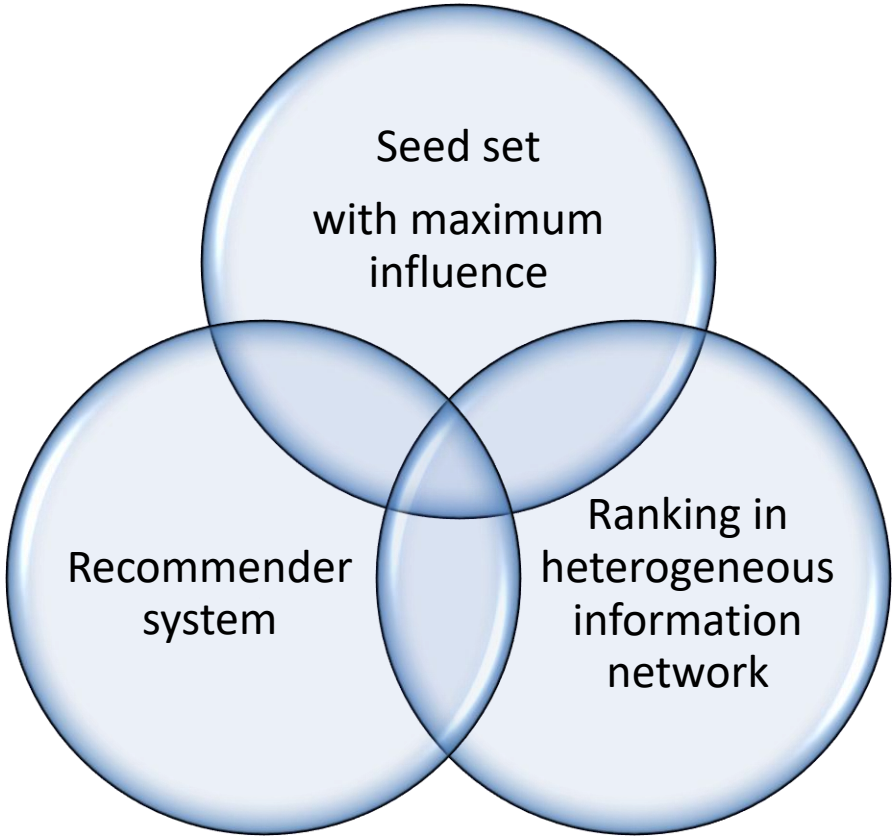
电子科技大学
University of Electronic Science and Technology of China

Complex Networks: From Ranking to Application

Discussion： Complex network group

Data Mining Lab, Big Data Research Center, UESTC
Email：yuzhongjing@foxmail.cn
http://staff.uestc.edu.cn/yuzhongjing

Outline

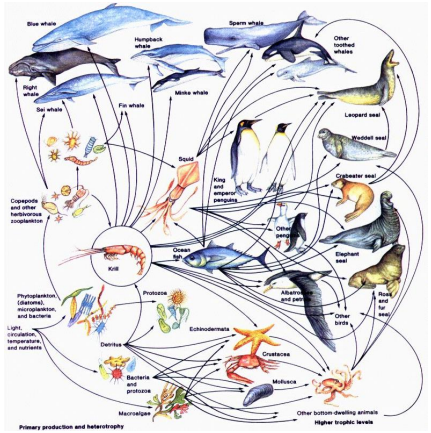


Importance **Ranking** in complex network

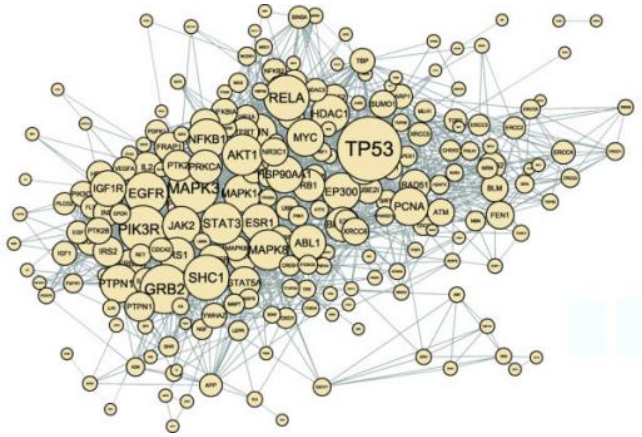
Background

Importance **ranking** in complex network (**vital nodes detection**)

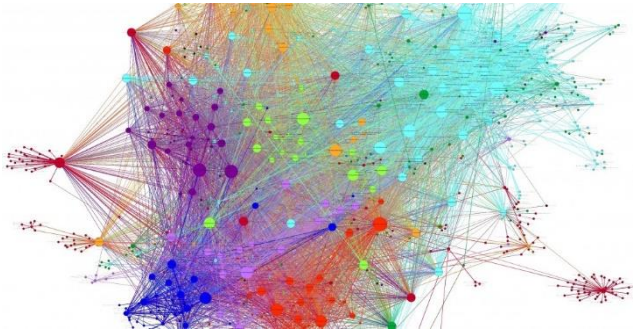
—which nodes are important?



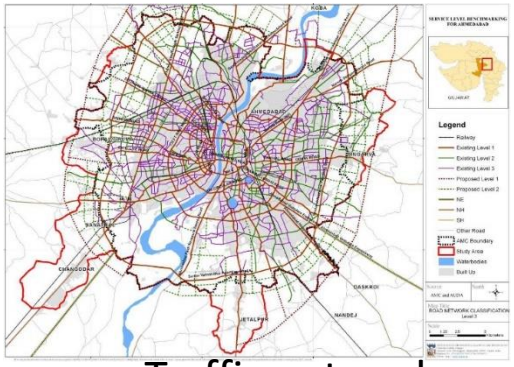
Food web



Protein network

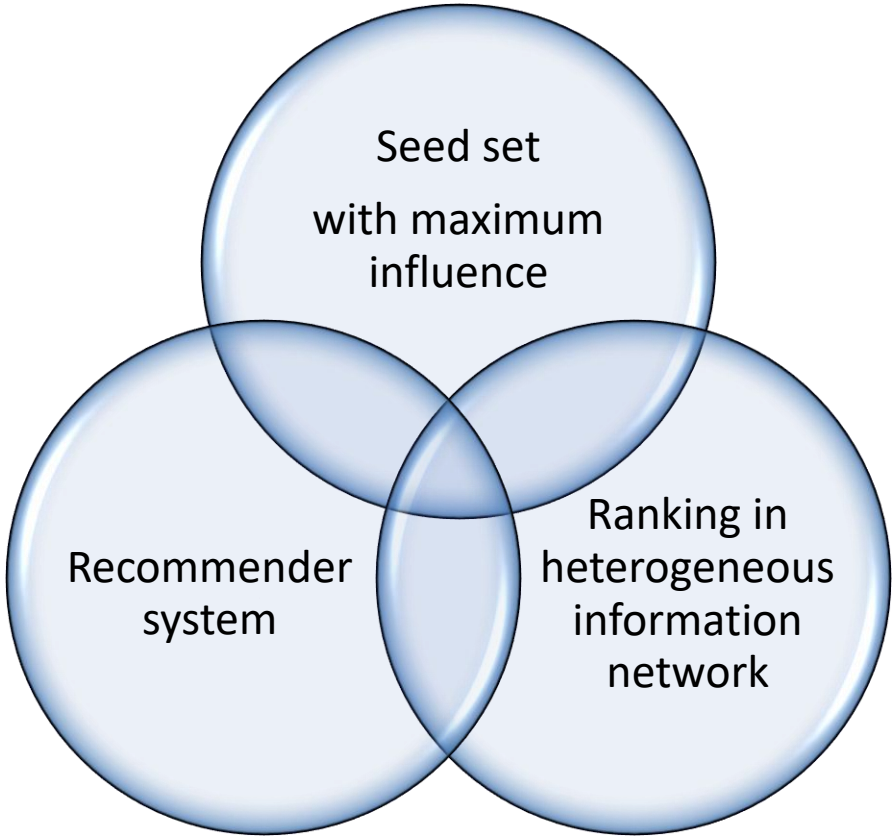


Social network



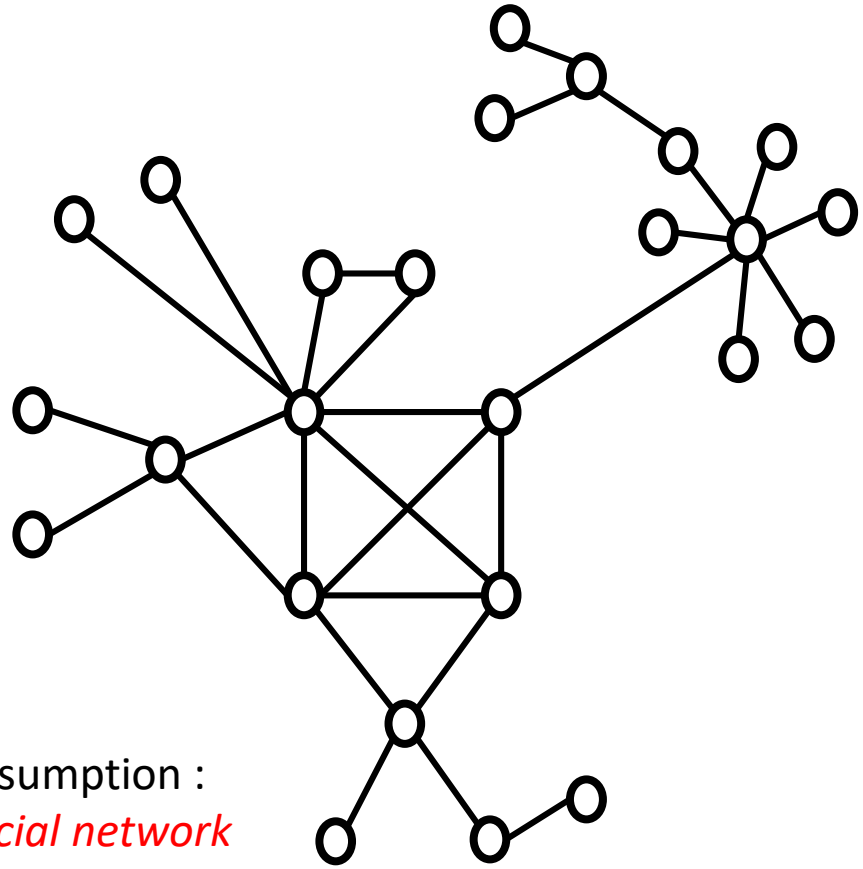
Traffic network

Outline



Importance **Ranking** in complex network

Question : ***Which nodes*** are important for you and ***why***?



Assumption :
social network

Node-based

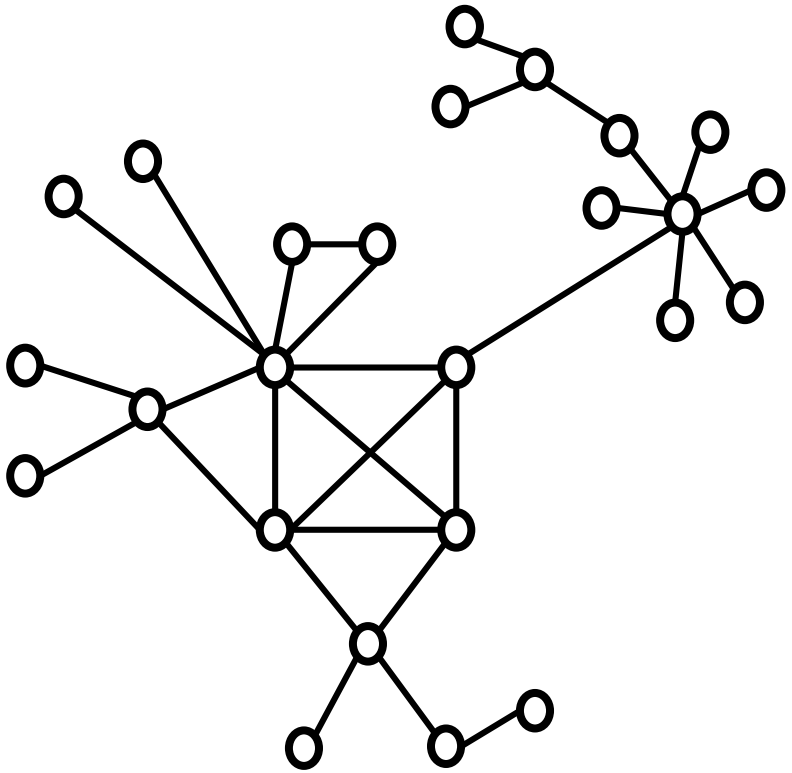
Edge-based

Neighbor-based Ideas(intuitive) :

One-order neighbors(degree) : degree centrality.

Two-order neighbors : semi-local centrality.

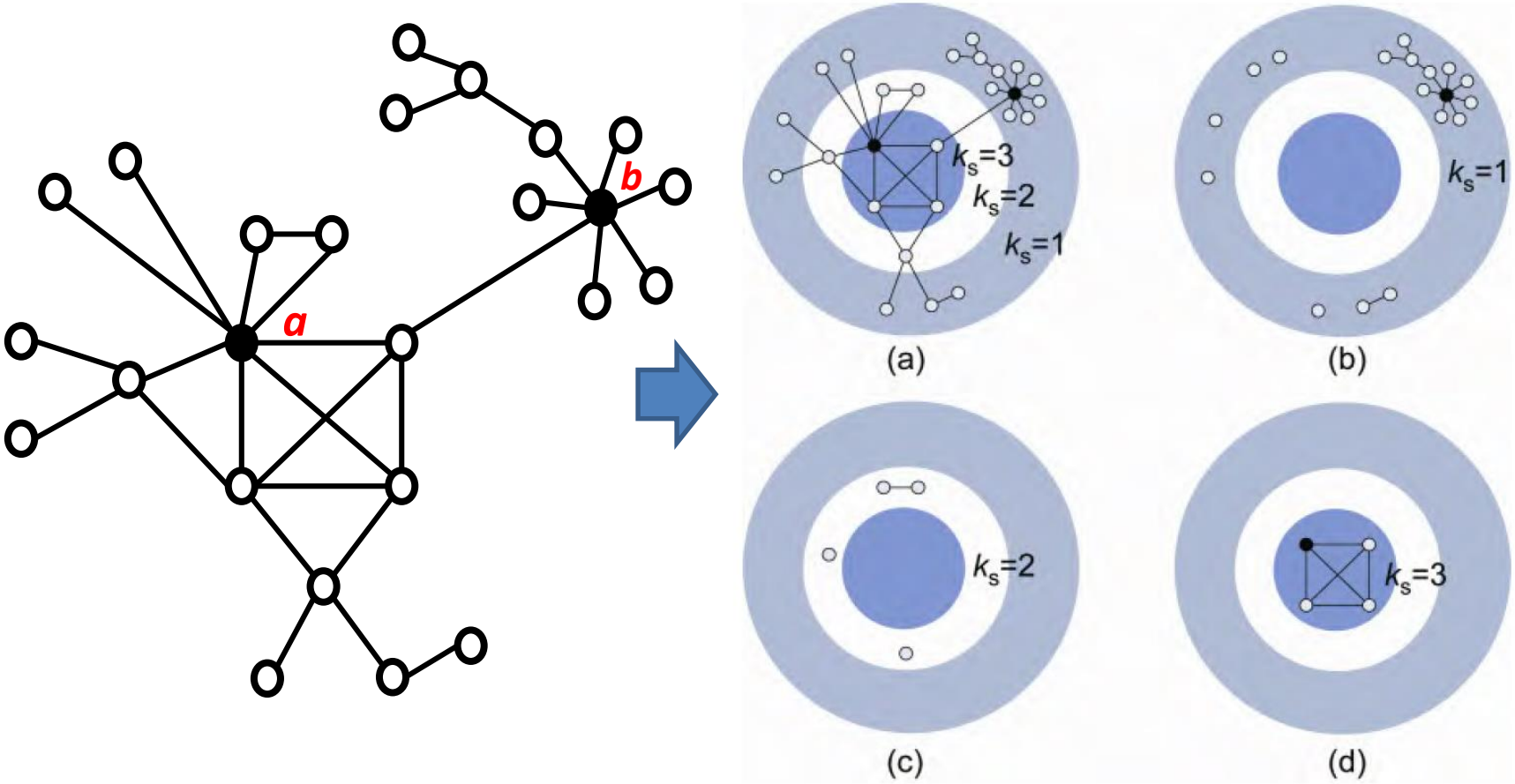
naive



Node with more neighbors is more important ?

Location-based Ideas :

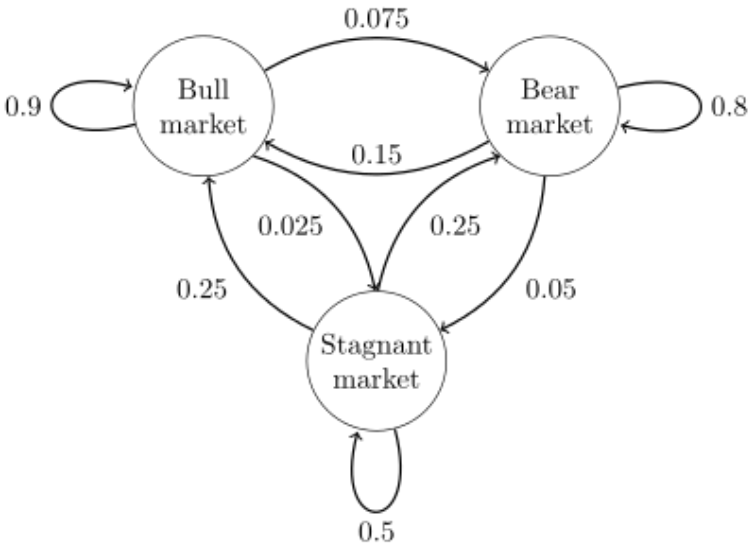
e.g. Node ***a*** and node ***b*** have same degree, which is more important?



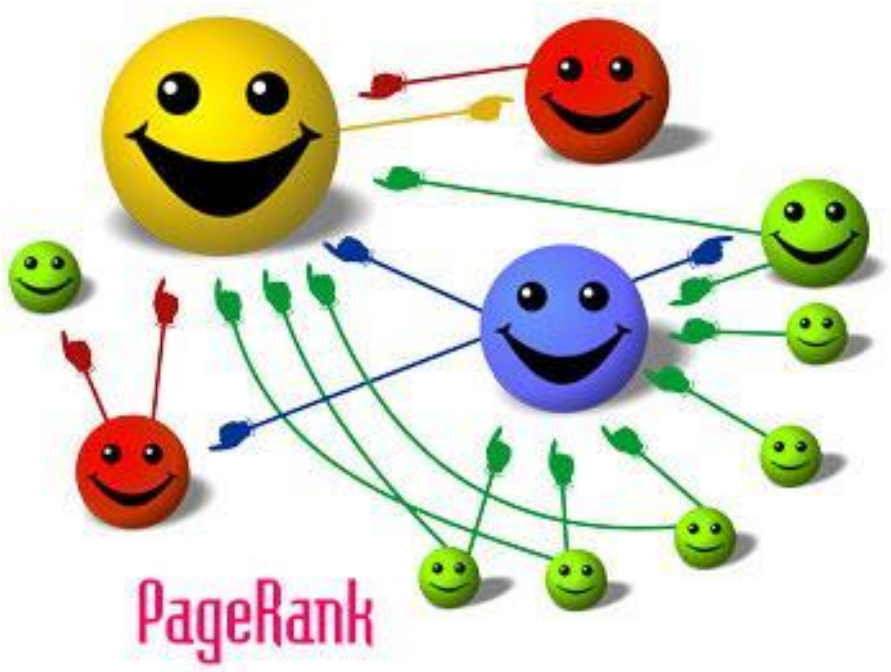
Node ***a*** is more important than node ***b***, since node ***a*** is closer the **network centre**.

Neighbor-based Ideas :

Not only the number of its neighbors matter, but also the importance of these neighbors is critical.



The Markov Chain



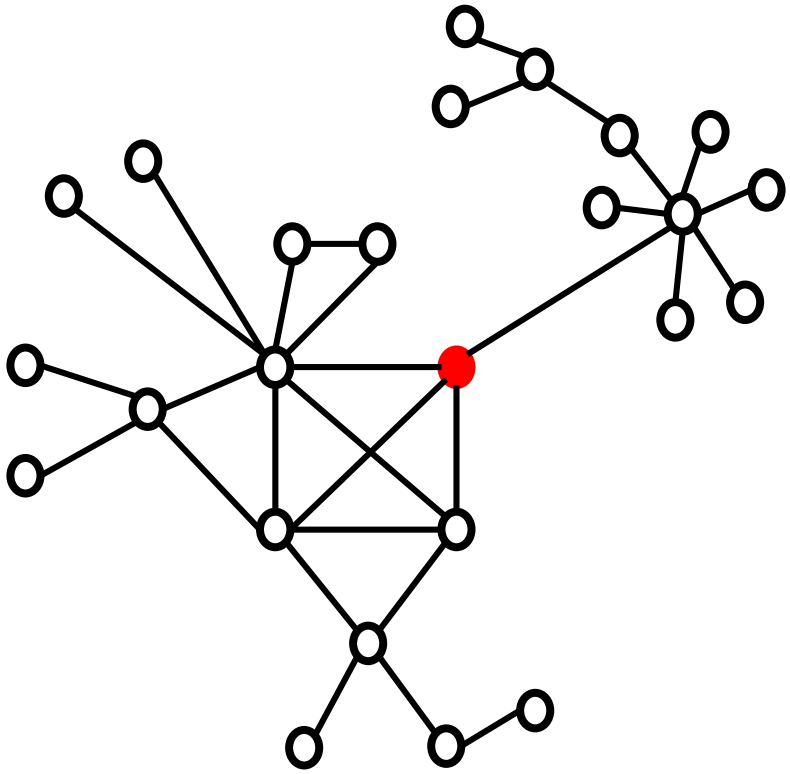
PageRank

Path-based Ideas :

Consider the ability of a node **to control information**.

The shortest path is addressed.

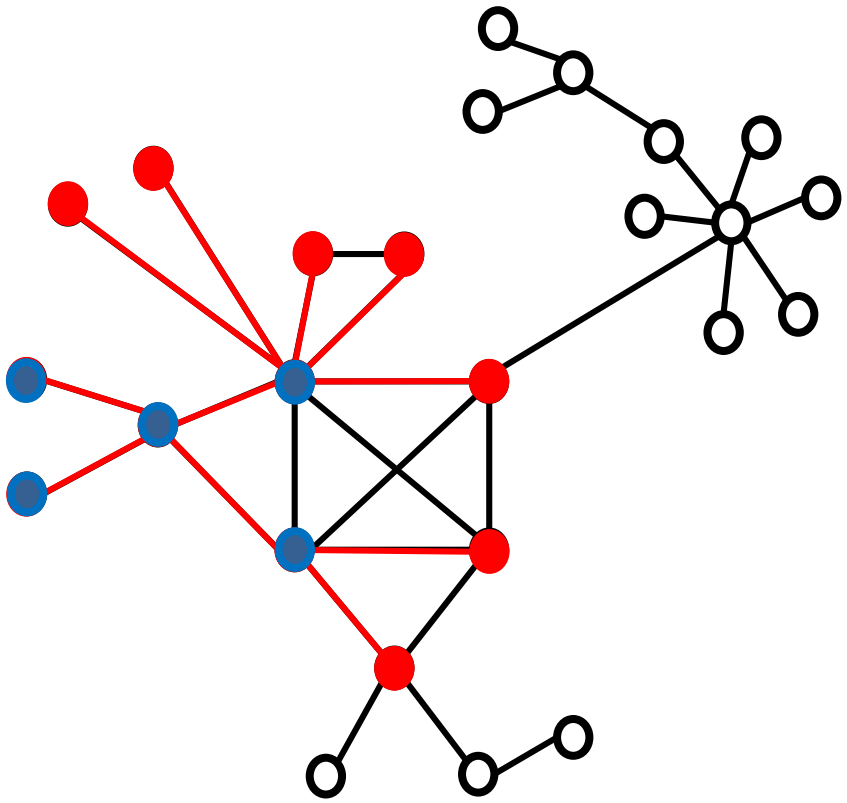
Which node is more important?



- Betweeness centrality
- Closeness centrality
- Eccentricity
- Katz centrality
- …

Diffusion-based Ideas :

The important nodes could diffuse a large scale(SIR model).



Terminal condition :

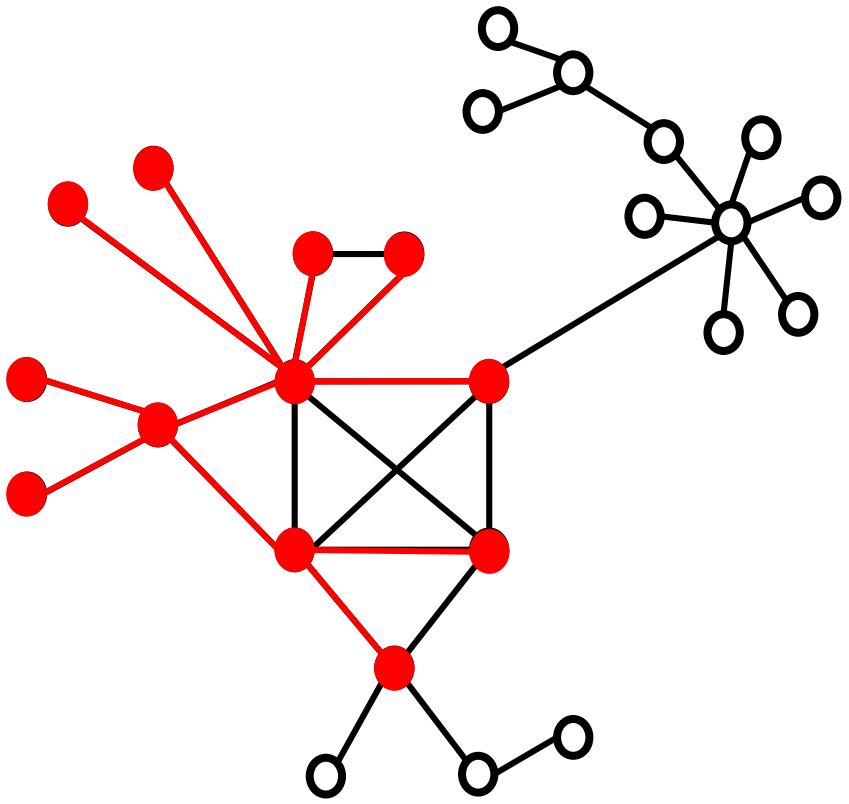
1. All nodes are infected.
2. no one is infected in a step

Infection probability : *r*;
Recover probability : *1*;

Some key conceptions : casecade, infection probability.

Diffusion-based Ideas :

The important nodes could diffuse a large scale(IC model).



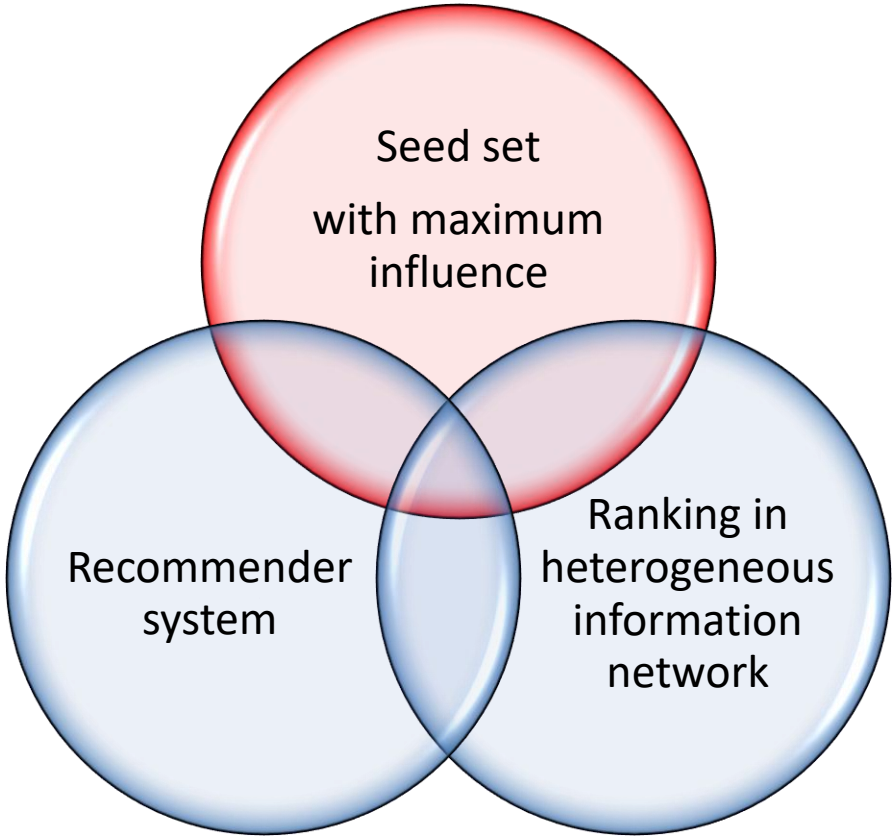
Terminal condition :

1. All nodes are infected.
2. no one is infected in a step

Infection probability : *r*;

Some key conceptions : LT model.

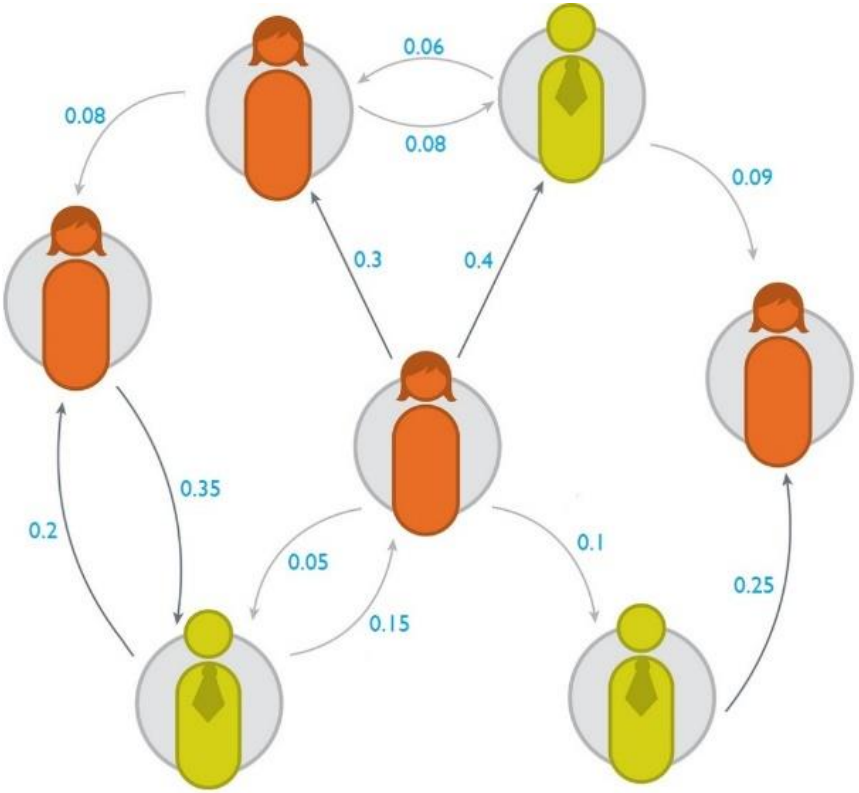
Outline



Importance **Ranking** in complex network

Concept of Influence Maximization

Input:

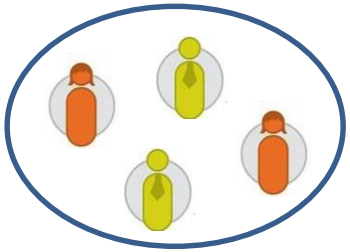


Social graph with influence probabilities of edges

Problem:

- Select *k* individuals such that by activating them, the expected spread of influence is maximized.

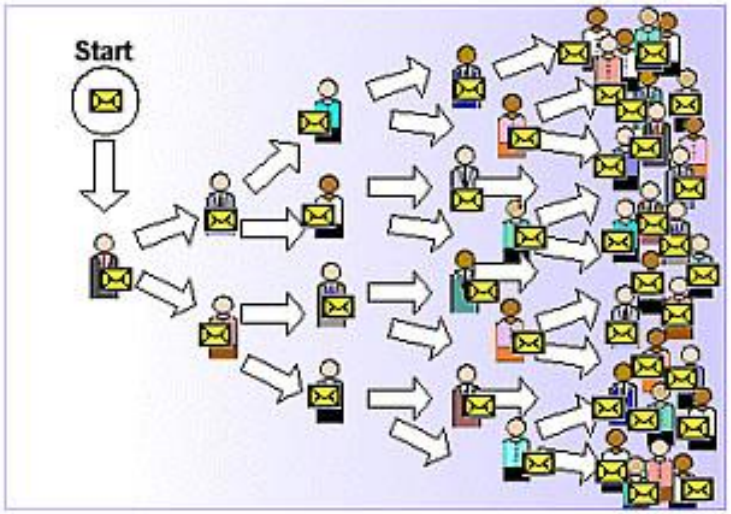
Output:



Seed set of size *k*

Example: Viral Marketing

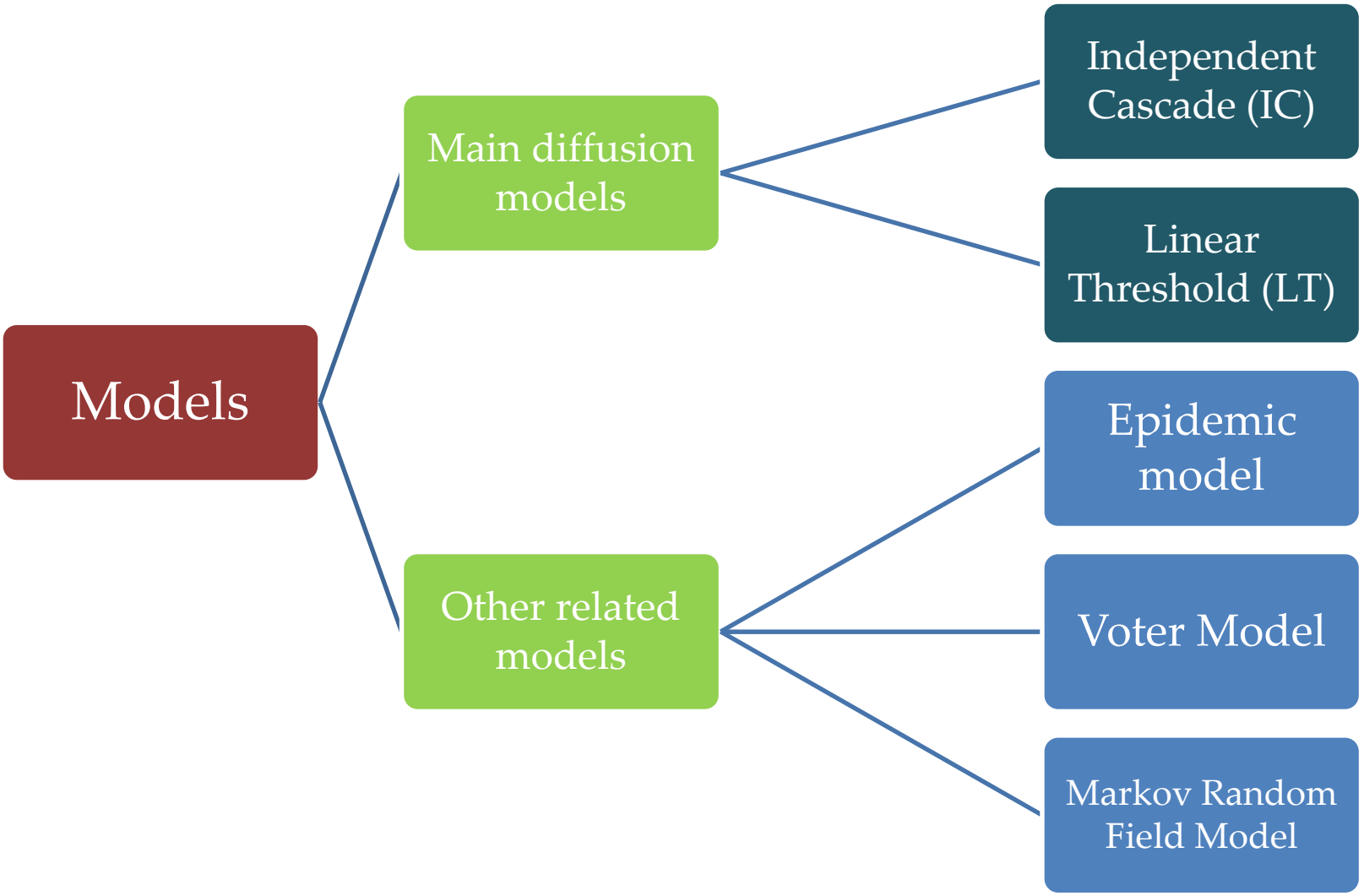
Identify influential customers



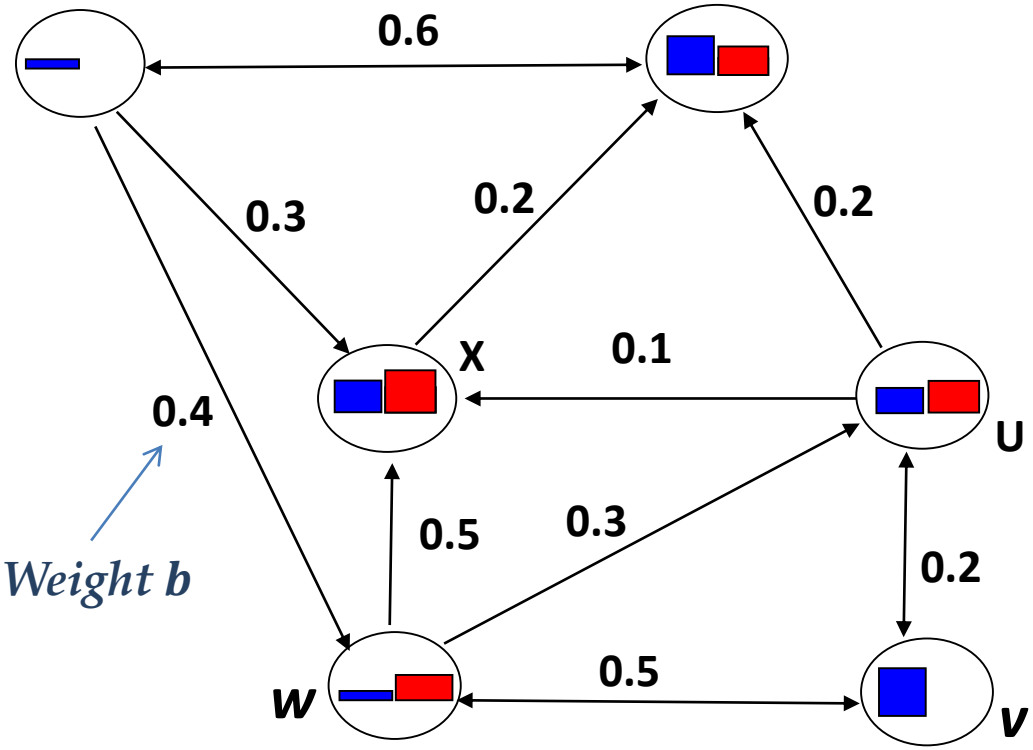
Convince them to adopt the product – Offer discount/free samples

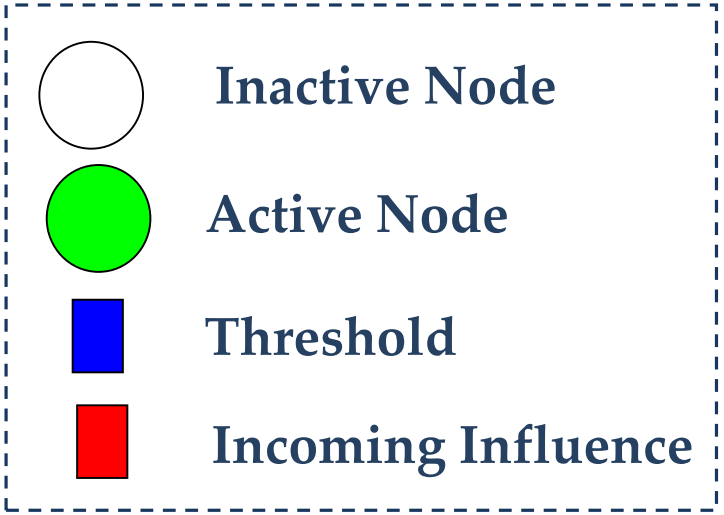
These customers endorse the product among their friends

Two Classical Propagation Models



Linear Threshold Model





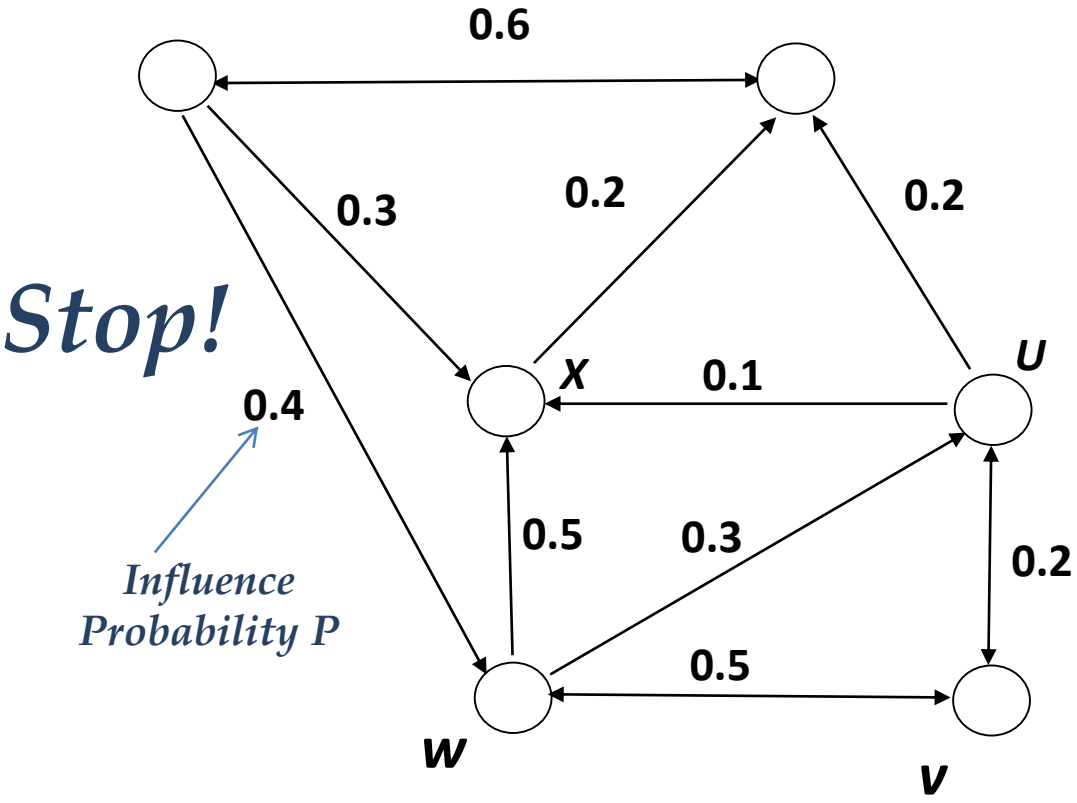
Stop!

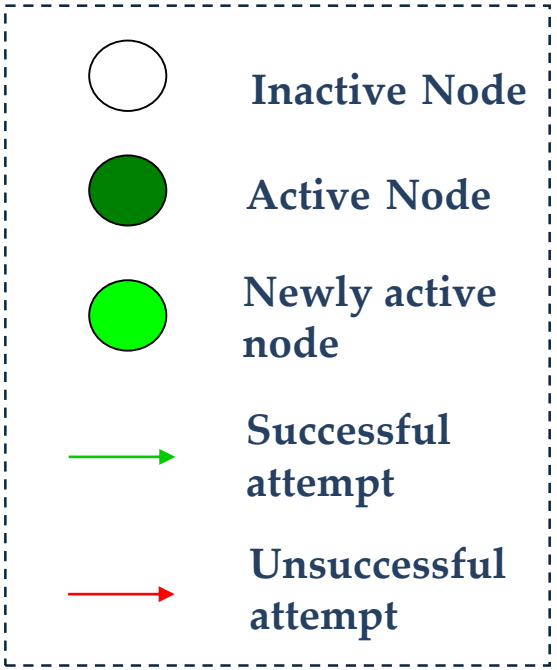
Node become active when the summation of incoming influences is greater or equal than its threshold.

$$\sum b(v, w) \geq \theta$$

Independent Cascade Model

In this case, let's say the global threshold G= 0.2





A node is activated once its incoming influence probability $P(v,w)$ is larger than global threshold G.

$$P(v,w) > G$$

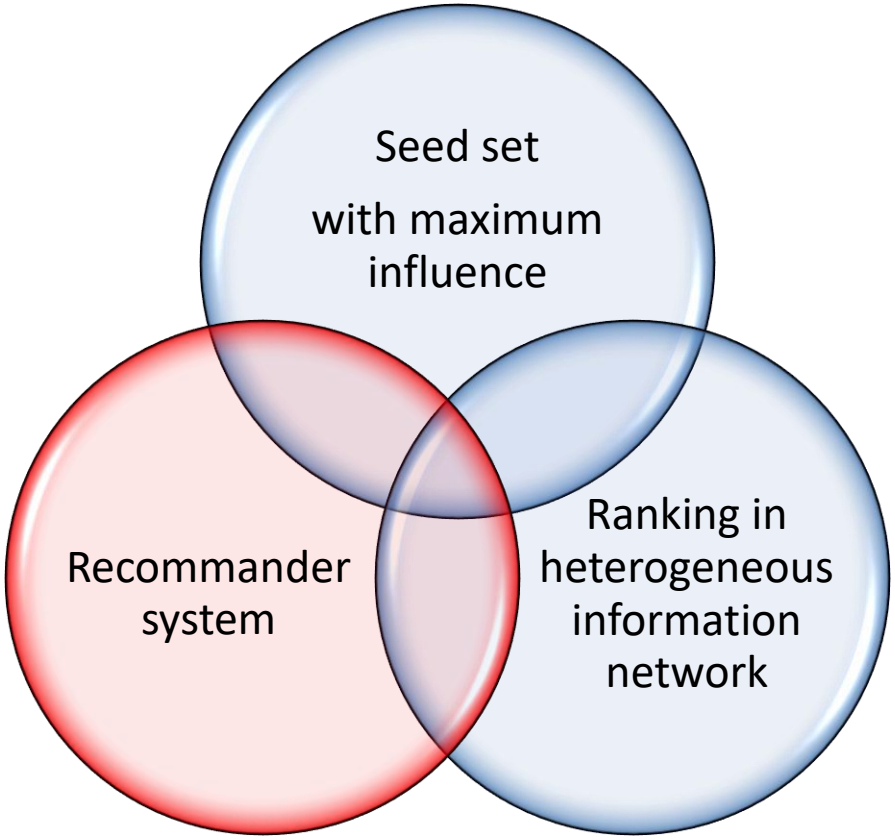
➢Based on the previous researches….

- Influence probability/weight on the edge are based on *assumption*. How can we get the real influence probability/weight?
- In a social network, does influence probability/weight remains constant over time?
- The influence only come from neighbours? Is there any other source of influence?

Issues and Discussion (Cont…)

- Nodes only have positive influence?
- Each node has only 2 influence status (Active and inactive). How about having 3 status (“Inactive”, “Influence & adopt” and “Influence & Impart”)?
- Focus only maximization on a single network based on the given constraint. How about maximize multiple social networks at once?

Outline



Importance **Ranking** in complex network

What's our goal?

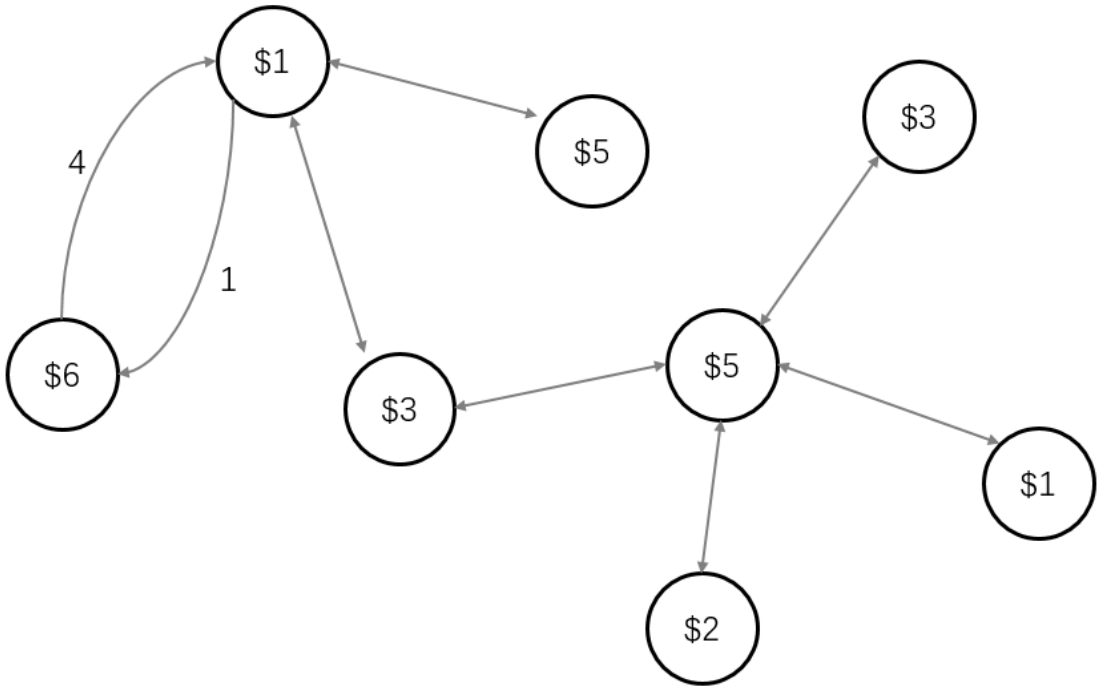
We want to

infer

The **Reserved Price**
For each user

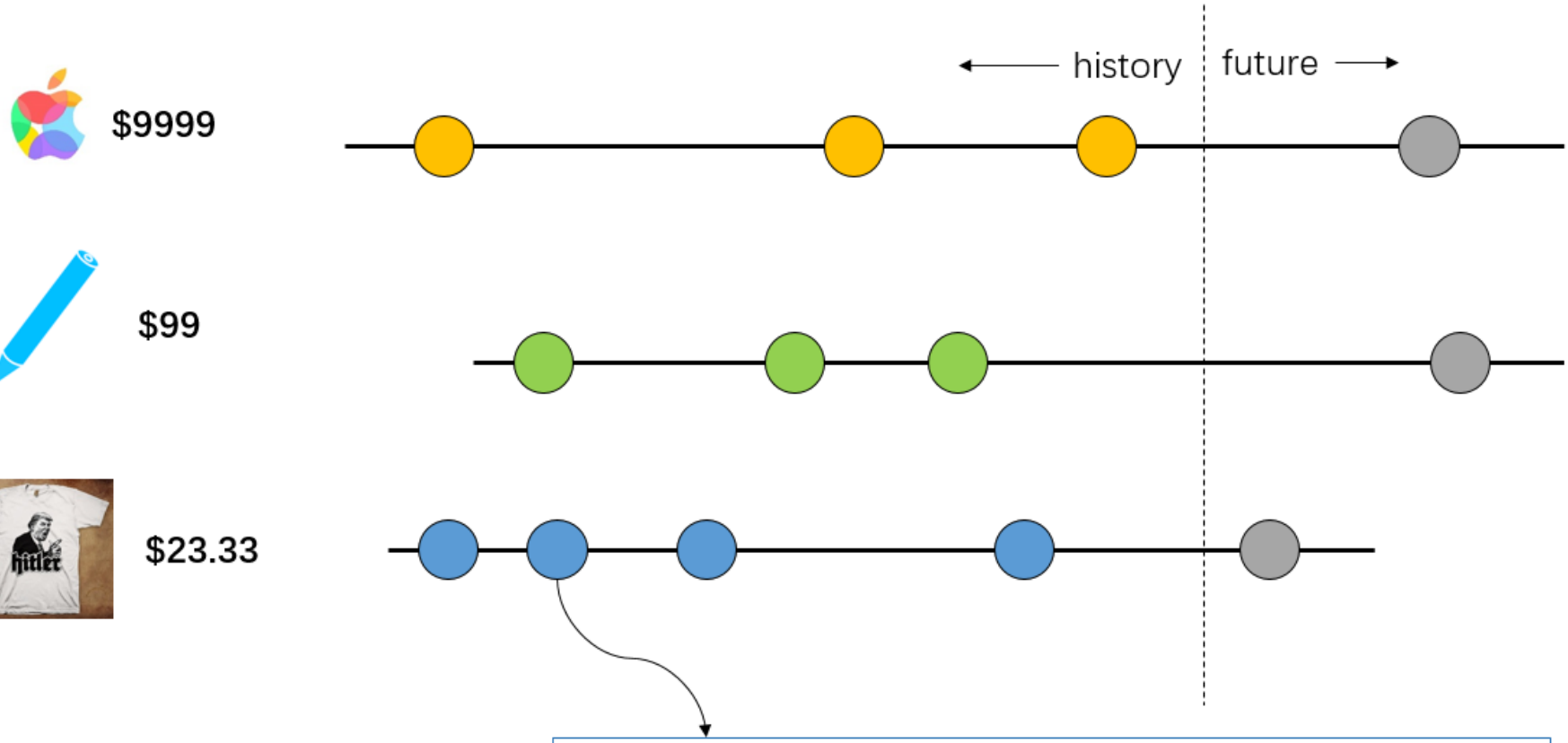
and

The Underlying
Monetary Network



What we have

Item cascades with **their sell price**



Why we need this?

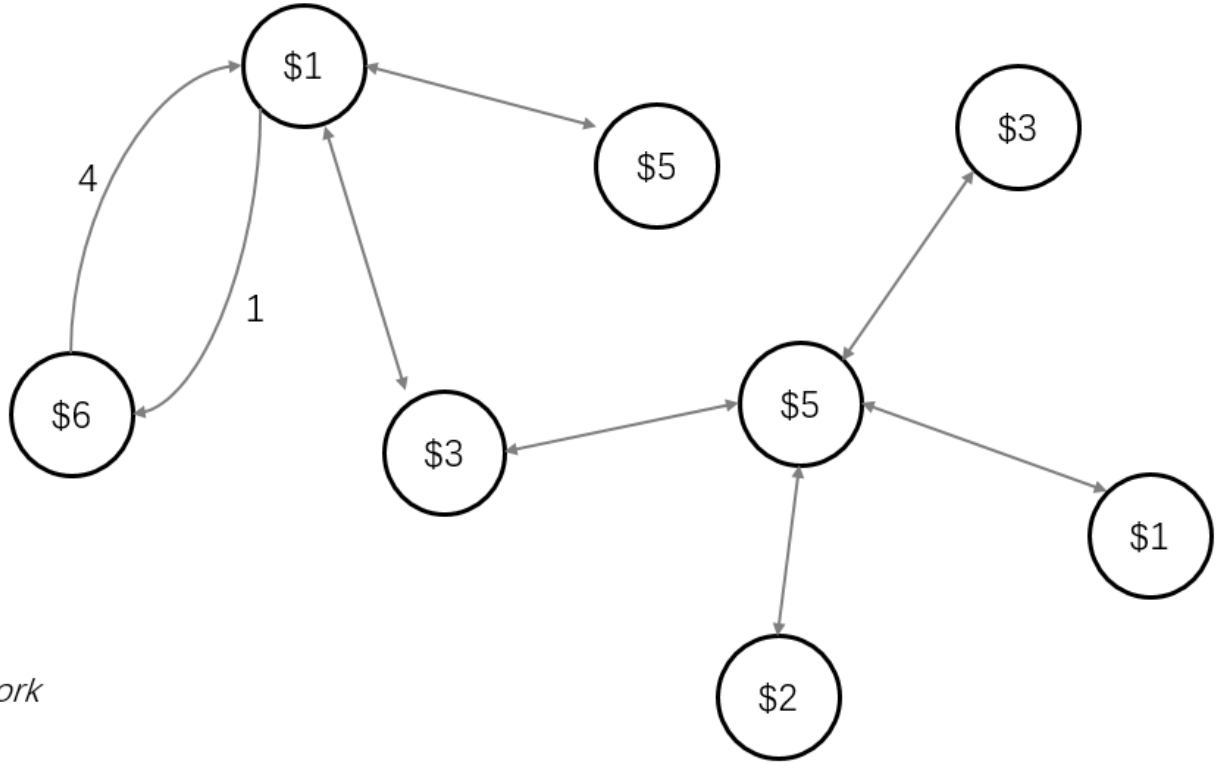
Reserved Price

Monetary Network

No.1

NEED

New metrics
for need
crucial in new RS



No.2

YouTube

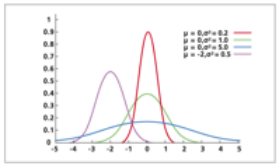
Promotion

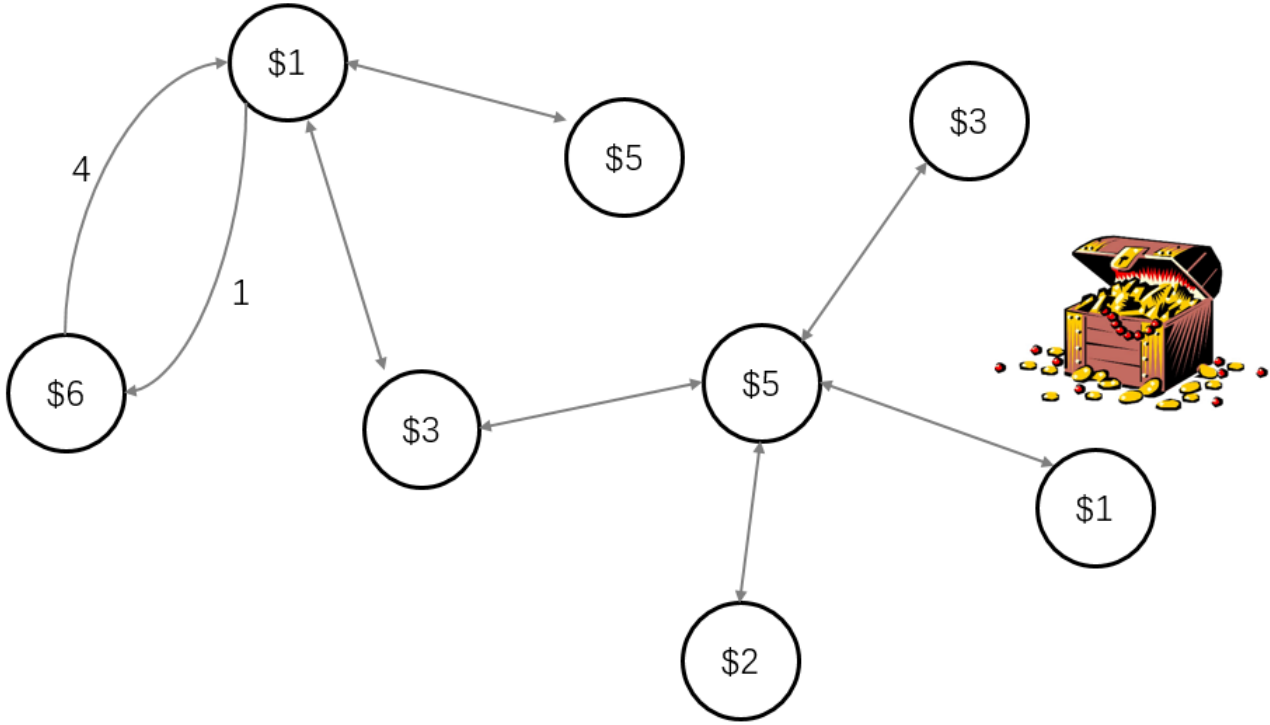
Spread a object
In large social network

How to get this?

Reserved Price

Monetary Network



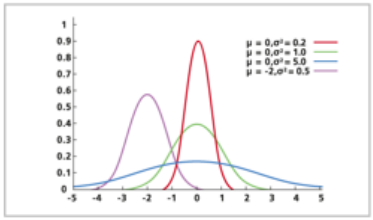


Find the Distribution

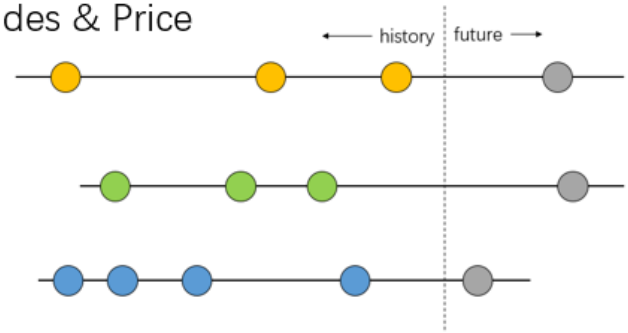
Reserved Price

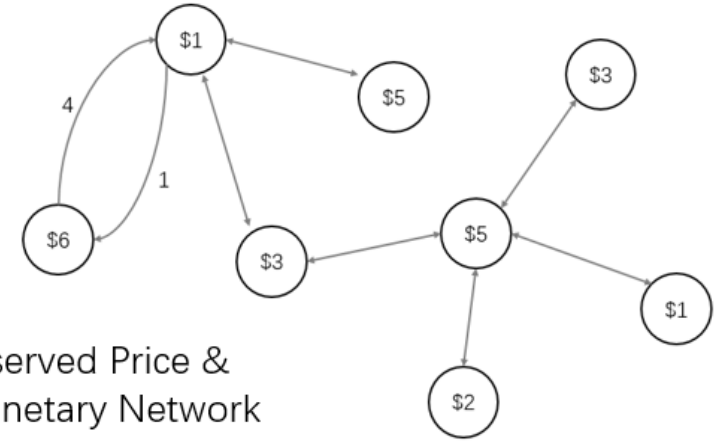
Monetary Network

Model / Dist.



Cascades & Price



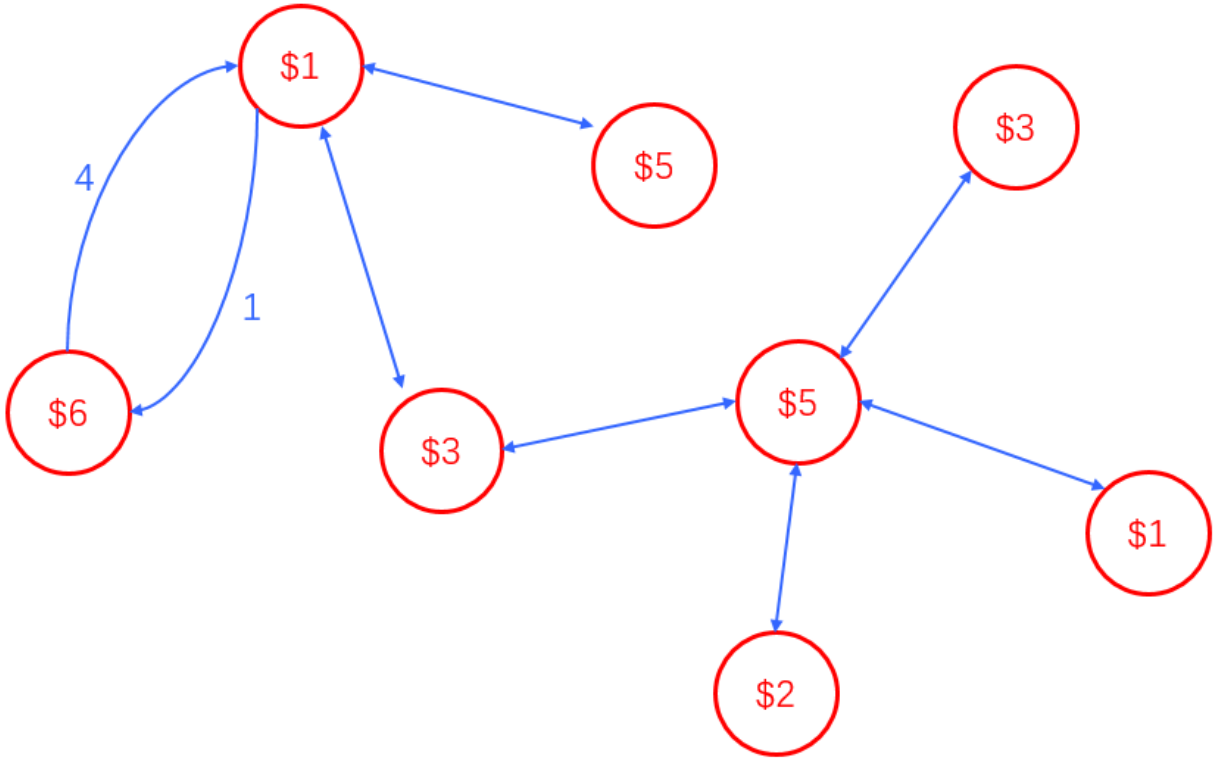


Reserved Price &
Monetary Network

Find the Distribution

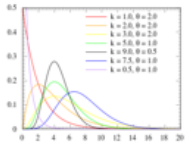
Reserved Price

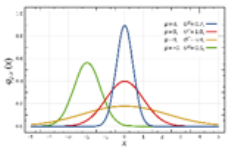
Monetary Network

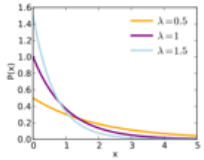


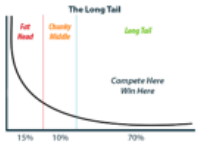
For Nodes

- Choosing the right one
- Parameter level









Find the Distribution

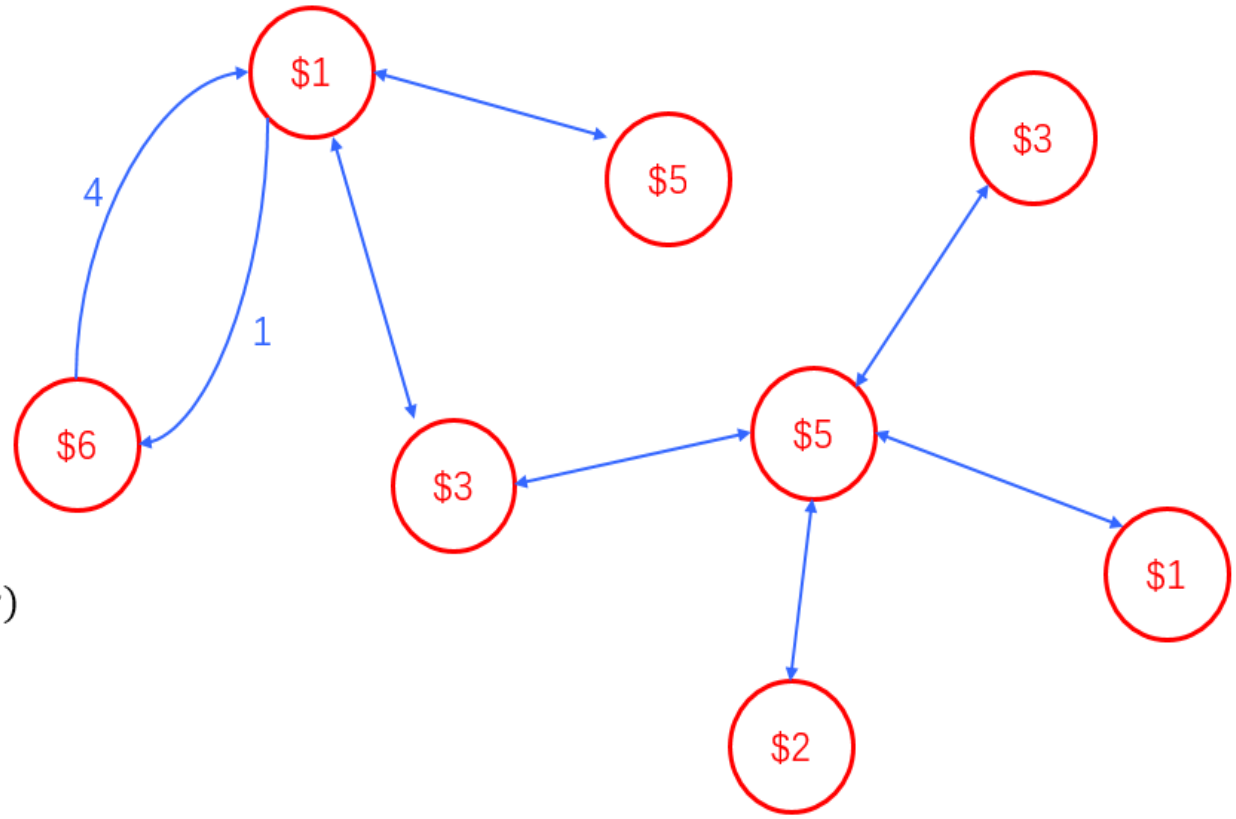
Reserved Price

Monetary Network

For Edges

- Aggregating the information
- Transfer into monetary concept

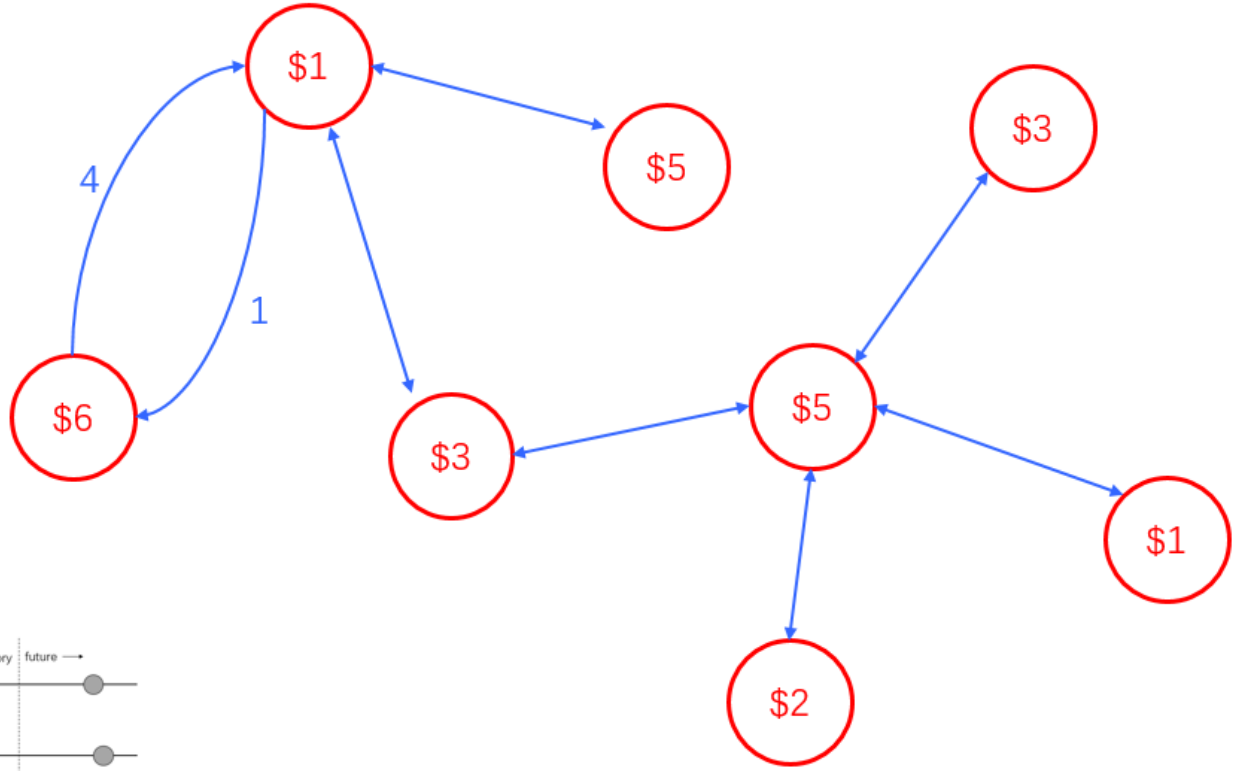
$$y = f(aggregate\ inf.of\ precursors)$$



How to Measure?

Reserved Price

Monetary Network



No

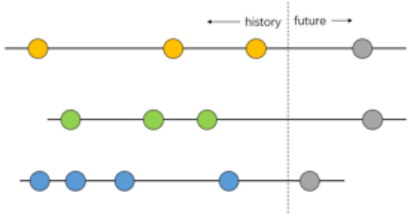
Real reserve price

No

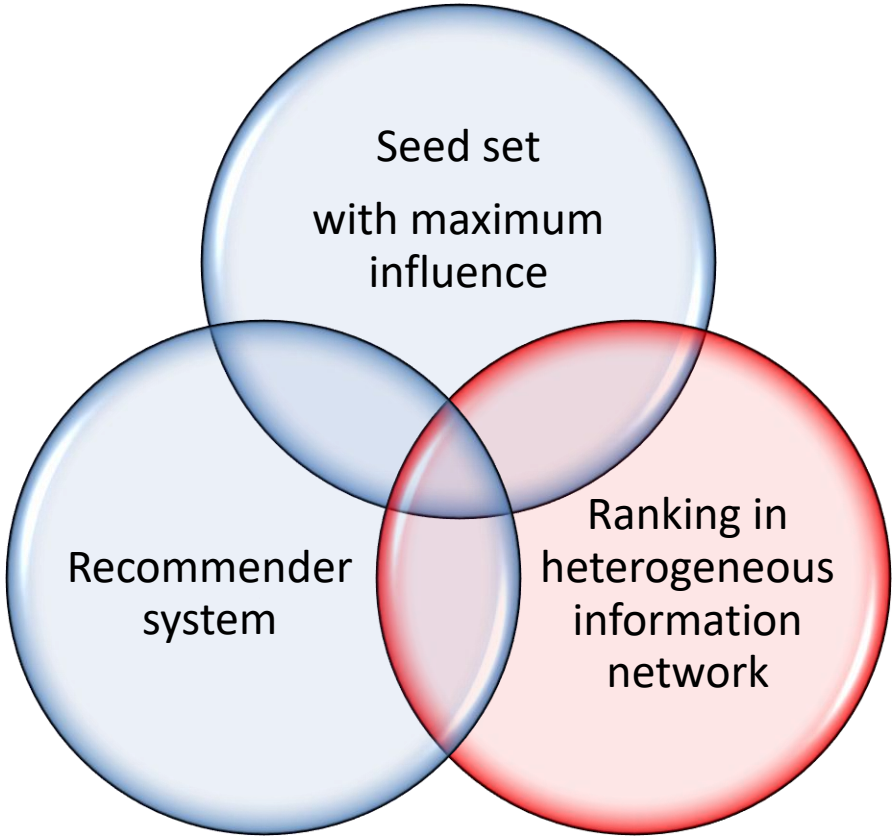
Real monetary network

Yes

Item cascades and sell price

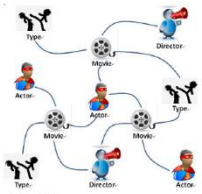


Outline

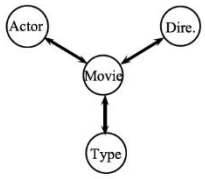


Importance **Ranking** in complex network

Hello, it's me again.
But this time, I
come with **ranking.**



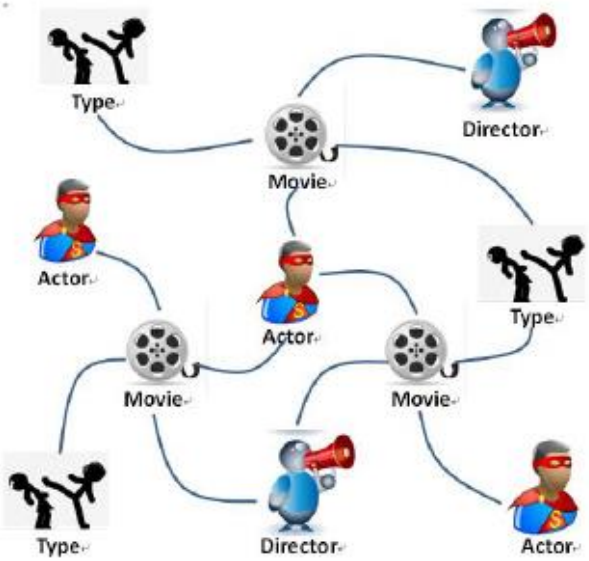
(a) Heterogeneous network of movie data



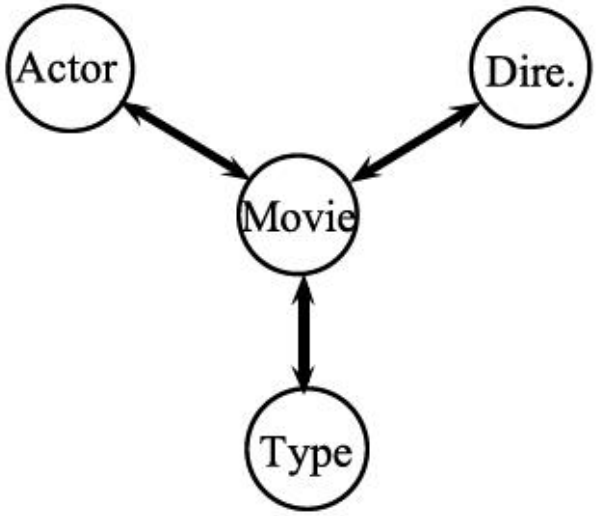
(b) Network schema

HIN:

Note: Heterogeneous information network (*abbrv.* *HIN*)



(a) Heterogeneous network of movie data

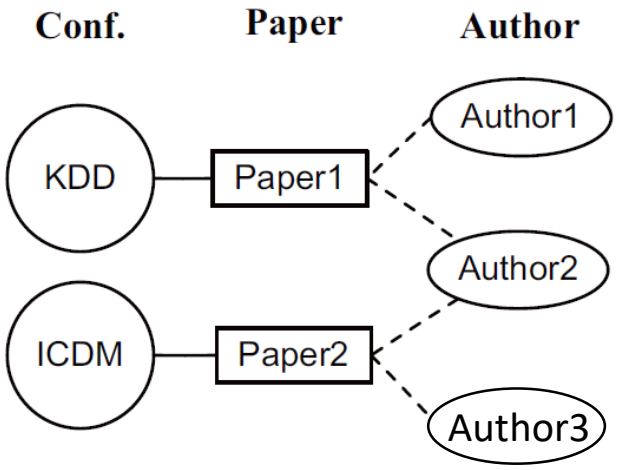


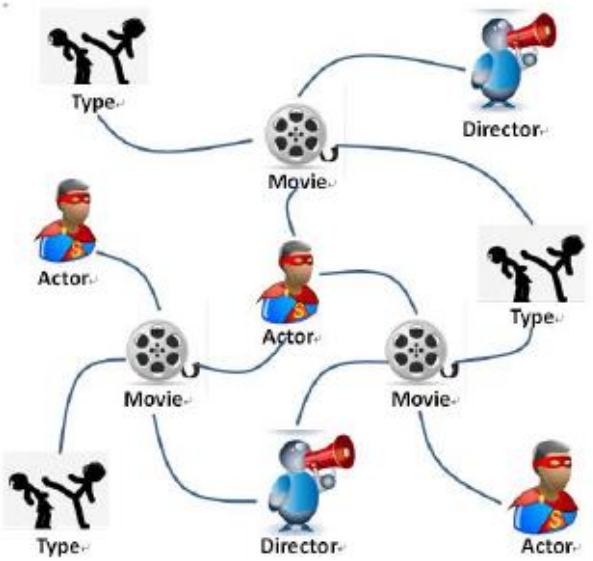
(b) Network schema

- **How to rank objects in *HIN*?**
  - **The ranking of each type has its practical significance.**

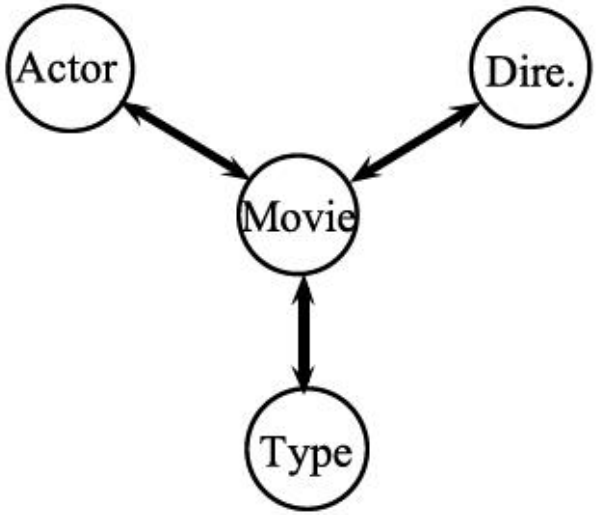
Some answers of Issue 1

- **Simple ranking**
  - Rules: Bigger degree of objects is, higher its ranking score is.
  - Eg: in DBLP author's rank scores is proportional to the numbers of papers accepted by a conference. (conference).
  - However: some people publish lots of papers in junk conferences.
- **Authority ranking**
  - Rules: high ranked type X is more likely related to high ranked typeY if X and Y are connected in the schema of *HIN*.
  - Eg: Jiawei Han & KDD in DBLP.
- **HRank**
- **Any Others ?**





(a) Heterogeneous network of movie data



(b) Network schema

- **How to utilize ranking results for other mining tasks in *HIN*?**

Some answers of issue 2

- Clustering
  - RankClus: iteratively ranking and clustering.
  - NetClus
- Classification
  - RankClass
- Recommender System
  - Ranking items according to similarity or relevance score

Table 1: A set of conferences from two research areas

DB/DM	{SIGMOD, VLDB, PODS, ICDE, ICDT, KDD, ICDM, CIKM, PAKDD, PKDD}
HW/CA	{ASPLOS, ISCA, DAC, MICRO, ICCAD, HPCA, ISLPED, CODES, DATE, VTS }

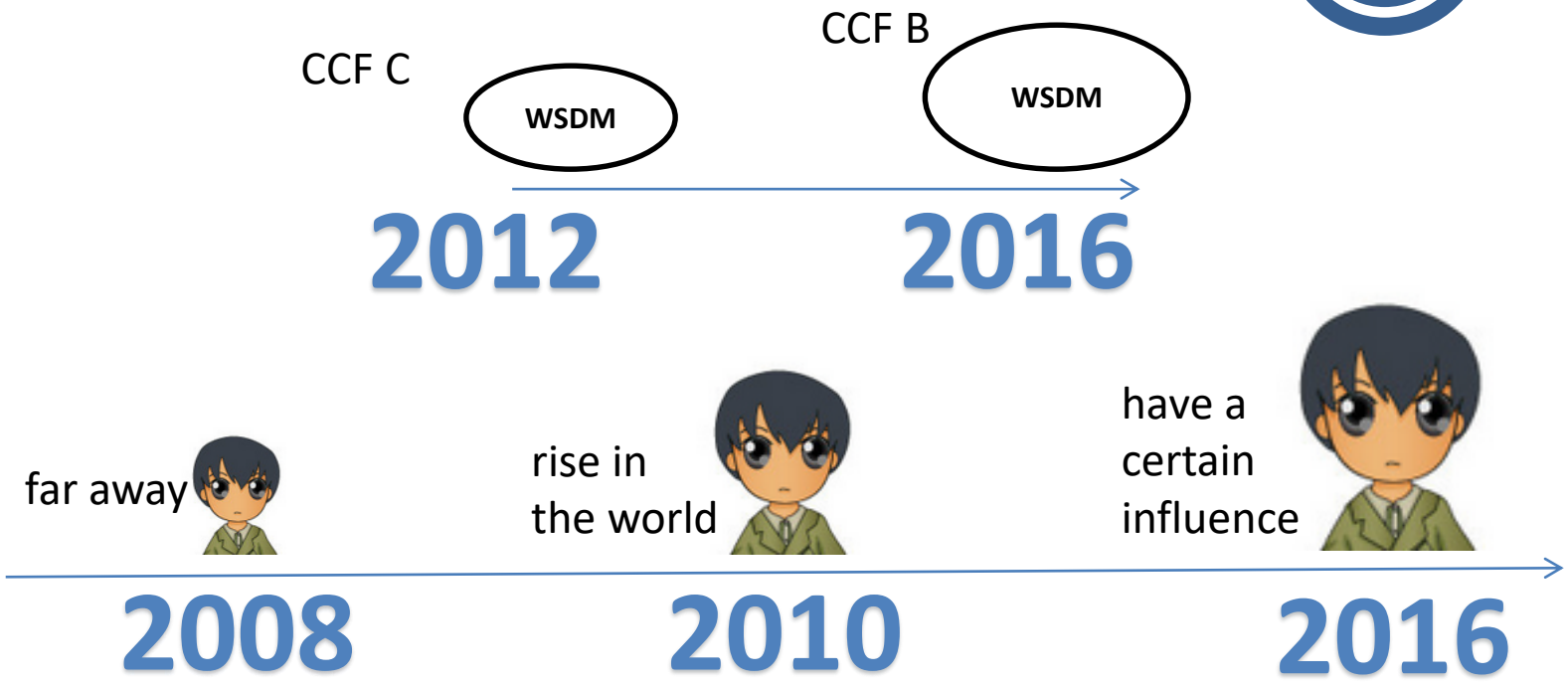
Table 2: Top-10 ranked conferences and authors in the mixed conference set

Rank	Conf.	Rank	Authors
1	DAC	1	Alberto L. Sangiovanni-Vincentelli
2	ICCAD	2	Robert K. Brayton
3	DATE	3	Massoud Pedram
4	ISLPED	4	Miodrag Potkonjak
5	VTS	5	Andrew B. Kahng
6	CODES	6	Kwang-Ting Cheng
7	ISCA	7	Lawrence T. Pileggi
8	VLDB	8	David Blaauw
9	SIGMOD	9	Jason Cong
10	ICDE	10	D. F. Wong

Table 3: Top-10 ranked conferences and authors in DB/DM set

Rank	Conf.	Rank	Authors
1	VLDB	1	H. V. Jagadish
2	SIGMOD	2	Surajit Chaudhuri
3	ICDE	3	Divesh Srivastava
4	PODS	4	Michael Stonebraker
5	KDD	5	Hector Garcia-Molina
6	CIKM	6	Jeffrey F. Naughton
7	ICDM	7	David J. DeWitt
8	PAKDD	8	Jiawei Han
9	ICDT	9	Rakesh Agrawal
10	PKDD	10	Raghu Ramakrishnan

Issue 3: an open problem



– **Time-Aware:** Evolving of Ranking in *HIN*.

- How to recommend to make a conference better(high rank score)?
- How to recommend to make a author more authoritative(high rank score)?

…

Conclusion

- Topic I: Importance Ranking on complex network
  - Node-based
  - Edge-based
- Topic 2: Influence Maximization
  - Learn influence probability/weight
  - Other source of influence
  - Node's status
  - Influence maximization in multiple networks
- Topic 3: Recommender system
  - Learn reserved price
  - Learn monetary network
  - How to aggregate
  - How to interpret
- Topic 4: Ranking in heterogeneous information network
  - How to rank
  - Why to rank
  - **Time-aware ranking**

Q&A