Inferring Strange Behavior from Connectivity Pattern in Social Networks

Meng Jiang, Peng Cui, Shiqiang Yang
(Tsinghua, Beijing)
Alex Beutel, Christos Faloutsos (CMU)

清華大學 Tsinghua University
Carnegie Mellon University

What is Strange Behavior?

- "Who-follows-whom" network with **billions** of edges: Twitter, Weibo, etc.

What is Strange Behavior?

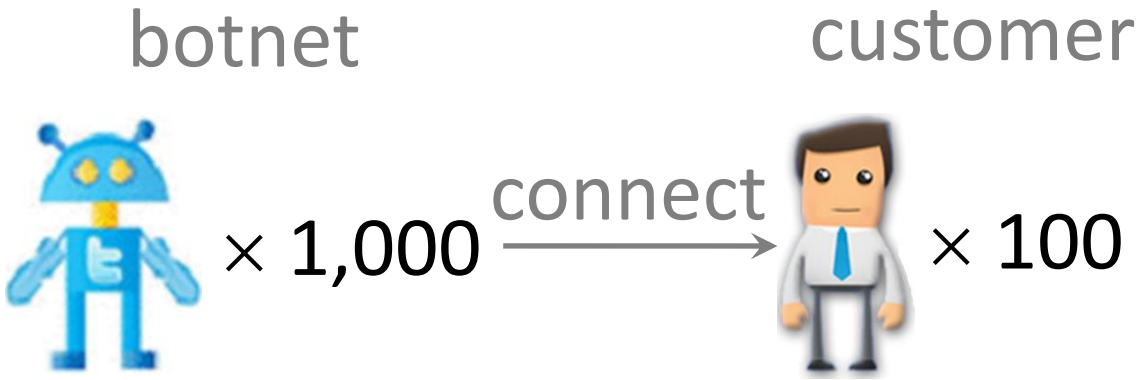
- Sell followers: “Become a Twitter Rockstar”

0.9 TWD per edge

清華大學 Tsinghua University

Carnegie Mellon University

What is Strange Behavior?



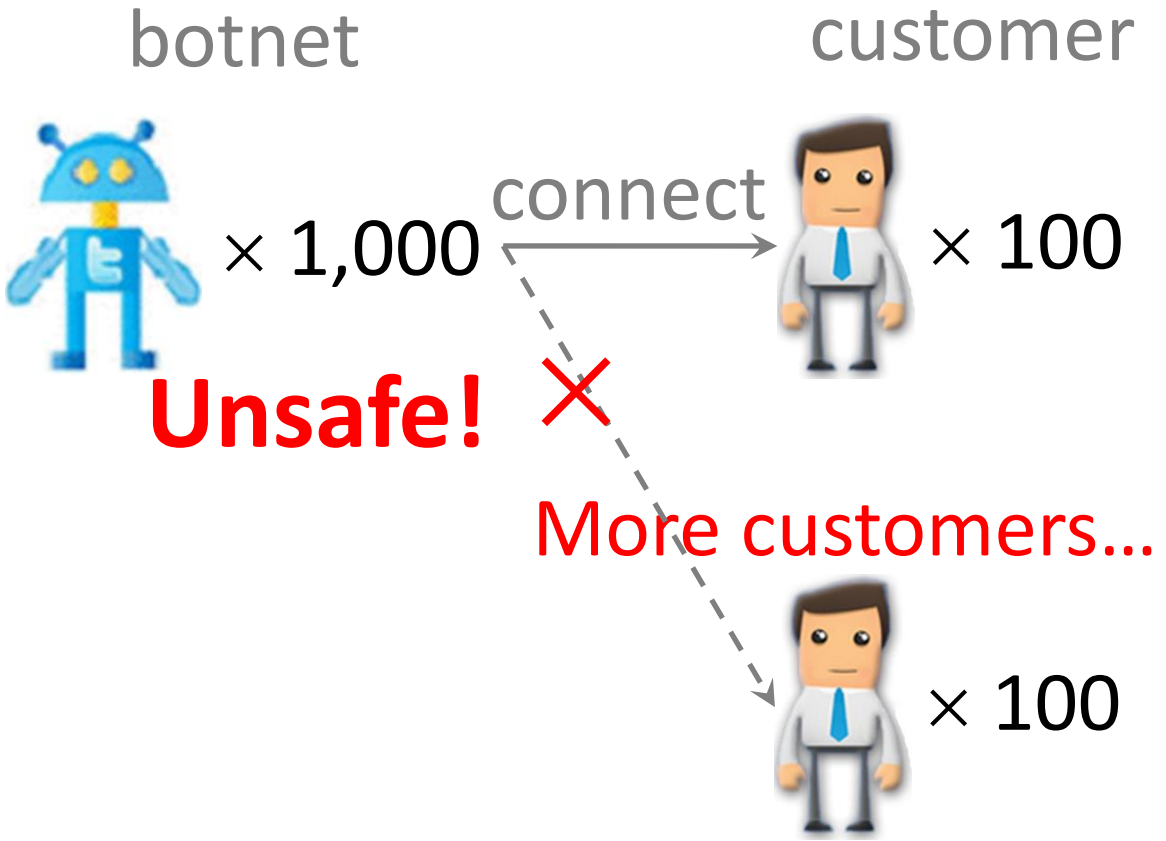
Tsinghua University
Carnegie Mellon University

What is Strange Behavior?

#follower↑+1,000

清華大學 Tsinghua University Carnegie Mellon University

What is Strange Behavior?



清華大學 Tsinghua University
Carnegie Mellon University

What is Strange Behavior?

botnet

customer

× 1,000 connect

× 100

More customers…

× 5,000 connect

× 100

清華大學 Tsinghua University

Carnegie Mellon University

What is Strange Behavior?

botnet customer


× 1,000 connect

× 100

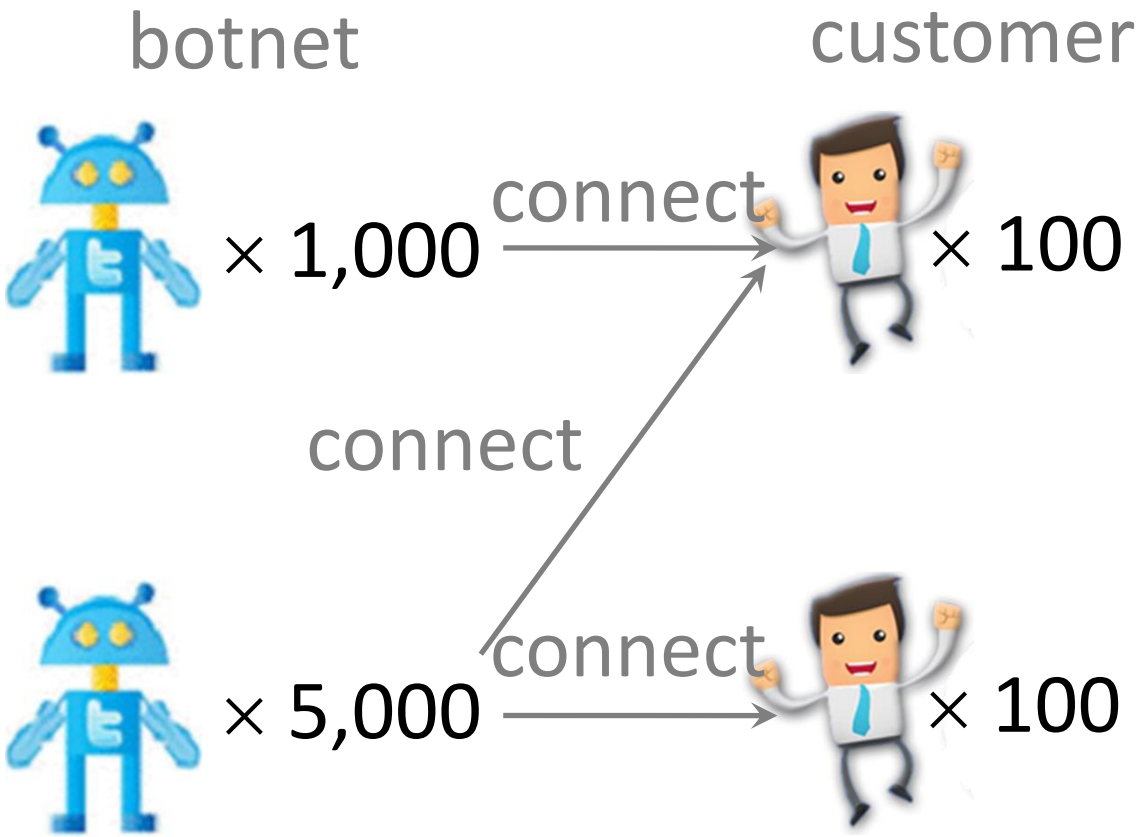
I want more followers…


× 5,000 connect

× 100

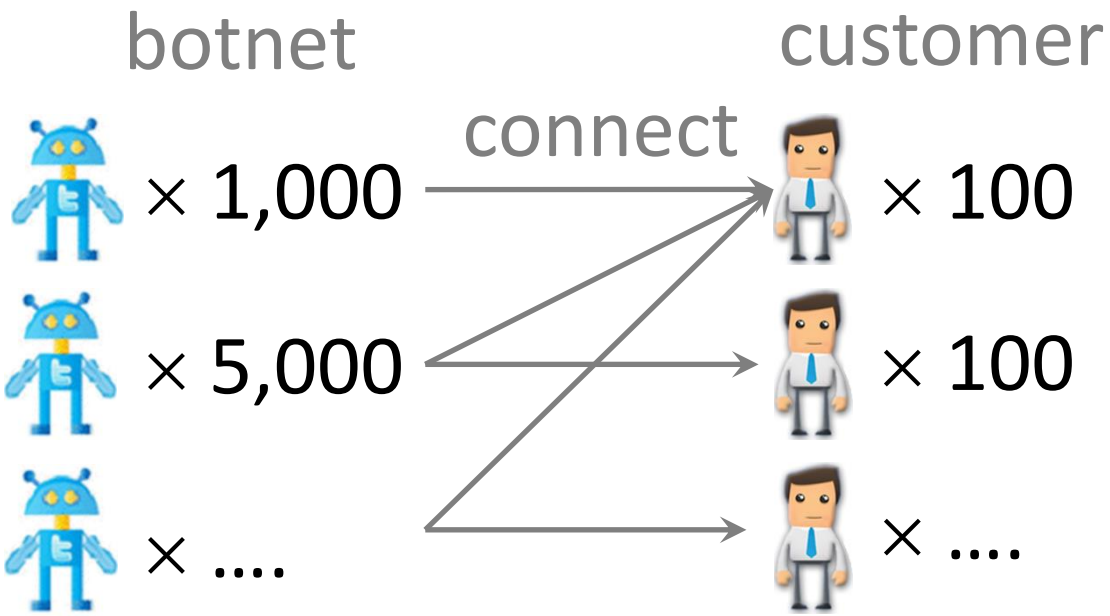
Tsinghua University Carnegie Mellon University

What is Strange Behavior?



Tsinghua University
Carnegie Mellon University

What is Strange Behavior?



More groups of customers

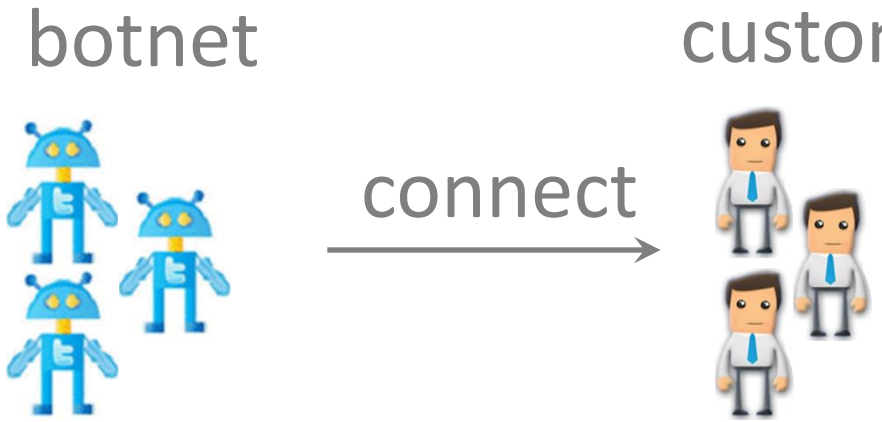
More groups of botnets

More companies….

清華大學 Tsinghua University

Carnegie Mellon University

What is Strange Behavior?



Detect dense biparitite cores!

How can we evade detection?

Some other activity!

清華大學 Tsinghua University

Carnegie Mellon University

What is Strange Behavior?

botnet

customer

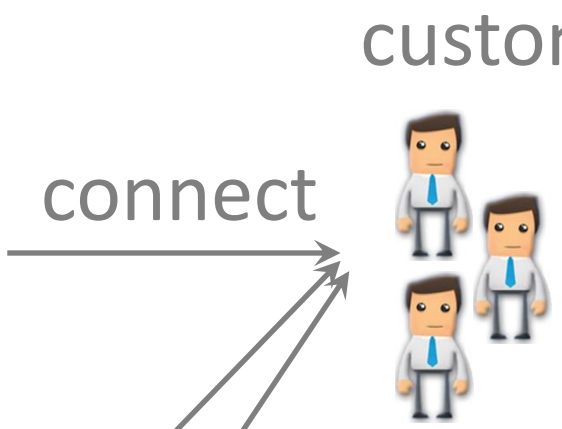
connect

"Camouflage":
may connect
to popular idols
to look normal

清華大學 Tsinghua University

Carnegie Mellon University

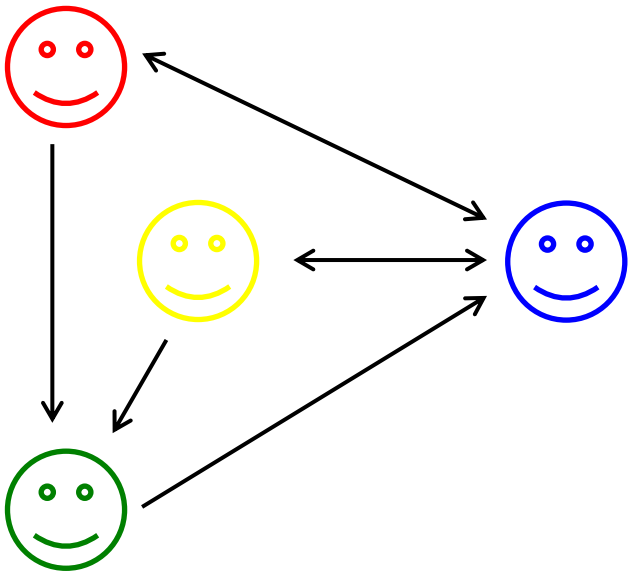
What is Strange Behavior?



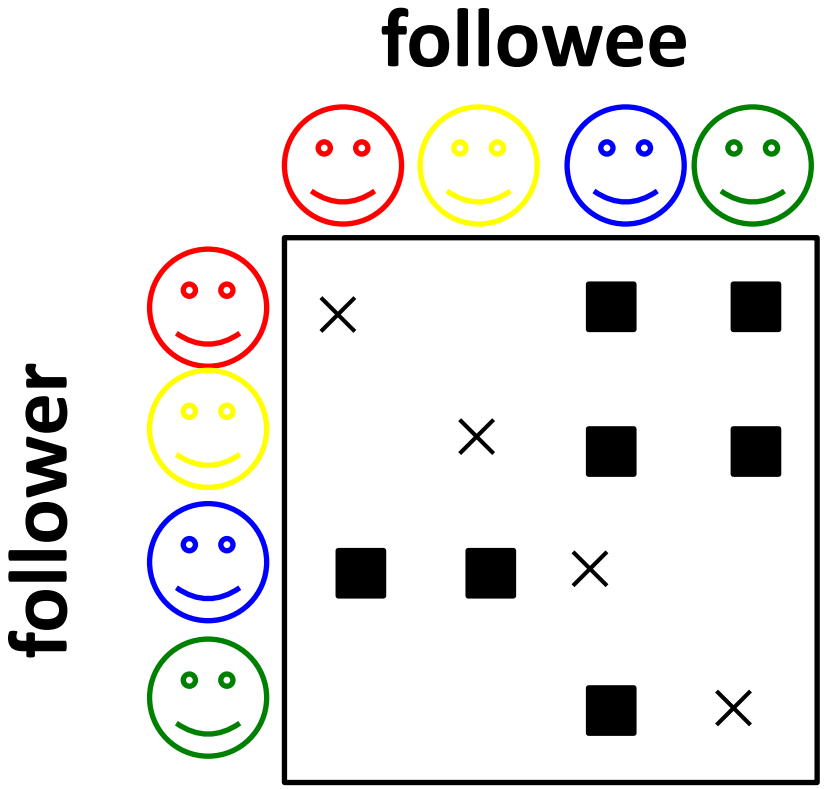
"Fame":
may have
a few honest
followers

Tsinghua University
Carnegie Mellon University

Adjacency Matrix Reminder



Graph Structure

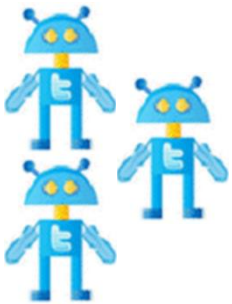


Adjacency Matrix

Strange ⟶ Lockstep Behavior

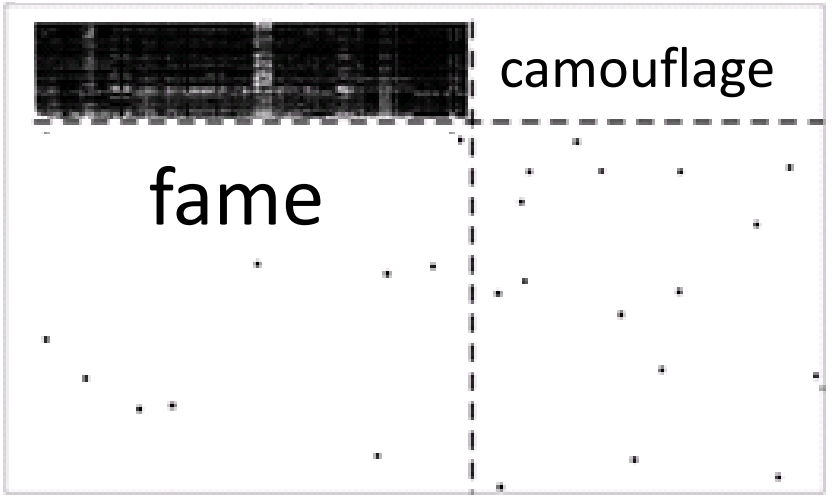
botnet

customer



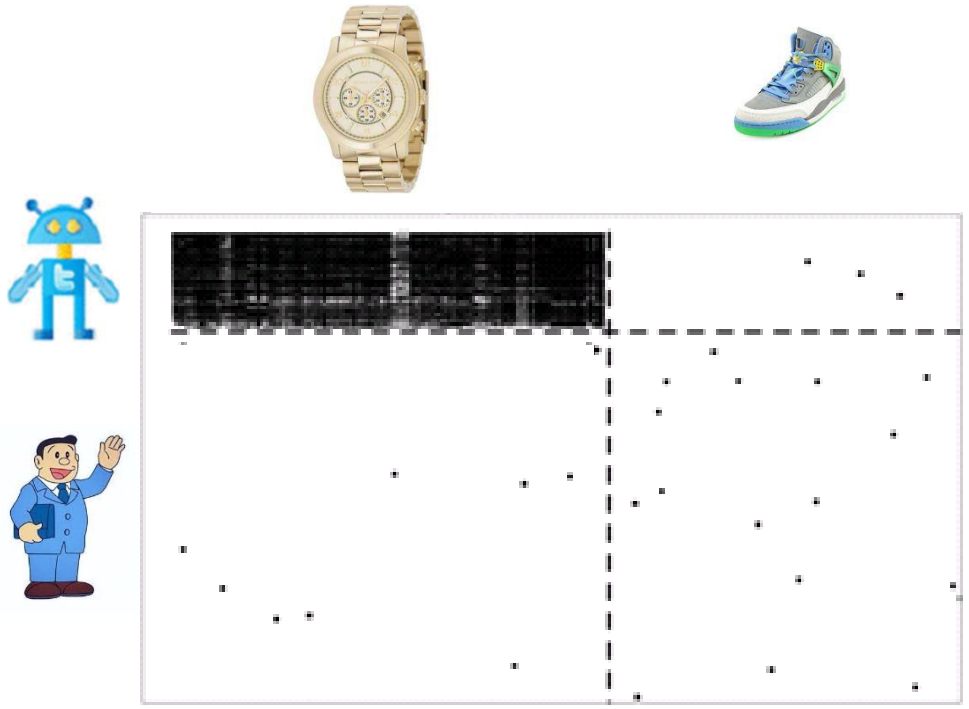
connect

- Groups
- Acting together
- Little other activity



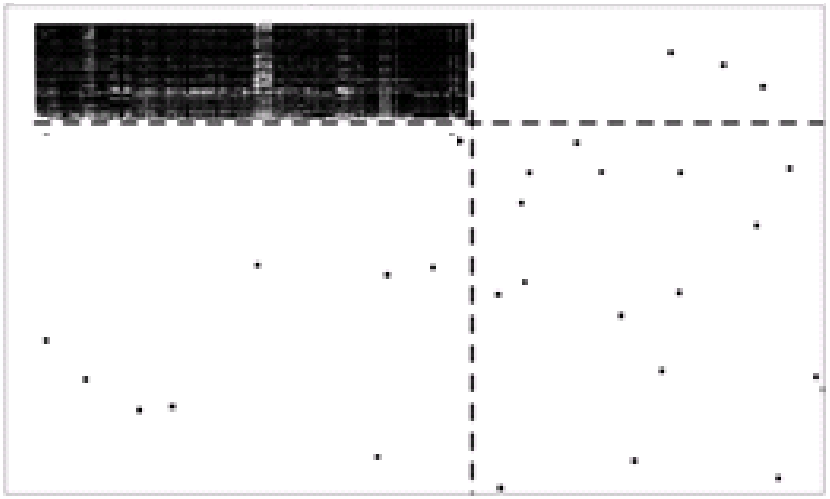
More Applications

- eBay reviews



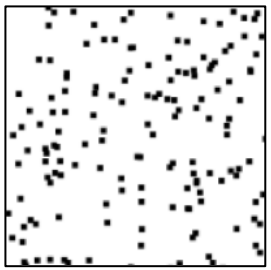
More Applications

- Facebook “Likes”

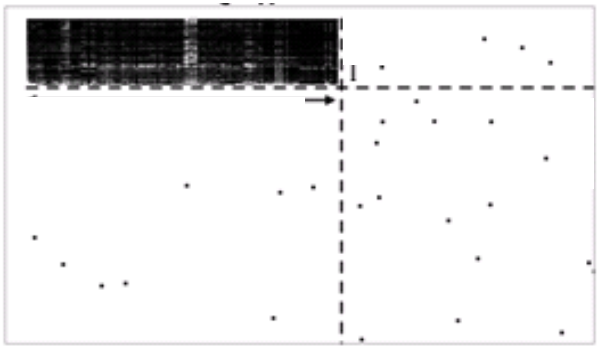


Problem Definition

- Given adjacency matrix



reordering

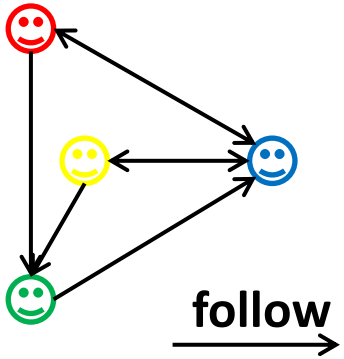


- Find Strange = “Lockstep” Behavior

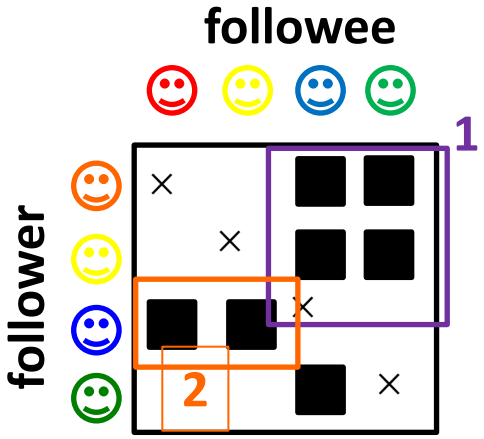
Outline

- Method
  - SVD Reminder
  - “Spectral Subspace Plot”
  - BP-based Algorithm
- Experiments
  - Dataset
  - Real Data
  - Synthetic Data

SVD Reminder

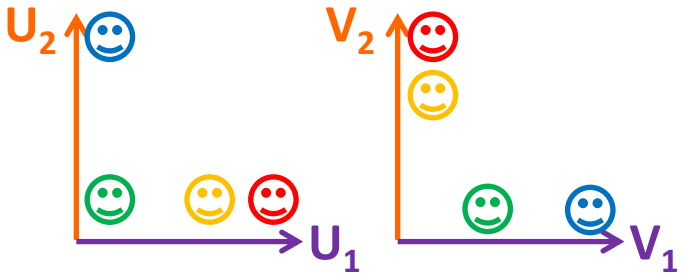


Graph Structure



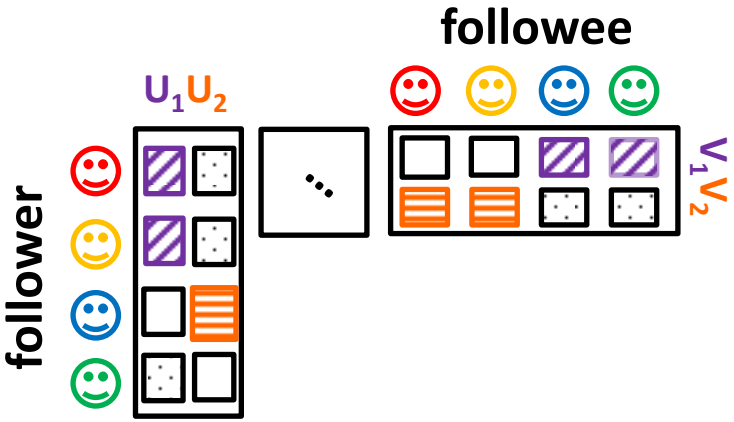
Adjacency Matrix

Pairs of singular vectors:



"Spectral Subspace Plot"

SVD: $A=USV^T$



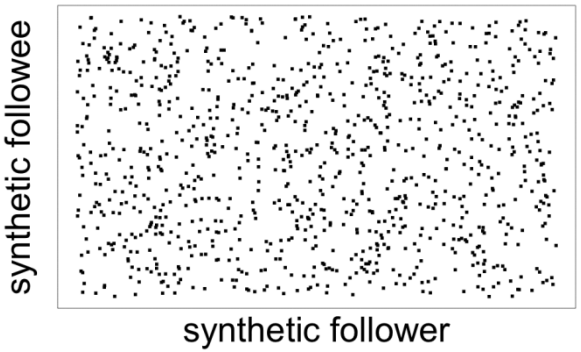
Outline

- Method
  - SVD Reminder
  - "Spectral Subspace Plot"
  - BP-based Algorithm
- Experiments
  - Dataset
  - Real Data
  - Synthetic Data

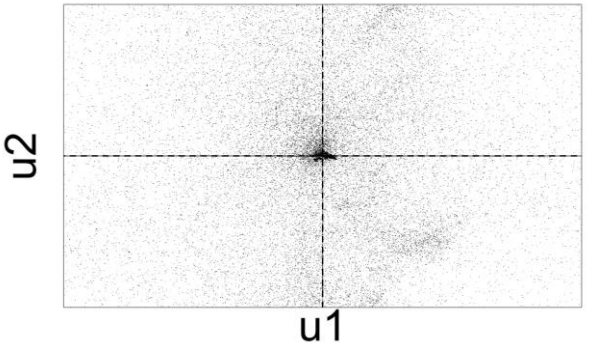
Lockstep and *Spectral Subspace Plot*

- Case #0: No lockstep behavior in random power law graph of 1M nodes, 3M edges
- Random ⟷ "Scatter"

Adjacency Matrix



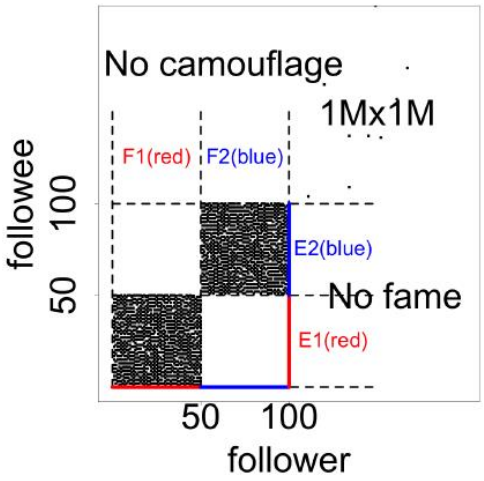
Spectral Subspace Plot



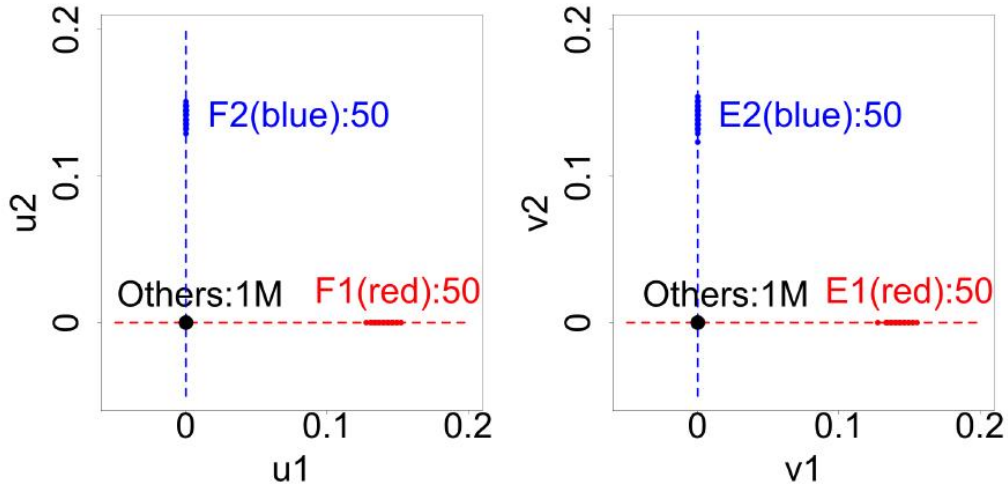
Lockstep and *Spectral Subspace Plot*

- Case #1: non-overlapping lockstep
- "Blocks" $\longleftrightarrow$ "Rays"

Adjacency Matrix



Spectral Subspace Plot

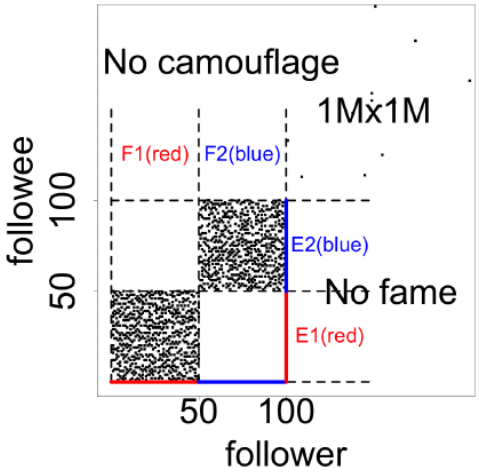


Rule 1 (short "rays"): two blocks, high density (90%), no "camouflage", no "fame"

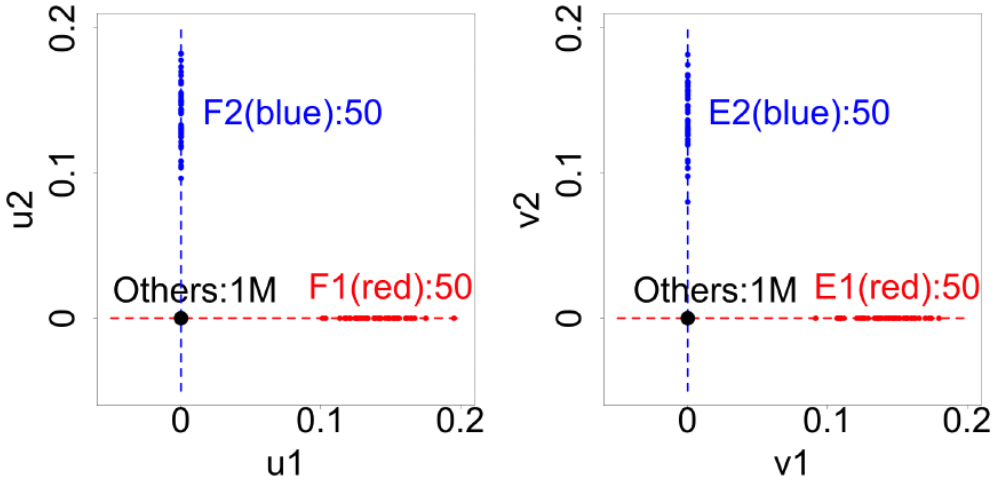
Lockstep and *Spectral Subspace Plot*

- Case #2: non-overlapping lockstep
- "Blocks; low density" ⟷ Elongation

Adjacency Matrix



Spectral Subspace Plot

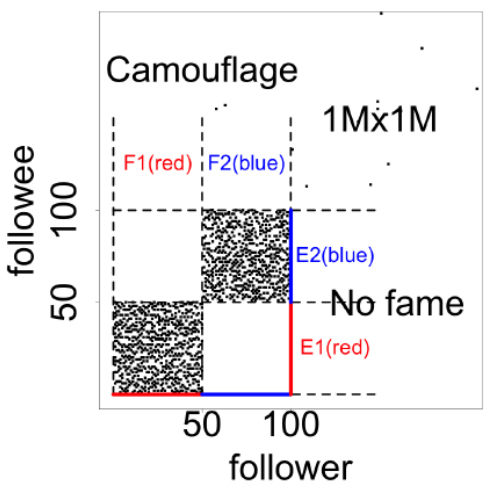


Rule 2 (long "rays"): two blocks, low density (50%), no "camouflage", no "fame"

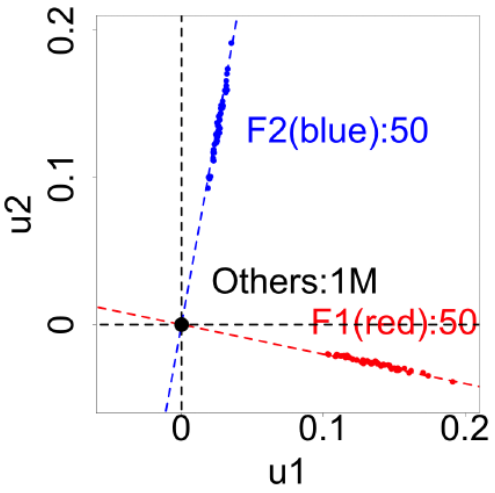
Lockstep and *Spectral Subspace Plot*

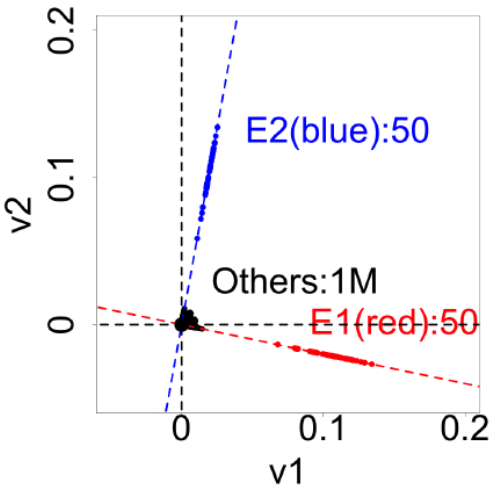
- Case #3: non-overlapping lockstep
- "**Camouflage**" (or "Fame") ⟷ Tilting "Rays"

Adjacency Matrix



Spectral Subspace Plot



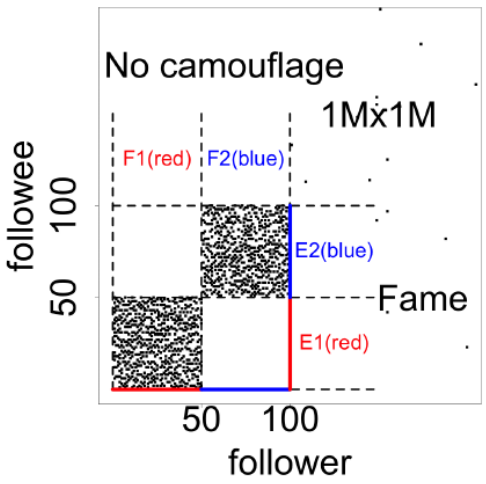


Rule 3 (tilting "rays"): two blocks, with "camouflage", no "fame"

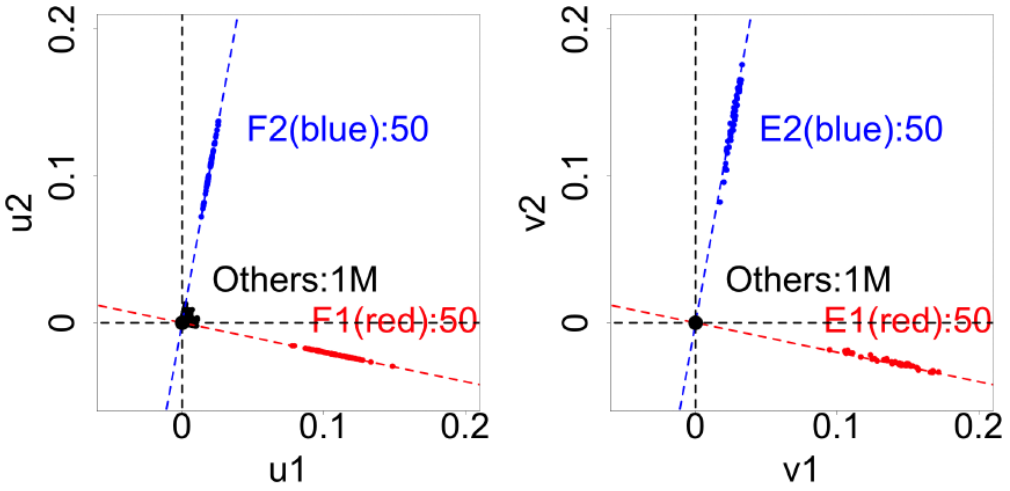
Lockstep and *Spectral Subspace Plot*

- Case #3: non-overlapping lockstep
- "Camouflage" (or "**Fame**") ⟷ Tilting "Rays"

Adjacency Matrix



Spectral Subspace Plot



Rule 3 (tilting "rays"): two blocks, no "camouflage", with "fame"

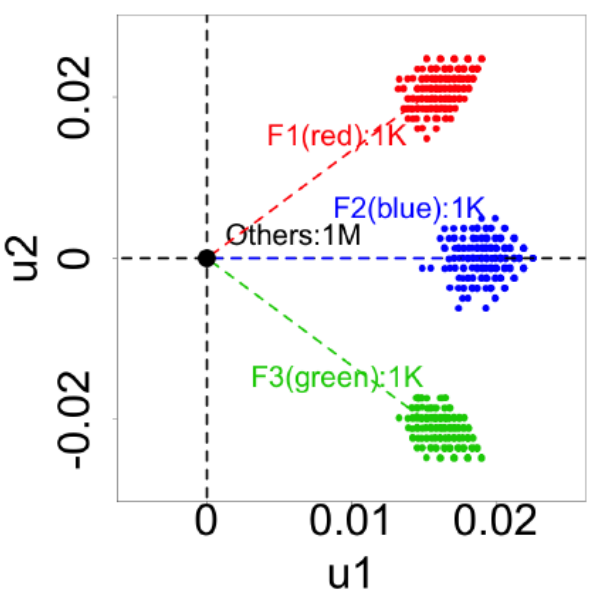
Lockstep and *Spectral Subspace Plot*

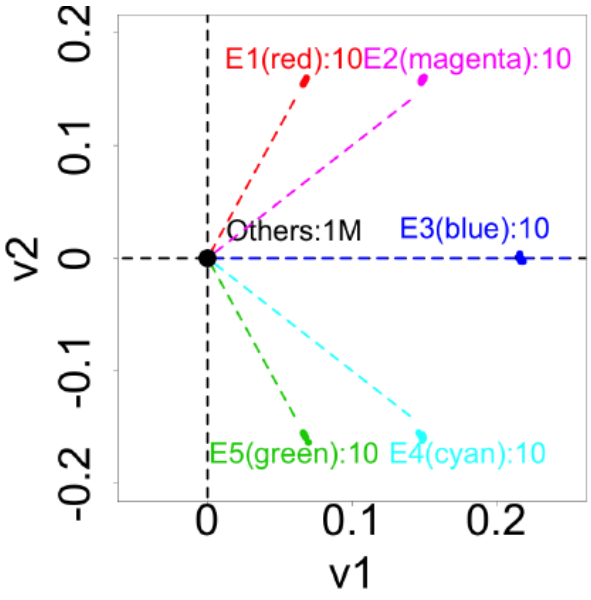
- Case #4: ? lockstep
- "?" ⟷ "Pearls"

Adjacency Matrix

?

Spectral Subspace Plot

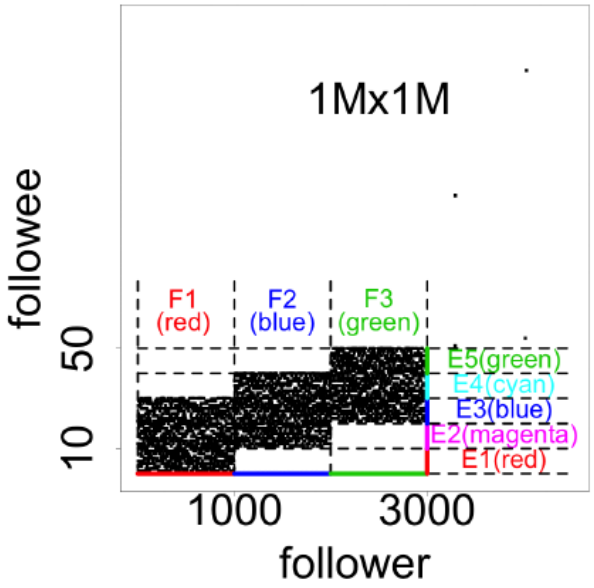




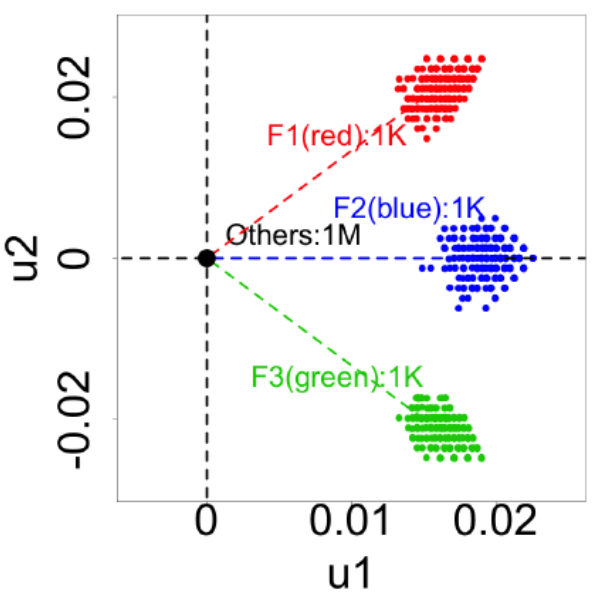
Lockstep and *Spectral Subspace Plot*

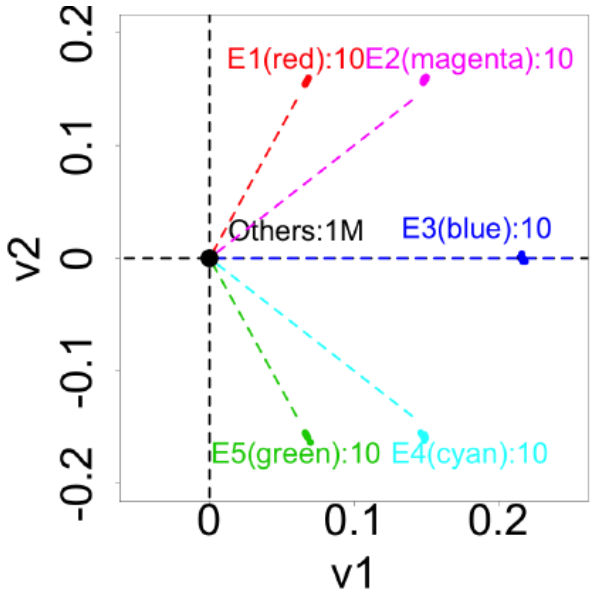
- Case #4: **overlapping** lockstep
- "**Staircase**" ⟷ "Pearls"

Adjacency Matrix



Spectral Subspace Plot





Rule 4 ("pearls"): a "staircase" of three partially overlapping blocks.

Outline

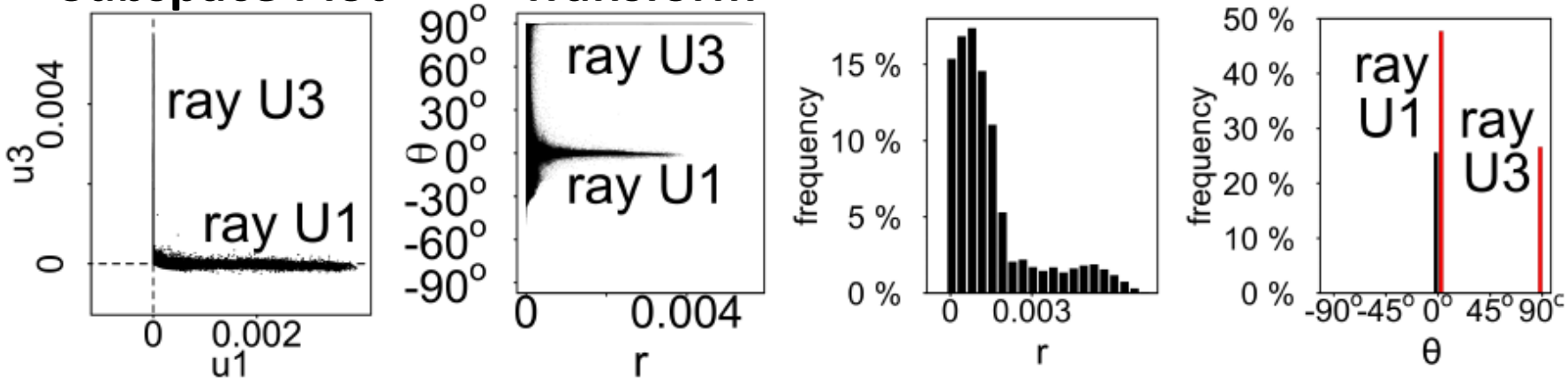
- Method
  - SVD Reminder
  - “Spectral Subspace Plot”
  - BP-based Algorithm
- Experiments
  - Dataset
  - Real Data
  - Synthetic Data

Algorithm

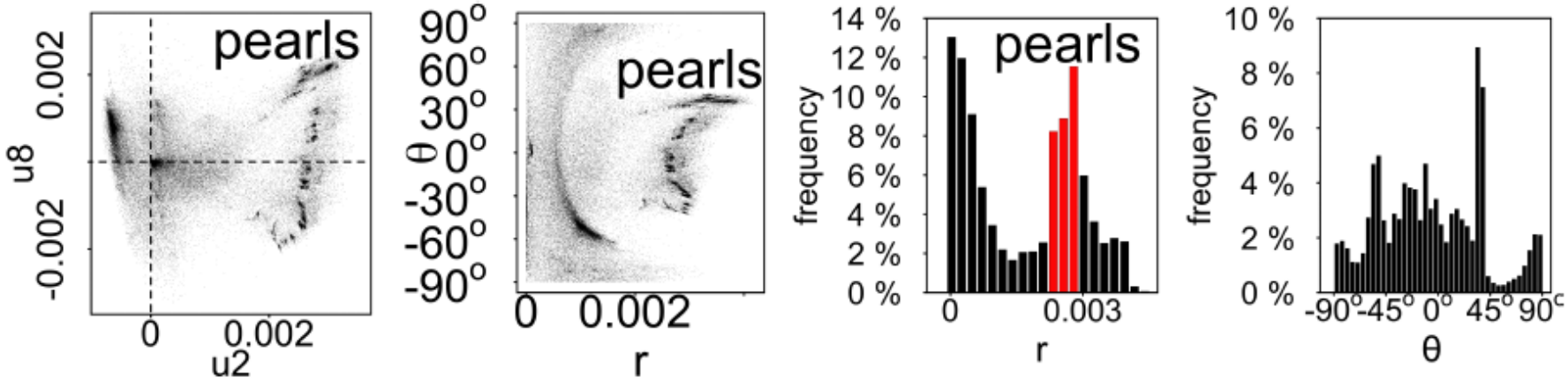
- Step 1: Seed selection
  - Spot “Rays” and “Pearls”
  - Catch seed followers
- Step 2: Belief Propagation
  - Blame followees with strange followers
  - Blame followers with strange followees

Automatically Spot "Rays" and "Pearls"

Spectral Subspace Plot | **Polar Coordinate Transform** | **Histograms**



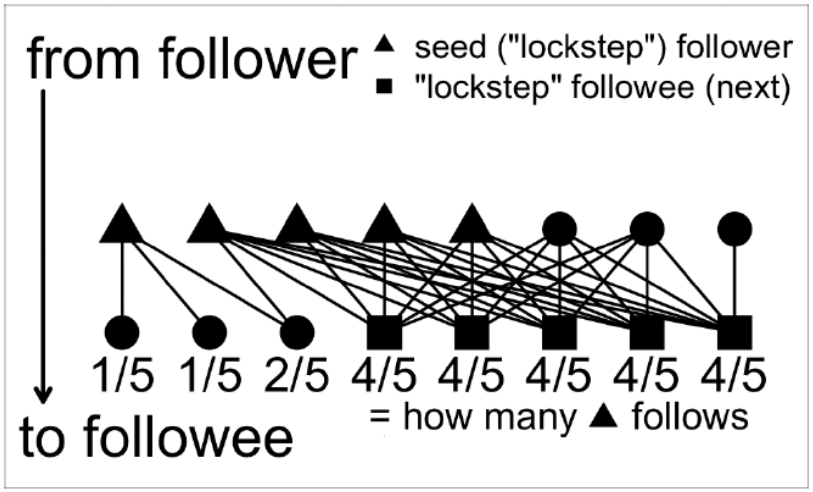
"rays" show two apparent spikes on θ frequency at 0^o and 90^o



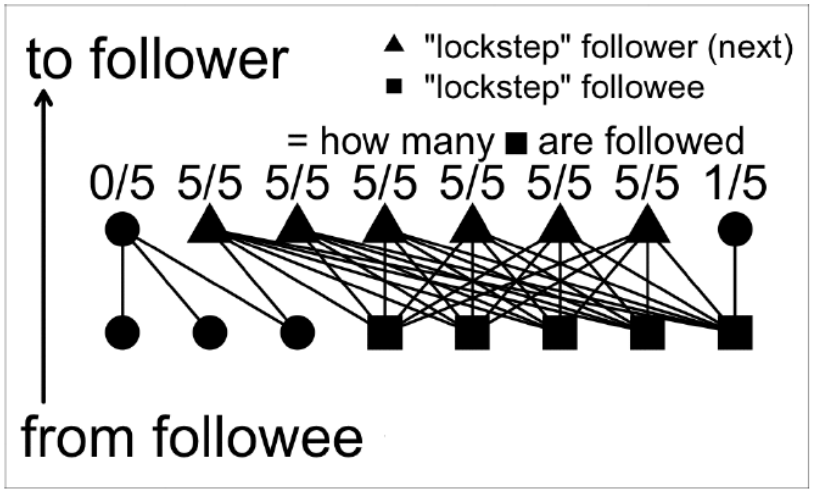
"pearls" show a spike on r frequency at a much-greater-than-zero value

BP-based Algorithm

- Blame followees with strange followers
- Blame followers with strange followees



(a) select "lockstep" followees: from (seed) followers to followees



(b) select "lockstep" followers: from followees to followers

Outline

- Method
  - SVD Reminder
  - “Spectral Subspace Plot”
  - BP-based Algorithm
- Experiments
  - Dataset
  - Real Data
  - Synthetic Data

Dataset

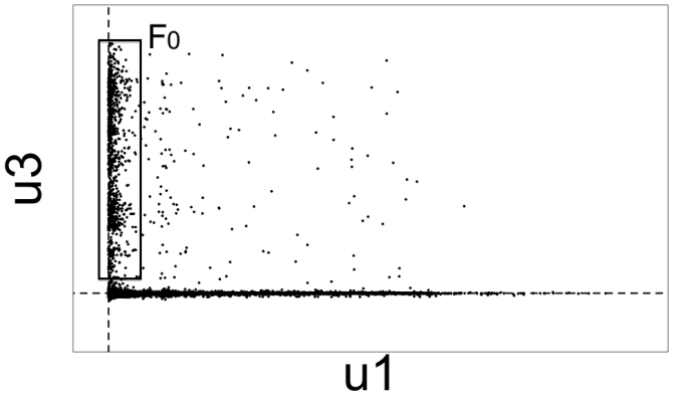
- Tencent Weibo
- 117 million nodes (users)
- 3.33 billion directed edges

Outline

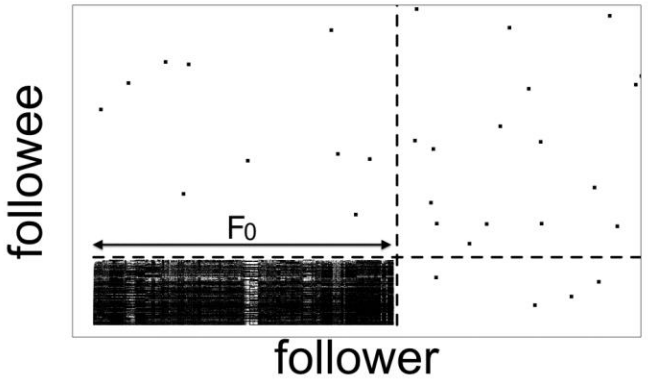
- Method
  - SVD Reminder
  - “Spectral Subspace Plot”
  - BP-based Algorithm
- Experiments
  - Dataset
  - Real Data
  - Synthetic Data

Real Data

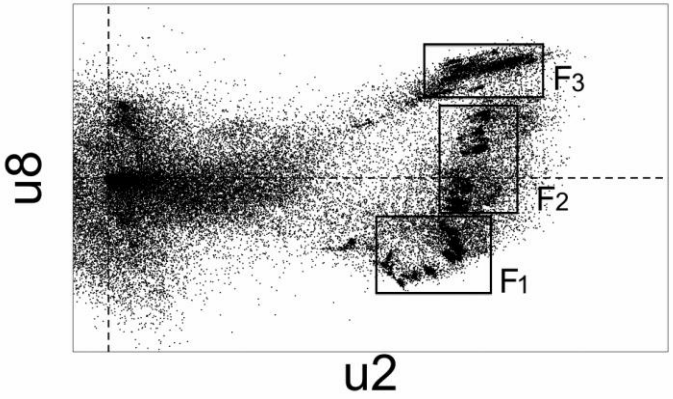
“Rays”



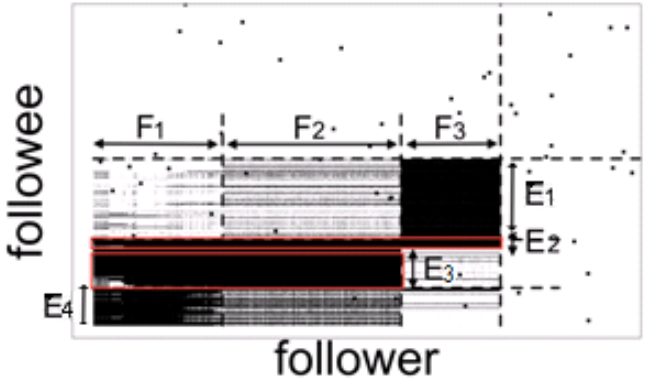
“Block”



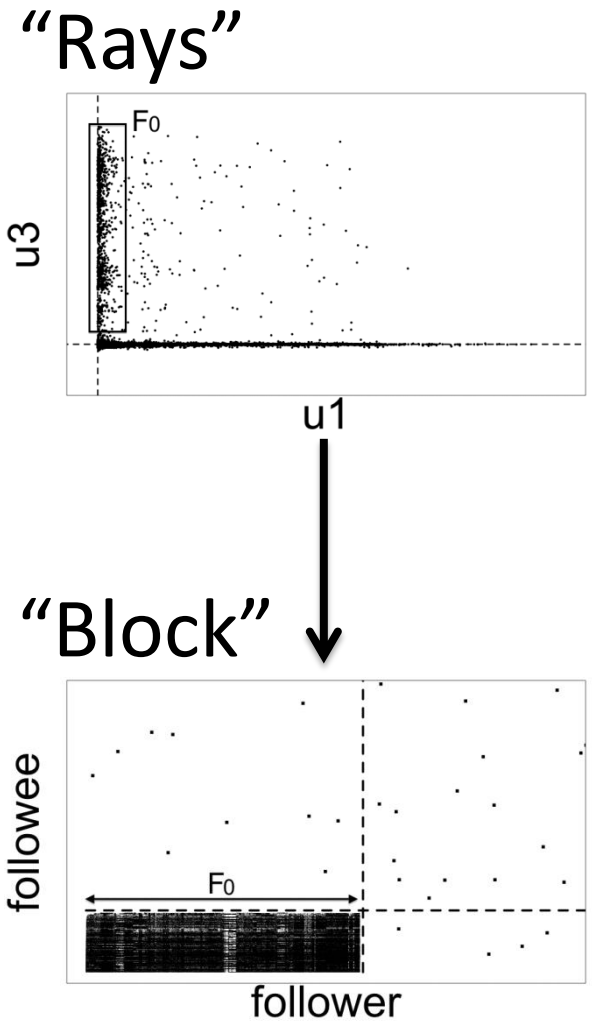
“Pearls”



“Staircase”



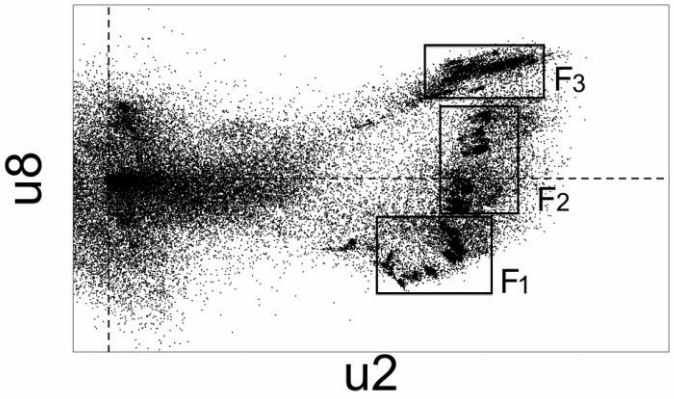
Real Data



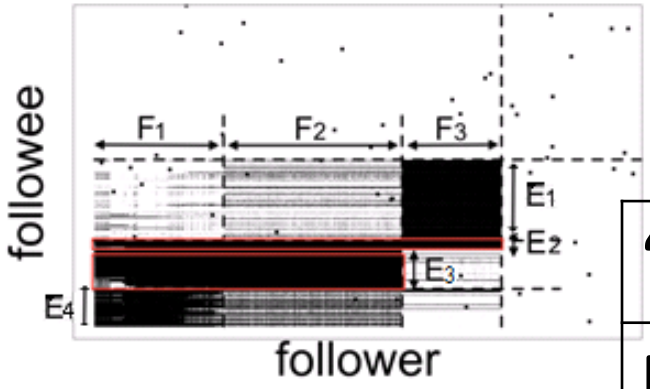
	“ray” F_0	“p
Num. seeds	100	
Size of block	$83,208 \times 30$	3, 1
Density	81.3%	
Camouflage	0.14%	
Fame	0.05%	
Out-degree	231±109	
In-degree	2.0±1.4	

Real Data

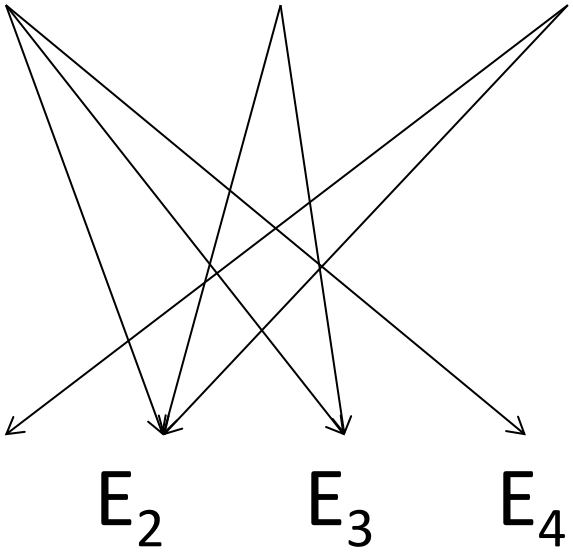
"Pearls"



"Staircase"



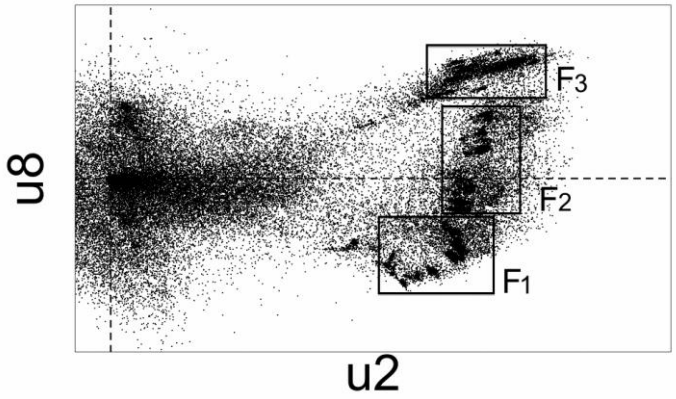
3,188 in F_1 7,210 in F_2 2,457 in F_3



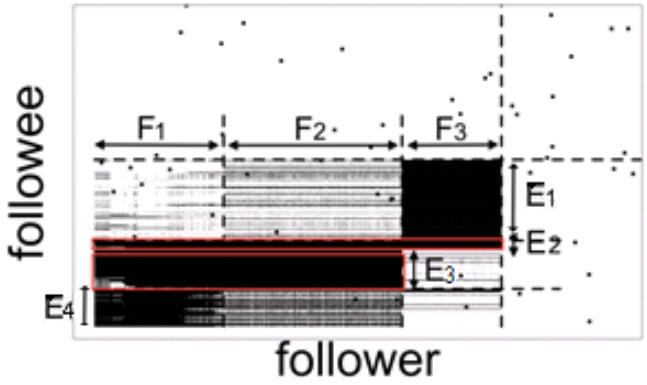
"F-E"	F_1-…	F_2-…	F_3-…
Density	91.3%	92.6%	89.1%

Real Data

"Pearls"



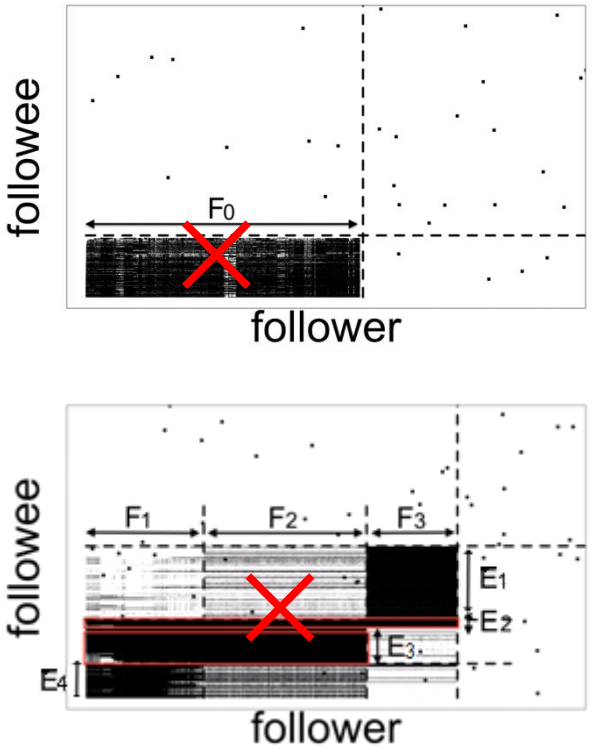
"Staircase"

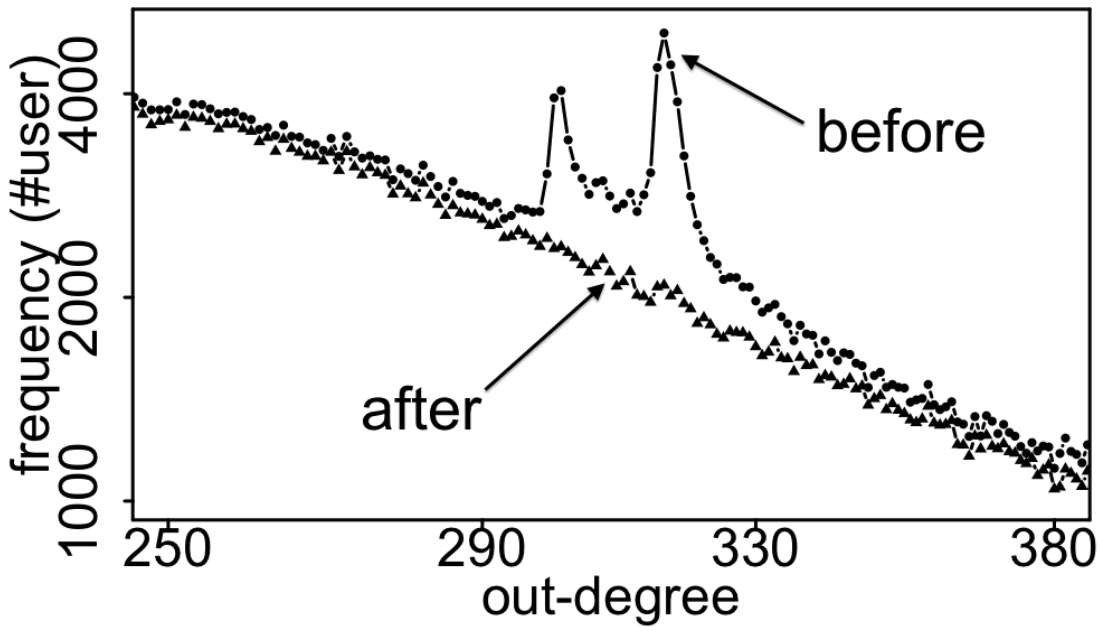


	"pearl" F_1	"pearl" F_2	"pearl" F_3	"pearl" Total
	1,239	107	990	—
)	$3,188 \times 135$	$7,210 \times 79$	$2,457 \times 148$	$10,952 \times 370$
	91.3%	92.6%	89.1%	43.1%
	0.06%	0.10%	0.05%	0.07%
	1.93%	1.94%	1.72%	1.73%
	310±7	312±7	304±5	310±7
	[illegible]	[illegible]	[illegible]	[illegible]

Real Data

- Spikes on the out-degree distribution



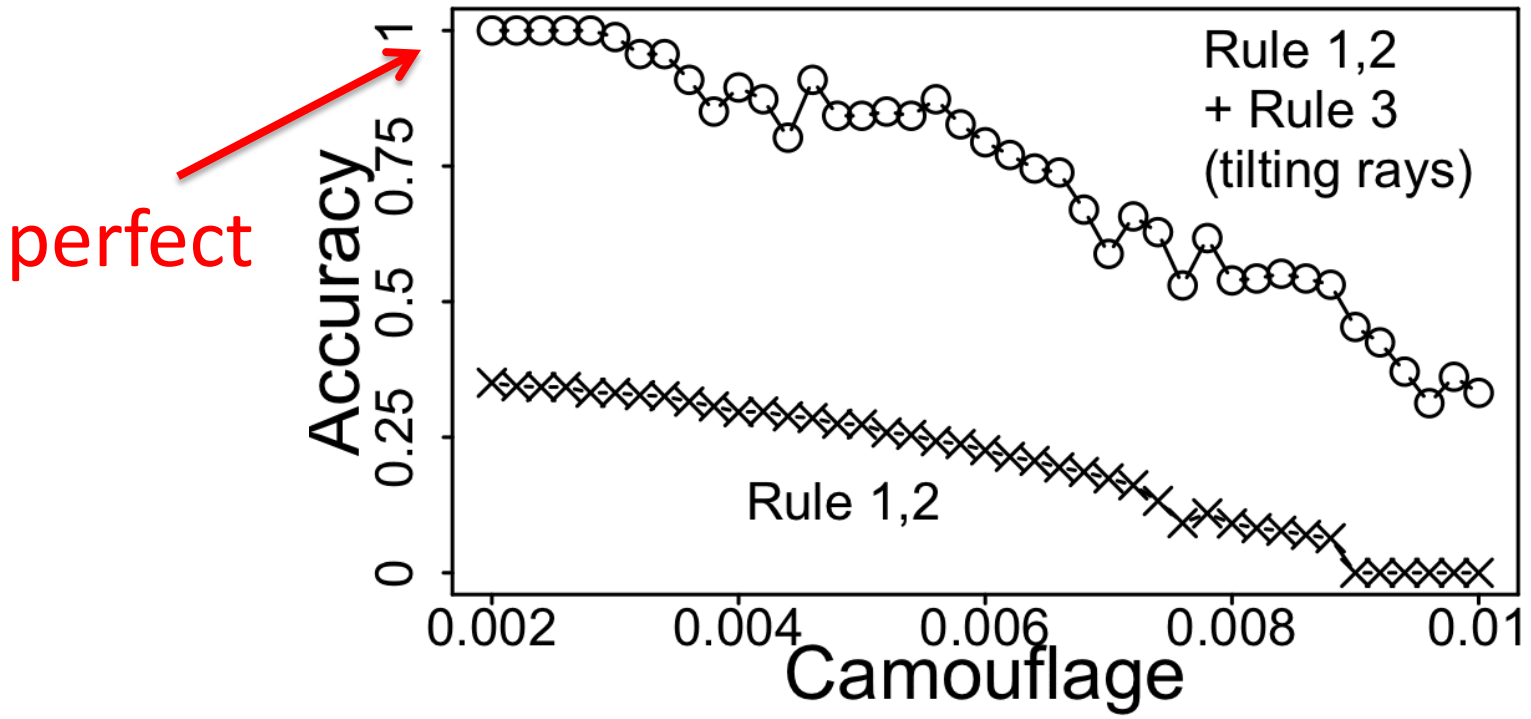


Outline

- Method
  - SVD Reminder
  - “Spectral Subspace Plot”
  - BP-based Algorithm
- Experiments
  - Dataset
  - Real Data
  - Synthetic Data

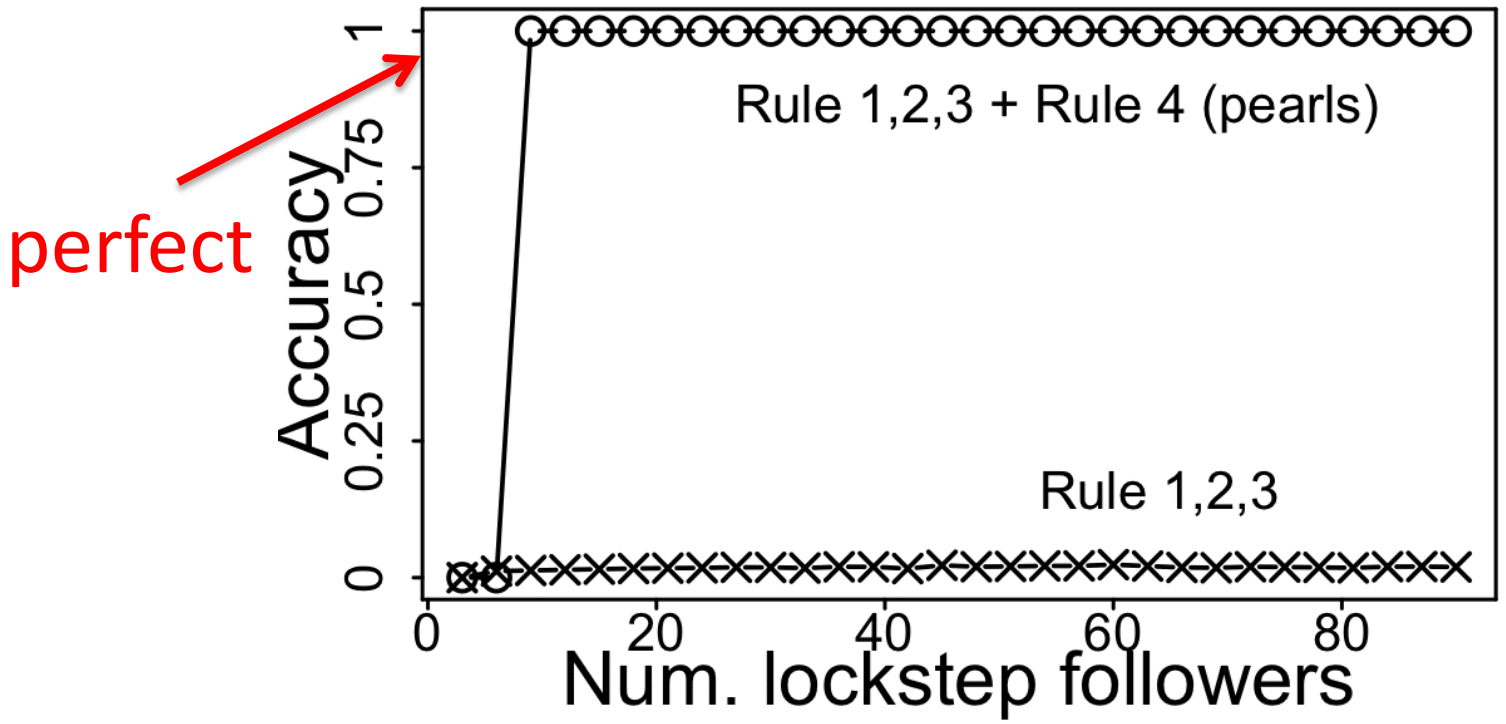
Synthetic Data

- Inject lockstep behavior with "**camouflage**"



Synthetic Data

- Inject **overlapping** lockstep behavior



Contributions

- Different types of lockstep behavior
- A handbook (rules) to infer lockstep behavior with connectivity patterns
- An algorithm to catch the suspicious nodes
- Remove spikes on out-degree distribution

清華大學 Tsinghua University Carnegie Mellon University

Thank you!