n$	relation coefficient which expresses the strength of the relationship
$n$	number of relation stored in the relation matrix $K$

personal attributes $a_i$	$\delta_i$	$g_i$
name	1	0.35
address	0	0.76
nickname	1	0.12
date of birth	1	0.81
pol. affiliation	1	0.73

# Related Work

- Jakobsson et al: Social Phishing, 2005
  - ❖ Core problem
- Fong, Anwar and Zhao: A Privacy Preservation Model for Facebook-Style Social Network Systems, 2009
  - ❖ Formalized access control model
- Narayanan and Shmatikov: De-anonymizing social networks, 2009
  - ❖ Linking twitter to flickr profiles with 12% error rate (↳Pseudonyms do not help!)

# Conclusion

- Status Quo
  - ❖ Social media main asset is personal data, but
  - ❖ Acquisition, correlation and misuse of such data is easily feasible for internet crime syndicates
- Mitigation
  - ❖ Protection of user data by new concepts is a chance but also the duty of the service providers
  - ❖ Increase users feeling for privacy needs

Thank you for your attention!  
Questions?

- ➔ IFIP/Primelife'09, Nice, France — 09.09.2009
- © Dominik Birk, Felix Gröbert, Christoph Wegener
- ☎ [felix@groeibert.org](mailto:felix@groeibert.org)
- ✂ [creativecommons.org/licenses/by-nc-nd/3.0/de](http://creativecommons.org/licenses/by-nc-nd/3.0/de)