1. Subgraph Analysis

a) Background

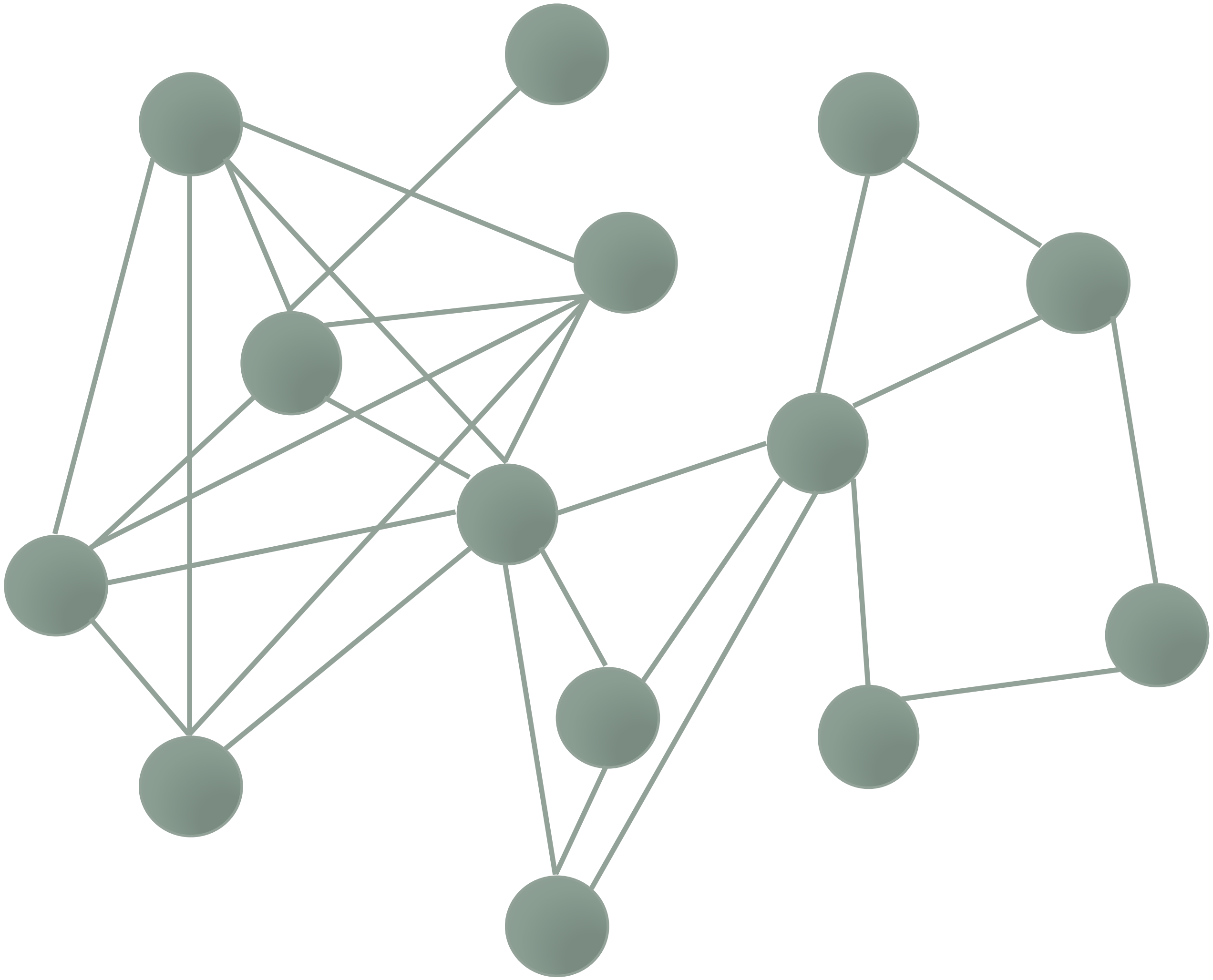
b) Normal Behavior

c) Abnormal Behavior

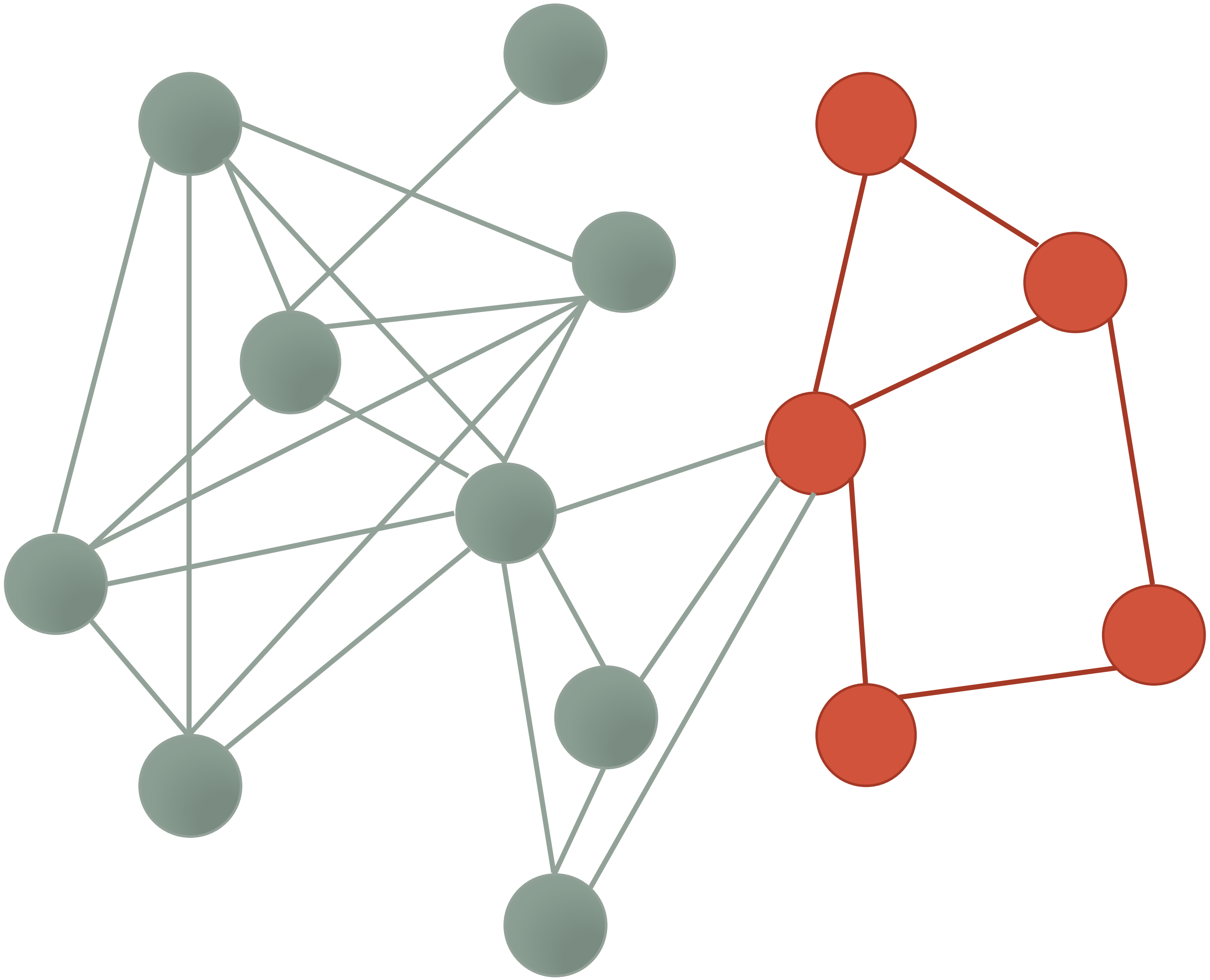
2. Propagation Methods

3. Latent Factor Models

What is a subgraph?



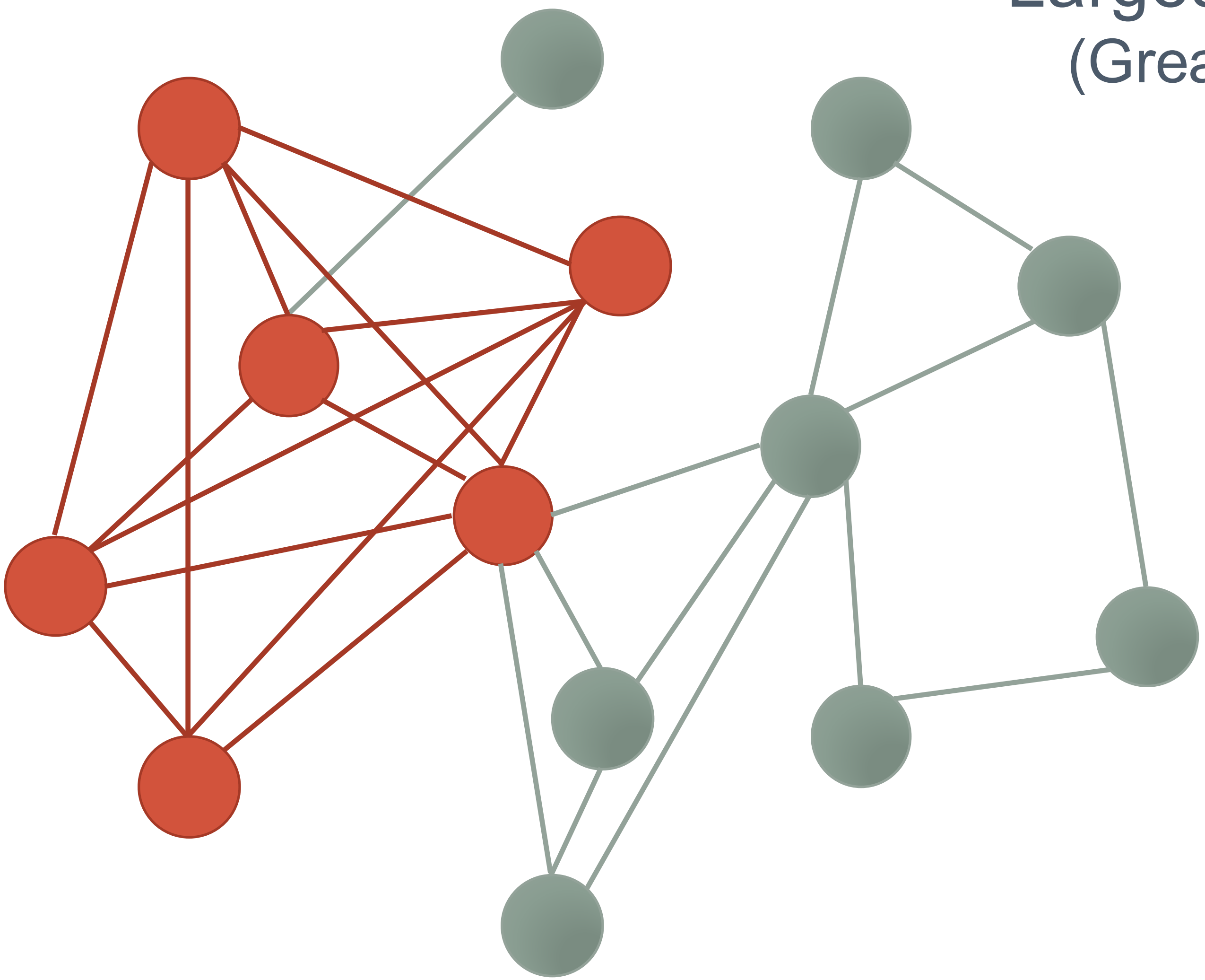
What is a subgraph?



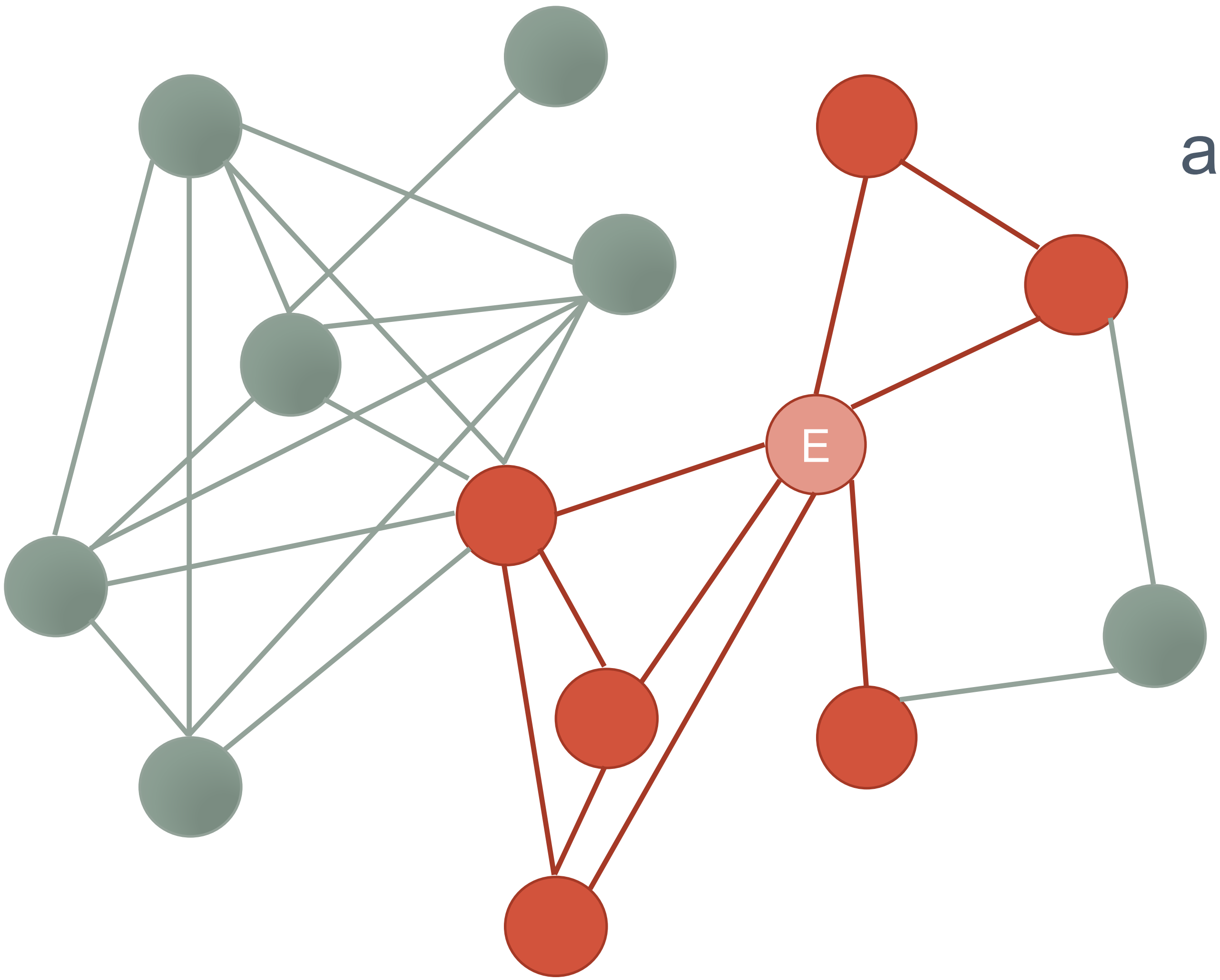
Subset of nodes and the edges between them

What are some useful subgraphs?

Largest dense subgraph
(Greatest average degree)

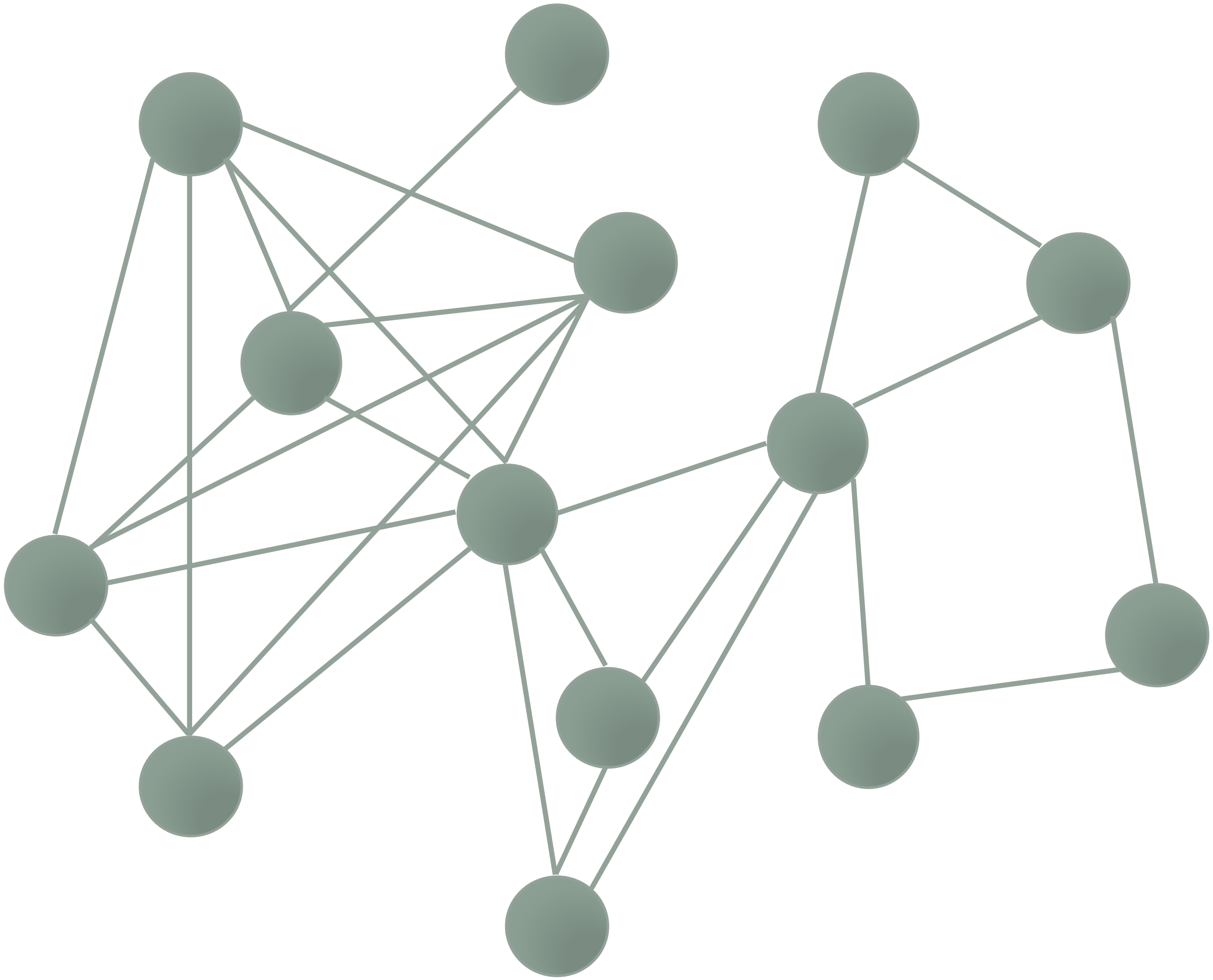


What are some useful subgraphs?

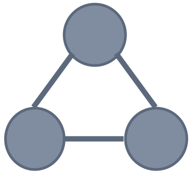


Ego-network:
the subgraph
among a node and
its neighbors

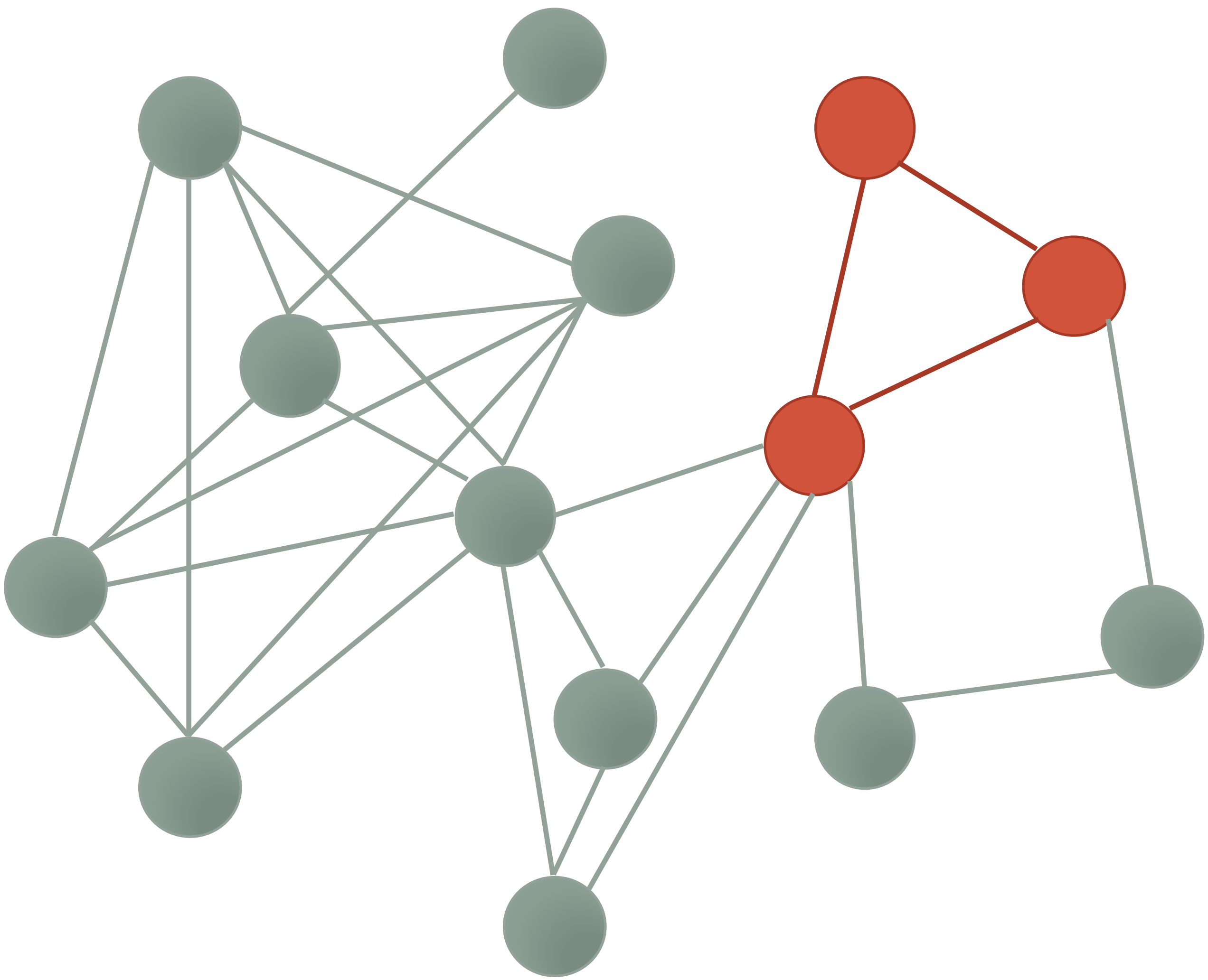
What are some useful subgraphs?



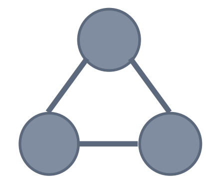
Graph queries:
find subgraphs of particular pattern



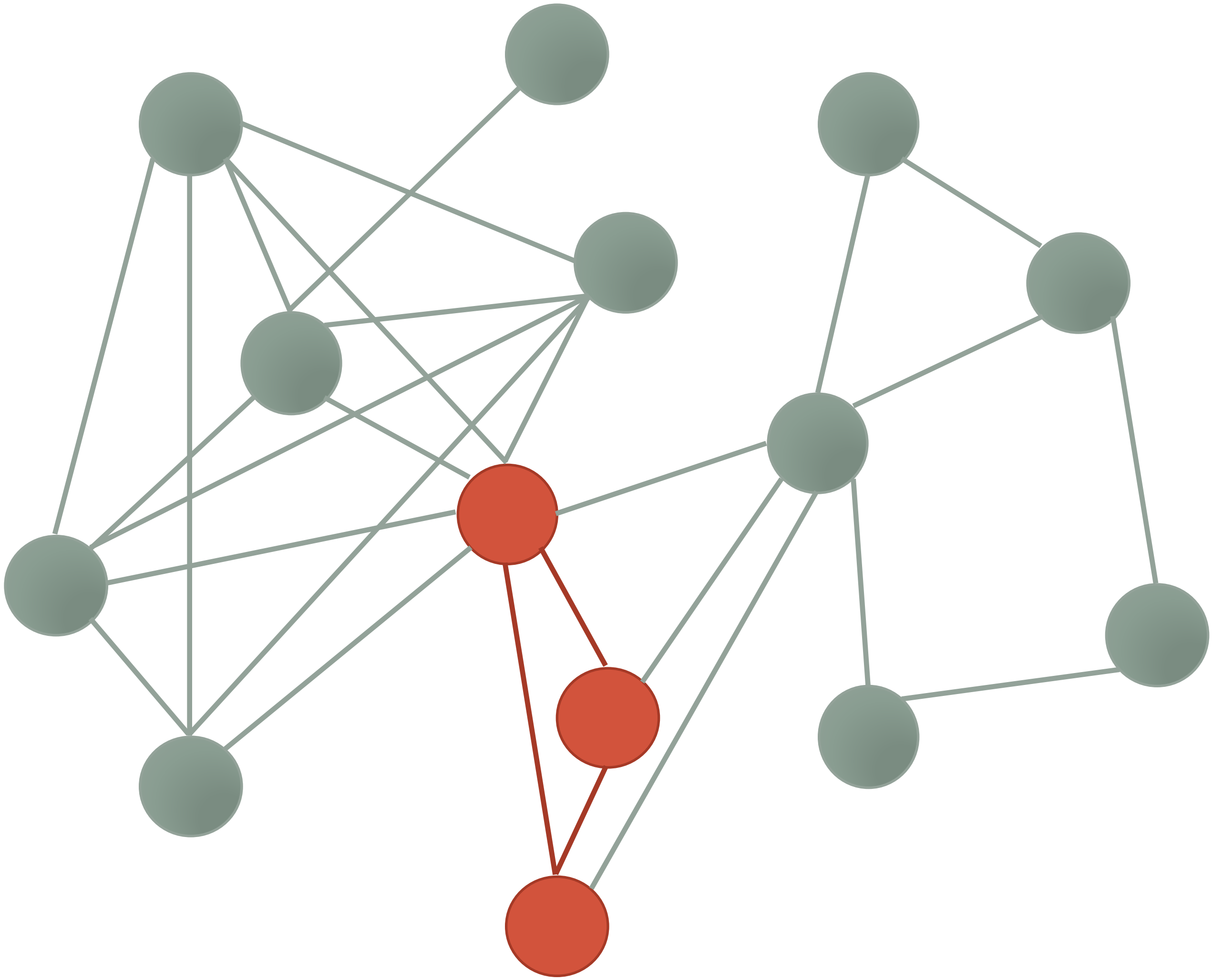
What are some useful subgraphs?



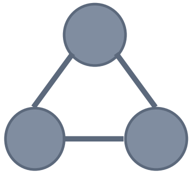
Graph queries:
find subgraphs of particular pattern



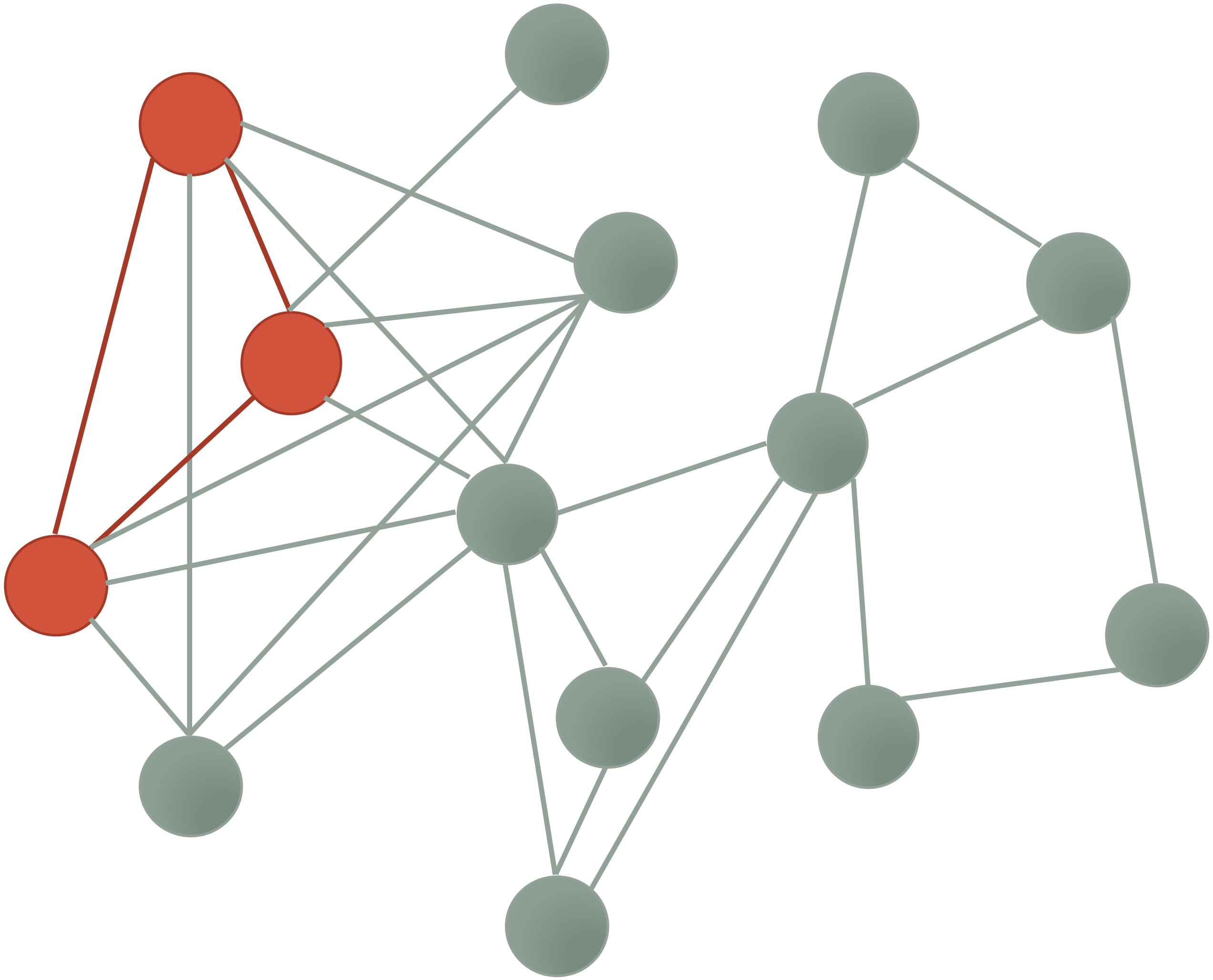
What are some useful subgraphs?



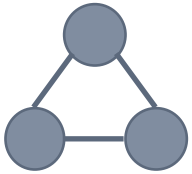
Graph queries: find subgraphs of particular pattern



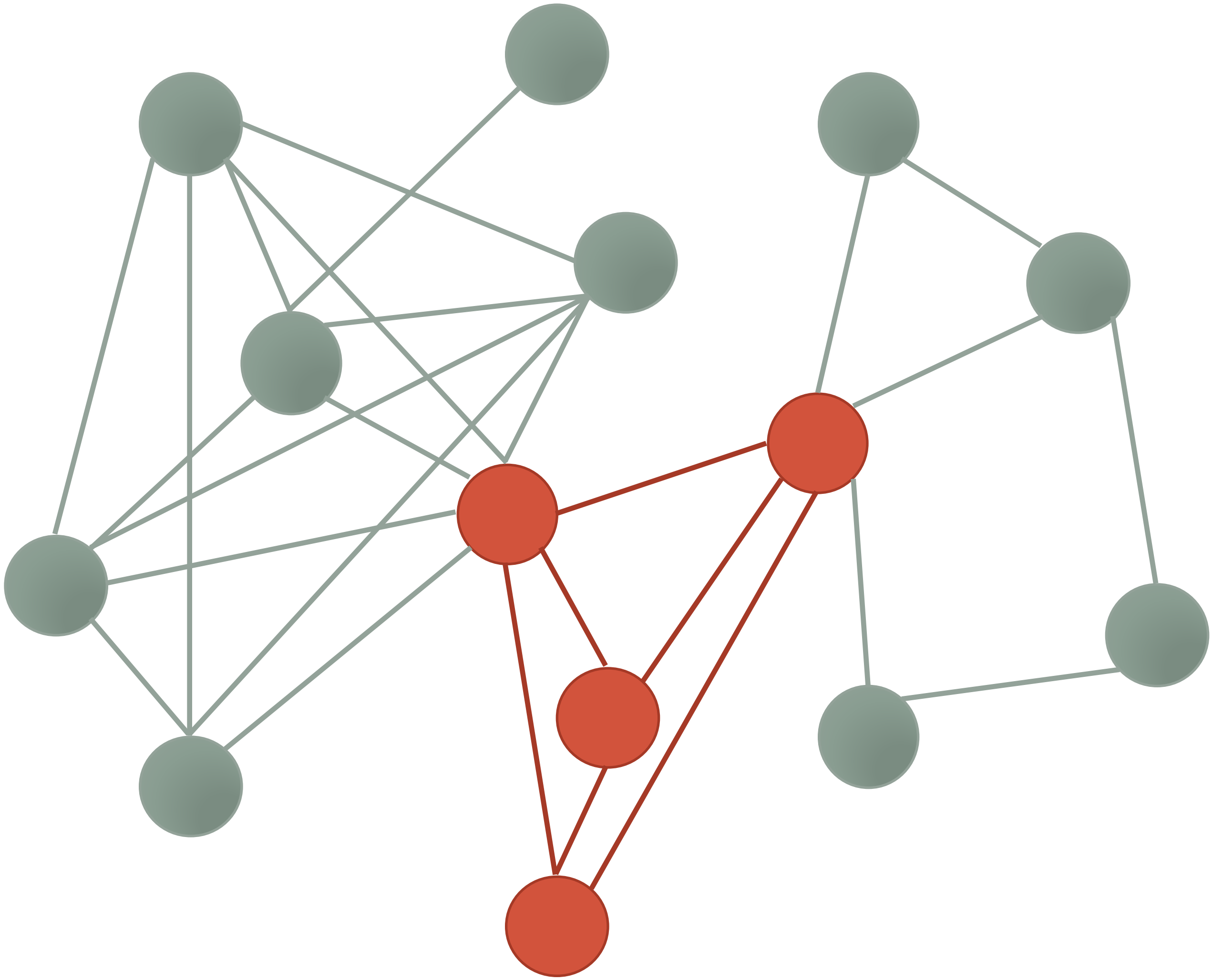
What are some useful subgraphs?



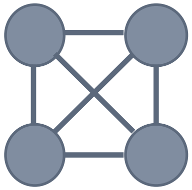
Graph queries:
find subgraphs of
particular pattern



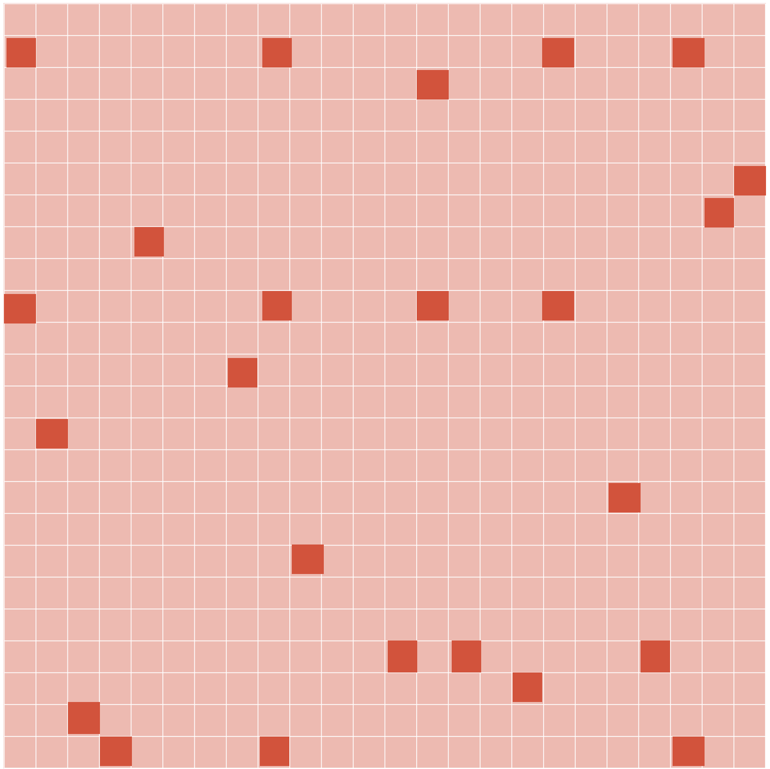
What are some useful subgraphs?



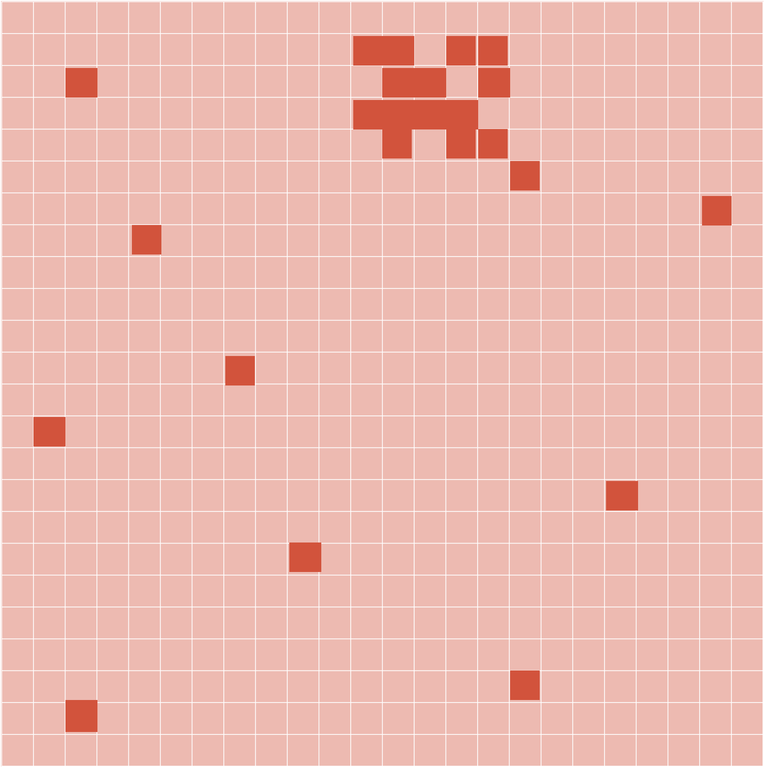
Graph queries:
find subgraphs of
particular pattern



Subgraphs as submatrices

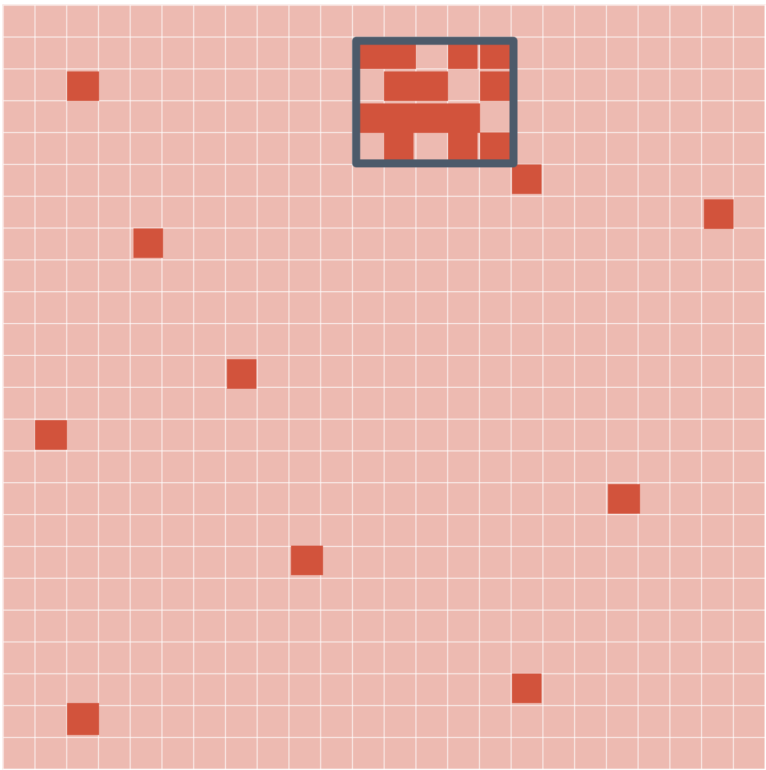


Subgraphs as submatrices

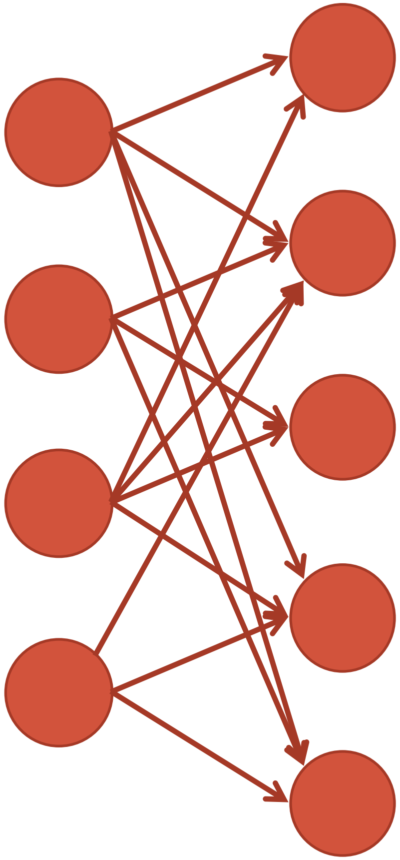


Rearrange to find dense regions!

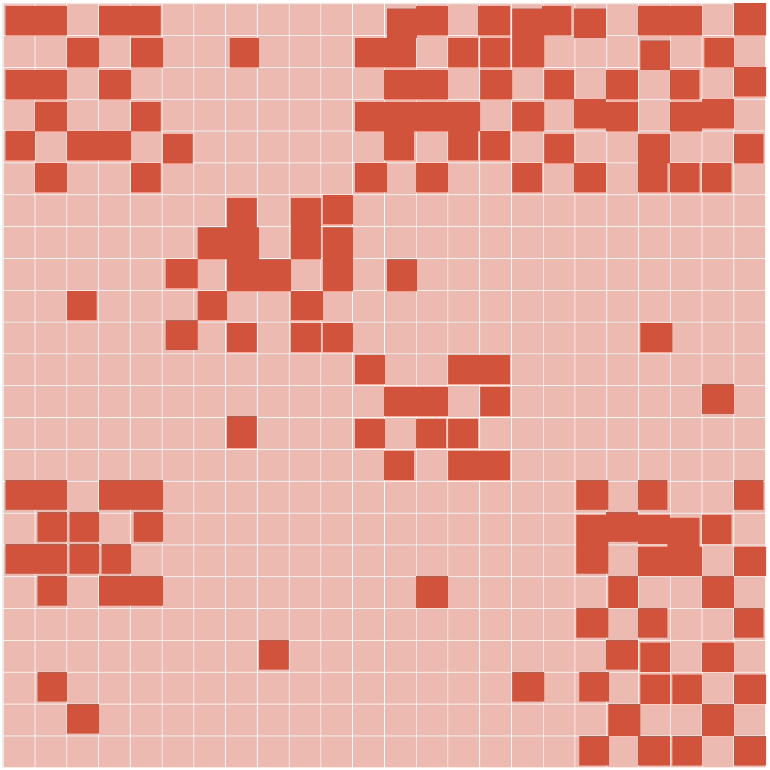
Subgraphs as submatrices



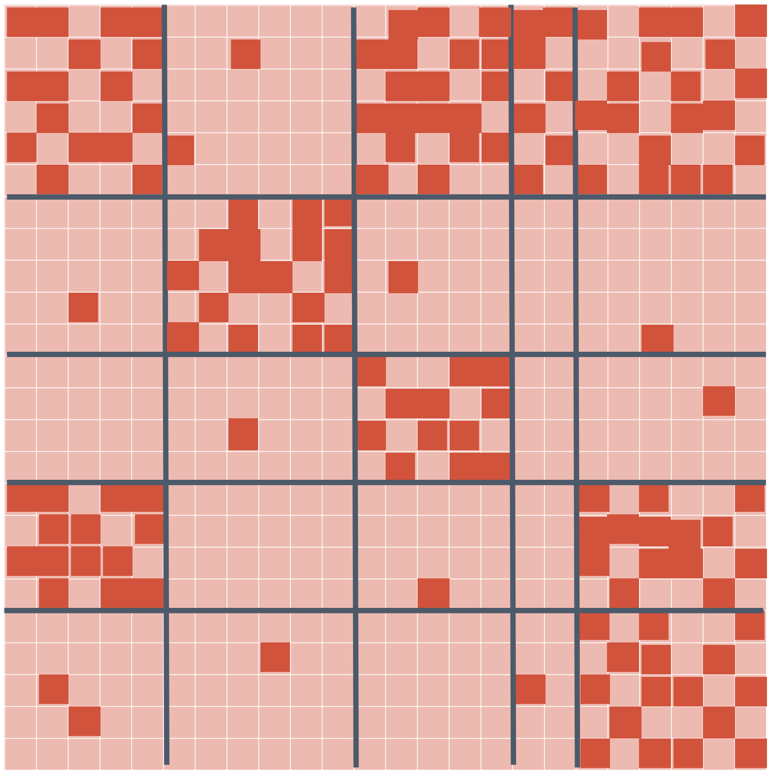
Near-Bipartite core



Subgraphs as submatrices



Subgraphs as submatrices



Co-clustering
and cross associations:
Partition matrix through clustering rows and columns.

Goal: Each block should have mostly similar cells

1. Subgraph Analysis

a) Background

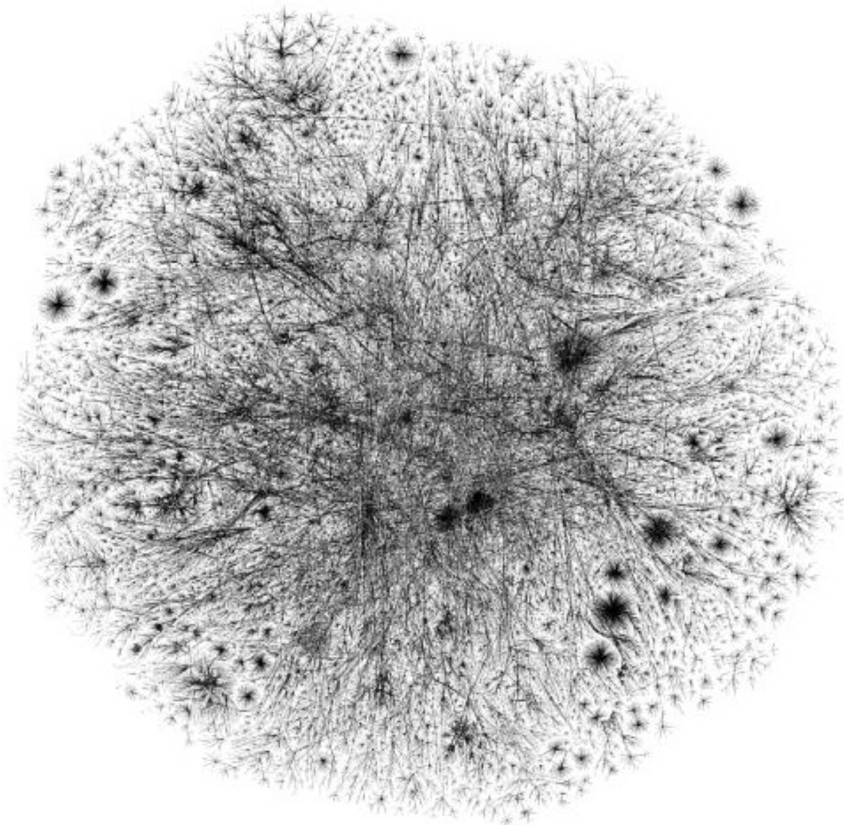
b) Normal Behavior

c) Abnormal Behavior

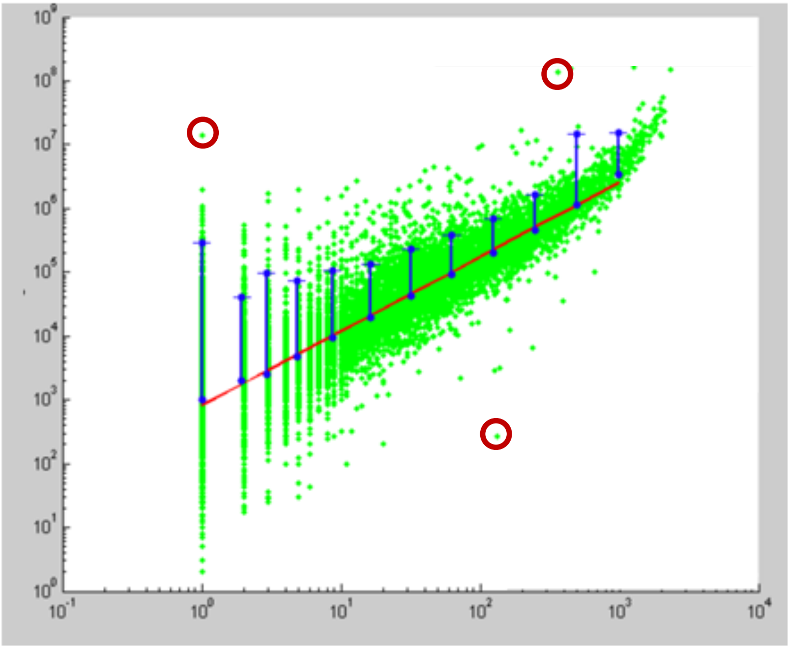
2. Propagation Methods

3. Latent Factor Models

Ego-net Patterns

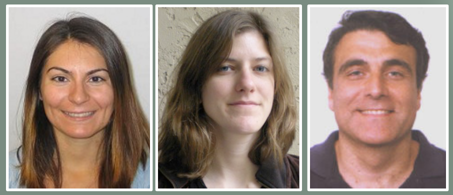


? ?



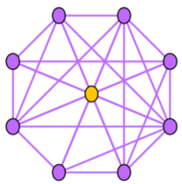
?

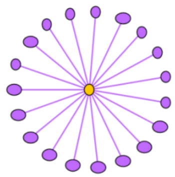
Oddball: Spotting anomalies in weighted graphs
Leman Akoglu, Mary McGlohon, Christos Faloutsos
PAKDD 2010



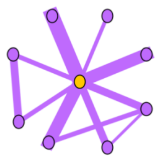
Ego-net Patterns

- N_i: number of neighbors (degree) of ego i
- E_i: number of edges in egonet i

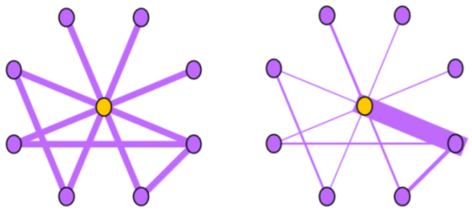




- W_i: total weight of egonet i

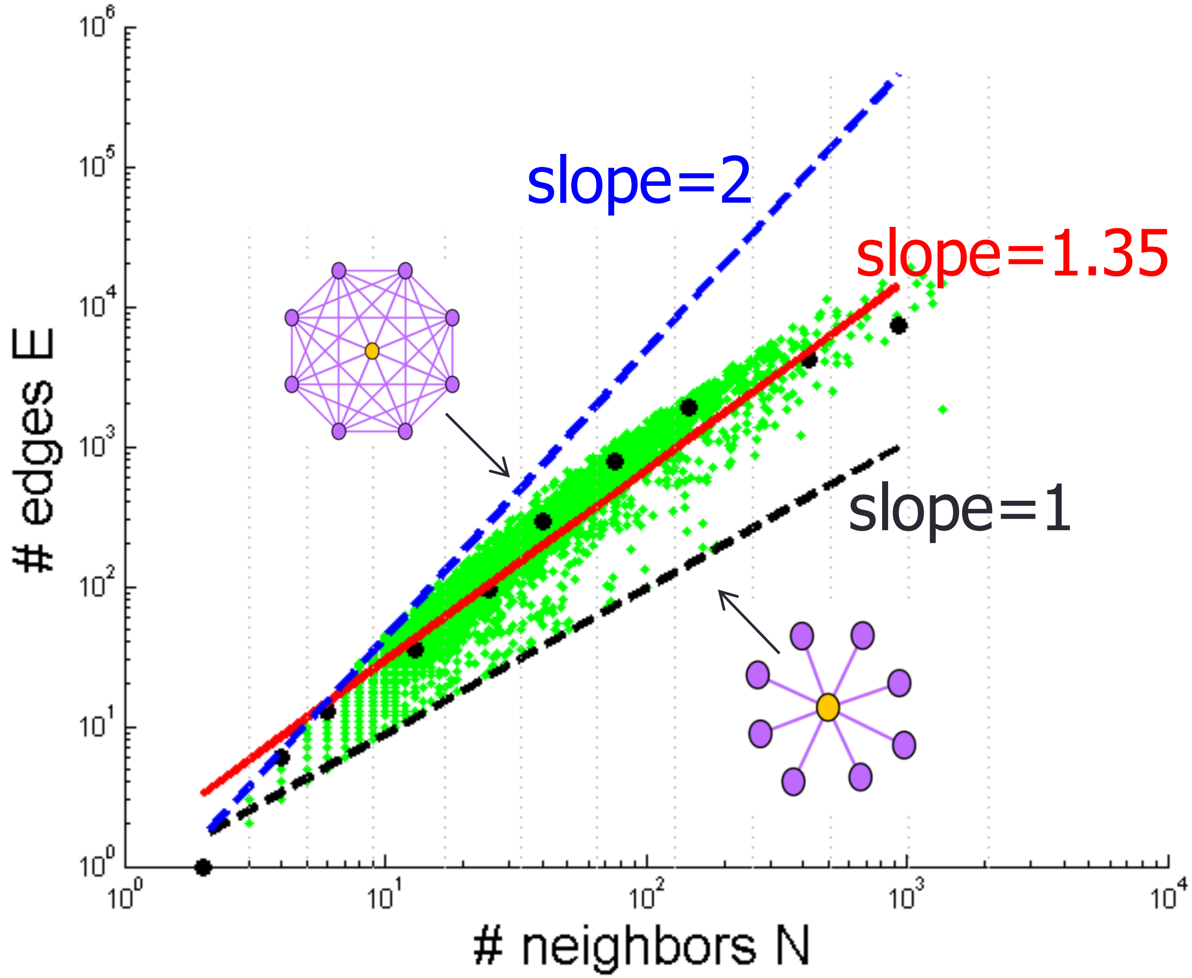


- $\lambda_{w,i}$: principal eigenvalue of the weighted adjacency matrix of egonet i



Oddball: Spotting anomalies in weighted graphs
Leman Akoglu, Mary McGlohon, Christos Faloutsos
PAKDD 2010

Pattern: Ego-net Power Law Density

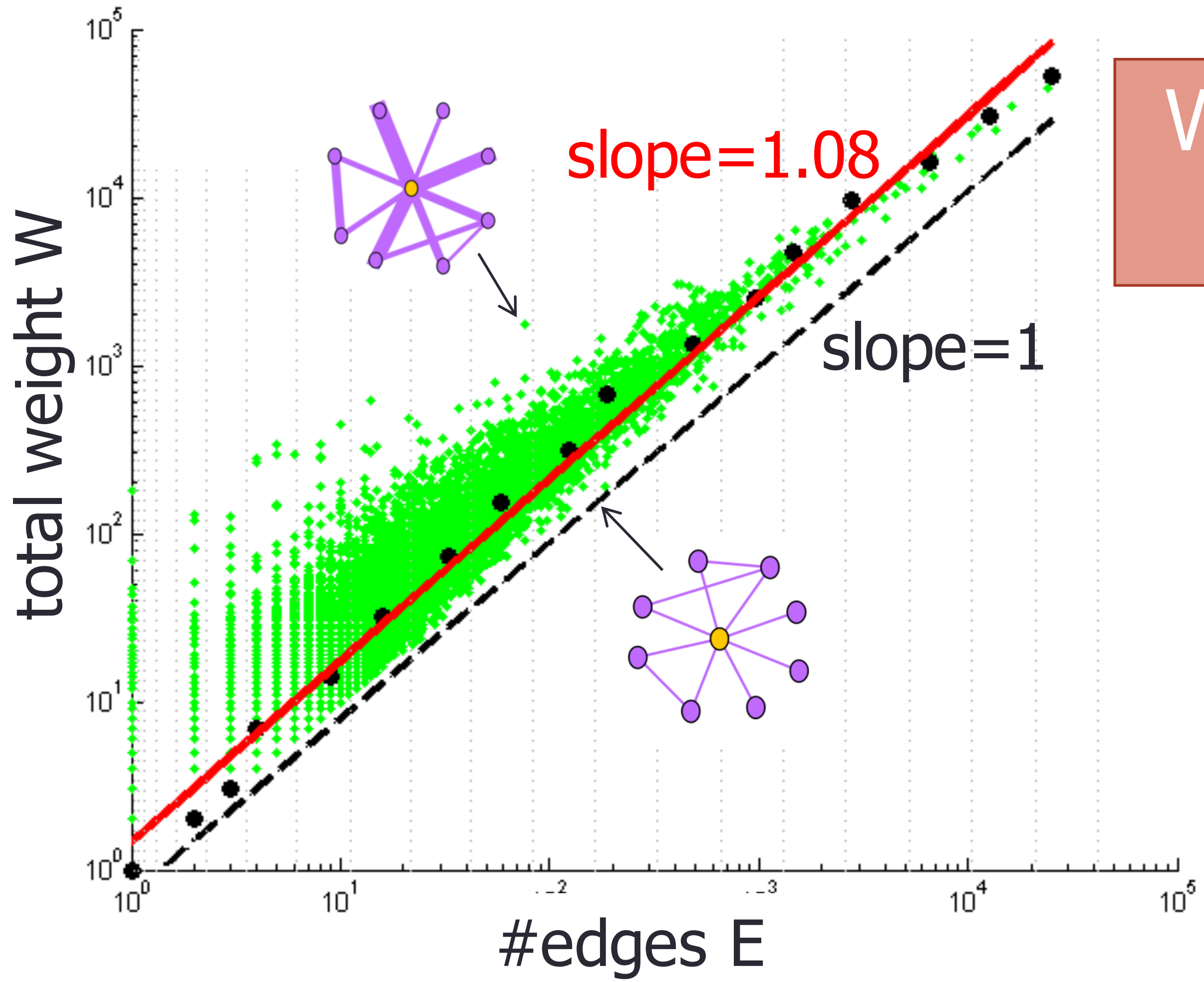


$E_i \propto N_i^{\alpha}$

$1 \leq \alpha \leq 2$

Oddball: Spotting anomalies in weighted graphs
Leman Akoglu, Mary McGlohon, Christos Faloutsos
PAKDD 2010

Pattern: Ego-net Power Law Weight

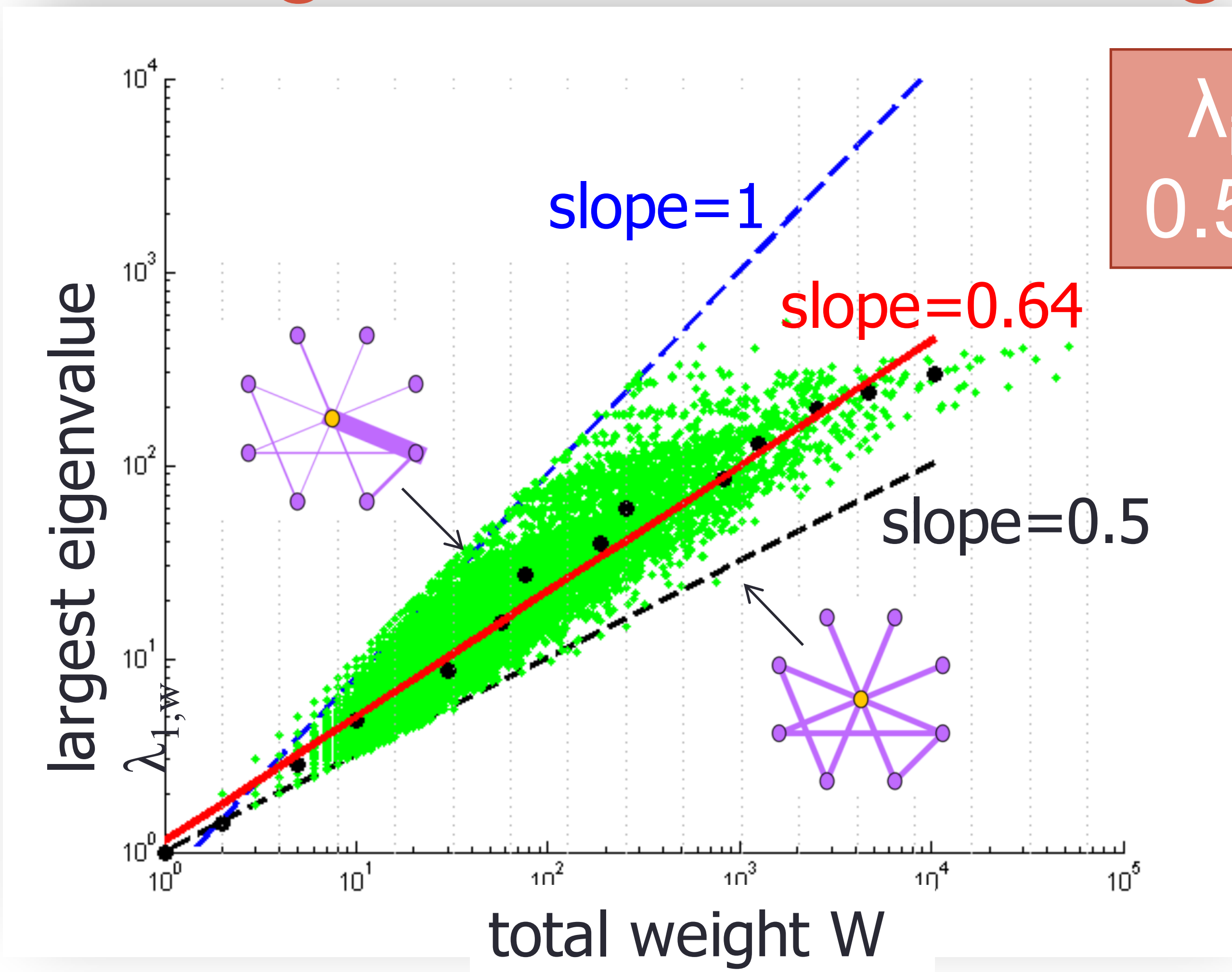


$W_i \propto E_i^{\alpha}$

$1 \leq \alpha$

Oddball: Spotting anomalies in weighted graphs
Leman Akoglu, Mary McGlohon, Christos Faloutsos
PAKDD 2010

Pattern: Ego-net Power Law Eigenvalue

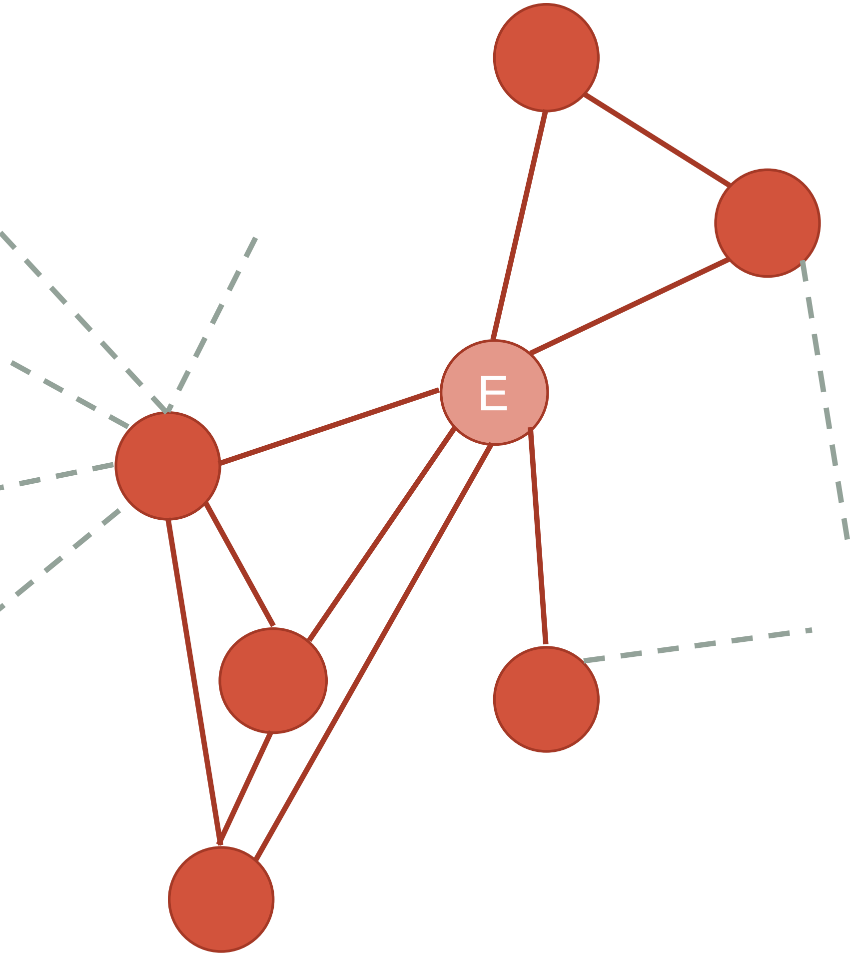


$\lambda_i \propto W_i^{\alpha}$

$0.5 \leq \alpha \leq 1$

Oddball: Spotting anomalies in weighted graphs
Leman Akoglu, Mary McGlohon, Christos Faloutsos
PAKDD 2010

Using graph patterns to find roles

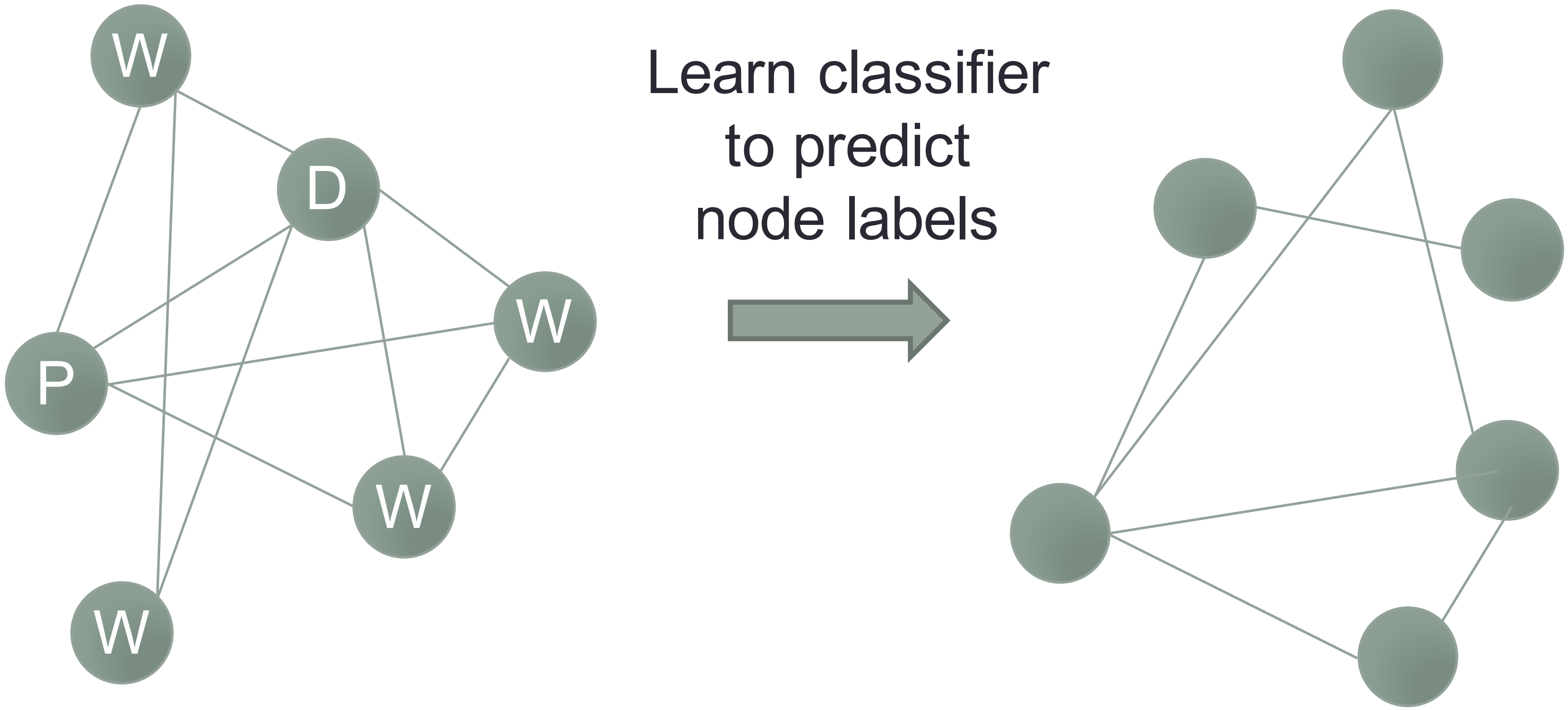


Useful node features:

- Degree
- Nodes in ego-net
- Edges in ego-net
- Edges leaving ego-net
- Mean of neighbor degree
- Sum of neighbor degree
- Expand recursively…

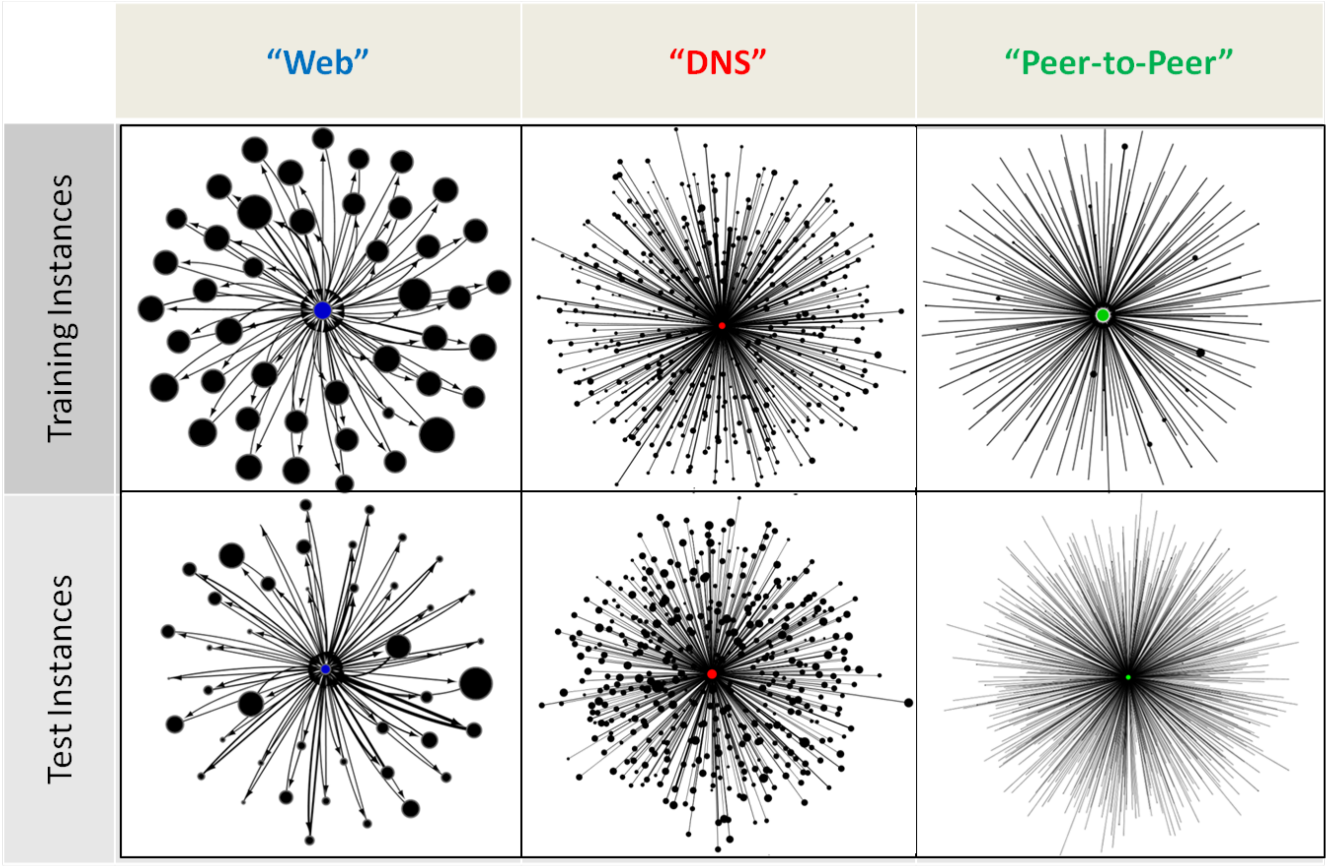
It's who you know: Graph mining using recursive structural features
K. Henderson, B. Gallagher, L. Li, L. Akoglu, T. Eliassi-Rad, H. Tong, C. Faloutsos
KDD 2011

Using graph patterns to find roles



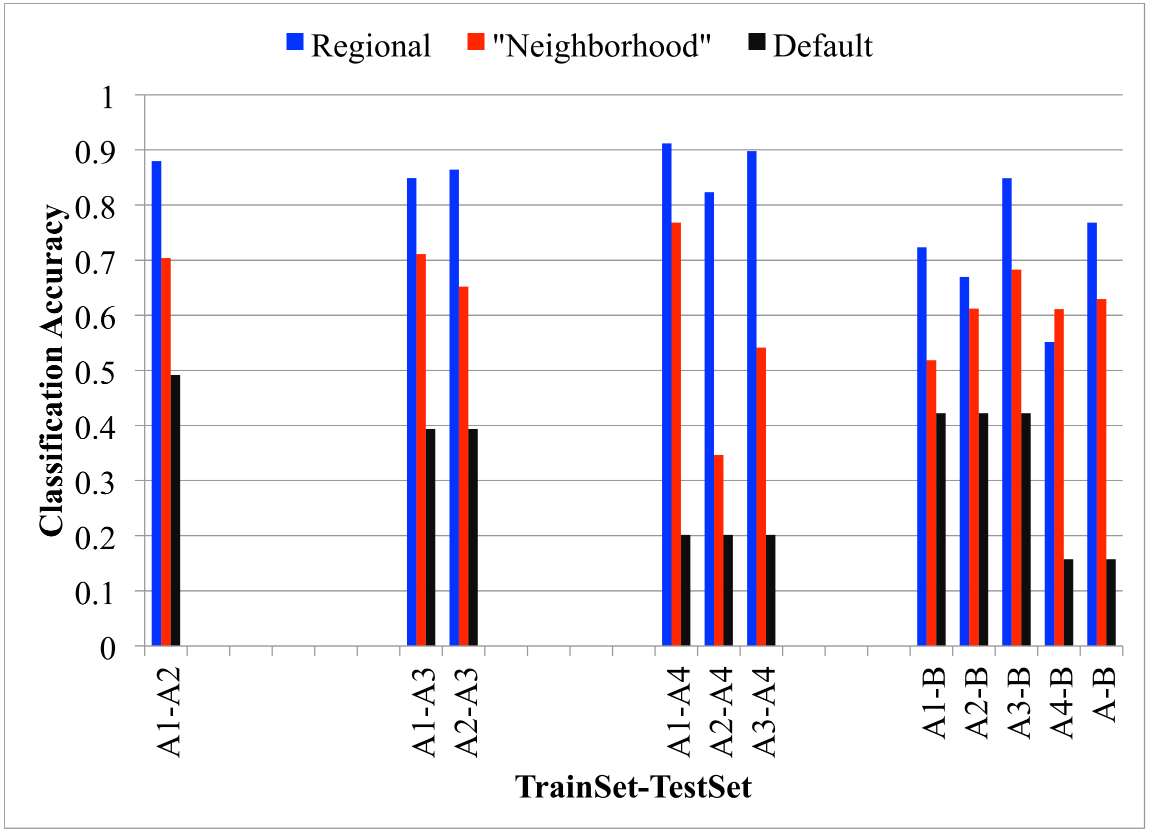
It's who you know: Graph mining using recursive structural features
Keith Henderson, Brian Gallagher, Lei Li, Leman Akoglu,
Tina Eliassi-Rad, Hanghang Tong, Christos Faloutsos
KDD 2011

Using graph patterns to find roles



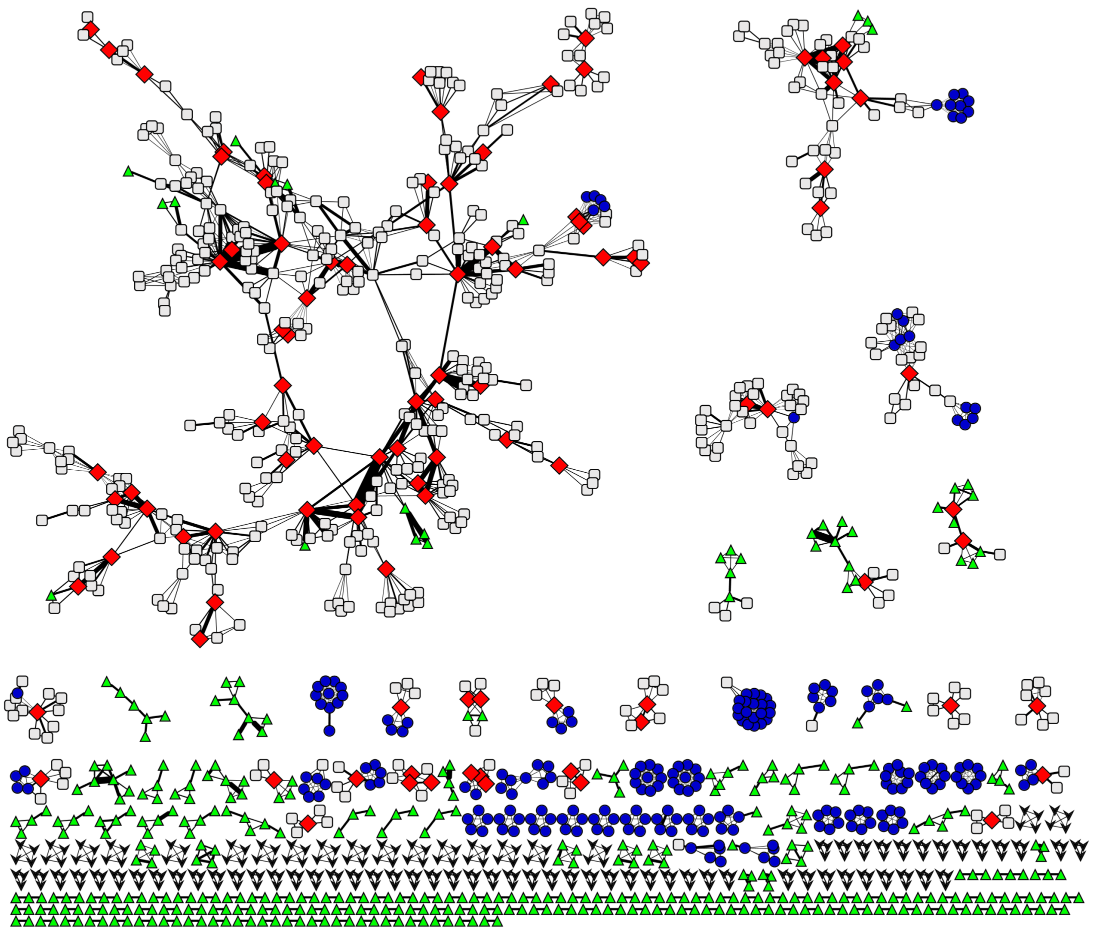
It's who you know: Graph mining using recursive structural features
Keith Henderson, Brian Gallagher, Lei Li, Leman Akoglu,
Tina Eliassi-Rad, Hanghang Tong, Christos Faloutsos
KDD 2011

Using graph patterns to find roles



It's who you know: Graph mining using recursive structural features
Keith Henderson, Brian Gallagher, Lei Li, Leman Akoglu,
Tina Eliassi-Rad, Hanghang Tong, Christos Faloutsos
KDD 2011

Using graph patterns to find roles

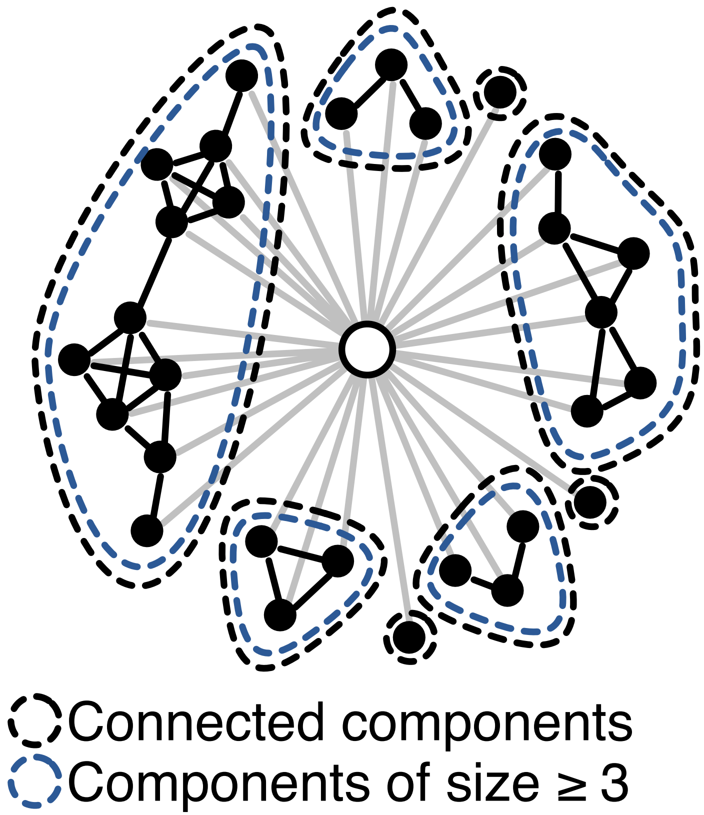


Use graph features to find similar types of behavior:

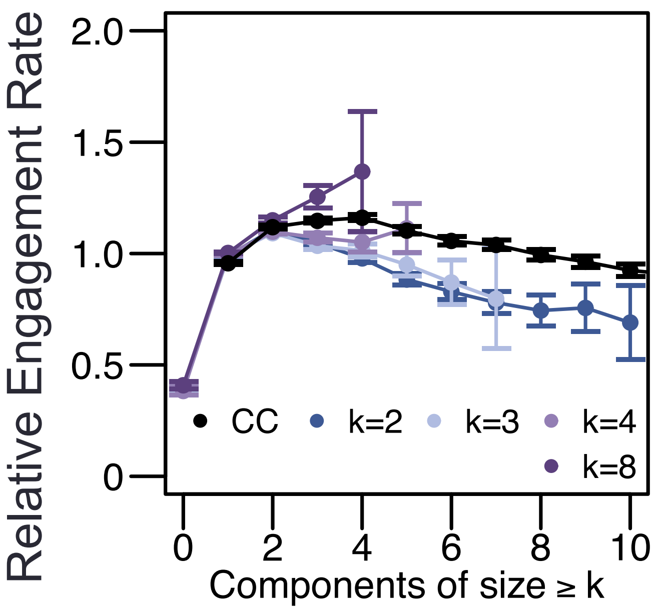
- Christos Faloutsos & Andrei Broder: tightly knit communities
- Albert-Laszlo Barabasi & Mark Newman: bridge communities
- John Hopcroft and Jon Kleinberg: mainstream
- Lada Adamic and Bernardo Huberman: elongated clusters

RolX: Structural Role Extraction & Mining in Large Graphs
K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, Sugato Basu, L.Akoglu, D. Koutra, C. Faloutsos, L. Li
KDD 2012

Using ego-nets to predict engagement



Number of connected components in egonet predicts engagement on Facebook

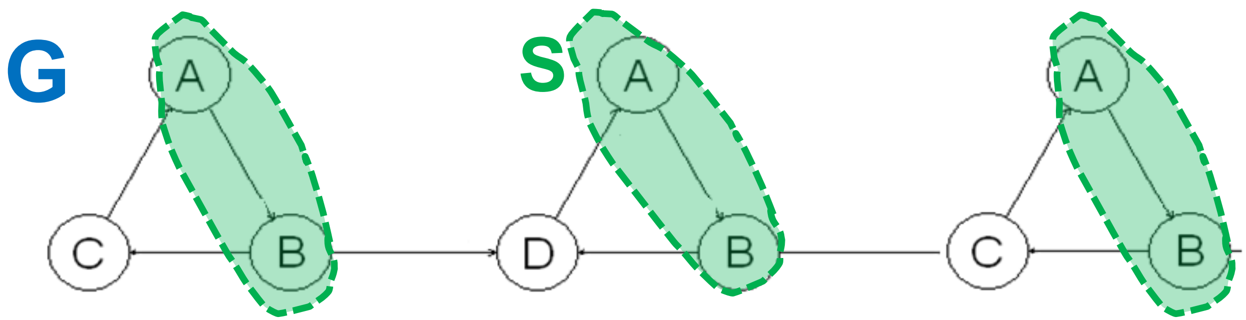


Structural diversity in social contagion
Johan Ugander, Lars Backstrom, Cameron Marlow, Jon Kleinberg
PNAS 2012

Attributed subgraph patterns

- SUBDUE: An algorithm for detecting repetitive patterns (substructures) within (single-attributed) graphs.
- The best substructure is the one that **minimizes**

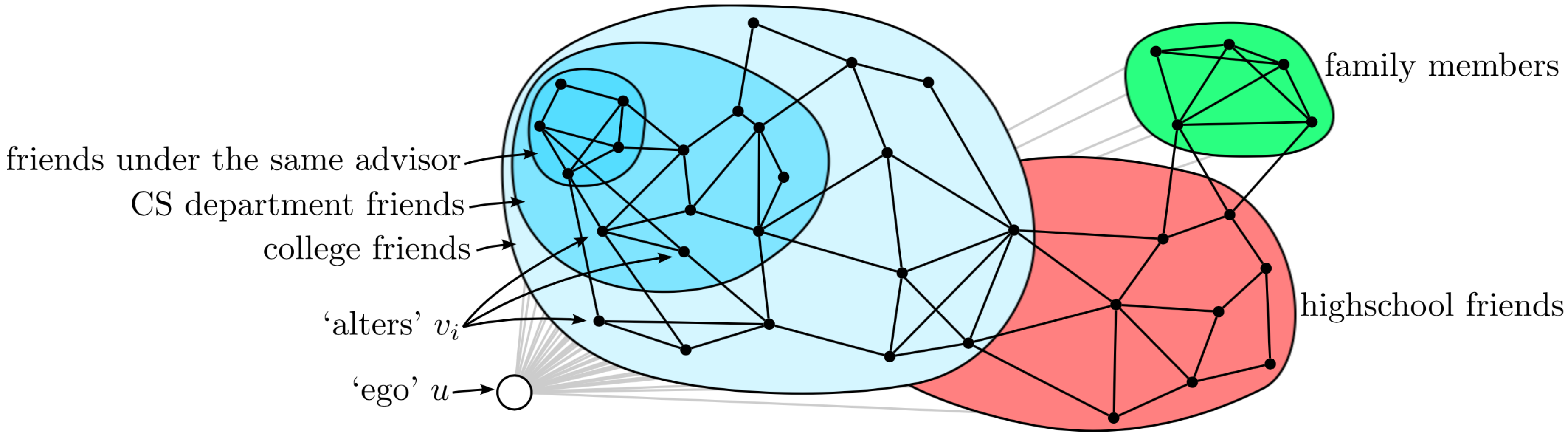
$$F1(\mathbf{S},\mathbf{G}) = DL(\mathbf{G} \mid \mathbf{S}) + DL(\mathbf{S})$$



- G: Entire graph, S: The substructure,
- DL(G|S) is the DL of G after compressing it using S,
- DL(S) is the description length of the substructure.

Graph-based Anomaly Detection
Caleb C. Noble and Diane Cook
KDD 2003

Friend groups within ego-nets

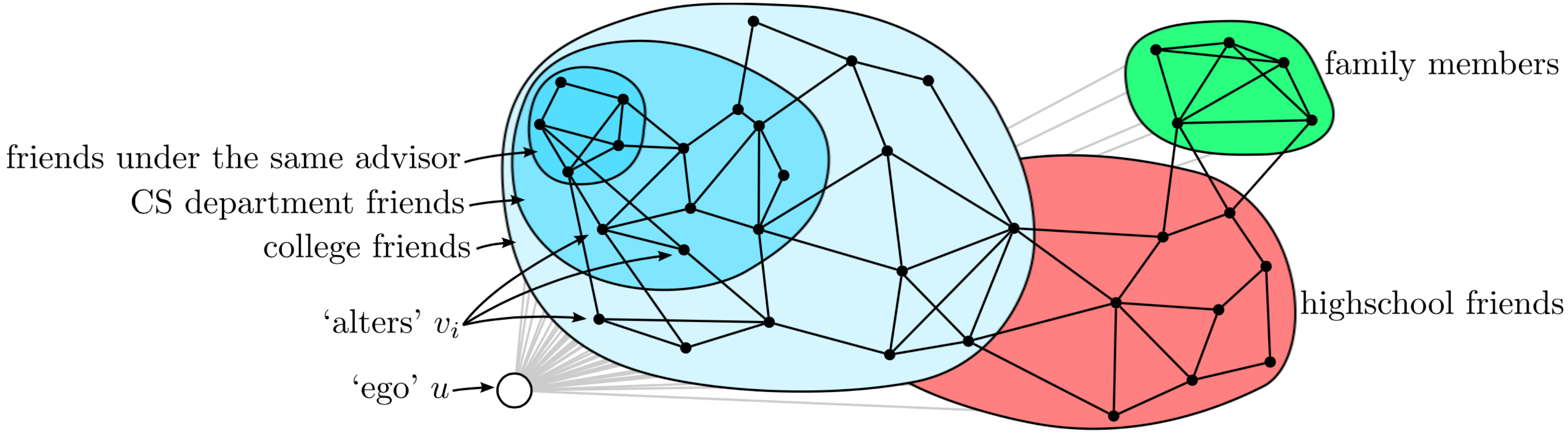


Learning to Discover Social Circles in Ego Networks
Julian McAuley, Jure Leskovec
NIPS 2012


Friend groups within ego-nets

Use node features to find clusters:

[Albert, Einstein, German, Princeton]

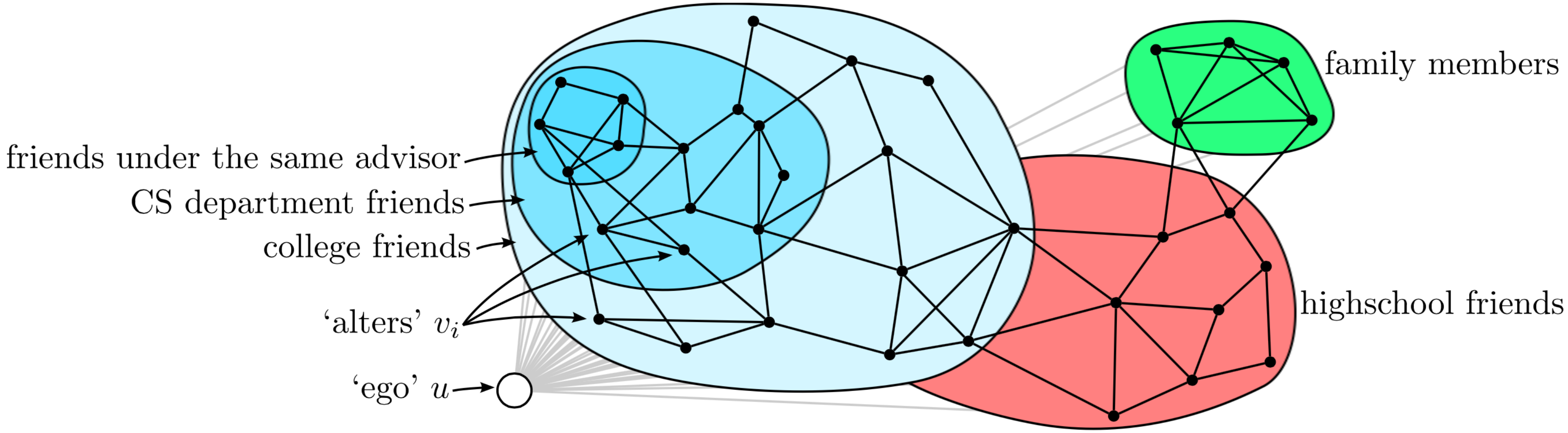


Learning to Discover Social Circles in Ego Networks
Julian McAuley, Jure Leskovec
NIPS 2012

Friend groups within ego-nets

Use node features to find clusters:

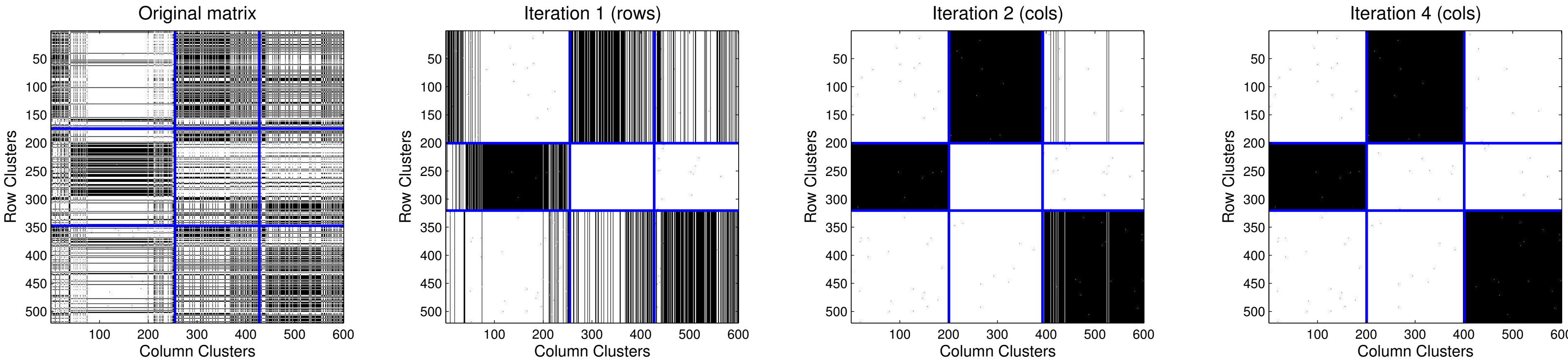
[Albert, Einstein, German, Princeton]



$$p((x,y) \in E) \propto \exp\Big\{ \underbrace{\sum_{C_k \supseteq \{x,y\}} \langle \phi(x,y), \theta_k \rangle}_{\text{circles containing both nodes}} - \underbrace{\sum_{C_k \nsupseteq \{x,y\}} \alpha_k \langle \phi(x,y), \theta_k \rangle}_{\text{all other circles}} \Big\}$$

Learning to Discover Social Circles in Ego Networks
Julian McAuley, Jure Leskovec
NIPS 2012

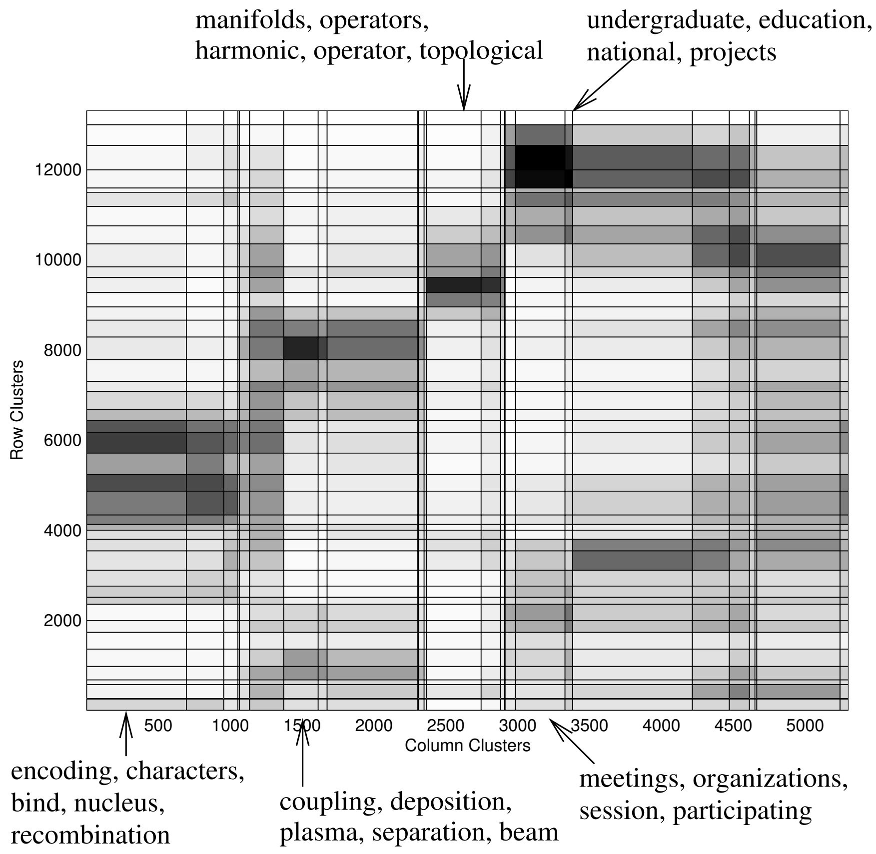
Modeling with Cross-Associations



Summarize binary matrices by
minimizing the number of bits to encode it.

Fully Automatic Cross-Associations
Deepayan Chakrabarti, Spiros Papadimitriou, Dharmendra S. Modha, Christos Faloutsos
KDD 2004

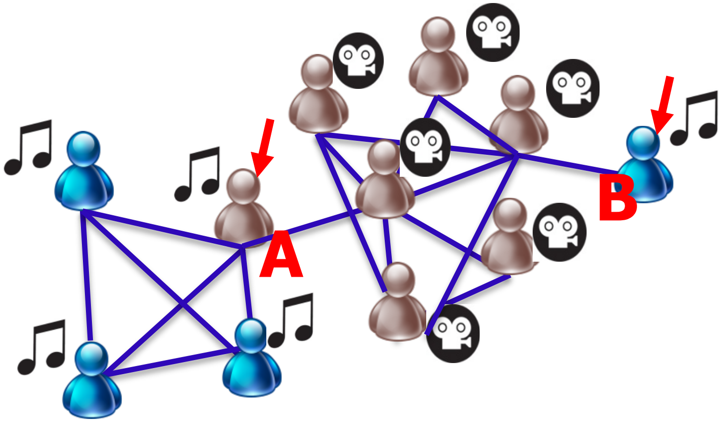
Modeling with Cross-Associations



Co-clustering of grant applications

Fully Automatic Cross-Associations
Deepayan Chakrabarti, Spiros Papadimitriou, Dharmendra S. Modha, Christos Faloutsos
KDD 2004

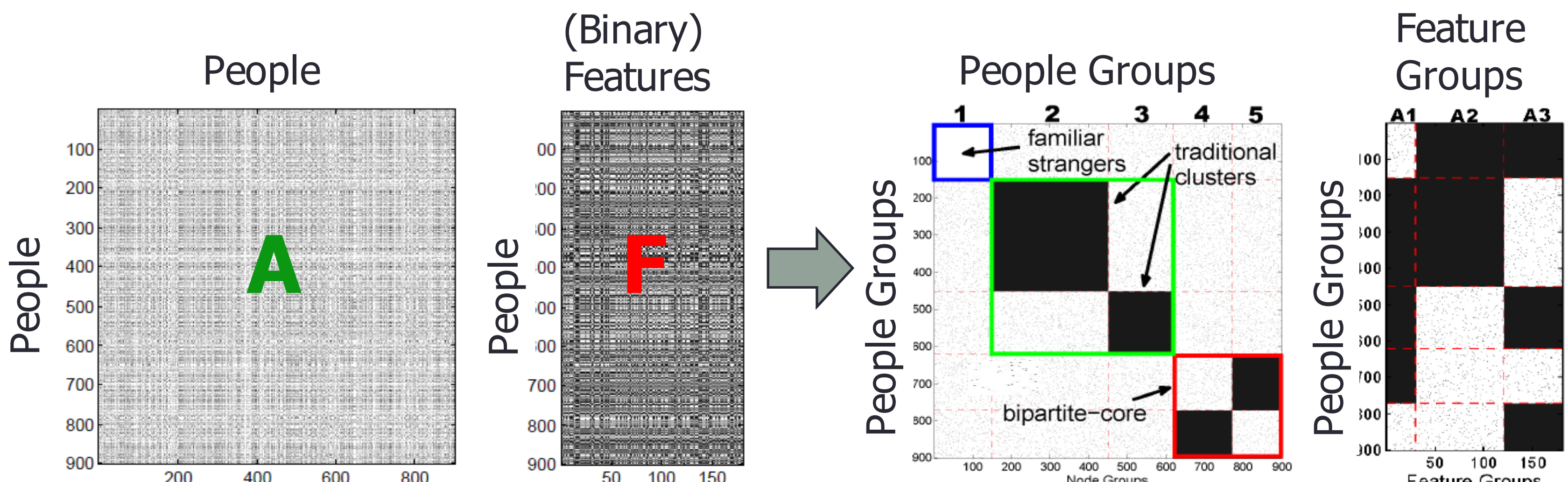
Joint co-clustering



- Cohesive clusters & anomalies

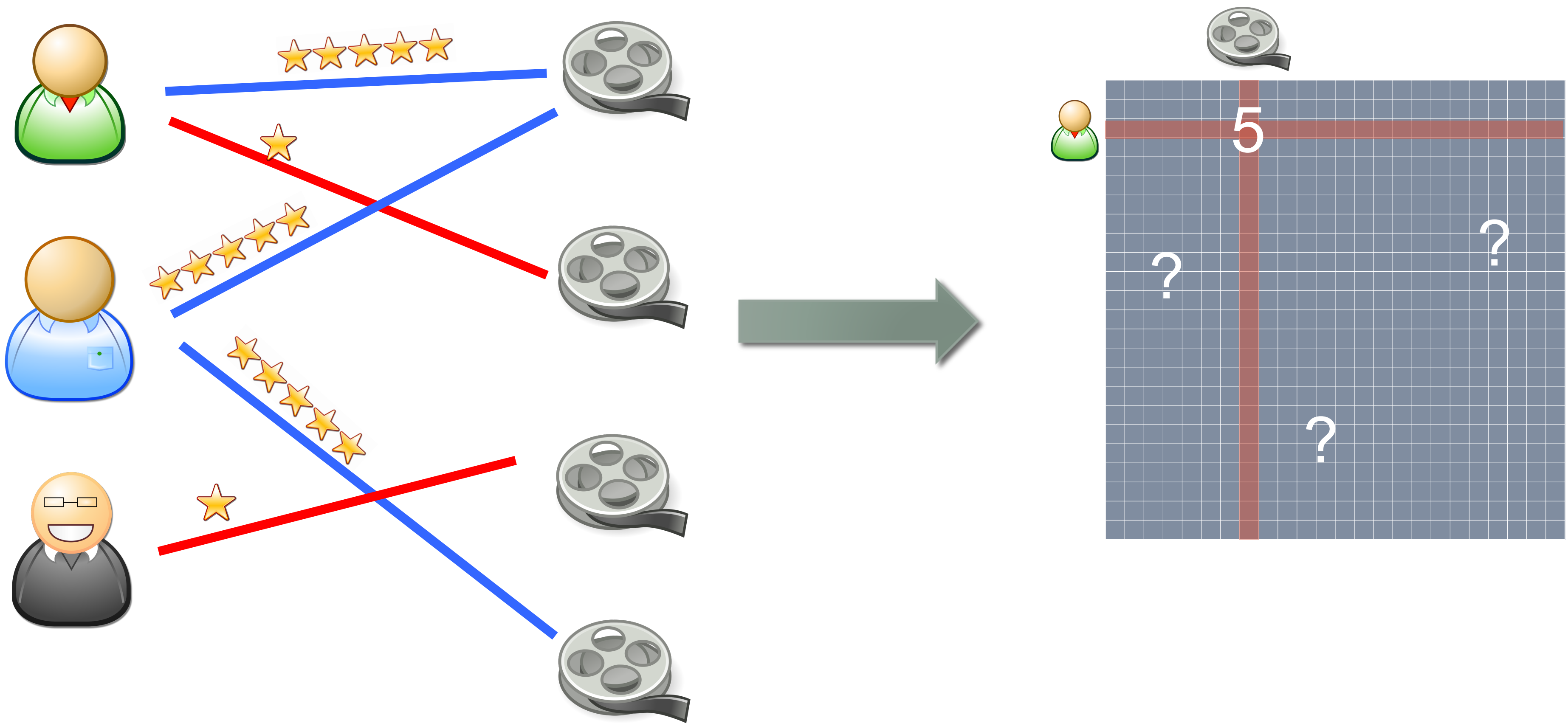
Given adjacency matrix **A** and feature matrix **F**

Find homogeneous blocks (clusters) in **A** and **F**



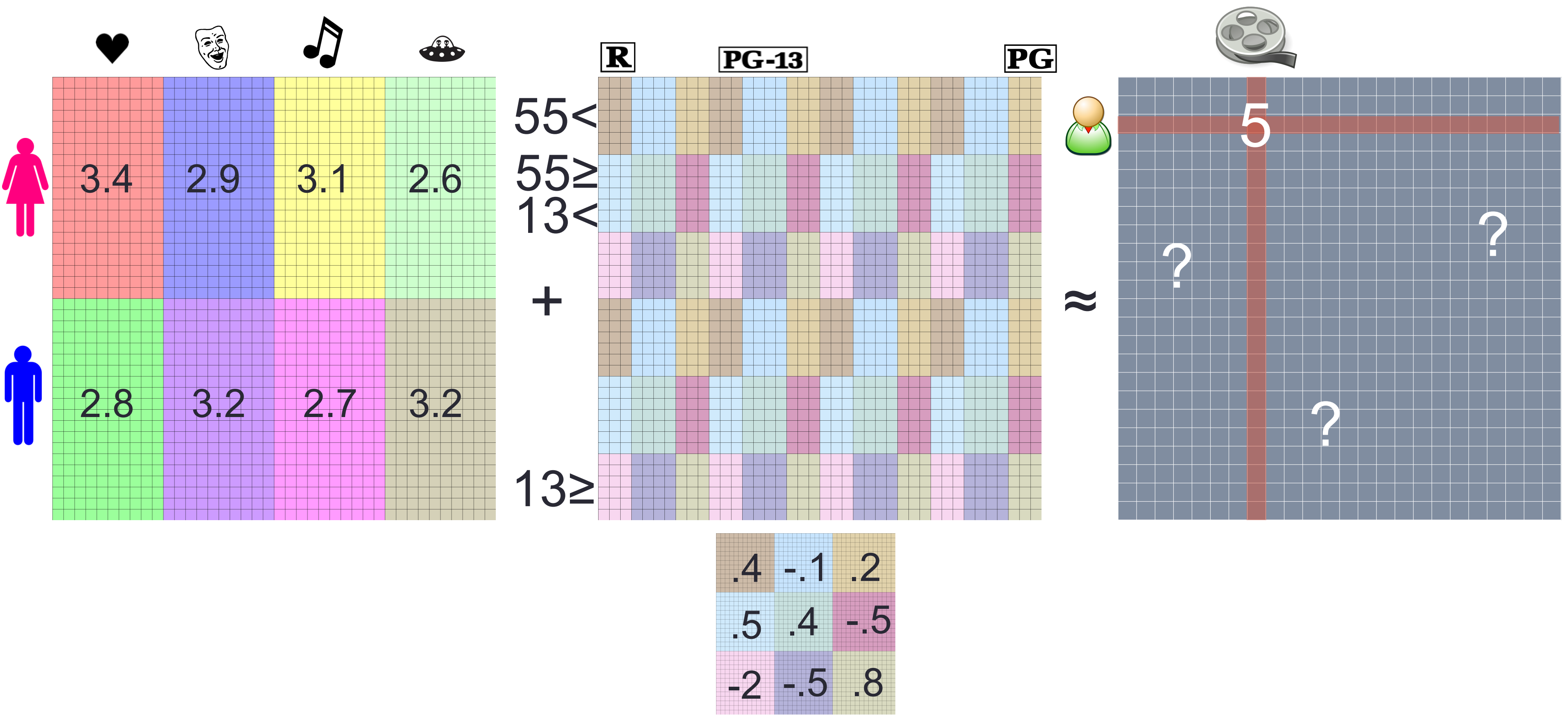
PICS: Parameter-free Identification of Cohesive Subgroups in Large Attributed Graphs. Leman Akoglu, Hanghang Tong, Brendan Meeder, Christos Faloutsos. *SDM* 2012

Prediction with Co-clustering



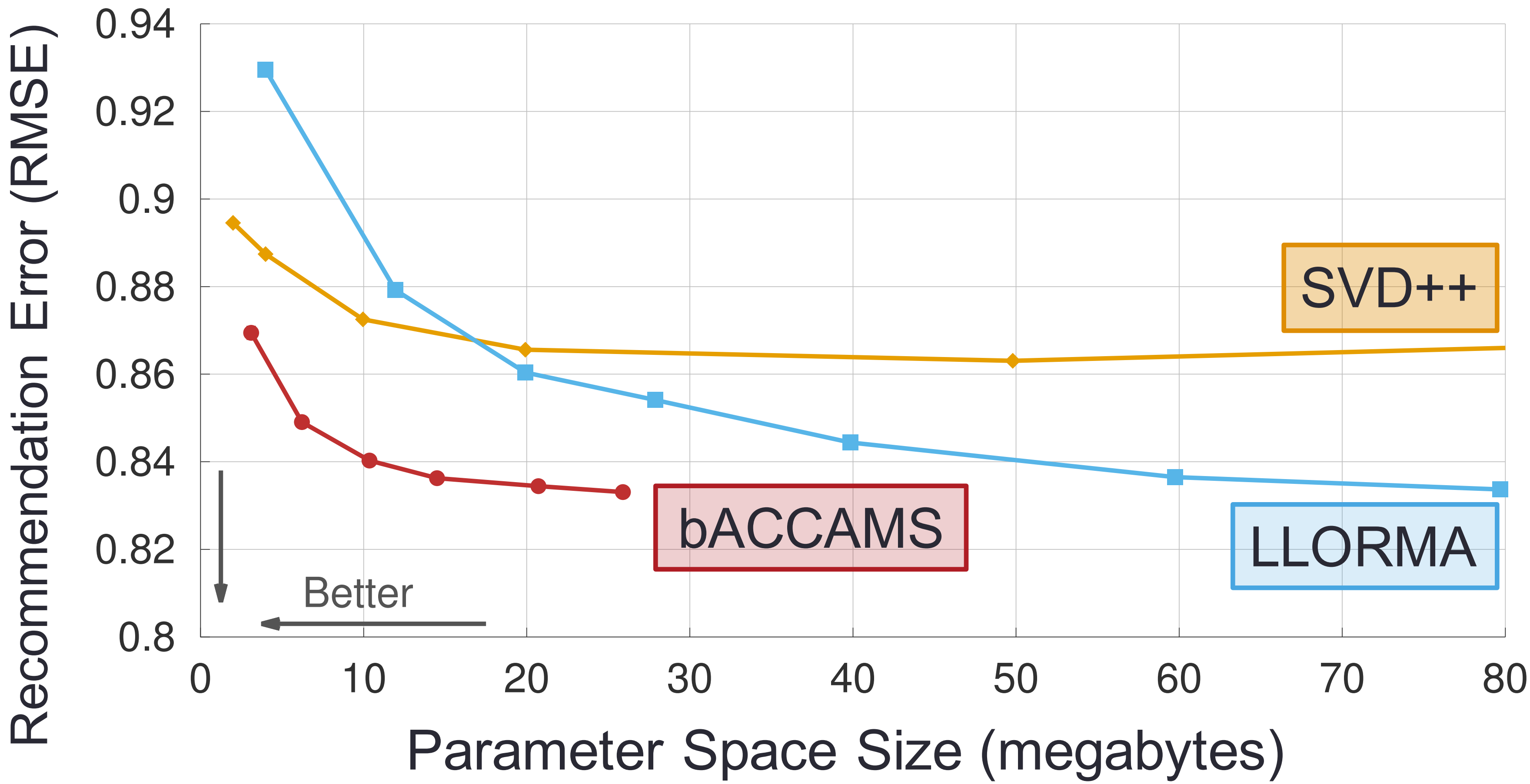
ACCAMS: Additive Co-clustering to Approximate Matrices Succinctly
Alex Beutel, Amr Ahmed, Alex Smola
WWW 2015

Prediction with Co-clustering



ACCAMS: Additive Co-clustering to Approximate Matrices Succinctly
Alex Beutel, Amr Ahmed, Alex Smola
WWW 2015

Modeling with Co-clustering



ACCAMS: Additive Co-clustering to Approximate Matrices Succinctly
Alex Beutel, Amr Ahmed, Alex Smola
WWW 2015

1. Subgraph Analysis

a) Background

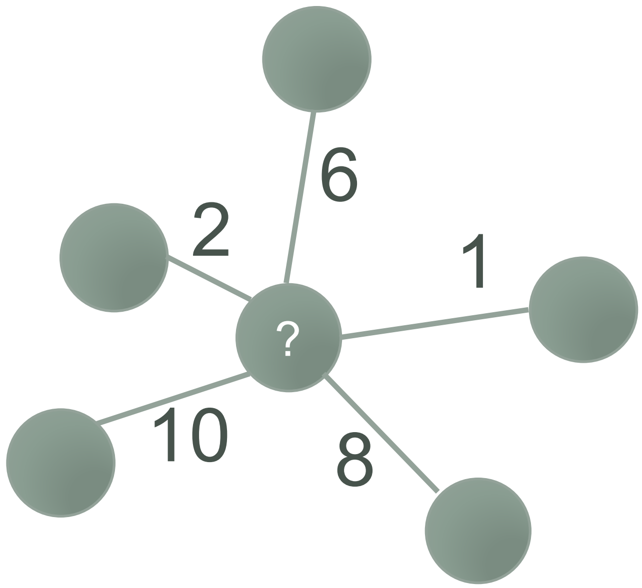
b) Normal Behavior

c) Abnormal Behavior

2. Propagation Methods

3. Latent Factor Models

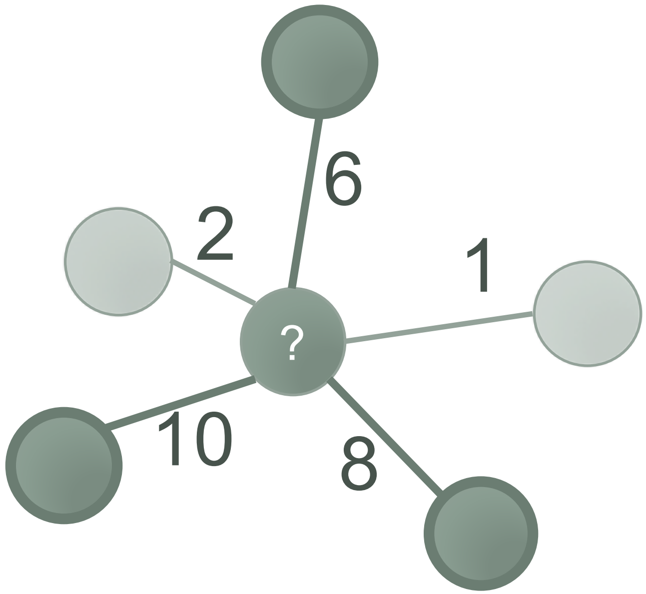
Fraud in Telecommunication Networks



- Community of Interest:
  - top-K connections

Communities of Interest
Corrinna Cortes, Daryl Pregibon, and Chris Volinsky
Springer, 2001

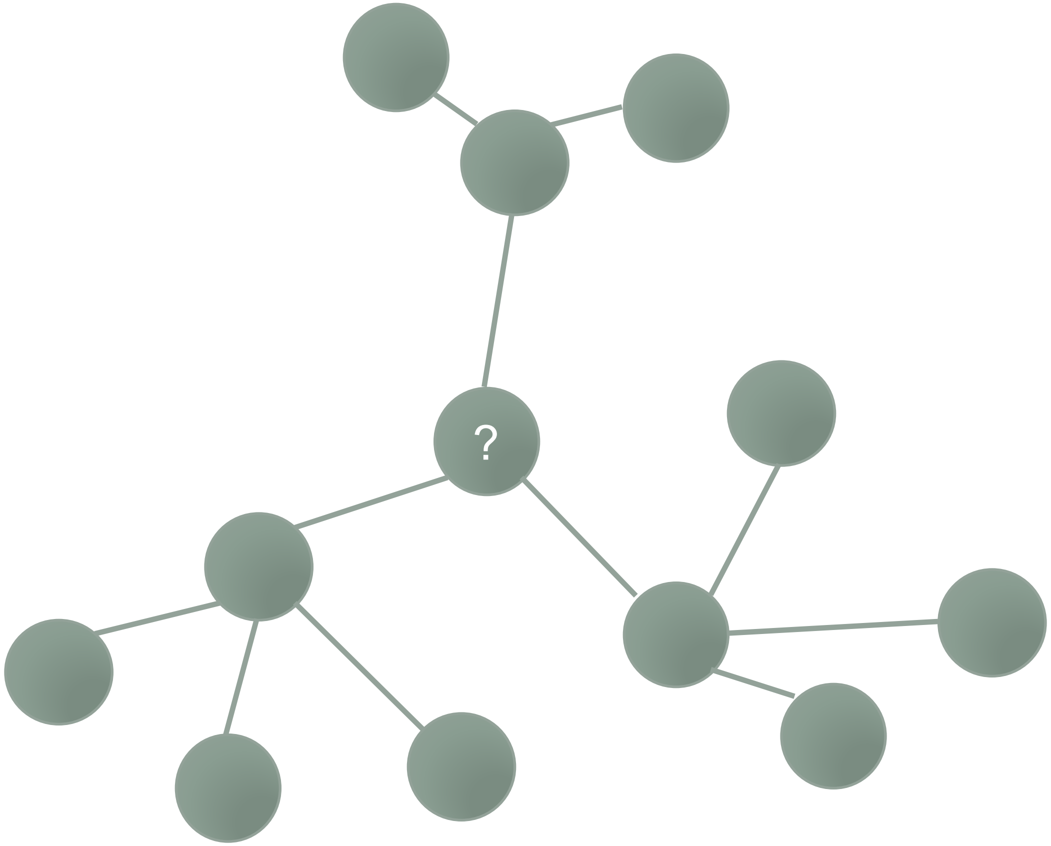
Fraud in Telecommunication Networks



- Community of Interest:
  - top-K connections

Communities of Interest
Corrinna Cortes, Daryl Pregibon, and Chris Volinsky
Springer, 2001

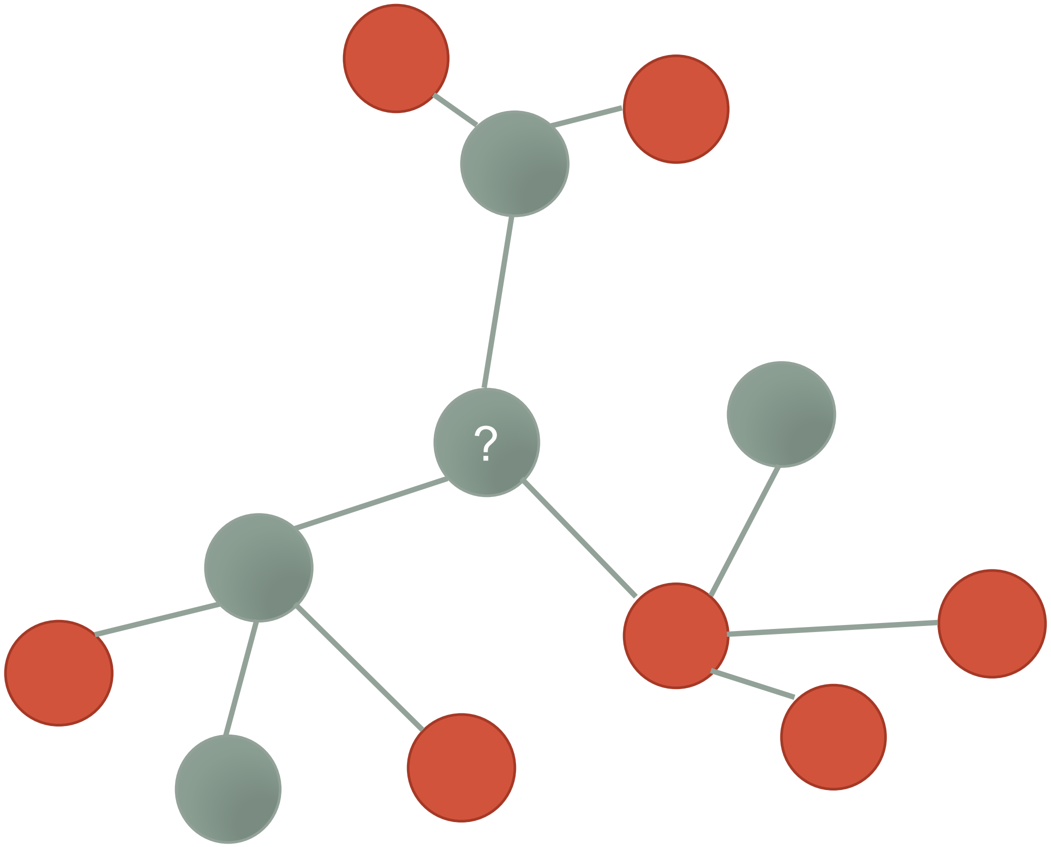
Fraud in Telecommunication Networks



- Community of Interest:
  - top-K connections
- d_2 community includes the COI for neighbors

Communities of Interest
Corrinna Cortes, Daryl Pregibon, and Chris Volinsky
Springer, 2001

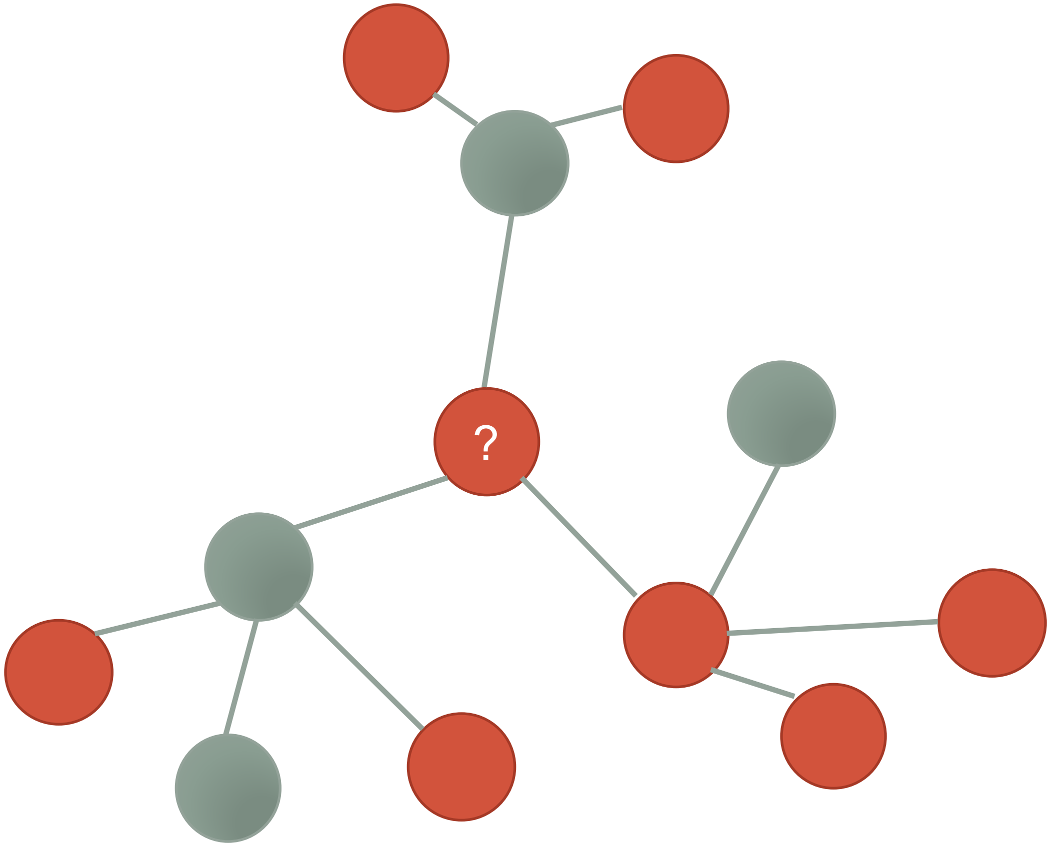
Fraud in Telecommunication Networks



- Community of Interest:
  - top-K connections
- d_2 community includes the COI for neighbors
- Label known fraudsters
- **Guilt-by-Association**
  - If most nodes in your d_2 community are fraudulent, you are probably fraudulent.

Communities of Interest
Corrinna Cortes, Daryl Pregibon, and Chris Volinsky
Springer, 2001

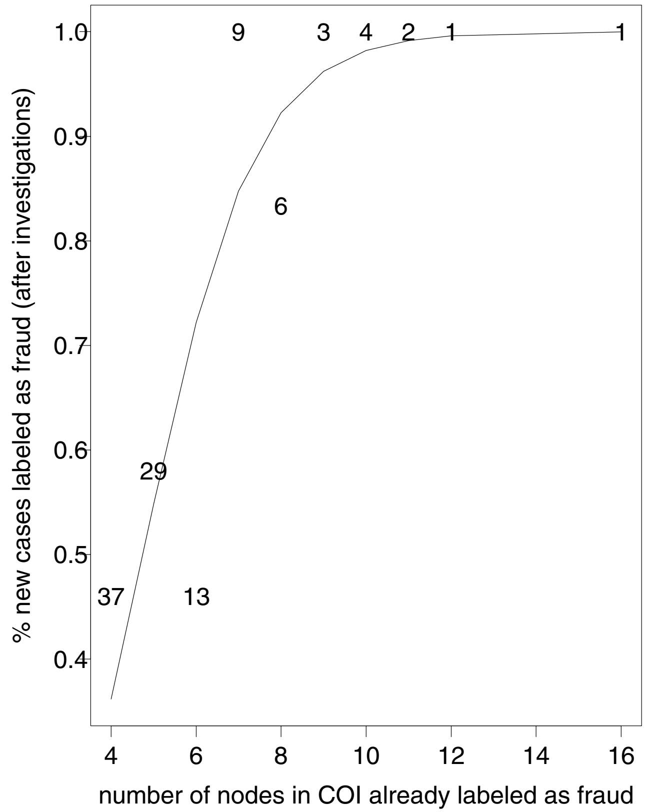
Fraud in Telecommunication Networks



- Community of Interest:
  - top-K connections
- d_2 community includes the COI for neighbors
- Label known fraudsters
- **Guilt-by-Association**
  - If most nodes in your d_2 community are fraudulent, you are probably fraudulent.

Communities of Interest
Corrinna Cortes, Daryl Pregibon, and Chris Volinsky
Springer, 2001

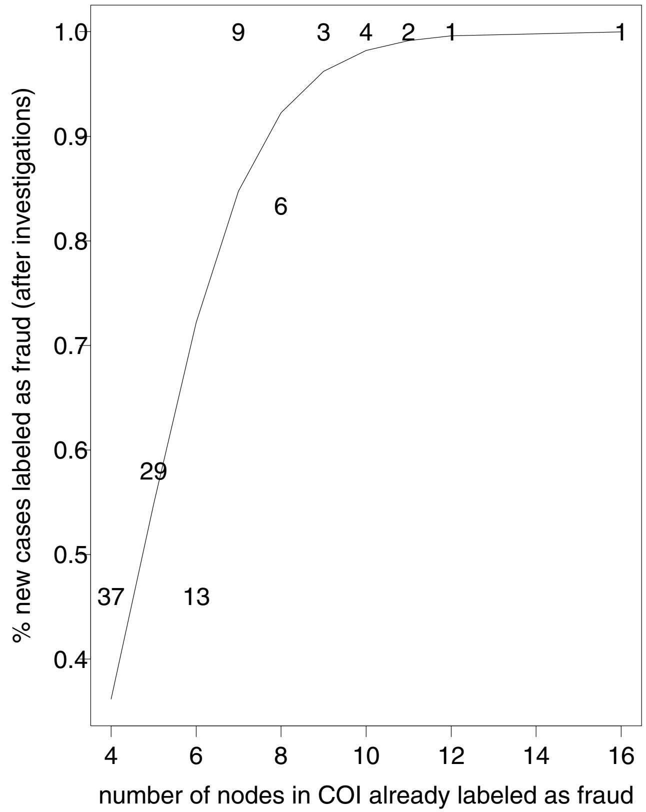
Fraud in Telecommunication Networks



- Community of Interest:
  - top-K connections
- d_2 community includes the COI for neighbors
- Label known fraudsters
- **Guilt-by-Association**
  - If most nodes in your d_2 community are fraudulent, you are probably fraudulent.

Communities of Interest
Corrinna Cortes, Daryl Pregibon, and Chris Volinsky
Springer, 2001

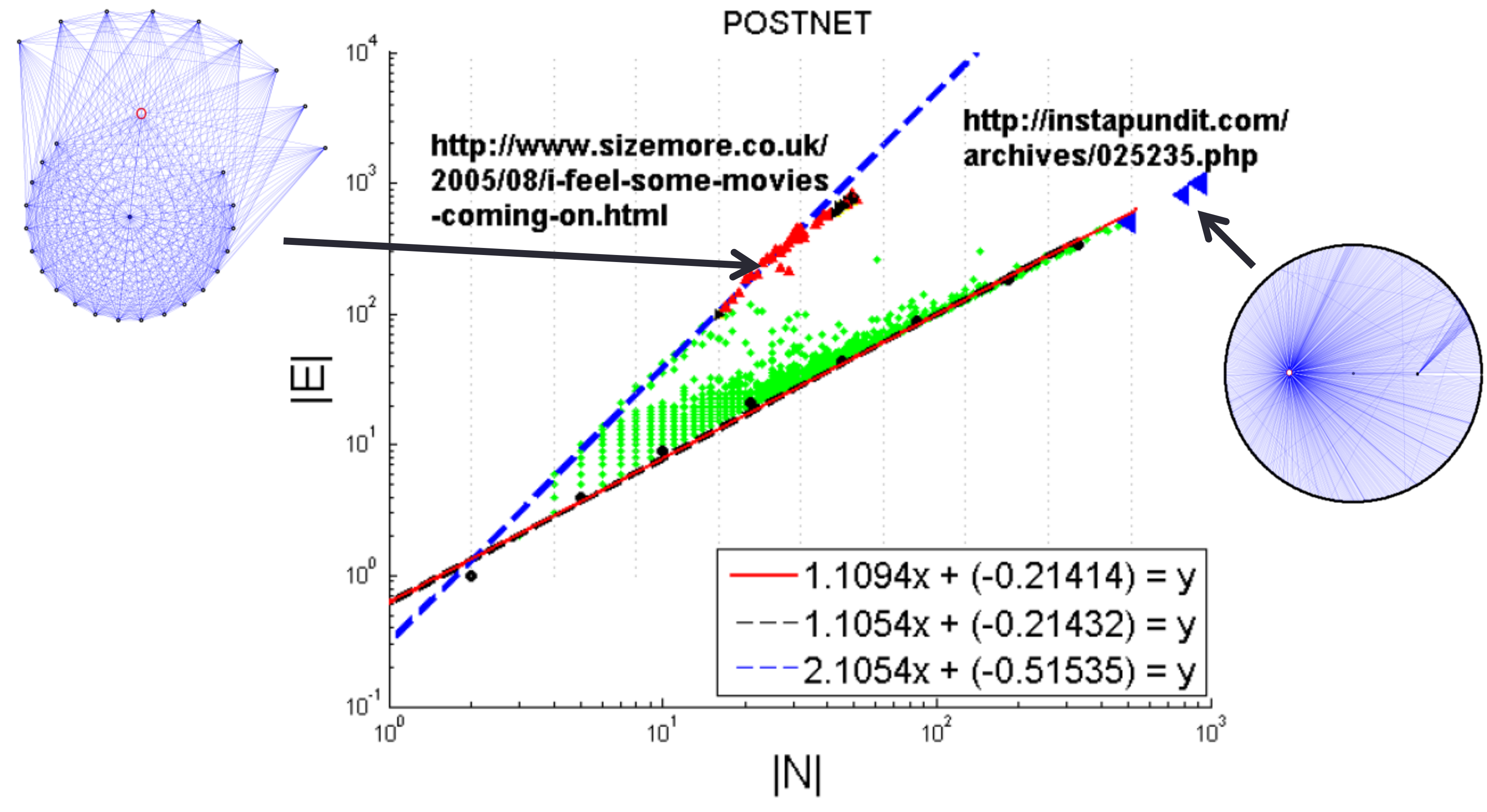
Fraud in Telecommunication Networks



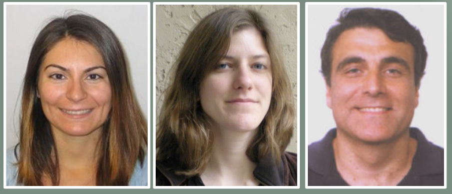
- Community of Interest:
  - top-K connections
- d_2 community includes the COI for neighbors
- Label known fraudsters
- **Guilt-by-Association**
  - **More "guilt-by-association" in next section**

Communities of Interest
Corrinna Cortes, Daryl Pregibon, and Chris Volinsky
Springer, 2001

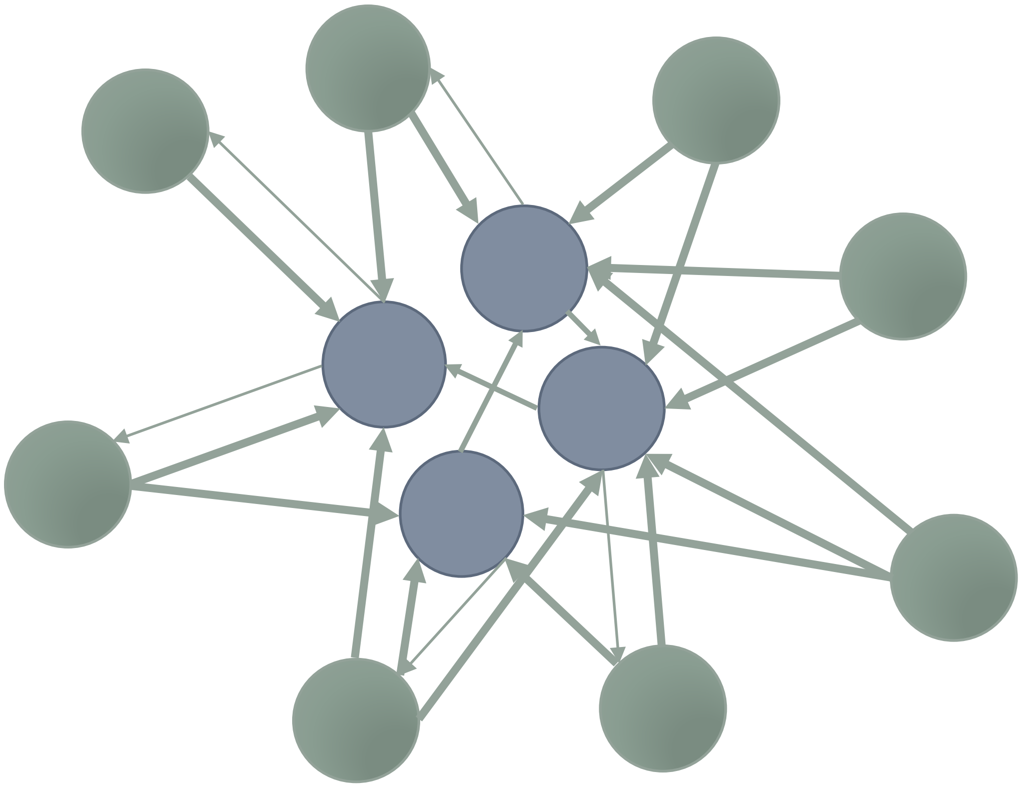
Pattern: Ego-net Power Law Density



Oddball: Spotting anomalies in weighted graphs
Leman Akoglu, Mary McGlohon, Christos Faloutsos
PAKDD 2010



Suspicious Subgraphs in Finance

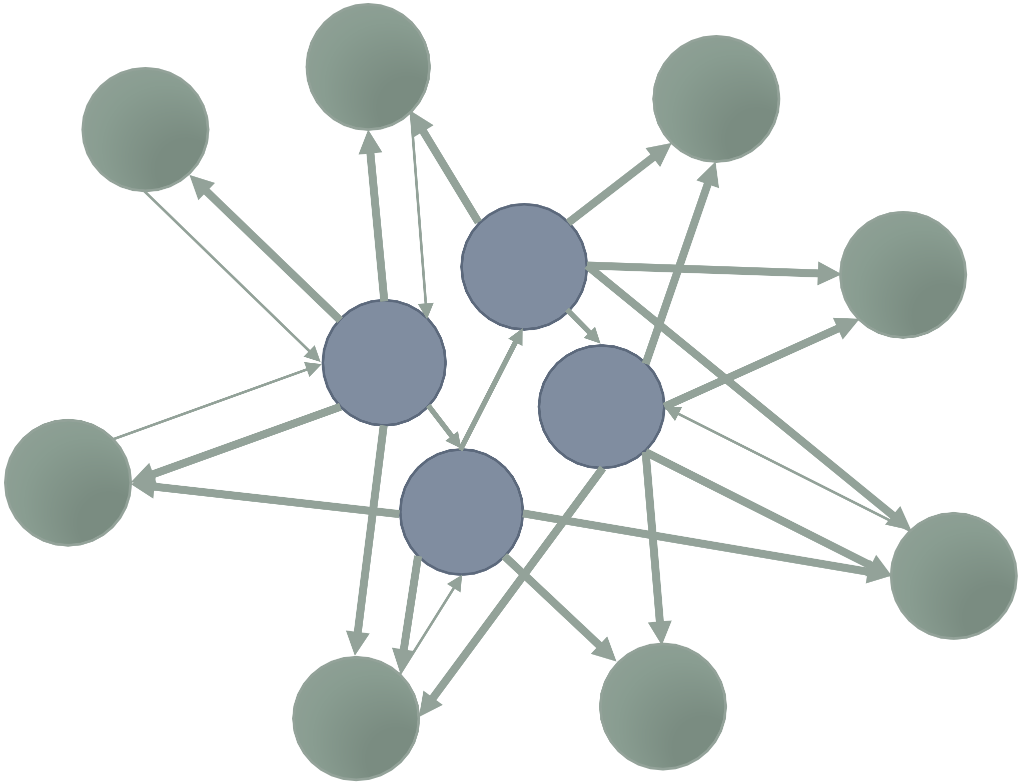


Blackhole:
Group of nodes with far more incoming weight than outgoing.

Could be indicative of trading ring buying up stock

Detecting Blackholes and Volcanoes in Directed Networks
Zhongmou Li, Hui Xiong, Yanchi Liu
ICDM 2010

Suspicious Subgraphs in Finance

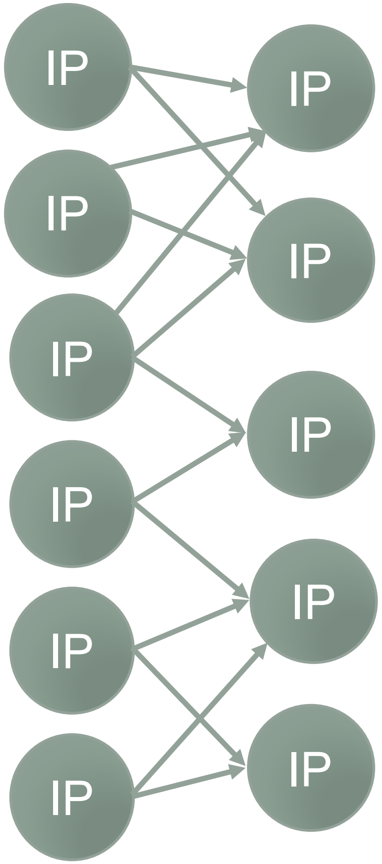


Volcano:
Group of nodes with far more outgoing weight than incoming.

Could be indicative of trading ring selling off inflated stock

Detecting Blackholes and Volcanoes in Directed Networks
Zhongmou Li, Hui Xiong, Yanchi Liu
ICDM 2010

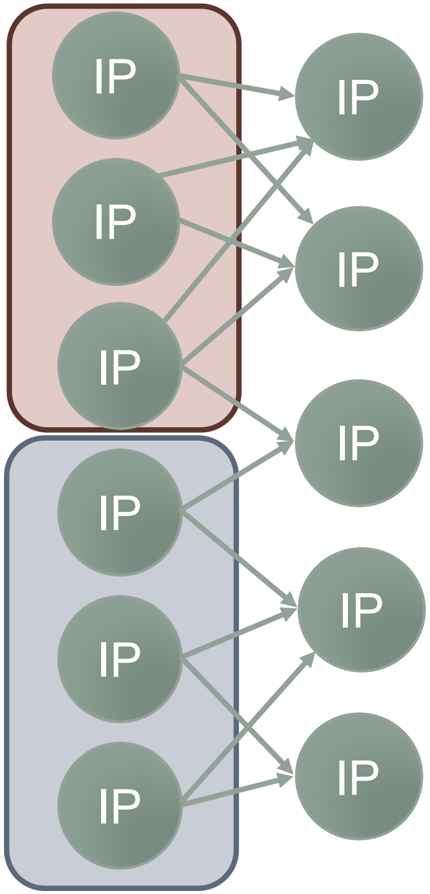
Graph Cuts for Intrusion Detection



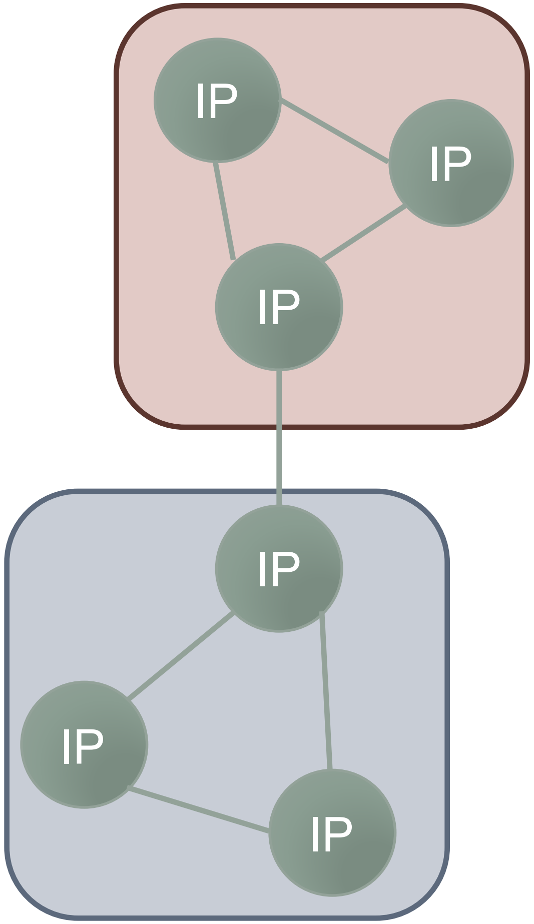
Bipartite graph between source IPs and destination IPs

Intrusion as (Anti)social Communication: Characterization and Detection
Qi Ding, Natallia Katenka, Paul Barford,
Eric Kolaczyk, Mark Crovella
KDD 2012

Graph Cuts for Intrusion Detection

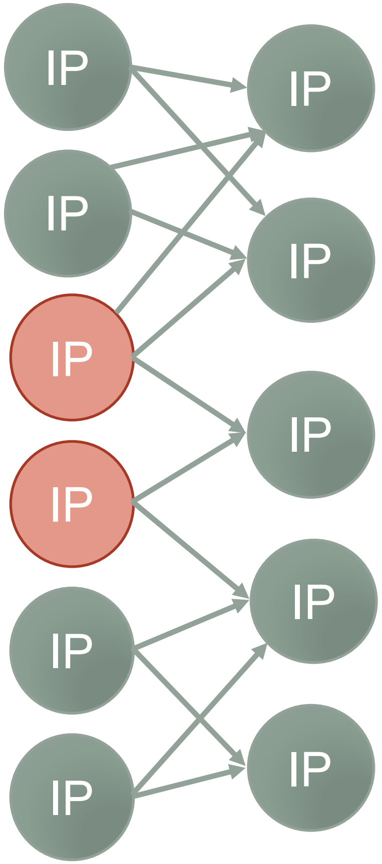


Connect source IPs if they connect to same destinations



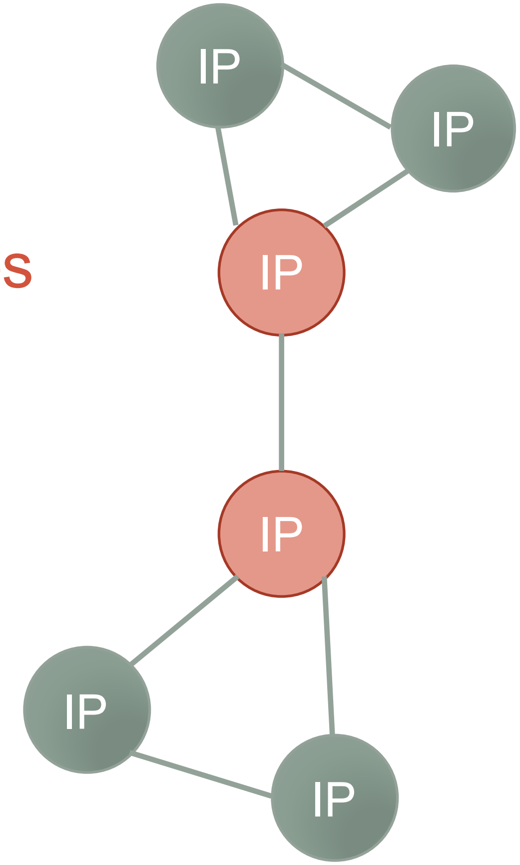
Intrusion as (Anti)social Communication: Characterization and Detection
Qi Ding, Natallia Katenka, Paul Barford,
Eric Kolaczyk, Mark Crovella
KDD 2012

Graph Cuts for Intrusion Detection



Nodes that
cross communities
are suspicious

Use min-cut to
find graph cuts



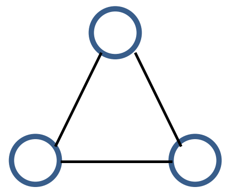
Intrusion as (Anti)social Communication: Characterization and Detection
Qi Ding, Natallia Katenka, Paul Barford,
Eric Kolaczyk, Mark Crovella
KDD 2012

Practitioner’s Guide to Detecting Fraud

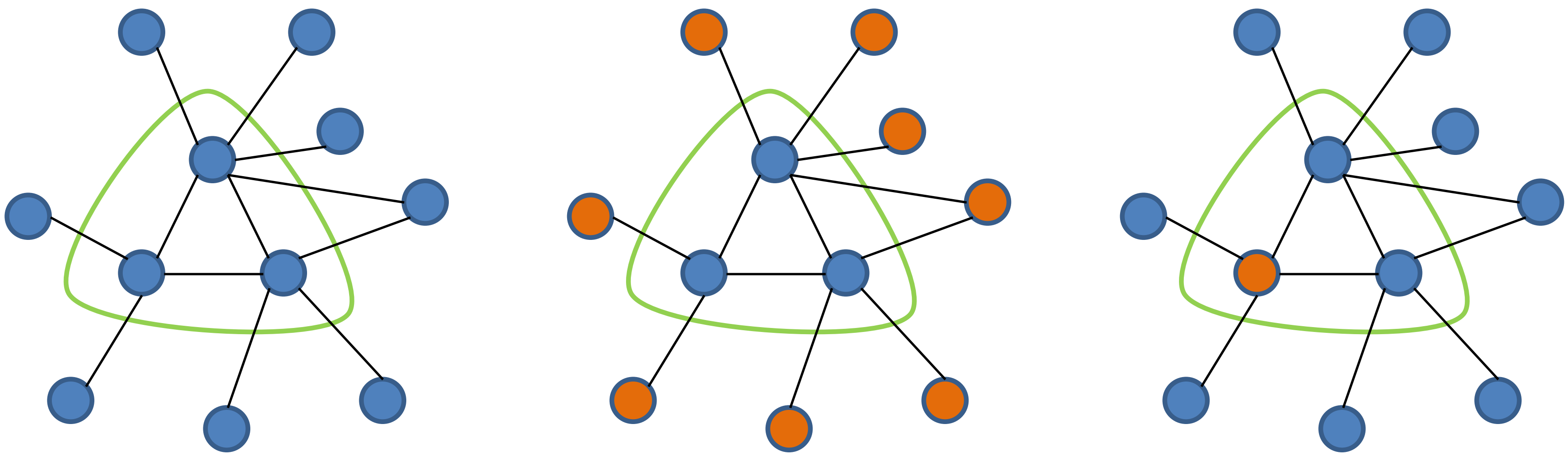
Method	Graph Type	Node Attributes	Edge Attributes	Seed Labels
COI	Undirected			✓
OddBall	Undirected			
Blackholes & Volcanoes	Directed			
(Anti)-Social	Bipartite			
SODA	Undirected	✓		
FocusCO	Undirected	✓		
CopyCatch	Bipartite		✓	
SynchroTrap	Bipartite+	✓	✓	
CrossSpot	Bipartite		✓	
Co-Clustering	Bipartite*		✓	

Outlier Detection in Attributed Subgraphs

User query:
3-author clique



- Data Mining Author
- Theory Author

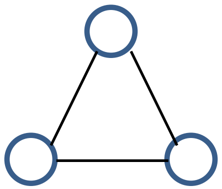


Normal | Anomalous | Anomalous

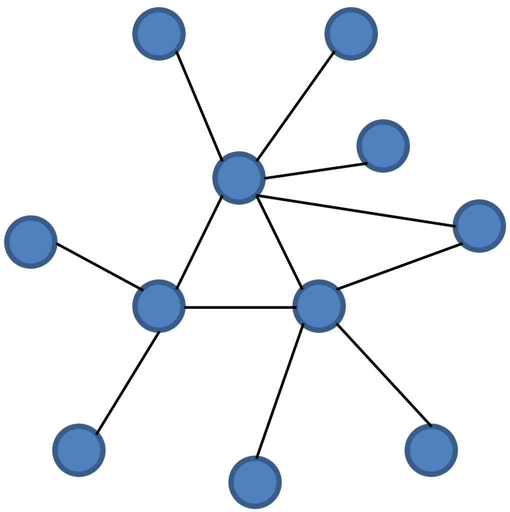
Local Learning for Mining Outlier Subgraphs from Network Datasets
Manish Gupta, Arun Mallya, Subhro Roy,
Jason Cho, Jiawei Han
SDM 2014 (slides adapted from Manish Gupta)

Outlier Detection in Attributed Subgraphs

User query:
3-author clique



Learn a Max-Margin SVM to predict which edges in the neighborhood exist based on node features.

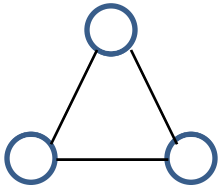


Normal

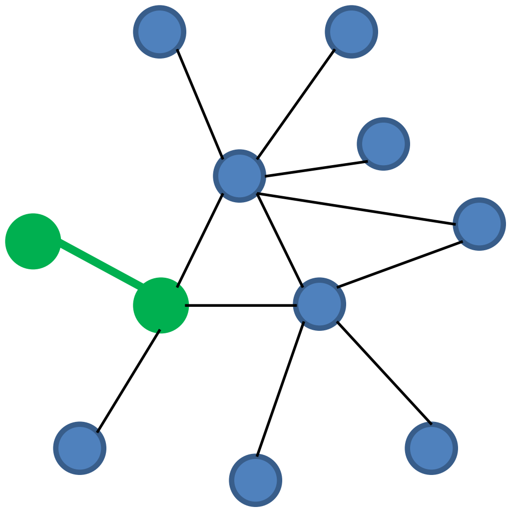
Local Learning for Mining Outlier Subgraphs from Network Datasets
Manish Gupta, Arun Mallya, Subhro Roy,
Jason Cho, Jiawei Han
SDM 2014 (slides adapted from Manish Gupta)

Outlier Detection in Attributed Subgraphs

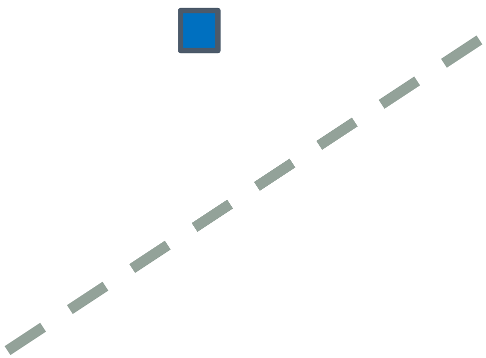
User query:
3-author clique



Learn a Max-Margin SVM to predict which edges in the neighborhood exist based on node features.



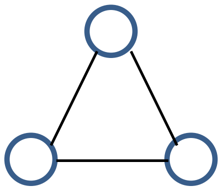
Normal



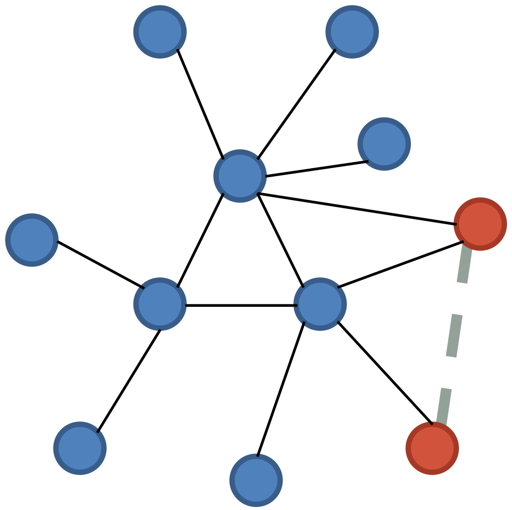
Local Learning for Mining Outlier Subgraphs from Network Datasets
Manish Gupta, Arun Mallya, Subhro Roy,
Jason Cho, Jiawei Han
SDM 2014 (slides adapted from Manish Gupta)

Outlier Detection in Attributed Subgraphs

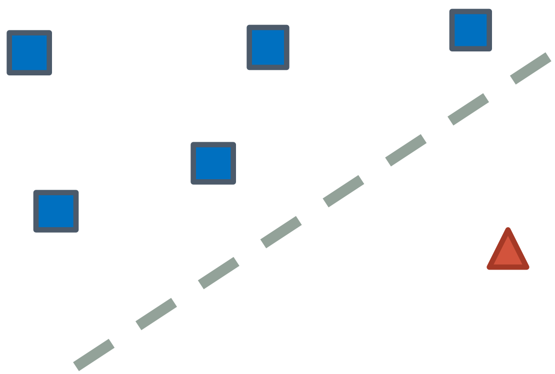
User query:
3-author clique



Learn a Max-Margin SVM to predict which edges in the neighborhood exist based on node features.



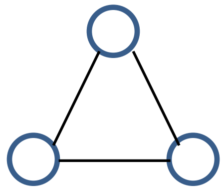
Normal



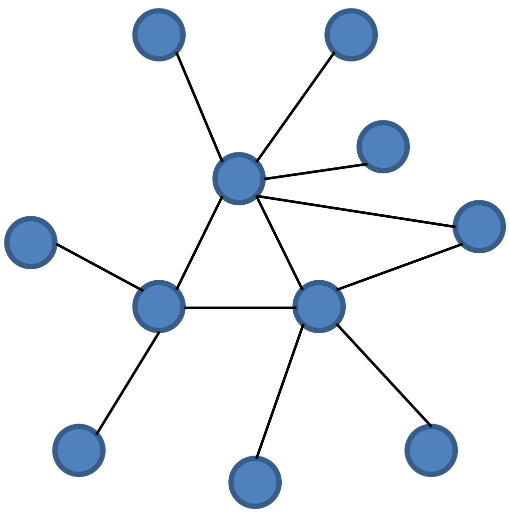
Local Learning for Mining Outlier Subgraphs from Network Datasets
Manish Gupta, Arun Mallya, Subhro Roy,
Jason Cho, Jiawei Han
SDM 2014 (slides adapted from Manish Gupta)

Outlier Detection in Attributed Subgraphs

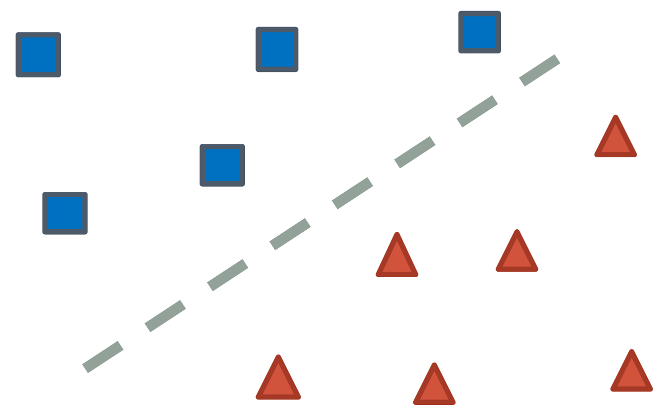
User query:
3-author clique



Learn a Max-Margin SVM to predict which edges in the neighborhood exist based on node features.

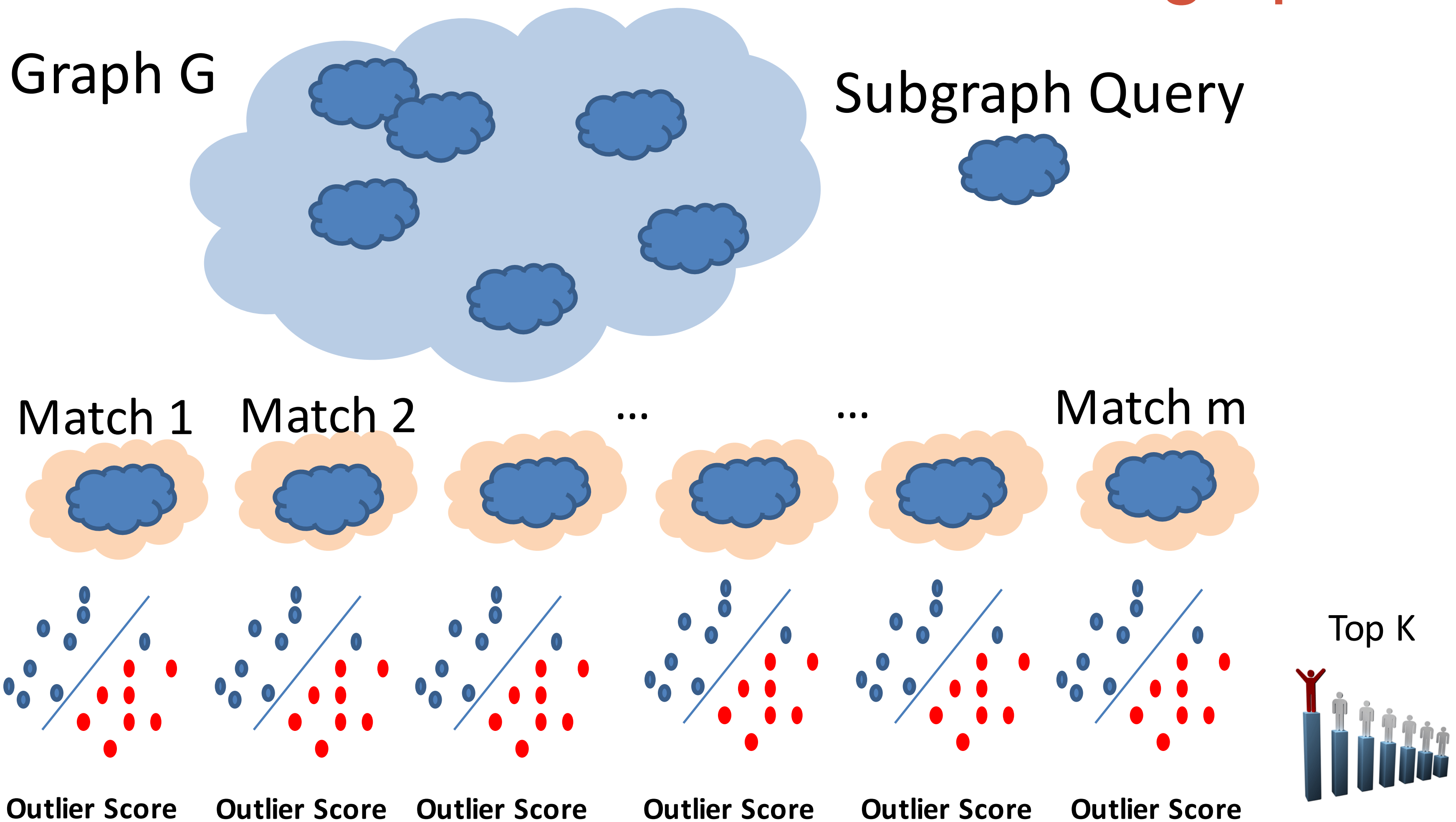


Normal



Local Learning for Mining Outlier Subgraphs from Network Datasets
Manish Gupta, Arun Mallya, Subhro Roy,
Jason Cho, Jiawei Han
SDM 2014 (slides adapted from Manish Gupta)

Outlier Detection in Attributed Subgraphs



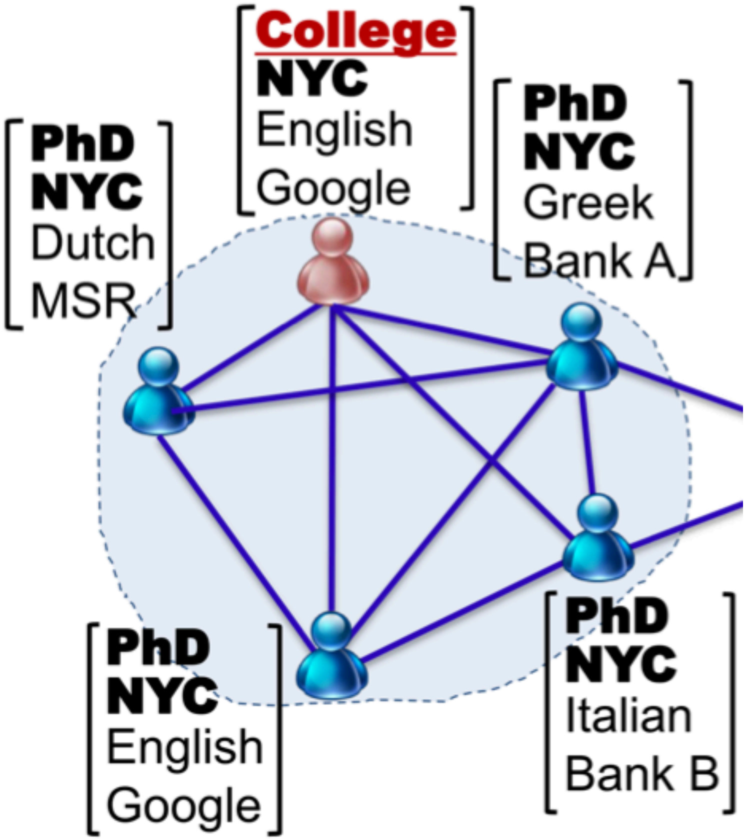
Local Learning for Mining Outlier Subgraphs from Network Datasets
Manish Gupta, Arun Mallya, Subhro Roy,
Jason Cho, Jiawei Han
SDM 2014 (slides adapted from Manish Gupta)

Clustering and Outlier Detection in Attributed Graphs

Given a graph with node attributes,

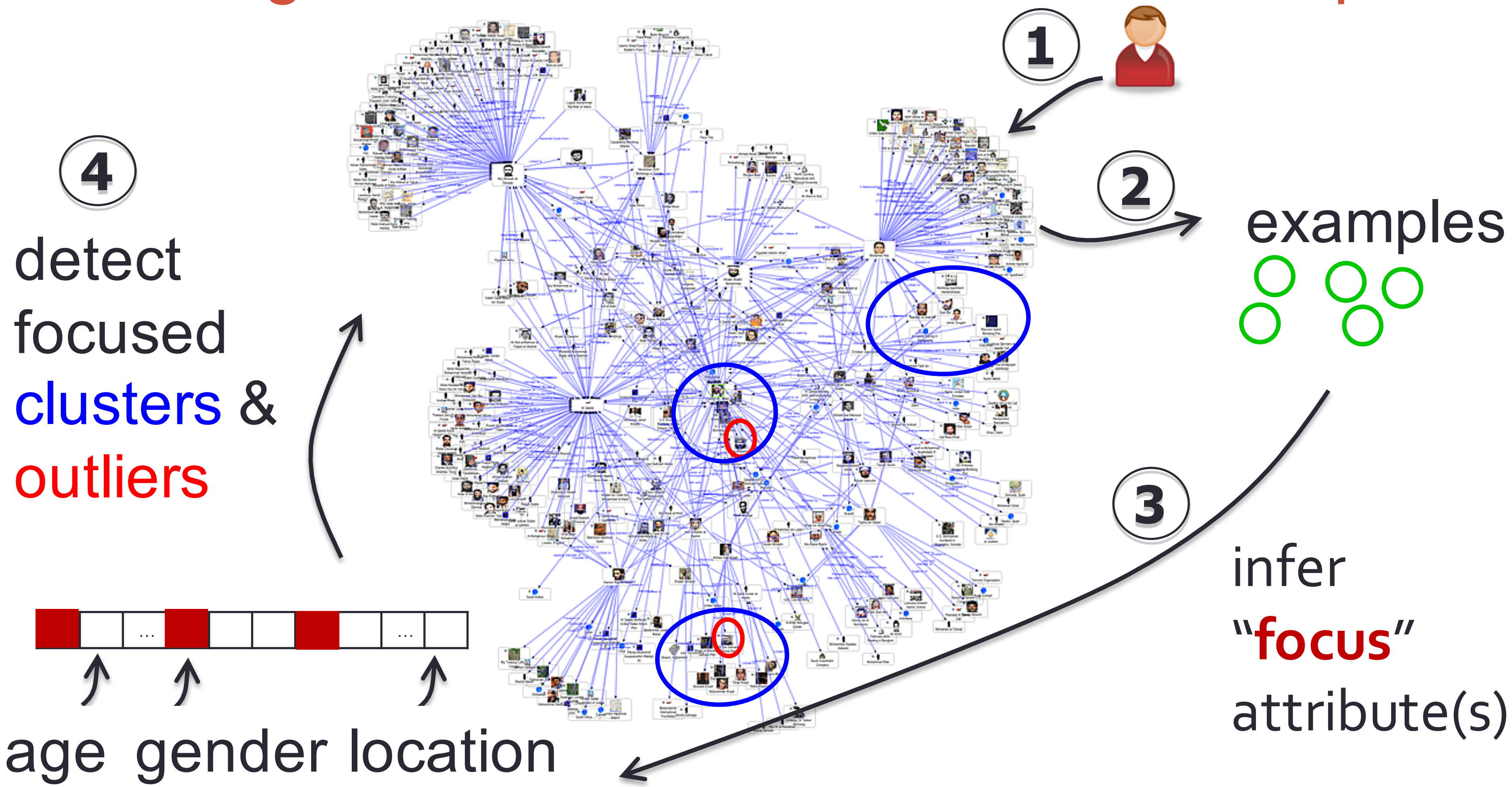
Find focused clusters that are dense and share attributes, and

Detect outliers, nodes whose attributes deviate from their cluster's attributes.



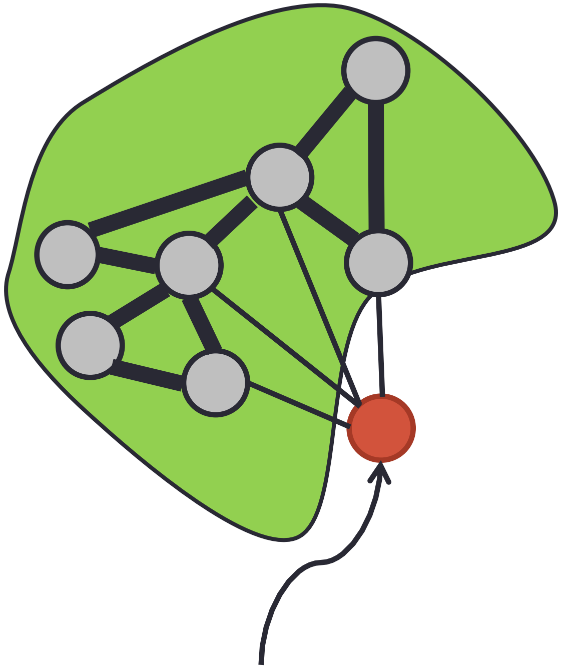
Focused Clustering and Outlier Detection in Large Attributed Graphs
Bryan Perozzi, Leman Akoglu, Patricia Iglesias Sanchez, Emmanuel Muller
KDD 2014 (slides adapted from Bryan Perozzi)

Clustering & Outlier Detection in Attributed Graphs



Focused Clustering and Outlier Detection in Large Attributed Graphs
Bryan Perozzi, Leman Akoglu, Patricia Iglesias Sanchez,
Emmanuel Muller
KDD 2014 (slides adapted from Bryan Perozzi)

Clustering & Outlier Detection in Attributed Graphs



Focused Outlier

1. Clustering objective: conductance $\phi^{(w)}$ weighted by focus

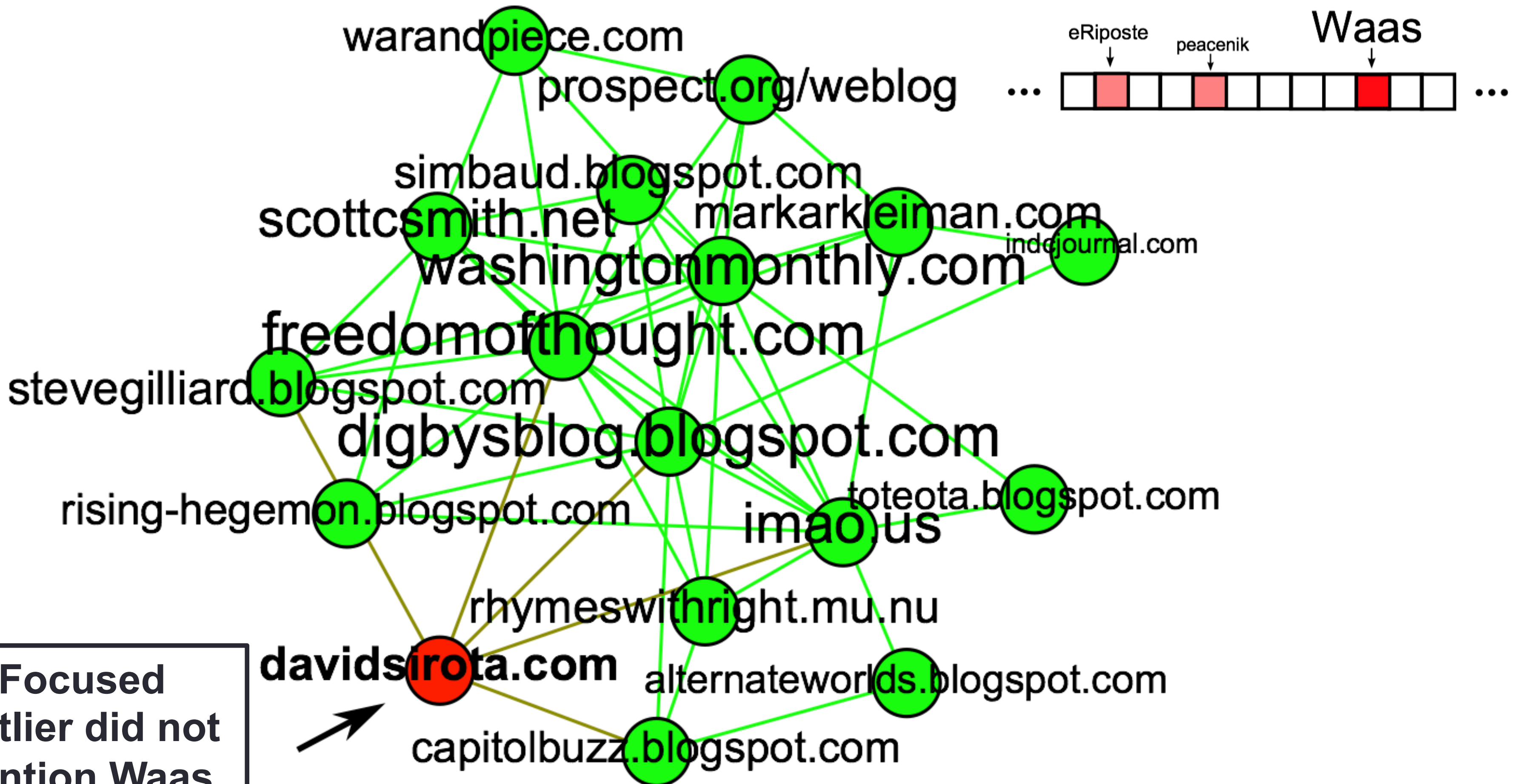
$$\phi^{(w)}(C,G) = \frac{W_{cut}(C)}{WVol(C)}$$

2. At each step in cluster expansion:
   - 2.1 - Examine boundary nodes
   - 2.2 - Add node with best $\Delta\,\phi^{(w)}$
   - 2.3 - Record best structural node

3. Focused Outliers:
   left-out best structural nodes

Focused Clustering and Outlier Detection in Large Attributed Graphs
Bryan Perozzi, Leman Akoglu, Patricia Iglesias Sanchez, Emmanuel Muller
KDD 2014
(slides adapted from Bryan Perozzi)

Clustering & Outlier Detection in Attributed Graphs

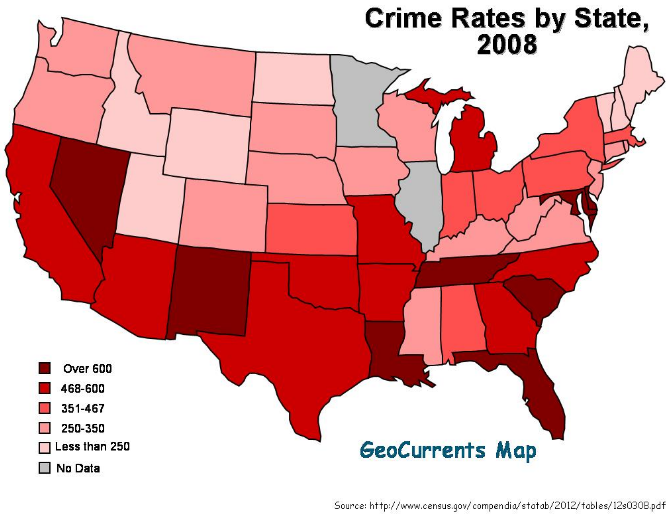


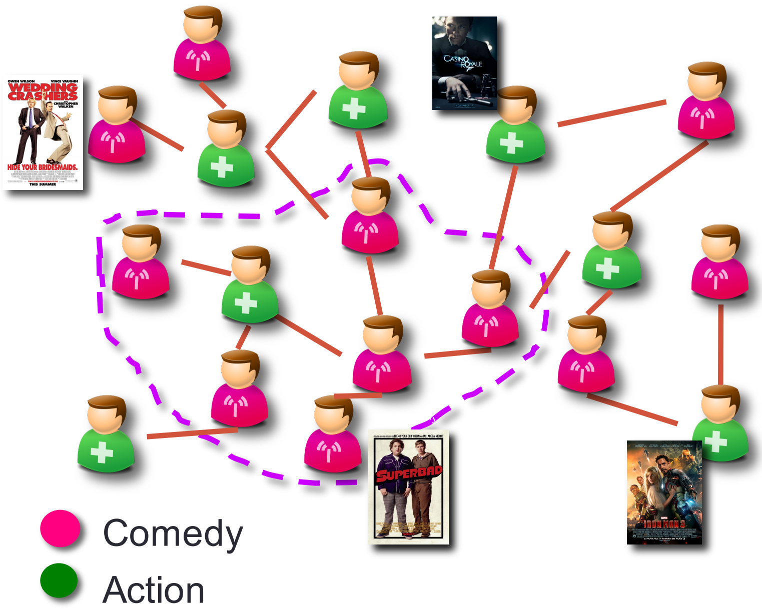
Focused Outlier did not mention Waas.

Focused Clustering and Outlier Detection in Large Attributed Graphs
Bryan Perozzi, Leman Akoglu, Patricia Iglesias Sanchez, Emmanuel Muller
KDD 2014

(slides adapted from Bryan Perozzi)

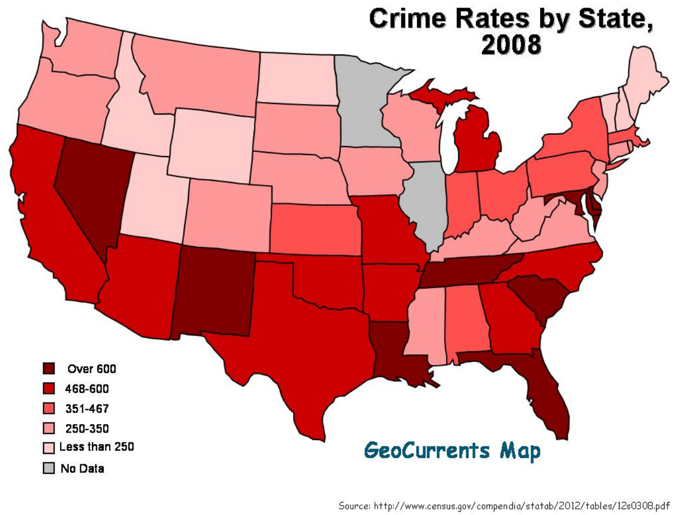
Anomalous-Attribute Subgraphs

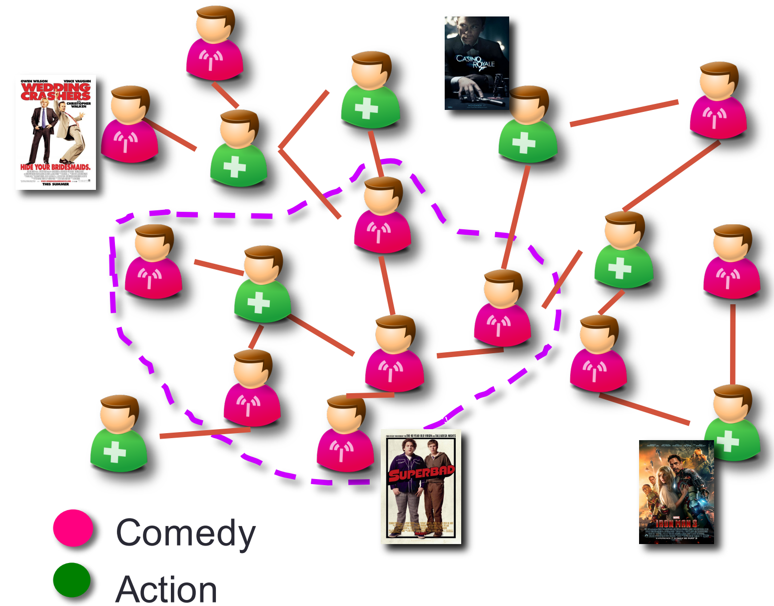


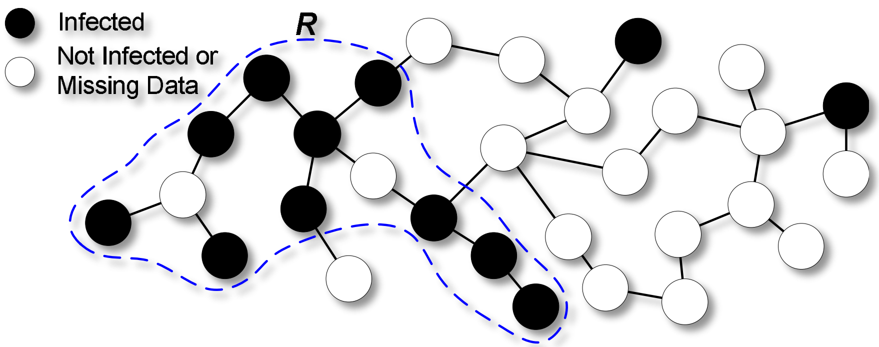


A Probabilistic Approach to Uncovering Attributed Graph Anomalies
Nan Li, Huan Sun, Kyle Chipman,
Jemin George, Xifeng Yan
SDM 2014

Anomalous-Attribute Subgraphs



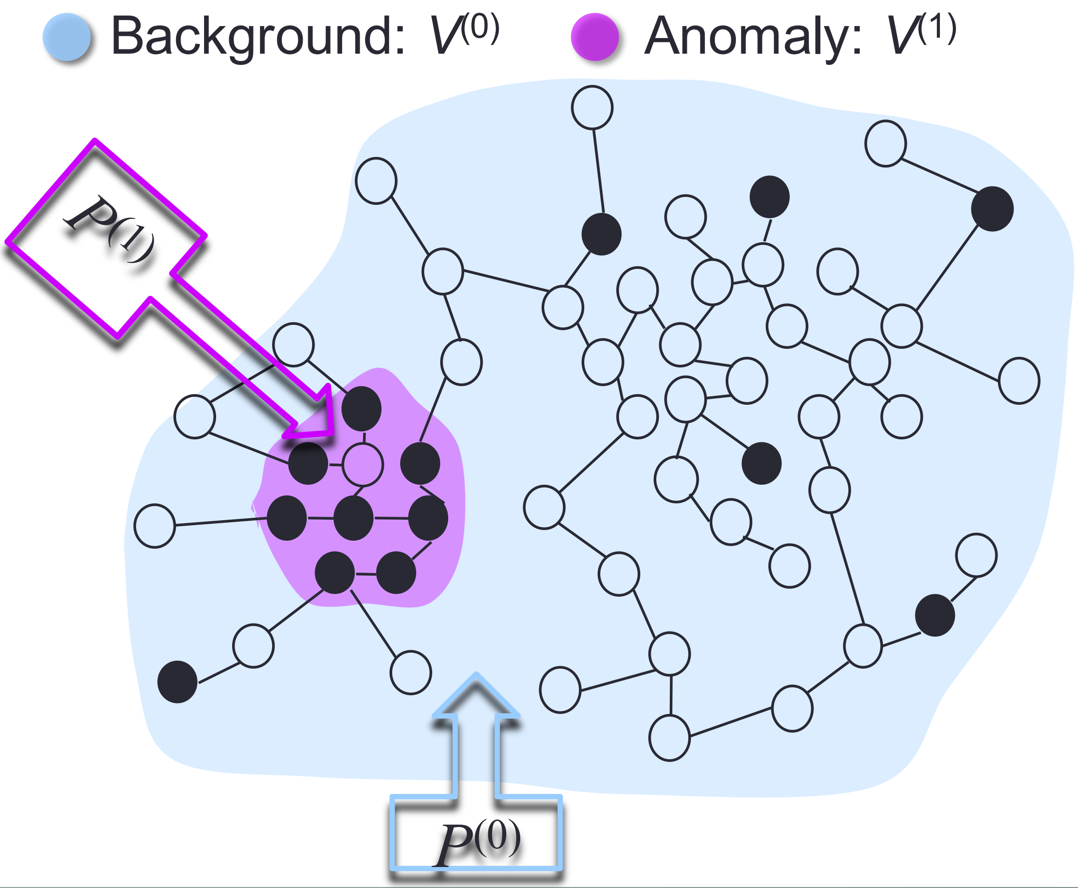




Subgraph with skewed attribute distribution

A Probabilistic Approach to Uncovering Attributed Graph Anomalies
Nan Li, Huan Sun, Kyle Chipman,
Jemin George, Xifeng Yan
SDM 2014

Anomalous-Attribute Subgraphs

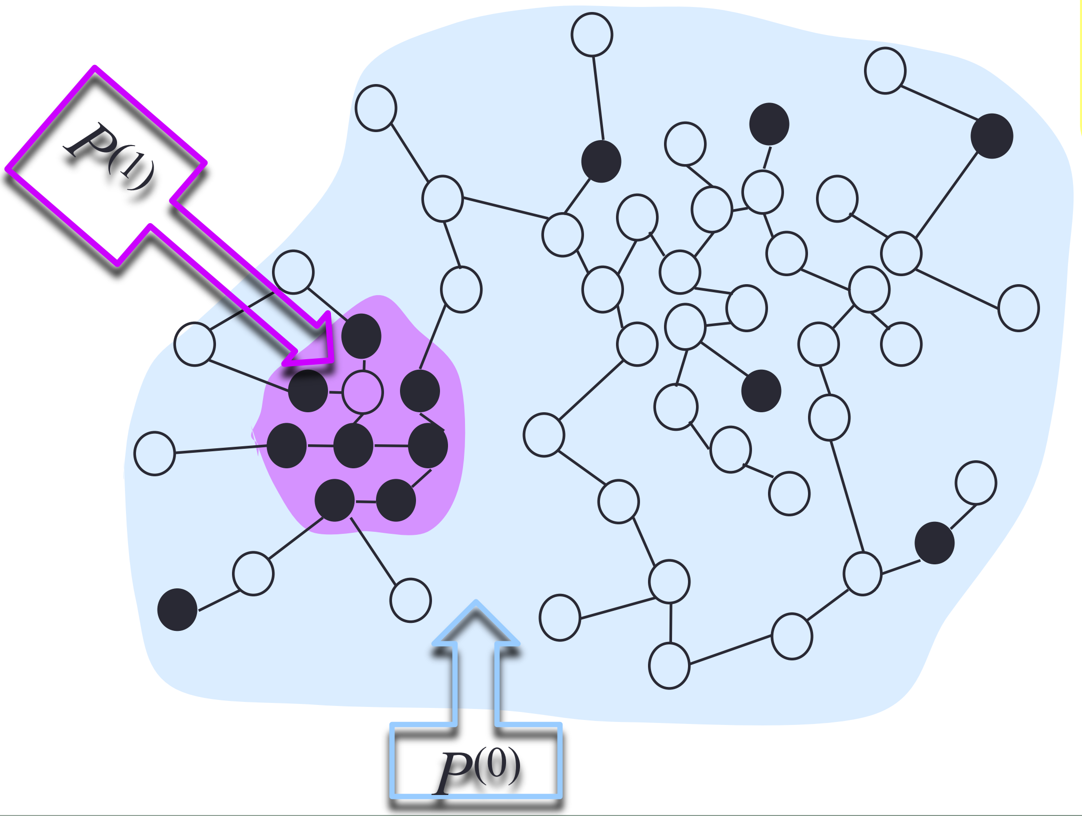


Two generative processes:
1) **anomaly distribution &**
2) **background distribution**

A Probabilistic Approach to Uncovering Attributed Graph Anomalies
Nan Li, Huan Sun, Kyle Chipman,
Jemin George, Xifeng Yan
SDM 2014

Anomalous-Attribute Subgraphs

Background: $V^{(0)}$ Anomaly: $V^{(1)}$



Two generative processes:

1) **anomaly distribution &**
2) **background distribution**

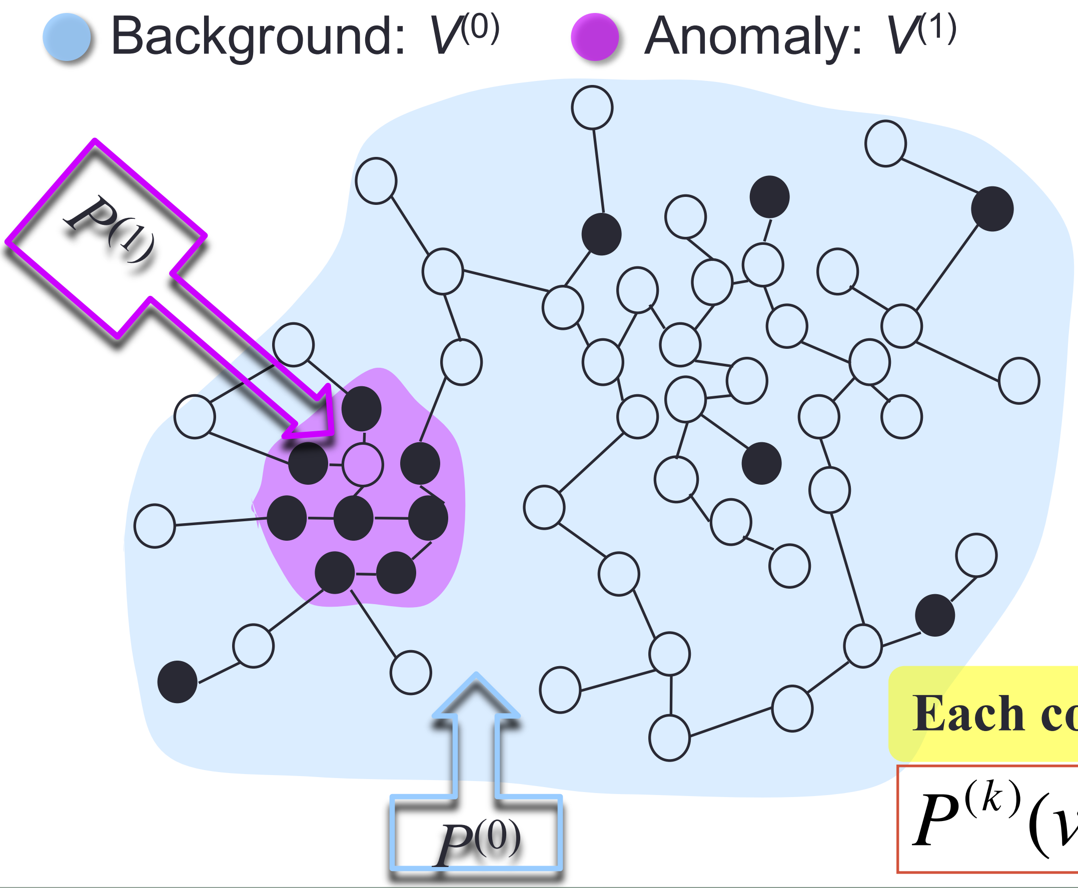
One overall mixture

$$P(v_i) = \sum_{k=0}^{1} \theta_i^{(k)} P^{(k)}(v_i)$$

With probability $\theta_i^{(0)}$, v_i belongs to the background component $V^{(0)}$, and with $\theta_i^{(1)}$ the anomaly component $V^{(1)}$.

A Probabilistic Approach to Uncovering Attributed Graph Anomalies
Nan Li, Huan Sun, Kyle Chipman,
Jemin George, Xifeng Yan
SDM 2014

Anomalous-Attribute Subgraphs



Two generative processes:

1) **anomaly distribution &**
2) **background distribution**

One overall mixture

$$P(v_i) = \sum_{k=0}^{1} \theta_i^{(k)} P^{(k)}(v_i)$$

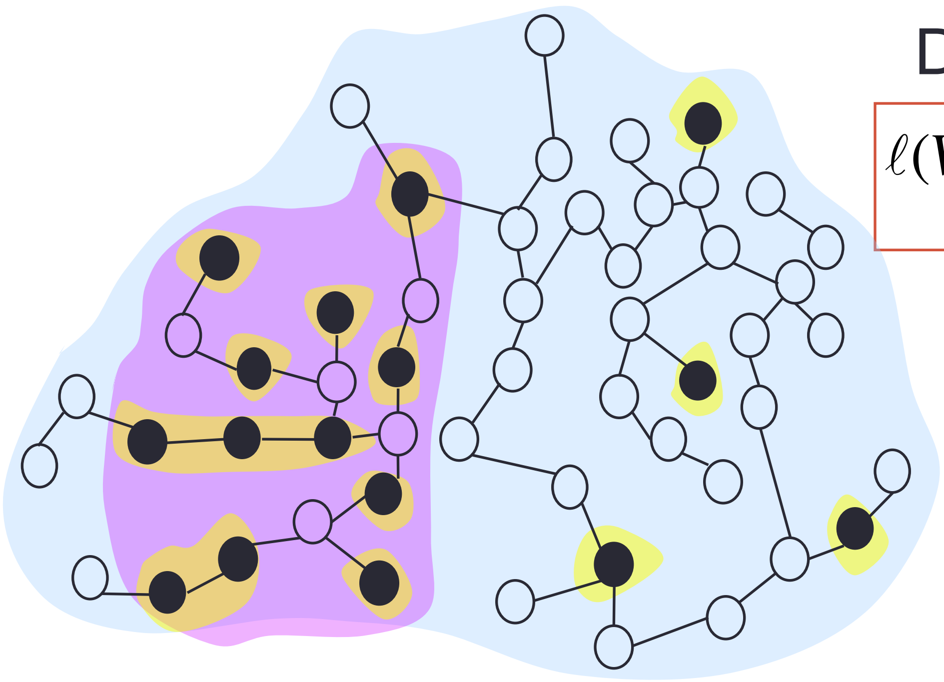
With probability $\theta_i^{(0)}$, v_i belongs to the background component $V^{(0)}$, and with $\theta_i^{(1)}$ the anomaly component $V^{(1)}$.

Each component is a Bernoulli distribution

$$P^{(k)}(v_i) = \boldsymbol{p}^{(k)}(1)^{X_i}\,(1 - \boldsymbol{p}^{(k)}(1))^{1-X_i}$$

A Probabilistic Approach to Uncovering Attributed Graph Anomalies
Nan Li, Huan Sun, Kyle Chipman,
Jemin George, Xifeng Yan
SDM 2014

Anomalous-Attribute Subgraphs

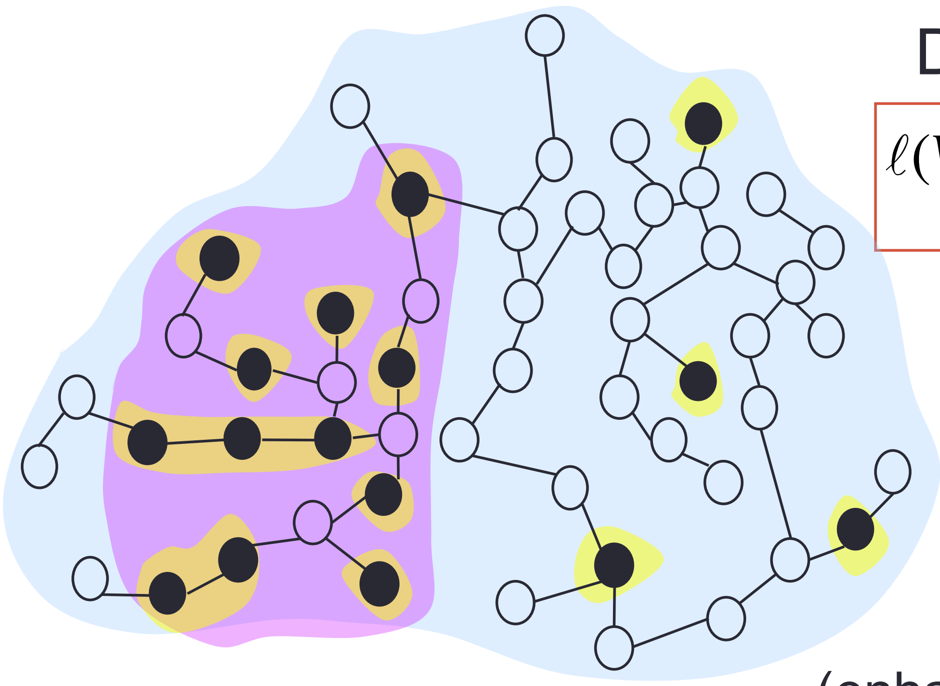


Data loglikelihood of vertex set V

$$\ell(V) = \sum_{v_i \in V} \log P(v_i) = \sum_{v_i \in V} \log \sum_k \theta_i^{(k)} P^{(k)}(v_i)$$

A Probabilistic Approach to Uncovering Attributed Graph Anomalies
Nan Li, Huan Sun, Kyle Chipman,
Jemin George, Xifeng Yan
SDM 2014

Anomalous-Attribute Subgraphs



Data loglikelihood of vertex set V

$$\ell(V) = \sum_{v_i \in V} \log P(v_i) = \sum_{v_i \in V} \log \sum_{k} \theta_i^{(k)} P^{(k)}(v_i)$$

Maximize:

$$\ell(V) - \lambda R_N(\Theta) + \gamma R_E(\Theta)$$

Network regularizer (enhances connectivity within each component)

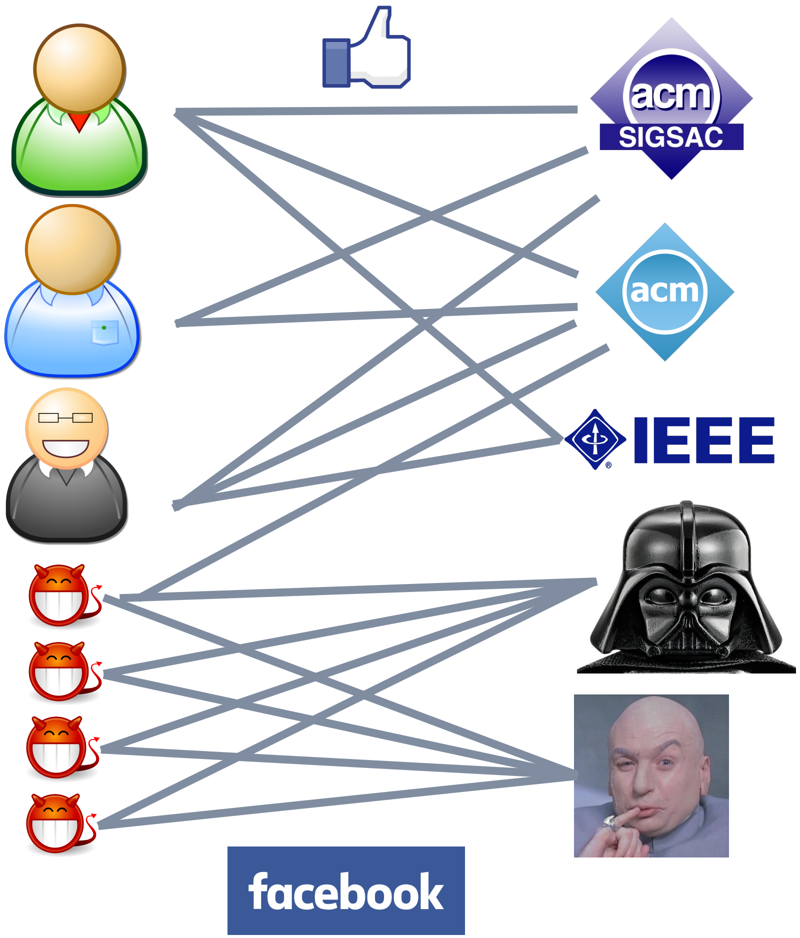
Entropy regularizer (enhances polarity of mixture weights)

A Probabilistic Approach to Uncovering Attributed Graph Anomalies
Nan Li, Huan Sun, Kyle Chipman,
Jemin George, Xifeng Yan
SDM 2014

Practitioner's Guide to Detecting Fraud

Method	Graph Type	Node Attributes	Edge Attributes	Seed Labels
COI	Undirected			✓
OddBall	Undirected			
Blackholes & Volcanoes	Directed			
(Anti)-Social	Bipartite			
SODA	Undirected	✓		
FocusCO	Undirected	✓		
CopyCatch	Bipartite		✓	
SynchroTrap	Bipartite+	✓	✓	
CrossSpot	Bipartite		✓	
Co-Clustering	Bipartite*		✓	

Lockstep Behavior in the Graph

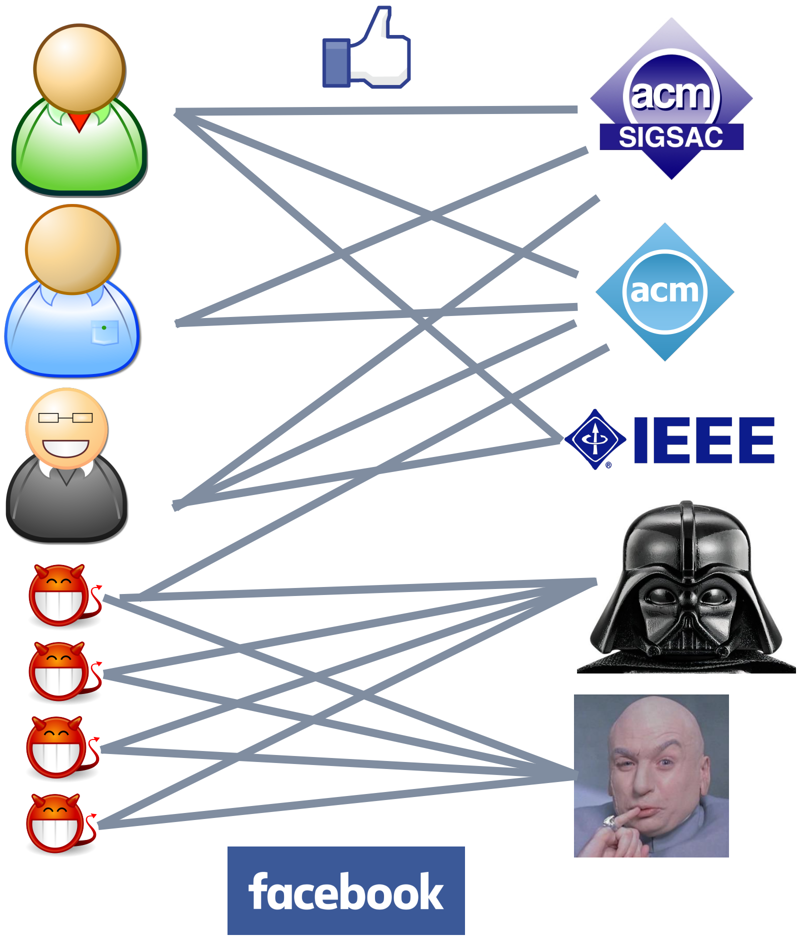


Dense group of data miner Page Likes

Dense group of purchased Page Likes

CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami, Christopher Palow, Christos Faloutsos
WWW, 2013

Lockstep Behavior in the Graph



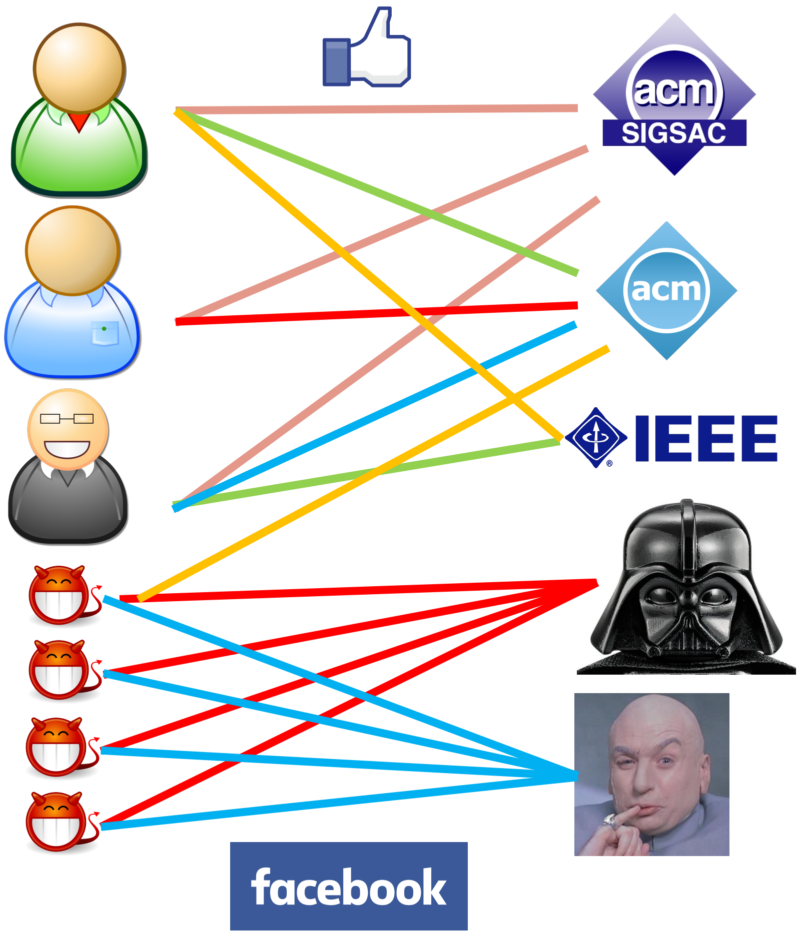
Dense group of data miner Page Likes

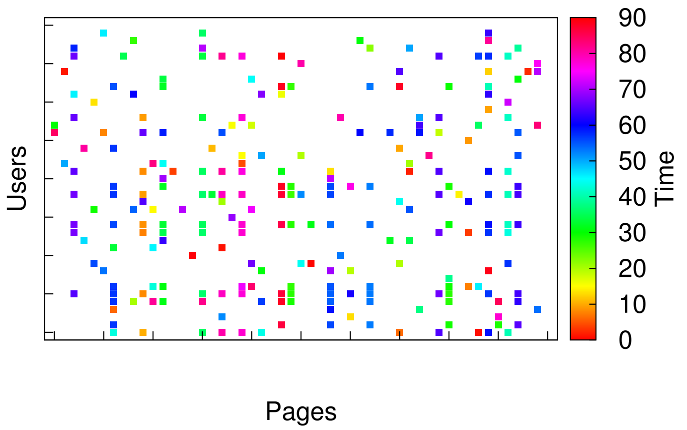
How can we tell which is fraudulent?

Dense group of purchased Page Likes

CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

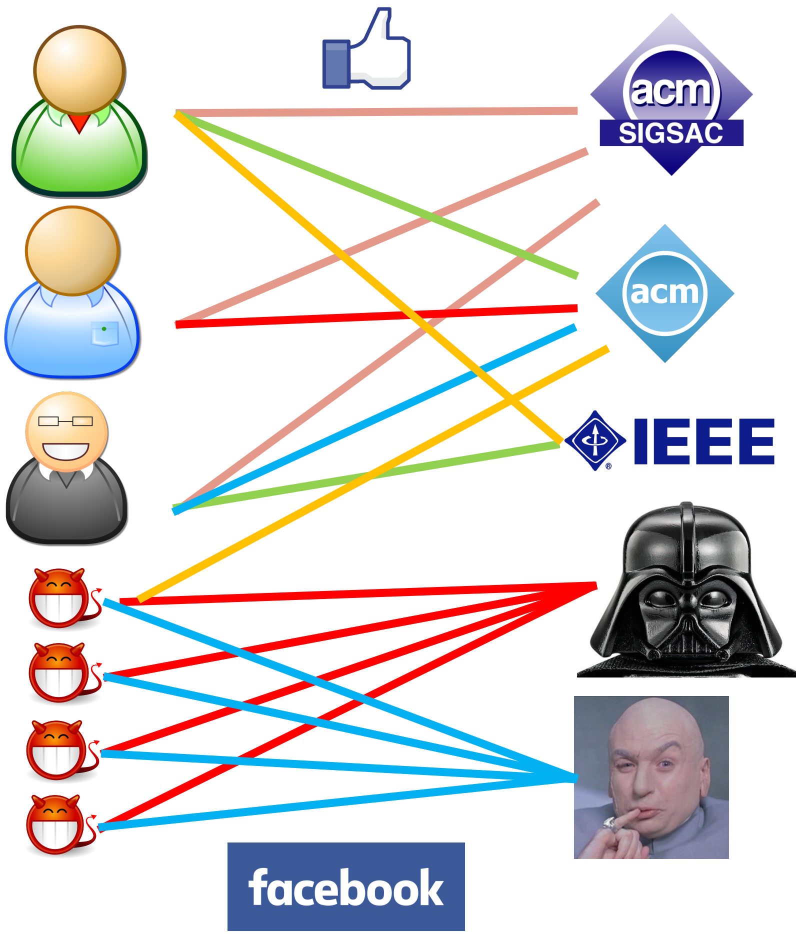
Lockstep Behavior in the Graph

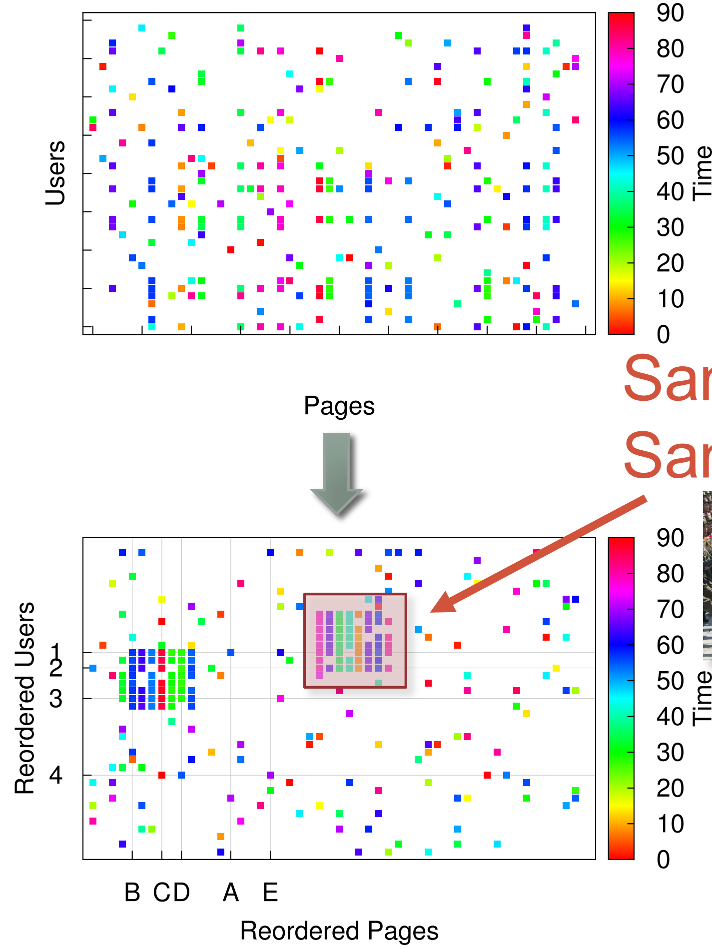




CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

Lockstep Behavior in the Graph



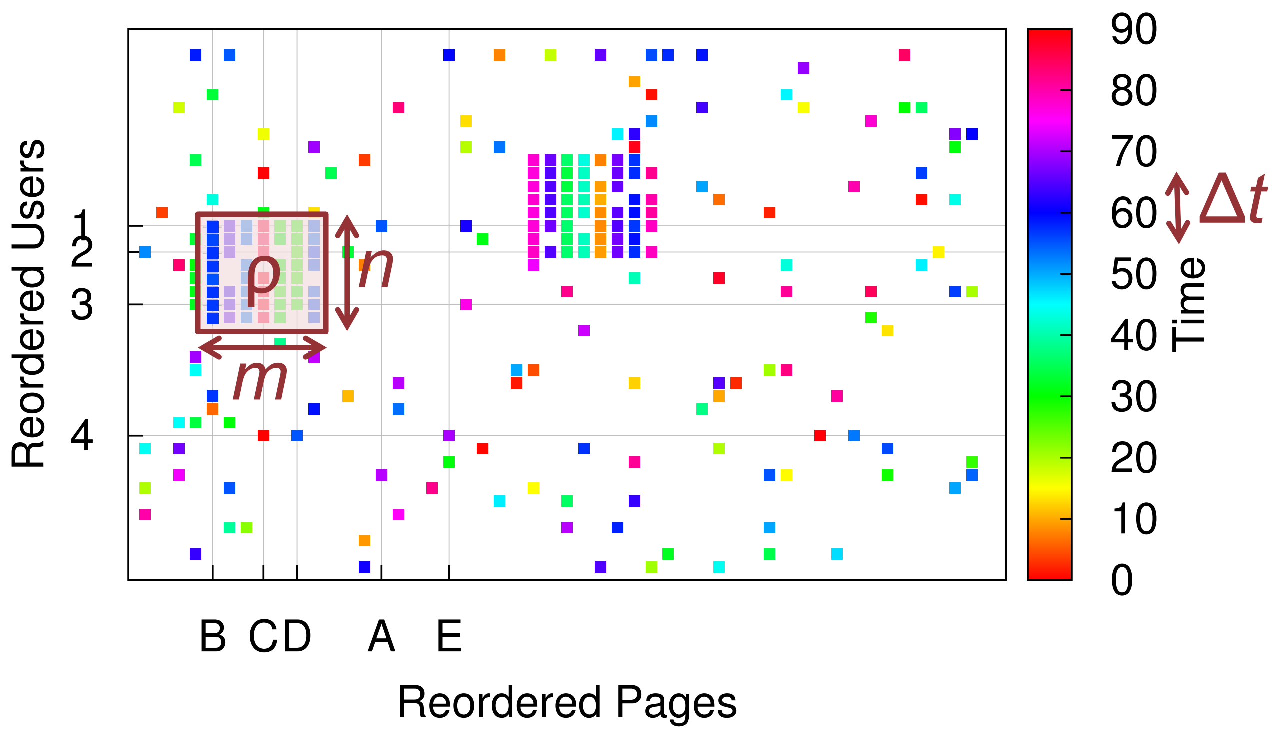


Same Likes,
Same Time

CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

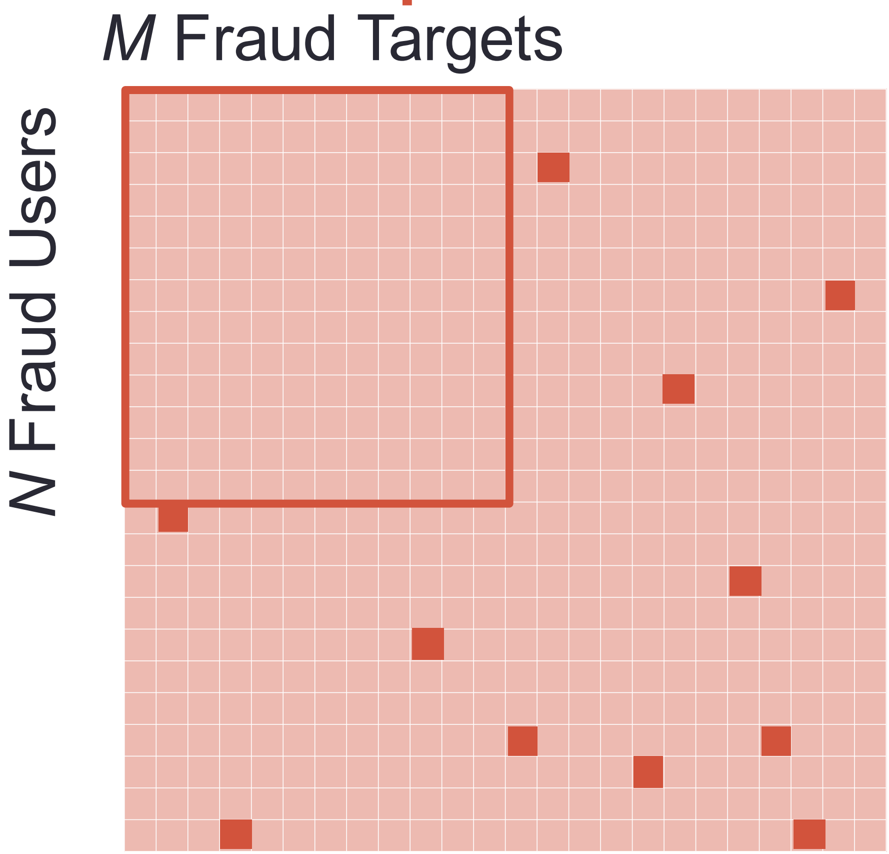
Lockstep Behavior in the Graph

Find [$n, m, \Delta t, \rho$]-Temporally Coherent Near Bipartite Cores (TNBC)



CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

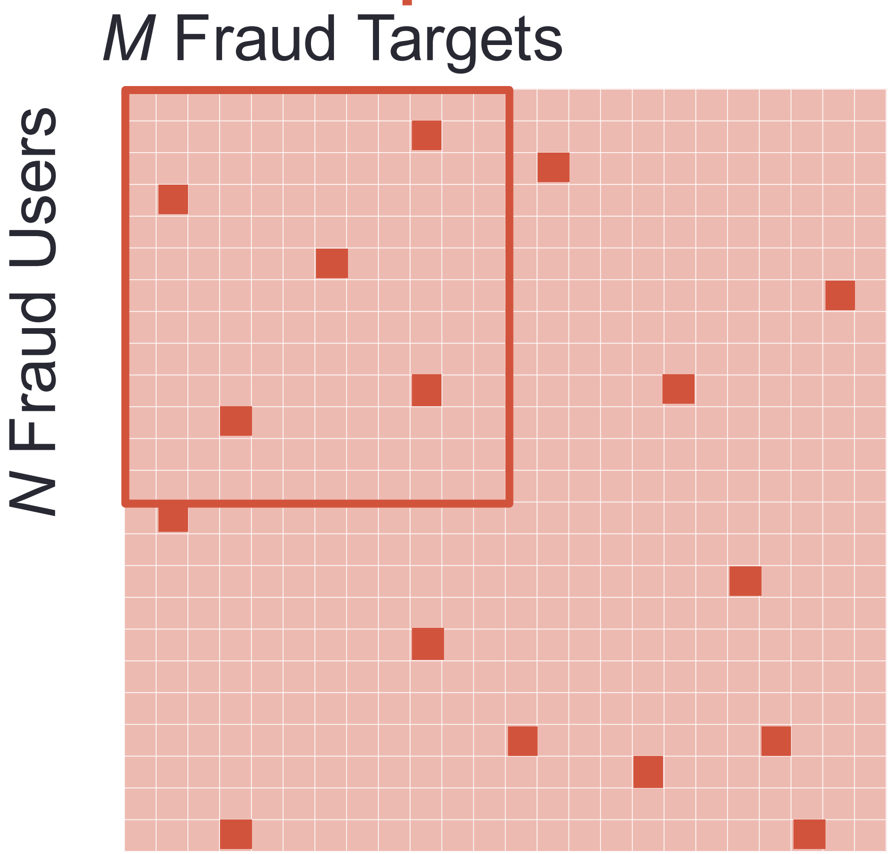
Lockstep Behavior in the Graph



How many edges can be added (by fraudsters) before creating an (r,s)-bipartite core?

CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

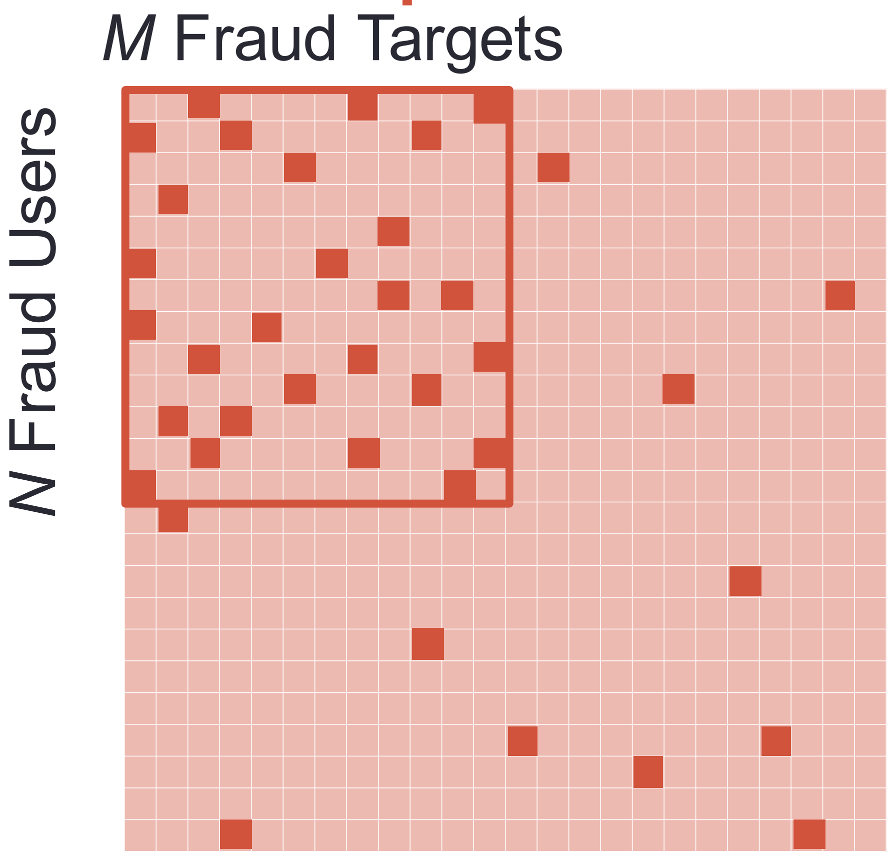
Lockstep Behavior in the Graph



How many edges can be added (by fraudsters) before creating an (r,s)-bipartite core?

CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

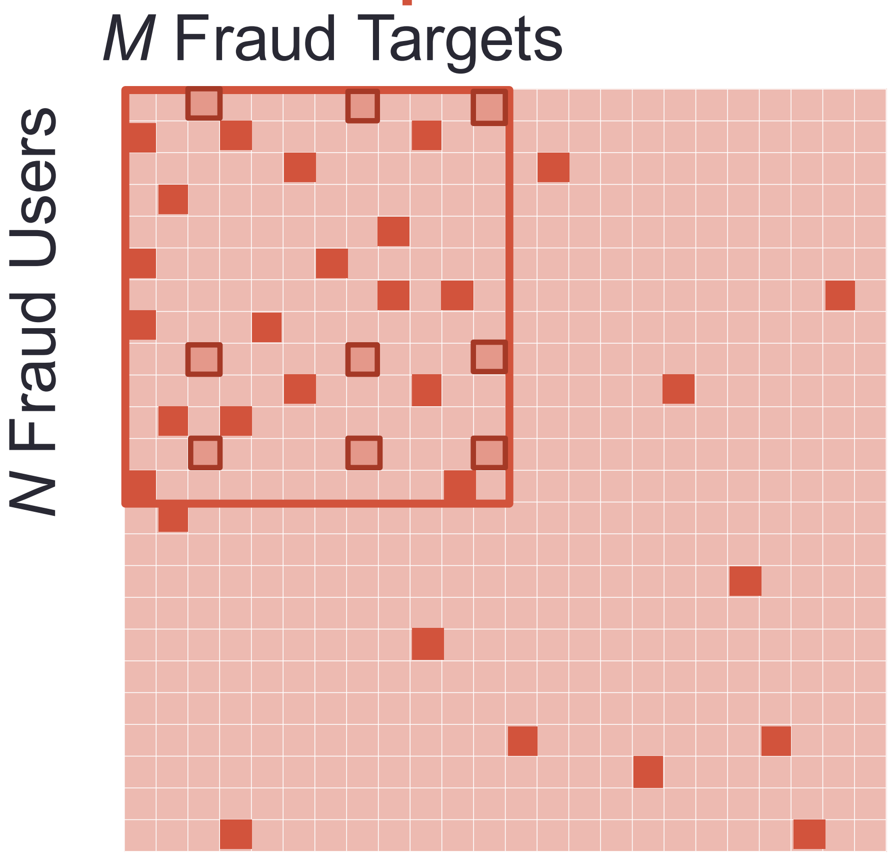
Lockstep Behavior in the Graph



How many edges can be added (by fraudsters) before creating an (r,s)-bipartite core?

CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

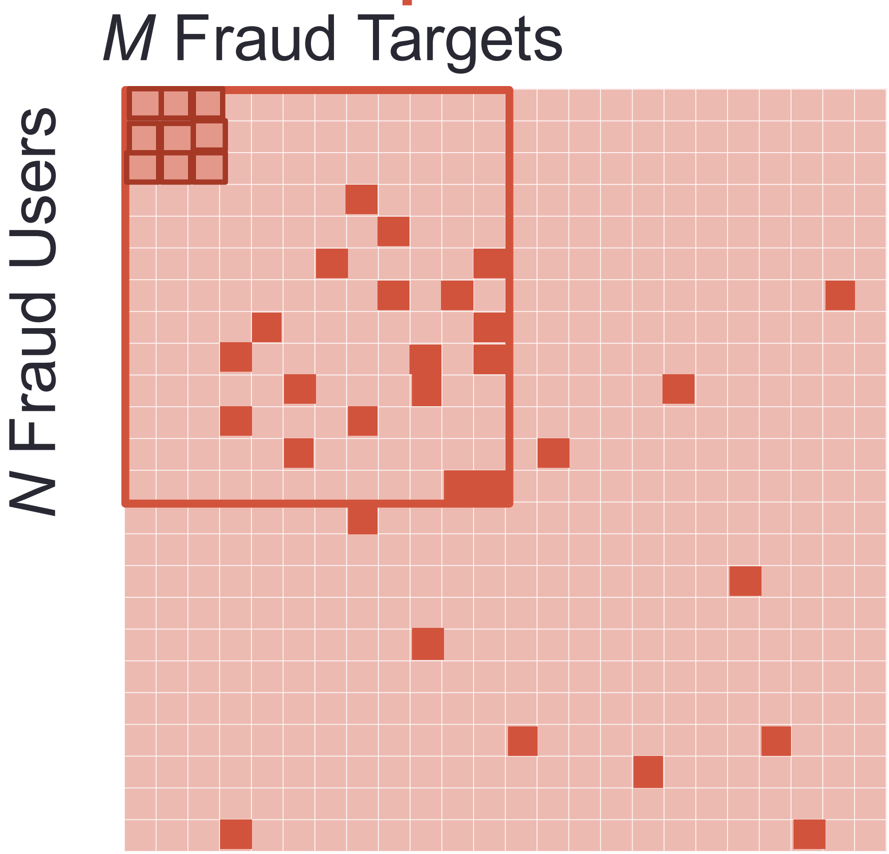
Lockstep Behavior in the Graph



How many edges can be added (by fraudsters) before creating an (r,s)-bipartite core?

CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

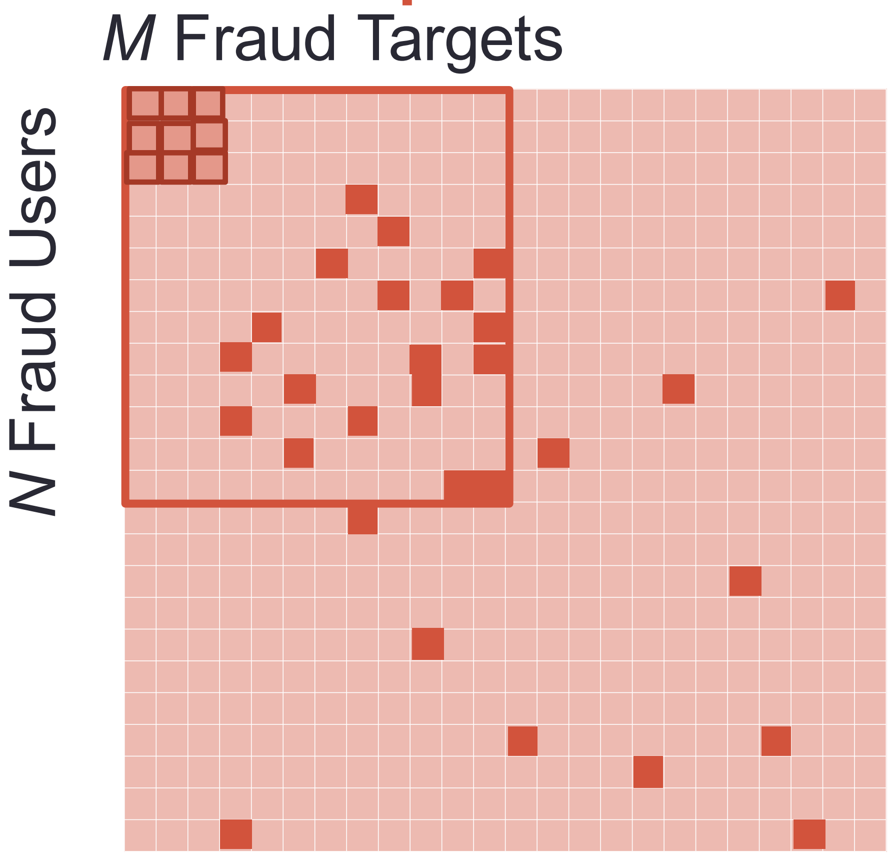
Lockstep Behavior in the Graph



How many edges can be added (by fraudsters) before creating an (r,s)-bipartite core?

CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

Lockstep Behavior in the Graph

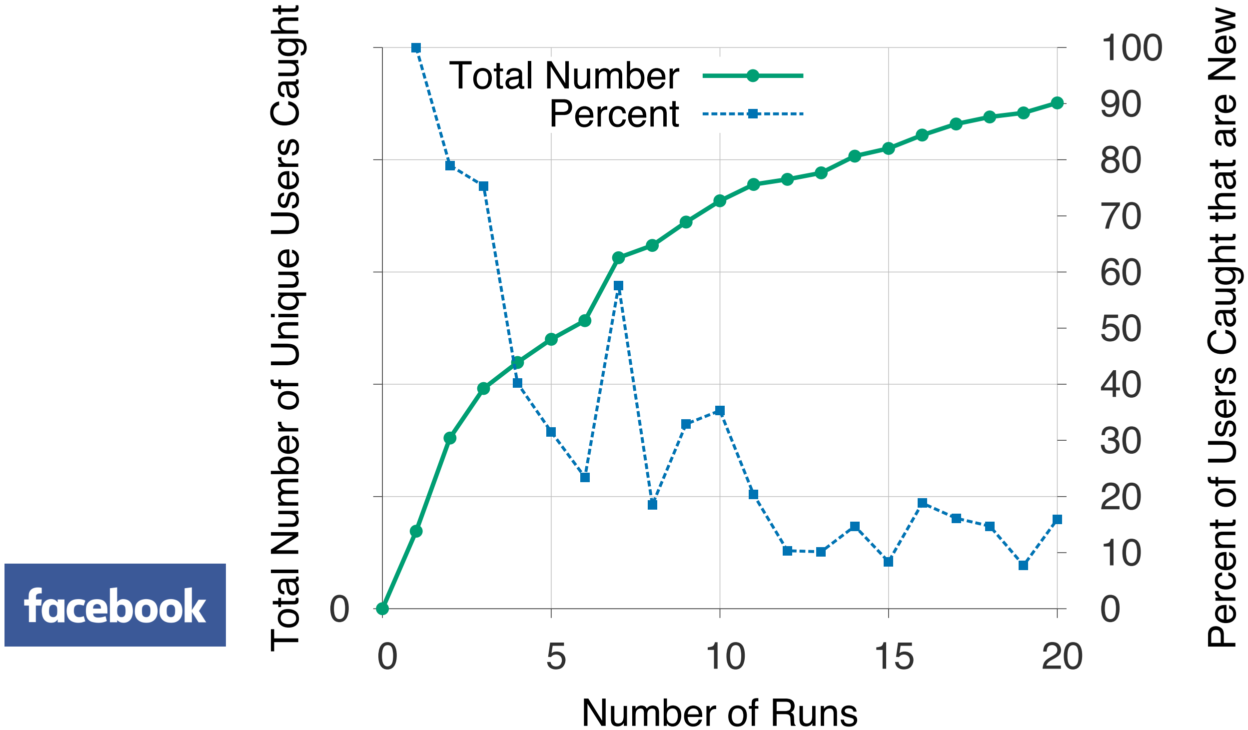


How many edges can be added (by fraudsters) before creating an (r,s)-bipartite core?

Zarankiewicz Problem
Proposed in 1951
and still open today

CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

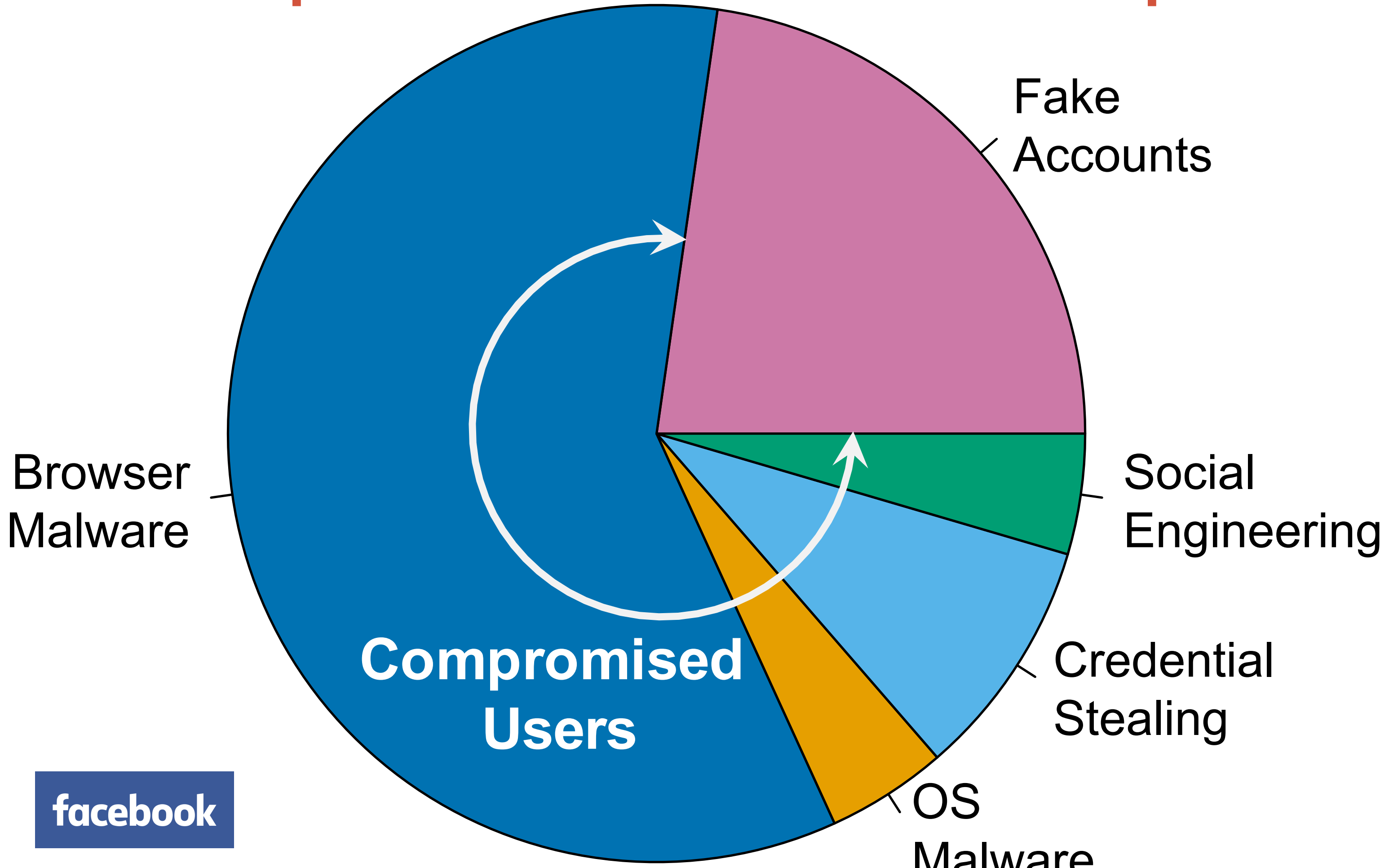
Lockstep Behavior in the Graph



CopyCatch works [quickly] – Few runs are enough

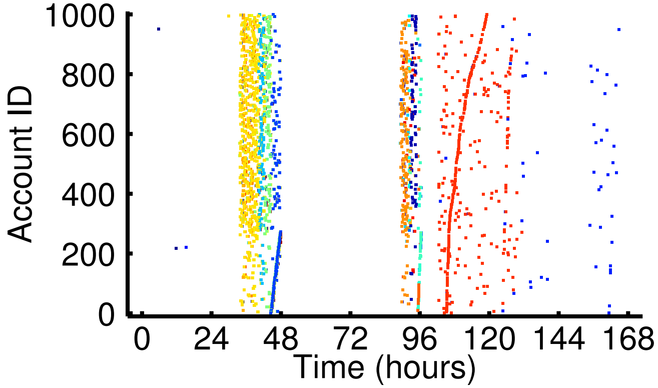
CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

Lockstep Behavior in the Graph

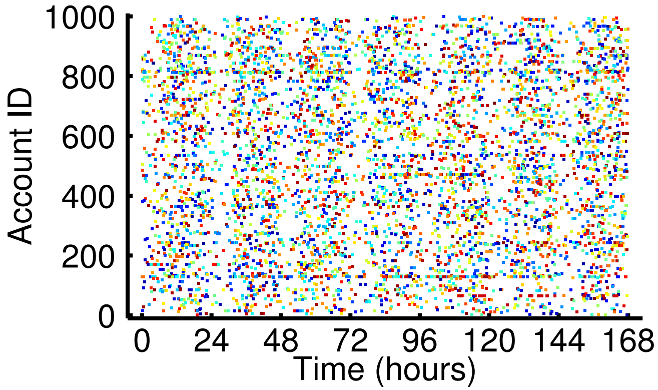


CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks
Alex Beutel, Wanhong Xu, Venkatesan Guruswami,
Christopher Palow, Christos Faloutsos
WWW, 2013

Lockstep Behavior in the Graph



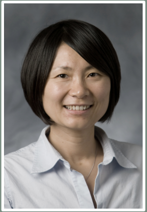
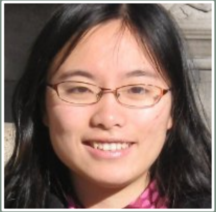
(a) Synchronized attack



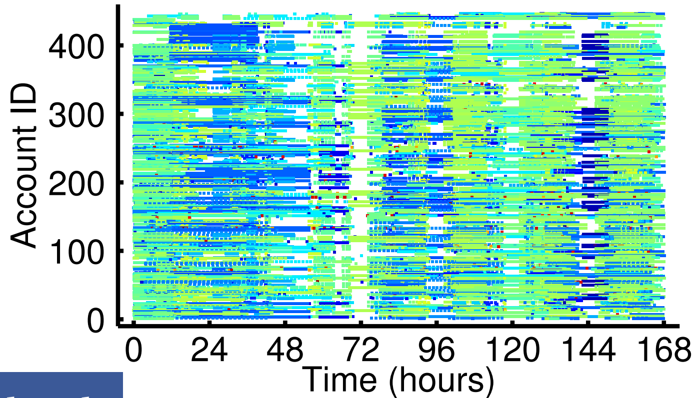
(b) Normal

Temporal lockstep behavior found in Instagram followers

Uncovering Large Groups of Active Malicious Accounts in Online Social Networks
Qiang Cao, Xiaowei Yang, Jieqi Yu, Christopher Palow
ACM *CCS* 2014

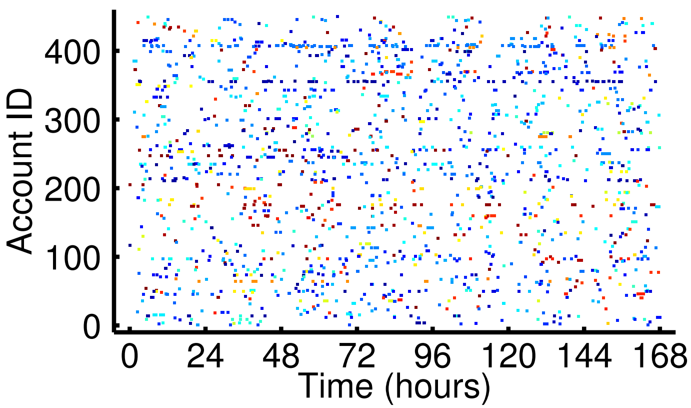





Lockstep Behavior in the Graph



facebook

(a) Synchronized attack



(b) Normal

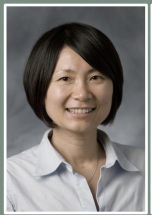
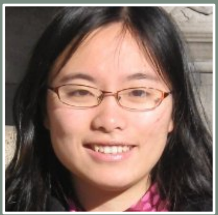
Accounts perform wide variety of synchronized tasks

Upload spammy photos

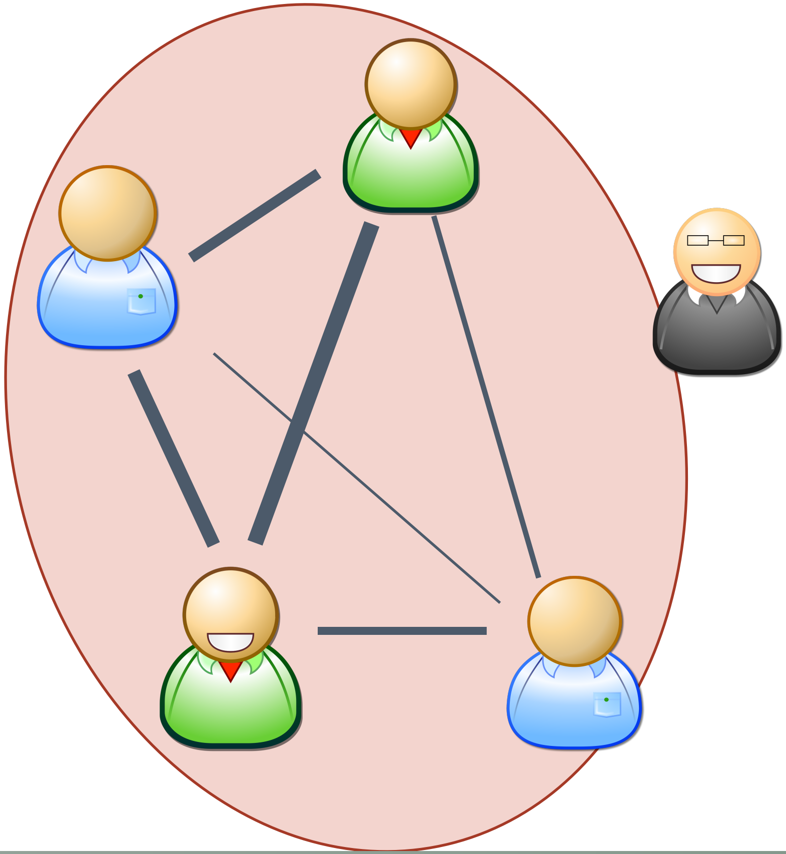
Share IP addresses (color)

Algorithmic Challenge:
Repeated actions

Uncovering Large Groups of Active Malicious Accounts in Online Social Networks
Qiang Cao, Xiaowei Yang, Jieqi Yu, Christopher Palow
ACM *CCS* 2014


Lockstep Behavior in the Graph



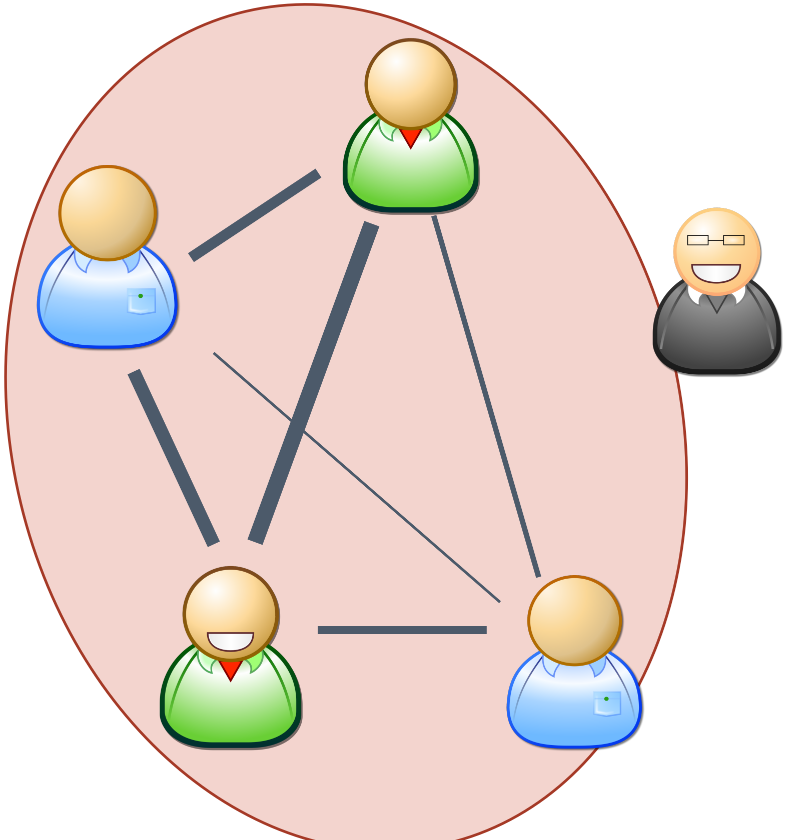
SynchroTrap

Define edge weight by similarity of actions in time window (including IP, action, etc.)

Cluster to find synchronized users

Uncovering Large Groups of Active Malicious Accounts in Online Social Networks
Qiang Cao, Xiaowei Yang, Jieqi Yu, Christopher Palow
ACM *CCS* 2014

Lockstep Behavior in the Graph



SynchroTrap

Bounds based on per-user rate limiting and set theory

If there are w time windows and L actions per object per time window:

$$\binom{w}{\lfloor w/2 \rfloor} wL$$

Uncovering Large Groups of Active Malicious Accounts in Online Social Networks
Qiang Cao, Xiaowei Yang, Jieqi Yu, Christopher Palow
ACM *CCS* 2014

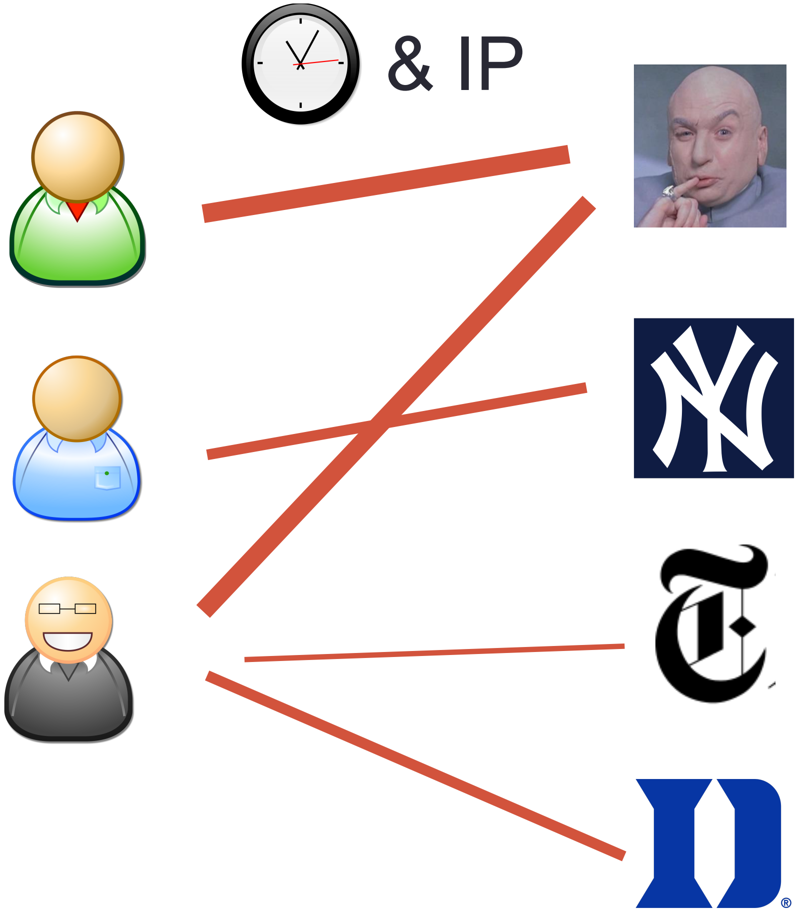
Lockstep Behavior in the Graph

Application	Page like	Instagram follow	App install	Photo upload	Login
Campaigns	201	531	74	29	321
Accounts	730K	589K	164K	120K	564K
Actions	357M	65M	4M	48M	29M
Precision	99.0%	99.7%	100%	100%	100%

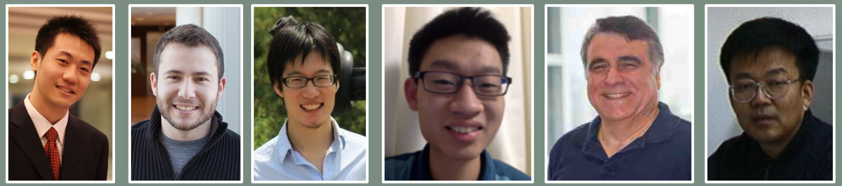
facebook

Uncovering Large Groups of Active Malicious Accounts in Online Social Networks
Qiang Cao, Xiaowei Yang, Jieqi Yu, Christopher Palow
ACM *CCS* 2014

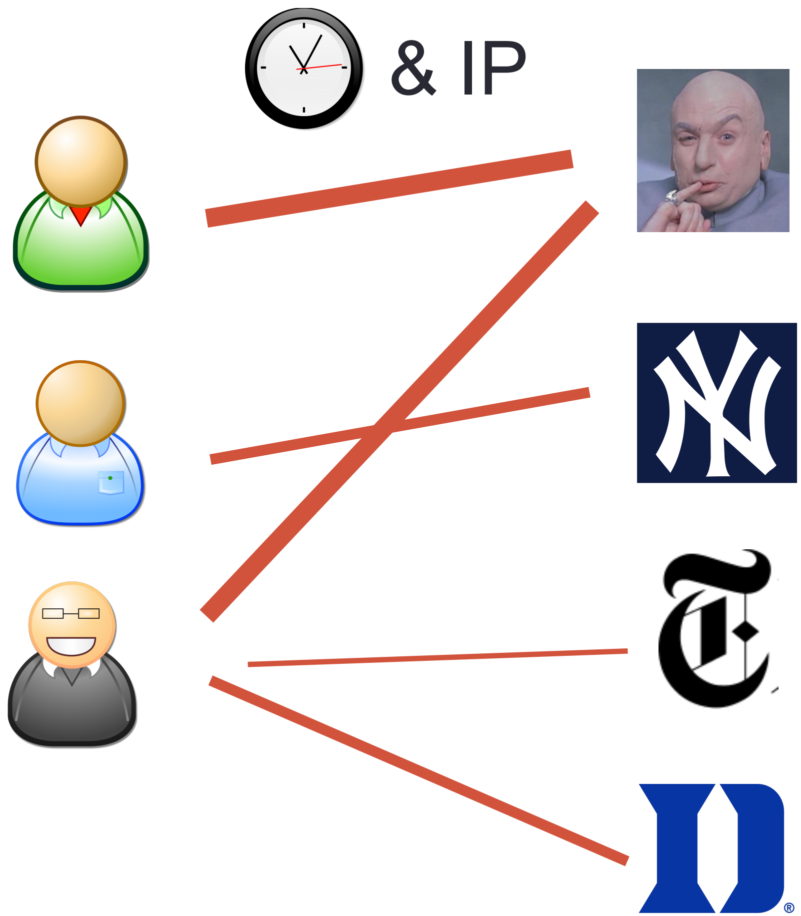
Lockstep Behavior on Some Attributes



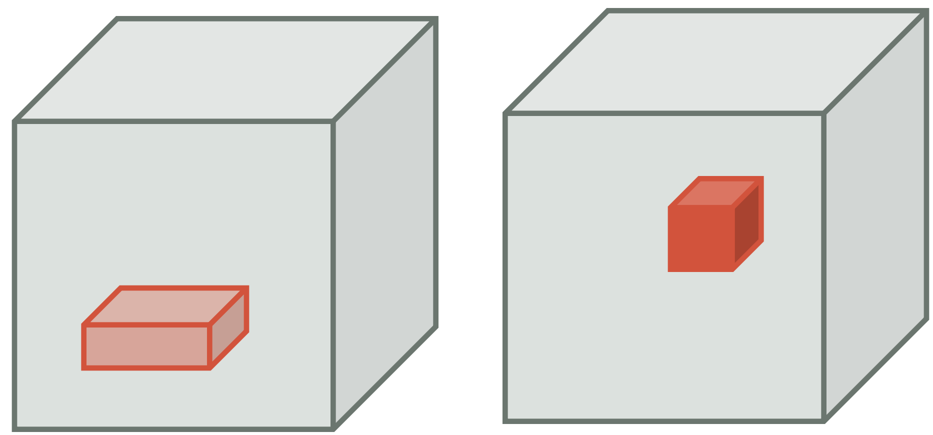
A General Suspiciousness Metric for Dense Blocks in Multimodal Data
Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi,
Shiqiang Yang, Christos Faloutsos
ICDM 2015



Lockstep Behavior on Some Attributes

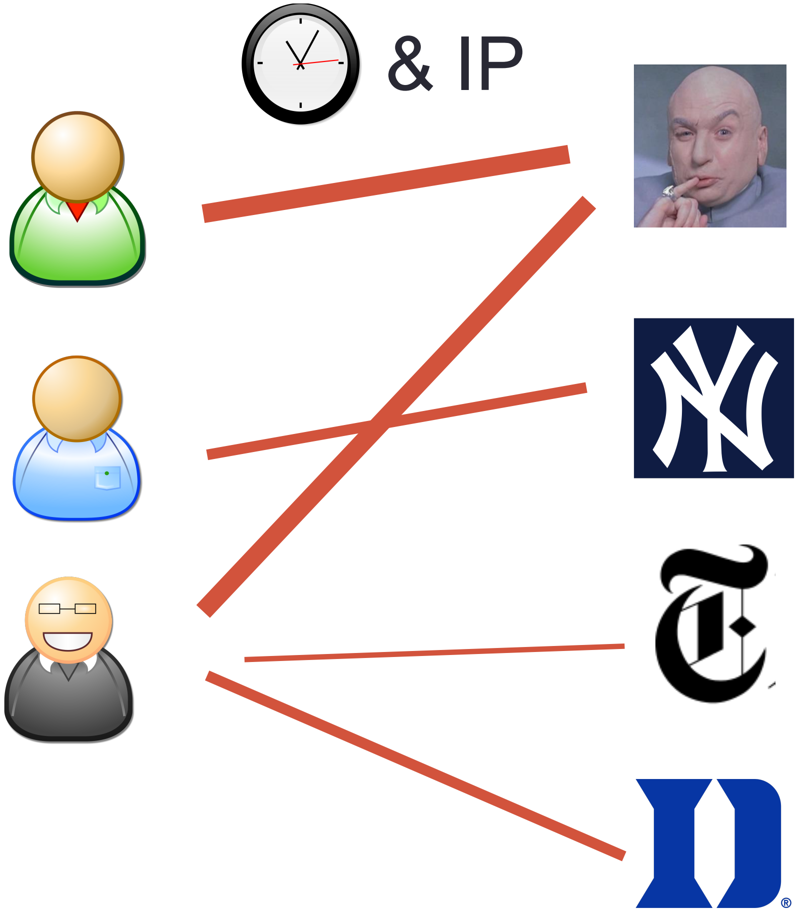


Which is more suspicious?

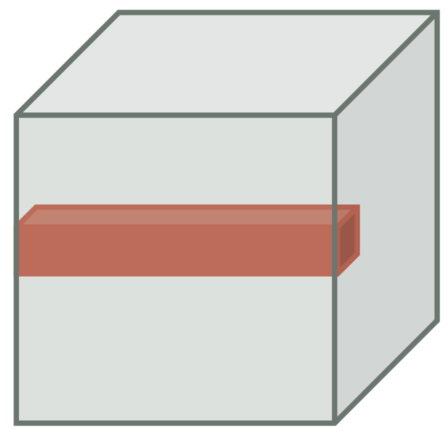


A General Suspiciousness Metric for Dense Blocks in Multimodal Data
Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi,
Shiqiang Yang, Christos Faloutsos
ICDM 2015

Lockstep Behavior on Some Attributes



Maybe suspiciously correlated in IP but not time



A General Suspiciousness Metric for Dense Blocks in Multimodal Data
Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi,
Shiqiang Yang, Christos Faloutsos
ICDM 2015

Lockstep Behavior on Some Attributes

Assume $X_i \sim \text{Poisson}(p)$

$$\hat{f}(\mathbf{n}, \rho, \mathbf{N}, p) = \left(\prod_{i=1}^{K} n_i\right) D_{KL}(\rho || p)$$

Suspiciousness metric is negative log-likelihood of sub-tensor

Satisfies many desirable properties

A General Suspiciousness Metric for Dense Blocks in Multimodal Data
Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi,
Shiqiang Yang, Christos Faloutsos
ICDM 2015

Lockstep Behavior on Some Attributes

	#	User×hashtag×IP×minute	Mass c	Suspiciousness
CROSSSPOT	1	582×3×294×**56,940**	5,941,821	111,799,948
	2	188×1×313×**56,943**	2,344,614	47,013,868
	3	75×1×2×2,061	689,179	19,378,403
HOSVD	1	2,001×1×4×135	77,084	2,931,982
	2	327×1×2×401	212,519	8,599,843
	3	851×2×4×337	103,873	3,903,703

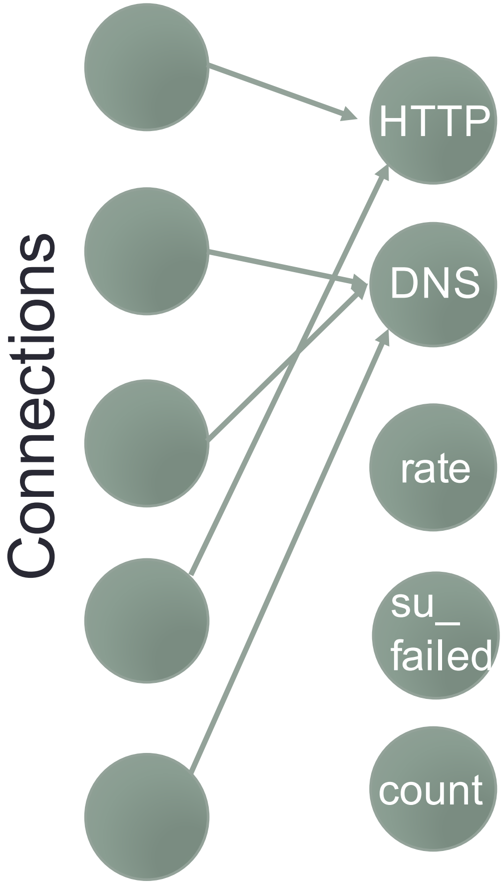
A General Suspiciousness Metric for Dense Blocks in Multimodal Data
Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi,
Shiqiang Yang, Christos Faloutsos
ICDM 2015

Lockstep Behavior on Some Attributes

User ID	Time	IP address (city, province)	Tweet text with hashtag
USER-D	11-18 12:12:51	IP-1 (Deyang, Shandong)	**#Snow#** the Samsung GALAXY SII QQ Service customized version...
USER-E	11-18 12:12:53	IP-1 (Deyang, Shandong)	**#Snow#** the Samsung GALAXY SII QQ Service customized version...
USER-F	11-18 12:12:54	IP-2 (Zaozhuang, Shandong)	**#Snow#** the Samsung GALAXY SII QQ Service customized version...
USER-E	11-18 12:17:55	IP-1 (Deyang, Shandong)	**#Li Ning - a weapon with a hero#** good support activities!
USER-F	11-18 12:17:56	IP-2 (Zaozhuang, Shandong)	**#Li Ning - a weapon with a hero#** good support activities!
USER-D	11-18 12:18:40	IP-1 (Deyang, Shandong)	**#Toshiba Bright Daren#** color personality test to find out your sense...
USER-E	11-18 17:00:31	IP-2 (Zaozhuang, Shandong)	**#Snow#** the Samsung GALAXY SII QQ Service customized version...
USER-D	11-18 17:00:49	IP-2 (Zaozhuang, Shandong)	**#Toshiba Bright Daren#** color personality test to find out your sense...
USER-F	11-18 17:00:56	IP-2 (Zaozhuang, Shandong)	**#Li Ning - a weapon with a hero#** good support activities!

A General Suspiciousness Metric for Dense Blocks in Multimodal Data
Meng Jiang, Alex Beutel, Peng Cui, Bryan Hooi,
Shiqiang Yang, Christos Faloutsos
ICDM 2015

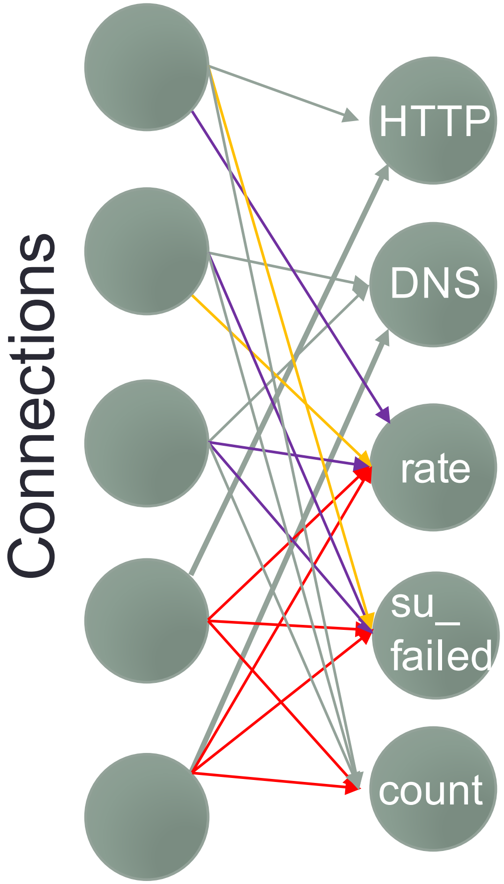
Co-clustering to find network fraud



Handles binary features
(edges without side information)
e.g., connection type

Network Anomaly Detection using Co-Clustering
Evangelos Papalexakis, Alex Beutel, Peter Steenkiste
ASONAM 2012

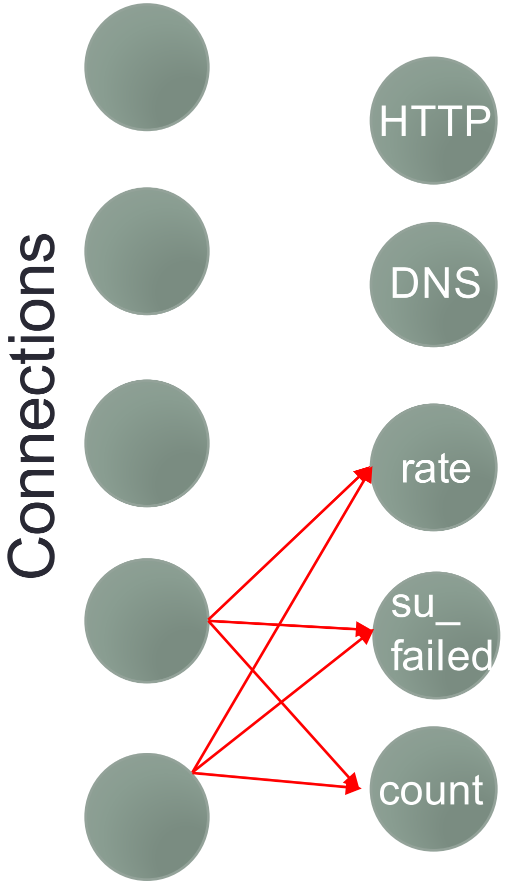
Co-clustering to find network fraud



As well as features with continuous values (edges with side information) e.g., round-trip time, number of requests, etc.

Network Anomaly Detection using Co-Clustering
Evangelos Papalexakis, Alex Beutel, Peter Steenkiste
ASONAM 2012

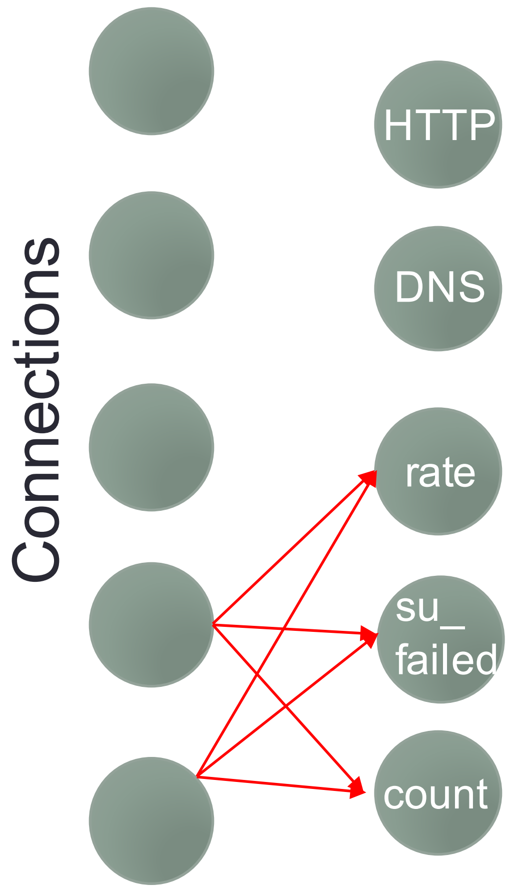
Co-clustering to find network fraud



Co-clustering finds groups of connections with very similar edges through partitioning all rows and columns

Network Anomaly Detection using Co-Clustering
Evangelos Papalexakis, Alex Beutel, Peter Steenkiste
ASONAM 2012

Co-clustering to find network fraud



Cluster	Number of Connections	Percent Normal	Percent Attacks
1	20,156	97.74%	2.26%
2	116,822	5.30%	94.70%
3	29,591	93.34%	6.66%
4	281,437	0.21%	99.79%
5	46,014	93.85%	6.15%

Each cluster is nearly all normal connections or all attacks

Network Anomaly Detection using Co-Clustering
Evangelos Papalexakis, Alex Beutel, Peter Steenkiste
ASONAM 2012

Practitioner's Guide to Detecting Fraud

Method	Graph Type	Node Attributes	Edge Attributes	Seed Labels
COI	Undirected			✓
OddBall	Undirected			
Blackholes & Volcanoes	Directed			
(Anti)-Social	Bipartite			
SODA	Undirected	✓		
FocusCO	Undirected	✓		
CopyCatch	Bipartite		✓	
SynchroTrap	Bipartite+	✓	✓	
CrossSpot	Bipartite		✓	
Co-Clustering	Bipartite*		✓	
PICS	Undirected	✓		

Recap

- **COI:** Guilt-by-Association
- **Oddball:** Unusually dense graphs are suspicious (along with other surprising patterns described in the paper)
- **Blackholes and Volcanos** can be indicative of trading rings
- **(Anti)social behavior** – In packet traces, cliques are normal and bridges connecting cliques are suspicious
- **SODA:** Attributed subnetwork anomalies
- **FocusCO:** Learn model of normal attributes among communities and find outliers in the community
- **CopyCatch:** Temporally near-bipartite cores are extra-suspicious
- **SynchroTrap:** Generalize CopyCatch to handle extra data like IP addresses and repeat actions
- **CrossSpot:** Find suspicious behavior in subset of action attributes
- **Co-clustering:** Global partitioning to find locally similar regions; can include edges with side information.